Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Temporada Académica de Verano (enero 2023)

Curso	**:**	**Probabilidad y Estadística**
Sigla	**:**	**EYP1113**
Profesores	**:**	**Ricardo Aravena C., Ricardo Olea O. y Felipe Ossa M.**

PAUTA INTERROGACIÓN 3

Pregunta 1

Las variaciones diaria (en %) del precio del dólar durante el año 2022 y lo que llevamos del 2023, se encuentran entre $-\theta$ y $+\theta$, ver Figura 1.

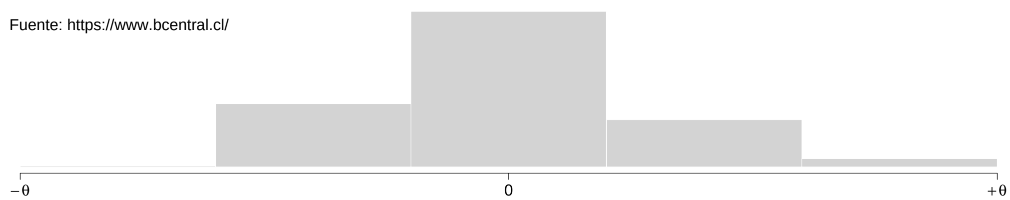

Figura 1: Variación diaria precio dólar 2022 - 2023.

Un resumen de las variaciones porcentuales se presenta a continuación:

```
  n  sum(Y) sum(Y^2)
262 7.45686 299.6222
```

Estime por el método de momentos el parámetro θ, bajo el supuesto que las variaciones diarias se comportan de manera independiente Beta$(2, 2)$.

Solución

Del enunciado se tiene que $X \sim$ Beta$(2, 2)$, con $\Theta_X = [-\theta, +\theta]$ y del formulario se deduce que

$$\text{E}(X) = 0 \quad \text{y} \quad \text{E}(X^2) = \frac{4^2\,\theta^2}{4^2 \cdot 5} \quad \textbf{[0.3 Ptos.]}$$

Como el 1er momento es cero (no depende de θ) se procede a igualar los 2dos momentos

$$\text{E}(X^2) = \frac{\theta^2}{5} = \frac{299.6222}{262} = \overline{X^2}. \quad \textbf{[0.3 Ptos.]}$$

Despejando

$$\hat{\theta} = \sqrt{\frac{5 \cdot 299.6222}{262}} = 2.39123 \quad \textbf{[0.4 Ptos.]}$$

Pregunta 2

La distribución de Pareto es una distribución de probabilidad continua con dos parámetros, que tiene aplicación en disciplinas como la sociología, la geofísica y la economía. Su función de densidad está dada por:

$$f_X(x) = \frac{k\,\lambda^k}{x^{k+1}},$$

con $x > \lambda$, $k > 0$ y $\lambda > 0$.

Se utilizó originalmente para describir la asignación de la riqueza entre individuos y en este caso el parámetro k es conocido popularmente como el índice de Pareto.

Si λ es conocido, determine a partir de una muestra aleatoria iid de tamaño n la distribución asintótica del estimador máximo verosímil $\hat{k}$ del parámetro k.

Para $n = 50$, $k = 2$ y $\lambda = 0.5$, calcule $P(\hat{k} > 2.5)$

Solución

La distribución asintótica del EMV de k está dada por

$$\hat{k} \overset{\text{aprox}}{\sim} \text{Normal}\left(k,\, \sqrt{\text{CCR}(k)}\right), \quad \textbf{[0.2 Ptos.]}$$

donde $\text{CCR}(k) = \dfrac{1}{\text{I}(k)}$ e $\text{I}(k) = -\text{E}\left(\frac{\partial^2}{\partial\, k^2} \ln\, L\right)$.

El logaritmo natural de la función de verosimilitud están dada por

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

Derivando parcialmente dos veces con respecto a k se tiene que

$$\textbf{[0.2 Ptos.]} \quad \frac{\partial^2}{\partial\, k^2} \ln\, L = -\frac{n}{k^2} \to \text{CCR}(k) = \frac{k^2}{n} \quad \textbf{[0.2 Ptos.]}$$

Para $n = 50$, $k = 2$ y $\lambda = 0.5$ se pide

$$P(\hat{k} > 2.5) = \texttt{1 - pnorm(2.5, mean = 2, sd = sqrt(4/50))} = 0.03854994 \quad \textbf{[0.2 Ptos.]}$$

Pregunta 3

La nota (N) de un alumno en el examen de un curso, puede modelarse como una combinación lineal entre su PPA (X), la nota alcanzada en la 1era interrogación del curso (Y) y el nivel de confianza adquirida por la asistencia a clases (Z), la cual se presenta a continuación:

$$N = 2 \cdot X - Y + 0.5 \cdot Z,$$

Suponga que X, Y y Z se comportan como variables aleatorias con distribución Normal, con las siguientes medidas:

```
        X   Y   Z     cor    X    Y    Z
mean  3.4 4.2 6.0       X 1.0  0.4  0.0
sd    0.5 1.0 2.0.      Y 0.4  1.0 -0.2
                        Z 0.0 -0.2  1.0
```

Determine la probabilidad que la nota en el examen sea más de un cinco.

