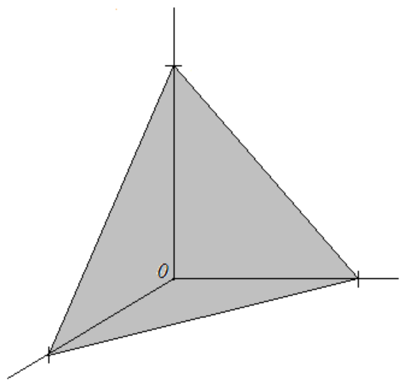
ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА



♦ ИЗДАТЕЛЬСТВО ТГТУ ♦

Министерство образования и науки Российской Федерации

ГОУ ВПО «Тамбовский государственный технический университет»

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания к выполнению контрольных работ
для студентов специальностей 080105, 080109, 080500

Тамбов
Издательство ТГТУ
2008

УДК 51:33
ББК В11я73
С606

Рекомендовано Редакционно-издательским советом ТГТУ

Рецензент

доктор экономических наук, профессор ТГТУ
Л.В. Пархоменко

С606 Теория вероятностей и математическая статистика : методические указания / сост. А.В. Солопахо. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2008. – 32 с. – 100 экз.

Рассмотрены методика и примеры решения типовых задач по темам «Вероятности случайных событий», «Теория случайных величин», «Оценка законов распределения», «Проверка статистических гипотез».

Предназначены для студентов специальностей 080105, 080109, 080500 всех форм обучения.

УДК 51:33
ББК В11я73

Учебное издание

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания

С о с т а в и т е л ь

СОЛОПАХО Александр Владимирович

Редактор Т.М. Глинкина
Компьютерное макетирование Т.Ю. Зотовой

Подписано в печать 04.07.2008
Формат $60 \times 84/16$. 1,86 усл. печ. л. Тираж 100 экз. Заказ № 313

Издательско-полиграфический центр ТГТУ
392000, Тамбов, Советская, 106, к. 14

ВВЕДЕНИЕ

Теория вероятностей – математическая дисциплина, объектом изучения которой являются случайные события. На теории вероятностей основывается *математическая статистика*, которую иногда считают даже частью теории вероятностей. Экономика и производственные процессы – одна из важнейших сфер применения теории вероятности и математической статистики.

Уже давно исследование и прогнозирование экономических явлений трудно представить без использования методик статистического оценивания, проверки гипотез, регрессионного анализа, трендовых и сглаживающих эконометрических моделей и других методов, опирающихся на теорию вероятностей. А с развитием общества мировая экономика все более усложняется и, следовательно, по законам развития динамических систем должен усиливаться статистический характер законов, описывающих социально-экономические явления.

Все это предопределяет необходимость овладения методами теории вероятностей и математической статистики как важнейшим инструментом анализа и прогнозирования экономических явлений и процессов.

Данное пособие содержит подборку примеров решения наиболее важных и типичных по содержанию задач, методические рекомендации по их решению, и изучению рассматриваемой дисциплины в целом. Пособие не имеет целью изложение теории, а содержит лишь вспомогательный материал, способствующий самостоятельному выполнению контрольных работ студентами, в особенности, заочного отделения. Последовательность рассмотрения материала, в основном, соответствует имеющемуся по данной дисциплине учебному пособию [1]. Необходимый теоретический материал можно найти и в других источниках [2, 4, 5] и т.д.

1. ИСЧИСЛЕНИЕ ВЕРОЯТНОСТЕЙ СЛУЧАЙНЫХ СОБЫТИЙ

Как и всякая математическая дисциплина, Теория вероятностей и математическая статистика является прежде всего стройной теорией, состоящей из определений, теорем и различных других абстрактных построений. Решение задач при ее изучении имеет целью проиллюстрировать соответствующие моменты теории. Однако, вероятностные и статистические примеры и задачи являются несомненно наиболее практически содержательными, часто находящими непосредственное продолжение в реальных ситуациях.

Следует отметить, что при решении задач теории вероятностей, особенно относящихся к рассматриваемым начальным, фундаментальным вопросам, интерес представляет не ответ, а именно процесс решения задачи. Ответ (результат), зачастую, интуитивно очевиден. Важно, формально строго, опираясь на известные в теории определения, теоремы и т.д., выстроить и четко описать схему решения задачи. В истории известны примеры, когда из-за неправильной схемы рассуждений весьма одаренные и известные математики допускали в вероятностных задачах грубейшие ошибки.

1.1. Общие понятия о случайном событии и его вероятности. Действия над случайными событиями

Теоретический материал, соответствующий данному вопросу, рассмотрен в [1, п. 1.1], а также в [2, п. 1.1, 1.2; 4, п. 1.1].

Любое случайное событие всегда связано с конкретным вероятностным опытом и не может рассматриваться само по себе, в отрыве от этого опыта. Это означает, что рассматривая случайное событие, необходимо четко определить, в чем заключается вероятностный опыт, каковы его возможные исходы, действительно ли данное событие следует изучать как результат этого опыта, и т.д.

***Пример* 1.1**. Определить множество исходов стохастического эксперимента, состоящего в сдаче студентом экзамена.

Решение. Очевидно, к этому множеству относятся исходы:

- получение оценки «отлично»;
- получение оценки «хорошо»;
- получение оценки «удовлетворительно»;
- получение оценки «неудовлетворительно».

Однако, в зависимости от ситуации, может оказаться целесообразным включение таких исходов, как

- опоздание студента на экзамен;
- неявка преподавателя, например по причине болезни
- и т.д.

Понятие случайного события в теории вероятностей *всегда* можно ассоциировать с понятием множества. Поэтому для понимания действий над случайными событиями следует разобраться с действиями над множествами. При этом целесообразно помнить, что теория множеств является самостоятельным разделом высшей математики [6, п. 1.1], а никак не частью теории вероятностей.

***Пример* 1.2**. Показать, что события A, $B - A$ и $\left(\overline{A+B}\right)$ несовместны, а их сумма составляет достоверное событие Ω.

Решение. Докажем несовместность каждой пары этих событий.

В соответствии с определением операции вычитания множеств события A и $B - A$, очевидно, являются несовместными, то есть

$$A\left(B-A\right)=\varnothing.$$

Далее, если некоторый элемент ω множества Ω таков, что $\omega \in A$, то $\omega\in(A + B)$, и значит $\omega\notin \left(\overline{A+B}\right)$, а тогда и

$$A\ \left(\overline{A+B}\right)=\varnothing.$$

Наконец, если $\omega \in (B-A)$, то $\omega \in (A+B)$ и $\omega \notin \left(\overline{A+B}\right)$, а значит

$$(B-A)\left(\overline{A+B}\right)=\varnothing,$$

т.е. указанные три события попарно несовместны.

Окончательно

$$A+(B-A)+\left(\overline{A+B}\right)=A+B+\left(\overline{A+B}\right)=\Omega,$$

т.е. эти события в сумме составляют Ω.

1.2. Схема с равновозможными исходами. Классическое определение вероятности

Теоретический материал, соответствующий данному вопросу, рассмотрен в [1, п. 1.2], а также в [2, п. 1.3; 4, п. 1.2].

Любому стохастическому эксперименту можно поставить в соответствие одну из трех возможных схем его проведения (см. также [1, п. 1.4, 1.5]). Схема с равновозможными исходами является простейшей из них. В реальной жизни немного экспериментов соответствует ей. Однако почти все азартные игры являются такими экспериментами.

Расчеты вероятностей событий, связанных с экспериментом, соответствующим схеме с равновозможными исходами, основаны на так называемом классическом определении вероятностей, которое имеет очень большую теоретическую и практическую значимость. Исходя из этого определения, если известно, что

- эксперимент соответствует данной схеме;
- заданы благоприятствующие исходы некоторых событий,

то можно формально строго находить и вероятности случайных событий, являющиеся результатом арифметических действий (операций) над этими событиями.

