Ю.Е. ВОСКОБОЙНИКОВ
Т.Т. БАЛАНЧУК

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

(с примерами в Excel)

НОВОСИБИРСК 2013

МИНИСТЕРСТВО ОБРАЗОВАНИЯ
И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ
АРХИТЕКТУРНО-СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ (СИБСТРИН)

Ю.Е. Воскобойников, Т.Т. Баланчук

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА (с примерами в Excel)

Рекомендовано УМО Российской академии естествознания по классическому университетскому и техническому образованию в качестве учебного пособия для студентов высших учебных заведений, обучающихся по направлениям подготовки 080200.62 «Менеджмент», 080100.62 «Экономика»

НОВОСИБИРСК 2013

УДК 519.2
ББК 22.17
В 762

Воскобойников Ю. Е.

Теория вероятностей и математическая статистика (с примерами в Excel) : учеб. пособие / Ю. Е. Воскобойников, Т. Т. Баланчук ; Новосиб. гос. архитектур.-строит. ун-т (Сибстрин). – Новосибирск : НГАСУ (Сибстрин), 2013. – 200 с.

ISBN 978-5-7795-0632-8

В учебном пособии рассмотрены основные понятия теории вероятностей, касающиеся случайных событий, дискретных и непрерывных случайных величин. Большое внимание уделено разделам математической статистики: точечному и интервальному оцениванию параметров случайных величин, проверке статистических гипотез. Учебное пособие содержит копии большого количества фрагментов документов Excel, которые позволят студентам не только лучше усвоить учебный материал, но и эффективно использовать программу Excel при выполнении курсовых и дипломных работ.

Учебное пособие предназначено студентам, обучающимся по направлениям 080200.62 «Менеджмент», 080100.62 «Экономика», 230400.62 «Информационные системы и технологии», а также магистрантам и аспирантам соответствующих специальностей.

Печатается по решению издательско-библиотечного совета НГАСУ (Сибстрин)

Рецензенты:

– А.В. Федоров, д-р физ.-мат. наук, профессор, заведующий лабораторией Института теоретической и прикладной механики СО РАН;
– М.С. Соппа, д-р физ.-мат. наук, профессор НГАСУ (Сибстрин)

ISBN 978-5-7795-0632-8

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

На практике найти точное решение возникшей математической задачи нередко не удается. Это происходит потому, что некоторые очень существенные факторы, влияющие на результат, носят случайный, труднопредсказуемый в единичном испытании характер. Например, если подбрасывать обычный игральный кубик, то заранее предсказать достоверно, какая грань из шести выпадет, невозможно. Однако методы теории вероятностей позволяют с большой степенью надежности определить примерное количество выпадений конкретной грани, если кубик подбрасывается достаточно большое число раз. Методы теории вероятностей эффективно применяются в самых различных областях – теоретических и прикладных исследованиях теории автоматического управления, теории надежности, теории ошибок измерений, теории массового обслуживания и др.

Очень важны подходы, устанавливающие и измеряющие корреляционную связь между случайными величинами.

Математическая статистика, в свою очередь, на основе изучения статистических данных (результатов измерений) с использованием математического аппарата теории вероятностей позволяет устанавливать закономерности, которым подчиняются массовые случайные явления в технике, экономике, обществе. Современная математическая статистика разрабатывает способы сбора и группировки статистических сведений, изучает методы их анализа. Это позволяет получать оценку неизвестной вероятности случайного события, оценку параметров распределений и величины взаимозависимости случайных величин, а также проводить проверку статистических гипотез.

Данный курс является неотъемлемой частью общей математической подготовки в полном соответствии с требованиями, отраженными в ФГОС3 по направлениям «Менеджмент» и «Экономика».

РАЗДЕЛ 1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Тема 1. Случайные события

1.1. Основные понятия теории вероятностей

Для первоначального интуитивного ознакомления дадим нестрогое описание понятий. *Вероятность события* – это мера нашей уверенности в том, что данное событие произойдет. ***Теория вероятностей*** – это математика случайности (хотя, с точки зрения обычного человека, математика и случайность мало совместимы).

Теория вероятностей применяется при изучении массовых процессов в окружающем мире. Например, можно сделать следующие обоснованные прогнозы:

– число покупателей какого-либо товара в магазине в последующий период;
– примерное число больных, которые обратятся в поликлинику с гриппом;
– период времени, в течение которого какое-либо устройство (например, телевизор) будет работать без сбоев.

Предметом рассмотрения в теории вероятностей также являются:

– расшифровка секретных сообщений (криптография);
– расчет игорного бизнеса;
– теория надежности и ошибок измерений;
– прогнозирование инвестиционных процессов;
– выбраковка продукции (определение ее качественного состава).

Опыт (*испытание*) – создание фиксированных условий для реализации какого-либо явления, например:

1) подбрасывание монеты;
2) подбрасывание кубика (игральной кости);
3) выбор случайной точки на отрезке [0, 1];
4) анкетирование случайно встреченного человека на улице (узнать и записать его вес в таблицу).

Назовем ***элементарными исходами*** ω_i различные взаимоисключающие результаты какого-либо опыта.

Пространство элементарных исходов $\mathbf{\Omega}$ – полная совокупность элементарных исходов опыта: $\omega_i \in \Omega$.

Запишем Ω для перечисленных выше примеров:

1) $\Omega = \{Г, Ц\}$ ($Г$ – герб, $Ц$ – цифра);
2) $\Omega = \{1, 2, 3, 4, 5, 6\}$;
3) Ω = {бесконечное число точек отрезка [0, 1]};
4) Ω = {бесконечное число точек отрезка [20, 200]}.

Случайное событие – это событие, которое может произойти, а может и не произойти вне зависимости от нашей воли (если мы можем повлиять на реализуемость события, то оно уже не может считаться случайным). Случайные события будем обозначать большими латинскими буквами A, B, C и т.д.

Установим связь между случайными событиями и элементарными исходами. Элементарный исход ω_i называется ***благоприятствующим*** случайному событию A, если при его реализации происходит событие A. Говорят, что такие элементарные исходы *порождают* событие A. Элементарные исходы, при реализации которых событие A не наступило, являются ***неблагоприятствующими*** событию A.

Пример 1.1.1. Определим случайное событие как выпадение четного числа очков на верхней грани при однократном подбрасывании кубика, т.е. $A = \{2, 4, 6\}$= {выпадение четной грани на кубике}. Найти элементарные исходы, *благоприятствующие* и *неблагоприятствующие* случайному событию A.

Решение. Введем элементарные исходы $\omega_i = \{\text{выпало } i \text{ очков}\}$. Тогда элементарные исходы $\{\omega_2, \omega_4, \omega_6\}$ являются *благоприятствующими* случайному событию A. Напротив, $\{\omega_1, \omega_3, \omega_5\}$ являются *неблагоприятствующими* событию A. Случайное событие A является подмножеством пространства элементарных исходов $\Omega : A \subset \Omega$. ☻

Пример 1.1.2. Опыт состоит в одновременном подбрасывании двух монет. Рассмотрим событие B = {выпадение *хотя бы* одного герба}. Требуется: 1) перечислить все элементарные исходы, т.е. построить Ω; 2) перечислить все исходы ω_i, благоприятствующие событию B.

Решение. Пронумеруем монеты. Тогда элементарные исходы:

1 М	2 М
Г	*Г*
Г	*Ц*
Ц	*Г*
Ц	*Ц*

$\omega_1 = \{Г, Г\}$, $\omega_2 = \{Г, Ц\}$,

$\omega_3 = \{Ц, Г\}$, $\omega_4 = \{Ц, Ц\}$.

$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$.

$B = \{\omega_1, \omega_2, \omega_3\}$. ☻

Задача 1.1.1. Два стрелка одновременно стреляют по одной мишени. Требуется: 1) перечислить все элементарные исходы, т.е. построить Ω; 2) перечислить все исходы ω_i, благоприятствующие событию B = {в мишени ровно одна пробоина}. При этом считаем, что если в мишень попали две пули, то и отверстий можно различить два. ♦

Замечание 1.1.1. Для решения подобных задач полезно разобраться, чем отличаются следующие события:

$$\{\text{в мишени} \begin{Bmatrix} \textit{не менее} \\ \textit{хотя бы} \end{Bmatrix} \begin{Bmatrix} \textit{по крайней мере} \\ \textit{более чем} \end{Bmatrix} \begin{Bmatrix} \textit{ровно} \\ \textit{не более чем} \end{Bmatrix} \text{1 пробоина}\}.$$ ●

Для наглядности пространство элементарных исходов Ω изображают некоторой областью на плоскости (рис. 1.1*а*). Любое событие A, связанное с этим испытанием, изображают как часть этой области – заштрихованное множество) (не обязательно круг). Очевидно, что каждый элемент этого множества есть элементарный исход, благоприятствующий событию A.

Невозможное событие (обозначается $\varnothing$) – событие, которое заведомо не произойдет в данном опыте. Например, при подбрасывании кубика невозможными событиями являются выпадения 0, 10, > 100 очков или одновременное выпадение 1 и 5.

События A и B называются ***несовместными***, если они не могут произойти одновременно в результате одного испытания, т.е. не имеют общих элементарных исходов. Это хорошо иллюстрирует рис. 1.1*б*.

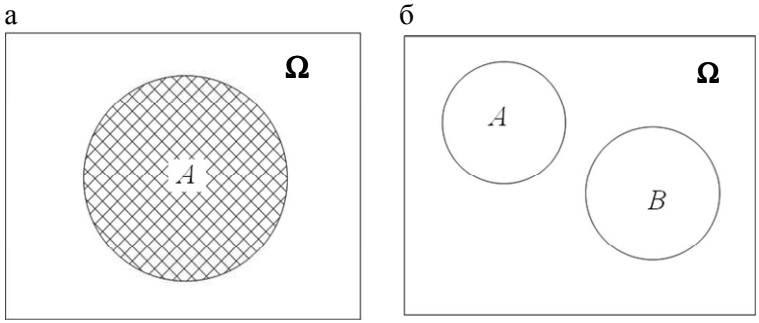


Рис. 1.1. Пространство элементарных исходов
и несовместные события

Достоверное событие (обозначается Ω) – событие, которое обязательно, наверняка произойдет в данном опыте. Например, при подбрасывании кубика выпадет число из отрезка [1, 6], < 100, > 0.

Противоположным к событию A называется событие, обозначаемое $\overline{A}$ и состоящее из таких элементарных исходов пространства Ω, которые не благоприятствуют событию A. Противоположное событие $\overline{A}$ происходит тогда и только тогда, когда не происходит событие A.

Справедливы следующие равенства:

$$\overline{\mathbf{\Omega}} = \varnothing\,;\ \overline{\varnothing} = \mathbf{\Omega}\,.$$

Пусть при подбрасывании кубика A = {2, 4}, тогда $\overline{A}$ = {1, 3, 5, 6}. Если при подбрасывании монеты D = {Г}, то $\overline{D}$ = {Ц}.

Несколько несовместных событий образуют ***полную группу событий***, если в результате испытаний появится только одно из этих событий.

1.2. Операции над случайными событиями

Сумма событий $C = A + B$ произойдет тогда и только тогда, когда произойдет событие A, событие B или оба вместе. На рис. 1.2*а* сумма событий $A + B$ показана заштрихованным множеством.

Пример 1.2.1. Пусть при подбрасывании кубика $A = \{2, 4\}$, $B = \{4, 6\}$. Определить сумму событий $C = A + B$.

Решение. Из определения суммы событий следует, что $C = \{2, 4, 6\}$. Таким образом, количество благоприятствующих элементарных исходов увеличилось. ☻

Произведение событий $C = A \cdot B$ – произойдет тогда и только тогда, когда произойдут оба события A и B одновременно. На рис. 1.2*б* произведение событий $A \cdot B$ показано заштрихованным множеством.

Пример 1.2.2. Пусть при подбрасывании кубика $A = \{2, 4\}$, $B = \{4, 6\}$. Определить произведение событий $C = A \cdot B$.

Решение. Из определения произведения событий следует, что $C = \{4\}$, так как только один элементарный исход $\omega_4 = \{4\}$ является благоприятствующим как событию A, так и событию B. Видно, что количество благоприятствующих элементарных исходов уменьшилось. ☻

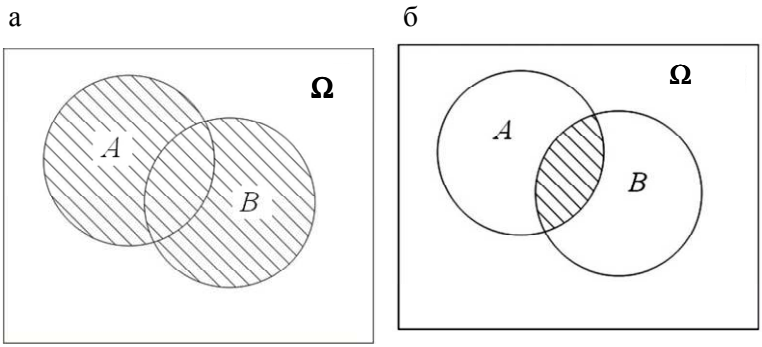


Рис. 1.2. Сумма $A + B$ и произведение $A \cdot B$ двух событий

Задача 1.2.1. Одновременно подбрасываются три монеты. A = {выпало < трех гербов}, B = {выпало ≥ двух цифр}. Найти элементарные исходы, благоприятствующие событию $C = A \cdot B$. *Ответ:* C = {*ГЦЦ*, *ЦЦЦ*, *ЦЦГ*, *ЦГЦ*}. ♦

Для противоположного события $\overline{A}$ справедливы равенства:

$$\boxed{\overline{A} + A = \Omega, \quad \overline{A} \cdot A = \varnothing}. \qquad (1.2.1)$$

Для несовместных (взаимоисключающих) событий имеет место

$$\boxed{A \cdot B = \varnothing} \qquad (1.2.2)$$

Пример 1.2.3. Два стрелка стреляют одновременно по одной мишени. Пусть A = {хотя бы один стрелок попал}, а B = {хотя бы один стрелок промахнулся}. Будут ли A и B несовместны?

Решение. Событию A благоприятствуют следующие исходы: $\{П,П\},\{П,Н\},\{Н,П\}$, где $П$ означает попадание стрелка в мишень, $Н$ – промах (не попал). Событию B благоприятствуют следующие исходы: $\{Н,Н\},\{П,Н\},\{Н,П\}$. Видно, что события A и B совместны, так как они могут произойти одновременно – элементарные исходы $\{П,Н\}, \{Н,П\}$. Следовательно, $A \cdot B \neq \varnothing$. ☻

Если события $A_1, A_2, ..., A_n$ образуют полную группу событий, то очевидно:

1. $$\boxed{A_1 + A_2 + \cdots + A_n = \Omega}. \qquad (1.2.3)$$

2. $$\boxed{A_i \cdot A_j = \varnothing}, \text{ если } i \neq j. \qquad (1.2.4)$$

1.3. Элементы комбинаторики

Комбинаторика применяется для подсчета количества комбинаций, вариантов и элементарных исходов в каком-либо опыте. Это очень важно при вычислении вероятности того или иного случайного события.

Правило произведения. Предположим, что переменная i может принимать одно из m значений, а переменная j – независимо от нее одно из n значений. Сколько упорядоченных пар (i, j) можно составить? Ответ получаем графически. Количество таких пар равно количеству клеток нижеприведенной таблицы $N = n \cdot m$.

j \ i	i_1	i_2	...	i_m
j_1	(i_1, j_1)	(i_2, j_1)		(i_m, j_1)
j_2	(i_1, j_2)	(i_2, j_2)		(i_m, j_2)
...				
j_n	(i_1, j_n)	(i_2, j_n)		(i_m, j_n)

Если независимых переменных не две, а больше, то правило произведения дает ответ, что количество упорядоченных комбинаций (i_1, i_2, ..., i_p) равно

$$\boxed{N = n_1 \cdot n_2 \cdot \ldots \cdot n_p}, \qquad (1.3.1)$$

где n_j – количество значений, принимаемых переменной i_j.

Пример 1.3.1. Опыт состоит в подбрасывании трех монет. Определить количество возможных элементарных исходов этого опыта.

Решение. Каждой из монет можно сопоставить переменную, которая принимает два значения: *Г* или *Ц*. Значит, $p = 3$, а каждая из i_1, i_2, i_3 принимает, независимо от других, ровно два значения, т.е. $n_1 = 2$, $n_2 = 2$, $n_3 = 2$. Количество возможных элементарных исходов этого опыта: $N = n_1 \cdot n_2 \cdot n_3 = 2^3 = 8$. Полученный результат можно проверить и методом перебора. ☻

Перестановки P_n – все *комбинации* из n различных элементов, отличающиеся друг от друга ***только порядком следования элементов.***

Формула для подсчета количества перестановок:

$$\boxed{P_n = n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-2) \cdot (n-1) \cdot n}. \qquad (1.3.2)$$

Пример 1.3.2. Даны три карточки с буквами А, В, С. Составить все перестановки этих карточек.

Решение. Это ABC, ACB, BAC, BCA, CAB, CBA – 6 вариантов. Проверим полученное число вариантов по формуле (1.3.2): $P_3 = 3! = 3 \cdot 2 \cdot 1 = 6$. ☻

Сочетания C_n^m – все комбинации, содержащие m предметов, входящих в набор из n различных предметов, $m \le n$. ***Порядок следования предметов здесь безразличен*** (комбинации отличаются только составом)**.** Формула для подсчета количества сочетаний:

$$\boxed{C_n^m = \frac{n!}{m!(n-m)!} = \frac{1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-2)(n-1)n}{1 \cdot 2 \cdot \ldots \cdot (m-2) \cdot (m-1) \cdot m \cdot 1 \cdot 2 \cdot \ldots \cdot (n-m-2) \cdot (n-m-1) \cdot (n-m)}}. \qquad (1.3.3)$$

Замечание 1.3.1. При вычислении числа сочетаний следует иметь в виду, что 0! = 1, и поэтому $C_n^0 = 1$, $C_n^n = 1$, $C_n^1 = n$. ●

Пример 1.3.3. Составим все сочетания по три карточки из набора из четырех карточек с буквами A, B, C, D.

Решение. Число комбинаций определим по формуле (1.3.3):

$$C_4^3 = \frac{1 \cdot 2 \cdot 3 \cdot 4}{1 \cdot 2 \cdot 3 \cdot 1} = 4.$$

Это ABC, ACD, ABD, BCD. ☻

Отметим, что всегда при вычислении количества сочетаний существует возможность сократить большое количество сомножителей. Например: $C_4^3 = \frac{\cancel{1} \cdot \cancel{2} \cdot \cancel{3} \cdot 4}{\cancel{1} \cdot \cancel{2} \cdot \cancel{3} \cdot 1} = 4.$

Размещением A_n^m из n элементов по m элементов называются комбинации, состоящие из m элементов и ***отличающиеся как составом элементов, так и порядком следования элементов в комбинации***. Число размещений определяется выражением

$$\boxed{A_n^m = C_n^m \cdot m! = \frac{n!}{(n-m)!}}. \qquad (1.3.4)$$

Пример 1.3.4. Определить количество размещений по три карточки из набора из четырех карточек с буквами A, B, C, D. Комбинации отличаются как составом, так и порядком следования, и число комбинаций равно

$$A_4^3 = \frac{4!}{3!(4-3)!} \cdot 3! = \frac{1 \cdot 2 \cdot 3 \cdot 4}{(1)!} = 24 .$$

Комбинаторная схема. Поясним эту схему следующим примером.

Пример 1.3.5. Пусть имеется совокупность предметов двух сортов: n – 1-го сорта и m – 2-го сорта (рис. 1.3). Требуется из этой совокупности сделать выборку (взять часть предметов). Сколько всего существует различных выборок – таких, что в них ровно p предметов 1-го сорта и q предметов 2-го сорта? Сколько существует всего выборок, содержащих $p + q$ предметов, если не следить за качественным составом выборки?

Рис. 1.3. Комбинаторная схема

Решение. Выборок с заданным качественным составом будет ровно $C_n^p \cdot C_m^q$. Если же не обращать внимание на качественный состав выборки, то их будет гораздо больше: C_{n+m}^{p+q}. ☻

Пример 1.3.6. В партии из 100 деталей есть 10 бракованных. Наугад выбираем из них 3 детали. Найти количество возможных комбинаций – таких, чтобы среди выбранных ровно две детали были годными, а одна бракованной.

Решение. $C_{90}^2 \cdot C_{10}^1$ = 4005 · 10 = 40050, а всего C_{100}^3 = = 98 · 99 · 100/(1 · 2 · 3) = 161 700. ☻

Вопросы и задачи для самопроверки

1. Что такое опыт?
2. Какое событие называется случайным?
3. Что такое элементарные исходы опыта?
4. Что такое полное пространство элементарных исходов?
5. Какой исход называется благоприятствующим случайному событию A?
6. Монета бросается два раза. Описать пространство элементарных исходов Ω и события:

   A = {выпал один герб};

   B = {выпал хотя бы один герб}.

6. Игральная кость бросается два раза. Описать пространство элементарных исходов $\mathbf{\Omega}$ и события:

   A = {сумма выпавших очков равна 8};

   B = {произведение выпавших очков равно 8};

   C = {сумма выпавших очков четна};

   D = {произведение выпавших очков четно}.

7. Дать определения суммы и произведения случайных событий.
8. Какие события называются противоположными?
9. Какие события называются несовместными?
10. Что означают следующие события:

    а) $A+\varnothing$; б) $A+\mathbf{\Omega}$; в) $A+A$; г) $A+\overline{A}$; д) $A\cdot A$;

    е) $A\cdot\overline{A}$; ж) $A\cdot\varnothing$; з) $A\cdot\mathbf{\Omega}$; и) $A+\overline{\mathbf{\Omega}}$?

    Ответы: ж) $A\cdot\varnothing=\varnothing$; з) $A\cdot\mathbf{\Omega}=A$; и) $A+\overline{\mathbf{\Omega}}=A$.

11. Имеются три произвольные события A, B, C. Записать события, заключающиеся в том, что из трех событий:

    а) произошло только событие A;

    б) произошло только одно событие;

    в) произошли только события A и B.

    Ответы: а) $A\cdot\overline{B}\cdot\overline{C}$; б) $A\cdot\overline{B}\cdot\overline{C}+\overline{A}\cdot B\cdot\overline{C}+\overline{A}\cdot\overline{B}\cdot C$;

    в) $A\cdot B\cdot\overline{C}$.

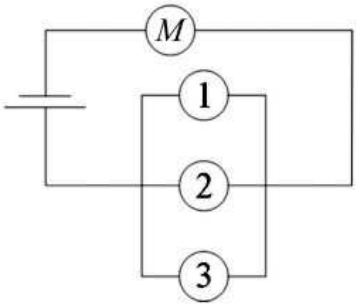
12. На станке изготовлено n деталей. Пусть события A_k (k = 1, 2, 3) заключаются в том, что k-я изготовленная деталь

имеет дефект. Записать через события A_k следующие события:
а) ни одна из трех деталей не имеет дефекта;
б) хотя бы одна деталь имеет дефект;
в) только одна деталь имеет дефект.

Ответы: а) $\overline{A}_1 \cdot \overline{A}_2 \cdot \overline{A}_3$; б) $A_1 + A_2 + A_3$;

в) $A_1 \cdot \overline{A}_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot A_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot \overline{A}_2 \cdot A_3$.

13. Сформулировать правило произведения для подсчета количества комбинаций.
14. Дать определения перестановок. Как подсчитать количество перестановок?
15. Дать определение сочетаний. Как подсчитать количество сочетаний?
16. Дать определение размещений. Как подсчитать количество размещений?
17. Почему число размещений больше числа сочетаний?
18. Когда можно применять комбинаторную схему?
19. На рисунке изображена электрическая цепь. Событие A заключается в том, что элемент M работает, событие B_k (k = 1, 2, 3) – работает k-й элемент. Событие C = {цепь замкнута}. Выразить события C и $\overline{C}$ через события A и B_k (k = 1, 2, 3).



Ответ: $C = A \cdot (B_1 \cdot \overline{B_2} \cdot \overline{B_3} + \overline{B_1} \cdot B_2 \cdot \overline{B_3} + \overline{B_1} \cdot \overline{B_2} \cdot B_3 + B_1 \cdot B_2 \cdot \overline{B_3} +$

$+ B_1 \cdot \overline{B_2} \cdot B_3 + \overline{B_1} \cdot B_2 \cdot B_3 + B_1 \cdot B_2 \cdot B_3) =$

$= A \cdot (B_1 + B_2 + B_3)$;

$\overline{C} = \overline{A} + \overline{B_1} \cdot \overline{B_2} \cdot \overline{B_3}$.

20. В партии из 50 деталей есть 5 бракованных. Наугад выбираем из них 4 детали. Найти количество возможных комбинаций – таких, чтобы среди выбранных деталей ровно две были годными и две – бракованными.
21. В партии из 50 деталей есть 5 бракованных. Наугад выбираем из них 4 детали. Найти количество всевозможных комбинаций, не обращая внимания на их качественный состав. Сравнить полученное количество комбинаций с количеством комбинаций, полученных в предыдущем примере. Объяснить причину отличия этих количеств.

Тема 2. Вероятность случайного события. Сложение и умножение вероятностей

2.1. Определение вероятности случайного события

Пусть для некоторого опыта множество элементарных исходов Ω состоит из конечного числа равновозможных элементарных исходов ω_i (i = 1, ..., n). *Равновозможные исходы* опыта – это события, у каждого из которых нет преимуществ для наступления перед другими событиями.

Пусть A – случайное событие, для которого благоприятствующими являются k исходов $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_k}$, т.е. $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_k}\}$, $0 \le k \le n$.

Тогда ***вероятностью события*** A называется число

$$\boxed{P(A) = \frac{\textit{количество благоприятных исходов}}{\textit{количество всех исходов}} = \frac{k(A)}{n}}. \quad (2.1.1)$$

Такое определение вероятности называют классическим (существуют также геометрическое и статистическое определения вероятности).

Пример 2.1.1. Заданы следующие случайные события:

A = {выпадение герба на монете};

B = {выпадение «5» на кубике};

C = {выпадение числа > 2 на кубике}.

Определить вероятности этих случайных событий.

Решение. Событию A благоприятствует только один элементарный исход $\{Г\}$ из возможных двух элементарных исходов $\{Г\},\{Ц\}$. Поэтому $P(A) = 1/2 = 0.5$. Аналогично событию B благоприятствует только один элементарный исход $\{5\}$ из возможных 6 элементарных исхода $\Omega = \{1, 2, 3, 4, 5, 6\}$. Поэтому $P(B) = 1/6 = 0.16667$. Событию C благоприятствуют четыре элементарных исхода $C = \{3, 4, 5, 6\}$ из возможных 6 элементарных исходов $\Omega = \{1, 2, 3, 4, 5, 6\}$. Поэтому $P(C) = 4/6 = 0.66667$. ☻

Задача 2.1.1. В урне 8 белых и 3 черных шара. Наугад достали 1 шар. Какова вероятность того, что он белый?

Ответ. $P = 8/11$. ♦

Пример 2.1.2. Монета бросается два раза. Найти вероятности событий:

A = {герб выпадет один раз};

B = {герб выпадет хотя бы один раз};

C = {герб не выпадет ни разу}.

Решение. Испытание – монета бросается два раза. Пространство элементарных исходов состоит из $n = 4$ равновозможных исходов: при первом бросании две возможности – герб ($Г$) и цифра ($Ц$), после этого при втором бросании также две возможности. По правилу произведения выполнить первое и второе действие вместе возможно $2 \cdot 2 = 4$ способами.

Число исходов, благоприятствующих событию A, равно 2: это исходы ($Г$,$Ц$) и ($Ц$,$Г$). Таким образом, $k(A) = 2$, тогда $P(A) = \frac{k(A)}{n} = \frac{2}{4} = \frac{1}{2}$. Для события B $k(B) = 3$ – это исходы ($Г$,$Ц$), ($Ц$,$Г$), ($Г$,$Г$). $P(B) = \frac{k(B)}{n} = \frac{3}{4}$. Для события C $k(C) = 1$ – это исход ($Ц$,$Ц$). $P(C) = \frac{1}{4}$. ☻

Задача 2.1.2. Игральная кость бросается один раз. Найти вероятности событий:

A = {выпадет четное число очков};

B = {выпадет число очков, кратное трем};

C = {выпадет не менее трех очков}.

Ответы: $P(A)=\frac{1}{2}; P(B)=\frac{1}{3}; P(C)=\frac{2}{3}$. ♦

Свойства вероятности

1. Вероятность случайного события есть неотрицательная величина, т.е. $P(A) \geq 0$.
2. Вероятность случайного события есть число между 0 и 1, т.е. $0 \leq P(A) \leq 1$.

   Вопрос: Достигаются ли крайние значения вероятностей 0 и 1?

   Ответ на этот вопрос дают свойства 3 и 4.
3. Вероятность достоверного события равна 1, т.е. $P(\Omega) = 1$.
4. Вероятность невозможного события равна 0, т.е. $P(\varnothing) = 0$.

Приведем полезные соотношения, часто используемые при вычислении вероятности случайного события:

$$P(A)=1-P(\bar{A}); \quad P(\bar{A})=1-P(A), \qquad (2.1.2)$$

Они непосредственно следуют из (1.2.1).

Пример 2.1.3. Игральная кость бросается три раза. Найти вероятности событий:

A = {появится одинаковое число очков};

B = {появится разное число очков}.

Решение. По принципу умножения $n = 6 \cdot 6 \cdot 6 = 216$ – число возможных исходов при троекратном бросании игральной кости.

Событие A= {(111), (222), …, (666)}, $k(A)=6$. $P(A)=\frac{6}{216}=\frac{1}{36}$.

Для нахождения $k(B)$ воспользуемся правилом умножения: на первой кости может появиться любое из шести очков; на второй – любое, кроме того, которое появилось на первой, т.е.

любое из пяти; на третьей – любое, кроме тех, которые появятся на первой и второй кости, т.е. любое из четырех.

Отсюда $k(B) = 6 \cdot 5 \cdot 4$, $P(B) = \dfrac{6 \cdot 5 \cdot 4}{6 \cdot 36} = \dfrac{5}{9}$. ☻

Замечание 2.1.1. В теории вероятностей выполняемые действия удобно представлять в виде модели извлечения элементов из некоторого множества $\{a_1, a_2, \ldots, a_n\}$, называемого ***генеральной совокупностью объема n***.

