Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2020

Interrogación 1 - Pauta

Pregunta 1: *Axiomas + Algebra + Independencia*

Si A, B y C son eventos posibles mutuamente independientes, entonces siempre se cumple que

(a) $(A \cap B)$ y C son independientes por ley asociativa y definición de independencia mutua.

(b) $(A \cup B)$ y C son independientes por ley distributiva, ley aditiva y definición de independencia mutua.

Pregunta 2: *Probabilidades Totales*

Estudios muestran que la prueba de aroma (KOR – Kit Olfativo Rápido desarrollado por la UC) logra identificar al $[B_1]\,\%$ de los Covid positivos. Sin embargo, un $[B_2]\,\%$ de personas sanas arrojan positivo debido por ejemplo a una anosmia (discapacidad olfativa). En cambio, el test de PCR es más certero ya que logra identificar al $[C_1]\,\%$ de los Covid positivos, y solo falla en un $[C_2]\,\%$ de las personas sanas.

Suponga que la UC ha diseñado el siguiente plan de retorno:

Aplicar el test de aroma a cada funcionario y si el teste da positivo este no regresa a sus labores. En cambio, si el test da negativo se aplica el test de PCR (que es más caro y lento). En caso que el test PCR resulte positivo, el funcionario no regresa a sus labores.

¿Cuál es la probabilidad que un funcionario al ser evaluado no pueda volver a sus labores? Suponga que el $[A]\,\%$ de los funcionarios al momento de realizarse el KOR tiene Covid.

Solución

Defina los siguientes eventos:

A: Enfermo de Covid al evaluar.
B: KOR +.
C: PCR +.
D: Funcionario no vuelve.

Se pide

$$P(D) = P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A}) + P(C \mid \overline{B} \cap A)P(\overline{B} \mid A)P(A) + P(C \mid \overline{B} \cap \overline{A})P(\overline{B} \mid \overline{A})P(\overline{A})$$

Del enunciado

$$P(C \mid \bar{B} \cap A) = C_1/100, \quad P(B \mid A) = B_1/100, \quad P(C \mid \bar{B} \cap \bar{A}) = C_2/100$$

$$P(B \mid \bar{A}) = B_2/100, \quad P(A) = A/100$$

Pregunta 3: *Teorema de Bayes*

Participación en el plebiscito
Un estudio muestra que el nivel de participación en el próximo plebiscito dependerá principalmente de la edad de los posibles votantes ante la incertidumbre del Covid en octubre.
De acuerdo a estudios, estos muestran un gran interés en participar en él. Cuando se analiza por grupo etario (¡30 años, 30-60 años, ¿60 años) los niveles de participación varían enormemente, dado que el nivel de participación va desde un [B1] %, [B2] % y [B3] %, respectivamente. Si usted sabe que las participaciones poblacionales de esos grupos etarios son de un 25 %, 60 % y 15 %, respectivamente.

Si un votante fue a votar en octubre, ¿cuál es la probabilidad que sea joven (¡30 años)?

Solución

$$\frac{B_1 \cdot 0{,}25}{B_1 \cdot 0{,}25 + B_2 \cdot 0{,}60 + B_3 \cdot 0{,}15}$$

Pregunta 4: *Conteo*

El domingo recién pasado Sebita cumplió 4 años y recibió varios regalos, ente ellos había un auto que requería 6 pilas AA. Lamentablemente, las pilas no venían incluida y en casa solo tenía 4 pilas AA nuevas.

Buscando, encontré 6 pilas AA que había apartado hace un tiempo para ir a dejarlas en un contenedor de pilas, ya que están descargadas, pero ahora eran útiles para hacer contacto.

Logré obtener otras dos de un control remoto, pero tienen media carga. Si Seba toma todas las pilas y las mezcla, ¿cuál es la probabilidad que el auto lo entretenga durante la tarde si tomo 6 pilas al azar?

