

ПРОКОНЬКИНА М. Ю.

**КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ ЖИЛИЩНЫХ УСЛОВИЙ
НАСЕЛЕНИЯ РЕСПУБЛИКИ МОРДОВИЯ**

Аннотация. Одной из задач статистического исследования является изучение взаимосвязей между явлениями и процессами, с целью определения факторов, оказывающих основное влияние на их вариацию. В статье проведен анализ обеспеченности населения региона жильем на основе корреляционно-регрессионного анализа. Установлено, что наиболее значимыми факторами являются индекс цен на первичном рынке жилья, а также размер тарифов на услуги ЖКХ.

Ключевые слова: рынок жилья, обеспеченность населения жильем, индекс цен на первичном рынке жилья, размер тарифов на услуги ЖКХ, корреляционно-регрессионный анализ.

PROKONKINA M. Yu.

**CORRELATION-REGRESSION ANALYSIS OF HOUSING CONDITIONS OF
MORDOVIA REPUBLIC POPULATION**

Abstract. One of the objectives of a statistical analysis is to study statistical relationships between phenomena and processes in order to identify factors that have a major impact on their variation. The article analyzes housing per capita in Mordovia Republic. The study is based on correlation and regression analysis. The author has proved the price index in primary housing market and the utility charges rate to be the key factors of housing per capita in Mordovia Republic.

Keywords: housing market, housing per capita, price index in primary housing market, utility charges rate, correlation and regression analysis.

Социально-экономические явления редко развиваются индивидуально, а чаще всего образуют целые комплексы, где одно самостоятельное явление влияет на другие или само подчиняется влиянию других явлений. Для достоверного отображения существующих социально-экономических явлений необходимо выявить имеющиеся взаимосвязи и дать им количественную оценку. Поэтому одна из задач статистического исследования состоит в изучении связей между наблюдаемыми явлениями и процессами, что позволяет определить факторы (признаки), оказывающие основное влияние на вариацию изучаемых явлений и процессов. Такая задача чаще всего решается методами корреляционно-регрессионного анализа, который включает в себя измерение тесноты, направления связи и установление аналитического выражения (формы) связи. Для этого строятся регрессионные модели влияния факторов на исследуемый процесс или явление.

Корреляционно-регрессионный анализ в рамках данного исследования проведен для выявления основных факторов, оказавших влияние на жилищные условия населения Республики Мордовия в 2000-2011 гг., характеризующиеся обеспеченностью населения жильем.

Абстрактно-логический анализ категорий сферы социально-экономических характеристик жилищных условий населения позволил выделить из массива данных государственной статистики следующие показатели для проведения корреляционно – регрессионного анализа обеспеченности населения региона жильем (Y - площадь жилищ, приходящаяся в среднем на одного жителя (на конец года), кв. м.):

X₁ – индекс цен на первичном рынке жилья, процентов;

X₂ – ввод в действие жилых домов, тыс. кв. м. общей площади;

X₃ – инвестиции в основной капитал организаций, осуществляющих строительную деятельность, в процентах к соответствующему периоду предыдущего года в сопоставимых ценах;

X₄ – реальные располагаемые денежные доходы, в процентах к предыдущему году;

X₅ – численность населения с денежными доходами ниже величины прожиточного минимума, в процентах от общей численности;

X₆ – удельный вес расходов домашних хозяйств на оплату ЖКХ, процентов;

X₇ – уровень зарегистрированной безработицы (зарегистрированные безработные к численности экономически активного населения), процентов;

X₈ – брачность (число браков на 1000 человек населения);

X₉ – удельный вес числа семей, состоявших на учете на получение жилья, в общем числе семей (на конец года), процентов.

Матрица исходных данных представлена в таблице 1.

Таблица 1

Исходные данные для корреляционно-регрессионного анализа

Год	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
2000	19,7	182,1	120,7	117,9	105,1	52,9	5,0	2,1	5,8	13,5
2001	20,0	183,5	116,7	155,9	109,1	49,5	5,6	2,0	6,3	13,3
2002	20,4	144,5	126,0	127,0	112,7	44,6	6,9	1,9	6,3	13,2
2003	21,1	153,7	126,6	167,6	112,2	36,1	7,1	1,8	7,1	10,2
2004	21,5	155,9	134,2	20,4	105,0	32,0	7,3	1,9	6,1	6,4
2005	22,0	180,7	144,9	71,3	110,7	30,1	8,8	1,5	6,6	6,7
2006	22,5	213,2	119,1	220,0	107,4	28,7	7,5	1,4	6,8	6,7
2007	22,9	242,8	124,1	45,6	114,7	24,6	6,9	1,3	7,8	7,3
2008	23,4	284,1	139,2	96,8	118,4	20,4	6,4	1,1	7,1	7,8
2009	23,9	276,6	86,7	51,7	101,5	19,7	7,7	1,7	7,0	8,2
2010	24,0	288,9	98,4	186,0	110,0	18,3	7,5	1,5	7,0	8,7
2011	24,5	298,0	101,9	143,7	96,7	20,2	9,0	1,2	7,7	8,3

Изучение взаимосвязи между явлениями начинается с установления ее тесноты, что производится методами корреляционного анализа. Основой измерения связей является матрица парных коэффициентов корреляции. В ППП «Statistica», используя значения таблицы 1, была получена матрица парных коэффициентов корреляции с указанием вероятности принятия гипотезы об их незначимости. На основании её данных определили зависимости между результативными и факторными признаками (таблица 2).

