TEMA Nº 1 → ESTIMACIÓN DE PARÁMETROS Y CONTRASTE DE HIPÓTESIS

1.- INFERENCIA ESTADÍSTICA

*Estudio de las muestras para conocer la población a la que representan. La inferencia siempre se hace en términos probabilísticos (afirmamos con una cierta probabilidad de éxito). **El error muestral** es la diferencia entre el resultado obtenido en la muestra y el que habríamos obtenido si se hubiese trabajado con la población.*
*Para cada característica de la muestra que evaluemos se obtiene lo que se conoce como **estadístico**: índices descriptivos de centralidad (Media), variabilidad (Varianza), etc. A partir de los estadísticos obtenidos en la muestra (lo concreto), se realizan afirmaciones sobre los parámetros de la población (lo general)*

DISTRIBUCIONES DE PROBABILIDAD →

Distribución Poblacional: *(Distribución de frecuencias que presenta la variable en la población sobre la que se quiere generalizar). Las medidas en la población se denominan parámetros poblacionales y se designan con letras griegas → (Parámetros: Media de la población = μ , Varianza de la población = σ^2, Proporción de la población = π)*

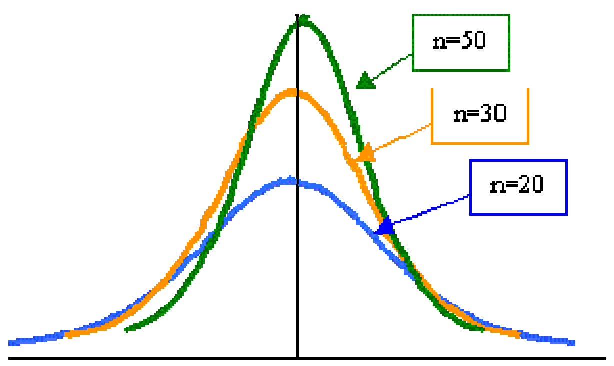
Distribución de la Muestra: *(Distribución de frecuencias que presenta la variable en la en la muestra con la que se trabaja). Las medidas realizadas en la muestra se denominan estadísticos y se designan con letras latinas mayúsculas → (Estadísticos: Media de la muestra = **Y**, Varianza de la muestra = S^2, Proporción de la muestra = **P**)*

Distribución Muestral de un estadístico*: (Distribución de frecuencias que presenta el estadístico que vamos a utilizar como base del proceso de inferencia para estimar los parámetros). Resulta de obtener todas las muestras posibles (de un determinado tamaño) de una población y medir en ellas una determinada característica. Las distribuciones muestrales que más vamos a utilizar son**: la media, la varianza y la proporción.***

DISTRIBUCIÓN MUESTRAL DE LA MEDIA →

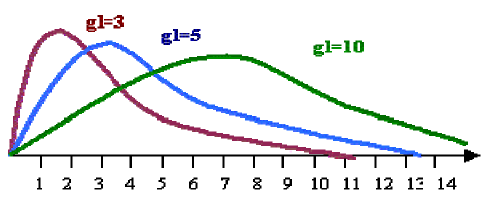
Teorema Central del límite: *Si una población tiene una media μ y una varianza σ^2 finitas, la distribución de las medias muestrales de tamaño "n" extraídas de manera aleatoria e independiente, se aproxima a la forma de una **distribución normal con varianza σ^2/n y media μ** conforme el tamaño de n se va incrementando.*

$$N(\mu, \sigma/\sqrt{n})$$



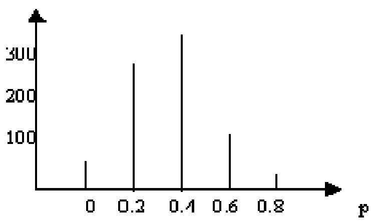
La Distribución muestral de la media *(media de todas las medias posibles): Se trata de una distribución de probabilidad conocida que viene recogida en las tablas. Es **Normal** N (0, 1) cuando lo es la distribución de la variable estudiada (al margen del tamaño de la muestra) y tiende a la normal a medida que n va aumentando (al margen de la distribución de la variable).*
*La distribución muestral de la media se ajusta a la distribución **t de Student** (con n-1 grados de libertad), si ignoramos la forma de la distribución de la variable y/o el tamaño de la muestra es pequeño.*

DISTRIBUCIÓN MUESTRAL DE LA VARIANZA →



La varianza es un índice de dispersión que permite determinar la homogeneidad de la variable de estudio. La ***distribución muestral de la varianza*** *se ajusta a la distribución Chi-cuadrado (con n-1 grados de libertad).La cuasi varianza muestral (*S^2_{n-1}*) es la mejor estimación de la varianza poblacional (*σ^2*). Tiende a la normal a medida que aumenta el tamaño de la muestra (mayor de 100 sujetos)*

DISTRIBUCIÓN MUESTRAL DE LA PROPORCIÓN →



Se trata de variables dicotómicas o dicotomizadas; normalmente éxito o fracaso. La ***distribución de la proporción poblacional*** *se ajusta al modelo binomial con parámetros* ***n*** *y* ***π****. La distribución binomial se aproxima a la normal a medida que el tamaño de la muestra va aumentado (Teorema Central del Límite), con parámetros →* $N(\pi,\ \sigma_p)$

