Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2021

Interrogación 2 - Pauta

Pregunta 1

Pese a las advertencias que la Comisión Médica de la ANFP envió hace unas semanas, con respecto a la prevención de contagios por COVID, algunos equipos profesionales fallaron en su cumplimiento. Hasta el momento en un equipo hay 100 personas entre jugadores, cuerpo técnico, staff medico y dirigentes en cuarentena preventiva por ser contacto estrecho hasta mañana miércoles.

Suponga que un test de antígenos clasifica como Covid+ a una persona con probabilidad `[p]` un día antes de finalizar la cuarentena preventiva.

¿Cuál es la probabilidad que hoy, después de realizar un examen de antígenos, al menos `[x]` de los que se encontraban en cuarentena preventiva sean clasificados como Covid+?

Solución

Sea X el número Covid+ que el test de antígenos reporto hoy entre los 100 que se encontraban en cuarentena preventiva.

$$X \sim \text{Bernoulli}(n = 100, p = \texttt{[p]})$$

Se pide $P(X \geq \texttt{[x]}) =$ `1 - pbinom(x-1, size = 100, prob = p)`.

```
## EJEMPLO:
n = 100
p = 0.03
x = 5
1 - pbinom(x-1, size = n, prob = p, lower.tail = T)
[1] 0.1821452
pbinom(x-1, size = n, prob = p, lower.tail = F)
[1] 0.1821452
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Responder correctamente el complemento* ± 0.005 *[0.5 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 2

Al finalizar esta evaluación, los alumnos procederán a escanear, guardar y subir sus respaldos.

Suponga que el tiempo que les tomará hacer todo esto se comporta como una variable aleatoria Log-Normal, con un tiempo mediano de `[x50]` minutos y un coeficiente de variación del [delta] %.

¿Cuál es la probabilidad que un alumno, al que se le cierra la prueba debido a que terminó el tiempo, no requiera mandar un mail acusando problemas técnicos por que no alcanzó a subir sus respaldos?

Solución

Sea X al tiempo que le toma a un alumno escanear, guardar y subir sus respaldos desde que se cierra la prueba.

$$X \sim \text{Log-Normal}(\lambda,\, \zeta)$$

Se pide

$$P(X \leq 15) = \text{plnorm}(15,\ \text{meanlog} = \text{lambda},\ \text{sdlog} = \text{zeta}),$$

ya que en las indicaciones de la prueba se establece que el cierre de la prueba es a las 20.45 horas y que la carga de archivos era hasta las 21.00 horas.

Del enunciado se tiene que

$$\lambda = \ln(\texttt{[x50]}) \quad \text{y} \quad \zeta = \sqrt{\ln\left(1 + \left(\frac{\texttt{[delta]}}{100}\right)^2\right)}$$

```
## EJEMPLO:
x50 = 7
delta = 35
lambda = log(x50)
zeta = sqrt(log(1+(delta/100)^2))
plnorm(15, meanlog = lambda, sdlog = zeta, lower.tail = T)
[1] 0.9875191
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 3

Durante las apelaciones de 1ra instancia, a cada ayudante se le asigna una sola pregunta a revisar, para mantener la misma mano en este proceso de revisión.

Suponga que una pregunta en particular, la probabilidad que una apelación logre puntaje completo es igual a `[p]`.

¿Cuál es la probabilidad que el ayudante asigne puntaje completo por 3ra vez, después de vigésima revisión en esa pregunta?

Solución

Sea X el número de pruebas revisadas hasta observar por 3ra vez un caso que amerita puntaje completo.

$$X \sim \text{Bin-Neg}(k = 3, p = \texttt{[p]})$$

Se pide $P(X > 20) =$ `1 - pnbinom(20-3, size = 3, prob = p)`.

```
## EJEMPLO:
p = 0.07
pnbinom(20-3, size = 3, prob = p, lower.tail = F)
[1] 0.8389971
1-pnbinom(20-3, size = 3, prob = p, lower.tail = T)
[1] 0.8389971
```

Sea Y el número de pruebas que ameritan puntaje completo entre las primeras 20 revisiones.

