Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2021

Examen - Pauta

Pregunta 1

Durante el fin de semana los profesores revisaron las apelaciones I2 de 2da instancia que aún estaban pendiente y las apelaciones I3. Como eran más de 200, se filtraron los 120 que estaban más cerca del umbral de eximición.

Suponga que entre estos, hay `[m]` que si o si alcanzaran (o sobrepasaran) el umbral de eximición.

Si los profesores, escogiendo al azar, solo alcanzaron a revisar 80 alumnos, ¿cuál es la probabilidad que hoy al menos cinco eximidos estén rindiendo el examen?

Solución

Alternativa 1: Definamos como A el evento "al menos cinco eximidos estén rindiendo el examen".

Tenemos que

$$\# S = \binom{120}{40} \quad \text{y} \quad \# \overline{A} = \sum_{x=0}^{4} \binom{\texttt{[m]}}{x} \cdot \binom{120 - \texttt{[m]}}{40 - x}$$

Por lo tanto

$$P(A) = 1 - \frac{\# \overline{A}}{\# S}$$

```
## Ejemplo
m = 10
N = 120
n = 40 ## 120-80
## P(A) = 1 - #A^c/#S
x = 0:4
1-sum(choose(m, x)*choose(N-m, n-x)/choose(N,n))
[1] 0.204076
```

[1.0 ptos.]

Alternativa 2: Definamos como X a la variable aleatoria número de eximidos en los no revisados.

$$X \sim \text{Hipergeometrica}(n = 40,\, N = 120,\, m = \texttt{[m]})$$

```
## Ejemplo
m = 10
N = 120
n = 40 ## 120-80
## P(X >= 5) = P(X > 4) = 1 - Fx(4)
1-phyper(4, m = m, n = N-m, k = n)
[1] 0.204076
```

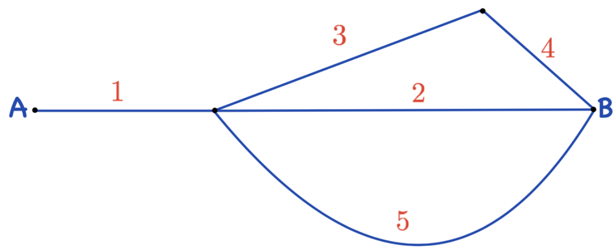
[1.0 ptos.]

Nota: Si entrega como resultado $P(X < 5)$, $P(X > 5)$ o $P(X \leq 5)$ asignar [0.5 ptos.]

Pregunta 2

Usted ha comprado recientemente un producto por e-commerce. El delivery que lleva el producto tiene 3 posibles rutas con idéntico tiempo de entrega entre A y B. Si él elige una ruta al azar entre: 1-3-4, 1-2 o 1-5, y en cada uno de los 5 tramo que componen estas las rutas, hay un probabilidad `[p]` de retrasarse, ¿cuál es la probabilidad que no cumpla con el tiempo de entrega? Suponga que los retrasos en cada tramo ocurren de manera independiente.

La siguiente Figura ilustra las tres rutas posibles:



Solución

Definamos como A el evento "no se cumple con el tiempo de entrega".

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

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

[1.0 Ptos.]

Nota: Si entrega el complemento asignar [0.5 ptos.]

Pregunta 3

Sean X y Y dos variables aleatorias independientes con distribución $\chi^2(2)$. Determine el coeficiente de variación aproximado de 1er orden de $X^{1/Y}$(X elevado a $1/Y$).

Solución

Tenemos que

$$Z = X^{1/Y} = g(X,Y) \approx \mu_X^{1/\mu_Y} + (X - \mu_X) \cdot \left(\frac{1}{\mu_Y} \cdot \mu_X^{\frac{1}{\mu_Y} - 1} \right) + (Y - \mu_Y) \cdot \left(-\mu_X^{\frac{1}{\mu_Y}} \cdot \frac{\ln(\mu_X)}{\mu_Y^2} \right)$$

Aplicando operador E() y Var() se tiene que

$$\mu_Z \approx \mu_X^{1/\mu_Y} = 1.414214$$

y

$$\sigma_Z^2 \approx \sigma_X^2 \cdot \left(\frac{1}{\mu_Y} \cdot \mu_X^{\frac{1}{\mu_Y} - 1} \right)^2 + \sigma_Y^2 \cdot \left[-\mu_X^{\frac{1}{\mu_Y}} \cdot \frac{\ln(\mu_X)}{\mu_Y^2} \right]^2 = 0.7402265 \quad \textbf{[0.4 Ptos.]}$$

