Estadística

Clase 4

Rodrigo Del Rosso

06 de Mayo de 2022

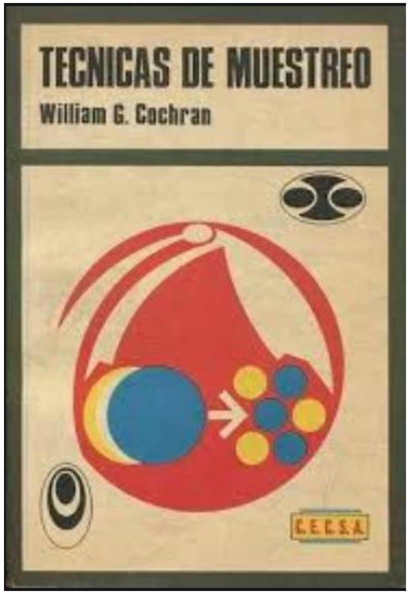
Definiciones Básicas

En la tarea estadística raramente se conocen cuales son las distintas funciones que describen probabilísticamente el comportamiento de la, o las variables, que representan las características de aquello que se está investigando.

A veces se está en condiciones de suponer formas funcionales de las distribuciones de probabilidad. Aún así, comúnmente se ignoran los valores de los parámetros de dichas distribuciones.

Uno de los propósitos de la Disciplina es poder inferir las distribuciones de probabilidad, sus propiedades y sus parámetros.

La técnica básica que se utiliza es el Muestreo.



Universo

Es el conjunto de Unidades Experimentales que poseen características comunes observables y que se utilizan para obtener información sobre un hecho particular. El mismo queda determinado cuando se definen los objetivos del trabajo que se llevará a cabo, como así también las Unidades Experimentales sobre las que se realizarán las observaciones.

Población

Se denomina así a cualquiera variable particular que se estudia a un Universo. Es decir, según esta definición cada Universo puede generar varias Poblaciones. Una por cada una de las variables cuya medición sea de interés para alcanzar los objetivos fijados.

Censo

Se denomina así a la medición en la totalidad de las Unidades Experimentales que conforman el Universo, de todas las variables que previamente hayan sido declaradas relevantes para la investigación a llevar a cabo.

$$N = Tamaño\ del\ Universo\ (o\ Población)$$

Los Universos pueden ser finitos o infinitos, dependiendo de la cantidad de elementos que los conforman y, consecuentemente, las poblaciones serán finitas o infinitas. El Muestreo consiste en observar y medir algunos elementos del Universo para obtener información necesaria para cumplir con los objetivos fijados para concretar el trabajo.

Muestra

Es un subconjunto o parte de una Población tomada de forma tal, que con ella se pueda hacer un juicio acerca de esa Población completa.

$$n = Tamaño\ de\ la\ Muestra$$

Fracción de Muestreo

Se denomina al cociente entre el tamaño de la muestra y el de la Población.

$$f_m = \frac{n}{N}$$

Inferencia Estadística

Se denomina así a cualquier afirmación que se realiza sobre una determinada población, basándose en los datos obtenidos con una muestra, pudiéndose obtener, a partir del cálculo de probabilidad, una determinada medida de la incertidumbre que se genera.

Se infiere la población a partir de la muestra

De esta manera las conclusiones a la que se arriba por estar basadas en ignorancias parciales, producen un cierto grado de duda, el cual podría estar controlado probabilísticamente si la muestra se selecciona mediante un método que garantice la aleatoriedad.

Introducción al Muestreo

Introducción al Muestreo

Para obtener una muestra que permita inferir adecuadamente la población en estudio, es necesario tener en cuenta algunas reglas y operaciones, las cuales dependerán de los objetivos fijados.

El **Muestreo** es el procedimiento mediante el cual se obtienen una o más muestras de una población dada.

La **Unidad de Muestreo** se denominada a cada unidad experimental, o grupo de unidades experimentales, que son tomadas para obtener una muestra.

El **Diseño Muestral** se denomina a un plan de muestreo específico donde se establecen cuáles serán los procedimientos a seguir para tomar una o más muestras.

Las unidades experimentales que intervienen en una muestra, pueden ser tomadas con mayor o menor grado de subjetividad por parte del sujeto que se encarga de realizar la muestra. De esta forma se originan distintos tipos de muestreo.

Tipos de Muestreo

- **Probabilístico**

  Cuando las Unidades Experimentales que componen la muestra son tomadas al azar, pudiéndose calcular a priori, la probabilidad que tiene cada muestra y, como consecuencia, cada Unidad Experimental, de ser la obtenida. Esto significa que el proceso de obtención de cada uno de los elementos es un Experimento Aleatorio.

- **Intencional**

  Cuando las Unidades Experimentales que componen la muestra son obtenidas siguiendo una regla o norma preestablecida. Es un tipo de muestreo subjetivo y carece de una base teórica satisfactoria.

- **Sin Norma**

  Cuando las Unidades Experimentales que componen la muestra se realiza sin una norma, regla o criterio definido. Se considera que la intención es cuasi-aleatoria, por lo tanto el muestreo es cuasi-objetivo.

