Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

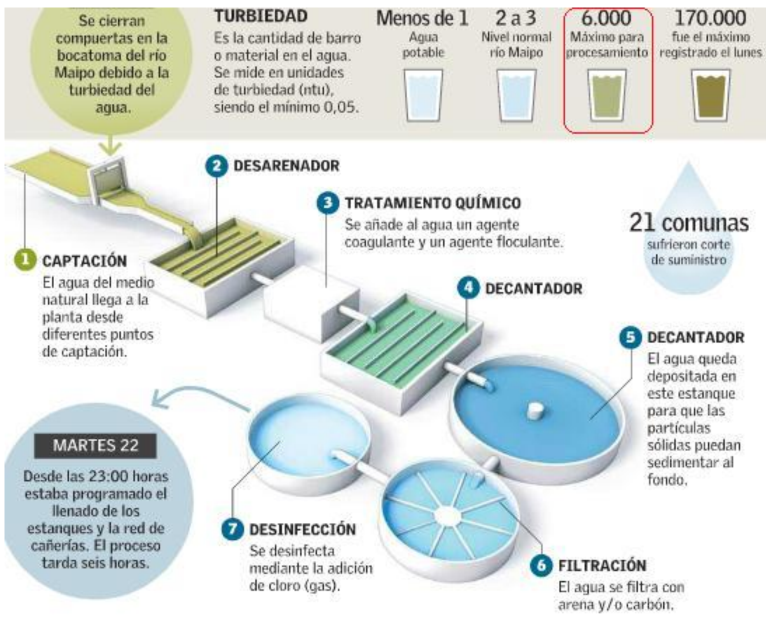
Temporada Académica de Verano 2013

Curso	:	**Probabilidad y Estadística**
Sigla	:	**EYP1113**
Interrogación	:	**3**
Profesores	:	**Ricardo Aravena (Sec. 1), Claudia Wehrhahn (Sec. 2) y Andrés Iturriaga (Sec. 3)**

PAUTA I3

Problema 1

Una empresa distribuidora de agua potable ha implementado un protocolo a objeto de monitorear y tomar decisiones respecto al procesamiento de agua que capta de un río (ver gráfica).



El departamento de control de calidad toma, cada una hora, 100 muestras (vasijas de 50 cc) en la bocatoma (1) a objeto de establecer el nivel de turbiedad del agua en la captación. Clasifica cada vasija de agua como “Turbia” o “No turbia” basado en un método rápido y sencillo (para cada vasija se calcula el cuociente entre el peso de los 50 cc obtenido en la bocatoma y el peso control de 50 cc de agua, si es mayor a 1,2 se clasifica como “turbia”). En caso de obtener, estadísticamente, una proporción mayor a un tercio de las vasijas como “turbias”, se cierran las compuertas de captación (pero continua el procesamiento de agua de los estanques).
Por otra parte, cada 4 horas levanta un muestreo de tamaño 25 en el estanque decantador (4). Estas muestras son analizadas en el laboratorio a fin de obtener el nivel medio de turbiedad (en ntu). Si el nivel medio, estadísticamente, no cumple la especificación máxima para su procesamiento, se procede a su detención. Basado en antecedentes históricos, usted encuentra para un día (21/ene) lo que sigue:

a) **[2 puntos]** de las muestras tomada a las 18.00 horas en la bocatoma(1), 42 de ellas fueron clasificadas como “Turbias”, ¿qué decisión debió tomarse? (use $\alpha = 1\,\%$). Sea explícito: Hipótesis, test, supuestos.

b) **[3 puntos]** las muestras tomadas a las 20.00 horas en el estanque decantador(4), presentan un promedio de turbiedad de 6.800 ntu y una desviación estándar de 1.800 ntu. Asumiendo normalidad respecto al nivel de turbiedad, ¿cuál debió ser la decisión frente a estos resultados? (use $\alpha = 5\,\%$). Sea explícito: hipótesis, test, supuestos. Obtenga el valor-p.

c) **[1 puntos]** Asumiendo σ=1.800 ntu, ¿cuántas muestras deben tomarse en el estanque decantador a objeto de estimar, con un 95 % de confianza, el nivel medio de turbiedad con un error no superior a 600 ntu?

