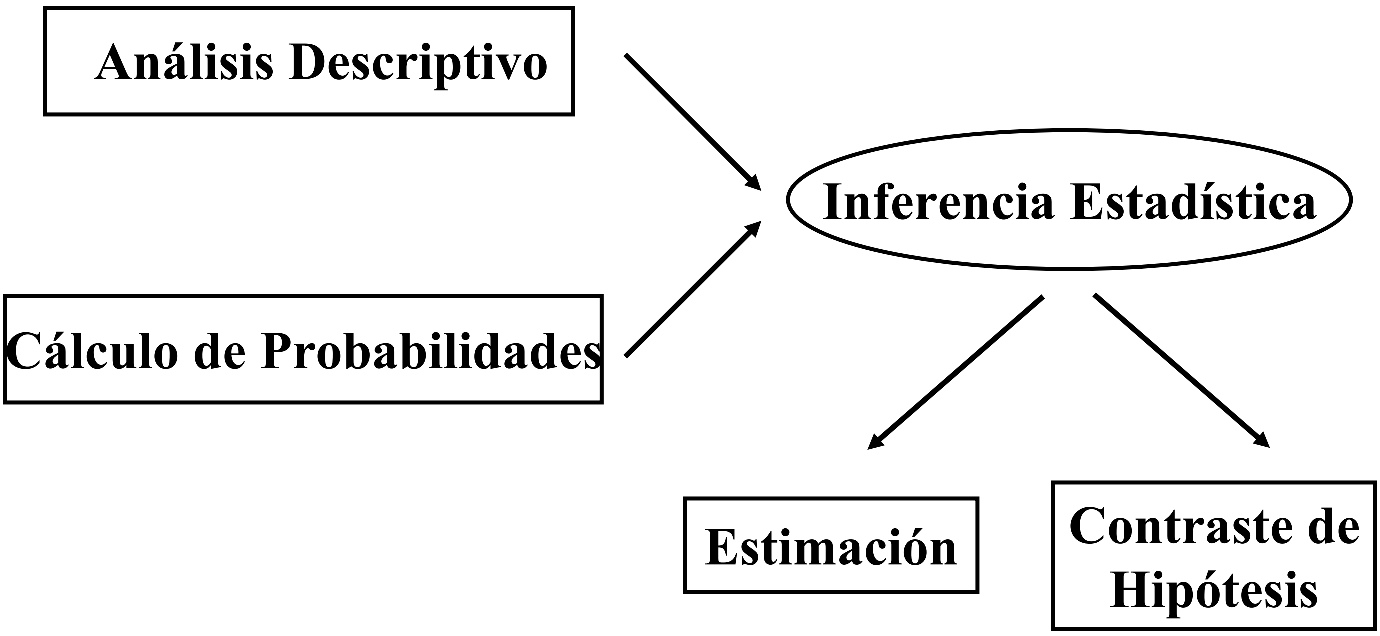
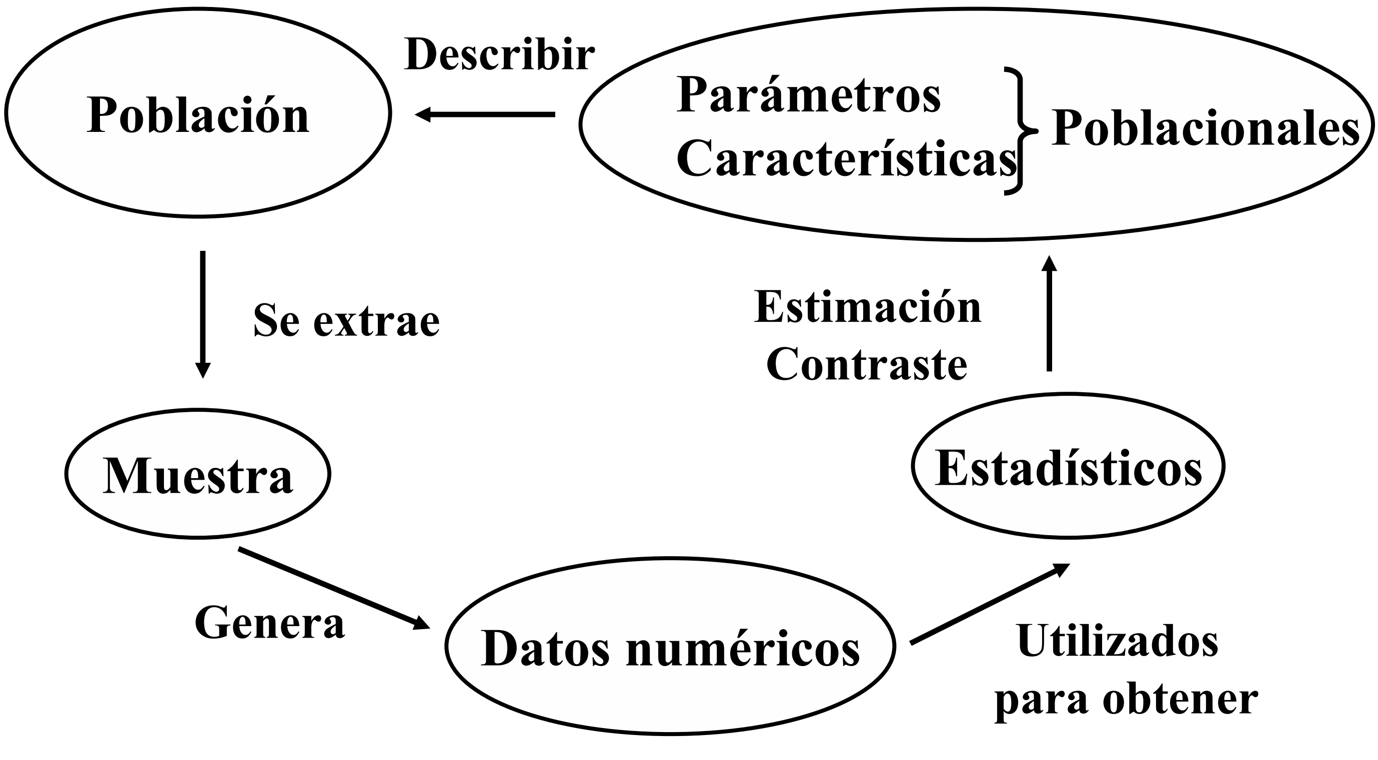
TEMA 6. INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

❖ 6.1. Introducción





❖ 6.2 Conceptos básicos

➢ **Población:** "Conjunto de elementos en los que se observa alguna característica común"

➢ **Observaciones:** "Valores que toma la característica observada en cada elemento de la población"

➢ **Parámetro: "**Característica numérica que describe una variable observada en la población"

➢ **Muestra:** "Conjunto de unidades representativas de una población"

➢ **Estadístico:** "Función de los valores de la muestra"

❑ La **inferencia estadística** esta basada en el estudio de las **muestras**

↓

❑ La muestra debe ser **representativa de la población** para extraer conclusiones validas sobre esta población

↓

❑ La muestra debe ser **aleatoria**

❖ 6.3 Muestreo aleatorio simple

➢ "Cada elemento de la población tiene la misma probabilidad de ser elegido para formar parte de la muestra y cada muestra del mismo tamaño tiene la misma probabilidad de ser seleccionada"

➢ **Muestra aleatoria simple de tamaño n:**

Sea una población donde observamos la variable aleatoria X.

