Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2018

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

Pauta Interrogación 3

Pregunta 1

Ayer martes, Estados Unidos celebró sus elecciones legislativas. Al cierre se observó que el nivel de participación (proporción de votantes) por distrito se comporto según una variable aleatoria Beta(α, α). Una muestra aleatoria de n distritos entregó el nivel de participación en cada caso. Obtenga el estimador de momentos del parámetro α y su error cuadrático medio aproximado de 1er orden.

Solución

Tenemos $X_1,\ldots, X_n$ variable aleatorias iid Beta(α, α) en [0, 1].

Como el 1er momento de este modelo no depende de α, se procede a igualar 2do momento teórico, $\text{E}(X^2)$, al 2do momento empírico, $\overline{X^2}$:

$$\mu_{X^2} = \text{E}(X^2) = \sigma_X^2 + \mu_X^2 = \frac{\alpha^2}{(2\,\alpha)^2\,(2\,\alpha+1)} + \left(\frac{\alpha}{2\,\alpha}\right)^2 = \frac{1}{2}\left(\frac{\alpha+1}{2\,\alpha+1}\right) = \overline{X^2} \qquad \textbf{[1.0 Ptos.]}$$

Por lo tanto, el estimador de momento $\hat{\alpha}$ de α es

$$\hat{\alpha} = \frac{2\,\overline{X^2}-1}{1-4\,\overline{X^2}} \qquad \textbf{[1.0 Ptos.]}$$

Se pide $\text{ECM}(\hat{\alpha}) = \text{Var}(\hat{\alpha}) + [\text{E}(\hat{\alpha}) - \alpha]^2$.

La aproximación de 1er orden del valor esperado está dada por:

$$\textbf{[1.0 Ptos.]} \qquad \text{E}(\hat{\alpha}) \approx \frac{2\,\mu_{\overline{X^2}}-1}{1-4\,\mu_{\overline{X^2}}} = \frac{2\,\mu_{X^2}-1}{1-4\,\mu_{X^2}} = \frac{2\cdot\frac{1}{2}\left(\frac{\alpha+1}{2\,\alpha+1}\right)-1}{1-4\cdot\frac{1}{2}\left(\frac{\alpha+1}{2\,\alpha+1}\right)} = \alpha \qquad \textbf{[1.0 Ptos.]}$$

entonces

$$\text{ECM}(\hat{\alpha}) \approx \text{Var}(\hat{\alpha})$$

Si definimos $Y_i = X_i^2$ y al ser variables aleatorias con idéntica distribución ($\mu = \mu_{Y_i}$, $\sigma = \sigma_{Y_i}$):

$$\hat{\alpha} = g(Y_1,\ldots,Y_n) = \frac{2\displaystyle\sum_{i=1}^n Y_i - n}{n - 4\displaystyle\sum_{i=1}^n Y_i} \approx \left(\frac{2\,\mu-1}{1-4\,\mu}\right) + \sum_{i=1}^n (Y_i-\mu)\left(-\frac{2}{n\,(1-4\,\mu)^2}\right) \qquad \textbf{[1.0 Ptos.]}$$

Luego, por independencia

$$\textbf{[0.5 Ptos.]} \quad \text{Var}(\hat{\alpha}) \approx \frac{4\,\sigma^2}{n\,(1-4\,\mu)^4} = \frac{(2\,\alpha+1)^4\,\sigma^2}{4\,n} = \frac{(2\,\alpha+1)^4}{4\,n}\cdot\left[\frac{\Gamma(4+\alpha)\,\Gamma(2\,\alpha)}{\Gamma(4+2\,\alpha)\,\Gamma(\alpha)} - \frac{1}{4}\left(\frac{\alpha+1}{2\,\alpha+1}\right)^2\right] \quad \textbf{[0.5 Ptos.]}$$

+ 1 Punto Base

Pregunta 2

En el Ironman los participantes tienen que cubrir 3 distancias, 3,86 km de natación, 180 km de ciclismo y 42,2 km de carrera a pie. Los atletas tienen diferentes técnicas que le permiten optimizar su tiempo total. Un estudio realizado el año 2000 por Whyte G, Lumley S, George K, et al. *Physiological profile and predictors of cycling performance in ultra-endurance triathletes* (J Sports Med Phys Fitness) muestra resultados y asociaciones para los tiempos por distancia de un conjunto de superatletas que terminaron los Ironman.

```
          mean sd  N
Natación    76 22 40
Bicicleta  385 25 40
Maratón    294 35 40

cor(Muestra):
          Natación Bicicleta Maratón
Natación       1.0       0.2     0.3
Bicicleta      0.2       1.0    -0.4
Maratón        0.3      -0.4     1.0
```

Asumiendo una distribución Normal:

(a) **[2.0 Ptos]** Calcule el tiempo máximo logrado en el Iroman por el 10 % de los mejores superatletas.

