Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2017

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

Problema 1

Con la puesta en marcha de la línea 6 del metro, la autoridad ha indicado que los tiempos de viaje en Santiago han disminuido significativamente. De hecho, afirma que la probabilidad que un viaje tenga una duración inferior a 30 min es superior al 50 %.

Como una forma de validar o refutar su afirmación, se considera una muestra de 45 viajes-tipo de Santiago con un promedio de 35 min y desviación estándar 30 min.

Asumiendo que los tiempos de viaje siguen una distribución exponencial, ¿para qué nivel de significancia la autoridad tiene razón?

Solución

Definamos como T a los tiempos de viaje, donde $T \sim \text{Exp}(\nu)$.

La afirmación dice $P(T < 30) > \frac{1}{2}$. **[0.5 Ptos.]**

Alternativa 1: Realizar un test de hipótesis sobre $P(T < t) = 1 - \exp(-30\,\nu) = g(\nu)$. **[0.5 Ptos.]**

$$\text{H}_0 : g(\nu) = \frac{1}{2} \quad \text{vs} \quad \text{H}_a : g(\nu) > \frac{1}{2} \quad \textbf{[0.5 Ptos.]}$$

Bajo H_0 se tiene $g(\nu) = \frac{1}{2} \to \nu = \frac{\ln(2)}{30} = 0.0231 = \nu_0$. **[0.5 Ptos.]**

$$\textbf{[0.5 Ptos.]} \quad \hat{\nu} \overset{\text{aprox}}{\sim} \text{Normal}\left(\nu, \sqrt{\frac{\nu^2}{n}}\right) \to g(\hat{\nu}) \overset{\text{aprox}}{\sim} \text{Normal}\left(g(\nu), \sqrt{\frac{\nu^2\,[g'(\nu)]^2}{n}}\right) \quad \textbf{[0.5 Ptos.]}$$

El estadístico de prueba está dado por

$$\textbf{[0.5 Ptos.]} \quad Z_0 = \frac{g(\hat{\nu}) - g(\nu_0)}{\sqrt{\dfrac{\nu_0^2\,[g'(\nu_0)]^2}{n}}} = \frac{0.576 - 0.5}{\dfrac{0.0231 \cdot 30 \cdot \exp(-0.0231 \times 30)}{\sqrt{45}}} \approx 1.47 \quad \textbf{[1.0 Ptos.]}$$

Luego

$$\textbf{[0.5 Ptos.]} \quad \text{valor-p} = P(Z > 1.47) = 1 - 0.9292 = 0.0708. \quad \textbf{[0.5 Ptos.]}$$

Por lo tanto, para todo $\alpha > 0.0708$ se valida la afirmación de la autoridad. **[0.5 Ptos.]**

Alternativa 2: Realizar un test sobre ν para las siguientes hipótesis

$$\text{H}_0 : \nu = 0.0231 \quad \text{vs} \quad \text{H}_a : \nu > 0.0231 \quad \textbf{[1.5 Ptos.]}$$

A partir de la distribución aproximada de $\hat{\nu}$

$$\hat{\nu} \overset{\text{aprox}}{\sim} \text{Normal}\left(\nu, \sqrt{\frac{\nu^2}{n}}\right) \quad \textcolor{blue}{\textbf{[1.0 Ptos.]}}$$

El estadístico de prueba está dado por

$$\textcolor{blue}{\textbf{[0.5 Ptos.]}} \quad Z_0 = \frac{\hat{\nu} - \nu_0}{\sqrt{\frac{\nu_0^2}{n}}} = \frac{1/35 - 0.0231}{\sqrt{\frac{0.0231^2}{45}}} \approx 1.59 \quad \textcolor{blue}{\textbf{[1.0 Ptos.]}}$$

Luego

$$\textcolor{blue}{\textbf{[0.5 Ptos.]}} \quad \text{valor-p} = P(Z > 1.59) = 1 - 0.9441 = 0.0559. \quad \textcolor{blue}{\textbf{[0.5 Ptos.]}}$$

