Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Temporada Académica de Verano (enero 2022)

Interrogación 1 - Pauta

Pregunta 1

Sea A, B y C, tres eventos asociados a un mismo experimento. Suponga que cumplen con las siguientes condiciones:

$$P(A) = \frac{1}{2},\ P(B) = \frac{1}{3},\ P(C) = \frac{1}{4}$$

$$P(A \cap B) = \frac{1}{5},\ P(A \cap C) = \frac{1}{6} \text{ y } P(B \cap C) = 0$$

Calcule $P(A \cap \overline{(B \cup C)})$.

Solución

Tenemos que

$$\begin{aligned}
P(A) = P(A \cap S) &= P\left(A \cap \left[(B \cup C) \cup \overline{(B \cup C)}\right]\right) \\
&= P(A \cap [B \cup C]) + P(A \cap \overline{[B \cup C]}), \quad \text{por axioma 3} \\
&= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C) + P(A \cap \overline{[B \cup C]}), \quad \text{por ley distributiva y aditiva} \qquad (1)
\end{aligned}$$

Por otra parte

$$0 = P(B \cap C) = P\left([B \cap C] \cap A\right) + P\left([B \cap C] \cap \overline{A}\right) \geq P(A \cap B \cap C) \rightarrow P(A \cap B \cap C) = 0 \text{ (por axioma 1)} \qquad (2)$$

Reemplazando (2) en (1) se tiene que

$$P(A \cap \overline{[B \cup C]}) = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C) = \frac{1}{2} - \frac{1}{5} - \frac{1}{6} + 0 = 0.1333333 \quad \textbf{[1.0 Puntos]}$$

Pregunta 2

Determine si las siguientes proposiciones son verdaderas o falsas:

(a) $P(A \cap B) \geq 1 - P(\bar{A}) - P(\bar{B})$. **[Verdadero]**

(b) Si $P(A \mid E) \geq P(B \mid E)$ y $P(A \mid \bar{E}) \geq P(B \mid \bar{E})$, entonces $P(A) \geq P(B)$. **[Verdadero]**

Solución

(a) Tenemos que

$$\begin{aligned}
P(A \cap B) &= P\left(\overline{\overline{A} \cup \overline{B}}\right), \quad \text{por De Morgan} \quad \textbf{[0.125 Puntos]} \\
&= 1 - P(\overline{A} \cup \overline{B}), \quad \text{por ley del complemento} \quad \textbf{[0.125 Puntos]} \\
&= 1 - P(\overline{A}) - P(\overline{B}) + P(\overline{A} \cap \overline{B}), \quad \text{por ley aditiva} \quad \textbf{[0.125 Puntos]} \\
&\geq 1 - P(\overline{A}) - P(\overline{B}), \quad \text{por axioma 1} \quad \textbf{[0.125 Puntos]}
\end{aligned}$$

Por lo tanto, la proposición es verdadera.

(b) Tenemos que

$$P(A \,|\, E) \geq P(B \,|\, E) \to P(A \cap E) \geq P(B \cap E) \quad \textbf{[0.15 Puntos]}$$

y

$$P(A \,|\, \overline{E}) \geq P(B \,|\, \overline{E}) \to P(A \cap \overline{E}) \geq P(B \cap \overline{E}) \quad \textbf{[0.15 Puntos]}$$

Como

$$P(A) = P(A \cap E) + P(A \cap \overline{E}) \geq P(B \cap E) + P(B \cap \overline{E}) = P(B) \quad \textbf{[0.20 Puntos]}$$

la proposición es verdadera.

Pregunta 3.1

El fin de semana 14 amigos se reúnen para jugar un partido de futbolito de siete jugadores por lado. Si en el grupo hay dos arqueros, seis defensas, cuatro delanteros y dos contención, ¿cuál es la probabilidad que si se arman los equipos al azar, al menos hay un equilibrio defensivo, es decir, cada equipo tiene 3 defensas y un arquero?

