TEMA 6
Teoría de la estimación

Estimadores. Condiciones de un buen estimador.

Teorema central del límite.

Estimaciones para medias aritméticas y variancias.

Estimación por Intervalo de Confianza.

Bibliografía

Estadística para administración y economía

Anderson, D.R., Sweeney,D. J. y Williams, T. A 2002 10a. ed. ISBN-13: 978-607-481-319-7 ISBN-10: 607-481-319-1

Bioestadística. Métodos y Aplicaciones

2002.Universidad de Málaga. España

INFERENCIA ESTADÍSTICA

La **inferencia** es un campo de la estadística que persigue la **obtención de conclusiones** de una población a partir de una muestra de la misma.
Se estimarán los Parámetros poblacionales (ej: μ; σ; σ^2, π) con estadísticos muestrales ($\bar{x}, S, S^2$, p).

Ejemplo

Se desea concluir algo respecto a la proporción de consumidores de café que prefieren cierta marca de café (Argentina). Se selecciona una muestra aleatoria grande y se calcula la proporción ***p*** de personas que prefieren la marca de café en cuestión (en la muestra). El valor ***p*** se utiliza ahora para hacer una inferencia respecto a la proporción verdadera π.

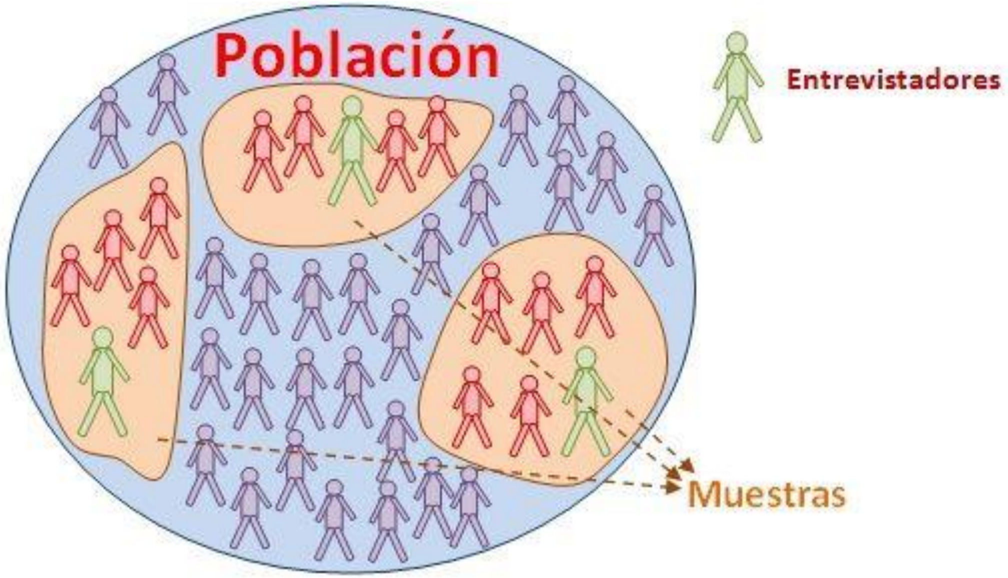
Población: Todos los consumidores de café de Argentina

Muestra: Un subconjunto de consumidores de café seleccionado al azar de la población mencionada.

p se calcula como un cociente entre la cantidad de personas que toman esa marca sobre la cantidad de consumidores de café.

Seguimos con el Ejemplo

p: es una función de los valores observados en la muestra aleatoria ; ya que es posible tomar muchas muestras aleatorias de la misma población, ***p*** variará de una a otra muestra. Si tomáramos muchas muestras de la misma población y de cada muestra calculamos un valor de ***p*** **,** entonces ***p*** es un una variable aleatoria y por lo tanto sus valores tiene una distribución.
Tal variable aleatoria se llama **estadístico o estimador**.
Cualquier función de las variables aleatorias que forman una muestra aleatoria se llama estadístico.



