Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2021

Interrogación 3 - Pauta

Pregunta 1

¿Cuales de las siguientes relaciones entre modelos siempre son correctas?

i. $X \sim \text{Log-Normal}(\lambda,\, \zeta) \to e^X \sim \text{Normal}(\lambda, \zeta)$

ii. $X \sim \text{Gamma}(k,\, \nu) \to c \cdot X \sim \text{Gamma}(c \cdot k, \nu), \quad c > 0$

iii. $X \sim \text{Normal}(\mu, \sigma) \to \left(\dfrac{X-\mu}{\sigma}\right)^2 \sim \text{Gamma}\left(1,\, \dfrac{1}{2}\right)$

Solución

Ninguna de las anteriores **[1.0 Ptos.]**, ya que

i. Si $X \sim \text{Log-Normal}(\lambda,\, \zeta) \to \ln(X) \sim \text{Normal}(\lambda, \zeta)$. **[1/3 Ptos.] si hay respaldo**

i. Si $X \sim \text{Gamma}(k,\, \nu) \to c \cdot X \sim \text{Gamma}(k, \nu/c),\ c > 0$. **[1/3 Ptos.] si hay respaldo**

iii. $X \sim \text{Normal}(\mu, \sigma) \to \left(\dfrac{X-\mu}{\sigma}\right)^2 \sim \text{Gamma}\left(\dfrac{1}{2}, \dfrac{1}{2}\right) = \chi^2(1)$. **[1/3 Ptos.] si hay respaldo**

Pregunta 2

Actualmente los laboratorio están analizando todas las muestras Covid + para chequear si presentan la variante Delta. Suponga que el tiempo X que tomó realizar un test PCR que resultó Covid+ y el tiempo Y que toma determinar si corresponde a la variante Delta tienen el siguiente comportamiento conjunto:

$$f_{X,Y}(x, y) = \beta^3\, y\, e^{-\beta\,(x+y)},$$

con $x > 0$, $y > 0$ y $\beta > 0$.

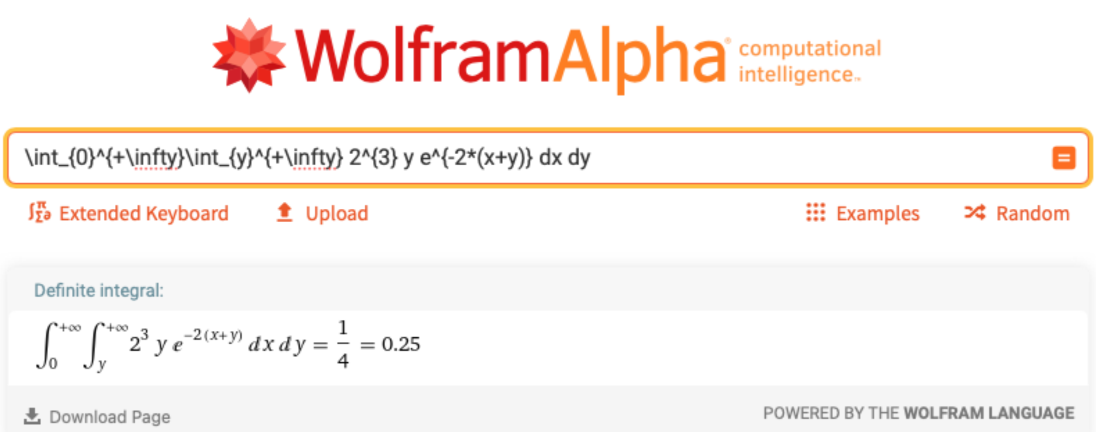
Para $\beta =$ `[beta]`, ¿cuál es la probabilidad que el tiempo que tomo determinar que el PCR era Covid+ sea mayor al tiempo que tomo determinar que correspondía a la variante Delta?

Solución

Se pide

$$\begin{aligned}
P(X > Y) &= \int_0^{+\infty}\int_y^{+\infty} \beta^3\, y\, e^{-\beta\,(x+y)}\, dx\, dy \\
&= \int_0^{+\infty} \beta^2\, y\, e^{-\beta\, y} \cdot \left\{\int_y^{+\infty} \beta\, e^{-\beta\, x}\, dx\right\} dy \\
&= \int_0^{+\infty} \beta^2\, y\, e^{-\beta\, y} \cdot e^{-\beta\, y}\, dy \\
&= \frac{1}{4}\int_0^{+\infty} \frac{(2\,\beta)^2}{\Gamma(2)}\, y^{2-1}\, e^{-2\,\beta\, y}\, dy \\
&= \frac{1}{4} \cdot 1, \quad \text{por área bajo todo el soporte de una Gamma}(2,\, 2\,\beta)
\end{aligned}$$

En WolframAlpha utilizando por ejemplo `[beta] = 2`:



[1.0 Ptos.]

