В. А. Малугин

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УЧЕБНОЕ ПОСОБИЕ ДЛЯ СПО

Рекомендовано Учебно-методическим отделом среднего профессионального образования в качестве учебного пособия для студентов образовательных учреждений среднего профессионального образования

Москва ▪ Юрайт ▪ 2018

УДК 519.2(075.32)
ББК 22.171я723
М19

***Автор*:**

Малугин Виталий Александрович — кандидат физико-математических наук, доцент кафедры математических методов анализа экономики экономического факультета Московского государственного университета имени М. В. Ломоносова.

***Рецензенты*:**

Маршев В. И. — доктор экономических наук, профессор кафедры управления организацией экономического факультета Московского государственного университета имени М. В. Ломоносова;

Афанасьев М. Ю. — доктор экономических наук, профессор, заведующий лабораторией прикладной эконометрики Центрального экономико-математического института Российской академии наук, ведущий научный сотрудник экономического факультета государственного академического университета гуманитарных наук.

Малугин, В. А.

М19 Математическая статистика : учеб. пособие для СПО / В. А. Малугин. — М. : Издательство Юрайт, 2018. — 218 с. — (Серия : Профессиональное образование).

ISBN 978-5-534-09872-3

В книге освещены основные идеи математической статистики, необходимые для полноценного освоения эконометрики и смежных экономико-математических дисциплин. Изложение сопровождается вопросами для повторения, решенными задачами на экономическую тематику и значительным числом содержательных экономико-статистических задач для самостоятельного решения.

Соответствует актуальным требованиям Федерального государственного образовательного стандарта среднего профессионального образования и профессиональным требованиям.

Для студентов среднего профессионального образования. Отдельные главы и задачи в них могут быть использованы в школьных группах повышенной нагрузки.

УДК 519.2(075.32)
ББК 22.171я723

ISBN 978-5-534-09872-3

Оглавление

Предисловие

> Наука начинается там, где начинаются измерения.
>
> *Д. И. Менделеев*

Математическая статистика наряду с теорией вероятностей представляет собой один из предметов второго уровня математического образования, опирающийся на математический анализ и линейную алгебру как на предметы первого уровня. К третьему уровню следует отнести специальные для определенной профессии математические дисциплины. Для экономистов такой дисциплиной является, например, эконометрика.

В экономическом учебном заведении, где математика является прикладной дисциплиной, все эти предметы преподаются на достаточно серьезном уровне. Тем не менее для каждого предмета ищется компромисс между строгостью и простотой. Данная книга предназначена для студентов, которые не выбрали математику в экономике своей основной специальностью, но готовы применять методы математической статистики в своей профессиональной деятельности. В книге рассматривается материал одного семестра обучения с вступительной лекцией по элементарной теории ошибок. Курс математической статистики был прочитан в течение 18 недель (одна лекция в неделю), материал проработан на 18 семинарах. Автор, не имея возможности прочитать на очередном занятии весь материал лекции, выбирал наиболее интересные и важные для понимания предмета страницы, публикуя предварительно в интернете лекцию целиком. Это давало возможность заинтересованным студентам слушать наиболее важные и часто достаточно сложные части материала, вначале ознакомившись с ним, что позволяло на лекции концентрировать внимание на содержании без необходимости вести параллельные подробные записи, ограничиваясь небольшими заметками.

В равной степени было обращено внимание на теорию и на методы решения задач. Основные теоремы не только формулируются, но и по возможности доказываются. Уровень строгости доказательства не всегда соответствует требованиям научной монографии, но, как нам представляется, удовлетворяет требованиям учебного пособия для студентов. Доказательства теорем и утверждений снабжены значками начала и конца рассуждений ◀ ▶. Большое число приведенных примеров, иллюстрирующих основные положения, большое число разобранных задач, в которых заложены наиболее важные идеи, помогут легче воспринять и глубже усвоить достаточно сложный учебный материал. Задачи для самостоятельного решения снабжены ответами, а также формулами, объясняющими идею решения. Встречаясь в будущем с экономическими задачами, специалист должен узнавать знакомые с университетских времен примеры и брать их решения за образец.

В результате изучения материала учебного пособия студент должен освоить:

трудовые действия

- методами определения характеристик случайных величин;
- навыками построения и исследования точечных и интервальных оценок статистических параметров;
- навыками построения и проверки статистических гипотез;
- методами решения экономических задач с использованием математической статистики;

необходимые умения

- вычислять различные характеристики случайной величины;
- использовать теоремы математической статистики при решении вероятностных задач;
- исследовать точечные оценки на несмещенность, состоятельность и эффективность;
- минимизировать ошибки первого и второго рода;
- применять методы математической статистики для исследования статистических параметров и проверки гипотез;

необходимые знания

- определений основных терминов математической статистики;
- основных формул и теорем математической статистики;
- приложений математической статистики в экономике;
- методов, используемые в математической статистике.

