Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Temporada Académica de Verano 2017

Curso	**:**	**Probabilidad y Estadística**
Sigla	**:**	**EYP1113**
Pauta	**:**	**I3**
Profesores	**:**	**Ricardo Aravena C. y Ricardo Olea O.**

Problema 1

Durante el presente mes nos hemos visto enfrentado a olas de calor en la zona central de Chile. Suponga que las temperaturas, durante las horas de sol día se comportan como variables aleatorias Exponenciales(ν) trasladadas en α. Si se toman n mediciones en un día y definimos como X e Y al mínimo y máximo registrado:

(a) **[2.0 Puntos.]** Determine la distribución de la amplitud registrada $Z = Y - X$ en un día. (Por simplicidad considere $n = 2$).

(b) **[2.0 Puntos.]** Calcule de manera aproximada la probabilidad que la mínima temperatura registrada diaria en los último 20 días en promedio superará los 20°. Asuma independencia entre los días y entre las mediciones. Considere 5 mediciones diarias, $\nu = 0.01$ y $\alpha = 5^\circ$.

(c) **[2.0 Puntos.]** Si la temperatura máxima día durante el mes de enero se comporta como Normal($\mu = 35^\circ$, $\sigma = 2^\circ$), la cual tiene una correlación igual a 0.5 con la temperatura máxima del día anterior, determine la probabilidad en dos días sucesivos durante este mes, la temperatura máxima no presente diferencias mayores a un grado.

Solución

(a) Tenemos que la conjunta del mínimo y máximo está dada por

$$f_{X,Y}(x,y) = 2\,(2-1)\,[F_X(y) - F_X(x)]^{2-2}\, f_X(y)\, f_X(x), \qquad x \leq y$$

Reemplazando

$$f_{X,Y}(x,y) = 2\,\nu^2\, e^{-\nu\,(x+y-2\,\alpha)}, \qquad x \leq y \quad \textbf{[0.4 Ptos.]}$$

Sea $Z = Y - X$, cuya función de densidad está dada por

$$\begin{aligned} f_Z &= \int_{z+\alpha}^{+\infty} 2\,\nu^2\, e^{-\nu\,(z+y+y-2\,\alpha)}\, dy \quad \textbf{[0.4 Ptos.]} \\ &= \nu\, e^{-\nu\, z} \int_{\alpha+z}^{+\infty} 2\,\nu\, e^{-2\,\nu\,\{y-(\alpha+z)\}}\, dy \quad \textbf{[0.4 Ptos.]} \\ &= \nu\, e^{-\nu\, z} \cdot 1, \quad z > 0 \quad \textbf{[0.4 Ptos.]} \end{aligned}$$

Por lo tanto,

$$Z \sim \text{Exponencial}(\nu) \quad \textbf{[0.4 Ptos.]}$$

(b) del formulario se tiene que

$$f_X(x) = n\,\nu\, e^{-n\,\nu\,(x-\alpha)}, \qquad x > \alpha$$

es decir, X distribuye Exponencial($n\,\nu$) trasladada en α **[0.5 Ptos.]**, donde

$$\textbf{[0.2 Ptos.]} \quad \mu_X = \frac{1}{n\,\nu} + \alpha \qquad \text{y} \qquad \sigma_X = \frac{1}{n\,\nu} \quad \textbf{[0.2 Ptos.]}$$

Luego

$$\overline{X}_m \overset{\text{aprox}}{\sim} \text{Normal}\left(\frac{1}{n\,\nu} + \alpha,\, \frac{1}{n\,\nu\,\sqrt{m}}\right) \quad \textcolor{red}{\textbf{[0.3 Ptos.]}}$$

$$\to \overline{X}_{20} \overset{\text{aprox}}{\sim} \text{Normal}\left(25,\, \frac{20}{\sqrt{20}}\right) \quad \textcolor{red}{\textbf{[0.2 Ptos.]}}$$

Se pide

$$\textcolor{red}{\textbf{[0.3 Ptos.]}} \quad P(\overline{X}_{20} > 20) = 1 - \Phi\left(\frac{20-25}{20/\sqrt{20}}\right) \approx 1 - \Phi(-1.12) = \Phi(1.12) = 0.8686 \quad \textcolor{red}{\textbf{[0.3 Ptos.]}}$$

(c) Supongamos ahora que Y_k es la máxima temperatura diaria registrada en el k-ésimo día de enero, la cual distribuye Normal(35, 2), con $k = 1, 2, \ldots, 31$.

