

## Исследование транспортных пассажиропотоков методами дисперсионного и регрессионного анализов

С.И. ЖОГАЛЬ<sup>1</sup>, С.П. ЖОГАЛЬ<sup>2</sup>, М.А. МАСЛОВСКАЯ<sup>1</sup>

На основе методов регрессионного и дисперсионного анализов определены факторы, существенно влияющие на динамику транспортных пассажиропотоков, построены регрессионные зависимости для пассажиропотоков.

**Ключевые слова:** транспортные пассажиропотоки, прикладной статистический анализ, регрессионный анализ, дисперсионный анализ.

On the basis of regression analysis and analysis of variance the factors that significantly affect the dynamics of transport passenger traffics, are identified and regression dependences for passenger traffics are developed.

**Keywords:** transport passenger traffics, applied statistical analysis, regression analysis, analysis of variance.

Изучение сложных объектов и процессов, являющихся по своей природе стохастическими, возможно лишь на основе применения методов прикладной статистики. Методы прикладной статистики, дисперсионного и регрессионного анализов являются весьма эффективными при исследовании прикладных задач в различных областях науки, техники и производства. Построение регрессионных моделей по опытным данным, выявление факторов, оказывающих существенное влияние на исследуемый процесс, позволяют строить прогнозы и принимать стратегические решения с использованием компьютерного моделирования и прогнозирования.

Рассмотрим задачу построения регрессионных моделей по имеющимся опытным данным при исследовании транспортных потоков. Разработка прогнозных моделей с помощью методов множественной регрессии является важной задачей нахождения математических зависимостей, отражающих статистическую связь транспортного потока с группой других факторов, от которых он зависит [1]. Для прогноза развития железнодорожных магистралей необходимо было построить регрессионные модели зависимости пассажиропотоков на железнодорожном и воздушном транспорте в районе тяготения магистрали Воронеж–Ростов-на-Дону. По данным исследования пассажиропотоков на железнодорожном и воздушном транспорте была составлена следующая таблица.

Таблица 1 – Распределение пассажиропотоков (тыс. чел.) по зонам дальности поездки, видам транспорта и годам

Дальность поездки (км)	2008 г.		2010 г.		2012 г.	
	ж/д	воздушный	ж/д	воздушный	ж/д	воздушный
0–300	191	14	210	17	273	20
301–500	392	12	408	23	575	17
501–800	301	30	313	25	443	21
801–1200	183	33	179	41	265	30
1201–2000	95	27	108	60	152	19
2001–3000	9	9	10	19	57	5

В качестве целевого фактор-признака  $y$ , зависящего от других факторов, будем рассматривать количество пассажиров, а в качестве объясняющих фактор-признаков:

$x_1$  – «код» зоны дальности поездки;

$x_2$  – «код» вида транспорта (железнодорожный или воздушный);

$x_3$  – «код» года совершения поездки.

Методами дисперсионного анализа установим степень и значимость влияния объясняющих факторов на целевой признак  $y$ . Для этого применим метод однофакторного дисперсионного анализа (однофакторной классификации) [2]–[4].

**Однофакторная классификация.** Пусть получено  $p$  выборок из  $p$  ( $p > 1$ ) генеральных совокупностей и необходимо сопоставить  $p$  значений выборочных средних. В этом случае можно говорить о разделении фактора  $A$  на  $p$  уровней или классов. Действие фактора  $A$  проявляется в  $p$  сравниваемых средних значениях. Будем исходить из следующей линейной модели, описывающей наблюдения:

$$y_{ij} = m + a_i + e_{ij}, \quad j = 1, 2, \dots, n_i, \quad i = 1, 2, \dots, p, \quad (1)$$

где  $n_i$  – объем выборки для  $i$ -го уровня фактора  $A$ ,  $m$  – общее среднее значение,  $a_i$  – эффект  $i$ -го уровня фактора  $A$ ,  $e_{ij}$  – независимые и нормально распределенные ошибки наблюдений с нулевым математическим ожиданием и дисперсией  $\sigma^2$ . Дополнительно к (1) добавляется условие

$$n_1 a_1 + n_2 a_2 + \dots + a_p n_p = 0,$$

которое означает, что  $a_i$  группируются около общего среднего  $m$  в виде положительных и отрицательных эффектов.

