НАУЧНЫЙ ЦЕНТР
БЕЗОПАСНОСТИ ДОРОЖНОГО ДВИЖЕНИЯ
МВД РОССИИ

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА, ОЦЕНКИ И ПРОГНОЗИРОВАНИЯ ДОРОЖНО-ТРАНСПОРТНОЙ АВАРИЙНОСТИ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ



МОСКВА 2025

МИНИСТЕРСТВО ВНУТРЕННИХ ДЕЛ
РОССИЙСКОЙ ФЕДЕРАЦИИ

НАУЧНЫЙ ЦЕНТР БЕЗОПАСНОСТИ ДОРОЖНОГО ДВИЖЕНИЯ

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА, ОЦЕНКИ И ПРОГНОЗИРОВАНИЯ ДОРОЖНО-ТРАНСПОРТНОЙ АВАРИЙНОСТИ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

Москва
2025

УДК 51-74
ББК 39.808.020.3+22.1(7)
М77

Рекомендовано к опубликованию редакционно-издательским советом
Научного центра БДД МВД России

О.Ю. Монина, кандидат технических наук, доцент

Рецензенты:
В.В. Афонин – доцент кафедры организации деятельности Госавтоинспекции ОрЮИ МВД России имени В.В. Лукьянова, кандидат юридических наук, доцент.
А.Б. Чубуков – заместитель Генерального директора ОАО «НИИАТ», кандидат технических наук.

Монина О.Ю. Статистические методы анализа, оценки и прогнозирования дорожно-транспортной аварийности. М.: ФКУ «НЦ БДД МВД России», 2025. 56 с.

В учебно-методическом пособии приведены основные понятия статистического анализа и рассмотрены вероятностно-статистические методы применительно к исследованию дорожно-транспортной аварийности, чтобы сделать более понятным использование математического аппарата для получения обоснований принятия управленческих решений и выбора мероприятий, направленных на повышение безопасности дорожного движения.

Целью издания является углубление теоретических знаний путем разъяснения основных понятий, сути методов математической статистики и обоснованности выбора из множества существующих для решения практических задач. Особое внимание уделено интерпретации (объяснению) полученных результатов как основы для принятия управленческих решений, формирования мероприятий, направленных на повышение безопасности дорожного движения.

Учебно-методическое пособие адресовано в первую очередь преподавателям и слушателям образовательных организаций МВД России, а также тем, кто интересуется анализом статистических данных в сфере безопасности дорожного движения.

УДК 51-74
ББК 39.808.020.3+22.1(7)

© Монина О.Ю., 2025
© ФКУ «НЦ БДД МВД России», 2025

ОГЛАВЛЕНИЕ

Введение 6

1. Анализ данных 8

2. Статистические данные, элементы теории вероятностей и математической статистики 9

3. Дорожно-транспортная аварийность 12

4. Источники данных 15

5. Практическая часть 19

5.1. Визуальный анализ данных 19

5.2. Вероятностно-статистические методы 21

5.2.1. Сбор и подготовка данных 22

5.2.2. Корреляционный анализ 24

5.2.3. Метод главных компонент 28

5.2.4. Регрессионный анализ 32

5.2.5. Оценка значимости факторов в модели 36

5.2.6. Определение доверительной вероятности 38

6. Прогнозирование дорожно-транспортной аварийности 43

7. Интерпретация полученных результатов и формирование выводов 47

8. Программные способы обработки статистических данных 48

Заключение 52

Список использованных источников 53




ПОЛИЦИЯ

Введение

Повышение безопасности дорожного движения, имея экономическую и социальную значимость, является одним из приоритетных направлений государственной политики Российской Федерации, в реализации которого задействованы федеральные органы исполнительной власти, исполнительные органы субъектов Российской Федерации, органы местного самоуправления, общественные объединения, юридические и физические лица.

Координация этой деятельности возложена на Министерство внутренних дел Российской Федерации, а полномочия в области обеспечения безопасности дорожного движения осуществляет входящая в систему Государственная инспекция безопасности дорожного движения Министерства внутренних дел Российской Федерации (Госавтоинспекция), одним из основных направлений деятельности которой является проведение мероприятий по предупреждению причин возникновения дорожно-транспортных происшествий (далее также – ДТП) и снижению тяжести их последствий.

В основе принятия решения лежит анализ информации (обработка данных) и часто оценивание вероятности наступления события или правильного выбора действий.

Безопасность дорожного движения, являясь областью решения задач, законодательно определена как совокупность общественных отношений, возникающих в процессе перемещения людей и грузов с помощью транспортных средств или без таковых в пределах дорог[1], зависит от соблюдения установленных правил, нарушение которых приводит к дорожно-транспортным происшествиям, при которых погибают или получают ранения люди, повреждаются транспортные средства, сооружения, грузы либо причиняется иной материальный ущерб.

Состояние безопасности дорожного движения оценивается количественными показателями, основными из которых являются число дорожно-транспортных происшествий, пострадавших в них.

Одним из необходимых условий повышения безопасности дорожного движения, как показывает отечественная и зарубежная практика, а также результаты научных исследований, является подробный анализ и выявление причин и условий возникновения дорожно-транспортных происшествий, а также различных факторов, влияющих на аварийность. Результаты анализа ложатся в основу формирования комплекса мероприятий, направленных на решение проблемы обеспечения безопасности дорожного движения.

В течение нескольких десятилетий в соответствии с правилами учета дорожно-транспортных происшествий в стране накапливалась информация о ДТП, и в настоящий момент сформирована и продолжает актуализироваться база данных – совокупность сведений об аварийности (статисти-

[1] Федеральный закон от 10 декабря 1995 г. № 196-ФЗ «О безопасности дорожного движения» // СПС «Консультант Плюс» (дата обращения: 15.08.2025).

ка), следовательно, есть основа для применения статистических методов.

Настоящее учебно-методическое пособие создано в соответствии с отмеченной в Концепции научного обеспечения деятельности органов внутренних дел Российской Федерации на период до 2030 года[2] необходимостью разработки и внедрения в практику работы ведомственных образовательных и научных организаций комплекса научно-методических рекомендаций по применению новых средств и методов научного познания, в том числе методов математического анализа больших массивов информации, моделирования и прогнозирования социальных процессов.

Пособие подготовлено в дополнение к методическим рекомендациям «Анализ дорожно-транспортной аварийности»[3]. Большая часть теории с примерами применения математических методов и моделей в сфере обеспечения безопасности дорожного движения изложена в методических рекомендациях[4] и дополнена в монографии[5].

В этом издании рассмотрено применение вероятностно-статистических методов для оценивания данных в сфере безопасности дорожного движения, выявления значимых воздействий (факторов) на показатели дорожно-транспортной аварийности. Устраняя причины или принимая их во внимание, можно способствовать повышению безопасности дорожного движения.

Применение математических методов требует специальных знаний. В пособии даны определения понятиям (терминам), кратко изложена теория, необходимая для понимания материала, разъяснена суть применяемых вероятностно-статистических методов, предложена последовательность действий при решении конкретных задач анализа дорожно-транспортной аварийности, приведены примеры, дана их интерпретация и варианты обоснования выбора мероприятий, направленных на повышение безопасности дорожного движения.

Работа проведена с целью совершенствования научного и учебно-методического обеспечения аналитической деятельности Госавтоинспекции, формирования дополнительных материалов для преподавателей и слушателей образовательных организаций МВД России.

---

[2] Приказ МВД России от 13 ноября 2020 г. № 767 «Об утверждении Концепции научного обеспечения деятельности органов внутренних дел Российской Федерации на период до 2030 года»// СПС «Консультант Плюс» (дата обращения: 15.08.2025).

[3] Баканов К.С., Наумов С.Б., Ляхов П.В. и др. Анализ дорожно-транспортной аварийности: методические рекомендации. М.: «ФКУ НЦ БДД МВД России», 2024. 236 с.

[4] Капитанов В.Т., Монина О.Ю., Чубуков А.Б. Прикладные математические методы для анализа аварийности: методические рекомендации. М.: НИЦ БДД МВД России, 2014. 97 с.

[5] Капитанов В.Т., Монина О.Ю., Сильянов В.В., Чубуков А.Б. Математические методы и модели в сфере обеспечения безопасности дорожного движения: монография. М.: Техполиграфцентр, 2020. 102 с.

1. Анализ данных

Анализ данных – это процесс исследования, очистки, преобразования и моделирования данных с целью извлечения полезной информации, получения выводов и принятия решений. Этот процесс включает множество методов и подходов, которые позволяют понять структуру данных, выявить закономерности, тенденции, аномалии и взаимосвязи. Чаще всего применяется к большим объемам информации, которые невозможно обработать вручную.

Выделяют различные типы анализа данных.

Описательный анализ. Методами расчета средних значений, медианы, моды, стандартного отклонения, визуализацией (графики, гистограммы, круговые диаграммы) описываются, суммируются основные характеристики данных, чтобы объяснить, что произошло.

Диагностический анализ проводится с целью понимания причин произошедших событий или явлений. Для этого применяют такие методы, как корреляционный анализ, анализ временных рядов, поиск зависимостей.

Прогнозный (предиктивный, от англ. predict – предсказывать, пророчить, прогнозировать) анализ осуществляется с целью предсказания будущих событий на основе исторических данных. Подходят методы регрессионного анализа и машинного обучения (например, деревья решений, нейронные сети).

Предписывающий анализ посвящен определению оптимальных (наилучших) действий для достижения желаемых результатов методами оптимизации, реализацией сценариев «что если», симуляцией.

Текстовый анализ заключается в извлечении информации из текстовых данных.

Причинно-следственный (каузальный, от лат. causa – причина) анализ – это методология исследования причинно-следственных отношений между переменными. В отличие от корреляционного анализа, который лишь выявляет взаимосвязь, каузальный анализ стремится установить направление влияния: что является причиной, а что следствием.

Статистический анализ нацелен на проверку гипотез и анализ значимости данных, например, предположения о влиянии некоторого воздействия на состояние объекта или изменение процесса.

Анализ временных рядов пригоден для изучения данных, изменяющихся во времени, и прогнозирования будущих значений. В этом случае выбор методов зависит от задач и осуществляется из длинного списка, в который входит и регрессионный анализ.

Следует отметить, что анализ данных применяется в различных областях: бизнесе, медицине, финансах, маркетинге, науке и многих других. Выделяют следующие его этапы: сбор данных, очистка данных, исследование и визуализация данных, моделирование и анализ, интерпретация результатов и формулировка выводов.

В каждом случае выбор типа и метода анализа зависит от целей исследования и типа данных, а системный подход к обработке данных позволяет выявлять риски и принимать обоснованные управленческие решения.

2. Статистические данные, элементы теории вероятностей и математической статистики

С обработкой и анализом результатов наблюдений, измерений, испытаний (эмпирическими данными) для обоснования принятия решений имеют дело специалисты во всех отраслях практической деятельности и почти во всех областях научных исследований. Поэтому существуют общие подходы: сформированы терминология и методы, из которых в зависимости от целей и исходных данных выбираются необходимые и достаточные для решения конкретной задачи в обозначенной сфере.

Кратко и в самом общем виде статистический метод можно описать несколькими словами – это сбор, обобщение, представление, анализ и интерпретация данных. Поскольку статистика как наука изучает множество явлений и характерные для них закономерности, то и метод статистики представляет собой совокупность приемов, пользуясь которыми статистика исследует свой предмет.

Статистический анализ – методология, которая помогает извлекать важную информацию из данных и интерпретировать ее для принятия обоснованных решений. Задача – преобразовать сырые данные в осмысленные выводы, которые можно будет использовать для выявления закономерностей, прогнозирования или оптимизации (нахождения наилучшего варианта) процессов.

Основа статистического анализа – математическая статистика. Это раздел математики, изучающий методы сбора (систематизации), обработки и интерпретации статистических данных для научных и практических выводов.

Математическая статистика предоставляет инструменты для работы с выборками, проверки гипотез и построения моделей. В машинном обучении статистический анализ играет ключевую роль, так как именно он лежит в основе многих алгоритмов, используемых для обучения моделей.

Под данными (data) понимаются зафиксированные сведения о событиях, объектах или явлениях, представленные в формализованном виде, пригодном для передачи, интерпретации или обработки с помощью вероятностно-статистических методов, в основе которых лежат математическая статистика и теория вероятностей.

Теория вероятностей – раздел математики, предназначенный для разработки и исследования свойств математических моделей (формул), имитирующих механизмы функционирования реальных явлений или систем, условия «жизни» которых включают неизбежность влияния большого числа случайных, то есть не поддающихся строгому учету и контролю факторов[6].

[6] Капитанов В.Т., Монина О.Ю., Сильянов В.В., Чубуков А.Б. Математические методы и модели в сфере обеспечения безопасности дорожного движения: монография. М.: Техполиграфцентр, 2020. 102 с.

Но случайность не является беспричинностью, она подчинена внутренним, скрытым законам, которые нужно открыть[7].

Для понимания последующей теории рассмотрим несколько терминов.

Генеральная совокупность – это множество всех объектов, относительно которых предполагается делать выводы в рамках конкретного исследования.

Выборка – набор наблюдаемых значений или множество объектов, выделенных из изучаемой совокупности.

Распределение случайной величины (вероятностное распределение) – представляет собой закон, который каждому значению случайной величины ставит в соответствие вероятность принятия величиной этого значения. Сумма всех вероятностей в распределении равна единице.

Вероятность – это число от 0 (невозможное событие) до 1 (достоверное событие), показывающее, насколько событие вероятно (шанс его наступления).

Оценивание – это процесс измерения качества, эффективности или соответствия объекта, процесса или результата заданным критериям. Оценка как результат оценивания помогает определить успешность деятельности и внести улучшения, бывает количественная (баллы, рейтинги, метрики (KPI) и качественная (экспертные заключения).

Прогнозирование – разработка прогноза (вероятностного суждения о будущем состоянии объекта исследования), специальное научное исследование конкретных перспектив дальнейшего развития какого-либо процесса.

Регрессионный анализ – статистический метод исследования влияния одной или нескольких независимых переменных на зависимую переменную.

Тренд – основные тенденции изменения временного ряда или направление преимущественного движения показателей.

Математическая модель – математическое (аналитическое) моделирование (по Ляпунову) – это опосредованное практическое или теоретическое исследование объекта, при котором непосредственно изучается не сам интересующий объект, а некоторая вспомогательная искусственная или естественная система (модель), находящаяся в некотором объективном соответствии с познаваемым объектом, способная замещать его в определенных отношениях и дающая при ее исследовании в конечном счете информацию о самом моделируемом объекте (например, математическая модель мячика, который можно представить как сферу, – уравнение сферы).

Аппроксимация – это замена одних результатов другими, близкими к исходным, но упрощенными. В математике это построение кривых, близких к имеющемуся набору значений.

Интерполяция (вариант аппроксимации) – это нахождение неизвестных промежуточных значений с помощью нескольких известных.

[7] Маркс К., Энгельс Ф. Людвиг Фейербах и конец классической немецкой философии. Сочинения. Изд. 2-е. М.: Политиздат, 1961. Т. 21. С. 306.

Интерполировать – это, по сути, вставить дополнительные значения в ряд.

Экстраполяция (вариант аппроксимации) – это метод нахождения значений за пределами интервала. На основе имеющихся значений строится предположение, какие значения этот показатель будет принимать в будущем (используется для прогнозирования).

Исходным материалом для статистического исследования служат статистические данные – сведения о числе объектов в какой-либо совокупности, обладающих теми или иными признаками.

Под статистическими данными понимают числовые или нечисловые значения контролируемых параметров (признаков) исследуемых объектов, которые получены в результате наблюдений (измерений, анализа, испытаний и т.д.) определенного числа признаков у каждой единицы, вошедшей в исследование.

Способы получения статистических данных и объемы выборок устанавливают исходя из постановок конкретной прикладной задачи. Часто исследуется совокупность количественных значений статистических показателей о социально-экономических, демографических и иных явлениях.

В формализованном виде результат наблюдения x_i исследуемого признака x (или совокупности исследуемых признаков X) у *i*-й единицы выборки отражает количественные и (или) качественные свойства обследованной единицы с номером *i* (*i*=1, 2, … , *n*, где *n* – объем выборки).

Вид обрабатываемых результатов наблюдений определяет направление прикладной статистики (статистика случайных величин, временных рядов, многомерный статистический анализ).

Результаты наблюдений обрабатывают методами прикладной статистики, соответствующими поставленной задаче. Используют, как правило, аналитические методы (основанные на численных расчетах). В отдельных случаях допустимо применение графических методов (визуального анализа).

Количество методов обработки данных велико, и они описаны в многочисленных книгах, статьях, нормативно-технических инструкциях и методиках. Многие методы прикладной статистики требуют проведения трудоемких расчетов, поэтому для их реализации используют компьютеры и программное обеспечение.

