УДК 656.11:519.87 (575.2) (04)

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВЛИЯНИЯ СКОРОСТИ И ИНТЕНСИВНОСТИ ДВИЖЕНИЯ НА КОЛИЧЕСТВО ДОРОЖНО-ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ НА АВТОДОРОГЕ БИШКЕК–ОШ

В.И. Компанцев, М.Т. Алсеитов

Приведено описание оперативной блок-схемы (алгоритма) решения задачи влияния зависимой переменной и независимых переменных дорожного движения на количество дорожно-транспортных происшествий на автодороге Бишкек–Ош.

Ключевые слова: дорожно-транспортное происшествие; скорость автомобиля; интенсивность движения.

В данной работе сформулирована задача выявления зависимости между количеством дорожно-транспортных происшествий (ДТП), скоростным режимом и интенсивностью движения на автодороге Бишкек–Ош. Выдвигается гипотеза о следующей причинно-следственной связи: количество ДТП зависит от скорости движения и его интенсивности. Составим алгоритм решения поставленной задачи, который представлен в виде оперативной блок-схемы (рисунок 1).

В блоке 1 вводятся исходные данные. Трасса Бишкек–Ош была разбита на равные участки длиной по 10 км, для которых были получены показатели количества ДТП (по данным ГУБДД), интенсивности и скорости движения. Кроме этого, была сформулирована задача выявления функциональной зависимости между общим количеством ДТП, скоростью и интенсивностью движения легковых автомобилей.

В качестве зависимой (эндогенной) переменной выступило количество ДТП; в качестве независимых переменных (экзогенных) – скорость и интенсивность. Проанализировав структуру ДТП, было выявлено, что в населенных пунктах, через которые проходит трасса Бишкек–Ош, наблюдаются специфичные ДТП, в которых помимо автомобилей, имеются и другие участники дорожного движения – пешеходы, велосипедисты, животные, гужевой транспорт и пр. Иными словами, в населенных пунктах, через которые проходит трасса, наблюдаются ДТП, связанные с близостью поселений к трассе Бишкек–Ош. Была введена дополнительная условная переменная, отражающая этот фактор. Так, для участков движения в населенных пунктах, через которые проходит трасса, данная переменная принимает значение 1, для остальных участков – 0.

В блоке 2 выбирается форма представленной модели линейной регрессии. В качестве формулы связи между переменными была выбрана линейная регрессия, как наилучшая для интерпретации причинно-следственных связей. На практике линия регрессии чаще всего ищется в виде линейной функции (линейная регрессия), наилучшим образом приближающей известную кривую [1]:

$$Y = b0 + b1X1 + b2X2 + \ldots + bNXN \quad (1)$$

к определенному виду.

В блоке 3 ставится задача определения неизвестных коэффициентов С(*1*), *C*(*2*), *C*(*3*) в следующей регрессии:

$$DTP = C(1)^{*}\ Vi + C(2)^{*}\ INTi + C(3)^{*}\ Nas\_p, \quad (2)$$

где $DTPi$ – количество ДТП на i участке; Vi – средняя скорость автомобилей, км/ч на i участке; $INTi$ – средняя интенсивность на i участке, ед/сут.; Nas_p – фиктивная переменная, для i участков – населенных пунктов, через которые проходит трасса, принимает значение 1, для прочих – 0.

Делается это с помощью метода наименьших квадратов, когда минимизируется сумма квадратов отклонений реальных данных от их оценок (имеются в виду оценки с помощью прямой линии, претендующей на то, чтобы представлять искомую регрессионную зависимость).

В блоке 4 происходит проверка независимых переменных, входящих в регрессию, на наличие мультиколлинеарности с помощью корреляции Пирсона. Мультиколлинеарность – тесная корреляционная взаимосвязь между отбираемыми для

Рисунок 1 – Оперативная блок-схема (алгоритм) решения задачи влияния скорости и интенсивности движения на количество ДТП на автодороге Бишкек–Ош

анализа факторами, совместно воздействующими на общий результат, которая затрудняет оценивание регрессионных параметров. Коэффициент корреляции Пирсона характеризует существование линейной зависимости между двумя величинами и рассчитывается по формуле [2]:

$$r_{xy} = \frac{\sum_{i=1}^{m}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{m}\left(x_i-\bar{x}\right)^2\sum_{i=1}^{m}\left(y_i-\bar{y}\right)^2}}, \qquad (3)$$

где $\bar{x}$, $\bar{y}$ – выборочные средние x^m и y^m.

В блоке 5 проводятся расчеты коэффициента корреляции. В результате расчетов получена корреляционная таблица 1. Как видно из данных таблицы 1, все коэффициенты корреляции между независимыми переменными меньше 0,7 (70 %), следовательно, мультиколлинеарность отсутствует, т.е. это означает, что линейная зависимость между независимыми факторами регрессионной модели отсутствует.

В блоке 6 с помощью инструмента – пакета программного обеспечения Eviews получена следующая спецификация модели:

$$DTP = 0{,}03853076817*V + 0{,}001091496694*INT + 7{,}578129244*Nas\_p. \qquad (4)$$

Таблица 1 – Результаты расчета коэффициентов корреляции

	V	INT	Nas_p
V	1		
INT	0,301413	1	
Nas_p	0,278546	0,564891	1

Пакет Eviews обеспечивает особо сложный и тонкий инструментарий обработки данных, позволяет выполнять регрессионный анализ, строить прогнозы в Windows-ориентированной компьютерной среде. С помощью этого программного средства можно очень быстро выявить наличие статистической зависимости в анализируемых данных и затем, используя полученные взаимосвязи, сделать прогноз изучаемых показателей [3].

