Estadística I
Tema 8: Contraste de hipótesis
SOLUCIONES

Grado en Administración de Empresas 08/09

1. a) Media muestral: $\bar{x} = \frac{618}{10} = 61.8$
   Tablas: $z_{\frac{0.05}{2}} = z_{0.025} = 1.96$
   I.C. al $(1-\alpha)\,\%$ para μ con σ conocida: $IC_{1-\alpha}(\mu) = [\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$.
   En este caso, $IC_{1-\alpha}(\mu) = (61.8 \mp 1.96 \cdot 1.265) = (61.8 \mp 2.479) = (59.321; 64.279)$.
   b) No podemos rechazar la H_0 porque el valor propuesto en dicha hipótesis se encuentra dentro del intervalo de confianza para un $\alpha = 0.05$.
   c) Como el intervalo de confianza para la media del apartado a) contiene el valor propuesto de la hipótesis nula del apartado b), y como la amplitud del intervalo aumenta si pasamos de un nivel de confianza de 95 % a 99 % manteniendo todo lo demás constante, sabemos, sin hacer ningún cálculo que el nuevo intervalo también contendrá dicho valor, por tanto, no podemos rechazar la hipótesis nula para un $\alpha = 0.01$.

2. Se define el nivel crítico p o p-valor de un contraste como la probabilidad de rechazar la hipótesis nula para la muestra particular observada, cuando H_0 es cierta.
   El contraste a plantear es:

$$H_0 : \mu \geq 800$$
$$H_1 : \mu < 800$$

   El enunciado nos dice que $X \sim N(\mu, 120)$, por tanto en este caso, el estadístico del contraste viene dado por

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0,1)$$

   Y la región de rechazo en este caso es:

$$RC_\alpha = \left\{ \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < z_{1-\alpha} \right\}$$

   El valor del estadístico del contraste para esta muestra particular es

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{776 - 800}{120/10} = -2$$

   Así pues, el p-valor es

$$p = P\left(\frac{\overline{X} - 800}{120/10} < -2 | H_0 \; cierta \right) = P(Z < -2) = P(Z > 2) = 0.0228.$$

   Es decir, para la muestra obtenida se rechazaría la hipótesis nula para todo nivel de significación mayor que 0.0228.

3. a) El contraste a plantear es:

$$H_0 : \mu \geq 16$$
$$H_1 : \mu < 16$$

Aunque no conocemos la distribución de la edad de los lectores de la revista (X), como el tamaño de muestra es grande (100¿30), podemos aplicar el teorema central del límite. En ese caso tenemos que las distribución del estadístico de contraste es aproximadamente normal:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} N(0,1)$$

Y la región de rechazo en este caso es:

$$RC_\alpha = \left\{ \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < z_{1-\alpha} = -z_\alpha \right\}$$

Como tenemos: $n = 100$, $\overline{x} = 17.2$, $s = 2.3$ y $-z_{0.05} = -1.96$, entonces

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} = 5.217 > -z_{0.05} = -1.96.$$

Por tanto, al nivel de significación 0.05 no se rechaza la hipótesis nula.

b) El p-valor de un contraste es la probabilidad de rechazar la hipótesis nula para la muestra particular observada, cuando H_0 es cierta.
Para el contraste del apartado a), el p-valor es

$$p = P\left(\frac{\overline{X} - 16}{2.3/10} < 5.217 | H_0 \ cierta \right) \overset{Z \sim N(0,1)}{=} P(Z < 5.217) \approx 1.$$

Como el p-valor obtenido es 1, eso quiere decir que para la muestra obtenida no se rechaza la hipótesis nula a ningún nivel de significación.

4. a) Suponemos que la muestra el aleatoria simple y además el número de camas en términos porcentuales es una variable aleatoria $N(\mu, \sigma^2)$. Como la varianza es desconocida, los respectivos intervalos de confianza pedidos son:

$$IC_{1-\alpha}(\mu) = [\overline{x} \mp t_{n-1, \frac{\alpha}{2}} \tfrac{s}{\sqrt{n}}], \qquad IC_{1-\alpha}(\sigma^2) = [\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}]$$

