

## Tema 4

# Muestreo y Estimación

### 4.1 Conceptos generales

Se llama **población** o **universo** al conjunto de todos los individuos o elementos objeto de estudio.

Una **muestra** es un subconjunto, más o menos representativo, de la población. Se llama **tamaño** de la muestra al número de elementos que la forman.

Se denomina **muestreo** al proceso por el cual extraemos una muestra de una población. Hay distintos tipos de muestreo que se adaptan a distintas situaciones experimentales. En éste y en los sucesivos temas trabajaremos bajo la hipótesis de que las muestras se recogen mediante un procedimiento de ‘muestreo aleatorio simple’.

Una **muestra aleatoria simple de tamaño  $n$**  se consigue eligiendo  $n$  elementos de la población al azar y sin reposición, esto es, se elige uno, de los que quedan se elige otro, y así hasta obtener los  $n$  elementos que forman la muestra.

#### 4.1.1 Función de densidad o probabilidad conjunta

Sea  $X$  una variable aleatoria que representa la medida de la característica bajo estudio y  $f_X(x; \theta)$  su función de densidad. Si se repite el experimento  $n$  veces en las mismas condiciones, las  $X_1, X_2, \dots, X_n$  medidas de  $X$  serán independientes y constituyen una muestra aleatoria simple de tamaño  $n$  procedente de la variable  $X$ .

Consideradas conjuntamente, las variables  $X_1, X_2, \dots, X_n$  forman una variable aleatoria  $n$ -dimensional,  $(X_1, X_2, \dots, X_n)$ , cuyas componentes siguen todas la misma distribución que  $X$  y por tanto, si  $X$  es una variable aleatoria continua,

$$f_{X_1}(x; \theta) = f_{X_2}(x; \theta) = \dots = f_{X_n}(x; \theta) = f_X(x; \theta)$$

con lo cual su función de densidad conjunta es

$$f(x_1, x_2, \dots, x_n; \theta) = f_X(x_1; \theta) \cdot f_X(x_2; \theta) \cdots f_X(x_n; \theta)$$

Si  $X$  es una variable discreta entonces cada una de las variables que forman la muestra tendrá asociada una función de probabilidad idéntica a la de  $X$  y podremos hablar de la función de probabilidad conjunta de  $(X_1, X_2, \dots, X_n)$  que vendrá definida como

$$P_X[x_1, \dots, x_n; \theta] = P[X_1 = x_1; \dots; X_n = x_n; \theta] = P[X = x_1; \theta] \cdots P[X = x_n; \theta]$$

## 4.2 Estimación puntual

**Definición 4.2.1.**— Dada una variable aleatoria  $X$ , que mide una característica en una población, un parámetro es una caracterización numérica de la distribución de la población que determina total o parcialmente la función de densidad (o de probabilidad) de la variable aleatoria  $X$ .

**Ejemplo 4.2.2.**— Si  $X \sim \mathcal{P}(\lambda)$ , la distribución está determinada cuando se conoce el valor de  $\lambda$ .

Si  $X \sim N(\mu, \sigma^2)$ , es necesario conocer los valores de  $\mu$  y de  $\sigma$  para determinar la distribución.

**Definición 4.2.3.**— Al conjunto de valores que puede tomar un parámetro  $\theta$  se le llama **espacio paramétrico** y lo denotaremos por  $\Theta$ .

**Ejemplo 4.2.4.**—

- Si  $X \sim \mathcal{P}(\lambda)$  entonces  $\Theta = \mathbb{R}^+$ .
- Si  $X \sim N(\mu, \sigma^2)$  entonces  $\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$ .

### 4.2.1 Estadísticos y estimadores.

**Definición 4.2.5.**— Sea  $X$  una variable aleatoria cuya función de densidad o de probabilidad es  $f(x, \theta)$  y  $X_1, \dots, X_n$  una m.a.s. Un estadístico es una función  $T(X_1, \dots, X_n)$  que no depende del parámetro  $\theta$ .

**Ejemplo 4.2.6.**— Si  $X \sim N(\mu, \sigma^2)$  con  $\mu$  y  $\sigma$  desconocidas, las siguientes funciones son estadísticos:

$$T_1(X_1, \dots, X_n) = \frac{X_1 + \dots + X_n}{n}$$

$$T_2(X_1, \dots, X_n) = \frac{X_1^2 + \dots + X_n^2}{n}$$

$$T_3(X_1, \dots, X_n) = \max\{X_1, \dots, X_n\}$$

Sin embargo,

$$T_4(X_1, \dots, X_n) = \frac{X_1 + \dots + X_n}{n\sigma}$$

no es un estadístico.

