Эконометрика

Задача 3.31

Подготовила: Власкина Анастасия

Постановка задачи

- В примере рассматриваются данные по стоимости квартир в Москве, собранные студентами первого курса РЭШ осенью 1997 г. Описание переменных представлено на слайде 3.

- Данные находятся в файле flat98s.xls

a) Постройте модель стоимости квартиры (или стоимости квадратного метра жилой площади квартиры) в зависимости от имеющихся факторов.

b) Проверьте гипотезу, что модели для 1,2,3,4-комнатных квартир различаются между собой, т.е. гипотезу, что рынок распадается на рынки однокомнатных, двухкомнатных и трех-четырехкомнатных квартир.

Описание переменных

Переменная	Описание
price	цена квартиры, тыс.долл.
rooms	количество жилых комнат
totsp	общая площадь, кв.м
livsp	жилая площадь, кв.м
kitsp	площадь кухни, кв.м
dist	расстояние до центра, км
metrdist	расстояние до ближайшей станции метро, мин
walk	1, если пешком от метро, 0 - иначе
brick	1, если дом кирпичный, 0 - иначе
tel	1, если есть телефон, 0 - иначе
bal	1, если есть балкон или лоджия, 0 - иначе
floor	0, если квартира находится на первом или последнем этаже, 1 - иначе

Решение(1/25)

Основываясь на том, что нам нужно ответить на вопрос о том распадается ли рынок на рынки однокомнатных, двухкомнатных и трех-четырехкомнатных квартир, введем бинарные переменные:

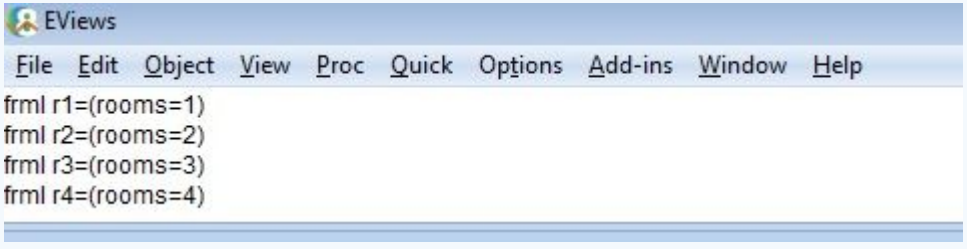
r1=(rooms=1) → однокомнатная квартира

r2=(rooms=2) → двухкомнатная квартира

r3=(rooms=3) → трехкомнатная квартира

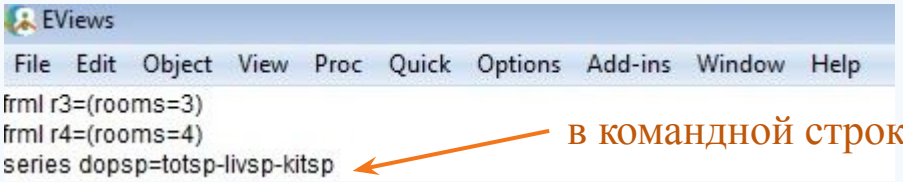
r4=(rooms=4) → четырехкомнатная квартира

в командной строке →



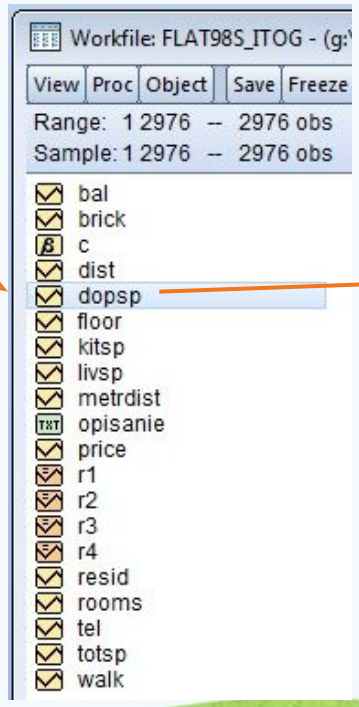
Решение(2/25)

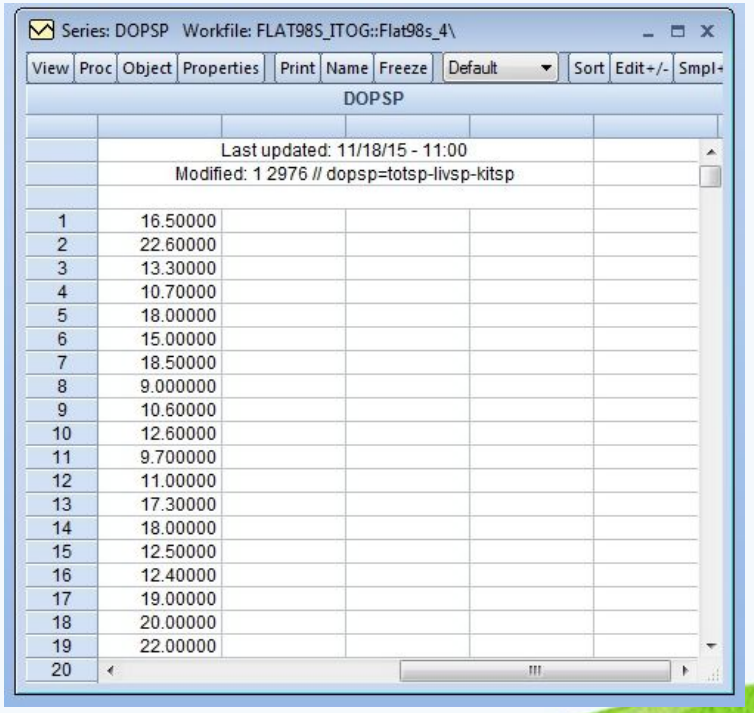
Добавим к уже имеющимся переменную *dopsp=totsp-livsp-kitsp*. Эта переменная имеет смысл площади дополнительных помещений (ванная, туалет, коридор и т.п.)