Por lo tanto, para todo $\alpha > 0.0559$ se valida la afirmación de la autoridad. **[0.5 Ptos.]**

Alternativa 3: Realizar un test sobre μ para las siguientes hipótesis

$$\text{H}_0 : \mu = 1/0.0231 \quad \text{vs} \quad \text{H}_a : \mu < 1/0.0231 \quad \textcolor{green}{\textbf{[1.5 Ptos.]}}$$

A partir de la distribución aproximada de $\hat{\mu}$

$$\hat{\mu} \overset{\text{aprox}}{\sim} \text{Normal}\left(\mu, \sqrt{\frac{\mu^2}{n}}\right) \quad \textcolor{green}{\textbf{[1.0 Ptos.]}}$$

El estadístico de prueba está dado por

$$\textcolor{green}{\textbf{[0.5 Ptos.]}} \quad Z_0 = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{\mu_0^2}{n}}} = \frac{35 - 1/0.0231}{\sqrt{\frac{(1/0.0231)^2}{45}}} \approx -1.28 \quad \textcolor{green}{\textbf{[1.0 Ptos.]}}$$

Luego

$$\textcolor{green}{\textbf{[0.5 Ptos.]}} \quad \text{valor-p} = P(Z < -1.28) = 1 - P(Z < 1.28) = 1 - 0.8997 = 0.1003. \quad \textcolor{green}{\textbf{[0.5 Ptos.]}}$$

Por lo tanto, para todo $\alpha > 0.1003$ se valida la afirmación de la autoridad. **[0.5 Ptos.]**

+ 1 Punto Base

Problema 2

La próxima semana se efectúa la 2da vuelta de la elección de una federación estudiantil y la competencia se vaticina cerrada (suponga que son las listas A y B).

Usted lleva a cabo una encuesta a 140 estudiantes en sus dos campus, cuyos resultados son:

	Campus 1	Campus 2
Número de entrevistados	60	80
Votaría Lista A	32	31
Votaría Lista B	18	34
No votaría	10	15

Considerando únicamente al votante probable:

(a) **[2.0 Puntos]** ¿Existe evidencia para afirmar que la lista A es mayoría? Use $\alpha = 5\,\%$.

(b) **[2.0 Puntos]** Con base a la información obtenida, ¿cuál debería ser el tamaño muestral de una segunda encuesta que permita estimar el apoyo a la lista A con un error no mayor al $6\,\%$ y un $90\,\%$ de confianza?

(b) **[2.0 Puntos]** Estime con un $90\,\%$ de confianza la diferencia en el apoyo a la lista B entre los dos campus.

Solución

(a) Considerando el total y solo al votante probable, se tiene

$$\hat{p}_A = \frac{63}{115} = 0.548 \quad \textbf{[0.5 Ptos.]}$$

Se pide testear

$$\mathrm{H}_0 : p_A = 0.5 \quad \text{vs} \quad \mathrm{H}_a : p_A > 0.5 \quad \textbf{[0.5 Ptos.]}$$

donde el estadístico de prueba bajo H_0 esta dado por

$$Z_0 = \frac{0.548 - 0.5}{\sqrt{\frac{0.5\times 0.5}{115}}} = 1.02 \quad \textbf{[0.5 Ptos.]}$$

Alternativa 1: Como $Z_0 = 1.02 \not> k_{1-\alpha} = 1.645$, no hay evidencia contra H_0, es decir no hay evidencia que sustente que la lista A sea mayoría. **[0.5 Ptos.]**

Alternativa 2: Como el valor-p $= 1 - \Phi(1.02) = 1 - 0.8461 = 0.8461 \not< 0.05 = \alpha$, no hay evidencia contra H_0, es decir no hay evidencia que sustente que la lista A sea mayoría. **[0.5 Ptos.]**

(b) De los datos se tiene: $\hat{p}_A = 0.548$ y que el $115/140 = 0.82$ de los entrevistados votaría. Así, el tamaño muestral mínimo estaría dado por