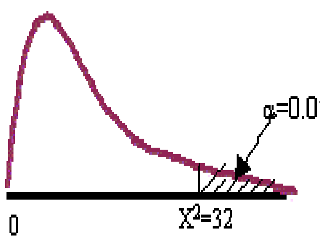
	D. Muestral (Media)	**D muestral (Varianza)**	**D. Muestral (Proporción)**
Media	$E(\bar{X}) = \mu$	$E(S^2_{n-1}) = \sigma^2$	$\mu_p = \pi$
Desviación Típica	$\sigma_x = \sigma / \sqrt{n}$ **Error Típico Media**	$\sigma_{S^2_{n-1}} = \sigma^2 \cdot \sqrt{2/n-1)}$ **Error Típico Cuasi Varianza**	$\sigma_p = \sqrt{\pi(1-\pi)/n}$ **Error Típico Proporción**
Tipificación	$Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ y $T = \dfrac{\bar{X}-\mu}{S_{n-1}/\sqrt{n}}$ **D. Normal D. T Student**	$X^2_{n-1} = \dfrac{(n-1)\,S^2_{n-1}}{\sigma^2}$ **Distribución** $X^2_{n-1\ gl}$	$Z = \dfrac{P-\pi}{\sqrt{\pi(1-\pi)/n}}$ **Distribución Normal**

PROBLEMAS EJEMPLO

El CI de los alumnos de un centro de E. Especial se distribuye normalmente con $\mu = 80$ *y* $\sigma = 10$*. Si de esta población extraemos una muestra aleatoria de 25 alumnos. ¿Cuál es la probabilidad de obtener una media mayor de 75 puntos?*

$$P(\bar{X} \geq 75) \rightarrow Z = \frac{75-80}{10/\sqrt{25}} \rightarrow Z = (-2'50) \text{ se corresponde con } p = 0'0062 \rightarrow 1 - 0'0062 = 0'9938$$

Los tiempos requeridos por un cierto autobús para alcanzar uno de sus destinos en una ciudad grande forman una distribución normal con una desviación típica $\sigma = 1$ *minuto. Si se elige al azar una muestra de 17 tiempos, encuentre la probabilidad de que la cuasi varianza muestral sea mayor que 2.*



Se busca el valor de ji-cuadrado correspondiente a $S^2_{n-1}=2$

$$X^2_{n-1} = \frac{(n-1)\,S^2_{n-1}}{\sigma^2} \rightarrow X^2_{n-1} = \frac{16 \cdot 2}{1^2} = 32$$

El valor de 32 se busca en la tabla Chi Cuadrado con 16 grados de libertad. A este valor le corresponde una probabilidad de 0,99. En consecuencia, 1-0´99 = 0´01→ $P(S^2_{n-1} > 2)$

Un partido político cree que el 60% del electorado está a favor de su programa. Como su líder encuentra que esta predicción es demasiado optimista decide hacer un sondeo con una muestra de 90 personas. ¿Cuál es la probabilidad de que como máximo 60 personas estén a favor de su partido?

$$P\,(60/90 = 0{'}67) \rightarrow \mathbf{Z = \frac{0{'}67 - 0{'}60}{\sqrt{0{'}6 \cdot 0{'}4 / 90}}} \rightarrow Z = (1{'}35) \text{ se corresponde con } p \mathbf{= 0{'}9115}$$

2.- ESTIMACIÓN DE PARÁMETROS

Generalmente se desconocen los parámetros de la población por lo que se hace necesario estimarlos a partir de los valores muestrales. Un estimador es un estadístico que utilizamos para estimar parámetros. Se pueden considerar dos formas de estimar los parámetros poblacionales: ***estimación puntual*** *(Conocida como método de los momentos de Pearson, se extrapola el estadístico de la muestra directamente a la población) y* ***estimación por intervalos*** *(establece un rango de valores dentro del cual estaría el valor del parámetro, con una determinada probabilidad)*

Para que el estimador represente correctamente al parámetro →r ***cuatro propiedades básicas:***

1.- ***Carencia de sesgo:*** *un estimador es insesgado o centrado cuando el valor del estadístico (en las infinitas muestras de tamaño "n" extraídas de una población) coincide con el valor del parámetro que queremos estimar →* ***U*** *es un estimador insesgado de* $\boldsymbol{\theta}$*, si →* $\mathbf{E\,(U) = \theta}$
La media, la proporción y la cuasi varianza de la muestra son estimadores insesgados de sus valores poblacionales
2.- ***Eficiencia (Precisión):*** *Se considera la inversa de la varianza de su distribución muestral. Cuanto mayor es el cociente, mayor es la eficiencia.* ***Eficiencia →*** $\boldsymbol{\theta = 1 / \sigma^2_0}$ *(A mayor varianza, menor eficiencia). La Media Aritmética es más eficiente que la Mediana. La Varianza es más eficiente que la Cuasi varianza. Cuando tenemos distintos estimadores y queremos determinar el más eficiente, se comparan sus eficiencias.* ***Ejemplo:*** *La varianza de la distribución muestral de dos estimadores es 2 y 1,5. Para un mismo tamaño muestral, la eficiencia relativa sería 1,5 / 2 = 0,75. Si el cociente fuera 1 serían iguales.*
3.- Suficiencia: *el estimador utiliza toda la información de la muestra para estimar el parámetro (****Ejemplo:*** *La media muestral sería suficiente para estimar la media poblacional. No lo sería la amplitud intercuartílica para estimar la varianza poblacional)*
4.- Consistencia: *El requisito mínimo que se le exige a un estimador es que sea consistente. Un estimador es consistente si, a medida que se dispone de más información (que aumenta el tamaño de la muestra), aumenta la probabilidad de que la estimación coincida con el parámetro. La media, la proporción y la varianza insesgada son consistentes porque son estimadores insesgados de los parámetros correspondientes y en sus límites valen cero (su sesgo y su varianza tienden a 0 a media que aumenta n)*

