

Неопределенность оценки рыночной стоимости, полученной по модели множественной регрессии

В работе [1] авторами была показана возможность применения известного в литературе соотношения для расчета доверительного интервала при оценке рыночной стоимости в условиях реальных распределений цен, заметно отличающихся от нормального. Там же получена модификация этого соотношения, позволяющая за счет учета конечного объема сделок на рынке¹ уменьшить ширину интервала неопределенности при оценке рыночной стоимости гомогенных товаров.

Естественно обобщить этот подход на оценку стоимости гетерогенных товаров, получаемые построением регрессионных моделей². В этом может помочь рекомендация Ф.Йетса³, согласно которой «формула для средней ошибки оценки средней или какой-либо другой оценки требует видоизменения, а именно введения множителя $\sqrt{1 - \frac{n}{N}}$ », в котором отношение объемов выборки n и генеральной совокупности N называют долей отбора.

При построении границ доверительного интервала для предсказанного моделью множественной регрессии среднего значения⁴ зависимой переменной \tilde{y}_0 , обычно используется формула в матричном выражении⁵:

$$\tilde{y}_0 \pm t_\alpha s \sqrt{x_0^T (X^T X)^{-1} x_0}, \quad (1)$$

где \tilde{y}_0 — предсказанное среднее значение зависимой переменной, s — остаточное среднеквадратическое отклонение регрессии, t_α — квантиль распределения Стьюдента уровня α , x_0 — вектор признак-факторов⁶ объекта оценки, X — матрица, у которой в первом столбце стоят единицы, а остальные столбцы составлены из признак-факторов объектов-аналогов.

¹ Совокупность всех представленных на рынке цен на данный товар рассматривается как генеральная совокупность цен.

² Модель среднего значения можно рассматривать как вырожденный случай множественной регрессии - при нулевом количестве объясняющих факторов.

³ [2], стр. 192

⁴ Среднее значение цен на данный товар на рассматриваемом рынке считается его рыночной стоимостью.

⁵ См., например, [3], стр. 119

⁶ Признак-факторами будем называть параметры объектов, учитываемые регрессионной моделью как влияющие факторы.

Для парной (однофакторной) регрессии выражение (1) сводится к виду

$$\tilde{y}_0 \pm t_\alpha s \sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right)} \quad (1')$$

Важно отметить, что выражения (1) и (1') получены в предположении неограниченно большого объема N нормально распределенной генеральной совокупности и пренебрежимо малой доле отбора $\frac{n}{N}$.

Следуя рекомендациям Йетса, используем формулу для интервала неопределенности регрессионного среднего, учитывающую существенную долю отбора и конечный объем генеральной совокупности (сделок на рынке)⁷:

$$\tilde{y}_0 \pm t_\alpha s \sqrt{1 - \frac{n}{N}} \sqrt{x_0^T (X^T X)^{-1} x_0} \quad (2)$$

Прибегнем к имитационному моделированию: возьмем с рынка N объектов-аналогов⁸ и примем их за «генеральную совокупность». Один из объектов выберем в качестве объекта оценки. По этой совокупности построим адекватную регрессионную модель и по ней получим «истинную» оценку стоимости для оцениваемого объекта.

Затем из «генеральной совокупности» случайным образом выберем n аналогов и определим по этой выборке параметры регрессионной модели такого же вида⁹, что и модель по генсовокупности. Затем по формулам (1) и (2) построим доверительные интервалы одного уровня $\alpha = 0,05$ для оценки стоимости и проверим, попадает ли в них «истинное» значение оценки¹⁰.

Проделав эту процедуру достаточно большое число раз, можно рассчитать частоту попадания «истинной» оценки стоимости в доверительный интервал, построенный по выборке из этой генеральной совокупности.

Применим этот алгоритм сначала для модели парной регрессии. Рассмотрим выборку из 25 цен на переносные электростанции различной мощности марки EUROPOWER фирмы Honda¹¹:

⁷ В дальнейшем, для единообразия с [1], будем называть интервал, рассчитанный по этой формуле, интервалом Стьюдента-Кокрена

⁸ Товаров с известными параметрами и ценами.

⁹ С таким же видом зависимостей моделируемой величины от влияющих факторов

¹⁰ Истинной считаем оценку, которую дает для оцениваемого объекта корректное уравнение регрессии, построенное по всей генеральной совокупности (в данном случае - по всем N аналогам)

¹¹ Цены взяты с сайта Яндекс-Маркет, см. табл. П1 Приложения

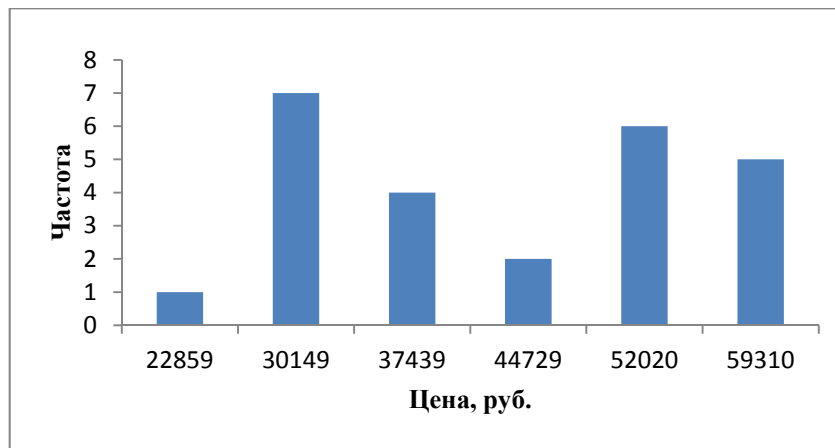


Рис.1 Гистограмма 25 цен на электростанции Honda

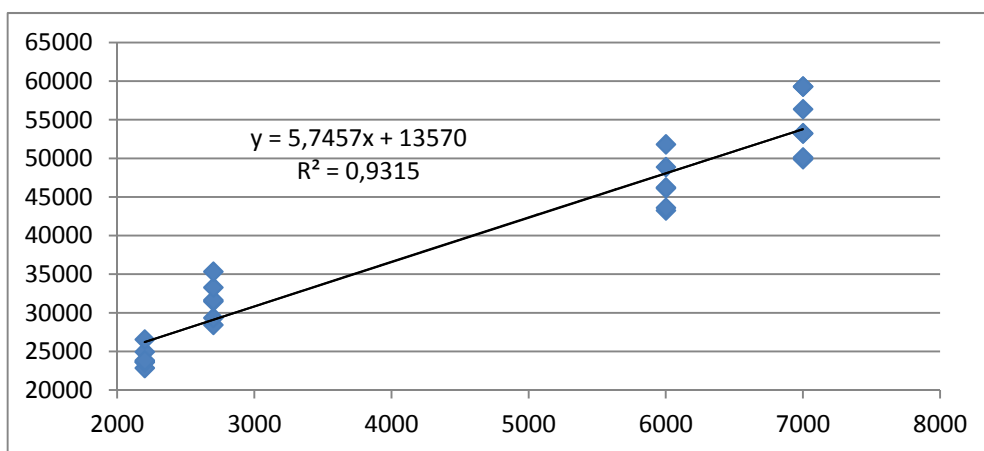


Рис.2 Зависимость цены на электростанции Honda от мощности

Зависимость средней цены от мощности электростанции хорошо приближает прямая линия. Адекватная полученным рыночным данным линейная регрессионная модель¹² имеет вид: $\tilde{y} = 5,75x + 13570$

В качестве объектов оценки последовательно приняты станции мощностью $X_1=2200$ Вт, $X_2=2700$ Вт, $X_3=6000$ Вт, $X_4=7000$ Вт¹³. Для каждого объекта рассчитывалась «истинная» оценка средней цены (стоимости).

Затем для каждого объекта оценки формировались две группы выборок по 13 и 18 объектов генсовокупности в каждой (доли отбора - 52% и 72% соответственно¹⁴). В каждой группе сгенерировано по 10 000 случайных выборок. По каждой выборке рассчитаны оценка стоимости и доверительный интервал для нее. Далее, в каждой из групп, соответствующей определенной доле отбора, подсчитывалась частота попадания «истинной» оценки в границы выборочных доверительных интервалов (табл.1).

¹² Характеристики модели см. в Приложении

¹³ То есть каждый из различающихся объектов, представленных в генеральной совокупности

¹⁴ Доли отбора выбирались близкими к 0,5 и 0,75

Таблица 1 Частость попадания «истинной» оценки стоимости электростанций Honda в доверительные интервалы, построенные по выборкам, отвечающим разным долям отбора

Доля отбора	Точка	Частость попадания в доверительный интервал	
		Стьюдента-Кокрена	Стьюдента
52%	X_1	99,2%	99,9%
	X_2	99,1%	99,9%
	X_3	92,9%	98,4%
	X_4	92,7%	98,5%
Средняя частость		96,0%	99,2%
72%	X_1	98,0%	100%
	X_2	98,2%	100%
	X_3	92,8%	99,7%
	X_4	92,7%	99,8%
Средняя частость		95,4%	99,9%

Как видим, частость попадания «истинного» значения в доверительный интервал Стьюдента-Кокрена, в среднем, оказывается близкой расчетной величине 95%, в то время как для «классического» доверительного интервала эта величина - заметно выше расчетной - 99-100%.

Это - ожидаемый результат, т.к. построение «классического» интервала Стьюдента основано на предположении о весьма большой (бесконечной) генеральной совокупности, в то время как реальные генсовокупности (объемы сделок на рынке) конечны и относительно невелики. Чем меньшее число членов однородной генеральной совокупности не учитывается выборкой, тем меньше дисперсия оценки среднего и, как следствие, уже доверительный интервал¹⁵.

В качестве второго примера *парной регрессии* рассмотрим аудиоплееры марки iAudio 10 фирмы Cowon с различным объемом памяти¹⁶:

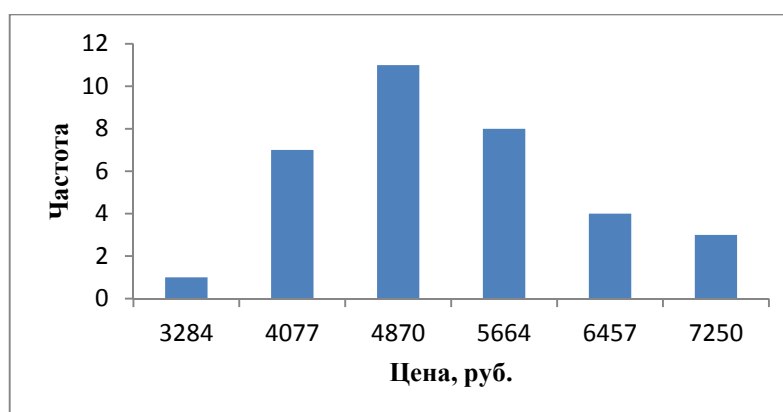


Рис.3 Гистограмма 31 цены на аудиоплееры Cowon

¹⁵ Граничный случай — если мы строим модель по всей генеральной совокупности, то получаем точное значение стоимости (т.е. с нулевой дисперсией оценки).