При этом качество расчетов зависит от обоснованно выбранного аппарата и достоверности анализируемых данных.

3. Дорожно-транспортная аварийность

Понятие «дорожное движение» законодательно определено[8] как совокупность общественных отношений, возникающих в процессе перемещения людей и грузов с помощью транспортных средств или без таковых в пределах дорог.

То есть это целенаправленное взаимодействие участников с помощью технических средств в границах некоторого пространства при различных условиях, которое можно описать общепринятой, но с некоторыми интерпретациями системой «Водитель (Человек) – Автомобиль (Транспортное средство) – Дорога – Среда» (ВАДС).

На обеспечение безопасности дорожного движения путем предупреждения и снижения тяжести последствий ДТП направлена деятельность федеральных органов исполнительной власти, исполнительных органов субъектов Российской Федерации, органов местного самоуправления, общественных объединений, юридических и физических лиц.

Их действия и решения должны быть целенаправленными, обоснованными и объяснимыми, то есть для достижения успешного результата нужно понимать, что делать, как и зачем.

Основополагающим элементом в принятии обоснованных решений являются результаты анализа информации о состоянии системы и наличие взаимодействия (взаимовлияния) ее элементов.

Для обоснования принятия управленческих решений и выбора мероприятий в сфере безопасности дорожного движения формулируются и решаются различные аналитические задачи.

К ним относятся анализ единичных ДТП, их совокупностей, дорожно-транспортной аварийности, оценка эффективности (соответствие между достигнутым результатом и использованными ресурсами, выраженное в натуральных или стоимостных показателях), а точнее результативности (уровень достижения запланированных результатов) мероприятий, направленных на повышение уровня безопасности дорожного движения.

Уровень безопасности дорожного движения характеризуется численными значениями. Согласно статье 9 Федерального закона от 10 декабря 1995 г. № 196-ФЗ «О безопасности дорожного движения»[9] основные показатели состояния безопасности дорожного движения в России:

> количество дорожно-транспортных происшествий, пострадавших в них граждан, транспортных средств, водителей транспортных средств;
>
> число нарушителей правил дорожного движения, административных правонарушений и уголовных преступлений в области дорожного движения;
>
> другие показатели, отражающие состояние безопасности дорожного движения и результаты деятельности по ее обеспечению.

[8] Федеральный закон от 10 декабря 1995 г. № 196-ФЗ «О безопасности дорожного движения» // СПС «Консультант Плюс» (дата обращения: 15.08.2025).

[9] Там же.

В зависимости от поставленной задачи любой из перечисленных показателей (в большинстве случаев обобщенная характеристика какого-либо объекта, процесса или его результата, понятия или их свойств, обычно выраженная в числовой форме) можно рассматривать в качестве критерия, на основании значения которого формируется оценка состояния безопасности дорожного движения.

В данном учебно-методическом пособии в качестве приоритетного (главного) критерия предлагается рассматривать абсолютное число погибших в ДТП, затем число раненых и количество ДТП, называя комплекс этих показателей безопасности дорожного движения на определенной территории за конкретный период времени дорожно-транспортной аварийностью, а сформированный массив данных – статистикой дорожно-транспортных происшествий.

В настоящее время действуют Правила учета дорожно-транспортных происшествий, утвержденные постановлением Правительства РФ от 19 сентября 2020 г. № 1502[10].

В Министерстве внутренних дел Российской Федерации учет ДТП осуществляется путем внесения сведений о ДТП в электронную карточку учета ДТП с использованием Единой информационно-аналитической системы обеспечения безопасности дорожного движения МВД России (ЕИАС БДД), автоматизированной информационно-управляющей системы Госавтоинспекции (АИУС), автоматизированной системы учета ДТП и многопараметрической информационно-аналитической системы прогнозирования и моделирования ситуации в области обеспечения безопасности дорожного движения (МИАС).

Формируемые базы данных являются основой анализа информации, характеризующей как отдельные ДТП, так и их некоторую совокупность.

Основным количественным мерилом безопасности дорожного движения является дорожно-транспортное происшествие, его характеристики, причины, условия. Наличие количественных данных – основа применения статистических методов.

Дорожно-транспортное происшествие возникает в результате формирования неких неблагоприятных случайных межэлементных взаимодействий в сложной системе ВАДС. Причины и обстоятельства совершения каждого ДТП являются результатом случайного стечения обстоятельств, поэтому происшествие можно рассматривать как случайное, но небеспричинное событие, а их совокупность (в течение упорядоченного периода времени или за фиксированный временной период, но в различных субъектах Российской Федерации) – как последовательность случайных событий.

В этом случае обоснованно применение методов теории вероятностей и математической статистики, математического моделирования, в частности, для анализа и выявления закономерностей, определяющих

[10] Постановление Правительства РФ от 19 сентября 2020 г. № 1502 «Об утверждении Правил учета дорожно-транспортных происшествий, об изменении и признании утратившими силу некоторых актов Правительства Российской Федерации» // Собрание законодательства РФ. 28.09.2020. № 39. Ст. 6077.

влияние различного рода факторов на возникновение дорожно-транспортных происшествий.

На основании результатов исследований[11, 12, 13] определено, что основными факторами являются численность населения и количество транспортных средств, зарегистрированных на территории субъекта Российской Федерации.

Представляет интерес выявить и оценить связь иных факторов на состояние критерия. Исходя из определения дорожного движения и описывающей его системы (ВАДС), полезными для исследования влияния могут быть показатели, характеризующие каждую составляющую и состояние регионов: демографическую, социальную, экономическую, природно-климатическую, а также принимаемые меры воздействия на участников дорожного движения. В зависимости от решаемой задачи могут рассматриваться наборы различных показателей.

В числе показателей, характеризующих особенности демографического и социального состояния регионов, могут быть численность населения, возрастной состав, численность экономически активного населения, численность пенсионеров.

Группа показателей, характеризующих экономическое развитие региона, включает валовый региональный продукт на душу населения, среднедушевые денежные доходы населения, численность населения с доходами ниже прожиточного минимума, потребительские расходы на душу населения и др.

Группа показателей, характеризующих социальный и культурный уровни региона, содержит данные о численности детей в дошкольных и школьных образовательных учреждениях, численности студентов средних специальных учебных заведений на 10 тыс. жителей, численности студентов вузов на 10 тыс. жителей; заболеваемости на 1 тыс. человек населения, численности зрителей театров на 1 тыс. человек населения, числе посещений музеев на 1 тыс. человек населения, розничной продаже различных видов алкогольных напитков и др.

Группа показателей, характеризующих транспортную систему, – численность зарегистрированных транспортных средств, пассажирооборот автобусов общего пользования, численность и средний возраст парка транспортных средств, в том числе по категориям, плотность автомобильных дорог общего пользования (протяженность/площадь), удельный вес автомобильных дорог с усовершенствованным покрытием в протяженности дорог с твердым покрытием общего пользования, число водителей, их возрастной состав, подготовка водителей и др.

Группа показателей, связанных с правонарушениями или гибелью людей, – количество зарегистрированных правонарушений в сфере безопасности дорожного движения; суммарный объем штрафов за админи-

[11] Капитанов В.Т., Монина О.Ю., Чубуков А.Б. Прикладные математические методы для анализа аварийности: методические рекомендации. М.: НИЦ БДД МВД России, 2014. 97 с.

[12] Капитанов В.Т., Чубуков А.Б., Монина О.Ю. и др. О факторах, определяющих дорожно-транспортную аварийность в регионах России // Наука и техника в дорожной отрасли. 2019. № 4. С. 18–23.

[13] Капитанов В.Т., Монина О.Ю., Чубуков А.Б. Модельная оценка взаимосвязи различных показателей, связанных с гибелью людей // Наука и техника в дорожной отрасли. 2016. № 2.

стративные правонарушения в сфере безопасности дорожного движения (начислено/оплачено), количество лишений права на управление транспортными средствами; уровень преступности по видам; число погибших при пожаре; количество суицидов (самоубийств); число погибших на воде; число зарегистрированных преступлений на 100 тыс. населения и др.

Целесообразно рассмотреть взаимосвязь дорожно-транспортной аварийности и природно-климатических условий субъектов Российской Федерации, например, соответствующую классификацию регионов, среднегодовую температуру воздуха в столице субъекта и другие доступные количественные данные.

4. Источники данных

Основным источником информации о дорожно-транспортной аварийности является база данных, автоматически формируемая в Многопараметрической информационно-аналитической системе Госавтоинспекции МВД России (МИАС) на основе обработки массива электронных карточек учета ДТП, являющихся документом первичного учета ДТП.

Перечень сведений, предусмотренных основными разделами электронной карточки учета дорожно-транспортного происшествия, утверждается приказом МВД России. В настоящее время действует приказ МВД России от 12 августа 2022 г. № 598[14], регламентирующий организацию учета, сбора и анализа сведений о дорожно-транспортных происшествиях (вместе с «Порядком организации учета, сбора и анализа сведений о дорожно-транспортных происшествиях, контроля за полнотой и достоверностью этих сведений»).

При этом карточка учета отдельного ДТП может содержать текстовую, числовую, графическую информацию, пригодную для изучения и вынесения суждения. Качественную информацию – текст, схемы, фотоматериалы можно оценить визуально, а при использовании соответствующих технологий – автоматизировано, в том числе с помощью искусственного интеллекта.

Количественную информацию (данные), например, число участников, их возраст, обозначаемые цифрами, причины, условия ДТП и содержание иных позиций, выбираемых из предложенных по принципу «да/нет», анализируют с помощью математических методов.

Одним из условий качественного анализа является достоверность данных. Поэтому для исследования, по возможности, отбираются официальные статистические показатели, публикуемые на официальных сайтах федеральных органов исполнительной власти, а также в Единой межведомственной информационно-статистической системе (ЕМИСС)[15], координируемой Федеральной службой государственной статистики.

[14] Приказ МВД России от 12 августа 2022 г. № 598 «Об организации учета, сбора и анализа сведений о дорожно-транспортных происшествиях и признании утратившим силу нормативного правового акта МВД России» (вместе с «Порядком организации учета, сбора и анализа сведений о дорожно-транспортных происшествиях, контроля за полнотой и достоверностью этих сведений») // СПС «Консультант Плюс» (дата обращения: 12.10.2025).

[15] Дубров А.М и др. Многомерные статистические методы. М., Финансы и статистика, 2011.

Это государственный информационный ресурс, объединяющий официальные государственные информационные статистические ресурсы, доступные на сайте службы www.fedstat.ru. Данные формируются 61 ведомством по 8 844 показателям (рис. 1) способом наблюдения, что сводит к минимуму воздействие на наблюдаемый объект и способствует повышению достоверности.

Например, по информации МВД России, формируется раздел, включающий данные об аварийности («Дорожно-транспортные происшествия») и иные, входящие в раздел «Безопасность дорожного движения» (рис. 2). В отсутствии доступа к МИАС можно воспользоваться этим источником.

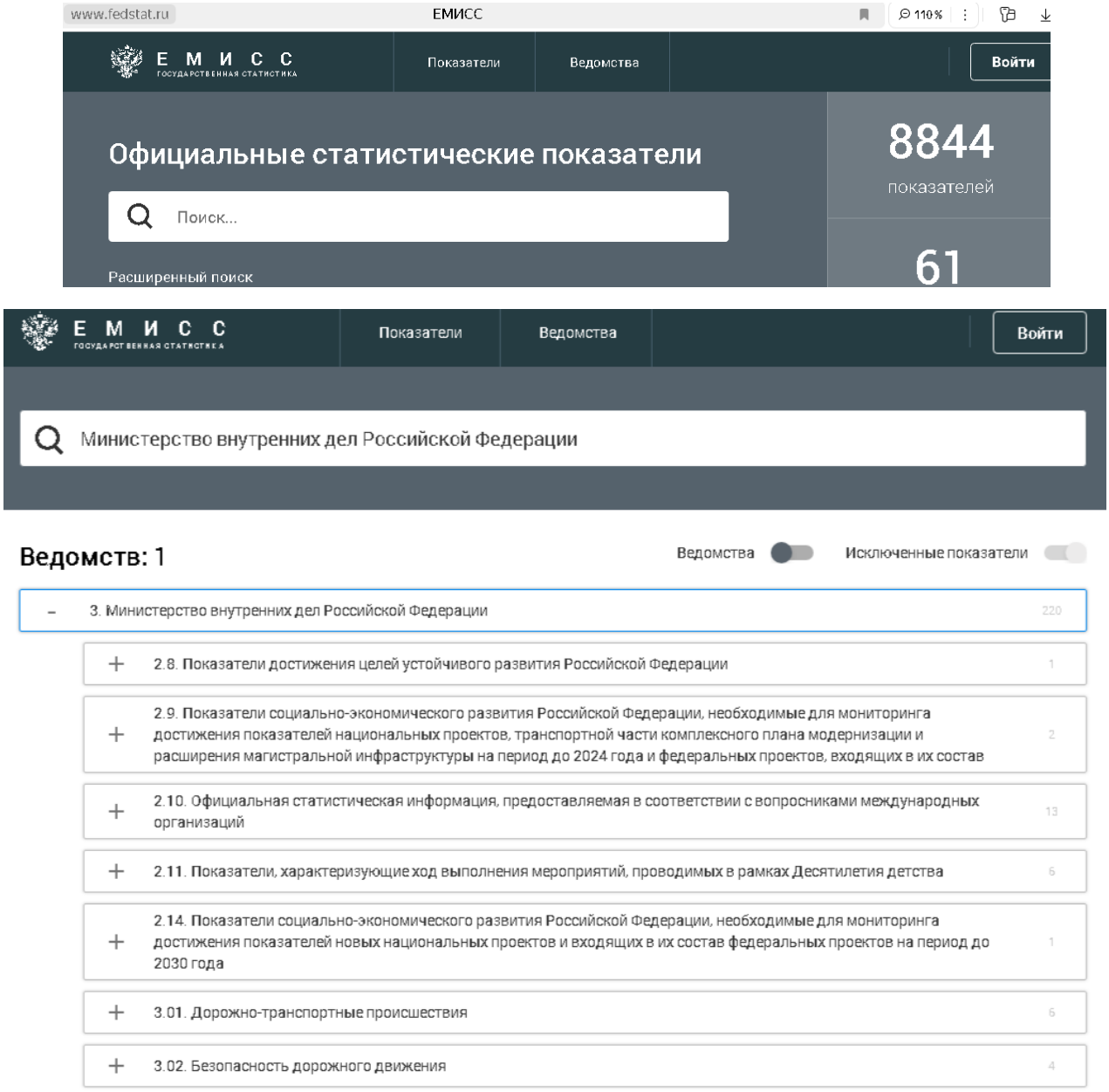


Рисунок 1. Единая межведомственная информационно-статистическая система

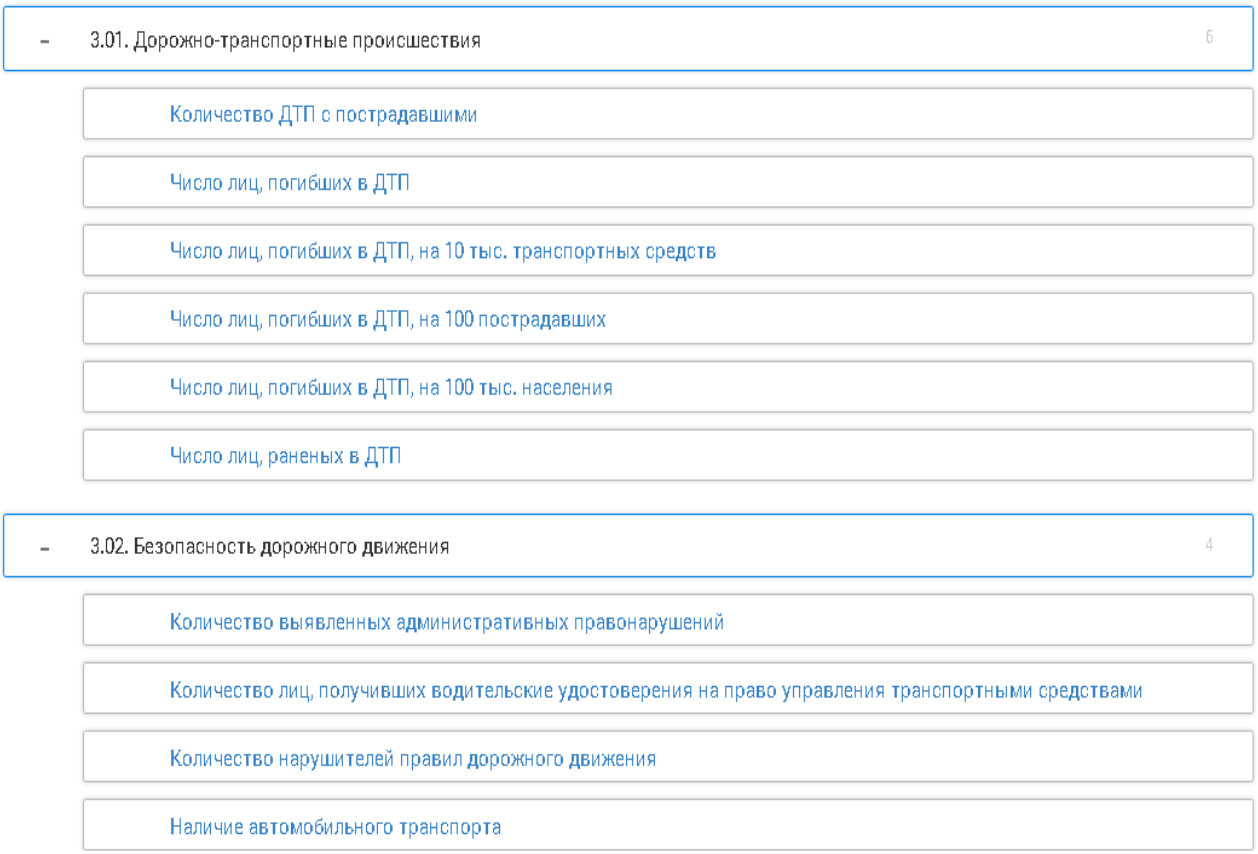


Рисунок 2. Содержание разделов «Дорожно-транспортные происшествия» и «Безопасность дорожного движения»

Еще одним источником данных для исследования может быть сводная информация об основных показателях дорожно-транспортной аварийности в субъектах Российской Федерации, дорожно-транспортной инфраструктуре и численности парка транспортных средств, а также социально-экономических показателях, непосредственно связанных или имеющих влияние на обеспечение безопасности дорожного движения, ежегодно формируемая в профилях безопасности дорожного движения субъектов Российской Федерации[16] (рис. 3).

