Conferencia 2. Estimación de Parámetros. Estimación Puntual y por Intervalos.

Tema 2. Inferencia Estadística.

Sumario:

- Inferencia Estadística.
- Estimación Puntual
- Propiedades de los Estimadores.
- Máxima Verosimilitud.
- Estimadores de la Media, Varianza y Proporción.
- Estimación por intervalos de confianza para la media, proporción y varianza.
- Anexo. Funciones de Distribución Muestrales.

Bibliografía:

- Introductory Statistics Prem S. Mann, Christopher Jay Lacke, 2010 (Capítulo 8)
- Introductory Statistics. Barbara Illowsky, Susan Dean 2018 (Capítulos 8).

Definiciones Fundamentales.

Estadística Inferencial: Son aquellas técnicas que permiten la toma de decisiones mediante las conclusiones a que se arriben cuando se analizan características numéricas. Dentro de las técnicas de la estadística inferencial están: Estimación de parámetros y Prueba de Hipótesis.

La mayoría de las investigaciones estadísticas se proponen llegar a generalizaciones a partir de la información contenida en muestras aleatorias acerca de la población de donde fueron obtenidos. En particular, estaremos interesados frecuentemente en el problema de hacer inferencias sobre los parámetros de las poblaciones, como la media o la desviación estándar. Para efectuar tales inferencias utilizaremos estadígrafos, como $\bar{X}$ y S^2, es decir, cantidades calculadas con base en las observaciones de la muestra.

Definición #1: Espacio Muestral. Conjunto de todas las muestras de tamaño fijo tomadas de una población.
Por ejemplo, si se tiene una población normal y se toman muestras de tamaño n puede afirmarse que el espacio muestral será $\mathbb{R}^n$, o sea el producto cartesiano de los números reales n veces ya que la población normal puede tomar cualquier valor en los $\mathbb{R}$. Si las muestras se toman de un experimento Bernoulli el espacio muestral seria $\{0,1\}^n$. Los espacios muestrales se denota por $\mathfrak{X}$.

Definición #2: Espacio Paramétrico. Es el conjunto de todos los valores posibles de él o los parámetros.

Por ejemplo, en una distribución normal el parámetro $\mu \in \mathbb{R}$ y $\sigma^2 \in \mathbb{R}_+^* = \mathbb{R}_+ \backslash \{0\}$, por tanto, los espacios paramétricos para cada uno serian $\mathbb{R}$ y $\mathbb{R}_+^*$ respectivamente. En conjunto para el parámetro (μ, σ^2) será $\mathbb{R}\ \mathbb{R}_+^*$. Los espacios paramétricos se denotan por la letra Θ.

Estimación puntual.

Definición #3 Estimador puntual: Un estimador del parámetro θ de una familia de distribuciones $\{P_\theta; \theta \in \Theta\}$ es una función $\hat{\theta}$ del espacio muestral en el espacio paramétrico tal que:

$$\hat{\theta}: \mathfrak{X} \to \Theta$$

$$(x_1, \dots, x_n) \to \hat{\theta}(x_1, \dots, x_n)$$

Una definición más informal y quizás más intuitiva seria:

Sea X una variable aleatoria y θ un parámetro desconocido de su distribución de probabilidad. Un estimador puntual de θ es un estadígrafo que se utiliza en sustitución de dicho parámetro. Llamémosle al estimador $\hat{\theta} = d(X_1, X_2, \dots, X_n)$,

Definición #4 Estadígrafo: Un Estadígrafo es una función de los resultados de la muestra que no depende de valores desconocidos.

Definición #5 Error cuadrático medio: Es la medición de la precisión del estimador y esta medida por:

$$ECM[d(X_1,X_2,\dots,X_n)] = E\left[\left(d(X_1,X_2,\dots,X_n)\right)^2\right] - E[d(X_1,X_2,\dots,X_n)]^2 + [E[d(X_1,X_2,\dots,X_n)] - \theta]^2$$

O

$$ECM[\hat{\theta}] = E[\hat{\theta}^2] - E[\hat{\theta}]^2 + [E[\hat{\theta}] - \theta]^2 = V(\hat{\theta}) + b(\hat{\theta})^2$$

De los términos:

1. $\hat{\theta} = d(X_1, X_2, \dots, X_n)$ estimador de θ evaluado en la muestra
2. $[E[\hat{\theta}^2] - E[\hat{\theta}]]^2 = V[\hat{\theta}]$
3. $[E[\hat{\theta}] - \theta]^2$ es el cuadrado del sesgo. El sesgo o *bias* de un estimador es la diferencia entre el valor esperado del estimador y el valor real del parámetro a ser estimado. Un estimador con sesgo cero es llamado insesgado o *unbiased*.