¹⁶ Цены взяты с сайта Яндекс-Маркет, см. табл. П2 Приложения

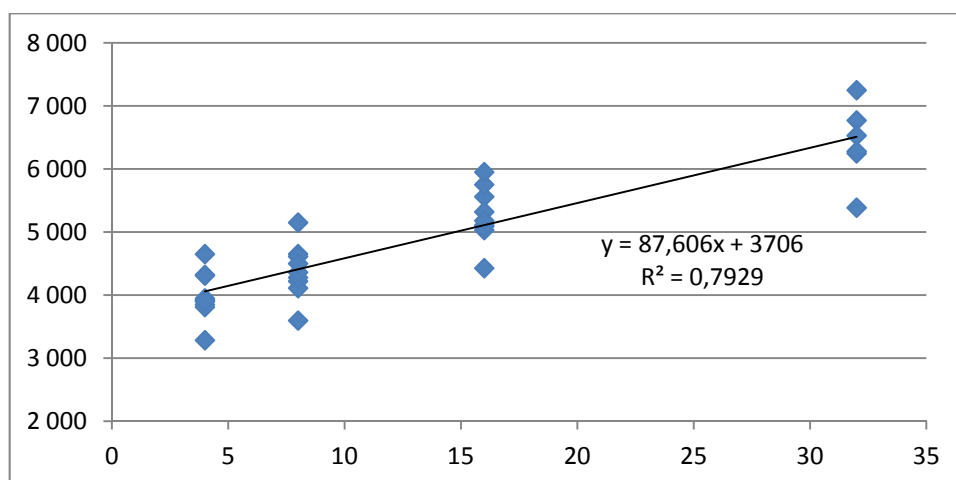


Рис.4 Зависимость цены на аудиоплееры Cowon iAudio 10 от объема памяти

Регрессионная модель с линейной зависимостью¹⁷, полученная по «генеральной совокупности», имеет вид: $\tilde{y} = 87,6x + 3706$.

За объекты оценки последовательно примем аудиоплееры с памятью $X_1=32$ Гб, $X_2=16$ Гб, $X_3=8$ Гб, $X_4=4$ Гб. Выполнив расчетные процедуры, аналогичные предыдущему примеру, получим (табл.2):

Таблица 2 Частость попадания «истинной» оценки стоимости плееров Cowon в доверительные интервалы, построенные по выборкам, отвечающим разным долям отбора

Доля отбора	Точка	Частость попадания в доверительный интервал	
		Стьюдента-Кокрена	Стьюдента
50%	X_1	87,9%	97,8%
	X_2	92,0%	98,5%
	X_3	97,3%	99,7%
	X_4	96,9%	99,6%
Средняя частость		93,5%	98,9%
74%	X_1	84,1%	99,6%
	X_2	90,5%	99,8%
	X_3	97,1%	100%
	X_4	96,6%	100%
Средняя частость		92,1%	99,8%

Как и в предыдущей модели, здесь наблюдается заметная разница в средних значениях частости попадания «истинных» оценок в доверительные интервалы, рассчитанные для одного и того же уровня надежности с учетом и без учета доли отбора.

Продолжим аналогичные расчеты для моделей многомерной регрессии.

¹⁷ Характеристики модели см. в Приложении

Начнем с *двухфакторной модели*, рассматривая выборку из 24 цен на карты памяти фирмы Kingston различных объема и скорости чтения¹⁸:

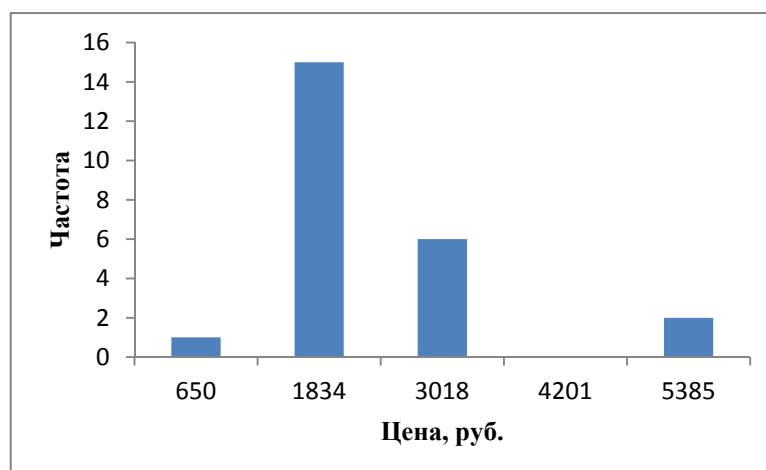


Рис.5 Гистограмма 24 цен на карты памяти Kingston.

По данной «генеральной совокупности» получена модель¹⁹ с линейными зависимостями вида:
$$\tilde{y} = 71,3z_1 + 69,8z_2 + 1312.1$$

Обозначим²⁰ $X = (z_1, z_2)$. Примем в качестве оцениваемого последовательно объекты с $X_1=(8, 20)$, $X_2=(8, 22)$, $X_3=(16, 15)$, $X_4=(16, 22)$, $X_5=(32, 15)$, $X_6=(32, 60)$, $X_7=(32, 22)$ и, проведя аналогичные расчеты, получим (табл.3):

Таблица 3 Частота попадания «истинной» оценки стоимости карт памяти Kingston в доверительные интервалы, построенные по выборкам, отвечающим разным долям отбора

Доля отбора	Точка	Частота попадания в доверительный интервал	
		Стьюдента-Кокрена	Стьюдента
50%	X_1	96,0%	99,3%
	X_2	96,3%	99,3%
	X_3	92,2%	98,2%
	X_4	93,5%	97,9%
	X_5	96,3%	99,6%
	X_6	93,4%	97,1%
	X_7	91,5%	97,5%
Средняя частота		94,2%	98,4%
71%	X_1	96,7%	100%
	X_2	97,0%	100%
	X_3	92,9%	99,9%
	X_4	95,5%	99,9%
	X_5	93,5%	100%

¹⁸ Цены взяты с сайта Яндекс-Маркет, см. табл. ПЗ Приложения

¹⁹ Характеристики модели см. в Приложении

²⁰ Z_1 - объем памяти, Гб; Z_2 – скорость чтения, Мб/с.

	X_6	92,7%	99,3%
	X_7	92,0%	99,6%
Средняя частота		94,3%	99,8%

И в этой модели наблюдаются близкие к заданному уровню 95% средние частоты попадания «истинных» оценок в доверительный интервал с учетом доли отбора, и повышенные средние значения этой частоты – для «классического» интервала Стьюдента.

Рассмотрим далее пример *трехфакторной модели* – цены на процессоры серии Phenom 2 фирмы AMD, различающиеся числом ядер, рабочей частотой и тепловыделением²¹

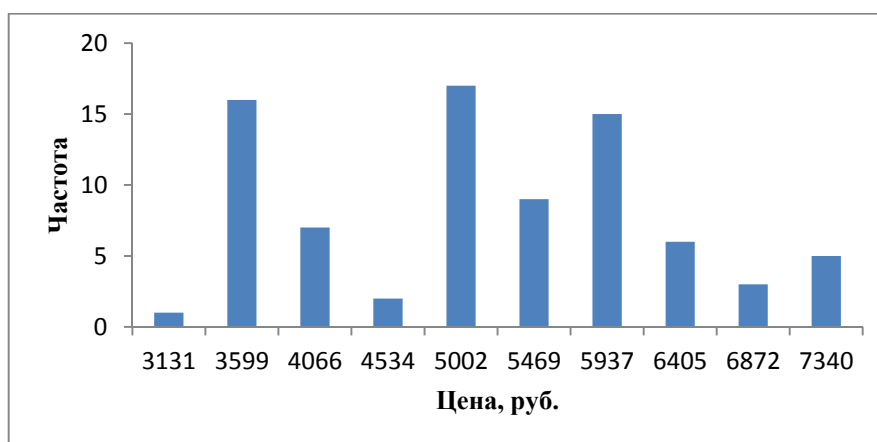


Рис.6 Гистограмма 81 цены на процессоры AMD Phenom 2.

По «генеральной совокупности» получена модель²² с линейными связями вида:

$$\hat{y} = 3,5z_1 - 8,4z_2 + 1203,8z_3 - 10880,5$$

Обозначим²³ $X = (z_1, z_2, z_3)$. В качестве объектов оценки последовательно возьмем объекты с $X_1=(2900, 95, 6)$, $X_2=(2800, 125, 6)$, $X_3=(2800, 95, 6)$, $X_4=(2600, 95, 6)$, $X_5=(3000, 125, 6)$, $X_6=(2700, 95, 6)$, $X_7=(3250, 125, 6)$, $X_8=(2900, 95, 4)$, $X_9=(3000, 95, 4)$.

Проведя расчетные процедуры, аналогичные описанным выше, получим (табл.4):

Таблица 4 Частота попадания «истинной» оценки стоимости процессора AMD Phenom 2 в доверительные интервалы, построенные по выборкам, отвечающим разным долям отбора

Доля отбора	Точка	Частота попадания в доверительный интервал	
		Стьюдента-Кокрена	Стьюдента
49%	X_1	91,2%	98,1%
	X_2	93,9%	99,0%

²¹ Цены взяты с сайта Яндекс-Маркет, см. табл. П4 Приложения

²² Характеристики модели см. в Приложении

²³ Z_1 - рабочая частота, МГц; Z_2 – тепловыделение, Вт; Z_3 – количество ядер;

	X_3	91,1%	98,0%
	X_4	92,1%	98,5%
	X_5	94,1%	99,1%
	X_6	91,8%	98,2%
	X_7	92,4%	98,4%
	X_8	99,1%	100%
	X_9	98,9%	100%
Средняя частота		93,8%	98,8%
74%	X_1	90,1%	99,8%
	X_2	93,8%	99,9%
	X_3	90,4%	99,8%
	X_4	91,0%	99,7%
	X_5	93,4%	99,9%
	X_6	90,7%	99,7%
	X_7	91,9%	99,8%
	X_8	99,0%	100%
	X_9	98,8%	100%
Средняя частота		93,2%	99,8%

И вновь мы наблюдаем уже привычное соотношение между средними частотами попадания «истинных» оценок в доверительные интервалы Стьюдента-Кокрена и Стьюдента.

Наконец, рассмотрим *модель с четырьмя факторами*. Сформируем выборку цен²⁴ на жесткие диски фирмы Seagate, различающиеся емкостью диска, скоростью вращения, количеством пластин и емкостью памяти буфера:

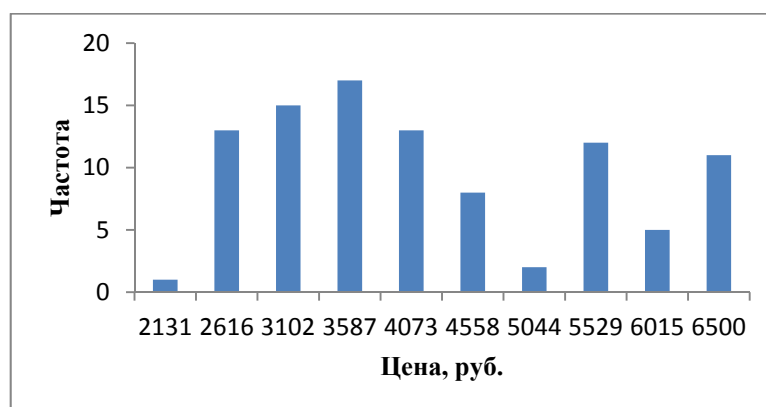


Рис.7 Гистограмма 97 цен на жесткие диски Seagate.