Métodos de Obtención de Muestras

En el curso se estudiará la aplicación de los Métodos Estadísticos en el Muestreo Probabilístico, por lo tanto, se supone que cada elemento que interviene en la muestra es tomada al azar.

Una muestra de tamaño n cuyos elementos se obtienen al azar de una población con función de densidad de probabilidad $f_X(x)$ o función de probabilidad $P[X = x]$, según corresponda, es un conjunto formado por n variables aleatorias que tienen la misma distribución de probabilidad.

$$Muestra: \{x_1; x_2; \ldots; x_n\}$$

$$f_X(x_1) = f_X(x_2) = \cdots = f_X(x_n)$$

$$P[X = x_1] = P[X = x_2] = \cdots = P[X = x_n]$$

Métodos de Obtención de Muestras

La naturaleza de las Poblaciones que serán objeto del muestreo puede ser de lo más variada.

Por ejemplo, puede tratarse de una población homogénea o heterogénea, las unidades experimentales pueden presentarse sistemáticamente con una determinada periodicidad, o pueden estar agrupadas formando conglomerados, etc.

Estas particularidades que presentan las poblaciones dan origen a los Métodos de Obtención de las muestras que resulten más adecuados, algunos de los cuales se describirán brevemente en las próximas slides.

- ***Muestreo simple al azar (sin reemplazo)***
- ***Muestreo estratificado al azar***
- ***Muestreo conglomerados polietápico***
- ***Muestreo sistemático al azar***

Muestreo Simple al Azar (*m.a.s.*)

CAPITULO 2

Muestreo Aleatorio Simple

Consiste en obtener al azar una muestra de n elementos de entre los N que constituyen el Universo.

Hay que tener en cuenta que todas las muestras posibles de tamaño n deben tener la misma probabilidad de ser tomadas, como así también, que todos los elementos que integrarán la muestra tengan, en el momento de cada extracción, la misma probabilidad de ser obtenidas. Disponemos de un Universo Finito de tamaño N.

Como cada una de las unidades que se extraen no se las repone, la cantidad de muestras distintas, igualmente posibles de tamaño n que pueden obtenerse es una combinación de N elementos tomados de a n. La probabilidad de que una muestra particular sea la obtenida es de acuerdo a la definición clásica,

$$\frac{n}{N} * \frac{n-1}{N-1} * \frac{n-2}{N-2} * \cdots * \frac{1}{N-n+1} = \frac{n!\,(N-n)!}{N!} = \frac{1}{\binom{N}{n}}$$

Muestreo Simple al Azar (*m.a.s.*)

Sobre la base de esto, por ser equiprobables todas las muestras de tamaño n, la probabilidad de que un elemento cualquiera de la población forme parte de la muestra, se puede calcular haciendo,

$$\frac{Cantidad\ de\ Muestras\ Favorables}{Cantidad\ de\ Muestras\ Posibles} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

Por lo tanto, todos los elementos tienen la misma probabilidad de ser extraídos.

Si la población es infinita y las variables muestrales son independientes, la probabilidad de que una de ellas asuma un determinado valor, es la misma que para cualquier valor que asuman las otras. Caso contrario, si hay dependencia estadística, dicha probabilidad dependerá del valor de las que fueron obtenidas. Las Unidades Experimentales son tomadas al azar, cualquiera sea la posición que ocupen en el Universo.

Muestreo Estratificado al Azar

Cuando la población presenta una gran variabilidad, es decir una población heterogénea, la utilización del Muestreo Simple al Azar puede proporcionar muestras no representativas y las conclusiones que surjan del análisis ellas no serán del todo confiables. Cuando ocurre esto el método de muestreo más adecuado es el Estratificado al Azar. Consiste en particionar al Universo en Estratos (o clases o subpoblaciones), dentro de los cuales si la variable debe presentar homogeneidad. De cada uno de los estratos se obtiene una muestra simple al azar.

$N\colon Tamaño\ del\ Universo$

$h\colon Cantidad\ de\ Estratos$

$N_j\colon Tamaño\ del\ j - ésimo\ estrato$

Se debe cumplir que,

$$\sum_{j=1}^{h} N_j = N$$

Muestreo Estratificado al Azar

Si el tamaño de la muestra es n y de cada uno de los estratos se toma una muestra de tamaño n_j entonces se debe cumplir que,

$$\sum_{j=1}^{h} n_j = n$$

Una cuestión importante a resolver es cuánto del total n deberá asignarse a cada uno de los estratos.