Pregunta 3

Considere una muestra aleatoria (iid) de tamaño [n] proveniente de una población Logistica (μ, σ), ¿Cuál es la probabilidad aproximada que el estimador de momentos de μ sea mayor a $\left(\mu + \dfrac{\sigma}{4}\right)$?

Solución

El estimador de momentos de μ es $\overline{X}$, el cuál aproximadamente por el teorema del límite central distribuye

$$\overline{X} \overset{\text{aprox}}{\sim} \text{Normal}\left(\mu,\, \frac{\sigma\,\pi}{\sqrt{3\,n}}\right) \textbf{[0.3 Ptos.]}$$

Se pide

$$P\left(\overline{X} > \mu + \frac{\sigma}{4}\right) \approx 1 - \Phi\left(\frac{\mu + \frac{\sigma}{4} - \mu}{\frac{\sigma\,\pi}{\sqrt{3\,n}}}\right) = 1 - \Phi\left(\frac{\sqrt{3\,n}}{4\,\pi}\right) = \texttt{1 - pnorm(sqrt(3*n)/(4*pi))} \quad \textbf{[0.7 Ptos.]}$$

Pregunta 4

¿Cuál es la mediana del mínimo entre [n] variables aleatorias iid $\chi^2(2)$?

Solución

Como $\chi^2(2) = \text{Gamma}\left(\frac{2}{2}, \frac{1}{2}\right) = \text{Exp}\,(1/2)$, entonces el mínimo entre `[n]` variables iid distribuye $\text{Exp}(\texttt{[n]}/2)$.

[0.3 Ptos.]

Por lo tanto, la mediana es: $\dfrac{2\,\ln(2)}{\texttt{[n]}}$. **[0.7 Ptos.]**

Pregunta 5

Sean X y Y dos variables aleatorias independientes con distribución Exp(2). Determine el coeficiente de variación aproximado de 1er orden de X^Y(X elevado a Y).

Solución

Tenemos que

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

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

$$\textbf{[0.2 Ptos.]} \quad \mu_Z = \mu_X^{\mu_Y} \quad \text{y} \quad \sigma_Z^2 = \sigma_X^2 \cdot \left(\mu_Y \cdot \mu_X^{\mu_Y - 1}\right)^2 + \sigma_Y^2 \cdot (\mu_X^{\mu_Y} \cdot \ln(\mu_X))^2 \quad \textbf{[0.2 Ptos.]}$$

Por lo tanto

$$\delta_Z \approx \sqrt{\sigma_X^2 \cdot (\mu_Y/\mu_X)^2 + \sigma_Y^2 \cdot [\ln(\mu_X)]^2} = \frac{\sqrt{1 + [\ln(2)]^2}}{2} \quad \textbf{[0.6 Ptos.]}$$

Pregunta 6

El archivo `ENS_Muestra.xlsx` contiene una muestra aleatoria de 350 casos con indicadores de salud. Considerando solo las personas de género "Masculino" y bajo el supuesto de normalidad, calcule:

$$P(\text{ COLESTEROL } > 4 \cdot \text{HDL})$$

Hint: Para la estimación de los parámetros puede utilizar las funciones de R `mean()`*,* `sd()`*,* `var()`*,* `cov()` *ó* `cor()`*.*

Solución

Se pide $P(4\,X - Y < 0)$, con Y: COLESTEROL e X: HDL.

Bajo Normalidad se tiene que

$$4\,X - Y \sim \text{Normal}(4\,\mu_X - \mu_y, \sqrt{4^2 \cdot \sigma_X^2 + \sigma_Y^2 - 2 \cdot 4 \cdot \rho \cdot \sigma_X \cdot \sigma_y}) \quad \textbf{[0.5 Ptos.]}$$

En R:

```
Base = rio::import("ENS_Muestra.xlsx")
Y = dplyr::filter(Base, SEXO == "Masculino")$COLESTEROL
X = dplyr::filter(Base, SEXO == "Masculino")$HDL

mu = 4*mean(X)-mean(Y)

sigma = sqrt(16*var(X)-2*4*cov(X,Y)+var(Y))
pnorm(0, mean = mu, sd = sigma)
[1] 0.5801775

sigma = sqrt(16*var(X)-2*4*cor(X,Y)*sd(X)*sd(Y)+var(Y))
pnorm(0, mean = mu, sd = sigma)
[1] 0.5801775

Z = 4*X-Y
mu = mean(Z)
sigma = sd(Z)
pnorm(0, mean = mu, sd = sigma)
[1] 0.5801775
```

[0.5 Ptos.]

