Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2021

Interrogación 3 - Pauta

Pregunta 1

Sea X el tiempo que le toma a Usted leer la prueba antes de comenzar e Y el tiempo que le tomará en responder esta pregunta.

Si $X \sim \text{Gamma}(k, \nu)$ e $Y \mid X = x \sim \text{Exponencial}(1/x)$, calcule la correlación entre estos tiempos si $k = 14$ y $\nu = 0.6$.

Solución

Se pide

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{\text{E}(X \cdot Y) - \mu_X \cdot \mu_Y}{\sigma_X \cdot \sigma_Y}$$

Del formulario se tiene que

$$\mu_X = \frac{k}{\nu} \quad \text{y} \quad \sigma_X = \frac{\sqrt{k}}{\nu}$$

Aplicando teorema de esperanzas iteradas (teorema de probabilidades totales) tenemos que

$$\mu_Y = \text{E}[\text{E}(Y \mid X)] = \text{E}(X) = \frac{k}{\nu}, \quad \textbf{[0.2 Ptos.]}$$

$$\sigma_Y^2 = \text{Var}[\text{E}(Y \mid X)] + \text{E}[\text{Var}(Y \mid X)] = \text{Var}(X) + \text{E}(X^2) = \frac{k}{\nu^2} + \frac{k\,(k+1)}{\nu^2} \quad \textbf{[0.2 Ptos.]}$$

y

$$\text{E}(X \cdot Y) = \text{E}[\text{E}(X \cdot Y \mid X)] = \text{E}(X^2) = \frac{k\,(k+1)}{\nu^2} \quad \textbf{[0.2 Ptos.]}$$

Reemplazando

$$\text{Corr}(X, Y) = \frac{1}{\sqrt{k+2}} = \frac{1}{\sqrt{14+2}} = \frac{1}{4} \quad \textbf{[0.4 Ptos.]}$$

Pregunta 2

El puntaje ponderado para ingresar a ingeniería, considerando únicamente la PTU está dado por:

Prueba de Comprensión Lectora: 17 %
Prueba de Matemática: 58 %
Prueba de Ciencias: 25 %

Un análisis del rendimiento de todos los postulantes a ING-UC determina que:

```
      C. Lectora   Matemáticas   Ciencia
mean       [mu1]         [mu2]      [mu3]
sd          [s1]          [s2]       [s3]

Correlación
             C. Lectora Matemáticas Ciencia
C. Lectora        1.00
Matemáticas       0.15        1.00
Ciencia           0.08        0.78    1.00
```

Asumiendo que cada prueba se comporta como una distribución normal, determine la probabilidad que el puntaje ponderado (considerando únicamente las componentes de la PTU) sea superior a 700 puntos.

Nota: En su respuesta considere todo el soporte de la distribución Normal. (No trunque en 850).

Solución

Definamos como X_1 al puntaje C. Lectora, X_2 al puntaje Matemáticas y X_3 al puntaje Ciencia.

Bajo Normalidad se tiene que

$$Z = 0.17 \cdot X_1 + 0.58 \cdot X_2 + 0.25 \cdot X_3 \sim \text{Normal}(\mu, \sigma)$$

con

$$\mu = 0.17 \cdot \texttt{[mu1]} + 0.58 \cdot \texttt{[mu2]} + 0.25 \cdot \texttt{[mu3]}$$

y

$$\sigma = \sqrt{0.17^2 \cdot \texttt{[s1]}^2 + 0.58^2 \cdot \texttt{[s2]}^2 + 0.25^2 \cdot \texttt{[s3]}^2 + 2 \cdot 0.17 \cdot 0.58 \cdot 0.15 \cdot \texttt{[s1]} \cdot \texttt{[s2]} + 2 \cdot 0.17 \cdot 0.25 \cdot 0.08 \cdot \texttt{[s1]} \cdot \texttt{[s3]} + 2 \cdot 0.25 \cdot 0.58 \cdot 0.78 \cdot \texttt{[s2]} \cdot \texttt{[s3]}}$$

Se pide

$$P(Z > 700) = \texttt{1 - pnorm(700, mean = mu, sd = sigma)} \quad \textbf{[1.0 Ptos.]}$$

```
## Ejemplo ##
a1 = 0.17
a2 = 0.58
a3 = 0.25
mu1 = 615
mu2 = 665
mu3 = 645
s1 = 58
s2 = 63
s3 = 32
mu = a1*mu1+a2*mu2+a3*mu3
sqrt(a1^2*s1^2+a2^2*s2^2+a3^2*s3^2+2*a1*a2*0.15*s1*s2+2*a1*a3*0.08*s1*s3+2*a2*a3*0.78*s2*s3)
1-pnorm(700, mu, sigma)
[1] 0.1433934
```

Nota: Si el alumnos calcula correctamente la probabilidad solicitada con σ bajo independencia [0.5 Ptos.]