Alternativa 1

$$\textbf{[1.0 Ptos.]} \quad n = \left(\frac{1.645 \cdot \sqrt{0.548 \cdot 0.452}}{0.06}\right)^2 = 186.2 \to n = \frac{186.2}{0.82} = 227.0732 \to n = 228 \quad \textbf{[1.0 Ptos.]}$$

Alternativa 2 (Varianza máxima)

$$\textbf{[1.0 Ptos.]} \quad n = \left(\frac{1.645 \cdot \sqrt{0.5 \cdot 0.5}}{0.06}\right)^2 = 187.9184 \to n = \frac{186.2}{0.82} = 229.1688 \to n = 230 \quad \textbf{[1.0 Ptos.]}$$

(c) Tenemos que

$$\textbf{[0.4 Ptos.]} \quad \hat{p}_{c1} = \frac{18}{50} = 0.36 \quad \text{y} \quad \hat{p}_{c2} = \frac{34}{65} = 0.523 \quad \textbf{[0.4 Ptos.]}$$

Usando la fórmula del IC para la diferencia de proporciones, se tiene

$$(0.360 - 0.523) \pm 1.645\sqrt{\frac{0.36 \cdot 0.64}{50} + \frac{0.523 \cdot 0.477}{65}} \quad \textbf{[0.4 Ptos.]}$$
$$-0.163 \pm 0.151 \quad \textbf{[0.4 Ptos.]}$$

Es decir, con un 90 % de confianza la diferencia de apoyo entre el Campus 1 y Campus 2 esta entre

$$[-31.4\,\% \,;\, -1.2\,\%] \quad \textbf{[0.4 Ptos.]}$$

o la diferencia entre el Campus 2 y Campus 1 esta entre

$$[+1.2\,\% \,;\, +31.4\,\%] \quad \textbf{[0.4 Ptos.]}$$

+ 1 Punto Base

Problema 3

Frente a la próxima construcción de la línea 7 del metro, se proponen dos tipos de máquinas topos (o tuneladoras) para la perforación, denominadas TBM1 y TBM2 (TBM-tunnel Boring Machine). Un requerimiento de la licitación es una avance superior a 2 mts por hora. Para tomar la decisión se realizan múltiples pruebas con cada una de las máquinas obteniendo lo que sigue:

	TBM1	TBM2
Número de pruebas	32,00	35,00
Promedio	2,06	2,12
Desv. Estándar	0,15	0,24

Suponiendo que el avance por hora sigue una distribución normal y con un nivel de significancia del 5 %:

(a) **[2.0 Puntos]** ¿Cumplen con el requerimiento mínimo cada una de las máquinas?

(b) **[2.0 Puntos]** ¿Existe evidencia que las variabilidades difieren?

(c) **[2.0 Puntos]** ¿Difieren los avances medios obtenidos por cada máquina?

Solución

(a) Para cada máquina se efectúa el siguientes test para medias

$$\text{H}_0 : \mu = 2 \quad \text{vs} \quad \text{H}_a : \mu > 2 \quad \textbf{[0.4 Ptos.]}$$

Alternativa 1: Para la maquina TBM1 el estadístico de prueba está dado por

$$T_0 = \frac{2.06 - 2.0}{0.15/\sqrt{32}} = 2.26, \quad \textbf{[0.4 Ptos.]}$$

Como $T_0 = 2.26 > t_{0.95}(31) = 1.645$ se rechaza H_0, es decir, la maquina TBM1 satisface el requerimiento de la licitación. **[0.4 Ptos.]**

Para la maquina TBM2 el estadístico de prueba está dado por

$$T_0 = \frac{2.12 - 2.0}{0.24/\sqrt{35}} = 2.96, \quad \textbf{[0.4 Ptos.]}$$

Como $T_0 = 2.96 > t_{0.95}(34) = 1.645$ se rechaza H_0, es decir, la maquina TBM2 satisface el requerimiento de la licitación. **[0.4 Ptos.]**

