Muestreo y estimación: problemas resueltos

BENITO J. GONZÁLEZ RODRÍGUEZ (bjglez@ull.es)
DOMINGO HERNÁNDEZ ABREU (dhabreu@ull.es)
MATEO M. JIMÉNEZ PAIZ (mjimenez@ull.es)
M. ISABEL MARRERO RODRÍGUEZ (imarrero@ull.es)
ALEJANDRO SANABRIA GARCÍA (asgarcia@ull.es)

Departamento de Análisis Matemático
Universidad de La Laguna

Índice

6. Problemas resueltos

6.1. Distribuciones muestrales

Ejercicio 6.1. *Si el contenido en gramos de un determinado medicamento sigue una distribución normal* $N(7.5, 0.3)$*, calcular la probabilidad de que en una muestra de tamaño* 5 *se obtenga que la media es menor que* 7.

RESOLUCIÓN. Sea $\overline{X}$ la media de la muestra. Con los datos que tenemos: $n = 5$, $\mu = 7.5$ y $\sigma = 0.3$. Como $\overline{X}$ sigue una distribución normal

$$N\left(7.5, \frac{0.3}{\sqrt{5}}\right) = N(7.5, 0.4474),$$

se tendrá que

$$\begin{aligned} P(\overline{X} < 7) &= P(Z < \frac{7-7.5}{0.4472}) = P(Z < -3.7269) = P(Z > 3.7269) \\ &= 1 - P(Z < 3.7269) = 1 - 0.9999 = 0.0001. \end{aligned}$$

□

Ejercicio 6.2. *Una fábrica de pasteles elabora, en su producción habitual, un* 3 % *de pasteles defectuosos. Un cliente recibe un pedido de* 500 *pasteles de la fábrica. Calcular la probabilidad de que encuentre más del* 5 % *de pasteles defectuosos.*

RESOLUCIÓN. Estamos tomando una muestra de tamaño $n = 500$, de una población donde la proporción de pasteles defectuosos es de $p = 0.03$. Teniendo en cuenta que la distribución muestral de una proporción se ajusta a una normal

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right),$$

se desprende que estaríamos ante una normal $N(0.03, 0.0076)$, con lo cual

$$P(\widehat{p} > 5) = P\left(Z > \frac{0.05-0.03}{0.0076}\right) = P(Z > 2.63) = 1 - P(Z \leq 2.63) = 1 - 0.9957 = 0.0043.$$

□

Ejercicio 6.3. *Un ascensor limita el peso de sus* 4 *ocupantes a* 300 *kilogramos. Si el peso de un individuo sigue una distribución normal* $N(71,7)$*, calcular la probabilidad de que el peso de* 4 *individuos supere los* 300 *kilogramos.*

RESOLUCIÓN. Considerando que el peso de cada persona presenta una distribución normal con $\mu = 71$ y $\sigma = 7$, al seleccionar una muestra de 4 personas tenemos que:

$$P(X_1+X_2+X_3+X_4 > 300) = P\left(\frac{X_1+X_2+X_3+X_4}{4} > \frac{300}{4}\right) = P(\overline{X} > 75) = P\left(Z > \frac{75-71}{\frac{7}{\sqrt{4}}}\right)$$
$$= P(Z > 1.143) = 1 - P(Z < 1.143) = 1 - 0.8735 = 0.1265.$$

□

Ejercicio 6.4. *Se prueba el rendimiento, en kilómetros por litro, de dos tipos de gasolina, encontrándose una desviación estándar de* 1.23 *kilómetros por litro para la primera gasolina y una desviación estándar de* 1.37 *kilómetros por litro para la segunda. Se prueba la primera gasolina en* 35 *vehículos, y la segunda en* 42.

a) *¿Cuál es la probabilidad de que la primera gasolina dé un rendimiento promedio mayor de* 0.45 *kilómetros por litro que la segunda gasolina?*

b) *¿Cuál es la probabilidad de que la diferencia en rendimientos promedio se encuentre entre* 0.65 *y* 0.83 *kilómetros por litro a favor de la primera gasolina?*