Таблица 2

Матрица парных коэффициентов корреляции

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
Y	1,00	0,88	-0,45	-0,05	-0,21	-0,97	0,66	-0,82	0,78	-0,70
X1		1,00	-0,56	0,05	-0,20	-0,79	0,31	-0,73	0,68	-0,39
X2			1,00	-0,24	0,61	0,30	-0,14	-0,01	-0,28	-0,11
X3				1,00	-0,01	0,13	-0,07	-0,04	0,03	0,30
X4					1,00	0,07	-0,35	-0,16	0,04	0,04
X5						1,00	-0,68	0,82	-0,75	0,81
X6							1,00	-0,54	0,52	-0,67
X7								1,00	-0,81	0,68
X8									1,00	-0,50
X9										1,00

По значениям, представленным в таблице 2 видно, что в ряде случаев связь является отрицательной, то есть с ростом показателей происходит уменьшение обеспеченности населения региона жильем. Исключение составляют три признака: ввод в действие жилых домов (X₁), удельный вес расходов домашних хозяйств на оплату ЖКХ (X₆) и показатель брачности (X₈). Полученные направления связи в ряде случаев выглядят нелогичными: отрицательная зависимость между Y и X₃, Y и X₄, и Y и X₉. Это объясняется временным лагом, т.е. текущее обеспечение населения жильем удовлетворяет потребности семей, вставших на учет в предыдущие годы (X₆) и инвестиции в организации, осуществляющие строительную деятельность, способствуют росту объемов строительства не в текущем периоде, а по истечении определенного времени. Кроме того, это может быть объяснено тем, что некоторые факторные признаки находятся в зависимости друг от друга, то есть являются мультиколлинеарными – значения коэффициентов корреляции между ними больше 0,8.

Для устранения обозначенных проблем и недопущения искажения коэффициентов уравнения регрессии часть мультиколлинеарных факторов были устранены из дальнейшего анализа (X₅, X₇ и X₉). После этого пересчитали матрицу парных коэффициентов корреляции.

Анализ парных коэффициентов корреляции, указывает на отсутствие мультиколлинеарности между факторами, следовательно, в процедуру последующего

регрессионного анализа следует включить все шесть оставшихся факторов. Процесс пошагового регрессионного анализа состоит из последовательности следующих шагов:

- 1) построение уравнения регрессии;
- 2) проверка значимости уравнения регрессии;
- 3) проверка значимости коэффициентов уравнения регрессии;
- 4) расчёт характеристик регрессии: бета-коэффициентов, коэффициентов эластичности и других.

Для описания влияния выделенных показателей на обеспеченность населения Республики Мордовия жильем строится регрессионная модель. Математически задача формулируется следующим образом: требуется найти аналитическое выражение зависимости экономического явления от определяющих его факторов, то есть необходимо найти функцию:

$$Y = f(X_1, X_2, \dots, X_k) \quad (1)$$

Цель состоит в том, чтобы раскрыть характер и степень влияния аргументов на функцию. Параметры уравнения регрессии находятся методом наименьших квадратов, сущность которого заключается в отыскании таких параметров a_n , для которых сумма квадратов отклонений расчетных значений результативного признака от эмпирических (фактических) будет минимальной. С помощью ППП «Statistica» получили результаты, представленные в таблице 3.

Таблица 3

Предварительные результаты регрессионного анализа

Характеристики параметров регрессионной модели				
Независимые факторы	бета-коэффициенты	Параметры уравнения	t-критерий Стьюдента	Вероятность ошибки t-критерия Стьюдента
Свободный параметр		9,773	2,957	0,032
X1	0,695	0,020	5,055	0,004
X2	-0,080	-0,008	-0,536	0,615
X3	-0,074	-0,002	-0,814	0,452
X4	0,127	0,035	0,914	0,403
X6	0,451	0,645	3,708	0,014
X8	0,047	0,127	0,324	0,759

На основе таблицы 2 составлено предварительное уравнение регрессии:

$$Y = 9,773 + 0,020 \cdot X_1 - 0,008 \cdot X_2 - 0,002 \cdot X_3 + 0,035 \cdot X_4 + 0,645 \cdot X_6 + 0,127 \cdot X_8 \quad (2)$$

