

parcial 1

Kevin Valencia Romero y Tatiana Mora Acosta

2022-03-24

1. La tabla mostrada contiene valores de población, en cientos de miles, de las diez ciudades más pobladas de cuatro países en el año 1967:

Argentina	EEUU	Holanda	Japón
29.66	77.81	8.68	110.21
7.61	35.50	7.31	32.14
6.35	24.79	6.02	18.88
4.10	20.02	2.64	16.38
3.80	16.70	1.75	13.37
2.75	9.39	1.72	11.92
2.70	9.38	1.51	10.71
2.69	8.76	1.42	7.80
2.51	7.63	1.31	7.70
2.44	7.50	1.29	7.00

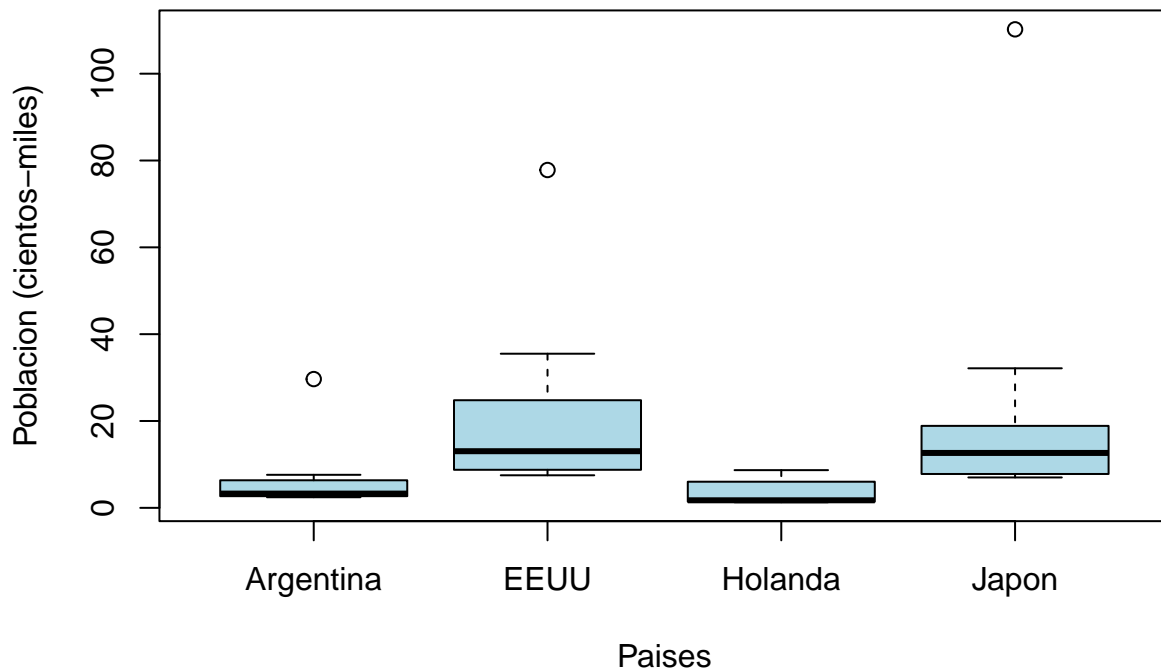
```
poblacion<- c(29.66,7.61,6.35,4.10,3.80,2.75,2.70,2.69,2.51,2.44,77.81,35.50,24.79,20.02,16.70,9.39,9.39)
```

- a) Construya un box-plot e identifique los puntos extremos. ¿Cuáles son las características más sobresalientes? ¿Hay outliers?

```
pais<-c("Argentina","Argentina","Argentina","Argentina","Argentina","Argentina","Argentina","Argentina")

boxplot(poblacion~pais, horizontal = FALSE, main = "Diagrama de cajas y bigotes paises",
        xlab = "Paises", ylab = "Poblacion (cientos-miles)",col = "lightblue")
```

Diagrama de cajas y bigotes paises



En Argentina la población de cientos de miles es menor que 10 y Holanda tiene una población un poco menor a la de Argentina, mientras que EEUU se lleva la mayor población entre todo el resto de los países aproximadamente tiene un poco menos de 40 y por último el país de Japón tiene un poco más de 30 de población pero no tiene tanta como la de EEUU.

Se puede apreciar que solo 3 países presentan datos atípicos uno por cada país, el país de Japón es el que tiene el dato atípico más alejado de los bigotes a diferencia de los otros países, y el que más cerca tiene el dato atípico a sus bigotes es Argentina.

- b) Compare los centros de cada población, sus dispersiones y su simetría. ¿Cuál es el país más homogéneamente habitado? ¿Cuáles son las características más sobresalientes? ¿Hay outliers?

Analizando el gráfico de EEUU se puede apreciar que a partir del cuartil Q2 hasta el Q3 los datos se encuentran muy dispersos más que el resto de países, en el país de Japón los datos entre cuartil Q1-Q2 y Q2-Q3 no son tan dispersos como el anterior y antes y después de la mediana tienen la misma dispersión, y en los países de Argentina y Holanda la dispersión es mucho menor que la de Japón, los dos países tienen poca dispersión después de la mediana.

Respecto a la simetría de cada uno de los países, podemos apreciar que el único país que es simétrico ya que la mediana se sitúa en el centro de la caja, a diferencia de los países como Argentina, EEUU y Holanda son asimétricamente positivos ya que la parte más larga de la caja es la parte superior a la mediana.

El país más homogéneamente habitado es Japón ya que tiene la misma dispersión de datos antes de la mediana y después de la mediana.

Como se está analizando los centros de cada población dentro de ello no hay datos atípicos, los únicos datos atípicos que apreciamos son los que tienen Argentina, EEUU, Japón que están por fuera de los límites.

