Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2016

Curso	**:**	**Probabilidad y Estadística**
Sigla	**:**	**EYP1113**
Interrogación	**:**	**3**
Profesores	**:**	**Ricardo Aravena C. y Ricardo Olea O.**

Problema 1

El fin de semana recién pasado muchos Santiaguinos aprovecharon de salir fuera de la capital, provocando atochamientos en los distintos puntos de salida. Usted ha conseguido información desde los peajes de Angostura (Ruta 5) y Lo Prado (Ruta 68) correspondiente al número de autos que salían desde las casetas y los tiempos de espera de los automovilistas entre las 18.00 y 21.00 hrs del del viernes pasado.

(a) **[3.0 Ptos.]** Al tomar un minuto cualquiera desde los registros enviados desde los dos peajes, usted ve que 41 vehículos pagaron en las 12 casetas de Lo Prado y 38 en las 16 casetas de Angostura. ¿Con esta información es posible afirmar que el peaje de Lo Prado es más eficiente? Utilice un nivel de significancia del 5 % y que el número de vehículos que pagan por minuto distribuyen Poisson.

(b) **[3.0 Ptos.]** Una encuesta sobre los tiempos de espera realizada a 18 conductores en Lo Prado y 14 en Angostura, justo después de pagar, arrojó que en promedio los conductores que pasaron por el peaje de Lo Prado demoraron 32 minutos y los conductores que pasaron por Angostura demoraron 48 minutos. Asumiendo que los tiempos de espera para pasar se comportan de acuerdo a una distribución Exponencial, ¿es posible afirmar que la Ruta 5 ha estado más congestionada, es decir, toma más tiempo cruzar? Utilice un nivel de significancia del 10 %.

Solución

(a) Definamos como $X_1, \ldots, X_n$ al número de vehículos que pagaron en el minuto de observación en las n casetas del peaje Lo Prado y como $Y_1, \ldots, Y_m$ al número de vehículos que pagaron en el minuto de observación en las m casetas del peaje Angostura.

Del enunciado se desprende que X_i's $\overset{\text{iid}}{\sim}$ Poisson(λ_X) e Y_i's $\overset{\text{iid}}{\sim}$ Poisson(λ_Y).

Se pide contrastar las siguientes hipótesis

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

Si H_0 es correcta ($\lambda_X = \lambda_Y = \lambda$) se tiene que

$$Z_0 = \frac{\hat{\lambda}_X - \hat{\lambda}_Y}{\sqrt{\hat{\lambda}\left(\frac{1}{n} + \frac{1}{m}\right)}} \overset{\text{aprox}}{\sim} \text{Normal}(0, 1) \quad \textbf{[0.5 Ptos.]}$$

con

$$\hat{\lambda}_X = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{41}{12}, \quad \textbf{[0.2 Ptos.]} \qquad \hat{\lambda}_Y = \frac{1}{m}\sum_{j=1}^{m} Y_j = \frac{38}{16}, \quad \textbf{[0.2 Ptos.]}$$

$$\hat{\lambda} = \frac{1}{n+m}\left(\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j\right) = \frac{41+38}{12+16} \quad \textbf{[0.4 Ptos.]}$$

Reemplazando, tenemos que $Z_0 = 1.623925$. **[0.5 Ptos.]**

Alternativa 1: El valor-p $= 1 - \Phi(Z_0) \approx 1 - \Phi(1.62) = 1 - 0.9474 = 0.0526 > 0.05 = \alpha$. **[0.4 Ptos.]**

Por lo tanto, no existe suficiente evidencia para rechazar H_0 y poder afirmar que el peaje de Lo Prado es más eficiente. **[0.3 Ptos.]**

Alternativa 2: El estadístico de prueba $Z_0 = 1.623925 < 1.645 = k_{1-\alpha}$. **[0.4 Ptos.]**

Por lo tanto, no existe suficiente evidencia para rechazar H_0 y poder afirmar que el peaje de Lo Prado es más eficiente. **[0.3 Ptos.]**

(b) Definamos como $X_1, \ldots, X_n$ los tiempos que demoraron en pasar los vehículos encuestados en el peaje de Lo Prado y como $Y_1, \ldots, Y_m$ los tiempos que demoraron en pasar los vehículos encuestados en el peaje de peaje Angostura.

