Разбор задачи 3.33 (Катышев, Магнус - Сборник задач по начальному курсу эконометрики

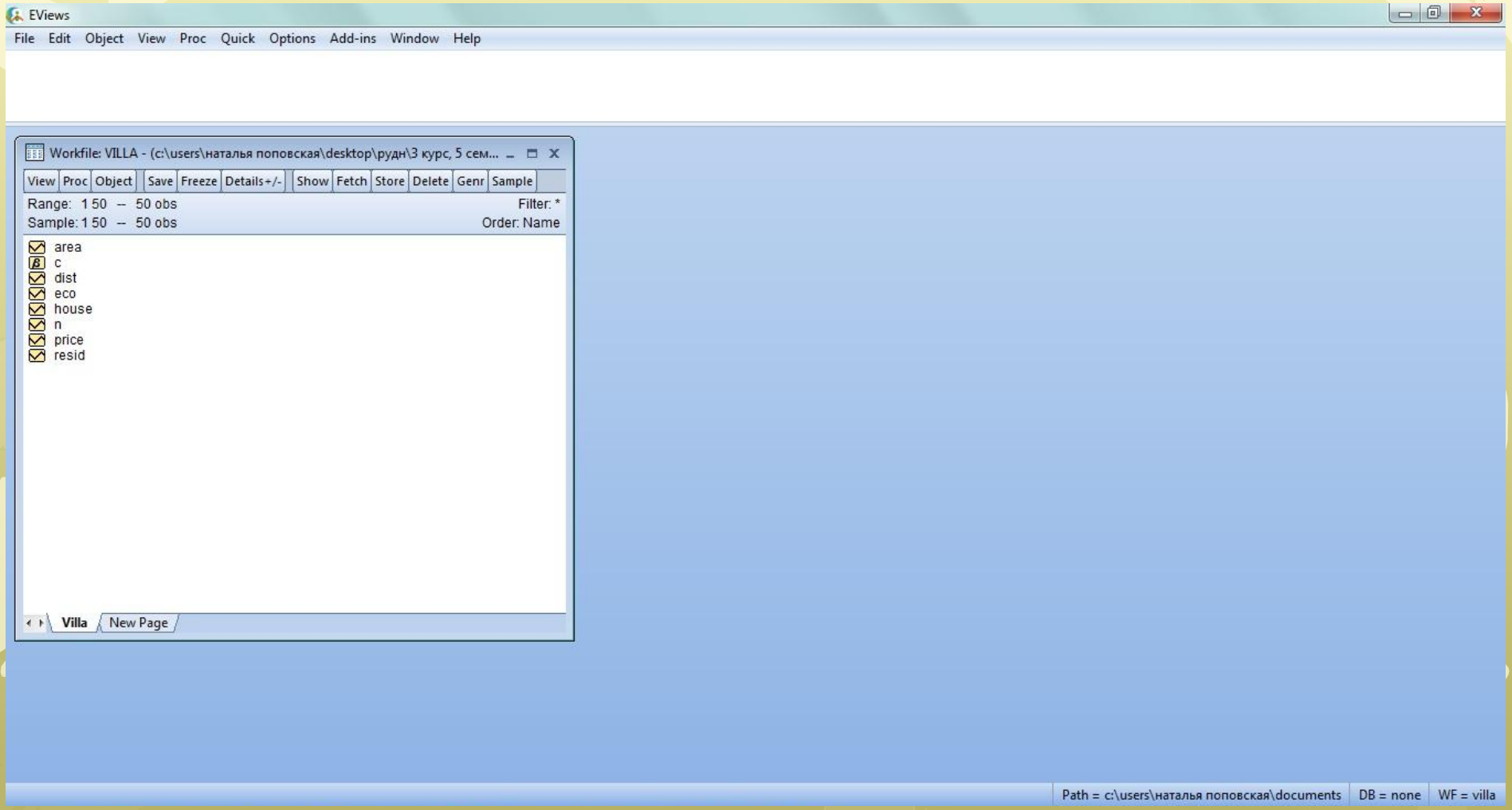
Подготовила презентацию Поповская Наталья, НБ-401

Формулировка задачи 3.33

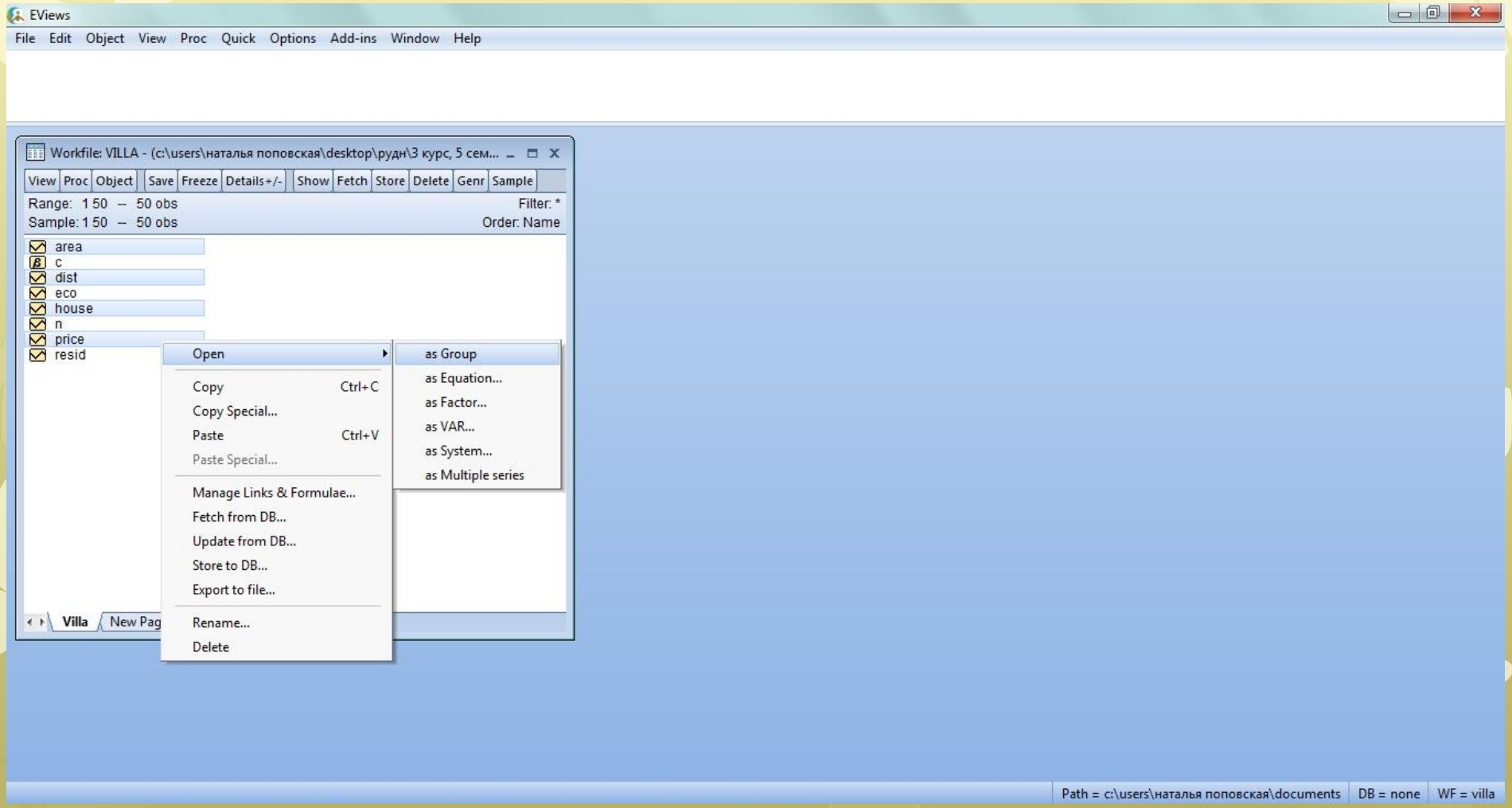
- Рассматривается информация о стоимости коттеджей в Московской области по Киевскому направлению (по данным строительной компании «Стройсервис», осень 1997 г.)
- Данные находятся в файле villa.xls. Переменные описаны в таблице 3.28.
- Подберите функциональную форму зависимости цены коттеджа от его параметров, учитывая такие факторы, как t-статистика и коэффициент детерминации R^2

Таблица 3.28	
Переменная	Описание
n	номер по порядку
price	цена в тыс. долл.
dist	расстояние от кольцевой автодороги в км
house	площадь дома в кв. м
area	площадь участка в сотках

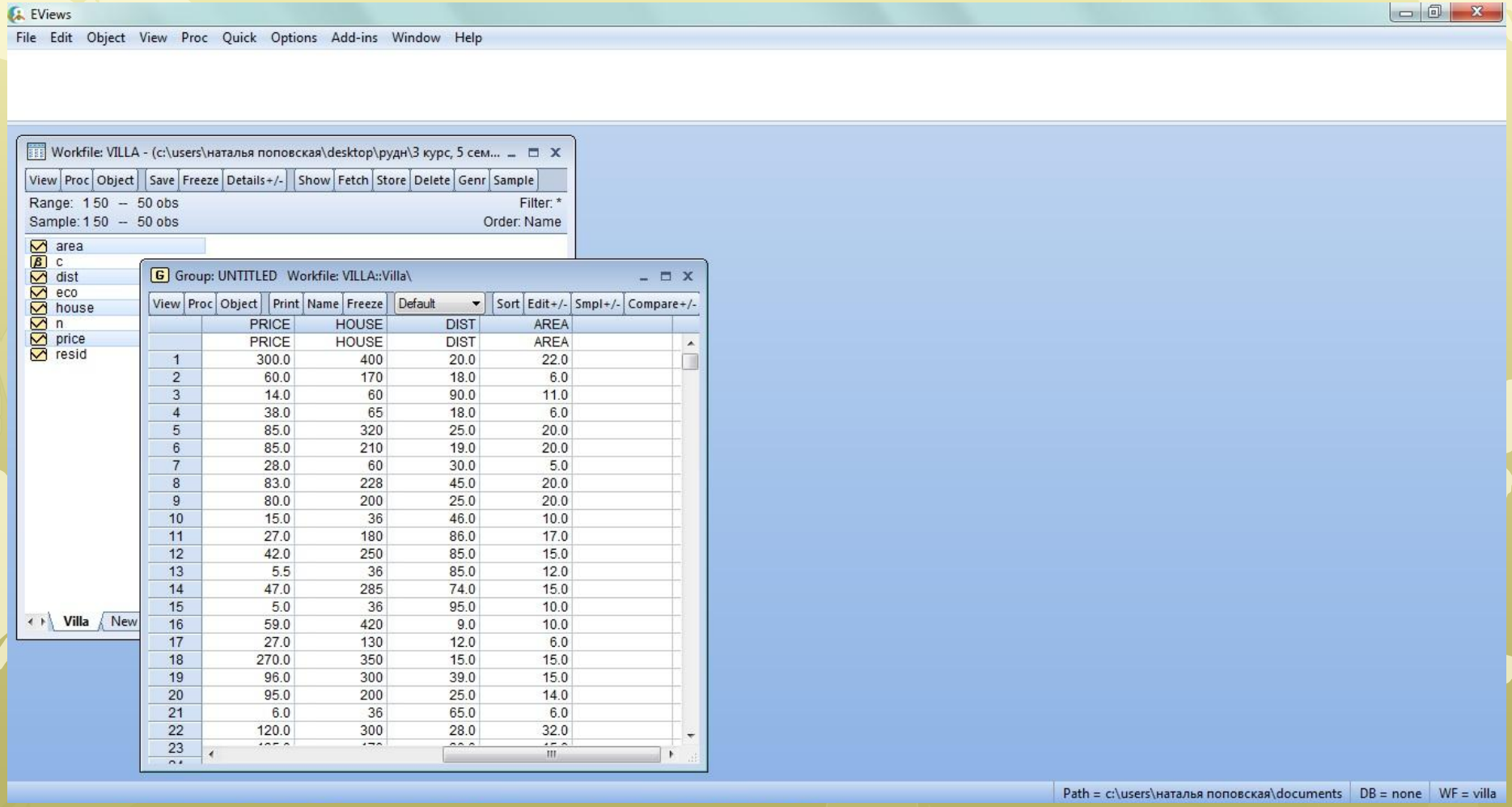
Открытие файла villa.wf1 в Eviews



Построение описательной статистики [1]

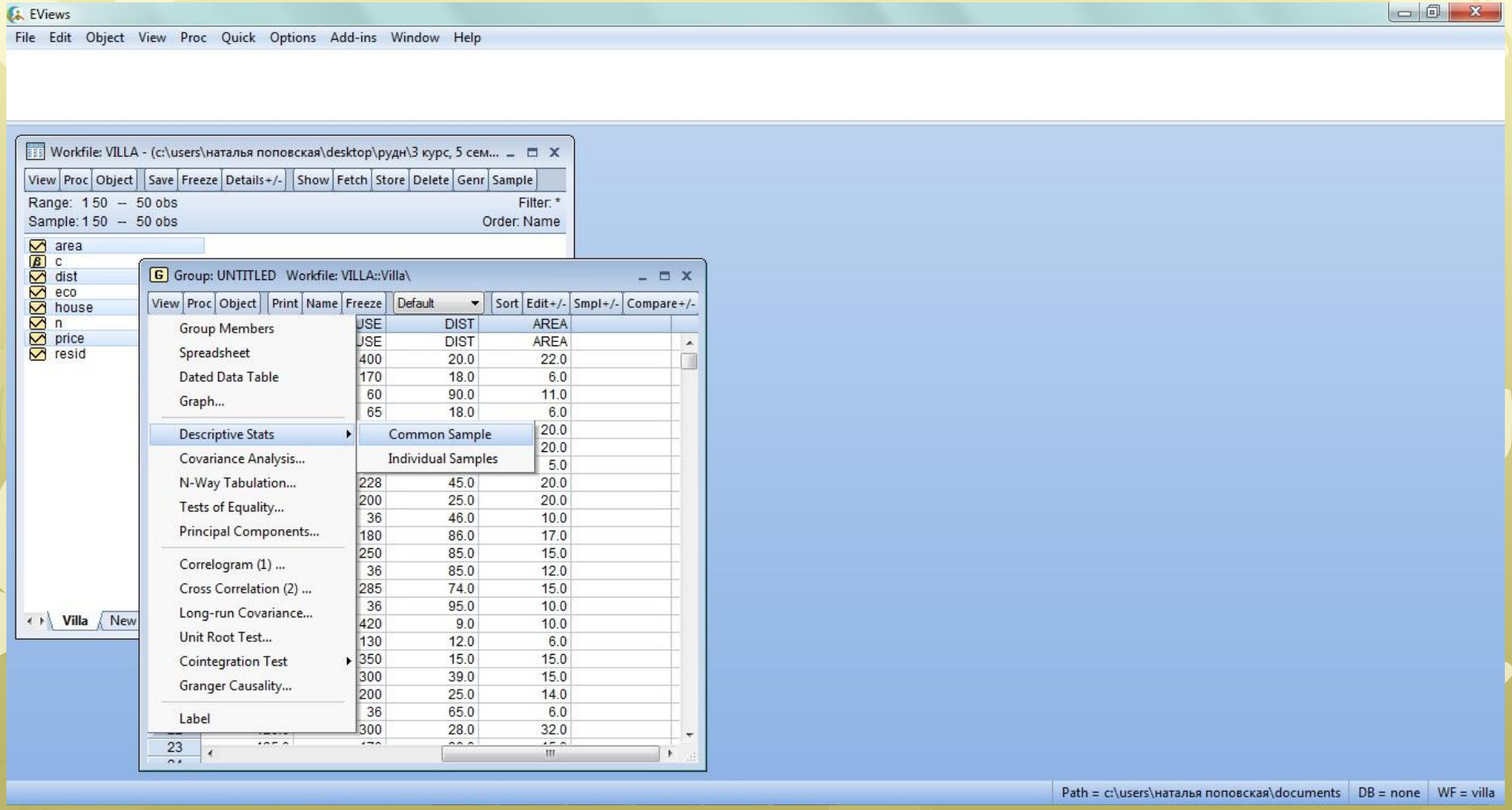


Построение описательной статистики [2]

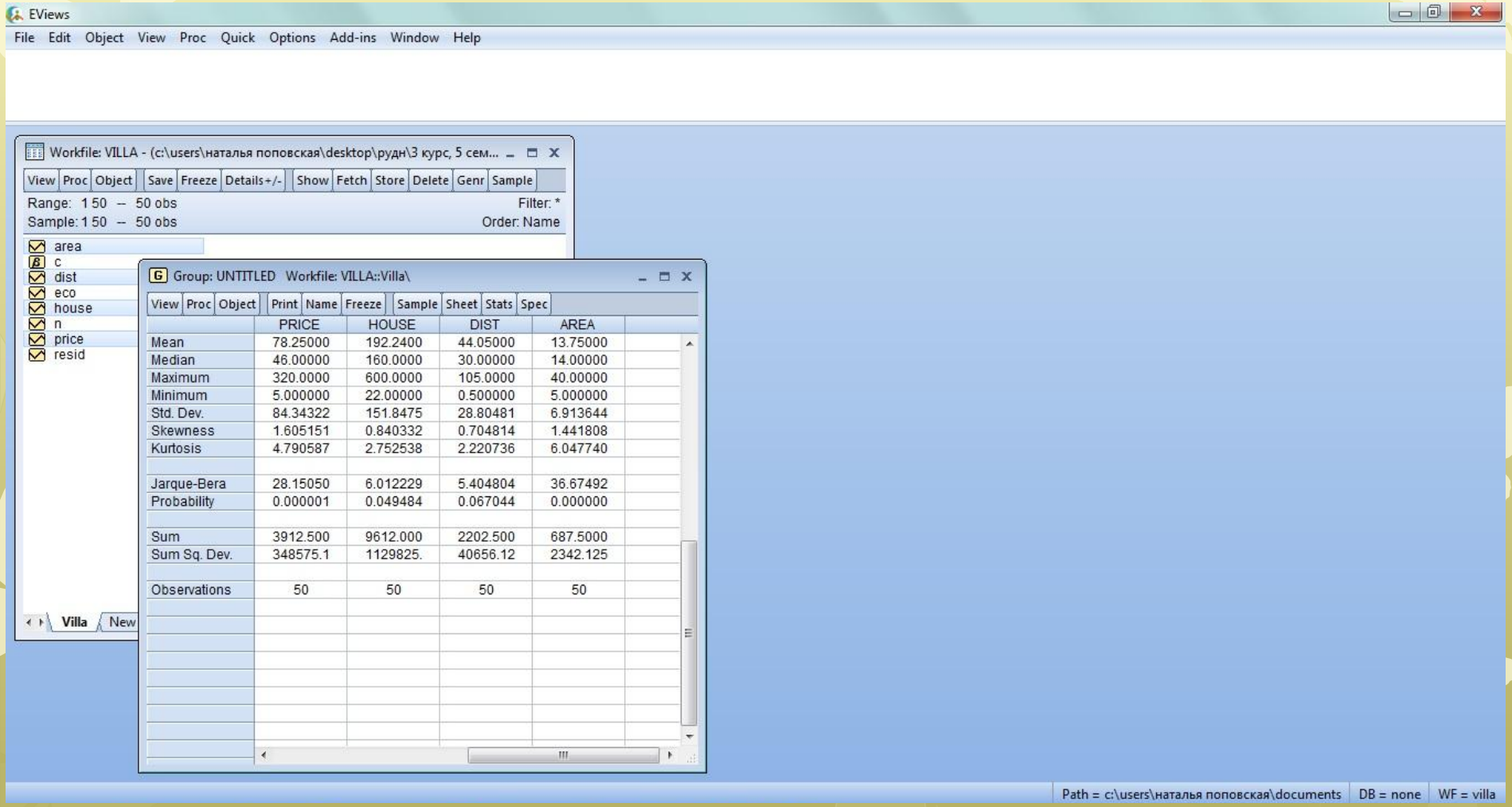


	PRICE	HOUSE	DIST	AREA
	PRICE	HOUSE	DIST	AREA
1	300.0	400	20.0	22.0
2	60.0	170	18.0	6.0
3	14.0	60	90.0	11.0
4	38.0	65	18.0	6.0
5	85.0	320	25.0	20.0
6	85.0	210	19.0	20.0
7	28.0	60	30.0	5.0
8	83.0	228	45.0	20.0
9	80.0	200	25.0	20.0
10	15.0	36	46.0	10.0
11	27.0	180	86.0	17.0
12	42.0	250	85.0	15.0
13	5.5	36	85.0	12.0
14	47.0	285	74.0	15.0
15	5.0	36	95.0	10.0
16	59.0	420	9.0	10.0
17	27.0	130	12.0	6.0
18	270.0	350	15.0	15.0
19	96.0	300	39.0	15.0
20	95.0	200	25.0	14.0
21	6.0	36	65.0	6.0
22	120.0	300	28.0	32.0
23				

Построение описательной статистики [3]