Es importante que el estimador tenga un pequeño valor del error cuadrático medio.

Propiedades de los estimadores:

1- Insesgadez: Se denomina sesgo de un estimador a la diferencia entre la esperanza (valor esperado) del estimador y el verdadero valor del parámetro a estimar. Es deseable que un estimador sea insesgado o centrado, esto es, que el sesgo sea nulo para que la esperanza del estimador sea igual al valor del parámetro que se desea estimar.

- Un estimador $\hat{\theta}$ es insesgado (centrado) cuando $E[\hat{\theta}] = \theta$
- Un estimador $\hat{\theta}$ es sesgado cuando $E[\hat{\theta}] = \theta + \epsilon(\hat{\theta}) \Rightarrow b(\hat{\theta}) = E[\hat{\theta}] - \theta$ al termino $b(\hat{\theta})$ al se le denomina sesgo.
- Un estimador $\hat{\theta}$ es asintóticamente insesgado si su posible sesgo tiende a cero al aumentar el tamaño muestral que se calcula: $\lim_{n\to\infty} b(\hat{\theta}) = 0$

En términos del ECM

$$ECM(\hat{\theta}) = V(\hat{\theta}) + \left(E[\hat{\theta}] - \theta\right)^2 = V(\hat{\theta}) + b(\hat{\theta})^2$$

- Si $b(\hat{\theta}) = 0$ entonces $ECM(\hat{\theta}) = V(\hat{\theta})$ Esto indica que en media el estimador no está lejos del valor real del parámetro. Que es lo ideal.

La utilización de la estimación puntual como si fuera el verdadero valor del parámetro conduce a que se pueda cometer un error más o menos grande.

Ejemplo 1. Sea el estimador $\hat{\theta}$ del parámetro μ de una variable X con distribución normal $N(\mu, \sigma^2)$ y tamaño de muestra n.

$$\hat{\theta} = \frac{1}{n+1}\sum_{i=1}^{n} x_i$$

Calcule los valores del sesgo y del error cuadrático medio. Analice los resultados.

$$b(\hat{\theta}) = E[\hat{\theta}] - \theta$$

Por tanto tenemos que calcular $E[\hat{\theta}]$

$$E[\hat{\theta}] = E\left(\frac{1}{n+1}\sum_{i=1}^{n} x_i\right) = \frac{1}{n+1}E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n+1}\sum_{i=1}^{n} E(x_i)$$

Como $X \sim N(\mu, \sigma^2)\ \ y\ \ x_i \in X\ \rightarrow E(x_i) = \mu\ \forall i\ i = 1, ..n$

$$E[\hat{\theta}] = \frac{1}{n+1}\sum_{i=1}^{n} E(x_i) = \frac{1}{n+1}(n\mu) = \frac{n\mu}{n+1}$$

Por tanto

$$b(\hat{\theta}) = E[\hat{\theta}] - \mu = \frac{n\mu}{n+1} - \mu = \frac{n\mu - (n+1)\mu}{n+1} = \frac{n\mu - n\mu - \mu}{n+1} = -\frac{\mu}{n+1} \xrightarrow[n\to\infty]{} 0$$

Por tanto el estimador $\hat{\theta}$ es asintóticamente insesgado.

2- Consistencia.

Si no es posible emplear estimadores de mínima varianza, el requisito mínimo deseable para un estimador es que a medida que el tamaño de la muestra crece, el valor del estimador tienda a ser el valor del parámetro poblacional, propiedad que se denomina consistencia.

Un **estimador $\hat{\theta}$ consistente** es un estimador asintóticamente insesgado cuya varianza tiende a cero al aumentar el tamaño muestral.

El estimador $\hat{\theta}$ es consistente cuando $\lim_{n\to\infty} E(\hat{\theta}) = \theta$ y $\lim_{n\to\infty} V(\hat{\theta}) = 0$

3- Eficiencia: La eficiencia de los estimadores está limitada por las características de la distribución de probabilidad de la muestra de la que proceden.

Un **estimador** es **eficiente** cuando se verifica que **Es insesgado y** Posee **varianza mínima.**

Un estimador es más eficiente o más preciso que otro estimador, si la varianza del primero es menor que la del segundo. Sean $\hat{\theta}_1$y $\hat{\theta}_2$dos estimadores insesgados, se dice que $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si se verifica que $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.