Pregunta 3

Suponga que el número de clientes que llegan a un local comercial durante un día cualquiera tiene distribución Poisson con tasa esperada igual a `[lambda]` clientes por día.

Además cada cliente que llega realiza una compra con probabilidad `[p]`.

Calcule aproximadamente la probabilidad de que el número promedio de clientes que compran durante `[t]` días sea mayor a `[x]` personas.

Solución

Definamos como $X_1, \ldots, X_{\texttt{[t]}}$ el número de clientes que comprarán diariamente en `[t]` días.

$$X_i \overset{\text{iid}}{\sim} \text{Poisson}(\texttt{[lambda]} \cdot \texttt{[p]})$$

Aplicando Teorema del Limite Central y corrección por continuidad de se tiene que:

$$P\left(\overline{X} > \texttt{[x]}\right) = P\left(\sum_{i=1}^{\texttt{[t]}} X_i > \texttt{[t]} \cdot \texttt{x}\right) = 1 - P\left(\sum_{i=1}^{\texttt{[t]}} X_i \leq \texttt{[t]} \cdot \texttt{x}\right)$$

$\approx$ `1 - pnorm(x * t + 0.5, mean = lambda * p * t, sd = sqrt(t * lambda * p))` **[1.0 Ptos.]**

```
## Ejemplo ##
lambda = 100
p = 0.20
t = 30
x = 21
1 - pnorm(x * t + 0.5, mean = lambda * p * t, sd = sqrt(t * lambda * p))
[1] 0.106537
```

Nota: No descontar puntaje si el alumnos contesta correctamente no aplicando corrección por continuidad. Si el alumnos contesta el complemento asignar [0.5 Ptos].

Pregunta 4

Sea X una variable aleatoria con distribución Gamma(k, ν) y defina $Y = 1/X$.

Obtenga el valor esperado de Y aproximado de 2 do orden y su varianza aproximada de 1er orden.

Para $k =$ `[k]` y $\nu =$ `[nu]` evalué el coeficiente de variación aproximado:

Solución

La aproximación de 2orden de Y está dada por:

$$Y = g(X) \approx \frac{1}{\mu_X} + (X - \mu_X) \cdot \left(-\frac{1}{\mu_X^2}\right) + \frac{(X - \mu_X)^2}{2} \cdot \left(\frac{2}{\mu_X^3}\right).$$

Valor esperado de 2do orden:

$$\mu_Y \approx \frac{\nu}{k} + \frac{k}{\nu^2} \cdot \frac{\nu^3}{k^3} = \frac{\nu\,(k+1)}{k^2}. \quad \textbf{[0.3 Ptos.]}$$

Varianza de 1er orden:

$$\sigma_Y^2 \approx \frac{k}{\nu^2} \cdot \frac{\nu^4}{k^4} = \frac{\nu^2}{k^3}. \quad \textbf{[0.3 Ptos.]}$$

Se pide

$$\delta_Y \approx \frac{\sigma_Y}{\mu_Y} = \frac{\sqrt{k}}{(k+1)} \quad \textbf{[0.4 Ptos.]}$$

```
## Ejemplo ##
k = 100
nu = 0.6
sqrt(k)/(k+1)
[1] 0.0990099
```

Pregunta 5

Considere una muestra aleatoria $X_1, \ldots, X_n$ proveniente de una distribución Exponencial (ν). Determine al valor absoluto del sesgo del siguiente estimador de ν:

$$\hat{\nu} = \frac{1}{\overline{X}}$$

Evalúe para $n =$ `[n]` y $\nu =$ `[nu]`.

Solución

Dado que $X_1, \ldots, X_n$ iid $\text{Exp}(\nu)$, entonces $\overline{X} \sim \text{Gamma}(n,\, n \cdot \nu)$.

Se pide

$$\text{E}(\hat{\nu}) = \int_0^\infty \frac{1}{y} \cdot \frac{(n \cdot \nu)^n}{\Gamma(n)}\, y^{n-1}\, e^{-n \cdot \nu\, y} dy = \frac{n \cdot \nu}{n-1} \quad \textbf{[0.5 Ptos.]}$$

Por lo tanto

$$|\text{Sesgo}| = |\text{E}(\hat{\nu}) - \nu| = \frac{\nu}{n-1} \quad \textbf{[0.5 Ptos.]}$$

```
## Ejemplo ##
n = 20
nu = 0.3
nu/(n-1)
[1] 0.01578947
```

Suponga que X es una variable aleatoria discreta con función de probabilidad dada por:

$$p_X(x) = \begin{cases} \theta^2 & , \text{ si } x = -1 \\ 2\theta(1-\theta) & , \text{ si } x = 0 \\ (1-\theta)^2 & , \text{ si } x = +1 \end{cases}$$

con $0 < \theta < 1$.