Estadísticos

- Promedio o Media
- Varianza
- Proporciones

Ramas de la Inferencia Estadística

Estimación de parámetros

Prueba de hipótesis

Existen **dos tipos** de estimaciones:

- Estimación puntual
- Estimación por intervalos de confianza

Tema 9

Estimación puntual

Sea x1, x2, . . .,xn una muestra aleatoria desde la distribución f(x;θ), la función $\hat{\theta}$(x1, x2, . . .,xn) es una estimación de θ .
$\hat{\theta}$ es un estimador puntual del parámetro θ .

Así, por ejemplo, la media muestral $\bar{X}$= 1/n($\sum X_i$) es una función de ***n*** variables aleatorias donde ***n*** es el tamaño de la muestra.

Propiedades clásicas de los buenos estimadores

- Insesgamiento
- Consistencia
- Eficiencia

Propiedades clásicas de los buenos estimadores

Insesgamiento

Un estimador $\hat{\theta}$ es un estimador **insesgado** para el parámetro θ si, para cualquier tamaño muestral, su valor esperado o esperanza matemática es igual al parámetro que estima.

Dicho de otra forma, si en promedio $\hat{\theta}$= θ (la distribución de $\hat{\theta}$está centrada en θ o no existe una tendencia persistente a subestimar o sobreestimar θ , diremos que $\hat{\theta}$ es un estimador insesgado.

Consistencia

Sea $\hat{\theta}$ un estimador del parámetro θ , si la P(| $\hat{\theta}$ -θ | > ε) tiende a 0, para ∀ ε > 0, cuando el tamaño de la muestra tiende a ∞, se dice que $\hat{\theta}$ es un estimador **consistente** del parámetro θ .

En otras palabras, esto significa que a medida que aumenta el tamaño de muestra aumenta la proximidad de $\hat{\theta}$ respecto θ . Un ejemplo clásico de estimador consistente es la media muestral $\bar{X}$.

Propiedades clásicas de los buenos estimadores

Eficiencia

Sea $\hat{\theta}$ un estimador insesgado para el parámetro θ, se dice que $\widehat{\theta}$ es eficiente si tiene la mínima varianza posible.

Por ejemplo, si X es una variable aleatoria con distribución normal, entonces la media y la mediana muestral son estimadores insesgados del parámetro μ y además ambos estimadores son consistentes.
Sin embargo, se puede comprobar que la varianza de la media muestral es menor que la varianza de la mediana, por lo tanto la media es más eficiente que la mediana.
Más aún, la media es el **estimador eficiente** en el sentido de que no existe ningún otro, que tenga menor varianza.

TEOREMA CENTRAL DEL LÍMITE

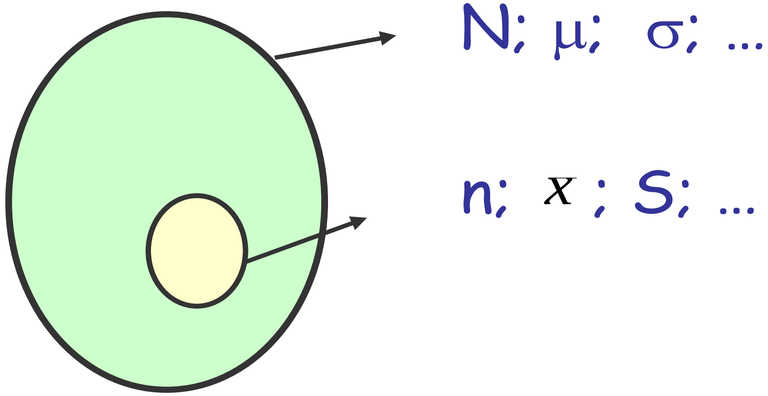
El teorema permite afirmar que si X es una variable que se distribuye normalmente con media μ y varianza σ^2 (X se distribuye $N(\mu , \sigma^2)$), entonces
entonces la distribución de la media muestral $\bar{X}$, obtenida de una muestra de tamaño n de la poblacional, también es normal con la misma Media μ y con varianza igual a σ^2/n .

