TEMA 8

INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA
DISTRIBUCIONES EN EL MUESTREO

EJEMPLOS

VNIVERSITAT DE VALÈNCIA
OpenCourseWare

PROGRAMACIÓN.

TEMA 8.- INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

1. Inferencia Estadística. Población y muestra.
2. Tipos de muestreo. Muestreo aleatorio.
3. Estadísticos y distribuciones asociadas.
3.1. **Distribución en el muestreo del estadístico MEDIA MUESTRAL.**
- **3.1.1** En una población cualquiera con varianza conocida.
- **3.1.2** En una población NORMAL con varianza conocida.
- **3.1.3** En una población NORMAL con varianza desconocida, t de Student.

3.2. Distribución en el muestreo de una **PROPORCIÓN MUESTRAL**
3.3. Distribución en el muestreo de la **diferencia de dos PROPORCIONES MUESTRALES**
3.4. **Distribución en el muestreo del estadístico VARIANZA MUESTRAL** χ^2 de Pearson.

Objetivos: *Comprender y asimilar que una muestra aleatoria de tamaño "n" puede ser considerada desde una doble perspectiva: como un repertorio de n valores y como n variables aleatorias independientes e idénticamente distribuidas. Igualmente un estadístico (como la media muestral por ejemplo) podrá ser considerado como un único valor o como una variable aleatoria. Se estudiarán las distribuciones de probabilidad asociadas a las variables aleatorias que conforman los estadísticos muestrales más importantes.*

BIBLIOGRAFÍA BÁSICA (teoría y problemas)

- ESTEBAN, J.; et Alt.: "Inferencia estadística", Garceta Editorial, 2011. Temas 3 (3.1 a 3.6, 3.8 y 3.9).
- NEWBOLD, P.: "Estadística para los Negocios y la Economía". Ed. Prentice Hall, Capítulo 6.

1. Población y muestra. Inferencia Estadística.

COLECTIVO o UNIVERSO: conjunto de elementos de los que queremos estudiar alguna propiedad. Puede ser finito (N elementos) o no finito.

POBLACIÓN: conjunto de datos asociados a los elementos del colectivo i generados por la variable aleatoria que mide la propiedad que queremos estudiar. Puede ser finita (N) o no finita.

MUESTRA: selección aleatoria de n datos de la población.

- *Concepto de muestra.* Se considerará la muestra desde una doble perspectiva:
  - como ***n*** valores: observaciones obtenidas de la población.
  - como ***n*** variables aleatorias independientes e idénticamente distribuidas: descripción teórica de todas las muestras posibles de tamaño ***n***.

- *Objeto de tomar una muestra*: hacer afirmaciones acerca de la distribución de la población.

- ***Estadístico muestral***: es un resumen de la muestra y, por tanto, está en función de los valores muestrales. También se estudiarán los estadísticos desde una doble acepción:

  - Como un único valor relativo a la muestra concreta obtenida.
  - *Como una variable aleatoria* que toma un valor para cada muestra. En este caso tendrá una distribución que esquematice su variabilidad.

- *Parámetro poblacional*: valor de la población que se pretende estimar.

INFERENCIA ESTADÍSTICA.

- *Inferencia Estadística Bayesiana*: basada en información muestral e información a priori (Teorema de Bayes).

- *Inferencia Estadística Clásica*: las propiedades inducidas sobre la población proceden exclusivamente de la información muestral.
  - *Estimación: teoría de la estimación estadística* (a partir de estadísticos muestrales se estiman parámetros poblacionales)
    - *Estimación puntual* (un solo valor para el parámetro estimado)
    - *Estimación por intervalos* (el parámetro estimado pertenece a un rango de valores según un nivel de confianza establecido)

  - *Contrastes de hipótesis*: *teoría de la decisión estadística* (se formulan hipótesis acerca de la población y se utiliza la información muestral para comprobarlas y aceptarlas o no)
    - *Paramétricos* (población con un modelo de probabilidad conocido)
    - *No paramétricos* (no se conoce el modelo de probabilidad poblacional)

2. Tipos de muestreo. Muestreo aleatorio.

TIPOS DE MUESTREO

No aleatorio u opinático: subjetivo

Muestreo aleatorio: al azar
Muestreo aleatorio simple (m.a.s): con remplazamiento
Muestreo aleatorio irrestricto: sin remplazamiento
Muestreo aleatorio estratificado: se obtiene una muestra simple sobre cada estrato.
Cfr.: ESTEBAN, J.; et Alt.: "Inferencia estadística", Garceta 2011 (pág.94-98)

Muestreo artificial: técnica bootstrap. Se obtienen muestras a partir de una función de distribución F(x).
Cfr.: J.C.Tuner. "Matemática moderna aplicada", Alianza Editorial, capítulo 5 Muestreo.

