

Laboratorio 1

Antonio J. Perán

29 de septiembre de 2017

Lectura de datos

```
titanic <- read.csv(file = "http://eio.usc.es/pub/pateiro/AE/data/titanic_es.csv",
                    header = T, sep = ";", dec = ",")
pandoc.table(head(titanic), caption = "Primeras seis observaciones del conjunto de datos")
```

Table 1: Primeras seis observaciones del conjunto de datos

clase	sobreviviente	sexo	edad	tarifa	embarque
primera	1	mujer	29	211.3	Southampton
primera	1	hombre	0.9167	151.6	Southampton
primera	0	mujer	2	151.6	Southampton
primera	0	hombre	30	151.6	Southampton
primera	0	mujer	25	151.6	Southampton
primera	1	hombre	48	26.55	Southampton

Ejercicios

1. ¿Cuántos pasajeros aparecen registrados en el conjunto de datos?

```
numpas <- nrow(titanic)
numpas
```

```
## [1] 1043
```

En la base de datos aparecen registrados 1043 pasajeros.

2. ¿Cuántos pasajeros sobrevivieron al naufragio?

```
numsv <- sum(titanic$sobreviviente)
numsv
```

```
## [1] 425
```

El número de pasajeros que sobrevivieron al naufragio es 425.

3. Calcula el porcentaje de pasajeros que sobrevivió al naufragio.

```
# numsv y numpas están definidas anteriormente
persv <- numsv/numpas*100
persv
```

```
## [1] 41
```

El porcentaje de pasajeros que sobrevivió al naufragio es 40.75%.

4. ¿Cuántos pasajeros viajaban en primera clase?

```
pas1clase <- sum(titanic$clase == "primera")
pas1clase
```

```
## [1] 282
```

El número de pasajeros que viajaban en primera clase es 282.

5. ¿Cuántos niños (menores de 12 años) aparecen registrados?

```
niñosmenos12 <- sum(titanic$edad < 12)
niñosmenos12
```

```
## [1] 91
```

Aparecen registrados 91 niños menores de 12 años.

6. Calcula las frecuencias relativas para las distintas clases de pasajeros.

```
frel <- table(titanic$clase)/nrow(titanic)
frel
```

```
##
## primera segunda tercera
## 0.27 0.25 0.48
```

Las frecuencias relativas según clase se muestran en la tabla superior.

7. Representa con un gráfico adecuado la distribución de las distintas clases de pasajeros.

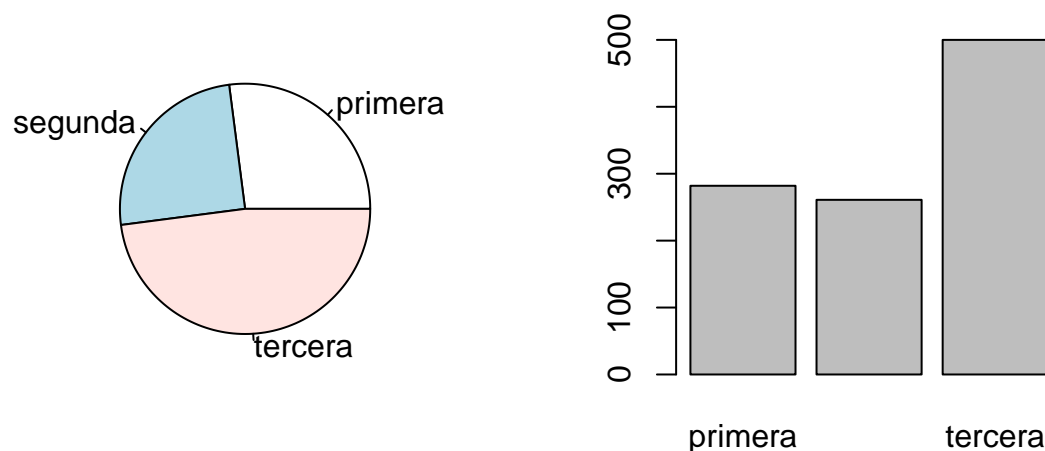
A continuación se muestran dos gráficos que representan la distribución de las distintas clases de pasajeros. En mi opinión, el más adecuado sería el gráfico de barras.

```
par(mfrow = c(1,2), oma = c(0, 0, 2, 0))

pie(table(titanic$clase))
barplot(table(titanic$clase))

title("Distribución de la variable clase.", outer = TRUE)
```

Distribución de la variable clase.