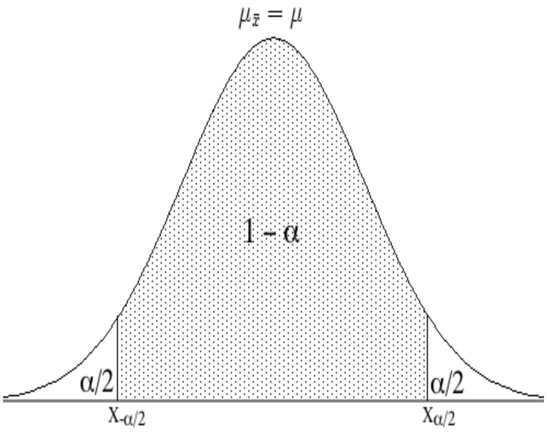
Cuando el valor del parámetro (población) coincide con el valor del estadístico (muestra), se considera que la estimación (inferencia sobre la población a partir de la muestra) es insesgada. Siempre que operemos se debe procurar que nuestros estimadores sean insesgados y tengan una varianza pequeña; estas dos características se denominan ***acuracidad.***

Resumen de las propiedades de los principales estadísticos				
	Carencia de sesgo	*Eficiencia*	*Suficiencia*	*Consistencia*
Media Aritmética	*SI*	*> Mediana*	*SI*	*SI*
Mediana	*NO*	*< Media*	*NO*	*SI*
Proporción	*SI*	*-----*	*SI*	*SI*
Varianza	*NO*	$> S^2_{n-1}$	*SI*	*SI*
Cuasi varianza	*SI*	$< S^2$	*SI*	*SI*

ESTIMACIÓN POR INTERVALOS

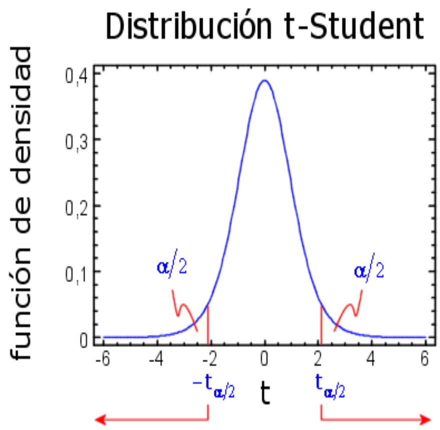
Una estimación por intervalos expresa el grado de confianza con el que se espera que esté el valor del parámetro dentro del intervalo, por lo que se suele llamar intervalo de confianza. La amplitud del intervalo nos indicará su precisión. ***A menor amplitud, más precisión****, más informativo es, más útil.*
Una estimación por intervalos ***depende de cuatro parámetros****: una estimación puntual del parámetro; una medida de variabilidad; una probabilidad (nivel de confianza) y un supuesto acerca de la distribución en la población.*

INTERVALO DE CONFIANZA PARA LA MEDIA



N(0,1)

Distribución Normal (Conocida σ^2)



$t_{n-1\ gl}$

Desconocida la σ^2 (Muestras pequeñas)

$\bar{X} + (Z_{\alpha/2}) \cdot (\sigma_X) < \mu < \bar{X} + (Z_{1-\alpha/2}) \cdot (\sigma_X)$	$\bar{X} + (t_{\alpha/2}) \cdot (S_{n-1}/\sqrt{n}) < \mu < \bar{X} + (t_{1-\alpha/2}) \cdot (S_{n-1}/\sqrt{n})$

<u>Problemas ejemplo</u>*: Se midieron los niveles de depresión en una muestra de 100 personas. Asumiendo un nivel de medida de intervalo y que la variable se distribuye normalmente en la población, se calculó la media de las puntuaciones y se obtuvo un valor de (Media = 8) y una cuasi desviación típica = 2.* ***Hallar los límites del intervalo de confianza*** *para la media de la población con un nivel de confianza del 99%.*

Datos *→ $\bar{X} = 8$ Distribución normal de la VD $\alpha = 0{,}01$ $S_{n-1} = 2$*

Desconocida la varianza poblacional → n = 100 (muestra grande: Distribución normal)

Intervalo de confianza → $\bar{X} \pm (Z\alpha \cdot \sigma_Y)$ = Límite superior y Límite inferior

$\alpha / 2 = 0{,}01/2 = 0{,}005$ $1 - (\alpha / 2) = 0{,}995$ $|Z\alpha| = 2{,}58$