Выборкой объема k из генеральной совокупности называется упорядоченная последовательность $(a_{i_1}, a_{i_2}, \ldots, a_{i_k})$ элементов генеральной совокупности. Выбор элементов может быть без возвращения (когда выбранный элемент не возвращается в генеральную совокупность) и с возвращением. Полученные в результате такого отбора выборки называются соответственно ***выборкой без возвращения*** *и* ***выборкой с возвращением***.

Число всех возможных ***выборок без возвращения*** объема k, полученных из генеральной совокупности объема n, равно числу размещений из n элементов по k:

$$\boxed{A_n^k = n(n-1)(n-2)\ldots(n-k+1)}.$$

Такие выборки ***отличаются друг от друга либо составом, либо порядком своих элементов***. В частном случае число таких всевозможных выборок объема n из генеральной совокупности объема n равно числу перестановок из n элементов:

$$\boxed{P_n = n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n}. \qquad (2.1.3)$$

Такие выборки ***отличаются только порядком своих элементов***.

Число всевозможных ***выборок с возвращением*** объема k, извлеченных из генеральной совокупности объема n, равно n^k. Такие выборки также отличаются либо составом, либо порядком своих элементов.

Число всевозможных выборок объема k, извлеченных из генеральной совокупности объема n и отличающихся друг от друга только составом своих элементов, равно числу сочетаний из n элементов по k:

$$\boxed{C_n^k = \frac{A_n^k}{P_k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)...(n-k+1)}{1\cdot 2\cdot 3\cdot ...\cdot k}}. \qquad (2.1.4)$$

Свойства сочетаний: $C_n^0 = 1$; $C_n^n = 1$; $C_n^1 = n$; $C_n^k = C_n^{n-k}$. Последним свойством рекомендуется пользоваться, когда $k > \frac{n}{2}$. Все вышесказанное необходимо учитывать при подсчете общего числа элементарных исходов. ●

Пример 2.1.4. В урне 8 белых и 3 черных шара. Наугад достали 2 шара. Какова вероятность того, что оба они белые?

Решение. Так как шары в урну не возвращаются, то неважно, по очереди достали два шара или оба сразу. Используем комбинаторную схему. Искомая вероятность определяется выражением:

$$P = \frac{C_8^2 \cdot C_3^0}{C_{11}^2} = \frac{28\cdot 1}{55} = 0.509. \quad ☻$$

Пример 2.1.5. В урне 8 белых и 3 черных шара. Наугад достают первый шар, и после определения его цвета шар возвращается в урну. Затем достают второй шар. Какова вероятность того, что оба вынутые шары белые?

Решение. Так как испытание проводится с возвращением вынутого шара (***выборка с возвращением***), то общее число элементарных исходов равно $n = 11\cdot 11 = 121$. Число элементарных исходов, благоприятствующих случайному событию (оба шара

белые), равно $k = 8 \cdot 8 = 64$. Тогда вероятность события равна $P = \frac{64}{121} = 0.529$.

2.2. Теорема сложения вероятностей

Пусть с некоторым испытанием связаны события *A* и *B*. Тогда вероятность того, что произойдет событие *A* + *B* (сумма случайных событий *A* и *B*) равна

$$\boxed{P(A+B) = P(A) + P(B) - P(A \cdot B)}. \qquad (2.2.1)$$

Наличие члена $P(A \cdot B)$ со знаком минус можно объяснить с помощью рис. 1.2*а*, на котором показан результат сложения двух событий *A* и *B*. Будем считать, что вероятность события *A* + *B* «равна» заштрихованной площади. Однако сумма площадей $P(A) + P(B)$ дважды включает площадь, соответствующую $P(A \cdot B)$, что обуславливает необходимость вычитания в формуле (2.2.1).

Следствия

1. Если события *A* и *B* несовместны, то

$$\boxed{P(A+B) = P(A) + P(B)}. \qquad (2.2.2)$$

2. Для трех событий *A*, *B*, *C*, связанных с некоторым испытанием, теорема сложения вероятностей имеет вид:

$$\boxed{\begin{gathered} P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - \\ -P(BC) + P(ABC). \end{gathered}} \qquad (2.2.3)$$

3. Если события *A*, *B* и *C* несовместны, то

$$\boxed{P(A+B+C) = P(A) + P(B) + P(C)}. \qquad (2.2.4)$$

Пример 2.2.1. Имеются две урны: в одной a белых и b черных шаров; в другой – c белых и d черных шаров. Из каждой урны вынимают по шару. Найти вероятность того, что оба шара будут одного цвета.

Решение. Обозначим событие A = {оба шара одного цвета} и представим его следующей суммой:

A = {оба шара одного цвета} = {оба шара либо белые, либо черные} = {оба шара белые} + {оба шара черные} = $B_1 + B_2$.

Так как события B_1, B_2 несовместны, то

$P(A) = P(B_1 + B_2) = P(B_1) + P(B_2)$. Далее найдем

$$P(B_1) = \frac{a \cdot c}{(a+b)\cdot(c+d)}; \; P(B_2) = \frac{b \cdot d}{(a+b)\cdot(c+d)}.$$

Тогда

$$P(A) = \frac{a \cdot c}{(a+b)\cdot(c+d)} + \frac{b \cdot d}{(a+b)\cdot(c+d)} = \frac{a \cdot c + b \cdot d}{(a+b)\cdot(c+d)}. \; ☻$$

Задача 2.2.1. В условиях примера 2.2.1 найти вероятность того, что оба шара будут одного цвета, при конкретных значениях: $a = 2$, $b = 6$, $c = 5$, $d = 6$. ♦

Пример 2.2.2. Студент проходит тестирование, решая одну за другой две задачи. Вероятность решить первую задачу – 0.7, а вторую – 0.8. Экзамен считается сданным, если хотя бы одна задача решена студентом успешно. Найти вероятность положительного результата тестирования.

Решение. Введем простые случайные события: A = {первая задача решена}, B = {вторая задача решена}. Считаем, что они независимы. Очевидно, что искомое событие C = {хотя бы одна задача решена} является *суммой* этих простых событий. Совместны ли события A и B? Да, совместны. Поэтому применяем теорему сложения вероятностей в общей форме:

$P(C) = P(A+B) = P(A) + P(B) - P(A \cdot B) = 0.7 +$
$+ 0.8 - 0.7 \cdot 0.8 = 0.94.$ ☻

2.3. Условная вероятность. Умножение вероятностей

Пусть события A и B определены на одном пространстве элементарных исходов $\mathbf{\Omega}$, причем $P(B) \neq 0$. ***Условной вероятностью*** события A при условии, что событие B произошло, называется

$$\boxed{P(A/B) = \frac{P(A \cdot B)}{P(B)}}. \qquad (2.3.1)$$

Условному событию A/B благоприятствуют исходы, принадлежащие событию A и событию B одновременно, из числа исходов, благоприятствующих событию B, причем все исходы равновозможные. Тогда, согласно классическому определению вероятности,

$$\boxed{P(A/B) = \frac{k(AB)}{k(B)}}, \qquad (2.3.2)$$

где $k(AB)$ – число исходов, благоприятствующих событию AB;
$k(B)$ – число исходов, благоприятствующих событию B.

События A и B называются независимыми, если

$$\boxed{P(A \cdot B) = P(A) \cdot P(B)}. \qquad (2.3.3)$$

Тогда для независимых событий имеем:

$$\boxed{P(A/B) = \frac{P(A \cdot B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)}. \qquad (2.3.4)$$

Словами это означает, что вероятность независимого события A не зависит от того, произошло событие B или не произошло.

Пусть события $A_1, A_2, \ldots, A_n$ определены на одном и том же пространстве элементарных исходов $\mathbf{\Omega}$. ***События $A_1, A_2, \ldots, A_n$ называются попарно независимыми***, если для любых $i \neq j$ $(1 \leq i,\ j \leq n)$ выполняется условие

$$\boxed{P(A_i \cdot A_j) = P(A_i) \cdot P(A_j)}. \qquad (2.3.5)$$

Пример 2.3.1. Одновременно бросаются две монеты. Определены следующие события:

A_1 = {выпадение герба на первой монете};

A_2 = {выпадение цифры на второй монете};

A_3 = {выпадение хотя бы одного герба};

A_4 = {выпадение хотя бы одной цифры}.

Являются ли попарно независимыми данные события?

Решение. Для решения будем использовать формулу (2.3.5). Вначале вычислим безусловные вероятности:

$$P(A_1) = P(A_2) = \frac{1}{2}; \quad P(A_3) = P(A_4) = \frac{3}{4}.$$

Затем начинаем проверку соотношения (2.3.5) для разных пар событий:

- $P(A_1A_2) = \frac{1}{4}$, так как $P(A_1)P(A_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, то A_1 и A_2 – независимые;
- $P(A_1A_4) = \frac{1}{4}$, но так как $P(A_1)P(A_4) = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$, то A_1 и A_4 – зависимые;
- $P(A_2A_3) = \frac{1}{4}$, но так как $P(A_2)P(A_3) = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$, то A_2 и A_3 – зависимые;
- $P(A_1A_3) = \frac{1}{2}$, но так как $P(A_1)P(A_3) = \frac{3}{8}$, то A_1 и A_3 – зависимые;
- $P(A_2A_4) = \frac{1}{2}$, но $P(A_2)P(A_4) = \frac{3}{8}$ – A_2 и A_4 – зависимые;
- $P(A_3A_4) = \frac{1}{2}$, но $P(A_3)P(A_4) = \frac{9}{16}$ – A_3 и A_4 – зависимые.

Таким образом, независимы только события A_1 и A_2. ☻

Пример 2.3.2. Пусть при подбрасывании кубика определены случайные события: A = {выпадение 2}, B = {выпадение четной грани}. Определим условные и безусловные вероятности $P(A)$, $P(B)$, $P(A/B)$, $P(B/A)$.

Решение. Применяя непосредственно формулу (2.1.1), получаем безусловные вероятности: $P(A) = 1/6$, $P(B) = 3/6 = 1/2$. Если наступило событие B, то число элементарных исходов, благоприятствующих этому событию, равно 3 $\{2, 4, 6\}$, но при этом событию A благоприятствует только один элементарный исход $\{2\}$. Поэтому $P(A/B) = 1/3$. Далее, если наступило событие A (выпало 2), то всегда наступит событие B = {выпадение четного числа очков} – достоверное событие. Поэтому $P(B/A) = 1/1 = 1$. ☻

Для несовместных событий A_1 и A_2 имеет место равенство, являющееся прямым следствием (2.2.2):

$$\boxed{P((A_1 + A_2)/B) = P(A_1/B) + P(A_2/B)}.$$

Теорема умножения вероятностей

Пусть A и B – события, относящиеся к одному и тому же испытанию. Тогда вероятность совместного наступления этих событий равна

$$\boxed{P(A \cdot B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B)}. \qquad (2.3.6)$$

Если события A и B независимы, то

$$\boxed{P(A \cdot B) = P(A) \cdot P(B)}. \qquad (2.3.7)$$

Замечание 2.3.1. Теорема умножения позволяет вычислять вероятности сложных событий, представляя их через простые события, вероятности которых найти гораздо легче. ●

Пример 2.3.3. В урне 8 белых и 3 черных шара. Наугад достали (без возврата) 2 шара. Какова вероятность того, что оба они белые?

Решение. Введем простые события: A = {второй шар белый}, B = {первый шар белый}. Тогда нужное нам сложное со-

бытие представляется через A и B: $C = A \cdot B$. Следовательно, по теореме умножения вероятностей $P(C) = P(A \cdot B) = P(B) \cdot P(A/B) = 8/11 \cdot 7/10 = 28/55$. Ранее (см. пример 2.1.4) эта задача была решена с помощью комбинаторной схемы и получен такой же ответ. ☻

Обобщенная теорема умножения вероятностей. Вероятность произведения n событий $A_1, A_2, \ldots, A_n$ равна произведению вероятности одного из них на условные вероятности других, вычисленные в предположении, что все предыдущие события произошли:

$$\boxed{\begin{gathered} P(A_1 \cdot A_2 \cdot \ldots \cdot A_n) = \\ = P(A_1) \cdot P(A_2 / A_1) \cdot P(A_3 / A_1 A_2) \cdot \ldots \cdot P(A_n / A_1 A_2 \ldots A_{n-1}) \end{gathered}}. \quad (2.3.8)$$

Если события $A_1, A_2, \ldots, A_n$ взаимно независимы, то

$$\boxed{P(A_1 \cdot A_2 \cdot \ldots \cdot A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n)}. \quad (2.3.9)$$

Пример 2.3.4. Рассмотрим ситуацию вынимания одного счастливого билета (вероятность не зависит от порядка, очередности вытягивания жребия). 20 человек разыгрывают один приз. Изготовлены таблички (19 пустых, а на одной – крест) и помещены в барабан. Участники по очереди случайным образом достают их. У кого больше вероятность вытянуть призовую табличку – у первого или последнего?

Решение. Введем события: A_i = {i-й участник вытянул приз}, тогда $P(A_1) = 1/20$. Воспользуемся обобщенной теоремой умножения вероятностей и очевидными соотношениями: $A_2 = \overline{A}_1 \cdot A_2$, $A_3 = \overline{A}_1 \cdot \overline{A}_2 \cdot A_3$, …и т.д. Тогда

$P(A_2) = P(\overline{A}_1 \cdot A_2) = P(\overline{A}_1) \cdot P(A_2/\overline{A}_1) = (19/20) \cdot (1/19) = 1/20$.

$P(A_3) = P(\overline{A}_1 \cdot \overline{A}_2 \cdot A_3) = P(\overline{A}_1) \cdot P(\overline{A}_2/\overline{A}_1)\, P(A_3/\overline{A}_1 \cdot \overline{A}_2) =$
$= (19/20) \cdot (18/19) \cdot (1/18) = 1/20$ и т.д.

Таким образом, каждый участник лотереи имеет одинаковую вероятность вытянуть призовую табличку. ☻

Задача 2.3.1. В связке 7 ключей, из которых лишь один открывает дверь. Какова вероятность того, что для открывания двери придется перепробовать ровно 4 ключа, (т.е. дверь откроется с 4-й попытки)? ♦

Пример 2.3.6. Ребенок играет с карточками, на которых написаны буквы: «Т» – 3 карточки, «У» – 1, «И» – 2, «Н» – 1, «С» – 1. Найти вероятность того, что он случайно выложит слово «институт».

Решение. Воспользуемся обобщенной теоремой умножения. Для этого определим вероятность события A_1 = {на первой карточке буква И} $P(A_1) = 2/8$. Условная вероятность события A_2 = {на второй карточке буква Н, если первая карточка – буква И} $P(A_2 / A_1) = 1/7$ (одна карточка с буквой Н и осталось 7 карточек). По аналогии находим другие условные вероятности. Получаем

P = (2/8) · (1/7) · (1/6) · (3/5) · (1/4) · (2/3) · (1/2) · (1/1) = 1/3360 ~ ~0.0003. ☻

Пример 2.3.7. Из трех орудий производится залп по цели. Вероятности попадания при одном выстреле для каждого из этих орудий равны соответственно 0.9, 0.8 и 0.6. Найти вероятности следующих случайных событий:

B = {в цель попали ровно два орудия};

C = {в мишени после залпа в точности одна пробоина};

D = {все орудия промахнулись}:

Решение. Введем простые события.

A_1 = {первое орудие попало}, $\overline{A}_1$ = {первое орудие промахнулось}, $P(A_1) = 0.9$, $P(\overline{A}_1) = 1 - 0.9 = 0.1$;

A_2 = {второе орудие попало}, $\overline{A}_2$ = {второе орудие промахнулось}, $P(A_2) = 0.8$, $P(\overline{A}_2) = 1 - 0.8 = 0.2$;

A_3 = {третье орудие попало}, $\overline{A}_3$ = {третье орудие промахнулось}, $P(A_3) = 0.6$, $P(\overline{A}_3) = 1 - 0.6 = 0.4$.

События B, C и D легко представить через них:

$B = A_1 \cdot A_2 \cdot \overline{A}_3 + A_1 \cdot \overline{A}_2 \cdot A_3 + \overline{A}_1 \cdot A_2 \cdot A_3$;

$C = A_1 \cdot \overline{A}_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot A_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot \overline{A}_2 \cdot A_3$;

$D = \overline{A}_1 \cdot \overline{A}_2 \cdot \overline{A}_3$.

Теперь, используя теоремы о сложении вероятностей и об умножении вероятностей, получаем:

$P(B) = P(A_1 \cdot A_2 \cdot \overline{A}_3 + A_1 \cdot \overline{A}_2 \cdot A_3 + \overline{A}_1 \cdot A_2 \cdot A_3)$ = {по теореме сложения вероятностей для несовместных событий} = $= P(A_1 \cdot A_2 \cdot \overline{A}_3) + P(A_1 \cdot \overline{A}_2 \cdot A_3) + P(\overline{A}_1 \cdot A_2 \cdot A_3)$ = {по теореме умножения вероятностей для независимых событий} = $P(A_1) \cdot P(A_2) \cdot P(\overline{A}_3) + P(A_1) \cdot P(\overline{A}_2) \cdot P(A_3) + P(\overline{A}_1) \cdot P(A_2) \cdot P(A_3) = 0.9 \cdot 0.8 \cdot 0.4 + 0.9 \cdot 0.2 \cdot 0.6 + 0.1 \cdot 0.8 \cdot 0.6 = 0.288 + 0.108 + 0.048 = 0.444$;

$P(C) = P(A_1 \cdot \overline{A}_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot A_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot \overline{A}_2 \cdot A_3)$ = {по теореме сложения вероятностей для несовместных событий} = $P(A_1 \cdot \overline{A}_2 \cdot \overline{A}_3) + P(\overline{A}_1 \cdot A_2 \cdot \overline{A}_3) + P(\overline{A}_1 \cdot \overline{A}_2 \cdot A_3)$ = {по теореме умножения вероятностей для независимых событий} = $P(A_1) \cdot P(\overline{A}_2) \cdot P(\overline{A}_3) + P(\overline{A}_1) \cdot P(A_2) \cdot P(\overline{A}_3) + P(\overline{A}_1) \cdot P(\overline{A}_2) \cdot P(A_3) = 0.9 \cdot 0.2 \cdot 0.4 + 0.1 \cdot 0.8 \cdot 0.4 + 0.1 \cdot 0.2 \cdot 0.6 = 0.072 + 0.012 + 0.032 = 0.116$;

$P(D) = P(\overline{A}_1 \cdot \overline{A}_2 \cdot \overline{A}_3)$ = {по теореме умножения вероятностей для независимых событий} = $P(\overline{A}_1) \cdot P(\overline{A}_2) \cdot P(\overline{A}_3) = 0.1 \cdot 0.2 \cdot 0.4 = 0.008$.

Последняя из найденных вероятностей позволяет, применяя формулу для вероятности противоположного события, легко найти вероятность события E = {цель поражена}. $P(E) = 1 - P(\overline{E}) = 1 - P(D) = 1 - 0.008 = 0.992$, так как $\overline{E} = D$. ☻

Замечание 2.3.2. Если в определении случайного события стоит формулировка «хотя бы один раз» или «не менее одного раза», то вычисление вероятности такого события (например, события A) лучше осуществлять через вероятность противоположного события $\overline{A}$ (которая вычисляется проще) по формуле

$$P(A) = 1 - P(\overline{A}). \qquad (2.3.10)$$

Противоположное событие в данном случае определяется формулировками «ни одного события» или «ни одного случая». Этот прием иллюстрируется следующим примером. ●

Пример 2.3.8. Студент знает 10 из 30 вопросов программы. Используя теорему умножения вероятностей, определить вероятность того, что из трех предложенных ему экзаменатором вопросов студент знает:

а) все три вопроса;

б) хотя бы один вопрос.

Решение. Пусть событие A = {студент знает все три вопроса}. Тогда по теореме умножения вероятностей

$$P(A)=\frac{10}{30}\cdot\frac{9}{29}\cdot\frac{8}{28}\approx 0,03.$$

Пусть событие B = {студент знает хотя бы один вопрос}.

Перейти к противоположному событию $\overline{B}$ = {студент не знает ни одного вопроса}, т.е. он три раза вытягивает незнакомые вопросы. $P(\overline{B})=\frac{20}{30}\cdot\frac{19}{29}\cdot\frac{18}{28}$. Тогда вероятность события B вычисляется по формуле:

$$P(B)=1-P(\overline{B})=1-\frac{20}{30}\cdot\frac{19}{29}\cdot\frac{18}{28}\approx 0,719.$$

Вопросы и задачи для самопроверки

1. Что такое вероятность случайного события и как она вычисляется?
2. Свойства вероятности случайного события.
3. Чему равны вероятности: $P(\varnothing), P(\Omega)$?
4. Бросаются две монеты. Найти вероятности следующих событий:

   A = {монеты лягут одинаковыми сторонами};

   B = {монеты лягут разными сторонами}.
5. В урне a белых и b черных шаров. Из урны вынимают наугад один шар. Найти вероятность того, что этот шар белый.
6. Игральная кость бросается один раз. Найти вероятность следующих событий:

   A = {появление четного числа очков};

   B = {появление не менее 5 очков};

C = {появление не более 5 очков};

Ответ: 1/2, 1/3, 5/6.

7. Сформулировать теорему сложения вероятностей (отдельно для случая несовместных событий).
8. Как вычислить вероятность противоположного события?
9. В урне 6 белых и 8 черных шаров. Из урны вынимаются два шара. Найти вероятность того, что шары будут:

   а) одного цвета;

   б) разных цветов.

   Рекомендация: использовать комбинаторную схему и учесть замечание 1.3.2.
10. Какая вероятность называется условной?
11. Дать определение независимости случайных событий.
12. Дать определение независимости событий через вероятности и условные вероятности.
13. Брошены две игральные кости. Найти условную вероятность того, что выпали две пятерки, если известно, что сумма выпавших очков делится на пять.

    Ответ: 1/3.
14. Сформулировать теорему умножения вероятностей.
15. Сформулировать теорему умножения вероятностей для независимых событий.
16. Студент знает 20 из 25 вопросов программы. Найти вероятность того, что студент не знает хотя бы один из трех предложенных ему вопросов.

    Ответ: ~0.515.
17. Из трех орудий произведен залп по цели. Вероятности попадания в цель при одном выстреле из первого, второго и третьего орудий равны соответственно 0.9, 0.8, 0.6. Найти вероятность того, что только одно орудие попадет в цель.

    Ответ: ~0.116.

Тема 3. Формула Бернулли. Формула полной вероятности. Формула Байеса

3.1. Повторение испытаний. Формула Бернулли

Предположим, что один и тот же опыт производится при неизменных условиях n раз. В каждом из этих опытов случайное событие A может произойти с вероятностью p и не произойти с вероятностью, соответственно, $q = (1 - p)$. Вероятность того, что в этих n испытаниях событие A произойдет ровно k раз ($0 \le k \le n$), вычисляется по формуле Бернулли:

$$\boxed{P_n(k) = C_n^k p^k (1-p)^{n-k}}. \qquad (3.1.1)$$

Пример 3.1.1. Найти вероятность того, что при 10 подбрасываниях монеты мы получим ровно 5 гербов.

Решение. A = {выпадение герба в одном единичном испытании}, вероятность этого события равна $p = 0.5$, $q = 0.5$, $n = 10$, $k = 5$. Тогда

$$P_{10}(5) = C_{10}^5 0.5^5 (1-0.5)^{10-5} = 36 \cdot 7/2^{10} = 0.245. \ ☻$$

Ценность формулы Бернулли состоит в том, что она дает ответ в случаях, когда из-за слишком большого числа элементарных исходов обычные комбинаторные способы подсчета вариантов неприменимы.

Пример 3.1.2. В мастерской работают 12 мастеров. Вероятность того, что мастер находится на рабочем месте, равна 0.8. Найти вероятность того, что, случайно зайдя в мастерскую, мы застанем на рабочих местах не менее 10 мастеров.

Решение. Введем случайные события:

A_1 = {на рабочем месте 10 мастеров}, $P(A_1) = P_{12}(10)$;

A_2 = {на рабочем месте 11 мастеров}, $P(A_2) = P_{12}(11)$;

A_3 = {на рабочем месте 12 мастеров}, $P(A_3) = P_{12}(12)$.

Тогда искомое случайное событие можно представить суммой $A = A_1 + A_2 + A_3$, а вероятность $P(A)$ вычислить по теореме сложения вероятностей для несовместных событий:

$$P(A) = P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3).$$

Применяя формулу Бернулли для вычисления вероятностей $P_{12}(10)$, $P_{12}(11)$, $P_{12}(12)$, получаем:

$$P(A) = P_{12}(10) + P_{12}(11) + P_{12}(12) =$$
$$= C_{12}^{10} 0.8^{10} (1-0.8)^{12-10} + C_{12}^{11} 0.8^{11} (0.2)^1 + C_{12}^{12} 0.8^{12} (0.2)^0 =$$
$$= 66 \cdot 0.8^{10} \cdot 0.04 + 12 \cdot 0.8^{11} \cdot 0.2 + 0.8^{12} =$$
$$= 66 \cdot 0.107 \cdot 0.04 + 12 \cdot 0.086 \cdot 0.2 + 0.069 = 0.558. \quad ☻$$

Пример 3.1.3. Найти вероятность того, что при пяти бросаниях монеты герб выпадет:

а) три раза;

б) не менее трех раз.

Решение. Испытание – подбрасывание монеты – повторяется $n = 5$ раз. «Успех» – выпадение герба, вероятность «успеха» $p = \frac{1}{2}$. Тогда $q = 1 - p = \frac{1}{2}$. По формуле Бернулли:

а) $P_5(3) = C_5^3 \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^2 = 10 \cdot \left(\frac{1}{2}\right)^5 = \frac{5}{16}$;

б) вероятность того, что герб выпадет не менее трех раз, определяется как сумма вероятностей $P_5(3) + P_5(4) + P_5(5) =$

$$= C_5^3 \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^2 + C_5^4 \left(\frac{1}{2}\right)^5 + C_5^5 \left(\frac{1}{2}\right)^5 = 16 \cdot \frac{1}{32} = \frac{1}{2}. \quad ☻$$

Пример 3.1.4. Завод производит в среднем 80 % продукции высшего сорта. Определить вероятность того, что среди взятых на проверку десяти изделий будет не менее двух изделий высшего сорта.

Решение. Введем величины $n = 10$, вероятность появления деталь высшего сорта $p = 0.8$, $q = 1 - p = 0.2$. Вероятность события A = {среди десяти изделий будет не менее двух изделий высшего сорта}будем определять через вероятность противоположного события $\overline{A}$ = {среди десяти изделий будет одно или ни одного изделия высшего сорта} (см. замечание 2.3.2). Вероятность этого события равна $P(\overline{A}) = P_{10}(0) + P_{10}(1) = 9 \cdot 0.2^9$. Тогда

$$P(A) = 1 - P(\overline{A}) = 1 - 9 \cdot 0,2^9 \approx 1. \quad ☻$$

3.2. Формула полной вероятности

Предположим, что все пространство элементарных исходов Ω можно разложить на n взаимоисключающих, ***несовместных гипотез (случайных событий)***:

$$\Omega = H_1 + H_2 + H_3 + \ldots + H_n,$$

и для всех H_k вероятности $P(H_k) > 0$. Очевидно, что гипотезы H_i образуют ***полную группу событий***, т.е. их попарные произведения $H_i \cdot H_j$, $i \neq j$ дают невозможные события $\varnothing$, а в сумме они дают Ω, поэтому

$$\sum_{k=1}^{n} P(H_k) = 1.$$

Событие A может произойти с одним из событий H_k ($k = 1, 2, \ldots, n$) с вероятностями $P(A/H_k) \geq 0$. Тогда вероятность события A определяется по ***формуле полной вероятности***:

$$\boxed{P(A) = \sum_{k=1}^{n} P(H_k) \cdot P(A/H_k)}. \qquad (3.2.1)$$

Пример 3.2.1. На фабрике первый станок производит 25 % деталей, второй – 35 %, а третий – 40 %. Брак в их продукции составляет соответственно 5, 4 и 3 %. Какова вероятность, что случайно отобранная на складе готовой продукции деталь дефектна?

Решение. Введем гипотезы и найдем их вероятности:

H_1 = {взятая деталь изготовлена на 1-м станке};

$P(H_1) = 0.25$, H_2 = {взятая деталь изготовлена на 2-м станке};

$P(H_2) = 0.35$, H_3 = {взятая деталь изготовлена на 3-м станке}, $P(H_3) = 0.40$.

Считая, что A = {взятая деталь дефектна}, вычислим условные вероятности: $P(A/H_1) = 0.05$, $P(A/H_2) = 0.04$, $P(A/H_3) = 0.03$. Гипотезы H_i образуют полную группу событий (убедитесь в этом). Тогда для нахождения $P(A)$ можно применить формулу полной вероятности:

$$P(A) = 0.25 \cdot 0.05 + 0.35 \cdot 0.04 + 0.4 \cdot 0.03 = 0.0385. \ ☻$$

Пример 3.2.2. Человек одет в костюм, имеющий два кармана: правый и левый. В правом кармане 5 десятирублевых купюр и 3 сторублевые купюры. В левом кармане – 9 десятирублевых купюр и 2 сторублевые купюры. Найти вероятность того, что, выбрав наугад карман и достав из него случайным образом купюру, он вынет «десятку».

Решение. Введем гипотезы и найдем их вероятности:

H_1 = {выбран правый карман}, $P(H_1) = 0.5$;

H_2 = {выбран левый карман}, $P(H_2) = 0.5$.

Считая, что A = {взятая купюра – «десятка»}, вычислим условные вероятности: $P(A/H_1) = 5/8$, $P(A/H_2) = 9/11$. Так как гипотезы H_i образуют полную группу событий, то применяем формулу полной вероятности:

$$P(A) = 0.5 \cdot 5 / 8 + 0.5 \cdot 9 / 11 = 0.722. \quad ☻$$

Пример 3.2.3. Завод выпускает изделия определенного типа, причем каждое изделие имеет дефект с вероятностью 0.1. Изделие осматривается контролером; он обнаруживает имеющийся дефект с вероятностью 0.95. Кроме того, с вероятностью 0.03 контролер может по ошибке забраковать бездефектное изделие. Найти вероятность того, что изделие на контроле будет забраковано.

