Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2017

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

Problema 1

Tratar de conseguir estacionamiento continua siendo un desafío para aquellos que llegan después de las 09:00 al campus, por esta razón usted decide llevar a cabo un estudio que busca determinar si es posible implementar corrales - estacionamientos reservados para estudiantes - al interior del campus. Para ello entrevista a 200 alumnos (que arriban conduciendo) durante una semana en los módulos 1 y 2.

En parte de su encuesta hay dos preguntas claves:

- Estaría dispuesto por pagar por tener estacionamiento: ____ (Si/No)
- Si su respuesta es previa es si, ¿cuánto estaría dispuesto a pagar mensualmente? ________ (CL$)

Al analizar los datos, determina qué solo 28 están dispuestos a pagar y un análisis descriptivo muestra un monto promedio de CL$20.000 con una desviación estándar de CL$4.500.

Basado en los resultados, responda

(a) ¿Es válida la afirmación que más del 10 % de los estudiantes que usan vehículos estarán dispuestos a pagar? Utilice un nivel de significancia del 5 %.

(b) Asumiendo que el pago se comporta como una distribución Normal, ¿existe evidencia que permita afirmar que la variabilidad (cuantificada por la desviación estándar) es superior a CL$3.000. Utilice un nivel de significancia del 10 %.

(c) Conforme al resultado en (b), determine el menor nivel de significancia para el cual rechaza la hipótesis que afirma que a lo más estarán dispuestos a pagar en promedio es CL$18.000.

(d) Ahora, asumiendo que el monto dispuesto a pagar se comporta como una distribución Gamma(k, ν), ¿existe evidencia que permita afirmar que el monto medio supera los CL$18.000 para un nivel de significancia del 5 %? Considere k conocido e igual a 20.5 y una prueba estadística aproximada.

Solución

(a) Tenemos $X_1, \ldots, X_n$ respuestas Bernoulli(p), con $n = 200$.

Se pide contrastar las siguientes hipótesis

$$\mathrm{H}_0 : p = p_0 \quad \text{vs} \quad \mathrm{H}_a : p > p_0 \quad \textbf{[0.3 Ptos.]}$$

con $p_0 = 0.10$.

Bajo H_0 el estadístico de prueba

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0\,(1 - p_0)}{n}}} \overset{\text{aprox}}{\sim} \text{Normal}(0,\, 1) \quad \textbf{[0.3 Ptos.]}$$

Evaluando, con $\hat{p} = \frac{28}{200}$ **[0.3 Ptos.]**, se tiene que $Z_0 = 1.886$. **[0.3 Ptos.]**

Alternativa 1: Como $Z_0 = 1.886 > k_{0.95} = 1.645$ se rechaza H_0, es decir, es válida la afirmación que más del 10 % de los estudiantes que usan vehículos estarán dispuestos a pagar. **[0.3 Ptos.]**

Alternativa 2: Como el valor-p es igual a $1 - \Phi(1.886) \approx 0.02975 < 0.05 = \alpha$ se rechaza H_0, Es decir, es válida la afirmación que más del 10 % de los estudiantes que usan vehículos estarán dispuestos a pagar. **[0.3 Ptos.]**

(b) Se piden un test sobre σ basado en una muestra aleatoria Normal.

$$\text{H}_0 : \sigma = \sigma_0 \quad \text{vs} \quad \text{H}_a : \sigma > \sigma_0 \quad \textbf{[0.3 Ptos.]}$$

con $\sigma_0 = 3.000$.