Al comparar estimadores si no son insesgados tenemos que utilizar el Error cuadrático Medio.

Si $CM(\hat{\theta}_1) < ECM(\hat{\theta}_2)$ entonces $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$.

4- **Suficiencia**: La suficiencia de un estimador fue definida por Fisher, quien además de dar una rigurosa a definición también planteo una de carácter intuitivo con la que trabajaremos, la más rigurosa se sale del ámbito de este curso.

Un estimador **suficiente** resume toda la información relevante suministrada por la muestra.

Por ejemplo el estimador $\bar{X}$ de la μ en una población normal es Suficiente.

Estimadores Puntuales.

Estimador para la media μ

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Estimador insesgado, consistente, eficiente y suficiente.

Estimador para la varianza σ^2

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1}$$

Estimador insesgado, consistente, eficiente y suficiente.

Estimador para la proporción (Binomial):

El problema de estimar proporciones se presenta cuando al tener una población se desea estimar que parte de ella, o que proporción, cumple determinada propiedad o cierto atributo, llamémosle a esta propiedad o atributo A. Para esto construimos de la muestra una variable aleatoria binomial de la siguiente forma $X = \begin{cases} 1 & si\ ocurre\ un\ suceso\ A \\ 0 & si\ no\ ocurre\ dicho\ suceso\ A \end{cases}$ entonces estimamos la proporción de la misma sumando todas las ocurrencias del suceso A. Este estimador se llama estimador para la proporción de una binomial o estimador de la proporción del suceso A, la proporción es el estimador de $P(A)$.

$$\bar{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{m}{n}$$

Siendo m la cantidad de veces que ocurre un suceso.

Estimador insesgado, consistente, eficiente.

Estimador De Máxima Verosimilitud

La estimación por máxima verosimilitud es un método de optimización que supone que la distribución de probabilidad de las observaciones es conocida.

Definición #6 Función de Verosimilitud: La función de verosimilitud de una muestra $(x_1, \dots, x_n)$ se define como la función de densidad conjunta de $(x_1, \dots, x_n)$.

En particular si la muestra es simple aleatoria de una población con función de densidad $f_\theta(x)$ la cual depende del parámetro θ la función de verosimilitud tiene la siguiente forma:

$$f_\theta(x_1)f_\theta(x_2)\dots.f_\theta(x_n)$$

La función de verosimilitud depende de los valores de la muestra y del parámetro θ, no obstante se demnota como $L(\theta)$ considerándose implícita la dependencia de los valores de la muestra. En resumen

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = f_\theta(x_1)f_\theta(x_2)\dots.f_\theta(x_n)$$

Definición #7 Estimador Máximo Verosímil: Sea $(x_1, \dots, x_n)$ una muestra simple aleatoria con remplazo y sea θ un parámetro desconocido de la distribución de la población el cual se desea estimar. Se dirá que el estimador $\hat{\theta}$ es el estimador máximo verosímil si $\hat{\theta}$ es tal que:

$$L(\hat{\theta}) = \max_{\theta\in\Theta} L(\theta)$$

La mayoría de las funciones de verosimilitud tienen propiedades analíticas suficientemente regulares tales que se puede afirmar que el estimador máximo verosímil es la solución de la ecuación

$$\frac{\partial L(\theta)}{\partial\theta} = 0$$

No obstante debe comprobarse que la segunda derivada de $L(\theta)$ en la solución hallada es negativa. En muchas ocasiones, es más práctico encontrar el estimador de máxima verosimilitud considerando la función log-verosimilitud $lnL(\theta)$ en lugar de la función de verosimilitud $L(\theta)$, ya que es más fácil de manejar por ser estrictamente creciente y sus máximos coinciden con los de $L(\theta)$.

Ejemplo 2: Hallando el estimador de Máxima Verosimilitud para la varianza y la media de una muestra de tamaño n que se extrae de una distribución normal con parámetros μ, σ^2

Como

$$f(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)}{2\sigma^2}}$$

Entonces

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{\sum_{i=1}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Calculando

$$\ln L(\mu,\sigma^2) = \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n + \ln e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} = \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= n\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = n\left(\ln 1 - \ln\sqrt{2\pi}\sigma\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= -n\ln\sqrt{2\pi}\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = -n\left(\ln\sqrt{2\pi} + \ln\sigma\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= -\ln\sqrt{2\pi} - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Para la varianza derivando primero con respecto a σ.

