Examen\_01\_hvg

Heidy Valdelamar Gonzalez

2022-09-12

## PRIMER EJERCICIO

## Cargo librerías ##  
library("tidyverse")

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.8 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library("fdth")

##   
## Attaching package: 'fdth'  
##   
## The following objects are masked from 'package:stats':  
##   
## sd, var

library("UsingR")

## Loading required package: MASS  
##   
## Attaching package: 'MASS'  
##   
## The following object is masked from 'package:dplyr':  
##   
## select  
##   
## Loading required package: HistData  
## Loading required package: Hmisc  
## Loading required package: lattice  
## Loading required package: survival  
## Loading required package: Formula  
##   
## Attaching package: 'Hmisc'  
##   
## The following objects are masked from 'package:dplyr':  
##   
## src, summarize  
##   
## The following objects are masked from 'package:base':  
##   
## format.pval, units  
##   
##   
## Attaching package: 'UsingR'  
##   
## The following object is masked from 'package:survival':  
##   
## cancer

library("formatR")  
  
## leer base de datos ##  
bd1 <- read\_csv("LDS\_C02\_NCBIRTH800.csv")

## Rows: 800 Columns: 14  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): hispmom  
## dbl (13): plural, sex, mage, weeks, marital, racemom, gained, smoke, drink, ...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## MAGE

#Creamos una tabla de datos#  
 tablamage <- table(bd1$mage)  
view(tablamage)  
  
# 1. Calculate the mean, median, standard deviation, IQR, and range. #  
mean(tablamage, trim = 0, na.rm = 0)

## [1] 28.57143

median(tablamage, trim = 0, na.rm = 0)

## [1] 32.5

sd(tablamage)

## [1] 15.60016

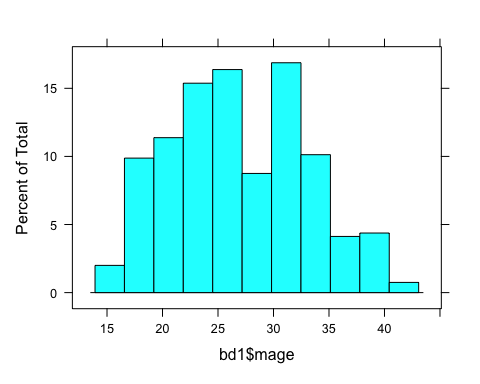
IQR(tablamage)

## [1] 24.25

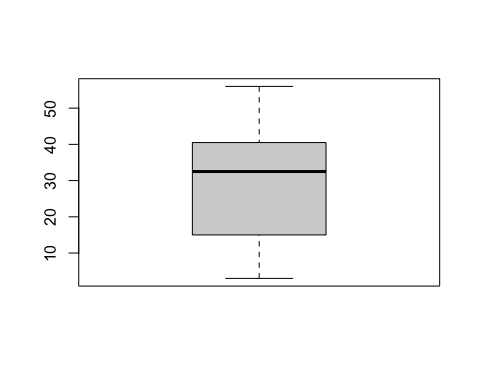
range(tablamage, trim = 0, na.rm = 0)

## [1] 0 56

# 2. For each, construct a histogram and comment on the shape of the distribution. #  
histogram(bd1$mage, prob=TRUE)



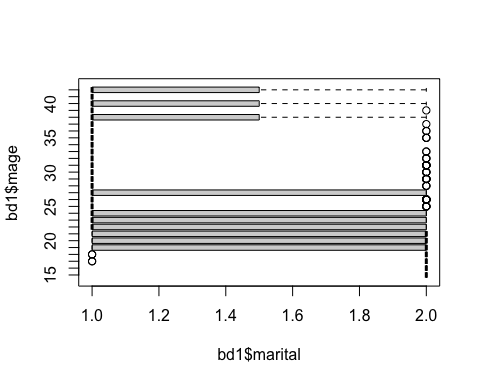
# 4. Construct box-plots for all four variables. #  
boxplot(tablamage)