Una muestra aleatoria simple, m.a.s., de tamaño n, es un conjunto de *n variables aleatorias* $X_1, X_2, ..., X_n$, que verifican:

$$X_1, X_2, \ldots, X_n \begin{cases} \text{Independientes entre sí} \\ \text{Cada } X_i \text{ con idénticas características que } X \end{cases}$$

❖ Muestreo aleatorio simple

◈ El muestreo aleatorio simple en poblaciones finitas se realiza "con reemplazamiento", es decir:

Se selecciona un elemento de la población al azar, se observa el valor de la variable aleatoria X, se devuelve a la población y se vuelve a seleccionar otro elemento. Así hasta obtener los n elementos. Este procedimiento garantiza la independencia de las observaciones

¤ La selección aleatoria de los elementos se realiza con una tabla de números aleatorios, o con algún procedimiento informático

❑ Pasos de un muestreo

Población en la que se observa la variable X

↓

Población

↓

Se decide extraer una muestra aleatoria simple de tamaño n, compuesta por las variables aleatorias $X_1, X_2, \ldots, X_n$

↓

Se seleccionan n elementos de la población

↓

Muestra

↓

Los elementos seleccionados generan n números

↓

$x_1, x_2, \ldots, x_n$, valores observados de *las variables aleatorias* $X_1, X_2, \ldots, X_n$

♦ Ejemplo en poblaciones finitas

▸ En un instituto se quiere realizar un estudio sobre el nivel de colesterol de los alumnos. Para ello, se decide extraer una muestra aleatoria simple de tamaño 10

✓ Población ⟶ Alumnos del instituto

✓ Variable aleatoria, X ⟶ Nivel de colesterol

✓ Muestra aleatoria simple, de tamaño 10 ⟶

Variables aleatorias $X_1, X_2, \ldots, X_{10}$

X_i, nivel de colesterol del i-ésimo alumno seleccionado

▪ Se seleccionan 10 alumnos y sus niveles de colesterol son: 129, 170, 135, 140, 225, 163, 131, 203, 187, 149

✓ Valores observados de las *variables aleatorias*

$$X_1, X_2, \ldots, X_{10}$$

$$\downarrow$$

$$x_1 = 129; \quad x_2 = 170; \quad x_3 = 135; \quad x_4 = 140; \quad x_5 = 225;$$
$$x_6 = 163; \quad x_7 = 131; \quad x_8 = 203; \quad x_9 = 187; \quad x_{10} = 149.$$

♦ Ejemplo en poblaciones infinitas

▸ Se analizan muestras de agua de un río para estudiar el índice de diversidad de especies. Este índice se utiliza para medir el efecto de una perturbación, como la contaminación del agua, en seres vivos. Puede determinarse la diversidad de la población antes y después de la perturbación. Si el índice tras la perturbación es mucho mas pequeño indica que la perturbación ha tenido efectos negativos. Para esto, se decide extraer una muestra aleatoria simple de tamaño 8

✓ Población ⟶ Posibles análisis del agua

✓ Variable aleatoria, X ⟶ Índice de diversidad

✓ Muestra aleatoria simple ⟶

Variables aleatorias $X_1, X_2, ..., X_8$

X_i :"*Índice de diversidad del i-ésimo análisis realizado*"

▪ Se realizan 8 análisis y sus índices de diversidad son:
1.92; 1.87; 1.35; 1.48; 2.13; 1.85; 2.07; 1.98

✓ Valores observados de las *variables aleatorias*

$$X_1, X_2, ..., X_8$$

$$\downarrow$$

$$x_1 = 1{,}92;\ x_2 = 1{,}87;\ x_3 = 1{,}35;\ x_4 = 1{,}48;$$
$$x_5 = 2{,}13;\ x_6 = 1{,}85,\ x_7 = 2{,}07;\ x_8 = 1{,}98$$

❖ 6.4 Distribuciones asociadas al muestreo

◈ 6.4.1 Distribución Chi-Cuadrado

➢ Sean n variables aleatorias, $X_1, X_2, \ldots X_n$, que verifican:

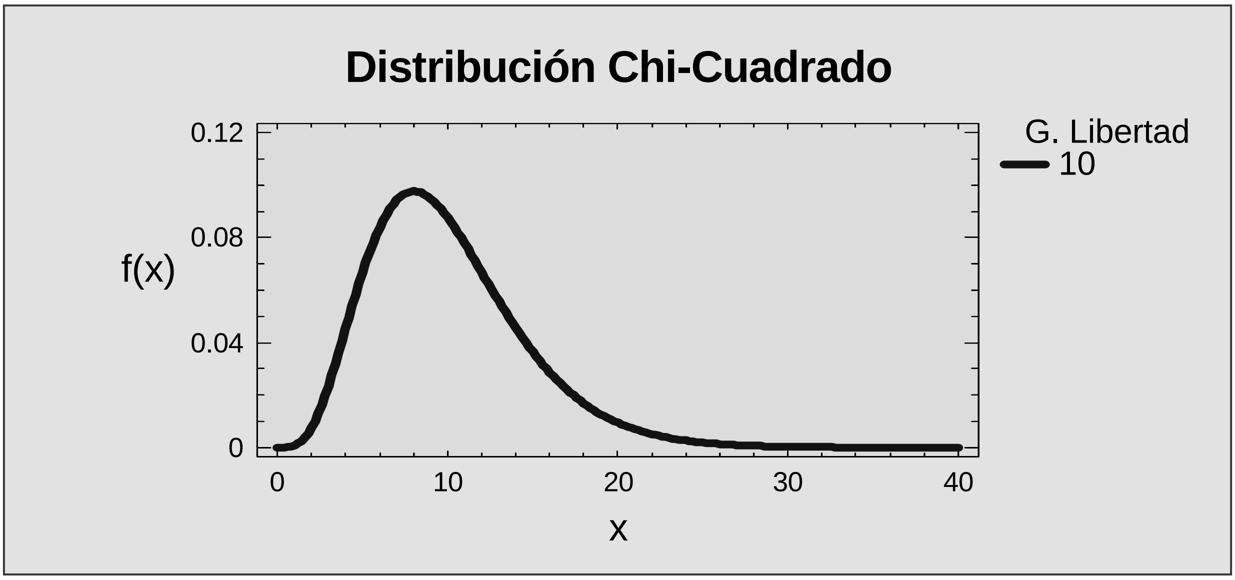
- Independientes entre sí
- $X_i \longrightarrow N(0; 1)$

