Tema 1: ESTIMACIÓN DE PARÁMETROS Y CONTRASTE DE HIPÓTESIS.

1.1.- Introducción

En la asignatura de primer curso "*Introducción al análisis de datos*" se han estudiado procedimientos para organizar, representar y describir un conjunto de datos, bien mediante la creación de tablas, gráficos, calculando medidas que nos informan de su tendencia central, variabilidad, forma, etc. De forma resumida, estos procedimientos nos proporcionan un conocimiento eficaz de las características de la muestra. En esta asignatura de segundo curso vamos a dar un paso adelante con el objetivo de utilizar esta información para que, mediante la inferencia y el contraste de hipótesis, podamos hacer generalizaciones referidas a la población a partir del análisis descriptivo de una, dos, o más muestras. Este conocimiento siempre será aproximado o, dicho con otras palabras, esta inferencia siempre será probabilística.

En este primer capítulo abordamos los fundamentos de la inferencia estadística, rama de la Estadística que permite realizar afirmaciones sobre una población a partir de los datos obtenidos en alguna de las muestras que se pueden extraer de la misma. En el proceso de inferencia hay que seguir unas pautas para que las afirmaciones que hagamos finalmente referidas a la población, y las correspondientes decisiones que tomemos respecto a ella, sean lo más racionales posibles. En este proceso inferencial se pueden distinguir básicamente los siguientes pasos: extracción de la muestra, medición de la(s) característica(s) objeto de nuestro interés, cálculo del estadístico apropiado en la muestra para inferir el parámetro de la población, y por último evaluación probabilística del error que podemos cometer al realizar dicha inferencia.

De manera resumida, explicaremos los fundamentos teóricos y los aspectos prácticos del proceso de inferencia, repasando un concepto fundamental, sin el cual no es posible comprender cómo se produce la inferencia, y que se conoce como **distribución muestral**, que ya fue tratado sucintamente en el tema 8 de la asignatura de "*Introducción al análisis de datos*", al cual remitimos al estudiante que por algún motivo no ahondó lo suficiente en este concepto. Posteriormente abordamos los procedimientos de estimación de parámetros, así como las propiedades que debe tener un estimador para que cumpla bien su función de estimar el parámetro que se desea conocer en la población[1].

Finalmente, explicamos con cierta amplitud la metodología general del contraste de hipótesis sobre los parámetros de una población, proceso íntimamente relacionado con el proceso de estimación de parámetros. En los epígrafes dedicados a los contrastes de hipótesis, además de la metodología, se tratan aspectos sustantivos de los contrastes tales como los posibles errores que se pueden cometer al hacer una inferencia, y un concepto que está en boga desde los años ochenta del pasado siglo, como es el de la magnitud o *tamaño del efecto*, y que ya es preceptivo referir en cualquier informe de investigación empírica.

En cualquier caso, el estudiante debe saber que la temática que se trata en este texto **asume conocimientos previos tratados en la asignatura de primer curso** de tal forma que se suponen adquiridos los conceptos básicos de análisis descriptivo de los datos, probabilidad, el cálculo de las probabilidades en las distribuciones discretas y continuas e, íntimamente relacionadas con éstas últimas, el concepto de distribución muestral. Adquiridos estos conceptos a los que nos hemos referido, en este primer tema marcamos los siguientes objetivos.

[1] El concepto de *parámetro* se explica detenidamente en el tema 9 sobre contrastes no paramétricos.

1.2.- Objetivos:

- Conocer cómo es la distribución muestral de los estadísticos media, varianza y proporción.
- Calcular intervalos de confianza de los parámetros poblacionales media, varianza y proporción.
- Calcular el tamaño de la muestra en función de la precisión de la estimación deseada.
- Comprender e interpretar la lógica de la metodología del contraste de hipótesis.
- Reconocer e identificar los errores y riesgos de todo contraste de hipótesis.

1.3.- Distribuciones muestrales

La inferencia estadística es una forma de razonamiento que va de lo concreto a lo general. El investigador, para confirmar o refutar las hipótesis teóricas que maneja, extrae una muestra representativa de la población objeto de estudio y sobre ella realiza las mediciones de las características relevantes para su investigación. Para cada característica evaluada se obtiene uno, o más, valores numéricos que se conocen como **estadísticos**, los cuales pueden ser cualquiera de los estudiados en la asignatura de primer curso (medidas de tendencia central, de posición, de variabilidad, de asimetría, de relación, de regresión, etc.). Y es a partir de los diversos estadísticos obtenidos en la muestra (lo concreto) que tiene que realizar afirmaciones sobre los valores de los **parámetros** de la población (lo general). Pero ¿cómo se realiza ese salto de lo concreto a lo general? Para entender este proceso es preciso, previamente, recordar lo que se conoce como distribución muestral, y para ello hay que situarse en un plano hipotético en el que pudiéramos trabajar con todas las posibles muestras del mismo tamaño, n, que se pueden extraer de una población de tamaño N (siendo, obviamente, $N > n$).

Razonando en este escenario hipotético en el que pudiéramos extraer todas las muestras de la población, en cada una de estas muestras se realizaría la medición de la o las variables de interés y se obtendría un estadístico (media, proporción, varianza, correlación, etc.) cuyo valor será diferente (o igual) al obtenido en cualquiera de las otras posibles muestras ya que, obviamente, depende de los datos que la componen. Es decir, el estadístico obtenido en cada una de las distintas muestras se comporta como una variable aleatoria, y sus diferentes valores forman una distribución de probabilidad que recibe el nombre de ***distribución muestral***. Como en toda distribución de probabilidad, también de la distribución muestral de uno de estos estadísticos obtenido para todas las muestras posibles, podemos obtener su media y su desviación típica. Esta última, al estar referida a la distribución muestral de un estadístico, recibe el nombre de **error típico** del estadístico.

De modo que el concepto de distribución muestral hay que distinguirlo de otros tipos de distribuciones, como son, la **distribución poblacional,** que se refiere a la distribución de los datos individuales en la población, y la **distribución en la muestra** que es la distribución de los datos individuales que constituyen la muestra.

Una vez que hemos repasado el concepto de distribución muestral, vamos a abordar de manera muy resumida cómo son las distribuciones muestrales de tres estadísticos ampliamente utilizados en la investigación social: la media, la proporción y la varianza, recordando que las dos primeras ya fueron tratadas en el curso anterior. Veremos cómo la forma que adopta la distribución muestral depende, entre otras cosas, de la forma que adopte la distribución poblacional.

1.3.1. Distribución muestral de la media

Consideremos una población formada por todos los estudiantes universitarios de una determinada comunidad de los que podemos conocer, a partir de sus datos de la matrícula, su edad. A partir de estos datos podemos calcular su edad media y la varianza de esta misma variable (edad), valores que representamos por $\mu\ y\ \sigma^2$, respectivamente (si dispusiéramos de más de una variable, sería recomendable indicar, mediante subíndices, a qué variable se corresponde cada media y varianza, de tal forma que en este caso podríamos indicarlo como $\mu_{edad}\ y\ \sigma^2_{edad}$). De esta población podemos extraer una muestra de, por ejemplo, 100 estudiantes y calcular su media y desviación típica que representamos por $\bar{Y}\ y\ S_Y$ si representamos la variable con la letra Y, o por $\bar{X}\ y\ S_X$ si la representamos con la letra X. Pero esta muestra no es la única posible. Se pueden extraer muchas otras muestras diferentes, todas ellas del mismo tamaño (n=100), y en cada una de ellas calcular su media y desviación típica que pueden variar de una muestra a otra, de tal manera que con las puntuaciones de todas las medias obtenidas en estas distintas muestras (véase Figura 1.1) se origina otra distribución que se llama ***distribución muestral de la media***. Con el mismo procedimiento se obtendría la distribución muestral de la desviación típica o de cualquier otro estadístico, como la proporción, la correlación de Pearson, etc. y corresponde a la distribución de probabilidad de un estadístico que se obtiene al calcularlo **en todas** las posibles muestras del mismo tipo y tamaño, n, extraídas de una población de tamaño N.

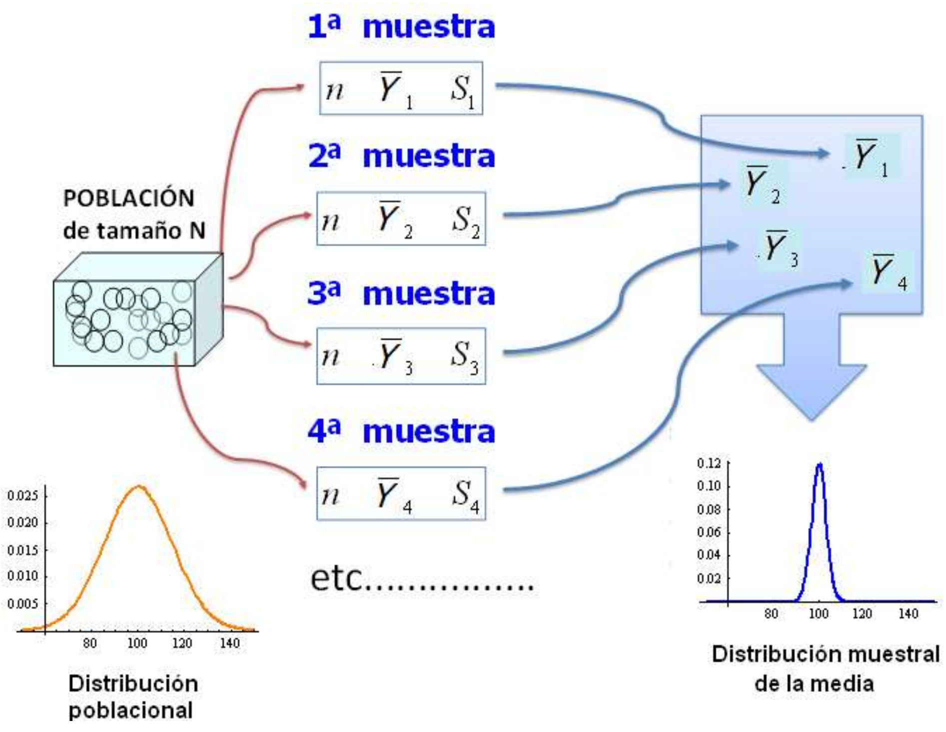


Figura 1.1: *Proceso de construcción de la distribución muestral para el estadístico media. A la izquierda aparece la representación de una variable en una población de tamaño N. Esta variable es normal con media 100 y desviación típica 15. A la derecha se muestra la distribución muestral teórica del estadístico Media calculado en todas las muestras posibles de tamaño n. Obsérvese que ambas distribuciones (la poblacional y la muestral) tienen la misma media pero la distribución muestral tiene una variabilidad muy inferior a la variabilidad de la distribución poblacional.*

Como se estudió en el tema 8 de la asignatura "*Introducción al análisis de datos*", podemos suponer que la distribución muestral de la media es normal, o se aproxima suficientemente a la normalidad, cuando se cumple **al menos una** de las siguientes condiciones:

- La variable en la población se distribuye normalmente.
- La variable en la población NO se distribuye normalmente, pero el tamaño de la muestra es igual o superior a 30 observaciones (Teorema Central del Límite).

En la asignatura "Introducción al Análisis de Datos", estudiamos que si se desconoce la forma de la distribución poblacional de la variable, la forma de la distribución muestral de la media depende del tamaño de la muestra. **El Teorema Central del Límite (TCL)** establece que sin importar la forma de la distribución poblacional, la distribución muestral de la media se aproximará a la normal a medida que aumenta el tamaño de la muestra. Y el tamaño que debe tener la muestra para que la distribución muestral se considere normal depende de la forma que tenga la distribución poblacional, de tal forma que cuanto más se aleje ésta de la distribución normal mayor tendrá que ser el tamaño de la muestra. Por otro lado, si asumimos que la mayoría de las variables que se utilizan en las ciencias sociales no se alejan en exceso de la distribución normal, vamos a considerar que una muestra es grande a partir de $n \geq 30$.

Cuando realizamos inferencia estadística sobre la media aritmética, **siempre ha de cumplirse al menos una de las dos condiciones descritas**, pero procederemos de forma diferente en función de si la varianza poblacional es conocida o desconocida.

1.- Así pues, si **conocemos la desviación típica poblacional** σ, y podemos asumir que la variable en la población se distribuye normalmente, o bien $n \geq 30$, entonces consideramos que la distribución muestral del estadístico media es también normal, cuya media y desviación típica (o error típico de la media) son, respectivamente:

$$\mu_{\bar{Y}} = \mu$$

$$\sigma_{\bar{Y}} = {}^{\sigma}/_{\sqrt{n}}$$

Obsérvese que para diferenciar los parámetros poblacionales ($\mu\ y\ \sigma$) de los parámetros de la distribución muestral de la media ($\mu_{\bar{Y}}$ y $\sigma_{\bar{Y}}$) hemos incluido en esta última un subíndice que señala el estadístico sobre el que se ha calculado la distribución muestral.

Obviamente, si tipificamos el valor del estadístico media $\bar{Y}$ que se distribuye normalmente, obtenemos la variable Z:

$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu}{{}^{\sigma}/_{\sqrt{n}}}$$

cuya distribución será normal, $N(0,1)$, lo cual permite conocer mediante las tablas de la curva normal la probabilidad asociada a cada valor del estadístico $\bar{Y}$ en la distribución muestral, o la distancia, en términos probabilísticos, desde la media de una muestra concreta, $\bar{Y}$, a la media de la población μ (que coincide con la media de la distribución muestral, $\mu_{\bar{Y}}$).

2.- No obstante, si, como suele ser habitual en la práctica investigadora, **se desconoce la varianza de la variable en la población**, pero podemos asumir que la distribución poblacional es normal o bien $n \geq 30$, los estudios realizados por W.S. Gosset al final del siglo XIX demostraron que en estas circunstancias la distribución muestral de la media es una distribución diferente de la normal, que se conoce con el nombre de **distribución t de Student** ([2]). Bajo estas condiciones Gosset demostró que la variable:

$$T = \frac{\bar{Y} - \mu}{S_{n-1}/\sqrt{n}} = \frac{\bar{Y} - \mu}{S_n/\sqrt{n-1}}$$

sigue el modelo *t* de Student con *n-1* grados de libertad, donde S_{n-1} y S_n son, respectivamente, la cuasidesviación típica y la desviación típica de la muestra, ambas vistas el curso anterior.

Recuerde el lector que, en la asignatura de Introducción al Análisis de Datos, se describían las características de las distribuciones Z y t, indicando que la distribución normal estándar es simétrica con media 0 y varianza 1 mientras que la distribución t de Student es también simétrica con media cero pero varianza igual a $n/(n-2)$. Es fácil deducir, por tanto, que a medida que aumenta el valor de *n*, la varianza de la distribución t se va aproximando a 1, y por tanto, la distribución t se irá aproximando a la normal de puntuaciones Z. En las tablas que manejamos en este curso, podemos consultar valores para distribuciones t de Student hasta 100 grados de libertad. Podemos comprobar que para dichos grados de libertad los valores que nos ofrece la tabla son muy parecidos a los de la curva normal tipificados, por lo que, **cuando los grados de libertad sean superiores a 100, podemos utilizar los valores de la tabla de curva normal como una aproximación a los valores de la t de Student**.

Veamos un ejemplo:

Ejemplo 1.1: Supongamos que en un determinado Estado la población de escolares es evaluada sobre conocimientos matemáticos básicos. Las puntuaciones en la población se distribuyen normalmente con media $\mu = 50$ y desviación típica $\sigma = 12$. Si de esta población se extrae una muestra aleatoria de 121 sujetos: ¿cuál es la probabilidad de obtener una media de 52 puntos o superior?; ¿cuál es la probabilidad de obtener una media que esté comprendida entre 48 y 51 puntos?

Partimos de una distribución poblacional normalmente distribuida con media 50 y desviación típica 12 por lo que la distribución muestral de la media es también normal (con media igual a 50 y desviación típica igual a: $12/\sqrt{121}$).

[2] Este fue el pseudónimo que tuvo que utilizar Gosset para poder publicar sus investigaciones sobre la distribución t ya que su contrato laboral con la cervecera Guinness le impedía publicar con su nombre verdadero.

Bajo estas condiciones, calculamos la puntuación típica correspondiente al valor 52, en la distribución muestral de medias:

$$z = \frac{52 - 50}{12/\sqrt{121}} = 1{,}83$$

y de acuerdo a la distribución normal tipificada, la probabilidad de obtener puntuaciones típicas de 1,83 o superior, o lo que es igual, de obtener una media de 52 puntos o superior es igual a 0,0336. Es decir, es poco probable encontrar de esa población una muestra de 121 elementos y que tenga 52 puntos de media o superior. En la Figura 1.2(a) se representa el área correspondiente a esta probabilidad.

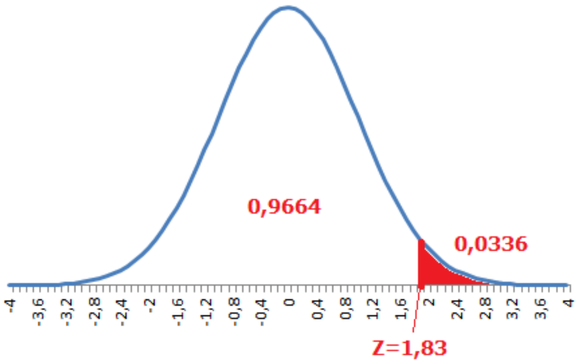


Figura 1.2 (a). *Probabilidad de encontrar valores iguales o mayores de 52 en la distribución muestral de media 50 y desviación típica:* $12/121$

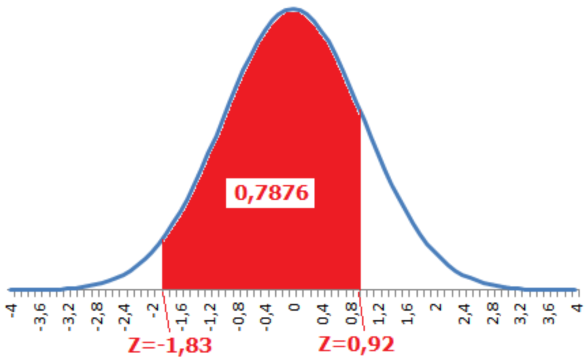


Figura 1.2 (b). *Probabilidad de encontrar valores comprendidos entre 48 y 51 en la distribución muestral de media 50 y desviación típica* $12/121$

Para la segunda cuestión hay que calcular las puntuaciones típicas correspondientes a 48 y 51, y determinar la probabilidad asociada a sus correspondientes valores z.

$$z = \frac{48 - 50}{12/\sqrt{121}} = -1{,}83 \qquad z = \frac{51 - 50}{12/\sqrt{121}} = 0{,}92$$

De acuerdo a la distribución normal tipificada la probabilidad, comprendida entre estas dos puntuaciones que se representan en la Figura 1.2 (b), es:

$$P(-1{,}83 \leq Z \leq 0{,}92) = P(Z \leq 0{,}92) - P(Z \leq -1{,}83) = 0{,}8212 - 0{,}0336 = 0{,}7876$$

1.3.2. Distribución muestral de la proporción

En el ámbito de las Ciencias Sociales es habitual dirigir nuestra atención a situaciones en las que no estamos interesados en la media de la muestra sino que queremos investigar la **proporción** de personas que votarán a un determinado partido político, que presentan un determinado síntoma, o que, en definitiva, cumplen una determinada condición a la que genéricamente llamaremos "éxito". En estas situaciones tenemos que apoyarnos en la distribución muestral de la proporción, la cual se genera con la misma lógica que la distribución muestral de la media, con la única diferencia de que al extraer todas las posibles muestras de tamaño n de la población, el estadístico que se calcula en cada una de ellas es la proporción $p = x/n$ donde x es el número de datos de la muestra que cumplen la condición designada como "éxito" y n es el tamaño de la muestra.

Entonces, si llamamos π a la proporción de casos que cumplen una determinada condición en una población de tamaño N y extraemos todas las posibles muestras aleatorias de tamaño *n*, en la que definimos la variable *p* = "Proporción de aciertos", la distribución muestral de la proporción es la distribución de probabilidad del conjunto de todas las proporciones, p, obtenidas en todas las muestras posibles de tamaño n, extraídas de una población de tamaño N. La variable aleatoria p, sigue el modelo de probabilidad *binomial*, cuya media y desviación típica son, respectivamente:

$$\mu_p = \pi$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Como sabemos por los temas ya estudiados en el primer curso, las probabilidades asociadas a cada valor de *p* se pueden buscar en las tablas de distribución binomial con parámetros *n* y π.

Por otra parte, la distribución binomial -igual que la χ^2, la *t* de Student o la *F* de Snedecor-Fischer- se aproxima a la normal a medida que aumenta el tamaño de la muestra, y por tanto se puede generar una nueva variable cuya distribución es la normal tipificada:

$$Z = \frac{p-\pi}{\sigma_p}$$

Ejemplo 1.2: Una escuela de educación primaria está compuesta por un 40% de niños y un 60% de niñas. Si se elige una muestra aleatoria de 20 alumnos, ¿cuál será la probabilidad de que haya más de 9 niños?

La probabilidad de que en una muestra de 20 alumnos haya más de 9 niños, siendo la proporción de éstos en la población $\pi = 0{,}40$, se obtiene recurriendo a la distribución binomial con parámetros $n = 20$ y $\pi = 0{,}40$, la probabilidad pedida es, utilizando la expresión de su función de distribución, la siguiente[3]:

$$P(y > 9) = 1 - P(y \leq 9) = 1 - \sum_{y=0}^{9} \binom{20}{y} 0{,}40^y 0{,}60^{20-y} = 1 - 0{,}7553 = 0{,}2447$$

Y utilizando la distribución normal, tipificamos la proporción de niños obtenida en la muestra p=9/20=0,45

[3] Valor que también podríamos obtener recurriendo a la tabla de la distribución binomial como se estudió en el Tema 6 de la asignatura de *Introducción al Análisis de Datos.*

$$Z = \frac{p - \pi}{\sigma_P} = \frac{0{,}45 - 0{,}40}{\sqrt{\frac{0{,}40 \cdot 0{,}60}{20}}} = \frac{0{,}05}{0{,}1095} = 0{,}46$$

$$P(Z \leq 0{,}46) = 0{,}6772$$
$$P(Z > 0{,}46) = 1 - P(Z \leq 0{,}46) = 1 - 0{,}6772 = 0{,}3228$$

Los resultados obtenidos por los dos procedimientos no coinciden pero la diferencia encontrada va desapareciendo a medida que aumenta el tamaño de la muestra, ya que el ajuste de la distribución binomial a la normal con este incremento de n es más preciso. Esta diferencia entre la probabilidad calculada mediante la distribución discreta binomial (de parámetros media igual a $n \times \pi$ y varianza igual a $n\pi(1-\pi)$) y la calculada mediante la curva normal se debe a que esta última es continua. Si en vez de utilizar el punto p = 0,45 correspondiente a 9 éxitos utilizamos el punto medio entre 9 y 10 éxitos (p = 9,5/20 = 0,475) y repetimos los pasos anteriores obtendríamos un valor de 0,2483, bastante cercano al inicial (0,2447). En la Figura 1.3 se muestra la diferencia entre ambas perspectivas. Parece obvio que la segunda es más aproximada, aunque dependa de introducir como aproximación un valor (y = 9,5) que no puede producirse jamás en la distribución binomial ya que ésta exige valores enteros.

$$Z = \frac{P - \pi}{\sigma_P} = \frac{0{,}475 - 0{,}40}{\sqrt{\frac{0{,}40 \cdot 0{,}60}{20}}} = 0{,}68$$

$$P(Z \leq 0{,}68) = 0{,}7517$$
$$P(Z > 0{,}68) = 1 - P(Z \leq 0{,}68) = 1 - 0{,}7517 = 0{,}2483$$

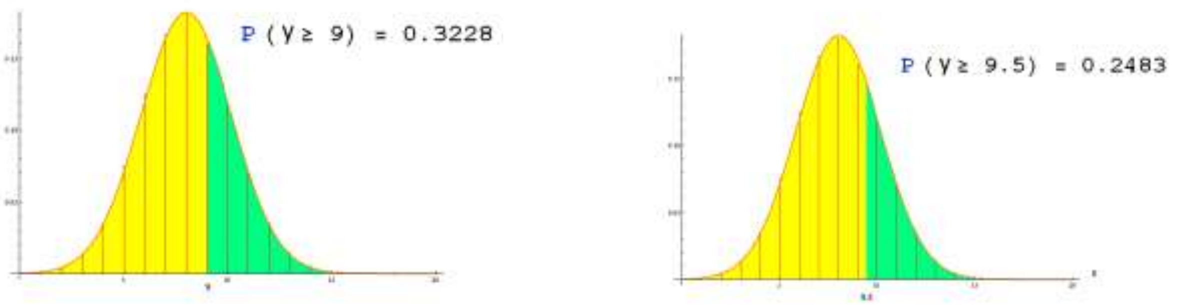


Figura 1.3: *Efecto de utilizar y = 9 ó y = 9,5 sobre las probabilidades para calcular la aproximación de la normal a la binomial. La curva continua es la curva normal con la misma media y desviación típica que la binomial. Las líneas verticales representan la función de probabilidad de la binomial.*

1.3.3. Distribución muestral de la varianza

La varianza es una medida de dispersión que permite determinar la variabilidad que presentan los datos para la variable objeto de estudio.

Recuérdese la relación entre la varianza (S_n^2) y la cuasivarianza (S_{n-1}^2):

$$\left.\begin{array}{l} S_n^2 = \dfrac{\sum(Y-\bar{Y})^2}{n} \Rightarrow \sum(Y-\bar{Y})^2 = n\cdot S_n^2 \\ S_{n-1}^2 = \dfrac{\sum(Y-\bar{Y})^2}{n-1} \Rightarrow \sum(Y-\bar{Y})^2 = (n-1)\cdot S_{n-1}^2 \end{array}\right\} \Rightarrow n\cdot S_n^2 = (n-1)\cdot S_{n-1}^2$$

Por lo que la cuasi-varianza de la muestra se puede calcular a partir de la varianza de la muestra:

$$S_{n-1}^2 = \frac{n}{n-1}\cdot S_n^2$$

No obstante, el proceso de construcción de una distribución muestral de varianzas no es tan inmediato como el de la media o el de la proporción, de modo que aquí nos limitaremos a describir cuál es la variable aleatoria, su distribución de probabilidad, sus medias -o valor esperado- así como su varianza y desviación típica.

La variable aleatoria que permite realizar afirmaciones sobre la varianza poblacional se puede generar a partir de la cuasivarianza o la varianza de la muestra:

$$X^2 = \frac{(n-1)S_{n-1}^2}{\sigma^2} = \frac{nS_n^2}{\sigma^2}$$

que se distribuyen según χ_{n-1}^2 (ji-cuadrado con $n-1$ grados de libertad). Es decir, mientras que en el primer caso de este capítulo calculábamos como estadístico de cada muestra, su media ($\bar{Y}$), en este caso para cada muestra calculamos el valor de X^2, para el cual necesitamos calcular la varianza (o cuasi-varianza) muestral, así como conocer el valor de σ en la población. La distribución de los valores de X^2 en todas las muestras posibles se distribuirá según χ_{n-1}^2 Teniendo en cuenta este modelo de probabilidad de la variable aleatoria así definida, su media y desviación típica son, respectivamente:

$$\mu_{X^2} = n-1$$

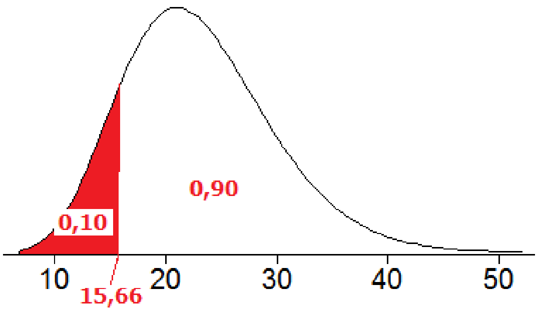
$$\sigma_{X^2} = \sqrt{2(n-1)}$$

Ejemplo 1.3: Supongamos que la altura (en centímetros) de los recién nacidos en Méjico se distribuye normalmente con media 48 cm y desviación típica 6 cm, es decir, N(48;6). Si se selecciona una muestra de 25 recién nacidos, ¿cuál es la probabilidad de que la desviación típica de la muestra tome un valor inferior a 4,75 centímetros?

Utilizando la desviación típica de la muestra, el valor de la variable aleatoria es:

$$X^2 = \frac{nS_n^2}{\sigma^2} = \frac{25 \cdot 4{,}75^2}{6^2} = 15{,}67$$

que es un valor de una distribución χ^2 con 24 grados de libertad.



Si buscamos en la tabla de probabilidades de la distribución χ^2_{24}, se observa que el valor 15,6587 que aparece en la tabla (el más aproximado a nuestro resultado) deja por debajo una probabilidad de 0,10. Por tanto, la probabilidad de que una muestra de 25 recién nacidos tenga una desviación típica inferior a 4,75 centímetros (o una varianza inferior a $4{,}75^2$) es aproximadamente de 0,10.

1.4. La estadística inferencial

Como se ha comentado en la introducción de este tema, la inferencia estadística permite inferir los parámetros de una, dos o más poblaciones a partir de la información recogida en las muestras. Esta inferencia o generalización de lo particular a lo general, la vamos a realizar mediante dos procedimientos íntimamente relacionados: la **estimación de parámetros** y el **contraste de hipótesis**. En ambos casos se trata de generalizar la información obtenida en una muestra a una población. Con la estimación tratamos de conocer el valor de uno o más parámetros correspondientes a una variable aleatoria poblacional, Y, a partir de los datos recogidos en una muestra. De forma alternativa, los procedimientos para el contraste de hipótesis (que son los más utilizados en la experimentación científica en el campo de las ciencias sociales y de la salud), nos permiten tomar una decisión sobre un valor hipotético que se formula como parámetro poblacional. El procedimiento se lleva a cabo analizando si determinadas características que hipotéticamente formulamos para definir la población pueden ser ciertas a partir de la información proporcionada por una muestra representativa de la misma.

Los procedimientos de contraste de hipótesis en los diseños de una, dos o más muestras que se verán en este curso se apoyan en el supuesto de que la muestra se ha seleccionado mediante muestreo aleatorio. Para ello, se tienen que cumplir dos condiciones: la muestra tiene que seleccionarse por algún procedimiento aleatorio y, en segundo lugar, todos los elementos de la población tienen la misma probabilidad de formar parte de la muestra. De esta forma, una muestra representativa es una reproducción, a escala, de la población a la que pertenece respecto a la o las variables que tratamos de estudiar. Por ejemplo si en la población de estudiantes de la UNED, el 60% son mujeres y de éstas el 40% tienen cargas laborales frente al 75% en los estudiantes varones, y queremos estudiar cómo las variables sexo y cargas laborales influyen en el rendimiento académico es necesario que la muestra recoja este mismo reparto de proporciones respecto al sexo y las cargas laborales. De no cumplirse esta condición, de los resultados observados en la muestra no se podrían hacer extrapolaciones válidas a la población general.

Aunque la estimación por intervalos y el contraste de hipótesis se tratan a continuación en epígrafes separados, veremos que son procedimientos complementarios de forma que los intervalos pueden aplicarse para el contraste de hipótesis y el contraste de hipótesis es una toma de decisión respecto al parámetro poblacional formulado.

1.4.1.- Estimación de parámetros. Propiedades de los estimadores.

Un estimador es un estadístico calculado en una muestra que se utiliza para estimar un parámetro poblacional. Para cada parámetro (v.g la media poblacional) pueden existir diferentes estimadores (v.g. la media aritmética, la media cuadrática, la mediana, la moda). Para que un estimador realice buenas estimaciones del parámetro poblacional es preciso que tenga las cuatro propiedades que de forma muy resumida expondremos en las siguientes líneas. Para desvincular las propiedades de los estimadores de un parámetro concreto, designaremos de forma genérica con **U** al parámetro poblacional, con $\widehat{U}$ a su valor estimado y con **u** a cualquier estadístico de la muestra que puede utilizarse como estimador. Por ejemplo, μ es el parámetro media poblacional y $\hat{\mu}$ su valor estimado. En este caso concreto, el estimador que se utiliza para estimar la media poblacional es el estadístico media aritmética de la muestra, $\hat{\mu} = \bar{Y}$. Es importante observar que, como hemos señalado al comienzo, podríamos haber elegido otros estadísticos muestrales como estimadores del parámetro media poblacional (v.g., la mediana, por citar algún otro de tendencia central). La cuestión es ¿cuál de los posibles estimadores deberíamos utilizar? Esto dependerá de la bondad de los mismos. Por lo tanto, es preciso saber qué hace que un estadístico, u, sea un buen estimador del parámetro[4] U, lo que dependerá de las siguientes propiedades:

Insesgado. Un buen estimador tiene que ser insesgado, lo cual supone que su valor esperado, $E(u) = U$. Es decir, la media de su distribución muestral, μ_u, debe coincidir con el parámetro que estima. La media muestral, tal como hemos visto, es un estimador insesgado de la media poblacional, y lo mismo ocurre con la proporción, la cuasi-varianza muestral y otros estadísticos que veremos a lo largo del curso. Sin embargo, por ejemplo, la varianza muestral es un estimador sesgado de la varianza poblacional ya que su valor esperado o media no coincide con la varianza poblacional, es decir: $E(S_n^2) \neq \sigma^2$, sin embargo, como veremos más adelante, $E(S_{n-1}^2) = \sigma^2$, por lo que la cuasivarianza muestral es un estimador insesgado de la varianza poblacional). Expresado, pues, de manera formal diremos que **u es un estimador insesgado de U, si su valor esperado o media coincide con el parámetro:** $\mu_u = U$

Eficiencia o precisión. Además de que un estimador coincida, en promedio, con su parámetro, es bueno que la distribución del estimador tenga poca variabilidad para que, de esta forma, se aleje poco del parámetro y en consecuencia sea más preciso. Por tanto, entre dos estimadores de un mismo parámetro, es más preciso el que tenga varianza más pequeña.

[4] Obsérvese que para denotar que un estadístico concreto es estimador de un parámetro, lo denotamos poniendo el acento circunflejo sobre el parámetro a estimar. De esta forma, conceptualmente no es lo mismo la media como estadístico de una muestra ($\bar{Y}$) que la media muestral como estimador de la media poblacional, es decir, $\hat{\mu} = \bar{Y}$. Aunque numéricamente valgan lo mismo, en el primer caso se la considera un simple índice descriptivo mientras que en el segundo se la considera un "representante" de la media poblacional y, además, un buen representante ya que nos sirve para inferir el valor μ desconocido.

Consistente. Como hemos visto al tratar la distribución muestral, raramente coincidirán los valores que el estimador adopta en muestras concretas con el parámetro, debido a las fluctuaciones del muestreo. Si pensamos en la distribución muestral de la media es fácil observar que al aumentar más y más el tamaño de la muestra el estimador se va aproximando al parámetro a la vez que su varianza tiende a cero. Con esta idea, decimos que un estimador consistente es aquel que se concentra en un rango cada vez más estrecho alrededor de su parámetro a medida que aumenta el tamaño de la muestra. Tomando como referencia las dos propiedades anteriores, se puede afirmar que una de las condiciones que hace consistente un estimador es que tanto su sesgo como su varianza tiendan a cero a medida que aumenta *n*.

Suficiencia. Un estimador es suficiente si al estimar el parámetro utiliza toda la información de la muestra relacionada con el parámetro. La media, la varianza y la proporción son estimadores suficientes de sus respectivos parámetros, porque en todos ellos se utiliza la información de todos los elementos de la muestra. No así la mediana que solo indica cuál es el valor central de la distribución.

Los estadísticos que hemos estudiado en cursos anteriores, cuando se aplican a los valores de las muestras que extraemos de la población -y que habitualmente se representan con letras del alfabeto latino- son los estimadores que podemos utilizar para estimar los parámetros poblacionales –representados con letras del alfabeto griego- y los que mejor cumplen con estas condiciones se muestran en la Tabla 1, de tal forma que en cada línea aparece el mejor estimador muestral de cada parámetro:

Estadísticos de la muestra	Parámetros en la población
Media: $\bar{Y}$	Media: μ
Cuasi-varianza: S_{n-1}^2	Varianza: σ^2
Proporción: p	Proporción: π
Correlación: r_{xy}	Correlación: ρ_{xy}
Ecuación de regresión: $Y' = B_0 + B_1 X$	Ecuación de regresión: $Y' = \beta_0 + \beta_1 X$

Tabla 1. *Estadísticos (muestra) y parámetros (población)*

Considerando estas propiedades que deben tener los (buenos) estimadores, la estimación de parámetros se realiza siguiendo dos procedimientos: la estimación puntual y la estimación por intervalos.

La estimación puntual consiste en utilizar el valor del estadístico calculado en la muestra como valor del parámetro que se desea estimar. Mediante este método se utiliza el estadístico obtenido en la muestra y se atribuye tal cual como parámetro de la población.

Sin embargo es poco probable que el valor del estadístico calculado en la muestra concreta coincida exactamente con el verdadero valor del parámetro y por ello es más interesante construir alrededor del estadístico de la muestra un intervalo, definido por su límite inferior y superior, que tenga en cuenta la precisión del estimador (su error típico) de forma que nos asegure, con una cierta probabilidad que el verdadero valor del parámetro se encuentra en esa franja de valores. A este método se le conoce como el cálculo de los ***intervalos de confianza*** en el ámbito de la estadística inferencial**.**

Por ejemplo, suponga que deseamos conocer el tiempo medio semanal que los estudiantes de psicología de la UNED dedican al estudio de una determinada asignatura. Mediante una encuesta realizada a una muestra representativa se obtiene una media de 6h/semanales. Este valor sería la estimación puntual para la media de todos los estudiantes. En otro caso, y mediante procedimientos que veremos más adelante podremos determinar que el tiempo medio que dedican los estudiantes al estudio es un valor comprendido entre 4,7h/semanales y 7,3 h/semanales con una probabilidad del 95%. Para llegar a estos resultados habremos utilizado los datos obtenidos en la muestra que ha sido encuestada y del conocimiento de las distribuciones muestrales de los estadísticos, con el doble objetivo tanto de asignar un valor del estadístico en la muestra que extraemos de la población, como estimación puntual de su parámetro, como para la estimación por intervalos.

1.4.1.1- Intervalo de confianza para la media

Para el cálculo del intervalo de confianza de la media hay que considerar las circunstancias bajo las cuales la distribución muestral de la media es una distribución normal o una distribución t de Student con *n*-1 grados de libertad. El cálculo del intervalo de confianza de la media aritmética (recordamos que siempre tiene que cumplirse una de las siguientes condiciones: distribución normal en la población, o bien, $\mathrm{n} \geq 30$) se realiza a partir de su estimador insesgado (la media muestral) sumando y restando a la media de la muestra una cantidad que recibe el nombre de error máximo de estimación, E_{max}. Para ilustrar el procedimiento nos apoyaremos en siguientes tres situaciones:

1.- **Varianza poblacional conocida $\boldsymbol{\sigma^2}$**. En estas circunstancias sabemos que la distribución muestral de la media es normal con media μ, y desviación típica (o error típico de la media) igual a la desviación típica poblacional dividida por la raíz cuadrada de n:

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Se trata, por tanto, de determinar dos valores que definen un intervalo dentro del cual estimamos que se encontrará la media poblacional, μ, con una determinada probabilidad, que representamos por $1-\alpha$, que se denomina **nivel de confianza**. Teniendo en cuenta las propiedades de la distribución normal si, por ejemplo, fijamos un nivel de confianza del $1-\alpha=0{,}95$ o del 95%, sabemos que entre los valores Z= -1,96 y Z=+1,96 a izquierda y derecha de la media de la distribución muestral, $\mu_{\bar{Y}}=\mu$, se encuentran el 95% de las medias de cualquier muestra, como se muestra en la Figura 1.4, expresadas en puntuaciones típicas Z.

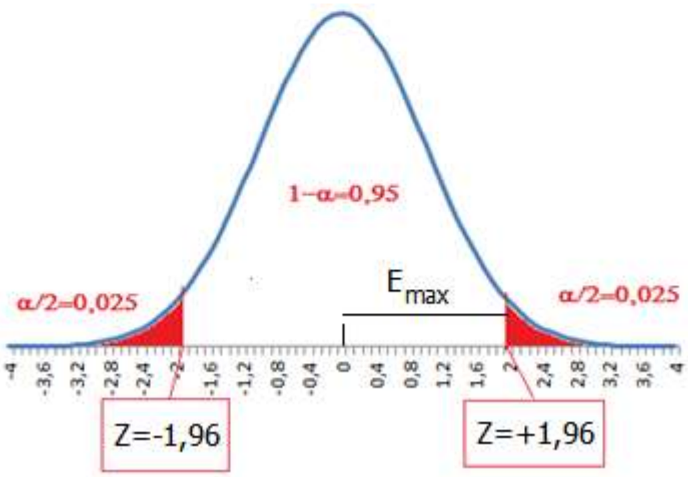


Figura 1.4. *Distribución muestral de medias con intervalo del 95% alrededor del valor esperado*

Es decir, en 95 de cada 100 muestras la media de la muestra, se encontrará comprendida entre las puntuaciones típicas: $z=-1{,}96$ y $z=+1{,}96$, de acuerdo a la siguiente expresión:

$$P\left(-1{,}96 \leq \frac{\bar{Y}-\mu}{\sigma_{\bar{Y}}} \leq +1{,}96\right) = 0{,}95$$

Siendo $\sigma_{\bar{Y}}$ la desviación típica de la distribución muestral de la media o error típico de la media, cuya expresión es:

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$

Nuestro interés se centra en estimar el parámetro poblacional, μ , a partir del estadístico de la muestra, $\bar{Y}$, que es un valor conocido. Para ello, de la expresión anterior tendríamos que despejar el valor de μ , y llegaríamos a:

$$P(\bar{Y} - 1{,}96\, \sigma_{\bar{Y}} \leq \mu \leq \bar{Y} + 1{,}96\, \sigma_{\bar{Y}}) = 0{,}95$$

Obsérvese que el intervalo de confianza se construye sumando y restando a la media de la muestra una cantidad que define el **error máximo de estimación** con un nivel de confianza del 95%, en este caso, y que representa la máxima diferencia que puede existir entre el estimador y el parámetro a estimar, en el caso de la media aritmética la diferencia entre: $\bar{X} - \mu$,

$$E_{max} = Z_{1-\alpha/2}\ \sigma_{\bar{Y}} = Z_{1-\alpha/2}\ \frac{\sigma}{\sqrt{n}}$$

A modo de resumen, en la Figura 1.5 se representa la distribución de probabilidad de las medias obtenidas al extraer todas las posibles muestras de tamaño *n* que se pueden extraer de una población. De todas estas muestras, en el 95% de ellas su media se encontrará dentro de la zona central y sólo un 5% estarán fuera de ese zona. Por lo tanto, partiendo de la media de una muestra que se encuentre dentro de la zona central -aunque no necesariamente coincidiendo con la media poblacional, μ, ya que varía de una muestra a otra- construimos un intervalo con la misma amplitud que tendrá una probabilidad del 95% de contener la media poblacional. Si partimos de la media de una muestra que se encuentra en la zona sombreada -fuera de la zona central del 95%- el intervalo de confianza que construyamos sobre ella no podrá incluir entre sus valores a la media de la población. Esto último sucederá, en promedio, en 5 de cada 100 muestras que extraigamos de la población.

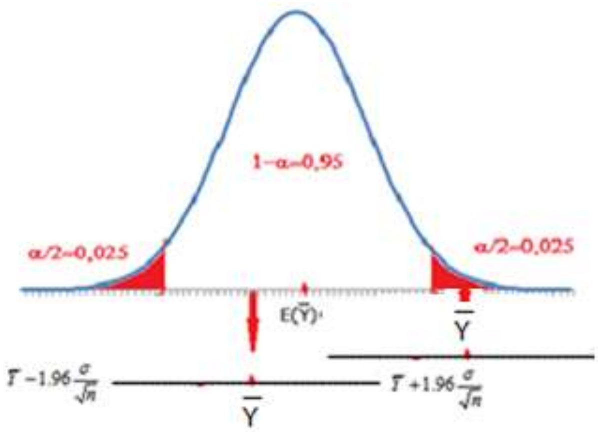


Figura 1.5. *Intervalo de confianza de la media con un NC del 95%*

2.- Varianza poblacional desconocida. En la práctica estadística no es frecuente que se conozca la varianza poblacional. Lo habitual es desconocer tal dato que tendremos que estimar a partir de la varianza o cuasi-varianza de la muestra como un estimador de la varianza poblacional. En estas circunstancias la distribución muestral de la media es la distribución t de Student y, por tanto, el intervalo de confianza para la media poblacional, estimado a partir de la media de la muestra, $\bar{Y}$, y con un nivel de confianza de $1-\alpha$ sigue el mismo razonamiento anterior pero apoyándose en una distribución t con $n-1$ grados de libertad. De acuerdo con ese mismo razonamiento, el intervalo de confianza de la media poblacional es:

$$P(\bar{Y} - E_{max} \leq \mu \leq \bar{Y} + E_{max}) = 1 - \alpha$$

La diferencia ahora es que utilizamos la distribución t como distribución muestral de la media, por lo que el error máximo de estimación es:

$$E_{max} = t_{1-\frac{\alpha}{2},n-1}\ \sigma_{\bar{Y}}$$

Los valores de t son los que dejan un intervalo central correspondiente a una probabilidad de $1-\alpha = 0{,}95$. Puesto que la varianza poblacional es desconocida, hay que estimarla a partir de su estimador (sesgado o insesgado) por lo que el error típico de la media, es:

$$\sigma_{\bar{Y}} = \frac{S_{n-1}}{\sqrt{n}} = \frac{S_n}{\sqrt{n-1}}$$

Quedando el error máximo de estimación de la siguiente forma:

$$E_{max} = t_{1-\frac{\alpha}{2},n-1}\ \frac{S_{n-1}}{\sqrt{n}}$$

$$E_{max} = t_{1-\frac{\alpha}{2},n-1}\ \frac{S_n}{\sqrt{n-1}}$$

Según se utilice la cuasi-desviación típica de la muestra (su estimador insesgado), en el primer caso, o la desviación típica de la muestra (estimador sesgado) en el segundo caso.

Ejemplo 1.4. En un experimento sobre atención, un psicólogo presenta durante 300 mseg un grupo de 16 letras del alfabeto (con una disposición de 4 filas y 4 columnas). Cada uno de los 12 sujetos que participan en el experimento debe verbalizar tantas letras como recuerde de cada presentación estimular. El promedio de letras bien recordadas es de 7 y la desviación típica insesgada (cuasi-desviación típica) es de 1,3. Asumiendo que la distribución en la población es normal (necesitamos este supuesto, porque la muestra es pequeña). ¿Entre qué límites se encontrará el verdadero promedio de palabras bien recordadas, con una probabilidad de 0,95?

Se asume que la distribución poblacional es normal con varianza desconocida, y la muestra es pequeña, por lo que la distribución muestral de la media es la t de Student. En la distribución t de Student con 11 gl, (Figura 1.6a) buscamos los valores que dejan en la zona central una probabilidad de 0,95. Estos valores son -2,201 y +2,201 que se incluyen en la expresión general:

$$P(\bar{Y} - E_{max} \leq \mu \leq \bar{Y} + E_{max})$$

$$E_{max} = t_{1-\frac{\alpha}{2},n-1} \cdot \frac{\sigma}{\sqrt{n}} = 2{,}201 \cdot \frac{1{,}3}{\sqrt{12}} = 0{,}826$$

$$P(7 - 0{,}826 \leq \mu \leq 7 + 0{,}826)$$

$$P(6{,}174 \leq \mu \leq 7{,}826) = 0{,}95$$

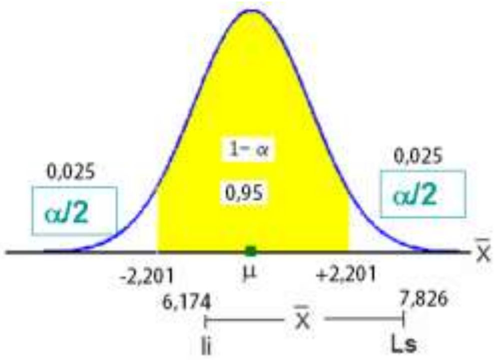


Figura 1.6a. *Intervalo de confianza de la media en la distribución t*

3.- Varianza poblacional desconocida (n>100). En este caso, en teoría seguimos trabajando con una distribución t de Student con $n-1$ grados de libertad, pero conociendo las propiedades de esta distribución, sabemos que cuanto mayor sea el valor de los grados de libertad, más se aproxima la distribución t a la distribución normal. En las tablas que manejamos, podemos consultar valores en distribuciones t hasta cien grados de libertad. Para valores superiores a dichos grados de libertad podemos considerar que las diferencias entre los valores Z y t son prácticamente despreciables, por lo que utilizaremos los valores Z de las tablas de curva normal, en sustitución de los valores t, cuando las muestras tengan un tamaño de $n > 100$.

Ejemplo 1.5. En un experimento sobre atención, un psicólogo presenta durante 300 mseg un grupo de 16 letras del alfabeto (con una disposición de 4 filas y 4 columnas). Cada uno de los 122 sujetos que participan en el experimento debe verbalizar tantas letras como recuerde de cada presentación estimular. El promedio de letras bien recordadas es de 7 y la desviación típica es de 1,3. ¿Entre qué límites se encontrará el verdadero promedio de palabras bien recordadas, con una probabilidad de 0,95?

Aunque se trata de una distribución t, (Figura 1.6 b) buscamos los valores que dejan en la zona central una probabilidad de 0,95 en la tabla Z, porque $n > 100$. Estos valores son -1,96 y +1,96 que se incluyen en la expresión general:

$$P(\bar{Y} - E_{max} \leq \mu \leq \bar{Y} + E_{max}) = 1 - \alpha$$

$$E_{max} = t_{1-\frac{\alpha}{2},n-1} \cdot \frac{S_n}{\sqrt{n-1}} = 1{,}96 \cdot \frac{1{,}3}{\sqrt{122-1}} = 0{,}232$$

$$P(7 - 0{,}232 \leq \mu \leq 7 + 0{,}232) = 0{,}95$$

$$P(6{,}768 \leq \mu \leq 7{,}232) = 0{,}95$$

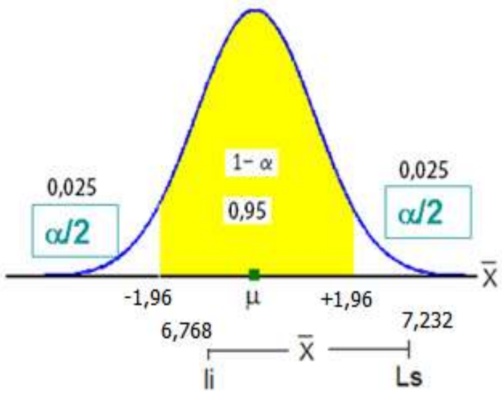


Figura 1.6b. *Intervalo de confianza de la media en la distribución t*

La interpretación correcta del intervalo de confianza es que dentro de él se encontrará, o no, el verdadero valor del parámetro, pero nos permite afirmar que si repitiésemos el proceso con muchas muestras del mismo tipo y tamaño, en una proporción igual a $(1 - \alpha)$ los intervalos así construidos contendrán al verdadero valor del parámetro (promedio de palabras recordadas en la población). Y esta interpretación es la que hay que mantener para todo intervalo de confianza de cualquier otro parámetro poblacional que vayamos a estimar, no cayendo en el error de interpretarlo en el sentido de que una proporción de personas igual a $(1 - \alpha)$ –en este ejemplo, el 95% de las personas- tienen un promedio de palabras recordadas comprendido entre 6,768 y 7,232.

1.4.1.2. Intervalo de confianza para la proporción

Sabemos que la distribución muestral de la proporción es una distribución binomial que se aproxima a la normal cuando se utilizan muestras grandes. Bajo estas condiciones, la distribución muestral de la proporción es normal con media y error típico iguales a:

$$\mu_p = \pi$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Cuando la proporción poblacional, π, es un valor desconocido hay que estimarlo a partir de su estimador insesgado, la proporción muestral, p, y el error típico de la distribución muestral de la proporción queda de la siguiente forma:

$$\sigma_p = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} = \sqrt{\frac{p(1-p)}{n}}$$

Teniendo en cuenta las propiedades de la distribución normal, si fijamos un nivel de confianza del $1-\alpha$ y siguiendo el mismo razonamiento utilizado para el caso de la media, llegaríamos a la siguiente expresión:

$$P(p - E_{max} \leq \pi \leq p + E_{max}) = 1 - \alpha$$

Siendo el error máximo de estimación:

$$E_{max} = Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Y de esta forma, la expresión final del intervalo de confianza de la proporción poblacional es la siguiente:

$$P(p - E_{max} \leq \pi \leq p + E_{max}) = 1 - \alpha$$

Ejemplo 1.6: Para dejar constancia real de las preferencias de los padres sobre la lengua vehicular en la que prefieren que se eduque a sus hijos, una determinada asociación de padres realiza una encuesta sobre una muestra de 800 familias residentes en una determinada autonomía bilingüe, encontrando que 280 familias son partidarios de que todas de las asignaturas se enseñen en Castellano. Con un nivel de confianza del 95% ¿entre que valores se encontrará la proporción de padres que en esa Comunidad son partidarios de que todas las asignaturas se impartan en Castellano?

La proporción de familias partidarias de la enseñanza en Castellano obtenida en la muestra es p=280/800 = 0,35. Al tratarse de una muestra grande, la distribución binomial se aproxima a la normal. Buscamos en la tabla de la distribución normal los valores Z que dejan una probabilidad central del 95% y son -1,96 y +1,96 (Figura 1.7) y aplicamos la siguiente expresión para calcular el error máximo de estimación:

$$P(p - E_{max} \leq \pi \leq p + E_{max}) = 1 - \alpha$$

$$E_{max} = Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 1{,}96\sqrt{\frac{0{,}35.0{,}65}{800}} = 0{,}033$$

$$P(0{,}35 - 0{,}033 \leq \pi \leq 0{,}35 + 0{,}033) = P(0{,}317 \leq \pi \leq 0{,}383) = 0{,}95$$

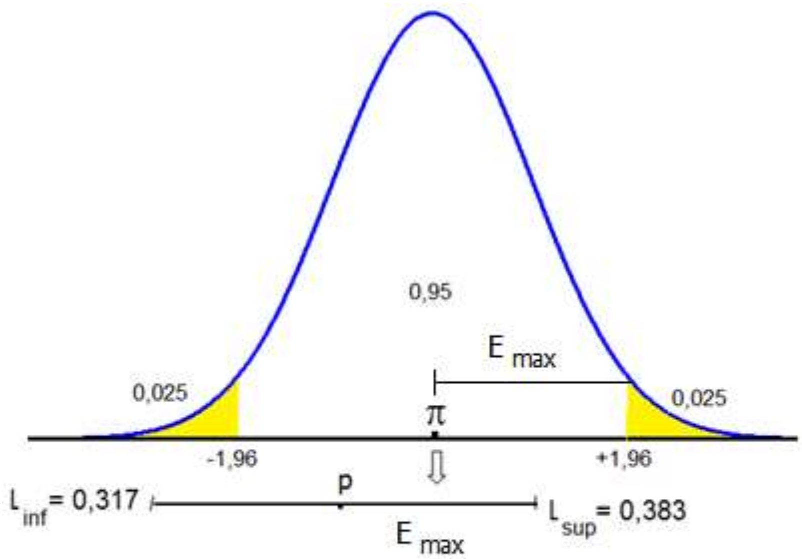


Figura 1.7. Intervalo de confianza de la proporción sobre una distribución normal.

En consecuencia, podemos decir que la proporción poblacional, π, es un valor comprendido entre 0,317 (31,7%) y 0,383 (38,3%) con una probabilidad, o nivel de confianza, del 95% (Figura 1.7):

$$P(0{,}317 \leq \pi \leq 0{,}383) = 0{,}95$$

1.4.1.3. Intervalo de confianza para la varianza

Cuando tratamos la distribución muestral de la varianza vimos que la variable aleatoria $\frac{(n-1)S_{n-1}^2}{\sigma^2} = \frac{nS_n^2}{\sigma^2}$ se distribuía según χ^2 con $n-1$ g.l. La Figura 1.8 es una representación genérica de esta distribución en la que se indica la probabilidad de que un valor de esa variable aleatoria, tomado al azar, se encuentre entre los dos valores que delimitan la zona mas clara, que vale $1 - \alpha$.

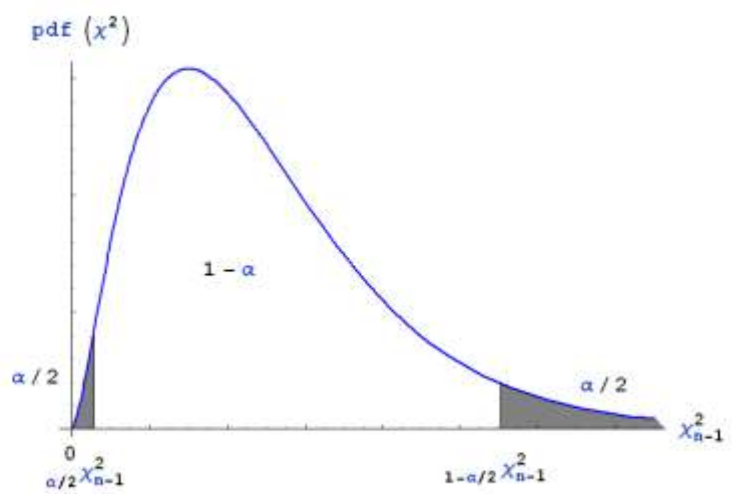


Figura 1.8. *Distribución* χ^2 *con n-1 grados de libertad*

Entonces, si fijamos dos valores: $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ de la distribución de probabilidad de χ^2_{n-1} de tal forma que la probabilidad, p, de que un valor tomado al azar de la variable aleatoria, $\frac{nS_n^2}{\sigma^2}$, , se encuentre en la zona delimitada entre estos dos valores (zona clara de la figura) sea igual a un valor que representamos por $1-\alpha$ y que representa el nivel de confianza. Escrito de otra forma:

$$P\left(\chi^2_{\alpha/2;n-1} \leq \frac{nS_n^2}{\sigma^2} \leq \chi^2_{1-\alpha/2;n-1}\right) = 1-\alpha$$

Para obtener el intervalo de confianza de la varianza hay que despejar de la expresión anterior el valor de la varianza poblacional. Veámoslo paso a paso. Primeramente al pasar a los lados de la desigualdad el valor de nS_n^2 tendríamos:

$$P\left(\frac{\chi^2_{\alpha/2;n-1}}{nS_n^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{1-\alpha/2;n-1}}{nS_n^2}\right) = 1-\alpha$$

y a continuación con el fin de aislar la varianza poblacional, σ^2, tenemos que invertir los miembros de la desigualdad lo que conlleva que varíe el sentido de la misma y tenemos:

$$P\left(\frac{nS_n^2}{\chi^2_{\alpha/2;n-1}} \geq \sigma^2 \geq \frac{nS_n^2}{\chi^2_{1-\alpha/2;n-1}}\right) = 1-\alpha$$

y ordenando esta desigualdad de menor a mayor, llegamos a la expresión del intervalo de confianza de la varianza poblacional:

$$P\left(\frac{nS_n^2}{\chi^2_{1-\alpha/2;n-1}} \leq \sigma^2 \leq \frac{nS_n^2}{\chi^2_{\alpha/2;n-1}}\right) = 1-\alpha$$

Es decir que los límites del intervalo de confianza para la varianza poblacional son:

$$l_{inf} = \frac{nS_n^2}{\chi^2_{1-\alpha/2;n-1}} \; ; \quad l_{sup} = \frac{nS_n^2}{\chi^2_{\alpha/2;n-1}}$$

Con las pertinentes modificaciones, se puede usar también la varianza insesgada (cuasi-varianza) siendo en este caso los límites inferior y superior los siguientes:

$$l_{inf} = \frac{(n-1)S_{n-1}^2}{\chi^2_{1-\alpha/2;n-1}} \; ; \quad l_{sup} = \frac{(n-1)S_{n-1}^2}{\chi^2_{\alpha/2;n-1}}$$

Cuando el tamaño de la muestra está por encima de 100 sujetos, la distribución muestral de la varianza (distribución chi-cuadrado) se puede aproximar a la normal que al ser una distribución simétrica respecto a su media permite obtener los límites del intervalo de confianza, sumando y restando al estimador, el error máximo de estimación, que para un nivel de confianza determinado es el siguiente:

$$P(S^2 - E_{max} \leq \sigma^2 \leq S^2 + E_{max}) = 1 - \alpha$$

$$E_{max} = Z_{1-\alpha/2} S^2 \sqrt{\frac{2}{n}}$$

Donde S^2 se refiere indistintamente a S_n^2 o S_{n-1}^2. De esta forma, la expresión final del intervalo de confianza para la varianza poblacional, cuando se trabaja con muestras grandes, sería la siguiente:

$$P(S^2 - E_{max} \leq \sigma^2 \leq S^2 + E_{max}) = 1 - \alpha$$

Ejemplo 1.7: Un grupo de 30 alumnos de enseñanza secundaria seleccionados al azar en una determinada Comunidad realizan un test de comprensión verbal de su lengua autónoma. Las puntuaciones obtenidas se distribuyen normalmente con media 120 y varianza 36. Con una probabilidad de 0’90, ¿entre que valores se encontrará la varianza en comprensión verbal de todos los alumnos de secundaria de esa Comunidad?

Buscamos en la tabla de la distribución chi-cuadrado y con $n - 1 = 29$ grados de libertad, los dos valores de la variable chi-cuadrado que dejan una probabilidad de 0,90 central. Estos valores son 17,708 y 42,557 tal y como se representan en la Figura 1.9.

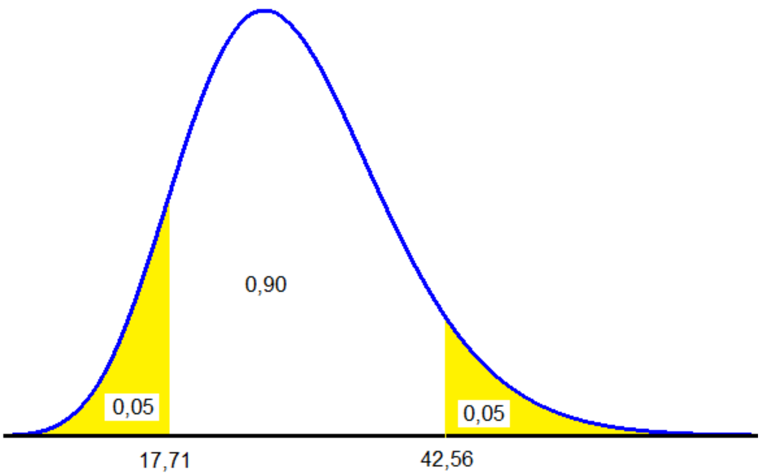


Figura 1.9. *Distribución chi-cuadrado con 29 g.l y valores que delimitan una probabilidad de 0,90 central*

$$l_{inf} = \frac{nS_n^2}{\chi^2_{1-\alpha/2;n-1}} = \frac{30 \cdot 36}{42{,}557} = 25{,}38; \qquad l_{sup} = \frac{nS_n^2}{\chi^2_{\alpha/2;n-1}} = \frac{30 \cdot 36}{17{,}708} = 60{,}99$$

Al mismo resultado llegaríamos utilizando la cuasi-varianza de la muestra. En este ejemplo, la varianza es 36 por lo que la cuasi-varianza vale:

$$S_{n-1}^2 = \frac{nS_n^2}{n-1} = \frac{30 \times 36}{29} = 37{,}24$$

Y los límites son:

$$l_{inf} = \frac{(n-1)S_{n-1}^2}{\chi^2_{1-\alpha/2;n-1}} = \frac{29 \times 37{,}24}{42{,}557} = 25{,}38\ ; \quad l_{sup} = \frac{(n-1)S_{n-1}^2}{\chi^2_{\alpha/2;n-1}} = \frac{29 \times 37{,}24}{17{,}708} = 60{,}99$$

1.4.2.- Amplitud del intervalo de confianza y su relación con el tamaño muestral.

La amplitud de un intervalo de confianza depende de dos factores: el nivel de confianza y el error típico de la distribución muestral del estadístico. Este segundo factor está en proporción inversa al tamaño de la muestra, de tal forma que cuanto mayor es el tamaño de la muestra, menor es el error típico del estadístico. Esta relación es fundamental, pues permite dar al intervalo de confianza el grado de precisión que se desee.

Para que el lector vea el proceso, vamos a ejemplificarlo con la media. El error típico de este estimador, cuando se desconoce la varianza poblacional, es: $\frac{S_{n-1}}{\sqrt{n}}$, y para obtener el error máximo de estimación se multiplica por el valor de la distribución *t* de Student (o la Z de la distribución normal, según la situación) correspondiente al nivel de confianza que se haya estipulado. Es decir, la distancia desde la media muestral a cualquiera de los límites, que vamos a llamar **error máximo de estimación** y lo designamos con E es:

$$E_{max} = t_{1-\alpha/2,n-1}\frac{S_{n-1}}{\sqrt{n}}$$

Si despejamos el tamaño de la muestra, n, y lo ponemos en función del resto de elementos el resultado es:

$$n = S_{n-1}^2 \frac{t^2_{1-\alpha/2;n-1}}{E^2_{max}}$$

Siguiendo un razonamiento similar, en el Cuadro 1.1 se resume el cálculo del tamaño de la muestra para los tres estadísticos básicos: media, varianza y proporción en función del nivel de confianza y del error máximo de estimación, E, que se quiera fijar.

Cuadro 1.1. *Calculo del tamaño de la muestra en función de la precisión de la estimación*

Media	Varianza poblacional conocida	$n = \sigma^2 \frac{z^2_{1-\alpha/2}}{E^2_{max}}$
	Varianza poblacional desconocida	$n = S^2_{n-1} \frac{t^2_{1-\alpha/2;n-1}}{E^2_{max}}$
Varianza (n>100)	$n = 2\,S^4 \frac{Z^2_{1-\alpha/2}}{E^2_{max}}$	
Proporción	$n = p(1-p) \frac{z^2_{1-\alpha/2}}{E^2_{max}}$	

En definitiva, las fórmulas del Cuadro 1.1 permiten al investigador calcular el tamaño de la muestra en función del error máximo, E, que esté dispuesto a admitir y del nivel de confianza $(1-\alpha)$ adoptado.

Veamos la aplicación con un sencillo ejemplo:

Ejemplo 1.8.Se desea calcular el tamaño de la muestra que se requiere utilizar en una encuesta electoral de manera que la precisión en la proporción de voto estimada, o error máximo de estimación, con un nivel de confianza del 95%, sea de $\pm$ 0,02.

Situándonos en la situación más desfavorable respecto del error típico de la proporción[5], se tiene que: $p = 1 - p = 0{,}5$. De acuerdo con esto, tendremos:

$$n = 0{,}5 \cdot 0{,}5 \cdot \frac{1{,}96^2}{0{,}02^2} = 2401$$

Con este número de sujetos, el investigador se asegura que la amplitud del intervalo de confianza será 0,04 (cuatro puntos porcentuales) con un nivel de confianza del 95%.

Podemos comprobar, que si el investigador quisiera trabajar con una precisión en la estimación igual a: $E_{max} = \pm 0{,}01$, entonces el tamaño de la muestra pasaría a ser de 9604 sujetos.

1.4.3. Contraste de hipótesis.

Una hipótesis estadística es una conjetura que se formula sobre una población y que puede someterse a prueba, o contrastación empírica, a partir de la información proporcionada por una muestra

[5] Si observa la fórmula del error típico de la distribución muestral de la proporción deducirá que alcanzará su valor máximo cuando p=q=0,5

representativa de esa población. Una vez que la hipótesis se ha contrastado con los datos de la muestra es el momento de tomar alguna decisión respecto a su resultado. El contraste de hipótesis es, pues, una parte esencial del método científico.

En general, siempre se parte de algún interrogante que se plantea en el ámbito de una investigación, a la luz de un determinado marco teórico, y debería formularse de una manera sencilla y clara: ¿votan las mujeres en mayor proporción a partidos de centro izquierda que a los de centro derecha?; en el proceso de trabajo manual ¿es más eficaz verbalizar las acciones durante la tarea que hacerlas en silencio?; ¿es más eficaz una terapia A que otra B para el tratamiento de la fobia de los niños a montar en ascensores?; ¿los salarios de hombres y mujeres son iguales por un mismo trabajo?

Una vez planteada la cuestión, hay que buscar una solución que adopte la forma de afirmación empíricamente verificable, es decir, debemos ser capaces de operativizar nuestras preguntas para que tengan entidad de hipótesis científicas. La mejor manera de hacerlo es plantearla en términos estadísticos; esto significa que las afirmaciones que se realicen estén relacionadas de alguna manera con una o más distribuciones de probabilidad. Por ejemplo, el salto entre una hipótesis científica como "¿los salarios de hombres y mujeres son iguales por un mismo trabajo?", se puede sustanciar como hipótesis estadística preguntando: "¿Es igual la media de salario de hombres y mujeres para un mismo trabajo?"; o también se podría preguntar en términos de otro estadístico tal como la Mediana, o en términos de una función de distribución. Es decir, una hipótesis científica se pueden plantear con diferentes hipótesis estadísticas, las cuales, al contrastarse dan respuesta a dicha hipótesis científica.

Las **hipótesis estadísticas** planteadas para dar respuesta a la hipótesis científica son: l**a hipótesis nula** y la **hipótesis alternativa**. La hipótesis nula se representa por H_0, y puede contener afirmaciones como las siguientes:

Hipótesis	Hipótesis estadística. Hipótesis Nula
La media de salario de hombres y mujeres para un mismo trabajo son iguales	$H_0: \mu_{Hombres} = \mu_{Mujeres}$
La proporción de mujeres que votan a partidos de centro-izquierda es del 60%.	$H_0: \pi = 0{,}6$
¿La varianza, respecto a un valor previo establecido, es al menos de 12 puntos?	$H_0: \sigma^2 \geq 12$
Las puntuaciones de un test de razonamiento numérico tienen distribución normal con media 50 y desviación típica 5.	H_0: la variable Y tiene distribución normal N(50,5).

En general, la hipótesis nula afirma que no existe diferencia entre el valor del estadístico obtenido en la muestra y el que formulamos como parámetro poblacional o, en otras palabras, que la diferencia observada entre estos dos valores es nula. Como la realidad es que estos valores casi nunca van a coincidir, lo que estamos afirmando es que la diferencia observada puede explicarse como resultado del azar. De otra forma, si se repitiese la investigación un número suficiente de veces con diferentes muestras del mismo tipo y tamaño extraídas aleatoriamente de la misma población, las diferencias observadas entre el estadístico calculado con los datos muestrales y el valor formulado en la hipótesis nula como parámetro poblacional, serían unas veces grandes, otras pequeñas, unas positivas, otras negativas, pero en conjunto tenderían a neutralizarse para finalmente ser cero.

Para cada hipótesis nula planteada, es preciso plantear otra, denominada **hipótesis alternativa**, representada por H_1, y que es la negación de la hipótesis nula, de tal forma que si la hipótesis nula es falsa la hipótesis alternativa tiene que ser *verdadera* o viceversa. Por tanto, estas dos hipótesis tienen que ser exhaustivas y mutuamente excluyentes. Para el conjunto de hipótesis nulas anteriores, las alternativas serían:

Hipótesis científica	Hipótesis estadística. Hipótesis Alternativa
La media de salario de hombres y mujeres para un mismo trabajo NO son iguales	$H_1: \mu_{Hombres} \neq \mu_{Mujeres}$
La proporción de mujeres que votan a partidos de centro-izquierda NO es del 60%.	$H_1: \pi \neq 0{,}6$
La varianza respecto al valor anterior establecido es menor de 12 puntos?	$H_1: \sigma^2 < 12$
Las puntuaciones de un test de razonamiento numérico NO tienen distribución normal con media 50 y desviación típica 5.	H_1: la variable Y NO tiene distribución normal N(50,5).

Dependiendo de cómo esté formulada la hipótesis nula se marca la **dirección del contraste**. Si, por ejemplo, la H_0 está planteada como igualdad de las medias de hombres y mujeres, mientras que la alternativa es simplemente su negación (las medias no son iguales) se dice que es un **contraste bilateral** porque H_1 admite que la diferencia pueda ser favorable a los hombres (un extremo de las posibles puntuaciones con respecto a la igualdad) o a las mujeres (el extremo contrario) Si, por el contrario, conocemos la dirección en que H_0 puede ser falsa, como por ejemplo en la hipótesis $H_0: \sigma^2 \geq 12$, o, en general, cuando en la investigación se plantea que un método de aprendizaje, un fármaco, un determinado proceso industrial, etc. tiene efecto positivo (o negativo) sobre lo que estamos estudiando, entonces tenemos un **contraste unilateral** en la medida en que indicamos la dirección esperada según H_1 de ese efecto. Igualmente en estos casos, para una hipótesis alternativa "El método A, favorece el aprendizaje", la hipótesis nula, su negación, sería: "El método A no favorece el aprendizaje".

En cualquier caso las hipótesis nula y alternativa son exhaustivas y mutuamente excluyentes, de tal forma que la negación de una conlleva la confirmación de la otra.

Una vez que se ha planteado la hipótesis, es preciso definir lo que se conoce como **medida de la discrepancia** y que, en general, cuando se trata de hacer contrastes sobre parámetros poblacionales, es una medida estandarizada dentro de alguna distribución de probabilidad, a semejanza de las vistas en los epígrafes de distribuciones muestrales. La medida de discrepancia no depende de las unidades en que esté medida la variable y su formulación habitual es:

$$\begin{pmatrix} \text{Estadístico de contraste} \\ o \\ \text{discrepancia} \end{pmatrix} = \frac{\begin{pmatrix} \text{valor del estadístico} \\ \text{en la muestra} \end{pmatrix} - \begin{pmatrix} \text{valor del parámetro} \\ \text{planteado en } H_0 \end{pmatrix}}{\begin{pmatrix} \text{desviación típica de la distribución} \\ \text{muestral del estadístico} \end{pmatrix}}$$

Además de definir la discrepancia es preciso considerar qué cantidad de ésta consideramos admisible para que la discrepancia observada no sea atribuible al azar. Es decir, debemos determinar, *a priori*, cuál será la diferencia máxima entre el estimador y el parámetro que estamos dispuestos a considerar compatible con la H_0, y esta decisión dependerá tanto de la distribución de probabilidad de la medida de discrepancia como de la dirección del contraste, como del riesgo que estamos dispuestos a asumir.

Como veremos en próximos apartados, este valor de la discrepancia se establece, también, en términos de probabilidad de obtener una diferencia entre el estadístico obtenido en la muestra y el parámetro formulado en la hipótesis igual o mayor que la observada. Esta probabilidad es la que se conoce como **nivel crítico p**, y en la mayor parte de las investigaciones se rechazará H_0 si este valor es menor de 0,05 o 0,01.

1.4.3.1 Metodología clásica del contraste de hipótesis

La metodología del contraste es fruto de los trabajos de Fisher, Neyman y Pearson y su lógica recuerda a la de un juicio en un estado de derecho, en el cual el acusado siempre es inocente (la hipótesis nula) hasta que las pruebas no demuestren lo contrario (la hipótesis alternativa). En los contrastes de hipótesis las pruebas son las evidencias recogidas en los datos muestrales provenientes de una investigación bien diseñada[6] y se parte de que la hipótesis nula es verdadera (presunción de inocencia). Si los datos aportan resultados significativamente diferentes de los planteados en la hipótesis nula, ésta es rechazada, y en caso contrario, no podremos hacerlo por no tener evidencias contra ella, de modo que la mantendremos como provisionalmente verdadera hasta que se encuentren nuevas evidencias.

Los procedimientos para el cálculo de intervalos de confianza -y buena parte de los contrastes de hipótesis que veremos en los siguientes temas- se basan en una serie de supuestos (v.gr., que la muestra procede de una población de puntuaciones que se distribuyen según una función de distribución poblacional conocida, como la curva normal, o sobre el nivel de medida de la variable, etc). Estos procedimientos y otros que no se han presentado todavía (ANOVA, regresión múltiple, etc.), se engloban en lo que se conoce como "**métodos paramétricos**" cuya denominación procede de la búsqueda de los parámetros subyacentes a unos datos asumiendo que éstos se distribuyen según una función de distribución poblacional concreta. Todos las pruebas paramétricos asumen una determinada forma (normal, binomial, F, etc.) para la distribución poblacional de los datos observados en la muestra. Pero a veces nos encontramos con situaciones en las que no podemos asumir los supuestos subyacentes a las pruebas paramétricas y necesitamos procedimientos cuya validez no dependa de esos supuestos. En este caso se nos hace necesario acudir a otro conjunto de técnicas que no exijan estos supuestos tan restrictivos. Por contraposición a los anteriores métodos, se los conoce como "**métodos no paramétricos**". Los contrastes de hipótesis no paramétricos se utilizan cuando no se cumplen los supuestos necesarios para realizar un contraste paramétrico, por ejemplo, cuando la variable dependiente no alcanza un nivel de medida de intervalo o razón, cuando la muestra es pequeña y no conocemos la forma de la distribución poblacional, etc.

[6] Tratado en la asignatura de Fundamentos de Investigación

Teniendo en consideración esta primera distinción entre las pruebas paramétricas y no paramétricas que se aplicarán en todo contraste, las etapas de un contraste de hipótesis las vamos a resumir en los siguientes puntos:

1.- Condiciones de la investigación y supuestos que cumplen los datos observados. A lo largo de este curso veremos que al diseñar cualquier investigación se puede trabajar con una, dos, tres o más muestras, las cuales pueden ser independientes o relacionadas, en las que se recoge información sobre una o más variables medidas con la misma o con diferentes escalas de medida (nominal, ordinal, de intervalo o de razón). Por otra parte, estos datos pueden provenir de poblaciones en las que la variable de estudio tiene una distribución de probabilidad conocida o desconocida. Todas estas características tanto del diseño como de los datos condicionan tanto la hipótesis que se puede someter a contrastación empírica como el procedimiento de análisis de datos más adecuado para someter a contrastación empírica la hipótesis.

2.- Formulación de la hipótesis nula y de la alternativa. Conforme al contexto de la investigación se formulan las hipótesis nula y alternativa, de las cuales se deriva un contraste bilateral o unilateral en función de sus objetivos. Por lo general la hipótesis científica, dirigida a encontrar resultados significativos, es la hipótesis alternativa que se aceptará como verdadera si la investigación aporta evidencias contra la hipótesis nula que es la que se somete a contrastación empírica.

3.- Estadístico de contraste. Representa una medida de la discrepancia entre la información proporcionada por los datos empíricos recogidos en la muestra y la proposición teórica planteada en la hipótesis nula. Esta medida es una variable aleatoria con una determinada distribución de probabilidad (normal, t, chi-cuadrado, etc.) que va a aportar información empírica sobre la afirmación formulada en H_0

4.- Regla de decisión. Una vez calculado el estadístico de contraste o discrepancia entre los datos empíricos observados en la muestra y los datos teóricos que planteamos en la hipótesis nula queda tomar una decisión respecto al rechazo o no de la hipótesis nula. Para ello, el investigador establece previamente el **nivel de significación**, α. Según Fisher, el nivel de significación, α, representa el máximo riesgo que el investigador está dispuesto a cometer al tomar la decisión errónea de rechazar una hipótesis nula verdadera. Por tanto, a la luz de sus resultados y del estadístico de contraste, el investigador calcula la probabilidad de obtener unos resultados como los observados en la muestra o más extremos. Esta probabilidad recibe el nombre de **nivel crítico p**. Si el nivel crítico p es muy pequeño en comparación con el nivel de significación, α, rechazamos la H_0 y en caso contrario la mantenemos.

El nivel de significación que suele utilizarse en la mayoría de las investigaciones es del 0,05, aunque en investigaciones más rigurosas se trabaja con un nivel de significación de 0,01. En cualquiera de los casos, se rechazaría la hipótesis nula siempre que la probabilidad de explicar los resultados obtenidos en relación a la hipótesis nula sea menor que el nivel de significación.

Otra alternativa a la hora de tomar la decisión de rechazar o no la hipótesis nula consiste en fijar el nivel de significación α, por lo que automáticamente se fija **el valor o valores críticos** de la distribución muestral que marcarán la máxima diferencia que podemos admitir, por simple azar, entre el valor teórico planteado en H_0 y el valor obtenido en la muestra. Este valor, o valores críticos, definen -en la distribución muestral del estadístico de contraste- los límites entre la zona de rechazo o no de la H_0.

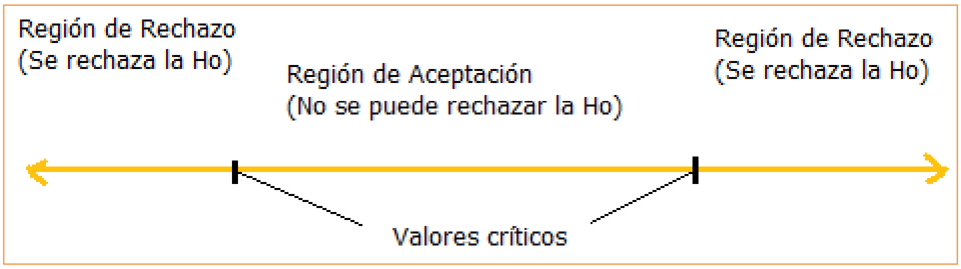
La ***zona de rechazo*** depende del nivel de significación, α, y es el área de la distribución muestral que corresponde a un valor de la discrepancia tan alejado de H_0 que la probabilidad de que se produzca es muy baja, si efectivamente H_0 es verdadera. En otras palabras, es aquella zona de la distribución muestral constituida por el conjunto de muestras para las cuales se rechaza la hipótesis nula H_0.

La ***región de no rechazo***, complementaria a la anterior, depende del nivel de confianza, $1-\alpha$, y es el área de la distribución muestral que corresponde a valores pequeños de la discrepancia tan poco alejados del valor formulado en la H_0 que la probabilidad de que se produzca es alta si efectivamente la H_0 es verdadera, por lo que no representa evidencia suficiente para rechazarla. En otras palabras, es aquella zona de la distribución muestral constituida por el conjunto de muestras para las cuales se mantiene la hipótesis nula H_0.

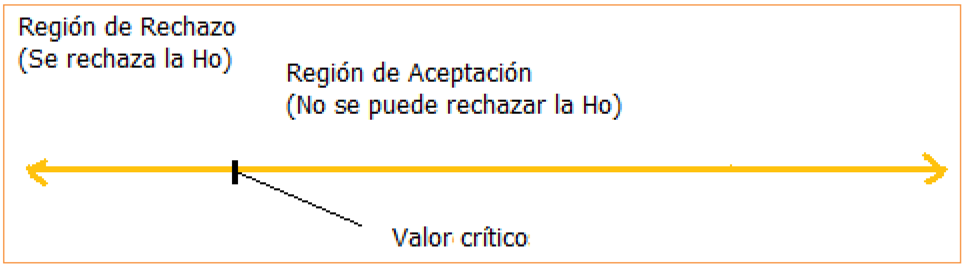
Por tanto, el valor o valores críticos corresponden a la máxima diferencia que cabe esperar por simple azar entre los datos empíricos obtenidos en la muestra y los datos teóricos que formulamos para la población, de tal forma que si el estadístico de contraste se sitúa en la zona de NO rechazo, podemos concluir que la diferencia observada no es significativa y se debe a los errores aleatorios por lo que no podemos rechazar la hipótesis nula con un determinado nivel de confianza.