Definamos la diferencia entre dos días consecutivos como Z_k:

$$Z_k = Y_k - Y_{k-1} \sim \text{Normal}\left(0,\, \sqrt{2^2 + 2^2 - 2\cdot 2\cdot 2\cdot 0.5}\right) \quad \textcolor{red}{\textbf{[0.5 Ptos.]}}$$

$$\sim \text{Normal}\,(0,\, 2)\,, \quad k = 2, \ldots, 31 \quad \textcolor{red}{\textbf{[0.5 Ptos.]}}$$

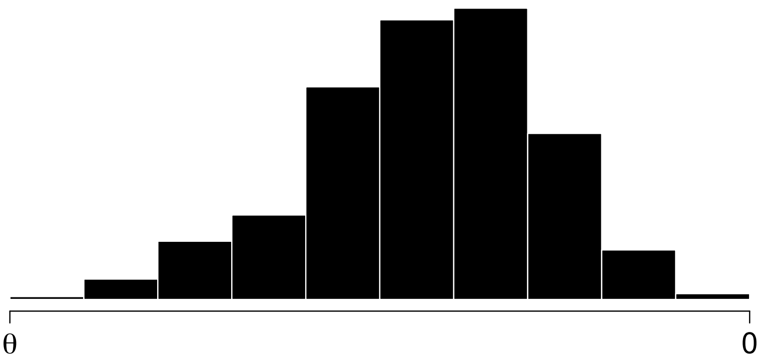
Se pide

$$\textcolor{red}{\textbf{[0.5 Ptos.]}} \quad P(|Z_k| < 1) = \Phi(1/2) - \Phi(-1/2) = 2\,\Phi(1/2) - 1 = 2\cdot 0.6915 - 1 = 0.383 \quad \textcolor{red}{\textbf{[0.5 Ptos.]}}$$

+ 1 Punto Base

Problema 2

Considere una muestra aleatoria de n temperaturas mínimas tomadas en invierno en el extremo sur del país, las cuales se presentan gráficamente



Un investigador propone un modelo Beta$(1, r)$ en el intervalo $[\theta, 0]$, con $\theta < 0$ y $r > 1$ como posible ajuste. Obtenga el estimador de momento del parámetro θ y proponga una distribución aproximada para él.

Solución

Definamos como $X_1, \ldots, X_n$ la muestra aleatoria Beta$(1, r)$ en el intervalo $[\theta, 0]$, con $\theta < 0$ y $r > 1$.

Del formulario se tiene que el primer momento teórico está dado por $\dfrac{r\,\theta}{1+r}$ e igualando al primer momento muestral $\overline{X}_n$, se tiene el estimador de momentos de θ está dado por

$$\hat{\theta} = \frac{(1+r)}{r} \cdot \overline{X}_n \quad \textbf{[2.0 Ptos.]}$$

Por teorema del límite central

$$\overline{X}_n \overset{\text{aprox}}{\sim} \text{Normal}(\mu_X,\, \sigma_X/\sqrt{n}) \quad \textbf{[1.0 Ptos.]}$$

con

$$\textbf{[1.0 Ptos.]} \quad \mu_X = \frac{r\,\theta}{1+r} \qquad \text{y} \qquad \sigma_X^2 = \frac{r\,\theta^2}{(1+r)^2\,(r+2)} \quad \textbf{[1.0 Ptos.]}$$

Por lo tanto

$$\hat{\theta} = \frac{r}{(1+r)} \cdot \overline{X}_n \overset{\text{aprox}}{\sim} \text{Normal}\left(\theta,\, |\theta|\sqrt{\frac{1}{r\,(r+2)\,n}}\right) \quad \textbf{[1.0 Ptos.]}$$

+ 1 Punto Base

Problema 3

Suponga que a su bandeja de entrada llegan correos según un proceso de Poisson con tasa esperada ν correos por hora y que con probabilidad p un correo cualquiera es SPAM. Interesa estudiar el tiempo Y transcurrido hasta recibir el próximo SPAM.

(a) **[3.0 Puntos.]** Si X representa el número de mail recibidos hasta observar un SPAM, obtenga a partir de la distribución condicional de Y dado X una aproximación de primer orden de la función generadora de momento marginal de Y. A partir de este resultado, obtenga el tiempo transcurrido esperado para recibir un SPAM.

(b) **[3.0 Puntos.]** A partir de la distribución marginal de Y, obtenga su función generadora de momento y su esperanza. Comente con respecto al resultado obtenido en (a).

