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 2

Pregunta 1

Dentro de los hogares de una cierta ciudad, la empresa distribuidora de energía eléctrica reveló que un 20 % de las casas consume menos que 500 kWh mensuales, mientras que sólo un 15 % consume más de 1600 kWh mensuales.

Suponga que el consumo mensual de energía eléctrica en una casa cualquiera de esta ciudad tiene distribución Log-Normal. ¿Cuál es la probabilidad de que una casa cualquiera consuma menos de 1000 kWh mensuales?

Solución

Sea X el consumo de un hogar cualquiera, el cuál es una variable aleatoria Log-Normal$(\lambda,\, \zeta)$.

Tenemos que el percentil $p \times 100\,\%$ está dado por

$$\ln(x_p) = \lambda + \zeta \cdot \texttt{qnorm}(p).$$

A partir del enunciado tenemos el siguiente sistema de ecuaciones:

$$\begin{aligned}\ln(500) &= \lambda + \zeta \cdot \texttt{qnorm(0.20)},\\ \ln(1600) &= \lambda + \zeta \cdot \texttt{qnorm(0.85)}.\end{aligned}$$

Resolviendo tenemos que

[0.4 Ptos.] $$\zeta = \frac{\ln(500) - \ln(1600)}{\texttt{qnorm(0.20)} - \texttt{qnorm(0.85)}} = 0.6193381 \to \lambda = 6.735856$$ **[0.3 Ptos.]**

Se pide

$$P(X < 1000) = \texttt{plnorm(1000, meanlog = 6.735856, sdlog = 0.6193381)} = 0.6093222$$ **[0.3 Ptos.]**

En R:

```
x20 = 500
x85 = 1600
zeta = (log(x20)-log(1600))/(qnorm(0.20)-qnorm(0.85))
lambda = log(x20)-zeta*qnorm(0.20)
plnorm(1000, meanlog = lambda, sdlog = zeta)
[1] 0.6093222
```

Pregunta 2

Muy atrás quedaron los tiempos en los que, sin importar qué tan llena iba la micro, los pasajeros pasaban de mano en mano su pago al conductor del bus, que devolvía el boleto y lo hacía pasar por el mismo camino de manos hasta su dueño. Hoy, esa costumbre parece perdida y a usted le encargan determinar el porcentaje de evasión. Para esto le pide a cien fiscalizadores que reporten el número de fiscalizaciones hasta detectar al k-ésimo infractor, con $k \in \mathbb{N}$, valor escogido por otro encargado, por lo cual para efectos de este problema es un parámetro desconocido.

El resumen estadístico obtenido fue el siguiente:

```
  n  mean     sd
100  9.76  3.6017
```

Solución

Sea X el numero reportado por un fiscalizador cualquier, el cual según enunciado distribuye Binomial-Negativa$(k,\, p)$.

Del formulario se tiene que

$$\mu_X = \frac{k}{p} \quad \text{y} \quad \sigma_X^2 = \frac{k\,(1-p)}{p^2}$$

Despejando p en términos de μ_X y σ_X^2 se tiene que

$$\textbf{[0.5 Ptos.]} \quad p = \frac{\mu_X}{\mu_X + \sigma_X^2} = \frac{9.76}{9.76 + 3.6017^2} = 0.4293461 \quad \textbf{[0.5 Ptos.]}$$

Observación: En el caso que estime k y lo redondea a 4, el valor de p será igual a 0.4098361, si no redondea el valor de p será 0.4293461.

Pregunta 3

El archivo `sismos2023.xlsx` contiene información de los sismos registrados en Chile en lo que va del año:

- `ORD: Orden cronológico.`
- `PROF: Profundidad en km.`
- `MAG: Magnitud en escala Richter.`
- `MIN: Tiempo en minutos transcurrido desde el sismo anterior.`

¿Cuál es la probabilidad que en las próximas 48 horas ocurra a lo más un sismo de 4 grados o más en la escala de Richter? Suponga que los sismos ocurren según el mismo proceso de Poisson en el tiempo.

Solución

Definamos como X_t el número de sismos de 4 o más grados ocurridos en t minutos.

Del enunciado se tiene que

$$X_t \sim \text{Poisson}(\nu\, p\, t),$$

con ν el valor esperado de sismos en un minuto y p la probabilidad que un sismos tenga una intensidad de 4 grados o más en la escala de Richter.