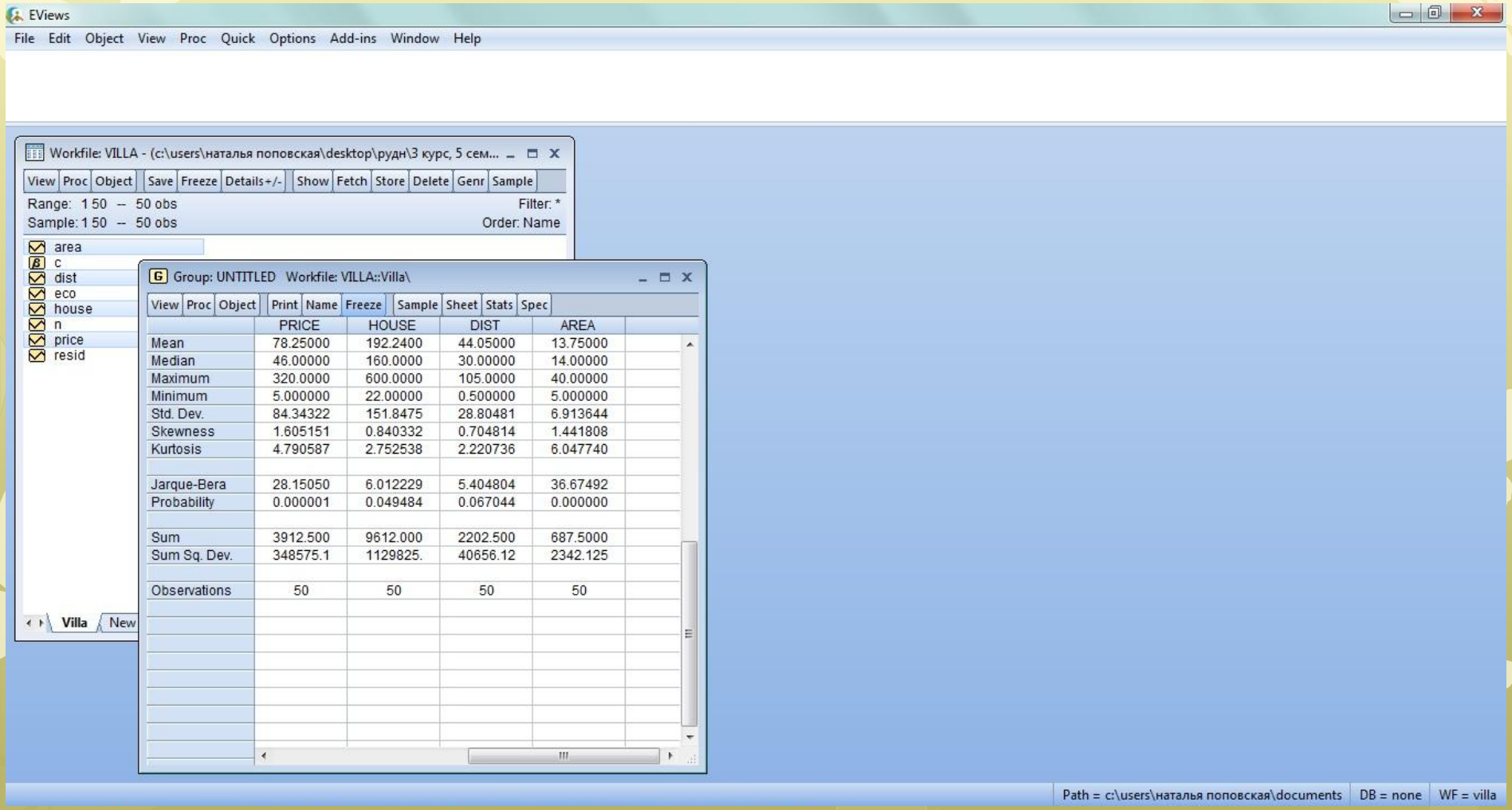


Построение описательной статистики [4]



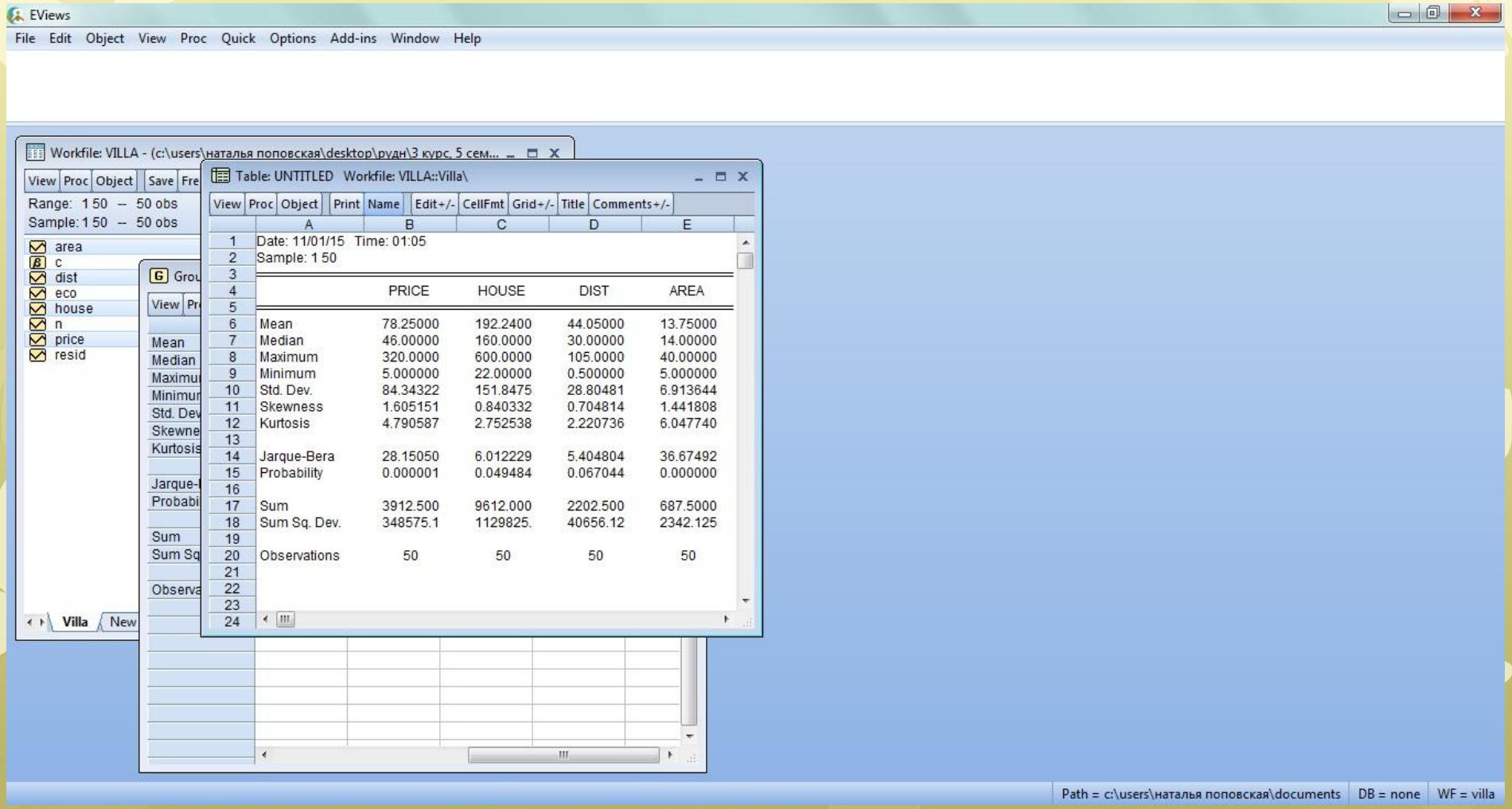
	PRICE	HOUSE	DIST	AREA
Mean	78.25000	192.2400	44.05000	13.75000
Median	46.00000	160.0000	30.00000	14.00000
Maximum	320.0000	600.0000	105.0000	40.00000
Minimum	5.000000	22.00000	0.500000	5.000000
Std. Dev.	84.34322	151.8475	28.80481	6.913644
Skewness	1.605151	0.840332	0.704814	1.441808
Kurtosis	4.790587	2.752538	2.220736	6.047740
Jarque-Bera	28.15050	6.012229	5.404804	36.67492
Probability	0.000001	0.049484	0.067044	0.000000
Sum	3912.500	9612.000	2202.500	687.5000
Sum Sq. Dev.	348575.1	1129825.	40656.12	2342.125
Observations	50	50	50	50

Сохранение через Freeze->Name [1]

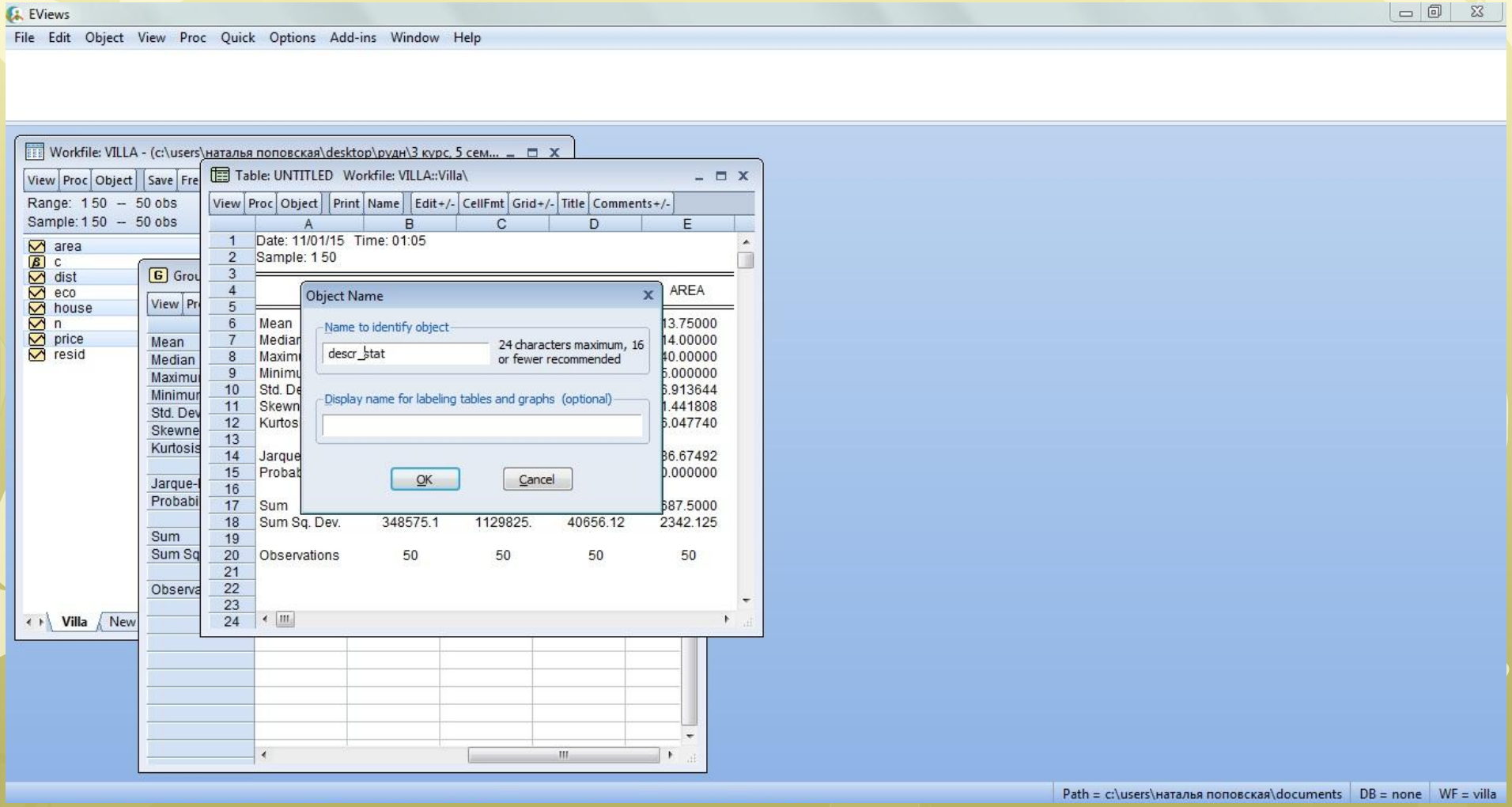


	PRICE	HOUSE	DIST	AREA
Mean	78.25000	192.2400	44.05000	13.75000
Median	46.00000	160.0000	30.00000	14.00000
Maximum	320.0000	600.0000	105.0000	40.00000
Minimum	5.000000	22.00000	0.500000	5.000000
Std. Dev.	84.34322	151.8475	28.80481	6.913644
Skewness	1.605151	0.840332	0.704814	1.441808
Kurtosis	4.790587	2.752538	2.220736	6.047740
Jarque-Bera	28.15050	6.012229	5.404804	36.67492
Probability	0.000001	0.049484	0.067044	0.000000
Sum	3912.500	9612.000	2202.500	687.5000
Sum Sq. Dev.	348575.1	1129825.	40656.12	2342.125
Observations	50	50	50	50

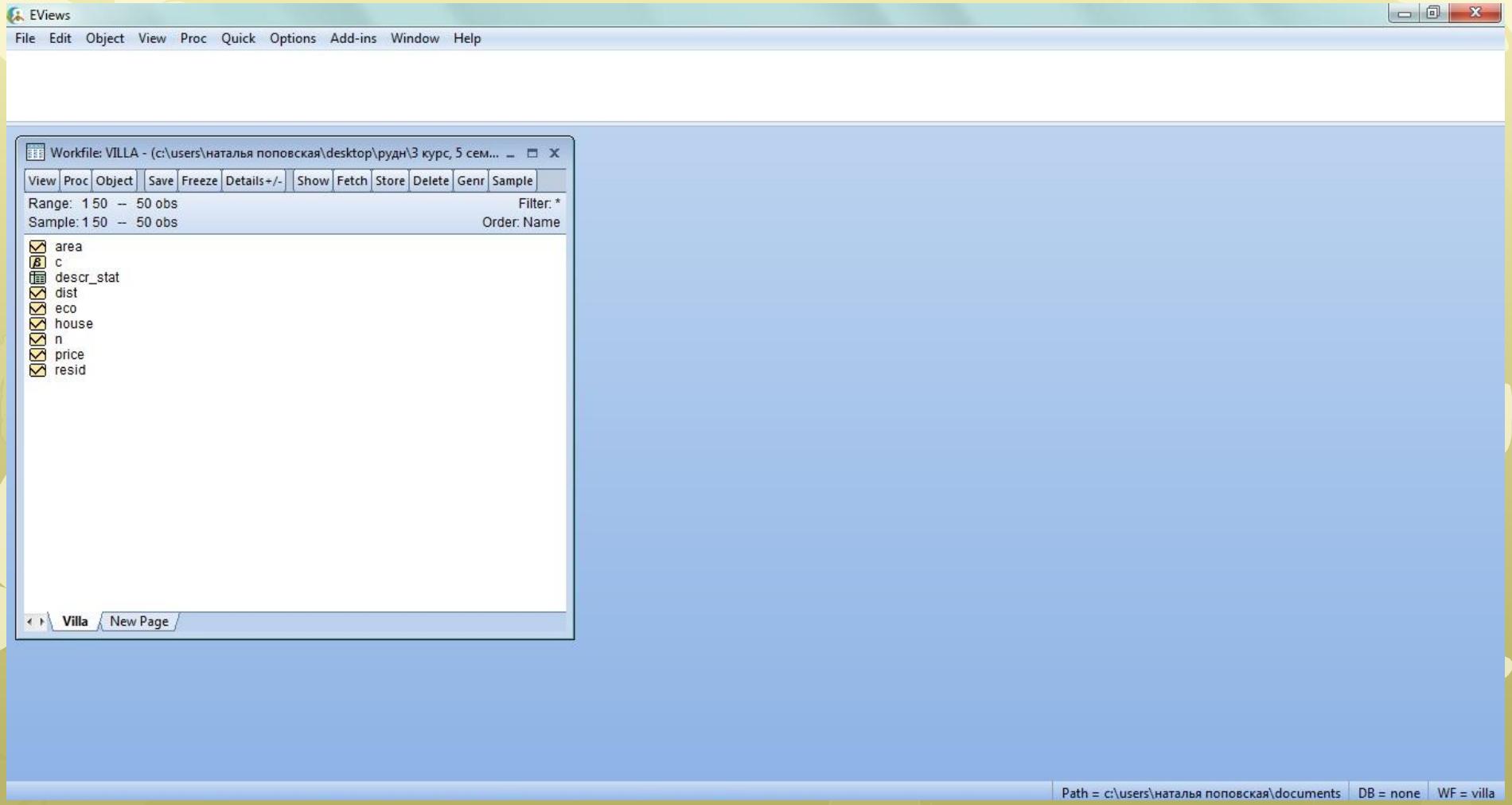
Сохранение через Freeze->Name [2]



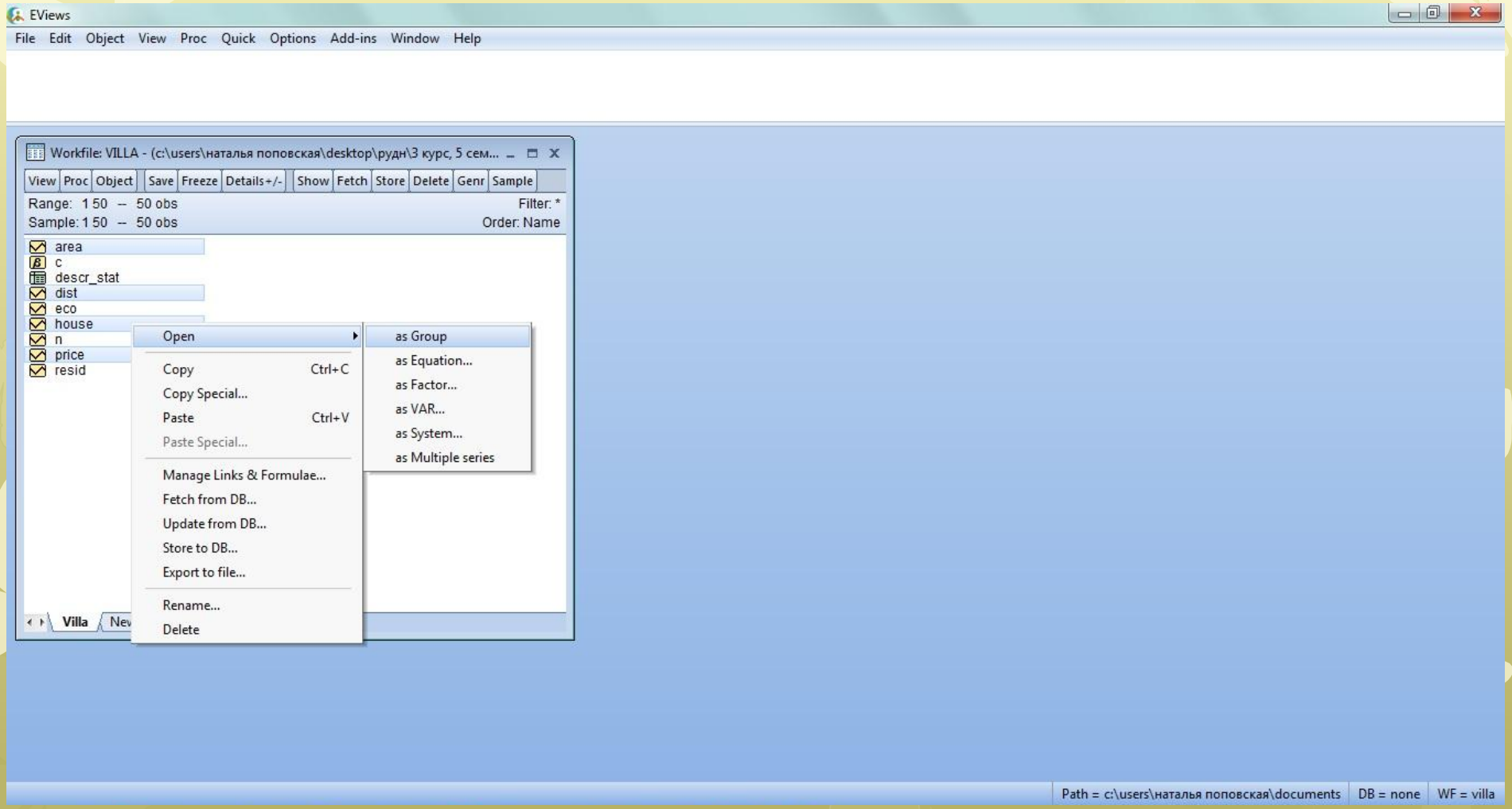
Сохранение через Freeze->Name [3]



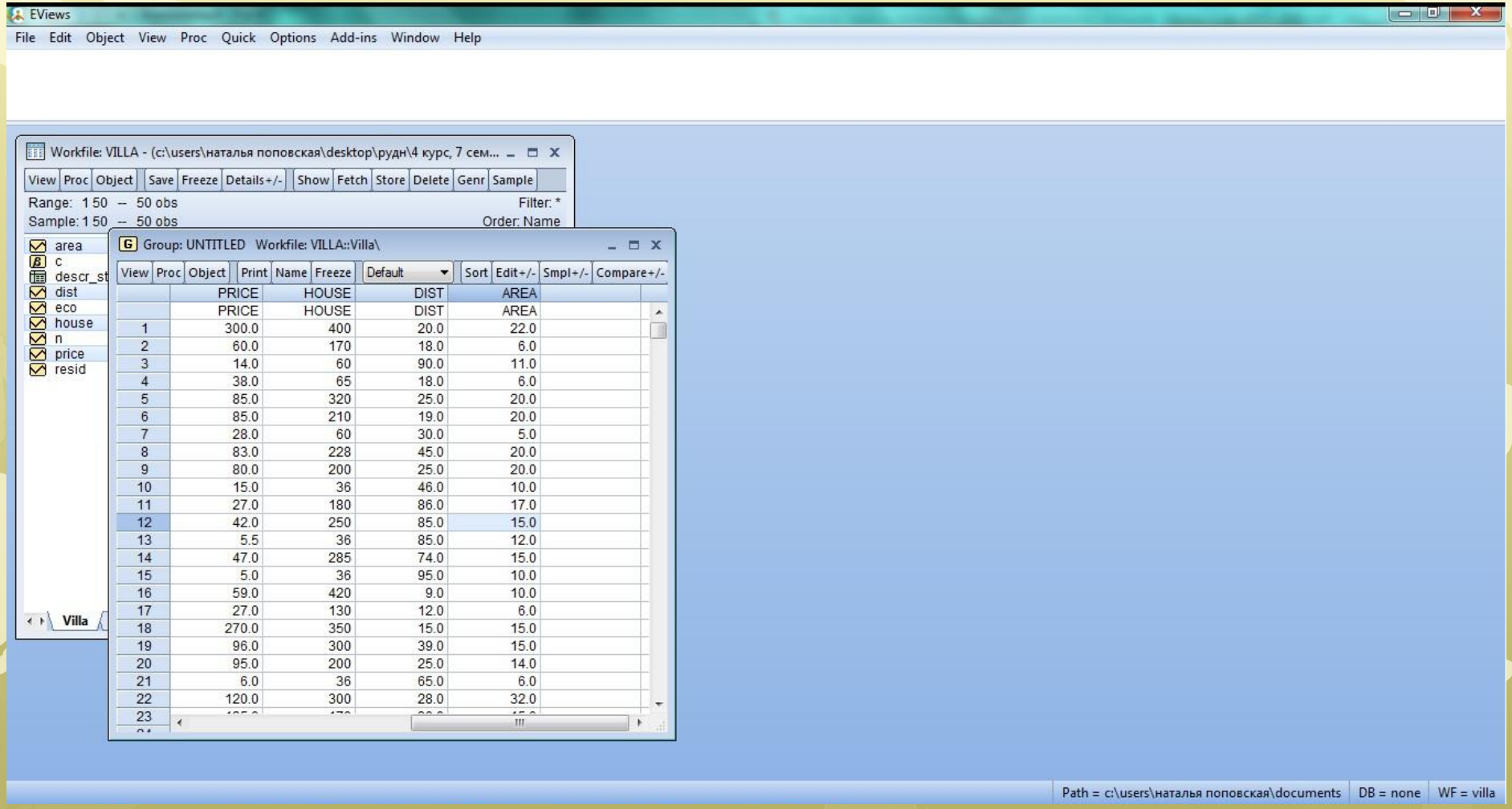
Сохранение через Freeze->Name [4]