Solución

(a) Tenemos que

$$\textbf{[0.3 Ptos.]} \quad X \sim \text{Geométrica}(p) \qquad \text{y} \qquad Y \,|\, X = x \sim \text{Gamma}(x, \nu) \quad \textbf{[0.3 Ptos.]}$$

Se pide a partir de la distribución de Y dado X una aproximación de primer orden de $M_Y(t)$.

$$\begin{aligned} M_Y(t) &= \text{E}\left(e^{Y\,t}\right) \quad \textbf{[0.3 Ptos.]} \\ &= \text{E}\left[\text{E}\left(e^{Y\,t} \,|\, X\right)\right] \quad \textbf{[0.3 Ptos.]} \\ &= \text{E}\left[\left(\frac{\nu}{\nu - t}\right)^X\right], \quad \text{para } t < \nu \quad \textbf{[0.4 Ptos.]} \\ &\approx \left(\frac{\nu}{\nu - t}\right)^{\mu_X} \quad \textbf{[0.3 Ptos.]} \\ &= \left(\frac{\nu}{\nu - t}\right)^{1/p} \quad \textbf{[0.3 Ptos.]} \end{aligned}$$

Derivando respecto a t y evaluando en cero se tiene que

$$\textbf{[0.4 Ptos.]} \quad M_Y^{(1)}(t) \approx \nu^{\frac{1}{p}} \cdot \left(-\frac{1}{p}\right) \cdot (\nu - t)^{-\frac{1}{p} - 1} \cdot (-1) \to M_Y^{(1)}(0) = \frac{1}{\nu\,p} = \text{E}(Y) \quad \textbf{[0.4 Ptos.]}$$

(b) Si Z es el número de mail tipo SPAM que llegan por hora, entonces por resultado visto en clases se tiene

$$Z \sim \text{Poisson}(\nu\,p) \quad \textbf{[0.8 Ptos.]}$$

Luego,

$$Y \sim \text{Exponencial}(\nu\,p) \quad \textbf{[0.8 Ptos.]}$$

Del formulario

$$\textbf{[0.4 Ptos.]} \quad M_Y(t) = \left(\frac{\nu\,p}{\nu\,p - t}\right) \to M_Y^{(1)}(t) = \nu\,p \cdot (-1) \cdot (\nu\,p - t)^{-2} \cdot (-1) \quad \textbf{[0.4 Ptos.]}$$

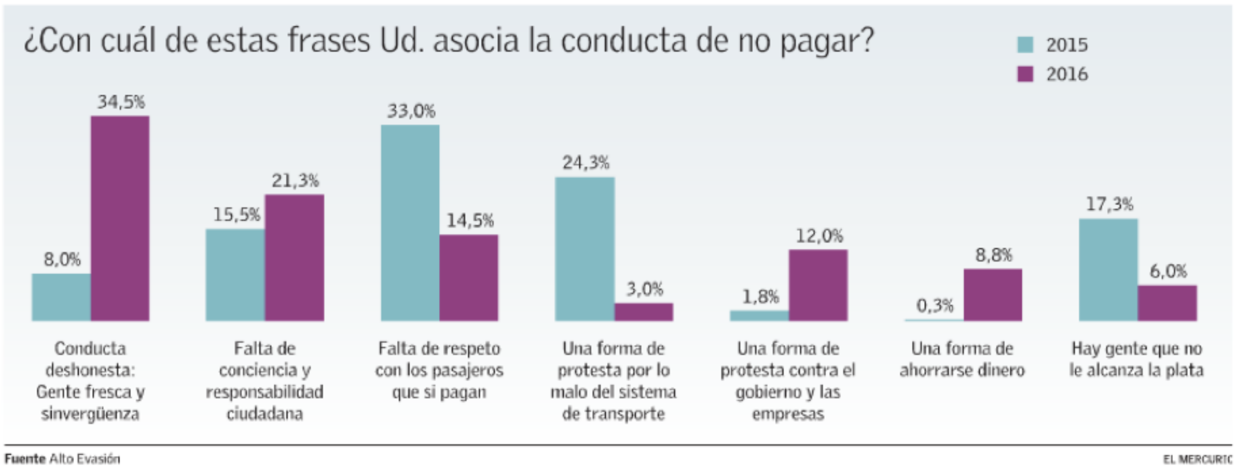
$$\to M_Y^{(1)}(0) = \frac{1}{\nu\,p} = \text{E}(Y) \quad \textbf{[0.3 Ptos.]}$$

Pese a que la versión aproximada de M_Y es distinta a la exacta, el valor esperado resulta idéntico. **[0.3 Ptos.]**

+ 1 Punto Base

Problema 4

La empresa ALTO Evasión recientemente publicó los resultados de una encuesta aplicada durante el año 2016 a 400 usuarios del Transantiago, entre múltiples resultados se encuentra lo que se presenta en la siguiente infografía.



(a) **[1.5 Puntos.]** ¿Existe evidencia que permita afirmar, con un 5 % significancia, que menos del 15 % asocia la conducta de no pagar como una forma de protestar contra el gobierno y las empresas?