A continuación responda las siguientes preguntas:

Pregunta 6.1

A partir de la siguiente muestra aleatoria: +1, +1 y 0, calcule y evalué el estimador de momentos de θ.

Solución

Tenemos que

$$\text{E}(X) = (-1) \cdot \theta^2 + 0 \cdot 2\,\theta\,(1-\theta) + (+1) \cdot (1-\theta)^2 = 1 - 2\theta = \overline{X} \to \hat{\theta} = \frac{1-\overline{X}}{2} = \frac{1}{6} \quad \textbf{[1.0 Ptos.]}$$

Pregunta 6.2

A partir de la siguiente muestra aleatoria: +1, -1 y -1, calcule y evalué el estimador de momentos de θ.

Solución

Tenemos que

$$\text{E}(X) = (-1) \cdot \theta^2 + 0 \cdot 2\,\theta\,(1-\theta) + (+1) \cdot (1-\theta)^2 = 1 - 2\theta = \overline{X} \to \hat{\theta} = \frac{1-\overline{X}}{2} = \frac{2}{3} \quad \textbf{[1.0 Ptos.]}$$

Pregunta 7.1

A partir de la siguiente muestra aleatoria: -1, +1 y -1, calcule y evalué el estimador máximo de verosimilitud de θ.

Solución

Tenemos que

$$\begin{aligned} L(\theta) = \theta^2 \cdot (1-\theta)^2 \cdot \theta^2 &\to \ln\, L(\theta) = 4\, \ln(\theta) + 2\, \ln(1-\theta) \\ &\to \frac{\partial}{\partial\theta} \ln\, L(\theta) = \frac{4}{\theta} - \frac{2}{1-\theta} \quad \textbf{[0.5 Ptos.]} \end{aligned}$$

Igualando a cero y despejando se tiene que

$$\hat{\theta} = \frac{2}{3} \quad \textbf{[0.5 Ptos.]}$$

Pregunta 7.2

A partir de la siguiente muestra aleatoria: 0, +1 y -1, calcule y evalué el estimador máximo de verosimilitud de θ.

Solución

Tenemos que

$$\begin{aligned} L(\theta) = \theta^2 \cdot 2\,\theta\,(1-\theta) \cdot (1-\theta)^2 &\to \ln\, L(\theta) = \ln(2) + 3\, \ln(\theta) + 3\, \ln(1-\theta) \\ &\to \frac{\partial}{\partial\theta} \ln\, L(\theta) = \frac{3}{\theta} - \frac{3}{1-\theta} \quad \textbf{[0.5 Ptos.]} \end{aligned}$$

Igualando a cero y despejando se tiene que

$$\hat{\theta} = \frac{1}{2} \quad \textbf{[0.5 Ptos.]}$$

Pregunta 8

Un fabricante de fibras textiles afirma que la nueva fibra para tapicería que están utilizando presenta una resistencia media superior a `[mu0]` kg, cuando estira.

Una muestra aleatoria de tamaño `[n]`, muestra una resistencia media de `[prom]` kg con una desviación estándar de `[s]` kg.

Realice la prueba de hipótesis correspondiente bajo el supuesto que los datos distribuyen Normal y entregue el valor-p

Solución

Tenemos las siguientes hipótesis

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

Bajo H_0

$$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim \text{t-Student}(n-1).$$

Para $\mathrm{H}_a : \mu > \mu_0$, el valor p se calcula como `1 - pt(T0, df = n-1)`. **[1.0 Ptos.]**

```
## Ejemplo ##
mu0 = 12
n = 20
prom = 12.5
s = 0.9
T0 = (prom-mu0)/(s/sqrt(n))
1-pt(T0, df = n-1)
[1] 0.01123042
```

Pregunta 9

En la elección presidencial de 1ra vuelta, se registraron los tiempos (en minutos) que les tomó a `[n]` votantes, escogidos al azar, para cumplir su deber cívico desde que entrega al presidente de mesa su CI y luego la retira.

```
mean(X) = [prom]
sd(X)   = [s]
```

Al analizar el error de estimación, con un 95 % de confianza, se determina que es demasiado grande y que en el próximo estudio debe ser reducido a la mitad.

Haciendo uso de la información disponible, ¿cuál es el tamaño muestral necesario?

Solución

Tenemos que el error de estimación de la muestra es

$$e.e = t_{0.975}(\texttt{[n]} - 1) \cdot \frac{\texttt{[s]}}{\sqrt{\texttt{[n]}}} \quad \textbf{[0.5 Ptos.]}$$

y se quiere en el próximo estudio reducir e.e a la mitad.

