cc www.pacorabadan.com

Tema 1. Estadísticos y sus distribuciones

Estadística II. Grado de Económicas

Dr. Francisco Rabadán Pérez

www.pacorabadan.com

1. Parámetros poblacionales y estadísticos muestrales

- En general diremos que los parámetros poblacionales son las características numéricas de la población.
  - En concreto, un parámetro es una concreción numérica de la distribución de la población.
  - El conocimiento del parámetro permite describir de forma parcial o total la función de probabilidad de la variable poblacional que estamos estudiando.

www.pacorabadan.com

1. Parámetros poblacionales y estadísticos muestrales (Ejemplo)

- Sabemos que la duración de cierto componente electrónico sigue una distribución exponencial, con la siguiente función de densidad:

$$f(x) = \theta e^{-\theta x} \; x > 0 \,,$$

pero esta función de densidad no estará totalmente determinada hasta que no sepamos el valor del parámetro θ.

- Del mismo modo ocurriría con las poblaciones:
  - $\mathrm{N}(\mu, \sigma)$,
  - $Poisson(\lambda)$,
  - $U[a, b]$, …

www.pacorabadan.com

1. Parámetros poblacionales y estadísticos muestrales (Ejemplo)

- Por tanto, el problema surge cuando conocemos la forma de la distribución de probabilidad de la población, pero ésta depende de un(os) parámetro(s) desconocido(s).
- Mediante el procedimiento de **muestreo aleatorio simple (m.a.s.)** seleccionaremos una muestra de tamaño n, $X = \{x_1, x_2, \dots, x_n\}$ y obtendremos una expresión o función de las variables muestrales que nos proporcione un valor aproximado del parámetro desconocido, es decir, una **estimación**.
- Estimaremos los parámetros poblacionales mediante estadísticos obtenidos empleando las observaciones de una muestra aleatoria, que llamaremos Estadísticos Muestrales.

www.pacorabadan.com

1. Parámetros poblacionales y estadísticos muestrales (Ejemplo)

- ESTADÍSTICO $T(X)$: Es cualquier función real de las variables aleatorias que integran la muestra, es decir, es una función de las observaciones muestrales que no contiene ningún valor o parámetro desconocido. En general lo representaremos por: $T = g(X)$ ó $T(X)$

www.pacorabadan.com

1. Parámetros poblacionales y estadísticos muestrales (Ejemplo)

	Parámetro poblacional	Estadístico muestral
Media	$E(\xi) = \mu = \alpha_1$	$a_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$
Varianza	$V(\xi) = \sigma^2 = \alpha_1 - \alpha_1^2$	$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i)^2\right) - \bar{x}^2 = a_1 - a_1^2$
Cuasivarianza	$V_1(\xi) = \frac{n}{n-1}\sigma^2$	$s_{1x}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}s_x^2$
Proporción	P o ∏ : número de éxitos dividido por el número de ensayos en una población binomial	P o π: número de éxitos en un m.a.s. sobre una población binomial respecto del número de encuestados (n).

www.pacorabadan.com

2. Distribución de los estadísticos muestrales

- Los estadísticos muestra les se calculan a partir de los valores $\boldsymbol{x_i}$ de una muestra aleatoria $X = \{x_1, x_2, \dots x_n\}$, y por tanto son variables aleatorias, y como una distribución de probabilidad.
- La distribución exacta de los estadísticos dependerá del tamaño de la muestra n.
- Nos interesará conocer las distribuciones de probabilidad de algunos estadísticos muestrales, en concreto:

a) Media muestral: $\bar{x}$

b) Varianza y cuasivarianza muestral: s_x^2 y s_{1x}^2

c) En poblaciones dicotómicas, la proporción muestra! (π ó p*).

www.pacorabadan.com

2. Distribución de los estadísticos muestrales

(Ejercicio)

En una empresa dedicada al transporte de mercancías se estudia el absentismo laboral, es decir, el número de días que ha faltado al trabajo un empleado elegido aleatoriamente de la plantilla total. Se sabe que durante el último año, han causado baja un día el 50% de los trabajadores, dos días el 40% y tres días el resto. Si se toma una m.a.s. de tamaño dos (X1 X2), se pide:

1. Distribución de probabilidad de la variable aleatoria "número de días que ha faltado al trabajo un empleado", así como su media y varianza.
2. Distribución de probabilidad del estadístico media muestra!, así como su esperanza y varianza.
3. Distribución de probabilidad del estadístico varianza muestral, y su esperanza.
4. Calcular la probabilidad de que el estadístico media muestra! sea menor que 2.