De forma similar si el estadístico de contraste alcanza la zona de rechazo indicaría que la diferencia observada entre los datos empíricos y los datos teóricos es muy poco probable que pueda atribuirse a errores aleatorios y concluimos que la diferencia observada es significativa, lo que nos lleva a rechazar la hipótesis nula con un determinado nivel de confianza.

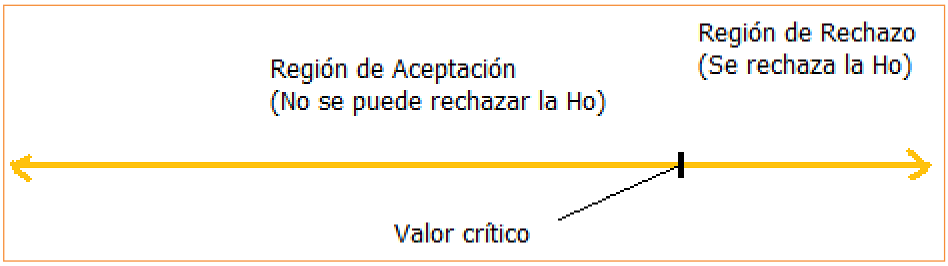
Aunque existen contrastes de hipótesis, que veremos posteriormente, en los que siempre se deja todo el nivel de significación en una parte de la distribución, en general y con independencia de la forma de la función de distribución del estadístico de contraste, si el contraste es bilateral tendremos tres zonas delimitadas por los dos valores críticos que se sitúan en el eje horizontal de la distribución muestral como las esquematizadas en el siguiente gráfico:



Si el contraste es unilateral izquierdo solo tendremos dos zonas, siendo la región de rechazo la situada en la parte izquierda de la distribución, como se representa en el siguiente gráfico esquemático:



De forma similar, si el contraste es unilateral derecho, la región de rechazo se situará en la parte derecha de la distribución muestral como se representa en el siguiente gráfico esquemático:



En cualquier caso, ya sea comparando el estadístico de contraste con el valor crítico o comparando el nivel crítico p con el nivel de significación α , la decisión que se toma respecto a la H_0 es la misma. Puesto que no hay verdades absolutas y siempre existe un riesgo de error, formalmente la hipótesis nula NUNCA se acepta, sino que la estrategia de la investigación es buscar evidencias para rechazarla.

5.- Conclusión. Formulada la hipótesis nula, que es la que sometemos a contrastación empírica asumiendo que es provisionalmente verdadera y una vez calculado el estadístico de contraste, se concluye rechazando o no la hipótesis nula (no hay un punto intermedio[7]). Si no tenemos evidencia suficiente para rechazarla, se está señalando que la hipótesis se mantiene porque es compatible con la evidencia muestral (el acusado en el juicio es inocente), y si se rechaza se quiere significar que la evidencia muestral no avala la hipótesis (las pruebas están en contra del acusado) y por tanto se rechaza.

6.- Interpretación. La conclusión simple y llana en términos de rechazo o no de la hipótesis nula tiene su correspondiente interpretación dentro del contexto de la investigación y de la hipótesis y objetivos que el investigador formula en su trabajo.

Ilustremos este razonamiento con un sencillo ejemplo, similar al que plantea R.A. Fisher en su libro *El Diseño de Experimentos*, en el cual refería la afirmación de una dama según la cual, cuando tomaba el té, podía detectar si se había vertido antes la leche o la infusión en la taza. Para refutar esta "facultad" de la dama podríamos realizar un contraste con los siguientes datos ficticios.

Ejemplo 1.9. Para contrastar la presunta "habilidad detectora" de la dama se preparan 16 tazas de té, siguiendo ambos procedimientos: en ocho se vierte primero la leche, y en otros ocho se vierte primero la infusión. La presentación se realiza al azar y la dama sólo tiene que decir cuál ha sido el procedimiento. Supongamos, por ejemplo, que la dama acierta en 12 ocasiones. ¿Es compatible este resultado muestral con la afirmación de la dama?. Como nivel de significación, tomaremos α = 0,05.

Seguiremos los 6 pasos del proceso pero utilizando sólo el nivel crítico p como regla de decisión, dejando el cálculo del estadístico de contraste para los siguientes temas.

[7] Cuando la medida de discrepancia cae justo en la región crítica de la zona de aceptación o rechazo, es difícil tomar una decisión sobre H_0. En estas circunstancias se suele coger nueva evidencia y proceder a un nuevo contraste.

1.- Condiciones y supuestos. 16 ensayos independientes con dos resultados posibles en cada uno: acierto o error, y la probabilidad del resultado permanece constante en todos ensayos.

2.- Formulación de las hipótesis nula y alternativa: Planteamos un contraste unilateral en el que la hipótesis nula presupone que la dama, en principio, no tiene dicha habilidad, y por tanto la proporción de veces que acertaría sería un valor igual a 0,5 o inferior (es decir, tendría la misma habilidad que el resto de los mortales). La hipótesis alternativa plantea que la dama si tiene esa habilidad y por tanto es capaz de acertar en más del 50% de los ensayos.

$$H_0: \pi \leq 0{,}5 \qquad H_1: \pi > 0{,}5$$

3.- Estadístico de contraste. Como estamos contrastando una proporción cuya distribución de probabilidad es la distribución binomial que estudiamos el curso pasado, vamos a calcular la probabilidad de que, bajo el supuesto de que la dama no tiene esa extraña facultad (y por tanto su proporción de aciertos es la de cualquier persona normal del 50% que se formula en la H_0) la dama haya sido capaz de detectar la diferencia en más de la mitad de la veces, concretamente en 12 o más ocasiones de los 16 ensayos realizados (formulado en la H_1):

$$P(y \geq 12) = \sum_{y=12}^{16} \binom{16}{y} 0{,}5^y 0{,}5^{n-y} = 1 - P(y < 12) = 1 - \sum_{y=0}^{11} \binom{16}{y} 0{,}5^y 0{,}5^{n-y} = 1 - 0{,}9616 = 0{,}0384$$

x	p
1	0,0002
2	0,0018
3	0,0085
4	0,0278
5	0,0667
6	0,1222
7	0,1746
8	0,1964
9	0,1746
10	0,1222
11	0,0667
12	0,0278
13	0,0085
14	0,0018
15	0,0002
16	0,0000

Cuadro 1.2. *Tabla de la distribución binomial para* $n = 16$ *y* $p = 0{,}5$

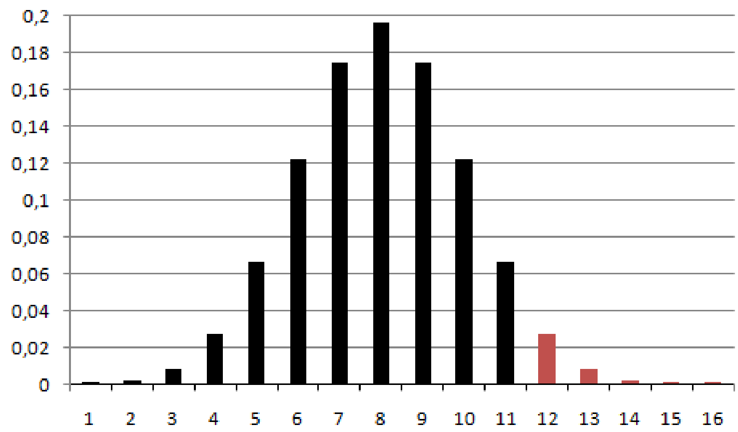


Figura 1.10. *Representación gráfica del ejemplo 1.6*

4.- Regla de decisión. Este resultado quiere decir que, bajo el supuesto de que la hipótesis nula es cierta y la probabilidad de acierto de la dama es de 0,5, la probabilidad de que en 16 ensayos la dama acierte en 12 ocasiones o más es de 0,0384, o que hay un 3,84% de probabilidades de que la dama, sin tener esa extraña habilidad, acierte por puro azar en 12 ocasiones o más. Como regla de decisión para rechazar o no la hipótesis nula comparamos esta probabilidad con el nivel de significación (0,05) y puesto que la probabilidad encontrada es menor que 0,05, rechazamos la H_0,

En la Figura 1.10 se representa la distribución binomial del cuadro 1.2 del ejemplo con el nivel crítico p representado por valores escritos y representados por barras rojas cuya suma es menor de 0,05. Por tanto, la regla de decisión bajo la interpretación de Fisher consiste en calcular la probabilidad de obtener unos resultados como los observados en la muestra o más extremos (nivel crítico p). Si esta probabilidad es muy pequeña en comparación con α, pueden ocurrir dos cosas: o bien la hipótesis nula es cierta (la dama no goza de tal habilidad) y se ha producido una situación muy poco probable (pero no imposible) o bien la hipótesis nula es falsa. Parece más lógico (o probable) inclinarse por esta segunda opción que descarta el azar como explicación del resultado obtenido y ante la evidencia que proporciona este resultado, el investigador opta por rechazar la hipótesis nula, asumiendo que esta afirmación tiene un cierto riesgo o probabilidad de error, que ha establecido en el 5%. Si, por el contrario, la probabilidad hubiese sido mayor que el nivel de significación, entonces no se podría descartar el azar, como explicación de la diferencia y se opta por no rechazar la hipótesis nula.

5.- Conclusión: Rechazamos la hipótesis nula, ya que el nivel crítico p es menor que 0,05. El nivel crítico p= 0,0384, nos indica la probabilidad de acertar por simple azar en 12 o más de las 16 ocasiones. Es un valor lo suficientemente pequeño que nos conduce a descartar el azar como explicación de este número de aciertos tan alto. En consecuencia, si descartamos el azar como explicación de que la dama acierte en 12 o más de los 16 ensayos es porque la dama si tiene realmente esa capacidad.

6.- Interpretación: Rechazar la hipótesis nula quiere decir que la dama tiene esa habilidad para distinguir si en una taza de té se ha puesto primero la leche o la infusión con un nivel de confianza del 95%.

Observe el lector que si se establece el nivel de significación en 0,01 la conclusión sería otra y sería necesario obtener evidencias más fuertes o una probabilidad mucho menor para descartar el azar como explicación de esta extraña habilidad de la señora.

En los siguientes temas seguiremos esta metodología pero utilizando, no solo el nivel crítico p, sino también y fundamentalmente el estadístico de contraste como medida de la discrepancia entre los valores teóricos que formulamos en la población y la información empírica que nos proporcionan los datos recogidos en la muestra, para los diseños de investigación que utilizan una, dos o más muestras. Y partir de estos estadísticos de contraste podremos calcular también el nivel crítico p.

1.4.3.2.- Errores al tomar una decisión en un contraste clásico de hipótesis

Hemos visto que el contraste de hipótesis es un proceso por el cual se toma una decisión acerca de lo que se afirma en la hipótesis nula. No obstante, una cosa es la decisión que se adopta sobre H_0 y otra es la propia naturaleza de H_0. Tenemos dos opciones posibles acerca de la decisión sobre H_0: o se Acepta o se Rechaza, y dos opciones sobre la naturaleza de H_0: o es verdadera o es falsa.

El error que supone rechazar una hipótesis nula cuando en realidad es verdadera se denomina **Error Tipo I**, y su probabilidad asociada es α. El error que supone aceptar una hipótesis nula cuando en realidad es falsa, siendo verdadera la hipótesis alternativa, se conoce como **Error Tipo II**, y su probabilidad asociada es β

Su complementario es $1 - \beta$ y corresponde a la **potencia del contraste** que es la decisión correcta de rechazar la hipótesis nula cuando es falsa.

Puede observarse que los errores y las decisiones correctas están probabilísticamente relacionados (cuanto menor es α mayor es $1 - \alpha$, por ejemplo). Estas ideas pueden verse reflejadas en el siguiente cuadro o tabla de doble entrada con dos alternativas en cada entrada como puede verse en el cuadro 1.3.

		Naturaleza de H_0	
		Verdadera	**Falsa**
Decisión sobre H_0	No rechazar	Decisión correcta. **Nivel de confianza** $1 - \alpha$	Decisión errónea **Error tipo II** β
	Rechazar	Decisión errónea **Error tipo I** α	Decisión correcta **Potencia del contraste** $1 - \beta$

Cuadro 1.3. *Decisiones sobre la hipótesis nula*

En todo contraste de hipótesis el investigador fija *a priori* el nivel de significación (error tipo I) y su complementario, el nivel de confianza, o decisión correcta de no rechazar una hipótesis nula que puede ser verdadera. En el tema siguiente veremos cómo se calcula el error tipo II y la potencia del contraste (pero exclusivamente para el contraste sobre la media y la proporción en los diseños de una muestra) y veremos también cómo este valor depende del nivel de significación, del tamaño de la muestra, n, y del tamaño del efecto.

¿Cuál de estos errores es más grave? Vamos a ilustrarlo con dos ejemplos. Suponga que comparamos un tratamiento nuevo, A, con otro antiguo, B, ya existente. La hipótesis nula dirá que los dos tratamientos son, al menos, igual de eficaces frente a la hipótesis alternativa según la cual el tratamiento A es mejor que el B:

$$H_0: A \leq B \qquad H_1: A > B$$

Imagine que rechazamos la hipótesis nula cuando en realidad es cierta, es decir, que concluimos que el tratamiento nuevo, A, es más eficaz cuando en realidad son iguales. En esta situación estaríamos cometiendo el error tipo I con una probabilidad α.

Si por el contrario, y como resultado del contraste de hipótesis, concluimos que los dos tratamientos son iguales (es decir, no rechazamos la hipótesis nula que es falsa) cuando en realidad el nuevo tratamiento es mejor que el anterior, estaríamos cometiendo un error tipo II con una probabilidad β. Es evidente que el error tipo II resulta más grave por cuanto impide beneficiarse de un tratamiento que es más eficaz. Mientras que el error tipo I, (sin entrar a considerar las consecuencias económicas de la decisión de cambiar a algo que es igual de eficaz) tiene consecuencias menos graves, al menos en el campo del progreso del conocimiento.

La realidad de la H_0

		Verdadera	Falsa
Decisión sobre H_0	No rechazar	**Correcto** El tratamiento no tiene efecto y así se decide. Probabilidad $1-\alpha$	**Error de tipo II** El tratamiento si tiene efecto pero no lo percibimos. Probabilidad $\boldsymbol{\beta}$
	Rechazar	**Error de tipo I** El tratamiento no tiene efecto pero se decide que sí. Probabilidad $\boldsymbol{\alpha}$	**Correcto** El tratamiento tiene efecto y el experimento lo confirma. Probabilidad $1-\beta$

Esta valoración puede cambiar dependiendo del contexto en el que nos situemos. Por ejemplo, aplicando el mismo razonamiento sobre la inocencia o culpabilidad de un imputado, las hipótesis nula y alternativa son: $H_0: Inocente \quad H_1: Culpable$

Partimos del supuesto de que la hipótesis nula es verdadera (el imputado es inocente) mientras no se tenga evidencia suficiente para condenarle (rechazar la H_0 y aceptar la H_1). En este contexto se introducen otras razones de índole moral para valorar qué es más grave: si tener un inocente en la cárcel o un culpable en la calle.

La realidad de la H_0

		Inocente	Culpable
Decisión sobre H_0	No rechazar H_0 Inocente	**Correcto** Es inocente. Probabilidad $1-\alpha$	**Error de tipo II** Es culpable y no se le condena. Probabilidad $\boldsymbol{\beta}$
	Rechazar H_0 Culpable	**Error de tipo I** Es inocente y se le condena. Probabilidad $\boldsymbol{\alpha}$	**Correcto** Es culpable. Probabilidad $1-\beta$

En cualquier caso, el hecho de no rechazar la H_0 puede deberse o bien a que sea realmente verdadera o sea falsa pero el experimento no tenía suficiente potencia para detectarlo. Debido a esto, aunque el análisis de los datos de nuestro diseño de investigación no produzca resultados significativos, nunca podremos aceptar la H_0 como verdadera. Como siempre, existe la posibilidad de que la H_0 sea falsa, pero el diseño de nuestra investigación no tiene la suficiente sensibilidad para detectarlo, tan solo podremos concluir que nuestros resultados no han aportado evidencia suficiente para rechazar la hipótesis nula. El análisis de la potencia del contraste nos proporciona información sobre el grado de confianza en los resultados no significativos que no ha permitido rechazar la hipótesis nula formulada en el diseño de nuestra investigación.

Por otra parte, y como veremos en temas posteriores, tanto en la investigación aplicada en el ámbito de la psicología, como en otras ciencias, un resultado estadísticamente *significativo* sobre la eficacia de un nuevo tratamiento puede no tener una *significación práctica* en el sentido de no representar un valor terapéutico real o importante. De aquí la recomendación, casi ineludible, de conocer adicionalmente el **tamaño del efecto** que expresa la magnitud de la diferencia observada entre la hipótesis nula (el valor teórico) y la hipótesis alternativa (el valor observado) expresado en una métrica común y que, por su importancia metodológica de cara a la validez de la conclusión estadística de la investigación que hemos diseñado, expondremos con algo más de detalle en temas posteriores.

1.5. **Ejercicios de autoevaluación.**

1. ¿Cuál de las siguientes afirmaciones referidas a la distribución muestral de la media es FALSA: a) Es normal cuando la población se distribuye normalmente y conocemos su varianza; b) Tiende a la normal cuando desconocemos la varianza poblacional pero trabajamos con muestras grandes; c) Siempre es normal con media igual a la media poblacional.
2. El nivel crítico p representa la probabilidad de: a) que la hipótesis nula sea verdadera; b) que siendo verdadera la hipótesis nula obtengamos unos datos como los observados o más extremos en la muestra; c) rechazar la hipótesis nula siendo verdadera.
3. A una muestra de 122 estudiantes universitarios, de los cuales 74 son mujeres se les pasa una prueba de memoria de palabras sin sentido, obteniendo una media de 15 palabras recordadas, con una varianza de 9. Trabajando con un nivel de confianza del 95%, ¿entre que valores se encontrará la media poblacional?
4. A una muestra de 122 estudiantes universitarios, de los cuales 74 son mujeres se les pasa una prueba de memoria de palabras sin sentido, obteniendo una media de 15 palabras recordadas, con una varianza de 9. Entre que valores se encontrará la varianza poblacional con un nivel de confianza del 95%?
5. A una muestra de 122 estudiantes universitarios, de los cuales 74 son mujeres se les pasa una prueba de memoria de palabras sin sentido, obteniendo una media de 15 palabras recordadas, con una varianza de 9. Si queremos estimar la varianza poblacional con un error máximo de estimación que no supere los dos puntos, ¿Cuál debería ser el tamaño de la muestra que debemos utilizar fijando el nivel de confianza en el 95%?.
6. A una muestra de 122 estudiantes universitarios, de los cuales 74 son mujeres se les pasa una prueba de memoria de palabras sin sentido, obteniendo una media de 15 palabras recordadas, con una varianza de 9. Trabajando con un nivel de confianza del 95%, ¿entre que valores se encontrará la proporción poblacional de mujeres universitarias?.
7. En un contraste de hipótesis es habitual que el investigador fije a priori el valor de α que representa la probabilidad de: a) conservar la hipótesis nula cuando no se encuentra en los datos de la muestra suficiente evidencia para rechazarla; b) rechazar la hipótesis nula siendo cierta; c) aceptar la hipótesis alternativa siendo cierta.

8. La probabilidad de rechazar la hipótesis nula siendo falsa, es: a) un valor conocido y fijado a priori por el investigador; b) un valor desconocido que representa el error tipo II; c) la potencia del contraste que es un valor desconocido a priori.
9. La amplitud del intervalo es más estrecho a medida que; a) aumenta el tamaño de la muestra; b) aumenta el nivel de confianza; c) aumenta el error típico del estadístico.
10. Se llama error típico de un estadístico a la desviación típica de: a) El parámetro poblacional; b) La distribución muestral de este estadístico; c) Los datos recogidos en la muestra.
11. En una muestra aleatoria de 100 personas se mide el pulso obteniendo una media de 65 ppm con desviación típica 5 ppm. ¿Cuánto vale el error máximo de estimación de la media poblacional con un nivel de confianza del 95%?: a) 0,5025; b) 0,997; c) 25,25

Soluciones.

1.- De las tres afirmaciones la c) es FALSA. La distribución muestral de la media es normal cuando la distribución en la población es normal o bien la muestra es superior a 30 observaciones y además conocemos la varianza poblacional. Si la varianza poblacional es desconocida entonces la distribución muestral de la media es la t de Student, que tiende a la normal cuando la muestra es grande.

2.- El nivel crítico p es la probabilidad asociada al estadístico de la muestra e indica la probabilidad de que, siendo cierta la hipótesis nula, encontrar valores tan extremos como los obtenidos en la muestra, luego la opción correcta es b.

3.- En este caso, dado que desconocemos la varianza poblacional, la distribución muestral de la media se distribuye según t de Student, con $n-1$ grados de libertad, pero dado que la muestra es grande ($n > 100$) utilizamos valores Z de la curva normal tipificada, que para un nivel de confianza del 95% son: $\pm 1{,}96$. Por lo tanto, los límites inferior y superior del intervalo son:

$$E_{max} = t_{1-\alpha/2,n-1}\frac{S}{\sqrt{n-1}} \approx z_{1-\alpha/2}\frac{S}{\sqrt{n-1}} = 1{,}96\frac{3}{\sqrt{122-1}} = 0{,}535$$

$$L_{inf} = \bar{Y} - E_{max} = 15 - 0{,}535 = 14{,}47$$

$$L_{sup} = \bar{Y} + E_{max} = 15 + 0{,}535 = 15{,}53$$

4.- Al tratarse de una muestra grande, la distribución muestral de la varianza se aproxima a la normal.

$$E_{max} = z_{1-\alpha/2}S^2\sqrt{\frac{2}{n}} = 1{,}96 \cdot 9\sqrt{\frac{2}{122}} = 2{,}2586 \rightarrow \begin{cases} L_{inf} = 9 - 2{,}2586 = 6{,}741 \\ L_{sup} = 9 + 2{,}2586 = 11{,}259 \end{cases}$$

Obsérvese que en este caso el error máximo de estimación que cometemos al estimar la varianza poblacional es de $\pm 2{,}258$ puntos. De ahí que si queremos mejorar la precisión de nuestra estimación, uno de los procedimientos puede ser aumentar el tamaño de la muestra, como veremos en el siguiente ejemplo.

5.-El tamaño de la muestra para un error máximo de estimación de ±2 puntos, es

$$n = 2S^4 \frac{z^2_{1-\alpha/2}}{E^2_{max}}$$

Conocemos la varianza de la muestra que en este ejemplo es 9, es decir; $S^2 = 9$ por tanto, $S^4 = (S^2)^2 = 9^2 = 81$:

$$n = 2S^4 \frac{z^2_{1-\alpha/2}}{E^2_{max}} = 2 \cdot 81 \cdot \frac{1{,}96^2}{2^2} = 155{,}6$$

Un valor muy similar si utilizamos la cuasivarianza de la muestra:

$$S^2_{n-1} = \frac{n}{n-1} S^2_n = \frac{122}{121} \cdot 9 = 9{,}07 \rightarrow S_{n-1} = 3{,}012$$

$$n = 2S^4 \frac{z^2_{1-\alpha/2}}{E^2_{max}} = 2 \cdot 3{,}012^4 \cdot \frac{1{,}96^2}{2^2} = 158{,}08$$

6.- El género es una variable dicotómica con distribución binomial siendo la proporción de mujeres obtenida en la muestra: $p = 74/122 = 0{,}6066$. Al tratarse de una muestra grande, esta distribución se aproxima a la normal y los límites del intervalo de confianza para π, son:

$$E_{max} = z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} = 1{,}96 \sqrt{\frac{0{,}6066(1-0{,}6066)}{122}} = 0{,}087$$

$$L_{inf} = p - E_{max} = 0{,}6066 - 0{,}087 = 0{,}52$$

$$L_{sup} = p + E_{max} = 0{,}6066 + 0{,}087 = 0{,}693$$

7.- El nivel de significación se representa por α y es un valor fijado a priori por el investigador como criterio de decisión. Los valores habituales son 0,05 y 0,01 y representan la probabilidad que desde un inicio asume el investigador de cometer el error tipo I, es decir, de rechazar una hipótesis nula que es verdadera. Su complementario 1-α, es el nivel de confianza o probabilidad de tomar la decisión correcta de no rechazar una hipótesis nula cuando no hay evidencia suficiente para hacerlo.

8.- La probabilidad de rechazar una hipótesis nula que es falsa es la potencia del contraste, cuyo valor es desconocido a priori pero se puede calcular, ya que depende del nivel de significación, α, del tamaño de la muestra y la magnitud del efecto.

9.- La amplitud del intervalo de confianza depende del tamaño de la muestra y del nivel de confianza, de forma que al aumentar el nivel de confianza el intervalo es más amplio y al aumentar el tamaño de la muestra disminuye el error típico del estadístico y con ello el intervalo de confianza, como puede deducirse al ver sus expresiones de cálculo.

10.- La opción A es incorrecta por varias razones: se refiere a la distribución poblacional, no a la distribución muestral; puede ser cualquier parámetro, no solo la desviación típica. La opción B es correcta. La opción C es incorrecta porque el error típico se refiere a la variabilidad de la distribución muestral del estadístico mientas que la desviación típica se refiere a la variabilidad de los valores recogidos en la muestra.

11.- Desconocemos la varianza (o la desviación típica) de la población y la forma de la distribución poblacional, por lo que nos apoyaremos en la distribución t con $n-1$ gl como distribución muestral. El valor de α es 0,05. Nos piden el Error Máximo de estimación. De forma general, el intervalo de confianza de la media se obtiene sumando y restando a la media de la muestra (estimación puntual del parámetro) el error máximo de estimación que corresponde a t veces el error típico de la media (o Z si se tratase de una población con distribución normal y varianza conocida) :

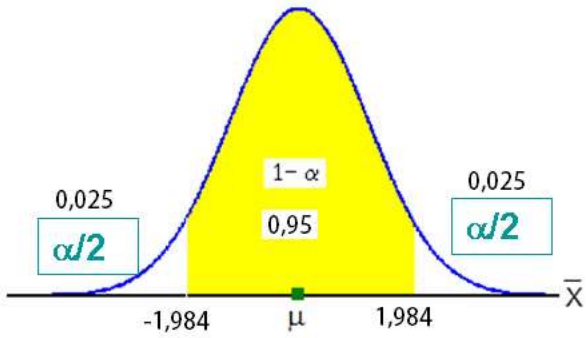
Estimador puntual $\pm$ [(Z o t) * (Error típico)]

Luego el valor que nos están pidiendo es el valor que está entre corchetes y que representa el error máximo de estimación:

Si no conocemos la varianza poblacional, la distribución muestral de la media es la distribución t de Student con $n-1$ grados de libertad y la varianza poblacional hay que estimarla a partir de la varianza o cuasi-varianza de la muestra. Por otro lado, los valores críticos de la distribución t con $n-1=100-1=99$ grados de libertad son: $|t_{0,025}| = |t_{0,975}| = 1{,}984$

Hemos buscando con 100 grados de libertad (el más cercano a 99, ya que este no existe en la tabla y hemos elegido una probabilidad de 0,975 que es la suma de: $\alpha/2 = 0{,}025$ y el Nivel de Confianza = 0,95 ($0{,}025 + 0{,}950 = 0{,}975$).

g.l.	0,550	0,600	0,650	0,700	0,750	0,800	0,850	0,900	0,950	0,975	0,990	0,995
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	[illegible]	31,821	63,657
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	[illegible]	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	[illegible]	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	[illegible]	2,457	2,750
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	[illegible]	2,423	2,704
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	[illegible]	2,403	2,678
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	[illegible]	2,390	2,660
70	0,126	0,254	0,387	0,527	0,678	0,847	1,044	1,294	1,667	[illegible]	2,381	2,648
80	0,126	0,254	0,387	0,526	0,678	0,846	1,043	1,292	1,664	[illegible]	2,374	2,639
90	0,126	0,254	0,387	0,526	0,677	0,846	1,042	1,291	1,662	1,987	2,368	2,632
100	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1,984	2,364	2,626



El error típico de la media, es: $\sigma_{\bar{Y}} = \frac{S_{n-1}}{\sqrt{n}} = \frac{S_n}{\sqrt{n-1}}$

Si utilizamos la cuasi-desviación típica de la muestra, primero debemos calcularla:

$$S_{n-1}^2 = \frac{nS_n^2}{n-1} = \frac{100 \cdot 5^2}{99} = 25{,}2525 \rightarrow S_{n-1} = \sqrt{25{,}2525} = 5{,}02519$$

Luego:

$$\sigma_{\bar{Y}} = \frac{S_{n-1}}{\sqrt{n}} = \frac{5{,}02519}{\sqrt{100}} = 0{,}502519$$

Y si utilizamos la desviación típica de la muestra:

$$\sigma_{\bar{Y}} = \frac{S_n}{\sqrt{n-1}} = \frac{5}{\sqrt{100-1}} = 0{,}502519$$

Obteniendo el mismo resultado con las dos expresiones del error típico de la media. Y el error máximo solicitado es:

$$E_{max} = 1{,}984 \cdot 0{,}502519 = 0{,}997$$

Tema 2. CONTRASTE DE HIPÓTESIS EN LOS DISEÑOS DE UNA MUESTRA

2.1.- Introducción

En las investigaciones que parten del conocimiento proporcionado por los datos recogidos de una muestra el objetivo es inferir las características de la población, de la cual los datos recogidos constituyen una muestra representativa. En este tipo de investigaciones la hipótesis a contrastar especifica una característica de la población, como las siguientes:

- Si un determinado parámetro poblacional puede tomar un valor concreto.
- Si entre las variables medidas en la muestra existe correlación en la población.
- La forma de la distribución de la variable Y en la población.
- Si los datos observados en la muestra son independientes entre sí, etc.

Los dos primeros casos se incluyen dentro de los contrastes paramétricos que, en una primera aproximación diremos que son todos aquellos que se relacionan con el estudio de un parámetro poblacional (media, varianza, proporción, correlación, etc.). Por su parte, las dos hipótesis siguientes se englobarían dentro de los contrastes no paramétricos que no se relacionan con parámetros. En el Tema 9 estudiaremos con más detalle la distinción entre contrastes paramétricos y no paramétricos, pero en cualquiera de estos casos nos encontramos ante un diseño de investigación en el que se utiliza la información proporcionada por una muestra.

En el tema anterior se han visto los procedimientos de inferencia estadística basados en la determinación del intervalo de confianza de algunos parámetros poblacionales (media, varianza y proporción) que se apoyan en el conocimiento de la distribución muestral del correspondiente estadístico obtenido en la muestra. Igualmente se ha expuesto la metodología de los contrastes de hipótesis describiendo y explicando una serie de pasos que nos conducen a la toma de una decisión a partir del cálculo de un estadístico de contraste considerado como una medida de la discrepancia entre unos datos teóricos formulados en la hipótesis nula a contrastar y unos datos empíricos obtenidos en una muestra. Como veremos, tanto la estimación por intervalos como el estadístico de contraste pueden utilizarse para contrastar hipótesis sobre parámetros poblacionales, aunque en la práctica de la investigación es más habitual calcular esta discrepancia o estadístico de contraste. En cualquier caso, tanto un procedimiento como el otro se apoyan en el conocimiento de la distribución muestral (que corresponde a la función de distribución de probabilidad de estos estadísticos) en la que nos basaremos para tomar decisiones respecto a un valor hipotético que planteemos en la hipótesis nula para el parámetro poblacional. Al finalizar este capítulo el estudiante deberá alcanzar los objetivos que se exponen en el siguiente epígrafe:

2.2.- Objetivos

- Plantear las hipótesis en función de los objetivos de la investigación.
- Distinguir entre el contraste unilateral y bilateral.
- Seleccionar el estadístico de contraste más adecuado a las hipótesis planteadas.
- Conocer la distribución muestral del estadístico seleccionado.
- Realizar los cálculos oportunos para someter a contrastación empírica las hipótesis planteadas.
- Relacionar el intervalo de confianza con el estadístico de contraste
- Interpretar el nivel crítico *p*.
- Determinar e interpretar el, o los, valores críticos de la distribución muestral.
- Tomar una decisión respecto a las hipótesis planteadas.

2.3.- Contraste sobre la media poblacional

Por lo que vimos en el tema anterior, sabemos que el intervalo de confianza de un parámetro poblacional es un rango de valores definido a partir del estadístico obtenido en la muestra y delimitado por sus límites inferior y superior. Este intervalo cubrirá el valor del parámetro poblacional con una probabilidad de $1-\alpha$, denominada "nivel de confianza". En concreto, el intervalo de confianza de la media nos delimita entre qué dos valores se encontrará la media poblacional, μ, con una probabilidad o nivel de confianza, previamente fijado. En el tema anterior se expuso el siguiente ejemplo (Ej. 1.4) para calcular el intervalo de confianza de la media poblacional:

Ejemplo 2.1: En un experimento sobre atención, un psicólogo presenta durante 300 mseg un grupo de 16 letras del alfabeto (con una disposición de 4 filas y 4 columnas). Cada uno de los 12 sujetos que participan en el experimento debe verbalizar tantas letras como recuerde. El promedio de letras bien recordadas es de 7 y la desviación típica insesgada (cuasi-desviación típica) es de 1,3. Suponiendo que la distribución en la población es normal. ¿Entre qué límites se encontrará el verdadero número de palabras bien recordadas, con una probabilidad de 0,95?

En las condiciones que se plantean en este ejercicio, recordará el lector que el intervalo de confianza construido en torno a la media muestral, y que contendrá el valor del parámetro con una probabilidad de 0,95, se construye a partir de la distribución muestral de la media, la cual sabemos que es una distribución t de Student. En términos formales, para calcular este intervalo de confianza utilizamos la expresión del intervalo de confianza:

$$E_{max} = t_{n-1;1-\alpha/2}\frac{S_{n-1}}{\sqrt{n}}$$

$$P(\bar{Y} - E_{max} \leq \mu \leq \bar{Y} + E_{max})$$

Podemos observar los límites superior e inferior que se representan en la Figura 2.1 y que se obtienen sumando y restando a la media de la muestra el error máximo de estimación:

$$E_{max} = t_{n-1;1-\alpha/2}\frac{S_{n-1}}{\sqrt{n}} = 2{,}201\frac{1{,}3}{\sqrt{12}} = 0{,}826$$

$$l_i = \bar{Y} - E_{max} = 7 - 0{,}826 = 6{,}174$$

$$l_s = \bar{Y} + E_{max} = 7 + 0{,}826 = 7{,}826$$

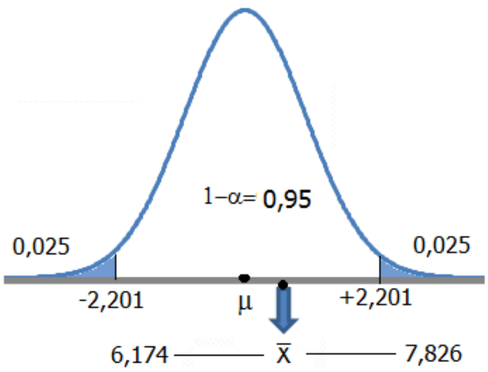


Figura 2.1: *Intervalo de confianza para la media*

El intervalo de confianza obtenido es (6,174; 7,826). Podemos afirmar al 95% de confianza que la media poblacional (desconocida) para el número de letras recordadas se encuentra entre los valores 6,174 y 7,826. Es decir, el intervalo de confianza de la media nos indica el conjunto de valores que podría tener la media poblacional con el nivel de confianza fijado previamente en el 95%. Por tanto, este intervalo se puede utilizar también para contrastar hipótesis sobre el valor que puede tomar este parámetro en la población. Así, si formulamos las hipótesis:

$$H_0: \mu = 8$$
$$H_1: \mu \neq 8$$

Tenemos que comprobar si el intervalo de confianza cubre (se solapa) o no al valor de la media poblacional planteada en la hipótesis nula. En caso negativo tomaremos la decisión de rechazar la H_0 con un nivel de significación, previamente fijado, que en este ejemplo es $\alpha = 0{,}05$. En caso contrario, diremos que no tenemos evidencia suficiente para rechazar la H_0, con un nivel de significación de $0{,}05$, y la mantendremos como provisionalmente verdadera hasta que no reunamos evidencia suficiente para rechazarla. De acuerdo con esta forma de proceder, podemos observar que el intervalo de confianza (6,174; 7,826) no cubre el valor de la media poblacional planteado en la hipótesis nula ($H_0: \mu = 8$), ya que el valor 8 planteado en esta hipótesis no se encuentra entre 6,174 y 7,826. Por tanto, rechazamos la hipótesis nula con un nivel de confianza del 95% (o con una probabilidad de 0,95).

Aunque el intervalo de confianza es un procedimiento para estimar parámetros poblacionales, puede también aplicarse para el contraste de hipótesis. Sin embargo, como se adelantaba en el tema anterior, es más frecuente aplicar otro procedimiento alternativo y que suele utilizarse habitualmente en los informes de investigación publicados. Este procedimiento alternativo consiste en obtener el estadístico de contraste como una medida más exacta de la discrepancia entre el valor planteado en la hipótesis nula y el valor obtenido en la muestra como estimación del parámetro. Esta medida de la discrepancia tiene una distribución de probabilidad conocida, por lo que suele ir acompañada de una probabilidad, a la que nos referiremos con el término de nivel crítico p, y que utilizaremos para tomar una decisión respecto a la hipótesis nula. Este nivel crítico p NO indica la probabilidad de que la H_0 sea verdadera, sino que nos informa sobre la probabilidad de obtener un resultado como el obtenido en la muestra, o más extremo, bajo el supuesto de que la hipótesis nula es verdadera. Se trata, por tanto, de una probabilidad condicionada: en el caso de que H_0 sea cierta (condición) nos indica la probabilidad de obtener un valor del estadístico de contraste igual o más extremo que el obtenido efectivamente en nuestra muestra. Esto se representa simbólicamente como: $P(Y \geq y_i | H_0) = p$. Lo que sigue a la barra vertical indica la condición de esta probabilidad (en este caso, la condición es que H_0 es cierta) y se lee como: la probabilidad de que, siendo H_0 cierta, se obtenga un valor del estadístico muestral igual o más extremo que el obtenido (y_i) es igual a p (el nivel crítico). Téngase en cuenta que la H_0 no es más que una conjetura sobre un valor del parámetro poblacional y difícilmente será verdadera. La finalidad de la investigación (y del contraste de hipótesis) es reunir información y evidencias suficientes para poder rechazarla.

2.3.1.- Conocida la varianza poblacional

Como se ha expuesto en los apartados anteriores, la inferencia estadística trata de estimar los parámetros poblacionales a partir de la información obtenida en la muestra. Sin embargo, en la actividad de la investigación real es poco probable que se conozca la varianza poblacional ya que conocerla supone que podemos acceder a todos los datos de la población en cuyo caso también podríamos calcular su media y sobraría cualquier tipo de inferencia o contraste sobre su valor. No obstante, existen casos en los que apoyados por los resultados de trabajos previos podemos asumir un determinado valor para la varianza poblacional como razonable. Si además podemos asumir que la distribución poblacional es normal, o bien trabajamos con muestras con $n \geq 30$, entonces la distribución muestral de la media es una distribución normal, y el estadístico de contraste para la media poblacional es:

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu_0}{\sigma / \sqrt{n}}$$

El estadístico Z se distribuye según la distribución normal tipificada, N(0;1). En esta ecuación:

- $\bar{Y}$ es la media obtenida en la muestra.
- μ_0 es el valor de la media poblacional formulado en la hipótesis nula.
- $\sigma_{\bar{Y}}$ es el error típico de la media o desviación típica de la distribución muestral de la media.
- σ es la desviación típica poblacional que suponemos conocida.
- n es el tamaño de la muestra que estamos utilizando para contrastar la hipótesis.

En consecuencia, el estadístico Z cuantifica la distancia de la media de la muestra, $\bar{Y}$, a la media poblacional, μ_0 , en unidades del error típico de la distribución muestral, $\sigma_{\bar{Y}}$.

Ejemplo 2.2: Por estudios previos conocemos que la población masculina de la tercera edad de una determinada Comunidad Autónoma, tiene un gasto medio en medicamentos de 215 euros/año con una desviación típica de 36 euros y queremos saber si la población femenina tiene el mismo gasto. Con tal finalidad analizamos el gasto medio de una muestra de 324 mujeres de la tercera edad de esa misma comunidad observando que la media es de 220 euros/año. Asumimos que esta variable se distribuye normalmente en la población, y que la varianza es la misma que en la población masculina. Fijando un nivel de confianza del 95%, contraste si el gasto de las mujeres es significativamente distinto de 215 euros/año.

Condiciones y supuestos: El estudio utiliza un diseño de una muestra de mujeres en la que la variable gasto medio se mide con escala de razón (variable cuantitativa) y sabemos que se distribuye normalmente en la población (aunque en este caso no haría falta este supuesto porque $n \geq 30$). Adicionalmente conocemos la desviación típica poblacional que es de 36 euros. Se trata de un contraste paramétrico bilateral ya que, *a priori*, no sabemos si el gasto de las mujeres es mayor o menor de 220 euros/año. Es decir, solo queremos contrastar que el gasto de las mujeres es diferente a esa cantidad, pero sin asumir que el sentido de esta diferencia sea positivo o negativo. Por tanto, debemos contemplar la posibilidad de que pueda serlo en un sentido u otro. Un contraste de este tipo se dice bilateral por razones obvias.

En definitiva, los supuestos que necesitamos en este caso son:

- Distribución normal en la población, o bien, $n \geq 30$.
- Varianza poblacional conocida.
- Variable dependiente con un nivel de medida de intervalo o razón.

Formulación de las hipótesis: La hipótesis de investigación es que "*las mujeres tienen un gasto medio en medicamentos distinto a los 215 euros/año*". Es decir, el investigador se ha planteado este estudio porque tiene razones para suponer que el gasto de medicamentos entre hombres y mujeres es distinto (*v.g.*, sabemos que por término medio las mujeres disfrutan de una mayor longevidad que los hombres pero con peor salud) y por ello plantea esta hipótesis de investigación. Sin embargo, normalmente la hipótesis estadística nula que plantea debe ser la contraria a su hipótesis de investigación. Por ello, la hipótesis nula debe plantearse en el sentido de que el gasto de las mujeres es de 215 euros/año, igual al de hombres, y la hipótesis alternativa que el gasto medio de las mujeres es un valor distinto a éste, es decir:

$$H_0: \mu = 215$$
$$H_1: \mu \neq 215$$

Partimos de que, provisionalmente, la hipótesis nula es verdadera, es decir, que las mujeres tienen un gasto de 215 euros/año y se trata de encontrar evidencia contra esta hipótesis a partir de la información proporcionada por una muestra representativa. Inicialmente se observa que, efectivamente, las mujeres parece que tienen un gasto diferente, pero la pregunta es: ¿la diferencia de 5 euros entre el valor observado en la muestra y el que planteamos en la hipótesis nula evidencia realmente un gasto distinto o son debidas a fluctuaciones aleatorias? El rechazo de la hipótesis nula y la consiguiente aceptación de la hipótesis alternativa, se deberá a que la diferencia observada es "*estadísticamente significativa*", es decir, es una diferencia real y evidente que no puede atribuirse al azar, a fluctuaciones aleatorias debidas al muestreo.

Estadístico de contraste: Para contrastar nuestra hipótesis vamos a calcular la discrepancia entre la evidencia observada de que el gasto medio es de 220 euros en la muestra de mujeres con el valor hipotéticamente establecido para la población general que plantea un gasto medio de 215 euros.

Calcularemos primero el error típico de la media (es decir, la desviación típica de la distribución muestral de todas las medias posibles en muestras de tamaño: $n = 324$):

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{36}{\sqrt{324}} = 2$$

Como la variable "*gasto anual en medicamentos*" se distribuye normalmente en la población y conocemos la desviación típica poblacional, la distribución muestral de la media es normal y el estadístico de contraste, como medida de esta discrepancia, es:

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}} = \frac{220 - 215}{2} = 2{,}5 \qquad (p = 0{,}0124)$$

Regla de decisión: En este contraste bilateral y trabajando con un nivel de confianza del 95%, los valores críticos a partir de los cuales rechazamos la hipótesis nula son $Z = \pm 1{,}96$ (Figura 2.2). Estos valores representan la máxima diferencia, en un sentido o en otro, atribuible al azar que puede existir entre los datos empíricos observados en la muestra y los datos teóricos que planteamos en la hipótesis nula. En la muestra el valor observado es 220 euros/año y el valor hipotético planteado es de 215 euros/año. Esta diferencia corresponde a 2,5 desviaciones típicas de la distribución muestral.

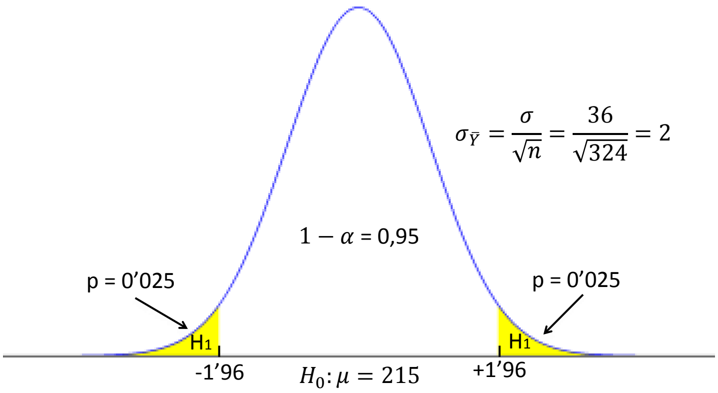


Figura 2.2: *Distribución muestral de la media y regiones de decisión para $H_0: \mu = 215$ para un nivel de confianza del 95%*

Conclusión: Con un nivel de confianza del 95%, el valor de este estadístico de contraste (Z = 2,5) sobrepasa la máxima diferencia que cabe esperar por simple azar que es de 1,96. Por tanto, debemos rechazar la hipótesis nula con un nivel de confianza del 95%. De otra forma, al valor del estadístico de contraste obtenido de Z=2,5 le corresponde un nivel crítico p de 0,0124 (Figura 2.3). Esta probabilidad[1] indica que, suponiendo verdadera la hipótesis de que las mujeres tienen un gasto medio de 215 euros/año, la probabilidad de observar un gasto medio de 220 euros/año o más extremo en una muestra de 324 mujeres es de 0,0124. Esta probabilidad es muy pequeña y menor que el nivel de significación "alfa" fijado en 0,05 ($0{,}0124 < 0{,}05 = \alpha$) lo que nos lleva a rechazar la hipótesis nula.

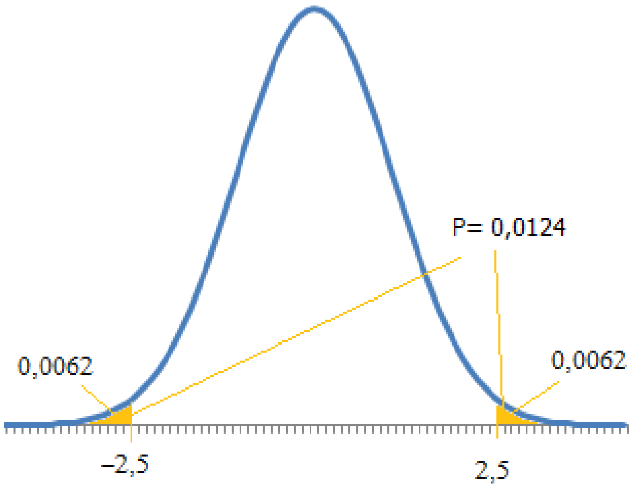


Figura 2.3: *Nivel crítico p para un contraste bilateral*

Interpretación: a la vista de los cálculos y de nuestra conclusión podemos decir que, con un nivel de confianza del 95%, el gasto de las mujeres difiere significativamente de 215 euros/año, que es el que realizan los hombres.

[1] Se busca en la tabla de la distribución normal, la probabilidad de P(Z≤-2,5) que es 0,0062. Al tratarse de un contraste bilateral tenemos que sumar la P(Z≥2,5) = 0,0062. La suma de estas dos probabilidades es el nivel crítico p resultante.

Observe el lector que igual de lícito sería afirmar que, con un nivel de confianza del 99%, no hay evidencia suficiente para rechazar la hipótesis nula ya que el valor de $p = 0{,}0124$ es mayor que el nivel de significación “alfa” de $0{,}01$. Es decir que la diferencia encontrada es significativa con “alfa”=0,05 pero no lo es con un nivel de significación de $0{,}01$. Dejamos al lector que llegue a la misma conclusión calculando nuevamente y comparando el valor del estadístico de contraste con los valores críticos para un nivel de confianza del 99%.

Estas conclusiones, aparentemente contradictorias, ponen de manifiesto la importancia de la replicación de la investigación para añadir más evidencia a favor o en contra de la hipótesis y, por otra parte, la exigencia de reflejar -en cualquier trabajo de investigación- el valor del estadístico de contraste y el nivel crítico p ($Z = 2{,}5$; $p = 0{,}0124$) con la finalidad de que el lector pueda interpretar por si mismo la magnitud de la discrepancia entre los datos y la hipótesis nula planteada y la seguridad a la hora de aceptar o rechazar la hipótesis[2].

2.3.2.- Desconocida la varianza poblacional

Ya se ha comentado que, en la práctica de la investigación social y de la Psicología, habitualmente se desconocen los parámetros poblacionales por lo que hay que estimarlos a partir de los estadísticos muestrales. Se estudió en el Tema 8 de Introducción al Análisis de datos y también en el tema anterior de este mismo texto, que si se desconoce la varianza poblacional y la forma de la distribución de la variable X en la población entonces la distribución muestral de la media es la distribución t de “Student”, siempre que podamos asumir que la distribución en la población es normal, o $n \geq 30$. En estas circunstancias el estadístico de contraste, como medida de la discrepancia, es:

$$T = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$

Que se distribuye según la t de “Student” con $n - 1$ grados de libertad y donde $\hat{\sigma}$ es el estimador de la desviación típica poblacional que se puede realizar a partir de la varianza o de la cuasi-varianza de la muestra, como se exponía en el punto 1.3.3 del Tema 1 y que nos conduce a las siguientes expresiones finales:

$$T = \frac{\bar{Y} - \mu_0}{\hat{\sigma}/\sqrt{n}} \Rightarrow \begin{cases} T = \dfrac{\bar{Y} - \mu_0}{S_{n-1}/\sqrt{n}} \\[2ex] T = \dfrac{\bar{Y} - \mu_0}{S_n/\sqrt{n-1}} \end{cases}$$

Aplicaremos a los datos del ejemplo 2.1 el contraste de hipótesis correspondiente:

[2] En este punto remitimos al lector al último epígrafe del Tema 1 que expone la relación entre el nivel crítico p y los errores del contraste.

Ejemplo 2.3: En un experimento sobre atención, un psicólogo presenta durante 300 mseg un grupo de 16 letras del alfabeto (con una disposición de 4 filas y 4 columnas). Cada uno de los 12 sujetos que participan en el experimento debe verbalizar tantas letras como recuerde. El promedio obtenido de letras bien recordadas es de 7 y la desviación típica insesgada (cuasi-desviación típica) de la muestra es de 1,3. Sabiendo que el recuerdo es una variable que se distribuye normalmente en la población y fijando el nivel de significación en 0,05, ¿Puede ser 8 la media de letras recordadas?

Condiciones y supuestos: una muestra aleatoria en la que recogemos datos medidos al menos con escala de intervalo y sabemos que la variable se distribuye normalmente en la población con varianza desconocida. En general, las condiciones necesarias para aplicar este contraste son:

- Variable dependiente con un nivel de medida de intervalo o razón.
- Distribución normal en la población o $n \geq 30$.
- Varianza poblacional desconocida.

Formulación de hipótesis: Se plantea un contraste bilateral

$$H_0: \mu = 8$$
$$H_1: \mu \neq 8$$

Partimos de que la H_0 es verdadera, es decir, que la media de palabras recordadas en este tipo de pruebas es de 8, y se trata de ver si los datos recogidos en una investigación bien diseñada y utilizando una muestra aleatoria arrojan evidencia a favor o en contra de la hipótesis nula.

El estadístico de contraste o discrepancia entre el estimador (media de la muestra) y el valor del parámetro formulado en la hipótesis nula, μ_0, es :

$$T = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu_0}{\frac{S_{n-1}}{\sqrt{n}}} = \frac{7 - 8}{1{,}3 / \sqrt{12}} = \frac{-1}{0{,}375} = -2{,}66$$

Obsérvese que, al desconocer la varianza poblacional, hemos utilizado S_{n-1}=1,3 muestral para calcular el estadístico de contraste.

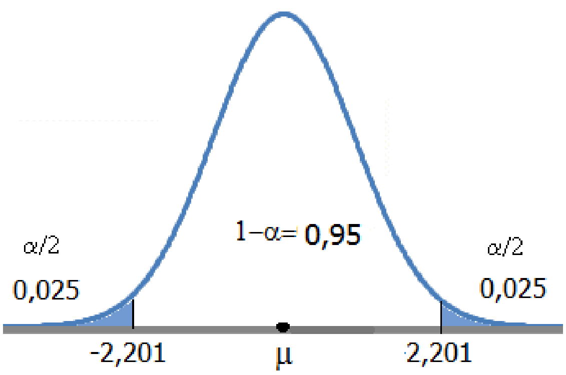


Figura 2.4: *Valores críticos de la distribución muestral para un nivel de confianza del 95%*

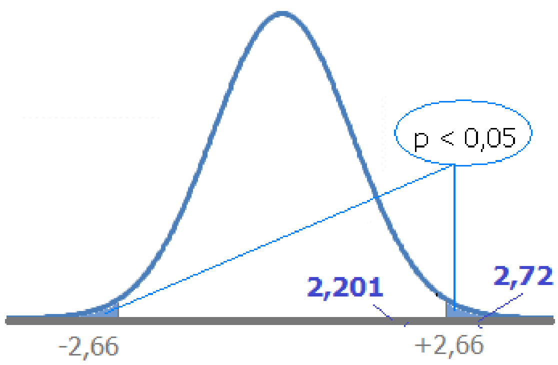
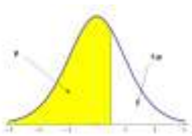


Figura 2.5: *Nivel crítico p asociado al estadístico de contraste t=2,66 en un contraste bilateral*

Regla de decisión: Con un nivel de confianza del 95% en un contraste bilateral, la máxima discrepancia que cabe esperar por simple azar entre el estimador y el valor planteado en la hipótesis nula es 2,201 (valores críticos). El valor del estadístico de contraste obtenido, supera este valor máximo (véase la Figura 2.4) lo que nos lleva a rechazar la hipótesis nula. El nivel crítico p asociado a este estadístico de contraste no aparece explícitamente en la tabla de la distribución t con 11 grados de libertad, pero podemos ver que es menor de 0,05 que resulta menor que el nivel de significación fijado en 0,05 (véase la Figura 2.5). La forma de buscar esta probabilidad en la tabla es la siguiente:



g.l.	0,550	0,600	0,650	0,700	0,750	0,800	0,850	0,900	0,950	0,975	0,990	0,995
	-	-	-	-	-	-	-	-	-	-	-	-
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,093	1,372	1,812	2,228	2,764	3,169
11	0,1286	0,2596	0,3956	0,5399	0,6974	0,8755	1,088	1,363	1,796	**2,201**	**2,718**	3,106
12	0,1283	0,2590	0,3947	0,5386	0,6955	0,8726	1,083	1,356	1,782	2,179	2,681	3,055
	-	-	-	-	-	-	-	-	-	-	-	-

Con 11 gl, el valor 2,66 se encuentra entre 2,201 y 2,718, (fig: 2.5 y tabla de t) por tanto: $0{,}025 > p > 0{,}01$ en una cola de la distribución y $0{,}05 > p > 0{,}02$ utilizando las dos colas de la distribución

Interpretación: A partir de la evidencia que proporcionan los datos de la investigación, debemos rechazar la hipótesis de que el número medio de palabras recordadas es de 8 con un nivel de confianza del 95%.

2.4.- Contraste sobre la proporción poblacional

El contraste paramétrico de hipótesis para una proporción poblacional sigue la misma lógica y procedimiento que el seguido para el contraste de la media. Sabemos que la proporción, o frecuencia relativa de aparición de una observación, es el cociente entre el número de veces que aparece la observación y el número total de observaciones. En el tema anterior se ha visto que la distribución muestral de la proporción es una distribución binomial y como se expuso en la asignatura de Introducción al Análisis de Datos, de primer curso. La distribución binomial se aproxima a la normal cuando el tamaño de la muestra es grande ($n > 25$ ó $np > 5$). En esta distribución muestral, la media y desviación típica (o error típico de la proporción) valen:

$$\mu_p = \pi$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

A partir de este supuesto y considerando, como se ha visto en el tema anterior, que la proporción observada en la muestra, p, es el estimador insesgado de la proporción poblacional, π, el intervalo de confianza para la proporción poblacional a partir de la proporción observada en una muestra se obtiene sumando y restando a la proporción observada en la muestra el error máximo de estimación (Figura 2.6):

$$E_{max} = z_{1-\alpha/2}\sigma_P = z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \qquad l_i = p - E_{max} \qquad l_s = p + E_{max}$$

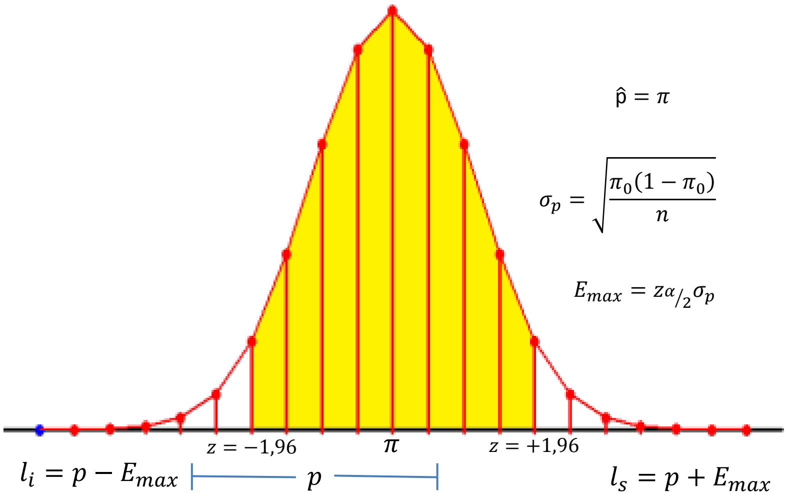


Figura 2.6: *Intervalo de confianza para la proporción poblacional, π.*

Igual que razonábamos para el caso de la media, para contrastar una hipótesis referida a un valor hipotéticamente establecido como proporción poblacional, π, podemos determinar el intervalo de confianza y comprobar si el valor planteado en la hipótesis nula se encuentra incluido o no por el intervalo.

De forma similar se puede determinar un estadístico de contraste para cuantificar la discrepancia entre el valor observado en la muestra y el planteado en la hipótesis nula. Para el caso de la proporción y sabiendo que la distribución muestral del estadístico, p, se aproxima a la normal cuando las muestras son grandes ($n > 25$ ó $np > 5$) , este estadístico es:

$$Z = \frac{p - \pi_0}{\sigma_P} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

Si la hipótesis nula es falsa esta discrepancia debe superar el valor crítico de la distribución muestral. De igual forma, el nivel crítico p asociado a esta discrepancia debe ser menor que el nivel de significación, α, para poder rechazar la hipótesis nula. En caso contrario no tendremos evidencia suficiente para poder rechazar la hipótesis nula planteada.

Ejemplo 2.4: Un investigador de estudios de mercado cree que más del 20% de los adolescentes cambian de móvil cada año. Con esta finalidad realiza una encuesta sobre una muestra de 150 adolescentes observando que 39 de ellos afirman haber cambiado de móvil en el último año. Con un nivel de confianza del 99%, ¿podemos admitir la hipótesis del investigador?

Condiciones y supuestos: El estudio utiliza un diseño de una muestra de 150 adolescentes en la que la variable "cambiar de móvil", es cualitativa y dicotómica ya que la respuesta solo puede ser "si" o "no". Cuando contabilizamos en cada muestra el número de participantes que contestan sí o no, entonces esta variable tiene una distribución binomial que, en las condiciones de este ejemplo, se aproxima a la normal por tratarse de una muestra grande. El investigador quiere demostrar que el porcentaje de adolescentes que cambia de móvil cada año es superior al 20%.

Planteamiento de las hipótesis: Se trata de un contraste unilateral ya que si la hipótesis alternativa dice que "la proporción supera el 0,20", la hipótesis nula dice que "la proporción es igual o no supera el 0,20". Por otra parte, observamos que a partir de los datos de la muestra el porcentaje de adolescentes que cambian de móvil es del 26% (o una proporción de 0,26). La hipótesis nula formula que la diferencia entre el valor observado en la muestra (26%) y el valor planteado para la proporción poblacional (20%) es nula. En otras palabras, que esta diferencia se debe a las fluctuaciones aleatorias porque la proporción poblacional es del 20% o menor.

$$H_0: \pi_0 \leq 0{,}20$$
$$H_1: \pi_1 > 0{,}20$$

Estadístico de contraste: Calculamos la discrepancia entre p y π medida en unidades de error típico de la proporción (asumiendo que H_0 es cierta).

$$Z = \frac{p - \pi_0}{\sigma_P} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} = \frac{0{,}26 - 0{,}20}{\sqrt{\frac{0{,}20 \cdot 0{,}8}{150}}} = 1{,}837 \approx 1{,}84$$

Siendo la proporción poblacional $p = 0{,}26$ un valor de la distribución muestral de la proporción, el estadístico $Z = 1{,}84$ indica que la distancia de $p = 0{,}26$ a $\pi_0 = 0{,}20$ es de $1{,}84$ desviaciones típicas de la distribución muestral.

Regla de decisión: Con un nivel de confianza del 99% y en un contraste unilateral, el valor crítico para rechazar la hipótesis nula es 2,33 (véase la Figura 2.7).

De otra forma, el nivel crítico p[3] asociado al estadístico de contraste obtenido es 0,0329 que es una probabilidad mayor que el nivel de significación establecido a priori $\alpha = 0{,}01$.

[3] Debe buscarse en las tablas de la distribución normal la probabilidad de obtener puntuaciones Z mayores que 1, 84. Recuérdese que las tablas de la distribución normal proporcionan probabilidades por debajo de una puntuación Z determinada. Por tanto, este valor es P(Z≥1,84)=1-P(Z≤1,84)= 1-0,9671=0,0329

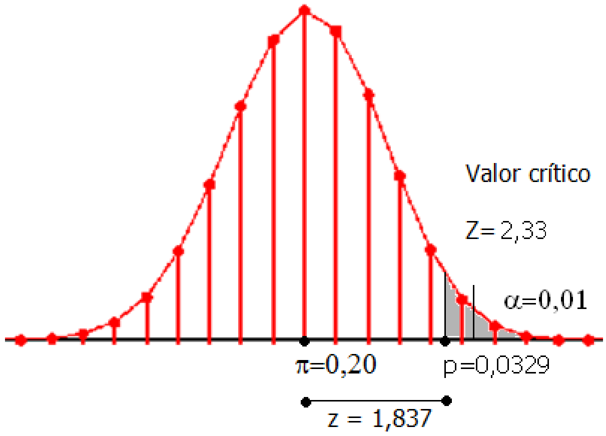


Figura 2.7: *Estadístico de contraste, nivel de significación y nivel crítico p para el contraste unilateral derecho de H_0: π=0,20*

Conclusión. Como el estadístico de contraste -o discrepancia encontrada entre los valores $p = 0{,}26$ y $\pi_0 = 0{,}20$ de $1{,}84$ no supera la máxima diferencia que puede esperarse por simple azar (el valor crítico 2,33), no tenemos evidencia suficiente para rechazar la hipótesis nula. De otra forma, el nivel crítico p de $0{,}0329$ es mayor que el nivel de significación $\alpha = 0{,}01$ por lo que no podemos rechazar la hipótesis nula.

Interpretación: A la luz de los datos obtenidos por el investigador, con un nivel de confianza del 99%, no hay evidencia suficiente para asumir que más del 20% de los adolescentes cambian de móvil cada año. Obsérvese, como los resultados si serían significativos si se adoptara un nivel de confianza del 95%.

2.5.- Contraste de hipótesis sobre la varianza poblacional

Ya sabemos que la inferencia y el contraste de hipótesis sobre cualquier parámetro requieren conocer cómo es su distribución muestral. En el tema anterior, vimos que si de una población donde la variable Y se distribuye normalmente con media μ y varianza σ^2 , se extraen todas las posibles muestras del mismo tipo y tamaño, y en cada muestra calculamos sus varianzas S_n^2, entonces se puede demostrar que la variable aleatoria:

$$\frac{nS_n^2}{\sigma^2}$$

Sigue una distribución chi-cuadrado con $n - 1$ grados de libertad (véase la Figura 2.8).

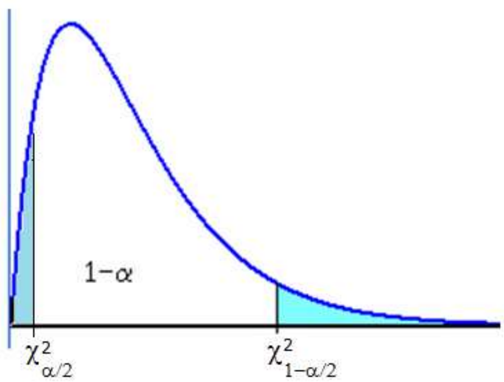


Figura 2.8: *Distribución muestral de la varianza y nivel de confianza*

De la misma forma, y por la relación existente entre varianza y cuasi-varianza, la expresión anterior también se puede expresar con referencia a la cuasi-varianza muestral, y sería:

$$\frac{(n-1)S_{n-1}^2}{\sigma^2}$$

que también se distribuye según chi-cuadrado con $n-1$ grados de libertad.

Con referencia a este principio vimos que el intervalo de confianza para la varianza poblacional viene definido por sus límites inferior y superior, que se calculan mediante la expresión:

$$l_i = \frac{nS_n^2}{\chi^2_{n-1;1-\alpha/2}} = \frac{(n-1)S_{n-1}^2}{\chi^2_{n-1;1-\alpha/2}}; \qquad l_s = \frac{nS_n^2}{\chi^2_{n-1;\alpha/2}} = \frac{(n-1)S_{n-1}^2}{\chi^2_{n-1;\alpha/2}}$$

Estos límites delimitan los valores entre los que se encontrará la varianza poblacional, con una probabilidad de $1-\alpha$.

Por otra parte, el estadístico de contraste o medida de la discrepancia entre el estimador y el parámetro es un cociente que recoge ambos valores, y adoptan las siguientes expresiones en función de que realizamos el cálculo con la varianza de la muestra o con la cuasi-varianza:

A partir de la varianza de la muestra:

$$\chi^2 = \frac{nS_n^2}{\sigma_0^2}$$

Y a partir de la cuasi-varianza de la muestra:

$$\chi^2 = \frac{(n-1)S_{n-1}^2}{\sigma_0^2}$$

Siendo σ_0 la desviación típica poblacional postulada en H_0. Con esta medida de la discrepancia, y a partir de la varianza obtenida en una muestra, comprobaremos la hipótesis acerca de la varianza poblacional de una variable normalmente distribuida. En este contraste se pueden dar los tres casos que pueden verse en la Figura 2.9:

Contraste bilateral	Unilateral derecho	Unilateral izquierdo:
$H_0: \sigma^2 = \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$	$H_0: \sigma^2 \geq \sigma_0^2$
$H_1: \sigma^2 \neq \sigma_0^2$	$H_1: \sigma^2 > \sigma_0^2$	$H_1: \sigma^2 < \sigma_0^2$

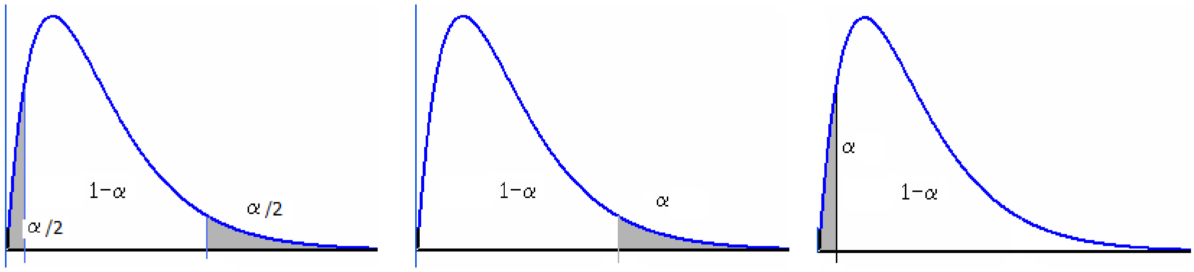


Figura 2.9: *Regiones de rechazo de la hipótesis nula en contrastes bilateral y unilaterales*.

Y siguiendo los pasos establecidos para todo contraste de hipótesis tal y como veremos en el siguiente ejemplo:

Ejemplo 2.5: El manual de un test para detectar niños con problemas de aprendizaje afirma que las puntuaciones del test se distribuyen normalmente y que la varianza de las puntuaciones disminuye con la edad, tomando el valor de 18,1 para los niños promedio de 5 años. Un psicólogo infantil considera que actualmente esta variabilidad ha aumentado y para probarlo, utiliza una muestra de 25 niños de 5 años a los que aplica el test obteniendo una desviación típica sesgada de 4,9 puntos. Trabajando con un nivel de significación de 0,01, contraste la hipótesis del investigador.

Condiciones y supuestos: El estudio utiliza un diseño de una muestra aleatoria de 25 niños a los que se les pasa un test. Asumimos que estas puntuaciones se miden, al menos en una escala de intervalo, y se distribuyen normalmente en la población con varianza 18,1, tal como indica el baremo del test. En la muestra se obtiene una desviación típica sesgada de 4,9.

En general, las condiciones que tienen que cumplirse son:

- Variable dependiente con un nivel de medida de intervalo o razón.
- Distribución normal en la población, o bien n ≥30.

Formulación de hipótesis: El investigador quiere probar que la varianza del test en los niños de 5 años es ahora mayor de 18,1 como afirma el manual. Por consiguiente, concreta una hipótesis nula contraria a la hipótesis que él desea probar de tal forma que si consigue rechazarla con los datos de la investigación, lo está haciendo con un elevado grado de confianza. Se trata por tanto de un contraste unilateral derecho.

$$H_0: \sigma^2 \leq 18{,}1$$
$$H_1: \sigma^2 > 18{,}1$$

Estadístico de contraste: Conociendo la desviación típica sesgada de la muestra, S_n, el estadístico de contraste, es:

$$\chi^2 = \frac{nS_n^2}{\sigma_0^2} = \frac{25 \cdot 4{,}9^2}{18{,}1} = 33{,}1629$$

Regla de decisión: En la distribución chi-cuadrado con n-1 = 25-1= 24 grados de libertad y un nivel de confianza del 99%, el valor crítico para rechazar la hipótesis nula es 42,98 (véase la Figura 2.10). El nivel crítico p hay que buscarlo en la tabla de la distribución chi-cuadrado con 24 gl y un chi-cuadrado igual a 33,16 y por aproximación es p= 0,10 que corresponde al valor 33,20 que es el que nos aparece más cercano a nuestro estadístico de contraste 33,1629. Rechazaremos la hipótesis nula si el estadístico de contraste es mayor que el valor crítico de 42,98 o si el nivel crítico p es menor que el nivel de significación de 0,01.

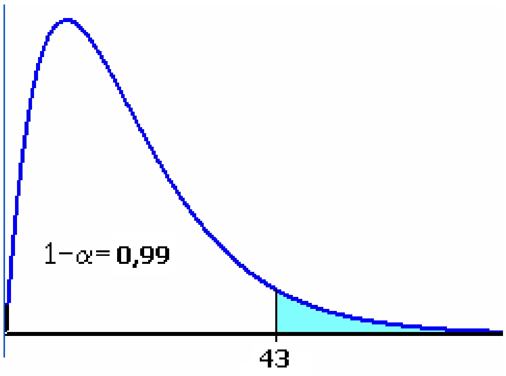


Figura 2.10: *Valor crítico de la distribución chi-cuadrado con 24 gl y un nivel de significación de 0,01*.

g.l.	-	0,750	0,900	0,950	0,975	0,990	0,995	0,999
-	-	-	-	-	-	-	-	-
23	-	27,14	32,01	35,17	38,08	41,64	44,18	49,73
24	-	28,24	**33,20**	36,42	39,36	**42,98**	45,56	51,18
25	-	29,34	34,38	37,65	40,65	44,31	46,93	52,62
-	-	-	-	-	-	-	-	-

Conclusión: Como el estadístico de contraste obtenido no supera el valor crítico, la evidencia aportada por nuestra muestra de estudio no resulta suficiente para rechazar la hipótesis nula. Igualmente, siendo el estadístico de contraste 33,16 buscamos en la distribución chi-cuadrado con 24 gl, el valor más próximo a éste, que es 33,20 y que se corresponde con un nivel crítico p de 0,10 que es mayor que el nivel de significación fijado en el 0,01. Por tando no tenemos evidencia suficiente para rechazar la hipótesis nula.

Interpretación: No tenemos evidencia suficiente para afirmar que la variabilidad de las puntuaciones obtenidas en el test para detectar problemas de aprendizaje en los niños de 5 es ahora mayor que la que figura en el manual del test.

Aclaración: Si el contraste se hubiera realizado a partir de la cuasi-varianza muestral en lugar de la varianza, el resultado del estadístico de contraste hubiera sido el mismo. Veámoslo, la cuasi-varianza de la muestra es:

$$S_{n-1}^2 = \frac{nS_n^2}{n-1} = \frac{25 \times 4{,}9^2}{25-1} = 25{,}01$$

Y en este caso, el estadístico de contraste, toma el mismo valor:

$$\chi^2 = \frac{(n-1)S_{n-1}^2}{\sigma_0^2} = \frac{(25-1)25{,}01}{18{,}1} = 33{,}16$$

2.6.- Cálculo de la Potencia del contraste

En el tema anterior se han expuesto los errores que se pueden cometer en todo contraste de hipótesis: rechazar una hipótesis nula que es verdadera (error tipo I o α) y no rechazar una hipótesis nula que es falsa (error tipo II o β). Allí se comentó que la potencia de un contraste estadístico es el complementario del error tipo II ($1-\beta$). Un aspecto importante de la investigación es conocer el valor que adopta la potencia ya que representa la probabilidad de poder detectar el efecto de interés que estamos buscando.

En este apartado vamos a ver, apoyándonos en el desarrollo de dos ejemplos, el procedimiento para calcular la potencia de un contraste paramétrico referido a la media y a la proporción poblacional en el diseño de una muestra. Pero téngase en cuenta que la potencia de un contraste se puede calcular en todo tipo de contraste de hipótesis, sea de la naturaleza que sea y para todo tipo de diseño de investigación de los que veremos a lo largo de este curso.

Ejemplo 2.6: Supongamos que la duración media de una lámpara de bajo consumo de una determinada marca es de 1000 horas con un desviación típica de 220 horas. La empresa que las fabrica introduce un nuevo proceso de fabricación y afirma que la vida media de las nuevas es superior a las antiguas. Vamos a suponer que como hipótesis alternativa única se plantea un promedio de duración de 1060 horas. Tomando un nivel de significación del 5%, determinar el error tipo II y la potencia de la prueba, si el estudio se realizara con un muestra de 100 lámparas.

Como se ha mencionado con anterioridad, cuando se realiza un contraste de hipótesis, ambas hipótesis han de ser exhaustivas y mutuamente excluyentes, no obstante, para calcular la potencia del contraste, se han de plantear dos hipótesis en las que sólo figura el signo "igual". En este ejemplo, las hipótesis son:

$$H_0: \mu = 1000$$
$$H_1: \mu = 1060$$

En la Figura 2.11 se plantea gráficamente la situación de este contraste unilateral. Una vez establecido en la distribución de la hipótesis nula el error tipo I (0,05) y que se corresponde con un valor crítico de Z = 1,64, se trata de determinar a qué valor corresponde en la distribución muestral de las duraciones medias de las lámparas antiguas. El resultado se obtiene de:

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}} \rightarrow 1{,}64 = \frac{\bar{Y} - 1000}{220/\sqrt{100}} \rightarrow \bar{Y} = 1000 + 1{,}64\frac{220}{\sqrt{100}} = 1036{,}1$$

Por tanto, una duración media de más de 1036,1 horas en una muestra de 100 lámparas nos conduciría a rechazar H_0. Para determinar el error tipo II (beta), debemos saber la puntuación típica que corresponde a esta media muestral pero referido a la media de la distribución de H_1, es decir, al valor planteado como hipótesis alternativa establecido en $\mu_1 = 1060$.

$$Z = \frac{1036{,}1 - 1060}{220/\sqrt{100}} = -1{,}09$$

En la distribución de H_1, la probabilidad de obtener un valor de Z igual o menor de -1,09 es 0,1379, que es la probabilidad de cometer un error tipo II. Y su complementario 1-0,1379=0,8621 es la potencia del contraste o probabilidad de que los resultados de la investigación permitan rechazar la hipótesis nula cuando es realmente falsa.

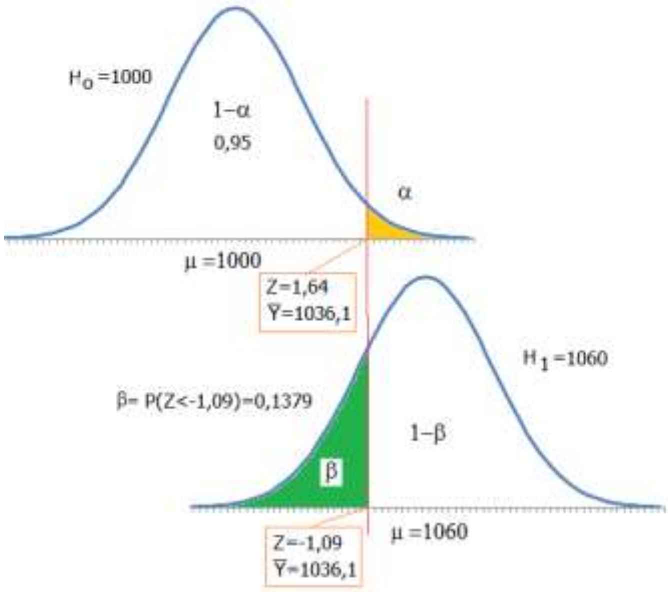


Figura 2.11: *Representación gráfica del ejemplo 2.6*

Resumiendo: si se rechazara la hipótesis nula de que el promedio de duración es de 1000 horas, pero en realidad esta hipótesis fuera verdadera (es decir, el nuevo proceso de fabricación no alarga la duración) entonces estaríamos cometiendo un error (tipo I) del 5%. Por otro lado, si se acepta la hipótesis nula, pero la alternativa es la verdadera, la probabilidad de cometer este error (tipo II) es del 13,79%. Por tanto, la potencia de la prueba es del 86,21% (1 – 0,1379 = 0,8621).

Realizaremos otro cálculo de la potencia del contraste recurriendo al ejemplo 1.8 del tema anterior que se resolvía aplicando la distribución binomial.

Ejemplo 2.7. Para contrastar la presunta "*habilidad detectora*" de la dama se preparan 16 tazas de té, siguiendo ambos procedimientos: en ocho se vierte primero la leche, y en otros ocho se vierte primero la infusión. La presentación se realiza al azar y la dama sólo tiene que decir cuál ha sido el procedimiento (primero la leche y después el té, o a la inversa). Supongamos, por ejemplo, que la dama acierta en 12 ocasiones. Vamos a utilizar este dato como hipótesis alternativa, para calcular la potencia de un contraste unilateral derecho con un nivel de significación de 0,05, es decir, veremos qué sucede bajo la hipótesis nula de que la señora no puede realizar esta discriminación ($\pi_0 = 0{,}5$) en relación a lo que sucedería si la señora puede, efectivamente, realizarla con una probabilidad superior al azar que, en este caso, hemos supuesto igual a 0,75.

$$H_0: \pi = 0{,}50$$
$$H_1: \pi = 0{,}75$$

Como vimos en el tema 1, concluimos que la dama no tiene esa habilidad si su probabilidad de acertar en n=16 ensayos es de aproximadamente 8 ocasiones (el 50% de los casos). ¿A partir de qué número de aciertos procederíamos a rechazar la hipótesis nula con un nivel de significación de 0,05?. Consultamos en la tabla de la distribución binomial para n=16 y p=0,5 el número de aciertos superiores a 8, el 50%, y cuya suma sea al menos igual o menor que el alfa fijado. Vemos que sólo rechazaríamos la hipótesis nula si la dama acierta en 12 o más ocasiones, ya que la suma de estas probabilidades vale:

$$P(y \geq 12) = P(y = 12) + P(y = 13) + P(y = 14) + P(y = 15) + P(y = 16) =$$
$$= 0{,}0278 + 0{,}0085 + 0{,}0018 + 0{,}0002 + 0{,}0000 = 0{,}0383 \leq \alpha$$

x	P=0,5
-	-
11	0,0667
12	0,0278
13	0,0085
14	0,0018
15	0,0002
16	0,0000

Tabla de la distribución binomial para N=16 y p=0,5

x	P=0,75
-	-
11	0,1802
12	0,2252
13	0,2079
14	0,1336
15	0,0534
16	0,0100

Tabla de la distribución binomial para N=16 y p=0,75

Sabiendo que la potencia corresponde a la probabilidad de rechazar la hipótesis nula cuando es falsa, es decir, cuando la dama sí tiene esa habilidad y que esta decisión se toma cuando es capaz de acertar en 12 o más ocasiones, la potencia del contraste se calcula procediendo de la siguiente forma:

Se calcula la probabilidad de acertar en 12 o más ocasiones cuando la dama sí tiene esa habilidad que, de acuerdo con la hipótesis alternativa hemos fijado en p=0,75[4] . Por consiguiente, acudimos a la tabla de la distribución binomial con n=16, p=0,75 (véase la Fig 2.12) y sumamos las probabilidades de:

$$P(y \geq 12) = P(y = 12) + P(y = 13) + P(y = 14) + P(y = 15) + P(y = 16) =$$
$$= 0{,}2252 + 0{,}2079 + 0{,}1336 + 0{,}0534 + 0{,}0100 = 0{,}6302 = 1 - \beta$$

[4] La tabla de la distribución binomial no refleja el valor p=0,75 pero la forma de razonar es la siguiente: Si la probabilidad de acertar es 0,75, la de fallar es 0,25. Por tanto, la probabilidad de tener 12 aciertos (con p=0,75) en N=16 ensayos es la misma que la probabilidad de tener 4 fallos (con p=0,25) en esos mismos 16 ensayos. Y esta probabilidad de p=0,25 sí que figura en la tabla binomial.

