TEMA Nº 10 ESTIMACIÓN DE PARÁMETROS Y CÁLCULO DEL TAMAÑO MUESTRAL

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

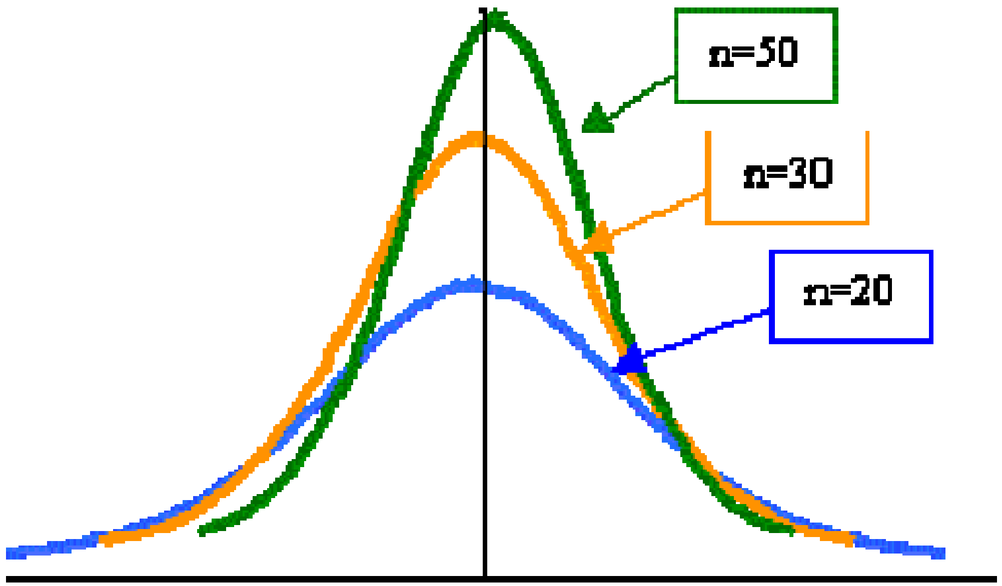
- ***Distribución Poblacional:*** *(Distribución de frecuencias que presenta la variable en la población sobre la que se quiere generalizar). Las medidas en la población se denominan parámetros poblacionales y se designan con letras griegas (Parámetros: Media de la población = µ , Varianza de la población =* σ***2,*** *Proporción de la población =* ***π****)*
- ***Distribución de la Muestra:*** *(Distribución de frecuencias que presenta la variable en la 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 =* ***S2,*** *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. Ejemplo pag 5 libro.

- Consideremos una población formada por todos los estudiantes universitarios de una comunidad de los que podemos conocer, a partir de sus datos de la matrícula, su edad. A partir de esos datos podemos calcular su edad media y varianza de esa misma variable edad, valores que representamos con **μ** y *σ2*, respectivamente.
- De esta población podemos extraer una muestra de por ejemplo 100 estudiantes y calcular su media y desviación típica $\dot{X}$ y S_x
- Se pueden extraer muchas otras muestras distintas del mismo tamaño y en cada una de ellas calcular su media y desviación típica, que puede variar de una muestra a otra, de tal manera que con las puntuaciones de todas las medias obtenidas en las distintas muestras, se origina otra que se llama distribución muestral de la media.
- Igualmente se puede obtener la distribución muestral de la desviación típica o de cualquier otro estadístico.