***Пример* 1.3.** Пусть в закрытой урне находится 20 пронумерованных шаров одинакового размера. Найти вероятности событий, что наудачу вытащенный шар имеет

- четный номер;
- номер больший, чем 11,

а также вероятности суммы, разности и произведения этих событий.

Решение. Обозначим: A – шар имеет четный номер; B – номер шара больше 11.

Элементарные исходы опыта обозначим через w_i – наудачу вытащенный шар имеет номер i, $i=\overline{1,20}$.

Ясно, что рассматриваемым событиям благоприятствуют следующие исходы:

$$A=\{w_2, w_4, ..., w_{20}\},\ B=\{w_{12}, w_{13}, ..., w_{20}\}.$$

По классическому определению вероятности

$$P(A)=\frac{10}{20}=\frac{1}{2},\ P(B)=\frac{9}{20}.$$

Далее имеем:

$$A+B=\{w_2, w_4, ..., w_{12}, w_{13}, ...w_{20}\} \Rightarrow P(A+B)=\frac{14}{20}=\frac{7}{10};$$

$$A-B=\{w_2, w_4, ..., w_{10}\} \Rightarrow P(A-B)=\frac{5}{20}=\frac{1}{4};$$

$$A\cdot B=\{w_{12}, w_{14}, ..., w_{20}\} \Rightarrow P(A\cdot B)=\frac{5}{20}=\frac{1}{4}.$$

1.3. Использование комбинаторных формул

Комбинаторика – ветвь математики, изучающая комбинации и перестановки предметов. Ее понятия используются при решении почти всех задач на схему с равновозможными исходами [1, п. 1.3; 2, п. 1.4; 4, п. 1.5].

***Пример* 1.4.** Наудачу взятый телефонный номер без нуля впереди состоит из 7 цифр. Найти вероятность того, что все цифры различны.

Решение. Число всех различных номеров по правилу произведения равно $n=9\cdot 10^6$, число номеров с различными цифрами равно $m=9\cdot A_9^6$. Согласно классической схеме вероятностей, искомая вероятность равна

$$P(A)=\frac{m}{n}=\frac{9\cdot A_9^6}{9\cdot 10^6}=\frac{9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4}{10^6}=0{,}06048.$$

$$P(A)=1/24.$$

Важно также разобраться с так называемой «урновой» схемой проведения эксперимента (см. пример 1.5).

***Пример* 1.5.** Пусть имеется 500 электрических лампочек. Завод-изготовитель гарантирует, что из них не более 2 бракованных. Оценить, какова вероятность, что из 5 выбранных лампочек нет ни одной бракованной.

Решение. Эксперимент соответствует «урновой» схеме. Используя соответствующую формулу [1, 8], при $N = 500$, $n = 5$, $M = 2$, $m = 0$, получаем

$$P = \frac{C_{498}^{5} C_{2}^{0}}{C_{500}^{5}} \leq 0{,}98.$$

1.4. Схема с неравновозможными исходами. Статистическое определение вероятности

Теоретический материал, соответствующий данному вопросу, рассмотрен в [1, п. 1.4], а также в [2, п. 1.7; 4, п. 1.3].

Большинство реальных вероятностных ситуаций, имеющих конечное (или счетное) количество элементарных исходов, соответствует именно схеме с неравновозможными исходами. Важным фактом, использующимся при решении многих соответствующих задач, является статистическое определение вероятности.

***Пример* 1.6.** Имеется статистика опозданий на лекции некоторого студента.

Время опоздания, мин.	Без опозданий	1–2	2 – 4	4 – 10	Более 10
Количество опозданий	30	20	10	5	5

Оценить вероятность, что на очередную лекцию студент опоздает более чем на 4 мин.

Решение. Рассмотрим ситуацию, как эксперимент с пятью неравновозможными элементарными исходами. Будем считать, что их вероятности, в соответствии со статистическим определением и имеющимися данными, можно принять равными

$$p_1 = \frac{30}{70},\ \ p_2 = \frac{20}{70},\ \ p_3 = \frac{10}{70},\ \ p_4 = \frac{5}{70},\ \ p_5 = \frac{5}{70}.$$

Тогда для искомой вероятности имеем

$$P(A) = p_4 + p_5 = \frac{10}{7} = \frac{1}{7}.$$

1.5. Схема с несчетным множеством исходов. Геометрическое определение вероятности

Соответствующий теоретический материал рассмотрен в [1, п. 1.5], а также в [2, п. 1.8; 4, п. 1.4].

Схема с нечетным множеством исходов является наиболее общей. Все изученные выше схемы проведения стохастических экспериментов можно рассмотреть как частный случай этой схемы. Это значит, что и любую реальную ситуацию можно рассмотреть в рамках данной схемы, поэтому очень важно хорошо уяснить ее *содержание*.

***Пример* 1.7.** Две точки случайным образом помещаются на отрезок длины l. Все положения точек при этом считаются равновозможными. Найти вероятность, что расстояние между точками окажется меньше, чем расстояние от любого конца отрезка до ближайшей точки.

Решение. На первый взгляд задача кажется одномерной, но на самом деле следует перейти к рассмотрению двухмерных множеств. Обозначим через x координату точки A, а через y – точки B. Для определенности будем считать, что A левее, чем B. Множество положений точек, соответствующих условию задачи, описывается неравенством

$$(y - x) < \min\{x,\, l - y\}.$$

Чертеж этого множества представлен на рис. 1.1: это выделенный треугольник.

Исходя из геометрического определения вероятности, находим ответ

$$P(A) = \frac{\frac{l}{\sqrt{2}}\frac{l}{\sqrt{2}}\operatorname{tg}(15^\circ)}{\frac{l^2}{2}} = \operatorname{tg}(15^\circ) \approx 0{,}268\ .$$

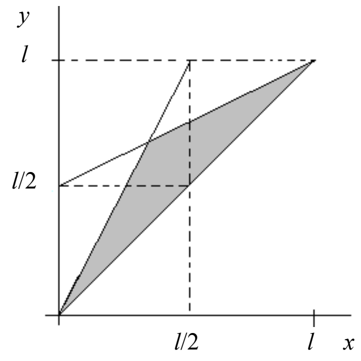


Рис. 1.1

1.6. Теоремы сложения и умножения вероятностей

Необходимый теоретический материал рассмотрен в [1, п. 1.6].

Соответствующие формулы являются основой решения очень широкого спектра задач.

***Пример* 1.8.** В лотерее на каждые 10 000 билетов разыгрываются 150 вещевых и 50 денежных выигрышей. Чему равна вероятность выигрыша, безразлично денежного или вещевого?

Решение. По классическому определению, вероятность вещевого выигрыша составляет

$$P(A)=\frac{150}{10\,000}=0{,}015\,,$$

аналогично – денежного

$$P(B)=\frac{50}{10\,000}=0{,}005\,.$$

События A и B, очевидно, несовместны. Тогда по теореме сложения вероятностей окончательно получаем

$$P(C)=P(A)+P(B)=0{,}015+0{,}005=0{,}02\,.$$

***Пример* 1.9.** Имеются два бизнес-проекта. Вероятность получения прибыли для первого бизнес-проекта равна 0,8, а для второго – 0,7. Найти общую вероятность получения прибыли.

Решение. Обозначим через A – успех первого проекта; B – успех второго.

Будем считать, что события A и B независимы. Тогда по теореме умножения вероятностей

$$P(w_1)=P(A\cdot B)=0{,}8\cdot 0{,}7=0{,}56\,;$$
$$P(w_2)=P(\overline{A}\cdot B)=(1-0{,}8)\cdot 0{,}7=0{,}14\,;$$
$$P(w_3)=P(A\cdot \overline{B})=0{,}8\cdot(1-0{,}7)=0{,}24\,;$$
$$P(w_4)=P(\overline{A}\cdot \overline{B})=(1-0{,}8)(1-0{,}7)=0{,}06\,.$$

Можно считать, что мы перешли к рассмотрению вероятностного эксперимента с исходами $w_1,\ w_2,\ w_3,\ w_4$.