Si X se distribuye $N(\mu , \sigma^2)$ entonces $\bar{X}$ se distribuye $N(\mu , \sigma^2 /n)$

Además, cuando el tamaño de la muestra es suficientemente grande, digamos mayores a n=30, se espera que la distribución del estadístico, "media muestral" $\bar{X}$ se aproxime a la distribución normal sin importar la la distribución de los valores de X.

Si $n>30$ entonces $\bar{X}$ se distribuye $N(\mu , \sigma^2 /n)$, siempre.

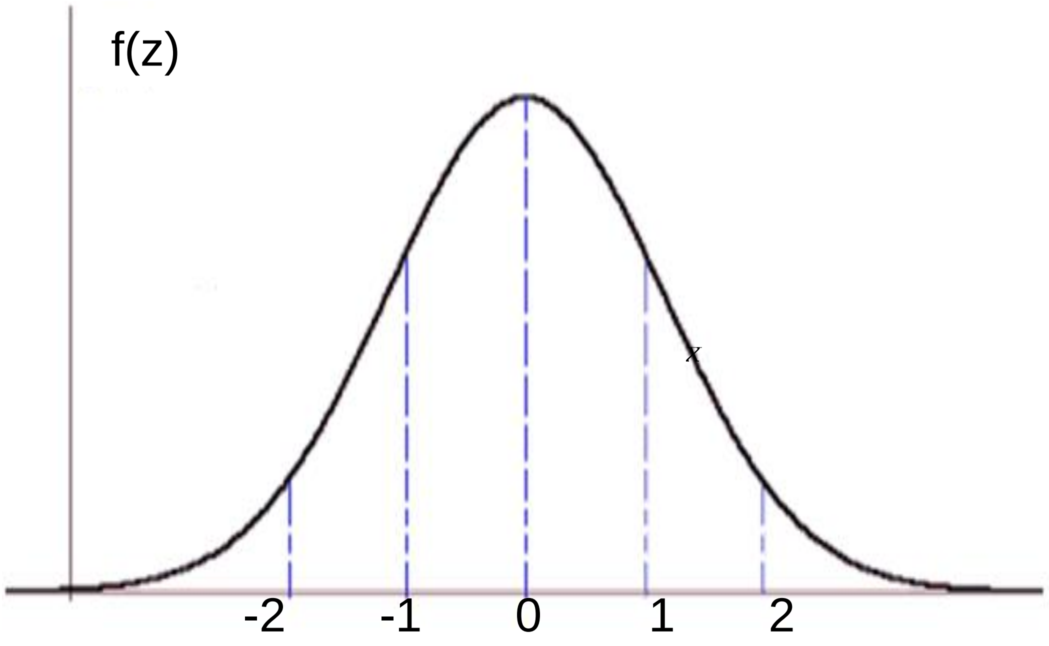
TEOREMA CENTRAL DEL LÍMITE



Si $\bar{X}$ se distribuye N(μ , σ^2 /n) se puede calcular
Z = ($\bar{X}$ -μ)/(σ /vn)

$$\frac{x-\mu}{\sigma/\sqrt{n}} \to Z(N(0,1))$$

$$n \to \infty$$



Por el Teorema Central del Límite sabemos entonces que $\bar{X}$ siempre se distribuye normal con parámetros N(μ , σ^2 /n)

ESTIMACIÓN POR INTERVALOS

Se refiere a la estimación de un parámetro mediante un intervalo aleatorio, llamado *intervalo de confianza,* cuyos puntos finales $L_1 < L_2$, son funciones de las variables aleatorias observadas de modo que la probabilidad de la desigualdad $P(L_1 \leq \theta \leq L_2)$ se satisfaga en términos de un número determinado $1 - \alpha$

Traduciendo con un ejemplo la definición anterior…

Si se quiere reportar el rendimiento de un producto, en vez de decir que la media estimada del rendimiento es 25 kg se podría decir que, el rendimiento promedio para ese producto está comprendido entre 23.5 y 26.5 kg, con una confianza del 95%,
El procedimiento que permite calcular los límites inferior y superior del intervalo antedicho se conoce como **Estimación por Intervalo** y el intervalo obtenido como **Intervalo de Confianza**.

