LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2008/09
SOLUCIÓN AL EXAMEN DE ESTADÍSTICA I
TIEMPO MÁXIMO = 3h

P1. (a) El valor del estadístico del contraste viene dado por,

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{35.67-36}{0.5355/\sqrt{10}} = -1.9487.$$

El p-valor del contraste se define como

$$p - valor = P(t_9 < -1.9487) + P(t_9 > 1.9487) = 2P(t_9 > 1.9487).$$

Se puede comprobar que $P(t_9 > 1.9487) \in (0.025, 0.05)$, por tanto el p-valor $\in (0.05, 0.1)$.

(b) Se rechaza la hipótesis nula siempre que el p-valor sea menor que el nivel de significación. Por tanto, para un nivel de significación igual al 5% no se rechaza la hipótesis nula. En cambio, para un nivel menor que el 5% (en particular para el 1%) se rechaza la hipótesis nula.

(c) El intervalo de confianza viene dado por,

$$I = \left[\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}\right] = \left[\frac{9\times 0.2868}{19.02}, \frac{9\times 0.2868}{2.7}\right] = [0.1357, 0.956].$$

(d) Para un nivel de significación del 5% se rechaza la hipótesis de que la varianza es igual a 1.5, al no encontrarse dicho valor en el intervalo de confianza al 95% de la varianza poblacional.

P2. _Puntuación de cada apartado del ejercicio:_ 0.25 puntos si los pasos a seguir son correctos; 0.25 puntos por la fórmula del estadístico del contraste; 0.25 puntos por los grados de libertad correctos; 0.25 por las conclusiones; 0.25 puntos por obtener los valores correctos.

(a) El contraste se realiza como sigue. En primer lugar los valores observados son:

$$\begin{array}{lll} O_{11} = 200, & O_{12} = 75, & O_{13} = 25, \\ O_{21} = 150, & O_{22} = 125, & O_{23} = 75, \\ O_{31} = 100, & O_{32} = 150, & O_{33} = 100, \end{array}$$

mientras que los valores esperados son:

$$E_{11} = \frac{300\times 450}{1000} = 135,\ E_{12} = \frac{300\times 350}{1000} = 105,\ E_{13} = \frac{300\times 200}{1000} = 60$$
$$E_{21} = \frac{350\times 450}{1000} = 157.5,\ E_{22} = \frac{350\times 350}{1000} = 122.5,\ E_{23} = \frac{350\times 200}{1000} = 70$$
$$E_{31} = \frac{350\times 450}{1000} = 157.5,\ E_{32} = \frac{350\times 350}{1000} = 122.5,\ E_{33} = \frac{350\times 200}{1000} = 70.$$

y, por lo tanto,

$$\chi^2 = \frac{(200-135)^2}{135} + \frac{(75-105)^2}{105} + \frac{(25-60)^2}{60} +$$
$$+\frac{(150-157.5)^2}{157.5} + \frac{(125-122.5)^2}{122.5} + \frac{(75-70)^2}{70} +$$
$$+\frac{(100-157.5)^2}{157.5} + \frac{(150-122.5)^2}{122.5} + \frac{(100-70)^2}{70} = 101.07,$$

que al ser mayor que $\chi^2_{2\times 2,0.05} = 9.48$, nos indica que al nivel de significación $\alpha = 0.05$, la dependencia es significativa.

(b) En este caso la tabla de contingencia se reduce a:

		P. Político			
		A	B	C	Total
Edad	Menor de 30 años	200	75	25	300
	Mayor de 50 años	100	150	100	350
	Total	300	225	125	650

En primer lugar los valores observados son:

$$O_{11} = 200, \quad O_{12} = 75, \quad O_{13} = 25,$$
$$O_{21} = 100, \quad O_{22} = 150, \quad O_{23} = 100,$$

mientras que los valores esperados son:

$$E_{11} = \frac{300\times 300}{650} = 138.46, \; E_{12} = \frac{300\times 225}{650} = 103.84, \; E_{13} = \frac{300\times 125}{650} = 57.69$$
$$E_{21} = \frac{350\times 300}{650} = 161.53, \; E_{22} = \frac{350\times 225}{650} = 121.15, \; E_{23} = \frac{350\times 125}{650} = 67.30,$$

y, por lo tanto,

$$\chi^2 = \frac{(200-138.46)^2}{138.46} + \frac{(75-103.84)^2}{103.84} + \frac{(25-57.69)^2}{57.69} +$$
$$+\frac{(100-161.53)^2}{161.53} + \frac{(150-121.15)^2}{121.15} + \frac{(100-67.30)^2}{67.30} = 100.08,$$

que al ser mayor que $\chi^2_{1\times 2,0.05} = 5.99$, nos indica que al nivel de significación $\alpha = 0.05$, la diferencia de proporciones es significativa.

