

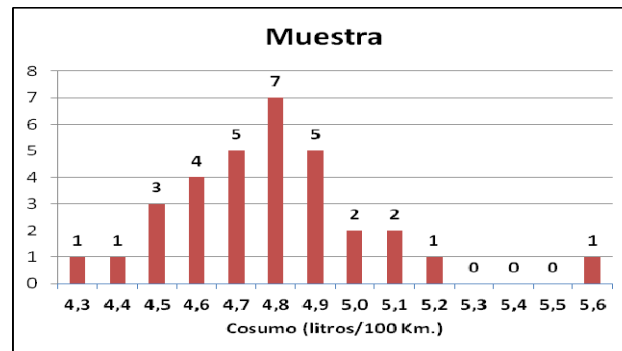
## MÉTODOS ESTADÍSTICOS DE LA INGENIERÍA

### PRIMERA CONVOCATORIA

### RESOLUCIÓN DEL EJERCICIO FINAL

#### Ejercicio 1: ESTADÍSTICA DESCRIPTIVA

Una empresa automovilística pretende cambiar el motor de uno de sus modelos de vehículo por otro con menor consumo. Para ello, se recogen en el siguiente diagrama de barras los consumos (en litros a los 100 km) de una muestra de estos nuevos motores:



- a) Determine los siguientes estadísticos muestrales: media, desviación típica, mediana y moda.

#### Solución

Estadísticos muestrales:

- Tamaño muestral:  $n = \sum_{i=1}^{i=14} n_i = 32$
- Media:  $\bar{x} = \frac{1}{n} \sum_{i=1}^{i=14} n_i \cdot x_i = 4.7875$
- Mediana:  $Me = 4.8$
- Moda:  $Mo = 4.8$
- Varianza:  $s^2 = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - \bar{x})^2 = 0.0623$
- Desviación típica:  $s = 0.2537$

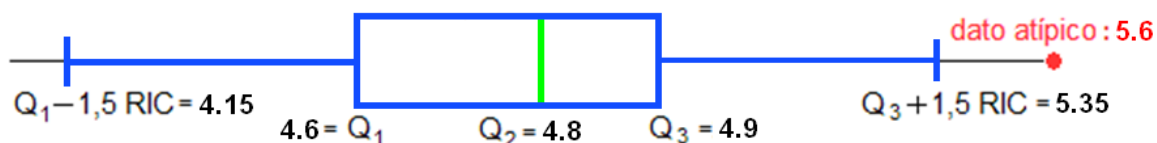
- b) Trace el *boxplot*.

#### Solución

Números resumen:

- Menor valor:  $L = 4.3$
- Mayor valor:  $H = 5.6$
- Mediana:  $Me = 4.8$
- Primer cuartil:  $Q_1 = 4.6$
- Tercer cuartil:  $Q_3 = 4.9$
- Rango intercuartílico:  $RIC = Q_3 - Q_1 = 4.9 - 4.6 = 0.3$
- Extremo inferior:  $LI = Q_1 - 1,5 \cdot RIC = 4.15$

- Extremo superior:  $LS = Q_3 + 1,5 \cdot RIC = 5.35$



- c) En la empresa se decide eliminar todos los datos atípicos de la muestra por considerarlos debidos a algún defecto mecánico o a una conducción ineficiente. Para esta **muestra corregida**, calcule la media y la desviación típica. Justifique la eliminación de datos atípicos.

**Solución**

Estadísticos muestrales de la muestra corregida:

- Dato atípico:  $x_{32} = 5.6 \notin [4.15, 5.35]$
- Tamaño muestral:  $n = \sum_{i=1}^{i=10} n_i = 31$
- Media:  $\bar{x} = \frac{1}{n} \sum_{i=1}^{i=10} n_i \cdot x_i = 4.7613$
- Varianza:  $s^2 = \frac{1}{n} \sum_{i=1}^{i=10} (x_i - \bar{x})^2 = 0.0424$
- Desviación típica:  $s = 0.2058$

**Ejercicio 2: CONTRASTE**

El motor que la empresa automovilística pretende retirar promedia un consumo de 4,9 litros a los 100 km. Para tomar una decisión se realizará un estudio a partir de la **muestra corregida** anterior, es decir, la presentada en el histograma eliminando el dato atípico (5,6).

- a) Se retirará el motor antiguo si se puede garantizar, con una confianza del 99%, que el nuevo modelo reduce el consumo. ¿Cuál será la decisión adoptada?

