LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2007/08
EXAMEN DE ESTADÍSTICA I
12 de Septiembre de 2008
TIEMPO MÁXIMO = 3 horas

1. **(3 puntos).** Según un estudio realizado, el número de viajeros (en miles de personas) que acuden diariamente a un cierto aeropuerto es una variable aleatoria X con distribución normal de media $\mu = 2.1$ y desviación típica $\sigma = 0.8$. Se selecciona al azar una muestra aleatoria simple de 200 días y se observa el número de viajeros que acude al aeropuerto. Se pide,

   a) **(1 punto).** Deducir la distribución de probabilidad de la media muestral del número de viajeros diarios ($\bar{X}$) así como las expresiones de su valor esperado y la varianza.
   Solución: Si X es una variable aleatoria normal con media μ y desviación σ, entonces la variable aleatoria media $\bar{X}$ también es normal, por ser combinación lineal de variables normales. Para calcular su media y desviación típica,

   $$E(\bar{X}) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu = 2.1.$$

   Por independencia entre los valores de la muestra,

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

   Por tanto,

   $$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

   En este caso, $\bar{X} \sim N\left(\mu = 2.1, \frac{\sigma}{\sqrt{n}} = \frac{0.8}{\sqrt{200}} = 0.0566\right)$.

   b) **(1 punto).** ¿Cuál es la probabilidad de que la media de viajeros diarios en los 200 días supere los 2200? ¿Y de que esté comprendida entre 1900 y 2200?
   Solución: se debe calcular,

   $$P(\bar{X} > 2.2) = P\left(\frac{\bar{X} - 2.1}{0.0566} > \frac{2.2 - 2.1}{0.0566}\right) = P(Z > 1.7668) \approx 0.0384,$$

   donde $Z \sim N(0, 1)$. Para la siguiente cuestión,

   $$\begin{aligned} P\left(1.9 < \bar{X} < 2.2\right) &= P\left(\frac{1.9 - 2.1}{0.0566} < \frac{\bar{X} - 2.1}{0.0566} < \frac{2.2 - 2.1}{0.0566}\right) \\ &= P\left(-3.5336 < Z < 1.7668\right) \\ &= P(Z < 1.7668) - P(Z < -3.5336) \\ &\approx (1 - 0.0384) - P(Z > 3.5336) \approx 0.9616. \end{aligned}$$

c) **(1 punto).** Suponiendo que μ fuese desconocido, se plantea la posibilidad de considerar el estimador $\widehat{\mu} = \frac{2X_1+X_{10}+X_{200}}{4}$. Calcular la eficiencia relativa de la media muestral respecto del estimador propuesto y explique el resultado obtenido.

Solución: La eficiencia relativa de $\bar{X}$ respecto de $\hat{\mu}$ se define como

$$EFR(\bar{X}/\hat{\mu}) = \frac{Var(\hat{\mu})}{Var(\bar{X})} = \frac{Var(\hat{\mu})}{\sigma^2/n},$$

donde $Var(\hat{\mu}) = \frac{1}{4^2}Var(2X_1+X_{10}+X_{200}) = \frac{1}{16}\left(4Var(X_1) + Var(X_{10}) + Var(X_{200})\right)$. *Pero por i.i.d,* $Var(X_1) = Var(X_{10}) = Var(X_{200}) = \sigma^2$, *así*

$$Var(\hat{\mu}) = \frac{6}{16}\sigma^2,$$

y por tanto,

$$EFR(\bar{X}/\hat{\mu}) = \frac{6\sigma^2/16}{\sigma^2/200} = 1200/16 = 75,$$

lo que lleva a concluir que el estimador $\bar{X}$ is 75 veces más eficiente que el estimador $\hat{\mu}$.