P3. (a) 1. Función de verosimilitud: **(0.4 puntos)**

$$\mathcal{L}(x_1,\ldots,x_n;\lambda) = f(x_1,\ldots,x_n;\lambda) \overset{indp.}{=} \prod_{i=1}^{n} f_{X_i}(x_i;\lambda)$$
$$\overset{i.d.}{=} \prod_{i=1}^{n} f_X(x_i;\lambda) = \prod_{i=1}^{n} \left(\lambda^2 x_i e^{-\lambda x_i}\right) = \lambda^{2n} \cdot \prod_{i=1}^{n} x_i \cdot e^{-\lambda \sum_{i=1}^{n} x_i}$$

2. Función soporte: **(0.3 puntos)**

$$\ell(\lambda) \;=\; \ln\left(\mathcal{L}(x_1,\ldots,x_n;\lambda)\right) = 2n\ln\lambda + \sum_{i=1}^{n}\ln(x_i) - \lambda\sum_{i=1}^{n}x_i$$

3. Derivada de la función soporte y cálculo del punto crítico: **(0.25 puntos)**

$$\frac{\partial}{\partial\lambda}\ell(\lambda) = \frac{2n}{\lambda} - \sum_{i=1}^{n}x_i$$

$$\frac{\partial}{\partial\lambda}\ell(\lambda) = 0 \quad\Leftrightarrow\quad \frac{2n}{\lambda} - \sum_{i=1}^{n}x_i = 0 \Leftrightarrow \lambda = \frac{2n}{\sum_{i=1}^{n}x_i} \Rightarrow \hat{\lambda}_{MV} = \frac{2n}{\displaystyle\sum_{i=1}^{n}x_i}$$

4. Comprobamos que el valor obtenido es un máximo: **(0.3 puntos)**

$$\frac{\partial^2}{\partial\lambda^2}\ell(\lambda) = -\frac{2n}{\lambda^2} < 0 \quad \forall\lambda \Rightarrow \frac{\partial^2}{\partial\lambda^2}\ell(\lambda)\,\big|_{\lambda=\hat{\lambda}_{MV}} < 0.$$

Por tanto el estimador máximo-verosímil de λ es: $\hat{\lambda}_{MV} = \dfrac{2}{\overline{x}}$.

(b) La estimación máximo-verosímil de λ para esta muestra es $\hat{\lambda}_{MV} = \dfrac{2}{\overline{x}} = \dfrac{2}{21.1} =$ 0.0948. **(0.5 puntos)**

(c) La varianza asintótica de $\hat{\lambda}_{MV}$ es **(0.35 puntos)**:

$$Var.asint[\hat{\lambda}_{MV}] \approx -\frac{1}{\frac{\partial^2}{\partial\lambda^2}\ell(\lambda)\,\big|_{\lambda=\hat{\lambda}_{MV}}}$$

Para la muestra del apartado b) **(0.4 puntos)**,

$$\begin{aligned} Var.asint[\hat{\lambda}_{MV}] &\approx -\frac{1}{-\frac{2n}{\hat{\lambda}_{MV}^2}} = \frac{\hat{\lambda}_{MV}^2}{2n} \\ &= \frac{\left(2n/\sum_{i=1}^{n}x_i\right)^2}{2n} = \frac{2n}{\overline{x}^2} = \frac{20}{21.1^2} = 0.0449 \end{aligned}$$

P4. (a) Las expresiones generales para el valor esperado de $\bar{X}$ y para su varianza son

$$\textbf{(0,5 puntos)}E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right] = \frac{1}{n}\sum_{i=1}^{n}E[X_i] = \frac{n\mu}{n} = \mu$$

y

$$\textbf{(0,25 puntos)}V[\bar{X}] = V\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}V[X_i] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

(0,25 puntos) Por el lema de Fisher, como $\{X_1, \ldots, X_n\}$ es una muestra aleatoria y X_i sigue una distribución $N(3.200, \sqrt{200^2})$, la distribución de $\bar{X}$ es una $N(3.200, \sqrt{\frac{200^2}{10}} = 63,2455)$.

(b) Si Y es el coste del minipiso de 50 m^2, la probabilidad que se pide es

$$\begin{aligned}
P(Y > 200.000) &= P(50X > 200.000) \\
&= P(X > 4.000)\textbf{(0,25 puntos)} \\
&= P(\frac{X - 3.500}{200} > \frac{4.000 - 3.500}{200})\textbf{(0,5 puntos)} \\
&= P(Z > 2,5) \\
&= 0,0062 = 0,62\%\textbf{(0,25 puntos)}.
\end{aligned}$$

(c) Utilizando la pista con $Y = \bar{X}$,

$$\begin{aligned}
E[\bar{X}^2] &= V[\bar{X}] + E[\bar{X}]^2 \\
&= \frac{\sigma^2}{n} + \mu^2\textbf{(0,25 puntos)}
\end{aligned}$$

por lo que $E[\bar{X}^2] \neq \mu^2$**(0,25 puntos)**, lo que implica que $\bar{X}^2$ es un estimador sesgado de μ^2.