La asignación del tamaño de la muestra a cada uno de los distintos estratos se llama “Afijación” y puede realizarse de alguna de las siguientes formas,

- *Afijación uniforme*
- *Afijación proporcional*
- *Afijación óptima*

Muestreo Estratificado al Azar

- *Afijación uniforme*

El tamaño de la muestra que le corresponde a cada estrato es igual para todos. Este tamaño se calcula, entonces, haciendo el cociente entre el tamaño de la muestra, n, y la cantidad de estratos, h.

$$n_1 = n_2 = \cdots = n_h = \frac{n}{h}$$

- *Afijación proporcional*

El tamaño de la muestra que le corresponde a cada estrato es proporcional al tamaño del estrato. Se determina mediante el producto entre la fracción de muestreo y el tamaño de cada estrato.

$$n_j = \frac{n}{N} * N_j$$

Muestreo Estratificado al Azar

- *Afijación óptima*

El tamaño de la muestra que le corresponde a cada estrato es proporcional al tamaño del estrato y al desvío estándar correspondiente. De esta manera se tiene en cuenta la falta de homogeneidad entre las subpoblaciones.

$$n_j = \frac{n * N_j * \sigma_j}{\sum_{i=1}^{h} N_i * \sigma_i}$$

Muestreo por Conglomerados Polietápico

Cuando se toman muestras mediante los métodos anteriores, las unidades de muestreo coinciden con las unidades experimentales.

El método es un caso más general que tiene como particularidad, que las unidades de muestreo está formada por un grupo de unidades experimentales.

El método consiste en agrupar elementos que conforman el Universo en conglomerados, de modo tal que entre ellos haya la suficiente homogeneidad, como para representar a la población.

El proceso de obtención de la muestra se hace sobre la base de estos conglomerados, que constituyen las unidades de muestreo y no sobre la base de las unidades físicas, con excepción de la última etapa.

Este método es adecuado para utilizarlo cuando las unidades experimentales, naturalmente, están agrupadas en conglomerados y cada una de ellas puede ser, a su vez, considerada como unidad de muestreo.

Muestreo por Conglomerados Polietápico

Generalmente los conglomerados son superficies o áreas en que se ha dividido el ámbito ocupado por el Universo.

Si es a una etapa, se toma una muestra simple al azar de conglomerados y en cada uno de ellos se miden todos los elementos que constituyen el conglomerado.

Si se trata de dos etapas de muestreo o bietápico, se toma una muestra simple al azar de conglomerados (1era etapa de muestreo) y de cada uno de estos conglomerados, se toma una muestra de unidades experimentales (2da etapa de muestreo).

Si se trata de tres etapas de muestreo (trietápico), primero se constituyen los conglomerados y dentro de cada uno de ellos se forman los subconglomerados, luego, se toma una muestra simple al azar de conglomerados (1era etapa de muestreo), de cada uno de ellos una muestra simple al azar de subconglomerados (2da etapa de muestreo) y por último, de cada subconglomerado una muestra simple al azar de unidades físicas (3era etapa de muestreo). Este procedimiento puede generalizarse obteniéndose un muestreo con varias etapas o Muestreo Polietápicos.

Muestreo Sistemático al Azar

Consiste en ordenar a las N Unidades Experimentales que conforman el Universo, de acuerdo a como se fueron presentando y obtener la muestra eligiendo, sistemáticamente un elemento cada c unidades, tomado el primero de ello en forma aleatoria. Si el Universo es finito, el número c es la parte entera del cociente entre el tamaño del universo y el de la muestra.

$$c = ent\left(\frac{N}{n}\right)$$

Si el Universo es infinito, el número c se elige arbitrariamente, sobre la base del buen saber y entender de quien realiza el trabajo.

Es adecuado para ser utilizado en aquellos casos en los cuales las Unidades Experimentales que forman el Universo se presentan con una determinada periodicidad. En esta situación, hay que evitar que el número c sea igual al período con que se presentan las Unidades Experimentales en el Universo, porque si ello ocurriese se perdería representatividad en la muestra.

Características Poblaciones y Muestrales

Parámetro Estadístico

Toda población es una variable aleatoria, por tal motivo su comportamiento probabilístico está explicado por una Función de Probabilidad (si la variable es discreta) o por una Función de Densidad (si la variable es continua).

Existen medidas que caracterizan a estas funciones y cumplen un papel importante en el análisis inferencial y se las define distinguiendo dos tipos de universos.

- **Universo Finito y Pequeño**

  Se denomina a toda medida que resume información calculada con las variables poblacionales.

- **Universo Finito y Grande (Infinito)**

  Se denomina a todo Parámetro Matemático de una función de probabilidad o de densidad de probabilidad que brinda información acerca de una Población.

Ejemplos de Parámetros

1. El total de elementos que presenta un determinado atributo
2. La proporción de elementos que presentan un determinado atributo
3. La media aritmética
4. La varianza
5. El desvío estándar
6. Etc.

Cuando se quiera representar a un parámetro indefinido se utilizará la letra griega Theta θ.

Parámetro: Media Poblacional μ

$$\mu = \sum_{i=1}^{N} \frac{X_i}{N}$$

Parámetro: Varianza Poblacional σ^2

$$\sigma^2 = \sum_{i=1}^{N} \frac{(X_i - \mu)^2}{N}$$

Parámetro: Proporción Poblacional p

$$p = \sum_{i=1}^{N} \frac{X_i}{N}$$

$$X_i = \begin{cases} 0 & si\ la\ unidad\ experimental\ medida\ NO\ tiene\ el\ atributo \\ 1 & si\ la\ unidad\ experimental\ medida\ SI\ tiene\ el\ atributo \end{cases}$$

Si $X \sim f_X(\theta)$ significa que la población X se distribuye con función de densidad de probabilidad $f_X(x)$ y parámetro θ.