Решение. Обозначим событие A = {изделие на контроле будет забраковано}. Введем следующие гипотезы и вычислим их вероятности:

H_1 = {изделие имеет дефект}, $P(H_1) = 0.1$;

H_2 = {изделие не имеет дефекта}, $P(H_2) = 1 - 0.1 = 0.9$.

Тогда по формуле полной вероятности имеем:

$$P(A) = 0.1 \cdot 0.95 + 0.9 \cdot 0.03 = 0{,}122. \quad ☻$$

3.3. Формула Байеса

Эта формула является в некотором смысле обратной к формуле полной вероятности (3.2.1). Пусть $H_1, H_2, ..., H_n$ – полная группа событий. В дальнейшем H_k будем называть k-й гипотезой. Вероятности $P(H_k)$ $(k = 1, 2, ..., n)$ называют ***априорными***

(доопытными) вероятностями. Спрашивается, как изменится вероятность гипотезы $P(H_k)$, если стало известно, что событие A произошло. Другими словами, нужно вычислить условные вероятности $P(H_k/A)$, которые называют ***апостаприорными вероятностями.*** Имеет место следующая ***формула Байеса***:

$$\boxed{P(H_k/A) = \frac{P(H_k)\cdot P(A/H_k)}{\sum_{k=1}^{n} P(H_k)\cdot P(A/H_k)}}, \ (k = 1, 2, \ldots, n). \qquad (3.3.1)$$

Пример 3.3.1. В эксперименте участвуют два стрелка. Хороший стрелок попадает в цель с вероятностью 0.9, плохой – с вероятностью 0.3. Подбрасывается монета, при выпадении герба стреляет хороший стрелок, при выпадении цифры – плохой. Наблюдатель не знает результата бросания монеты и не видит, кто стреляет. Какова вероятность того, что стрелял хороший стрелок, если наблюдатель зафиксировал попадание в цель?

Решение. Пусть A = {зафиксировано попадание в цель}. Сформулируем следующие гипотезы, составляющие полную группу:

H_1 = {стрелял хороший стрелок};

H_2 = {стрелял плохой стрелок}.

Очевидно, что априорные вероятности этих гипотез равны: $P(H_1) = P(H_2) = \frac{1}{2}$ (бросание монеты). Условные вероятности $P(A/H_1) = 0.9;\ P(A/H_2) = 0.3$. Применяя формулу (3.3.1), получаем:

$$P(H_1/A) = \frac{P(H_1)\cdot P(A/H_1)}{P(A)} = \frac{\frac{1}{2}\cdot 0.9}{\frac{1}{2}\cdot 0.9 + \frac{1}{2}\cdot 0.3} = 0.75. \ ☻$$

Пример 3.3.2. Изделие проверяется на стандартность одним из двух товароведов. Первый товаровед проверяет 60 % всех изделий, а второй – 40 %. Вероятность того, что стандартное изделие будет признано нестандартным первым товароведом, равна 0.02, а вторым – 0.1. Стандартное изделие при про-

верке было признано стандартным. Найти вероятность того, что это изделие проверял второй товаровед.

Решение. Произошло событие A = {стандартное изделие при проверке признано стандартным}. Определим следующие гипотезы, составляющие полную группу и их вероятности:

H_1 = {изделие проверял первый товаровед}, $P(H_1) = 0.6$;

H_2 = {изделие проверял второй товаровед}, $P(H_2) = 0.4$.

Далее определим условные вероятности:

$P(A/H_1) = 1 - 0.02 = 0.98$ – вероятность того, что первый контролер признает изделие стандартным;

$P(A/H_2) = 1 - 0.1 = 0.90$ – вероятность того, что второй контролер признает изделие стандартным.

Тогда по формуле (3.3.1) получаем:

$$P(H_2/A) = \frac{0.4 \cdot 0.9}{0.6 \cdot 0.98 + 0.4 \cdot 0.9} \approx 0.38,$$

т.е. с вероятностью 0.38 деталь была проверена и признана стандартной вторым контролером. ☻

Вопросы и задачи для самопроверки

1. В какой ситуации применима формула Бернулли?
2. Дать определение основных параметров формулы Бернулли – n, p, k.
3. Известно, что 97 % выпускаемых заводом изделий отвечает стандарту. Найти вероятность того, что среди 10 проверенных изделий будет не более 2 непригодных.
4. Прибор состоит из трех блоков, исправность каждого из которых необходима для работы прибора. Блоки выходят из строя независимо друг от друга с вероятностью 0.2 каждый. Найти вероятность того, что прибор выйдет из строя.
5. Что называется полной группой событий?
6. Сформулировать формулу полной вероятности.
7. Магазин получил три партии изделий. В двух партиях все изделия доброкачественные, а в одной – 15 % бракованных. Для контроля из наугад выбранной партии берется одно из-

делие. Какова вероятность того, что это изделие бракованное?

Ответ: $P(\{\text{изделие бракованное}\}) = 0.05$.

8. Чем гипотезы в формуле полной вероятности отличаются от элементарных исходов?
9. Некоторое изделие выпускается двумя заводами, причем объем продукции второго завода в три раза превосходит объем продукции первого. Доля брака у первого завода 1 %, у второго – 2 %. Изделия, выпущенные заводами за одинаковый промежуток времени, перемешали и в таком виде пустили в продажу. Какова вероятность того, что приобретенное изделие окажется бракованным?

   Ответ: 0.0175.
10. Вероятности перегорания первой, второй и третьей ламп равны соответственно 0.1, 0.2 и 0.3. Вероятности выхода прибора из строя при перегорании одной, двух и трех ламп равны соответственно 0.3, 0.6 и 0.9. Найти вероятность выхода прибора из строя.

    Ответ: 0.18.
11. По самолету производится три одиночных выстрела. Вероятность попадания при первом выстреле равна 0.4, при втором – 0.5, при третьем – 0.8. Для вывода самолета из строя достаточно трех попаданий. При двух попаданиях самолет выходит из строя с вероятностью 0.6, при одном попадании – с вероятностью 0.2. Найти вероятность того, что в результате трех выстрелов самолет будет выведен из строя.

    Ответ: 0.492. ♦
12. Что такое апостериорная вероятность гипотезы H_i?
13. Запишите и поясните формулу Байеса.
14. Из 18 стрелков 5 стрелков (первая группа) попадают в мишень с вероятностью 0.8; 7 стрелков (вторая группа) – с вероятностью 0.7; 4 стрелка (третья группа) – с вероятностью 0.6; 2 стрелка (четвертая группа) – с вероятностью 0.5. Наудачу выбранный стрелок произвел выстрел, но в мишень не попал. Из какой группы, вероятнее всего, был этот стрелок?

Рекомендация: нужно определить четыре условные вероятности $P(H_k/A)$, наибольшая из которых укажет на номер группы.

Ответ: наиболее вероятно, что промахнулся стрелок из второй группы.

15. Изделие проверяется на стандартность одним из двух товароведов. Первый товаровед проверяет 70 % всех изделий, а второй – 30 %. Вероятность того, что стандартное изделие будет признано нестандартным первым товароведом, равна 0.05, а вторым – 0.1. Стандартное изделие при проверке было признано стандартным. Найти вероятность того, что это изделие проверял первый товаровед.

    Рекомендация: посмотрите решение примера 3.3.2.

Тема 4. Дискретные случайные величины

4.1. Определение дискретной случайной величины

Ранее были изучены свойства случайных событий. Имеется частный класс случайных событий, когда каждому элементарному исходу можно сопоставить некоторое определенное *число*.

Случайной величиной называется функция, определенная на пространстве элементарных исходов Ω и принимающая действительные значения, т.е.

$$X\colon\ \omega_i \to X(\omega_i) \in R.$$

Будем обозначать случайные величины большими буквами, находящимися в конце латинского алфавита.

Пример 4.1.1. При подбрасывании кубика естественным образом возникает случайная величина X = {значения, которые могут выпасть на кубике} = {1, 2, 3, 4, 5, 6}. ☻

Пример 4.1.2. Пусть опыт состоит в подбрасывании двух монет. Будем фиксировать число выпавших гербов. Получим случайную величину Z, имеющую значения Z = {0, 1, 2}. ☻

Дискретной случайной величиной называется случайная величина, имеющая конечное число значений. В общем случае к

классу дискретных относят и случайные величины, имеющие бесконечное число значений, но которые можно перенумеровать с помощью натурального ряда. Говорят, что такие случайные величины имеют счетное число значений.

Безусловно, в рассмотренных примерах 4.1.1, 4.1.2 имеем дело с дискретными случайными величинами.

Закон распределения дискретной случайной величины. **Дискретная случайная величина полностью определяется своим** *законом распределения*. **Закон распределения (иногда его называют** *ряд распределения*) **– это таблица следующего вида:**

X	x_1	x_2	...	x_i	...	x_n
p	p_1	p_2	...	p_i	...	p_n

,

где x_i (первая строка таблицы) – все значения случайной величины, выстроенные в порядке возрастания; p_i (вторая строка таблицы) – вероятности соответствующих значений случайной величины.

Важным свойством закона распределения, часто используемым для проверки и решения задач, является равенство единице полной вероятности или суммы всех p_i:

$$\boxed{\sum p_i = 1}. \qquad (4.1.1)$$

Словами, это означает, что обязательно в опыте появится какое-либо значение дискретной случайной величины, заданное законом распределения.

Построим закон распределения для случайных величин из рассмотренных примеров.

Пример 4.1.3. Пусть случайная величина X = {значения, которые могут выпасть на верхней грани кубика} = {1, 2, 3, 4, 5, 6}. Построить закон распределения этой случайной величины.

Решение. Все значения случайной величины равновероятны (вспомним определение вероятности случайного события (2.1.1)) и вероятность каждого из возможных значений

x_k $(k = 1, 2, ..., 6)$ равна $p_k = p = \frac{1}{6}$. Поэтому закон распределения задается следующей таблицей:

X	1	2	3	4	5	6
p	1/6	1/6	1/6	1/6	1/6	1/6

Выполним проверку: $\sum_{k=1}^{6} p_k = 1$. ☻

Пример 4.1.4. Пусть случайная величина Z = {число выпавших гербов при подбрасывании двух монет} (см. пример 4.1.2). Построить закон распределения этой случайной величины.

Решение. Первоначально определим возможные значения случайной величины Z. Очевидно, что при подбрасывании двух монет герб может появиться 2 раза, один раз или вообще не появиться. Поэтому $Z = \{0, 1, 2\}$. Вероятность значения 0 определяется по теореме умножения вероятностей двух независимых событий – при каждом подбрасывании выпала цифра $P(Z = 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Вероятность значения 1 определяется как сумма вероятностей двух событий – на первой и второй монете один герб и одна цифра: $P(Z = 1) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$. Наконец, вероятность значения 2 определяется по теореме умножения вероятностей двух независимых событий – при каждом подбрасывании выпала герб: $P(Z = 2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Поэтому закон распределения задается следующей таблицей:

Z	0	1	2
p	1/4	1/2	1/4

Выполним проверку: $\sum_{k=1}^{3} p_k = 1$. ☻

4.2. Биномиально распределенные случайные величины

Рассмотрим ситуацию повторения некоторых фиксированных испытаний. В каждом из этих n одинаковых опытов случайное событие A может произойти с вероятностью p, а может и не произойти – с вероятностью $1-p$. Введем случайную величину S_n как количество появлений события A в этой цепочке испытаний (или как число опытов, в которых событие A произошло). Такая дискретная случайная величина называется распределенной по биномиальному закону, или ***биномиальной случайной величиной.*** Случайная величина S_n принимает целые неотрицательные значения $i = 0, 1, 2, ..., n$ с вероятностями $P_n(i)$, которые вычисляются по формуле Бернулли:

$$\boxed{P(S_n = i) = P_n(i) = C_n^i p^i (1-p)^{n-i}}.$$

Поэтому закон распределения случайной величины S_n задается таблицей:

S_n	0	1	2	$\dots i \dots$	n
p	$(1-p)^n$	$np(1-p)^{n-1}$	$C_n^2 p^2 (1-p)^{n-2}$	$C_n^i p^i (1-p)^{n-i}$	p^n

Пример 4.2.1. В урне 2 белых и 4 черных шара. Выбираем наугад с возвращением 2 шара. Требуется составить закон распределения случайной величины Y – количества белых шаров в этой выборке.

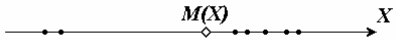
Решение. Очевидно, что $Y = S_2$ – биномиальная случайная величина, при этом $p = 2/6 = 1/3$, $1 - p = 2/3$, следовательно, закон распределения этой случайной величины представим таблицей:

$Y = S_2$	0	1	2
p	4/9	4/9	1/9

Выполним проверку: $\sum_{k=1}^{3} p_k = 1$. ☻

4.3. Математическое ожидание случайной величины

Математическое ожидание случайной величины имеет смысл среднего значения, вокруг которого группируются все значения этой случайной величины, которые она может принимать. В механике – это центр тяжести системы точек. Нередко данное значение является также и наиболее вероятным, самым *ожидаемым*. Представим геометрический смысл математического ожидания на схеме, где точками показаны возможные значения случайной величины:



Для дискретной случайной величины X математическое ожидание определяется и вычисляется по формуле

$$\boxed{M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i}. \qquad (4.3.1)$$

З*амечание 4.3.1.* Само математическое ожидание ***не является случайной величиной***. Значение математического ожидания удовлетворяет следующему неравенству:

$$\boxed{\min\{x_i\} \le M(X) \le \max\{x_i\}}. \qquad (4.3.2)$$

Для вычисления математического ожидания достаточно знать закон распределения случайной величины. ●

Пример 4.3.1. Найдем математическое ожидание случайной величины X – значения грани, выпавшей на кубике. Используя построенный ранее закон распределения этой случайной величины (см. пример 4.1.3), получаем:

$$M(X) = 1/6 + 2/6 + 3/6 + 4/6 + 5/6 + 6/6 = 21/6 = 3.5.$$

Найденное математическое ожидание удовлетворяет неравенству (4.3.2). ☻

Пример 4.3.2. Найдем математическое ожидание случайной величины Z, возникающей при подбрасывании двух монет и равной числу выпавших гербов (см. пример 4.1.4):

$$M(Z) = 0\cdot 1/4 + 1\cdot 1/2 + 2\cdot 1/4 = 1.$$

Найденное математическое ожидание удовлетворяет неравенству (4.3.2). ☻

Свойства математического ожидания. Приведем без доказательства (которое легко выполняется с учетом (4.3.1)) следующие свойства математического ожидания:

1. $M(\text{const}) = \text{const}$.
2. $M(CX) = CM(X)$.
3. $M(X + Y) = M(X) + M(Y)$ для любых двух случайных величин X и Y.
4. Пусть случайные величины X и Y *независимы*. Тогда $M(X \cdot Y) = M(X) \cdot M(Y)$.

Пояснение. Случайные величины X и Y называются *независимыми* тогда и только тогда, когда случайные события $\{X = x_i\}$ и $\{Y = y_j\}$ при любых i и j являются независимыми. Из данного определения следует, что при любых i и j вероятность события $\{X = x_i\}$ не зависит от того, произошло событие $\{Y = y_j\}$ или нет. Другими словами, независимые случайные величины X и Y не могут влиять друг на друга, взаимовлияние отсутствует. ●

Пример 4.3.3. Пусть требуется найти математическое ожидание суммы значений, выпавших на двух одновременно подброшенных кубиках.

Решение. Пусть случайная величина X = {значения, которые могут выпасть на верхней грани первого кубика}, случайная величина Y = {значения, которые могут выпасть на верхней грани второго кубика}. Найти $M(X + Y)$ суммы этих двух случайных величин можно двумя способами. *Первый способ* (сложный) – построить закон распределения случайной величины $Z = X + Y$, а затем вычислить искомое математическое ожидание. *Второй способ* (простой) – используя свойство 3. Из примера 4.3.1 следует, что $M(X) = M(Y) = 3.5$. Тогда $M(X + Y) = 3.5 + 3.5 = 7$. ☻

Математическое ожидание биномиальной случайной величины также легко получить, пользуясь свойством 3:

$$\boxed{M(S_n) = n \cdot p}. \qquad (4.3.3)$$

Пример 4.3.4. Дана случайная величина $Z = 3 + 2 \cdot X$, где X – случайная величина с $M(X) = 5$. Вычислить математическое ожидание величины Z.

Решение. Используя свойства 1 и 2 математического ожидания, имеем:

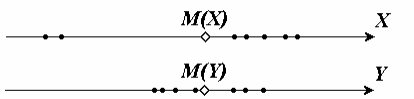
$$M(Z) = M(3 + 2 \cdot X) = M(3) + M(2 \cdot X) =$$
$$= 3 + 2 \cdot M(X) = 3 + 2 \cdot 5 = 13.$$

☻

4.4. Дисперсия случайной величины

Дисперсия случайной величины является мерой разброса (рассеяния) значений этой случайной величины вокруг ее математического ожидания.

Рассмотрим геометрическую схему, где точками показаны возможные значения случайных величин X и Y:



Показано, что эти величины имеют одинаковые математические ожидания, но разный разброс вокруг этих математических ожиданий. Вопрос: какая случайная величина имеет большую дисперсию? *Ответ*: $D(X) > D(Y)$ (убедитесь в этом).

Для дискретной случайной величины X дисперсия определяется и вычисляется по формуле

$$\boxed{D(X) = \sum_{i=1}^{n} (x_i - M(X))^2 p_i}. \qquad (4.4.1)$$

Часто, особенно при ручном счете, используется следующая формула:

$$\boxed{\begin{aligned} &D(X) = M(X^2) - (M(X))^2 = \sum_{i=1}^{n} x_i^2 p_i - (M(X))^2 = \\ &= x_1^2 \cdot p_1 + x_2^2 \cdot p_2 + \ldots + x_n^2 \cdot p_n - (M(X))^2. \end{aligned}} \qquad (4.4.2)$$

Для вычисления дисперсии достаточно знать закон распределения случайной величины и вычислить значение математического ожидания случайной величины.

Замечание 4.4.1. Значение дисперсии случайной величины не является случайной величиной. ●

Замечание 4.4.2. Математическое ожидание и дисперсия являются ***числовыми характеристиками*** случайной величины, так как они характеризуют статистические свойства случайной величины числовыми значениями. ●

Пример 4.4.1. Найдем дисперсию случайной величины X – значения грани, выпавшей на кубике.

Решение. Используя построенный ранее закон распределения этой случайной величины (см. пример 4.1.3) и вычисленное ее математическое ожидание (см. пример 4.3.1), получаем:

$$D(X) = 1/6 + 4/6 + 9/6 + 16/6 + 25/6 + 36/6 - 3.5^2 =$$
$$= 91/6 - 12.25 \cong 15.17 - 12.25 = 2.92. \quad ☻$$

Пример 4.4.2. Найти дисперсию случайной величины Z, возникающей при подбрасывании двух монет и равной числу выпавших гербов.

Решение. Используя построенный ранее закон распределения этой случайной величины (см. пример 4.1.4)) и вычисленное ее математическое ожидание (см. пример 4.3.2), получаем:

$$D(Z) = 0 \cdot 1/4 + 1 \cdot 1/2 + 4 \cdot 1/4 - 1^2 = 1/2. \quad ☻$$

Свойства дисперсии. Приведем следующие свойства:

1. $D(X) \geq 0$.
2. $D(\text{const}) = 0$.
3. $D(CX) = C^2 D(X)$.

Эти три свойства доказываются по определению (4.4.1).

4. Пусть случайные величины X и Y – *независимы*. Тогда

$$D(X + Y) = D(X) + D(Y).$$

Пример 4.4.3. Пусть требуется найти дисперсию суммы значений, выпавших на двух одновременно подброшенных кубиках.

Решение. Пусть случайная величина X = {значения верхней грани первого кубика}, случайная величина Y = {значения верхней грани второго кубика}. Дисперсии этих случайных величин одинаковы и равны 2.92. Так как эти величины независимы, то на основании свойства 4

$$D(X + Y) = 2.92 + 2.92 = 5.84. \quad ☻$$

Дисперсию биномиальной случайной величины также легко получить, пользуясь свойством 4:

$$\boxed{D(S_n) = n\,p\,(1 - p) = npq}. \qquad (4.4.3)$$

На практике в прикладных исследованиях чаще используется тесно связанная с дисперсией величина $\sigma(X) = \sqrt{D(X)}$, называемая ***среднеквадратическим отклонением*** рассматриваемой случайной величины. Удобство ее использования состоит в том, что она измеряется в тех же самых единицах, что и случайная величина X.

Пример 4.4.4. Случайные величины X и Y независимы и имеют следующие числовые характеристики: $M(X) = 1$, $D(X) = 3$, $M(Y) = 2$, $D(Y) = 4$. Вычислить математическое ожидание и дисперсию случайной величины $U = 3 \cdot X - 2 \cdot Y$.

Решение. Сначала вычислим математическое ожидание $M(U)$, используя известные свойства математического ожидания (см. также пример 4.3.4):

$$M(U) = M(3 \cdot X - 2 \cdot Y) = 3 \cdot M(X) - 2 \cdot M(Y) = 3 \cdot 1 - 2 \cdot 2 = -1.$$

Затем вычислим дисперсию $D(U)$, используя известные свойства дисперсии:

$$D(U) = D(3 \cdot X - 2 \cdot Y) = 3^2 \cdot D(X) + (-2)^2 \cdot D(Y) = \\ = 9 \cdot 3 + 4 \cdot 4 = 43. \quad ☻$$

4.5. Функция распределения случайной величины

Функцией распределения $F(x)$ случайной величины X называется функция, равная вероятности события

$A = \{$*значения случайной величины X меньше заданного значения x*$\}$ и определяемая формулой:

$$\boxed{F(x) = P(X < x)}. \qquad (4.5.1)$$

Как и закон распределения, функция распределения целиком и полностью, исчерпывающим образом описывает все свойства и поведение рассматриваемой случайной величины. Функция распределения или закон распределения случайной величины дает полную информацию о ней.

Пример 4.5.1. Построить функцию распределения дискретной случайной величины X – значения грани, выпавшей на кубике.

Решение. Будем использовать закон распределения этой случайной величины, полученный в примере 4.1.3. Выполним следующие рассуждения:

$P(\{X < 1\}) = 0$ (нет значений случайной величины, меньших 1), поэтому при $x \le 1$ $F(x) = 0$;

$P(\{X < 2\}) = 1/6$ (только одно значение (1) удовлетворяет неравенству $X < 2$), поэтому при $1 < x \le 2$ $F(x) = 1/6$;

$P(\{X < 3\}) = 1/6 + 1/6$ (только два значения – 1 и 2 – удовлетворяют неравенству $X < 3$), поэтому при $2 < x \le 3$ $F(x) = 2/6 = 1/3$;

$P(\{X < 4\}) = 1/6 + 1/6 + 1/6$ (только три значения – 1, 2, 3 – удовлетворяют неравенству $X < 4$), поэтому при $3 < x \le 4$ $F(x) = 3/6 = 1/2$;

$P(\{X < 5\}) = 1/6 + 1/6 + 1/6 + 1/6$ (только четыре значения – 1, 2, 3, 4 – удовлетворяют неравенству $X < 5$), поэтому при $4 < x \le 5$ $F(x) = 4/6 = 2/3$;

$P(\{X < 6\}) = 1/6 + 1/6 + 1/6 + 1/6 + 1/6$ (только пять значений – 1, 2, 3, 4, 5 – удовлетворяют неравенству $X < 5$), поэтому при $5 < x \le 6$ $F(x) = 5/6$;

$P(\{X < t\})_{t>6} = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6$ (все шесть возможных значений – 1, 2, 3, 4, 5, 6 – удовлетворяют заданному неравенству), поэтому при $6 < x$ $F(x) = 1$.

Графическое представление вычисленной функции распределения показано на рис. 4.1. Стрелки показывают, что соответствующие значения достигаются при стремлении аргумента функции распределения справа (предел справа). ☻

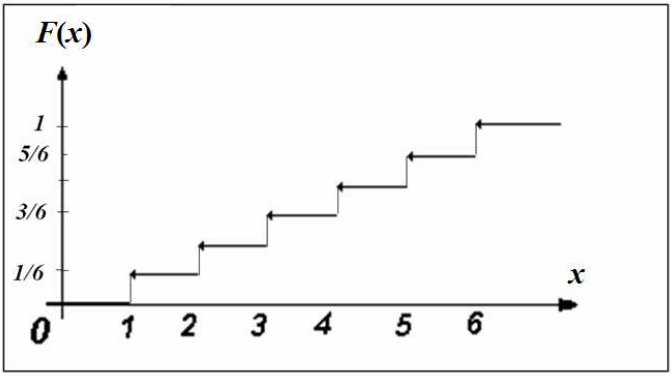


Рис. 4.1. График функции распределения (к примеру 4.5.1)

Замечание 4.5.1. Анализируя построенные аналитическое и графическое задания функции распределения, следует заметить, что:

- функция распределения является кусочно-постоянной;
- функция распределения является неубывающей (т.е. ее значения не уменьшаются при увеличении аргумента);
- точки разрыва этой функции (скачкообразного увеличения ее значений) совпадают с возможными значениями дискретной случайной величины;
- величина скачка значений функции равна вероятности появления соответствующего возможного значения случайной величины. ●

Задача 4.5.2. Студенту предложен тест из трех задач. Каждую из них он может решить с вероятностью 0.4. Тестирование заканчивается, как только будет впервые правильно решена какая-либо из задач. Определим дискретную случайную величину

X – количество попыток, которые сделал студент при решении этих задач. Требуется найти закон распределения, математическое ожидание, дисперсию этой случайной величины, построить ее функцию распределения, определить вероятность того, что число попыток будет не более двух.

Решение. Сначала построим закон распределения случайной величины X – количество попыток, которые сделал студент при решении задач. Очевидно, что значения этой случайной величины – 1, 2, 3. Определим вероятности этих значений. Значение $X = 1$ (студент решил первую задачу) может появиться с вероятностью $P = 0.4$. Значение $X = 2$ (не решил первую, решил вторую) имеет $P = 0.6 \cdot 0.4 = 0.24$ (используем теорему умножения вероятностей независимых событий). Значение $X = 3$ (не решил первую и вторую, решил третью или не решил ни одной из трех задач) имеет вероятность $P = 0.6 \cdot 0.6 \cdot 0.4 + 0.6^3 = 0.144 + 0.216 = 0.36$. Проверим условие (4.1.1): $0.4 + 0.24 + 0.36 = 1.0$. Таким образом, закон распределения имеет вид:

X	1	2	3
p	0.4	0.24	0.36

Вычислим математическое ожидание:

$$M(X) = 0.4 + 0.48 + 1.08 = 1.96.$$

Вычислим дисперсию по формуле (4.4.2):

$$D(X) = 0.4 + 0.96 + 3.24 - 3.8416 = 0.7584.$$

Построим функцию распределения, вычислив следующие вероятности:

$P(\{X < 1\}) = 0$, поэтому при $x \le 1$ $F(x) = 0$;

$P(\{X < 2\}) = 0.4$, поэтому при $1 < x \le 2$ $F(x) = 0.4$;

$P(\{X < 3\}) = 0.4 + 0.24$, поэтому при $2 < x \le 3$ $F(x) = 0.64$;

$P(\{X < t\})_{t>3} = 0.4 + 0.24 + 0.36$, поэтому при $x > 3$ $F(x) = 1$.

График этой функции приведен на рис. 4.2.

Определим вероятность того, что число попыток, которые сделал студент при решении тестовых задач, будет не более двух. Из закона распределения следует, что

$$P(\{X \le 2\}) = P(\{X = 1\}) + P(\{X = 2\}) = 0.4 + 0.24 = 0.64. \quad ☻$$

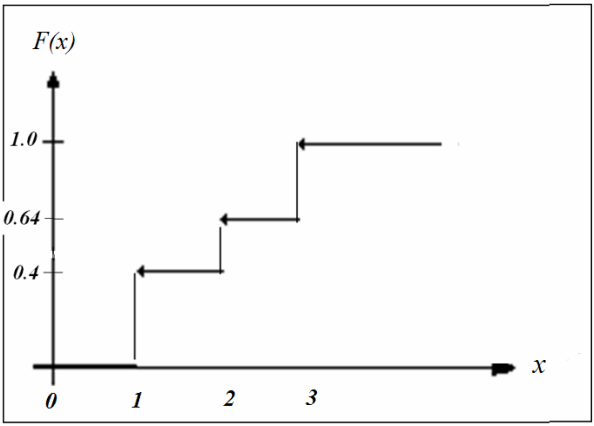


Рис. 4.2. График функции распределения (к примеру 4.5.2)

Свойства функции распределения. Приведем очевидные свойства функции распределения:

1. $0 \le F(x) \le 1$.

2. Достигаются ли эти крайние значения? В пределе – всегда (функция распределения определена на всей числовой оси): $F(-\infty) = 0$, $F(\infty) = 1$.

3. $F(x)$ – монотонно неубывающая функция (может только сохранять значения или возрастать и не может убывать). Это легко доказать по определению функции распределения. Действительно, для убывания функции распределения необходимо, чтобы вероятность какого либо значения случайной величины была меньше нуля, что противоречит определению вероятности.

4. Вероятность попадания случайной величины X на заданный интервал $[\alpha, \beta)$ определяется по следующей формуле:

$$\boxed{P\left(X \in [\alpha, \beta)\right) = P(\alpha \le X < \beta) = F(\beta) - F(\alpha)}. \qquad (4.5.2)$$

Пример 4.5.3. Пусть случайная величина Z имеет закон распределений, задаваемый следующей таблицей:

Z	1	2	3
p	$p_1 = ?$	0.20	0.30

Определить неизвестную вероятность, с которой появляется значение $Z = 1$, а также $M(Z)$, $D(Z)$.

Решение. Для вычисления неизвестной вероятности обратимся к свойству (4.1.1). Получаем: $p_1 + 0.20 + 0.30 = 1$. Откуда $p_1 = 1 - 0.20 - 0.30 = 0.50$. Таким образом, значение $Z = 1$ появляется с вероятностью 0.50.