**Solución**

Datos:

- tamaño muestral:  $n = 31$
- media poblacional:  $\mu = 4,9 \text{ l./100km}$
- valor de la media muestral:  $\bar{x} = 4.7613 \text{ l./100km}$
- desviación típica muestral:  $s = 0.2058 \text{ l./100km}$
- cuasidesviación típica muestral:  $\hat{s} = 0.2092 \text{ l./100km}$
- grados de libertad:  $gl = v = 31 - 1 = 30$
- nivel de confianza:  $1 - \alpha = 0.99$
- nivel de significación:  $\alpha = 0.01$

Hipótesis:

- nula:  $H_0 : \mu = 4.9 \text{ litros/100km}$
- alternativa:  $H_1 : \mu < 4.9 \text{ litros/100km}$

Estadístico del contraste:  $\bar{X} \approx N\left(\mu_0; \frac{\hat{s}}{\sqrt{n}}\right) \Rightarrow \bar{X} \approx N(4.9; 0.0376)$

Valor del estadístico del contraste:  $\bar{x} = 4.7613 \text{ l./100km}$

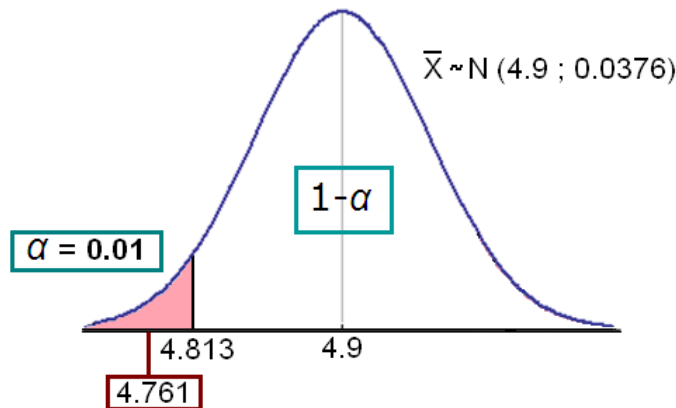
Región crítica:  $RC \equiv \frac{\bar{x} - \mu_0}{\frac{\hat{s}}{\sqrt{n}}} < z_\alpha$

- $z_\alpha = z_{0.01} = -2.3263 \Rightarrow z_\alpha = z_{0.99} = 2.3263$
- **Excel:** =DISTR.NORM.ESTAND.INV(0,99)=2.3263

$$RC \equiv \left\{ \bar{X} \in \mathbb{R} / \bar{x} \leq \mu_0 + \frac{\hat{s}}{\sqrt{n}} \cdot z_{0.01} \right\}$$

$$RC \equiv \left\{ \bar{X} \in \mathbb{R} / \bar{x} < 4.8126 \right\}$$

**Contraste.** La hipótesis nula se rechaza frente a la alternativa ya que  $\bar{x} \in RC$ . Por tanto, se adopta la decisión de cambiar el modelo de motor antiguo ya que el consumo medio del nuevo modelo es  $\mu < 4.9$  litros cada 100 km.



b) Determine el máximo nivel de significación que hubiera permitido aceptar (no rechazar) la hipótesis nula.

**Solución**

Cálculo del p-valor:

- $P(\bar{X} < 4.7613 \mid \bar{X} \approx N(4.9; 0.0376)) = 0.00011$
- **Excel:** =DISTR.NORM(4,7613;4,9;0,0376;1)=0.00011

Por lo tanto:  $p = 0.00011 \xrightarrow{\times 100} 0.011\%$

El máximo nivel de significación será:  $p = 0.00011 = \alpha_{\max}$

Se reduce el error de tipo I (rechazar la hipótesis nula siendo cierta).

**Ejercicio 3: ESTIMACIÓN**

La empresa automovilística también quiere incluir en su estudio una estimación de la proporción de motores nuevos que consumen por debajo del promedio de 4,9 litros a los 100 km. a partir de

la **muestra corregida** anterior, es decir, la presentada en el histograma eliminando el dato atípico (5,6).

a) Halle ese intervalo con una confianza del 99%.

**Solución**

Datos:

- tamaño muestral:  $n = 31$
- estimación puntual de la proporción:  $\hat{p} = \frac{21}{31} = 0.6774$
- $z_{\frac{\alpha}{2}} = z_{0.005} = 2.5758$
- **Excel:** =DISTR.NORM.ESTAND.INV(0,995)=2.5758

Intervalo pedido:

$$IC = \left( \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right)$$

$$IC = (0.4612, 0.8937)$$

b) ¿Cuál debería ser el tamaño muestral para que el intervalo tuviera una amplitud de 0.20 manteniendo el nivel de confianza?