Los parámetros no son constantes matemáticas, son números reales que pueden asumir cualquier valor que se encuentre dentro de un conjunto específico.

Este conjunto de números reales se llama **Espacio Parámetrico** y se simboliza con la letra griega Omega Ω.

Si fuese posible realizar un censo, entonces se podría conocer los verdaderos valores de los parámetros correspondientes a las poblaciones que son objeto de la investigación, pero si por alguna razón el censo es impracticable, estos valores son desconocidos.

En estos casos a los parámetros se los puede inferir, mediante determinadas funciones que se generan con las variables muestrales.

Estadígrafo y Estimador

Se definió una muestra aleatoria de tamaño n, como un conjunto formado por n variables aleatorias, todas con la misma distribución de probabilidad.

$$Muestra\colon \{x_1; x_2; \ldots; x_n\}$$

Se llama **Estadígrafo** a toda función escalar $h(x_1; x_2; \ldots; x_n)$ generada con las variables muestrales.

Dado que son funciones generadas por variables aleatorias, también son variables aleatorias. Luego existirá una función de densidad de probabilidad o de probabilidad, lo que corresponda, que describa su comportamiento probabilístico, como así también es posible que tengan una Esperanza Matemática y Varianza finita.

Se llama **Estimador** de un parámetro θ a todo Estadígrafo que proporcione información acerca de dicho parámetro.

Se llama **Estadígrafo de Transformación** a aquel Estadígrafo que permite transformar al estimador, en una variable que tenga una determinada distribución de probabilidad.

Para tener una adecuada información acerca de un parámetro desconocido θ, perteneciente a una población X, finita o infinita, se debería obtener una muestra aleatoria de tamaño n.

$$\{x_1; x_2; \dots; x_n\}$$

Y con ella generar un Estadígrafo específico,

$$\hat{\theta} = g(x_1; x_2; \dots; x_n)$$

Si puede proporcionar información acerca del parámetro θ entonces $\hat{\theta}$ es un estimador del mismo.

Estimación

Los métodos que se utilizan para asacar conclusiones sobre el parámetro acerca de las variables muestrales son dos,

- **Estimación Puntual**

  Del parámetro θ es un método de estimación que consiste en calcular el valor numérico único que asume el estimador, luego de tomar la muestra y realizar las mediciones correspondientes. Este valor numérico se llama Punto de Estimación.

$$\hat{\theta}_p = g(x_1; x_2; \dots; x_n)$$

- **Estimación por Intervalo**

  Del parámetro θ es un método de estimación que consiste en calcular, con los datos de la muestra, los límites de un conjunto cerrado y acotado de números reales. Este conjunto se denomina Intervalo de Estimación.

$$L_i(\hat{\theta}_p) \leq \theta \leq L_s(\hat{\theta}_p)$$

Distribución de los Estimadores

Recordando que por definición todo Estimador es una Variable Aleatoria, se define formalmente lo siguiente,

Sea $X \sim f_X(\theta)$ una población X cuya distribución de probabilidad tiene un parámetro estadístico desconocido θ.

Sea $\{x_1; x_2; \dots; x_n\}$ una muestra de tamaño n

Sea $\hat{\theta} = g(x_1; x_2; \dots; x_n)$ un estimador de dicho parámetro cuya Esperanza Matemática y Varianza sean finitas. En términos formales,

$$E[\hat{\theta}] < \infty \qquad Var[\hat{\theta}] < \infty$$

Se denomina **Distribución de Probabilidad del Estimador** $\hat{\theta}$ a aquella función de densidad de probabilidad o de probabilidad (según corresponda) que describe su comportamiento probabilístico.

Sesgo

Se denomina a la diferencia entre la Esperanza Matemática del Estimador y el parámetro a estimar. Formalmente,

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

Error Medio Cuadrático

Se denomina a la Esperanza Matemática del cuadrado de la diferencia entre el Estimador $\hat{\theta}$ y el parámetro θ.

$$EMC(\hat{\theta}) = E[\hat{\theta} - \theta]^2$$

Mide la variabilidad del estimador con respecto al parámetro que está estimando. Se demuestra que,

$$EMC(\hat{\theta}) = Bias(\hat{\theta})^2 + Var(\hat{\theta})$$

Propiedades de los Buenos Estimadores

De todos los estimadores que se pueden construir para inferir a un determinado parámetro, los mejores son aquellos que cumplen con ciertas condiciones generales o propiedades, con las cuales es posible garantizar la buena información que puedan proporcionar, acerca del parámetro a estimar. Las propiedades más importantes que debe tener un estimador $\hat{\theta}$ para ser considerado un "buen" estimador del parámetro θ son las siguientes,

- *Insesgamiento (Insesgadez)*
- *Eficiencia*
- *Consistencia*
- *Suficiencia*
- *Etc.*

Estimador Insesgado

El Estimador $\hat{\theta}$ es un estimador insesgado del parámetro θ si y sólo si, la Esperanza Matemática del Estimador $\hat{\theta}$ es igual al parámetro θ.

$$E[\hat{\theta}] = \theta$$

En otras palabras, el **estimador es insesgado** si su sesgo es igual a cero.

$$Bias(\hat{\theta}) = 0$$

En algunas ocasiones, esta propiedad puede no cumplirse directamente. Sin embargo, sí podría cumplirse si se hiciese crecer indefinidamente al tamaño de la muestra.