Del enunciado se desprende que X_i's $\overset{\text{iid}}{\sim}$ Exponencial(ν_X) e Y_i's $\overset{\text{iid}}{\sim}$ Exponencial(ν_Y).

Se pide contrastar las siguientes hipótesis

$$\text{H}_0 : \mu_X = \mu_Y \quad \text{vs} \quad \text{H}_a : \mu_X < \mu_Y \quad \textbf{[0.5 Ptos.]}$$

con $\mu_X = \frac{1}{\nu_X}$ y $\mu_Y = \frac{1}{\nu_Y}$.

Si H_0 es correcta ($\mu_X = \mu_Y = \mu$ o $\nu_X = \nu_Y = \nu$) se tiene que

$$Z_0 = \frac{\hat{\mu}_X - \hat{\mu}_Y}{\sqrt{\hat{\mu}^2\left(\frac{1}{n} + \frac{1}{m}\right)}} = \frac{\frac{1}{\hat{\nu}_X} - \frac{1}{\hat{\nu}_Y}}{\sqrt{\frac{1}{\hat{\nu}^2}\left(\frac{1}{n} + \frac{1}{m}\right)}} \overset{\text{aprox}}{\sim} \text{Normal}(0, 1) \quad \textbf{[0.5 Ptos.]}$$

con

$$\hat{\mu}_X = \overline{X}_n = 32, \quad \textbf{[0.2 Ptos.]} \qquad \hat{\mu}_Y = \overline{Y}_m = 48, \quad \textbf{[0.2 Ptos.]}$$

$$\hat{\mu} = \frac{n \cdot \overline{X}_n + m \cdot \overline{Y}_m}{n + m} = \frac{18 \cdot 32 + 14 \cdot 48}{18 + 14} \quad \textbf{[0.4 Ptos.]}$$

Reemplazando, tenemos que $Z_0 = -1.151279$. **[0.5 Ptos.]**

Alternativa 1: El valor-p $= \Phi(Z_0) \approx \Phi(-1.15) = 1 - \Phi(1.15) = 1 - 0.8749 = 0.1251 > 0.10 = \alpha$. **[0.4 Ptos.]**

Por lo tanto, no existe suficiente evidencia para rechazar H_0 y poder afirmar que en el peaje de Angostura estuvo más congestionado. **[0.3 Ptos.]**

Alternativa 2: El estadístico de prueba $Z_0 = -1.151279 > -1.28 = k_\alpha$. **[0.4 Ptos.]**

Por lo tanto, no existe suficiente evidencia para rechazar H_0 y poder afirmar que en el peaje de Angostura estuvo más congestionado. **[0.3 Ptos.]**

+ 1 Punto Base

Problema 2

A usted le encargan la construcción de un estimador insesgado para ζ^2 a partir de una muestra aleatoria proveniente de una Población Log-Normal(λ, ζ), basandose en el estimador de máxima verosimilitud de ζ. Proponga una distribución exacta para ambos estimadores, obtenga sus errores cuadráticos medios y comente.

Solución

Sea $X_1, \ldots, X_n$ una muestra aleatoria proveniente de una población Log-Normal(λ, ζ).

El estimador máximo verosímil de ζ^2 esta dado por

$$\hat{\zeta}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\ln(X_i) - \hat{\lambda}\right]^2 \quad \textbf{[0.5 Ptos.]}$$

con $\hat{\lambda} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\ln(X_i)$.

Notemos que

$$\ln(X_i) \sim \text{Normal}(\lambda,\, \zeta) \to \frac{n\,\hat{\zeta}^2}{\zeta^2} = \sum_{i=1}^{n}\left[\frac{\ln(X_i) - \hat{\lambda}}{\zeta}\right]^2 \sim \chi^2(n-1) = \text{Gamma}\left(\frac{n-1}{2},\, \frac{1}{2}\right)$$
$$\to \hat{\zeta}^2 \sim \text{Gamma}\left(\frac{n-1}{2},\, \frac{n}{2\,\zeta^2}\right) \quad \textbf{[0.5 Ptos.]}$$

con

$$\textbf{[0.5 Ptos.]} \quad \text{E}(\hat{\zeta}^2) = \frac{(n-1)\,\zeta^2}{n} \quad \text{y} \quad \text{Var}(\hat{\zeta}^2) = \frac{2\,\zeta^4\,(n-1)}{n^2} \quad \textbf{[0.5 Ptos.]}$$