$$\frac{\partial lnL(\mu,\sigma^2)}{\partial\sigma} = -\frac{n}{\sigma} - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}\frac{1}{\sigma^2} = -\frac{n}{\sigma} - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}\frac{-2}{\sigma^3} = -\frac{n}{\sigma} - \frac{1}{2}\frac{-2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

Para la media derivando primero con respecto a μ.

$$\frac{\partial lnL(\mu,\sigma^2)}{\partial\mu} = -\frac{1}{2\sigma^2}\frac{\partial\sum_{i=1}^{n}(x_i-\mu)^2}{\partial\mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial(x_i-\mu)^2}{\partial\mu}$$

Como

$$\frac{\partial(x_i-\mu)^2}{\partial\mu} = 2(x_i-\mu)\partial(x_i-\mu) = 2(x_i-\mu)(-1) = -2(x_i-\mu)$$

Entonces

$$\frac{\partial lnL(\mu,\sigma^2)}{\partial\mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial(x_i-\mu)^2}{\partial\mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}-2(x_i-\mu) = \frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

Resolviendo la ecuación 1 para un valor de $\hat{\mu}$ tenemos que

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\hat{\mu}) = 0 \rightarrow \sum_{i=1}^{n}(x_i-\hat{\mu}) = 0 \rightarrow \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\hat{\mu} = 0 \rightarrow \sum_{i=1}^{n}x_i - n\hat{\mu} = 0 \rightarrow \sum_{i=1}^{n}x_i = n\hat{\mu}$$

Por tanto encontramos que $\hat{\mu}$ es el estimador de la media que coincide con $\bar{X}$

$$\frac{\sum_{i=1}^{n}x_i}{n} = \hat{\mu} = \bar{X}$$

Resolviendo la ecuación 2 para un valor $\hat{\sigma}$ tenemos que

$$-\frac{n}{\hat{\sigma}}+\frac{1}{\hat{\sigma}^3}\sum_{i=1}^{n}(x_i-\mu)^2=0\rightarrow -\frac{n}{\hat{\sigma}}+\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\hat{\sigma}^3}=0\rightarrow\frac{-n\hat{\sigma}^2+\sum_{i=1}^{n}(x_i-\mu)^2}{\hat{\sigma}^3}=0$$

$$\rightarrow -n\hat{\sigma}^2+\sum_{i=1}^{n}(x_i-\mu)^2=0\rightarrow\sum_{i=1}^{n}(x_i-\mu)^2=n\hat{\sigma}^2$$

Por tanto encontramos que $\hat{\sigma}^2$ es el estimador de la varianza, note que es diferente a S^2

$$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}=\hat{\sigma}^2$$

Aunque S^2 no sea máximo verosímil se prefiere pues es insesgado consistente y eficiente.

El siguiente paso es hallar las segundas derivadas, evaluarlas en el estimador y ver si sus valores son negativos, solo entonces podemos decir que los estimadores son máximos verosímil.

$$\frac{\partial lnL(\mu,\sigma^2)}{\partial\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0\quad(1)$$

$$\frac{\partial lnL(\mu,\sigma^2)}{\partial\sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2=0\quad(2)$$

Para comprobar que de hecho son EMV, es necesario verificar que la segunda derivada de en la solución hallada es negativa.

$$\frac{\partial^2 lnL(\mu,\sigma^2)}{\partial^2\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=\frac{1}{\sigma^2}(-n)<0$$

Porque $\sigma^2>0$ y $n>0$ porque n es el tamaño de la muestra siempre positivo.

$$\frac{\partial^2 lnL(\mu,\sigma^2)}{\partial^2\sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2=\frac{n}{\sigma^2}-\frac{3}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

Sustituyendo

$$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}=\hat{\sigma}^2$$

Tenemos

$$\frac{\partial^2 lnL(\mu,\sigma^2)}{\partial^2\sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2=\frac{n}{\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}\right)^2}-\frac{3}{\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}\right)^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

Sustituyendo $\sum_{i=1}^{n}(x_i-\mu)^2=A$ para hacer los cálculos más simples.

$$\frac{\partial^2 lnL(\mu,\sigma^2)}{\partial^2\sigma}=\frac{n}{\frac{A^2}{n^2}}-\frac{3}{\frac{A^4}{n^4}}A=\frac{n^3}{A^2}-\frac{3n^4}{A^4}A=\frac{n^3}{A^2}-\frac{3n^4}{A^3}=\frac{n^3A^3-3n^4A^2}{A^6}$$

Analizando el resultado A es un valor positivo, por tanto $n^3A^3>0$, $3n^4A^2>0$ y $A^6>0$, además como A es positiva y $n^4>n^3$ entonces $n^3A^3-3n^4A^2<0$

Estimación por intervalos.

