

Laboratorio: Introducción a la estadística usando R

Edgar Seijo Otero

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

Comando:

```
npasajeros <- nrow(titanic)
npasajeros
```

Resultado:

```
1043
```

2. ¿Cuántos pasajeros sobrevivieron al naufragio?

Comando:

```
nsupervivientes <- sum(titanic$sobreviviente == 1)
nsupervivientes
```

Resultado:

```
425
```

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

Comando:

```
psupervivientes <- nsupervivientes/npasajeros*100
psupervivientes
```

Resultado:

```
40.74784
```

4. ¿Cuántos pasajeros viajaban en primera clase?

Comando:

```
nprimera = sum(titanic$clase == "primera")
nprimera
```

Resultado:

```
282
```

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

Comando:

```
nmenores = sum(titanic$edad < 12)
nmenores
```

Resultado:

```
91
```

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

Comando:

```
fabsolutas <- table(titanic$clase)
frelativas <- fabs/sum(fabs)
frelativas
```

Resultado:

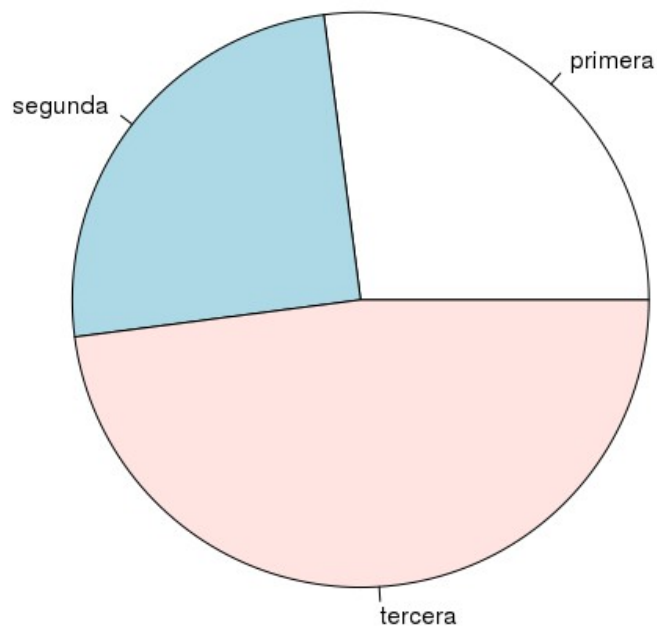
```
    primera    segunda    tercera
0.2703739  0.2502397  0.4793864
```

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

Comando:

```
pie(frelativas)
```

Resultado:

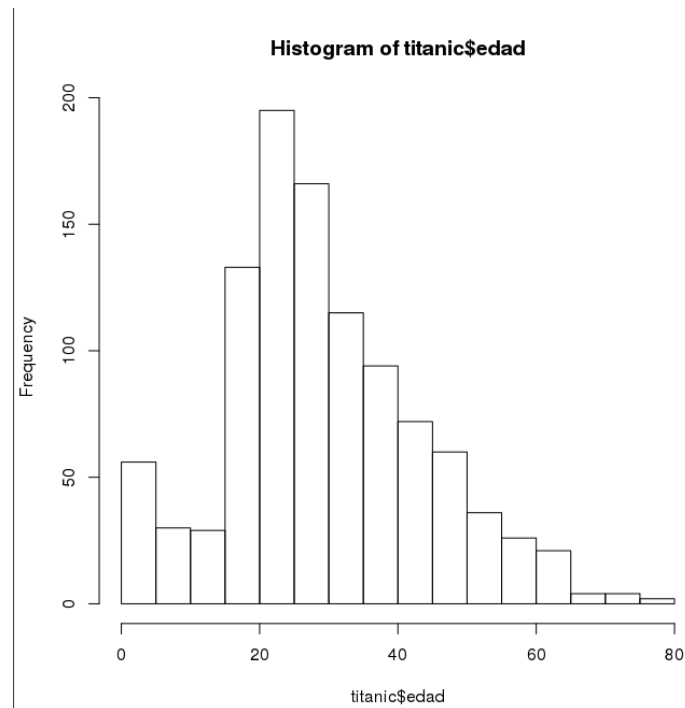


8. Representa con un gráfico adecuado la distribución de la edad de los pasajeros registrados.

Comando:

```
hist(titanic$edad)
```

Resultado:



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

Comando:

```
mean(titanic$edad)  
median(titanic$edad)
```

Resultado:

```
29.8132  
28
```

La edad media de los pasajeros era de 29,8132 años y la mediana de 28.

