



Бизнис статистика

Аудиторски вежби 2

Претставување на податоци со
табели и графици

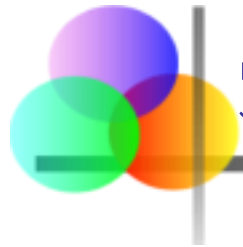


Задача 1.

Следните податоци се просечните неделни хипотекарни каматни стапки за период од 40 недели.

7.29, 7.23, 7.11, 6.78, 7.47, 6.69, 6.77, 6.57, 6.80, 6.88,
6.98, 7.16, 7.30, 7.24, 7.16, 7.03, 6.90, 7.16, 7.40, 7.05,
7.28, 7.31, 6.87, 7.68, 7.03, 7.17, 6.78, 7.08, 7.12, 7.31,
7.40, 6.35, 6.96, 7.29, 7.16, 6.97, 6.96, 7.02, 7.13, 6.84

Да се направи табела со распределба на честоти за овие податоци. Да се определат средните точки на интервалите, релативните фреквенции и кумулативните фреквенции. Потоа, добиените вредности графички да се претстават со хистограми и полигони.



Задача 1: решение

Најмалиот податок е 6.35, а најголемиот 7.68, значи рангот на податоците е $7.68 - 6.35 = 1.33$.

Обем на примерокот: $n = 40$.

Број на интервали во кои треба да се групираат податоците: $r = 1 + 3.322 \cdot \log n = 1 + 3.322 \cdot \log 40 \approx 6.322$, ќе направиме поделба во 7 интервали.

Ширина на интервалите: $1.33 / 7 = 0.19 \approx 0.2$, во формирање на интервалите ќе почнеме од 6.30 до 7.7.

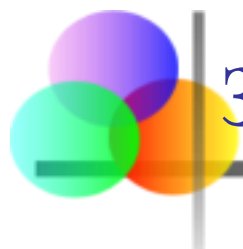


Задача 1: решение - табела

Подредениот примерок е: 6.35, 6.57, 6.69, 6.77, 6.78, 6.78, 6.80, 6.84, 6.87, 6.88, 6.90, 6.96, 6.96, 6.97, 6.98, 7.02, 7.03, 7.03, 7.05, 7.08, 7.11, 7.12, 7.13, 7.16, 7.16, 7.16, 7.17, 7.23, 7.24, 7.28, 7.29, 7.29, 7.30, 7.31, 7.31, 7.40, 7.40, 7.47, 7.68.

Се добива следната табела:

| Инт. | честоти | Рел. честоти | Кум. честоти | Рел. кум. честоти | Рел. честоти % | Рел. Кум. честоти % |
|------------|---------|-----------------|-----------------|----------------------|----------------------|---------------------------|
| [6.3, 6.5) | 1 | 0.025 | 1 | 0.025 | 2.5 | 2.5 |
| [6.5, 6.7) | 2 | 0.050 | 3 | 0.075 | 5.0 | 7.5 |
| [6.7, 6.9) | 7 | 0.175 | 10 | 0.250 | 17.5 | 25.0 |
| [6.9, 7.1) | 10 | 0.250 | 20 | 0.500 | 25.0 | 50.0 |
| [7.1, 7.3) | 13 | 0.325 | 33 | 0.825 | 32.5 | 82.5 |
| [7.3, 7.5) | 6 | 0.150 | 39 | 0.975 | 15.0 | 97.5 |
| [7.5, 7.7) | 1 | 0.025 | 40 | 1 | 2.5 | 100.0 |
| Вкупно | 40 | 1 | | | | |



Задача 1: решение - табела со R

```
d = c(7.29, 7.23, 7.11, 6.78, 7.47, 6.69, 6.77, 6.57, 6.80, 6.88, 6.98, 7.16, 7.30, 7.24, 7.16,  
7.03, 6.90, 7.16, 7.40, 7.05, 7.28, 7.31, 6.87, 7.68, 7.03, 7.17, 6.78, 7.08, 7.12, 7.31, 7.40, 6.35,  
6.96, 7.29, 7.16, 6.97, 6.96, 7.02, 7.13, 6.84)
```

```
n = length(d)
```

```
sort(d)
```

```
breaks1 = seq(6.3, 7.7, by=0.2)
```

```
d.int = cut(d, breaks1, right=FALSE)
```

```
freq = table(d.int)
```

```
Rfreq = freq/n
```

```
Cumfreq = cumsum(freq)
```

```
R_Cumfreq = cumsum(freq)/n
```

```
R_Cumfreq2 = cumsum(Rfreq)
```

```
Pfreq = Rfreq*100
```

```
P_Cumfreq = R_Cumfreq*100
```

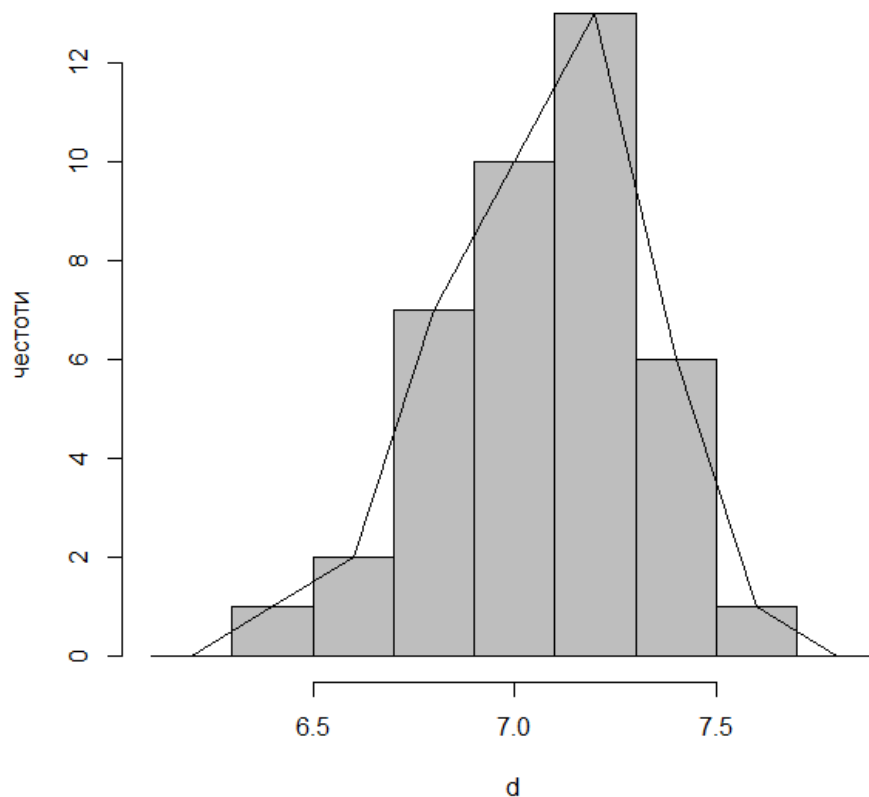
```
d.table = cbind(freq, Rfreq, Cumfreq, Pfreq, P_Cumfreq)
```



