

УДК 33:311 (045)

А.А. Мухин

ПРИМЕНЕНИЕ МЕТОДОВ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ В ОЦЕНКЕ ФАКТОРНЫХ ЗАВИСИМОСТЕЙ ЧИСЛЕННОСТИ НАСЕЛЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

В методологии применение методов статистического моделирования в оценке факторных зависимостей численности населения Российской Федерации используются шесть параметров: рождаемость, браки, разводы, количество зарегистрированных преступлений, ввод в действие домов, численность врачей в приведенных единицах. Построена математическая модель оценки влияния факторов на параметр – рождаемость, проведен выбор факторных признаков для построения двухфакторной регрессионной модели за 1999–2014 гг. Рассчитаны параметры модели, проведена оценка параметров регрессии с помощью метода наименьших квадратов, оценка качества модели множественной регрессии, оценка точности модели с помощью средней ошибки аппроксимации, применен критерий Дарбина-Уотсона для проверки адекватности выбранного вида уравнения тренда, рассчитаны коэффициент множественной корреляции, коэффициент детерминации R^2 , критерий Фишера, эластичность Y , бета- и дельта-коэффициенты, критерий Стьюдента в оценке статистической значимости коэффициентов уравнения множественной регрессии, доверительный интервал, величина отклонения от линии регрессии, точечный и интервальный прогноз результирующего показателя при $\alpha = 0,1$. Выявлено, что в наибольшей степени на рост населения оказывает влияние такие факторы, как «количество браков» и «ввод в действие домов».

Ключевые слова: модель множественной регрессии, критерий Дарбина-Уотсона, коэффициент детерминации R^2 , критерий Фишера, бета- и дельта-коэффициенты, критерий Стьюдента.

Одним из важнейших инструментов государственного и муниципального управления является экономико-математический аппарат в изучении связей социально-экономических явлений. Используем статистические данные Российской Федерации для *оценки факторных зависимостей численности населения*. Раскроем алгоритм действий:

1. Выбор факторных признаков для построения двухфакторной регрессионной модели.
2. Расчет параметров модели в матричной форме.
3. Оценка точности модели с помощью средней ошибки аппроксимации.
4. Применение критерия Дарбина-Уотсона для проверки адекватности выбранного вида уравнения тренда.
5. Проведение оценки значимости уравнения регрессии.
6. Оценка статистической значимости коэффициентов уравнения множественной регрессии с помощью t-критерия Стьюдента.
7. Оценка влияния факторов на зависимую переменную по модели.
8. Построение точечного и интервального прогноза результирующего показателя на два года вперед ($\alpha = 0,1$).

1. *Осуществление выбора факторных признаков для построения двухфакторной регрессионной модели.*

Рождаемость, тыс. чел. – это зависимая переменная Y . В качестве независимых, объясняющих переменных выбраны: браки X_1 , разводы X_2 , коэффициент преступности X_3 , ввод в действие домов X_4 , численность врачей X_5 . Все данные отражены в относительных единицах: на 1000 чел населения (табл. 2).

Анализ матрицы коэффициентов парной корреляции начинаем с ее первого столбца, где расположены коэффициенты корреляции, отражающие тесноту связи зависимой переменной Y «Рождаемость». Анализ матрицы показывает, что рождаемость (тыс. чел.) имеет прямую тесную связь с коэффициентом браков ($r_{yx_1} = 0,87$), а также с вводом в действие домов, млн. м² ($r_{yx_4} = 0,94$).

В результате перечисленных действий будет получена *матрица коэффициентов парной корреляции* (табл. 3).