Bajo H_0 se tiene que

$$C_0 = \frac{(n-1)\, S^2}{\sigma_0^2} \sim \chi^2(n-1) \quad \textbf{[0.3 Ptos.]}$$

Reemplazando con $n = 28$ y $S = 4.500$ se tiene que $C_0 = 60.75$ **[0.3 Ptos.]**

Alternativa 1: Como $C_0 = 60.75 > c_{0.90}(27) = 36.741$ se rechaza H_0, es decir, la variabilidad es superior a 3.000. **[0.6 Ptos.]**

Alternativa 2: Como el valor-p $= 1 - P(C \leq 60.75) < 1 - 0.995 = 0.005 < 0.10 = \alpha$ se rechaza H_0, es decir, la variabilidad es superior a 3.000. **[0.3 Ptos.]**

(c) En este caso el test de para μ, distribución normal con σ desconocido, ya que en (b) se rechazo el valor 3.000

$$\text{H}_0 : \mu = \mu_0 \quad \text{vs} \quad \text{H}_a : \mu > \mu_0 \quad \textbf{[0.3 Ptos.]}$$

con μ_0 =18.000.

Bajo H_0 el estadístico de prueba

$$T_0 = \frac{\overline{X}_n - \mu_0}{S/\sqrt{n}} \sim \text{t-Student}(n-1) \quad \textbf{[0.3 Ptos.]}$$

Evaluando, con n = 28, $\overline{X} = 20.000$ y $S = 4.500$, se tiene que $T_0 = 2.351779$. **[0.3 Ptos.]**

Luego, el menor nivel de significancia para el cual se rechaza H_0 en favor de H_a es el valor-p **[0.3 Ptos.]**, que para este caso se encuentra entre

$$1\,\% < \text{valor-p} < 2.5\,\% \quad \textbf{[0.3 Ptos.]}$$

(d) Si la muestra proviene de una distribución Gamma(k, ν) con k conocido y $\mu = \frac{k}{\nu}$, entonces se pide el siguiente test sobre μ

$$\text{H}_0 : \mu = 18.000 \quad \text{vs} \quad \text{H}_a : \mu > 18.000$$

equivalente sobre ν a

$$\text{H}_0 : \nu = \frac{k}{18.000} \quad \text{vs} \quad \text{H}_a : \nu < \frac{k}{18.000} \quad \textbf{[0.3 Ptos.]}$$

Como

$$\hat{\nu} \overset{\text{aprox}}{\sim} \text{Normal}\left(\nu, \sqrt{\frac{1}{I_n(\nu)}}\right)$$

entonces, bajo H_0

$$Z_0 = \sqrt{I_n(\nu_0)}\,(\hat{\nu} - \nu_0) = \sqrt{\frac{n\,k}{\nu_0^2}}\,(\hat{\nu} - \nu_0) \overset{\text{aprox}}{\sim} \text{Normal}(0,\,1) \quad \textbf{[0.3 Ptos.]}$$

reemplazado con $k = 20.5$, μ_0 =18.000 $\to \nu_0 = \frac{k}{\mu_0} = 0.001138889$ y $\hat{\nu} = \frac{k}{\overline{X}} = 0.001025$, se tiene que $Z_0 = -2.39583$ **[0.6 Ptos.]**

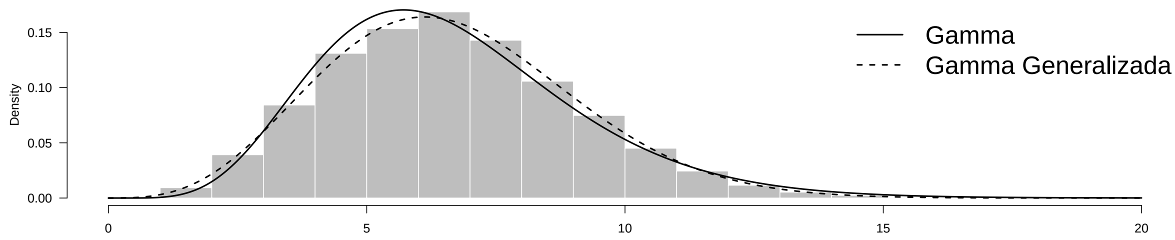
Alternativa 1: Como $Z_0 = -2.39583 < k_{0.05} = -1.645$ se rechaza H_0, es decir, existe evidencia que permite afirmar que el monto medio supera los CL\$18.000 para un nivel de significancia del 5 %. **[0.3 Ptos.]**