DISTRIBUCIÓN MUESTRAL DE LA MEDIA

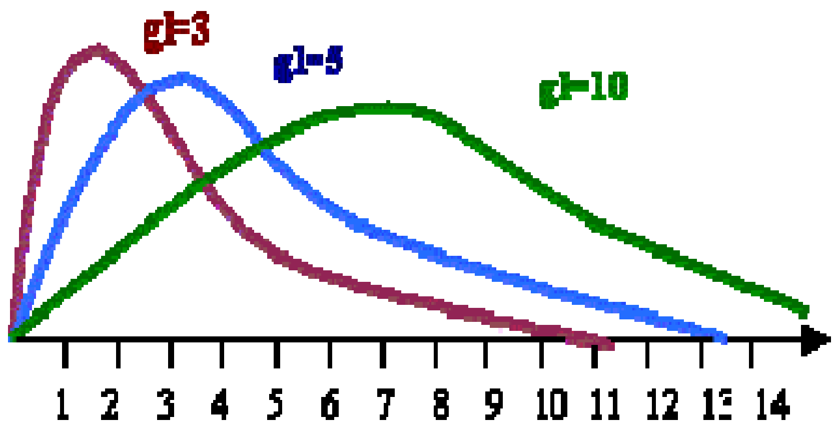
- ***Teorema Central del límite:*** *Si una población tiene una media μ y una varianza σ^2 finitas, la distribución de las medias muestrales de tamaño "n" extraídas de manera aleatoria e independiente, se aproxima a la forma de una* ***distribución normal con desviación típica σ/√n y media μ*** *conforme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

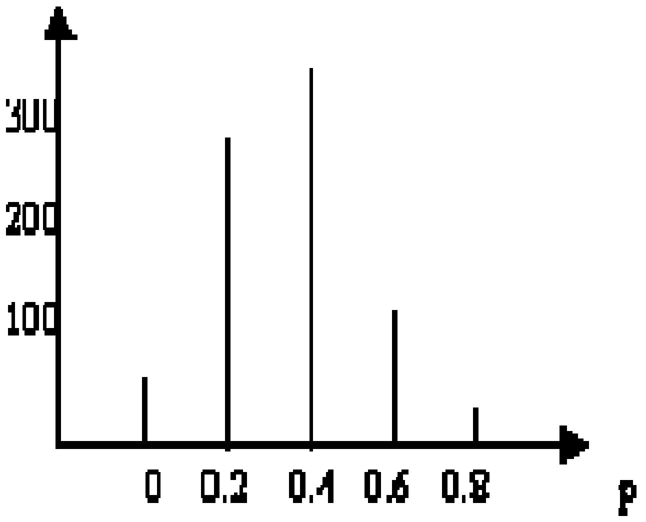
La distribución muestral de la media se ajusta a la distribución ***t de Student*** *(con n-1 grados de libertad), si ignoramos la forma de la distribución de la variable y/o el tamaño de la muestra es pequeño.*

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 (π, σ_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

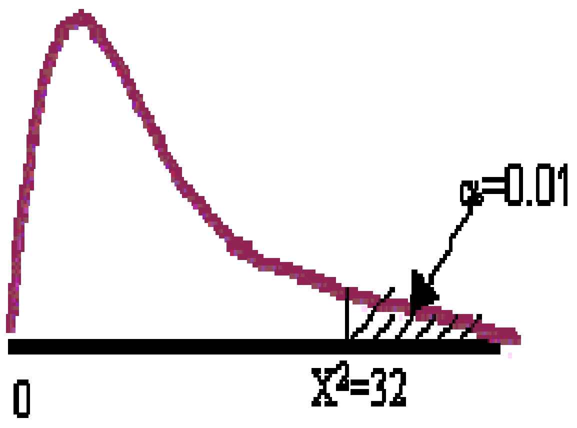
El CI de los alumnos de un centro de E. Especial se distribuye normalmente con μ = 80 y σ =10. Si de esta población extraemos una muestra aleatoria de 25 alumnos. ¿Cuál es la probabilidad de obtener una media mayor de 75 puntos?

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

PROBLEMAS EJEMPLO

La varianza sería 1,882

*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 **σ** =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 $S^2_{n-1}=2$

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

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 $\rightarrow P(S^2_{n-1} > 2)$

PROBLEMAS EJEMPLO

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?