Un estadístico es función de la muestra, y por tanto es una variable aleatoria.

**Definición 4.2.7.**— Un estadístico  $T(X_1, \dots, X_n)$  se dice que es un estimador de  $\theta$  si toma valores en el espacio paramétrico de  $\theta$ . Para una realización  $x_1, \dots, x_n$  de una muestra, el valor de  $T(x_1, \dots, x_n)$  se llama estimación de  $\theta$ .

**Ejemplo 4.2.8.**— Sea  $X$  una variable aleatoria con  $E[X] = \mu$  y  $\text{Var}(X) = \sigma^2$  y sea  $X_1, \dots, X_n$  es una m.a.s. de  $X$ . Entonces:

- $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  es un estimador de  $\mu$  y se denomina **media muestral**.

- $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$  es un estimador de  $\sigma^2$  y se denomina **varianza muestral**.
- $S_c^2 = \frac{n}{n-1} S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$  también es un estimador de  $\sigma^2$  y se denomina **cuasivarianza muestral**.

Se verifica que  $(n-1)S_c^2 = nS^2$  y  $S^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$ .

#### 4.2.2 Propiedades deseables en los estimadores

Entre las propiedades deseables de un buen estimador se encuentran las siguientes: **Carencia de sesgo, Consistencia, Eficiencia y Suficiencia**

##### • Carencia de Sesgo

Un buen estimador debería, en promedio, determinar el verdadero valor del parámetro.

**Definición 4.2.9.**— Se dice que  $T$  es un **estimador insesgado** de  $\theta$  si  $E[T] = \theta$ . En otro caso se dice que  $T$  es un **estimador sesgado** y se llama **sesgo** de  $T$  a  $|E[T] - \theta|$ .

**Ejemplo 4.2.10.**— Sea  $X$  una variable aleatoria de la que sabemos que  $E[X] = \mu$  y  $Var(X) = \sigma^2$ . Entonces:

- $\bar{X}$  es un estimador insesgado de  $\mu$ .
- $S^2$  es un estimador sesgado de  $\sigma^2$ .
- $S_c^2$  es un estimador insesgado de  $\sigma^2$ .

##### • Consistencia

Es deseable que a medida que aumente el tamaño de la muestra mejoren las estimaciones obtenidas.

**Definición 4.2.11.**— Diremos que  $T$  es un **estimador consistente** de cara a estimar el parámetro  $\theta$  si

$$\forall \epsilon > 0 \quad \lim_{n \rightarrow \infty} P(|T - \theta| < \epsilon) = 1$$

**Teorema 4.2.12.**— Sea  $T$  un estimador del parámetro  $\theta$ . Si se verifica que

$$\lim_{n \rightarrow \infty} E[T] = \theta \quad \text{y} \quad \lim_{n \rightarrow \infty} Var(T) = 0$$

entonces  $T$  es un estimador consistente para  $\theta$ .

**Ejemplo 4.2.13.**—

- Si  $X$  es una variable aleatoria tal que  $E(X) = \mu$ , entonces  $\bar{X}$  es un estimador consistente de  $\mu$ .
- Si  $X \sim N(\mu, \sigma^2)$  entonces  $S^2$  y  $S_c^2$  son estimadores consistentes de  $\sigma^2$ .

##### • Eficiencia

**Definición 4.2.14.**— Sean  $T_1$  y  $T_2$  dos estimadores insesgados de  $\theta$ . Diremos que  $T_1$  es más eficiente que  $T_2$  si  $Var(T_1) < Var(T_2)$ .

Entre los estimadores insesgados de  $\theta$ , será preferible tomar el de menor varianza.

### • Suficiencia

Un estadístico sintetiza en un número toda la información contenida en la muestra. Sería pues interesante que en ese proceso de síntesis no se perdiera la información que la muestra contiene acerca del valor del parámetro.