Al realizar estimaciones puntuales como estamos buscando un valor único es poco probable que coincida con el verdadero valor del parámetro. Es incluso imposible definir con que error exactamente estamos estimando. Este problema lo soluciona en gran medida la estimación por intervalos de confianza.

Definición #8 Intervalos de Confianza. Sea $(x_1, \dots x_n)$ una muestra aleatoria de una población con función de densidad f_θ. Sea $\theta_1 = \theta_1(x_1, \dots x_n)$ y $\theta_2 = \theta_2(x_1, \dots x_n)$ dos estadígrafos tales que $\theta_1 \leq \theta_2$ y que satisfacen que $P(\theta_1 \leq \theta \leq \theta_2) = 1 - \alpha$, entonces el intervalo aleatorio $[\theta_1, \theta_2]$ es llamado intervalo de confianza del $(1 - \alpha) * 100\%$ para θ.

- La constante $(1 - \alpha)$ se llama Nivel de Confianza.
- Los valores θ_1, θ_2 son llamados límites de confianza inferior y superior respectivamente.
- Una vez que estemos trabajando con calores numéricos lo correcto es decir que el intervalo $[\theta_1, \theta_2]$ contiene a θ con una confianza del $(1 - \alpha) * 100\%$

Media de una población Normal

Las principales aplicaciones de los Intervalos de Confianza se han producido siempre en las estimaciones de las medias poblacionales en particular cuando se está en presencia de una población normal.

Media con varianza conocida:

El procedimiento para encontrar los extremos del intervalo de confianza es similares para todos los parámetros de las distribuciones, por tanto ilustramos en el caso de la media de la distribución normal, como encontrarlos, mediante el cálculo de los percentiles de la distribución asociada al estadígrafo correspondiente al parámetro para el cual se halla la estimación puntual.

Como la distribución normal es simétrica con respecto a la media en vez de encontrar dos valores θ_1, θ_2, queremos encontrar dos valores $\theta_2 = -\theta_1$ tales que:

$$P(|\bar{X} - \mu| < \theta_2) = 1 - \alpha$$

En este caso θ_2 representa el error de la estimación puntual, o sea la diferencia entre el valor encontrado de la estimación puntual y el valor real del parámetro. Por tanto queremos encontrar:

$$P(\theta_1 < \bar{X} - \mu < \theta_2) = 1 - \alpha \tag{3}$$

Como la distribución muestral del promedio es $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Entonces tipificamos en 3 utilizando $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

$$P\left(\frac{\theta_1}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

$$P\left(\frac{\theta_1}{\frac{\sigma}{\sqrt{n}}} \leq Z \leq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

$$P\left(Z \leq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) - P\left(Z \leq \frac{-\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

$$P\left(Z \leq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) - P\left(Z \geq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

$$P\left(Z \leq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) - 1 + P\left(Z \leq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

$$2 * P\left(Z \leq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) - 1 = 1 - \alpha$$

$$2 * P\left(Z \leq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) = 2 - \alpha$$

$$P\left(Z \leq \frac{\theta_2}{\frac{\sigma}{\sqrt{n}}}\right) = \frac{2-\alpha}{2} = \frac{2}{2} - \frac{\alpha}{2} = 1 - \frac{\alpha}{2}$$

Luego

$$\frac{\theta_2}{\frac{\sigma}{\sqrt{n}}} = Z_{1-\frac{\alpha}{2}}$$

Despejando

$$\theta_2 = Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Por tanto $\theta_1 = -Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

Por tanto

$$P\left(-Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Pero nos interesa el intervalo de μ no de $\bar{X} - \mu$ por tanto despejamos $\bar{X}$ y luego multiplicamos por -1

$$P\left(-\bar{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\bar{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Por tanto

$$\mu \in \left[\bar{X} - Z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}; \bar{X} + Z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right]$$

Donde $Z_{1-\frac{\alpha}{2}}$ es un percentil de la distribución normal estándar y $e = Z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$ es el error de la estimación puntual.

Media con varianza desconocida $n \leq 30$:

$$\mu \in \left[\bar{X} - t(n-1)_{1-\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}\,; \bar{X} + t_{1-\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}\right]$$

Como la varianza es desconocida trabajamos con su estimación S^2 y además trabajamos con las estimaciones de la distribución t de student en vez de la distribución normal. Donde $t(n-1)_{1-\frac{\alpha}{2}}$ es un percentil de la distribución t de student o sea es un número que buscamos en la tabla de la t de student. Donde $(n-1)$ son los grados de libertad.

