

Исследование влияния различных факторов на число пострадавших в дорожно-транспортных происшествиях с помощью методов регрессионного анализа

Горяинов А.В. *

ФГБОУ ВО МАИ (НИУ), Москва, Россия,
e-mail: agoryainov@gmail.com

Поляков Е.А. **

ФГБОУ ВО МАИ (НИУ), Москва, Россия,
e-mail: polyakovevgeny96@mail.ru

В работе рассматриваются зависимости суммарного числа погибших и раненых от различных характеристик ДТП и построение моделей оценки числа погибших и раненых.

Ключевые слова: множественная регрессия, метод наименьших квадратов, дорожно-транспортное происшествие.

Для цитаты:

Горяинов А.В., Поляков Е.А. Исследование влияния различных факторов на число пострадавших в дорожно-транспортных происшествиях с помощью методов регрессионного анализа // Моделирование и анализ данных. 2020. Том 10. № 1. С. 64–74. DOI: 10.17759/mda.2020100103

1. ВВЕДЕНИЕ

Исследование факторов, приводящих к гибели и ранению людей в ДТП является актуальной темой так, как только за первое полугодие 2019 года на территории Москвы зарегистрировано более 4,7 тыс. дорожно-транспортных происшествий, в результате которых погиб 231 человек, более 5,9 тыс. человек получили ранения. Работа посвящена анализу зависимости суммарного числа погибших и раненых в ДТП от различных факторов – характеристик, описывающих, при каких обстоятельствах произошло ДТП.

***Горяинов Александр Владимирович**, кандидат физико-математических наук, доцент, Московский авиационный институт (национальный исследовательский университет) (ФГБОУ ВО МАИ (НИУ)), Москва, Россия, e-mail: agoryainov@gmail.com

****Поляков Евгений Андреевич**, бакалавр прикладной математики, студент магистратуры, Московский авиационный институт (национальный исследовательский университет) (ФГБОУ ВО МАИ (НИУ)), Москва, Россия, e-mail: polyakovevgeny96@mail.ru



Целью исследования является определение характеристик, наиболее сильно влияющих на число пострадавших в ДТП, и построение модели зависимости числа пострадавших от этих характеристик. Предполагается с помощью метода наименьших квадратов найти оценки неизвестных параметров в множественной модели регрессии и на основе построенной модели провести анализ факторов, сильнее всего способствующих гибели и ранению людей при ДТП.

2. СБОР ДАННЫХ

Рассмотрены ДТП, информация о которых была предоставлена центром городских проектов «Штаб». Была получена таблица с данными о 732-х ДТП, произошедших в Москве за август 2019 года. Характеристики ДТП являются как количественными, так и номинальными. В табл. 1 приведено описание всех известных характеристик.

Таблица 1

Характеристики ДТП

Параметр	Описание	Вид характеристики
Time	время	Количественная
DTP_V	вид дтп	Номинальная. Таблица 2.
POG	погибшие	Количественная
RAN	раненые	Количественная
osv	освещение	Номинальная. Таблица 3.
velo	был ли среди участников велосипедист	Бинарная
moto	был ли среди участников мотоциклист	Бинарная
days_of_week	день недели	Номинальная
kids	был ли среди участников ребенок	Бинарная
public_tr	был ли среди транспортных средств общественный транспорт	Бинарная
truck	был ли среди транспортных средств грузовой транспорт	Бинарная
road_quality	выявленные недостатки содержания дорог	Номинальная. Таблица 4
location	место дтп	Номинальная. Таблица 5
weather	погода	Номинальная. Таблица 6
alco	находился ли водитель в состоянии алкогольного опьянения	Бинарная
dr_viol	нарушения водителей	Номинальная. Таблица 7
ped_viol	нарушения пешеходов	Номинальная. Табл. 8
tr_c	количество транспортных средств	Количественная
ptpnt_c	количество участников	Количественная
date	будний/выходной	Бинарная
District	Удалённость от центра	Порядковая. Районы разделены на четыре группы по удаленности от центра Москвы.