2. Avianca se encuentra estudiando la situación de venta de puesto en el vuelo Armenia – Bogotá puesto que ha identificado que el 2% de las personas que reservaron puesto, no se presentan a la hora del embarque. Teniendo en cuenta que el avión A320 utilizado en esta ruta tiene una capacidad para 180 pasajeros, la empresa decide vender 181 tiquetes. ¿Cuál es la probabilidad de que todas las personas que llegan a embarque puedan tener su silla asegurada y no se presenten quejas por “sobre-vender” sillas ante la aeronáutica civil? Asuma que las llegadas de los pasajeros siguen una distribución Poisson.

```
lambda <- 180 * 0.98
```

```
#la probabilidad de que todas las personas que llegan a embarque puedan tener su silla asegurada y no s  
probabilidad2 <- ppois(180,lambda);probabilidad2
```

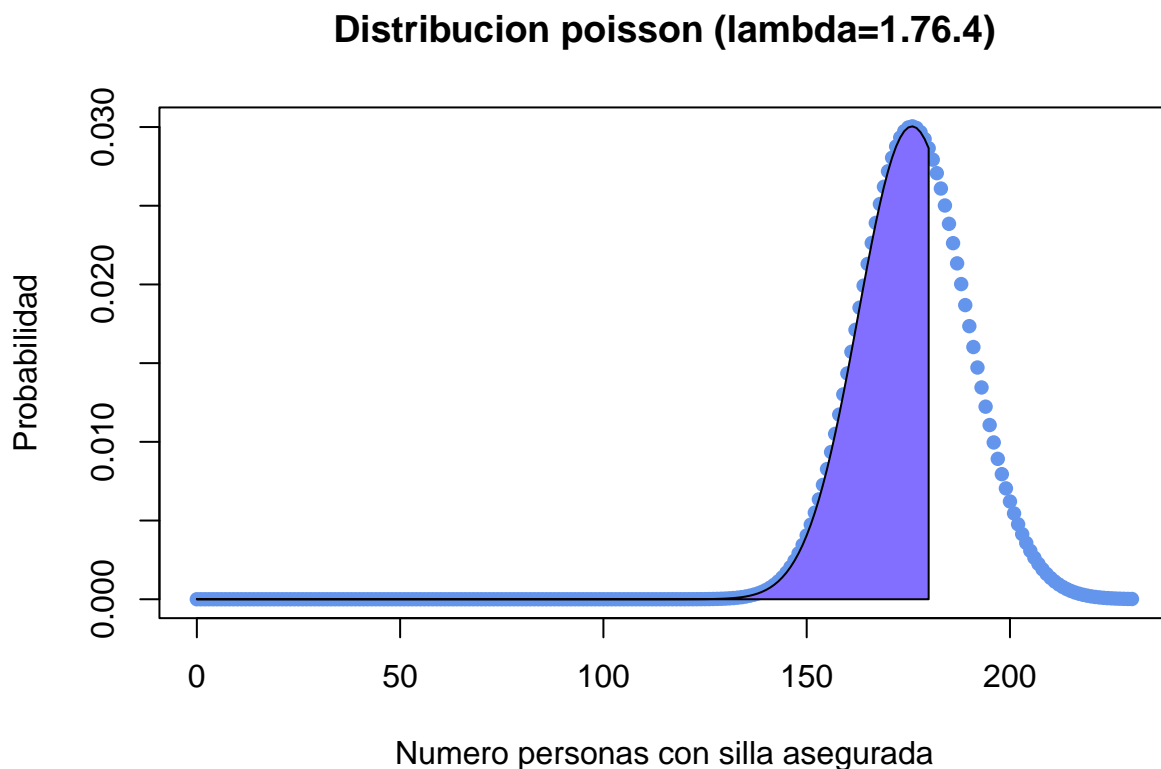
```
## [1] 0.6255245
```

```
longitud2<-c(0:230)
```

```
probabilidadPunto2<-dpois(longitud2,lambda)
```

```
plot(longitud2,probabilidadPunto2, main="Distribucion poisson (lambda=1.76.4)",  
xlab="Numero personas con silla asegurada",  
ylab="Probabilidad",  
pch=16,  
col = "cornflowerblue")
```

```
polygon(c(longitud2[longitud2 <= 180 ], 180),c(probabilidadPunto2[longitud2 <= 180 ], 0),  
col = "slateblue1",  
border = 1)
```



3. En la planta envasadora de Coca-Cola FEMSA ubicada en Tocancipá se realiza el embotellado de la Coca Cola presentación de 400 mL (en promedio), con una desviación estándar de 5 mL (mililitros). Si se sabe por experiencia que este proceso de embotellado sigue una distribución normal, y también se sabe que todo producto con más de 415 mL es declarado como NO CONFORME. Determine el porcentaje de las botellas de gaseosa que son declaradas producto NO CONFORME

```
media<- 400
desvE<- 5

probabilidad3<- 1-( pnorm (415, media, desvE));probabilidad3
```

```
## [1] 0.001349898
```

```
#El porcentaje de las botellas de gaseosa que son declaradas producto NO CONFORME es
respuesta<-(probabilidad3*100);respuesta
```

```
## [1] 0.1349898
```

```
longitud3<-c(0:450)
probabilidadPunto3<-dnorm(0:450,media,desvE)

plot(longitud3,probabilidadPunto3, main="Distribucion normal (media=400, desvE=5)",
xlab="Porcentaje de botellas NO CONFORME",
ylab="Probabilidad",
pch=16,
col = "cornflowerblue")

polygon(c(longitud3[longitud3 >= 416 ], 416),c(probabilidadPunto3[longitud3 >= 416], 0),
col = "slateblue1",
border = 1)
```

Distribucion normal (media=400, desvE=5)