➢ Definimos la variable aleatoria X como:

$$X = X_1^2 + X_2^2 + \ldots + X_n^2$$

❖ La variable aleatoria X sigue una distribución Chi-Cuadrado con n grados de libertad

$$X \longrightarrow \chi_n^2$$



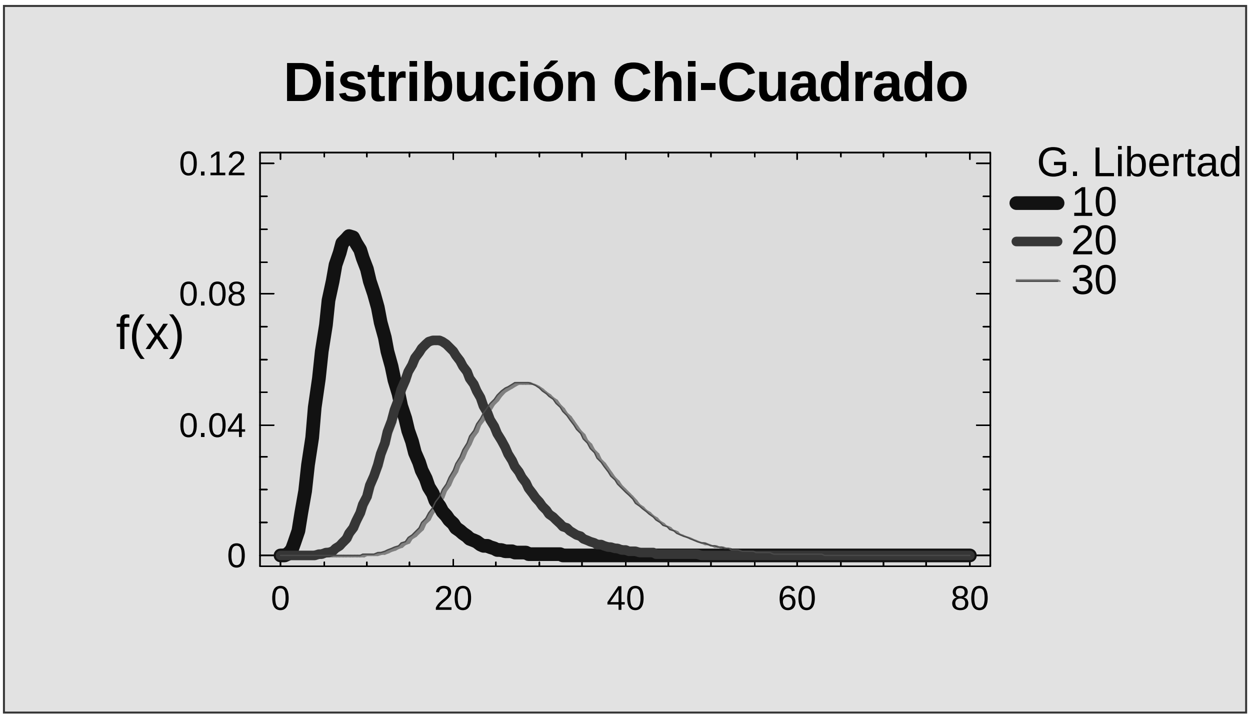
✓ Esperanza matemática

$$\boxed{\mathrm{E}\left[\chi_n^2\right] = n}$$

✓ Varianza

$$\boxed{\mathrm{Var}\left[\chi_n^2\right] = n}$$

✓ Para valores grandes de *n*, la distribución Chi-Cuadrado se aproxima a la distribución Normal. La aproximación se considera aceptable para *n* > 30



◈ 6.4.2 Distribución t de Student

➢ Sean las variables aleatorias, *Y* y *Z*, que verifican:

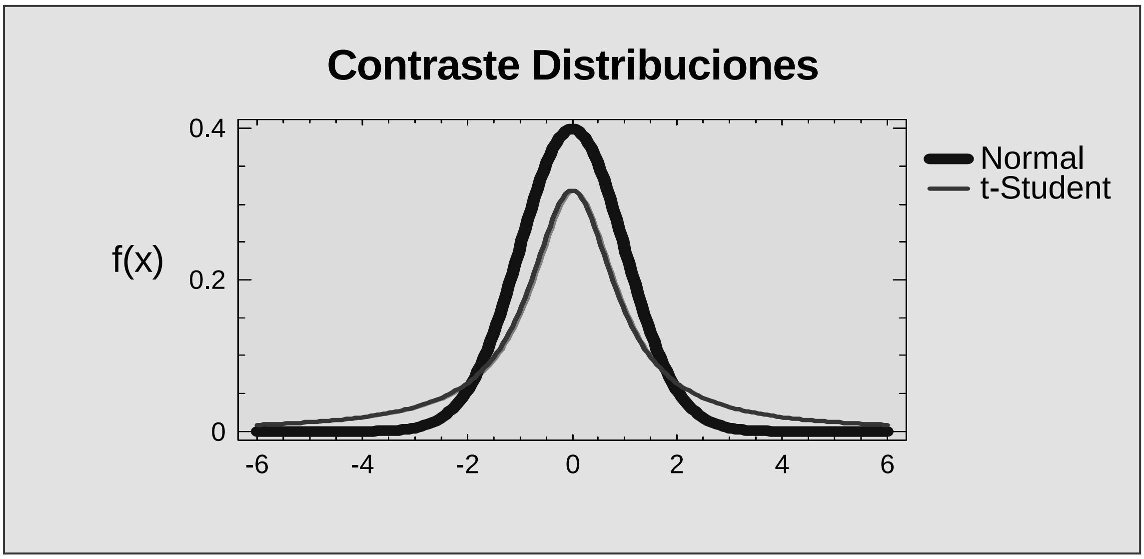
$$\left.\begin{array}{lll} \blacksquare\ Z & \longrightarrow & N(0;1) \\ \blacksquare\ Y & \longrightarrow & \chi_n^2 \end{array}\right\} \text{Independientes}$$

➢ Definimos la variable aleatoria *X* como:

$$X = \frac{Z}{\sqrt{Y/n}}$$

❖ La variable aleatoria *X* sigue una distribución *t* de Student con *n* grados de libertad

$$X \rightarrow t_n$$



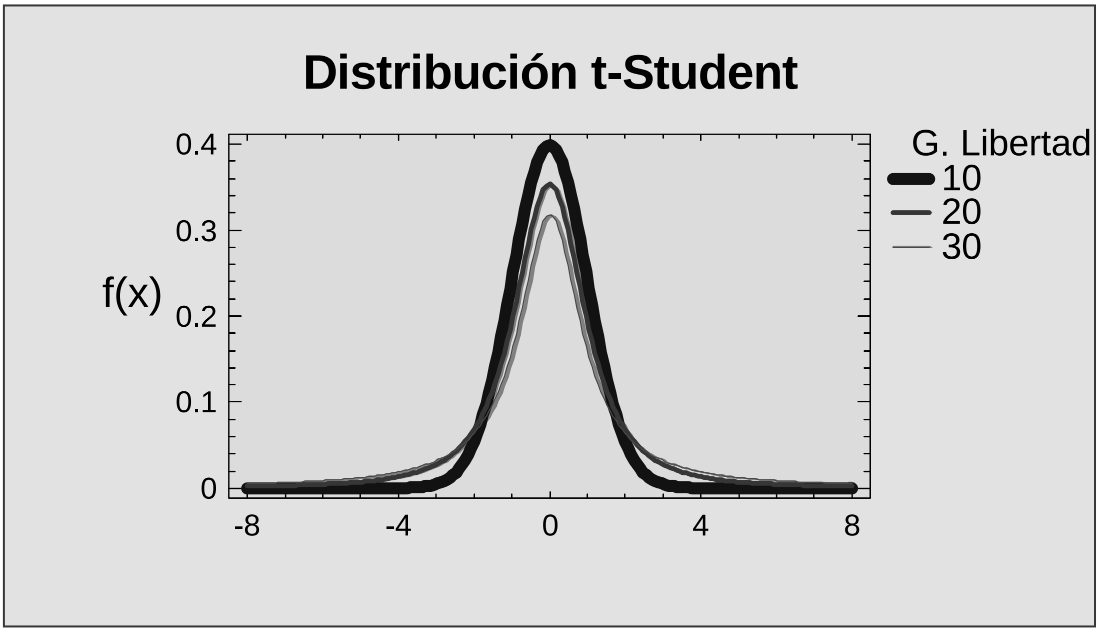
✓ Esperanza matemática

$$\boxed{\mathrm{E}\left[\,t_n\,\right] = 0}$$

✓ Varianza

$$\boxed{\mathrm{Var}\left[t_n\right] = \frac{n}{n-2}}$$

✓ Para valores grandes de *n*, la distribución *t* de Student se aproxima a la distribución Normal. La aproximación se considera aceptable para *n* > 30