Error Típico → $\sigma_X = S_{n-1} / \sqrt{n}$ → $\sigma_X = 2 / 10$ → 0,20

Error máximo de estimación → $|Z\alpha| \cdot \sigma_X = (2{,}58 \cdot 0{,}20) = 0{,}516$

Límites del intervalo de confianza → ***8 ± (2,58 · 0,20) = [7,484 y 8,516]***

Hallar los límites del intervalo de confianza *suponiendo que la muestra está compuesta por 25 personas, a un nivel de confianza del 95%.*

Datos *→ $\bar{X} = 8$ Distribución normal de la VD $\alpha = 0{,}05$ $S_{n-1} = 2$*

Desconocida la varianza poblacional → n = 25 (muestra pequeña: Distribución t de Student)

Intervalo de confianza → $\bar{X} \pm (t_{\alpha} \cdot S_{n-1} / \sqrt{n})$ *= Límite superior y Límite inferior*

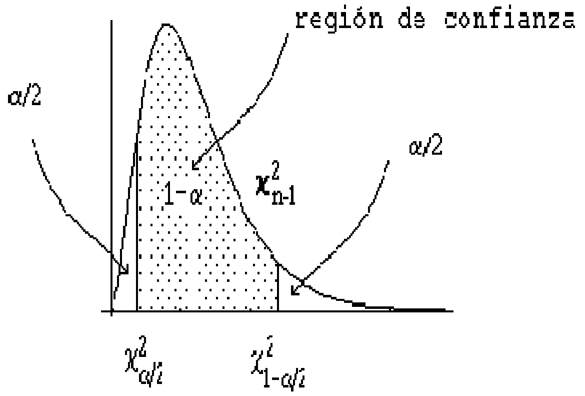
$\alpha / 2 = 0{,}05/2 = 0{,}025$ $\quad 1 - (\alpha / 2) = 0{,}975 \quad$ $|t\ \alpha| = 2{,}06$ *(Tablas t de Student)*

Error Típico → $S_{n-1} / \sqrt{n} \rightarrow 2/5 \rightarrow 0{,}4$

Error máximo de estimación → $|t_{\alpha}| \cdot S_{n-1} / \sqrt{n} = (2{,}06 \cdot 0{,}4) = 0{,}824$

Límites del intervalo de confianza → ***8 ± (2,06 · 0,4) = [7,176 y 8,824]***

INTERVALO DE CONFIANZA PARA LA VARIANZA



Distribución X^2_{n-1} *(Chi Cuadrado con n-1 gl)*

$$L_i = \frac{(n-1)\, S^2_{n-1}}{X^2_{(n-1)(1-\alpha/2)}} \quad \text{y} \quad Ls = \frac{(n-1)\, S^2_{n-1}}{X^2_{(n-1)\alpha/2}}$$

$$P\,(L_i < \sigma^2 < L_s) = 1 - \alpha$$

Los valores de X^2 ***se obtienen en la tabla*** X^2_{n-1} ***(Chi-Cuadrado con n-1 grados de libertad)***

Problemas Ejemplo: *En una muestra aleatoria de 20 sujetos, extraída de una población normal, se ha obtenido una media de 24 puntos y una cuasi desviación típica de 10,8 puntos. Averiguar los límites del intervalo de confianza para la* ***varianza de la población****, suponiendo α = 0,05.*

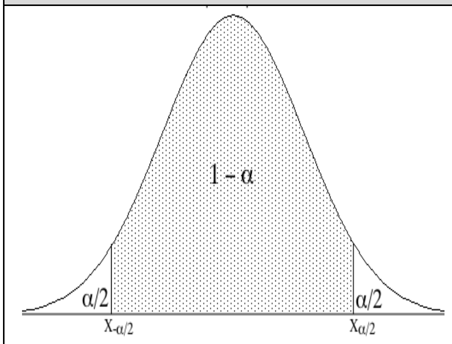
Estimamos la varianza de la población a partir de la varianza de la muestra (estimación por intervalo):

$$(n-1) \cdot S^2_{n-1} / (\chi^2_{\alpha/2}) < \sigma^2 < (n-1) \cdot S^2_{n-1} / (\chi^2_{1-\alpha/2})$$

Según las tablas χ^2_{n-1} (χ^2_{19}) *→ para (α / 2 = 0,025) → 8,91 y para (1 - α / 2 = 0,975) → 32,9*

$(19 \cdot 10{,}8^2) / 32{,}9 < \sigma^2 < (19 \cdot 10{,}8^2) / 8{,}91$ *→ Límites del intervalo de confianza →* ***[67,36 y 248,7]***

INTERVALO DE CONFIANZA PARA LA PROPORCIÓN



$$P + (Z_{\alpha/2}) \cdot (S_P) < \pi < P + (Z_{1-\alpha/2}) \cdot (S_P)$$

Donde : $S_P = \sqrt{P\,(1-P)/n}$

$P\,(L_i < \pi < L_s) = 1 - \alpha \qquad P = (L_i + L_s)/2$

$Z_{\alpha/2} \cdot S_p$ ***= Error máximo de estimación***

S_p ***= Error Típico (distribución muestral de la Proporción)***

Problema Ejemplo: *Para comprobar la eficacia en la aplicación de un tratamiento, se someten al mismo 64 pacientes. Finalizado el periodo de aplicación, se observó que remitió la enfermedad en 50 casos. Con un nivel de significación del 92% (α = 0,08), estime por intervalo el porcentaje de efectividad del tratamiento objeto de estudio.*

