

Pauta Evaluación 3

Natalie Julian

Visualización de información sobre fallos cardíacos

El archivo de datos **heart** contiene datos de pacientes de distintas variables, incluyendo la muerte o no muerte por falla cardíaca. La información es la siguiente:

- **age** Indica la edad del paciente en años
- **diabetes** Indica si el paciente presentaba (1) diabetes o no (0)
- **high_pressure** Indica si el paciente presentaba presión arterial alta (1) o no (0)
- **serum_sodium** Indica el nivel de sodio sérico en la sangre
- **smoking** Indica si el paciente fumaba (1) o no
- **death** Indica si el paciente presentó muerte por falla cardíaca (1) o no (0)

En las siguientes preguntas se le pide generar gráficos donde se visualicen los datos de manera informativa y adecuada, y además, **debe entregar inferencias y comentarios para cada uno de estos gráficos**. Puede apoyarse utilizando estadísticas descriptivas o tablas si así lo desea.

- a) Con el fin de analizar una posible asociación entre presentar presión alta y presentar muerte por fallo cardíaco, realice dos gráficos, uno para pacientes que no murieron por fallo cardíaco y otro para pacientes que sí murieron por fallo cardíaco. En cada gráfico debe visualizarse la fracción de pacientes que presentaban y no presentaban presión arterial alta. Es decir, para el grupo de pacientes que presentaron fallo cardíaco visualizar la fracción de casos con presión alta y sin presión alta, asimismo para el grupo de pacientes que no presentaron fallo cardíaco. ¿Se observan diferencias? ¿Para qué grupo se observa menor el número de casos con presión arterial alta? Comente.

Una solución: Gráfico de torta

realizar los gráficos (no es necesario que se muestren los porcentajes, lo importante es que se visualice la fracción, proporción, área de cada grupo), estos gráficos deben llevar título

```
library(rio)
heart <- import("heart.csv")

#Para fallecidos cuántos tuvieron presion alta y cuántos no (porcentaje)
table(heart$high_pressure[which(heart$death==1)]/nrow(heart[which(heart$death==1),]))

##
##      0      1
## 0.59375 0.40625

#Para no fallecidos cuántos tuvieron presion alta y cuántos no (porcentaje)
table(heart$high_pressure[which(heart$death==0)]/nrow(heart[which(heart$death==0),]))

##
##      0      1
## 0.6748768 0.3251232
```

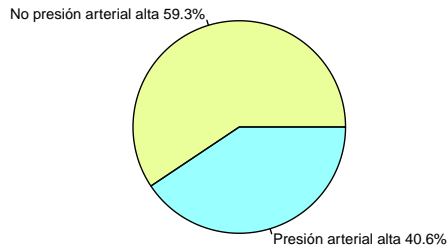
```

par(mfrow=c(1,2))
pie(table(heart$high_pressure[which(heart$death==1)]),
     labels=c("No presión arterial alta 59.3%", "Presión arterial alta 40.6%"),
     col=rainbow(2, alpha=0.4, start=0.2),
     main="Casos de muerte por falla cardiaca",
     cex=0.7)

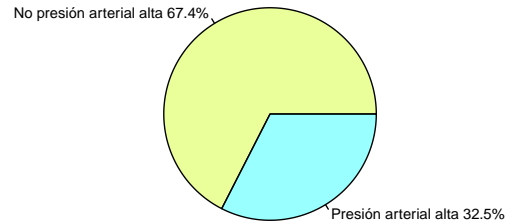
pie(table(heart$high_pressure[which(heart$death==0)]),
     labels=c("No presión arterial alta 67.4%", "Presión arterial alta 32.5%"),
     col=rainbow(2, alpha=0.4, start=0.2),
     main="Casos de no muerte por falla cardiaca",
     cex=0.7)

