

УДК 519.86

DOI 10.17513/snt.40058

ОЦЕНКА КАЧЕСТВА МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ В СРЕДЕ SPSS STATISTICS

Гусарова О.М., Денисов Д.Э.

*ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации»,
Смоленск, e-mail: omgusarova@mail.ru*

Целями исследования являются обобщение и систематизация методологических подходов и методики оценки качества математических моделей для решения широкого круга прикладных задач с использованием современных цифровых технологий. В ходе исследования использовались методы математической статистики и теории вероятностей, метод статистических наблюдений и выборочных совокупностей, методы стохастического моделирования, базирующиеся на аппарате корреляционно-регрессионного анализа. Статистические расчеты и анализ полученных результатов выполнены в среде SPSS Statistics. В качестве информационной базы для построения математических моделей использовались официальные статистические данные социально-экономического развития региональных субъектов в интервале 2010–2022 гг. В ходе исследования рассмотрены общесистемные подходы к построению многофакторных регрессионных моделей, выполнен обзор основных принципов построения системы факторов-регрессоров. Систематизированы критерии оценки качества регрессионных моделей. Приведены математические зависимости оценки критериев исследования качества проектируемых моделей. Обозначены рекомендуемые интервалы для определения статистической значимости регрессионных уравнений и факторов-регрессоров. Приведены фрагменты программирования для реализации функций корреляционно-регрессионного анализа в среде SPSS Statistics. Выполнены расчеты и обоснование результатов корреляционной зависимости и регрессионной статистики при построении математических моделей для показателей уровня развития регионального субъекта. Практическая значимость исследования заключается в возможности применения результатов исследования при разработке стратегий развития региональных субъектов на краткосрочную и долгосрочную перспективу с использованием математического моделирования и современных цифровых технологий.

Ключевые слова: математические модели, многофакторная регрессия, критерии оценки качества регрессионных уравнений

EVALUATION OF THE QUALITY OF MATHEMATICAL MODELS IN THE SPSS STATISTICS ENVIRONMENT

Gusarova O.M., Denisov D.E.

*Financial University under the Government of the Russian Federation, Smolensk,
e-mail: om.gusarova@mail.ru*

This scientific article is dedicated to the generalization and systematization of some methodological approaches and methods of assessing the quality of mathematical models for solving a wide range of applied aims by using some modern digital technologies. While analysis, the method of statistical observations and sample populations, special methods of stochastic modeling based on the apparatus of correlation-regression analysis were put into practice. Statistical calculations and analysis of the results obtained are performed in the SPSS Statistics environment. During the study, the system-wide approaches to the construction of multivariate regression models were put into practice, and the basic principles of constructing a system of several factors-regressors were studied accordingly. The criteria for estimating the quality of regression models are systematized. Mathematical dependencies on the evaluation of the criteria for studying the quality of the designed models are given accordingly. Recommended intervals for determining the statistical significance of regression equations and regressor factors are indicated. Fragments of programming for the implementation of correlation-regression analysis functions in the SPSS Statistics environment are shown. Calculations and substantiation of the results of correlation-regression analysis in the construction of mathematical models for indicators of socio-economic development of a regional subject are put into practice. The practical value of the article under discussion is in the possibility of using the results of the research of the development of strategies in the development of regional centers in the short and long term by means of using mathematical modeling and modern digital technologies.

Keywords: mathematical models, multivariate regression, criteria for assessing the quality of regression equations

В стратегии развития Российской Федерации до 2030 года и на перспективу до 2036 года основными целями национального развития отечественной экономики являются «укрепление экономического суверенитета, увеличение численности населения страны и повышение уровня жизни граждан» [1, с. 1]. Авторы настоящей публикации в более ранних научных трудах

исследовали динамику ряда показателей уровня развития ряда региональных субъектов [2, 3]. В ходе проведенных исследований с целью выявления тенденций развития хозяйствующих субъектов были построены математические модели, характеризующие особенности развития региональных субъектов. При этом систематически возникал вопрос об уровне надежности и достовер-

ности построенных математических моделей и степени доверия к полученным результатам исследования.

Целями исследования являются обобщение и систематизация методологических подходов и методики оценки качества математических моделей для решения широкого круга прикладных задач с использованием современных цифровых технологий.

Материалы и методы исследования

При проведении исследования в качестве материалов использовались статистические данные по показателям социально-экономического развития региональных субъектов в интервале 2010–2022 гг.

Для получения достоверных результатов применялись методы математической статистики и теории вероятностей, метод статистических наблюдений и выборочных совокупностей, методы стохастического моделирования с использованием инструментария корреляционно-регрессионного анализа. Статистические расчеты и анализ полученных результатов осуществлялись в среде SPSS Statistics.

