Задания

Задача 1:

Работа программы:

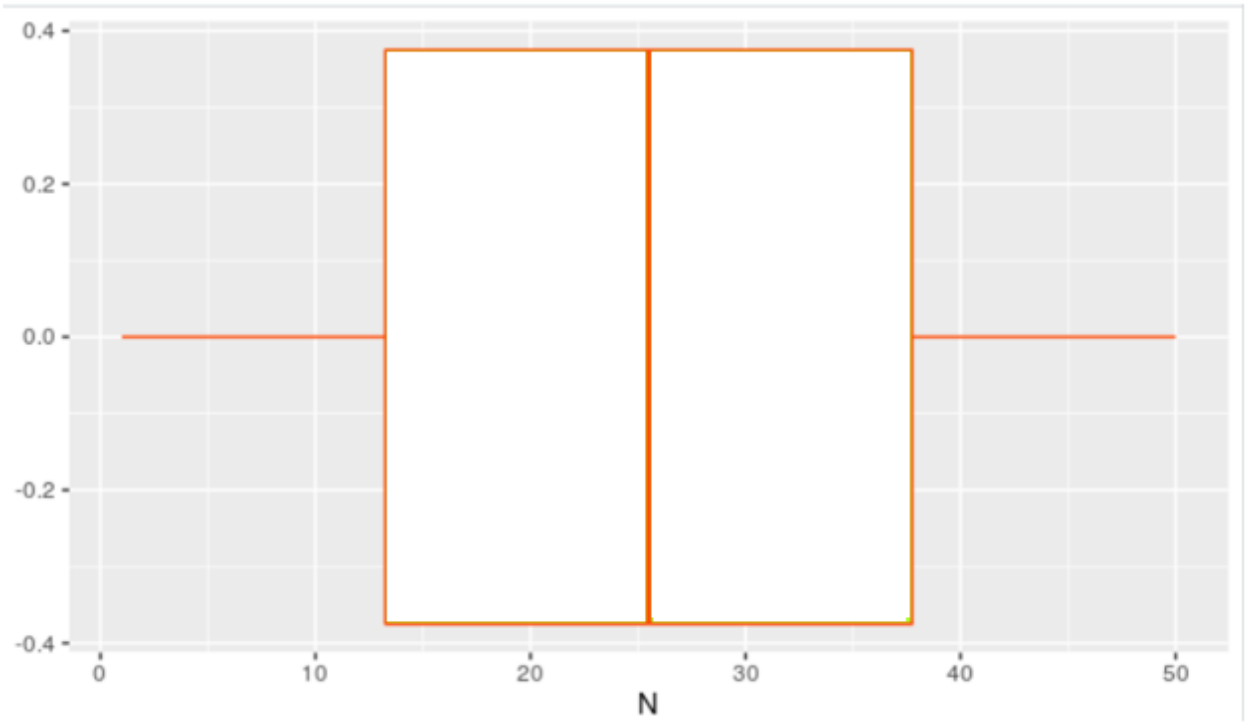
1

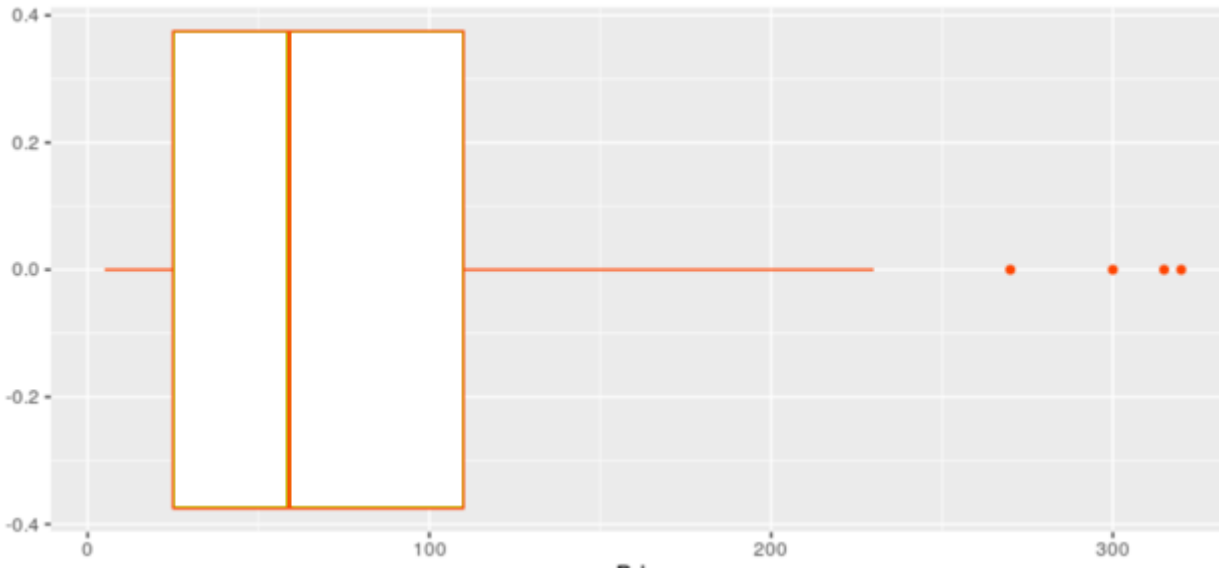
Untitled1* | villa2 | df | .Rhistory