Определим математическое ожидание и дисперсию:

$$M(Z) = 1 \cdot 0.5 + 2 \cdot 0.2 + 3 \cdot 0.3 = 1.8;$$

$$D(Z) = M(Z^2) - (M(Z))^2 = 1 \cdot 0.5 + 2^2 \cdot 0.2 + 3^2 \cdot 0.3 - (1.8)^2 =$$
$$= 4 - 3.24 = 0.76.$$

Вычислим функцию распределения, рассуждая также как в примере 4.5.2. Получаем:

$$F(x) = \begin{cases} 0, & x \le 1; \\ 0.5, & 1 < x \le 2; \\ 0.7, & 2 < x \le 3; \\ 1.0, & x > 3. \end{cases}$$

График этой функции распределения показан на рис. 4.3.

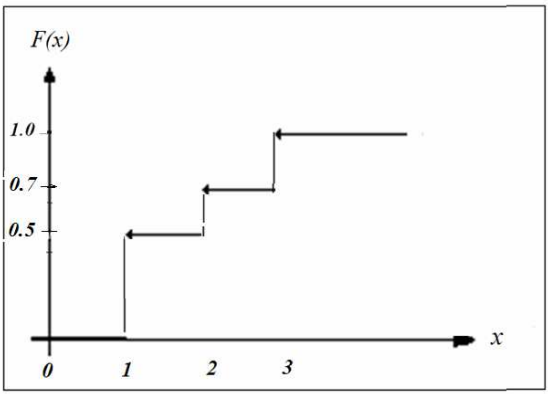


Рис. 4.3. График функции распределения (к примеру 4.5.3)

Вопросы и задачи для самопроверки

1. Чем случайная величина отличается от случайного события?
2. Дать определение закона (ряда) распределения дискретной случайной величины.
3. Кубик подбрасывается 3 раза. Составить закон распределения случайной величины X – количество выпавших единиц.
4. Какая случайная величина называется биномиально распределенной?
5. Дать определение математического ожидания случайной величины и объяснить его теоретико-вероятностный смысл.
6. Указать свойства математического ожидания.
7. Дать определение дисперсии случайной величины и объяснить теоретико-вероятностный смысл этого понятия.
8. Кубик подбрасывается 3 раза. Используя закон распределения случайной величины X – количество выпавших единиц (найденный в задаче 3), найти $M(X), D(X)$.

   Ответ: $M(X)=\frac{1}{2}$, $D(X)=\frac{5}{12}$.
9. Сформулировать свойства дисперсии.
10. Вероятность, что деталь стандартна, – 0.9. Из партии деталей с возвращением извлекается 4 детали. Случайная величина X – количество стандартных деталей среди извлеченных. Найти закон распределения случайной величины X, а также $M(X)$, $D(X)$.
11. *Рекомендация:* при вычислении закона распределения воспользуйтесь примером 4.2.1.
12. Что называется функцией распределения случайной величины?
13. Сформулировать свойства функции распределения.
14. Построить функцию распределения случайной величины X, определенной в задании 8. Проверьте выполнение свойств функции распределения.
15. Построить функцию распределения случайной величины X, определенной в задании 10. Проверьте выполнение свойств функции распределения.

Тема 5. Непрерывные случайные величины

5.1. Плотность вероятности непрерывной случайной величины

Непрерывной случайной величиной называется случайная величина, значения которой покрывают сплошным образом целый интервал на числовой оси. Например, температура, замеренная в какой-либо точке пространства. Ее значения изменяются непрерывным образом. Другой пример – вес *случайно взятого* человека.

Непрерывные случайные величины исчерпывающим образом описываются:

1) ***функцией распределения*** (имеющей тот же смысл, что и для дискретной случайной величины, но являющейся непрерывной функцией), определяемой выражением:

$$\boxed{F(x) = P(X < x)}; \qquad (5.1.1)$$

2) ***функцией плотности вероятности***:

$$\boxed{p(x) = \frac{dF(x)}{dx}}. \qquad (5.1.2)$$

Из определения (5.1.2) следует, что, зная $p(x)$, можно найти $F(x)$:

$$\boxed{F(x) = \int_{-\infty}^{x} p(z)dz}. \qquad (5.1.3)$$

Формулы (5.1.2), (5.1.3) часто используются при решении задач с непрерывными случайными величинами.

На рис. 5.1 графически показана взаимосвязь между этими функциональными характеристиками непрерывной случайной величины.

Свойства функции плотности вероятности. Приведем основные свойства функции плотности распределения:

1. Плотность распределения $p(x)$ является неотрицательной функцией:

$$\boxed{p(x) \geq 0} \qquad (5.1.4)$$

как производная от неубывающей функции.

2. Интеграл, т.е. площадь под графиком плотности вероятности, равен единице:

$$\boxed{\int_{-\infty}^{+\infty} p(x)dx = 1}, \qquad (5.1.5)$$

3. Вероятность попадания случайной величины на заданный интервал $[\alpha, \beta)$ вычисляется по формуле

$$\boxed{P(X \in [\alpha, \beta)) = P(\alpha \le x < \beta) = F(\beta) - F(\alpha)}, \qquad (5.1.6)$$

или

$$\boxed{P(X \in [\alpha, \beta)) = P(\alpha \le x < \beta) = \int_{\alpha}^{\beta} p(x)dx}. \qquad (5.1.7)$$

Интеграл в правой части формулы (5.1.7) определяет площадь криволинейной фигуры, ограниченной сверху кривой $p(x)$, слева прямой $y = \alpha$, справа – $y = \beta$, внизу – осью абсцисс.

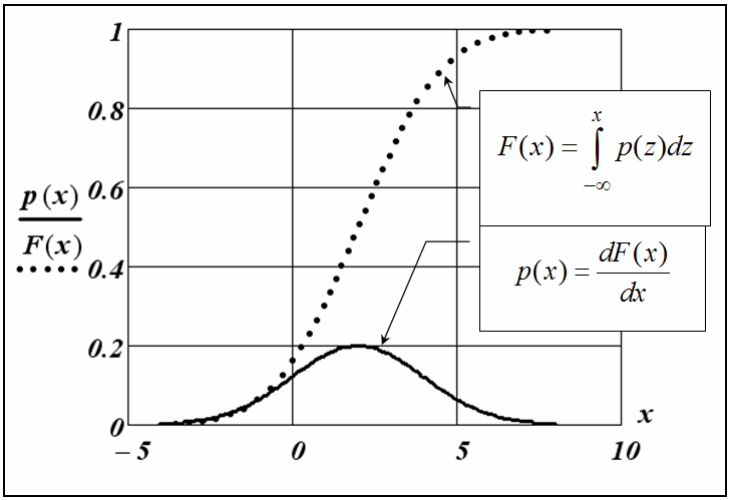


Рис. 5.1. Графики плотности распределения и функции распределения

Замечание 5.1.1. В силу непрерывности $F(x)$ вероятность того, что непрерывная случайная величина примет одно определенное значение (например, $X = x_1$), равна нулю. Действительно, из (5.1.6) следует, что вероятность

$$P(x_1 \le X < x_1 + \Delta x) = F(x_1 + \Delta x) - F(x_1).$$

Устремим $\Delta x \to 0$. Тогда в точке x_1 разность $F(x_1 + \Delta x) - F(x_1)$ также стремится к нулю; следовательно, $P(X = x_1) = 0$. Используя этот факт, легко убедиться в справедливости для непрерывной случайной величины следующих равенств:

$$\begin{aligned} &P(\alpha \le X < \beta) = P(\alpha < X < \beta) = \\ &= P(\alpha < X \le \beta) = P(\alpha \le X \le \beta). \end{aligned} \tag{5.1.8}$$

Таким образом, соотношения (5.1.6), (5.1.7) можно использовать для вычисления вероятностей, перечисленных в равенствах (5.1.8). ●

Пример 5.1.1. На рис. 5.2 сплошной кривой обозначен график плотности $p(x)$. Определим два события:

- событие A – значение случайной величины x попало в интервал $[-5, 0]$, т.е. $A = \{x \in [-5, 0]\}$;
- событие B – значение случайной величины x попало в интервал $[0, 2]$, т.е. $B = \{x \in [0, 2]\}$.

Вопрос: какое событие имеет большую вероятность?

Решение. Так как вероятность попадания случайной величины в заданный интервал равна площади соответствующей криволинейной фигуры (см. свойство (5.1.7) и соотношения (5.1.8)), то вероятность события A определяется площадью S_1 фигуры с правой «штриховкой», а вероятность события B определяется площадью S_2 фигуры с левой «штриховкой» (см. рис. 5.2). Из сравнения этих площадей видно, что $P(A) > P(B)$. ☻

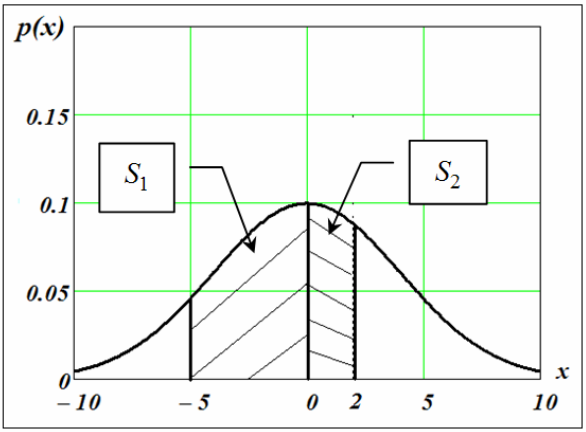


Рис. 5.2. К вычислению вероятностей событий A, B (пример 5.1.1)

Замечание 5.1.2. При решении задач с непрерывными случайными величинами будут полезны следующие формулы дифференциального и интегрального исчисления:

а) $$\boxed{\frac{d}{dx}(\text{const}) = 0}; \qquad (5.1.9)$$

б) $$\boxed{\frac{d}{dx}\left(x^n\right) = n \cdot x^{n-1}}; \qquad (5.1.10)$$

в) $$\boxed{\int_{\alpha}^{\beta} x^n dx = \frac{1}{n+1} x^{n+1}\Big|_{\alpha}^{\beta} = \frac{1}{n+1}\left(\beta^{n+1} - \alpha^{n+1}\right)}; \qquad (5.1.11)$$

г) $$\boxed{\int_{\alpha}^{\beta} (f_1(x) + f_2(x))dx = \int_{\alpha}^{\beta} f_1(x)dx + \int_{\alpha}^{\beta} f_2(x)dx}. \qquad (5.1.12)$$

5.2. Математическое ожидание и дисперсия

Математическое ожидание непрерывной случайной величины определяется по формуле

$$\boxed{M(X)=\int_{-\infty}^{+\infty} xp(x)dx}. \qquad (5.2.1)$$

Если плотность распределения $p(x)$ отличается от нуля только на конечном интервале $[a,b]$ (говорят, что функция $p(x)$ финитна на $[a,b]$), то вместо бесконечных интервалов интегрирования используют конечные:

$$\boxed{M(X)=\int_{a}^{b} xp(x)dx}. \qquad (5.2.2)$$

Математическое ожидание непрерывной случайной величины имеет те же свойства, что математическое ожидание дискретной случайной величины, а именно:

1. $M(\text{const}) = \text{const}$.
2. $M(CX) = CM(X)$.
3. $M(X + Y) = M(X) + M(Y)$, для любых двух случайных величин X и Y.
4. Пусть случайные величины X и Y *независимы*. Тогда $M(X\cdot Y) = M(X) \cdot M(Y)$.

Дисперсия непрерывной случайной величины определяется по формуле

$$\boxed{D(X)=\int_{-\infty}^{+\infty} (x-M(X))^2 p(x)dx}. \qquad (5.2.3)$$

На практике часто используют другую формулу:

$$\boxed{D(X)=M(X^2)-[M(X)]^2=\int_{-\infty}^{+\infty} x^2 p(x)dx-M^2(X)}. \qquad (5.2.4)$$

Если плотность распределения финитна на интервале $[a,b]$, то после замены пределов интегрирования имеем:

$$\boxed{D(X)=\int_a^b (x-M(X))^2 p(x)dx}; \qquad (5.2.5)$$

$$\boxed{D(X)=M(X^2)-[M(X)]^2=\int_a^b x^2 p(x)dx-M^2(X)}. \qquad (5.2.6)$$

Свойства дисперсии непрерывной случайной величины:

1. $D(X) \geq 0$.
2. $D(\text{const}) = 0$.
3. $D(CX) = C^2D(X)$.
4. Пусть случайные величины X и Y – *независимы*. Тогда $D(X + Y) = D(X) + D(Y)$; $D(X - Y) = D(X) + D(Y)$.

Пример 5.2.1. Непрерывная случайная величина X задана плотностью распределения:

$$p(x)=\begin{cases} cx, & 0 \leq x \leq 2; \\ 0, & x \notin [0,2], \end{cases}$$

т.е. плотность распределения финитна на интервале $[0, 2]$. Нужно:

- вычислить константу c;
- вычислить функцию распределения $F(x)$;
- вычислить $M(X), D(X)$;
- построить графики $p(x)$, $F(x)$.

Решение. Вычисление константы осуществляем на основе условия: $\int_{-\infty}^{+\infty} p(x)dx = 1$. Подставляя конкретное выражение, имеем уравнение:

$$\int_0^2 cxdx = c\int_0^2 xdx = 1.$$

Выполним интегрирование (см. формулу (5.1.11)):

$$c\int_0^2 xdx = c\cdot\frac{x^2}{2}\Big|_0^2 = \frac{c}{2}\cdot(2^2-0^2) = 2\cdot c = 1.$$

Тогда $c=\frac{1}{2}$, и плотность распределения имеет вид:

$$p(x)=\begin{cases}\frac{1}{2}x, & 0\le x\le 2;\\ 0, & x\notin[0,2].\end{cases}$$

Функцию распределения вычисляем по формуле (5.1.3):

$$F(x)=\int_{-\infty}^{x} p(z)dz = \int_0^x \frac{1}{2}zdz = \frac{x^2}{4} \text{ для } 0\le x\le 2 .$$

Окончательно имеем:

$$F(x)=\begin{cases}0, & x<0;\\ \frac{1}{4}x^2, & 0\le x\le 2;\\ 1, & x>2.\end{cases}$$

На рис. 5.3 сплошной кривой показан график плотности распределения, а штрихпунктирной – график функции распределения. Определим математическое ожидание:

$$M(X)=\int_0^2 x\cdot p(x)dx = \int_0^2 \frac{1}{2}x^2dx = \frac{1}{2}\cdot\left(\frac{1}{3}x^3\Big|_0^2\right) = \frac{8}{6} = 1\frac{1}{3}.$$

Вычислим дисперсию по формуле (5.2.6). Первоначально определим $\int_{-\infty}^{+\infty} x^2 p(x)dx = \frac{1}{2}\int_0^2 x^3dx = \frac{1}{2}\cdot\left(\frac{x^4}{4}\Big|_0^2\right) = \frac{16}{8} = 2.$ Затем вычисляем:

$$D(X)=\int_{-\infty}^{+\infty} x^2 p(x)dx - M^2(X) = 2-\left(\frac{4}{3}\right)^2 = \frac{18-16}{9} = \frac{2}{9}. \quad ☻$$

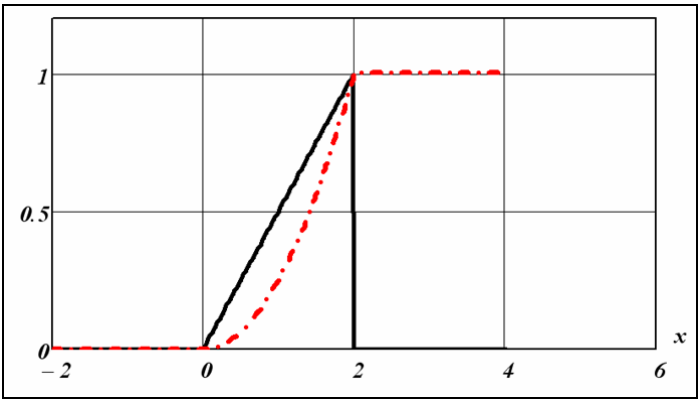


Рис. 5.3. Графики функций $p(x)$, $F(x)$ (к примеру 5.2.1)

Пример 5.2.2. Непрерывная случайная величина X имеет функцию распределения:

$$F(x)=\begin{cases}0, x\le 0;\\ kx,\ \ x\in[0,4];\\ 1,\ \ x>4.\end{cases}$$

Необходимо: вычислить неизвестную константу k, плотность распределения $p(x)$, построить графики $p(x)$, $F(x)$, найти $M(Z)$, $D(Z)$ и определить вероятность события $A=\{2\le X\le 3\}$.

Решение. Определение k будем осуществлять исходя из непрерывности $F(x)$ в точке $x=4$ и равенства $F(4)=1$, из которого получаем уравнение $k\cdot 4=1$. Следовательно, $k=0.25$, а функция распределения примет вид:

$$F(x)=\begin{cases}0, & x\le 0;\\ \frac{1}{4}x, & x\in[0,4];\\ 1, & x>4.\end{cases}$$

Для вычисления плотности распределения $p(x)$ найдем первую производную функции $F(x)$, используя для этого формулы (5.1.9), (5.1.10):

$$p(x)=\begin{cases}0, & x\le 0;\\ \frac{1}{4}, & x\in[0,4];\\ 0, & x>4.\end{cases}$$

На рис. 5.4 приведены графики функций $F(x)$ (сплошная кривая) и $p(x)$ (точечная кривая). Плотность распределения постоянна на интервале $[0,4]$, такое распределение случайной величины называют ***равномерным распределением.***

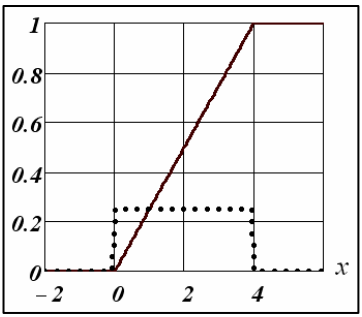


Рис. 5.4. К примеру 5.2.2

Проверим условие $\int_0^4 p(x)dx=1$. Площадь прямоугольника равна *основание* умножить на *высоту* = $4\cdot\frac{1}{4}=1$, т.е. условие выполнено. Вычислим математическое ожидание:

$$M(X)=\int_0^4\frac{1}{4}xdx=\frac{1}{4}\cdot\frac{1}{2}x^2\Big|_0^4=\frac{1}{8}(16-0)=2.$$

Дисперсию вычисляем по формуле (5.2.4). Сначала определим значение интеграла

$$\int_0^4 \frac{1}{4}x^2 dx = \frac{1}{4}\cdot\frac{1}{3}x^3\Big|_0^4 = \frac{1}{12}(64-0) = \frac{64}{12} = 5\frac{1}{3}.$$

Тогда дисперсия

$$D(X) = 5\frac{1}{3} - (2)^2 = 5 - 4 = 1\frac{1}{3} = \frac{4}{3}.$$

Вероятность события $A = \{2 \le X \le 3\}$ будем определять по формуле (5.1.7):

$$P(A) = F(3) - F(2) = \frac{1}{4}\cdot 3 - \frac{1}{4}\cdot 2 = \frac{1}{4}.$$

Заметим, что эту вероятность можно было также определить через интеграл $\int_2^3 \frac{1}{4}dx = \frac{1}{4}\cdot x\Big|_2^3 = \frac{1}{4}$, но предыдущий способ проще. ☻

5.3. Нормальное распределение случайной величины

Нормально распределенной (или нормальной) случайной величиной X (обозначаемой как $N(a, \sigma)$), называется непрерывная случайная величина, имеющая плотность вероятности следующего вида:

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}}. \qquad (5.3.1)$$

Параметры a и σ носят очень простой смысл – это математическое ожидание и среднеквадратичное отклонение (корень из дисперсии):

$$a = M(N(a,\sigma)), \quad D(N(a,\sigma)) = \sigma^2, \ \sigma = \sqrt{D(N(a,\sigma))}. \qquad (5.3.2)$$

График плотности вероятности нормально распределенной случайной величины носит название кривой Гаусса. Эта кривая

имеет вертикальную ось симметрии, проходящую через точку а. На рис. 5.4 кривой 1 показана кривая плотности распределения с математическим ожиданием $a = M(N(a,\sigma)) = 2$. Хорошо видна симметрия относительно точки x = 2. Максимальное значение плотность распределения достигает в точке x = a, величина плотности в этой точке равна $\frac{1}{\sqrt{2\pi}\cdot\sigma}$. При увеличении дисперсии максимум функции уменьшается, но увеличивается ее протяженность по оси x. Поэтому условие $\int p(x)dx = 1$ будет выполняться при любой дисперсии.

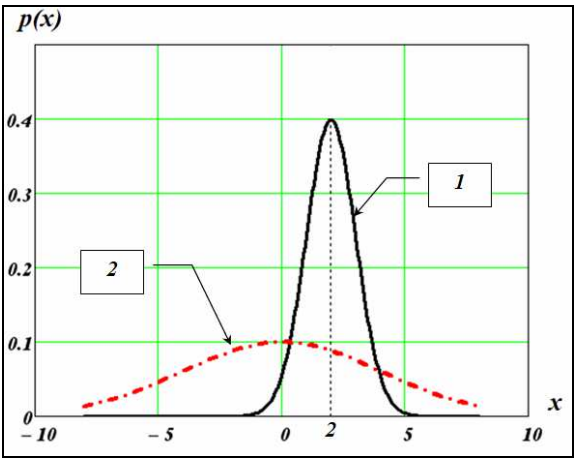


Рис. 5.4. Графики плотностей нормально распределенной случайной величины

Из определения дисперсии следует, что чем «шире» кривая плотности нормального распределения, тем больше дисперсия.

Пример 5.3.1. На рис. 5.4 представлены кривые двух нормальных распределений (номера *1* и *2*). Какое из распределений имеет:

- большее математическое ожидание;
- большую дисперсию?

Решение. Из свойства симметрии кривой распределения относительно математического ожидания следует, что распределение *1* имеет большее математическое ожидание. Кривая распределения *2* имеет большую протяженность по оси *x*, и поэтому распределение *2* имеет большую дисперсию. ☻

Вопрос: ***Почему нормально распределенные случайные величины нашли широкое применение в практике?***

Ответ:

1. Многие реально существующие в природе, технике и обществе случайные величины очень хорошо моделируются с помощью нормальных случайных величин. Это, например, ошибка результатов измерений теодолитом в геодезии, разброс скоростей и энергий молекул в газе, рост или вес случайно взятого человека.
2. Величина, которая определяется взаимодействием большого числа независимых друг от друга причин и факторов, также подчиняется нормальному распределению.

Особую роль среди нормально распределенных случайных величин играет ***нормированная случайная величина*** $N(0, 1)$ с нулевым математическим ожиданием и единичной дисперсией.

Плотность распределения величины $N(0, 1)$ имеет вид:

$$\boxed{p_N(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}}, \qquad (5.3.3)$$

а ее график – на рис. 5.5. Видно, что вне интервала $[-3, 3]$ значения функции $p_N(x)$ практически равны нулю, и появление значений случайной величины вне этого интервала является практически невозможным событием.

Вероятность события $P(0 \le N(0,1) \le x) = \Phi(x)$, где $\Phi(x)$ – интегральная функция Лапласа, определяемая выражением

$$\boxed{\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{o}^{x} e^{-\frac{x^2}{2}} dx}. \qquad (5.3.4)$$

Очевидно, что значения функции Лапласа меняется от 0 до 0.5. На рис. 5.5 заштрихованной площадью показана величина $\Phi(2) = 0.4772$. Таблица значений этой функции имеется в каждом учебнике по теории вероятностей и математической статистике, и она приведена в прил. 1. Функция $\Phi(x)$ имеет свойство:

$$\boxed{\Phi(-x) = -\Phi(x)}, \qquad (5.3.5)$$

которое часто используется в расчетах.

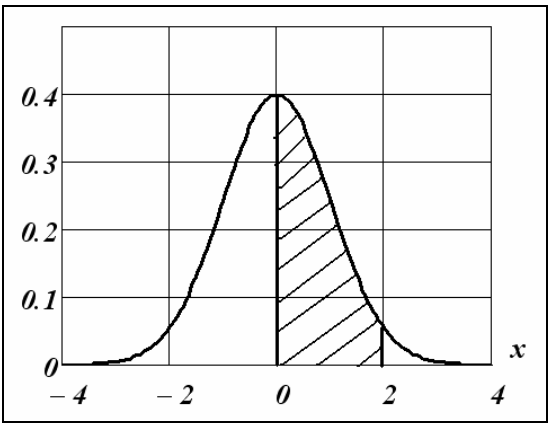


Рис. 5.5. График плотности распределения величины $N(0, 1)$

Функция Лапласа позволяет вычислить вероятность попадания случайной величины $N(a,\sigma)$ в заданный интервал $[\alpha,\beta]$, выполнив преобразование границ интервала:

$$\boxed{P(\alpha \le N(a,\sigma) \le \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)}. \qquad (5.3.6)$$

Правило трех сигм. Как показал предыдущий пример, подавляющая часть значений нормальной случайной величины достаточно компактно находятся в небольшой окрестности своего математического ожидания. Пользуясь таблицами интегральной функции Лапласа, определим радиус такой окрестности $M(N(a, \sigma))$, чтобы вероятность выхода значения случайной величины из этой окрестности была очень мала. Примем радиус 3σ и найдем вероятность события

$$A = \left(X \in \left[a - 3\sigma, a + 3\sigma\right]\right) = \left(a - 3\sigma \le X \le a + 3\sigma\right),$$

где случайная величина X имеет нормальное распределение $N(a, \sigma)$. Для вычисления этой вероятности воспользуемся выражением (5.3.6) и свойством (5.3.5):

$$P\left(a - 3\sigma \le X \le a + 3\sigma\right) = \Phi\left(\frac{a + 3\sigma - a}{\sigma}\right) - \Phi\left(\frac{a - 3\sigma - a}{\sigma}\right) =$$

$$= \Phi(3) - \Phi(-3) = \Phi(3) + \Phi(3) = 2 \cdot \Phi(3) = 2 \cdot 0.499 = 0.998.$$

Значение функции $\Phi(3) = 0.499$ можно взять из табл. П1 приложения. Словами это означает, что практически все значения (например, 997 значений из 1000) будут лежать в трехсигмовой окрестности математического ожидания, и это свойство получило название ***правило трех сигм***.

Иногда этим свойством пользуются в обратном смысле: если подавляющая часть значений исследуемой случайной величины локализуется в трехсигмовой окрестности математического ожидания, то делается вывод, что эта случайная величина имеет нормальное распределение.

Пример 5.3.2. Случайная величина X имеет нормальное распределение $N(a,\sigma)$. Определить вероятность попадания этой случайной величины в интервал $\left[a - 2\sigma, a + 2\sigma\right]$.

Решение. Для вычисления этой вероятности вновь воспользуемся выражением (5.3.6) и свойством (5.3.5):

$$P\left(a-2\sigma \le X \le a+2\sigma\right)=\Phi\left(\frac{a+2\sigma-a}{\sigma}\right)-\Phi\left(\frac{a-2\sigma-a}{\sigma}\right)=$$

$$=\Phi\left(2\right)-\Phi\left(-2\right)=\Phi\left(2\right)+\Phi\left(2\right)=2\cdot\Phi\left(2\right)=2\cdot 0.4772=0.9544.$$

Значение функции $\Phi(2)=0.4772$ возьмем из табл. П1. Можно сказать, что с вероятностью 0.95 значения случайной величины $N(a,\sigma)$ находятся в интервале $\left[a-2\sigma, a+2\sigma\right]$. Другими словами, из 1000 значений случайной величины примерно 954 будет находиться в указанном интервале. Это свойство можно назвать ***правилом двух сигм***. ☻

В качестве графической интерпретации правил двух и трех сигм на рис. 5.6 приведены 200 значений нормально распределенной величины X = N(0, 1). Видно, что только 2 значения из 200 находятся вне интервала трех сигм $\left[-3, 3\right]$, а большинство значений находятся внутри двух сигм $\left[-2, 2\right]$.

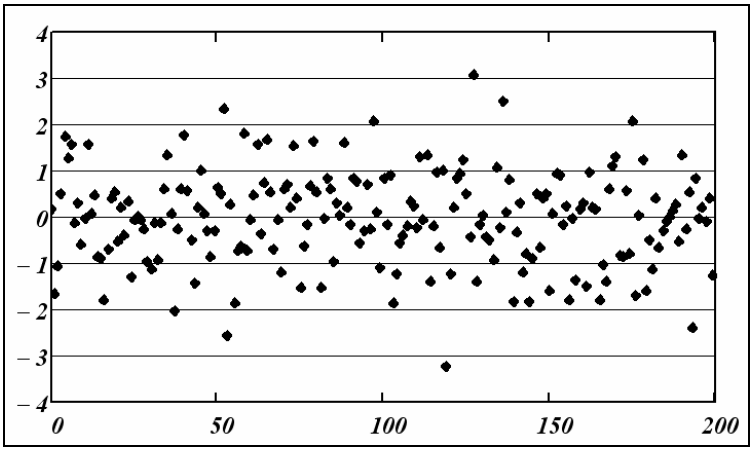


Рис. 5.6. Значения случайной величины N(0, 1) (пример 5.3.2)

Для вычисления вероятности $P(\alpha \le N(a,\sigma) \le \beta)$ можно использовать функцию Excel НОРМРАСП, обращение к которой имеет вид:

$$\text{НОРМРАСП}(x;a;\sigma;ind). \qquad (5.3.7)$$

Если параметр $ind = 0$, то результатом работы функции является значение плотности нормального распределения в заданной точке x:

$$p(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Если параметр $ind = 1$, то результатом работы функции является значение функции распределения:

$$F(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\int_{-\infty}^{x} e^{-\frac{(z-a)^2}{2\sigma^2}}dz. \qquad (5.3.8)$$

Из выражения (5.3.8) следует, что вероятность $P(\alpha \le N(a,\sigma) \le \beta)$ можно вычислить следующим фрагментом Excel:

$$P(\alpha \le N(a,\sigma) \le \beta) =$$

$$=\text{НОРМРАСП}(\beta;a;\sigma;1)\text{–НОРМРАСП}(\alpha;a;\sigma;1). \qquad (5.3.9)$$

Пример 5.3.3. Для закупки и последующей продажи мужских зимних курток было проведено выборочное обследование мужского населения города N в возрасте от 18 до 65 лет в целях определения его среднего роста. В результате было установлено, что средний рост (математическое ожидание) равен 176 см, среднеквадратическое отклонение $\sigma = 6$. Необходимо определить, какой процент общего числа закупаемых курток должны составлять куртки 5-го роста (182–186 см). Предполагается, что рост мужского населения города N распределен по нормальному закону.