Тогда получаем

$$P(C)=P(w_1)+P(w_2)+P(w_3)=0{,}94\,.$$

1.7. Формулы условной вероятности, полной вероятности. Формула Байеса

Необходимый теоретический материал рассмотрен в [1, п. 1.7] и др.

***Пример* 1.10.** На предприятии изготавливают некоторые изделия на трех поточных линиях. На первой производят 20 % всех изделий, на второй – 30 %, на третьей – 50 %. Каждая линия характеризуется следующими процентами годности изделия: 95, 98, 97 %. Требуется определить вероятности того, что:

1. Взятое наугад изделие окажется бракованным.
2. Бракованное изделие изготовлено на 1, 2, 3-й линиях.

Решение. Обозначим: A_1, A_2, A_3 – изделие изготовлено на 1, 2, 3-й линиях. Согласно условию:

$$P(A_1)=0{,}2;\quad P(A_2)=0{,}3;\quad P(A_3)=0{,}5.$$

Обозначим: B – взятое наугад изделие оказалось бракованным. Согласно условию

$$P(B/A_1)=0{,}05;\quad P(B/A_2)=0{,}02;\quad P(B/A_3)=0{,}03.$$

Используя формулу полной вероятности, находим

$$P(B)=P(B/A_1)\,P(A_1)+P(B/A_2)\,P(A_2)+P(B/A_3)\,P(A_3)=$$
$$=0{,}2\cdot 0{,}05+0{,}3\cdot 0{,}02+0{,}5\cdot 0{,}03=0{,}031.$$

Отметим, что последняя величина фактически выражает общий уровень брака по предприятию.
Вероятность того, что взятое бракованное изделие изготовлено на той или иной линии, находим по формуле Байеса:

$$P(A_1/B)=\frac{P(A_1)P(B/A_1)}{P(B)}=\frac{0{,}02\cdot 0{,}05}{0{,}031}=\frac{10}{31};$$

$$P(A_2/B)=\frac{P(A_2)P(B/A_2)}{P(B)}=\frac{0{,}006}{0{,}031}=\frac{6}{31};$$

$$P(A_3/B)=\frac{P(A_3)P(B/A_3)}{P(B)}=\frac{0{,}015}{0{,}031}=\frac{15}{31}.$$

1.8. Последовательности испытаний. Схема Бернулли

Необходимый теоретический материал рассмотрен в [1, п. 1.9] и др.

Данная схема проведения вероятностного эксперимента, являясь частным случаем схемы с неравновозможными исходами, очень важна для исследований всевозможных массовых явлений, состоящих из серии одинаковых независимых опытов. Примеры таких явлений в изобилии можно найти в сферах экономики, социологии и т.п. Начнем с наиболее простых примеров.

***Пример* 1.11.** Монету бросают 6 раз. Найти вероятность того, что герб выпадет: а) менее двух раз; б) не менее двух раз.

Решение. Ясно, что данный эксперимент можно считать соответствующим схеме Бернулли при $n=6$. Тогда

а) $P(A)=P_6(0)+P_6(1)=C_6^0\cdot\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^6+C_6^1\cdot\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^5=$

$$=(1+6)\cdot\left(\frac{1}{2}\right)^6=\frac{7}{64};$$

б) $P(\overline{A})=1-P(A)=1-\frac{7}{64}=\frac{57}{64}$.

***Пример* 1.12**. Перерасход кассовых средств в каждый отдельный день возможен с одной и той же вероятностью p. Какова должна быть эта вероятность, чтобы с вероятностью не менее 0,9 перерасход происходил не более чем один раз в пять рабочих дней?

Решение. Аналогично предыдущей задачи, вероятность не более чем одного перерасхода за пять дней

$$P(A)=P_5(0)+P_5(1)=C_5^0\cdot p^0\cdot(1-p)^5+C_5^1\cdot p\cdot(1-p)^4=$$

$$=(1-p+6\cdot p)\cdot(1-p)^4<0{,}1.$$

Решаем уравнение

$$(6-5\cdot q)\cdot q^4=0{,}1$$

графически, получаем

$$q\approx 0{,}3973\ \Rightarrow\ p=1-q=0{,}6027.$$

Таким образом, вероятность перерасхода в день не должна превышать $p=0{,}6027$.

1.9. Локальная и интегральная теоремы Муавра-Лапласа

Необходимый теоретический материал рассмотрен в [1, п. 1.10] и др.

При большом количестве опытов n, что весьма естественно для рассматриваемых массовых процессов, непосредственное использование формулы Бернулли становится практически невозможным. Поэтому широко используются упрощенные приближенные соотношения.

***Пример* 1.13**. Вероятность того, что покупатель, зашедший в магазин, совершит покупку, равна 0,2. Найти вероятность того, что из 400 покупателей покупку осуществят от 70 до 100.

Решение. По условию $n=400$, $m_1=70$, $m_2=100$, $q=0{,}8$, $p=0{,}2$.

Условие применимости интегральной формулы Муавра-Лапласа

$$npq>15-20$$

существенно перевыполняется. Тогда

$$x_1=\frac{70-400\cdot 0{,}2}{\sqrt{400\cdot 0{,}2\cdot 0{,}8}}=-1{,}25,\ x_2=2{,}5.$$

Используя таблицы, получаем

$$P_{400}(70,\,100)\approx\Phi(2{,}5)-\Phi(-1{,}25)=\Phi(2{,}5)+\Phi(1{,}25)\approx$$

$$\approx 0{,}4938 + 0{,}3944 = 0{,}8882.$$

Разнообразные интересные задачи решаются с помощью формулы

$$\gamma = P\left\{\left|\frac{m}{n} - p\right| \le \varepsilon\right\} \approx 2\cdot\Phi\left(\varepsilon\cdot\sqrt{\frac{n}{p\cdot q}}\right),$$

которая следует из интегральной теоремы. В частности, она фактически позволяет решать задачи уже *математической статистики* по оценке различных параметров, например вероятностей событий.

***Пример* 1.14**. Было произведено 500 опытов, в которых 300 раз произошло интересующее нас событие. В каких пределах с вероятностью 0,9 лежит истинная вероятность p этого события.

Решение. В соответствии с формулой имеем, что должно быть

$$2\cdot\Phi\left(\varepsilon\cdot\sqrt{\frac{500}{p\cdot q}}\right) \ge 0{,}9,$$

тогда из таблиц значений интегральной функции Лапласа следует, что

$$\varepsilon\cdot\sqrt{\frac{500}{p\cdot q}} \ge 1{,}645.$$

В качестве центра искомого интервала, в котором лежит неизвестная истинная вероятность p, естественно принять наблюдаемую частоту

$$w = \frac{300}{500} = \frac{3}{5} = 0{,}6.$$

Тогда должно быть

$$\varepsilon\cdot\sqrt{\frac{500}{(0{,}6-\varepsilon)\cdot(1-(0{,}6-\varepsilon))}} \ge 1{,}645.$$

Решим соответствующее уравнение графически: $\varepsilon = 0{,}0365$.

Таким образом, с вероятностью 0,9 истинная вероятность p лежит в пределах $0{,}6 \pm 0{,}0365$.

2. ОСНОВЫ ТЕОРИИ СЛУЧАЙНЫХ ВЕЛИЧИН

Теория случайных величин (с. в.) является наиболее практически важным разделом теории вероятностей. В частности потому, что в этом разделе вводятся многие важнейшие понятия, необходимые для математической статистики.

2.1. Определение случайной величины. Задание дискретной случайной величины

Необходимый теоретический материал рассмотрен в [1, п. 2.1] и др.

В природе дискретные с.в. встречаются относительно реже непрерывных, однако можно привести целый ряд важных примеров их использования.

***Пример* 2.1**. В цеху имеются 8 токарных станков, использующихся с одинаковой интенсивностью. Вероятность, что в данный момент используется тот или другой станок одинакова и равна 0,4. Построить ряд распределения случайной величины числа использующихся в данный момент станков.