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

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)*

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* ***θ,*** *si* ***E (U) = θ*** *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 θ = 1 / σ2*** *(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.*

cuatro propiedades básicas

- ***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 medida 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

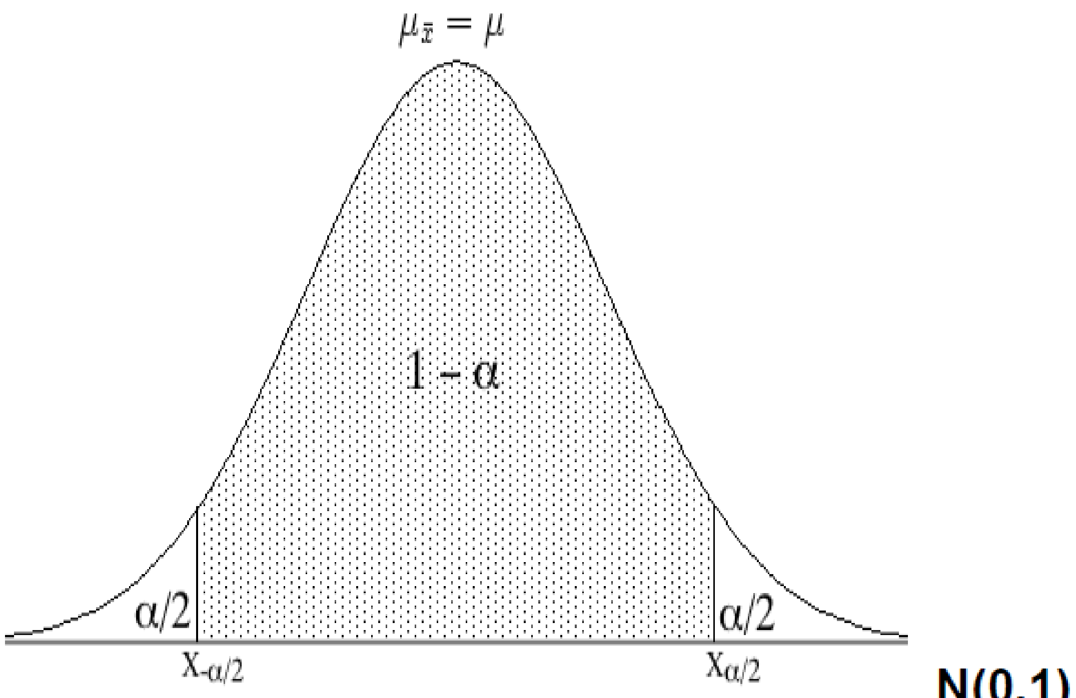
	Carencia de sesgo	Eficiencia	Suficiencia	Consistencia
Media Aritmética	SI	> Mediana	SI	SI
Mediana	NO	< Media	NO	SI
Proporción	SI	-----	SI	SI
Varianza	NO	$> S^2_{n-1}$	SI	SI
Cuasi varianza	SI	$< S^2$	SI	SI

ESTIMACIÓN POR INTERVALOS

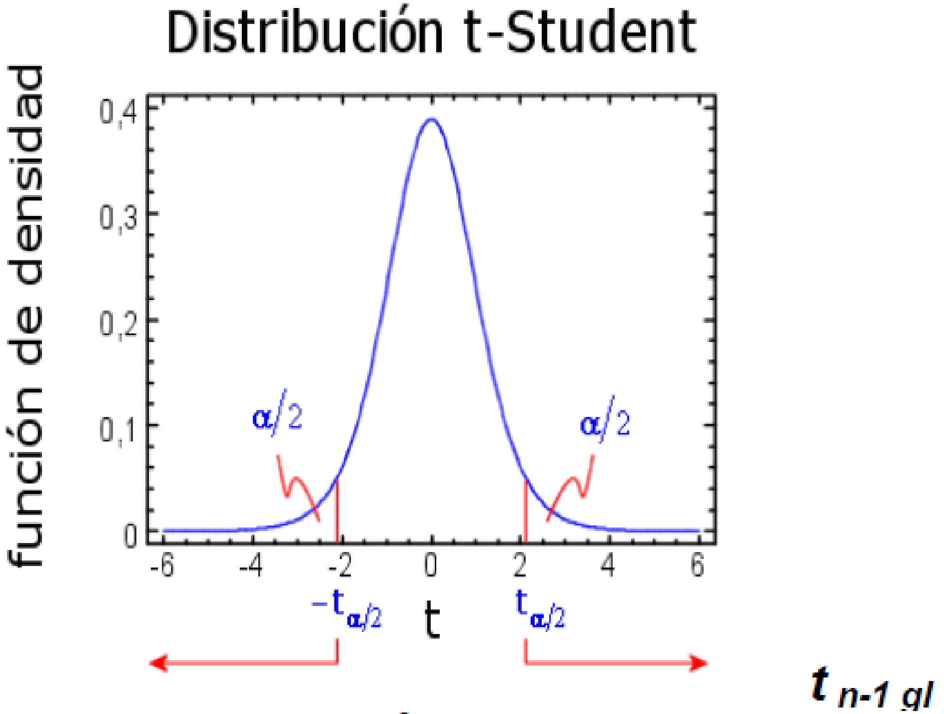
- *Una estimación por intervalos expresa el grado de confianza con el que se espera que esté el valor del parámetro dentro del intervalo, por lo que se suele llamar intervalo de confianza.*
- *La amplitud del intervalo nos indicará su precisión. **A menor amplitud, más precisión**, más informativo es, 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