Построение корреляционной матрицы [1]

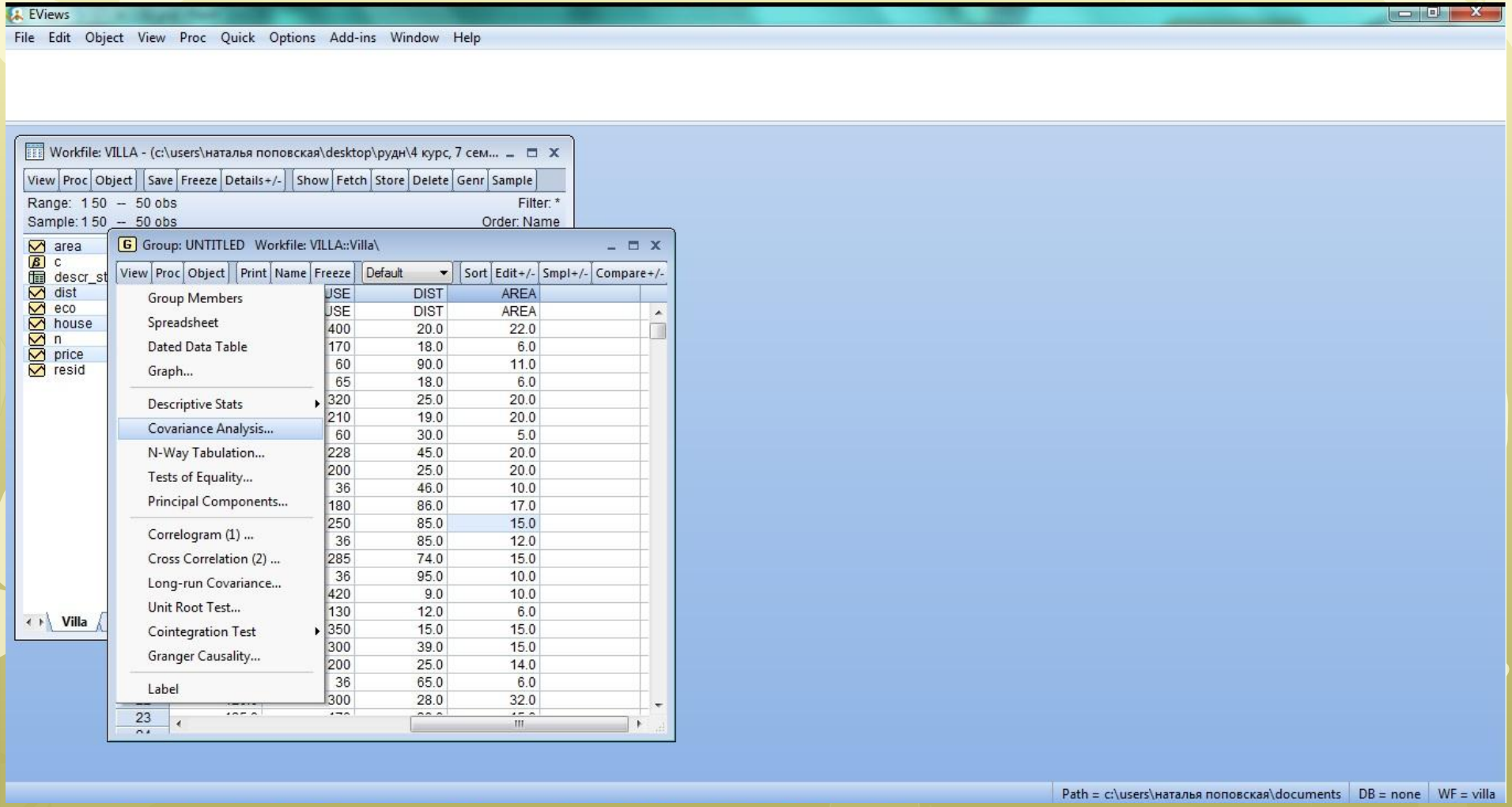


Построение корреляционной матрицы [2]

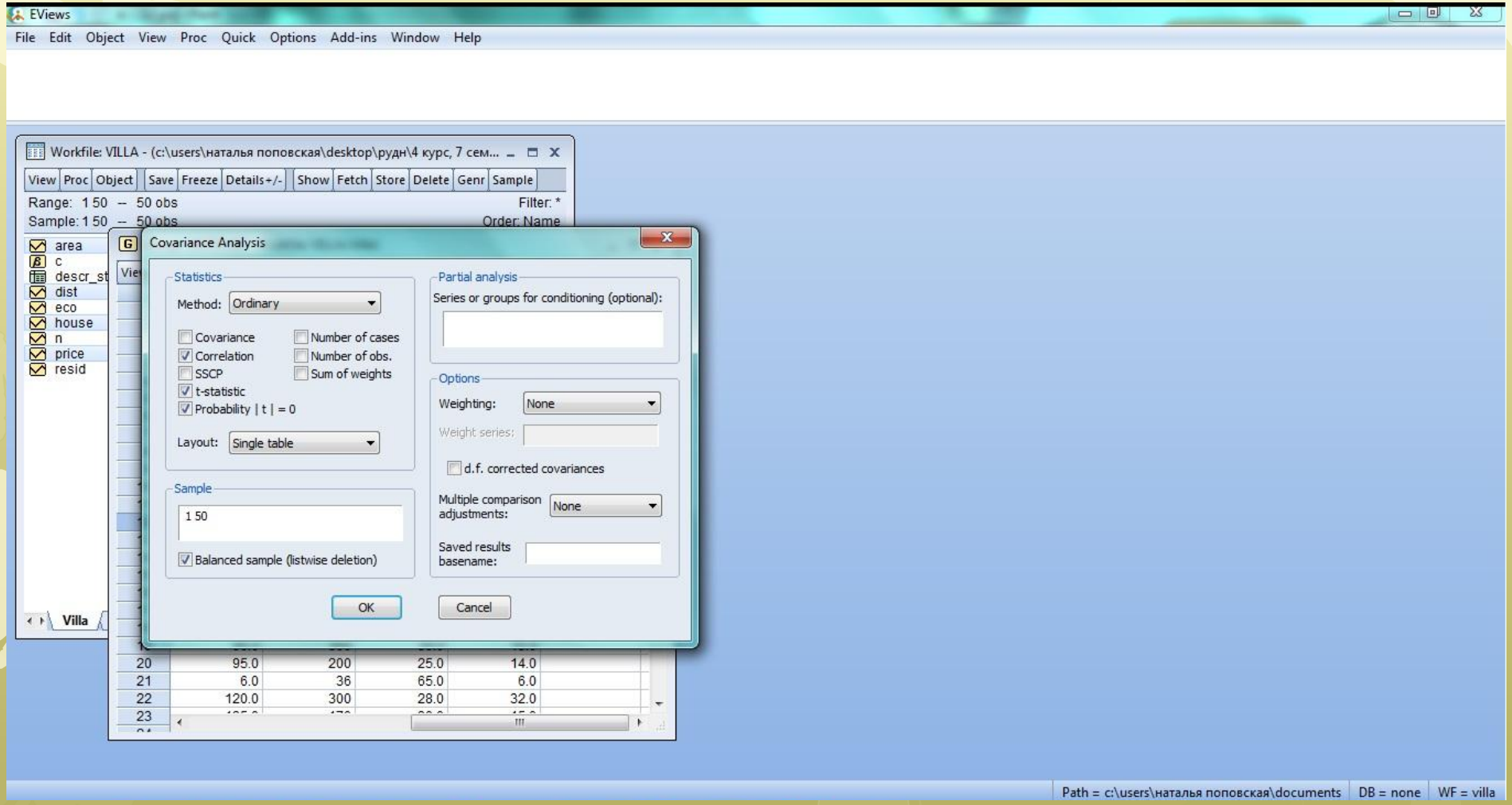


	PRICE	HOUSE	DIST	AREA
	PRICE	HOUSE	DIST	AREA
1	300.0	400	20.0	22.0
2	60.0	170	18.0	6.0
3	14.0	60	90.0	11.0
4	38.0	65	18.0	6.0
5	85.0	320	25.0	20.0
6	85.0	210	19.0	20.0
7	28.0	60	30.0	5.0
8	83.0	228	45.0	20.0
9	80.0	200	25.0	20.0
10	15.0	36	46.0	10.0
11	27.0	180	86.0	17.0
12	42.0	250	85.0	15.0
13	5.5	36	85.0	12.0
14	47.0	285	74.0	15.0
15	5.0	36	95.0	10.0
16	59.0	420	9.0	10.0
17	27.0	130	12.0	6.0
18	270.0	350	15.0	15.0
19	96.0	300	39.0	15.0
20	95.0	200	25.0	14.0
21	6.0	36	65.0	6.0
22	120.0	300	28.0	32.0
23				

Построение корреляционной матрицы [3]



Построение корреляционной матрицы [4]



Построение корреляционной матрицы [5]

EViews

File Edit Object View Proc Quick Options Add-ins Window Help

Workfile:

View Proc Ob

Range: 1 50

Sample: 1 50

area
c
descr_st
dist
eco
house
n
price
resid

Villa

Group: UNTITLED Workfile: VILLA::Villa\

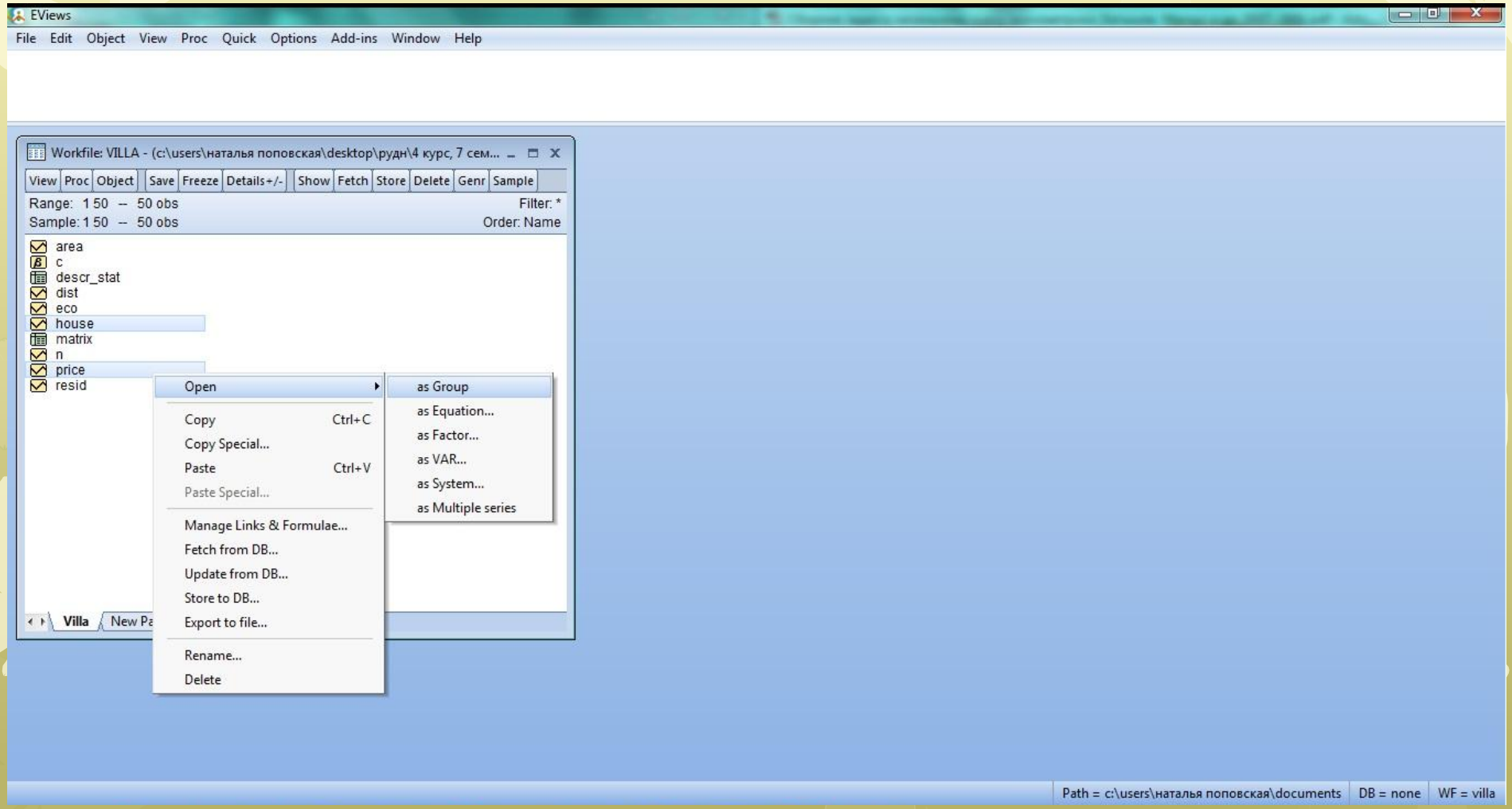
View Proc Object Print Name Freeze Sample Sheet Stats Spec

Covariance Analysis: Ordinary
Date: 11/01/15 Time: 20:22
Sample: 1 50
Included observations: 50

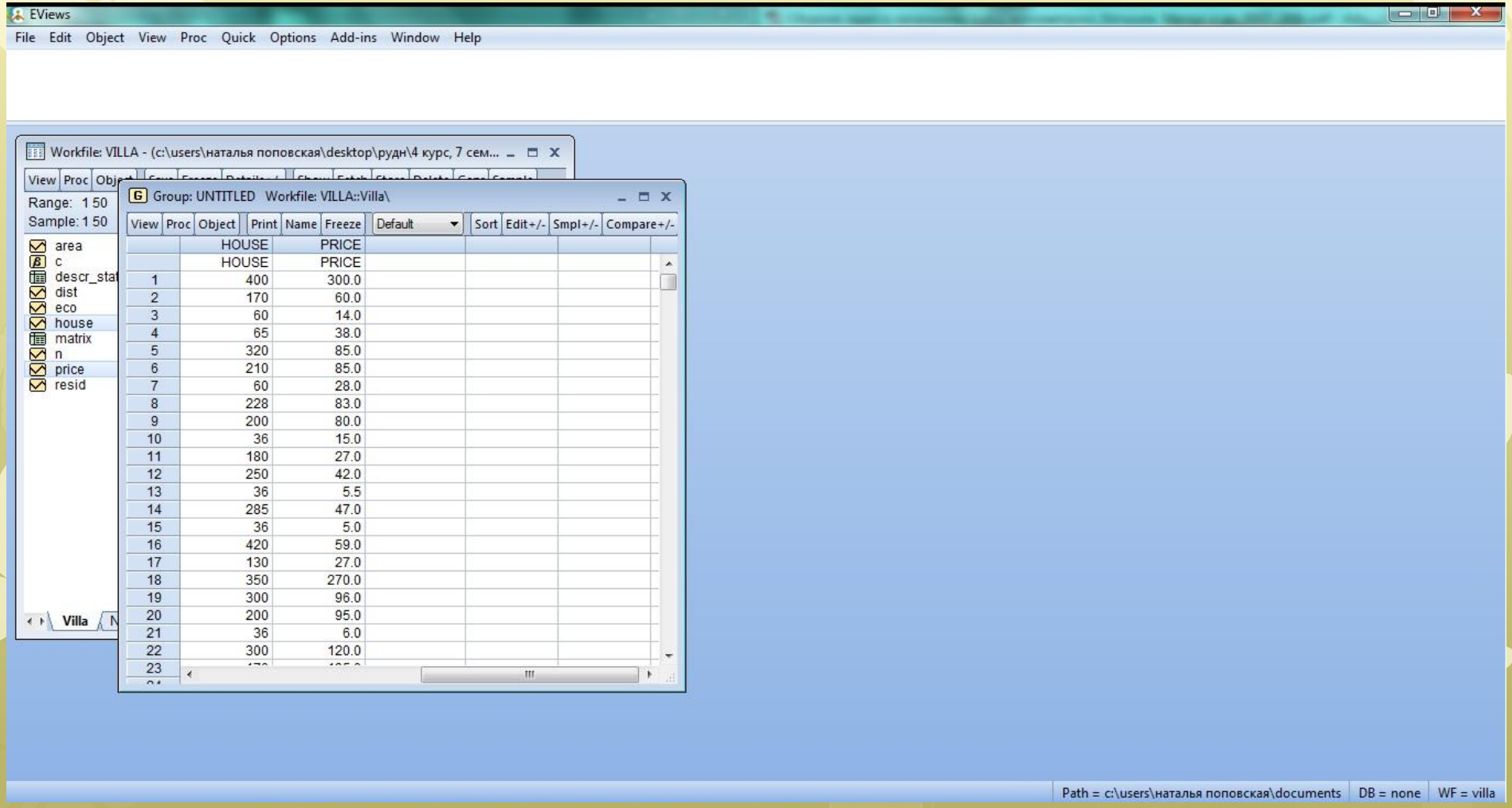
Correlation t-Statistic Probability	PRICE	HOUSE	DIST	AREA
PRICE	1.000000 ----- -----			
HOUSE	0.673724 6.316383 0.0000	1.000000 ----- -----		
DIST	-0.547239 -4.529862 0.0000	-0.541391 -4.461229 0.0000	1.000000 ----- -----	
AREA	0.576627 4.889787 0.0000	0.603486 5.243553 0.0000	-0.159648 -1.120448 0.2681	1.000000 ----- -----

Path = c:\users\наталья поповская\documents DB = none WF = villa

Построение диаграммы рассеяния [1, house-price]

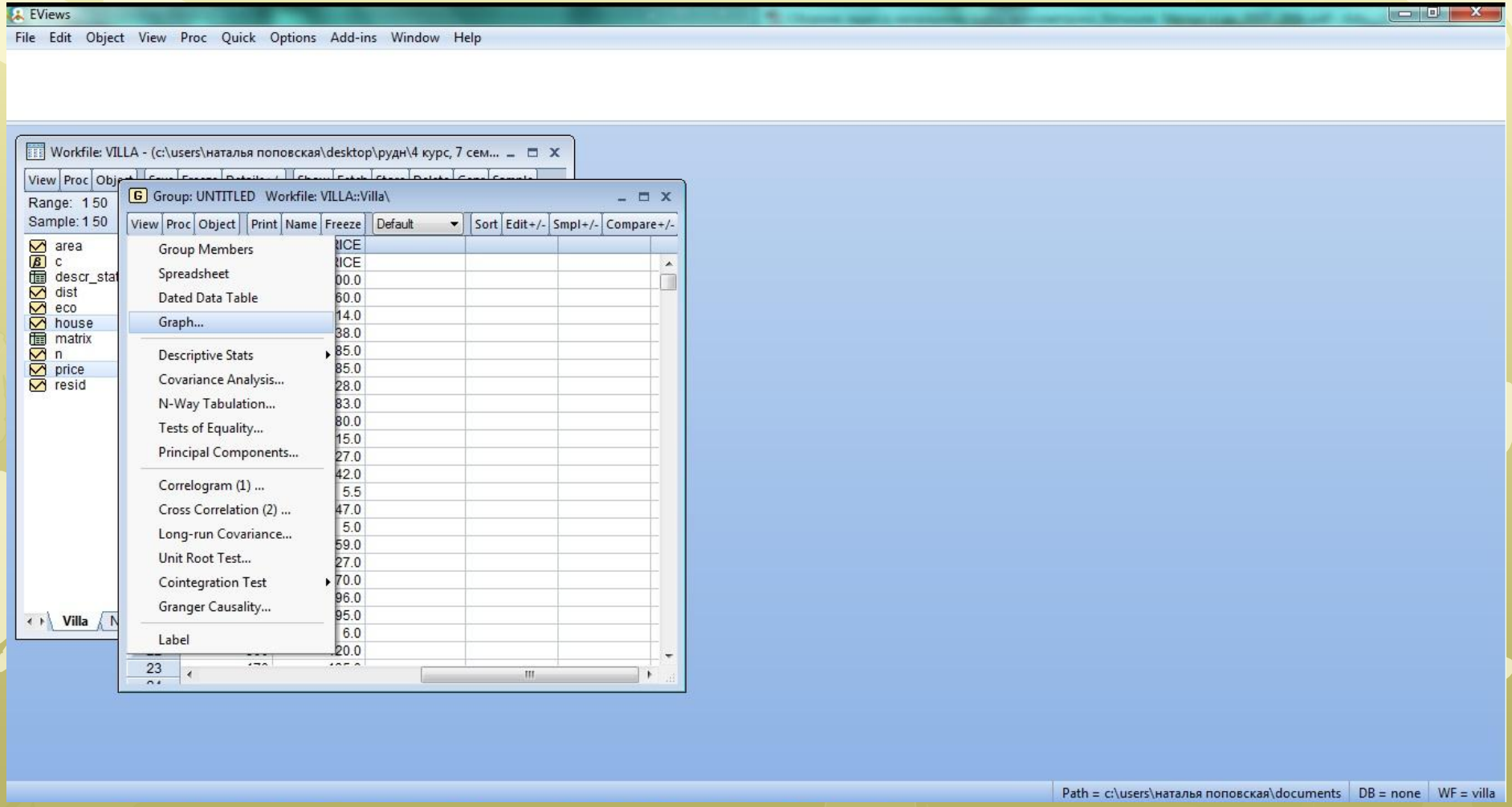


Построение диаграммы рассеяния [2, house-price]