Por lo tanto

$$\begin{aligned} \delta_Z &= \frac{\sigma_Z}{\mu_Z} \\ &\approx \sqrt{\sigma_X^2 \cdot \left(\frac{1}{\mu_Y \cdot \mu_X} \right)^2 + \sigma_Y^2 \cdot \left[\frac{\ln(\mu_X)}{\mu_Y^2} \right]^2} \\ &= \sqrt{4 \cdot \left(\frac{1}{2 \cdot 2} \right)^2 + 4 \cdot \left[\frac{\ln(2)}{2^2} \right]^2} \\ &= 0.6083693 \quad \textbf{[0.6 Ptos.]} \end{aligned}$$

Pregunta 4

Una institución financiera mensualmente evalúa a sus clientes y marca a algunos como pre-aprobados para créditos.

Suponga que un cliente que fue pre-aprobado durante el mes, presenta con probabilidad `[p]` sus antecedentes para acceder al crédito.

Si los clientes pre-aprobados mensualmente se comportan de acuerdo a una distribución Poisson(`[lambda]`). ¿Cuál es el coeficiente de variación del número de clientes pre-aprobados en el mes que presentarán sus antecedentes para acceder al crédito? Asuma que los clientes actúan independientemente.

Solución

Sean X los clientes pre-aprobados en el mes e Y los que presentan sus antecedentes.

$$X \sim \text{Poisson}(\texttt{[lambda]}) \quad \text{e} \quad Y \,|\, X = x \sim \text{Binomial}(n = x,\ \texttt{[p]})$$

Por teorema de probabilidades totales se tiene que

$$Y \sim \text{Poisson}(\texttt{[lambda]} \cdot \texttt{[p]}) \quad \textbf{[0.5 Ptos.]}$$

Se pide

$$\delta_Y = \frac{1}{\sqrt{\texttt{[lambda]} \cdot \texttt{[p]}}}$$

[0.5 Ptos.]

Pregunta 5

El consumo de electricidad mensual en un hogar está muy relacionado en el período de invierno con las temperaturas mínimas, debido principalmente al uso de calefacción.

Suponga que el consumo de electricidad mensual X, en kWh, y la temperatura mínima promedio de un mes Y, en °C, distribuyen conjuntamente $\text{Normal}_2\left(\mu_X = 1200, \mu_Y = 8, \sigma_X = 500, \sigma_Y = 10, \rho = -0.65\right).$

Determine la probabilidad que en un hogar el consumo sea mayor a 1500 kWh, dado se pronostica para este mes una temperatura minima promedio de `[y]`°C.

Solución

Tenemos que

$$X \,|\, Y = \texttt{[y]} \sim \text{Normal}\left(\mu_X + (\texttt{[y]} - \mu_Y) \cdot \frac{\sigma_X \cdot \rho}{\sigma_Y}, \, \sigma_X \cdot \sqrt{1-\rho^2}\right) \quad \textbf{[0.5 Ptos.]}$$

Se pide

$P(X > 1500 \,|\, Y = \texttt{[y]}) =$ `1 - pnorm(1500, mean = 1200 + (y - 8) * 500 * (-0.65)/10, sd = 500 * sqrt(1-(-0.65)^2))`

[0.5 Ptos.]

Pregunta 6

Un supervisor diariamente llega tarde a su trabajo en promedio 35 min. Suponga que el tiempo de retraso de este supervisor se comporta como una variable aleatoria Exponencial.

Descontando los días feriados y vacaciones, el año laboral de este supervisor consiste en 231 jornadas laborales.

¿Cuál es la probabilidad aproximada que dicho supervisor sume más de `[k]` horas de atraso en un año laboral? Asuma independencia entre los tiempos de retrasos y que no se ausenta ninguno de los 231 días.

Solución

Sea X el tiempo de atraso en un día laboral.

$$X \sim \text{Exponencial}(\nu = 1/35)$$

Por teorema central del límite se tiene que

$$P\left(\sum_{i=1}^{231} X_i > \texttt{[k]} \cdot 60\right) \approx \texttt{1 - pnorm(k * 60, mean = 231/(1/35), sd = sqrt(231/(1/35)^2))}$$

[1.0 Ptos.]

Nota: Si entrego el valor exacto (uso `pgamma`) asignar puntaje completo. Si respondió el complemento asignar [0.5 ptos.]

Pregunta 7

Considere una muestra aleatoria (iid) de tamaño [n] proveniente de una población Log-Normal (λ, ζ), donde λ es conocido. ¿Cuál es la probabilidad exacta que el estimador máximo verosímil de ζ sea a lo más un 80 % del valor (desconocido) de ζ?

