Estadística I
Tema 7: Intervalos de confianza
SOLUCIONES

Grado en Administración de Empresas 08/09

1. $X =$ "salarios percibidos". $E[X] = \mu$, $Var[X] = \sigma^2$, ambas desconocidas. Mediante muestreo aleatorio simple se toma una muestra donde

$$n = 100, \quad \overline{x} = \frac{1}{100}\sum_{i=1}^{9} x_i \cdot n_i = 778, \quad s = \sqrt{\frac{1}{99}\left(\sum_{i=1}^{9} x_i^2 \cdot n_i - 100 \cdot \overline{x}^2\right)} = 193.4.$$

a) Como no conocemos la distribución de X, pero el tamaño de muestra es grande, sabemos que por el TCL la cantidad pivotal es

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$$

y el intervalo de confianza para μ es

$$IC_{1-\alpha}(\mu) = \left[\bar{x} \mp z_{\alpha/2}\frac{s}{\sqrt{n}}\right].$$

En nuestro caso

$$IC_{0.95}(\mu) = \left[778 \mp z_{0.025}\frac{193.4}{\sqrt{100}}\right] = [740.09\,,\, 815.90]\,.$$

b) Las hipótesis del contraste son:

$$H_0 : \mu = 713$$
$$H_1 : \mu > 713$$

El estadístico del contraste viene dado por (puesto que tenemos m.a.s. y n grande)

$$Z = \frac{\overline{X} - 713}{S/\sqrt{n}} \overset{H_0}{\sim} N(0,1)$$

Y la región de rechazo en este caso es:

$$RC_\alpha = \left\{\frac{\bar{x} - 713}{s/\sqrt{n}} > z_\alpha\right\}$$

El valor del estadístico del contraste para esta muestra particular es

$$\frac{\bar{x} - 713}{s/\sqrt{n}} = \frac{778 - 713}{193.4/10} = 3.36.$$

Como no nos dan el nivel de significación, calculamos el p-valor:

$$p = P\left(\frac{\overline{X} - 713}{S/10} > 3.36 | H_0 \; cierta\right) = P(Z > 3.36) \approx 0.$$

Como el p-valor del contraste es prácticamente 0, para cualquier valor de significación vamos a rechazar la hipótesis nula. Es decir, podemos concluir que los salarios percibidos por los trabajadores de dicha empresa han aumentado.

2. $X =$ "notas del test de aptitud". $X \sim N(\mu, 28.2)$. Mediante muestreo aleatorio simple se toma una muestra donde

$$n = 9, \quad \overline{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = 122, \quad s = \sqrt{\frac{1}{8}\left(\sum_{i=1}^{9} x_i^2 - 9 \cdot \overline{x}^2\right)} = 21.58.$$

a) En este caso la cantidad pivotal es

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

y el intervalo de confianza para μ es

$$IC_{1-\alpha}(\mu) = \left[\bar{x} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

En nuestro caso

$$IC_{0.90}(\mu) = \left[122 \mp z_{0.05}\frac{28.2}{\sqrt{9}}\right] = [106.54\,,\ 137.46]\,.$$

b) El intervalo al 95 % será mayor, puesto que a mayor nivel de confianza, mayor longitud del intervalo (a mayor α, mayor es el valor de $z_{\alpha/2}$).

c) La longitud del intervalo es $2 \cdot z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Al nivel de confianza 0.95, si queremos un intervalo con longitud de a lo sumo 10:

$$2 \cdot z_{0.05}\frac{\sigma}{\sqrt{n}} \le 10 \iff z_{0.05}\frac{\sigma}{\sqrt{n}} \le 5 \iff z_{0.05}\frac{\sigma}{5} \le \sqrt{n} \iff z_{0.05}^2\frac{\sigma^2}{25} \le n$$

$$\iff n \ge z_{0.05}^2\frac{\sigma^2}{25} = 1.645^2\,\frac{28.2^2}{25} = 86.08.$$

Por lo tanto el tamaño de muestra mínimo necesario será 87.