LÍMITES DE CONFIANZA PARA μ

Para demostrar cómo se obtienen los límites de confianza en primer debemos determinar la confianza deseada, la cual se puede expresar como porcentaje o proporción.

Luego, basándonos en el teorema central del límite, sabemos que $\bar{X}$ tiene distribución normal obtendremos valores de probabilidad de la Normal y si tenemos datos de muestras de tamaño n<30 debemos trabajar con T-Student con parámetro=n-1.

La confianza se expresa como lo opuesto a error permitido el cual se simboliza como α , entonces la confianza es (1- α); esto es, si el error es 0.05, la confianza es (1-0.05) = 0.95.

En la próxima hoja veremos el caso donde la confianza deseada es 0.95.

Se parte del conocimiento de los valores de Z que encierran el 0.95 de probabilidad, dichos valores son -1.96 y 1.96. Luego, se reemplaza Z por su fórmula, se despeja μ , con la particularidad que es una desigualdad.

LÍMITES DE CONFIANZA PARA μ

$$P(-1{,}96 \leq Z \leq 1{,}96) = 0{,}95$$

$$P\left(-1{,}96 \leq \frac{\overline{x}-\mu}{Sx/\sqrt{n}} \leq 1{,}96\right) = 0{,}95$$

$$P(-1{,}96^{*}\,Sx/\sqrt{n} \leq x-\mu \leq 1{,}96^{*}\,Sx/\sqrt{n}) = 0{,}95$$

$$P(-x-1{,}96^{*}\,Sx/\sqrt{n} \leq -\mu \leq -x+1{,}96^{*}\,Sx/\sqrt{n}) = 0{,}95$$

Luego como queremos encontrar que entre que valores se encuentra μ , no $-\mu$ multiplicamos por (-1)

$$P(x+1{,}96^{*}\,Sx/\sqrt{n} \geq \mu \geq x-1{,}96^{*}\,Sx/\sqrt{n}) = 0{,}95$$

Reordenamos

$$P(\underbrace{x-1{,}96^{*}\,Sx/}_{Li}\sqrt{n} \leq \mu \underbrace{\leq x+1{,}96^{*}}_{Ls}\,Sx/\sqrt{n}) = 0{,}95$$

Li:Límite Inferior

Li:Límite Superior

Ejemplo

Con una muestra de n = 15; $\bar{x}$ = 18,2 y S^2_x = 40,4 determinar los límites de confianza del 95% para el promedio de la población.

Cuando el tamaño de la población es n<30, entonces los valores de probabilidad debemos obtener de la T de Student con 14 grados de libertad.

En una distribución T de Student con 14 (15-1) grados de libertad, los valores entre los cuales se encuentra el 95% de probabilidad no son -1.96 y 1.96, sino son -2.14 y 2.14.

Para μ:

$$P(-t_{5\%} \leq \frac{x-\mu}{Sx} \leq t_{5\%}) = 0{,}95 \qquad ; \; S\overline{x} = Sx/\sqrt{n}$$

$$P(\overline{x} - t_{5\%}.Sx/\sqrt{n} \leq \mu \leq \overline{x} + t_{5\%}.Sx/\sqrt{n}) = 0{,}95$$

$$P\left(18{,}2 - 2{,}14.\frac{6{,}36}{3{,}87} \leq \mu \leq 18{,}2 + 2{,}14.\frac{6{,}36}{3{,}87}\right) = 0{,}95$$