Alternativa 2: Como el valor-p $\approx \Phi(-2.396) = 1 - 0.9917 = 0.0083 < 0.05 = \alpha$ se rechaza H_0, es decir, existe evidencia que permite afirmar que el monto medio supera los CL\$18.000 para un nivel de significancia del 5 %. **[0.3 Ptos.]**

+ 1 Punto Base

Problema 2

Una distribución de probabilidad que permite mayor grado de libertad en su forma para modelar datos con soporte en $\mathbb{R}_0^+$ que la clásica Gamma, es la Gamma-Generalizada(k, ν, β), tal como se muestra la figura en la que se ajustan ambas distribuciones a un conjunto de mediciones.



Esta este modelo tiene una función de densidad dada por

$$f(x) = \frac{\beta\,\nu^{\beta\,k}}{\Gamma(k)}\,x^{\beta\,k-1}\,e^{-(\nu\,x)^\beta}, \quad x \geq 0$$

con $k > 0$, $\nu > 0$ y $\beta > 0$.

Obtenga el estimador de momento y de máxima verosimilitud de ν, calcule sus valores esperados aproximados de 1er orden y comente. Por simplicidad suponga que k y β son parámetros conocidos.

Solución

Consideremos una muestra aleatoria $X_1, \ldots, X_n$ iid proveniente de una población Gamma-Generalizada(k, ν, β).

Se pide estimar ν bajo el supuesto que k y β son parámetros conocidos por:

Método de Momentos

Tenemos que el 1er momento de una Gamma-Generalizada está dado por

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^{+\infty} x \cdot \frac{\beta\,\nu^{\beta\,k}}{\Gamma(k)}\,x^{\beta\,k-1}\,e^{-(\nu\,x)^\beta}\,dx \quad \textbf{[0.3 Ptos.]}\\
&= \int_0^{+\infty} \frac{1}{\nu\,\Gamma(k)}\,y^{\left(k+\frac{1}{\beta}\right)-1}\,e^{-y}\,dy, \qquad \text{con } y = (\nu\,x)^\beta \quad \textbf{[0.3 Ptos.]}\\
&= \frac{\Gamma\left(k+\frac{1}{\beta}\right)}{\nu\,\Gamma(k)} \int_0^{+\infty} \frac{1}{\Gamma\left(k+\frac{1}{\beta}\right)}\,y^{\left(k+\frac{1}{\beta}\right)-1}\,e^{-y}\,dy \quad \textbf{[0.3 Ptos.]}\\
&= \frac{\Gamma\left(k+\frac{1}{\beta}\right)}{\nu\,\Gamma(k)} \cdot 1, \quad \text{por área bajo la curva sobre } \mathbb{R}_0^+ \text{ de una Gamma} \quad \textbf{[0.3 Ptos.]}\\
&= \frac{\Gamma\left(k+\frac{1}{\beta}\right)}{\nu\,\Gamma(k)} \quad \textbf{[0.3 Ptos.]}
\end{aligned}$$

Nota: El alumno(a) también podría haber formado una Weibull para llegar al mismo resultado, en este caso distribuir los 1.5 Ptos.

Luego, igualando al 1er momento empírico $\overline{X}_n$, se tiene que el estimador de momento $\widetilde{\nu}$ de ν es

$$\textbf{[0.5 Ptos.]} \quad \widetilde{\nu} = \frac{\Gamma\left(k+\frac{1}{\beta}\right)}{\overline{X}_n \cdot \Gamma(k)} \to \mathrm{E}(\widetilde{\nu}) \approx \frac{\Gamma\left(k+\frac{1}{\beta}\right)}{\mu_{\overline{X}} \cdot \Gamma(k)} = \frac{\Gamma\left(k+\frac{1}{\beta}\right)}{\mathrm{E}(X) \cdot \Gamma(k)} = \nu \quad \textbf{[1.0 Ptos.]}$$