Решение. Будем считать, что станки используются независимо друг от друга. Тогда ситуация соответствует схеме Бернулли, и вероятность того или иного числа использующихся станков можно найти по формуле Бернулли. Получим следующий ряд распределения.

k_i	0	1	2	3	4	5	6	7	8
p_i	0,016796	0,08958	0,209019	0,278692	0,232243	0,123863	0,041288	0,007864	0,000655

Распределение случайной величины k числа успехов в схеме Бернулли называется *биномиальным*. Оно нередко встречается в различных задачах.

Часто целесообразнее говорить об относительной частоте числа успехов $\frac{k}{n}$.

Такая с. в. называется дробно-биномиальной.

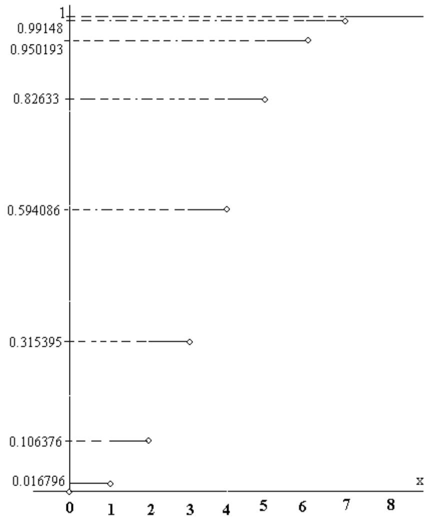


Рис. 2.1

2.2. Непрерывная случайная величина. Функция распределения

Необходимый теоретический материал рассмотрен в [1, п. 2.2] и др.

***Пример* 2.2.**

1) Построим график функции распределения с. в. из предыдущего примера (см. рис. 2.1). Отметим, что график построен в точной пропорции вероятностям, и поэтому предпоследней ступеньки не видно (она сливается с единицей);

2) С. в. X задана функцией распределения

$$F(x)=\begin{cases} 0 \text{ при } x\le -1; \\ \dfrac{x+1}{3} \text{ при } -1<x\le 2; \\ 1 \text{ при } x>2. \end{cases}$$

Найти вероятность того, что в результате испытания X примет значение, заключенное в интервале (0, 1).

Решение. В соответствии с определением функции распределения имеем:

$$P\{x\in(0,\,1)\}=F(1)-F(0)=\frac{1}{3}.$$

В данном случае мы имеем абсолютно непрерывную с. в. Если бы это было не так, то могло бы оказаться необходимым рассмотреть односторонние пределы.

2.3. Функция плотности распределения случайной величины

Необходимый теоретический материал рассмотрен в [1, п. 2.3] и др.

Понятие функции плотности распределения с.в. является важнейшим во всей теории. Без его усвоения невозможно изучение последующих фактов, в частности методик математической статистики.

***Пример* 2.3**. Функция плотности распределения с. в. X имеет следующий кусочно-аналитический вид:

$$P(x)=\begin{cases} 0,\ x<-c; \\ x+c,\ -c\le x<0; \\ -x+c,\ 0\le x\le c; \\ 0,\ x>c. \end{cases}$$

Найти:

1) при каком значении c это возможно;

2) функцию распределения $F(x)$ с. в. X;

3) вероятность $P\{x\in[-0{,}5;\,1{,}2]\}$.

Построить графики $P(x)$ и $F(x)$.

Решение. Построим график плотности (см. рис. 2.2). Область, площадь которой соответствует искомой вероятности, выделена более темным цветом.

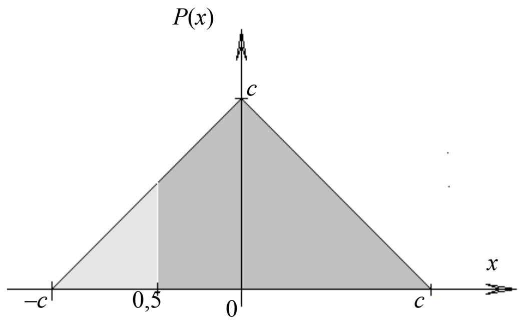


Рис. 2.2

Далее находим:

1. Используем одно из важнейших необходимых свойств плотности:

$$\int_{-\infty}^{\infty} P(x)dx = 1 .$$

В данном случае и без интегрирования, из простейших геометрических соображений, ясно, что площадь под графиком указанной на рис. 2.2 плотности составляет

$$\int_{-\infty}^{\infty} P(x)dx = c^2 ,$$

а это означает, что $c = 1$, т.е. плотность должна иметь вид:

$$P(x) = \begin{cases} 0, & x < -1; \\ x+1, & -1 \le x < 0; \\ -x+1, & 0 \le x \le 1; \\ 0, & x > 1. \end{cases}$$

2. В соответствии со свойствами $F(x)$ и $P(x)$, интегрируя, получаем

$$F(x) = \int_{-\infty}^{x} P(t)dt = \begin{cases} 0, \ x < -1; \\ \int_{-1}^{x} (t+1)dt, \ -1 \le x < 0; \\ \frac{1}{2} + \int_{0}^{x} (-t+1)dt, \ 0 \le x \le 1; \\ 1, \ x > 1. \end{cases} = \begin{cases} 0, \ x < -1; \\ \frac{(x+1)^2}{2}, \ -1 \le x < 0; \\ 1 - \frac{(x-1)^2}{2}, \ 0 \le x \le 1; \\ 1, \ x > 1. \end{cases}$$

3. В соответствии с определением плотности имеем

$$P\{x \in [-0,5; 1,2]\} = \int_{-0,5}^{1,2} P(x)dx = \int_{-0,5}^{0} (x+1)dx + \int_{0}^{1} -(x+1)dx + \int_{1}^{1,2} 0 \cdot dx =$$

$$= \frac{(x+1)^2}{2}\Bigg|_{-0,5}^{0} + \frac{1}{2} + 0 = \frac{3}{8} + \frac{1}{2} + 0 = \frac{7}{8}.$$

Заметим, что поскольку в данном случае известна функция распределения, то для нахождения вероятности проще было бы использовать именно ее, тогда

$$P\{x \in [-0,5; 1,2]\} = F(1,2) - F(-0,5) = 1 - \frac{1}{8} = \frac{7}{8}.$$

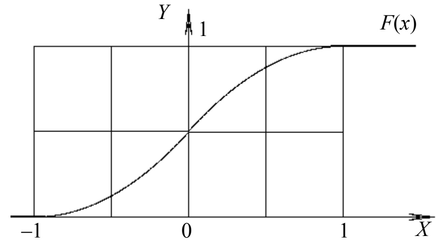


Рис. 2.3

В завершение строим график функции распределения (рис. 2.3).

***Пример* 2.4**. Известно, что $X \sim N(3,\ 4)$, т.е. с. в. X имеет нормальное распределение с параметрами $M(X)=3,\ \ D(X)=4$. Найти

$$P\{x \in [-0{,}5;\ 2{,}5]\}.$$

Решение. Нормальное распределение является наиболее практически важным распределением. Поэтому необходимо уметь работать с ним.

В соответствии с определением нормальной плотности имеем

$$P\{x \in [-0{,}5;\ 2{,}5]\} = \frac{1}{\sqrt{2\pi}\cdot 2}\int_{-0{,}5}^{2{,}5} e^{-\frac{(x-3)^2}{2\cdot 4}}\,dx = \ldots,$$

но этот интеграл не выражается аналитически. Однако, сделав замену

$$z = \frac{x-a}{\sigma} = \frac{x-3}{2},$$

получаем

$$\ldots = \frac{1}{\sqrt{2\pi}}\int_{\frac{-0{,}5-3}{2}}^{\frac{2{,}5-3}{2}} e^{-z^2}\,dz = \Phi(t)\Bigg|_{-\frac{7}{4}}^{-\frac{1}{4}} = \ldots,$$

где $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_0^t e^{-z^2}\,dz$ – так называемая интегральная функция Лапласа, таблицы значений которой приводятся во многих справочниках и учебниках. При их использовании следует помнить о нечетности этой функции.