$$Y \sim \text{Binomial}(n = 20, p = \texttt{[p]})$$

Se pide $P(Y \leq 2) =$ `pbinom(3-1, size = 20, prob = p)`.

```
## EJEMPLO:
p = 0.07
pbinom(3-1, size = 20, prob = p, lower.tail = T)
[1] 0.8389971
1-pbinom(3-1, size = 20, prob = p, lower.tail = F)
[1] 0.8389971
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Responder correctamente el complemento* ± 0.005 *[0.5 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 4

Estudios muestran que, en espacios cerrados y expuestos al virus por varios minutos, un `[p]` % de las personas se contagian de Covid19.

Suponga que durante una evaluación presencial un total de `[N]` alumnos con antígenos negativos, son expuestos (involuntariamente) al contagio, dado que uno de los cuidadores esta contagiado (y se ha paseado por todas las salas).

Una vez concluida la evaluación, una muestra de `[n]` alumnos son sometidos a un test de antígenos certero.

Determine la probabilidad que al menos tres alumnos se hayan contagiados de Covid19

Solución

Sea X el número de contagiados en la muestra.

$$X \sim \text{Hipergeometrica}(n = \texttt{[n]}, N = \texttt{[N]}, m = \texttt{[N]} \cdot \texttt{[p]}/100)$$

Se pide $P(X \geq 3) =$ `1 - phyper(2, m = N * p/100, n = N - N * p/100, k = n)`.

```
## EJEMPLO:
N = 150
p = 30
m = N * p/100
n = 12
1-phyper(2, m = m, n = N-m, k = n, lower.tail = T)
[1] 0.7580152
phyper(2, m = m, n = N-m, k = n, lower.tail = F)
[1] 0.7580152
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Responder correctamente el complemento* ± 0.005 *[0.5 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 5

El número de correos que llegan diariamente al mail UC, se comporta como una variable aleatoria Poisson(`[lambda]`). Sin embargo, un `[p]` % de ellos corresponden a Spam.

Determine el tiempo esperado (en horas) entre cuatro Spam.

Solución

Sea X_t el número de Spam que llegan en t horas y T el tiempo, en horas, entre cuatro Spam.

Del enunciado se tiene que

$$X_t \sim \text{Poisson}\left(\frac{\texttt{[lambda]} \cdot (\texttt{[p]}/100)}{24} \cdot t\right) \quad \text{y} \quad T \sim \text{Gamma}\left(k = 3,\, \nu = \frac{\texttt{[lambda]} \cdot (\texttt{[p]}/100)}{24}\right)$$

Se pide $\text{E}(T) = \dfrac{k}{\nu}$.

```
## EJEMPLO:
k = 3
lambda = 40
p = 29
nu = lambda*(p/100)/24
k/nu
[1] 6.206897
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Respuesta correcta considerando Gamma(4, ν) y margen de error* ± 0.005 *[0.5 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 6

Un estudio realizado a partir del padrón histórico automotriz muestra que el tiempo, en meses, entre inscripciones de vehículos que realizan RUT asociados a personas naturales que son parte de este padrón, se comporta como una variable aleatoria Weibull, cuya mediana es de `[x50]` meses. Además, se sabe que solo al 20 % de ellos les toma más de `[x80]` meses volver a inscribir un nuevo vehículo.

Determine el IQR de los tiempos, en meses, entre inscripciones.

Solución

Sea T el tiempo, en meses, entre inscripciones de vehículos que realiza un RUT asociados a una personas natural que es parte del padrón actual.

$$T \sim \text{Weibull}(\eta,\, \beta)$$

Se pide

$$\text{IQR} = \texttt{qweibul(0.75, shape = } \beta\texttt{, scale = } \eta\texttt{) - qweibul(0.25, shape = } \beta\texttt{, scale = } \eta\texttt{).}$$

Del formulario y enunciado se tiene que

$$\ln(\texttt{[x50]}) = \ln(\eta) + \frac{1}{\beta} \cdot \ln\left(-\ln(1 - 0.50)\right)$$
$$\ln(\texttt{[x80]}) = \ln(\eta) + \frac{1}{\beta} \cdot \ln\left(-\ln(1 - 0.80)\right)$$

Resolviendo el sistema de ecuaciones:

$$\beta = \frac{\ln\left(-\ln(1 - 0.80)\right) - \ln\left(-\ln(1 - 0.50)\right)}{\ln(\texttt{[x80]}) - \ln(\texttt{[x50]})} \quad \text{y} \quad \eta = \exp\left\{\ln(\texttt{[x80]}) - \frac{1}{\beta} \cdot \ln(-\ln(1 - 0.80))\right\}$$

```
## EJEMPLO:
x50 = 45
x80 = 80
beta = (log(-log(1-0.80))-log(-log(1-0.50)))/(log(x80)-log(x50))
eta = exp(log(x80)-(1/beta)*log(-log(1-0.80)))
qweibull(0.75, shape = beta, scale = eta)-qweibull(0.25, shape = beta, scale = eta)
[1] 47.56571
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 7

Suponga que X e Y son dos variables aleatorias discretas con la siguiente función de probabilidad conjunta:

$$p_{X,Y}(x,y) = \frac{n!}{y!\,(x-y)!\,(n-x)!}\left(\frac{p}{1-p}\right)^x \left(\frac{q}{1-q}\right)^y (1-p)^n(1-q)^x,$$

para $x \in \mathbb{N}_0$, $y \in \mathbb{N}_0$, $0 \leq y \leq x \leq n$, $0 < p < 1$ y $0 < q < 1$.