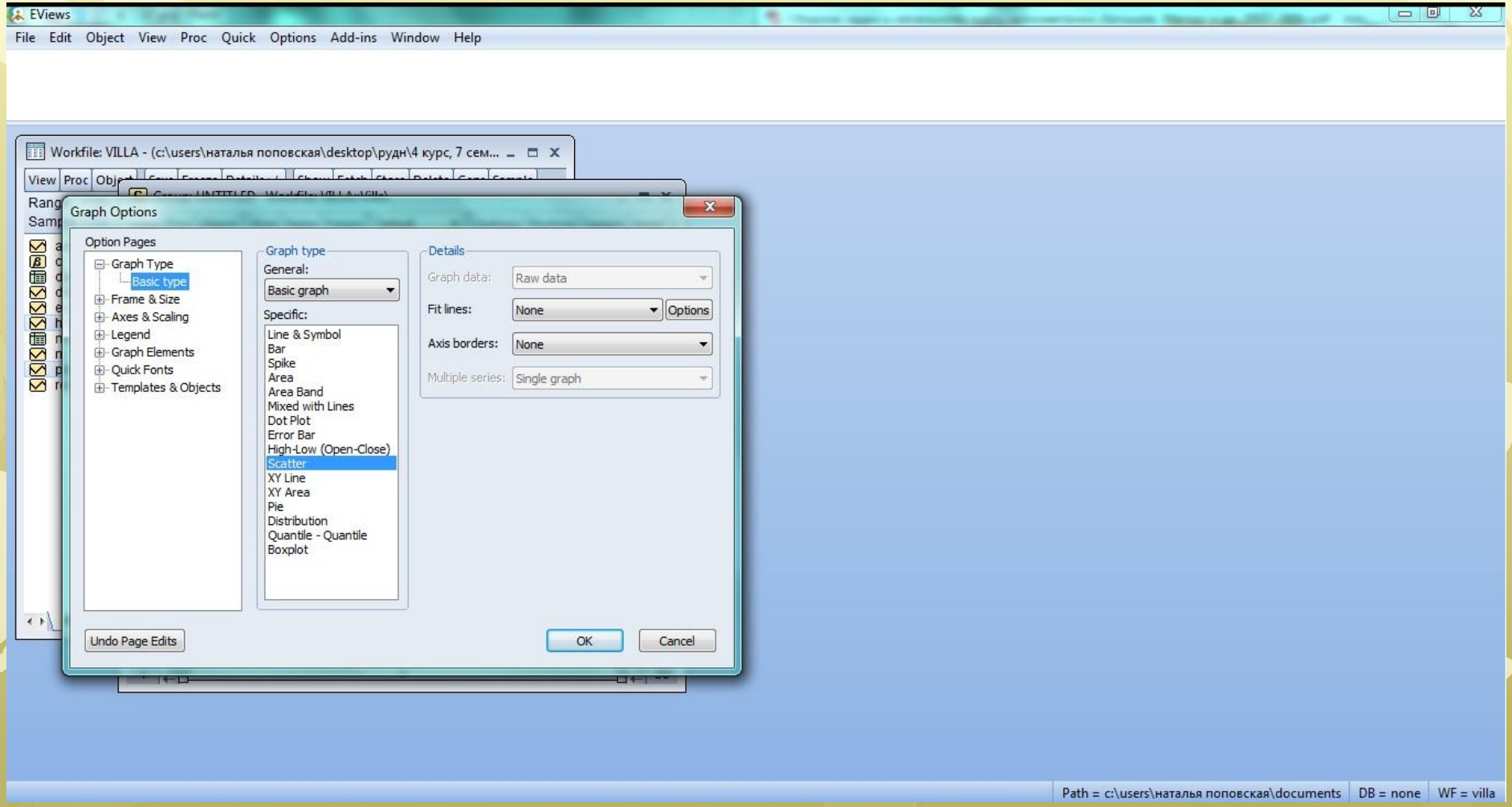


	HOUSE	PRICE
	HOUSE	PRICE
1	400	300.0
2	170	60.0
3	60	14.0
4	65	38.0
5	320	85.0
6	210	85.0
7	60	28.0
8	228	83.0
9	200	80.0
10	36	15.0
11	180	27.0
12	250	42.0
13	36	5.5
14	285	47.0
15	36	5.0
16	420	59.0
17	130	27.0
18	350	270.0
19	300	96.0
20	200	95.0
21	36	6.0
22	300	120.0
23		

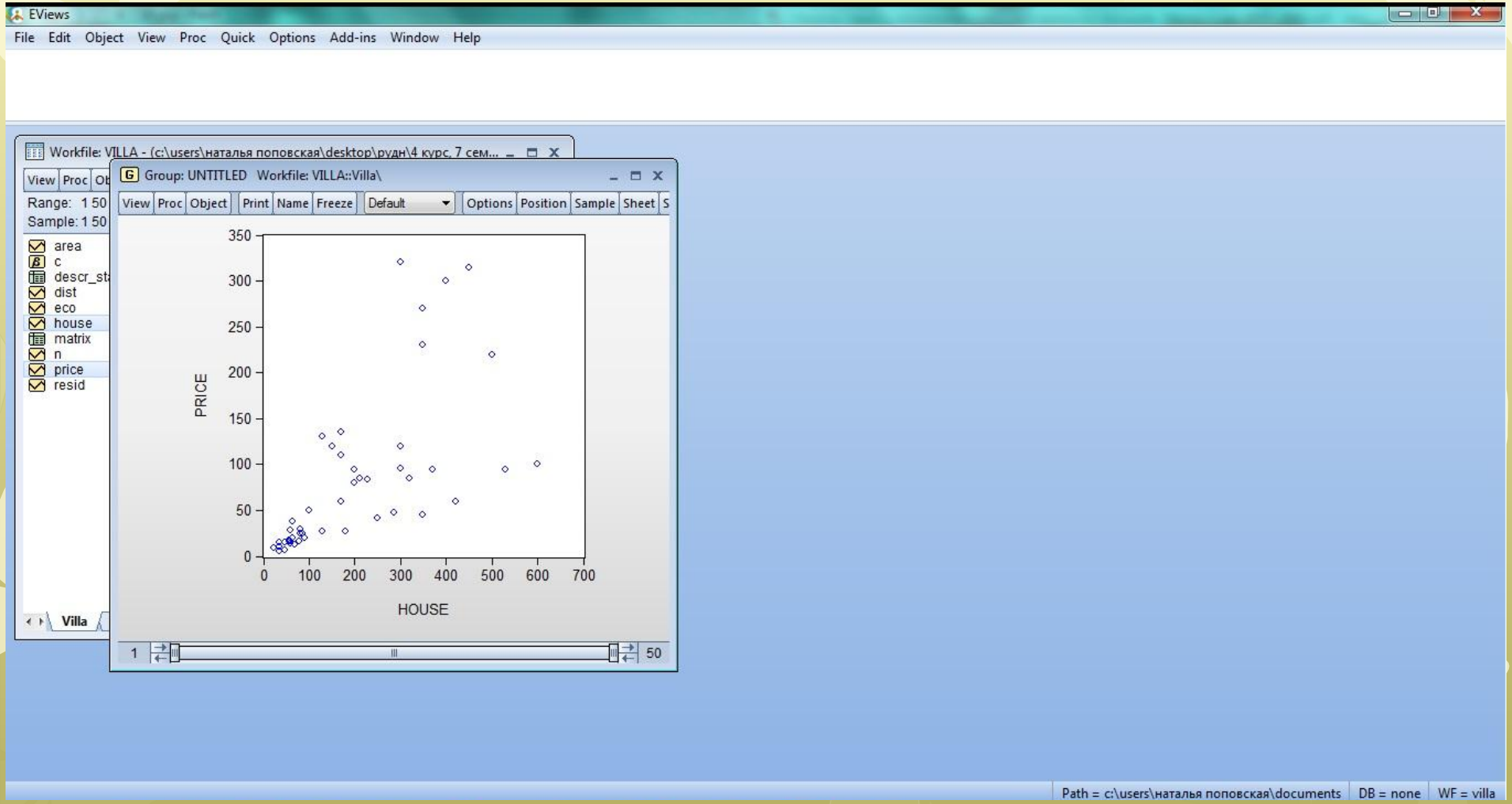
Построение диаграммы рассеяния [3, house-price]



Построение диаграммы рассеяния [4, house-price]



Построение диаграммы рассеяния [5, house-price]



Создание lnprice и lnhouse в командной строке командой genr lnprice=log(price) и genr lnhouse=log(house)

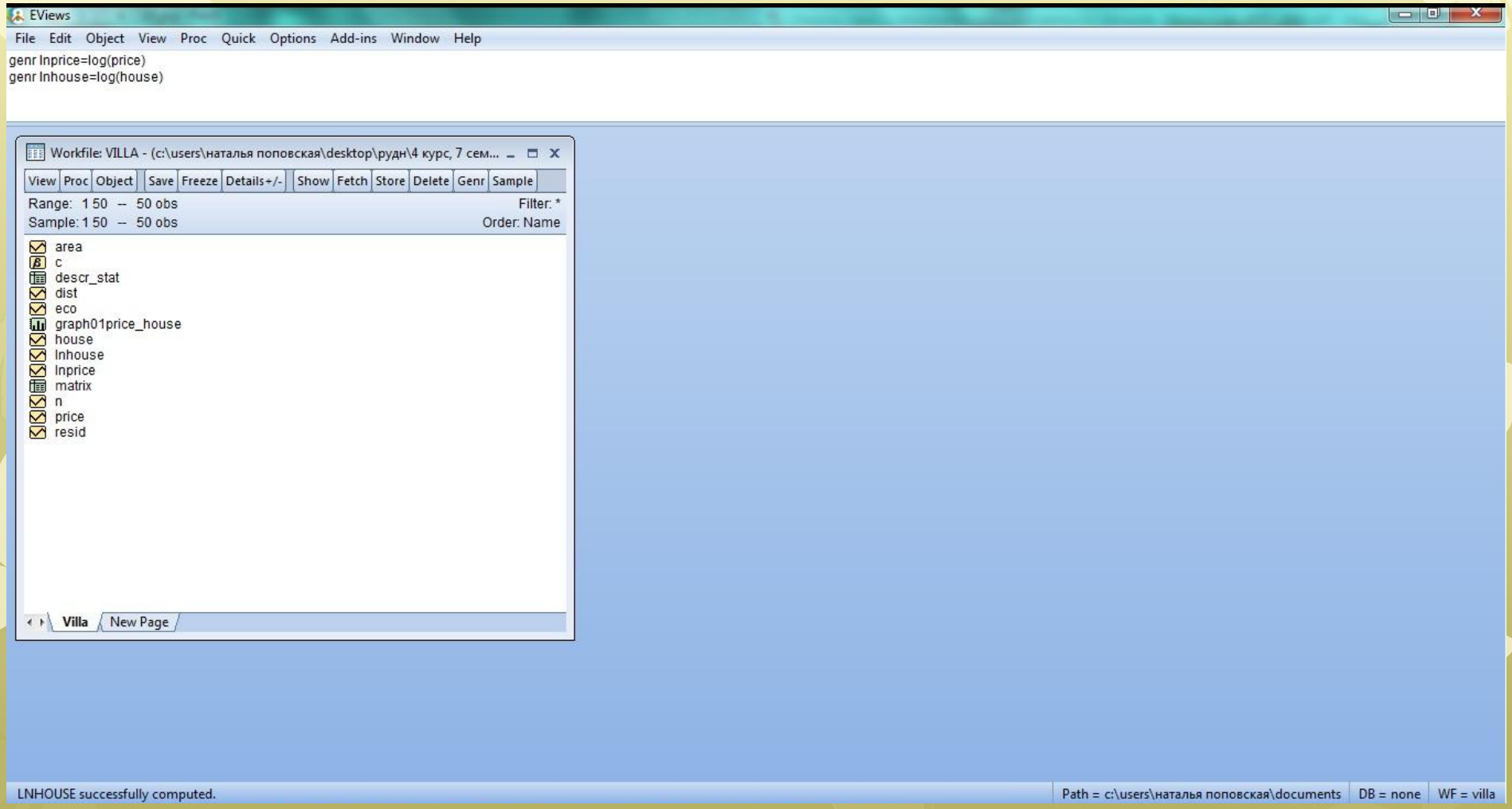


Диаграмма рассеяния lnhouse-price

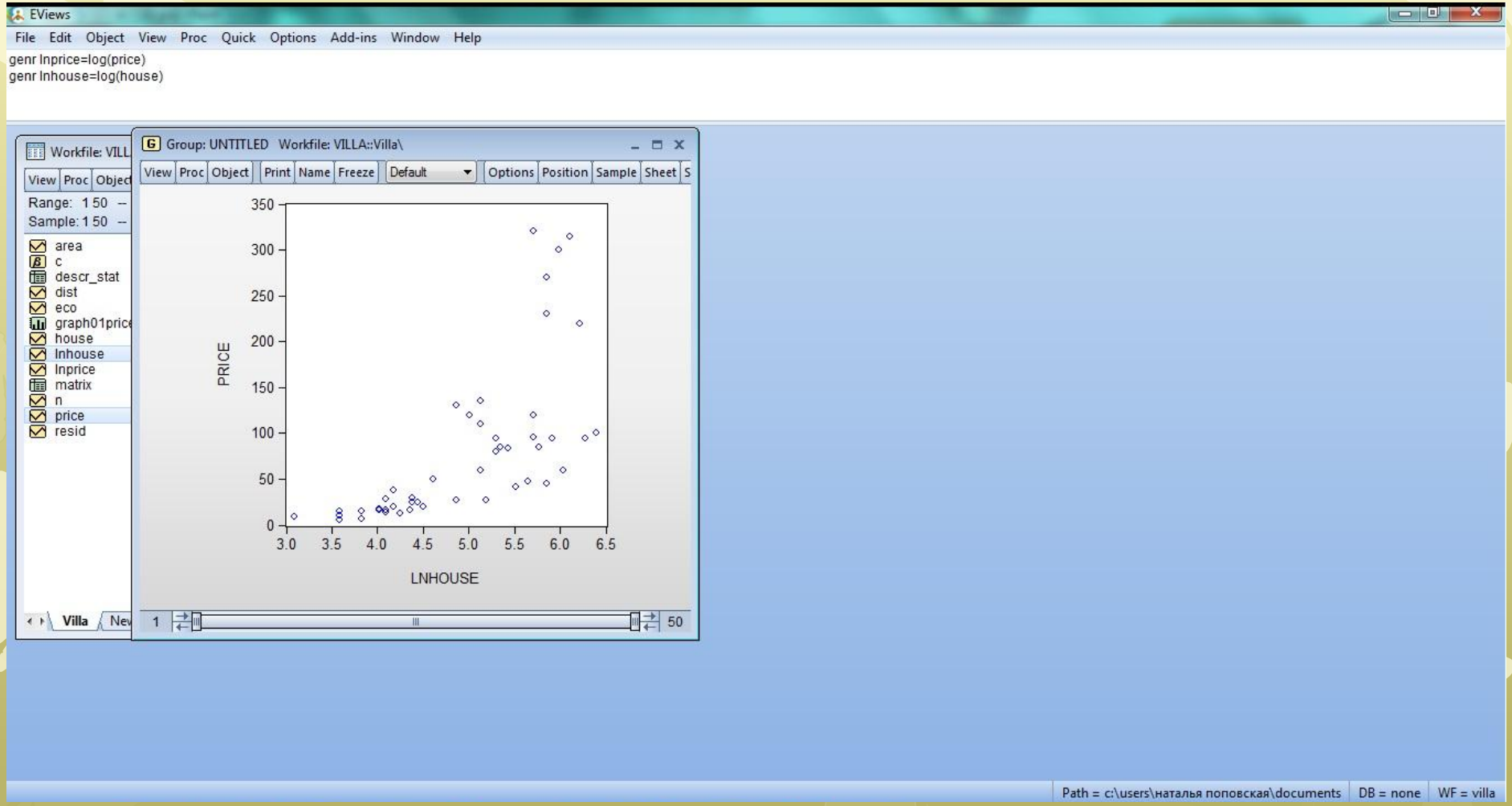


Диаграмма рассеяния house-lnprice

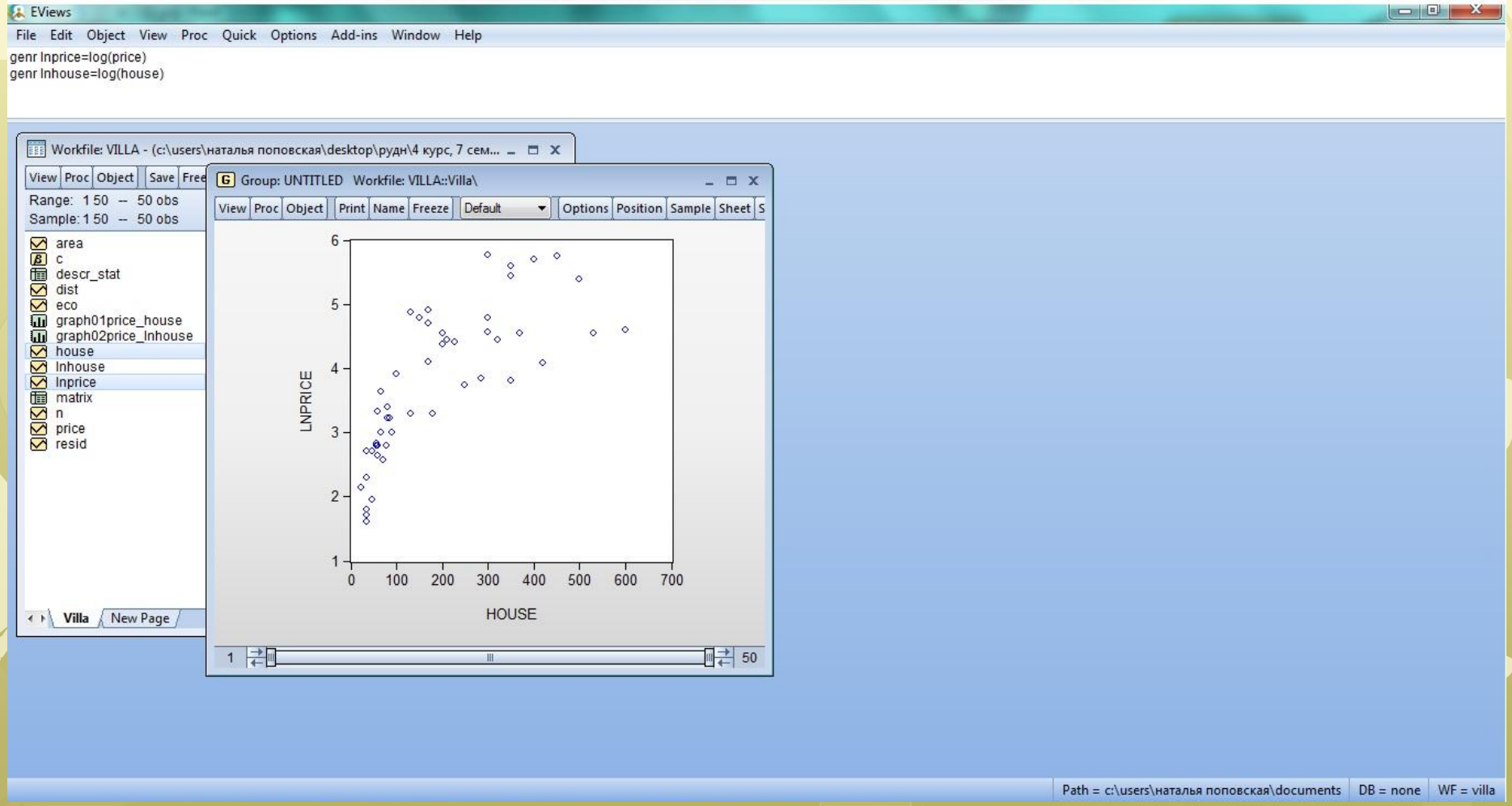
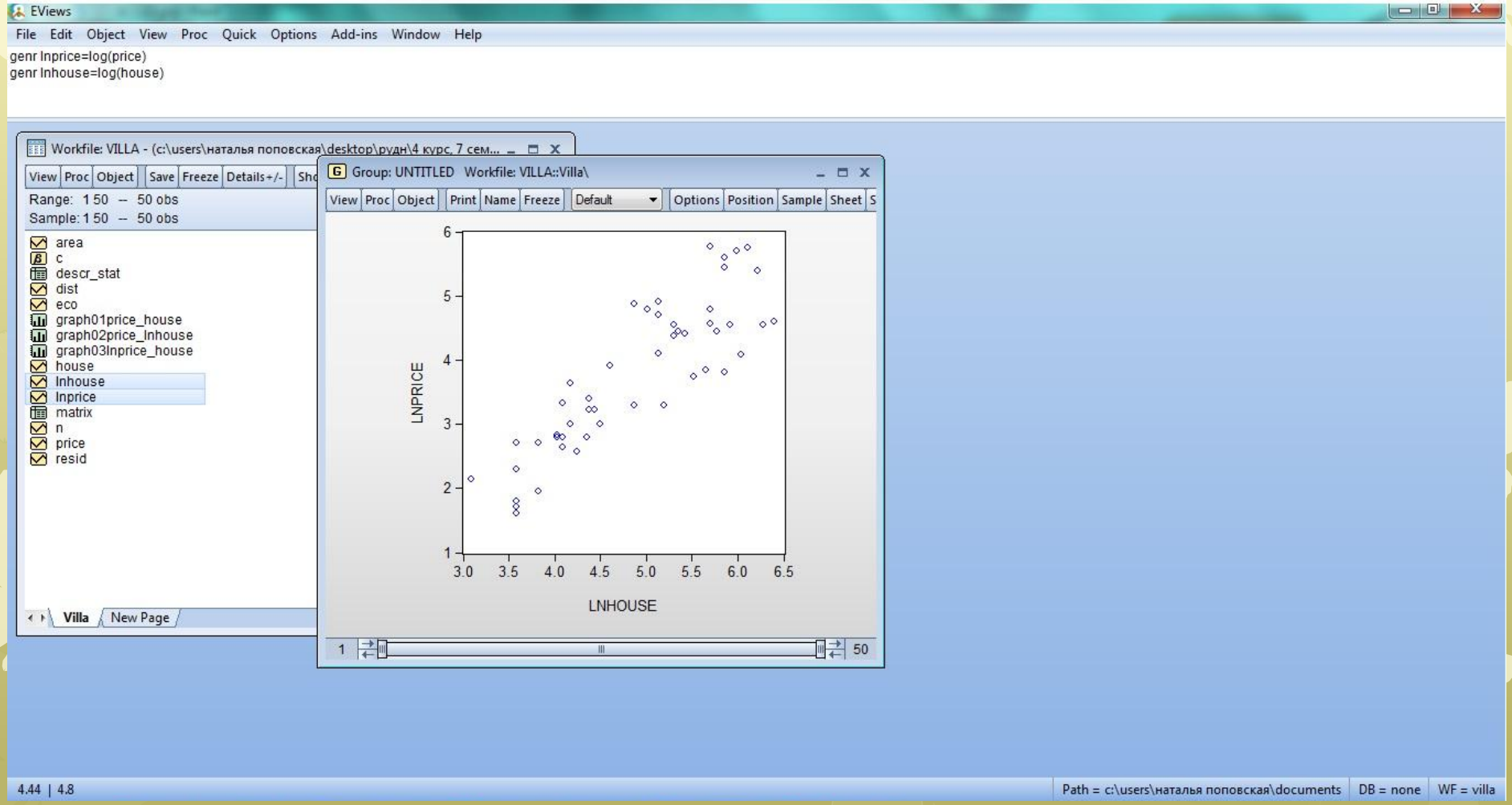
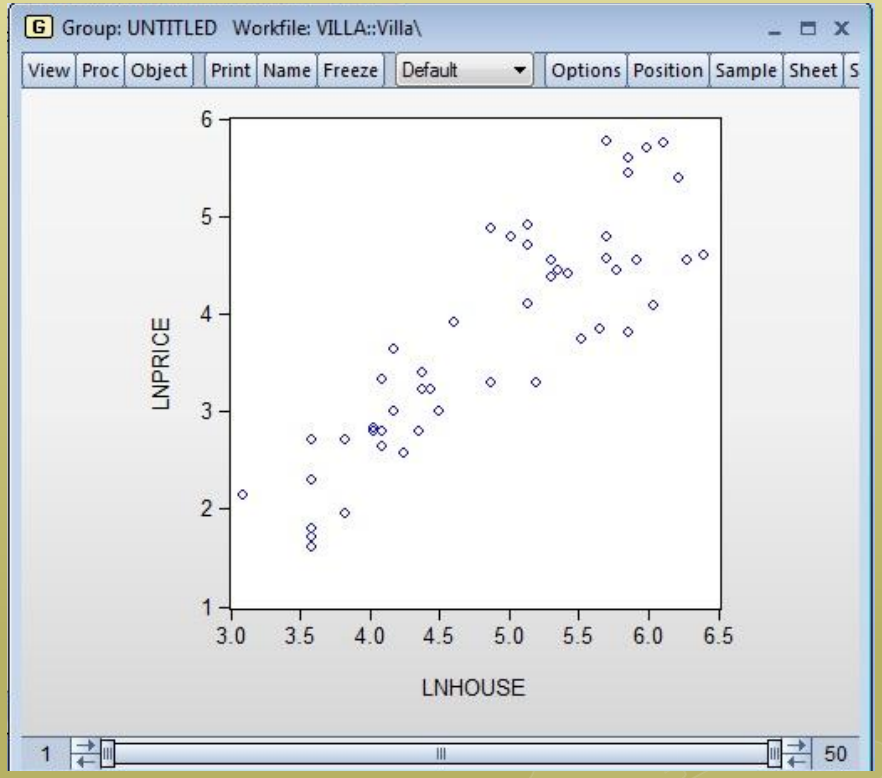


