

# **Очень краткое введение в математическую статистику для оценщиков**

К. А. Мурашев

8 октября 2021 г.

Какую бы работу не выполнял оценщик, во всех случаях он имеет дело с информацией и данными. Часто эти данные представляют собой числа либо могут быть формализованы иным образом. В любом случае требуется алгоритмическая обработка входных данных и преобразование их в информацию, а в некоторых случаях — в знания. Целью данного фрагмента является формирование общих представлений об основных понятиях и методах математической статистики, необходимых современному оценщику. Автор постарался прибегать к минимальному числу формул и сложных определений, хотя это и не вполне получилось. Поскольку конечной целью всей работы является цифровизация оценочной деятельности, в тексте приводятся короткие листинги на языках R и Python, позволяющие реализовать то, о чём говорится в тексте.

## **1. Что есть математическая статистика?**

У термина «статистика» существует несколько определений. Статистикой называют:

- данные, количественно описывающие тот или иной аспект окружающего мира: например данные об уровне безработицы, заболеваемости коронавирусом или доходах граждан, т. е. такие данные, которые описывают явление целиком;
- количественные данные, относящиеся к какому-либо одному субъекту либо результатам его деятельности: например количество выполненных оценщиком отчётов об оценке за календарный год;
- результаты исследования отдельных выборок: например итоги социологических опросов или результаты анализа рынка недвижимого имущества;
- конкретные методы анализа данных с помощью математических методов;

- т. н. статистики критерия, т. е. конкретные числовые значения отдельных вычислений, например статистика критерия Шапиро—Уилка;
- область знаний, которая разрабатывает и использует математические методы для описания данных и формирования суждений о них.

**Первый** тип статистики как правило не имеет прямого отношения к деятельности оценщика. Подобные сведения чаще всего могут быть получены из открытых источников. Кроме того, даже в случае недоверия к ним, у оценщика всё равно отсутствуют инструменты для получения подобных данных самостоятельно. **Второй** тип — скорее всего также не является особым предметом интереса оценщика. Определение стоимости объекта осуществляется методом аналогии путём сравнения с наблюдениями (предложениями либо сделками), тогда как погружение в свойства только самого объекта не позволяют определить его стоимость. **Третий** тип статистики отсылает нас к фундаментальному принципу: исследовать генеральную совокупность путём изучения выборки из неё. Собственно это и является предметом данной работы, а также профессиональной деятельности оценщика: формирование предсказания свойства объекта (его стоимости) на основе изучения выборки. **Четвёртый** — является ключевым с узко практической точки зрения. Процесс предсказания неизвестных свойств объектов на основе известных с учётом знаний, полученных при изучении выборки, по мнению многих, и есть статистика. Данный подход не является ошибочным, однако его вряд ли можно считать полным. Умение применять конкретные методы является необходимым, но недостаточным условием успешной работы. В настоящее время практически все вычисления выполняются программными средствами.<sup>1</sup> В связи с этим важность навыков ручного применения тех или иных методов сведена к минимуму. Вместо этого на первый план выходят навыки планирования оценочного статистического эксперимента, постановки задачи, поиск источников данных, их сбор и предобработка, общее понимание применяемых методов, выбор между ними, а также интерпретация полученного результата. **Пятый** тип означает результаты применения конкретных методов. В настоящее время чаще используются не сами статистики критериев, а универсальный показатель — р-значение (p-value). И, наконец, **шестое** определение означает обширную область человеческих знаний, в рамках которой существуют конкретные методы и результаты их применения. Данное определение может быть заменено термином «математическая статистика», подчёркивающим отличие от других значений общего термина «статистика». Именно в этом значении мы и будем использовать данный термин на протяжении всей работы по ознакомлению со статистическими основами оценки стоимости. Таким образом, если в тексте прямо не указано иное слово «статистика» следует понимать как «математическая статистика».

---

<sup>1</sup>Автор в своей работе использует языки программирования Python и R и рекомендует поступать также, однако существуют и иные средства: Julia, SPSS, PSPP, Stata и многие другие вплоть до табличных процессоров.

## 2. Генеральная совокупность и выборка

Генеральной совокупностью называется всё множество объектов, в отношении которого необходимо сделать те или иные выводы. В случае оценки, например торгово-развлекательного центра, расположенного на проспекте Просвещения в Санкт-Петербурге, генеральной совокупностью будут являться все торговые центры, расположенные на территории Санкт-Петербургской городской агломерации, независимо от того, выставлены они в данный момент на продажу или нет. Для того, чтобы понять, что является генеральной совокупностью, необходимо ответить на один простой вопрос: «на какое множество объектов можно обобщить полученные результаты исследования?». Очевидно, что независимо от того, выставлен ли какой-либо из существующих в агломерации ТРЦ на продажу или нет, совершались ли с ним в последнее время сделки или нет, все выводы, сделанные относительно множества «торгово-развлекательные центры, расположенные в СПбГА,<sup>2</sup> могут быть распространены на него равно как и на вообще любой объект данного типа, находящийся в указанных границах.