в командной строке

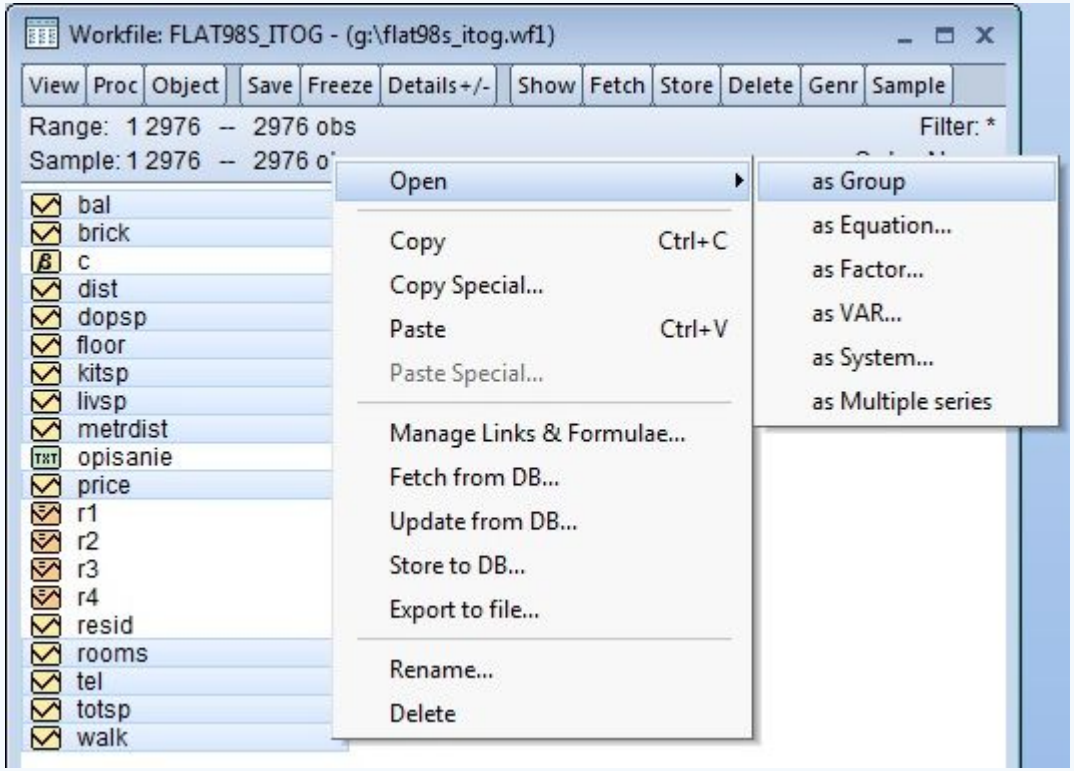
Создастся переменная





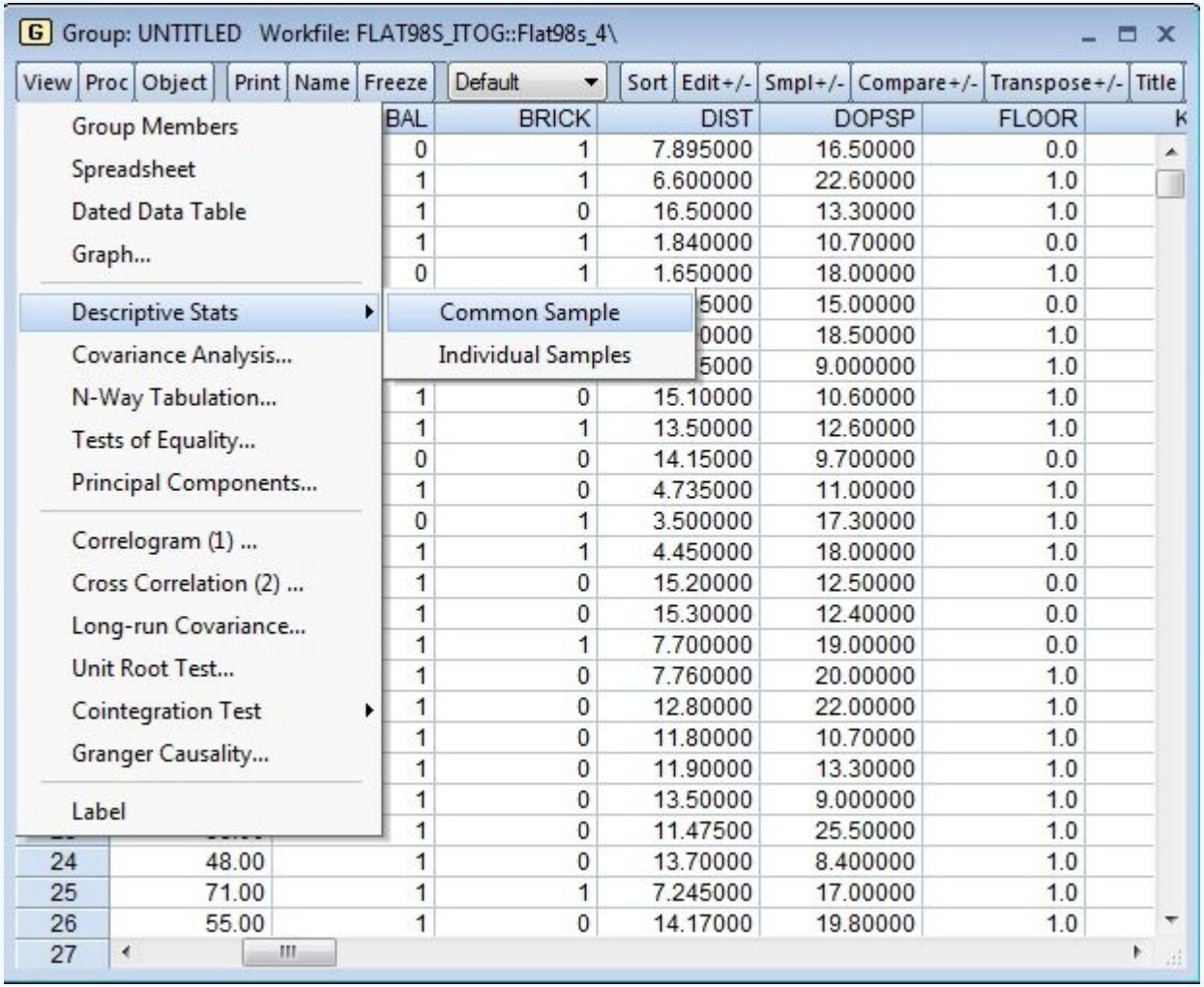
Решение(3/25)

Рассмотрим описательные статистики. Для этого удерживая клавишу Ctrl выделяем: *price, bal, brick, dist, dopsp, floor, kitsp, livsp, metrdist, rooms, tel, totsp, walk* > правой кнопкой мыши > *Open* > *as Group*.



Решение(4/25)

В открывшемся окне: *View > Descriptive Stats > Common Sample*



Решение(5/25)

В результате получим таблицу

Group: UNTITLED Workfile: FLAT98S_ITOG::Flat98s_4\

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

	PRICE	BAL	BRICK	DIST	DOPSP	FLOOR	KITSP	LIVSP	METRDIST	ROOMS	TEL	TOTSP	WALK
Mean	53.00318	0.718414	0.355511	9.392105	11.38555	0.699093	8.282359	29.93767	8.674933	1.786290	0.876680	49.60558	0.731519
Median	42.00000	1.000000	0.000000	9.600000	10.00000	1.000000	8.500000	22.00000	10.00000	1.000000	1.000000	40.00000	1.000000
Maximum	902.0000	1.000000	1.000000	21.87500	58.00000	3.500000	30.00000	119.0000	25.00000	4.000000	1.000000	180.0000	1.000000
Minimum	18.00000	0.000000	0.000000	0.400000	-6.100000	0.000000	0.000000	10.00000	0.000000	1.000000	0.000000	18.00000	0.000000
Std. Dev.	35.14978	0.449848	0.478748	4.501414	5.944704	0.461924	2.138460	14.23098	3.627481	0.915613	0.328859	19.78882	0.443244
Skewness	7.436734	-0.971220	0.603714	0.090998	1.715972	-0.778565	1.213983	1.297490	0.363777	0.744593	-2.291213	1.325726	-1.044832
Kurtosis	135.0242	1.943268	1.364471	2.086369	8.175067	2.164432	10.62610	4.646815	3.732162	2.301035	6.249659	4.869131	2.091675
Jarque-Bera	2188799.	606.3295	512.4720	107.6127	4781.384	387.2303	7942.506	1171.294	132.1091	335.5718	3913.306	1304.958	643.7774
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Sum	157737.5	2138.000	1058.000	27950.90	33883.40	2080.500	24648.30	89094.50	25816.60	5316.000	2609.000	147626.2	2177.000
Sum Sq. Dev.	3675633.	602.0309	681.8696	60281.61	105135.0	634.7876	13604.70	602499.7	39146.89	2494.081	321.7416	1165002.	584.4835
Observations	2976	2976	2976	2976	2976	2976	2976	2976	2976	2976	2976	2976	2976

Mean – среднее значение

Median – медиана

Maximum – максимум

Minimum – минимум

Std. Dev- стандартное отклонение

Observation – количество наблюдений

Skewness – ассиметрия

Observation – количество наблюдений

Решение(6/25)

При анализе получившейся статистики можно сделать выводы:

минимальное значение переменной kitsp равна нулю, а это значит, что существуют квартиры без кухни.

+

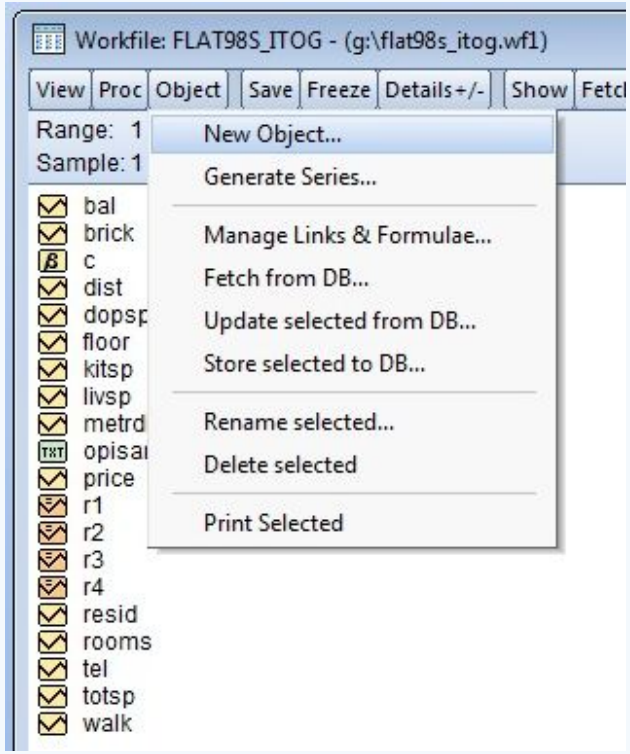
минимальное значение переменной dopsp отрицательное, чего не может быть в реальной жизни.