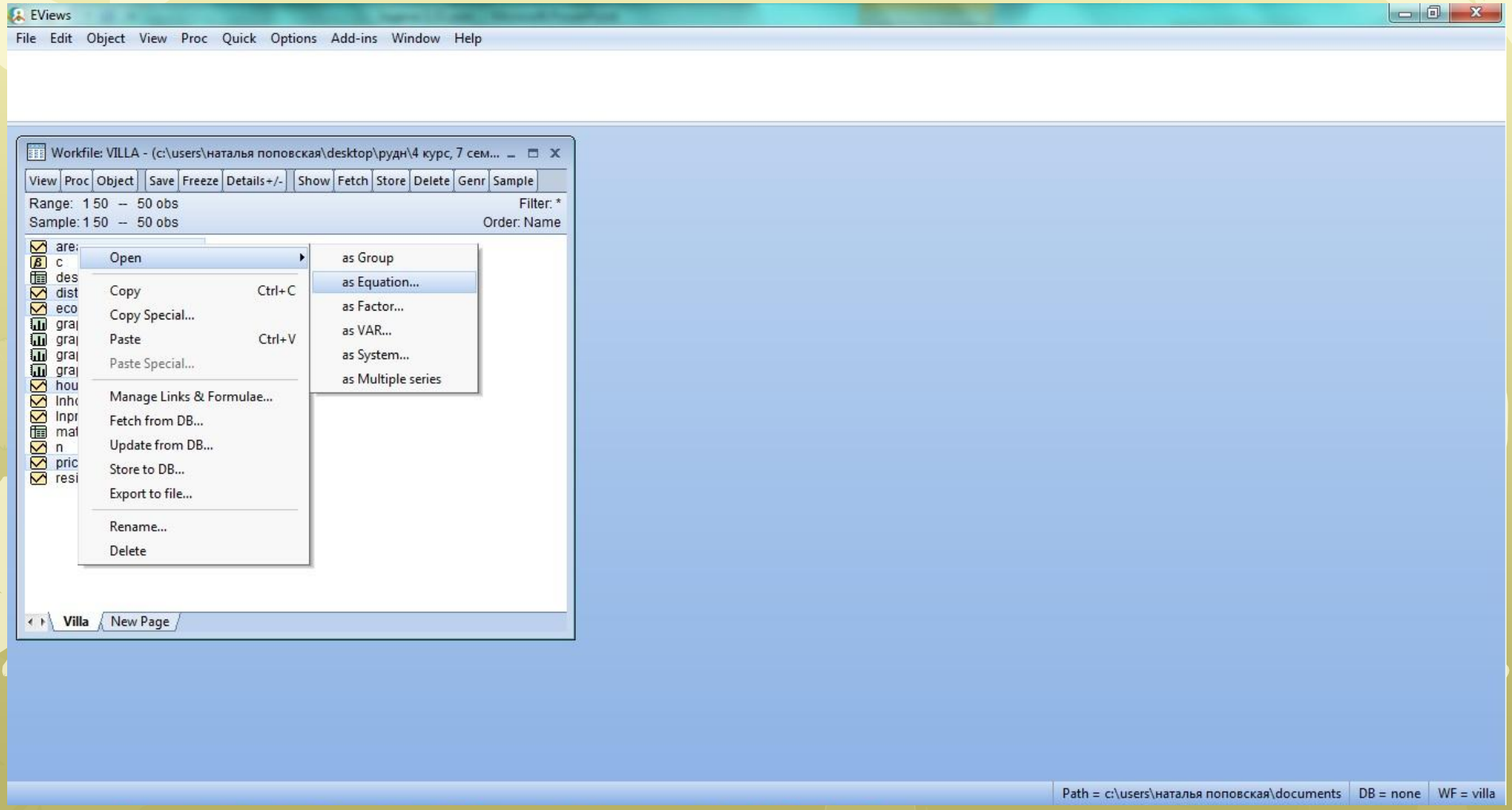
Диаграмма рассеяния lnhouse-lnprice



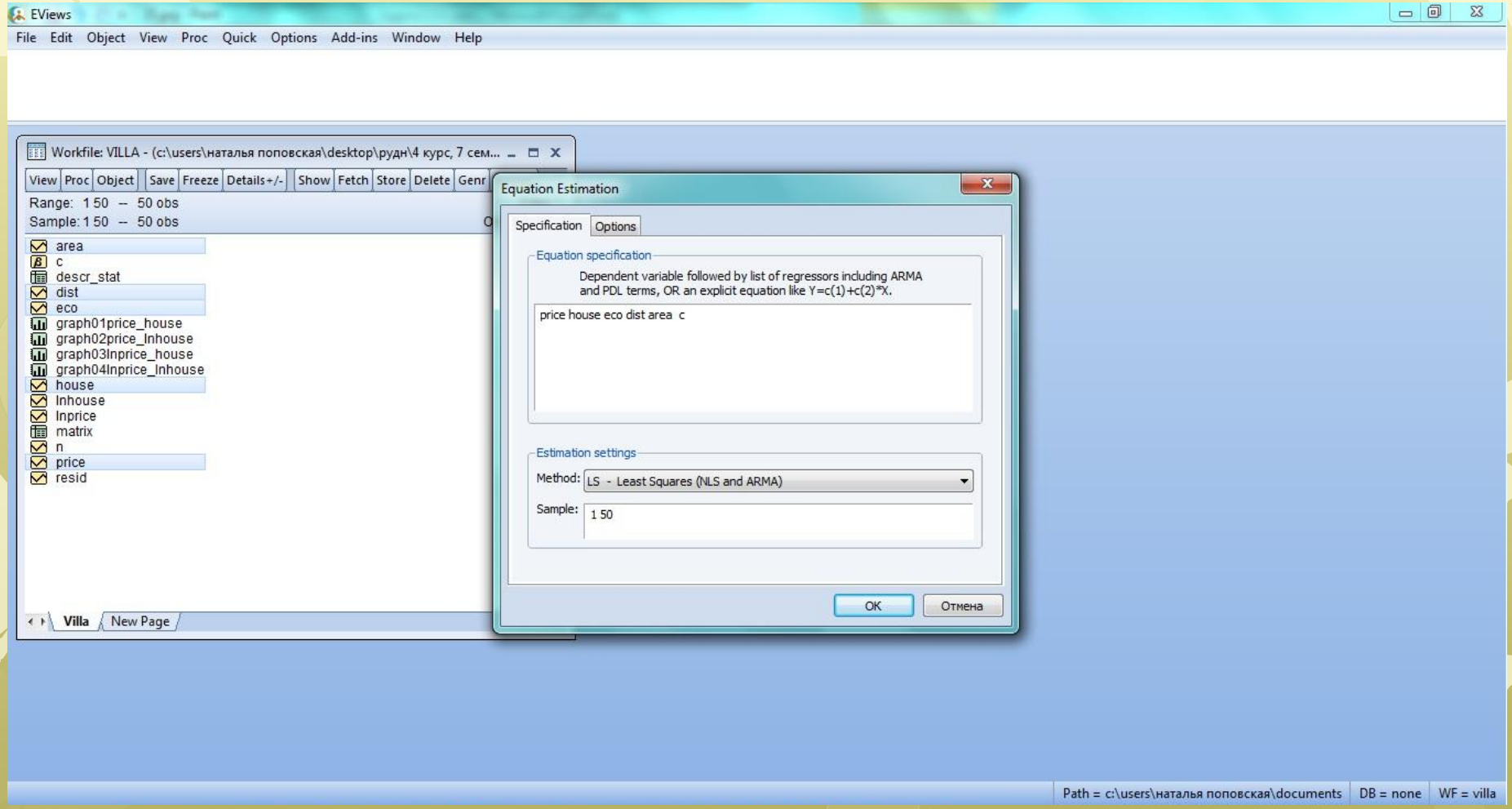
- Проанализировав диаграммы рассеяния, мы приходим к выводу, что самой хорошей функциональной формой будет логарифмическая функция(4-я диаграмма рассеяния lnhouse-lnprice)
- Перейдем к построению моделей



1. Линейная модель. Построение [1]

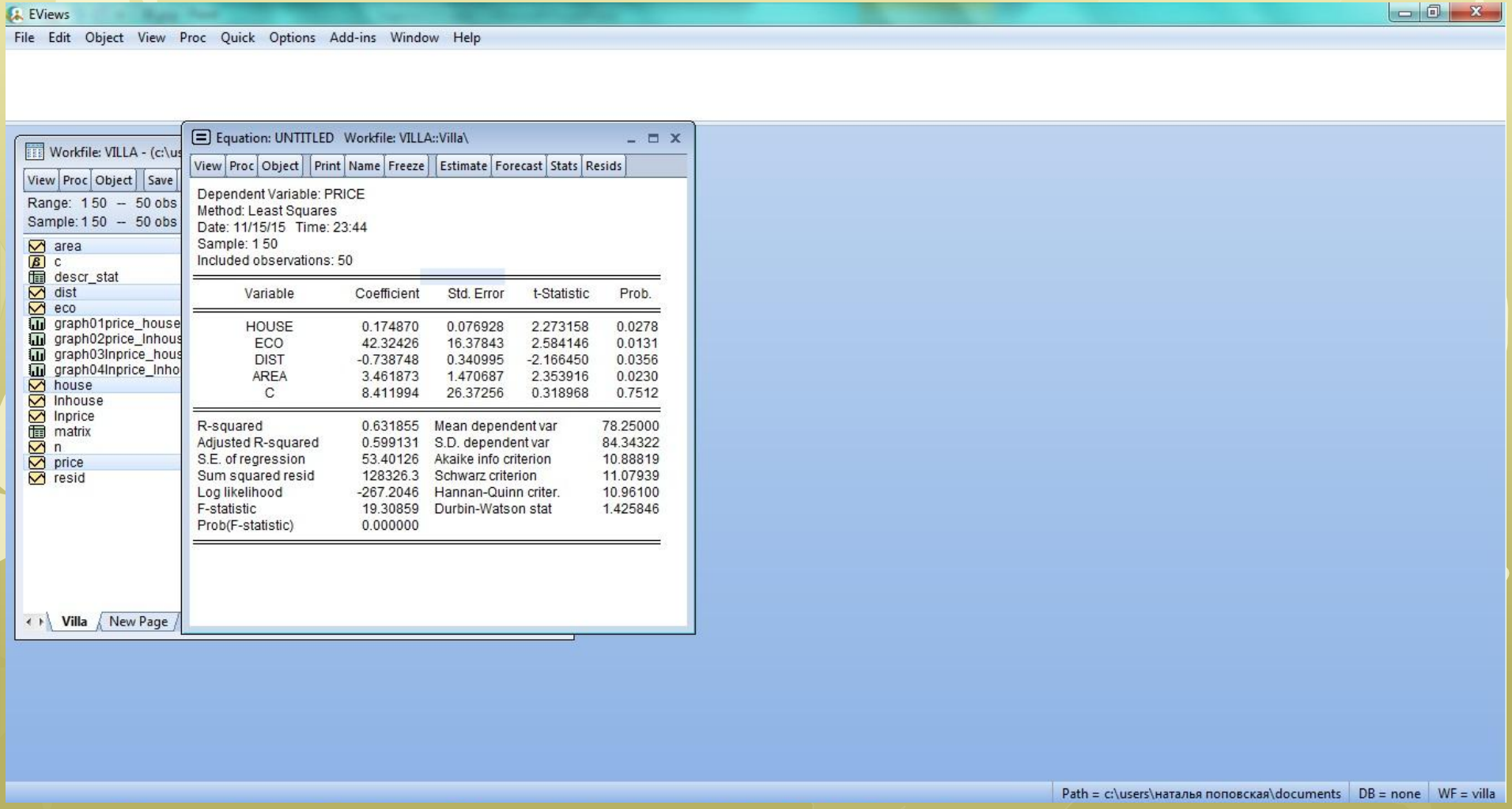


1. Линейная модель. Построение [2]

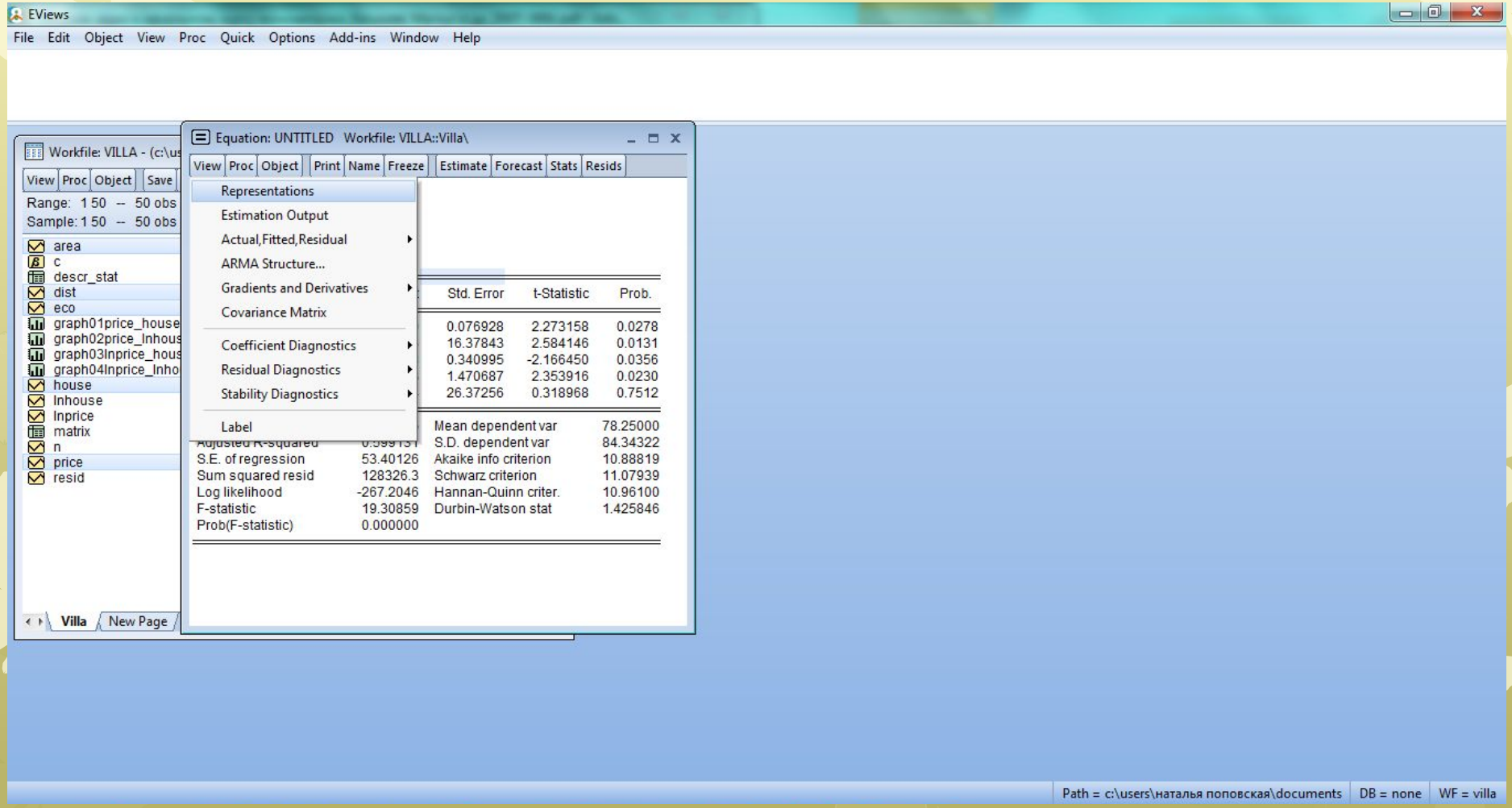


1. Линейная модель. Построение [3]

- В линейную модель включаем переменные без логарифмов. Все коэффициенты значимы (Prob<0.05, у Const не учитываем).R^2=0,631855, adj ^2=0.599131, модель значима



1. Линейная модель
Вывод уравнения [1]



1. Линейная модель
Вывод уравнения [2]. Интерпретация [1]

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$
- При возрастании x_j на 1 единицу (своего измерения), y возрастает на β_j единиц (своего измерения)

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Estimation Command:
==========================
LS PRICE HOUSE ECO DIST AREA  C

Estimation Equation:
==========================
PRICE = C(1)*HOUSE + C(2)*ECO + C(3)*DIST + C(4)*AREA + C(5)

Substituted Coefficients:
==========================
PRICE = 0.174869789136*HOUSE + 42.3242635228*ECO - 0.738748051679
*DIST + 3.46187330361*AREA + 8.41199372691
```

1. Линейная модель. Интерпретация [2]

- dist – при увеличении расстояния на 1 км цена коттеджа падает на 739$
- house – при увеличении площади дома на 1 кв.м цена коттеджа увеличивается на 175$

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Estimation Command:
==========================
LS PRICE HOUSE ECO DIST AREA  C

Estimation Equation:
==========================
PRICE = C(1)*HOUSE + C(2)*ECO + C(3)*DIST + C(4)*AREA + C(5)

Substituted Coefficients:
==========================
PRICE = 0.174869789136*HOUSE + 42.3242635228*ECO - 0.738748051679
*DIST + 3.46187330361*AREA + 8.41199372691
```

1. Линейная модель. Интерпретация [3]