При написании учебника автор, с одной стороны, пытался удовлетворить требованиям, предъявляемым к преподаванию курса математической статистики Федеральным государственным образовательным стандартом. С другой стороны, автор соблюдал преемственность в использовании идей и методов, теории и задач, наработанных предыдущими лекторами, читавшими этот курс на экономическом факультете МГУ и представленных ими в пособиях для студентов.

Глава 1
ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ОШИБОК

В результате освоения материала данной главы студент должен:

знать

- основные погрешности при совершении измерений и их классификацию;
- основные идеи, заложенные Гауссом при создании теории ошибок;

уметь

- вычислять погрешности прямых и косвенных наблюдений;
- получать максимально надежные результаты при вычислительной обработке наблюдений;

владеть

- навыками вычислительной культуры.

1.1. Погрешности наблюдений и измерений

Теория ошибок — это дисциплина, которая вносит свой вклад в получение научно обоснованных максимально надежных результатов наблюдений и измерений и в проведение первичной вычислительной обработки полученных данных с минимальными потерями в информации.

Проводя наблюдения или делая измерения, исследователь совершает ошибки. Существует мало наблюдений или измерений, которые дают истинную картину или значение величины. Обычно это наблюдения за дискретными элементами. Непрерывную величину невозможно измерить без ошибки. Садовод собрал в корзинку малину. Число ягод в ней — случайная величина, но их можно пересчитать. Вес ягод — другая случайная величина, указать точное значение веса невозможно. Чтобы измерить длину забора, рабочий использует механическую рулетку, ошибка измерения которой 0,5 см = $5 \cdot 10^{-3}$ м. Поэтому забор длиной в 10,00 м при многократных измерениях окажется имеющим длину 10,01 м, 9,99 м, 10,005 м и т.д.

Точность измерений — характеристика качества измерений, отражающая степень близости результатов измерений к истинному значению измеряемой величины. Можно увеличивать точность измерения. Но совсем не ясно, о чем идет речь, если поставить задачу об измерении, например, длины стержня с погрешностью до размеров атома или даже электрона. Принципиально неограниченная точность измерения длин имеет смысл для абстрактных прямолинейных отрезков геометрии, а не для реальных тел, имеющих атомистическую структуру. В науке и производстве встречаются задачи по измерению длины с высокой, но, конечно, ограниченной точностью. Развитие таких технологий, как производство интегральных схем с микронными и субмикронными размерами, требует измерений с точностью до ангстрема, т.е. до 10^{-10} м. На XVII Генеральной конференции по мерам и весам было принято считать метром длину пути, проходимого светом в вакууме за 1/299 792 458 с. Хотя точность измерения метра велика, она ограничена величиной $3 \cdot 10^{-9}$ м.

Итак, получить абсолютно точно *истинное значение* величины в большинстве опытов невозможно. Вместо истинного значения используется так называемое *действительное значение* величины. Оно может представлять собой единичное измерение, как это происходит в спорте, или так называемую статистическую оценку, которая является наиболее вероятным значением, полученным по результатам научно обоснованной статистической обработки нескольких измерений. Действительное значение должно быть настолько близко к истинному, чтобы в поставленной задаче оно могло быть использовано вместо истинного.

В математической статистике истинное значение обозначается обычно через θ или x_0 и называется *теоретическим значением* параметра, поскольку его в некоторых задачах можно вычислить. Действительное значение будем определять через $\hat{\theta}_n$, где число n указывает на число измерений, и называется *выборочным значением* параметра, поскольку оно рассчитывается по наблюдениям $x_1, x_2, \ldots, x_n$ (по выборке). К величине $\hat{\theta}_n$ предъявляются жесткие требования, при выполнении которых $\hat{\theta}_n$ удовлетворительно описывает величину θ.

Модуль разности Δx между истинным значением измеряемой величины θ и ее действительным значением $\hat{\theta}_n$ называется *погрешностью* (ошибкой) или отклонением:

$$|\theta - \hat{\theta}_n| = \Delta x.$$

Погрешность указывает на то, с какой точностью по результатам наблюдений мы определяем наблюдаемую величину. Она не может быть вычислена по приведенной формуле в силу незнания нами величины θ, если, конечно, не удастся величину θ рассчитать теоретически.

Существуют разные подходы к оценке Δx. Во всех случаях необходимо провести серию наблюдений и получить значения $x_1, x_2, \ldots, x_n$ исследуемой величины.