Pregunta 7

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

Solución

Con μ conocido, el siguiente pivote basado en el estimador máximo verosímil de σ:

$$C = \frac{\texttt{[n]} \cdot \hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{\texttt{[n]}} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(\texttt{[n]}) \quad \textbf{[0.5 Ptos.]}$$

Se pide

$$P(\hat{\sigma} \leq 0.8 \cdot \sigma) = P\left(C \leq 0.8^2 \cdot \texttt{[n]}\right) = \texttt{pchisq(0.8 * 0.8 * n, df = n)} \quad \textbf{[0.5 Ptos.]}$$

La base Prototipo.xlsx contiene información sobre un estudio en que se probaron 35 motores fabricados bajo un nuevo prototipo y 40 fabricados de acuerdo al estándar. Se registró el RENDIMIENTO (km/lts) de cada motor según tipo de PROTOTIPO (nuevo o estándar).

A continuación responda las siguientes 4 preguntas:

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

Pregunta 8.1

El grupo de ingenieros a cargo de probar los motores, estiman que el rendimiento del “nuevo” prototipo debería tener un rendimiento promedio mayor que 23 km/lts. Considerando un nivel de significante del 1 %, realice una prueba de hipótesis bajo el supuesto de normalidad y responda:

Valor-p

```
X <- dplyr::filter(Base, PROTOTIPO == "nuevo")$RENDIMIENTO
mu0 <- 23
t.test(x = X, mu = mu0, alternative="greater")$p.value
[1] 0.02672312
```

[0.8 Ptos.]

Basado en el valor-p y el nivel de significancia, ¿está de acuerdo con lo afirmado por sus colegas? *No*

[0.2 Ptos.]

Pregunta 8.2

El grupo de ingenieros a cargo de probar los motores, estiman que el rendimiento del “nuevo” prototipo debería tener un rendimiento promedio mayor que 23 km/lts. Considerando un nivel de significante del 5 %, realice una prueba de hipótesis bajo el supuesto de normalidad y responda:

Valor-p

```
X <- dplyr::filter(Base, PROTOTIPO == "nuevo")$RENDIMIENTO
mu0 <- 23
t.test(x = X, mu = mu0, alternative="greater")$p.value
[1] 0.02672312
```

[0.8 Ptos.]

Basado en el valor-p y el nivel de significancia, ¿está de acuerdo con lo afirmado por sus colegas? *Si*

[0.2 Ptos.]

Nota: Conclusiones sin valores-p de respaldo tienen [0.0 Ptos.]. Conclusiones correctas, pero a valor-p incorrecto, tienen [0.2 Ptos.].

Pregunta 9.1

¿Existe evidencia para afirmar que la dispersión de los rendimientos en motores de prototipo estándar es menor a 0.57 km/ lts? Utilice un nivel de significancia del 5 %, realice una prueba de hipótesis bajo el supuesto de normalidad y responda:

Valor-p

```
X <- dplyr::filter(Base, PROTOTIPO == "estandar")$RENDIMIENTO
sigma0 <- 0.57
TeachingDemos::sigma.test(x = X, sigma = sigma0, alternative="less")$p.value
[1] 0.05797018
```

[0.8 Ptos.]

Basado en el valor-p y el nivel de significancia, ¿está de acuerdo con la afirmación? *No*

[0.2 Ptos.]

Pregunta 9.2

¿Existe evidencia para afirmar que la dispersión de los rendimientos en motores de prototipo estándar es menor a 0.57 km/lts? Utilice un nivel de significancia del 10 %, realice una prueba de hipótesis bajo el supuesto de normalidad y responda:
Valor-p

```
X <- dplyr::filter(Base, PROTOTIPO == "estandar")$RENDIMIENTO
sigma0 <- 0.57
TeachingDemos::sigma.test(x = X, sigma = sigma0, alternative="less")$p.value
[1] 0.05797018
```

[0.8 Ptos.]

Basado en el valor-p y el nivel de significancia, ¿está de acuerdo con la afirmación? *Si*

[0.2 Ptos.]

Nota: Conclusiones sin valores-p de respaldo tienen [0.0 Ptos.]. Conclusiones correctas, pero a valor-p incorrecto, tienen [0.2 Ptos.].

Pregunta 10.1

Se sugiere que el rendimiento de los motores basados en el nuevo prototipo se podrían comportar como variables aleatorias: Normal, Log-Normal o Gamma. Estime por el método de momentos los tres modelos y evalué cuál de estos ajusta mejor al porcentaje de rendimientos mayores a 23.5 km/lts.

```
X <- dplyr::filter(Base, PROTOTIPO == "nuevo")$RENDIMIENTO
```

Resultado empírico

```
mean(X > 23.5)
[1] 0.3142857
```

[0.4 Ptos.]