```

Casos de muerte por falla cardiaca



Casos de no muerte por falla cardiaca



realizar comentario general entre ambos gráficos

Observando ambos gráficos se observa que para el grupo de no fallecidos existe una mayor proporción de casos sin presión arterial alta, de todas formas esta diferencia no es tan radical.

realizar alguna inferencia adicional: relaciona el contexto del ejercicio o genera conjeturas adicionales o complementa con estadísticas adicionales y comenta, etcétera

Dado el contexto podríamos pensar que quizás pudiera existir alguna asociación entre el evento morir de falla cardiaca y presentar presión arterial alta. Observando los gráficos y cifras se puede observar que el porcentaje de pacientes con presión arterial alta aumenta aproximadamente un 8% en los casos de muerte por falla cardiaca, además, en ambos grupos (casos de muerte y no muerte por falla cardiaca) se observa que hay una mayor proporción de casos no presión arterial alta.

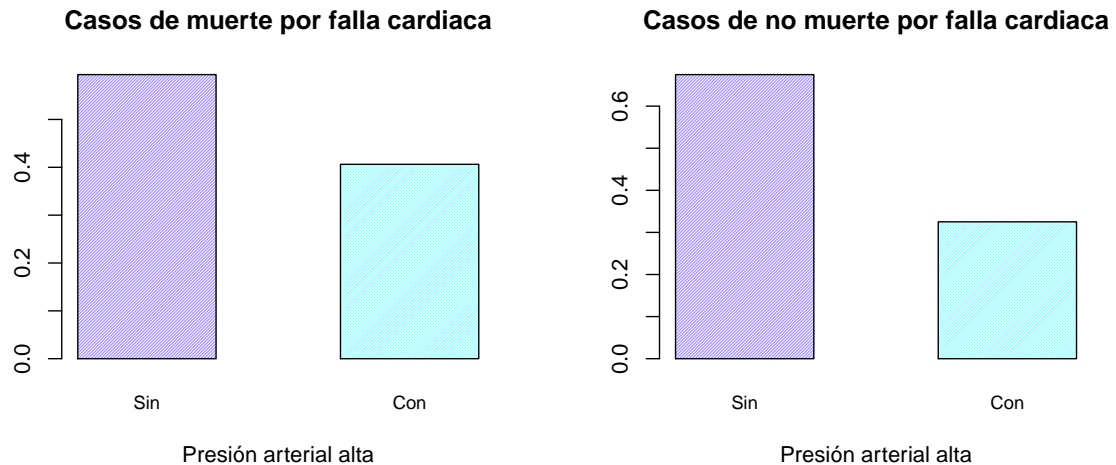
Otra solución con gráfico de barras

```

par(mfrow=c(1,2))
barplot(table(heart$high_pressure[which(heart$death==1)]/nrow(heart[which(heart$death==1)]),
         names.arg=c("Sin", "Con"),
         col=rainbow(2, alpha=0.4, start=0.7),
         main="Casos de muerte por falla cardiaca",
         cex.names = 0.8,
         xlab="Presión arterial alta",
         space=0.9,
         density=60)

```

```
barplot(table(heart$high_pressure[which(heart$death==0)]/nrow(heart[which(heart$death==0)]),
  names.arg=c("Sin", "Con"),
  col=rainbow(2, alpha=0.4, start=0.7),
  main="Casos de no muerte por falla cardiaca",
  cex.names=0.8,
  xlab="Presión arterial alta",
  space=0.9,
  density=60)
```