(0,032) (0,004) (0,615) (0,452) (0,403) (0,014) (0,759)

$$R = 0.9817 \quad R^2 = 0.9638 \quad F = 22.183$$

Необходимо определить, является ли данное уравнение статистически значимым. Это

можно сделать с применением F-критерия Фишера, рассчитываемого по формуле:

$$F_{расч} = \frac{D_{общ}^2}{D_{ост}^2} \quad (3)$$

Далее полученное по этой формуле значение сравнивается с теоретическим значением, получаемым из специальных таблиц для заданных условий значимости (вероятности ошибки, равной в нашем случае 5%, и степеней свободы k и $n-k-1$). Если фактическое значение оказывается больше табличного, то гипотеза о статистической незначимости уравнения отклоняется в пользу противоположной. Проверка с помощью F-критерия Фишера показала, что уравнение является статистически значимым, так как его расчетное значение оказалось больше табличного ($22,183 > 4,9502$), то есть оно объективно характеризует существующие взаимосвязи. Однако при этом только для одного фактора выполняется проверка на статистическую значимость, которая осуществляется с помощью t-критерия Стьюдента по формуле:

$$t_{a_j} = \frac{a_j}{S_{a_j}} \quad (4)$$

где S_a – это квадратный корень из элементов главной диагонали ковариационной матрицы.

Находится табличное значение статистики Стьюдента при уровне значимости p и степенях свободы ($n-k-1$). Коэффициент регрессии признается значимым, если расчётное значение t-критерия больше табличного.

Последний столбец таблицы 3 показывает, что вероятность для 4 из 6 признаков превышает допустимые 5%, поэтому методом пошаговой регрессии будем исключать из модели по одному факторному признаку. В результате получили новую модель (таблица 4).

Таблица 4

Итоговый результат регрессионного анализа

Характеристики регрессионной модели				
Множественный коэффициент корреляции R		0,9742		
Множественный коэффициент детерминации R ²		0,9491		
F-критерий Фишера Ft		83,923 (4,2564)		
Характеристики параметров регрессионной модели				
Независимые факторы	Параметры уравнения	Бета-коэффициенты	t-критерий Стьюдента	Вероятность ошибки t-критерия Стьюдента
Свободный параметр	13,184		16,266	0,000
X1	0,021	0,750	9,478	0,000
X6	0,614	0,430	5,425	0,000

В результате уравнение регрессии имеет следующий вид:

$$Y = 13,1843 + 0,021 \cdot X_1 + 0,614 \cdot X_6$$
$$(16,266) \quad (9,478) \quad (5,425)$$
$$R = 0.9742 \quad R^2 = 0.9491 \quad F = 83.923$$
(5)

Проведенная проверка с помощью F-критерия Фишера показала, что сравнение является статистически значимым (расчетное значение оказалось больше табличного – $83,923 > 4,2564$). Также статистически значимыми являются и все оставшиеся в модели признаки. Оставшиеся в модели два признака на 94,9% определяют вариацию результативного показателя (об этом свидетельствует величина множественно коэффициента детерминации, равного 0,949). Теснота связи между этими признаками весьма высокая, что подтверждает значение множественного коэффициента корреляции (0,974).

Следовательно, на обеспеченность населения региона жильем оказывают влияние два фактора: индекс цен на первичном рынке жилья (X_1) и удельный вес расходов домашних хозяйств на оплату ЖКХ (X_6). Причем оба фактора оказывают положительное влияние на результативный показатель, т.е. при увеличении индекса цен на первичном рынке жилья на 1 ед., площадь жилищ, приходящаяся в среднем на одного жителя возрастет на 0,021 кв.м. в среднем, а рост удельного веса расходов домашних хозяйств на оплату ЖКХ (X_6) на 1 % приведет к увеличению площади жилищ, приходящейся в среднем на одного жителя на 0,615 кв.м в среднем.

Вычислим нормированные коэффициенты регрессии (коэффициенты эластичности) определяющие меру влияния вариации соответствующего фактора на вариацию результативного признака, и показывающие, на сколько процентов изменится величина обеспеченности жильем, если значение какого-либо из трех признаков увеличится на 1%: для X_1 коэффициент эластичности составляет 2,11%, а для X_6 – 19,79%. Таким образом, на величину площади жилищ, приходящуюся в среднем на одного жителя Республики Мордовия, наибольшее влияние оказывает удельный вес расходов домашних хозяйств на оплату ЖКХ (X_6).

В результате приведенного анализа получили следующие результаты: уравнение и образующие его факторы статистически значимы, весьма высоким является значение коэффициента детерминации (94,9%), т.е. оставшиеся в модели три признака в значительной степени определяют вариацию результативного показателя. Установлено, что жилищные условия населения Республики Мордовия, характеризуемые уровнем обеспеченности жильем, зависят от ценообразования на рынке жилья, характеризующимся индексом цен на первичном рынке жилья и покупательской способности населения региона, определяемой удельным весом расходов домашних хозяйств на оплату ЖКХ.