По этой «генеральной совокупности» получена модель²⁵ с линейными связями вида:

$$\hat{y} = 0,3z_1 + z_2 + 914,4z_3 + 7,3z_4 - 6067,9$$

²⁴ Цены взяты с сайта Яндекс-Маркет, см. табл. П5 Приложения

²⁵ Характеристики модели см. в Приложении

Обозначим²⁶ $X = (z_1, z_2, z_3, z_4)$. В качестве объектов оценки возьмем объекты с $X_1=(3000, 7200, 3, 64)$, $X_2=(500, 7200, 1, 16)$, $X_3=(1000, 7200, 1, 64)$, $X_4=(2000, 5900, 3, 64)$, $X_5=(1500, 5900, 3, 64)$, $X_6=(2000, 7200, 4, 64)$. Результаты расчетных процедур, аналогичных выполненным выше, представлены в табл.5

Таблица 5 Частота попадания «истинной» оценки стоимости жесткого диска Seagate в доверительные интервалы, построенные по выборкам с разными долями отбора

Доля отбора	Точка	Частота попадания в доверительный интервал	
		Стьюдента-Кокрена	Стьюдента
51%	X_1	84,0%	94,9%
	X_2	98,6%	100%
	X_3	98,2%	99,9%
	X_4	94,1%	99,2%
	X_5	97,2%	99,5%
	X_6	99,7%	100%
Средняя частота		95,3%	98,9%
74%	X_1	84,3%	99,0%
	X_2	98,4%	100%
	X_3	97,8%	100%
	X_4	94,0%	99,9%
	X_5	97,6%	99,9%
	X_6	99,8%	100%
Средняя частота		95,3%	99,8%

Как и в предыдущих случаях, в данной модели средняя частота попадания близка к расчетной для интервала, полученного с учетом доли отбора, и существенно превышает расчетный уровень 95% - для «классического» интервала.

Подведем промежуточные итоги.

Для всех построенных на реальной рыночной информации о *гомогенных* товарах регрессионных моделей с линейными связями и числом факторов от одного до четырех, средняя частота попадания «истинной» оценки стоимости в выборочный доверительный интервал с учетом доли отбора (Стьюдента-Кокрена) лежит в пределах 92,1 - 96,0% (близких к расчетному уровню 95%). В этих же условиях средняя частота попадания в «классический» интервал Стьюдента лежит в пределах 98,4 - 99,9%, близких к 100%.

Перейдем теперь к *объектам недвижимости*. Особенности оценки недвижимости в нашем рассмотрении являются нелинейные зависимости стоимости от влияющих факторов, а также то, что объект оценки, как правило, имеет набор

²⁶ Z_1 - емкость диска, Гб; Z_2 – скорость вращения, об/мин; Z_3 – количество пластин; Z_4 – емкость буфера, Мб.

значений признак-факторов, не совпадающий с аналогичным набором ни с одним из аналогов²⁷.

Рассмотрим пример из реального отчета об оценке, где объектом оценки является двухкомнатная квартира, расположенная на третьем этаже дома (старый фонд) в Центральном районе Санкт-Петербурга.

«Генеральная совокупность» цен представлена 23 квартирами-аналогами²⁸ и, в отличие от рассмотренных выше примеров, не включает объект оценки.

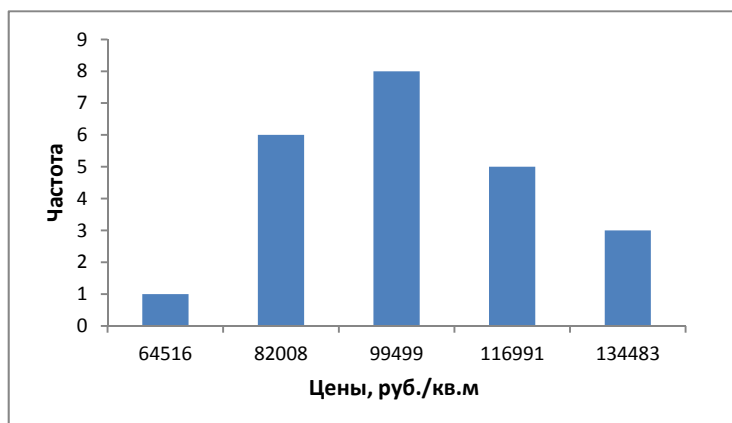


Рис.8 Гистограмма удельных цен по 23 квартирам (Центральный район Петербурга).

По «генеральной совокупности» квартир-аналогов построена *восьмифакторная* регрессионная модель с нелинейными связями²⁹, по которой получена «истинная» оценка стоимости для оцениваемой квартиры.

Далее из «генсовкупности» были сформированы все возможные выборки³⁰ по 17 аналогов (для отбора – 74%)³¹. По каждой выборке строилась регрессионная модель с теми же, что и в «генеральной» модели, зависимостями от влияющих факторов. По каждой «выборочной» модели определялась оценка удельной стоимости для объекта оценки и доверительные интервалы для нее по формулам (1), (2). Затем рассчитывались частоты попадания «истинной» оценки удельной стоимости объекта оценки в доверительные интервалы для нее.

За неимением возможности обеспечить долю отбора на уровне 50%, аналогичное моделирование для оцениваемой квартиры было проведено по выборкам с большими долями отбора³² - 18/23, 20/23 и 22/23.

В результате были получены следующие значения частот (табл.6):

²⁷ В рассмотренных выше моделях для движимого имущества зависимости были линейными и объектами оценки поочередно выступали все аналоги.

²⁸ см. табл. П6 Приложения

²⁹ Характеристики модели см. в Приложении

³⁰ Количество выборок в группе равно $C_{23}^{17} = 100947$.

³¹ Близкую к 50% долю отбора реализовать не удалось из-за недостатка степеней свободы. Как известно, необходимо, чтобы число аналогов превышало число объясняющих переменных, как минимум, в два раза, см., например, <http://www.appraiser.ru/default.aspx?SectionId=41&Id=1577>.

³² Количество выборок в группах равно $C_{23}^{18} = 33649$, $C_{23}^{20} = 1771$ и $C_{23}^{22} = 23$ соответственно.

Таблица 6 Частота попадания «истинной» оценки стоимости в доверительные интервалы, построенные по выборкам (квартира в Центральном районе Петербурга).

Доля отбора	Частота попадания в доверительный интервал	
	Стьюдента-Кокрена	Стьюдента
74%	96,2%	100%
78%	94,6%	100%
87%	96,5%	100%
96%	91,3%	100%
Средняя частота	94,7%	100%

Чтобы корректно сопоставлять результаты моделирования оценки недвижимости с аналогичными результатами для гомогенных товаров, в качестве оцениваемых объектов были дополнительно приняты два аналога, близких³³ к оцениваемой квартире.

Рассчитав предварительно «истинные» оценки стоимости для этих аналогов по «генеральной» модели и моделируя выборки с той же долей отбора (74%), получаем следующие значения частоты (табл.7):

Таблица 7 Частоты попадания «истинной» оценки стоимости в выборочные доверительные интервалы, построенные для объекта оценки и двух близких квартир-аналогов

Доля отбора	Номер аналога	Частота попадания в доверительный интервал	
		Стьюдента-Кокрена	Стьюдента
74%	2	92,1%	99,8%
	13	96,3%	100%
	Объект оценки	96,2%	100%
Средняя частота		94,9%	99,9%

Рассмотрим *другой пример оценки квартиры*, позволяющий провести моделирование с обеими интересующими нас долями отбора - близкими 50% и 75%.

В качестве объекта оценки выберем типовую двухкомнатную квартиру, расположенную на среднем этаже дома в одном из районов новостроек Петербурга – Приморском, в зоне влияния станции метрополитена «Комендантский проспект», и сформируем «генеральную совокупность» из предложений к продаже 31 аналогичной квартиры в ближайшем окружении³⁴.

³³ Близость аналогов, для простоты, определена по модулю разности между «истинным» значением оценки стоимости и ценой аналога; одна из цен выбрана меньше, а другая - больше «истинной» оценки.

³⁴ см. Табл. П7 Приложения. Источник: портал www.bn.ru, май 2012 г.

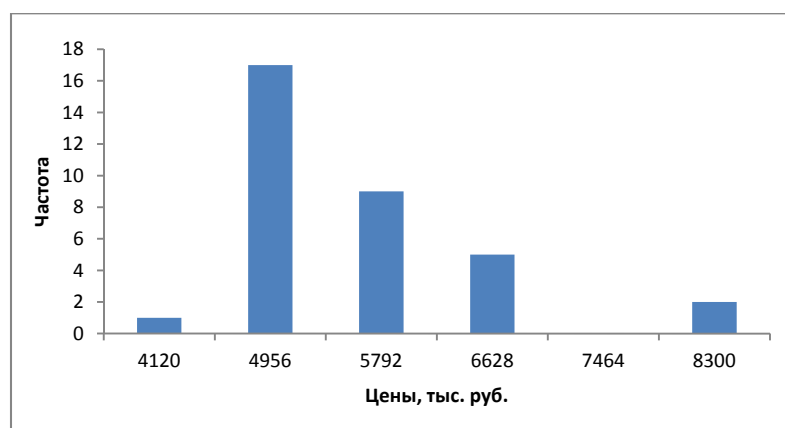


Рис.9 Гистограмма распределения цен по 31 квартире (Приморский район Петербурга)

По «генеральной совокупности» построена *шестифакторная* регрессионная модель с нелинейными связями³⁵, по которой получена «истинная» оценка стоимости для оцениваемой квартиры.

Затем из генеральной совокупности случайным образом сформированы 2 группы выборок (по 10 тыс. каждая)³⁶ – по 16 и по 23 аналога. По каждой выборке получены оценка стоимости оцениваемой квартиры и доверительные интервалы для нее. Затем в каждой группе выборок рассчитана частота попадания «истинной» стоимости объекта оценки в выборочные доверительные интервалы (табл.8):

Таблица 8 Частоты попадания «истинной» стоимости оцениваемой квартиры в выборочные доверительные интервалы (Приморский район).

Доля отбора	Частота попадания в доверительный интервал	
	Стьюдента-Кокрена	Стьюдента
52%	96,8%	99,5%
74%	96,9%	100%
Средняя частотность	96,8%	99,8%

Проведя попутно аналогичное моделирование для промежуточных значений доли отбора 19/31 и 22/31, получим аналогичные результаты (табл.9):

Таблица 9 Частоты попадания «истинной» стоимости оцениваемой квартиры в выборочные доверительные интервалы (Приморский район).

Доля отбора	Частота попадания в доверительный интервал	
	Стьюдента-Кокрена	Стьюдента
61%	96,9%	99,9%
71%	96,9%	100%
Средняя частотность	96,9%	99,9%

Как и в предыдущем примере с квартирой в Центральном районе, для получения результатов, сопоставимых с результатами предыдущего моделирования,

³⁵ Характеристики модели см. в Приложении

³⁶ Количество всех сочетаний в этом случае велико (от десятков до сотен миллионов), поэтому здесь случайным образом генерирована лишь часть из них.

