

# Домашна работа 1

## Статистика практикум

Петко Каменов. ФН :45546

### Задача 1.

а) `length(age[age<24])` . Резултат: 373

б) `length(age[age<24&smoke==1])` . Резултат: 167

в) Код: `gestLen = length(gestation)`

```
SortedLenOfGestation = sort(gestation)
```

```
mean(wt[SortedLenOfGestation[1:20]],na.rm=T)
```

```
mean(wt[SortedLenOfGestation[21:gestLen]],na.rm=T)
```

Резултат:

- Средно тегло на бебетата 20 майките с най-кратка бременност: 122.25 унции
- Средно тегло на останалите бебета: 122.6562 унции

г) Код: `wt1[which(wt1==999)]=NA`

```
inc[which(inc==98|inc==99)]=NA
```

```
boxplot(wt1 ~ inc,
```

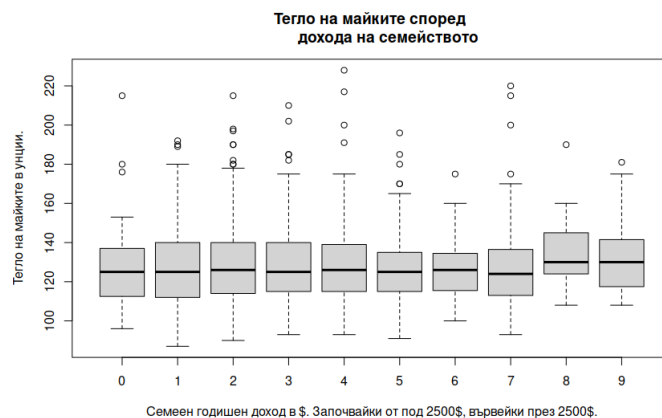
```
ylab="Тегло на майките в унции.",
```

```
xlab="Семеен годишен доход в $. Започвайки от под 2500$, вървейки през 2500$.",
```

```
main="Тегло на майките според
```

```
дохода на семейството")
```

Резултат:



На таблицата може да видим средните стойности за теглата на майките според доходите, двата квантила 25 и 75, както и минималното и максимално тегло. Интересно е, че средното тегло е едно и също почти при всички доходи с изключение на хората с най-високи годишни такива. Също така се вижда, че най-леките майки не са наблюдавани в най-бедната група, а в тези с малко повече доходи, категории 1,2,3,4,5.

д) Код:

```
smoke[which(smoke==9)]=NA

white = (race==0 | race==1 | race==2 | race==3 | race==4 |
race==5)

black = (race==7)

others = (race==6 | race ==8 | race == 9)

lbls <- c("Никога", "В момента", "Допреди бременността",
"Някога да, сега не")

tableWhite = table(smoke, white)
tableBlack = table (smoke,black)
tableOthers = table (smoke,others)

pie(tableWhite[,2],labels = lbls, main = "Пушене
на белите майки")

pie(tableBlack[,2],labels=lbls, main = "Пушене на черните
майки")

pie(tableOthers[,2],labels=lbls, main = "Пушене
на другите майки")
```

Резултат:

Пушене на черните майки



Пушене на белите майки



Пушене на другите майки

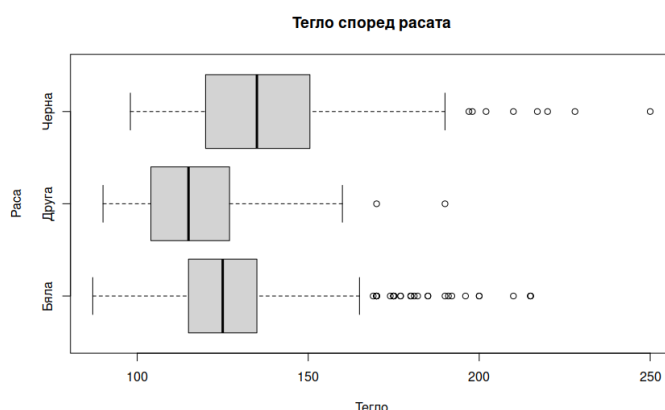


На трите графики може да наблюдаваме, че най-големият процент пушачки е при белите майки, следвани от черните и накрая другите. Черните пък са най-склонни да се откажат да пушат след бременността си, но това се дължи и на факта, че те са повече пушачи от „другите майки“. Белите пък от своя страна са най-голям процент от тези които някога са пушили, но сега вече не.

е) Код:

```
f = cut(race, c((-1:5), (6), (7), (8), (9)))
levels(f) = c("Бяла", "Бяла", "Бяла", "Бяла", "Бяла",
              "Бяла", "Друга", "Черна", "Друга", "Друга")
wt1[which(wt1==999)] = NA
plot(wt1 ~ f, horizontal=T, main="Тегло
      според расата", xlab="Раса", ylab="Тегло")
```

Резултат:

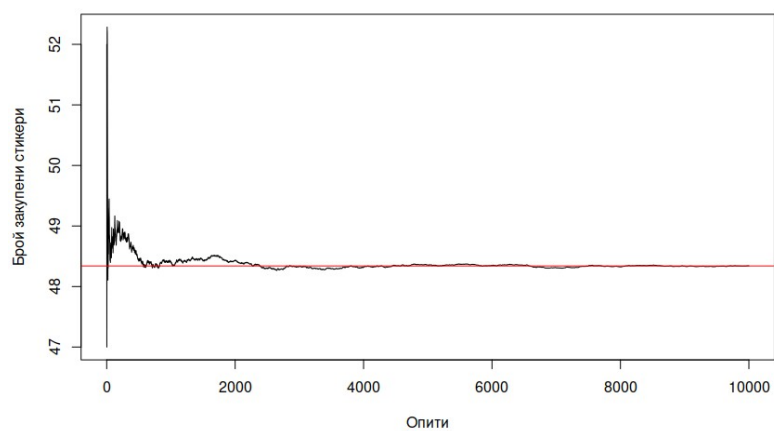


Теглото на майките от негроидната раса, както може да бъде видно от графиката, са не само с най-високо средно тегло, но и с най-високо минимално и максимално такова. На второ място по средно тегло са белите, а на трето останалите раси. От друга страна пък, майките с най-ниско тегло са не от останалите раси, а от бялата раса.

Задача 2. Код:

```
46 # Заг. 2
47 allstickers = function(arr) {
48   for(i in 1:20) {
49     if(arr[i]==F) return(F);
50   }
51   return(T);
52 }
53
54
55 stickers = function() {
56   N = 20;
57   x = sample(1:20,N,replace=T);
58   arr = rep.int(F,20);
59   for(i in 1:N) {
60     arr[x[i]] = T;
61   }
62   N = N + 1;
63   while (!allstickers(arr)) {
64     x = sample(1:20,N,replace=T);
65     arr = rep.int(F,20);
66     for(i in 1:N) {
67       arr[x[i]] = T;
68     }
69     N = N + 1;
70   }
71   return(N);
72 }
73
74 prob.stickers = function( n ) {
75   x = rep.int(0,n)
76   c = 0
77   for(i in 1 : n ) {
78     c = c + stickers()
79     x[i] = c / (1 * i)
80   }
81   return(x)
82 }
83 x = prob.stickers(10000)
84 plot(x,type="l",ylab = "Брой закупени стикери",xlab = "Опити")
85 abline(h = 48.34, col = 'red')
```

Резултат:



След използването на Монте-Карло метода за апроксимация на нужните опити за събиране на всички 20 стикера, виждаме че тази стойност е между 48 и 49 пакета с бонбони.