Информацию о природно-климатическом состоянии регионов можно почерпнуть в Строительных нормах и правилах (СНиП), результатах научных исследований.

[16] Баканов К.С., Ляхов П.В. и др. Профили безопасности дорожного движения субъектов Российской Федерации 2024: статистический сборник. М.: ФКУ «Научный центр БДД МВД России», 2025. 104 с.

17

Республика Коми

Административный центр – Сыктывкар
Площадь – 416 774 км²
Количество населенных пунктов – 757

Аварийность

Социальный риск / ранг	8,50 /	19
Транспортный риск / ранг	1,93 /	19
Дорожный риск / ранг	0,76 /	30

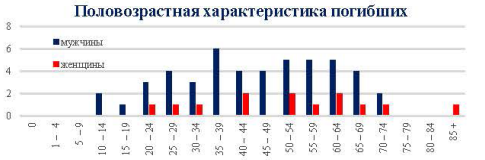
Основные показатели аварийности



Помощь пострадавшим



Половозрастная характеристика погибших



Детский дорожно-транспортный травматизм

	ДТП	У.В.	погибло	У.В.
Всего с участием	136	15,9%	2	3,3%
Пассажиры	57	41,9%	2	100%
Пешеходы	47	34,6%	0	0
Водители	19	14,0%	0	0
Доля от всех умерших детей / ранг			1,9% /	8
ЮИД (чел.)			2 452 /	2,9%

Правонарушения в сфере БДД

Количество преступлений		458
Возбуждено дел (КоАП РФ)		465 500
Из них:		
Превышение скорости	295 370 /	63,5%
Выезд в нарушение ПДД на полосу встречного движения	4 381 /	0,9%
Нарушение ПДД пешеходами	8 498 /	1,8%
Непредоставление преимущества в движении пешеходам	4 983 /	1,1%
Управление ТС в состоянии опьянения	2 120 /	0,5%
Из них:		
В состоянии наркотического опьянения	121 /	5,7%
Сумма оплаченных штрафов(тыс. руб.)		284 898

Основные факторы аварийности

	ДТП	У.В.	погибло	У.В.
Управление ТС с признаками опьянения	77	9,0%	15	24,6%
Наезд на пешеходов	210	24,5%	10	16,4%
Несоответствие скорости	119	13,9%	9	14,8%
Выезд на полосу встречного движения	115	13,4%	34	55,7%

Транспорт

Парк транспортных средств / ранг	315 665 /	65
Из них:		
Легковые	267 069 /	84,6%
Грузовые	25 845 /	8,2%
Автобусы	4 777 /	1,5%
Мототранспорт	17 974 /	5,7%
Автомобилизация / ранг	441,6 /	41
Плотность движения / ранг	39,2 /	30

Улично-дорожная сеть

Протяженность дорог (км) / ранг	8 049,8 /	65
Из них:		
Федеральные (км)	350,2 /	4,4%
Региональные и межмуниципальные (км)	4 723,6 /	58,7%
Местные (км)	2 976,0 /	37,0%
Плотность дорог (км/км²) / ранг	19,3 /	76

Демография

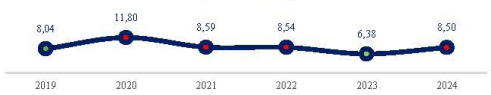
Население / ранг	714 785 /	62
ОПЖ (лет) / ранг	70,4 /	67
Смертность от алкоголя (социальный риск) / ранг	85,6 /	84
Продажа алкоголя на душу населения (л) / ранг	11,2 /	83

Экономика

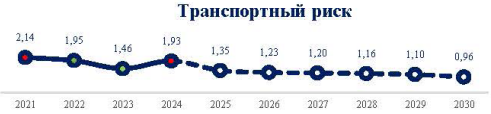
ВРП (тыс. руб.) / ранг	1 450,4 /	12
Безработица (тыс. чел./%)	13,4 /	3,7%
Рег. проект (млн) / ранг	0,2 /	54
Пассажирооборот		480 578,4
Грузооборот (млн тонн-км)		1 523,0
Дорожный фонд (млрд руб./%)	6,6 /	5,3%

Относительные показатели аварийности

Социальный риск



Транспортный риск



НАУЧНЫЙ ЦЕНТР БДД МВД РОССИИ

ПРОФИЛИ СУБЪЕКТОВ РФ
2024

Рисунок 3. Профиль безопасности дорожного движения субъекта Российской Федерации

5. Практическая часть

Следует обратить внимание на то, что поиск решения должен быть осознанным, целенаправленным, адресным и с пониманием «физики» процесса. То есть важно знание структуры, взаимосвязи элементов, правил изменений изучаемой системы или сферы.

5.1. Визуальный анализ данных

Визуальный анализ данных – это процесс исследования данных с помощью графических представлений (графиков, диаграмм, карт и др.). Он позволяет быстро выявлять закономерности, аномалии и тенденции, которые трудно обнаружить при анализе сырых чисел. Визуальное представление данных помогает лучше понять информацию и сделать акцент на важнейших аспектах.

Визуальный анализ направлен:

на понимание данных – выявление структуры, распределения и взаимосвязей;

на обнаружение аномалий – поиск выбросов и ошибок в данных;

на коммуникацию результатов – наглядное представление информации для принятия решений.

Существуют различные инструменты и технологии визуального анализа: графики, диаграммы, карты в GIS-системах для геопространственной визуализации, интерактивные инструменты и библиотеки программирования.

Преимуществом является ускорение анализа данных, а также простота восприятия сложной информации, недостатком – риск субъективной интерпретации и ограничения при работе с большими данными.

Визуальный анализ данных – мощный инструмент для исследования и представления информации. При правильном применении он значительно упрощает работу с данными и помогает в принятии решений.

Этот вид анализа применяется в различных сферах (бизнес, наука, социальные исследования), а также для анализа дорожно-транспортной аварийности и информации, касающейся дорожно-транспортных происшествий (рис. 4, 5).

Большое внимание графическому представлению информации и визуальному анализу уделено в методических рекомендациях «Анализ дорожно-транспортной аварийности»[17].

[17] Баканов К.С., Наумов С.Б., Ляхов П.В. и др. Анализ дорожно-транспортной аварийности: методические рекомендации. М.: ФКУ «НЦ БДД МВД России», 2024. 236 с.

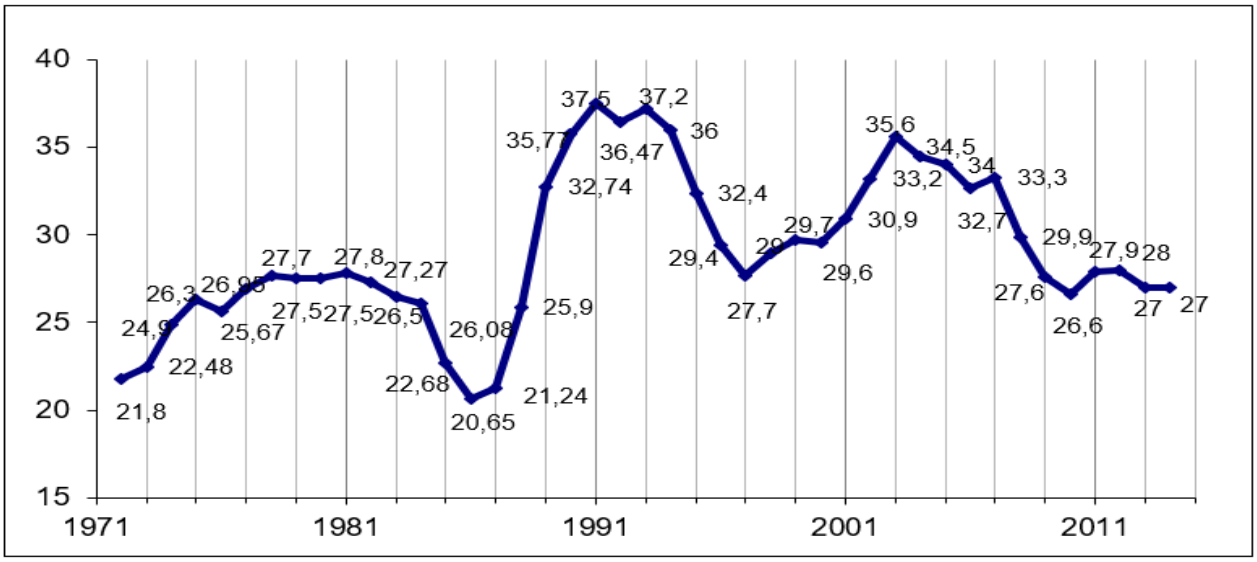


Рисунок 4. График числа погибших
в ДТП в России с 1972 по 2014 год

Например.

Если данные разместить в таблице, то сложно выделить продолжительные изменения и пики (экстремумы). На графике видны периоды увеличения показателя, например, с 1972 по 1981 год, с 1986 по 1991 год, а затем с 1997 по 2003 год, сменяющиеся периодами снижения, соответственно с 1981 по 1986, с 1993 по 1997 год, а затем и с 2003 по 2014 год (и по настоящее время).

На графике локальные минимумы отмечались в 1972, 1986, 1997 и в 2010 годах. Аналогично выделяются максимумы с абсолютным – 37,5 тыс. погибших в 1991 году. На основе анализа происходящих в стране событий и принимаемых мер по снижению дорожно-транспортной аварийности, можно экспертно оценить, что влияло на показатель безопасности дорожного движения.



Рисунок 5. ДТП, связанные с нарушением скоростного режима
в России в 2024 году

Дорожно-транспортные происшествия, связанные с нарушением скоростного режима движения, включают события, произошедшие из-за

несоответствия скорости конкретным условиям движения (92%) и превышения установленной скорости движения (8%)[18] – на круговой диаграмме видно это соотношение.

На гистограмме представлена в пропорции тяжесть последствий при первом (13,5) и втором нарушении (9,8). Очевидно, что случаев превышения установленной скорости движения весомо меньше, но тяжесть последствий значительно выше, чем из-за несоответствия скорости конкретным условиям движения. Однако, что влияет на нарушения и тяжесть из графического представления данных, непонятно.

Таким образом, метод визуализации, имея свои преимущества, при работе с большими данными для выявления взаимозависимости показателей процессов не работает.

5.2. Вероятностно-статистические методы

Вероятность – это число, указывающее, насколько возможным является результат случайного события.

Вероятность события – это число, заключенное в интервале от 0 до 1. Вероятность невозможного события равна 0, вероятность достоверного события равна 1. Если событие наступает в некоторых случаях, его вероятностью будет число, лежащее в интервале от 0 до 1. Но следует понимать, что даже если вероятность некоторого события очень высока, это не гарантирует, что событие абсолютно точно произойдет в каком-то конкретном случае.

Вероятностно-статистические методы позволяют решать задачи в условиях неопределенности, учитывать действие множества случайных факторов в массовых явлениях случайного характера и выявлять закономерности при взаимодействии большого числа случайных факторов.

Характеристики аварийности, полученные при обработке данных достаточно большой группы дорожно-транспортных происшествий, подчиняются статистическим закономерностям.

Их существование определяется тем, что даже если каждое конкретное ДТП возникает в результате сочетания многих факторов[19] и потому может носить случайный характер, то для достаточно больших совокупностей ДТП действие этих факторов принимает устойчивый характер, что создает предпосылки для создания математических моделей.

Методология вероятностно-статистического анализа включает следующие основные этапы, каждый из которых направлен на достижение достоверных и интерпретируемых результатов:

1. Постановка задачи, включающая определение цели исследования, совокупности и типа данных.

2. Сбор и первичная подготовка данных.

3. Описательный анализ.

[18] Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: Физматлит, 2006. 816 с.

[19] Фактор – показатель, влияющий на изменение значения критерия.

4. Проверка статистических гипотез.

5. Корреляционный анализ.

6. Регрессионный анализ.

7. Вероятностное моделирование.

8. Визуализация и интерпретация результатов, разработка рекомендаций для практического применения.

5.2.1. Сбор и подготовка данных

Первым этапом исследования является сбор и подготовка исходных данных, формирование массива исходных данных, их предварительная обработка, которая включает очистку, преобразование и структурирование данных для повышения их качества и пригодности для статистического анализа.

Сбор и загрузка включают:

получение данных из различных источников;

проверку на полноту и согласованность структуры.

Очистка данных подразумевает:

обработку пропущенных значений путем удаления строк/столбцов с пропусками (например, данные, относящиеся к Республике Крым, г. Севастополю до 2015 года);

замену средним, медианой или значениями, полученными методом интерполяции (при отсутствии данных, но необходимости их учета в исследовании);

устранение выбросов (например, в отдельных исследованиях целесообразно исключить из общей последовательности большие значения показателей для г. Москвы или незначительные для отдельных регионов);

исправление ошибок (коррекция опечаток, дубликатов и несоответствующих форматов).

Преобразование данных: нормализация и стандартизация, соотнесение размерности (число погибших может рассматриваться в единицах, численность зарегистрированных транспортных средств и населения – в тысячах единиц), логарифмирование и другие преобразования.

Выполнение указанных действий улучшает качество моделей, повышает точность статистических тестов, позволяет избежать ложных выводов из-за «мусорных» данных.

Таким образом, предварительная обработка данных – важный этап, влияющий на достоверность статистического анализа. Грамотная подготовка данных экономит время, снижает риск ошибок и повышает качество результатов.

Пример таблицы с исходными данными в разрезе субъектов Российской Федерации представлен в таблице 1, по России – в таблице 2. Перечень показателей может быть различен в зависимости от решаемой задачи – чем больше выборка (количество субъектов, временной период), тем достовернее будет результат анализа.