3. $X =$ "posesión de teléfono en los hogares del citado país". $X \sim \mathcal{B}(p)$, donde p es la proporción de hogares que disponen de teléfono. Mediante muestreo aleatorio simple se toma una muestra donde

$$n = 200, \quad \hat{p} = \overline{x} = \frac{1}{200}\sum_{i=1}^{200} x_i = \frac{160}{200} = 0.8.$$

Tenemos una distribución de Bernouilli y un tamaño de muestra suficientemente grande para poder aplicar el Teorema Central del Límite, por lo tanto, el intervalo de confianza será:

$$IC_{1-\alpha}(p) = \left[\widehat{p} \mp z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right].$$

En nuestro caso,

$$IC_{0.95}(p) = \left[0.8 \mp z_{0.025}\sqrt{\frac{0.8 \cdot 0.2}{200}}\right] = [0.74\,,\ 0.86]\,.$$

Como $0.9 \notin IC_{0.95}(p)$, para un nivel de significación de 0.05, se rechazaría la hipótesis nula de que el 90 % de los hogares poseen teléfono.

4. X = "duración del préstamo en la biblioteca de la universidad". $E[X] = \mu$, $Var[X] = \sigma^2$, ambas desconocidas. Mediante muestreo aleatorio simple se toma una muestra donde

$$n = 100, \quad \overline{x} = 18, \quad s = 8.$$

Tenemos una distribución desconocida pero el tamaño de muestra es suficientemente grande para poder aplicar el Teorema Central del Límite, por lo tanto, el intervalo de confianza será:

$$IC_{1-\alpha}(\mu) = \left[\bar{x} \mp z_{\alpha/2} \frac{s}{\sqrt{n}}\right]$$

En nuestro caso,

$$IC_{0.99}(\mu) = \left[18 \mp z_{0.005} \frac{8}{\sqrt{100}}\right] = \left[18 \mp 2.575 \frac{8}{10}\right] = [15.94\,,\, 20.06]\,.$$

5. X = "calificación de las condiciones de confort en la universidad". $E[X] = \mu$, $Var[X] = \sigma^2$, ambas desconocidas. La distribución de X es discreta pero desconocida. Mediante muestreo aleatorio simple se toma una muestra donde

$$n = 172, \quad \overline{x} = 4.38, \quad s = 0.70.$$

a) Como no conocemos la distribución de X, pero el tamaño de muestra es grande, sabemos que por el TCL la cantidad pivotal es

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0, 1)$$

y el intervalo de confianza para mu es

$$IC_{1-\alpha}(\mu) = \left[\bar{x} \mp z_{\alpha/2} \frac{s}{\sqrt{n}}\right].$$

En nuestro caso

$$IC_{0.99}(\mu) = \left[4.38 \mp z_{0.005} \frac{0.7}{\sqrt{172}}\right] = [4.24\,,\, 4.52]\,.$$

b) La longitud del intervalo es $2 \cdot z_{\alpha/2} \frac{s}{\sqrt{n}}$. Si todos los valores permanecen iguales salvo el tamaño de muestra, tendríamos:
Longitud para $n = 100$: $2 \cdot z_{0.005} \frac{0.7}{\sqrt{100}} = 0.36$.
Longitud para $n = 300$: $2 \cdot z_{0.005} \frac{0.7}{\sqrt{300}} = 0.21$.

c) Cuando todo permace fijo, y lo único que varía es el tamaño de muestra, a mayor tamaño de muestra, mayor información, y por tanto menor incertidumbre, es decir para un mismo nivel de confianza vamos a obtener intervalos con longitud menor. Eso es lo que se refleja en los tres casos considerados (para $n = 172$, la longitud fue $4.52 - 4.24 = 0.28$).

6. X = "presencia de defectos en un periódico". $X \sim \mathcal{B}(p)$, donde p es la proporción de periódicos que se imprimen con defectos. Mediante muestreo aleatorio simple se toma una muestra donde

$$n = 100, \quad \hat{p} = \overline{x} = \frac{1}{100}\sum_{i=1}^{100} x_i = \frac{35}{100} = 0.35.$$

a) Tenemos una distribución de Bernouilli y un tamaño de muestra suficientemente grande para poder aplicar el Teorema Central del Límite, por lo tanto, el intervalo de confianza será:

$$IC_{1-\alpha}(p) = \left[\widehat{p} \mp z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right].$$