Solución

Tenemos $\#S = \binom{14}{7}$ formas de armar los equipos. **[0.4 Puntos]**

El equilibrio defensivo ocurre en $\# A = \binom{2}{1} \cdot \binom{6}{3} \cdot \binom{6}{3}$ casos. **[0.4 Puntos]**

Por lo tanto,

$$P(A) = \frac{\# A}{\# S} = \texttt{choose(2,1)*choose(6,3)*choose(6,3)/choose(14,7)} = \texttt{0.2331002}$$ **[0.2 Puntos]**

Pregunta 3.2

El fin de semana 14 amigos se reúnen para jugar un partido de futbolito de siete jugadores por lado. Si en el grupo hay dos arqueros, seis defensas, cuatro delanteros y dos contención, ¿cuál es la probabilidad que si se arman los equipos al azar, al menos hay un equilibrio ofensivo, es decir, cada equipo tiene 2 delanteros y un contención?

Solución

Tenemos $\#S = \binom{14}{7}$ formas de armar los equipos. **[0.4 Puntos]**

El equilibrio ofensivo ocurre en $\# B = \binom{4}{2} \cdot \binom{2}{1} \cdot \binom{8}{4}$ casos. **[0.4 Puntos]**

Por lo tanto,

$$P(B) = \frac{\# B}{\# S} = \texttt{choose(4,2)*choose(2,1)*choose(8,4)/choose(14,7)} = \texttt{0.2447552}$$ **[0.2 Puntos]**

Pregunta 4

Actualmente la población de cinco años o más ha tenido la oportunidad de inocularse (vacunarse) contra el COVID y según estadísticas provenientes del MINSAL el 10 % no se ha vacunado, el 5 % tiene una dosis, el 25 % dos dosis y el 60 % dos dosis + refuerzo. Suponga que independientemente del estado de vacunación, la exposición al COVID ocurre en el 20 % de los casos. La probabilidad de contagio en el caso que una persona este expuesta al virus es de `[B1]` si presenta dos dosis + refuerzo, `[B2]` si solo tiene dos dosis, `[B3]` cuando tiene una dosis y `[B4]` en el caso que no este vacunado.

¿Cuál es la probabilidad que una persona se contagie de covid?

Nota: Si no hay exposición al COVID, entonces la probabilidad de contagio es cero.

Solución

Consideremos los siguientes eventos:

A_1: No vacunado.

A_2: Solo una dosis.

A_3: Solo dos dosis.

A_4: Dos dosis + refuerzo.

E: Persona expuesta al virus.

B: Persona se contagia.

Del enunciado, se tiene que

$$P(A_1) = 0.10, \quad P(A_2) = 0.05,\ P(A_3) = 0.25,\ P(A_4) = 0.60$$

$$P(E \mid A_1) = P(E \mid A_2) = P(E \mid A_3) = P(E \mid A_4) = 0.20$$

$$P(B \mid A_1 \cap \overline{E}) = P(B \mid A_2 \cap \overline{E}) = P(B \mid A_3 \cap \overline{E}) = P(B \mid A_4 \cap \overline{E}) = 0$$

$$P(B \mid A_1 \cap E) = \texttt{[B4]},\ P(B \mid A_2 \cap E) = \texttt{[B3]},\ P(B \mid A_3 \cap E) = \texttt{[B2]},\ P(B \mid A_4 \cap E) = \texttt{[B1]}$$

[0.5 Puntos]

Se pide $P(B)$.

Por teorema de probabilidades totales se tiene que

$P(B) = \texttt{[B4]} \times 0.20 \times 0.10 + \texttt{[B3]} \times 0.20 \times 0.05 + \texttt{[B2]} \times 0.20 \times 0.25 + \texttt{[B1]} \times 0.20 \times 0.60$ **[0.5 Puntos]**

Por ejemplo,

```
B1 = 0.1
B2 = 0.3
B3 = 0.5
B4 = 0.9
B4*0.20*0.10 + B3*0.20*0.05 + B2*0.20*0.25 + B1*0.20*0.60
[1] 0.05
```

Nota: Respuesta con una diferencia de ± 0.0005 *asignar puntaje completo.*

Pregunta 5

Considerando la información entregada en la pregunta anterior. Si una persona no presenta contagio, ¿cuál es la probabilidad que no esté vacunada?

Solución

Se pide $P(A_1 \,|\, \overline{B})$.