RESOLUCIÓN. Tenemos que $\sigma_1 = 1.23$ km/L y $\sigma_2 = 1.37$ km/L, con $n_1 = 35$ y $n_2 = 42$.

a) En este caso:

$$P(\overline{X}_1 - \overline{X}_2 > 0.45) = P\left(Z > \frac{0.45-0}{\sqrt{\frac{1.23^2}{35}+\frac{1.37^2}{42}}}\right) = P(Z > 1.52)$$
$$= 1 - P(Z < 1.52) = 1 - 0.9358 = 0.0642.$$

b) Ahora:

$$P(0.65<\overline{X}_1-\overline{X}_2<0.83)=P\left(\frac{0.65-0}{\sqrt{\frac{1.23^2}{35}+\frac{1.37^2}{42}}}<Z<\frac{0.83-0}{\sqrt{\frac{1.23^2}{35}+\frac{1.37^2}{42}}}\right)=P(2.19<Z<2.80)$$
$$=P(Z<2.80)-P(Z<2.19)=0.9974-0.9857=0.0117.$$

□

6.2. Intervalos de confianza

Ejercicio 6.5. *Se ha comprobado que la concentración promedio de zinc que se saca del agua de un río a partir de una muestra de mediciones de zinc en* 36 *sitios diferentes es de* 2.6 *gramos por mililitro. Encontrar los intervalos de confianza del* 95 % *y* 99 % *para la concentración media de zinc en el río, suponiendo que la desviación típica de la población es* 0.3.

RESOLUCIÓN. En el primer caso $1-\alpha=0.95$, es decir, $\alpha=0.05$, de donde $Z_{\alpha/2}=1.96$. Como $\sigma=0.3$, $n=36$ y $\overline{X}=2.6$, podemos afirmar que

$$IC=2.6\pm\frac{1.96\cdot 0.3}{\sqrt{36}}=(2.50,2.70),$$

y $\mu\in IC$ al 95 %.

En el segundo caso $1-\alpha=0.99$, es decir, $\alpha=0.01$, de donde $Z_{\alpha/2}=2.575$. Con el resto de los datos igual que en el caso anterior, sigue que

$$IC=2.6\pm\frac{2.575\cdot 0.3}{\sqrt{36}}=(2.47,2.73),$$

de donde se concluye que $\mu\in IC$ al 99 %. □

Ejercicio 6.6. *Determinar un intervalo de confianza al nivel* $\alpha=0.05$ *para la probabilidad de que un recién nacido sea niño, si en una muestra de tamaño* 123 *se han contabilizado* 67 *niños.*

RESOLUCIÓN. Como $\alpha = 0.05$, $Z_{\alpha/2} = Z_{0.025} = 1.96$ y además $\widehat{p} = 67/123 = 0.54$, con $n = 123$ se tiene que

$$IC = \left(0.54 \pm 1.96\sqrt{\frac{0.54 \cdot 0.46}{123}}\right) = (0.046, 0.633)$$

contendrá a la proporción de recién nacidos con una probabilidad del 95 %. □

Ejercicio 6.7. *El encargado del departamento de producción de una fábrica recibe un lote de* 2000 *piezas necesarias para el montaje de un artículo. El fabricante de las piezas asegura que en este lote no hay más de* 100 *piezas defectuosas.*

a) *¿Cuántas piezas hay que examinar para que, con un nivel de confianza del* 95 %*, el error que se cometa en la estimación de la proporción de piezas defectuosas no sea mayor que* 0.05*?*

b) *Si se toma una muestra de* 100 *piezas elegidas al azar y se encuentran* 4 *defectuosas, determinar un intervalo de confianza para la proporción de defectuosas al nivel del* 95 %*.*

RESOLUCIÓN. *a*) Para este caso debemos tomar

$$n \geq \left(\frac{Z_{\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})}}{E}\right)^2 \qquad \text{(proporción } p \text{ desconocida)}.$$

Como $\widehat{p} = 100/2000 = 0.05$, $E = 0.05$, $1-\alpha = 0.05$ y $Z_{\alpha/2} = 1.96$, se concluye que

$$n \geq 1.96^2\,\frac{0.05 \cdot 0.95}{0.0025} \simeq 72.99,$$

es decir, se deberían analizar 73 piezas.

