ESTADÍSTICA INFERENCIAL

IV. DISTRIBUCIONES MUESTRALES

Wendy Plata Alarcón
wplata@espol.edu.ec

Guayaquil, mayo de 2017

Poblaciones y Muestras

Muestra Aleatoria de una Población Finita

Sea X una población finita y de tamaño N, que tiene una distribución f; una Muestra de tamaño n, tomada de X, es denominada Muestra Aleatoria, cuando y solo cuando, al tomarla, todo subconjunto de tamaño n de la Población X, tiene igual probabilidad de constituirla.

$$\text{P(constituir la muestra)} = \frac{1}{\binom{N}{n}} = \frac{(N-n)!n!}{N!}$$

Poblaciones y Muestras

- Si la población X es infinita, con Distribución o Densidad f, una **Muestra** de tamaño **n**, $\boldsymbol{X^T = (X_1 \ \ X_2 \ldots X_n)}$es **Aleatoria**, si las n variables $\boldsymbol{X_1, X_2, \ldots, X_n}$ que la constituyen son ***Independientes e Idénticamente Distribuidas*** *(iid)*, lo cual significa que por independencia, la distribución conjunta f_x es:

$$\boldsymbol{f_X(X) = f_X(X_1 \, X_2 \ldots X_n) = f_1(X_1) f_2(X_2) \ldots \ f_n(X_n)}$$

$$\boldsymbol{= \prod_{i=1}^{n} f_i(X_i)}$$

Marco Muestral

- Representación simbólica de la Población Objetivo; deberá ser exhaustiva y actualizada.

Valores esperados de la Media Aritmética

$$\mathrm{E}[\bar{X}] = \mathrm{E}[X] = \mu$$

- **Población Infinita**

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

- **Población Finita**

$$Var(\bar{X}) = \left[\frac{N-n}{N-1}\right]\frac{\sigma^2}{n}$$

- $\frac{N-n}{N-1}$ se denomina Factor de Corrección para Poblaciones Finitas

Distribución Muestral de la Media Aritmética

$$\bar{X} = R^n \rightarrow R \;\; tal\ que,$$

$$\bar{X}(X_1\ X_2\ \ldots X_n) = \frac{S_n}{n}$$

$$E[\bar{X}] = E\left[\frac{S_n}{n}\right] = \frac{1}{n}E(S_n) = \frac{1}{n}n\mu = \mu$$

$$\text{Var}[\bar{X}] = Var\left[\frac{S_n}{n}\right] = \frac{1}{n^2}Var(S_n) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Teorema del Límite Central (TLC)

- Sea X una Población con Media $\boldsymbol{\mu}$ y Varianza $\boldsymbol{\sigma}^2$ ambas finitas, de la que se toma una Muestra Aleatoria $\boldsymbol{X}^T = (\boldsymbol{X}_1 \ \boldsymbol{X}_2 \ldots \boldsymbol{X}_n)$ de tamaño **n** siendo $\bar{\boldsymbol{X}}$ la Media Aritmética de la Muestra; bajo estas condiciones, la Variable Aleatoria $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ **"tiende"** a una Variable Normal Estándar, a medida que **n** tiende a infinito.

Condiciones para aplicar el TLC

- La población de la que se toma la muestra puede ser cualquiera.
- El tamaño de la muestra debe ser grande. ($n>=30$)
- La varianza debe ser conocida.

Distribución JI-CUADRADO

Sean $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$ variables independientes que siguen una distribución N(0,1).

