

Экзамен. Часть 2. Работа в статистических пакетах.

Фамилия Имя Отчество __Татаринов_Никита_Алексеевич_____

Задание. За каждый пункт 3 балла. Итоговый балл за эту часть 21 балл.

В файле **housing.xls** содержатся данные по рынку недвижимости города Виндзор, Канада:

Переменная	Описание переменной
lnprice	Логарифм цены продажи дома, тыс. долларов США
lotsize	Общая площадь дома, квадратные футы
bedrooms	Число спален в доме
bathrms	Число ванных комнат в доме
stories	Количество этаже в доме (не включая подвал).
garagepl	Число мест в гараже
driveway	Dummy-переменная, принимающая значение 1, если у дома есть выделенная подъездная дорога, и 0 – иначе.
recroom	Dummy-переменная, принимающая значение 1, если в доме есть комната отдыха, и 0 – иначе.
fullbase	Dummy-переменная, принимающая значение 1, если в доме есть оборудованное подвальное помещение, и 0 иначе.
gashw	Dummy-переменная, принимающая значение 1, если в доме используется система подогрева воды на газе, и 0 – иначе.
airco	Dummy-переменная, принимающая значение 1, если в доме имеется система центрального кондиционирования, и 0 – иначе.
prefarea	Dummy-переменная, принимающая значение 1, если дом располагается в элитном районе, и 0 – иначе.

Расчеты можно делать в Excel, Gretl или иной программе. Результаты всех тестов должны быть проинтерпретированы (отвергается или не отвергается гипотеза, на каком уровне значимости, при какой альтернативной, какой делается вывод).

В качестве ответа на вопросы скопируйте результаты соответствующих тестов из статистических пакетов и вставьте в этот документ после самих вопросов.

Обращаем внимание, что наборы данных индивидуальные.

Pdf документ с решением необходимо приложить к ответу на вопрос в SmartLMS

- 1) Основной целью исследователя была оценка следующей исходной модели ценообразования:

$$\ln price = \beta_1 + \beta_2 \cdot lotsize + \beta_3 \cdot bedrooms + \beta_4 \cdot bathrms + \beta_5 \cdot stories + \beta_6 \cdot garagepl + \beta_7 \cdot driveway + \beta_8 \cdot recroom + \beta_9 \cdot fullbase + \beta_{10} \cdot gashw + \beta_{11} \cdot airco + \varepsilon$$

Интерпретируйте результаты. Выпишите уравнение оцененной модели.

Ответ:

$$\hat{\ln price} = 10.0 + 4.24e-05 \cdot lotsize + 0.0249 \cdot bedrooms + 0.180 \cdot bathrms + 0.0859 \cdot stories +$$

(0.0991) (5.09e-06) (0.0188) (0.0239) (0.0131)

$$+ 0.0436 \cdot garagepl + 0.287 \cdot driveway + 0.0293 \cdot recroom + 0.131 \cdot fullbase + 0.101 \cdot gashw +$$

(0.0129) (0.0843) (0.0292) (0.0264) (0.0631)

$$+ 0.183 \cdot airco$$

(0.0246)

Model 1: OLS, using observations 1-296				
Dependent variable: lnprice				
	coefficient	std. error	t-ratio	p-value
-----	-----	-----	-----	-----
const	10.0239	0.0991164	101.1	2.51e-225 ***
lotsize	4.24057e-05	5.08859e-06	8.333	3.36e-15 ***
bedrooms	0.0248892	0.0187698	1.326	0.1859
bathrms	0.180495	0.0238772	7.559	5.58e-13 ***
stories	0.0859377	0.0131301	6.545	2.76e-10 ***
garagepl	0.0435899	0.0129032	3.378	0.0008 ***
driveway	0.286829	0.0843088	3.402	0.0008 ***
recroom	0.0292578	0.0292182	1.001	0.3175
fullbase	0.131370	0.0263969	4.977	1.12e-06 ***
gashw	0.101139	0.0631136	1.602	0.1102
airco	0.182804	0.0245679	7.441	1.19e-12 ***
Mean dependent var	11.20456	S.D. dependent var	0.322312	
Sum squared resid	9.739819	S.E. of regression	0.184864	
R-squared	0.682183	Adjusted R-squared	0.671032	
F(10, 285)	61.17434	P-value(F)	4.64e-65	
Log-likelihood	85.28646	Akaike criterion	-148.5729	
Schwarz criterion	-107.9790	Hannan-Quinn	-132.3199	

Модель является полулогарифмической.

Для каждого j-го фактора в такой модели (отбросив дамми переменные) при увеличении этого фактора x_j на 1 единицу объясняемая переменная y изменится на $(e^{\beta_j} - 1) \cdot 100\%$.

Дамми переменные в данной модели только на константу, то есть они только смещают уравнение модели (не изменяют наклон).

2) Выполните тест на адекватность этой модели и сделайте выводы.

Ответ: значение F-статистики равно 61.17434, p-value равно 4.64e-65 – значит, модель является адекватной на любом разумном уровне значимости (нулевая гипотеза теста на адекватность – равенство всех переменных 0, что не выполняется, так как незначимы 3 из 10 переменных).

3) Проведите тест Рамсея с одним вспомогательным регрессором (только с квадратом). Сделайте вывод.