Результаты исследования и их обсуждение

Для построения математических моделей показателей социально-экономического развития региональных субъектов представляется целесообразным использование инструментария корреляционно-регрессионного анализа, базирующегося на методах математической статистики и теории вероятностей.

Для оценки уровня социально-экономического развития регионального субъекта может быть использована многофакторная регрессионная модель, которая может быть записана в общем виде:

$$y(t) = f(x_1, x_2, x_3 \dots x_m) + \varepsilon(t), \quad (1)$$

где $y(t)$ – значение результативного признака, характеризующего уровень социально-экономического развития регионально-го субъекта;

f – функция зависимости результативного признака и экзогенных факторов-регрессоров;

$x_1, x_2, x_3 \dots x_m$ – набор экзогенных факторов-регрессоров;

$\varepsilon(t)$ – стохастическая компонента, характеризующая различного рода технические погрешности и влияние случайных факторов.

Определение параметров уравнения множественной линейной регрессии может быть осуществлено с использованием метода наименьших квадратов. Важным

моментом при осуществлении спецификации модели является построение системы показателей для выбора факторов-регрессоров. При этом необходимо провести оценку корреляционных зависимостей между эндогенной переменной и экзогенными факторами-признаками.

Корреляционная матрица имеет вид:

$$R = \begin{pmatrix} r_{yy} & r_{yx_1} & \dots & r_{yx_m} \\ r_{yx_1} & r_{x_1x_1} & \dots & r_{x_1x_m} \\ r_{yx_2} & r_{x_2x_1} & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ r_{yx_m} & r_{x_mx_1} & \dots & r_{x_mx_m} \end{pmatrix}, \quad (2)$$

где r_{yxi} – коэффициент корреляции, характеризующий влияние соответствующего фактора-регрессора на зависимую переменную;

r_{xixj} – коэффициент парной корреляции между экзогенными переменными, рассматриваемыми в качестве факторов-регрессоров, оценивающий величину их взаимного влияния [4].

Для определения величины коэффициента парной корреляции между результативным признаком и фактором-регрессором может быть также использована формула:

$$r_{yx} = \sqrt{\frac{\sum (\tilde{y}_x - \bar{y})^2}{\sum (y - \bar{y})^2}}, \quad (3)$$

где $\sum (\tilde{y}_x - \bar{y})^2$ – сумма квадратов отклонений моделируемых значений результативного признака, обусловленная влиянием фактора-регрессора, от среднего значения y , рассчитанного по эмпирическим данным;

$\sum (y - \bar{y})^2$ – общая сумма квадратов отклонений эмпирических значений признака у относительно его среднего значения.

Метод наименьших квадратов позволяет определить параметры модели регрессии. В общем случае нахождение параметров множественной регрессии сводится к нахождению минимума функции G , характеризующей сумму отклонений эмпирических данных от моделируемых значений результативного признака с использованием регрессионной модели:

$$G = \sum_{i=1}^n f(y_i - \hat{y}_i) = \sum_{i=1}^n f(e_i), \quad (4)$$

где f – функция, характеризующая математическую форму измерения разброса эмпирических данных относительно моделируемых значений. В большинстве случаев в качестве математической формы используется ква-

длат разброса эмпирических и моделируемых значений результативного признака.

После осуществления этапа параметризации регрессионной модели необходимо осуществить исследование качества и достоверности разработанной регрессионной модели. С этой целью необходимо провести анализ ряда следующих характеристик.

Для оценки общего качества уравнения регрессии традиционно используется коэффициент детерминации, значение которого может быть определено по формуле:

$$R^2 = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_y^2} = 1 - \frac{ESS / n}{TSS / n} = 1 - \frac{ESS}{TSS}, \quad (5)$$

где

$$ESS = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

– величина дисперсии остатков, характеризующих отклонения фактических и расчетных (моделируемых) значений эндогенной переменной;

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 = n\hat{\sigma}_y^2 \quad (7)$$

– величина общей дисперсии, характеризующей величину отклонений фактических значений признака-результата относительно среднего значения результативного признака.

Значение коэффициента детерминации может изменяться в пределах $[0-1]$, и, чем ближе его значение к 1, то есть чем меньше соотношение остаточной и общей дисперсии, тем более высоким признается общее качество построенного уравнения регрессии.

Для оценки статистической значимости уравнения регрессии и достоверности полученных результатов рекомендуется использовать величину F-критерия Фишера, значение которого может быть рассчитано как отношение факторной и остаточной дисперсии результативного признака:

$$F = QSS / ESS \quad (8)$$

где

$$QSS = \sum (\hat{y}_x - \bar{y})^2 \quad (9)$$

– факторная дисперсия.