Distribución Normal (Conocida σ^2)



Desconocida la σ^2 (Muestras pequeñas)

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

Error en el problema, es el valor absoluto de α½

Problemas ejemplo: *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 → **$8 \pm (2{,}58 \cdot 0{,}20) = [7{,}484$ 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 *X = 8 Distribución normal de la VD α = 0,05 Sn- 1 = 2*

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

Intervalo de confianza X ± (t α · Sn- 1 / √ n) = Límite superior y Límite inferior

α / 2 = 0,05/2 = 0,025

1 - (α / 2) = 0,975

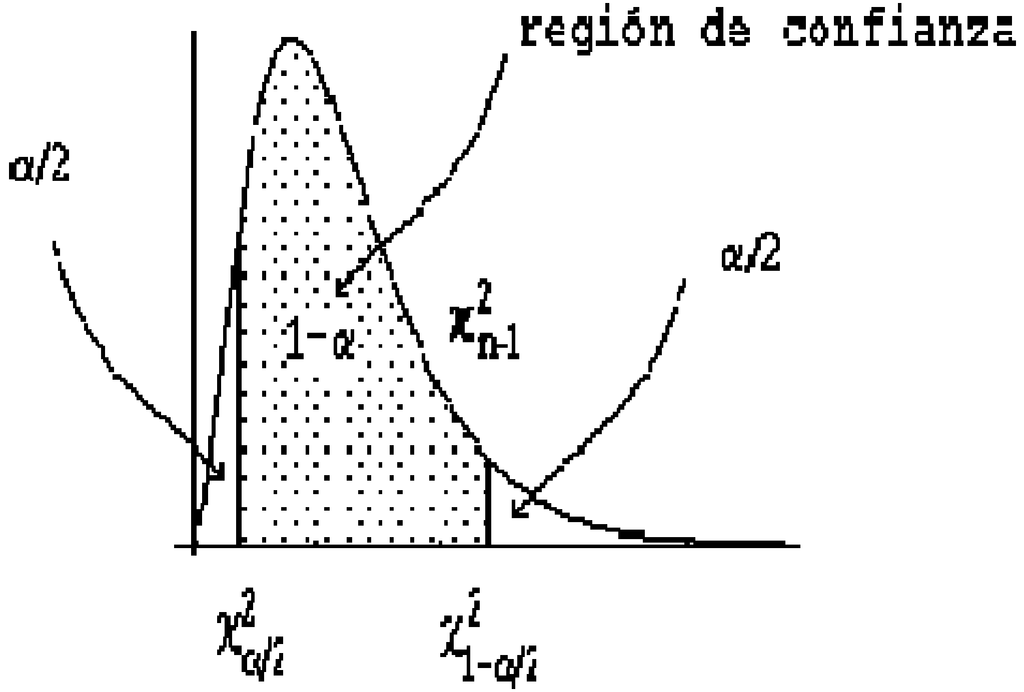
|t α / 2 | = 2,06 (Tablas t de Student)

