Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Temporada Académica de Verano (enero 2021)

Interrogación 2 - Pauta

Pregunta 1 [1.0 Punto]

Como han de saber, el puntaje de la nueva PDT (ex PSU) se puede modelar de acuerdo a una distribución Normal. Usted orgulloso de su establecimiento de origen, revisa antecedentes de los egresados que dieron la PDT-Matemáticas y obtiene que el 75 % obtuvo 650 o menos puntos, y un rápido análisis muestra que el coeficiente de variación (c.o.v.) es de un 25 %.

¿Cuál es la probabilidad que un alumno de su establecimiento que rindió la PDT obtenga más de 700 puntos en matemáticas?

Solución

Tenemos que

$$x_{0.75} = \mu + \sigma \cdot \Phi^{-1}(0.75) = 650 \qquad \text{y} \qquad \sigma = 0.25 \cdot \mu$$

En R, el script es:

```
  x75 = 650
   mu = x75 / (1 + 0.25 * qnorm(0.75))
sigma = 0.25 * mu
1 - pnorm(700, mean = mu, sd = sigma)
[1] 0.1505527
```

Pregunta 2 [1.0 Punto]

La información respecto al proceso de vacunación muestra niveles altos de adherencia en Chile. Sin embargo, en países europeos han tenido problemas dado el alto nivel de rechazo a vacunarse.

Datos muestran que, en Francia, el 55 % de los adultos mayores de residencias geriátricas han rechazado vacunarse.

Suponga que un TENS ("vacunador") concurre a una residencia geriátrica en Francia, donde residen 35 adultos mayores, con solo 12 dosis.

¿Cuál es la probabilidad de lograr vacunar a todos los adultos mayores residentes que lo desean. Suponga que cada adulto mayor decide vacunarse o no de manera independiente.

Solución

Tenemos que $p = 0.45$, representa la probabilidad que un adulto mayor en Francia NO rechace vacunarse.

Definamos como X el número de adultos mayores en la residencia que quieren vacunarse y por la independencia entre los adultos en su decisión de vacunarse o no, entonces

$$X \sim \text{Binomial}(n = 35,\, p = 0.45)$$

Se pide $P(X \leq 12)$.

En R, el script es:

```
p = 0.45
pbinom(12, size = 35, prob = p)
[1] 0.1343571
```

Pregunta 3 [1.0 Punto]

Un laboratorio que trabaja 24×7 recibe muestras de PCR para su análisis. Suponga que las últimas 1.000 muestras analizadas presentaron una positividad del 8 %.

¿Cuál es la probabilidad qué al revisar 20 resultados entre las últimas mil muestra analizadas, al menos 3 arrojen PCR positivo?

Solución

Tenemos una población de 1000 muestras últimamente analizadas, de las cuales 80 son PCR positivo

Al revisar 20 entre las últimas mil muestras analizadas, se tiene que el número de PCR positivo X en la muestra distribuye:

- Hipergeométrica($n = 20$, $N = 1000$, $m = 80$) si se asume un muestreo sin reemplazo.
- Binomial($n = 20$, $p = 0.08$) si asume un muestreo con reemplazo.

Se pide $P(X \geq 3)$.

En R, el script es:

```
N = 1000
m = N * 0.08
n = 20
1-phyper(2, m = m, n = N-m, k = n)
[1] 0.2107529
1-pbinom(2, size = n, prob = m/N)
[1] 0.2120538
```

Pregunta 4 [1.0 Punto]

Suponga que el proceso de recepción de muestras para PCR en un laboratorio durante el día (24×7) sigue proceso de Poisson con coeficiente de variación del 12 %.

¿Cuál es la probabilidad que durante hoy viernes 15 de enero se obtengan al menos 10 muestras positivas? Suponga que la positividad actual es de 8 %.

Solución

Tenemos que el número de muestras X que llegan en un día se comporta como una variable aleatoria Poisson$\left(\lambda = \left(\frac{1}{0.12}\right)^2\right)$.

Si definimos Y el número de muestras PCR positivo, entonces

$$Y \sim \text{Poisson}(\lambda \cdot p)$$

con $p = 0.08$.

Se pide $P(Y \geq 10)$.

En R, el script es:

```
lambda = (1/0.12)^2
     p = 0.08
1-ppois(9, lambda = lambda*p)
[1] 0.05671
```

También se podría definir T_{10} el tiempo (en días) hasta la décima muestra con PCR positivo.

$$T_{10} \sim \text{Gamma}(k = 10, \nu = \lambda \cdot p)$$

En este caso se pide $P(T_{10} < 1)$.