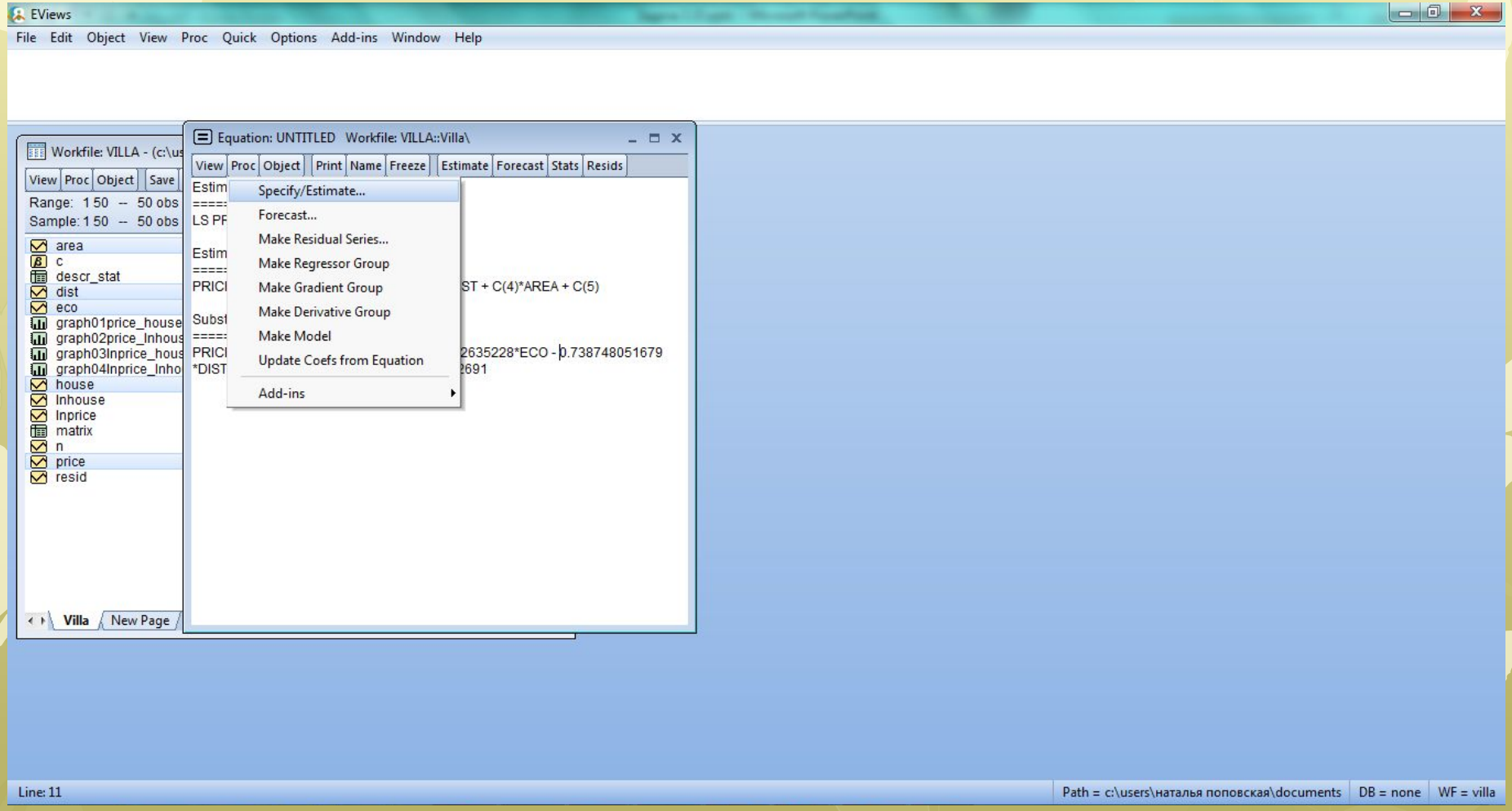
- eco – если рядом есть реки и озера, то цена возрастает на 42 тыс $
- area – при увеличении площади участка на 1 сотку цена увеличивается на 3462 $

Equation: UNTITLED Workfile: VILLA::Villa\

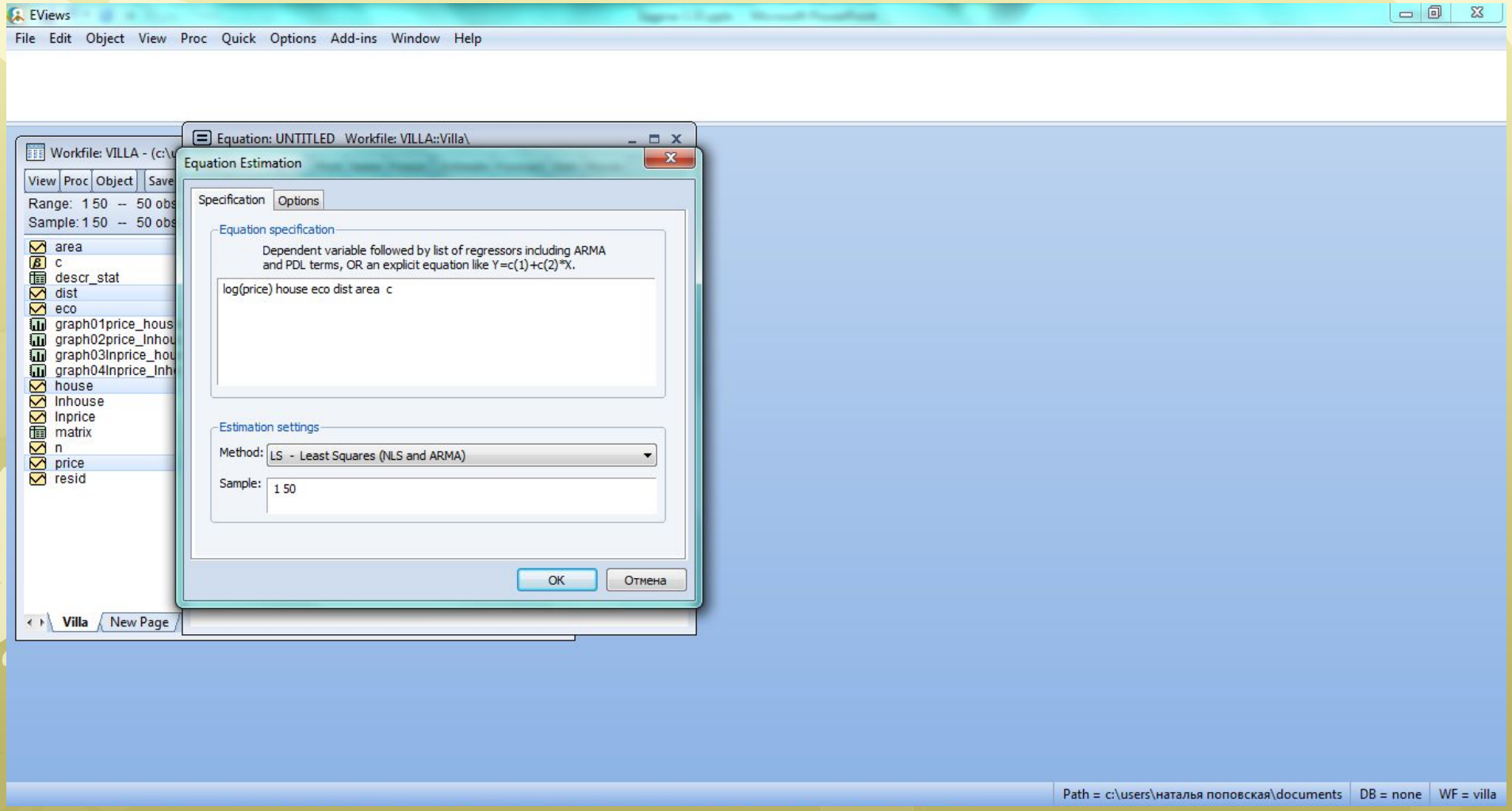
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Estimation Command:
==========================
LS PRICE HOUSE ECO DIST AREA  C

Estimation Equation:
==========================
PRICE = C(1)*HOUSE + C(2)*ECO + C(3)*DIST + C(4)*AREA + C(5)

Substituted Coefficients:
==========================
PRICE = 0.174869789136*HOUSE + 42.3242635228*ECO - 0.738748051679
*DIST + 3.46187330361*AREA + 8.41199372691
```

2. Полулогарифмическая модель (log(y)). Построение [1]



2. Полулогарифмическая модель (log(y)). Построение [2]



2. Полулогарифмическая модель (log(y)). Построение [3]

- Коэффициенты значимы (Prob<0.05), R^2=0.782721, adjR^2=0.763408, заметим, что они выше, чем у линейной модели. Модель значима.

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(PRICE)
Method: Least Squares
Date: 11/16/15 Time: 00:49
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
HOUSE	0.002941	0.000802	3.666452	0.0006
ECO	0.440003	0.170806	2.576037	0.0133
DIST	-0.015559	0.003556	-4.375340	0.0001
AREA	0.035805	0.015337	2.334464	0.0241
C	3.192330	0.275032	11.60712	0.0000

R-squared	0.782721	Mean dependent var	3.784721
Adjusted R-squared	0.763408	S.D. dependent var	1.144939
S.E. of regression	0.556907	Akaike info criterion	1.761801
Sum squared resid	13.95653	Schwarz criterion	1.953004
Log likelihood	-39.04503	Hannan-Quinn criter.	1.834612
F-statistic	40.52685	Durbin-Watson stat	2.099562
Prob(F-statistic)	0.000000		

2. Полулогарифмическая модель (log(y)). Вывод уравнения. Интерпретация [1]

- $\ln(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$
- При изменении x_j на 1 единицу, y меняется на (e^{β_j} -1)*100% (при малых $-0.2 < \beta_j < 0.2$ это примерно равно βj *100%)

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Estimation Command:
==========================
LS LOG(PRICE) HOUSE ECO DIST AREA  C

Estimation Equation:
==========================
LOG(PRICE) = C(1)*HOUSE + C(2)*ECO + C(3)*DIST + C(4)*AREA + C(5)

Substituted Coefficients:
==========================
LOG(PRICE) = 0.00294145354071*HOUSE + 0.440002745825*ECO -
0.0155593082047*DIST + 0.0358045585599*AREA + 3.19232963887
```

2. Полулогарифмическая модель (log(y)). Интерпретация [2]

- house - при изменении площади дома на 1 кв.м цена меняется на 0.29% (т.к. $-0.2<\beta_j<0.2$)
- eco – если рядом есть реки и озера, то цена увеличивается на 55%

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Estimation Command:
==========================
LS LOG(PRICE) HOUSE ECO DIST AREA  C

Estimation Equation:
==========================
LOG(PRICE) = C(1)*HOUSE + C(2)*ECO + C(3)*DIST + C(4)*AREA + C(5)

Substituted Coefficients:
==========================
LOG(PRICE) = 0.00294145354071*HOUSE + 0.440002745825*ECO -
0.0155593082047*DIST + 0.0358045585599*AREA + 3.19232963887
```

2. Полулогарифмическая модель (log(y)). Интерпретация [3]

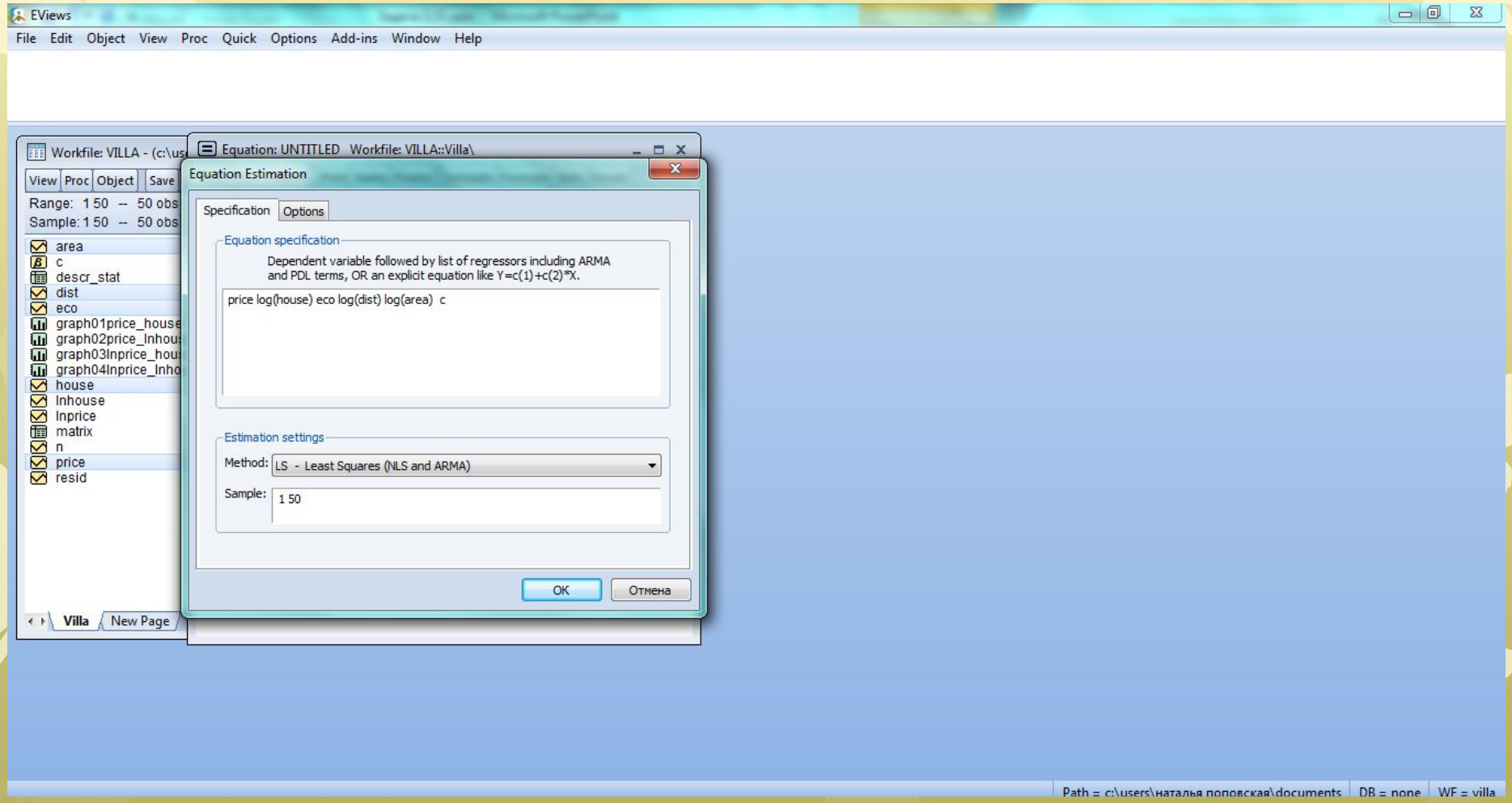
- dist – при увеличении расстояния на 1 км цена снижается на 1.6% (т.к. $-0.2<\beta_j<0.2$)
- area – при увеличении площади участка на 1 сотку цена меняется на 3.6%

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Estimation Command:
==========================
LS LOG(PRICE) HOUSE ECO DIST AREA  C

Estimation Equation:
==========================
LOG(PRICE) = C(1)*HOUSE + C(2)*ECO + C(3)*DIST + C(4)*AREA + C(5)

Substituted Coefficients:
==========================
LOG(PRICE) = 0.00294145354071*HOUSE + 0.440002745825*ECO -
0.0155593082047*DIST + 0.0358045585599*AREA + 3.19232963887
```

3. Полулогарифмическая модель (log(x)). Построение [1]



3. Полулогарифмическая модель (log(x)). Построение [2]

- Коэффициенты значимы (Prob<0.05, у Const не учитываем). R^2=0.641281, adj R^2=0.609395, заметим, что R^2 ниже, чем у полулогарифмической (log(y)) , но выше, чем у линейной. Модель значима.

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

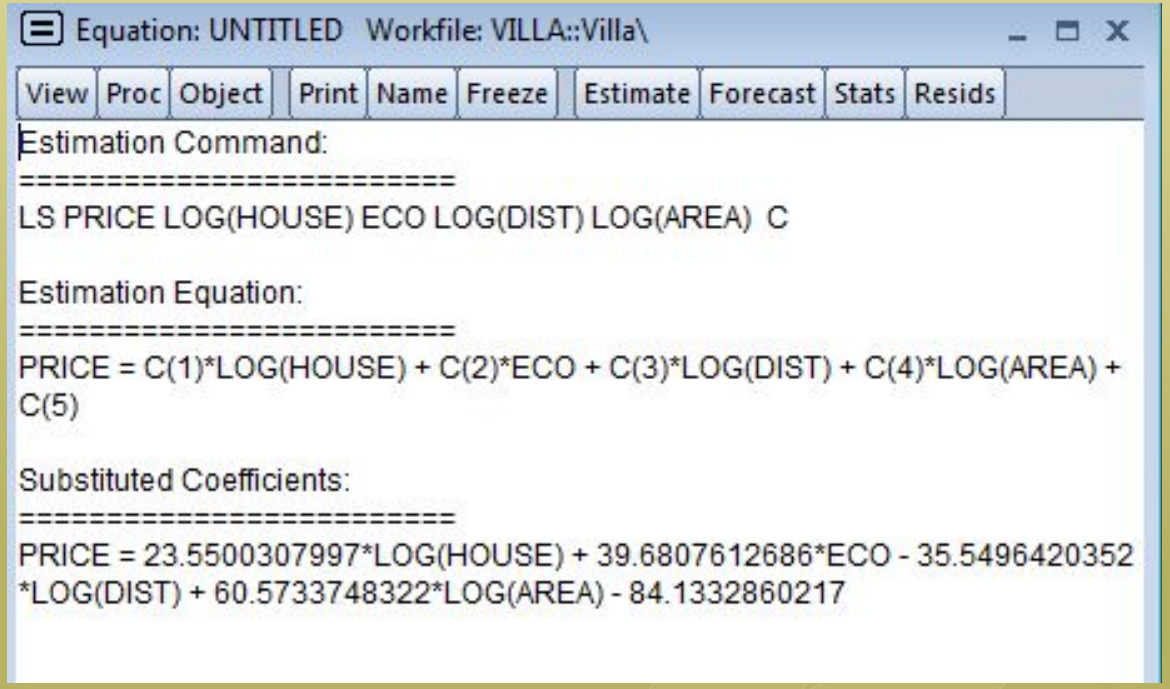
Dependent Variable: PRICE
Method: Least Squares
Date: 11/16/15 Time: 01:16
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(HOUSE)	23.55003	11.19622	2.103391	0.0411
ECO	39.68076	14.31986	2.771029	0.0081
LOG(DIST)	-35.54964	9.666786	-3.677504	0.0006
LOG(AREA)	60.57337	18.14896	3.337568	0.0017
C	-84.13329	67.45095	-1.247325	0.2187

R-squared	0.707498	Mean dependent var	78.25000
Adjusted R-squared	0.681498	S.D. dependent var	84.34322
S.E. of regression	47.59995	Akaike info criterion	10.65818
Sum squared resid	101959.0	Schwarz criterion	10.84938
Log likelihood	-261.4545	Hannan-Quinn criter.	10.73099
F-statistic	27.21125	Durbin-Watson stat	1.708642
Prob(F-statistic)	0.000000		

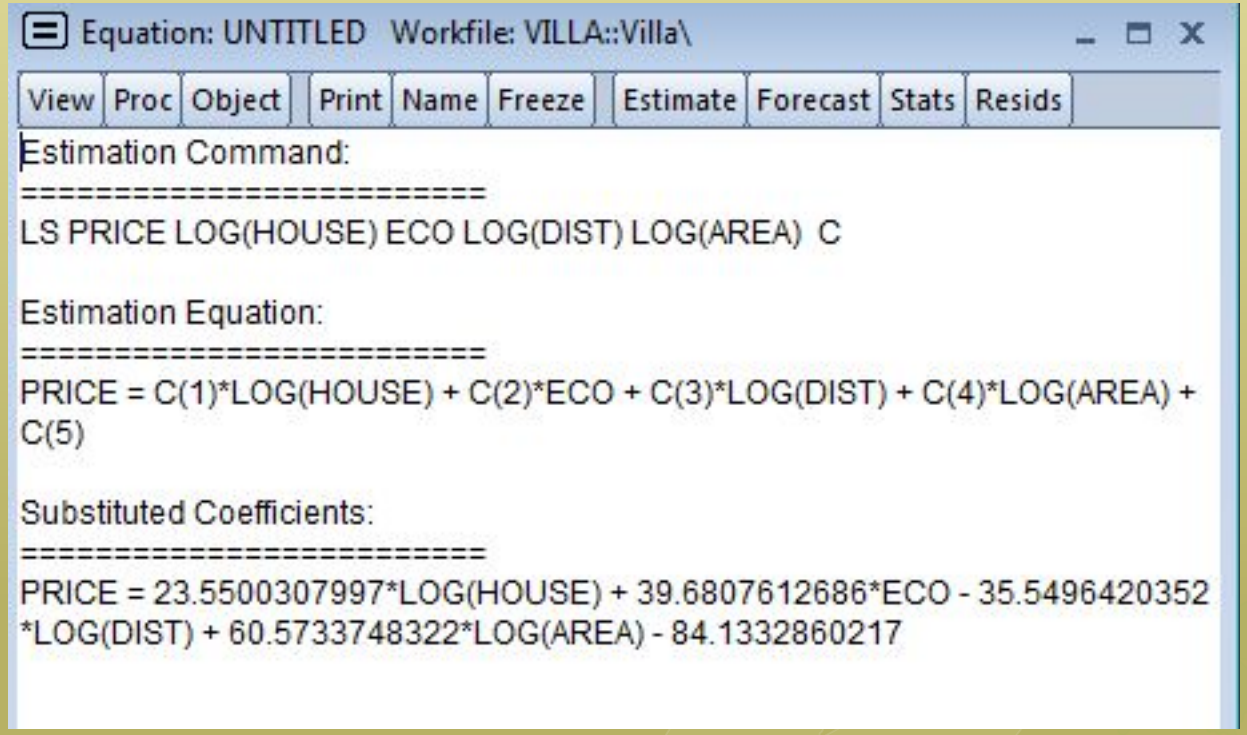
3. Полулогарифмическая модель (log(x)). Вывод уравнения. Интерпретация [1]

- $y = \beta_0 + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \ldots + \beta_n \ln(x_n)$
- При изменении x_j на 1 %, y меняется в среднем на $\beta_j/100$ единиц измерения



3. Полулогарифмическая модель (log(x)). Вывод уравнения. Интерпретация [2]

- house – при увеличении площади дома на 1 кв.м цена увеличивается на 0.24 тыс $
- dist – при увеличении расстояния на 1 км цена уменьшится на 0.36 тыс $



3. Полулогарифмическая модель (log(x)). Вывод уравнения. Интерпретация [3]