En este caso, $\bar{x} = 3.925, s^2 = 9.7335, s = 3.12, n = 8, \alpha = .1, t_{n-1, \frac{\alpha}{2}} = 1,895, \chi^2_{n-1;\alpha/2} = 14.07$ y $\chi^2_{n-1;1-\alpha/2} = 2.17$. Por tanto, los intervalos finales son:

$$IC_{1-\alpha}(\mu) = [1.83,\ 6.01],$$

$$IC_{1-\alpha}(\sigma^2) = [4.84, 31.38] \Rightarrow IC_{1-\alpha}(\sigma) = [2.2, 5.6] \text{ (tomando la raíz cuadrada).}$$

b) El contraste de hipótesis a formular es

$$H_0 : \mu \leq 3.5$$
$$H_1 : \mu > 3.5$$

El estadístico del contraste:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}$$

Y la región de rechazo en este caso es:

$$RC_\alpha = \left\{ \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{7,\alpha} = 1.895 \right\}$$

El valor del estadístico del contraste para la muestra particular obtenida es $\dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = 0.38$, y por tanto no se puede rechazar la hipótesis nula.

c) Suponemos que ambas poblaciones son normales e independientes, y además, $Y_1, ..., Y_7$ m.a.s. El contraste a plantear es:

$$H_0 : \mu_x = \mu_y$$
$$H_1 : \mu_x \neq \mu_y$$

El estadístico del contraste en este caso es (varianzas poblacionales iguales):

$$T = \frac{\overline{X} - \overline{Y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \overset{H_0}{\sim} t_{n_1+n_2-2}$$

y la región de rechazo sería

$$RC_\alpha = \left\{ |\bar{x} - \bar{y}| > t_{n_1+n_2-2;\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\} =$$
$$= \left\{ |\bar{x} - \bar{y}| > 3.012 \cdot 2.664 \cdot \sqrt{\frac{1}{8} + \frac{1}{7}} \right\} = \{|\bar{x} - \bar{y}| > 4.15\}$$

Como $|\bar{x} - \bar{y}| = 0.18$, no existe evidencia estadística para rechazar la hipótesis nula.

5. a) Hipótesis asumidas: $X_1, ..., X_n$ es una m. a. s. (v. a. i. i. d.) normales.
El contraste a plantear es:

$$H_0 : \mu = 15$$
$$H_1 : \mu \neq 15$$

El estadístico del contraste:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}$$

Y la región de rechazo en este caso es:

$$RC_\alpha = \left\{ \left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| > t_{n-1;\frac{\alpha}{2}} = 2.306 \right\}$$

Como tenemos: $n = 9$, $\quad \overline{x} = 15.308$, $\quad s^2 = 0.193 \quad y \quad s = 0.44$, rechazamos H_0 a nivel del $5\,\% \Leftrightarrow \left|\frac{\bar{x}-\mu_0}{\frac{s_x}{\sqrt{n}}}\right| > t_{n-1;\frac{\alpha}{2}} \Leftrightarrow \left|\frac{15.308-15}{\frac{0.44}{\sqrt{9}}}\right| > t_{8;0.025} \Leftrightarrow 2.1 > 2.306$
Luego la muestra no proporciona evidencia suficiente para rechazar la hipótesis nula al nivel de significación $\alpha = 0.05$.

b) Por la dualidad intervalos de confianza-contrastes de hipótesis, 15 (valor del parámetro apuntado en H_0) ha de pertenecer al intervalo de confianza descrito en el enunciado.

c) Se define el nivel crítico p o p-valor de un contraste como la probabilidad de rechazar la hipótesis nula para la muestra particular observada, cuando H_0 es cierta.
Así pues, el p-valor es

$$p = P\left(|T| > \left| \frac{15.308 - 15}{\frac{0.44}{\sqrt{9}}} \right| |H_0 \; cierta \right) = P(|t_8| > 2.1) = 2P\,(t_8 > 2.1)$$

Como $P\,(t_8 > 2.1) \in (0.025, 0.05)$, entonces $p \in (0.05, 0.1)$.

El nivel p-valor es el menor nivel de significación para el cuál se rechaza la hipótesis nula con la muestra particular obtenida. Como $0.1 > p$-valor, para un nivel de significación $\alpha = 0.10$ también rechazaríamos la hipótesis nula.

6. Como $n < 30$ es necesario asumir normalidad para la variable de interés, X =cantidad de pasta que contiene cada paquete. Además la muestra debe ser m.a.s.

   a) Sean μ el parámetro media poblacional y σ la desviación típica poblacional. El contraste a plantear sería:

   $$H_0 : \mu \geq 500$$
   $$H_1 : \mu < 500$$

   La discrepancia observada (el valor del estadístico del contraste) es:

   $$Z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = -2.985$$

   y el p-valor es entonces $P(Z < -2.985) = 0.0014$, por lo que rechazamos la hipótesis nula. Podríamos estar cometiendo un error de tipo I con probabilidad menor de 0.0014.

   b) El contraste a plantear es:

   $$H_0 : \sigma \leq 3$$
   $$H_1 : \sigma > 3$$

   La discrepancia observada (el valor del estadístico del contraste) es:

   $$\chi^2_{19} = \frac{(n-1)S^2}{\sigma^2} = 33.7778$$

   y el valor, entonces está comprendido entre: $0.025 < p - valor < 0.01$.
   Entonces, rechazamos H_0 ya que el p-valor $< \alpha$.

