Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2021

Interrogación 1 - Pauta

Pregunta 1:

Sean A, B y C, eventos contenidos en el espacio muestral. Responda V (verdadero) o F (falso) a las siguientes afirmaciones:

(a) Si $A \cap B = \phi$ entonces $\bar{A}$ y $\bar{B}$ son independientes.

(b) Si A y B son independientes, entonces $P(A \cap B) \leq P(A)$.

Solución

(a) La afirmación es FALSA **[0.50 Ptos]**, ya que los tres eventos están contenidos en el espacio muestra y al tener intersección vacía, entonces A y B son disjuntos.

(b) La afirmación es VERDADERA **[0.50 Ptos]**, no solo para A y B independientes, si no para cualquier par de eventos

$$P(A) = P(A \cap S) = P([A \cap B] \cup [A \cap \overline{B}]) = P(A \cap B) + P(A \cap \overline{B}) \geq P(A \cap B)$$

por axiomas 3 y 1.

Pregunta 2:

Suponga que durante la prueba de hoy, `[m]` alumnos tendrán algún problema técnico, de salud o fuerza mayor. Si en el link 2 hay 150 alumnos conectados y los que se presentaron hoy a rendir la prueba fueron 447, cuál es la probabilidad que al menos cuatro de estos alumnos se encuentre en esta sala.

Solución

Definamos el evento A: "al menos cuatro alumnos se encuentra en esta sala".

Como los alumnos escogían una sala al azar, tenemos que $\# S = \binom{447}{150}$. **[0.40 Ptos]**

Se pide $P(A) = 1 - P(\overline{A})$.

Los casos favorables del complemento de A son:

$$\# \overline{A} = \binom{\mathtt{m}}{0}\binom{447-\mathtt{m}}{150} + \binom{\mathtt{m}}{1}\binom{447-\mathtt{m}}{149} + \binom{\mathtt{m}}{2}\binom{447-\mathtt{m}}{148} + \binom{\mathtt{m}}{3}\binom{447-\mathtt{m}}{147} \quad \textbf{[0.50 Ptos]}$$

Luego

$$P(A) = 1 - \frac{\# \overline{A}}{\# S} \quad \textbf{[0.10 Ptos]}$$

Por ejemplo en R:

```
m = 10
1-sum(choose(m,0:3)*choose(447-m, 150-0:3))/choose(447,150)
[1] 0.4469048

m = 20
1-sum(choose(m,0:3)*choose(447-m, 150-0:3))/choose(447,150)
[1] 0.9459579
```

Pregunta 3:

Mucho se ha hablado sobre la efectividad del proceso vacunatorio. En particular, se ha dado ha conocer la fuerte caída en las hospitalizaciones de las personas con las dos dosis.

Usted tiene acceso a datos y determina que el 20 % de las personas, vacunadas con Sinovac, que contraen Covid19 requieren de hospitalización. Mientras que solo el 5 % de las personas, vacunadas con Pfizer, que contraen Covid19 requieren de hospitalización.

Por otra parte, análisis muestran que el `[B1]` % de las personas vacunadas con Sinovac contraen Covid19, mientras que las personas vacunadas con Pfizer solo el `[B2]` % contrae Covid19. De las personas vacunadas con dos dosis, el `[A]` % de las personas han sido vacunadas con Pfizer.

Una persona que ha recibido sus dos dosis, ¿cuál es la probabilidad que no requiera hospitalización por Covid19?

Solución

Definamos los siguientes eventos:

A: Persona vacunada con dos dosis Pfizer.
B: Persona vacunada con dos dosis contrae Covid.
C: Persona vacunada con dos dosis que contrae Covid requiere hospitalización.

Del enunciado se tiene que:

$$P(A) = \frac{\texttt{[A]}}{100}, \quad P(B \,|\, \overline{A}) = \frac{\texttt{[B1]}}{100}, \quad P(B \,|\, A) = \frac{\texttt{[B2]}}{100}$$

$$P(C \,|\, A \cap B) = 0.05 \quad \text{y} \quad P(C \,|\, \overline{A} \cap B) = 0.20$$

