

АНАЛИЗ ФАКТОРОВ, ВЛИЯЮЩИХ НА РОСТ ДОХОДОВ НАСЕЛЕНИЯ В РОССИЙСКИХ РЕГИОНАХ

ANALYSIS OF FACTORS INFLUENCING THE GROWTH OF POPULATION INCOME IN THE RUSSIAN REGIONS

V. Zaernyuk
V. Shenderov
Yu. Zabaykin
M. Kostin

Annotation

In connection with the crisis processes observed in the world and the national economy, the problem of the standard of living of the population, assessing the factors that affect the income of the population, have not only scientific but also practical significance, has recently acquired a particular urgency. In the article, based on the correlation-regression analysis, the key factors determining the greatest impact on the incomes of the population in the context of the Russian regions were investigated and identified. As independent variables, the impact of the gross regional product, investment in fixed assets, the population, the average annual number of employed in the economy, consumer spending per capita on average, fixed assets in the regional economy, agricultural output, the introduction of the total area of residential buildings, turnover of retail trade, financial performance of organizations. The sampled sample is represented by fifty-five Russian regions. The quantitative dependence of the monetary incomes of the population in the context of the federal districts of the Russian Federation is revealed.

Keywords: population incomes, forecast model, correlation analysis.

Заернюк Виктор Макарович
Д.э.н., доцент, Российский государственный
геологоразведочный университет
им. Серго Орджоникидзе

Шендеров Владимир Исаакович
К.т.н., Российский государственный
геологоразведочный университет
им. Серго Орджоникидзе
Забайкин Юрий Васильевич

К.э.н., доцент, Российский государственный
геологоразведочный университет
им. Серго Орджоникидзе
Костин Михаил Петрович
К.э.н., Московский государственный
университет технологий и управления
им. К.Г. Разумовского

Аннотация

В связи с кризисными процессами, наблюдаемыми в мировой и отечественной экономике, проблема, касающаяся уровня жизни населения, оценки факторов, влияющих на получение населением доходов, имеют не только научную, но и практическую значимость, приобретает в последнее время особую актуальность. В статье на основе корреляционно-регрессионного анализа исследованы и выявлены ключевые факторы, обуславливающие наибольшее влияние на доходы населения в разрезе российских регионов. В качестве независимых переменных рассматривалось влияние валового регионального продукта, инвестиций в основной капитал, численности населения, среднегодовой численности занятых в экономике, потребительские расходы в среднем на душу населения, основных фондов в экономике региона, продукции сельского хозяйства, ввода в действие общей площади жилых домов, оборота розничной торговли, финансовых результатов деятельности организаций. Исследуемая выборка представлена пятидесяти пятью российскими регионами. Выявлена количественная зависимость денежных доходов населения в разрезе федеральных округов Российской Федерации.

Ключевые слова:

Доходы населения, прогнозная модель, корреляционный анализ.

Методология

Оценка проводилась методом наименьших квадратов. Проведена проверка набора данных на мультиколлинеарность, проверка однородности данных с помощью теста Чоу, проверка на гомоскедастичность на основе тестов Гольдфельда – Куандта, Бреуша – Пагана и Уайта.

Результаты исследования.

При проверке исходного набора наблюдения на мультиколлинеарность установлена сильная линейная зависимость между факторными переменными, которая была устранена в ходе анализа корреляционной матрицы.

В результате исследования можно утверждать, что доходы населения тесно зависят от численности населения и уровня потребительских расходов.

Статистическая оценка полученного уравнения по отмеченным тестам позволяет с уверенностью на 95 % утверждать о том, что полученная прогнозная модель является статистически значимой, что дает авторам основание рекомендовать её для использования в практике прогнозирования.

Переход к рыночным отношениям в России сильно обострил проблемы распределения доходов. Реализация шоковых экономических реформ спровоцировала крупномасштабные глубокие изменения в национальной экономике.