Se pide un estimador insesgado a partir de $\hat{\zeta}^2$

$$\widetilde{\zeta}^2 = \frac{n}{(n-1)}\,\hat{\zeta}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\ln(X_i) - \hat{\lambda}\right]^2 \ \textbf{[0.5 Ptos.]} \sim \text{Gamma}\left(\frac{n-1}{2},\, \frac{n-1}{2\,\zeta^2}\right) \quad \textbf{[0.5 Ptos.]}$$

con

$$\textbf{[0.5 Ptos.]} \quad \text{E}(\hat{\zeta}^2) = \zeta^2 \quad \text{y} \quad \text{Var}(\hat{\zeta}^2) = \frac{2\,\zeta^4}{n-1} \quad \textbf{[0.5 Ptos.]}$$

Se pide

$$\textbf{[0.5 Ptos.]} \quad \text{ECM}(\hat{\zeta}^2) = \frac{\zeta^4(2n-1)}{n^2} \quad \text{y} \quad \text{ECM}(\widetilde{\zeta}^2) = \frac{2\,\zeta^4}{n-1} \quad \textbf{[0.5 Ptos.]}$$

dividiendo vemos que

$$\frac{\text{ECM}(\hat{\zeta}^2)}{\text{ECM}(\widetilde{\zeta}^2)} = 1 - \frac{1}{n} \quad \textbf{[0.5 Ptos.]}$$

es decir, el estimador máximo verosímil pese a que es un estimador sesgado es más eficiente. **[0.5 Ptos.]**

+ 1 Punto Base

Problema 3

Existe molestias por parte de los estudiantes de San Joaquín en relación a las dificultades de encontrar estacionamiento después de las 09.00 hrs. Usted, buscando responder algunas inquietudes recoge información sobre los tiempos que le toma a alumnos(as) que concurren en auto al campus en estacionar.

Tipo	No Molestos	Molestos
Tamaño muestra	30	20
Promedios	12.8	15.4
Desv. Estándar	4.0	6.8

¿Es válido afirmar que a los molestos les toma un mayor tiempo la búsqueda de estacionamiento? Asuma independencia y normalidad de los tiempos registrados. Use un nivel de significancia del 10 %.

Solución

Se pide realizar una prueba de hipótesis para la comparación de medias entre No Molestos (X) y Molestos (Y), con varianzas poblacionales desconocidas (¿iguales o distintas?)

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

Alternativa 1: Si H_0 es correcta, se tiene que

$$F_0 = \frac{S_X^2}{S_Y^2} \sim F(29,\, 19) \quad \textbf{[0.5 Ptos.]}$$

Se rechaza la igualdad de varianzas si: $F_0 < F_{\alpha/2}(29,\, 19)$ o $F_0 > F_{1-\alpha/2}(29,\, 19)$, con $\alpha = 0.10$. **[0.5 Ptos.]**

Reemplazando tenemos que

$$F_0 = 0.3460208, \quad F_{0.95}(29,\, 19) = 2.08, \quad F_{0.05}(29,\, 19) = \frac{1}{F_{0.95}(19,\, 29)} = \frac{1}{1.96} = 0.5102041 \quad \textbf{[0.5 Ptos.]}$$

Es decir, existe suficiente evidencia para rechazar la igualdad de varianzas. **[0.5 Ptos.]**

Alternativa 2: Si H_0 es correcta, se tiene que

$$F_0 = \frac{S_Y^2}{S_X^2} \sim F(19,\, 29) \quad \textbf{[0.5 Ptos.]}$$

Se rechaza la igualdad de varianzas si: $F_0 < F_{\alpha/2}(19,\, 29)$ o $F_0 > F_{1-\alpha/2}(19,\, 29)$. **[0.5 Ptos.]**

Reemplazando tenemos que

$$F_0 = 2.89, \quad F_{0.95}(19,\, 29) = 1.96, \quad F_{0.05}(19,\, 29) = \frac{1}{F_{0.95}(29,\, 19)} = \frac{1}{2.08} = 0.4807692 \quad \textbf{[0.5 Ptos.]}$$

Es decir, existe suficiente evidencia para rechazar la igualdad de varianzas. **[0.5 Ptos.]**

Se pide contrastar las siguientes hipótesis

$$\mathrm{H}_0 : \mu_X = \mu_Y \quad \text{vs} \quad \mathrm{H}_a : \mu_X < \mu_Y \quad \textbf{[0.5 Ptos.]}$$