Решение. Определить процент в этой задаче – значит найти вероятность случайного события $A = \left(182 \le N(176, 6) \le 186\right)$.

Для вычислений этой вероятности в ячейке Excel запрограммируем следующее выражение:

=НОРМРАСП(186;176;6;1)–НОРМРАСП(182;176;6;1).

Получаем вероятность P = 0.111.

Ответ: куртки 5-го роста должны составлять приблизительно 11 % общего числа закупаемых курток.

5.4. Предельные теоремы теории вероятностей

Сходимость по вероятности. Пусть имеется последовательность случайных величин X_1, X_2, …, X_n, … Говорят, что данная последовательность сходится по вероятности к случайной величине Y (обозначается $X_n \underset{n\to\infty}{\overset{P}{\to}} Y$), если

$$\boxed{\lim_{n\to\infty} P(\{|X_n - Y| < \varepsilon\}) = 1} \qquad (5.4.1)$$

при любом сколь угодно малом $\varepsilon > 0$.

Лемма Чебышева. Для любой случайной величины X справедливо неравенство (называемое неравенством Чебышева)

$$\boxed{P(|X - M(X)| > \varepsilon) \le \frac{D(X)}{\varepsilon^2}}, \qquad (5.4.2)$$

дающее *количественную* оценку вероятности отклонения случайной величины от своего математического ожидания при любом распределении случайной величины.

Пример 5.4.1. Математическое ожидание случайной величины равно 5, дисперсия – 0.01. Оценить вероятность того, что случайная величина будет отличаться от математического ожидания не более чем на 0.5.

Решение. Введем случайное событие $A = (|X - 5| \le 0.5)$ и противоположное событие $\overline{A} = (|X - 5| > 0.5)$. Для вычисления вероятности события $\overline{A}$ обратимся к неравенству (5.4.2), под-

ставив в него конкретные числовые характеристики случайной величины:

$$P(\bar{A}) = P(|X-5| > 0.5) \le \frac{0.01}{0.25} = 0.04.$$

Тогда искомая вероятность события A определяется как

$$P(A) \ge 1 - P(\bar{A}) = 1 - 0.04 = 0.96,$$

т.е. $P\left(|X-5| \le 0.5\right) \ge 0.96$. ☻

Закон больших чисел в форме Чебышева. Пусть имеется последовательность попарно независимых случайных величин $X_1, X_2, \ldots, X_n, \ldots$ с ограниченными дисперсиями ($D(X_i) \le$ C для всех i). Тогда среднее арифметическое этих случайных величин сходится по вероятности к среднему арифметическому их математических ожиданий:

$$\boxed{(X_1 + X_2 + \ldots + X_n)/n \underset{n\to\infty}{\overset{P}{\to}} (\sum_{i=1}^{n} M(X_i))/n} \qquad (5.4.3)$$

и, следовательно, асимптотически уже не является случайной величиной.

Следствия из закона больших чисел

Следствие 1. Пусть $X_1, X_2, \ldots, X_n \ldots$ – попарно независимые одинаково распределенные случайные величины с одинаковыми числовыми характеристиками: $M(X_i) = M(X)$, $D(X_i) = \sigma^2$. Из закона больших чисел следует, что среднее арифметическое $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ большого количества значений случайной величины (называемое выборочным средним) сходится по вероятности к математическому ожиданию:

$$\boxed{\bar{X} \underset{n\to\infty}{\overset{P}{\to}} M(X)}. \qquad (5.4.4)$$

Пример 5.4.2. Определить, сколько нужно сделать измерений диаметра деревьев, чтобы средний размер отличался от истинного не более чем на 2 см с вероятностью не менее 0.95.

Решение. Для применения неравенства Чебышева (5.4.2) необходимо знать математическое ожидание среднего размера диаметра деревьев $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Из предела (5.4.4) следует, что математическим ожиданием выборочного среднего является математическое ожидание случайной величины X, т.е. $M(X)$. Поэтому обратимся к неравенству (5.4.2) и перепишем его для противоположного события:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - M(X)\right| \le \varepsilon\right) > 1 - \frac{D(X)}{n\varepsilon^2}. \qquad (5.4.5)$$

Подставим в это выражение $\varepsilon = 2, D(X) = 10, P \ge 0.95$. Имеем $1 - \frac{100}{4n} \ge 0.95$ или $0.05 \ge \frac{100}{4n}$. Из последнего неравенства следует ответ $n \ge 500$. ☻

Заметим, что выражение (5.4.5) есть упрощенная формулировка закона больших чисел, пригодная для решения практических задач.

Следствие 2. Теорема Бернулли. Данная теорема позволяет на основе наблюдений над случайной величиной приближенно определить вероятность случайного события *A*.

Рассмотрим цепочку *n* одинаковых опытов, в каждом из которых событие *A* может произойти с вероятностью *p*. Введем биномиальную случайную величину S_n как количество появлений события *A* в этой цепочке испытаний (или как число опытов, в которых событие *A* произошло). Тогда относительная частота $\frac{S_n}{n}$ появления события *A* в *n* независимых опытах сходится по вероятности к константе *p*, т.е.

$$\boxed{\frac{S_n}{n} \underset{n\to\infty}{\overset{P}{\longrightarrow}} p}. \qquad (5.4.6)$$

Значит, приближенное значение вероятности случайного события A можно найти, если путем наблюдений определить относительную частоту появления этого события в достаточно большом числе испытаний.

Пример 5.4.3. Сколько следует провести независимых испытаний, чтобы выполнение неравенства $\left|\frac{S_n}{n} - p\right| \le 0.08$ превысило 0.75, если вероятность появления случайного события в испытании равна $p = 0.8$?

Решение. Для применения неравенства Чебышева (5.4.2) необходимо знать математическое ожидание величины $\frac{S_n}{n}$. Из (5.4.6) следует, что им является вероятность события p. Тогда перепишем неравенство Чебышева для противоположного события:

$$P\left(\left|\frac{S_n}{n} - p\right| \le \varepsilon\right) > 1 - \frac{p(1-p)}{n\varepsilon^2}. \qquad (5.4.7)$$

Подставляя $\varepsilon = 0.08,\ p = 0.8$, получаем

$$P\left(\left|\frac{S_n}{n} - p\right| \le 0.08\right) > 1 - \frac{0.8 \cdot 0.2}{n \cdot 0.0064} > 0.75.$$

Из этого неравенства следует:

$$0.25 > \frac{0.16}{0.0064n}, \text{ или } \ n > 100.$$

Ответ: $n > 100$. ☻

Заметим, что выражение (5.4.7) есть упрощенная формулировка закона больших чисел для частного случая – испытания по схеме Бернулли.

Центральная предельная теорема (теорема Ляпунова). Пусть X_1, X_2, …, X_n, … – последовательность непрерывных одинаково распределенных случайных величин с одинаковыми числовыми характеристиками $M(X) = a$, $D(X) = \sigma^2$ (А.М. Ляпунов дал некоторые довольно общие условия, при которых X_i могут быть и не одинаковы). Тогда для суммы этих случайных величин $Y_n = X_1 + X_2 + ... + X_n$ асимптотически (т.е. при $n \to \infty$) имеет место

$$\boxed{\frac{Y_n - an}{\sqrt{n\sigma^2}} \underset{n\to\infty}{\overset{P}{\to}} N(0,1)}, \qquad (5.4.8)$$

т.е. величина $\frac{Y_n - an}{\sqrt{n\sigma^2}}$ при $n \to \infty$ подчиняется нормальному распределению с нулевым средним и единичной дисперсией.

Эта теорема еще раз показывает важность для практических приложений изучения нормально распределенных случайных величин.

Вопросы и задачи для самопроверки

1. Какая случайная величина называется непрерывной случайной величиной?
2. Как связаны между собой функция распределения и функция плотности вероятности?
3. Сформулировать свойства плотности вероятности.
4. Как вычисляется математическое ожидание?
5. Свойства математического ожидания.
6. Как вычисляется дисперсия непрерывной случайной величины?
7. Какими свойствами обладает дисперсия непрерывной случайной величины?
8. Пусть $F(x)$ – функция распределения случайной величины X:

$$F(x)=\begin{cases}0, & x\le 0;\\ \dfrac{x^2}{4}, & 0<x\le 2;\\ 1, & x>2.\end{cases}$$

Вычислить плотность распределения $p(x)$; построить графики функций $F(x)$, $p(x)$; вычислить числовые характеристики $M(X), D(X)$ и вероятность $P(X\ge 1)$.

Ответ: $M(X)=1\frac{1}{3}$; $D(X)=\frac{2}{9}$; $P=\frac{3}{4}$.

9. Пусть случайная величина имеет плотность распределения

$$p(x)=\begin{cases}Cx, & x\in[0,1];\\ 0, & x\notin[0,1].\end{cases}$$

Найти постоянную C, функцию распределения $F(x)$; построить графики функций $F(x)$, $p(x)$; найти вероятность $P(X>0.5)$.

Ответ: $C=2$; $F(x)=\begin{cases}x, & x<0;\\ x^2, & 0\le x\le 1;\\ 1, & x>1;\end{cases}$ $P=0.75$.

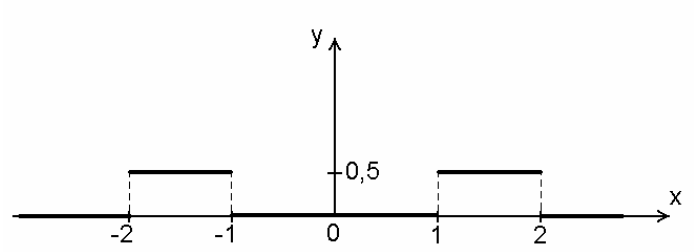
10. Функция плотности случайной величины X имеет вид:

$$p(x)=\begin{cases}x, & x\in[0,1];\\ 2-x, & x\in[1,2];\\ 0, & x\notin[0,2].\end{cases}$$

Необходимо найти $F(x)$, $M(X^3)$, $P(0.5\le X\le 1.5)$ и построить графики функций $F(x)$, $p(x)$.

Ответ: $F(x)=\begin{cases} 0, & x<0; \\ \dfrac{x^2}{2}, & 0\le x\le 1; \\ \dfrac{(3-x)(x-1)+1}{2}, & 1\le x\le 2; \\ 1, & x>2; \end{cases}$ $P=0.75.$

11. На рисунке изображен график функции плотности случайной величины X.



Найти функцию распределения случайной величины X. Построить ее график.

Ответ: $F(x)=\begin{cases} 0, & x<-2; \\ (x+2)/2, & -2\le x\le -1; \\ 0.5, & -1\le x\le 1; \\ 0.5+(x-1)/2, & 1\le x\le 2; \\ 1, & x>2. \end{cases}$

12. Как записывается плотность случайной величины, распределенной по нормальному закону?
13. Случайные величины N_1, N_2, …, N_{50} распределены нормально с параметрами a = 2, σ = 1. Найти математическое ожидание и дисперсию случайной величины $X = \dfrac{N_1+N_2+...+N_{50}}{50}$,

если $N_1, N_2, \ldots, N_{50}$ – независимые случайные величины.
Ответ: $M(X) = 2, \quad D(X) = 0.02.$

14. Дать определение интегральной функции Лапласа. Указать ее свойства и способ практического применения.
15. Как вычисляется вероятность попадания нормально распределенной величины $N(2, 4)$ в интервал $[1, 5]$?
16. Что такое правило трех сигм?
17. Какова вероятность попадания случайной величины $N(a, \sigma)$ в интервал $[a - 2\sigma, a + 2\sigma]$?

Тема 6. Двумерные случайные величины

Пусть X и Y – случайные величины, определенные на одном и том же пространстве элементарных исходов Ω.

Под ***двумерной случайной величиной*** будем понимать упорядоченную пару случайных величин (X, Y). Случайные величины X и Y называются ***составляющими двумерной случайной величины***. Значениями двумерной случайной величины являются точки (x, y) на плоскости.

Рассмотрим отдельно два случая, когда обе составляющие одновременно дискретные или непрерывные случайные величины. В первом случае двумерную случайную величину называют дискретной, во втором – непрерывной.

6.1. Дискретные двумерные случайные величины

Законом распределения дискретной двумерной случайной величины (X, Y) называют перечень возможных значений этой величины, т.е. пар чисел (x_i, y_j) и вероятностей $p_{i,j} = P(X = x_i, Y = y_j)$. Если количество значений, составляющих X и Y, конечно, то распределение задают в виде табл. 6.1.

Таблица 6.1

$Y \backslash X$	x_1	x_2	...	x_n
y_1	$p_{1,1}$	$p_{2,1}$	...	$p_{n,1}$
y_2	$p_{1,2}$	$p_{2,2}$	...	$p_{n,1}$
...	...	...	...	...
y_m	$p_{1,m}$	$p_{2,m}$	...	$p_{n,m}$

В первой строке помещены все возможные значения случайной величины X, в первом столбце – все возможные значения случайной величины Y, в клетках – вероятности $p_{i,j}$.

Зная закон распределения двумерной дискретной случайной величины, можно найти закон распределения составляющих X и Y. Можно показать, что

$$\boxed{P(X = x_i) = \sum_{j=1}^{m} p_{i,j} = p_{i,\bullet}}, \qquad (6.1.1)$$

где точка в записи $p_{i,\bullet}$ означает индекс, по которому было выполнено суммирование. Аналогично,

$$\boxed{P(Y = y_j) = \sum_{i=1}^{n} p_{i,j} = p_{\bullet,j}}. \qquad (6.1.2)$$

Из последнего выражения следует, что составляющая Y имеет следующий закон распределения:

Y	y_1	y_2	...	y_m
P	$p_{\bullet,1}$	$p_{\bullet,2}$	...	$p_{\bullet,m}$

Отметим, что сумма всех вероятностей $p_{i,j}$ равна 1 и поэтому

$$\boxed{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = \sum_{i=1}^{n} p_{i,\bullet} = \sum_{j=1}^{m} p_{\bullet,j} = 1}. \qquad (6.1.3)$$

Составляющие X и Y двумерной дискретной случайной величины будут независимыми тогда и только тогда, когда для любых $i = \overline{1,\ n}$, $j = \overline{1,\ m}$ выполняется равенство

$$\boxed{p_{i,j} = p_{i,\bullet} \cdot p_{\bullet,j}}. \qquad (6.1.4)$$

Задание. Сравните это равенство с определением независимости случайных событий (2.3.3).

Рассмотрим условные распределения составляющих X и Y. Предположим, что случайная величина Y приняла значение y_j. При этом составляющая X может принять одно из значений x_1, x_2, ..., x_n. Обозначим условную вероятность $P(X = x_i\ /\ Y = y_j)$ через $p(x_i/y_j)$. ***Условным распределением*** составляющей X при $Y = y_j$ называется совокупность условных вероятностей $p(x_1\ /y_j)$, ..., $p(x_n\ /\ y_j)$. Аналогично определяется условное распределение составляющей Y.

По определению условной вероятности имеем:

$$\boxed{p(x_i\ /\ y_j) = \frac{P(X = x_i,\ Y = y_j)}{P(Y = y_j)} = \frac{p_{i,j}}{p_{\bullet,j}}}; \qquad (6.1.5)$$

$$\boxed{p(y_j\ /\ x_i) = \frac{P(X = x_i,\ Y = y_j)}{P(X = x_i)} = \frac{p_{i,j}}{p_{i,\bullet}}}. \qquad (6.1.6)$$

для $i = \overline{1,\ n}$, $j = \overline{1,\ m}$.

Пример 6.1.1. Закон распределения двумерной дискретной случайной величины задан табл. 6.2. Найти условный закон распределения составляющей X при условии, что составляющая Y приняла значение 0.

Решение. Из табл. 6.2 (используя формулу (6.1.5)) находим вероятности:

$$p_{\bullet,1} = 0.1 + 0.2 + 0.3 = 0.6; \quad p(3/0) = \frac{0.1}{0.6}; \quad p(5/0) = \frac{0.2}{0.6} = \frac{1}{3};$$

$$p(7/0) = \frac{0.3}{0.6} = \frac{1}{2}. \quad ☻$$

Таблица 6.2

Y \ X	3	5	7
0	0.10	0.20	0.30
1	0.05	0.15	0.20

Определим среднее значение составляющей X при условии, что составляющая Y приняла значение y_j. Оно называется ***условным математическим ожиданием*** и вычисляется по формуле

$$\boxed{M(X/Y = y_j) = \sum_{i=1}^{n} x_i\, p(x_i / y_j)}. \qquad (6.1.7)$$

По аналогии условное математическое ожидание при $X = x_i$ определяется выражением

$$\boxed{M(Y/X = x_i) = \sum_{j=1}^{m} y_j\, p(y_j / x_i)}. \qquad (6.1.8)$$

Пример 6.1.2. Вычислить условное математическое ожидание составляющей X при условии, что $Y = 0$ (распределение двумерной случайной величины задано табл. 6.2).

Решение. Для этого будем использовать формулу (6.1.6) и условные вероятности, вычисленные в примере 6.1.1:

$$M(X/Y=0)=3\cdot P(3/0)+5\cdot P(5/0)+7\cdot P(7/0)=$$
$$=3\cdot 1/6+5\cdot 1/3+7\cdot 1/2=34/6=17/3.$$ ☻

6.2. Двумерные непрерывные случайные величины

Рассмотрим двумерную случайную величину (X, Y). Составляющие X и Y являются непрерывными случайными величинами. Введем статистические характеристики двумерной случайной величины, которые аналогичны характеристикам одномерной (см. тему 5).

Пусть x, y – произвольные действительные числа. Функция распределения $F(x, y)$ пары случайных величин (X, Y) определяется как вероятность:

$$\boxed{F(x, y) = P(X < x, Y < y)}. \qquad (6.2.1)$$

Множество точек на плоскости, координаты которых удовлетворяют неравенствам $X < x$ и $Y < y$, образуют угол, заштрихованный на рис. 6.1. Поэтому значение функции распределения $F(x, y)$ представляет вероятность попадания случайной точки (X, Y) в заштрихованную область.

Приведем следующие свойства функции распределения пары любых случайных величин (X, Y):

1. $0 \le F(x, y) \le 1.$
2. $\lim\limits_{x\to-\infty} F(x, y) = \lim\limits_{y\to-\infty} F(x, y) = \lim\limits_{\substack{x\to-\infty\\ y\to-\infty}} F(x, y) = 0.$
3. $\lim\limits_{\substack{x\to\infty\\ y\to\infty}} F(x,y) = 1.$
4. $F(x, y)$ – неубывающая функция по каждому аргументу, т.е. $x_2 > x_1$ влечет $F(x_2,y) \ge F(x_1,y)$; $y_2 > y_1$ влечет $F(x, y_2) \ge F(x, y_1)$.

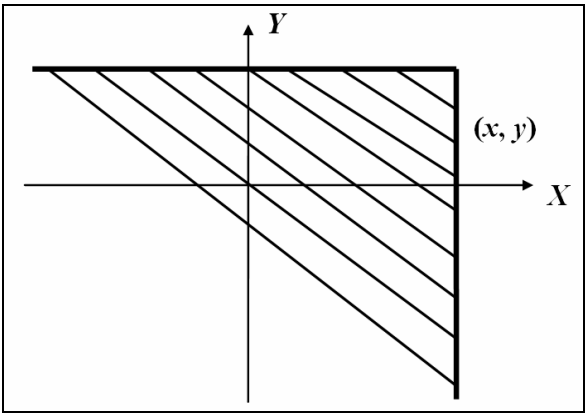


Рис. 6.1. К определению функции распределения $F(x, y)$

Плотность распределения $p(x, y)$ двумерной случайной величины (X, Y) и функция распределения связаны между собой соотношениями:

$$\boxed{F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} p(u,v)du\,dv}; \qquad (6.2.2)$$

$$\boxed{p(x,y)=\frac{\partial^2 F(x,y)}{\partial x \partial y}}. \qquad (6.2.3)$$

Очевидно, что плотность распределения должна удовлетворять условию:

$$\boxed{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)dx\,dy=1}. \qquad (6.2.4)$$

Вероятность попадания случайной величины в некоторую область D равна двойному интегралу по области D от функции плотности распределения $p(x, y)$:

$$\boxed{P((X,Y)\in D)=\iint_{D} p(x,y)dx\,dy}. \qquad (6.2.5)$$

Дадим геометрическую трактовку этому соотношению. Рассмотрим объемную фигуру, основанием которой является область D, а высота ограничена функцией $p(x, y)$. Тогда вероятность попадания случайной величины (X, Y) равна объему этой фигуры, и этот объем может принимать любое значение из интервала $[0, 1]$.

Двумерная случайная величина (X, Y) распределена ***равномерно в области D*** на плоскости, если плотность распределения вероятности имеет вид:

$$P(x, y) = \begin{cases} C, & (x, y) \in D; \\ 0, & (x, y) \notin D, \end{cases}$$

где константа находится из условия (6.2.4): $C = 1/S$, S – площадь области D.

Пример 6.2.1. Двумерная случайная величина (X, Y) распределена равномерно в круге радиуса R с центром в начале координат. Найти плотность распределения вероятности $p(x, y)$, а также вероятность попадания случайной точки (X, Y) в квадрат, вписанный в круг.

Решение. Так как площадь круга $S = \pi R^2$, то двумерная плотность вероятности определяется формулой

$$p(x, y) = \begin{cases} 1/\pi R^2, & x^2 + y^2 \le R^2 \\ 0, & x^2 + y^2 > R^2. \end{cases}$$

Вероятность попадания случайной точки в квадрат можно вычислить по формуле (6.2.5):

$$P = \iint\limits_{квадрат} 1/\pi R^2 dx\, dy = 1/\pi R^2 \iint\limits_{квадрат} dx\, dy = S_{квадрат} / \pi R^2.$$

Так как площадь вписанного в круг квадрата $S_{квадрат}$ равна $2R^2$, то искомая вероятность $P = 2/\pi$. ☻

Зная плотность распределения двумерной дискретной случайной величины, можно найти плотности распределения

составляющих X и Y. Можно показать, что плотность распределения составляющей X определяется как

$$\boxed{p_1(x) = \int_{-\infty}^{+\infty} p(x,v)dv},$$

а составляющей Y:

$$\boxed{p_2(y) = \int_{-\infty}^{+\infty} p(u,y)du}.$$

Критерий независимости непрерывных случайных величин. Для того чтобы непрерывные случайные величины X и Y были независимы, необходимо и достаточно, чтобы в точках непрерывности функций $p(x, y)$, $p_1(x)$, $p_2(y)$ выполнялось равенство

$$\boxed{p(x, y) = p_1(x) \cdot p_2(y)}. \qquad (6.2.6)$$

6.3. Коэффициент корреляции

При изучении пары случайных величин, естественно, возникает вопрос о связи между составляющими. Напомним, что независимые случайные величины X и Y не могут влиять друг на друга, взаимовлияние отсутствует. В свою очередь, между зависимыми случайными величинами обязательно существует статистическая взаимосвязь, и говорят, что они в этом случае влияют друг на друга. Пусть, например, X – число курильщиков по данным из ряда регионов, а Y – число зарегистрированных в этих регионах больных раком легких и туберкулезом. Очевидно, эти случайные величины будут связаны между собой (положительное взаимовлияние: если X растет, то и Y растет). Бывает и отрицательная взаимосвязь: например, при повышении X – цены, устанавливаемой на товар, – как правило, уменьшается величина Y – эффективный спрос на него.

Для характеристики взаимосвязи случайных величин служат корреляционный момент и коэффициент корреляции.

Корреляционным моментом $\mu_{X,Y}$ случайных величин X и Y называется математическое ожидание случайной величины $(X - M(X)) \cdot (Y - M(Y))$, т.е.

$$\boxed{\mu_{X,Y} = M\left[(X - M(X)) \cdot (Y - M(Y))\right]}. \qquad (6.3.1)$$

Из определения математического ожидания для дискретных случайных величин получаем

$$\boxed{\mu_{X,Y} = \sum_i \sum_j (x_i - m_x)(y_j - m_y)\, p_{i,j}}, \qquad (6.3.2)$$

где $m_x = M(X)$; $m_y = M(Y)$; $p_{i,j} = P(X = x_i,\ Y = y_j)$.

Для непрерывных случайных величин получаем:

$$\boxed{\mu_{X,Y} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - m_x)(y - m_y)\, p(x, y)\, dx\, dy}. \qquad (6.3.3)$$

Пример 6.3.1. Распределение дискретной случайной величины (X, Y) задано табл. 6.3. Найти корреляционный момент $\mu_{X,Y}$.

Таблица 6.3

Y \ X	–1	0	1
0	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{8}$
1	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$

Решение. Определим математическое ожидание $M(X)$, $M(Y)$. Для этого вычислим следующие вероятности составляющих (см. формулы (6.1.1) и (6.1.2)).

Составляющая X: $p_{1,\bullet} = \frac{1}{8} + \frac{1}{6} = \frac{7}{24}$; $p_{2,\bullet} = \frac{5}{12}$; $p_{3,\bullet} = \frac{7}{24}$.

Составляющая Y: $p_{\bullet,1} = \frac{1}{3}$, $p_{\bullet,2} = \frac{2}{3}$.

Тогда получаем:

$$M(X) = -1 \cdot \frac{7}{24} + 0 \cdot \frac{5}{12} + 1 \cdot \frac{7}{24} = 0; \quad M(Y) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

Используя формулу (6.3.2), найдем $\mu_{X,Y}$:

$$\mu_{X,Y} = \sum_{i=1}^{3} \sum_{j=1}^{2} (x_i - 0)\left(y_i - \frac{2}{3}\right) p_{i,j} = (-1) \cdot \left(\frac{2}{3}\right) \cdot \frac{1}{8} +$$

$$+ (-1) \cdot \left(1 - \frac{2}{3}\right) \cdot \frac{1}{6} + 1 \cdot \left(-\frac{2}{3}\right) \cdot \frac{1}{8} + 1 \cdot \left(1 - \frac{2}{3}\right) \cdot \frac{1}{6} = 0.$$

Итак, корреляционный момент $\mu_{X,Y} = 0$. ☻

Свойства корреляционного момента

1. Для любых случайных величин X и Y имеет место равенство

$$\boxed{\mu_{X,Y} = M(XY) - M(X)M(Y)}. \qquad (6.3.4)$$

2. Корреляционный момент двух независимых случайных величин равен нулю.

Случайные величины, корреляционный момент которых равен нулю, называются ***некоррелированными***.

Заметим, что обращение свойства 2, вообще говоря, неверно, т.е. из равенства нулю коэффициента корреляции не следует в общем случае независимость случайных величин. Так, случайные величины, распределение которых задано табл. 6.2, зависимы. Чтобы проверить это, достаточно вычислить $P(X = -1) = 7/24$ и $P(Y = 0) = 1/3$, а затем убедиться, что $P(X = -1,\ Y = 0) \neq$ $\neq P(X = -1)P(Y = 0)$, или $\frac{7}{72} \neq \frac{9}{72}$. Тем не менее, корреляционный момент $\mu_{X,Y} = 0$.

Другой важной характеристикой связи случайных величин является ***коэффициент корреляции***. В отличие от корреляционного момента, он является безразмерной величиной.

Коэффициентом корреляции ρ_{XY} случайных величин X и Y называется отношение корреляционного момента $\mu_{X,Y}$ к произведению их среднеквадратических отклонений:

$$\boxed{\rho_{X,Y} = \frac{\mu_{X,Y}}{\sigma_X \sigma_Y}}, \qquad (6.3.5)$$

или

$$\boxed{\rho_{X,Y} = \frac{M\left[(X - M(X))(Y - M(Y))\right]}{\sigma_X \sigma_Y}}. \qquad (6.3.6)$$

Свойства коэффициента корреляции. Приведем некоторые свойства коэффициента корреляции.

1. $-1 \le \rho_{XY} \le 1$, ($|\rho_{XY}| \le 1$).

2. Если случайные величины X и Y являются независимыми (не влияют друг на друга), то $\rho_{XY} = 0$.

3. Если $\rho_{XY} = +1$ (положительная корреляция), то между X и Y существует функциональная линейная зависимость $Y = aX + b$, где $a > 0$. Если $\rho_{XY} = -1$ (отрицательная корреляция), то между X и Y существует функциональная обратная зависимость $Y = aX + b$, где $a < 0$. На рис. 6.2 показаны графики функциональных зависимостей при разных знаках коэффициента a.

4. Из некоррелированности составляющих двумерной нормально распределенной случайной величины следует ***их независимость***. Важность этого объясняется тем, что наблюдаемые во многих случаях практики двумерные случайные величины, по крайней мере с некоторым приближением, можно считать нормальными. Это тот случай, когда ***понятия независимости и некоррелированности совпадают***.

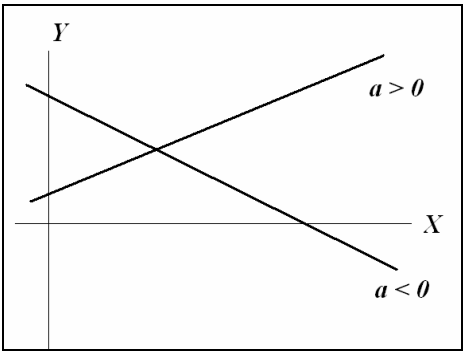


Рис. 6.2. Функциональные зависимости при разных знаках коэффициента a

Для иллюстрации статистического смысла коэффициента корреляции приведем диаграммы рассеяния (геометрические места точек (x_i, y_j) на плоскости) для разных коэффициентов корреляции. На рис 6.3*а* диаграмма рассеяния соответствует коэффициенту корреляции, равному 0, на рис. 6.3*б* коэффициент корреляции равен 0.94. Видно, что при увеличении модуля $|\rho_{XY}|$ расположение точек на плоскости стремится к прямой линии (см. рис. 6.2).