Поэтому в настоящее время исследования авторитетных ученых 1,2,3,4,5,7 , касающиеся уровня жизни населения в целом, и в частности, его доходов и факторов, на них влияющих, имеют не только научную, но и особую практическую значимость. Актуальность данной темы можно выразить одним словосочетанием – "рыночная система" 6 . И действительно, сейчас, когда человечество находится в XXI веке, нельзя недооценивать значение доходов населения. Конечно же, в более раннее время люди тоже думали о своём благосостоянии, но сейчас этот вопрос стоит как никогда остро. Дело в том, что в последнее десятилетие наша страна вступила в последнюю, заключительную фазу рыночных преобразований. Всё это оказывается на уровне жизни населения, на их доходах и расходах.

Существующее в данный момент неравенство получаемых доходов и заработной платы типично для всех современных государств, вне зависимости от уровня благосостояния их населения. Одним из самых значимых последствий экономических преобразований в нашей стране явилось изменение модели распределения доходов между разными категориями населения.

Доходы населения могут меняться под воздействием довольно большого количества факторов: социально-политических, демографических, профессиональных, статусных, экономических, географических и других. Так, доход любого человека зависит, в первую очередь, от социальных факторов. Власть и имеет возможность определять направление социальных программ, порядок получения доходов в рамках определенной экономической стратегии.

В настоящей работе рассмотрим пример построения эконометрической модели социальных процессов и ее анализ. Будем исследовать зависимость одного из основных социальных показателей – "среднедушевые доходы населения" в зависимости от ряда факторов, ха-

рактеризующих уровень социально-экономического развития в регионе.

Основные источники доходов населения могут быть представлены:

- ◆ факторными доходами (заработной платой);
- ◆ доходами от собственности (арендной платой, процентами, дивидендами);
- ◆ доходами от предпринимательской деятельности (прибылью);
- ◆ трансфертными платежами (получение пенсий, пособий, стипендий и т. д.);
- ◆ другими поступлениями (страховыми возмещениями, доходами от продажи иностранной валюты и пр.).

Исследование проводилось на базе статистических данных Росстата 8 . Для анализа были доступны наблюдения официальной статистики по 82 российским регионам за 2015 г. В основу формирования рабочей выборки для целей анализа авторами приняты следующие соображения. Прежде всего, выборка среднедушевых доходов населения должна, с одной стороны, быть достаточно представительной, с другой – не содержать регионы с малыми доходами. Кроме того, представляется нецелесообразным включение в состав выборки рассматриваемый показатель по крупным регионам, таким как Москва, Московская область, Санкт-Петербург и другие, так как их присутствие в выборке, по нашему убеждению, исказит основные результаты анализа. На наш взгляд, следует рассмотреть 55 российских регионов со среднедушевыми доходами населения в диапазоне начиная от 16,1 тыс. рублей в месяц и выше.

В качестве результирующего показателя рассмотрим "Y" – Среднедушевые денежные доходы (в месяц), руб.

Для анализа зависимости между среднедушевыми денежными доходами и основными макроэкономическими показателями в российских регионах были отобраны объясняющие переменные:

- X₁ – Численность населения, тыс. человек;
- X₂ – Среднегодовая численность занятых в экономике, тыс. человек;
- X₃ – Потребительские расходы в среднем на душу населения (в месяц), руб.;
- X₄ – Валовой региональный продукт, млн. руб.;
- X₅ – Основные фонды в экономике, млн. руб.;
- X₆ – Продукция сельского хозяйства, млн. руб.;
- X₇ – Ввод в действие общей площади жилых домов, тыс. м²;
- X₈ – Оборот розничной торговли, млн. руб.;
- X₉ – Сальдированный финансовый результат деятельности организаций, млн. руб.;
- X₁₀ – Инвестиции в основной капитал, млн. руб.

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Вычислим матрицу парных корреляций с помощью инструментария пакета Excel "Анализ данных – Корреляция". Значимость коэффициентов корреляции определяем по критерию Стьюдента. Для уровня значимости $\alpha=0,05$ критическое значение коэффициентов корреляции по модулю составляет 0,266.