Для оценки величины F-критерия Фишера также можно применить следующую формулу:

$$F = \frac{R^2}{1 - R^2} \times (n - m - 1), \quad (10)$$

где n – длина временного ряда результативного признака, отражающая количество исследуемых наблюдений,

m – количество факторных признаков, включенных в уравнение регрессии.

Полученное расчетным путем значение F-критерия Фишера соотносят с его критическим значением, при этом необходимо учесть заданный уровень значимости α (0,05 или 0,01) и число свободы $v_1 = m - 1$, $v_2 = n - m$. Если $F_{факт} > F_{табл}$, то полученное уравнение регрессии признается статистически значимым, в противном случае уравнение регрессии признается статистически не значимым, и необходимо пересмотреть набор факторов-регрессоров [5].

Третьим направлением в оценке качества регрессионного уравнения является оценка статистической значимости факторов-регрессоров, которая может быть выполнена с использованием t-критерия Стьюдента по следующей формуле:

$$t_{b_i} = \frac{b_i}{\mu_{b_i}}, \quad (11)$$

где b_i – коэффициенты регрессии при факторных признаках;

μ_{b_i} – величина стандартной ошибки коэффициента регрессии при соответствующем факторе-регрессоре.

Расчетное значение t-критерия Стьюдента сравнивается с критическим значением критерия, определенным с учетом степеней свободы $n - m$ и заданной вероятности расчетов α (0,01; 0,05). Если расчетное значение t-критерия превышает критическое, то делается вывод о статистической значимости соответствующего параметра регрессии и статистической значимости фактора-регрессора и правомерности его включения в регрессионное уравнение [6].

Данный математический аппарат реализован в среде SPSS Statistics, которая была использована при анализе уровня социально-экономического развития Смоленской области в динамике за 2010–2022 годы. В качестве результативного признака $Y(t)$ по усмотрению исследователей был выбран показатель «валовой региональный продукт на душу населения», характеризующий уровень развития региона в целом. В качестве факторов-регрессоров по усмотрению исследователей выбраны следующие экзогенные переменные: стоимость основных фондов (X_1 , млн руб.), численность населения региона (X_2 , тысяч), численность занятого населения (X_3 , тысяч), средняя заработная плата по региону (X_4 , тыс. руб.), инвестиции в основные фонды (X_5 , млн руб.).