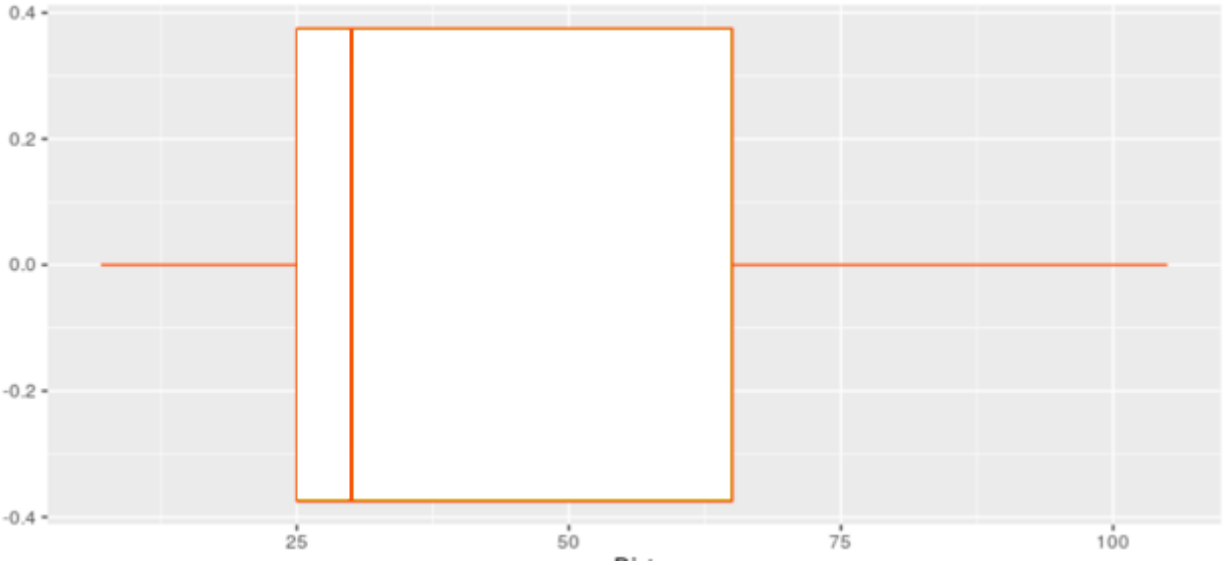
Filter

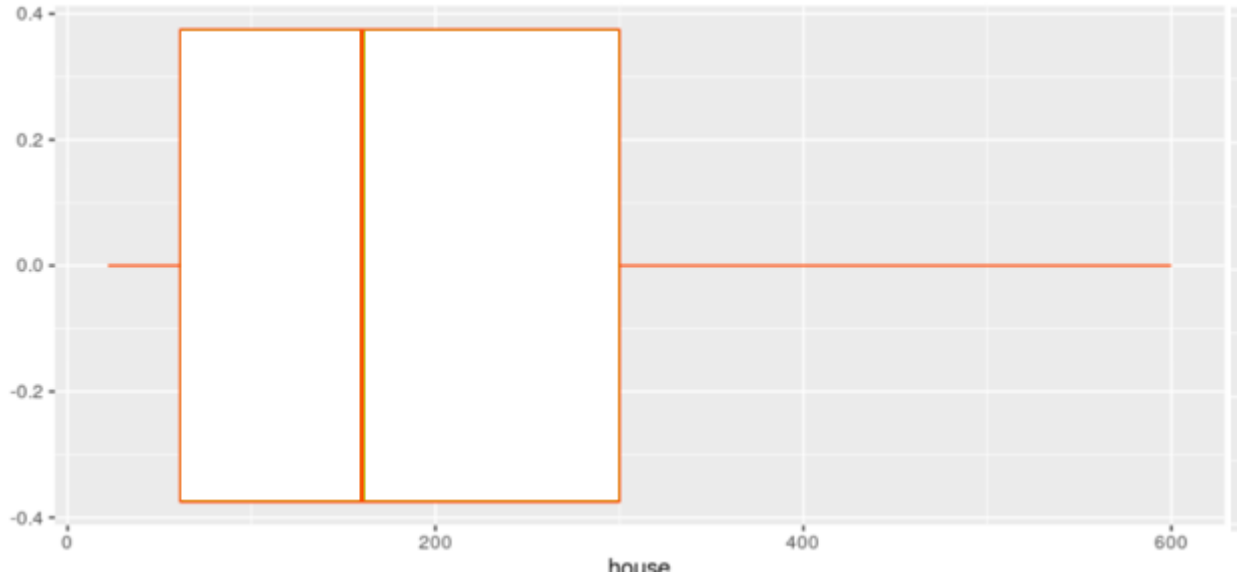
	N	Price	Dist	house	area	Eco
1	1	300	20	400	22	1
2	2	60	18	170	6	0
3	3	14	90	60	11	1
4	4	38	18	65	6	1
5	5	85	25	320	20	0
6	6	85	19	210	20	0
7	7	28	30	60	5	1
8	8	83	45	228	20	0
9	9	80	25	200	20	1
10	10	15	46	36	10	1
11	11	27	86	180	17	0