Si H_0 es correcto se tiene que

$$T_0 = \frac{(\overline{X}_n - \overline{Y}_m)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim \text{t-Student}(\nu)$$

con

$$\nu = \left[\frac{\left(S_X^2/n + S_Y^2/m\right)^2}{\frac{\left(S_X^2/n\right)^2}{n-1} + \frac{\left(S_Y^2/m\right)^2}{m-1}} \right]$$

Reemplazando se tiene que $T_0 = -1.54137$ **[0.5 Ptos.]** y $\nu = [27.80744] = 27$. **[1.0 Ptos.]** (el alumno en vez de truncar podría haber aproximado a 28, no descontar puntaje)

Alternativa 1: El valor-p $= P(T < -1.54137) \to 5\,\% <$ valor-p $< 10\,\% = \alpha$, para $T \sim$ t-Studente(ν). **[1.0 Ptos.]**

Por lo tanto, existe suficiente evidencia para rechazar H_0, es decir se apoya la idea que a los Molestos les toma un mayor tiempo la búsqueda de estacionamiento. **[0.5 Ptos.]**

Alternativa 2: El estadístico de prueba $T_0 = -1.54137 < -1.314 = t_{0.10}(27)$ (el alumno podría haber utilizado $-1.313 = t_{0.10}(28)$, no descontar puntaje). **[1.0 Ptos.]**

Por lo tanto, existe suficiente evidencia para rechazar H_0, es decir se apoya la idea que a los Molestos les toma un mayor tiempo la búsqueda de estacionamiento. **[0.5 Ptos.]**

+ 1 Punto Base

Problema 4

Las series financieras, por ejemplo los retornos diarios (ganancia porcentual) de la acción de Cencosud tiene algunas características especiales, como son la volatilidad y colas pesadas, las cuales se pueden observar a continuación:

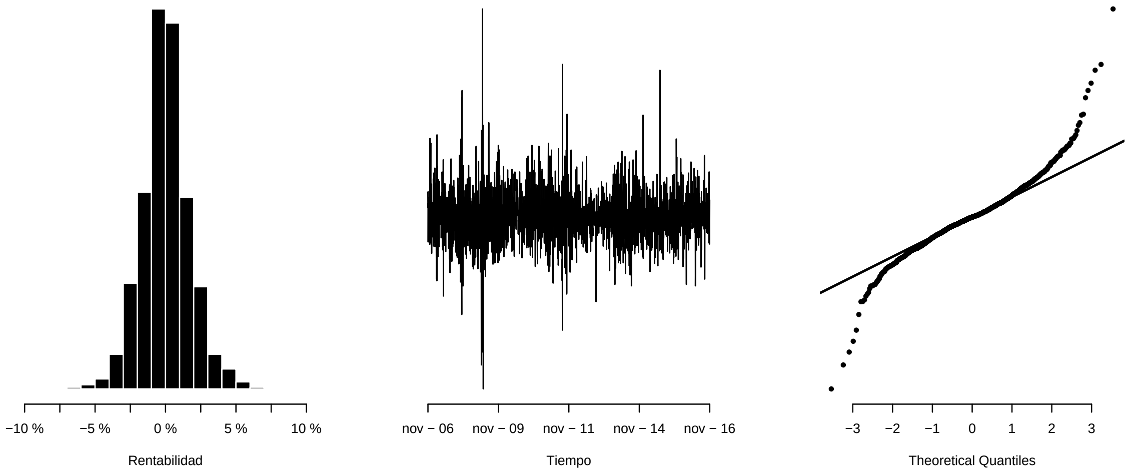

Figura 1: Rentabilida Diaria Cencosud nov06 - nov16

Una modelo muy popular que logra representar estas característica es el siguiente

$$Y_t = \mu + X_t, \quad X_t = (\alpha + \beta\, X_{t-1}^2)^{1/2} \cdot Z_t, \quad \{Z_t\} \overset{\text{iid}}{\sim} \text{Normal}(0,1), \quad t \in \mathbb{Z}$$

Suponga que β es conocido e igual a $1/2$. Además considere que $\text{E}(X_s^k) = \text{E}(X_r^k)$ para todo $s \in \mathbb{Z}$, $r \in \mathbb{Z}$ y $k \in \mathbb{N}$.