Таблица 1

Исходные статистические данные (фрагмент)

2016-2020	Погибло в ДТП, всего	Ранено в ДТП, всего	ДТП, всего	Население (тыс) на 1 янв след года	Транспорт (тыс)	Бензин только	Плотность автомоб дорог	Доля АД общего пользования, отвечающих нормативным требованиям на конец года регионального или мужмеципального значения	Средний возраст ТС, лет, общий	Средний возраст ТС, PC легковые	Средний возраст ТС, LCV легкие коммерческие	Средний возраст ТС, HCV грузовые	Средний возраст ТС, BUS автобусы	Расходы консолидированного бюджета, млн. руб 2015-2018 2019	Возбуждено дел об адм правонарушениях всего, тыс.	**В том числе**. с помощью ФВФ,тыс
Источник, Форма, Строка	гибдд.рф	гибдд.рф	гибдд.рф		гибдд.рф (МИАС, ЕМИСС)	fuelbroker.ru	Росстат, ЕМИСС	ЕМИСС 1.24.5. Наличие путей сообщения	autostat.ru (2020 - расчет)					gks.ru/bgd/regl/b19_14p/Main.htm 2019 год - сайт казначейства	МИАС, 1-БДД, приказ 79, раздел 1, 560, 1	МИАС,1-БДД, приказ 79, раздел 1, 560, 6
	1	2	3	4	5	9	6	7	8	8а	8б	8в	8г	9	10	10а
Алтайский край	257	4061	3206	2365	834,063	0,52	220,66	48,8	17,4	16,8	16,0	23,6	18,0	94291,4	665	167
Амурская область	100	1663	1249	801	294,716	0,2	34,1409	76,1	19,4	18,6	21,1	22,0	14,3	54292,8	267	43
Архангельская область	147	2050	1580	1121	399,614	0,22	29,4708	13,8	11,7	11,1	12,5	18,0	12,4	81717,5	409	178
Астраханская область	120	1675	1305	1018	351,116	0,18	88,0796	36,2	14,6	13,9	13,9	23,4	18,8	40457,8	242	50
Белгородская область	212	1624	1368	1552	617,27	0,36	703,694	65,7	13,5	12,9	14,1	19,4	14,6	85180,6	1071	836
Брянская область	173	1623	1321	1220	309,182	0,21	306,765	47,1	13,9	13,5	13,6	18,5	15,8	57338,3	469	196
Владимирская область	315	3270	2476	1389	482,979	0,26	335,625	47,2	12,4	12,0	12,3	17,9	16,5	60359,6	523	227
Волгоградская область	345	3393	2686	2535	1174,039	0,32	140,872	37,5	13,9	13,2	14,1	21,3	16,6	108067,2	851	542
Вологодская область	122	2144	1591	1183	487,027	0,24	117,785	19,9	12,1	11,5	12,7	18,3	15,1	62535,4	1003	701
Воронежская область	489	3921	3088	2335	933,748	0,64	327,471	66,9	14	13,4	13,3	21,2	16,7	104595,4	2276	1963
Еврейская автономная обл.	24	306	249	164	48,303	0,03	67,5262	45	21	20,3	22,6	23,9	16,0	12090,2	53	22
Забайкальский край	207	1867	1395	1078	389,682	0,14	33,5376	25,4	18,8	18,4	17,4	23,7	16,6	55405,2	454	315
Ивановская область	106	2094	1543	1023	382,991	0,09	332,944	33,6	13,6	12,8	13,1	21,4	17,0	38486,9	305	148
Иркутская область	365	4310	3482	2408	780,458	0,68	29,9666	26,5	18,1	17,4	18,1	22,4	19,8	152046,7	772	359
Кабардино-Балкарская Респ.	139	833	664	864	261,786	0,06	564,832	56,6	15,9	15,2	14,7	24,5	20,9	34117,4	1055	782
Калининградская область	169	1540	1377	986	440,451	0,22	513,066	31,5	18,6	18,8	17,4	18,5	20,1	88326,4	496	203
Калужская область	206	2423	1830	1014	392,315	0,51	321,349	50	14,2	13,1	13,2	23,2	19,8	62966,4	871	722
Камчатский край	69	729	551	314	203,421	0,07	4,18738	41,5	21,9	21,6	21,6	24,5	21,3	72624,4	235	145
Карачаево-Черкесская Респ.	87	742	533	466	128,476	0,05	340,783	41,5	14,6	13,4	16,0	22,5	19,0	21895,9	196	124
Кемеровская область	341	4107	3054	2708	920,011	0,61	173,51	34,8	13,8	13,3	15,1	19,4	12,8	141464,5	1307	401
Кировская область	172	2026	1618	1291	461,658	0,25	113,786	26	13,9	13,2	14,5	20,5	13,5	57766,5	527	344
Костромская область	50	920	725	648	239,29	0,07	132,834	17	13,2	12,6	13,1	20,4	13,8	30946,8	483	369
Краснодарский край	1072	7550	6386	5570	2144,989	2,03	447,032	67,6	13,6	13,0	14,2	19,7	14,9	260865,8	3649	2609
Красноярский край	437	5322	4130	2875	1059,884	0,58	11,6302	67	15,8	15,2	16,3	20,7	16,5	238822,4	1040	396
Курганская область	157	1329	1074	854	329,201	0,15	131,67	30,5	15,5	14,9	14,9	22,7	14,8	40739,4	290	154
Курская область	181	2156	1616	1122	441,416	0,27	356,73	43,2	14	13,3	14,5	20,6	15,4	54858,7	748	587

Таблица 2

Причины смерти в Российской Федерации

	Погибло	Умышл.	Тяжк.тел.	Разбойн.	Грабежи	Погибло	Суициды	Умерло от всех причин	
	в ДТП	убийства	поврежд.	нападн.		при пожаре	п. 3.4.2	ед	тыс
1992	36,47	23	53,6	30,5	164,9	10,758	46,1	1 807 441	1807,441
1993	37,2	29,2	67	40,1	184,4	13,874	56,1	2 129 339	2129,339
1994	36	32,3	67,8	37,4	148,5	15,668	61,9	2 301 366	2301,366
1995	32,4	31,7	61,5	36,3	140,6	14,973	60,9	2 203 811	2203,811
1996	29,4	29,4	53,4	34,6	121,4	15,877	57,8	2 082 249	2082,249
1997	27,7	29,3	46,1	34,3	112,1	13,832	55	2 015 779	2015,779
1998	29	29,5	45,2	38,5	122,4	13,749	51,7	1 988 744	1988,744
1999	29,7	31,1	47,7	41,1	139	14,901	57,3	2 144 316	2144,316
2000	29,6	31,8	49,8	39,4		16,298	56,9	2 225 332	2225,332
2001	30,9	33,6	55,7	44,8		18,321	57,3	2 254 896	2254,896
2002	33,2	32,3	58,5	47,1		19,988	55,7	2 332 272	2332,272
2003	35,6	31,6	57,1	48,7	198	19,303	53,1	2 365 826	2365,826
2004	34,5	31,5	57,4	55,4	251,4	18,868	49	2 295 402	2295,402
2005	34	30,8	57,9	63,7	344,4	18,412	46,6	2 303 935	2303,935
2006	32,7	27,5	51,4	59,8	357,3	17,223	46,1	2 166 703	2166,703
2007	33,3	22,2	43,3	45,3	295	16,066	41,3	2 080 445	2080,445
2008	29,9	20,1	45,4	35,4	244	15,165	38,4	2 075 954	2075,954
2009	27,6	17,7	43,1	30,1	205	13,946	37,6	2 010 543	2010,543
2010	26,6	15,6	39,7	24,5	165	13,061	33,2	2 028 516	2028,516
2011	27,9	14,3	38,5	20,1	128	12,019	30,6	1 925 720	1925,72
2012	28	13,1	37,1	18,6	110,1	11,652	29,7	1 906 335	1906,335
2013	27	12,2	34,8	16,4	92,1	10,601	28,1	1 871 809	1871,809
2014	27	11,8	32,9	14,2	76,2	10,138	26,6	1 912 347	1912,347
2015	23,1	11,4	30,2	13,4	71,7	9,405	26	1 908 541	1908,541
2016	20,3	10,3	27,4	11,4	61,5	8,76	23,1	1 891 015	1891,015
2017	19,1	9,6	24,6	9,1	56,9	7,782	20,1	1 824 340	1824,34
2018	18,2	8,5	23,2	7,5	50,1	7,913	18,2	1 817 710	1817,71
2019	17,0	7,9	21,5	6,7	45,8	8,567	17,192	1 800 677	1800,677
2020	16,2	7,6	20,0	5,3	38,4	8,313	17,192	2 124 479	2124,479

В последней таблице данные для показателя «Грабежи» с 2000 по 2002 год отсутствуют и для дальнейшего анализа могут быть интерполированы (оценены). Следует помнить, что при этом точность анализа снижается. Второй вариант – исключить показатель.

До обработки данных значения некоторых из них нормализуются, то есть приводятся к заданному диапазону (в таблице 2 показатель «Умерло от всех причин»).

5.2.2. Корреляционный анализ

Корреляционный анализ – один из основных методов математической статистики, используемый для изучения взаимосвязей между случайными величинами. Его применяют для выявления из многих показателей потенциально значимых, называемых факторами.

Корреляция (соотношение) – термин, применяемый в различных областях науки и техники для обозначения взаимозависимости, взаимного соответствия, соотношения понятий, предметов, функций.

В математической статистике – вероятностная или статистическая зависимость, не имеющая строго функционального характера. О корреляции говорят, когда изменению одной величины сопутствует изменение другой величины. Если при изменении значения одной переменной систематически меняется значение другой (увеличивается или уменьшается), и эта закономерность прослеживается на большом количестве наблюдений, то такие переменные считаются коррелирующими.

Часто рассматриваются наборы статистических данных, характеризующих субъекты в течение нескольких лет по различным показателям. Таким образом, формируются последовательности случайных событий, статистическая связь которых исследуется с помощью корреляционного анализа. Рассчитывается коэффициент корреляции, являющийся статистической мерой, которая количественно выражает силу и направление взаимосвязи между двумя переменными (последовательностями значений).

При анализе числовых данных применяют два коэффициента:

Коэффициент корреляции Пирсона является наиболее распространенным и широко используемым показателем линейной связи между двумя количественными переменными. Он применяется, когда данные имеют нормальное распределение и между переменными предполагается линейная зависимость.

Ранговый коэффициент корреляции Спирмена предназначен для анализа связи между порядковыми переменными или количественными данными, которые не подчиняются нормальному распределению. Коэффициент Спирмена – мера ранговой корреляции, он считает не значения (числа), а их порядок в выборке. Коэффициент Спирмена показывает, насколько эти ранги совпадают.

Значения коэффициентов корреляции Пирсона и Спирмена различаются незначительно, но выбор одного из них показывает корректность применения математического аппарата.

Выделяют положительную корреляцию, когда при увеличении одной переменной растет и другая, и отрицательную корреляцию, когда одна переменная растет, а другая падает.

Значения коэффициента корреляции бывают от –1 до +1. Если значение ближе к единице – связь сильная положительная, ближе к минус единице – сильная отрицательная, ближе к нулю или нулевая – слабая или отсутствует. Обычно считается, что сильная связь начинается с 0,7. Однако это не совсем так, поскольку зависит от размера выборки анализируемых данных и уровня значимости.

Оценка взаимосвязи показателей и критерия осуществляется путем сравнения соответствующих коэффициентов корреляции с критическим

(пороговым) значением, зависящим от размерности выборки данных и уровня значимости (таблицы приведены во многих справочниках, напр-мер таблица 3[20]). Критическое значение является граничным значением, выше которого гипотеза независимости факторов отвергается.

В теории вероятностей и математической статистике понятие вероятности (классическое) формализуется как числовая характеристика события, принимающая значения от 0 до 1, а уровень значимости задается в долях от единицы, наиболее часто используемый – 0,05. При необходимости более точных исследований может быть 0,01 и меньше. В научно-популярных текстах и быту допустима процентная запись, соотносящая единицу и 100%.

Таблица 3

Критические значения коэффициента ранговой корреляции Спирмена

n	Уровень значимости α			
	0,05	0,025	0,01	0,005
5	0,9			
6	0,829	0,886	0,943	
7	0,714	0,786	0,893	
8	0,643	0,738	0,833	0,881
9	0,6	0,683	0,783	0,833
10	0,564	0,648	0,745	0,794
11	0,523	0,623	0,736	0,818
12	0,497	0,591	0,703	0,78
13	0,475	0,566	0,673	0,745
14	0,457	0,545	0,646	0,716
15	0,441	0,525	0,623	0,689
16	0,425	0,507	0,601	0,666
17	0,412	0,49	0,582	0,645
18	0,399	0,476	0,564	0,625
19	0,388	0,462	0,549	0,608
20	0,377	0,45	0,534	0,591
21	0,368	0,438	0,521	0,576
22	0,359	0,428	0,508	0,562
23	0,351	0,418	0,496	0,549
24	0,343	0,409	0,485	0,537
25	0,336	0,4	0,475	0,526
26	0,329	0,392	0,465	0,515
27	0,323	0,385	0,456	0,505
28	0,317	0,377	0,448	0,496
29	0,311	0,37	0,44	0,487
30	0,305	0,364	0,432	0,478

При значении коэффициента корреляции ниже критического, показатель из рассмотрения исключается. Выбираются показатели, имеющие наибольшие значения корреляции с критерием и в дальнейшем рассматриваются в качестве факторов.

[20] Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: Физматлит, 2006. 816 с.

Коэффициент корреляции является объективным показателем, свидетельствующим о наличии или отсутствии связи между переменными, и измеряющим выраженность этой связи, но не доказывающим причинно-следственную связь (рис. 6). Наличие корреляции является основанием для продолжения исследования.

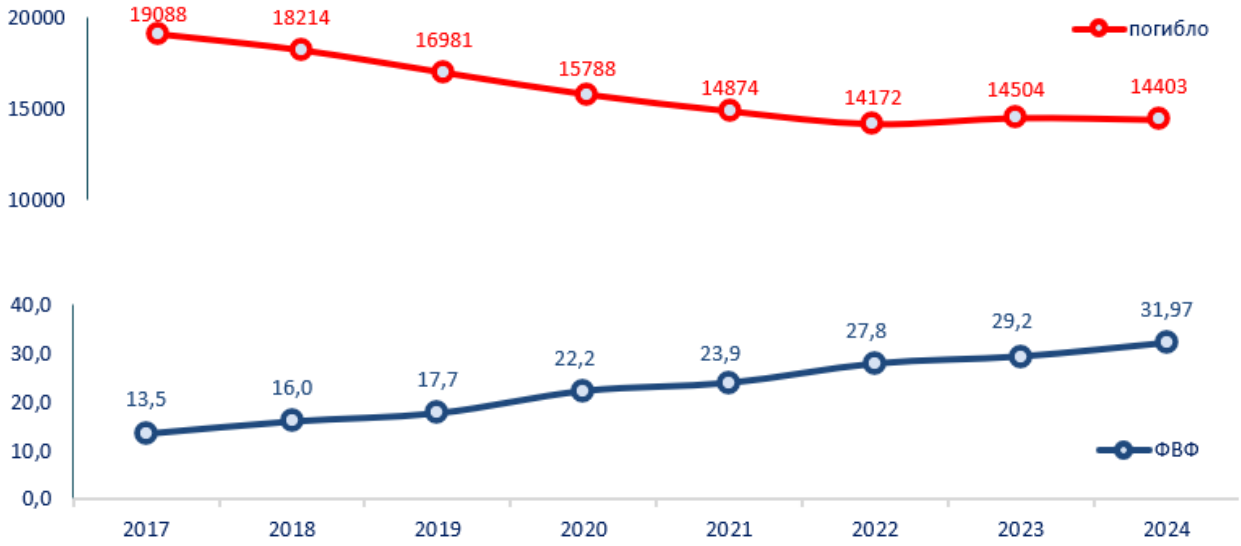


Рисунок 6. Количество комплексов фотовидеофиксации административных правонарушений и число погибших в ДТП

Увеличение количества комплексов ФВФ, возможно, способствует снижению числа погибших, но это не единственный фактор, и говорить о зависимости числа погибшихв ДТП напрямую от развития системы автоматической фиксации административных правонарушений только на основании корреляции неправильно.

Пример. Проведение корреляционного анализа и выявление потенциально значимых факторов.

Для оценки степени и характера влияния правоприменительной деятельности Госавтоинспекции на число погибших в ДТП по вине водителей в состоянии опьянения (критерий)[21] собран массив данных для 82 субъектов за 5 лет (2015-2019 годы), в который входят почти 40 различных показателей, относящихся тем или иным образом к безопасности дорожного движения.

Кроме численности населения, зарегистрированных транспортных средств, плотности автомобильных дорог и иных «очевидных» показателей, рассмотрены показатели, характеризующие сферу оборота алкогольной продукции и «Рейтинг трезвости регионов России», а также показатели правоприменительной деятельности Госавтоинспекции.

[21] Монина О.Ю., Приходько В.Е. Оценка влияния правоприменительной деятельности Госавтоинспекции и иных факторовна состояние аварийности по вине водителей с признаками опьянения: информационно-аналитический отчет. План НИОКР ФКУ «НЦ БДД МВД России». 2020. Тема 13.

По каждому сформирована последовательность из 410 случайных событий с неизвестной функцией распределения.

С помощью корреляционного анализа определяется коэффициент корреляции Спирмена, поскольку закон распределения случайных событий, а именно числа погибших и значений показателей, неизвестен.

По его величине (пороговое значение на уровне значимости 0,05 для массива данных размерности 410 равно 0,0968[22]) из перечня показателей выявляются наиболее значимые – факторы (табл. 4).

В соответствии с математической статистикой с целью формирования максимально возможного независимого набора факторов для построения регрессионной модели, чтобы более корректно (чисто) было отслежено их влияние, из уравнения, включающего все выбранные по величине корреляции факторы, удаляются малозначимые, а также из двух значительно взаимно коррелирующих факторов удаляется тот, который сильнее связан с оставшимися, особенно, если взаимная корреляция больше 0,7. Применяется метод главных компонент[23].