Alternativa 2: Para la maquina TBM1 el estadístico de prueba está dado por

$$T_0 = \frac{2.06 - 2.0}{0.15/\sqrt{32}} = 2.26, \quad \textbf{[0.4 Ptos.]}$$

Como valor-p $= P(T > T_0) \approx 1 - \Phi(Z_0) = 1 - 0.9881 = 0.0119 < 0.05 = \alpha$ se rechaza H_0, es decir, la maquina TBM1 satisface el requerimiento de la licitación. **[0.4 Ptos.]**

Para la maquina TBM2 el estadístico de prueba está dado por

$$T_0 = \frac{2.12 - 2.0}{0.24/\sqrt{35}} = 2.96, \quad \textbf{[0.4 Ptos.]}$$

Como valor-p $= P(T > T_0) \approx 1 - \Phi(Z_0) = 1 - 0.9985 = 0.0015 < 0.05 = \alpha$ se rechaza H_0, es decir, la maquina TBM1 satisface el requerimiento de la licitación. **[0.4 Ptos.]**

(b) Se pide un test para comparación de varianzas

$$\mathrm{H}_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad \mathrm{H}_a : \sigma_1^2 \neq \sigma_2^2 \quad \textbf{[0.5 Ptos.]}$$

Bajo H_0 el estadístico

$$\textbf{[0.5 Ptos.]} \quad F = \frac{0.15^2}{0.24^2} = 0.39 \sim F(31,\, 34) \quad \textbf{[0.5 Ptos.]}$$

Como $F = 0.39 \not> F_{0.975}(31,\, 34) = 2.00$, pero si $F = 0.39 < 1/F_{0.975}(34,\, 31) = 1/2.03 = 0.493$ se rechaza H_0, es decir las varianzas difieren. **[0.5 Ptos.]**

(c) Se pide un test para comparaciones de medias, con σ desconocidas y distintas.

$$\mathrm{H}_0 : \mu_1 = \mu_2 \quad \text{vs} \quad \mathrm{H}_a : \mu_1 \neq \mu_2 \quad \textbf{[0.5 Ptos.]}$$

El estadístico de prueba bajo H:$_0$ está dado por

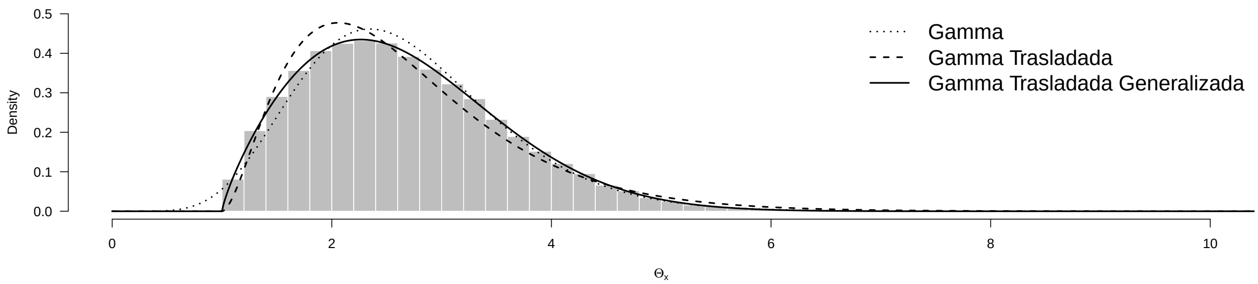
$$T_0 = \frac{2.06 - 2.12}{\sqrt{\dfrac{0.15^2}{32} + \dfrac{0.24^2}{35}}} = -1.238 \quad \textbf{[0.5 Ptos.]}$$

Como $|T_0| = 1.238 \not> t_{0.975}(52) = k_{0.975} = 1.96$ no hay evidencia contra H_0, es decir, no hay evidencia que los avances medios difieran. **[1.0 Ptos.]**

+ 1 Punto Base

Problema 4

Una distribución de probabilidad que permite mayor grado de libertad en su forma para modelar datos con soporte distinto a $\mathbb{R}_0^+$ como lo hace clásica Gamma, es la Gamma-Generalizada$(k,\, \nu,\, \beta)$ trasladada en α. La figura muestra el ajuste de extensiones de una Gamma a un conjunto de mediciones, donde la Gamma-Generalizada trasladada es la que logra un mejor resultado.