Media con varianza desconocida $n > 30$:

$$\mu \in \left[\bar{X} - Z_{1-\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}\,; \bar{X} + Z_{1-\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}\right]$$

Los intervalos atormente expuestos para la media poblacional tienen como restricción fundamental que requieren una distribución normal en la población. Cuando esta restricción no se cumple el teorema central del límite garantiza que para muestras grandes la distribución muestral de $\bar{X}$ es aproximadamente normal y entonces pueden emplearse los intervalos anteriores en poblaciones no normales, teniendo en cuenta que el nivel de confianza en estos casos es aproximado. Una de las aplicaciones más importantes es la estimación de proporciones o equivalentemente la estimación del parámetro p de la distribución Bernoulli.

Proporción.

El problema de estimar proporciones se presenta cuando al tener una población se desea estimar que parte de ella o proporción cumple determinada propiedad o tiene cierto atributo A. En este caso se considera a cada individuo de la población como una variable aleatoria con dos valores 1 si cumple la propiedad 0, si no. En este caso es una variable aleatoria con distribución Bernoulli y parámetro p su intervalo de confianza se calcula de la siguiente forma:

$$P(A) \in \left[\bar{p} - \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} Z_{1-\frac{\alpha}{2}}\,; p + \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} Z_{1-\frac{\alpha}{2}}\right]$$

Donde $Z_{1-\frac{\alpha}{2}}$ es el percentil de la distribución N y $\bar{p}$ es el estimador de la proporción.

Varianza de una población Normal.

Por último el estimador de la varianza que es el único no simétrico y que utiliza el percentil de la distribución χ^2.

$$\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\,; \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right]$$

Donde $\chi^2_{1-\frac{\alpha}{2}}$ es el percentil de la distribución Chi-cuadrado y S^2 es el estimador de la varianza.

Anexo. Funciones de Distribución Muestrales (Chi-Cuadrado, t-Student y F-Fisher)

Funciones de Distribución Muestrales

Al trabajar con muestras y funciones de variables aleatorias con distribución normal, tenemos que analizar qué ley de distribución siguen los estadígrafos que son funciones de las variables aleatorias, empleadas posteriormente en la inferencia estadística.

Definición #A1 Distribución Chi-cuadrado (χ^2)

Si χ^2 es una variable aleatoria continua cuya función de densidad probabilística es:

$$f(u) = \begin{cases} \dfrac{1}{\Gamma\left(\frac{n-2}{2}\right)} \cdot \left(\dfrac{1}{2}\right)^{\frac{n-2}{2}} \cdot u^{\left(\frac{n-2}{2}-1\right)} \cdot e^{-\frac{1}{2}x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se dice que la variable aleatoria X tiene distribución χ^2, con $\mathrm{n}-2$ grados de libertad, se denota $\mathrm{X} \sim \chi^2(\mathrm{n}-2)$

Una aplicación de esta ley de distribución es cuando tenemos n variables $Z_i \sim N(0,1)$ entonces

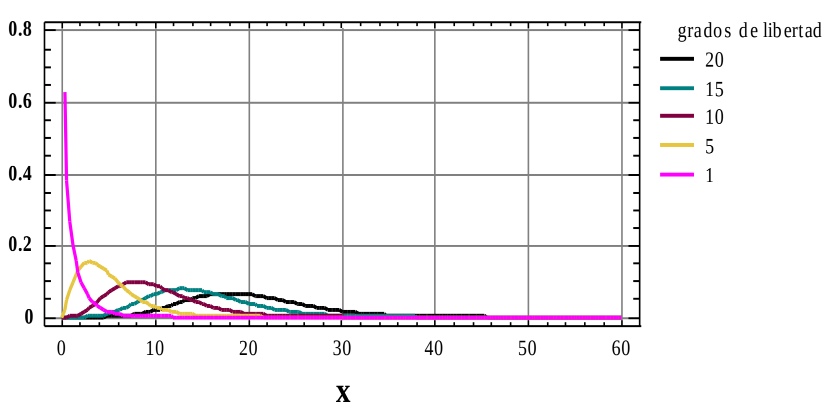
$$\chi^2 = \sum_{i=1}^{n} Z_i^2$$

Definición #A2 Grados de libertad.