Задача 1 - Хистограм на честоти

| Инт. | честоти |
|------------|---------|
| [6.3, 6.5) | 1 |
| [6.5, 6.7) | 2 |
| [6.7, 6.9) | 7 |
| [6.9, 7.1) | 10 |
| [7.1, 7.3) | 13 |
| [7.3, 7.5) | 6 |
| [7.5, 7.7) | 1 |

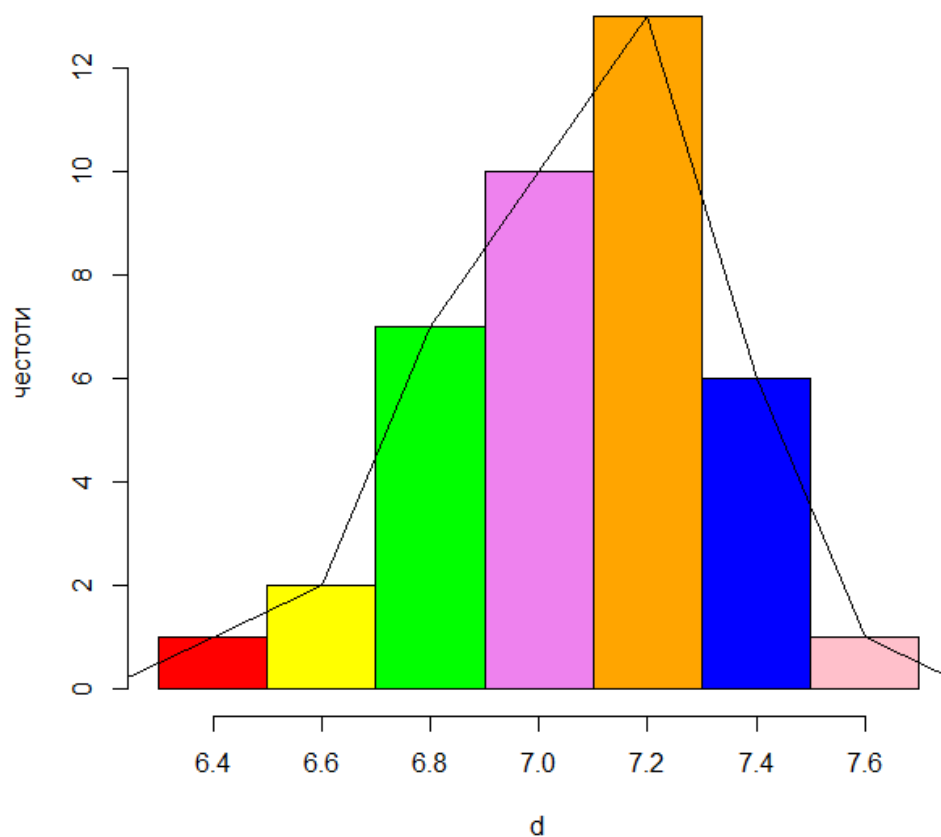
Неделни хипотекарни каматни стапки

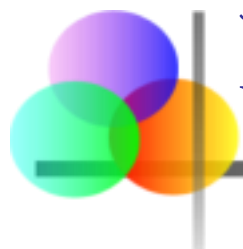




Задача 1 - Хистограм на честоти

Неделни хипотекарни каматни стапки

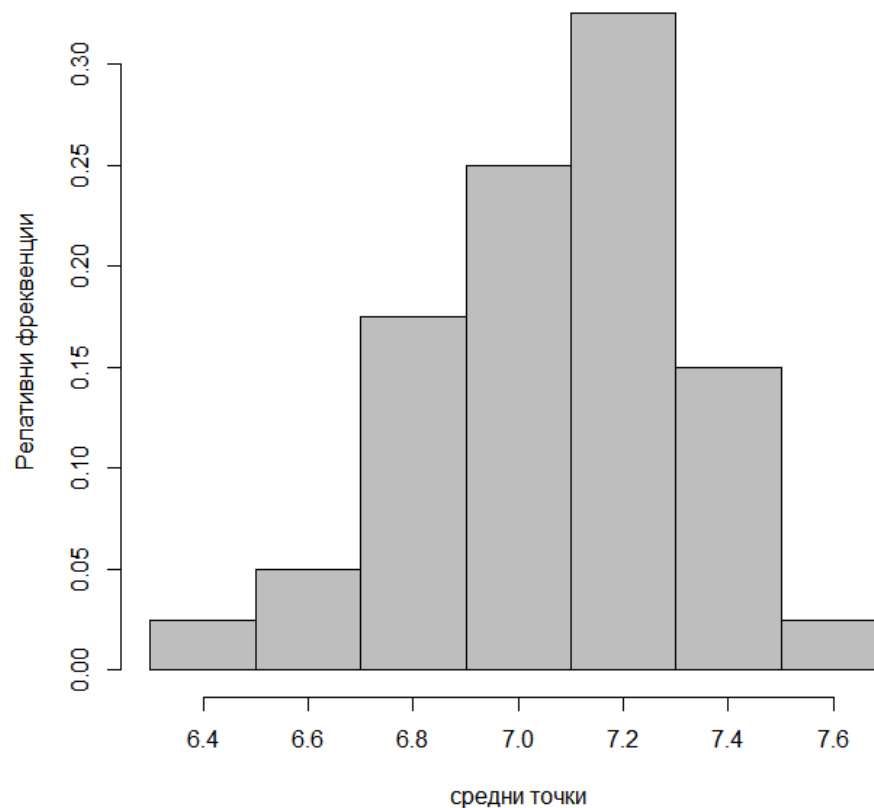


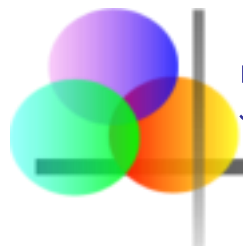


Задача 1: Хистограм на релативни честоти

| Инт. | Рел. честоти |
|------------|--------------|
| [6.3, 6.5) | 0.025 |
| [6.5, 6.7) | 0.050 |
| [6.7, 6.9) | 0.175 |
| [6.9, 7.1) | 0.250 |
| [7.1, 7.3) | 0.325 |
| [7.3, 7.5) | 0.150 |
| [7.5, 7.7) | 0.025 |
| Вкупно | 1 |

