Задания

Задача 1:

Работа программы:

1

	N	Price	Dist	house	area	Eco
1	1	300.0	20.0	400	22.0	1
2	2	60.0	18.0	170	6.0	0
3	3	14.0	90.0	60	11.0	1
4	4	38.0	18.0	65	6.0	1
5	5	85.0	25.0	320	20.0	0
6	6	85.0	19.0	210	20.0	0
7	7	28.0	30.0	60	5.0	1
8	8	83.0	45.0	228	20.0	0

2

3
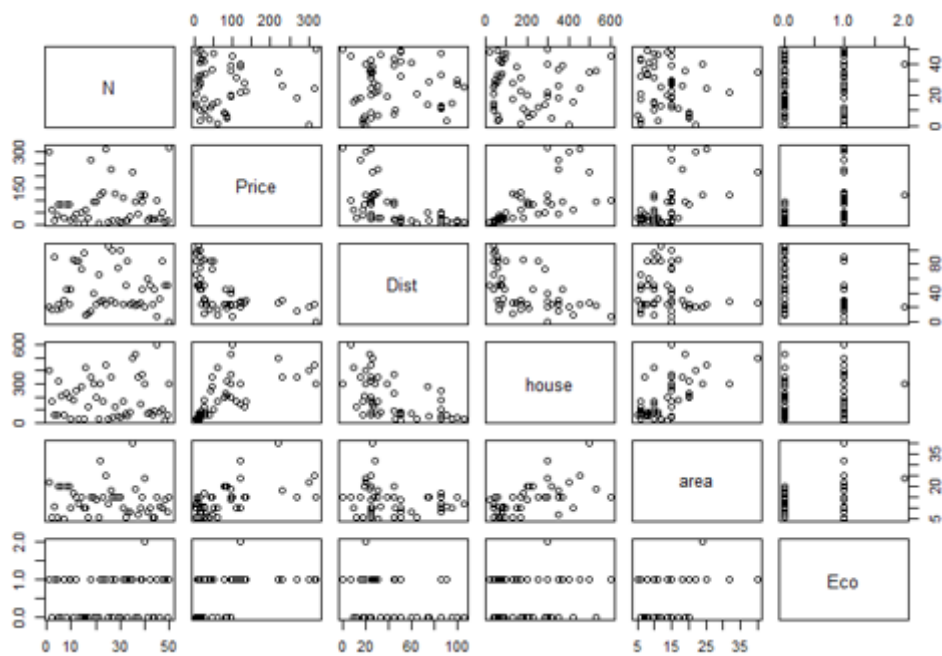
Name	Type	Value
lw	list [12] (S3: lm)	List of length 12
coefficients	double [3]	-21.318 0.285 3.263
residuals	double [50]	135.7 13.4 -17.6 21.2 -50.0 -18.7 ...
effects	double [50]	-553.31 397.77 -125.91 1.91 -65.66 -30.84 ...
rank	integer [1]	3
fitted.values	double [50]	164.3 46.6 31.6 16.8 135.0 103.7 ...
assign	integer [3]	0 1 2
qr	list [5] (S3: qr)	List of length 5
df.residual	integer [1]	47
xlevels	list [0]	List of length 0
call	language	lm(formula = df$Price ~ df$house + df$area, data = df)
terms	formula	df$Price ~ df$house + df$area
model	list [50 x 3] (S3: data.frame)	A data.frame with 50 rows and 3 columns

```
Call:
lm(formula = df$Price ~ df$house + df$area, data = df)

Residuals:
   Min    1Q Median    3Q    Max
-98.36 -26.75 -12.94  12.58 207.01

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -21.31804   19.35152  -1.102 0.276236
df$house      0.28457    0.07189   3.958 0.000254 ***
df$area       3.26273    1.57901   2.066 0.044334 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 60.93 on 47 degrees of freedom
Multiple R-squared:  0.4994,   Adjusted R-squared:  0.4781
F-statistic: 23.44 on 2 and 47 DF,  p-value: 8.678e-08

[1] 0.4993823
[1] 557.7775
[1] 565.4256
```

Характеристики модели как t-статистики и коэффициент детерминации R2 и скорректированный коэффициент детерминации: Multiple R-squared: 0.4994, Adjusted R-squared: 0.4781

4

```
df$house  df$area
1.572809 1.572809
```

Модель не мультиколлинеарна. Vif<5

5

```
model<-lm(formula=log(df$Price)~log(df$house)+log(df$area)+log(df$Dist)+df$Eco)
summary(model)
Call:
lm(formula = log(df$Price) ~ log(df$house) + log(df$area) + log(df$Dist) +
    df$Eco)

Residuals:
     Min      1Q   Median      3Q     Max
-1.11093 -0.30496 -0.03132  0.36833  0.77331

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.25746    0.62597  -0.411 0.682808
log(df$house)  0.81753    0.10425   7.842 5.85e-10
log(df$area)   0.30355    0.17021   1.783 0.081278
log(df$Dist)  -0.27524    0.09005  -3.057 0.003759
df$Eco         0.45746    0.12509   3.657 0.000666

(Intercept)
log(df$house) ***
log(df$area)  .
log(df$Dist)  **
df$Eco        ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4431 on 45 degrees of freedom
Multiple R-squared:  0.8624,   Adjusted R-squared:  0.8502
F-statistic: 70.52 on 4 and 45 DF,  p-value: < 2.2e-16

    report<-summary(model)
```

```
report$r.squared
 [1] 0.8624272

AIC(model)
 [1] 67.23872
BIC(model)
 [1] 78.71085
vif(model)
log(df$house)    log(df$area)    log(df$Dist)         df$Eco
    2.181372        1.657737        1.629860        1.153851
mean(vif(model))
 [1] 1.655705
```

Выбираем эту модель Multiple R-squared: 0.6245,Adjusted R-squared: 0.5818, так как наименьшее значение по Шварцу из предложенных 4 моделей = 78.71085

Листинг:

```
library(readxl)
library(car)
setwd("C:/Users/Админ/Desktop/Компьютинг")
df<-read_excel("Villa.xlsx")
View(df)
plot(df)
lw<-lm(formula=df$Price~df$house+df$area,df)
report<-summary(lw)
report
report$r.squared
AIC(lw)
BIC(lw)
vif(lw)
model<-lm(formula=log(df$Price)~log(df$house)+log(df$area)+log(df$Dist)+df$Eco)
summary(model)
report<-summary(model)
report$r.squared
AIC(model)
BIC(model)
vif(model)
mean(vif(model))
```