**Solución**

Con los datos anteriores:

$$z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} = 0.1 \Rightarrow 2.5758 \cdot \sqrt{\frac{0.2185}{n}} = 0.1 \Rightarrow \sqrt{\frac{0.2185}{n}} = 0.0388$$

$$n = 144.98 \Rightarrow n = 145$$

**EJERCICIO 4: PROBABILIDAD**

a) Una urna contiene dos bolas rojas, otra urna contiene dos bolas blancas y una tercera urna contiene una bola roja y otra blanca. Se elige una urna al azar y se extrae una bola que resulta ser roja. ¿Cuál es la probabilidad de que la otra bola de esa urna sea también roja?

**Solución**

Se consideran los siguientes sucesos elementales:

- $U_i$  : se escoge la  $i$ -ésima urna ( $i=1,2,3$ )  $\rightarrow P(U_i) = \frac{1}{3}$
- $B_i$  : la bola  $i$ -ésima es blanca ( $i=1,2$ )
- $R_i$  : la bola  $i$ -ésima es roja ( $i=1,2$ )

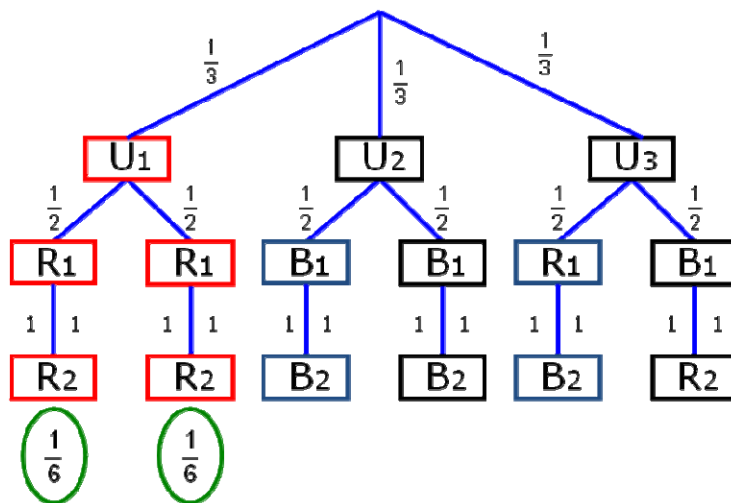
La probabilidad pedida es la siguiente:

$$P(R_1) = \sum_{i=1}^{i=3} P(U_i) \cdot P(R_1 | U_i) = \frac{1}{3} \sum_{i=1}^{i=3} P(R_1 | U_i) = \frac{1}{3} \left( 1 + 0 + \frac{1}{2} \right) = \frac{1}{3} \cdot \frac{3}{2} = \frac{1}{2}$$

$$P(R_2 | R_1) = \frac{P(R_1 \cap R_2)}{P(R_1)} = \frac{\frac{1}{6} + \frac{1}{6}}{\frac{1}{2}} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Una solución correcta del problema se obtiene intuitivamente con ayuda de un diagrama en árbol y una adecuada notación a la hora de representar el espacio muestral.

Un primer experimento es elegir al azar una de tres urnas con la siguiente composición: urna 1 (ROJA, ROJA), urna 2 (BLANCA, BLANCA) y urna 3: (ROJA, BLANCA). El segundo experimento consiste en extraer una bola de una de las urnas, donde se pueden encontrar los casos representados en la segunda división en ramas del árbol. El tercer experimento (última rama) es el tipo de bola que queda en la urna cuando se ha extraído la primera bola.



El enunciado nos pone una condición (una de las bolas es roja), con lo que nos pide una probabilidad condicional. Por tanto, quedan tres posibilidades equiprobables (como observamos en las ramas finales del árbol):

- que hayamos tomado la única bola roja en la urna 3 (ROJA, BLANCA); en este caso, la otra bola que queda en la urna es blanca
- que hayamos tomado la primera bola roja en la urna 1 con dos bolas rojas; en este caso, la otra bola es roja
- que hayamos tomado la segunda bola roja en la urna 1 con dos bolas rojas; en este caso, la otra bola, también, es roja

Dicho de otro modo, tenemos dos casos en que la urna elegida sea la primera (ROJA, ROJA) si la bola observada es roja y solo uno de que la urna sea la tercera (ROJA, BLANCA). En consecuencia, sabiendo que una bola es roja, la probabilidad de que la otra sea roja es el doble ( $\frac{2}{3}$ ) que la probabilidad de que sea blanca ( $\frac{1}{3}$ ).

- b) Un candado de combinación básico para maletas consta de tres discos en cada uno de los cuales figuran los diez dígitos del 0 al 9. Un empleado de una empresa de *handling* de cierto aeropuerto pretende abrir una maleta cerrada con uno de estos candados sin causarle destrozos. Calcula que dispone de cinco minutos para la operación y que puede obtener una combinación del candado cada cuatro segundos. ¿Cuál es la probabilidad de que abra la maleta?