Таблица 1

Показатели социального развития России за период 1999-2014 гг. (на 1000 чел населения)

Го- ды	Рождае- мость, на 1000 чел населения Y	Браки, на 1000 чел. населения X_1	Разводы, на 1000 чел. населения X_2	Зарегистрировано преступлений, всего, тыс. X_3	Ввод жилых домов, млн. м ² X_4	Численность врачей, тыс. чел. X_5
1999	8,3	6,2	3,6	124,7	21,7	4,6
2000	8,7	6,2	4,3	201,0	20,6	4,6
2001	9	6,9	5,3	202,9	21,7	4,6
2002	9,7	7,1	5,9	173,5	23,2	4,7
2003	10,2	7,6	5,5	190,1	25,1	4,7
2004	10,4	6,8	4,4	200,7	28,4	4,8
2005	10,2	7,5	4,2	247,8	30,4	4,8
2006	10,4	7,8	4,5	270,1	35,4	4,9
2007	11,3	8,9	4,8	251,9	43,0	5,0
2008	12	8,3	4,9	225,9	45,1	5,0
2009	12,3	7,4	4,2	209,9	42,0	5,0
2010	12,5	8,4	4,9	184,1	40,9	5,0
2011	12,6	8,5	4,5	168,3	43,6	5,1
2012	13,3	9,2	4,7	161,0	45,9	4,9
2013	13,2	8,5	4,5	153,9	49,2	4,9
2014	13,3	8,5	4,7	152,5	58,2	4,9

Таблица 2

Матрица коэффициентов парной корреляции

Факторы	Рождаемость, тыс. чел. Y	Браки, на 1000 чел населения X_1	Разводы, на 1000 чел населения X_2	Зарегистрировано преступлений – всего, тыс X_3	Ввод жилых домов, млн. м ² X_4	Численность врачей, человек, тыс. чел. X_5
Y	1					
X_1	0,87	1				
X_2	0,02	0,20	1			
X_3	-0,15	0,06	0,07	1		
X_4	0,94	0,85	-0,10	-0,08	1	
X_5	0,83	0,80	-0,10	0,17	0,81	1

Таблица 3

Матрица коэффициентов парной корреляции

Факторы	Рождаемость на 1000 чел. населения Y	Браки, на 1000 чел населения X_1	Ввод жилых домов, млн. м ² X_4
Y	1		
X_1	0,87	1	
X_4	0,94	0,85	1

2. Расчет параметров модели в матричной форме.

В результате применения двух подходов к выбору факторов в модель пришли к выводу о необходимости включения двух факторов: браки, ввод в действие домов ($n = 16, k = 2$). Оценка параметров регрессии осуществляется методом наименьших квадратов в матричной форме по формуле:

$$A = (X'X)^{-1}X'Y$$

с использованием данных, приведенных в табл. 4 (для вычисления a_0 добавлен столбец X_0).

Оценку параметров регрессии выполним методом наименьших квадратов, используя функции матричной алгебры.

Таблица 4

Оценка параметров регрессии

№ п.п.	Годы	Рождаемость, на 1000 чел населения Y	Браки, на 1000 чел населения X_1	Ввод жилых домов, млн. м ² X_2
1	1999	8,3	6,2	21,7
2	2000	8,7	6,2	20,6
3	2001	9	6,9	21,7
4	2002	9,7	7,1	23,2
5	2003	10,2	7,6	25,1
6	2004	10,4	6,8	28,4
7	2005	10,2	7,5	30,4
8	2006	10,4	7,8	35,4
9	2007	11,3	8,9	43,0
10	2008	12	8,3	45,1
11	2009	12,3	7,4	42,0
12	2010	12,5	8,4	40,9
13	2011	12,6	8,5	43,6
14	2012	13,3	9,2	45,9
15	2013	13,2	8,5	49,2
16	2014	13,3	8,5	58,2

Используем матричное исчисление:

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \left[\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{2,1} & \dots & x_{16,1} \\ x_{1,2} & x_{2,2} & \dots & x_{16,2} \end{pmatrix} \begin{pmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ \dots & \dots & \dots \\ 1 & x_{16,1} & x_{16,2} \end{pmatrix} \right]^{-1} \times \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{2,1} & \dots & x_{16,1} \\ x_{1,2} & x_{2,2} & \dots & x_{16,2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_{16} \end{pmatrix}$$

$$(X'X) = \begin{pmatrix} 16 & 123,8 & 574,5 \\ 123,8 & 970,8 & 4582,8 \\ 574,5 & 4582,8 & 22661 \end{pmatrix},$$

$$(X'X)^{-1} = \begin{pmatrix} 8,54 & -1,48 & -0,08 \\ -1,48 & 0,28 & -0,02 \\ 0,08 & -0,02 & 0,0017 \end{pmatrix},$$

$$A = (X'X)^{-1} X'Y = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 3,756 \\ 0,453 \\ 0,107 \end{pmatrix}$$

Уравнение регрессии зависимости рождаемости от количества браков и ввода в действие домов можно записать в следующем виде:

$$y = 3,756 + 0,453x_1 + 0,107x_2$$

Расчетные значения Y определяются путем последовательной подстановки в эту модель значений факторов, взятых для каждого наблюдения.