A partir de los tiempos (en minutos) entre sismos, los cuales distribuyen Exponencial(ν) se tiene que

$$\nu = \frac{1}{\texttt{mean(MIN)}} = 0.01317521 \quad \textbf{[0.3 Ptos.]}$$

y a partir de las magnitudes obtenemos la frecuencia relativa de 4 o más grados:

$$p = \texttt{mean(MAG >= 4)} = 0.09473684 \quad \textbf{[0.3 Ptos.]}$$

Se pide

$$P(X_{48\cdot 60} \leq 1) = \texttt{ppois(1, lambda = 0.01317521 * 0.09473684 * 48 * 60)} = 0.1262063 \quad \textbf{[0.4 Ptos.]}$$

Alternativamente se puede definir como T_2 al tiempo, en minutos, transcurrido hasta el 2do sismo de 4 o más grados, el cual se comporta como una variable aleatoria Gamma$(2,\, \nu\, p)$.

En este caso

$$\begin{aligned} P(X_{48\cdot 60} \leq 1) &= P(T_2 > 48 \cdot 60) \\ &= \texttt{1 - pgamma(48*60, shape = 2, rate = 0.01317521 * 0.09473684)} \\ &= 0.1262063 \quad \textbf{[0.4 Ptos.]} \end{aligned}$$

Observación: Si el alumno responde $P(X_{48\cdot 60} = 0) = 0.02746749 = P(T_1 > 48 \cdot 60)$, no descontar puntaje.

En R:

```
Data <- rio::import("sismos2023.xlsx")
nu = 1/mean(Data$MIN)
p = mean(Data$MAG >= 4)
lambda <- nu*p
1-pgamma(48*60, shape = 2, rate = lambda)
[1] 0.1262063
ppois(1, lambda = lambda*48*60)
[1] 0.1262063
```

Pregunta 4

Sean X e Y dos variables aleatorias continuas y se dispone de la siguiente información:

$$f_{X\,|\,Y=y}(x) = \frac{1}{y},\ 0 < x < y \quad \text{y} \quad f_Y(y) = \frac{y}{8},\ 0 < y < 4.$$

Calcule la probabilidad marginal $P(X > 2)$.

Solución

Alternativa 1: Calcular con integral doble, primero con respecto a y y luego x.

$$\begin{aligned}
P(X > 2) &= \int_2^4 \int_x^4 f_{X|Y=y}(x) \cdot f_Y(y) dy dx = \int_2^4 \int_x^4 \frac{1}{y} \cdot \frac{y}{8} dy dx \quad \textbf{[0.3 Ptos.]} \\
&= \int_2^4 \int_x^4 \frac{1}{8} dy dx = \int_2^4 \frac{4-x}{8} dx \quad \textbf{[0.3 Ptos.]} \\
&= \frac{1}{2} \cdot (4-2) - \frac{4^2 - 2^2}{16} = 1 - \frac{16-4}{16} \quad \textbf{[0.3 Ptos.]} \\
&= \frac{1}{4} \quad \textbf{[0.1 Ptos.]}
\end{aligned}$$

Alternativa 2: Calcular con integral doble, primero con respecto a x y luego y.

$$\begin{aligned}
P(X > 2) &= \int_2^4 \int_2^y f_{X|Y=y}(x) \cdot f_Y(y) dx dy = \int_2^4 \int_2^y \frac{1}{y} \cdot \frac{y}{8} dx dy \quad \textbf{[0.3 Ptos.]} \\
&= \int_2^4 \int_2^y \frac{1}{8} dx dy = \int_2^4 \frac{y-2}{8} dy \quad \textbf{[0.3 Ptos.]} \\
&= \frac{4^2 - 2^2}{16} - \frac{1}{4} \cdot (4-2) = \frac{3}{4} - \frac{1}{2} \quad \textbf{[0.3 Ptos.]} \\
&= \frac{1}{4} \quad \textbf{[0.1 Ptos.]}
\end{aligned}$$

Alternativa 3: Encontrar marginal de X y luego integrar.

$$f_X(x) = \int_x^4 f_{X|Y=y}(x) \cdot f_Y(y) dy = \int_x^4 \frac{1}{y} \cdot \frac{y}{8} dy = \int_x^4 \frac{1}{8} dy = \frac{4-x}{8}, \quad 0 < x < 4 \quad \textbf{[0.5 Ptos.]}$$

Entonces

$$P(X > 2) = \int_2^4 f_X(x) dx = \int_2^4 \frac{4-x}{8} dx = \frac{1}{2} \cdot (4-2) - \frac{4^2 - 2^2}{16} = 1 - \frac{3}{4} = \frac{1}{4} \quad \textbf{[0.5 Ptos.]}$$

Pregunta 5

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.