8. Representa con un gráfico adecuado la distribución de la edad de los pasajeros registrados.

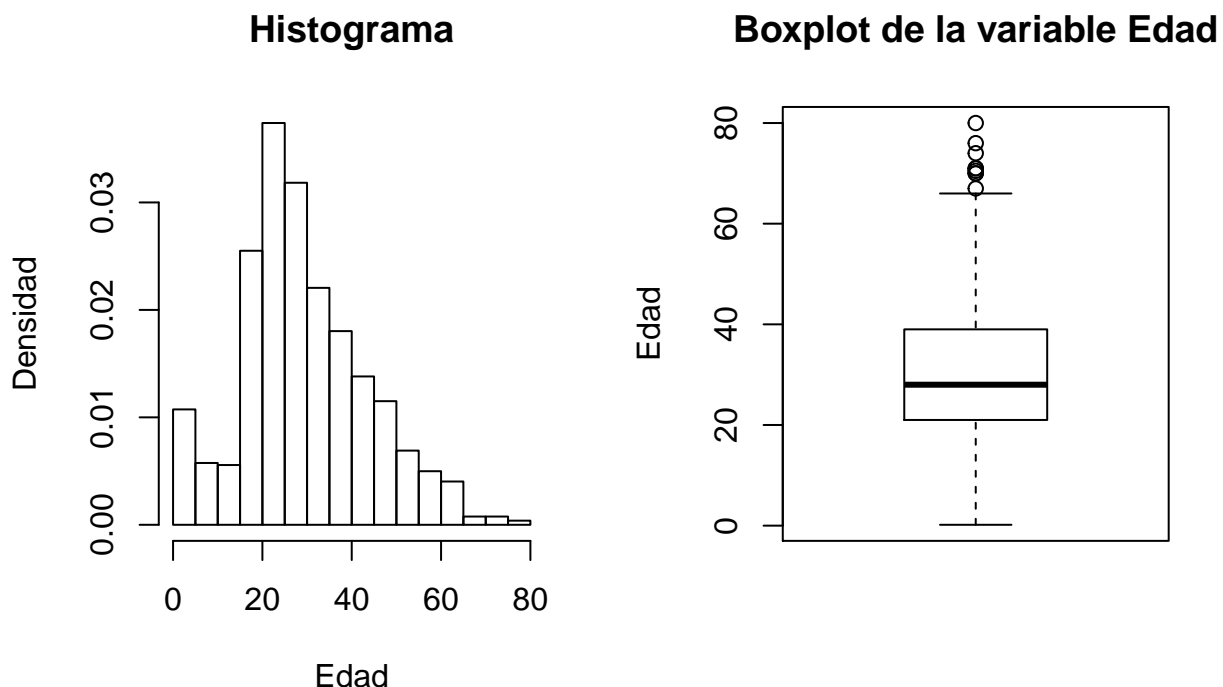
Para representar la distribución de una variable continua podemos usar, o bien un histograma, que nos daría una idea aproximada de la forma de la función de densidad, o bien un boxplot o gráfico de cajas. A continuación se muestra un histograma y un boxplot que representan la distribución de la edad de los pasajeros.

```
par(mfrow = c(1,2), oma = c(0, 0, 2, 0))

hist(titanic$edad, freq = FALSE, main = 'Histograma', xlab = "Edad", ylab = "Densidad" )
boxplot(titanic$edad, main = "Boxplot de la variable Edad", ylab = "Edad")

title("Distribución de la variable edad.", outer = TRUE)
```

Distribución de la variable edad.



9. ¿Cuál era la edad media de los pasajeros? ¿Y la mediana?

```
meanEdad <- mean(titanic$edad)
medianEdad <- median(titanic$edad)
```

La edad media de los pasajeros es de 29.81 años, mientras que la mediana es de 28 años.

10. ¿Cuál era el precio medio de un pasaje en el Titanic?

```
meanBillete <- mean(titanic$tarifa)
```

El precio medio del billete era de 36.6 libras.