En R, el script es:

```
lambda = (1/0.12)^2
     p = 0.08
pgamma(1, shape = 10, rate = lambda * p)
[1] 0.05671
```

Pregunta 5 [1.0 Punto]

Diversos estudios muestran que los alumnos que rinden la PDT (ex PSU) la desarrollan en forma lineal, donde el tiempo utilizado en cada pregunta se comporta como una variable aleatoria Log-Normal. Suponga que el tiempo medio asignado a cada una de las 10 preguntas de probabilidad y estadística es de 4 min con una desviación estándar de 2 min. Determine la probabilidad que a lo más en dos de ellas el tiempo utilizado sea superior a 7 min. Asuma que cada pregunta se responde de manera independiente.

Solución

Definamos como T el tiempo que toma responder una pregunta de probabilidad y estadística en la PDT.

$$T \sim \text{log-Normal}(\lambda,\, \zeta)$$

con

$$\zeta = \sqrt{\ln(1 + \delta_X^2)} \quad \text{y} \quad \lambda = \ln(\mu_X) - 0.5 \cdot \zeta^2$$

Definamos X el número de preguntas que toma más de 7 min entre las 10 que aparece en la PDT.

$$X \sim \text{Binomial}(n = 10,\, p)$$

con $p = P(T > 7)$ y se pide $P(X \leq 2)$.

En R, el script es:

```
   mu.x = 4
sigma.x = 2
delta.x = sigma.x/mu.x
   zeta = sqrt(log(1+delta.x^2))
 lambda = log(mu.x)-0.5*zeta^2
      p = 1 - plnorm(7, meanlog = lambda, sdlog = zeta)
pbinom(2, size = 10, prob = p)
[1] 0.9628481
```

Pregunta 6 [1.0 Punto]

Suponga que el tiempo en recorrer una distancia en vehículo, condicionado al número de semáforos en rojo que se le presentan en el trayecto distribuye Gamma. Si el número de semáforos en rojo que se le presentan en el trayecto distribuye Binomial-Negativa . Calcule el coeficiente de variación, incondicional al número de semáforos en rojo, del tiempo que toma en recorrer una distancia en vehículo. Utilice $\nu = [\texttt{nu}]$, $p = [\texttt{p}]$ y $k = [\texttt{k}]$.

Solución

Tenemos que

$$X \sim \text{Binomial-Negativa}(k,\, p) \quad \text{y} \quad T \,|\, X = x \sim \text{Gamma}(x,\, \nu)$$

Se pide δ_T, para lo cual obtendremos en una primera instancia $f_T(t)$:

$$\begin{aligned}
f_T(t) &= \sum_{x=k}^{\infty} \binom{x-1}{k-1} (1-p)^{x-k}\, p^k \cdot \frac{\nu^x}{\Gamma(x)}\, t^{x-1}\, e^{-\nu\, t} \\
&= \sum_{x=k}^{\infty} \frac{(x-1)!}{(k-1)!\,(x-k)!} (1-p)^{x-k}\, p^k \cdot \frac{\nu^x}{(x-1)!}\, t^{x-1}\, e^{-\nu\, t} \\
&= \frac{(\nu\, p)^k\, t^{k-1}\, e^{-\nu\, t}}{(k-1)!} \sum_{y=0}^{\infty} \frac{[(1-p)\,\nu\, t]^y}{y!} \\
&= \frac{(\nu\, p)^k\, t^{k-1}\, e^{-\nu\, t}}{(k-1)!} \cdot e^{\nu\, t - \nu\, p\, t} \\
&= \frac{(\nu\, p)^k}{\Gamma(k)}\, t^{k-1}\, e^{-\nu\, pt}, \quad t > 0
\end{aligned}$$

Como

$$T \sim \text{Gamma}(k,\, \nu\, p) \to \delta_T = \frac{1}{\sqrt{k}}$$

En R, el script es:

```
## Ejemplo
k = 3
1/sqrt(k)
[1] 0.5773503
```

Pregunta 7 [1.0 Punto]

Suponga que durante el periodo de apelación de 1ra instancia de la nota I_2, el número de solicitudes que llegarán distribuirá Binomial(n, p) y que el número de solicitudes que subirán su calificación entre las apelaciones recibidas también distribuye Binomial con probabilidad de éxito q.

Evalúe la covarianza entre el número de solicitudes y el número de solicitudes que suben su calificación para $n =$ [n], $p =$ [p] y $q =$ [q].