Неделни хипотекарни каматни стапки

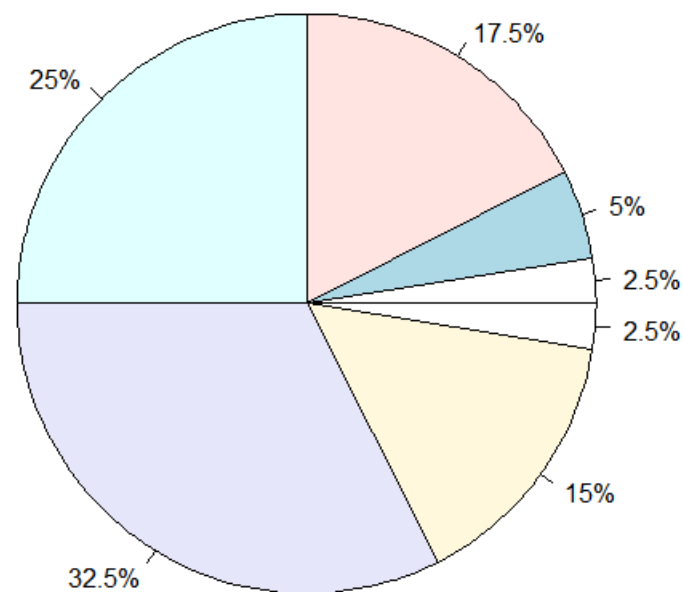




Задача 1: Пита график за честота во %

Честоти во проценти

| Инт. | честоти % |
|------------|--------------|
| [6.3, 6.5) | 2.5 |
| [6.5, 6.7) | 5.0 |
| [6.7, 6.9) | 17.5 |
| [6.9, 7.1) | 25.0 |
| [7.1, 7.3) | 32.5 |
| [7.3, 7.5) | 15.0 |
| [7.5, 7.7) | 2.5 |





Задача 1: решение – Графици R

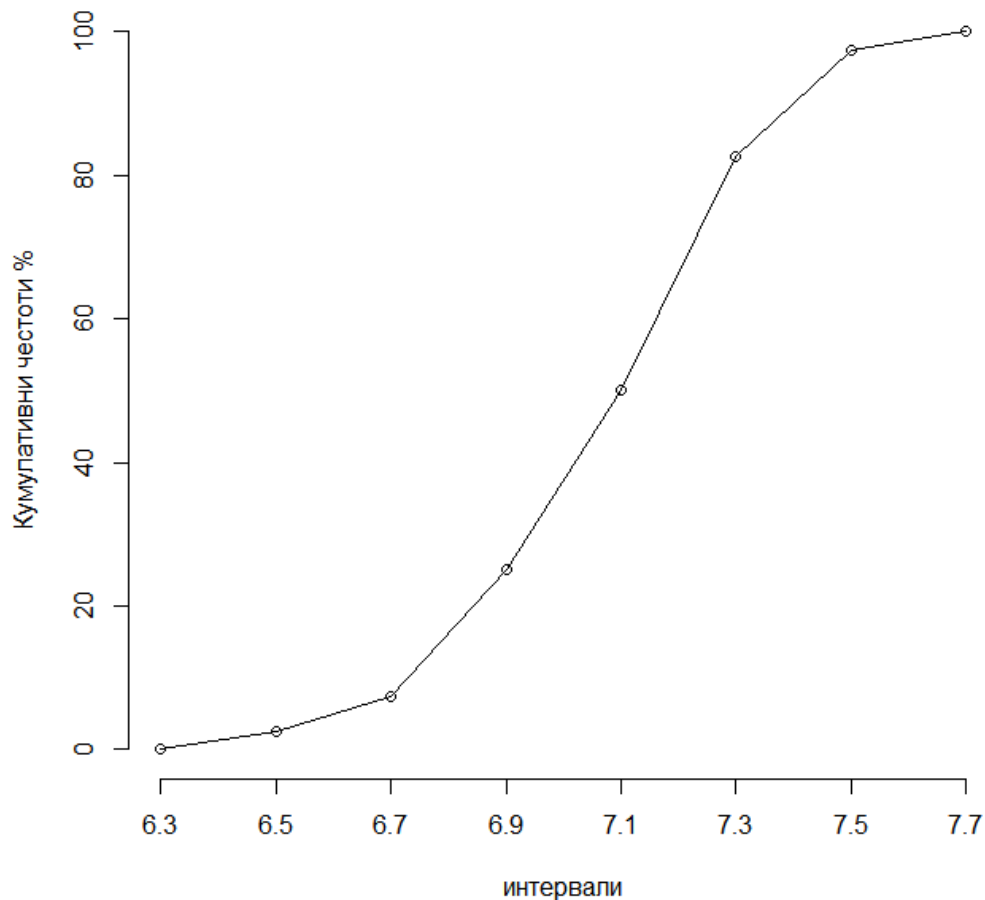
```
breaks2 = seq(6.1, 7.9, by=0.2)
breaks1 = seq(6.3, 7.7, by=0.2)
d.int = cut(d, breaks2, right=FALSE)
freq = table(d.int)
mid = c()
for (i in 1:length(breaks2)-1)
  {mid = c(mid, (breaks2[i]+breaks2[i+1])/2)} # определяне на средните точки

h1 = hist(d, right = FALSE, breaks = breaks2, col = "grey", main = "Неделни хипотекарни
каматни стапки", ylab = "честоти")
colors = c("red", "yellow", "green", "violet", "orange", "blue", "pink")
h1 = hist(d, right = FALSE, breaks = breaks1, col = colors, main = "Неделни хипотекарни
каматни стапки", ylab = "честоти")
lines(mid,freq)
h1$counts = h1$counts / sum(h1$counts)
plot(h1, freq = TRUE, col = "grey",ylab = "Релативни фреквенции",
xlab = "средни точки", main = "Неделни хипотекарни каматни стапки")
procent_freq = paste(Pfreq,"%",sep = "")
procent_freq
pie(Pfreq,labels = procent_freq, main = "Честоти во проценти")
```