Для проведения регрессионного анализа используем инструменты MS Excel.

В результате перечисленных действий будет получен протокол выполнения регрессионного анализа.

Таблица 5

Протокол выполнения регрессионного анализа

Регрессионная статистика		Дисперсионный анализ				
Множественный R	0,95	df	SS	MS	F	
R -квадрат	0,91	Регрессия	2	39	19,5	61
Нормированный R -квадрат	0,89	Остаток	13	4,1	0,32	
Стандартная ошибка	0,56	Итого	15	43,2		
Наблюдения	16					

	Коэффициенты	Стандартная ошибка	t -статистика
Y -пересечение	3,756	1,65	2,28
Браки, X_1	0,453	0,3	1,52
Ввод в действие жилых домов в расчете на 1000 чел. кв. м, X_2	0,107	0,02	4,49

Рассмотрим содержание протокола регрессионного анализа (табл. 5), в котором отражены основные итоги расчетов. Пояснения таблицы «Регрессионная статистика» приведены в табл. 6, а «Дисперсионный анализ» – в табл. 7.

В протоколе выполнения регрессионного анализа (табл. 5) отображены коэффициенты уравнения регрессии, a_0 , a_1 , a_2 стандартные ошибки коэффициентов, t -статистика, используемая для проверки значимости коэффициентов уравнения регрессии, индекс корреляции, критерий Фишера и другие параметры.

Таблица 6

Регрессионная статистика

Наименование показателя в отчете Excel	Принятые наименования	Формула
Множественный R	Коэффициент множественной корреляции, индекс корреляции	$R = \sqrt{R^2}$
R -квадрат	Коэффициент детерминации R^2	$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
Нормированный R -квадрат	Скорректированный R^2	$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}$
Стандартная ошибка	Среднеквадратическое отклонение от модели	$S_e = \sqrt{\frac{1}{n-k-1} \sum_{i=1}^n e_i^2}$
Наблюдения	Количество наблюдений n	n

Таблица 7

Дисперсионный анализ

	<i>df</i> – число степеней свободы	SS – сумма квадратов	MS – среднее значение	<i>F</i> -критерий Фишера
Регрессия	$k = 2$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / k$	$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$
Остаток	$n - k - 1 = 4$	$\sum_{i=1}^n e_i^2$	$\sum_{i=1}^n e_i^2 / (n - k - 1)$	
Итого	$n - 1 = 6$	$\sum_{i=1}^n (y_i - \bar{y})^2$		

Данные матричного исчисления и протокол уравнения регрессии совпали. Зависимость валового регионального продукта от численности работников и основных фондов можно записать в следующем виде:

$$y = 3,756 + 0,453x_1 + 0,107x_2$$

Оценим качество модели.

Значение коэффициентов детерминации и множественной корреляции можно найти в таблице «Регрессионная статистика» (табл. 6).

Коэффициент детерминации:

$$R^2 = R^2_{y;x_1,x_2} = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{39,25}{43,16} = 0,91.$$

Коэффициент детерминации показывает, что около 91 % вариации зависимой переменной учтено в модели и обусловлено влиянием включенных факторов.

Коэффициент множественной корреляции:

$$R = \sqrt{R^2} = \sqrt{0,91} = 0,95$$

Он показывает высокую тесноту связи зависимой переменной *Y* с двумя включенными в модель объясняющими факторами.