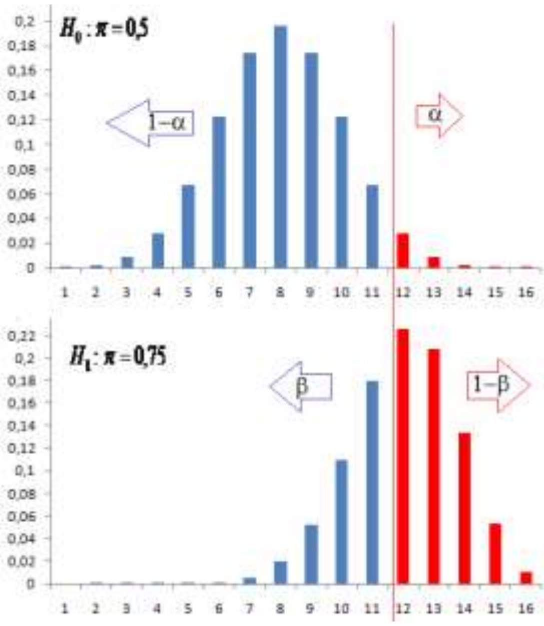


Figura 2.12: *Representación gráfica del ejemplo 2.7*

A partir de estos ejemplos, el lector puede deducir que para calcular la potencia de un contraste se necesita que la hipótesis nula y la alternativa sean simples, es decir, que establezcan un único valor como parámetro poblacional en vez de un rango de valores como hacíamos en el contraste de hipótesis. En los ejemplos que se ha desarrollado, y en el caso concreto de la media, los cálculos se han realizado para los valores $\mu = 1000$, en la H_0, y $\mu = 1060$ en la H_1. Cuando la hipótesis alternativa es compuesta, es decir, plantea más de un valor como media poblacional (H_1: $\mu \neq 1000$) la **potencia del contraste**, o *probabilidad de rechazar una hipótesis nula que en realidad es falsa*, varía en función de dos factores: la distancia entre el valor de la hipótesis nula y la hipótesis alternativa, y el tamaño muestral. De este modo para un mismo valor del error tipo I, se pueden confeccionar lo que se denominan **curvas de potencia**, las cuales permiten fácilmente localizar la potencia de un contraste según sea el valor que puede tomar H_1 y el tamaño de la muestra. En la Figura 2.13 se representan diversas curvas de potencia para los datos del ejemplo, de acuerdo a diferentes tamaños muestrales y a diferentes valores de H_1. Se puede ver en la Figura 2.13 que para $H_1 = 1060$ y un tamaño muestral de 100, la potencia, efectivamente, está por encima de 0,85 en el gráfico (el valor exacto es 0,8621).

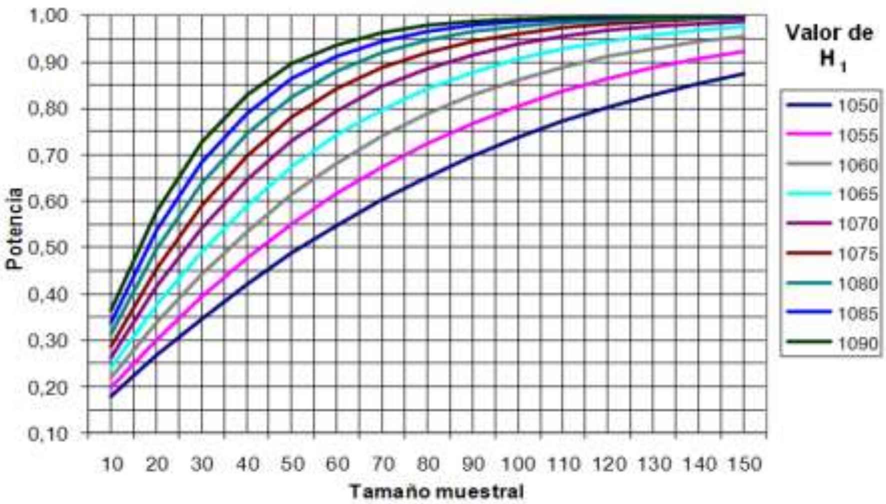


Figura 2.13: *Potencia en función de H_1 y tamaño muestral, con H_0 = 1000 y Error Típico =* $220/\sqrt{n}$

2.7.- Nivel crítico p y errores en los contrastes

En las pruebas clásicas de contrastes que hemos explicado, es preciso establecer el error tipo I (nivel de significación α) antes de realizar el contraste, de modo que este valor no influya en la decisión final que se toma. Este error es, pues, el máximo riesgo que estamos dispuestos a admitir al tomar una decisión respecto a la hipótesis nula. No obstante, establecer previamente un nivel de error tipo I, presenta algún inconveniente que puede ser decisivo en la decisión que se tome.

Como hemos visto en los ejemplos 2.2 y 2.4 de este tema, la decisión que se tome sobre H_0 puede depender del nivel de significación que se establezca, y se puede dar la circunstancia de que sea rechazada con un nivel del 5% y no serlo con el 1%. Si bien es cierto que hay un acuerdo en el ámbito científico acerca de que "alfa" debe ser un valor pequeño (aunque el valor concreto depende mucho del área de investigación, siendo usual en Psicología el 0,05, en otros ámbitos los editores de las revistas científicas llegan a pedir valores de α tan pequeños como 0,01 o inferiores), es más difícil determinar cuán pequeño debe ser, ya que en parte dependerá de factores, alguno de los cuales, como señalan Wonnacott y Wonnacott (1999), pueden ser simplemente las creencias previas sobre los procesos de toma de decisión que se han realizado anteriormente sobre la misma o parecida cuestión, y también sobre las consecuencias que se deriven al tomar una decisión errónea, y ésta se puede tomar tanto rechazando una hipótesis nula que es verdadera (error tipo I) como aceptando una hipótesis nula que es falsa (error tipo II). Además, es preciso tener en cuenta que una disminución del primero (α) provoca un aumento automático del segundo (β).

Debido, pues, a estos inconvenientes, en el análisis de datos moderno hace ya un tiempo que se ha introducido el denominado nivel crítico p, que se define como el nivel de significación más pequeño al que una hipótesis nula puede ser rechazada con la medida de discrepancia obtenida. Es decir, el nivel crítico p es la probabilidad asociada a la medida de discrepancia que hemos obtenido a partir de la información obtenida en nuestra muestra y cuantifica la probabilidad de obtener unos datos como los obtenidos en la investigación o más extremos bajo el supuesto de que la hipótesis nula es verdadera.

En los contrastes bilaterales de parámetros (o "two tail" en inglés que, literalmente, significa "dos colas" haciendo referencia a los dos extremos de la distribución de probabilidad correspondiente) de una distribución muestral simétrica (*v.g.* la distribución normal, la t de Student o la binomial cuando p=0,5), el valor del nivel crítico p se obtiene multiplicando por dos la probabilidad asociada a los valores mayores o menores (según en qué parte de la cola caiga el valor del estadístico de contraste como medida de discrepancia).

Al utilizar como criterio para la decisión el nivel crítico p no hay que establecer previamente un nivel de significación, y ésta se toma en función del valor de p. Si p es pequeño se rechazará H_0, y si es grande se aceptará H_0. Obviamente, como señalan Pardo y San Martin (1994), persiste el problema de determinar qué es grande y qué pequeño. Entonces para tomar una decisión hay que recurrir al criterio del grado de cercanía o alejamiento de p a, por ejemplo, el valor 0,05. Si es claramente inferior, se rechaza H_0, si es claramente superior se acepta H_0, y si está en torno a ese valor, se vuelve a tomar nueva evidencia muestral y se repite el contraste.

No obstante, el empleo del nivel crítico p como criterio de decisión tampoco está exento de problemas, ya que, al igual que las medidas de discrepancia observada entre H_0 y la evidencia muestral, depende del tamaño de la muestra utilizada, y es por ello, que, desde la década de los ochenta del siglo pasado se han explorado nuevas medidas, independientes del tamaño muestral, que explicamos en otros temas.

2.8.- Resumen

Como se ha explicado en los diseños de una muestra, todo contraste de hipótesis tiene unos pasos que se pueden fijar con más o menos detalle. De acuerdo con los que se han establecido en este texto, para determinar el procedimiento de análisis de datos más adecuado que se debe utilizar para contrastar una hipótesis de un diseño de investigación, los pasos a seguir serían:

Condiciones y supuestos: Los procedimientos para el contraste de hipótesis que veremos a lo largo del programa de este curso requieren el cumplimiento de unos supuestos a la hora de seleccionar el estadístico de contraste más adecuado al diseño de la investigación, y se refieren al número de muestras utilizadas y su tamaño, el nivel de medida de la o las variables incluidas en la hipótesis, la forma de su distribución en la población, la varianza poblacional conocida o desconocida, etc.

Formulación de hipótesis: Las hipótesis de investigación se traducen en hipótesis estadísticas. Por lo general, la hipótesis del investigador trata de encontrar resultados significativos, es decir, diferencias significativas entre la teoría y los datos, y por esta razón se corresponde con la hipótesis alternativa. Por el contrario, la hipótesis nula afirma que tales diferencias no existen y es la hipótesis que se supone provisionalmente verdadera y que se contrasta con la evidencia que proporcionan los datos de la investigación. Si las hipótesis se refieren a parámetros poblacionales podemos plantear una hipótesis direccional o bidireccional (en donde θ representa un parámetro poblacional genérico):

Contraste bilateral	Unilateral derecho	Unilateral izquierdo
$\boldsymbol{H_0: \theta = \theta_0}$ $\boldsymbol{H_1: \theta \neq \theta_0}$	$\boldsymbol{H_0: \theta \leq \theta_0}$ $\boldsymbol{H_1: \theta > \theta_0}$	$\boldsymbol{H_0: \theta \geq \theta_0}$ $\boldsymbol{H_1: \theta < \theta_0}$

Regla de decisión: Seleccionamos el estadístico de contraste que representa la discrepancia entre el estadístico obtenido a partir de los datos observados en la muestra y el valor planteado en la hipótesis como parámetro poblacional. Este estadístico de contraste tiene una determinada distribución de probabilidad (su distribución muestral) que nos permite fijar los valores críticos que determinan la zona de rechazo de la hipótesis nula. Se ha explicado que estos valores críticos representan la máxima diferencia que puede observarse entre los datos observados en la muestra y los datos teóricos planteados en la hipótesis nula, bajo el supuesto de que ésta es cierta. Esta diferencia o discrepancia entre los datos teóricos y los datos empíricos se puede cuantificar igualmente, en términos de probabilidad: el nivel crítico p.

Calculamos el estadístico de contraste y el nivel crítico p asociado a este valor, que indica la probabilidad de que, siendo cierta la hipótesis nula, obtengamos unos datos iguales o más extremos a los observados en la muestra.

Concluimos respecto al rechazo o no de la hipótesis nula, bien comparando el estadístico de contraste con el valor crítico o comparando el nivel crítico p con el nivel de significación. Si el nivel crítico p es menor que el nivel de significación establecido *a priori*, rechazamos la hipótesis nula. En esta situación también observaremos que el estadístico de contraste supera la máxima diferencia que cabe esperar por simple azar. En caso contrario diremos que no hay evidencias suficiente para rechazar la hipótesis nula por lo que la conservamos o mantenemos con un determinado nivel de confianza.

Interpretamos esta conclusión con referencia a los objetivos e hipótesis de la investigación.

2.9.- Ejercicios de autoevaluación.

1. La distribución muestral de la media es una distribución t de Student, cuando: a) se desconoce la varianza poblacional y el tamaño de la muestra es $n \geq 30$; b) se conoce la varianza poblacional pero se utilizan muestras pequeñas; c) la variable de estudio no se distribuye normalmente en la población, se conoce su varianza y se utilizan muestras grandes.
2. La distribución muestral de la media es una distribución normal, cuando: a) La distribución en la población es normal y conocemos su varianza; b) se desconoce la varianza poblacional pero se utilizan muestras pequeñas; c) la variable de estudio no se distribuye normalmente en la población y se utilizan muestras pequeñas.
3. El nivel crítico p, representa: a) la probabilidad de rechazar una hipótesis nula que es verdadera; b) la probabilidad de error al tomar una decisión sobre la hipótesis nula; c) la probabilidad de obtener unos resultados como los obtenidos en nuestra investigación o más extremos, suponiendo cierta la hipótesis nula.
4. ¿Cuál de las siguientes alternativas es INCORRECTA: a) el valor crítico puede ser negativo; b) el estadístico de contraste puede ser negativo; c) el nivel de significación puede ser negativo.

Para dejar constancia real de las preferencias de los padres sobre la lengua vehicular en la que prefieren que se eduque a sus hijos, una determinada asociación de padres realiza una encuesta sobre una muestra de 800 familias residentes en una determinada autonomía bilingüe, encontrando que 280 familias son partidarios de que todas de las asignaturas se enseñen en castellano y 168 manifiestan su deseo de que la mayoría de las asignaturas se impartan en castellano. Se fija un nivel de significación "alfa" del 0,05 (5%) y la asociación de padres quiere dejar evidencia de que más de la mitad de los padres quiere escolarizar a sus hijos en colegios en los que la presencia del castellano en la enseñanza sea, al menos, mayoritaria:

5. La hipótesis nula es: a) $H_0: \pi \leq 0{,}5$; b) $H_0: \pi \geq 0{,}5$; c) $H_0: \pi = 0{,}5$.
6. El valor del estadístico de contraste, es: a) 2,28; b) 1,96; c) 3,39
7. La máxima diferencia atribuible al azar entre los datos observados en la muestra y los datos teóricos planteados en la hipótesis nula es: a) 1,96; b) $1{,}64$; c) $\pm 1{,}96$
8. Suponiendo cierta la hipótesis nula, la probabilidad de encontrar unos resultados como los observados en la muestra es: a) 0,9997; b) 0,0003 ;c) 0,0006

9. La conclusión de este estudio, es: a) Rechazar la hipótesis nula porque p<α; b) No se puede rechazar la hipótesis nula porque p>α; c) No rechazar la hipótesis nula porque el estadístico de contraste no supera la máxima diferencia que cabe esperar por simple azar.
10. De una población en la que la variable de estudio tiene una distribución normal con varianza 225, se extrae una muestra aleatoria de 25 observaciones. Si fijamos el nivel de significación en 0,10, ¿cuánto valdrá la potencia del contraste de $H_1: \mu = 23$ frente a $H_0: \mu = 20$, para un contraste unilateral derecho: a) 0,6163; b) 0,3897; c) 0,2652.
11. De una población en la que la variable de estudio tiene una distribución normal con varianza 225, se extrae una muestra aleatoria de 25 observaciones. Si fijamos el nivel de significación en 0,10, ¿cuánto valdrá la potencia del contraste de $H_1: \mu = 23$ frente a $H_0: \mu = 20$, para un contraste bilateral: a) 0,2652; b) 0,7348; c) 0,3887.

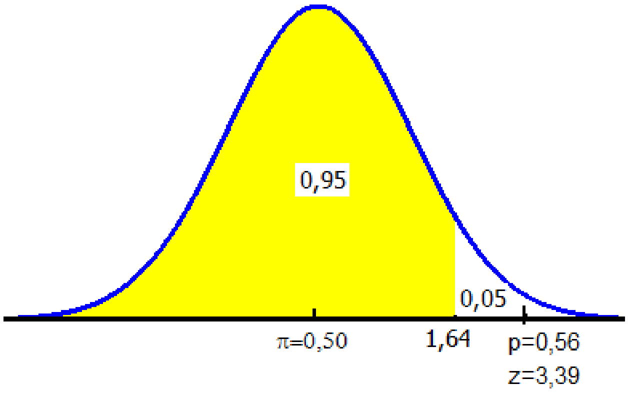
Soluciones

1. Si se desconoce la varianza poblacional, la distribución muestral de la media es la distribución t de Student si $n \geq 30$. La respuesta correcta es la a)
2. Cuando la población se distribuye normalmente y se conoce su varianza, la distribución muestral de la media es normal. La respuesta correcta es la a)
3. El nivel crítico p indica la probabilidad de obtener unos determinados resultados supuesta verdadera la hipótesis nula. Si esta probabilidad es muy pequeña se rechaza la hipótesis nula. La respuesta correcta es la c)
4. Tanto el nivel crítico p como el nivel de significación son probabilidades que nunca pueden ser negativos. Sus valores, expresados en tanto por uno, están comprendidos entre 0 y 1. La respuesta correcta es la c)
5. La hipótesis del investigador es la hipótesis alternativa que pretende demostrar que "más de la mitad de los padres desean escolarizar a sus hijos en colegios en los que la presencia del castellano es mayoritaria. Por tanto la hipótesis nula, negación de la anterior corresponde a la alternativa a).
6. La respuesta correcta es la c)

$$P = \frac{168 + 280}{800} = 0{,}56$$

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} = \frac{0{,}56 - 0{,}50}{0{,}01768} = 3{,}39$$

7.- La respuesta correcta es la b). La distribución muestral de la proporción es una distribución binomial que tiende a la normal. El valor de z que deja por debajo una probabilidad de 0,95 es z=1,64 que corresponde al valor crítico de este ejemplo.

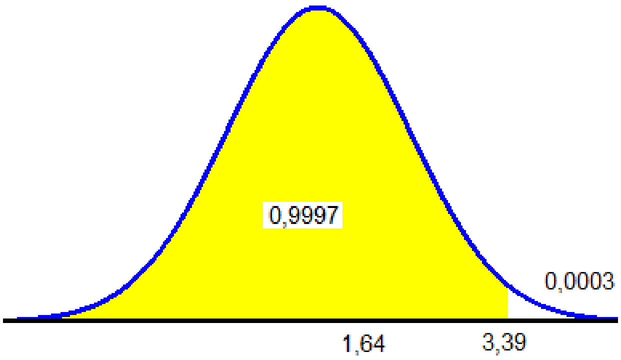


8.- La respuesta correcta es la b). Para z=3,39 encontramos que en la distribución normal:

$$P(z \leq 3{,}39) = 0{,}9997$$

Por tanto:

$$P(z > 3{,}39) = 1 - P(z \leq 3{,}39) = 1 - 0{,}9997 = 0{,}0003$$



9. La solución correcta es la a). Ya que suponiendo cierta la hipótesis nula, la probabilidad de encontrar en una muestra de 800 personas a 448 a favor de esta opción es de un 3 por 10.000. Esta probabilidad es tan pequeña que nos lleva a rechazar la hipótesis nula.
De otra forma, la discrepancia entre la proporción obtenida en la muestra y el valor teórico planteado en la hipótesis nula (3,39) es mayor que la máxima discrepancia que se puede admitir por simple azar (1,64) lo que nos lleva rechazar la hipótesis nula.

10.- Para calcular la potencia del contraste: $H_0: \mu = 20$ frente a $H_1: \mu = 23$ el primer paso es buscar en la distribución muestral de la media formulada en la Ho el valor de Z que deja por debajo una probabilidad de 0,90, (nivel de confianza en un contraste unilateral) y es Z= 1,28. A esta puntuación le corresponde, en la distribución de la Ho, una media muestral $\bar{Y}$ de 23,84.
Segundo paso: En la distribución de H_1, con media de 23, a la puntuación $\bar{Y} = 23{,}84$ le corresponde una puntuación típica de 0,28. Buscamos en la tabla de la distribución normal las probabilidades correspondientes a esta puntuación típica que vale 0,6103. De forma gráfica, el razonamiento es el siguiente:

Primer paso:

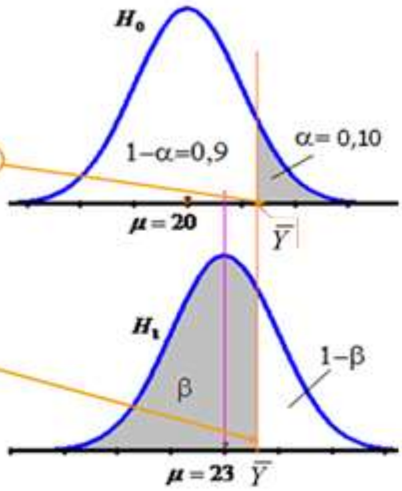
$$\left.\begin{array}{l}\alpha = 0{,}1 \Rightarrow Z = 1{,}28 \\ \sigma_{\bar{Y}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{15}{5} = 3\end{array}\right\} \begin{array}{l} Z = \dfrac{\bar{Y} - \mu}{\sigma_{\bar{Y}}} \Rightarrow \\ 1{,}28 = \dfrac{\bar{Y} - 20}{3} \Rightarrow \bar{Y} = 23{,}84 \end{array}$$

Segundo paso:

$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}} = \frac{23{,}84 - 23}{3} = 0{,}28$$

$$\beta = P(Z \le 0{,}28) = 0{,}6103$$

$$1 - \beta = P(Z \ge 0{,}28) = 1 - 0{,}6103 = 0{,}3897$$



11.- De una población con distribución normal y varianza de 225 se extrae una muestra de n=25 sujetos y con un alfa=0,10 queremos calcular la potencia de un contraste bilateral, siendo las hipótesis nula y alternativa, las siguientes: $H_0: \mu = 20$ frente a $H_1: \mu = 23$

Como en la situación anterior, pero ahora con un contraste bilateral, el primer paso es buscar en la distribución muestral de la media formulada en la Ho los valores de Z que deja entre ambos una probabilidad de 0,90, y es: $z = \pm 1{,}64$. A cada una de estas puntuaciones le corresponde, en la distribución de la H_0, una media muestral $\bar{Y}$ de 15,08 y 24,92 respectivamente.
Segundo paso: En la distribución de H_1, con media de 23, calculamos la puntuación típica que corresponde a cada una de estas medias y son: Z=-2,64 y Z= 0,64. El error tipo II (beta) es la probabilidad entre ambas puntuaciones típicas Z, y la potencia del contraste su complementario. De forma gráfica, el razonamiento es el siguiente:

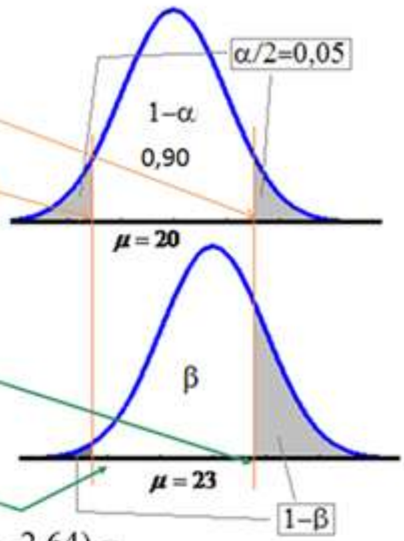
Primer paso:

$$\alpha = 0{,}1 \Rightarrow Z = \pm 1{,}64 \qquad \sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = 3$$

$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}} \Rightarrow \left\{ \begin{array}{l} 1{,}64 = \dfrac{\bar{Y}_s - 20}{3} \Rightarrow \bar{Y}_s = 24{,}92 \\ -1{,}64 = \dfrac{\bar{Y}_i - 20}{3} \Rightarrow \bar{Y}_i = 15{,}08 \end{array} \right.$$

Segundo paso:

$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}} = \left\langle \begin{array}{l} \dfrac{24{,}92 - 23}{3} = 0{,}64 \\ \dfrac{15{,}08 - 23}{3} = -2{,}64 \end{array} \right.$$



$$\beta = P(-2{,}64 \le Z \le 0{,}64) = P(Z \le 0{,}64) - P(Z \le -2{,}64) =$$
$$= 0{,}7389 - 0{,}0041 = 0{,}7348 \qquad 1 - \beta = 1 - 0{,}7348 = 0{,}2652$$

Tema 3. ANÁLISIS DE DATOS PARA DISEÑOS DE DOS GRUPOS. MUESTRAS INDEPENDIENTES.

3.1.- Introducción.

En temas anteriores hemos visto los contrastes de hipótesis más habituales para diseños de una sola muestra cuya utilidad es indiscutible, pero generalmente será conveniente, y en algunos casos necesario, utilizar más de una muestra.

Veamos un ejemplo típico en el que utilizamos un diseño de una muestra. Un profesor opina que aplicando un nuevo método de enseñanza, podría lograr que sus estudiantes comprendieran mejor su asignatura, lo que se traduciría en un incremento de la nota media a final del curso. Sabe que la nota media y la varianza de las calificaciones de años anteriores son: $\mu = 5,\ \sigma^2 = 4$, y considera que estos son los datos de la población. Aplica el nuevo método de enseñanza a una muestra de 36 sujetos obteniendo una nota media: $\bar{Y} = 7$. Se pregunta si los datos de la muestra son compatibles con los datos de cursos pasados. Realiza un contraste de hipótesis y como el estadístico de contraste es significativo ($Z = 6,\ p < 0{,}0001$) concluye que el nuevo método de enseñanza es más eficaz que el tradicional.

Ahora bien, en los resultados que ha encontrado el profesor pueden influir otros factores. Por ejemplo, si ha sido él quien aplicó el nuevo método de enseñanza, es posible que haya trabajado con más entusiasmo que en cursos anteriores, o bien los alumnos del último curso accedieron a su asignatura mejor preparados, o quizás los estudiantes han tenido menos trabajo en otras asignaturas y han podido dedicar más tiempo a la suya. Si hubiese formado dos grupos las conclusiones serían más claras al estar menos sujetas a explicaciones rivales. Para uno de ellos el método de enseñanza sería el tradicional (grupo de control), mientras que el otro grupo (experimental) estaría formado por los sujetos que aprendieron la asignatura con el nuevo método. Finalmente, nuestro profesor realizaría un contraste de hipótesis para comprobar si existen diferencias entre el grupo experimental y el grupo de control.

También es muy común en psicología el diseño de dos muestras cuando queremos comprobar la eficacia de un tratamiento. En este caso medimos la variable dependiente (por ejemplo, ansiedad) antes y después del tratamiento y comparamos ambas medias para comprobar si la terapia ha sido eficaz.

En otras ocasiones el mismo problema que queremos investigar nos obliga a utilizar dos muestras, porque queremos estudiar diferencias entre dos poblaciones diferentes, como puede ser entre hombres y mujeres, entre ambiente rural y urbano, entre dos clases sociales, etc.

Al trabajar con dos muestras en los Temas 3 y 4, utilizaremos subíndices para distinguir el tamaño muestral, la media y varianza de cada una de ellas. También, para facilitar la legibilidad de las fórmulas y distinguir entre varianza $(S_1^2\ y\ S_2^2)$ y cuasivarianza $\left(\hat{S}_1^2\ y\ \hat{S}_2^2\right)$, representaremos a la segunda mediante un acento circunflejo.

Dividiremos el estudio del análisis de datos en diseños de dos muestras en dos temas. En el Tema 3 trabajaremos con muestras independientes y en Tema 4 con muestras relacionadas. Comenzaremos el presente tema aprendiendo a distinguir entre los dos tipos de muestras (punto 3.3), para ver a continuación como llevar a cabo contrastes de hipótesis paramétricos para dos medias en muestras independientes (punto 3.4).

A continuación trataremos el contraste de hipótesis para dos varianzas (punto 3.5) y dos proporciones (punto 3.6) en muestras independientes. Finalizaremos el tema con un apartado sobre la magnitud del efecto (punto 3.7), que cada vez está cobrando más importancia en los informes de investigación, para finalizar con un resumen del tema (punto 3.8) y con ejercicios de autocomprobación (punto 3.9).

Los dos temas que veremos a continuación, no deberían plantearnos muchos problemas si hemos asimilado correctamente los anteriores, puesto que matemáticamente los conceptos son los mismos y seguiremos el mismo esquema que en el Tema 2, fijándonos prácticamente los mismos objetivos aplicándolos al estudio de dos muestras, es decir:

3.2.- Objetivos del tema.

- ✓ Distinguir entre muestras independientes y relacionadas.
- ✓ Plantear las hipótesis en función de los objetivos de la investigación.
- ✓ Distinguir entre contraste unilateral y bilateral.
- ✓ Seleccionar el estadístico de contraste más adecuado a las hipótesis planteadas.
- ✓ Conocer la distribución muestral del estadístico seleccionado.
- ✓ Realizar los cálculos oportunos para someter a contrastación empírica las hipótesis planteadas.
- ✓ Relacionar el intervalo de confianza con el estadístico de contraste.
- ✓ Interpretar el nivel crítico p.
- ✓ Determinar e interpretar el o los valores críticos de la distribución muestral.
- ✓ Tomar una decisión respecto a las hipótesis planteadas.
- ✓ Conocer, comprender e interpretar la magnitud del efecto.

3. 3.- Muestras independientes o relacionadas.

En todas las técnicas estadísticas que vemos en este curso **suponemos que las observaciones dentro de una muestra son independientes**, es decir, que no existe relación entre ellas. Por lo tanto, dentro de un grupo el valor de una determinada puntuación no nos informa en absoluto del valor de otras puntuaciones dentro del mismo grupo. Veamos un par de ejemplos. Un psicobiólogo dispone de 10 ratas para realizar un experimento. Quiere formar dos grupos, que tras ser sometidos a diferentes niveles de estrés correrán el mismo laberinto. Todas las ratas están en la misma jaula. Las 5 primeras que coge forman el primer grupo y las 5 restantes el segundo. Probablemente la primera rata que atrapa el investigador es la más "despistada", la siguiente es un poco menos despistada y así sucesivamente, de forma que el primer grupo está formado por las ratas más torpes, e independientemente del tratamiento la media del primer grupo es superior (tardan más en correr el laberinto). En este caso las puntuaciones dentro de cada grupo están relacionadas y sabiendo en qué orden fue capturada una rata puedo predecir su nivel de "torpeza". Veamos otro ejemplo. Me voy 20 días a Río de Janeiro de vacaciones en la mejor época posible, y el tiempo empeora progresivamente. Calculo la temperatura media de esos días y concluyo que la temperatura en Río es muy fría. Los resultados son significativos. Pero si he tenido la mala suerte de que mi viaje coincida con que la peor borrasca del siglo haya comenzado justo al aterrizar en Río, quizás observe que la temperatura ha descendido día tras día, de forma que conociendo la temperatura de un día cualquiera de mis vacaciones, puedo predecir que la del día siguiente será inferior. La conclusión a la que he llegado es errónea porque los datos que he tomado no son independientes. Si bien existen contrastes de hipótesis para comprobar la independencia de las observaciones (que no se verá en este curso), para garantizar la independencia de los datos dentro de un grupo, lo mejor que puede hacerse es seleccionar los elementos de la muestra de forma aleatoria.

Por otro lado, cuando trabajamos **con dos muestras (o más de dos), las muestras pueden ser independientes o relacionadas**. Son independientes cuando no existe relación entre los sujetos de una y otra, lo que podremos garantizar si los sujetos son asignados aleatoriamente a cada una de las muestras.

Tenemos **muestras relacionadas** cuando cada observación en una muestra tiene su pareja en la otra. El caso más evidente es cuando son los **mismos sujetos** los que pasan por diferentes condiciones experimentales. Como comentábamos anteriormente, si queremos probar la eficacia de una terapia contra la ansiedad, podemos seleccionar a un grupo de sujetos en los que medimos su nivel de ansiedad antes del experimento; aplicamos la terapia, y volvemos a medirlo después de la terapia para comparar las medias en ansiedad antes y después. En otras ocasiones no son los mismos sujetos los que se repiten en las muestras, pero hay una **relación sujeto a sujeto** en ambas. Por ejemplo, si disponemos de 10 parejas de hermanos gemelos, podemos formar dos grupos de 5 personas donde cada dos hermanos son asignados, aleatoriamente, a grupos distintos. También podemos contar con padres e hijos, maridos y mujeres, etc. Por último, también podemos utilizar pares de **sujetos que están equiparados en variables que pueden influir en el diseño de la investigación**. Por ejemplo, supongamos que para probar la eficacia de dos métodos de enseñanza, queremos controlar la influencia del cociente intelectual, por lo que tomamos pares de sujetos con un CI semejante formando cada uno de ellos parte de muestras diferentes.

Recordamos que las ventajas y problemas de los diseños de dos grupos son tratados en la asignatura Fundamentos de Investigación (Tema 5).

3.4.- Contrastes de hipótesis sobre dos medias en muestras independientes.

En este apartado veremos tres contrastes de hipótesis sobre dos medias para muestras independientes en función de los supuestos que hagamos sobre las varianzas poblacionales. Comenzaremos por suponerlas conocidas, pero si no es así, que es lo más habitual, podemos suponer que son iguales o diferentes. Utilizaremos un ejemplo para ilustrar el procedimiento estadístico en cada uno de los casos, y para no repetir innecesariamente las mismas fórmulas, cambiaremos a lo largo de los ejemplos la hipótesis alternativa que, como sabemos, puede ser bilateral, unilateral derecha o unilateral izquierda.

En todo contraste de hipótesis el proceso de inferencia estadística se realiza sobre una distribución teórica que denominamos distribución muestral. Comenzaremos con un ejemplo, con fines obviamente didácticos, cuyo objetivo es, simplemente, que el alumno comprenda cómo se compone la distribución muestral en el caso de dos medias con muestras independientes.

3.4.1.- Distribución muestral de la diferencia de medias para dos muestras independientes (OPCIONAL).

Supongamos que tenemos dos poblaciones, y que cada una de ellas, para que el ejemplo sea lo más corto posible, contiene 3 observaciones. Denotaremos las puntuaciones mediante la letra latina Y. Presentamos las puntuaciones, media y varianza de dichas poblaciones:

Población 1. $\{Y_{11} = 2; Y_{12} = 5; Y_{13} = 8\}$. $\mu_1 = 5$, $\sigma_1^2 = 6$.
Población 2. $\{Y_{21} = 3{,}5; Y_{22} = 5; Y_{23} = 6{,}5\}$. $\mu_2 = 5$, $\sigma_2^2 = 1{,}5$.

Donde el primer subíndice hace referencia a la población a la que pertenecen y el segundo al orden que cada puntuación ocupa en su población. En la Tabla 3.1 se muestra el cálculo de la media aritmética de todas las sub-muestras de tamaño “n = 2” con reposición para la Población 1, y que formarán la distribución muestral de la media para dicha población en muestras de tamaño “n = 2”.

Tabla 3.1

Media aritmética de todas las muestras posibles de tamaño n = 2 para la Población 1.

	$Y_{11} = 2$	$Y_{12} = 5$	$Y_{13} = 8$
$Y_{11} = 2$	$\frac{2+2}{2} = 2$	$\frac{2+5}{2} = 3{,}5$	$\frac{2+8}{2} = 5$
$Y_{12} = 5$	$\frac{5+2}{2} = 3{,}5$	$\frac{5+5}{2} = 5$	$\frac{5+8}{2} = 6{,}5$
$Y_{13} = 8$	$\frac{8+2}{2} = 5$	$\frac{8+5}{2} = 6{,}5$	$\frac{8+8}{2} = 8$

La distribución muestral de la media para la Población 1, está compuesta por los 9 valores de la Tabla 3.1, que ordenados son: {2; 3,5; 3,5; 5; 5; 5; 6,5; 6,5; 8}, con media y varianza:

$$\mu_1 = \frac{\sum Y}{n} = \frac{45}{9} = 5 \qquad \sigma^2_{\bar{y}_1} = \frac{\sum Y^2}{n} - \bar{Y}^2 = \frac{252}{9} - 5^2 = 3$$

Como vimos en el primer tema, estos valores también podríamos calcularlos en función de la media y varianza de la población.

$$\mu_1 = 5; \;\; \sigma^2_{\bar{y}_1} = \frac{\sigma^2}{n} = \frac{6}{2} = 3$$

Repitiendo el proceso seguido para la Población 1, obtenemos la distribución muestral para la Población 2, formada por los valores: {3,5; 4,25; 4,25; 5; 5; 5; 5,75; 5,75; 6,5}, con media y varianza, respectivamente: $\mu_2 = 5 \;\; y \;\; \sigma^2_{\bar{y}_2} = 0{,}75$.

La distribución muestral de las diferencias $\bar{Y}_1 - \bar{Y}_2$, la compondremos al emparejar todas las muestras de la Población 1 con todas las muestras de la Población 2. En la Tabla 3.2 están reflejadas todas las opciones posibles.

Tabla 3.2

Distribución muestral de las diferencias $\bar{Y}_1 - \bar{Y}_2$. Todas las posibles muestras de la Población 1 pueden emparejarse, aleatoriamente, con cualquier muestra de la Población 2.

		Distribución muestral de la Población 1								
		2	**3,5**	**3,5**	**5**	**5**	**5**	**6,5**	**6,5**	**8**
Distribución muestral. Población 2	**3,5**	-1,5	0	0	1,5	1,5	1,5	3	3	4,5
	4,25	-2,25	-0,75	-0,75	0,75	0,75	0,75	2,25	2,25	3,75
	4,25	-2,25	-0,75	-0,75	0,75	0,75	0,75	2,25	2,25	3,75
	5	-3	-1,5	-1,5	0	0	0	1,5	1,5	3
	5	-3	-1,5	-1,5	0	0	0	1,5	1,5	3
	5	-3	-1,5	-1,5	0	0	0	1,5	1,5	3
	5,75	-3,75	-2,25	-2,25	-0,75	-0,75	-0,75	0,75	0,75	2,25
	5,75	-3,75	-2,25	-2,25	-0,75	-0,75	-0,75	0,75	0,75	2,25
	6,5	-4,5	-3	-3	-1,5	-1,5	-1,5	0	0	1,5

Con los valores de la Tabla 3.2, podemos comprobar que la media y varianza de los datos que contiene son igual a:

$$\mu_1 - \mu_2 = 0, \quad \sigma^2_{\bar{Y}_1 - \bar{Y}_2} = 3{,}75$$

Otra forma de obtener la varianza de la distribución muestral es:

$$\sigma^2_{\bar{Y}_1 - \bar{Y}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{6}{2} + \frac{1{,}5}{2} = 3{,}75$$

En los contrastes que veremos a continuación vamos a suponer que las poblaciones de las que proceden las muestras que utilizaremos se distribuyen normalmente, o bien que $n_1 \geq 30$ y $n_2 \geq 30$. Esto nos garantiza que las distribuciones muestrales de la media en ambos casos también se distribuyen normalmente, y si esto es así, también se distribuirá normalmente la distribución muestral de las diferencias entre medias (observe el lector que la diferencia $\mu_1 - \mu_2$ en el ejemplo anterior es igual a cero, que es lo que generalmente postulará la hipótesis nula, por lo que habitualmente nos limitaremos a este caso).

3.4.2.- Varianzas poblacionales conocidas.

Ejemplo 3.1. Un psicólogo escolar utiliza un test de comprensión verbal recientemente traducido del inglés, que proporciona puntuaciones en un nivel de medida de intervalo. Se sabe, por investigaciones anteriores, que las varianzas en la población son para niños y niñas $\sigma_1^2 = 36 \ y \ \sigma_2^2 = 49$ respectivamente. Las investigaciones anteriores también indican que la media es la misma en ambos grupos, pero este último aspecto no ha sido comprobado con muestras españolas. El psicólogo considera que la traducción del test no es muy acertada y puede provocar diferencias que en realidad no se deben a la comprensión verbal, por lo que selecciona aleatoriamente una muestra de 100 niños y otra muestra de 200 niñas obteniendo una media igual a 20 para los niños e igual a 17,5 para las niñas. Con un nivel de confianza del 95%. ¿Podemos afirmar que la puntuación media en el test de comprensión verbal es la misma para niños y niñas?

Condiciones y supuestos. Tenemos un diseño de dos muestras independientes (niños y niñas), seleccionadas de dos poblaciones con varianzas conocidas (el psicólogo asume que las varianzas de las poblaciones de niños y niñas son las que reflejan las investigaciones anteriores), donde la variable dependiente (comprensión verbal) proporciona puntuaciones en una escala de intervalo. Aunque no sabemos si las poblaciones se distribuyen normalmente, trabajamos con muestras que son lo suficientemente grandes ($n_1 \geq 30$ y $n_2 \geq 30$). En definitiva se cumplen los siguientes supuestos:

- Variable dependiente con un nivel de medida de intervalo o razón.
- Dos poblaciones que se distribuyen normalmente, o bien, $n_1 \geq 30$ y $n_2 \geq 30$.
- Varianzas poblacionales conocidas.

Formular las hipótesis. En este caso el psicólogo piensa que pueden existir diferencias pero no tiene una hipótesis previa sobre la dirección de las mismas, por lo que planteamos un contraste de hipótesis bilateral:

$$\begin{array}{l} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{array} \quad , \quad \text{o bien:} \quad \begin{array}{l} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{array}$$

Estadístico de contraste y su distribución muestral. Conocemos las varianzas de las dos poblaciones y trabajamos con muestras grandes, lo que nos permite asumir la normalidad de la distribución muestral de las diferencias entre medias. Siendo el grupo 1 el de niños y el 2 el de niñas, el estadístico de contraste es igual a:

$$Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(20 - 17{,}5) - 0}{\sqrt{\frac{36}{100} + \frac{49}{200}}} = 3{,}21$$

Observamos que la fórmula del estadístico de contraste sigue el mismo esquema general visto en el Tema 1, cuantificando la discrepancia entre la diferencia de medias observada entre las dos muestras frente a una diferencia nula planteada en la hipótesis nula medida en unidades de desviación típica. Por tanto, en el numerador tenemos la diferencia entre el valor del estadístico en la muestra $(\bar{Y}_1 - \bar{Y}_2)$ respecto del valor del parámetro que postula la hipótesis nula $(\mu_1 - \mu_2)$.

Habitualmente la hipótesis nula, como en este caso, especificará que no existe diferencia entre las medias poblacionales, por lo que el término $\mu_1 - \mu_2$, es igual a cero. Por este motivo, generalmente calcularemos el estadístico de contraste mediante la siguiente ecuación:

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad \text{(Ecuación 3.1)}$$

Podemos calcular el nivel crítico p en la tabla de curva normal, que como sabemos es la probabilidad de obtener un valor como el observado o más extremo, suponiendo que la hipótesis nula es cierta.

Primero buscamos la probabilidad de encontrar valores superiores a 3,21:

$$z = 3{,}21 \xrightarrow{Tabla\ de\ curva\ normal} p = 0{,}9993 \rightarrow 1 - 0{,}9993 = 0{,}0007$$

Y como el contraste es bilateral multiplicamos por dos el valor obtenido:

$$\textit{Nivel crítico } p = 2 \cdot 0{,}0007 = 0{,}0014$$

Establecer la regla de decisión en función del nivel de confianza. El nivel de significación es del 5% y el contraste es bilateral, por lo que los valores críticos que delimitan cuándo mantenemos y cuándo rechazamos la hipótesis nula son las puntuaciones típicas: $z = \pm 1{,}96$. En el Figura 3.1 representamos los datos del problema.

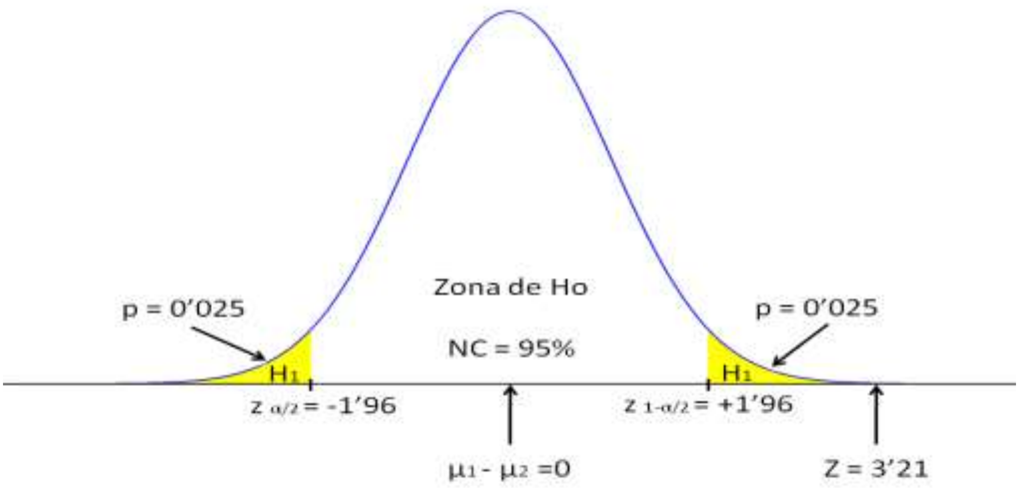


Figura 3.1. *Distribución muestral de las diferencias entre medias según especifica la hipótesis nula. Varianzas poblacionales conocidas.*

Conclusión. Vemos claramente en la Figura 3.1, que el estadístico de contraste (Z = 3,21) no pertenece a la zona de valores compatibles con H_0 que definen las puntuaciones $\pm 1{,}96$ $(3{,}21 > 1{,}96)$, por lo que rechazamos la hipótesis nula. En otras palabras, el estadístico de contraste (la discrepancia observada) supera la diferencia que cabría esperar por simple azar. En general, en un contraste bilateral, mantendremos la hipótesis nula cuando el estadístico de contraste no alcance el valor crítico: $z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}$, y la rechazaremos cuando: $Z < z_{\frac{\alpha}{2}}\ o\ bien\ Z > z_{1-\frac{\alpha}{2}}$

Si utilizamos el nivel crítico p para concluir qué decisión tomar con respecto a H_0, llegamos a la misma conclusión, puesto que 0,0014 < 0,05 (en general, $p < \alpha$). Como se ha expuesto en los temas anteriores, el comparar el nivel crítico con el nivel de significación nos proporciona más información que la comparación del estadístico de contraste con el valor crítico, puesto que vemos claramente que es muy improbable que, siendo la hipótesis nula verdadera, obtengamos dos muestras cuyas medias tengan una diferencia como la observada. El resultado sería significativo incluso a un nivel de confianza superior al 99%.

Interpretar el resultado en función del contexto de la investigación. Las sospechas del psicólogo parecen fundadas. Las diferencias entre niños y niñas en fluidez verbal son significativas, y pueden deberse a la deficiente traducción del test.

Intervalo de confianza. Si estuviéramos interesados en calcular el intervalo de confianza, lo haríamos mediante la expresión:

$$(\bar{Y}_1 - \bar{Y}_2) \pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad \text{(Ecuación 3.2)}$$

Que en nuestro caso queda:

$$(20 - 17{,}5) \pm 1{,}96\sqrt{\frac{36}{100} + \frac{49}{200}} \longrightarrow 2{,}5 \pm 1{,}52 \longrightarrow (0{,}98;\ 4{,}02)$$

Es decir, con un nivel de confianza del 95% la diferencia entre la media de los niños y la media de las niñas en el test de fluidez verbal oscila entre 0,98 y 4,02 puntos a favor de los primeros. Como el contraste de hipótesis planteado es bilateral, también podemos decidir si mantenemos o rechazamos la hipótesis nula examinando los límites del intervalo de confianza. Efectivamente dicho intervalo de confianza no contiene el valor cero, por lo que rechazamos hipótesis nula: $\mu_1 - \mu_2 = 0$.

3.4.3.- Varianzas poblacionales desconocidas pero supuestas iguales.

Ejemplo 3.2. En un estudio sobre depresión en personas mayores llevado a cabo en un centro geriátrico, se quiere comprobar si las personas ingresadas que no reciben visitas de sus familiares tienen una puntuación media en depresión superior a aquellas personas cuyos familiares les visitan con frecuencia. Para comprobar esta hipótesis, se seleccionaron aleatoriamente 41 personas que no reciben visitas obteniéndose una puntuación media de 20 puntos con una cuasivarianza igual a 100, mientras que en una muestra aleatoria de 31 personas que si reciben visitas con frecuencia la media fue igual a 15 con una cuasivarianza igual a 90. Suponiendo que las varianzas en la población son iguales para ambos grupos, y con un nivel de confianza del 99% ¿podemos decir que los datos obtenidos avalan la hipótesis de partida?

Condiciones y supuestos. Los requisitos en este caso son iguales que en el caso anterior, excepto que no conocemos las varianzas poblacionales, si bien las suponemos iguales. Comprobamos pues que se cumplen los siguientes puntos:

- Variable dependiente con un nivel de medida de intervalo o razón. Suponemos que el test de depresión proporciona medidas en una escala de intervalo.
- No sabemos si la distribución en la población es normal, pero salvamos este obstáculo utilizando dos muestras con 30 o más observaciones en cada una de ellas.
- Varianzas poblacionales desconocidas y supuestas iguales. Veremos posteriormente cómo contrastar diferencias entre dos varianzas. En cualquier caso, la diferencia entre las varianzas de las muestras es pequeña.

Formular las hipótesis. Partimos de la idea de que la depresión media es superior en las personas que no reciben visitas de sus familiares (Grupo 1) respecto de las personas que reciben con frecuencia visitas de sus familiares (Grupo 2), por lo que realizamos un contraste unilateral derecho. Las hipótesis en este caso han de ser:

$$\begin{array}{l} H_0: \mu_1 - \mu_2 \leq 0 \\ H_1: \mu_1 - \mu_2 > 0 \end{array} \quad \text{o bien,} \quad \begin{array}{l} H_0: \mu_1 \leq \mu_2 \\ H_1: \mu_1 > \mu_2 \end{array}$$

Estadístico de contraste y su distribución muestral. El estadístico de contraste en este caso se distribuye según t de Student con $n_1 + n_2 - 2$ grados de libertad, y adopta la siguiente expresión:

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Como comentábamos anteriormente, el término $\mu_1 - \mu_2$, habitualmente es igual a cero, por lo que calcularemos el estadístico de contraste, mediante la siguiente ecuación.

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{(Ecuación 3.3)}$$

Con los datos del ejemplo tendremos: $n_1 + n_2 - 2 = 41 + 31 - 2 = 70$ grados de libertad, siendo el estadístico de contraste igual a:

$$T = \frac{20 - 15}{\sqrt{\frac{40 \cdot 100 + 30 \cdot 90}{41 + 31 - 2}\left(\frac{1}{41} + \frac{1}{31}\right)}} = \frac{5}{2{,}33} = 2{,}15$$

Establecer la regla de decisión en función del nivel de confianza. Buscamos en las tablas de t de Student el valor crítico, que en este caso es igual a la puntuación que supera al 99% de la distribución para 70 grados de libertad: $t_{70;0,99} = 2{,}381$ (véase la Figura 3.2)

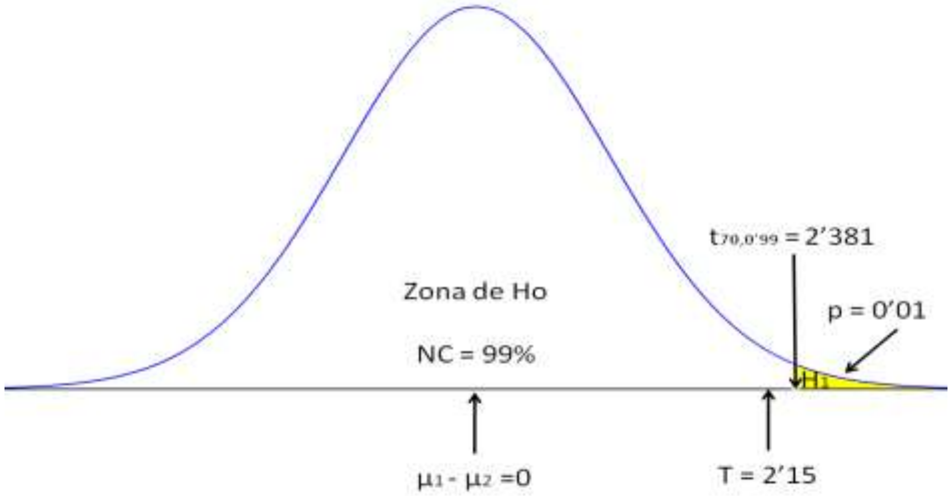


Figura 3.2. *Distribución muestral de las diferencias entre medias según especifica la hipótesis nula. Varianzas poblacionales desconocidas pero supuestas iguales ($\sigma_1^2 = \sigma_2^2$).*

El nivel crítico p es igual a p = 0,0175. No podemos calcularlo exactamente en las tablas del apéndice, pero podemos utilizarlas para hallar un valor aproximado. Observamos en la tabla t de Student, que para 70 grados de libertad nuestro estadístico de contraste se encuentra entre las puntuaciones 1,994 y 2,381 (1,994 < 2,15 < 2,381) que dejan por encima de si respectivamente las proporciones: 0,025 y 0,01, luego el nivel crítico p se encontrará entre estos dos valores (0,01 < p < 0,025).

Conclusión. Como podemos apreciar en el Figura 3.2, el valor del estadístico de contraste no supera al valor crítico ($2{,}15 < 2{,}381$) por lo que la diferencia encontrada no es significativa con un nivel de confianza del 99%. En general, y como en situaciones anteriores, en un contraste unilateral derecho mantendremos la hipótesis nula cuando el estadístico de contraste no supere el valor crítico, es decir, si $T < t_{n_1+n_2-2,1-\alpha}$, y la rechazaremos en caso contrario, es decir, cuando $T > t_{n_1+n_2-2,1-\alpha}$. Si comparamos el nivel crítico p con el nivel de significación, llegamos a la misma conclusión (0,0175 > 0,01).

Interpretar el resultado en función del contexto de la investigación. Al nivel de confianza del 99% los resultados no indican que la puntuación media en depresión es mayor en el grupo de sujetos que no reciben visitas respecto de los que sí las reciben. Pero los resultados sí son significativos al nivel de confianza del 95%, como apreciamos al comparar el nivel de significación con el nivel crítico. Quizás fuera conveniente profundizar en la relación entre ser visitado o no por los familiares y puntuar más alto en depresión en las personas que permanecen ingresadas en centros geriátricos.

Intervalo de confianza. Utilizamos para su cálculo la expresión que puede verse en la Ecuación 3.4:

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{n_1+n_2-2;1-\frac{\alpha}{2}} \sqrt{\frac{(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \qquad \text{(Ecuación 3.4)}$$

Que en nuestro caso queda:

$$(20-15) \pm 2{,}648\sqrt{\frac{40 \cdot 100 + 30 \cdot 90}{41 + 31 - 2}\left(\frac{1}{41} + \frac{1}{31}\right)} \longrightarrow 5 \pm 6{,}16 \longrightarrow (-1{,}16;\ 11{,}16)$$

Como vimos en temas anteriores, en el caso en el que los grados de libertad sean superiores a g.l. = 100, no podemos consultar en las tablas que manejamos los valores de la distribución T de Student. Pero dado que a medida que aumentan los grados de libertad la distribución T se parece cada vez más a la normal tipificada, para g.l. > 100 podemos utilizar la tabla de curva normal, dado que la diferencia entre los valores T y Z es muy pequeña.

3.4.4.- Varianzas poblacionales desconocidas y supuestas distintas.

Ejemplo 3.3. Un laboratorio desarrolla un fármaco con el que se pretende reducir la ansiedad. Para comprobarlo, se extrajeron dos muestras aleatorias de cinco observaciones cada una que suponemos procedentes de poblaciones que se distribuyen normalmente con distinta varianza. A los sujetos de la primera muestra se les administró el fármaco y los de la segunda una sustancia placebo. Posteriormente se les midió la ansiedad a todos los sujetos mediante un test en el que cuanto más elevada es la puntuación mayor es la ansiedad. Los resultados de ambas muestras fueron:

Grupo 1 (con fármaco): 10; 20; 30; 20; 5
Grupo 2 (sin fármaco): 30; 50; 30; 60; 20

Con un nivel de confianza del 95%, ¿podemos afirmar que el fármaco efectivamente reduce la ansiedad?

Condiciones y supuestos. Al igual que en los ejemplos anteriores, necesitamos que la variable dependiente esté medida a nivel de intervalo. En cuanto a las poblaciones de las que proceden las varianzas, necesitamos suponerlas normalmente distribuidas porque el tamaño de las muestras es pequeño (con $n_1 \geq 30$ y $n_2 \geq 30$, no es necesario suponer distribuciones normales en ambas poblaciones). En este caso tampoco conocemos las varianzas poblacionales, aunque ahora las suponemos distintas.

Formular las hipótesis. De acuerdo con la hipótesis del laboratorio esperamos que la puntuación media sea inferior en el Grupo 1, por lo que hemos de plantear un contraste de hipótesis unilateral izquierdo.

$$\begin{array}{l} H_0: \mu_1 - \mu_2 \geq 0 \\ H_1: \mu_1 - \mu_2 < 0 \end{array} \quad \text{o bien,} \quad \begin{array}{l} H_0: \mu_1 \geq \mu_2 \\ H_1: \mu_1 < \mu_2 \end{array}$$

Estadístico de contraste y su distribución muestral. El estadístico de contraste lo calculamos mediante la Ecuación 3.5:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}} \qquad \text{(Ecuación 3.5)}$$

Observe el lector que hemos omitido el término $\mu_1 - \mu_2$, porque en este caso, como es habitual, es igual a cero, por lo que omitiremos este término en contrastes posteriores.

El estadístico de contraste sigue una distribución muestral cuyos grados de libertad calculamos mediante la Ecuación 3.6, cuyo resultado se redondea prescindiendo de los decimales:

$$gl = \frac{\left(\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}\right)^2}{\frac{\left(\hat{S}_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(\hat{S}_2^2/n_2\right)^2}{n_2 - 1}} \qquad \text{(Ecuación 3.6)}$$

Con los datos del ejemplo 3.3 tenemos que calcular las medias y varianzas insesgadas de ambas muestras ya que no nos son proporcionadas directamente en el enunciado del ejercicio.

Primero calculamos las varianzas de ambos grupos:

$$\bar{Y}_1 = \frac{\sum Y_1}{n_1} = \frac{85}{5} = 17, \quad S_1^2 = \frac{\sum Y_1^2}{n_1} - (\bar{Y}_1)^2 = \frac{1825}{5} - 17^2 = 76$$

$$\bar{Y}_2 = \frac{\sum Y_2}{n_2} = \frac{190}{5} = 38, \quad S_2^2 = \frac{\sum Y_2^2}{n_2} - (\bar{Y}_2)^2 = \frac{8300}{5} - 38^2 = 216$$

Las cuasivarianzas o varianzas insesgadas, serán:

$$\hat{S}_1^2 = S_1^2 \cdot \frac{n_1}{n_1 - 1} = 76 \cdot \frac{5}{4} = 95$$

$$\hat{S}_2^2 = S_2^2 \cdot \frac{n_2}{n_2 - 1} = 216 \cdot \frac{5}{4} = 270$$

Con lo que calculamos el estadístico de contraste y los grados de libertad.

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}} = \frac{17 - 38}{\sqrt{\frac{95}{5} + \frac{270}{5}}} = -2{,}46$$

$$gl = \frac{\left(\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}\right)^2}{\frac{\left(\hat{S}_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(\hat{S}_2^2/n_2\right)^2}{n_2 - 1}} = \frac{\left(\frac{95}{5} + \frac{270}{5}\right)^2}{\frac{(95/5)^2}{5-1} + \frac{(270/5)^2}{5-1}} = 6,50 \approx 6$$

Establecer la regla de decisión en función del nivel de confianza. Buscamos en las tablas t de Student el valor que supera a una proporción igual a 0,05 para 6 grados de libertad, obteniendo un valor igual a: $t_{6;0,05} = -1,943$. En el Figura 3.3 representamos los datos del ejemplo.

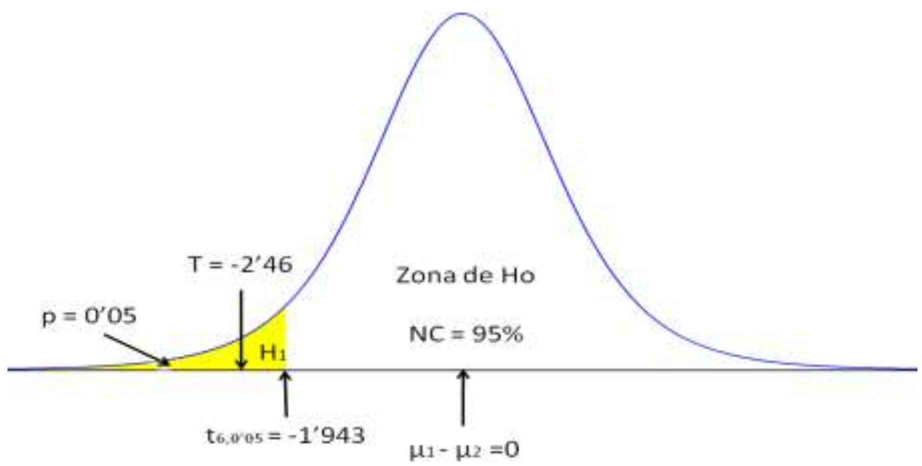


Figura 3.3. *Distribución muestral de las diferencias entre medias según especifica la hipótesis nula. Varianzas poblacionales desconocidas pero supuestas distintas ($\sigma_1^2 \neq \sigma_2^2$).*

Conclusión. El valor del estadístico de contraste es una puntuación más extrema que el valor crítico que hemos buscado en la tabla t de Student ($-2,46 < -1,943$), por lo que rechazamos la hipótesis nula. Con la misma lógica que en todos los contrastes, en general en un contraste unilateral izquierdo mantendremos la hipótesis nula cuando se cumpla que, $T > t_{g.l.,\alpha}$ y la rechazaremos si $T < t_{g.l.,\alpha}$

En cuanto al nivel crítico p, en la tabla t de Student, para 6 grados de libertad, tenemos que: ($-3,143 < -2,46 < -2,447$), por lo que deducimos que el nivel crítico p estará comprendido entre las probabilidades de encontrar valores iguales o inferiores a estas dos puntuaciones, es decir: ($0,01 < p < 0,025$).

Interpretar el resultado en función del contexto de la investigación. Con un nivel de confianza del 95% la diferencia de medias es significativa, por lo que concluimos que el fármaco reduce la ansiedad.

3.4.5.- Consideraciones sobre los contrastes de hipótesis en dos muestras independientes.

En el primer contraste de hipótesis incluíamos en los supuestos que las varianzas poblacionales son conocidas, lo que difícilmente podremos asumir en un caso práctico. Si no conocemos las medias de las poblaciones con las que trabajamos, difícilmente podremos considerar que sí conocemos sus varianzas.

Lo más habitual, por lo tanto, será asumir que las varianzas poblacionales son desconocidas, y en este caso el contraste más utilizado es la prueba T descrita en el punto 3.4.3, en el que suponemos varianzas poblacionales iguales. Este supuesto, al que denominaremos **homocedasticidad**, es muy común en otras técnicas estadísticas, como veremos en los temas en que compararemos las medias de más de dos grupos (Análisis de la Varianza) o en el Análisis de Regresión. De hecho, algunos manuales de estadística tan sólo describen este procedimiento para contrastar diferencias entre dos medias de muestras independientes. La cuestión estriba en que podamos asumir la normalidad de la distribución muestral de las diferencias, lo que podremos garantizar si las muestras que utilizamos son grandes. Si la distribución muestral es normal y los tamaños de ambas muestras son iguales, podemos despreocuparnos de las varianzas poblacionales y suponer sin más que son iguales, sin que por ello peligre la validez del contraste de hipótesis que estamos realizando.

Ahora bien, habrá casos en los que la opción más acertada será suponer varianzas poblacionales distintas, y por lo tanto tendremos que utilizar el estadístico de contraste visto en el apartado 3.4.4. En la literatura científica sobre este tema se proponen diferentes procedimientos para ajustar los grados de libertad de la distribución muestral. No es nuestro objetivo profundizar en esta cuestión, que el lector interesado puede consultar en la bibliografía recomendada en este texto, por lo que nos hemos limitado a describir la solución propuesta por Welch (1938) que posiblemente sea la más utilizada. En definitiva, el procedimiento de Welch nos ofrece un valor inferior para los grados de libertad en relación a si tomamos "$n_1 + n_2 - 2$", el contraste por lo tanto es más conservador, siendo más difícil rechazar la hipótesis nula.

Muchos investigadores sugieren que ha de realizarse previamente un contraste de hipótesis sobre la igualdad las varianzas, de manera que si aceptamos la hipótesis nula ($H_0: \sigma_1^2 = \sigma_2^2$) las supondremos iguales y en caso contrario diferentes. Veremos a continuación cómo llevar a cabo dicho contraste, que tampoco está exento de problemas que también podemos consultar en la bibliografía recomendada.

3.5.- Contraste de hipótesis sobre dos varianzas en muestras independientes.

Ejemplo 3.4. Según Eysenck (1981), hombres y mujeres tienen la misma puntuación media en cociente intelectual (CI), pero distinta varianza, siendo esta superior para los hombres. Para comprobar la hipótesis de Eysenck, seleccionamos aleatoriamente una muestra de 41 hombres y otra de 31 mujeres. Tras aplicar un test de inteligencia en ambas muestras, observamos que la cuasivarianza en el grupo de hombres es igual a 289, mientras que en el de mujeres vale 225. Con un nivel de confianza del 99% ¿avalan estos datos la hipótesis de Eysenck?

Condiciones y supuestos. Asumimos que las puntuaciones que nos proporcionan los tests de inteligencia miden este constructo en una escala de intervalo, y que la variable medida se distribuye normalmente tanto en la población de hombres como en la de mujeres. En general, los supuestos necesarios son:

- Variable dependiente con un nivel de medida de intervalo o razón.
- Dos poblaciones con variables normalmente distribuidas, o bien $n_1 \geq 30$ y $n_2 \geq 30$.

Formular las hipótesis. Plantearemos un contraste unilateral derecho, en el que la hipótesis alternativa corresponderá a la sugerida por Eysenck, e indicará que la variabilidad en inteligencia es superior en el grupo de hombres. Estadísticamente planteamos las hipótesis de la siguiente forma:

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} \leq 1 \qquad \text{o bien:} \qquad H_0 : \sigma_1^2 \leq \sigma_2^2$$
$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1 \qquad \qquad \qquad H_1 : \sigma_1^2 > \sigma_2^2$$

Elección del estadístico de contraste y su distribución muestral. El estadístico de contraste, sigue una distribución muestral "F" de Fisher, y es calculado según la Ecuación 3.7 :

$$F = \frac{\hat{S}_1^2}{\hat{S}_2^2} = \frac{289}{225} = 1{,}28 \qquad \text{(Ecuación 3.7)}$$

Los grados de libertad del numerador y denominador son, respectivamente: $n_1 - 1 = 41 - 1 = 40$, y $n_2 - 1 = 31 - 1 = 30$.

El cálculo del nivel crítico p mediante las tablas de las que disponemos, generalmente no podrá ser muy aproximado. Observamos en dichas tablas que el primer valor que nos ofrecen para 40 y 30 grados de libertad es igual a 1,573, al que supera una proporción igual a 0,10, luego con las tablas, tan sólo podemos saber que el nivel crítico p es mayor que 0,10 (p > 0,10). Con un programa informático adecuado concluiríamos que el valor exacto de p es 0,2432.

Establecer la regla de decisión en función del nivel de confianza. A un nivel de confianza del 99% para 40 y 30 grados de libertad, el valor crítico es igual a 2,299. La Figura 3.4 refleja los datos del problema.

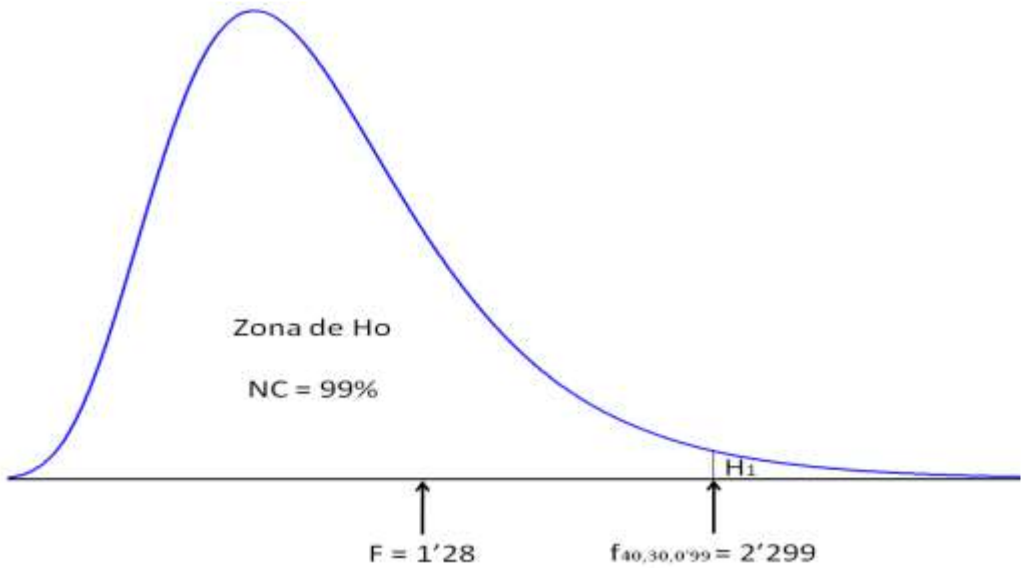


Figura 3.4. *Distribución muestral del cociente de dos varianzas en muestras independientes.*

Conclusión. A la vista de los resultados mantenemos la hipótesis nula a un nivel de confianza del 99%, puesto que el valor del estadístico de contraste es inferior al valor crítico, que en una distribución F de Fisher con 40 y 30 grados de libertad deja por encima a una proporción igual a 0,01. Deducimos del valor del nivel crítico p, que los resultados obtenidos están lejos de ser significativos para cualquier nivel de confianza razonable. Concluimos por lo tanto que la varianza de hombres y mujeres en inteligencia es la misma.

Interpretar el resultado en función del contexto de la investigación. Eysenck (1981) afirmaba que el hecho de que los hombres mostrasen mayor variabilidad en inteligencia, implica que hay más hombres que mujeres con CI muy altos y con CI muy bajos. Literalmente afirmaba "Esto está de acuerdo con la observación común de que la mayoría de los genios en ciencias, en artes o en otras ocupaciones, así como con defectos mentales, son hombres". Los datos que manejamos no avalan la hipótesis de Eysenck al no mostrar diferencias significativas en cuanto a la variabilidad en la inteligencia de ambos grupos.

Propiedad recíproca de la distribución F. Aunque en este contraste no ha sido necesario utilizarla, recordamos la propiedad recíproca de la distribución F que vimos en la asignatura Introducción al Análisis de Datos, y que nos sirve para calcular probabilidades que no aparecen en la tabla:

$$f_{n_1-1,n_2-1,\alpha/2} = \frac{1}{f_{n_2-1,n_1-1,1-\alpha/2}}$$

3.6.- Contrastes de hipótesis sobre dos proporciones en muestras independientes.

Tenemos ahora una variable dependiente dicotómica o dicotomizada, en la que podemos distinguir entre dos sucesos que denominamos "éxito" y "fracaso", y nos preguntamos si la proporción de éxitos difiere o no en dos poblaciones distintas. Este es un caso muy común en Psicología, donde con frecuencia trabajamos directamente con datos dicotómicos. Por ejemplo: hombre y mujer, acierto o fracaso en una determinada tarea, posición a favor o en contra, recuperación o no de una enfermedad tras aplicar una determinada terapia, etc. Otras veces dicotomizamos una variable que originalmente es continua, como aprobados y suspensos, admitidos y no admitidos en función de las puntuaciones en un test, etc.

Una de las ventajas de trabajar con datos dicotómicos reside en que los supuestos en los que nos basamos no son tan fuertes como en el caso de variables continuas. Tan sólo necesitamos que los tamaños de las muestras sean razonables $(n_1 \geq 30 \; y \; n_2 \geq 30)$ para poder asumir que la distribución muestral de las diferencias de proporciones es normal.

El estadístico de contraste que aplicaremos en este apartado dependerá de cómo planteemos las hipótesis estadísticas. En primer lugar veremos cómo proceder cuando queramos comprobar si la diferencia entre dos proporciones es igual, mayor o menor que cero. Posteriormente trataremos el caso en el que estemos interesados en probar si la diferencia de proporciones es igual, mayor o menor que un determinado valor distinto de cero.

Ejemplo 3.5. En unas determinadas oposiciones se presentaron 200 graduados por la UNED, de los que aprobaron 70, mientras que de 300 candidatos de otras universidades aprobaron 60. Con un nivel de confianza del 95%, ¿podemos afirmar que la proporción de aprobados es la misma entre los graduados de la UNED y los de otras universidades?

Condiciones y supuestos. Tenemos una variable dicotómica y dos muestras grandes que proceden de poblaciones independientes. En general, los supuestos necesarios son:

- Observaciones aleatorias e independientes.
- Variable dependiente dicotómica o dicotomizada.
- Muestras grandes. $n_1 \geq 30 \; y \; n_2 \geq 30$.

Formular las hipótesis. Queremos comprobar si la proporción de aprobados es igual o diferente en dos poblaciones distintas, es decir, como no tenemos ninguna hipótesis de partida, planteamos un contraste bilateral.

$$\begin{matrix} H_0: \pi_1 - \pi_2 = 0 \\ H_1: \pi_1 - \pi_2 \neq 0 \end{matrix} \quad o \; bien, \quad \begin{matrix} H_0: \pi_1 = \pi_2 \\ H_1: \pi_1 \neq \pi_2 \end{matrix}$$

Estadístico de contraste y su distribución muestral. Al plantear la hipótesis nula la igualdad de proporciones, el estadístico de contraste que debemos utilizar es el que muestra la Ecuación 3.8:

$$Z = \frac{p_1 - p_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{(Ecuación 3.8)}$$

Bajo el supuesto de que la hipótesis nula es verdadera y las dos proporciones poblacionales son iguales, la proporción P es la proporción ponderada de éxitos obtenidos en las dos muestras (véase la Ecuación 3.9), como mejor estimador de la proporción poblacional, π:

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \qquad \text{(Ecuación 3.9)}$$

Calculamos en primer lugar las proporciones muestrales y el término P:

$$p_1 = \frac{70}{200} = 0{,}35, \qquad p_2 = \frac{60}{300} = 0{,}20, \qquad P = \frac{200 \cdot 0{,}35 + 300 \cdot 0{,}20}{200 + 300} = 0{,}26$$

La proporción P, no es más que la proporción total de éxitos, que también podemos calcular como:

$$P = \frac{70 + 60}{200 + 300} = 0{,}26$$

El estadístico de contraste es igual a:

$$Z = \frac{p_1 - p_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0{,}35 - 0{,}20}{\sqrt{0{,}26(1-0{,}26)\left(\frac{1}{200} + \frac{1}{300}\right)}} = 3{,}75$$

Establecer la regla de decisión en función del nivel de confianza. Para un nivel de confianza del 95% y un contraste bilateral, los valores críticos son $z = \pm 1{,}96$. En la Figura 3.5 representamos los datos del problema.

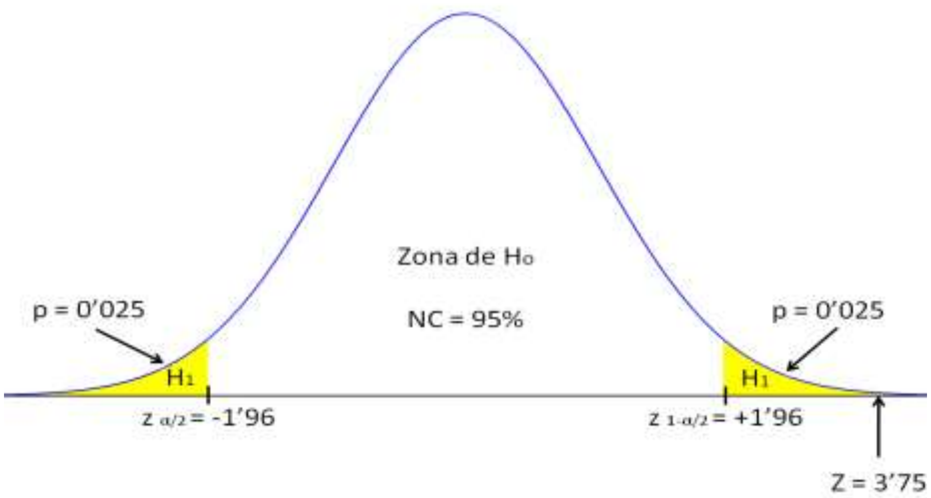


Figura 3.5. *Distribución muestral de dos proporciones en muestras independientes según especifica la hipótesis nula.*

Conclusión. A un nivel de confianza del 95% rechazamos la hipótesis nula, puesto que el estadístico de contraste cae fuera del intervalo que definen los valores críticos $\pm 1{,}96$, como podemos apreciar en el Figura 3.5. En cuanto al nivel crítico p, acudimos a las tablas de curva normal, observando que el valor mayor que podemos consultar es 3,59, y que a este valor le corresponde una probabilidad igual a $0{,}0002$. Como el contraste es bilateral, el nivel crítico p es igual o inferior a: $0{,}0002 \times 2 = 0{,}0004$. Con un programa informático adecuado, podemos comprobar que el nivel crítico es igual a $0{,}0002$. En cualquier caso, el resultado supera ampliamente el nivel de confianza fijado por el investigador.

Interpretar el resultado en función del contexto de la investigación. Los resultados muestran que la proporción de opositores que aprueban es diferente entre los graduados de la UNED y otras universidades.

Ejemplo 3.6. Según los datos que maneja el director de una academia especializada en preparar a sus alumnos para unas determinadas oposiciones, la proporción de aprobados entre los titulados de la UNED es 0,15 puntos superior a la proporción de aprobados de los titulados procedentes de otras universidades. El director sospecha que el presente curso dicha proporción será superior a 0,15. Para comprobar esta hipótesis extrae una muestra aleatoria de 60 alumnos procedentes de la UNED y otra de 100 alumnos procedentes de otras universidades. Somete a ambas muestras a un examen con el temario de las oposiciones, que es superado por 33 alumnos de la UNED y 30 de otras universidades. Con un nivel de confianza del 95%, ¿podemos afirmar que los datos que maneja el director de la academia son correctos?

Condiciones y supuestos. Tenemos dos muestras independientes con una variable dependiente dicotómica, y estamos interesados en comprobar si la diferencia entre las proporciones poblacionales es superior a 0,15. La variable dependiente es dicotómica y ambas muestras superan las 30 observaciones.

Hipótesis. Planteamos un contraste unilateral derecho.

$$H_0: \pi_1 - \pi_2 \leq 0{,}15$$
$$H_1: \pi_1 - \pi_2 > 0{,}15$$

Estadístico de contraste y su distribución muestral. Según la Ecuación 3.10:

$$Z = \frac{(p_1 - p_2) - D}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \qquad \text{(Ecuación 3.10)}$$

Con los datos del ejemplo (siendo D el valor propuesto en H_0):

$$Z = \frac{(0{,}55 - 0{,}30) - 0{,}15}{\sqrt{\frac{0{,}55(1-0{,}55)}{60} + \frac{0{,}30(1-0{,}30)}{100}}} = 1{,}27$$

Establecer la regla de decisión en función del nivel de confianza. Al nivel de confianza del 95%, para un contraste unilateral derecho, el valor crítico es igual a 1,64.

Conclusión. Dado que el estadístico de contraste es menor que el valor crítico, mantenemos la hipótesis nula. Buscando en las tablas de curva normal la probabilidad de encontrar puntuaciones típicas iguales o superiores a 1,27, deducimos que el nivel crítico p es igual a: $p = 0{,}102$.

Interpretación del resultado en función del contexto de la investigación. Con un nivel de confianza del 95%, y a pesar de que la proporción de aprobados entre los alumnos de la UNED supera a la de otras universidades en una proporción igual a 0,25, no podemos afirmar, con los datos que tenemos, que en la población esta proporción será superior a 0,15.

3.7.- Tamaño del efecto.

Como ya se indicó previamente, la magnitud o tamaño del efecto es el nombre que se da a una familia de índices que miden el efecto que tiene un tratamiento. Es decir, es un índice que se aplica cuando hay implicados al menos dos grupos, uno de tratamiento y otro de control. Difiere de los contrastes clásicos en que es independiente del tamaño muestral. Este tipo de índices son de uso frecuente en el ámbito del meta-análisis aplicado a la psicología, educación, etc.

Vamos a ver un ejemplo que nos ayude a comprender la importancia de estudiar el tamaño del efecto. Supongamos que un profesor elabora un material que si bien consume muchas horas de estudio, pretende ayudar a sus alumnos a mejorar el rendimiento académico. Lógicamente, nuestro profesor parte de la hipótesis de que las notas serán superiores si los alumnos utilizan el material al que nos referimos. Para comprobar su hipótesis, el profesor utiliza dos muestras aleatorias de 900 alumnos cada una, obteniendo una nota media igual a 5,5 en el grupo que ha utilizado el nuevo material (Grupo 1), y una nota media igual a

5 en el grupo que no ha utilizado el nuevo material (Grupo 2). Vamos a suponer que conocemos las varianzas poblacionales y que ambas valen 12,5. Plantemos un contraste unilateral derecho cuyas hipótesis son:

$$H_0: \mu_1 - \mu_2 \leq 0$$
$$H_1: \mu_1 - \mu_2 > 0$$

Siendo el valor del estadístico de contraste:

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{5{,}5 - 5}{\sqrt{\frac{12{,}5}{900} + \frac{12{,}5}{900}}} = 3$$

Acudiendo a la tabla de curva normal comprobamos que el nivel crítico p vale: 0,0013. Por lo tanto los resultados son significativos, superando con creces un nivel de confianza del 99%.

Ahora bien, el incremento de la nota (es decir, el tamaño del efecto) es muy pequeño, por lo que, teniendo en cuenta que el nuevo material consume muchas horas de estudio, ¿merece la pena emplear dicho material? Si tomamos una decisión teniendo en cuenta únicamente que se han obtenido resultados significativos, concluiríamos que sí merece la pena, pero si valoramos el tamaño del efecto (un incremento de tan sólo medio punto), probablemente concluiríamos que no merece la pena utilizar el nuevo material. Por otro lado, es obvio que los resultados son significativos porque el tamaño de las muestras es muy grande, porque, por ejemplo, con muestras $n_1 = n_2 = 100$, el mismo efecto (medio punto de mejora en el rendimiento académico) no habría resultado significativo.

Es importante, por lo tanto, desarrollar medidas que cuantifiquen cuál es el tamaño del efecto, medidas que, por otro lado, han de ser independientes del tamaño muestral.

Aunque hay una amplia variedad de fórmulas, vamos a ilustrar el proceso básico de obtención del efecto con el denominado índice "*d*". Básicamente este índice no es más que la estandarización de una diferencia en dos medias, una que sería la del grupo al que se ha aplicado un determinado tratamiento, y otra, la del llamado grupo control. La fórmula del índice viene dada en la Ecuación 3.11.

$$d = \frac{|\bar{Y}_{Tratamiento} - \bar{Y}_{Control}|}{\hat{\sigma}} \qquad \text{(Ecuación 3.11)}$$

Siendo $\hat{\sigma}$ la desviación típica conjunta de ambos grupos, pero cuando sus varianzas son homogéneas. Para el caso en que las varianzas no sean homogéneas, hay otras fórmulas más genéricas en donde se sustituye $\hat{\sigma}$ por el promedio ponderado de las desviaciones típicas insesgadas de ambos grupos (véase la Ecuación 3.12), es decir:

$$\sqrt{\frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}} \qquad \text{(Ecuación 3.12)}$$

Para interpretar el resultado de este índice, y teniendo en cuenta que es un medida estandarizada, Cohen (1988) propuso una gradación de la magnitud del efecto en "pequeño: d = 0,2", "mediano: d= 0,5" y "grande: d = 0,8 o superior". Veamos un sencillo ejemplo.