Из табл. 1 видно, что доходы населения имеет значимую при уровне значимости $\alpha=0,05$ линейную зависимость с X_3 , X_5 и X_{10} . С факторами X_1 , X_2 , X_4 , X_6 , X_7 , X_8 , X_9 связь не значима.

Проведем первичный анализ исходных данных на присутствие в экзогенных переменных модели мультиколлинеарности. Для этого в табл. 2 найдем близкие по модулю к единице коэффициенты корреляции между объясняющими переменными

$$(|r_{x_i x_j}| > 0,7).$$

Таблица 1.

Корреляционная таблица.

	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Y	1,000	-0,242	0,185	0,924	0,256	0,372	-0,255	-0,248	-0,049	-0,069	0,314
X_1	-0,242	1,000	0,974	-0,093	0,728	0,566	0,653	0,805	0,953	0,210	0,669
X_2	0,185	0,974	1,000	-0,034	0,790	0,618	0,637	0,762	0,930	0,240	0,677
X_3	0,924	-0,093	-0,034	1,000	0,346	0,433	-0,102	-0,103	0,150	-0,031	0,397
X_4	0,256	0,728	0,790	0,346	1,000	0,887	-0,102	0,633	0,760	0,372	0,887
X_5	0,372	0,566	0,618	0,433	0,887	1,000	0,242	0,429	0,429	0,233	0,866
X_6	-0,255	0,653	0,637	-0,102	0,455	0,242	1,000	0,429	0,646	0,492	0,378
X_7	-0,248	0,805	0,762	-0,103	0,633	0,429	0,671	1,000	0,793	0,241	0,556
X_8	-0,049	0,953	0,930	0,150	0,760	0,610	0,646	0,793	1,000	0,187	0,722
X_9	-0,069	0,210	0,240	-0,031	0,372	0,233	0,492	0,241	0,187	1,000	0,204
X_{10}	0,314	0,669	0,677	0,397	0,887	0,866	0,378	0,556	0,722	0,204	1,000

Таблица 2.

Корреляционная таблица R.

	X_1	X_3	X_5	X_6	X_{10}
X_1	1,000	-0,093	0,566	0,653	0,210
X_3	-0,093	1,000	0,433	-0,102	-0,031
X_5	0,566	0,433	1,000	0,242	0,233
X_6	0,653	-0,102	0,242	1,000	0,492
X_{10}	0,210	-0,031	0,233	0,492	1,000

Если такие присутствуют, то из двух переменных, которым соответствует данный коэффициент корреляции, оставим ту, которая имеет наибольшую связь с результативным показателем "Y":

Исходя из этого, получим набор независимых переменных $\{X_1, X_3, X_5, X_6, X_9\}$ в линейной модели множественной регрессии будет отсутствовать мультиколлинеарность, что даст более достоверные результаты, и при этом сохранится максимальная зависимость результирующего фактора от объясняемых переменных.

Проверка набора данных на мультиколлинеарность

Проверка мультиколлинеарности факторов может быть проведена методом испытания гипотезы о независимости переменных H_0 :

$$\text{Det} |R| = 1.$$

Определитель $\text{Det} R = 0,156$. Считаем статистику:

$$\begin{aligned} \chi^2_{\text{факт}} &= \left[n - 1 - \frac{1}{6} \cdot (2 \cdot m + 5) \lg \text{Det} R \right] = \\ &= \left[55 - 1 - \frac{1}{6} \cdot (2 \cdot 5 + 5) \lg 0,156 \right] = 56,02 \\ \chi^2_{\text{мабл}} &\left(0,95; \frac{1}{2} \cdot n \cdot (n-1) \right) = \\ &= \chi^2_{\text{мабл}} \left(0,95; \frac{1}{2} \cdot 55 \cdot (55-1) \right) = 1396,51 \\ \chi^2_{\text{факт}} &< \chi^2_{\text{мабл}} \end{aligned}$$

Гипотеза H_0 принимается. Это означает, что $\text{Det}|R|$ близок к единице и недиагональные ненулевые коэффициенты корреляции не оказывают влияние на коллинеарность факторов. То есть мультиколлинеарность можно считать недоказанной.