Ejemplo 8.1 MUESTREO ALEATORIO SIMPLE (m.a.s.): Da lugar a variables aleatorias independientes e idénticamente distribuidas.

Supóngase que se pide la opinión a un colectivo sobre la política medioambiental del gobierno.

Se supondrá conocido el parámetro poblacional $p = 0{,}40$ que indica la proporción de personas que está a favor.

Sea la v. a. $X = \begin{cases} 0 & en\, contra \quad P(X=0) = 0{,}60 \\ 1 & a\, favor \quad P(X=1) = 0{,}40 \end{cases}$

$X \sim Bernoulli(p = 0{,}40)$

$$E(X) = p =$$

$$Var(X) = p(1-p) =$$

Muestreo aleatorio simple sobre esa población:

(X_1, X_2) representa todas las posibles muestras de tamaño 2.

Como el muestreo es con remplazamiento,

$\Rightarrow (X_1, X_2)$ son v.a. **independientes** $\Rightarrow P(X_1, X_2) = P(X_1) \cdot P(X_2)$

Por tanto, la distribución conjunta quedará:

$n = 2$:

(X_1, X_2)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$P(X_1, X_2)$	$0,36$	$0,24$	$0,24$	$0,16$

A partir de la distribución conjunta se puede obtener fácilmente la distribución de la suma

$Y = \sum_{i=1}^{2} X_i = X_1 + X_2$ que mide el número de opiniones a favor:

$$P(Y = 0) = P(0;0) = 0{,}36$$
$$P(Y = 1) = P(0;1) + P(1;0) = 0{,}48$$
$$P(Y = 2) = P(1;1) = 0{,}16$$

Además, las dos variables $X_1 \ i \ X_2$ tienen la misma distribución, es decir,

$X_i \sim Bernoulli(p = 0{,}40) \rightarrow$ son idénticamente distribuidas.

$$X_i: opinión\ del\ i - esimo\ encuestado$$

$$X_i = \begin{cases} 0 & en\ contra \quad P(X_i = 0) = 0{,}60 \\ 1 & a\ favor \quad P(X_i = 1) = 0{,}40 \end{cases}$$

X_1	X_2	$P(X_1, X_2)$
0	0	0,36
0	1	0,24
1	0	0,24
1	1	0,16

$P(X_1 = 0) = 0{,}36 + 0{,}24 = 0{,}60$

$P(X_1 = 1) = 0{,}24 + 0{,}16 = 0{,}40$

y lo mismo para X_2

Así pues, son idénticamente distribuidas: $X_1 \sim Bernoulli(p = 0{,}40)$
$X_2 \sim Bernoulli(p = 0{,}40)$

3. Estadísticos y distribuciones asociadas.

ESTADÍSTICOS MUESTRALES:

$X_1, \ldots, X_n$ muestra aleatoria simple de tamaño ***n***

- Media muestral: $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$

- Varianza muestral: $S_n^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{n}$

- Cuasi-varianza muestral: $S_{n-1}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{n-1}$ $S_{n-1}^2 = \frac{n}{n-1} S_n^2$

- Proporción muestral: $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$

SE SUPONDRÁ MUESTREO ALEATORIO SIMPLE (m.a.s.) EN TODO EL TEMA.

3.1. Distribución en el muestreo del estadístico MEDIA MUESTRAL.

PARÁMETROS DEL ESTADÍSTICO MEDIA MUESTRAL $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$

$$E(\bar{X}_n) = \mu$$

$$Var(\bar{X}_n) = \frac{\sigma^2}{n} \quad DT(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

Para cualquier tamaño muestral.

3.1.1. En una población cualquiera con varianza σ^2 ***conocida***

PARA ***n GRANDE***

Por TCL $\bar{X}_n \sim Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

(En este contexto, es una distribución aproximada)

Es decir: $\Rightarrow Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma} \cdot \sqrt{n} \sim Normal(0, 1)$

3.1.2.En una población NORMAL con varianza σ^2 conocida.

