LEC/LADE/LECD/LADED
CURSO 2007/08
SOLUCIONES HOJA DE PROBLEMAS 3
INTERVALOS DE CONFIANZA

Problema 1.

(a) Se trata del caso de una variable aleatoria binomial (caso de las proporciones). En este caso,

$$n_1 = 100 \quad \widehat{p}_1 = \tfrac{49}{100} \quad \alpha = 0,10 \Longrightarrow \tfrac{\alpha}{2} = 0,05 \Longrightarrow Z_{\frac{\alpha}{2}} = 1,645$$

IC del 90% $\equiv \left(\widehat{p}_1 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}_1 \widehat{q}_1}{n}}\right) \equiv \left(0,49 \pm 1,645 \sqrt{\frac{0,49*0,51}{100}}\right) \approx (0,4077; 0,5722)$.

(b) Por dualidad intervalos-contrastes, como $0,5$ pertenece al intervalo de confianza del 90%, también pertenecerá al del 95%. Por tanto, no hay evidencia estadística para rechazar $H_0 : p = 0,5$ para un nivel de significación del 5%

(c) $\begin{array}{ll} n_1 = 100 & \widehat{p}_1 = \frac{49}{100} = 0,49 \\ n_2 = 200 & \widehat{p}_2 = \frac{33}{200} = 0,165 \end{array} \quad \alpha = 0,05 \Longrightarrow \frac{\alpha}{2} = 0,025 \Longrightarrow Z_{\frac{\alpha}{2}} = 1,96$

IC del 95%
$\equiv \left(\widehat{p}_1 - \widehat{p}_2 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}_1 \widehat{q}_1}{n} + \frac{\widehat{p}_2 \widehat{q}_2}{n}}\right) \equiv \left(0,49 - 0,165 \pm 1,96 \sqrt{\frac{0,49*0,51}{100} + \frac{0,165*0,835}{200}}\right) \equiv$
$\equiv (0,325 \pm 0,1106) \equiv (0,214; 0,436)$
Por dualidad intervalos-contrastes:
Como $0 \notin (0,206; 0,443) \Longrightarrow$ Rechazamos la afirmación de los dirigentes con $\alpha = 0,05$

Problema 2.

a) $\alpha = 0.02$

$\mu \in \left[\overline{X} \pm t_{n-1,1-\alpha/2} \frac{\widehat{s}}{\sqrt{n}}\right] = [25.6 - 2.624 \cdot \frac{2.3}{\sqrt{15}}] = [24.042, 27.158]$

Suponiendo normalidad del contenido de nicotina de los cigarrillos y una m.a.s.

b) No, porque está centrado en $\overline{X}$.

c) Con $\sigma = 2.5$, $\mu \in \left[\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$

Longitud$=2z_{0.025} \frac{\sigma}{\sqrt{n}} = 4 \Rightarrow 4z_{0.025}^2 \frac{\sigma^2}{n} = 16 \Rightarrow n = \frac{4}{16}(2.5)^2(1.96)^2 = 6.002\,5 \approx 6$

Problema 3.

a) $\widehat{p} = \frac{69}{230} = 0.3$.

$\alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow z_{\frac{\alpha}{2}} = 1.645.$

$n = 230.$

Para n suficientemente grande $\frac{\sqrt{n}(\widehat{p}-p)}{\sqrt{\widehat{p}(1-\widehat{p})}} \sim N(0,1)$.

$\widehat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 0.3 \pm 1.645 \times 0.0302 = 0.3 \pm 0.049679.$

Con lo que el intervalo resultante es $(0.2503, 0.3497)$.

b) $longitud = 0.357 - 0.243 = 0.114.$

$0.114 = 2z_{\frac{\alpha}{2}} \sqrt{\frac{0.3\times(1-0.3)}{230}} \Rightarrow z_{\frac{\alpha}{2}} = 1.8874 \Rightarrow \frac{\alpha}{2} = 1 - 0.9706 \Rightarrow 1-\alpha = 0.9412.$

Se trata de un intervalo al 94.12%.

c) $longitud = 2z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$

El intervalo de confianza alcanzará su longitud máxima cuando $\widehat{p} = \frac{1}{2}$, de modo que calcularemos n para este caso que es el más desfavorable. Así:

$longitud = 2z_{\frac{\alpha}{2}} \sqrt{\frac{1}{4n}}.$

$1-\alpha = 0.9 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow z_{\frac{\alpha}{2}} = 1.645.$

$0.15 = 2 \times 1.645 \sqrt{\frac{1}{4n}} \Rightarrow \sqrt{n} = \frac{1.645}{0.15} = 10.967 \Rightarrow n = 10.967^2 = 120.28.$

Por tanto se tomaría $n = 121$.