1. Разыскиваются наибольшее $x_{\max}$ и наименьшее $x_{\min}$ экспериментальные значения. Погрешность вычисляется по формуле

$$\Delta x = \frac{x_{\max} - x_{\min}}{2}.$$

2. Вычисляется среднее значение измеряемой величины

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Погрешность разыскивается как среднее линейное отклонение:

$$\Delta x = \frac{|x_1 - \overline{x}| + |x_2 - \overline{x}| + \ldots + |x_n - \overline{x}|}{n}.$$

3. Погрешность вычисляется как среднее квадратическое отклонение:

$$\Delta x = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1}}.$$

4. Если в измерениях делается акцент на определении погрешности измерительного прибора, а значение измеряемой величины $\theta \equiv a$ известно, то погрешность прибора вычисляется как

$$\Delta x = \sqrt{\frac{(x_1 - a)^2 + (x_2 - a)^2 + \ldots + (x_n - a)^2}{n}}.$$

С развитием общества требования к величине погрешности измерений увеличиваются. Например, в беге на 100 м время раньше фиксировалось по ручному секундомеру с погрешностью 0,1 с. Со временем мировые рекорды росли на все меньшую величину, и улучшить мировой рекорд на 0,1 с стало чрезвычайно сложно. Десятки спортсменов официально показывали одно и то же время и являлись соавторами мировых рекордов. Переход на фиксацию электронным секундомером с погрешностью 0,001 с позволил решить эту проблему.

Примеры мировых рекордов в беге на 100 м:

1968—1975 гг. — 9,9 ± 0,1 с;

2005 г. — 9,768 ± 0,001 с;

2006 г. — 9,763 ± 0,001 с;

2009 г. — 9,578 ± 0,001 с.

Примеры показывают, что в измерениях необходимо указывать на погрешность измерения. Запись 9,578 ± 0,001 с означает, что истинное значение мирового рекорда лежит в интервале от 9,577 до 9,579 с с определенной заранее рассчитанной вероятностью, в данном случае равной 0,997.

1.2. Классификация погрешностей

Погрешности классифицируются по разным признакам.

1. По форме представления погрешности делятся на абсолютные и относительные.

Абсолютная погрешность — разность между измеренным и действительным значением измеренной величины: $\Delta x_{абс} = x_i - \hat{\theta}_n$, где $i = 1, 2, \ldots, n$.

Абсолютная погрешность измерения меняется от наблюдения к наблюдению, она измеряется в тех же единицах измерения, что и сама величина.

Относительная погрешность — погрешность измерения, выраженная отношением абсолютной погрешности измерения к действительному значению измеряемой величины:

$$\delta_x = \frac{\Delta x_{абс}}{\hat{\theta}_n}.$$

Относительная погрешность является безразмерной величиной; ее численное значение может указываться, например, в процентах.

2. По причине возникновения погрешности делятся на инструментальные, методические и операторные.

Инструментальные погрешности — погрешности, которые определяются погрешностями применяемых средств измерений и вызываются несовершенством принципа действия, неточностью градуировки шкалы и т.д.

Методические погрешности — погрешности, обусловленные несовершенством метода, а также упрощениями, положенными в основу методики.

Субъективные/операторные/личные погрешности — погрешности, обусловленные степенью внимательности, сосредоточенности, подготовленности и другими качествами оператора.

3. По характеру воздействия погрешности делятся на случайные ошибки, систематические ошибки и промахи.

Случайные ошибки — это ошибки, которые возникают при воздействии одновременно множества источников помех. Попадание в центр мишени зависит не только от умения стрелка, но и от множества факторов, участвующих в процессе стрельбы: силы бокового ветра, веса пули, изношенности ствола, амплитуды дрожания рук и т.д. Воздействие факторов меняется от выстрела к выстрелу, уменьшается точность попадания. Было замечено, что:

1) чем меньше по абсолютной величине случайная ошибка, тем она чаще встречается при измерениях;

2) одинаковые по абсолютной величине случайные ошибки одинаково часто встречаются при измерениях;

3) при проведении равноточных наблюдений величина случайной погрешности по абсолютной величине не превосходит некоторого максимального значения;

4) среднее значение случайных ошибок стремится к нулю при неограниченном возрастании числа измерений.

Если измерения или наблюдения повторяются многократно, ошибки, вызванные воздействием случайных факторов, могут быть учтены с помощью математической статистики и сведены к минимуму. Уже с Птолемея на точность наблюдений стали обращать внимание.

Существенный вклад в теорию ошибок принадлежит К. Ф. Гауссу. Рассмотрим его основные идеи.

Пусть x_i, $i = 1, 2, \ldots, n$, — числовые значения проведенных наблюдений или результат измерения некоторой величины, истинное значение которой x_0. Очевидно, что результат измерения отличается от истинного значения измеряемой величины: $x_i = x_0 + \varepsilon_i$, где ε_i — ошибка измерения, принимающая равновероятно как положительные значения, так и отрицательные. Следует сказать, что на переломе веков (1790—1810), когда делались первые шаги по созданию теории ошибок, понятие случайной величины еще не было введено, рассуждения велись в других терминах. Пусть ε_i, $i = 1, 2, \ldots, n$, есть значения случайной величины ε с плотностью распределения $p_\varepsilon(x)$, где аргумент x принимает значения ошибок измерения. Вероятность того, что при проведении n измерений возникнут ошибки $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, равна

$$\Omega = p_\varepsilon(\varepsilon_1) \cdot p_\varepsilon(\varepsilon_2) \cdot \ldots \cdot p_\varepsilon(\varepsilon_n).$$

Истинное значение измеряемой величины x_0 неизвестно, но можно в качестве x_0 подобрать такое значение $\hat{x}$, которое максимизирует величину Ω, поскольку чем лучше набор ошибок относительно $\hat{x}$ описывается плотностью распределения, тем больше будет величина произведения. Эти рассуждения привели к постановке задачи исследования функции Ω на максимум.