Auxiliary regression for RESET specification test
OLS, using observations 1-296
Dependent variable: lnprice

	coefficient	std. error	t-ratio	p-value
const	18.4396	13.6116	1.355	0.1766
lotsize	0.000123162	0.000130712	0.9422	0.3469
bedrooms	0.0704859	0.0761026	0.9262	0.3551
bathrms	0.530232	0.566156	0.9365	0.3498
stories	0.250155	0.265924	0.9407	0.3477
garagepl	0.127141	0.135748	0.9366	0.3498
driveway	0.823316	0.871788	0.9444	0.3458
recroom	0.0839271	0.0931324	0.9012	0.3683
fullbase	0.378647	0.400808	0.9447	0.3456
gashw	0.292187	0.315386	0.9264	0.3550
airco	0.529967	0.562026	0.9430	0.3465
yhat^2	-0.0848401	0.137217	-0.6183	0.5369

Test statistic: F = 0.382283,
with p-value = P(F(1,284) > 0.382283) = 0.537

Ответ: p-value равно 0.537, то есть нулевая гипотеза не отвергается на любом разумном уровне значимости – значит, спецификация модели является правильной (то есть нет необходимости добавлять в модель квадраты факторов).

4) Протестируйте наличие мультиколлинеарности двумя способами. Сделайте вывод.

Variance Inflation Factors
Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

lotsize	1.199
bedrooms	1.392
bathrms	1.333
stories	1.532
garagepl	1.153
driveway	1.022
recroom	1.239
fullbase	1.381
gashw	1.017
airco	1.242

VIF(j) = $1/(1 - R(j)^2)$, where R(j) is the multiple correlation coefficient between variable j and the other independent variables

```

Correlation Coefficients, using the observations 1 - 296
5% critical value (two-tailed) = 0.1140 for n = 296

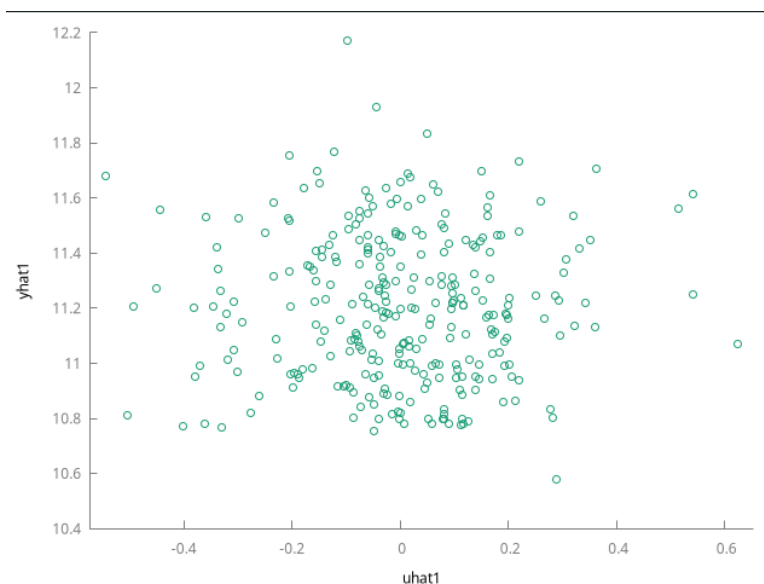
      lotsize      bedrooms      bathrms      stories      garagepl
1.0000      0.1642      0.1663      -0.0066      0.3375 lotsize
      1.0000      0.4205      0.3608      0.1354 bedrooms
      1.0000      0.3425      0.0974      0.0974 bathrms
      1.0000      0.0122      0.0122 stories
      1.0000      1.0000      1.0000 garagepl

      driveway      recroom      fullbase      gashw      airco
0.0546      0.0820      0.0334      0.0542      0.1646 lotsize
-0.0382      0.0952      0.1218      -0.0009      0.2218 bedrooms
-0.0165      0.0783      0.0664      -0.0004      0.2607 bathrms
0.0693      0.0492      -0.2527      -0.0470      0.3584 stories
0.0299      -0.0061      -0.0271      -0.0052      0.1437 garagepl
1.0000      0.0682      -0.0124      0.0232      -0.0031 driveway
      1.0000      0.3908      0.0041      0.1105 recroom
      1.0000      0.0332      -0.0695 fullbase
      1.0000      -0.1008 gashw
      1.0000 airco

```

Ответ: мультиколлинеарность не наблюдается (так как нет значений VIF больше 10 и нет близких к 1 значений в корреляционной матрице).

5) Постройте график «остатки-прогнозы». Сделайте вывод.



Ответ:

6) Протестируйте наличие гетероскедастичности одним из изучаемых в рамках курса тестов. Сделайте вывод.

```

Unadjusted R-squared = 0.251404

Test statistic: TR^2 = 74.415584,
with p-value = P(Chi-square(55) > 74.415584) = 0.041685

```

Ответ: проблема гетероскедастичности присутствует (p-value равно 0.041658, то есть на 5% и 10% уровнях значимости нулевая гипотеза о гомоскедастичности отвергается).

- 7) Прodelайте тест Чoу для выбора между моделями ценообразования для элитных и иных районов. Сделайте вывод.

```
Chow test for structural difference with respect to prefarea  
F(11, 274) = 3.15338 with p-value 0.0005
```

Ответ: p-value равно 0.0005, то есть нулевая гипотеза отвергается – значит, две выборки не принадлежат к одной генеральной совокупности (лучше сделать 2 модели или 1 модель с дамми переменными).