

Ejercicios tema 2

Juan Andrés Peraira Pérez

13 de abril de 2018

Ejercicio 1: Utiliza los datos “iris” que corresponden a mediciones (en centímetros) de 4 variables: largo y ancho de los pétalos y sépalos; para 50 flores de 3 especies distintas de plantas Iris setosa, versicolor, y virginica.

Queremos responder a las siguientes preguntas:

¿Cuántos datos (o casos) tenemos para cada especie? y ¿qué porcentaje representan del total de casos? Realice los gráficos pertinentes para cada tipo de variable (cualitativa vs. cuantitativa).

```
datos<-iris #-- cargamos los datos  
head(datos)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1           5.1         3.5         1.4         0.2   setosa  
## 2           4.9         3.0         1.4         0.2   setosa  
## 3           4.7         3.2         1.3         0.2   setosa  
## 4           4.6         3.1         1.5         0.2   setosa  
## 5           5.0         3.6         1.4         0.2   setosa  
## 6           5.4         3.9         1.7         0.4   setosa
```

```
attach(datos) #-- activamos las variables
```

¿Cuál es la media clásica y robusta del ancho del sépalo para cada especie?. Realice diagrama de cajas.

```
#-- Media clásica del ancho del sépalo para cada especie.
```

```
media_anchos_sepalo<-aggregate(Sepal.Width~Species, datos, mean)  
media_anchos_sepalo
```

```
##      Species Sepal.Width  
## 1     setosa      3.428  
## 2 versicolor      2.770  
## 3  virginica      2.974
```

```
#-- Media robusta del ancho del sépalo para cada especie.
```

```
library(WRS2)
```

```
## Warning: package 'WRS2' was built under R version 3.4.4
```

```
media_robusta_anchos_sepalo<-aggregate(Sepal.Width~Species, datos, mest)  
media_robusta_anchos_sepalo
```

```
##      Species Sepal.Width  
## 1     setosa      3.418994  
## 2 versicolor      2.782828  
## 3  virginica      2.962500
```

¿Qué correlaciones existen entre las distintas medidas tomadas a cada planta? Realice un gráfico bidimensional para observarlo.

Ejercicio 2: Utiliza los datos “Davis” (paquete “car”) para calcular el IMC como se indicó en el tema 2 ($IMC = \text{Peso} / \text{Estatura}^2$). Realia:

Gráfico de barras y de sectores para las categorías del IMC por sexo.

```

#-- cargamos los datos
library(car)

## Warning: package 'car' was built under R version 3.4.4
## Loading required package: carData
## Warning: package 'carData' was built under R version 3.4.4

datos<-Davis
head(datos)

##   sex weight height repwt repht
## 1  M     77     182     77    180
## 2  F     58     161     51    159
## 3  F     53     161     54    158
## 4  M     68     177     70    175
## 5  F     59     157     59    155
## 6  M     76     170     76    165

attach(datos) #-- Activamos las variables

#-- Creamos la función
imc=function(w,h){w/(h/100)^2}
#-- Calculamos el imc para los datos
datos_imc<-imc(datos$weight,datos$height)
# creamos las categorías de IMC
imcc_datos=cut(datos_imc, breaks=c(0, 15, 18.5, 25, 30))
#-- Creamos la tabla
imcfrec=table(imcc_datos)
cbind(imcfrec)

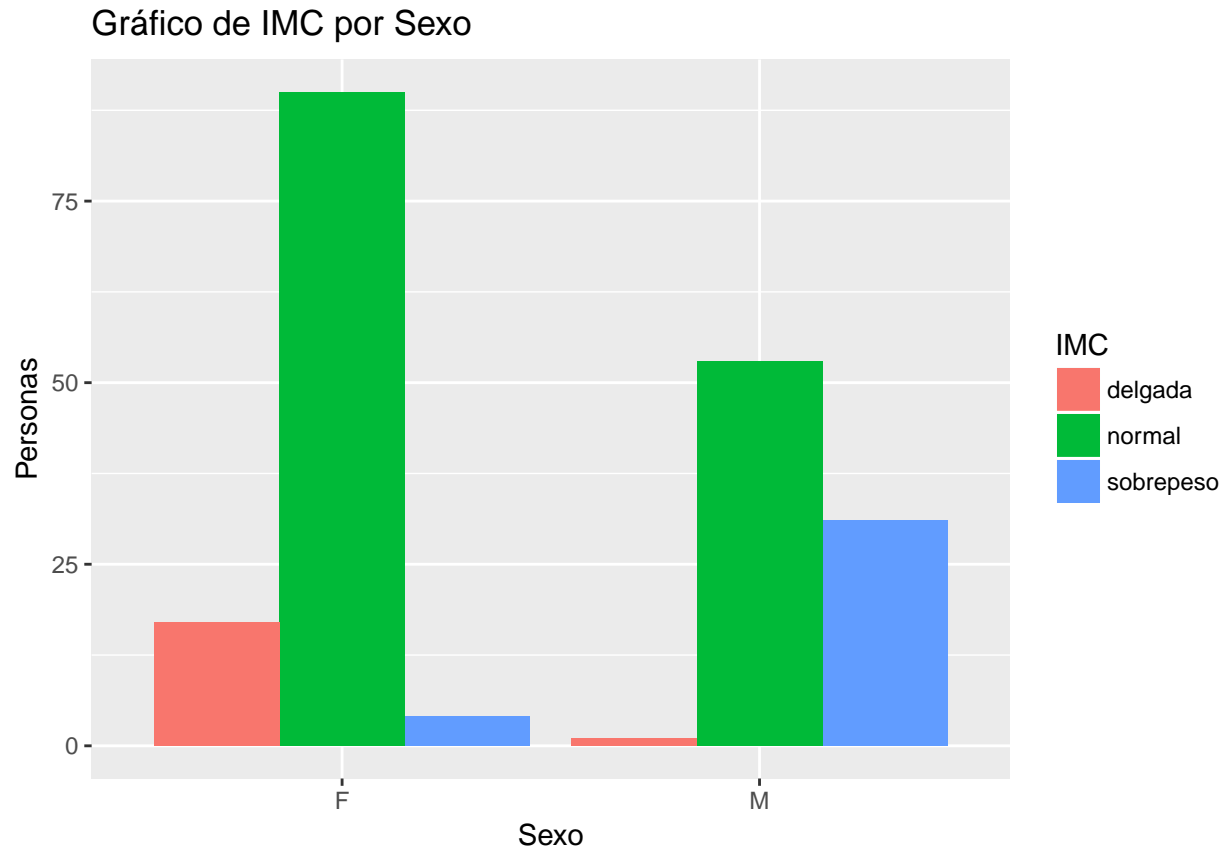
##           imcfrec
## (0,15]           0
## (15,18.5]        18
## (18.5,25]       143
## (25,30]         35

#-- agregamos las etiquetas
levels(imcc_datos)=c("infrapeso","delgada","normal","sobrepeso","obesidad")
#-- Creamos el data frame
datos_davis<-data.frame(datos, IMC=datos_imc, IMCc=imcc_datos)
# eliminamos los errores
datos_davis<-datos_davis[!is.na(datos_davis$IMCc),]
head(datos_davis)

##   sex weight height repwt repht      IMC      IMCc
## 1  M     77     182     77    180 23.24598   normal
## 2  F     58     161     51    159 22.37568   normal
## 3  F     53     161     54    158 20.44674   normal
## 4  M     68     177     70    175 21.70513   normal
## 5  F     59     157     59    155 23.93606   normal
## 6  M     76     170     76    165 26.29758 sobrepeso

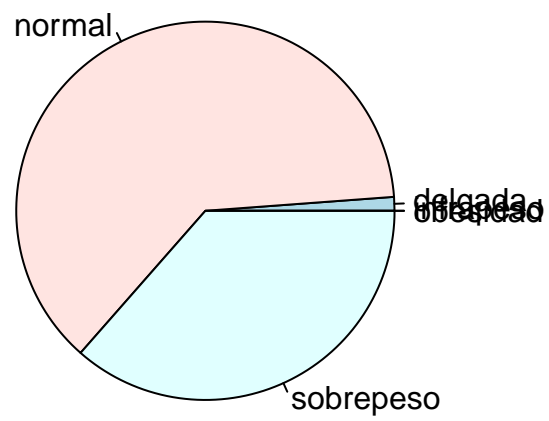
```

```
library(ggplot2)
library(gridExtra)
ggplot(datos_davis,aes(x=factor(sex),fill=factor(IMCc))) +
  geom_bar(stat = "count", position="dodge")+
  labs(title = "Gráfico de IMC por Sexo") +
  labs(fill = "IMC") +
  labs(aes(x="Sexo",y="Personas"))
```



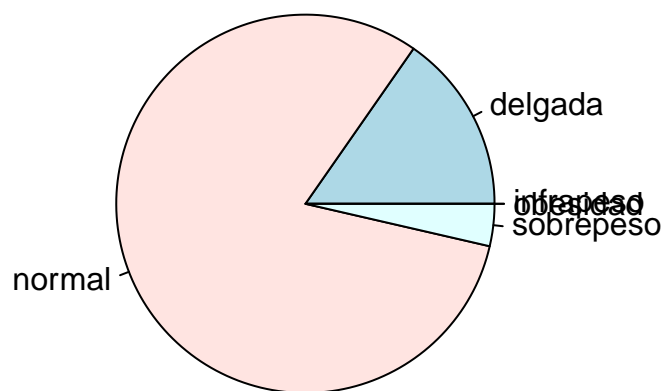
```
-- Gráfico de sectores
-- Realizaremos dos gráficos separando los hombres y las mujeres

-- Gráfico para los hombres
datos_hombres<-subset(datos_davis,datos_davis$sex=="M")
grafico_tarta<-pie(table(datos_hombres$IMCc))
```



-- Gráfico para las mujeres

```
datos_mujeres<-subset(datos_davis,datos_davis$sex=="F")  
grafico_tarta<-pie(table(datos_mujeres$IMCc))
```



Gráficos de cajas e histogramas para la variable IMC numérica, también por sexo. ¿Existe algún outlier?, ¿cuáles?.

Interpreta los resultados.