Pontificia Universidad Católica de Chile
Facultad de Ciencias Económicas y Administrativas

Segundo Semestre 2016

Curso	**:**	**Inferencia Estadística**
Sigla	**:**	**EAS201a**
Profesores	**:**	**Rafael Águila (Sec 01 - Sec 02), Osvaldo Ferreiro (Sec 03)**
		Victor Correa (Sec 04) y Ricardo Olea (Sec 05)

Pauta Interrogación 1

Problema 1

Una máquina embotelladora de bebidas, puede ser regulada de tal manera que las descargas en promedio sean igual a μ cm^3 por botella. Se ha observado que la cantidad de llenado dispensado por la máquina tiene una distribución Normal(μ, σ^2), con σ igual a 1 cm^3.

Una muestra de 9 botellas fue seleccionada aleatoriamente de la máquina en un día cualquiera, registrándose los cm^3 de llenado para cada una.

(a) **[2.0 Ptos.]** Calcular la probabilidad que el promedio muestral se encuentre ± 0.3 cm^3 respecto de la media poblacional.

(b) **[2.0 Ptos.]** ¿Cuántas observaciones deben ser incluidas en la muestra si se desea que el promedio muestral este en ± 0.3 cm^3 respecto de la media población con una probabilidad del 95 %?

Supóngase ahora que se tiene una muestra de 10 botellas.

(c) **[2.0 Ptos.]** Determine un valor k, tal que el estimador máximo verosímil de σ^2 sea inferior a este con probabilidad 0.9.

Solución

(a) Tenemos que

$$\overline{X}_n \sim \text{Normal}\left(\mu,\, \frac{\sigma^2}{n}\right) \quad \textbf{[0.5 Ptos.]}$$

con n igual a 9 y σ igual a uno.

Se pide

$$\textbf{[0.5 Ptos.]} \quad P\left(\mu - 0.3 \leq \overline{X}_n \leq \mu + 0.3\right) = \Phi(0.3{\cdot}3) - \Phi(-0.3{\cdot}3) = 2\,\Phi(0.9) - 1 = 2{\cdot}0.8159 - 1 = 0.6318 \quad \textbf{[1.0 Ptos.]}$$

(b) Se pide

$$\Phi(0.3 \cdot n) - \Phi(-0.3 \cdot \sqrt{n}) = 2\,\Phi(0.3 \cdot \sqrt{n}) - 1 = 0.95 \to \Phi(0.3 \cdot \sqrt{n}) = 0.975 \quad \textbf{[1.0 Ptos.]}$$

Es decir,

$$\textbf{[0.5 Ptos.]} \quad 0.3 \cdot \sqrt{n} = 1.96 \to n = 42.68444 \approx 43 \quad \textbf{[0.5 Ptos.]}$$

(c) El estimador de máxima verosimilitud para σ^2 está dado por

$$\widehat{\sigma^2} = \frac{1}{10}\sum_{i=1}^{10}(X_i - \overline{X}_{10})^2 \quad \textbf{[0.3 Ptos.]}$$

Notemos que

$$\frac{10 \cdot \widehat{\sigma^2}}{1^2} \sim \chi^2(10-1) \quad \textbf{[0.5 Ptos.]}$$

y se pide

$$\begin{aligned} P\left(\widehat{\sigma^2} \leq k\right) &= 0.90 \quad \textbf{[0.3 Ptos.]} \\ \to P\left(10 \cdot \widehat{\sigma^2} \leq 10\,k\right) &= 0.90 \quad \textbf{[0.3 Ptos.]} \end{aligned}$$

$$\begin{aligned} \to 10\,k &= 14.68366 \quad \textbf{[0.5 Ptos.]} \\ \to k &= 1.468366 \quad \textbf{[0.1 Ptos.]} \end{aligned}$$

+ 1 Punto Base

Problema 2

Parte I

Ud. ya está trabajando y como parte de ello debe entrevistar candidatos para un trabajo temporal y breve. Se tiene una probabilidad π de que un candidato cumpla los requisitos. Al primer candidato que cumple los requerimientos se le asigna el trabajo. Considere la variable aleatoria Y: "nº de entrevistas hasta la contratación de la persona seleccionada". Ud. ha realizado y realizará este proceso de entrevistas-selección muchas veces. Ud. selecciona una muestra aleatoria (m.a.) de n repeticiones de este proceso, con lo que selecciona una m.a. de tamaño n de Y.

(a) **[2.5 Ptos.]** Obtenga el Estimador Máximo Verosímil (EMV) de π.

(b) **[1.0 Ptos.]** Ud. desea estimar la probabilidad de tener que realizar más de 5 entrevistas en un próximo proceso de entrevistas-selección. Obtenga el EMV de dicha probabilidad.