$$P(18{,}2 - 2{,}14.1{,}64 \leq \mu \leq 18{,}2 + 2{,}14.1{,}64) = 0{,}95$$

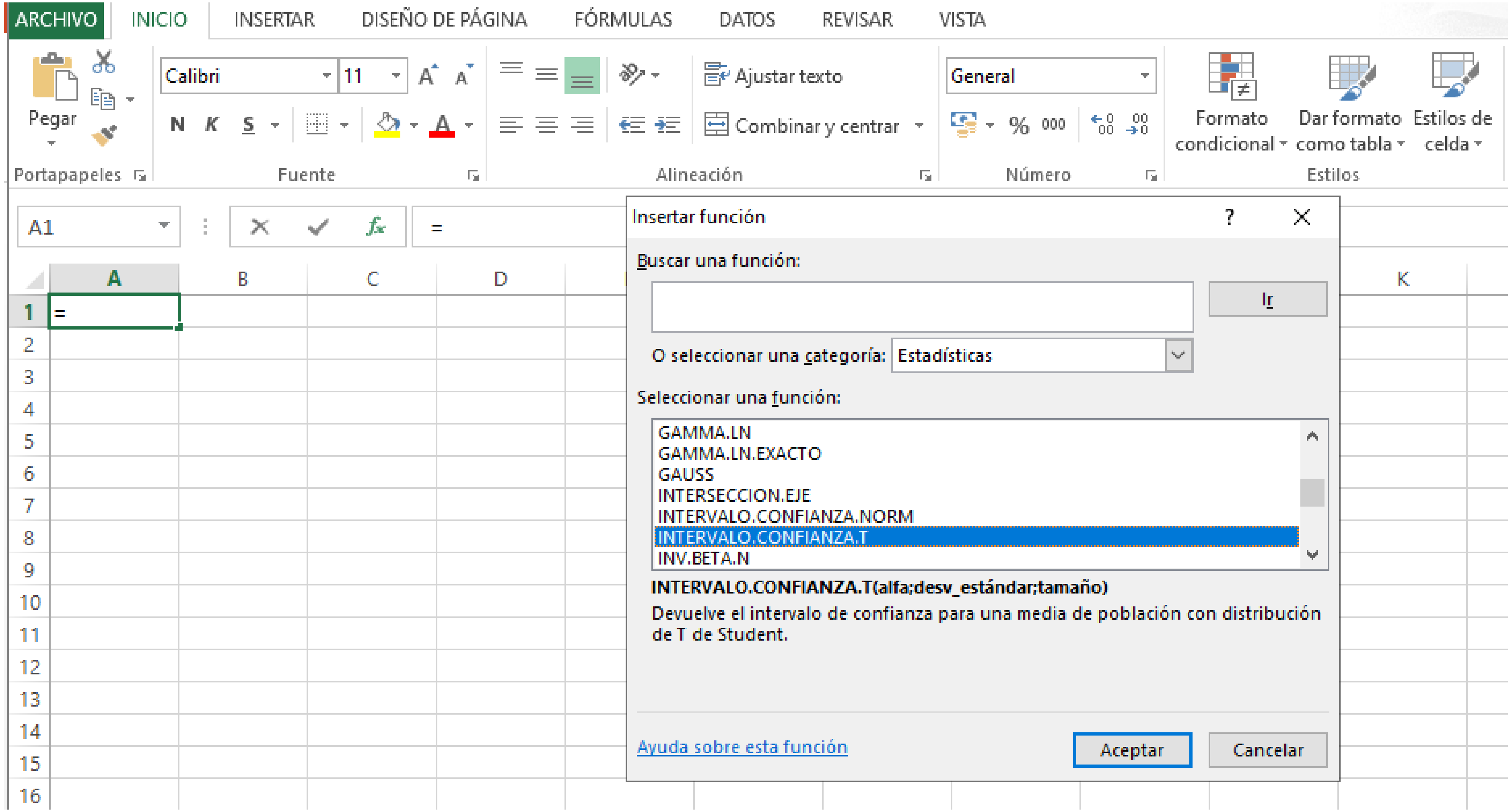
$$P(18{,}2 - 3{,}51 \leq \mu \leq 18{,}2 + 3{,}51) = 0{,}95$$

$$P(14{,}69 \leq \mu \leq 21{,}71) = 0{,}95$$

Los límites entre los cuales se encuentra μ son (14,69 y 21,71) y esa afirmación se realiza con una confianza de 0.95.