Задача 1: решение - Графици

Полигон на кумулативни честоти во % (Ogive)



```
P_Cumfreq0 = c(0, P_Cumfreq)
plot(breaks1, P_Cumfreq0, axes = F,
     main = "Полигон на кумулативни
            честоти во % (Ogive)",
     xlab = "интервали",
     ylab = "Кумулативни честоти %")
axis(side = 1, at = breaks1)
axis(side = 2)
```

```
lines(breaks1, P_Cumfreq0)
```

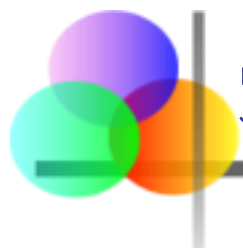


Задача 2.

Менаџерот за човечки ресурси во една голема компанија сака да направи анализа за отсуството од работа на 500 вработени во компанијата во текот на изминатата година. Истражувачот ги организира податоците во следната табела со распределба на фреквенции за да му помогне на менаџерот за човечки ресурси во анализирањето на податоците.

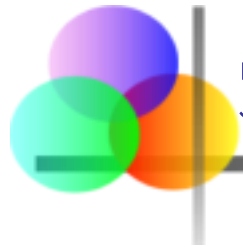
| Интервал | Честота |
|-----------|---------|
| $[0, 2)$ | 218 |
| $[2, 4)$ | 207 |
| $[4, 6)$ | 56 |
| $[6, 8)$ | 11 |
| $[8, 10)$ | 8 |

Да се одреди средната точка за секој интервал, релативната фреквенција и кумулативната фреквенција.



Задача 2: решение

| Интервал | Средна точка | Честота | Релативна честота | Релативна честота (%) | Кумулативна честота | Рел. Кумулативна честота (%) |
|----------|--------------|---------|-------------------|-----------------------|---------------------|------------------------------|
| [0, 2) | 1 | 218 | 0.436 | 43.6 | 218 | 43.6 |
| [2, 4) | 3 | 207 | 0.414 | 41.4 | 425 | 85.0 |
| [4, 6) | 5 | 56 | 0.112 | 11.2 | 481 | 96.2 |
| [6, 8) | 7 | 11 | 0.022 | 2.2 | 492 | 98.4 |
| [8, 10] | 9 | 8 | 0.016 | 1.6 | 500 | 100.0 |
| Вкупно | | 500 | | | | |



Задача 3.

Следните податоци ги претставуваат трошоците (во долари) за испраќање по пошта на 30 писма од страна на компанијата.

3.67, 2.75, 9.15, 5.11, 3.32, 2.09, 1.83, 10.94, 1.93, 3.89,
7.20, 2.78, 6.72, 7.80, 5.47, 4.15, 3.55, 3.53, 3.34, 4.95,
5.42, 8.64, 4.84, 4.10, 5.10, 6.45, 4.65, 1.97, 2.84, 3.21

Користејќи ги доларите како стебла, центите како листови, да се конструира стебло-лист дијаграм.



Задача 3: Решение

3.67, 2.75, 9.15, 5.11, 3.32, 2.09, 1.83, 10.94, 1.93, 3.89,
7.20, 2.78, 6.72, 7.80, 5.47, 4.15, 3.55, 3.53, 3.34, 4.95,
5.42, 8.64, 4.84, 4.10, 5.10, 6.45, 4.65, 1.97, 2.84, 3.21

Примерокот се подредува во неопаѓачки редослед:

| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|-------|
| 1.83 | 1.93 | 1.97 | 2.09 | 2.75 | 2.78 | 2.84 | 3.21 | 3.32 | 3.34 |
| 3.53 | 3.55 | 3.67 | 3.89 | 4.10 | 4.15 | 4.65 | 4.84 | 4.95 | 5.10 |
| 5.11 | 5.42 | 5.47 | 6.45 | 6.72 | 7.20 | 7.80 | 8.64 | 9.15 | 10.94 |



Задача 3: Решение

| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|-------|
| 1.83 | 1.93 | 1.97 | 2.09 | 2.75 | 2.78 | 2.84 | 3.21 | 3.32 | 3.34 |
| 3.53 | 3.55 | 3.67 | 3.89 | 4.10 | 4.15 | 4.65 | 4.84 | 4.95 | 5.10 |
| 5.11 | 5.42 | 5.47 | 6.45 | 6.72 | 7.20 | 7.80 | 8.64 | 9.15 | 10.94 |

1 | 83 93 97
2 | 09 75 78 84
3 | 21 32 34 53 55 67 89
4 | 10 15 65 84 95
5 | 10 11 42 47
6 | 45 72
7 | 20 80
8 | 64
9 | 15
10 | 94



Задача 3: Решение

```
1 | 83 93 97
2 | 09 75 78 84
3 | 21 32 34 53 55 67 89
4 | 10 15 65 84 95
5 | 10 11 42 47
6 | 45 72
7 | 20 80
8 | 64
9 | 15
10 | 94
```

```
cost = c(3.67, 2.75, 9.15, .... , 3.21)
```

Во R има функција `stem()`, но не може да се дефинира кој дел од бројот е стебло, а кој лист. За овој пример прво бројот го заокружува на една децимала, потоа целиот дел го зема за стебло, а лист е децималата.