(cc) www.pacorabadan.com

2. Distribución de los estadísticos muestrales

(Ejercicio)

POBLACIÓN

ξ: nº dias falta al trabajo	x=1	x=2	x=3
P(ξ=xi)	0,50	0,40	0,10

PARÁMETROS POBLACIONALES:

$$E(\xi) = \mu = \sum_{\forall xi} x_i P(\xi = x_i) = 1 * 0{,}50 + 2 * 0{,}40 + 3 * 0{,}10 = 1{,}6$$

$$V(\xi) = \sigma^2 = \sum_{\forall xi} {x_i}^2 P(\xi = x_i) - \mu^2 = (1^2 * 0{,}50 + 2^2 * 0{,}40 + 3^2 * 0{,}10) - (1{,}6)^2 = 0{,}44$$

$V_1(\xi) = (n/n\text{-}1)*\sigma^2 = (2/1)*0{,}44 = 0{,}88$

www.pacorabadan.com

2. Distribución de los estadísticos muestrales

(Ejercicio)

M.A.S. n=2

Muestras $(x_1\ x_2)$	Media muestral $\overline{x_i} = \frac{(x_1 + x_2)}{2}$	Varianza muestral $S^2_{xi} = \sum_{i=1}^{2} \frac{(x_i - \overline{x})^2}{2}$	Probabilidades $P(X=x_1; X=x_2)$
(1,1)	$\overline{x_i} = \frac{(1+1)}{2} = 1$	$S^2_{xi} = \frac{(1-1)^2 + (1-1)^2}{2} = 0$	$P(x_1 = 1; x_2 = 1) = 0{,}5 * 0{,}5 = 0{,}25$
(1,2)	$\overline{x_i} = \frac{(1+2)}{2} = 1{,}5$	$S^2_{xi} = \frac{(1-1{,}5)^2 + (2-1{,}5)^2}{2} = 0{,}25$	$P(x_1 = 1; x_2 = 2) = 0{,}5 * 0{,}4 = 0{,}20$
(1,3)	2	1	0,05
(2,1)	1,5	0,25	0,20
(2,2)	2	0	0,16
(2,3)	2,5	0,25	0,04
(3,1)	2	1	0,05
(3,2)	2,5	0,25	0,04
(3,3)	3	0	0,01

www.pacorabadan.com

2. Distribución de los estadísticos muestrales

(Ejercicio)

Según los anteriores resultados, la distribución de probabilidad del estadístico media muestral $\bar{x}$ sería:

$\bar{x}_i$	1	1,5	2	2,5	3
P_i	0,25	0,40	0,26	0,08	0,01

Esperanza y varianza de la media muestral:

$$E(\bar{x}) = \sum_{\forall xi} \bar{x}_i P(\xi = \bar{x}_i) = 1*0{,}25 + 1{,}5*0{,}40 + 2*0{,}26 + 2{,}5*0{,}08 + 3*0{,}01 = 1{,}6$$

$$V(\bar{x}) = \left[\sum_{\forall xi} \bar{x}_i^{\,2} P(\xi = \bar{x}_i) \right] - (E(\bar{x}))^2 = (1^2*0{,}25 + 1{,}5^2*0{,}40 + 2^2*0{,}26 + 2{,}5^2*0{,}08 + 3^2*0{,}01) - (1{,}6)^2 = 0{,}22$$

RELACIONES ENTRE LOS PARÁMETROS POBLACIONALES Y EL ESTADÍSTICO MEDIA MUESTRAL:

Se cumple que: $E(\bar{x}) = \mu = E(\xi)$ $\qquad V(\bar{x}) = \dfrac{\sigma^2}{n}$

La distribución de probabilidad del estadístico varianza muestral S^2_x sería:

S^2_{xi}	0	0,25	1
P_i	0,42	0,48	0,10

Esperanza de la varianza muestral: $E(S_x^{\,2}) = (0*0{,}42)+(0.25*0{,}48)+(1*0{,}10) = 0{,}22$

www.pacorabadan.com

2. Distribución de los estadísticos muestrales

(Ejercicio)

RELACIONES ENTRE LOS PARÁMETROS POBLACIONALES Y EL ESTADÍSTICO VARIANZA MUESTRAL:

Se cumple que: $E(S_x^2) = \frac{n-1}{n}\sigma^2$

La varianza de la Varianza Muestral se obtiene como:
$V(S_x^2)=(0^2*0{,}42)+(0.25^2*0{,}48)+(1^2*0{,}10)-(0.22^2) = 0{,}0816$

www.pacorabadan.com

3. Distribución de los estadísticos muestrales para poblaciones normales $N(\mu, \sigma)$

3.1 Distribución de la media muestral cuando se conoce la varianza poblacional (σ^2)

Al estudiar la propiedad aditiva o reproductiva de una población Normal, decíamos que si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes, distribuidas cada una de ellas según una $N(\mu_i; \sigma_i)$ para i=1,2,...,n,

entonces la variable $Z=X_1+X_2+\ldots+X_n$ seguía una distribución Normal de parámetros $N(\sum_{i=1}^{n}\mu_i; \sqrt{\sum_{i=1}^{n}\sigma_i^2})$

Ahora vamos a particularizarlo para el caso del muestreo aleatorio simple (m.a.s.). Obtenemos una muestra $X=(x_1, x_2, \ldots, x_n)$ procedente de una población $N(\mu; \sigma)$ con σ conocida.

Si analizamos las consecuencias del m.a.s. sabemos que:

Muestreo aleatorio (probabilístico) $\Rightarrow L(\xi) \cong L(x_i) \Rightarrow L(\xi) \equiv L(x_1) \equiv L(x_2) \equiv \ldots \equiv L(x_n)$
Es decir, si $\xi \to N(\mu; \sigma) \Rightarrow x_i \to N(\mu; \sigma)$

www.pacorabadan.com

3. Distribución de los estadísticos muestrales para poblaciones normales $N(\mu, \sigma)$

Muestreo aleatorio simple $\Rightarrow$ x_i independientes $\Rightarrow$ $L(x_1, x_2, \ldots, x_n) = L(x_1)*L(x_2)*\ldots*L(x_n)$

Es decir, $L(x_i) \cong \dfrac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ para $X=(x_1, x_2, \ldots, x_n)$

Además la media muestral se obtiene como: $\bar{x}$ ó $a_x = \dfrac{\sum_{i=1}^{n} x_i}{n}$ por tanto, la variable $n\bar{x}$ ó $na_x = \sum_{i=1}^{n} x_i$ formada por la suma de las n variables muestrales independientes que siguen una normal $x_i \rightarrow N(\mu; \sigma)$ se distribuye como: $n\bar{x} \rightarrow N(\sum_{i=1}^{n} \mu; \sqrt{\sum_{i=1}^{n} \sigma^2}) \equiv N(n\mu; \sqrt{n\sigma^2})$

Entonces, la distribución de probabilidad del estadístico media muestral, la variable $\bar{x}$ ó $a_x = \dfrac{\sum_{i=1}^{n} x_i}{n}$ se distribuye como: $\bar{x} \rightarrow N(\mu; \dfrac{\sigma}{\sqrt{n}})$.

www.pacorabadan.com

3. Distribución de los estadísticos muestrales para poblaciones normales $N(\mu, \sigma)$

EJEMPLO:

En cierta Universidad los estudiantes de segundo ciclo de Carrera trabajan en prácticas antes de finalizar sus estudios. La remuneración de las prácticas sigue una Normal de media de 6000 € y desviación típica 600 €. Para una m.a.s. de 6 alumnos que están trabajando en prácticas, ¿cuál es la probabilidad de que el salario medio de la muestra sea mayor que 6350 €?