Otra solución (comando plotrix)

```
#install.packages("plotrix")
library(plotrix)

frecuencias<-data.frame(table(heart$high_pressure[which(heart$death==1)]))

slices <- frecuencias$Freq

par(mfrow=c(1,2))

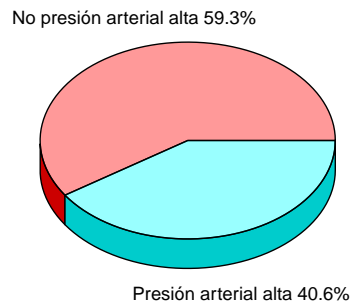
pie3D(slices,labels=c("No presión arterial alta 59.3%", "Presión arterial alta 40.6%"),
  main="Muerte por falla cardiaca", col=rainbow(2, alpha=0.4), theta=pi/3, labelcex=0.8)

frecuencias<-data.frame(table(heart$high_pressure[which(heart$death==0)]))

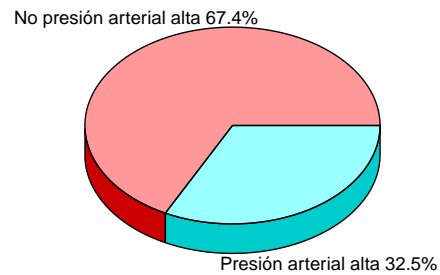
slices <- frecuencias$Freq

pie3D(slices,labels=c("No presión arterial alta 67.4%", "Presión arterial alta 32.5%"),
  main="No muerte por falla cardiaca", col=rainbow(2, alpha=0.4), theta=pi/3, labelcex=0.8)
```

Muerte por falla cardiaca



No muerte por falla cardiaca



- b) La edad suele ser una variante de interés en bioestadística, puesto que, a edades avanzadas suelen desencadenarse patologías en los pacientes asociadas al desgaste del cuerpo debida al tiempo. ¿Entre qué rango de edad se encuentran los pacientes? Muestre en un gráfico la distribución de pacientes por edad, es decir, que se observe la cantidad de pacientes para cada una de las edades. Este gráfico debe estar ordenado por frecuencia absoluta de mayor a menor. ¿En qué edades se concentra la mayoría de los pacientes? Comente.

Una solución

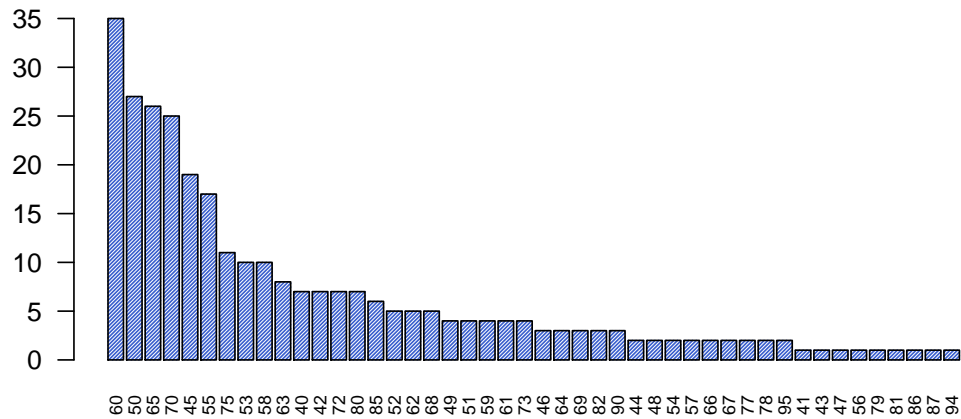
crear un gráfico de barras simple (no ordenado)(debe incluir título)

crear un gráfico de barras ordenado

```
df<-data.frame(table(heart$age))  
df<-df[order(df$Freq, decreasing=TRUE),]
```

```
barplot(df$Freq,  
        names.arg = df$Var1,  
        main="Conteo de pacientes por edad",  
        las=2,  
        col="royalblue3",  
        density=60,  
        cex.names = 0.7)
```

Conteo de pacientes por edad



realizar comentario general del gráfico

Se observa mayor cantidad de pacientes en particular en una edad, que sería la edad de 60 años. Se poseen pacientes de diversas edades en los datos, esto se puede notar gráficamente viendo la cantidad de barras.

realizar alguna inferencia adicional: relaciona el contexto del ejercicio o genera conjeturas adicionales o complementa con estadísticas adicionales y comenta, etcétera