Por probabilidades totales se tiene que

$$P(C) = P(A \cap B \cap C) + P(\overline{A} \cap B \cap C) \quad \textbf{[0.20 Ptos]}$$

$$P(\overline{C}) = 1 - P(C \,|\, A \cap B) \cdot P(B \,|\, A) \cdot P(A) - P(C \,|\, \overline{A} \cap B) \cdot P(B \,|\, \overline{A}) \cdot P(\overline{A}) \quad \textbf{[0.30 Ptos]}$$

$$P(\overline{C}) = 1 - 0.05 \cdot \frac{\texttt{[B2]}}{100} \cdot \frac{\texttt{[A]}}{100} - 0.20 \cdot \frac{\texttt{[B1]}}{100} \cdot \left(1 - \frac{\texttt{[A]}}{100}\right) \quad \textbf{[0.50 Ptos]}$$

Por ejemplo en R:

```
B1 = 65
B2 = 15
A  = 20
1-(0.05*(B2/100)*(A/100) + 0.20*(B1/100)*(1-A/100))
[1]  0.8945
```

Pregunta 4:

Diversos rumores en el congreso nacional apuntan a una nueva posible colusión del gas debido a las fuertes alzar en el periodo de pandemia.

La encuesta presupuestaria indica que de las personas que consume gas licuado, el `[A1]` % prefiere el cilindro de 11 kilos, un `[A2]` % utiliza el de 15 kilos, mientras que un `[A3]` % utiliza el de 45 kilos y el resto compra el de 5 kilos.

Al analizar los precios, un proveedor grande de Chile indica que en el último mes los formatos de 11 y 15 kilos han aumentado sus costos en el `[B1]` % y `[B2]` % de los distribuidores asociados, respectivamente, mientras que en los de 5 kilos el `[B4]` % de los distribuidores no subió sus precios, y finalmente en 45 kilos todas las distribuidoras mantuvieron los precios.

¿Cuál es la probabilidad que un cliente que se vio afectado con el alza de precios, sea un consumidor habitual de 15 kilos?

Solución

Definamos los siguientes eventos:

A_1: Cliente usa cilindro 11 kilos.
A_2: Cliente usa cilindro 15 kilos.
A_3: Cliente usa cilindro 45 kilos.
A_4: Cliente usa cilindro 5 kilos.
B: Cliente afectado por alza de precio.

Se pide

$$P(A_2 \mid B) = \frac{P(B \mid A_2) \cdot P(A_2)}{P(B \mid A_1) \cdot P(A_1) + P(B \mid A_2) \cdot P(A_2) + P(B \mid A_3) \cdot P(A_3) + P(B \mid A_4) \cdot P(A_4)} \quad \textbf{[0.50 Ptos]}$$

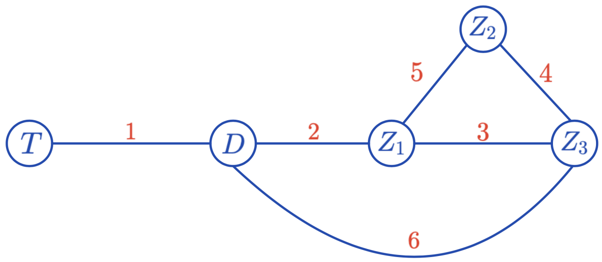
$$= \frac{(\texttt{[B2]}/100) \cdot (\texttt{[A2]}/100)}{(\texttt{[B1]}/100) \cdot (\texttt{[A1]}/100) + (\texttt{[B2]}/100) \cdot (\texttt{[A2]}/100) + 0 \cdot (\texttt{[A3]}/100) + (1 - \texttt{[B4]}/100) \cdot (1 - \texttt{[A1]}/100 - \texttt{[A2]}/100 - \texttt{[A3]}/100)} \quad \textbf{[0.50 Ptos]}$$

Por ejemplo en R:

```
A1 = 40/100
A2 = 35/100
A3 = 20/100
A4 = 1 - A1 - A2 - A3
B1 = 20/100
B2 = 30/100
B3 =  0/100
B4 = 85/100
B2*A2/(B1*A1+B2*A2+B3*A3+(1-B4)*A4)
[1] 0.5454545
```

Pregunta 5:

Una empresa de transmisión eléctrica T entrega energía a una empresa distribuidora D, la cual distribuye a través de sus redes a tres zonas (Z_1, Z_2, Z_3). Ultimamente, el robo de cable de la red de distribución se ha incrementado y cuando se producen estos robos, provocan apagones en algunas zonas.



Suponga que se ha estimado la probabilidad de robo de cable por mes (y por tanto, suspensión de la transmisión de energia) en cada una de las 6 lineas de distribución, cuyos valores son respectivamente [p1], [p2], [p3], [p4], [p5] y [p6]].

