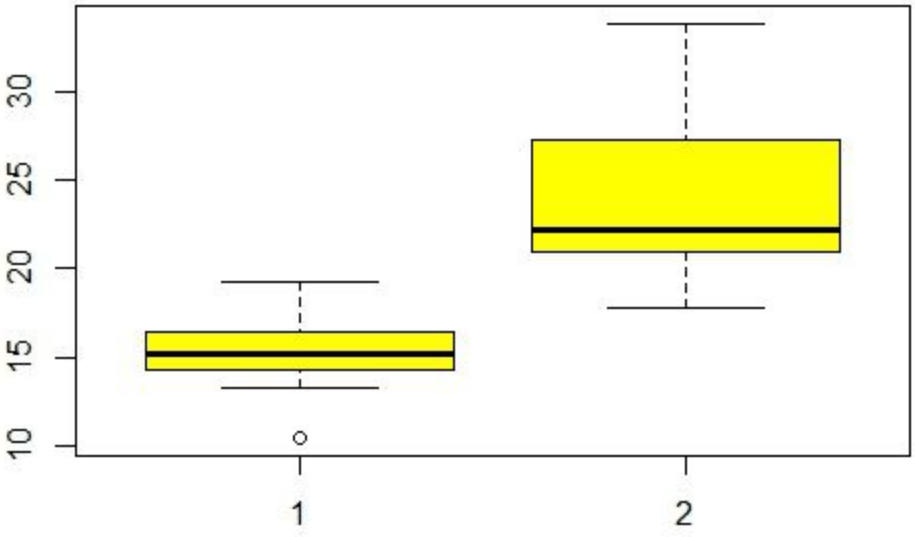
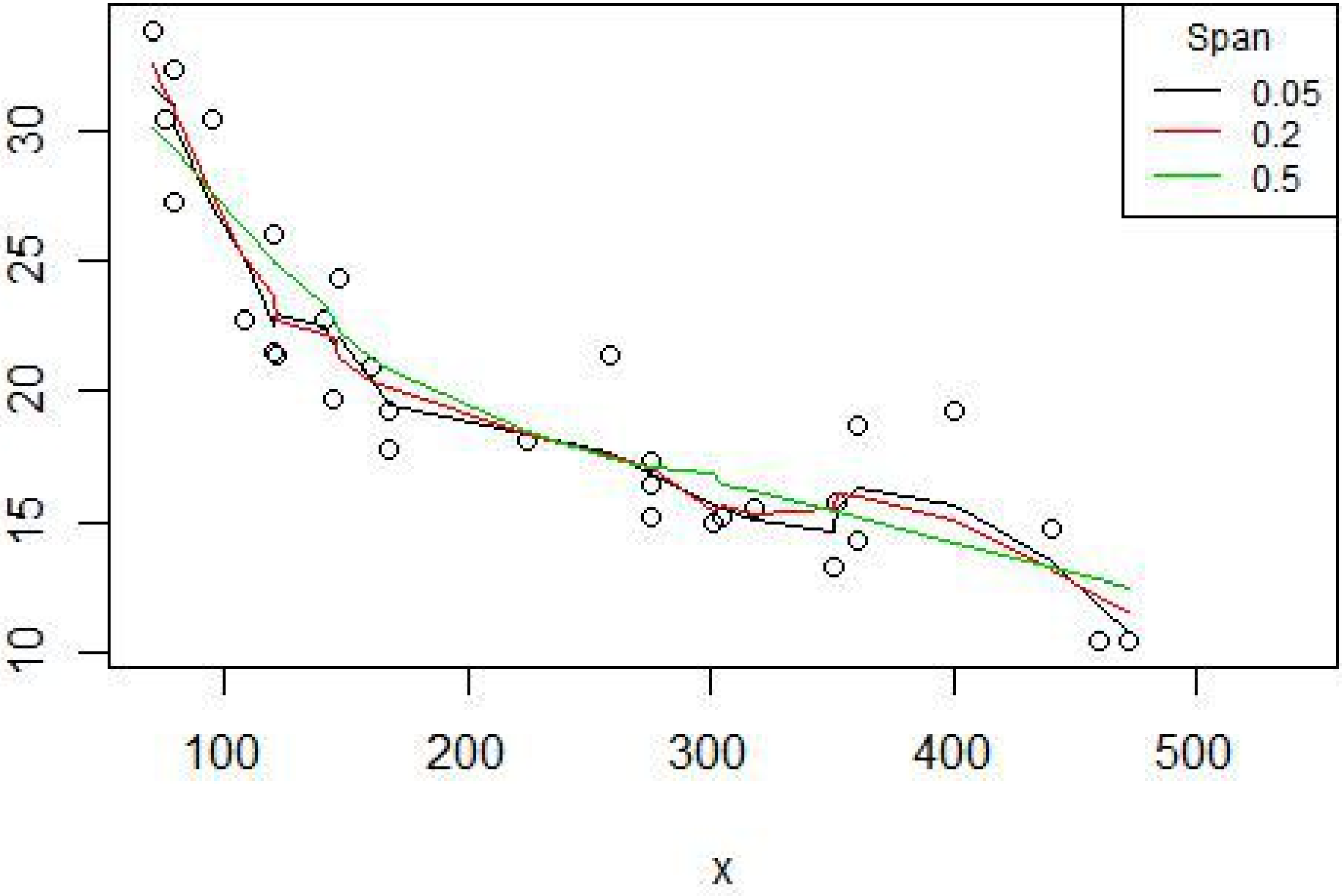
```
> car #OK
       mpg cyl
 [1,] 21.0   6 160.0 110 3.90 2.620 16.46 0 1 4 4
 [2,] 21.0   6 160.0 110 3.90 2.875 17.02 0 1 4 4
 [3,] 22.8   4 108.0  93 3.85 2.320 18.61 1 1 4 1
 [4,] 21.4   6 258.0 110 3.08 3.215 19.44 1 0 3 1
 [5,] 18.7   8 360.0 175 3.15 3.440 17.02 0 0 3 2
 [6,] 18.1   6 225.0 105 2.76 3.460 20.22 1 0 3 1
 [7,] 14.3   8 360.0 245 3.21 3.570 15.84 0 0 3 4
 [8,] 24.4   4 146.7  62 3.69 3.190 20.00 1 0 4 2
 [9,] 22.8   4 140.8  95 3.92 3.150 22.90 1 0 4 2
[10,] 19.2   6 167.6 123 3.92 3.440 18.30 1 0 4 4
[11,] 17.8   6 167.6 123 3.92 3.440 18.90 1 0 4 4
[12,] 16.4   8 275.8 180 3.07 4.070 17.40 0 0 3 3
[13,] 17.3   8 275.8 180 3.07 3.730 17.60 0 0 3 3
[14,] 15.2   8 275.8 180 3.07 3.780 18.00 0 0 3 3
[15,] 10.4   8 472.0 205 2.93 5.250 17.98 0 0 3 4
[16,] 10.4   8 460.0 215 3.00 5.424 17.82 0 0 3 4
[17,] 14.7   8 440.0 230 3.23 5.345 17.42 0 0 3 4
[18,] 32.4   4  78.7  66 4.08 2.200 19.47 1 1 4 1
[19,] 30.4   4  75.7  52 4.93 1.615 18.52 1 1 4 2
[20,] 33.9   4  71.1  65 4.22 1.835 19.90 1 1 4 1
[21,] 21.5   4 120.1  97 3.70 2.465 20.01 1 0 3 1
[22,] 15.5   8 318.0 150 2.76 3.520 16.87 0 0 3 2
[23,] 15.2   8 304.0 150 3.15 3.435 17.30 0 0 3 2
[24,] 13.3   8 350.0 245 3.73 3.840 15.41 0 0 3 4
[25,] 19.2   8 400.0 175 3.08 3.845 17.05 0 0 3 2