Datos: *Para (α / 2 = 0,04) → Z = (- 1,75) y Para (1 - α / 2 = 0,96) → Z = (+ 1,75)*

Proporción muestral → (50 / 64) = 0,781

Intervalo de confianza → $P \pm |Z_{\alpha/2}| \cdot (S_P)$ = Límites Superior e Inferior

$S_P = \sqrt{P(1-P)/n} \rightarrow S_P = \sqrt{0{,}781 \cdot (1-0{,}781)/64} = 0{,}0517$

$P \pm |Z_{\alpha/2}| \cdot (S_P) \rightarrow 0{,}781 \pm (1{,}75 \cdot 0{,}0517) =$ ***[0,87 y 0,69]***

Con un margen de error del 8% el tratamiento será efectivo entre el 69% y el 87% de los casos.

AMPLITUD DEL INTERVALO DE CONFIANZA Y RELACIÓN CON EL TAMAÑO MUUESTRAL

*La amplitud del intervalo de confianza depende de dos factores: el nivel de confianza y el error típico de la distribución muestral del estadístico (La suma de ambos **E = Error máximo de estimación**). Cuanto mayor es el tamaño de la muestra mayor es la precisión del intervalo y mayor la precisión de la estimación. Cuanto menor es el error típico, menor es el intervalo de confianza y, por tanto, más preciso (para reducirlo se aumenta el tamaño muestral)*

ESTADÍSTICO	***SUPUESTOS***	***FÓRMULA***
MEDIA	***Var. Poblacional conocida***	$n = (\sigma^2 \cdot Z^2_{\alpha/2}) / E^2$
	Var. Poblacional desconocida y muestra grande	$n = (S^2_{n-1} \cdot Z^2_{\alpha/2}) / E^2$
	Var. Poblacional desconocida y muestra pequeña	$n = (S^2_{n-1} \cdot {}_{\alpha/2}t^2_{n-1}) / E^2$
VARIANZA	***Error Típico (muestras grandes)***	$n = (2S^4_{n-1} \cdot Z^2_{\alpha/2}) / E^2$
PROPORCIÓN	***Error Típico (muestras grandes)***	$n = P \cdot (1-P) \cdot Z^2_{\alpha/2} / E^2$

Problemas Ejemplo: *Por experiencias anteriores se sabe que las estaturas de los soldados tienen una varianza de 64 cm. ¿Qué tamaño debe tener la muestra para que la media estimada no se aleje más de ± 1, 5 puntos de la media poblacional?: considere (α = 0,02).*

Para (α / 2 = 0,01) → $Z_{Tablas} = \pm 2{,}33$ Varianza poblacional conocida $\sigma^2 = 64$

$n = (\sigma^2 \cdot Z^2_{\alpha/2}) / E^2$ *máximo de estimación* → $n = (64 \cdot 2{,}33^2) / 1{,}5^2 = 154{,}42 \approx$ ***154 soldados***

Con un margen de error del 2% debemos tomar una muestra de 154 soldados.

➔ *Un estudio sobre la proporción de fumadores entre el personal de un hospital estableció que sólo fumaban el 35%. Si el análisis se efectuó con un nivel de confianza del 95%, ¿Qué tamaño debió tener la muestra para que la proporción estimada no se aleje más de ± 0,15 puntos de la proporción poblacional?: considere (α = 0,05).*

Para (α / 2 = 0,025) → $Z_{Tablas} = \pm 1{,}96$

$n = P \cdot (1-P) \cdot Z^2_{\alpha/2} / E^2$ *máximo de estimación* → $n = 0{,}35 \cdot 0{,}65 \cdot 1{,}96^2 / 0{,}15^2 =$ ***38,84 ≈ 39***

Con un margen de error del 5% debemos tomar una muestra de 39 personas.

3.- CONTRASTE DE HIPÓTESIS

Hipótesis estadística: *proposición (afirmación) sobre algún aspecto de la distribución de una población (parámetro, forma, etc.) que puede someterse a prueba a través de una muestra aleatoria de esa población. No tiene por qué suponerse interés científico.*

Contraste de hipótesis: *Procedimiento por el cual decidimos si una* ***propuesta sobre la población*** *puede aceptarse o no. Su finalidad es generalizar un resultado muestral a la población de la que procede la muestra. Siempre se formulan dos hipótesis (****exhaustivas y mutuamente excluyentes****); de tal modo, que el rechazo de una implica la aceptación de la otra.*
H_0 *(hipótesis nula): se acepta provisionalmente como verdadera y se somete a contraste.*
H_1 *(hipótesis alternativa): se acepta al rechazar la hipótesis nula.*

Dependiendo de cómo se formule la hipótesis hablamos de ***dirección del contraste****.*