Solución

Definamos X al número de solicitudes que llegaran e Y al número de solicitudes que suben su calificación.

$$X \sim \text{Bimomial}(n,\, p) \quad \text{e} \quad Y|X = x \sim \text{Binomial}(x,\, q)$$

Se pide $\text{Cov}(X, Y)$.

Por formulario se tiene que

$$\text{E}(X) = n\,p \quad \text{y} \quad \text{Var}(X) = n\,p\,(1-p)$$

y al aplicar teorema de esperanzas iterada:

$$\begin{aligned} \text{E}(Y) &= \text{E}[\text{E}(Y \mid X)] = n\,p\,q \\ \text{E}(X \cdot Y) &= \text{E}[\text{E}(Y \cdot X \mid X)] = n\,p\,q\,(1 - p + n\,p) \end{aligned}$$

Luego

$$\text{Cov}(X, Y) = \text{E}(X \cdot Y) - \text{E}(X) \cdot \text{E}(Y) = n\,p\,q\,(1-p)$$

En R, el script es:

```
## Ejemplo
n = 111
q = 0.43
p = 0.45
n*q*p*(1-p)
[1] 11.81318
```

Pregunta 8 [1.0 Punto]

Considere las siguientes variables aleatorias continuas e independientes:

$$X \sim \text{Exponencial}(\nu) \quad \text{e} \quad Y \sim \text{Gamma}(2,\, \nu)$$

- Calcule $P(X > Y)$.
- Calcule $P(X < Y)$.

Solución

$$\begin{aligned}
P(X > Y) &= \int_0^\infty \int_y^\infty \nu\, e^{-\nu\, x} \cdot \nu^2\, y\, e^{-\nu\, y}\, dx\, dy \\
&= \int_0^\infty \nu^2\, y\, e^{-\nu\, y} \cdot \left\{ \int_y^\infty \nu\, e^{-\nu\, x}\, dx \right\} dy \\
&= \int_0^\infty \nu^2\, y\, e^{-\nu\, y} \cdot e^{-\nu\, y}\, dy \\
&= \frac{1}{4} \int_0^\infty (2\,\nu)^2\, y\, e^{-2\,\nu\, y}\, dy \\
&= \frac{1}{4} \cdot 1, \quad \text{por área bajo todo el soporte de una Gamma}(2,\, 2\,\nu) \\
&= 0.25 \\
P(X < Y) &= \int_0^\infty \int_0^y \nu\, e^{-\nu\, x} \cdot \nu^2\, y\, e^{-\nu\, y}\, dx\, dy \\
&= \int_0^\infty \nu^2\, y\, e^{-\nu\, y} \cdot \left\{ \int_0^y \nu\, e^{-\nu\, x}\, dx \right\} dy \\
&= \int_0^\infty \nu^2\, y\, e^{-\nu\, y} \cdot \left\{ 1 - e^{-\nu\, y} \right\} dy \\
&= \int_0^\infty \nu^2\, y\, e^{-\nu\, y}\, dy - \frac{1}{4} \int_0^\infty (2\,\nu)^2\, y\, e^{-2\,\nu\, y}\, dy \\
&= 1 \cdot 1 - \frac{1}{4} \cdot 1, \quad \text{por área bajo todo el soporte de una Gamma}(2,\, \nu) \text{ y Gamma}(2,\, 2\,\nu) \text{ resp.} \\
&= 1 - \frac{1}{4} \\
&= 0.75
\end{aligned}$$

Pregunta 9 [1.0 Punto]

Suponga que el consumo eléctrico anual de una ciudad por casa durante 2019 y 2020 se modela, en conjunto, como una variable aleatoria Normal-Bivariada. Según los registros de la empresa distribuidora, en el 2019 el consumo medio anual por casa fue de 12.1 megawatts-hora (MWh) con una desviación estándar 3 MWh; en 2020, el consumo medio anual por casa fue de 16.7 MWh con una desviación estándar de 4.2 MWh.

La covarianza entre el consumo de ambos años es igual a 5.67 MWh.

Entregue la probabilidad de que una casa cualquiera en esta ciudad haya tenido un consumo eléctrico anual en 2019 bajo la media de ese año y un consumo en 2020 sobre la media de ese año.

Solución

Definamos como X al consumo anual 2019 e Y al del año 2020.

Se tiene que

$$(X, Y)' \sim \text{Normal-Bivaraida}(\mu_X,\, \mu_Y,\, \sigma_X,\, \sigma_Y,\, \rho)$$

Se pide $P(X < \mu_X,\, Y > \mu_Y)$.