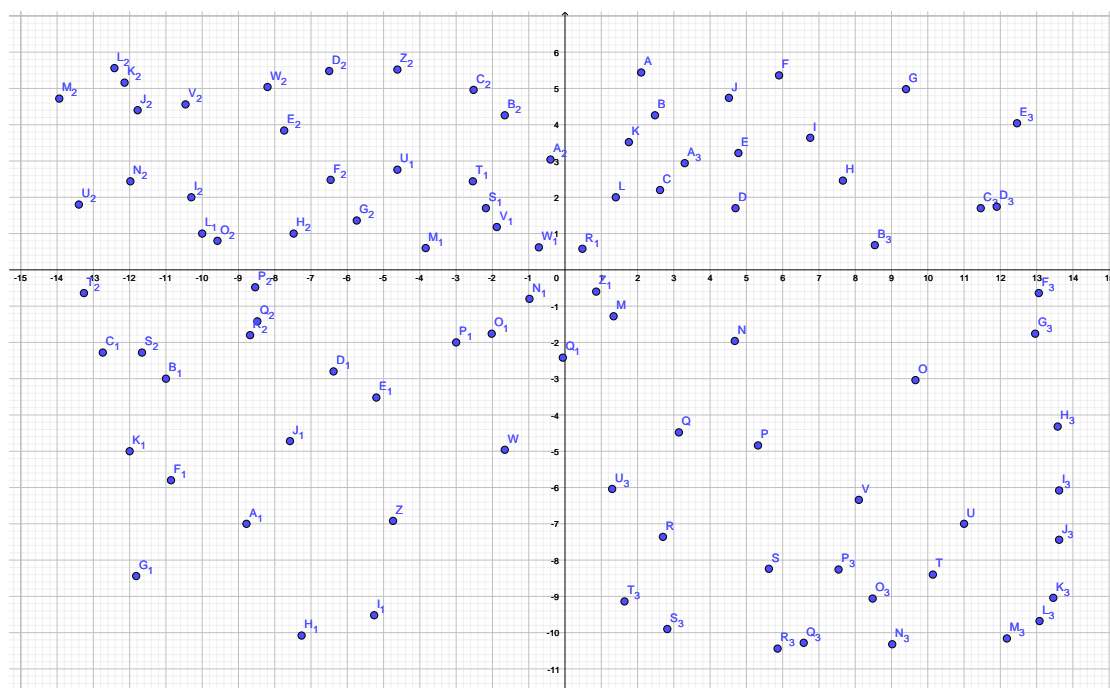


Рис. 1. Генеральная совокупность

Таким образом, процесс предсказания стоимости конкретного объекта происходит в два этапа и на основе двух принципов:

---

<sup>2</sup>Санкт-Петербургская городская агломерация. Включает в себя Санкт-Петербург, большую часть Всеволожского, части Выборгского, Кировского, Тосненского, Гатчинского и Ломоносовского районов Ленинградской области.

- 1) формирование выводов о свойствах всего открытого рынка на основе изучения выборки — **принцип «от частного к общему»**;
- 2) формирование вывода о стоимости конкретного объекта на основе установленных свойств открытого рынка, к которому он относится — **принцип «от общего к частному»**.

В случае необходимости формирования общих выводов о свойствах рынка достаточно проведение только первой части анализа.



3. Типы данных
4. Меры центральной тенденции
5. Меры изменчивости
6. Квантили распределения
7. Распределения
  - 7.1. Нормальное распределение
  - 7.2. Логарифмически нормальное распределение
  - 7.3. Равномерное распределение
  - 7.4. Экспоненциальное распределение
  - 7.5. Нормальное распределение
  - 7.6. Распределение Вейбулла
  - 7.7. Нормальное распределение
  - 7.8. Гамма распределение
  - 7.9. Бета распределение
  - 7.10. Распределение  $\chi^2$  (Распределение Пирсона)
  - 7.11. Распределение Стьюдента (t-распределение)
  - 7.12. Распределение Фишера (F-распределение)
  - 7.13. Логистическое распределение
  - 7.14. Распределение Парето
8. Проверка распределения на нормальность
9. Центральная предельная теорема
10. Доверительные интервалы
11. Сравнение средних (t-критерий Стьюдента)
12. Однофакторный дисперсионный анализ
13. ANOVA
14. A/B тесты
15. Корреляционный анализ
  - 15.1. Параметрические методы
  - 15.2. Непараметрические методы
16. Регрессионный анализ
  - 16.1. GLM

лишь весьма ограниченный круг вопросов. В списке источников информации дана некоторая подборка материалов, которые, по мнению автора, могут быть полезными в дальнейшем освоении вопросов математической статистики. Поскольку оценщики, как правило, очень занятые люди, освоение вряд ли можно ожидать, что кто-то захочет сразу изучать десятки материалов. В связи с этим автор рекомендует ознакомление в первую очередь со следующими работами:

- В. Савельев. *Статистика и коттики* [18]. Прекрасная книга для тех, кто только начинает погружение в область математической статистики.
- С. Бослаф. *Статистика для всех* [14]. Одна из лучших книг по статистике для тех, кто не учился по профилю и хочет освоить её методы на уровне, достаточном для применения в профессиональной деятельности.
- А. Кобзарь. *Прикладная математическая статистика* [11]. Данный материал представляет собой обширное пособие, предназначенное для тех, кто уже владеет некоторой базой и хочет внедрить применение методов математической статистики на профессиональном уровне.