2. **(3 puntos).** En una agencia inmobiliaria deciden realizar un estudio para ajustar las mejores condiciones laborales de sus nuevos comerciales. La hipótesis que se plantea es que los agentes comerciales que cobran sus horas por comisión de ventas, tienen un mayor volumen de ventas en promedio, que aquellos que tienen un sueldo fijo cada mes. Para comprobarlo la empresa seleccionó una muestra de 10 agentes inmobiliarios cuyo sueldo era fijo y una muestra independiente de 10 agentes cuyo sueldo dependía de las ventas realizadas. Los datos siguientes representan su volumen de venta en miles de euros durante un mes de trabajo,

   Agentes con sueldo fijo:

   $$23,\ 22,\ 20,\ 19,\ 21,\ 22,\ 19,\ 20,\ 21,\ 20$$

   Agentes con sueldo por comisión:

   $$20,\ 23,\ 25,\ 26,\ 22,\ 25,\ 24,\ 21,\ 21,\ 20$$

   a) **(2 puntos).** Expresando los supuestos necesarios y asumiendo que las varianzas poblacionales son iguales, ¿hay evidencia con un 95% de confianza, de que los agentes que cobran un sueldo por comisión producen una venta promedio superior a la de los que cobran un sueldo fijo?

   Solución: Asumiendo normalidad en las poblaciones y suponiendo que las muestras son m.a.s, se denotará por X la variable aleatoria que representa el volumen de ventas en miles de euros durante un mes de trabajo para los agentes con sueldo fijo, e Y lo equivalente para los agentes con sueldo por comisión.

Sean μ_x y μ_y las respectivas medias poblacionales; entonces el contraste de hipótesis a plantear es

$$\begin{aligned} H_0 &: \mu_x - \mu_y \geq 0 \\ H_1 &: \mu_x - \mu_y < 0 \end{aligned}$$

Utilizamos el estadístico de diferencias de medias para poblaciones normales con varianzas desconocidas pero iguales:

$$t = \frac{\bar{x} - \bar{y} - d_0}{s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{20.7 - 22.7 - 0}{1.83\sqrt{\frac{1}{10} + \frac{1}{10}}} = -2.44$$

donde, para calcular el valor de s_p, se ha utilizado que $s_x = 1.337$ y $s_y = 2.214$. Se rechazará H_0 cuando $t_{18,0.05} \geq -2.44$. En este caso, $t_{18,0.05} = -1.734$, por tanto se puede afirmar con un 95% de confianza que los agentes que cobran por comisión producen en término medio un mayor volumen de ventas que los que tienen un sueldo fijo.

b) **(1 punto).** Acotar superior e inferiormente el p-valor del contraste del apartado a).
Solución: El valor del estadístico del contraste es $t = -2.44$ y por tanto el p-valor es $P(t_{18} < -2.44) = P(t_{18} > 2.44)$. Como $P(t_{18} > 2.101) = 0.025$, $P(t_{18} > 2.552) = 0.01$ y $2.44 \in (2.101, 2.552)$, se concluye que el p-valor está acotado por 0.01 y 0.025.

3. **(2 puntos).** A partir de una muestra aleatoria simple de 150 bebedores de cerveza, se obtiene la siguiente tabla que refleja sus preferencias por sexos,

	Cerveza Ligera	Cerveza Clara	Cerveza Oscura
Hombre	20	40	20
Mujer	30	30	10

Con un nivel de significación igual al 5%, ¿se puede rechazar la hipótesis de independencia entre la preferencia de cerveza y el sexo del consumidor? Acotar superior e inferiormente el p-valor del contraste realizado.
Solución: El contraste a realizar es, H_0 : existe independencia entre la preferencia de cerveza y el sexo del consumidor, frente a H_1 : la preferencia de cerveza y el sexo del consumidor son dependientes. La tabla de valores esperados $E_{i,j}$, asumiendo cierta H_0 es:

	Cerveza Ligera	Cerveza Clara	Cerveza Oscura	TOTAL
Hombre	26.67	37.33	16	80
Mujer	23.33	32.67	14	70
TOTAL	50	70	30	150

Se rechaza la hipótesis nula si el valor de $\sum_i \sum_j \frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}}$ *supera el valor de* $\chi_{2,0.05} = 5.99$*. En este caso,*

$$\sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(20-26.67)^2}{26.67} + \ldots + \frac{(10-14)^2}{14} \approx 6.13.$$