Применяя метод наименьших квадратов (МНК), связанный с минимизацией суммы квадратов ошибок:

$$S = \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - m - a_i)^2 \quad (2)$$

получаем следующие оценки параметров:

$$\begin{aligned} \hat{m} = y_{..} &= \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} \\ \hat{a}_i = y_{i.} - y_{..} &= \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} - \hat{m}, \quad i = 1, 2, \dots, p \\ n &= n_1 + n_2 + \dots + n_p \end{aligned} \quad (3)$$

Результаты измерений при однофакторной классификации удобно представлять в виде следующей таблицы:

Таблица 2 – Результаты измерений при однофакторной классификации с повторными наблюдениями

$i$	$j$	Число наблюдений по выборкам				Среднее по выборкам
		1	2	...	$n_i$	
Уровни фактора $A$	1	$y_{11}$	$y_{12}$	...	$y_{1n_1}$	$y_{1.}$
	2	$y_{21}$	$y_{22}$	...	$y_{2n_2}$	$y_{2.}$
	...	...	...	...	...	...
	$p$	$y_{p1}$	$y_{p2}$	...	$y_{pn_p}$	$y_{p.}$
						$y_{..}$

Смысл дисперсионного анализа при однофакторной классификации заключается в проверке гипотезы:

$$H_A : a_1 = a_2 = \dots = a_p = 0, \quad (4)$$

означающей равное воздействие всех  $p$  уровней фактора, т. е. отсутствие существенного влияния фактора  $A$  на выходную переменную  $y$ .

Для проверки данной гипотезы потребуются следующие величины:

$$y_{..} = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij}, \quad y_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

$$S_A = \sum_{i=1}^p n_i (y_{i\cdot} - y_{\cdot\cdot})^2 \quad (5)$$

$$S_G = \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - y_{\cdot\cdot})^2,$$

$$S_R = S_G - S_A.$$

Если справедлива гипотеза  $H_A$ , то отношение

$$F = \frac{s_A^2}{s_R^2}, \quad (6)$$

где

$$s_A^2 = \frac{S_A}{(p-1)} \quad (7)$$

$$s_R^2 = \frac{S_R}{(n-p)}$$

имеет F-распределение с числами степеней свободы  $\phi_1 = p-1, \phi_2 = n-p$ . Здесь  $s_R^2$  – несмещенная оценка дисперсии экспериментальной ошибки. Гипотеза  $H_A$  отвергается при выбранном уровне значимости  $\alpha$ , если  $F > F_{kp}$ , где  $F_{kp}$  критическое значение F-распределения. В противном случае гипотеза принимается, т. е. делается вывод о том, что результаты наблюдений не зависят от того, каковы значения фактора  $A$  и, следовательно, фактор  $A$  практически не влияет на выходную величину  $y$ .

**Применение однофакторной классификации при исследовании степени влияния факторов на формирование пассажиропотоков.** Методом однофакторной классификации установлена степень влияния объясняющих признаков на пассажиропотоки железнодорожного и воздушного транспорта. На основе анализа произведенных расчетов и их объединения получены результирующие дисперсионные таблицы по видам транспорта.

Таблица 3 – Дисперсионная таблица по железнодорожному транспорту

Факторы	SG = 410388	SA/SG	F-отношение	Степени свободы
Дальность	SA = 365754.667	0.891241	19.667	5
				12
Год	SA = 35803	0.087242	0.71685	2
				15

Сравнивая полученные значения F-отношения с критическими значениями таблицы Фишера, несложно установить, что фактор года не значимо влияет на величину пассажиропотока по железнодорожному транспорту, в то время как фактор дальности поездки весьма значим.