**Ejemplo 4.2.15.**— Si  $X \sim Be(p)$  y tomamos una muestra aleatoria simple de tamaño tres,  $X_1, X_2, X_3$  el conjunto de los posibles valores muestrales es:

$$\Lambda = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$$

Al aplicar el estadístico  $T(X_1, X_2, X_3) = X_1 + X_2 + X_3$  a cada una de estas muestras se obtiene el valor 0, 1, 2 ó 3. Puesto que  $p$  representa la probabilidad de éxito en cada realización de  $X$  parece razonable estimar  $p$  como la proporción de éxitos obtenidos en las  $n$  realizaciones del experimento, y para ello necesitamos conocer el número de éxitos que aparecen en la muestra. Si consideramos el anterior estadístico  $T = X_1 + X_2 + X_3$  podemos observar que si conocemos su valor sabremos inmediatamente el número de éxitos que hay en la muestra. En cambio esto no ocurre si consideramos el estadístico  $T_1 = X_1 \cdot X_2 \cdot X_3$ , por lo que no podríamos utilizarlo si lo que queremos hacer es estimar el valor de  $p$ . Esto se debe a que, mientras  $T$  conserva la información que la muestra posee acerca de  $p$ , no ocurre lo mismo con  $T_1$ .

La cuestión fundamental está en si, de cara a estimar un parámetro determinado, el estadístico con el que estemos trabajando conserva la información que la muestra contiene acerca de dicho parámetro. Esto se recoge en la siguiente definición.

**Definición 4.2.16.**— Diremos que un estadístico  $T(X_1, \dots, X_n)$  es **suficiente** de cara a estimar el parámetro  $\theta$  si la distribución conjunta de la muestra condicionada al valor del estadístico no depende de  $\theta$ . Esto es, si la variable de la que procede la muestra es discreta entonces

$$P(X_1 = x_1; X_2 = x_2; \dots; X_n = x_n | T=t)$$

no depende de  $\theta$  para ningún valor de  $t$  y si es continua entonces

$$f(x_1, \dots, x_n | T=t)$$

no depende de  $\theta$  para ningún valor de  $t$ .

**Teorema 4.2.17.**— Sea  $X$  una variable aleatoria continua (discreta) con función de densidad  $f(x, \theta)$  (o de probabilidad  $P[X = x, \theta]$ , si  $X$  es discreta) y  $X_1, \dots, X_n$  una muestra aleatoria simple procedente de  $X$ . El estadístico  $T = T(X_1, \dots, X_n)$  es un estadístico suficiente de cara a estimar el parámetro  $\theta$ , si la función de densidad (probabilidad) conjunta de la muestra se puede descomponer como:

$$f(X_1, \dots, X_n; \theta) = g(T, \theta) \cdot h(X_1, \dots, X_n)$$

$$\text{ó} \\ P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta] = g(T, \theta) \cdot h(X_1, X_2, \dots, X_n),$$

donde  $g$  depende de  $\theta$  y del estadístico  $T$ , y  $h$  depende sólo de la muestra o bien es una constante.

**Ejemplo 4.2.18.**— Si  $X_1, \dots, X_n$  es una muestra aleatoria simple de una variable  $X$ , entonces

- Si  $X \sim Be(p)$ , el estadístico  $T = \sum X_i$  es suficiente para estimar  $p$ .
- Si  $X \sim \mathcal{P}(\lambda)$ , entonces  $T(X_1, X_2, \dots, X_n) = \sum_{i=1}^n X_i$  es suficiente para estimar  $\lambda$ .
- Si  $X \sim N(0, \sigma^2)$ , entonces  $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i^2$  es suficiente para estimar  $\sigma^2$ .

### 4.2.3 Algunos estimadores usuales

La siguiente tabla recoge estimadores para los parámetros de algunas de las distribuciones más usuales.

Distribución	Estimador
$Be(p)$	$\hat{p} = \bar{X}$
$Ge(p)$	$\hat{p} = 1/\bar{X}$
$\mathcal{P}(\lambda)$	$\hat{\lambda} = \bar{X}$
$N(\mu, \sigma^2)$	$\hat{\mu} = \bar{X}$
	$\hat{\sigma}^2 = S^2$ (No insesgado)
	$\hat{\sigma}^2 = S_c^2$ (Inssegado)
$Exp(\lambda)$	$\hat{\lambda} = 1/\bar{X}$

### 4.3 Distribuciones asociadas al muestreo en poblaciones Normales

**Definición 4.3.1.**— Una variable aleatoria  $X$  se dice que sigue una *distribución normal* de parámetros  $\mu$  y  $\sigma^2$ ,  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$ , y se denota por  $X \sim N(\mu, \sigma^2)$ , si su f.d.d. es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \forall x \in \mathbb{R}$$