Si $n =$ `[n]`, $p =$ `[p]` y $q =$ `[q]`, calcule $F_Y(3)$.

Solución

Por teorema de probabilidades totales se tiene que

$$p_Y(y) = \sum_{x=y}^{n} \frac{n!}{y!\,(x-y)!\,(n-x)!}\left(\frac{p}{1-p}\right)^x \left(\frac{q}{1-q}\right)^y (1-p)^n(1-q)^x$$

y se pide $F_Y(3) = p_Y(0) + p_Y(1) + p_Y(2) + p_Y(3)$.

Alterativa 1: Aplicando el teorema de probabilidades totales directamente en R o Wolfram.

```
## EJEMPLO:
p = 0.15
q = 0.30
n = 60

y = 0
x = y:n
p0 = sum(factorial(n)*(p/(1-p))^x*(q/(1-q))^y*(1-p)^n*(1-q)^x/(factorial(y)*factorial(x-y)*factorial(n-x)))

y = 1
x = y:n
p1 = sum(factorial(n)*(p/(1-p))^x*(q/(1-q))^y*(1-p)^n*(1-q)^x/(factorial(y)*factorial(x-y)*factorial(n-x)))

y = 2
x = y:n
p2 = sum(factorial(n)*(p/(1-p))^x*(q/(1-q))^y*(1-p)^n*(1-q)^x/(factorial(y)*factorial(x-y)*factorial(n-x)))

y = 3
x = y:n
p3 = sum(factorial(n)*(p/(1-p))^x*(q/(1-q))^y*(1-p)^n*(1-q)^x/(factorial(y)*factorial(x-y)*factorial(n-x)))

p0+p1+p2+p3
[1] 0.715676
```

Alterativa 2: Mostrando previamente que $Y \sim$ Benoulli(`[n]`, `[p]` $\cdot$ `[q]`) y luego utilizar `pbinom(...)`.

```
## EJEMPLO:
p = 0.15
q = 0.30
n = 60

pbinom(3, size = n, prob = p * q, lower.tail = T)
[1] 0.715676
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*

- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 8

Una nueva tienda de un centro comercial, que abre los siete días de la semana, promociona compras 2×1 mediante cupones que se reparten en los pasillos.

Suponga que los clientes ingresan a la tienda según un proceso de Poisson con tasa esperada de 3 clientes cada 10 minutos y que cada cliente posee un cupón 2×1 con probabilidad `[p]`.

¿Cuál es la probabilidad que en una hora ingresen al menos `[x]` clientes con derecho a un 2×1?

Solución

Sea X_t el número de clientes que ingresan en t minutos a la tienda con derecho a un 2×1.

$$X_t \sim \text{Poisson}\left(\lambda = \frac{3}{10} \cdot \texttt{[p]} \cdot t\right)$$

Se pide $P(X_{60} \geq \texttt{[x]}) =$ `1 - ppois(x-1, lambda = 3*p*60/10, lower.tail = T)`.

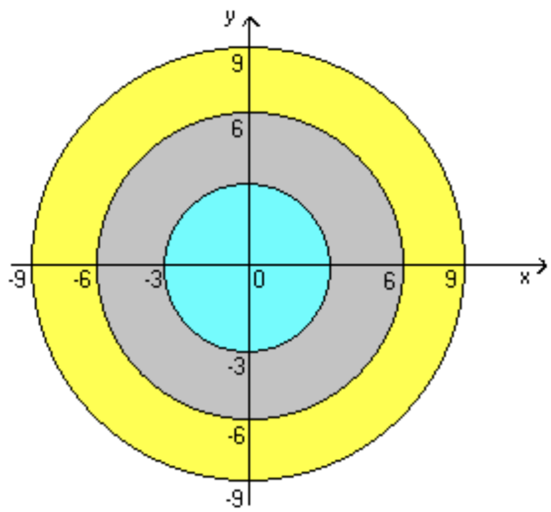
```
## EJEMPLO:
p = 0.15
nu = 3/10
t = 60
lambda = nu*t*p
x = 7
1-ppois(x-1, lambda = lambda, lower.tail = T)
[1] 0.02056945
ppois(x-1, lambda = lambda, lower.tail = F)
[1] 0.02056945
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Responder correctamente el complemento* ± 0.005 *[0.5 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 9

En un juego del tiro al blanco, se debe lanzar un dardo hacia un tablero circular. Consideremos el siguiente tablero con radios 3,6 y 9 cm, como muestra la siguiente figura:



Suponga que un jugador lanza un dardo y cae en las coordenadas X e Y del plano, las cuales pueden ser consideradas como variables aleatorias independientes con distribución Uniforme(-9, +9).