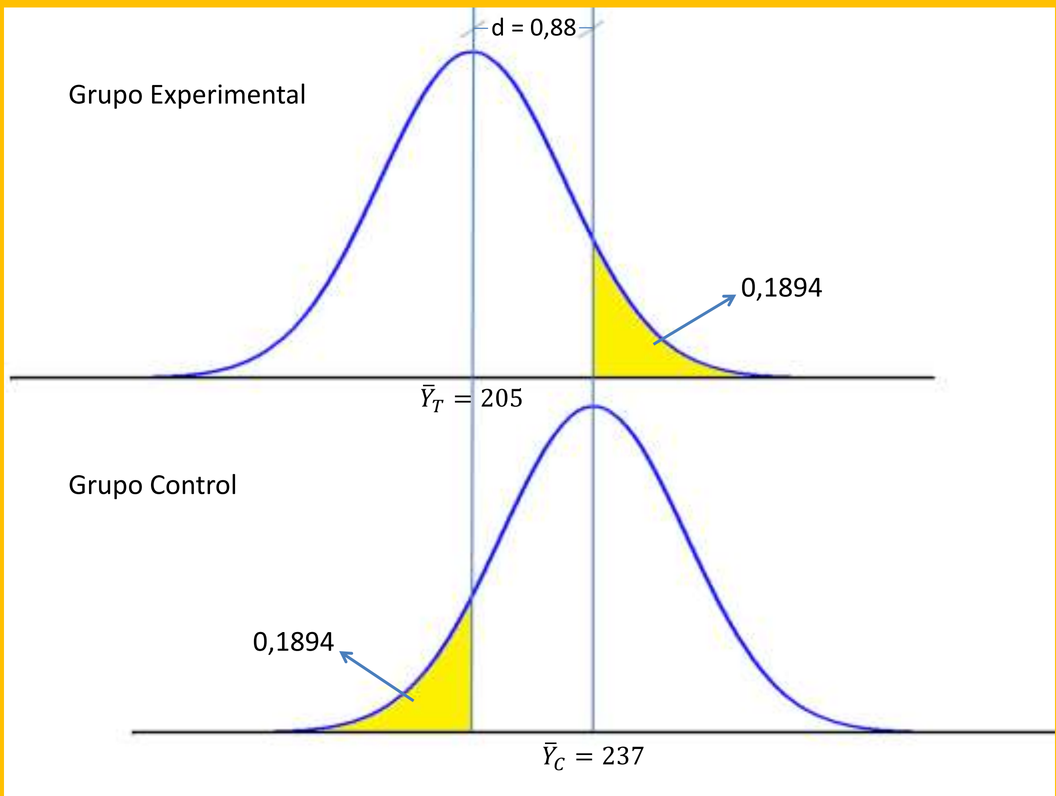
Ejemplo 3.7. Se realiza un experimento por el que se trata de estudiar si la verbalización del proceso facilita la realización de tareas manuales complejas. Se seleccionan aleatoriamente 60 sujetos y se asignan 30 a cada uno de dos grupos: el experimental, en el cual los sujetos verbalizan la tarea, y el de control, en el que los sujetos realizan la tarea en silencio. Como variable dependiente se registra el tiempo en segundos que se requiere para completar la tarea. Para el grupo control se obtuvo una media igual a 237 y cuasidesviación típica igual a 38 y para el experimental media igual a 205 y cuasidesviación típica igual a 35. El índice d de Cohen para cuantificar la mejora en rapidez que se produce al verbalizar la tarea es:

$$d = \frac{|\bar{Y}_{Tratamiento} - \bar{Y}_{Control}|}{\sqrt{\dfrac{(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2}{n_1+n_2-2}}} = \frac{|205-237|}{\sqrt{\dfrac{29 \cdot 35^2 + 29 \cdot 38^2}{30+30-2}}} = 0,876 \approx 0,88$$

0,88 es la distancia estandarizada entre las medias de los dos grupos, y su probabilidad asociada es 0,8106, lo que indica que el 81,06% de los sujetos del grupo experimental tardan menos tiempo que el promedio de los sujetos que no verbalizan. Sólo un 18,94% de los niños que no verbalizan tardan menos tiempo que el promedio de los que sí lo hacen. En la Figura 3.6 se observa la situación del grupo de control respecto del experimental.

Figura 3.6

Magnitud del efecto del grupo experimental respecto del grupo control



3.8.- Resumen.

En el resumen del Tema 2 veíamos cuáles son los pasos que hemos de seguir en cualquier contraste de hipótesis. Dicho resumen sigue vigente puesto que hemos seguido los mismos pasos para aplicar los contrastes que hemos desarrollado. Además son pocos los conceptos estadísticos nuevos que hemos introducido. Por ejemplo, en el caso de dos medias procedentes de muestras independientes, cambia la manera de obtener la distribución muestral, pero al igual que en el Tema 2, llegamos a una distribución muestral que se distribuye según la curva normal, cuando conocemos las varianzas poblacionales o T de Student, cuando desconocemos las varianzas poblacionales. También cambian las hipótesis postuladas que, lógicamente, ahora plantean diferencias entre dos medias poblacionales.

Gran parte de los conceptos tratados en este tema ya los hemos visto en la asignatura Fundamentos de Investigación, por lo que recomendamos repasar el Tema 5 de dicha asignatura.

Entre los conceptos nuevos destacamos la distinción entre muestras independientes y relacionadas, que es imprescindible para aplicar el análisis de datos adecuado cuando trabajamos con dos muestras. También es nuevo el apartado dedicado a la magnitud del efecto, índice que es independiente del tamaño muestral y que complementa al valor del estadístico de contraste ayudándonos a interpretar la magnitud de las diferencias observadas.

Cada uno de los contrastes tratados ha sido ilustrado a través de un ejemplo, por lo que no hemos visto cómo proceder cuando cambian las hipótesis planteadas. Por ello, a continuación presentamos en las Tablas 3.4 a la 3.7 el resumen de cada una de las pruebas de este tema.

Tabla 3.4. *Contrastes de hipótesis para dos medias en muestras independientes.*

SUPUESTOS	- Observaciones independientes. - Nivel de medida de intervalo o razón. - Distribuciones normales en la población ó $(n_1 \geq 30 \;\; y \;\; n_2 \geq 30)$.		
Varianzas poblacionales	**Conocidas**	**Desconocidas y supuestas iguales**	**Desconocidas y supuestas distintas**
HIPÓTESIS	*Igual en los tres casos. En función de la hipótesis científica plantearemos un contrate:* Bilateral: $H_0: \mu_1 - \mu_2 = 0$; $H_1: \mu_1 - \mu_2 \neq 0$	Unilateral izquierdo: $H_0: \mu_1 - \mu_2 \geq 0$; $H_1: \mu_1 - \mu_2 < 0$	Unilateral derecho: $H_0: \mu_1 - \mu_2 \leq 0$; $H_1: \mu_1 - \mu_2 > 0$
ESTADÍSTICO DE CONTRASTE	$Z = \dfrac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$T = \dfrac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\dfrac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$	$T = \dfrac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\dfrac{\hat{S}_1^2}{n_1} + \dfrac{\hat{S}_2^2}{n_2}}}$
DISTRIBUCIÓN MUESTRAL	Normal tipificada $N(0{,}1)$	"t" de Student $g.l. = n_1 + n_2 - 2$	"t" de Student $gl = \dfrac{\left(\dfrac{\hat{S}_1^2}{n_1} + \dfrac{\hat{S}_2^2}{n_2}\right)^2}{\dfrac{\left(\hat{S}_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(\hat{S}_2^2/n_2\right)^2}{n_2 - 1}}$

REGLA DE DECISIÓN

Distribución normal (varianzas conocidas).

	bilateral	Unilateral Derecho	Unilateral Izquierdo
Mantener H_0	$z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}$	$Z < z_{1-\alpha}$	$Z > z_{\alpha}$
Rechazar H_0	$Z < z_{\frac{\alpha}{2}}$ ó $Z > z_{1-\frac{\alpha}{2}}$	$Z > z_{1-\alpha}$	$Z < z_{\alpha}$

Distribución "t" de Student (varianzas desconocidas).

	bilateral	Unilateral Derecho	Unilateral Izquierdo
Mantener H_0	$t_{g.l.,\frac{\alpha}{2}} < T < t_{g.l.,1-\frac{\alpha}{2}}$	$T < t_{g.l.,1-\alpha}$	$T > t_{g.l.,\alpha}$
Rechazar H_0	$T < t_{g.l.,\frac{\alpha}{2}}$ ó $T > t_{g.l.,1-\frac{\alpha}{2}}$	$T > t_{g.l.,1-\alpha}$	$T < t_{g.l.,\alpha}$

Tabla 3.5. *Contrastes de hipótesis para dos proporciones en muestras independientes.*

SUPUESTOS	- Observaciones independientes. - Variable dependiente dicotómica o dicotomizada. - $n \geq 30$	
HIPÓTESIS	*En función de la hipótesis científica plantearemos un contrate:* Bilateral: $H_0: \pi_1 - \pi_2 = D$, $H_1: \pi_1 - \pi_2 \neq D$ Unilateral izquierdo: $H_0: \pi_1 - \pi_2 \geq D$, $H_1: \pi_1 - \pi_2 < D$ Unilateral derecho: $H_0: \pi_1 - \pi_2 \leq D$, $H_1: \pi_1 - \pi_2 > D$	
ESTADÍSTICO DE CONTRASTE	$\boldsymbol{D = 0}$ $Z = \dfrac{p_1 - p_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ Donde: $P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$	$\boldsymbol{D \neq 0}$ $Z = \dfrac{(p_1 - p_2) - D}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$
DISTRIBUCIÓN MUESTRAL	Normal tipificada. $N(0,1)$	
REGLA DE DECISIÓN	(ver tabla siguiente)	

REGLA DE DECISIÓN

	bilateral	Unilateral Derecho	Unilateral Izquierdo
Mantener H_0	$z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}$	$Z < z_{1-\alpha}$	$Z > z_{\alpha}$
Rechazar H_0	$Z < z_{\frac{\alpha}{2}}$ ó $Z > z_{1-\frac{\alpha}{2}}$	$Z > z_{1-\alpha}$	$Z < z_{\alpha}$

<table>
<tr><td colspan="2">Tabla 3.6. Contrastes de hipótesis para dos varianzas en muestras independientes.</td></tr>
<tr><td>SUPUESTOS</td><td>- Variable dependiente con un nivel de medida de intervalo o razón.
- Dos poblaciones con variables normalmente distribuidas, o bien $n_1 \geq 30$ y $n_2 \geq 30$.</td></tr>
<tr><td>HIPÓTESIS</td><td>
<table>
<tr><td>Bilateral</td><td>Unilateral izquierdo</td><td>Unilateral derecho</td></tr>
<tr><td>$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$
$H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$</td><td>$H_0: \frac{\sigma_1^2}{\sigma_2^2} \geq 1$
$H_1: \frac{\sigma_1^2}{\sigma_2^2} < 1$</td><td>$H_0: \frac{\sigma_1^2}{\sigma_2^2} \leq 1$
$H_1: \frac{\sigma_1^2}{\sigma_2^2} > 1$</td></tr>
</table>
</td></tr>
<tr><td>ESTADÍSTICO DE CONTRASTE</td><td>$$F = \frac{\hat{S}_1^2}{\hat{S}_2^2}$$</td></tr>
<tr><td>DISTRIBUCIÓN MUESTRAL</td><td>"f" de Fisher
$\mathrm{g.l.\,numerador} = \mathrm{n}_1 - 1$
$\mathrm{g.l.\,denominador} = \mathrm{n}_2 - 1$</td></tr>
<tr><td>REGLA DE DECISIÓN</td><td>
<table>
<tr><td></td><td>Mantener H_0</td><td>Rechazar H_0</td></tr>
<tr><td>Bilateral</td><td>$f_{n_1-1,n_2-1,\alpha/2} < F < f_{n_1-1,n_2-1,1-\alpha/2}$</td><td>$F < f_{n_1-1,n_2-1,\alpha/2}$
O bien:
$F > f_{n_1-1,n_2-1,1-\alpha/2}$</td></tr>
<tr><td>Unilateral Derecho</td><td>$F < f_{n_1-1,n_2-1,1-\alpha}$</td><td>$F > f_{n_1-1,n_2-1,1-\alpha}$</td></tr>
<tr><td>Unilateral Izquierdo</td><td>$F > f_{n_1-1,n_2-1,\alpha}$</td><td>$F < f_{n_1-1,n_2-1,\alpha}$</td></tr>
</table>
</td></tr>
</table>

3.9.- Ejercicios de autocomprobación.

1. En un caso práctico en el que estamos interesados en contrastar la diferencia entre las medias de dos poblaciones distintas, lo más habitual es que las varianzas poblacionales sean: a) conocidas; b) desconocidas; c) ninguna de las anteriores es correcta.

2. Si la magnitud del efecto es grande: a) la diferencia entre dos medias es significativa; b) el tamaño de las muestras es grande; c) la diferencia de medias puede ser significativa o no dependiendo del tamaño de las muestras.

3. Si en un contraste unilateral trabajamos con un nivel de confianza del 95%, llamamos a la probabilidad "$1 - 0.95 = 0.05$": a) nivel de significación; b) nivel crítico p; c) las dos opciones anteriores son verdaderas.

4. En un contraste bilateral sobre dos varianzas al nivel de confianza del 95%, los tamaños muestrales son: $n_1 = 21 \ y \ n_2 = 31$, por lo que los valores críticos son: a) 0,426 y 2,195; b) 0,456 y 2,195; c) 0,426 y 2,347.

5. En un contraste de hipótesis unilateral sobre dos proporciones hemos obtenido un estadístico de contrate igual a $Z = 1'80$. ¿Con qué nivel de confianza es significativo? a) 0,95; b) 0,99; c) ninguna de las anteriores.

En un estudio sobre sensibilidad gustativa, un investigador desea comprobar si la sensibilidad gustativa para la fructosa es menor en los fumadores que en los no fumadores, para lo que dispone de dos muestras aleatorias. Tras calcular el umbral absoluto para esta sustancia en ambos grupos, los resultados son:

Fumadores: $\bar{Y}_1 = 2'3, \quad \hat{S}_1^2 = 5'2, \ \ n_1 = 34$

No fumadores: $\bar{Y}_2 = 1'25, \ \ \hat{S}_2^2 = 4'3, \ \ n_2 = 38$

Recordando que a mayor umbral absoluto menor sensibilidad, asumiendo que las varianzas poblacionales son iguales y con un nivel de confianza del 95%, conteste a las siguientes cuestiones:

6. La hipótesis nula es: a) $\mu_1 - \mu_2 \neq 0$; b) $\mu_1 - \mu_2 \leq 0$; c) $\mu_1 - \mu_2 \geq 0$

7. El valor del estadístico de contraste es igual a: a) 1,667; b) 1,994; c) 2,046.

8. La magnitud del efecto es igual a: a) 0,05; b) 0,483; c) 1,05.

9. Utilizando las tablas del texto, el valor crítico está comprendido entre los valores: a) 0,025 y 0,05; b) 0,01 y 0,025; c) 0,01 y 0,005.

10. La conclusión de este estudio es: a) mantener la hipótesis nula porque $p < \alpha$; b) rechazar la hipótesis nula porque $p < \alpha$; c) rechazar la hipótesis nula porque el estadístico de contraste es menor que el valor crítico.

Soluciones.

1. La solución es "b". En un caso real difícilmente conoceremos las varianzas poblacionales.

2. La respuesta correcta es "c". Aunque el tamaño del efecto sea grande, si la muestra es pequeña podemos obtener resultados no significativos.

3. la respuesta correcta es "a", el nivel crítico p es una probabilidad asociada al estadístico de contraste.

4. Tenemos que buscar en la tabla f de Fisher los valores que superan las probabilidades 0,025 y 0,975 con 20 y 30 grados de libertad. En la tabla encontramos:

$$f_{20;30;975} = 2{,}195$$

Y aplicando la propiedad recíproca de la distribución f:

$$f_{20;30;0{,}025} = \frac{1}{f_{30;20;0{,}975}} = \frac{1}{2{,}349} = 0{,}426$$

Luego la respuesta correcta es "a".

5. La solución es "a". El estadístico de contraste supera al valor crítico al nivel de confianza del 95% (1,80 > 1,64), pero no supera el valor crítico correspondiente al nivel de confianza del 99% (1,80 < 2,33).

6. El investigador sospecha que el umbral absoluto es mayor en los fumadores (menor sensibilidad), que será lo que postula la hipótesis alternativa, luego la hipótesis nula es la que especifica la opción "b".

7. Tras realizar los cálculos oportunos comprobamos que la opción correcta es "c"

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{2{,}3 - 1{,}25}{\sqrt{\frac{33\cdot 5{,}2 + 37\cdot 4{,}3}{70}\left(\frac{1}{34}+\frac{1}{38}\right)}} = 2{,}046$$

8. La opción correcta es "b".

$$d = \frac{|\bar{Y}_{Tratamiento} - \bar{Y}_{Control}|}{\sqrt{\frac{(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2}{n_1+n_2-2}}} = \frac{|2{,}3 - 1{,}25|}{\sqrt{\frac{33\cdot 5{,}2 + 37\cdot 4{,}3}{70}}} = 0{,}483$$

9. Buscando en las tablas t de Student para 70 grados de libertad, observamos que el estadístico de contraste (T = 2,046) está comprendido entre los valores 1,994 y 2,381 que dejan por encima de sí las probabilidades 0,025 y 0,01, respectivamente. Luego la respuesta correcta es "b".

10. La respuesta correcta es "b".

Tema 4. ANÁLISIS DE DATOS PARA DISEÑOS DE DOS GRUPOS. MUESTRAS RELACIONADAS.

4.1.- Introducción.

Ya sabemos con lo que nos vamos a enfrentar en este tema, puesto que vimos en el anterior en qué consisten las muestras relacionadas y somos conscientes de su amplia utilización en la investigación en Psicología.

Las muestras relacionadas tienen una ventaja sobre las independientes que consiste en que nos ayudan a reducir la varianza de error, de manera que cuanto mayor sea la relación entre ambas muestras, menor será la varianza de la distribución muestral de las diferencias, obteniendo por lo tanto un estadístico de contraste mayor.

Comenzaremos por ilustrar cómo se compone la distribución muestral de las diferencias para muestras relacionadas (punto 4.3.1), para tratar, mediante dos ejemplos, el contraste para dos medias relacionadas cuando asumimos que conocemos la varianza poblacional de las diferencias (punto 4.3.2) y cuando asumimos, como es más habitual, que la desconocemos (punto 4.3.3). Pasaremos después al contraste de dos proporciones en muestras relacionadas (punto 4.4), para finalizar con el resumen (punto 4.5) y los ejercicios de autocomprobación (punto 4.6).

Los objetivos, aplicados ahora al caso de dos muestras relacionadas, son los mismos que en el tema anterior.

4.2.- Objetivos del tema.

- ✓ Plantear las hipótesis en función de los objetivos de la investigación
- ✓ Distinguir entre contraste unilateral y bilateral
- ✓ Seleccionar el estadístico de contraste más adecuado a las hipótesis planteadas
- ✓ Conocer la distribución muestral del estadístico seleccionado
- ✓ Realizar los cálculos oportunos para someter a contrastación empírica las hipótesis planteadas
- ✓ Relacionar el intervalo de confianza con el estadístico de contraste
- ✓ Interpretar el nivel crítico p
- ✓ Determinar e interpretar el o los valores críticos de la distribución muestral
- ✓ Tomar una decisión respecto a las hipótesis planteadas

4.3.- Contraste de hipótesis sobre dos medias en muestras relacionadas.

Aunque contamos con dos medias como en el tema anterior, la distribución muestral es diferente, por lo que comenzaremos este apartado con un breve ejemplo, cuyo fin es exclusivamente didáctico y que nos servirá para comprender cómo se compone la distribución muestral de las diferencias.

4.3.1.- Distribución muestral para dos medias relacionadas (OPCIONAL).

Supongamos que tenemos una población compuesta por 4 sujetos a los que medimos la variable dependiente antes y después de una terapia, siendo sus puntuaciones y sus diferencias las que podemos ver en la Tabla 4.1

Tabla 4.1
Ejemplo. Datos de cuatro sujetos antes y después de la terapia.

	Sujeto 1	Sujeto 2	Sujeto 3	Sujeto 4
Antes	8	6	4	10
Después	3	5	6	8
Diferencia	8 – 3 = 5	6 - 5= 1	4 – 6 = -2	10 – 8 = 2

La población sobre la que vamos a trabajar está formada por las diferencias entre las puntuaciones “antes” y “después”, o sea, por los valores, una vez ordenados: {-2; 1; 2; 5}. La media y varianza de la población de diferencias es:

$$\mu_d = \frac{6}{4} = 1'5, \qquad \sigma_d^2 = \frac{34}{4} - 1'5^2 = 6'25$$

Sobre la población de diferencias, tomamos todas las muestras de un determinado tamaño y las medias de todas ellas formarán la distribución muestral de las diferencias. Por ejemplo, para muestras de tamaño n=2, los resultados de todas las muestras con reposición pueden verse en la Tabla 4.2:

Tabla 4.2
Todas las posibles medias aritméticas para muestras de tamaño n=2

	-2	**1**	**2**	**5**
-2	-2	-0’5	0	1’5
1	-0’5	1	1’5	3
2	0	1’5	2	3’5
5	1’5	3	3’5	5

Por lo que la distribución muestral de las diferencias está formada por los valores:

{-2; -0’5; -0’5; 0; 0; 1; 1’5; 1’5; 1’5; 1’5; 2; 3; 3; 3’5; 3’5; 5}

La media y varianza de la distribución muestral podemos calcularla con los datos de la Tabla 4.1, o bien a través de los parámetros poblacionales como hemos visto en casos anteriores:

$$\mu_d = \frac{6}{4} = 1'5, \qquad \sigma_{\bar{Y}_1 - \bar{Y}_2}^2 = \frac{\sigma_d^2}{n} = \frac{6'25}{2} = 3'125$$

Podemos apreciar cómo el contraste de hipótesis sobre dos medias relacionadas es muy parecido al de una sola media. Aunque ahora partimos de dos muestras, al estar estas relacionadas, finalmente tenemos una sola variable (las diferencias entre cada par de puntuaciones). Para apreciar más claramente esta similitud, vamos a comparar el estadístico de contraste en ambos casos cuando la varianza poblacional es conocida (véase la Tabla 4.3.).

Tabla 4.3

Estadísticos de contraste para una media y dos relacionadas

Una media	Dos medias relacionadas
$$Z=\frac{\bar{Y}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$ Donde: $\bar{Y}$ = media de la muestra. σ^2= varianza de la población. μ_0 = media que postula la hipótesis nula	$$Z=\frac{\bar{D}-\mu_d}{\sqrt{\frac{\sigma_d^2}{n}}}$$ Donde: $\bar{D}$ = media de las diferencias de ambas muestras. σ_d^2 = varianza de la población de diferencias. μ_d = media de diferencias que postula la hipótesis nula. Habitualmente igual a cero

4.3.2.- Conocida la varianza poblacional de las diferencias.

Ejemplo 4.1. Un psicólogo escolar está interesado en estudiar si la presión de los padres para el rendimiento escolar es igual en chicos y en chicas. Toma una muestra aleatoria de 36 parejas de hermanos (chico y chica), y mediante un test que proporciona medidas en una escala de intervalo mide la variable "presión para el rendimiento escolar" en todos los individuos. La media para los chicos (Grupo 1) fue igual a 21 y para el grupo de chicas fue igual a 19. Suponemos que conocemos la varianza poblacional de las diferencias y que es igual a 64. A un nivel de confianza del 99%. ¿Es igual la presión para el rendimiento escolar en chicos y chicas?

Condiciones y supuestos. Las muestras de chicos y chicas son relacionadas dado que están compuestas por parejas de hermanos. Conocemos la varianza de las diferencias en la población, la variable dependiente está medida a un nivel de intervalo, y no sabemos si la población de diferencias se distribuye normalmente, pero la muestra supera las 30 observaciones. En general, los supuestos son:

- Variable dependiente con un nivel de medida de intervalo o razón.
- Población de diferencias normalmente distribuida o n ≥ 30.
- Varianza poblacional de las diferencias conocida.

Formular las hipótesis. El psicólogo no tiene una hipótesis previa sobre las diferencias debidas al género en la variable "presión para el rendimiento escolar", por lo que planteamos un contraste bilateral.

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{o bien,} \quad H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 - \mu_2 \neq 0 \qquad\qquad\quad H_1: \mu_1 \neq \mu_2$$

Elección del estadístico de contrate y su distribución muestral. El estadístico de contraste se distribuye normalmente, según la expresión de la Ecuación 4.1:

$$Z = \frac{\overline{D} - \mu_d}{\sqrt{\frac{\sigma_d^2}{n}}} = \frac{(21 - 19) - 0}{\sqrt{\frac{64}{36}}} = 1'5 \qquad \text{(Ecuación 4.1)}$$

Mediante la tabla de curva normal del apéndice deducimos que la probabilidad de encontrar valores superiores a una puntuación típica de 1'5 es igual a: 0'0668. Al ser el contraste bilateral, el nivel crítico p es: $2 \cdot 0'0668 = 0'1336$.

Establecer la regla de decisión en función del nivel de confianza. Al nivel de confianza del 99%, los valores críticos que delimitan la zona en la que mantenemos H_0 son: $z = \pm 2'58$. La Figura 4.1 muestra los datos del problema.

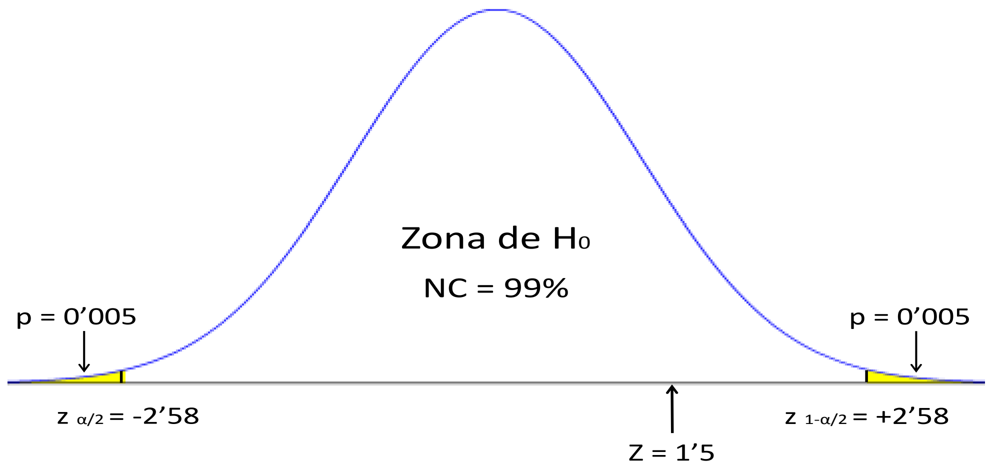


Figura 4.1.*Distribución muestral de las diferencias postulada por la hipótesis nula.*

Conclusión. Al nivel de confianza del 99% no existen diferencias significativas entre las medias de chicos y chicas, puesto que el estadístico de contraste se encuentra comprendido en el intervalo que definen los valores críticos, por lo que mantenemos la hipótesis nula.

Interpretar el resultado en función del contexto de la investigación. Según los datos que manejamos, no existen diferencias en cuanto a la presión para el rendimiento escolar entre chicos y chicas.

Intervalo de confianza. Viene dado por la expresión de la Ecuación 4.2:

$$\bar{D} \pm z_{\alpha/2}\sqrt{\frac{\sigma_d^2}{n}} \qquad \text{(Ecuación 4.2)}$$

Que con nuestros datos vale:

$$2 \pm 2'58\sqrt{\frac{64}{36}} \longrightarrow (-1'44;\ 5'44)$$

Observamos que el intervalo de confianza contiene al cero que es lo que postula la hipótesis nula.

4.3.3.- Desconocida la varianza poblacional de las diferencias.

Al igual que comentábamos en temas anteriores, lo más habitual en un caso práctico es que desconozcamos la varianza de la población, siendo el proceso muy parecido al que acabamos de ver. Simplemente tenemos que sustituir la varianza poblacional de las diferencias (σ_d^2) por su estimador, que es la cuasivarianza de las diferencias en la muestra. El estadístico de contraste en este caso se distribuye según t de Student con n - 1 grados de libertad.

Ejemplo 4.2. Un psicólogo que trabaja en una empresa imparte un curso sobre asertividad. El objetivo del curso consiste en fomentar esta habilidad en los directivos que forman parte de su departamento. Antes del curso mide la asertividad mediante un test que proporciona medidas en una escala de intervalo, y en el que las puntuaciones altas indican un comportamiento asertivo. Al finalizar el curso el psicólogo aplica de nuevo el test de asertividad a los asistentes. Las puntuaciones antes y después del curso fueron las siguientes:

Sujeto	1	2	3	4	5	6	7	8	9	10
Antes	18	24	25	24	27	30	24	31	24	28
Después	24	23	34	22	34	40	35	31	27	30

Con un nivel de confianza del 95%. Suponiendo que en la población la distribución de las diferencias es normal, ¿Podemos decir que el curso realizado por el psicólogo ha incrementado la asertividad de los directivos?

Antes de realizar el contraste de hipótesis vamos a calcular las diferencias y los cuadrados de las mismas entre las condiciones “Antes” y “Después” para cada par de sujetos. Estos cálculos, necesarios para calcular la media y la cuasivarianza insesgada, aparecen en la Tabla 4.4.

Tabla 4.4 *Puntuaciones antes y después en asertividad*										
Sujeto	1	2	3	4	5	6	7	8	9	10
Antes	18	24	25	24	27	30	24	31	24	28
Después	24	23	34	22	34	40	35	31	27	30
d_i	-6	1	-9	2	-7	-10	-11	0	-3	-2
d_i^2	36	1	81	4	49	100	121	0	9	4

La media y la varianza insesgada de las puntuaciones diferencia valen:

$$\bar{D} = \frac{\sum d_i}{n} = -4{,}5, \qquad S_d^2 = \frac{\sum d_i^2}{n} - (\bar{D})^2 = \frac{405}{10} - (-4{,}5)^2 = 20{,}25$$

$$\hat{S}_d^2 = S_d^2 \cdot \frac{n}{n-1} = 20{,}25 \cdot \frac{10}{9} = 22'5$$

Condiciones y supuestos. Según el enunciado del problema, la variable dependiente está medida a un nivel de intervalo. Hemos de suponer que la población de las diferencias sigue una distribución normal porque la muestra es pequeña y no conocemos su varianza. En general se tendrá que cumplir:

- Variable dependiente con un nivel de medida de intervalo o razón.
- Población de diferencias que se distribuye normalmente, o bien n ≥ 30.
- Varianza poblacional de las diferencias desconocida.

Asumiendo que se cumplen estos supuestos, pasamos a plantear las hipótesis.

Hipótesis. El psicólogo tiene la idea de que su curso incrementará las puntuaciones en asertividad, o sea, que la media en esta variable será menor antes del curso (grupo 1) que después del curso (grupo 2), por lo que podemos plantear un contraste unilateral izquierdo de la siguiente forma.

$$\begin{matrix} H_0: \mu_1 - \mu_2 \geq 0 \\ H_1: \mu_1 - \mu_2 < 0 \end{matrix} \quad \text{o bien,} \quad \begin{matrix} H_0: \mu_1 \geq \mu_2 \\ H_1: \mu_1 < \mu_2 \end{matrix}$$

También podemos referirnos en las hipótesis directamente a la población de diferencias:

$$\begin{matrix} H_0: \mu_d \geq 0 \\ H_1: \mu_d < 0 \end{matrix}$$

Elección del estadístico de contraste y su distribución muestral. El estadístico de contraste se distribuye según la t de Student con n-1 grados de libertad, y lo calculamos según la expresión:

$$T = \frac{\bar{D} - \mu_d}{\sqrt{\frac{\hat{S}_d^2}{n}}} = \frac{-4'5 - 0}{\sqrt{\frac{22'5}{10}}} = -3 \tag{4.3}$$

Para averiguar el nivel crítico de forma aproximada, buscamos en la tabla t de Student del apéndice entre que valores se encuentra el estadístico de contraste, en este caso para 9 grados de libertad observamos: $-3'25 < -3 < -2'821$, por lo que el nivel crítico p se encontrará entre los valores: $0'005 < p < 0'01$ (con un programa informático se puede calcular el nivel crítico de una forma más precisa, en este caso $p = 0{,}0075$).

Establecer la regla de decisión en función del nivel de confianza. Los grados de libertad del ejemplo son: $n - 1 = 9$. Acudiendo a la tabla t de Student, el valor crítico que delimita cuando mantenemos o rechazamos la hipótesis nula, a un nivel de confianza del 95%, es -1'833. En la Figura 4.2 representamos los datos del ejemplo.

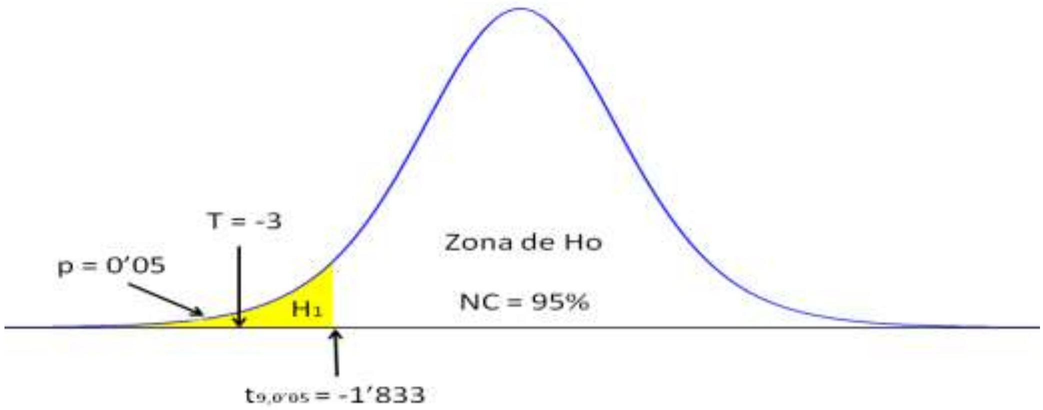


Figura 4.2. *Distribución muestral de diferencias según especifica la hipótesis nula.*

Conclusión. A un nivel de confianza del 95%, rechazamos la hipótesis nula puesto que el estadístico de contraste es más extremo que el valor crítico (máxima diferencia que cabe esperar por simple azar), por lo que concluimos que la media en asertividad de los directivos es inferior antes que después del curso. En cualquier caso, los datos obtenidos superan con creces el nivel de confianza que de antemano había fijado el psicólogo tal y como queda de manifiesto en el valor del nivel crítico p.

Interpretar el resultado en función del contexto de la investigación. El curso realizado por el psicólogo ha obtenido los resultados esperados, demostrando su utilidad para fomentar la asertividad en los directivos de su departamento.

Intervalo de confianza. Podemos calcularlo mediante la Ecuación 4.4:

$$\overline{D} \pm t_{1-\frac{\alpha}{2},\ n-1}\sqrt{\frac{\hat{S}_d^2}{n}} \qquad \text{(Ecuación 4.4)}$$

Que aplicándola a nuestros datos:

$$-4{,}5 \pm 2'262\sqrt{\frac{22'5}{10}} \longrightarrow (-1'107;\ -7'893)$$

Con lo que interpretaríamos, a un nivel de confianza del 95% que la media en asertividad es menor antes del curso (ambos límites son negativos y no contienen el valor cero plateado en la hipótesis nula), observándose un aumento después del mismo, que oscila entre 1'107 y 7'893 puntos en la puntuación media del test.

4.4.- Contraste de hipótesis sobre dos proporciones en muestras relacionadas.

Emplearemos este contraste de hipótesis para comparar dos proporciones poblacionales cuyos valores se estiman a partir de los datos observados en la misma muestra de sujetos, siendo la variable dependiente dicotómica o dicotomizada. Como se ha indicado anteriormente, la variable dependiente presenta dos opciones mutuamente excluyentes como, por ejemplo: "Sí", "No"; "A favor" o "En contra", "Aprobado", "Suspenso", etc., donde se trata de comparar la proporción de éxitos en las dos condiciones que presenta la variable independiente en función de las hipótesis establecidas. También se indicó con anterioridad, que una de las ventajas de trabajar con datos dicotómicos consiste en que los supuestos no son tan estrictos como en el caso de las variables continuas. Tan sólo es preciso que los tamaños de las muestras sean lo suficientemente grandes para poder asumir la normalidad de la distribución muestral de las diferencias entre proporciones.

Generalmente se empleará este contraste de hipótesis cuando se miden los datos de la muestra en la variable dependiente en **dos momentos temporales distintos**, como por ejemplo "antes" y "después" del tratamiento o "tratamiento A" vs. "tratamiento B".

4.4.1.- Estadístico Z.

El razonamiento para el contraste de hipótesis sobre dos proporciones para muestras relacionadas es el mismo que el visto en el caso de dos muestras independientes, (Tema 3, epígrafe 3.6), y se basa en que la distribución muestral de la diferencia de proporciones es una distribución normal cuando se trabaja con muestras grandes, siendo el estadístico de contraste una puntuación típica con la siguiente expresión general:

$$Z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sigma_{p_1 - p_2}}$$

Para aplicar este estadístico con el fin contrastar la hipótesis nula : $H_0\colon \pi_1 = \pi_2$ frente a la alternativa $H_1\colon \pi_1 \neq \pi_2$, el primer paso consiste en organizar los datos en una tabla de doble entrada de 2x2, denominada Tabla de Contingencia, como la siguiente:

		Después o tratamiento B		
		Éxito	Fracaso	Total
Antes o Tratamiento A	Éxito	a	b	$a+b$
	Fracaso	c	d	$c+d$
	Total	$a+c$	$b+d$	N

Los respuestas de los N sujetos recogidos en esta tabla de contingencia se reparten en las cuatro celdillas con a, b, c y d observaciones.

En la muestra, la proporción de sujetos con respuesta de "éxito" antes de introducir la variable independiente es un estimador insesgado de la proporción poblacional y corresponde a:

$$P_{Antes} = \frac{a+b}{N}$$

De igual forma, la proporción de sujetos con respuesta de "éxito" después de introducir la variable independiente, es:

$$P_{Después} = \frac{a+c}{N}$$

La hipótesis nula a contrastar es que las dos proporciones poblacionales (estimadas a partir de los datos de la muestra) son iguales, es decir:

$$H_0: \pi_{Antes} = \pi_{Después} \qquad \text{o bien:} \qquad H_0: \pi_{Antes} - \pi_{Después} = 0$$
$$H_1: \pi_{Antes} \neq \pi_{Después} \qquad \qquad H_1: \pi_{Antes} - \pi_{Después} \neq 0$$

Si las dos proporciones poblacionales son iguales, entonces, su diferencia a partir de sus estimadores muestrales es:

$$P_{Antes} - P_{Después} = \frac{a+b}{N} - \frac{a+c}{N} = \frac{b-c}{N}$$

Es decir que si la H_0 es cierta entonces $b - c = 0$, es decir las frecuencias de las celdillas b y c son iguales.

Se puede demostrar que el error típico de la distribución muestral de la diferencia de dos proporciones, $p_1 - p_2$, para muestras relacionadas, es:

$$\sigma_{p_1 - p_2} = \frac{1}{N}\sqrt{b + c - \frac{(b-c)^2}{N}}$$

Como se ha indicado, si las dos proporciones son iguales, entonces $b - c = 0$, y la expresión anterior del error típico de la distribución muestral quedaría simplificada a:

$$\sigma_{p_1 - p_2} = \frac{1}{N}\sqrt{b+c}$$

Siendo el estadístico de contraste Z:

$$Z = \frac{\left(P_{Antes} - P_{Después}\right) - \left(\pi_{Antes} - \pi_{Después}\right)}{\sigma_{p_1 - p_2}} = \frac{\left(\frac{a+b}{N} - \frac{a+c}{N}\right) - 0}{\frac{1}{N}\sqrt{b+c}} = \frac{\frac{b-c}{N}}{\frac{1}{N}\sqrt{b+c}}$$

Finalmente:

$$Z = \frac{b-c}{\sqrt{b+c}} \qquad \text{(Ecuación 4.5)}$$

A partir de esta expresión, y como se ha indicado anteriormente, se puede observar que para la aplicación del **estadístico Z solo se tiene en cuenta** a los individuos que aportan información, que son **aquellos que han obtenido resultados distintos bajo una u otra condición**. Estos individuos son los que ocupan las casillas b y c, es decir, aquellos cuya puntuación es diferente en las dos condiciones en las que medimos la variable dependiente o respuesta.

Por tanto, de la muestra inicial de N sujetos $(a+b+c+d)$ solo utilizaremos los n sujetos de las casillas b y c $(n = b + c)$. Si el tamaño de esta muestra de n sujetos es suficientemente grande $(n > 25)$ podremos aplicar el estadístico Z o el estadístico de McNemar que veremos más adelante.

Ejemplo 4.3. Un empresario, antes de introducir en el mercado un determinado producto "X", toma una muestra aleatoria de 500 sujetos de la población a la que quiere dirigirse y les pregunta si comprarían o no dicho producto. A continuación les muestra las posibles ventajas que aporta el producto "X" y les vuelve a preguntar si lo comprarían. Antes de la demostración de las ventajas del producto "X", 400 personas declaran que no lo comprarían, mientras que después de la demostración son 380 personas las que no lo comprarían. Por otro lado 60 personas estarían dispuestas a comprar el producto "X" tanto antes como después de la demostración. ¿Podemos afirmar al nivel de confianza del 99% que, la demostración del producto "X" ha sido eficaz?

En primer lugar construimos la siguiente tabla con los datos proporcionados en el enunciado:

		¿Compraría "X" **después** de la demostración?		
		SI	NO	
¿Compraría "X" **antes**?	SI	$a = 60$	b	$a+b$
	NO	c	d	400
		$a+c$	380	500

A continuación completamos fácilmente el resto de la tabla.

Sujetos que comprarían "X" antes de la demostración: $a + b = 500 - 400 = 100$

Sujetos que comprarían "X" después de la demostración: $a + c = 500 - 380 = 120$

Una vez deducidas las frecuencias marginales:

$$c = 120 - 60 = 60; b = 100 - 60 = 40$$

Y finalmente:

$$d = 380 - 40 = 340$$

Tabla 4.5

Número de personas que compraría "X" antes y después de la demostración.

		¿Compraría "X" **después** de la demostración?		
		SI	NO	
¿Compraría "X" **antes**?	SI	$a = 60$	$\boldsymbol{b = 40}$	100
	NO	$\boldsymbol{c = 60}$	$d = 340$	400
		120	380	500

En este contraste **sólo nos interesan los sujetos que han cambiado de opinión** tras la presentación de "X" (casillas "b" y "c"). En general, aquellos sujetos cuya puntuación es diferente en los dos momentos en los que medimos la variable dependiente. Una vez seleccionados dichos sujetos, nos preguntamos si el número de cambios es el mismo en las dos direcciones (hipótesis nula) o si, por el contrario, la mayor parte de los sujetos que cambian de opinión lo hacen en una dirección determinada (hipótesis alternativa).

Supongamos que los datos obtenidos hubieran sido diferentes, de forma que hubieran cambiado de opinión 100 sujetos, 50 de los cuales no comprarían "X" inicialmente y sí lo harían después de la demostración. Por lo tanto, los otros 50 sujetos habrían cambiado de opinión, pero de forma opuesta. La tabla de contingencia quedaría en este caso:

		¿Compraría "X" **después** de la demostración?		
		SI	NO	
¿Compraría "X" **antes**?	SI	$a = 60$	$\boldsymbol{b = 50}$	**110**
	NO	$\boldsymbol{c = 50}$	$d = 340$	390
		110	390	500

Por lo que la demostración llevada a cabo por el empresario no habría sido eficaz, dado que el número de personas que estaría dispuesto a comprar "X" sería el mismo antes y después de la demostración.

Los datos obtenidos, Tabla 4.5, muestran que 60 sujetos han cambiado de opinión en la dirección que pretendía el empresario (casilla c) puesto que declararon que no comprarían "X" antes de la demostración y sí después. Por otro lado, tenemos 40 sujetos que han cambiado de opinión en la dirección contraria (casilla b). Se trata de comprobar si el primer grupo de sujetos supera estadísticamente al segundo. En total son 100 $(b + c)$ los sujetos que cambian de opinión.

Condiciones y supuestos. Contamos con 100 observaciones independientes y una variable dicotómica, donde definimos el éxito como: "No antes, Sí después" y el fracaso como "Sí antes, No después". En general tendremos:

- Variable dependiente dicotómica o dicotomizada
- Muestra con "*b+c"* observaciones independientes, donde *b+c > 25*.

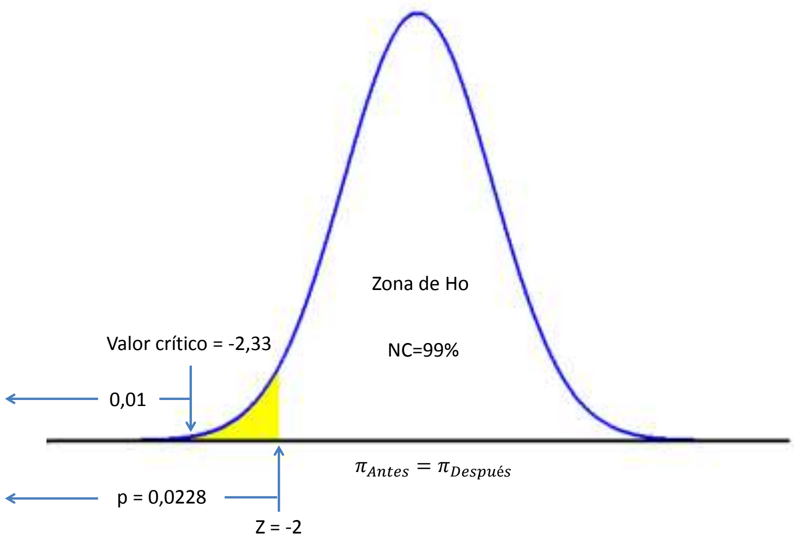
Hipótesis. La hipótesis nula especifica que la proporción de éxitos es igual o menor que la de fracasos, y la alternativa que la proporción de éxitos es superior a la de fracasos.

$$H_0: \pi_{Antes} \geq \pi_{Después}$$
$$H_1: \pi_{Antes} < \pi_{Después}$$

Estadístico de contraste y su distribución muestral. Aplicando el estadístico Z (Ecuación 4.5):

$$Z = \frac{b - c}{\sqrt{b + c}} = \frac{40 - 60}{\sqrt{40 + 60}} = -2$$

Regla de decisión: En un contraste unilateral izquierdo, con un nivel de confianza de 0,99, el valor crítico es: $z = -2{,}33$. Como el estadístico de contaste $Z = -2$, es superior al valor crítico, no tenemos evidencia suficiente para rechazar la hipótesis nula que se mantiene como provisionalmente verdadera.



Podemos consultar en las tablas de curva normal el nivel crítico, que es: $p = 0{,}0228$, que al ser mayor que el nivel de significación $(\alpha = 0{,}01)$, no nos permite rechazar H_0 con un nivel de confianza del 99%

Interpretar el resultado en el contexto de la investigación. Los resultados obtenidos no superan el nivel de confianza fijado de antemano, pero por otro lado, como se puede apreciar al examinar el nivel crítico, sí son significativos al nivel de confianza del 95%. En cualquier caso, de un total de 400 personas que no comprarían el producto "X" inicialmente, tan sólo 60 han cambiado de opinión en la dirección que pretendía el empresario, y 40 personas han cambiado de opinión en la dirección contraria. El empresario del ejemplo debería considerar si es rentable la campaña publicitaria que ha diseñado, puesto que en total, tan sólo se incrementa en 20 personas el número de posibles clientes que ganaría tras la demostración del producto "X".

OPCIONAL. Obsérvese que este contraste podríamos haberlo realizado con los conocimientos adquiridos en Tema 2. Si consideramos que tenemos una única muestra de 100 observaciones, definiendo el "éxito" como: "no antes" y "si después", se pueden plantear las siguientes hipótesis:

$$H_0: \pi_{Éxitos} \leq 0'5$$
$$H_1: \pi_{Éxitos} > 0'5$$

Siendo el estadístico de contraste igual a:

$$Z = \frac{p - \pi_{Éxitos}}{\sqrt{\frac{\pi_{Éxitos}(1-\pi_{Éxitos})}{n}}} = \frac{\frac{60}{100} - 0'5}{\sqrt{\frac{0'5 \cdot 0'5}{100}}} = 2 \xrightarrow{nivel\ crítico} p = 0'0228$$

4.4.2.- Estadístico de McNemar.

Otra prueba estadística muy utilizada en Psicología es el test de McNemar, cuyo estadístico de contraste es:

$$X^2 = \frac{(b-c)^2}{b+c} \tag{4.6}$$

Observamos que este estadístico es el cuadrado del anterior, por lo que, como podríamos deducir de los conocimientos adquiridos en la asignatura Introducción al Análisis de Datos (Tema 7), se distribuye según Chi cuadrado con un grado de libertad.

Los supuestos y las hipótesis para aplicar el test de McNemar son iguales a las tratadas para el estadístico Z, y con los datos del Ejemplo 4.3:

$$X^2 = \frac{(b-c)^2}{b+c} = \frac{(40-60)^2}{40+60} = 4$$

El nivel crítico, calculado con un programa informático es igual a $p = 0'0455$ (mediante las tablas deduciríamos que está comprendido entre: $0{,}025 < p < 0{,}05$, puesto que el estadístico de contraste se encuentra entre los valores 3,84 y 5,0239).

Podemos apreciar que el nivel crítico p es mayor para Chi cuadrado que para Z, lo que sucederá siempre que el contraste sea unilateral, siendo en este caso más difícil rechazar la hipótesis nula con el test de McNemar. Si el contraste fuera bilateral el nivel crítico sería el mismo para ambos estadísticos.

En el ejemplo que hemos desarrollado, el contraste es unilateral, lo que no plantearía ningún problema si utilizamos el estadístico Z, pero si empleamos el estadístico χ^2 sólo podremos plantear una **hipótesis estadística** bilateral, porque al elevar al cuadrado la diferencia $b - c$, este estadístico no nos informa de la dirección de las diferencias. No obstante, **el investigador puede interpretar los datos**, y por lo tanto la dirección de las diferencias, fijándose en cuál de las dos casillas "b" o "c" presenta una frecuencia mayor.

La diferencia entre los estadísticos Z y X^2, consiste en que el estadístico X^2 es insensible a la dirección del cambio producido. En nuestro ejemplo, el valor de ambos estadísticos es:

$$Z = \frac{b-c}{\sqrt{b+c}} = \frac{40-60}{\sqrt{40+60}} = -2$$

$$X^2 = \frac{(b-c)^2}{b+c} = \frac{(40-60)^2}{40+60} = +4$$

Observamos que al elevar $(b-c)$ al cuadrado X^2 es siempre positivo. Es decir X^2 puede detectar que existe un cambio pero no en qué dirección, por eso no podemos plantear una hipótesis estadística unilateral con dicho estadístico.

Por otro lado, con χ^2 no tiene sentido rechazar la hipótesis nula si obtenemos un valor muy pequeño, lo que nos indicaría que el número de sujetos que cambian de opinión en ambas direcciones son muy parecidos. De hecho, como podemos apreciar en la Ecuación 4.6, χ^2 valdrá cero cuando "$b = c$" que es precisamente lo que postula la hipótesis nula. Es decir, al aplicar la prueba de McNemar siempre dejamos todo el nivel de significación (α) en la parte derecha de la distribución. En la Figura 4.3 mostramos cuándo mantener o rechazar H_0. En temas posteriores veremos que sucede algo análogo en otras pruebas estadísticas, por ejemplo cuando aplicamos un Análisis de la Varianza.

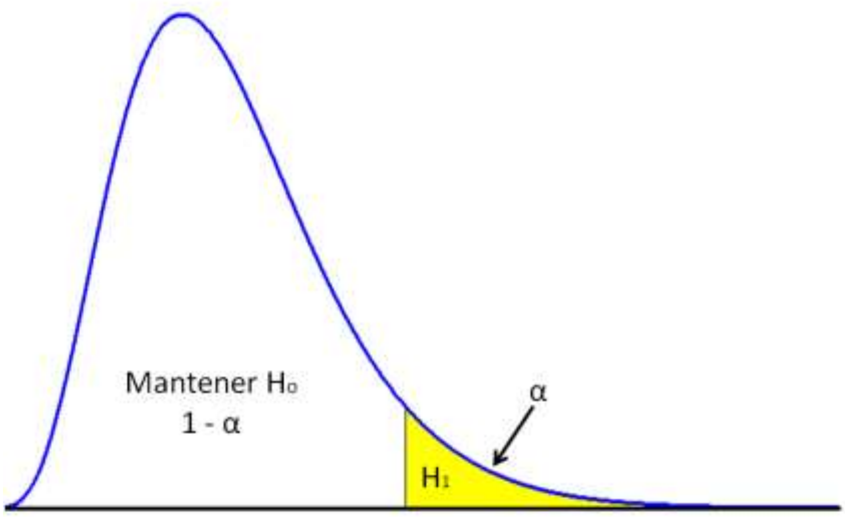


Figura 4.3. *Aplicación de la prueba de McNemar, en la que siempre queda α en la parte derecha.*

Establecer regla de decisión en función del nivel de confianza. Buscando en las tablas elos valor críticos al nivel de confianza del 99%, tenemos que:

$$4 < 6'6349 \longrightarrow Mantenemos\ H_0$$

4.5.- Resumen del tema.

El esquema seguido vuelve a ser el mismo que en los dos anteriores, por lo que el resumen del Tema 2 en el que se fijan los pasos de todo contraste de hipótesis sigue siendo válido. Además, si introducíamos pocos conceptos nuevos en el tema anterior, en este caso son todavía menos. Hemos comenzado por el contraste sobre dos medias relacionadas donde cambia la manera de obtener la distribución muestral de las diferencias, que por otro lado es muy similar a cuando trabajábamos con una única muestra. Finalmente estudiamos dos estadísticos para el contraste sobre dos proporciones en muestras relacionadas.

En las Tablas 4.6 y 4.7 que presentamos a continuación ofrecemos un resumen de las técnicas desarrolladas a lo largo de este tema.

Tabla 4.6. *Contrastes de hipótesis para dos medias en muestras relacionadas.*

<table>
<tr><td>SUPUESTOS</td><td colspan="2">- Observaciones independientes.
- Nivel de medida de intervalo o razón.
- Distribuciones normales en la población de diferencias o bien: $n \geq 30$.</td></tr>
<tr><td>Varianza poblacional</td><td>Conocida</td><td>Desconocida</td></tr>
<tr><td>HIPÓTESIS</td><td colspan="2">Igual en los dos casos. En función de la hipótesis científica plantearemos un contrate:
Bilateral: $H_0: \mu_1 - \mu_2 = 0$; $H_1: \mu_1 - \mu_2 \neq 0$
Unilateral izquierdo: $H_0: \mu_1 - \mu_2 \geq 0$; $H_1: \mu_1 - \mu_2 < 0$
Unilateral derecho: $H_0: \mu_1 - \mu_2 \leq 0$; $H_1: \mu_1 - \mu_2 > 0$</td></tr>
<tr><td>ESTADÍSTICO DE CONTRASTE</td><td>$Z = \frac{\bar{D} - \mu_d}{\sqrt{\frac{\sigma_d^2}{n}}}$</td><td>$T = \frac{\bar{D} - \mu_d}{\sqrt{\frac{\hat{S}_d^2}{n}}}$</td></tr>
<tr><td>DISTRIBUCIÓN MUESTRAL</td><td>Normal tipificada: $N(0,1)$</td><td>"t" de Student: $g.l. = n - 1$</td></tr>
<tr><td>REGLA DE DECISIÓN</td><td colspan="2">

Distribución normal (varianza conocida).

	Bilateral	Unilateral Derecho	Unilateral Izquierdo
Mantener H_0	$z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}$	$Z < z_{1-\alpha}$	$Z > z_{\alpha}$
Rechazar H_0	$Z < z_{\frac{\alpha}{2}}$ ó $Z > z_{1-\frac{\alpha}{2}}$	$Z > z_{1-\alpha}$	$Z < z_{\alpha}$

Distribución "t" de Student (varianza desconocida).

	Bilateral	Unilateral Derecho	Unilateral Izquierdo
Mantener H_0	$t_{g.l.,\frac{\alpha}{2}} < T < t_{g.l.,1-\frac{\alpha}{2}}$	$T < t_{g.l.,1-\alpha}$	$T > t_{g.l.,\alpha}$
Rechazar H_0	$T < t_{g.l.,\frac{\alpha}{2}}$ ó $T > t_{g.l.,1-\frac{\alpha}{2}}$	$T > t_{g.l.,1-\alpha}$	$T < t_{g.l.,\alpha}$

</td></tr>
</table>

<table>
<tr><td colspan="4">Tabla 4.7. Contrastes de hipótesis sobre dos proporciones en muestras relacionadas.</td></tr>
<tr><td>SUPUESTOS</td><td colspan="3">- Variable dependiente dicotómica o dicotomizada que medimos en la misma muestra en dos ocasiones.
- Muestra con “$b + c$” observaciones independientes, donde $b + c$ > 25 (b y c representan a los sujetos cuya puntuación es distinta en las dos ocasiones.</td></tr>
<tr><td>HIPÓTESIS</td><td>Bilateral
H_0: $\pi_{Antes} = \pi_{Después}$
H_1: $\pi_{Antes} \neq \pi_{Después}$</td><td>Unilateral izquierdo (sólo para estadístico Z)
H_0: $\pi_{Antes} \geq \pi_{Después}$
H_1: $\pi_{Antes} < \pi_{Después}$</td><td>Unilateral derecho (sólo para estadístico Z)
H_0: $\pi_{Antes} \leq \pi_{Después}$
H_1: $\pi_{Antes} > \pi_{Después}$</td></tr>
<tr><td>ESTADÍSTICO DE CONTRASTE</td><td colspan="2">$Z = \frac{b - c}{\sqrt{b + c}}$</td><td>$\chi^2 = \frac{(b - c)^2}{b + c}$</td></tr>
<tr><td>DISTRIBUCIÓN MUESTRAL</td><td colspan="2">$N(0,1)$</td><td>χ^2 con un grado de libertad</td></tr>
<tr><td>REGLA DE DECISIÓN</td><td colspan="2">Igual que en casos anteriores</td><td>$X^2 < \chi^2_{1,1-\alpha} \longrightarrow$ Mantener H_0
$X^2 > \chi^2_{1,1-\alpha} \longrightarrow$ Rechazar H_0</td></tr>
</table>

4.6.- Ejercicios de autocomprobación.

1. En cuál de los siguientes diseños empleamos muestras relacionadas: a) cuando los mismos sujetos son sometidos a dos condiciones experimentales; b) cuando disponemos de parejas de hermanos gemelos que son asignados aleatoriamente a grupos distintos; c) las dos opciones anteriores son verdaderas.
2. La varianza de la distribución muestral es mayor: a) en un diseño de dos muestras independientes; b) en un diseño de dos muestras relacionadas; c) son iguales en ambos casos.
3. Si trabajamos con dos grupos siendo los tamaños de las muestras: $n_1 = 30$ y $n_2 = 35$, sabemos que: a) se trata de un contraste paramétrico; b) las muestras están relacionadas; c) las muestras son independientes.
4. En un contraste sobre dos proporciones en muestras relacionadas realizamos el contraste: a) con los sujetos que obtienen puntuaciones diferentes en dos ocasiones; b) con los sujetos que obtienen la misma puntuación en dos ocasiones; c) utilizamos todas las puntuaciones de los sujetos en las dos ocasiones.
5. En un contraste sobre dos proporciones en muestras relacionadas sólo podemos plantear una hipótesis estadística tanto unilateral como bilateral: a) con el estadístico Z; b) mediante la prueba de McNemar; c) ninguna de las anteriores es correcta.

Un psicólogo escolar considera que una determinada técnica de relajación puede disminuir la ansiedad de los estudiantes para realizar los exámenes. Para comprobar esta hipótesis toma una muestra aleatoria de 31 estudiantes a los que mide su ansiedad mediante un test antes de un examen de matemáticas (Condición 1), posteriormente les enseña técnicas de relajación y vuelve a aplicar el test antes del siguiente examen de matemáticas (Condición 2) a los mismos alumnos. La media en ansiedad antes y después de aprender técnicas de relajación fueron, respectivamente: 12’18 y 10. Sabiendo que a mayor puntuación mayor ansiedad, que la cuasivarianza de las diferencias es igual a 25 y con un nivel de confianza del 95% conteste a las siguientes cuestiones:

6. La hipótesis alternativa es: a) $\mu_1 - \mu_2 \neq 0$; b) $\mu_1 - \mu_2 < 0$; c) $\mu_1 - \mu_2 > 0$.
7. El valor del estadístico de contrate es: a) 1’697; b) 2’428; c) -1’697.
8. El valor crítico es: a) 1’697; b) 2’428; c) -1’697.
9. Suponiendo que la hipótesis nula es cierta, la probabilidad de encontrar unos resultados como los obtenidos en la muestra, o más extremos, es, aproximadamente: a) 0’25; b) 0’01; c) 0’005.
10. La conclusión del estudio es: a) mantener la hipótesis nula porque el estadístico de contraste es menor que el nivel crítico; b) rechazar la hipótesis nula porque el estadístico de contraste es mayor que el valor crítico; c) mantener la hipótesis nula porque $p < \alpha$.

Un Psicólogo ha construido una prueba de Eficiencia Lectora para la detección de estudiantes con dificultades en la identificación de palabras y quiere realizar un estudio comparativo con otra prueba estándar para comprobar si con las dos pruebas se obtienen resultados similares. Aplica las dos pruebas “A” y “B” a una muestra de 200 estudiantes con edades comprendidas entre los 3 y 12 años de los cuales 80 son detectados por la prueba A y 90 por la prueba B, y de estos 65 sujetos también son detectados con la prueba A.

11. La hipótesis nula, es: a) $H_0: \pi_{test\ A} \leq \mu_{test\ B}$; b) $H_0: \pi_{test\ A} = \mu_{test\ B}$; c) $H_0: \pi_{test\ A} > \mu_{test\ B}$
12. El valor del estadístico de contraste de McNemar vale: a) 2,5; b) 1,58; c) 3.
13. El nivel crítico p asociado al estadístico de contraste Z, es: a) 0,0571; b) 0,9429; c) 0,1141.

Soluciones

1. "c".

2. "a".

3. Si $\mathrm{n}_1 \neq \mathrm{n}_2$ la respuesta correcta siempre es "C". Con muestras relacionadas cada sujeto de una muestra tiene su pareja (o él mismo) en la otra muestra, por lo que trabajando con muestras relacionadas siempre se cumplirá que: $\mathrm{n}_1 = \mathrm{n}_2$

4. "a".

5. "a"

6. El psicólogo espera que la media en ansiedad sea menor tras enseñar a técnicas de relajación a los alumnos, por lo que la respuesta correcta es "c".

7. La respuesta correcta es "b".

$$T = \frac{\overline{D} - \mu_d}{\sqrt{\frac{\hat{S}_d^2}{n}}} = \frac{(12'18 - 10) - 0}{\sqrt{\frac{25}{31}}} = 2'428$$

8. Buscando en la tabla t de Student con 30 grados de libertad y un nivel de confianza del 95% comprobamos que la respuesta correcta es "a".

9. Consultado las tablas observamos que entre los valores que contiene para 30 g.l. el valor que más se aproxima al estadístico de contraste es 2'457, siendo la probabilidad de encontrar valores más extremos que dicho valor igual a 0'01, luego la respuesta es"b".

10. "b".

11. El investigador solo quiere comparar las dos pruebas pero sin una idea previa de que una sea mejor que la otra, por lo que plantea un contraste bilateral siendo la hipótesis nula la que figura en la alternativa b).

12. Con los datos del enunciado, marcados en negrita, completamos la siguiente tabla de contingencia para realizar los cálculos:

		Test B		
		Si	N0	
Test A	si	**65**	15	**80**
	no	25	95	120
		90	110	**200**

El estadístico de McNemar, es:

$$X^2 = \frac{(b-c)^2}{b+c} = \frac{(15-25)^2}{15+25} = +2{,}5$$

Por lo que la respuesta correcta es la a).

13. Primero calculamos el estadístico Z:

$$Z = \frac{b-c}{\sqrt{b+c}} = \frac{15-25}{\sqrt{15+25}} = -1{,}58$$

Obsérvese que se obtendría el mismo resultado, con la siguiente expresión general del estadístico Z, aunque de forma algo más laboriosa:

$$Z = \frac{(P_{test\ A} - P_{test\ B}) - (\pi_{test\ A} - \pi_{test\ B})}{\sigma_{p_1-p_2}} = \frac{\left(\frac{a+b}{N} - \frac{a+c}{N}\right) - 0}{\frac{1}{N}\sqrt{b+c}} =$$

$$= \frac{\left(\frac{65+15}{200} - \frac{65+25}{200}\right) - 0}{\frac{1}{200}\sqrt{15+25}} = \frac{0{,}40 - 0{,}45}{0{,}0316} = -1{,}58$$

Puede comprobarse, además, que $Z^2 = -1{,}582^2 = 2{,}5$ es igual al valor obtenido al aplicar la prueba de McNemar.

Finalmente, para $Z = -1{,}58$ el nivel crítico, en un contraste bilateral es $p = 2 \cdot 0{,}0571 = 0{,}1142$ que corresponde a la alternativa c).

La interpretación de estos resultados indican que las diferencias observadas entre las dos pruebas no son significativas y proporcionan los mismos resultados para la detección de estudiantes con dificultades en la identificación de palabras.

Tema 5. DISEÑOS DE MÁS DE DOS GRUPOS INDEPENDIENTES.

5.1.- Introducción

En los temas precedentes hemos estudiado contrastes de hipótesis sobre la media que se corresponden con diseños realizados con uno y dos grupos (tanto si se trata de dos grupos independientes como de dos grupos relacionados). Sin embargo, no siempre nos encontramos con diseños tan simples. Con mucha frecuencia, lo que tenemos que comparar son más de dos grupos pues ello nos dará una idea más completa de la relación que se pueda establecer entre nuestras variables. En este caso se utiliza el Análisis de Varianza.

El Análisis de Varianza (***An****alysis* ***of*** ***Va****riance* o ANOVA) es una técnica de análisis estadístico que se utiliza para comparar las **medias** de más de dos grupos, aunque su nombre hace referencia al estudio de la variabilidad observada en los datos como veremos a lo largo del tema.

Antes de introducirnos en esta técnica de análisis veamos un ejemplo que nos ayude a entender la lógica y los mecanismos de este procedimiento. En el Tema 3 (ejemplo 3.3, apartado 3.4.4) veíamos un ejemplo en el que queríamos probar el efecto del consumo de un determinado fármaco en la reducción de la ansiedad y teníamos un grupo experimental y un grupo control. A la hora de analizar los resultados, realizamos un contraste de medias y, en este caso, el estadístico de contraste es: $T = -2{,}46$.

Dado que la probabilidad de encontrar un valor igual o menor al que hemos obtenido es inferior a $0{,}05$ decidimos rechazar la hipótesis nula de igualdad de medias poblacionales, concluyendo que el fármaco en cuestión sí influye en la ansiedad que presentan los sujetos.

Ahora bien, en lugar de limitarnos a este experimento tan sencillo podríamos ir un poco más allá y estar interesados, no sólo en la existencia de diferencias entre los grupos como consecuencia de la administración o no del fármaco, sino que además nos puede interesar averiguar si existe un efecto distinto en función, por ejemplo, de dos cantidades distintas del fármaco suministrado.

De este modo, podemos complicar un poco más el diseño y, en lugar de dos grupos, podemos construir aleatoriamente tres grupos: el primero, sin tomar el fármaco; el segundo, con una dosis de 0,05 miligramos; y el tercero, con una dosis de $0{,}10$ miligramos. Supongamos que en esta ocasión los datos que hemos obtenido son los del ejemplo 5.1 que aparece a continuación.

Ejemplo 5.1. En este caso vamos a trabajar con tres grupos porque, además de estar interesados en si el fármaco influye o no en la ansiedad, queremos averiguar si influye de forma distinta en función de que se suministre 0,05 o 0,10 mg. Supongamos que hemos obtenido los siguientes resultados:

	Sin fármaco	Con 0,05 mg de fármaco	Con 0,10 mg de fármaco
	30	10	20
	50	20	40
	30	30	10
	60	20	10
	20	5	10
n_j	5	5	5
$\sum$	190	85	90
$\bar{Y}_j$	38	17	18
$\hat{S}_j^2$	270	95	170

Nota: *recuérdese que el acento circunflejo sobre el símbolo de la varianza muestral representa la varianza insesgada muestral o cuasivarianza.*

Disponemos de una herramienta conocida, la prueba *t* de Student, que podríamos utilizar, pero ello conllevaría una serie de inconvenientes que debemos tener en cuenta. En primer lugar, y aunque no es el inconveniente más importante, el número de comparaciones dos a dos que hay que realizar aumenta con el número de grupos o niveles (con tres grupos son tres comparaciones, con cuatro son seis, con cinco son diez, etc.). La regla general para el cálculo del número de comparaciones a realizar son las **combinaciones**[1] del número de grupos tomados de dos en dos (siendo *a* el número de muestras o grupos, las comparaciones posibles son: $\frac{a(a-1)}{2}$. En nuestro caso, como tenemos 3 muestras, el número de comparaciones serían tres, es decir: $\frac{3(3-1)}{2} = 3$.

En segundo lugar, con el aumento del número de comparaciones también aumenta la probabilidad de cometer el error de tipo I (α), esto es, se incrementa la probabilidad de rechazar la H_0 siendo cierta. Podemos considerar este segundo inconveniente como el más grave debido a que toda la Estadística Inferencial clásica[2] se basa en el control del error tipo I. Si nosotros estamos interesados en trabajar con un nivel de significación de $\alpha = 0{,}01$, sabemos que ésta es la probabilidad de rechazar la H_0 cierta en cada uno de los contrastes que realicemos y no deseamos que esa probabilidad se vea aumentada. Pero esto es lo que sucederá si realizamos comparaciones dos a dos.

[1] Dado un conjunto A de m elementos, se denomina combinaciones de orden n (n<m) a las distintas agrupaciones de n elementos elegidos del conjunto A que pueden realizarse de modo que dos agrupaciones son distintas si se distinguen únicamente por uno o más de los elementos que la componen. Es decir, dos subconjuntos de n elementos se consideran la misma combinación si sólo difieren en el orden pero no en la identidad de los elementos.

[2] Existen otras alternativas a la inferencia estadística "clásica", la más importante de las cuales es la Inferencia Bayesiana. Este curso no puede entrar en su explicación.

Así, por ejemplo, si realizamos tres contrastes, y asumiendo que los contrastes son independientes, la probabilidad de cometer al menos un error de tipo I en los tres vendría dado por:

$$1-(1-\alpha)^{a}=1-(1-0{,}01)^{3}=0{,}0297\approx 0{,}03$$

Podemos observar que α ha aumentado de 0,01 a 0,03 aproximadamente. Si en lugar de *3* tuviésemos *4* grupos ($a = 4$), la probabilidad de cometer al menos un error de tipo I sería (trabajando con un $\alpha = 0{,}01$) de 0,0394. Si tuviésemos 5 grupos o muestras ($a = 5$) la probabilidad sería de 0,049. Como vemos, a medida que aumenta el número de grupos a comparar nos estamos alejando del nivel de riesgo inicial con el que queríamos trabajar.

Para superar estos inconvenientes disponemos de una herramienta que realiza la comparación, a la vez, de todos los grupos: el Análisis de Varianza (también conocido como ANOVA y, en algunas publicaciones, como ANVAR), al que le dedicaremos este tema y los dos siguientes.

5.2.- Objetivos del tema

En este tema vamos a ocuparnos de una de las técnicas de análisis estadístico más utilizadas en Psicología: el Análisis de Varianza (ANOVA o ANVAR).

Se trata de introducir al alumno en los **fundamentos** y la lógica del mismo así como en la utilización de su **terminología**. Para ello, comenzaremos con el **modelo de un solo factor** (en temas posteriores se verá el modelo de dos factores), con el que resulta más fácil entender, intuitivamente, la lógica de este tipo de contraste estadístico. Como veremos, en el ANOVA hablamos de factor para referirnos a una variable independiente.

Veremos, también, las **condiciones** que deben cumplir los datos para que se pueda utilizar el Análisis de Varianza e incluiremos, en este tema, el concepto de **comparaciones múltiples.**

5.3.- Conceptos básicos del análisis de varianza

Antes de entrar en el fundamento del ANOVA, veamos algunas ideas y **terminología** que deben ser manejadas con precisión.

Siguiendo con nuestro ejemplo, si nosotros comparamos los grupos que han tomado las distintas dosis del fármaco (0; 0,05 mg y 0,10 mg), la **variabilidad** que aparezca entre ellos puede deberse tanto a los efectos del fármaco como a la influencia de otros factores que no hayamos podido controlar, dado que por muy perfecto que sea el diseño, las variables que existen entre los sujetos y su entorno son tantas que es imposible controlarlas todas. Aunque, por ejemplo, aislásemos e incomunicásemos a los sujetos, esto podría influir de forma distinta en cada uno de ellos.

Así pues, a la hora de realizar el estudio hay que ser conscientes de ello, por lo que podemos considerar la variabilidad que se observa entre las puntuaciones, después de haber introducido la variable independiente, como formada por dos partes o componentes:

a) La que se debe al efecto del **factor estudiado**, en nuestro ejemplo, a las distintas dosis del fármaco.
b) La que se debe a los factores extraños y no controlados, que es lo que recibe el nombre de **error experimental**, dado que introduce una fuente de error en nuestro diseño.

La tarea estará en discernir qué variabilidad corresponde a cada parte y éste es el cometido del Análisis de Varianza.

En la terminología del ANOVA, las variables independientes que se estudian reciben el nombre de **factores** y las categorías en que se dividen, el de **niveles**. Así, en nuestro ejemplo, tenemos una sola variable independiente (un solo factor, el fármaco) con tres niveles (0; 0,05 y 0,10 mg, es decir, las distintas dosis suministradas).

Hemos hablado, siguiendo con el ejemplo, de distintas categorías de la variable independiente o niveles: el grupo control no ha tomado nada (0 mg), el segundo ha tomado 0,05 mg y el tercero 0,10 mg.. Ante estos mismos niveles del factor o variable independiente, nosotros podríamos plantearnos de dos formas la investigación:

a) Nuestro interés puede ser probar que a mayor cantidad del fármaco la ansiedad disminuye.
b) O bien nuestro interés puede ser probar que con 0,10 mg los sujetos presentan menos ansiedad que con 0,05 mg y, con esta dosis, menos que sin nada.

Estos dos planteamientos podrían parecer iguales pero no lo son y suponen dos formas distintas de probar la hipótesis nula en el ANOVA. Veamos por qué

Para el primer planteamiento, no nos ceñimos a las dosis concretas del fármaco en cuestión y los niveles actúan como una muestra de todos los posibles niveles que se pudiesen establecer (las posibles dosis que pudiésemos suministrar). A la hora de plantear las conclusiones nos daría lo mismo utilizar estas dosis en el experimento que otras siempre que se pudiese establecer un sentido ascendente o descendente. A este tipo de diseños en los que los niveles son una muestra de todos los posibles niveles del factor y nuestras conclusiones van a ser para todos ellos, se conoce como diseño de **efectos aleatorios** o **modelo aleatorio**.

Para el segundo planteamiento, sin embargo, nuestras conclusiones estarán restringidas a los niveles establecidos en el diseño. Sabemos que pueden existir más niveles pero sólo nos interesan estos. Este tipo de diseños se conocen como de **efectos fijos** o **modelos fijos**.

La interpretación de uno y otro, como vemos, está en función de la hipótesis que se plantea. Más adelante veremos también las implicaciones que suponen a la hora del cálculo.

En cuanto al **número de sujetos**, si los grupos o muestras son de distinto número de elementos, el modelo recibe el nombre de **modelo no equilibrado** y, en el caso de que todas las muestras sean iguales, tendremos un **modelo equilibrado**.

5.4.- Fundamentos del análisis de varianza

Como se ha visto en la introducción, en ocasiones es necesario comparar más de dos muestras, aunque esta comparación podría realizarse de dos en dos por los procedimientos vistos en los temas anteriores (Z o T), el procedimiento resultaría tedioso y aumentaría la probabilidad del error de tipo I. La técnica a utilizar, en estos casos, es el Análisis de Varianza cuya finalidad es contrastar la diferencia de medias entre varias muestras (dos o más).

La hipótesis nula que se establece consiste en afirmar que no existe diferencia entre las medias de los distintos grupos o muestras en la variable dependiente, aunque el modo de contrastar dicha hipótesis va a ser a través de la variabilidad observada en las puntuaciones, de ahí el nombre de esta técnica.

Cuando se diseña un experimento, se intenta minimizar la influencia de las variables extrañas, tanto las conocidas como las desconocidas (todas aquellas variables que están influyendo en la variable dependiente pero que no son la que se está manipulando). Si no se realiza un control de éstas, se producen sesgos sistemáticos en los resultados que se confunden con los efectos que pudieran deberse a la variable independiente. Un modo de controlar el efecto de variables extrañas desconocidas es la aleatorización de los sujetos, tanto en su obtención (muestreo), como en su asignación a las condiciones experimentales o niveles. Sin embargo, esta aleatorización no asegura que las diferencias observadas entre los niveles no sean fruto del efecto conjunto de la variable independiente y de factores azarosos introducidos en el propio proceso.

Nunca hay certeza absoluta para atribuir en exclusiva las diferencias entre los niveles o tratamientos sólo al efecto de éstos, y siempre cabe que se atribuya una porción al azar. Algunas de las variables que pueden perturbar y contribuir al error experimental se aleatorizan, pero hay otras muy importantes que no se pueden tratar de la misma manera: diferencias individuales, ambiente en que se desarrolla el experimento, la mejor o peor dicción que tenga el experimentador cuando explica al sujeto lo que hay que hacer, y por tanto que no todos los sujetos lo entiendan de la misma manera, el error de medida, un juicio equivocado sobre la conducta objeto de estudio, error al transcribir los datos del experimento, etc. Suponemos que todos estos factores influyen de manera no sistemática en el error experimental y que son independientes de los efectos de los tratamientos que, repetimos, son los únicos que nos interesan.

Como se ha señalado, una fuente del error (en el sentido de la existencia de variables extrañas o diferentes a la variable independiente que estoy manipulando y que están afectando a la variable dependiente) son las características (de personalidad, biológicas, etc.) de los propios sujetos. Si tomamos las puntuaciones de todos los sujetos que pertenecen en un mismo grupo o nivel experimental, no es lógico esperar que todas las puntuaciones sean iguales, pues ellos mismos se ven afectados por todas las fuentes de variabilidad no controlada. Por lo tanto, la propia variabilidad de los sujetos sometidos **al mismo tratamiento** nos proporciona una buena estimación del error experimental. Si extendemos el alcance de este argumento a los sujetos de los demás tratamientos, vemos que se refuerza esta idea del error experimental. Como en principio no podemos suponer que el error sea mayor en un tratamiento que en otro, se puede obtener una mejor y más estable estimación del error (variabilidad entre las puntuaciones) promediando las estimaciones del error que se obtienen para cada tratamiento.

Ahora cambiemos el enfoque y pensemos en los tratamientos por separado, en los cuales las asignaciones de sujetos se han hecho de manera aleatoria. Si la hipótesis nula fuera verdadera (que todas las medias de los tratamientos son iguales), tampoco se nos ocurriría pensar que las medias de los grupos que usamos en el diseño (que son, conviene recordarlo, muestras aleatorias de poblaciones) son necesariamente iguales ya que estamos trabajando con muestras. Al no ser iguales, y ser verdadera la hipótesis nula de igualdad, de nuevo hay que recurrir a explicar estas diferencias entre las medias de los grupos como un efecto del error experimental. Es decir, todas las fuentes de variabilidad no sistemática que provocan las diferencias entre los sujetos, también influyen en las diferencias que se producen entre los medias muestrales de los grupos.

El Análisis de Varianza proporciona al investigador "argumentos estadísticos" para decidir si las diferencias que se observan entre los tratamientos son enteramente, o solo parcialmente, debidas al azar.

Decíamos en el apartado anterior que la variabilidad observada entre las puntuaciones en la variable dependiente debíamos considerarla compuesta por dos componentes: la variabilidad atribuible a los distintos tratamientos experimentales y la variabilidad debida al error experimental.

Pues bien, en el estudio de estas dos variabilidades se fundamenta el Análisis de Varianza. En el modelo más simple, el de un sólo factor, se trataría de realizar dos estimaciones, independientes, de la varianza general o común (desconocida) por medio de:

- La varianza atribuible a los distintos niveles del factor en estudio, que es lo que conocemos como **varianza intergrupos.**
- La varianza atribuible al error, que es lo que conocemos como varianza del error o, también, **varianza intragrupos**.

De la comparación entre ambas varianzas obtenemos la aceptación o rechazo de la hipótesis nula.

Tomemos de nuevo el ejemplo anterior. Hemos distribuido, aleatoriamente a 15 sujetos en tres grupos, de cinco sujetos cada uno, y les hemos administrado tres dosis distintas de un determinado fármaco para reducir la ansiedad (se supone que inicialmente los quince sujetos puntuaban alto en esta variable y que el efecto del fármaco, si existe, será reducir estas puntuaciones). Dado que los sujetos han sido distribuidos aleatoriamente entre los distintos niveles suponemos que, inicialmente, son semejantes en cuanto a la variable estudiada en todos los niveles. Por lo tanto, si después de suministrarles el tratamiento presentan diferencias hay que pensar que el tratamiento ha influido. Pero, ¿cómo descartar el efecto de factores extraños si, como hemos dicho, la variabilidad que se presenta entre las puntuaciones contiene un componente de error?

El razonamiento del Análisis de Varianza es el siguiente: lógicamente, dentro de un mismo nivel y como el efecto del factor es único (todos los sujetos han tomado la misma dosis del fármaco, en nuestro ejemplo) las puntuaciones deberían ser semejantes, es decir, presentar una variabilidad prácticamente nula. En tanto en cuanto esta variabilidad, dentro de los grupos, sea grande es que están influyendo factores extraños y no controlados en el diseño. Así pues, la varianza dentro de cada grupo o nivel (varianza intragrupos) podemos considerarla como la varianza de error (la que nos mide el error experimental). Por otro lado, si los distintos niveles del factor objeto de estudio producen distintos efectos sobre la variable dependiente, la variabilidad entre los grupos o niveles debería aumentar (la varianza intergrupos).

El punto esencial es que si la varianza intergrupos (debida al factor manipulado) es significativamente mayor que la varianza intragrupos (debida a factores de error), se admite que hay diferencias entre los grupos, es decir, que los distintos niveles del factor objeto de estudio han influido de forma distinta sobre la variable dependiente. En nuestro ejemplo, que las distintas dosis del fármaco influyen de forma distinta sobre la ansiedad. La herramienta que tenemos para probar si ambas varianzas (la intergrupos y la intragrupos) difieren significativamente es la distribución F en la que se basa el Análisis de Varianza.

5.5.- Análisis de varianza de un factor

5.5.1.- Modelo de Efectos fijos.

En este modelo nos interesa estudiar la influencia de un solo factor (al que llamaremos genéricamente **factor A** o simplemente **A**), que tiene distintos niveles, como en el ejemplo que estamos comentando: tenemos una variable dependiente, la puntuación en ansiedad en algún test válido (mide realmente la ansiedad) y fiable (está libre de error en la medida de lo posible), sobre la que vamos a hacer actuar (comprobar el efecto que produce) un factor o variable independiente, que puede presentarse bajo un cierto número de niveles (en este caso las dosis), y nos interesa estudiar, concretamente, el efecto de esos niveles y no otros.