**Solución**

Aplicación de la regla de Laplace:

- **Casos posibles.** En una combinación del candado importa el orden de los dígitos y puede tener dígitos repetidos, por tanto, se trata de variaciones con repetición de diez elementos tomados de tres en tres:

$$VR_{10,3} = 10^3 = 1000$$

- **Casos favorables.** El operario puede introducir 15 combinaciones por minuto con lo que en el plazo de cinco minutos que se ha fijado puede introducir 75 combinaciones.

Por lo tanto, siendo el suceso A, “abrir el candado”:

$$P(A) = \frac{\text{casos favorables}}{\text{casos posibles}} = \frac{75}{1000} = 0,075$$

### Ejercicio 5: VARIABLE ALEATORIA

- a) Se considera el experimento aleatorio consistente en el lanzamiento simultáneo de dos dados. ¿Cuántos lanzamientos deben realizarse para que la probabilidad de obtener un doble seis (dos seises) al menos una vez sea superior a 0.5? Describa previamente la variable aleatoria correspondiente.

#### Solución

Descripción del éxito y el fracaso del experimento así como de la variable aleatoria:

- **S (éxito)**: “sacar doble seis en un lanzamiento”  $\rightarrow P(S) = \frac{1}{36} = p$
- **F (fracaso)**: “NO sacar doble seis en un lanzamiento”  $\rightarrow P(F) = \frac{35}{36} = q$
- **X**: “número de dobles seises obtenidos”  $\rightarrow X \approx B\left(n; \frac{1}{36}\right)$

Entonces:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{n}{0} \cdot p^0 \cdot q^n = 1 - \left(\frac{35}{36}\right)^n > \frac{1}{2} \Rightarrow \left(\frac{35}{36}\right)^n < \frac{1}{2}$$

Tomando logaritmos neperianos:

$$n \cdot \ln\left(\frac{35}{36}\right) < \ln \frac{1}{2} \Rightarrow n < \frac{-0.69314718}{-0.02817088} = 24.6050977$$

Por lo tanto:

$$n = 25$$

- b) Una empresa de electrodomésticos a punto de cerrar vende su último stock de 1000 televisores a una cadena hotelera a un precio muy rebajado dado que 20 presentan serios desperfectos. Se seleccionan al azar 25 de esos televisores y se envían al nuevo hotel que la cadena va a inaugurar. ¿Cuántos aparatos con defectos pueden esperarse en el envío? Interprete el resultado. ¿Cuál es la probabilidad de que ninguno de los televisores enviados al hotel presente desperfectos?

#### Solución

Descripción del éxito y el fracaso del experimento así como de la variable aleatoria:

- **S (éxito)**: “televisor con algún defecto serio”  $\rightarrow P(S) = \frac{20}{1000} = 0.02 = p$
- **F (fracaso)**: “televisor SIN ningún defecto serio”  $\rightarrow P(F) = 0.98 = 1 - p = q$

- $X$ : “número de televisores defectuosos enviados”  $\rightarrow X \approx H(1000; 25; 0.02)$

Esperanza matemática:  $E[X] = n \cdot p = 25 \cdot 0.02 = 0.5$

Como, evidentemente, no puede recibirse medio televisor defectuoso el resultado se debe interpretar como un promedio que pertenece a envíos repetidos hechos en las condiciones dadas.

Sean:

- número total de televisores:  $N = 1000$
- tamaño muestral:  $n = 25$
- número de “éxitos” en la muestra:  $i = 0$
- número total de “éxitos”:  $r = 20$

Entonces:

$$P(X=i) = \frac{\binom{r}{i} \cdot \binom{N-r}{n-i}}{\binom{N}{n}} \Rightarrow P(X=0) = \frac{\binom{20}{0} \cdot \binom{980}{25}}{\binom{1000}{25}} = \frac{980!}{25! \cdot 955!} \cdot \frac{1000!}{25! \cdot 975!}$$

$$P(X=0) = \frac{980! \cdot 25! \cdot 975!}{1000! \cdot 25! \cdot 955!} = 0.5997196$$

En este caso, el modelo hipergeométrico resulta muy similar al modelo binomial:

$$H\left(N; n; \frac{r}{N}\right) \xrightarrow[\frac{n}{N} < 0.1]{N \rightarrow \infty} B\left(n; p = \frac{r}{N}\right)$$

Por tanto:

$$P(X=0) = \binom{25}{0} \cdot p^0 \cdot q^{25} = (0.98)^{25} = 0.60346473$$

- **Excel:** =DISTR.HIPERGEOM(0;25;20;1000)=0,5997196
- **Excel:** =DISTR.BINOM(0;25;0,02;1)=0.60346473