Были составлены также аналогичные дисперсионные таблицы для пассажиропотока по воздушному транспорту, а также по суммарному пассажиропотоку по двум видам транспорта. Составленные таблицы позволяют судить о степени значимости влияния соответствующих факторов и их взаимодействий на изменение пассажиропотока. Для установления этого факта достаточно найти соответствующие рассчитанным по дисперсионным таблицам степеням свободы критические значения F-распределения и сравнить их с полученными расчетными F-отношениями.

Расчёты значимости влияния факторов показали, что по воздушному транспорту на пассажиропоток оказывают значимое влияние оба фактора – дальность перелёта и год поездки, а на суммарный пассажиропоток значимо влияют все три фактора – дальность поездки, год поездки и вид транспорта.

**Применение методов регрессионного анализа при исследовании транспортных пассажиропотоков.** Были построены регрессионные модели [1], [2], [4], которые могут быть применены при прогнозировании пассажиропотоков (переменная  $y$ ) в зависимости от вышеперечисленных факторов:

$x_1$  – «кода» зоны дальности поездки;

$x_2$  – «кода» вида транспорта (железнодорожный или воздушный);

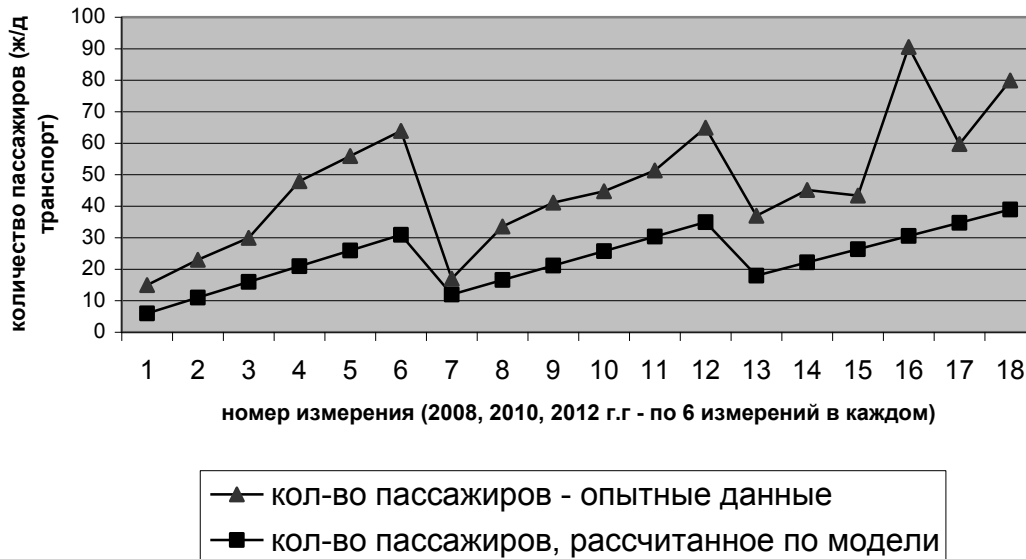
$x_3$  – «кода» года совершения поездки.

**Регрессионная модель по воздушным пассажиропотокам:**

$$y = -5,422 + 5,39x_1 + 6,4x_3 - 0,4x_1x_3, \quad (8)$$

R-Sq (ADJ.) = 0,7849 – степень объяснения поведения фактора  $y$  посредством регрессионной модели (8) составляет 78,49 %.