Suponga que para que al auto funciones durante la tarde, se requiere la mitad al menos de carga. Notar que las dos pilas del control remoto equivalen en carga a una nueva.

Solución

Tenemos que $\# S = \binom{12}{6}$ y $\# A = \binom{4}{2} \cdot \binom{2}{2} \cdot \binom{6}{2} + \binom{4}{3} \cdot \binom{8}{3} + \binom{4}{4} \cdot \binom{8}{2}$.

```
(choose(4,4)*choose(8,2)+choose(4,3)*choose(8,3)+choose(4,2)*choose(2,2)*choose(6,2))/choose(12,6)
[1] 0.3701299
```

Pregunta 5: *Percentil*

Una alternativa al modelo Exponencial es una función de densidad definida por

$$f_X(x) = \alpha\beta(1+\alpha x)^{-(\beta+1)}$$

con $x > 0, \alpha > 0$ y $\beta > 0$ El percentil [p] % teórica de este modelo, si $\alpha =$[a] y $\beta =$[b], es:

Solución

Definamos como x_p al percentil $p \times 100\,\%$

Tenemos que

$$F_X\left(x_p\right) = \int_0^{x_p} \alpha\beta(1+\alpha x)^{-(\beta+1)}dx = 1 - \left(1+\alpha x_p\right)^{-\beta} = p$$

Despejando

$$x_p = \frac{(1-p)^{-1/\beta} - 1}{\alpha}$$

Pregunta 6: *Generadora de momento y coeficiente de variación*

El número de interrupciones, X, en el servicio de internet de un proveedor puede ser modelada por una variable aleatoria cuya función de probabilidad está dada por

$$p_X(x) = \left(\frac{\alpha}{1+\alpha}\right)^x \cdot \frac{1}{(1+\alpha)}$$

con $x \in \mathbb{N}_0$.

Si $\alpha =$[x] entonces el coeficiente de variación es:

Solución

Tenemos que

$$\begin{aligned}
M_X(t) &= \sum_{x=0}^{\infty} e^{tx} \left(\frac{\alpha}{1+\alpha}\right)^x \cdot \frac{1}{(1+\alpha)} \\
&= \frac{1}{(1+\alpha)} \sum_{x=0}^{\infty} \left(\frac{e^t \alpha}{1+\alpha}\right)^x \\
&= \frac{1}{(1+\alpha)} \cdot \frac{1}{\left[1 - \left(\frac{e^t\alpha}{1+\alpha}\right)\right]} \\
&= \left(1 + \alpha - e^t \alpha\right)^{-1}
\end{aligned}$$

Derivando dos veces e igualando a cero se tiene que

$$\mathrm{E}(X) = \alpha, \quad \mathrm{E}\left(X^2\right) = \alpha + 2\alpha^2 \to \mathrm{Var}(X) = \alpha + \alpha^2$$

Por lo tanto

$$\delta_X = \frac{\sqrt{\alpha + \alpha^2}}{\alpha} = \sqrt{\frac{1+\alpha}{\alpha}}$$

Pregunta 7: *Normal*

Frente al plebiscito próximo una de las grandes interrogantes es el tiempo que cada elector ocupará en realizar la acción de votar – entregar su cédula de identidad, firmar, ingresar a la cámara secreta, ejercer su derecho (seleccionar sus alternativas en cada cédula: Apruebo – Rechazo; Convención mixta – Convención constituyente), introducir sus votos en las urnas respectivas y recuperar su cédula de identidad. Un especialista afirma que ha evaluado su modelo y ha determinado que la probabilidad que demore más un minuto es de un tercio, valor que coincide con el coeficiente de variación. Asuma un distribución Normal.