выберем в качестве оцениваемых объектов дополнительно два аналога (16, 17)³⁷, близких к оцениваемой квартире. Проведя аналогичные расчеты, получим (табл.10):

Табл.10 Частоты попадания «истинный» стоимости объекта в выборочные доверительные интервалы (Приморский район)

Доля отбора	Номер аналога	Частота попадания в доверительный интервал	
		Стьюдента-Кокрена	Стьюдента
52%	16	98,4%	99,4%
	17	97,0%	98,8%
	Объект оценки	96,8%	99,5%
Средняя частотность		97,4%	99,2%
74%	16	97,4%	100%
	17	92,1%	99,9%
	Объект оценки	96,9%	100%
Средняя частотность		95,5%	99,97%

Таким образом, при построении регрессионных моделей на реальных данных рынка *недвижимости* наблюдается такое же соотношение частоты попадания «истинных» оценок в доверительные интервалы, что и на рынке *гомогенных* товаров: в интервал, построенный с учетом известной доли отбора – близкие к расчетному уровню 95%, в «классический» интервал - близкие к 100%.

Подведем окончательные итоги расчетов, сведя результаты исследования в единую таблицу³⁸:

Табл.10 Сводные по различным регрессионным моделям результаты расчетов частоты попадания истинной оценки стоимости в доверительные интервалы

Число факторов модели	1		2	3	4	6	8
Объекты	Плееры	Электростанции	Карты памяти	Процессоры	Жесткие диски	Квартиры	Квартиры
Доверительный интервал	Стьюдента-Кокрена (с учетом доли отбора)						
Доли отбора 49-52%	93,5%	96,0%	94,2%	93,8%	95,3%	97,7%	-
Доли отбора 71-74%	92,1%	95,4%	94,3%	93,2%	95,3%	94,7%	94,2%
Доверительный интервал	Стьюдента (классический)						
Доли отбора 49-52%	98,9%	99,2%	98,4%	98,8%	98,9%	99,1%	-
Доли отбора 71-74%	99,8%	99,9%	99,8%	99,8%	99,8%	99,9%	99,9%

³⁷ См. таблицу П7 Приложения

³⁸ В таблице приведены значения долей отбора и средних частоты попадания «истинных» значений оценки стоимости в доверительные интервалы.

Средние значения частоты попадания в доверительный интервал по всем объектам оценки и долям отбора составляют:

- для интервала Стьюдента-Кокрена (с учетом доли отбора) - **94,6%**;
- для интервала Стьюдента (с пренебрежимо малой долей отбора) – **99,4%**.

Таким образом, проведенное локальное исследование возможности учета в интервальной оценке регрессионной модели отношения объемов выборки и генеральной совокупности (рынка) при ограниченных объемах рыночных данных показывают:

1. Применение формулы (2), учитывающей конечную величину доли отбора цен с рынка, для расчета интервальной оценки стоимости объекта, получаемой по регрессионным моделям, вполне оправдано — *интервал Стьюдента-Кокрена* обеспечивает надежность на уровнях, близких к расчетному (95%). Этот интервал при известной или предполагаемой доле отбора *может служить «оценкой снизу»* величины рассчитываемого доверительного интервала.
2. В этих же условиях классический *доверительный интервал Стьюдента* обеспечивает надежность попадания в него стоимости объекта на уровнях 99-100%. Поэтому он *может рассматриваться как гарантированная «оценка сверху»* рассчитываемого доверительного интервала.
3. Обсуждаемые оценки «надежности» (средних частостей) попадания в доверительные интервалы получены в условиях *различных*:
 - *рынков* – гомогенных и гетерогенных (недвижимость) товаров;
 - *распределений цен* (существенно отличающихся от нормального: асимметрия от -0,02 до 1,97; эксцесс от -1,56 до 3,72);
 - *размерности регрессионных моделей* (числа влияющих факторов);
 - *функций влияния факторов модели на зависимую величину* - линейной и нелинейных;
 - *методов оцифровки* (линеаризации) нелинейных зависимостей;
 - *показателей зависимой переменной* – стоимости всего объекта и удельного показателя стоимости;
 - *распределений остатков моделей* (отличающихся от нормального: асимметрия от -0,27 до 0,92; эксцесс от -0,95 до 1,2; отношение «интервал/СКО» - 3,3-4,6);
 - *показателей качества моделей, полученных по «генеральным совокупностям» и дающих «истинную» оценку стоимости*: R^2 – 0,72 – 0,97, средняя ошибка аппроксимации – 3,4 -14,6%;

- *долей отбора* – отношения объема выборки рыночных данных к объему генеральной совокупности (объему сделок на рынке) - 49-96%.

Все это позволяет рассматривать полученные результаты как достаточно универсальные, т.е. не связанные с конкретными характеристиками моделей по любому из упомянутых условий.

Литература

1. Зельдин М.А., Баринов Н.П., Аббасов М.Э. Доверительный интервал для среднего по выборке из конечной генеральной совокупности, Регистр оценщиков, №11, 2012, с. 70-75; Бюллетень RWAY, №211 (октябрь 2012); <http://www.appraiser.ru/default.aspx?SectionId=35&Id=3657>
2. Йетс Ф. Выборочный метод в переписях и обследованиях, перевод с английского, М.: «Статистика», 1965. — 434 с.
3. Дрейпер Н., Смит Г. Прикладной регрессионный анализ, том 1. М.: Финансы и статистика, 1986. — 366 с.
4. Кокрен У. Методы выборочного исследования. М.: «Статистика», 1976 – 440с.

Опубликовано:

Материалы VI Поволжской научно-практической конференции «Статистические методы массовой и индивидуальной оценки», Нижний Новгород, 13–15.06.2013

Информационно-аналитический бюллетень рынка недвижимости RWAY №221 (август 2013)

Таблица П1 «Генеральная совокупность» по электростанциям Honda³⁹.

№	Максимальная мощность, Вт	Цена	№	Максимальная мощность, Вт	Цена
1	2 200	24 949	13	6 000	48 888
2	2 200	23 649	14	6 000	43 268
3	2 200	23 632	15	6 000	46 148
4	2 200	22 859	16	6 000	43 608
5	2 200	26 559	17	6 000	51 828
6	2 200	23 835	18	6 000	46 244
7	2 700	33 289	19	7 000	56 380
8	2 700	28 439	20	7 000	49 890
9	2 700	31 495	21	7 000	53 213
10	2 700	35 339	22	7 000	59 310
11	2 700	29 359	23	7 000	59 310
12	2 700	31 662	24	7 000	50 140
			25	7 000	53 278

Показатели регрессионного уравнения⁴⁰, полученного по генеральной совокупности

	A1	A0
Функция ЛИНЕЙН():	5,75	13570,1
	0,32	1632,7
<i>R-квадрат</i>	0,932	3379,7
	313,0	23,0
	3574896333	262717813,8
<i>t Stat</i>	17,7	8,3
<i>P-value</i>	7E-15	
<i>Фактор значим на</i>	100%	
<i>Знак коэффициента.</i>	логичен	

Гистограмма остатков⁴¹ модели (электростанции Honda)



³⁹ Получено по данным http://market.yandex.ru/compare.xml?hid=226667&CAT_ID=7320593&CMD=-CMP=7353485,7353481,7353475,7353476 Коэффициент асимметрии для цен равен 0,04; коэффициент эксцесса — -1,56, отношение Интервал/СКО — 2,88.

⁴⁰ Дополнительно: средняя ошибка аппроксимации равна 7,2%, максимальная - 17,7%.

⁴¹ Коэффициент асимметрии – 0,44; коэффициент эксцесса – 0,91; отношение Интервал/СКО – 3,3

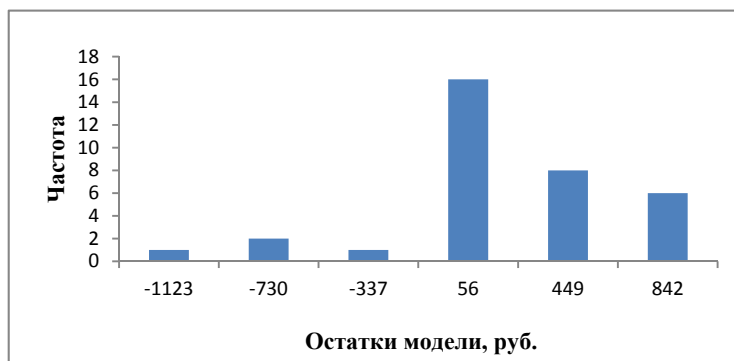
Таблица П2 «Генеральная совокупность» по аудиоплеерам Cowon⁴².

№	Объём встроенной памяти, Гб	Цена	№	Объём встроенной памяти, Гб	Цена
1	32	5 386	18	8	3 595
2	32	6 770	19	8	5 150
3	32	6 530	20	8	4 110
4	32	6 247	21	8	4 610
5	32	7 250	22	8	4 277
6	32	6 280	23	8	4 360
7	16	5 030	24	8	4 500
8	16	5 110	25	4	3 920
9	16	5 560	26	4	3 850
10	16	5 752	27	4	4 320
11	16	5 950	28	4	3 907
12	16	5 087	29	4	3 947
13	16	4 427	30	4	3 810
14	16	5 320	31	4	4 650
15	16	5 180	32	4	3 284
16	8	4 220	33	4	4 310
17	8	4 652	34	4	3 900

Показатели регрессионного уравнения⁴³, полученного по генеральной совокупности:

	A1	A0
Функция ЛИНЕЙН():	87,6	3706,0
	7,91	130,1
<i>R-квадрат</i>	0,793	453,3
	122,5	32,0
	25180766,5	6576589,0
<i>t Stat</i>	11,1	28,5
<i>P-value</i>	1,8E-12	
<i>Фактор значим на</i>	100%	
<i>Знак коэффициента</i>	логичен	

Гистограмма остатков⁴⁴ модели (плееры)



⁴² Получено по данным http://market.yandex.ru/compare.xml?hid=90560&CAT_ID=137555&CMD=-CMP=7701101,7701100,7701098,7701097 Коэффициент асимметрии для цен равен 0,7; коэффициент эксцесса — -0,2; отношение Интервал/СКО — 4,04 .

⁴³ Дополнительно: средняя ошибка аппроксимации равна 7,0%, максимальная - 23,5%.

⁴⁴ Коэффициент асимметрии — 0,45; коэффициент эксцесса — 0,27; отношение Интервал/СКО — 4,4

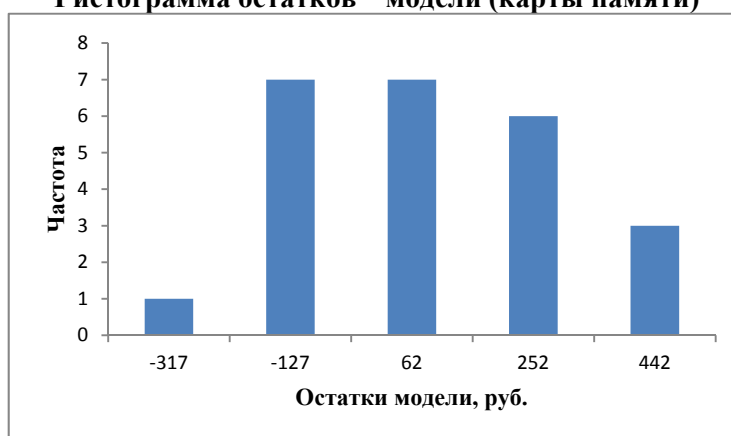
Таблица ПЗ «Генеральная совокупность» по картам памяти Kingston⁴⁵.