На первом этапе найдем вид функции $p_\varepsilon(\varepsilon_i)$. При замене x_0 на $\hat{x}$ и в предположении дифференцируемости функции получаем дифференциальное уравнение

$$\frac{p'_\varepsilon(x_1-\hat{x})}{p_\varepsilon(x_1-\hat{x})}+\frac{p'_\varepsilon(x_2-\hat{x})}{p_\varepsilon(x_2-\hat{x})}+\ldots+\frac{p'_\varepsilon(x_n-\hat{x})}{p_\varepsilon(x_n-\hat{x})}=0.$$

Не вдаваясь в подробности последующих преобразований, укажем, что дальнейшие предположения позволяют свести уравнение к виду:

$$\frac{p'_\varepsilon(x)}{xp_\varepsilon(x)}=-\frac{1}{\sum\limits_{i=1}^{n}\varepsilon_i^2 p_\varepsilon(\varepsilon_i)}\equiv-\frac{1}{\sigma^2}=\text{const}.$$

где переменная x представляет разность между измеренным значением и величиной $\hat{x}$.

Его решение $p_\varepsilon(x)=\alpha e^{-\frac{x^2}{2\sigma^2}}$.

Подбор коэффициента α перед экспонентой можно провести из условия

$$\int\limits_{-\infty}^{\infty}\alpha e^{-\frac{x^2}{2\sigma^2}}dx=1.$$

Отсюда $\alpha = \frac{1}{\sqrt{2\pi}\sigma}$.

Окончательно, плотность распределения ошибок измерения оказалась равной

$$p_\varepsilon(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}.$$

На втором этапе найдем $\hat{x}$ — действительное значение измеряемой величины. Для одного измерения функция плотности ошибок, или вероятность того, что измерение будет сделано с ошибкой ε_1, равна

$$\left.\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}\right|_{x=\varepsilon_1=x_1-\hat{x}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-\hat{x})^2}{2\sigma^2}}.$$

Для n наблюдений получаем вероятность

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-\hat{x})^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_2-\hat{x})^2}{2\sigma^2}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_n-\hat{x})^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i-\hat{x})^2}{2\sigma^2}}.$$

Максимум вероятности достигается в такой точке $\hat{x}$, в которой функция $\sum_{i=1}^{n}(x_i-\hat{x})^2 = \sum_{i=1}^{n}\varepsilon_i^2$ минимальна. Исследование на минимум относительно параметра $\hat{x}$ дает его оценку

$$\hat{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

Таким образом, при проведении n наблюдений сумма квадратов ошибок измерений станет минимальной, если в качестве действительного значения измеряемого параметра $\hat{x}$ взять среднее значение $\bar{x}$ по всем наблюдениям.

На третьем этапе перейдем к нахождению истинного значения x_0 измеряемой величины. При увеличении количества измерений найдем предел среднего значения случайной погрешности

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n}\varepsilon_i}{n}.$$

Предел будет равен нулю, потому что в числителе сумма случайных ошибок будет конечной величиной, так как положительные и отрицательные случайные ошибки при сложении будут компенсироваться. Вспомним, что вначале погрешность измерений ε_i вводилась, как $\varepsilon_i = x_i - x_0$. Тогда

$$0 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(x_i - x_0).$$

Отсюда $x_0 = \lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \lim\limits_{n\to\infty} \overline{x} = \lim\limits_{n\to\infty} \hat{x}$, т.е. оценка параметра $\hat{x}$ величиной $\overline{x}$ при $n \to \infty$ дает наиболее близкое к истинному значение измеряемой величины.

Систематические ошибки — это ошибки, которые постоянно либо преувеличивают, либо преуменьшают результаты измерений вследствие воздействия определенных факторов, систематически влияющих на эти измерения и изменяющих их в одном направлении.

Из них можно выделить:

- систематические ошибки, природа которых известна, а величина может быть найдена, следовательно, исключена введением поправок. Это поправки, уточняющие теорию, поправки за счет постоянного воздействия и т.д.;
- систематические ошибки неизвестного происхождения (недостаточно разработанная теория, сложный эксперимент), но их величина не превышает некоторого определенного значения. В этом случае при записи ответа может быть указано их максимальное значение;
- систематические ошибки, о существовании которых ничего не известно, но результат стабильно завышается или занижается. Ошибки такого типа можно выявить с помощью дополнительных исследований, путем проведения измерений совсем другим методом и в других условиях, «улучшением» эксперимента;
- систематические ошибки, обусловленные классом точности приборов. В предположении, что случайная величина (показание прибора) имеет нормальное распределение, диапазон отклонения показаний от действительного значения измеряемой величины не будет превышать 3σ с вероятностью 0,997, где σ — среднее квадратическое отклонение. Величина 3σ называется точностью прибора. Чаще всего класс точности приборов считается основным источником систематических ошибок. В электроизмерительных приборах наиболее распространены классы от 0,05 до 4. Это число показывает максимально возможную погрешность прибора, выраженную в процентах от $A_{\max}$ — наибольшего значения величины, измеряемой в данном диапазоне работы прибора:

$$\text{Класс точности прибора} = \frac{3\sigma}{A_{\max}} \cdot 100\%.$$

Так, для вольтметра, работающего в диапазоне измерений 0—30 В, класс точности 1,0 означает, что указанная погрешность при положении стрелки в любом месте шкалы не превышает 0,3 В. Это точность показаний прибора. Соответственно, среднее квадратическое отклонение прибора составляет 0,1 В. На хороших электроизмерительных приборах цена деления шкалы согласована с классом данного прибора.

Термин «точность» используется при вычислениях. Например, говорят, что вычисления сделаны с точностью до второй значащей цифры после запятой, если числа соответственно округляются.

Промахи обычно резко выделяются на фоне остальных измерений. Можно считать некоторое измерение промахом, если вероятность случайного появления такого значения достаточно мала. Рассмотрим эту проблему подробнее. В соответствии с центральной предельной теоремой среднее значение случайных величин $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ при $n \to \infty$ ведет себя как нормально распределенная случайная величина. Вероятность появления значения, отклоняющегося от среднего арифметического $\bar{x}$ более, чем на 3σ, не превышает величины 0,003, т.е. подобные события маловероятны. Поэтому измерения x_i, для которых выполняется неравенство $|x_i - \bar{x}| > 3\sigma$, должны быть выделены для последующего решения вопроса, что с ними делать. Если имеется много наблюдений или измерений, то такие измерения x_i следует отбросить. Удаление сильно отклоняющихся значений не приведет к существенной ошибке. Исключив промахи, можно улучшить результат.

Если наблюдения редки или получение их сопряжено с большими затратами усилий, следует проверить наличие промахов, проследив всю цепочку получения данного наблюдения x_i, его хранения и первичной обработки. При невозможности обнаружить ошибку следует проверить, как сильно x_i меняет окончательный результат. Если наблюдение появилось с вероятностью менее 0,01, т.е. менее, чем в одном случае из ста, его можно отбросить. Оно мало повлияет на числовые характеристики закона распределения. Если вероятность больше 0,1, то x_i следует оставить. При числе наблюдений $n > 100$ число значений x_i становится больше 10. Отбрасывание аномальных наблюдений в этом случае может привести к потере неожидаемого исследователем важного эффекта. В истории науки известно много случаев, когда результаты, выглядевшие сначала как промахи, приводили к научным открытиям. Иной подход к фильтрации наблюдений может рассматриваться как «подгонка» данных под нужные выводы.

1.3. Погрешности косвенных наблюдений

Многие исследуемые величины невозможно измерять напрямую или наблюдать явно. Они вычисляются как функции аргументов, значения которых получены в результате прямых измерений. Такая процедура наблюдений называется *косвенными измерениями*, а ошибки вычисляемой по результатам наблюдений функции определяются ошибками входящих в формулы прямых измерений.

Пусть для вычисления величины y проведены однократные наблюдения величин x_1^0, x_2^0, …, x_n^0, которые связаны зависимостью $y = y(x_1^0$, x_2^0, …, $x_n^0)$. При измерениях получены значения $x_1^0, x_2^0, ..., x_n^0$ с малыми ошибками Δx_1, Δx_2, …, Δx_n.

Разложим функцию $y(x_1, x_2, \dots, x_n)$ в ряд Тейлора в окрестности точки $x_1^0, x_2^0, ..., x_n^0$, оставив линейные слагаемые:

$$y(x_1, x_2, ..., x_n) - y(x_1^0, x_2^0, ..., x_n^0) \approx$$
$$\approx y'_{x_1}\Big|_{x_{i=1,\dots,n}=x_i^0} \Delta x_1 + y'_{x_2}\Big|_{x_{i=1,\dots,n}=x_i^0} \Delta x_2 + \ldots + y'_{x_n}\Big|_{x_{i=1,\dots,n}=x_i^0} \Delta x_n.$$

Приращение $y(x_1, x_2, ..., x_n) - y(x_1^0, x_2^0, ..., x_n^0) = \Delta y$ называется *ошибкой косвенных наблюдений*. Обычно находят ее максимально возможное значение $\Delta y_{\max}$, которое возникает, если приращения аргументов (ошибки измерений) подбирать со знаками, увеличивающими вычисляемую погрешность, или слагаемые просто брать по модулю.

Рассмотрим примеры нахождения погрешности.

Пример 1.1. Найти погрешность вычисления функции в точке x^0, если $y = f(x)$.

