Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2019

Curso	**:**	**Probabilidad y Estadística**
Sigla	**:**	**EYP1113**
Profesores	**:**	**Ricardo Aravena C. (Sec 01 - Sec 03) y Ricardo Olea O (Sec 02 - Sec 04).**

Pauta Interrogación 3

Problema 1

Un equipo de psicólogos lleva a cabo una investigación respecto a la atención y comprensión de un tópico específico dentro de una clase en un curso universitario.

El equipo busca verificar un conjunto de hipótesis:

- Hip1: En cursos masivos, menos de dos tercios de los alumnos logran comprender un tópico especifico.
- Hip2: Frente a una consulta específica del tópico, los alumnos que lograron una buena comprensión responderán acertadamente con (asuma que el tiempo sigue una distribución normal):
  - un tiempo medio menor a 20 seg,
  - una desviación estándar mayor a 10 seg

Para ello considera un curso masivo con 124 alumnos, quienes son consultados al final respecto a la comprensión del tópico especifico. Del total, solo 72 consideran haber comprendido el tópico especifico. Posterior a la clase, los alumnos que indican haber comprendido el tópico son sometidos a una evaluación con la aplicación Kahoot!. Esta aplicación permite medir el acierto y tiempo (en segundos) utilizado para una o más consultas.

El resultado para una consulta especifica fue:

```
            Respuesta
     -----------------------
     | Correcta | Incorrecta
----------------------------
N    |       28 |         44
mean |       16 |         24
sd   |       12 |         18
----------------------------
```

Indique y desarrolle las respectivas pruebas para verificar o refutar las hipótesis planteadas. Use valor-p para su decisión y un nivel de significancia del 5 %.

Solución

Logro 1: *Hipótesis 1 + Estadístico de prueba*

$$\text{H}_0 : p = 2/3 \qquad \text{vs} \qquad \text{H}_a : p < 2/3 \qquad \textbf{[0.5 Ptos]}$$

Estadístico de prueba

$$Z_0 = \frac{72/124 - 2/3}{\sqrt{\frac{(1/3)\cdot(2/3)}{124}}} = -2{,}03 \overset{\text{aprox}}{\sim} \text{Normal}(0,1) \qquad \textbf{[0.5 Ptos]}$$

Logro 2: *Valor-p + Conclusión sobre hipótesis 1*

$$\text{valor-p} = \Phi(-2{,}03) = 0{,}0212 < 0{,}05 = \alpha \qquad \textbf{[0.5 Ptos]}$$

Se rechaza H_0 en favor de H_a, es decir, es válida la afirmación que menos de 2/3 de los alumnos logran la comprensión del tópico. **[0.5 Ptos]**

Logro 3: *Hipótesis 2 + Estadístico de prueba*

$$H_0 : \mu = 20 \qquad \text{vs} \qquad H_a : \mu < 20 \qquad \textbf{[0.5 Ptos]}$$

Estadístico de prueba

$$T_0 = \frac{16 - 20}{12/\sqrt{28}} = -1{,}76 \overset{\text{aprox}}{\sim} \text{t-Student}(27) \qquad \textbf{[0.5 Ptos]}$$

Logro 4: *Valor-p + Conclusión sobre hipótesis 2*

$$\text{valor-p} = P(T < -1{,}76) \to 2{,}5\,\% < \text{valor-p} < 5\,\% = \alpha \qquad \textbf{[0.5 Ptos]}$$

Se rechaza H_0 en favor de H_a, es decir, es válida la afirmación que el tiempo medio es menor a 20 seg.
[0.5 Ptos]

Logro 5: *Hipótesis 3 + Estadístico de prueba*

$$H_0 : \sigma = 10 \quad \text{vs} \quad H_a : \sigma > 10 \qquad \textbf{[0.5 Ptos]}$$

Estadístico de prueba

$$C_0 = \frac{(28 - 1)\,12^2}{10^2} = 38{,}88 \sim \chi^2(27) \qquad \textbf{[0.5 Ptos]}$$

Logro 6: *Valor-p + Conclusión sobre hipótesis 3*

$$\text{valor-p} = P(C > 38{,}99) \to \alpha = 5\,\% < \text{valor-p} < 10 \qquad \textbf{[0.5 Ptos]}$$

No se rechaza H_0 en favor de H_a, es decir, no es válida la afirmación que hay una desviación estándar mayor a 10 seg. **[0.5 Ptos]**