В данном случае получаем

$$\ldots = \Phi\left(-\frac{1}{4}\right) - \Phi\left(-\frac{7}{4}\right) = -\Phi\left(\frac{1}{4}\right) + \Phi\left(\frac{7}{4}\right) \approx -0{,}09871 + 0{,}45994 = 0{,}36123.$$

2.4. Математическое ожидание и дисперсия случайной величины

Почти вся математическая статистика занимается оценкой и анализом всего двух параметров: математического ожидания и дисперсии. Из этого очевидна важность данных понятий.

Необходимый теоретический материал рассмотрен в [1, п. 2.4 – 2.5] и др.

***Пример* 2.5**. Найти математическое ожидание и дисперсию биномиально распределенной с. в. X.

Решение. Проще это сделать следующим образом. Представим X как сумму

$$X = Z_1 + Z_2 + \ldots + Z_n$$

одинаковых независимых с. в. Z_i, каждая из которых принимает значение 1 с вероятностью p и значение 0 с вероятностью $q=(1-p)$. Тогда очевидно,

$$M(Z_i) = 1\cdot p + 0\cdot(1-p) = p$$

и

$$D(Z_i) = p - p^2 = p\cdot q.$$

Используя свойства математического ожидания и дисперсии, без труда получаем:

$$M(X) = M(Z_1 + Z_2 + \ldots + Z_n) = M(Z_1) + M(Z_2) + \ldots + M(Z_n) = n\cdot p;$$

$$D(X) = D(Z_1 + Z_2 + \ldots + Z_n) = D(Z_1) + M(Z_2) + \ldots + D(Z_n) = n\cdot p\cdot q.$$

***Пример* 2.6**. Найти математическое ожидание и дисперсию с. в. из примера 2.3.

Решение. Вспоминая содержательный смысл параметра «математическое ожидание», делаем очевидный (из вида соответствующей плотности распределения) вывод, что в данном случае оно равно нулю. Поэтому рассчитаем только дисперсию. Сначала находим

$$M(X^2) = \int_{-\infty}^{\infty} x^2\cdot P(x)dx = \int_{-1}^{0} x^2\cdot(x+1)dx + \int_{0}^{1} x^2\cdot(-x+1)dx =$$

$$= \left(\frac{x^4}{4} + \frac{x^3}{3}\right)\Bigg|_{-1}^{0} + \left(-\frac{x^4}{4} + \frac{x^3}{3}\right)\Bigg|_{0}^{1} = \frac{7}{12} + \frac{1}{12} = \frac{2}{3}.$$

Окончательно

$$D(X)=M\left(X^2\right)-[M(X)]^2=\frac{2}{3}-0^2=\frac{2}{3}.$$

2.5. Независимость случайной величины и коэффициент корреляции

Выявление зависимости между различными экономическими факторами и измерение силы этой зависимости является важным инструментом самых разных экономических исследований. Необходимый теоретический материал рассмотрен в [1, п. 2.6] и др.

***Пример* 2.7**. Найдем коэффициент корреляции между с. в. X и $Y=X^2$, если область возможных значений с. в. X представляет собой отрезок $\left[-1,1\right]$.

Решение. В соответствии с формулой коэффициента корреляции имеем

$$r_{XY}=\frac{M\left((X-M(X))\left(X^2-M\left(X^2\right)\right)\right)}{\sqrt{D(X)\cdot D\left(X^2\right)}}=$$

$$=\frac{M\left(X^3-X^2\cdot M(X)-X\cdot M\left(X^2\right)+M(X)\cdot M\left(X^2\right)\right)}{\sqrt{D(X)\cdot D\left(X^2\right)}}=$$

$$=\frac{M\left(X^3\right)-M\left(X^2\right)\cdot M(X)-M(X)\cdot M\left(X^2\right)+M(X)\cdot M\left(X^2\right)}{\sqrt{D(X)\cdot D\left(X^2\right)}}=$$

$$=\frac{M\left(X^3\right)-M\left(X^2\right)\cdot M(X)}{\sqrt{D(X)\cdot D\left(X^2\right)}}.$$

Но при указанной области значений с. в. X легко доказать, что $M\left(X\right)=0$ и $M\left(X^3\right)=0$, а тогда и $r_{XY}=0$, в то время, как очевидно, что с. в. X и Y полностью взаимозависимы. Данный пример показывает ограниченность возможностей коэффициента корреляции как меры взаимозависимости между с. в.

2.6. Лемма Маркова и неравенство Чебышева

Необходимый теоретический материал рассмотрен в [1, п. 2.7] и др.

Особенностью соответствующих неравенств является то, что они позволяют оценивать (хотя зачастую достаточно грубо) вероятности тех или иных значений с. в., имея минимум информации об этих с. в.

***Пример* 2.8**. Среднее число покупателей в магазине за день составляет 300 человек. Оценить вероятность того, что в очередной день их число превысит 500 человек.

Решение. В данном случае неизвестен ни вид закона распределения этой с. в., ни ее параметры, кроме математического ожидания $M\left(X\right)=300$.

Кроме того, ясно, что это неотрицательная с. в. Применяя неравенство Маркова, имеем

$$P\{\,x>500\,\}\le\frac{300}{500}=\frac{3}{5}.$$

***Пример* 2.9**. Вероятность попадания по мишени в каждом отдельном выстреле $p=0{,}7$. Оценить вероятность, что из 100 выстрелов в цель попадут от 50 до 70.

Решение. Если считать, что результаты выстрелов независимы, то количество попаданий это биномиально распределенная с. в. Для этого распределения известно, что

$$M\left(X\right)=n\cdot p=100\cdot 0{,}7=70\,,\;D\left(X\right)=n\cdot p\cdot q=100\cdot 0{,}7\cdot 0{,}3=21\,.$$

Тогда по неравенству Чебышева имеем

$$P\{\,60\le x\le 80\,\}=P\{|x-70|\le 10\,\}=P\{|x-M(X)|\le\varepsilon\,\}>1-\frac{D(X)}{\varepsilon^2}=$$

$$=1-\frac{21}{100}=0{,}79.$$

Интересно решить ту же задачу с помощью интегральной теоремы Лапласа.

2.7. Многомерные случайные величины

Как известно, практическая важность теории с. в. основана на том, что с. в. является *математической моделью* параметров реальных процессов и объектов. Последние редко характеризуются одним отдельным параметром, а сразу некоторым их набором. Например, случайным образом отобранное коммерческое предприятие, очевидно, в той или иной мере полно может быть охарактеризовано только, если имеются данные о его прибыли, обороте, количестве работников и т.д. и т.п. Отсюда естественным образом возникают многомерные с. в.

Изучение любых математических многомерных объектов всегда сопряжено с той трудностью, что невозможно дать соответствующим фактам геометрических иллюстраций. Поэтому всегда стараются основные факты рассмотреть на частном случае одномерных объектов. То же справедливо и для теории с.в. Однако имеется ряд моментов, которые невозможно проиллюстрировать на одномерных с.в.

Необходимый теоретический материал рассмотрен в [1, п. 2.9] и др.

***Пример* 2.10**. Имеется двухмерная дискретная с. в. $Z=(X,\ Y)$, которая задана рядом распределения

x_i/y_i	$y_1=-3$	$y_2=-1$	$y_3=0$
$x_1=2$	0,1	0,15	0,2
$x_2=4$	0,05	q	0,25

При каком q это возможно? Рассчитать ковариационную матрицу этой с. в.