En esta situación se hace alusión al estimador $\hat{\theta}$ **es asintóticamente insesgado** del parámetro θ si se cumple lo siguiente,

$$\lim_{n \to \infty} E[\hat{\theta}] = \theta$$

Si el Estimador $\hat{\theta}$ del del parámetro θ es insesgado o asintóticamente insesgado, entonces el **Error Cuadrático Medio** es igual a la varianza del Estimador directamente o en el límite cuando n tiende a infinito.

Recordar que,

$$EMC(\hat{\theta}) = Bias(\hat{\theta})^2 + Var(\hat{\theta})$$

Entonces si el estimador es directamente insesgado o en forma asintóticamente,

$$EMC(\hat{\theta}) = Var(\hat{\theta})$$

UNIVERSIDAD AUSTRAL | INGENIERÍA

Estimador Eficiente

Otra de las condiciones con las que debería cumplir un estimador, para que pueda ser considerado un "Buen Estimador" es que su variabilidad con respecto al parámetro, se mide con el Error Medio Cuadrático debería ser mínima. Formalmente,

El Estimador $\hat{\theta}$ es un estimador eficiente del parámetro θ si y sólo si, se cumple que el estimador Estimador $\hat{\theta}$ tiene la menor varianza que puede tener un estimador del parámetro θ.

La búsqueda de un estimador que posea esta cualidad se facilita con la aplicación del Teorema de Cramer-Rao en el que se demuestra que la varianza del Estimador $\hat{\theta}$ necesariamente debe satisfacer la siguiente desigualdad,

$$Var(\hat{\theta}) \geq \frac{1}{nE\left[\frac{\partial ln f_X(\theta)}{\partial \theta}\right]^2}$$

Esta desigualdad se conoce como Cota de Cramer-Rao e indica cual es la menor varianza que puede tener un estimador.

Eficiente relativa

Dado dos estimadores insesgados de un mismo parámetro θ,

$$\hat{\theta}_1 = g_1(x_1; x_2; \dots; x_n)$$

$$\hat{\theta}_2 = g_2(x_1; x_2; \dots; x_n)$$

Se dice que el estimador $\hat{\theta}_1$ tiene una eficiencia relativa mayor al estimador $\hat{\theta}_2$ si y sólo si la varianza del estimador $\hat{\theta}_1$ es menor a la varianza del estimador $\hat{\theta}_2$.

Si los estimadores son sesgados, entonces la eficiencia relativa entre los estimadores se establece comparando el Error Cuadrático Medio de cada uno de ellos.

Consistencia

El Estimador $\hat{\theta}$ es un Estimador Consistente del parámetro θ, si y solo, si se cumple que el estimador $\hat{\theta}$ converge en probabilidad al parámetro θ, cuando el tamaño de la muestra n tiende a infinito.

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right|<\delta\right)=1 \qquad para\, todo\, \delta>0$$

En otras palabras, un estimador es consistente si a medida que el tamaño de la muestra crece indefinidamente (tiende a infinito), la probabilidad de que la diferencia entre el estimador y el valor del parámetro que pueda hacerse tan pequeña como se quiera, tiende a la unidad.

Suficiencia

El Estimador $\hat{\theta}$ es un Estimador Suficiente del parámetro θ, si y solo, si se cumple que el estimador $\hat{\theta}$ utiliza toda la información relevante acerca del parámetro θ contenida en la muestra aleatoria.

Grados de Libertad

Es una expresión introducida por Ronald Fisher, dice que, de un conjunto de observaciones, los **grados de libertad** están dados por el número de valores que pueden ser asignados de forma arbitraria, antes de que el resto de las variables tomen un valor automáticamente, producto de establecerse las que son libres, esto, con el fin de compensar e igualar un resultado el cual se ha conocido previamente.

En otras palabras, se denomina así a la cantidad de variables libres o estadísticamente independientes, que intervienen en un problema o en una distribución asociada a un problema. En la generación de los distintos estadígrafos, es frecuente que por alguna circunstancia, no todas las variables muestrales que intervienen en él, sean estadísticamente independientes. Los grados de libertad, entonces, serán inferiores al tamaño de la muestra. En un estadístico calculado en base a n datos, se refiere al número de cantidades independientes que se necesitan en su cálculo, menos el número de restricciones que vinculan a las observaciones y el estadístico. Simbólicamente se representa por $gl = n - k$.