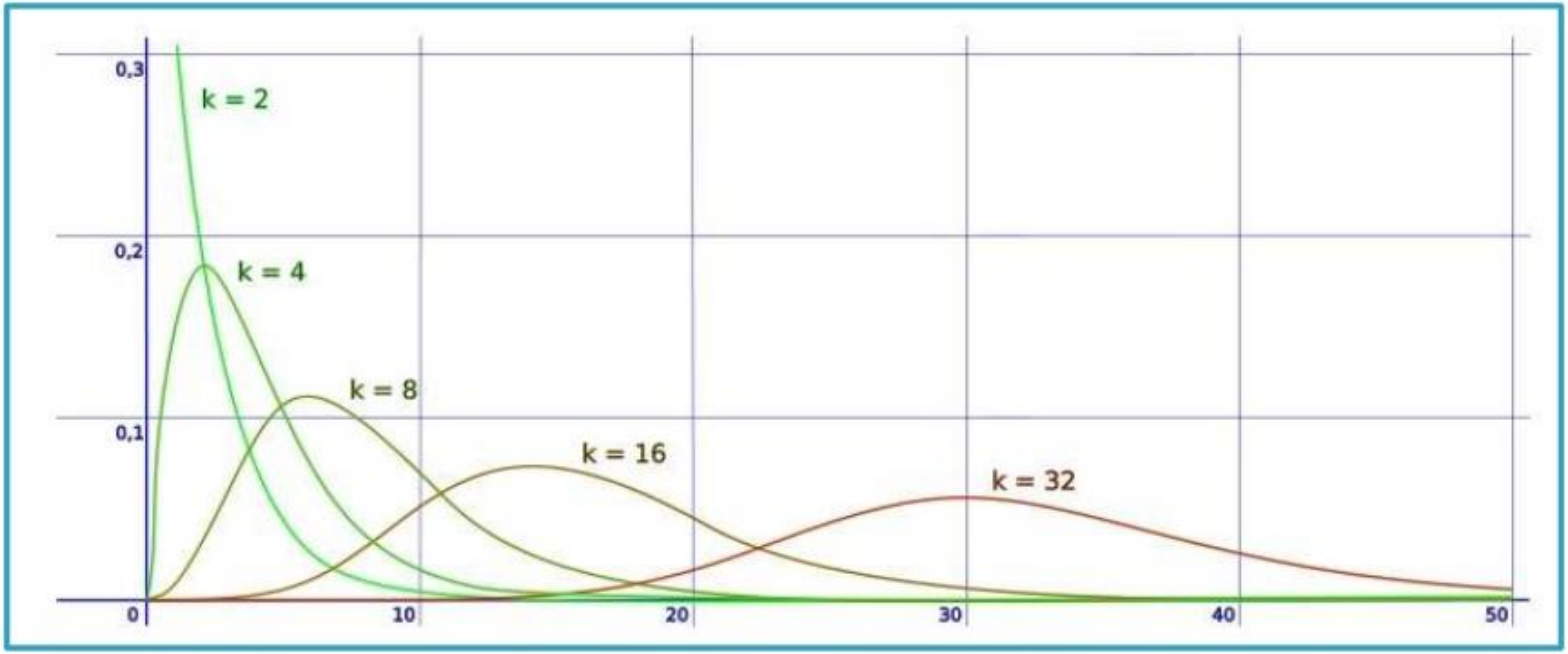
Sea ***X*** una variable definida como:

$$X = x_1^2 + x_2^2 + \ldots + x_n^2 = \sum_{i=1}^{n} x_i^2$$

En este caso, se dice que ***X*** se distribuye como una JI-CUADRADO con n grados de libertad:

$$\boldsymbol{X} \sim \chi^2(\mathrm{n})$$

Distribución JI-CUADRADO



Distribución F-FISHER

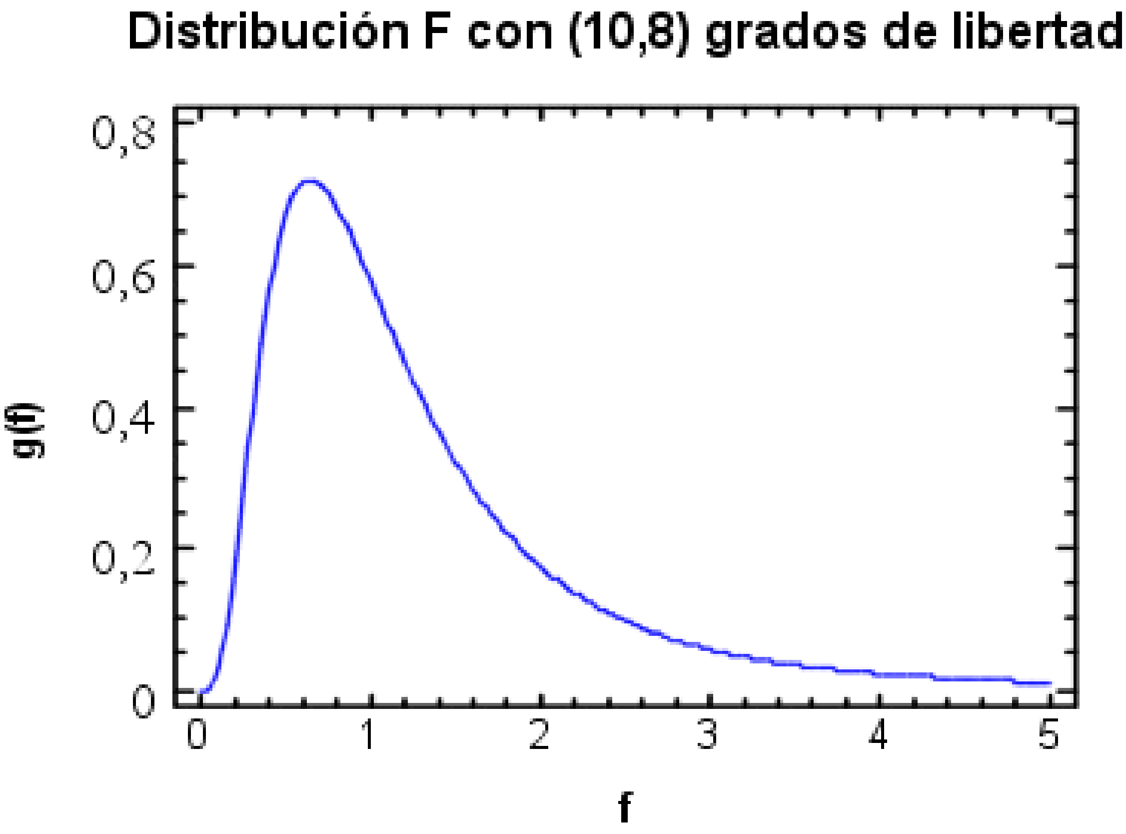
Sean U y V dos variables independientes, tal que:

$$\boldsymbol{U} \sim \chi^2(v_1) \qquad \text{y} \qquad \boldsymbol{V} \sim \chi^2(v_2)$$

Bajo estas condiciones, la variable F se define como:

$$F = \frac{\chi^2(v_1)/_{v_1}}{\chi^2(v_2)\,/_{v_2}}$$

Distribución F-FISHER



Estimación por Intervalos

Definición 4.4.2 Un **intervalo de confianza** del $(1-\alpha)100\,\%$ para el parámetro θ

$$\hat{\theta}_L < \theta < \hat{\theta}_U$$

es el valor que toma el estimador por intervalo de θ

$$\widehat{\Theta}_L < \theta < \widehat{\Theta}_U$$

para una probabilidad $1-\alpha$ dada.

Ejemplo: ¿Qué es preferible? Saber con un nivel de confianza del 90% que un parámetro se encuentra entre 2 y 4 o saber con un nivel de confianza del 99,99% que el mismo parámetro se encuentra entre -10000 y 10000?

Intervalo de confianza para la Media de muestras grandes

- Si se cumplen las condiciones del TLC:

$$\overline{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- Si la muestra es grande y la varianza desconocida, se la estima con la varianza muestral.

$$\overline{x} - Z_{\alpha/2} \cdot \frac{\mathbf{s}}{\sqrt{n}} < \mu < \overline{x} + Z_{\alpha/2} \cdot \frac{\mathbf{s}}{\sqrt{n}}$$

Media para muestras pequeñas

La muestra ha sido tomada de una población Normal

$$\overline{x} - t_{\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}}$$

Teoremas para Población Normal

Teorema 7.1

Sea $\boldsymbol{X}^{\boldsymbol{T}} = (\boldsymbol{X_1}\ \ \boldsymbol{X_2} \dots \boldsymbol{X_n})$ una Muestra Aleatoria de tamaño n con Media Aritmética $\overline{\boldsymbol{X}}$ y Varianza s^2, que es tomada de la población X que es Normal, X ~ N(μ, σ^2), bajo estas condiciones:

a) $\overline{\boldsymbol{X}} \sim N(\mu, \sigma^2/n)$; y,

b) $$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

Teoremas para Población Normal

Teorema 7.2

Sea $\boldsymbol{X}^{\boldsymbol{T}} = (\boldsymbol{X}_{\boldsymbol{1}}\ \ \boldsymbol{X}_{\boldsymbol{2}} \ldots \boldsymbol{X}_{\boldsymbol{n}})$ una Muestra Aleatoria de tamaño n con Media Aritmética $\bar{\boldsymbol{X}}$ y Varianza s^2, que es tomada de la población X que es Normal, X $\sim$ N(μ, σ^2), bajo estas condiciones la variable:

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim T_{(n-1)}$$

Referencias

- **Zurita, G.** (2010), "*Probabilidad y Estadística: Fundamentos y Aplicaciones*", Segunda Edición, Escuela Superior Politécnica del Litoral, Instituto de Ciencias Matemáticas, Guayaquil-Ecuador.