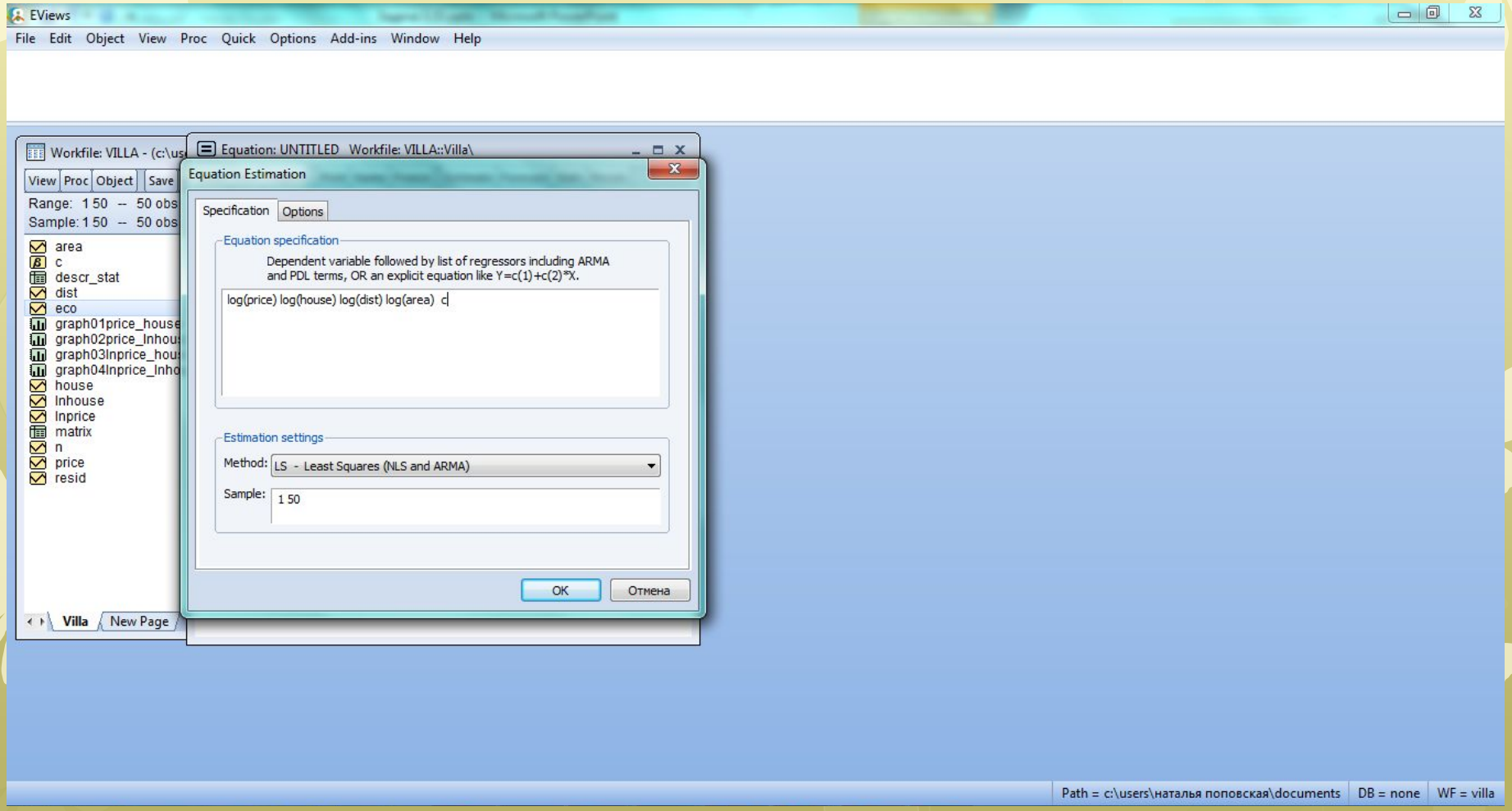
- area – при увеличении площади участка на 1 сотку цена увеличится на 0.6 тыс $
- eco – если рядом есть реки и озера, то цена увеличивается на 40 тыс $ (у eco не стоит log, т.к. принимает значения только 0 и 1)

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Estimation Command:
==========================
LS PRICE LOG(HOUSE) ECO LOG(DIST) LOG(AREA)  C

Estimation Equation:
==========================
PRICE = C(1)*LOG(HOUSE) + C(2)*ECO + C(3)*LOG(DIST) + C(4)*LOG(AREA) +
C(5)

Substituted Coefficients:
==========================
PRICE = 23.5500307997*LOG(HOUSE) + 39.6807612686*ECO - 35.5496420352
*LOG(DIST) + 60.5733748322*LOG(AREA) - 84.1332860217
```

4. Логарифмическая модель. Построение [1]



4. Логарифмическая модель. Построение [2]

- Мы не взяли в модель eco, т.к. это фиктивная переменная (принимает значения только 0 и 1)Коэффициенты значимы (Prob<0.05, у Const не учитываем). R^2=0.821542, adjR^2=0.809904, коэффициенты выше, чем у других моделей. Модель значима.

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(PRICE)
Method: Least Squares
Date: 11/16/15 Time: 01:31
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(HOUSE)	0.791142	0.117151	6.753190	0.0000
LOG(DIST)	-0.358343	0.098157	-3.650692	0.0007
LOG(AREA)	0.449922	0.186367	2.414167	0.0198
C	0.034687	0.699384	0.049597	0.9607

R-squared	0.821542	Mean dependent var	3.784721
Adjusted R-squared	0.809904	S.D. dependent var	1.144939
S.E. of regression	0.499194	Akaike info criterion	1.524973
Sum squared resid	11.46294	Schwarz criterion	1.677935
Log likelihood	-34.12433	Hannan-Quinn criter.	1.583222
F-statistic	70.58799	Durbin-Watson stat	1.776973
Prob(F-statistic)	0.000000		

4. Логарифмическая модель. Интерпретация [1]

- $\ln(y) = \beta_0 + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \ldots + \beta_n \ln(x_n)$
- При изменении x_j на 1 %, y меняется на β_j %

4. Логарифмическая модель. Интерпретация [2]

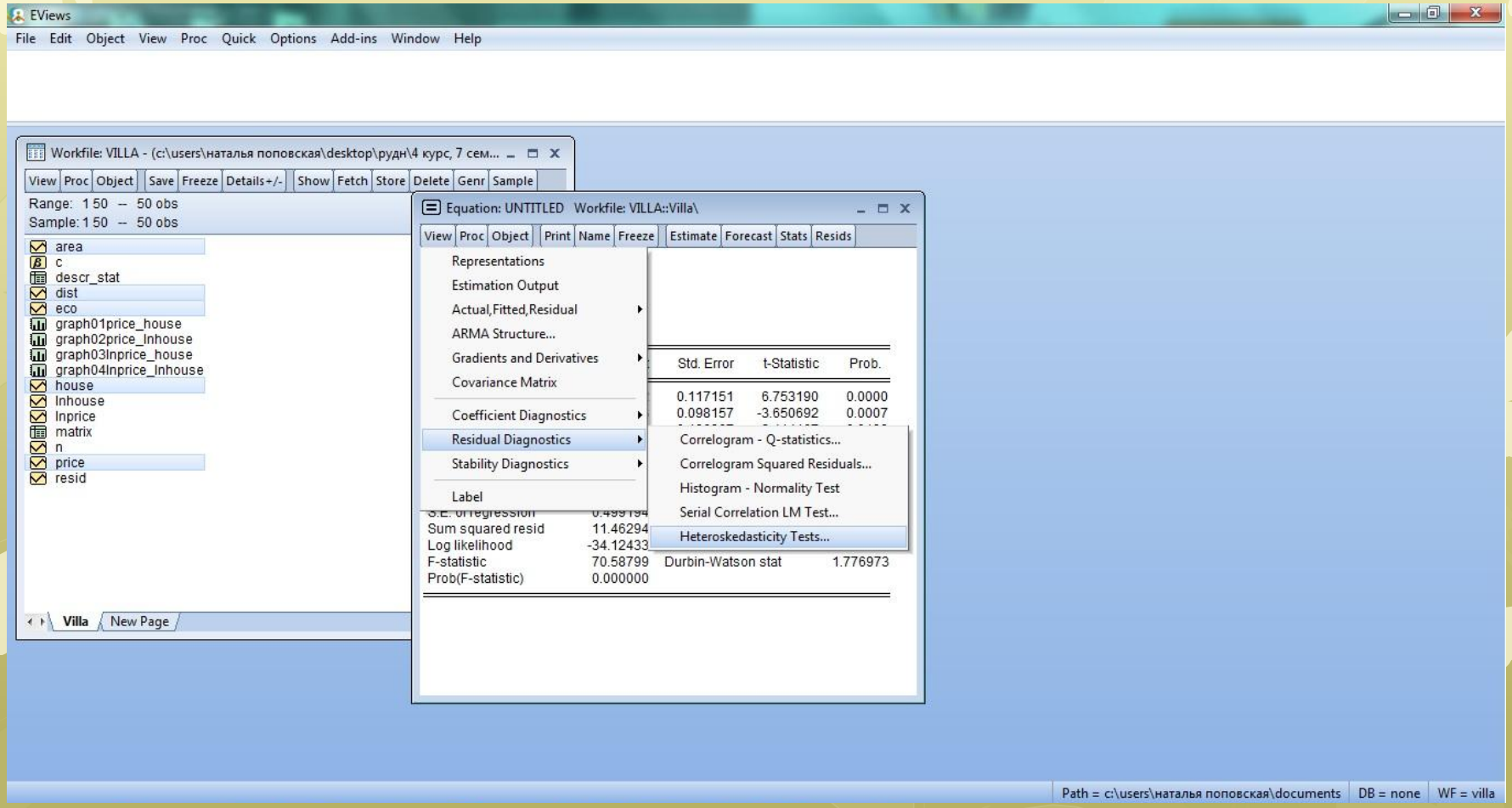
- house – при увеличении площади дома на 1 % цена увеличивается на 0.79 %
- dist – при увеличении расстояния на 1 % цена уменьшается на 0.36 %
- area – при увеличении площади участка на 1 % цена увеличится на 0.45 %

Equation: UNTITLED Workfile: VILLA::Villa\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

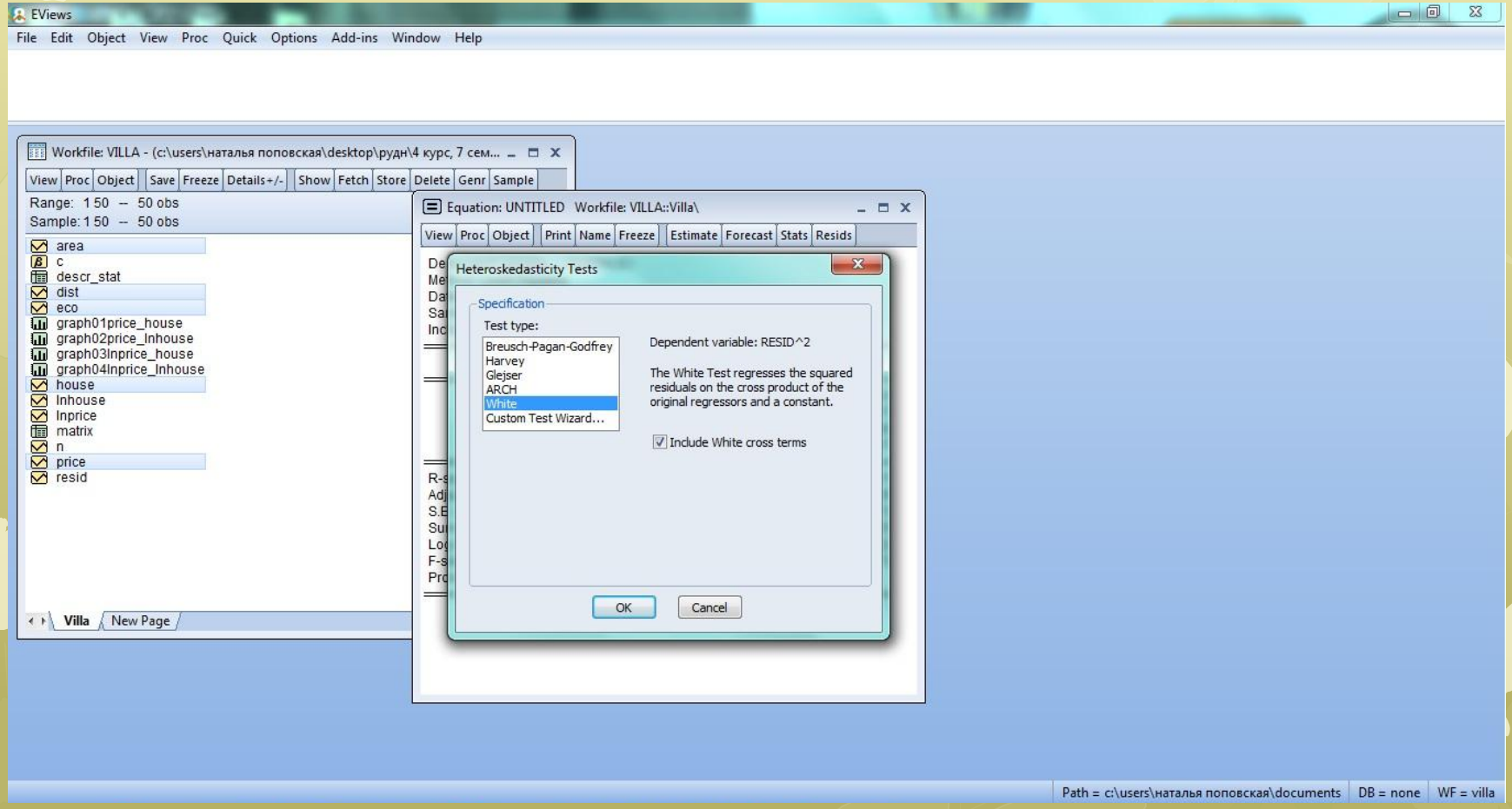
```
Estimation Command:
==========================
LS LOG(PRICE) LOG(HOUSE) LOG(DIST) LOG(AREA)  C

Estimation Equation:
==========================
LOG(PRICE) = C(1)*LOG(HOUSE) + C(2)*LOG(DIST) + C(3)*LOG(AREA) + C(4)