7. a) Si asumimos que la variable X = medida de polución atmosférica, sigue una distribución normal, $X \sim N(\mu, \sigma)$, y que además la muestra recogida es aleatoria simple, entonces el intervalo de confianza $1 - \alpha$ para σ^2 viene dado por:

   $$I_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}} \right].$$

   Dada la muestra, calculamos $\overline{x} = 3.04$, $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2 = 1.4093$, $n = 10$, $\chi^2_{9;0.025} = 19.02$ y $\chi^2_{9;1-0.025} = 2.70$. Sustituyendo lo anterior en la expresión para el intervalo, tenemos finalmente que,

   $$I_{95\,\%}(\sigma^2) = [0.67, 4.70].$$

   b) Tenemos dos v.a. X = polución atmosférica en la primera ciudad e Y = polución atmosférica en la segunda ciudad. Puesto que los tamaños de muestra $n_1 = n_2 = 10$ son pequeños y no podemos aplicar el Teorema Central del límite, tenemos que suponer que ambas variables tienen distribución normal.

   $$X \sim N(\mu_1, \sigma_1) \qquad Y \sim N(\mu_2, \sigma_2)$$

   El contraste a plantear es:

   $$H_0 : \sigma_1 = \sigma_2$$
   $$H_1 : \sigma_1 \neq \sigma_2$$

   y el estadístico del contraste, y su distribución bajo la hipótesis nula sería

   $$T = \frac{S_1^2}{S_2^2} \overset{H_0}{\sim} F_{n_1-1,n_2-1}.$$

   La región crítica o de rechazo vendría dada por:

   $$RC_\alpha = \left\{ \frac{s_1^2}{s_2^2} > F_{n_1-1,n_2-1;\alpha/2} \text{ ó } \frac{s_1^2}{s_2^2} < F_{n_1-1,n_2-1;1-\alpha/2} \right\}$$

En nuestro caso: $n_1 = n_2 = 10$, $s_1^2 = 1.4093$, $s_2^2 = 1.5^2$, $F_{9,9;0.05} = 3.18$ y $F_{9,9;0.95} = \frac{1}{F_{9,9;0.05}} = \frac{1}{3.18} = 0.3145$. Como

$$\frac{s_1^2}{s_2^2} = \frac{1.4093}{1.5^2} = 0.6264 \in [0.3145, 3.18]$$

no rechazamos H_0 a nivel de significación 0.1.

8. a) El intervalo de confianza para la media de una distribución normal con varianza σ^2 conocida, viene dado por $[\bar{x} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$. En este caso sería: $[120 \mp 1.96\frac{20}{\sqrt{25}}] = [112.16, 127.84]$.
El contraste a realizar es

$$H_0 : \mu = 112$$
$$H_1 : \mu \neq 112$$

Como 112 no pertenece al IC obtenido para μ al 95 %, para un nivel de significación del 5 % se rechaza H_0. Y si se rechaza al 5 %, entonces también se rechaza al 10 %.

b) El contraste a realizar es

$$H_0 : \sigma = 20$$
$$H_1 : \sigma > 20$$

La hipótesis nula también se puede expresar como $H_0 : \sigma =\leq 20$ sin que varía ni el procedimiento de contraste ni las conclusiones que se van a obtener.
El estadístico del contraste es

$$(n-1)S^2/\sigma_0^2 \overset{H_0}{\sim} \chi^2_{n-1}$$

y la región crítica será

$$RC_{0.01} = \left\{(n-1)s^2/\sigma_0^2 > \chi^2_{n-1,\alpha} = \chi^2_{24,0.01} = 43\right\}$$

El valor del estadístico del contraste para la muestra particular obtenida es $(n-1)s^2/\sigma_0^2 = 30.73$ (que no pertenece a la región crítica) y por tanto, al nivel de significación $\alpha = 0.01$, no existe evidencia para rechazar la hipótesis nula, es decir, no puede concluirse que haya habido un aumento significativo de la varianza.