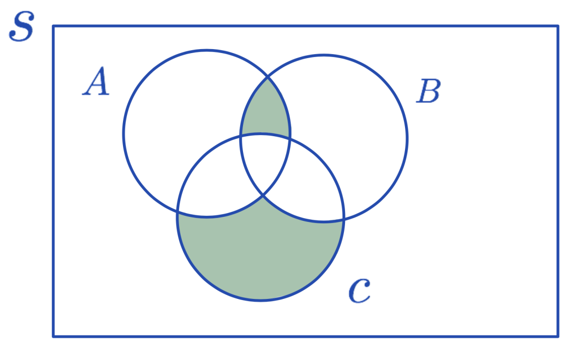
Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2021

Interrogación 1 - Pauta

Pregunta 1

Considere el siguiente diagrama de Venn



¿Cuál de las alternativas representa al evento compuesto por los puntos muestrales contenidos en las zonas de color verde?

Solución

Los dos eventos que aparecen en la figura se pueden generar a partir de las siguientes operaciones: $\left[\overline{(A \cup B)} \cap C\right]$ y $\left[(A \cap B) \cap \overline{C}\right]$.

Aplicando De Morgan se tiene que la unión de ambos está dada por:

$$\left[\overline{(A \cup B)} \cap C\right] \cup \left[(A \cap B) \cap \overline{C}\right] = \left[(\overline{A} \cap \overline{B}) \cap C\right] \cup \left[(A \cap B) \cap \overline{C}\right] = \left[\overline{A} \cap \overline{B} \cap C\right] \cup \left[A \cap B \cap \overline{C}\right]$$

Pauta Apelación

- *Respuesta correcta. [+1.0 Puntos]*
- *Si alumno marca más de una alternativa. [+0.0 Puntos]*
- *Para otros casos, ingresar apelación y equipo de ayudantes evaluará caso a caso.*

Pregunta 2

Para esta interrogación, los 926 alumnos que actualmente aparecen en la lista oficial del curso fueron asignados al azar a cuatro link de Zoom con las siguientes restricciones:

Link 1: n_1 alumnos.

Link 2: n_2 alumnos.

Link 3: n_3 alumnos.

¿Cuál es la probabilidad que los amigos A y B sean asignados al mismo link?

Solución

Definamos como C el evento en que los amigos quedan asignados en el mismo link.

- Número de formas de realizar la asignación es: $\dfrac{n!}{n_1!\,n_2!\,n_3!\,n_4!}$.
- Maneras que queden asignados en el link 1: $\dfrac{(n-2)!}{(n_1-2)!\,n_2!\,n_3!\,n_4!}$.
- Maneras que queden asignados en el link 2: $\dfrac{(n-2)!}{n_1!\,(n_2-2)!\,n_3!\,n_4!}$.
- Maneras que queden asignados en el link 3: $\dfrac{(n-2)!}{n_1!\,n_2!\,(n_3-2)!\,n_4!}$.
- Maneras que queden asignados en el link 4: $\dfrac{(n-2)!}{n_1!\,n_2!\,n_3!\,(n_4-2)!}$.

con

$$n = 926 \quad \text{y} \quad n_4 = n - n_1 - n_2 - n_3$$

Por lo tanto

$$P(C) = \frac{\# C}{\# S} = \frac{\displaystyle\sum_{i=1}^{4} n_i\,(n_i - 1)}{926 \times 925}$$

```
## Ejemplo

n = 926
n1 = 230
n2 = 240
n3 = 220
n4 = n - n1 - n2 - n3
(n1*(n1-1)+n2*(n2-1)+n3*(n3-1)+n4*(n4-1))/(n*(n-1))
[1] 0.2494542
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+1.0 Puntos]*
- *Respuesta incorrecta, sobre el margen de error* ± 0.005*: [+0.6 Puntos] sí los casos favorables estén correctamente calculados y [+0.2 Puntos] si los casos totales están correctos.*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje desglosado en el item anterior.*
- *Si el alumno realizo el ejercicios mediante simulaciones y frecuencia empírica, explicar código de respaldo para su revisión.*

Pregunta 3

Una planta automotriz produce tres tipos de vehículos: Sedán, SUV y Deportivos. El `[p]` % de la producción corresponde a Sedán, un 20 % a SUV y el resto son Deportivos.

Por otra parte, los motores eléctricos han incrementado su participación y actualmente corresponden a la mitad de los producidos en esta planta.

Si 2/3 de los sedanes y él [q] % de los SUV son eléctricos. ¿Cuál es la probabilidad que un auto Deportivo producido por esta planta sea eléctrico?

Solución

Definamos los eventos A_1: Sedán, A_2: SUV, A_3: Deportivos y E: Eléctrico.

Del enunciado se tiene que

$$P(E) = 0.50, \quad P(A_1) = \frac{\texttt{[p]}}{100}, \quad P(A_2) = 0.20, \quad P(A_3) = 1 - P(A_1) - P(A_2)$$

$$P(E \,|\, A_1) = \frac{2}{3}, \quad \text{y} \quad P(E \,|\, A_2) = \frac{\texttt{[q]}}{100}$$

Se pide $P(E \,|\, A_3)$.

Por teorema de probabilidades totales se tiene que

$$P(E \,|\, A_3) = \frac{P(E) - P(E \,|\, A_1) \cdot P(A_1) + P(E \,|\, A_2) \cdot P(A_2)}{P(A_3)}$$