Grados de libertad = número de mediciones – 1 = $n-1$

Propiedades Distribución Chi-cuadrado (χ^2)

1. La variable χ^2 siempre toma valores positivos.
2. Es asimétrica, lo cual se confirma en los gráficos de sus funciones de densidad.
3. Si $\mathrm{X} \sim \chi^2(\mathrm{n}-2)$ entonces $\mathrm{E(X)} = \mathrm{n}-2$ y $\mathrm{V(X)} = 2(\mathrm{n}-2)$.



Ejemplo 1: Utilice la tabla de la distribución Chi-cuadrado y calcule la probabilidad correspondiente.

$$P(X \leq 23.7)\ si\ X \sim \chi^2(14)$$

Se busca en la tabla "Valores críticos de la Distribución Chi-Cuadrado". Buscamos la fila que corresponde al grado de libertad 14 y en esa fila comenzamos a buscar el valor que más se acerca a 23.7 una vez encontrado miramos el valor de la probabilidad que encabeza esa columna.

$$P(X \leq 23.7) = 0.95$$

gl	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995
1									
…	…	…	…	…	…	…	…	…	…
14						23.6848			
…	…	…	…	…	…	…	…	…	…
60									

Al número 23.7 se le llama percentil por lo que si se les pide que hallen el percentil de una variable $X \sim {\chi_{0.95}}^2(14)$ entonces se busca en la tabla los 14 grados de libertad con la probabilidad 0.95 y entonces tenemos el percentil.

Distribución t de Student.

Es una distribución de probabilidad que surge del problema de estimar la medida de una población normalmente distribuida cuando el tamaño de la muestra es pequeño.

Si X y Z son variables aleatorias independientes, donde $\mathrm{Z} \sim \mathrm{N}(0,1)$ y $\mathrm{X} \sim \chi^2(\mathrm{n})$, entonces la variable aleatoria:

$$\mathrm{T} = \frac{\mathrm{Z}}{\sqrt{\frac{\mathrm{X}}{\mathrm{n}-1}}}$$

Sigue una distribución t-student con $\mathrm{n}-1$ grados de libertad.

Definición #A3 Distribución t de Student.

Sea t una variable aleatoria continua cuya función de densidad probabilística es:

$$\mathrm{f(t)} = \frac{\Gamma\left(\frac{\mathrm{n}}{2}\right)}{\Gamma\left(\frac{\mathrm{n}-1}{2}\right)} \cdot \frac{1}{\sqrt{\mathrm{n}-1\pi}} \cdot \frac{1}{\left(1+\frac{\mathrm{t}^2}{\mathrm{n}-1}\right)^{\frac{\mathrm{n}}{2}}} \quad \mathrm{t} \in \mathbb{R}$$

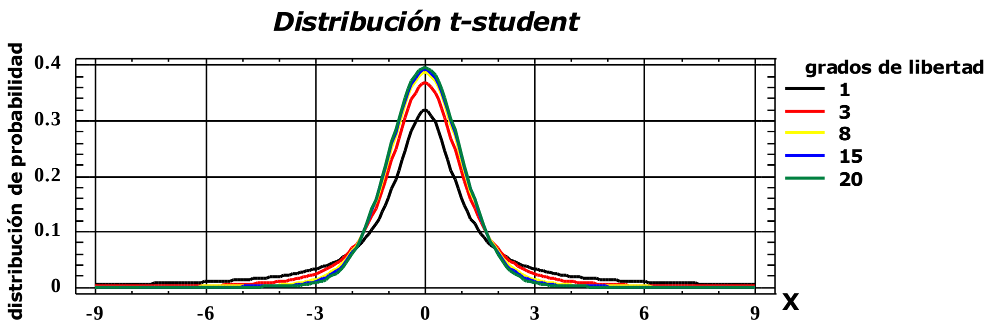
Se dice entonces que t tiene una distribución t-student con k grados de libertad y se denota como $t \sim t(n-1)$.

Nota: La función $\Gamma(\mathrm{x})$ Gamma de x, es una integral impropia que se calcula por métodos recursivos cuando $x \in \mathbb{R}$.

Propiedades

1. La distribución t-student es simétrica respecto a 0 y el gráfico de la función de densidad se asemeja al de la normal estándar.
2. $\mathrm{E(t)} = 0,\ \mathrm{V(t)} = \frac{\mathrm{n}}{\mathrm{n}-2}$ con $\mathrm{n} > 2$.
3. A medida que n aumenta, la dispersión de la curva t-student correspondiente disminuye.