Эмпирические и рассчитанные по регрессионной модели (8) пассажиропотоки (воздушный транспорт)

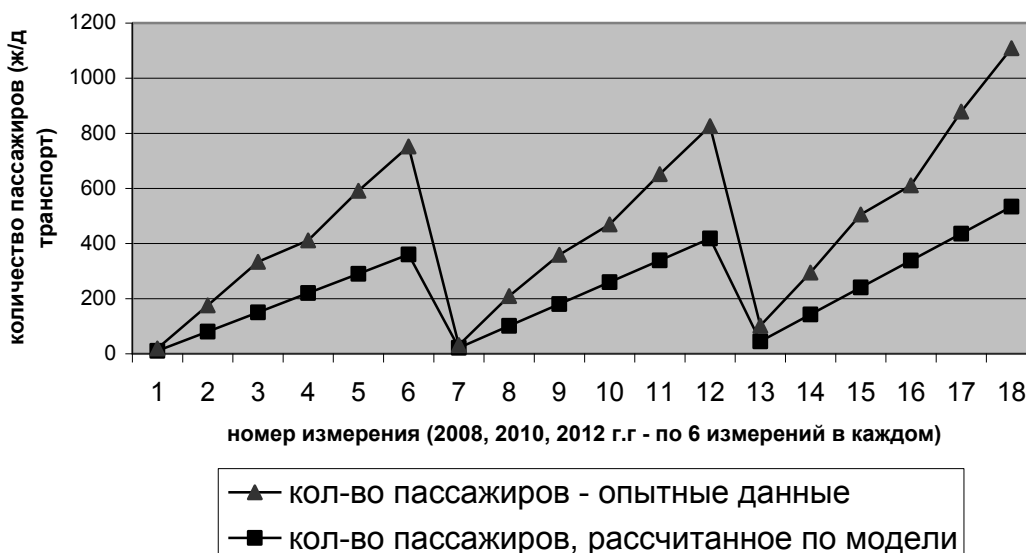


**Регрессионная модель по железнодорожным пассажиропотокам:**

$$y = -61,933 + 60,671x_1 + 2,143x_3 + 9,296x_1x_3, \quad (9)$$

R-Sq (ADJ.) = 0,8695 – степень объяснения поведения фактора  $y$  посредством регрессионной модели (9) составляет 86,95 %.

Эмпирические и рассчитанные по регрессионной модели (9) пассажиропотоки (ж/д транспорт)



**Общая регрессионная модель по воздушному и железнодорожному пассажиропотокам:**

$$y = -14,283 + 29,414x_1 + 79,85x_2 - 0,798x_3 - 40,457x_1x_2 + 4,969x_1x_3 - 18,063x_2x_3 \quad (10)$$

R-Sq (ADJ.) = 0,6834 – степень объяснения поведения фактора  $y$  посредством общей регрессионной модели (10) составляет 68,34 %.

**Заключение.** Построенные регрессионные зависимости могут быть эффективными при принятии стратегических решений по модернизации транспортных маршрутов, целесообразности введения в эксплуатацию на исследуемых направлениях скоростных железнодорожных магистралей.

**Литература**

1. Правдин, Н.В., Дыканюк, М.Л., Негрей, В.Я. Прогнозирование грузовых потоков / Н.В. Правдин, М.Л. Дыканюк, В.Я. Негрей. – М. : Транспорт, 1987. – 247 с.
2. Жогаль, С.П., Жогаль, С.И., Максимей, И.В. Основы регрессионного анализа и планирования эксперимента : Учебное пособие / С.П. Жогаль, С.И. Жогаль, И.В. Максимей. – Гомель, 1997. – 92 с.
3. Невзорова, А.Б., Жогаль, С.И., Скойбеда, А.Т. Исследование статистическими методами влияния влажности древесины на её прочностные свойства / А.Б. Невзорова, С.И. Жогаль, А.Т. Скойбеда // Вести Национальной Академии наук Беларуси. Серия физико-технических наук. – 2002. – № 1. – С. 89–92.
4. Тюрин, Ю.Н., Макаров, А.А. Анализ данных на компьютере / Ю.Н. Тюрин, А.А. Макаров. – М. : ИНФРА–М, 2003. – 544 с.

<sup>1</sup>Белорусский государственный университет транспорта

<sup>2</sup>Гомельский государственный университет им. Ф. Скорины

Поступила в редакцию 15.09.2015