=

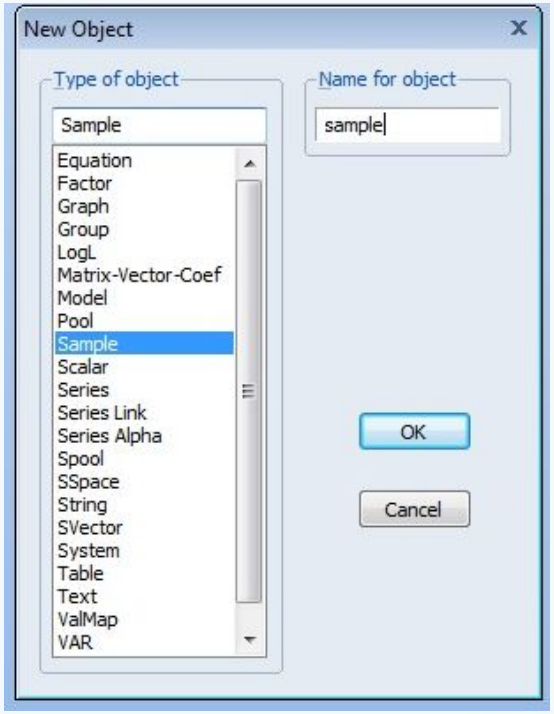
Исключим наблюдения, в которых площадь кухни меньше 5 кв. м, и в которых площадь дополнительных помещений меньше 0.

Решение(7/25)

Для того, чтобы исключить наблюдения, мы создадим переключатель: *Object > New Object*..



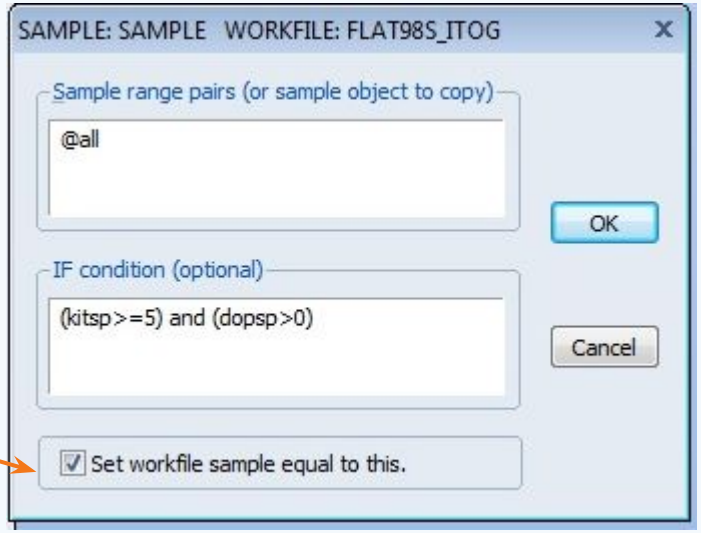
В открывшемся окне: Выберем *Type of object* равном значению *Sample*. И назовем переключатель в графе *Name for object* .



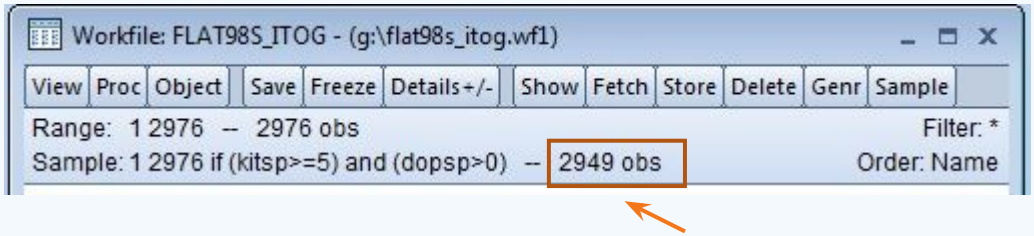
Решение(8/25)

В открывшемся окне: в графе *IF condition (optional)* введем наше условие: *(kitsp>=5) and (dopsp>0)*

и поставим галочку

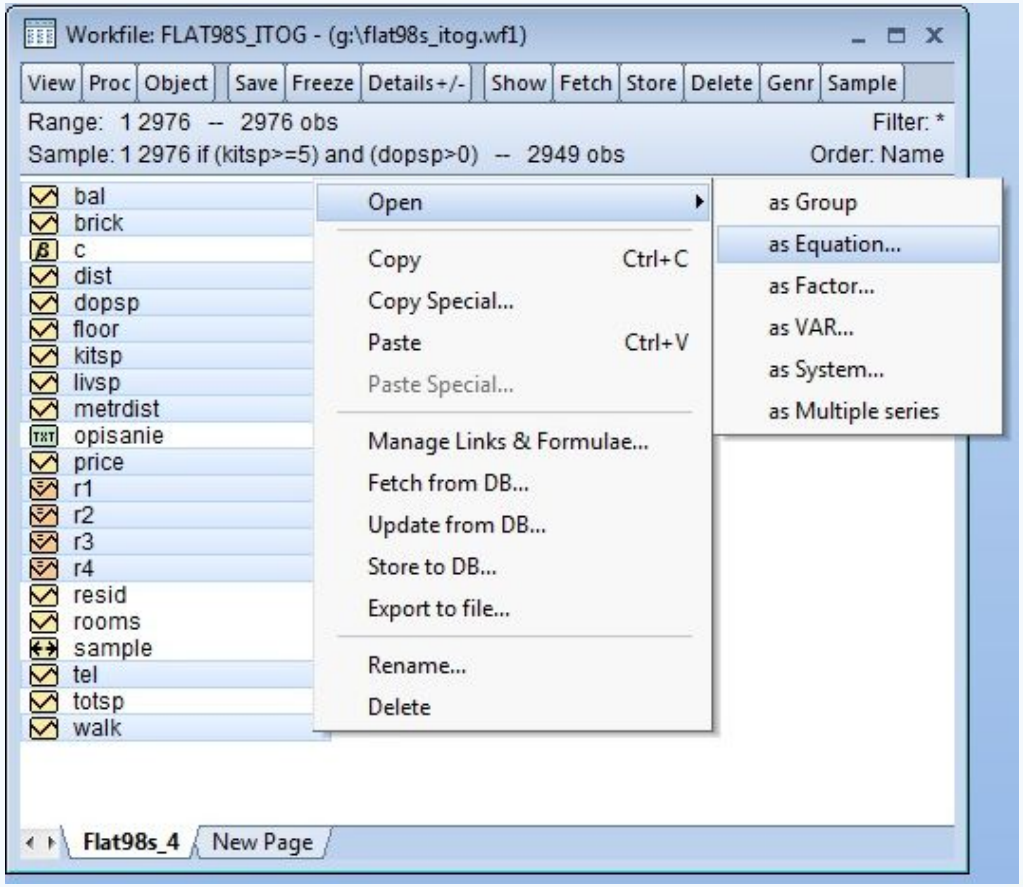


Таким образом, мы отбросили все неправдоподобные наблюдения и получили *2949* наблюдений, для которых определены все переменные.



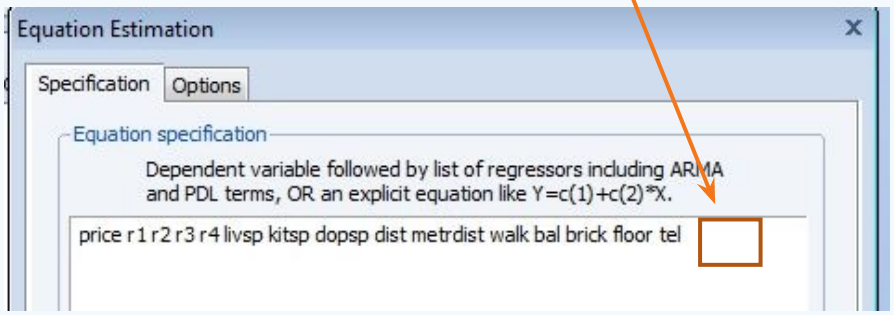
Решение(9/25)

а) Строим модель стоимости квартиры (модель 1). Для этого удерживая клавишу Ctrl выделяем: *price, r1, r2, r3, r4, livsp, kitsp, dopsp, dist, metrdist, walk, bal, brick, floor, tel* > правой кнопкой мыши > *Open* > *as Equation*..