11. La frase “Mujeres y niños primero” hace referencia a un protocolo histórico por el que las mujeres y los niños debían ser los primeros en ser salvados en una emergencia. Según los datos, ¿qué porcentaje de mujeres sobrevivió al naufragio? ¿qué porcentaje de hombres sobrevivió al naufragio? ¿qué porcentaje de niños (menores de 12 años) sobrevivió al naufragio?

```

supervivientes <- titanic[titanic$sobreviviente == 1, ]
persvsexo      <- table(supervivientes$sexo)/table(titanic$sexo)*100

niñossv        <- nrow(supervivientes[supervivientes$edad < 12, ])
niños          <- nrow(titanic[titanic$edad < 12, ])
persvniños     <- niñossv/niños*100

```

El porcentaje de mujeres que sobrevivió fue del 75.13%, el de hombres del 20.55%, y el de niños menores a 12 años del 56.04%.

12. ¿Es la tasa de supervivencia de los hombres que viajaban en primera clase superior a la de los hombres que viajaban en segunda? ¿Es superior a las de los hombres que viajaban en tercera clase?

```

hombresv       <- supervivientes[ supervivientes$sexo == "hombre", ]
hombres        <- titanic[ titanic$sexo == "hombre", ]
TSVhombresclase <- table(hombresv$clase)/table(hombres$clase)*100
TSVhombresclase

```

```

##
## primera segunda tercera
##      35      15      17

```

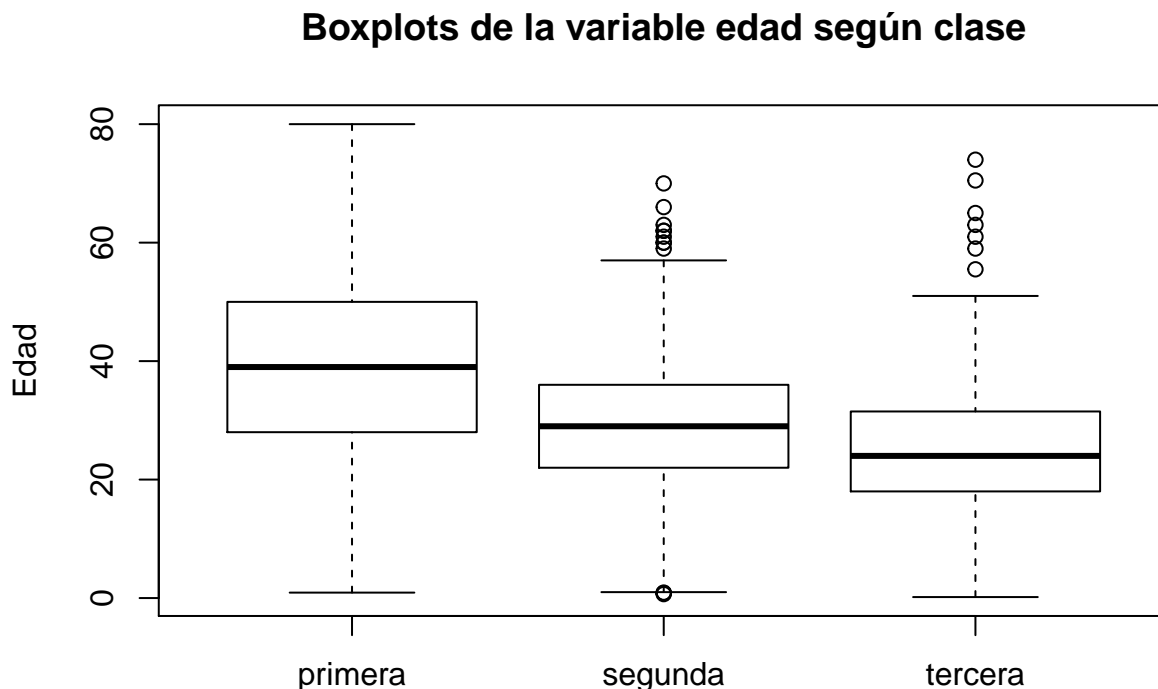
En la tabla de arriba se muestran las tasas de supervivencia de los hombres según clase en %. Según estos resultados, la tasas de supervivencia de los hombres según clase estarían ordenadas, de mayor a menos como: primera, tercera, segunda.