PARA ***n cualquier tamaño***

Por reproductividad Normal $\bar{X}_n \sim Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

(En este contexto, es una distribución exacta)

Es decir: $\Rightarrow Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma} \cdot \sqrt{n} \sim Normal(0,1)$

Ejemplo 8.2 Si de una población Normal de media μ = 10 y desviación típica σ = 3 se extraen muestras aleatorias de tamaño n = 16 $(X \sim N(10;3) \to n = 16)$ ¿cuál es la probabilidad de que la media muestral tome valores entre 8,53 y 11,47?

Solución: 0,95

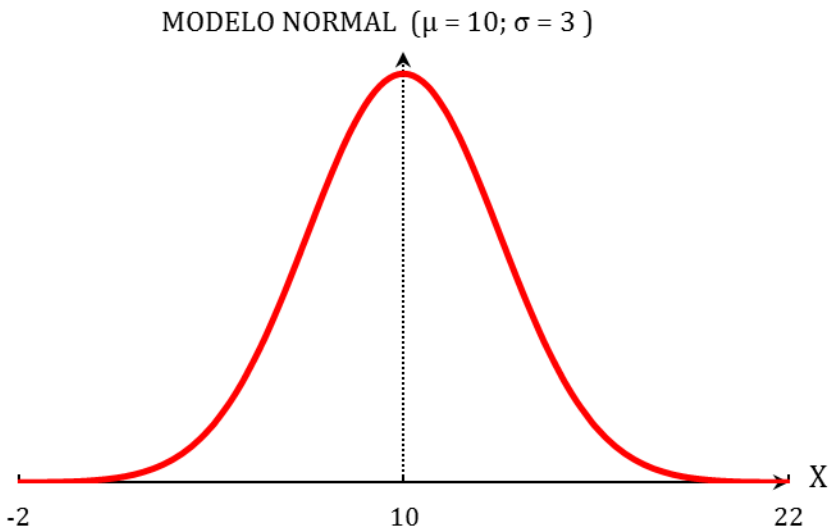
Solución:

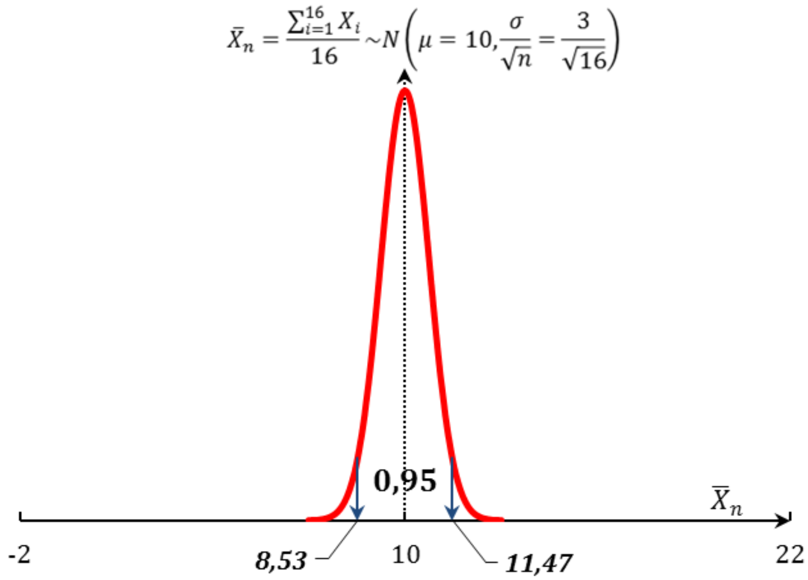
Para calcular la probabilidad pedida, $P(\bar{X}_n \in [8{,}53; 11{,}47]) = P(8{,}53 \leq \bar{X}_n \leq 11{,}47)$, *se utilizará distribución en el muestreo de la media muestral en una población Normal de varianza* σ^2 *conocida:*

$\bar{X}_n \sim Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \to \bar{X}_n \sim Normal\left(\mu = 10, \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{16}}\right) \to Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - 10}{3/\sqrt{16}} \sim Normal(0,1)$

$P(8{,}53 \leq \bar{X}_n \leq 11{,}47) = P\left(\frac{8{,}53-10}{3/\sqrt{16}} \leq \frac{\bar{X}_n - 10}{3/\sqrt{16}} \leq \frac{11{,}47-10}{3/\sqrt{16}}\right) =$

$= P(-1{,}96 \leq Z \leq 1{,}96) = 2F(1{,}96) - 1 = 2 \cdot 0{,}9750 - 1 = 0{,}95$





En el 95% de las muestras de tamaño n = 16 de una población Normal (10; 3), la variable aleatoria media muestral $\bar{X}_n$ tomará valores entre 8,53 y 11,47 alrededor de la constante media poblacional μ = 10.