(b) **[2.0 Ptos]** Obtenga la probabilidad que el tiempo natación-maratón sea mayor al tiempo de bicicleta.

(c) **[2.0 Ptos]** ¿Existe evidencia para afirmar que los superatletas demoran menos de 6 horas y media en el tramo de bicicleta? Use un nivel de significancia del 10 %.

Solución

(a) Definamos como X, Y y Z los tiempos de natación, bicicleta y maratón respectivamente, todos con distribución Normal.

$$T = X + Y + Z \sim \text{Normal}(\mu_T, \sigma_T) \qquad \textbf{[0.3 Ptos.]}$$

con

$$\mu_T = \mu_X + \mu_Y + \mu_Z \qquad \textbf{[0.3 Ptos.]}$$

y

$$\sigma_T = \sqrt{\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 + 2 \cdot 0{,}2 \cdot \sigma_X \cdot \sigma_Y + 2 \cdot 0{,}3 \cdot \sigma_X \cdot \sigma_Z - 2 \cdot 0{,}4 \cdot \sigma_Y \cdot \sigma_Z} \qquad \textbf{[0.5 Ptos.]}$$

Reemplazando los datos empíricos como estimadores de los parámetros

$$\textbf{[0.3 Ptos.]} \qquad \mu_T = 755 \qquad \text{y} \qquad \sigma_T = 48{,}12484 \qquad \textbf{[0.3 Ptos.]}$$

Se pide, percentil 10 % igual a

$$t_{10\,\%} = \mu_T + \sigma_T \cdot \Phi^{-1}(0{,}1) \approx 755 - 48{,}12484 \cdot 1{,}28 = 693{,}4002 \text{ (11 horas y 33.4 minutos)} \qquad \textbf{[0.3 Ptos.]}$$

(b) Sea

$$W = X - Y + Z \sim \text{Normal}(\mu_W, \sigma_W) \qquad \textbf{[0.3 Ptos.]}$$

con

$$\mu_W = \mu_X - \mu_Y + \mu_Z \qquad \textbf{[0.3 Ptos.]}$$

y

$$\sigma_W = \sqrt{\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 - 2 \cdot 0{,}2 \cdot \sigma_X \cdot \sigma_Y + 2 \cdot 0{,}3 \cdot \sigma_X \cdot \sigma_Z + 2 \cdot 0{,}4 \cdot \sigma_Y \cdot \sigma_Z} \qquad \textbf{[0.3 Ptos.]}$$

Reemplazando los datos empíricos como estimadores de los parámetros

$$\textbf{[0.3 Ptos.]} \qquad \mu_W = -15 \qquad \text{y} \qquad \sigma_W = 57{,}23635 \qquad \textbf{[0.3 Ptos.]}$$

Se pide,

$$\textbf{[0.3 Ptos.]} \quad P(W > 0) = 1 - \Phi\left(\frac{0-(-15)}{57{,}23635}\right) = 1 - \Phi(0{,}2620712)$$
$$\approx 1 - \Phi(0{,}26) = \Phi(0{,}26) = 1 - 0{,}6026 = 0{,}3974 \quad \textbf{[0.2 Ptos.]}$$

(c) Se pide contrastar

$$\mathrm{H}_0 : \mu = \mu_0 \qquad \text{vs} \qquad \mathrm{H}_a : \mu < \mu_0 \quad \textbf{[0.4 Ptos.]}$$

con $\mu_0 = 390$.

Bajo H_0 se tiene que

$$Z_0 = \frac{\overline{X}_n - \mu_0}{S/\sqrt{n}} \sim \text{t-Student}(n-1) \quad \textbf{[0.4 Ptos.]}$$

Reemplazando $\overline{X}_n = 385$, $S = 25$ y $n = 40$ se tiene que $T_0 = -1{,}264911$ **[0.4 Ptos.]**, lo que implica que:

$$\text{valor-p} \approx \Phi(-1{,}26) = 1 - \Phi(1{,}26) = 1 - 0{,}8961653 = 0{,}1038347 > 0{,}10 = \alpha \quad \textbf{[0.4 Ptos.]}$$