ξ: Remuneración o salario de alumnos en prácticas en €. $\xi \to N(6000; 600)$

m.a.s. n=6, X=(x_1,....,x_6); por tanto $\bar{x} \to N(6000; \frac{600}{\sqrt{6}}) \equiv N(6000;244{,}95)$

$$P(\bar{x} \geq 6350) = P(\frac{\bar{x}-6000}{244{,}95} \geq \frac{6350-6000}{244{,}95}) = P(\xi^* \geq 1{,}43) = 0{,}0764$$

www.pacorabadan.com

3. Distribución de los estadísticos muestrales para poblaciones normales $N(\mu, \sigma)$

3.2 Distribución de la varianza muestral (Fisher-Cochran) para poblaciones Normales

La distribución de la varianza muestral y la cuasivarianza muestral se puede aproximar a una chi-cuadrado como: $\frac{nS^2}{\sigma^2} = \chi^2_{n-1}$ ó bien $\frac{(n-1)S_1^2}{\sigma^2} = \chi^2_{n-1}$

EJEMPLO:
La empresa empaquetadora de paquetes de azúcar está interesada en disminuir la variabilidad de los pesos de paquetes de azúcar, respecto del contenido real al contenido marcado en la etiqueta. Si se supone que el peso se distribuye como una Normal, hallar aquellos valores k_1 y k_2, para una m.a.s. de 25 paquetes de azúcar, que verifican:

$$P\left(\frac{S_x^2}{\sigma^2} \leq k_1\right) = 0{,}05 \quad \text{y} \quad P\left(\frac{S_x^2}{\sigma^2} \geq k_2\right) = 0{,}05$$

Sabemos que n=25, como la población es normal, $\frac{25S^2}{\sigma^2} = \chi^2_{24}$, por tanto:

$$P\left(\frac{25S_x^2}{\sigma^2} \leq 25k_1\right) = 0{,}05\,;\, P(\chi^2_{24} \leq 25\text{k}_1) = 0{,}05 \quad \text{y} \quad P\left(\frac{S_x^2}{\sigma^2} \geq k_2\right) = 0{,}05; P(\chi^2_{24} \geq 25k_2) = 0{,}05$$

Buscando en las tablas de chi-cuadrado esas probabilidades, los valores de tablas son: $25k_1=12{,}6$ y $25k_2=36{,}4$ despejando $k_1=0{,}504$ y $k_2=1{,}456$.

www.pacorabadan.com

3. Distribución de los estadísticos muestrales para poblaciones normales $N(\mu, \sigma)$

3.3 Distribución de la media muestral cuando se desconoce la varianza poblacional

Sabemos que $\bar{x} \rightarrow N(\mu; \frac{\sigma}{\sqrt{n}})$ y también que $\frac{nS^2}{\sigma^2} = \chi^2_{n-1}$ y σ es desconocida, tenemos que emplear una distribución de probabilidad que no dependa de σ, es decir, elaborar una t de Student de la siguiente forma:

$$t_{n-1} = \frac{\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} = \frac{\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\frac{nS^2}{\sigma^2}}{n-1}}} = \frac{\bar{x}-\mu}{\sqrt{\frac{S^2}{n-1}}}, \text{ entonces } \bar{x} = (t_{n-1}\frac{S}{\sqrt{n-1}}) + \mu$$

www.pacorabadan.com

3. Distribución de los estadísticos muestrales para poblaciones normales $N(\mu, \sigma)$

EJEMPLO:

En cierta Universidad los estudiantes de segundo ciclo de Carrera trabajan en prácticas antes de finalizar sus estudios. La remuneración de las prácticas sigue una Normal de media de 6 mil € y desviación típica desconocida. Para una m.a.s. de 10 alumnos que están trabajando en prácticas, ¿cuál es la probabilidad de que el salario medio de la muestra sea mayor que 6,5 miles €? Datos obtenidos de la muestra: $\Sigma xi=64$ €; $\Sigma xi^2=430$.