3. Точность модели оценим с помощью средней ошибки аппроксимации:

Наблюдение	<i>y</i>	\hat{y}_x	Остатки $ y - \hat{y}_x $	$\frac{ y - \hat{y}_x }{y} \cdot 100$
1	8,3	8,9	0,6	7,1
2	8,7	8,8	0,1	0,8
3	9	9,2	0,2	2,3
4	9,7	9,5	0,2	2,5
5	10,2	9,9	0,3	3,1
6	10,4	9,9	0,5	5,0
7	10,2	10,4	0,2	2,0
8	10,4	11,1	0,7	6,5
9	11,3	12,4	1,1	9,6
10	12	12,3	0,3	2,8
11	12,3	11,6	0,7	5,7
12	12,5	11,9	0,6	4,5
13	12,6	12,3	0,3	2,6
14	13,3	12,8	0,5	3,5
15	13,2	12,9	0,3	2,5
16	13,3	13,8	0,5	4,0
Сумма				64,6

$$E_{\text{отн}} = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{y_i} \cdot 100\% = \frac{64,6}{16} = 4\%$$

Вывод: $E_{\text{отн}} = 4\%$ – хороший уровень точности модели.

4. *Использование критерия Дарбина-Уотсона для проверки адекватности выбранного вида уравнения тренда.*

При проведении аналитического выравнивания зачастую бывает трудно заранее определить подходящий вид уравнения тренда, особенно если эмпирические данные графически явно не демонстрируют свою приверженность к какой-либо аналитической функции. Тогда поступают следующим образом: строят несколько уравнений тренда, затем для каждого из них вычисляют остаточную дисперсию и модель с наименьшей величиной остаточной дисперсии признают лучшей из имеющихся на данный момент. Но можно пойти и другим путем, а именно: проверить остаточные величины (отклонения фактических значений показателя от выравненных по уравнению тренда) на автокорреляцию. Под *автокорреляцией* здесь понимается корреляционная зависимость остаточных величин. Если она отсутствует, значит, подобранная модель полностью описывает содержащуюся закономерность в изменении показателя, и, следовательно, адекватно отражает тенденцию в его развитии.

$$d = \frac{\sum_{t=1}^n (\varepsilon_{t+1} - \varepsilon_t)^2}{\sum_{t=1}^n \varepsilon_t^2},$$

где ε_t – отклонения фактического уровня ряда от его выравненного значения ($\varepsilon_t = y_t - \bar{y}_t$).

Величина d изменяется от 0 до 4. При $d = 2$ автокорреляция в остаточных величинах отсутствует, следовательно, модель адекватна. При $d = 0$ имеется полная положительная, а при $d = 4$ – полная отрицательная автокорреляция.

Реально расчетные значения будут находиться между этими границами. Сделать конечный вывод можно с помощью дальнейшей проверки гипотезы об отсутствии автокорреляции в остатках. Для этого расчетное значение d сравнивается с критическими значениями критерия.

Таблица 8

Расчетная таблица для вычисления значения критерия Дарбина-Уотсона

Условное обозначение времени t	Год	y	\hat{y}_x	$\varepsilon_t = y_t - \bar{y}_t$	ε_{t+1}	ε_t^2	$\varepsilon_{t+1} - \varepsilon_t$	$(\varepsilon_{t+1} - \varepsilon_t)^2$
-8	1999	8,3	8,9	-1	0	0	0	0
-7	2000	8,7	8,8	0	0	0	0	0
-6	2001	9	9,2	0	0	0	0	0
-5	2002	9,7	9,5	0	0	0	0	0
-4	2003	10,2	9,9	0	1	0	0	0
-3	2004	10,4	9,9	1	0	0	-1	0
-2	2005	10,2	10,4	0	-1	0	0	0
-1	2006	10,4	11,1	-1	-1	0	0	0
1	2007	11,3	12,4	-1	0	1	1	1
2	2008	12	12,3	0	1	0	1	1
3	2009	12,3	11,6	1	1	0	0	0
4	2010	12,5	11,9	1	0	0	0	0
5	2011	12,6	12,3	0	1	0	0	0
6	2012	13,3	12,8	1	0	0	0	0
7	2013	13,2	12,9	0	-1	0	-1	1
8	2014	13,3	13,8	-1	0	0	0	0
Сумма		177,4	177,7			4,2		3,9

Тогда расчетное значение критерия равно:

$$d = \frac{\sum_{t=1}^n (\varepsilon_{t+1} - \varepsilon_t)^2}{\sum_{t=1}^n \varepsilon_t^2} = \frac{3,9}{4,2} = 0,9.$$

Полученная величина находится в интервале от 0 до 2.