# 6. Construct side-by-side box-plots of the variable of MAGE for women for women who are and are not married. Do you see differences in ages in the two groups? Which group has more variability?#  
dftmage <- data\_frame(bd1$marital, bd1$mage)

## Warning: `data\_frame()` was deprecated in tibble 1.1.0.  
## Please use `tibble()` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was generated.

view(dftmage)  
boxplot(bd1$marital ~ bd1$mage, horizontal=TRUE)



#Más variedad que está del grupo que está casado#

## WEEKS

#Creamos una tabla de datos#  
 tablaweeks <- table(bd1$weeks)  
view(tablaweeks)  
  
# 1. Calculate the mean, median, standard deviation, IQR, and range. #  
mean(tablaweeks, trim = 0, na.rm = 0)

## [1] 34.73913

median(tablaweeks, trim = 0, na.rm = 0)

## [1] 5

sd(tablaweeks)

## [1] 54.63367

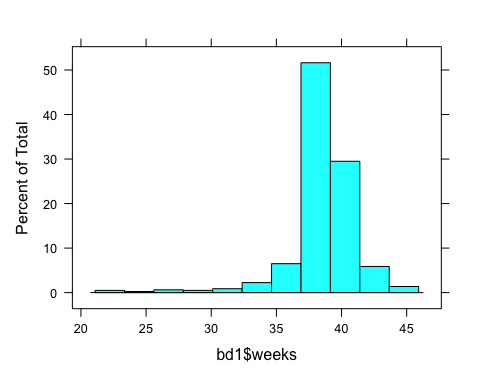
IQR(tablaweeks)

## [1] 27.5

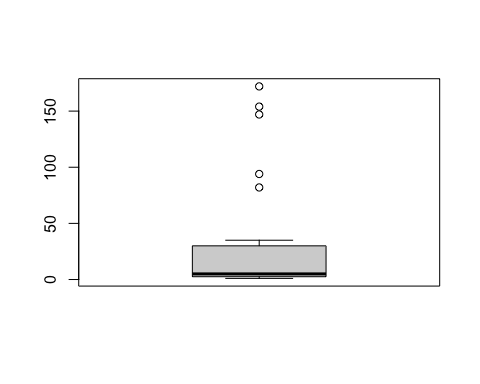
range(tablaweeks, trim = 0, na.rm = 0)

## [1] 0 172

# 2. For each, construct a histogram and comment on the shape of the distribution. #  
histogram(bd1$weeks, prob=TRUE)



# 4. Construct box-plots for all four variables. #  
boxplot(tablaweeks)



## GAINED

#Creamos una tabla de datos#  
 tablagained<- table(bd1$gained)  
view(tablagained)  
  
# 1. Calculate the mean, median, standard deviation, IQR, and range. #  
mean(tablagained, trim = 0, na.rm = 0)

## [1] 11.42647

median(tablagained, trim = 0, na.rm = 0)

## [1] 7.5

sd(tablagained)

## [1] 12.70834

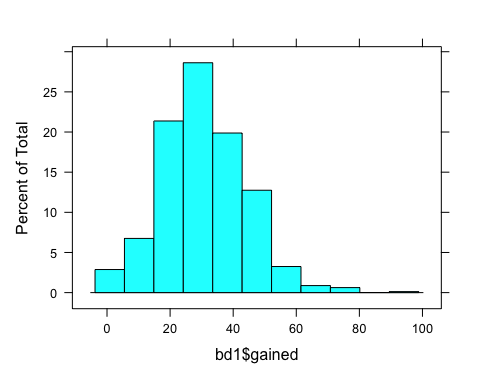
IQR(tablagained)