+ 1 Punto Base

Problema 2

La distribución Gamma-Inversa(α, β) es una alternativa a las clásicas distribuciones con asimetría positiva. Su función de densidad está dada por la siguiente expresión:

$$f_X(x) = \frac{x^{-(\alpha+1)}\, e^{-1/(\beta\, x)}}{\Gamma(\alpha)\, \beta^{\alpha}},$$

con $x \geq 0$, $\alpha > 2$ y $\beta > 0$. En base a una muestra aleatoria de tamaño n, independiente con idéntica distribución, se proponen dos estimadores para β con α conocido:

$$\widehat{\beta} = \frac{n}{(\alpha-1)\displaystyle\sum_{i=1}^{n} X_i} \qquad \text{y} \qquad \widetilde{\beta} = \frac{1}{n\,\alpha}\sum_{i=1}^{n}\frac{1}{X_i}.$$

Obtenga el ECM aproximado de 1er orden de $\widehat{\beta}$ y exacto para $\widetilde{\beta}$. ¿Cuál es más eficiente? Justifique.

Hint: Si $X \sim$ Gamma-Inversa(α, β), entonces $\mu_X = \dfrac{1}{\beta\,(\alpha-1)}$ y $\sigma_X^2 = \dfrac{1}{\beta^2\,(\alpha-1)^2\,(\alpha-2)}$.

Solución

Logro 1: *Valor esperado de* $\widehat{\beta}$

Tenemos que, por identica distribución

$$\textbf{[0.5 Ptos]} \qquad \mathrm{E}(\widehat{\beta}) \approx \frac{n}{(\alpha-1)\displaystyle\sum_{i=1}^{n}\mu_{X_i}} = \frac{n}{(\alpha-1)\,n\,\mu_X} = \beta. \qquad \textbf{[0.5 Ptos]}$$

Logro 2: *Varianza y ECM de* $\widehat{\beta}$

Por independencia e idéntica distribución + aproximación de 1er orden + insesgamiento

$$\textbf{[0.5 Ptos]} \qquad \mathrm{Var}(\widehat{\beta}) \approx n\,\sigma_X^2\left[-\frac{n\,\alpha}{(\alpha-1)^2\,n^2\,\mu_X^2}\right]^2 = \frac{\alpha^2\,\beta^2}{n\,(\alpha-1)^2\,(\alpha-2)} = \mathrm{ECM}(\widehat{\beta}). \qquad \textbf{[0.5 Ptos]}$$

Logro 3: *Distribución exacta* $\widetilde{\beta} \sim$ *Gamma*$(n\,\alpha, n\,\alpha/\beta)$

Por teorema de cambio de variables y suma de variables independientes

$$\begin{aligned} Y_i = 1/X_i \to Y_i &\sim \text{Gamma}(\alpha, 1/\beta) \qquad \textbf{[0.5 Ptos]} \\ \to \sum_{i=1}^{n} Y_i &\sim \text{Gamma}(n\,\alpha, 1/\beta) \\ \to \widetilde{\beta} &\sim \text{Gamma}(n\,\alpha, n\,\alpha/\beta) \qquad \textbf{[0.5 Ptos]} \end{aligned}$$

Logro 4: *Valor esperado de* $\widetilde{\beta}$

$$\mathrm{E}(\widetilde{\beta}) = \beta. \qquad \textbf{[1.0 Ptos]}$$

Logro 5: *Varianza y ECM de* $\widetilde{\beta}$

Por insesgamiento

$$\mathrm{Var}(\widetilde{\beta}) = \frac{\beta^2}{n\,\alpha} = \mathrm{ECM}(\widetilde{\beta}). \qquad \textbf{[1.0 Ptos]}$$

Logro 6: *Mayor eficiencia* $\widetilde{\beta}$

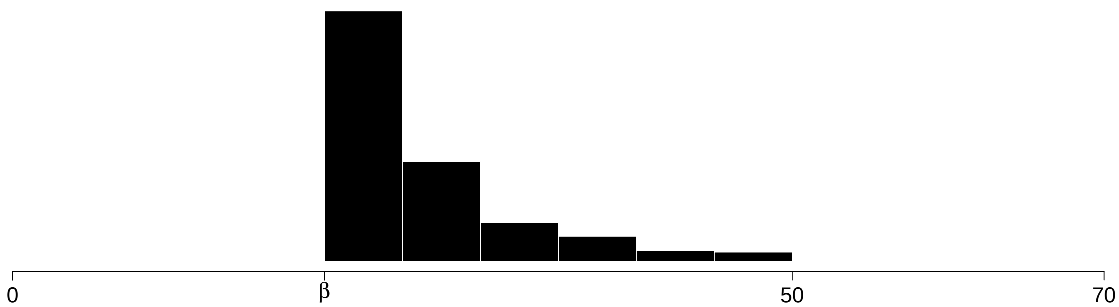
$$\frac{\text{ECM}(\widehat{\beta})}{\text{ECM}(\widetilde{\beta})} = \frac{\alpha^2}{(\alpha-1)^2\,(\alpha-2)} > 1, \qquad \textbf{[0.5 Ptos]}$$

es decir, $\widetilde{\beta}$ es más eficiente como estimador de β. **[0.5 Ptos]**