```
## Ejemplo

p = 0.60
q = 0.35
p1 = p
p2 = 0.2
p3 = 1 - p1 - p2
q1 = 2/3
q2 = q
q3 = (0.5 - p1*q1 - p2*q2)/p3
q3
[1] 0.15
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+1.0 Puntos]*
- *Respuesta incorrecta, sobre el margen de error* ± 0.005*: [+0.8 Puntos] sí aplica correctamente el teorema de probabilidades totales.*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje desglosado en el item anterior.*

Pregunta 4

Hoy en día un gran porcentaje de las compras se realizan utilizando tarjeta.

Estudios de un banco muestran que 1/3 de las compras que se realizan con tarjeta, son de débito (contado "inmediato"), mientras que el resto de las compras son con tarjeta de crédito (TC). Entre las compras con TC, un `[p]` % son en una cuota y el resto elige pagar en dos o más cuotas (con posibles intereses).

Respecto a los montos, el `[q]` % de los pagos con débito son inferiores a M$100. En cambio, los pagos con TC a una cuota la mitad es inferior a M$100, mientras que sólo el 20 % de los pagos en dos o más cuotas son inferiores a M$100.

Usted realiza una compra que supera los M$100, ¿cuál es la probabilidad que sea con TC en una sola cuota?

M$: miles de pesos.

Solución

Definamos los siguientes eventos

A: Pago con tarjeta de débito.
B: Pago en una cuota.
C: Monto inferior a M$100.

Del enunciado se tiene que

$$P(A) = 1/3, \quad P(B \,|\, \overline{A}) = \frac{\texttt{[p]}}{100}, \quad P(C \,|\, A) = \frac{\texttt{[q]}}{100}, \quad P(C \,|\, \overline{A} \cap B) = 1/2, \quad P(C \,|\, \overline{A} \cap \overline{B}) = 0.20$$

Se pide $P(\overline{A} \cap B \,|\, \overline{C})$.

Por teorema de probabilidades totales se tiene que

$$\begin{aligned} P(\overline{C}) &= P(\overline{C} \cap A) + P(\overline{C} \cap \overline{A}) \\ &= P(\overline{C} \cap A) + P(\overline{C} \cap \overline{A} \cap B) + P(\overline{C} \cap \overline{A} \cap \overline{B}) \\ &= P(\overline{C} \,|\, A)\, P(A) + P(\overline{C} \,|\, \overline{A} \cap B)\, P(B \,|\, \overline{A})\, P(\overline{A}) + P(\overline{C} \,|\, \overline{A} \cap \overline{B})\, P(\overline{B} \,|\, \overline{A})\, P(\overline{A}) \end{aligned}$$

y por teorema de Bayes

$$P(\overline{A} \cap B \,|\, \overline{C}) = \frac{P(\overline{C} \,|\, \overline{A} \cap B)\, P(B \,|\, \overline{A})\, P(\overline{A})}{P(\overline{C})}$$

```
## Ejemplo

p = 66
q = 70
(1-1/2)*(p/100)*(2/3)/((1-q/100)*(1/3) + (1-1/2)*(p/100)*(2/3) + (1-0.20)*(1-p/100)*(2/3))
[1] 0.4388298
```