## [1] 13.25

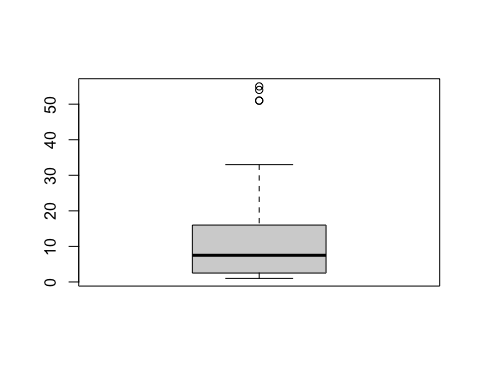
range(tablagained, trim = 0, na.rm = 0)

## [1] 0 55

# 2. For each, construct a histogram and comment on the shape of the distribution. #  
histogram(bd1$gained, prob=TRUE)



# 4. Construct box-plots for all four variables. #  
boxplot(tablagained)



## TOUNCES

#Creamos una tabla de datos#  
 tablatounces<- table(bd1$tounces)  
view(tablatounces)  
  
# 1. Calculate the mean, median, standard deviation, IQR, and range. #  
mean(tablatounces, trim = 0, na.rm = 0)

## [1] 6.722689

median(tablatounces, trim = 0, na.rm = 0)

## [1] 4

sd(tablatounces)

## [1] 6.439976

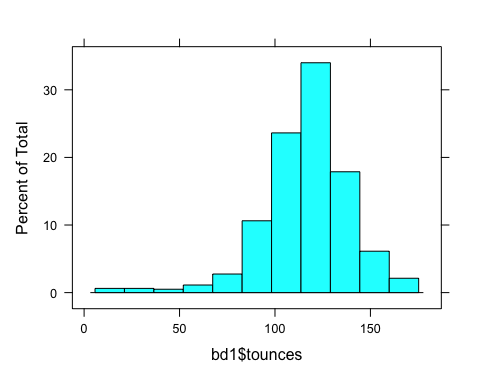
IQR(tablatounces)

## [1] 9

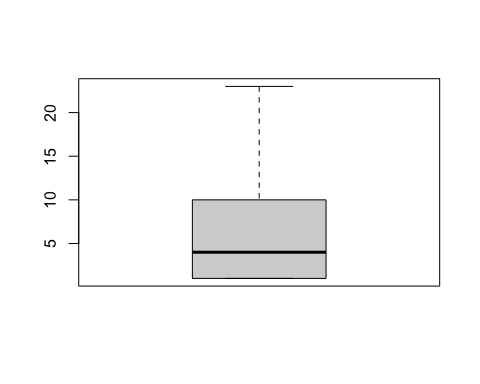
range(tablatounces, trim = 0, na.rm = 0)

## [1] 0 23

# 2. For each, construct a histogram and comment on the shape of the distribution. #  
histogram(bd1$tounces, prob=TRUE)



# 4. Construct box-plots for all four variables. #  
boxplot(tablatounces)



## TGRAMS

#Creamos una tabla de datos#  
 tablatgrams<- table(bd1$tgrams)  
view(tablatgrams)  
  
# 1. Calculate the mean, median, standard deviation, IQR, and range. #  
mean(tablatgrams, trim = 0, na.rm = 0)

## [1] 6.722689

median(tablatgrams, trim = 0, na.rm = 0)

## [1] 4

sd(tablatgrams)

## [1] 6.439976

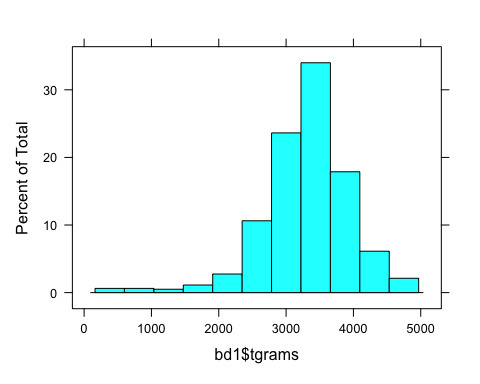
IQR(tablatgrams)