13. Representa mediante un diagrama de cajas las edades de los pasajeros agrupadas por clase. ¿Qué observas?

```

boxplot(titanic$edad ~ titanic$clase, main = "Boxplots de la variable edad según clase",
        ylab = "Edad")

```



Los pasajeros de primera clase son los de mayor edad media y un mayor rango intercuartílico. Los pasajeros de segunda y tercera clase son similares, con una menor media de edad con respecto a los de primera clase y un menor rango intercuartílico. Por otro lado, entre los pasajeros de segunda y tercera clase encontramos diversos valores atípicos.

14. Calcula la media y la varianza de las edades en cada clase.

```
meanClass <- with(titanic, tapply(edad, clase, mean))
varClass  <- with(titanic, tapply(edad, clase, var))
```

La media de edad en cada clase es de: `xtable(meanClass)`

La varianza de la edad en cada clase es de: 211.29, 186.01, 140.73

15. En el barco se proporcionaba asistencia especial al 5% de los pasajeros de mayor edad. ¿Qué edad debía tener un pasajero para recibir asistencia especial? ¿Qué edad tenía el pasajero más mayor?

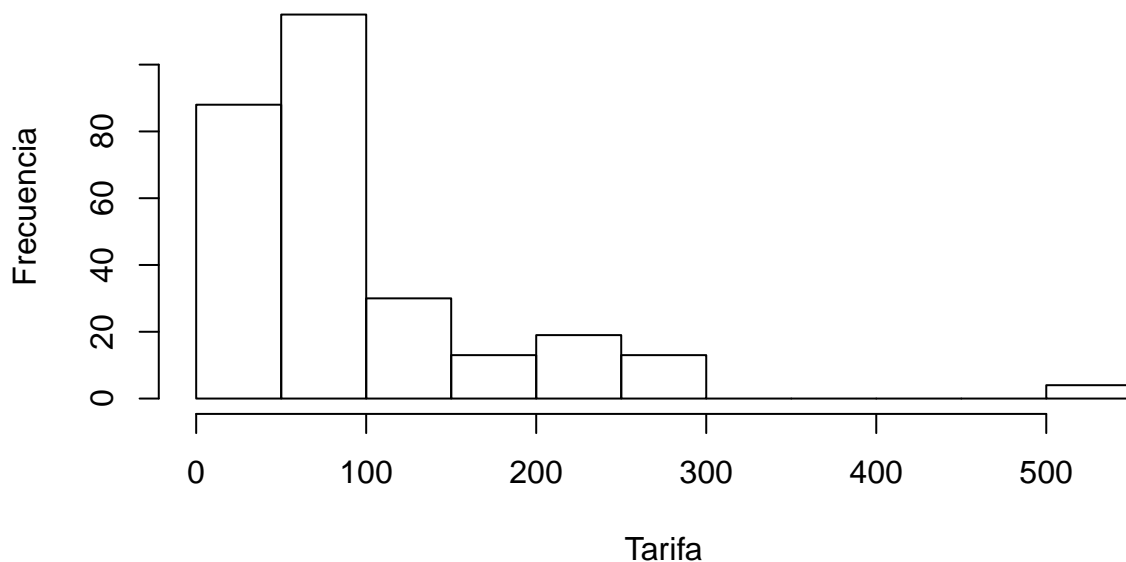
```
edadasisesp <- quantile(titanic$edad, 0.95)
maxedad     <- max(titanic$edad)
```

Los pasajeros que recibían asistencia especial son los que tenían una edad igual o superior a 56 años. El pasajero de mayor edad tenía 80 años.

16. Dibuja un histograma para el precio del pasaje en primera clase.

```
hist(titanic$tarifa[titanic$clase == "primera"],
     main = "Histograma de la variable tarifa para pasajeros de primera clase",
     xlab = "Tarifa", ylab = "Frecuencia")
```

Histograma de la variable tarifa para pasajeros de primera clase



17. ¿Cuál es la moda para el puerto de embarque?

```
frecEmbarque <- table(titanic$embarque)

# Buscamos el nombre del puerto con mayor frecuencia de embarque
moda <- dimnames(frecEmbarque)[[1]][which(frecEmbarque == max(frecEmbarque))]
```

La moda para el puerto de embarque es Southampton.