Determine la probabilidad que la zona Z_3 sufra al menos un apagón este mes.

Solución

Definamos como E_1 al evento "NO hay robo en línea i-ésima" y A al evento NO hay "apagón en Z_3".

$$A = (E_1 \cap E_2 \cap E_3) \cup (E_1 \cap E_2 \cap E_5 \cap E_4) \cup (E_1 \cap E_6) \quad \textbf{[0.30 Ptos]}$$

Aplicando probabilidad, ley aditiva y ley distributiva

$$\begin{aligned} P(A) = {} & P(E_1 \cap E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_5 \cap E_4) + P(E_1 \cap E_6) - P(E_1 \cap E_2 \cap E_3 \cap E_4 \cap E_5) \\ & - P(E_1 \cap E_2 \cap E_3 \cap E_6) - P(E_1 \cap E_2 \cap E_4 \cap E_5 \cap E_6) + P(E_1 \cap E_2 \cap E_3 \cap E_4 \cap E_5 \cap E6) \quad \textbf{[0.40 Ptos]} \end{aligned}$$

Suponiendo independencia entre robos:

$$\begin{aligned} P(A) = {} & (1 - \texttt{[p1]}) \cdot (1 - \texttt{[p2]}) \cdot (1 - \texttt{[p3]}) \\ & + (1 - \texttt{[p1]}) \cdot (1 - \texttt{[p2]}) \cdot (1 - \texttt{[p4]}) \cdot (1 - \texttt{[5]}) \\ & + (1 - \texttt{[p1]}) \cdot (1 - \texttt{[p6]}) \\ & - (1 - \texttt{[p1]}) \cdot (1 - \texttt{[p2]}) \cdot (1 - \texttt{[p3]}) \cdot (1 - \texttt{[p4]}) \cdot (1 - \texttt{[p5]}) \\ & - (1 - \texttt{[p1]}) \cdot (1 - \texttt{[p2]}) \cdot (1 - \texttt{[p3]}) \cdot (1 - \texttt{[p6]}) \\ & - (1 - \texttt{[p1]}) \cdot (1 - \texttt{[p2]}) \cdot (1 - \texttt{[p4]}) \cdot (1 - \texttt{[p5]}) \cdot (1 - \texttt{[p6]}) \\ & + (1 - \texttt{[p1]}) \cdot (1 - \texttt{[p2]}) \cdot (1 - \texttt{[p3]}) \cdot (1 - \texttt{[p4]}) \cdot (1 - \texttt{[p5]}) \cdot (1 - \texttt{[p6]}) \quad \textbf{[0.20 Ptos]} \end{aligned}$$

Se pide $P(\overline{A}) = 1 - P(A)$. **[0.10 Ptos]**

Por ejemplo en R:

```
p1 = 0.7; q1 = 1-p1
p2 = 0.6; q2 = 1-p2
p3 = 0.6; q3 = 1-p3
p4 = 0.6; q4 = 1-p4
p5 = 0.6; q5 = 1-p5
p6 = 0.7; q6 = 1-p6
1-(q1*q2*q3 + q1*q2*q4*q5 + q1*q6 - q1*q2*q3*q4*q5 - q1*q2*q3*q6 - q1*q2*q4*q5*q6 + q1*q2*q3*q4*q5*q6)
[1] 0.868336
```

Pregunta 6:

En la formulación de modelos para determinar indicadores de pobreza es comúnmente utilizada la siguiente función de densidad:

$$f_X(x) = \frac{\beta}{x^2} \cdot e^{-\beta/x}$$

con $x > 0$ y $\beta > 0$.

Para esta distribución determine la moda y evalúe en $\beta =$ `[beta]`.

Solución

Tenemos que

$$f_X'(x) = \beta\left[-2\,x^{-3} \cdot e^{-\beta/x} + x^{-2} \cdot e^{-\beta/x} \cdot \left(-\beta \cdot x^{-2}\right)\right] = -\frac{\beta\, e^{-\beta/2}}{x^3}\left(2 - \frac{\beta}{x}\right) \quad \textbf{[0.50 Ptos]}$$

Igualando f' a cero se tiene que la Moda es igual a $\dfrac{\beta}{2}$. **[0.50 Ptos]**

Pregunta 7:

El número de alumnos afectados por corte en el sumistro eléctrico durante la prueba se puede modelar con una variable aleatoria que tiene la siguiente función generadora de momentos

$$M_X(t) = \left(\frac{\beta}{\beta + (1 - e^t)}\right)^{\alpha}$$

con $t \in \mathbb{R}$.

