Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2023

Curso : Probabilidad y Estadística
Sigla : EYP1113
Profesores : Ricardo Aravena C., Ricardo Olea O.,
Felipe Ossa M. y Alejandro Sepulveda P.

INTERROGACIÓN 2

Pregunta 1

Suponga que el número de ingresos clandestinos diarios que ocurren en Colchane, frontera con Bolivia en la Región de Tarapacá, es una variable aleatoria Poisson cuyo coeficiente de variación es de 17.5 %.

(a) El personal de frontera (que debería evitar los ingresos) almuerza diariamente entre 13:00 y 15:00 hrs. Determine la probabilidad que, durante el horario de almuerzo, en los próximos cinco días, al menos en dos días ingresen clandestinamente más de tres personas.

(b) Si usted, mediante un drone, registra los ingresos clandestinos en Colchane, y comienza a grabar a las 09.00 hrs, ¿cuál es la probabilidad que registre al menos un ingreso antes de las 09:45 hrs, si ya revisó los primeros 20 minutos de grabación y no registró ingresos clandestinos?

Solución

(a) Sea X_t el número de ingresos clandestinos en Colchane en t días, variable aleatoria que según enunciado distribuye Poisson$(\nu \cdot t)$, con

$$\nu = \frac{1}{0.175^2} = 32.65306.$$

Sea p la probabilidad que durante el horario de almuerzo ingresen clandestinamente más de tres personas.

$$p = P(X_{2/24} > 3) = 1 - P(X_{2/24} \leq 3) = 1 - \sum_{x=0}^{3} \frac{(\nu \cdot 2/24)^x \, e^{-\nu \cdot 2/24}}{x!} = 0.2905625$$

Definamos como Y el número de días, entre los próximos cinco, en que ingresan más de tres personas clandestinamente durante el horario de almuerzo.

$$Y \sim \text{Binomial}(n = 5,\, p)$$

Se pide

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - \sum_{y=0}^{1} \binom{5}{y} p^y \, (1-p)^{5-y} = 0.452275$$

(b) Alternativa 1: Sea T el tiempo, en minutos, transcurrido hasta el 1er ingreso clandestino desde el inicio de la grabación (09.00 horas).

$$T \sim \text{Exponencial}\left(\frac{\nu}{24 \cdot 60}\right)$$

Se pide

$$\begin{aligned}
P(T \leq 45 \,|\, T > 20) &= 1 - P(T > 45 \,|\, T > 20) \\
&= 1 - \frac{P(\{T > 45\} \cap \{T > 20\})}{P(T > 20)} \\
&= 1 - \frac{P(T > 45)}{P(T > 20)} \\
&= 1 - \exp\left(-\frac{\nu}{24 \cdot 60} \cdot (45 - 20)\right) \\
&= 1 - \exp\left(-\frac{\nu}{24 \cdot 60} \cdot 25\right), \quad \text{por propiedad de carencia de memoria} \\
&= 0.432715
\end{aligned}$$

Alternativa 2: Sea Y_t el número de ingresos clandestinos en Colchane en t minutos

$$\text{Poisson}\left(\frac{\nu}{24 \cdot 60} \cdot t\right)$$

Se pide

$$\begin{aligned}
P(Y_{45} \geq 1 \,|\, Y_{20} = 0) &= \frac{P(\{Y_{20} = 0\} \cap \{Y_{45} \geq 1\})}{P(Y_{20} = 0)} \\
&= \frac{P(\{Y_{20} = 0\} \cap \{Y_{45-20} \geq 1\})}{P(Y_{20} = 0)} \\
&= \frac{P(Y_{20} = 0) \cdot P(Y_{45-20} \geq 1)}{P(Y_{20} = 0)}, \quad \text{por independencia entre eventos} \\
&= P(Y_{45-20} \geq 1) \\
&= 1 - P(Y_{45-20} = 0) \\
&= 1 - \exp\left(-\frac{\nu}{24 \cdot 60} \cdot (45 - 20)\right) \\
&= 1 - \exp\left(-\frac{\nu}{24 \cdot 60} \cdot 25\right) \\
&= 0.432715
\end{aligned}$$

Asignación de Puntaje:

Logro 1: Obtener en (a) el valor $\nu = 32.65306$ en días, 1.360544 en horas o 0.02267574 en minutos. **[1.0 Ptos]**