№	Объем памяти	Скорость чтения (Мб/с)	Цена	№	Объем памяти	Скорость чтения (Мб/с)	Цена
1	8	20	950	13	16	22	1070
2	8	20	890	14	16	22	1390
3	8	22	660	15	16	22	1185
4	8	22	1010	16	16	22	1100
5	8	22	740	17	32	15	1940
6	8	22	658	18	32	15	2130
7	8	22	650	19	32	60	5385
8	16	15	1030	20	32	60	5034
9	16	15	1066	21	32	22	2395
10	16	22	1255	22	32	22	2190
11	16	22	1120	23	32	22	2949
12	16	22	1690	24	32	22	2500

Показатели регрессионного уравнения⁴⁶, полученного по генеральной совокупности:

	A2 (объем памяти)	A1 (скорость чтения)	A0
Функция ЛИНЕЙН():	69,8	71,3	-1312,1
	4,4	5,0	123,9
R-квадрат	0,970	225,8	#Н/Д
	344,9	21,0	#Н/Д
	35177856,3	1070863,7	#Н/Д
t Stat	15,96	14,18	10,59
P-value	3E-13	3E-12	
Фактор значим на	100%	100%	
Знак коэф-та	логичен	логичен	

Гистограмма остатков⁴⁷ модели (карты памяти)



⁴⁵ Получено по данным http://market.yandex.ru/compare.xml?hid=91032&CAT_ID=432805&CMD=-CMP=6118189,6273901,6857848,7311578,6935885,6935886,6857849,7265361,6857847,6935887,6118188
Коэффициент асимметрии для цен — 1,97; коэффициент эксцесса — 3,72; отношение Интервал/СКО — 3,77.

⁴⁶ Дополнительно: средняя ошибка аппроксимации равна 14,6%, максимальная - 31%.

⁴⁷ Коэффициент асимметрии — 0,45; коэффициент эксцесса — 0,27; отношение Интервал/СКО — 4,4

Таблица П4 «Генеральная совокупность» по процессорам AMD⁴⁸

№	Частота	Тепловыделение	Кол-во ядер	Цена	№	Частота	Тепловыделение	Кол-во ядер	Цена
1	2900	95	6	5074	42	3000	125	6	6072
2	2900	95	6	5164	43	3000	125	6	6739
3	2900	95	6	5198	44	3000	125	6	6924
4	2900	95	6	5523	45	2700	95	6	4663
5	2900	95	6	5785	46	2700	95	6	4690
6	2900	95	6	6370	47	2700	95	6	4804
7	2900	95	6	6510	48	2700	95	6	4857
8	2800	125	6	4357	49	2700	95	6	4890
9	2800	125	6	4636	50	2700	95	6	5340
10	2800	125	6	4734	51	3250	125	6	6028
11	2800	125	6	4740	52	3250	125	6	6311
12	2800	125	6	4760	53	3250	125	6	7340
13	2800	125	6	4772	54	3250	125	6	6668
14	2800	125	6	4969	55	3250	125	6	6886
15	2800	125	6	5020	56	3250	125	6	6992
16	2800	125	6	5200	57	3250	125	6	7242
17	2800	125	6	5500	58	2900	95	4	3285
18	2800	95	6	4710	59	2900	95	4	3315
19	2800	95	6	4800	60	2900	95	4	3374
20	2800	95	6	5275	61	2900	95	4	3390
21	2800	95	6	5501	62	2900	95	4	3409
22	2800	95	6	5580	63	2900	95	4	3490
23	2600	95	6	4860	64	2900	95	4	3500
24	2600	95	6	4428	65	2900	95	4	3513
25	2600	95	6	4752	66	2900	95	4	3590
26	2600	95	6	4611	67	2900	95	4	3667
27	2600	95	6	4719	68	2900	95	4	4030
28	2600	95	6	5022	69	3000	95	4	3131
29	2600	95	6	5687	70	3000	95	4	3139
30	2600	95	6	5751	71	3000	95	4	3202
31	3000	125	6	5400	72	3000	95	4	3458
32	3000	125	6	5723	73	3000	95	4	3490
33	3000	125	6	5732	74	3000	95	4	3810
34	3000	125	6	5750	75	3000	95	4	3951
35	3000	125	6	5786	76	3000	95	4	3190
36	3000	125	6	5809	77	3000	95	4	3462
37	3000	125	6	5867	78	3000	95	4	3384
38	3000	125	6	5870	79	3000	95	4	3774
39	3000	125	6	5920	80	3000	95	4	3780

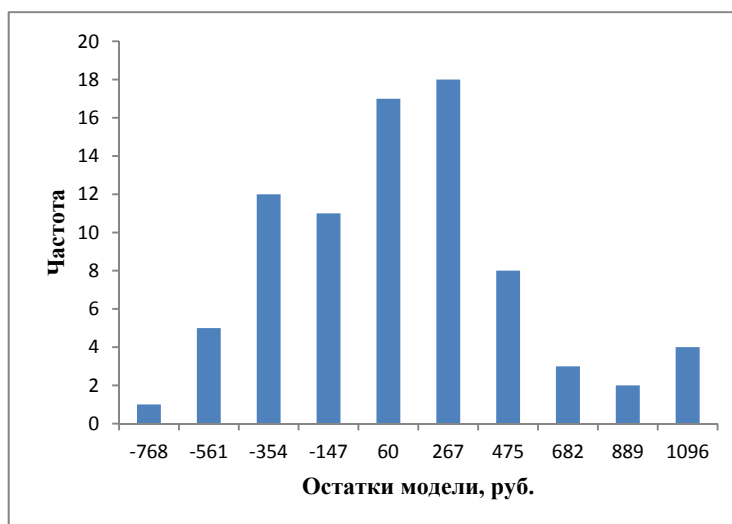
⁴⁸ Получено по данным http://market.yandex.ru/guru.xml?CMD=-RR=0,0,0,0-PF=1801946~EQ~sel~7971913-PF=2141160768~EQ~sel~x487954812-PF=2141160768~EQ~sel~x487954814-VIS=70-CAT_ID=651600-EXC=1-PG=10&hid=91019 Коэффициент асимметрии для цен — 0,11; коэффициент эксцесса — -0,86; отношение Интервал/СКО — 3,72.

40	3000	125	6	5970	81	3000	95	4	3963
41	3000	125	6	6070					

Показатели регрессионного уравнения⁴⁹, полученного по генеральной совокупности:

	A3 (кол-во ядер)	A2 (тепловыделение)	A1 (частота)	A0
Функция ЛИНЕЙН():	1203,8	-8,4	3,5	-10880,5
	71,0	4,8	0,4	1028,8
<i>R-квадрат</i>	0,869	417,6	#Н/Д	#Н/Д
	169,7	77,0	#Н/Д	#Н/Д
	88769821,1	13426530,6	#Н/Д	#Н/Д
<i>t Stat</i>	16,94	1,75	9,44	10,58
<i>P-value</i>	1E-27	0,085	2E-14	
<i>Фактор значим на</i>	100%	91,5%	100%	
<i>Знак коэф-та</i>	логичен	логичен	логичен	

Гистограмма остатков⁵⁰ модели (процессоры)



⁴⁹ Дополнительно: средняя ошибка аппроксимации равна 6,5%, максимальная - 18,7%.

⁵⁰ Коэффициент асимметрии – 0,58; коэффициент эксцесса – 0,43; отношение Интервал/СКО – 4,6

Таблица П5 «Генеральная совокупность» по жестким дискам Seagate⁵¹.

№	Ем-кость	Ско-рость вращения	Количе-ство пла-стин	Бу-фер	Цена	№	Ем-кость	Ско-рость вращения	Количе-ство пла-стин	Бу-фер	Цена
1	3000	7200	3	64	5033	49	1000	7200	1	64	3200
2	3000	7200	3	64	5070	50	1000	7200	1	64	3260
3	3000	7200	3	64	5090	51	1000	7200	1	64	3400
4	3000	7200	3	64	5110	52	1000	7200	1	64	3600
5	3000	7200	3	64	5138	53	2000	5900	3	64	3740
6	3000	7200	3	64	5201	54	2000	5900	3	64	3778
7	3000	7200	3	64	5328	55	2000	5900	3	64	3790
8	3000	7200	3	64	5330	56	2000	5900	3	64	3790
9	3000	7200	3	64	5370	57	2000	5900	3	64	3940
10	3000	7200	3	64	5403	58	2000	5900	3	64	3945
11	3000	7200	3	64	5442	59	2000	5900	3	64	3982
12	3000	7200	3	64	5445	60	2000	5900	3	64	4080
13	3000	7200	3	64	5470	61	2000	5900	3	64	4086
14	3000	7200	3	64	5610	62	2000	5900	3	64	4090
15	3000	7200	3	64	6080	63	2000	5900	3	64	4140
16	3000	7200	3	64	6190	64	2000	5900	3	64	4214
17	3000	7200	3	64	6428	65	2000	5900	3	64	4228
18	500	7200	1	16	2131	66	2000	5900	3	64	4230
19	500	7200	1	16	2232	67	2000	5900	3	64	4620
20	500	7200	1	16	2250	68	1500	5900	3	64	3210
21	500	7200	1	16	2350	69	1500	5900	3	64	3229
22	500	7200	1	16	2366	70	1500	5900	3	64	3259
23	500	7200	1	16	2380	71	1500	5900	3	64	3351
24	500	7200	1	16	2429	72	1500	5900	3	64	3374
25	500	7200	1	16	2430	73	1500	5900	3	64	3429
26	500	7200	1	16	2450	74	1500	5900	3	64	3450
27	500	7200	1	16	2480	75	1500	5900	3	64	3490
28	500	7200	1	16	2490	76	1500	5900	3	64	3510
29	500	7200	1	16	2500	77	1500	5900	3	64	3530
30	500	7200	1	16	2590	78	1500	5900	3	64	3530
31	500	7200	1	16	2690	79	1500	5900	3	64	3560
32	500	7200	1	16	2712	80	1500	5900	3	64	3600
33	500	7200	1	16	2790	81	1500	5900	3	64	3640
34	500	7200	1	16	2850	82	1500	5900	3	64	3648
35	500	7200	1	16	3060	83	1500	5900	3	64	3730
36	1000	7200	1	64	2590	84	1500	5900	3	64	3800
37	1000	7200	1	64	2675	85	1500	5900	3	64	4200

⁵¹ Ссылка на данные: http://market.yandex.ru/compare.xml?hid=91033&CAT_ID=686672&CMD=-RR=0.0.0.0-CMP=7776547,7627770,7773181,6842727,6842726,7075659-VIS=70-CAT_ID=686672-EXC=1-PG=10.