•**Propiedades:**

1.  $f$  es simétrica respecto a  $x = \mu$ .
2. Se verifica que si  $Z = \frac{X-\mu}{\sigma}$ , entonces  $Z \sim N(0, 1)$
3.  $E[X] = \mu$  y  $Var[X] = \sigma^2$
4.  $P[Z \leq -z] = 1 - P[Z \leq z]$

Se denota por  $z_{1-\alpha}$  al valor tal que  $P[Z \leq z_{1-\alpha}] = 1 - \alpha$

**Definición 4.3.2.**— Si  $X_1, \dots, X_n$  son variables aleatorias independientes idénticamente distribuidas según  $N(0, 1)$ , entonces  $X_1^2 + \dots + X_n^2$  sigue una **distribución Chi-cuadrado** con  $n$  grados de libertad. Lo denotaremos  $X \sim \chi_n^2$

•**Propiedades:**

1. Si  $X \sim \chi_n^2$  e  $Y \sim \chi_m^2$ ,  $X$  e  $Y$  independientes, entonces  $X + Y \sim \chi_{n+m}^2$
2.  $E[\chi_n^2] = n$  y  $Var[\chi_n^2] = 2n$

Se denota  $\chi_{n,1-\alpha}^2$  al valor tal que  $P[X \leq \chi_{1-\alpha,n}^2] = 1 - \alpha$

**Definición 4.3.3.**— Sean  $X \sim N(0,1)$  e  $Y \sim \chi_n^2$ ,  $X$  e  $Y$  independientes, entonces  $T = \frac{X}{\sqrt{Y/n}}$  sigue una **distribución T-Student** con  $n$  grados de libertad. Lo denotaremos  $T \sim t_n$

•**Propiedades:**

1.  $E[T] = 0$  y  $Var[T] = \frac{n}{n-2}$
2. Si  $T \sim t_n$ , entonces  $\lim_{n \rightarrow +\infty} T \sim N(0,1)$ . La aproximación es buena para  $n \geq 30$ .
3. Se denota por  $t_{n,1-\alpha}$  al valor tal que  $P[T \leq t_{n,1-\alpha}] = 1 - \alpha$ . Además,  $t_{n,1-\alpha} = -t_{n,\alpha}$ .

**Definición 4.3.4.**— Sean  $X \sim \chi_n^2$  e  $Y \sim \chi_m^2$ ,  $X$  e  $Y$  independientes, entonces  $F = \frac{X/n}{Y/m}$  sigue una **distribución F-Snedecor** con  $n$  y  $m$  grados de libertad. Lo denotaremos como  $F \sim \mathcal{F}_{n,m}$

•**Propiedades**

1.  $E[F] = \frac{m}{m-2}$  y  $Var[F] = \frac{m^2(2n+2m-4)}{n(m-2)^2(m-4)}$
2. Si  $F \sim \mathcal{F}_{n,m}$ , entonces  $\frac{1}{F} \sim \mathcal{F}_{m,n}$

Se denota por  $f_{n,m,1-\alpha}$  al valor tal que  $P[F \leq f_{n,m,1-\alpha}] = 1 - \alpha$ .

Se verifica que  $f_{n,m,\alpha} = 1/f_{m,n,1-\alpha}$ .

## 4.4 Estimación por regiones de confianza

Aunque bajo ciertas condiciones los estimadores puntuales gozan de buenas propiedades y proporcionan buenas estimaciones no nos permitirán obtener, en general, el valor exacto del parámetro a estimar.

Nuestro objetivo ahora es encontrar un intervalo aleatorio del que podamos afirmar, con una probabilidad prefijada, que contiene el verdadero valor del parámetro.

**Definición 4.4.1.**— Sea  $X$  una variable aleatoria cuya función de distribución depende de un parámetro  $\theta$ , y  $l_1(X_1, \dots, X_n)$  y  $l_2(X_1, \dots, X_n)$  dos estadísticos tales que  $l_1 \leq l_2$ . Un intervalo de la forma  $(l_1(X_1, \dots, X_n), l_2(X_1, \dots, X_n))$  se llama **intervalo aleatorio de límite inferior  $l_1$  y límite superior  $l_2$** .

