Week 2 tasks

teodor dyakov

10/18/2020

library(MASS)  
  
#намираме първо всички стойности в колоната Smoke  
unique(survey$Smoke)

## [1] Never Regul Occas Heavy <NA>   
## Levels: Heavy Never Occas Regul

#Задача 1  
  
#а) - вероятноста случайно избран човек да е редовен пушач:  
  
#броя на хората за които Smoke != N/A  
n\_people = nrow(survey[!is.na(survey$Smoke),])  
  
sum(survey$Smoke == 'Regul', na.rm = TRUE)/n\_people

## [1] 0.0720339

#втори начин:  
table = prop.table(table(survey$Smoke),)  
table['Regul']

## Regul   
## 0.0720339

#б) - вероятността случайно избран човек да е мъж и редовен пушач  
  
maleSmokers = sum(survey$Smoke == 'Regul' & survey$Sex == 'Male', na.rm = TRUE)  
  
maleSmokers/(sum(!is.na(survey$Sex) & !is.na(survey$Smoke)))

## [1] 0.05106383

#втори начин  
table = prop.table(table(survey$Sex,survey$Smoke),)  
table['Male', 'Regul']

## [1] 0.05106383

#в) - вероятността случайно избран мъж да е редовен пушач  
  
#броят на всички мъже   
maleNumber = sum(survey$Sex == 'Male' & !is.na(survey$Smoke), na.rm = TRUE)  
maleSmokers/maleNumber

## [1] 0.1025641

#втори начин  
table = prop.table(table(survey$Sex,survey$Smoke), 1)  
table['Male', 'Regul']

## [1] 0.1025641

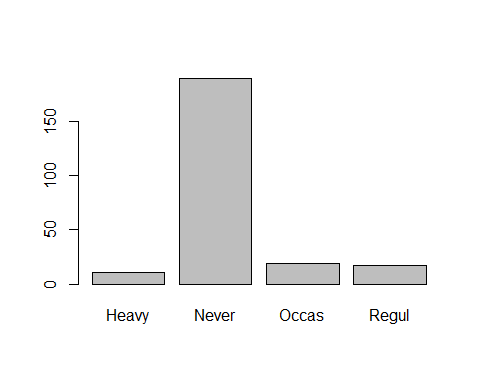
#г) - вероятноста случайно избран редовен пушач да е мъж  
  
#броят на всички редовни пушачи  
regulCount = sum(survey$Smoke == 'Regul', na.rm = TRUE)  
maleSmokers/regulCount

## [1] 0.7058824

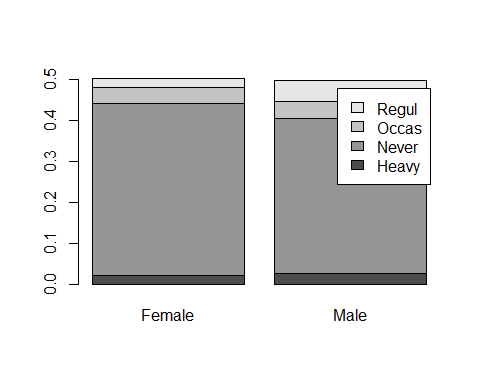
#втори начин  
table = prop.table(table(survey$Sex,survey$Smoke), 2)  
table['Male','Regul']

## [1] 0.7058824

#Зад 2  
#барплот на броя на типовете пушачи  
barplot(table(survey$Smoke))



#пушене в зависимост от пола  
table = prop.table(table(survey$Smoke,survey$Sex),)  
barplot(table, legend.text = TRUE)



#Зад 3  
#некви статистики  
summary(survey$Height)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 150.0 165.0 171.0 172.4 180.0 200.0 28

heightMean = mean(survey$Height, na.rm = TRUE)  
heightSd = sd(survey$Height, na.rm = TRUE)  
  
  
maleHeights = survey[survey$Sex == 'Male', 'Height']  
summary(maleHeights)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 154.9 172.8 180.0 178.8 185.0 200.0 13

sd(maleHeights, na.rm = TRUE)

## [1] 8.380252

femaleHeights = survey[survey$Sex == 'Female', 'Height']  
summary(femaleHeights)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 150.0 162.6 166.8 165.7 170.0 180.3 17

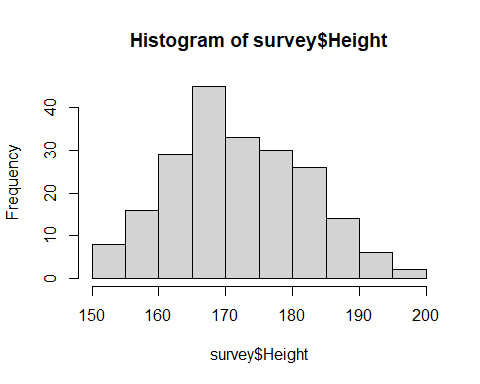
sd(femaleHeights, na.rm = TRUE)