$n = Cantidad\ de\ Observaciones$
$k = Cantidad\ de\ Restricciones$

Grados de Libertad

Supongamos que contamos con una muestra aleatoria simple de tamaño 4,

$$\{x_1; x_2; x_3; x_4\} = \{10; 30; 70; 40\} \Rightarrow \bar{X} = 37{,}5$$

Dado que la suma de las desviaciones respecto a la media aritmética debe ser nula. Formalmente,

$$\sum_{i=1}^{n} (x_i - \bar{X}) = 0$$

Solamente se podrán asignar libremente valores a $n-1$ variables y en nuestro ejemplo,

$$(10 - 37{,}5) = -27{,}5 \quad (30 - 37{,}5) = -7{,}5 \quad (70 - 37{,}5) = 32{,}5$$

La última desviación no puede ser cualquier número real, solamente podrá ser aquel valor que haga cero la suma de las desviaciones. Formalmente,

$$(40 - 37{,}5) = 2{,}5$$

La suma del cuadrado de los desvíos tiene 3 grados de libertad.

Algunos Estimadores Importantes

Parámetro: Media Poblacional μ y Estimador: Media Muestral $\overline{X}$

$$\mu = \sum_{i=1}^{N} \frac{X_i}{N} \Rightarrow \hat{\mu} = \overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

Parámetro: Varianza Poblacional σ^2 y Estimador: Varianza Muestral S_X^2

$$\sigma^2 = \sum_{i=1}^{N} \frac{(X_i - \mu)^2}{N} \Rightarrow \hat{\sigma}^2 = S_X^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

Parámetro: Proporción Poblacional p y Estimador: Proporción Muestral $\hat{p}$

$$p = \sum_{i=1}^{N} \frac{X_i}{N} \Rightarrow \hat{p} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$X_i = \begin{cases} 0 & si\ la\ unidad\ experimental\ medida\ NO\ tiene\ el\ atributo \\ 1 & si\ la\ unidad\ experimental\ medida\ SI\ tiene\ el\ atributo \end{cases}$$

Parámetro: Media Poblacional μ y Estimador: Media Muestral $\overline{X}$

$$\mu = \sum_{i=1}^{N} \frac{X_i}{N} \qquad \hat{\mu} = \overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

Supuestos

- *Población Normal, Varianza Poblacional Conocida y Población Infinita*

Sea X una Población que se distribuye normalmente con Esperanza Matemática y Varianza Poblacional igual a μ y σ^2 respectivamente. Bajo estas condiciones se demuestra que,

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Supuestos

- *Población Normal, Varianza Poblacional Conocida y Población Finita*

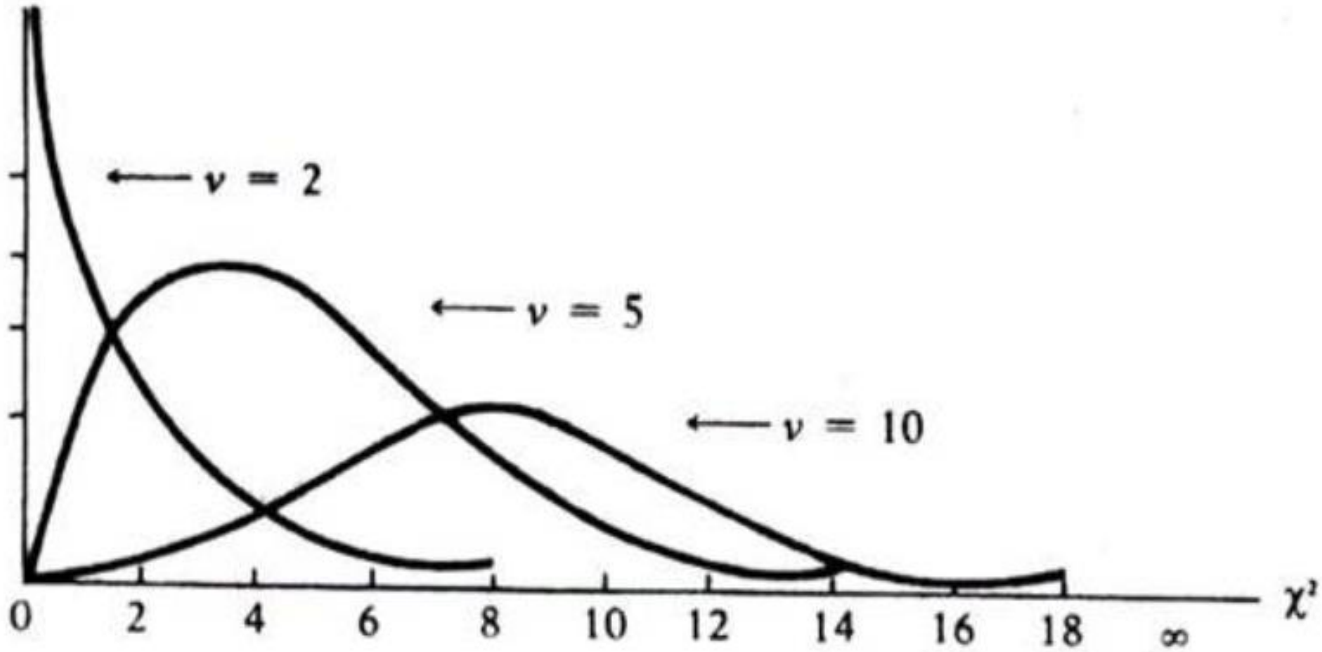
Sea X una Población que se distribuye normalmente con Esperanza Matemática y Varianza Poblacional igual a μ y σ^2 respectivamente. Bajo estas condiciones se demuestra que,

$$\boldsymbol{\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) \Rightarrow Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \sim N(0,1)}$$