que es equivalente a

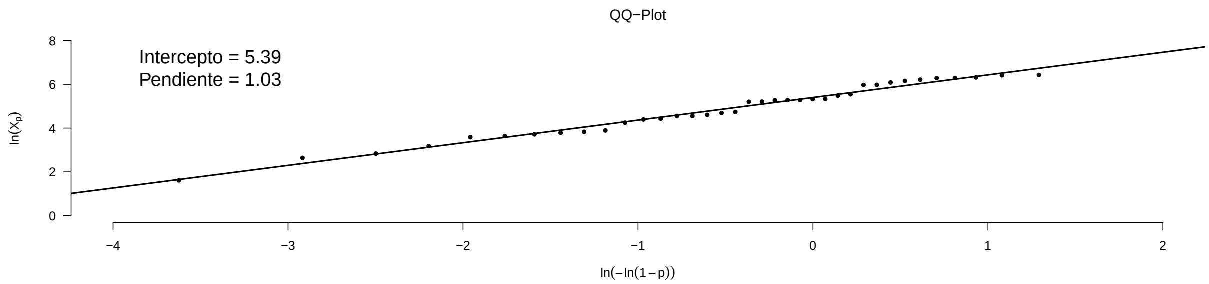
$$\textbf{[0.4 Ptos.]} \qquad T_0 = -1{,}264911 > -1{,}28 \approx k_{0,10} \qquad \textbf{[0.4 Ptos.]}$$

Por lo tanto no existe suficiente evidencia al 10 % de significancia para rechazar H_0 y apoyar la hipótesis que los superatletas demoran menos de 6 horas y media en el tramo de bicicleta. **[0.4 Ptos.]**

+ 1 Punto Base

Pregunta 3

Posterior al sismo de 6,2° que sacudió a la ciudad de Iquique el 1° de noviembre a las 19.20 hrs cuyo epicentro estuvo localizado a 37 km al sur de Camiña, se observó una seguidilla de sismos de menor intensidad. Los tiempos, en minutos, entre sismos se comportaron según una distribución Weibull(η, β) tal como se aprecia en el gráfico de probabilidad construido con los siguientes 37 sismos posteriores.



Asumiendo independencia entre los sismos, ¿cuál era la probabilidad que el tiempo mínimo entre sismos para los 37 registrados en la zona, post el sismo de 6,2°, fuese menor a 10 minutos?

Solución

Del formulario se deduce que los 37 tiempos entre sismos distribuyen Weibull($\eta = 219{,}2034$, $\beta = 0{,}9708738$). **[2.0 Ptos.]**

Alternativa 1: Reemplazando la función de densidad y de probabilidad acumulada Weibull en la expresión correspondiente a la densidad del mínimo se tienen que

$$f_{Y_1}(y) = \frac{\beta}{\eta\, n^{-1/\beta}} \left(\frac{y}{\eta\, n^{-1/\beta}}\right)^{\beta-1} \exp\left[-\left(\frac{y}{\eta\, n^{-1/\beta}}\right)^{\beta}\right] \to Y_1 \sim \text{Weibull}(\eta\, n^{-1/\beta},\, \beta) \qquad \textbf{[2.0 Ptos.]}$$

con n igual a 37. *(Nota: Si utiliza $n = 36$ no descontar puntaje)*

Se pide

$$\textbf{[1.0 Ptos.]} \qquad P(Y_1 < 10) = 1 - \exp\left[-\left(\frac{10}{\eta\, n^{-1/\beta}}\right)^{\beta}\right] = 0{,}8422511 \qquad \textbf{[1.0 Ptos.]}$$

Alternativa 2: Por definición

$$\textbf{[1.0 Ptos.]} \qquad F_{Y_1}(y) = 1 - [1 - F_X(y)]^n = 1 - \exp\left[-\left(\frac{10}{\eta}\right)^{\beta}\right]^n \qquad \textbf{[1.0 Ptos.]}$$

con n igual a 37. *(Nota: Si utiliza $n = 36$ no descontar puntaje)*

Se pide

$$\textbf{[1.0 Ptos.]} \qquad P(Y_1 < 10) = F_{Y_1}(10) = 0{,}8422511 \qquad \textbf{[1.0 Ptos.]}$$

Nota: Si el alumno aproxima β por 1 y trabaja con un modelo Exponencial($\nu = 1/\eta$) y aplica alternativa 1 o 2, NO descontar puntaje. En este caso la respuesta sería $P(Y_1 < 10) = 0{,}8150981$

+ 1 Punto Base

Pregunta 4

Nuevamente el fin de semana se produjeron "tacos" en diversas plazas de peajes. Usted, interesado en validar o refutar algunas afirmaciones, entrevista a más de un centenar de alumnos que sufrieron las consecuencias del regreso masivo. De los 124 entrevistados, 75 indican que los tacos fueron superiores a vividos anteriores en fines de semana largos. Por otra parte, 32 reportaron con precisión, el tiempo X (en minutos) que estuvieron esperando en las plazas de peaje:

```
 sum(X) sum(X^2) sum(Y) sum(Y^2)
315.156 3192.772 72.679  166.174
```

con $Y = \ln(X)$.