Ejemplo 8.3 Si de una población Normal de media μ y desviación típica σ = 3 se extraen muestras aleatorias de tamaño n = 16 ($X \sim N(\mu; 3) \rightarrow n = 16$) ¿en qué porcentaje de muestras se verificará que la diferencia entre la media de la población μ y la media de la muestra $\bar{X}_n$ sea menor que 1,47?

Solución: 95%

Solución: (el planteamiento es análogo al ejemplo 8.3 pero para cualquier μ)

Para calcular la probabilidad pedida, $P(|\bar{X}_n - \mu| \leq 1{,}47) = P(-1{,}47 \leq \bar{X}_n - \mu \leq 1{,}47)$, se utilizará distribución en el muestreo de la media muestral en una población Normal de varianza σ^2 conocida:

$$\bar{X}_n \sim Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \rightarrow \bar{X}_n \sim Normal\left(\mu, \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{16}}\right) \rightarrow Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - \mu}{3/\sqrt{16}} \sim Normal(0, 1)$$

Así, $P(|\bar{X}_n - \mu| \leq 1{,}47) = P(-1{,}47 \leq \bar{X}_n - \mu \leq 1{,}47) = P\left(\frac{-1{,}47}{3/\sqrt{16}} \leq \frac{\bar{X}_n - \mu}{3/\sqrt{16}} \leq \frac{1{,}47}{3/\sqrt{16}}\right) =$
$= P(-1{,}96 \leq Z \leq 1{,}96) = 2F(1{,}96) - 1 = 2 \cdot 0{,}9750 - 1 = 0{,}95$

3.1.3. En una población NORMAL con varianza σ^2 desconocida.

PARA ***n* cualquier tamaño**

Se utilizará la siguiente v.a. $\Rightarrow T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n-1}} = \frac{\bar{X}_n - \mu}{S_n} \cdot \sqrt{n-1} \sim t_{n-1}$

t de Student con (n-1) grados de libertad.

Para hacer estimaciones y contrastes sobre la media poblacional μ.

Ejemplo 8.4 Partiendo de una población $N(\mu, \sigma^2)$ con σ^2 desconocida y extrayendo m.a.s. de tamaño 26 ¿cuál es el porcentaje de muestras en las que se verificará que la diferencia entre la media de la población y la media de la muestra sea menor que 0,824? De una muestra aleatoria concreta de 26 datos, se ha obtenido que $S_n^2 = 4$.