◈ 6.4.3 Distribución F de Snedecor

➢ Sean las variables aleatorias, *Y* y *W*, que verifican:

- Y $\longrightarrow \chi^2_n$
- W $\longrightarrow \chi^2_m$

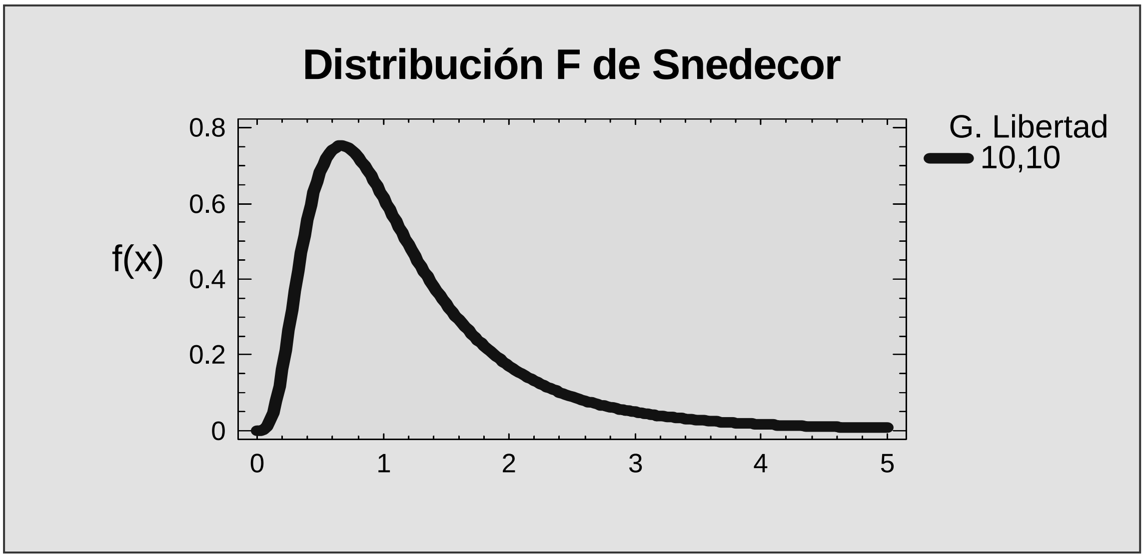
$\left.\begin{array}{l} Y \longrightarrow \chi^2_n \\ W \longrightarrow \chi^2_m \end{array}\right\}$ Independientes

➢ Definimos la variable aleatoria *X* como:

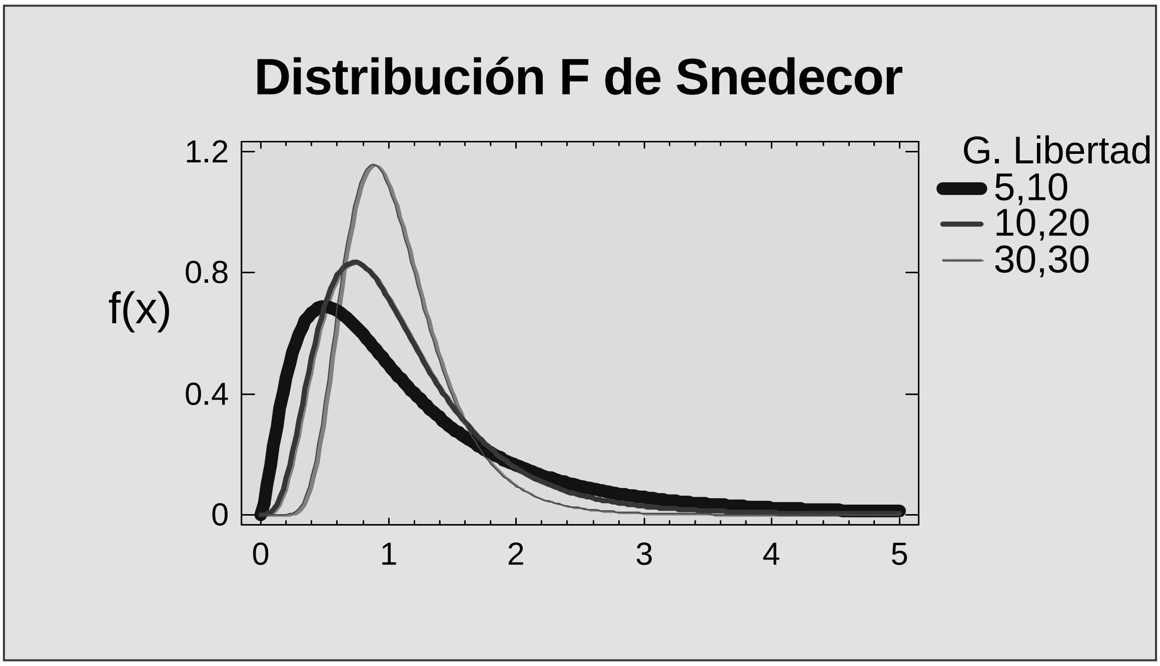
$$X = \frac{Y/n}{W/m}$$

❖ La variable aleatoria *X* sigue una distribución *F* de Snedecor con *n* y *m* grados de libertad

$$X \longrightarrow F_{n,\, m}$$



✓ Para valores grandes de n y m, la distribución F de Snedecor se aproxima a la distribución Normal.



❖ 6.5 Distribución de estadísticos muestrales

◈ 6.5.1 Concepto de estadístico y distribución muestral

➢ **Estadístico:** "Una función de los valores de la muestra". Es una variable aleatoria, cuyos valores dependen de la muestra seleccionada. Su distribución de probabilidad, se conoce como "*Distribución muestral del estadístico*"

➢ Sea una población donde se observa la variable aleatoria *X*. Esta variable *X*, tendrá una distribución de probabilidad, que puede ser conocida o desconocida, y ciertas características o parámetros poblacionales

Estadísticos muestrales

⇓

Inferencia

⇓

Parámetros poblacionales

➢ Sea una población donde se observa la variable aleatoria X

$$\mathrm{E}[X] = \mu; \qquad \mathrm{Var}[X] = \sigma^2$$

➢ Consideramos una muestra aleatoria simple, m.a.s., de tamaño n, formada por las v.a. $X_1, X_2, \ldots X_n$

$$X_1, X_2, \ldots, X_n \begin{cases} \text{Independie ntes entre sí} \\ \mathrm{E}[X] = \mu \\ \mathrm{Var}[X] = \sigma^2 \end{cases}$$

➢ Definimos los siguientes estadísticos muestrales:

✓ Media muestral:

$$\boxed{\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}}$$

✓ Varianza muestral:

$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}_0\right)^2}{n}}$$

✓ Cuasi-Varianza muestral:

$$\boxed{S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}}$$

➢ Consideramos todas las posibles muestras de tamaño *n*

Muestra 1	Muestra 2		Muestra j
X_1^1	X_1^2		X_1^j
X_2^1	X_2^2		X_2^j
$\vdots$	$\vdots$		$\vdots$
X_i^1	X_i^2		X_i^j
$\vdots$	$\vdots$		$\vdots$
X_n^1	X_n^2		X_n^j
$\bar{x}_1$	$\bar{x}_2$	…	$\bar{x}_j$ …

❖ La variable aleatoria $\overline{X}$ toma los valores: $\bar{x}_1, \bar{x}_2, .., \bar{x}_j ..$

✓ Su distribución de probabilidad

"Distribución de la media muestral"

✓ Esperanza matemática: $E\left[\overline{X}\right] = \mu_{\overline{X}}$

✓ Varianza: $Var\left[\overline{X}\right] = \sigma^2_{\overline{X}}$

➢ Los estadísticos muestrales, media, varianza y cuasi-varianza verifican las siguientes propiedades:

✓ Media muestral:

$$E\left[\overline{X}\right] = \mu_{\overline{X}} = \mu$$

$$Var\left[\overline{X}\right] = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

✓ Varianza muestral:

$$E\left[\hat{\sigma}^2\right] = \frac{n-1}{n}\sigma^2$$

✓ Cuasivarianza muestral:

$$E\left[S^2\right] = \sigma^2$$

❑ **Estas propiedades se verifican siempre, cualquiera que sea la distribución de la variable *X***

♦ Ejemplo en poblaciones infinitas

▸ Sea una v.a. X con valores: 1, 3, 5. Consideramos una m.a.s. de tamaño 2. Obtener:

1.- Media y varianza de la v.a. X

2.- Media y varianza de la v.a. $\overline{\mathrm{X}}$

1.-

X	$P(X)$
1	1 / 3
3	1 / 3
5	1 / 3

$$\boxed{\mu = E[X] = 3 \qquad \sigma^2 = \frac{8}{3}}$$

$$\mu = E[X] = 1\times\frac{1}{3} + 3\times\frac{1}{3} + 5\times\frac{1}{3} = \frac{9}{3} = 3$$

$$\sigma^2 = E\left[X^2\right] - E[X]^2 =$$

$$= 1^2\times\frac{1}{3} + 3^2\times\frac{1}{3} + 5^2\times\frac{1}{3} - 3^2 = \frac{8}{3}$$

2.-

x_1	x_2	$\bar{x}$
1	1	1
1	3	2
1	5	3
3	1	2
3	3	3
3	5	4
5	1	3
5	3	4
5	5	5

$\bar{X}$	$P\left(\bar{X}\right)$
1	$1/9$
2	$2/9$
3	$3/9$
4	$2/9$
5	$1/9$

$$\boxed{\begin{gathered} E\left[\bar{X}\right] = 3 = \mu \\ Var\left[\bar{X}\right] = \frac{4}{3} = \frac{8/3}{2} = \frac{\sigma^2}{n} \end{gathered}}$$

$$\mu_{\bar{X}} = E\left[\bar{X}\right] = 1 \times \frac{1}{9} + 2 \times \frac{2}{9} + \ldots + 5 \times \frac{1}{9} = 3$$

$$Var\ [\ \bar{X}\ ] = \sigma_{\bar{X}}^2 = E\left[\bar{X}^2\right] - E\left[\bar{X}\right]^2 =$$

$$= 1^2 \times \frac{1}{9} + \ldots + 5^2 \times \frac{1}{9} - 3^2 = \frac{4}{3}$$

◈ 6.5.2. Distribución de la media muestral de una población Normal

➢ Sea una población donde se observa la variable aleatoria X. Supongamos que $X \longrightarrow N(\mu, \sigma)$

➢ Consideramos una muestra aleatoria simple, m.a.s., de tamaño *n*, formada por las v.a., $X_1, X_2, \ldots, X_n$

$$X_1, X_2, \ldots, X_n \begin{cases} \text{Independientes entre si} \\ X_i \longrightarrow N(\mu, \sigma) \end{cases}$$

❖ Distribución de la media muestral

♦ Caso A. Varianza poblacional, σ^2, conocida

♦ Caso B. Varianza poblacional, σ^2, desconocida

♦ Caso C. Varianza poblacional, σ^2, desconocida. Muestras grandes

❖ Distribución de la media muestral

♦ Caso A. Varianza poblacional, σ^2, conocida

- La variable aleatoria media muestral:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Tiene distribución Normal

$$\boxed{\overline{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

- Por lo tanto

$$\boxed{Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \rightarrow N(0;1)}$$

♦ Caso B. Varianza poblacional, σ^2, desconocida

▪ El estadístico *T*, definido como: $$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

tiene una distribución *t* de Student con *n – 1* g. l.

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \rightarrow t_{n-1}$$

♦ Caso C. Varianza poblacional, σ^2, desconocida. Muestras grandes, *n* > 30

▪ El estadístico *T*, definido como: $$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

tiene una distribución Normal, $T \longrightarrow N(0\,;\,1)$

❖ Teorema Central del Limite

- Sea $X_1, X_2, \ldots, X_n$, una m.a.s., de tamaño *n* de una población con distribución de probabilidad no especificada, con media μ y desviación típica σ

- La variable aleatoria *Z*, definida como:

$$\boxed{Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}}$$

tiene una distribución, aproximadamente, $N(0, 1)$

✓ La aproximación es aceptable para $n > 30$

♦ Ejemplo: Distribución de la media muestral Varianza poblacional conocida

▸ Se está estudiando el tiempo transcurrido entre la polinización y la fertilización, *X*, en una especie de coníferas. Supongamos que la variable *X* está normalmente distribuida con una media de 6 meses y una desviación típica de 2 meses. Consideramos una m.a.s. de tamaño 25.

Obtener la probabilidad de que el tiempo medio transcurrido en la muestra entre la polinización y la fertilización sea como máximo de 6,3 meses

$$X : \text{"Tiempo transcurrido"} \rightarrow N(\mu;\sigma) = N(6;2)$$

$$Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X}-6}{\frac{2}{\sqrt{25}}} = \frac{\overline{X}-6}{0.4} \rightarrow N(0;\ 1)$$

$$P(\overline{X} \leq 6.3) = P\left(\frac{\overline{X}-6}{0.4} \leq \frac{6.3-6}{0.4}\right) = P(Z \leq 0.75) =$$

$$= 1 - P(Z \geq 0.75) = 1 - 0.2266 = 0.7734$$

♦ Ejemplo: Distribución de la media muestral Varianza poblacional desconocida

▸ Se está realizando un estudio sobre la calidad del aire en una zona. Uno de los indicadores de la calidad del aire es el número medio de microgramos de partículas en suspensión por metro cúbico. Supongamos que la variable *X*: "Número de microgramos de partículas", está normalmente distribuida.

Se hacen 16 mediciones, en las que se obtiene una cuasidesviación típica de 10.8585 unidades. Obtener la probabilidad de que la media muestral no difiera de la media poblacional en más de 8 unidades.