Consideremos que tenemos *a* niveles (los niveles del factor en cuestión varían desde *i = 1, …, a*) y *n* elementos, o sujetos, medidos en cada nivel (los elementos varían desde *j = 1, …, n* dentro de cada grupo asumiendo un diseño equilibrado). Pues bien, la puntuación Y_{ij} que es la puntuación del sujeto que ocupa el lugar *j* dentro del nivel *i*, está formada por:

- Un componente al que llamamos α_i y que será común a todos los elementos sometidos a ese nivel del factor.
- El componente del error experimental, formado, según hemos visto, por todos los factores no controlados en el experimento, y que llamamos ε_{ij} (que puede afectar a todos los sujetos y en todos los niveles, de ahí los dos subíndices).
- Una constante, común para todos los valores de la variable dependiente, que llamaremos μ dado que es la media de la población.

La acción de estos tres componentes se supone lineal, de forma que el modelo sería:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Donde:

ε_{ij} (el componente de error) es una variable aleatoria distribuida según la curva normal con media 0 y desviación típica σ, es decir, N(0, σ).

$\sum_{i=1}^{a} n_i\alpha_i = 0$ bajo la H_0 de que no hay efecto.

Esta es, pues, la estructura de una observación cuando estamos analizando un solo factor.

Antes de continuar con la exposición, veamos la **terminología** estadística que vamos a utilizar:

Tabla 5.1. *Sistema de notación para el ANOVA de un factor A.*

	Nivel a_1	Nivel a_2	Nivel a_i	Nivel a_a	Total
	$Y_{1,1}$	$Y_{2,1}$	$Y_{i,1}$	$Y_{a,1}$	
	$Y_{1,2}$	$Y_{2,2}$	$Y_{i,2}$	$Y_{a,2}$	
	$Y_{1,3}$	$Y_{2,3}$	$Y_{i,3}$	$Y_{a,3}$	
	$Y_{1,4}$	$Y_{2,4}$	$Y_{i,4}$	$Y_{a,4}$	
	$Y_{1,j}$	$Y_{2,j}$	$Y_{i,j}$	$Y_{a,j}$	
	Y_{1,n_1}	Y_{2,n_2}	Y_{i,n_j}	Y_{a,n_a}	
Suma	A_1	A_2	A_i	A_a	$T = \sum Ai$
Media	$\bar{Y}_{A_1}$	$\bar{Y}_{A_2}$	$\bar{Y}_{A_i}$	$\bar{Y}_{A_a}$	$\bar{Y}_T = {}^{T}/_{N}$
Nº obs.	n_1	n_2	n_i	n_a	$N = \sum n_i$

En definitiva:

- $\boldsymbol{Y_{ij}}$ es la puntuación del sujeto *j* en el nivel *i*.
- $\boldsymbol{\alpha_i}$ es un elemento común a todos los sujetos experimentales sometidos a un mismo nivel del factor, y es el efecto que queremos medir.
- $\boldsymbol{n_i}$ es el número de sujetos u observaciones en el nivel *i*.
- **N** es el número total de sujetos.
- $\boldsymbol{\bar{Y}_{A_i}}$ es la media aritmética de cada nivel (*i*).
- $\boldsymbol{\bar{Y}_T}$ es la media aritmética total, o media de todas las puntuaciones. Es decir, el resultado de sumar todas las puntuaciones de cada nivel y en todos los niveles, y dividirlo por el número total de sujetos.

Se demuestra que la media total ($\bar{Y}_T$) es una estimación insesgada de la media de la población μ:

$$\bar{Y}_T \approx \mu$$

Y que la media de cada nivel es una estimación sesgada de la media de la población, con un sesgo aditivo consistente en el efecto del factor:

$$\bar{Y}_{A_i} \approx \mu + \alpha_i$$

Por lo tanto:

$$\bar{Y}_{A_i} - \bar{Y}_T \approx \mu + \alpha_i - \mu$$

Luego:

$$\bar{Y}_{A_i} - \bar{Y}_T \approx \alpha_i$$

¿Cuándo diremos que los tratamientos, o niveles del factor, influyen de forma distinta sobre la variable dependiente? Lo que nos estamos preguntando es si los α_i son iguales dado que, si recordamos lo dicho previamente, α_i es el elemento común a todos los sujetos experimentales sometidos a un mismo nivel del factor que estemos estudiando. Es decir, queremos saber si:

$$\alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_a$$

Pero, como hemos dicho que bajo la H_0 de que no hay efecto:

$$\sum_{i=1}^{a} n_i \alpha_i = 0$$

Lo fundamental es probar si alguno de los α_i es distinto de cero ya que en este caso se podrá rechazar la H_0.

Las **hipótesis estadísticas** que queremos probar en este tipo de problemas son:

$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$
$H_1: \mu_1 \neq \mu_2 \neq \cdots \neq \mu_a$ al menos para un par de μ_i

Respecto a los **supuestos** de este contraste de hipótesis hablaremos de ellos a continuación, aunque adelantamos que son:

- Independencia de las observaciones.
- Normalidad de las distribuciones.
- Homogeneidad de las varianzas (homocedasticidad).

Estadístico de contraste

El estadístico de contraste para poner a prueba la hipótesis nula de igualdad de medias entre todos los niveles del factor manipulado consiste en el cociente entre la varianza entre los grupos (estimada mediante la media cuadrática inter-grupos o MC_{inter}) y la varianza dentro de los grupos (estimada mediante la media cuadrática intra-grupos o MC_{intra}). Luego veremos cómo se calculan ambas medias cuadráticas. Lo más importante es que, como se estudió en Fundamentos de Investigación, cualquier cociente entre varianzas, bajo el supuesto de la veracidad de la H_0, se distribuirá según la distribución F:

$$F = \frac{MC_{inter}}{MC_{intra}}$$

Los grados de libertad son $(a - 1)$ para el numerador y $(N - a)$ para el denominador. Veamos cómo llegar a este estadístico de contraste.

Variabilidad del sistema

Partimos de lo que denominaremos suma de cuadrados total que, si recordamos la fórmula de la varianza, correspondería al numerador de la varianza total:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_T\right)^2$$

Sumando y restando la media de cada grupo, obtenemos:

$$\sum_{i=1}^{a}\sum_{j=1}^{n_j}(Y_{ij}-\bar{Y}_T)^2=\sum_{i=1}^{a}\sum_{j=1}^{n_j}(Y_{ij}-\bar{Y}_T+\bar{Y}_{A_i}-\bar{Y}_{A_i})^2$$

Reagrupando el segundo término de la igualdad:

$$\sum_{i=1}^{a}\sum_{j=1}^{n_j}(Y_{ij}-\bar{Y}_T+\bar{Y}_{A_i}-\bar{Y}_{A_i})^2=\sum_{i=1}^{a}\sum_{j=1}^{n_j}[(Y_{ij}-\bar{Y}_{A_i})+(\bar{Y}_{A_i}-\bar{Y}_T)]^2$$

Podemos demostrar que el doble producto es igual a cero, con lo que:

$$\sum_{i=1}^{a}\sum_{j=1}^{n_j}(Y_{ij}-\bar{Y}_T)^2=\sum_{i=1}^{a}n_i(\bar{Y}_{A_i}-\bar{Y}_T)^2+\sum_{i=1}^{a}\sum_{j=1}^{n_j}(Y_{ij}-\bar{Y}_{A_i})^2$$

En la igualdad a la que hemos llegado, podemos observar que la suma de cuadrados total se descompone en la suma de dos componentes independientes:

$$SC_{Total}=SC_{Inter}+SC_{Intra}$$

Donde:

SC_{Total} es la suma de cada puntuación menos la media total al cuadrado.

SC_{Inter} es la suma de la media de cada nivel menos la media total al cuadrado (ponderado por el número de sujetos de cada nivel), es decir, nos está midiendo la **variabilidad entre los niveles**, o variabilidad debida al efecto del factor (A) que estamos estudiando.

SC_{Intra} es la suma de cada puntuación menos la media de su nivel al cuadrado, es decir, nos está midiendo la **variabilidad dentro de cada nivel** o variabilidad debida al error experimental. Representa las desviaciones de las puntuaciones de cada sujeto respecto de la media de su grupo.

Si estas sumas de cuadrados (que se pueden contemplar como los numeradores de la fórmula de la varianza) las dividimos por sus respectivos grados de libertad, obtendremos la varianza **entre** los niveles y la varianza **dentro** de los niveles.

Recordemos que:

$$\text{Varianza} = \frac{\textit{suma de cuadrados de las desviaciones de la media}}{\textit{grados de libertad}}$$

Los grados de libertad (gl) asociados a las SC se refieren al número de observaciones independientes menos las estimaciones que haya habido que realizar. Los grados de libertad de SC_{Inter} son *(a-1)* pues tenemos *a* puntuaciones independientes (los grupos) a los que hay que restar *1* gl por la estimación que realizamos de la media poblacional a partir de la media total. Los grados de libertad de SC_{Intra} son *(N-a)*, pues tenemos *N* puntuaciones independientes a las que hay que restar las a estimaciones que realizamos de las medias de los grupos.

En definitiva, logramos descomponer la varianza total en dos componentes:

1- Varianza entre los grupos, que llamaremos media cuadrática intergrupos y que refleja la varianza atribuible a los distintos niveles del factor en estudio:

$$MC_{Inter} = \frac{SC_{Inter}}{a-1}$$

2- Varianza dentro de los grupos, que llamaremos media cuadrática intragrupos, y que refleja la variabilidad atribuible al error:

$$MC_{Intra} = \frac{SC_{Intra}}{N-a}$$

De la comparación de ambas varianzas obtenemos la aceptación o rechazo de la hipótesis nula. Por otro lado, como hemos visto en el Tema 3, podemos contrastar si dos varianzas son iguales o diferentes mediante el estadístico F, por lo que, como hemos adelantado:

$$F = \frac{MC_{inter}}{MC_{intra}}$$

Se demuestra que la esperanza matemática de la varianza dentro de los niveles (la media cuadrática intragrupos) es un estimador insesgado de la varianza poblacional, mientras que la varianza entre los niveles (media cuadrática intergrupos) es un estimador sesgado de la varianza poblacional, con un sesgo aditivo (un sumando) función del efecto del factor (al que hemos llamado α_i), es decir, cuanto más efecto ejerza la variable independiente sobre la dependiente, mayor será el sesgo en la estimación de MC_{inter}:

$$E(MC_{Inter}) = \sigma^2 + \frac{1}{(a-1)}\sum_{i=1}^{a} n_i\alpha_i^2$$

$$E(MC_{Intra}) = \sigma^2$$

Por lo tanto si dividimos la MC_{Inter} entre la MC_{Intra} el cociente, en teoría, deberá ser igual o mayor de uno.

Si MC_{Inter} es muy grande (más de lo que cabría esperar por efecto del azar en el muestreo) respecto a MC_{Intra} admito que algún $\alpha_i \neq 0$ pues, en teoría, si todos los $\alpha_i = 0$ las dos esperanzas que componen la razón entre MC_{Inter} y MC_{Intra} son iguales y por lo tanto el cociente sería igual a la unidad.

Tabla del Análisis de Varianza.

Los resultados de un Análisis de Varianza suelen presentarse en tablas del tipo que puede verse en la Tabla 5.2:

Tabla 5.2
Tabla del Análisis de Varianza

Fuentes de Variación	Sumas de cuadrados	Grados de libertad	Medias Cuadráticas	F
Inter	SC_{Inter}	$a-1$	$MC_{Inter} = \frac{SC_{Inter}}{a-1}$	$\frac{MC_{Inter}}{MC_{Intra}}$
Intra	SC_{Intra}	$N-a$	$MC_{Intra} = \frac{SC_{Intra}}{N-a}$	$(g.l. = a-1, N-a)$
Total	SC_{Total}	$N-1$		

5.5.2.- Modelo de Efectos aleatorios

Todo lo dicho hasta ahora en el apartado anterior sobre el modelo de un factor, se refería al modelo de efectos fijos. Decíamos en el apartado de conceptos básicos que un modelo es de efectos fijos cuando el investigador establece como niveles del factor sólo aquellos en los que está interesado en estudiar. Si el modelo es de efectos aleatorios, consideramos que los *i* niveles del factor son una muestra aleatoria de todos los posibles niveles. Esta muestra aleatoria de niveles se supone distribuida según $N(0, \sigma_A)$, siendo *A* el factor en cuestión.

Aunque el planteamiento es distinto, cuando se trata de un solo factor, este hecho no tiene consecuencias para el cálculo. El estadístico de contraste sigue siendo el mismo:

$$F = \frac{MC_{Inter}}{MC_{Intra}}$$

donde F se distribuye según la distribución F de Snedecor con $(a-1)$ y $(N-a)$ grados de libertad.

5.5.3.- Cálculo del ANOVA mediante el método clásico

A continuación desarrollaremos un ejemplo mediante las fórmulas expuestas anteriormente. Dichas fórmulas son empleadas en muchos libros de texto y tienen la ventaja de ser más intuitivas respecto a qué es lo que nos indican. Por motivos didácticos, utilizaremos un ejemplo no equilibrado.

Ejemplo 5.2. Supongamos que tenemos una muestra aleatoria de nueve sujetos que puntúan alto en ansiedad con los que formamos tres grupos en función de tres dosis distintas de cierta droga (0,05 mg, 0,10 mg y 0,20 mg). Queremos comprobar si estas dosis influyen en el estado de ansiedad de los sujetos. Suponemos que se cumplen los supuestos de independencia de las observaciones, normalidad de las distribuciones y homocedasticidad. Los resultados fueron:

$a_1 = 0{,}05$	$a_2 = 0{,}10$	$a_3 = 0{,}20$
5	6	2
8	7	4
4	8	
3		

Nivel de confianza NC = 95%.

Condiciones y supuestos. Suponiendo que las puntuaciones en ansiedad están medidas en una escala de intervalo o de razón, el enunciado nos dice que se cumplen los supuestos para aplicar el ANOVA.

- Independencia de las observaciones.
- Normalidad de las distribuciones en todos los niveles.
- Homocedasticidad (varianzas iguales en todos los niveles).

Hipótesis. Las hipótesis que debemos plantear son:

$H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: \mu_1 \neq \mu_2 \neq \mu_3$ al menos para un par de μ_i

Estadístico de contraste y distribución muestral. Tenemos que realizar una serie de operaciones para deducir el valor del estadístico de contraste. Comenzamos calculando las sumas de las puntuaciones y de sus cuadrados para todos los niveles.

	a_1	a_2	a_3	a_1^2	a_2^2	a_3^2
	5	6	2	25	36	4
	8	7	4	64	49	16
	4	8		16	64	
	3			9		
Sumatorio	20	21	6	114	149	20

A continuación calculamos la suma de todas las puntuaciones y la suma de todos los cuadrados:

$$\sum\sum Y_{ij} = 20 + 21 + 6 = 47 \qquad \sum\sum Y_{ij}^2 = 114 + 149 + 20 = 283$$

Observamos que hemos omitido los subíndices y superíndices de los sumatorios. Si tenemos un doble sumatorio sin superíndices ni subíndices, sabemos que hemos de realizar los cálculos correspondientes para todas las puntuaciones.

Ya podemos calcular las sumas de cuadrados intergrupo, intragrupo o error y total mediante las siguientes fórmulas:

$$SC_{Total}=\sum_{i=1}^{a}\sum_{j=1}^{n_j}\left(Y_{ij}-\bar{Y}_T\right)^2=\sum\sum Y_{ij}^2-\frac{\left(\sum\sum Y_{ij}\right)^2}{N}$$

$$SC_{Inter}=\sum_{i=1}^{a}n_i\left(\bar{Y}_{A_i}-\bar{Y}_T\right)^2=\left[\sum_{i=1}^{a}\frac{\left(\sum Y_{ij}\right)^2}{n_i}\right]-\frac{\left(\sum\sum Y_{ij}\right)^2}{N}$$

$$SC_{Intra}=\sum_{i=1}^{a}\sum_{j=1}^{n_j}\left(Y_{ij}-\bar{Y}_{A_i}\right)^2=SC_{Total}-SC_{Inter}$$

Utilizamos las fórmulas del último término de cada igualdad, porque los cálculos serán, en general, mucho más fáciles de realizar.

$$SC_{Total}=\sum\sum Y_{ij}^2-\frac{\left(\sum\sum Y_{ij}\right)^2}{N}=283-\frac{47^2}{9}=37{,}56$$

$$SC_{Inter}=\left[\sum_{i=1}^{a}\frac{\left(\sum Y_{ij}\right)^2}{n_i}\right]-\frac{\left(\sum\sum Y_{ij}\right)^2}{N}=\frac{20^2}{4}+\frac{21^2}{3}+\frac{6^2}{2}-\frac{47^2}{9}=19{,}56$$

$$SC_{Intra}=SC_{Total}-SC_{Inter}=37{,}56-19{,}56=18$$

Una vez completada la tabla de ANOVA, comprobamos que el valor del estadístico de contraste es $F=3{,}26$. La distribución muestral sigue un modelo F de Fisher con 2 y 6 grados de libertad asociados al numerador y denominador respectivamente.

F.V.	**S.C.**	**g.l.**	**M.C.**	**F**
Inter	19,56	$a-1=3-1=2$	$\frac{19{,}56}{2}=9{,}78$	$F=\frac{9{,}78}{3}=3{,}26$
Intra	18	$N-a=9-3=6$	$\frac{18}{6}=3$	
Total	37,56	$N-1=9-1=8$	$\frac{37{,}56}{8}=4{,}69$	

Establecer la regla de decisión en función del nivel de confianza. Cuando realizamos un análisis de la varianza y aunque la hipótesis alternativa siempre será bilateral, dejaremos todo el nivel de significación (α) por la parte derecha de la distribución. Por lo tanto, tenemos que buscar en las tablas la puntuación que supera una proporción igual a $1-\alpha$. En nuestro caso $1-\alpha=0{,}95$. Para este nivel de confianza con 2 y 6 grados de libertad, el valor de la tabla es: $F_{0{,}95;2{,}6}=5{,}143$.

Conclusión. Con un nivel de confianza del 95% no podemos rechazar la hipótesis nula, puesto que el estadístico de contraste (3,26) es menor que el valor crítico (5,143).

En cuanto al nivel crítico p, con un programa informático adecuado, comprobaríamos que es igual a $p=0{,}1101$. Con las tablas que manejamos tan sólo podríamos concluir que el nivel crítico es $p>0{,}10$, puesto que para 2 y 6 grados de libertad observamos que la puntuación que supera una proporción de 0,90 es igual a $3{,}463$.

Interpretar el resultado en función del contexto de la investigación. Con un nivel de confianza del 95% no podemos afirmar que los tratamientos suministrados influyan de forma distinta sobre la ansiedad.

5.5.4.- Cálculo del ANOVA mediante las razones básicas. En los siguientes temas (6 y 7) trabajaremos con modelos de ANOVA más complejos, por lo que los cálculos a realizar serán más laboriosos. Para facilitar dichos cálculos, los realizaremos mediante la notación que introducimos en el siguiente ejemplo.

Ejemplo 5.3. Supongamos un grupo de fumadores que quieren dejar este hábito y están dispuestos a participar en un ensayo en el que se prueban diversas terapias: terapia conductual (a_1), hipnosis (a_2) y administración de un fármaco (a_3). En total se seleccionan 18 sujetos que consumen 2 cajetillas diarias (40 cigarrillos) antes del tratamiento, los cuales son asignados aleatoriamente a cada uno de los grupos. En el transcurso de la experiencia un sujeto del grupo a_1 abandona la terapia. La variable dependiente fue el número de cigarrillos fumados el día después de seis meses de terapia. Los resultados fueron:

a_1	a_2	a_3
4	7	7
3	9	5
7	10	5
8	11	5
9	14	6
	12	11

Nivel de confianza: 95%.

Condiciones y supuestos. La variable dependiente está medida en una escala de razón. Hemos de suponer que las observaciones son independientes, las distribuciones normales para cada grupo y las varianzas homogéneas.

Hipótesis.

$$H_0: \mu_1 = \mu_2 = \mu_3$$
$H_1: \mu_1 \neq \mu_2 \neq \mu_3$ al menos para un par de μ_i

Estadístico de contraste y distribución muestral. Al igual que en el ejemplo anterior, comenzamos calculando la suma de todas las puntuaciones y la suma de los cuadrados.

	a_1	a_2	a_3	a_1^2	a_2^2	a_3^2
	4	7	7	16	49	49
	3	9	5	9	81	25
	7	10	5	49	100	25
	8	11	5	64	121	25
	9	14	6	81	196	36
		12	11		144	121
Sumatorio	$A_1 = 31$	$A_2 = 63$	$A_3 = 39$	219	691	281

Representaremos por A_i a la suma de cada nivel y utilizaremos la letra T (total) para la suma de todas las puntuaciones, por lo que:

$$T = \sum A_i = 31 + 63 + 39 = 133.$$

A continuación calculamos las siguientes razones básicas:

$$[Y] = \sum\sum Y^2 = 219 + 691 + 281 = 1191$$

$$[A] = \frac{A_1^2}{n_1} + \frac{A_2^2}{n_2} + \frac{A_3^2}{n_3} = \frac{31^2}{5} + \frac{63^2}{6} + \frac{39^2}{6} = 1107{,}2$$

$$[T] = \frac{T^2}{N} = \frac{133^2}{17} = 1040{,}53$$

A partir de las razones básicas calculamos fácilmente las sumas de cuadrados:

$$SC_{Inter} = [A] - [T] = 1107{,}2 - 1040{,}53 = 66{,}67$$

$$SC_{Intra} = [Y] - [A] = 1191 - 1107{,}2 = 83{,}8$$

$$SC_{Total} = [Y] - [T] = 1191 - 1040{,}53 = 150{,}47$$

A continuación resumimos los resultados en una tabla del ANOVA, donde aparece el estadístico de contraste que se distribuye según la F de Fisher con 2 y 14 grados de libertad.

F.V.	S.C.	g.l.	M.C.	F
Inter	66,67	2	33,33	5,56
Intra	83,8	14	5,99	
Total	150,47	16		

Regla de decisión en función del nivel de confianza. Con un nivel de confianza del 95% el valor que consultamos en las tablas es: $\boldsymbol{F_{0{,}95;2;14} = 3{,}739}$. Rechazamos la hipótesis nula puesto que $\boldsymbol{5{,}56 > 3{,}739}$.

El nivel crítico *p*, calculado con un programa informático, vale $p = 0{,}0167$. Con las tablas habríamos llegado a la conclusión de que: $0{,}01 < p < 0{,}025$.

Interpretar el resultado en función del contexto de la investigación. Con un nivel de confianza del 95% concluimos que las tres terapias no son igualmente eficaces.

5.6.- Comparaciones múltiples

Un objetivo de toda investigación consiste en obtener el máximo de resultados e interpretaciones a partir de las observaciones de que disponemos. Cuando un investigador aplica un Análisis de Varianza a un conjunto de observaciones para detectar si hay diferencias entre los distintos grupos sometidos a tratamientos distintos, o lo que es lo mismo, si hay diferencias entre los resultados de la variable dependiente en función de distintos niveles de la variable independiente (o de distintos niveles de distintas variables independientes como veremos en el tema siguiente), lo único que obtiene es un razón F que, en caso de ser un resultado significativo (es decir que no podamos mantener la H_0 de que todas las medias son iguales) nos lleva a una H_1 que indica que **al menos** entre dos medias hay diferencias que no son debidas al azar. Pero no nos dice dónde existen esas diferencias significativas.

Las comparaciones múltiples permiten establecer una información más exacta sobre la importancia de cada uno de los niveles de la variable independiente.

Al plantearnos el análisis de las comparaciones múltiples entre las medias de los distintos niveles de la variable independiente, cabe distinguir dos situaciones básicas:

1.- La primera se refiere a la situación más común en la que el investigador, una vez realizado el Análisis de Varianza y rechazada la H_0, desea conocer entre qué medias hay diferencias no debidas al azar. Se trata de las comparaciones **no planificadas, *a posteriori* o *post hoc***.

2.- La segunda se refiere a cuando el investigador no está interesado en realizar un Análisis de Varianza para probar todas las medias sino sólo en algunas comparaciones entre los niveles del factor, no en todas las posibles, y sabe de antemano qué comparaciones le interesan. Se trata de **comparaciones planificadas o *a priori***.

El objeto de las comparaciones múltiples es, como parte del Análisis de Varianza, reducir la cantidad de error Tipo I que cometeríamos si comparásemos dos a dos todas las muestras. Por lo tanto, aunque comparemos las muestras dos a dos, no recurrimos a la prueba *t* estudiada en temas precedentes, sino que aplicaremos pruebas específicas que aprovechan los resultados del Análisis de Varianza y que nos aseguran que no se incrementa el error de tipo I (α).

5.6.1.- Comparaciones planificadas o *a priori*

Como decíamos en el apartado 5.6, en ocasiones el investigador no está interesado en comparar dos a dos todas las medias, sino sólo algunas combinaciones de los niveles del factor pero, dado que si se realizan comparaciones dos a dos con la prueba *t* se incrementa el error de tipo I (según hemos visto al inicio del tema), realiza un Análisis de Varianza, aunque no es su resultado el que le interesa, sino esas comparaciones respecto a las hipótesis específicas diseñadas de antemano. De ahí su nombre de comparaciones planificadas o *a priori*. De hecho, en los resultados de la investigación no suele citarse, siquiera, el valor F obtenido en el análisis de varianza.

Ejemplo 5.4. Supongamos que estamos interesados en conocer el efecto que sobre la reducción de la ansiedad tiene el realizar terapias conductuales (vamos a suponer 4 terapias distintas). No nos interesa comparar las terapias entre sí. Tomamos a un grupo de sujetos, que puntúan semejante en ansiedad y los distribuimos aleatoriamente en cinco grupos: uno de ellos será el grupo control y los otros cuatro serán tratados cada uno de ellos con una terapia distinta. Al finalizar la terapia medimos la variable dependiente (ansiedad) de los cinco grupos.

Tendríamos que comparar, para probar nuestra hipótesis, las puntuaciones del grupo control con las de los otros cuatro grupos (cuatro comparaciones).

Hemos visto en la Introducción del tema cómo a medida que aumenta el número de grupos a comparar nos alejamos del nivel de riesgo inicial con el que queríamos trabajar (aumentamos la probabilidad de cometer un error de tipo I). Es por ello que recurrimos al Análisis de Varianza, aunque desechemos parte de sus resultados.

Dado el alcance del curso, no vamos a detenernos en este tipo de comparaciones, pero sí apuntarlas para que el estudiante conozca su existencia.[3]

5.6.2.- Comparaciones no planificadas, *a posteriori* o *post hoc*

Las comparaciones no planificadas, *a posteriori* o *post hoc* (simplemente comparaciones múltiples para algunos autores) son aquellas, como hemos dicho, que se deciden después de que el investigador haya obtenido los resultados del Análisis de Varianza y haya rechazado la hipótesis nula. Aunque existen distintas técnicas para realizar estas comparaciones, aquí vamos a estudiar sólo una de ellas: la prueba de comparaciones múltiples de Scheffé que es una de las más utilizadas.

5.6.3.- Prueba de comparaciones múltiples de Scheffé

Esta prueba propuesta por Scheffé permite no sólo comparar las medias de los niveles del Análisis de Varianza dos a dos, sino también realizar comparaciones complejas, esto es, permite comparar la media de

[3] Remitimos al estudiante interesado en el tema de las comparaciones múltiples al libro clásico de Keppel (1973) *Design and anlysis: A researcher's handbook*. Editado por Prentice Hall, Inc.

un nivel con un conjunto de medias de otros niveles. También permite comparar un conjunto de medias de distintos niveles con otro conjunto de medias de otros niveles.

Esta prueba fija la tasa de error de tipo I en el α al que estemos trabajando, sin aumentarlo en todas las posibles comparaciones que realicemos y obtiene un valor al que llama diferencia mínima o rango crítico (***C**ritical* ***R**ange* de Scheffé) por encima de la cual diremos que hay diferencias entre las medias o entre los grupos de medias que estemos comparando. Esta diferencia mínima se calcula según la fórmula:

$$CR_{Scheffe} = \sqrt{(a-1)\cdot F_{(a-1),(N-a)}}\sqrt{MC_{Intra}\left[\sum\left(c_i^2/n_i\right)\right]}$$

Donde c_i son los coeficientes de las combinaciones lineales que podemos establecer entre las distintas medias a comparar y el resto de valores se pueden obtener de la tabla del ANOVA. En cada combinación de coeficientes c_i la suma de los mismos debe ser igual a cero. Por ejemplo, si tenemos tres niveles en el factor y queremos hacer todas las comparaciones posibles dos a dos de las medias, resultarán tres comparaciones y los coeficientes serían:

	a_1	**a_2**	**a_3**	**Medias a comparar**
	1	-1	0	$\bar{Y}_1 - \bar{Y}_2$
Coeficientes	1	0	-1	$\bar{Y}_1 - \bar{Y}_3$
	0	1	-1	$\bar{Y}_2 - \bar{Y}_3$

Observamos que cuando uno de los coeficientes es 0 esto significa que elimina a ese grupo de la comparación. Cuando las comparaciones implican a más de dos grupos, los valores de los coeficientes deben reflejar los grupos a comparar y el tipo de comparación. En la Tabla 5.3, se muestran algunos ejemplos de comparaciones en un diseño de un factor con 5 niveles.

Tabla 5.3. *Ejemplos de coeficientes para comparación de medias de grupos*

	Hipótesis sobre comparaciones	a_1	a_2	a_3	a_4	a_5	$\sum c_i$
		Coeficientes c_i					
Ej. 1	$H_0: \mu_1 = \frac{\mu_2+\mu_3}{2}$	2	-1	-1	0	0	0
Ej. 2	$H_0: \mu_5 = \frac{\mu_2+\mu_4}{2}$	0	-1	0	-1	2	0
Ej. 3	$H_0: \frac{\mu_1+\mu_5}{2} = \frac{\mu_2+\mu_3+\mu_4}{3}$	3	-2	-2	-2	3	0
Ej. 4	$H_0: \mu_2 = \frac{\mu_3+\mu_4+\mu_5}{3}$	0	3	-1	-1	-1	0
Ej. 5	$H_0: \mu_1 = \frac{\mu_2+\mu_3+\mu_4+\mu_5}{4}$	4	-1	-1	-1	-1	0

Observe el estudiante que los coeficientes están relacionados con el número de grupos a cada lado de la expresión de H_0. Así por ejemplo, en el Ej. 3 de la Tabla 5.3, el más complicado quizás, se trata de comparar si la media de las medias de los grupos 1 y 5 es igual a la media de las medias de los grupos 2, 3 y 4. La misma expresión formulada en la hipótesis nula, se puede escribir de la siguiente manera:

$$H_0: 3(\mu_1 + \mu_5) - 2(\mu_2 + \mu_3 + \mu_4) = 0$$

Si se multiplican los factores 2 y 3 por la expresión dentro de su correspondiente paréntesis, se ve claramente el valor de los coeficientes que afectan a cada media de grupo.

Ejemplo 5.5. Supongamos que tenemos una muestra aleatoria de 21 sujetos que puntúan alto en ansiedad y los distribuimos aleatoriamente en tres grupos de 7 sujetos cada uno. Les suministramos tres dosis distintas de una cierta droga: 0,05 mg, 0,10 mg y 0,20 mg. Queremos ver si estas dosis influyen en el estado de ansiedad de los sujetos. Se cumplen los supuestos de independencia de las observaciones, normalidad de las distribuciones y homocedasticidad. La variable dependiente es de intervalo. Una vez realizado el Análisis de Varianza, los resultados fueron:

F.V.	S.C.	g.l.	M.C.	F
Intergrupo	164,95	2	82,475	5,069
Intragrupo	292,86	18	16,27	
Total	457,81	20		

$$\bar{Y}_1 = 11{,}7143 \qquad \bar{Y}_2 = 15{,}43 \qquad \bar{Y}_3 = 8{,}57$$

Si acudimos a las tablas de la distribución F, el valor crítico para 2 y 18 grados de libertad, trabajando con un nivel de confianza del 95%, es 3,555. Comparando nuestro resultado con el de las tablas vemos que el estadístico de contraste supera al valor crítico (5,069 > 3,555) por lo que rechazaremos la H_0 de igualdad de medias. Ahora bien, ¿entre qué pares de medias está la diferencia que hace que rechacemos la hipótesis nula?

Aplicamos la prueba de comparaciones múltiples de Scheffé:

$$CR_{Scheff\acute{e}} = \sqrt{(a-1)F_{(a-1),(N-a)}}\sqrt{MC_{Intra}\left(\sum_{i=1}^{a}\left(c_i^2/n_i\right)\right)}$$

Si queremos comparar todas las medias entre sí, tenemos tres posibles comparaciones, y establecemos tres combinaciones lineales:

Media 1 con media 2: $(1)\bar{Y}_1 + (-1)\bar{Y}_2 + (0)\bar{Y}_3$
Media 1 con media 3: $(1)\bar{Y}_1 + (0)\bar{Y}_2 + (-1)\bar{Y}_3$
Media 2 con media 3: $(0)\bar{Y}_1 + (1)\bar{Y}_2 + (-1)\bar{Y}_3$

Observamos, que cuando queremos comparar todas las medias dos a dos, siempre tendremos que, por ejemplo, para las medias "i" y "j":

$$\sum_{i=1}^{a}\left(c_i^2/n_i\right)=\frac{1^2}{n_i}+\frac{(-1)^2}{n_j}$$

Y si el modelo es equilibrado, como en nuestro ejemplo, para todas las comparaciones hemos de calcular:

$$\sum_{i=1}^{a}\left(c_i^2/n_i\right)=\frac{1^2}{n}+\frac{(-1)^2}{n}=\frac{2}{n}$$

Dado que nuestro ejemplo es equilibrado donde $n=7$, para las tres comparaciones que hemos de realizar, tendremos que calcular:

$$\sum_{i=1}^{a}\left(c_i^2/n_i\right)=\frac{2}{n}=\frac{2}{7}$$

La diferencia mínima o rango crítico por encima de la cual diremos que hay diferencias entre las medias es igual a:

$$CR_{Scheffé}=\sqrt{2\cdot 3{,}555}\sqrt{16{,}27\cdot\frac{2}{7}}=5{,}749$$

Finalmente realizamos las comparaciones para cada par de medias:

$$|\bar{Y}_1-\bar{Y}_2|=|11{,}7143-15{,}43|=3{,}7157<5{,}749\rightarrow H_0$$
$$|\bar{Y}_1-\bar{Y}_3|=|11{,}7143-8{,}57|=3{,}1443<5{,}749\rightarrow H_0$$
$$|\bar{Y}_2-\bar{Y}_3|=|15{,}43-8{,}57|=6{,}86>5{,}749\rightarrow H_1$$

Como vemos, la única diferencia significativa (responsable de que hayamos rechazado la hipótesis nula en el análisis de varianza ómnibus) se da entre los grupos 2 y 3, ya que la diferencia de medias entre estos dos grupos supera el valor CR de Scheffé.

Decíamos que una de las posibilidades que ofrece la prueba de Scheffé es que nos permite realizar comparaciones más complejas que la simple comparación, dos a dos, de las muestras. Veamos un ejemplo.

Ejemplo 5.6. Supongamos, siguiendo con el mismo ejemplo, que de las tres dosis distintas la primera (0,05), a la que llamaremos *A*, es la dosis tradicional que se suele suministrar mientras que las otras dos, a las que llamaremos *B* y *C,* son dosis que están en estudio y, al comprobar que el resultado del Análisis de Varianza es significativo (se rechaza la H_0). Se pretende probar si el efecto de la dosis tradicional es distinto al efecto de las otras dosis experimentales, tomadas en su conjunto.

La H_0 que queremos probar es que la media del grupo *A* es igual a la media de los grupos *B* y *C.*

$$H_0 : \mu_A = \frac{\mu_B + \mu_C}{2}$$

O bien:

$$H_0 : \mu_A - \frac{\mu_B + \mu_C}{2} = 0$$

Si aplicamos la $CR_{Scheffé}$:

a) $n_i = 7; \quad a = 3; \quad \bar{Y}_1 = 11{,}7143; \quad \bar{Y}_2 = 15{,}43; \quad \bar{Y}_3 = 8{,}57$

b) El valor de la distribución F para un $\alpha = 0{,}05$ con 2 grados de libertad en el numerador y 18 grados de libertad en el denominador, es igual a 3,555.

c) $MC_{error} = 16{,}27$

d) Dado que sólo queremos realizar una comparación, sólo tenemos que establecer una combinación lineal para ver sus coeficientes (c_i). Dicha combinación será:

$$(2)\bar{Y}_1 + (-1)\bar{Y}_2 + (-1)\bar{Y}_3$$

Por lo que:

$$CR_{Scheffé} = \sqrt{2 \cdot 3{,}555}\sqrt{16{,}27\left(\left({2^2}/{7}\right) + \left({-1^2}/{7}\right) + \left({-1^2}/{7}\right)\right)} = 9{,}96$$

Esta es la diferencia mínima o rango crítico (***C****ritical* ***R****ange* de Scheffé), por encima de la cual (en valores absolutos) diremos que hay diferencias entre los grupos de medias que estemos comparando. Por lo tanto, realizamos la comparación:

$$|(2)(11{,}7143) + (-1)(15{,}43) + (-1)(8{,}57)| = 0{,}5714$$

e) Si comparamos $0{,}5714 < 9{,}96$ concluimos que no podemos rechazar la hipótesis nula en este contraste *a posteriori* y, por lo tanto interpretamos, con un nivel del confianza del 95%, que el método tradicional no difiere de los otros dos métodos tomados conjuntamente.

5.7. Supuestos del análisis de varianza

Tras ver en qué consiste el Análisis de Varianza, vamos a analizar ahora las condiciones que deben cumplir los datos para que pueda aplicarse el ANOVA. Estas son:

a) <u>Independencia</u>. Es decir, que las distintas muestras o grupos a comparar hayan sido obtenidas aleatoriamente. Esto implica que tanto las muestras como las observaciones deben ser independientes.
b) <u>Normalidad de las distribuciones</u>. Las muestras o grupos que comparamos deben proceder de poblaciones que se distribuyan normalmente en la variable estudiada.

c) <u>Homogeneidad de las varianzas</u> (homocedasticidad). Los grupos a comparar deben proceder de poblaciones que no difieran significativamente en sus varianzas para la variable estudiada.

Dado el alcance del temario no vamos a detenernos en el desarrollo de las distintas técnicas estadísticas que se suelen utilizar para probar que se cumplen estos supuestos. Para probar la independencia de las observaciones puede utilizarse, por ejemplo, el test de rachas; para probar la normalidad las pruebas más utilizadas tenemos la prueba de Chi-cuadrado de Pearson, la prueba de Kolmogorov-Smirnov o la prueba de Lilliefors; y para probar la homocedasticidad se suele acudir a las pruebas de Cochran, de Bartlett o el test de Levene.

Aún insistiendo en la importancia de probar que se cumplen los supuestos, antes de proceder a realizar un Análisis de Varianza, resulta imposible (por la extensión del curso) desarrollarlos aquí. Los paquetes estadísticos para el cálculo del Análisis de Varianza comienzan probando los supuestos.

5.8.- Resumen

En este capítulo hemos iniciado el estudio del análisis de los diseños de más de dos muestras, comenzando por el caso más sencillo: el de un factor (variable independiente) con más de dos niveles o grupos, en el que cada nivel está integrado por un conjunto distinto de sujetos, asignados de manera aleatoria. Esta técnica de análisis se conoce como Análisis de Varianza (ANOVA)

- La técnica de análisis difiere de la seguida en el diseño de dos grupos, en el cual se contrasta la hipótesis nula de diferencias de medias, mediante la comparación directa de éstas. En el ANOVA, aunque la hipótesis estadística de partida es semejante a "las medias de los grupos son iguales", el enfoque del análisis es muy diferente pues, como hemos visto, se hace a través del estudio de dos variabilidades como estimadores de la variabilidad total del sistema (de ahí su nombre).

- Una vía de estimación es a través de la variabilidad que se da en las puntuaciones individuales de cada grupo y otra es la variabilidad que se da entre los promedios generales de cada grupo. Si estas dos estimaciones son similares se deduce que no hay un efecto diferencial entre los grupos, pero si no lo son, la conclusión es que, además de la variabilidad intrínseca de un experimento, fruto del error asistemático, hay otra fuente de variabilidad fruto de las propias diferencias, estás sistemáticas, de los tratamientos entre sí.

- El estadístico F es el que, a la postre, determina si la comparación de estas variabilidades (denominadas Medias Cuadráticas) es o no estadísticamente significativa. Esta prueba es la que se conoce como prueba general o prueba *ómnibus* y sirve para el contraste de la hipótesis nula general de igualdad de medias de las poblaciones objeto de estudio.

- Realizada la prueba general, si resulta significativa, se daría un paso más en la dirección de establecer entre qué grupos o combinaciones de grupos (dependiendo del tipo de contraste) se producen estas diferencias. Para estos contrastes hay una serie de estadísticos (Scheffé, entre otros) que informan de la significación estadística de las diferencias.

- También hemos visto el modelo teórico en que se basa el ANOVA, y los supuestos subyacentes: independencia; normalidad y homocedasticidad.

5.9.- Ejercicios de autoevaluación

Se diseñó un experimento[4] para evaluar el efecto de diferentes tipos de entrenamiento para que los niños adquieran la habilidad del concepto "triángulo equilátero". En total participaron en la experiencia 40 niños de tres años de una escuela infantil, que fueron asignados aleatoriamente a cada uno de los siguientes cuatro grupos:

- **Nivel a_1**. Condición visual. Se mostraron 30 figuras de madera de forma consecutiva y la instrucción era mirarlos pero no tocarlos.
- **Nivel a_2**. Condición visual y motora. Los niños observan los bloques y además pueden jugar con ellos. Se les pide que realicen ejercicios táctiles específicos tales como desplazar el dedo índice por el perímetro de las piezas.
- **Nivel a_3**. Condición visual y verbal. Los niños miran los bloques y se les hace notar las diferencias en color, forma, tamaño y grosor.
- **Nivel a_4**. Los niños en este nivel no realizan actividad alguna.

Todas las actividades se desarrollaron de manera individual. El día después del entrenamiento, a cada niño se les enseñaba durante cinco segundos una figura y luego se le pedía que identificara esa pieza entre un conjunto de otras 7 piezas. La tarea se repitió seis veces usando diferentes figuras de referencia. Como variable dependiente se tomó el número de piezas correctamente identificadas. Los resultados se muestran en la siguiente tabla:

a_1	a_2	a_3	a_4
0	2	2	1
1	3	4	0
3	4	5	2
1	2	3	1
1	1	2	1
2	1	1	2
2	2	3	1
1	2	3	0
1	3	2	1
2	4	4	3

Nivel de confianza: 95%

[4] Los datos y el diseño del experimento están adaptados del libro ya referido de Kirk, R.E. *Experimental Designs. Procedure for de Behavioral Sciences*. Pág 206, que a su vez toma de una investigación de Nelson, G.K. (1976) *Concomitant effects of visual, motor, and verbal experiences in young children's concept development. Journal of Educational Psychology*, *68*, 466-473.

Preguntas

1. ¿Cuál es el resultado de las Media Cuadrática Intra-grupos (MC_{Intra})? A) 1,0361; B) 4,1536; C) 37,30
2. ¿Cuál es el resultad de la Suma de Cuadrado Inter-grupos (SC_{Inter})? A) 56,975; B) 19,675; C) 37,300
3. ¿Cuál es el resultado del estadístico de la prueba ómnibus? A) 6,330; B) 2,866; C) 1,314
4. Si queremos contrastar la hipótesis $\left(H_0: \mu_3 - \frac{\mu_1+\mu_2+\mu_4}{3} = 0\right)$, ¿cuál sería, aproximadamente, el valor de la diferencia crítica ($CR_{Scheff\acute{e}}$) según la Prueba de Scheffé? A) 3,265; B) 1,372; C) 0,956

Soluciones

Condiciones y supuestos. La variable dependiente tiene un nivel de medida de razón, suponemos que las observaciones son independientes, que las distribuciones se distribuyen normalmente en cada grupo y que las varianzas son homogéneas.

Hipótesis.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ al menos para un par de μ_i

Estadístico de contraste. Se han de realizar los cálculos necesarios para obtener el estadístico F de Fisher.

	a_1	a_2	a_3	a_4	a_1^2	a_2^2	a_3^2	a_4^2
	0	2	2	1	0	4	4	1
	1	3	4	0	1	9	16	0
	3	4	5	2	9	16	25	4
	1	2	3	1	1	4	9	1
	1	1	2	1	1	1	4	1
	2	1	1	2	4	1	1	4
	2	2	3	1	4	4	9	1
	1	2	3	0	1	4	9	0
	1	3	2	1	1	9	4	1
	2	4	4	3	4	16	16	9
Sumas	14	24	29	12	26	68	97	22

Razones básicas:

$$T = \sum A_i = 14 + 24 + 29 + 12 = 79 \qquad [T] = \frac{79^2}{40} = 156,025$$

$$[A] = \frac{14^2}{10} + \frac{24^2}{10} + \frac{29^2}{10} + \frac{12^2}{10} = 175,7 \qquad [Y] = 26 + 68 + 97 + 22 = 213$$

Sumas de cuadrados:

$SC_{Inter} = [A] - [T] = 19{,}675$ $\quad SC_{Intra} = [Y] - [A] = 37{,}3$ $\quad SC_{Total} = [Y] - [T] = 56{,}975$

F.V.	S.C.	g.l.	M.C.	F
Inter	19,675	3	6,558	6,330
Intra	37,3	36	1,036	
Total	56,975	39		

Pregunta 1. A

Pregunta 2. B

Pregunta 3. A

Pregunta 4. A **Prueba de Scheffé.** Coeficientes para la prueba

	a₁	**a₂**	**a₃**	**a₄**	
c	-1	-1	3	-1	
c^2	1	1	9	1	$\sum c^2 = 12$

$$CR_S = \sqrt{\left((a-1)(F_{a-1;N-a}\right)\left(MS_{Intra}\sum\frac{c_i^2}{n_i}\right)} = \sqrt{(4-1)\mathbf{2{,}866}\left(1{,}036\frac{12}{10}\right)} = 3{,}269$$

Nota. El valor F: $F_{0,95;3;36} = 2{,}866$ ha sido calculado con un programa informático. Con las tablas del texto habríamos buscado con los grados de libertad más próximos: $F_{0,95;3;36} \approx F_{0,95;3;30} = 2{,}922$

Tema 6. DISEÑOS DE MÁS DE DOS GRUPOS CON MUESTRAS RELACIONADAS

6.1.- Introducción

En este capítulo aprenderemos a analizar los resultados obtenidos en un diseño experimental en donde se ha manipulado una única variable independiente (es decir, un único factor en la terminología ya estudiada en el Capítulo 5) con más de dos niveles pero de forma intra-sujeto. Esto significa que todos los participantes (o unidades de observación) han pasado por todos los niveles del factor. A este tipo de diseños también se les conoce como diseños de medidas repetidas en el sentido de que a cada sujeto se le repite la medición de la variable dependiente en diversas condiciones, tantas como niveles tenga el factor manipulado. También se les conoce como diseños de medidas dependientes debido a que las puntuaciones de un mismo sujeto muestran dependencia estadística entre ellas. Esta particularidad tendrá consecuencias importantes para el análisis debido a que las observaciones realizadas sobre cada unidad de observación se encuentran relacionadas.

6.2.- Objetivos

Al finalizar este capítulo, el alumno podrá ser capaz de:

- Diferenciar un diseño de un factor intra o intersujeto.
- Reconocer las ventajas e inconvenientes de cada uno de estos diseños.
- Aplicar los cálculos necesarios para evaluar la significatividad de un factor manipulado intra-sujeto, es decir, para determinar si el factor afecta a la variable dependiente.
- Informar de los resultados de un ANOVA de un factor intra-sujeto.

6.3.- Diseños de un factor intra-sujetos

Es posible que el alumno haya observado por sí mismo que, en algunos casos, un factor (es decir, una variable independiente) puede manipularse de manera distinta a la creación de grupos distintos, que es la forma en que se ha presentado hasta el momento la manipulación de una variable independiente y que se estudió en el capítulo previo. En ciertos casos, en vez de formar un grupo distinto de unidades de observación (usualmente los participantes en un experimento pero que también pueden ser empresas comerciales, departamentos, terapias, etc.) y someter a cada uno de los sujetos de estos grupos a un único nivel del factor, se puede pensar en someter a cada unidad de observación a todos los niveles del factor. En este caso diremos que el factor se ha manipulado intra-sujetos y lo representaremos introduciendo el nombre o símbolo del factor entre paréntesis.

Ejemplo 6.1.- Durante todo este capítulo trabajaremos con un fenómeno extraído de la Psicología de la Atención utilizado para evaluar defectos atencionales. Nos referimos al *efecto Stroop*. El efecto *Stroop* se define en función de la diferencia de los tiempos de reacción medios existente entre tres condiciones experimentales consistentes en presentar palabras de color (*v.g.*, verde) escritas en tintas de diferentes colores (*v.g.*, tinta roja, verde o azul). La tarea del sujeto es nombrar lo más rápidamente la tinta en que se encuentran escritas las palabras, no la palabra escrita. En estos experimentos, lo más usual es que existan tres condiciones (niveles del factor): congruente, incongruente y neutral. En la condición congruente se presentan palabras de color escritas en el mismo color que representan (*v.g.*, verde, rojo, azul). En la condición incongruente, las palabras y el color no coinciden (*v.g.*, verde, rojo, azul). El hecho de que el sujeto tenga que decir en voz alta y rápida "rojo" cuando se le presenta la palabra "verde" provoca un incremento del tiempo de reacción medio así como un incremento en el número de errores de esta condición en relación a la condición neutral. En la condición neutral, se suelen presentar estímulos semánticamente neutros (*v.g.*, xxxxx, xxxxx, xxxxx), aunque hay otras posibilidades. La tarea del sujeto sigue siendo nombrar el color de la tinta. La investigación sigue tratando hoy en día de identificar el mecanismo que produce este efecto (véase el artículo de Melara y Algom, "*Driven by information: A Tectonic Theory of Stroop Effects*" de 2003, para ver un modelo formal que trata de explicar los efectos experimentales observados utilizando este paradigma).

Siguiendo con el fenómeno y la situación experimental mostrada en el Ejemplo 6.1, cuando realizamos un experimento clásico para demostrar el efecto *Stroop* podemos realizarlo, a grandes rasgos, de dos maneras distintas.

Con un **diseño inter-sujetos** visto en el Tema 5, utilizaríamos tres grupos distintos de participantes y cada grupo sería sometido a una única condición, congruente, incongruente o neutral. En consecuencia, cada sujeto tendría única y exclusivamente una puntuación (la media del tiempo de reacción –TR– de todos aquellos ensayos de la condición a la que haya sido asignado).

Con un **diseño intra-sujetos**, por el contrario, presentaríamos a todos los participantes las tres condiciones experimentales. En consecuencia, obtendríamos para cada participante tres puntuaciones: la media del TR en la condición congruente, la media del TR en la condición incongruente y la media del TR en la condición neutral.

Para diferenciar simbólicamente este tipo de diseños (intra-sujetos) de los anteriores (inter-sujetos) introduciremos el factor experimental entre paréntesis en combinación con el factor sujetos. Por consiguiente, en nuestro caso lo representaríamos como (A × S). Es decir cuando veamos los símbolos A × S representaremos un experimento inter-sujetos, mientras que un experimento intra-sujetos lo representaremos mediante (A × S).

El sistema que utilizaremos para denotar las puntuaciones individuales, los sumatorios de cada condición y demás cálculos, es idéntico al utilizado en el diseño inter-sujetos y, para un seguimiento sencillo del procedimiento de análisis, presentaremos el capítulo al hilo de unos datos hipotéticos referentes al ejemplo presentado.

Ejemplo 6.2. Supongamos que en un experimento de Stroop clásico con tres condiciones hemos obtenido las puntuaciones medias de TR que se observan en la siguiente tabla:

Tabla 6.1

	Niveles del Factor A (Condición de Stroop)		
	Condición congruente	**Condición incongruente**	**Condición neutral**
	a_1	a_2	a_3
Participante 1	0,545	0,832	0,620
Participante 2	0,630	0,736	0,635
Participante 3	0,610	0,663	0,680
Participante 4	0,680	0,715	0,660
Participante 5	0,590	0,880	0,700
Participante 6	0,600	0,790	0,670

Con la letra mayúscula A designamos al factor, representando el número de niveles por la misma letra pero en minúscula. En nuestro caso a = 3.

Cada valor numérico que aparece en el grueso de la Tabla 6.1 representa la media de TR para varios ensayos (por ejemplo la media en 50 ó 100 ensayos para cada uno de los valores). Si en el experimento (ficticio) se hubiesen utilizado 100 ensayos por condición, esto significaría que cada sujeto habría tenido que realizar 300 ensayos (100 para la condición congruente, 100 para la incongruente y 100 para la neutral). Esto señala un inconveniente de los diseños intra-sujetos, a saber, a igualdad de condiciones, los participantes son sometidos a muchos más ensayos que en el diseño inter-sujetos, lo cual puede introducir efectos de fatiga y/o aprendizaje.

En nuestro ejemplo las condiciones experimentales se presentaron en diferente orden para cada sujeto (a unos sujetos se les presentaría primero la condición congruente, luego la incongruente y por último la neutral; otros pasarían primero por la neutral, luego la congruente y por último, la neutral; otros pasarán primero por la condición incongruente, luego la neutral y por último, la congruente). A este procedimiento metodológico para evitar que el orden de los tratamientos interfiera con los resultados se le conoce como contrabalanceo de las condiciones experimentales. No obstante, esta necesidad de contrabalanceo sólo se presenta si utilizamos un diseño en que todos los estímulos del mismo tipo se presentan agrupados. En el caso de que todos los estímulos se presentaran aleatoriamente, el control metodológico del contrabalanceo no tendría sentido.

Resulta útil considerar la Tabla 6.1 como el total entrecruzamiento de dos factores: el factor A que estamos manipulando (condición de Stroop) y los sujetos (Factor S), asemejándose en este sentido al diseño de dos factores que se verá en el Tema 7. En el diseño inter-sujetos este entrecruzamiento no existía debido a que cada sujeto sólo participaba en un nivel de A.

Se admite que los niveles absolutos de actuación de los sujetos pueden diferir (algunos serán mejores y otros peores, es decir, más rápidos y con menos errores o más lentos y con más errores) pero el patrón de respuesta ante las tres condiciones no debe diferir significativamente entre distintos sujetos si consideramos que A no interactúa con S. Esto es lo que significa que no exista interacción entre los factores A y S. Si distintos sujetos respondieran de forma diferente a las condiciones experimentales, tendríamos una interacción. En el diseño intra-sujetos se admite (o se asume) que no existe interacción entre el factor y los sujetos.

Como ya vimos en el tema 5, el análisis del ANOVA exige dividir los componentes de la **variabilidad total** observada en la variable dependiente (σ^2_{Total} o σ^2_T), **en varios componentes aditivos**. En nuestro caso encontramos tres componentes que puedan aportar variabilidad a los datos:

En primer lugar, la **variabilidad del factor que estamos manipulando** si efectivamente afecta al tiempo de reacción medio, ya que existe la posibilidad de que la variable independiente manipulada, el factor A, no afecte a la variable dependiente Y (en nuestro ejemplo, el TR).

En segundo lugar la **variabilidad de los sujetos**. Los participantes difieren en muchas características individuales y, en consecuencia, no se comportan exactamente igual ante las condiciones experimentales presentadas; en definitiva, también son una fuente de variabilidad.

Finalmente, la **variabilidad debida a la interacción entre el factor y los sujetos**. Es decir, la posibilidad de que no todos los sujetos respondan con el mismo patrón de respuestas ante las tres condiciones experimentales.

Por consiguiente buscamos la parte de variabilidad que aporta el factor manipulado (σ^2_A), la parte de variabilidad que aportan los sujetos o participantes (σ^2_S), y la parte de variabilidad que aporta la interaccion entre ambos ($\sigma^2_{(A\times S)}$), a la variabilidad total. Para realizar el tratamiento del análisis de datos necesitamos conocer las siguientes sumas de cuadrados:

$$SC_{Total} = SC_A + SC_S + SC_{(A\times S)}$$

La **hipótesis nula** (H_0) que ponemos a prueba afirma que no existen diferencias entre las medias de los diferentes tratamientos. La **hipótesis alternativa** (H_1) afirma que, al menos para una comparación entre un par de tratamientos, esas diferencias son reales:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$$

$H_1: \mu_1 \neq \mu_2 \neq \cdots \neq \mu_a$ *al menos para un par de tratamientos*

En cuanto a los **supuestos** que deben cumplirse para poder aplicar correctamente el ANOVA de un factor intra-sujetos son:

- La variable dependiente (a la que denotaremos genéricamente por Y) debe estar medida, al menos, a un nivel de intervalo.
- Las puntuaciones de la variable dependiente Y en cada nivel del factor deben ser independientes entre sí.
- Las puntuaciones de la variable dependiente Y en cada nivel del factor deben distribuirse según la curva normal. No obstante, el ANOVA es robusto ante el incumplimiento de este supuesto, por lo que se pueden encontrar estudios donde se aplicó el ANOVA como técnica de análisis sin cumplirse el supuesto (aunque no es lo más recomendable). Existen contrastes de hipótesis denominados de "bondad de ajuste" que se podrían aplicar para comprobar este supuesto y que no veremos en este curso.
- Las varianzas de las puntuaciones para los distintos niveles del factor deber ser iguales entre sí.

Estos supuestos son similares a los que se plantearon en el diseño de un factor inter-sujetos. En el diseño intra-sujetos hemos de añadir un supuesto adicional:

- Las covarianzas entre todos los niveles del factor deben ser iguales entre sí.

Los últimos dos supuestos pueden representarse conjuntamente mediante una tabla cuadrada (*a x a)* que está en función del número de niveles del factor. En nuestro caso:

	Congruente (C)	Incongruente (I)	Neutral (N)
Congruente (C)	σ^2_{CC}	$Cov(C,I)$	$Cov(C,N)$
Incongruente (I)	$Cov(I,C)$	σ^2_{II}	$Cov(I,N)$
Neutral (N)	$Cov(N,C)$	$Cov(N,I)$	σ^2_{NN}

En esta tabla observamos que la diagonal negativa (las celdillas CC, II y NN) representan las varianzas de cada condición (varianzas de la condición congruente, incongruente y neutral, respectivamente; hemos introducido los subíndices CC, II y NN porque la varianza de una variable es lo mismo que la covarianza entre una variable y ella misma). Hemos de asumir por lo tanto que:

$$\sigma_{CC}^2 = \sigma_{II}^2 = \sigma_{NN}^2$$

El incumplimiento de este supuesto es particularmente grave para el ANOVA debido a que esta técnica funciona dividiendo la varianza en diferentes componentes. Su evaluación se realiza, entre otros, mediante el test de Levene, que no veremos en este curso.

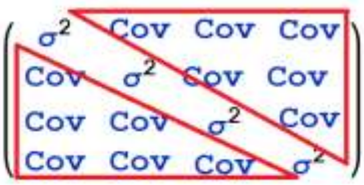
El resto de celdillas representan las covarianzas entre las puntuaciones de cada par de condiciones, recuérdese que en general: $Cov(X,Y) = Cov(Y,X)$, luego en nuestro ejemplo:

$$Cov(I,C) = Cov(C,I); \quad Cov(C,N) = Cov(N,C); \quad Cov(I,N) = Cov(N,I).$$

Por lo tanto, el último supuesto ha de plantear:

$$Cov(I,C) = Cov(C,N) = Cov(N,I)$$

Estos dos últimos supuestos se denominan en la literatura como "*simetría compuesta*" debido a que, en último término, conducen a una matriz de datos (una tabla) simétrica en relación a la diagonal negativa y compuesta solamente por dos valores (una varianza y una covarianza si se cumplen los supuestos).



En cuanto al estadístico de contraste para poner a prueba la hipótesis nula consiste en el cálculo de una razón F, es decir, de un cociente entre varianzas como se hizo en el Tema 5. En este caso partimos de una matriz de datos:

Matriz de datos					
	Factor A				
	a_1	a_2	…	a_a	**Suma**
Sujeto 1	$Y_{1,1}$	$Y_{2,1}$	…	$Y_{a,1}$	S_1
Sujeto 2	$Y_{1,2}$	Y_{22}	…	$Y_{a,2}$	S_2
Sujeto 3	$Y_{1,3}$	$Y_{2,3}$	…	$Y_{a,3}$	S_3
……….	…	…	…	…	…
Sujeto s	$Y_{1,s}$	$Y_{2,s}$	…	$Y_{a,s}$	S_s
Suma	A_1	A_2	….	A_a	$T = \sum A_i = \sum\sum Y_{ij}$

Los cálculos pertinentes exigen calcular primero lo que llamaremos **razones básicas**, similares a las que se presentaron en el capítulo anterior, aunque ahora muestran la particularidad del tipo de diseño. Estas razones son:

- [A] o el sumatorio de las puntuaciones de cada nivel al cuadrado dividido por el número total de participantes (s). Es decir:

$$[A] = \frac{\sum_{i=1}^{a} A_i^2}{s}$$

- [S] o el sumatorio de las puntuaciones de cada sujeto al cuadrado dividido por el número de condiciones:

$$[S] = \frac{\sum_{j=1}^{s} S_j^2}{a}$$

- [AS] o el sumatorio de cada puntuación al cuadrado:

$$[AS] = \sum_{i=1}^{a} \sum_{j=1}^{s} (Y_{ij})^2$$

Obsérvese que, a diferencia de las anteriores razones, no se divide por ningún elemento ya que cada celda de la tabla (AS_{ij}) consta de una única puntuación.

- [T] o el total de los elementos de la tabla elevado al cuadrado dividido por el producto entre el del número de sujetos (s) por el número de condiciones experimentales (a).

$$[T] = \frac{\left(\sum_{i=1}^{a} \sum_{j=1}^{s} Y_{ij} \right)^2}{a \cdot s} = \frac{T^2}{N}$$

Es importante darse cuenta que hemos calculado una razón básica para cada uno de los elementos que creemos están aportando variabilidad a los datos mas la variabilidad total. Es decir, si recordamos la fórmula: $SC_{Total} = SC_A + SC_S + SC_{(A \times S)}$ vemos que hemos calculado una razón básica para cada uno de los elementos de dicha fórmula.

Además, es interesante observar que la expresión matemática de cada una de estas razones puede derivarse fácilmente observando que el numerador viene dado por el cuadrado de los elementos a los que hace referencia.

La razón básica [A] viene dada por el cuadrado de los sumatorios de cada condición experimental (A_i) y dividida por el número de elementos que han servido para calcular cada A_i, es decir, el número de sujetos.

La razón básica [S] viene dado por el cuadrado de los sumatorios de los sujetos (S_j) dividida por el número de elementos que han servido para calcular cada S_j (el número de condiciones experimentales).

La razón básica [AS] viene dada simplemente por el cuadrado de las puntuaciones directas y se divide por 1 ya que cada puntuación directa sólo integra un elemento.

Por último, la razón básica [T] viene dada por el c uadrado del sumatorio del total y dividida por el número de elementos que han servido para su cálculo. Este número de elementos viene dado por el producto entre el número de condiciones (*a*) y el número de participantes (*s*).

Una vez que disponemos de estas razones básicas, procedemos a calcular las sumas de cuadrados (es decir, los numeradores de cada una de las fuentes de variación):

$$SC_A = [A] - [T]$$

$$SC_S = [S] - [T]$$

$$SC_{(A\times S)} = [AS] - [A] - [S] + [T]$$

$$SC_{Total} = [AS] - [T]$$

A continuación hemos de calcular las medias cuadráticas (cuasivarianzas) dividiendo las sumas de cuadrados por sus correspondientes grados de libertad (g.l.) para calcular el estadístico F que nos permitirá comprobar si los resultados obtenidos son significativos, es decir, si se puede rechazar H_0 (significativo) o no. Al igual que en el Tema 5 ordenamos los cálculos en una tabla de ANOVA.

Fuentes de variación	Sumas de cuadrados	Grados de libertad	Medias cuadráticas	F
Factor (A)	SC_A	$a-1$	$MC_A = \frac{SC_A}{a-1}$	$F = \frac{MC_A}{MC_{(A\times S)}}$
Sujetos (S)	SC_S	$s-1$	$MC_S = \frac{SC_S}{s-1}$	
Error ($A \times S$)	$SC_{(A\times S)}$	$(as)-a-s+1$	$MC_{(A\times S)} = \frac{SC_{(A\times S)}}{(as)-a-s+1}$	
Total	SC_T	$(as)-1$		

Observamos que hemos utilizado como varianza de error la media cuadrática correspondiente a la interacción (A $\times$ S). En los diseños intra-sujetos la fuente de error se considera que es la producida por la interacción entre el factor y el sujeto (A $\times$ S) debido a que refleja la inconsistencia con la que los sujetos se comportan bajo los diferentes tratamientos, inconsistencia que si todos los sujetos se comportaran igual con respecto a la variable independiente, no existiría. Si no hubiese inconsistencia indicaría que todos los sujetos mostrarían el efecto del factor en la misma medida o con el mismo patrón, que es precisamente uno de los supuestos de este análisis. También resulta obvio que si $MC_{(A \times S)}$ fuese exactamente cero, no podríamos calcular la razón F del factor manipulado o de los sujetos ya que esto implicaría dividir MC_A por 0, operación prohibida en matemáticas, como muy bien sabe el lector. Afortunadamente, en la práctica es muy difícil encontrar esta situación ya que siempre existirá varianza debida a la interacción entre los sujetos y el factor, aunque sea pequeña.

Obsérvese también que no hemos calculado la razón F para el factor "S" o Sujetos. Esto es debido a que estamos interesados únicamente en el factor "condición de Stroop". En este tipo de experimentos no nos interesan las diferencias individuales. Esta información puede interesarle, en todo caso, al psicólogo diferencial ya que él estudia porqué los sujetos difieren entre sí cuando son sometidos a la misma condición experimental pero no al psicólogo que busca leyes generales.

Por último, para tomar una decisión sobre la significatividad del factor manipulado (condición de Stroop) debemos comparar la F obtenida (F muestral) con la F crítica que obtenemos de las Tablas de la razón F conociendo α (probabilidad de error tipo I que estamos dispuestos a cometer) así como los grados de libertad del numerador y del denominador de la F muestral. Si la F empírica del factor A (la que obtenemos siguiendo todos los pasos mencionados hasta ahora) es igual o superior a la F crítica entonces rechazamos H_0, lo cual significa que hay, al menos, dos medias que difieren entre sí en nuestros datos. En caso contrario (la F empírica es inferior a la F crítica) significa que no podemos rechazar la hipótesis de que, en la población, las medias de los distintos niveles del factor sean idénticas. El valor de la F crítica nos separa los valores posibles de F en dos segmentos excluyentes: a la izquierda el que define la región de aceptación de H_0 y a la derecha, la región de rechazo de H_0, tal y como puede verse en la Figura 6.1 que presenta el caso genérico (véase explicación a pie de figura).

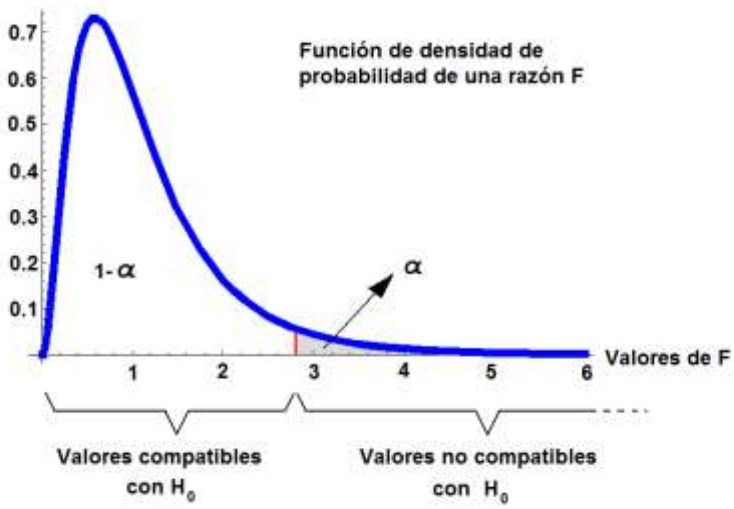


Figura 6.1: Distribución de densidad de probabilidad para una F genérica. Para poder dibujar la gráfica se ha utilizado una F con 6 y 15 grados de libertad (se han asignado unos valores concretos ya que no se puede dibujar la gráfica de una Función de Densidad de Probabilidad, o FDP, si no se concretan sus parámetros). En estas condiciones y con un α de 0.05, la F crítica vale 2.8. Este valor viene reflejado por la línea vertical que divide la función de densidad de probabilidad en dos partes, aquellos valores de F inferiores a 2.8 y que, siendo H_0 cierta, tienen una probabilidad de producirse superior a α y, por otra parte, aquellos valores de F superiores a 2.8 y que, siendo H_0 cierta, tienen una probabilidad de producirse inferior a α. En este último caso, sospechamos que H_0 es falsa y la rechazamos.

6.3.1. Análisis de datos mediante razones básicas.

Veamos con los datos de nuestro ejemplo cómo llevar a cabo el análisis de la varianza.

Condiciones y supuestos En nuestro ejemplo estos supuestos se concretan en que:

- La variable dependiente en nuestro ejemplo es el TR medio medido en segundos, que tiene un nivel de medida de razón.
- Suponemos que las puntuaciones para cada condición experimental son independientes entre sí.
- Suponemos que las poblaciones en todas las condiciones se distribuyen según la curva normal.
- Suponemos igualdad de varianzas y covarianzas.

Formulación de las hipótesis:

$$H_0: \mu_{Congruente} = \mu_{Incongruente} = \mu_{neutral}$$

$$H_1: \mu_{Congruente} \neq \mu_{Incongruente} \quad \text{y/o}$$
$$\mu_{Congruente} \neq \mu_{Neutral} \quad \text{y/o}$$
$$\mu_{Incongruente} \neq \mu_{Neutral}$$

Estadístico de contraste: En nuestro ejemplo tenemos 3 niveles ($a = 3$), seis sujetos ($s = 6$) y un número total de observaciones igual a 18 ($as = 6 \cdot 3 = 18$) ya que el diseño es equilibrado. Comenzamos calculando las sumas para todos los niveles y todos los sujetos, así como sus cuadrados.

	congruente a_1	incongruente a_2	neutral a_3	S_i	S_i^2
S_1	0,545	0,832	0,620	1,997	3,988
S_2	0,630	0,736	0,635	2,001	4,004
S_3	0,610	0,663	0,680	1,953	3,814
S_4	0,680	0,715	0,660	2,055	4,223
S_5	0,590	0,880	0,700	2,170	4,709
S_6	0,600	0,790	0,670	2,060	4,244
A_i	3,655	4,616	3,965	$\sum A_i = 12{,}236$	$\sum S_j^2 = 24{,}982$
A_i^2	13,359	21,307	15,721	$\sum A_i^2 = 50{,}388$	

Calculamos las razones básicas:

$$[A] = \frac{\sum A^2}{s} = \frac{13{,}359 + 21{,}307 + 15{,}721}{6} = \frac{50{,}388}{6} = 8{,}398$$

$$[S] = \frac{\sum S^2}{a} = \frac{3{,}988 + 4{,}004 + 3{,}814 + 4{,}223 + 4{,}709 + 4{,}244}{3} = \frac{24{,}982}{3} = 8{,}327$$

$$[T] = \frac{\left(\sum\sum Y_{ij}\right)^2}{as} = \frac{12{,}236^2}{18} = 8{,}318$$

$$[AS] = 0{,}545^2 + 0{,}630^2 + \cdots + 0{,}670^2 = 8{,}444$$

Calculamos las sumas de cuadrados:

$$SC_A = [A] - [T] = 8{,}398 - 8{,}318 = 0{,}080$$

$$SC_S = [S] - [T] = 8{,}327 - 8{,}318 = 0{,}009$$

$$SC_{(A\times S)} = [AS] - [A] - [S] + [T] = 8{,}444 - 8{,}398 - 8{,}327 + 8{,}318 = 0{,}037$$

$$SC_{Total} = [AS] - [T] = 8{,}444 - 8{,}318 = 0{,}126$$

Construimos la tabla de ANOVA:

FV	SC	g.l.	MC	F
Factor (A)	0,080	2	0,04010	10,888
Sujetos (S)	0,009	5	0,00190	
Error $(A \times S)$	0,037	10	0,00368	
Total	0,126	17	0,00744	

Regla de decisión. Consultamos el valor crítico en la tabla F de Fisher, que para 2 y 10 grados de libertad y un nivel de confianza del 95% es igual a: F = 4,103. Dado que el estadístico de contraste es superior al valor crítico (10,888 > 4,103) rechazamos la hipótesis nula. Mediante un programa informático podríamos calcular que el nivel crítico p es igual a: p = 0,0031. Por lo tanto el estadístico de contraste también supera un nivel de confianza del 99%.

Conclusión. Existen diferencias significativas en cuanto el tiempo de reacción medio entre al menos un par de medias para las condiciones Congruente, Incongruente y Neutral del efecto Stroop.

6.4.- Ejercicios de autoevaluación

Enunciados

1. El contrabalanceo de los niveles de un factor es una exigencia en los diseños: A) test-retest; B) inter-sujetos; C) intra-sujeto.

2. En un diseño de un factor intra-sujeto, los participantes: A) pasan por todos los niveles del factor; B) solamente pasan por un nivel del factor; C) no muestran efectos de aprendizaje.

3. La ventaja de utilizar un diseño intra-sujetos en relación a un diseño inter-sujetos consiste en que: A) el término error es usualmente inferior; B) el sujeto es su propio control; C) ambas opciones son correctas.

4. Una tabla(A $\times$ S) se compone de: A) la media de la variable dependiente para cada nivel del factor; B) los valores de la variable dependiente en cada nivel del factor y para cada sujeto; C) el sumatorio de la variable dependiente para cada sujeto.

5. Si disponemos de 8 sujetos y 4 condiciones experimentales, el denominador de la razón básica [T] es igual a: A) 8; B) 32; C) 4.

6. En los diseños intra-sujetos se considera que la varianza de error viene dada por: A) el factor manipulado; B) los sujetos; C) la interacción entre el factor manipulado y los sujetos.

Soluciones.

1.- C 2.- A 3.- C 4.- B 5.- B 6.- C

6.5.- Ejercicios propuestos.

1. Se realizó un estudio para comparar la productividad de tres trabajadores en cuatro máquinas de ensamblaje idénticas. Se recogieron los registros de producción para cada operario en cuatro días seleccionados al azar. Los resultados fueron:

	Día 1	Día 2	Día 3	Día 4
Operario 1	56	66	79	72
Operario 2	79	69	74	58
Operario 3	75	76	90	80

Evalúe la hipótesis de que los tres trabajadores fueron igualmente productivos (NC=95%).

2. Se realizó un experimento para determinar el efecto de cuatro productos químicos sobre la resistencia a la rotura en varias telas. Se quería obtener esta información porque estos productos químicos se utilizaban como parte normal del tratamiento para el prensado final de la tela. Si su efecto no resultaba significativo, entonces no habrían implicaciones en la manufactura del producto pero si lo resultaba, se podrían estudiar procedimientos alternativos que no afectaran a la calidad de la tela. Por ello, se seleccionaron cinco telas distintas y se realizó un experimento en orden aleatorio con cada tipo de tela y cada producto químico. Los resultados obtenidos se muestran a continuación:

Producto químico	**Tipo de tela**				
	1	**2**	**3**	**4**	**5**
1	1,3	1,6	0,5	1,2	1,1
2	2,2	2,4	0,4	2,0	1,8
3	1,8	1,7	0,6	1,5	1,3
4	3,9	4,4	2,0	4,1	3,4

Evalúe si existen diferencias entre las medias de resistencia al estiramiento en función de los productos químicos utilizados (NC=95%).

6.6.- Solución a los ejercicios propuestos.

La hipótesis en este caso no se plantea sobre los días (factor en las columnas) sino sobre los sujetos. También podemos emplear el ANOVA de medidas repetidas para comprobar si existen diferencias entre los operarios. Por consiguiente, tenemos que realizar el ANOVA usual pero en la tabla del ANOVA no nos interesa el factor A (que en este caso sería el día) sino el factor S, que es el que pondremos a prueba.

Suponiendo que se cumplen los supuestos necesarios, las hipótesis que planteamos son:

$$H_0: \mu_{Operario\ 1} = \mu_{Operario\ 2} = \mu_{Operario\ 3}$$

$$H_1: \mu_{Operario\ 1} \neq \mu_{Operario\ 2} \qquad \text{y/o}$$

$$\mu_{Operario\ 1} \neq \mu_{Operario\ 2} \qquad \text{y/o}$$

$$\mu_{Operario\ 2} \neq \mu_{Operario\ 3} \qquad \text{y/o}$$

Calculamos las sumas para todos los niveles, para todos los sujetos y los cuadrados.

	Día					
	1	**2**	**3**	**4**	S_i	S_i^2
Operario 1	56	66	79	72	273	74529
Operario 2	79	69	74	58	280	78400
Operario 3	75	76	90	80	321	103041
A_i	210	211	243	210	874	
A_i^2	44100	44521	59049	44100		

Se calculan las razones básicas:

$$[A] = \frac{\sum A^2}{s} = \frac{44100 + 44521 + 59049 + 44100}{3} = \frac{191770}{3} = 63923{,}33$$

$$[S] = \frac{\sum S^2}{a} = \frac{74529 + 78400 + 103041}{4} = \frac{255970}{4} = 63992{,}5$$

$$[T] = \frac{\left(\sum \sum Y_{ij}\right)^2}{as} = \frac{874^2}{12} = 63656{,}33$$

$$[AS] = 56^2 + 66^2 + \cdots + 80^2 = 64660$$

Calculamos las sumas de cuadrados:

$$SC_A = [A] - [T] = 63923{,}33 - 63656{,}33 = 267$$

$$SC_S = [S] - [T] = 63992{,}5 - 63656{,}33 = 336{,}17$$

$$SC_{(A\times S)} = [AS] - [A] - [S] + [T] = 64660 - 63923{,}33 - 63992{,}5 + 63656{,}33 = 400{,}5$$

$$SC_{Total} = [AS] - [T] = 64660 - 63656{,}33 = 1003{,}67$$

Construimos la tabla de ANOVA:

FV	SC	g.l.	MC	F
Factor (A)	267	3	89	1,33
Sujetos (S)	**336,17**	**2**	**168,08**	**2,52**
Error $(A \times S)$	400,50	6	66,75	
Total	1003,67	11		

El valor de la F crítica con 2 y 6 grados de libertad (ya que estamos evaluando los sujetos, no el día) a un $\alpha = 0{,}05$ vale 5,143. Debido a que la F empírica es inferior a la F crítica (2,52 < 5,143) no podemos rechazar la hipótesis nula de que los tres operarios son igual de eficientes.

2.- Suponiendo que se cumplen los supuestos necesarios, las hipótesis que se plantean son:

$$H_0: \mu_{Producto\ 1} = \mu_{Producto\ 2} = \mu_{Producto\ 3} = \mu_{Producto\ 4}$$

$H_1: \mu_i = \mu_j$ Al menos para un par de productos

Comenzamos organizando los datos como los utilizamos habitualmente para realizar los cálculos, ya que nos interesa determinar el efecto del Producto Químico utilizando el tipo de tela como "unidad de observación", sobre la que realizan 4 evaluaciones repetidas, una para cada producto químico.

	Producto químico					
Tipo de tela	**1**	**2**	**3**	**4**	S_i	S_i^2
1	1,3	2,2	1,8	3,9	9,2	84,64
2	1,6	2,4	1,7	4,4	10,1	102,01
3	0,5	0,4	0,6	2,0	3,5	12,25
4	1,2	2,0	1,5	4,1	8,8	77,44
5	1,1	1,8	1,3	3,4	7,6	57,76
$\boldsymbol{A_i}$	5,7	8,8	6,9	17,8	39,2	
$\boldsymbol{A_i^2}$	32,49	77,44	47,61	316,84		

Se calculan las razones básicas:

$$[A] = \frac{\sum A^2}{s} = \frac{32{,}49 + 77{,}44 + 47{,}61 + 316{,}84}{5} = \frac{474{,}38}{5} = 94{,}876$$

$$[S] = \frac{\sum S^2}{a} = \frac{84{,}64 + 102{,}01 + 12{,}25 + 77{,}44 + 57{,}76}{4} = \frac{334{,}1}{4} = 83{,}525$$

$$[T] = \frac{\left(\sum\sum Y_{ij}\right)^2}{as} = \frac{39{,}2^2}{20} = 76{,}832$$

$$[AS] = 1{,}3^2 + \cdots + 3{,}4^2 = 102{,}52$$

Calculamos las sumas de cuadrados:

$$SC_A = [A] - [T] = 94{,}876 - 76{,}832 = 18{,}044$$

$$SC_S = [S] - [T] = 83{,}525 - 76{,}832 = 6{,}693$$

$$SC_{(A\times S)} = [AS] - [A] - [S] + [T] = 102{,}52 - 94{,}876 - 83{,}525 + 76{,}832 = 0{,}951$$

$$SC_{Total} = [AS] - [T] = 102{,}52 - 76{,}832 = 25{,}688$$

Construimos la tabla de ANOVA:

FV	SC	g.l.	MC	F
Factor (A)	**18,044**	**3**	$\mathbf{6{,}015}$	$\mathbf{75{,}89}$
Sujetos (S)	6,693	4	1,673	
Error $(A \times S)$	0,951	12	0,079	
Total	25,688	19		

Conclusión:

El valor de la F crítica con 3 y 12 grados de libertad (ya que estamos evaluando los productos químicos) a un $\alpha = 0{,}05$ vale 3,490. Debido a que la F empírica es superior a la F crítica (75,89 > 3,490) debemos rechazar la hipótesis nula de que los cuatro productos químicos tengan el mismo efecto sobre la resistencia a la rotura de las telas sobre las que se aplica. Hay al menos dos de ellos que difieren en el efecto que producen sobre las telas.

Tema 7: DISEÑOS CON MÁS DE DOS GRUPOS INDEPENDIENTES.

ANÁLISIS DE LA VARIANZA DE DOS FACTORES

Índice

7.1 Introducción

En los dos capítulos precedentes se han estudiado dos diseños de más de dos grupos, uno con grupos completamente aleatorizados (cap. 5) y otro con grupos de medidas repetidas (cap. 6). En los primeros, las unidades de observación (v.g., sujetos) se asignan aleatoriamente a cada uno de los niveles del factor, de modo que cada sujeto sólo recibe uno de los posibles tratamientos establecidos en el diseño. En los segundos, cada sujeto pasa por todos los tratamientos del diseño. El primer tipo de diseño se conoce como un diseño completamente aleatorizado y el segundo de medidas repetidas. No obstante, aunque ambos tipos de diseños de un solo factor tienen una utilidad muy amplia, no permiten abordar cuestiones complejas que se dan muy a menudo en el mundo “real”, como son las posibles interacciones que se pueden dar cuando se manipulan varios factores a la vez y su incidencia sobre la variable objeto de estudio.