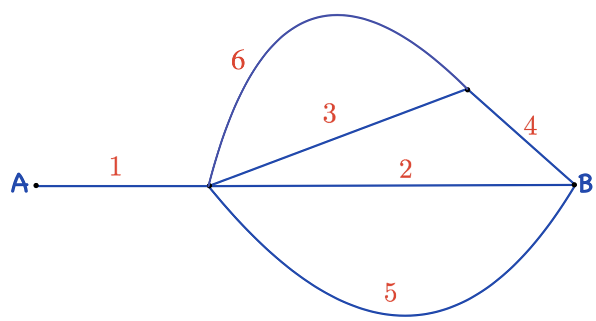
Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+1.0 Puntos]*
- *Respuesta incorrecta, sobre el margen de error* ± 0.005: *[+0.6 Puntos] sí aplica correctamente el teorema de probabilidades totales y [+0.2 Puntos] por aplicar teorema de Bayes correctamente.*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje desglosado en el item anterior.*

Pregunta 5

Usted ha comprado recientemente un producto por e-commerce. El delivery que lleva el producto tiene 4 posibles rutas con idéntico tiempo de entrega entre A y B. Si él elige una ruta al azar entre: 1-3-4, 1-2, 1-6-4 o 1-5, y en cada uno de los 6 tramos que componen estas rutas, hay un probabilidad [p] de retrasarse, ¿cuál es la probabilidad que cumpla con el tiempo de entrega? Suponga que los retrasos en cada tramo ocurren de manera independiente.

La siguiente Figura ilustra las cuatro rutas posibles:



Solución

Definamos como A el evento "cumpla con el tiempo de entrega".

Por teorema de probabilidades totales y la independencia entre los tramos se tiene que

$$P(A) = \frac{1}{4}\left[(1-\texttt{[p]})^3 + (1-\texttt{[p]})^3 + (1-\texttt{[p]})^2 + (1-\texttt{[p]})^2\right] = \frac{(1-\texttt{[p]})^2\,(2-\texttt{[p]})}{2}$$

```
## Ejemplo

p  = 0.4
0.5 * (1-p)^2 * (2-p)
[1] 0.288
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+1.0 Puntos]*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje.*

Pregunta 6

Ayer lunes 20 de septiembre, los medios de comunicación informaron que un análisis de aguas servidas permitió detectar el primer caso de variante Delta del Covid-19 en la comuna de Arauco, Región del Biobío.

Gracias al sistema de vigilancia del Sars Cov 2 en aguas servidas, la autoridad sanitaria logró detectar una zona particular con presencia de la variante Delta, para poder realizar una búsqueda activa de casos.

Al empadronar la zona se registraron `[N]` viviendas y se procedió a seleccionar `[n]` de estas al azar (sin reemplazo) para realizar exámenes PCR a sus integrantes. ¿Cuál es la probabilidad que al menos dos viviendas entre las seleccionadas tengan integrantes con PCR+ variante Delta? Suponga que el número de viviendas con presencia Delta en esta zona es de `[m]`.

Solución

La cantidad de maneras de seleccionar `[n]` viviendas es $\# S = \binom{\texttt{N}}{\texttt{n}}$.

Definamos como A al evento al menos dos viviendas PC R+ Delta son seleccionadas.

Se pide $P(A) = 1 - \dfrac{\# \overline{A}}{\# S}$, donde

$$\# \overline{A} = \binom{\texttt{m}}{\texttt{0}} \binom{\texttt{N-m}}{\texttt{n}} + \binom{\texttt{m}}{\texttt{1}} \binom{\texttt{N-m}}{\texttt{n-1}}$$

```
## Ejemplo

N = 55
m = 6
n = 12
1-(choose(m,0)*choose(N-m,n) + choose(m,1)*choose(N-m,n-1))/choose(N,n)
[1] 0.3912443
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+1.0 Puntos]*
- *Respuesta incorrecta, sobre el margen de error* ± 0.005*: [+0.6 Puntos] sí los casos favorables estén correctamente calculados y [+0.2 Puntos] si los casos totales están correctos.*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje desglosado en el item anterior.*
- *Si el alumno realizo el ejercicios mediante simulaciones y frecuencia empírica, explicar código de respaldo para su revisión.*

Pregunta 7

Un estudiante de otra Universidad que está por tomar un curso catalogado como difícil por sus pares y la probabilidad de de aprobación histórica es de `[p]`.

En el caso de reprobar, lo debe tomar al siguiente semestre y en el caso que nuevamente repruebe, podrá hacer el curso por tercera y última vez, debido a la regla que esta universidad tiene establecida en su reglamento.

Suponiendo que la probabilidad de aprobación o reprobación permanece constante, determine la probabilidad que el estudiante repruebe más de una vez.