¿Cuántos segundos le tomará al menos al cuarto de electores "más lento" el votar?

```
> mu = 60/(1+(1/3)*qnorm(2/3))
> sigma = mu*(1/3)
> qnorm(3/4, mu, sigma)
[1] 64.26316
```

Cuántos segundos al menos le tomará al sexto de electores "más lento" el votar?

```
> mu = 60/(1+(1/3)*qnorm(2/3))
> sigma = mu*(1/3)
> qnorm(5/6, mu, sigma)
[1] 69.38625
```

¿Cuántos segundos le tomará a lo más al quinto de electores "más veloces" el votar?

```
> mu = 60/(1+(1/3)*qnorm(2/3))
> sigma = mu*(1/3)
> qnorm(1/5, mu, sigma)
[1] 37.74789
```

¿Cuántos segundos le tomará a lo más al tercio de electores "más veloces" el votar?

```
> mu = 60/(1+(1/3)*qnorm(2/3))
> sigma = mu*(1/3)
> qnorm(1/3, mu, sigma)
[1] 44.93402
```

Pregunta 8: *Log-Normal*

Frente al plebiscito próximo una de las grandes interrogantes es el tiempo que cada elector ocupará en realizar la acción de votar – entregar su cédula de identidad, firmar, ingresar a la cámara secreta, ejercer su derecho (seleccionar sus alternativas en cada cédula: Apruebo – Rechazo; Convención mixta – Convención constituyente), introducir sus votos en las urnas respectivas y recuperar su cédula de identidad. Un especialista afirma que ha evaluado su modelo y ha determinado que la probabilidad que demore más un minuto es de un tercio, valor que coincide con el coeficiente de variación. Asuma un distribución Log-Normal.

¿Cuántos segundos al menos le tomará al quinto de electores "más lento" el votar?

```
> zeta = sqrt(log(1+(1/3)^2))
> lambda = log(60)-zeta*qnorm(2/3)
> qlnorm(4/5, meanlog = lambda, sdlog = zeta)
[1] 68.56058
```

Frente al plebiscito próximo una de las grandes interrogantes es el tiempo que cada elector ocupará en realizar la acción de votar – entregar su cédula de identidad, firmar, pasar al mesón de votación, ejercer su derecho (seleccionar sus alternativas en cada cédula: Apruebo – Rechazo; Convención mixta – Convención constituyente), cerrar los votos, introducirlos en las respectivas urnas y recuperar su cédula de identidad. Ensayos previos muestran que estos tiempo se comportan como variables aleatorias de tipo Log-Normal. Si la media registrada fue de 45 segundos, con una desviación estándar de 15 segundos, ¿cuánto demoraría a lo más el tercio más veloz?

```
> zeta = sqrt(log(1+delta^2))
> lambda = log(45)-0.5*zeta^2
> qlnorm(1/3, meanlog = lambda, sdlog = zeta)
[1] 37.12057
```

Pregunta 9: *DataI1.xlsx*

La base de datos DataI1.xlsx [Pulse para descargar] contiene una muestra aleatoria de consumo de energía eléctrica (en kWh) mensual de 200 mil clientes residenciales. La toma de muestra fue realizada entre enero y diciembre del año 2019 en las 33 comunas de la región metropolitana.

A partir de la mediana observada en septiembre en la comuna de Maipú ajuste un modelo Exponencial (visto en clases como ejemplo) a esos consumos mensuales y entregue el cociente entre el percentil 45 % estimado vs el empírico.

```
library(readxl)
library(dplyr)
Base = as.data.frame(read_excel("DataI1.xlsx"))
X = filter(Base, Mes == 9, Comuna == "Maipú")$Energia
lambda = log(2)/median(X)
qexp(0.45, rate = lambda)/quantile(X, probs = 0.45)
      45%
0.9170849
quantile(X, probs = 0.45)/qexp(0.45, rate = lambda)
     45%
1.090412
```