Решение(10/25)

В открывшемся окне: т.к. мы используем в модели бинарные переменные r1, r2, r3 и r4, то константу (c) нужно убрать из модели



Equation: UNTITLED Workfile: FLAT98S_ITOG::Flat98s_4\

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

Dependent Variable: PRICE
Method: Least Squares
Date: 11/18/15 Time: 11:45
Sample: 1 2976 IF (KITSP>=5) AND (DOPSP>0)
Included observations: 2949

Variable	Coefficient	Std. Error	t-Statistic	Prob.
R1	-32.56215	2.992250	-10.88216	0.0000
R2	-46.46225	3.414595	-13.60696	0.0000
R3	-56.05606	4.410791	-12.70884	0.0000
R4	-70.93242	5.843748	-12.13817	0.0000
LIVSP	1.704005	0.086923	19.60358	0.0000
KITSP	2.455811	0.216543	11.34100	0.0000
DOPSP	1.930557	0.098571	19.58548	0.0000
DIST	-1.105346	0.096511	-11.45308	0.0000
METRDIST	-0.498152	0.109835	-4.535473	0.0000
WALK	4.144865	0.894992	4.631175	0.0000
BAL	3.362497	0.861062	3.905058	0.0001
BRICK	3.425905	0.918129	3.731397	0.0002
FLOOR	4.890328	0.831261	5.883025	0.0000
TEL	7.571699	1.192188	6.351094	0.0000

R-squared	0.660610	Mean dependent var	53.19951
Adjusted R-squared	0.659106	S.D. dependent var	35.23535
S.E. of regression	20.57254	Akaike info criterion	8.890528
Sum squared resid	1242179.	Schwarz criterion	8.918961
Log likelihood	-13095.08	Hannan-Quinn criter.	8.900764
Durbin-Watson stat	1.603057		

Все коэффициенты значимы (prob<0.05), т.е. выбрасывать переменные не нужно, модель качественная.

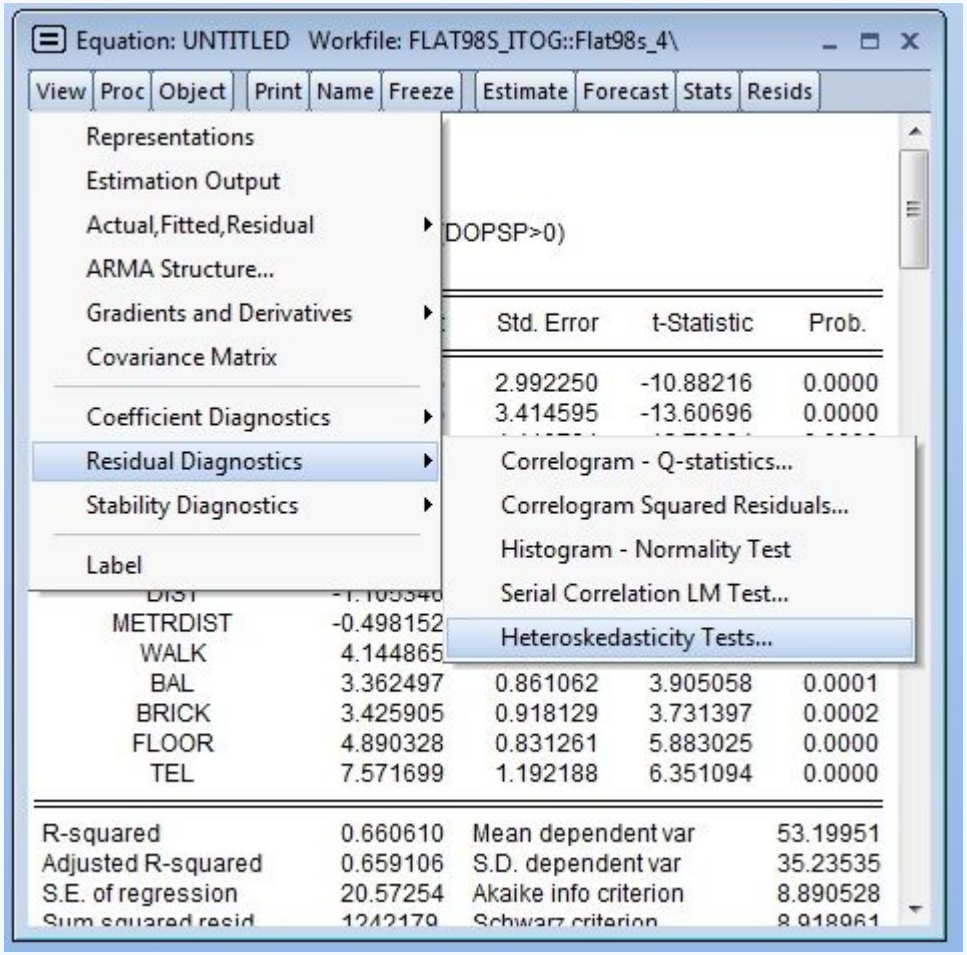
Решение(11/25)

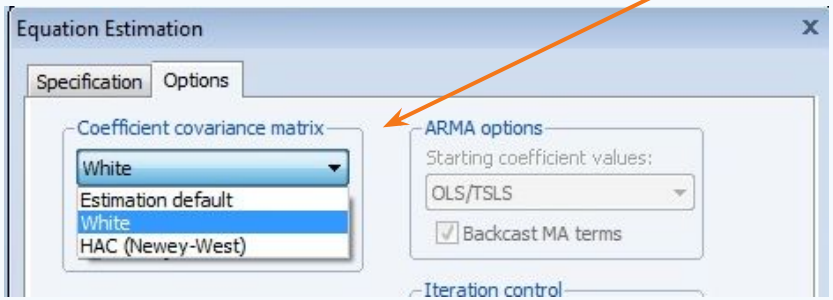
Проверим построенную модель на гетероскедастичность:

H0:гомоскедастичность

H1:гетероскедастичность

Для этого: *View > Residul Diagnostics > Heteroskedasticity Test (Test type: White)*





Решение(12/25)

В полученной таблице смотрим

- Если эти значения <0.05, то H1 верна (существует гетороскедастичность) - >нужна подправка
- Если эти значения >0.05, то H0 верна (существует гомоскедастичность) -> подправка не нужна.

Equation: UNTITLED Workfile: FLAT98S_ITOG::Flat98s_4\

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

Heteroskedasticity Test: White

F-statistic	82.95772	Prob. F(94,2854)	0.0000
Obs*R-squared	2158.874	Prob. Chi-Square(94)	0.0000
Scaled explained SS	315228.6	Prob. Chi-Square(94)	0.0000

<0.05

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/06/15 Time: 20:54
Sample: 1 2976 IF (KITSP>=5) AND (DOPSP>0)
Included observations: 2949
Collinear test regressors dropped from specification

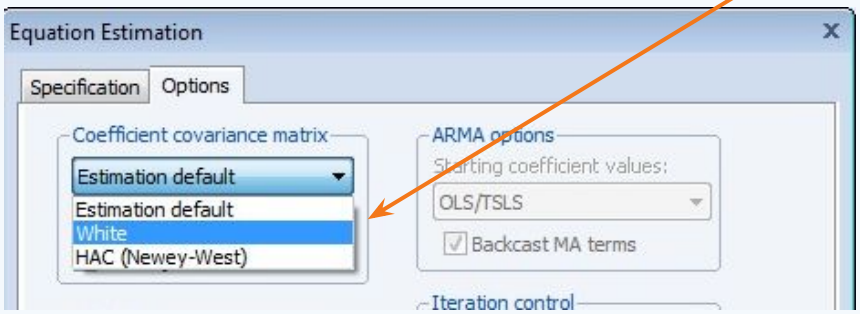
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	206107.6	8367.805	24.63103	0.0000
R1^2	-181024.7	6918.654	-26.16473	0.0000
R1*LIVSP	-1918.904	122.0724	-15.71939	0.0000
R1*KITSP	2074.253	320.8390	6.465090	0.0000
R1*DOPSP	-2207.245	135.7178	-16.26350	0.0000
R1*DIST	326.7382	164.1261	1.990776	0.0466
R1*METRDIST	-43.74676	184.2374	-0.237448	0.8123
R1*WALK	-332.6621	1403.102	-0.237090	0.8126
R1*BAL	-2278.874	1299.125	-1.754161	0.0795
R1*BRICK	5430.527	1339.216	4.055003	0.0001
R1*FLOOR	-4814.840	1366.459	-3.523588	0.0004
R1*TEL	-5835.900	1587.830	-3.675394	0.0002