Substituted Coefficients:
==========================
LOG(PRICE) = 0.791141669725*LOG(HOUSE) - 0.358342794639*LOG(DIST) +
0.449921513669*LOG(AREA) + 0.0346874551441
```

Проверка логарифмической модели на гетероскедастичность[1]



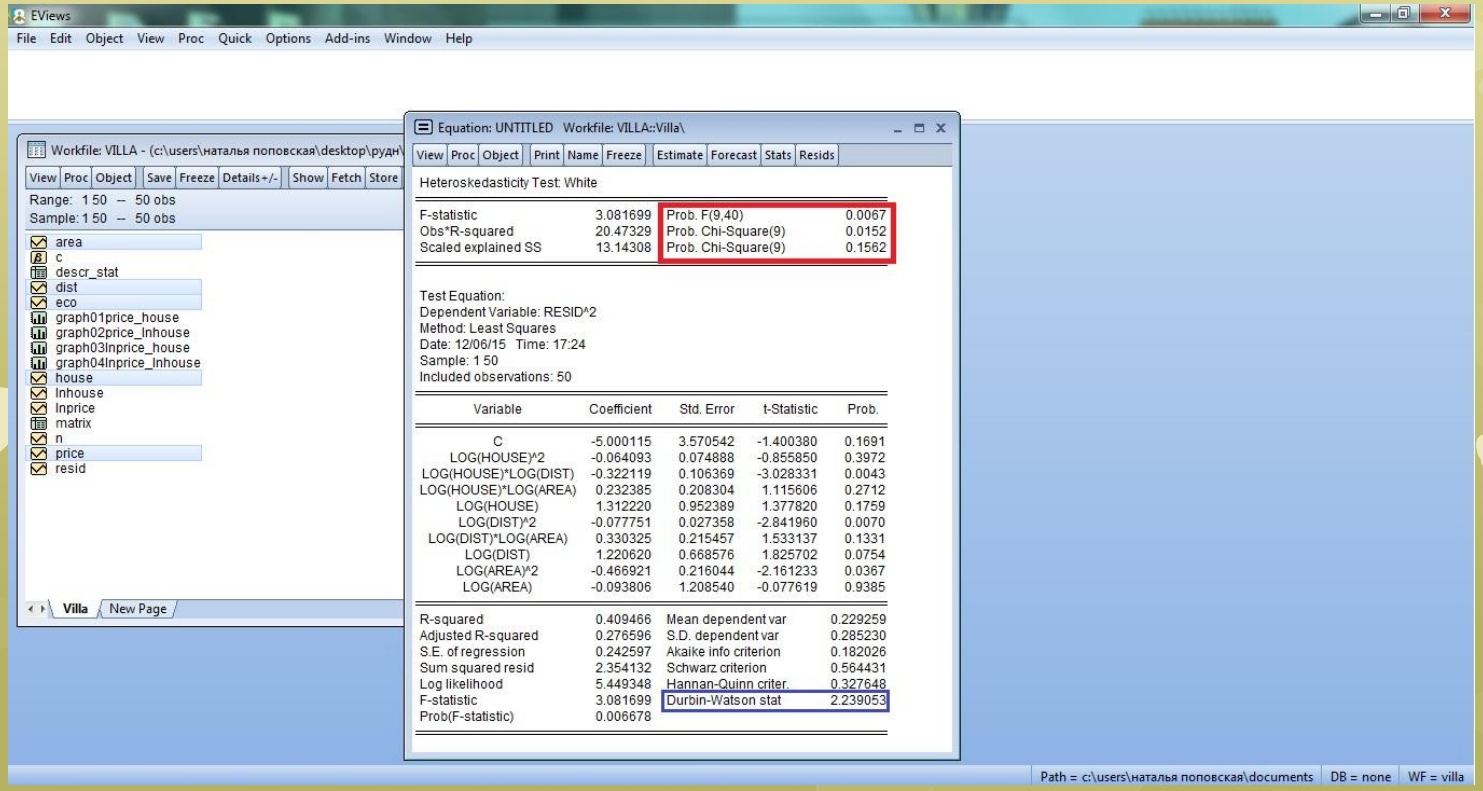
Проверка логарифмической модели на гетероскедастичность[2]

- Выбираем проверку по White.

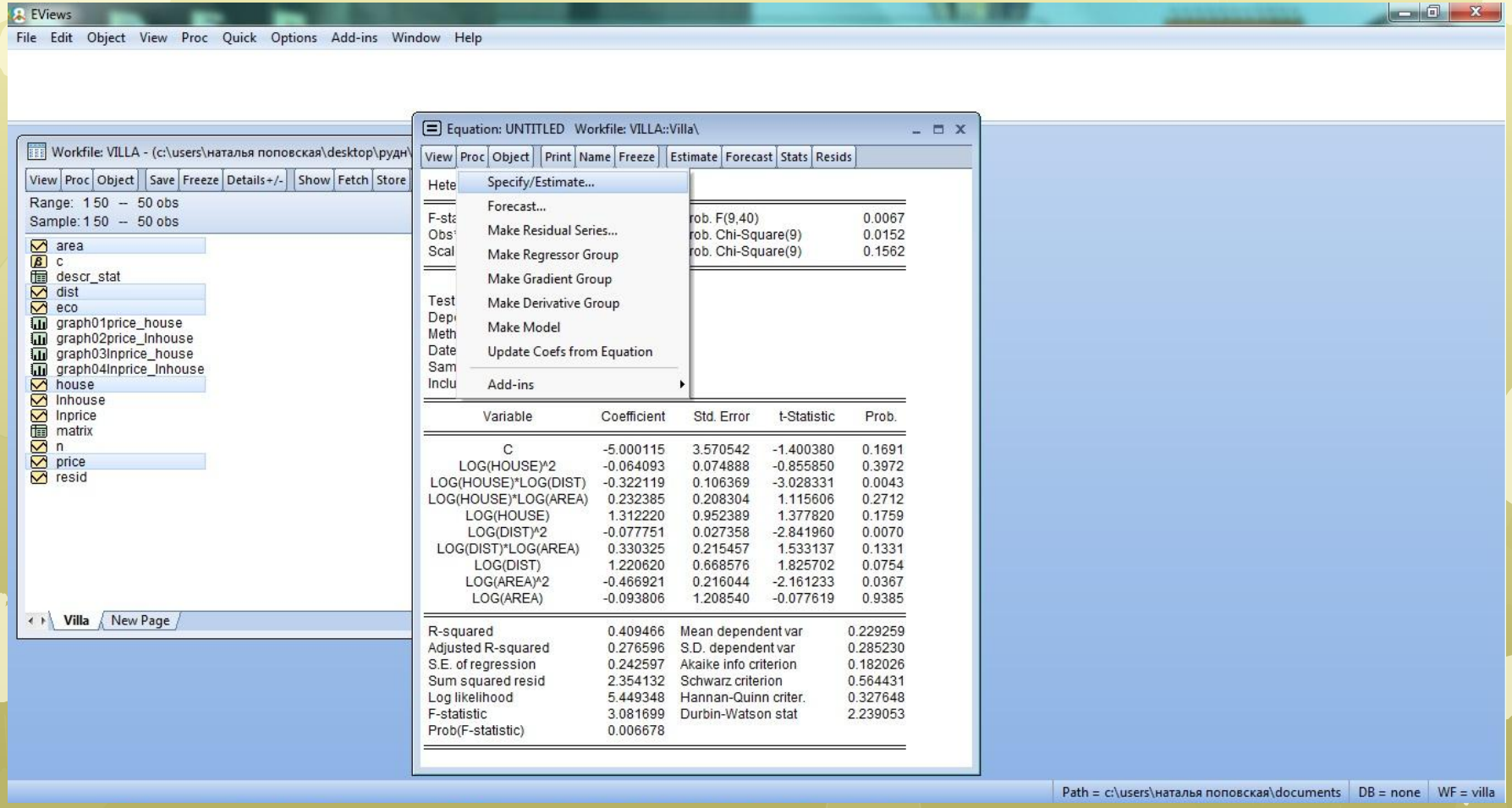


Проверка логарифмической модели на гетероскедастичность[3]

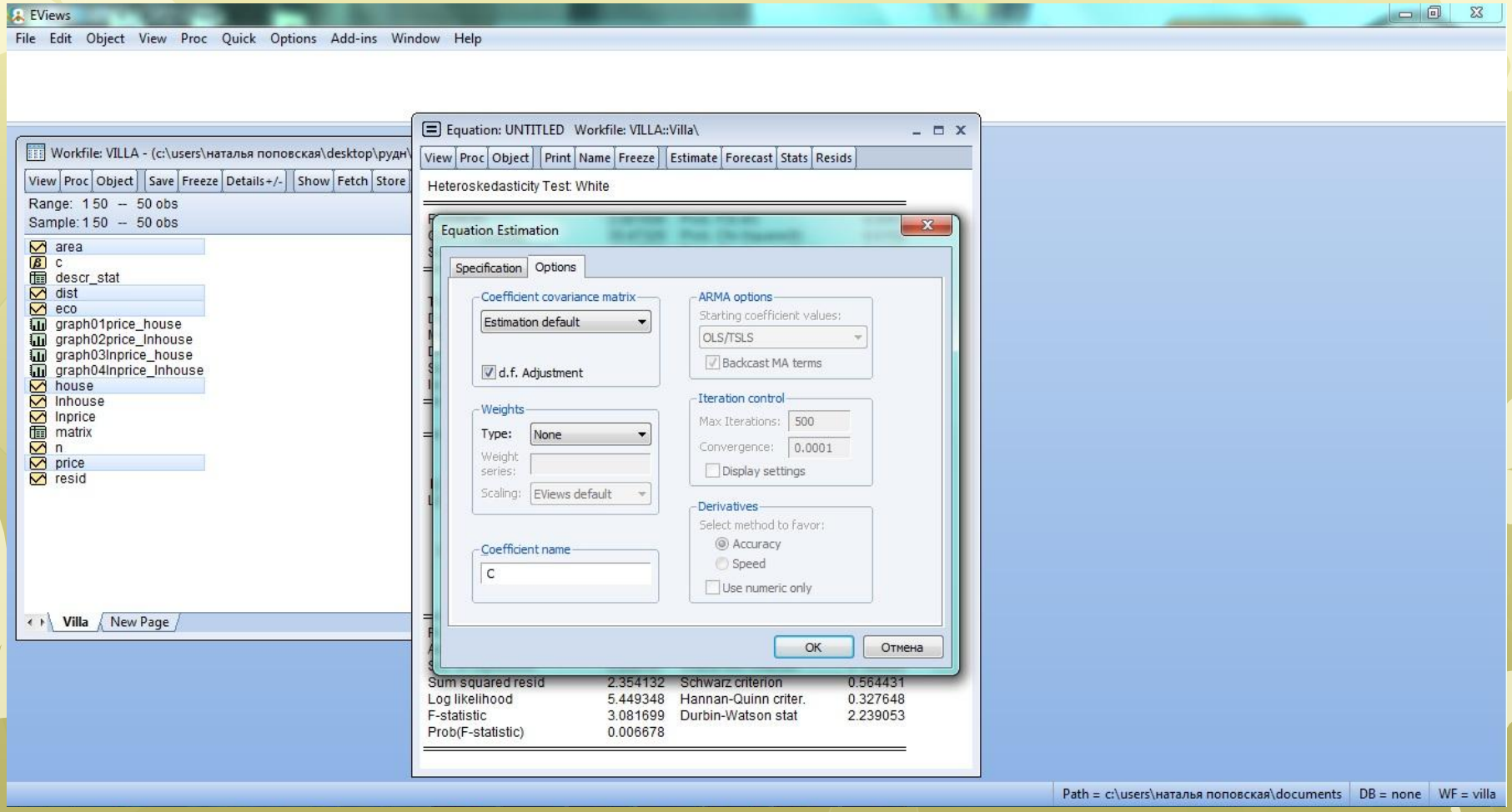
- Гетероскедастичность – непостоянство дисперсии остатков
- H0: Остатки гомоскедастичны, σ^2=Const
- H1: Остатки гетероскедастичны σ^2 ≠ Const.
- Присутствуют Prob.<0.05, значит принимает гипотезу H1 (гетероскедастичность есть), смотрим коэффициент Durbin-Watson, сравниваем с 1.5(2.239053>1.5)



Подправка [1]

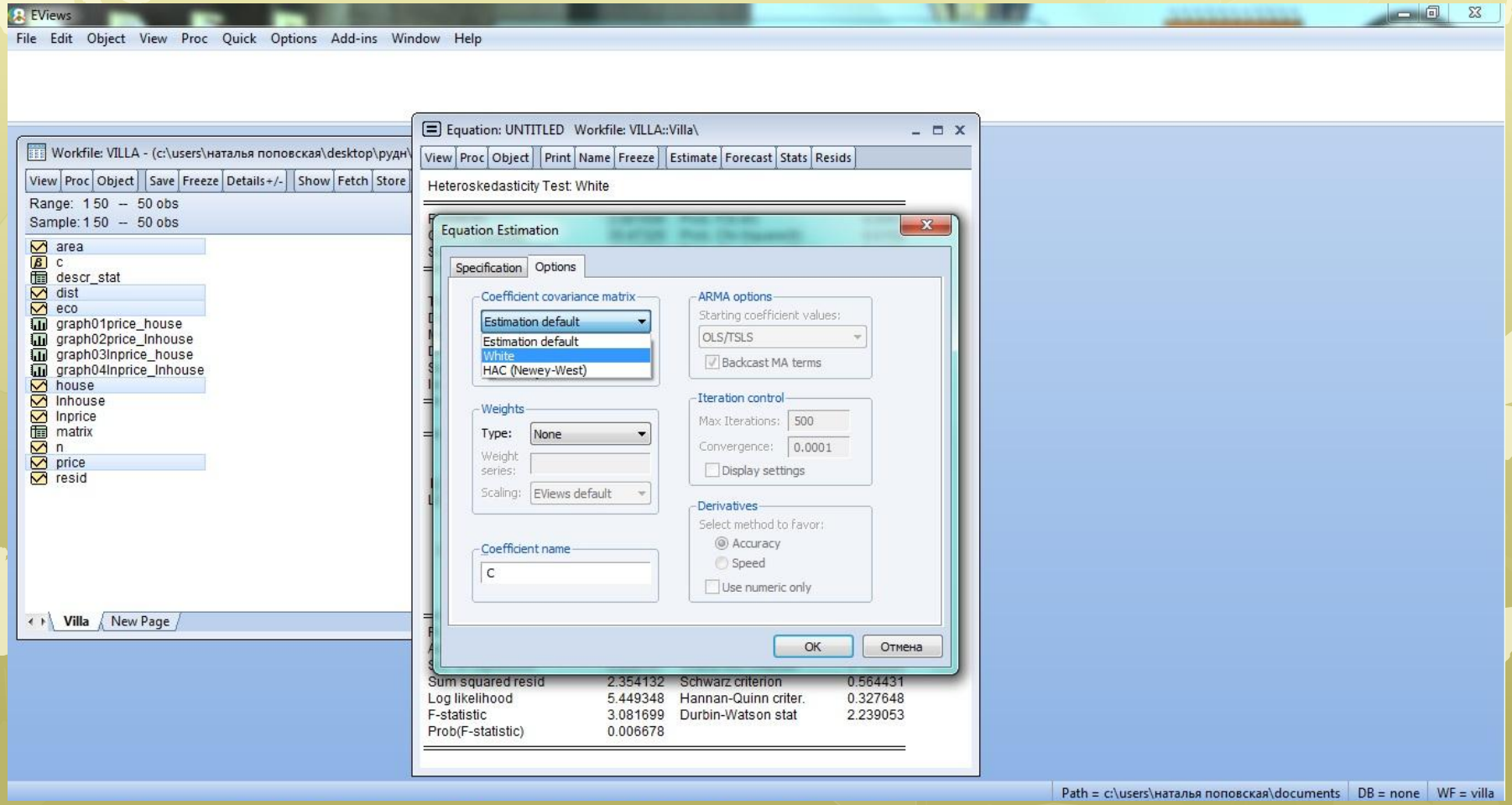

EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Workfile: VILLA - (c:\users\наталья поповская\desktop\рудн\
View Proc Object Save Freeze Details+/- Show Fetch Store
Range: 1 50 -- 50 obs
Sample: 1 50 -- 50 obs
area
c
descr_stat
dist
eco
graph01price_house
graph02price_lnhouse
graph03lnprice_house
graph04lnprice_lnhouse
house
lnhouse
lnprice
matrix
n
price
resid
Villa
New Page
Equation: UNTITLED Workfile: VILLA::Villa\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Specify/Estimate...
Forecast...
Make Residual Series...
Make Regressor Group
Make Gradient Group
Make Derivative Group
Make Model
Update Coefs from Equation
Add-ins
rob. F(9,40) 0.0067
rob. Chi-Square(9) 0.0152
rob. Chi-Square(9) 0.1562
Variable Coefficient Std. Error t-Statistic Prob.
C -5.000115 3.570542 -1.400380 0.1691
LOG(HOUSE)^2 -0.064093 0.074888 -0.855850 0.3972
LOG(HOUSE)*LOG(DIST) -0.322119 0.106369 -3.028331 0.0043
LOG(HOUSE)*LOG(AREA) 0.232385 0.208304 1.115606 0.2712
LOG(HOUSE) 1.312220 0.952389 1.377820 0.1759
LOG(DIST)^2 -0.077751 0.027358 -2.841960 0.0070
LOG(DIST)*LOG(AREA) 0.330325 0.215457 1.533137 0.1331
LOG(DIST) 1.220620 0.668576 1.825702 0.0754
LOG(AREA)^2 -0.466921 0.216044 -2.161233 0.0367
LOG(AREA) -0.093806 1.208540 -0.077619 0.9385
R-squared 0.409466 Mean dependent var 0.229259
Adjusted R-squared 0.276596 S.D. dependent var 0.285230
S.E. of regression 0.242597 Akaike info criterion 0.182026
Sum squared resid 2.354132 Schwarz criterion 0.564431
Log likelihood 5.449348 Hannan-Quinn criter. 0.327648
F-statistic 3.081699 Durbin-Watson stat 2.239053
Prob(F-statistic) 0.006678
Path = c:\users\наталья поповская\documents
DB = none
WF = villa

Подправка [2]


EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Workfile: VILLA - (c:\users\наталья поповская\desktop\рудн\
View Proc Object Save Freeze Details+/- Show Fetch Store
Range: 1 50 -- 50 obs
Sample: 1 50 -- 50 obs
area
c
descr_stat
dist
eco
graph01price_house
graph02price_lnhouse
graph03lnprice_house
graph04lnprice_lnhouse
house
lnhouse
lnprice
matrix
n
price
resid
Villa
New Page
Equation: UNTITLED Workfile: VILLA::Villa\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Heteroskedasticity Test: White
Equation Estimation
Specification Options
Coefficient covariance matrix
Estimation default
d.f. Adjustment
Weights
Type: None
Weight series:
Scaling: EViews default
Coefficient name
C
ARMA options
Starting coefficient values:
OLS/TSLS
Backcast MA terms
Iteration control
Max Iterations: 500
Convergence: 0.0001
Display settings
Derivatives
Select method to favor:
Accuracy
Speed
Use numeric only
OK
Отмена
Sum squared resid 2.354132 Schwarz criterion 0.564431
Log likelihood 5.449348 Hannan-Quinn criter. 0.327648
F-statistic 3.081699 Durbin-Watson stat 2.239053
Prob(F-statistic) 0.006678
Path = c:\users\наталья поповская\documents
DB = none
WF = villa

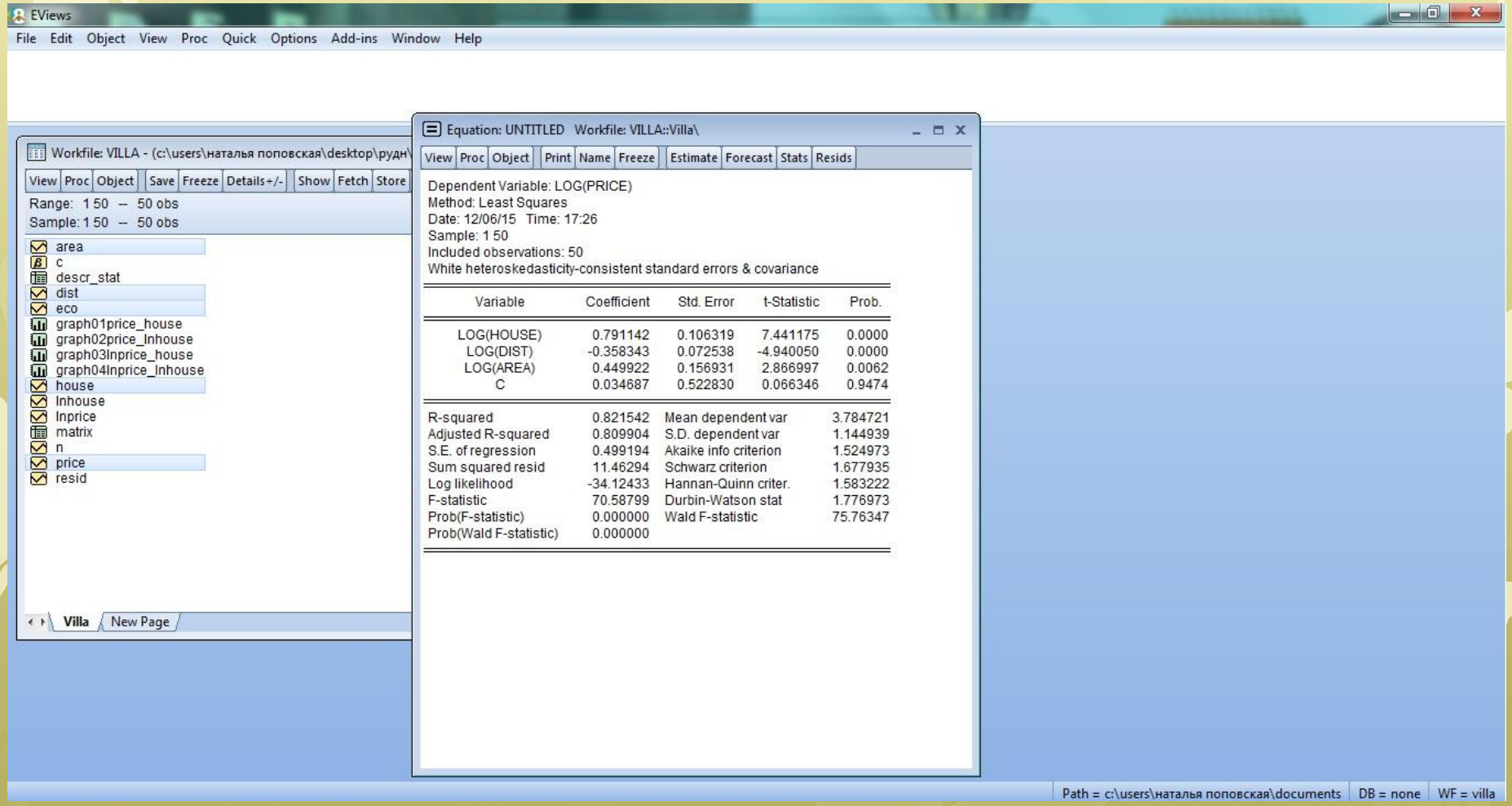
Подправка [3]

- Т.к. коэффициент Durbin-Watson>1.5, то берем подправку по White, в ином случае(D-W<1.5) – Newey-West.

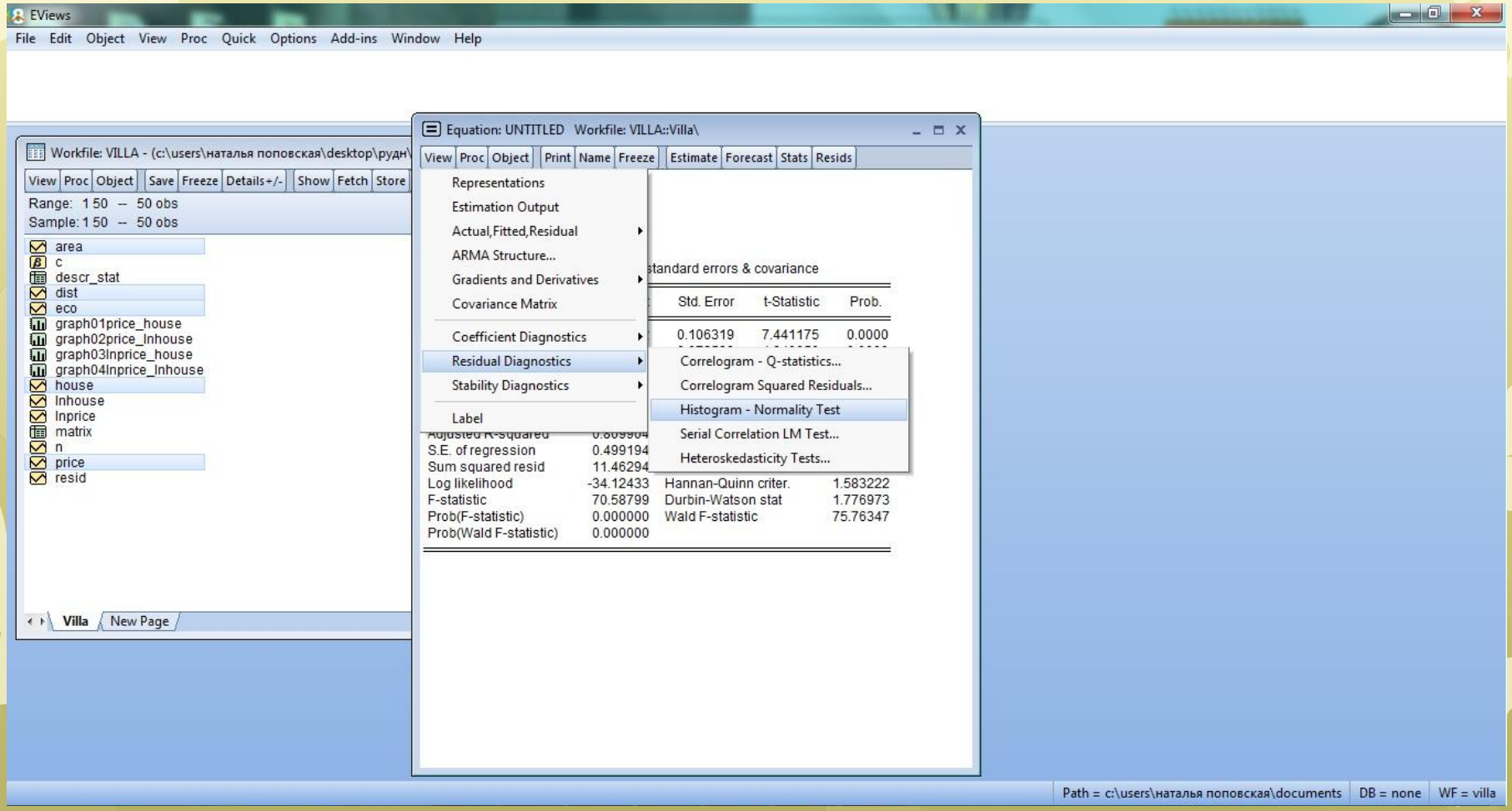


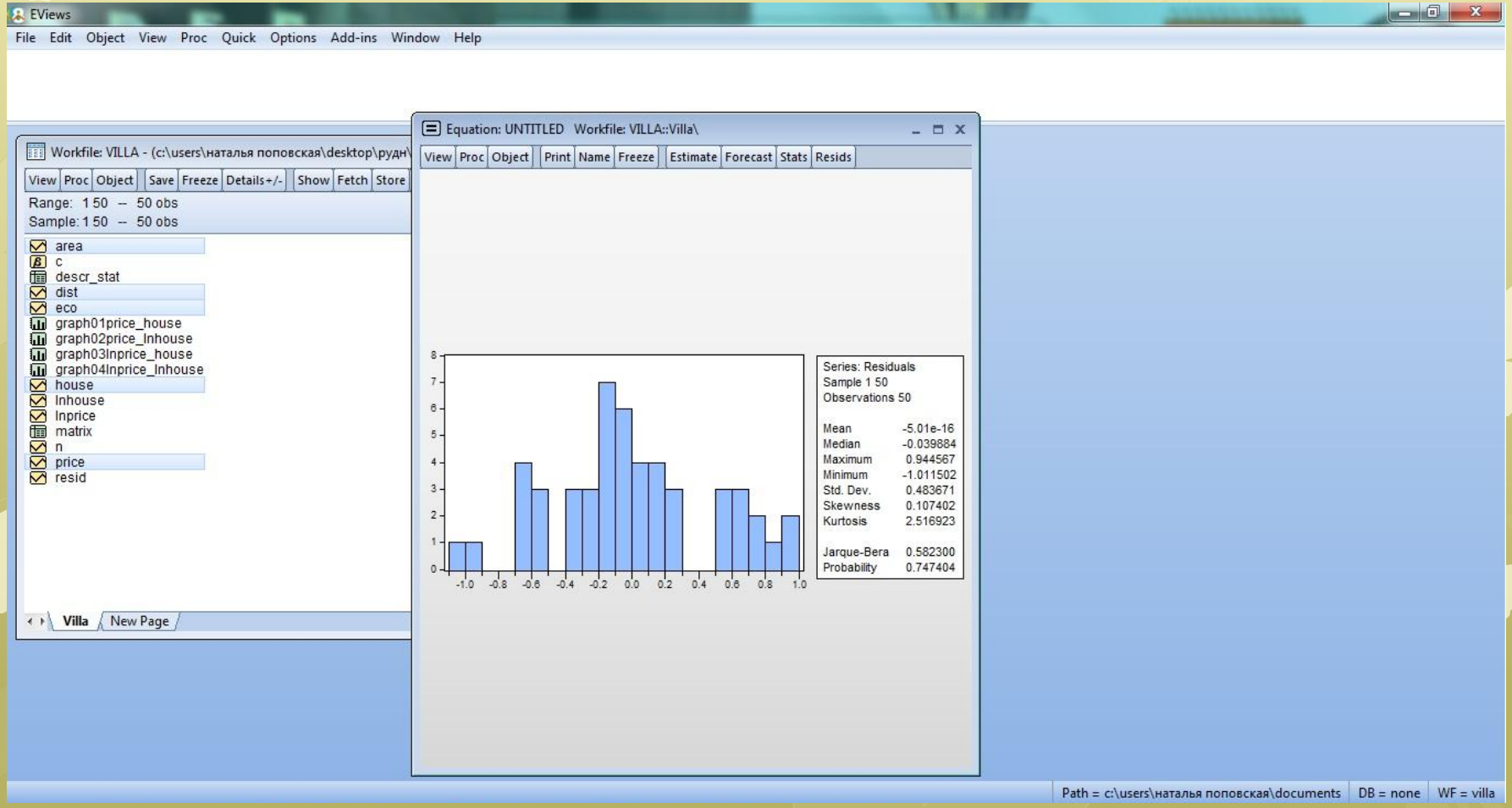
Подправка [4]

- Probability log(area) и log(dist) стали ближе к нулю, то есть стали лучше значимости коэффициентов.



Проверка на нормальность[1]



Проверка на нормальность[2]

- H0: нормальное распределение
- H1: ненормальное распределение (Prob<0.05)
- Probability > 0.05 -> распределение нормальное, Skewness близок к нулю, что хорошо.