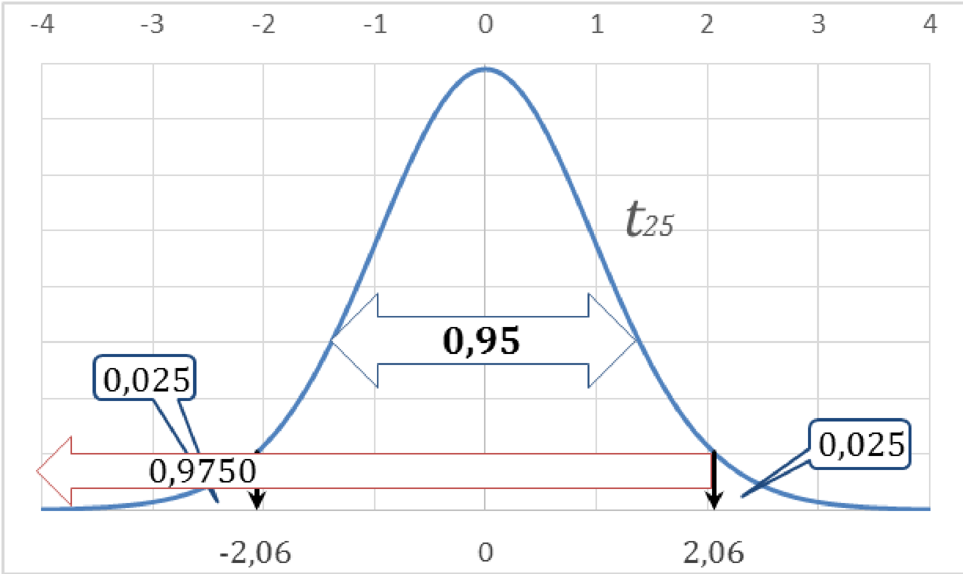
Solución: 0,95

Para calcular la probabilidad pedida, $P(|\bar{X}_n - \mu| \leq 0{,}824) = P(-0{,}824 \leq \bar{X}_n - \mu \leq 0{,}824)$*, se utilizará distribución en el muestreo de la media muestral en una población Normal de varianza* σ^2 *desconocida:*

$\rightarrow T = \frac{\bar{X}_n - \mu}{s_n/\sqrt{n-1}} \sim t_{n-1} \rightarrow T = \frac{\bar{X}_n - \mu}{2/\sqrt{25}} \sim t_{25}$

Así, $P(|\bar{X}_n - \mu| \leq 0{,}824) = P(-0{,}824 \leq \bar{X}_n - \mu \leq 0{,}824) = P\left(\frac{-0{,}824}{2/\sqrt{25}} \leq \frac{\bar{X}_n - \mu}{2/\sqrt{25}} \leq \frac{0{,}824}{2/\sqrt{25}}\right) =$

$$= P(-2{,}06 \leq t_{25} \leq 2{,}06) = 2F(2{,}06) - 1 = 2 \cdot 0{,}9750 - 1 = 0{,}95$$


-4
-3
-2
-1
0
1
2
3
4
t25
0,95
0,025
0,9750
0,025
-2,06
0
2,06

3.2. Distribución en el muestreo de una PROPORCIÓN MUESTRAL.

POBLACIÓN: parámetro $p \Rightarrow$ proporción con que se da cierto fenómeno en la población.

m.a.s. de tamaño ***n*,** PARA ***n GRANDE***

$$X_i = \begin{cases} 0 & fracàs \quad P(X_i = 0) = 1 - p \\ 1 & èxit \quad P(X_i = 1) = p \end{cases}$$
$$X_i \sim Bernoulli(p) \quad i = 1, 2, \cdots, n$$

Sea la v.a. $X = \sum_{i=1}^{n} X_i$: nº de veces que repite el fenómeno en la muestra.

$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ proporción con que se observa el fenómeno en la muestra.

$E(\hat{p}) = p \quad Var(\hat{p}) = \frac{p(1-p)}{n}$

Por el TCL

$$\hat{p} \sim Normal\left(\mu = p, \sigma = \sqrt{\frac{p(1-p)}{n}}\right) \Rightarrow Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim Normal(0,1)$$

Ejemplo 8.5 Supóngase que el 10% de los estudiantes de la Universitat de València viene a la Universitat en bibicleta (*p = 0,10*). ¿Cuál es la probabilidad de que más del 9% de los estudiantes de muestras de n = 200 se desplace a la Universitat en bicicleta?

Solución: 0,6808

Solución:

La probabilidad pedida, $P(\hat{p} > 0{,}09)$*, se calculará utilizando la distribución en el muestreo de una proporción muestral:* $\hat{p} \sim Normal\left(\mu = p, \sigma = \sqrt{\frac{p(1-p)}{n}}\right) \Rightarrow Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\hat{p}-0{,}1}{\sqrt{\frac{0{,}1\times 0{,}9}{200}}} \sim Normal(0,1)$

Así, $P(\hat{p} > 0{,}09) = P\left(\frac{\hat{p}-0{,}1}{\sqrt{\frac{0{,}1\times 0{,}9}{200}}} > \frac{0{,}09-0{,}1}{\sqrt{\frac{0{,}1\times 0{,}9}{200}}}\right) = P(Z > -0{,}471) = P(Z \leq 0{,}471) = F(0{,}47) = 0{,}6808$

Es decir, aproximadamente en el 68% de las muestras de 200 estudiantes se obtendría que más del 9% de ellos viene a la Universitat en bicicleta.