**Definición 4.4.2.**— Un intervalo aleatorio  $(l_1(X_1, \dots, X_n), l_2(X_1, \dots, X_n))$  se llama **intervalo de confianza para  $\theta$  a nivel de confianza  $(1 - \alpha) \cdot 100\%$  con  $0 < \alpha < 1$  si**

$$P[l_1(X_1, \dots, X_n) < \theta < l_2(X_1, \dots, X_n)] = 1 - \alpha.$$

La siguiente tabla recoge la expresión de intervalos a nivel de confianza  $(1 - \alpha) \cdot 100\%$  para la media y la varianza de poblaciones normales y para la diferencia de medias y el cociente de varianzas de dos poblaciones normales independientes. En las expresiones de los intervalos de confianza para los parámetros de una población normal, el tamaño muestral se representa por  $n$ . En el caso de los intervalos asociados a los parámetros de dos poblaciones normales, los tamaños muestrales de ambas poblaciones se representan por  $n_x$  y  $n_y$ .

# Intervalos de confianza, en poblaciones normales, al $(1 - \alpha) \cdot 100\%$

Parámetro	Casos	Intervalo
$\mu$	$\sigma$ conocida	$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$
	$\sigma$ desconocida	$\bar{X} \pm \frac{S_c}{\sqrt{n}} \cdot t_{n-1, 1-\frac{\alpha}{2}}$
$\sigma^2$		$\left( \frac{(n-1)S_c^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{(n-1)S_c^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right)$ ó $\left( \frac{nS^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{nS^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right)$
$\mu_x - \mu_y$	$\sigma_x, \sigma_y$ conocidas	$(\bar{X} - \bar{Y}) \pm \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \cdot z_{1-\frac{\alpha}{2}}$
$\mu_x - \mu_y$	$\sigma_x = \sigma_y$ desconocidas *	$(\bar{X} - \bar{Y}) \pm \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \cdot S_p \cdot t_{n_x+n_y-2, 1-\frac{\alpha}{2}}$
$\mu_x - \mu_y$	$\sigma_x \neq \sigma_y$ desconocidas **	$(\bar{X} - \bar{Y}) \pm \sqrt{\frac{S_{c_x}^2}{n_x} + \frac{S_{c_y}^2}{n_y}} \cdot t_{g, 1-\alpha/2}$
$\sigma_y^2/\sigma_x^2$		$\left( \frac{S_{c_y}^2}{S_{c_x}^2} \frac{1}{f_{n_y-1, n_x-1, 1-\frac{\alpha}{2}}}, \frac{S_{c_y}^2}{S_{c_x}^2} f_{n_x-1, n_y-1, 1-\frac{\alpha}{2}} \right)$

$$* S_p^2 = \frac{(n_x-1)S_{c_x}^2 + (n_y-1)S_{c_y}^2}{n_x+n_y-2}$$

$$** g = \frac{\left( \frac{S_{c_x}^2}{n_x} + \frac{S_{c_y}^2}{n_y} \right)^2}{\frac{(S_{c_x}^2/n_x)^2}{n_x+1} + \frac{(S_{c_y}^2/n_y)^2}{n_y+1}} - 2$$

## 4.5 ANEXO. Distribuciones de funciones muestrales

### 4.5.1 Distribución de la media muestral

**Definición 4.5.1.**— Sea  $X_1, \dots, X_n$  una muestra aleatoria simple procedente de una variable aleatoria  $X$ . Se define la media muestral como:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

*Nota:*  $\bar{X}$  es una variable aleatoria, no una constante.

**Teorema 4.5.2.**— Si  $X_1, \dots, X_n$  es una muestra aleatoria simple procedente de una variable aleatoria  $X$  con media  $\mu$  y varianza  $\sigma^2$ , se tiene que:  $E[\bar{X}] = \mu$  y  $\text{Var}[\bar{X}] = \frac{\sigma^2}{n}$ .

Se define la desviación típica de la muestra (error estándar de la media) como  $S_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

**Teorema 4.5.3.**— Sea  $X_1, \dots, X_n$  una muestra aleatoria simple procedente de una variable aleatoria  $X \sim N(\mu, \sigma^2)$ . Entonces

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

En consecuencia:

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

**Teorema 4.5.4.**— Si  $X_1, \dots, X_n$  es una muestra aleatoria simple de una variable aleatoria  $X$ , de distribución no especificada, con media  $\mu$  y varianza  $\sigma^2$  finita, entonces:

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

En general, la aproximación es buena para  $n \geq 30$

### 4.5.2 Distribución de la varianza muestral

**Definición 4.5.5.**— Si  $X_1 \dots X_n$  es una muestra aleatoria simple procedente de una variable aleatoria  $X$ , se define la **varianza muestral** como:

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

y la **cuasivarianza muestral** como:

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

Se verifica que  $(n-1)S_c^2 = nS^2$  y  $S^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$

**Teorema 4.5.6.**— Sea  $X_1, \dots, X_n$  una muestra aleatoria simple procedente de una población  $N(\mu, \sigma^2)$  de media y varianzas desconocidas. Entonces:

1.  $\bar{X}$  y  $S_c^2$  son independientes.