El nuevo tamaño muestral es:

$$m = \left(\frac{k_{0.975} \cdot \texttt{[s]}}{e.e/2} \right)^2 \quad \textbf{[0.5 Ptos.]}$$

```
## Ejemplo ##
n = 100
prom = 5.1
s = 1.1
ee = qt(0.975, df = n-1)*s/sqrt(n)
m = (qnorm(0.975)*s/(ee/2))^2
trunc(m)+1
[1] 391
```

Nota: Si el alumno calcula el error de estimación con qnorm() en vez de qt() no descontar puntaje.

La base de datos [Peliculas.xlsx] contiene la siguiente información de 7.661 películas: Nombre, Categoría, Género, Año, Fecha estreno, Puntuación, Votos, Director, Escritor, Protagonista, País, Presupuesto, Ganancia, Compañía y Duración.

A continuación responda las siguientes tres preguntas:

Pregunta 10

Ajuste una distribución Gamma y Log-Normal por el método de momentos a los tiempo de duración (en min) de las películas animadas (Genero Animación) y responda:

Log-Normal($\lambda =$ 4.5159139, $\zeta =$ 0.1272098) **[0.4 Ptos.]**

Gamma($k =$ 61.2971127, $\nu =$ 0.6647978) **[0.4 Ptos.]**

¿Cuál de estos dos modelos ajusta mejor la proporción de películas cuya duración es superior a las 2 horas?
Log-Normal **[0.2 Ptos.]**

```
Base <- rio::import("Peliculas.xlsx")
X <- dplyr::filter(Base, Género == "Animación")$"Duración"
mme_ln <- fitdistrplus::fitdist(data = X, distr = "lnorm", method = "mme")
mme_ln$estimate
  meanlog     sdlog
4.5159139 0.1272098
mme_gamma <- fitdistrplus::fitdist(data = X, distr = "gamma", method = "mme")
mme_gamma$estimate
     shape       rate
61.2971127  0.6647978
mean(X>120)
[1] 0.01775148
pgamma(120, shape = mme_gamma$estimate[1], rate = mme_gamma$estimate[2], lower.tail = F)
[1] 0.01393085
plnorm(120, meanlog = mme_ln$estimate[1], sdlog = mme_ln$estimate[2], lower.tail = F)
[1] 0.01638536
```

Pregunta 11.1

Existe evidencia para afirmar que los tiempos de duración medios de las películas de terror (Genero Terror) son menores a 97 min. Asumiendo Normalidad, realice la prueba de hipótesis correspondiente con un 5 % de significancia y responda:

Estadístico de Prueba = `-1.205007` **[0.4 Ptos.]**

Valor-p = `0.1145441` **[0.4 Ptos.]**

¿Existe evidencia? `NO` **[0.2 Ptos.]**

```
X <- dplyr::filter(Base, Género=="Terror")$"Duración"
mu0 = 97
t.test(x = X, mu = mu0, alternative = "less")$statistic
        t
-1.205007
t.test(x = X, mu = mu0, alternative = "less")$p.value
[1] 0.1145441
```

Pregunta 11.2

Existe evidencia para afirmar que los tiempos de duración medios de las películas de terror (Genero Terror) son superiores a 95 min. Asumiendo Normalidad, realice la prueba de hipótesis correspondiente con un 5 % de significancia y responda:

Estadístico de Prueba = `2.259388` **[0.4 Ptos.]**

Valor-p = `0.01226478` **[0.4 Ptos.]**

¿Existe evidencia? `SI` **[0.2 Ptos.]**

```
X <- dplyr::filter(Base, Género=="Terror")$"Duración"
mu0 = 95
t.test(x = X, mu = mu0, alternative = "greater")$statistic
       t
2.259388
t.test(x = X, mu = mu0, alternative = "greater")$p.value
[1] 0.01226478
```

Pregunta 12

¿Existe evidencia que permita afirmar que la desviación estándar de la puntuación de las películas de acción (Genero Acción) desde el año 2002 es inferior a 1 punto? Asumiendo que las puntuaciones provienen de una distribución Normal y un nivel de signifcancia del 2 %, realice la prueba de hipótesis correspondiente y responda:

Estadístico de Prueba = `1038.498` **[0.4 Ptos.]**

Valor-p = `0.04787016` **[0.4 Ptos.]**

¿Existe evidencia? `NO` **[0.2 Ptos.]**

```
X <- dplyr::filter(Base, Género=="Acción", Año>2001)$"Puntuación"
TeachingDemos::sigma.test(X, sigma = 1, alternative="less")$statistic
X-squared
 1038.498
TeachingDemos::sigma.test(X, sigma = 1, alternative="less")$p.value
[1] 0.04787016
```