Бинарные признаки кодировались числами 1 (присутствует) и 0 (отсутствует). При кодировании числами номинальных характеристик был осуществлён переход от номинальной шкалы к порядковой: градации характеристик были на основе экспертных оценок упорядочены по степени опасности для жизни участников ДТП. Результат кодирования числами номинальных характеристик приведён в табл. 2 – 8.

Таблица 2

Значения характеристики «Вид ДТП»

Вид ДТП	Код
Не определено или иной вид ДТП	-1
Возгорание вследствие технической неисправности движущегося или остановившегося ТС, участвующего в дорожном движении.	1
Падение пассажира	2
Наезд на велосипедиста	3
Наезд на пешехода	4
Опрокидывание	5
Столкновение	6
Наезд на препятствие	7
Наезд на стоящее ТС	8

Таблица 3

Значения характеристики «Освещение»

Освещение	Код
Не определено	-1
Светлое время суток	0
В темное время суток, освещение включено	1
Сумерки	2
В темное время суток, освещение отсутствует	3

Таблица 4

Значения характеристики «Выявленные недостатки содержания дорог»

Выявленные недостатки содержания дорог	Код
Не установлены	0
Дефекты покрытия	1
Отсутствие, плохая различимость горизонтальной разметки проезжей части	2
Неисправность светофора	3
Неисправное освещение	4
Отсутствие пешеходных ограждений в необходимых местах	5
Отсутствие дорожных ограждений в необходимых местах	6
Сужение проезжей части; наличие препятствий, затрудняющих движение транспортных средств	7
Отсутствие дорожных знаков в необходимых местах, отсутствие временных ТСОД в местах проведения работ	8
Неправильное применение, плохая видимость дорожных знаков	9
Отсутствие, плохая различимость вертикальной разметки	10



Таблица 5

Значения характеристики «Место ДТП»

Место ДТП	Код
Не определено или иное место	-1
Автостоянка, отделенная от проезжей части	1
Перегон (нет объектов на месте ДТП)	2
Подход к мосту, эстакаде, путепроводу	3
Мост, эстакада, путепровод, тоннель	4
Регулируемый пешеходный переход	5
Нерегулируемый пешеходный переход	6
Регулируемый перекресток	7
Нерегулируемый перекресток неравнозначных улиц или дорог, автостоянка, не отделенная от проезжей части, выезд с прилегающей территории	8
Внутридворовая территория, пешеходная зона, тротуар, пешеходная дорожка	9
Остановка общественного транспорта	10

Таблица 6

Значения характеристики «Погода»

Погода	Код
Не определено	-1
Ясно	0
Пасмурно	1
Дождь	2
Туман	3
Пасмурно и Дождь	4

Таблица 7

Значения характеристики «Нарушения водителей»

Нарушения водителей	Код
Не определено	-1
Нарушения ПДД велосипедистом	1
Нарушение правил остановки	2
Нарушение правил погрузки, выгрузки и крепления грузов	3
Нарушение правил проезда остановок трамвая	4
Нарушение требований дорожных знаков	5
Несоблюдение условий, разрешающих движение транспорта задним ходом	6
Эксплуатация ТС с техническими неисправностями, при которых запрещается их эксплуатация	7
Нарушения правил перевозки пассажиров	8



Нарушения водителей	Код
Нарушение правил проезда пешеходного перехода	9
Нарушение правил пользования общественным транспортом	10
Несоблюдение очередности проезда перекрестков	11
Неправильный выбор дистанции	12
Несоблюдение бокового интервала	13
Другие нарушения ПДД водителем	14
Нарушение правил перестроения или обгона	15
Нарушение правил расположения ТС на проезжей части	16
Несоответствие скорости конкретным условиям движения, превышение установленной скорости движения	17
Нарушение требований линий разметки	18
Нарушение требований сигналов светофора	19
Выезд на полосу встречного движения, выезд на полосу встречного движения с разворотом, поворотом налево или объездом препятствия	20

Таблица 8

Значения характеристики «Нарушения пешеходов»