Учиться следует всю жизнь. Оценочная деятельность трансформируется, и, спустя несколько лет, данное пособие будет казаться детской раскраской. Автор желает читателям непрерывного совершенствования и становления в качестве настоящих специалистов в сфере цифровой оценки XXI века.

## Источники информации

- [1] Alan Stuart Maurice G. Kendall. *Статистические выводы и связи. Перевод с английского.* Л. И. Гальчук, А. Т. Терёхин под ред. А. Н. Колмогорова. Пер. А. Т. Терёхин Л. И. Гальчук. Примеч. А. Н. Колмогоров. 1973.
- [2] С. А. Айвазян. *Прикладная статистика: исследование зависимостей.* 1985.
- [3] Г. Смит Н. Дрейпер. *Прикладной регрессионный анализ.* 1987.
- [4] Ю. В. Прохоров. *Вероятность и математическая статистика. Энциклопедия.* Moskva: Nauch. izd-vo "Bolqshaya Rossijskaya Encziklopediya, 1999. ISBN: 5-85270-265-X.
- [5] Госстандарт России. *ГОСТ Р 50779.10-2000 Статистические методы. Вероятность и основы статистики. Термины и определения.* Москва, 2000.
- [6] Госстандарт России. *ГОСТ Р 50779.11-2000 (ИСО 3534.2-93) Статистические методы. Статистическое управление качеством. Термины и определения (Переиздание).* Москва, 2000.
- [7] Госстандарт России. *ГОСТ Р ИСО 5725-1-2002 Точность (правильность и прецизионность) методов и результатов измерений. Часть 1. Основные положения и определения.* Москва, 2002.

- [8] Е. В. Чимитова Б. Ю. Лемешко. «О выборе числе интервалов в критериях согласия типа Хи-квадрат». В: *Заводская лаборатория. Диагностика материалов* (2003).
- [9] В. Р. Бараз. *Корреляционно-регрессионный анализ связи показателей коммерческой деятельности с использованием программы Excel*. Федеральное агентство по образованию. ГОУ ВПО «Уральский государственный технический университет — УПИ», 2005.
- [10] Т. М. Сизова. *Статистика. Учебное пособие*. 2005.
- [11] А. И. Кобзарь. *Прикладная математическая статистика*. 2006.
- [12] Т. А. Лёзина В. Л. Аббакумов. *Бизнес-анализ информации. Статистические методы*. 2009. ISBN: 978-5-282-02918-5.
- [13] О. Р. Никитин. *Статистические методы обработки параметров радиосигналов*. 2012.
- [14] Сара Бослаф. *Статистика для всех*. 2015.
- [15] А. М. Шихалёв. *Регрессионный анализ. Парная линейная регрессия*. Науч. отч. Казан. ун-т, 2015.
- [16] Михаил Хальман. «Регрессионный анализ.» В: *Прикладной статистический анализ данных*. 2017.
- [17] Н. В. Казанцева. *Математическое моделирование в программных пакетах Excel и MathCad : учеб.-метод. пособие*. 2018.
- [18] Владимир Савельев. *Статистика и коттики*. Русский. Москва: Издательство АСТ, 2018, с. 122. ISBN: 978-5-17-106143-2.
- [19] Институт биоинформатики. *Основы статистики. Часть 1*. Русский. 2020. URL: <https://stepik.org/course/76/info> (дата обр. 07.10.2021).
- [20] Институт биоинформатики. *Основы статистики. Часть 2*. Русский. 2020. URL: <https://stepik.org/course/524/info> (дата обр. 07.10.2021).
- [21] Институт биоинформатики. *Основы статистики. Часть 3*. Русский. 2020. URL: <https://stepik.org/course/2152/info> (дата обр. 07.10.2021).
- [22] Computer Science Center. *Введение в математический анализ*. URL: <https://stepik.org/course/95/info>.
- [23] Computer Science Center. *Лекбез по дискретной математике*. URL: <https://stepik.org/course/91/info>.
- [24] Computer Science Center. *Линейная алгебра*. URL: <https://stepik.org/course/2461/info>.
- [25] Computer Science Center. *Математическая статистика*. URL: <https://stepik.org/course/326/info>.
- [26] Computer Science Center. *Математический анализ (часть 1)*. URL: <https://stepik.org/course/716/info>.



- [27] Computer Science Center. *Математический анализ (часть 2)*. URL: <https://stepik.org/course/711/info>.
- [28] Computer Science Center. *Основы теории графов*. URL: <https://stepik.org/course/126/info>.
- [29] Computer Science Center. *Теория вероятностей*. URL: <https://stepik.org/course/3089/info>.
- [30] Татьяна Кабанова. *Статистика для гуманитариев*. URL: <https://stepik.org/course/83603/info>.
- [31] Игорь Клейнер. *Теория вероятностей в удовольствие (курс 1 основы)*. URL: <https://stepik.org/course/102057/info>.