В блоке 7 проводится анализ полученной модели на непротиворечивость. Коэффициенты при всех переменных положительны. Это означает, что связь между экзогенными и эндогенными переменными положительна, это не противоречит логике. Иными словами, полученные коэффициенты говорят о том, что:

- чем выше скорость автомобиля, тем больше ДТП;
- чем больше интенсивность движения автомобиля, тем больше ДТП;
- на участках дороги, где трасса проходит через населенные пункты, количество ДТП больше чем вне их.

В блоке 8 осуществляется проверка значимости всего уравнения регрессии по критерию Фишера (F-статистика). С помощью критерия Фишера оценивают качество регрессионной модели в целом и по параметрам.

Для этого выполняется сравнение полученного значения F и табличного значения F. F фактический определяется из отношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы [4]:

$$F_{факт} = \frac{\sum(\hat{y}-\bar{y})^2 / m}{\sum(y-\hat{y})^2 / (n-m-1)}, \qquad (5)$$

где n – число наблюдений; m – число параметров при факторе x.

F табличный – это максимальное значение критерия под влиянием случайных факторов при текущих степенях свободы и уровне значимости а. Уровень значимости а – вероятность не принять гипотезу при условии, что она верна. Как правило, а принимается равной 0,05 или 0,01. Если $F_{табл}$ > $F_{факт}$, то признается статистическая незначимость модели, ненадежность уравнения регрессии.

Нулевая гипотеза о том, что коэффициенты регрессии незначимы (или равны 0) принимается в случае, если вероятность F-статистики больше выбранного уровня значимости в 5 %. В нашем случае ее вероятность равна 0, следовательно, нулевая гипотеза отвергается, т.е. полученные коэффициенты значимы.

В блоке 9 для оценки статистической значимости модели по параметрам рассчитываются t-критерии Стьюдента. Оценка значимости модели с помощью этого критерия проводится путем сравнения их значений с величиной случайной ошибки:

$$t_b = \frac{b}{m_b}, \qquad (6)$$

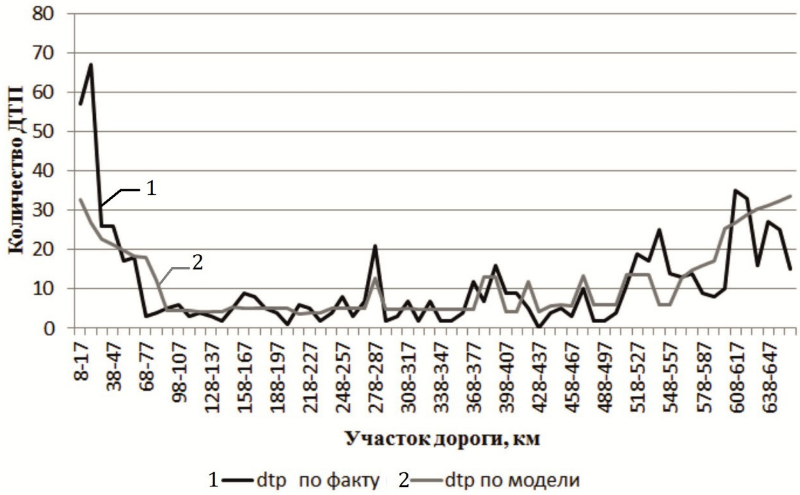

Рисунок 2 – График количества ДТП на участках дороги по факту и по модели

$$t_a = \frac{a}{m_a}, \quad (7)$$

$$t_r = \frac{r}{m_r}. \quad (8)$$

Сравнивая фактическое и табличное значения t-статистики, принимается или отвергается гипотеза о значимости модели по параметрам.

Нулевая гипотеза о том, что коэффициент регрессии незначим (или равен 0) принимается в случае, если вероятность t-статистики больше выбранного уровня значимости в 5 %. В нашем случае вероятность t-статистик для всех переменных меньше 5 %, следовательно, нулевая гипотеза отвергается, полученные коэффициенты значимы.

В блоке 10 проверяем гипотезу о равенстве математического ожидания остатков нулю с помощью t-статистики. Нулевая гипотеза (принимается в случае, если вероятность t-статистики больше выбранного уровня значимости в 5 %) – математическое ожидание остатков равно 0. В нашем случае вероятность t-статистики для ряда остатков больше выбранного уровня значимости в 5 % (0,05) и равна 97,33 % (0.9733), следовательно, математическое ожидание остатков равно 0. Это означает, что разница между реальными значениями ДТП и значениями ДТП по модели в основном аналогичны (рисунок 2).

В блоке 11 получаем значения количества ДТП на участках дороги по модели. Поскольку данный ряд не является временным, необходимость проведения тестов на наличие автокорреляции отпадает.

Таким образом, исходная гипотеза о наличие причинно – следственной связи “количество ДТП зависит от скорости движения и его интенсивности” подтверждается проделанной выше работой, т.е. чем выше скорость, тем больше ДТП и чем выше интенсивность движения, тем также выше количество ДТП.

Литература

1. *Дрейпер Норман, Смит Гарри.* Прикладной регрессионный анализ. Множественная регрессия. 3-е изд. М.: Диалектика, 2007. 912 с.
2. *Айвазян С.А.* Прикладная статистика. Основы эконометрики. Том 2. М.: Юнити-Дана, 2001. 432 с.
3. *Князевский В.С., Молчанов И.Н.* Статистические расчеты на компьютере с использованием ППП Microstat. Ростов н/Д: РГЭА, 1996. 86 с.
4. *Кобзарь А.И.* Прикладная математическая статистика. М.: Физматлит, 2006. 816 с.