Solución

a) Sea P = proporción de vasijas clasificadas como "Turbias". Entonces corresponde a un test sobre proporciones. Es decir, la distribución subyacente es Bernoulli de parámetro P. **[0,2]**

$$Hipotesis: H_0: P = 1/3 \quad \text{vs.} \quad H_a: P > 1/3 \qquad [\mathbf{0,6}]$$

$$test: z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,4 - 0,333}{\sqrt{\frac{0,333 \cdot 0,667}{100}}} = \frac{0,067}{0,0471} = 1,42 \qquad [\mathbf{0,6}]$$

Como z_{calc} = 1,42 no es mayor que z_α = 2,326 no hay evidencia en contra la hipótesis nula. **[0,3]**
Es decir, la decision debió ser NO CERRAR LAS COMPUERTAS. **[0,3]**

b) Sea μ: Nivel medio de turbiedad. Supuestos: los datos subyacentes son normales con desviación estándar desconocida. Test sobre μ **[0,4]**.

$$Hipotesis: H_0: \mu = 6000 \quad \text{vs.} \quad H_a: \mu > 6000 \qquad [\mathbf{0,8}]$$

$$Test: t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{6800 - 6000}{1800/\sqrt{25}} = \frac{800}{360} = 2,22 \qquad [\mathbf{0,8}]$$

Como t_{calc} = 2,22 es mayor que $t_{5\,\%}(25-1)$ = 1,711 hay evidencia en contra la hipótesis nula. **[0,3]**
Es decir, la decision debió ser DETENER EL PROCESAMIENTO DE AGUA. **[0,2]**
valor − p = $P(T(24) > 2,22)$, mirando e interpolando en la tabla, se tiene:

$$1,0\,\% < valor - p < 2,5\,\% \qquad [\mathbf{0,5}]$$

c) En este caso, se tiene σ=1.800 ntu, w=600 ntu, y $z_{1-\alpha/2} = z_{\alpha/2} = 1,96$, entonces: **[0,3]**

$$n = \left(\frac{z_{\alpha/2}\sigma}{w}\right)^2 = \left(\frac{1,96 \cdot 1800}{600}\right)^2 = (5,88)^2 = 34,6 \qquad [\mathbf{0,6}]$$

Es decir, se deben tomar (como mínimo) 35 muestras. **[0,1]**

Problema 2

Suponga que en un determinado paradero los buses del transantiago llegan de acuerdo a un proceso de Poisson y que usualmente llegan 6 buses en una hora. Sean X_i, para $i = 1, \ldots, 10$, los tiempos entre llegada de los primeros 10 buses. El objetivo de este problema es determinar el valor esperado (en minutos) de la diferencia entre el tiempo máximo y mínimo de las llegadas entre buses, esto es,

$$E\left(X_{(10)} - X_{(1)}\right)$$

donde $X_{(10)} = \text{máx}\{X_1, \ldots, X_{10}\}$ y $X_{(1)} = \text{mín}\{X_1, \ldots, X_{10}\}$.

(a) **[2.5 puntos]** Obtenga la función de distribución acumulada, la función de densidad de probabilidad y valor esperado de $X_{(1)}$.

(b) **[2.5 puntos]** Determine la función de distribución acumulada de $X_{(10)}$. Usando que el valor esperado para una variable aleatoria positiva Y se puede calcular con la fórmula

$$E(Y) = \int_0^\infty P(Y > y)dy,$$

calcule el valor esperado de $X_{(10)}$.
Indicación: Puede ser útil recordar que $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$.

(c) **[1 puntos]** Calcule $E\left(X_{(10)} - X_{(1)}\right)$.

Solución

Notemos que $X_1, \ldots, X_{10} \overset{\text{iid}}{\sim} Exp(\lambda)$ con $\lambda = \frac{6}{60} = \frac{1}{10}$.

(a) Comencemos calculando la función de distribución acumulada de $X_{(1)}$. En efecto,

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - P(X_1 > x, \ldots, X_{10} > x) = 1 - [P(X_1 > x)]^{10} \\ &= 1 - e^{-10\lambda x}, x > 0. \qquad \textbf{[1.0pts]} \end{aligned}$$

Entonces la función de densidad de probabilidad es:

$$f_{X_{(1)}} = \frac{dF_{X_{(1)}}(x)}{dx} = 10\lambda e^{-10\lambda x}, x > 0. \qquad \textbf{[1.0pts]}$$

Por lo tanto, $X_{(1)} \sim Exp(10\lambda)$ y $E(X_{(1)}) = \frac{1}{10\lambda} = 1$. **[0.5pts]**

(b) Comencemos calculando la función de distribución acumulada de $X_{(10)}$. En efecto,

$$\begin{aligned} F_{X_{(10)}}(x) &= P(X_{(10)} \leq x) = P(X_1 \leq x, \ldots, X_{10} \leq x) = [P(X_1 \leq x)]^{10} \\ &= \left[1 - e^{-\lambda x}\right]^{10}, x > 0. \qquad \textbf{[1.0pts]} \end{aligned}$$

Usando la fórmula para el valor esperado de una v.a. positiva se tiene que

$$E(X_{(10)}) = \int_0^\infty \left(1 - \left[1 - e^{-\lambda x}\right]^{10}\right) dx. \qquad \textbf{[0.5pts]}$$