Решение. Имеем

$$\Delta y = | f'_x\big|_{x=x^0} \cdot \Delta x |.$$

Пример 1.2. Найти погрешность вычисления функции, если $y = x_1 x_2$.

Решение. Имеем

$$\Delta y = x_1 x_2 - x_1^0 x_2^0 = (x_1^0 + \Delta x_1)(x_2^0 + \Delta x_2) - x_1^0 x_2^0 = x_1^0 \Delta x_2 + x_2^0 \Delta x_1 + \Delta x_1 \Delta x_2.$$

Максимальная ошибка косвенных наблюдений принимается равной

$$\Delta y_{\max} = | x_1^0 \Delta x_2 | + | x_2^0 \Delta x_1 |.$$

1.4. Погрешности, возникающие при первичной обработке данных

Данные, собираемые для решения какой-либо задачи, могут быть получены из разных источников, могут дополнять друг друга, могут быть измерены различными приборами. Культура вычислений требует, чтобы после фильтрации данных (удаления промахов и вычитания систематических ошибок) они были приведены к единообразному виду.

1. В случае использования приближения десятичной дробью числа следует писать одинаковое число знаков после запятой:

Получены числа	Следует писать
0,235	0,235
1/5	0,200
2/3	0,667

2. Для получения обоснованной точности измерений, сделанных приборами с разными классами точности или при наличии нескольких значений, вычисленных с разной точностью, необходимо уменьшать ошибку, имеющую наибольшую величину. Например:

$$1{,}234+0{,}55+0{,}7\approx 1{,}23+0{,}55+\underbrace{0{,}72}_{\substack{\text{Измеренное или вычисленное}\\ \text{значение после уточнения}}}=2{,}50.$$

Получившийся в конце нуль отбрасывать не следует. Он указывает, что получено число с точностью до второй значащей цифры после запятой.

3. Точность обработки числового материала должна быть согласована с точностью самих измерений. Вычисления, проведенные с бóльшим числом знаков, чем это необходимо, создают ложное впечатление о большой точности измерений. В то же время не следует ухудшать результаты измерений, грубо округляя измерения. Точность данных, поступающих в совместную обработку, будет определяться точностью наиболее грубых данных.

4. Последняя значащая цифра в любом приводимом результате должна быть того же порядка величины (находиться в той же десятичной позиции), что и ошибка измерения. Если погрешность равна 0,03, результат измерения 0,557 следует округлить до 0,56.

5. Точность полученного значения определяется не последней цифрой в десятичной записи числа, а известной или вычисленной погрешностью. Результат необходимо приводить с этой погрешностью $0{,}56 \pm 0{,}03$. Если $3\sigma = 0{,}03$, то истинное значение измерения лежит в интервале от 0,53 до 0,59 с вероятностью 0,997.

Подводя итог краткому введению в теорию ошибок, можно сказать, что идеи и методы элементарной теории ошибок позволяют провести наблюдения и вычисления с научно обоснованным результатом, наиболее близким к истинному значению величины.

Вопросы и задания для повторения

1. Что такое истинное значение величины и как точно оно может ли быть оно измерено?
2. Что называется погрешностью измерения? Как она может вычисляться?
3. Укажите признаки, по которым классифицируются погрешности.
4. Как предложил Гаусс минимизировать случайные ошибки?

5. Каким образом в теории ошибок Гаусса при проведении измерений может быть достигнуто истинное значение измеряемой величины?

6. Что такое класс точности прибора?

7. Как вычисляются погрешности при косвенных измерениях?

8. Привести несколько идей, обеспечивающих культуру вычислений.

Примеры решения задач

Задача 1.1. Найти погрешность вычисления функции, если $y = x_1 \pm \pm x_2$.

Решение. Имеем

$$\Delta y = x_1 \pm x_2 - (x_1^0 \pm x_2^0) = x_1 - x_1^0 \pm (x_2 - x_2^0) = \Delta x_1 \pm \Delta x_2.$$

Максимальная ошибка косвенных наблюдений равна

$$\Delta y_{\max} = |\Delta x_1| + |\Delta x_2|.$$

Задача 1.2. Найти погрешность вычисления функции, если $y = \dfrac{x_1}{x_2}$.

Решение. Имеем

$$\Delta y = \frac{x_1}{x_2} - \frac{x_1^0}{x_2^0} = \frac{x_1^0 + \Delta x_1}{x_2^0 + \Delta x_2} - \frac{x_1^0}{x_2^0} =$$

$$= \frac{x_2^0 \Delta x_1}{(x_2^0 + \Delta x_2)x_2^0} - \frac{x_1^0 \Delta x_2}{(x_2^0 + \Delta x_2)x_2^0} \approx \frac{1}{x_2^0}\Delta x_1 - \frac{x_1^0}{(x_2^0)^2}\Delta x_2.$$

Максимальная ошибка косвенных наблюдений равна

$$\Delta y_{\max} = \left|\frac{1}{x_2^0}\Delta x_1\right| + \left|\frac{x_1^0}{(x_2^0)^2}\Delta x_2\right|.$$

Задача 1.3. Найти погрешность вычисления функции, если $y = {x_1}^{x_2}$.