+ 1 Punto Base

Problema 3

Durante el reciente fenómeno acaecido en Concepción-Chile (tornados y trombas marinas) un especialista realizó, con un anemómetro, mediciones de velocidad de viento (m/s). La siguiente Figura y salidas de **R** muestran el comportamiento empírico de las mediciones obtenidas:




```
 N   min(x)   max(x)   sum(x) sum(log(x))
40 25.02385 50.04845 1259.933    137.4285
```

Un posible modelo que podría ajustar bien este comportamiento empírico esta dado por la siguiente función de densidad: $f(x) = \dfrac{\theta\,\beta^\theta}{x^{\theta+1}}$, para $x > \beta$ y $\theta > 1$.

(a) Obtenga el estimador de momentos y de máxima verosimilitud de θ. Por simplicidad, considere β conocido e igual 25 m/s.

(b) Con base a la información, es posible afirmar que la velocidad media supera los 30 m/s, es decir, puede ser catalogado como fuerte temporal-huracán. Sea explícito y utilice un nivel de significancia el 5 %.

Solución

(a) **Logro 1:** *Estimador de Momento*

$$E(X) = \int_\beta^\infty x\,\frac{\theta\,\beta^\theta}{x^{\theta+1}}\,dx = \frac{\theta\,\beta}{\theta-1}\int_\beta^\infty \frac{(\theta-1)\,\beta^{\theta-1}}{x^{(\theta-1)+1}}\,dx = \frac{\theta\,\beta}{\theta-1}\cdot 1 = \frac{\theta\,\beta}{\theta-1}. \qquad \textbf{[0.5 Ptos]}$$

Luego, igualando al primer momento empírico se tiene que el estimador de momento de θ es

$$\widehat{\theta} = \frac{\overline{X}}{\overline{X}-25} = 4{,}847145 \qquad \textbf{[0.5 Ptos]}$$

Logro 2: *Verosimilitud y Log-Verosimilitud*

Por independencia e idéntica distribución

$$L(\theta) = \theta^n\,\beta^{n\,\theta}\left(\prod_{i=1}^n X_i\right)^{-(\theta+1)} \qquad \textbf{[0.5 Ptos]}$$

$$\ln\,L(\theta) = n\,\ln(\theta) + n\,\theta\,\ln(\beta) - (\theta+1)\sum_{i=1}^n \ln(X_i) \qquad \textbf{[0.5 Ptos]}$$

Logro 3: *Estimador Máximo Verosímil*

Derivando parcialmente $\ln\,L(\theta)$ una vez con respecto a θ e igualando a cero:

$$\frac{\partial}{\partial\,\theta}\,\ln\,L(\theta) = \frac{n}{\theta} + n\,\ln(\beta) - \sum_{i=1}^n \ln(X_i) = 0 \qquad \textbf{[0.5 Ptos]}$$

se tiene que

$$\widetilde{\theta} = \frac{n}{\displaystyle\sum_{i=1}^{n} \ln(X_i) - n\, \ln(25)} = 4{,}611766 \qquad \textbf{[0.5 Ptos]}$$

(b) **Logro 4:** *Información de Fisher y distribución asintótica de* $\widetilde{\theta}$

Tenemos que

$$\frac{\partial^2}{\partial\, \theta^2} \ln\, L(\theta) = -\frac{n}{\theta^2} \to I(\theta) = \frac{n}{\theta^2}. \qquad \textbf{[0.5 Ptos]}$$

Luego

$$\widetilde{\theta} \overset{\text{aprox}}{\sim} \text{Normal}\left(\theta,\, \frac{\theta}{\sqrt{n}}\right). \qquad \textbf{[0.5 Ptos]}$$

Logro 5: *Hipótesis*

$$\text{H}_0 : g(\theta) = 30 \qquad \text{vs} \qquad \text{H}_a : g(\theta) > 30$$

con $g(\theta) = \dfrac{25\,\theta}{\theta - 1}$, lo que es equivalente a testear

$$\text{H}_0 : \theta = 6 \qquad \text{vs} \qquad \text{H}_a : \theta < 6 \qquad \textbf{[1.0 Ptos]}$$