По табл. 9 находим критические значения критерия Дарбина-Уотсона.

Таблица 9

Критические значения критерия Дарбина-Уотсона для 0,05% уровня значимости

Количество уровней ряда <i>n</i>	Нижняя граница <i>d</i> ₁	Верхняя граница <i>D</i> ₂
16	1,10	1,37

Расчетное значение *d* находится в интервале от 0 до 2, при *d* < *d*₁ имеется положительная автокорреляция.

Положительная автокорреляция вызывается направленным постоянным воздействием некоторых неучтенных в модели факторов. Положительная автокорреляция означает, что отклонения от равновесия имеют тенденцию сохраняться от периода к периоду.

5. Осуществление оценки значимости уравнения регрессии.

Оценим значимость уравнения регрессии. Проверку значимости проведем на основе вычисления *F*-критерия Фишера:

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)} = \frac{0,91 / 2}{(1 - 0,91) / (16 - 2 - 1)} = 65.$$

Табличное значение *F*-критерия можно найти с помощью функции *F*.ОБР.ПХ (2010 Excel). При доверительной вероятности 0,9 при *v*₁ = *k* = 2 и *v*₂ = *n* - *k* - 1 = 16 - 2 - 1 = 13 оно составляет 2,67.

В нашем примере *F*_{расч.} > *F*_{табл.}, то есть 65 > 2,67, и уравнение регрессии следует признать адекватным.

*6. Оценка с помощью *t*-критерия Стьюдента статистическую значимость коэффициентов уравнения множественной регрессии.*

Анализ статистической значимости параметров модели (коэффициентов регрессии) проводится с использованием *t*-статистики путем проверки нулю *j*-го параметра уравнения (кроме свободного члена).

$$t_{a_j} = \hat{a}_j / S_{a_j},$$

где *S*_{*a*_{*j*}} – это стандартное (среднеквадратическое) отклонение коэффициента уравнения регрессии *a*_{*j*}.

Величина *S*_{*a*_{*j*}} представляет собой квадратный корень из произведения несмещенной оценки дисперсии *S*_{*ε*}² и *j*-го диагонального элемента матрицы, обратной матрице системы нормальных уравнений:

$$S_{a_j} = \sqrt{b_{jj}},$$

где *b*_{*jj*} – диагональный элемент матрицы (*X*'*X*)⁻¹.

Значения *t*-критерия для коэффициентов уравнения регрессии *a*₀, *a*₁, *a*₂ вычислим по формуле

$$t_{a_j} = \hat{a}_j / S_{a_j} = \hat{a}_j / (S_{\epsilon} \sqrt{b_{jj}}) \quad (1)$$

$$(X'X)^{-1} = \begin{pmatrix} 27,27 & -2,06 & -0,00383 \\ -2,06 & 0,19 & 0,00019 \\ -0,038 & 0,00019 & 0,0000008 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 8,54 & -1,48 & -0,08 \\ -1,48 & 0,28 & -0,02 \\ 0,08 & -0,02 & 0,0017 \end{pmatrix},$$

$$b_{11} = 8,54, b_{22} = 0,28, b_{33} = 0,0017$$

$$t_{a_j} = \hat{a}_j / S_{a_j} = \hat{a}_j / (S_e \sqrt{b_{jj}});$$

$$t_{a_0} = 3,756 / (0,56\sqrt{8,54}) = 2,3,$$

$$t_{a_1} = 0,453 / (0,56\sqrt{0,28}) = 1,53,$$

$$t_{a_2} = 0,107 / (0,56\sqrt{0,0017}) = 4,6.$$

Табличное значение t -критерия Стьюдента можно найти с помощью функции СТЬЮДРАСПОБР (0,1; 13) (Excel 2010); при 10 %-м уровне значимости и 13 степенях свободы ($16 - 2 - 1 = 13$) оно составляет 1,77.

Так как $|t_{\text{расч.}}| > t_{\text{табл.}}$, то коэффициенты a_0, a_2 существенны (значимы), коэффициент a_1 находится рядом с критическим значением. Будем считать, что все коэффициенты существенны.