Problema 4.

a) (1 punto) Intervalo de confianza al $100(1-\alpha)\%$ para p:

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

(Supuestos necesarios: muestra aleatoria simple y tamaño muestral "grande").

Quantities needed:
$\hat{p} = 0.43$, $\alpha = 0.10$, $z_{0.025} = 1.96$, $n = 400$. We have

$$\left[0.43 \mp 1.96\sqrt{\frac{0.43(1-0.43)}{400}}\right] = [0.43 \mp 1.96(0.025)] = [0.43 \mp 0.049] = [0.381, 0.479]$$

b) (0.5 puntos) No es correcto. No hay nada aleatorio en p, es un valor fijo que puede o no estar contenido en el intervalo. Una vez que la muestra es seleccionada no podemos hablar de probabilidades.

c) (0.5 puntos) Se tiene

$$length = 2z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{1}{\sqrt{n}}2z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}$$

Así, la mitad del intervalo es:

$$\frac{length}{2} = \frac{1}{2}\frac{1}{\sqrt{n}}2z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})} = \frac{1}{\sqrt{4n}}2z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}$$

Incrementando el tamaño muestral por 4, se puede reducir la longitud a la mitad, así que necesitamos aumentarlo a $4 \cdot 400 = 1600$ observaciones.

Problema 5.

a) $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 114.85.$

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) = 142.14.$

b) Suponemos normalidad en la población.

$\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1} \Rightarrow I.C. \equiv \left(\overline{x} - t_{n-1;\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \overline{x} + t_{n-1;\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$

$s = \sqrt{142.14} = 11.92 \qquad t_{6;0.025} = 2.447$

$I.C. \equiv \left(114.85 - 2.447\frac{11.92}{\sqrt{7}}, 114.85 + 2.447\frac{11.92}{\sqrt{7}}\right) = (103.82, 125.88).$

c) Suponiendo de nuevo normalidad en la población:

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \Rightarrow I.C. \equiv \left(\frac{(n-1)s^2}{\chi^2_{n-1;\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}\right)$

$\chi^2_{6;0.1} = 10.645 \qquad \chi^2_{6;0.9} = 2.20$

$I.C. \equiv (80.12, 387.65)$

d) Exactamente igual que en el apartado anterior, sólo que ahora:

$\chi^2_{6;0.05} = 12.59 \qquad \chi^2_{6;0.95} = 1.63$

$I.C.$ para $\sigma^2 \equiv (67.73, 521.63)$

$I.C.$ para $\sigma \equiv \left(\sqrt{67.73}, \sqrt{521.63}\right) \equiv (8.2298, 22.839)$

Problema 6.

a) $\widehat{p} = \frac{35}{100} = .35$, $\alpha = 0.1$

$I.C.(p) = \widehat{p} \pm Z_{\alpha/2}\sqrt{\left(\frac{\widehat{p}(1-\widehat{p})}{n}\right)} = 0.35 \pm 0.79 = (0.271, 0.429)$

Con un nivel de confianza del 90% podemos afirmar que la proporción de periódicos con defectos está entre el 27.1% y el 42.9% de los periódicos impresos.

b)

$e = 0.05$ $\alpha = 0.1$, $\widehat{p} = 0.35$

$\left|Z_{\alpha/2}\sqrt{\left(\frac{\widehat{p}(1-\widehat{p})}{n}\right)}\right| \leq e \Rightarrow \frac{Z^2_{\alpha/2}}{e^2}\ \widehat{p}(1-\ \widehat{p}) \leq n$

Sustituimos,

$n \geq \frac{1.645^2 * (0.35)(0.65)}{(0.05)^2} = 246.24$

Entonces necesitaríamos una muestra de 247 periódicos.

c) En el peor caso, si no tuviésemos información previa, $\widehat{p} = 0.5$

$n = \frac{1.645^2 * (0.5)^2}{(0.05)^2} = 270.6$

Necesitaríamos una muestra de 271 periódicos.

Problema 7.

a) $IC(p_1, 95\%) = \hat{p}_1 \pm 1.96\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}} = [0.5192, 0.5808]$

Como dicho intervalo no contiene el valor p=0.5 no podemos afirmar al nivel de significación del 5% que sólo la mitad de los encuestados eran favorables al nuevo aeropuerto.

b) Al nivel de significación del 10% el intervalo centrado en torno a la estimación puntual $\hat{p}_1 = 0.55$ será aún más pequeño, por lo que tampoco contendrá el valor p=0.5, llegándose así a la misma conclusión que en el apartado anterior.

c) Si suponemos que las muestras de encuestados en los años 2002 y 2003 son independientes, se tendrá:

$IC(p_2 - p_1.95\%) = \hat{p}_2 - \hat{p}_1 \pm 1.96\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}} =$

$= 0.6 - 0.55 \pm 1.96\sqrt{\dfrac{0.55(0.45)}{1000} + \dfrac{0.6(0.4)}{1500}} = 0.05 \pm 0.0396 = [0.0104, 0.0896].$

Como este intervalo sólo contiene valores positivos, podemos afirmar al nivel de significación del 5% que la campaña publicitaria en favor del nuevo aeropuerto pudo tener efecto.