El rango de edad de los pacientes en los datos es de 40 años a 95 años. Las edades con mayor número de pacientes son 60 años con 35 pacientes, 50 años con 27 pacientes y 65 años con 26 pacientes. Si se comparan las estadísticas de la edad en los grupos de muerte y no muerte por falla cardíaca, se puede ver que por ejemplo la mediana para el grupo no muerte por falla cardíaca es de 60 mientras que para el grupo de muerte por falla cardíaca es de 65, podemos decir que la edad del 50% de personas del grupo de no muerte por falla cardíaca se encuentra debajo de la edad del 50% del grupo muerte por falla cardíaca.

```
head(df)
```

```
##      Var1 Freq
## 21     60   35
## 11     50   27
## 26     65   26
## 31     70   25
## 6      45   19
## 16     55   17
```

```
range(heart$age)
```

```
## [1] 40 95
```

```
range(df$Freq)
```

```
## [1] 1 35

summary(heart$age)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  40.00   51.00   60.00   60.83   70.00   95.00

summary(heart$age[which(heart$death==0)])

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  40.00   50.00   60.00   58.76   65.00   90.00

summary(heart$age[which(heart$death==1)])

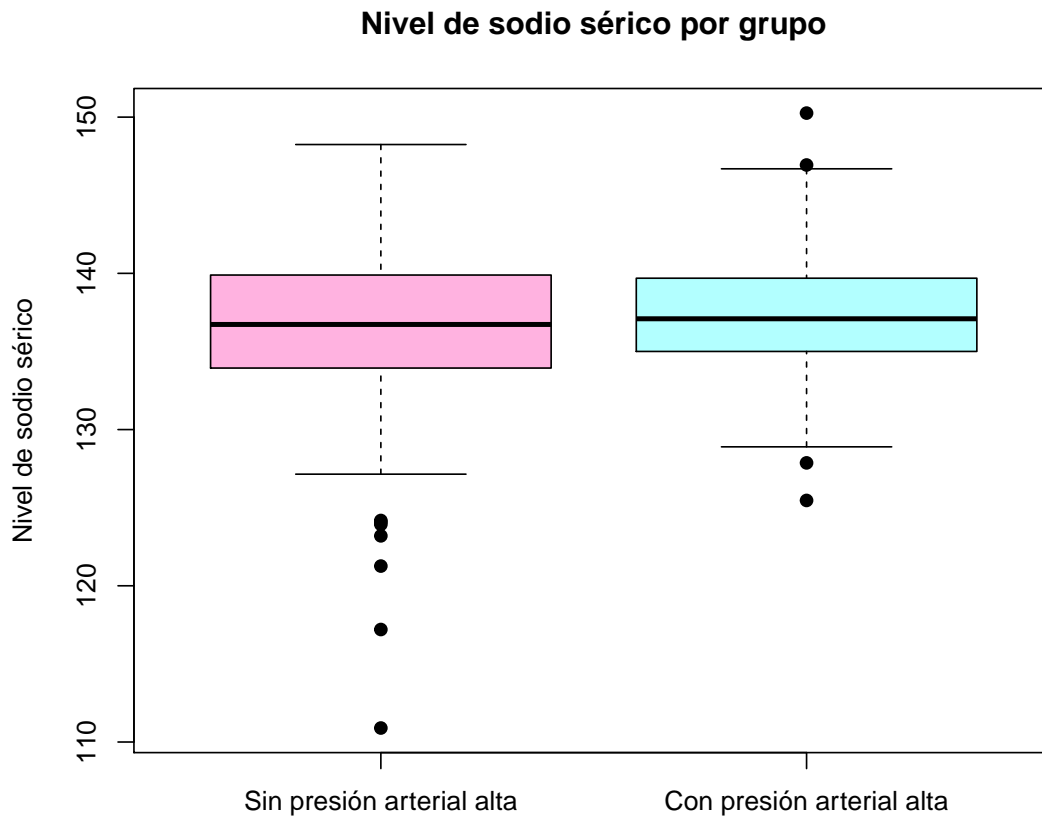
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  42.00   55.00   65.00   65.21   75.00   95.00
```