Линейное регрессионное уравнение

Для построения линейного регрессионного уравнения используем инструментарий Анализ данных – Регрессия.

Регрессионный анализ по пяти факторам.

Регрессионная статистика	
Множественный R	0,944
R-квадрат	0,891
Нормированный R-квадрат	0,880
Стандартная ошибка	1831,843
Наблюдения	55,000

Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	5,000	1349889625,322	269977925,064	80,455*	0,000
Остаток	49,000	164426844,183	3355649,881		
Итого	54,000	1514316469,505			

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
$Y_{\text{перес.}}$	3282,720	1512,081	2,171*	0,035	244,079	6321,360
X_1	-1,355	0,646	-2,099*	0,041	-2,652	-0,058
X_3	1,278	0,091	14,032*	0,000	1,095	1,461
X_5	0,001	0,001	1,685	0,098	0,000	0,002
X_6	-0,012	0,010	-1,117	0,269	-0,032	0,009
X_9	0,001	0,007	0,125	0,901	-0,013	0,015

Значимо при $\alpha=0,05$

Уравнение множественной регрессии:

$$\tilde{Y} = 3282,720 - 1,355 X_1 + 1,278 X_3 + 0,001 X_5 - 0,012 X_6 + 0,001 X_9$$

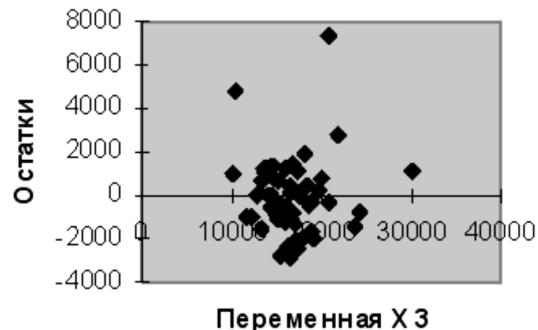
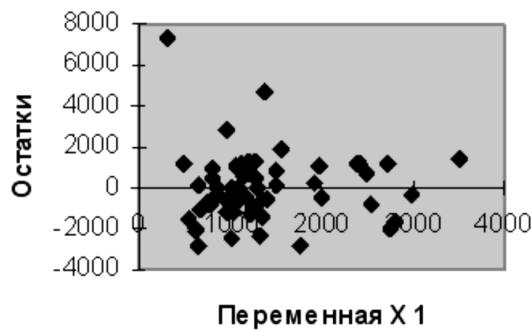
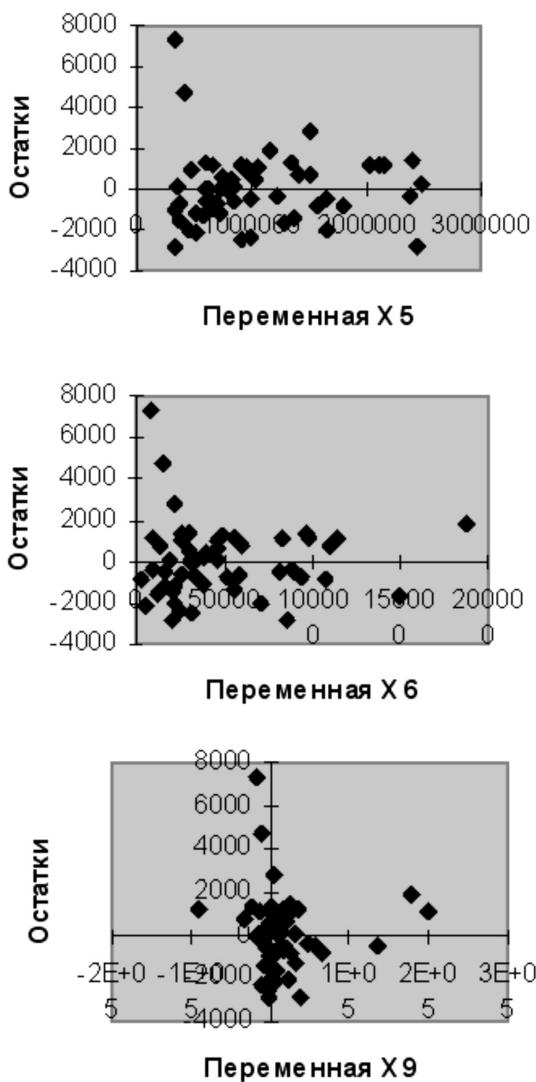