Solución

Si aplicamos logaritmo natural a la muestra, se convierte en una muestra aleatoria iid Normal(λ, ζ), por lo cual el pivote $C_{\texttt{n}}$ basado en el estimador máximo verosímil de ζ con λ conocido

$$C_{\texttt{n}} = \frac{\texttt{[n]}\,\hat{\zeta}^2}{\zeta^2} \sim \chi^2(\texttt{[n]})$$

Se pide

$$P(\hat{\zeta} \leq 0.8 \cdot \zeta) = P(C_{\texttt{n}} \leq 0.8^2 \cdot \texttt{[n]}) = \texttt{pchisq(0.8*0.8*n, df = n)}$$

[1.0 Ptos.]

Nota: Si utilizo una $\chi^2(\texttt{[n]} - 1)$ o calculó el complemento asignar [0.5 ptos.]

El archivos BandaAncha.xlsx contiene información sobre la velocidad media de subida (en Mbps) durante periodos de alto tráfico en 50 ciudades de Chile y 4 operadores: M, E, W y C.

A continuación responda las siguientes cuatro preguntas:

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

Pregunta 8.1

Ajuste una distribución Log-Normal y Gamma por el método de momentos a las velocidades medias del operador M y a partir de un test de bondad de ajuste KS (Kolmogorov-Smirnov) responda lo siguiente:

valor-p para ajuste Log-Normal = 0.7439375 **[0.4 Ptos.]**

valor-p para ajuste Gamma = 0.3641442 **[0.4 Ptos.]**

¿Qué modelo ajusta mejor? Log-Normal **[0.2 Ptos.]**

```
X = dplyr::filter(Base, Operador == "M")$Velocidad
par = fitdistrplus::fitdist(data = X, method = "mme", distr = "lnorm")$estimate
ks.test(X, "plnorm", meanlog = par[1], sdlog = par[2])$p.value
[1] 0.7439375
par = fitdistrplus::fitdist(data = X, method = "mme", distr = "gamma")$estimate
ks.test(X, "pgamma", shape = par[1], rate = par[2])$p.value
[1] 0.3641442
```

Pregunta 8.2

Ajuste una distribución Log-Normal y Gamma por el método de máxima verosimilitud a las velocidades medias del operador E y a partir de un test de bondad de ajuste KS (Kolmogorov-Smirnov) responda lo siguiente:

valor-p para ajuste Log-Normal = 0.7949498 **[0.4 Ptos.]**

valor-p para ajuste Gamma = 0.1932944 **[0.4 Ptos.]**

¿Qué modelo ajusta mejor? Log-Normal **[0.2 Ptos.]**

```
X = dplyr::filter(Base, Operador == "E")$Velocidad
par = fitdistrplus::fitdist(data = X, method = "mle", distr = "lnorm")$estimate
ks.test(X, "plnorm", meanlog = par[1], sdlog = par[2])$p.value
[1] 0.7949498
par = fitdistrplus::fitdist(data = X, method = "mle", distr = "gamma")$estimate
ks.test(X, "pgamma", shape = par[1], rate = par[2])$p.value
[1] 0.1932944
```

Pregunta 9.1

Ajuste una distribución Weibull por el método de máxima verosimilitud a las velocidades medias del operador W y a partir de un test de bondad de ajuste χ^2 basado en los quintiles responda lo siguiente:

valor-p = 0.1194606 **[0.8 Ptos.]**

¿El modelo ajusta considerando un nivel de significancia del 10 %? SI **[0.2 Ptos.]**

Nota: Para el cálculo del estadístico de prueba, considere intervalos cerrados por la derecha: (LI - LS], excepto el 1ero que deberá también cerrarlo por la izquierda: [LI - LS].

LI: Límite inferior
LS: Límite superior

```
X = dplyr::filter(Base, Operador == "W")$Velocidad
par = fitdistrplus::fitdist(data = X, method = "mle", distr = "weibull")$estimate
x = quantile(X, prob = seq(0,1,.2))
x[1] = 0
x[6] = Inf
p = diff(pweibull(x, shape = par[1], scale = par[2]))
O = hist(X, plot = F, breaks = x, right = T)$count
O
[1] 10 10 10 10 10
X2 = chisq.test(x = O, p = p)$statistic
1-pchisq(X2, df = 5-1-2)
X-squared
0.1194606
```

Pregunta 9.2

Ajuste una distribución Weibull por el método de máxima verosimilitud a las velocidades medias del operador C y a partir de un test de bondad de ajuste χ^2 basado en los quintiles responda lo siguiente:

valor-p = 0.08234826 **[0.8 Ptos.]**

¿El modelo ajusta considerando un nivel de significancia del 10 %? NO **[0.2 Ptos.]**

Nota: Para el cálculo del estadístico de prueba, considere intervalos cerrados por la derecha: (LI - LS], excepto el 1ero que deberá también cerrarlo por la izquierda: [LI - LS].