b) Ahora $\widehat{p} = 4/100 = 0.04$, así que

$$\begin{aligned} IC &= \left(\widehat{p} - Z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right) \\ &= \left(0.04 - 1.96\sqrt{\frac{0.04 \cdot 0.96}{100}}, 0.04 + 1.96\sqrt{\frac{0.04 \cdot 0.96}{100}}\right) \\ &= (0.002, 0.078). \end{aligned}$$

El fabricante aseguraba que la proporción de defectuosos en el lote era de $0.05 \in (0.002, 0.078$. Puesto que esta afirmación no contradice los resultados que hemos encontrado en la muestra analizada, decidimos confiar en el fabricante y aceptar el lote. □

Ejercicio 6.8. *El peso de los terneros de una granja se distribuye normalmente, con desviación típica de* 10 *kilogramos. Se toma al azar una muestra de* 35 *de ellos para transportarlos en un camión. Sabiendo que el peso medio resulta ser de* 140 *kilogramos, determinar un intervalo de confianza al* 8 % *de nivel de significación en el que oscilará el peso de los* 35 *terneros.*

RESOLUCIÓN. La variable $X =$'peso de los terneros' sigue una distribución normal $N(\mu, 10)$, con μ desconocida. Al ser el tamaño de la muestra mayor que 30 podemos considerar que la suma de los pesos, esto es, $35X$, sigue una distribución normal $N(35 \cdot 140, 10\sqrt{35})$, con lo que, al nivel de confianza del 92 %, tendríamos el intervalo:

$$IC = n\overline{X} \pm Z_{\alpha/2}\,\sigma\sqrt{n} = 35 \cdot 140 \pm 1.75 \cdot 10\sqrt{35} = 4900 \pm 103.53 = (4796.5, 50053.53).$$

□

Ejercicio 6.9. *Dos compañías A y B fabrican el mismo tipo de cable. Un distribuidor desea conocer la diferencia promedio de la resistencia a la rotura de los mismos, para lo cual toma muestras de* 100 *cables de A y* 50 *cables de B. La muestra de los cables de la compañía A arroja una resistencia promedio a la rotura de* 4500 *kilogramos, mientras que los cables de la compañía B arrojan una resistencia promedio a la rotura de* 4000 *kilogramos. Se sabe, por experiencia, que la desviación típica de la resistencia a la rotura es de* 300 *kilogramos para la compañía A y de* 200 *kilogramos para la compañía B. Se pide estimar, con un nivel de confianza del* 95 %*, el intervalo de confianza de la diferencia de medias de la resistencia a la rotura entre los dos cables, si la resistencia a la rotura se distribuye normalmente para ambas compañías.*

RESOLUCIÓN. Por los datos aportados sabemos que $n_1 = 100$, $n_2 = 50$, $\mu_1 = 4500$, $\mu_2 = 4000$, $\sigma_1 = 300$, $\sigma_2 = 200$, $\alpha = 0.05$ y $Z_{\alpha/2} = 1.96$. De aquí sigue que:

$$IC = (4500 - 4000) \pm 1.96\sqrt{\frac{300^2}{100} + \frac{200^2}{50}} = (419.19, 580.81).$$

Es decir,

$$P(419.12 < \mu_1 - \mu_2 < 580.81) = 0.95.$$

Podemos interpretar la expresión anterior diciendo que, con un nivel de confianza del 95 %, la diferencia de los promedios a la rotura de ambos cables fluctúa entre 419.19 y 580.81 kg. □

6.3. Contraste de hipótesis

Ejercicio 6.10. *Se quiere comprobar si una máquina destinada al llenado de envases de agua mineral ha sufrido un desajuste. Una muestra aleatoria de* 10 *envases de esta máquina ha proporcionado los siguientes resultados:*

$$0.49 \quad 0.52 \quad 0.51 \quad 0.48 \quad 0.53 \quad 0.55 \quad 0.49 \quad 0.50 \quad 0.52 \quad 0.49.$$