Рисунок 1/1. Графики подбора по переменным X_1, X_3

Рисунок 1/2. Графики подбора по переменным X_5, X_6, X_9

Интерпретация коэффициентов регрессии

- ◆ Коэффициент при переменной X_1 , показывает, что при увеличении численности населения на 1 тыс. чел. доходы в среднем снижаются на 1,355 руб.
- ◆ Коэффициент при переменной X_3 , показывает, что при увеличении потребительских расходов в расчете на душу населения на 1 руб. среднедушевые доходы в среднем увеличиваются на 1,278 руб.
- ◆ Коэффициент при переменной X_5 , показывает, что при увеличении основных фондов на 1 млн. руб. среднедушевые доходы в среднем увеличиваются на 0,001 руб.
- ◆ Коэффициент при переменной X_6 , показывает, что при увеличении продукции сельского хозяйства на 1 млн. руб. среднедушевые доходы в среднем снижаются на 0,012 руб.

◆ Коэффициент при переменной X_3 , показывает, что при увеличении сальдированного финансового результата деятельности организаций на 1 млн. руб. среднедушевые доходы в среднем увеличиваются на 0,001 руб.

Из приведенных данных видно, что при $\alpha=0,05$ только коэффициенты регрессии по объясняющим переменным X_1 и X_3 статистически значимы. Остальные коэффициенты при $\alpha=0,05$ не значимы.

Коэффициент детерминации равен:

$$R^2 = 1 - \frac{\sum(Y_t - \hat{Y}_t)^2}{\sum(Y_t - \bar{Y})^2} = 1 - \frac{164426844,183}{1514316469,505} = 0,891$$

Следовательно, в 89,1% случаев изменение доходов связано с учтенными модели факторами.

Индекс корреляции данного регрессионного уравнения составляет

$$R = \sqrt{R^2} = \sqrt{0,891} = 0,944,$$

следовательно, связь между рассматриваемыми показателями тесная.

Оценим надежность производственной функции по критерию Фишера:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m} = \frac{0,891}{1-0,891} \cdot \frac{55-5-1}{5} = 80,455$$

Поскольку

$$F > F_{KP}(5; 55; 0,05) = 2,40,$$

то можно сказать, что построенное регрессионное уравнение адекватно и статистически надежно.

Скорректированный коэффициент детерминации равен:

$$\hat{R}^2 = 1 - (1 - 0,891) \frac{55-1}{55-5-1} = 0,880.$$

Высокое значение скорректированного коэффициента детерминации говорит о хорошем качестве модели.

Проверка на однородность данных

Тест Чоу

Отсортируем исходные данные по возрастанию Y переменной и, используя метод наименьших квадратов, построим модель по данным выборки, содержащей 55 наблюдений, и найдем для нее ESS_R .

$$\tilde{Y} = 1966,189 - 1,162X_1 + 1,380X_3.$$

$$ESS_R = 185\,216\,776,18.$$

Далее принятую в расчет выборку наблюдений разбиваем на две подвыборки и строим для каждой из них линейные регрессии, определяя при этом величины суммы квадратов отклонений значений \hat{Y}_t от регрессионных значений

$$\hat{Y}_t, ESS_1, \hat{Y},$$

ESS_1 и, соответственно ESS_2 .