Aplicando la definición de probabilidad condicional y ley del complemento

$$P(A_1 \,|\, \overline{B}) = \frac{P(A_1 \cap \overline{B})}{P(\overline{B})} = \frac{P(A_1 \cap \overline{B})}{1 - P(B)} \quad \textbf{[0.4 Puntos]}$$

Por otra parte, tenemos que

$$\begin{aligned} P(A_1 \cap \overline{B}) &= P(A_1 \cap \overline{B} \cap E) + P(A_1 \cap \overline{B} \cap \overline{E}) \\ &= P(\overline{B} \,|\, A_1 \cap E) \cdot P(E \,|\, A_1) \cdot P(A_1) + P(\overline{B} \,|\, A_1 \cap \overline{E}) \cdot P(\overline{E} \,|\, A_1) \cdot P(A_1) \\ &= (1 - \texttt{[B4]}) \times 0.20 \times 0.10 + 1.00 \times 0.80 \times 0.10 \quad \textbf{[0.4 Puntos]} \end{aligned}$$

Por lo tanto

$$P(A_1 \,|\, \overline{B}) = \frac{(1 - \texttt{[B4]}) \times 0.20 \times 0.10 + 1.00 \times 0.80 \times 0.10}{1 - (\texttt{[B4]} \times 0.20 \times 0.10 + \texttt{[B3]} \times 0.20 \times 0.05 + \texttt{[B2]} \times 0.20 \times 0.25 + \texttt{[B1]} \times 0.20 \times 0.60)} \quad \textbf{[0.2 Puntos]}$$

Por ejemplo,

```
B1 = 0.1
B2 = 0.3
B3 = 0.5
B4 = 0.9
((1-B4)*0.20*0.10 + 1.00*0.80*0.10)/(1-(B4*0.20*0.10+B3*0.20*0.05+B2*0.20*0.25+B1*0.20*0.60))
[1] 0.08631579
```

Nota: Respuesta con una diferencia de ± 0.0005 *asignar puntaje completo.*

Pregunta 6

Suponga que X es una variable aleatoria cuya función de densidad esta definida como sigue:

$$f_X(x) = \frac{(\alpha+1)x^{\alpha}}{2^{\alpha+1}},$$

con $0 < x < 2$ y $\alpha > -1$.

Para $\alpha =$ `[alpha]` calcule el IQR de X.

Nota: Respuesta con una diferencia de ± 0.0005 *asignar puntaje completo.*
Solución

Tenemos que

$$F_X(x_p) = \int_0^{x_p} \frac{(\alpha+1)\,x^{\alpha}}{2^{\alpha+1}}\,dx = \frac{(\alpha+1)\,x_p^{\alpha+1}}{2^{\alpha+1}\,(\alpha+1)} = p \to x_p = 2 \cdot p^{1/(\alpha+1)} \quad \textbf{[0.8 Puntos]}$$

Se pide

$$\text{IQR} = x_{0.75} - x_{0.25} \quad \textbf{[0.2 Puntos]}$$

Por ejemplo,

```
alpha = 1.5
2*0.75^(1/(alpha+1)) - 2*0.25^(1/(alpha+1))
[1] 0.6339041
```

Pregunta 7.1

Hoy [m] alumnos se encuentran realizando la prueba desde la sala CS103 ubicada en el edifico de ciencias de la salud. Si [N] alumnos hoy se conectan a rendir la prueba, ¿cuál es la probabilidad que al realizar un muestreo sin reemplazo de tamaño [n], [x] alumnos se encuentren conectados desde la sala CS103?

Solución

Como la muestra es sin reemplazo, el número de conectados desde la sala CS103 en la muestra distribuye

$$\text{Hipergeometrica}(n = \texttt{[n]}, N = \texttt{[N]}, m = \texttt{[m]}). \quad \textbf{[0.5 Puntos]}$$

Se pide

$$p_X(x) = \texttt{dhyper(x, m = m, n = N-m, k = n)} \quad \textbf{[0.5 Puntos]}$$

Por ejemplo,

```
N = 150
m = 55
n = 35
x = 10
dhyper(x, m = m, n = N-m, k = n)
[1] 0.08602675
```

Pregunta 7.2

Hoy [m] alumnos se encuentran realizando la prueba desde la sala CS103 ubicada en el edifico de ciencias de la salud. Si [N] alumnos hoy se conectan a rendir la prueba, ¿cuál es la probabilidad que al realizar un muestreo con reemplazo de tamaño [n], [x] alumnos se encuentren conectados desde la sala CS103?