4. A medida que $n \to \infty$, la distribución t-student se aproxima a la normal estándar.



Ejemplo 2: Utilice la tabla de la distribución t-student y calcule la probabilidad correspondiente.

$$P\ (T \geq 2.26\ )\ si\ T \sim t\ (10)$$

Se busca en la tabla "Valores críticos de t Distribución t-Student". Buscamos la fila que corresponde al grado de libertad 10 y en esa fila comenzamos a buscar el valor que más se acerca a 2.26 una vez encontrado miramos el valor de la probabilidad que encabeza esa columna. $P(T \geq 2.26\ ) = 0.975$

gl	Puntos percentiles						
	0.7	0.8	0.9	0.95	0.975	0.99	0.995
1							
…	…	…	…	…	…	…	…
10					2.2281	2.7638	
…	…	…	…	…	…	…	…
∞							

Al número 2.26 se le llama percentil por lo que si se les pide que hallen el percentil de una variable $T \sim {t_{0.99}}^2(14)$ entonces se busca en la tabla los 10 grados de libertad con la probabilidad 0.99, como esta no se encuentra explicita podemos tomar el valor de probabilidad más cercano o sea 0.975 o 0.99 en este caso 0.99 es más exacto y entonces tenemos un aproximado el percentil. Que en este caso es 2.76.

Distribución F de Fisher.

Si X_1 y X_2 son variables aleatorias independientes, con distribución χ^2 con n y m grados de libertad respectivamente, entonces la variable aleatoria:

$$F = \frac{\frac{X_1}{n}}{\frac{X_2}{m}}$$

Sigue una distribución F con n y m grados de libertad.

Definición #A4 Distribución F de Fisher.

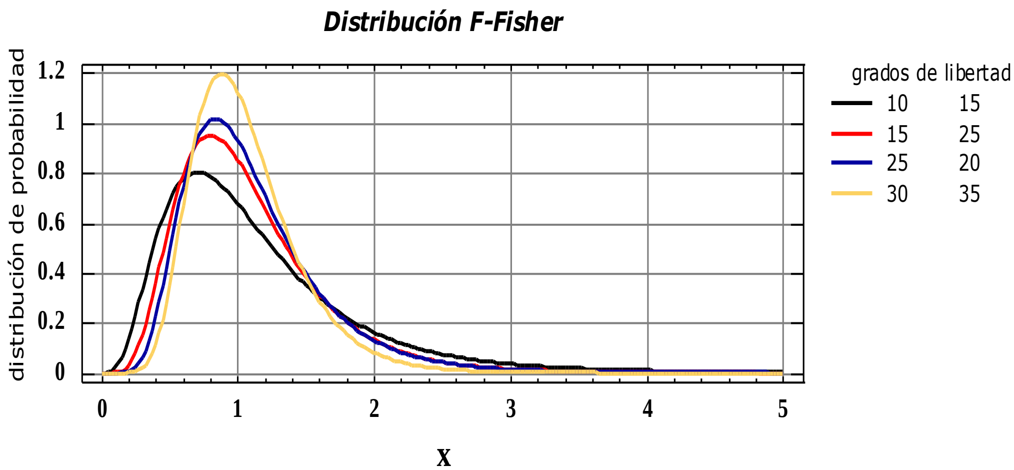
Sea X una variable aleatoria continua cuya función de densidad probabilística es:

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\frac{m+n}{2}\right) \cdot \left(\frac{m}{n}\right)^{m/2} \cdot x^{\left(\frac{m}{2}-1\right)}}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right) \cdot (mx+n)^{\frac{m+n}{2}}} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se dice que X tiene una distribución F-Fisher con n y m grados de libertad, se denota $X \sim F(n, m)$

Propiedades

1. $F(n, m)_{\alpha} = \frac{1}{F(m,n)_{1-\alpha}}$



Ejemplo 3: Utilice la tabla de la distribución F-Fisher y calcule la probabilidad correspondiente

Busque cual es el valor de $F \sim F(\ 2, 4\ )$ para un nivel de significación del 5%.
Se busca en la tabla "Valores críticos de la Distribución $F(N, M)$". Como podrán notar esta tabla es diferente pues especifica el nivel de significación de la variable teniendo 3 o más tablas cada una con un valor de probabilidad. Para buscar primero buscamos la tabla que corresponde a la probabilidad indicada, luego la fila que corresponde a la N en este caso 2 y por columnas buscamos el valor de M que en este caso es 4 por tanto en la celda 2,4 de la tabla encontraremos el valor de la Variable F para cuando la probabilidad de que $P(F \geq 19.49) = 0.95$.

N|M	1	2	3	4	5	6	7
1							
2				19.491			
…	…	…	…	…	…	…	…
∞							

En este caso también llamamos al valor 19.49 el percentil.