Método de Máxima Verosimilitud

Tenemos que la verosimilitud está dada por

$$L(\nu) = \prod_{i=1}^{n} \frac{\beta\,\nu^{\beta\,k}}{\Gamma(k)}\,X_i^{\beta\,k-1}\,e^{-(\nu\,X_i)^\beta} = \frac{\beta^n\,\nu^{n\,\beta\,k}}{[\Gamma(k)]^n}\left(\prod_{i=1}^{n} X_i\right)^{\beta\,k-1} \exp\left[-\sum_{i=1}^{n}(\nu\,X_i)^\beta\right] \quad \textbf{[0.4 Ptos.]}$$

Aplicando logaritmo natural y derivando parcialmente con respecto a ν se tiene que

$$\frac{\partial}{\partial\,\nu}\ln\,L(\nu) = \frac{n\,k\,\beta}{\nu} - \sum_{i=1}^{n}\beta\,(\nu\,X_i)^{\beta-1}\,X_i \quad \textbf{[0.4 Ptos.]}$$

al igualar a cero y despejar se tiene que

$$\textbf{[0.5 Ptos.]} \quad \hat{\nu} = \left(\frac{n\,k}{\sum_{i=1}^{n} X_i^\beta}\right)^{1/\beta} = \left(\frac{k}{\overline{X^\beta}}\right)^{1/\beta} \to \mathrm{E}(\hat{\nu}) \approx \left(\frac{k}{\mu_{\overline{X^\beta}}}\right)^{1/\beta} = \left(\frac{k}{\mathrm{E}(X^\beta)}\right)^{1/\beta} \quad \textbf{[0.5 Ptos.]}$$

Notemos que si $Y = X^\beta$, entonces

$$\textbf{[0.3 Ptos.]} \quad f_Y(y) = f_X(y^{1/\beta})\cdot\left|\frac{1}{\beta}\,y^{1/\beta-1}\right| = \frac{\nu^{\beta\,k}}{\Gamma(k)}\,y^{k-1}\,e^{-\nu^\beta\,y} \to X^\beta = Y \sim \text{Gamma}\left(k,\,\nu^\beta\right) \quad \textbf{[0.3 Ptos.]}$$

Por lo tanto

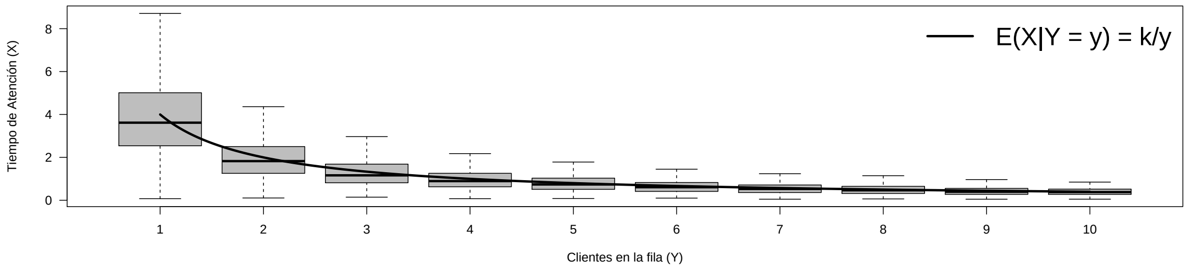
$$\mathrm{E}(\hat{\nu}) \approx \left(\frac{k}{k/\nu^\beta}\right)^{1/\beta} = \nu \quad \textbf{[0.3 Ptos.]}$$

Es decir, ambos estimadores son aproximadamente insesgados. **[0.3 Ptos.]**

+ 1 Punto Base

Problema 3

Buscando información para mejorar la percepción que tienen sus clientes, un supermercado encargo un análisis de los tiempos X de atención de sus cajeros cuando pasaba la mercadería por los sensores y cobraba la cuenta. La figura a continuación muestra los diagramas de cajas de estos tiempos, cuando en la fila había solo una persona (el atendido) o más (hay cola).