En nuestro caso,

$$IC_{0.90}(p) = \left[0.35 \mp z_{0.05}\sqrt{\frac{0.35\cdot 0.65}{100}}\right] = [0.27\,,\,0.43]\,.$$

b) El error de estimación es $z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$. Si utilizamos la información muestral, es decir, suponemos que $\hat{p}$ va a valer aproximadamente 0.35 en cualquier muestra que tomemos, entonces a un nivel de confianza del 90 % tenemos que

$$z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq 0.05 \iff z^2_{\alpha/2}\,\frac{\widehat{p}(1-\widehat{p})}{n} \leq 0.05^2 \iff z^2_{\alpha/2}\,\frac{\widehat{p}(1-\widehat{p})}{0.05^2} \leq n$$

$$\iff 1.645^2\,\frac{0.35\cdot 0.65}{0.05^2} \leq n \iff n \geq 1.645^2\,\frac{0.35\cdot 0.65}{0.05^2} = 246.25.$$

El tamaño de muestra mínimo necesario para obtener un error de estimación de a lo sumo el 5 % sería de 247.

c) En este caso no podemos suponer que $\hat{p}$ va a valer aproximadamente 0.35 en cualquier muestra, y por tanto como desconocemos $\hat{p}$ hemos de ponernos en el caso más desfavorable, es decir, cuando es igual a 1/2. Entonces a un nivel de confianza del 90 % tenemos que

$$z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq 0.05 \iff z^2_{\alpha/2}\,\frac{\widehat{p}(1-\widehat{p})}{n} \leq 0.05^2 \iff z^2_{\alpha/2}\,\frac{\widehat{p}(1-\widehat{p})}{0.05^2} \leq n$$

$$\iff 1.645^2\,\frac{\widehat{p}(1-\widehat{p})}{0.05^2} \leq n \iff n \geq 1.645^2\,\frac{0.25}{0.05^2} = 270.60.$$

El tamaño de muestra mínimo necesario para obtener un error de estimación de a lo sumo el 5 % sería en este caso de 271.

7. X = "inflación en el año en curso". Mediante muestreo aleatorio simple se toma una muestra donde

$$n = 7, \quad \overline{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = 2.36, \quad s = \sqrt{\frac{1}{6}\left(\sum_{i=1}^{7} x_i^2 - 7\cdot\overline{x}^2\right)} = 0.60.$$

Puesto que no nos dicen nada sobre la distribución de X y que el tamaño de muestra no es suficientemente grande para poder aplicar el TCL, debemos suponer que $X \sim N(\mu, \sigma)$.

a) Asumiendo normalidad en la población, la cantidad pivotal es

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

y el intervalo de confianza para μ es

$$IC_{1-\alpha}(\mu) = \left[\bar{x} \mp t_{n-1;\alpha/2}\,\frac{s}{\sqrt{n}}\right].$$

En nuestro caso

$$IC_{0.99}(\mu) = \left[2.36 \mp t_{6;0.005}\,\frac{0.6}{\sqrt{7}}\right] = \left[2.36 \mp 3.707\,\frac{0.6}{\sqrt{7}}\right] = [1.52\,,\,3.20]\,.$$

b) Asumiendo normalidad en la población, la cantidad pivotal es

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}$$

y el intervalo de confianza para σ^2 es

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

En nuestro caso

$$IC_{0.99}(\sigma^2) = \left[\frac{6 \cdot 0.6^2}{\chi^2_{6;0.005}}, \frac{6 \cdot 0.6^2}{\chi^2_{6;0.995}}\right] = [0.12\,,\,3.19]\,.$$

El intervalo de confianza para σ es: $IC_{0.99}(\sigma) = [0.34\,,\,1.79]\,.$

c) La longitud del intervalo es $2 \cdot t_{n-1;\alpha/2}\,\frac{s}{\sqrt{n}}$. Al nivel de confianza 0.95, Si queremos un intervalo con longitud de a lo sumo 1, con el mismo tamaño de muestra:

$$2t_{6;\alpha/2}\,\frac{s}{\sqrt{n}}1 \iff t_{6;\alpha/2}\,\frac{s}{\sqrt{n}} \leq 0.05 \iff t_{6;\alpha/2} \leq 0.05\frac{\sqrt{n}}{s} = 0.5\frac{\sqrt{7}}{0.6} = 2.21$$

$$t_{6;\alpha/2} = 2.21 \iff \alpha/2 \in [0.025, 0.05] \iff \alpha \in [0.05, 0.1] \iff 1-\alpha \in [0.9, 0.95]$$

Por tanto, para garantizar que la longitud sea a lo sumo 1, eligiremos un nivel de confianza del 90 %.