2. Si  $Y = \frac{(n-1)S_c^2}{\sigma^2}$ , entonces  $Y \sim \chi_{n-1}^2$

3. Si  $Y = \frac{nS_c^2}{\sigma^2}$ , entonces  $Y \sim \chi_{n-1}^2$

**Proposición 4.5.7.**— Sea  $X_1, \dots, X_n$  una muestra aleatoria simple procedente de una población  $N(\mu, \sigma^2)$  de media y varianzas desconocidas. Entonces:

1.  $\frac{\bar{X} - \mu}{S_c} \sqrt{n} \sim t_{n-1}$

2.  $\frac{\bar{X} - \mu}{S} \sqrt{n-1} \sim t_{n-1}$

### 4.5.3 Distribución de la diferencia de medias muestrales

Sea  $X_1, \dots, X_{n_x}$  una muestra aleatoria simple procedente de una variable aleatoria  $X$  e  $Y_1, \dots, Y_{n_y}$  una muestra aleatoria simple procedente de una variable aleatoria  $Y$ , tales que  $X$  e  $Y$  son independientes y además  $X \sim N(\mu_x, \sigma_x^2)$  e  $Y \sim N(\mu_y, \sigma_y^2)$ . Entonces,

si  $\sigma_x^2 = \sigma_y^2 = \sigma^2$  conocida,

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim N(0, 1)$$

si  $\sigma_x^2 \neq \sigma_y^2$ , siendo ambas conocidas,

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \sim N(0, 1)$$

si  $\sigma_x^2 = \sigma_y^2$  pero ambas son desconocidas,

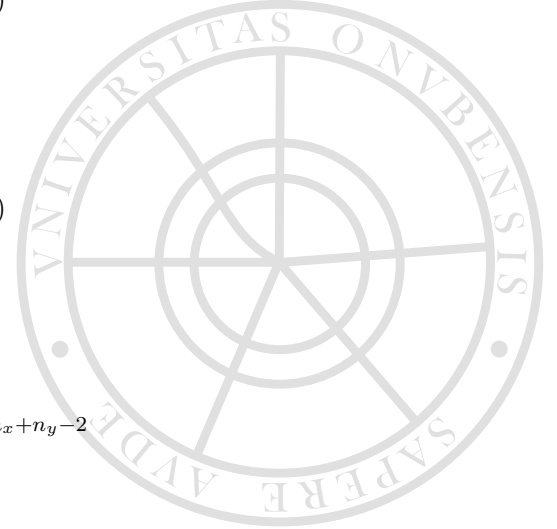
$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{(n_x-1)S_{c_x}^2 + (n_y-1)S_{c_y}^2}{n_x+n_y-2} \left( \frac{1}{n_x} + \frac{1}{n_y} \right)}} \sim t_{n_x+n_y-2}$$

si  $\sigma_x^2 \neq \sigma_y^2$ , ambas desconocidas,

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_{c_x}^2}{n_x} + \frac{S_{c_y}^2}{n_y}}} \sim t_g$$

siendo

$$g = \frac{\left( \frac{S_{c_x}^2}{n_x} + \frac{S_{c_y}^2}{n_y} \right)^2}{\frac{(S_{c_x}^2/n_x)^2}{n_x+1} + \frac{(S_{c_y}^2/n_y)^2}{n_y+1}} - 2$$



#### 4.5.4 Distribución del cociente de varianzas muestrales

Sea  $X_1, \dots, X_{n_x}$  una muestra aleatoria simple procedente de una variable aleatoria  $X$  e  $Y_1, \dots, Y_{n_y}$  una muestra aleatoria simple procedente de una variable aleatoria  $Y$ , tales que  $X$  e  $Y$  son independientes y además  $X \sim N(\mu_x, \sigma_x^2)$  e  $Y \sim N(\mu_y, \sigma_y^2)$ . Entonces:

$$\frac{S_{c_x}^2/\sigma_x^2}{S_{c_y}^2/\sigma_y^2} \sim \mathcal{F}_{n_x-1, n_y-1}$$

En el caso en que  $\sigma_x = \sigma_y$  entonces:

$$\frac{S_{c_x}^2}{S_{c_y}^2} \sim \mathcal{F}_{n_x-1, n_y-1}$$