Determine el coeficiente de asimetría y evalué para $\alpha =$ `[alpha]` y $\beta =$ `[beta]`.

Solución

Tenemos que

$$M_X^{(k)}(0) = \mathrm{E}(X^k)$$

A partir de esto se tiene que

$$\textbf{[0.30 Ptos]} \quad \mu_X = \mathrm{E}(X) = \frac{\alpha}{\beta}, \quad \sigma_X^2 = \mathrm{E}(X^2) - \mu_X^2 = \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2} \quad \textbf{[0.30 Ptos]}$$

$$\theta_X = \frac{\mathrm{E}(X^3) - 3\,\mathrm{E}(X^2)\,\mu_X + 2\,\mu_X^3}{\sigma_X^3} = \frac{1 + \frac{2}{\beta}}{\sqrt{\frac{\alpha}{\beta}\left(1 + \frac{1}{\beta}\right)}} \quad \textbf{[0.40 Ptos]}$$

Pregunta 8:

La resistencia del hormigón puede ser modelada por una distribución Normal. La normativa indica que un producto debe cumplir con el requerimiento técnico de resistencia superior a `[xp]`MPa.

Si el producto que proveniente de una empresa presenta una resistencia media de 24MPa y la probabilidad de ser rechazado es de un `[p]` %, ¿cuál es el coeficiente de variación (c.o.v.) del producto con respecto a su resistencia?

Solución

Sea X la resistencia del hormigón que distribuye Normal(μ, σ).

Del enunciado tenemos que

$$\textbf{[0.30 Ptos]} \quad \mu = 24 \quad \text{y} \quad \sigma = \frac{x_p - \mu}{\Phi^{-1}(p)} \quad \textbf{[0.30 Ptos]}$$

Se pide $\delta_X = \dfrac{\sigma}{\mu}$. **[0.40 Ptos]**

Por ejemplo en R:

```
mu = 24
p  = 0.11
xp = 20
sigma = (xp-mu)/qnorm(p)
sigma/mu
[1] 0.1358849
```

Pregunta 9:

La tensión (o voltaje) en un sistema de transmisión puede ser considerada como una variable aleatoria Log-Normal, tal que la probabilidad de entregar un voltaje superior a `[x]`150 voltios es de `[q]`0.85, con un coeficiente de variación del `[delta]`25 %.

Las vivienda que reciben un voltaje sobre 250 voltios o bajo 130 voltios, quedan expuestas a que sus artefactos eléctricos presenten problemas. Si desde la central de la empresa distribuidora un encargado monitorea 4 viviendas al azar, ¿cuál es la probabilidad que los artefactos eléctricos de las 4 casas queden expuestos a problemas?

Solución

Sea X el voltaje que recibe una vivienda, el cuál distribuye Log-Normal(λ, ζ).

Del enunciado se tiene que

$$\textbf{[0.30 Ptos]} \quad \zeta = \sqrt{\ln[1 + (\texttt{delta}/100)^2]} \quad \text{y} \quad \lambda = \ln(\texttt{x}) - \zeta \cdot \Phi^{-1}(1 - \texttt{q}/100) \quad \textbf{[0.30 Ptos]}$$

Se pide p^4 **[0.20 Ptos]**, con $p = P(\{X < 130\} \cup \{X > 250\})$. **[0.20 Ptos]**

Por ejemplo en R:

```
x = 150
q = 85
delta = 25
zeta = sqrt(log(1+(delta/100)^2))
lambda = log(x)-zeta*qnorm(1-q/100)
p = plnorm(130, lambda, zeta)+(1-plnorm(250, lambda, zeta))
p^4
0.001679948
```

Para las empresas distribuidoras de Gas es clave poder relacionar la demanda (venta en kg) con factores externos que permitan predecir su comportamiento, y así mejorar la logística con respecto al stock que deben manejar.

En esta parte utilizaremos los datos que se encuentra en el archivo `Demanda Gas.xlsx`, el cuál contiene ventas y precio diario por kg de gas en una empresa distribuidora. También dispone de información meteorológicas como min, max y promedio de la temperatura que se registraron diariamente, el porcentaje de nubosidad y también algunas variables calendario, por ejemplo, semana, mes y año.

La 1ra hoja contiene la data que van a utilizar y en la 2da viene un mayor detalle que describe las variables disponibles.