На следниот слајд е дадена функција со која може да се дефинира дијаграмот.

```
stem(cost)
1 | 89
2 | 01888
3 | 2335679
4 | 1278
5 | 01145
6 | 57
7 | 28
8 | 6
9 | 2
10 | 9
```



Задача 3: Решение

```
cost = sort(cost)
```

```
library(data.table)
```

```
# leftDigits - pozicija na '|' vo odnos na decimalnata tocka, rounding - br. na decimali,
```

```
myStem <- function(x, leftDigits, rounding = 1) {  
  data = data.table("x" = x)  
  data[, left := floor(x/10^leftDigits)]  
  data[, right := (round(x - left*10^leftDigits, rounding))*10^rounding]  
  data = data[, paste(sort(right), collapse = " "), by = left]  
  data[, out := paste(left, " | ", V1), by = left]  
  cat(data$out, sep = "\n")  
}  
myStem(cost, 0, 2)
```



Задача 4.

Дали парите потрошени за рекламирање се во релација со вкупните приходи од продажба на компанијата? Следните податоци ги претставуваат парите (во милион долари) потрошени за рекламирање и приходите од продажби за различни компании во дадена индустрија за последната година. Да се нацрта график на расејување за податоците од двете обележја и да се дискутира односот (врската) помеѓу двете обележја.

| Рекламирање | Продажба |
|-------------|----------|
| 4.2 | 155.7 |
| 1.6 | 87.3 |
| 6.3 | 135.6 |
| 2.7 | 99.0 |
| 10.4 | 168.2 |
| 7.1 | 136.9 |
| 5.5 | 101.4 |
| 8.3 | 158.2 |



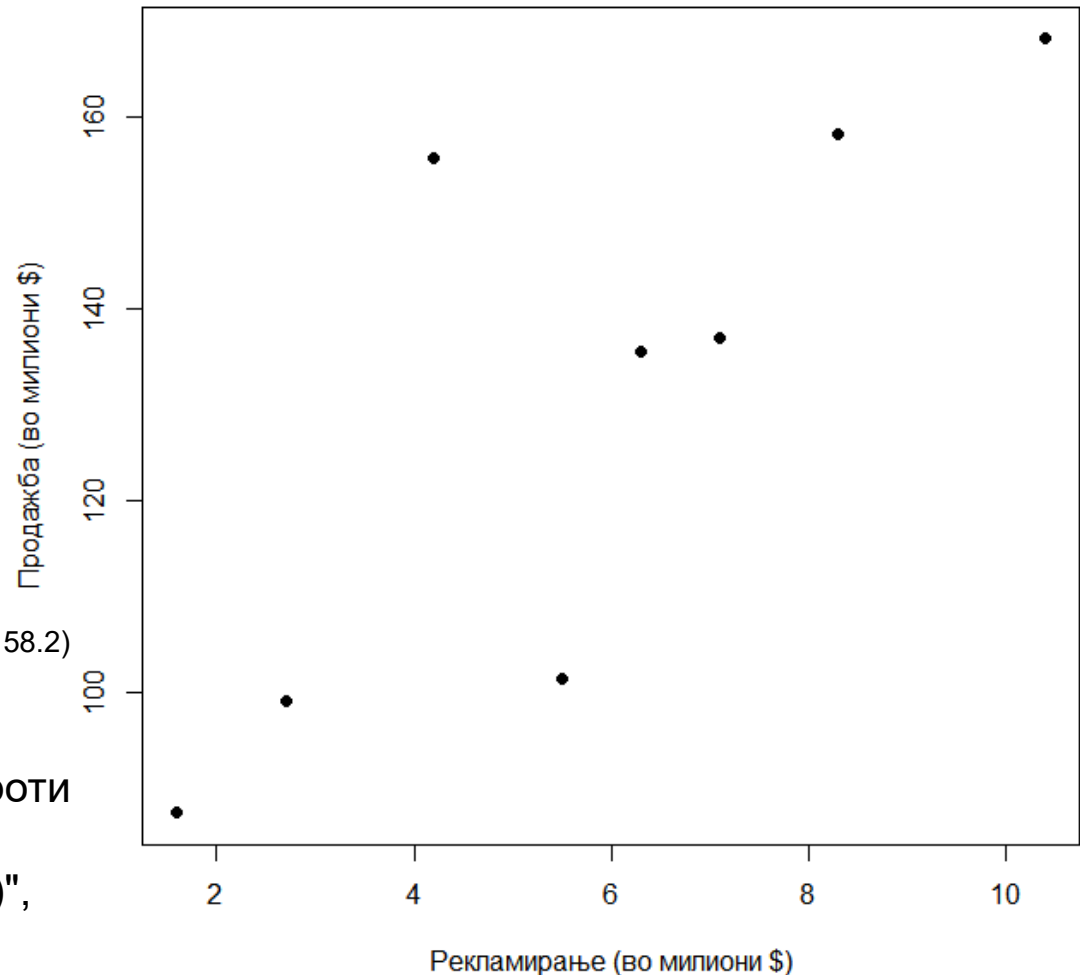
Задача 4: Решение

| Реклами | Продажба |
|---------|----------|
| 4.2 | 155.7 |
| 1.6 | 87.3 |
| 6.3 | 135.6 |
| 2.7 | 99.0 |
| 10.4 | 168.2 |
| 7.1 | 136.9 |
| 5.5 | 101.4 |
| 8.3 | 158.2 |

```
rek = c(4.2, 1.6, 6.3, 2.7, 10.4, 7.1, 5.5, 8.3)  
prod = c(155.7, 87.3, 135.6, 99.0, 168.2, 136.9, 101.4, 158.2)
```

```
plot(rek,prod,pch = 19,  
main = "Пари за рекламирање наспроти  
продажба",  
xlab = "Рекламирање (во милиони $)",  
ylab = "Продажба (во милиони $)")
```