Нарушения пешеходов	Код
Не определено	-1
Переход через проезжую часть вне пешеходного перехода в зоне его видимости либо при наличии в непосредственной близости подземного надземного пешеходного перехода	1
Иные нарушения	2
Неожиданный выход из-за сооружений деревьев	3
Нахождение на проезжей части без цели её перехода	4
Переход через проезжую часть в неустановленном месте при наличии в зоне видимости перекрёстка	5
Переход проезжей части в запрещённом месте оборудованном пешеходными ограждениями	6
Неподчинение сигналам регулирования	7
Движение вдоль проезжей части при наличии и удовлетворительном состоянии тротуара	8
Неожиданный выход из-за стоящего ТС	9
Движение вдоль проезжей части попутного направления вне населенного пункта при удовлетворительном состоянии обочины	10

3. ОБЗОР ИСТОЧНИКОВ

Очевидно, что помимо характеристик, представленных в таблице 1, на число погибших и раненых в ДТП влияют также и другие факторы. В их числе в работах [3], [4], [6] посвящённых анализу причин гибели и ранения людей в ДТП, называются



техническое состояние автомобиля, переход дорог пешеходами в состоянии алкогольного опьянения, не использование ремней безопасности, близость медицинских учреждений к месту ДТП и др. Многие из этих характеристик требуют более подробного изучения, и их выяснение – трудоемкий процесс.

4. АНАЛИЗ ДАННЫХ

Расчеты в работе проведены в программной среде SPSS.

Множественная регрессия является расширением простой линейной регрессии. С помощью простой регрессии оценивалась степень влияния одной независимой переменной (предиктора) на зависимую переменную (критерий). В отличие от простой регрессии ($Y = \beta_0 + \beta_1 X_1$), множественная регрессия исследует влияние двух и более предикторов на критерий ($Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$).

Применение множественной регрессии позволяет исследователю ответить на вопрос, насколько хорошо оцененное уравнение аппроксимирует данные, есть ли значимая линейная связь, а также каковы оцененные значения коэффициентов для уравнения наилучшего предсказания. Кроме того, может быть определена относительная важность независимых переменных в предсказании зависимой переменной.

Рассмотрим множественную модель регрессии:

$$\begin{aligned}
 SUM_i = & \beta_0 + \beta_1 DTP_V_i + \beta_2 Time_i + \beta_3 osv_i + \beta_4 velo_i + \beta_5 moto_i \\
 & + \beta_6 days_of_week_i + \beta_7 kids_i + \beta_8 public_tr_i + \beta_9 truck_i \\
 & + \beta_{10} road_quality_i + \beta_{11} location_i + \beta_{12} weather_i + \beta_{13} alco_i \\
 & + \beta_{14} dr_viol_i + \beta_{15} ped_viol_i + \beta_{16} tr_c_i + \beta_{17} ptpnt_c_i \\
 & + \beta_{18} District_i + \beta_{19} date_i + \varepsilon_i
 \end{aligned} \quad (1)$$

$$i=1, \dots, N,$$

Где N – число ДТП, ε_i независимые и одинаково распределённые случайные величины.

С помощью $\beta_0 \dots \beta_{19}$ метода наименьших квадратов были найдены оценки неизвестных параметров которые представлены во втором столбце таблицы 9. В третьем столбце представлены средние квадратические отклонения соответствующих оценок.

Таблица 9

Результаты оценивания параметров модели (1)

Переменная	β	СКО	t	p
Константа	-0,412	0,124	-3,311	0,001
date	0,019	0,040	0,460	0,645
Time	-0,013	0,020	-0,646	0,519
District	-0,006	0,017	-0,351	0,726
DTP_V	0,111	0,017	6,492	0,000
osv	0,057	0,039	1,487	0,137



Переменная	β	СКО	t	p
velo	0,658	0,102	6,465	0,000
moto	0,247	0,053	4,655	0,000
days_of_week	0,004	0,009	0,441	0,659
kids	-0,099	0,061	-1,606	0,109
public_tr	0,013	0,069	0,185	0,853
truck	0,162	0,059	2,757	0,006
road_quality	0,000	0,008	0,052	0,959
location	-0,002	0,006	-0,259	0,796
weather	0,017	0,030	0,552	0,581
alco	0,027	0,145	0,183	0,855
dr_viol	0,012	0,004	3,092	0,002
ped_viol	-0,011	0,012	-0,843	0,399
tr_c	-0,664	0,031	-21,632	0,000
ptpnt_c	0,819	0,020	41,229	0,000