(a) **[3.0 Ptos]** Muestran los datos evidencia que permita afirmar que la mayoría considera que los tacos fueron superiores a los anteriores. Determine el valor-p y conclusión para un nivel de significancia del 1 %.

(b) **[3.0 Ptos]** ¿Qué puede decir frente a la afirmación: "La mediana de los tiempos fue menor a 10 min y 30 segundos"? Use un nivel de significancia del 10 % y comportamiento Log-Normal para los tiempos con un c.o.v del 20 % .

Solución

(a) Se pide contrastar

$$\text{H}_0 : p = p_0 \qquad \text{vs} \qquad \text{H}_a : p > p_0 \qquad \textbf{[0.5 Ptos.]}$$

con $p_0 = 0{,}5$.

Bajo H_0 se tiene que

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0\,(1-p_0)}{n}}} \overset{\text{aprox}}{\sim} \text{Normal}(0,1) \qquad \textbf{[0.5 Ptos.]}$$

Reemplazando $\hat{p} = \frac{75}{124} = 0{,}6048387$, $p_0 = 0{,}5$ y $n = 124$ se tiene que $Z_0 = 2{,}334869$. **[0.5 Ptos.]**

Luego

$$\text{valor-p} = 1 - \Phi(2{,}334869) \approx 1 - \Phi(2{,}34) = 1 - 0{,}9904 = 0{,}0096 < 0{,}01 = \alpha \qquad \textbf{[1.0 Ptos.]}$$

Por lo tanto existe suficiente evidencia al 1 % de significancia para rechazar H_0 y apoyar la hipótesis que la mayoría considera que los tacos fueron superiores a vividos anteriores en fines de semana largos. **[0.5 Ptos.]**

(b) Del enunciado tenemos que los tiempos distribuyen Log-Normal(λ, ζ), con $\zeta = \sqrt{\ln(1 + 0{,}2^2)} = 0{,}1980422$. **[0.5 Ptos.]**

Alternativa 1

Se pide un test sobre la mediana e^λ.

$$\text{H}_0 : e^\lambda = 10{,}5 \qquad \text{vs} \qquad \text{H}_a : e^\lambda < 10{,}5 \qquad \textbf{[0.5 Ptos.]}$$

Bajo H_0 se tiene que

$$Z_0 = \frac{e^{\hat{\lambda}} - 10{,}5}{\zeta\, 10{,}5/\sqrt{n}} \overset{\text{aprox}}{\sim} \text{Normal}(0,1) \qquad \textbf{[0.5 Ptos.]}$$

con $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i) = 2{,}271219$. **[0.5 Ptos.]**

Reemplazando se tiene que

[0.5 Ptos.] $Z_0 = -2{,}200222 \to \text{valor-p} \approx \Phi(-2{,}2) = 1-\Phi(2{,}2) = 1-0{,}9861 = 0{,}0139 < 0{,}10 = \alpha$ **[0.5 Ptos.]**

que es equivalente a

$$\textbf{[0.5 Ptos.]} \qquad Z_0 = -2{,}200222 < -1{,}28 \approx k_{0{,}10} \qquad \textbf{[0.5 Ptos.]}$$

Por lo tanto, existe suficiente evidencia para rechazar H_0 y apoyar que la mediana es inferior a 10 minutos y 30 segundos. **[0.5 Ptos.]**

Alternativa 2

Se pide un test sobre la mediana λ.

$$H_0 : \lambda = \ln(10{,}5) \qquad \text{vs} \qquad H_a : \lambda < \ln(10{,}5) \qquad \textbf{[0.5 Ptos.]}$$

Bajo H_0 se tiene que

$$Z_0 = \frac{\hat{\lambda} - \ln(10{,}5)}{\zeta/\sqrt{n}} \overset{\text{aprox}}{\sim} \text{Normal}(0, 1) \qquad \textbf{[0.5 Ptos.]}$$

con $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i) = 2{,}271219$. **[0.5 Ptos.]**

Reemplazando se tiene que

[0.5 Ptos.] $Z_0 = -2{,}289574 \to \text{valor-p} \approx \Phi(-2{,}29) = 1-\Phi(2{,}29) = 1-0{,}9890 = 0{,}011 < 0{,}10 = \alpha$ **[0.5 Ptos.]**

que es equivalente a

$$\textbf{[0.5 Ptos.]} \qquad Z_0 = -2{,}289574 < -1{,}28 \approx k_{0{,}10} \qquad \textbf{[0.5 Ptos.]}$$

Por lo tanto, existe suficiente evidencia para rechazar H_0 y apoyar que la mediana es inferior a 10 minutos y 30 segundos. **[0.5 Ptos.]**

+ 1 Punto Base