[26,] 27.3   4  79.0  66 4.08 1.935 18.90 1 1 4 1
[27,] 26.0   4 120.3  91 4.43 2.140 16.70 0 1 5 2
[28,] 30.4   4  95.1 113 3.77 1.513 16.90 1 1 5 2
[29,] 15.8   8 351.0 264 4.22 3.170 14.50 0 1 5 4
[30,] 19.7   6 145.0 175 3.62 2.770 15.50 0 1 5 6
[31,] 15.0   8 301.0 335 3.54 3.570 14.60 0 1 5 8
[32,] 21.4   4 121.0 109 4.11 2.780 18.60 1 1 4 2
```


10
15
20
25
30
1
2

```
> # car <- mtcars
> with_8_cyl = car[which(car[,2] == 8), 1] # все, у которых 8 цилиндров
> with_4_or_6_cyl = car[-which((car[,2] == 8)), 1] # все, у которых не 8
> length(with_8_cyl) # есть 14 машин с 8 цилиндрами
[1] 14
> length(with_4_or_6_cyl) # и 18 --- с 4 или 6 цилиндрами
[1] 18
> boxplot(with_8_cyl, with_4_or_6_cyl, col="yellow")
> wilcox.test(with_8_cyl, with_4_or_6_cyl, exact=FALSE, alternative="greater")

        Wilcoxon rank sum test with continuity correction

data:  with_8_cyl and with_4_or_6_cyl
W = 4.5, p-value = 1
alternative hypothesis: true location shift is greater than 0

> # Медианы не совпадают, причем p-значение равно 1
>
```

```
> length(with_8_cyl) # для этих 14 машин надо найти максимально похожие среди оставшихся
18
[1] 14
> features = cyl
> for (i in 1:9){
+   features = cbind(features, mtcars[,i+2])
+ }
> scaled <- scale(features)
> with_4_or_6_all = car[-which((car[,2] == 8)), ]
> not_eight <- scale(with_4_or_6_all)