Notemos que

$$1 - \left[1 - e^{-\lambda x}\right]^{10} = 1 - \sum_{k=0}^{10} \binom{10}{k} (-1)^k e^{-k\lambda x} = \sum_{k=1}^{10} \binom{10}{k} (-1)^{k+1} e^{-k\lambda x}.$$

Por lo tanto,

$$\begin{aligned} E(X_{(10)}) &= \sum_{k=1}^{10} \binom{10}{k} (-1)^{k+1} \int_0^\infty e^{-k\lambda x} dx = \sum_{k=1}^{10} \binom{10}{k} \frac{(-1)^{k+1}}{k\lambda} \int_0^\infty k\lambda e^{-k\lambda x} dx \qquad \textbf{[0.5pts]} \\ &= \sum_{k=1}^{10} \binom{10}{k} \frac{(-1)^{k+1}}{k\lambda} \approx 29{,}28968. \qquad \textbf{[0.5pts]} \end{aligned}$$

(c) Finalmente,

$$E\left(X_{(10)} - X_{(1)}\right) = E(X_{(10)}) - E(X_{(1)}) = 29{,}28968 - 1 = 28{,}28968. \qquad \textbf{[1.0pts]}$$

Problema 3

Los chilenos, fanáticos del fútbol, quieren que la selección chilena participe en el mundial de fútbol Brasil 2014. Actualmente es Sampaoli quien está a cargo de esta misión y sabe que debe contar con los mejores jugadores para lograr este objetivo. Una de las características que espera que tengan los jugadores de su plantel es que el cansancio no afecte su rendimiento y se ha preocupado de contar con jugadores que cumplan este requerimiento. Se sabe que los kilómetros que corre un jugador de fútbol durante los 90 minutos de un partido tiene función de densidad dada por

$$f_X(x,\theta) = \begin{cases} \frac{x+1}{\theta(\theta+1)} \exp\left\{-\frac{x}{\theta}\right\}, & x > 0,\, \theta > 0; \\ 0 & \text{en otro caso} \end{cases}$$

En el partido de Chile v/s Haití se monitorearon n jugadores, es decir $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e identicamente distribuidas con función densidad dada por la expresión anterior.

(a) **[1.5 puntos]** Mostrar que la media de los kilómetros corridos por un jugador de fútbol durante un partido es $\dfrac{\theta(2\theta+1)}{\theta+1}$.

(b) **[1.5 puntos]** Se propone el siguiente estimador para θ, $\widehat{\theta}_1 = \left(\frac{\theta+1}{2\theta+1}\right)\left(\frac{1}{n-1}\right)\sum_{i=1}^{n} X_i$. Determinar si es o no un estimador insesgado o asintóticamente insesgado. Si no es insesgado, encontrar su sesgo.
Recordar que el sesgo de un estimador es Sesgo$(\widehat{\theta}) = E(\widehat{\theta}) - \theta$.

(c) **[1.5 puntos]** Encontrar el estimador de máxima verosimilitud de θ y su distribución asintótica.

(d) **[1.5 puntos]** Encontrar el EMV de la cantidad media de kilómetros que corre un jugador y su distribución asintótica.
Observación: Si a es una constante conocida, entonces $a = \frac{\theta(2\theta+1)}{\theta+1} \Leftrightarrow \theta = \frac{a-1+\sqrt{a^2-2a+9}}{4}$

Solución

(a) Nos piden obtener la experanza de X

$$\begin{aligned} E(X) &= \int_0^\infty \frac{1}{\theta(\theta+1)} x(x+1)e^{-x/\theta}dx \qquad \textbf{[0.5pts]} \\ &= \frac{1}{\theta(\theta+1)}\left[\int_0^\infty x^2 e^{-x/\theta}dx + \int_0^\infty x e^{-x/\theta}dx\right] \\ &= \frac{1}{\theta(1+\theta)}\left[\frac{\Gamma(3)}{(1/\theta)^3} + \frac{\Gamma(2)}{(1/\theta)^2}\right] \qquad \textbf{[0.5pts]} \\ &= \frac{1}{\theta(1+\theta)}[2\theta^3+\theta^2] = \frac{\theta(2\theta+1)}{\theta+1} \qquad \textbf{[0.5pts]} \end{aligned}$$

(b) Calculamos $E(\widehat{\theta}_1)$

$$\begin{aligned} E(\widehat{\theta}_1) &= E\left(\left(\frac{\theta+1}{2\theta+1}\right)\left(\frac{1}{n-1}\right)\sum_{i=1}^{n} X_i\right) \\ &= \left(\frac{\theta+1}{2\theta+1}\right)\left(\frac{1}{n-1}\right)\sum_{i=1}^{n} E(X_i) \qquad \textbf{[0.5pts]} \\ &= \left(\frac{\theta+1}{2\theta+1}\right)\left(\frac{1}{n-1}\right) n\frac{\theta(2\theta+1)}{\theta+1} \\ &= \frac{n}{n-1}\theta \qquad \textbf{[0.5pts]} \end{aligned}$$