β – оценки коэффициентов регрессионного уравнения;

СКО – среднее квадратическое отклонения оценки;

t – эмпирическое значение t -критерия для проверки гипотезы о статистической значимости соответствующего коэффициента;

p – p -значение для проверки гипотезы о статистической значимости соответствующего коэффициента (вероятность ошибочного принятия гипотезы о существовании ненулевых коэффициентов регрессии).

Таблица 10

Сводка для модели (1)

R^2	Скорректированный R^2	СКО	Дарбин-Уотсон
0,743	0,736	0,466	1,931

Коэффициент множественной детерминации R^2 составляет 0,743, что говорит, о том, что 74,3% дисперсии зависимой переменной объясняется влиянием независимой переменной (чем ближе R^2 к 1, тем лучше построенная нами модель). Такое значение говорит о том, что модель учитывает большую часть факторов влияющих на число пострадавших в ДТП.

Стандартная ошибка оценки зависимой переменной 0,466.

Значение статистики Дарбина-Уотсона не выходит за границы [1,5;2,5], поэтому можно говорить о том, что автокорреляции нет.

Таблица 11

Дисперсионный анализ для модели (1)

	Сумма квадратов	ст. св.	Средний квадрат	F	Значимость
Регрессия	447,579	19	23,557	108,551	0,000
Остаток	154,729	713	0,217		
Всего	602,308	732			



Таблица 10 показывает суммы квадратов отклонений, значение статистики F -критерия Фишера, p -значение для Критерия Фишера, по которому можно судить о достоверности построенной связи переменных.

После исключения статистически незначимых характеристик (*days_of_week*, *road_quality*, *district*, *alco*, *date*, *public_tr*, *weather*, *location*, *time*) была построена новая модель регрессии с меньшим количеством переменных. Исключение переменных проводилось по одному и в зависимости от поведения скорректированного коэффициента детерминации были приняты решения о том, исключать переменную или нет. В итоге была получена следующая модель:

$$\begin{aligned}
 SUM_i = & \beta_0 + \beta_1 DTP_V_i + \beta_2 osv_i + \beta_3 velo_i + \beta_4 moto_i + \beta_5 kids_i \\
 & + \beta_6 truck_i + \beta_7 dr_viol_i + \beta_8 ped_viol_i + \beta_9 tr_c_i + \beta_{10} ptpnt_c_i \quad (2) \\
 & + \varepsilon_i, i = 1, \dots, N,
 \end{aligned}$$

Таблица 12

Результаты оценивания параметров модели (2)

Переменная	β	СКО	t	p
Константа	-0,421	0,077	-5,446	0,000
DTP_V	0,112	0,015	7,267	0,000
osv	0,073	0,034	2,178	0,030
velo	0,658	0,099	6,626	0,000
moto	0,242	0,052	4,676	0,000
kids	-0,101	0,060	-1,673	0,095
truck	0,163	0,058	2,819	0,005
dr_viol	0,012	0,004	3,112	0,002
tr_c	-0,665	0,030	-22,109	0,000
ptpnt_c	0,822	0,019	43,131	0,000
ped_viol	-0,010	0,012	-0,841	0,400

Таблица 13

Сводка для модели (2)

R^2	Скорректированный R^2	СКО	Дарбин-Уотсон
0,743	0,739	0,463	1,926

Таблица 14

Дисперсионный анализ для модели (2)

	Сумма квадратов	ст. св.	Средний квадрат	F	Значимость
Регрессия	447,229	10	44,723	208,215	0,000
Остаток	155,080	722	0,215		
Всего	602,308	732			



Коэффициент детерминации в новой модели практически не изменился, что подтверждает оправданность исключения девяти объясняющих переменных из модели. Таким образом можно сделать вывод о незначимости влияния на число пострадавших в ДТП календарной даты, дня недели, времени дня, качества дорожного покрытия, удалённости от центра города и погоды.