Решение. Поскольку сумма вероятностей всех возможных пар значений этой с. в. должна быть равна единице, то

$$q=1-(0{,}1+0{,}15+0{,}2+0{,}05+0{,}25)=0{,}25\ .$$

Далее находим

$$M(X)=2\cdot(0{,}1+0{,}15+0{,}2)+4\cdot(0{,}05+0{,}25+0{,}25)=3{,}1$$

и

$$M(Y)=(-3)\cdot(0{,}1+0{,}05)+(-1)\cdot(0{,}15+0{,}25)+0\cdot(0{,}2+0{,}25)=-0{,}85\ .$$

Тогда

$$D(X)=2^2\cdot 0{,}45+4^2\cdot 0{,}55-3{,}1^2=0{,}99\ ,$$

$$D(Y)=(-3)^2\cdot 0{,}15+(-1)^2\cdot 0{,}4+0^2\cdot 0{,}45-0{,}85^2=1{,}0275\ ,$$

$$\begin{aligned}\operatorname{cov}(X,Y)&=M((X-M(X))(Y-M(Y)))=\\&=0{,}1\cdot(2-3{,}1)\cdot(-3-(-0{,}85))+0{,}15\cdot(2-3{,}1)\cdot(-1-(-0{,}85))+\\&+0{,}2\cdot(2-3{,}1)\cdot(0-(-0{,}85))+0{,}05\cdot(4-3{,}1)\cdot(-3-(-0{,}85))+\\&+0{,}25\cdot(4-3{,}1)\cdot(-1-(-0{,}85))+0{,}25\cdot(4-3{,}1)\cdot(0-(-0{,}85))=0{,}135.\end{aligned}$$

Таким образом, искомая матрица

$$V(Z)=\begin{pmatrix}0{,}99 & 0{,}135\\ 0{,}135 & 1{,}0275\end{pmatrix}.$$

***Пример* 2.11.** Имеется двухмерная с. в. $Z=(X,\ Y)$, плотность распределения которой имеет вид

$$P(x,y)=\begin{cases}h\cdot(1-x-y) & \text{при } (x,y)\in Q=\{0\le x\le 1,\ 0\le y\le 1,\ \ x+y\le 1\};\\ 0 & \text{при } (x,y)\notin Q.\end{cases}$$

При каком h это возможно? Рассчитать ковариационную матрицу этой с. в.

Решение. Построим график заданной плотности (рис. 2.4).

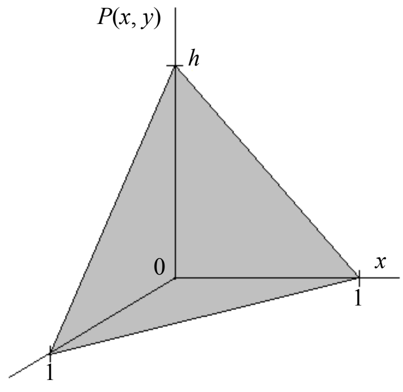


Рис. 2.4

В соответствии со свойством плотности должно быть $\int_Q P(x,y)\,ds = 1$.

В данном случае имеем

$$\int_Q P(x,y)ds = \int_0^1 dy \int_0^{1-y} h\cdot(1-x-y)dx = \int_0^1 h\cdot\left(x - \frac{x^2}{2} - y\cdot x\right)\Bigg|_0^{1-y} dy =$$

$$= \int_0^1 h\cdot\left((1-y) - \frac{(1-y)^2}{2} - y\cdot(1-y)\right)dy = h\cdot\int_0^1\left((1-y) - \frac{(1-y)^2}{2} - y\cdot(1-y)\right)dy =$$

$$= h\cdot\int_0^1 \frac{(1-y)^2}{2}dy = -h\cdot\frac{(1-y)^3}{6}\Bigg|_0^1 = \frac{h}{6} = 1 \;\Rightarrow\; h = 6.$$

Далее находим математическое ожидание элемента X

$$M(X) = \int_0^1\left(x\cdot\int_0^{1-x} 6\cdot(1-x-y)dy\right)dx = \frac{1}{4}.$$

В силу симметричности распределения очевидно, что и $M(Y) = \frac{1}{4}$.

Тогда

$$D(X) = \int_0^1\left(x^2\cdot\int_0^{1-x} 6\cdot(1-x-y)dy\right)dx - [M(X)]^2 = \frac{1}{10} - \frac{1}{16} = \frac{3}{80} \text{ и } D(Y) = \frac{3}{80}.$$

Для коэффициента ковариации имеем

$$\operatorname{cov}(X,Y) = \iint_Q (x - M(X))\cdot(y - M(Y))\cdot P(x,y)\,dxdy =$$

$$= \int_0^1\int_0^{1-y}\left(x - \frac{1}{4}\right)\cdot\left(y - \frac{1}{4}\right)\cdot 6\cdot(1-x-y)\,dxdy = -\frac{1}{80}.$$

Таким образом, искомая матрица имеет вид

$$V(Z) = \begin{pmatrix} \frac{3}{80} & -\frac{1}{80} \\ -\frac{1}{80} & \frac{3}{80} \end{pmatrix}.$$

3. ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Математическая статистика является наиболее широко востребованной на практике отраслью математики. Анализ данных и методики получения обоснованных выводов используются в самых разных отраслях человеческой деятельности, от инженерии до медицины и психологии. Экономика и менеджмент являются яркими примерами.

3.1. Точечные и интервальные оценки математического ожидания и дисперсии

Отправной точкой, а иногда и самой целью всех статистических исследований является оценка тех или иных параметров распределения изучаемых с. в. Теория даже этой простейшей (и основополагающей) задачи является весьма содержательной. Так, например, очень важно уяснить недостаточность *точечных* оценок и понять необходимость использования *интервальных*.

Необходимый теоретический материал рассмотрен в [1, п. 3.1 – 3.9] и др.

***Пример* 3.1**. Химическое предприятие получает сырье (апатитовый концентрат) от некоторого поставщика *A*. Качество сырья характеризуется концентрацией в нем полезного вещества (фосфата), а также стабильностью этого показателя, поскольку процент содержания полезного вещества определяет технологию ведения процесса, и при его значительном изменении требуется переналадка оборудования, иначе происходит снижение качества конечного продукта.

Для очередной партии сырья поставщика *A* были проведены анализы концентрации фосфата по 15 опытам (табл. 3.1). Рассчитать выборочные математическое ожидание и дисперсию содержания полезного вещества в сырье поставщика *A*. Построить для них доверительные интервалы, с уровнем доверия $p = 95$ %.

Таблица 3.1

Номер опыта	Концентрация, %
1	76,77
2	76,04
3	75,01
4	73,97
5	75,58
6	71,81
7	72,24
8	75,1
9	75,2
10	79,44
11	73,31
12	74,55
13	74,7
14	76,88
15	77,66

Таблица 3.2

Номер опыта	Концентрация, %
1	75,61
2	76,16
3	79,89
4	75,58
5	76,2
6	77,29
7	75,56
8	74,41
9	80,25
10	76,15
11	79,71
12	78,12

Решение. Обозначим случайную величину концентрации фосфата в апатите поставщика *A* через X. Используя данные анализов, находим:

– выборочное среднее

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{76{,}77 + 76{,}04 + 75{,}01 + ... + 76{,}88 + 77{,}66}{15} = 75{,}217\ ;$$

– выборочную несмещенную оценку дисперсии

$$S_X^2 = \frac{1}{n-1}\sum^{n}(x_i - \overline{X})^2 = \frac{1}{14}\left((76{,}77 - 75{,}217)^2 + ... + (77{,}66 - 75{,}217)^2\right) = 4{,}015664.$$

Строим доверительный интервал для теоретического математического ожидания, используя формулу

$$\overline{X} - t_{\alpha/2}(n-1)\frac{S_X}{\sqrt{n}} \le a \le \overline{X} + t_{\alpha/2}(n-1)\frac{S_X}{\sqrt{n}}\ .$$

При доверительной вероятности $p = 0{,}95$, для соответствующего уровня значимости $\alpha = 0{,}05$, по таблицам значений критических точек распределения Стьюдента, при числе степеней свободы $(n-1) = 14$ находим, что $t_{\alpha/2}(n-1) = t_{0{,}025}(14) = 2{,}1448$.

Таким образом, с вероятностью $p = 0{,}95$ истинное среднее содержание фосфата в сырье поставщика *A* лежит в пределах $74{,}107 \le a \le 76{,}327$.