> dst <- rep(0, length(with_4_or_6_cyl))
> res <- rep(0, length(with_8_cyl))
> for (i in 1:length(with_8_cyl)){
+   x = scaled[i,]
+   for(j in 1:length(with_4_or_6_cyl)){
+     dst[j] = dist(rbind(x, not_eight[j,]))
+   }
+   res[i] = mpg[which.min(dst)]
+ }
There were 50 or more warnings (use warnings() to see the first 50)
> wilcox.test(res, mpg[1:length(with_8_cyl)], alternative = "two.sided", paired = TRUE)

        Wilcoxon signed rank test with continuity correction

data:  res and mpg[1:length(with_8_cyl)]
V = 80, p-value = 0.01746
alternative hypothesis: true location shift is not equal to 0

Warning messages:
1: In wilcox.test.default(res, mpg[1:length(with_8_cyl)], alternative = "two.sided",  :
  не могу подсчитать точное p-значение при наличии повторяющихся наблюдений
2: In wilcox.test.default(res, mpg[1:length(with_8_cyl)], alternative = "two.sided",  :
  не могу высчитать точное p-значение при наличии нулей
> # p-значение меньше 0.05, поэтому на 5%-ном уровне значимости гипотеза о равенстве
> # mpg в группах 1 и 2 уверенно отвергается
>
```

```
+   list(mse1, res1, res2)
+ }
>
> all_possible_mse(disp, mpg)
[[1]]
[1] 4.231023 4.741002 6.112540

[[2]]
[1] 4.231023 1.000000

[[3]]
[1] "span = " "0.05"

> # Лучшая по MSE модель с разными span -- модель с параметром span = 0.05
>
```


Span
0.05
0.2
0.5
10
15
20
25
30
100
200
300
400
500
x