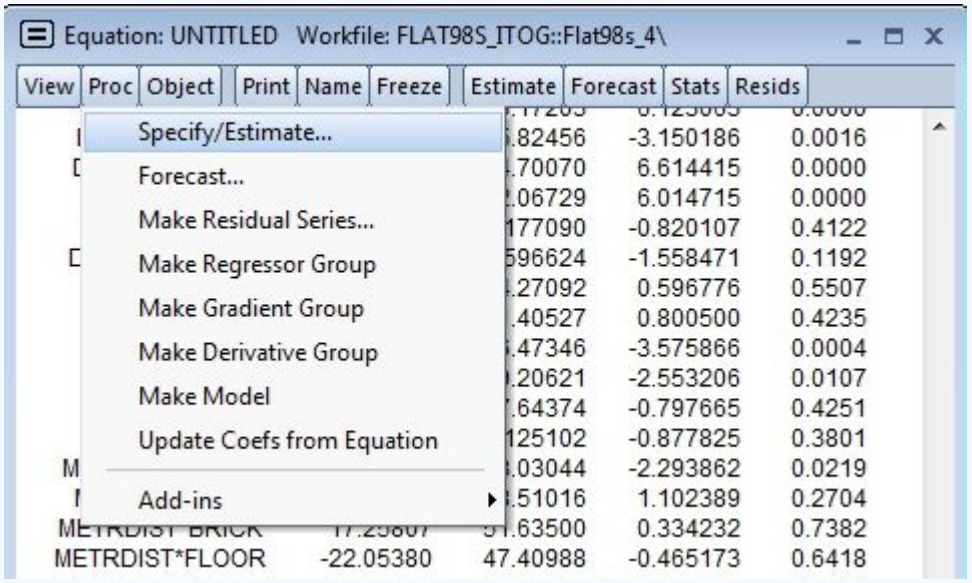
Решение(13/25)

Делаем подправку

Proc > Specify/Estimate… > Options (Coefficient covariance matrix: если Durbin-Watson <1.5, то Newey-West, если >=1.5, то White)

R-squared	0.732070	Mean dependent var	421.2203
Adjusted R-squared	0.723245	S.D. dependent var	7233.754
S.E. of regression	3805.497	Akaike info criterion	19.35796
Sum squared resid	4.13E+10	Schwarz criterion	19.55090
Log likelihood	-28448.32	Hannan-Quinn criter.	19.42742
F-statistic	82.95772	Durbin-Watson stat	1.796538
Prob(F-statistic)	0.000000		





Решение(12/25)

Таким образом, получаем подправленные значения

Все коэффициенты значимы (prob<0.05), т.е. выбрасывать переменные не нужно, модель качественная.

Equation: UNTITLED Workfile: FLAT98S_ITOG::Flat98s_4\

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

Dependent Variable: PRICE
Method: Least Squares
Date: 12/06/15 Time: 20:56
Sample: 1 2976 IF (KITSP>=5) AND (DOPSP>0)
Included observations: 2949
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
R1	-32.56215	9.316877	-3.494965	0.0005
R2	-46.46225	16.08248	-2.888998	0.0039
R3	-56.05606	23.87694	-2.347707	0.0190
R4	-70.93242	34.20673	-2.073639	0.0382
LIVSP	1.704005	0.488645	3.487205	0.0005
KITSP	2.455811	0.479998	5.116297	0.0000
DOPSP	1.930557	0.413019	4.674253	0.0000
DIST	-1.105346	0.102729	-10.75981	0.0000
METRDIST	-0.498152	0.091649	-5.435412	0.0000
WALK	4.144865	0.745548	5.559487	0.0000
BAL	3.362497	0.852881	3.942518	0.0001
BRICK	3.425905	1.585676	2.160533	0.0308
FLOOR	4.890328	0.645479	7.576277	0.0000
TEL	7.571699	1.443516	5.245317	0.0000

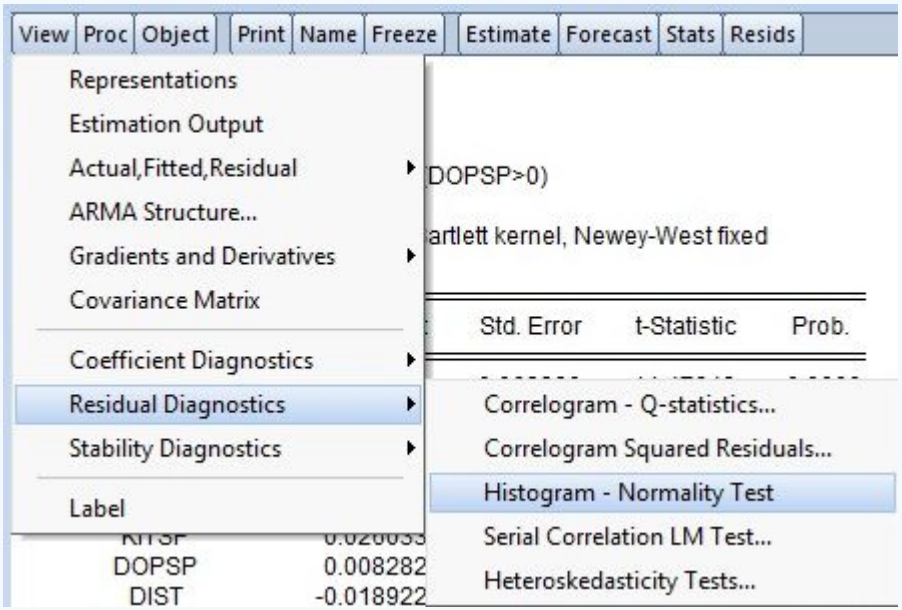
R-squared	0.660610	Mean dependent var	53.19951
Adjusted R-squared	0.659106	S.D. dependent var	35.23535
S.E. of regression	20.57254	Akaike info criterion	8.890528
Sum squared resid	1242179.	Schwarz criterion	8.918961
Log likelihood	-13095.08	Hannan-Quinn criter.	8.900764
Durbin-Watson stat	1.603057		

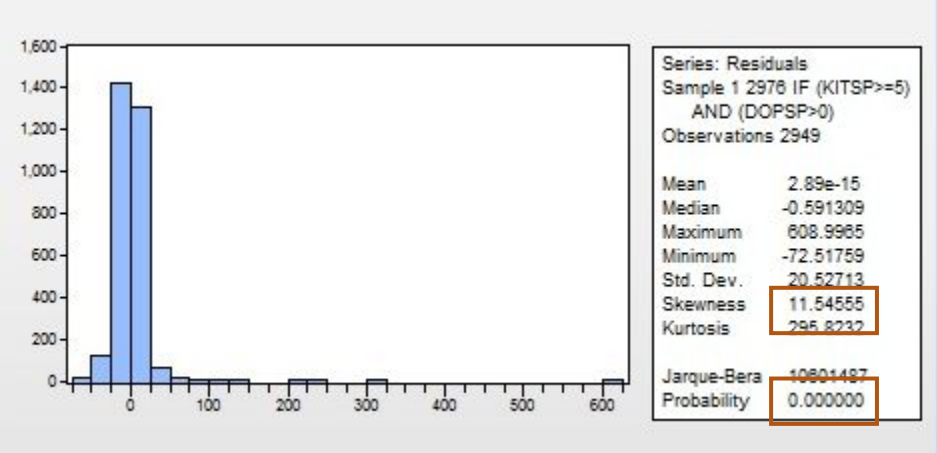
Решение(15/25)

Проверим построенную модель на нормальность. Для этого:

View > Residual Diagnostics > Histogram-Normality Test

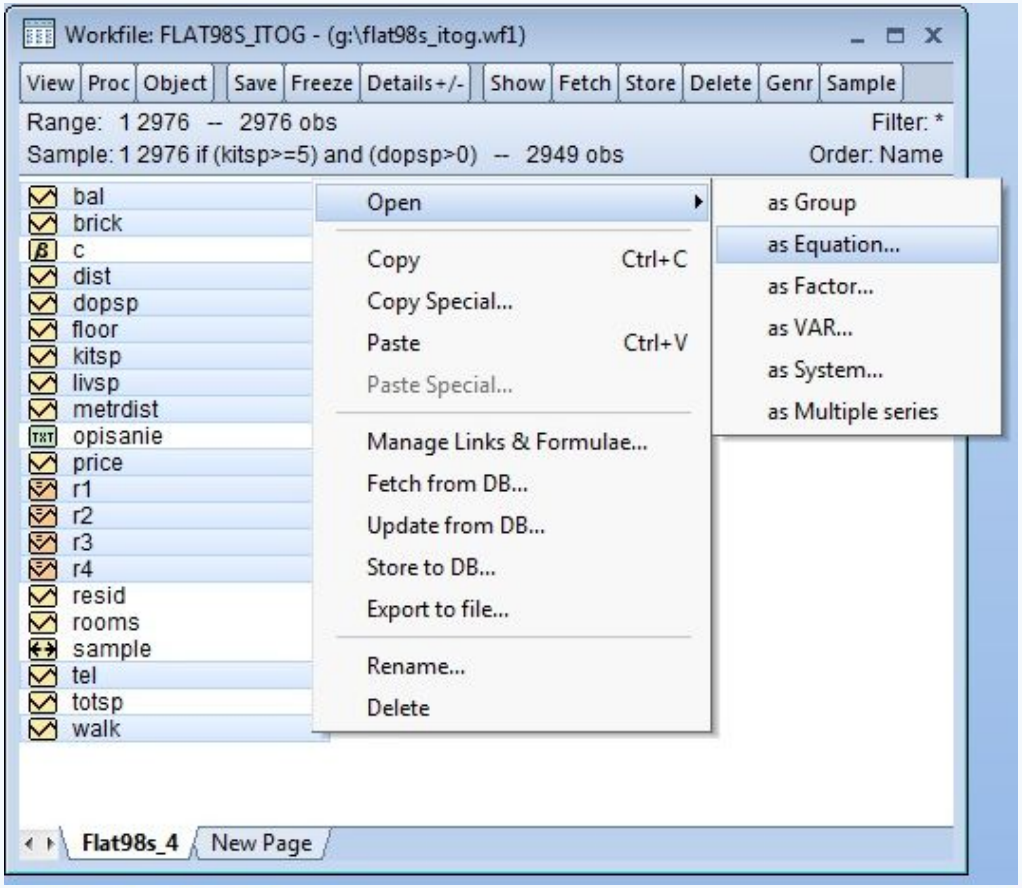
- если Probability<0.05, то не является нормальным
- Skewness(ассиметрия)=0, то является нормальным.





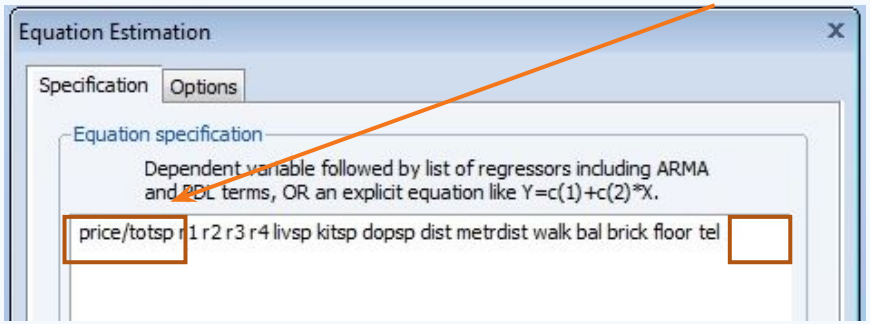
Решение(16/25)

Строим модель стоимости квадратного метра жилой площади квартиры (модель 2). Для этого удерживая клавишу Ctrl выделяем: *price, r1, r2, r3, r4, livsp, kitsp, dopsp, dist, metrdist, walk, bal, brick, floor, tel* > правой кнопкой мыши > *Open* > *as Equation*..



Решение(17/25)

В открывшемся окне: т.к. мы используем в модели бинарные переменные r1, r2, r3 и r4, то константу (c) нужно убрать из модели; переменную price меняем на price/totsp



Все коэффициенты значимы (кроме переменной *livsp*), т.е. выбрасывать переменные не нужно, модель качественная. Коэффициенты при *kitsp* и *dopsp* положительные, что означает, что квадратный метр кухни и комнаты стоит дороже, чем квадратный метр комнаты.

Equation: UNTITLED Workfile: FLAT98S_ITOG::Flat98s_4\

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

Dependent Variable: PRICE/TOTSP
Method: Least Squares
Date: 11/18/15 Time: 11:49
Sample: 1 2976 IF (KITSP>=5) AND (DOPSP>0)
Included observations: 2949

>0.05

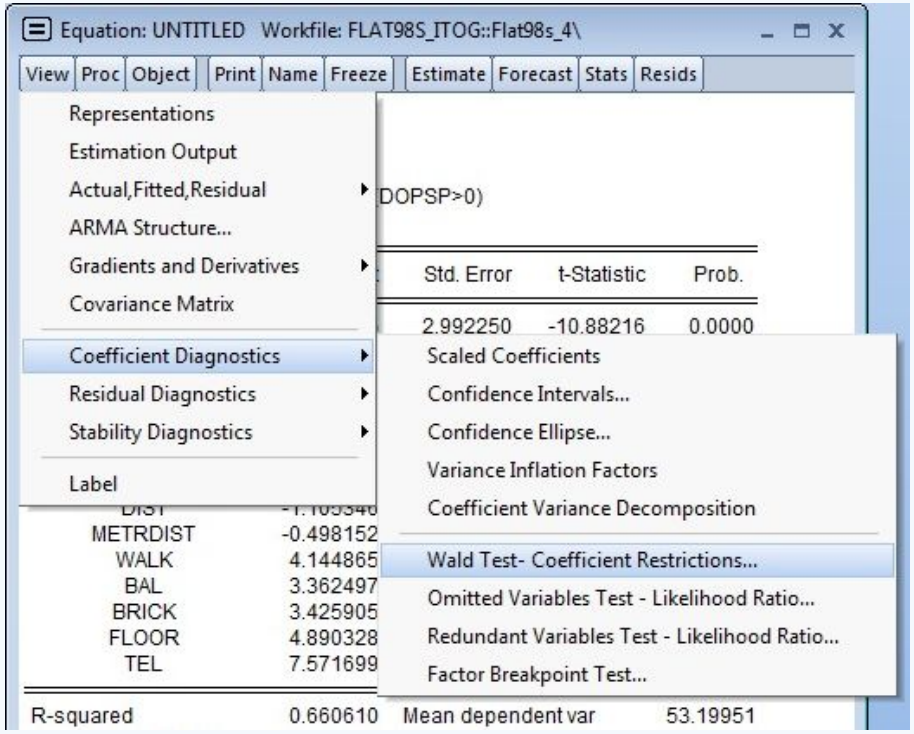
Variable	Coefficient	Std. Error	t-Statistic	Prob.
R1	0.727378	0.033171	21.92826	0.0000
R2	0.641006	0.037853	16.93418	0.0000
R3	0.641667	0.048896	13.12305	0.0000
R4	0.542133	0.064781	8.368654	0.0000
LIVSP	0.000970	0.000964	1.007106	0.3140
KITSP	0.026033	0.002401	10.84499	0.0000
DOPSP	0.008282	0.001093	7.579068	0.0000
DIST	-0.018922	0.001070	-17.68596	0.0000
METRDIST	-0.008171	0.001218	-6.710772	0.0000
WALK	0.079539	0.009922	8.016851	0.0000
BAL	0.056795	0.009545	5.950031	0.0000
BRICK	0.094680	0.010178	9.302409	0.0000
FLOOR	0.085444	0.009215	9.272228	0.0000
TEL	0.088048	0.013216	6.662187	0.0000

R-squared	0.344667	Mean dependent var	1.042347
Adjusted R-squared	0.341765	S.D. dependent var	0.281097
S.E. of regression	0.228059	Akaike info criterion	-0.113692
Sum squared resid	152.6514	Schwarz criterion	-0.085259
Log likelihood	181.6395	Hannan-Quinn criter.	-0.103456
Durbin-Watson stat	1.491978		

Решение(18/25)

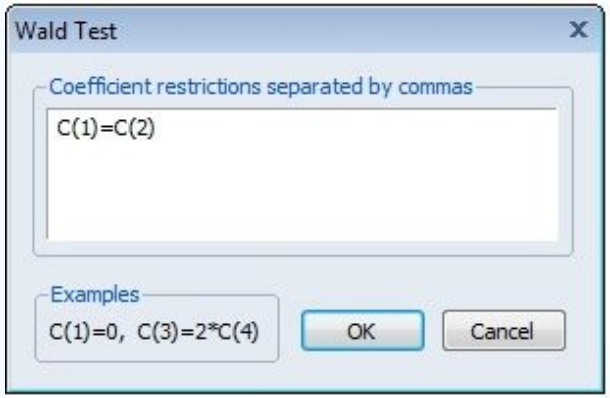
b) Проверим гипотезу, что модели для 1,2,3,4-комнатных квартир различаются между собой.