Solución

Tenemos que

$$N = 2 \cdot X - Y + 0.5 \cdot Z \sim \text{Normal}(\mu,\, \sigma),$$

con

$$\begin{aligned} \mu &= 2 \cdot \mu_X - \mu_Y + 0.5 \cdot \mu_Z \\ &= 2 \cdot 3.4 - 4.2 + 0.5 \cdot 6.0 \\ &= 5.6 \quad \textbf{[0.3 Ptos.]} \end{aligned}$$

y

$$\begin{aligned} \sigma &= \sqrt{2^2 \cdot \sigma_X^2 + (-1)^2 \cdot \sigma_Y^2 + 0.5^2 \cdot \sigma_Z^2 + 2 \cdot 2 \cdot (-1) \cdot \sigma_X \cdot \sigma_Y \cdot \rho_{X,Y} + 2 \cdot (-1) \cdot 0.5 \cdot \sigma_Y \cdot \sigma_Z \cdot \rho_{Y,Z}} \\ &= \sqrt{2^2 \cdot 0.5^2 + 1^2 + 0.5^2 \cdot 2^2 + 2 \cdot 2 \cdot (-1) \cdot 0.5 \cdot 1 \cdot 0.4 + 2 \cdot (-1) \cdot 0.5 \cdot 1 \cdot 2 \cdot (-0.2)} \\ &= 1.612452 \quad \textbf{[0.4 Ptos.]} \end{aligned}$$

Se pide

$P(N > 5) =$ `1-pnorm(5, mean = 5.6, sd = 1.612452)` $= 0.6450924$ **[0.3 Ptos.]**

Pregunta 4

No es ningún secreto que las apuestas deportivas están en auge. De hecho, se han hecho tan populares que ahora hay más formas que nunca de apostar. Suponga que el próximo fin de semana 36 amigos apostaran de manera independiente y al azar por uno de los finalistas del Abierto de Australia. Si el vencedor paga dos veces la apuesta, ¿cuál es la probabilidad aproximada que entre todos sumen 80 dólares o más de ganancia, si cada uno apuesta 10 dólares?.

Solución

Definamos como X_i la ganancia del i-ésimo amigo y del enunciado se tiene que

$$p_X(-10) = p_X(+10) = 1/2 \to \mu_X = 0 \quad \text{y} \quad \sigma_X^2 = 100. \quad \textbf{[0.2 Ptos.]}$$

Se pide $P(Y \geq 80)$, con $Y = \sum_{i=1}^{36} X_i$.

Notemos que $\Theta_Y = \{-360, -340, \ldots - 20, 0, +20, +40, +60, +80, +100, \ldots, +360\}$. **[0.2 Ptos.]**

Aplicando Teorema Central del Límite y corrección por continuidad se tiene que

$$\begin{aligned}
P(Y \geq 80) &= 1 - P(Y < 80) \\
&= 1 - P(Y \leq 60) \quad \textbf{[0.2 Ptos.]} \\
&= 1 - P(Y < 70) \quad \textbf{[0.2 Ptos.]} \\
&\approx \texttt{1 - pnorm(70, mean = 0, sd = sqrt(36*100))} \\
&= 0.1216725 \quad \textbf{[0.2 Ptos.]}
\end{aligned}$$

Alternativamente se puede definir la ganancias como

$$Y = 20 \cdot Z - 360,$$

con $Z \sim$ Binomial$(n = 36, p = 1/2)$:

- Valor aproximado aplicando Teorema del Limite Central y corrección por continuidad

$$P(Y \geq 80) = P(Z \geq 22) = \texttt{1 - pnorm(21.5, mean = 36/2, sd = sqrt(36/4))} = 0.1216725 \quad \textbf{[0.6 Ptos.]}$$

- Valor exacto utilizando modelo Binomial

$$P(Y \geq 80) = P(Z \geq 22) = \texttt{1 - pbinom(21, size = 36, prob = 1/2)} = 0.1214925 \quad \textbf{[0.6 Ptos.]}$$