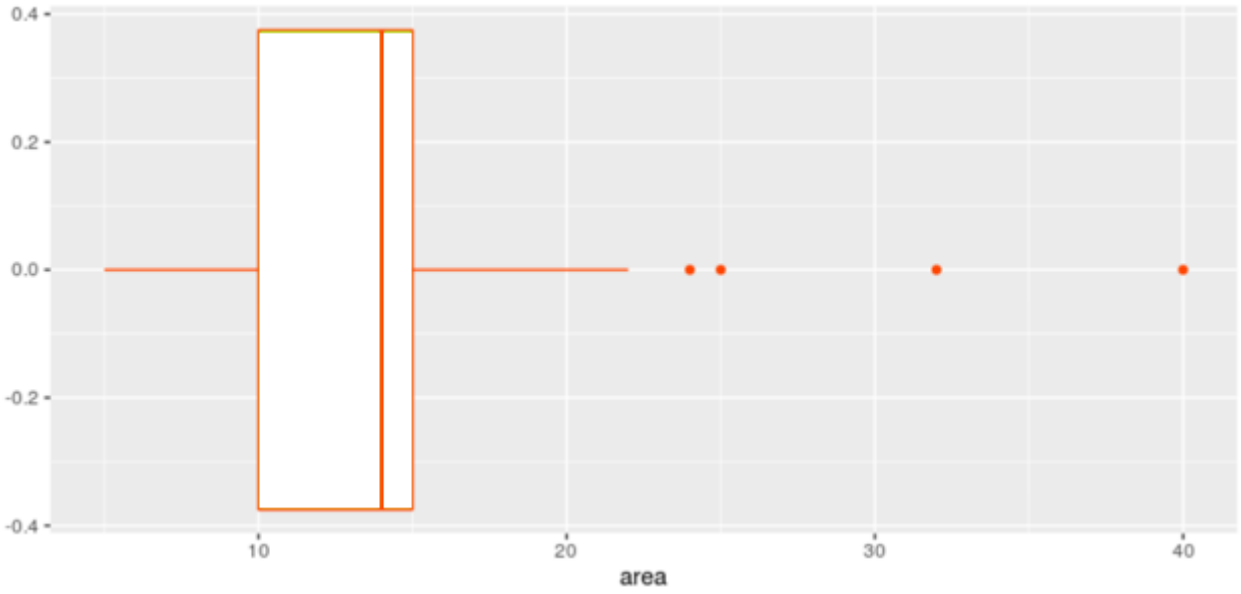
2




0.4
0.2
0.0
-0.2
-0.4
0
100
200
300


0.4
0.2
0.0
-0.2
-0.4
25
50
75
100


0.4
0.2
0.0
-0.2
-0.4
0
200
400
600
house



Значения выбросы присутствуют в переменных Price и Area

3

```
Call:
lm(formula = Price ~ log(house) + log(area) + Eco, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-51.659 -15.497  -1.039  10.310  55.508

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -152.553     27.683  -5.511 3.73e-06 ***
log(house)    29.310      5.779   5.071 1.39e-05 ***
log(area)     22.474     10.862   2.069  0.04621 *
Eco           23.519      8.171   2.878  0.00687 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 25.04 on 34 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.6468,    Adjusted R-squared:  0.6157
F-statistic: 20.76 on 3 and 34 DF,  p-value: 8.025e-08

> AIC(model1)
[1] 358.3836
> BIC(model1)
[1] 366.5715
```