Для первой выборки:

$$\tilde{Y} = 10543,181 + 0,045X_1 + 0,643X_3,$$

$$ESS_1 = 30752641,840.$$

Для второй выборки:

$$\tilde{Y} = 756,207 - 1,446X_1 + 1,478X_3.$$

$$ESS_2 = 88800322,542.$$

$$ESS_{UR} = ESS_1 + ESS_2 = 119\,552\,964,38.$$

Вычислим F – статистику:

$$F = \frac{(ESS_R - ESS_{UR})/(k+1)}{ESS_{UR}/(m+n-2k-2)}$$

$$F = \frac{(185\,216\,776,18 - 119\,552\,964,38)/3}{(119\,552\,964,38) / (60 - 2*2 - 2)} = 8,06$$

Найдем критическую точку распределения Фишера исходя из выбранного уровня значимости α .

$$F_c(3, 50-2*2-2) = 2,82.$$

Поскольку $F > F_c$, то отвергаем гипотезу об однородности данных. Следовательно следует рассматривать две выборки раздельно.

$$\tilde{Y} = 10543,181 + 0,045X_1 + 0,643X_3:$$

По первой выборке модель имеет вид:

Таким образом, для регионов с небольшими доходами населения при увеличении численности населения на 1 тыс. чел. среднедушевые доходы в среднем увеличиваются на 0,045 руб., а переменной X_3 на 1 руб. – среднедушевые доходы в среднем увеличиваются на 0,643 руб.

По второй выборке модель имеет вид:

$$\tilde{Y} = 756,207 - 1,446X_1 + 1,478X_3:$$

Получается, что для регионов с большими доходами населения при увеличении численности населения на 1 тыс. чел. среднедушевые доходы в среднем снижаются на 1,446 руб., а при увеличении переменной X_3 1 руб. среднедушевые доходы в среднем увеличиваются на 1,478 руб.

Таким образом, чем больше доходы населения, тем сильнее снижаются доходы при росте населения и тем быстрее растут доходы при росте потребительских расходов.

Проверка на гомоскедастичность

Проведем проверку по тесту Гольдфельда – Куандта. Для этого упорядочим все переменные Y по возрастанию факторов x_1 и x_3 .

По переменной x_1 :

В результате получены две совокупности с малыми и большими значениями x_1 . По каждым совокупностям выполним расчет регрессии и рассчитаем сумму квадратов отклонений. Найдем отношение полученных остаточных сумм квадратов (в числителе должна быть большая сумма):

$$F = 1,645.$$

Вывод о наличии гомоскедастичности остатков можно сделать с помощью F -критерия Фишера при уровне значимости $\alpha = 0,05$ с двумя степенями свободы

$$k_1 = k_2 = \frac{n - C - 2 \cdot p}{2} = \frac{55 - 13 - 2 \cdot 3}{2} = 18,$$

где p – число параметров уравнения регрессии:

$$F_{табл}(0,05; 18; 18) = 2,217.$$

Поскольку $F_{табл} > F$ можно отвергнуть наличие гомоскедастичности в остатках регрессии. Исходя из этого в остатках отсутствует гетероскедастичность по переменной x_1 .

По переменной x_3 :

Найдем отношение полученных остаточных сумм квадратов (в числителе должна быть большая сумма):

$$F = 2,646.$$

Аналогично как и по переменной X_1 в условиях когда $F_{табл} < F$ гомоскедастичность в остатках регрессии отвергается. Исходя из этого в остатках отсутствует гетероскедастичность по переменной x_3 .

Проверка по тесту Бреуша – Пагана

Находим $e_t^2 / \hat{\sigma}^2$,

где $\hat{\sigma}^2 = RSS/n$, RSS – сумма квадратов остатков.

Оцениваем вспомогательную линейную регрессию квадратов стандартизованных остатков на константу и фактор x_1 и x_3 , от которых может зависеть дисперсия ошибок:

$$\epsilon_t^2 / \sigma^2 = \gamma_0 + \gamma_1 x_3$$

Статистика теста рассчитывается как

$$ESS/2 = 5,610/2 = 2,805,$$

где ESS – объяснённая сумма квадратов вспомогательной регрессии.