Parámetro: Varianza Poblacional σ^2 y Estimador: Varianza Muestral S_X^2

Distribución Chi (Ji) Cuadrado

$$J = \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$



Parámetro: Varianza Poblacional σ^2 y Estimador: Varianza Muestral S_X^2

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$$

$$(n-1)\hat{\sigma}^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

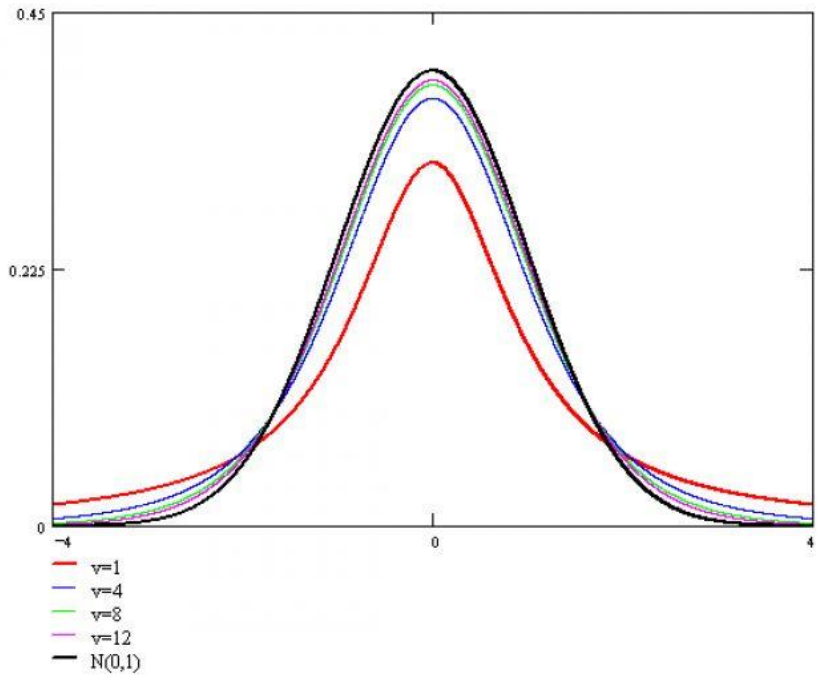
$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

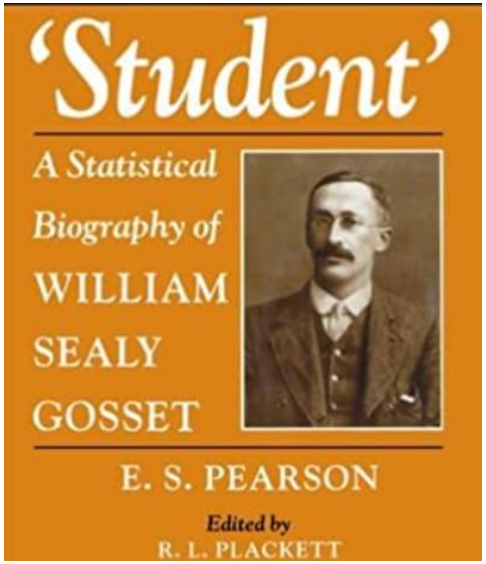
$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

Parámetro: Media Poblacional μ y Estimador: Media Muestral $\bar{X}$

Sea X una Población que se distribuye normalmente con Esperanza Matemática μ y Varianza Poblacional (desconocida) σ^2 respectivamente. Bajo estas condiciones se demuestra que la distribución de probabilidad utilizada es la T-Student,

$$T = \frac{N(0,1)}{\sqrt{\frac{\chi_n^2}{n}}} \sim t_n$$





Supuestos

- *Población Normal, Varianza Poblacional Desconocida y Población Infinita*

Sea X una Población que se distribuye normalmente con Esperanza Matemática μ y Varianza Poblacional (desconocida) σ^2 respectivamente. Bajo estas condiciones se demuestra que la distribución de probabilidad utilizada es la T-Student,

$$\boldsymbol{T = \frac{\dfrac{\bar{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{\dfrac{(n-1)\hat{\sigma}^2}{\sigma^2}}{n-1}}} = \frac{\bar{X}-\mu}{\dfrac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1}}$$

- *Población Normal, Varianza Poblacional Desconocida y Población Finita*

$$\boldsymbol{T = \frac{\bar{X}-\mu}{\dfrac{\hat{\sigma}}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}} \sim t_{n-1}}$$

Parámetro: Proporción Poblacional $\boldsymbol{p}$ y Estimador: Proporción Muestral $\widehat{\boldsymbol{p}}$

- *Población Normal y Infinita*

$$Z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \longrightarrow N(0,1)$$

- *Población Normal y Finita*

$$Z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)}} \longrightarrow N(0,1)$$