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{\overline{X} - \mu}{14/\sqrt{16}} = \frac{\overline{X} - \mu}{3.5} \rightarrow t_{n-1} = t_{15}$$

$$P\left(\left| \overline{X} - \mu \right| \le 8 \right) = P\left(-8 \le \overline{X} - \mu \le 8 \right) =$$

$$= P\left(\frac{-8}{\frac{10.8585}{\sqrt{16}}} \le \overline{X} - \mu \le \frac{8}{\frac{10.8585}{\sqrt{16}}} \right) = P(-2.947 \le t_{15} \le 2.947) =$$

$$= 1 - 2P(t_{15} \ge 2.947) = 1 - 2 \times 0.005 = 1 - 0.01 = 0.99$$

2. Se hacen 36 mediciones en las que se obtiene una cuasidesviación típica de 12 unidades. Obtener la probabilidad de que la media muestral no difiera de la media poblacional en más de 5 unidades.

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}} = \frac{\overline{X} - \mu}{12 / \sqrt{36}} = \frac{\overline{X} - \mu}{2} \rightarrow t_{35} \cong N(0;1)$$

$$P\left(\left|\overline{X} - \mu\right| \leq 5\right) = P\left(-5 \leq \overline{X} - \mu \leq 5\right) =$$

$$= P\left(\frac{-5}{2} \leq \frac{\overline{X} - 5}{2} \leq \frac{5}{2}\right) = P(-2.5 \leq Z \leq 2.5) =$$

$$= 1 - 2 \times 0.00621 = 0.98758$$

♦ Ejemplo: Teorema central del límite

▸ Supongamos que el nº de barriles de petróleo que produce un pozo al día es una v.a. con distribución no especificada. Si se observa la producción en 64 días y se sabe que la desviación típica del nº de barriles por día es 16, obtener la probabilidad de que la media muestral se encuentre a no más de 4 barriles del verdadero valor de la producción media diaria

$$\left.\begin{array}{l} X_1 : \text{" Nº de barriles el día 1"} \\ X_2 : \text{" Nº de barriles el día 2"} \\ \qquad\vdots \\ X_i : \text{" Nº de barriles el día } i\text{"} \\ \qquad\vdots \\ X_{64} : \text{" Nº de barriles el día 64"} \end{array}\right\} \sigma_{X_i} = 16$$

$$n = 64 > 30 \Rightarrow \overline{X} = \frac{\sum X_i}{n} \rightarrow N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

$$\overline{X} \rightarrow N\left(\mu; \frac{16}{\sqrt{64}}\right) = N(\mu; 2)$$

$$\overline{X} \to N\left(\mu;\, \frac{\sigma}{\sqrt{n}}\right) = N\left(\mu;\, \frac{16}{\sqrt{64}}\right) = N(\mu;\, 2)$$

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{\overline{X} - \mu}{2} \to N(0;\, 1)$$

$$P\left(\left|\overline{X} - \mu\right| \le 4\right) = P\left(-4 \le \overline{X} - \mu \le 4\right) =$$

$$= P\left(\frac{-4}{2} \le \frac{\overline{X} - \mu}{2} \le \frac{4}{2}\right) = P\left(-2 \le Z \le 2\right) =$$

$$= 1 - 2P\left(Z \ge 2\right) = 1 - 2 \times 0.0228 = 0.9544$$

◈ 6.5.3. Distribución de la varianza muestral de una población Normal

➢ Sea una población donde se observa la variable aleatoria X. Supongamos que $X \longrightarrow N(\mu, \sigma)$

➢Consideramos una muestra aleatoria simple, m.a.s., de tamaño *n*, formada por las v.a., $X_1, X_2, ..., X_n$

$$X_1, X_2, \ldots, X_n \begin{cases} \text{Independientes entre si} \\ X_{\mathbf{i}} \longrightarrow N(\mu, \sigma) \end{cases}$$

❖ Distribución de la varianza muestral

♦ Caso A. Media poblacional, μ, conocida (*)

♦ Caso B. Varianza poblacional, μ, desconocida

(*) Este caso no se incluye en los contenidos del curso

❖ Distribución de la varianza muestral

♦ Media poblacional, μ, desconocida

▪ El estadístico χ^2, definido como:

$$\chi^2 = \frac{n\hat{\sigma}^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2}$$

tiene una distribución Chi-Cuadrado con *n* – 1 grados de libertad

$$\boxed{\chi^2 = \frac{n\hat{\sigma}^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \rightarrow \chi^2_{n-1}}$$

♦ Ejemplo: Distribución de la varianza muestral

▸ Se considera una medición física realizada con un instrumento de precisión, donde el interés se centra en la variabilidad de la lectura. Se sabe que la medición es una v.a. con distribución Normal y desviación típica 4 unidades. Se toma una m.a.s. de tamaño 25.

Obtener la probabilidad de que el valor de la varianza muestral sea mayor de 12.16 unidades cuadradas.

$$X_i : \text{"Medición"} \rightarrow N(\mu; 4)$$

$$n = 25$$

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

$$P\left(\hat{\sigma}^2 \geq 12.16\right) = P\left(\frac{n\hat{\sigma}^2}{\sigma^2} \geq \frac{n\,12.16}{\sigma^2}\right) =$$

$$= P\left(\chi^2_{n-1} \geq \frac{25 \times 12.16}{16}\right) = P\left(\chi^2_{n-1} \geq 19\right) = 0.75$$

◈6.5.4. Distribución de la diferencia de medias muestrales de dos poblaciones Normales independientes

➢ Sean las variables aleatorias X e Y tales que

$$\left.\begin{array}{l} X \longrightarrow N(\mu_X, \sigma_X) \\ Y \longrightarrow N(\mu_Y, \sigma_Y) \end{array}\right\} \text{ Independientes}$$

Consideramos:

$$\left.\begin{array}{c} \text{m.a.s. de tamaño } n_X \text{ de } X \\ X_1, X_2, \ldots, X_{n_X} \end{array}\right\} \quad \overline{X},\ S_X^2$$

$$\left.\begin{array}{c} \text{m.a.s. de tamaño } n_Y \text{ de } Y \\ Y_1, Y_2, \ldots, Y_{n_Y} \end{array}\right\} \quad \overline{Y},\ S_Y^2$$

$$\overline{X} = \frac{1}{n_X} \sum_{i=1}^{n_X} X_i \qquad S_X^2 = \frac{1}{n_X - 1} \sum_{i=1}^{n_X} \left(X_i - \overline{X}\right)^2$$

$$\overline{Y} = \frac{1}{n_Y} \sum_{i=1}^{n_Y} Y_i \qquad S_Y^2 = \frac{1}{n_Y - 1} \sum_{i=1}^{n_Y} \left(Y_i - \overline{Y}\right)^2$$

❖ Distribución de la diferencia de medias

♦ **Caso A. Varianzas poblacionales conocidas**

♦ **Caso B. Varianzas poblacionales desconocidas, pero iguales**

♦ **Caso C. Varianzas poblacionales desconocidas, distintas o no, con $n_X, n_Y > 30$**

❖ Distribución de la diferencia de medias

♦ Caso A. Varianzas poblacionales conocidas

- La variable aleatoria, $\overline{X} - \overline{Y}$, tiene distribución Normal

$$N\left((\mu_X - \mu_Y), \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \right)$$

- Por lo tanto

$$\boxed{Z = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \rightarrow N(0;1)}$$