7. Оценка влияния факторов на зависимую переменную по модели.

Для каждого коэффициента регрессии вычислим коэффициент эластичности и бета-коэффициент соответственно по формулам (2) и (3):

$$\mathcal{E}_j = \hat{a}_j \bar{x}_j / \bar{y} \quad (2)$$

$$y = 3,756 + 0,453x_1 + 0,107x_2,$$

$$\mathcal{E}_1 = 3,756 \cdot 0,453 / 11,09 = 0,32,$$

$$\mathcal{E}_2 = 0,107 \cdot 35,91 / 11,09 = 0,34,$$

$$\beta_j = \hat{a}_j S_{x_j} / S_y, \quad (3)$$

$$\beta_1 = 0,453 \cdot 0,93 / 1,7 = 0,23,$$

$$\beta_2 = 0,107 \cdot 11,64 / 1,7 = 0,73.$$

Коэффициент эластичности \mathcal{E}_1 показывает, что при увеличении браков на 1 %, рождаемость увеличится на 0,32 тыс. чел., а коэффициент β_1 показывает, что при увеличении количества браков на 0,93 (на 1000 чел населения), рождаемость увеличится на 0,4 тыс. чел. ($0,23 \cdot 1,7 = 0,4$).

Коэффициент эластичности \mathcal{E}_2 показывает, что при увеличении ввода в действие домов на 1 %, рождаемость увеличится на 0,34 тыс. чел., а коэффициент β_2 показывает, что при увеличении браков на 11,64 (на 1000 человек населения), количество рождаемости увеличится 1,24 тыс. ($0,73 \cdot 1,7 = 1,24$).

Долю влияния каждого фактора в суммарном влиянии всех факторов можно оценить по величине дельта-коэффициента Δ_j :

$$\Delta_j = \frac{r_{y,x_j} \beta_j}{R^2},$$

$$\Delta_1 = \frac{0,87 \cdot 0,23}{0,91} = 0,25,$$

$$\Delta_2 = \frac{0,94 \cdot 0,73}{0,91} = 0,75.$$

8. Построение точечного и интервального прогноза результирующего показателя на два года вперед ($\alpha = 1$).

Прогнозируемое значение переменной получается при подстановке в уравнение регрессии ожидаемых значений объясняющих факторов X .

В нашей задаче необходимо построить прогноз объема реализации на два года вперед. (y_{17}, y_{18}).

Уравнение аппроксимации полиномиального тренда 5-й степени рождаемости в РФ.

$$y = -1,005x^5 - 0,0011x^4 + 0,0234x^3 - 0,2036x^2 + 1,0709x - 7,3125.$$

Надежность уравнения аппроксимации: $R^2 = 0,98$.

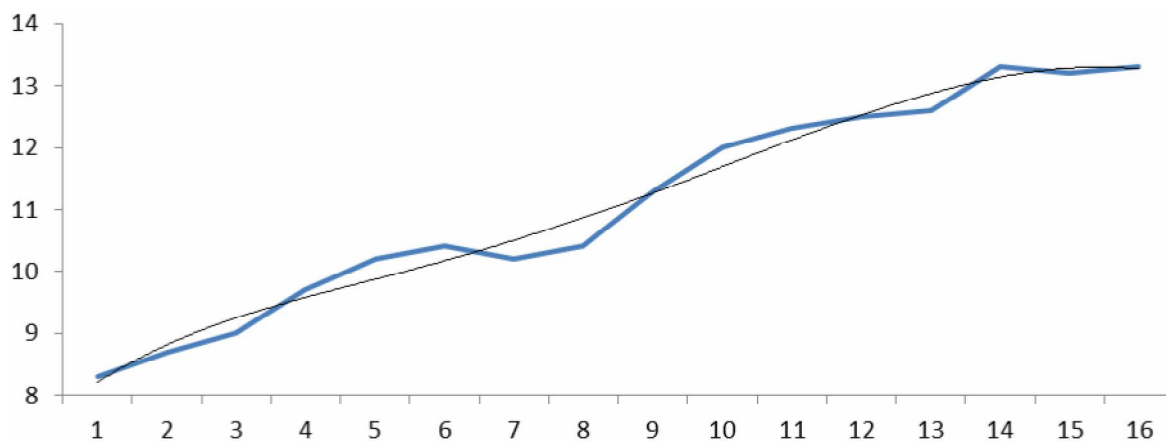


Рис. 1. Динамика рождаемости в РФ 1999–2014 гг., тыс. чел.