Parte II

Considere una muestra aleatoria de tamaño m proveniente de una población Uniforme$(-\theta, \theta)$, con $\theta > 0$.

(c) **[2.5 Ptos.]** Determine el estimador de momento del parámetro θ.

Solución

a) Se tiene Y~$G(\pi)$. **(0,5 pto.)**

$P_Y(y) = (1-\pi)^{(Y-1)} \cdot \pi$, y=1, 2, 3, 4, ...

Función de Verosimilitud:

$L(\pi) = \prod_{i=1}^{n} P_Y(y) = \prod_{i=1}^{n} (1-\pi)^{(y_i-1)} \cdot \pi = \pi^n \cdot \prod_{i=1}^{n} (1-\pi)^{(\sum_{i=1}^{n} y_i - n)}$ **(0,5 pto.)**

$l(\pi) = n \cdot \ln(\pi) + (\sum_{i=1}^{n} y_i - n) \cdot \ln(1-\pi) = n \cdot \ln(\pi) + (n\bar{y} - n) \cdot \ln(1-\pi) =$

$= n \cdot (\ln(\pi) + (\bar{y}-1) \cdot \ln(1-\pi))$ **(0,5 pto.)**

$\frac{\partial \ln(\pi)}{\partial \pi} = \frac{n}{\pi} + \frac{n \cdot (\bar{y}-1)}{1-\pi} \cdot (-1) = 0$ **(0,5 pto.)**

$\rightarrow \frac{n}{\pi} = \frac{n \cdot (\bar{y}-1)}{1-\pi} \rightarrow n \cdot \pi \cdot (\bar{y}-1) = n \cdot (1-\pi) \rightarrow \hat{\pi}_{MV} = \bar{Y}^{(-1)}$ **(0,5 pto.)**

b) $P(Y > 5) = 1 - P(Y \leq 5) = 1 - \pi \cdot \sum_{i=1}^{5} (1-\pi)^{(i-1)} = 1 - \pi \cdot \frac{1-(1-\pi)^5}{\pi} = (1-\pi)^5$

(0,5 pto.)

Por la propiedad de invarianza de los EMV:

$\widehat{(P > 5)}_{MV} = (1 - \bar{Y}^{(-1)})^5 = (\frac{\bar{Y}-1}{\bar{Y}})^5$ **(1,0 pto.)**

c) $E(Y) = \frac{(\theta + (-\theta))}{2} = 0 \rightarrow$ No se puede formar la primera ecuación de momentos **(0,5 pto.)**

$E(Y^2) = Var(Y) + (E(Y))^2 = Var(Y) + 0 \rightarrow E(Y^2) \equiv \frac{(\theta-(-\theta))^2}{12} = \frac{(2\theta)^2}{12} = \frac{\theta^2}{3}$

(0,5 pto.)

Así, $E(Y^2) = \frac{\theta^2}{3} \equiv \overline{Y^2}$

$\rightarrow \theta^2 = 3 \cdot \overline{Y^2} \rightarrow \hat{\theta}_{MV} = \sqrt{3} \cdot \sqrt{\overline{Y^2}}$ **(0,5 pto.)**

+ 1 Punto Base

Problema 3

La unidad de Business Intelligence de una empresa está midiendo el flujo de ventas de un cierto producto muy costoso. Debido a que mucha gente llega a la tienda, y algunos se interesan pero no lo llevan, es que para los expertos ha sido muy difícil proponer una distribución para el número de ventas mensuales de dicho producto. Finalmente, la unidad de análisis obtiene una muestra de tamaño cien, y propone tres distribuciones. En el anexo encontrará toda la información necesaria.

(a) Esboce un posible gráfico de cajón con bigotes para el número de ventas mensuales y comente.

(b) Comente sobre las tres distribuciones y decida cuál será la más apropiada. ¿Cuál es el valor del objeto `L.pois`?

(c) Según un análisis histórico, se prueba que los datos provienen de una distribución Poisson con parámetro λ igual a 8.5. El experto ofrece dos estimadores `E1` y `E2`, de los cuales se tiene información de la distribución muestral. Decida justificadamente qué estimador prefiere.

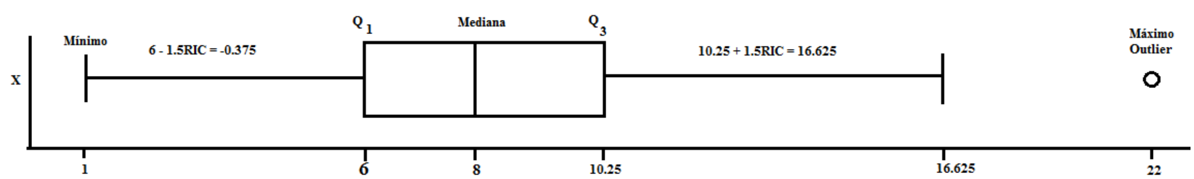
Solución

(a) Para construir el boxplot, deberá identificar $Q_1 = 6$ y $Q_3 = 10.25$ y crear la caja, con la Mediana como linea central.