ξ: Remuneración o salario de alumnos en prácticas en €. $\xi \rightarrow N(6; \sigma)$ σ desconocida.

m.a.s. n=10, $X=(x_1,....,x_{10})$; por tanto $\bar{x} \rightarrow N(6; \frac{\sigma}{\sqrt{10}})$ y $\frac{10S^2}{\sigma^2} = \chi_9^2$ entonces:

$$t_9 = \frac{\frac{\bar{x}-69}{\frac{\sigma}{\sqrt{10}}}}{\sqrt{\frac{\chi_9^2}{9}}} = \frac{\frac{\bar{x}-6}{\frac{\sigma}{\sqrt{10}}}}{\sqrt{\frac{\frac{10S^2}{\sigma^2}}{9}}} = \frac{\bar{x}-6}{\sqrt{\frac{S^2}{9}}} \Rightarrow \bar{x} = (t_9 \frac{S}{\sqrt{9}}) + 6$$

www.pacorabadan.com

3. Distribución de los estadísticos muestrales para poblaciones normales $N(\mu, \sigma)$

Calculamos el valor de la varianza muestral y de la media muestral, con los datos del m.a.s. realizado:

$$S_x^2 = \frac{\sum_{i=1}^{10} x_i^2}{10} - (\bar{x})^2 = \frac{430}{10} - \left(\frac{64}{10}\right)^2 = 2{,}04 \qquad \bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{64}{10} = 6{,}4$$

$$P(\bar{x} \geq 6{,}5) = P(t_9 \frac{\sqrt{2{,}04}}{\sqrt{9}} + 6 \geq 6{,}5) = P(t_9 \frac{\sqrt{2{,}04}}{\sqrt{9}} \geq (6{,}5 - 6)) = P(t_9 \geq 0{,}5 * 0{,}476) = P(t_9 \geq 0{,}238)$$

Buscamos en tablas de la t de Student el valor 0,238 (el más aprox.) y la probabilidad buscada es 0,10.

www.pacorabadan.com

4. Distribución de los estadísticos muestrales para dos poblaciones normales $N(\mu_i, \sigma_i\ )$

4.1 Distribución de la diferencia de medias muestrales cuando se conocen las varianzas poblacionales $\sigma_1^2\ \sigma_2^2$

Sea $X=(x_1,\ x_2,\ldots,x_n)$ una m.a.s. procedente de una población $N(\mu_1;\ \sigma_1)$ donde su media muestral se distribuye según $\overline{x_1} \rightarrow N(\mu_1; \frac{\sigma_1}{\sqrt{n}})$.

Sea $Y=(y_1,\ y_2,\ldots,y_n)$ otra m.a.s. independiente, procedente de una población $N(\mu_2;\ \sigma_2)$ donde su media muestral se distribuye según $\overline{y}_2 \rightarrow N(\mu_2; \frac{\sigma_2}{\sqrt{m}})$

La distribución de probabilidad del estadístico diferencia de medias muestrales tendrá la siguiente distribución:

$$(\overline{x} - \overline{y}) \rightarrow N(\mu_1 - \mu_2; \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}).$$

www.pacorabadan.com

4. Distribución de los estadísticos muestrales para dos poblaciones normales $N(\mu_i, \sigma_i)$

EJEMPLO:
Los alimentos precocinados de un fabricante A tienen una duración media de 1400 horas, con una desviación típica de 200 horas. Los de otro fabricante B tienen una caducidad media de 1200 horas, con desviación típica de 100 horas. Si tomamos muestras al azar de 125 cajas de alimentos precocinados de cada fabricante, ¿cuál es la probabilidad de que los alimentos precocinados de A tengan una duración media de 250 horas más que los del fabricante B?