Таблица 4

	Погибло в ДТП	Численность населения	Природные условия	Температура воздуха	Продажа винодел. продукции	Плотность дорог	Возбуждено дел об АП водителями	Возбуждено дел об АП водителям и в состоянии опьянения	Объем продажи бензина	Возбуждено дел об АП, всего	Расход консолидир. бюджета	Отказ от мед. освидет.	По результатам автоматич. фиксации	Нарушения ПДД лицом (повторно)	Погибло по вине водителей в состоянии опьянения
	y	x1	x2	x3	x4	x5	x8	x9	x10	x11	x12	x13	x14	x15	y`
y		0,9007	0,4303	0,3002	0,7960	0,3984	0,7992	0,7520	0,7941	0,7922	0,6734	0,8224	0,6802	0,7390	0,7753
x1	0,9007		0,3726	0,2086	0,8837	0,3407	0,7992	0,7968	0,8557	0,8739	0,8175	0,8377	0,7608	0,7631	0,7219
x2	0,4303	0,3726		0,6763	0,3810	0,7242	0,8717	0,1219	0,4027	0,4675	0,0995	0,2326	0,5232	0,0677	0,3558
x3	0,3002	0,2086	0,6763		0,0719	0,7837	0,2093	-0,1270	0,2275	0,2876	-0,0633	0,1699	0,3319	-0,1489	0,0394
x4	0,7960	**0,8837**	0,3810	0,0719		0,1956	0,1066	0,8003	0,8133	0,7909	0,7914	0,7299	0,7117	0,7494	0,7216
x5	0,3984	0,3407	**0,7242**	**0,7837**	0,1956		0,7801	0,0066	0,3038	0,4433	0,0337	0,2239	0,4765	-0,0482	0,1896
x8	0,7992	**0,7992**	**0,8717**	0,2093	0,1066	**0,7801**		0,7877	0,7900	0,7516	0,7547	0,8207	0,5739	0,7409	0,6078
x9	0,7520	**0,7968**	0,1219	-0,1270	**0,8003**	0,0066	**0,7877**		0,7480	0,6499	0,7424	0,7736	0,5190	0,9210	0,7221
x10	0,7941	**0,8557**	0,4027	0,2275	**0,8133**	0,3038	**0,7900**	**0,7480**		0,7564	0,7532	0,7769	0,6531	0,7236	0,6719
x11	0,7922	**0,8739**	0,4675	0,2876	**0,7909**	0,4433	**0,7516**	0,6499	**0,7564**		0,7194	0,6893	0,9519	0,6199	0,6444
x12	0,6734	**0,8175**	0,0995	-0,0633	**0,7914**	0,0337	**0,7547**	**0,7424**	**0,7532**	**0,7194**		0,6932	0,6169	0,7233	0,5829
x13	0,8224	**0,8377**	0,2326	0,1699	**0,7299**	0,2239	**0,8207**	**0,7736**	**0,7769**	0,6893	0,6932		0,5507	0,8235	0,6771
x14	0,6802	**0,7608**	0,5232	0,3319	**0,7117**	0,4765	0,5739	0,5190	0,6531	**0,9519**	0,6169	0,5507		0,4982	0,5795
x15	0,7390	**0,7631**	0,0677	-0,1489	**0,7494**	-0,0482	**0,7409**	**0,9210**	**0,7236**	0,6199	**0,7233**	**0,8235**	0,4982		0,7512
y`	0,7753	0,7219	0,3558	0,0394	0,7216	0,1896	0,6078	0,7221	0,6719	0,6444	0,5829	0,6771	0,5795	**0,7512**	

Значения коэффициентов корреляции показателей и факторов

5.2.3. Метод главных компонент

Казалось бы, чем больше показателей (переменных), тем больше информации и тем лучше для исследователя. Однако это не всегда так. Чем больше информации, тем сложнее ее анализировать.

[22] Айвазян С.А., Бухштабер В.М., Енюков И.С. и др. Прикладная статистика. Классификация и снижение размерности / Под ред. С.А. Айвазяна. М.: Финансы и статистика, 1989. 607 с.

[23] Монина О.Ю., Приходько В.Е., Капитанов В.Т. О влиянии на показатели аварийности правоприменительных мер в отношении водителей в состоянии опьянения // Безопасность дорожного движения. 2021. № 1. С. 41–45.

Анализируя данные, иногда целесообразно создать и исследовать упрощенную модель, максимально точно описывающую явление (объект). В некоторых случаях одновременное наличие его признаков избыточно, и они сильно зависят друг от друга (коррелируют).

В качестве выхода можно несколько признаков выразить, например, через один, и работать с более простой моделью. В этом случае избежать потери информации, скорее всего не удастся, но минимизировать ее помогает метод главных компонент. Он применяется для снижения размерности решаемой задачи (минимизации набора факторов с учетом их взаимной корреляции) и получения значимых результатов.

Метод главных компонент – это математический инструмент, один из основных методов ортогонального линейного преобразования, отображающего данные из исходного пространства признаков в новое пространство меньшей размерности, с целью уменьшения размерности при минимальной потере информации.

Суть метода заключается в том, что с каждой главной компонентой связана определенная доля общей дисперсии (показатель разброса значений в выборке или совокупности относительно среднего арифметического) исходного набора данных[24]. В свою очередь, дисперсия, являющаяся мерой изменчивости данных, может отражать уровень их информативности.

Первая ось новой системы координат строится таким образом, чтобы дисперсия данных вдоль нее была максимальная. Вторая ось строится ортогонально первой так, чтобы дисперсия данных вдоль нее была бы максимальной из оставшихся возможных и т.д. Первая ось называется первой главной компонентой, вторая – второй и т.д.

Задача метода главных компонент заключается в том, чтобы построить новое пространство признаков меньшей размерности, дисперсия между осями которой будет перераспределена так, чтобы максимизировать дисперсию по каждой из них.

Для этого выполняется последовательность следующих действий:

вычисляется общая дисперсия исходного пространства признаков путем суммирования взаимных дисперсий переменных, которые определяются из ковариационной матрицы;

вычисляются собственные векторы и собственные значения ковариационной матрицы (квадратная матрица, которая количественно оценивает, как изменяются две переменные относительно друг друга), определяющие направления главных компонент и величину связанной с ними дисперсии;

производится снижение размерности – диагональные элементы ковариационной матрицы соответствуют дисперсии в исходной системе координат, а ее собственные значения – в новой.

[24] Монина М.Ю. Комплексная система оценки факторов дорожно-транспортной аварийности для принятия управленческих решений: дис. ... канд. т. наук. М., 2024. 131 с.

Разделив дисперсию, связанную с каждой главной компонентой, на сумму дисперсий по всем компонентам, получаем долю дисперсии, связанную с каждой компонентой.

После этого отбрасывается столько главных компонент, чтобы доля оставшихся составляла 80-90%, что оказывается достаточным в большинстве случаев.

В специальной литературе[25] отмечается, что использование главных компонент наиболее плодотворно в ситуациях, в которых показатели имеют общую физическую природу. Исходя из этого, показатели предлагается разбить на группы – демографические, социально-экономические, природно-климатические, административные и дорожно-транспортные и т.д. и использовать алгоритм построения главных компонент, изложенный в работе С.А. Айвазяна, В.М. Бухштабера, И.С. Енюкова, Л.Д. Мешалкина Прикладная статистика. Классификация и снижение[26].

Для выявления степени влияния факторов на критерий (число погибших в ДТП) требуется сформировать из них максимально независимый набор. С этой целью:

из двух значительно коррелирующих факторов удаляется тот, который сильнее связан с другими факторами и коэффициент корреляции которого с числом погибших существенно меньше;

исходя из общей природы (или взаимозависимости), связанные факторы объединяются на основе главных компонент.

Пример. Продолжение рассмотренного выше (оценка степени и характера влияния правоприменительной деятельности Госавтоинспекции на число погибших в ДТП по вине водителей в состоянии опьянения)[27].

В рассматриваемом примере, исходя из изложенного, рейтинговые факторы исключаются, и остается один, характеризующий экономическую составляющую – расходы консолидированного бюджета (x_{12}). Далее, учитывая его высокую корреляцию с численностью населения, этот фактор также исключается.

Таким образом, из этой группы в качестве фактора для дальнейшего рассмотрения оставлен один – объем продажи бензина (x_{10}), который косвенно характеризует пробег транспортных средств.

Из группы показателей, относящихся к «алкогольной» составляющей, выбран фактор – розничная продажа винодельческой продукции (входит в общий алкоголь), тыс. дал. (x_4), (коэффициент корреляции 0,7960; 0,7216).

Сформирована группа факторов, характеризующих показатели правоприменительной деятельности Госавтоинспекции как объект исследования:

количество возбужденных дел об административных правонарушениях в области дорожного движения (тыс.), (x_{11}), (коэффициент корреляции 0,7922; 0,6444);

---

[25] Айвазян С.А., Бухштабер В.М., Енюков И.С. и др. Прикладная статистика. Классификация и снижение размерности / Под ред. С.А. Айвазяна. М.: Финансы и статистика, 1989. 607 с.

[26] Там же.

[27] Монина О.Ю., Приходько В.Е., Капитанов В.Т. О влиянии на показатели аварийности правоприменительных мер в отношении водителей в состоянии опьянения // Безопасность дорожного движения. 2021. № 1. С. 41–45.

количество возбужденных дел об административных правонарушениях, совершенных водителями транспортных средств (тыс.), (x_8), (коэффициент корреляции 0,7992; 0,6078);

количество возбужденных дел об административных правонарушениях, совершенных водителями транспортных средств в состоянии опьянения (суммарно алкогольного и наркотического) (ед.), (x_9), (входит в x_8), (коэффициент корреляции 0,752; 0,7221);

количество возбужденных дел об административных правонарушениях, совершенных водителями, отказавшимися от медицинского освидетельствования (ед.), (x_{13}), (входит в x_8), (коэффициент корреляции 0,8224; 0,6771);

количество возбужденных дел об административных правонарушениях по результатам автоматической фиксации правонарушений[28] (тыс.), (x_{14}), (входит в x_{11}), (коэффициент корреляции 0,6802; 0,5795);

нарушение ПДД лицом, подвергнутым административному наказанию (возбуждено дел о нарушении ПДД лицом в состоянии опьянения (повторно), ст. 264.1 УК РФ) (ед), (x_{15}), (коэффициент корреляции 0,739; 0,7512).

Всего выбрано 13 наиболее значимых факторов, в том числе 6, характеризующих правоприменительную деятельность Госавтоинспекции (табл. 5).

Далее для выявления функциональной связи, при которой определенному значению показателя или их набору соответствует только одно значение выбранного критерия, проводится регрессионный анализ.

[28] Приводится формулировка в соответствии с формой «№ 1-БДД» (560) федерального статистического наблюдения «Сведения о состоянии безопасности дорожного движения».

Таблица 5

Факторы, в значительной степени влияющие на число погибших в ДТП

Переменная в уравнении	Фактор
x_1	численность населения
x_2	природные условия
x_3	температура в столице субъекта
x_4	продажа винодельческой продукции
x_5	плотность дорог
x_8	количество возбужденных дел об административных правонарушениях, совершенных водителями транспортных средств
x_9	количество возбужденных дел об административных правонарушениях, совершенных водителями транспортных средств в состоянии опьянения
x_{10}	объем продажи бензина
x_{11}	количество возбужденных дел об административных правонарушениях в области дорожного движения
x_{12}	расходы бюджета
x_{13}	количество возбужденных дел об административных правонарушениях, совершенных водителями, отказавшимися от медицинского освидетельствования
x_{14}	количество возбужденных дел об административных правонарушениях по результатам автоматической фиксации правонарушений
x_{15}	нарушение ПДД лицом, подвергнутым административному наказанию (возбуждено дел о нарушении ПДД лицом в состоянии опьянения (повторно))

5.2.4. Регрессионный анализ

Регрессионный анализ – один из базовых методов статистики, широко применяется для предсказания числовых значений на основе входных данных, а также исследования влияния одной или нескольких независимых переменных (факторов) на зависимую (критерий).

Создание регрессионной модели (уравнения) является итерационным процессом (повторяется несколько раз, чтобы постепенно улучшить результат), направленным на поиск эффективных независимых переменных (x_i), чтобы объяснить зависимые (y_i). При этом должны учитываться теоретические аспекты, экспертное мнение в этой области и здравый смысл.

После выявления с помощью корреляционного анализа связей переменных и определения степени этих связей, как правило, ставится задача нахождения их математической зависимости с использованием регрессионного анализа.

Для этого выбирается класс функций, связывающих k аргументов $x_1,...,x_k$ и функцию $y = f(x_1,...,x_k)$ от этих аргументов.

Функция $f(x_1,...,x_k)$, описывающая зависимость среднего значения величины y от значений аргументов, называется функцией (уравнением) регрессии.

В учебно-методическом пособии рассматриваются функциональные связи, которые являются либо линейными, либо могут быть сведены к линейным. При этом следует иметь в виду, что, если многомерная случайная величина $(y,x_1,...,x_k)$ подчиняется нормальному закону, то уравнение регрессии имеет линейный вид.

Предполагается, что число погибших в ДТП y зависит от n факторов, принимающих при каждом наблюдении конкретное значение. Величина y полагается случайной величиной, имеющей нормальное распределение с центром распределения в точке $\overline{y}$, являющейся математическим ожиданием в нашем случае числа погибших в ДТП и представляемой в общем виде:

$$\overline{y} = \sum_{i=1}^{n} a_i \cdot x_i + a_0, \quad (1)$$

где $a_0, a_i (i = 1,...,n)$ – параметры;

n – число факторов;

x_i – значения рассматриваемых факторов ($i = 1,...,n$).

Для каждой из k реализаций модель можно представить формулой в виде

$$y_k = \sum_{i=1}^{n} a_i \cdot x_i^k + a_0 + \delta_k, \quad (2)$$

где k – номер реализации;

y_k – значение показателя аварийности в k-ом наблюдении (реализации).

δ_k – ошибки наблюдения, взаимно некоррелированные случайные величины с нулевым математическим ожиданием и нормальным распределением (погрешность оценки, обусловленная, например, наличием неучтенных факторов в модели, так называемый свободный член уравнения; чем меньше его значение, тем точнее построена модель).

При этом предполагается случайный характер коэффициентов в общем уравнении (2), распределенных по нормальному закону (наиболее распространенный тип распределения случайных величин, играющий большую роль в теории вероятностей).

Оценки неизвестных параметров в уравнении (1) находятся методом наименьших квадратов путем минимизации суммы квадратов отклонений фактических и расчетных значений.

Пример. Продолжение рассмотренного выше (оценка влияния правоприменительной деятельности Госавтоинспекции на число погибших в ДТП по вине водителей в состоянии опьянения)[29].

В практической задаче для выявления характера связи факторов и числа погибших требуется создание модели, отражающей эту взаимосвязь. Для этого проводится регрессионный анализ и формируется модель, учитывающая функциональную связь показателя аварийности от нескольких наиболее значимых факторов с учетом степени их влияния.

Связь числа погибших по вине водителей в состоянии или с признаками опьянения (y') с рассматриваемыми факторами (x_i, приведены в таблице 5) описывается регрессионной моделью вида:

$$y` = 1{,}0992x_1 + 19{,}8356x_2 + 0{,}1413x_4 - 0{,}3676x_5 + 0{,}2557x_8 + 0{,}7928x_9 + 0{,}1024x_{10} + 0{,}5048x_{13} - 0{,}1315x_{14} + 0{,}949x_{15} - 52{,}819 .$$

С помощью разработанной модели (регрессионного уравнения) рассчитываются модельные значения критерия для субъектов Российской Федерации, и оценивается совпадение (связь) с фактическими значениями, а именно, подставляя в формулу численные значения факторов, характеризующих субъекты, получаем значение показателя для региона (некоторое среднее значение). Коэффициент ранговой корреляции расчетных и фактических (для анализируемых лет) данных – 0,774.

Графически расположение фактических и рассчитанных значений для субъектов Российской Федерации представлено на рисунке 7. Сравнивая расчетное (модельное) значение с фактическим, можно оценить ситуацию в регионе: факт ниже расчетного – благополучно; фактическое значение превышает расчетное – есть потенциал для снижения, величина которого определяется степенью отклонения.

Анализируемые факторы входят в уравнение с различными знаками (минус, плюс), то есть способствуют либо уменьшению, либо увеличению значения критерия (число погибших в ДТП по вине водителей в состоянии или с признаками опьянения), а величина коэффициента переменной (значение параметра, коэффициент регрессии) определяет, насколько изменение существенно. В уравнении со знаком минус – только плотность дорог, количество возбужденных дел об административных правонарушениях по результатам автоматической фиксации правонарушений и свободный член.