A partir de la media observada en mayo en la comuna de La Florida ajuste un modelo Exponencial (visto en clases como ejemplo) a esos consumos mensuales y entregue el cociente entre el percentil 60 % estimado vs el empírico.

```
library(readxl)
library(dplyr)
Base = as.data.frame(read_excel("DataI1.xlsx"))
X = filter(Base, Mes == 5, Comuna == "La Florida")$Energia
lambda = 1/mean(X)
qexp(0.60, rate = lambda)/quantile(X, probs = 0.60)
     60%
1.047496
quantile(X, probs = 0.60)/qexp(0.60, rate = lambda)
      60%
0.9546572
```

A partir de la media observada en mayo en la comuna de Las Condes ajuste un modelo Exponencial (visto en clases como ejemplo) a esos consumos mensuales y entregue el cociente entre el percentil 40 % estimado vs el empírico.

```
library(readxl)
library(dplyr)
Base = as.data.frame(read_excel("DataI1.xlsx"))
X = filter(Base, Mes == 5, Comuna == "Las Condes")$Energia
lambda = 1/mean(X)
qexp(0.40, rate = lambda)/quantile(X, probs = 0.40)
      40%
0.8228445
quantile(X, probs = 0.40)/qexp(0.40, rate = lambda)
     40%
1.215297
```

A partir de la mediana observada en diciembre en la comuna de La Florida ajuste un modelo Exponencial (visto en clases como ejemplo) a esos consumos mensuales y entregue el cociente entre el percentil 60 % estimado vs el empírico.

```
library(readxl)
library(dplyr)
Base = as.data.frame(read_excel("DataI1.xlsx"))
X = filter(Base, Mes == 5, Comuna == "Las Condes")$Energia
lambda = log(2)/median(X)
qexp(0.60, rate = lambda)/quantile(X, probs = 0.60)
     60%
1.118259
quantile(X, probs = 0.60)/qexp(0.60, rate = lambda)
      60%
0.8942473
```

A partir de la mediana observada en mayo en la comuna de Las Condes ajuste un modelo Exponencial (visto en clases como ejemplo) a esos consumos mensuales y entregue el cociente entre el percentil 40 % estimado vs el empírico.

```
library(readxl)
library(dplyr)
Base = as.data.frame(read_excel("DataI1.xlsx"))
X = filter(Base, Mes == 5, Comuna == "Las Condes")$Energia
lambda = log(2)/median(X)
qexp(0.40, rate = lambda)/quantile(X, probs = 0.40)
      40%
```

```
0.8834708
quantile(X, probs = 0.40)/qexp(0.40, rate = lambda)
     40%
1.131899
```

A partir de la media observada en septiembre en la comuna de Maipú ajuste un modelo Exponencial (visto en clases como ejemplo) a esos consumos mensuales y entregue el cociente entre el percentil 45 % estimado vs el empírico.

```
library(readxl)
library(dplyr)
Base = as.data.frame(read_excel("DataI1.xlsx"))
X = filter(Base, Mes == 9, Comuna == "Maipú")$Energia
lambda = 1/mean(X)
qexp(0.45, rate = lambda)/quantile(X, probs = 0.45)
     45%
0.774295
quantile(X, probs = 0.45)/qexp(0.45, rate = lambda)
     45%
1.291497
```

Pregunta 10: *DataI1.xlsx*

La base de datos DataI1.xlsx [Pulse para descargar] contiene una muestra aleatoria de consumo de energía eléctrica (en kWh) mensual de 200 mil clientes residenciales. La toma de muestra fue realizada entre enero y diciembre del año 2019 en las 33 comunas de la región metropolitana.

Uilizando el percentil 25 % y la mediana ajuste una distribución Log- Normal a los datos.

¿Según este modelo cuál debería ser el IQR?

```
library(readxl)
library(dplyr)
Base = as.data.frame(read_excel("DataI1.xlsx"))
X = Base$Energia
lambda = log(median(X))
zeta = (log(quantile(X,0.25))-lambda)/qnorm(.25)
qlnorm(0.75,lambda,zeta)-qlnorm(0.25,lambda,zeta)
[1] 167.3846
```