Por tanto que se rechaza la hipótesis nula con nivel de significación del 5%*; no hay evidencia estadística suficiente para concluir que la preferencia de cerveza y el sexo del consumidor son independientes.*

El p-valor del contraste en este caso es $P(\chi^2_2 > 6.13)$*. Dicho valor está acotado por* $[P(\chi^2_2 > 7.38), P(\chi^2_2 > 5.99)] = [0.025, 0.05]$*.*

4. **(2 puntos).** Considere una muestra aleatoria de n observaciones $X_1, X_2, \ldots, X_n$ independientes, pero no idénticamente distribuidas, cuyas densidades son

$$f(x_i) = i\theta \exp\{-i\theta x_i\}, \quad \theta > 0, \quad x_i > 0, \quad i = 1, 2, \ldots, n.$$

Se desea estimar el parámetro desconocido θ mediante máxima verosimilitud.

a) **(0.6 puntos).** Sin realizar ningún cálculo, escriba los tres pasos principales del método de estimación puntual de máxima verosimilitud.
Solución:
Paso 1) Calcular la función de verosimilitud, $l(\theta)$*, y/o log-verosimilitud,* $L(\theta) = \ln(l(\theta))$*.*
Paso 2a) (maximizar) Despejar θ *de la ecuación*

$$\frac{dL(\theta)}{d\theta} = 0,$$

y denotar la solución $\hat{\theta}$*.*
Paso 2b) (maximizar) Verificar que

$$\frac{d^2 L(\theta)}{d\theta^2}|_{\theta=\hat{\theta}} < 0.$$

b) **(1 punto).** Encuentre el estimador máximo verosímil de θ.
Solución:

$$\begin{aligned}
l(\theta) \quad &=_{indep.} \prod_{i=1}^{n} \Big(i\theta \exp\{-i\theta x_i\} \Big) \\
&= \theta^n \cdot \prod_{i=1}^{n} i \cdot \prod_{i=1}^{n} \exp\{-i\theta x_i\} \\
&= \theta^n \cdot \prod_{i=1}^{n} i \cdot \exp\{-\theta \sum_{i=1}^{n} i x_i\}, \\
L(\theta) = \ln(l(\theta)) \quad &= n\ln(\theta) + \sum_{i=1}^{n} \ln(i) - \theta \sum_{i=1}^{n} i x_i.
\end{aligned}$$

$$\begin{aligned}
\frac{dL(\theta)}{d\theta} &= \frac{n}{\theta} - \sum_{i=1}^{n} i x_i. \\
\frac{dL(\theta)}{d\theta} &= 0 \Leftrightarrow \\
\theta &= \frac{n}{\sum_{i=1}^{n} i x_i} \\
\hat{\theta} &= \frac{n}{\sum_{i=1}^{n} i x_i}
\end{aligned}$$

$$\frac{d^2L(\theta)}{d\theta^2}|_{\theta=\hat{\theta}} = -\frac{n}{\hat{\theta}^2} < 0$$

ya que $n > 0$ y $\hat{\theta}^2 > 0$ ($\hat{\theta}^2$ no puede ser 0).

c) **(0.4 puntos).** Obtenemos las siguientes 20 observaciones $(x_1, x_2, \ldots, x_{20})$,

1.62, 0.27, 0.02, 0.19, 0.06, 0.13, 0.07, 0.06, 0.10, 0.56,
0.03, 0.08, 0.10, 0.06, 0.01, 0.04, 0.02, 0.11, 0.09, 0.02,

para las que

$$\sum_{i=1}^{20} i x_i = 20.18.$$

Basándose en dicha información muestral, ¿cuál sería el estimador máximo verosímil de θ^2 ?

Solución:

Por la propiedad de invarianza, el estimador máximo verosímil de θ^2 es

$$(\hat{\theta})^2 = \left(\frac{n}{\sum_{i=1}^{n} i x_i} \right)^2.$$

Por tanto, el estimador de θ^2 es

$$\left(\frac{20}{20.18} \right)^2 \approx 0.991^2 \approx 0.982.$$