Solución

Definamos como X el número de veces que reprueba el curso, con $\Theta_X = \{0, 1, 2, 3\}$. Como la probabilidad de aprobación permanece constante, se tiene que

$$p_X(x) = \begin{cases} \mathtt{p}, & x = 0 \\ (1 - \mathtt{p})\,\mathtt{p}, & x = 1 \\ (1 - \mathtt{p})^2\,\mathtt{p}, & x = 2 \\ (1 - \mathtt{p})^3, & x = 3 \end{cases}$$

Se pide $P(X > 1) = (1 - \mathtt{p})^2\,\mathtt{p} + (1 - \mathtt{p})^3 = (1 - \mathtt{p})^2$.

```
## Ejemplo

p = 0.9
(1-p)^2
[1] 0.01
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+1.0 Puntos]*
- *Respuesta incorrecta, sobre el margen de error* ± 0.005*: [+0.4 Puntos] sí aplica correctamente independencia y [+0.4 Puntos] si considera todos los casos favorables.*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje desglosado en el item anterior.*

Pregunta 8

Considere una variable aleatoria X cuya distribución tiene la siguiente densidad

$$f_X(x) = \frac{a\,b}{a-b}\left(e^{-b\,x} - e^{-a\,x}\right), \quad x > 0$$

con $a =$ `[a]` y $b =$ `[b]`. Calcule la esperanza de X.

Solución

Se pide

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^\infty x \cdot \frac{a\,b}{a-b}\left(e^{-b\,x} - e^{-a\,x}\right)\,dx \\
&= \frac{a\,b}{a-b}\left\{\int_0^\infty x \cdot e^{-b\,x} - \int_0^\infty x \cdot e^{-a\,x}\right\} \\
&= \frac{a\,b}{a-b}\left\{\frac{1}{b}\int_0^\infty x \cdot b\,e^{-b\,x} - \frac{1}{a}\int_0^\infty x \cdot a\,e^{-a\,x}\right\} \\
&= \frac{a\,b}{a-b}\left\{\frac{1}{b^2} - \frac{1}{a^2}\right\} \\
&= \frac{a+b}{a\,b}
\end{aligned}$$

```
## Ejemplo

a = 5
b = 3
(a+b)/(a*b)
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+1.0 Puntos]*
- *Respuesta incorrecta, sobre el margen de error* ± 0.005*: [+0.4 Puntos] hasta 3ra igualdad, [+0.4 Puntos] si llega a la expresión correcta.*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje desglosado en el item anterior.*

Pregunta 9

Según un estudio, la temperatura máxima T (en grados Celsius) de cada día durante la primavera tiene una distribución con la siguiente función de densidad:

$$f_T(t) = \begin{cases} \dfrac{1}{10}, & 21 < t < 31 \\ & \\ 0, & \text{e.o.c.} \end{cases}$$

Asuma que las temperaturas máximas observadas diariamente son independientes. ¿Cuál es la probabilidad de que tanto mañana como pasado mañana se registren temperaturas máximas sobre los `[t]`°C?

Solución

Definamos como T a la temperatura máxima de un día cualquier de primavera, la cual se comporta como una variable aleatoria Uniforme(21, 31), por lo tanto:

$$P(T > \texttt{t}) = 1 - \frac{\texttt{t} - 21}{31 - 21}.$$

Se pide

$$P(\{T_{\text{mañana}} > \texttt{t}\} \cap \{T_{\text{pasado mañana}} > \texttt{t}\})$$

Por independencia se tiene que

$$\begin{aligned} P(\{T_{\text{mañana}} > \texttt{t}\} \cap \{T_{\text{pasado mañana}} > \texttt{t}\}) &= P(\{T_{\text{mañana}} > \texttt{t}\}) \cdot P(\{T_{\text{pasado mañana}} > \texttt{t}\}) \\ &= \left\{1 - \frac{\texttt{t} - 21}{31 - 21}\right\}^2 \end{aligned}$$

```
## Ejemplo

t = 24
p = 1 - (t-21)/(31-21)
p^2
[1] 0.49
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+1.0 Puntos]*
- *Respuesta incorrecta, sobre el margen de error* ± 0.005*: [+0.4 Puntos] por aplicar correctamente independencia y [+0.4 Puntos] por calcular correctamente* $P(T > \texttt{[t]})$.
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje desglosado en el item anterior.*

El web scraping o raspado web, es una técnica utilizada mediante programas de software para extraer información de sitios web. Usualmente, estos programas simulan la navegación de un humano en la World Wide Web ya sea utilizando el protocolo HTTP manualmente, o incrustando un navegador en una aplicación.

El conjunto de datos [Marvel.csv] se creó con el propósito de la visualización datos relevantes extraídos desde https://www.imdb.com/list/ls031310794/ correspondiente a las películas que componen el UCM (Universo Cinematográfico de Marvel):

