Група 2: Иван Петров, 3MI0700035

Задача 1:

# Прочетете предоставените данни (file2.csv) с помощта на функцията "read.csv" в R

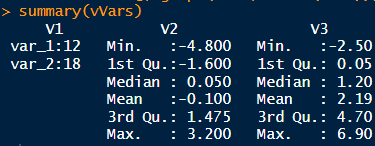
data = read.csv("file2.csv")

# Едномерен описателен анализ за всяка променлива, започваща с буквата "V"

data$V1 <- as.factor(data$V1)

vVars <- data[, grepl("^V", names(data))]

summary(vVars)

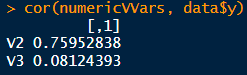


# Изследвате влиянието на всяка променлива, започваща с “V”, върху променливата “y”.

# Корелация между непрекъснати променливи и “y”

numericVVars <- vVars[, grepl("^V[^1]", names(vVars))]

cor(numericVVars, data$y)

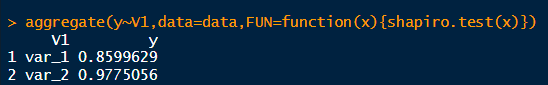


**Корелацията между V3 и Y е незначителна за статистиката**

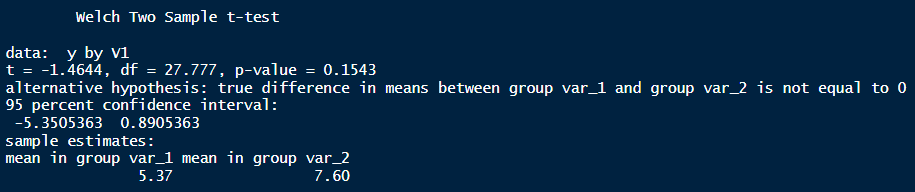
# t.test За категорийните променливи върху “y”

aggregate(y~V1,data=data,FUN=function(x){shapiro.test(x)})

t.test(y ~ V1, data=data)



**Имаме нормално разпределение на Y за кетегорийната променлива V1**



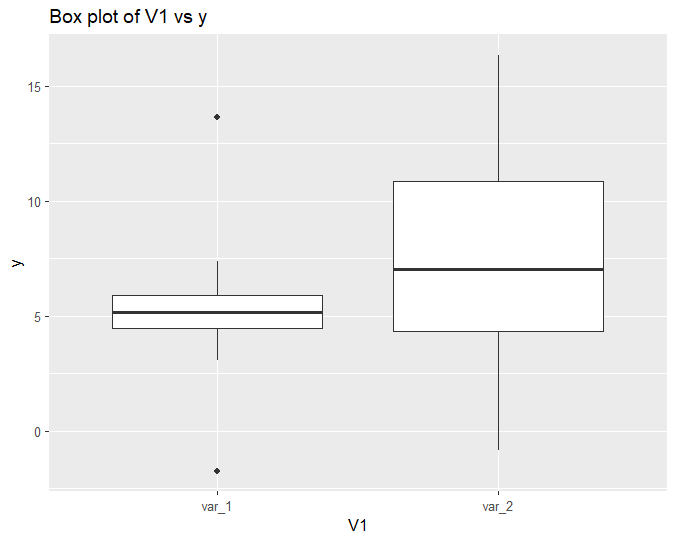
**Не можем да отхвърлим H0-хипотеза**

library(ggplot2)

ggplot(data, aes(x = V1, y = y)) +

geom\_boxplot() +

labs(title = "Box plot of V1 vs y")

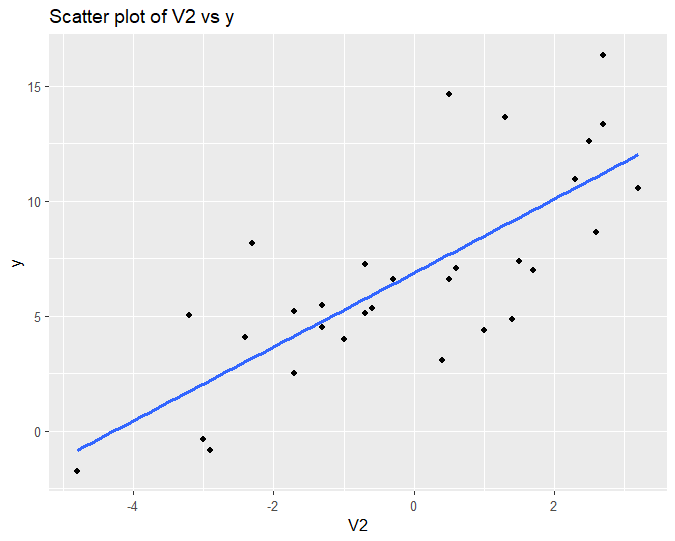


ggplot(data, aes(x = V2, y = y)) +

geom\_point() +

labs(title = "Scatter plot of V2 vs y") +

geom\_smooth(method = "lm", se = FALSE)



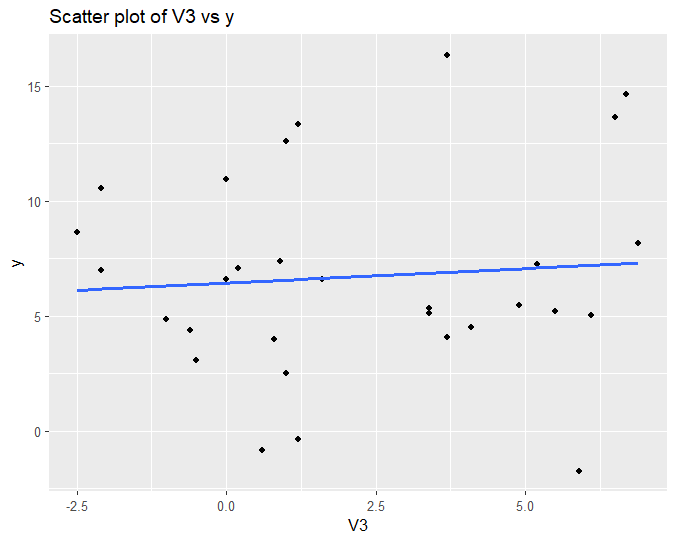
**V2 има много outlier-и спрямо y**

ggplot(data, aes(x = V3, y = y)) +

geom\_point() +

labs(title = "Scatter plot of V3 vs y") +

geom\_smooth(method = "lm", se = FALSE)



Задача 2 – Разпределения

1. Едър зърнопроизводител от Добрич обработва 700 000 декара. През последните 30 години той забелязал, че средният добив от един декар е 650 кг пшеница, със стандартно отклонение 40 килограма.

# Стандартно нормално разпределение

mu = 650

sigma = 40

x = 700

1. Определете вероятността средният добив от декар да надскочи 700 килограма?

pnorm(q = 700, mean = mu, sd = sigma)



1. Определете минималният добив от декар в 95% от случаите?

qnorm(p = 0.95, mean = mu, sd = sigma)



1. Агент на “LR” е установил, че вероятността едно лице да си купи тяхна козметика, ако бъде посетено у дома е 37%. Агентът си поставил за цел да посети 10 случайно избрани лица през днешния ден.

# Биномно разпределение

p=0.37

n=10

1. точно 4-ма да си купят козметика?

dbinom(x=4,size=10,prob=0.37)



1. Между 3 и 8 човека (включително) да си купят козметика?

pbinom(q=8,size=10,prob=0.37)-pbinom(q=3,size=10,prob=0.37)