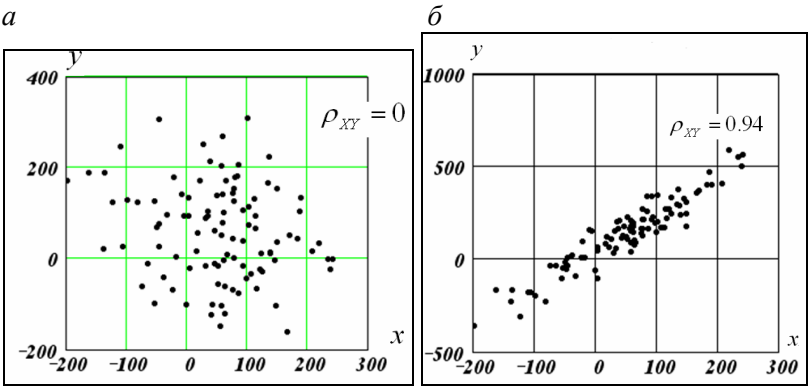


Рис. 6.3. Диаграммы рассеяния при разных ρ_{XY}

Пример 6.3.2. Двумерная случайная величина (X, Y) задана своим законом распределения (в табл. 6.4 под наклонной чертой показана вероятность пары значений двумерной дискретной величины, указанной над наклонной чертой). Требуется найти коэффициент корреляции данной системы случайных величин.

Таблица 6.4

Y \ X	0	–1
3	(3, 0) / 1/2	(3, –1) / 1/12
2	(2, 0) / 1/12	(2, –1) / 1/3

Таблица 6.5

X	3	2
p	7/12	5/12

Таблица 6.6

Y	0	–1
p	7/12	5/12

Решение. Составим законы распределения составляющих X и Y, пользуясь формулами (6.1.1) и (6.1.2). Вычисленные законы распределения приведены в табл. 6.5 и 6.6. Пользуясь этими вероятностями, вычислим числовые характеристики:

$M(X) = 3\cdot 7/12 + 2\cdot 5/12 = 31/12$;

$M(Y) = 0\cdot 7/12 + (-1)\cdot 5/12 = -5/12$;

$D(X) = 9\cdot 7/12 + 4\cdot 5/12 - 961/144 = (996 - 961)/144 = 35/144$;

$D(Y) = 0\cdot 7/12 + 1\cdot 5/12 - 25/144 = (60 - 25)/144 = 35/144$;

$M(X\cdot Y) = 3\cdot 0\cdot 1/2 + 3\cdot(-1)\cdot 1/12 + 2\cdot 0\cdot 1/12 + 2\cdot(-1)\cdot 1/3 =$

$= -1/4 - 2/3 = -(3 + 8)/12 = -11/12$.

По формуле (6.3.6) окончательно получаем

$$\rho_{XY} = (M(X\cdot Y) - M(X)\cdot M(Y)) / \sqrt{D(X)\cdot D(Y)} =$$

$$= (-11/12 + (5/12)\cdot(31/12)) / \sqrt{35/144\cdot 35/144} =$$

$$= (155/144 - 11/12)\cdot 144/35 = (23/144)\cdot(144/35) = 23/35 \approx 0.66.$$

Таким образом, коэффициент корреляции $\rho_{XY} \approx 0.66$, и что позволяет сделать вывод о наличии статистической связи между составляющими X, Y. ☻

Вопросы и задачи для самопроверки

1. Какая случайная величина называется двумерной случайной величиной?
2. Как задается закон распределения дискретной двумерной случайной величиной?
3. Как определить законы распределения каждой составляющей дискретной двумерной случайной величины?
4. Что такое условное математическое ожидание дискретной двумерной случайной величиной и как оно вычисляется?
5. Случайная величина X принимает значения 0, 1 или 2 с вероятностями, соответственно 0.2, 0.7 и 0.1, а не зависящая от нее случайная величина Y – значения –1, 0, 1 с вероятностями соответственно 0.3, 0.5, 0.2. Необходимо описать закон распределения двумерной случайной величины (X, Y) и законы распределения каждой составляющей двумерной случайной величины (X, Y).
6. В условиях предыдущей задачи вычислить условные математические ожидания $M(X \mid Y=-1)$, $M(X \mid Y=1)$, $M(Y \mid X=2)$.
7. Как определяется функция распределения непрерывной двумерной случайной величины? Каковы ее свойства?
8. Как определяется функция плотности распределения непрерывной двумерной случайной величины? Каковы ее свойства?
9. Как связаны между собой функция распределения и функция плотности вероятности непрерывной двумерной случайной величины?
10. Как определить плотности распределения каждой составляющей непрерывной двумерной случайной величины?
11. Дать определение независимости случайных величин.
12. Назвать критерий независимости составляющих дискретных и непрерывных случайных величин?
13. Как определяется корреляционный момент ρ_{XY} ?
14. Определить коэффициент корреляции. Объяснить его теоретико-вероятностный смысл.

15. Свойства коэффициента корреляции.
16. В каких случаях значения коэффициента корреляции равны +1 и –1?
17. Какова связь между независимостью и некоррелированностью случайных величин?
18. Закон распределения дискретной двумерной случайной величины задан следующей таблицей:

X \ Y	–1	0	1
–1	1/8	1/12	7/24
1	5/24	1/6	1/8

Определить коэффициент корреляции между составляющими этой случайной величины и сделать вывод о статистической взаимосвязи между X, Y.

19. Привести пример случайных величин, для которых понятия независимости и некоррелированности совпадают.

РАЗДЕЛ 2. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Тема 7. Основные понятия математической статистики

7.1. Задачи математической статистики

Как было показано в разделе 1, основные задачи теории вероятностей связаны с вычислением вероятностей случайных событий по известной функции распределения $F(x)$ или по функции плотности вероятности $p(x)$ (например, вычислить вероятность попадания значений случайной величины на заданный интервал $[\alpha, \beta]$: $P(X \in [\alpha, \beta])$). Однако на практике для реальных случайных величин эти статистические характеристики не всегда известны.

Поэтому возникает проблема определения неизвестных статистических характеристик случайного события или случайной величины, и основой для этого являются результаты выполненных опытов над исследуемым явлением, т.е. набор *статистических (экспериментальных) данных*. Решением подобных проблем занимается математическая статистика.

Математическая статистика – наука, изучающая методы исследования закономерностей в массовых случайных явлениях и процессах по данным, полученным из конечного числа наблюдений за ними.

Построенные на основании этих методов закономерности относятся не к отдельным испытаниям, а представляют собой утверждения об общих вероятностных характеристиках данного процесса. Такими характеристиками могут быть вероятности, плотности распределения вероятностей, математические ожидания, дисперсии и т.п.

Основными задачами математической статистики являются:

1. ***Оценка неизвестной функции распределения и функции плотности.*** По результатам n независимых испытаний над случайной величиной X получены ее значения $x_1, x_2, ..., x_n$. Тре-

буется оценить, хотя бы приближенно, неизвестные функции распределения $F(x)$ и плотности $p(x)$.

2. ***Оценка неизвестных параметров распределения.***

3. ***Проверка статистических гипотез*** относительно числовых характеристик случайной величины, плотности распределения.

Исходным материалом математической статистики являются *статистические данные*. Требуется наилучшим образом обработать информацию с целью получения по возможности более правильного решения о природе рассматриваемого статистического явления. Например, с некоторого участка было срезано n колосков и определено количество зерен в каждом из них. Как по результатам этих наблюдений определить среднее количество зерен в колосе на всем исследуемом участке, насколько однородны колоски? Ясно, что нельзя говорить о точном определении математического ожидания и дисперсии по конечному числу наблюдений. Не приходится говорить и о приближенном в обычном смысле определении этих характеристик. Говоря о приближенном определении какой-либо величины, обычно подразумевают, что можно указать пределы погрешности, за которые ошибка не выйдет. Однако из-за случайности результатов наблюдений полной гарантии, что ошибка не выйдет за данные пределы, быть не может. Поэтому в математической статистике говорят не о приближенных значениях неизвестных характеристик в обычном смысле, а об их приближенных значениях в вероятностном смысле, или об ***оценках неизвестных характеристик.***

Решение задач математической статистики обусловливает существенный объем вычислений, связанный с численной реализацией необходимого вычислительного алгоритма и графической интерпретацией результатов решения. Этому моменту в учебной литературе уделяется крайне мало внимания, что затрудняет использование методов математической статистики на практике.

Поэтому в данном разделе делается акцент на *изложении численных методик решения задач математической стати-*

стики в вычислительной среде табличного процессора Excel. Для каждой из рассматриваемых задач математической статистики кроме теоретических положений даются фрагменты документов Excel, реализующих алгоритмы решения задачи. При этом алгоритм решения может быть реализован путем программирования необходимых выражений в ячейках электронной таблицы или путем обращения к стандартным функциям или модулям Excel.

В учебном пособии будут использоваться обе рассмотренные возможности реализации требуемого вычислительного алгоритма. Поэтому предполагается, что читатель имеет достаточные навыки для реализации вычислений в Excel с использованием:

- программирования арифметических выражений в ячейках электронной таблицы;
- функций Excel (в основном математических и статистических).

Замечание 7.1.1. При описании обращения к той или иной стандартной функции Excel в качестве *формальных параметров* используются *имена переменных*, определенные в тексте пособия. При обращении к функции в качестве *фактических параметров* могут использоваться *константы, адреса ячеек, диапазоны адресов и арифметические выражения*. Например, описание функции для вычисления среднего арифметического значения (выборочного среднего) имеет вид:

$$\text{СРЗНАЧ}(x_1; x_2; ...; x_m),$$

где $x_1, x_2, ..., x_m$ – формальные параметры, количество которых не превышает 30 ($m \le 30$). Для вычисления среднего значения величин, находящихся в ячейках B3, B4, B5, B6, C3, C4, C5, C6, обращение к функции в соответствующей ячейке имеет вид:

=СРЗНАЧ(B3:B6;C3:C6),

т.е. в качестве фактических параметров используются два диапазона ячеек. ●

Замечание 7.1.2. Так как в запрограммированной ячейке выводится результат вычислений и не видно самого запрограммированного выражения, то в некоторых случаях рядом с результатом приводится (в другой ячейке) запрограммированное выражение (своеобразный комментарий к выполняемым вычислениям). В случаях, когда не очевидно, к какой ячейке относится приводимое выражение, используется стрелка, указывающая на нужную ячейку. ●

7.2. Выборочная совокупность и обработка ее элементов

***Генеральная совокупность* (*ГС*)** – это все однородные объекты (элементы), предназначенные для исследования. Каждой генеральной совокупности соответствует случайная величина (например, X), определяемая изучаемым признаком объекта. Например, для определения плотности распределения диаметра прошлифованного валика необходимо располагать набором возможных значений его диаметра. Количество объектов ГС называется ее объемом (N).

***Выборочная совокупность* (*ВС*),** или ***выборка* –** все случайно отобранные из ГС объекты, предназначенные для непосредственного исследования. Число объектов в выборочной совокупности называется объемом выборки (n).

Для того чтобы по измеренным значениям некоторого количественного показателя можно было достаточно уверенно судить обо всей совокупности, полученная выборка должна быть ***репрезентативной (представительной)***, т.е. правильно отражать пропорции генеральной совокупности. Предположим, например, что вся совокупность состоит из равного большого количества белых и черных шаров, помещенных в ящик, на дне которого имеется отверстие. Если черные шары сосредоточены в нижней части ящика, а белые – в верхней, то, открывая некоторое небольшое количество раз заслонку в отверстии ящика, получим выборку только из черных шаров. На основании такого способа отбора шаров нельзя сделать правильные выводы о содержании всей совокупности шаров, т.е. такая выборка не будет

репрезентативной. Выборка будет репрезентативной лишь тогда, когда все объекты генеральной совокупности будут иметь *одинаковую вероятность попасть в выборку*. Для этого шары в нашем примере должны быть перемешаны.

Требование ***репрезентативности*** означает, что:

1) должен быть обеспечен полностью ***случайный выбор*** n объектов из генеральной совокупности;
2) выборка должна иметь ***достаточно большой объем*** (желательно $n > 40–50$).

После получения выборочной совокупности все ее объекты обследуются по отношению к определенной случайной величине, т.е. обследуемому признаку объекта. В результате этого получают n значений $x_1, x_2, ..., x_n$, которые представляют собой множество чисел, расположенных в беспорядке. Анализ таких данных весьма затруднителен, и для изучения закономерностей полученные данные подвергаются определенной обработке.

Простейшая операция – ***ранжирование*** опытных данных, результатом которого являются значения, расположенные в порядке ***неубывания.*** Если среди элементов встречаются одинаковые, то они объединяются в одну группу.

Значение случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных, называется ***вариантом,*** а изменение этого значения – ***варьированием.*** Варианты будем обозначать строчными буквами с соответствующими порядковому номеру группы индексами $x^{(1)}, x^{(2)}, ..., x^{(m)}$, где m – число групп. При этом $x^{(1)} < x^{(2)} < ... < x^{(m)}$. Такую упорядоченную последовательность называют ***дискретным вариационным рядом.***

Численность отдельной группы сгруппированного ряда данных называется ***частотой*** n_i, где i – индекс варианта, а отношение частоты данного варианта к общей сумме частот называется ***частностью*** (или ***относительной частотой***) и обозначается ω_i, т.е.

$$\boxed{\omega_i = n_i / \sum_{i=1}^{m} n_i}, \qquad (7.2.1)$$

$i = 1, ..., m$. При этом имеют место следующие равенства:

$$\boxed{\sum_{i=1}^{m} n_i = n, \quad \sum_{i=1}^{m} \omega_i = 1}. \qquad (7.2.2)$$

Если набор данных $x_1, x_2, ..., x_n$ не содержит одинаковых значений, то число групп $m = n$, все $n_i = 1$, и дискретный вариационный ряд имеет вид:

$$x^{(1)} < x^{(2)} < ... < x^{(n-1)} < x^{(n)},$$

а относительная частота

$$\boxed{\omega_i = 1/n}, \quad i = 1, ..., n. \qquad (7.2.3)$$

Пример 7.2.1. На телефонной станции в течение 10 минут велись наблюдения над случайной величиной X – число неправильных соединений в минуту. Была получена следующая выборка объемом 10, т.е. (n = 10): 3; 1; 3; 1; 4; 1; 2; 4; 0; 3.

Необходимо построить дискретный вариационный ряд и определить его характеристики.

Решение. Результаты вычисления приведены в табл. 7.1.

Таблица 7.1

Индекс	i	1, 2, 3, 4, 5
Вариант	$x^{(i)}$	0, 1, 2, 3, 4
Частота	n_i	1, 3, 1, 3, 2; $\sum n_i = 10$
Относительная частота	ω_i	$\frac{1}{10}, \frac{3}{10}, \frac{1}{10}, \frac{3}{10}, \frac{2}{10}$; $\sum \omega_i = 1$

Вариационный ряд: $0<1<2<3<4$; $m=5$. ☻

Если число возможных значений дискретной случайной величины достаточно велико или наблюдаемая случайная величина является ***непрерывной***, то строят ***интервальный (или группированный) вариационный ряд,*** под которым понимают ***упорядоченную совокупность интервалов*** варьирования значений случайной величины с соответствующими частотами или относительными частотами попаданий в каждый из них значений случайной величины.

Как правило, частичные интервалы, на которые разбивается весь интервал варьирования, имеют одинаковую длину h и представимы в виде:

$$[z_i, z_i + h), \quad i = 1, 2, ..., m, \tag{7.2.4}$$

где m – число интервалов.

Длину h следует выбирать так, чтобы построенный ряд не был громоздким, но в то же время позволял выявлять характерные изменения случайной величины. Для вычисления h рекомендуется использовать следующую формулу:

$$\boxed{h = \frac{x_{\max} - x_{\min}}{1 + [3.222 \lg n]}}, \tag{7.2.5}$$

где $x_{\max}, x_{\min}$ – наибольшее и наименьшее значения выборки; $[x]$ означает целую часть числа x. Если окажется, что h – дробное число, то за длину интервала следует принять либо ближайшую простую дробь, либо ближайшую целую величину. При этом необходимо выполнение условий:

$$z_1 \le x_{\min}; \quad z_{m+1} = z_m + h \ge x_{\max}. \tag{7.2.6}$$

Из формулы (7.2.5) следует, что число интервалов m определяется по формуле Стерджеса:

$$\boxed{m = [3,222 \lg n] + 1}. \tag{7.2.7}$$

Середину каждого i-го интервала обозначим как z_i^*, т.е.

$$z_i^* = \frac{z_i + z_{i+1}}{2}.$$

Схема разбиения диапазона выборки на интервалы показана на рис. 7.1.

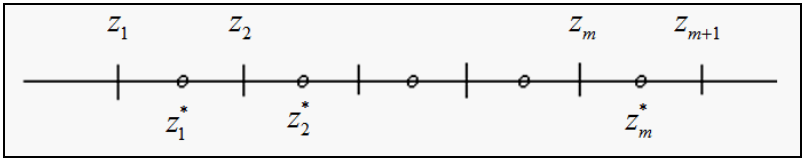


Рис. 7.1. Схема разбиения диапазона выборки на интервалы

Пример 7.2.2. При изменении диаметра валика после шлифовки была получена выборка объемом n = 14, представленная в табл. 7.2.

Таблица 7.2

20.3	15.4	17.2	19.2	23.3	18.1	21.9
15.3	12.8	13.2	20.4	16.5	19.7	20.5

Необходимо построить интервальный вариационный ряд, состоящий из четырех интервалов.

Решение. Так как наибольшая варианта равна 23.3, а наименьшая 12.8, то вся выборка попадает в интервал (12, 24). Длина каждого частичного интервала равна $\frac{24-12}{4} = 3$. Получаем следующие четыре интервала: $[12, 15)$; $[15, 18)$; $[18, 21)$; $[21, 24)$, а соответствующий интервальный вариационный ряд представлен в табл. 7.3. Здесь же приведены результаты вычислений характеристик n_i, ω_i. ☻

Таблица 7.3

X	12–15	15–18	18–21	21–24
z_i^*	13.5	16.5	19.5	22.5
n_i	2	4	3	5
ω_i	$\frac{2}{14}$	$\frac{4}{14}$	$\frac{3}{14}$	$\frac{5}{14}$

Заметим, что интервальный вариационный ряд, включая в себя строки z_i^* и ω_i, является аналогом закона распределения дискретной модели исследуемой генеральной совокупности. Действительно, z_i^* можно интерпретировать как значение дискретной случайной величины, а ω_i – вероятность соответствующего значения случайной величины.

7.3. Выборочная функция распределения. Гистограмма

В теории вероятностей для характеристики распределения случайной величины X служит функция распределения $F(x) = P(X < x)$, равная вероятности события $\{X < x\}$, где x – любое действительное число.

Одной из основных характеристик выборки является ***выборочная (эмпирическая) функция распределения***, определяемая формулой

$$\boxed{F_n^*(x) = \frac{n_x}{n}}, \qquad (7.3.1)$$

где n_x – количество элементов выборки, меньших x. Другими словами, $F_n^*(x)$ есть относительная частота появления события $A = \{X < x\}$ в n независимых испытаниях. Главное различие между $F(x)$ и $F_n^*(x)$ состоит в том, что $F(x)$ определяет вероятность события A, а выборочная функция распределения $F_n^*(x)$ – относительную частоту этого события.

Из определения (7.3.1) имеем следующие свойства выборочной функции распределения $F_n^*(x)$:

1. $0 \le F_n^*(x) \le 1.$ (7.3.2)
2. $F_n^*(x)$ – неубывающая функция.
3. $F_n^*(-\infty) = 0;\ F_n^*(\infty) = 1.$

Напоминаем, что такими же свойствами обладает и функция распределения $F(x)$ (вспомните эти свойства и сравните).

Функция $F_n^*(x)$ является «ступенчатой», имеются разрывы в точках, которым соответствуют наблюдаемые значения вариантов. Величина скачка равна относительной частоте варианта.

Аналитически $F_n^*(x)$ задается следующим соотношением:

$$\boxed{F_n^*(x) = \begin{cases} 0 \quad \text{при} \quad x \le x^{(1)}; \\ \sum_{j=1}^{i-1} \omega_j \quad \text{при} \quad x^{(i-1)} < x \le x^{(i)}, \quad i = 2, ..., m; \\ 1 \quad \text{при} \quad x > x^{(m)}, \end{cases}} \qquad (7.3.3)$$

где ω_i – соответствующие относительные частоты, определяемые выражением (7.2.1); $x^{(i)}$ – элементы вариационного ряда (варианты).

Замечание 7.3.1. В случае интервального вариационного ряда под $x^{(i)}$ понимается середина i-го частичного интервала. ●

Перед вычислением $F_n^*(x)$ полезно построить дискретный или интервальный вариационный ряд.

Пример 7.3.1. По данным примера 7.2.1 построить выборочную функцию распределения.

Решение. Используя данные таблицы примера 7.2.1, вычислим значения $F_{10}^*(x)$ и занесем их в табл. 7.4.

Таблица 7.4

x	$F_{10}^*(x)$
$x \le 0$	0
$0 < x \le 1$	$\omega_1 = \frac{1}{10}$
$1 < x \le 2$	$\omega_1 + \omega_2 = \frac{4}{10}$
$2 < x \le 3$	$\omega_1 + \omega_2 + \omega_3 = \frac{5}{10}$
$3 < x \le 4$	$\omega_1 + \omega_2 + \omega_3 + \omega_4 = \frac{8}{10}$
$x > 4$	$\omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 = \frac{10}{10} = 1$

Используя вычисленные значения $F_{10}^{*}(x)$, построим график выборочной функции распределения (рис. 7.2). Из графика видно, что $F_{10}^{*}(x)$ удовлетворяет вышеперечисленным свойствам. ☻

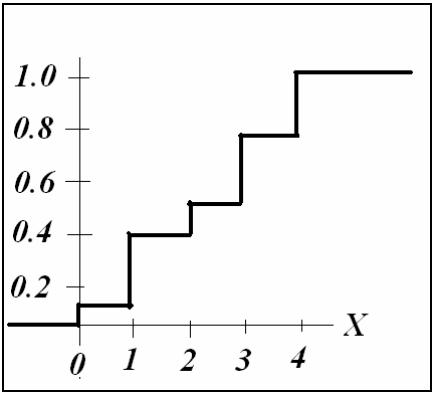


Рис. 7.2. График выборочной функции распределения

В случае интервального вариационного ряда под $x^{(i)}$ понимается середина *i*-го частичного интервала, т.е.

$$\boxed{x^{(i)} = z_i^*},\ i = 1, ..., m\ .$$

Пример 7.3.2. Построить выборочную функцию распределения по группированному вариационному ряду примера 7.2.2.

Решение. Используя данные таблицы примера 7.2.2, вычислим значения $F_{14}^{*}(x)$ и занесем их в табл. 7.5 (за варианты принимаются середины отрезков $x^{(i)} = z_i^*$). Используя вычисленные значения $F_{14}^{*}(x)$, построим график выборочной функции распределения (рис. 7.3). Из графика видно, что $F_{14}^{*}(x)$ удовлетворяет вышеперечисленным свойствам. ☻

Таблица 7.5

x	$F_{14}^{*}(x)$
$x \le 13.5$	0
$13.5 < x \le 16.5$	$\omega_1 = \dfrac{2}{14}$
$16.5 < x \le 19.5$	$\omega_1 + \omega_2 = \dfrac{6}{14}$
$19.5 < x \le 22.5$	$\omega_1 + \omega_2 + \omega_3 = \dfrac{9}{14}$
$x > 22.5$	$\omega_1 + \omega_2 + \omega_3 + \omega_4 = \dfrac{14}{14} = 1$

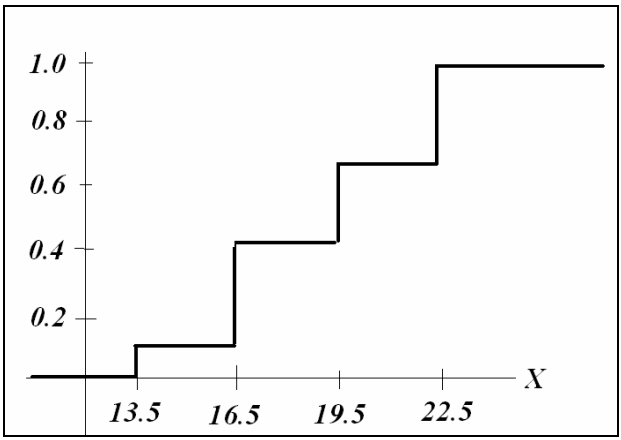


Рис. 7.3. График выборочной функции распределения $F_{14}^{*}(x)$

Напомним, что $F_n^{*}(x)$ равна относительной частоте появления события $A = \{X < x\}$ и, следовательно, при любом значении x величина $F_n^{*}(x)$ является случайной. Тогда конкретной вы-

борке $(x_1, x_2, ..., x_n)$ объема n соответствует функция распределения $F_n^*(x)$, которая в силу своей случайности будет отличаться от $F_n^*(x)$, построенной по другой выборке из той же генеральной совокупности. Возникает вопрос: зачем нужна такая характеристика, меняющаяся от выборки к выборке? Ответ получаем на основе следующих рассуждений.

По теореме Бернулли (см. раздел 5.4) относительная частота появления события A в n независимых опытах сходится по вероятности к вероятности $P(X < x)$ этого события при увеличении n. Следовательно, при больших объемах выборки выборочная функция распределения $F_n^*(x)$ близка к теоретической функции $F(x)$. Точнее, имеет место следующая теорема.

Теорема Гливенко. Для любого действительного числа x и любого $\varepsilon > 0$ имеет место сходимость по вероятности

$$\lim_{n\to\infty} P\left(\left|F_n^*(x) - F(x)\right| > \varepsilon\right) = 0. \qquad (7.3.4)$$

Таким образом, функция $F_n^*(x)$ ***является оценкой*** $F(x)$, т.е. позволяет приближенно определить неизвестную функцию $F(x)$. Точность оценивания возрастает при увеличении объема выборки.

В качестве ***оценки плотности распределения вероятности непрерывной случайной величины*** используют ***гистограмму относительных частот.***

Гистограммой относительных частот называется система прямоугольников, каждый из которых основанием имеет i-й интервал интервального вариационного ряда; площадь, равную относительной частоте ω_i; высота y_i определяется по формуле

$$\boxed{y_i = \frac{\omega_i}{h_i}}, \quad i = 1, 2, ..., m, \qquad (7.3.5)$$

где $h_i = z_{i+1} - z_i$ – длина i-го частичного интервала. Если длина частичных интервалов одинакова, то $h_i = h$.

Очевидно, что ***сумма площадей всех прямоугольников гистограммы относительных частот равна*** *1*. Действиительно,

$$\boxed{\sum_{i=1}^{m} S_i = \sum_{i=1}^{m} y_i h_i = \sum_{i=1}^{m} \frac{\omega_i}{h_i} h_i = \sum_{i=1}^{m} \omega_i = 1}. \qquad (7.3.6)$$

Возникает вопрос: ***почему в качестве оценки плотности распределения берут гистограмму относительных частот?*** Для ответа проведем следующие рассуждения. Площадь прямоугольника ω_i равна относительной частоте попадания элементов выборочной совокупности объема n в i-й интервал. По теореме Бернулли эта частота при больших значениях объема выборки n близка к вероятности:

$$p_i = P(z_i \le X < z_{i+1}) = \int_{z_i}^{z_{i+1}} p(x)dx. \qquad (7.3.7)$$

Пусть y_i – высота i-го прямоугольника. По теореме о среднем интеграл, выражающий вероятность в формуле (7.3.7), можно записать в виде:

$$p_i = \int_{z_i}^{z_{i+1}} p(x)dx = (z_{i+1} - z_i) \cdot p(u_i), \qquad (7.3.8)$$

где u_i – некоторое число из промежутка $[z_i, z_{i-1})$. Так как $\omega_i = (z_{i+1} - z_i) y_i$, то при больших значениях n и правильном формировании интервалов имеет место приближенное равенство:

$$(z_{i+1} - z_i) y_i \approx (z_{i+1} - z_i) \cdot p(u_i),,$$

из которого получаем:

$$\boxed{y_i \approx p(u_i)}. \qquad (7.3.9)$$

Практически это означает, что график плотности распределения генеральной совокупности X проходит вблизи верхних границ прямоугольников, образующих гистограмму. Поэтому

при больших объемах выборок и правильном выборе длины интервалов гистограмма напоминает график плотности распределения $p(x)$.

Пример 7.3.3. Построить гистограмму относительных частот выборочной совокупности (объем выборки 1000) нормально распределенной случайной величины $X = N(-20, 10)$ (сама выборка не показана).

Решение. В соответствии с формулой (7.2.5) принимаем количество интервалов $m = 11$, длина интервала равна 5.9, строим интервальный вариационный ряд и находим высоты y_i по формуле $y_i = \omega_i / 5.9$. График построенной гистограммы приведен на рис. 7.4. Здесь же точечной кривой показан график плотности $p(x)$ нормально распределенной случайной величины $X = N(-20, 10)$. Видно, что гистограмма достаточно хорошо описывает форму и значения распределения случайной величины. ☻

Замечание 7.3.1. Для того чтобы увеличить точность оценивания плотности распределения, необходимо увеличить количество наблюдений (объем выборки $n \to \infty$) при одновременном уменьшении шага разбиения диапазона выборки $h \to 0$. В этом случае можно сколь угодно точно приблизить ступенчатую гистограмму к теоретической кривой функции плотности вероятности. ●

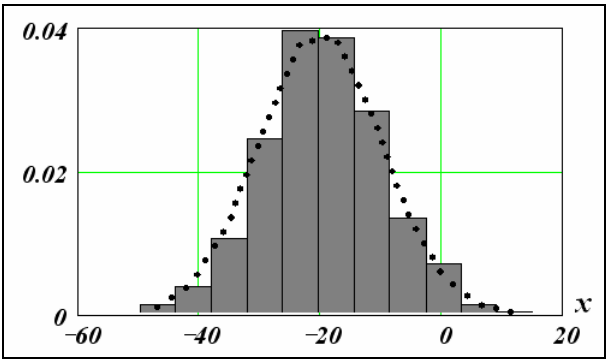


Рис. 7.4. Гистограмма и плотность распределения

7.4. Вычисление характеристик выборки в Excel

Вычисление частот. Для вычисления частот n_i можно использовать ***функцию ЧАСТОТА***, обращение к которой имеет вид:

=ЧАСТОТА(*массив_данных*;*массив_границ*)

,

где *массив_данных* – адреса ячеек, для которых вычисляется частота n_i; *массив_границ* – адреса ячеек, в которых размещаются упорядоченные по возрастанию значения z_j, $j = 1, 2, ..., m+1$, где m – число интервалов.

