Yahriel\_Salinas-Reyes\_7340\_HW1\_FA24.rmd

Yahriel Salinas-Reyes

2024-09-02

## clean R environment

rm(list = ls(all=TRUE))  
graphics.off()  
shell("cls")

## Include Packages

library(datasets)   
library(ggplot2)  
library(MASS)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':  
##   
## select

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

# Question 1

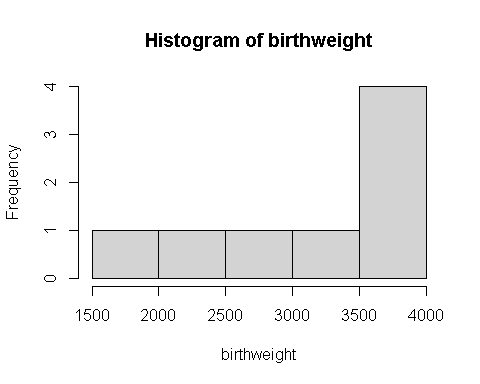
# Question 1  
birthweight <- c(3600,1700,4000,3900,3100,3800,2200,3000)  
summary(birthweight)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1700 2800 3350 3162 3825 4000

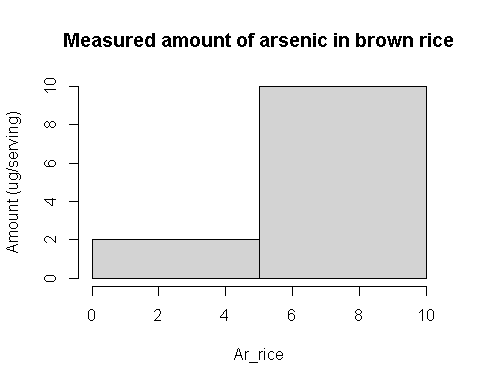
sd(birthweight)

## [1] 839.9617

hist(birthweight)



# Question 2  
Ar\_rice <- c(6.1, 5.4, 6.9, 4.9, 6.6, 6.3, 6.7, 8.2, 7.8, 1.5, 5.4, 7.3)  
hist(Ar\_rice,   
 main = "Measured amount of arsenic in brown rice",   
 # xlab = ""  
 ylab = "Amount (ug/serving)",  
 breaks = 2)



# Question 3  
father <- c(73.0, 75.5, 75.0, 75.0, 75.0, 74.0, 74.0, 73.0, 73.0, 78.5)  
son <- c(74.0, 73.5, 71.0, 70.5, 72.0, 76.5, 74.0, 71.0, 72.0, 73.2)  
summary(father)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 73.00 73.25 74.50 74.60 75.00 78.50

sd(father)

## [1] 1.66333

var(father)

## [1] 2.766667

str(father)

## num [1:10] 73 75.5 75 75 75 74 74 73 73 78.5

summary(son)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 70.50 71.25 72.60 72.77 73.88 76.50

sd(son)

## [1] 1.833667

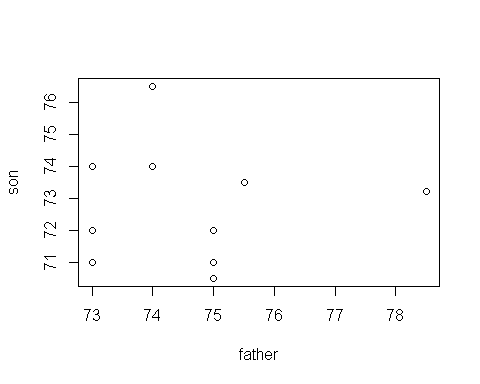
var(son)

## [1] 3.362333

str(son)

## num [1:10] 74 73.5 71 70.5 72 76.5 74 71 72 73.2

plot(father,son)



cor(father,son)

## [1] -0.01712208

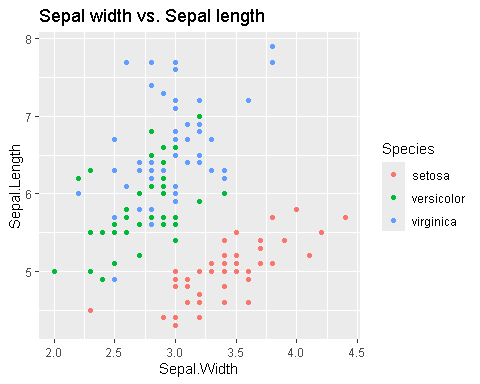
# Question 4  
mydata <- iris  
head(mydata)

## 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

str(mydata)

## 'data.frame': 150 obs. of 5 variables:  
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

ggplot(mydata, aes(x=Sepal.Width, y=Sepal.Length, col=Species))+  
 geom\_point()+  
 ggtitle("Sepal width vs. Sepal length")



# Question 5  
# 5.1  
## Given  
mean <- 0  
sd <- 1  
## Z0.01 is the z-score separating the bottom 99% from the top 1%  
Z0.01 <- qnorm(p=.01, mean, sd, lower.tail=F)  
Z0.01

## [1] 2.326348

# 5.2  
## Find the Z score for the bottom 2%  
cutoff\_low <- qnorm(p=0.02, mean, sd, lower.tail=T)  
## Find the Z score for the top 2%  
cutoff\_high <- qnorm(p=0.98, mean, sd, lower.tail=T)  
## Output the results  
cutoff\_low

## [1] -2.053749

cutoff\_high

## [1] 2.053749

# Question 6  
## 6.1 Probability of the bone density test score between -2.75 and 2.75  
prob\_1 <- pnorm(2.75, mean, sd, lower.tail=T) - pnorm(-2.75, mean, sd, lower.tail=T)  
prob\_1

## [1] 0.9940405

##6.2Probability of the bone density test score between -2.0 and 2.0  
prob\_2 <- pnorm(2.0, mean, sd, lower.tail=T) - pnorm(-2.0, mean, sd, lower.tail=T)  
prob\_2

## [1] 0.9544997

# Example sub-setting with “[]” and “==”

# sub-setting with "[]" and "=="  
#attach(mydata)  
  
#hist(Sepal.Length [Species == "setosa"])  
#detach(mydata)

## Don’t forget to click “Knit”

# END