

Научная статья  
https://doi.org/10.23672/SAE.2024.4.4.043  
УДК 311.2



## РЕАЛИЗАЦИЯ МЕТОДОВ СТАТИСТИЧЕСКОГО АНАЛИЗА С ИСПОЛЬЗОВАНИЕМ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

**Чикин С.Н.**

*ФГБОУ ВО «Новосибирский государственный университет экономики и управления»*

**Аннотация.** В связи с активным развитием цифровизации, использование информационных технологий в статистическом анализе приобретает особую актуальность. Автор выделяет основные преимущества использования информационных технологий в статистическом анализе, такие как повышение скорости обработки данных, улучшенная точность результатов и возможность автоматизации процессов анализа.

**Цель:** практическое применение информационных технологий при проведении статистических исследований.

**Задачи:** исследование методов и инструментов, которые позволяют современным исследователям эффективно проводить статистический анализ информации, а также практическое применение статистических методов на примере различных информационных технологий, таких как программное обеспечение для анализа данных, базы данных и средства визуализации результатов.

**Методы:** научная диалектика, анализ и синтез, системный подход, систематизация научных исследований по реализации статистических методов анализа с использованием IT инструментов.

**Результаты.** Обоснована необходимость успешной реализации методов статистического анализа с применением информационных технологий в различных областях, что подтверждает их значимость и перспективность.

**Ключевые слова:** случайные события, факторный и результативный параметры, корреляционный, регрессионный и дисперсионный анализ данных, статистика.

## IMPLEMENTATION OF STATISTICAL ANALYSIS METHODS USING INFORMATION TECHNOLOGIES

**Sergey N. Chikin**

*Novosibirsk State University of Economics and Management*

**Abstract.** In connection with the active development of digitalization, the use of information technologies in statistical analysis is becoming particularly relevant. The author highlights the main advantages of using information technology in statistical analysis, such as increased data processing speed, improved accuracy of results and the ability to automate analysis processes.

**Object:** practical application of information technology in statistical research.

**Research object:** research of methods and tools that allow modern researchers to effectively conduct statistical analysis of information, as well as the practical application of statistical methods using the example of various information technologies, such as data analysis software, databases and results visualization tools.

**Methods:** scientific dialectics, analysis and synthesis, systems approach, systematization of scientific research on the implementation of statistical methods of analysis using IT tools.

**Findings:** The need for successful implementation of statistical analysis methods using information technologies in various fields is substantiated, which confirms their significance and promise.

**Key words:** random events, factor and result parameters, correlation, regression and variance analysis of data, statistics.

### **Введение.**

Актуальность применения информационных технологий для статистического анализа данных обусловлена, в первую очередь, растущим объемом информации, которая поступает для обработки из различных источников. Современное

общество переживает этап цифровизации всех сфер жизнедеятельности, в результате которого появляется огромное количество структурированных и неструктурированных данных, требующих анализа. Эффективная обработка такого массива

информации без использования ИТ-инструментов становится практически невозможной. Кроме того, быстрота и точность анализа с использованием информационных технологий значительно превосходят традиционные методы обработки данных [4].

Второй аспект актуальности связан с необходимостью принятия обоснованных управленческих, экономических и социальных решений. В условиях высокой конкуренции и быстро меняющейся экономической ситуации, важность принятия правильных решений, основанных на анализе больших данных возрастает. Информационные технологии позволяют не только быстро собирать и анализировать данные, но и представлять их в удобном для восприятия виде, что является ключевым фактором в процессе принятия решений на всех уровнях управления [1].

В профессиональной деятельности, а также в научных исследованиях люди часто сталкиваются с событиями и явлениями с непредсказуемым исходом. При этом приходится принимать важные решения на их основе. В быту такие решения принимаются на основе здравого смысла, интуиции, предыдущего опыта. Однако в практической сфере решения принимаются на основе детального анализа имеющейся информации.

### Результаты.

Разработанные математической статистической методы анализа данных, связанные с влиянием случайных воздействий, позволяют выявлять закономерности на фоне случайности, делать обоснованные выводы и прогнозы, а также оценивать вероятность их выполнения.

В данной статье рассматривается реализация данных методов с использованием информационных технологий.

Стохастические зависимости отличаются непредсказуемостью, многие объекты имеют параметры, характеризующиеся количественными показателями. Некоторые из показателей влияют друг на друга. Такие показатели называют факторными. Каждой величине факторного показателя или аргумента может соответствовать некоторое значение результативного показателя или функции [6].

При анализе факторных показателей, мы часто сталкиваемся с понятием корреляции. Корреляция позволяет нам оценить степень взаимосвязи между двумя количественными переменными. Сильная корреляция указывает на то, что изменение одного показателя склонно сопровождаться изменением другого в определённом

направлении. Однако важно помнить, что корреляция не говорит нам о причинно-следственной связи – она только обнаруживает связь без объяснения причин такой зависимости.