[29] Монина О.Ю., Приходько В.Е., Капитанов В.Т. О влиянии на показатели аварийности правоприменительных мер в отношении водителей в состоянии опьянения // Безопасность дорожного движения. 2021. № 1. С. 41–45.

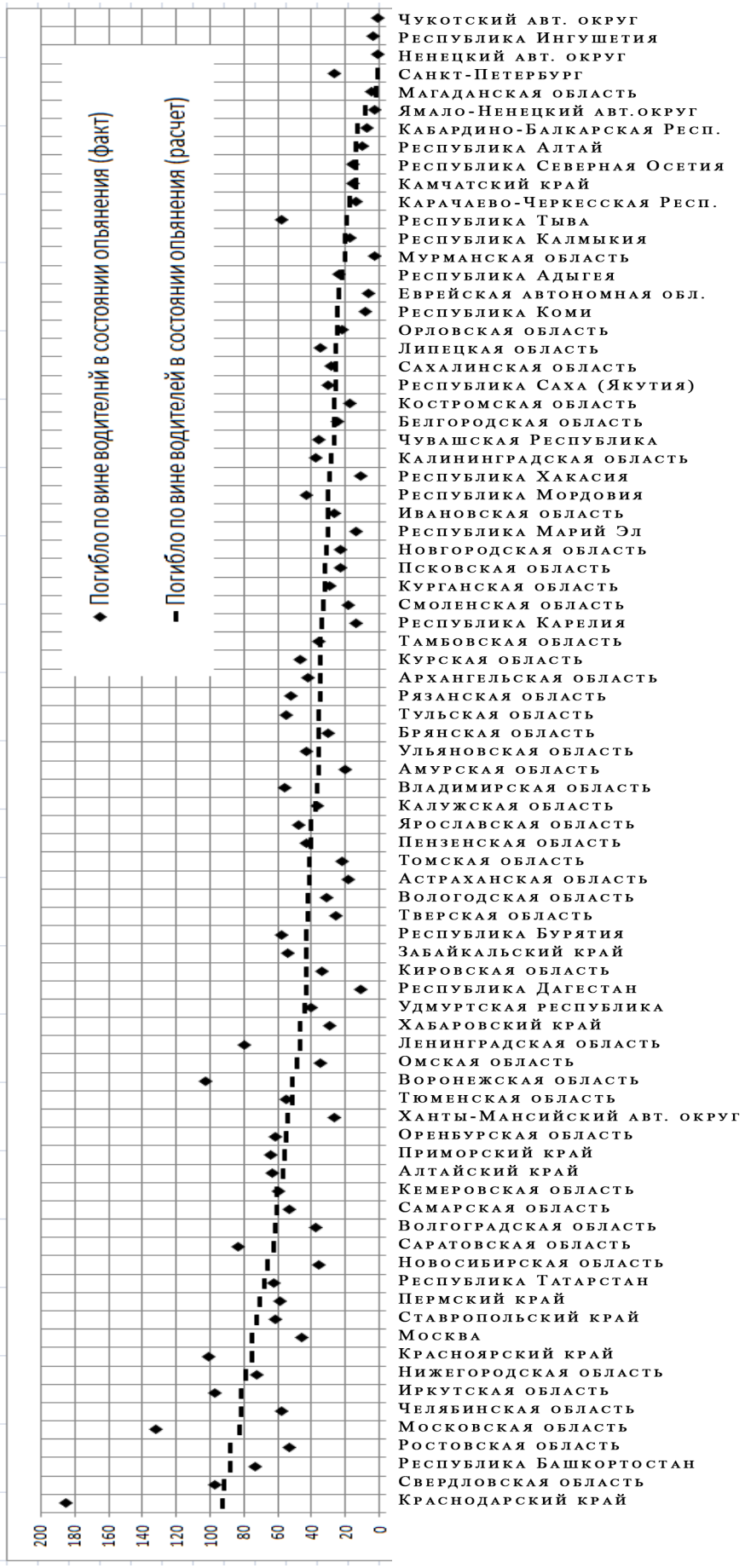


Рисунок 7. Фактические и модельные (рассчитанные по формуле (4) значения числа погибших в ДТП по вине водителей в состоянии опьянения в субъектах Российской Федерации в 2019 году

5.2.5. Оценка значимости факторов в модели

Значимость факторов в модели рассматривается в два этапа.

На первом этапе для характеристики важности фактора по сравнению с другими, оказывающими влияние на изучаемое явление (объект, критерий), рассчитывается его весовой коэффициент. Определяется сумма всех значений по каждому показателю, затем умножается на коэффициент переменной в уравнении регрессии, ему соответствующей, складываются значения с одинаковыми знаками, затем в обеих группах находится вес в процентах каждого фактора.

Пример. Чтобы оценить степень влияния факторов на число погибших в ДТП, использовано ранее полученное уравнение регрессии, связывающее число погибших в ДТП с набором факторов, выделенных из большего списка по результатам корреляционного анализа, применения метода главных компонент[30]:

$$y' = 1.3731368 \cdot y_1 - 7.9096146 \cdot x_9 + 4.54522087 \cdot x_{10} + 0.9232238 \cdot x_2$$
$$- 0.00472717 \cdot x_{11} + 0.548287544 \cdot x_{13} - 40.5744 \cdot x_3 + 0.838 \cdot x_4 - 65.6456 \cdot x_1$$
$$- 1.9224 \cdot x_5 + 63.7125 \cdot x_6 - 0.0175580478\ 6 \cdot x_7 + 31.9557.$$

Уравнение регрессии разработано на основе четырехлетнего массива данных, характеризующих 82 субъекта Российской Федерации.

С этой целью все факторы разбиты на две группы – способствующие росту аварийности и снижающие аварийность. Совокупное воздействие факторов в каждой группе принято за 100 процентов. Роль каждого оценивалась пропорционально, исходя из суммарного воздействия всех факторов. В таблице 6 приведен фрагмент результатов расчета веса показателей для регионов России (2018 год). Выбраны те, относительно которых вынесено суждение о влиянии того или иного фактора.

[30] Капитанов В.Т., Чубуков А.Б., Монина О.Ю. и др. О факторах, определяющих дорожно-транспортную аварийность в регионах России // Наука и техника в дорожной отрасли. 2019. № 4. С. 18–23.

Таблица 6

Вес показателей для регионов России в 2018 году, % (фрагмент)

Субъект Российской Федерации	Население и транспорт	Природа	Температура	Алкоголь	Качество жизни	Бензин	Плотность дорог	Возраст парка ТС	Инвестиционный риск	Инвестиционный потенциал	Уровень социально-экон. положения	Правоприменение
	y_1	x_9	x_{10}	x_2	x_4	x_6	x_{11}	x_{13}	x_3	x_1	x_5	x_7
Волгоградская обл.	81,45	3,85	7,17	1,90	1,47	4,16	0,86	11,18	2,50	43,39	19,95	22,13
Воронежская область	76,64	4,11	5,39	2,90	2,03	8,93	1,78	9,58	1,35	44,02	19,21	24,07
Еврейская АО	41,79	24,66	3,62	21,05	6,25	2,62	1,04	45,18	12,00	21,85	13,98	5,96
Камчатский край	57,06	11,30	7,78	13,99	6,14	3,74	0,05	34,82	8,09	33,06	19,69	4,29
Карачаево-Черкесская Респ.	50,95	12,10	26,75	5,32	2,66	2,20	5,12	29,72	10,64	32,14	17,99	4,40
Краснодарский край	77,39	2,04	5,05	1,13	1,01	13,38	1,38	5,47	0,73	58,60	13,46	20,36
Красноярский край	84,56	3,07	0,46	2,72	1,47	7,72	0,05	8,90	1,85	65,42	16,72	7,05
Магаданская область	45,98	17,29	23,47	26,24	10,50	0,00	0,06	28,29	6,41	22,71	17,94	1,13
Москва	87,64	1,12	0,01	0,70	0,69	8,34	0,00	0,97	0,21	65,59	4,19	29,04
Московская область	86,21	1,65	0,02	1,29	0,96	8,32	0,98	1,94	0,31	51,81	6,85	38,12
Ненецкий авт. округ	21,67	25,81	25,04	39,44	13,09	0,00	0,02	21,01	7,97	13,11	32,51	0,34
Нижегородская обл.	84,85	3,46	2,90	2,81	1,54	4,44	1,16	6,13	1,36	52,26	16,43	22,66
Республика Адыгея	46,71	11,17	26,07	7,91	3,97	4,16	6,90	25,07	5,75	31,89	23,39	7,01
Республика Алтай	51,12	18,34	7,13	17,95	5,46	0,00	0,83	42,08	12,16	20,98	16,85	7,10
Респ. Ингушетия	48,79	11,43	30,27	3,26	2,88	3,38	10,62	23,57	12,67	29,62	13,03	10,49
Республика Коми	74,28	7,78	27,64	9,47	3,30	5,18	0,11	13,37	3,89	29,12	20,80	5,07
Республика Саха	79,27	6,99	30,10	7,60	3,06	3,07	0,01	11,20	2,36	37,73	14,49	4,09
Респ. Татарстан	85,40	2,96	2,18	2,26	1,61	5,60	1,43	4,46	0,86	55,04	17,42	20,79
Республика Тыва	60,10	18,28	27,75	18,47	3,14	0,00	0,26	28,25	14,05	17,11	9,01	3,57
Рязанская область	74,18	7,47	7,47	5,60	3,00	2,27	2,07	13,48	2,61	36,03	22,09	23,72
Самарская область	83,67	3,55	3,45	2,30	1,54	5,48	1,24	5,77	1,45	45,31	16,71	29,52
Санкт-Петербург	82,69	2,11	2,95	1,43	1,33	8,48	0,01	3,15	0,61	70,90	12,87	12,47
Сахалинская область	62,55	11,29	6,98	12,74	5,06	1,38	0,23	23,79	3,57	37,38	32,76	2,27
Свердловская обл.	84,57	2,38	0,71	1,88	1,17	9,30	0,38	5,09	1,17	53,36	14,79	25,21
Челябинская область	84,41	2,94	1,53	2,51	1,39	7,21	1,09	8,34	2,17	61,63	19,43	7,35
Чукотский авт. округ	23,20	20,25	45,11	42,83	13,73	0,00	0,01	20,68	7,76	16,49	9,82	0,13
Ямало-Ненецкий АО	68,40	8,45	19,26	11,01	6,26	5,88	0,02	8,60	2,23	42,32	25,33	2,24

красным выделены показатели, способствующие росту, синим – снижению аварийности

Можно отметить, что наиболее весомый фактор, обеспечивающий снижение аварийности, – инвестиционный потенциал региона. По этому фактору выделяются Красноярский край, Москва, Санкт-Петербург, Челябинская область.

Существенная роль правоприменительной деятельности Госавтоинспекции отмечена в следующих регионах: Волгоградская, Воронежская, Московская, Нижегородская, Рязанская, Самарская, Свердловская области, Краснодарский край, г. Москва, Республика Татарстан, г. Санкт-Петербург.

Природный фактор играет значительную роль в Еврейской автономной области, Ненецком и Чукотском автономном округах. По температурному фактору выделяются Карачаево-Черкесская Республика, республики Адыгея, Ингушетия, Калмыкия, Коми, Саха (Якутия), Тыва, Чукотский и Ненецкий автономный округа. По роли алкоголя можно отметить Камчатский край, Магаданскую и Сахалинскую области, республики Алтай и Тыва, Чукотский, Ямало-Ненецкий, Ненецкий автономный округа.

Представленная в таблице весовая оценка позволяет сформировать суждение о наиболее существенных факторах, воздействующих на рассматриваемый процесс, и рассмотреть с учетом этого возможные мероприятия по снижению уровня аварийности в регионах.

Однако весовой характеристики недостаточно для того, чтобы сделать уверенный вывод о влиянии факторов на аварийность. На втором этапе проводится доверительная оценка коэффициентов в уравнении регрессии.

5.2.6. Определение доверительной вероятности

Для оценки точности статистических выводов (результатов) используют доверительную вероятность (1-a) – это вероятность того, что истинное значение оцениваемого параметра попадает в рассчитанный доверительный интервал, то есть диапазон значений, в котором с заданной вероятностью (чаще используют 90%, 95%, 99% или a=0,1; 0,05; 0,01) находится истинное значение.

В уравнении регрессии линейного вида величина y полагается случайной величиной, имеющей нормальное распределение с центром распределения в точке $\bar{y}$, являющейся математическим ожиданием в нашем случае числа погибших в ДТП и представляемой в общем виде:

где $a_0, b_i \ (i = 1,...,n)$ – параметры;

n – число факторов;

x_i – значения рассматриваемых факторов $(i = 1,...,n)$.

$$\bar{y} = \sum_{i=1}^{n} b_i \cdot x_i + a_0. \qquad (3)$$

Для каждой из k реализаций модель можно представить в виде

$$y_k + \Delta_k) = \sum_{i=1}^{n} (b_i + \alpha_i) \cdot x_i^k + (a_0 + \sigma) , \quad (4)$$

где k – номер реализации;

y_k – значение показателя аварийности в k-ом наблюдении (реализации).

$\delta, \Delta_k, \alpha_i$ – взаимно некоррелированные случайные величины с нулевым математическим ожиданием и нормальным распределением (например, погрешность оценки, обусловленная наличием в модели неучтенных факторов).

Ставится задача при заданной доверительной вероятности P_0 найти отклонение ε, обеспечивающее выполнение условия (вероятность отклонения должна быть не меньше заданной):

$$P\langle |y_k + \Delta_k - \varepsilon| \rangle \geq P_0 .$$

Оценки неизвестных параметров в уравнении (3) находятся методом наименьших квадратов путем минимизации суммы квадратов отклонений фактических и расчетных значений[31].

Адекватность регрессионной модели – это способность модели предсказывать результаты процесса в некоторой области с требуемой точностью. Цель проверки – установить, насколько хорошо модель описывает реальный процесс и насколько качественно ее можно использовать. Адекватность моделей подтверждается близкими (высокий коэффициент корреляции) фактическими и расчетными (модельными) значениями. В простых случаях адекватность устанавливают визуально: сравнивают фактические значения с результатами модельной функции в тех же точках.

Алгоритм проверки адекватности регрессионной модели состоит в последовательности действий, суть которой заключается в подстановке в модель данных, не использованных для ее построения, и проверке на совпадение с фактической величиной. Сложность заключается в том, что надо иметь такие данные.

1. Рассчитывается оценка остаточной дисперсии фактических и расчетных данных σ.

2. Проводится независимая оценка дисперсии σ_1 по данным, не использованным при построении регрессионной модели.

3. Вычисляется дисперсионное отношение:

$$F = \frac{\sigma}{\sigma_1} ,$$

[31] Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. М.: Наука, 1969. 512 с.

4. При заданном уровне значимости α и числам степеней свободы, указанных выше двух выборок по таблицам распределения Снедекора, определяется величина F_0:

$$P\{F > F_0\} = \alpha .$$

5. Если $F < F_0$, то модель считается адекватной, в противном случае адекватность отвергается[32].

Пример. Уравнение регрессии числа погибших $\bar{y}$ (критерий) и валового внутреннего продукта (показатель) имеет вид (на основе доступных данных 1992-2022 годов)[33]:

$$\overline{y_i} = -195{,}625 \cdot x_{i-2} + 36144{,}18 ,$$

где x_i – валовый внутренний продукт (в текущих ценах, в трлн руб.), соответствующий i-му году.

Наилучшее совпадение фактических и модельных значений при сдвиге показателя на 2 года (x_{i-2}). Коэффициент корреляции расчетных и фактических значений равен 0,9732, ранговый коэффициент корреляции – 0,9532. Графически результаты представлены на рисунке 8.



Рисунок 8. Расчетные (модельные) и фактические значения числа погибших в ДТП в Российской Федерации (со сдвигом на два года)

Полученная регрессионная модель дает оценку влияния на число погибших в среднем, вместе с тем не полностью описывает статистические данные, и наблюдается расхождение расчетных и фактических значений, что можно объяснить неполным набором факторов (не все выявлены), а также неточностью исходных данных, используемых для анализа. Эта

[32] Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику. М.: Изд-во ЛКИ, 2010. 600 с.

[33] Монина О.Ю. Научно обоснованный расчет прогнозных значений целевых показателей (транспортный риск, социальный риск) федерального проекта «Безопасность дорожного движения» национального проекта «Безопасные и качественные дороги» на 2025-2030 годы с учетом прогнозируемого социально-экономического развития Российской Федерации: аналитическая справка. План-график НИОКР ФКУ «НЦ БДД МВД России». Т. 10. 2024.