8. X = "intención de compra del producto en Andalucía". $X \sim \mathcal{B}(p_1)$.
Y = "intención de compra del producto en Galicia". $Y \sim \mathcal{B}(p_2)$.
Mediante muestreo aleatorio simple se toman dos muestras donde

$$n_1 = 100, \quad \hat{p_1} = \overline{x} = \frac{1}{100}\sum_{i=1}^{100} x_i = \frac{49}{100} = 0.49.$$

$$n_2 = 200, \quad \hat{p_2} = \overline{y} = \frac{1}{200}\sum_{i=1}^{200} y_i = \frac{33}{200} = 0.165.$$

a) Tenemos una distribución de Bernouilli y un tamaño de muestra suficientemente grande para poder aplicar el Teorema Central del Límite, por lo tanto, el intervalo de confianza será:

$$IC_{1-\alpha}(p_1) = \left[\widehat{p_1} \mp z_{\alpha/2}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n}}\right].$$

En nuestro caso,

$$IC_{0.90}(p_1) = \left[\widehat{p_1} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p_1}(1-\widehat{q_1})}{n}}\right] = \left[0,49 \pm 1,645\sqrt{\frac{0,49 * 0,51}{100}}\right] \approx [0.41; 0.57]\,.$$

b) Suponiendo que las dos variables son independientes, como tenemos tamaños de muestra grandes, el intervalo de confianza será:

$$IC_{1-\alpha}(p_1 - p_2) = \left[\widehat{p_1} - \widehat{p_2} \mp z_{\alpha/2}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}\right].$$

En nuestro caso,

$$IC_{1-\alpha}(p_1-p_2)=\left[0.49-0.165\pm 1.96\sqrt{\frac{0.49\cdot 0.51}{100}+\frac{0.165\cdot 0.835}{200}}\right]=[0.214\,,\,0.436]\,.$$

Por la dualidad entre intervalos de confianza y contrastes de hipótesis bilaterales (en este caso $H_0\,:\,p_1-p_2=0\quad H_1\,:\,p_1-p_2\neq 0$) tenemos que, como $0\notin[0.214\,,\,0.436]\Longrightarrow$ Rechazamos la afirmación de los dirigentes para $\alpha=0.05$.

9. X = "medida de polución atmosférica". Mediante muestreo aleatorio simple se toma una muestra donde

$$n=10,\quad \overline{x}=\frac{1}{10}\sum_{i=1}^{10}x_i=3.04,\quad s^2=\frac{1}{6}\left(\sum_{i=1}^{7}x_i^2-7\cdot\overline{x}^2\right)=1.4093.$$

Puesto que no nos dicen nada sobre la distribución de X y que el tamaño de muestra no es suficientemente grande para poder aplicar el TCL, debemos suponer que $X\sim N(\mu,\sigma)$.

a) El intervalo de confianza $1-\alpha$ para σ^2 viene dado por:

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

En nuestro caso $\overline{x}=3.04, s^2=\frac{1}{9}\sum_{i=1}^{10}(x_i-\overline{x})^2=1.4093$, $n=10$, $\chi^2_{9;0.025}=19.02$ y $\chi^2_{9;1-0.025}=2.70$. Por tanto, $I_{95\,\%}(\sigma^2)=[0.67,4.70]$.

b) Y = "medida de polución atmosférica en la segunda ciudad". Mediante muestreo aleatorio simple se toma una muestra donde

$$n=12,\quad s_2=1.5$$

. De nuevo, tenemos que asumir normalidad: $Y\sim N(\mu_2,\sigma_2)$. Denotaremos por μ_1 y σ_1 a la esperanza y la desviación típicas de X respectivamente. Asumiendo normalidad entre las dos poblaciones, el intervalo de confianza $1-\alpha$ para el cociente de varianzas viene dado por:

$$IC_{1-\alpha}(\sigma_1^2/\sigma_2^2)=\left[\frac{s_1^2/s_2^2}{F_{n_1-1,n_2-1;\alpha/2}},\frac{s_1^2}{s_2^2}F_{n_2-1,n_1-1;\alpha/2}\right]$$

En nuestro caso:

$$IC_{0.9}(\sigma_1^2/\sigma_2^2)=\left[\frac{1.4093/1.5^2}{F_{9,11;0.05}},\frac{1.4093}{1.5^2}F_{11,9;0.05}\right]=[0.22\,,\,1.94].$$