Для получения прогнозных оценок зависимостей переменных по модели на два года вперед

$$y = -1,005x^5 - 0,0011x^4 + 0,0234x^3 - 0,2036x^2 + 1,0709x - 7,3125$$

подставим в нее найденные прогнозные значения факторов x_1 и x_2 .

Упреждение	Прогноз, тыс. чел
1	11,5
2	12,1

На рис. 2 приведен результат построения тренда для временного ряда «Ввод в действие жилых домов в расчете на 1000 чел, кв. м».

$$X_2 = 0,001x^5 - 0,0342x^4 + 0,3691x^3 - 1,1446x^2 + 0,8275x + 21,62.$$

Надежность уравнения аппроксимации: $R^2 = 0,98$; по данному уравнению построен прогноз на два шага вперед:

Упреждение	Прогноз
1	40,1
2	42,5

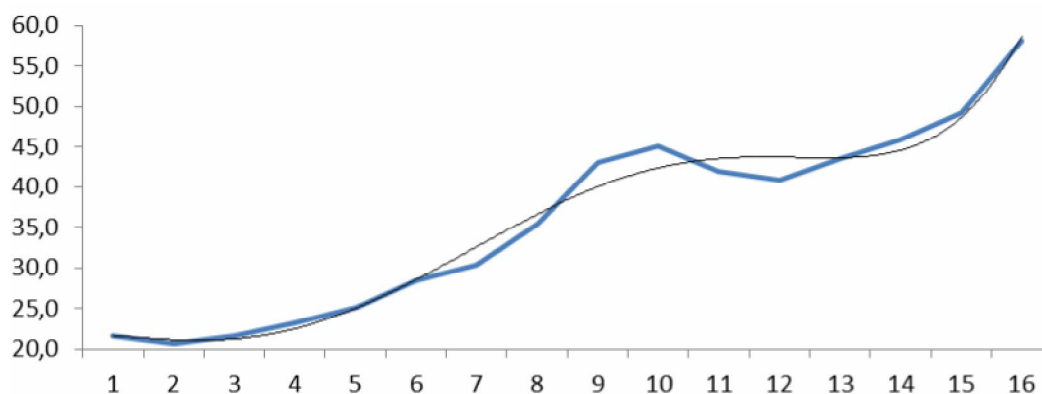


Рис. 2. Динамика ввода в действие жилых домов в расчете на 1000 чел, кв. м. в РФ 1999–2014 гг., млн. м²

Для получения прогнозных оценок зависимостей переменных по модели

$$y = 3,756 + 0,453x_1 + 0,107x_2$$

подставим в нее найденные прогнозные значения факторов X_1 и X_2 :

$$y_{17} = 3,756 + 0,453 \cdot 11,5 + 0,107 \cdot 40,1 = 13,26,$$

$$y_{18} = 3,756 + 0,453 \cdot 12,1 + 0,107 \cdot 42,5 = 13,78.$$

Доверительный интервал прогноза будет иметь следующие границы:

$$Y_{\text{прогн}(n+l)} + U_{(l)} \text{ (верхняя граница),}$$

$$Y_{\text{прогн}(n+l)} - U_{(l)} \text{ (нижняя граница),}$$

где $U(l) = S_{\varepsilon} t_{\alpha} \sqrt{1 + X'_{\text{прогн}} (X'X)^{-1} X_{\text{прогн}}}$, $S_{\varepsilon} = 0,565$.