неточность корректируется на основе использования предположения о вероятностном (случайном) характере коэффициентов в уравнении регрессии.

Учет этого расхождения обеспечивается формулой общего вида (2), предполагающей случайный характер коэффициентов уравнения регрессии и свободного члена.

То есть нужно не только знать среднее значение, но и оценить степень концентрации данных относительно рассчитанного среднего.

Соответствующие теоретические положения приведены в работах по математической статистике[34].

В примере, имея прогнозные значения валового внутреннего продукта, можно оценить средние прогнозные значения числа погибших, а при необходимости и доверительный интервал. Исходя из прогноза роста ВВП с 2024 по 2026 год в 2%, а с 2027 по 2030 год в 3%, рассчитано среднее значение числа погибших (табл. 7).

В свою очередь, расхождение между модельным и фактическим значением объясняется случайным характером рассматриваемых величин и оценивается с помощью доверительных интервалов для коэффициентов при факторах уравнений и указывается в виде «коэффициент переменной ± доверительный интервал». Вероятность влияния фактора на критерий рассмотрена с доверительной вероятностью 0,95, а также 0,975; 0,98 и 0,99.

Доверительные оценки случайных величин находятся путем следующих расчетов[35]:

$$s^2 = \frac{\sum_i (y_i - \bar{y})^2}{n},$$

$$s_\alpha = \frac{s}{\sqrt{n-2}},$$

$$z = \sum_i (x_i - x)^2,$$

$$s_\alpha = \sqrt{\frac{1}{n-2}\frac{n \cdot s^2}{z}}, \qquad -195{,}625 \pm t \cdot s_\alpha = -195{,}382 \pm 30{,}337,$$

$\bar{x}, \bar{y}$ – математические ожидания.

Отсюда доверительный интервал для α:

где t – распределение Стьюдента с $n-2$ степенями свободы.

[34] Дубров А.М и др. Многомерные статистические методы. М., Финансы и статистика, 2011.

[35] Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. М.: Наука, 1969. 512 с.

В рассматриваемом случае t = 2,093 для вероятности 0,95.

Аналогично рассчитывается доверительный интервал для δ.

Уравнение регрессии с доверительными оценками числа погибших и валового внутреннего продукта:

$$y_i \pm \Delta_i = -(195{,}625 \pm 22{,}196) \cdot (x_{i-2} - 53{,}7) + (25638 \pm 762{,}05)\,.$$

Доверительная оценка для числа погибших в ДТП y_0 рассчитывается по формуле:

$$y_0 \pm \frac{t \cdot s}{n-2} \sqrt{\frac{n^2 \cdot (x_0 - \bar{x})^2}{n \cdot \sum_i x_i^2 - (\sum_i x_i)^2} + 1}\,, \qquad (5)$$

x_0, y_0 – соответственно значения валового внутреннего продукта и числа погибших в ДТП.

С вероятностью 0,95 прогнозируемое исходя из валового внутреннего продукта среднее значение числа погибших в ДТП рассчитывается по формуле (5) и находится в интервале от минимального до максимального значения (табл. 7, рис. 9).

Таблица 7

Среднее значение, доверительное отклонение от среднего значения, интервалы прогнозируемого значения числа погибших в ДТП в Российской Федерации

Год	Среднее значение	Доверительное отклонение с веростностью 0,95	Минимальное значение	Максимальное значение
2023	14702	1456	13246	16158
2024	13408	1583	11825	14991
2025	12589	1665	10924	14254
2026	12118	1713	10405	13831
2027	11637	1762	9875	13399
2028	11148	1812	9335	12960
2029	10648	1864	8784	12511
2030	9883	1943	7939	11826

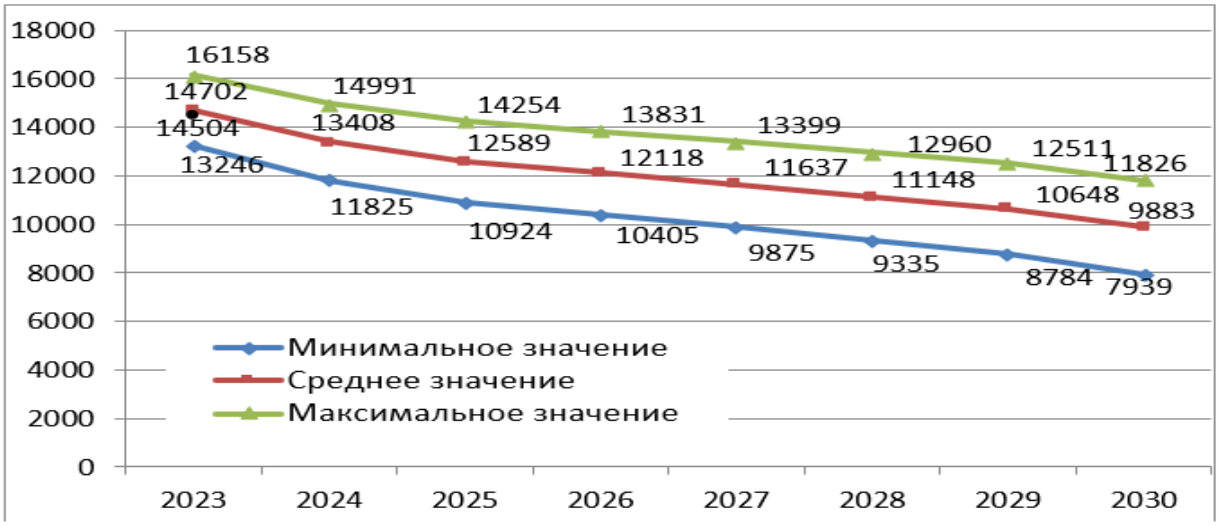


Рисунок 9. Среднее, интервалы прогнозируемого (с вероятностью 0,95) и фактическое (2023 год) значения числа погибших в ДТП в Российской Федерации

В указанный коридор от минимального до максимального значения попадет 95% значений. Для иных вероятностей коридор будет шире, то есть разброс значений больше. На графике видно, что фактическое значение числа погибших в ДТП в 2023 году практически совпадает с прогнозным значением, полученным с помощью разработанной математической модели.

6. Прогнозирование дорожно-транспортной аварийности

Под прогнозированием понимается процесс предсказания тенденций и перспектив развития состояния аварийности на основе знания истории и выявленных закономерностей его развития в прошлом. Это состояние чаще всего оценивается числом погибших в ДТП, но могут использоваться и другие критерии.

Выделяют методы прогнозирования, базирующиеся на экспертных оценках (имеющемся опыте, знаниях), экстраполяции данных, корреляционном и регрессионном анализе, математическом моделировании[36].

В сфере безопасности дорожного движения существенное значение имеет выявление закономерностей, в частности, трендов. Наряду с общепринятыми и уже программно-реализованными методами трендового анализа существуют и более сложные.

В зависимости от того, на сколько шагов j вперед осуществляется прогнозирование, оно подразделяется на краткосрочное $j \le 3$, среднесрочное $3 < j \le 6$ и долгосрочное $j > 6$.

Обычно статистический анализ используется для построения прогноза для первых двух видов прогнозирования. Для долгосрочного

[36] Капитанов В.Т., Монина О.Ю., Сильянов В.В., Чубуков А.Б. Математические методы и модели в сфере обеспечения безопасности дорожного движения: монография. М.: Техполиграфцентр, 2020, 102 с.

прогноза требуются не только статистические экстраполяционные методы, но, например, экспертные оценки характера протекания процесса на много шагов вперед.

Метод экстраполяции – это один из способов прогноза, который основывается на оценке будущих событий, учитывая анализ показателей, которые имели место в прошлые годы. То есть, если подобрать функцию, максимально точно описывающую последовательность событий в прошлом, допускается предположение, что и в будущем процесс будет развиваться соответствующим образом[37].

При этом чем больше прошлый период (обычно, не менее 5-8 лет), тем точнее можно определить развитие процесса. Однако, исходя из теории и практики, долгосрочные прогнозы, как правило, не сбываются. В лучшем случае с той или иной вероятностью можно предсказать развитие событий на 2-3 года вперед. А затем модель (функцию) надо корректировать.

Для получения достоверного прогноза необходимо наличие статистических данных и модели, удовлетворяющей ряду требований: она не должна меняться с изменением временного масштаба; должен учитываться временной фактор, то есть данные, относящиеся к намного более поздним моментам времени по сравнению с рассматриваемыми моментами, практически не должны влиять на результат.

В исследовании дорожно-транспортной аварийности существенное значение имеет выявление закономерностей, в частности, трендов. Под трендом данных об аварийности понимаются основные тенденции изменения временного ряда или направление преимущественного движения показателей.

Результат применения критерия инверсий выявил появление тренда в данных об аварийности в России последние годы[38].

То есть в последовательности случайных событий (число погибших в ДТП по годам) есть некая закономерность, что позволяет воспользоваться экстраполяционной моделью (первый подход) для прогнозирования значений показателей аварийности.

Для задачи прогнозирования значения $\overline{x_i}$ числа погибших в ДТП в момент времени предлагается использовать следующую модель[39]:

$$x_i = f \cdot x_{i-k} + f_2 \cdot \overline{x_{i-k-1}}, (i = k+2,...,n),$$

где f, f_2 – параметры[40];

x_i – фактическое значение наблюдаемой величины (числа погибших в ДТП) в момент времени i;

x_i – расчетное значение наблюдаемой величины в момент времени;

[37] Капитанов В.Т., Монина О.Ю., Сильянов В.В., Чубуков А.Б. Математические методы и модели в сфере обеспечения безопасности дорожного движения: монография. М.: Техполиграфцентр, 2020, 102 с.

[38] Капитанов В.Т., Монина О.Ю., Чубуков А.Б. Модельная оценка взаимосвязи различных показателей, связанных с гибелью людей // Наука и техника в дорожной отрасли. 2016. № 2.

[39] Редкозубов С.А. Статистические методы прогнозирования в АСУ. М., Энергоиздат, 1981.

[40] Параметр – это заданный буквой коэффициент в уравнении или неравенстве, который может принимать некоторые числовые значения. В зависимости от определенных значений параметра могут изменяться решения уравнения или неравенства, а также их количество.

n – число наблюдений;
k – глубина прогноза;
i – момент времени.

Каждому значению k соответствует свое значение f_k, f_{2k}.
Прогнозное значение рассматриваемой величины в момент времени ($n+k$) находится по формуле (прогноз на глубину k).

При $f_{2k} = 1 - f_k$ $x_{k+n} = \alpha_k \cdot x_n + (1-\alpha_k)x_{n-1}$.

Выше обозначено $\alpha_k = f_k$.

В частном случае при прогнозировании на один год формула имеет вид:

$$x_i = f \cdot x_{i-1} + f_2 \cdot x_{i-2} \qquad (i = 3,\ldots,n).$$

Параметры f, f_2 в формуле находятся путем решения оптимизационной задачи

$$\sum_{i=k+2}^{n} (\overline{x_i} - x_i)^2 \underset{f,f_2}{\rightarrow} \min .$$

Для этой цели использован метод Нелдера и Мида (подробно описан в работе Д. Химмельблау Прикладное нелинейное программирование[41]), ориентированный на поиск оптимума целевой функции произвольного вида.

Пример. Для получения прогнозного значения погибших в 2018 и 2019 годах[42] были использованы значения числа погибших в ДТП с 2000 по 2017 год и программа оптимизации, с помощью которых рассчитаны значения параметров f и f_2 и получен конкретизированный вариант модели ($\overline{x_i} = f \cdot x_{i-1} + f_2 \cdot x_{i-2}$) при $f_2 = 1 - f$ следующего вида:

$$x_i = 1{,}30849 \cdot x_{i-1} - 0{,}30849 \cdot x_{i-2} - 0{,}38957,$$

где:

$\overline{x_i}$ – прогнозируемое значение;
x_{i-1} – предыдущее значение (значение на год раньше);
x_{i-2} – значение на 2 шага назад (на два года раньше).

В результате в 2018 году ожидалось 18 218 (фактическое значение 18 214) погибших, а в 2019 году – 17 555 (фактическое значение 16 891, погрешность – 3,38%).

[41] Химмельблау Д. Прикладное нелинейное программирование. М.: Мир, 1975. 536 с.

[42] Капитанов В.Т., Сильянов В.В., Чубуков А.Б., Монина О.Ю. О прогнозе дорожно-транспортной аварийности в Российской Федерации в 2019 году // Наука и техника в дорожной отрасли. 2019. № 3.

Для определения значения показателя в 2020 году использованы данные с 2012 по 2019 год и программа оптимизации, получен вариант модели:

$(i = 3,...,n)$,

где

$\overline{x_i}$ – прогнозируемое значение в момент t;
x_{i-1} – фактическое значение годом раньше;
x_{i-2} – фактическое значение двумя годами раньше.

Расчетные и фактические данные для восьмилетнего периода представлены на рисунке 10, где прогнозируемое для 2020 года значение – 15 558 человек (предполагалось дальнейшее снижение числа погибших в ДТП на 8,38 %). Фактическое значение – 15 788 (отмечается расхождение в 1,48%).

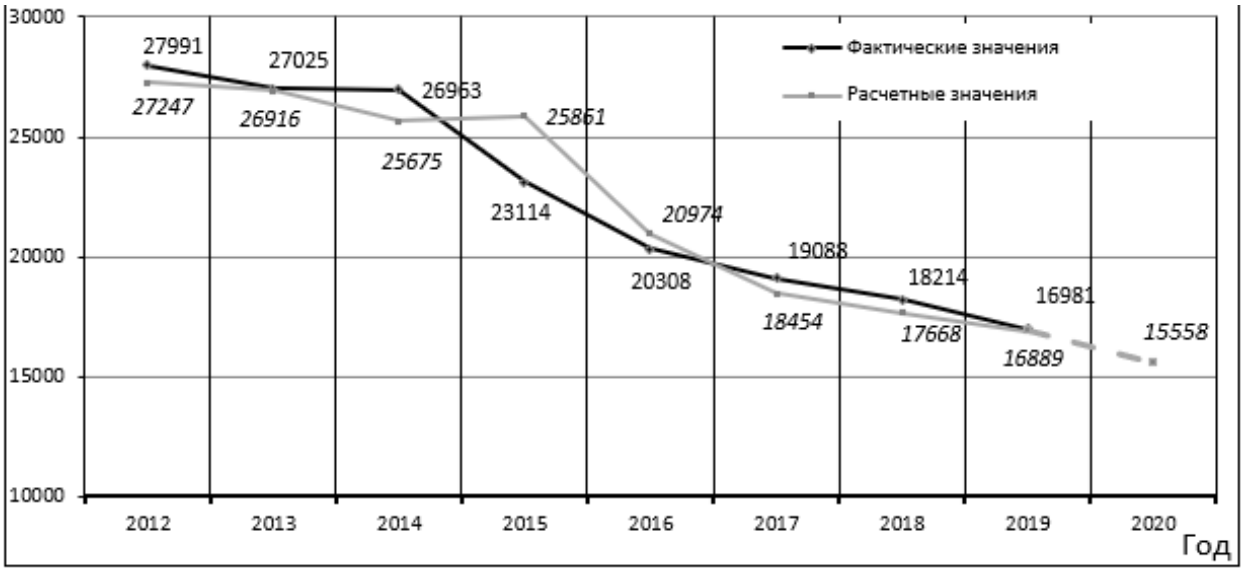


Рисунок 10. Фактические и расчетные значения числа погибших в ДТП в Российской Федерации с прогнозом на 2020 год

С помощью вышеизложенной теории оценен диапазон возможного предельного изменения рассматриваемого показателя сверху и снизу, то есть лучший и худший варианты:

ограничения снизу рассчитываются по модели:

$$\overline{x_i} = 0{,}72459 \cdot x_{i-1} + 0{,}27541 \cdot x_{i-2} - 3{,}86606,$$

аналогичная формула для ограничения сверху:

$$\overline{x_i} = 1{,}31611 \cdot x_{i-1} - 0{,}31611 \cdot x_{i-2} + 0{,}24335 .$$

На рисунке 11 представлены фактические значения числа погибших в ДТП и кривые, соответствующие возможным максимальным и минимальным значениям. Диапазон предельного изменения числа погибших в ДТП в 2020 году возможен от 13 455 до 16 835. Это означает, что, исходя из истории, вряд ли без каких-то особых воздействий на процесс удастся выйти из этого диапазона.