```
NombrePelicula
FechaLanzamientoEEUU               : Mes Día, Año
Director
Productor
Duracion                           : Hora Minutos
Genero                             : Acción, Aventura, Drama, etc.
CalificacionIMDB                   : 0 a 10
PromedioCriticas                   : 0 a 100
RepartoPresupuesto                 : (en US$)
RecaudacionEEUUCanada              : (en US$)
RecaudacionGlobal                  : (en US$)
RecaudacionEEUUCanadaPrimeraSemana : (en US$)
NominacionesPremiosOscar           : 0,1,2,3,4,5,6,7
PremiosOscarGanados                : 0,1,2,3
FaseMCU                            : 1,2,3
```

Pregunta 10

Para la variable en millones de dólares, ¿cuál es la diferencia en mediana y desviación estándar entre las películas que tienen al menos una nominación al premio Oscar vs. las que no tienen nominaciones?

Diferencia entre medianas = 114.3759

Diferencia entre desviación estándar = 338.6903

```
Data <- rio::import("Marvel.csv")
Data$RecaudacionGlobal <- Data$RecaudacionGlobal/1000000
Data$Filtro <- Data$NominacionesPremiosOscar > 0
median(Data$RecaudacionGlobal[Data$Filtro == TRUE])-median(Data$RecaudacionGlobal[Data$Filtro == FALSE])
sd(Data$RecaudacionGlobal[Data$Filtro == TRUE])-sd(Data$RecaudacionGlobal[Data$Filtro == FALSE])
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+0.5 Puntos] cada respuesta.*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje al citar sus respaldos.*
- *Para otros casos, ingresar apelación y equipo de ayudantes evaluará caso a caso.*

Pregunta 11

¿Qué Fase del UCM es más heterogénea con respecto al total que recaudó cada película en EE.UU. y Canadá (`RecaudacionEEUUCanada`)?

Fase = `3`

Indique la medida estadística utilizada = `Coeficiente de Variación`

Valor medida estadística utilizada = `0.6603804`

```
Data <- rio::import("Marvel.csv")
delta <- function(x){sd(x, na.rm = T)/mean(x, na.rm = T)}
delta(Data$RecaudacionEEUUCanada[Data$FaseMCU == 1])
delta(Data$RecaudacionEEUUCanada[Data$FaseMCU == 2])
delta(Data$RecaudacionEEUUCanada[Data$FaseMCU == 3])
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+0.2 Puntos] por Fase correcta, [+0.4 Puntos] por ingresar medida estadística correcta y [+0.4 Puntos] por valor correcto correspondiente a la Fase más Heterogénea.*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje desglosado anteriormente al citar sus respaldos.*
- *Para otros casos, ingresar apelación y equipo de ayudantes evaluará caso a caso.*

Pregunta 12

Considere la recaudación (en millones de dólares) de las películas que componen él UCM en EE.UU. y Canadá durante la primera semana de exhibición (`RecaudacionEEUUCanadaPrimeraSemana`).

En clases, a modo de ilustración, se analizaron algunas medidas descriptivas teóricas del modelo Exponencial(λ).

A partir de la mediana de lo recaudado por las películas, asigne un valor para el parámetro lambda y entregue la diferencia absoluta entre la probabilidad empírica y teórica estimada del siguiente evento:

$$\{\texttt{RecaudacionEEUUCanadaPrimeraSemana} > 100\}$$

```
Data   <- rio::import("Marvel.csv")
X      <- Data$RecaudacionEEUUCanadaPrimeraSemana/1000000
lambda <- log(2)/median(X)
abs((1-(1-exp(-lambda*100)))-mean(X>100))
[1] 0.03131784
```

Pauta Apelación

- *Respuesta correcta, con margen de error* ± 0.005. *[+1.0 Puntos]*
- *Si el alumno NO ingreso respuesta en Canvas, solo podrá apelar a la mitad del puntaje*
- *Para otros casos, ingresar apelación y equipo de ayudantes evaluará caso a caso.*