Строим доверительный интервал для теоретической дисперсии:

$$\frac{(n-1)\cdot S_X^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)\cdot S_X^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\ .$$

По таблицам значений критических точек χ^2-распределения находим, что

$$\chi^2_{0{,}975}(14) = 26{,}119, \quad \chi^2_{0{,}025}(14) = 5{,}629.$$

Таким образом, с вероятностью $p = 0{,}95$ истинная дисперсия содержания фосфата в сырье поставщика *A* лежит в пределах

$$2{,}1524 \le \sigma^2 \le 9{,}9874\ .$$

3.2. Точечные и интервальные оценки генеральной доли признака

Помимо математического ожидания и дисперсии – практически важнейших параметров, во многих ситуациях интерес представляет информация о доли объектов из некоторой совокупности, обладающих заданным признаком. Например, оценка доли потребителей, предпочитающих определенный товар и т.д.

Необходимый теоретический материал рассмотрен в [1, п. 3.5, 3.10] и др.

***Пример* 3.2**. В некотором городе проживает 100 тысяч избирателей. Из 1000 случайным образом опрошенных избирателей 352 заявили, что пойдут на очередные выборы. Построить 99 %-ный доверительный интервал для доли избирателей, которые собираются голосовать.

Решение. При подобных социологических опросах, как правило, уже опрошенный гражданин сообщает об этом, поэтому выборку следует считать бесповторной. Хотя очевидно, что в силу значительного объема генеральной совокупности, это не будет иметь существенного значения.

Находим точечную оценку соответствующей доли:

$$w = \frac{m}{n} = \frac{352}{1000} = 0{,}352 .$$

По таблицам критических границ стандартных статистических распределений находим соответствующую двухстороннюю границу стандартного гауссовского распределения $u_{0,01/2} = 2{,}567$.

Далее – соответствующий объему выборки поправочный коэф-фициент

$$\theta = \frac{100\,000 - 1000}{100\,000} = 0{,}99 .$$

И по известной формуле

$$p_{1,2} = \frac{1}{1 + \frac{u_{\alpha/2}^2}{n} \cdot \theta} \cdot \left[w + \frac{u_{\alpha/2}^2}{2 \cdot n} \cdot \theta \pm u_{\alpha/2} \cdot \sqrt{\frac{w \cdot (1-w)}{n} \cdot \theta + \left(\frac{u_{\alpha/2}}{2 \cdot n} \cdot \theta \right)^2} \right] \quad -$$

требуемые границы $p_1 = 31{,}4\,\%$ и $p_1 = 39{,}2\,\%$.

В этих пределах с вероятностью 0,99 лежит доля избирателей, которые на момент опроса собираются принимать участие в голосовании.

4. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Получение из статистических данных обоснованных выводов, которые часто могут являться основанием для выработки конкретных управленческих решений, почти всегда сводится к проверке статистических гипотез о параметрах распределения исследуемых с. в. Соответствующие методики – неотъемлемая часть арсенала математической статистики.

4.1. Проверка параметрических гипотез

Необходимый теоретический материал рассмотрен в [1, п. 4.1 – 4.4] и др. При его изучении особое внимание следует обратить на усвоение общей схемы проверки статистических гипотез. Использующаяся при этом идея является относительно непростой, но освоившие ее получают ключ к целому спектру весьма разнообразных методик анализа данных.

***Пример* 4.1.** По данным примера 3.1 проверить утверждение поставщика *A*, что его сырье имеет среднюю концентрацию фосфата *a* = 75 % при среднеквадратическом отклонении σ = 2,5 %.

Конкурент поставщика *A* – поставщик *B* представил для проверки свое сырье. Данные по 12-ти опытам приведены в табл. 3.2.

Согласуются ли они с утверждением поставщика *B* о том, что в его сырье выше концентрация фосфата, чем у поставщика *A*, при меньшей дисперсии?

Решение. Пользуясь построенными интервалами, проверяем гипотезы:

– $H_0 : a = 75$ при альтернативной гипотезе $H_1 : a \neq 75$. Поскольку это значение принадлежит построенному интервалу $74{,}107 \le a \le 76{,}327$, гипотеза принимается. Данное заявление поставщика следует считать истинным;

– $H_0 : \sigma^2 = 2{,}5^2 = 6{,}25$ при альтернативной гипотезе $H_1 : \sigma \neq 2{,}5$. Поскольку это значение принадлежит построенному интервалу $2{,}1524 \le \sigma^2 \le 9{,}9874$, гипотеза принимается. Данное заявление поставщика следует считать истинным.

Обозначим случайную величину концентрации фосфата в апатите поставщика *B* через Y . Используя данные анализов сырья поставщика *B*, находим:

– выборочное среднее

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{75{,}61 + 76{,}16 + ...79{,}71 + 78{,}12}{12} = 77{,}0775 ;$$

– выборочную несмещенную оценку дисперсии

$$S_Y^2 = \frac{1}{n-1} \sum^{n} (y_i - \overline{Y})^2 = 3.847584 .$$

Видим, что выборочное среднее содержание фосфата в сырье второго поставщика выше, чем для первого. Для проверки статистической значимости этого превосходства рассматриваем гипотезу $H_0 : \overline{Y} - \overline{X} = 0$ при альтернативной гипотезе $H_1 : \overline{Y} - \overline{X} > 0$. Вычисляем расчетную величину (выборочную статистику):

$$t_{\text{расч}} = \frac{\overline{Y}-\overline{X}}{\sqrt{((n-1)\cdot S_x^2+(m-1)\cdot S_y^2)/(m+n-2)}}\cdot\sqrt{\frac{n\cdot m}{n+m}} =$$

$$= \frac{77{,}0775-75{,}217}{\sqrt{(14\cdot 4{,}015664+11\cdot 3{,}847584)/25}}\cdot\sqrt{\frac{15\cdot 12}{15+12}} = 2{,}419156\,.$$

По таблицам находим критическое значение t-распределения $t_{0,05}(25)=1{,}7081$, соответствующее 5 %-ному уровню значимости проверки односторонней гипотезы. Расчетное значение превосходит критическое, следовательно, нулевая гипотеза отклоняется, т.е. содержание полезного вещества в сырье второго поставщика отличается от первого статистически значимо. В этом смысле его сырье действительно лучше.

Видим, что выборочная дисперсия содержания фосфата в сырье второго поставщика ниже, чем для первого. Для проверки статистической значимости этой разницы рассматриваем гипотезу $H_0: \dfrac{S_X^2}{S_Y^2}=1$ при альтернативной гипотезе $H_1: \dfrac{S_X^2}{S_Y^2}>1$. Находим расчетное значение

$$F_{\text{расч}} = \frac{S_X^2}{S_Y^2} = \frac{4{,}015664}{3{,}847584} = 1{,}043684.$$

По таблицам распределения Фишера находим соответствующее критическое значение $F_{0,05}\left(\dfrac{14}{11}\right)=2{,}74$, видим – расчетное значение не превосходит критическое, следовательно, нулевая гипотеза не отклоняется. В этом смысле его сырье не лучше.

4.2. Проверка непараметрических гипотез. Критерий согласия Пирсона

Методики проверки непараметрических гипотез находят широкое применение в весьма разнообразных задачах, иногда даже будто бы и не относящихся к с. в. Но нередко их использование и непосредственно для проверки вида распределения изучаемой с. в. Особенно это относится к гипотезе нормальности распределения с. в. Напомним, что если изучаемая с. в. не имеет нормального распределения, то, строго говоря, к ней не применимы стандартные методики математической статистики.

Необходимый теоретический материал рассмотрен в [1, п. 4.6] и др.

***Пример* 4.2**. В табл. 4.1 представлены результаты выборочного обследования 100 малых предприятий по величине X-соотношения заемных и собственных средств. Построить гистограмму и проверить гипотезу о нормальности распределения выборки.