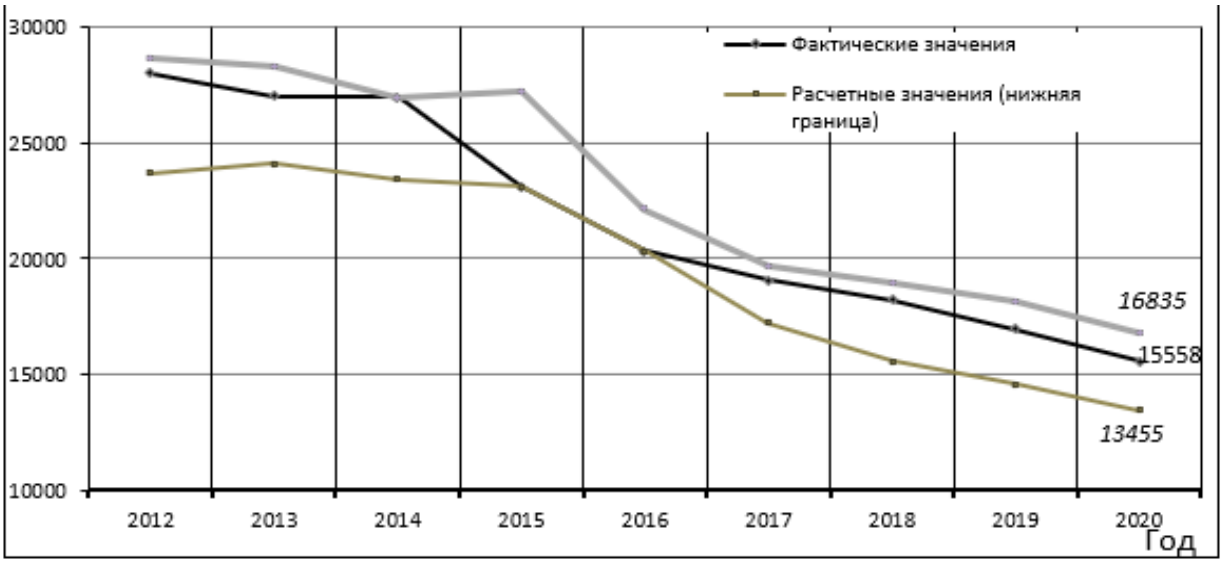


Рисунок 11. Фактические значения и диапазон предельного изменения числа погибших в ДТП в Российской Федерации с прогнозом на 2020 год

Второй подход для прогноза основывается на знании зависимости значения критерия от показателя, и предполагается, что в будущем эта зависимость не изменится, и, имея прогнозные значения показателя, можно получить предполагаемое значение критерия (пример в п. 5.2.6.).

Сложность реализации этого подхода заключается в выявлении показателя с достоверными прогнозными значениями (долгое время можно было опираться на прогнозные значения уровня инфляции, характеризующие изменения в экономической сфере).

7. Интерпретация полученных результатов и формирование выводов

Интерпретация результатов и формирование выводов – завершающий и ответственный этап статистического исследования, от правильности выполнения которого зависит практическая ценность всей проделанной работы.

Этот процесс требует не только статистической грамотности, но и глубокого понимания предметной области исследования. Самое важное – не только суметь посчитать, выполнить математические действия, но и понять, что получилось, и объяснить, что означают полученные значения, является ли решение оптимальным (наилучшим).

На этапе интерпретации результаты анализируются строго в контексте поставленных задач. Важно избегать подмены корреляции причинностью;

игнорирования сторонних факторов, влияющих на результат; преувеличения ценности значимых статистически, но незначимых практически результатов; представления только удобных результатов без учета противоречивых. Возможно обращение за экспертной оценкой, привлечение специалистов предметной области.

При формировании выводов выделяются ключевые значения (закономерности), осуществляется группировка по значимости (различая статистическую и практическую значимость), проводится сравнение с теоретическими ожиданиями, отмечаются неожиданные результаты. Использование визуализации помогает сделать выводы более наглядными.

В результате приходим к пониманию того, что сделано именно то, что нужно, и целесообразно остановиться или продолжить исследования (расчеты).

Правильная интерпретация результатов и формирование выводов требуют системного подхода, критического мышления и глубокого понимания как статистических методов, так и предмета исследования. Качественно выполненные выводы превращают статистические данные в практически полезное знание, служащее основой для принятия обоснованных решений.

8. Программные способы обработки статистических данных

Чтобы извлечь из данных ценные выводы, необходимо уметь их обрабатывать. Здесь на помощь приходят программные инструменты, которые позволяют автоматизировать рутинные задачи, провести сложные вычисления и визуализировать результаты. В этом разделе приведены основные программные способы обработки статистических данных, отмечены их особенности и область применения.

Раньше статистика была уделом математиков и исследователей, которые вручную проводили расчеты на бумаге или с помощью калькуляторов. Программные способы обработки данных важны, потому что сегодня объемы данных растут экспоненциально, и ручная обработка становится невозможной.

Программные инструменты позволяют:

- обрабатывать большие массивы данных (BigData);
- автоматизировать рутинные задачи (например, очистку данных);
- проводить сложные анализы (например, регрессионный анализ или кластеризацию);
- визуализировать результаты для лучшего понимания;
- интегрировать (объединять) данные из разных источников.

Основные программные инструменты для обработки данных:

Электронные таблицы (Excel, GoogleSheets) – это первый шаг в мир обработки данных. Они просты в использовании и подходят для базовых задач:

- фильтрация и сортировка данных;
- построение графиков и диаграмм;
- проведение простых статистических расчетов (среднее, медиана, стандартное отклонение).

Однако для сложных анализов и больших объемов данных электронные таблицы могут быть недостаточно мощными.

Статистические пакеты (SPSS, SAS, Stata) – программы, созданные специально для статистического анализа. Они предлагают готовые методы:

- для проведения регрессионного анализа;
- для тестирования гипотез;
- для работы с временными рядами;
- для кластеризации и классификации данных.

Основное преимущество таких пакетов – удобный интерфейс и обширная библиотека статистических методов. Однако они часто требуют лицензий и могут быть дорогими.

Языки программирования (Python, R) – это продвинутый инструмент для обработки данных. Они гибкие, бесплатные и поддерживают огромное количество библиотек.

Python широко используется благодаря своей простоте и универсальности. Библиотеки Pandas, NumPy и SciPy позволяют обрабатывать данные, а Matplotlib и Seaborn – визуализировать их. Scikit-learn используется для машинного обучения.

R создан специально для статистического анализа. Он предлагает богатый набор функций для работы с данными и визуализации. Пакеты dplyr, ggplot2 и caret активно используются аналитиками.

Инструменты бизнес-аналитики (Power BI, Tableau) – предназначены для визуализации и анализа данных и позволяют:

подключаться к различным источникам данных (базы данных, облачные хранилища);

создавать интерактивные дашборды (дашборд – инструмент для визуализации данных, который собирает, структурирует и анализирует информацию в графиках, диаграммах или таблицах);

проводить простой анализ данных (например, вычисление агрегированных значений (объединенных из различных источников).

Такие инструменты особенно полезны для презентации результатов руководителям и заинтересованным сторонам.

Для работы с большими объемами данных требуются системы управления базами данных и специализированные платформы для работы с BigData (PosgreSQL, Hadoop, Spark):

PosgreSQL используется для работы с реляционными базами данных. Он позволяет извлекать, фильтровать и агрегировать данные.

Hadoop и Spark предназначены для распределенной обработки данных. Они используются в задачах машинного обучения и анализа больших массивов данных.

Программные инструменты помогают на каждом этапе работы с данными:

Сбор данных из разных источников: базы данных, электронные таблицы и др. (Application public interface (API), веб-скрапинг, CSV).

Часто данные содержат ошибки, пропуски или дубликаты. Программные инструменты помогают автоматизировать процесс очистки данных. Например, библиотека Pandas в Python позволяет удалять дубликаты, заполнять пропуски и преобразовывать типы данных.

Анализ данных включает применение статистических методов: вычисление средних значений, дисперсии, корреляции, регрессионный анализ и т.д. Программные инструменты позволяют проводить сложные расчеты за считанные секунды.

Визуализация данных, представление в виде графиков и диаграмм помогают понять структуру данных и сделать выводы. Например, библиотека Matplotlib в Python позволяет создавать графики, гистограммы и диаграммы рассеяния.

Интерпретация и прогнозирование. На основе анализа данных можно строить прогнозы и принимать решения. Например, библиотека Scikit-learn позволяет строить модели машинного обучения для прогнозирования.

Выбор программного инструмента зависит от задач и уровня подготовки сотрудника:

- для базового анализа подойдут электронные таблицы или Tableau;
- для сложного статистического анализа – R или SPSS;
- для работы с большими данными и машинным обучением – Python, SQL и Spark;
- для визуализации и презентации – Power BI или Tableau.

С развитием технологий обработка данных становится все более автоматизированной. Появляются новые инструменты, которые позволяют автоматически строить модели машинного обучения. Растет популярность облачных платформ (AWS, GoogleCloud, Azure), которые предлагают готовые решения для работы с данными.

Можно воспользоваться российским пакетом офисных приложений для работы с документами «МойОфис», который состоит в том числе из приложений для работы с текстом и электронными таблицами.

Программные способы обработки статистических данных дают возможности для анализа и принятия решений. Независимо от квалификации важно выбрать подходящие инструменты и развивать свои навыки.

Заключение

Статистический анализ помогает принимать обоснованные решения на основе данных. Важно правильно выбирать методы, проверять гипотезы и визуализировать результаты. Статистический анализ дорожно-транспортной аварийности – важный инструмент для снижения значений ее показателей.

Предложенный математический аппарат является классическим (стандартным) для анализа статистических данных в различных сферах. В учебно-методическом пособии изложены теоретические основы статистических методов, необходимые для понимания действий обработки массивов данных дорожно-транспортной аварийности и значений показателей, характеризующих сферу безопасности дорожного движения, приведены примеры расчетов и способы интерпретации полученных результатов для обоснования выбора мероприятий, нацеленных на повышение безопасности дорожного движения.

Набор предложенных методов для этой сферы и конкретных задач разработан и апробирован при проведении научных исследований в Научном центре БДД МВД России, сформирован в методику.

Это издание не является справочником и не содержит исчерпывающую информацию или пошаговые инструкции обработки статистических данных, но дает представление о решении практических задач анализа, оценивания и прогнозирования дорожно-транспортной аварийности на уровне субъектов и страны в целом.

Даны определения понятиям, используемым в работе, разъяснена суть применяемых вероятностно-статистических методов применительно к анализу данных в сфере безопасности дорожного движения.

Полагается, что изложенный материал позволяет расширить кругозор, способствует осознанности действий и дает представление о сути математических методов как инструмента для применения в обозначенной сфере деятельности.

Можно посчитать самостоятельно или обратиться к специалисту, но очень важно оценить полученные результаты расчетов, понять и смочь объяснить, а не принимать просто на веру.

Хочется надеяться, что стали понятнее сложные методы и, возможно, появился интерес к дальнейшему самостоятельному изучению возможностей статистического анализа для поиска решений, способствующих снижению дорожно-транспортной аварийности. Те, кого заинтересовала математическая составляющая, могут обратиться к учебникам и пособиям или использовать пакеты прикладных программ.

Список использованных источников

1. Федеральный закон от 10 декабря 1995 г. № 196-ФЗ «О безопасности дорожного движения» // СПС «Консультант Плюс».

2. Федеральный закон от 29 ноября 2007 г. № 282-ФЗ «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» // СПС «Консультант Плюс».

3. Постановление Правительства РФ от 19 сентября 2020 г. № 1502 «Об утверждении Правил учета дорожно-транспортных происшествий, об изменении и признании утратившими силу некоторых актов Правительства Российской Федерации» // Собрание законодательства РФ. 28.09.2020. № 39. Ст. 6077.

4. Приказ МВД России от 13 ноября 2020 г. № 767 «Об утверждении Концепции научного обеспечения деятельности органов внутренних дел Российской Федерации на период до 2030 года» // СПС «Консультант Плюс».

5. Приказ МВД России от 12 августа 2022 г. № 598 «Об организации учета, сбора и анализа сведений о дорожно-транспортных происшествиях и признании утратившим силу нормативного правового акта МВД России» (вместе с «Порядком организации учета, сбора и анализа сведений о дорожно-транспортных происшествиях, контроля за полнотой и достоверностью этих сведений») // СПС «Консультант Плюс».

6. Айвазян С.А., Бухштабер В.М., Енюков И.С. и др. Прикладная статистика. Классификация и снижение размерности / Под ред. С.А. Айвазяна. М.: Финансы и статистика, 1989. 607 с.

7. Баканов К.С., Ляхов П.В. и др. Профили безопасности дорожного движения субъектов Российской Федерации 2024: статистический сборник. М.: ФКУ «Научный центр БДД МВД России», 2025. 104 с.

8. Баканов К.С., Ляхов П.В., Айсанов А.С. и др. Дорожно-транспортная аварийность в Российской Федерации в 2024 году: информационно-аналитический обзор. М.: ФКУ «НЦ БДД МВД России», 2025, 148 с.

9. Баканов К.С., Наумов С.Б., Ляхов П.В. и др. Анализ дорожно-транспортной аварийности: методические рекомендации. М.: «ФКУ НЦ БДД МВД России», 2024. 236 с.

10. Дубров А.М. и др. Многомерные статистические методы. М., Финансы и статистика, 2011.

11. Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику. М.: Изд-во ЛКИ, 2010. 600 с.

12. Капитанов В.Т., Монина О.Ю., Сильянов В.В., Чубуков А.Б. Математические методы и модели в сфере обеспечения безопасности дорожного движения: монография. М.: Техполиграфцентр, 2020. 102 с.

13. Капитанов В.Т., Сильянов В.В., Чубуков А.Б., Монина О.Ю. О прогнозе дорожно-транспортной аварийности в Российской Федерации в 2019 году // Наука и техника в дорожной отрасли. 2019. № 3.

14. Капитанов В.Т., Монина О.Ю., Чубуков А.Б. Прикладные математические методы для анализа аварийности: методические рекомендации. М.: НИЦ БДД МВД России, 2014. 97 с.

15. Капитанов В.Т., Монина О.Ю., Чубуков А.Б. Модельная оценка взаимосвязи различных показателей, связанных с гибелью людей // Наука и техника в дорожной отрасли. 2016. № 2.

16. Капитанов В.Т., Чубуков А.Б., Монина О.Ю. и др. О факторах, определяющих дорожно-транспортную аварийность в регионах России // Наука и техника в дорожной отрасли. 2019. № 4. С. 18–23.

17. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: Физматлит, 2006. 816 с.

18. Маркс К., Энгельс Ф. Людвиг Фейербах и конец классической немецкой философии. Сочинения. Изд. 2-е. М.: Политиздат, 1961. Т. 21. С. 306.

19. Монина М.Ю. Комплексная система оценки факторов дорожно-транспортной аварийности для принятия управленческих решений: дис. ... канд. т. наук. М., 2024. 131 с.

20. Монина О.Ю. Научно обоснованный расчет прогнозных значений целевых показателей (транспортный риск, социальный риск) федерального проекта «Безопасность дорожного движения» национального проекта «Безопасные и качественные дороги» на 2025-2030 годы с учетом прогнозируемого социально-экономического развития Российской Федерации: аналитическая справка. План-график НИОКР ФКУ «НЦ БДД МВД России». Т. 10. 2024.

21. Монина О.Ю., Приходько В.Е. Оценка влияния правоприменительной деятельности Госавтоинспекции и иных факторов на состояние аварийности по вине водителей с признаками опьянения: информационно-аналитический отчет. План НИОКР ФКУ «НЦ БДД МВД России». Т. 13. 2020.

22. Монина О.Ю., Приходько В.Е., Капитанов В.Т. О влиянии на показатели аварийности правоприменительных мер в отношении водителей в состоянии опьянения // Безопасность дорожного движения. 2021. № 1. С. 41–45.

23. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. М.: Наука, 1969. 512 с.

24. Редкозубов С.А. Статистические методы прогнозирования в АСУ. М., Энергоиздат, 1981.

25. Химмельблау Д. Прикладное нелинейное программирование. М.: Мир, 1975. 536 с.

План-график выпуска научных изданий
ФКУ «НЦ БДД МВД России»
на 2025 г. поз. 33

Ольга Юрьевна Монина

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА, ОЦЕНКИ И ПРОГНОЗИРОВАНИЯ ДОРОЖНО-ТРАНСПОРТНОЙ АВАРИЙНОСТИ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

Редактор Е.В. Марцога
Компьютерная верстка Д.В. Самуйлов

Федеральное казенное учреждение
«Научный центр безопасности дорожного движения
Министерства внутренних дел Российской Федерации»
121293, г. Москва, ул. Поклонная, д. 17.
Тел. (495) 184-32-23.
Официальный сайт: https://нцбдд.мвд.рф

Подписано в печать 25.12.2025
Формат 60x84/16. Тираж 10 экз.

Отпечатано в ФКУ «НЦ БДД МВД России»
125195, г. Москва, Ленинградское шоссе, д. 59

Для заметок