Из гистограммы стандартизированных остатков, приведённой на рисунке 1, можно сделать вывод об отсутствии выбросов и оправданности применения метода наименьших квадратов.

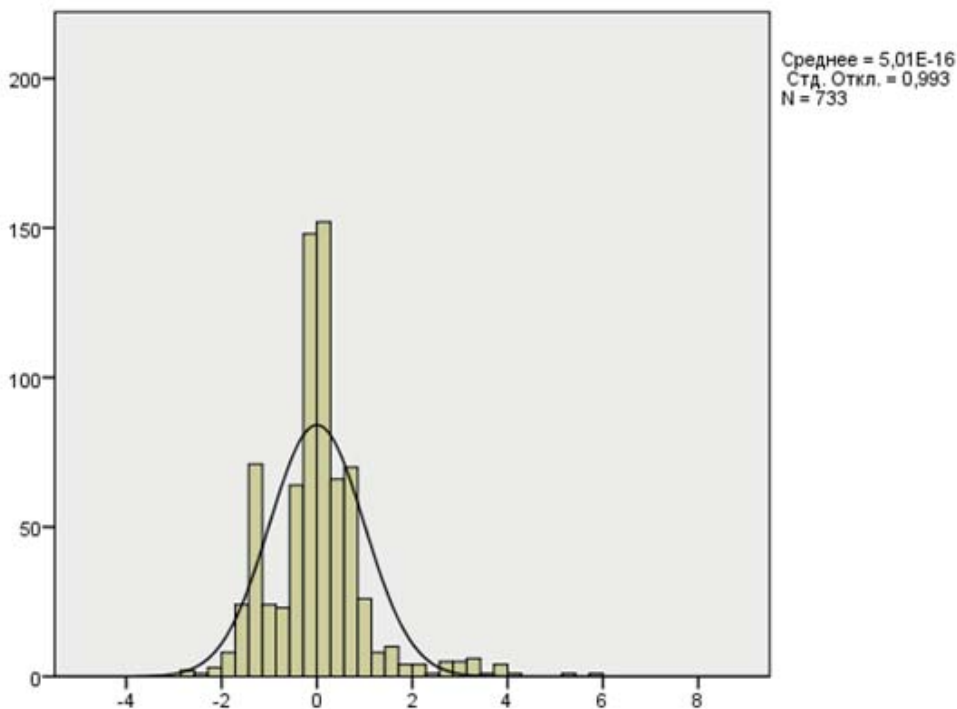


Рис.1. Гистограмма стандартизированных остатков

График $P-P$ стандартизированных остатков, приведённый на рисунке 2, также говорит в пользу предположения о нормальном распределении остатков. При нормальном распределении график представляет собой прямую $y = x$, наблюдаемый график достаточно слабо отклоняется от неё.

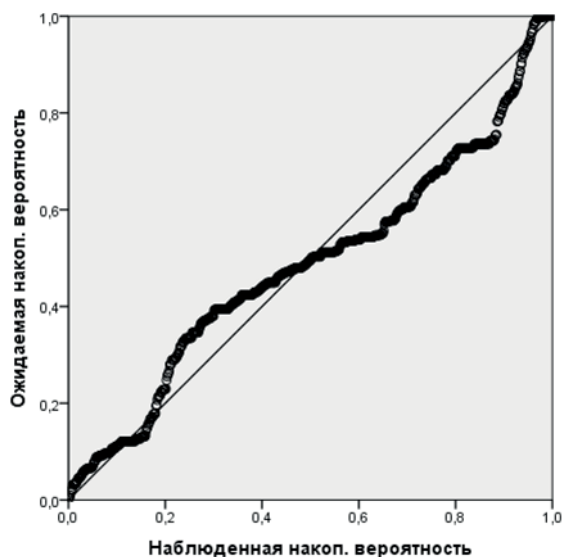


Рис.2. График P-P стандартизованных остатков

5. ЗАКЛЮЧЕНИЕ

В работе методами регрессионного анализа было проанализировано влияние документально фиксируемых характеристик ДТП на количество погибших и раненых в результате происшествия. С одной стороны, исследование показало, что некоторые характеристики значимо не влияют на количество пострадавших. С другой стороны, можно сделать вывод о том, что имеющихся в отчётах о происшествиях характеристик недостаточно для приемлемо полного объяснения всех причин гибели и ранения людей при ДТП. Для выяснения и документальной фиксации неучтённых в исследовании факторов необходимы дополнительные исследования.