Planteamiento equivalente: la v.a. $X = \sum_{i=1}^{200} X_i$ *que mide el nº de estudiantes que acude en bicicleta de muestras de n = 200, por el TCL, se distribuye aproximadamente como una Normal de parámetros:*

$E(X) = np = 200 \cdot 0{,}1 = 20 \quad Var(X) = np(1-p) = 200 \cdot 0{,}1 \cdot 0{,}9 = 18$

$\Rightarrow Z = \frac{X-20}{\sqrt{18}} \sim Normal(0,1)$*. Como el 9% de 200 es 18:*

$P(\hat{p} > 0{,}09) = P(X > 18) = P\left(\frac{X-20}{\sqrt{18}} > \frac{18-20}{\sqrt{18}}\right) = P(Z > -0{,}471) = P(Z \leq 0{,}471) = F(0{,}47) = 0{,}6808$

3.3. Distribución en el muestreo de la diferencia de dos PROPORCIONES MUESTRALES.

$$\begin{matrix} \textit{Población } X_1 \\ p_1: \textit{proporción poblacional} \end{matrix} \rightarrow m.a.s.\ n_1 \qquad \begin{matrix} \textit{Población } X_2 \\ p_2: \textit{proporción poblacional} \end{matrix} \rightarrow m.a.s.\ n_2$$

$$\begin{matrix} n_1\ i\ n_2\ GRANDES\ independientes \\ X_1: n^{\underline{o}}\ de\ veces\ que\ se\ repite\ el\ fenómeno\ en\ la\ muestra\ 1 \\ X_2: n^{\underline{o}}\ de\ veces\ que\ se\ repite\ el\ fenómeno\ en\ la\ muestra\ 2 \end{matrix} \qquad \begin{matrix} \hat{p}_1 = \frac{X_1}{n_1} \quad \hat{p}_2 = \frac{X_2}{n_2} \\ proporciones\ muestrales \end{matrix}$$

Se sabe por apartado 3.2 que $\hat{p}_1$ y $\hat{p}_2$ son NORMALES, por tanto:

$$(\hat{p}_1 - \hat{p}_2) \sim Normal\left((p_1 - p_2);\ \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

$$\Longrightarrow Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim Normal(0,1)$$

Ejemplo 8.6 Supóngase que el porcentaje de estudiantes que está a favor del proceso de Bolonia es del 40% en la Universitat de València (UV) ($p_1 = 0{,}40$) y del 35% en la Universitat de Barcelona (UB) ($p_2 = 0{,}35$). Se toman dos muestras independientes por m.a.s. de tamaños $n_1 = 150$ en UV y $n_2 = 180$ en UB. ¿Cuál es la probabilidad de que el porcentaje de estudiantes de la muestra de la UV que está a favor de Bolonia sea inferior al porcentaje de estudiantes de la muestra de la UB? ¿Y de que sea mayor?

Solución: 0,1762 y 0,8238 respectivamente

Solución: se pide $P(\hat{p}_1 < \hat{p}_2) = P\big((\hat{p}_1 - \hat{p}_2) < 0\big)$

Aplicando el resultado anterior $\Longrightarrow Z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}} = \frac{(\hat{p}_1-\hat{p}_2)-(0{,}40-0{,}35)}{\sqrt{\frac{0{,}4\cdot 0{,}6}{150}+\frac{0.35\cdot 0{,}65}{180}}} \sim Normal(0,1)$

Así:

$$P\big((\hat{p}_1 - \hat{p}_2) < 0\big) = P\left[\frac{(\hat{p}_1-\hat{p}_2)-(0{,}40-0{,}35)}{\sqrt{\frac{0{,}4\cdot 0{,}6}{150}+\frac{0.35\cdot 0{,}65}{180}}} < \frac{0-0{,}05}{\sqrt{\frac{0{,}4\cdot 0{,}6}{150}+\frac{0.35\cdot 0{,}65}{180}}}\right] = P[Z < -0{,}934] = 1 - F(0{,}93) =$$

$$= 1 - 0{,}8238 = 0{,}1762$$

Es decir, que en el 17,62% de las muestras de esos tamaños que se puedan tomar en ambas universidades, el porcentaje de estudiantes a favor de Bolonia en la muestra de la UV será inferior al porcentaje a favor de la muestra de la UB. En el 82,38% de las muestras, ese porcentaje será mayor.