Con Excel

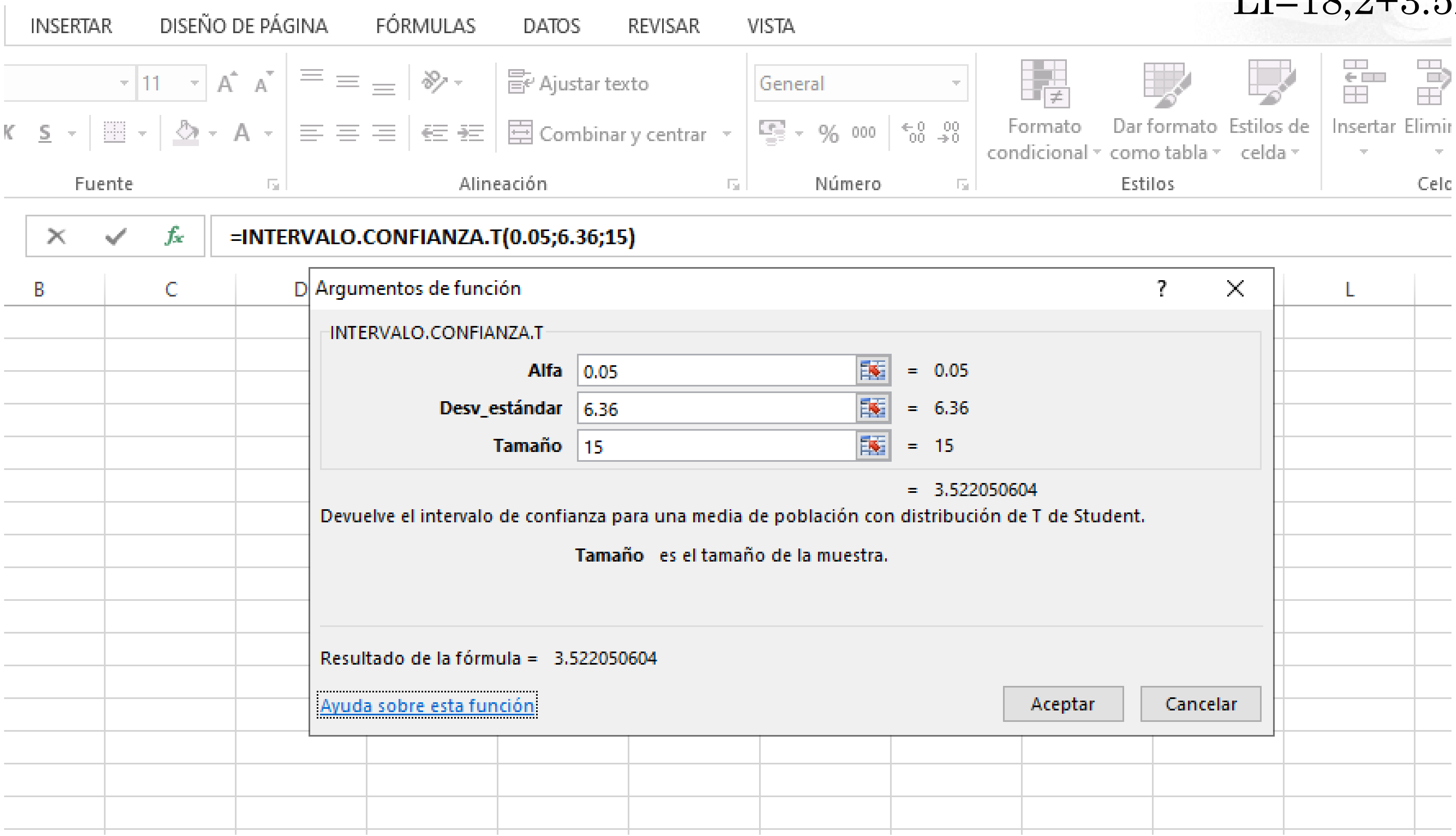
Excel tiene una función denominada INTERVALO.CONFIANZA.T y otra función INTERVALO.CONFIANZA.Z , esta función arroja que valores debe restarse y sumarse $\bar{x}$ para obtener los límites.