Esta este modelo tiene una función de densidad dada por:

$$f(x) = \frac{\beta\, \nu^{\beta\, k}}{\Gamma(k)}\, (x - \alpha)^{\beta\, k - 1} \exp\left\{-[\nu\, (x - \alpha)]^{\beta}\right\}, \quad x \geq \alpha$$

con $k > 0$, $\nu > 0$, $\beta > 0$, $\alpha \in \mathbb{R}$ y $\text{E}[(X - \alpha)^r] = \dfrac{\Gamma\left(k + \frac{r}{\beta}\right)}{\nu^r\, \Gamma(k)}$.

Para una muestra aleatoria $X_1,\, X_2, \ldots,\, X_n$ y asumiendo que los parámetros k, ν y β son conocidos:

(a) **[3.0 Puntos]** Obtenga el estimador de momento $\hat{\alpha}$ de α y calcule su error cuadrático medio.

(b) **[2.0 Puntos]** El estimador máximo verosímil $\widetilde{\alpha}$ de α es el $\text{mín}\{X_1, \ldots, X_n\}$. A partir de su distribución aproximada, obtenga su error cuadrático medio, por simplicidad entregue su aproximación de 1er orden.

(c) **[1.0 Puntos]** Considere que $k = 2$, $\beta = 1$ y $\nu = 1$. Muestre que en este caso el estimador máximo verosímil no resulta más eficiente que el estimador de momento. ¿Por qué pasa esto?

Solución

(a) Consideremos una muestra aleatoria $X_1, \ldots, X_n$ proveniente de una población Gamma-Generalizada$(k,\, \nu,\, \beta)$ trasladada en α, con k, ν y β parámetros conocidos. Se pide estimar α por método de momentos.

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

$$\text{E}(X) = \frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu\, \Gamma(k)} + \alpha \quad \textbf{[0.4 Ptos.]}$$

igualando al 1er momento empírico $\overline{X}_n$, se tiene que el estimador de momento $\hat{\alpha}$ de α es

$$\textbf{[0.4 Ptos.]} \quad \hat{\alpha} = \overline{X}_n - \frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu \cdot \Gamma(k)} \to \text{E}(\hat{\alpha}) = \frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu\, \Gamma(k)} + \alpha - \frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu \cdot \Gamma(k)} = \alpha \quad \textbf{[0.4 Ptos.]}$$

Por otra parte, el 2do momento teórico de una Gamma-Generalizada trasladada está dado por

$$\text{E}(X^2) = \frac{\Gamma\left(k + \frac{2}{\beta}\right)}{\nu^2\, \Gamma(k)} + 2\, \alpha\, \frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu\, \Gamma(k)} - \alpha^2, \quad \textbf{[0.4 Ptos.]}$$

Por lo tanto,

$$\textbf{[0.4 Ptos.]} \quad \text{Var}(X) = \frac{\Gamma\left(k + \frac{2}{\beta}\right)}{\nu^2\,\Gamma(k)} - \left[\frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu\,\Gamma(k)}\right]^2 \to \text{Var}(\hat{\alpha}) = \frac{1}{n}\left\{\frac{\Gamma\left(k + \frac{2}{\beta}\right)}{\nu^2\,\Gamma(k)} - \left[\frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu\,\Gamma(k)}\right]^2\right\} \quad \textbf{[0.4 Ptos.]}$$

$$= \text{ECM}(\hat{\alpha}) \quad \textbf{[0.2 Ptos.]}$$

ya que $\hat{\alpha}$ es insesgado para α. **[0.4 Ptos.]**

(b) Tenemos que

$$\textbf{[0.2 Ptos.]} \quad \widetilde{\alpha} \overset{\text{aprox}}{\sim} \text{Normal}\left(\alpha,\, \sqrt{\frac{1}{I_n(\alpha)}}\right) \to \text{ECM}(\widetilde{\alpha}) = \frac{1}{I_n(\alpha)} \quad \textbf{[0.2 Ptos.]}$$

donde, $I_n(\alpha) = -\text{E}\left[\frac{\partial^2}{\partial\alpha^2}\ln L(\alpha)\right]$. **[0.2 Ptos.]**