♦ Caso B. Varianzas poblacionales desconocidas, pero iguales

❖ $\sigma_X^2 = \sigma_Y^2$

▪ El estadístico *T*, definido como:

$$T = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

donde:

$$S_p^2 = \frac{(n_X - 1)\, S_X^2 + (n_Y - 1)\, S_Y^2}{n_X + n_Y - 2}$$

tiene una distribución *t* de Student con $n_X + n_Y - 2$ grados de libertad

$$T = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \rightarrow t_{n_X + n_Y - 2}$$

♦ Caso C. Varianzas poblacionales desconocidas distintas o no, con n_X, $n_Y > 30$

▪ El estadístico Z, definido como:

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}}$$

tiene distribución Normal

$$\boxed{Z = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \to \mathrm{N}\left(0;1\right)}$$

♦ Ejemplo: Distribución de la diferencia de medias Varianzas poblacionales conocidas

▸ Los niveles de radiación latente en dos regiones A y B siguen distribuciones Normales independientes de medias 0.48 y 0.4663 y varianzas 0.2 y 0.01 rem por año, respectivamente. Se realizan 25 mediciones en la región A y 100 en la B. Obtener la probabilidad de que la media de la muestra A sea como máximo 0.2 rem superior a la media de la muestra B.

$$X : \text{"Nivel radiación latente en A"}$$
$$Y : \text{"Nivel radiación latente en B"}$$

$$X \to N\left(0.48; \sqrt{0.2}\right); \; n_X = 25$$
$$Y \to N\left(0.4663; \sqrt{0.01}\right); \; n_Y = 100$$

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}} \to N(0;1)$$

$$Z=\frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}}\to N(0;1)$$

$$P\left(\overline{X}\le\overline{Y}+0.2\right)=P\left(\overline{X}-\overline{Y}\le 0.2\right)=$$

$$=P\left(\frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}}\le\frac{0.2-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}}\right)=$$

$$=P\left(Z\le\frac{0.2-0.0137}{\sqrt{0.008+0.0001}}\right)=$$

$$=P\left(Z\le 2.07\right)=1-P\left(Z\ge 2.07\right)=$$

$$=1-0.0192=0.9808$$

♦ Ejemplo: Distribución de la diferencia de medias. Varianzas poblacionales desconocidas, pero iguales

▸ Se está realizando un estudio sobre la calidad del aire en dos zonas A y B. Un indicador de la calidad es el número de microgr. de partículas en suspensión por m^3 de aire, que suponemos siguen distribuciones Normales independientes de media 62.237 en A, 61.022 en B y varianzas iguales. En la zona A se realizan 12 mediciones, obteniéndose una cuasi-varianza de 8.44 $microgr^2$ y en la B 15 mediciones, con una cuasi-varianza de 9.44 $microgr^2$. Obtener la probabilidad de que la media muestral de A sea como mínimo tres unidades superior a la media muestral de B.

$$X : \text{"Calidad del aire en A"}; \quad X \to N(62.237; \sigma)$$
$$Y : \text{"Calidad del aire en B"}; \quad Y \to N(61.022; \sigma)$$

$$n_X = 12; \qquad s_X^2 = 8.44$$
$$n_Y = 15; \qquad s_Y^2 = 9.44$$

$$X: \text{"Calidad del aire en A"}; \quad X \to N(62.237; \sigma)$$
$$Y: \text{"Calidad del aire en B"}; \quad Y \to N(61.022; \sigma)$$

$$n_X = 12; \qquad s_X^2 = 8.44$$
$$n_Y = 15; \qquad s_Y^2 = 9.44$$

$$S_p^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{(n_X + n_Y - 2)} = 9$$

$$P(\overline{X} - \overline{Y} \geq 3) =$$

$$= P\left(\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \geq \frac{3 - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \right) =$$

$$= P\left(t_{25} \geq \frac{3 - 1.015}{3\sqrt{\frac{1}{12} + \frac{1}{15}}} \right) = P(t_{25} \geq 1.708) = 0.05$$

♦ Ejemplo: Distribución de la diferencia de medias Varianzas poblacionales desconocidas. Muestras grandes

▸ Se estudia el efecto de un vertido tóxico en un río, comparando el índice de biodiversidad I.B-D. antes y después del vertido.

Supongamos que los I.B-D. siguen distribuciones Normales. Antes del vertido se habían realizado 35 pruebas y se obtuvo una media de 1.9 y una cuasi-desviación típica de 0.4. Después del vertido se realizan 40 pruebas y se obtiene una media de 1.7 y una cuasi-desviación típica de 0.7.

Obtener la probabilidad de que la media poblacional antes del vertido sea como máximo 0.5 unidades inferior a la media poblacional después del vertido.

$$X: \text{"I.B-D antes del vertido"} \rightarrow N(\mu_X ; \sigma_X)$$

$$Y: \text{"I.B-D después del vertido"} \rightarrow N(\mu_Y ; \sigma_Y)$$

$$n_X = 35; \quad \overline{X} = 1.9; \quad S_X = 0.4$$

$$n_Y = 40 \quad \overline{Y} = 1.7; \quad S_Y = 0.7$$

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_Y\right)}{\sqrt{\dfrac{S_X^2}{n_X} + \dfrac{S_Y^2}{n_Y}}} \quad \rightarrow \ N(0;1)$$

$$X:\text{"I.B - D antes del vertido"} \rightarrow N(\mu_X ; \sigma_X)$$
$$Y:\text{"I.B - D después del vertido"} \rightarrow N(\mu_Y ; \sigma_Y)$$

$$n_X = 35; \quad \overline{X} = 1.9; \quad S_X = 0.4$$
$$n_Y = 40 \quad \overline{Y} = 1.7; \quad S_Y = 0.7$$

$$P(\overline{X} - \overline{Y} \leq 0.2) =$$

$$= P\left(\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \leq \frac{0.5 - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \right) =$$

$$= P\left(Z \leq \frac{0.5 - (1.9 - 1.7)}{\sqrt{\frac{0.4^2}{35} + \frac{0.7^2}{40}}} \right) =$$

$$= P(Z \leq 2.313) = 1 - P(Z \geq 2.313) =$$

$$= 1 - 0.0104 = 0.9896$$

◈6.5.5. Distribución del cociente de varianzas muestrales de dos poblaciones Normales independientes

➢ Sean las variables aleatorias X e Y tales que

$$\left.\begin{array}{l} X \longrightarrow N(\mu_X, \sigma_X) \\ Y \longrightarrow N(\mu_Y, \sigma_Y) \end{array}\right\} \text{Independientes}$$

Consideramos:

$$\left.\begin{array}{c} \text{m.a.s. de tamaño } n_X \text{ de } X \\ X_1, X_2, \ldots, X_{n_X} \end{array}\right\} \bar{X},\ S_X^2$$

$$\left.\begin{array}{c} \text{m.a.s. de tamaño } n_Y \text{ de } Y \\ Y_1, Y_2, \ldots, Y_{n_Y} \end{array}\right\} \bar{Y},\ S_Y^2$$

$$\bar{X} = \frac{1}{n_X}\sum_{i=1}^{n_X} X_i \qquad S_X^2 = \frac{1}{n_X - 1}\sum_{i=1}^{n_X}\left(X_i - \bar{X}\right)^2$$

$$\bar{Y} = \frac{1}{n_Y}\sum_{j=1}^{n_Y} Y_j \qquad S_Y^2 = \frac{1}{n_Y - 1}\sum_{j=1}^{n_Y}\left(Y_j - \bar{Y}\right)^2$$

❖ Distribución del cociente de varianzas muestrales

▪ El estadístico *F*, definido como:

$$F = \frac{S_X^2 \big/ \sigma_X^2}{S_Y^2 \big/ \sigma_Y^2} = \frac{S_X^2 \times \sigma_Y^2}{S_Y^2 \times \sigma_X^2}$$

tiene una distribución F de Snedecor con $n_X - 1, n_Y - 1$, grados de libertad

$$\boxed{F = \frac{S_X^2 \times \sigma_Y^2}{S_Y^2 \times \sigma_X^2} \to F_{n_X - 1,\, n_Y - 1}}$$

♦ Ejemplo: Distribución del cociente de varianzas muestrales

▸ Se está comparando la variabilidad de los I.B-D de dos ríos A y B, que suponemos siguen distribuciones Normales. Se realizan 16 mediciones en el río A y se obtiene una cuasi-varianza de 9.52, y 18 mediciones en el río B y se obtiene una cuasivarianza de 7.