BILATERAL O BIDIRECCIONAL	***UNILATERAL IZQUIERDO***	***UNILATERAL DERECHO***
➔ *La media (μ) es algún valor Igual o distinto a (X)* ***Ho : $\mu = X$*** ***$H_1 : \mu \neq X$***	➔ *La media (μ) es algún valor No inferior a (X).* ***Ho : $\mu \geq X$*** ***$H_1 : \mu < X$***	➔ *La media (μ) es algún valor No superior a (X).* ***Ho : $\mu \leq X$*** ***$H_1 : \mu > X$***

La decisión de utilizar contrastes unilaterales o bilaterales depende de la información o la idea del investigador sobre la tendencia de la variable en la población. Esta decisión influye en el nivel de significación (α) y en la interpretación de los resultados. El tipo de contraste es similar para la media (μ), la proporción (π) y la varianza (σ^2). En los tres tipos de contraste, ***el signo igual está en la H_0****. Las hipótesis siempre se hacen sobre los parámetros poblacionales.*

<u>Estadístico de Contraste (medida de discrepancia):</u> *Instrumento para tomar decisiones sobre la hipótesis nula, con una cierta probabilidad. La elección de un estadístico de contraste depende de las características de la variable en la población (forma, parámetros,...), de los datos muestrales (forma de obtenerlos, nivel de medida,...) y de la hipótesis estadística que se quiere contrastar (generada a partir de la hipótesis de investigación y los supuestos).*

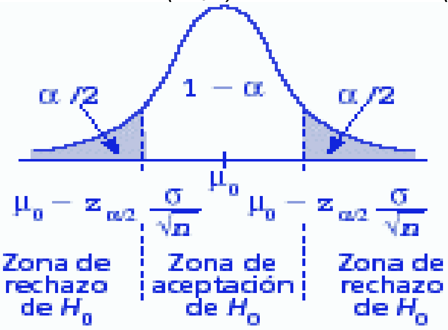
Estadístico de Contraste (Discrepancia)
Estadístico de Contraste Discrepancia = $\dfrac{\text{Valor del estadístico en la muestra} - \text{Valor del parámetro en la } H_0}{\text{Error típico de estimación}}$

Se trata de averiguar la ***discrepancia*** *entre los datos empíricos observados en la muestra y los datos teóricos que planteamos en la H_0.*

<u>Regla de decisión</u>: *Tomando en consideración* ***el valor o los valores críticos*** *(máxima diferencia que cabe esperar por simple azar entre los datos empíricos y los teóricos) y a partir del nivel de significación. Acumulado en los contrastes unilaterales (α) y dividido por dos en los bilaterales (α/2)*

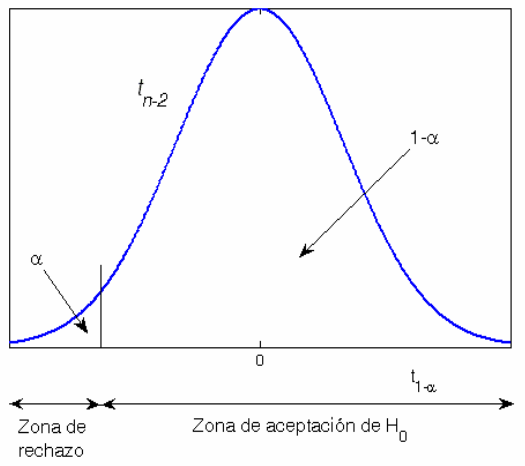
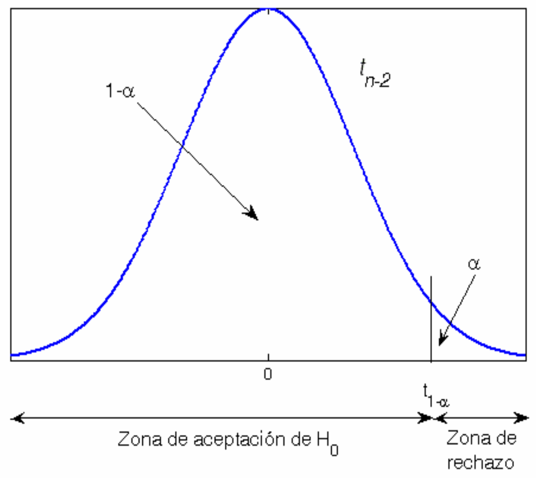
CONTRASTE BILATERAL

Valor crítico ($Z_{\alpha/2}$) *Valor crítico ($Z_{\alpha/2}$)*



Valores α / 1- α / Z más frecuentes

α	*0´05*	*0´01*	*0´001*
1 - α	*0´95*	*´99*	*0´999*
$Z_{\alpha/2}$	*- 1´96*	*- 2´58*	*- 3´29*
$Z_{1-\alpha/2}$	*1´96*	*2´58*	*3´29*
α → Nivel de significación			
1 – α → Nivel de confianza			
$Z_{\alpha/2}$ → Z (Tabla III negativa)			
$Z_{1-\alpha/2}$ → Z (Tabla IV positiva)			

CONTRASTE UNILATERAL IZQUIERDO	***CONTRASTE UNILATERAL DERECHO***
Zona Rechazo H_0 ***Zona Aceptación H_0 $(1 - \alpha)$*** |------------|--------------------------------------| ***α VALOR CRÍTICO***	***Zona Aceptación H_0 $(1 - \alpha)$*** ***Zona Rechazo H_0*** |---------------------------------|-------------| ***VALOR CRÍTICO α***
	