A continuación responda las siguientes tres preguntas.

```
library(rio)
Base = import("Demanda_Gas.xlsx")
head(Base)
```

Pregunta 10.1:

Para la estación del año "spring" considere solo las temperaturas mínimas diarias registradas en las semanas 4 y 5 de un mes.

(a) ¿Cuál fue el porcentaje de días en que la temperatura mínima fue superior a 14 grados?

(b) Utilizando la mediana y el percentil 80 %, ajuste a las mínimas temperaturas registradas una distribución Log-Normal(λ, ζ).

(c) Utilizando la media y la desviación estándar, ajuste ahora una distribución Gamma(k, ν).

(d) ¿Cuál de los dos modelos ajusta mejor el resultado observado en (a)? (Responda Log-Normal o Gamma)

(e) Ingrese el valor calculado en (d) que respalda su respuesta anterior.

Solución

(a) 82.58 %. **[0.20 Ptos]**

(b) $\lambda = 2.778819$ y $\zeta = 0.1272522$. **[0.20 Ptos]**

(c) $\mu = 15.97052$ y $\sigma = 1.915335$. **[0.20 Ptos]**

(d) Normal. **[0.20 Ptos]**

(e) 0.8482155. **[0.20 Ptos]**

En R:

```
X = subset(Base, SEMANA_MES >= 4 & ESTACION ==  "spring")$MIN

lambda = log(median(X))
zeta = (log(quantile(X, prob = 0.80))-lambda)/qnorm(0.80)

cbind(lambda, zeta)
  lambda      zeta
2.778819 0.1272522

mu = mean(X)
sigma = sd(X)

cbind(mu, sigma)
      mu    sigma
15.97052 1.915335

M = cbind(mean(X>14),
1-plnorm(14, meanlog = lambda, sdlog = zeta),
1-pnorm(14, mean = mu, sd = sigma))
colnames(M) = c("Empírico", "Log-Normal", "Normal")
rownames(M) = c("P(X > 14)")
M
           Empírico Log-Normal    Normal
P(X > 14) 0.8258065  0.8639647 0.8482155

## Mejor ajuste Normal
```

Pregunta 10.2:

Para la estación del año "summer" considere solo demandas diarias registradas en las semanas 4 y 5 de un mes.

(a) ¿Cuál fue el porcentaje de días en que la demanda diaria fue superior a 15 mil kg?

(b) Utilizando la media y la desviación estándar, ajuste a la demanda una distribución Log-Normal.

(c) Utilizando la mediana y el percentil 80 %, ajuste ahora una distribución Normal.

(d) ¿Cuál de los dos modelos ajusta mejor el resultado observado en (a)? (Responda Normal o Log-Normal)

(e) Ingrese el valor calculado en (d) que respalda su respuesta anterior.

Solución

(a) 13.77 %. **[0.20 Ptos]**

(b) $\lambda = 9.095432$ y $\zeta = 0.4957532$. **[0.20 Ptos]**

(c) $\mu = 9375$ y $\sigma = 4648.172$. **[0.20 Ptos]**

(d) Log-Normal. **[0.20 Ptos]**

(e) 0.1469366. **[0.20 Ptos]**

En R:

```
X = subset(Base, SEMANA_MES >= 4 & ESTACION ==  "summer")$DEMANDA

mu.x = mean(X)
sigma.x = sd(X)

delta.x = sigma.x/mu.x
zeta = sqrt(log(1+delta.x^2))
lambda = log(mu.x)-0.5*zeta^2

cbind(lambda,zeta)
  lambda      zeta
9.095432 0.4957532

mu = median(X)
sigma = (quantile(X, prob = 0.80)-mu)/qnorm(0.80)

cbind(mu, sigma)
  mu    sigma
9375 4648.172

M = cbind(mean(X>15000),
1-plnorm(15000, meanlog = lambda, sdlog = zeta),
1-pnorm(15000, mean = mu, sd = sigma))
colnames(M) = c("Empírico", "Log-Normal", "Normal")
rownames(M) = c("P(X > 15000)")
M
              Empírico Log-Normal    Normal
P(X > 15000) 0.1377246  0.1469366 0.1131101

## Mejor ajuste Log-Normal
```

Pregunta 11.1:

Considere los promedios mensuales de los precios diarios por kg de gas y responda:

IQR = 156.36 **[0.20 Ptos]**
Mediana = 1140.74 **[0.20 Ptos]**
Coeficiente de variación = 0.10 **[0.20 Ptos]**
Moda = 1136.27 (Utilice el método que por defecto aplica la función) **[0.20 Ptos]**
Percentil 65 % = 1204.56 **[0.20 Ptos]**