Si el jugador al lanzar cae dentro del tablero, ¿cuál es la probabilidad de que haya caído en el anillo gris?

Solución

Por independencia se tiene que

$$f_{X,Y}(x,y) = \frac{1}{18} \cdot \frac{1}{18}$$

Se pide $P(A \,|\, B) = \dfrac{P(A \cap B)}{P(B)}$, con A el evento caer en zona gris y B el evento caer en el tablero.

Como la densidad conjunta es constante se tiene que las probabilidades es simplemente el área multiplicada por la densidad:

$$P(A \cap B) = \frac{\pi \cdot 6^2 - \pi \cdot 3^2}{18^2} \quad \text{y} \quad P(B) = \frac{\pi \cdot 9^2}{18^2}$$

Por lo tanto

$$P(A \,|\, B) = \frac{1}{3}$$

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

El archivo [PSU2019.xlsx] contiene 11.312 registros correspondiente a establecimientos particulares pagados de la Región Metropolitana. La información disponible es la siguiente:

- $SEXO$: M (masculino) o F (femenino)
- PSU_L: Puntaje prueba lenguaje.
- PSU_M: Puntaje prueba matemáticas.
- PSU_H: Puntaje prueba historia.
- PSU_C: Puntaje prueba ciencias.
- $RANKING$: Puntaje trayectoria enseñanza media.

Esta base de datos tiene bastantes registros faltantes y a modo de ejemplo algunas funciones que podrían ser de utilidad son:

- `mean(X, na.rm = T)`: Omite los NA's al calcular un promedio.
- `is.na(X)`: Función lógica que entrega TRUE si el valor es un dato faltante y FALSE en caso contrario.
- `c(na.omit(X))`: Si X es un vector, elimina los datos faltantes, si agrega `c()` dejará al vector sin atributos que podrían genera algún problema al correr una función
- `na.omit(Data)`: Si Data es un data frame (Base de Datos), esta función elimina todas las filas que tienen al menos un registro faltante.

```
Base <- rio::import("psu2019.xlsx")
```

Pregunta 10

Suponga que los puntajes obtenido en matemática se comportan como una variable aleatoria Log-Normal.

Basado en los percentiles [p] % y [q] %, calcule la probabilidad que un alumno cualquiera obtenga más de [x] puntos en la prueba PSU matemática.

Solución

```
## EJEMPLO:
X = Base$PSU_M
p = 0.25
q = 0.85
xp = quantile(X, p, na.rm = T)
xq = quantile(X, q, na.rm = T)
zeta = (log(xp)-log(xq))/(qnorm(p)-qnorm(q))
lambda = log(xp)-zeta*qnorm(p)
x = 690
plnorm(x, meanlog = lambda, sdlog = zeta, lower.tail = F)
[1] 0.258903
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 11

Suponga que los puntajes en la prueba de lenguaje y ranking de enseñanza media se comportan según una Normal Bivariada.

Considerando solo los registros que tienen información en ambos casos, calcule la probabilidad que un alumno cualquiera obtenga sobre `[x]` puntos en Ranking y Lenguaje al mismo tiempo.

Solución

```
## EJEMPLO:
X = Base$PSU_L
Y = Base$RANKING
Data = na.omit(cbind(X,Y))
mu = apply(Data, 2, mean)
sigma = cov(Data)
x = 710
mvtnorm::pmvnorm(lower = c(x,x), upper = c(Inf, Inf), mean = mu, sigma = sigma)[1]
[1] 0.09827398
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*

Pregunta 12

Suponga que los puntajes en la prueba de lenguaje y ciencia se comportan según una Normal Bivariada. Considerando solo los registros que tienen información en ambos casos, calcule la probabilidad que un alumno obtenga más de `[x]` puntos en lenguaje, dado que presenta `[y]` puntos en ciencias.

Solución

```
## EJEMPLO:
X = Base$PSU_L
Y = Base$PSU_C
Data = as.data.frame(na.omit(cbind(X,Y)))
mu.x = mean(Data$X)
mu.y = mean(Data$Y)
sigma.x = sd(Data$X)
sigma.y = sd(Data$Y)
rho = cor(Data$X, Data$Y)
x = 680
y = 790
mu = mu.x+(y - mu.y)*sigma.x*rho/sigma.y
sigma = sigma.x*sqrt(1-rho^2)
pnorm(x, mean = mu, sd = sigma, lower.tail = F)
[1] 0.8188135
```

Puntaje:

- *Respuesta correcta con margen de error* ± 0.005 *[1.0 puntos].*
- *Sin respuesta en canvas, pero respaldo correcto* ± 0.005 *[0.5 puntos].*
- *Para otros casos correctamente justificados por parte del alumno, los ayudantes podrán asignar máximo [0.5 puntos].*