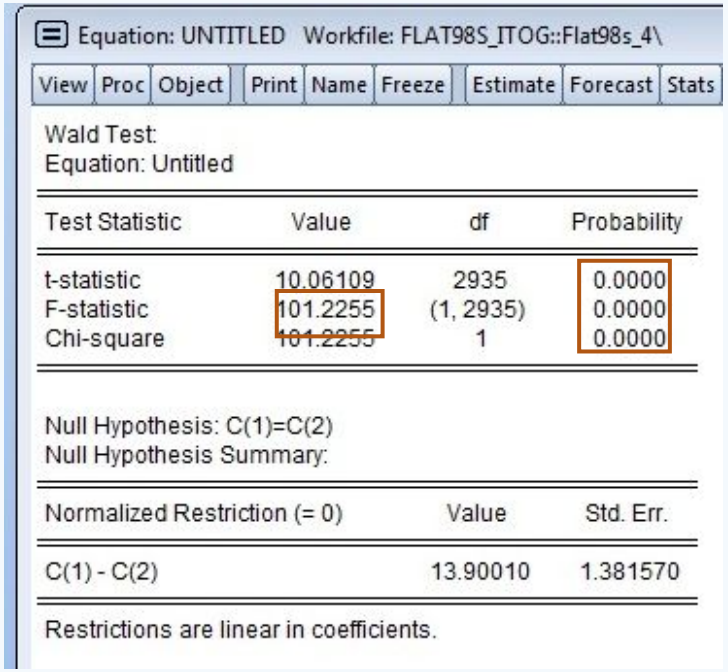
Сравним рынки попарно в модели 1 (на вид коэффициенты при r1, r2, r3, r4 разные). Для этого: *View > Coefficient Diagnostics > Wald Test – Coefficient Restrictions..*



Решение(19/25)

В открывшемся окне прописываем условия: C(1)=C(2), C(2)=C(3), C(1)=C(3), C(1)=C(4), C(2)=C(4), C(3)=C(4) {C(i),где i – порядковый номер нужного коэффициента в уравнении}





Такие же действия проделываем для модели 2

Решение(20/25)

Результаты модели 1 (модель стоимости квартиры)

H_0	F-статистика	p-значение
r1=r2	101,2255	0,0000
r2=r3	28,97807	0,0000
r1=r3	88,07842	0,0000
r1=r4	15,74625	0,0001
r2=r4	6,655218	0,0099
r3=r4	12,64181	0,0004

Результаты модели 2 (модель стоимости квадратного метра жилой площади квартиры)

H_0	F-статистика	p-значение
r1=r2	31,80476	0,0000
r2=r3	0,001120	0,9733 >0.05
r1=r3	9.539405	0,0020
r1=r4	83,02069	0,0000
r2=r4	50,09535	0,0000
r3=r4	34,70355	0,0000

Итоги:

Как видим, модель 1 разбивает рынок на четыре категории по числу комнат в квартире. Модель 2 разбивает рынок только на три группы (однокомнатные квартиры, четырехкомнатные и остальные квартиры)

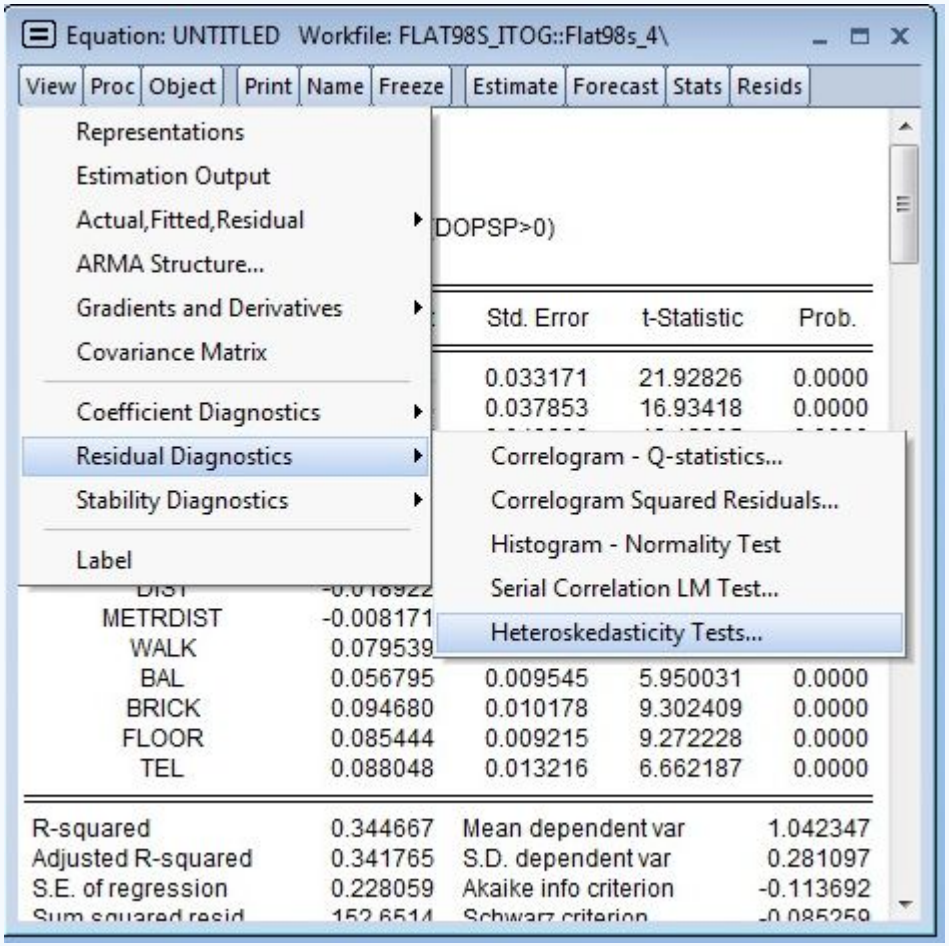
Решение(21/25)

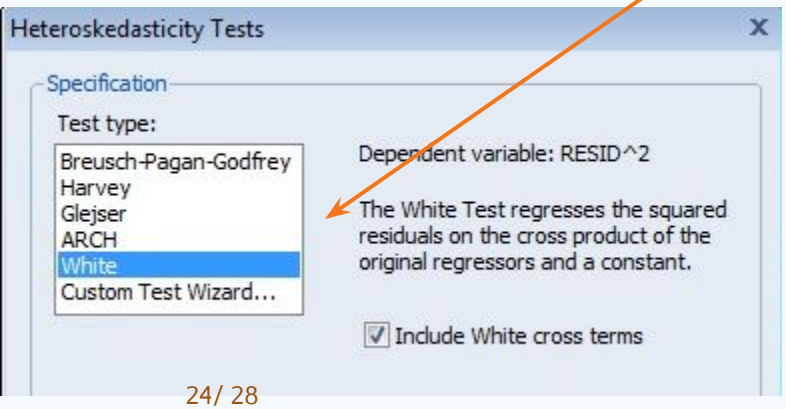
Проверим построенную модель на гетероскедастичность:

H0:гомоскедастичность

H1:гетероскедастичность

Для этого: *View > Residual Diagnostics > Heteroskedasticity Tests (Test type: White)*





Решение(22/25)

В полученной таблице смотрим

- Если эти значения <0.05, то H1 верна (существует гетороскедастичность) ->нужна подправка
- Если эти значения >0.05, то H0 верна (существует гомоскедастичность) -> подправка не нужна.