## [1] 9

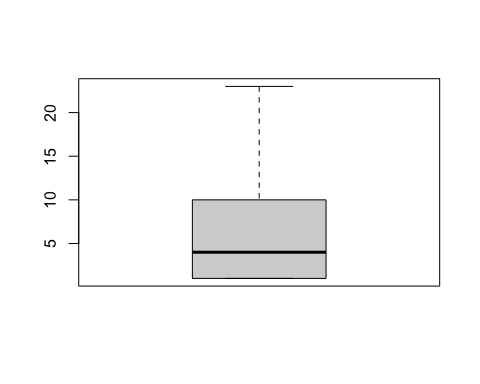
range(tablatgrams, trim = 0, na.rm = 0)

## [1] 0 23

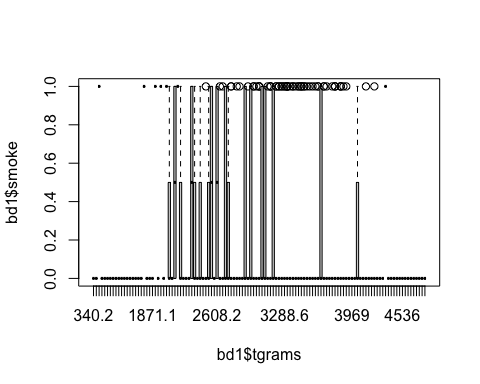
# 2. For each, construct a histogram and comment on the shape of the distribution. #  
histogram(bd1$tgrams, prob=TRUE)



# 4. Construct box-plots for all four variables. #  
boxplot(tablatgrams)



# 5. Construct side-by-side box-plots for the variable of TGRAMS for women who admitted to smoking and women who did not admit to smoking, Which group has more variability? #  
dftgrams <- data\_frame(bd1$smoke, bd1$tgrams)  
view(dftgrams)  
boxplot(bd1$smoke ~ bd1$tgrams)



#más variedad grupo que no fuma#

## SEGUNDO EJERCICIO

# a) Calcule la probabilidad de que un postulante tomado de forma aleatoria de este grupo:

# 1. Tuviera un examen de calificación baja.

totalbaja = 200  
total = 1000  
  
P\_Baja = (totalbaja/total)  
 P\_Baja

## [1] 0.2

# 2. Sea graduado de una preparatoria de calidad superior #

totalsuperior = 500  
total = 1000  
  
P\_Superior = (totalsuperior/total)  
 P\_Superior

## [1] 0.5

# 3. Tuviera una calificación baja en su examen y sea graduado de una preparatoria de calidad superior.#

totalbajasuperior = 55  
total = 1000  
P\_Bajasup = (totalbajasuperior/total)  
 P\_Bajasup

## [1] 0.055

# 4. Tuviera una calificación alta en su examen o sea graduado de una preparatoria de calidad superior. #

# 4. Tuviera una calificación alta en su examen o sea graduado de una preparatoria de calidad superior. #  
 totalaltasuperior = 300  
total = 1000  
  
P\_Altasup = (totalaltasuperior/total)  
 P\_Altasup

## [1] 0.3

#b) Calcule las siguientes probabilidades:

#1. P(A) 2. P(H) 3. P(M)

#P(A)  
 totalA = 300  
total = 1000  
  
P\_A = (totalA/total)  
 P\_A

## [1] 0.3

#P(H)  
 totalH = 3901  
total = 1000  
  
P\_H = (totalH/total)  
 P\_H

## [1] 3.901

#P(M)  
 totalM = 3901  
total = 1000  
  
P\_M = (totalM/total)  
 P\_M

## [1] 3.901

#4. P(A|H) 5. P(MyP) 6. P(H|S)

#P(A|H)  
totalA = 300  
totalH = 3901  
P\_AH = (totalA/totalH)  
P\_AH

## [1] 0.07690336

#P(MyP)  
totalM = 70  
totalP = 200  
P\_MyP = (totalM/totalP)  
P\_MyP

## [1] 0.35

#P(H|S)  
totalH = 3901  
totalS = 220  
P\_HS = (totalH/totalS)  
P\_HS

## [1] 17.73182