LI: Límite inferior
LS: Límite superior

```
X = dplyr::filter(Base, Operador == "C")$Velocidad
par = fitdistrplus::fitdist(data = X, method = "mle", distr = "weibull")$estimate
x = quantile(X, prob = seq(0,1,.2))
x[1] = 0
x[6] = Inf
p = diff(pweibull(x, shape = par[1], scale = par[2]))
O = hist(X, plot = F, breaks = x, right = T)$count
O
[1] 10 10 10 10 10
X2 = chisq.test(x = O, p = p)$statistic
1-pchisq(X2, df = 5-1-2)
 X-squared
0.08234826
```

Pregunta 10

Para cada ciudad, promedie las velocidades de subida de los cuatro operadores y ajuste mediante un gráfico de probabilidad una distribución Log-Normal a estos promedios:

$\lambda = 2.3056564$ **[0.5 Ptos.]**
$\zeta = 0.5003953$ **[0.5 Ptos.]**

```
X = aggregate(Velocidad ~ Ciudad, data = Base, FUN = mean)[,2]
xp = sort(X)
N = length(X)
p = 1:N/(N+1)
par = lm(log(xp)~qnorm(p))$coef
par
(Intercept)    qnorm(p)
  2.3056564   0.5003953
```

Pregunta 11

Para los promedios por ciudad, calculados anteriormente, existe evidencia para afirmar que la velocidad media es mayor a 10 Mbps? Suponga que los datos distribuyen Gamma$(4, \nu)$, realice la prueba de hipótesis correspondiente y responda:

valor-p = 0.04091037 **[0.8 Ptos.]** 0.06062555 **[0.8 Ptos.]**

¿Apoyaría la afirmación, considerando un nivel de significancia del 1 %? NO **[0.2 Ptos.]**

```
X = aggregate(Velocidad ~ Ciudad, data = Base, FUN = mean)[,2]

## H0: mu = mu0 vs Ha: mu > mu0
mu0 = 10
k   = 4
nu0 = k/mu0
n   = length(X)
sigma = sqrt(k/nu0^2)
TeachingDemos::z.test(x = X, mu = mu0, stdev = sigma, alternative = "greater")$p.value
[1] 0.04091037
TeachingDemos::z.test(x = mean(X), mu = mu0, stdev = sigma/sqrt(n), alternative = "greater")$p.value
[1] 0.04091037

## H0: nu = nu0 vs Ha: nu < nu0
mu0 = 10
k   = 4
nu0 = k/mu0
n   = length(X)
CCR = nu0^2/(n*k)
nu.hat = k/mean(X)
TeachingDemos::z.test(x = nu.hat, mu = nu0, stdev = sqrt(CCR), alternative = "less")$p.value
[1] 0.06062555
```

Pregunta 12

A continuación se presenta la salida de un modelo de regresión lineal simple:

```
lm(formula = Imacec ~ Pib, data = Base)

Coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept) *********  1.8855300  ******    0.125
Pib         0.0007133  0.0000149  47.887   <2e-16

Residual standard error: 4.156 on 191 degrees of freedom
Multiple R-squared: ****, Adjusted R-squared:  0.9227
F-statistic: ***** on 1 and 191 DF,  p-value: < 2.2e-16
```

Ingrese los valores faltantes:

Estimate = 2.9053840 **[0.2 Ptos.]**

t value = 1.5408845 **[0.2 Ptos.]**

Multiple R-squared = 0.9231026 **[0.2 Ptos.]**

F-statistic = 2293.1647690 **[0.2 Ptos.]**

¿Considerando un nivel de significancia del 1 %, hay regresión? SI **[0.2 Ptos.]**

```
M = rbind(qt(1-0.125/2, 191), 1.8855300*qt(1-0.125/2, 191), 47.887^2, 1-(1-0.9227)*191/192)
rownames(M) = c("t value", "Estimate", "F-statistic", "Multiple R-squared")
colnames(M) = "Respuesta"
M
                      Respuesta
t value               1.5408845
Estimate              2.9053840
F-statistic        2293.1647690
Multiple R-squared    0.9231026
```