Logro 6: *Estadístico de prueba + conclusión (valor-p o valor crítico)*

Bajo H_0 se tiene que

$$Z_0 = \frac{\widetilde{\theta} - 6}{6/\sqrt{n}} = -1{,}463327 \overset{\text{aprox}}{\sim} \text{Normal}(0,1). \qquad \textbf{[0.5 Ptos]}$$

Como

$$\text{valor-p} \approx \Phi(-1{,}46) = 0{,}0721 > 0{,}05 = \alpha \quad \text{ó} \quad Z_0 = -1{,}46 > -1{,}645 = k_{0{,}05} = -1{,}645,$$

no es posible afirmar que la velocidad media supera los 30 m/s. **[0.5 Ptos]**

Nota: Alternativamente se puede calcular Z_0 a partir de la distribución asintótica de $g(\widetilde{\theta})$ y concluir. $Z_0 = 2{,}026$ y valor-p $\approx 0{,}0214 < 0{,}05 = \alpha$, es decir, es posible afirmar que la velocidad media supera los 30 m/s.

+ 1 Punto Base

Problema 4

Airbnb es una empresa que ofrece una plataforma de software dedicada a la oferta de alojamientos mediante la cual los anfitriones pueden publicitar y contratar el arriendo de sus propiedades. Los arrendatarios y huéspedes pueden valorarse mutuamente, como referencia para futuros usuarios. Airbnb tiene una oferta de 2.000.000 de propiedades en 192 países y 33.000 ciudades. Desde su creación en noviembre de 2008 hasta junio de 2012 se realizaron 10 millones de reservas. En esta oportunidad, usted dispone de una muestra de 25 ofertas de Airbnb y realiza varias pruebas de hipótesis con respecto al precio X, en €, por noche:

```
                                                  statistic    p.value
sigma.test(X, sigma = 130, alternative = "greater")   21.134660 0.63076898
z.test(X, mu = 130, sd(X), alternative = "greater")   -1.296794 0.90264902
t.test(X, mu = 130, alternative = "less")             -1.296794 0.10351587
z.test(X, mu = 130, sd(X), alternative = "two.sided") -1.296794 0.19470195
sigma.test(X, sigma = 130, alternative = "two.sided") 21.134660 0.73846204
z.test(X, mu = 130, sd(X), alternative = "less")      -1.296794 0.09735098
sigma.test(X, sigma = 130, alternative = "less")      21.134660 0.36923102
t.test(X, mu = 130, alternative = "two.sided")        -1.296794 0.20703173
t.test(X, mu = 130, alternative = "greater")          -1.296794 0.89648413
```

Si los precios se comportan según una distribución Normal:

(a) ¿Existe evidencia al 10 % de significancia para afirmar que la desviación estándar de los precios es superior a 130 €? Indique hipótesis y test utilizado.

(b) ¿Cuál es el mínimo nivel de significancia que apoya la hipótesis que la media de los precios es menor a 130 €? Indique hipótesis y test utilizado.

Solución

(a) **Logro 1:** *Hipótesis* **[1.0 Ptos]**

Se pide testear

$$\mathrm{H}_0 : \sigma = \sigma_0 \quad \text{vs} \quad \mathrm{H}_a : \sigma > \sigma_0$$

con $\sigma_0 = 130$.

Logro 2: *Identificar Test* **[1.0 Ptos]**

Se debe realizar un test chi-cuadrado con $C_0 = 21{,}135 \sim \chi^2(24)$.

Logro 3: *Responder utilizando salida de **R*** **[1.0 Ptos]**

Desde la salida de **R** tenemos que el valor-p es igual a 63,08 %, el cual es mayor a $\alpha = 10\,\%$, por lo tanto no existe suficiente evidencia para rechazar H_0 y apoyar la hipótesis que la desviación estándar de los precios es superior a 130 €.

(b) **Logro 4:** *Hipótesis* **[1.0 Ptos]**

Se pide testear

$$\mathrm{H}_0 : \mu = \mu_0 \quad \text{vs} \quad \mathrm{H}_a : \mu < \mu_0$$

con $\mu_0 = 130$.

Logro 5: *Identificar Test* **[1.0 Ptos]**

Se debe realizar un test t-Student, ya que la σ es desconocido, con $T_0 = -1{,}297 \sim$ t-Student(24).

Logro 6: *Responder utilizando salida de **R*** **[1.0 Ptos]**

Desde la salida de **R** tenemos que el valor-p es igual a 10,35 %, que corresponde mínimo nivel de significancia que rechaza H_0 y apoya la hipótesis que la media de los precios es menor a 130 €.

+ 1 Punto Base