Logro 2: Obtener en (a) $p = 0.2905625$ **[1.0 Ptos]**

Logro 3: Obtener en (a) Obtener $P(Y \geq 2) = 0.452275$ **[1.0 Ptos]**

Logro 4: Define en (b) correctamente el modelo a utilizar: Exponencial o Poisson **[1.0 Ptos]**

Logro 5: Aplica en (b) correctamente la carencia de memoria en el caso Exponencial e independencia en le caso Poisson. **[1.0 Ptos]**

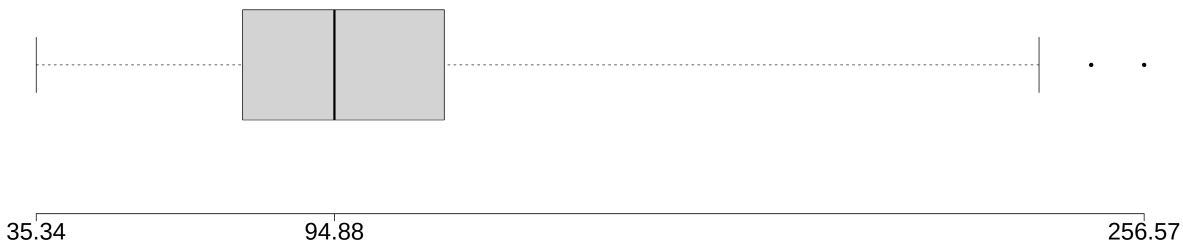
Logro 6: Obtener en (b) el resultado final igual a 0.432715 **[1.0 Ptos]**

+ 1 Punto Base

Nota: Cada logro, dependiendo el desarrollo planteado, puede tener puntaje parcial en el caso de no lograrlo completamente.

Pregunta 2

Durante marzo las plantas de revisión técnica presentan largas esperas y a partir de una muestra aleatoria de tiempos de espera se obtiene la siguiente información:




```
   c.o.v  skewness kurtosis
0.318178 0.9226519 1.189797
```

Bajo el supuesto que estos tiempos distribuyen Log-Normal:

(a) ¿Cuál sería el IQR de los tiempos de espera en marzo?

(b) ¿Cuál es la probabilidad que los tiempos de espera en marzo superen los 150 minutos?

(c) Cuál es la diferencia absoluta entre el skewness de la Log-Normal estimada y la skewness empírica de los tiempos de espera de marzo que contiene la muestra?

Solución

Sea T el tiempo de espera, el cual distribuye Log-Normal$(\lambda,\, \zeta)$.

Del enunciado se desprende que

$$\lambda = \ln(94.88) = 4.552613 \quad \text{y} \quad \zeta = \sqrt{\ln(1 + 0.318178^2)} = 0.3105388.$$

(a) Se pide

$$\text{IQR} = e^{\lambda + \zeta \cdot \Phi^{-1}(0.75)} - e^{\lambda + \zeta \cdot \Phi^{-1}(0.25)} \approx e^{\lambda + \zeta \cdot 0.675} - e^{\lambda - \zeta \cdot 0.675} = 40.06821.$$

(b) Se pide

$$P(T > 150) = 1 - \Phi\left(\frac{\ln(150) - \lambda}{\zeta}\right) \approx 1 - \Phi(1.47) = 0.0708.$$

(c) Del formulario se tiene que

$$\text{E}(T^k) = e^{k\,\lambda + (k\,\zeta)^2/2}.$$

Reemplazando

$$\theta_X = \frac{\text{E}(T^3) - 3 \cdot \text{E}(T^2) \cdot \mu_T + 2\,\mu_T^3}{\left[\text{E}(T^2) - \mu_T^2\right]^{3/2}} = 0.9867456.$$

Por lo tanto,

$$|\texttt{skewness} - \theta_X| = 0.06409367.$$

Alternativamente, el alumno podrían haber asignado a T, tiempo de espera, una distribución Log-Normal(λ, ζ) trasladado en α.