ξ_A: Duración alimentos precocinados fabricante A, en horas. $\xi_A \to N(1400;200)$, tomamos una m.a.s. n=125; la media muestral sigue $\overline{x_A} \to N(1400;\frac{200}{\sqrt{125}})$

ξ_B: Duración alimentos precocinados fabricante B, en horas. $\xi_B \to N(1200;100)$, tomamos una m.a.s. m=125; la media muestral sigue $\overline{x_B} \to N(1200;\frac{100}{\sqrt{125}})$

El estadístico diferencia de medias muestrales se distribuye:

$$(\overline{x_A} - \overline{x_B}) \to N(1400-1200;\sqrt{\frac{200^2}{125}+\frac{100^2}{125}}) \cong N(200;20)$$

La probabilidad que nos piden hallar es:

$$P\left[\ (\overline{x_A} \geq 250 + \overline{x_B})\right] = P[(\overline{x_A} - \overline{x_B}) \geq 250\ \ ] = P(\xi^* \geq \frac{250-200}{20}) = P(\xi^* \geq 2{,}5) = 0{,}0062$$

www.pacorabadan.com

4. Distribución de los estadísticos muestrales para dos poblaciones normales $N(\mu_i, \sigma_i)$

4.2 Distribución de la diferencia de medias muestrales cuando las varianzas poblacionales son desconocidas pero iguales

$\xi_1 \rightarrow N(\mu_1; \sigma)$, extraemos una m.a.s. de tamaño n, donde $\bar{x}_1 \rightarrow N(\mu_1; \frac{\sigma}{\sqrt{n}})$; $\frac{nS_x^2}{\sigma^2} = \chi^2_{n-1}$

$\xi_2 \rightarrow N(\mu_2; \sigma)$, extraemos una m.a.s. de tamaño m, donde $\bar{y}_2 \rightarrow N(\mu_2; \frac{\sigma}{\sqrt{m}})$ $\frac{mS_y^2}{\sigma^2} = \chi^2_{m-1}$

$(\bar{x}_1 - \bar{y}_2) \rightarrow N(\mu_1 - \mu_2; \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}) = N(\mu_1 - \mu_2; \sqrt{\sigma^2 \frac{n+m}{nm}})$ como σ es desconocida, tenemos que encontrar una distribución que no dependa de σ, esto es, la t de Student.

$$t_{n+m-2} = \frac{\dfrac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\dfrac{n+m}{nm}}}}{\sqrt{\dfrac{\chi^2_{n+m-2}}{n+m-2}}} = \ldots\ldots = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{nS_x^2 + mS_y^2}} \frac{\sqrt{n+m-2}\sqrt{nm}}{\sqrt{n+m}}$$

www.pacorabadan.com

4. Distribución de los estadísticos muestrales para dos poblaciones normales $N(\mu_i, \sigma_i)$

EJEMPLO:

Sea la variable aleatoria ξ con distribución N(2; σ). Mediante muestreo aleatorio simple se seleccionan dos m.a.s. independientes de tamaño n=5, m=4. Calcule la probabilidad de que la media muestral de la primera muestra sea superior a la de la segunda como mínimo en 4,38 unidades, sabiendo que S^2_x=2,298 y S^2_y=3,23.

$\xi_1 \rightarrow N(2;\sigma)$, extraemos una m.a.s. de tamaño n=5, donde $\overline{x}_1 \rightarrow N(2; \frac{\sigma}{\sqrt{5}})$; $\frac{5*2{,}298}{\sigma^2} = \chi^2_4$