Para comprender lo anterior, supongamos que un departamento policial de una gran ciudad está interesado en mejorar la actitud de los nuevos oficiales hacia las minorías radicadas en la ciudad. Los responsables piensan que la mejora dependerá de la duración del curso que se les imparte sobre relaciones humanas, pero no descartan que también sea importante la zona de la ciudad donde se impartirá el curso. Para dar respuesta a esta inquietud, contactan con un consultor estadístico para que diseñe el experimento, de modo que se puedan tomar decisiones posteriormente a partir de los resultados. Al ser dos los factores o variables independientes (duración del curso y zona de la ciudad), el consultor elabora un diseño en el que considera la zona de la ciudad donde se va a impartir el curso y también la duración del mismo. En cada uno de los factores establece tres niveles: en el Factor A (zona) el curso se imparte o en un barrio de clase alta, o de clase media, o en un barrio económicamente deprimido; respecto de la duración del curso establece tres niveles: de 5, de 10 o de 15 horas. Una vez elaborado el diseño, realiza una selección al azar de 45 policías que van a participar en los cursos y asigna 5 a cada combinación de zona y duración. Como variable dependiente se toma la puntuación alcanzada en un test, previamente validado, sobre actitudes hacía los grupos minoritarios.

Con un diseño factorial de este tipo, no sólo se pueden alcanzar conclusiones sobre la incidencia que tenga la duración del curso o la ubicación de la oficina policial donde se imparte, que en términos de los diseños experimentales se conoce como *efectos principales del factor*, sino que, además, se pueden llegar a conclusiones sobre si la duración y la zona están relacionados de algún modo, lo cual se conoce como *efectos de interacción* entre los factores. En general, un diseño factorial es más eficiente que varios diseños simples. Es más económico, en el sentido de que proporciona más información con menor número de sujetos, esfuerzo y tiempo. En este capítulo vamos a estudiar la técnica de análisis de diseños de dos factores completamente aleatorizados, presentando las posibilidades de contrastes que estos diseños permiten.

7.2 Objetivos del tema

En este tema vamos a aprender los siguientes aspectos de un diseño factorial de 2 factores completamente aleatorizados:

- La disposición de los datos y la notación.
- Identificar los diversos efectos que están presenten en el diseño (efectos principales, de interacción y simples).
- Método de cálculo para el contraste de la significación estadística de estos efectos.
- Métodos para el cálculo de los efectos simples presentes en el diseño.
- Métodos de cálculo para las comparaciones por pares de los efectos simples significativos.

7.3 ¿Qué información proporciona un diseño factorial?

Pensemos en el diseño expuesto en la introducción. Tal como está planteado, los responsables de la formación de los oficiales de policía pueden contrastar el efecto que tiene la zona donde se imparte el curso, al margen de la duración del mismo. Para ello sólo se tendrían en cuenta los datos de los 15 oficiales asignados a cada una de las zonas (cinco por cada uno de los tres grupos de tiempo) –ver Figura 7.1-. También podrían evaluar el efecto que tiene la duración del curso, al margen de la zona, para lo cual sólo se tendrían en cuenta los datos de los 15 oficiales asignados a cada tiempo de duración de curso, con independencia de la zona en que se imparten. Con esto, estaríamos contrastando los denominados *efectos principales*, que serán tantos como factores hay implicados en el diseño.

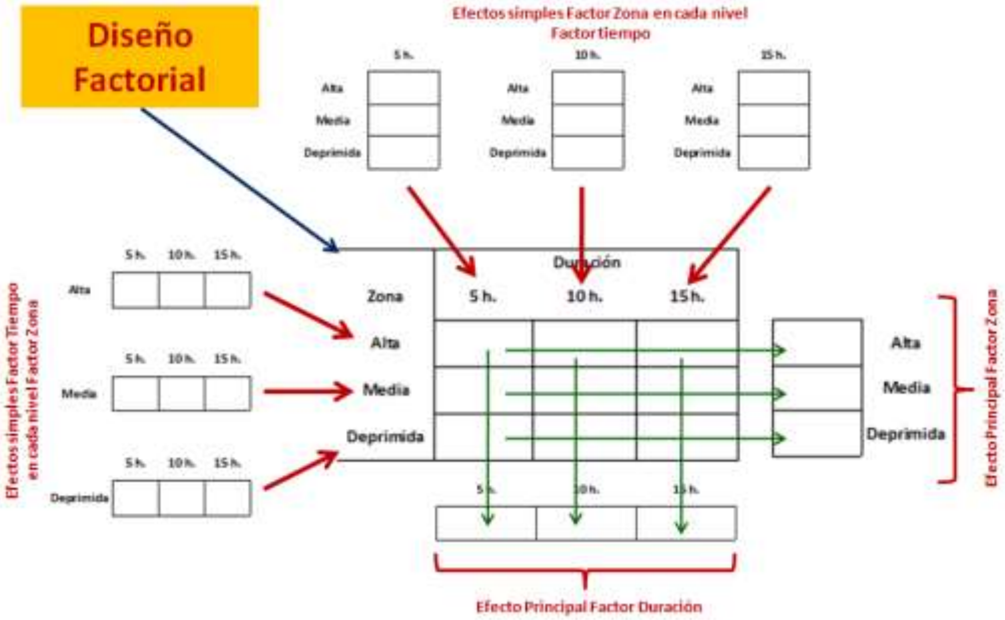


Figura 7.1 *Posibles contrastes en un diseño factorial de dos factores*

Además de estos efectos principales (tantos como factores haya), cuyos contrastes no difieren de los estudiados en el capítulo 5, hay un conjunto de contrastes más focalizados de cada factor con cada nivel del otro factor, que representan los llamados *efectos simples*, en el sentido de que se contrastan los tratamientos de un factor en cada nivel del otro factor. Aplicado al ejemplo de la introducción, un efecto simple es el de la zona en que se imparte el curso cuando esté sólo dura 5 horas. Otro efecto simple es el del tiempo de duración del curso cuando éste se imparte en una zona de clase media. En total, hay tantos efectos simples como la suma de los niveles de cada factor (en nuestro caso 3+3 = 6). En cada uno de estos contrastes sólo entran en juego los sujetos asignados aleatoriamente a cada una de los tres niveles de un factor condicionado a un nivel del otro factor. Para el caso del ejemplo de la introducción, cada nivel estaría compuesto por 5 oficiales.

Un tercer y último efecto es el que se produce por los cruces de los tratamientos (o niveles) de cada uno de los factores, que está relacionado con los denominados efectos simples y que se conoce como *efecto de interacción*. Aunque luego lo veremos con datos, imaginemos, en el ejemplo de las actitudes ante las minorías por parte de los oficiales policiales, que para el caso de la zona alta, la actitud de los oficiales sube conforme aumenta la duración del curso, observándose para otras zonas un patrón de respuesta similar, aumentando de forma parecida las puntuaciones en la variable dependiente a medida que aumenta la duración del curso. En este caso, las líneas del gráfico de medias de la VD respecto de la duración, para cada una de las zonas, serían más o menos paralelas. Por el contrario, si se observaran comportamientos diferentes de la VD respecto de la duración en función de la zona, las líneas del gráfico de medias tenderían a cruzarse, o se cruzarían de hecho, en algún lugar del plano, tal como de hecho sucede con los datos del ejemplo que veremos en el próximo epígrafe. En este último caso, diríamos que se ha producido una interacción (que habría que confirmar analíticamente como veremos más adelante).

La presencia de interacción en un diseño factorial obliga a ir más allá de las conclusiones que se sacan a partir de los efectos principales, pues éstos, para cada factor, deben ser interpretados teniendo en cuenta los niveles del otro factor. Hay muchas definiciones de interacción, todas ellas equivalentes pero con énfasis en distintos aspectos de la misma. En el Cuadro 7.1 hacemos una relación de éstas para hacerse una idea de la importancia que se le concede a este concepto.

Cuadro 7.1 *Definiciones sobre el concepto de interacción en un diseño factorial*

- Una interacción está presente cuando los patrones de diferencias asociados con una variable independiente cambian con los diferentes niveles de la otra variable independiente. Para entender esto, supongamos que en un diseño experimental tenemos que uno de los factores está compuesto por tres niveles compuestos por un grupo control y dos tratamientos experimentales. Es posible que consideremos realizar comparaciones entre los dos tratamientos experimentales y entre el control y cada uno de los tratamientos, o entre el control y una combinación de los tratamientos. Si está presente una interacción, el patrón de estas diferencias no será el mismo en cada nivel del otro factor.
- Una interacción está presente cuando los efectos simples de una variable independiente no son los mismos en todos los niveles de la otra variable independiente.
- Una interacción está presente cuando los efectos principales de una variable independiente no son representativos de los efectos simples de esa misma variable.
- Una interacción está presente cuando las diferencias entre las medias de las celdas que representan el efecto de un factor en algún nivel del otro factor no son iguales a las correspondientes diferencias en otro nivel de este factor.
- Una interacción está presente cuando los efectos de una de las variables independientes están condicionalmente relacionado con los niveles de la otra variable independiente (Cohen, 1983).
- Una interacción está presente cuando una variable independiente no tiene un efecto constante en todos los niveles de la otra variable independiente (Pedhazur, 1982).

El estudiante puede observar que hay diferencias más semánticas que conceptuales o estadísticas en estas definiciones. En la diversos gráficos de la Figura 7.2 se proponen algunos ejemplos de datos de diseños factoriales 3x3 (cada dígito indica el número de niveles del factor correspondiente), donde se muestran gráficos de medias de tratamientos, dos de ellos con interacción y uno sin interacción. Las medias de los tratamientos están representadas encima de cada gráfica.

A modo de resumen y antes de empezar con el ejemplo numérico, la notación y la exposición del modelo estadístico, veremos algunas cuestiones básicas en un diseño factorial.

- Un diseño factorial consiste en un conjunto de diseños simples de un factor en el cual la misma variable independiente es manipulada en combinación con una segunda variable independiente.
- Los efectos simples de una variable independiente se refieren a las diferencias entre las medias para cada uno de los componentes del experimento. Si tenemos un factor *A*, las diferencias

observadas en el nivel b_1, del factor B, se denominan como efectos simples del factor A en el nivel b_1.

- Las interacciones se definen en términos de las comparaciones entre un conjunto de efectos simples. Están presentes cuando se encuentra que los efectos simples asociados con una variable independiente no son los mismos en todos los niveles de la otra variable independiente.
- Los efectos principales de una variable independiente, o factor, se refieren a los efectos promedio totales de una variable, y se obtienen combinando el conjunto completo de componentes experimentales presentes en ese factor.

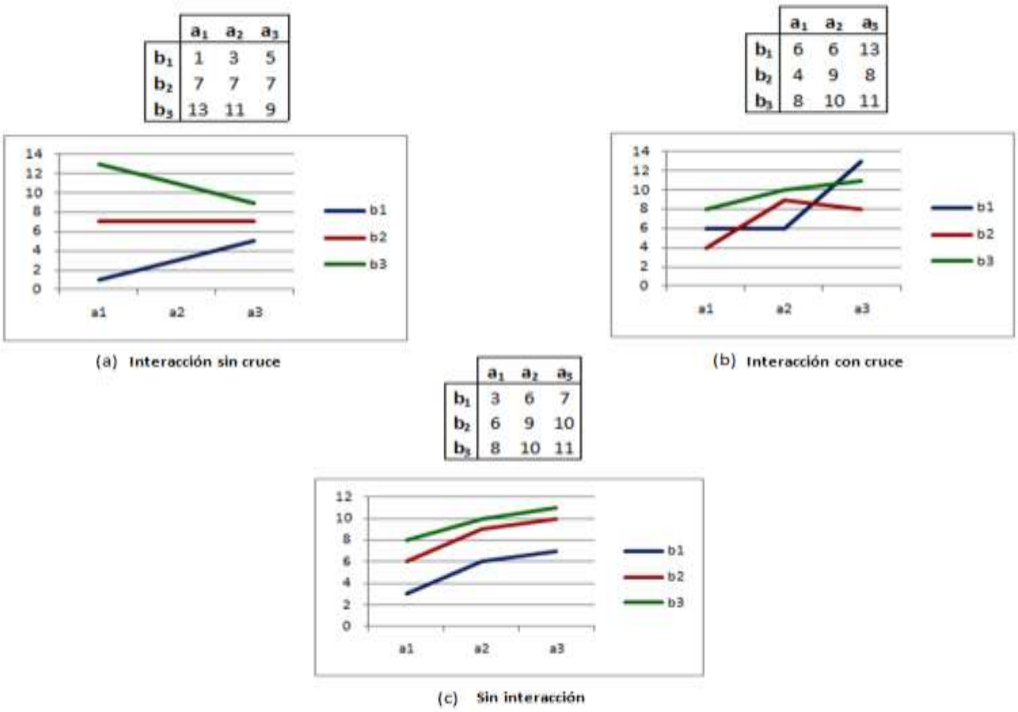


Figura 7.2 *Gráficos con interacción (a y b) y sin interacción entre dos factores (c)*

7.4 Reglas para el cálculo de los efectos principales y del efecto de interacción

En el capítulo 5 se ha explicado el procedimiento de cálculo para los efectos de los tratamientos de un factor a partir de las llamadas razones básicas, y a partir de éstas, el cálculo de las sumas de cuadrados (*SC*), las medias cuadráticas (*MC*) y el valor del estadístico *F*. También se explicó cómo la variabilidad total del sistema (la suma de cuadrados total) se puede partir en dos sumas de cuadrados independientes; por una parte, la que refleja la variabilidad entre los tratamientos (suma de cuadrados entre-grupos) y por otra la que refleja la variabilidad dentro de los tratamientos (suma de cuadrados intra-grupos). En los diseños factoriales de 2 factores (que denotaremos genéricamente como los factores *A* y *B*) se sigue un patrón de análisis similar, pero en esta ocasión la suma de cuadrados entre-grupos se divide a su vez en tres componentes que reflejan: (1) la suma de cuadrados entre tratamientos del factor A (SC_A), que refleja los efectos principales del factor *A*; (2) la suma de cuadrados entre tratamientos del factor *B* (SC_B), que refleja los efectos principales del factor *B*; y (3) la suma de cuadrados que representa la interacción entre *A* y *B* ($SC_{A\times B}$). En este capítulo sólo consideraremos los diseños equilibrados, es decir aquellos que tienen el mismo número de casos o sujetos en cada cruce de tratamientos.

7.4.1 Diseño y notación

Habitualmente, los datos en un diseño factorial de dos factores se presentarán en forma de columnas con las dos primeras filas representando las diferentes combinaciones de los niveles de los factores (Figura 7.3a). Para los datos del ejemplo seguido en el texto, tenemos un diseño factorial de 3x3 (al tener 3 niveles cada factor), siendo el número de cruces de tratamientos entre los factores *A* y *B* igual a 9 (producto de los niveles de los factores: 3 x 3 = 9).

Una observación genérica se representa como Y_{ijk}, siendo i el nivel genérico del factor *A*, j el nivel genérico del factor *B*, y k la observación genérica dentro de la combinación de tratamientos AB_{ij}. A partir de la matriz de datos inicial (Figura 7.3a) se obtiene la matriz *AB* (Figura 7.3b) en la que se calcula la suma para cada combinación de tratamientos de los dos factores (AB_{ij}), las sumas marginales A_i y B_j para los factores A y B respectivamente, así como la suma de todas las observaciones representada por T .

Matriz de Datos

Combinaciones de tratamientos

a_1	a_1	a_1	a_2	a_2	a_2	a_3	a_3	a_3
b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3
$Y_{1,1,1}$	$Y_{1,2,1}$	$Y_{1,3,1}$	$Y_{2,1,1}$	$Y_{2,2,1}$	$Y_{2,3,1}$	$Y_{3,1,1}$	$Y_{3,2,1}$	$Y_{3,3,1}$
$Y_{1,1,2}$	$Y_{1,2,2}$	$Y_{1,3,2}$	$Y_{2,1,2}$	$Y_{2,2,2}$	$Y_{2,3,2}$	$Y_{3,1,2}$	$Y_{3,2,2}$	$Y_{3,3,2}$
$Y_{1,1,3}$	$Y_{1,2,3}$	$Y_{1,3,3}$	$Y_{2,1,3}$	$Y_{2,2,3}$	$Y_{2,3,3}$	$Y_{3,1,3}$	$Y_{3,2,3}$	$Y_{3,3,3}$
$Y_{1,1,4}$	$Y_{1,2,4}$	$Y_{1,3,4}$	$Y_{2,1,4}$	$Y_{2,2,4}$	$Y_{2,3,4}$	$Y_{3,1,4}$	$Y_{3,2,4}$	$Y_{3,3,4}$
$Y_{1,1,5}$	$Y_{1,2,5}$	$Y_{1,3,5}$	$Y_{2,1,5}$	$Y_{2,2,5}$	$Y_{2,3,5}$	$Y_{3,1,5}$	$Y_{3,2,5}$	$Y_{3,3,5}$
$AB_{1,1}$	$AB_{1,2}$	$AB_{1,3}$	$AB_{2,1}$	$AB_{2,2}$	$AB_{2,3}$	$AB_{3,1}$	$AB_{3,2}$	$AB_{3,3}$

Figura 7.3(a) *Disposición inicial de los datos en un diseño factorial*

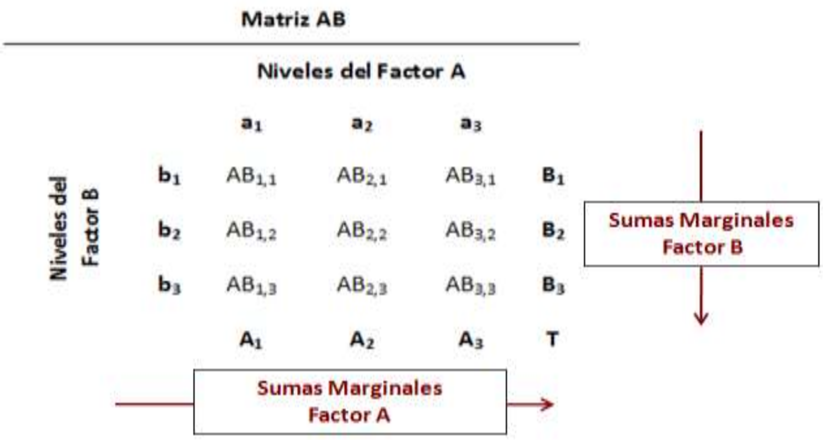


Figura 7.3(b) *Disposición de los datos en un diseño factorial para su ulterior análisis*

A partir de los sumatorios representados en la Figura 7.3(b), se pueden obtener fácilmente la media total, las medias para los niveles de cada factor, y las combinaciones de niveles de ambos factores:

$$\bar{Y}_T = \frac{T}{abn} \qquad \bar{Y}_{A_i} = \frac{A_i}{bn} \qquad \bar{Y}_{B_j} = \frac{B_j}{an} \qquad \bar{Y}_{AB_{ij}} = \frac{AB_{ij}}{n}$$

7.4.2 Variabilidad del sistema

En el capítulo 5 ya hemos tenido ocasión de comprobar que la suma de cuadrados total es igual a la suma de cuadrados entre-grupos (o tratamientos, o inter-grupo) y la suma de cuadrados intra-grupo:

$$SC_{Total} = SC_{Inter} + SC_{Intra} \qquad \text{Ecuación 7.1}$$

Esta igualdad se mantiene en un diseño factorial aunque con algunas consideraciones sobre los componentes en los que está basada ahora la suma de cuadrados entre-grupos (SC_{Inter}).

En un diseño de un único factor, SC_{Inter} está basada en las desviaciones de las medias de cada tratamiento respecto a la media total, es decir, siendo *A* el factor, y prescindiendo de los sumatorios propios en el cálculo de las sumas de cuadrados, SC_{Inter} está basada en las desviaciones $\bar{Y}_{A_i} - \bar{Y}_T$. Sin embargo, en un diseño factorial de dos factores, SC_{Inter} está basada en las desviaciones de las medias de cada tratamiento conjunto ***AB***, respecto de la media total, es decir, $\bar{Y}_{AB_{ij}} - \bar{Y}_T$.

Si pensamos en un grupo de sujetos que reciben una combinación de tratamientos *A* y *B*, la desviación respecto de la media total del sistema puede estar influida por tres componentes: el Factor *A*, el Factor *B*, y la interacción entre *A* y *B*. A su vez, cada una de estas influencias puede ser expresada de la siguiente forma:

$$\bar{Y}_{AB_{ij}} - \bar{Y}_T = \left(\bar{Y}_{A_i} - \bar{Y}_T\right) + \left(\bar{Y}_{B_j} - \bar{Y}_T\right) + \left(\bar{Y}_{AB_{ij}} - \bar{Y}_{A_i} - \bar{Y}_{B_j} + \bar{Y}_T\right) \qquad \text{Ecuación 7.2}$$

A partir de aquí, y con leves transformaciones algebraicas, se concluye que la desviación de cualquier puntuación individual respecto de la media total del conjunto de datos del diseño se puede dividir en cuatro componentes de desviación

$$Y_{ijk} - \bar{Y}_T = \left(\bar{Y}_{A_i} - \bar{Y}_T\right) + \left(\bar{Y}_{B_j} - \bar{Y}_T\right) + \left(\bar{Y}_{AB_{ij}} - \bar{Y}_{A_i} - \bar{Y}_{B_j} + \bar{Y}_T\right) + \left(Y_{ijk} - \bar{Y}_{AB_{ij}}\right) \qquad \text{Ecuación 7.3}$$

El primer componente es el relacionado con los efectos del tratamiento A_i, el segundo es el relacionado con los efectos del tratamiento B_j, el tercero es el efecto de la interacción entre ambos factores, y el cuarto es la desviación de la puntuación del sujeto respecto de la combinación de tratamientos que le ha sido asignada aleatoriamente. Tenemos por lo tanto, cuatro fuentes de variabilidad, tres de ellas relacionadas con los tratamientos o combinación de tratamientos, y una relacionada con su propio grupo que sería únicamente la expresión del error experimental, fuentes que denominaremos, *A*, *B*, *AB* y *S/AB,* respectivamente. A partir de aquí, solo quedarían por realizar los sumatorios y elevar al cuadrado para obtener las correspondientes sumas de cuadrados para el análisis de varianza.

Tenemos ahora que determinar los grados de libertad para cada una de estas fuentes. Para los efectos principales se sigue la misma regla del número de tratamientos menos *1*. Por tanto:

$$gl_A = a - 1 \qquad y \qquad gl_B = b - 1 \qquad \text{Ecuación 7.4}$$

Los grados de libertad de la interacción es el resultado del producto de los grados de libertad asociados con los factores A y B.

$$gl_{AB} = (a-1)(b-1) \qquad \text{Ecuación 7.5}$$

Los grados de libertad debidos al error experimental, es decir, a las diferencias individuales dentro de cada tratamiento, son:

$$gl_{S/AB} = ab(n-1) \qquad \text{Ecuación 7.6}$$

7.4.3 Proceso de cálculo del ANOVA de dos factores

Se han de calcular las razones básicas (Tabla 7.1) para realizar el cálculo posterior de las sumas de cuadrados (Tabla 7.2).

Tabla 7.1. *Cálculo de las razones básicas*

Cantidad Básica	Sumas y cuadrados implicados	Razón básica	Código
A	$\sum A^2$	$\frac{\sum A^2}{bn}$	$[A]$
B	$\sum B^2$	$\frac{\sum B^2}{an}$	$[B]$
AB	$\sum\sum(AB)^2$	$\frac{\sum\sum(AB)^2}{n}$	$[AB]$
Y	$\sum\sum\sum Y^2$	$\sum\sum\sum Y^2$	$[Y]$
T	T^2	$\frac{T^2}{abn}$	$[T]$

A partir de estas razones, obtener las sumas de cuadrados es sencillo. En la Tabla 7.2, se puede ver la tabla resumen del ANOVA con las fórmulas de cómputo a partir de las razones básicas. Los totales para SC_A son las sumas marginales del factor *A*, los totales para SC_B son las sumas marginales del factor *B*, ambos representados en la matriz *AB* de la figura 7.3(b), y las sumas para la razón [*AB*] se obtienen directamente de las sumas de la matriz de datos representadas en la misma figura.

Tabla 7.2 *Fórmulas de cálculo y tabla del ANOVA de dos factores*

Fuente de Variación	Sumas de Cuadrados	grados de libertad	Medias Cuadráticas	F
A	$SC_A = [A] - [T]$	$a - 1$	$MC_A = \frac{SC_A}{gl_A}$	$\frac{MC_A}{MC_{S/AB}}$
B	$SC_B = [B] - [T]$	$b - 1$	$MC_B = \frac{SC_B}{gl_B}$	$\frac{MC_B}{MC_{S/AB}}$
A x B	$SC_{AB} = [AB] - [A] - [B] + [T]$	$(a - 1)(b - 1)$	$MC_{AB} = \frac{SC_{AB}}{gl_{AB}}$	$\frac{MC_{AB}}{MC_{S/AB}}$
S/AB (Intra)	$SC_{Intra} = [Y] - [AB]$	$ab(n-1)$	$MC_{S/AB} = \frac{SC_{S/AB}}{gl_{S/AB}}$	
Total	$SC_T = [Y] - [T]$	$abn - 1$		

7.4.4 Desarrollo del ANOVA de 2 factores con un ejemplo numérico

Vamos ahora a desarrollar la técnica con los datos del ejemplo referido en la introducción sobre mejora de las actitudes hacia las minorías por parte de jóvenes oficiales de policía. En total hay 45 observaciones repartidas entre las 9 combinaciones de tratamientos, que son los que se producen por el cruce de los 3 tratamientos de cada factor. Llamaremos factor *A* a las zonas donde se realiza el curso, con tres niveles, *alta*, *media* y *deprimida* (*a=3*) y factor *B* a la duración del curso, también con tres niveles, *5*, *10* y *15 horas* (*b=3*). Los datos obtenidos son:

a_1 b_1	a_1 b_2	a_1 b_3	a_2 b_1	a_2 b_2	a_2 b_3	a_3 b_1	a_3 b_2	a_3 b_3
24	44	38	30	35	26	21	41	42
33	36	29	21	35	32	18	39	52
37	25	28	39	27	36	10	50	53
29	27	47	26	31	46	31	36	49
42	43	48	34	22	45	20	34	64

Condiciones y supuestos. Al igual que en el análisis de la varianza para un factor visto en el tema 5, el nivel de medida de la variable dependiente ha de ser de intervalo o razón. Suponemos que las observaciones son independientes, que en la población las distribuciones son normales para todos los grupos y que las varianzas son homogéneas (homocedasticidad).

Hipótesis. Al trabajar con dos factores se plantean sobre el factor *A*, el factor *B* y la interacción:

Para el factor A: $H_0: \mu_{A_1} = \mu_{A_2} = \mu_{A_3}$

$H_1: \mu_{A_1} \neq \mu_{A_2} \neq \mu_{A_3}$ al menos para un par de tratamientos

Para el factor B: $H_0: \mu_{B_1} = \mu_{B_2} = \mu_{B_3}$

$H_1: \mu_{B_1} \neq \mu_{B_2} \neq \mu_{B_3}$ al menos para un par de tratamientos

Para la interacción: $H_0: No\ existe\ interacción$
$H_1: Existe\ interacción$

Estadístico de contraste y distribución muestral. Se han de calcular tres estadísticos de contraste, uno para cada una de las hipótesis nulas. En la Tabla 7.3 se muestran las sumas totales de las puntuaciones de cada tratamiento. A continuación se calcula la matriz *AB* donde se presentan los resultados de las sumas para cada combinación de tratamientos, las sumas marginales y la suma total.

Tabla 7.3. *Datos del ejemplo numérico*

	a1 b_1	**a1 b_2**	**a1 b_3**	**a2 b_1**	**a2 b_2**	**a2 b_3**	**a3 b_1**	**a3 b_2**	**a3 b_3**
	24	44	38	30	35	26	21	41	42
	33	36	29	21	35	32	18	39	52
	37	25	28	39	27	36	10	50	53
	29	27	47	26	31	46	31	36	49
	42	43	48	34	22	45	20	34	64
Suma	165	175	190	150	150	185	100	200	260
Media	33	35	38	30	30	37	20	40	52
Desv.Típic	6,229	7,874	8,509	6,229	4,980	7,642	6,723	5,550	7,127

Matriz AB

	Factor A			
Factor B	a_1 (zona alta)	a_2 (zona media)	a_3 (zona deprimida)	Suma
b_1 (5 h.)	165	150	100	415
b_2 (10 h.)	175	150	200	525
b_3 (15 h.)	190	185	260	635
Suma	530	485	560	1575

Los resultados de las razones básicas a partir de estos datos son:

$$[\boldsymbol{Y}] = 24^2 + 33^2 + 37^2 + \cdots + 53^2 + 49^2 + 64^2 = \mathbf{60265} \qquad [\boldsymbol{T}] = \frac{1575^2}{3 \cdot 3 \cdot 5} = \mathbf{55125}$$

$$[\boldsymbol{A}] = \frac{530^2 + 485^2 + 560^2}{3 \cdot 5} = \mathbf{55315} \qquad [\boldsymbol{B}] = \frac{415^2 + 525^2 + 635^2}{3 \cdot 5} = \mathbf{56738{,}333}$$

$$[\boldsymbol{AB}] = \frac{165^2 + 150^2 + \cdots + 185^2 + 260^2}{5} = \mathbf{58155}$$

A partir de estos valores se calculan las sumas de cuadrados:

$$\boldsymbol{SC_A} = [A] - [T] = 55315 - 55125 = \mathbf{190} \qquad \boldsymbol{SC_B} = [B] - [T] = 56738{,}333 - 55125 = \mathbf{1613{,}333}$$

$$\boldsymbol{SC_{AB}} = [AB] - [A] - [B] + [T] = 58155 - 55315 - 56738{,}333 + 55125 = \mathbf{1226{,}667}$$

$$\boldsymbol{SC_{S/AB}} = [Y] - [AB] = 60265 - 58155 = \mathbf{2110} \qquad \boldsymbol{SC_T} = [Y] - [T] = 60265 - 55125 = \mathbf{5140}$$

Se comprueba que $SC_T = SC_A + SC_B + SC_{AB} + SC_{S/AB}$.

A continuación se presenta en la Tabla 7.4, el resumen del ANOVA con todos los resultados:

Tabla 7.4. *Resumen del Análisis de Varianza*

Fuente de Variación	Sumas de Cuadrados	grados de libertad	Medias Cuadráticas	F
A	190	2	95	1,621
B	1613,333	2	806,667	13,763*
A x B	1226,667	$2 \cdot 2 = 4$	306,667	5,232*
Intra (S/AB)	2110	$3 \cdot 3 \cdot (5-1) = 36$	58,611	
Total	5140			

La distribución muestral para todos de los estadísticos de contraste (columna F) sigue el modelo *F* de Fisher con sus correspondientes grados de libertad. En este caso para los factores *A* y *B* son 2 y 36 grados de libertad y para la interacción 4 y 36 grados de libertad.

Regla de decisión. Mediante un programa informático adecuado (con las tablas se tomarían los grados de libertad más próximos), podemos comprobar que los valores críticos son, al nivel de confianza del 95%: $F_{0,95;2;36} = 3{,}2595$ Para los factores *A* y *B* y $F_{0,95;4;36} = 2{,}6335$ para la interacción.

Observamos que rechazamos la hipótesis nula para el factor *B* (13,763 > 3,2595) y para la interacción (5,232 > 2,6335), lo que queda reflejado en la Tabla 7.4 con un asterisco para los valores *F* correspondientes, pero mantenemos la hipótesis nula para el factor *A* (1,621 < 3,2595).

Conclusión. Los efectos principales del factor *B* (duración del curso) y de la interacción son significativos. Conviene, no obstante, representar gráficamente las medias de los tratamientos cada uno de los factores, para examinar los efectos principales, y también las medias de los tratamientos para interpretar el sentido de la interacción a través de la representación de los efectos simples.

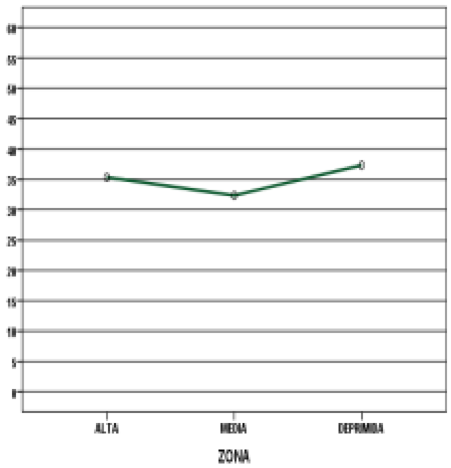


Figura 7.4 (a) *Medias de los efectos principales del factor Zona*

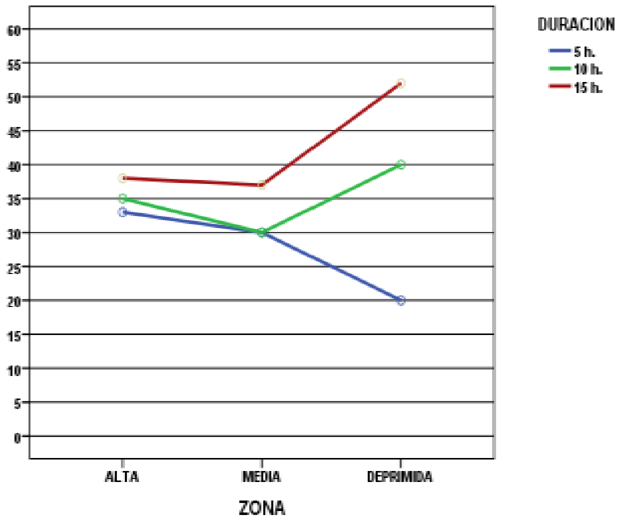


Figura 7.4 (b) *Medias de los efectos simples del factor Zona para los niveles del factor Duración del curso*

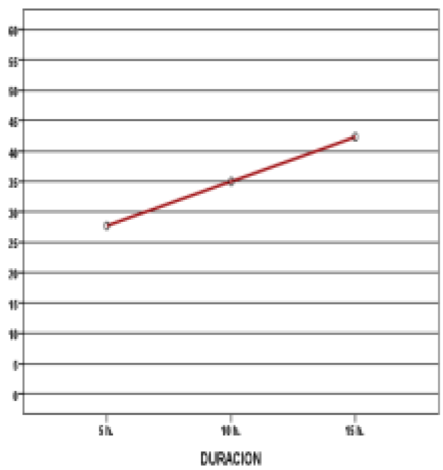


Figura 7.4 (c) *Medias de los efectos principales del factor Duración del curso*

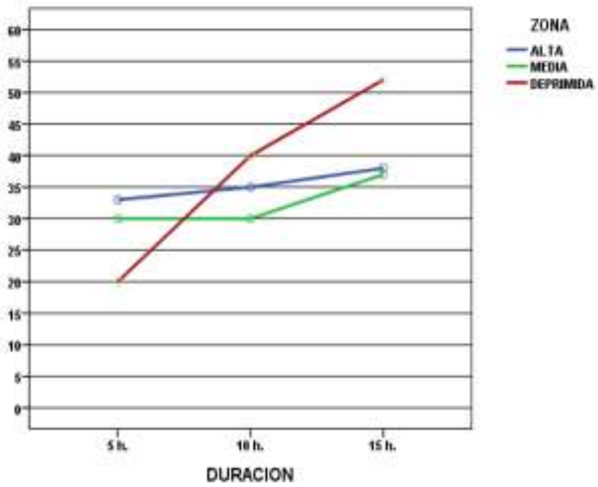


Figura 7.4 (d) *Medias de los efectos simples del factor Duración del curso para los niveles del factor Zona*

A partir de los gráficos es posible anticipar, antes de proceder al análisis de las comparaciones, que la interacción resulta significativa (en todos ellos se ha puesto la misma escala de valores en el eje de ordenadas, de 0 a 60, para facilitar la comparación).

Observe el estudiante los gráficos 7.4 (a) y 7.4 (b), que representan, el primero, los efectos principales del factor Zona, y el segundo, los efectos simples del factor Zona en función de los tres niveles del factor Duración. Como demuestra la tabla del ANOVA, los tratamientos del efecto principal "Zona" no resultan ser estadísticamente diferentes, pero la Figura 7.4(b) muestra que esta ausencia de diferencias no parece darse en todos los niveles del otro factor. Quizás no haya diferencias cuando la duración del curso es 10 horas, pero sí cuando la duración del curso es de 5 o 15 horas.

En el caso del factor Duración (Figuras 7.4c y 7.4d) es posible que haya también un efecto de interacción. Si bien los resultados no parecen diferir para las zonas alta y media, se observa que para la zona deprimida las puntuaciones en la variable dependiente aumentan al incrementarse la duración del curso.

Posteriormente veremos si estas conjeturas que hemos realizado sobre las gráficas tienen respaldo estadístico al realizar los contrastes sobre los efectos simples.

7.5 El modelo estadístico

El modelo estadístico que subyace en un diseño factorial de 2 factores completamente aleatorizados es un modelo lineal en el cual se especifican los componentes que contribuyen a explicar cualquier puntuación Y_{ijk}. Este modelo se puede expresar del siguiente modo:

$$Y_{ijk} = \mu_T + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad \text{Ecuación 7.7}$$

Donde:

- μ_T es la media total de la población
- α_i es el promedio del efecto del factor "*A*" en el nivel "*i*"
- β_j es el promedio del efecto del factor "B" en el nivel "j"
- $(\alpha\beta)_{ij}$ es el efecto de la interacción para la combinación de tratamientos "*i*" y "*j*" de los factores *A* y *B* respectivamente.
- ε_{ijk} es el error experimental asociado con cada puntuación

Por lo tanto, las hipótesis estadísticas también se pueden expresar:

Factor A	Factor B	Interacción AxB
$H_0: \alpha_i = 0$ Para todo *i* $H_1: \alpha_i \neq 0$ Al menos para algún *i*	$H_0: \beta_j = 0$ Para todo *j* $H_1: \beta_j \neq 0$ Al menos para algún *j*	$H_0: \alpha\beta_{ij} = 0$ Para todo *ij* $H_1: \alpha\beta_{ij} \neq 0$ Al menos para algún *ij*

7.6 Análisis de la interacción

La prueba anterior en donde se evaluó la significatividad de los efectos principales y la interacción entre factores se conoce como prueba *ómnibus* (global) debido a que no diferencia entre niveles, esto es, el efecto de un factor principal con tres niveles puede ser significativo según el test anterior pero éste no nos indicará si las tres comparaciones posibles entre niveles son todas ellas significativas o solo lo son un subconjunto de las mismas.

Una vez realizada la prueba ómnibus, el siguiente paso es comprobar si el efecto de interacción resulta significativo. La secuencia lógica de análisis sería la que se muestra en el diagrama de flujo de la Figura 7.5

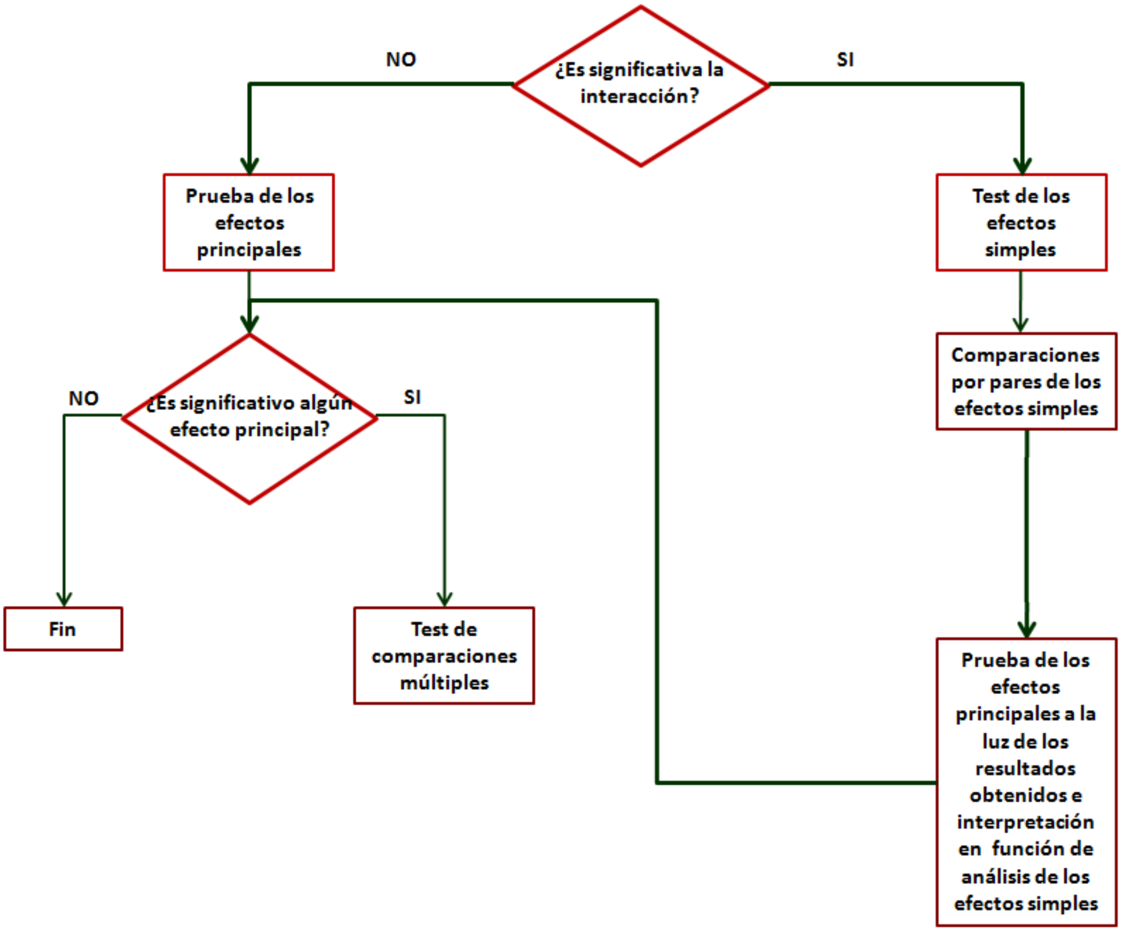


Figura 7.5 *Secuencia de análisis en un diseño factorial*

El que el efecto de interacción sea significativo, no supone que no prestemos atención a los efectos principales, sólo que su interpretación, caso de resultar significativos, debe hacerse, como señala el diagrama, teniendo en cuenta los resultados obtenidos en el análisis de los efectos simples, y sobre todo el marco teórico en el que se desarrollan las hipótesis de investigación.

Retomando nuestro ejemplo, comenzaremos por el contraste de los efectos simples. Para este análisis partimos de la matriz *AB*, que contiene las sumas de las puntuaciones de las combinaciones de tratamientos. En la Tabla 7.5 reflejamos esta matriz para los datos que nos sirven de ejemplo. La manera de enfocar este análisis es convertir cada columna o fila de la matriz *AB* en un diseño de un solo factor y luego obtener las sumas de cuadrados entre grupos del mismo modo que se ha hecho con las sumas de cuadrados de los efectos principales.

Tabla 7.5 *Matriz AB. Datos del ejemplo*

	Factor A			
Factor B	**a_1 (zona alta)**	**a_2 (zona media)**	**a_3 (zona deprimida)**	**Suma**
b_1 (5 h.)	165	150	100	415
b_2 (10 h.)	175	150	200	525
b_3 (15 h.)	190	185	260	635
Suma	530	485	560	1575

Para estudiar el factor A en la condición b_1 del factor B, se calculan las razones básicas:

$$[A\ en\ b_1] = \frac{\sum AB_{i1}^2}{n} \qquad\qquad [B_1] = \frac{B_1^2}{an}$$

Siendo la suma de cuadrados:

$$SC_{A\ en\ b_1} = [A\ en\ b_1] - [B_1], \quad \text{(con } a - 1 \text{ grados de libertad).}$$

Las medias cuadráticas para los efectos simples se obtienen de la misma manera que en el análisis general, dividiendo la suma de cuadrados por los grados de libertad. El denominador de la F es la media cuadrática intra-grupos del análisis general, es decir:

$$F = \frac{MC_{A\ en\ b_1}}{MC_{S/AB}} = \frac{\frac{SC_{A\ en\ b_1}}{a-1}}{MC_{S/AB}}$$

Del mismo modo se calcularían los efectos simples para el factor B en relación a los niveles del factor A.

Para ilustrar cómo se realizan los cálculos, por ejemplo, para el factor A en el nivel b_1:

$$SC_{A\ en\ b_1} = \frac{\sum AB_{i1}^2}{n} - \frac{B_1^2}{an} = \frac{(165)^2 + (150)^2 + (100)^2}{5} - \frac{415^2}{3 \cdot 5} = 463{,}33$$

Y para el factor B en el nivel a_2:

$$SC_{B\ en\ a_2} = \frac{\sum AB_{2j}^2}{n} - \frac{A_2^2}{bn} = \frac{(150)^2 + (150)^2 + (185)^2}{5} - \frac{485^2}{3 \cdot 5} = 163{,}33$$

Al haber tres niveles por factor, los grados de libertad serán 2 (a – 1 = b – 1 = 2) para ambas sumas de cuadrados y las medias cuadráticas serán el resultado del cociente entre las SC y sus grados de libertad. En la Tabla 7.6 se refleja el resultado del contraste para todos los efectos simples con su significación estadística señalada con un asterisco en la columna de la derecha. El estadístico de contraste, en este ejemplo, es igual en todos los casos: $F_{0,95;2;36} = 3{,}2595$ (calculado con un programa informático).

Tabla 7.6 *Tabla resumen ANOVA con los resultados de los efectos simples*

FV	SC	gl	MC	F	Sig.
SC_{AB1}	463,33	2	231,667	3,953	**
SC_{AB2}	250	2	125	2,133	
SC_{AB3}	703,33	2	351,667	6	**
SC_{BA1}	63,33	2	31,667	0,540	
SC_{BA2}	163,33	2	81,667	1,393	
SC_{BA3}	2613,33	2	1306,667	22,294	**
$SC_{S/AB}$	2110	36	58,611		

Siguiendo la secuencia del diagrama de la Figura 7.5, pasaríamos a efectuar las comparaciones por pares dentro de los niveles donde se han detectado diferencias significativas y que no veremos en este curso.

7.6.2 ¿Cómo se actúa cuando no es significativo el efecto de la interacción?

Cuando se analiza un diseño factorial, se hace la prueba ómnibus para los efectos principales y el efecto de interacción. Si ésta última no es significativa, es preciso rehacer el análisis focalizándolo sólo sobre los efectos principales. Esto tiene consecuencias en la tabla del ANOVA, en el sentido de que aumenta la suma de cuadrados del error en la misma cuantía que la suma de cuadrados de la interacción, y también los grados de libertad en la misma cuantía que los grados de libertad del efecto de interacción. En la tabla 7.7 se puede ver la tabla resumen del ANOVA con el efecto de interacción incluido (a) y la misma sin el efecto de interacción (b).

Tabla 7.7 (a) *Tabla resumen del ANOVA **con** el efecto de interacción*

FV	SC	gl	MC	F	Sig.
SC_A	190	2	95	1,62	
SC_B	1613,33	2	806,67	13,76	**
SC_{AB}	1226,67	4	306,67	5,23	**
$SC_{S/AB}$	2110	36	58,61		
SC_T	5140	44			

Tabla 7.7 (b) *Tabla resumen del ANOVA* ***sin*** *el efecto de interacción*

FV	SC	gl	MC	F	Sig.
SC_A	190	2	95	1,14	
SC_B	1613,33	2	806,67	9,67	**
$SC_{S/AB}$	3336,67	40	83,42		
SC_T	5140	44			

Cuando el efecto de interacción no forma parte del modelo, el valor de la suma de cuadrados del error es igual a la suma de cuadrados del error cuando entra la interacción y la suma de cuadrados de la interacción. Observe que 3336,67 = 2110 + 1226,67 y los grados de libertad 40 = 36 + 4. Los valores de F para los efectos principales son inferiores, por el aumento de la media cuadrática del error que pasa de 58,611 a 83,417.

7.7 Resumen

En este capítulo hemos visto el procedimiento de análisis de los diseños factoriales de 2 factores completamente aleatorizados, que son útiles por la economía de sujetos que supone y el número de contrastes que se pueden realizar con un conjunto no muy grande de datos.

Todos los diseños factoriales se pueden descomponer en un conjunto de efectos: los denominados efectos principales de cada factor y los efectos de interacción entre los factores. Los dos primeros son los efectos de los tratamientos de cada factor con independencia del otro factor, mientras los segundos son los efectos que se producen por el cruce de los factores

En el proceso de análisis, la prueba *ómnibus* informa de la significación estadística tanto de los efectos principales como de la interacción. Si los primeros son significativos, se procede a las comparaciones múltiples entre los tratamientos. Si son los segundos significativos, es preciso analizar cuáles de los diferentes efectos simples que explican este efecto.

En el procedimiento de análisis de los efectos simples es muy similar al seguido para un diseño de un solo factor completamente aleatorizado. El total de efectos simples no es independiente de los efectos principales ni de los efectos de interacción, más bien estos últimos están contenidos en el conjunto de efectos simples.

7.8 Ejercicios de autoevaluación

Supongamos "**el nivel de aspiración**" como variable dependiente en un experimento, en el que la tarea experimental consiste en un juego aparentemente difícil que implica ciertas habilidades motoras y en el que el sujeto considera que la puntuación que se le otorga depende del nivel de ejecución alcanzado. En realidad todos los sujetos reciben la misma puntuación, y se les pide que predigan la puntuación que obtendrán en el próximo grupo de ensayos. Antes de efectuar la predicción, se "informa" al sujeto sobre la puntuación obtenida comparándola con un grupo de referencia ficticio. En una condición experimental (Factor *B*), al sujeto se le dice que su puntuación está por *encima del promedio* (b_1) del grupo de referencia; en la segunda se le dice que está *en el promedio* (b_2), y en la tercera que está por *debajo del promedio* (b_3). Además, hay dos grupos de referencia con los que comparar a los sujetos (Factor *A*): el primer grupo se compone de *varones universitarios* (a_1) y el segundo es de *atletas profesionales* (a_2).

La variable dependiente es la puntuación que el sujeto predice que va a obtener en el siguiente conjunto de ensayos, que se considera como una cuantificación de su nivel de aspiración. Cada sujeto fue entrevistado por separado y no se permitía la comunicación con los otros sujetos hasta que no acabó el experimento. Se asignaron 10 sujetos a cada grupo de forma aleatoria, y las puntuaciones predichas por los sujetos fueron las siguientes (NC = 95%):

	a_1			**a_2**	
b_1	**b_2**	**b_3**	**b_1**	**b_2**	**b_3**
52	28	15	38	43	23
48	35	14	42	34	25
43	34	23	42	33	18
50	32	21	35	42	26
43	34	14	33	41	18
44	27	20	38	37	26
46	31	21	39	37	20
46	27	16	34	40	19
43	29	20	33	36	22
49	25	14	34	35	17

7.8.1 Preguntas

1. En el diseño planteado sólo una de las siguientes tres afirmaciones es cierta: A) Los dos efectos principales son significativos; B) Sólo el efecto principal del Factor B es significativo; C) Son significativos el efecto principal del factor B y el efecto de interacción
2. El resultado de la razón básica [AB] para el cálculo de las sumas de cuadrados es: A) 66228,8; B) 54789,5; C) 45791,9.
3. El valor de la razón F para contrastar el efecto principal del factor A es: A) 34,007; B) 0,358; C) 4,574.
4. El valor de La Media Cuadrática Error $\left(MC_{S/AB}\right)$ es: A) 11,911; B) 25,346; C) 4,225.
5. ¿Cuántos efectos simples hay en este diseño? A) 6; B) 5; C)4.
6. El valor de la F teórica con el que comparar el valor razón F obtenida en el contraste de la interacción, es, aproximadamente: A) 3,168; B) 4,019; C) 5,346.
7. ¿Cuál es el valor de la F observada para el contraste de la interacción? A) 34,007; B) 2,356; C) 8,254.
8. De todos los efectos simples posibles, ¿cuántos de ellos resultan ser estadísticamente significativos? A) dos; B) tres; C) todos.
9. La suma de las sumas de cuadrados de todos los efectos simples del Factor A, vale en total A) 814,4; B) 5804,27; C) 810,13.

7.8.2 Soluciones ejercicios de autoevaluación

Comencemos con algunas tablas intermedias, tanto de los datos como de las matrices AB, una con los sumatorios y otra con las medias (resaltado en negrilla estás las diferentes respuestas con resultados numéricos).

	a_1			a_2		
	b_1	**b_2**	**b_3**	**b_1**	**b_2**	**b_3**
	52	28	15	38	43	23
	48	35	14	42	34	25
	43	34	23	42	33	18
	50	32	21	35	42	26
	43	34	14	33	41	18
	44	27	20	38	37	26
	46	31	21	39	37	20
	46	27	16	34	40	19
	43	29	20	33	36	22
	49	25	14	34	35	17
Media	46,4	30,2	17,8	36,8	37,8	21,4
Suma	464	302	178	368	378	214

Matriz AB de sumas			
	a_1	**a_2**	**Suma**
b_1	464	368	832
b_2	302	378	680
b_3	178	214	392
Suma	944	960	1904

Matriz AB de Medias			
	a_1	**a_2**	**Medias**
b_1	46,400	36,800	41,600
b_2	30,200	37,800	34,000
b_3	17,800	21,400	19,600
Medias	31,467	32,000	31,733

Razones Básicas

$$[\boldsymbol{T}] = \frac{1904^2}{2 \cdot 3 \cdot 10} = \mathbf{60420,2667}$$

$$[\boldsymbol{Y}] = 52^2 + 48^2 \ldots + 22^2 + 17^2 = \mathbf{66872}$$

$$[\boldsymbol{A}] = \frac{944^2 + 960^2}{3 \cdot 10} = \mathbf{60424,5333}$$

$$[\boldsymbol{B}] = \frac{832^2 + 680^2 + 392^2}{2 \cdot 10} = \mathbf{65414,4}$$

$$[\boldsymbol{AB}] = \frac{464^2 + 368^2 + \cdots + 178^2 + 214^2}{10} = \mathbf{66228,8}$$

A partir de estas razones se construye la siguiente tabla resumen de ANOVA, en donde ya hemos incorporado los efectos simples.

FV	SC	gl	MC	$F_{empírica}$	$F_{teórica}$	Sig.
SC_A	4,267	1	4,267	0,358	4,020	
SC_B	4994,133	2	2497,067	209,642	3,168	***
SC_{AB}	810,133	2	405,067	34,007	3,168	***
SC_{AB1}	460,800	1	460,800	38,687	4,020	***
SC_{AB2}	288,800	1	288,800	24,246	4,020	***
SC_{AB3}	64,800	1	64,800	5,440	4,020	***
SC_{BA1}	4113,867	2	2056,933	172,690	3,168	***
SC_{BA2}	1690,400	2	845,200	70,959	3,168	***
$SC_{S/AB}$	643,200	54	11,911			
SC_T	6451,733	59				

*** Estadísticamente significativos con probabilidad p < 0,05

El cálculo para, por ejemplo, la suma de cuadrados de uno de los efectos principales sería:

$$SC_{AB1} = \frac{464^2 + 368^2}{10} - \frac{832^2}{2 \cdot 10} = \mathbf{460,800}$$

Siendo las sumas de cuadrados de todos los efectos simples del Factor A:

$$460{,}80 + 288{,}80 + 64{,}80 = 814{,}4$$

Los valores F teóricos han sido calculados con un programa informático. Si sólo manejamos las tablas, por ejemplo, para la interacción, buscaríamos con 2 y 60 grados de libertad (los más próximos a 2 y 54) obteniendo un valor igual a: 3,150.

7.8.3 Respuestas

1	2	3	4	5	6	7	8	9
C	A	B	A	B	A	A	C	A

TEMA 8. ANÁLISIS DE REGRESIÓN

8.1.- Introducción

Este capítulo trata sobre análisis de correlación y regresión, procedimiento que puede ser usado siempre que una variable dependiente cuantitativa pueda ser expresada como función de una variable, o de una combinación de variables independientes. El primer caso se conoce como Análisis de Regresión Simple (ARS) y el segundo como Análisis de Regresión Múltiple (ARM).

La manera en la que se relacionan la VI y la VD puede ser muy diversa. En el caso del ARS pueden darse relaciones lineales, exponenciales, potenciales, polinómicas, etc. En este texto únicamente se tratarán las relaciones de carácter lineal, es decir, aquellas en las que la VD se puede expresar genéricamente de la siguiente forma: $Y_i' = B_0 + BX_i$. Para el ARM sólo estudiaremos el caso en el que la VD se puede expresar como una combinación lineal de dos variables independientes.

El análisis de regresión, si bien es una técnica de análisis de datos idónea para los diseños ex post facto[1], también se puede aplicar a situaciones en las que se manipulan condiciones experimentales. Por tanto, las variables independientes pueden tener una ocurrencia natural (sexo, Cociente Intelectual, tiempo que se tarda en aprender una lista de palabras, introversión, ansiedad, etc.), o pueden ser variables manipuladas en un laboratorio. En resumen, "*casi cualquier información que tenga interés para el estudio de la VD puede ser objeto de incorporación en este tipo de análisis*"[2].

El punto 8.3 trata sobre ARS. Se utiliza un ejemplo para el desarrollo de este apartado, comenzando por recordar el procedimiento de cálculo para estudiar la relación lineal entre dos variables y los coeficientes de la recta de regresión. A continuación se repasa la interpretación de los coeficientes de regresión y del coeficiente de determinación (r_{xy}^2). Todas estas cuestiones fueron tratadas en la asignatura Introducción al Análisis de Datos de primer curso. Los contenidos específicos de Diseños de Investigación y Análisis de Datos se verán en el punto 8.3.3.

El apartado 8.4 se dedica al ARM con dos variables independientes, donde mediante un ejemplo, se estudiarán las ecuaciones de regresión lineal múltiple, el ajuste del modelo de regresión lineal múltiple y los coeficientes de correlación semiparcial y parcial.

8.2.- Objetivos

- Elaborar un **modelo** de regresión simple, e interpretar los coeficientes del mismo.
- Determinar si el modelo es suficientemente explicativo (bondad de ajuste).
- Realizar inferencias sobre el coeficiente de correlación y los coeficientes de la recta de regresión.
- Elaborar un modelo de regresión lineal múltiple con dos variables predictoras.
- Calcular la bondad del modelo de regresión múltiple y la correlación de dos variables cuando se excluye el influjo de otras variables

[1] Como se explica en la asignatura Fundamentos de Investigación, "los diseños ex post facto se caracterizan porque el investigador no puede manipular intencionalmente la variable independiente, ni asignar aleatoriamente a los participantes a los diferentes niveles de la misma … en estos diseños, el investigador selecciona a los sujetos en función de que posean o no determinadas características"

[2] Cohen, J, Cohen, P. , West, S. G.y Aiken, L. S. *Applied Multiple Regression/Correlation. Analysis for the Behavorial Sciences*. 3ª Ed. Lawrence Erlbaum Assoc. N, Jersey, 2003.

UNED

8.3.- Análisis de Regresión Simple

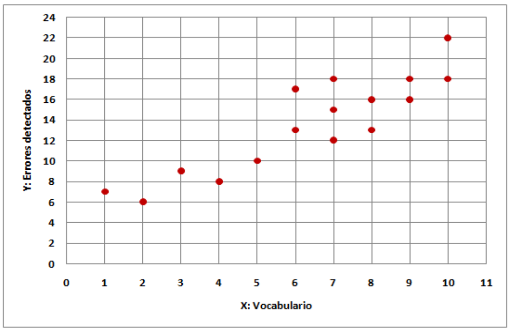
En la asignatura de primer curso, Introducción al Análisis de Datos en Psicología (Tema 4), se estudiaron índices para cuantificar la relación lineal existente entre *X* (VI) e *Y* (VD), así como el cálculo e interpretación de los coeficientes de la recta de regresión. Recordemos cómo calcular todos estos índices mediante un ejemplo.

Supongamos (Tabla 8.1) que se dispone de las puntuaciones de 16 escolares en dos variables: una prueba de vocabulario (variable *X*) y el número de errores ortográficos detectados dentro de un texto (variable *Y*). En la Figura 8.1, se obtiene la representación gráfica de todos los pares de valores en ambas variables, lo que nos sirve para observar si existe una tendencia de carácter lineal entre *X* e *Y*. En caso afirmativo, tendrá sentido calcular la ecuación de regresión lineal simple para predecir el número de errores ortográficos.

Tabla 8.1. *Datos de 16 escolares en una prueba de vocabulario (X) y número de errores ortográficos detectados en un texto (Y)*

Sujeto	X	Y	Sujeto	X	Y
1	3	9	9	10	22
2	1	7	10	2	6
3	7	12	11	5	10
4	9	18	**12**	**7**	**18**
5	10	18	**13**	**9**	**16**
6	8	13	14	6	13
7	4	8	15	7	15
8	6	17	16	8	16

Figura 8.1 *Diagrama de dispersión de los datos de la tabla 8.1*



Al inspeccionar el diagrama de dispersión (Figura 8.1) se observa que hay una tendencia lineal y positiva. A medida que un escolar puntúa más alto en la prueba de vocabulario (*X*), también **suele** detectar más errores ortográficos (*Y*). Obviamente es una **tendencia**, porque no se cumple estrictamente para todos los sujetos. Por ejemplo, el sujeto nº 12 presenta una puntuación en *X* inferior a la del sujeto nº 13, pero detecta más errores (*Y*) que este último. Aun así, se aprecia que la tendencia global de los datos es claramente directa o positiva, y aproximadamente lineal.

A continuación se obtienen las sumas de *X* e *Y*, de los cuadrados de ambas variables y del producto *XY* (Tabla 8.2). Con estas sumas se calculan medidas de tendencia central (medias) y variabilidad (varianzas y cuasivarianzas) para cada variable, así como medidas de relación lineal entre ambas (covarianza y coeficiente de correlación de Pearson) y los coeficientes de la recta de regresión de *Y* sobre *X*.

Tabla 8.2

Desarrollo para el cálculo del coeficiente de correlación de Pearson y ecuaciones de regresión

Sujetos	X	Y	XY	X²	Y²
1	3	9	27	9	81
2	1	7	7	1	49
3	7	12	84	49	144
4	9	18	162	81	324
5	10	18	180	100	324
6	8	13	104	64	169
7	4	8	32	16	64
8	6	17	102	36	289
9	10	22	220	100	484
10	2	6	12	4	36
11	5	10	50	25	100
12	7	18	126	49	324
13	9	16	144	81	256
14	6	13	78	36	169
15	7	15	105	49	225
16	8	16	128	64	256
Suma	**102**	**218**	**1561**	**764**	**3294**

Medias:

$$\bar{X} = \frac{102}{16} = 6{,}375 \qquad \bar{Y} = \frac{218}{16} = 13{,}625$$

Varianzas y desviaciones típicas:

$$S_x^2 = \frac{\sum X^2}{n} - \bar{X}^2 = \frac{764}{16} - 6{,}375^2 = 7{,}109 \rightarrow S_x = 2{,}666$$

$$S_y^2 = \frac{\sum Y^2}{n} - \bar{Y}^2 = \frac{3294}{16} - 13{,}625^2 = 20{,}234 \rightarrow S_y = 4{,}498$$

Cuasivarianzas y cuasidesviaciones típicas:

$$\hat{S}_x^2 = \frac{nS_x^2}{n-1} = \frac{16 \cdot 7{,}109}{15} = 7{,}583 \rightarrow \hat{S}_x = 2{,}754$$

$$\hat{S}_y^2 = \frac{nS_y^2}{n-1} = \frac{16 \cdot 20{,}234}{15} = 21{,}583 \rightarrow \hat{S}_y = 4{,}646$$

Covarianza y coeficiente de correlación de Pearson:

$$S_{xy} = \frac{\sum XY}{n} - \bar{X}\bar{Y} = \frac{1561}{16} - 6{,}375 \cdot 13{,}625 = 10{,}703 \qquad r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{10{,}703}{2{,}666 \cdot 4{,}498} = 0{,}892$$

También podemos calcular Pearson con la siguiente fórmula:

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} = \frac{(16)(1561) - (102)(218)}{\sqrt{[(16)(764) - 102^2][(16)(3294) - 218^2]}} = 0{,}892$$

Pendiente de la recta de regresión:

$$B = r_{xy}\frac{S_y}{S_x} = 0{,}892\frac{4{,}498}{2{,}666} = 1{,}505$$

O bien:

$$B = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{16 \cdot 1561 - 102 \cdot 218}{16 \cdot 764 - 102^2} = 1{,}505$$

Ordenada en el origen de la recta de regresión en puntuaciones directas:

$$B_0 = \bar{Y} - B\bar{X} = 13{,}625 - 1{,}505 \cdot 6{,}375 = 4{,}027$$

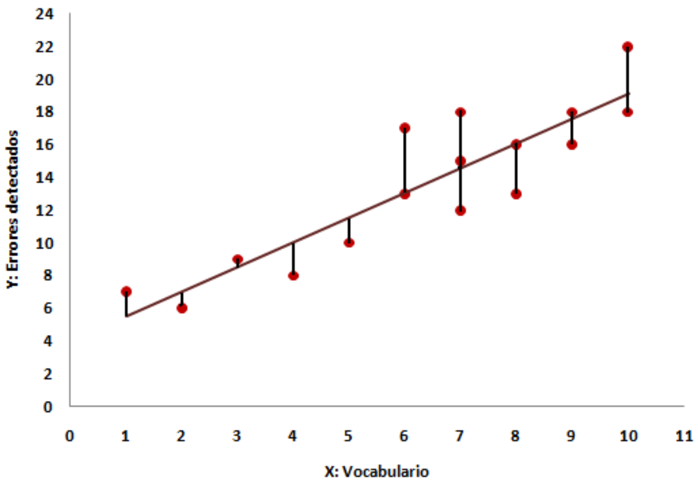
Finalmente, se puede expresar la recta de regresión en puntuaciones directas, diferenciales o típicas:

Directas	Diferenciales	Típicas
$Y' = BX_i + B_0$	$y'_i = Bx_i = B(X_i - \bar{X})$	$z'_i = r_{xy}Z_{x_i} = r_{xy}\left(\frac{X_i - \bar{X}}{S_x}\right)$

8.3.1 Coeficientes de la regresión lineal simple

Cuando utilizamos el modelo lineal para estimar cada valor *Y* a partir de *X*, generalmente se comete **un error en la estimación de la VD (*Y*)** para cada uno de los sujetos, ya que el valor pronosticado (*Y'*) y el valor medido (*Y*) no suelen coincidir. La diferencia entre ambos es el error de estimación ($\varepsilon = Y - Y'$). La Figura 8.2 muestra este error para cada uno de los sujetos, que viene dado por la magnitud o longitud de la línea vertical que separa cada dato de la predicción realizada por la recta de regresión. Se observa, que en algunos casos el valor que se obtiene con la recta de ajuste (la estimación, *Y'*) coincide con el verdadero valor de la VD (representado por los puntos), aunque en la mayoría de los casos no es así.

Figura 8.2 *Errores después del ajuste de una recta*



Al método de ajuste de una recta de regresión se le conoce como **ajuste por mínimos cuadrados,** ya que el objetivo es encontrar los valores *B* y B_0 que hacen más pequeño el error al cuadrado. Es decir, se trata de encontrar los valores de *B* y B_0 con los que la siguiente expresión toma un valor mínimo:

$$\sum (Y - Y')^2 = \sum \varepsilon^2$$

Siendo las ecuaciones que minimizan el error cuadrático las calculadas en el apartado anterior.

Una característica importante de la recta de regresión calculada por mínimos cuadrados, consiste en que proporciona estimaciones insesgadas de la VD en el sentido de que la media de los valores pronosticados es igual a la media de los valores observados. Es decir:

$$\frac{\sum Y_i}{n} = \frac{\sum Y_i'}{n} \rightarrow \bar{Y} = \bar{Y}'$$

Construida la recta de ajuste podemos expresar la variable dependiente, *Y*, como una función de la variable independiente, *X*, mediante la siguiente expresión:

$$Y = B_0 + BX + \varepsilon$$

Donde $\varepsilon = Y - Y'$ representa el error de predicción.

En el análisis de regresión simple el coeficiente "protagonista" es el factor ***B***, conocido como **pendiente de la recta**, y cuantifica el incremento que se produce en la estimación de la variable dependiente (*Y'*) cuando la independiente (*X*) aumenta en una unidad.

En la Figura 8.3 se ve de manera gráfica el significado de B en nuestros datos. La estimación de *Y* para un valor *X = 4*, proporciona el valor *10,049*, y para una *X = 5*, el valor es *11,555*. La diferencia entre estos valores al aumentar *X* en una unidad (de 4 a 5) es lo que aumenta *Y'* y ese es el valor de la pendiente. En el caso del ejemplo que ilustra esta explicación, la pendiente nos dice que los escolares, con cada punto más que obtienen en la prueba de vocabulario detectan, en promedio, *1,5* errores más en la prueba de lectura.

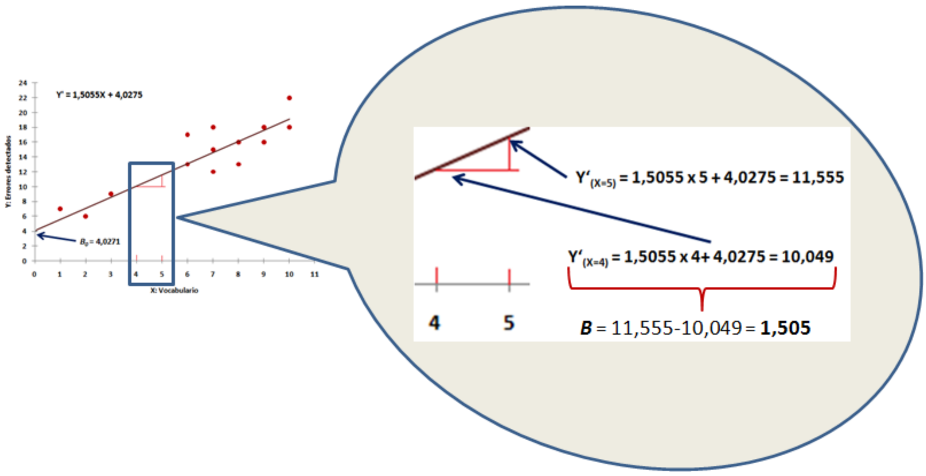


Figura 8.3 *Interpretación gráfica de la pendiente de la recta de regresión*

La **constante de la recta** de regresión, ***B₀***, señala el punto en el que ésta corta al eje de ordenadas, por lo que se denomina ordenada en el origen. Es decir, refleja el valor estimado de *Y* cuando *X* es igual a *0*. Generalmente no es un coeficiente interpretable, excepto cuando el valor *0* se encuentra dentro del rango de valores de la VI. La recta de regresión se construye con los valores de dicho rango, por lo que fuera del mismo, es posible que la función que estima la relación entre *X* e *Y* cambie de forma.

8.3.2 Bondad de Ajuste de la Recta de Regresión

La expresión Bondad de Ajuste, se refiere a cómo de "explicativa" es la recta respecto de los datos sobre los que se ha ajustado. Para explicar este concepto, veamos que sucede en uno de los 16 sujetos del ejemplo que estamos siguiendo (Figura 8.4).

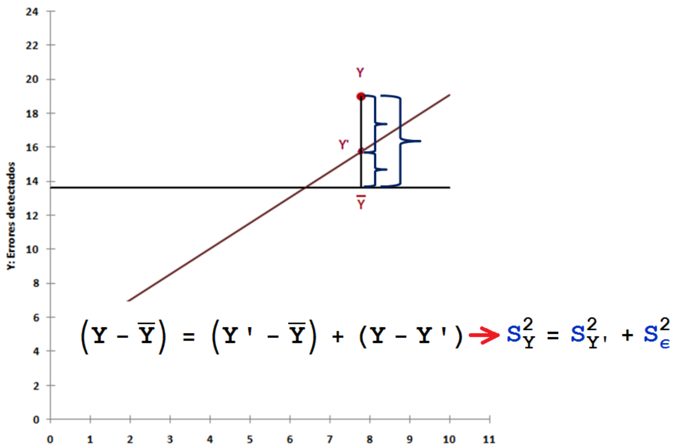


Figura 8.4 *Descomposición de la suma de cuadrados de la VD*

Imagine el lector que **sólo se conocen las puntuaciones de los sujetos en la prueba de detección de errores (*Y*)**, y se desea hacer una estimación para un sujeto concreto. Si no se conocen las puntuaciones en la prueba de vocabulario (*X*), se otorga, como mejor estimación, la media del grupo en la variable dependiente a todos los sujetos. Es decir, el error cometido para cada sujeto concreto será: $(Y_i - \bar{Y})$, y el error de predicción global vendrá dado por la expresión: $\sum(Y_i - \bar{Y})^2$.

Pero si **se conocen las puntuaciones de los sujetos en la prueba de vocabulario (*X*)**, y existe relación lineal entre *X* e *Y*, se pueden estimar las puntuaciones de los sujetos en la prueba de detección de errores (*Y*) con mayor precisión, siendo el error cometido para cada sujeto, en este caso: $(Y_i - Y')$.

Se puede descomponer el error original, cuando no contamos con la variable independiente $(Y - \bar{Y})$, en dos partes:

$$(Y - \bar{Y}) = (Y' - \bar{Y}) + (Y - Y')$$

Del error original, por lo tanto, hemos reducido una parte $(Y' - \bar{Y})$, pero aún queda otra parte $(Y - Y')$, sin explicar. Dichas partes son independientes entre sí (su correlación vale cero), por lo tanto:

$$\sum(Y - \bar{Y})^2 = \sum(Y' - \bar{Y})^2 + \sum(Y - Y')^2$$

El término $\sum(Y-\bar{Y})^2$ representa la suma de cuadrados de Y, o suma de cuadrados total ($SC_y = SC_T$). El término en el que se reduce el error original $\left(\sum(Y'-\bar{Y})^2\right)$ se denomina suma de cuadrados de la regresión $\left(SC_{reg}\right)$, siendo la suma de cuadrados de error o residual: $SC_{res} = \sum(Y-Y')^2$. En definitiva, al utilizar la ecuación de regresión, la suma de cuadrados de Y se descompone de la siguiente forma:

$$SC_T = SC_{reg} + SC_{res}$$

Dividiendo las sumas de cuadrados por el número total de observaciones, se obtienen la varianza total de Y (S_y^2), la varianza de las puntuaciones pronosticadas ($S_{y'}^2$) y la varianza de los errores (S_ε^2).

$$\frac{\sum(Y-\bar{Y})^2}{n} = \frac{\sum(Y'-\bar{Y})^2}{n} + \frac{\sum(Y-Y')^2}{n} \rightarrow \; S_y^2 = S_{Y'}^2 + S_\varepsilon^2$$

A partir de la esta ecuación se puede establecer una serie de relaciones. La primera representa ***la proporción de la varianza de la VD explicada por la varianza de la VI***. La cuantía de esta proporción es el cuadrado del coeficiente de correlación de Pearson entre la VD y la VI (esto solo sirve para el caso de la Regresión Lineal Simple).

$$\boxed{r_{xy}^2 = \frac{S_{y'}^2}{S_y^2}} = \frac{\sum(Y'-\bar{Y})^2}{\sum(Y-\bar{Y})^2} = \frac{SC_{Reg}}{SC_y} \rightarrow S_{y'}^2 = r_{xy}^2 S_y^2$$

$$\boxed{\left(1-r_{xy}^2\right) = \frac{S_\varepsilon^2}{S_y^2}} = \frac{\sum(Y-Y')^2}{\sum(Y-\bar{Y})^2} = \frac{SC_{Residuos}}{SC_y} \rightarrow S_\varepsilon^2 = S_y^2\left(1-r_{xy}^2\right) \rightarrow \boxed{S_\varepsilon = S_y\sqrt{1-r_{xy}^2}}$$

En resumen, r_{xy}^2 (que también designaremos como R^2), sirve para evaluar la bondad de ajuste de la recta de regresión, y se denomina ***Coeficiente de Determinación***, reflejando la proporción de la variabilidad de la VD que es imputada (o explicada por) la variabilidad de la VI. Su complemento, $\left(1-r_{xy}^2\right)$, se denomina ***Coeficiente de Alienación***, y es la parte residual de la variabilidad de la VD, atribuible a otros factores no relacionados linealmente con la VD.

También se puede interpretar r_{xy}^2 como ***la proporción en que se reduce el error de la VD cuando empleamos la recta de regresión para estimarla***.

Se puede representar la varianza compartida mediante diagramas de Venn, donde la varianza de cada variable es representada por círculos de área igual a la unidad. La intersección de ambos representa la proporción de varianza común $\left(r_{xy}^2\right)$.

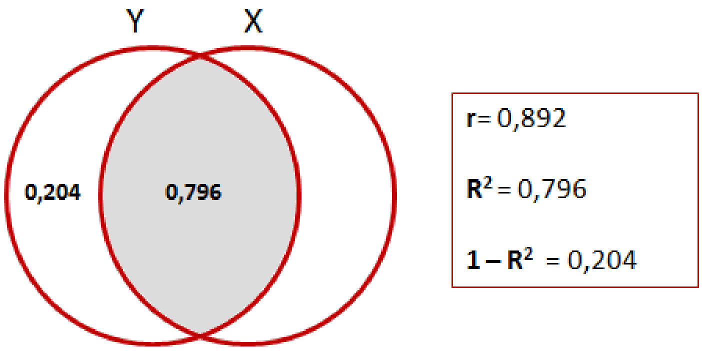


Figura 8.5 *Diagrama de Venn con la representación de la proporción de varianza compartida*

Otro indicador del ajuste, además de R^2, es el error típico, que es el estimador insesgado de la desviación típica de error:

$$\hat{S}_\varepsilon = \sqrt{\frac{\sum (Y - Y')^2}{n-2}}$$

Los cálculos con los datos del ejemplo, se podrían realizar, por ejemplo:

$$r_{xy}^2 = 0{,}892^2 = 0{,}796 \qquad r_{xy}^2 = \frac{S_{y'}^2}{S_y^2} \rightarrow S_{y'}^2 = r_{xy}^2 S_y^2 = 0{,}892^2 \cdot 20{,}234 = 16{,}113$$

$$S_y^2 = S_{y'}^2 + S_\varepsilon^2 \rightarrow S_\varepsilon^2 = S_y^2 - S_{y'}^2 = 20{,}234 - 16{,}113 = 4{,}121 \rightarrow S_\varepsilon = 2{,}030$$

$$SC_T = SC_y = nS_y^2 = 16 \cdot 20{,}234 = 323{,}75 \qquad SC_{reg} = r_{xy}^2 SC_y = 0{,}796 \cdot 323{,}75 = 257{,}816$$

$$SC_{error} = \left(1 - r_{xy}^2\right) SC_y = 65{,}934 \rightarrow \hat{S}_\varepsilon = \frac{65{,}934}{14} = 4{,}71$$

8.3.3.- Inferencias sobre correlación y regresión

Para que sean válidas las inferencias proporcionadas por la recta de regresión, se deben de cumplir los siguientes supuestos:

1. Independencia de las observaciones.

2. Homocedasticidad. Las varianzas de las distribuciones de los errores, condicionadas a los diferentes valores de la VI, deben ser iguales.
3. Normalidad de las distribuciones condicionadas con media *Y'*.
4. Independencia entre los valores estimados, *Y'*, y los errores de estimación, ε. Expresado en términos de coeficiente de correlación de Pearson, $r_{y'\varepsilon} = 0$. Esto es así debido a que los errores se distribuyen de manera aleatoria.

En la Figura 8.6 se representan los supuestos 2 y 3.

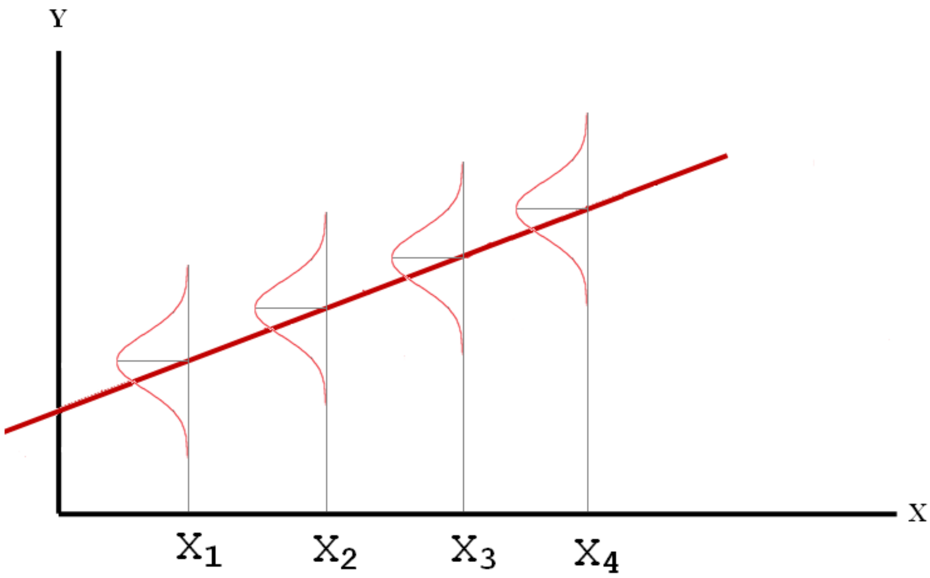


Figura 8.6 *Representación supuestos 2 y 3 en el ARS*

8.3.3.1.- Contraste sobre el coeficiente de correlación de Pearson.

El contraste de hipótesis que se presenta a continuación, se utiliza para comprobar si es significativo el coeficiente de correlación de Pearson, es decir, si existe relación lineal entre *X* e *Y*. Con los datos del ejemplo y un nivel de confianza del 99%:

Condiciones y supuestos. Tenemos dos variables medidas en una escala de intervalo o razón que se distribuyen normalmente en la población. En el caso del ejemplo de este capítulo, hemos de suponer que las variables prueba de vocabulario (*X*) y número de errores ortográficos (*Y*) se distribuyen normalmente en la población, dado que la muestra es menor de 30 observaciones.

Formulación de hipótesis. La hipótesis nula postula que en la población el coeficiente de correlación de Pearson es igual a cero, mientras que la hipótesis alternativa indica que la relación lineal entre *X* e *Y* es significativa:

$$H_0: \ \rho = 0$$
$$H_1: \ \rho \neq \ 0$$

Estadístico de contraste y distribución muestral. El estadístico de contraste se distribuye según t de Student con $n-2$ grados de libertad, y viene dado por la siguiente fórmula:

$$T = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

Con los datos del ejemplo:

$$T = \frac{0{,}892\sqrt{16-2}}{\sqrt{1-0{,}892^2}} = 7{,}399 \approx 7{,}40$$

Establecer regla de decisión en función del nivel de confianza. Para un nivel de confianza del 99% en un contraste bilateral, el valor crítico obtenido en las tablas t de Student es igual a: 2,977. Dado que: $7{,}40 > 2{,}977$ rechazamos la hipótesis nula, concluyendo que la relación entre X e Y es significativa.

Mediante un programa informático adecuado se comprueba que el nivel crítico es: $p < 0{,}0001$. Con las tablas se llega a la conclusión que el nivel crítico es: $p < 0{,}005$, que es la probabilidad de obtener valores superiores a 2,977 en una distribución t de Student con 14 grados de libertad.