При использовании этой функции необходимо помнить:

1. Функция ЧАСТОТА вводится как формула массива, т.е. ***предварительно выделяется интервал ячеек***, в который будут помещены вычисленные частоты (число ячеек должно быть на 1 больше числа границ), затем вводится функция ЧАСТОТА с соответствующими аргументами, потом ***одновременно нажимаются клавиши*** **[*Ctrl*] + [*Shift*] + [*Enter*].**
2. Функция ЧАСТОТА игнорирует пустые ячейки и текстовые данные.
3. Если *массив_границ* не содержит возрастающих значений границ интервалов, то осуществляется автоматическое вычисление границ интервалов равной ширины, причем число интервалов равно корню квадратному из числа элементов *массива_данных*.

Результатом работы является массив значений, определяемый по следующему правилу: первый элемент равен числу n_0 элементов *массива_данных* меньше z_1; последний элемент равен числу n_{m+1} элементов *массива_данных* больше z_{m+1}; остальные элементы определяются как числа n_j элементов x_i *массива_данных*, удовлетворяющих условию

$$z_j < x_i \le z_{j+1}, \ j = 1, 2, ..., m .$$

Другими словами, кроме m значений частот n_j, $j = 1, 2, ..., m$, соответствующих m интервалам, вычисляются частоты n_0 (число

значений x_i, лежащих левее z_1) и n_{m+1} (число значений x_i, лежащих правее z_{m+1}).

Для подсчета количества элементов выборки (т.е. объема выборки) использовалась ***функция СЧЕТ***, обращение к которой имеет вид:

СЧЕТ(*массив_данных*),

где *массив_данных* – адреса ячеек или числовые константы.

Результатом работы является количество числовых величин в *массиве_данных*. При этом игнорируются пустые ячейки, логические значения, тексты и значения ошибок.

Пример 7.4.1. При изменении диаметра валика после шлифовки была получена следующая выборка объемом $n = 55$, приведенная в табл. 7.4.

Таблица 7.4

20.3	15.4	17.2	19.2	23.3	18.1	21.9
15.3	16.8	13.2	20.4	16.5	19.7	20.5
14.3	20.1	16.8	14.7	20.8	19.5	15.3
19.3	17.8	16.2	15.7	22.8	21.9	12.5
10.1	21.1	18.3	14.7	14.5	18.1	18.4
13.9	19.8	18.5	20.2	23.8	16.7	20.4
19.5	17.2	19.6	17.8	21.3	17.5	19.4
17.8	13.5	17.8	11.8	18.6	19.1	

Необходимо вычислить частоты и частности для семи заданных интервалов [10, 12); [12, 14); [14, 16); [16, 18); [18, 20); [20, 22); [22, 24), используя функцию ЧАСТОТА.

Решение. Начиная с ячейки А3 (рис. 7.5), введем в столбец А 55 элементов выборки (диапазон А3:А57) – на рисунке показаны только первые элементы выборки. Затем, начиная с ячейки В3, введем границы заданных интервалов. После подготовки этих данных выделим ячейки С3:С11, введем выражение =ЧАСТОТА(А3:А57;В3:В10) и нажмем одновременно клавиши [Ctrl] + [Shift] + [Enter]. В ячейках С3:С11 появится результат выполнения функции (см. рис. 7.5).

	A	B	C	D	E	F	G
1							
2	Выборочные значения	Границы интервалов	Частоты	Частности			
3	20,3	**10**	*0*	*0*			
4	15,3	**12**	*2*	*2/55*			
5	14,3	**14**	*4*	*4/55*	=C4/СЧЁТ(A$3:A$57)		
6	19,3	**16**	*8*	*8/55*			
7	10,1	**18**	*12*	*12/55*			
8	13,9	**20**	*15*	*3/11*			
9	19,5	**22**	*11*	*1/5*			
10	17,8	**24**	*3*	*3/55*			
11	15,4		*0*	*0*			
12	16,8		*55*	*1*			
13	20,1					=СУММ(D3:D11)	
14	17,8			=СУММ(C3:C11)			
15	21,1						
16	19,8	={ЧАСТОТА(A3:A57;B3:B10)}					
17	17,2						
18	13,5						
19	17,2						
20	13,2						

Рис. 7.5. Фрагмент вычисления частот и частностей

Для вычисления относительных частот ω_j (частностей) необходимо частоты поделить на число элементов выборки. Эти вычисления реализованы в ячейках D3:D11 (см. рис. 7.5). В знаменателе запрограммированных в этих ячейках выражений стоит обращение к функции СЧЁТ, описанной выше. Для контроля правильности вычисления частот и частностей в ячейках C12, D12 определены суммы (см. рис. 7.5):

$$\sum_{j=0}^{m+1=9} n_j = 55, \quad \sum_{j=0}^{m+1=9} \omega_j = 1. \ ☻$$

Вычисление ненормированной гистограммы частот. Заметим, что у ранее определенной гистограммы относительных

частот сумма площадей прямоугольников равна 1, а высота прямоугольников равна $y_j = n_j / (nh_j)$, где n – объем выборки; h_j – длина j-го интервала (т.е. выполнено нормирование – деление на длину интервала h_j).

Иногда в статистической (особенно зарубежной) литературе под гистограммой понимают систему прямоугольников, каждый из которых основанием имеет j-й интервал, а высота равна n_j. Очевидно, что сумма высот всех прямоугольников равна n. Такую гистограмму будем называть ***ненормированной*** (нет деления на h_j) ***гистограммой частот*** (нет деления на объем выборки n). Если высота прямоугольников определяется по формуле n_j / h_j, то такую гистограмму будем называть ***нормированной гистограммой частот.*** У такой гистограммы сумма площадей прямоугольников равна n.

Для построения ненормированной гистограммы частот необходимо:

- в табличном процессоре Excel 2003 обратиться к пункту **Сервис** строки меню Excel, а затем щелкнуть на команду *Анализ данных*, в появившемся окне диалога *Анализ данных* выбрать режим ***Гистограмма*** и нажать ОК.
- в табличном процессоре Excel 2007 или Excel 2010 обратиться к пункту **Данные** строки меню Excel, затем щелкнуть на команду *Анализ данных*, в появившемся окне диалога *Анализ данных* выбрать режим ***Гистограмма*** и нажать ОК.

Появится окно гистограммы, показанное на рис. 7.6. В окне задаются следующие параметры:

Входной интервал: – адреса ячеек, содержащие выборочные данные, по которым будет строиться гистограмма;

Интервал карманов: (необязательный параметр) – адреса ячеек, содержащие границы интервалов (кармана). Эти значения должны быть введены в возрастающем порядке.

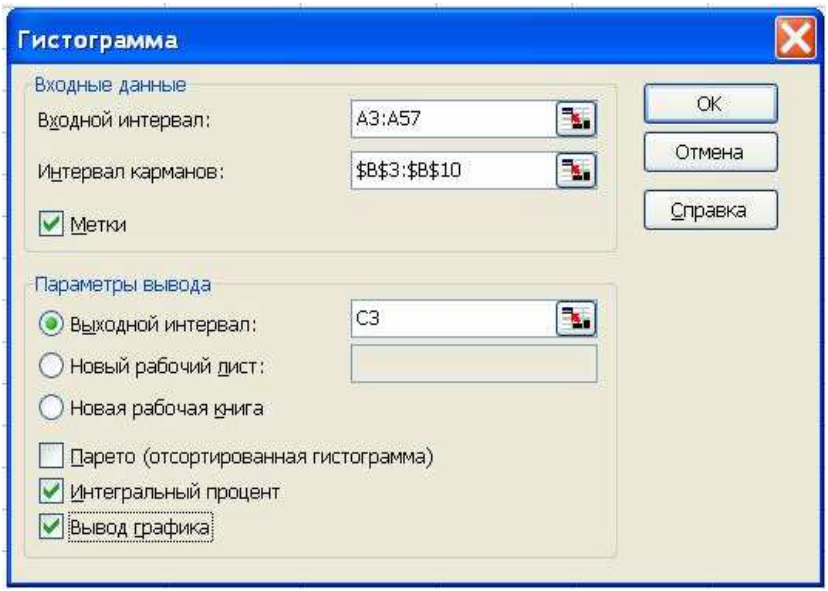


Рис. 7.6. Диалоговое окно режима ***Гистограмма***

Метки – флажок, включаемый, если первая строка во входных данных содержит заголовки. Если заголовки отсутствуют, то флажок следует выключить.

Выходной интервал: / Новый рабочий лист: / Новая рабочая книга. Включенный переключатель *Выходной интервал* требует ввода адреса верхней ячейки, начиная с которой будут размещаться вычисленные частоты n_j. В положении переключателя *Новый рабочий лист:* открывается новый лист, в котором начиная с ячейки A1 размещаются частоты n_j. В положении переключателя *Новая рабочая книга* открывается новая книга, на первом листе которой начиная с ячейки A1 размещаются частоты n_j.

Парето (*отсортированная гистограмма*) устанавливается в активное состояние, чтобы представить значения n_j в порядке их убывания. Если параметр выключен, то n_j приводятся в порядке следования интервалов.

Интегральный процент устанавливается в активное состояние для расчета выраженных в процентах накопленных относительных частот, т.е. вычисляются величины $\left(\sum_{j=1}^{i-1}\omega_j\right)\cdot 100\,\%$ (процентный аналог значений выборочной функции распределения при $x_i = z_j$, $j = 1, 2, ..., m+1$).

Вывод графика – устанавливается в активное состояние для автоматического создания встроенной диаграммы на листе, содержащем частоты n_j.

При использовании режима ***Гистограмма*** модуля *Анализ данных* необходимо помнить:

1. Частоты n_j вычисляются как количество элементов x_i выборки, удовлетворяющих условию

$$z_j < x_i \le z_{j+1}.$$

2. Если границы интервалов не заданы, т.е. в поле *Интервал карманов* отсутствуют адреса ячеек, то автоматически будет создан набор интервалов с одинаковой длиной $h = \dfrac{x_{\max} - x_{\min}}{[m]-1}$, где $[m]$ – целая часть величины $m = 1 + 3.322\cdot\lg n$, n – объем выборки.

Пример 7.4.2. По выборке примера 7.4.1 построить ненормированную гистограмму частот, используя режим ***Гистограмма*** модуля *Анализ данных*.

Решение. Начиная с ячейки А3 (рис. 7.7), введем в столбец А55 элементов выборки (диапазон А3:А57).

	A	B	C	D	E
2	Выборочные значения	Границы интервалов			
3	20,3	**10**	*10*	*Частота*	*Интегральный %*
4	15,3	**12**	**12**	**2**	**3,70%**
5	14,3	**14**	**14**	**4**	**11,11%**
6	19,3	**16**	**16**	**8**	**25,93%**
7	10,1	**18**	**18**	**12**	**48,15%**
8	13,9	**20**	**20**	**15**	**75,93%**
9	19,5	**22**	**22**	**10**	**94,44%**
10	17,8	**24**	**24**	**3**	**100,00%**
11	15,4		**Еще**	**0**	**100,00%**
12	16,8				
13	20,1				
14	17,8				
15	21,1				
16	19,8				
17	17,2				
18	13,5				
19	17,2				
20	13,2				
21	16,8				
22	16,2				
23	18,3				

Рис. 7.7. Фрагмент построения гистограммы (к примеру 7.4.2)

Затем обратимся к пункту **Сервис**, команде *Анализ данных*, режиму ***Гистограмма***. В появившемся диалоговом окне ***Гистограмма*** установим значения параметров, показанные на рис. 7.6, и после этого щелкнем на кнопку ОК. В ячейках D4:D11 выводятся вычисленные значения n_j, а в ячейках E4:E11 – значения интегрального процента. На этом же листе строится диаграмма, на которой отображаются вычисленные характеристики. ☻

Замечание 7.4.2. Как правило, гистограммы изображаются в виде смежных прямоугольных областей. Поэтому столбики гистограммы на рис. 7.7 целесообразно расширить до соприкосновения друг с другом. Для этого в табличном процессоре Excel 2003 необходимо щелкнуть мышью на диаграмме, далее на панели инструментов ***Диаграмма*** раскрыть список инструментов и

выбрать элемент *Ряд ‘Частота’*, после чего щелкнуть по кнопке **Формат ряда**. В появившемся одноименном диалоговом окне необходимо активизировать закладку ***Параметры*** и в поле *Ширина зазора* установить значение 0.

В табличном процессоре Excel 2007 или Excel 2010 необходимо щелкнуть правой кнопкой мыши на диаграмме и в появившемся контекстном меню выполнить команду *Формат ряда данных …*, а затем в появившемся диалоговом окне установить параметры, показанные на рис. 7.8. ●

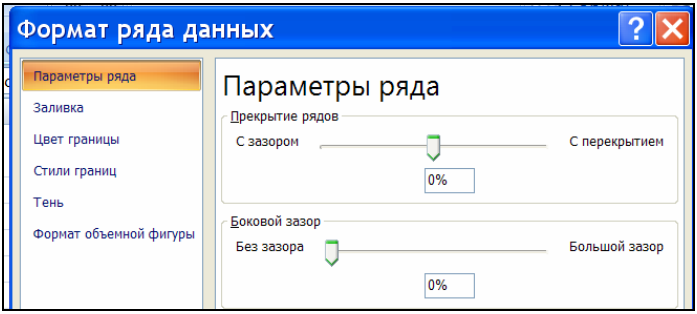


Рис. 7.8. Установка параметров в окне *Формат ряда данных*

На рис. 7.9 показана гистограмма, полученная из гистограммы рис. 7.7 путем действий, описанных в замечании 7.4.2.

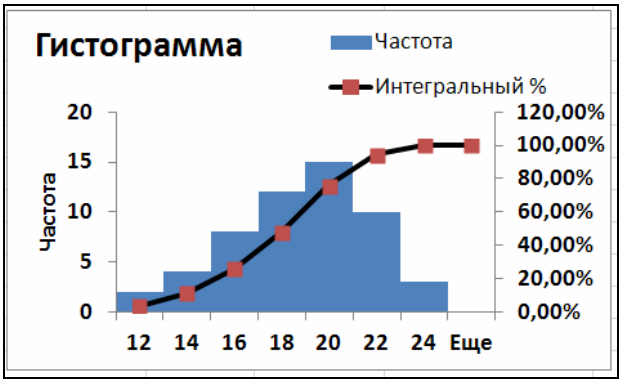


Рис. 7.9. График преобразованной гистограммы

Замечание 7.4.3. Ненормированная гистограмма частот не может служить оценкой для плотности распределения случайной величины, из значений которой была сформирована выборка, из-за того, что сумма площадей прямоугольников $\neq 1$. В качестве такой оценки нужно рассматривать гистограмму относительных частот. ●

Вычисление гистограммы относительных частот. Для вычисления такой гистограммы нужно сначала вычислить относительные частоты (частности), используя для этого функции ЧАСТОТА и СЧЁТ, а затем полученные значения поделить на длину h_j соответствующего интервала, т.е. получить высоту соответствующего прямоугольника $y_j = \omega_j / h_j$. По вычисленным значениям построить гистограмму. Для получения соприкасающихся прямоугольников гистограммы выполнить операции, описанные в замечании 7.4.2, для соответствующего элемента.

Пример 7.4.3. По выборке примера 7.4.1 построить гистограмму относительных частот.

Решение. Как и в примере 7.4.1, введем в столбце A выборочные значения и, используя функцию ЧАСТОТА, вычислим частоты. В столбце D запрограммируем вычисление высоты прямоугольников гистограммы по формуле $y_j = {n_j}/{n \cdot h_j}$, где $h_j = 2$, $n = 55$ и введем в E4:E10 значения середин интервалов. Для проверки правильности вычислений высот в ячейке D11 вычислена сумма $\sum y_j$. Очевидно, что площадь всех прямоугольников равна $2 \cdot \sum y_j = 1$. В заключение по данным столбцов E, D строим гистограмму (рис. 7.10). ☻

	A	B	C	D	E	F
1					=C4/55/2	
2	Выборочные значения	Границы интервалов	Частоты	Высота	Середины интервалов	
3	20,3	**10**	0			
4	15,3	**12**	2	0,018	11	
5	14,3	**14**	4	0,036	13	
6	19,3	**16**	8	0,073	15	
7	10,1	**18**	12	0,109	17	
8	13,9	**20**	15	0,136	19	
9	19,5	**22**	11	0,100	21	
10	17,8	**24**	3	0,027	23	
11	15,4		0	***0,500***		
12	16,8					
13	20,1					
14	17,8					
15	21,1					
16	19,8					
17	17,2					
18	13,5					
19	17,2					
20	13,2					
21	16,8					
22	16,2					
23	18,3					
24	18,5					

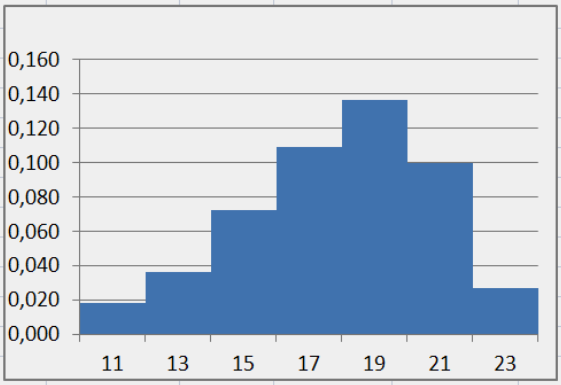


Рис. 7.10. Построение гистограммы относительных частот

Вопросы и задачи для самопроверки

1. Чем отличаются генеральная и выборочная совокупности?
2. Что такое дискретный вариационный ряд и как он строится?
3. Что такое интервальный вариационный ряд и как он строится?
4. Как вычисляются характеристики n_i и ω_i? Каковы их свойства?
5. На телефонной станции проводились наблюдения над числом X неправильных соединений в минуту. Наблюдения в течение часа дали следующие 60 значений:

   3; 1; 3; 1; 4; | 1; 2; 4; 0; 3; | 0; 2; 2; 0; 1; |1; 4; 3; 1; 1;
   4; 2; 2; 1; 1; | 2; 1; 0; 3; 4; | 1; 3; 2; 7; 2; |0; 0; 1; 3; 3;
   1; 2; 1; 2; 0; | 2; 3; 1; 2; 5; | 1; 2; 4; 2; 0; | 2; 3; 1; 2; 5.

   Необходимо: выполнить операции ранжирования и построить вариационный ряд (7 вариант), вычислить характеристики n_i и ω_i, $i = 1, ..., 7$.
6. Что такое выборочная функция распределения и как она строится?
7. Свойства выборочной функции распределения.
8. По данным задачи 5 необходимо построить выборочную функцию распределения.
9. Какую характеристику выборочной совокупности вычисляет функция Excel ЧАСТОТА и какие параметры имеет эта функция?
10. Что принимается в качестве оценки плотности распределения случайной величины?
11. По данным задачи 5 необходимо построить гистограмму относительных частот, используя функцию Excel ЧАСТОТА.
12. Используя режим Гистограмма, построить по данным задачи 5 гистограмму относительных частот (см. пример 7.4.3).

Тема 8. Точечные оценки параметров генеральной совокупности

8.1. Точечные оценки и их свойства

Большинство случайных величин, рассмотренных в курсе теории вероятностей, имели распределения, зависящие от одного или нескольких параметров. Так, биномиальное распределение зависит от параметров p и n, нормальное – от параметров a и σ и т.п. Одной из основных задач математической статистики является оценивание этих параметров по наблюдаемым данным, т.е. по выборочной совокупности.

Выборочная характеристика, используемая в качестве приближенного значения неизвестного параметра генеральной совокупности, называется *точечной оценкой* этого параметра. В этом определении слово «точечная» означает, что значение оценки представляет собой число или точку на числовой оси.

Замечание 8.1.1. При построении точечной оценки наряду с генеральной совокупностью X будем рассматривать n независимых случайных величин, обозначаемых той же буквой, что и генеральная совокупность, и имеющих точно такое же распределение, как генеральная совокупность. Итак, $X_1, X_2, ..., X_n$ – n независимых экземпляров X. Если $F(x)$ – функция распределения генеральной совокупности X, то у каждой случайной величины X_i функция распределения также равна $F(x)$. Понятно, что получить n значений случайной величины X – все равно, что получить одно значение n-мерной случайной величины ($X_1, X_2, ..., X_n$). Поэтому каждую выборку $x_1, x_2, ..., x_n$ объема n можно рассматривать как одно значение n-мерной случайной величины ($X_1, ..., X_n$). ●

Обозначим через θ некоторый неизвестный параметр генеральной совокупности, а через $\hat{\theta}$ – точечную оценку этого параметра. Оценка $\hat{\theta}$ есть некоторая функция $\varphi(X_1, X_2, ..., X_n)$ от n независимых экземпляров $X_1, X_2, ..., X_n$ генеральной совокупно-

сти, где n – объем выборки. Поэтому оценка $\hat{\theta}$, как функция случайных величин, также является случайной, и свойства $\hat{\theta}$ можно исследовать с использованием понятий теории вероятностей.

Чтобы оценка $\hat{\theta}$ была хорошим приближением к оцениваемой характеристике θ генеральной совокупности, к ней предъявляются следующие требования.

Несмещенность – оценка $\hat{\theta}$ является несмещенной, если ее математическое ожидание $M(\hat{\theta})$ равно оцениваемому параметру θ:

$$\boxed{M(\hat{\theta}) = \theta}. \qquad (8.1.1)$$

Поясним смысл этого равенства следующим примером. Имеются два алгоритма вычисления оценок для параметра θ. Значения оценок, построенных первым алгоритмом по различным выборкам объема n генеральной совокупности, приведены на рис. 8.1*а*, а с использованием второго алгоритма – на рис. 8.1*б*. Видим, что среднее значение оценок на рис. 8.1*а* совпадает с θ, и, естественно, такие оценки предпочтительнее по сравнению с оценками на рис. 8.1*б*, которые концентрируются слева от значения θ и для которых $M(\hat{\theta}) < \theta$, т.е. эти оценки являются смещенными.

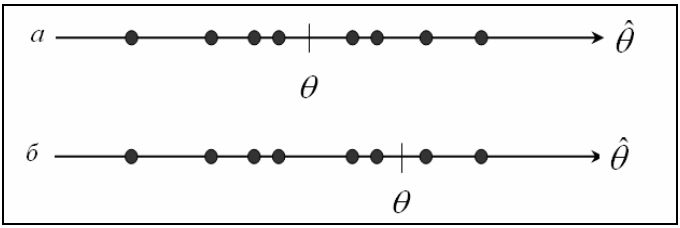


Рис. 8.1. К определению несмещенности оценок

Состоятельность. Оценка $\hat{\theta}$ называется *состоятельной*, если для любого $\varepsilon > 0$ при $n \to \infty$

$$P\left(\left|\hat{\theta} - \theta\right| < \varepsilon \right) \to 1. \qquad (8.1.2)$$

Поясним смысл этого предельного соотношения. Пусть ε – очень малое положительное число. Тогда (8.1.3) означает, что чем больше число наблюдений *n*, тем больше уверенность (вероятность) в незначительном отклонении $\hat{\theta}$ от неизвестного параметра θ. Очевидно, что «хорошая» оценка должна быть состоятельной, иначе она не имеет практического смысла, так как увеличение объема исходной информации не будет приближать нас к «истинному» значению θ.

Эффективность. Предположим, что имеются две состоятельные и несмещенные оценки $\hat{\theta}^{(1)}$ и $\hat{\theta}^{(2)}$ одного и того же параметра θ. Как из двух этих оценок выбрать лучшую? Каждая из них является случайной величиной, и нельзя предсказать индивидуальное значение оценки в каждом частном случае. Однако, рассматривая в качестве меры разброса оценки $\hat{\theta}$ около значения параметра θ величину дисперсии $D(\hat{\theta}) = M(\hat{\theta} - \theta)^2$, можно теперь точно охарактеризовать эффективность оценок $\hat{\theta}^{(1)}$ и $\hat{\theta}^{(2)}$. Если $\hat{\theta}^{(1)}$ и $\hat{\theta}^{(2)}$ – несмещенные оценки параметра θ и если $D(\hat{\theta}^{(1)}) < D(\hat{\theta}^{(2)})$, то оценка $\hat{\theta}^{(1)}$ является более эффективной, чем оценка $\hat{\theta}^{(2)}$. Следовательно, из всех несмещенных оценок $\hat{\theta}$ ***эффективной*** (наилучшей) оценкой является та, которая обладает наименьшей дисперсией. На рис. 8.2 показаны значения оценок $\hat{\theta}^{(1)}$ и $\hat{\theta}^{(2)}$. Видно, что разброс оценки $\hat{\theta}^{(1)}$ относительно ее математического ожидания θ меньше, чем у оценки $\hat{\theta}^{(2)}$.

Следовательно, оценка $\hat{\theta}^{(1)}$ является эффективной, если сравнивать только две оценки $\hat{\theta}^{(1)}$ и $\hat{\theta}^{(2)}$.

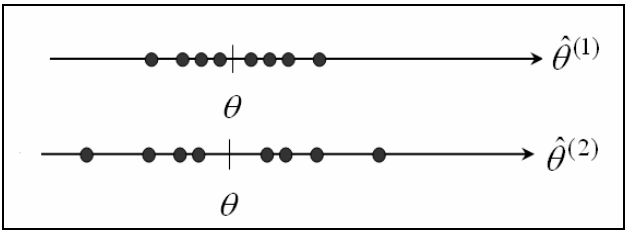


Рис. 8.2. К определению эффективности оценок

8.2. Выборочное среднее и выборочная дисперсия

Рассмотрим две точечные оценки для числовых характеристик математического ожидания и дисперсии и установим свойства этих оценок.

Выборочным средним $\overline{X}_в$ называется случайная величина, определенная формулой

$$\boxed{\overline{X}_в = \frac{X_1 + X_2 + ... + X_n}{n}}. \qquad (8.2.1)$$

Так как конкретная выборка $x_1, ..., x_n$ является реализацией значений случайных величин $X_1, ..., X_n$ (см. замечание 8.1.1), то среднее значение данной выборки определяется через наблюдаемые значения $x_1, ..., x_n$ по формуле:

$$\boxed{\overline{x}_в = \frac{x_1 + x_2 + ... + x_n}{n}} \qquad (8.2.2)$$

и является одной из реализаций случайной величины $\overline{X}_в$. Другими словами, $\overline{x}_в$ ***есть одно из значений случайной величины*** $\overline{X}_в$.

Если данные представлены в виде вариационного ряда, то для вычисления выборочного среднего целесообразно применить одно из следующих соотношений:

- для дискретного вариационного ряда

$$\boxed{\overline{x}_в = \frac{\sum_{i=1}^{m} x^{(i)} n_i}{\sum_{i=1}^{m} n_i} = \sum_{i=1}^{m} x^{(i)} \omega_i}; \qquad (8.2.3)$$

- для интервального вариационного ряда

$$\boxed{\overline{x}_в = \frac{\sum_{i=1}^{m} z_i^* n_i}{\sum_{i=1}^{m} n_i} = \sum_{i=1}^{m} \omega_i z_i^*}, \qquad (8.2.4)$$

где ω_i – частность (относительная частота), соответствующая i-й варианте или i-му частичному интервалу; z_i^* – середина i-го частичного интервала, т.е. $z_i^* = \frac{(z_i + z_{i+1})}{2}, \quad i = 1, 2, ..., m.$

Сравним математическое ожидание дискретной случайной величины X, вычисляемое по формуле

$$M(X) = \sum_{i=1}^{m} x_i p_i, \qquad (8.2.5)$$

и значение выборочного среднего, определяемое (8.2.3). Прежде всего, очевидна их внешняя схожесть. Однако в формуле (8.2.5) x_i – возможные значения случайной величины, а p_i – вероятности. В формуле (8.2.4) $x^{(i)}$ – варианты случайной величины, полученные в результате наблюдений; ω_i – их относительная частота. Далее, математическое ожидание не является случайной величиной, а выборочное среднее – случайная величина, значение которой меняется от выборки к выборке. Несмотря на это, выборочное среднее при определенных условиях выступает как «хорошая» оценка математического ожидания.

Пример 8.2.1. В результате обработки выборки объемом $n = 60$ был получен дискретный вариационный ряд, приведенный в табл. 8.1.

Таблица 8.1

Индекс	i	1, 2, 3, 4, 5, 6, 7
Вариант	$x^{(i)}$	0, 1, 2, 3, 4, 5, 7
Частота	n_i	8, 17, 16, 10, 6, 2, 1
Частность	ω_i	$\frac{8}{60}, \frac{17}{60}, \frac{16}{60}, \frac{10}{60}, \frac{6}{60}, \frac{2}{60}, \frac{1}{60}$

Необходимо вычислить значение выборочного среднего по этой выборке.

Решение. Используя дискретный вариационный ряд (см. табл. 8.1) и соотношение (8.2.3), имеем:

$$\overline{x}_в = 0 \cdot \frac{8}{60} + 1 \cdot \frac{17}{60} + 2 \cdot \frac{16}{60} + 3 \cdot \frac{10}{60} + 4 \cdot \frac{6}{60} + 5 \cdot \frac{2}{60} + 7 \cdot \frac{1}{60} = 2.0. \ ☻$$

Приведем некоторые свойства выборочного среднего $\overline{X}_в$.

1. Оценка $\overline{X}_в$ ***является несмещенной, состоятельной и эффективной оценкой*** для математического ожидания генеральной совокупности.

2. Если случайная величина X подчиняется нормальному закону распределения $N(a, \sigma)$ с параметрами a и σ, то $\overline{X}_в$ также ***подчиняется нормальному закону*** с параметрами a и $\frac{\sigma}{\sqrt{n}}$, т.е. $\overline{X}_в = N(a, \frac{\sigma}{\sqrt{n}}))$. Следовательно, оценкой для параметра a является $\hat{a} = \overline{X}_в$, и эта оценка удовлетворяет требованиям несмещенности, состоятельности и эффективности. Видно, что

при увеличении объема выборки n дисперсия оценки $D(\hat{a}) = \dfrac{\sigma^2}{n}$ уменьшается, т.е. оценка состоятельна.