Obtener la probabilidad de que la varianza en el río B sea como mínimo el doble de la varianza en el río A.

$$X : \text{"I.B - D en el río A"} \rightarrow N(\mu_X ; \sigma_X)$$
$$Y : \text{"I.B - D en el río B"} \rightarrow N(\mu_Y ; \sigma_Y)$$

$$F = \frac{S_X^2 \times \sigma_Y^2}{S_Y^2 \times \sigma_X^2} \rightarrow F_{n_X - 1,\, n_Y - 1}$$

$$P\left(\sigma_Y^2 \geq 2\sigma_X^2\right) = P\left(\frac{\sigma_Y^2}{\sigma_X^2} \geq 2\right) = P\left(\frac{S_X^2 \sigma_Y^2}{S_Y^2 \sigma_X^2} \geq 2\frac{S_X^2}{S_Y^2}\right) =$$

$$= P\left(F_{15,17} \geq 2 \times \frac{9.52}{7}\right) = P\left(F_{15,17} \geq 2.72\right) = 0.025$$

◈ 6.5.6. Distribución de la proporción muestral

➢ Consideramos una variable aleatoria $X \longrightarrow \mathrm{B}(n;\ p)$, donde “*p*” es la proporción de “éxitos” en la población

➢ Para tamaños grandes de *n*, $n > 30$, la distribución Binomial se aproxima a una distribución Normal :

$$X \to N\left(np;\ \sqrt{npq}\right)$$

▪ Definimos el estadístico proporción muestral como:

$$\boxed{\hat{p} = \frac{X}{n}}$$

❖ Distribución de la proporción muestral

- El estadístico proporción muestral :

$$\boxed{\hat{p} = \frac{X}{n}}$$

- Verifica que:

$$\boxed{\hat{p} \rightarrow N\left(p; \sqrt{\frac{pq}{n}}\right)}$$

Por lo tanto:

$$\boxed{Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \rightarrow N(0; 1)}$$

♦ Ejemplo: Distribución de la proporción muestral

▸ Se quiere probar una terapia de grupo para dejar de fumar. Para ello se toma una m.a.s. de 50 fumadores. Se sabe que las personas que llevan al menos 10 años fumando tienen más dificultades para dejar de fumar, y que el 38% de los fumadores llevan al menos 10 años fumando. Por ello, se decide separar unos de otros si entre los fumadores elegidos más de un 19% llevan más de 10 años fumando. Obtener la probabilidad de que se decida separarlos.

p :"Proporción de fumadores con ≥ 10 años, en la población

$\hat{p}$:"Proporción de fumadores con ≥ 10 años, en la muestra

$$\hat{p} \to N\left(p; \sqrt{\frac{pq}{n}}\right) = N\left(0.38; \sqrt{\frac{0.38 \times 0.62}{50}}\right) = N(0.38; 0.068)$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{\hat{p} - 0.38}{0.068} \to N(0;1)$$

$$P(\hat{p} \geq 0.19) = P\left(\frac{\hat{p} - 0.38}{0.0686} \geq \frac{0.19 - 0.38}{0.0686}\right) = P(Z \geq -2.769) =$$

$$= 1 - P(Z \leq -2.769) = 1 - P(Z \geq 2.769) = 1 - 0.0028 = 0.9972$$

◈ 6.5.7. Distribución de la diferencia de proporciones muestrales

➢ Sean las variables aleatorias X e Y tales que

$$\left.\begin{aligned} X &\to B(n_X\,;\,p_X) \\ Y &\to B(n_Y\,;\,p_Y) \end{aligned}\right\} \text{ Independientes}$$

Para n_X y n_Y grandes, se verifica:

$$X \longrightarrow N\left(n_X\, p_X\,;\,\sqrt{n_X\, p_X\, q_X}\right)$$

$$Y \longrightarrow N\left(n_Y\, p_Y\,;\,\sqrt{n_Y\, p_Y\, q_Y}\right)$$

▪ Definimos las proporciones muestrales como:

$$\boxed{\begin{aligned} \hat{p}_X &= \frac{X}{n_X} \\ \hat{p}_Y &= \frac{Y}{n_Y} \end{aligned}}$$

❖ Distribución de la diferencia de proporciones muestrales

▪ Definimos el estadístico diferencia de proporciones muestrales:

$$\hat{p}_X - \hat{p}_Y \,; \text{donde} : \begin{cases} \hat{p}_X = \dfrac{X}{n_X} \\ \\ \hat{p}_Y = \dfrac{Y}{n_Y} \end{cases}$$

▪ Se verifica que:

$$\boxed{Z = \frac{(\hat{p}_X - \hat{p}_Y) - (p_X - p_Y)}{\sqrt{\dfrac{p_X q_X}{n_X} + \dfrac{p_Y q_Y}{n_Y}}} \rightarrow N(0;1)}$$

♦ Ejemplo: Distribución de la diferencia de proporciones muestrales

▸ Se sabe que en una población el 28% de las mujeres y el 25% de los hombres son fumadores. Se extraen muestras de 42 mujeres y 40 hombres. Determinar la probabilidad de que las mujeres fumadoras superen a los hombres fumadores en al menos el 4%.

p_X: “Proporción de mujeres fumadoras en la población

p_Y: “Proporción de mujeres fumadoras en la población

$\hat{p}_X$:“Proporción de mujeres fumadoras en la muestra

$\hat{p}_Y$: “Proporción de mujeres fumadoras en la muestra

$$Z=\frac{(\hat{p}_X-\hat{p}_Y)-(p_X-p_Y)}{\sqrt{\frac{p_Xq_X}{n_X}+\frac{p_Yq_Y}{n_Y}}}=\frac{(\hat{p}_X-\hat{p}_Y)-(0.28-0.25)}{\sqrt{\frac{0.28\times 0.72}{42}+\frac{0.25\times 0.75}{40}}}$$

$$=\frac{(\hat{p}_X-\hat{p}_Y)-0.03}{0.0974}\to N(0;1)$$

$$P(\hat{p}_X\geq\hat{p}_Y+0.04)=P(\hat{p}_X-\hat{p}_Y\geq 0.04)=$$

$$=P\left(\frac{(\hat{p}_X-\hat{p}_Y)-0.03}{0.0974}\geq\frac{0.04-0.03}{0.0974}\right)=$$

$$=P\left(Z\geq\frac{0.04-0.03}{0.974}\right)=P(Z\geq 0.0103)=0.4602$$