En este caso, desde enunciado, se desprende que

$$\alpha = 35.34, \quad \text{y} \quad \lambda = \ln(94.88 - \alpha) = 4.086648.$$

El calculo de ζ en este caso requiere resolver la siguiente ecuación:

$$\texttt{c.o.v} \cdot \left(e^{\lambda+\zeta^2/2} + \alpha\right) = \left(e^{2\lambda+2\zeta^2} - e^{2\lambda+\zeta^2}\right)^{1/2} \to \zeta \approx 0.294006 \quad \text{(resuelto de manera numérica)}.$$

Es posible que por simplicidad utilicen la relación c.o.v. vs δ_X de la Log-Normal no trasladada

$$\zeta = \sqrt{\ln(1 + 0.318178^2)} = 0.3105388.$$

(a) Se pide

$$\text{IQR} = \left[e^{\lambda+\zeta\cdot\Phi^{-1}(0.75)} + \alpha\right] - \left[e^{\lambda+\zeta\cdot\Phi^{-1}(0.25)} + \alpha\right] \approx e^{\lambda+\zeta\cdot 0.675} - e^{\lambda-\zeta\cdot 0.675} = 25.14399.$$

(b) Se pide

$$P(T > 150) = 1 - \Phi\left(\frac{\ln(150-\alpha) - \lambda}{\zeta}\right) \approx 1 - \Phi(2.11) = 0.0174.$$

(c) Del formulario se puede deducir que

$$\mathsf{E}[(T-\alpha)^k] = e^{k\,\lambda + (k\,\zeta)^2/2}.$$

Reemplazando

$$\theta_X = \frac{\mathsf{E}(T^3) - 3\cdot\mathsf{E}(T^2)\cdot\mu_T + 2\,\mu_T^3}{[\mathsf{E}(T^2) - \mu_T^2]^{3/2}} = 0.9867456.$$

Por lo tanto,

$$|\texttt{skewness} - \theta_X| = 0.06409367.$$

Asignación de Puntaje:

Logro 1: Obtener $\lambda = 4.552613$ o 4.086648 y $\zeta = 0.3105388$ o 0.294006. **[1.0 Ptos]**

Logro 2: Responder IQR = 40.06821 en el caso Log-Normal o Responder IQR = 25.14399 en el caso Log-Normal Trasladado. **[1.0 Ptos]**

Logro 3: Relación acumulada Log-Normal con Normal(0,1). **[1.0 Ptos]**

Logro 4: Uso correcto tabla Normal(0,1) respondiendo 0.0708 (caso Log-Normal) o 0.01744 (caso Log-Normal Trasladada). **[1.0 Ptos]**

Logro 5: Por obtener coeficiente de asimetría igual a 0.9867456, este resultado es idéntico para Log-Normal y Log-Normal trasladada. **[1.0 Ptos]**

Logro 6: Por obtener diferencia absoluta igual a 0.06409367. **[1.0 Ptos]**

+ 1 Punto Base

Nota: Cada logro, dependiendo el desarrollo planteado, puede tener puntaje parcial en el caso de no lograrlo completamente.

Pregunta 3

La distribución de probabilidad Dagum es una distribución continua utilizada para modelar variables aleatorias que toman valores positivos y se ajustan a una forma de cola pesada, es decir, tienen una probabilidad significativa de tomar valores extremos.

Esta distribución se utiliza comúnmente para modelar la distribución de ingresos, la distribución de tamaños de empresa, la distribución de valores de propiedades inmobiliarias, entre otros. También se utiliza en finanzas para modelar la duración de los precios de los activos financieros, como las acciones y los bonos.

Sea X una variable aleatoria Dagum$(\alpha,\, \beta,\, \gamma)$ cuya función de probabilidad acumulada está dada por:

$$F_X(x) = \left[1 + \left(\frac{x}{\beta}\right)^{-\alpha}\right]^{-\gamma}, \; x > 0, \, \alpha > 0, \, \beta > 0 \text{ y } \gamma > 0$$

Obtenga la mediana y moda de este modelo. Evalúe para $\alpha = 2.5$, $\beta = 2$ y $\gamma = 3$.

Solución

Sea $x_{50\,\%}$ el percentil $50\,\%$ (mediana).

$$F_X(x_{50\,\%}) = \left[1 + \left(\frac{x_{50\,\%}}{\beta}\right)^{-\alpha}\right]^{-\gamma} = \frac{1}{2} \to x_{50\,\%} = \beta\,\left((\frac{1}{2})^{-1/\gamma} - 1\right)^{-1/\alpha}$$

Reemplazando en $p = 0.5$, $\alpha = 2.5$, $\beta = 2$ y $\gamma = 3$, se tiene que $x_{50\,\%} = 3.428415$.