Estimación mejor modelo

```
par = fitdistrplus::fitdist(data = X, distr = "norm", method = "mme")$estimate
1-pnorm(23.5, mean = par[1], sd = par[2])
[1] 0.2977296
par = fitdistrplus::fitdist(data = X, distr = "lnorm", method = "mme")$estimate
1-plnorm(23.5, meanlog = par[1], sdlog = par[2])
[1] 0.2946315
par = fitdistrplus::fitdist(data = X, distr = "gamma", method = "mme")$estimate
1-pgamma(23.5, shape = par[1], rate = par[2])
[1] 0.2956845
```

[0.4 Ptos.]

Mejor ajuste: *Normal* **[0.2 Ptos.]**

Pregunta 10.2

Se sugiere que el rendimiento de los motores basados en el prototipo estándar se podrían comportar como variables aleatorias: Gamma, Log-Normal o Weibull. Estime por el método de máxima verosimilitud los tres modelos y evalué cuál de estos ajusta mejor al porcentaje de rendimientos mayores a 21.8 km/lts.

```
X <- dplyr::filter(Base, PROTOTIPO == "estandar")$RENDIMIENTO
```

Resultado empírico

```
mean(X > 21.8)
[1] 0.275
```

[0.4 Ptos.]

Estimación mejor modelo

```
par = fitdistrplus::fitdist(data = X, distr = "gamma", method = "mle")$estimate
1-pgamma(21.8, shape = par[1], rate = par[2])
[1] 0.2701989
par = fitdistrplus::fitdist(data = X, distr = "lnorm", method = "mle")$estimate
1-plnorm(21.8, meanlog = par[1], sdlog = par[2])
[1] 0.2694802
par = fitdistrplus::fitdist(data = X, distr = "weibull", method = "mle")$estimate
1-pweibull(21.8, shape = par[1], scale = par[2])
[1] 0.3248864
```

[0.4 Ptos.]

Mejor ajuste: *Gamma* **[0.2 Ptos.]**
Nota: Mejor modelo sin respaldo [0.0 Ptos.].

Pregunta 11.1

Suponga que los rendimientos de los motores basados en el prototipo estándar distribuye Gamma($k = 2177$, ν). ¿Existe evidencia para afirmar que los rendimientos es estos motores presentan un rendimiento medio mayor a 21.4 km/lts? Entregue el valor del estadístico de prueba que le permitía evaluar esta afirmación.

```
X <- dplyr::filter(Base, PROTOTIPO == "estandar")$RENDIMIENTO
k = 2177
mu0 = 21.4
n = length(X)
Z0 = (mean(X)-mu0)/sqrt(mu0^2/(n*k))
Z0
[1] 1.651281
Z0 = (k/mean(X)-k/mu0)/sqrt((k/mu0)^2/(n*k))
Z0
[1] -1.642092
```

Pregunta 11.2

Suponga que los rendimientos de los motores basados en el nuevo prototipo distribuye Gamma($k = 1645$, ν). ¿Existe evidencia para afirmar que los rendimientos es estos motores presentan un rendimiento medio menor a 23.4 km/lts? Entregue el valor del estadístico de prueba que le permitía evaluar esta afirmación

```
X <- dplyr::filter(Base, PROTOTIPO == "nuevo")$RENDIMIENTO
k = 1645
n = length(X)
mu0 = 23.4
Z0 = (mean(X)-mu0)/sqrt(mu0^2/(n*k))
Z0
[1] -2.088924
Z0 = (k/mean(X)-k/mu0)/sqrt((k/mu0)^2/(n*k))
Z0
[1] 2.107269
```

Pregunta 12

Ultimamente ha estado en discusión los permisos humanitarios que el gobierno extiende para viajar en estos tiempos pandémicos. Un especialista afirma que la proporción de permisos entregados supera el tercio de las solicitudes.

Usted, solicita a través de transparencia a la Subsecretaria de Prevención de Delito el resumen de todas las solicitudes de la ultima semana. Los datos entregados por ellos muestran que de un total de 124 solicitudes, solo en 50 casos se ha accedido a entregar este permiso humanitario.

Utilizando un nivel de significancia del 5 %, realice una prueba de hipótesis correspondiente y responda:
Valor-p

```
n = 124
x = 50
p0 = 1/3
TeachingDemos::z.test(x = x/n, mu = p0, stdev = sqrt(p0*(1-p0)/n),alternative = "greater")$p.value
[1] 0.04936912
prop.test(x = x, n = n, p = p0, alternative = "greater", correct = F)$p.value
[1] 0.04936912
```

[0.8 Ptos.]

Basado en el valor-p y el nivel de significancia, ¿está de acuerdo con la afirmación? *Si*

[0.2 Ptos.]

Nota: Conclusiones sin valores-p de respaldo tienen [0.0 Ptos.]. Conclusiones correctas, pero a valor-p incorrecto, tienen [0.2 Ptos.].