$\xi_2 \rightarrow N(2;\sigma)$, extraemos una m.a.s. de tamaño m=4, donde $\overline{y}_2 \rightarrow N(2; \frac{\sigma}{\sqrt{4}})$ $\frac{4*3{,}23}{\sigma^2} = \chi^2_3$

$$t_7 = \frac{\dfrac{(\overline{x_1} - \overline{y_2}) - (2-2)}{\sigma\sqrt{\dfrac{5+4}{5*4}}}}{\sqrt{\dfrac{\chi^2_7}{7}}} = \ldots\ldots = \frac{(\overline{x}_1 - \overline{y_2}) - 0}{\sqrt{5*2{,}298 + 4*3{,}23}} \frac{\sqrt{7}\sqrt{5*4}}{\sqrt{5+4}} = \frac{(\overline{x_1} - \overline{y_2})}{\sqrt{24{,}41}} * 3{,}9439 = (\overline{x}_1 - \overline{y_2}) * 0{,}79825$$

$$P(\overline{x_1} > \overline{y_2} + 4{,}38) = P[(\overline{x}_1 - \overline{y_2}) > 4{,}38] = P(\frac{t_7}{0{,}79825} > 4{,}38) = P(t_7 > 3{,}5) = 0{,}01$$

www.pacorabadan.com

5. Estadísticos muestrales de poblaciones dicotómicas (binomiales)

5.1. Distribución de la proporción muestral, cuando el tamaño de la muestra es grande: n>30

Sea una población $\xi \to B(1;P)$, donde llamamos proporción poblacional al parámetro P, desconocido:

$$P = \frac{\text{Nº elementos de la población que cumplen cierta característica}}{\text{Total elementos de la población}}$$

Recordamos que en las poblaciones binomiales, clasificamos los elementos en dos grupos mutuamente excluyentes, aquellos que cumplen la característica observada y aquellos que no, únicamente toman dos valores, X=1 para aquellos que sí tienen dicha característica y X=0 para los que no.

La esperanza y varianza de una B(1; P) se obtienen como: $E(\xi) = p; \quad V(\xi) = pq = p(1-p)$

Mediante una m.a.s. de tamaño n, $X=(x_1, x_2, \ldots, x_n)$ se estima el parámetro P mediante $\hat{p} = \pi$, proporción muestral, siendo:

$$\hat{p} = \pi = \frac{\text{Nº elementos de la muestra que cumplen cierta característica}}{\text{Tamaño de la muestra}} = \frac{\sum_{i=1}^{n} x_i}{n},$$

en donde Σxi lo componen únicamente aquellos valores de la muestra que tienen la característica observada, es decir, que xi=1.

www.pacorabadan.com

5. Estadísticos muestrales de poblaciones dicotómicas (binomiales)

5.1. Distribución de la proporción muestral, cuando el tamaño de la muestra es grande: n>30

Calculamos la esperanza y varianza de la variable aleatoria proporción muestral como:

$$E(\hat{p}) \text{ ó } E(\pi) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n} np = p$$

recordamos m.a.s. xi indeptes, L(xi)≡L(ξ)

$$V(\hat{p}) \text{ ó } V(\pi) = V\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2} npq = \frac{pq}{n} = \frac{p(1-p)}{n}$$

Para un tamaño de la muestra grande, n>30, podemos aproximar la distribución de la proporción muestral a una Normal, aplicando el teorema central del límite:

$\pi \xrightarrow[TCL]{} N(p; \sqrt{\frac{p(1-p)}{n}})$ y cuando p desconocido, $N(\pi; \sqrt{\frac{\pi(1-\pi}{n}})$,

cc www.pacorabadan.com

5. Estadísticos muestrales de poblaciones dicotómicas (binomiales)

5.1. Distribución de la proporción muestral, cuando el tamaño de la muestra es grande: n>30

EJEMPLO:
Se desea conocer la seguridad de las instalaciones eléctricas en edificios de más de 40 años de antigüedad, porque se estima que el 30% del conjunto de las viviendas tienen una instalación eléctrica insegura. Se toma una m.a.s. de 250 viviendas construidas hace más de 40 años, ¿cuál es la probabilidad de que el porcentaje de hogares de esa muestra con una insegura instalación eléctrica sea del 25% al 35%?