En R, el script es:

```
library(mvtnorm)
   mu.x = 12.1
sigma.x = 3
   mu.y = 16.7
sigma.y = 4.2
    rho = 5.67/(sigma.x*sigma.y)
SIGMA = matrix(c(sigma.x^2, rep(rho*sigma.x*sigma.y,2), sigma.y^2),2,2)
pmvnorm(lower = c(-Inf, mu.y), upper = c(mu.x, Inf), sigma = SIGMA, mean = c(mu.x, mu.y))[1]
[1] 0.175712
```

En la industria salmonera uno de los aspectos más importantes y que es un % importante del costo del proceso de crianza de las piezas de Salmones es el alimento. Las piezas que se exportan en formato HG (sin cabeza) presentan variaciones en su peso esperado, debido principalmente a ineficiencia en la alimentación, ya que los Salmones comen hasta saciarse y eso implicará en el futuro un exceso de grasa, la cual al entrar al proceso de preparación y limpieza se desecha, afectando el margen económico y las estimaciones de los Kg comprometidos para la venta.

Se tomaron 200 muestras en un centro de cultivo, registrando en terreno el PESO fuera del agua (en Kg) y su LARGO (en cm). Posteriormente en la planta durante el procesamiento de las piezas, se registro el peso de la GRASA y de la pieza en formato HG lista para exportación.

A continuación responda las últimas tres preguntas:

```
library(readxl)
Base = as.data.frame(read_excel("IngresoPlanta.xlsx"))
```

Pregunta 10 [1.0 Punto]

Para la variable GRASA, ajuste una distribución Gamma utilizado el coeficiente de variación y media de los datos.

Calcule el error porcentual del IQR estimado vs el empírico:

$$\text{Error porcentual} = \left(\frac{\text{IQR estimado}}{\text{IQR empírico}} - 1\right) \times 100\,\%$$

Solución

Tenemos que

$$X \sim \text{Gamma}(k,\, \nu)$$

y piden estimar los parámetros a partir de las siguientes relaciones teóricas:

$$\delta_X = \frac{1}{\sqrt{k}} \quad \text{y} \quad \mu_X = \frac{k}{\nu}$$

En R, el script es:

```
    X = Base$GRASA
delta = sd(X)/mean(X)
    k = 1/delta^2
   nu = k/mean(X)
  iqr = qgamma(0.75, shape = k, rate = nu)-qgamma(0.25, shape = k, rate = nu)
iqr/IQR(X)-1
[1] -0.05033095
```

Pregunta 11 **[1.0 Punto]**

Para la variable HG, ajuste una distribución Normal utilizando los percentiles 20 % y 80 %. Calcule la probabilidad estimada que el peso en formato HG sea mayor a 3 kilos.

Solución

Tenemos que

$$X \sim \text{Normal}(\mu, \sigma)$$

Se pide $P(X > 3)$ y la estimación de los parámetros es a partir de las siguientes relaciones teóricas:

$$x_{0.20} = \mu + \sigma \cdot \Phi^{-1}(0.20) \quad \text{y} \quad x_{0.80} = \mu + \sigma \cdot \Phi^{-1}(0.80)$$

En R, el script es:

```
    X = Base$HG
  x20 = quantile(X, prob = 0.20)
  x80 = quantile(X, prob = 0.80)
sigma = (x20-x80)/(qnorm(0.20)-qnorm(0.80))
   mu = x20 - sigma * qnorm(0.20)
1-pnorm(3, mean = mu, sd = sigma)
[1] 0.2162055
```

Pregunta 12 [1.0 Punto]

Para las variables PESO y HG, obtenga los parámetros de la distribución Normal Bivariada y calcule:

$$P(\text{PESO} > 4.5, \text{HG} > 3.0)$$

Solución

En R, el script es:

```
library(mvtnorm)
      X = Base$PESO
      Y = Base$HG
   mu.x = mean(X)
   mu.y = mean(Y)
sigma.x = sd(X)
sigma.y = sd(Y)
    rho = cor(X,Y) ## También podría utilizar cov(X,Y)*sigma.x*sigma.y
SIGMA = matrix(c(sigma.x^2, rep(rho*sigma.x*sigma.y,2), sigma.y^2),2,2)
pmvnorm(lower = c(4.5, 3.0), upper = c(Inf, Inf), sigma = SIGMA, mean = c(mu.x, mu.y))[1]
[1] 0.1302922
```