Nivel crítico (p) → *Calcular la probabilidad de obtener unos resultados como los observados en la muestra bajo el supuesto de que la H_0 es cierta.*

Zona de rechazo *(depende de que el contraste sea bilateral o unilateral). Teniendo en cuenta el riesgo adoptado se determinan los valores del estadístico y se decide si la hipótesis nula es verdadera o falsa. Nunca puede afirmarse categóricamente que la hipótesis es cierta o falsa, se afirma con un determinado nivel de probabilidad. Así, ¿Cómo decidir sobre la hipótesis nula?* → ***Dos métodos equivalentes****:*

- *A través del* ***nivel crítico*** → *si* ***p es ≤ α****, rechazaremos H_0, y si* ***p > α****, no la rechazaremos.*
- *A través del* ***valor o los valores críticos****. El criterio alfa divide la distribución de probabilidad del estadístico de contraste en dos zonas: la zona de aceptación de H_0 y la zona de rechazo de H_0 (región crítica). Si el valor muestral del estadístico de contraste cae en la zona de aceptación no lo rechazaremos, y sí lo haremos cuando cae en la zona de rechazo.*
- *A través del* ***intervalo de confianza*** *que delimita dos valores entre los cuales se encuentra o no el la hipótesis. Si está entre ellos se acepta la hipótesis nula, si no es así, se rechaza.*

Ho (Cierta) → ***(p)*** *es mayor que* ***(α)*** → *El valor se encuentra en el intervalo de confianza. El valor del estadístico está entre los valores críticos.*
Ho (Falsa) → ***(p)*** *es menor o igual que* ***(α)*** → *El valor está fuera del intervalo de confianza. El valor del estadístico supera o es inferior a los valores críticos.*

El ***nivel de significación (α)*** *se fija de antemano, mientras que el* ***nivel crítico p*** *es consecuencia del resultado obtenido al aplicar el estadístico de contraste.*
El contraste bilateral es más conservador que el contraste unilateral (es más difícil rechazar la H_0)

Conclusión e interpretación*: Contrastada la H_0, considerada de forma provisional como verdadera, y calculado el estadístico de contraste, se concluye rechazando o no la H_0. Después se interpreta el resultado en el contexto de la investigación.*

3.1.- RESUMEN DEL PROCEDIMIENTO (CONTRASTE DE HIPÓTESIS)
***1.-** Supuestos (descripción de las **características de la variable en la población**, y de los **datos de la muestra):** Nº de muestras con el que trabajamos; son independientes o relacionadas; seleccionadas de forma aleatoria o no; cómo son las escalas de medida utilizadas para las variables, la distribución de la que provienen los datos es conocida o desconocida, etc.*
***2.**- Formulación de las **Hipótesis Estadísticas** (exhaustivas y mutuamente excluyentes) de acuerdo con la hipótesis científica y las características anteriores.*
***3.**- Elección del **estadístico de contraste** apropiado.*
***4.- Fijar α (regla de decisión).** La zona de rechazo de la H_0 estará formada por todos los valores del estadístico cuya probabilidad de ser obtenidos es muy pequeña bajo el supuesto de que la H_0 es cierta.*
***5.- Cálculo del estadístico de contraste** con los datos obtenidos en la muestra.*
6.-** Decisión sobre el **rechazo o no de la H_0. ***Mediante el nivel crítico (p) ------------------ Mediante el ó los valores críticos (α)***
***7.- Conclusión e interpretación** de los resultados en el contexto de la investigación.*

PROBLEMA EJEMPLO → CONTRASTE DE HIPÓTESIS

Queremos contrastar la hipótesis que supone que la media poblacional en un test de atención de los estudiantes de la ESO es de 35 puntos. La distribución de la variable en la población es normal con varianza = 225. Extraemos una muestra aleatoria de 144 alumnos y obtenemos: una media = 32 puntos. Fijamos un nivel de significación α = 0,05.

***Supuestos →** La variable (atención) está medida a nivel de razón. Se distribuye normalmente en la población. Se trata de una muestra de observaciones aleatorias e independientes.*

Hipótesis Estadísticas → $H_0: \mu = 35$ $\quad$ $H_1: \mu \neq 35$ $\quad$ *(Contraste Bilateral)*

Estadístico de Contraste → $Z = (\overline{Y} - \mu) / (\sigma / \sqrt{n}$ *(varianza poblacional conocida)*

***Regla de decisión:** Nivel de Significación → (α = 0,05) / Zona de rechazo de la H_0 (Contraste bilateral) → Todos los valores iguales o menores que $Z_{\alpha/2}$ = (- 1,96) y todos los valores iguales o mayores que $Z_{1-\alpha/2}$ = (+ 1,96), según la distribución normal.*

***Cálculo del Estadístico de Contraste** (datos de la muestra) →* $Z = (32 - 35) / (15 / \sqrt{144} = (- 2,4)$

Toma de decisión sobre la H_0:

- ***Nivel crítico:** Como p = 2 · [Z ≥ l- 2,4 l] = 2 · (0,0082) = 0,0164 es menor que α = 0,05, rechazamos la hipótesis nula. Por tanto → 0,0164 cae en la zona de rechazo de la H_0.*
- ***Valores críticos:** Como (-2,4) < (-1,96) rechazamos hipótesis nula.*
- ***Intervalo de confianza**: 32 ± (1,96) · (1,25) = 29,55 < μ < 34,45. Como 35 no está en el intervalo de confianza se rechaza la H_0.*

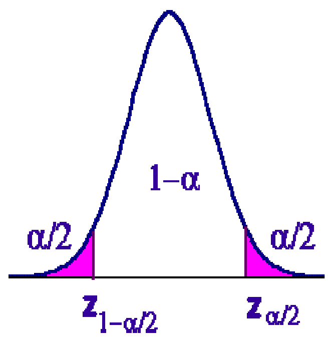
***Interpretación de resultados →** Para un nivel de significación α = 0,05, podemos rechazar la hipótesis nula (la media en atención de los estudiantes de la ESO no es de 35 puntos).*

ERRORES EN LA TOMA DE DECISIONES

Los contrastes de hipótesis están basados en estadísticos ***(medidas de discrepancia)*** *y tienen una distribución de probabilidad conocida; así, todas las decisiones llevan aparejadas una probabilidad de ocurrencia. Al tomar una decisión sobre una H_0 pueden darse cuatro situaciones (dos de ellas son decisiones acertadas y dos erróneas). Siempre debemos proponernos que* ***(1 - α) y (1 - β)*** *sean altos y que* ***α y β*** *sean bajos.*

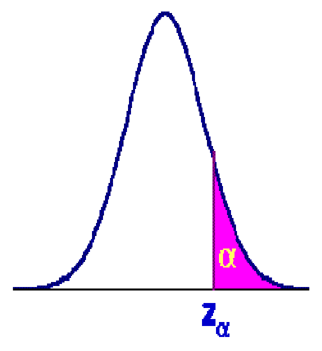
Decisión	***Ho (cierta)***	***H_1 (cierta)***
Acepto H_1	***ERROR TIPO I Probabilidad → α***	***DECISIÓN CORRECTA Probabilidad → 1 – β***
Acepto Ho	***DECISIÓN CORRECTA Probabilidad → 1 – α***	***ERROR TIPO II Probabilidad → β***

Siempre se cumple → ***$\alpha + (1 - \alpha) = 1$ y $\beta + (1 - \beta) = 1$***

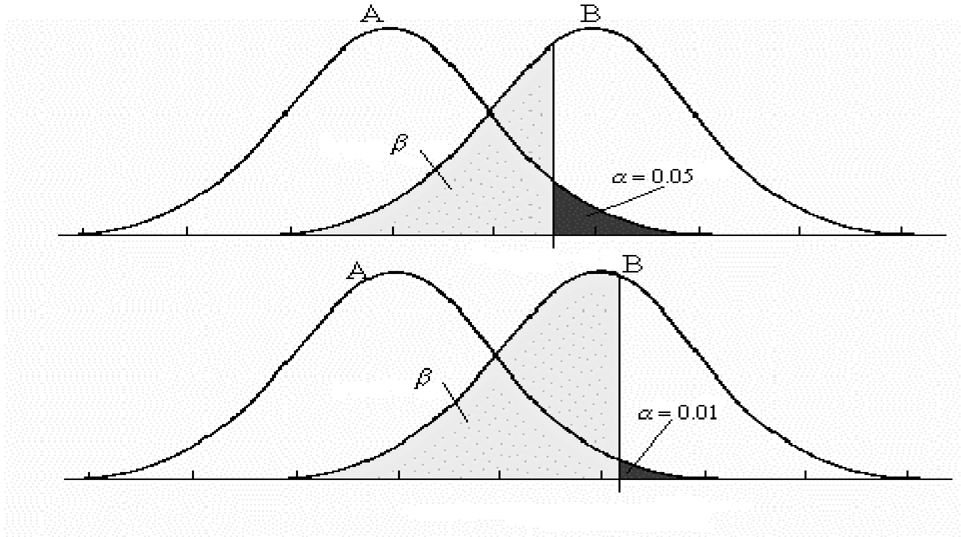


Error tipo I* (α)** *→ Bajo la distribución de H_0 representa la probabilidad de rechazar una H_0 verdadera = la probabilidad de aceptar una H_1 falsa / (*α = probabilidad de cometer error de tipo I*** *= nivel de significación)*

***Nivel de confianza* (1–α)** *→ Representa la probabilidad de aceptar una H_0 verdadera = la probabilidad de rechazar una H_1 falsa.*



Error tipo II* (β)** *→ Bajo la distribución de H_1 cuando aceptamos erróneamente la H_0 siendo falsa (*β =*** *probabilidad de cometer error de tipo II)*
Potencia de contraste* (1–β)** *→ Representa la probabilidad de rechazar la H_0 siendo falsa = la probabilidad de aceptar una H_1 verdadera (probabilidad de obtener un resultado estadísticamente significativo)* ***1 – β = *Probabilidad de no cometer Error tipo II.*



Tamaño del efecto: *Expresa la magnitud de la diferencia observada entre la H_0 (el valor teórico) y la H_1 (el valor observado)*