(b) **[1.5 Puntos.]** Si se desea reiterar la encuesta de tal forma de estimar, con un error no mayor al 4 % y un 90 % de confianza, la proporción de usuarios que considera que la conducta de no pagar es deshonesta y la cataloga como gente fresca y sinvergüenza, basado en los resultados previos, ¿a cuántos usuarios del Transantiago debe entrevistar como mínimo?.

Como resultado de la fiscalización de 20 buses (se denominan bus-viaje) se obtiene que el no pago promedio alcanza a los \$16.700 con una desviación estándar de \$4.500. Basado en estos resultados y asumiendo que el monto de no pago sigue una distribución Normal:

(c) **[1.5 Puntos.]** ¿Es posible afirmar que la desviación estándar es superior a \$4.000 (resultado del año 2015) para un nivel de significancia del 10 %?

(d) **[1.5 Puntos.]** Si el 2015 el no pago promedio por bus-viaje alcanzó a \$14.500, ¿para qué valores de significancia es posible afirmar que este año se ha incrementado significativamente?.

Solución

(a) Sea p la proporción que asocia el no pago como forma de protesta.

Se pide contrastar las siguientes hipótesis:

$$\mathrm{H}_0 : p = 0.15 \qquad \text{vs} \qquad \mathrm{H}_a : p < 0.15 \quad \textbf{[0.3 Ptos.]}$$

del enunciado tenemos $n = 400$ y $\hat{p} = 0.12$. **[0.3 Ptos.]**

El estadístico de prueba bajo H_0 es

$$Z_0 = \frac{0.12 - 0.15}{\sqrt{\dfrac{0.15 \cdot (1 - 0.15)}{400}}} = -1.68 \quad \textbf{[0.3 Ptos.]}$$

Alternativa 1: Como $Z_0 < k_{0.05} = -1.645$, entonces existe evidencia para rechazar H_0. **[0.3 Ptos.]**

Alternativa 2: Como valor-p $= P(Z < Z_0) = \Phi(-1.68) = 0.0465 < \alpha = 0.05$, entonces existe evidencia para rechazar H_0. **[0.3 Ptos.]**

Por lo tanto, puedo afirmar que menos del 15 % asocia la conducta de no pagar a una forma de protesta. **[0.3 Ptos.]**

(b) Según el estudio anterior la proporción p de usuarios que considera que la conducta de no pagar es deshonesta y la cataloga como gente fresca y sinvergüenza es 0.345.

Tomando este resultado como base, el tamaño muestral para el nuevo estudio necesitaría más

$$n = \left(\frac{1.645 \cdot \sqrt{0.345 \cdot 0.655}}{0.04}\right)^2 = 382.18 \quad \textbf{[1.5 Ptos.]}$$

para poder cumplir con el requerimiento técnico de confianza 90 % y margen de error 0.04.

(c) Tenemos una muestra de tamaño 20, con los siguietes estadísticos muestrales

$$\overline{X} = 16700 \qquad \text{y} \qquad S = 4500 \quad \textbf{[0.2 Ptos.]}$$

Se pide contrastar las siguientes hipótesis:

$$\text{H}_0 : \sigma = 4000 \qquad \text{vs} \qquad \text{H}_a : \sigma > 4000 \quad \textbf{[0.3 Ptos.]}$$

El estadístico de prueba bajo H_0 es

$$C_0 = \frac{(20-1) \cdot 4500^2}{4000^2} = 24.047 \quad \textbf{[0.4 Ptos.]}$$

Alternativa 1: Como $C_0 \not> c_{0.90} = 27.204$, entonces no existe evidencia para rechazar H_0. **[0.3 Ptos.]**

Alternativa 2: Como valor-p $= P(C > C_0) \to \alpha = 10\,\% <$ valor-p $< 90\,\%$, entonces No existe evidencia para rechazar H_0. **[0.3 Ptos.]**

Por lo tanto, no es posible afirmar que la desviación estándar es superior a \$4.000. **[0.3 Ptos.]**

(d) Se pide valores de α para rechazar H_0 y apoyar H_a definidas como

$$\text{H}_0 : \mu = 14500 \qquad \text{vs} \qquad \text{H}_a : \mu > 14500 \quad \textbf{[0.4 Ptos.]}$$

El estadístico de prueba bajo H_0 es

$$T_0 = \frac{16500 - 14500}{4500/\sqrt{20}} = 2.186 \quad \textbf{[0.4 Ptos.]}$$

De la tabla t-Student se observa que

$$1\,\% < \text{valor-p} < 2.5\,\% \quad \textbf{[0.4 Ptos.]}$$

Por lo tanto, para todo α mayor al 2.5 % (y posiblemente para valores de α sobre un 2 %) se rechaza H_0 en favor de H_a **[0.3 Ptos.]**

+ 1 Punto Base