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

Comando:

```
mean(titanic$tarifa)
```

Resultado:

```
36.60302
```

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?

Comando:

```
## Porcentaje de mujeres supervivientes

# Numero total de mujeres
nmujeres <- sum(titanic$sexo == "mujer")
# Numero de mujeres supervivientes
nmujeressup <- sum(titanic$sexo == "mujer" & titanic$sobreviviente == 1)
# Porcentaje
pmujeressup <- nmujeressup/nmujeres*100

## Porcentaje de hombres supervivientes

# Numero total de hombres
nhombres <- sum(titanic$sexo == "hombre")
# Numero de hombres supervivientes
nhombressup <- sum(titanic$sexo == "hombre" & titanic$sobreviviente == 1)
# Porcentaje
phombressup <- nhombressup/nhombres*100

## Porcentaje de menores de 12 supervivientes

# Numero total de menores de 12 calculado anteriormente
# Numero de menores de 12 supervivientes
nmenoressup <- sum(titanic$edad < 12 & titanic$sobreviviente == 1)
# Porcentaje
pmenoressup <- nmenoressup/nmenores*100

pmujeressup
phombressup
pmenoressup
```

Resultado:

```
75.12953
20.54795
56.04396
```

Sobrevivieron al naufragio un 75,13% de las mujeres, un 20,55% de los hombres y un 56,04% de los niños.

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?

Comando:

```
tapply(titanic$sexo == 'hombre' & titanic$sobreviviente == 1,
titanic$clase, sum)/table(titanic$clase[titanic$sexo == 'hombre'])*100
```

Resultado:

```
primera segunda tercera
35.09934 14.55696 16.95402
```

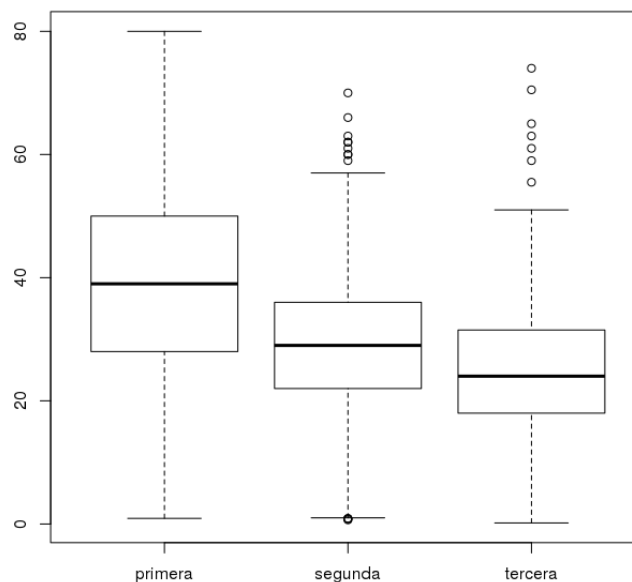
Podemos observar que la tasa de supervivencia de los hombres que viajaban en primera es superior a la de aquellos que viajaban en segunda y tercera. En el caso de los que viajaban en segunda, la tasa es menor (aunque similar) a la de los de tercera clase.

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

Comando:

```
boxplot(titanic$edad ~ titanic$clase)
```

Resultado:



Se observa que la edad de los pasajeros tiende a ser más baja cuanto más baja es la clase (la media de edad en primera se sitúa en torno a los 40 años, en segunda cerca de los 30 y en tercera hacia los 20 años). Esto podría deberse a que la mayor parte de pasajeros jóvenes eran de familias modestas que buscaban nuevas oportunidades laborales al otro lado del Atlántico.

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

Comando:

```
# Media
tapply(titanic$edad, titanic$clase, mean)
# Varianza
tapply(titanic$edad, titanic$clase, var)
```

Resultado:

```
primera segunda tercera
39.08304 29.50670 24.74500
211.2852 186.0122 140.7283
```

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?

Comando:

```
quantile(titanic$edad, seq(0, 1, by=0.05))[20]  
max(titanic$edad)
```

Resultado:

```
95%  
56
```

```
80
```

A través del cálculo del cuantil 0,95, obtenemos la edad que deja a su izquierda el 95% de las edades de los pasajeros. Por lo tanto, recibirían asistencia especial los pasajeros con edades a partir de 56 años.