Пари за рекламирање наспроти продажба



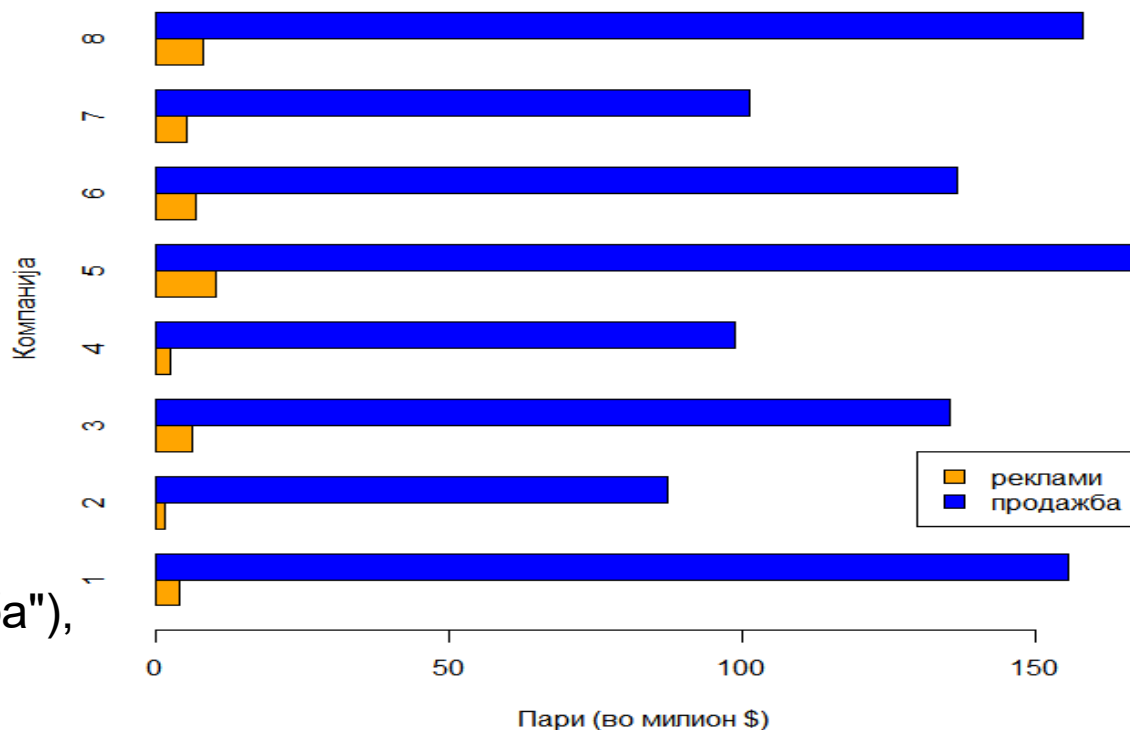
Задача 4:Дополнување

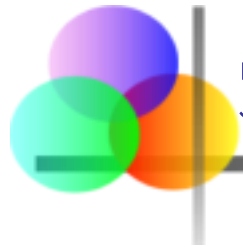
Паралелен столбест дијаграм во R



```
реклами = c(4.2, 1.6, 6.3, 2.7, 10.4, 7.1, 5.5, 8.3)
продажба = c(155.7, 87.3, 135.6, 99.0, 168.2, 136.9, 101.4, 158.2)
tabela = rbind(реклами,продажба)
```

```
barplot(tabela, beside = TRUE,
  col = c("orange","blue"),
  names.arg = c(1:8),
  horiz = TRUE,
  ylab = "Компанија",
  xlab = "Пари (во милион $)")
legend(130, 7,
  legend = c("реклами","продажба"),
  fill=c("orange","blue"))
```





Задача 5.

Во табелата се дадени податоци по колку пари (во долари) троши просечен студент за различни ставки.

| | |
|------------------------|--------|
| Електроника | 211.89 |
| Облека и додатоци | 134.4 |
| Мебел | 90.9 |
| Факултетски материјали | 68.47 |
| Останато | 93.72 |

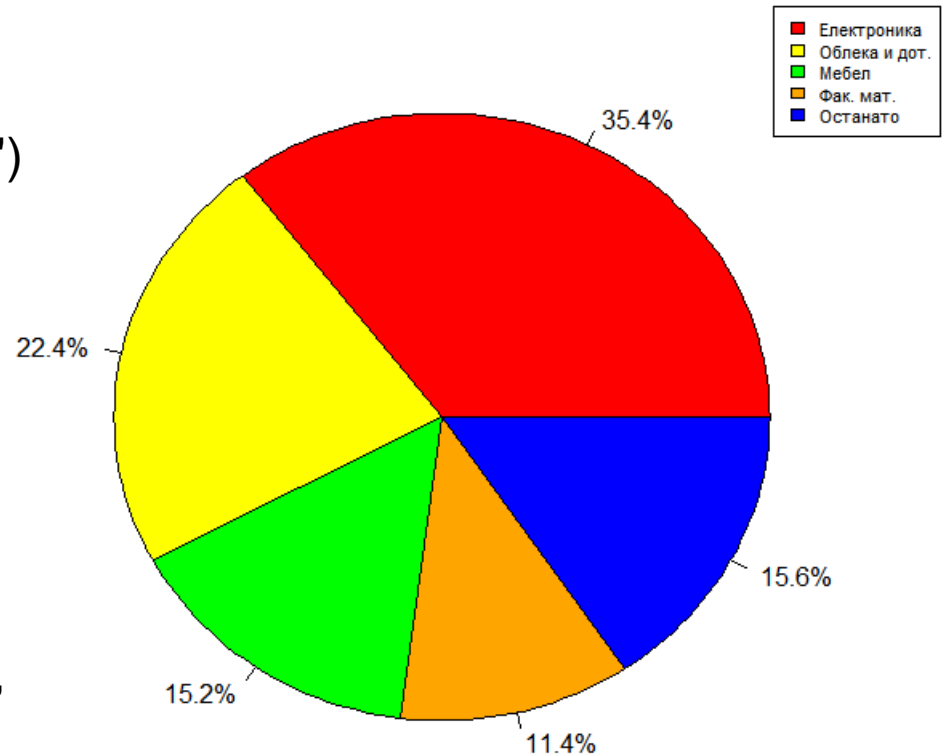
Податоците да се претстават графички.