## [1] 6.151777

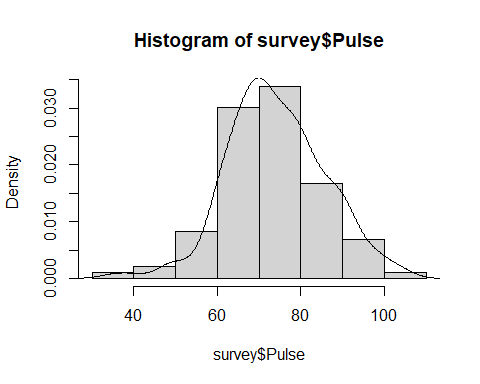
#каква част от студентите се различават от ср. височина с <= 1 стандартно отклонение?  
sum(abs(survey$Height - heightMean) <= heightSd, na.rm = TRUE)/sum(!is.na(survey$Height))

## [1] 0.6842105

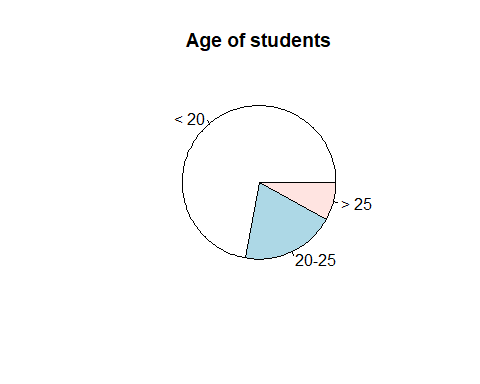
#Зад 4  
hist(survey$Height)



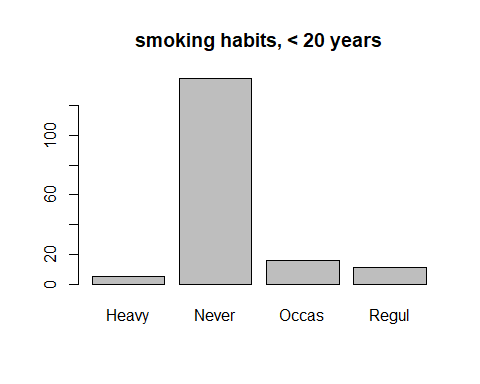
#Зад 5  
#хистограма на пулса на студентите  
hist(survey$Pulse, prob = TRUE)  
  
#графика плътността на разпределението на пулса   
lines(density(survey$Pulse, na.rm = TRUE))



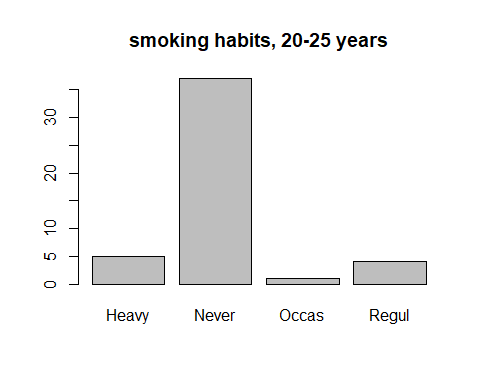
#Зад 6  
group1 = survey[survey$Age < 20,]  
group2 = survey[survey$Age >= 20 & survey$Age < 25,]  
group3 = survey[survey$Age >= 25,]  
  
pie(x = c(nrow(group1), nrow(group2), nrow(group3)), labels = c('< 20', '20-25', '> 25'), main = 'Age of students')



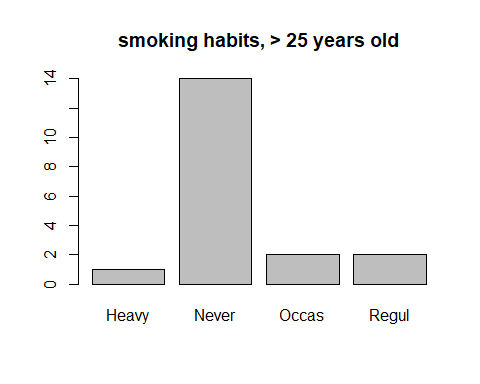
barplot(table(group1$Smoke), main = 'smoking habits, < 20 years')



barplot(table(group2$Smoke), main = 'smoking habits, 20-25 years')



barplot(table(group3$Smoke), main = 'smoking habits, > 25 years old')



#втори начин:  
survey['gr'] = findInterval(survey$Age, c(20,25))  
barplot(prop.table(table(survey$Smoke, survey$gr),2), legend = TRUE, names.arg = c('<20','20-25','>=25'),main = 'Smoking habits by age')