18. Los pasajeros que pagaban menos por su billete debían dormir en los camarotes más modestos del barco. En concreto, el 5% de los pasajeros que menos pagaban viajaban en dichos camarotes. ¿Cuál era el precio máximo que se pagaba por uno de estos camarotes?

```
preciomax <- quantile(titanic$tarifa, 0.05)
```

El precio máximo que se pagaba era de `preciomax` libras.

19. Construye una tabla de contingencia para la distribución de la clase del pasajero por sexo.

```
table(titanic$clase, titanic$sexo)
```

```
##
##           hombre mujer
## primera      151   131
## segunda      158   103
## tercera      348   152
```

20. Representa en un mismo gráfico la distribución de la clase del pasajero por sexo.

```
primera <- titanic[titanic$clase == "primera", ]
segunda <- titanic[titanic$clase == "segunda", ]
tercera <- titanic[titanic$clase == "tercera", ]

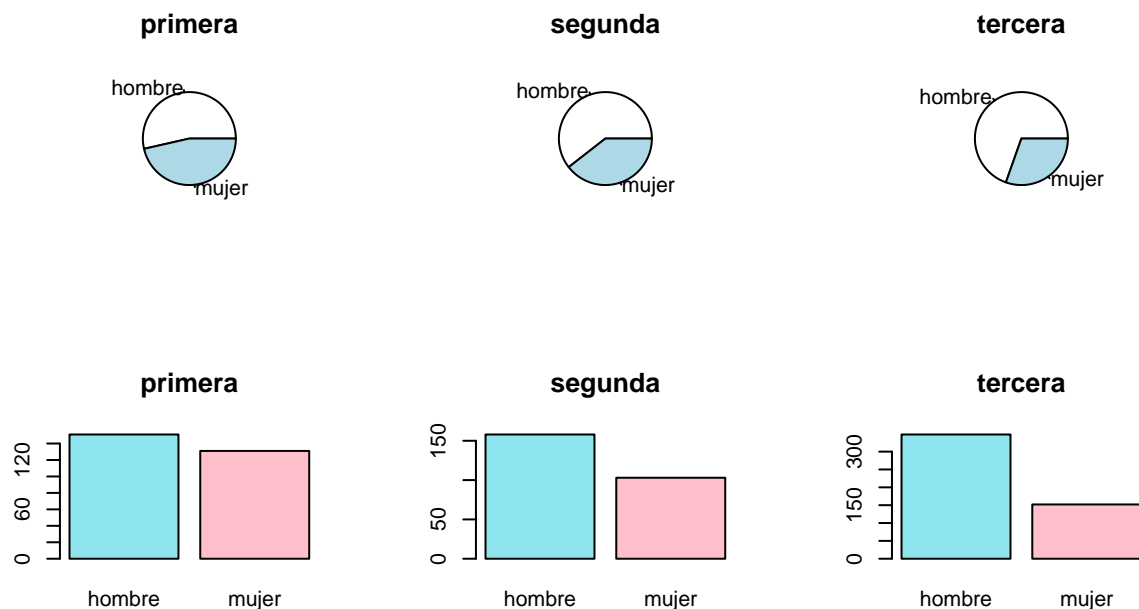
par(mfrow = c(2,3), oma = c(0, 0, 2, 0))

pie(table(primera$sexo), main = "primera")
pie(table(segunda$sexo), main = "segunda")
pie(table(tercera$sexo), main = "tercera")

barplot(table(primera$sexo), main = "primera", col = c("cadetblue2", "pink"))
barplot(table(segunda$sexo), main = "segunda", col = c("cadetblue2", "pink"))
barplot(table(tercera$sexo), main = "tercera", col = c("cadetblue2", "pink"))

title("Distribución de la clase del pasajero por sexo", outer = TRUE)
```

Distribución de la clase del pasajero por sexo



21. Calcula los cuartiles para la edad.

```
quantile(titanic$edad)
```

```
##      0%    25%    50%    75%   100%  
##  0.17 21.00 28.00 39.00 80.00
```

22. ¿Pagaban billete los menores de un año?

```
any(titanic$tarifa[titanic$edad < 1] > 0)
```

```
## [1] TRUE
```

Sí, los menores de un año también pagaban billete.