La verosimilitud, su logaritmo natural y las primeras dos derivadas con respecto a α son:

$$L(\alpha) = \frac{\beta^n\,\nu^{n\,\beta\,k}}{[\Gamma(k)]^n} \times \left[\prod_{i=1}^n (X_i - \alpha)\right]^{\beta\,k-1} \times \exp\left[-\nu^\beta \sum_{i=1}^n (X_i - \alpha)^\beta\right] \quad \textbf{[0.2 Ptos.]}$$

$$\ln L(\alpha) = n\,\ln(\beta) + n\,\beta\,k\,\ln(\nu) + (\beta\,k - 1)\sum_{i=1}^n \ln(X_i - \alpha) - \nu^\beta \sum_{i=1}^n (X_i - \alpha)^\beta - n\,\ln[\Gamma(k)] \quad \textbf{[0.2 Ptos.]}$$

$$\frac{\partial}{\partial\alpha}\ln L(\alpha) = -(\beta\,k - 1)\sum_{i=1}^n \frac{1}{(X_i - \alpha)} + \nu^\beta \sum_{i=1}^n \beta\,(X_i - \alpha)^{\beta-1} \quad \textbf{[0.2 Ptos.]}$$

$$\frac{\partial^2}{\partial\alpha^2}\ln L(\alpha) = -(\beta\,k - 1)\sum_{i=1}^n \frac{1}{(X_i - \alpha)^2} - \nu^\beta \sum_{i=1}^n \beta\,(\beta - 1)\,(X_i - \alpha)^{\beta-2} \quad \textbf{[0.2 Ptos.]}$$

Luego, como la muestra es iid, entonces la aproximación de 1er orden para $I_n(\alpha)$ es

$$I_n(\alpha) \approx (\beta\,k - 1)\,\frac{n}{(\mu_X - \alpha)^2} + \nu^\beta\,n\,\beta\,(\beta - 1)\,(\mu_X - \alpha)^{\beta-2} \quad \textbf{[0.2 Ptos.]}$$

$$= n\left\{\frac{(\beta\,k - 1)\,\nu^2\,\Gamma^2(k)}{\Gamma^2\left(k + \frac{1}{\beta}\right)} + \nu^\beta\,\beta\,(\beta - 1)\left[\frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu\,\Gamma(k)}\right]^{\beta-2}\right\} \quad \textbf{[0.2 Ptos.]}$$

Por lo tanto

$$\text{ECM}(\widetilde{\alpha}) \approx \frac{1}{n}\cdot\left\{\frac{(\beta\,k - 1)\,\nu^2\,\Gamma^2(k)}{\Gamma^2\left(k + \frac{1}{\beta}\right)} + \nu^\beta\,\beta\,(\beta - 1)\left[\frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu\,\Gamma(k)}\right]^{\beta-2}\right\}^{-1} \quad \textbf{[0.2 Ptos.]}$$

(c) Reemplazando en (a) y (b) tenemos que

$$\textbf{[0.4 Ptos.]} \quad \text{ECM}(\hat{\alpha}) = \frac{2}{n} \qquad \text{y} \qquad \text{ECM}(\widetilde{\alpha}) = \frac{4}{n} \quad \textbf{[0.4 Ptos.]}$$

Es decir, el estimador de momento es más eficiente. Esto ocurre por la doble aproximación que se hizo en el EMC del estimador máximo verosímil. **[0.2 Ptos.]**

+ 1 Punto Base