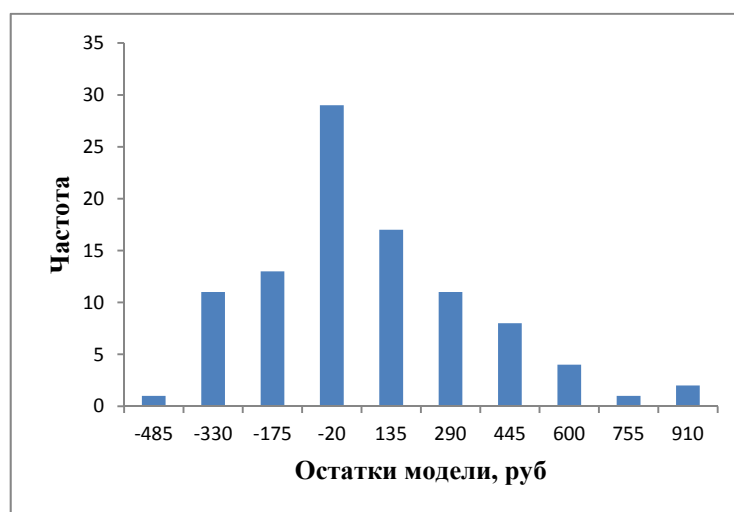
Коэффициент асимметрии для цен равен 0,54; коэффициент эксцесса — -0,98; отношение Интервал/СКО – 3,4.

38	1000	7200	1	64	2736	86	2000	7200	4	64	5901
39	1000	7200	1	64	2875	87	2000	7200	4	64	5913
40	1000	7200	1	64	2930	88	2000	7200	4	64	5950
41	1000	7200	1	64	2946	89	2000	7200	4	64	5960
42	1000	7200	1	64	3020	90	2000	7200	4	64	6035
43	1000	7200	1	64	3040	91	2000	7200	4	64	6120
44	1000	7200	1	64	3050	92	2000	7200	4	64	6154
45	1000	7200	1	64	3060	93	2000	7200	4	64	6229
46	1000	7200	1	64	3100	94	2000	7200	4	64	6290
47	1000	7200	1	64	3130	95	2000	7200	4	64	6400
48	1000	7200	1	64	3190	96	2000	7200	4	64	6470
						97	2000	7200	4	64	6500

Показатели регрессионного уравнения⁵², полученного по генеральной совокупности:

	A4 (Буфер)	A3 (Количество пластин)	A2 (Скорость вращения)	A1 (Емкость)	A0
Функция ЛИНЕЙН():	7,3	914,4	1,0	0,3	-6067,9
	2,3	48,2	0,1	0,1	456,2
<i>R-квадрат</i>	0,950	294,6	#Н/Д	#Н/Д	#Н/Д
	433,0	92,0	#Н/Д	#Н/Д	#Н/Д
	150287634,4	7982877,9	#Н/Д	#Н/Д	#Н/Д
<i>t Stat</i>	3,180	18,986	17,277	4,957	13,301
<i>P-value</i>	0,002	1,4E-33	1,2E-30	3,3E-06	
<i>Фактор значим на</i>	99,8%	100%	100%	100%	
<i>Знак коэф-та</i>	логичен	логичен	логичен	логичен	

Гистограмма остатков⁵³ модели (жесткие диски)



⁵² Дополнительно: средняя ошибка аппроксимации равна 5,9%, максимальная - 18%.

⁵³ Коэффициент асимметрии – 0,69; коэффициент эксцесса – 0,48; отношение Интервал/СКО – 4,6

Таблица П6 «Генеральная совокупность» квартир в Центральном районе Петербурга⁵⁴.

№	Адрес	Цена/ кв.м	Комнат	S общая	S кухни	Тип здания	Этаж	Состояние	Планировка	До метро
1	Конная ул. 10	100 000	2	32,0	6,5	СФ	Пер.	Отл.	Одн.	750
2	Конная ул. 1	108 788	2	33,0	7,7	СФКР	Ср.	Хор.	Двус.	680
3	Гончарная ул. 9	104 651	2	43,0	5,5	СФ	Пер.	Хор.	Одн.	200
4	Невский пр. 115	134 483	1	43,5	8,0	СФКР	Пос.	Отл.	Одн.	710
5	Невский пр. 111	91 324	2	43,8	6,0	СФ	Пос.	Хор.	Одн.	530
6	Конная ул. 1	86 957	1	46,0	7,0	СФКР	Ср.	Т.р.	Одн.	680
7	Тележная ул. 13	93 117	2	49,4	10,0	СФКР	Ср.	Хор.	Одн.	620
8	Конная ул. 12	92 000	1	50,0	7,0	СФ	Ср.	Хор.	Одн.	730
9	Полтавская ул. 14	78 534	2	57,3	8,0	СФКР	Ср.	Т.р.	Одн.	590
10	Невский пр. 133	121 777	1	57,4	9,5	СФКР	Ср.	Отл.	Одн.	560
11	Исполкомская ул. 10	104 839	3	62,0	4,0	СФКР	Пер.	Отл.	Одн.	490
12	Невский пр. 153	96 875	2	64,0	8,0	СФ	Ср.	Хор.	Одн.	220
13	Конная ул. 15	114 706	2	68,0	5,8	СФКР	Ср.	Отл.	Одн.	520
14	Бакунина пр. 13	121 429	2	70,0	9,0	СФКР	Ср.	Отл.	Двус.	740
15	Полтавская ул. 12	69 014	2	71,0	9,3	СФКР	Ср.	Т.р.	Одн.	610
16	Исполкомская ул. 5	74 722	2	72,0	14,0	СФ	Ср.	Т.р.	Одн.	390
17	Херсонская ул. 10	85 135	3	74,0	12,8	СФ	Ср.	Хор.	Одн.	600
18	Херсонская ул. 10	64 516	3	77,5	15,0	СФ	Пер.	Т.р.	Одн.	600
19	Тележная ул. 13	76 923	3	78,0	9,0	СФ	Пос.	Хор.	Одн.	620
20	Невский пр. 172	98 765	3	81,0	8,2	СФКР	Ср.	Уд.	Двус.	380
21	Невский пр. 173	82 927	2	82,0	11,0	СФКР	Ср.	Т.р.	Одн.	70
22	Тележная ул. 23	72 222	3	90,0	10,0	СФКР	Пос.	Т.р.	Одн.	270
23	Невский пр. 93	76 333	3	90,0	12,3	СФ	Ср.	Т.р.	Одн.	490
00	Тележная, 25	?	2	50,3	8,3	СФ	Ср.	Отл.	Одн.	530

⁵⁴ Источник данных: отчет об оценке, Коэффициент асимметрии распределения цен — 0,47; коэффициент эксцесса — -0,42; отношение Интервал/СКО — 3,8.

Принятые сокращения:

Тип здания: СФ - старый фонд, СФКР – старый фонд после капремонта;

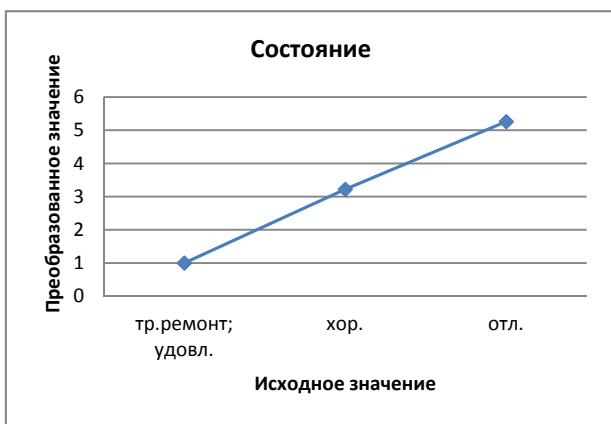
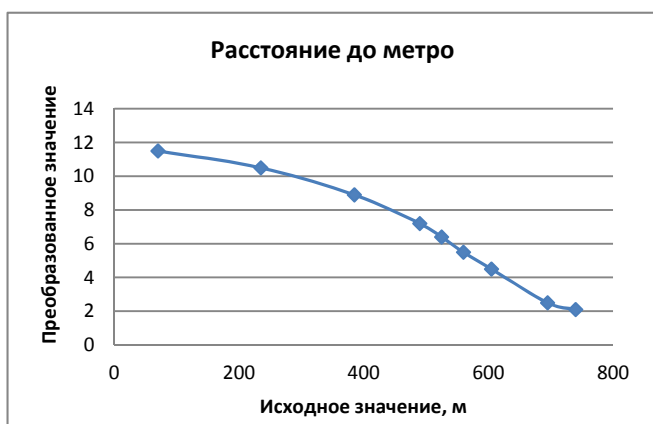
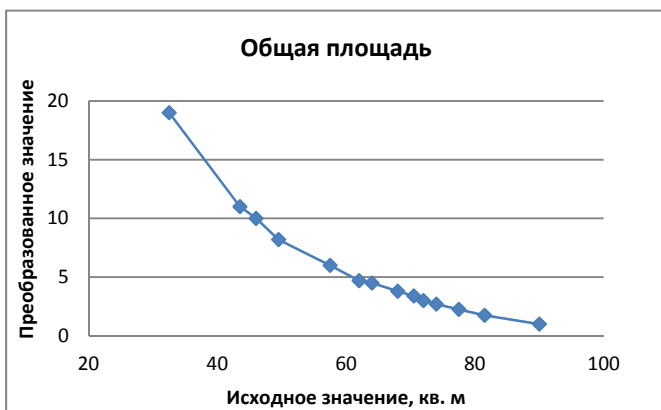
Этаж: Пер. – первый, Ср. – средний, Пос. – последний;

Состояние: Т.р. – требует ремонта, Уд. – удовлетворительное, Хор. – хорошее, Отл. – отличное;

Планировка: Одн. – односторонняя, Двус. – двусторонняя (окна на улице и во двор).

Цена 1 кв.м измеряется в тыс. руб., расстояние до метро - в метрах, площади - в кв.м.

Принятые в модели зависимости средней удельной цены от влияющих факторов (Центральный район, старый фонд):



Объекты сравнения и их оцифрованные характеристики (Центральный район Петербурга).