Con los Cuartiles, se obtiene $RIC = 4.25$. Luego, deberá calcular el largo de los bigotes:

En el caso inferior, deberá ser $\max\{\text{Minimo}, Q_1 - 1.5RIC\} =$ Mínimo igual a 1.

En el caso superior, será $\min\{\text{Máximo}, Q_3 + 1.5RIC\} = 10.625$. Luego, el Máximo será un outiler.



Comentarios:
Hay un 25 % de sucursales que vendieron a lo más 4 productos.

Hay un 25 % de sucursales que vendieron a lo menos 10 productos.

La sucursal que más vendió productos llegó a vender 22 de ellos.

Es posible determinar asimetría, de donde se concluye que hay más dispersión sobre el 50 % de los datos, y el primer 50 % está más concentrado.

Al menos hay un dato atípico (máximo=22), sin embargo podría haber otros.

(b) A través de los comandos se logra identificar que las tres distribuciones son Poisson, Binomial Negativa y Geométrica.

La distribución Geométrica es la que peor se ajusta a estos datos.

La distribución Poisson y BinNeg son similares: la BinNeg subestima las probabilidades empíricas en la parte de más concentración, entregando valores bajo el fin de la barra. Por otro lado, la distribución

Poisson sobreestima las probabilidades de la parte central, ubicando sus valores por sobre las barras.

Puede justificar también por el contexto de la pregunta, y elegir Poisson o BinNeg.

El objeto `L.pois` es el EMV para el parámetro de una distribución Poisson. Sabemos que si $X \sim$Poisson(λ), entonces

$$\texttt{L.pois} = \hat{\lambda} = \bar{X} = 8.48$$

(c) Se conoce el valor real del parámetro: $\lambda = 8.5$.

El estimador $E2$ posee menor sesgo, por sobre $E1$.

El estimador $E1$ posee menor varianza, por sobre $E2$

Se debe decidir a partir de ECM:

$$\begin{aligned} ECM(E1) =& 0.200848 + (8.1375 - 8.5)^2 = 0.3322543 \\ ECM(E2) =& 0.570044 + (8.5525 - 8.5)^2 = 0.5728003 \end{aligned}$$

Por lo tanto, el mejor estimador sería $E1$

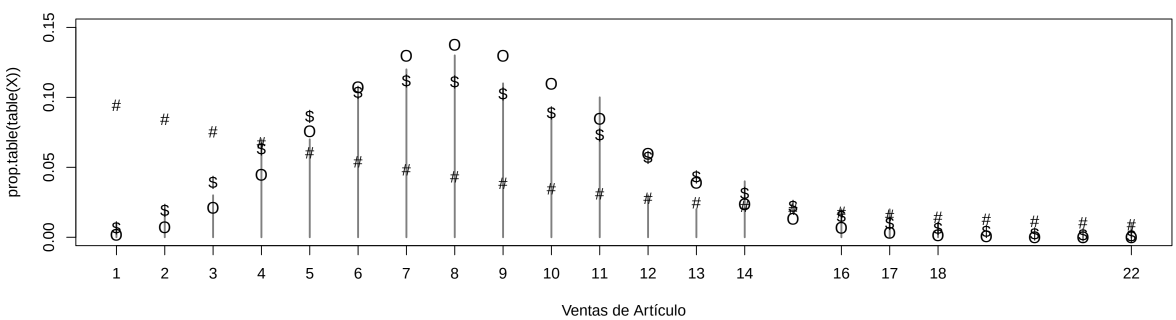
+ 1 Punto Base

Anexo Problema 3

```
# X: Numero de artículos vendidos
summary(X)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    6.00    8.00    8.48   10.25   22.00
sd(X)
[1] 3.713149

L.pois=fitdistr(X,"Poisson")$estimate
L.bn=fitdistr(X,"negative binomial")$estimate
L.geo=fitdistr(X,"geometric")$estimate

plot(prop.table(table(X)))
points(0:22,dpois(0:22,L.pois), pch="O")
points(0:22,dnbinom(0:22, size=L.bn[1], mu=L.bn[2]), pch="$")
points(0:22,dgeom(0:22, L.geo), pch="#")
```




```
# E1 y E2 vectores de estimadores de distintas muestras aleatorias.
mean(E1); mean(E2)
[1] 8.1375     8.5525
var(E1);  var(E2)
[1] 0.200848    0.570044
```