Observación: Si el alumno define que $\theta_X = \{-10, +20\}$, entonces $P(Y \geq 80) = P(Z \geq 15)$ y la respuesta bajo el modelo Binomial estaría dada por 0.8785075, mientras que la respuesta aproximada utilizando Teorema Central del Límite + corrección por continuidad sería 0.8783275. En estos casos no descontar puntaje

Pregunta 5

Un emprendimiento tiene dos puntos de atención y las ventas diarias en estos distribuyen Gamma$(k,\, \nu)$ y Log-Normal$(\lambda,\, \zeta)$ respectivamente. Dado que los puntos de atención se encuentran en lugares muy distantes, se puede suponer que las ventas se comportan de manera independientes entre estos puntos. Calcule el coeficiente de variación, aproximado de 1er orden, del porcentaje de ventas del emprendimiento correspondientes al primer punto durante un día cualquiera. Evalúe este coeficiente para $k = 300$, $\nu = 0.005$, $\lambda = 11$ y $\zeta = 0.30$.

Solución

Definamos como X a las ventas del 1er punto de atención e Y a las ventas del 2do.

$$X \sim \text{Gamma}(k,\, \nu) \quad \text{y} \quad Y \sim \text{Log-Normal}(\lambda,\, \zeta)$$

Para $Z = \dfrac{100 \cdot X}{X + Y}$ se pide $\delta_X = \dfrac{\sigma_Z}{\mu_Z}$ aproximado de 1er orden.

Del formulario se tiene que

$$\mu_Z \approx \frac{100 \cdot \mu_X}{\mu_X + \mu_Y} \quad \textbf{[0.3 Ptos.]}$$

y por la independencia descrita en el enunciado

$$\sigma_Z^2 \approx \sigma_X^2 \cdot \left(\frac{100 \cdot \mu_Y}{(\mu_X + \mu_Y)^2}\right)^2 + \sigma_Y^2 \cdot \left(-\frac{100 \cdot \mu_X}{(\mu_X + \mu_Y)^2}\right)^2, \quad \textbf{[0.5 Ptos.]}$$

con

$$\mu_X = \frac{k}{\nu}, \quad \mu_Y = e^{\lambda + \zeta^2/2}, \quad \sigma_X^2 = \frac{k}{\nu^2} \quad \text{y} \quad \sigma_Y^2 = \mu_Y^2 \left(e^{\zeta^2} - 1\right).$$

Reemplazando $k = 300$, $\nu = 0.005$, $\lambda = 11$ y $\zeta = 0.30$, se tiene que

$$\delta_Z \approx \frac{100 \cdot 0.07802962}{100 \cdot 0.4892766} = 0.1594796 \quad \textbf{[0.2 Ptos.]}$$

Pregunta 6

Suponga que las nuevas pruebas de competencias matemáticas, que son parte de la nueva PAES, M_1 y M_2, pueden ser modeladas por una distribución Normal-Bivariada.

El archivo `paes.xlsx` contiene resultados de alumnos de distintos establecimientos de la Capital, que además de los puntajes en esas pruebas incluye el género y año de egreso de los alumnos.

¿Cuál es el pronóstico para el puntaje de M_2 de alguien que obtuvo 700 puntos en la M_1?

Solución

Definamos como X al puntaje M_1 e Y al puntaje M_2.

Se pide

$$\mathsf{E}(Y \mid X = 700) = \mu_Y + (700 - \mu_X) \cdot \frac{\sigma_Y \cdot \rho}{\sigma_X} \quad \textbf{[0.5 Ptos.]}$$

```
Data <- rio::import("paes.xlsx")
X <- Data$M1
Y <- Data$M2

## Estimación de parámetros utilizando mean(), sd() y cor()
mu.x <- mean(X)
sigma.x <- sd(X)
mu.y <- mean(Y)
sigma.y <- sd(Y)
rho <- cor(X,Y)
mu.y + (700 - mu.x) * sigma.y * rho / sigma.x
[1] 683.6854

## Estimación de parámetros utilizando fitdist() y cor()
par.x <- fitdistrplus::fitdist(data = X, distr = "norm", method = "mle")$estimate
mu.x <- par.x[1]
sigma.x <- par.x[2]
par.y <- fitdistrplus::fitdist(data = Y, distr = "norm", method = "mle")$estimate
mu.y <- par.y[1]
sigma.y <- par.y[2]
rho <- cor(X,Y)
mu.y + (700 - mu.x) * sigma.y * rho / sigma.x
[1] 683.6854
```

[0.5 Ptos.]