Решение. Разложение в ряд Тейлора приводит к формуле

$$\Delta y = x_2^0 (x_1^0)^{x_2^0 - 1}\Delta x_1 + (x_1^0)^{x_2^0} \ln(x_1^0)\Delta x_2.$$

Максимальная ошибка равна

$$\Delta y_{\max} = (x_1^0)^{x_2^0 - 1} |\, x_2^0 \Delta x_1 \,| + (x_1^0)^{x_2^0} |\ln(x_1^0)\Delta x_2 \,|.$$

Задачи для самостоятельного решения

1.1. Решить уравнение $\dfrac{p'(x)}{xp(x)} = -\dfrac{1}{\sigma^2}$.

1.2. Исследовать функцию $p(x) = c \cdot e^{-(x-\mathrm{a})^2}$ на максимум.

1.3. В какой точке функция $p(\hat{x}) = \alpha e^{-\frac{\sum_{i=1}^{n}(x_i - \hat{x})^2}{2\sigma^2}}$ имеет максимальной значение?

1.4. На весах с диапазоном работы 0…100 кг и классом точности 0,1 взвешивается груз. Определить точность взвешивания.

1.5. Электронные весы с диапазоном измерения 0…200 кг показывают вес человека с точностью 50 г. Найти класс точности весов.

1.6. Зная, что показания весов с диапазоном измерения 0…200 кг и классом точности 0,15 распределены по нормальному закону, указать, в каких пределах с вероятностью 0,95 находится вес человека, если весы показали 65 кг 300 г?

1.7. Проведены многократные измерения каждой из трех величин. Получены точечные оценки соответственно 2,371, 1,12 и 3,1. Найдена их сумма. Какое значение суммы является математически корректным: 6,591, 6,59 или 6,6?

1.8. При исследовании параметра y проведены однократные измерения величин x_1, x_2. Получены значения x_1^0, x_2^0 с погрешностями Δx_1, Δx_2 соответственно. Рассматриваемые величины связаны зависимостью $y = x_1^{x_2}$. Какой будет максимально возможная погрешность при нахождении величины y?

1.9. Величины x_1, x_2, x_3 связаны зависимостью $x_3 = \sin(x_1 \cdot x_2)$. Удалось измерить x_2 и x_3. Получились значения x_2^0, x_3^0 с погрешностями Δx_2, Δx_3 соответственно. Какой будет максимально возможная погрешность при нахождении величины x_1?

1.10. Каждая из величин x_1, x_2 измерена отдельным прибором. Получены результаты x_1^0 и x_2^0 с квадратическими отклонениями σ_1 и σ_2 соответственно. С какой точностью можно определить величину $x_1 \cdot x_2$?

1.11. Погрешность Δx_n, возникающая при проведении n измерений, отличается от теоретической погрешности Δx, обусловленной законом распределения. Квадрат погрешности для трех измерений оказался равным $\frac{4}{3}$, для 5 – $\frac{8}{5}$, для 10 – $\frac{18}{10}$. Какова ожидаемая погрешность измерений при проведении а) 100 наблюдений; б) n наблюдений, где $n \gg 100$?

1.12. В забеге на 100 м в первой группе лучшее время оказалось равным 10,00 с, во второй группе — 9,95 сек. Время измерялось секундомером с точностью $3\sigma = 0{,}03$ с. Победа была присуждена бегуну из второй группы. Какова вероятность того, что судьи ошиблись? Показания секундомера распределены нормально.

Глава 2
НЕСМЕЩЕННОСТЬ, СОСТОЯТЕЛЬНОСТЬ И ЭФФЕКТИВНОСТЬ ТОЧЕЧНЫХ ОЦЕНОК

В результате освоения материала данной главы студент должен:

знать

- основные понятия математической статистики;
- основные выборочные характеристики;
- требования, предъявляемые к точечным оценкам;
- теорему Рао — Фреше — Крамера;
- понятие асимптотических оценок;
- понятие количества информации и энтропии;
- теорему о минимуме информации Фишера, достигаемой на нормальном распределении;

уметь

- исследовать точечные оценки на несмещенность, состоятельность и эффективность;
- вычислять информационную энтропию случайной величины;

владеть

- навыками исследования точечных оценок.

2.1. Основные понятия математической статистики

Слово «статистика» происходит от лат. *status* — состояние дел. В науку термин «статистика» ввел в 1746 г. немецкий ученый Г. Ахенваль[1], предложив заменить название курса «Государствоведение», преподававшегося в университетах Германии, на «Статистика», положив тем самым начало развитию статистики как науки и учебной дисциплины.

Следует разделить предмет статистики и математической статистики.

Статистика — отрасль знаний, которая занимается сбором сведений, их сортировкой, группировкой, проводит изучение массива

[1] Ахенваль Готфрид (1719—1772) — знаменитый статистик, профессор Геттингенского университета.