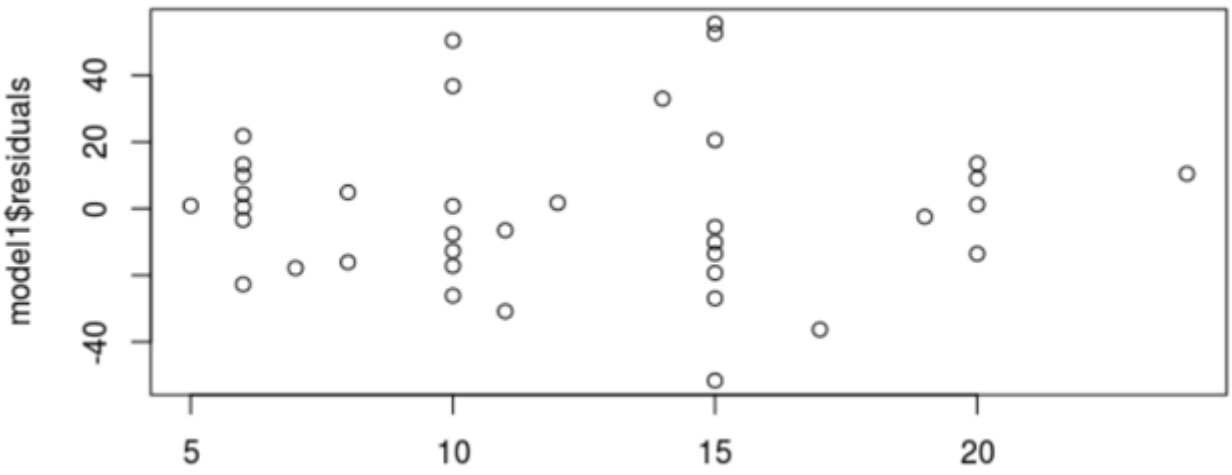
```
Call:
lm(formula = log(Price) ~ log(house) + area + log(Dist) + Eco,
    data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-0.96510 -0.28279  0.00793  0.33007  0.75994

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.94274    0.98275   0.959  0.34439
log(house)   0.66752    0.14061   4.747 3.87e-05 ***
area         0.03514    0.01896   1.853  0.07286 .
log(Dist)   -0.32726    0.14134  -2.315  0.02695 *
Eco          0.46722    0.14883   3.139  0.00356 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4512 on 33 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.7931,    Adjusted R-squared:  0.768
F-statistic: 31.62 on 4 and 33 DF,  p-value: 7.245e-11

> AIC(model2)
[1] 54.00221
> BIC(model2)
[1] 63.82772

Call:
lm(formula = log(Price) ~ log(house) + log(area) + log(Dist) +
    Eco, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-1.01341 -0.23816  0.01685  0.33683  0.72636

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.5130     0.9036   0.568  0.57403
log(house)    0.6516     0.1426   4.568 6.54e-05 ***
log(area)     0.4329     0.2221   1.949  0.05987 .
log(Dist)    -0.3562     0.1461  -2.438  0.02034 *
Eco           0.4584     0.1483   3.091  0.00404 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.449 on 33 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.7951,    Adjusted R-squared:  0.7703
F-statistic: 32.02 on 4 and 33 DF,  p-value: 6.162e-11

> AIC(model3)
[1] 53.62383
> BIC(model3)
[1] 63.44935
```

```
Call:
lm(formula = Price ~ house + area * Dist + Eco, data = dat)

Residuals:
   Min     1Q Median     3Q    Max
-34.95 -13.55  -8.60  10.09  53.03

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.13946   20.62515   0.734  0.46828
house        0.04408    0.03819   1.154  0.25697
area         4.29303    1.55411   2.762  0.00943 **
Dist        -0.47831    0.49751  -0.961  0.34355
Eco         17.89675    8.09572   2.211  0.03433 *
area:Dist   -0.01218    0.03760  -0.324  0.74807
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.96 on 32 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.6959,    Adjusted R-squared:  0.6484
F-statistic: 14.64 on 5 and 32 DF,  p-value: 1.74e-07
```

```
> AIC(model4)
[1] 356.7019
> BIC(model4)
[1] 368.165
```

Выбираем первую модель по критериям Шварца и Акайке, тк она имеет наибольшие значения

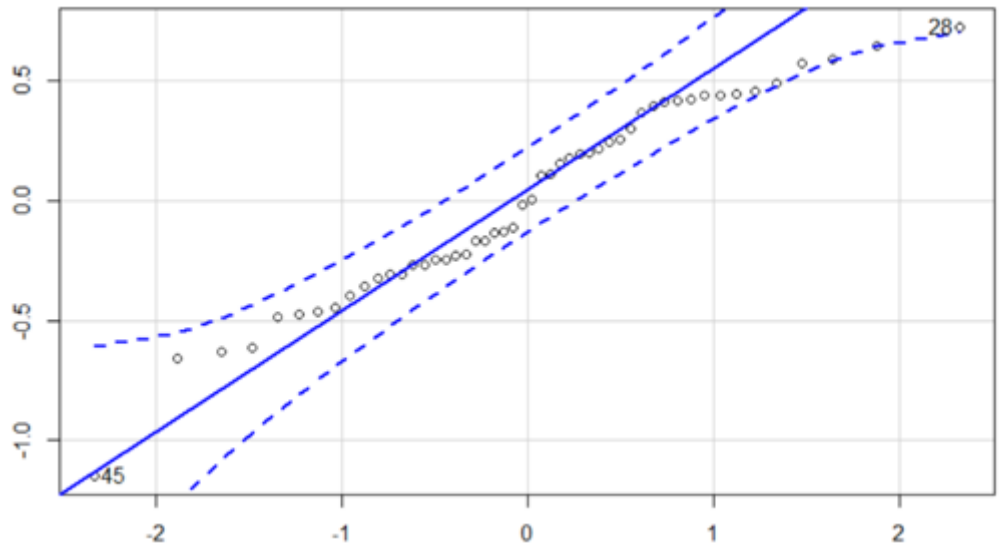
4




```
        studentized Breusch-Pagan test

data:  model1
BP = 7.812, df = 3, p-value = 0.05006
```

```
        Breusch-Pagan test

data:  model1
BP = 8.718, df = 3, p-value = 0.03329
```