Таблица 4.1

5,56	5,45	5,48	5,45	5,39	5,37	5,46	5,59	5,61	5,31
5,46	5,61	5,11	5,41	5,31	5,57	5,33	5,11	5,54	5,43
5,34	5,53	5,46	5,41	5,48	5,39	5,11	5,42	5,48	5,49
5,36	5,40	5,45	5,49	5,68	5,51	5,50	5,68	5,21	5,38
5,58	5,47	5,46	5,19	5,60	5,63	5,48	5,27	5,22	5,37
5,33	5,49	5,50	5,54	5,40	5,58	5,42	5,29	5,05	5,79
5,79	5,65	5,70	5,71	5,85	5,44	5,47	5,48	5,47	5,55
5,67	5,71	5,73	4,97	5,35	5,72	5,49	5,61	5,57	5,69
5,54	5,39	5,32	5,21	5,73	5,59	5,38	5,25	5,26	5,81
5,27	5,64	5,20	5,23	5,33	5,37	5,24	5,55	5,60	5,51

Решение. Находим:

– минимальный и максимальный элементы выборки

$$x_{\min}=4{,}97\,;\ x_{\max}=5{,}85\,;$$

– по формуле Стерджерса число интервалов группирования

$$\log_2 100+1\approx 7{,}62\ \Rightarrow\ k=8\,;$$

– тогда ширина интервалов группирования

$$\Delta = \frac{5{,}85 - 4{,}97}{8} = 0{,}11\,;$$

– частоты m_i (табл. 4.2) для найденных интервалов. Теперь строим гистограмму (рис. 4.1);

– выборочные оценки математического ожидания и дисперсии

$$\overline{X} = 5{,}4602\,;\; S^2 = 0{,}030741374\,;$$

– теоретические частоты $n \cdot p_i$ для найденных интервалов (см. табл. 4.2), соответствующие нормальному распределению с параметрами

$$a = 5{,}4602\,;\; \sigma^2 = 0{,}030741374\,.$$

Таблица 4.2

Наблюдаемые частоты m_i	2	4	11	21	27	19	12	4
Теоретические частоты $n \cdot p_i$	1,247375	4,658727	11,878948	20,688017	24,615444	20,011614	11,114688	4,216273

Теперь находим расчетное значение критерия согласия Пирсона:

$$T_n = \sum_{i=1}^{k} \frac{(m_i - n \cdot p_i)^2}{n \cdot p_i} = 0{,}980738\,.$$

По таблицам критических границ распределения Пирсона

$$\chi^2_\alpha(m - l - 1) = \chi^2_{0,05}(8 - 2 - 1) = 11{,}1\,.$$

Видим, расчетное значение значительно меньше критического, – нет ни каких оснований отвергать нулевую гипотезу. Выборка действительно соответствует нормальной совокупности.

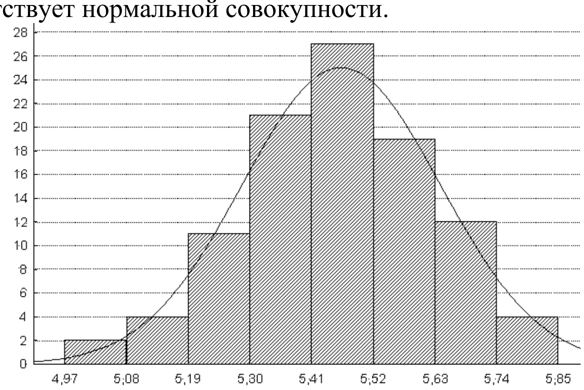


Рис. 4.1.

4.3. Однофакторный дисперсионный анализ

Однофакторный дисперсионный анализ используется для проверки наличия влияния на результирующий показатель факторов, которые не выражаются численно. Однофакторный случай является простейшим, но на нем вполне можно проиллюстрировать данную методику.

Необходимый теоретический материал рассмотрен в [1, п. 4.5] и др.

***Пример* 4.3**. Инвестор изучает вопрос о наиболее выгодном вложении инвестиций. Имеются данные по 50 инвестиционным проектам, по четырем отраслям (табл. 4.3).

Помогите инвестору принять решение о предпочтительности той или иной отрасли. Принять доверительную вероятность $p = 0{,}95$.

Решение. Видим, что для решения задачи следует выяснить, значимо ли на разброс данных влияет фактор отрасли. Для этого можно воспользоваться методикой однофакторного дисперсионного анализа, т.е. следует выяснить, статистически значимо ли межгрупповая дисперсия данных превосходит внутригрупповую.

Таблица 4.3

Отрасль	Электроника	Перерабатывающая промышленность	Легкая промышленность	Издательская деятельность
ин- вес- тици он-	25	33	13	16
	20	25	12,7	25
	–5	26	12,1	11

	15	21	10,9	80
	9	16	10,3	45
	24	30	10,1	60
	14	25	10,9	35
	9	28	–	23
	–	17	–	40

Вычисляем выборочные значения:

– групповых средних

$$\overline{X}_1 = \frac{25+...+9}{8} = 13,87; \quad \overline{X}_2 = \frac{33+...+17}{9} = 24,55;$$

$$\overline{X}_3 = \frac{13+...+10,9}{7} = 11,42; \quad \overline{X}_4 = \frac{16+...+40}{9} = 37,22.$$

Видим, что средняя рентабельность в издательской деятельности заметно выше, однако, является ли это результатом общей случайности разброса данных или нет, пока утверждать нельзя;

– общей средней $\overline{X} = \dfrac{25+...+40}{8+9+7+9} = 22,63$;

– внутригрупповой дисперсии $S_1^2 = \dfrac{1}{33-4}\sum_{i=1}^{4}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_j)^2 = 167,13$;

– межгрупповой дисперсии

$$S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m} k_i\,(\overline{X}_i - \overline{X})^2 = \frac{1}{4-1}\Big(8\cdot(13,87-22,63)^2 + ... + 8\cdot(37,22-22,63)^2\Big) = \quad = \frac{3441,2}{3} = 1147,0\,.$$

Итак, для проверки того, что межгрупповая дисперсия статистически значимо превосходит внутригрупповую, рассматриваем гипотезу $H_0 : \dfrac{S_2^2}{S_1^2} = 1$ при альтернативной гипотезе $H_1 : \dfrac{S_2^2}{S_1^2} > 1$. Для этого используем обычный критерий Фишера. Находим расчетную величину

$$F_{\text{расч}} = \frac{S_2^2}{S_1^2} = \frac{1147,0}{167,13} = 6,8634\,.$$

И сравниваем ее с табличным значением односторонней критической точки распределения Фишера при соответствующем числе степеней свободы числителя и знаменателя,

$$F_{0,05}\left(\frac{3}{29}\right) = 2,934\,.$$

Видим – расчетное значение превосходит критическое, следовательно, нулевая гипотеза отклоняется, т.е. фактор отрасли значимо влияет на рентабельность инвестиций. Теперь есть все основания предпочесть издательскую деятельность.

СПИСОК ЛИТЕРАТУРЫ

1. Солопахо, А.В. Теория вероятностей и математическая статистика: краткий курс для экономистов : учеб. пособие / А.В. Солопахо. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2007.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман. – М. : Высшая школа, 1977.
3. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике / В.Е. Гмурман. – М. : Высшая школа, 1979.
4. Кремер, Н.Ш. Теория вероятностей и математическая статистика / Н.Ш. Кремер. – М. : ЮНИТИ, 2000.
5. Колемаев, В.А. Теория вероятностей и математическая статистика / В.А. Колемаев, В.Н. Калинина. – М. : Инфра-М, 1997.
6. Красс, М.С. Математика для экономических специальностей : учебник / М.С. Красс. – М. : ИНФРА-М, 1999.
7. Хили, Дж. Статистика. Социологические и маркетинговые исследования / Дж. Хили ; пер. с англ. – СПб. : Питер, 2005.
8. Айвазян, С.А. Прикладная статистика и основы эконометрики / С.А. Айвазян, В.С. Мхитарян. – М. : ЮНИТИ, 1998.