Problema 8.

(a) . Dada la función de masa del enunciado, calculamos la función de verosimilitud para una muestra aleatoria simple $x_1, \ldots, x_n$,

$$l(x_1, \ldots, x_n|\theta) = l(\theta) = \frac{1}{\prod_{i=1}^{n} x_i!} e^{-n\theta}\theta^{x_1+\ldots+x_n}$$

Por tanto, la función soporte viene dada por:

$$L(\theta) = log(l(x_1, \ldots, x_n|\theta)) = -n\theta + \sum_{i=1}^{n} x_i log(\theta) + \text{constante.}$$

Derivamos para obtener el punto crítico:

$$\frac{dL(\theta)}{d\theta} = -n + \sum_{i=1}^{n} x_i/\theta = 0 \Rightarrow \widehat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

Comprobamos que se trata de un máximo:

$$\frac{d^2L(\theta)}{d\theta^2} = -\sum_{i=1}^{n} x_i/\theta^2 < 0, \forall\theta,$$

luego efectivamente se trata de un máximo y podemos concluir que el estimador máximo verosímil es la media muestral, $\widehat{\theta}_{MV} = \overline{x}$.

(b) Con los datos de la tabla tenemos que

$$\widehat{\theta}_{MV} = \overline{x} = \frac{0 + 10 + \ldots + 6}{40} = 2.2$$

(c) El intervalo de confianza asintótico es de la forma, $\widehat{\theta}_n \mp z_{\alpha/2}\sqrt{Var(\widehat{\theta}_n)}$. Además,

$$Var(\widehat{\theta}_n) = -(\frac{d^2L(\widehat{\theta}_n)}{d\widehat{\theta}_n^2})^{-1} = \frac{\widehat{\theta}_n^2}{\sum_{i=1}^{n} x_i} = \frac{2.2^2}{88} = 0.055$$

Sustituyendo los valores en el intervalo obtenemos:

$$IC_{95\%} = 2.2 \mp 1.96\sqrt{.055} = (1.74, 2.66)$$

Problema 9.

a) Vamos a calcular el estimador MV de λ. La función de verosimilitud es:

$$l(\lambda) = \left(\frac{1}{\Pi_{i=1}^{n} x_i!} \right) \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}.$$

La función soporte queda:

$$L(\lambda) \simeq \log(\lambda) \sum_{i=1}^{n} x_i - n\lambda.$$

La primera derivada de la función soporte queda:

$$\frac{dL(\lambda)}{d\lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n \Rightarrow \text{ si } \frac{dL(\lambda)}{d\lambda} = 0 \Rightarrow \widehat{\lambda}_{MV} = \overline{X}.$$

b)La esperanza y varianza del estimador de máxima verosimilitud son:

$$E\left(\widehat{\lambda}_{MV}\right) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \lambda,$$

$$Var\left(\widehat{\lambda}_{MV}\right) = Var\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{\lambda}{n}.$$

De las propiedades de los estimadores MV ya sabemos que:

$$\widehat{\lambda}_{MV} \overset{a}{\sim} N(\lambda, \frac{\lambda}{n}).$$

La **varianza asintótica** es:

$$Var\left(\widehat{\lambda}_{MV}\right) = \frac{\widehat{\lambda}_{MV}}{n} = \frac{\overline{X}}{n}.$$

Por consiguiente, la expresión del intervalo de confianza (aproximado) del $(1-\alpha)\%$ para λ queda:

$$\lambda \in \overline{X} \pm z_{\alpha/2} \sqrt{\frac{\overline{X}}{n}}$$

- c) Sustituyendo los datos del problema,

$$\lambda \in 3.15 \pm 1.96\sqrt{\frac{3.15}{100}}$$

es decir,

$$\lambda \in 3.15 \pm 0.347\,87$$

es decir,

$$\lambda \in (2.8021,\ \ 3.4979)$$

- d) Se rechaza al 5% de nivel de significación las hipótesis nulas que estén fuera de ese intervalo. En particular, no se rechaza que $\lambda = 3$, y sí que $\lambda = 4$.