Задача 5. Решение

```
trosoci = matrix(c(211.89, 134.4,  
93.72, 90.9, 68.47),nrow = 5,byrow =  
T)  
names = c("Електроника","Облека и  
дот.,""Останато","Мебел","Фак. мат.")  
dimnames(trosoci) = list(names)  
n = sum(trosoci)  
procent = round(trosoci/n*100,1)  
p = data.frame(trosoci,procent)  
colors = c("red", "yellow", "green",  
"orange", "blue")  
  
pie(procent,  
labels = paste(procent, "%", sep = ""),  
col = colors,  
main = "Студентски трошоци")  
legend("topright", names, cex = 0.7,  
fill = colors)
```

Студентски трошоци

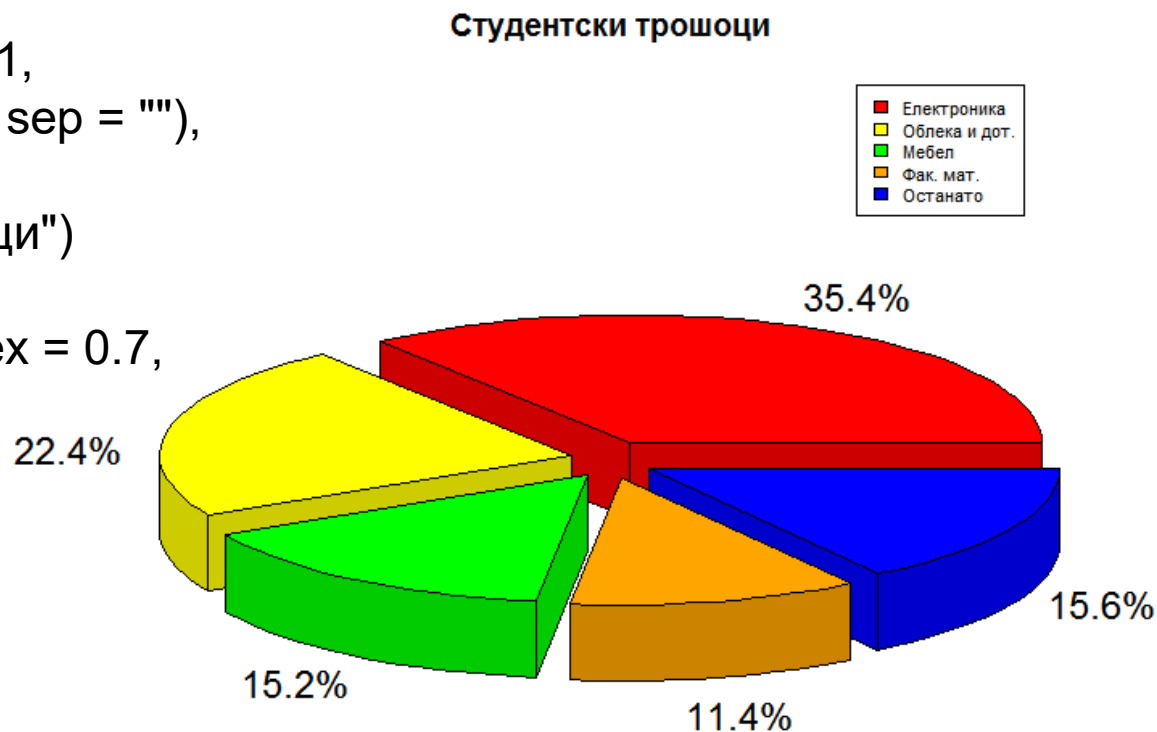


Задача 5. Решение

```
library(plotrix)
```

```
pie3D(procent, explode = 0.1,  
labels = paste(procent, "%", sep = ""),  
col = colors,  
main = "Студентски трошоци")
```

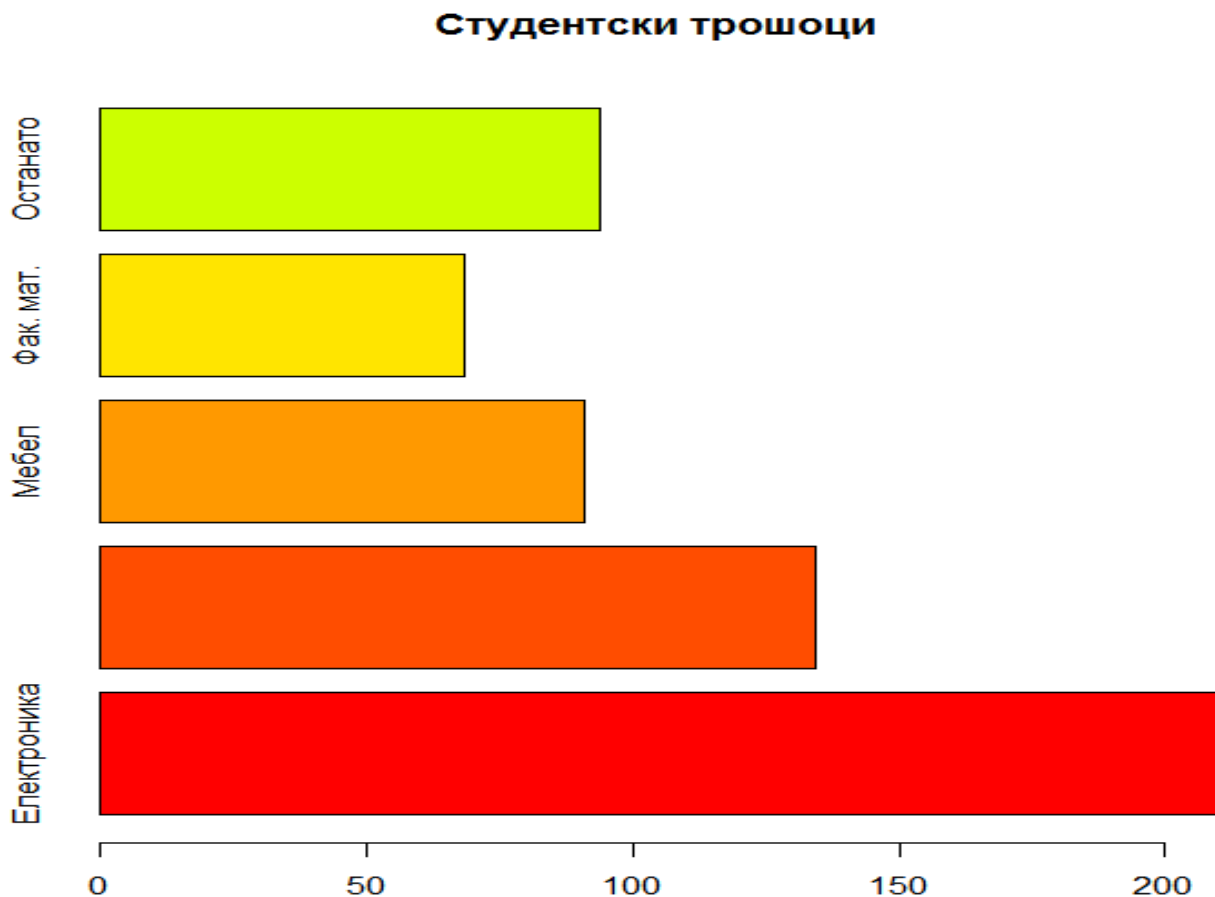
```
legend("topright", names, cex = 0.7,  
fill = colors)
```