$\xi \rightarrow B(1;P)$

www.pacorabadan.com

5. Estadísticos muestrales de poblaciones dicotómicas (binomiales)

5.1. Distribución de la proporción muestral, cuando el tamaño de la muestra es grande: n>30

x=1 – éxito – viviendas con instalación eléctrica insegura P=% edificios con instalac. eléct. inseg.; P*=0,30
x=0 – fracaso – viviendas con instalación eléctrica segura.

$X=(x_1,.....,x_{250})$, se toma una m.a.s. n=250 viviendas, donde habrá viviendas que tengan instalación eléctrica insegura (cumplen la característica observada), y por tanto su valor xi=1, y aquellas que no por lo que su valor es xi=0.

La proporción de viviendas de la muestra con instalación eléctrica insegura es la proporción muestral, P* ó π y como el tamaño de la muestra n es grande,

$$\pi \xrightarrow[TCL]{} N(0{,}30;\sqrt{\frac{0{,}30*0{,}70}{250}}) \cong N(0{,}30;0{,}0289)$$ por tanto, la probabilidad que me piden calcular es:

$$P(0{,}25 \leq P^* \leq 0{,}35) = P(\frac{0{,}25-0{,}30}{0{,}0289} \leq \xi^* \leq \frac{0{,}35-0{,}30}{0{,}0289}) = P(-1{,}73 \leq \xi^* \leq +1{,}73) = \boxed{0{,}91638827}$$

www.pacorabadan.com

5. Estadísticos muestrales de poblaciones dicotómicas (binomiales)

5.2. Distribución de la diferencia de las proporción muestrales, cuando el tamaño de la muestra es grande: n>30 m>30

$$\xi_1 \to B(1;P_1) \quad \xi_2 \to B(1;P_2)$$

$$(\pi_1 - \pi_2) \xrightarrow[TCL]{} N[\ ](p_1 - p_2); \sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}\ ] \ \mathrm{p}_1, \mathrm{p}_2 \text{ desconocidos}$$

$$N(\pi_1 - \pi_2; \sqrt{\frac{\pi_1(1-\pi_1)}{n} + \frac{\pi_2(1-\pi_2}{m}})$$

www.pacorabadan.com

5. Estadísticos muestrales de poblaciones dicotómicas (binomiales)

5.2. Distribución de la diferencia de las proporción muestrales, cuando el tamaño de la muestra es grande: n>30 m>30

EJEMPLO:

El porcentaje de votantes con preferencia en un determinado partido es del 5% en una región A y del 10% en otra región B. Consultados 100 electores de la región A y 150 de la región B, determinar la probabilidad de que el porcentaje de electores consultados favorables a dicho partido en la segunda región supere en más del 0,02 al porcentaje de electores favorables a dicho partido en la primera.

$\xi_A \rightarrow B(1;P_A)$ $\quad P_A^*= 0{,}05$, m.a.s. n=100 $\quad P_A$ ó $\pi_A \xrightarrow[TCL]{} N(0{,}05;\sqrt{\frac{0{,}05*0{,}95}{100}}) \cong N(0{,}05;0{,}0218)$

$\xi_B \rightarrow B(1;P_B)$ $\quad P_B^*= 0{,}10$, m.a.s. m=150 $\quad P_B$ ó $\pi_B \xrightarrow[TCL]{} N(0{,}10;\sqrt{\frac{0{,}10*0{,}90}{150}}) \cong N(0{,}10;0{,}0245)$

$$(\pi_A - \pi_B) \xrightarrow[TCL]{} N\left[-0{,}05;\sqrt{\frac{0{,}05*0{,}95}{100}+\frac{0{,}10*0{,}90}{150}}\right] \cong N(-0{,}05;0{,}032787)$$

$$P[\pi_A > \pi_B + 0{,}02] = P[(\pi_A - \pi_B) > 0{,}02] = P(\xi^* > \frac{0'22 - (-0{,}05)}{0{,}032787}) = P(\xi^* > 1{,}52) = 0{,}0643$$

www.pacorabadan.com

Bibliografía

- Martín Pliego, Fundamentos de Inferencia Estadística, Editorial AC, 3ªEd, 2005

- Otros recursos en www.pacorabadan.com