Для выявления корреляционной зависимости обозначенной системы показателей

была использована функция «Корреляция» SPSS Statistics. Фрагмент реализации данной операции представлен следующим образом:

```
GET DATA
  /TYPE=XLSX
  /FILE='C:\Users\Desktop\!2024\Стат данные.xlsx'
  /SHEET=name 'Лист1'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME Наборданных2 WINDOW=FRONT.
CORRELATIONS
  /VARIABLES=YBPII X1COC X2ЧН X3ЧЗН X4СЗП X5ИОС
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Фрагмент полученной матрицы корреляционных зависимостей выбранной системы показателей представлен в таблице 1.

Таблица 1

Итоги расчетов корреляционных зависимостей системы показателей

	Y	X1	X2	X3	X4	X5
Y	1	,955**	-,993**	-,880**	,995**	,814**
X1	,955**	1	-,961**	-,872**	,961**	,767**
X2	-,993**	-,961**	1	,917**	-,995**	-,829**
X3	-,880**	-,872**	,917**	1	-,903**	-,855**
X4	,995**	,961**	-,995**	-,903**	1	,828**
X5	,814**	,767**	-,829**	-,855**	,828**	1

Источник: составлено авторами по [7].

** – коэффициент корреляции признается значимым.

Таблица 2

Результаты построения многофакторной регрессии в SPSS Statistics

Сводка для модели						
Модель	R	R-квадрат	Скорректированный R-квадрат	F		
1	,996 ^a	,993	,989	267,900		
а. Предикторы: (константа), X5, X1, X3, X2						
Итоги регрессии						
	Коэффициенты	Стандартная ошибка	Бета-коэффициенты	t-статистика	Значимость	
1	(Константа)	3891,711	496,502	-	7,838	,000
	X1	-,001	,032	-,004	-,033	,975
	X2	-4,162	,505	-1,162	-8,248	,000
	X3	,721	,290	,210	2,488	,038
	X5	,001	,001	,033	,533	,608
а. Зависимая переменная: Y, BPII						

Источник: составлено авторами

По результатам расчетов можно утверждать, что все выбранные факторы-регрессоры имеют достаточно сильную корреляци-

онную связь с результативным признаком, что подтверждается значением коэффициентов корреляционной связи в интервале

(0,814–0,995). Следует отметить, что фактор-регрессор X_4 (средняя заработная плата по региону) по усмотрению исследователей, исходя из экономического смысла, из рассмотрения был исключен. С факторами-регрессорами X_2 , X_3 (численность населения региона и численность занятого населения) наблюдается тесная отрицательная связь резульгативного признака Y (ВРП на душу населения), что свидетельствует об ориентированности экономики региона на интенсивный путь развития, связанный с внедрением инноваций и современных технологий. Показатели X_1 (стоимость основных фондов) и X_5 (инвестиции в основные фонды) имеют устойчивую положительную корреляцию с эндогенной переменной Y (величина ВРП на душу населения). Результаты регрессионной статистики по итогам построения многофакторной модели представлены в таблице 2.

На основании полученных результатов можно записать уравнение множественной регрессии:

$$Y = 3891,711 - 0,001 * X_1 - 4,162 * X_2 + 0,721 * X_3 + 0,001 * X_5, \quad (12)$$

которое по F -критерию Фишера, равному 267,9, признается статистически значимым и имеет высокое качество, что подтверждают такие характеристики, как коэффициент детерминации R -квадрат = 0,993, значение которого достаточно близко к 1, и достаточно высокое значение коэффициента $R = 0,996$ множественной корреляции, характеризующее тесноту связи факторов-регрессоров и эндогенной переменной.

В дополнение к данной модели множественной регрессии, исходя из целей и задачи исследования, могут быть построены следующие модели: многофакторная регрессия только со статистически значимыми факторами-регрессорами, модели парной регрессии от каждого из рассмотренных факторов, что позволит осуществить проведение более детального исследования каждой переменной, входящей в систему рассматриваемых показателей. В дополнение к этому может быть осуществлен трендовый анализ факторов-регрессоров, что позволит в автоматическом режиме в среде SPSS Statistics построить прогнозы по исследуемым показателям и таким способом определить прогнозные значения эндогенной переменной, формируемой под воздействием факторов-регрессоров. По усмотрению исследователей может быть определен другой набор экзогенных переменных для получения различных вариантов регрессионных моделей и повышения достоверности результатов. Выбор лучшей по качеству модели должен

осуществляться на основании экономико-статистического анализа рассматриваемой системы показателей, но обязательно с учетом математических методов оценки качества регрессий.

Заключение

Рассмотренные методологические аспекты и методика оценки качества математических моделей актуальны и могут быть использованы для решения широкого круга прикладных задач, в том числе для моделирования и исследования динамики показателей, характеризующих уровень социально-экономического развития региональных субъектов.

Практическая значимость исследования заключается в том, что инвариантные многофакторные регрессионные модели могут быть применены для определения перспективных значений ряда показателей регионов, что актуально при разработке стратегий развития региональных субъектов на краткосрочную и долгосрочную перспективу с использованием математического моделирования и инструментария корреляционно-регрессионного анализа. Использование для построения математических моделей и проведения дальнейшего анализа среды SPSS Statistics, позволяющей автоматизировать большую часть статистических расчетов, отражает общую тенденцию расширения сфер применения цифровых технологий в научных исследованиях и практической деятельности.

Список литературы

1. Указ Президента Российской Федерации от 07.05.2024 № 309 «О национальных целях развития Российской Федерации на период до 2030 г. и на перспективу до 2036 года». URL: <http://publication.pravo.gov.ru/document/0001202405070015?ysclid=1xa7zi8vdo7775326> (дата обращения: 11.04.2024).
2. Гусарова О.М., Денисов Д.Э., Сулеменков А.В. Математическое моделирование и численные методы оценки эффективности малого и среднего бизнеса // Современные наукоемкие технологии. 2023. № 10. С. 32-38. DOI: 10.17513/snt.39788.
3. Гусарова О.М., Денисов Д.Э. Цифровые трансформации как фактор стимулирования развития бизнеса // Фундаментальные исследования. 2022. № 5. С. 40-45. DOI: 10.17513/fr.43251.
4. Орлова И.В. Оценивание параметров регрессионной модели с помощью ковариаций переменных и интерпретация коэффициентов регрессии // Мягкие измерения и вычисления. 2023. Т. 71, № 10-1. С. 5-12.
5. Задаев С.А., Орлова И.В. Опыт применения эконометрического инструментария для прогнозирования показателей национальных целей развития РФ // Фундаментальные исследования. 2022. № 10-1. С. 54-59. DOI: 10.17513/fr.43343.
6. Орлова И.В. К вопросу об оценке качества эконометрических моделей // Фундаментальные исследования. 2022. № 3. С. 92-99. DOI: 10.17513/fr.43220.
7. Росстат. Официальный сайт Федеральной службы государственной статистики. [Электронный ресурс]. URL: <https://rosstat.gov.ru/> (дата обращения 25.04.2024).