CON EXCEL

Debemos calcular $S_{\bar{X}}$, que es √(40,4/15)= 6.36

Excel arroja como resultado 3.52 luego el
LI=18,2-3.52=14.68
LI=18,2+3.52=21.72



LÍMITES DE CONFIANZA PARA σ^2

$$P\left(\left|S^2-\sigma^2\right|<k\sqrt{\frac{2\sigma}{(n-1)}}\right)\cong 1-\alpha$$

Solucionando la desigualdad dentro del paréntesis

(k es un multiplicador que se obtiene de X2)

$$\frac{(n-1)S^2}{X_1^{\ 2}}<\sigma^2<\frac{(n-1)S^2}{X_2^{\ 2}}$$

Que nos permite obtener los límites de confianza

para σ^2 que corresponden al coeficiente de confianza $1-\alpha$

LÍMITES DE CONFIANZA PARA σ^2

Si la muestra es pequeña, la distribución muestral de la varianza muestral S^2 es aproximadamente Chi-cuadrado y el intervalo de confianza se calcula para σ^2 mediante la expresión anterior, el multiplicador de confianza surgirá de la distribución Chi-Cuadrado:

Ejemplo 1

Dada una muestra de n =15 con $\bar{x} = 24$ y $S_x = 9{,}5$. Calcular los límites de confianza del 95% σ^2. Los denominadores surgen de la distribución X^2 con 14 grados de libertad que encienrran el 0.95 de confianza.

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

Siendo n-1 =14 , S^2 =9,5^2=90.25, y los denominadores surgen de la distribución X^2 con 14 grados de libertad son (5.6287 y 26.12) , esos valores se reemplazan asi, en el denominador de LI va 26.12 y en el LS=5.6287

La Varianza está entre 48.32 y 224.17, con una confianza del 95%.

EJEMPLO 2

Con una muestra de n = 15; $\bar{x}$ = 18,2 y S^2_x = 40,4 determinar los límites de confianza del 95% para la varianza de la población de la que proviene.

Para σ^2:

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

$$P\left(\frac{14*40,4}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{14*40,4}{\chi^2_{1-\frac{\alpha}{2}}}\right) = 0,95$$

$$P\left(\frac{565,6}{\chi^2_{(14;0,025)}} \leq \sigma^2 \leq \frac{565,6}{\chi^2_{(14;0,975)}}\right) = 0,95$$

$$P\left(\frac{565,6}{26,1189} \leq \sigma^2 \leq \frac{565,6}{5,6287}\right) = 0,95$$

ESTIMACIÓN POR INTERVALOS

Donde $SE(\hat{\theta})$ equivale al error estándar del estimador $\hat{\theta}$

$1-\alpha$ se llama coeficiente de confianza y

k es el multiplicador de confianza correspondiente

La desigualdad anterior se puede escribir como :

$$P(\left|\hat{\theta}-\theta\right| < kSE(\hat{\theta})) = 1-\alpha$$

$$\left|\theta-\hat{\theta}\right| < kSE(\hat{\theta})$$

$$-kSE(\hat{\theta}) < \theta-\hat{\theta} < +kSE(\hat{\theta})$$

$$\hat{\theta}-kSE(\hat{\theta}) < \theta < \hat{\theta}+kSE(\hat{\theta})$$

$$P\left[\hat{\theta}-kSE(\hat{\theta}) < \theta < \hat{\theta}+kSE(\hat{\theta})\right] = 1-\alpha$$

$\hat{\theta}-kSE(\hat{\theta}) = L_1$ Límite inferior de confianza

$\hat{\theta}+kSE(\hat{\theta}) = L_2$ Límite superior de confianza

Es la forma genérica de representar los lìmites de confianza de un estimador $\hat{\theta}$.