Por otra parte, la moda es el valor de x donde la densidad tiene pendiente igual a cero.

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{\alpha\,\gamma}{\beta}\left(1 + \left(\frac{x}{\beta}\right)^{-\alpha}\right)^{-(\gamma+1)} \left(\frac{x}{\beta}\right)^{-(\alpha+1)}$$

$$f_X'(x) = \frac{\alpha\,\gamma}{\beta}\left[\frac{(\gamma+1)\,\alpha}{\beta}\left(1 + \left(\frac{x}{\beta}\right)^{-\alpha}\right)^{-(\gamma+2)} \left(\frac{x}{\beta}\right)^{-2\,(\alpha+1)} - \frac{(\alpha+1)}{\beta}\left(1 + \left(\frac{x}{\beta}\right)^{-\alpha}\right)^{-(\gamma+1)} \left(\frac{x}{\beta}\right)^{-(\alpha+2)}\right]$$

Igualando f' a cero se tiene que

$$\text{moda} = \beta\,\left(\frac{\gamma\,\alpha - 1}{\alpha + 1}\right)^{1/\alpha} = 2.561935$$

Asignación de Puntaje:

Logro 1: Igualar $F_X(x_{50\,\%}) = 1/2$. **[1.0 Ptos]**

Logro 2: A partir de $F_X^{-1}(1/2)$, expresar $x_{50\,\%}$ en términos de los parámetros α, β y γ. **[1.0 Ptos]**

Logro 3: Evaluar y obtener $x_{50\,\%} = 3.428415$. **[1.0 Ptos]**

Logro 4: Obtener f_X. **[1.0 Ptos]**

Logro 5: Obtener f_X'. **[1.0 Ptos]**

Logro 6: Despejar x desde $f_X'(x) = 0$ y obtener moda = 2.561935. **[1.0 Ptos]**

+ 1 Punto Base

Nota: Cada logro, dependiendo el desarrollo planteado, puede tener puntaje parcial en el caso de no lograrlo completamente.

Pregunta 4

Entre las indicaciones que los profesores están proyectando en este momento, se encuentra escribir su nombre en cada página de su cuadernillo (si aún no lo ha realizado, ahora es el momento).

A cada uno de los nueve ayudantes correctores los profesores pedirán que registren el número de alumnos revisados hasta observar el k-ésimo alumno que no siguió la indicación, con k un parámetro desconocido para usted, pero no para los ayudantes.

Suponga que que los ayudantes, una vez finalizada la corrección de la prueba, informan que revisaron 66.66667 alumnos en promedio hasta detectar al k-ésimo alumno que no siguió la indicación y que el coeficiente de variación resultante fue de 0.6964194. Proponga un modelo de probabilidad que describa el valor que los ayudantes reportarán e indique el valor de los parámetros.

Solución

Sea X el valor que los ayudantes reportarán.

Del enunciado se deduce que

$$X \sim \text{Binomial-Negativa}(k,\, p).$$

Del formulario se tiene que

$$\mu_X = \frac{k}{p} \quad \text{y} \quad \delta_T = \sqrt{\frac{1-p}{k}}.$$

Como

$$\mu_X = 66.66667 \quad \text{y} \quad \delta_X = 0.6964194 \rightarrow k = 2 \quad \text{y} \quad p = 0.03$$

Asignación de Puntaje:

Logro 1: Por definir correctamente que la variable aleatoria distribuye Binomial-Negativa. **[1.0 Ptos]**

Logro 2: Por relación μ_X con respecto a los parámetros k y p. **[1.0 Ptos]**

Logro 3: Por relación δ_X (c.o.v.) con respecto a los parámetros k y p. **[1.0 Ptos]**

Logro 4: Por planteamiento del sistema de ecuaciones a resolver. **[1.0 Ptos]**

Logro 5: Obtener $k = 2$. **[1.0 Ptos]**

Logro 6: Obtener $p = 0.03$. **[1.0 Ptos]**

+ 1 Punto Base

Nota: Cada logro, dependiendo el desarrollo planteado, puede tener puntaje parcial en el caso de no lograrlo completamente.