полученных данных (статистическое исследование), находит полезную информацию для создания количественных и качественных выводов о случайной величине на основе данных массового наблюдения. Решая эти задачи, статистика в большинстве использует готовые привнесенные извне идеи, методы, формулы. Термин «статистика» употребляют также для простого обозначения набора количественных данных.

Про статистику Марк Твен в своем произведении «Главы моей автобиографии» написал: «Существуют три вида обмана: ложь, наглая ложь и статистика». Видимо, он имел ввиду тенденциозную подборку данных, подгонку под известный результат, создание общих выводов на основе нескольких примеров и т.д. Чтобы изъять статистику из этой триады, существует математическая статистика.

Математическая статистика — наука, разрабатывающая математические методы систематизации и использования статистических данных для научных и практических выводов. Она научно обосновывает методы и формулы, разрабатывает приемы статистического наблюдения и анализа статистических данных, которыми следует руководствоваться, ведя наблюдения и их обработку. Математическая статистика опирается на теорию вероятностей, дающую возможность оценить надежность и точность выводов, делаемых на основании ограниченного статистического материала, а также на математический анализ и линейную алгебру как инструменты исследования.

Среди задач, решаемых наукой математической статистикой, перечислим следующие:

1) задачи определения закона распределения случайной величины по статистическим данным;

2) оценка по выборочным характеристикам исследуемых параметров и функций от них;

3) задачи проверки статистических гипотез.

Познакомимся с основными понятиями, которыми оперирует математическая статистика.

Генеральной совокупностью называется выделенное множество объектов произвольной природы, обладающих признаками, доступными для наблюдения и количественного измерения. Объекты, входящие в генеральную совокупность, называются ее *элементами*, а их общее число — ее *объемом* N.

Например, генеральной совокупностью может быть население определенного города, а изучаемой случайной величиной — доход отдельного жителя. Для построения функции распределения дохода, нахождения среднего дохода и других характеристик желательно опросить каждого жителя города. Но для многотысячного или миллионного города это невозможно сделать. И тогда опрашивается некоторое количество жителей, т.е. делаются *наблюдения*: говоря языком математической статистики, из генеральной совокупности извлекаются случайным образом элементы $x_1, x_2, \ldots, x_n$.

Выборкой называется набор некоторого числа наблюдений из генеральной совокупности. Число сделанных наблюдений называется *объемом выборки*. Проводится статистический анализ выборки, по результатам которого определяется нужная нам числовая характеристика случайной величины. По выборке мы находим *выборочную,* или *эмпирическую,* характеристику, которая отличается от *генеральной,* или *теоретической*, характеристики, вычисляемой по всей генеральной совокупности. Задача статистики — построить наблюдения и провести анализ таким образом, чтобы вычисленная по ним *выборочная* характеристика была максимально близка *генеральной* характеристике.

Для этого должны быть выполнены некоторые условия.

1. Выборка должна *репрезентативной,* т.е. должно выполняться соответствие характеристик выборки характеристикам всей генеральной совокупности. Репрезентативность определяет, насколько возможно по выборке обобщить результаты исследования на всю генеральную совокупность.

2. Выборка должна быть *случайной,* т.е. все элементы генеральной совокупности имеют одинаковую вероятность попасть в выборку.

3. Выборка должна быть *массовой*. Имея в наблюдениях один-два элемента, невозможно сделать объективные выводы.

После того как выборка сделана, проводится первичная обработка наблюдений. Ряд наблюдений упорядочивается по возрастанию. Полученный ряд называется *вариационным рядом*. Некоторые наблюдения могут оказаться имеющими одинаковое значение. Вводится понятие варианты. *Вариантами* называют различные численные значения наблюдений.

Вариационный ряд представляется графически в виде полигона, гистограммы или графика накопленных частот.

Полигоном частот называется ломаная линия, отрезки которой соединяют точки (x_1, n_1), (x_2, n_2), …, (x_k, n_k), где x_i, $i = 1, 2, \dots, k$, — варианты (члены вариационного ряда); n_i — количества появлений наблюдения x_i в выборке. Полигон используется для изображения выборки в случае дискретных случайных величин.

Гистограмма представляет собой ступенчатую фигуру, состоящую из прямоугольников с основаниями определенной длины Δx, высоты которых равны относительным частотам $\frac{n_{\Delta x}}{n}$, где $n\Delta_x$ — число наблюдений, попавших в промежуток Δx.

При $n \to \infty$ и $\Delta x \to 0$ гистограмма сходится по вероятности в каждой точке x к кривой плотности распределения $p_\xi(x)$ случайной величины ξ, т.е. для любого $\varepsilon > 0$ и любого x

$$\lim_{\substack{n\to\infty \\ \Delta x\to 0}} P\left\{\left|\frac{n_{\Delta x}}{n} - p_\xi(x)\right| \geq \varepsilon\right\} = 0,$$

поэтому используется для изображения эмпирической плотности распределения в случае непрерывных случайных величин.