Solución

Como la muestra es con reemplazo, el número de conectados desde la sala CS103 en la muestra distribuye

$$\text{Binomial}(n = \texttt{[n]}, p = \texttt{[m]/[N]}). \quad \textbf{[0.5 Puntos]}$$

Se pide

$$p_X(x) = \texttt{dbinom(x, size = n, prob = m/N)} \quad \textbf{[0.5 Puntos]}$$

Por ejemplo,

```
N = 150
m = 55
n = 35
x = 10
dbinom(x, size = n, prob = m/N)
[1] 0.08858229
```

Volcán Nevados de Chillán

Entre el 3 y 9 de enero de 2022, el complejo volcánico nevados del Chillán ha presentado 19 episodios de actividad volcánica con sismos asociados al fracturamiento de roca, los cuales pueden o no presentar actividad superficial (emisión de material piroclástico).

A continuación responda las siguientes tres preguntas:

Pregunta 8

Bajo el supuesto que que actualmente los episodios mantienen la frecuencia observada y que ocurren según un proceso de Poisson, ¿cuál es la probabilidad que el tiempo transcurrido entre tres episodios supere las `[t]` horas?

Solución

Definamos como X_t al número de episodios en t horas, el cual distribuye $\text{Poisson}\left(\frac{19}{7 \times 24} \cdot t\right)$ según información descrita anteriormente.

Si T describe el tiempo, en horas, transcurrido entre tres episodios Poisson, entonces

$$T \sim \text{Gamma}\left(k = 2,\ \nu = \frac{19}{7 \times 24}\right) \quad \textbf{[0.5 Puntos]}$$

Se pide

$$P(T > \texttt{[t]}) = \texttt{1 - pgamma(t, shape = 2, rate = 19/(7*24))} \quad \textbf{[0.5 Puntos]}$$

Por ejemplo,

```
nu = 19/(7*24)
t = 10
1 - pgamma(t, shape = 2, rate = nu)
[1] 0.6877132
```

Nota: Respuesta con una diferencia de ± 0.0005 *asignar puntaje completo. Si responde correctamente con k igual a 3 asignar 0.8 puntos*

Pregunta 9

Hasta el momento el `[p]` % de los episodios presentan actividad superficial. Si esta frecuencia permanece constante, ¿cuál es la probabilidad, a partir de ahora, que se observe actividad superficial por 2da vez después del décimo episodio sísmico?

Solución

Definamos como X el número de episodios observados desde ahora hasta la ocurrencia del 2do que presente actividad superficial, el cuál se comporta como una variable aleatoria

$$X \sim \text{Bin-Neg}(k = 2,\, p = \texttt{[p]}/100). \quad \textbf{[0.5 Puntos]}$$

Se pide

$$P(X > 10) = 1 - P(X \leq 10) = \texttt{1 - pnbinom(10-2, size = 2, prob = p /100)} \quad \textbf{[0.5 Puntos]}$$

Por ejemplo,

```
p = 0.20
k = 2
1-pnbinom(10-k, size = k, prob = p)
[1] 0.3758096
```

Alternativamente podemos definir Y como el número de episodios con actividad superficial entre los próximos 10. En este caso la variable aleatoria se comporta como

$$Y \sim \text{Binomial}(n = 10,\, p = \texttt{[p]}/100). \quad \textbf{[0.5 Puntos]}$$

Se pide $P(Y \leq 1) =$ `pbinom(2-1, size = 10, prob = p/100)`. **[0.5 Puntos]**

Por ejemplo,

```
p = 0.20
k = 2
pbinom(k-1, size = 10, prob = p)
[1] 0.3758096
```

Nota: Respuesta con una diferencia de ± 0.0005 *asignar puntaje completo.*

Pregunta 10

Hasta el momento la altura que alcanza la columna de material para cada episodio presentan una mediana iguala 800 metros con un coeficiente de variación de `[delta]` %. Bajo el supuesto que la altura se comporta como una variable aleatoria Log-Normal, ¿cuál es la probabilidad que la próxima columna de material supere el `[x]` kilómetros?

Solución

Sea X la altura, en kilómetros, que alcanza la columna de material durante un episodio, la cual según enunciado se comporta como una variable aleatoria Log-Normal(λ, ζ).