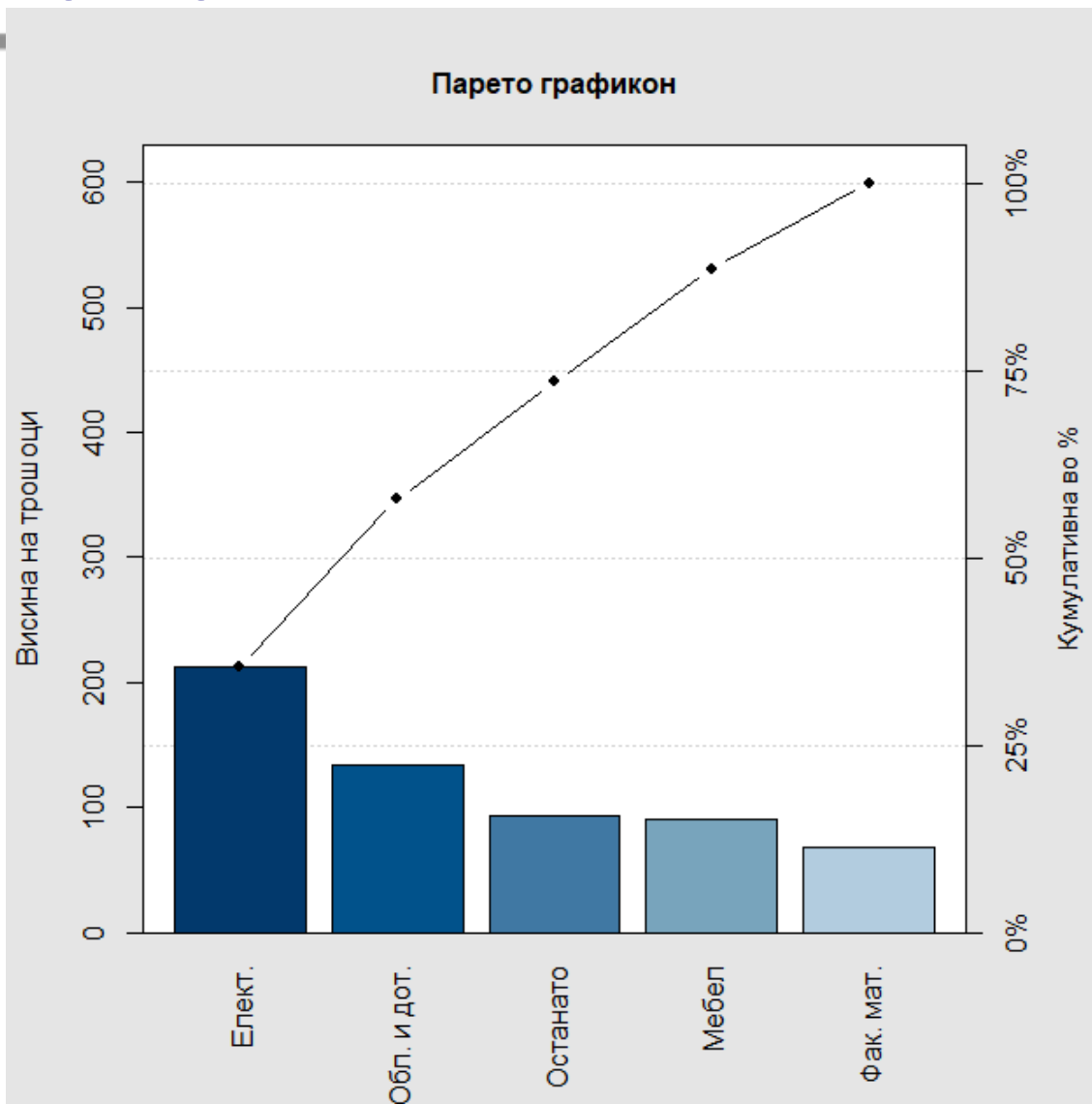
Задача 5. Решение

```
barplot(c(211.89, 134.4, 90.9, 68.47, 93.72),main = "Студентски трошоци",  
horiz = TRUE, col = rainbow(20),names.arg = names)
```



Задача 5. Решение

| | трошоци | честоти | процент | кум.(%) |
|-------------|---------|---------|---------|---------|
| Елект. | 211.89 | 0.354 | 35.4 | 35.4 |
| Обл. и дод. | 134.40 | 0.224 | 22.4 | 57.8 |
| Останато | 93.72 | 0.156 | 15.6 | 73.4 |
| Мебел | 90.90 | 0.152 | 15.2 | 88.6 |
| Фак. Мат. | 68.47 | 0.114 | 11.4 | 100.0 |





Задача 5. Решение

```
trosoci = matrix(c(211.89, 134.4, 93.72, 90.9, 68.47),nrow = 5,byrow = T)
names = c("Елект.", "Обл. и дот.", "Останато", "Мебел", "Фак. мат.")
dimnames(trosoci) = list(names)
df = data.frame(trosoci)
library(qcc)
pareto.chart(df$trosoci, names.arg = names, ylab = "Висина на трошоци", ylab2 =
"Кумулативна во %", main = "Парето графикон")
```

Pareto chart analysis for df\$trosoci

| | Frequency | Cum.Freq. | Percentage | Cum.Percent. |
|---|-----------|-----------|------------|--------------|
| A | 211.89000 | 211.89000 | 35.35153 | 35.35153 |
| B | 134.40000 | 346.29000 | 22.42317 | 57.77470 |
| C | 93.72000 | 440.01000 | 15.63616 | 73.41086 |
| D | 90.90000 | 530.91000 | 15.16567 | 88.57653 |
| E | 68.47000 | 599.38000 | 11.42347 | 100.00000 |