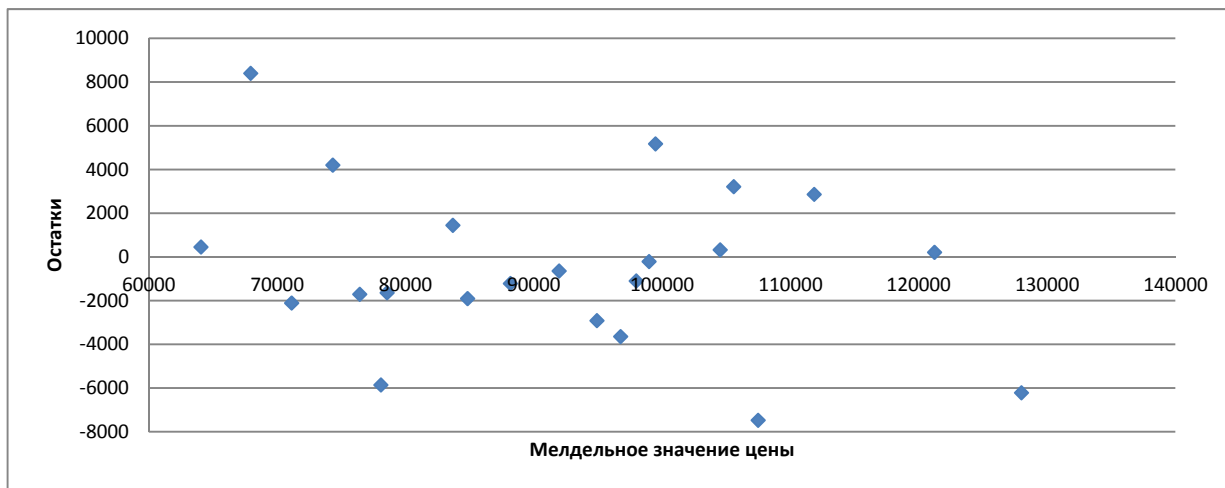
	Y	X1	X2	X3	X4	X5	X6	X7	X8
№ аналога	Цена 1 кв.м.	Количество комнат	Общая площадь	Площадь кухни	Тип здания (капремонт)	Этаж	Состояние	Планировка	До метро
1	100000	1	19,0	7,0	1	1,0	5,3	1	1,0
2	108788	1	19,0	7,5	2	3,9	3,2	1	2,5
3	104651	1	11,0	6,0	1	1,0	3,2	1	10,5
4	134483	3,5	11,0	7,5	2	1,9	5,3	1	2,5
5	91324	1	11,0	6,0	1	1,9	3,2	1	6,4
6	86957	3,5	10,0	7,0	2	3,9	1,0	1	2,5
7	93117	1	8,2	7,6	2	3,9	3,2	1	4,5
8	92000	3,5	8,2	7,0	1	3,9	3,2	1	1,0
9	78534	1	6,0	7,5	2	3,9	1,0	1	4,5
10	121777	3,5	6,0	7,5	2	3,9	5,3	1	5,5
11	104839	1	4,7	1,0	2	1,0	5,3	1	7,2
12	96875	1	4,5	7,5	1	3,9	3,2	1	10,5
13	114706	1	3,8	6,0	2	3,9	5,3	1	6,4
14	121429	1	3,4	7,5	2	3,9	5,3	2	1,0
15	69014	1	3,4	7,5	2	3,9	1,0	1	4,5
16	74722	1	3,4	10,3	1	3,9	1,0	1	8,9
17	85135	1	2,7	8,6	1	3,9	3,2	1	4,5
18	64516	1	2,3	13,8	1	1,0	1,0	1	4,5
19	76923	1	2,3	7,5	1	1,9	3,2	1	4,5
20	98765	1	1,8	7,5	2	3,9	1,0	2	8,9
21	82927	1	1,8	7,8	2	3,9	1,0	1	11,5
22	72222	1	1,0	7,6	2	1,9	1,0	1	10,5
23	76333	1	1,0	8,6	1	3,9	1,0	1	7,2
00	?	1	8,2	7,5	1	3,9	5,3	1	6,4

Показатели регрессионного уравнения⁵⁵, полученного по генеральной совокупности:

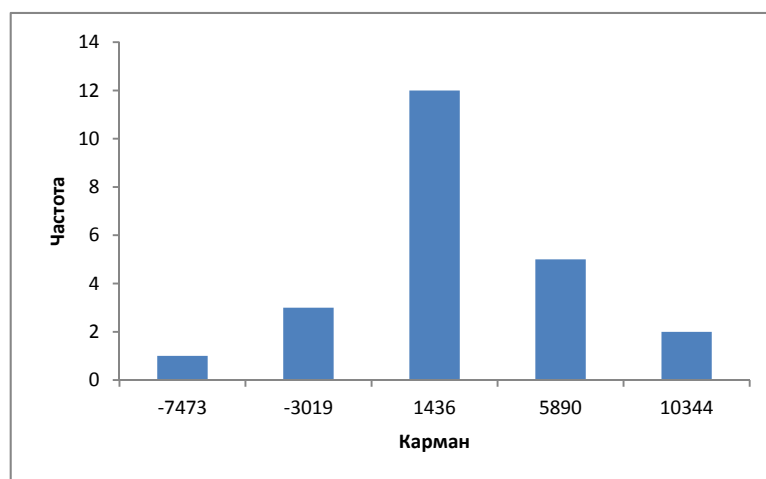
	A8	A7	A6	A5	A4	A3	A2	A1	A0
Функция ЛИНЕЙН():	2204,0	20190,2	8832,4	1701,4	7339,2	1052,9	1233,9	5545,4	-6770,6
	484,4	4561,0	907,4	1090,3	2725,0	720,3	286,6	1385,3	12309,6
R-квадрат	0,946	5414,1	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
	30,43	14	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
	7,137E09	41,04E07	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
t Stat	4,550	4,427	9,734	1,560	2,693	1,462	4,305	4,003	
P-value	5E-04	0,001	1E-07	0,141	0,017	0,166	0,001	0,001	
Фактор значим на	100%	99,9%	100%	85,9%	98,3%	83,4%	99,9%	99,9%	
Знак коэф-та.	логичен	логичен	логичен	логичен	логичен	логичен	логичен	логичен	

⁵⁵ Дополнительно: средняя ошибка аппроксимации равна 3,4%, максимальная – 12,4%.

График остатков регрессионной модели (Центральный район)



Гистограмма остатков⁵⁶ модели (Центральный район).



⁵⁶ Коэффициент асимметрии — 0,63; коэффициент эксцесса — 0,62; отношение Интервал/СКО — 4,13.

Таблица П7 «Генеральная совокупность» квартир в Приморском районе Петербурга⁵⁷.

№	Адрес	Цена, тыс. руб.	Расстояние до метро	Состояние	Этаж	Тип дома	S общая	S жилая	S кухни
1	Авиаконструкторов пр. 14	4490	1500	Хор.	Ср.	504Д	58	30	11,5
2	Авиаконструкторов пр. 16	5250	1575	Хор.	Ср.	К/М	64,4	31,6	18,3
3	Авиаконструкторов пр. 20	5500	1870	Хор.	Ср.	137	68	31,4	11
4	Авиаконструкторов пр. 27 к. 2	4120	1610	Хор.	Ср.	504	45,3	28	7
5	Авиаконструкторов пр. 38	5850	2890	Евро	Ср.	К	62	33	9
6	Байконурская ул. 28	5630	1520	Отл.	Ср.	К	59	33	10,5
7	Богатырский пр. 36	5900	1200	Отл.	Ср.	К/М	65	37	9
8	Богатырский, 53, к.1	4900	2150	Хор.	Ср.	600.11	60	32,9	11
9	Долгоозерная ул. 1	4600	1250	Хор.	Ср.	К	49,1	29,5	8
10	Ильюшина ул. 1	4900	80	Хор.	Ср.	137	60	32	10,5
11	Камышовая ул. 40 к. 1	6100	2120	Евро	Ср.	К	66	37	12
12	Камышовая ул. 6 к. 1	4900	440	Отл.	Ср.	137	52,7	29	10,7
13	Камышовая ул. 9 к. 1	4400	380	Т.р.	Пос.	137	55,2	32,4	10,2
14	Комендантский пр. 16 к. 1	5630	190	Хор.	Ср.	121	62	31,9	12,7
15	Комендантский пр. 21 к. 1	5800	600	Отл.	Ср.	К	56	30	9
16	Комендантский пр. 21 к. 2	4550	710	Хор.	Ср.	137	58	31	7,4
17	Комендантский пр. 21-2	4800	710	Хор.	Пос.	137	52,5	30,2	7,8
18	Комендантский пр. 22	4350	490	Хор.	Ср.	137	52	31	8
19	Комендантский пр. 27 к. 2	4250	1320	Отл.	Ср.	504	45,2	28,3	7
20	Королева пр. 47 к. 1	5500	1700	Отл.	Ср.	137	63	34	13
21	Королева пр. 49 к. 1	4600	1840	Хор.	Ср.	137	57	32,5	10
22	М. Новикова ул. 2/1	4630	550	Хор.	Пос.	137	51,7	29,6	7,9
23	М. Новикова ул. 2/1	5100	550	Отл.	Пос.	137	52	28	7
24	Планерная ул. 79	4700	3300	Евро	Ср.	504Д	57,5	30	20
25	Сизова пр. 20 к. 1	5600	930	Хор.	Ср.	137	56	32,6	10
26	Сизова пр. 19	5700	1180	Отл.	Ср.	К/М	59	34	10,5
27	Туристская ул. 38	5100	2450	Отл.	Пос.	137	55,2	35	15
28	Уточкина ул. 1 к. 1	4900	380	Отл.	Ср.	137	56,4	32,4	10,1
29	Уточкина ул. 6 к. 1	4400	550	Хор.	Пос.	504	45,2	28,4	7
30	Уточкина ул. 6 к. 1	4400	550	Ср.	Ср.	504	48	28,2	7,2
31	Шуваловский пр., 51 к.2	4600	2600	Т.р.	Ср.	600.11	59	32,6	9
ОО	Комендантский пр. 28 к. 1	?	830	Отл.	Ср.	К	58,7	32,6	9,7

⁵⁷ Источник данных: портал www.bn.ru, май 2012 г.

Принятые сокращения:

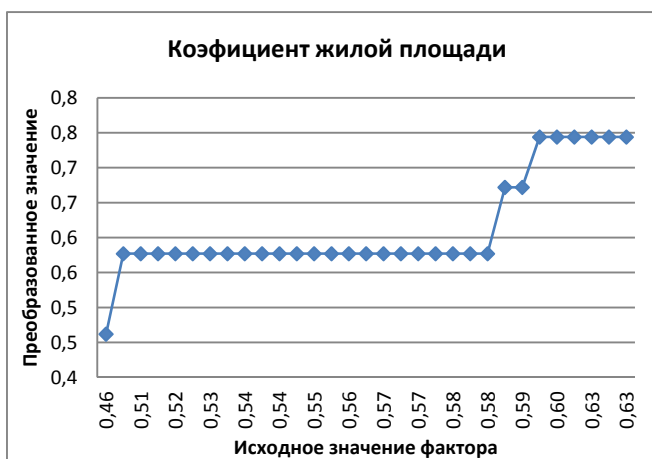
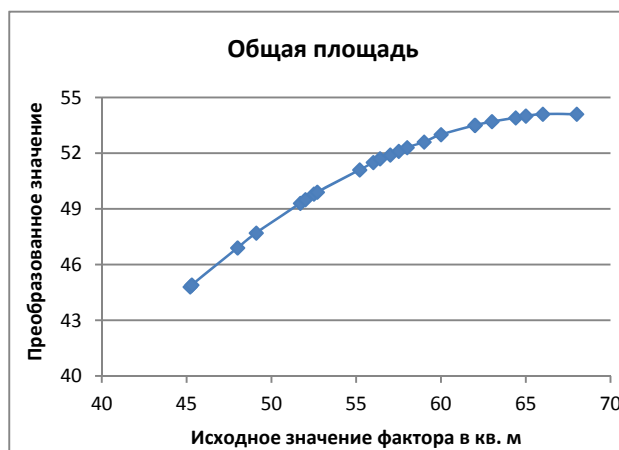
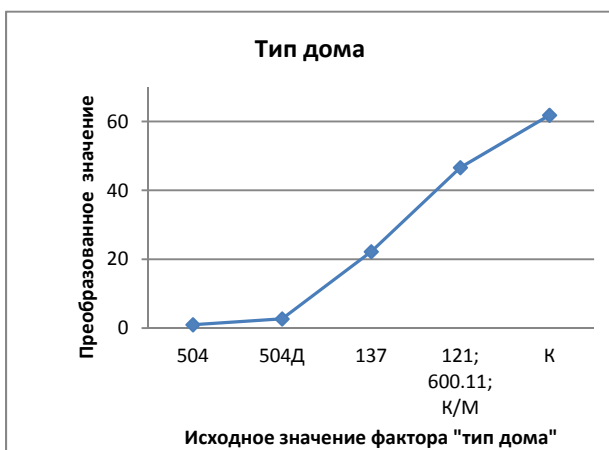
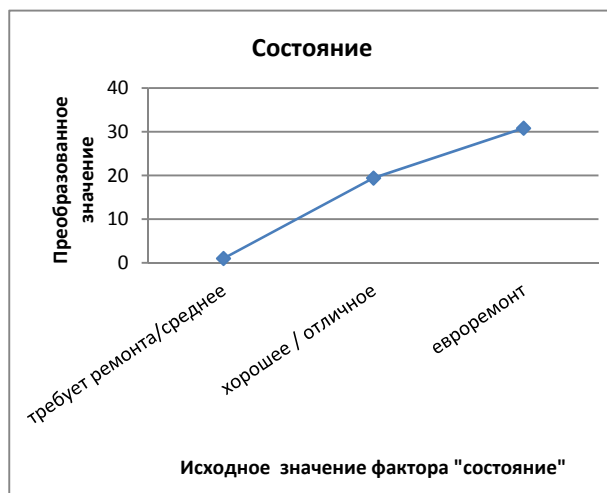
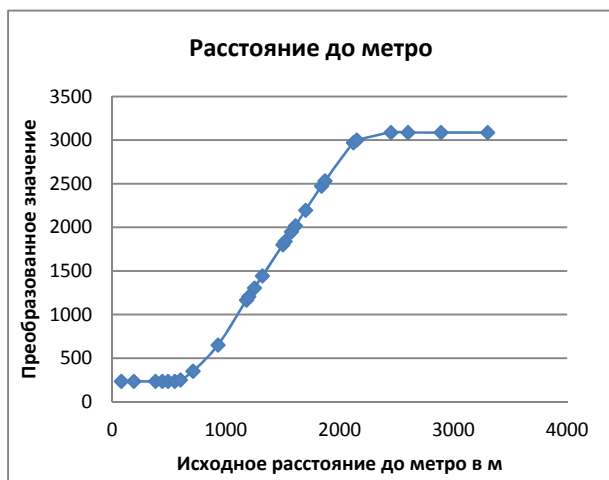
Тип здания: К/М – кирпично-монолитный, К – кирпичный, число – номер серии ;