A partir de la mediana y coeficiente de variación se tiene que

$$\textbf{[0.2 Puntos]} \quad \lambda = \ln(800/1000) \quad \text{y} \quad \zeta = \sqrt{\ln(1 + (\texttt{[delta]}/100)^2)} \quad \textbf{[0.2 Puntos]}$$

Se pide

$$P(X > \texttt{[x]}) = \texttt{1 - plnorm(x, meanlog = } \lambda\texttt{, sdlog = } \zeta\texttt{)} \quad \textbf{[0.6 Puntos]}$$

Por ejemplo,

```
x50 = 800
delta = 35
lambda = log(x50/1000)
zeta = sqrt(log(1+(delta/100)^2))
1-plnorm(1.3, meanlog = lambda, sdlog = zeta)
[1] 0.07661405
```

Nota: Respuesta con una diferencia de ± 0.0005 *asignar puntaje completo.*

Indice Nasdaq

NASDAQ (National Association of Securities Dealers Automated Quotation) es el segundo mercado de valores y bolsa de valores automatizada y electrónica más grande de los Estados Unidos, siendo la primera la Bolsa de Nueva York, con más de 3800 compañías y corporaciones.

La base de datos `NASDAQ.xlsx` contiene la información bursátil del ultimo año y a partir de ella responda las siguientes dos preguntas:

```
Base = rio::import("NASDAQ.xlsx")
```

Pregunta 11

El retorno diario se define como la diferencia relativa entre el valor de cierre de hoy con respecto al día anterior, la cual ya se encuentra en la base de datos en la columna Return". Se pide ajustar a partir de la media y desviación estándar de los retornos, los parámetros correspondientes a una distribución Normal y Logistica.

¿Cuál de estos dos modelos ajusta mejor la probabilidad empírica (%) de retornos diarios superiores a 0.01 (rentabilidad del 1%)?

Probabilidad Normal = `0.2061036` **[0.4 Puntos]**

Probabilidad Logistica = `0.1843207` **[0.4 Puntos]**

Mejor ajuste = LOGISTICA **[0.2 Puntos]**

Solución

```
X = Base$Return
mu = mean(X)
sigma = sd(X)
1 - pnorm(0.01, mean = mu, sd = sigma)
[1] 0.2061036
sigma = sd(X)*sqrt(3)/pi
1 - plogis(0.01, location = mu, scale = sigma)
[1] 0.1843207
mean(X > 0.01)
[1] 0.173913
```

Pregunta 12

Considere la variable "Volume" en millones, es decir, divida previamente la variable por un millón.

A partir de los percentiles `[p]` % y `[q]` % ajuste los parámetros de una distribución Weibull y calcule la probabilidad que el "Volume" diario supere los 6 mil millones.

Solución

Se X la variable "Volume" en millones a la cuál se debe ajustar una distribución Weibull(η, β) a partir de x_p y x_q.

Del formulario se tiene que

$$\ln(x_p) = \ln(\eta) + \frac{1}{\beta}\,\ln(-\ln(1 - [\text{p}]/100))$$

y

$$\ln(x_q) = \ln(\eta) + \frac{1}{\beta}\,\ln(-\ln(1 - [\text{q}]/100))$$

Despejando los parámetros se tiene que

$$\beta = \frac{\ln(-\ln(1 - \texttt{[p]}/100)) - \ln(-\ln(1 - \texttt{[q]}/100))}{\ln(x_p) - \ln(x_q)} \quad \textbf{[0.4 Puntos]}$$

y

$$\eta = \exp\left\{\ln(x_p) - \frac{1}{\beta}\,\ln(-\ln(1 - [\text{p}]/100))\right\} \quad \textbf{[0.4 Puntos]}$$

Se pide

$$P(X > 6000) = \texttt{1 - pweibull(600, shape = }\beta\texttt{, scale = }\eta\texttt{)} \quad \textbf{[0.2 Puntos]}$$

Por ejemplo

```
Y = Base$Volume/1000000
p = 50
q = 60
beta = (log(-log(1-p/100))-log(-log(1-q/100)))/(log(quantile(Y,p/100))-log(quantile(Y,q/100)))
eta = exp(log(quantile(Y,p/100))-(1/beta)*log(-log(1-p/100)))
1 - pweibull(6000, shape = beta, scale = eta)
[1] 0.1138506
```

Nota: Respuesta con una diferencia de ± 0.0005 *asignar puntaje completo.*