Распределение Стьюдента при 16 степенях свободы и 90 % точности:

$$t_{\alpha} = \text{СТЮДРАСПОБР}(0,1; 13) = 1,77.$$

На первый шаг ($l = 1$):

$$X'_{\text{прогн}} = (1; 11,5; 40,1);$$

$$(X'X)^{-1} = \begin{pmatrix} 8,54 & -1,48 & 0,08 \\ -1,48 & 0,29 & -0,018 \\ 0,08 & -0,018 & 0,001 \end{pmatrix};$$

$$U(1) = 2,1.$$

На второй шаг ($l = 2$):

$$X'_{\text{прогн}} = (1; 11,5; 40,1);$$

$$U(2) = 2,3.$$

Результаты прогнозных оценок модели регрессии при $\alpha = 1$ представим в табл. 10.

Таблица 10

Упреждение	Прогноз	Нижняя граница	Верхняя граница
1	13,3	11,2	15,4
2	13,8	11,5	16,1

С вероятностью 90 % можно утверждать, что рождаемость (на 1000 чел. населения) в следующие два года будет лежать в пределах, указанных в табл. 10.

Таким образом, в наибольшей степени на рождаемость и одновременно на численность населения РФ оказывают влияние такие факторы, как браки и ввод в действие жилых домов.

СПИСОК ЛИТЕРАТУРЫ

1. Постановление Правительства РФ от 02.06.2008 № 420 «О Федеральной службе государственной статистики» // Собрание законодательства РФ, 09.06.2008, № 23, ст. 2710. URL: <http://www.gks.ru>.
2. Войтович В.Ю. Модернизация правового регулирования местного самоуправления: проблемы, теория и практика // Вестн. Удм. ун-та. Сер. «Экономика и право». 2013. № 2-3. С. 128-131.
3. Войтович В.Ю., Войтович В.В. Роль экспертных исследований в гражданском и арбитражном процессах. Ижевск: КнигоГрад, 2009.
4. Основы теории права учеб.-метод. пособие/ В.Ю Войтович. Ижевск, 2005.
5. Мухина И.А. Социально-экономическая статистика: учебник. М.:ФЛИНТА, 2011.
6. Мухин А.А. Статистические методы в оценке динамики численности населения Удмуртской Республики / А.А. Мухин, И.А. Мухина // Региональная экономика: теория и практика. 2009. № 12. С. 58-59.
7. Орлова И.В., Половников В.А. Экономико-математические методы и модели: компьютерное моделирование: учеб. пособие. 3-е изд., перераб. и доп. М.: НИЦ ИНФРА-М, 2013. 389 с.

Поступила в редакцию 18.03.16

A.A. Mukhin

THE APPLICATION OF STATISTICAL MODELING METHODS TO THE ESTIMATION OF FACTOR DEPENDENCIES OF THE RUSSIAN FEDERATION POPULATION

In the methodology of the application of statistical modeling methods to the estimation of factor dependencies of the Russian Federation population there are six parameters: births, marriages, divorces, the number of reported crimes, the commissioning of houses, and the number of doctors in the reduced units. A mathematical model of the factors' impact on the parameter – the birth rate – has been developed; the factor signs for constructing a two-factor regression model for the period 1999-2014 have been chosen. The parameters of the model have been calculated; the estimation of regression parameters using the method of least squares has been carried out; the quality of multiple regression model has been evaluated; the model accuracy with an average error of approximation has been estimated; the criterion of Durbin-Watson to check the adequacy of the selected type of the trend equation has been used. The following values have been calculated: the coefficient of multiple correlation, the coefficient of determination R^2 , Fisher's exact test, elasticity Y , beta and delta coefficients, Student's criterion in the assessment of statistical significance of the coefficients of a multiple regression equation, the confidence interval, the deviation from the regression line, the point and interval prediction of the resulting index at $\alpha=0,1$. It has been revealed that the greatest effect on the population growth is exerted by such factors as the number of marriages and commissioning of houses.

Keywords: multiple regression model, criterion of Durbin-Watson, coefficient of determination R^2 , Fisher's exact test, beta and delta coefficients, Student's criterion.

Мухин Алексей Арьевич,
кандидат юридических наук, доцент
ФГБОУ ВО «Удмуртский государственный университет»
426034, Россия, г. Ижевск, ул. Университетская, 1 (корп. 4)

Mukhin A.A.,
candidate of law, associate professor
Udmurt State University
Universitetskaya st., 1/4, Izhevsk, Russia, 426034