Этаж: Ср. – средний, Пос. – последний;

Состояние: Т.р. – требует ремонта, Уд. – удовлетворительное, Хор. – хорошее, Отл. – отличное, Евро - евроремонт;

Цена измеряется в тыс. руб., расстояние до метро - в метрах, площади - в кв. м.

Принятые в модели зависимости средней удельной цены от влияющих факторов (Приморский район):



Примечание: Для факторов «Коэффициент жилой площади» и «Площадь кухни» использовались негладкие, «ступенчатые» зависимости. Модели, полученные с такой заменой переменных можно корректно применять лишь для объектов, исходные значения параметров которых находятся в пределах ступеней (не в промежутках между ними), так как имеется неоднозначность в определении преобразованных значений.

Объекты сравнения и их оцифрованные характеристики (Приморский район Петербурга)⁵⁸.

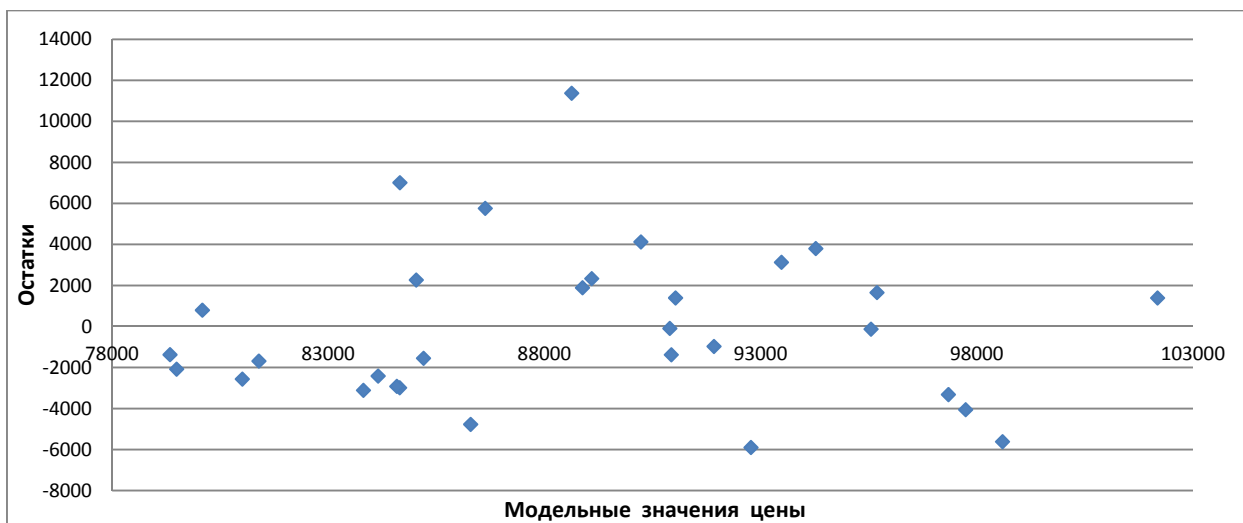
№ аналога	Адрес	Y	X1	X2	X3	X4	X5	X6
		Цена 1 кв.м.	Расстояние до метро	Состояние	Тип дома	Общая площадь	Кэф. жилой площади	Площадь кухни
1	Авиаконструкторов пр. 14	77413,8	1798,9	19,4	2,7	52,3	0,577	12,8
2	Авиаконструкторов пр. 16	81521,7	1947,5	19,4	46,6	53,9	0,577	15,4
3	Авиаконструкторов пр. 20	80882,4	2531,8	19,4	22,2	54,1	0,462	12,8
4	Авиаконструкторов пр. 27 к. 2	90949,2	2016,8	19,4	1,0	44,9	0,744	7,0
5	Авиаконструкторов пр. 38	94354,8	3086,5	30,8	61,8	53,5	0,577	12,8
6	Байконурская ул. 28	95423,7	1838,5	30,8	61,8	52,6	0,577	12,8
7	Богатырский пр. 36	90769,2	1204,6	30,8	46,6	54,0	0,577	12,8
8	Богатырский, 53, к.1	81666,7	3000,0	19,4	46,6	53,0	0,577	12,8
9	Долгоозерная ул. 1	93686,4	1303,7	19,4	61,8	47,7	0,744	7,0
10	Ильюшина ул. 1	81666,7	234,0	19,4	22,2	53,0	0,577	12,8
11	Камышовая ул. 40 к. 1	92424,2	2970,0	30,8	61,8	54,1	0,577	15,4
12	Камышовая ул. 6 к. 1	92979,1	234,0	30,8	22,2	49,9	0,577	12,8
13	Камышовая ул. 9 к. 1	79710,1	234,0	1,0	22,2	51,1	0,672	12,8
14	Комендантский пр. 16 к. 1	90806,5	234,0	19,4	46,6	53,5	0,577	15,4
15	Комендантский пр. 21 к. 1	103571,4	250,0	30,8	61,8	51,5	0,577	12,8
16	Комендантский пр. 21 к. 2	78448,3	350,0	19,4	22,2	52,3	0,577	7,0
17	Комендантский пр. 21-2	91428,6	350,0	19,4	22,2	49,8	0,577	7,0
18	Комендантский пр. 22	83653,8	234,0	19,4	22,2	49,5	0,744	7,0
19	Комендантский пр. 27 к. 2	94026,5	1442,3	30,8	1,0	44,8	0,744	7,0
20	Королева пр. 47 к. 1	87301,6	2195,1	30,8	22,2	53,7	0,577	15,4
21	Королева пр. 49 к. 1	80701,8	2472,4	19,4	22,2	51,9	0,577	12,8
22	М. Новикова ул. 2/1	89555,1	234,0	19,4	22,2	49,3	0,577	7,0
23	М. Новикова ул. 2/1	98076,9	234,0	30,8	22,2	49,5	0,577	7,0
24	Планерная ул. 79	81739,1	3086,5	30,8	2,7	52,1	0,577	15,4
25	Сизова пр. 20 к. 1	100000,0	650,0	19,4	22,2	51,5	0,577	12,8
26	Сизова пр. 19	96610,2	1165,0	30,8	46,6	52,6	0,577	12,8
27	Туристская ул. 38	92391,3	3086,5	30,8	22,2	51,1	0,744	15,4
28	Уточкина ул. 1 к. 1	86879,4	234,0	30,8	22,2	51,7	0,577	12,8
29	Уточкина ул. 6 к. 1	97345,1	234,0	19,4	1,0	44,8	0,744	7,0
30	Уточкина ул. 6 к. 1	91666,7	234,0	1,0	1,0	46,9	0,672	7,0
31	Шуваловский пр., 51 к.2	77966,1	3086,5	1,0	46,6	52,6	0,577	12,8
ОО	Комендантский пр. 28 к. 1	?	513,6	30,8	61,8	52,5	0,6	12,8

⁵⁸ Коэффициент асимметрии распределения цен — -0,02; коэффициент эксцесса — -1,03; отношение Интервал/СКО — 4,49.

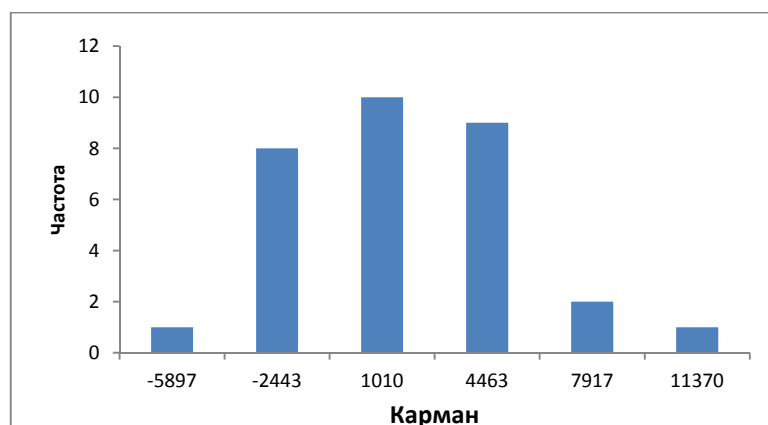
Показатели регрессионного уравнения⁵⁹, полученного по генеральной совокупности:

	A6	A5	A4	A3	A2	A1	A0
Функция ЛИНЕЙН():	967,43	-30446,93	-3232,58	221,94	349,91	-1,93	249833,74
	445,0	18495,1	754,2	49,0	92,5	0,8	44092,0
<i>R-квадрат</i>	0,716	4304,1	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
	10,06	24	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
	1118587565	444600245	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
<i>t Stat</i>	2,17	1,65	4,29	4,53	3,78	2,38	5,67
<i>P-value</i>	0,040	0,113	3E-04	1E-04	0,001	0,026	8E-06
Фактор значим на	96,0%	88,7%	100%	100%	99,9%	97,4%	100%
Знак коэф-та	логичен	логичен	логичен	логичен	логичен	логичен	

График остатков регрессионной модели (Приморский район)



Гистограмма остатков⁶⁰ модели (Приморский район)



⁵⁹ Дополнительно: средняя ошибка аппроксимации равна 3,4%, максимальная – 12,8%.

⁶⁰ Коэффициент асимметрии – 0,92; коэффициент эксцесса – 1,2; отношение Интервал/СКО – 4,5.