Литература

1. Горяинова Е.Р., Панков А.Р., Платонов Е.Н. Прикладные методы анализа статистических данных: учеб. пособие; Нац. исслед. ун-т «Высшая школа экономики». – М.: Изд. дом Высшей школы экономики, 2012.
2. Демиденко Е.З. Линейная и нелинейная регрессии. М.: Финансы и статистика, 1981.
3. Есина М.Г., Лобова А.А., Хонгорова О.В., Шарабанова И.Ю. Статистический анализ дорожно-транспортных происшествий как сферы общественных отношений // Russian Journal of Education and Psychology. 2017. №6-2.
4. Махова О.А. Теоретические вопросы прикладного анализа смертности из-за ДТП // Статистика и экономика. 2014. №6.
5. Себер Дж. Линейный регрессионный анализ. – М.: Мир, 1980.
6. Фаттахов Т.А. Дорожно-транспортный травматизм в России и его факторы // Социальные аспекты здоровья населения. 2015. №4.



Study of the Influence of Various Factors on the Number of Victims in Road Accidents Using Regression Analysis Methods

Goryainov A.V.*

Moscow Aviation Institute (National Research University), Moscow, Russia,
e-mail: agoryainov@gmail.com

Polyakov E.A.**

Moscow Aviation Institute (National Research University), Moscow, Russia,
e-mail: polyakovevgeny96@mail.ru

In the paper is considered the dependence of the total number of dead and injured on various characteristics of the accident and the construction of models for estimating the total number of dead and injured.

Keywords: multiple regression, least squares method, traffic accident.

For citation:

Goryainov A.V., Polyakov E.A. Study of the Influence of Various Factors on the Number of Victims in Road Accidents Using Regression Analysis Methods. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2020. Vol. 10, no. 1, pp. 64–74. DOI: 10.17759/mda.2020100103 (In Russ., abstr. In Engl.)

References

1. Goryainova E.R., Pankov A.R., Platonov E.N. Applied Methods of Statistical Data Analysis (Prikladnye metody analiza statisticheskikh dannykh); National Research University “Higher School of Economics”. M.: Publishing house of the Higher School of Economics, 2012.
7. Demidenko E.Z. Linear and Nonlinear Regression (Lineynaya i Nelineynaya Regressii). M.: Finansy i statistika, 1981.
8. Esina M.G., Lobova A.A., Khongorova O.V., Sharabanova I.Yu. Statistical Analysis of Traffic Accidents as a Sphere of Public Relations (Statisticheskii Analiz Dorozhno-Transportnykh Proisshestviy Kak Sfery Obshhestvennykh Otnosheniy). Russian Journal of Education and Psychology. 2017. №6-2.
9. Makhova O.A. Theoretical Aspects of Applied Analysis of Deaths in Traffic Accidents (Teoreticheskie Voprosy Prikladnogo Analiza Amernosti iz-za DTP). Statistika i ekonomika. 2014. №6.
10. Seber Dzh. Linear Regression Analysis (Lineynyi Regressionnyi Analiz). M.: Mir, 1980.
11. Fattakhov T.A. Traffic Traumatism in Russia and its Factors (Dorozhno-transportnyi travmatizm v Rossii i ego faktory). Socialnye aspekty zdorovya naseleniya. 2015. №4.

***Goryainov Alexander Vladimirovich**, PhD (physics-mathematics), associate professor, Moscow Aviation Institute (National Research University), Moscow, Russia, e-mail: agoryainov@gmail.com

****Polyakov Evgeny Andreevich**, bachelor of applied mathematics, master’s degree student, Moscow Aviation Institute (National Research University), Moscow, Russia, e-mail: polyakovevgeny96@mail.ru