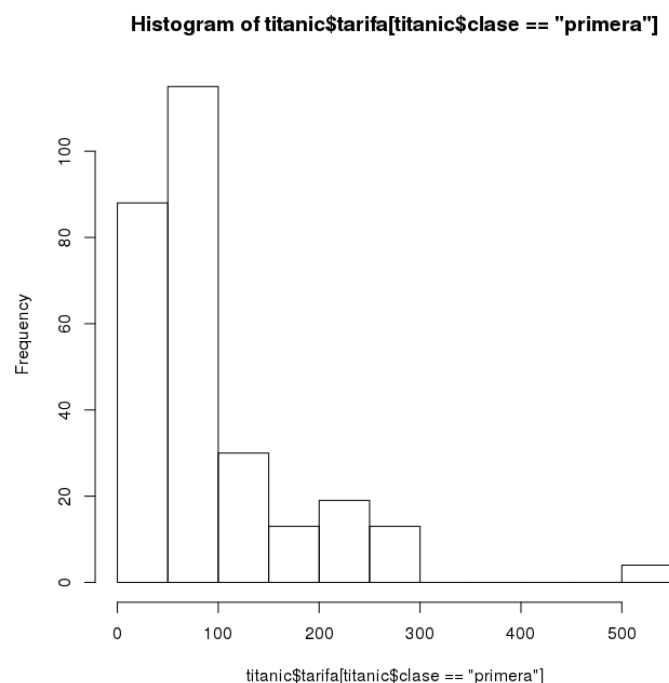
El pasajero más mayor tenía 80 años.

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

Comando:

```
hist(titanic$tarifa[titanic$clase == "primera"])
```

Resultado:



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

Comando:

```
which(table(titanic$embarque) == max(table(titanic$embarque)))
```

Resultado:

```
Southampton  
3
```

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?

Comando:

```
quantile(titanic$tarifa, seq(0, 1, by=0.05))[2]
```

Resultado:

```
5%  
7.2292
```

A través del cálculo del cuantil 0,05, obtenemos el precio del billete que deja a su izquierda el 5% de los precios más bajos. Por lo tanto, el precio máximo del billete para un camarote modesto era de 7,2292 libras.

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

Comando:

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

Resultado:

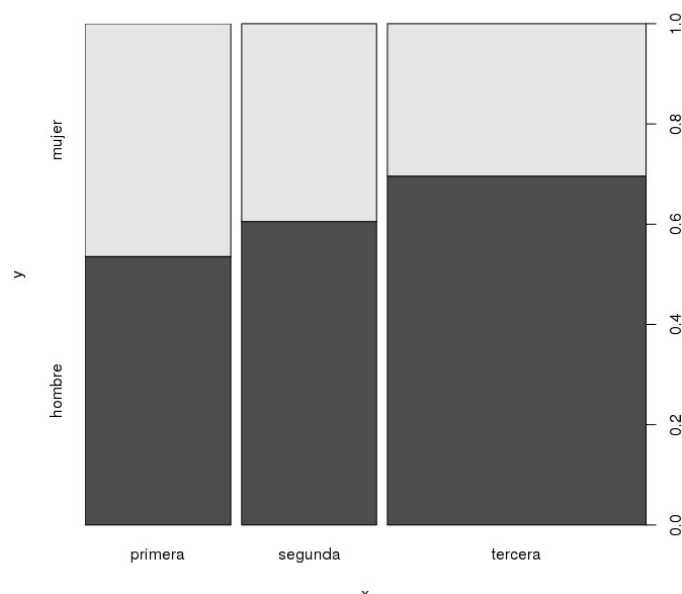
	hombre	mujer
primera	151	131
segunda	158	103
tercera	348	152

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

Comando:

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

Resultado:



21. Calcula los cuartiles para la edad.

Comando:

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

Resultado:

0%	25%	50%	75%	100%
0.1667	21.0000	28.0000	39.0000	80.0000

Q1: 21

Q2: 28

Q3: 39

22. ¿Pagaban billete los menores de un año?

Comando:

```
titanic$tarifa[titanic$edad < 1]
```

Resultado:

151.5500	29.0000	14.5000	18.7500	27.7500	9.3500	19.2583	19.2583
14.4000	20.5750	13.7750	8.5167				

Podemos observar, sacando los datos de los precios de los billetes para los pasajeros menores de un año, que si que tenían que pagar billete.