Como $\lim_{n\to\infty} E(\widehat{\theta}_1) = \theta$ concluimos que $\widehat{\theta}_1$ es un estimador asintóticamente insesgado para θ. **[0.5pts]**

(c) Encontramos el EMV de θ. Para esto

$$\begin{aligned}
L(x_1,\ldots,x_n,\theta) &= \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \frac{(x_i+1)}{\theta(\theta+1)} e^{-x_i/\theta} \\
&= [\theta(\theta+1)]^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i} \prod_{i=1}^{n}(x_i+1) \ \textbf{[0.25pts]} \\
\Rightarrow \ln(L(x_1,\ldots,x_n,\theta)) &= -n\ln(\theta) - n\ln(\theta+1) + \sum_{i=1}^{n}\ln(x_1+1) - \frac{1}{\theta}\sum_{i=1}^{n} x_i \\
\Rightarrow & \quad \frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} - \frac{n}{\theta+1} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0 \ \textbf{[0.25pts]} \\
\Rightarrow & \quad \frac{n}{\theta} + \frac{n}{\theta+1} = \frac{1}{\theta^2}\sum_{i=1}^{n} x_i \\
\Rightarrow & \quad 2n\theta^2 + \theta(n - \sum_{i=1}^{n} x_i) - \sum_{i=1}^{n} x_i = 0 \qquad \textbf{[0.125pts]} \\
\Rightarrow & \quad \widehat{\theta} = \frac{\overline{x} - 1 \pm \sqrt{\overline{x}^2 + 6\overline{x} + 1}}{4} \qquad \textbf{[0.125pts]}
\end{aligned}$$

Ahora calculamos la información de Fisher:

$$\begin{aligned}
I_n(\theta) &= -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right) = -E\left(\frac{n}{\theta^2} + \frac{n}{(\theta+1)^2} - \frac{2\theta\sum_{i=1}^{n} X_i}{\theta^4}\right) \\
&= -\frac{n}{\theta^2} - \frac{n}{(\theta+1)^2} + \frac{2\sum_{i=1}^{n} E(X_i)}{\theta^3} = -\frac{n}{\theta^2} - \frac{n}{(\theta+1)^2} + \frac{2}{\theta^3}\frac{n\theta(2\theta+1)}{\theta+1} \ \textbf{[0.25pts]} \\
&= \frac{-n(\theta+1)^2 - n\theta^2 + 2n(2\theta+1)(\theta+1)}{\theta^2(\theta+1)^2} \\
&= \frac{-n\theta^2 - 2n\theta - n - n\theta^2 + 2n(2\theta^2+3\theta+1)}{\theta^2(\theta+1)^2} = \frac{2n\theta^2 + 4n\theta + n}{\theta^2(\theta+1)^2} \ \textbf{[0.25pts]}
\end{aligned}$$

Luego el EMV tiene distribución asintótica

$$\widehat{\theta} \dot{\sim} N\left(\theta, \sqrt{\frac{\theta^2(\theta+1)^2}{2n\theta^2 + 4n\theta + n}}\right) \qquad \textbf{[0.25pts]}$$

(d) La cantidad media de metros que corre un jugador en un partido es $E(X) = \frac{\theta(2\theta+1)}{\theta+1} = g(\theta)$ y por invarianza del EMV, un estimador de máxima verosimilitud para $E(X)$ es

$$\widehat{E(X)} = \frac{\widehat{\theta}(2\widehat{\theta}+1)}{\widehat{\theta}+1} \quad \textbf{[0.3pts]}$$

con $\widehat{\theta}$ el EMV encontrado en (c).
La distribución asintótica es $\widehat{E(X)} \dot{\sim} N\left(g(\theta), \sqrt{\frac{[g'(\theta)]^2}{I_n(\theta)}}\right)$ **[0.3pts]** con

$$\begin{aligned}
g(\theta) &= \frac{\theta(2\theta+1)}{\theta+1} \\
I_n(\theta) &= \frac{2n\theta^2 + 4n\theta + n}{\theta^2(\theta+1)^2} \\
g'(\theta) &= \frac{[4\theta+1][\theta+1] - 2\theta^2 + \theta}{(\theta+1)^2} = \frac{2\theta^2 + 4\theta + 1}{(\theta+1)^2} \ \textbf{[0.6pts]}
\end{aligned}$$

Por lo tanto la distribución asintótica de $\widehat{E(X)}$ es

$$\widehat{E(X)} \dot{\sim} N\left(\frac{\theta(2\theta+1)}{\theta+1}, \sqrt{\frac{\theta^2(2\theta^2+4\theta+1)}{n(\theta+1)^2}}\right) \ \textbf{[0.3pts]}$$