Según este modelo, ¿cuál sería la probabilidad de obtener simultáneamente sobre los 700 puntos en M_1 y bajo los 750 en M_2?

Solución

En R:

```
Data <- rio::import("paes.xlsx")

## Matriz sigma utilizando cov()
sigma <- cov(Data[,c("M1","M2")])

## Matriz sigma utilizando sd() y cor()
sigma.x <- sd(Data[,c("M1")])
sigma.y <- sd(Data[,c("M2")])
rho.xy  <- cor(Data[,c("M1")],Data[,c("M2")])
sigma = matrix(c(sigma.x^2,rho.xy*sigma.x*sigma.y,rho.xy*sigma.x*sigma.y, sigma.y^2),
ncol = 2, byrow = T)

sigma
         [,1]     [,2]
[1,] 18505.66 18853.72
[2,] 18853.72 28217.86
```

[0.4 Ptos.]

```
## Vector de medias utilizando apply()
mu <- apply(Data[,c("M1","M2")],2,mean)

## Vector de medias
mu <- c(mean(Data[,c("M1")]), mean(Data[,c("M2")]))

mu
      M1       M2
614.4017 596.4771
```

[0.2 Ptos.]

```
## Probabilidad solicitada
mvtnorm::pmvnorm(lower = c(700,-Inf), upper = c(+Inf, 750), mean = mu, sigma = sigma)[1]
[1] 0.1202166
```

[0.4 Ptos.]

Observación: Si el alumno considera 0 y 1000 como límites extremos, no descontar puntaje.

Pregunta 6

El período de incubación T de la nueva cepa de Covid19 (tiempo entre el momento de contraer el virus y el comienzo de los síntomas) puede ser modelada de acuerdo a una distribución Log-Logística de parámetros μ y σ.

En cambio, el tiempo de duración de los síntomas depende del período de incubación, digamos t, se comporta como una distribución Gamma$(k,\, 1/t)$.

Determine la covarianza entre el período de incubación y la duración de los síntomas. Evalúe para $\mu = 3$, $\sigma = 1/4$ y $k = 3$.

Solución

Definamos T al tiempo de incubación e Y al tiempo de duración de síntomas.

Del enunciado se tiene que

$$T \sim \text{Log-Logística}(\mu,\, \sigma) \quad \text{e} \quad Y \,|\, T = t \sim \text{Gamma}(k,\, 1/t)$$

Se piden Cov$(T,\, Y) = $ E$(T \cdot Y) - $ E$(T) \cdot $ E(Y).

A partir del formulario y aplicando esperanza iterada se tiene que

$$\text{E}(T) = e^{\mu} \cdot \Gamma(1+\sigma) \cdot \Gamma(1-\sigma) \quad \textbf{[0.2 Ptos.]}$$
$$\text{E}(Y) = \text{E}[\text{E}(Y \,|\, T)] = \text{E}(k\,T) = k \cdot e^{\mu} \cdot \Gamma(1+\sigma) \cdot \Gamma(1-\sigma) \quad \textbf{[0.3 Ptos.]}$$

Por otra parte,

$$\text{E}(T \cdot Y) = \text{E}[\text{E}(T \cdot Y \,|\, T)] = \text{E}[T \cdot \text{E}(Y \,|\, T)] = \text{E}(k \cdot T^2) = k \cdot \text{E}(T^2) = k \cdot e^{2\mu} \cdot \Gamma(1+2\sigma) \cdot \Gamma(1-2\sigma) \quad \textbf{[0.3 Ptos.]}$$

Por tanto,

$$\text{Cov}(T,\, Y) = k \cdot e^{2\mu} \cdot \Gamma(1+2\sigma) \cdot \Gamma(1-2\sigma) - k \cdot e^{2\mu} \cdot \Gamma^2(1+\sigma) \cdot \Gamma^2(1-\sigma).$$

Reemplazando para $\mu = 3$, $\sigma = 1/4$ y $k = 3$ se tiene Cov$(T,\, Y) = 407.9824$. **[0.2 Ptos.]**

En R:

```
mu <- 3
sigma <- 1/4
k <- 3
k*exp(2*mu)*gamma(1+2*sigma)*gamma(1-2*sigma)-k*exp(2*mu)*gamma(1+sigma)^2*gamma(1-sigma)^2
[1] 407.9824
```