- c) Interesa visualizar los niveles de sodio sérico en la sangre para el grupo de pacientes con presión arterial alta y pacientes sin presión arterial alta, de modo de observar y determinar si existen diferencias. Realice un gráfico en el que se pueda diferenciar los niveles de sodio sérico para ambos grupos (presión arterial alta y no presión arterial alta), en este gráfico deben observarse los cuartiles de los niveles de sodio sérico para cada grupo, de modo que puedan compararse. ¿Observa diferencias? Comente.

Una solución:

realizar ambos gráficos, idealmente juntos para permitir comparaciones

```
boxplot(heart$serum_sodium~heart$high_pressure,
        xlab="",
        ylab="Nivel de sodio sérico",
        main="Nivel de sodio sérico por grupo",
        names=c("Sin presión arterial alta", "Con presión arterial alta"),
        col=rainbow(2, alpha=0.3, start=0.9),
        pch=19)
```



comentar sobre el gráfico

En el gráfico se puede ver que ambos boxplots se encuentran bastante alineados, lo que quizás podría decir que no hay muchas diferencias en el nivel de sodio sérico para el grupo de presión arterial alta y no presión arterial alta. Al menos en el centro de sus cajas (mediana) se ven muy similares. También es posible ver que el rango que recorre el sodio sérico es de 110 a 150 aproximadamente, tomando para el grupo de presión arterial alta valores de aproximadamente 125 a 150 (rango está más acotado a valores altos).

```
median(heart$serum_sodium[which(heart$death==0)])

## [1] 137.2605

median(heart$serum_sodium[which(heart$death==1)])

## [1] 135.6346
```

comentar sobre elementos del gráfico, interpretar a qué corresponden, incorporar estadísticas, generar conjeturas, explicar

Pueden comentar sobre los outliers, incluso determinar cuáles son. Notar que para el

grupo de sin presión arterial alta hay varios outliers inferiores, es decir, valores que se alejan considerablemente del centro de los datos. Para el caso de presión arterial alta se observan 4 outliers, pero estos son inferiores y superiores.

- d) Al grupo de estudio le interesa conocer más detalles de la distribución del sodio sérico. Realice un gráfico en el que se pueda visualizar la distribución de esta variable por intervalos, en particular utilice 30 intervalos para construir el gráfico. ¿Cómo es la forma de la distribución de los niveles de sodio sérico en la sangre? ¿Observa simetría? Comente. Sobreponga una curva de una distribución normal. Recuerde incluir la leyenda en su gráfico de modo de ser más informativo. Comente.

Una solución:

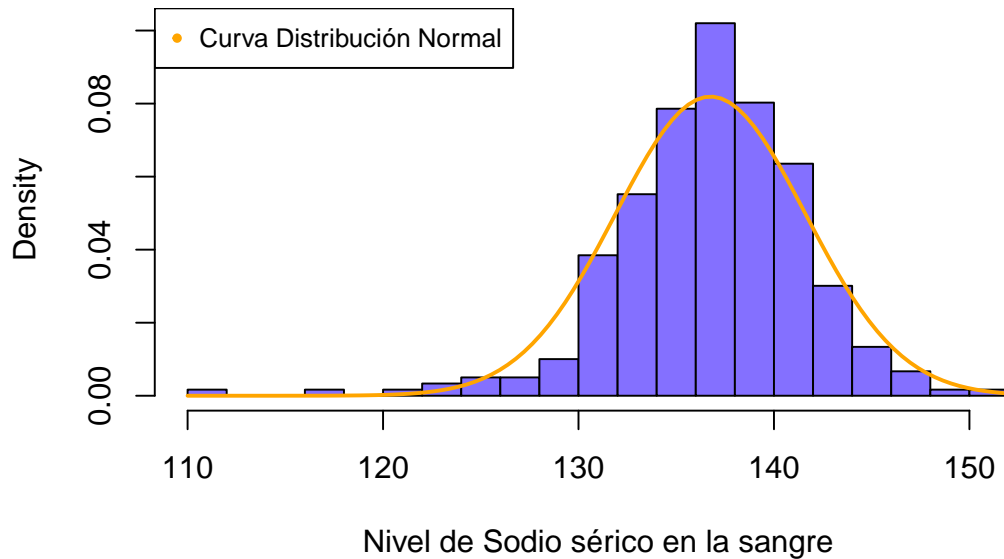
hacer el gráfico con todos sus elementos correspondientes

```
hist(heart$serum_sodium,
     breaks=20,
     freq=FALSE,
     col="lightslateblue",
     main="Histograma de nivel de sodio sérico en la sangre",
     xlab="Nivel de Sodio sérico en la sangre")

curve(dnorm(x,
            mean=mean(heart$serum_sodium),
            sd=sd(heart$serum_sodium)),
      col="orange",
      lwd=2, add=TRUE)

legend("topleft",
      legend="Curva Distribución Normal",
      col="orange",
      cex=0.8,
      pch=20)
```


