

Борислав Тодоров
Второ домашно
Статистика
Курс 3 Група 3
45421

Зад. 1

```
> x = morley[morley$Expt == "4",]
```

#взимаме само данните от експеримент 4

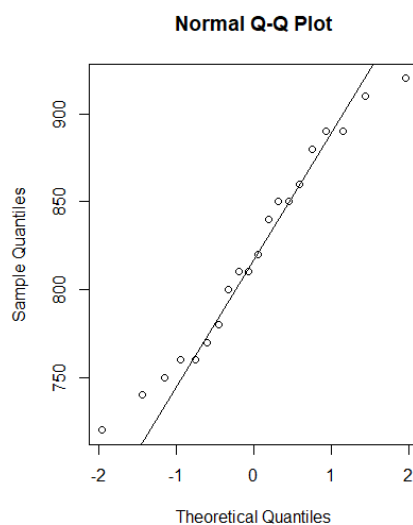
```
> shapiro.test(x$Speed)
```

#правим тест на скоростите

#p- value- то е 0.5667, т.е. данните са нормално разпределени ($p > 0.05$)

```
> qqnorm(x$Speed)
```

```
> qqline(x$Speed)
```



#точките са разположени около линията => изглежда окей

```
> t.test(x$Speed,coef.level=0.97)
```

#792.3996 - 848.6004

Зад. 2

```
> candies = c(83,35,42,48)
```

#вектор с броя бонбони в пакет от всеки цвят

```
> Pcandies = c(0.4,0.2,0.2,0.2)
```

#вектор със съответните им вероятности

#h0- сините бонбони се срещат с 2 пъти по- голяма вероятност от останалите

#h1- всички бонбони се срещат с еднаква вероятност

```
> chisq.test(candies,p=Pcandies)
```

#тестът връща p- value = 0.565 (> 0.05)

#=> h0 =>можем да приемем твърдението на производителя

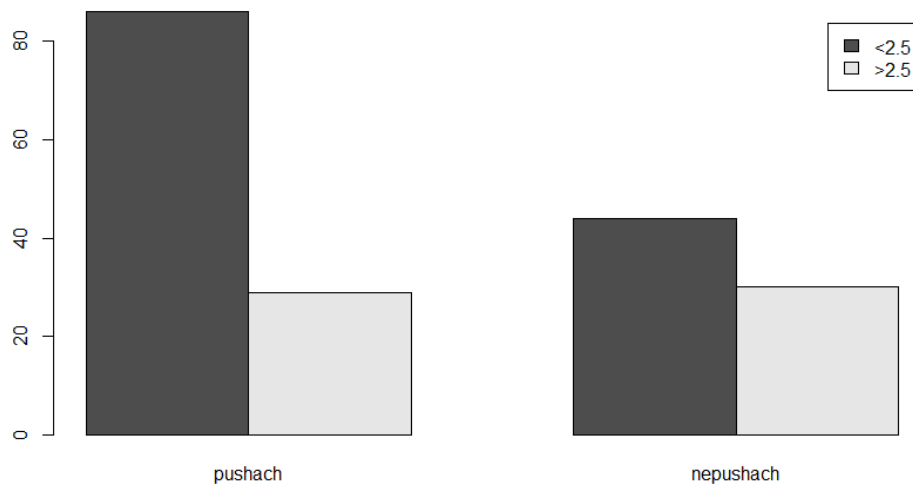
Зад. 3

```
> data = birthwt[,c(1,5)]
```

#взимаме само колоните за “дали бебето е <2.5” и “дали майката е пушач”

```
> t=table(data)
```

```
> barplot(t, beside = T, legend.text = c("<2.5", ">2.5"), names.arg = c("pushach", "nepushach"))
```



#h0- не са зависими

#h1- зависими са

```
> chisq.test(t)
```

#p-value = 0.03958 (< 0.05) => ЛЗ => h1

#=> има зависимост между тютюнопушенето и раждането на деца с по- малко тегло от 2.5кг

Зад. 4

```
> data=read.csv("DomR2.txt")
```

```
#прочитаме информацията от файла
```

```
> l = lm(data$y ~ data$x1 + data$x2 + data$x3)
```

```
> summary(l)
```

```
#запазваме линейния модел и разглеждаме всичките данни  
#за него
```

```
#забелязваме, че има голяма вероятност x2 да е 0
```

```
#=> правим нов линеен модел без него
```

```
> l = lm(data$y ~ data$x1 + data$x3)
```

```
> summary(l)
```

```
#наблюдаваме, че стойността на Adjusted R-squared  
(0.8567) се покачва, т.е. моделът се подобрява
```

```
> l3 = lm(data$y ~ data$x1 + data$x3 - 1)
```

```
> summary(l3)
```

```
#махаме и свободния коефициент, понеже той също е  
#проблемен (като x2) ; Adjusted R-squared (0.8729) се  
#покачва дори повече
```

```
> anova(l,l3)
```

```
#сравняваме разпределението на двата линейни модела
```

```
#Pr = 0.6503 (> 0.05) => това е търсеният модел
```

```
#=>остават променливите x1 и x3
```

```
> coef = -1
```

```
> s = summary(l3)
```

```
> est = s$coefficients[1,1]
```

```
> err = s$coefficients[1,2]
```

```
> t = (est - coef)/err
```

```
#взимаме съответните стойности от таблицата coefficients
```

```
#останалото- формули
```

```
#h0- коефициентът пред x1 е -1
```

```
#h1- коефициентът пред x1 не е -1
```

```
> x1Len=length(data$x1)
```

```
> if (t>0) {
```

```
+ pVal = 2*(1-pt(t, df = x1Len - 2))
```

```
+ } else {
```

```
+ pVal = 2*(pt(t, df=x1Len - 2))
```

```
+ }
```

```
> pVal
```

```
[1] 0.008207128 (< 0.05)
```

```
#=>приемаме алтернативата
```

```
#=>коефициентът пред x1 не е -1
```