Данная статистика асимптотически имеет распределение $\chi^2(p)$

где p – количество переменных при $t=1$.

$$\chi^2(p) = 3,841$$

Поскольку $2,805 < 3,841$, то остатки гомоскедастичны.

Проверка по тесту Уайта

Построим регрессию:

$$\epsilon^2 = \beta_0 + \beta_1 X_1 + \beta_2 X_3 + \beta_3 X_1^2 + \beta_4 X_3^2 + \beta_5 X_1 X_3 .$$

LM-статистика для данной регрессии составила:

$$n R^2 = 55 * 0,168 = 9,25.$$

По таблице распределения $\chi^2_{kp} (\alpha; K)$, где α – уровень значимости.

В нашем случае $\alpha = 0,05$; $K = [k^2 + 3k] / 2$,

где k – это число регрессоров в исходной модели; $k = 2$, следовательно, $K = 7$.

$$\chi^2_{kp} = 14,06.$$

Поскольку

$$nR^2 = 9,25 > \chi^2_{kp} (\alpha; K),$$

то есть основания отклонять гипотезу H_0 об отсутствии гетероскедастичности.

ЗАКЛЮЧЕНИЕ

В работе получена эконометрическая модель зависимости социального показателя – доходы населения от различных социально-экономических показателей, характеризующих уровень развития регионов. Осуществлена проверка выполнимости предпосылок гетероскедастичности и мультиколлинеарности, являющихся основными условиями при построении и анализе множественной регрессионной модели.

При проверке исходного набора наблюдения на мультиколлинеарность установлена сильная линейная зависимость между факторными переменными, которая была устранена в ходе анализа корреляционной матрицы. В результате можно отметить, что доходы населения тесно в наибольшей степени зависят от численности населения в регионе и от уровня потребительских расходов, другие факторы, включенные в модель менее значимы.

Представленная в работе методика эконометрического исследования социальных процессов является эффективной и ее можно рекомендовать для использования в аналогичных исследованиях.

ЛИТЕРАТУРА

1. Колмаков, И.Б. Инструментальные средства реализации прогноза поляризации денежных доходов населения / И.Б Колмаков // Современная экономика: концепции и модели инновационного развития. М., 2016, Российский экономический университет им. Г.В. Плеханова. С. 94–101.
2. Овчарова, Л.Н., Попова, Д.О. Доходы и расходы российских домашних хозяйств: что изменилось в массовом стандарте потребления / Л.Н. Овчарова // Мир России: Социология, этнография. – 2013. Т. 22. – № 3. – С. 3–34.
3. Кабашова, Е.В. Статистическое моделирование в исследовании благосостояния населения / Е.В. Кабашова // Современные технологии управления. Сборник трудов. – М.: 2014, МЦНИП. – С. 356–364.
4. Китова О.В., Колмаков И.Б., Доможаков М.В., Кривошеева Я.В., Пеньков И.А. Гибридные распределенные регрессионные интеллектуальные системы прогноза показателей социально-экономического развития России / О.В. Китова // Вестник Российской экономического университета им. Г.В. Плеханова. – 2017. – № 2 (92). – С. 147–161.
5. Великанова, Т.Б. Средний, медианный и модальный уровень денежных доходов населения в целом по России и по субъектам Российской Федерации / Т.Б. Великанова // – Режим доступа: http://www.gks.ru/free_doc/new_site/population/bednost/tabl/tab-bed1-2-6.htm (дата обращения 29.10.2017).
6. Блищенко, Д.К., Бондаренко, О.В. Дифференциация доходов населения и ее влияние на экономику страны / Д.К. Блищенко // [Электронный ресурс]. – Режим доступа: <http://www.scienceforum.ru/2014/pdf/3264.pdf> (дата обращения 25.10.2017).
7. Забайкин, Ю.В. Применение экономико-математических методов в производственном планировании / Ю.В. Забайкин // Кант. – 2017. – № 2 (23). – С. 140–147.
8. Регионы России. Социально-экономические показатели. – 2015: Стат. сб. / Росстат. – М.: – 2015. – 1226 с.