Histograma de nivel de sodio sérico en la sangre



comentar lo observado, pueden comentar sobre la forma de la distribución, qué tan simil se observa la curva de la distribución normal, si está sesgada o inclinada, etcétera

BONUS) Muchas veces más que estudiar la edad del paciente en sí, interesa saber si pertenece a un rango etario u otro, pues se ha visto que por rango etario se esperan comportamientos similares, permitiendo caracterizar a los pacientes mediante categorías basadas en la edad. Agrupe a los pacientes por edad en los siguientes rangos etarios:

- Si el paciente tiene entre 40 y 50 años (incluyendo dichos valores) pertenezca al rango etario *Adulto I*
- Si el paciente tiene entre 51 y 65 años (incluyendo dichos valores) pertenezca al rango etario *Adulto II*
- Si el paciente tiene entre 66 y 70 años (incluyendo dichos valores) pertenezca al rango etario *Vejez I*
- Si el paciente tiene más de 70 años pertenezca al rango etario *Vejez II*
- Si no cumple ninguna de las anteriores que pertenezca al rango etario *No especifica*

Muestre en una tabla de doble entrada la cantidad de pacientes que murieron por falla cardiaca por rango etario. Comente.

Una solución:

```

etario<-ifelse(heart$age>=40&heart$age<=50, "Adulto I",
              ifelse(heart$age>=51&heart$age<=65, "Adulto II",
                    ifelse(heart$age>65&heart$age<=70, "Vejez I",
                          ifelse(heart$age>70, "Vejez II", "No especifica"))))

table(etario, heart$death)

##
## etario      0      1
## Adulto I    55    19
## Adulto II  101    35
## Vejez I     26    11
## Vejez II    21    31

prop.table(table(etario, heart$death)) #Muestra distribucion entre el total

##
## etario      0      1
## Adulto I    0.18394649 0.06354515
## Adulto II   0.33779264 0.11705686
## Vejez I     0.08695652 0.03678930
## Vejez II    0.07023411 0.10367893

prop.table(table(etario, heart$death), margin=1) #Muestra distribucion por fila

##
## etario      0      1
## Adulto I    0.7432432 0.2567568
## Adulto II   0.7426471 0.2573529
## Vejez I     0.7027027 0.2972973
## Vejez II    0.4038462 0.5961538

```

hacer la variable rango etario, se pudo haber usado for, if, else, u otras funciones

comentar sobre la tabla de doble entrada, se puede ver que la mayor cantidad de personas que falleció por fallo cardiaco pertenecía al grupo Adulto II, pero si se ven en porcentajes por ejemplo, en Vejez II es donde murió mayor proporción de personas por fallo cardiaco. También se puede conjeturar, la idea es comentar.