Interpretar los resultados en función del contexto de la investigación. Existe relación lineal entre las variables prueba de vocabulario (*X*) y número de errores ortográficos (*Y*).

Podemos observar que si el coeficiente de correlación de Pearson es distinto de cero, también será distinta de cero la pendiente de la recta de regresión de *Y* sobre *X*, dado que ambos índices se relacionan según la siguiente ecuación:

$$B = r_{xy}\frac{S_y}{S_x}$$

8.3.3.2.- Contraste para el coeficiente de regresión B (ANOVA).

En el caso de la regresión lineal simple, también se puede contrastar si existe relación lineal entre *X* e *Y* utilizando la descomposición de la variabilidad total vista en el apartado 8.3.2, ordenando los datos en una tabla como las utilizadas en los temas sobre Análisis de Varianza:

Tabla 8.3

Tabla ANOVA para el contraste de la regresión

Fuentes de variación	Sumas de cuadrados	Grados de libertad	Medias cuadráticas	F
Regresión	SC_{Reg}	1	$MC_{Reg} = \frac{SC_{Reg}}{1}$	$F = \frac{MC_{Reg}}{MC_{Res}}$
Residual	SC_{Res}	$n-2$	$MC_{Res} = \frac{SC_{Res}}{n-2}$	
Total	$SC_T = SC_y$	$n-1$		

Completando la Tabla 8.3 con datos del ejemplo:

FV	SC	gl	MC	F
Regresión	257,816	1	257,816	54,7
Residual	65,934	14	4,709	
Total	323,750	15		

También podríamos calcular el estadístico F mediante la siguiente expresión:

$$F = \frac{R^2/1}{(1-R^2)/(N-2)} = \frac{0{,}796}{(1-0{,}796)/(16-2)} = 54{,}7$$

El estadístico de contraste F resulta significativo, pues la probabilidad de encontrar un valor F igual o mayor, con 1 y 14 grados de libertad es: $p = 3{,}358 \cdot 10^{-6}$ (valor calculado mediante un programa informático).

El estadístico de contraste T visto en el punto 8.3.3.1 y el estadístico F están relacionados según la siguiente expresión:

$$t_n^2 = F_{1,n}$$

8.3.3.3.- Contraste para el coeficiente de regresión B.

Para determinar si hay evidencia estadística de que la pendiente es diferente de cero, es decir si la pendiente es significativamente diferente a una línea horizontal, se utiliza el siguiente contraste ($NC = 95\%$):

Condiciones y supuestos. Como se indicó previamente, los supuestos son: Independencia de las observaciones, homocedasticidad, normalidad de las distribuciones condicionadas e independencia entre los valores estimados y los errores de estimación.

Formulación de hipótesis. Normalmente se pretende comprobar si la pendiente de la recta de regresión en la población es distinta de cero, por lo que las hipótesis son:

$$\begin{aligned} H_0: &\quad \beta = 0 \\ H_1: &\quad \beta \neq 0 \end{aligned}$$

Estadístico de contraste y distribución muestral. El estadístico de contraste se distribuye según t de Student con $n-2$ grados de libertad, y viene dado por la siguiente expresión:

$$T = \frac{B-\beta}{\sigma_B} = \frac{B-\beta}{\frac{S_y}{S_x}\sqrt{\frac{1-r_{xy}^2}{(n-2)}}}$$

(Nota: en esta fórmula se obtiene el mismo resultado empleando desviaciones típicas o cuasidesviaciones típicas)

El valor β es el que especifica la hipótesis nula. Normalmente interesa comprobar: $H_0: \beta = 0$. Aplicando este contraste a la pendiente de los datos del ejemplo, el valor del estadístico es:

$$T = \frac{1{,}5055 - 0}{\frac{4{,}498}{2{,}666}\sqrt{\frac{1-0{,}8924^2}{(16-2)}}} = \frac{1{,}5055}{0{,}2035} \approx 7{,}4$$

Observe el lector que el valor obtenido en este caso es igual al estadístico T utilizado en el punto 8.3.3.1. Efectivamente, siempre que: $H_0: \beta = 0$:

$$T = \frac{B-0}{\frac{S_y}{S_x}\sqrt{\frac{1-r_{xy}^2}{(n-2)}}} = \frac{r_{xy}\frac{S_y}{S_x}}{\frac{S_y}{S_x}\sqrt{\frac{1-r_{xy}^2}{(n-2)}}} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

Establecer regla de decisión en función del nivel de confianza. Para un nivel de confianza del 95% en un contraste bilateral: $7{,}40 > 2{,}145$, luego rechazamos la hipótesis nula, concluyendo que la pendiente de la ecuación de regresión es distinta de cero, siendo el nivel crítico: $p < 0{,}0001$ (calculado con un programa informático).

Interpretar el resultado en función del contexto de investigación. Existe relación lineal entre la prueba de vocabulario (*X*) y el número de errores ortográficos detectados en un texto (*Y*), de manera que podemos pronosticar los valores de la VD en función de los valores de la VI.

Intervalo de confianza. El intervalo de confianza para la pendiente de la recta de regresión se puede calcular mediante la siguiente expresión:

$$B \pm \left(t_{(n-2;1-\alpha/2)}\right)(\sigma_B)$$

Aplicando la fórmula a los resultados del ejemplo se obtienen, para un nivel de confianza del 95%, los siguientes límites:

$$1{,}5055 \pm (2{,}145)\left(\frac{4{,}498}{2{,}666}\sqrt{\frac{1-0{,}8924^2}{(16-2)}}\right) = \begin{cases} 1{,}942 \\ 1{,}069 \end{cases}$$

8.3.3.4.- Contraste para el coeficiente de regresión B_0.

También se puede comprobar si el intercepto es distinto de cero, aunque en este caso, ya se ha señalado que en la mayor parte de los estudios suele ser ignorado. El estadístico de contraste se distribuye según t de Student con $n-2$ grados de libertad, y viene dado por la expresión:

$$T = \frac{B_0 - 0}{\sigma_{B_0}} = \frac{B_0 - 0}{\hat{S}_\varepsilon\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{(n-1)\hat{S}_X^2}}}$$

Siendo $\hat{S}_\varepsilon$ el ***Error Típico***, cuyo valor es la raíz cuadrada de la Media Cuadrática (MC) de los Residuos de la tabla del ANOVA para el contraste de la regresión, que representa al estimador de la varianza residual en la población para el caso de la regresión bivariada. Como en el caso de la pendiente, el estadístico *T* tiene la misma distribución con los mismos grados de libertad.

Aplicando el contraste a los datos del ejemplo:

$$T = \frac{4,0275 - 0}{\sqrt{4,7096}\sqrt{\frac{1}{16} + \frac{6,375^2}{(16-1)7,583}}} = \frac{4,0275}{1,4061} = 2,864$$

Con $\alpha = 0,05$ en un contraste bilateral rechazamos la hipótesis nula de que el intercepto es igual a 0 ya que en este caso, para 14 grados de libertad los valores críticos son: -2,14 y 2,14

Para el intercepto, la fórmula de cálculo del intervalo de confianza es:

$$B_0 \pm (t_{(n-2;1-\alpha/2)})(\sigma_{B_0})$$

Aplicando la expresión a los datos del ejemplo los límites son:

$$4,0275 \pm (2,145)\left(\sqrt{4,7096}\sqrt{\frac{1}{16} + \frac{6,375^2}{(16-1)7,583}}\right) = \begin{cases} 7,043 \\ 1,012 \end{cases}$$

8.4.- Análisis de Regresión Múltiple

Como se ha señalado en el epígrafe de Introducción, en este tema sólo tratamos modelos lineales de explicación del comportamiento de una VD en función de una o varias VI. Ya hemos desarrollado la técnica de **Análisis de Regresión Lineal Simple**, y en este epígrafe ampliamos dicho modelo para dos VI o variables predictoras. Como en el caso de una sola variable predictora, se va a desarrollar con el mínimo aparato matemático posible. La técnica de cálculo con el modelo de dos variables independientes es relativamente sencilla y se puede desarrollar con un calculadora científica, aunque su modelo matemático, el mismo que el del **Modelo Lineal General** (MGL), del cual los modelos de regresión y los modelos de análisis de la varianza son parte, requiere para su desarrollo algebra de matrices, por lo que queda fuera del alcance de este texto. Dado que, en la actualidad, todos estos procedimientos de análisis se realizan con programas informáticos de análisis estadístico, el interés estriba en saber leer e interpretar correctamente los resultados del análisis.

8.4.1.- Regresión con dos Variables Independientes

Para la explicación vamos a servirnos de un ejemplo numérico que hace menos abstracto el modelo. Supongamos que un psicólogo escolar quiere determinar qué factores pueden influir en el rendimiento en matemáticas en uno de los cursos de educación secundaria. Supone que el tiempo que dedican al estudio en general es importante, y quizás también su capacidad para el razonamiento abstracto. Para llevar a cabo esta investigación, selecciona al azar una muestra de 15 estudiantes del colegio y registra el tiempo semanal de estudio (variable X_1) y les administra, además, un test de razonamiento abstracto (variable X_2). Las notas obtenidas por estos 15 escolares en el último examen que han realizado de matemáticas le sirven como variable dependiente (*Y*). Los datos son los que se muestran en la Tabla 8.4

Tabla 8.4

Datos para el desarrollo del análisis con dos VI

Sujeto	**Horas Estudio (X_1)**	**Test Razonamiento (X_2)**	**Punt. Matemáticas (Y)**
1	8	19	54
2	9	18	52
3	6	14	34
4	9	24	63
5	9	19	46
6	9	16	44
7	12	17	50
8	9	14	52
9	6	23	57
10	11	21	53
11	10	17	56
12	13	19	67
13	9	24	57
14	9	19	54
15	11	17	51

El modelo de estimación lineal de la VD con dos VI's, constará de dos coeficientes de regresión, uno para cada VI, y una constante que será el valor estimado para la VD cuando son nulas las dos VI. No obstante, como ya hemos explicado anteriormente, la constante, si no está el valor cero dentro del rango de valores de las variables predictoras no se toma en consideración en el análisis. Es decir, si X_1= 0 y X_2 = 0 no forman parte de los rangos admitidos empíricamente por ambas variables, no tiene sentido considerar el valor que adoptaría la constante en esos casos. El modelo de estimación es:

$$Y' = B_1X_1 + B_2X_2 + B_0$$

Por lo que la VD se puede expresar como:

$$Y = Y' + \varepsilon \ = \ B_1X_1 + B_2X_2 + B_0 + \varepsilon$$

Siendo B_1 el coeficiente de regresión parcial para X_1, B_2 el coeficiente de regresión parcial para X_2, B_0 el intercepto, y ε los residuos una vez que se ha determinado la función de estimación de la VD. Al igual que en regresión simple, estos coeficientes son los que hacen mínimo el error cuadrático de predicción, es decir, minimizan las diferencias cuadráticas entre *Y* e *Y'*.

Antes de calcular los coeficientes de regresión parciales de la ecuación, llamados así para remarcar cual es el peso o efecto de una VI cuando el resto de las VI permanecen constantes, se presentan en la Tabla 8.6 los estadísticos descriptivos de cada una de las variables y los coeficientes de correlación entre las variables dos a dos (también llamados bivariados). Hemos simplificado la notación de los coeficientes de correlación (r_{y1} representa la correlación entre la variable *Y* y el predictor X_1, y el resto siguen la misma pauta).

Tabla 8.5

Estadísticos descriptivos de los datos de la Tabla 8.4

	X_1	X_2	Y
Medias	9,33	18,73	52,67
Varianzas	3,422	9,396	56,222
D. típicas	1,850	3,065	7,498
CuasiVarianzas	3,667	10,067	60,238
Cuasi D. Típicas	1,915	3,173	7,761

$$r_{y1} = 0{,}441 \rightarrow r_{y1}^2 = 0{,}194$$

$$r_{y2} = 0{,}628 \rightarrow r_{y2}^2 = 0{,}395$$

$$r_{12} = -0{,}043 \rightarrow r_{12}^2 = 0{,}002$$

Las ecuaciones de regresión lineal múltiple se pueden expresar en puntuaciones directas, diferenciales o típicas:

	Directas	Diferenciales	Típicas
Ec. de regresión	$Y' = B_1X_1 + B_2X_2 + B_0$	$y' = B_1x_1 + B_2x_2$	$z_{y'} = \beta_1 z_1 + \beta_2 z_2$

Donde a los coeficientes en puntuaciones típicas (β_1 y β_2) se les denomina coeficientes de regresión estandarizados. Para el cálculo de los coeficientes de regresión parcial se pueden utilizar las siguientes fórmulas:

	Directas o diferenciales	Típicas
Coeficiente para X_1	$B_1 = \beta_1 \frac{S_y}{S_1} = \frac{\sum x_1 y \sum x_2^2 - \sum x_2 y \sum x_1 x_2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$	$\beta_1 = \frac{r_{y1} - r_{y2} r_{12}}{1 - r_{12}^2} = B_1 \sqrt{\frac{\sum x_1^2}{\sum y^2}}$
Coeficiente para X_2	$B_2 = \beta_2 \frac{S_y}{S_2} = \frac{\sum x_2 y \sum x_1^2 - \sum x_1 y \sum x_1 x_2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$	$\beta_2 = \frac{r_{y2} - r_{y1} r_{12}}{1 - r_{12}^2} = B_2 \sqrt{\frac{\sum x_2^2}{\sum y^2}}$

Con los resultados de la Tabla 8.6 se calculan en primer lugar los coeficientes de regresión estandarizados:

$$\beta_1 = \frac{0{,}441 - (0{,}628)(-0{,}043)}{1 - (-0{,}043)^2} = 0{,}469 \qquad \beta_2 = \frac{0{,}628 - (0{,}441)(-0{,}043)}{1 - (-0{,}043)^2} = 0{,}649$$

A continuación, se obtienen fácilmente los coeficientes sin estandarizar:

$$B_1 = \beta_1 \frac{S_y}{S_1} = 0{,}469 \frac{7{,}498}{1{,}85} = 1{,}899 \qquad B_2 = \beta_2 \frac{S_y}{S_2} = 0{,}649 \frac{7{,}498}{3{,}065} = 1{,}587$$

Siendo la constante para la ecuación en puntaciones directas:

$$B_0 = \bar{Y} - B_1\bar{X}_1 - B_2\bar{X}_2 = 52{,}67 - (1{,}899)(9{,}33) - 1{,}587(18{,}73) = 5{,}217$$

Obtenidos los coeficientes, las ecuaciones de regresión en puntuaciones directas y típicas son:

$$Y' = 1{,}899X_1 + 1{,}587X_2 + 5{,}217 \qquad z'_Y = 0{,}469z_1 + 0{,}649z_2$$

Al ser dos las variables independientes, las estimaciones quedan situadas en un plano, que se conoce como plano de regresión. Algunas de las puntuaciones de la VD estarán por encima del plano y otras por debajo, y esas distancias de cada punto de la VD al plano forman los residuos del modelo de estimación (véase Figura 8.7).

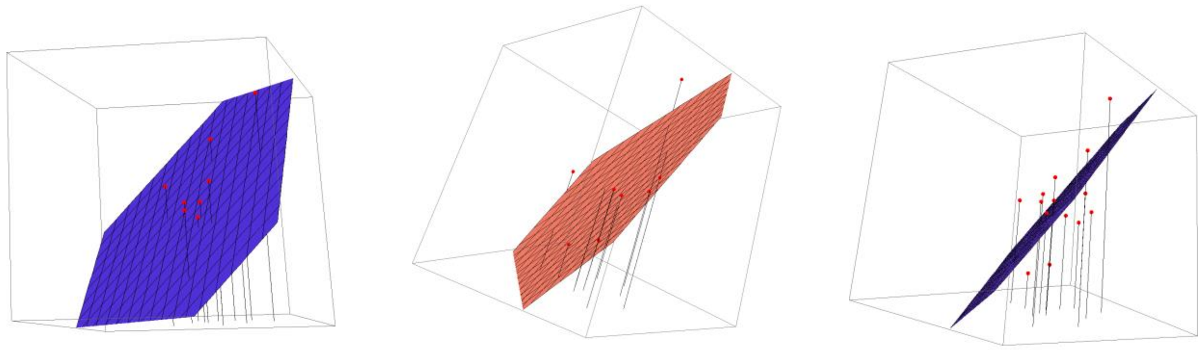


Figura 8.7: *tres vistas del conjunto de puntos y el plano de regresión. La zona azul representa el plano visto "desde arriba", la zona naranja representa el plano visto "desde abajo" . La tercera gráfica intenta visualizar todos los puntos, tanto los que están situados por encima como los que están situados por debajo del plano. En este caso, el plano se ve en "escorzo". Los datos están representados por puntos rojos.*

El modelo ajustado, *Y'*, ya arroja una primera interpretación: cuando permanece constante X_2, por cada hora de estudio, la puntuación en matemáticas aumenta en promedio, 1,899 puntos, y cuando permanece constante X_1, por cada punto más en razonamiento abstracto, aumenta 1,587 la puntuación en matemáticas

8.4.2.- Ajuste del modelo. Medidas de asociación

En regresión simple, el ajuste del modelo viene dado por el coeficiente de determinación que es el cuadrado del coeficiente de correlación de Pearson entre la VD y la VI, y ese coeficiente informa de qué porción de la variabilidad de la VD es explicada por, o atribuida a, la variabilidad de la VI. En el caso de la regresión múltiple, las preguntas básicas que hay que responder son las siguientes:

- ¿Estiman bien la VD el **conjunto** de VI's?
- ¿Cuánta variabilidad explica cada variable individualmente una vez que las otras variables han aportado lo suyo?

Para responder a la primera pregunta disponemos del **coeficiente de correlación múltiple** (*R*), que correlaciona la VD con una combinación lineal de variables independientes. Su cuadrado (R^2) es el **coeficiente de determinación**, que indica la proporción de variabilidad de la variable dependiente explicada por el conjunto de variables independientes.

$$R_{y.12} = \sqrt{\frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2}} = \sqrt{\beta_1 r_{y1} + \beta_2 r_{y2}}$$

Con los datos del ejemplo, el valor de $R_{y.12}$ es:

$$R_{y.12} = \sqrt{(0{,}469)(0{,}441) + 0{,}649(0{,}628)} = 0{,}7836$$

El coeficiente de determinación múltiple es el cuadrado del coeficiente de correlación múltiple:

$$R_{y.12}^2 = 0{,}7836^2 = 0{,}614$$

La interpretación de $R_{y.12}^2$ es similar a r_{xy}^2. Es decir, la combinación de las dos variables (tiempo de estudio y razonamiento abstracto) se atribuyen el 61,4% de la variabilidad de las puntuaciones obtenidas en matemáticas, y por tanto el 38,6% restante se debe a otros factores no relacionados linealmente con las variables independientes.

El estimador insesgado de ρ^2 en la población se denomina ***R*² Ajustado,** siendo igual a:

$$\hat{R}_{y.12}^2 = 1 - (1 - R_{y.12}^2)\frac{n-1}{n-p-1}$$

siendo n, el número de observaciones y p, el número de variables independientes o predictoras. Para el caso de ejemplo, el valor de R^2 Ajustado es:

$$\hat{R}^2_{y.12} = 1 - (1 - 0{,}614)\frac{15 - 1}{15 - 2 - 1} = 0{,}5498$$

8.4.3.- Correlación Semiparcial y Parcial

La segunda de las preguntas que hacíamos al comienzo del epígrafe anterior, es cómo determinar la contribución de cada variable independiente a la explicación de la dependiente. La respuesta a esta pregunta la proporciona la llamada **correlación semiparcial**, *sr*, y su cuadrado, sr^2. Antes de explicar qué son esas nuevas correlaciones que acaban de entrar en escena, piense el lector que cuando en un modelo intervienen más de dos variables, las correlaciones que se calculan entre las variables dos a dos, no son correlaciones "puras", en el sentido de que no miden relaciones entre esas dos variables al margen del influjo que las otras variables del modelo puedan tener sobre cada una de ellas. Estas correlaciones que se calculan entre dos variables (correlaciones bivariadas) se denominan **correlaciones de orden cero**, y a través del valor obtenido no se puede saber qué parte de la varianza de la VD es capaz de explicar independientemente cada una de las VI's, puesto que entre éstas también puede haber relación. Por lo tanto, para saber qué parte de la VD explica cada VI al margen de las otras VI's, es necesario eliminar el influjo que sobre cada VI tienen el resto de las VI's, para así poder determinar el influjo único que esa VI tiene sobre la VD. Esta relación entre cada VI y la VD habiendo eliminado el influjo del resto de las VI's sobre cada VI es lo que se llama **Coeficiente de Correlación Semiparcial**.

¿Cómo se calcula este coeficiente? Ya sabemos, por todo lo explicado hasta el momento, que en un modelo de regresión hay una proporción de varianza explicada y una proporción de varianza no explicada que es la varianza de los residuos. La varianza explicada lo es en función de una cierta combinación de las variables independientes; por consiguiente, si en un modelo, por ejemplo, con dos predictoras X_1 y X_2, se ajusta una regresión de la 1 sobre la 2, se extraen los residuos y, por último, los correlaciono con la VD, habré calculado el coeficiente de correlación semiparcial entre X_1 y la VD habiendo eliminado el influjo de X_2 sobre la VD. Por otra parte, si se ajusta una regresión simple entre X_2 y X_1 (obsérvese el cambio de subíndices en relación a la frase anterior), se extraen los residuos y éstos se correlacionan con la VD, habré calculado la correlación entre el predictor X_2 y la VD, habiendo eliminado el influjo de X_1 sobre la VD.

Para llevar a cabo este cálculo de los coeficientes de correlación semiparcial no es necesario proceder como hemos explicado en el párrafo anterior; hay fórmulas muy sencillas para ello, a partir de las correlaciones de orden cero.

$$sr_1 = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{1 - r_{12}^2}} \qquad sr_2 = \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{1 - r_{12}^2}}$$

Elevando al cuadrado estos valores se tiene la contribución que cada VI tiene sobre la VD habiendo eliminado el influjo de las otras VI's. En la Figura 8.8 se observa gráficamente, mediante un Diagrama de Venn, estás contribuciones expresadas en forma de área compartida

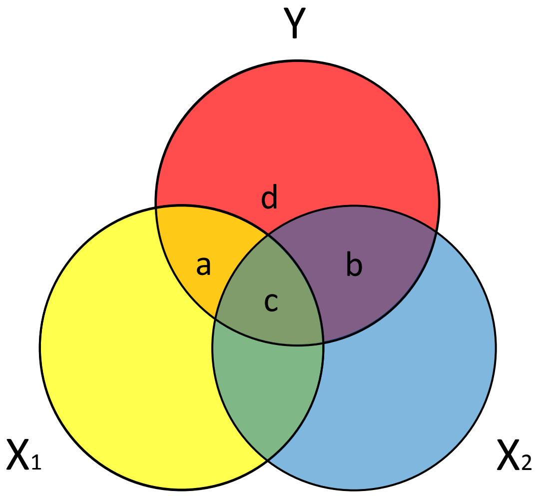


Figura 8.8 *Diagrama de Venn para un modelo de regresión con dos variables independientes*

Tomando como referencia el diagrama de la Figura 8.8, las equivalencias entre las zonas designadas con letras y los cuadrados de los coeficientes de correlación semiparcial, son las siguientes:

$$a = sr_1^2 = R_{y.12}^2 - r_{y2}^2 \qquad\qquad b = sr_2^2 = R_{y.12}^2 - r_{y1}^2$$

Siendo:

$$R_{y.12}^2 = a + b + c \qquad\qquad r_{y1}^2 = a + c \qquad\qquad r_{y2}^2 = b + c$$

Para el ejemplo numérico que sirve de base a la explicación, los cálculos de los coeficientes de correlación semiparcial son los siguientes:

$$sr_1 = \frac{0{,}4406 - (0{,}6285)(-0{,}0431)}{\sqrt{1 - (-0{,}0431)^2}} = 0{,}4681 \qquad sr_2 = \frac{0{,}6285 - (0{,}4406)(-0{,}0431)}{\sqrt{1 - (-0{,}0431)^2}} = 0{,}6481$$

Estos valores elevados al cuadrado dan la proporción de varianza compartida por cada predictora habiendo eliminado el influjo de la otra predictora sobre la misma.

$$a = sr_1^2 = 0{,}4681^2 = 0{,}2191 \qquad\qquad b = sr_2^2 = 0{,}6481^2 = 0{,}4200$$

El valor $0{,}4681^2$ (0,2191) es ***a*** en el diagrama de la Figura 8.8, y $0{,}6481^2$ (0,4200) es ***b***. Estos dos valores representan la contribución exclusiva que cada variable hace a la explicación de la dependiente. La porción ***c***, es la proporción de varianza de la VD estimada conjuntamente (es decir, de forma redundante) por las dos variables. Sin embargo esta proporción es de muy difícil interpretación.

El otro coeficiente que se calcula en los modelos de regresión, y que además sirve para determinar cuál es la primera variable que se incorpora al modelo cuando se realiza variable a variable[3], es el denominado **coeficiente de correlación parcial**, *pr*. La diferencia con el semiparcial es que en el parcial se elimina el influjo de los predictores tanto de la VI objeto de correlación como de la VD. Es decir, es una correlación entre residuos.

En el modelo de dos variables, si se ajusta una recta entre Y y X_2, y nos quedamos con los residuos, y si se ajusta una recta entre X_1 y X_2, y nos quedamos también con los residuos, podemos correlacionar ambos residuos. De esta forma obtendremos la correlación parcial entre Y y X_1. A partir de aquí se ve claro que esta es la correlación "pura" entre dos variables, puesto que de ambas se ha extraído el influjo de terceras variables. Al igual que en la correlación semiparcial, no es necesario el cálculo de los residuos, pues se pueden obtener a partir de los correlaciones de orden cero entre pares de variables.

$$pr_1 = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{(1 - r_{y2}^2)(1 - r_{12}^2)}} \qquad\qquad pr_2 = \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{(1 - r_{y1}^2)(1 - r_{12}^2)}}$$

El cuadrado de estos coeficientes (p.e. pr_1) se interpreta como la proporción de la varianza de la VD (Y) no asociada con X_2 que sí está asociada a X_1.

Otra manera de calcular esta proporción de varianza es por medio de las porciones representadas en el diagrama de Venn de la Figura 8.8.

$$\begin{aligned} pr_1^2 &= \frac{a}{a+d} = \frac{R_{y.12}^2 - r_{Y2}^2}{1 - r_{Y2}^2} \\ pr_2^2 &= \frac{b}{b+d} = \frac{R_{y.12}^2 - r_{y1}^2}{1 - r_{y1}^2} \end{aligned} \tag{8.35}$$

Aplicando las fórmulas a los datos del ejemplo, los coeficientes son:

$$pr_1 = \frac{0{,}441 - (0{,}628)(-0{,}043)}{\sqrt{1 - 0{,}628^2}\sqrt{1 - (-0{,}043)^2}} = 0{,}6018 \ \rightarrow pr_1^2 = 0{,}6018^2 = 0{,}3622$$

[3] Hay varios métodos para la introducción de variables en el análisis de regresión. Uno de estos métodos es el denominado *Stepwise* (Pasos Sucesivos) y en él se introduce en primer lugar la variable con mayor correlación con el criterio, y a partir de ahí, sucesivamente la variable que mayor correlación parcial tenga con el criterio. El proceso de introducción de variable se detiene cuando la siguiente variable independiente que va a entrar no aporta un plus significativo a la explicación de la VD.

$$pr_2 = \frac{0{,}628 - (0{,}441)(-0{,}043)}{\sqrt{1-0{,}441^2}\sqrt{1-(-0{,}043)^2}} = 0{,}7219 \quad \rightarrow pr_2^2 = 0{,}7219^2 = 0{,}5211$$

Si se hubiera realizado una regresión paso a paso, es decir, introduciendo las variables por su relación con la VD, la primera que habría entrado en el modelo hubiera sido la variable X_2 (en el ejemplo, Razonamiento abstracto) que es la que presenta mayor correlación con la VD.

En resumen, por los resultados del coeficiente de correlación parcial y semiparcial al cuadrado en el modelo obtenido, está clara la contribución de ambas variables a la explicación de la puntuación en matemáticas. El cuadrado de los coeficientes *pr* señala la proporción de varianza de una VI asociada con la parte de la VD que no está asociada con la otra VI. En nuestro caso es mayor la de razonamiento abstracto que la de tiempo de estudio (52,11% y 36,22%, respectivamente). Además, el modelo es bueno (luego veremos su significación estadística, por medio de los contrastes) porque ambas variables independientes tienen una buena relación con la dependiente, y sin embargo, entre ellas no hay apenas relación (es, pues, un modelo casi ideal[4]). ¿Cómo se manifiesta numéricamente la ausencia de relación entre las variables independientes?, pues sencillamente en que el coeficiente de determinación, R^2 (0,6141), tiene un valor aproximado (siempre menor) que la suma de los cuadrados de los coeficientes de correlación semiparcial (0,2191+0,4200 = 0,6391 < 0.6141). La diferencia entre ambos valores es la parte redundante del diagrama de Venn (zona *c*) que el modelo de regresión elimina cuando se ajusta con el conjunto completo de variables independientes.

8.5.- Resumen

El análisis de correlación y regresión trata de determinar cómo un conjunto de variables, que llamamos independientes, predictoras o explicativas, pueden explicar el comportamiento de la variable objeto de estudio, que llamamos dependiente o criterio. Ello se ha realizado en tres pasos:

- Ajuste del modelo de regresión para estimar la VD. Sólo se han tratado ajustes de modelo lineales, es decir, modelos en que la VD es una función lineal de la o las VI's. Cuando sólo hay una VI, el modelo se conoce como de **Regresión Lineal Simple** y cuando hay varias VI's, como de Regresión Lineal Múltiple.
- Cálculo de la bondad del modelo ajustado. El estadístico que cuantifica el ajuste se denominado **coeficiente de determinación** y su valor oscila entre 0 y 1, e informa de la proporción en que la o las VI's explican la VD. En el caso de la regresión simple, este valor es el cuadrado del coeficiente de correlación de Pearson, y en el caso de la regresión múltiple este valor es el cuadrado del coeficiente de correlación múltiple. La parte no explicada por el modelo de regresión es aquella que no está relacionada linealmente con la VD.
- Contraste de significación de los estadísticos del modelo en el caso de la regresión lineal simple.

[4] Los datos del ejemplo son ficticios y han sido simulados para lograr este efecto de correlación media-alta de las variables predictoras con la VD y ausencia de correlación entre las predictoras. En análisis de regresión, cuando las VI's correlacionan se dice que hay "colinealidad", y cuanto mayor es ésta peor es el modelo de regresión.

8.6.- Ejercicio de Autoevaluación

Todas las preguntas están relacionadas con datos de una investigación (ficticia, con datos simulados) en la que se trata de determinar la influencia que sobre el resultado en las pruebas para acceder a un puesto de trabajo especializado tienen una serie de variables, como son los días que asisten a tutoría en una escuela de formación para ese tipo de profesionales (variable X_1), y la expectativa de empleo que manifiestan los sujetos (variable X_2), variables todas ellas cuantitativas o métricas. Como variable dependiente se toma, como se ha señalado, el resultado en una prueba en términos de puntuación obtenida (variable Y). Los datos de 25 personas son los siguientes:

X_1	X_2	Y
31	9	108
41	6	86
20	9	80
41	7	79
40	9	96
28	9	79
41	9	98
37	8	86
41	6	89
39	11	92
56	9	111
43	11	102
42	10	89
36	7	90
36	13	112
32	7	83
49	8	104
45	11	98
20	10	88
33	11	106
39	13	110
19	10	92
27	12	92
17	11	81
29	13	103

Para facilitar los cálculos, presentamos los estadísticos descriptivos de cada variable, y la matriz de correlaciones.

Estadísticos descriptivos

	X_1	X_2	Y
Suma	882	239	2354
Media	35,28	9,56	94,16
Desv. Típica	9,5143	2,0412	10,3293
Varianza	90,5216	4,1664	106,6944

Matriz de correlaciones de orden cero

	X_1	X_2	Y
X_1		-0,231	0,436
X_2			0,504
Y			

Preguntas

1. ¿Cuál es la ecuación de regresión para la predecir el comportamiento de la variable *Y* a partir de la variable *X_1*? A) Y' = 77,465 + 0,473X_1; B) Y' = 35,465 + 0,573X_1; C) Y' = 77,465 + 0,743X_1

2. ¿Cuál es la ecuación de regresión para la predecir el comportamiento de la variable Y a partir de la variable X_2? A) Y' = 44,236 + 1,873X_2; B) Y' = 69,768 + 2,551X_2; C) Y' = 77,465 + 0,743X_1

3. El coeficiente de correlación múltiple del modelo Y' = B_0 + B_1X_1 + B_2X_2 para los datos propuestos es: A) 0,874; B) 0,759; C) 0,576

4. El coeficiente R^2 ajustado para los datos es: A) 0,594; B) 0,512; C) 0,538

5. Siguiendo el método de Pasos Sucesivos (Stepwise) para lograr el mejor ajuste, ¿qué cambio se produce en R^2 cuando se incorpora la primera variable? A) 0,322; B) 0,254; C) 0,222

6. La ecuación de regresión múltiple estandarizada para los datos es: A) $z'_y = 0{,}423z_1 + 1{,}436z_2$; B) $z'_y = 1{,}014z_1 + 0{,}872z_2$; C) $z'_y = 0{,}583z_1 + 0{,}639z_2$

7. La correlación entre la variable dependiente *Y* y la predictora *X_1*, una vez que se ha eliminado el influjo de X_2 sobre ambas variables, es: A) 0,659; B) 0,567; C) 0,621

8. ¿Cuál es la proporción de la varianza de *Y* asociada a *X_2*, y no asociada a *X_1*. A) 0,234; B) 0,342; C) 0,477

Solución ejercicios de autoevaluación

Debajo de las respuestas están las operaciones necesarias, a partir de los estadísticos y la matriz de correlaciones.

Pregunta 1 A

$$B_1 = r_{y1}\frac{S_y}{S_{y_1}} = 0{,}436\frac{10{,}3293}{9{,}5143} = 0{,}473$$

$$B_0 = \bar{Y} - B_1\bar{X}_1 = 94{,}16 - (0{,}473)(35{,}28) = 77{,}465$$

Pregunta 2 B

$$B_1 = r_{y2}\frac{S_y}{S_{x_2}} = 0{,}504\frac{10{,}3293}{2{,}0412} = 2{,}5514$$

$$B_0 = \bar{Y} - B_1\bar{X}_2 = 94{,}16 - (2{,}5514)(9{,}56) = 69{,}768$$

Pregunta 3. B

$$R_{y.12} = \sqrt{\frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2}} = \sqrt{\frac{0{,}436^2 + 0{,}504^2 - 2(0{,}436)(0{,}504)(-0{,}231)}{1 - (-0{,}231)^2}} = 0{,}759$$

Pregunta 4. C

$$\hat{R}_{y.12}^2 = 1 - \left(1 - R_{y.12}^2\right)\frac{n-1}{n-p-1} = 1 - (1 - 0{,}759^2)\frac{25-1}{25-2-1} = 0{,}538$$

Pregunta 5. A

$$R_{y.12}^2 - r_{y2}^2 = 0{,}759^2 - 0{,}504^2 = 0{,}322$$

El método *Stepwise*, la primera variable en entrar en el modelo sería la X_2 pues es la que más correlaciona con Y

Pregunta 6. C

$$\beta_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2} = \frac{0{,}436 - (0{,}504)(-0{,}231)}{1 - (-0{,}231)^2} = 0{,}583$$

$$\beta_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2} = \frac{0{,}504 - (0{,}436)(-0{,}231)}{1 - (-0{,}231)^2} = 0{,}639$$

Pregunta 7. A

Se trata del coeficiente de correlación parcial entre las variable *Y* y X_1.

$$pr_1 = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{1 - r_{y2}^2}\sqrt{1 - r_{12}^2}} = \frac{0{,}436 - (0{,}504)(-0{,}231)}{\sqrt{1 - (0{,}504)^2}\sqrt{1 - (-0{,}231)^2}} = 0{,}657$$

Pregunta 8. C

$$pr_2^2 = \left(\frac{r_{y2} - r_{y1}r_{12}}{\sqrt{1 - r_{y1}^2}\sqrt{1 - r_{12}^2}}\right)^2 = \left(\frac{0{,}504 - (0{,}436)(-0{,}231)}{\sqrt{1 - (0{,}436)^2}\sqrt{1 - (-0{,}231)^2}}\right)^2 = 0{,}477$$

2013/2014

UNED

ANÁLISIS NO PARAMÉTRICO

[TEMA 9]

Índice

9.1.- Introducción: contrastes paramétricos vs. no paramétricos

Los procedimientos estudiados para el análisis de los diferentes tipos de diseños en el resto de capítulos del texto se han basado en la asunción de una serie de supuestos (*v.g.*, que la muestra procede de una población de puntuaciones que se distribuyen según la curva normal, o que el nivel de medida de la variable dependiente era de intervalo o de razón, etc.). Todos estos procedimientos de análisis se engloban en lo que se conoce como "*métodos paramétricos*" ya que todos los cálculos matemáticos que se realizan (sumatorio de puntuaciones, divisiones, potenciación, etc.) dependen del supuesto de que las puntuaciones proceden de una familia de distribuciones paramétrica particular. El objetivo de estos métodos es identificar a un elemento de esta familia paramétrica como el candidato más probable del que proceden nuestras observaciones.

Veamos varios ejemplos de lo que estamos diciendo, haciendo especial hincapié en el significado que tiene en la anterior definición el concepto de "paramétrico". Si en un contraste de hipótesis realizamos el supuesto de distribución según la curva normal, tendríamos toda una familia (el concepto de "familia" en matemáticas es idéntico al de "conjunto") de funciones de densidad de probabilidad (*f.d.p.*) que tienen la misma forma (campana de Gauss). Por consiguiente, cada elemento de este conjunto se diferenciaría del resto no por su forma (todas son normales) sino por sus parámetros μ y σ (es decir, por su media y su desviación típica). La estadística paramétrica funciona realizando supuestos sobre los valores de μ y σ. Por ejemplo, hemos visto que suponiendo la normalidad y asumiendo que la varianza de dos poblaciones de puntuaciones son idénticas, se puede poner a prueba la hipótesis de que la μ de una población tiene un valor concreto. En este sentido, podemos decir que la estadística paramétrica toma en consideración un espacio de parámetros o "espacio paramétrico". En el caso del supuesto de normalidad este espacio paramétrico vendría dado por todos los valores posibles para μ y σ. Este espacio se puede representar gráficamente mediante el área que resulta de la intersección entre los intervalos $(-\infty;+\infty)$ y $[0;+\infty)$ respectivamente (véase la Figura 9.1). Es decir, si el eje de abscisas representa el valor de las medias y el eje de ordenadas representa el correspondiente a las desviaciones típicas, y recordando que la desviación típica no puede ser negativa, entonces la media puede ser, teóricamente, cualquier valor entre +Infinito y –Infinito, que se representa como el intervalo $(-\infty;+\infty)$, mientras que la varianza, al tener que ser igual o superior a cero vendría dada por el intervalo de valores $[0;+\infty)$ en donde se incluye teóricamente el 0. En la Figura 9.1 se ha representado este "espacio" de valores posibles para los parámetros μ y σ, para este ejemplo particular, como el área en blanco. El área inferior dibujada en gris está prohibida porque representaría σ negativos. Dado un conjunto de datos empíricos (muestrales), el objetivo de la estadística paramétrica sería encontrar aquellos parámetros poblacionales (μ y σ) de la curva normal que mejor ajustan los datos.

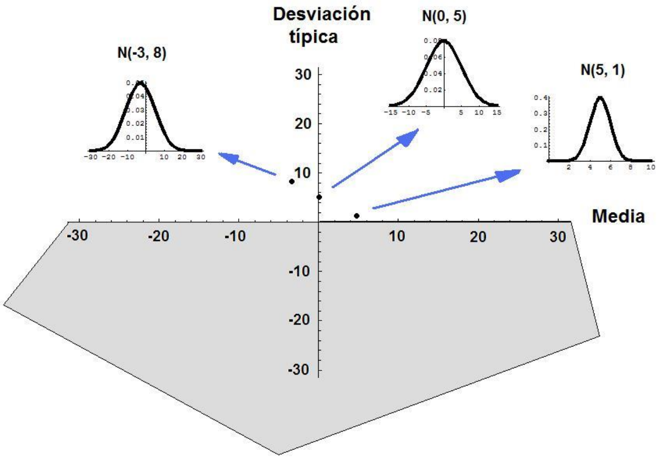


Figura 9.1: Espacio paramétrico de funciones normales. El eje de abscisas representa la media (μ) y el eje de ordenadas representa la desviación típica (σ) de funciones gaussianas. La mitad horizontal inferior no pertenece al espacio paramétrico debido a que representa valores negativos para la desviación típica, y esto es imposible. Cada punto del espacio paramétrico factible representa una curva normal concreta con sus valores μ y σ . En el gráfico se han representado tres puntos y sus correspondientes curvas.

La anterior discusión ha tratado de mostrar gráficamente que los <u>parámetros</u> de una distribución normal son μ y σ así como que los métodos englobados dentro de la estadística paramétrica tratan de identificar los valores poblacionales más probables para estos parámetros dados los valores de la variable dependiente observados en la muestra (es decir, dados los datos). A continuación traducimos la anterior discusión en términos analíticos. Sabemos que cualquier distribución normal viene dada por la ecuación 1:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{Ecuación 1}$$

En esta ecuación podemos identificar:

- Un conjunto de constantes (1, 2, $\pi \approx 3.14159$, $e \approx 2.71828$)
- El <u>argumento</u> de la función (x), que representa la "variable independiente" (gráficamente es el eje de abscisas).
- Dos valores que pueden variar de una función a otra (identificados mediante los símbolos μ y σ).
- El resultado de la función $f(x)$ que gráficamente es el eje de ordenadas.

En cualquier función de densidad de probabilidad (y la Ecuación 1 muestra una de ellas, la correspondiente a la f.d.p. de la curva normal), una vez que hayamos identificado y descartado las constantes y el argumento (o argumentos en el caso de que la f.d.p. sea multidimensional) de la función, lo que resta son los parámetros: valores numéricos que identifican a la función. Por ejemplo, si supiéramos que μ vale 15 y que σ vale 5, la Ecuación 1 estaría completamente especificada y tendría la expresión:

$$f(x) = \frac{1}{5\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-15}{5}\right)^2}$$

Ahora, una vez dada un valor concreto del argumento (x), podríamos operar y calcular su imagen [f(x)].

En la Figura 9.2 aparece la concreción, en el espacio paramétrico de valores de μ y σ, de esta función (representado por el punto en rojo) así como la forma concreta de la función de densidad de probabilidad para este caso.

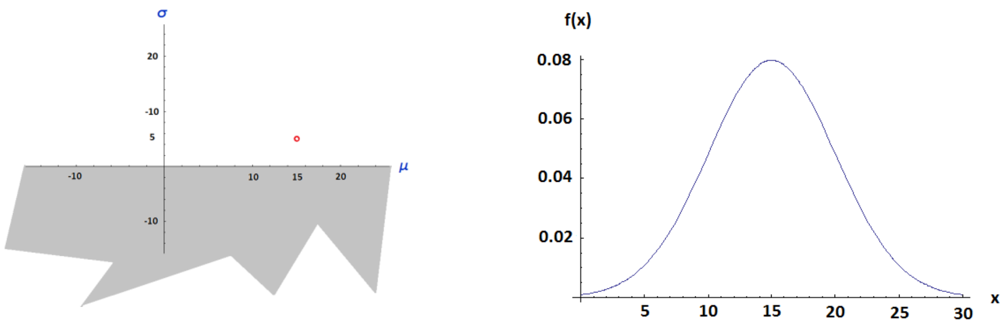


Figura 9.2: A mano izquierda podemos ver, en el espacio paramétrico de valores de μ y σ, el punto que define la f.d.p. con $\mu = 15$ y $\sigma = 5$(en rojo). A mano derecha podemos ver la forma concreta de la curva normal para estos mismos valores de μ y σ.

En esta situación los parámetros han adquirido un valor concreto y el único símbolo sin especificar numéricamente que quedaría en la ecuación sería el argumento (x) de la función. El resto serían constantes y parámetros. Esta capacidad de los parámetros para identificar la expresión concreta de la f.d.p. les confiere una gran importancia en la estadística paramétrica y el interés que tienen los métodos que tratan de encontrar qué valores numéricos tienen los parámetros en la población dados los valores de las puntuaciones muestrales.

Ahora nos podemos preguntar por qué la mediana no aparece como un parámetro de la distribución normal. Sabemos que toda función normal tiene mediana (o cualquier otro estadístico de sesgo, curtosis, etc., que se estudiaron en Estadística Descriptiva). De forma sucinta podemos decir que la mediana no es un parámetro de la función normal (ni ningún otro estadístico de posición) porque no aparece en la expresión analítica que la define. Es un valor numérico que caracteriza a la distribución de la que procede y que puede calcularse a partir de procedimientos matemáticos conociendo la fórmula pero que no aparece explícitamente en la misma fórmula.

Pongamos otro ejemplo de familia paramétrica para comprender el concepto de estadística paramétrica y, por contraposición, la estadística no paramétrica. Si asumimos que el fenómeno que estamos estudiando se distribuye según una distribución Binomial, entonces tendremos que realizar una búsqueda en el espacio de todas las funciones binomiales para saber de qué función binomial concreta proceden nuestros datos. Toda función binomial tiene dos parámetros: *n* y *p* (el número de ensayos o veces que se repite un experimento de Bernouilli y la probabilidad de éxito en cada ensayo asumiendo que se mantiene constante, respectivamente). Sabemos que *n* tiene que ser un número natural (1, 2, 3, …) y es igual al número de elementos muestreados (es decir, el tamaño muestral). Por su parte, *p* se encuentra en el intervalo cerrado [0, 1] ya que se trata de una probabilidad y puede tomar cualquier valor entre 0 y 1, incluyendo estos valores. Por ello el espacio de búsqueda paramétrico viene dado por los intervalos [0, 1] para cada *n* natural. En la Figura 9.2 se ha representado este espacio paramétrico mediante un conjunto de líneas verticales que se extienden verticalmente desde 0 hasta 1 (parámetro *p*) y que parten de cada valor natural del eje de abscisas (parámetro *n*). Dado un valor concreto de n, el espacio de búsqueda sería el intervalo desde 0 hasta 1. Obsérvese que cualquier valor externo a estas líneas no forma parte del espacio paramétrico de la f.d.p Binomial así como que hemos terminado el diagrama en n=15 por simplificar la Figura pero que, en realidad, se extendería hacia la derecha de forma infinita.

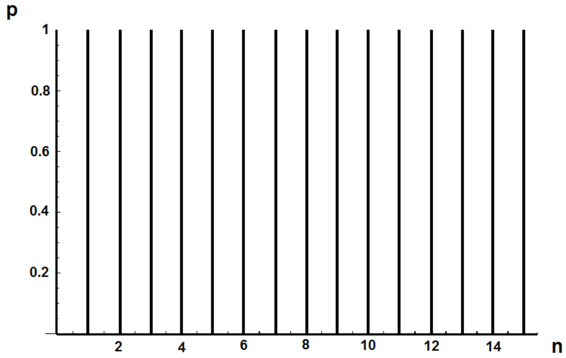


Figura 3: Espacio paramétrico de las funciones binomiales. Para cada valor de *n* (número de datos o tamaño muestral), el objetivo de la estadística paramétrica sería encontrar el valor *p* que mejor ajusta los datos. Este valor *p*, como probabilidad que es, se encuentra entre 0 y 1.

De la misma forma que antes, podemos decir que *p* y *n* son los parámetros de la distribución Binomial porque aparecen en su expresión analítica (véase Ecuación 2):

$$f(x) = \binom{n}{x} p^x q^{n-x} \qquad \text{Ecuación 2}$$

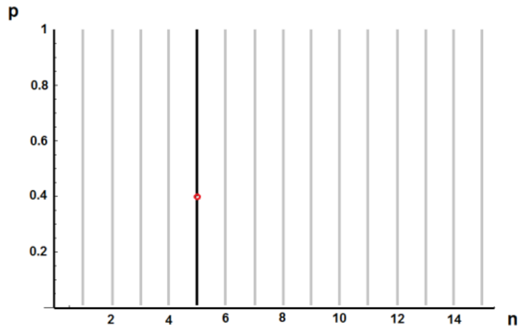
En esta expresión podemos observar:

- El argumento de la función (x), con la limitación de que sólo puede adoptar valores naturales (recuérdese que los números naturales son el conjunto infinito {1, 2, 3, …} mientras que los números enteros incluyen el cero y los valores negativos por lo que el argumento no puede adoptar ciertos valores del conjunto de los enteros).
- Los parámetros *n* y *p*.
- El valor q es una simple derivación de p ya que q = 1 – p, y por lo tanto no se considera un parámetro.

Podemos decir que la aparición de una variable que no sea la variable independiente (o argumento) o una constante en la expresión analítica de una función de densidad de probabilidad (f.p.d) la caracteriza como un **parámetro** de esa distribución porque define a la distribución. Por eso, en la función de probabilidad de la Binomial podemos observar que son parámetros *n*, *p* y *q* ya que x es la variable independiente o argumento. Por su parte, como p (probabilidad de éxito) y *q* (probabilidad de fracaso) están relacionadas mediante la expresión p = 1 – q, los parámetros en realidad se reducen a dos: *n* y *p*. Por su parte, los parámetros caracterizan unívocamente a la función: dados valores numéricos concretos para los parámetros, estos identifican un punto del espacio paramétrico y una expresión analítica concreta. En el ejemplo anterior, si supiéramos que con un tamaño muestral de 5 (*n* = 5), *p* valiese 0.4, entonces la Ecuación 2 sería:

$$f(x) = \binom{5}{x} 0.4^x 0.6^{5-x}$$

la cual sólo depende de valores numéricos constantes (5, 0.4 y 0.6) y del argumento (x). Gráficamente la f.d.p. resultante puede verse en la Figura 9.3:



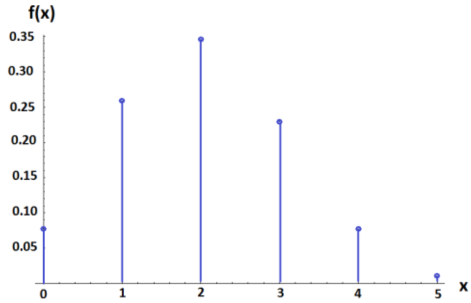


Figura 9.4: A mano izquierda puede verse la especificación de la Binomial con parámetros n = 5 y p = 0.4 en rojo y en su espacio paramétrico (al haberse especificado n = 5, se han dibujado el resto de líneas verticales en gris para señalar que no son valores permisibles ya que se corresponden con otros valores de n). A mano derecha puede verse la función de probabilidad concreta especificada por los anteriores parámetros.

Este sería el objetivo de toda la estadística paramétrica: conociendo los valores muestrales obtenidos en un estudio (los datos) y realizando un conjunto de supuestos razonables, encontrar los valores concretos de los parámetros que caracterizan la distribución poblacional de la que se han extraído estos valores. En definitiva, la denominación de "*técnicas paramétricas*" procede de la búsqueda de los *parámetros* subyacentes a unos datos asumiendo que éstos se distribuyen en la población según una función de probabilidad o de densidad de probabilidad concreta (hablaremos genéricamente de f.d.p. o función de densidad de probabilidad). Todos los tests paramétricos asumen una determinada forma (normal, binomial, etc.) para la distribución poblacional de los datos observados en la muestra (variable dependiente) y esta forma depende de unos parámetros distintos y propios de cada f.d.p., como hemos visto previamente (la normal depende de los parámetros μ y σ^2 mientras que la Binomial depende de los parámetros n y p).

Pero a veces nos encontramos con datos, poblaciones o situaciones en que no podemos asumir los supuestos subyacentes a los tests paramétricos y necesitamos procedimientos cuya validez no dependa de esos supuestos. En este caso parece obvio concluir que necesitamos acudir a otro conjunto de técnicas que no exijan buscar los parámetros de la distribución poblacional de los datos por la sencilla razón de que, por causas de distinta índole, no podemos asumir una forma concreta (normal, binomial, etc.) de la f.d.p.. Por contraposición a los anteriores métodos, se los conoce como "*métodos no paramétricos*". Es por ello que, de forma genérica, podemos identificar un contraste de hipótesis como "no paramétrico" porque se realiza con datos procedentes de una población en la que la variable de estudio no tiene una distribución de probabilidad **conocida**.

En relación a la diferenciación entre contrastes paramétricos y no paramétricos, el estudiante debe diferenciar entre los supuestos realizados por el método de inferencia de las pruebas (criterio que diferencia las técnicas paramétricas de las no paramétricas) de las distribuciones muestrales utilizadas en los cálculos para el contraste de hipótesis en cualquiera de estas técnicas (tanto paramétricas como no paramétricas). Aunque la estadística no paramétrica no incluya entre sus supuestos a una f.d.p. concreta como la distribución poblacional de la variable dependiente, no obstante, los estadísticos que se calculan en la estadística no paramétrica sí se distribuirán según una u otra distribución paramétrica concreta. Por ejemplo, cuando veamos en este mismo tema la prueba de los signos comprobaremos que no se hace ningún supuesto sobre cómo se distribuyen los datos observados en la población de la cual se ha extraído la muestra de estudio (en este sentido es un contraste no paramétrico) pero, sin embargo, utilizará la distribución binomial (como distribución muestral del estadístico) para realizar los cálculos de los niveles de probabilidad asociados a la hipótesis nula (véase la Figura 9.5). La distribución binomial es una distribución paramétrica pero la prueba de los signos no la utiliza como un *supuesto* necesario para poder aplicarla, sino como una herramienta para calcular niveles de probabilidad.

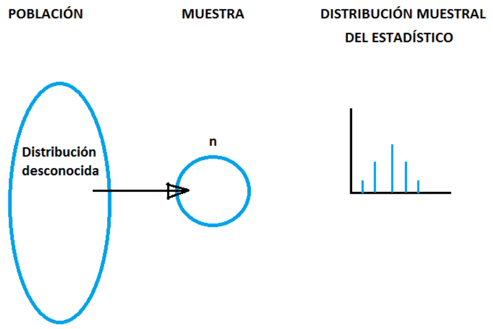


Figura 9.5: la característica de las pruebas no paramétricas es que no hacen supuestos sobre la forma poblacional de la variable (distribución paramétrica desconocida, parte izquierda de la figura) incluso aunque en su análisis de la probabilidad del valor del estadístico obtenido en la muestra haga uso de distribuciones muestrales paramétricas (parte derecha de la figura).

Otro aspecto importante a señalar es que todos los tests no paramétricos que se estudiarán a continuación no hacen uso de los valores de las puntuaciones muestrales observadas sino de estadísticos calculados usualmente a partir de sus rangos u órdenes. Es por ello que usualmente se trabaja con la mediana (valor de la variable observada que deja por debajo de sí el 50% de los casos). Para solventar los pocos supuestos que realizan, estos tests no trabajan con las puntuaciones originales o directas sino con puntuaciones de categoría inferior a estas (los rangos). Esta característica va a implicar que los tests no paramétricos tendrán menor potencia que sus correspondientes paramétricos y será más difícil rechazar H_0 siendo falsa.

Nota: el punto 9.2 no entra en el temario debido a que se mete en disquisiciones bastante sutiles, pero desde mi punto de vista, resulta interesante leerlo al menos.

9.2.- Tests no paramétricos: ¿una distinción sencilla?

Sin embargo, aunque la distinción anterior entre los procedimientos paramétricos y no paramétricos pueda parecer directa y simple (tests paramétrico: aquel en que después de realizar el supuesto razonable de que, en la población, la variable bajo estudio se distribuye según una función de densidad concreta, buscamos los parámetros que la caracterizan; tests no paramétrico, aquellos en donde esta condición no se cumple), resulta difícil proporcionar una definición más precisa del concepto de "*técnicas no paramétricas*" ya que no existe un consenso absoluto al respecto. La explicación anterior ha tratado de aclarar pedagógicamente la diferencia entre ambos tipos de técnicas, incidiendo en la aclaración de los términos utilizados (paramétrico *vs*. no paramétrico). No obstante, la literatura técnica diferencia estas técnicas de formas muy diversas. Así, por ejemplo, Ross (2004) las define como aquellas pruebas que se aplican cuando los datos proceden de una distribución de probabilidad cuya forma no viene especificada, es decir, cuando no podemos asumir una forma concreta de la distribución poblacional para los datos (normal, exponencial, binomial, etc.). En este caso, no podemos buscar sus parámetros. Esta es, básicamente, la definición que hemos utilizado en nuestra exposición anterior. Es por ello que muchos autores también denominan a estas pruebas como "pruebas sin supuesto distribucional" (*distribution-free*) en vez de "no paramétricas".

No obstante, a veces se generaliza inadecuadamente la anterior definición y se afirma que los tests no paramétricos no hacen ningún tipo de supuesto (obsérvese que en la definición anterior no se implicaba esto ya que solo se eliminaba el supuesto de una distribución concreta). Esta idea es incorrecta ya que en este caso se debería hablar de tests libres de supuestos (*assumption-free*) y no de tests no paramétricos. No obstante, la verdad que subyace a esta sutil distinción (la que diferencia a los tests que no realizan supuestos distribucionales –*distribution-free*– y aquellos libres de supuestos –*assumption-free*–) es que los supuestos realizados por los contrastes no paramétricos son menos rígidos que los realizados por los contrastes paramétricos (Marascuilo y McSweeney, 1977). Es decir que los contrastes no paramétricos sí realizan supuestos aunque estos no están referidos a la forma específica que adopta la distribución poblacional. Por ejemplo, un supuesto muy utilizado en algunas pruebas no paramétricas es el supuesto de distribución simétrica de la variable dependiente en la población. Debido a que existen otras muchas distribuciones distintas de la normal que también son simétricas, este supuesto no nos limita a trabajar con una distribución normal (que es simétrica). No obstante, sí nos elimina del área de búsqueda todas las distribuciones asimétricas (*v.g.*, la F o la chi-cuadrado). Esta es, quizás, una de las razones por las que algunos autores han introducido otros términos que pretenden establecer matizaciones a medio camino entre las técnicas paramétricas y las no paramétricas (*v.g.*, técnicas semiparamétricas).

Otros autores que se enfrentan a la tarea de diferenciar entre técnicas paramétricas y no paramétricas hacen hincapié simplemente en que, de forma genérica, estas últimas realizan supuestos menos restrictivos o rígidos que las técnicas paramétricas (Daniel, 1990). Desde esta perspectiva, la distinción entre paramétrico *vs*. no paramétrico no es una distinción cualitativa sino cuantitativa. Hay procedimientos que se acercan más al extremo paramétrico y otros al extremo no paramétrico, existiendo otros en puntos intermedios del continuo. Aunque esta clasificación cuantitativa de los contrastes es una definición más vaga que las anteriores, resulta muy útil ya que existen técnicas estadísticas entre cuyos objetivos iniciales explícitos se encuentra el realizar los menos supuestos posibles sobre la procedencia de los datos (v.g., el Análisis de Componentes Independientes o ICA).

Una tercera opinión insiste en que lo que caracteriza a las técnicas no paramétricas es el nivel de medida de los datos. Las técnicas no paramétricas se suelen utilizar cuando las escalas utilizadas para medir la variable dependiente, es decir, los datos recogidos en la muestra, son de tipo nominal u ordinal o bien cuando las escalas sean de tipo de intervalo/razón pero han sido recodificadas en variables de tipo nominal u ordinal. Esto no quiere decir que no puedan utilizarse técnicas no paramétricas para variables medidas en escalas de intervalo o razón pero la disminución de potencia que esto produciría en relación a las técnicas paramétricas no recomienda esta opción.

Nosotros estamos de acuerdo con Wasserman (2006) cuando subraya que el punto esencial de las técnicas no paramétricas consiste en que los métodos estadísticos desarrollados en este área tratan de mantener los supuestos lo menos restrictivos o rígidos posibles. Hemos de señalar que cuanto menos restrictivos sean los supuestos que se realicen a la hora de seleccionar la prueba que se aplicará para el contraste de hipótesis, más amplias serán las posibilidades que tendremos que contemplar. Esto se comprende fácilmente si se observa que cuando hacemos el supuesto de que la distribución de los datos se distribuyen en la población según la curva normal, sólo tendremos que buscar dentro de un espacio paramétrico de valores μ y σ relativamente restringido (véase la Figura 9.1). Sin embargo, si no hacemos restricción alguna sobre la forma de la distribución de las puntuaciones, como sucede en las técnicas no paramétricas, tenemos que admitir que estamos buscando en un espacio mucho más amplio que incluye no solo funciones normales sino exponenciales, logarítmicas, etc. Este espacio puede llegar a ser infinito por lo que también se conoce a las técnicas no paramétricas como "técnicas paramétricas de dimensionalidad infinita" (*infinite-dimensional parameter* -Arkadi Nemirovski, 2000-). Esta definición suaviza la opinión de que la denominación de "técnicas no paramétricas" ha sido desafortunada (Noether, 1991). Lo que sí es cierto es que el concepto de "no paramétrico" no tiene una definición precisa y universalmente aceptada.

Nota: repetimos que la información contenida en este recuadro (9.2) es de ampliación de conocimientos y no formará parte de la evaluación de la asignatura.

Nota importante: al igual que el contenido del apartado 9.2, el contenido del resto del capítulo es de estudio opcional ya que no entrará como materia de examen ni como parte de ninguna de las PECs. No obstante, recomendamos al estudiante que guarde estos apuntes, incluso aunque no los estudie ahora, porque en su trabajo profesional es muy posible que se encuentre trabajos que utilizan estadísticos no paramétricos. Estos son usuales en Psicología Clínica ya que en esta se trabaja con poblaciones especiales con pocos sujetos y en donde, por tanto, se incumplen los supuestos paramétricos.

9.3.- Contrastes no paramétricos para diseños de una muestra

Como se ha explicado previamente, aunque la "estadística no paramétrica" no asuma ciertos supuestos referidos a la distribución de la variable en la población, los estadísticos de contraste que se aplican para calcular la discrepancia entre la teoría y los datos sí se distribuirán de acuerdo a una distribución de probabilidad determinada que nos informará de la máxima discrepancia (valores críticos) que, a un determinado nivel de confianza, se puede admitir entre los datos teóricos y los empíricos y el nivel p-crítico asociado a ese estadístico de contraste. En la siguiente prueba, aplicable para los diseños de una única muestra, veremos que el estadístico de contraste que se aplica en estos casos se ajusta a la distribución binomial. Se trata de la prueba de los signos.

9.3.1.- La prueba de los signos

Es una prueba no paramétrica equivalente a la prueba *t* de una muestra que se aplica cuando no se cumplen los supuestos que ésta exige. Supongamos que nos disponemos a recoger un conjunto de *n* datos procedentes de una muestra. Las unidades de observación, es decir, los elementos a los que someteremos a medición, pueden ser personas, colegios, ratas, amebas, etc. dependiendo solamente de nuestra área de investigación. Para cada unidad de observación realizaremos la medición de una variable en la que estamos interesados (*v.g.*, CI en personas, nivel de ruido ambiental en colegios, número de ensayos para aprender una tarea en ratas, etc.). Como en el caso anterior, la variable que midamos la denotaremos genéricamente por X.

El conjunto de observaciones procederá de una población de puntuaciones que siguen una función de distribución continua pero desconocida a la que denominaremos D. Pero si no conocemos ni podemos asumir la forma concreta de D, entonces no podemos utilizar las pruebas paramétricas. Tendremos que utilizar, por consiguiente, pruebas no paramétricas. Esto quiere decir que debemos admitir que D puede adoptar cualquier forma (normal, exponencial, gamma, etc.) y que, además, el procedimiento estadístico que utilicemos no puede depender de la forma concreta que adopte la distribución D de la variable en la población. En consecuencia, el procedimiento que utilicemos debe ser válido para cualquier forma que pueda adoptar la distribución D en la realidad.

En estos casos, en vez de elegir un parámetro para realizar el contraste (cosa imposible de hacer ya que si no especificamos una distribución concreta, no podemos saber los parámetros de los que depende), se suele poner a prueba un estadístico de posición, en concreto, la mediana. Ya hemos visto que de toda distribución se puede calcular su mediana aunque no se conozcan sus parámetros y sabemos que la mediana es un índice de tendencia central. Así que, supongamos que estamos interesados en poner a prueba la hipótesis de que la mediana poblacional de D, a la que llamaremos η (símbolo griego eta minúscula) es igual a un valor específico al que llamaremos m_0. Es decir, estamos poniendo a prueba el siguiente contraste de hipótesis para la mediana poblacional:

$$H_0 : \eta = m_0$$
$$H_1 : \eta \neq m_0$$

Si recordamos el concepto de mediana y asumiendo que H_0 es cierta, esto significa que en el valor m_0 (un valor concreto de la variable medida X que queremos contrastar) aproximadamente la mitad de los datos de la muestra deben quedar por debajo del mismo y la mitad restante por encima, excepto variaciones debidas al azar. En otras palabras, en la hipótesis nula estamos planteando que m_0 es el percentil 50 de la distribución (la mediana). Y esto sin importar el tipo de distribución concreta que pueda adoptar D en la realidad.

Como hemos dicho, es esencial darse cuenta de que **si H_0 es cierta** entonces cada observación será menor que m_0 con probabilidad 0.5 (o superior a m_0 con probabilidad 1- 0.5 = 0.5). Es decir, que si H_0 es cierta la probabilidad de que cada observación sea superior o inferior a m_0 será similar al experimento aleatorio consistente en arrojar una moneda insesgada (no trucada) y determinar si cae cara o cruz. Ya se vio en la asignatura de Introducción al Análisis de Datos que el modelo estadístico que subyace a un proceso dicotómico del tipo cara-cruz, repetido varias veces (*n* ensayos), se modela mediante una variable aleatoria de Bernouilli con parámetro *p* (probabilidad de éxito). En nuestro caso, el valor de *p* deberá ser 0.5 siempre y cuando H_0 sea cierta. En consecuencia, partiendo de los valores de X_i obtenidos empíricamente en el experimento u observación, podemos calcular otra variable aleatoria, a la que llamaremos I, que se distribuirá según Bernouilli según la fórmula:

$$I_i = \begin{cases} - & si \quad X_i < m_0 \\ + & si \quad X_i \geq m_0 \end{cases}$$

Esto significa que cada valor concreto de X_i se transforma en un signo negativo si es menor que m_0 (el valor propuesto para la mediana en H_0) o en un signo + si es igual o superior a m_0. El valor de I_i nos señala si ese valor se encuentra por encima o por debajo del valor m_0 propuesto en H_0. Si H_0 es cierta, entonces la nueva variable aleatoria I_i será una variable aleatoria de Bernouilli con $p = 0.5$. Ahora podemos sumar el número de signos negativos y positivos para toda la muestra. Llamemos S_- al estadístico resultante de contabilizar el número de signos negativos y S_+ al estadístico resultante de contabilizar el número de signos positivos. Utilizaremos como estadístico S el menor de estos valores (y, por tanto, el más discrepante con H_0). Este valor S se distribuye según la binomial con parámetros *p* y *n* (en donde *p* valdrá 0.5 si H_0 es cierta y *n* será el número de datos de la muestra). En la Tabla de la Binomial (Tablas I y II) podemos buscar la probabilidad de obtener este valor de S y lo compararemos con α si el contraste es unilateral y con $\alpha / 2$ si el contraste es bilateral.

Una vez explicada la lógica del test de los signos, procedemos a desarrollar más detenidamente su mecánica.

Supuestos: Los supuestos para poder aplicar correctamente el estadístico de los signos son:

- Asumimos que la variable de interés es continua. Esto significa que teóricamente no deberían producirse diferencias nulas cuando calculamos $X_i - m_0$ (es decir, que no deberíamos encontrarnos con valores de X_i y m_0 idénticos, en cuyo caso la diferencia sería 0 y no podríamos asignarle un signo negativo o positivo al mismo). No obstante, como en la práctica se producen debido a la precisión limitada de nuestros instrumentos de medida, el procedimiento usual es descartar aquellas observaciones que generen estas diferencias nulas y reducir *n* en consecuencia.
- La variable de interés viene medida, como mínimo, en una escala ordinal. Obsérvese que estamos diferenciando entre la "variable de interés" real (que asumimos continua según el supuesto anterior) y la variable tal y como la medimos (mediante una escala ordinal como mínimo). Así, por ejemplo, teóricamente podríamos considerar la "belleza" de una obra de arte como una variable continua (esto es discutible pero asumamos que podemos hacerlo) pero cuando la medimos podemos utilizar simplemente una escala que distinga únicamente tres valores ordenados: 0 ("me parece fea"), 1 ("normal") y 2 ("me parece bella"), aunque obviamente podríamos utilizar escalas ordinales con un mayor número de categorías.
- La muestra se ha extraído aleatoriamente de una población con mediana desconocida.

A continuación presentamos de forma detallada una situación y un ejemplo de los cálculos a seguir.

Ejemplo del test de los signos: Sabiendo que la distribución de los tiempos de reacción (TR) tiene una distribución asimétrica negativa, un investigador decide aplicar un contraste no paramétrico para contrastar la hipótesis de que la mediana en un experimento de TR simple es de 450 ms. Ha sometido a 10 participantes a un experimento de TR simple, calculando <u>para cada uno</u> de ellos la media de TR en 100 ensayos. Los datos obtenidos han sido

$\{X_1 = 430, X_2 = 480, X_3 = 510, X_4 = 500, X_5 = 390, X_6 = 455, X_7 = 440, X_8 = 470, X_9 = 475, X_{10} = 465\}$

El investigador publicará sus resultados en una revista científica que exige un nivel de confianza del 95%.

Condiciones y supuestos: El investigador utiliza un diseño de una única muestra y decide aplicar una prueba no paramétrica <u>debido</u> a la asimetría de la distribución de los TR y no debido al nivel de medida de la variable dependiente, que en este caso sería de razón. La variable subyacente (tiempo) es una variable continua aunque la precisión de nuestros instrumentos la puede convertir en discreta. Así, por ejemplo, si se realiza con un ordenador que tenga una precisión del milisegundo, no podremos diferenciar entre valores como 532.3 de 532.4 ya que el ordenador nos medirá en ambos casos 532 milisegundos.

Formulación de las hipótesis: La hipótesis de investigación es que la mediana de los TR en la población es de 450 ms, por lo que la hipótesis alternativa dirá que no es de 450 ms. Se trata, por tanto, de un contraste bilateral:

$$H_0 : \eta = 450$$
$$H_1 : \eta \neq 450$$

Obsérvese en este caso que el propio enunciado de la hipótesis del investigador ya nos está indicando que estamos interesados en el estadístico *mediana*, y por tanto se descartan los contrastes paramétricos.

Estadístico de contraste: Para contrastar la hipótesis, transformamos los datos observados en signos positivos y negativos, tal y como se recoge en la Tabla 9.1.

Tabla 9.1

Sujeto	X_i	$(X_i\text{-}m_0) = (X_i\text{-}450)$	Signo de las diferencias X_i y m_0
1	X_1=430	-20	-
2	X_2=480	30	+
3	X_3=510	60	+
4	X_4=500	50	+
5	X_5=390	-60	-
6	X_6=455	5	+
7	X_7=440	-10	-
8	X_8=470	20	+
9	X_9=475	25	+
10	X_{10}=465	15	+
			S_-=3; S_+=7

En la Tabla 9.1 vemos, en la segunda columna, los datos directos observados (la media del TR simple en 100 ensayos) para los 10 participantes (primera columna). En la tercera columna vemos el resultado de restar cada puntuación de la variable dependiente (TR) de la mediana planteada en H_0. En la cuarta columna sólo aparecen los signos (positivos y negativos) correspondientes a la diferencia anterior. Por último, hemos calculado el número de signos positivos y negativos que aparecen en la cuarta columna. Observamos que $S_- = 3$ y que $S_+ = 7$ (su suma nos proporciona el número de observaciones, es decir, $S_- + S_+ = 3+7=10$). Elegimos como estadístico de contraste (y lo denotamos como S) el valor más pequeño de ambos, en este caso, $S_- = 3$. Es decir, utilizaremos como estadístico S = 3.

Para contrastar nuestra hipótesis debemos preguntarnos por la probabilidad asociada con un valor de signos de S = 3 (o más extremo) según la función de probabilidad binomial con n = 10 (el número de participantes) y p = 0.5 (véase la Figura 9.6). Observamos en la Figura 9.6 que el valor más probable para S si H_0 es cierta, es X = 5. Conforme X es más extremo, la probabilidad de obtener estos valores decrece. Hemos representado en rojo los valores iguales o más extremos que S = 3 (es decir, 3, 2, 1 y 0). La suma de estas probabilidades nos responderá a la pregunta de cuál es la probabilidad de obtener un valor igual o más extremo que el obtenido (S=3) bajo el supuesto de que H_0 es cierta.

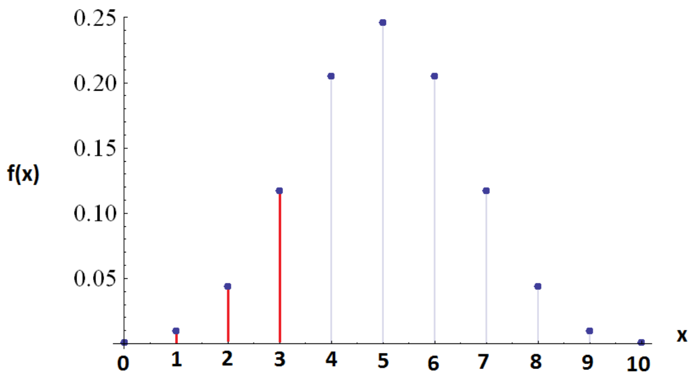


Figura 9.6: Función de probabilidad de una distribución binomial con n = 10 y p = 0.5.

La probabilidad de obtener un valor del estadístico de contraste S menor o igual al obtenido en nuestra muestra viene dado por la suma de los niveles de probabilidad para *n* entre 0 y 3, es decir:

$$P(S \leq 3) = p(0) + p(1) + p(2) + p(3) =$$
$$0.0010 + 0.0098 + 0.0439 + 0.1172 = 0.1719$$

Podemos obtener estos valores directamente de la Tabla 1 (véase la Figura 9.7):

n=10 p=0.5

9	9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0008	0,0020
10	0	0,9044	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025	0,0010
10	1	0,0914	0,3151	0,3874	0,3474	0,2684	0,1877	0,1211	0,0725	0,0403	0,0207	0,0098
10	2	0,0042	0,0746	0,1937	0,2759	0,3020	0,2816	0,2335	0,1757	0,1209	0,0763	0,0439
10	3	0,0001	0,0105	0,0574	0,1298	0,2013	0,2503	0,2668	0,2522	0,2150	0,1665	0,1172
10	4	0,0000	0,0010	0,0112	0,0401	0,0881	0,1460	0,2001	0,2377	0,2508	0,2384	0,2051
10	5	0,0000	0,0001	0,0015	0,0085	0,0264	0,0584	0,1029	0,1536	0,2007	0,2340	0,2461
10	6	0,0000	0,0000	0,0001	0,0012	0,0055	0,0162	0,0368	0,0689	0,1115	0,1596	0,2051
10	7	0,0000	0,0000	0,0000	0,0001	0,0008	0,0031	0,0090	0,0212	0,0425	0,0746	0,1172
10	8	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0014	0,0043	0,0106	0,0229	0,0439
10	9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0016	0,0042	0,0098
10	10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0010
11	0	0,8953	0,5688	0,3138	0,1673	0,0859	0,0422	0,0198	0,0088	0,0036	0,0014	0,0005

Figura 9.7: Extracto de la Tabla 1 del Formulario para n = 10 y p = 0.5. En rojo se representan los valores iguales o más extremos que X = 3 (izquierda) así como sus probabilidades (derecha).

En consecuencia, sabemos que si H_0 es cierta, la probabilidad de obtener un número de signos negativos igual o inferior a 3 vale 0.1719. En principio parece una probabilidad lo suficientemente elevada como para ser compatible con H_0. Pero para tomar una decisión de manera correcta es necesario compararla con el nivel de significación elegido (si no se indica expresamente en el enunciado, se asumirá por defecto que vale 0.05).

Regla de decisión: La decisión a tomar respecto a la H_0 la hacemos comparando este valor de probabilidad (nivel p-crítico que en nuestro caso vale 0.1719) con el nivel de significación, α o error tipo I, que estemos dispuestos a cometer ($\alpha / 2$ en el caso de un contraste bilateral).

Conclusión: La hipótesis planteada fue bilateral. Por consiguiente, si utilizamos un α de 0.05 podemos observar que $\alpha / 2 < P(S \leq 3)$ ya que 0.025 < 0.1719. Esta desigualdad nos informa que el nivel p-crítico es más probable que el valor máximo de probabilidad del error tipo I que estamos dispuestos a cometer. Esto nos conduce a no rechazar H_0. Como ya sabemos, sólo en el caso de que $\alpha/2$ fuese mayor que la probabilidad de obtener un estadístico S igual o más extremo que el calculado (el nivel p-crítico) rechazaríamos H_0.

Interpretación: La comparación entre la probabilidad empírica de obtener un valor igual o más extremo que el obtenido para el estadístico de contraste (nivel p-crítico) con el nivel de significación nos llevan a no rechazar H_0: no existe evidencia para sospechar que la mediana de los TR sea diferente de 450 ms.

Ya se vio en cursos anteriores que la distribución de probabilidad binomial se aproxima a la normal conforme *n* aumenta. Por ello cuando el número de elementos muestrales es lo suficientemente elevado, puede utilizarse el estadístico Z como aproximación a la binomial mediante la fórmula:

$$Z = \frac{S - \bar{S}}{\sigma_S} \qquad \Rightarrow \qquad Z = \frac{S - \frac{n}{2}}{\sqrt{n/4}}$$

No obstante, no hay acuerdo preciso acerca del tamaño muestral que garantiza o recomienda la utilización de Z. Así, Daniel (1990) defiende que *n* debería ser igual o superior a 12. Sin embargo Siegel y Castellan (1988) recomiendan que la aproximación se utilice cuando *n* sea mayor que 25. Para los efectos prácticos de este curso, utilizaremos un tamaño muestral intermedio entre estas dos recomendaciones (tal y como hacen Lubin, Maciá y Rubio, 2000). Así, elegiremos utilizar la aproximación normal de la binomial cuando *n* sea superior a 20. Entre otras razones prácticas para seguir este criterio, podemos citar que las tablas de la Binomial usuales suelen finalizar con este valor. No obstante, en una época de ordenadores baratos y fáciles de utilizar, con un gran número de algoritmos y *software* al alcance del estudiante, las aproximaciones de una distribución a otra puede estar de más.

9.3.2.- La prueba de Wilcoxon para una muestra

Hemos visto que la prueba de los signos solo hace uso de la información presente en la muestra acerca de si cada valor muestral observado, X_i, es superior o inferior a la mediana (m_0). Esto se reflejaba en los signos de las diferencias $X_i - m_0$. En cualquier caso, la prueba de los signos no utiliza la magnitud de esta diferencia entre X_i y m_0 en sus cálculos, es decir, daba igual que, por ejemplo, $X_i - m_0 = -1$ o que fuese $X_i - m_0 = -25$ ya que en ambos casos, se disponía y se contabilizaba un único signo negativo. Por tanto, la magnitud de la diferencia es una información presente en los datos pero que el test de los signos no utiliza. Ante este problema de falta de sensibilidad del test de los signos, Frank Wilcoxon (1945, 1949) desarrolló un procedimiento que utiliza ambas informaciones: la dirección o signo de la diferencia y la magnitud de esta diferencia, permitiendo poner a prueba hipótesis sobre la mediana pero, además, pone a prueba la hipótesis de que la distribución sea simétrica (que el conjunto de datos superior a m_0 es igual al conjunto de datos inferior a m_0).

Los **supuestos** que utiliza la prueba de Wilcoxon para una muestra son los siguientes:

- La muestra se ha seleccionado al azar de la población que representa.
- La variable dependiente medida es de intervalo o de razón. Obsérvese que este supuesto difiere del correspondiente supuesto realizado en el test de los signos.

- La distribución poblacional subyacente es simétrica. En el caso de que este supuesto no se cumpla, se puede utilizar la prueba de los signos binomial (Daniel, 1990), en la que no entraremos.

El test de Wilcoxon se recomienda especialmente como sustituto de la prueba *t* cuando se sospeche que el supuesto de normalidad se incumple de forma obvia.

La hipótesis nula bilateral es similar a la formulada para la prueba de los signos:

$$H_0 : \eta = m_0$$
$$H_1 : \eta \neq m_0$$

Si H_0 es cierta y recordando que la mediana deja por debajo de sí a la mitad de los datos muestrales (ya recordamos esta característica de la mediana en el test de los signos), esto se traduce en que la suma de los órdenes (o rangos) de las puntuaciones existentes por encima de la mediana deberá ser aproximadamente igual a la suma de los órdenes de las puntuaciones por debajo de la mediana, es decir, $\Sigma R_+ \approx \Sigma R_-$. El incumplimiento de esta desigualdad conducirá al rechazo de H_0.

Cálculo de los rangos promediados: Recordemos que el rango R (u orden) de una observación X_i se obtiene disponiendo todas las observaciones en orden desde la más pequeña a la mayor. A cada observación directa se le asocia su orden, es decir, a la observación más pequeña se le asocia el rango u orden 1, a la siguiente observación el 2 y así sucesivamente. A la mayor observación se le proporciona el rango *n* (ya que será la última). Este procedimiento para asignar rangos a las observaciones directas asume que no hay empates, es decir, que no hay dos medidas exactamente iguales. Esto sólo podría suceder si las mediciones originales se realizasen con una precisión infinita, lo cual no es el caso en las situaciones reales. Las observaciones reales tienen una precisión finita y, por consiguiente, se pueden producir empates. Podemos pensar en dos estudiantes, Juan y Pedro, a los que se les ha sometido a un experimento de TR y podrían tener un TR medio de 453.5432… ms. y 453.5382… ms. respectivamente. Pero si el cronómetro que se utilice para su medición sólo puede precisar hasta el milisegundo significa que obtendrá para ambos el mismo valor de 453 ms., aunque en realidad tienen diferente puntuación siempre y cuando pudiéramos medirla con una precisión inferior al milisegundo. Este argumento puede ampliarse a cualquier nivel de precisión que podamos alcanzar con un instrumento de medida real. Es por ello que en la realidad y debido a la precisión finita de nuestros instrumentos de medida tendremos empates. En estos casos los rangos iniciales (1, 2, 3, …, n) se reemplazan por los rangos promediados.