La linea es una función que representa la evolución del valor esperado del tiempo de atención condicionado a la cantidad de clientes en la fila. El especialista indica que un modelo Gamma$(k,\,\nu)$ con coeficiente de variación del 50 % sería adecuado para los tiempos de atención, pero de todas maneras está condicionado al número de clientes Y en la fila. Si el número de clientes en una fila es una variable aleatoria Geométrica(p), determine entonces la distribución del tiempo de servicio Z definido como $Z = X \cdot Y$.

Solución

Tenemos del enunciado que

$$\textbf{[0.5 Ptos.]} \quad X \,|\, Y = y \sim \text{Gamma}(k,\, y) \quad \text{y} \quad Y \sim \text{Geométrica}(p) \quad \textbf{[0.5 Ptos.]}$$

con $k = 4$ **[0.3 Ptos.]**, ya que el c.o.v. de una Gamma en este caso no depende de y, y es igual a $1/\sqrt{k}$.

Si $Z = X \cdot Y \to X = \frac{Z}{Y}$ **[0.5 Ptos.]**, y del formulario

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(g^{-1}, y) \left| \frac{\partial}{\partial z} g^{-1} \right| dy \quad \textbf{[0.3 Ptos.]}$$

Como X es continua e Y discreta, la densidad conjunta es $f_{X,Y}(x, y) = f_{X\,|\,Y=y}(x) \cdot p_Y(y)$ **[1.0 Ptos.]**.

Luego, la integral con respecto a y cambia por una suma sobre todo el soporte Θ_Y:

$$\begin{aligned}
f_Z(z) &= \sum_{y \in \Theta_Y} f_{X\,|\,Y=y}\left(g^{-1}\right) \cdot p_Y(y) \cdot \left| \frac{\partial}{\partial z} g^{-1} \right| \quad \textbf{[1.0 Ptos.]} \\
&= \sum_{y \in \Theta_Y} f_{X\,|\,Y=y}(z/y) \cdot p_Y(y) \cdot \frac{1}{y} \quad \textbf{[0.3 Ptos.]} \\
&= \sum_{y \in \Theta_Y} \frac{y^k}{\Gamma(k)} \left(\frac{z}{y}\right)^{k-1} e^{-z} \cdot p_Y(y) \cdot \frac{1}{y} \quad \textbf{[0.3 Ptos.]} \\
&= \frac{1}{\Gamma(k)}\, z^{k-1}\, e^{-z} \sum_{y \in \Theta_Y} p_Y(y) \quad \textbf{[0.3 Ptos.]} \\
&= \frac{1}{\Gamma(k)}\, z^{k-1}\, e^{-z} \cdot 1 \quad \textbf{[0.3 Ptos.]} \\
&= \frac{1}{\Gamma(k)}\, z^{k-1}\, e^{-z} \quad \textbf{[0.3 Ptos.]}
\end{aligned}$$

Por lo tanto,

$$Z \sim \text{Gamma}(4,\, 1) \quad \textbf{[0.4 Ptos.]}$$

+ 1 Punto Base

Problema 4

Durante los meses fríos de invierno, las temperaturas mínimas diarias que entrega una estación de monitoreo se comportan como variables aleatorias cuya distribución de probabilidad está dada por la siguiente función de densidad

$$f(x) = \nu\, e^{\nu\,(x+\alpha)}, \quad x < -\alpha \quad \text{y} \quad \alpha \in \mathbb{R}$$

Para una muestras aleatorias de tamaño n de temperaturas mínimas históricas medidas en meses fríos de invierno, proponga una distribución aproximada del promedio de estas temperaturas.

Solución

Consideremos una muestra aleatoria $X_1, \ldots, X_n$ iid proveniente de una distribución con función de densidad $f(x)$.

Aplicando teorema del límite central **[1.0 Ptos.]** se tiene que

$$\overline{X}_n \overset{\text{aprox}}{\sim} \text{Normal}\left(\mu,\, \frac{\sigma}{\sqrt{n}}\right) \quad \textbf{[0.5 Ptos.]}$$

donde $\mu = \text{E}(X_i)$ y $\sigma^2 = \text{Var}(X_i)$.