POBLACIÓN NORMAL	Población Infinita	Población Finita
Media Poblacional μ σ^2 *conocida*	$Z=\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0{,}1)$	$Z=\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}\sim N(0{,}1)$
Media Poblacional μ σ^2 *desconocida*	$T=\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}\sim t_{n-1}$	$T=\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}\sim t_{n-1}$
Varianza Poblacional σ^2	$\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2_{n-1}$	
Proporción Poblacional p	$Z=\dfrac{\bar{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\longrightarrow N(0{,}1)$	$Z=\dfrac{\bar{p}-p}{\sqrt{\frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)}}\longrightarrow N(0{,}1)$

Aplicación de la Fracción de Muestreo

Se denomina al cociente entre el tamaño de la muestra y el de la Población.

$$f_m = \frac{n}{N}$$

Mide la proporción del tamaño de la muestra con respecto al de la Población.

A fines prácticos se considerará que si la fracción de muestreo es a lo sumo el 10%, la población puede considerarse infinita y consecuentemente, el factor $\left(\frac{N-n}{N-1}\right)$ que se utiliza en la varianza de la media muestral y la proporción muestral se puede considerar que tiende a 1.

$$Si\ f_m \leq 0{,}10 \Rightarrow N \rightarrow \infty \Rightarrow \left(\frac{N-n}{N-1}\right) \rightarrow 1$$

Ejercicio n° 1 (Guía)

En la Facultad de Ciencias Económicas, la altura de los estudiantes se distribuye normalmente con una media de 174 cm y desvío 20 cm. Se toma un curso de 50 alumnos al azar:

¿Cuál es la probabilidad de que la altura promedio de la muestra sea inferior a 172 cm?

$$X = Altura\ de\ los\ Estudiantes\ de\ la\ FCE\ (en\ cms)$$

$$X \sim N(\mu = 174, \sigma = 20)$$

$$n = 50 \Rightarrow \bar{X} \sim N\left(174, \frac{20}{\sqrt{50}}\right)$$

$$P(\bar{X} \leq 172) = F[\bar{X} = 172] = F\left[Z = \frac{172 - 174}{\frac{20}{\sqrt{50}}}\right] = F[Z = -0{,}71] = 0{,}23885$$

Ejercicio n° 1 (Guía)

En la Facultad de Ciencias Económicas, la altura de los estudiantes se distribuye normalmente con una media de 174 cm y desvío 20 cm. Se toma un curso de 50 alumnos al azar:

¿De qué tamaño deberá ser la muestra si se quiere que esta probabilidad sea de 0,20?

$$X = Altura\ de\ los\ Estudiantes\ de\ la\ FCE\ (en\ cms)$$

$$X \sim N(\mu = 174, \sigma = 20)$$

$$n =? \qquad P(\bar{X} \leq 172) = F[Z = z] = 0{,}20 \Rightarrow z = -0{,}841621$$

$$-0{,}841621 = \frac{172 - 174}{\frac{20}{\sqrt{n}}} \Rightarrow n = \frac{(-0{,}841621)^2 20^2}{(172 - 174)^2} \cong 70{,}83 \cong 71$$

Ejercicio n° 3 (Guía)

Los precios de los artículos que vende un supermercado, se distribuyen normalmente con media US$ 4 y desvío US$ 0,75. Se toma al azar una muestra de 50 artículos. ¿Cuál es la probabilidad de que el precio promedio de los artículos de la muestra esté entre US$ 3,9 y US$ 4,2?

$$X = Precio\ de\ los\ artículos\ vendidos\ (en\ US\$)$$

$$X \sim N(\mu = 4, \sigma = 0{,}75)$$

$$n = 50 \Rightarrow \bar{X} \sim N\left(4, \frac{0{,}75}{\sqrt{50}}\right)$$

$$P(3{,}9 \leq \bar{X} \leq 4{,}2) = F[\bar{X} = 4{,}2] - F[\bar{X} = 3{,}9]$$

$$P(3{,}9 \leq \bar{X} \leq 4{,}2) = F\left[Z = \frac{4{,}2 - 4}{\frac{0{,}75}{\sqrt{50}}}\right] - F\left[Z = \frac{3{,}9 - 4}{\frac{0{,}75}{\sqrt{50}}}\right] = F[Z = 1{,}89] - F[Z = -0{,}94]$$

$$PP(3{,}9 \leq \bar{X} \leq 4{,}2) = 0{,}97062102 - 0{,}17360878 = 0{,}79701224$$

Próxima Clase

En la próxima semana veremos

- Intervalos de Confianza

Por tal motivo es necesario e indispensable que tengan en claro cada uno de los temas, antes de avanzar con los siguientes temas de la materia.

Recomendación

Leer al menos alguno de los siguientes capítulos

Capítulo 5 del libro de Canavos

Capítulo 5 del libro de Chao

Capítulo 6 del libro de Anderson

Consultar dudas en el foro del Campus Virtual


TO BE
CONTINUED....

Preguntas, Sugerencias y Comentarios

RDelRosso-ext@austral.edu.ar

¡Muchas Gracias!