Equation: UNTITLED Workfile: FLAT98S_ITOG::Flat98s_4\

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

Heteroskedasticity Test: White

F-statistic	23.64561	Prob. F(94,2854)	0.0000
Obs*R-squared	1291.138	Prob. Chi-Square(94)	0.0000
Scaled explained SS	20242.64	Prob. Chi-Square(94)	0.0000

<0.05

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/06/15 Time: 21:02
Sample: 1 2976 IF (KITSP>=5) AND (DOPSP>0)
Included observations: 2949
Collinear test regressors dropped from specification

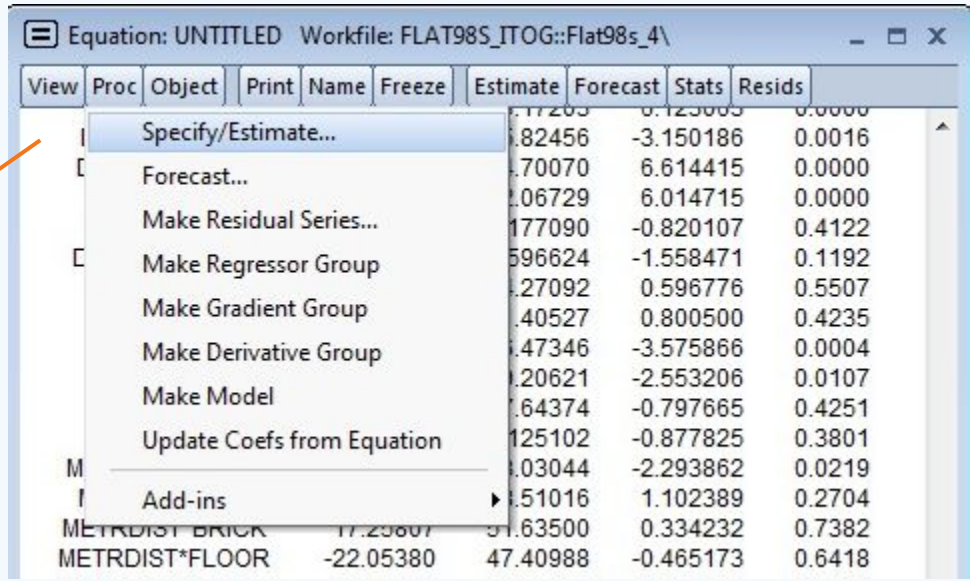
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.482789	0.488093	11.23309	0.0000
R1^2	-4.928436	0.403564	-12.21228	0.0000
R1*LIVSP	-0.054425	0.007120	-7.643507	0.0000
R1*KITSP	0.082257	0.018714	4.395366	0.0000

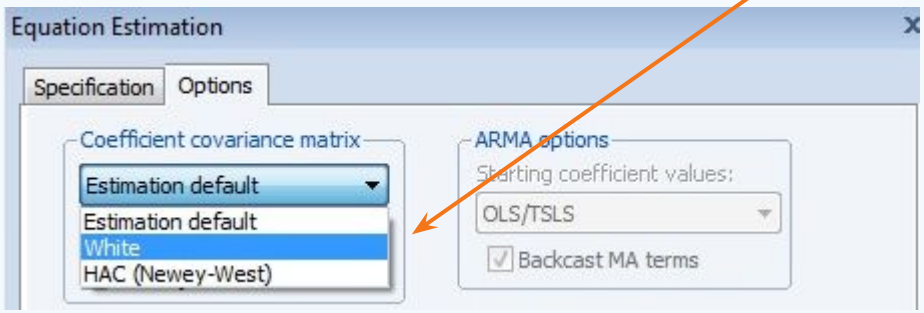
Решение(23/25)

Делаем подправку

Proc > Specify/Estimate… > Options (Coefficient covariance matrix: если Durbin-Watson <1.5, то Newey-West, если >=1.5, то White)

R-squared	0.437822	Mean dependent var	0.051764
Adjusted R-squared	0.419306	S.D. dependent var	0.291292
S.E. of regression	0.221974	Akaike info criterion	-0.140828
Sum squared resid	140.6237	Schwarz criterion	0.052110
Log likelihood	302.6513	Hannan-Quinn criter.	-0.071368
F-statistic	23.64561	Durbin-Watson stat	1.937363
Prob(F-statistic)	0.000000		





Решение(24/25)

Таким образом, получаем подправленные значения

Все коэффициенты значимы (prob<0.05) (кроме переменной *livsp*), т.е. выбрасывать переменные не нужно, модель качественная.

Equation: UNTITLED Workfile: FLAT98S_ITOG::Flat98s_4\

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

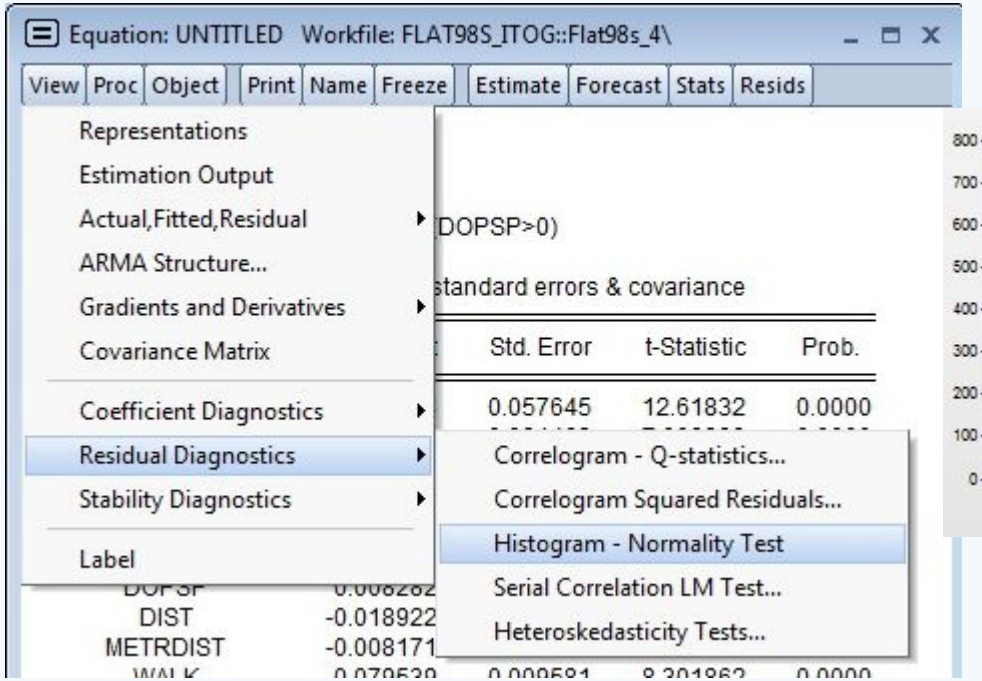
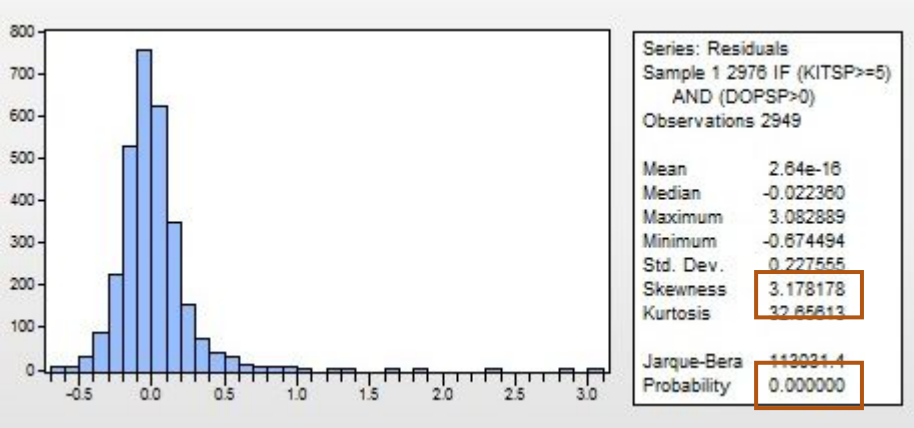
Dependent Variable: PRICE/TOTSP
Method: Least Squares
Date: 12/06/15 Time: 21:36
Sample: 1 2976 IF (KITSP>=5) AND (DOPSP>0)
Included observations: 2949
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
R1	0.727378	0.057645	12.61832	0.0000
R2	0.641006	0.091463	7.008388	0.0000
R3	0.641667	0.134183	4.782015	0.0000
R4	0.542133	0.190853	2.840573	0.0045
LIVSP	0.000970	0.002800	0.346557	0.7289
KITSP	0.026033	0.003411	7.631724	0.0000
DOPSP	0.008282	0.002669	3.102390	0.0019
DIST	-0.018922	0.001203	-15.73326	0.0000
METRDIST	-0.008171	0.001121	-7.291571	0.0000
WALK	0.079539	0.009581	8.301862	0.0000
BAL	0.056795	0.009445	6.013471	0.0000
BRICK	0.094680	0.012307	7.692920	0.0000
FLOOR	0.085444	0.008431	10.13407	0.0000
TEL	0.088048	0.015316	5.748835	0.0000

R-squared	0.344667	Mean dependent var	1.042347
Adjusted R-squared	0.341765	S.D. dependent var	0.281097
S.E. of regression	0.228059	Akaike info criterion	-0.113692
Sum squared resid	152.6514	Schwarz criterion	-0.085259
Log likelihood	181.6395	Hannan-Quinn criter.	-0.103456
Durbin-Watson stat	1.491978		

Решение(25/25)

Проверим построенную модель на нормальность. Для этого:

View > Residal Diagnostic > Histogram-Normality Test

- если Probability<0.05, то не является нормальным
- Skewness(ассиметрия)=0, то является нормальным.