Supongamos que disponemos de las siguientes observaciones de TR: {450, 500, 732, 322, 450}. Si disponemos estas puntuaciones en orden y les asignamos a cada una el número de orden tendríamos (véase la Tabla 9.2):

Tabla 9.2

Puntuación directa	450	500	732	322	450
Rango u orden	2	4	5	1	3
Rango promediado	(2+3)/=2,5	4	5	1	(2+3)/=2,5

En la Tabla 9.2 podemos observar en la primera fila los datos originales de TR. En la segunda fila podemos observar su orden (a 322 se le asigna el orden 1 al ser la puntuación más pequeña de todas, a 450 se le asigna el orden 2 al ser la segunda puntuación, …, a la puntuación 732 se le asigna el rango 5 al ser la puntuación más elevada). Pero el problema estriba en que hay dos puntuaciones iguales a 450. ¿A cuál debería otorgársele el rango 2 y a cuál el 3? Como inicialmente no tenemos ninguna razón para discriminar entre dos puntuaciones que son idénticas, inicialmente se le ha proporcionado a una de ellas (a cualquiera) el rango 2 y a la otra el rango 3 (no hay forma de identificar cual es cual, así que da igual cual tenga el rango 2 y cual el 3). Pero como no tiene sentido proporcionar a una única puntuación directa (450) dos rangos distintos (el 2 y el 3), se les proporciona un único rango (2.5) que es la media entre los rangos otorgados inicialmente (la media entre los rangos 2 y 3 es (2+3)/2 = 2.5). Y si tuviéramos tres puntuaciones empatadas, el procedimiento sería similar. Por ejemplo, asumamos que las puntuaciones originales hubiesen sido {322, 450, 450, 500, 450}, podemos repetir los cálculos anteriores tal y como puede verse en la Tabla 9.3.

Tabla 9.3

Puntuación directa	322	450	450	450	500
Rango u orden	1	2	3	4	5
Rango promediado	1	(2+3+4)/3=3	(2+3+4)/3=3	(2+3+4)/3=3	5

Las puntuaciones ordenadas de menor a mayor serían: 322, 450, 450, 450, 500 (de hecho, por comodidad, ya están ordenadas en la Tabla 9.3) Observamos que en este caso, como tenemos tres puntuaciones empatadas, debemos calcular la media de tres puntuaciones, por lo que dividimos entre 3 la suma de los rangos que asignamos a las puntuaciones empatadas.

El procedimiento que seguimos en el test de Wilcoxon calcula primero la diferencia entre cada puntuación directa X_i y la mediana postulada en H_0. De esta forma obtendremos un conjunto de *n* puntuaciones diferencia a las que denominaremos Y_i para diferenciarlas de las originales ($Y_i = X_i - m_0$). El signo de Y_i dependerá de que X_i sea mayor (signo positivo) o menor (signo negativo) que m_0. Para cada puntuación Y_i registraremos también su orden o rango utilizando los siguientes criterios:

- Ordenamos los valores absolutos de los Y_i, es decir, sin considerar el signo de Y_i. Además, eliminamos cualquier Y_i que valga 0 porque coincide con la m_0 lo que debería ser imposible según el supuesto de distribución continua subyacente.
- Cuando hayan valores de Y_i que en término absoluto (recuérdese que el valor absoluto de una puntuación, Y_i, se simboliza encerrando entre barras verticales esta puntuación, es decir, $|Y_i|$ y que representa a esta puntuación siempre con valor positivo) se encuentren empatados se proporcionará a cada valor empatado el promedio de los órdenes. A diferencia de otros procedimientos que utilizan rangos, la prueba de Wilcoxon exige que se asigne el orden 1 a la puntuación con el valor absoluto más bajo y el rango mayor (n) a la puntuación diferencia con el valor absoluto más alto.
- A continuación se suman los valores absolutos de las puntuaciones diferencia que tienen signos positivos para producir una puntuación que denotaremos por ΣR_+ (sumatorio de los rangos positivos) y los valores absolutos de las puntuaciones diferencia que tienen signos negativos para producir ΣR_- (sumatorio de los rangos negativos, obsérvese el signo del subíndice en ambos casos).

Si H_0 es cierta, es de esperar que los valores ΣR+ y ΣR- sean similares entre sí. Bajo el supuesto anterior, la media ($\overline{R}$) y la varianza (σ_R^2) de los rangos valen respectivamente:

$$\overline{R} = \frac{n(n+1)}{4}$$

$$\sigma_R^2 = \frac{n(n+1)(2n+1)}{24}$$

En la Figura 9.8 podemos ver el conjunto de observaciones directas {3, 5, 7, 9.5, 12, 15, 22} representados por un diagrama de puntos (en azul) sobre el eje X. Si quisiéramos poner a prueba la hipótesis de que la mediana vale 8 en la población de la que se han extraído estos datos (la mediana viene representada por la línea vertical roja situada en la vertical de X = 8) deberíamos calcular los rangos no de las observaciones directas sino de las diferencias entre cada puntuación y la mediana. Estas diferencias vienen representadas por las flechas horizontales grises que parten de la mediana. Cada una de ellas viene etiquetada con su longitud (y sin considerar el signo, es decir, sin considerar el que se encuentre hacia la izquierda de la mediana o hacia la derecha). Esta longitud se corresponde matemáticamente con el valor absoluto. Son estos valores absolutos (la distancia sin considerar el signo) los que se ordenan en el test de Wilcoxon, no las puntuaciones directas. Siempre y cuando H_0 sea cierta, resulta razonable esperar que la suma de los rangos inferiores a m_0 sea aproximadamente igual a la suma de rangos superiores a m_0. Vemos que la distancia más pequeña con respecto a la mediana se corresponde con la puntuación directa 7 (distancia = 1), a la cual le otorgamos el rango 1. La siguiente distancia se corresponde con la puntuación directa 9.5 (distancia 1.5) y por tanto le otorgamos el rango 2. Y así sucesivamente. Las distancias ordenadas de menor a mayor son: {1, 1.5, 3, 4, 5, 7, 14}. Los rangos son, respectivamente, {1, 2, 3, 4, 5, 6, 7}. En la Tabla 3 aparecen las puntuaciones directas, las distancias (en valor absoluto) a la mediana y el rango otorgado a estas distancias.

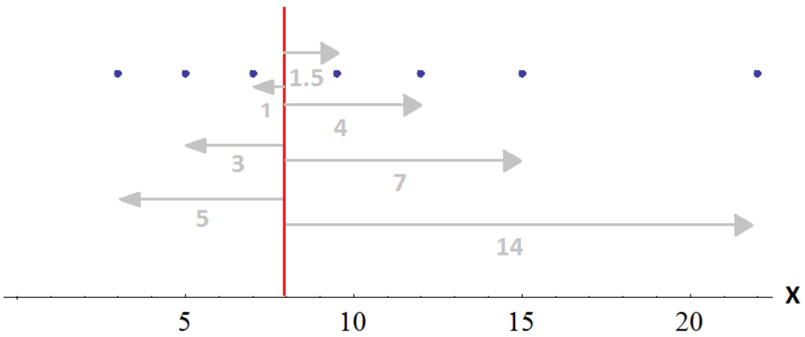


Figura 9.8: Diagrama de puntos de las puntuaciones {3, 5, 7, 9.5, 12, 15, 22} así como la magnitud de la distancia entre cada punto y la mediana (flechas grises).

Tabla 9.3

Puntuaciones directas	7	5	3	9.5	12	15	22
Distancia	1	3	5	1.5	4	7	14
Rango de las distancias	1º	3º	5º	2º	4º	6º	7º

$\Sigma R_{-}=1+3+5=9$ $\Sigma R_{+}=2+4+6+7=19$

Obsérvese que en la Tabla 9.3 hemos reordenado las puntuaciones directas para poner todas aquellas puntuaciones inferiores a la mediana a mano derecha y las puntuaciones superiores a mano derecha. De esta forma nos resultará sencillo realizar el sumatorio de los rangos correspondientes a aquellas puntuaciones inferiores a la mediana y de los rangos correspondientes a las puntuaciones superiores a la mediana. La pregunta a responder es si la diferencia entre ΣR_{+} y ΣR_{-} es lo suficientemente elevada como para sospechar que H_0 no es cierta (recordemos que si H_0 es cierta, es de esperar que estos dos sumatorios sean similares). Si el valor de ΣR_{+} es significativamente mayor que el valor de ΣR_{-} nos señala que existe una alta probabilidad de que la muestra se derive de una población con una mediana mayor que el valor planteado en la H_0 para la población, y a la inversa si $\Sigma R+$ es significativamente menor que el valor de $\Sigma R-$. Cuando estos valores sean equivalentes ambos tendrán un valor de $n(n+1)/4$ excepto por desviaciones debidas al azar.

La Tabla del estadístico W de Wilcoxon se encuentra tabulada en función de los valores inferiores y superiores que pueda adoptar cualquiera de las sumas de rangos. Si, como suele ser habitual, se utiliza como estadístico de contraste o discrepancia entre los rangos observados y los que cabría esperar si la H_0 fuese verdadera, el valor más pequeño de las dos sumas de rangos (ΣR_{+} y ΣR_{-}) obtenidas en nuestra muestra, deberíamos buscar en la Tabla del estadístico W de Wilcoxon el valor numérico situado bajo la columna rotulada W_{izq}. A este estadístico (de Wilcoxon) lo representaremos por W. En nuestro ejemplo, el estadístico W adoptará el valor de 9 ya que es el valor inferior del conjunto {9, 19}. Para poder interpretar este valor es necesario calcular la probabilidad de obtener un valor igual o tan extremo como el obtenido. Si esta probabilidad (nivel p-crítico) es inferior al nivel de significación α , rechazaremos H_0. Existen diferentes formas de realizar este paso. La más directa sería utilizar un software que proporcionara el valor p-crítico exacto. En su defecto, recurrimos a las tablas de valores críticos de Wilcoxon. En el ejemplo que estamos desarrollando, para un $\alpha = 0.05$ obtenemos un valor del estadístico W crítico de 4 (véase la Figura 9.9). Solamente si W hubiese sido inferior a 4 se podría haber rechazado la H_0. De hecho, podemos observar que las Tablas del estadístico W de Wilcoxon están diseñadas para establecer un intervalo de valores de W superiores e inferiores ($W_{izq.}$ y $W_{der.}$) alrededor del cual deben encontrarse los valores de ΣR_{-} y ΣR_{+}. Si utilizamos este criterio vemos que, según las tablas a un alpha de 0.05 (el valor más cercano que aparece para n = 7 es, en realidad, de 0.055), los valores que limitan la región de aceptación de H_0 son 4 y 24. Como los valores de ΣR_{-} y $\Sigma R+$ son, respectivamente, 9 y 19, podemos observar que se encuentran dentro del límite de aceptación.

Tamaño muestral

α

N	$W_{izq.}$	P	$W_{der.}$
7	0	0.008	28
	1	.016	27
	2	.023	26
	3	.039	25
	4	.055	24
	5	.078	23
	6	.109	22
	7	.148	21
	8	.188	20
	9	.234	19
	10	.289	18
	11	.344	17
	12	.406	16
	13	.469	15
	14	.531	14

Figura 9.9: Valores críticos del test de Wilcoxon para n = 7 y un $\alpha = 0.055$. Estos valores aparecen en la misma línea que el α buscado y delimitan una gama de valores dentro de la cual deben encontrarse ΣR_- y ΣR_+.

Para rechazar H_0, el valor obtenido de W en la muestra no debe alcanzar el valor crítico $W_{izq.}$ que refleja la tabla para un nivel de significación especificado tal y como se desarrolla en el siguiente ejemplo.

Ejemplo: Una muestra aleatoria de estudiantes fueron entrevistados para determinar el número real de horas que dedicaban al estudio del examen final de una asignatura concreta. Los resultados fueron

{25, 12, 10, 15, 20, 17, 30, 125, 3}

Si estudios anteriores han mostrado que la mediana del número de horas que se estudia es de 15 horas ¿apoya la evidencia empírica actual esta observación? Trabaje a un $\alpha = 0.05$. El gráfico de estos datos aparece en la Figura 10 con sus correspondientes distancias hasta la mediana planteada. El hecho de que haya un dato extremo (120) impide visualizar correctamente el grueso de los datos por lo que en la parte inferior se han ampliado para poder visualizarse mejor

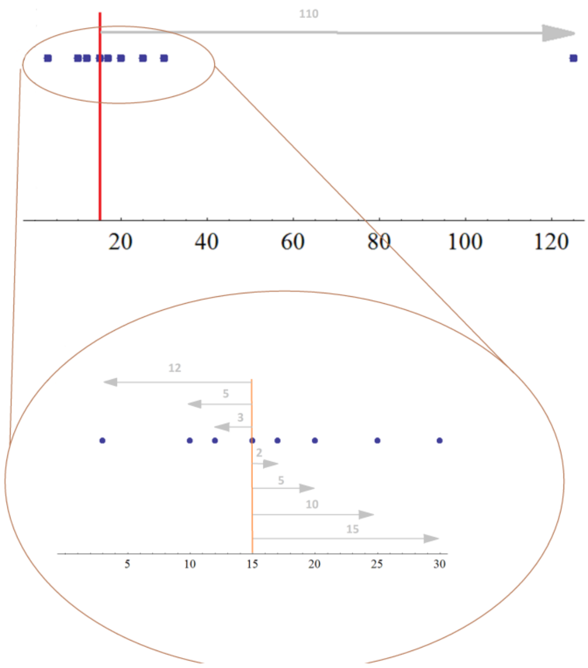


Figura 9.10: representación gráfica de los datos del ejemplo 2.4 con ampliación de la zona central para que puedan identificarse los rangos del grueso de datos.

Condiciones y supuestos: El investigador utiliza una única muestra para realizar un contraste sobre la mediana, estadístico que no representa un parámetro de una distribución poblacional y aún tratándose de una variable medida al menos con escala de intervalo, desconoce la distribución de la variable en la población. Por otra parte, puede verse un aspecto muy interesante en la gráfica de los datos: tenemos un dato muy alejado del resto de datos (un *outlier*). Esta puede ser otra de las razones por las que se ha adoptado por una técnica no paramétrica (si hubiéramos utilizado un contraste de la media, este valor extremo habría desplazado notablemente hacia la derecha la estimación de la media).

Formulación de las hipótesis: La hipótesis de investigación es que la mediana de las horas de estudio es de 15 horas. Se trata, por tanto, de un contraste bilateral

$$H_0: \quad \eta = 15$$
$$H_1: \quad \eta \neq 15$$

Estadístico de contraste: Para contrastar la hipótesis, calculamos la diferencia entre cada puntuación y la mediana, sin tener en cuenta el signo. Posteriormente ordenamos estas diferencias asignándoles un rango de menor a mayor. Observemos que tenemos un dato que coincide con la mediana (m_0=15) por lo que debemos eliminarlo pasando de tener 9 datos a 8 (n = 8). Además, tenemos dos datos que están a la misma distancia de m_0 (10 y 20 se encuentran a 5 unidades en valor absoluto de m_0=15) por lo que tendremos que utilizar el cálculo de los rangos ponderados para este par de datos. Además, tenemos solo 3 datos a la izquierda de m_0 y 5 a la derecha con los rangos alternando a derecha e izquierda.

Intuitivamente no tenemos una base firme para sospechar que H_0 sea cierta o falsa, pero necesitamos confirmación analítica.

Veamos analíticamente los cálculos a realizar ordenando previamente los datos en la Tabla 9.4.

Tabla 9.4

Sujeto	X_i	$\lvert X_i - m_0 \rvert$	Rango	Rango con signo
1	3	$\lvert 3-15 \rvert = 12$	6	-6
2	10	$\lvert 10-15 \rvert = 5$	(3+4)/2=3.5	-3.5
3	12	$\lvert 12-15 \rvert = 3$	2	-2
4	15	$\lvert 15-15 \rvert = 0$	Eliminado	Eliminado
5	17	$\lvert 17-15 \rvert = 2$	1	1
6	20	$\lvert 20-15 \rvert = 5$	(3+4)/2=3.5	3.5
7	25	$\lvert 25-15 \rvert = 10$	5	5
8	30	$\lvert 30-15 \rvert = 15$	7	7
9	125	$\lvert 125-15 \rvert = 110$	8	8

$\Sigma R_- = 11.5$; $\Sigma R_+ = 24.5$

Obsérvese que los rangos han sido asignados a las diferencias en valor absoluto {2, 3, 5, 5, 10, 12, 15, 110}, obteniendo los rangos {1, 2, 3.5, 3.5, 5, 6, 7, 8}. Los rangos de las puntuaciones {3, 10, 12}, al ser inferiores a la mediana postulada, tienen signo negativo y se suman (sin considerar el signo). Esto produce un valor de ΣR_- igual a 2+3.5+6 = 11.5. Por su parte, los rangos de las puntuaciones {17, 20, 25, 30, 125}, al ser superiores a la mediana postulada, tienen signo positivo y se suman (sin considerar el signo). Esto produce un valor de y ΣR_+ igual a 1+3.5+5+7+8 = 24.5.

En consecuencia los estadísticos[1] valen $\Sigma R_- = 11.5$ y $\Sigma R_+ = 24.5$. Utilizaremos como estadístico W el valor de ΣR_- ya que es el más pequeño de ambos. Luego W = 11.5.

Regla de decisión: Ahora debemos acudir a las tablas de Wilcoxon para un contraste bilateral con $\alpha = 0.05$ (al ser bilateral utilizaremos $\alpha / 2$) para extraer los valores críticos (véase Figura 9.11). Para n = 8 en un contraste bilateral el valor crítico es W = 4. Si utilizamos los valores críticos superior e inferior, observamos que valen, respectivamente, 4 y 32.

[1] Obsérvese que el signo negativo de ΣR- se ha utilizado simplemente como señal para sumar estos valores pero con signo positivo, es decir, en vez de sumar (-6)+(-3.5)+(-2) sumamos 6+3.5+2 ya que el signo negativo solo es un indicador de que estos rangos van juntos en un único sumatorio.

Tamaño muestral

α

8	0	.004	36
	1	.008	35
	2	.012	34
	3	.020	33
	4	.027	32
	5	.039	31
	6	.055	30
	7	.074	29
	8	.098	28
	9	.125	27
	10	.156	26
	11	.191	25
	12	.230	24
	13	273	23
	14	.320	22
	15	.371	21
	16	.422	20
	17	.473	19
	18	.527	18

Figura 9.11: Valores críticos superior e inferior del test de Wilcoxon (en azul) para un $\alpha/2 = 0.025$ (se ha utilizado el valor de α más próximo) con n = 8.

Conclusión: Debido a que la H_0 solo puede rechazarse si W ≤ 4, concluimos que a partir de los datos observados en la muestra, no hay evidencia para rechazar la H_0. Llegaríamos a la misma conclusión si utilizamos los valores superior e inferior de la tabla y observamos que los valores empíricos $\Sigma R_- = 11.5$ y $\Sigma R_+ = 24.5$ son interiores al rango extraído de la tabla, que va de 4 a 32. Luego son compatibles con la H_0 a un α bilateral de 0.05.

Interpretación: No hay evidencia suficiente para rechazar que el número real de horas semanales que los estudiantes emplean para repasar el examen final es de 15.

9.4.- Contrastes no paramétricos para diseños de dos grupos independientes

9.4.1.- Test de Mann-Whitney-Wilcoxon

Cuando no podamos asumir los supuestos necesarios para realizar un contraste paramétrico sobre dos medias en muestras independientes, utilizaremos el test de Mann-Whitney-Wilcoxon (MWW), con el que pondremos a prueba la igualdad o desigualdad de las medianas de las poblaciones de las que proceden las muestras, pudiendo plantearse, al igual que en el caso de las medias, contrates unilaterales o bilaterales. También podemos emplear el test MWW para reducir o eliminar el impacto de los valores atípicos (*outliers*) ya que este test utiliza para su cómputo los valores ordinales (rangos) de los datos y no los valores directos. Como los rangos no se ven afectados por la magnitud de las puntuaciones, la presencia de *outliers* no afectará a los resultados que se obtengan de la aplicación del test MWW mientras que sí lo harían si utilizáramos el correspondiente test paramétrico que utiliza estadísticos muestrales (media) fuertemente afectados por los valores atípicos.

No obstante, este test, desarrollado independientemente por Wilcoxon (1945) y Whitney (1947) no está exento de supuestos, siendo necesario para su aplicación que la variable dependiente empleada sea, al menos, de nivel ordinal y que las distribuciones subyacentes a partir de las que se han extraído las muestras tengan la misma forma. Dicha forma no tiene porque ser normal, si bien, el supuesto de igualdad de forma conlleva el supuesto de homocedasticidad. En su lado positivo, el test MWW no se muestra tan afectado por la violación del supuesto de homocedasticidad como lo es el test paramétrico correspondiente.

Veamos, a partir de un ejemplo, cómo aplicar el test de Mann-Whitney-Wilcoxon.

Ejemplo: Un equipo de psicólogos evolutivos ha estudiado como adquieren los niños la capacidad de empatizar con otros. Para este fin han desarrollado un test para medir la empatía en niños pequeños. Los estudios preliminares han mostrado que los chicos son más lentos en desarrollar esta capacidad y, además, la desarrollan en menor cuantía que las chicas. En uno de estos estudios se eligieron dos grupos al azar, uno de niños (Grupo 1) y otro de niñas (Grupo 2) y las puntuaciones obtenidas, en una escala de intervalo, fueron:

Grupo 1: 13; 13; 25; 13; 18; 8

Grupo 2: 27; 21; 15; 21; 23; 30; 16

Determine si estos resultados apoyan la afirmación de que los niños tienen menor empatía que las chicas ($\alpha = 0.05$).

Condiciones y supuestos. Aunque la variable dependiente está medida a un nivel de intervalo, no sabemos cómo es la forma de las distribuciones de niños y niñas en la población. Esto significa que no debemos utilizar un tests paramétrico que asuma una forma concreta para las mismas. Por otro lado, el tamaño de las muestras es pequeño ($n_1 = 6 \ y \ n_2 = 7$), y por lo tanto no podemos asumir que la distribución muestral de las diferencias entre las puntuaciones de ambos grupos sea normal. Todas estas razones nos inducen a utilizar un contraste no paramétrico aplicando el test MWW ya que tenemos dos muestras independientes (la puntuación obtenida por uno de los chicos es independiente de la puntuación obtenida por cualquiera de las chicas).

En general, para aplicar el test MWW, necesitamos que se cumplan las siguientes condiciones:

- Variable dependiente medida en una escala al menos ordinal.
- Distribuciones poblacionales con la misma forma, aunque se desconozca la misma.

Formular las hipótesis. Queremos contrastar si los niños (Grupo 1) tienen menor empatía que las niñas (Grupo 2). Utilizaremos la mediana como índice de tendencia central ya que no podemos utilizar la media al ser ésta un estadístico impropio para variables ordinales. En términos formales, representando la mediana poblacional mediante η (eta minúscula), lo que nos indica la hipótesis planteada por el experimentador es que η_1 (la mediana para la población de niños) es menor en la variable empatía que la mediana para la población de niñas (η_2). Esto significa que el investigador plantea que $\eta_1 < \eta_2$ para la variable empatía. Si en esta desigualdad restamos un mismo valor a ambos lados del signo, la desigualdad se mantiene. Al hacerlo con η_2 obtenemos $\eta_1 - \eta_2 < \eta_2 - \eta_2$. Como el segundo término de la desigualdad es 0, tenemos que $\eta_1 - \eta_2 < 0$ y, por lo tanto, se plantea que la diferencia de medianas entre el primer grupo (niños) y el segundo (niñas) es negativa. Por eso se dice que en este caso se trata de un contraste unilateral izquierdo (los valores compatibles con la hipótesis alternativa, H_1, están a la izquierda en la representación usual de la recta real $\mathbb{R}$).

Obsérvese que la hipótesis que ha llevado al investigador a plantear este experimento es precisamente $\eta_1 < \eta_2$ (o equivalentemente $\eta_1 - \eta_2 < 0$) y coincide con H_1. Si el investigador hubiera creído que la empatía no difiere entre niños y niñas (la H_0) no tendría razón alguna para embarcarse en este trabajo experimental. Como es usual, la hipótesis nula se plantea como la contraria a la hipótesis alternativa y es la que realmente se pone a prueba. De esta forma, si H_0 se rechaza, se hará con un elevado grado de confianza en la decisión de rechazo.

La discusión anterior nos lleva a considerar que el contraste es no paramétrico (las hipótesis planteadas hacen referencia a las medianas poblacionales) y H_1 plantea que esta mediana es menor para niños (grupo 1) que para niñas (grupo 2). Las hipótesis quedan entonces de la siguiente forma:

Formas (equivalentes) de H_0 y H_1

$$H_0 : \eta_1 \geq \eta_2 \qquad\qquad H_0 : \eta_1 - \eta_2 \geq 0$$
$$H_1 : \eta_1 < \eta_2 \qquad\qquad H_1 : \eta_1 - \eta_2 < 0$$

Estadístico de contrate. Para calcular el estadístico del test MWW debemos combinar las dos muestras y transformar las observaciones directas en rangos. A continuación se calcula la suma de los rangos pertenecientes a cada muestra por separado. La lógica del estadístico MWW consiste en que si la mediana de la población de la que se ha extraído la primera muestra (grupo 1) es inferior a la mediana de la segunda población (de la que se ha extraído la segunda muestra o grupo 2) deberíamos esperar que la suma de los órdenes pertenecientes a la primera muestra fuese inferior a la suma de los órdenes pertenecientes a la segunda muestra ya que las puntuaciones del grupo 1 obtendrían, por término medio, rangos inferiores a las puntuaciones del grupo 2 (un razonamiento similar se realizó en el estadístico de Wilcoxon pero en aquel caso en una única muestra). Comenzamos, por lo tanto, asignado rangos a todas las puntuaciones (véase la Tabla 9.5).

Tabla 9.5

Asignación de rangos a las puntuaciones de empatía

	Grupo 1						Grupo 2						
Sujeto	1	2	3	4	5	6	7	8	9	10	11	12	13
Empatía	13	13	25	13	18	8	27	21	15	21	23	30	16
Rango	3	3	11	3	7	1	12	8’5	5	8’5	10	13	6

Recordemos cómo ordenamos las puntuaciones cuando hay empates. En el Grupo 1 hay tres puntuaciones iguales (sujetos 1, 2 y 4 con puntuaciones todas iguales a 13), a las que les corresponderían los rangos 2, 3, 4, y a las que asignamos el rango medio de estos tres valores:

$$\frac{2+3+4}{3} = 3$$

De igual forma, las puntuaciones de los sujetos 8 y 10 son iguales a 21, y ocupan los rangos octavo y noveno, por lo que les asignamos el rango medio 8.5 (es decir, (8+9)/2).

Este primer paso ha asignado los rangos de las puntuaciones globalmente y, por tanto, sin considerar el grupo al que pertenecían (chicos o chicas). A continuación separamos los rangos pertenecientes al grupo de chicos y al grupo de chicas y calculamos la suma de los rangos obtenidos por cada grupo, es decir, la suma de los rangos para las puntuaciones del grupo de chicos, a la que llamaremos S_1, y la suma de los rangos para las puntuaciones del grupo de chicas, a la que llamaremos S_2:

Grupo 1: $S_1 = 3 + 3 + 11 + 3 + 7 + 1 = 28$

Grupo 2: $S_2 = 12 + 8'5 + 5 + 8'5 + 10 + 13 + 6 = 63$

Por último obtenemos los estadísticos U_1 y U_2 que nos van a servir para buscar en la tabla del test MWW las probabilidades buscadas. Pero antes, recordemos qué pasaría en el caso hipotético de que todas las puntuaciones del primer grupo (con n_1 puntuaciones) hubiesen obtenido los rangos inferiores, es decir, si las puntuaciones del grupo 1 fuesen todas inferiores a las puntuaciones del grupo 2. En este caso, la suma de rangos que se hubiese obtenido para S_1 se podría calcular con la fórmula:

$$\frac{1}{2}n_1(n_1+1)$$

Como el grupo S_1 tiene 6 puntuaciones, la suma de rangos en este caso hipotético alcanzaría un valor de $\frac{1}{2}6(6+1)=21$. Por su parte, la suma de rangos que se hubiese obtenido para S2 en el caso hipotético de que todas las puntuaciones del segundo grupo (con n2 puntuaciones) hubiesen obtenido los rangos inferiores, alcanzaría un valor de $\frac{1}{2}n_2(n_2+1)=\frac{1}{2}7(7)=28$. El simple hecho de que los grupos puedan no tener el mismo número de puntuaciones lleva a que los rangos puedan diferir debido a ello y no debido a las diferencias en los rangos de los dos grupos. Para contrarrestar esta posibilidad lo que se hace es restar de los valores S_1 y S_2 obtenidos estos valores mínimos de sumatorios de rangos en donde sólo se considera el número de puntuaciones en cada grupo. Este razonamiento nos conduce a las Ecuaciones 2 y 3, en las que directamente obtenemos los estadísticos buscados U_1 y U_2:

$$U_1 = S_1 - \frac{1}{2}n_1(n_1+1) = 28 - \frac{1}{2}6\cdot 7 = 28-21=7$$

$$U_2 = S_2 - \frac{1}{2}n_2(n_2+1) = 63 - \frac{1}{2}7\cdot 8 = 63-28=35$$

Finalmente elegimos como estadístico U el valor inferior de U_1 o U_2, en nuestro caso el mínimo de 35 y 7 es 7. Por consiguiente, asignamos a U el valor 7. Este valor nos permitirá buscar en las tablas la probabilidad correspondiente a este estadístico.

Establecer la regla de decisión en función del nivel de confianza. La Tabla U de Mann-Whitney-Wilcoxon del apéndice (véase Figura 9.12) expresa los valores críticos U en función del número de sujetos de cada grupo, del nivel de confianza y del tipo de contraste (bilateral o unilateral). Acudiendo a dicha tabla con $\alpha = 0.05,\ n_1 = 6\ y\ n_1 = 6$ en un contraste unilateral, obtenemos un valor crítico igual a 8.

Tabla de valores críticos para el estadístico U de Mann-Whitney-Wilcoxon
(valores unilaterales 0.05)

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1																		
2					0	0	0	1	1	1	1	2	2	2	3	3	3	4
3			0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9
4			0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16
5		0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22
6		0	2	3	5	[illegible]	8	10	12	14	16	17	19	21	23	25	26	28
7		0	2	4	[illegible]	8	11	13	15	17	19	21	24	26	28	30	33	35
8		1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41
9		1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48
10		1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55
11		1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61
12		2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68

Figura 9.12

Conclusión. La significación, es decir, el rechazo de H_0, se alcanza si el estadístico de contraste (valor U obtenido del procedimiento de cálculo previo) es igual o menor que el valor crítico extraído de la tabla al nivel de significación especificado. Como 7 < 8, rechazamos la hipótesis nula al nivel de confianza del 95%.

Interpretar el resultado en el contexto de la investigación. Al nivel de confianza del 95%, los resultados apoyan la afirmación de que los niños tienen menor empatía que las niñas de acuerdo con los estudios preliminares llevados a cabo por el equipo de psicólogos evolutivos.

Muestras grandes (aproximación a la distribución normal). Cuando n_1 ó n_2 son superiores a 20 podemos utilizar el Teorema del Límite Central para demostrar que el estadístico que puede verse en la Ecuación 3, se distribuye según la curva normal tipificada, es decir, según una *z*:

$$Z = \frac{U_i - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}}} \qquad \text{Ecuación 3}$$

y en donde el término $n_1 \cdot n_2 / 2$ representa el valor medio esperado de U si H_0 es cierta, es decir, si realmente los dos grupos tienen idéntica mediana. El denominador representa el error típico de la distribución muestral del estadístico U. Aunque con los datos del ejemplo anterior no tenemos suficientes sujetos para utilizar de forma razonable la aproximación a la distribución normal, lo haremos para ilustrar su cálculo. Observamos que en la Ecuación 3 se ha puesto U_i, ya que podemos utilizar U_1 o U_2 de forma indistinta para realizar el contraste. Efectivamente:

$$Z_1 = \frac{U_1 - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}}} = \frac{7 - 21}{\sqrt{\frac{42 \cdot 14}{12}}} = -2$$

$$Z_2 = \frac{U_2 - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}}} = \frac{35 - 21}{\sqrt{\frac{42 \cdot 14}{12}}} = 2$$

Si buscamos en las tablas de la curva normal para $\alpha = 0.05$ con un contraste unilateral izquierdo (véase en la Figura 9.13) encontramos una Z crítica de -1.64 (por la izquierda).

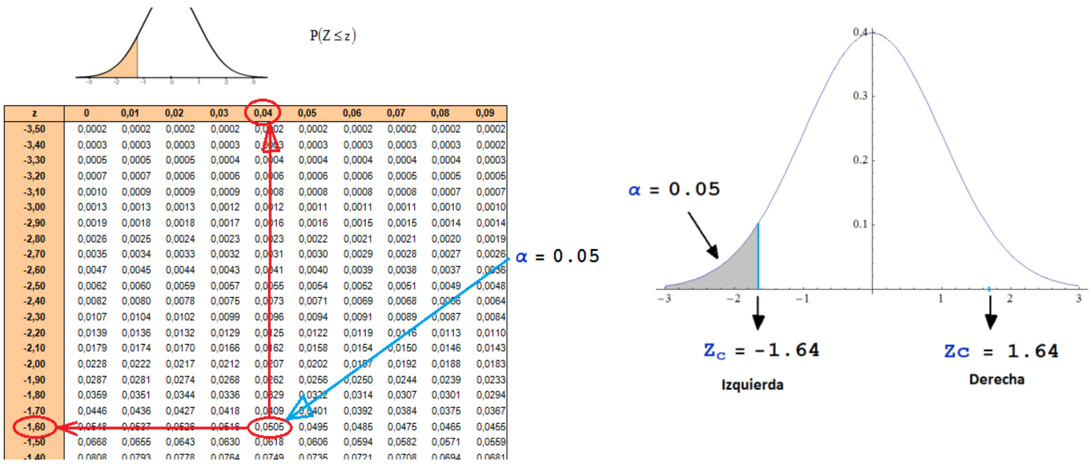


Figura 13: A mano derecha se muestra la curva normal tipificada (media 0 y desviación típica 1) con el valor de Z crítico (Z_c= -1.64) que deja por debajo de sí el 0.05 del área de la distribución. A mano izquierda se muestra la búsqueda de Z a partir del valor de α .

El resultado de Z_1 nos informa que el valor U_1 (Grupo 1) es inferior al valor medio esperado si H_0 es cierta (es decir, si Z_1 = -2 y Z crítica por la izquierda -1.64, resulta -2 < -1.64), lo cual nos llevaría a rechazar H_0. Por su parte, Z_2 = 2 indica que U_2 es mayor que el valor medio esperado según H_0 (2 > 1.64). En ambos casos, el valor de U_i queda en la región de rechazo de H_0.

Dos casos especiales de la aproximación de la normal se producen si el valor Z obtenido está cerca de los valores críticos tabulados. Estos dos casos son la corrección por continuidad y la corrección por empates. En ambos casos la Ecuación 3 debe modificarse para resolver estas cuestiones. El lector interesado puede consultar dichos casos en la bibliografía recomendada.

9.5.- Contrastes no paramétricos para diseños de dos grupos dependientes

9.5.1 Test de Wilcoxon

Cuando se incumplan uno o más de los supuestos de los tests paramétricos que se han visto en capítulos anteriores, podremos utilizar como procedimiento alternativo el test de Wilcoxon para dos muestras relacionadas debido a que sus supuestos son menos estrictos. Tanto el test de los signos como el test de Wilcoxon se han aplicado al caso de una única muestra. Los casos más interesantes surgen, sin embargo, cuando tenemos dos o más muestras ya que entonces podemos realizar comparaciones entre grupos o condiciones experimentales distintas.

Ejemplo: Un investigador educativo está interesado en el efecto que tiene el ruido ambiental sobre la comprensión lectora. Dispone de seis estudiantes a los que somete a dos condiciones experimentales, una con ruido y otra sin ruido. Las condiciones se contrabalancean aleatoriamente entre sujetos (es decir, la mitad de las sujetos pasan primero la condición sin ruido y después la condición con ruido mientras que la otra mitad de estudiantes pasan primero la condición con ruido y después la condición sin ruido) de tal forma que el efecto que pueda tener el orden de las condiciones sobre los resultados afecte por igual a ambas condiciones. Las puntuaciones obtenidas fueron las siguientes:

Sujeto	1	2	3	4	5	6
Grupo 1. Sin Ruido	76	89	65	95	98	35
Grupo 2. Con Ruido	60	49	75	91	22	50

Con un nivel de confianza del 95% el investigador desea responder a la pregunta de si el ruido ambiental disminuye las puntuaciones en comprensión lectora.

Condiciones y supuestos. Puesto que todos sujetos pasan por las dos condiciones experimentales, tenemos dos muestras relacionadas ya que las puntuaciones pertenecientes al mismo sujeto obtenidas en las dos condiciones experimentales están relacionadas entre sí. El tamaño de la muestra es muy pequeño (n = 6) y no sabemos cómo es la distribución de las diferencias en la población, por lo que utilizamos una prueba no paramétrica. El test apropiado para este caso es el test de Wilcoxon. Si bien los supuestos no son tan restrictivos como en las pruebas paramétricas, para aplicar Wilcoxon tenemos que asumir que la distribución de las diferencias de los rangos es simétrica. En general, los supuestos necesarios para aplicar apropiadamente este test son:

- La variable dependiente está medida al menos a un nivel ordinal.
- La distribución de las diferencias de los rangos es simétrica.

Formular las hipótesis. Los investigadores en Psicología Ambiental saben que los factores ambientales (v.g., ruido, humedad, iluminación, etc.) afectan a la conducta. En concreto, los estudios del nivel de ejecución en escuelas que se encuentran en ambientes ruidosos vs. no ruidosos han mostrado que los alumnos de escuelas que se encuentran en ambientes ruidosos (v.g., urbanos) obtienen resultados inferiores. Es por ello que el investigador plantea este trabajo bajo la hipótesis teórica de que el grupo con ruido actuará peor en una prueba específica de comprensión lectora. Asumiendo, como es usual, que una puntuación superior en el test significa una mejor comprensión lectora, esto significa que la hipótesis alternativa o H_1 afirma que la mediana del grupo de estudiantes sometidos al ruido (grupo 2) es inferior a la mediana del grupo de estudiantes no sometidos al ruido (grupo 1), es decir, $\eta_1 > \eta_2$. Por consiguiente, podemos establecer esta hipótesis alternativa también como $\eta_1 - \eta_2 > \eta_2 - \eta_2 = 0$. Esto significa que H_1 plantea que la diferencia entre la mediana del primer grupo menos la del segundo será superior a 0. En consecuencia, se está planteando un contraste unilateral derecho ya que se plantea que la diferencia de medianas será un valor positivo (superior a 0) en H_1. La hipótesis nula o H_0 abarcará el conjunto de valores complementario, en el conjunto de los números reales ($\mathbb{R}$), al anterior (véase Figura 9.14).

$$H_0 : \eta_1 \leq \eta_2$$
$$H_1 : \eta_1 > \eta_2$$

ó bien

$$H_0 : \eta_1 - \eta_2 \leq 0$$
$$H_1 : \eta_1 - \eta_2 > 0$$

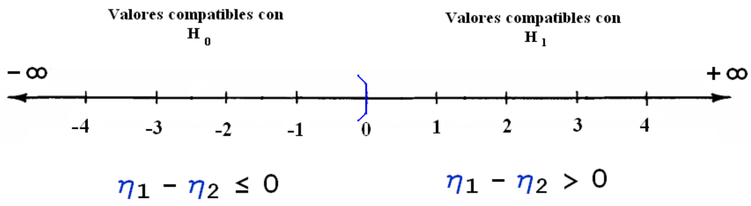


Figura 9.14: A mano izquierda aparece la expresión simbólica de las hipótesis nula y alternativa para el Ejemplo. A mano derecha aparece la recta real con la partición en dos subconjuntos: aquel compatible con H_0 (valores negativos o cero de la diferencia entre las medianas poblacionales del grupo 1 menos la del grupo 2) y aquel compatible con H_1 (valores positivos de la diferencia entre las medianas poblacionales del grupo 1 menos la del grupo 2).

Obsérvese que la asignación del grupo sin ruido como el grupo 1 y el grupo con ruido como el grupo 2 es lo que ha convertido el contraste en unilateral derecho. Al ser esta asignación arbitraria (no hay nada que impida cambiar la asignación de grupo 1 vs. grupo 2), también puede considerarse arbitrario el que el contraste sea unilateral derecho.

Estadístico de contraste. Llamaremos W al estadístico proporcionado por la prueba de Wilcoxon y, al igual que los contrastes paramétricos para dos medias en muestras relacionadas, utiliza las diferencias entre cada par de puntuaciones observadas, calculando posteriormente los rangos de los valores absolutos de dichas diferencias (utilizando los rangos promediados en el caso de que hubiesen empates y eliminando cualquier diferencia nula disminuyendo el valor del tamaño de la muestra). En la Tabla 9.6 se muestran los cálculos necesarios para obtener el estadístico de contraste.

Tabla 9.6 *Aplicación del test de Wilcoxon al ejemplo*					
Sujeto	Sin Ruido X_1	Con Ruido X_2	Diferencia $D = X_1 - X_2$	Rango de las diferencias $\lvert D \rvert$	Rango con signo
1	76	60	16	4	4
2	89	49	40	5	5
3	65	75	-10	2	-2
4	95	91	4	1	1
5	98	22	76	6	6
6	35	50	-15	3	-3
					$\sum R_- = 5$ $\sum R_+ = 16$

La primera columna de la Tabla 9.6 muestra los sujetos. La segunda columna muestra las puntuaciones directas obtenidas por cada sujeto en la primera condición (sin ruido, muestra de puntuaciones 1). La tercera columna muestra las puntuaciones directas obtenidas por cada sujeto en la segunda condición (con ruido, muestra de puntuaciones 2). En la cuarta columna calculamos la diferencia entre cada par de puntuaciones para cada sujeto. El siguiente paso sería ordenar las puntuaciones de la cuarta columna asumiendo que todas tienen signo positivo, es decir, calculando el valor absoluto ($\lvert D \rvert$) de las puntuaciones obtenidas en la cuarta columna. El valor absoluto de las puntuaciones diferencia nos daría el conjunto de puntuaciones {16, 40, 10, 4, 76, 15} que ordenadas darían {4, 10, 15, 16, 40, 76}. Es decir, la puntuación 4 tendría el rango 1, la puntuación 10 el rango 2, …, la puntuación 76 tendría el rango 6. Como vemos, hemos obtenido los rangos sin considerar el signo (quinta columna). En la sexta columna volvemos a retomar el signo obtenido para las diferencias D y asignamos un signo negativo a los rangos que proceden de diferencias negativos (las diferencias positivas ya tienen un signo positivo en sus rangos, por lo que las dejamos tal y como están).

Por último se calcula el valor absoluto de la suma de los rangos con signo positivo y el valor absoluto de la suma de rangos con signo negativo. Si H_0 es verdadera, esperaríamos encontrar que la suma de rangos que procede de puntuaciones positivas sería parecida a la suma de los rangos que procede de las puntuaciones negativas. Sin embargo, si observamos un alejamiento suficientemente elevado de esta expectativa en los datos, se dudará de la veracidad de H_0. Para evaluar si este alejamiento es estadísticamente significativo, tendremos que buscar en las Tablas del test de Wilcoxon los valores de W compatibles con H_0 a un cierto nivel de probabilidad. Para ello y de la misma forma que hicimos en el test de Wilcoxon para una muestra, el estadístico W para dos muestras relacionadas se calcula como el valor más pequeño de los dos sumatorios ($\sum R_- \;\; o \;\; \sum R_+$). En nuestro caso: W = 5 (el mínimo de 5 y 16).

Establecer la regla de decisión en función del nivel de confianza. Trabajando con un nivel de confianza del 95% y para contraste unilateral, acudimos a la tabla de Wilcoxon con α = 0’05 y un tamaño muestral de n = 6, en la que obtenemos un valor crítico igual a: $w_{0'05,\,6} = 2$ (véase la Figura 9.15).

Obsérvese que aunque hemos utilizado un contraste unilateral derecho, hemos elegido en la Tabla de Wilcoxon el valor de w_{izq}. La razón es que en el paso anterior hemos elegido como W el valor más pequeño de los sumatorios de rangos y por tanto, tiene sentido elegir en la tabla el valor w_{izq} ya que es el más pequeño. No obstante, si atendemos al carácter unilateral derecho del contraste, podríamos pensar en utilizar como valor crítico w_{der}, que en este caso sería 19. En el paso siguiente veremos que cualquiera de las opciones es válida pero la primera suele resultar más cómoda por lo que resulta preferible.

Tamaño muestral

N	W_{izq}	P	W_{der}
⋮	⋮	⋮	⋮
6	0	.016	21
	1	.031	20
	2	.047	19
	3	.078	18
	4	.109	17
	5	.156	16
	6	.219	15
	7	.281	14
	8	.344	13
	9	.422	12
	10	.500	11

$\alpha = 0.047 \approx 0.05$

Figura 9.15: Valor del estadístico de contraste W de Wilcoxon para n = 6 y un α unilateral de 0.05 (se ha elegido 0.047 como valor de p ya que es el más cercano al valor de α nominal). Se elige el valor de w_{izq}. por comodidad ya que se eligió como estadístico de contraste el menor de los sumatorios de rangos.

Conclusión. Dado que el estadístico de contrate supera al valor crítico, es decir: 5 > 2 (en general, $W > w_{0'05,\,6}$), no podemos rechazar la hipótesis nula al nivel de confianza del 95%. Si hubiéramos elegido w_{der} en la tabla, podemos observar que llegamos a la misma conclusión ya que $w_{der} = 19 > 16 = \sum R_+$. Es decir, según la Tabla, los valores críticos que delimitan la región de aceptación (o no rechazo) de H_0 son 2 y 19. Como los dos sumatorios de rangos (5 y 16) caen dentro de este intervalo, no se rechaza H_0.

Interpretar el resultado en función del contexto de la investigación. Con los datos de los que disponemos, no podemos afirmar que el ruido ambiental disminuya las puntuaciones en comprensión lectora. Ante este resultado negativo (se dice que un resultado es negativo cuando no podemos rechazar H_0) y dado que existe en la literatura del área información de que el ruido afecta a la ejecución, el investigador se tendrá que plantear si su resultado es fiable (repetible) o no. Dado que la muestra utilizada es muy pequeña, el investigador haría bien en optar por incrementar el tamaño muestral e intentar eliminar los factores por los que no pudo efectuar un análisis paramétrico que suelen resultar más potentes. Otra opción, complementaria a la anterior, sería calcular un índice del tamaño del efecto conseguido con sus 6 sujetos y calcular el tamaño muestral que debería tener para conseguir que esa diferencia fuera significativa. Si el tamaño muestral calculado no fuese excesivo, el investigador debería plantearse incrementarlo adecuadamente.

Muestras grandes. En el caso de que tengamos más de 30 sujetos, no podríamos utilizar la Tabla del estadístico de Wilcoxon para obtener los valores críticos. En este caso y de forma similar a como hicimos en el test de Wilcoxon para una muestra, la Ley de los Grandes Números nos permite utilizar una aproximación para tamaños muestrales superiores a 20 utilizando la distribución normal tipificada z tal y como puede verse en la Ecuación 4.

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \qquad \text{Ecuación 4}$$

9.6.- Contrastes no paramétricos para diseños con más de dos grupos independientes.

9.6.1.- Test de Kruskal-Wallis

Este test es una ampliación del test de Mann-Whitney para más de dos muestras independientes. Para seguir con la notación del correspondiente test paramétrico, denotaremos el número de tratamientos a comparar mediante *a*, siendo *a* necesariamente mayor que 2 ya que en caso contrario deberíamos aplicar el test MWW. Por tanto, es un test que pone a prueba las medianas muestrales con datos medidos a nivel ordinal (o transformados a nivel ordinal a partir de formatos superiores, es decir, a partir de escalas de intervalo o de razón). Los supuestos necesarios para poder aplicar adecuadamente este test son los siguientes:

- Las muestras deben haberse seleccionado al azar de la población que representan.
- Las *a* muestras son independientes entre sí.
- La variable dependiente es una variable continua aunque posteriormente se transforma en una escala ordinal. No obstante, en la práctica este test también se emplea con variables aleatorias propiamente discretas.
- La función de densidad de probabilidad poblacional adoptada por cada distribución puede ser de cualquier tipo (normal, t de Student, χ^2, etc.)
- Debe cumplirse el supuesto de homogeneidad de varianzas.

Un aspecto positivo del test de Kruskal-Wallis es que, al trabajar a partir de datos ordinales, disminuye o se elimina el impacto que pueden tener los valores extremos (*outliers*). En su aspecto negativo, el test de Kruskal-Wallis es menos potente que el correspondiente ANOVA.

Explicaremos el procedimiento de análisis mediante un ejemplo con datos ficticios.

Ejemplo: Un psicólogo está interesado en estudiar el efecto que la privación del sueño tiene sobre la ansiedad. Un grupo de 21 sujetos se asigna al azar a cada uno de los tres tratamientos: (a_1) sin interrupción del sueño; (a_2) a los participantes se les despierta dos veces durante la noche mientras están en sueño profundo medido según el electroencefalograma (EEG) y (a_3) a los participantes se les despierta cuatro veces durante la noche mientras están en sueño profundo medido también según el EEG. El procedimiento se repitió durante cinco noches y cada día se les pasaba un test para determinar el nivel de ansiedad. Puntuaciones altas en el test indican alta ansiedad. Antes de terminar el experimento, se produjo mortalidad experimental ya que abandonaron un sujeto del grupo a_1 y dos del a_3, razón por la cual los tamaños muestrales no son iguales.

Las puntuaciones promedio en el test de ansiedad para cada sujeto en cada grupo se muestran en la Tabla 9.7.

Tabla 9.7

Datos de ejemplo para el Test de Kruskal-Wallis

Número de interrupciones del sueño

Ninguna a_1	**Dos** a_2	**Cuatro** a_3
7	10	15
7	9	11
3	11	12
6	10	9
5	7	10
8	6	
	8	

El investigador desea poner a prueba la hipótesis de que cuanto más tiempo se prive de sueño a los sujetos más alto puntuarán en la escala de ansiedad. Como hicimos en el caso de dos muestras independientes, consideramos todas las puntuaciones en conjunto y reemplazamos cada observación directa por sus rangos u órdenes. Así, reemplazamos la más pequeña de todas las puntuaciones por 1, la siguiente más pequeña por 2, y así sucesivamente. En la Tabla 9.8 aparecen las puntuaciones directas ordenadas (como hemos dicho, sin considerar el grupo) así como su rango y su rango promediado:

Tabla 9.8

Puntuación directa (ordenada)	3	5	6	6	7	7	7	8	8	9	9	10	10	10	11	11	12	15

Rango	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Rango promediado	1	2	3.5	3.5	6	6	6	8.5	8.5	10.5	10.5	13	13	13	15.5	15.5	17	18

En caso de que existan empates entre puntuaciones directas (como sucede en el ejemplo) se promedian los rangos implicados. De esta forma, el valor 6 aparece dos veces como puntuación directa, y la media de los rangos (no promediados) que ocupan son el 3 y el 4. Por consiguiente, su rango promediado es $(3+4)/2=3.5$. De la misma forma, el valor directo 7 está repetido 3 veces (existen tres valores igual a 7 en el conjunto de datos directos) y ocupan las posiciones o rangos 5, 6 y 7. Por consiguiente su rango promediado vale $(5+6+7)/3=6$. El resto de casos de datos directos repetidos está tratado de forma idéntica. A continuación sustituimos las puntuaciones directas de la Tabla 7 por sus rangos promediados y calculamos las sumas y las medias de éstos para cada tratamiento tal y como puede verse en la Tabla 9.

Tabla 9.9

a_1	a_2	a_3
6	13	18
6	10,5	15,5
1	15,5	17
3,5	13	10,5
2	6	13
8,5	3,5	
	8,5	
$R_{a_1}=27$	$R_{a_2}=70$	$R_{a_3}=74$
$\overline{R}_{a_1}=27/6=4.5$	$\overline{R}_{a_2}=70/7=10$	$\overline{R}_{a_3}=74/5=14.8$

Observamos que el grupo a_1 ha obtenido una media de rangos inferior al grupo a_2 (4.5 vs. 10, respectivamente) y éste, a su vez, inferior a la del grupo a_3 (10 vs. 14.8, respectivamente). Esto significa que trabajando con los rangos en vez de con las puntuaciones directas, las medias de los rangos parecen ir en la dirección propuesta por el investigador.

La suma de todos los rangos en un conjunto de 18 datos sabemos que vale (recuérdese que la suma de los *n* primeros números naturales se calcula directamente mediante la relación

$n(n+1)/2$):

$$1+2+3+...+18=\frac{18(18+1)}{2}=171$$

Mientras que el investigador propuso y realizó el experimento pensando que la ansiedad aumentaba conforme el número de horas de privación de sueño aumentaba, la H_0 debe plantearse como la inexistencia de esta relación entre ansiedad y horas de privación de sueño. Si H_0 fuera cierta, deberíamos asumir que los rangos individuales se han distribuido aleatoriamente entre las tres condiciones experimentales (a_1, a_2 y a_3) y, por tanto, en promedio la suma de rangos en cada condición, debería ser proporcional al número de observaciones por grupo. Siendo 171 el valor de la suma de rangos para 18 puntuaciones y 6/18 la proporción de observaciones en el primer grupo, 7/18 la proporción de observaciones en el segundo grupo y 5/18 la proporción de observaciones en el tercer grupo (obsérvese que $\frac{6}{18}+\frac{7}{18}+\frac{5}{18}=\frac{18}{18}=1$), el producto entre el total de la suma de rangos por su proporción esperada en cada grupo nos proporciona la suma de rangos esperada en cada grupo si H_0 fuese cierta:

$$[R_{a_1}=171\frac{6}{18}=57;R_{a_2}=171\frac{7}{18}=66{,}5;R_{a_3}=171\frac{5}{18}=47{,}5].$$

Es decir, si H_0 es cierta, esperaríamos en promedio una suma de rangos de 57 puntos en el grupo a_1 aunque empíricamente hemos obtenido un valor de 27 (lo cual parece mostrar una diferencia apreciable entre lo que se esperaría si H0 fuese cierta y lo que realmente hemos encontrado); en el grupo a_2 esperaríamos una suma de rangos de 66.5 aunque empíricamente hemos obtenido un valor de 70 (la diferencia entre 66.5 y 70 parece bastante pequeña); por último, en el grupo a_3 esperaríamos una suma de rangos de 47.5 aunque empíricamente hemos obtenido un valor de 74 (lo cual muestra en este grupo una diferencia que también parece apreciable). La pregunta que debemos plantear es ¿son estas diferencias entre lo que esperaríamos si H_0 es cierta y lo que hemos encontrado empíricamente lo suficientemente elevadas para sospechar que H_0 es falsa?

Un test básico que se podría plantear es como el de la prueba de Chi-cuadrado. Como ya sabemos, en este test se comparan los rangos esperados, en el caso de que no hubiera diferencias entre las medias de las poblaciones, y los rangos observados en la muestra. No obstante, el estadístico propuesto por Kruskal-Wallis es más elaborado que la prueba Chi-Cuadrado y permite posteriormente comparaciones múltiples en el caso de la prueba ómnibus (la prueba ómnibus o prueba global es aquella que busca diferencias entre los grupos sin diferenciar entre ellos; en nuestro caso es el propio test de Kruskal-Wallis ya que llevará al rechazo de H_0 si existe alguna diferencia entre cualquiera de los grupos pero sin indicarnos entre qué grupos ha encontrado esta diferencia) lleve al rechazo de H_0. Recuérdese que la prueba ómnibus o prueba global, es el test que busca alguna diferencia, sin importar entre qué grupos se produce. Así, el test de Kruskal-Wallis es un test ómnibus ya que indicará el rechazo de H_0 si existe diferencia entre a_1 vs. a_2, entre a_1 vs. a_3 , a_2 vs. a_2 o alguna combinación de ella, pero no nos señalará entre qué grupos se encuentra esta diferencia. Solamente nos indicará que existe "alguna" diferencia. Es tarea nuestra investigar posteriormente dónde se encuentran estas diferencias mediante las comparaciones múltiples o *a posteriori*. La formulación del test es:

$$H = \frac{12}{N(N+1)}\left(\sum \frac{R_a^2}{n_a}\right) - 3(N+1)$$

siendo N el total de datos, R_a las sumas de rangos obtenidas empíricamente y n_a el tamaño muestral en cada grupo. El estadístico *H* se distribuye según χ^2 con *a-1* grados de libertad. El resultado de este estadístico para los datos que sirven de ejemplo es:

$$H = \frac{12}{18\,(18+1)}\left(\frac{27^2}{6} + \frac{70^2}{7} + \frac{74^2}{5}\right) - 3(18+1) = 10.2526$$

Al ser el valor crítico de las distribución χ^2, con 3-1 = 2 grados de libertad y un α = 0.05, igual a 5.99, se debe rechazar la hipótesis nula de que la privación del sueño no influye en el estado de ansiedad ya que nuestro estadístico se encuentra en la región de rechazo de H_0 (véase la Figura 9.16.a y 9.16.b).

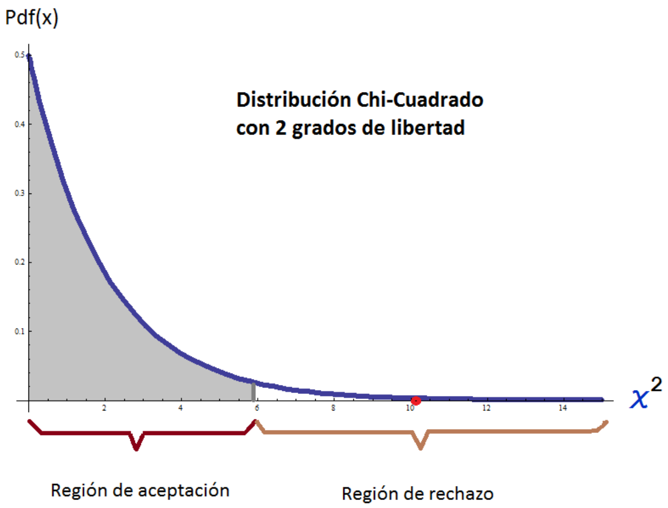
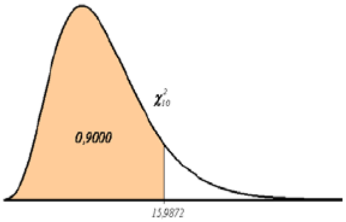


Figura 9.16.a: el valor de chi-cuadrado que separa la región de aceptación de la región de rechazo (χ_2^2 =5.99) está a la izquierda del estadístico H encontrado (punto rojo) igual a 10.25. Luego H se encuentra en la región de rechazo de H_0.



$P(X \leq \chi^2_{gl})$

g.l.	Probabilidad							
	0,005	0,010	0,025	0,050	0,100	0,900	0,950	0,
1	0,0000	0,0002	0,0010	0,0039	0,0158	2,7055	3,[illegible]415	
2	~~0,0100~~	~~0,0201~~	~~0,0506~~	~~0,1026~~	~~0,2107~~	~~4,[illegible]~~	5,9915	
3	0,0717	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	
4	0,2070	0,2971	0,4844	0,7107	1,0636	7,7794	9,4877	1

Figura 9.16.b: forma de encontrar el valor crítico de χ^2_2 (chi-cuadrado con dos grados de libertad) con un $\alpha = 0.05$ ya que, en este caso, $1-\alpha = 0.95$.

Ya hemos indicado que el test de Chi-Cuadrado es un test ómnibus porque se convierte en significativo (rechazo de H_0) en cuanto exista una única diferencia significativa entre un par de grupos. Pero no indica en dónde se encuentra esta diferencia. Por consiguiente, el siguiente paso cuando rechazamos H_0 es determinar entre qué par de tratamientos se producen las diferenciaas. Como tenemos tres tratamientos, tenemos asimismo tres comparaciones: a_1 vs a_2, a_1 vs a_3 y a_2 vs a_3. El test ómnibus nos ha indicado que, como mínimo, una de estas comparaciones es significativa. Ahora debemos encontrar cuál de ellas lo es.

Para lograrlo seguiremos un proceso similar al de las **comparaciones múltiples** en el ANOVA. En primer lugar hay que determinar cuál es la diferencia mínima crítica por encima de la cual las diferencias entre promedios de pares de rangos resultarán significativas. Cuando el diseño es equilibrado, esta diferencia crítica será la misma para todas las comparaciones dos a dos entre grupos, pero con diseños no equilibrados, habrá una diferencia crítica diferente para cada par de tratamientos con tamaños muestrales desiguales. La fórmula para su cálculo es la siguiente:

$$MV_{KW} = z_{adj}\sqrt{\frac{N(N+1)}{12}\left(\sum \frac{1}{n_a}\right)}$$

siendo z_{adj} el percentil $1-\dfrac{\alpha}{2\cdot c}$ de la distribución normal tipificada, en donde *c* es el número de comparaciones a realizar y siendo el contraste bilateral . Si se realizan comparaciones unilaterales el valor de z_{adj} se calcula como $1-\dfrac{\alpha}{c}$. Serán significativas las diferencias de rangos que cumplan la condición:

$$\left|\overline{R}_i - \overline{R}_j\right| > MV_{KW}$$

Para los datos del ejemplo numérico que estamos desarrollando, habrá que determinar una diferencia mínima crítica por cada par de tratamientos al tener un tamaño muestral desigual. A continuación aparecen los cálculos para las tres diferencias posibles en nuestro diseño (a_1 vs a_2, a_1 vs a_3 y a_2 vs a_3):

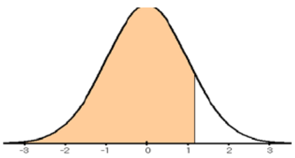
$$MV_{KW}(a_1 - a_2) = 2{,}39\sqrt{\frac{18(18+1)}{12}\left(\frac{1}{6}+\frac{1}{7}\right)} = 7.0985$$

$$MV_{KW}(a_1 - a_3) = 2{,}39\sqrt{\frac{18(18+1)}{12}\left(\frac{1}{6}+\frac{1}{5}\right)} = 7{,}7260$$

$$MV_{KW}(a_2 - a_3) = 2{,}39\sqrt{\frac{18(18+1)}{12}\left(\frac{1}{7}+\frac{1}{5}\right)} = 7{,}4710$$

Al haber en total c = 3 comparaciones, el valor el valor de z_{adj} para el percentil

$1-\dfrac{\alpha}{2\cdot c} = 1-\dfrac{0.05}{2\times 3} = 0{,}9916$ vale 2.39, como puede comprobarse en las tablas de la distribución normal tipificada (véase la Figura 9.17).



P(Z ≤ z)

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
…	…	…	…	…	…	…	…	…	…	…
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952

Figura 9.17

Y las diferencias entre los rangos medios de los grupos son:

$$\left|\overline{R}_{a_1} - \overline{R}_{a_2}\right| = |4{,}5 - 10| = 5{,}5 < 7.0985 \qquad no\,significativa\,(n.s.)$$

$$\left|\overline{R}_{a_1} - \overline{R}_{a_3}\right| = |4{,}5 - 14{,}8| = 10{,}3 > 7.7260 \qquad significativa$$

$$\left|\overline{R}_{a_2} - \overline{R}_{a_3}\right| = |10 - 14{,}8| = 4{,}8 < 7.4710 \quad no\,significativa\,(n.s.)$$

Observamos que de las tres comparaciones, sólo la de los grupos a_1 y a_3 es significativa en términos de diferencias en puntuación en ansiedad.

9.7.- Contrastes no paramétricos para diseños con más de dos grupos dependientes.

9.7.1.- Test de Friedman

El test de Friedman es la alternativa no paramétrica al ANOVA de un factor intra-sujetos. Por consiguiente, la hipótesis que se pone a prueba con este test no hace referencia a la media poblacional de la variable dependiente (a la que llamaremos Y) sino a la mediana poblacional (que denotaremos por la letra griega eta η) cuando disponemos de un factor con más de dos niveles y que ha sido manipulado de manera intra-sujeto. Como todos los tests no paramétricos, el test de Friedman hará uso de los rangos de las puntuaciones originales por lo que se perderá información con respecto a la utilización de las puntuaciones originales. Esto significa que la utilización de este test no paramétrico solo está justificada cuando alguno de los supuestos básicos del ANOVA paramétrico no se cumple de forma obvia y extrema. Esta es la razón por la que muchos investigadores se muestran poco proclives a la utilización del test de Friedman. No obstante, la utilización del ANOVA paramétrico sin cumplirse los supuestos del mismo tampoco está justificada.

Por su parte, la utilización de la alternativa no paramétrica puede deberse bien a que la variable dependiente ha sido medida en una escala ordinal (en este caso se incumpliría el supuesto del test paramétrico que exige una variable dependiente medida a nivel de intervalo o de razón) o bien porque los datos han sido transformados desde escalas de nivel superior a escalas de nivel inferior por el incumplimiento de algún otro supuesto.

Condiciones y supuestos: los supuestos en los que se basa el ANOVA no paramétrico de Friedman son mucho más generales que los necesarios para aplicar el ANOVA paramétrico, asegurando de esta forma que los conjuntos de datos que cumplen estas condiciones y a los que se puede aplicar, sea mucho mayor. La contrapartida radica en que será más difícil encontrar significativo el factor manipulado (recuérdese que factor es sinónimo de variable independiente en el área del ANOVA). Estos supuestos son:

- La muestra de unidades de observación (usualmente, participantes en un experimento) ha sido seleccionada al azar de la población que representa.

- La variable dependiente que se mide es una variable aleatoria continua que es posteriormente traducida a una escala ordinal. No obstante, este supuesto se viola frecuentemente aplicándose también a variables aleatorias discretas.

Además, es importante asegurar que los efectos del orden de presentación de las distintas condiciones se ha contrabalanceado o aleatorizado. Al utilizar rangos en vez de las puntuaciones originales, esta exigencia usual del método experimental tiene más importancia, si cabe, si se utiliza este procedimiento de análisis que en su alternativa paramétrica.

Formulación de las hipótesis: La hipótesis nula y alternativa tienen la misma forma que ya vimos en el ANOVA paramétrico exceptuando que el parámetro poblacional al que vamos a inferir nuestros resultados es la mediana, no la media. Por consiguiente:

$$H_0 : \eta_1 = \eta_2 = \eta_3 = ... = \eta_a$$
$$H_1 : \eta_1 \neq \eta_2 \neq \eta_3 \neq ... \neq \eta_a$$

La hipótesis alternativa debería leerse en el sentido de que hay, al menos, un par de medianas poblacionales que difieren entre sí del total de *a(a-1)/2* comparaciones posibles (siendo *a* el número de niveles en que se ha manipulado el factor) y no en el sentido de que todas las medianas poblacionales deben ser diferentes. Ya vimos que este razonamiento es lo que caracteriza a los tests ómnibus o globales: nos indican si existe alguna comparación significativa de entre todas las posibles (con que sólo haya una significativa, el test global será también significativo), pero no nos indica cuál o cuáles lo son.

Para comparar los resultados que obtendríamos en un test paramétrico vs. el correspondiente no paramétrico, retomamos los datos del primer ejemplo del tema 6. Comprobamos previamente que se cumplen los supuestos:

- Asumimos que los 6 sujetos del ejemplo han sido seleccionados al azar de la población de adultos sanos.

- La variable dependiente (tiempo de reacción medio) que se ha medido en tres condiciones ($a = 3$) es una variable aleatoria continua. Por consiguiente, si queremos aplicar el test de Friedman tendremos que utilizar los rangos de las puntuaciones para cada sujeto y no las puntuaciones directas. En la Tabla 9 aparece repetida la Tabla 6.1

con los datos originales pero, además, se han introducido tres columnas nuevas que codifican los rangos de las puntuaciones para cada sujeto. Esto significa que los rangos no se calculan para el conjunto global de puntuaciones (como en los tests anteriores) sino para cada sujeto: un conjunto de rangos para cada sujeto. Es decir, ordenamos las puntuaciones obtenidas por <u>cada sujeto</u> y le asignamos a cada una su rango, de menor a mayor (el valor 1 a la puntuación más pequeña, el 2 a la siguiente y así sucesivamente). De esta forma, para el primer sujeto, observamos que las puntuaciones originales siguen el orden ascendente $a_1 = 0.545 < a_3 = 0.620 < a_2 = 0.832$, luego a la condición congruente le asignaremos el orden 1, a la condición neutral el orden 2 y a la condición incongruente el orden 3 por lo que las puntuaciones originales 0.545, 0.832 y 0.620 se ven transformadas en los rangos 1, 3 y 2. Este mismo proceso se repite independientemente para cada sujeto.

La obtención de los rangos de menor a mayor podría invertirse y considerar los rangos de mayor a menor ya que el test de Friedman produciría el mismo valor, pero procederemos calculando los rangos de menor a mayor por ser más intuitivo.

Si hubiesen empates se asignaría a cada puntuación empatada la media de los rangos originales. En el ejemplo de la Tabla 9 no se han producido empates ya que con TR es muy difícil que dos puntuaciones sean exactamente iguales, pero en otras situaciones esto puede resultar factible.

En la Tabla 9.10 también aparece el sumatorio de los rangos para cada condición experimental así como la media de los mismos, valores todos ellos que necesitaremos posteriormente para aplicar el estadístico de contraste.

Tabla 9.10

Niveles del factor "Condición de Stroop" (A)

	Condición congruente a_1		Condición incongruente a_2		Condición neutral a_3	
Participante	TR	Rango	TR	Rango	TR	Rango
1	0.545	1	0.832	3	0.620	2
2	0.630	1	0.736	3	0.635	2
3	0.610	1	0.663	2	0.680	3

4	0.680	2	0.715	3	0.660	1
5	0.590	1	0.880	3	0.700	2
6	0.600	1	0.790	3	0.670	2
$\sum_{i=1}^{I} R_i$		$R_{a_1} = 7$		$R_{a_2} = 17$		$R_{a_2} = 12$
$\overline{R}_a = \frac{\sum_{i=1}^{6} R_i}{I}$		$\overline{R}_{a_1} = 7/6 = 1.17$		$\overline{R}_{a_2} = 2.8\widehat{3}$		$\overline{R}_{a_3} = 2$

Estadístico de contraste: el estadístico de contraste utilizado para evaluar la H_0 es el estadístico de Friedman que se distribuye según χ^2 con *a-1* grados de libertad:

$$\chi^2 = \frac{12}{n \cdot a(a+1)} \left[\sum_{i=1}^{a} R_{a_i}^2 \right] - 3 \cdot n \cdot (a+1)$$

siendo a el número de niveles del factor, n el número de sujetos y $\sum_{i=1}^{a} R_{a_i}^2$ el sumatorio de rangos al cuadrado para cada nivel del factor.

Aplicando el estadístico χ^2 a los datos del ejemplo de la Tabla 9.10 obtendríamos un valor de 8.33 para el estadístico de contraste.

$$\chi^2 = \frac{12}{n \cdot a \cdot (a+1)} \left[\sum_{i=1}^{a} (R_{a_i})^2 \right] - 3 \cdot n \cdot (a+1)$$
$$= \frac{12}{6 \cdot 3 \cdot (3+1)} \left[7^2 + 17^2 + 12^2 \right] - 3 \cdot 6 \cdot (3+1) \approx 8.3\widehat{3}$$

Regla de decisión: para evaluar si el valor muestral del estadístico de contraste es compatible con la hipótesis nula o no, es necesario confrontarlo con el valor crítico de χ^2 con *a-1* grados de libertad al nivel especificado de confianza (α). En nuestro caso disponemos de una distribución muestral del estadístico χ^2 con 3-1 = 2 grados de libertad, cuya función de densidad de probabilidad (f.d.p.) puede verse en la Figura 9.18:

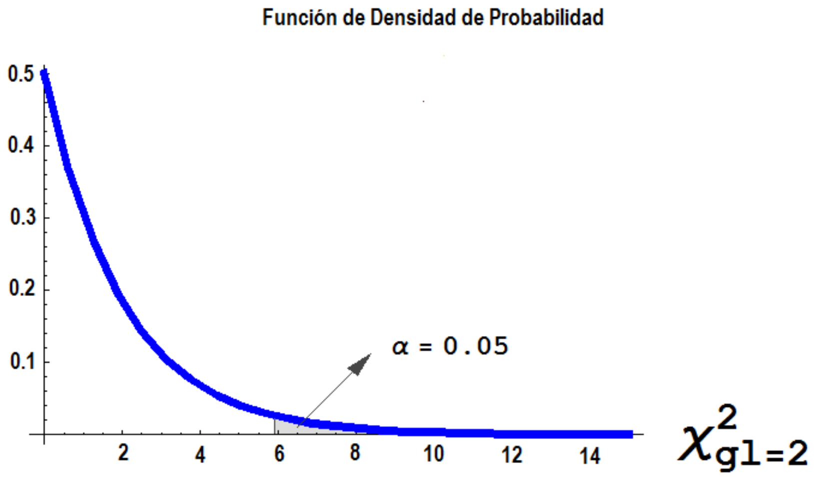
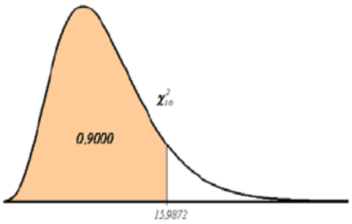


Figura 9.18: Distribución de densidad de probabilidad para la distribución χ^2 con dos grados de libertad. En la Figura se ha señalado en color gris el área de la función que deja por encima de sí el 0.05 del área total. El valor de χ^2 que separa la zona de aceptación de H_0 de la zona de rechazo vale 5.99. El pequeño número de grados de libertad es el responsable de que la forma de esta distribución no tenga la forma "normativa" de χ^2 que puede verse en la Figura 6.1 (Tema 6).

Según la tabla de la distribución Chi-Cuadrado, para un nivel de confianza del 95% ($\alpha = 1 - 0.95 = 0.05$), el valor que separa la zona de aceptación de la zona de rechazo vale 5.99 (véase Figura 9.19).



$$P\left(X \le \chi^2_{gl}\right)$$

g.l.	Probabilidad							
	0,005	0,010	0,025	0,050	0,100	0,900	0,950	0,9
1	0,0000	0,0002	0,0010	0,0039	0,0158	2,7055	3,8415	5,
2	0,0100	0,0201	0,0506	0,1026	0,2107	4,60[illegible]	5,9915	7,
3	0,0717	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	9,
4	0,2070	0,2971	0,4844	0,7107	1,0636	7,7794	9,4877	11,
5	0,4117	0,5543	0,8312	1,1455	1,6103	9,2364	11,0705	12,
6	0,6757	0,8721	1,2373	1,6354	2,2041	10,6446	12,5916	14,
7	0,9893	1,2390	1,6899	2,1673	2,8331	12,0170	14,0671	16,

Figura 9.19

En consecuencia, como el valor muestral de χ^2 obtenido supera este valor (5.99 < 8.33), rechazamos H_0.

Interpretación: la interpretación de estos resultados indica que para el estadístico de tendencia central “mediana”, los resultados del experimento muestran que, al menos, existen dos condiciones del factor “Condición de Stroop” que difieren estadísticamente entre sí en los datos de la Tabla 9. Estos resultados podrían resumirse indicando que el factor principal “Condición de Stroop” ha resultado significativo según el test de Friedman, $\chi^2_2 = 6.16, p < 0.05$.

Como precisiones añadidas, debemos indicar que si se produjeran un número elevado de empates en los rangos, debería utilizarse una corrección en la Fórmula para el cálculo de de χ^2. Asimismo, como hemos rechazado H_0 en el test ómnibus, el paso siguiente sería realizar un análisis de comparaciones *a posteriori* para determinar entre qué niveles del factor “Condición de Stroop” se han producido las diferencias. El test de Friedman nos ha indicado que existen diferencias entre dos o más niveles del factor pero no nos ha indicado dónde se han producido. No obstante, estas precisiones superan el nivel del curso y el alumno debe consultar la bibliografía pertinente (*v.g.*, Sheskin, 2003).

9.6.2.-Test de Cochran

Hasta el momento hemos visto los tests paramétrico y no paramétrico para cuando disponemos de más dos condiciones experimentales manipuladas intra-sujetos. El primero se utilizaba cuando, cumpliéndose el resto de supuestos, la escala de la variable dependiente era de nivel superior (intervalo o de razón), mientras que el segundo se utilizaba cuando se producían violaciones importantes de los supuestos del ANOVA o cuando la escala de medida no llegaba al nivel de intervalo (es decir, cuando la escala era o se había transformado en ordinal). La pregunta lógica que se puede plantear a continuación es si existen tests para aplicar cuando la escala es del tipo más bajo posible: escala nominal o categórica. Esta situación se puede resolver mediante el test Q de Cochran.

El test Q de Cochran se utiliza cuando en un experimento se ha manipulado un factor o variable independiente con 2 o más condiciones experimentales pero la variable dependiente que se ha medido es dicotómica (sí/no, pasa/no pasa, apto/no apto/, enferma/no enferma, etc.).

Condiciones y supuestos: Los supuestos son similares a los que ya tuvimos ocasión de contemplar en el test no paramétrico de Friedman:

- Se asume que la muestra de los *n* sujetos sometidos a las *a* condiciones experimentales han sido seleccionados al azar de la población que representan.
- Las puntuaciones de los sujetos en la variable dependiente implican una variable dicotómica con dos categorías mutuamente exclusivas.

Debiéndose cumplir también el resto de condiciones que citamos en el test de Friedman, podemos observar que la diferencia básica entre ambos tests (Friedman y Cochran) es el nivel de medida de la variable dependiente.

Ejemplo: Un investigador desea investigar el efecto del formato en el que se presentan las cuestiones en un test de evaluación educativa de un idioma extranjero sobre la ejecución de los sujetos. Para ello, decide presentar cuestiones sobre el conocimiento del idioma extranjero en tres tipos de formato: papel y lápiz, en formato exclusivamente auditivo y en formato audio-visual. El investigador decide calificar a cada sujeto exclusivamente como Apto o No Apto. Somete a los tres tipos de formatos pero con distintas cuestiones a 12 sujetos participantes y los resultados pueden verse en la Tabla 9.11. En la misma se ha codificado la categoría "Apto" mediante un 1 en el grueso de la tabla mientras que la categoría "No Apto" se ha codificado mediante un 0. Por otra parte, la columna codificada como R_j indica el sumatorio de valores para cada sujeto (señalizado mediante el subíndice *j*). De esta forma, el primer sujeto que superó los tests de papel y lápiz así como el audio-visual pero no el auditivo, obtendría un 2 ya que ha obtenido un 1 en el test de papel y lápiz, un 1 en el audio-visual y un 0 en auditivo. La suma de estas puntuaciones es la puntuación alcanzada por este sujeto. El segundo participante en el experimento sólo superó el test auditivo obteniendo un 1, y así sucesivamente. Por otra parte, se ha calculado para cada condición (denotadas por a_1, a_2 y a_3) el sumatorio de valores codificándose como A_i. Esto significa que A_1 es la suma de los valores 1 existentes para la primera condición (segunda columna), A_2 es la suma de los valores 1 existentes para la segunda condición (tercera columna) y A_3 lo mismo para la tercera condición.

Tabla 9.11

Sujetos	Formato de la evaluación			R_j	R_j^2
	Papel y lápiz a_1	Auditivo a_2	Audio-visual a_3		
1	1	0	1	2	4
2	0	1	0	1	1
3	1	0	1	2	4
4	1	1	1	3	9
5	0	0	0	0	0
6	1	1	0	2	4
7	0	1	1	2	4
8	1	0	1	2	4
9	0	1	1	2	4

10	0	1	0	1	1
11	1	1	1	3	9
12	1	1	0	2	4

$$A_1 = \sum_{j=1}^{12} a_1 = 7 \quad A_2 = \sum_{j=1}^{12} a_2 = 8 \quad A_3 = \sum_{j=1}^{12} a_3 = 7$$

De estos valores nos interesa calcular, para la aplicación posterior del estadístico de contraste, los siguientes valores.

El sumatorio de los A_i al cuadrado (valor al que llamaremos C):

$$C = \sum_{i=1}^{a} A_i^2 = 7^2 + 8^2 + 7^2 = 162$$

El sumatorio del número de éxitos por sujeto (al que llamaremos T):

$$T = \sum_{j=1}^{12} R_j = 22$$

Y el sumatorio del número de éxitos por sujeto elevado al cuadrado (al que llamaremos D)

$$D = \sum_{j=1}^{12} R_j^2 = 48$$

Nota: la codificación de la Tabla 9.11 es tal que el valor 1 significa que el sujeto ha aprobado la evaluación en ese formato mientras que un valor de 0 significa que el sujeto no ha superado la evaluación en ese formato.

Formulación de las hipótesis: Debido a que la variable dependiente se ha medido en una escala dicotómica, ahora no podemos formular las hipótesis nula y alternativa en términos de tendencia central (media o mediana) debido a que para una variable dicotómica sólo resulta adecuado hablar del porcentaje de casos positivos o negativos e incluso el sentido de positivo/negativo es arbitrario (aunque una vez codificado en un sentido debe respetarse para interpretarse correctamente el análisis). Por todo ello introduciremos el símbolo π (pi minúscula, no confundir con el número Pi) para representar la proporción de casos positivos (usualmente codificados por 1) en la población. Como disponemos de varias poblaciones (cada una de las poblaciones representada por las condiciones experimentales o niveles del factor) debemos diferenciarlas mediante un subíndice. Utilizaremos el subíndice "a_i. En consecuencia, la hipótesis nula plantea que la proporción poblacional de todas las condiciones experimentales es la misma:

$$H_0: \pi_{a_1} = \pi_{a_2} = \pi_{a_3} = \ldots = \pi_a$$
$$H_1: \pi_{a_1} \neq \pi_{a_2} \neq \pi_{a_3} \neq \ldots \neq \pi_a$$

Advertimos, de nuevo que la hipótesis alternativa debería leerse en el sentido de que hay, al menos, un par de proporciones poblacionales que difieren entre sí del total de *a(a-1)/2* comparaciones posibles y no en el sentido de que todas las proporciones poblacionales deben ser diferentes entre sí para que H_0 sea cierta.

Estadístico de contraste: El estadístico de contraste es la Q de Cochran:

$$Q = \frac{(a-1)\left[a \cdot C - T^2\right]}{a \cdot T - D}$$

siendo C, T y D los valores previamente calculados inmediatamente después de la Tabla 10 y *a* el número de condiciones experimentales o niveles del factor manipulado.

Para nuestros datos de ejemplo, el estadístico Q alcanza el valor:

$$Q = \frac{(a-1)\left[a \cdot C - T^2\right]}{a \cdot T - D} = \frac{(3-1)\left[3 \cdot 162 - 22^2\right]}{3 \cdot 22 - 48} = 0.2\widehat{2}$$

Regla de decisión: El estadístico Q se diseñó con el objetivo de que se distribuyera según χ^2 en todas las muestras posibles, de tal forma que serán las tablas de χ^2 con *a-1* grados de libertad y al nivel de confianza precisado en el ejercicio que estemos resolviendo las que nos permitirán determinar el valor crítico con el que comparar el valor Q obtenido. Para nuestro ejemplo y trabajando a un nivel de confianza del 95%, tendríamos que el valor de χ^2 crítico es $\chi^2_{gl=2} = 5.99$, el mismo valor que para el test de Friedman ya que los parámetros son idénticos (los parámetros se refiere a los grados de libertad -2 en este caso- y el nivel de confianza).

En este caso, no podemos rechazar H_0 ya que el valor de Q es inferior al valor crítico, es decir, Q < χ^2 ya que 0.22 < 5.99. Solamente si Q hubiera superado el valor tabular de χ^2 habríamos podido rechazar H_0.

Interpretación: En este caso, el formato en el que se le presentan las preguntas de test a los sujetos es indiferente con respecto a la superación de la evaluación ya que los tres formatos no difieren entre sí en el número de éxitos alcanzados por los participantes.