Suponiendo que la cantidad de agua mineral que este tipo de máquinas deposita en cada envase sigue una distribución normal de media 0.5 *litros y desviación típica* 0.02 *litros, se desea contrastar si el contenido medio de los envases de esta máquina es de* 0.5 *litros, con un nivel de significación del* 5 %*.*

a) *Plantear las hipótesis nula y alternativa del contraste.*

b) *Determinar la región crítica del contraste.*

c) *Realizar el contraste.*

RESOLUCIÓN. *a*) Hipótesis nula $H_0 : \mu = 0.5$; hipótesis alternativa $H_1 : \mu \neq 0.5$.

b) En nuestro caso: $\mu = 0.5$, $\sigma = 0.02$, $\alpha = 0.05$ y $Z_{\alpha/2} = 1.96$. Por tanto, la región de aceptación será

$$\left(\mu_0 - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = (0.4876, 0.5124),$$

y la región crítica, $\mathbb{R}\backslash(0.4876, 0.5124)$.

c) Se tiene:

$$\overline{X} = \frac{1}{n}\sum \overline{X}_i = \frac{5.08}{10} = 0.508.$$

Como $0.508 \in (0.4876, 0.5124)$, aceptamos la hipótesis nula $\mu = 0.5$ litros a un nivel de significación del 5 %, pudiendo concluir que la máquina no ha sufrido desajustes. □

Ejercicio 6.11. *Un laboratorio farmacéutico sostiene que uno de sus productos es efectivo en un* 90 % *para reducir una alergia en* 8 *horas. El medicamento dio buen resultado en* 160 *personas de una muestra de* 200*.*

Determinar si la afirmación del laboratorio es cierta.

RESOLUCIÓN. Tenemos una muestra de tamaño $n = 200$. Sea n_A el número de individuos a los que se le reduce la alergia en 8 horas, y sea p la proporción de individuos de la población que presentan dicha característica. Entonces $\widehat{p} = n_A/n$ es un estimador de p. En este caso, $\widehat{p}$ sigue una distribución normal

$$N\left(p, \sqrt{\frac{p(1-p)}{200}}\right).$$

Además, $\alpha = 0.1$ y $Z_{\alpha/2} = Z_{00.025} = 1.96$. Con estos datos, planteamos el contraste:

$$\begin{cases} H_0 : & p = 0.9\,(= p_0), \\ H_1 : & p \neq 0.9. \end{cases}$$

La región de aceptación es:

$$\left(p_0 + Z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}, \mu_0 + Z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}\right) = \left(0.9 - 1.96\sqrt{\frac{0.9\cdot 0.1}{200}}, 0.9 + 1.96\sqrt{\frac{0.9\cdot 0.1}{200}}\right)$$
$$= (0.859, 0.941).$$

Como $\widehat{p} = 0.8 \notin (0.859, 0.941)$, rechazamos H_0 con un nivel de significación del 5 %. Luego, $p \neq 0.9$, por lo que la afirmación del laboratorio es falsa. □

Ejercicio 6.12. *Se quiere contrastar el contenido de azúcar de distintos cargamentos de remolacha. Se sabe que el contenido medio de azúcar para remolacha de regadío es del* 18 % *con una media superior para el secano, siendo la desviación típica del* 6 % *en ambos casos. Se toma una muestra de* 20 *cargamentos. ¿Qué valor de la media permitirá tomar la decisión sobre si la remolacha es de secano o de regadío, al nivel del* 5 %*?*

RESOLUCIÓN. En estas condiciones se plantea:

$$\begin{cases} H_0 : & \mu \leq 18\,(= \mu_0), \\ H_1 : & \mu > 18. \end{cases}$$

Los datos que aporta el enunciado nos permiten afirmar que $n = 20$, $\alpha = 0.05$, $Z_\alpha = 1.645$ y $\sigma = 6$. La región de aceptación será:

$$\left(-\infty, \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}}\right) = \left(-\infty, 18 + 1.645\,\frac{6}{\sqrt{20}}\right) = (-\infty, 20.2).$$

Por consiguiente, si $\overline{X} < 20.2$ aceptamos remolacha de regadío con una confianza del 95 %. $\square$