Другой важной задачей корреляционного анализа является определение влияния факторов на величину результата. Для решения этой задачи подбирается необходимый тип математической модели, которая лучше отражает характер исследуемой связи (прямой, обратной и других). Подбор математической модели или уравнения играет важную роль в корреляционном анализе, потому что от него зависит ход решения задачи и результаты расчетов [9].

Эффективным статистическим методом изучения взаимосвязей переменных, из которых одна рассматривается как зависимая, а другие – как независимые, является регрессионный анализ. В его основе лежит уравнение регрессии, которое показывает, каким будет в среднем изменение зависимой переменной  $y$ , если независимые переменные  $x_i$  примут конкретные значения [8].

Регрессионный анализ не только выявляет силу взаимосвязи между переменными, но и предоставляет возможность проверки гипотез о характере этих взаимосвязей. Используя статистические тесты, такие как  $t$ -критерий Стьюдента для коэффициентов регрессии, исследователи могут оценить, насколько значимо влияние независимых переменных на зависимую переменную. Это позволяет отличать значимые предикторы от случайных колебаний, обеспечивая тем самым более надежное основание для прогнозирования и принятия решений [3].

С другой стороны, дисперсионный анализ (ANOVA) используется, когда необходимо оценить влияние одной или нескольких категориальных независимых переменных на непрерывную зависимую переменную. ANOVA позволяет нам проверить гипотезу о том, что средние значения групп, сформированных на основе категориальных переменных, различаются. Это критически важно во многих областях исследований, включая клинические испытания, где изучается эффективность лекарств, или в маркетинговых исследованиях, где анализируется влияние рекламных кампаний на поведение потребителей [5].

### Обсуждение.

Для реализации методов математической статистики с помощью информационных технологий используется такой мощный инструмент проведения статистического анализа как табличный процессор Microsoft Excel.

На вкладке Данные ленты Microsoft Excel

расположена надстройка Анализ данных, применяемая для проведения статистического анализа. Надстройка Анализ данных содержит такие статистические инструменты, как Корреляция, Регрессия, Однофакторный дисперсионный анализ и Двухфакторный дисперсионный анализ.

Кроме этого, при выполнении статистических расчетов применяются статистические функции табличного процессора.

Для корреляционного анализа данных используют статистическую функцию КОРРЕЛ (массив1; массив2).

Здесь:

– Массив1 - это интервал ячеек одной случайной величины.

– Массив2 - это второй диапазон ячеек со значениями другой случайной величины.

Параметры  $a_0$  и  $a_1$  линейной регрессионной модели  $y = a_0 + a_1x$ , можно определить с помощью функций НАКЛОН и ОТРЕЗОК.

Функция НАКЛОН определяет угловой коэффициент наклона линии тренда, а функция ОТРЕЗОК - точку пересечения линии тренда с осью ОУ.

Для дисперсионного анализа данных, при котором происходит сравнение внутригрупповых и межгрупповых дисперсий, можно применять статистические функции Дисп.В, Дисп.Г, диспа, диспра для выборки или генеральной совокупности [7].

#### **Заключение.**

Таким образом, табличный процессор Microsoft Excel предоставляет большое количество возможностей для определения статистических закономерностей массивов данных.

Сочетание мощного функционала и дру-

жественного интерфейса привело к тому, что табличный процессор Microsoft Excel является одним из наиболее популярных средств проведения статистических расчетов [1].

Microsoft Excel не только облегчает выполнение статистических расчетов, но и делает их визуализацию более наглядной за счёт широкого спектра инструментов для создания графиков и диаграмм. Так, для отображения трендов и изменений в данных во времени часто используются линейные графики. Гистограммы и круговые диаграммы позволяют наглядно представить распределение частот или пропорций. Также, с помощью условного форматирования можно выделять значимые статистические отклонения и тенденции прямо в таблицах данных [10].

Для более сложного и углубленного анализа данных, такого как многомерный статистический анализ или анализ временных рядов, могут потребоваться специализированные программные продукты, такие как SPSS, SAS, или программирование на языке R или Python с использованием библиотеки Pandas и других. Однако даже в таких случаях, предварительная обработка и первичный анализ данных часто осуществляются в Excel, благодаря его доступности и широкому распространению в бизнес-среде и академическом сообществе [2].

Исследование статистических показателей позволяет установить степень связи между ними, определить характер этой связи или показать достоверность различий между показателями. Вследствие этого определяется роль каждого влияющего параметра в формировании уровня исследуемых показателей, углубляются знания об изучаемых объектах и закономерностях их развития.

#### **Конфликт интересов**

Не указан.

#### **Рецензия**

Все статьи проходят рецензирование в формате double-blind peer review (рецензенту неизвестны имя и должность автора, автору неизвестны имя и должность рецензента). Рецензия может быть предоставлена заинтересованным лицам по запросу.

#### **Conflict of Interest**

None declared.

#### **Review**

All articles are reviewed in the double-blind peer review format (the reviewer does not know the name and position of the author, the author does not know the name and position of the reviewer). The review can be provided to interested persons upon request.