Obtenga y evalúe en base a los datos históricos los estimadores de momento de μ y α.

```
   n  mean(Y) mean(Y^2) mean(Y^3) mean(Y^4)
2493 0.032187  3.378454  2.352926  100.6172
```

Solución

Alternativa 1:

Tenemos que

$$\begin{aligned}
\text{E}(Y_t) &= \mu + \text{E}(X_t) \quad \textbf{[0.5 Ptos.]} \\
&= \mu + \text{E}\left[(\alpha + \beta\, X_{t-1}^2)^{1/2} \cdot Z_t\right] \quad \textbf{[0.4 Ptos.]} \\
&= \mu + \text{E}\left[(\alpha + \beta\, X_{t-1}^2)^{1/2}\right] \cdot \text{E}\,(Z_t)\,, \quad \text{por la independencia de los } \{Z_t\} \quad \textbf{[0.4 Ptos.]} \\
&= \mu + \text{E}\left[(\alpha + \beta\, X_{t-1}^2)^{1/2}\right] \cdot 0 \quad \textbf{[0.4 Ptos.]} \\
&= \mu + 0 \quad \textbf{[0.4 Ptos.]} \\
&= \mu \quad \textbf{[0.4 Ptos.]}
\end{aligned}$$

Luego

$$\textbf{[0.5 Ptos.]} \quad \text{E}(Y_t^2) = \mu^2 + 2\,\mu\,\text{E}(X_t) + \text{E}(X_t^2) = \mu^2 + \text{E}(X_t^2) \quad \textbf{[0.5 Ptos.]}$$

Alternativa 2:

Tenemos que

$$\mathrm{E}(Y_t) = \mu + \mathrm{E}(X_t) \quad \textbf{[0.5 Ptos.]}$$

y como una aproximación (1er orden) para X_t está dada por

$$X_t \approx (\alpha + \beta\,\mu_X^2)^{1/2} \cdot \mu_Z + (X_{t-1} - \mu_X) \cdot \frac{1}{2}(\alpha + \beta\,\mu_X^2)^{-1/2} \cdot \mu_Z + (Z_t - \mu_Z) \cdot (\alpha + \beta\,\mu_X^2)^{1/2} \quad \textbf{[1.0 Ptos.]}$$

con $\mu_X = \mathrm{E}(X_{t-1})$ y $\mu_Z = \mathrm{E}(Z_t) = 0$, entonces

$$\mathrm{E}(X_t) \approx 0 \to \mathrm{E}(Y) \approx \mu \quad \textbf{[1.0 Ptos.]}$$

Luego

$$\textbf{[0.5 Ptos.]} \quad \mathrm{E}(Y_t^2) = \mu^2 + 2\,\mu\,\mathrm{E}(X_t) + \mathrm{E}(X_t^2) \approx \mu^2 + \mathrm{E}(X_t^2) \quad \textbf{[0.5 Ptos.]}$$

Como

$$\begin{aligned}
\mathrm{E}(X_t^2) &= \mathrm{E}\left[(\alpha + \beta\,X_{t-1}^2) \cdot Z_t^2\right] \quad \textbf{[0.3 Ptos.]} \\
&= \mathrm{E}\left(\alpha + \beta\,X_{t-1}^2\right) \cdot \mathrm{E}\left(Z_t^2\right), \quad \text{por la independencia de los } \{Z_t\} \quad \textbf{[0.5 Ptos.]} \\
&= \mathrm{E}\left[(\alpha + \beta\,X_{t-1}^2)\right] \cdot 1 \quad \textbf{[0.3 Ptos.]} \\
&= \alpha + \beta\,\mathrm{E}\left(X_{t-1}^2\right) \quad \textbf{[0.3 Ptos.]} \\
\to \mathrm{E}(X_t^2) &= \frac{\alpha}{1-\beta} = 2\,\alpha, \quad \text{ya que} \quad \mathrm{E}(X_s^k) = \mathrm{E}(X_r^k) \quad \textbf{[0.5 Ptos.]}
\end{aligned}$$

Entonces

$$\textbf{[0.3 Ptos.]} \quad \hat{\mu} = \overline{Y} = 0.032187 \quad \text{y} \quad \hat{\alpha} = \frac{\overline{Y^2} - (\overline{Y})^2}{2} = 1.688709 \quad \textbf{[0.3 Ptos.]}$$

+ 1 Punto Base