Alternativa 1

Los parámetros μ y σ se obtiene por definición:

$$\begin{aligned}
\mu &= \text{E}(X_i) \quad \textbf{[0.2 Ptos.]}\\
&= \int_{-\infty}^{-\alpha} x \cdot \nu\, e^{\nu\,(x+\alpha)}\, dx \quad \textbf{[0.3 Ptos.]}\\
&= \int_{+\infty}^{0} \left(\frac{u}{\nu} + \alpha\right) \cdot e^{-u}\, du, \qquad u = -\nu\,(x+\alpha) \quad \text{y} \quad du = -\nu\, dx \quad \textbf{[0.7 Ptos.]}\\
&= -\frac{1}{\nu} \int_{0}^{+\infty} u\, e^{-u} - \alpha \int_{0}^{+\infty} e^{-u}\, du \quad \textbf{[0.6 Ptos.]}\\
&= -\left(\frac{1}{\nu} + \alpha\right) \quad \textbf{[0.2 Ptos.]}
\end{aligned}$$

y

$$\begin{aligned}
\sigma^2 &= \text{E}(X_i^2) - \mu^2 \quad \textbf{[0.2 Ptos.]}\\
&= \int_{-\infty}^{-\alpha} x^2 \cdot \nu\, e^{\nu\,(x+\alpha)}\, dx - \left(\frac{1}{\nu} + \alpha\right)^2 \quad \textbf{[0.4 Ptos.]}\\
&= -\int_{+\infty}^{0} \left(\frac{u}{\nu} + \alpha\right)^2 \cdot e^{-u}\, du - \left(\frac{1}{\nu} + \alpha\right)^2 \quad \textbf{[0.4 Ptos.]}\\
&= \frac{1}{\nu^2} \int_{0}^{+\infty} u^2\, e^{-u} + \frac{2\,\alpha}{\nu} \int_{0}^{+\infty} u\, e^{-u}\, du + \alpha^2 \int_{0}^{+\infty} e^{-u}\, du - \left(\frac{1}{\nu} + \alpha\right)^2 \quad \textbf{[0.4 Ptos.]}\\
&= \frac{2}{\nu^2} + \frac{2\,\alpha}{\nu} + \alpha^2 - \left(\frac{1}{\nu} + \alpha\right)^2 \quad \textbf{[0.2 Ptos.]}\\
&= \frac{1}{\nu^2} \quad \textbf{[0.2 Ptos.]}\\
\to \sigma &= \frac{1}{\nu} \quad \textbf{[0.2 Ptos.]}
\end{aligned}$$

Alternativa 2

Tenemos que si $Y = -(X + \alpha)$, con X una variable aleatoria con función de densidad $f(x)$ propuesta en el enunciado, entonces Y distribuye Exponencial(ν) ya que

$$f_Y(x) = f_X(-Y - \alpha)\,|1| = \nu\, e^{-\nu\, y}, \quad y > 0 \quad \textbf{[2.0 Ptos.]}$$

Luego

$$\textbf{[0.5 Ptos.]} \quad \mathrm{E}(Y) = \frac{1}{\nu} = \mathrm{E}(-X - \alpha) = -\mathrm{E}(X) - \alpha \to \mathrm{E}(X) = -\left(\frac{1}{\nu} + \alpha\right) = \mu \quad \textbf{[0.5 Ptos.]}$$

y

$$\mathrm{Var}(Y) = \frac{1}{\nu^2} = \mathrm{Var}(-X - \alpha) = \mathrm{Var}(X) = \sigma^2 \to \sigma = \frac{1}{\nu} \quad \textbf{[1.0 Ptos.]}$$

Por lo tanto,

$$\overline{X}_n \overset{\text{aprox}}{\sim} \text{Normal}\left(-\frac{1}{\nu} - \alpha,\, \frac{1}{\nu\,\sqrt{n}}\right) \quad \textbf{[0.5 Ptos.]}$$

+ 1 Punto Base