#### **Литература:**

1. Алиакберова, Л. Ф. Современные программные продукты для анализа данных / Л. Ф. Алиакберова. — Текст: непосредственный // Молодой ученый. — 2021. — № 37 (379). — С. 13-17.
2. Античко, А. В. Инструменты для анализа данных: сравнение Python, R и других популярных платформ / А. В. Античко. — Текст: непосредственный // Молодой ученый. — 2023. — № 33 (480). — С. 14-16.
3. Береснева, И.В. Применение методов корреляционного анализа в Microsoft Excel для выявления зависимостей между выборками значений двух случайных величин // Синтез науки и образования

как инструмент решения глобальных проблем современности: сборник статей Международной научно-практической конференции Екатеринбург, 10 июня 2023 г. - Уфа: Аэтерна, 2023. - С. 129 -132.

4. Борздова, Т. В. Основы статистического анализа и обработка данных с применением Microsoft Excel: учеб. пособие / Т. В. Борздова. - Минск: ГИУСТ БГУ, 2021. - 75 с

5. Мамедова, Н. Г. Анализ данных бизнес-аналитики для организаций в современном мире / Н. Г. Мамедова. — Текст: непосредственный // Молодой ученый. — 2022. — № 24 (419). — С. 63-65.

6. Муратов, А.Н., Бекасов, Р.Д., Нурутдинов, А.О., Нурисламов Р.Н., Муродов Э.Р., Нишионов А.Б. Математическая статистика // Мировая наука. - - 2020. - №7 (40) – С. 231 – 240.

7. Петяркина В.Н. Значение информационных технологий при обработке статистических данных // Экономика и социум. - 2015. - №6 (19). – С. 872-874.

8. Шаталина А.М., Родионова Е.В. Использование корреляционно-регрессионного метода при анализе предприятий // Математическое и компьютерное моделирование в экономике, страховании и управлении рисками № 3, 2018. – С.272-276.

9. Ширинкина Е.В. Методы интеллектуального анализа данных и образовательной аналитики // Современное образование. - 2022. - №1. – С. 51 - 67.

10. Яргин С. В. Использование математической статистики для выявления манипуляций с количественными данными и образцами / С. В. Яргин. — Текст: непосредственный // Молодой ученый. — 2012. — № 9 (44). — С. 41-44.

#### References:

1. Aliakberova L. F. Modern software products for data analysis / L. F. Aliakberova. — Text: direct // Young scientist. — 2021. — № 37 (379). — Pp. 13-17.

2. Antipko, A.V. Tools for data analysis: a comparison of Python, R and other popular platforms / A.V. Antipko. — Text: direct // Young scientist. — 2023. — № 33 (480). — Pp. 14-16.

3. Beresneva I.V. Application of correlation analysis methods in Microsoft Excel to identify dependencies between samples of values of two random variables // Synthesis of science and education as a tool for solving global problems of our time: collection of articles of the International Scientific and practical Conference Yekaterinburg, June 10, 2023 - Ufa: Aeterna, 2023. - pp. 129 -132.

4. Borzdova, T. V. Fundamentals of statistical analysis and data processing using Microsoft Excel: textbook. the manual / T. V. Borzdova. - Minsk: GIUST BSU, 2021. - 75 c

5. Mammadova N. G. Analysis of business intelligence data for organizations in the modern world / N. G. Mammadova. — Text: direct // Young scientist. — 2022. — № 24 (419). — Pp. 63-65.

6. Muratov A.N., Bekasov R.D., Nurutdinov A.O., Nurislamov R.N., Murodov E.R., Nishonov A.B. Mathematical statistics // World science. - - 2020. - No.7 (40) – pp. 231-240.

7. Petyarkina V.N. The importance of information technologies in statistical data processing // Economy and society. - 2015. - №6 (19). – Pp. 872-874.

8. Shatalina A.M., Rodionova E.V. The use of the correlation and regression method in the analysis of enterprises // Mathematical and computer modeling in economics, insurance and risk management No. 3, 2018. – pp. 272 - 276.

9. Shirinkina E.V. Methods of data mining and educational analytics // Modern education. - 2022. - No.1. – pp. 51-67.

10. Yargin S. V. The use of mathematical statistics to identify manipulations with quantitative data and samples / S. V. Yargin. — Text: direct // Young scientist. — 2012. — № 9 (44). — Pp. 41-44.

#### Информация об авторе:

**Чикин Сергей Николаевич**, кандидат экономических наук, доцент, доцент кафедры статистики, ФГБОУ ВО «Новосибирский государственный университет экономики и управления», доцент кафедры статистики, [uprav551@mail.ru](mailto:uprav551@mail.ru)

**Sergey N. Chikin**, PhD in Economics, Associate Professor, Associate Professor of the Department of Statistics, Novosibirsk State University of Economics and Management, Associate Professor of the Department of Statistics.