3.4. Distribución en el muestreo del estadístico VARIANZA MUESTRAL S_n^2

Parámetros del estadístico varianza muestral:

$$E\left(S_n^2\right) = \frac{n-1}{n} \cdot \sigma^2$$

$$E\left(S_n^2\right) \xrightarrow[n\to\infty]{} \sigma^2$$

***<u>Población Normal con media μ desconocida</u>*.**

En este caso la distribución de la varianza muestral es una ji – cuadrado con (n-1) grados de libertad:

$$Y = \frac{nS_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

Para hacer estimaciones y contrastes de σ^2

Ejemplo 8.7 De una población con distribución normal N(μ,2) de la que se extraen muestras aleatorias simples de tamaño n = 25, ¿Cuál es la probabilidad de que la varianza muestral sea menor que 5,8272 y mayor que 2,216?

Solución: 0,09

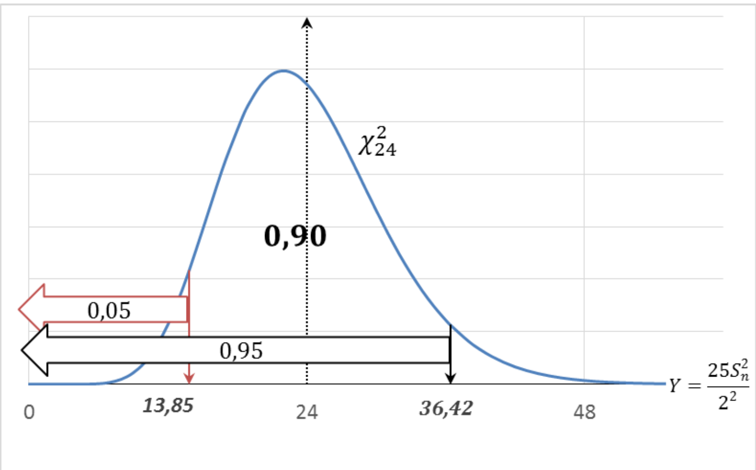
Solución:

Para calcular probabilidades sobre la variable aleatoria varianza muestral S_n^2 *se utilizará la distribución en el muestreo de la variable* $Y = \frac{nS_n^2}{\sigma^2}$ *que, en una población Normal de media μ desconocida, se distribuye como una ji – cuadrado con n – 1 grados de libertad:* $Y = \frac{nS_n^2}{\sigma^2} \sim \chi_{n-1}^2 \rightarrow Y = \frac{25S_n^2}{2^2} \sim \chi_{24}^2$.

Así, la probabilidad pedida queda:

$$P(S_n^2 \in [2{,}216; 5{,}8272]) = P(2{,}216 < S_n^2 < 5{,}8272) = P\left(\frac{25 \cdot 2{,}216}{2^2} < \frac{25S_n^2}{2^2} < \frac{25 \cdot 5{,}8272}{2^2}\right) =$$

$$= P(13{,}85 < \chi_{24}^2 < 36{,}42) = F(36{,}42) - F(13{,}85) = 0{,}95 - 0{,}05 = 0{,}90$$



Esto significa que, al extraer muestras aleatorias de tamaño 25 de una población normal de varianza 4 $(N(\mu;\ \sigma^2 = 4))$*, en el 90% de esas muestras la variable varianza muestral* S_n^2 *tomará valores entre 2,216 y 5,8272.*

EN RESUMEN:

TEOREMA DE COCHRAN – FISHER (*Distribución de la media y varianza muestrales)*

Si las v.a. $X_1, X_2, \cdots, X_n$ representan m.a.s de tamaño $\boldsymbol{n}$ de una población con **distribución normal** $\boldsymbol{N(\mu, \sigma)}$, el teorema de Cochran–Fisher demuestra que:

1. Respecto a la media muestral $Z = \frac{\bar{X}_n - \mu}{\sigma} \cdot \sqrt{n} \sim Normal(0, 1)$ (σ conocida)
2. Respecto a la varianza muestral $Y = \frac{nS_n^2}{\sigma^2} \sim \chi_{n-1}^2$ (μ desconocida)
3. La media muestral y la varianza muestral son variables aleatorias independientes.

Como consecuencia de este teorema, se puede demostrar fácilmente que la variable aleatoria $T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n-1}} = \frac{\bar{X}_n - \mu}{S_n} \cdot \sqrt{n-1} \sim t_{n-1}$ (σ desconocida)