Solución

En R:

```
X = c(tapply(X = Base$PRECIO, INDEX = list(Base$MES,Base$AÑO), mean))
X = c(tapply(X = Base$PRECIO, INDEX = list(Base$MES+100*Base$AÑO), mean))
X = aggregate(PRECIO ~ MES*AÑO, data = Base, FUN = mean)$PRECIO

Tabla = round(rbind(IQR(X, na.rm = T),median(X, na.rm = T),sd(X, na.rm = T)/mean(X, na.rm = T),
modeest::mlv(X, na.rm = T),quantile(X, prob = 0.65, na.rm = T)),2)
rownames(Tabla) = c("IQR", "Mediana", "c.o.v.", "Moda", "Percentil 65%")
colnames(Tabla) = "PRECIO"
Tabla
               PRECIO
IQR            156.36
Mediana       1140.74
c.o.v.           0.10
Moda          1136.27
Percentil 65% 1204.56
```

Nota: Si el alumno obtiene las medidas correctamente con toda la data o con los 12 promedio mensuales, no descontar puntaje.

Pregunta 11.2:

Considere los promedios mensuales de los kilos demandados diariamente de gas y responda:

IQR = 1924.25 **[0.20 Ptos]**
Media = 10050.68 **[0.20 Ptos]**
Coeficiente de variación = 0.14 **[0.20 Ptos]**
Moda = 10366.54 (Utilice el método que por defecto aplica la función) **[0.20 Ptos]**
Percentil 56 % = 10293.46 **[0.20 Ptos]**

Solución

En R:

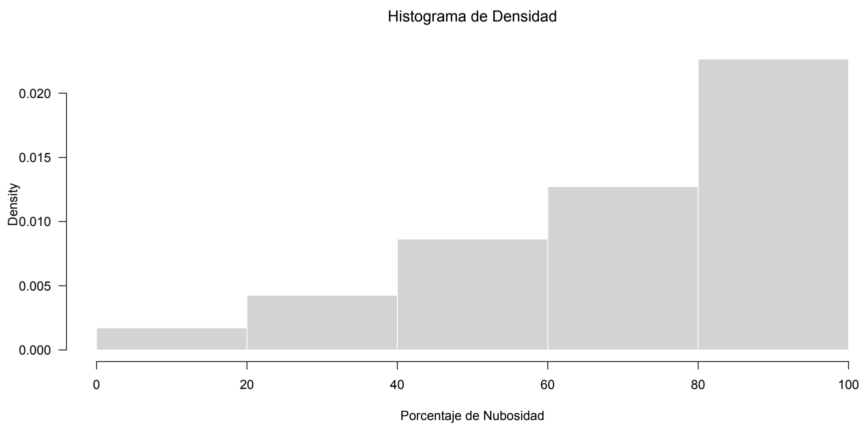
```
X = c(tapply(X = Base$DEMANDA, INDEX = list(Base$MES,Base$AÑO), mean))
X = c(tapply(X = Base$DEMANDA, INDEX = list(Base$MES+100*Base$AÑO), mean))
X = aggregate(DEMANDA ~ MES*AÑO, data = Base, FUN = mean)$DEMANDA

Tabla = round(rbind(IQR(X, na.rm = T),mean(X, na.rm = T),sd(X, na.rm = T)/mean(X, na.rm = T),
modeest::mlv(X, na.rm = T),quantile(X, prob = 0.56, na.rm = T)),2)
rownames(Tabla) = c("IQR", "Media", "c.o.v.", "Moda", "Percentil 56%")
colnames(Tabla) = "DEMANDA"
Tabla
               DEMANDA
IQR            1924.25
Media         10050.68
c.o.v.            0.14
Moda          10366.54
Percentil 56% 10293.46
```

Nota: Si el alumno obtiene las medidas correctamente con toda la data o con los 12 promedio mensuales, no descontar puntaje.

Pregunta 12:

Para poder construir el siguiente gráfico



Se utilizó la siguiente linea de comando. Complete los campos faltantes

```
[hist](X, [breaks] = 5, [las] = 1, [freq] = F, [main] = "Histograma de Densidad",
[font.main] = 1, [border] = "white", [xlab] = "Porcentaje de Nubosidad")
```

[0.125 Ptos] cada campo correctamente llenado.