p-value < 0.05: гипотезу о гомоскедастичности отвергаем.




```
lag Autocorrelation D-W Statistic p-value
  1      0.05633208       1.861936   0.614
Alternative hypothesis: rho != 0
```

p-value > 0.05 = автокорреляция присутствует, остаточные значения имеют нормальное распределение

Листинг:

```
df <- villa[, 1:6]
View(df)
#---------------------------------
ggplot(df, aes(x = N)) + geom_boxplot(binwidth=1, colour = "orangered", fill="
white")
ggplot(df, aes(x = Price)) + geom_boxplot(binwidth=1, colour = "orangered", fill="
white")
ggplot(df, aes(x = Dist)) + geom_boxplot(binwidth=1, colour = "orangered", fill="
white")
ggplot(df, aes(x = house)) + geom_boxplot(binwidth=1, colour = "orangered", fill="
white")
ggplot(df, aes(x = area)) + geom_boxplot(binwidth=1, colour = "orangered", fill="
white")
#---------------------------------
boxplot.stats(df$Price)$out
ind <- which(df$Price %in% boxplot.stats(dat$Price)$out)
ind
df <- df[-ind, ]

ggplot(df, aes(x = Price)) + geom_boxplot(binwidth=1, colour = "orangered", fill="
white")
#---------------------------------
boxplot(df$area)
boxplot.stats(df$area)$out
ind <- which(df$area %in% boxplot.stats(dat$area)$out)
ind
df <- df[-ind, ]

ggplot(df, aes(x = area)) + geom_boxplot(binwidth=1, colour = "orangered",
fill="white")
#---------------------------------
```

```
summary(lm(formula=Price~log(house)+log(area)+Eco, data= df))
AIC(lm(formula=Price~log(house)+log(area)+Eco, data= df))
BIC(lm(formula=Price~log(house)+log(area)+Eco, data= df))
#---------------------------------
summary(lm(formula=log(Price)~log(house)+area+log(Dist)+Eco, data= df))
AIC(lm(formula=log(Price)~log(house)+area+log(Dist)+Eco, data= df))
BIC(lm(formula=log(Price)~log(house)+area+log(Dist)+Eco, data= df))
#---------------------------------
summary(lm(formula=log(Price)~log(house)+log(area)+log(Dist)+Eco, data= df))
AIC(lm(formula=log(Price)~log(house)+log(area)+log(Dist)+Eco, data= df))
BIC(lm(formula=log(Price)~log(house)+log(area)+log(Dist)+Eco, data= df))
#---------------------------------
summary(lm(formula=Price~house+area*Dist+Eco, data= df))
AIC(lm(formula=Price~house+area*Dist+Eco, data= df))
BIC(lm(formula=Price~house+area*Dist+Eco, data= df))
#---------------------------------
plot(dat$area, model1$residuals)
bptest(model1, varformula = NULL, studentize = TRUE, df)
bptest(model1, varformula = NULL, studentize = FALSE, df)
qqPlot(residuals(model1), xlab="Квантили нормального распределения",
ylab="Наблюдаемые квантили")
dwt(model1)
```