Error Típico Sn- 1 / √ n = 2 / 5 = 0,4

Error máximo de estimación |t α / 2 | · Sn- 1 / √ n = (2,06 · 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)

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

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

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

INTERVALO DE CONFIANZA PARA LA VARIANZA

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 $\alpha = 0{,}05$.*

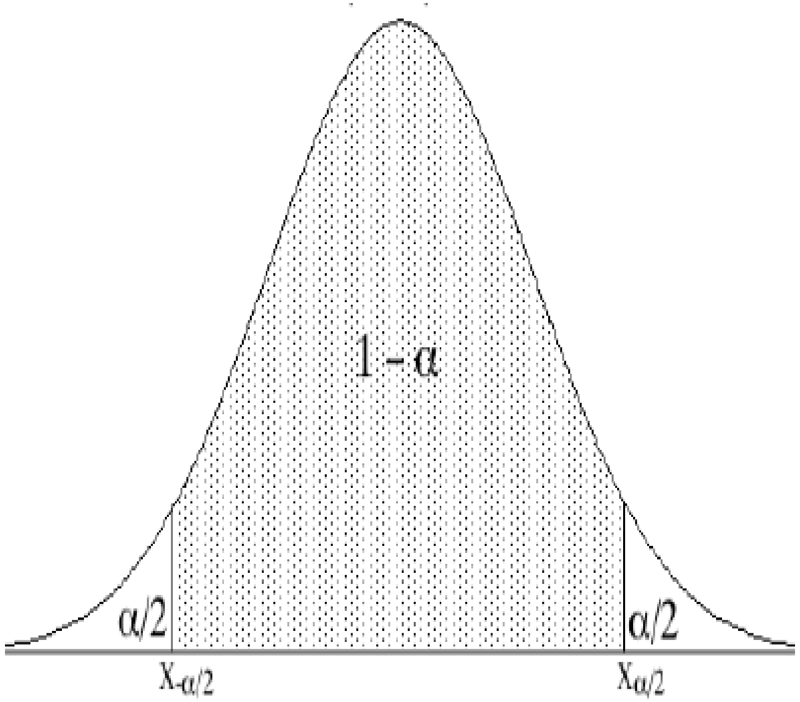
Estimamos la varianza de la población a partir de la varianza de la muestra (estimación por intervalo):

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

Según las tablas χ^2_{n-1} (χ^2_{19}) → para ($\alpha / 2 = 0{,}025$) → 8,91 y para ($1 - \alpha / 2 = 0{,}975$) → 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 : $S_P = \sqrt{P(1 - P) / n}$

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

$Z_{\alpha/2} \cdot S_p$ = *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* $(\alpha / 2 = 0{,}04) \rightarrow Z = (-1{,}75)$ *y Para* $(1 - \alpha / 2 = 0{,}96) \rightarrow Z = (+1{,}75)$

Proporción muestral $\rightarrow (50 / 64) = 0{,}781$

Intervalo de confianza $\rightarrow P \pm |Z_{\alpha/2}| \cdot (S_P)$ *= 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_{max} = 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)*

Cálculo de la n

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 $(\alpha / 2 = 0{,}01) \rightarrow Z_{Tablas} = \pm\ 2{,}33$ *Varianza poblacional conocida* $\sigma^2 = 64$

$n = (\sigma^2 \cdot Z^2_{\alpha/2}) / E^2$ *máximo de estimación* $\rightarrow n = (64 \cdot 2{,}33^2) / 1{,}5^2 = 154{,}\ 42 \approx$ ***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 $(\alpha / 2 = 0{,}025) \rightarrow Z_{Tablas} = \pm\ 1{,}96$

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

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

Distributing prohibited | Downloaded by Adolfo Navascués Navascues (adolfonavascues@gmail.com)

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.*
- ***H0*** *(hipótesis nula): se acepta provisionalmente como verdadera y se somete a contraste.*
- ***H1*** *(hipótesis alternativa): se acepta al rechazar la hipótesis nula*.