В качестве точечной ***оценки дисперсии*** $D(X)$ случайной величины X используют выборочную дисперсию:

$$\boxed{D_в = \sum_{i=1}^{n} \frac{(X_i - \overline{X}_в)^2}{n}}. \qquad (8.2.6)$$

Так как конкретная выборка $x_1, ..., x_n$ является реализацией значений случайных величин $X_1, ..., X_n$ (см. замечание 8.1.1), то выборочная дисперсия данной выборки определяется через наблюдаемые значения $x_1, ..., x_n$ по формуле

$$\boxed{d_в = \sum_{i=1}^{n} \frac{(x_i - \overline{x}_в)^2}{n}}. \qquad (8.2.7)$$

Если данные представлены в виде вариационного ряда, то целесообразно для вычислений $d_в$ вместо (8.2.7) использовать следующие соотношения:

- для дискретного вариационного ряда

$$d_в = \frac{\sum_{i=1}^{m} (x^{(i)} - \overline{x}_в)^2 n_i}{n} = \sum_{i=1}^{m} (x^{(i)} - \overline{x}_в)^2 \omega_i \; ; \qquad (8.2.8)$$

- для интервального вариационного ряда

$$d_в = \frac{\sum_{i=1}^{m} (z_i^* - \overline{x}_в)^2 n_i}{n} = \sum_{i=1}^{m} (z_i^* - \overline{x}_в)^2 \omega_i \; , \qquad (8.2.9)$$

где ω_i, z_i^* – те же, что и в формулах (8.2.3), (8.2.4).

Можно показать справедливость следующих выражений, являющихся аналогами (8.2.7), (8.2.8), (8.2.9) соответственно:

$$d_в = \frac{1}{n}\sum_{i=1}^{n}(x^{(i)})^2 - (\overline{x}_в)^2; \qquad (8.2.10)$$

$$d_в = \sum_{i=1}^{m}(x^{(i)})^2 \omega_i - (\overline{x}_в)^2; \qquad (8.2.11)$$

$$d_в = \sum_{i=1}^{m}(z_i^*)^2 \omega_i - (\overline{x}_в)^2. \qquad (8.2.12)$$

Приведенные соотношения (8.2.10)–(8.2.12) оказываются более удобными для программной реализации вычислений значения $d_в$. Однако если дисперсия σ^2 существенно меньше квадрата математического ожидания, т.е. $\sigma^2 << (M(x))^2$, то из-за ошибок округления при машинном счете по этим формулам возможна ситуация $d_в < 0$. Тогда следует воспользоваться формулами (8.2.7)–(8.2.9).

Сравним формулу (8.2.7) с формулой дисперсии дискретной случайной величины:

$$D(X) = \sum_{i=1}^{m}(x_i - M(X))^2 p_i\,. \qquad (8.2.13)$$

Различие между этими формулами состоит в том, что: а) величина $D(X)$ не случайна, $d_в$ – значение случайной величины, которое может меняться от выборки к выборке; б) в формуле (8.2.13) x_i – возможные значения случайной величины X; p_i – их вероятности; $M(X)$ – математическое ожидание. В формуле (8.2.8) $x^{(i)}$ – варианты случайной величины; ω_i – их относительные частоты; $\overline{x}_в$ – значения выборочного среднего. Несмотря на различия, между этими двумя формулами много общего. Во-первых, обе они являются мерой рассеивания. Во-вторых, кроме внешнего сходства формул, соответствующие дисперсии обладают схожими свойствами. В-третьих, выборочная дисперсия при определенных условиях является хорошей оценкой для дисперсии $D(X)$.

Пример 8.2.2. В результате обработки выборки объемом $n = 60$ был получен дискретный вариационный ряд, приведенный в табл. 8.1. Необходимо вычислить значение выборочной дисперсии.

Решение. Воспользуемся формулой (8.2.11). Первоначально, используя дискретный вариационный ряд (см. табл. 8.1), вычислим:

$$\sum_{i=1}^{7}(x^{(i)})^2\omega_i = 0\cdot\frac{8}{60}+1\cdot\frac{17}{60}+4\cdot\frac{16}{60}+9\cdot\frac{10}{60}+$$

$$+16\cdot\frac{6}{60}+25\cdot\frac{2}{60}+49\cdot\frac{1}{60}=6.09.$$

Так как значение $\overline{x}_в$ было вычислено в примере 8.2.1 ($\overline{x}_в = 2.0$), то

$$d_в = \sum_{i=1}^{7}(x^{(i)})^2\omega_i - (\overline{x}_в)^2 = 6.09 - 4.0 = 2.09.. \quad ☻$$

Выборочная дисперсия $D_в$ является состоятельной, но ***смещенной оценкой.*** Можно показать, что

$$\boxed{M(D_в) = \frac{n-1}{n}D(X)}, \qquad (8.2.14)$$

следовательно, $D_в$ – смещенная оценка для дисперсии генеральной совокупности.

Формула (8.2.14) позволяет указать состоятельную и несмещенную оценку для генеральной дисперсии. Для этого рассмотрим случайную величину

$$\boxed{S^2 = \frac{n}{n-1}D_в = \sum_{i=1}^{n}\frac{(X_i - \overline{X}_в)^2}{n-1}}, \qquad (8.2.15)$$

называемую ***исправленной дисперсией***. Заметим, что для выборок большого объема множитель $\frac{n}{n-1}$ близок к 1, поэтому слу-

чайные величины S^2 и $D_в$ мало отличаются друг от друга. Однако для выборок малого объема это отличие может быть существенным.

Возникает вопрос: будет ли оценка S^2 эффективной? Ответ отрицательный: оценка S^2, будучи несмещенной оценкой дисперсии $D(X)$, ***не является эффективной оценкой***. Относительное увеличение дисперсии оценки S^2 по сравнению с минимально возможной дисперсией определяется величиной $\frac{n}{n-1} > 1$, но при достаточно больших n это увеличение пренебрежимо мало.

Так как конкретная выборка $x_1, ..., x_n$ является реализацией значений случайных величин $X_1, ..., X_n$ (см. замечание 8.1.1), то выборочная исправленная дисперсия s^2 определяется через наблюдаемые значения $x_1, ..., x_n$ по формуле

$$\boxed{s^2 = \frac{n}{n-1} d_в = \sum_{i=1}^{n} \frac{(x_i - \overline{x}_в)^2}{n-1}}. \qquad (8.2.16)$$

8.3. Точечные оценки вероятности события и коэффициента корреляции

Обозначим через $p(A)$ неизвестную вероятность события A в одном испытании. Для оценивания $p(A)$ проведем n независимых испытаний, в которых событие A произошло m раз. Тогда случайная величина

$$\boxed{\hat{p} = \frac{m}{n}} \qquad (8.3.1)$$

является относительной частотой события A. Свойства этой точечной оценки определяет следующее утверждение.

Относительная частота $\hat{p} = m/n$ появления события A в n испытаниях есть ***состоятельная, несмещенная и эффективная оценка вероятности*** $p(A)$.

В теории вероятностей было определено, что случайные величины X и Y называются независимыми, если при любых i и j вероятность события $\{X = x_i\}$ не зависит от того, произошло ли событие $\{Y = y_j\}$, и наоборот. Таким образом, независимые случайные величины не могут влиять друг на друга. Если случайные величины зависимы, то между ними существует статистическая связь. Количественной мерой взаимосвязи случайных величин является коэффициент корреляции (см. раздел 6.3):

$$\boxed{\rho_{X,Y} = \frac{M\left[(X - M(X))(Y - M(Y))\right]}{\sigma_X \sigma_Y}}.$$

Точечной оценкой коэффициента корреляции является выборочный коэффициент корреляции:

$$\boxed{R_{XY} = \frac{\overline{XY}_в - \overline{X}_в \cdot \overline{Y}_в}{\sqrt{D_{Xв} \cdot D_{Yв}}}}, \qquad (8.3.2)$$

где оценка корреляционного момента $\overline{XY}_в$ вычисляется по формуле

$$\boxed{\overline{XY}_в = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} X_i Y_j}. \qquad (8.3.3)$$

Так как конкретная выборка $x_1, ..., x_n$ является реализацией значений случайных величин $X_1, ..., X_n$ (см. замечание 8.1.1), то выборочное значение коэффициента корреляции r_{XY} определяется по формуле

$$\boxed{r_{XY} = \frac{\overline{xy}_в - \overline{x}_в \cdot \overline{y}_в}{\sqrt{d_{Xв} \cdot d_{Yв}}}}, \qquad (8.3.4)$$

где

$$\boxed{\overline{xy}_в = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j}, \qquad (8.3.5)$$

а выборочные значения $\overline{x}_в$, $\overline{y}_в$, $d_{Xв}$, $d_{Yв}$ вычисляются по выборкам из генеральных совокупностей X, Y по выше приведенным формулам.

8.4. Вычисление точечных оценок в Excel

Рассмотрим несколько функций Excel, полезных при вычислении точечных оценок.

Функция СРЗНАЧ вычисляет выборочное среднее (см. раздел 8.2):

$$\boxed{\overline{X}_в = \frac{X_1 + X_2 + ... + X_n}{n}}. \qquad (8.4.1)$$

Обращение к функции имеет вид:

=СРЗНАЧ(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числа или адреса ячеек, содержащие значения выборки $x_1, x_2, ..., x_n$. Ячейки, содержащие текстовые, логические данные или пустые, при вычислении выборочной дисперсии игнорируются.

Функция ДИСП вычисляет исправленную дисперсию (см. раздел 8.2):

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_в)^2}. \qquad (8.4.2)$$

Обращение к функции имеет вид:

=ДИСП(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числа или адреса ячеек, содержащие значения выборки $x_1, x_2, ..., x_n$. Ячейки, содержащие текстовые, логические данные или пустые, при вычислении выборочной дисперсии игнорируются.

Функция ДИСПР вычисляет выборочную дисперсию по формуле

$$\boxed{D_в = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_в)^2}. \qquad (8.4.3)$$

Обращение к функции имеет вид:

=ДИСПР(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числа или адреса ячеек, значения выборки $x_1, x_2, ..., x_n$. Ячейки, содержащие текстовые, логические данные или пустые, при вычислении выборочной дисперсии игнорируются.

Функция КВАДРОТКЛ вычисляет сумму квадратов:

$$\boxed{\sum_{i=1}^{n}(X_i - \bar{X}_в)^2}. \qquad (8.4.4)$$

Обращение к функции имеет вид:

=КВАДРОТКЛ (*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числа или адреса ячеек, содержащие значения выборки $x_1, x_2, ..., x_n$.

Пример 8.4.1. По выборке, приведенной в табл. 8.4, вычислить выборочное среднее и исправленную дисперсию (8.4.1).

Таблица 8.4

20.3	15.4	17.2	19.2	23.3	18.1	21.9
15.3	16.8	13.2	20.4	16.5	19.7	20.5
14.3	20.1	16.8	14.7	20.8	19.5	15.3
19.3	17.8	16.2	15.7	22.8	21.9	12.5
10.1	21.1	18.3	14.7	14.5	18.1	18.4
13.9	19.8	18.5	20.2	23.8	16.7	20.4
19.5	17.2	19.6	17.8	21.3	17.5	19.4
17.8	13.5	17.8	11.8	18.6	19.1	–

Решение. Начиная с ячейки А3, введем в столбец А 55 элементов выборки (рис. 8.3). Затем, используя функции СРЗНАЧ, КВАДРОТКЛ, ДИСП (как показано на рис. 8.3), вычислим требуемые оценки. Видно ожидаемое совпадение вычисленных значений исправленной дисперсии с помощью функций КВАДРОТКЛ, ДИСП. ☻

Функция СТАНДОТКЛОН вычисляет величину

$$\boxed{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_в)^2}}.$$

Обращение к ней имеет вид:

$$\boxed{=\text{СТАНДОТКЛОН}(\textit{арг1; арг2; …; арг30})},$$

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащие значения выборки $x_1, x_2, ..., x_n$.

Функция СТАНДОТКЛОНП вычисляет величину

$$\boxed{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_в)^2}}.$$

Обращение к ней имеет вид:

=СТАНДОТКЛОНП(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащие значения выборки $x_1, x_2, ..., x_n$.

	A	B	C	D
1				
2	Выборочные значения			
3	20,3		=СРЗНАЧ(A3:A57)	
4	15,3			
5	14,3		***17,907***	
6	19,3			
7	10,1	=КВАДРОТКЛ(A3:A57)/(55-1)		
8	13,9			
9	19,5		***8,760***	
10	17,8			
11	15,4		***8,760***	
12	16,8			
13	20,1		=ДИСП(A3:A57)	
14	17,8			
15	21,1			
16	19,8			

Рис. 8.3. Фрагмент вычисления точечных оценок

Функция МОДА вычисляет наиболее часто встречающееся значение в заданных аргументах функции, т.е. значение, встречающееся в выборке с максимальной частотой.

Обращение к функции имеет вид:

=МОДА(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащие значения выборки $x_1, x_2, ..., x_n$.

Если в заданных значениях аргументов *нет повторяющихся значений*, то функция возвращает признак ошибки #Н/Д.

Функция МЕДИАНА вычисляет значение выборки, приходящееся на середину упорядоченной выборочной совокупности. Если выборка имеет четное число элементов, то значение функции будет равно среднему двух значений, находящихся посередине упорядоченной выборочной совокупности. Например, медиана выборки (200, 236, 250, 305, 337, 220) будет равна (236 + 250) / 2 = 243.

Обращение к функции имеет вид:

=МЕДИАНА(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащие значения выборки $x_1, x_2, ..., x_n$.

Пример 8.4.2. По выборке, приведенной в табл. 8.4, вычислить выборочное среднее $\overline{x}_в$ и выборочную дисперсию $d_в$ двумя способами:

Способ 1. Программируя в ячейках Excel необходимые вычисления.

Способ 2. Используя функции Excel СРЗНАЧ, ДИСПР.

Решение. Начиная с ячейки А3, введем в столбец А 55 элементов выборки (диапазон А3:А57). На рис. 8.4 в столбце *А* показаны не все элементы выборки. Запрограммируем выражения (8.2.2), (8.2.7), используя функции СУММ, КВАДРОТКЛ с аргументами, указанными на рис. 8.4. Затем вычислим характеристики: выборочные средние и дисперсию с использованием статистических функций СРЗНАЧ, ДИСПР (см. рис. 8.4). Как и следовало ожидать, результаты вычислений двумя способами совпали. ☻

	A	B	C	D	E
1					
2	Выборочные значения				
3	20,3		Программирование		
4	15,3				
5	14,3	*17,907*			
6	19,3		=СУММ(A3:A57)/55		
7	10,1		=КВАДРОТКЛ(A3:A57)/55		
8	13,9	*8,601*			
9	19,5				
10	17,8				
11	15,4		Стандартные функции Excel		
12	16,8				
13	20,1		=СРЗНАЧ(A3:A57)		
14	17,8	*17,907*			
15	21,1	*8,601*			
16	19,8		=ДИСПР(A3:A57)		
17	17,2				

Рис. 8.4. Вычисление выборочных среднего и дисперсии

Вопросы и задания для самопроверки

1. Что называется точечной оценкой параметра генеральной совокупности?
2. Дать определение несмещенности и состоятельности, привести примеры точечных оценок, обладающих этими свойствами.
3. Чем вызвана необходимость введения исправленной выборочной дисперсии?
4. Что такое эффективность точечной оценки?
5. Является ли исправленная выборочная дисперсия эффективной оценкой?

6. Как повысить точность получаемых нами оценок различных параметров генеральной совокупности?
7. Какими свойствами обладает выборочный коэффициент корреляции?
8. Как в табличном процессоре Excel вычислить выборочное среднее (как минимум привести два способа)?
9. Какую точечную оценку вычисляет функция ДИСП?
10. Какую точечную оценку вычисляет функция ДИСПР? В чем отличие от функции ДИСП?
11. Как, используя функцию СТАНДОТКЛОН, вычислить выборочную дисперсию $D_в$?

Тема 9. Интервальные оценки параметров распределения генеральной совокупности

9.1. Определение интервальных оценок

Точечные оценки дают приближенное значение неизвестного параметра θ и используются в тех случаях, когда требуется поставить некоторое число $\hat{\theta}$ вместо неизвестного θ, т.е. заменить θ его оценкой $\hat{\theta}$. При этом возникает вопрос: как сильно может отличаться это $\hat{\theta}$ от истинного значения θ? Другими словами, можно ли указать такую величину δ, которая с заранее заданной вероятностью γ, близкой к единице, гарантировала бы выполнение неравенства $\left|\hat{\theta}-\theta\right|<\delta$? То есть нельзя ли указать такой интервал $(\hat{\theta}-\delta;\hat{\theta}+\delta)$, который с заранее заданной вероятностью γ, близкой к единице, «накрывал» бы неизвестное истинное значение θ? Здесь заранее заданная вероятность γ называется ***доверительной вероятностью (или надежностью)***, а сам интервал $(\hat{\theta}-\delta;\hat{\theta}+\delta)$ – ***доверительным интервалом*** (или ***интервальной оценкой***) для параметра θ. На практике обычно задают γ из набора значений γ = 0.9; 0.95; 0.99; 0.995 и т.д.

Доверительный интервал ***имеет случайные границы*** (так как $\hat{\theta}$ – случайная величина). Величина δ – точность оценки. Она, как правило, зависит от выборочных данных и поэтому тоже носит случайный характер.

Таким образом, доверительный интервал определяется следующим соотношением:

$$\boxed{P(\hat{\theta}-\delta<\theta<\hat{\theta}+\delta)=\gamma}. \qquad (9.1.1)$$

Утверждение $\hat{\theta}-\delta<\theta<\hat{\theta}+\delta$ с вероятностью γ является практически истинным, в то время как событие $(\theta=\hat{\theta})$, как правило, имеет вероятность, равную нулю.

В общем случае интервальной оценкой (доверительным интервалом) для параметра θ называется интервал $\left(\hat{\theta}_H,\hat{\theta}_B\right)$, нижняя $\hat{\theta}_H$ и верхняя $\hat{\theta}_B$ границы которого определяются соотношением

$$\boxed{P(\hat{\theta}_H<\theta<\hat{\theta}_B)=\gamma}. \qquad (9.1.2)$$

Очевидно, что для интервальной оценки (9.1.1)

$$\hat{\theta}_H=\hat{\theta}-\delta\,,\ \ \hat{\theta}_B=\hat{\theta}+\delta\,.$$

Следует отметить, что чем меньше доверительный интервал, тем лучше оценка параметра. Величина этого интервала, как показано в математической статистике, зависит от объема выборки *n* и задаваемой доверительной вероятности γ (надежности интервальной оценки). Величина $\alpha=1-\gamma$ называется ***уровнем значимости*** (уровень значимости – это минимальная вероятность, начиная с которой событие считается практически невозможным).

Общая теория построения интервальных оценок заключается в определении *случайной величины, зависящей от оцениваемого параметра*. Зная распределение этой случайной величины,

находят соответствующие доверительные границы и сам доверительный интервал с требуемой точностью. Посмотрим, как эта идея реализуется для различных параметров.

9.2. Интервальные оценки математического ожидания нормального распределения

Рассмотрим построение интервальной оценки для двух случаев:

- ***Случай 1.*** Известно среднеквадратическое отклонение σ или дисперсия нормального распределения $D(X)=\sigma^2$.
- ***Случай 2.*** Параметр σ или $D(X)=\sigma^2$ неизвестны.

Случай 1. Пусть генеральная совокупность X распределена по нормальному закону $N(a,\sigma)$, причем параметр σ известен, а параметр a требуется оценить с надежностью γ. Можно показать (см. свойства выборочного среднего $\overline{X}_в$), что случайная величина $\dfrac{(\overline{X}_в - a)\sqrt{n}}{\sigma}$ распределена по закону $N(0,1)$, т.е. имеет нормальное распределение с нулевым математическим ожиданием и единичным среднеквадратическим отклонением. На рис. 9.1 изображен график функции плотности этой случайной величины, т.е. кривая $p(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Выберем число x_γ так, чтобы заштрихованная площадь под кривой $p(x)$ была равна γ, т.е.

$$\boxed{P\left(-x_\gamma < \frac{(\overline{X}_в - a)\sqrt{n}}{\sigma} < x_\gamma\right) = \gamma}. \qquad (9.2.1)$$

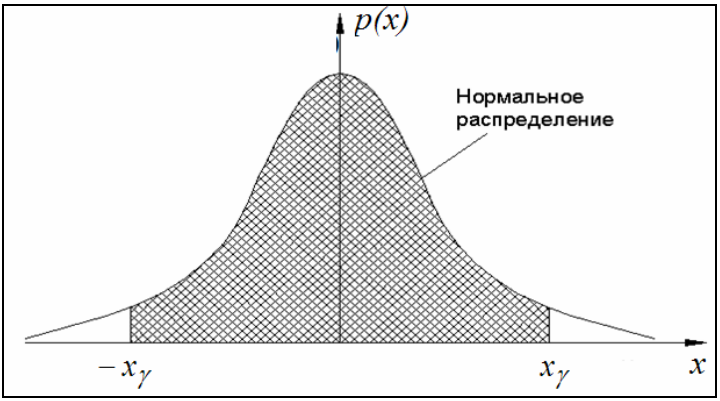


Рис. 9.1. К построению доверительных интервалов

Это значение легко находится с использованием интегральной функции Лапласа $\Phi(x)=\frac{1}{2\pi}\int\limits_0^x e^{-\frac{t^2}{2}}dt$. Действительно,

$$P(-x_\gamma < N(0,1) < x_\gamma) = \Phi(x_\gamma) - \Phi(-x_\gamma) = 2\Phi(x_\gamma) = \gamma. \qquad (9.2.2)$$

Значение x_γ, удовлетворяющее нелинейному уравнению

$$\boxed{\Phi(x_\gamma) = \frac{\gamma}{2}}, \qquad (9.2.3)$$

находится по табл. П1. Вычисляем значение $\gamma/2$ (например, 0.95/2 = 0.475) и для этого значения функции $\Phi(x)$ в столбце x находим соответствующее значение x_γ (например, для 0.95/2 = = 0.475 $x_\gamma = 1.96$). Так как $\sigma > 0$, то события $-x_\gamma < \frac{\left(\overline{X}_в - a\right)\sqrt{n}}{\sigma} < x_\gamma$ и $\overline{X}_в - \frac{x_\gamma\sigma}{\sqrt{n}} < a < \overline{X}_в + \frac{x_\gamma\sigma}{\sqrt{n}}$ эквивалентны, а значит, их вероятности равны

$$P\left(\overline{X}_в - \frac{x_\gamma\sigma}{\sqrt{n}} < a < \overline{X}_в + \frac{x_\gamma\sigma}{\sqrt{n}}\right) = \gamma.$$

Таким образом, для математического ожидания a построен доверительный интервал (интервальная оценка)

$$\boxed{\left(\overline{X}_в - \frac{x_\gamma \sigma}{\sqrt{n}},\ \overline{X}_в + \frac{x_\gamma \sigma}{\sqrt{n}}\right)}, \qquad (9.2.4)$$

левая граница которого $\overline{X}_в - \frac{x_\gamma \sigma}{\sqrt{n}}$, правая – $\overline{X}_в + \frac{x_\gamma \sigma}{\sqrt{n}}$, а точность – $\delta = \frac{x_\gamma \sigma}{\sqrt{n}}$. Центр этого интервала находится в точке с координатой $\overline{X}_в$, а длина интервала $2\frac{x_\gamma \sigma}{\sqrt{n}}$. Если объем выборки неограниченно возрастает, то интервал стягивается в одну точку $\overline{X}_в$, которая является состоятельной и несмещенной оценкой для параметра a.

Для вычисления корня x_γ уравнения (9.2.3) можно использовать функцию Excel НОРМСТОБР следующим образом:

$$\boxed{x_\gamma = \text{НОРМСТОБР}((\gamma+1)/2)}. \qquad (9.2.5)$$

Вычисление величины $\delta = x_\gamma \sigma / \sqrt{n}$ осуществляется с помощью функции ДОВЕРИТ:

$$\boxed{\delta = x_\gamma \sigma / \sqrt{n} = \text{ДОВЕРИТ}(1-\gamma; \sigma; n)}, \qquad (9.2.6)$$

где σ – известное среднеквадратичное отклонение; n – объем выборки. После вычисления точности δ интервальную оценку (9.2.4) можно записать в виде:

$$\boxed{\left(\overline{X}_в - \delta, \overline{X}_в + \delta\right)}. \qquad (9.2.7)$$

Видно, что границы интервальной оценки (9.2.4) случайны, так как зависят от случайной величины выборочного среднего

$\overline{X}_в$. Для вычисления интервальной оценки по выборке $x_1, x_2, ..., x_n$ нужно вместо случайной величины $\overline{X}_в$ использовать ее выборочное (уже не случайное) значение $\overline{x}_в$ (см. формулу (8.2.2)), и тогда интервальная оценка принимает вид:

$$\boxed{\left(\overline{x}_в - \frac{x_\gamma \sigma}{\sqrt{n}},\ \overline{x}_в + \frac{x_\gamma \sigma}{\sqrt{n}}\right)}. \qquad (9.2.8)$$

Пример 9.2.1. По выборке объема $n = 9$ найдено среднее значение $\overline{x}_в = 1.5$. Считая, что генеральная совокупность распределена по нормальному закону с $\sigma = 2$, определить интервальную оценку для математического ожидания с надежностью $\gamma = 0.95$.

Решение. Используя табл. П1, находим, что $\Phi(x_\gamma) = \frac{0.95}{2} = 0.475$ при $x_\gamma = 1.96$. Это же значение $x_\gamma = 1.96$ можно получить, используя функцию (9.2.5) с параметрами: НОРМСТОБР (1.95/2). Далее $\delta = 1.96 \cdot \frac{2}{\sqrt{9}} = 1.31$ и доверительный интервал (9.2.4) имеет границы $(\overline{X}_в - 1.31,\ \overline{X}_в + 1.31)$. Таким образом, с вероятностью 0.95 можно быть уверенным в том, что интервал

$$(\overline{X}_в - 1.31,\ \overline{X}_в + 1.31)$$

накроет параметр *a*, или, другими словами, с вероятностью 0.95 значение $\overline{X}_в$ дает значение параметра *a* с точностью δ = 1.31. Вычислим выборочную интервальную оценку (9.2.8), подставив в формулу значение $\overline{x}_в = 1.5$:

$$(1.5 - 1.31,\quad 1.5 + 1.31) = (0.19,\ 2.81). \quad ☻$$

Случай 2. Определим теперь интервальную оценку для неизвестного параметра *a* нормально распределенной генеральной совокупности *X* в том случае, когда *генеральная дисперсия*

$D(X)$ *неизвестна*, т.е. построим доверительный интервал для параметра *a*, если параметр σ неизвестен. В отличие от предыдущего случая, вместо случайной величины $\frac{(\overline{X}_в - a)\sqrt{n}}{\sigma}$, распределенной по закону $N(0,1)$, рассмотрим случайную величину $\frac{(\overline{X}_в - a)\sqrt{n-1}}{\sqrt{D_в}}$. Можно показать, что эта величина имеет распределение Стьюдента T_{n-1} с $n-1$ степенями свободы. Случайную величину, имеющую распределение T_{n-1}, можно представить в виде:

$$T_{n-1} = \frac{N(0,1)}{\sqrt{\chi^2_{n-1}}}\sqrt{n-1},$$

где $\chi^2_{n-1} = N_1^2 + ... + N_{n-1}^2$ – так называемое распределение хи-квадрат (распределение Пирсона) с $n-1$ степенями свободы; N_i – случайная величина, имеющая распределение $N(0, 1)$.

При заданном значении γ, пользуясь табл. П2, вычислим значение $t(\gamma, n-1)$ из условия

$$P\left(-t(\gamma, n-1) < \frac{(\overline{X}_в - a)\sqrt{n-1}}{\sqrt{D_в}} < t(\gamma, n-1)\right) = \gamma, \qquad (9.2.9)$$

где γ – надежность интервальной оценки. Значение $t(\gamma, n-1)$ можно также вычислить, используя функцию Excel СТЬЮДРАСПОБР в виде:

$$\boxed{t(\gamma, n-1) = \text{СТЬЮДРАСПОБР}(1-\gamma; n-1)}. \qquad (9.2.10)$$

Замена случайной величины $\frac{(\overline{X}_в - a)\sqrt{n}}{\sigma}$ на случайную величину $\frac{(\overline{X}_в - a)\sqrt{n-1}}{\sqrt{D_в}}$ вызвана тем, что закон распределения последней случайной величины известен и в ее запись не входит

неизвестный в данном случае параметр σ. Из условия (9.2.9) получаем:

$$P\left(\overline{X}_в - \frac{t(\gamma, n-1)\sqrt{D_в}}{\sqrt{n-1}} < a < \overline{X}_в - \frac{t(\gamma, n-1)\sqrt{D_в}}{\sqrt{n-1}}\right) = \gamma.$$

Таким образом, интервальная оценка надежности γ для неизвестной математического ожидания a имеет вид:

$$\boxed{\left(\overline{X}_в - \frac{t(\gamma, n-1)\sqrt{D_в}}{\sqrt{n-1}}, \overline{X}_в + \frac{t(\gamma, n-1)\sqrt{D_в}}{\sqrt{n-1}}\right)}. \qquad (9.2.11)$$

Если известна исправленная дисперсия S^2, то интервальная оценка имеет вид:

$$\boxed{\left(\overline{X}_в - \frac{t(\gamma, n-1)S}{\sqrt{n}}, \overline{X}_в + \frac{t(\gamma, n-1)S}{\sqrt{n}}\right)}, \qquad (9.2.12)$$

а точность интервальной оценки можно определить соотношениями:

$$\boxed{\delta = \frac{t(\gamma, n-1)\sqrt{D_в}}{\sqrt{n-1}}} \text{ или } \boxed{\delta = \frac{t(\gamma, n-1)}{\sqrt{n}} S}. \qquad (9.2.13)$$

Как и в предыдущем случае, центр интервала находится в точке $\overline{X}_в$, но длина интервала $2\delta = 2\frac{t(\gamma, n-1)}{\sqrt{n}} S$ является случайной величиной, принимающей тем меньшие значения, чем больше значение n. Это объясняется тем, что *наличие большей информации* $x_1, ..., x_n$ *о генеральной совокупности X позволяет сузить интервал.*

Для вычисления интервальных оценок по выборке $x_1, x_2, ..., x_n$ нужно вместо случайных величин $\overline{X}_в, D_в, S$ использовать их выборочные (уже не случайные) значения $\overline{x}_в, d_в, s$ и тогда интервальные оценки принимают вид:

$$\boxed{\left(\overline{x}_в - \frac{t(\gamma, n-1)\sqrt{d_в}}{\sqrt{n-1}},\ \overline{x}_в + \frac{t(\gamma, n-1)\sqrt{d_в}}{\sqrt{n-1}}\right)}; \qquad (9.2.14)$$

$$\boxed{\left(\overline{x}_в - \frac{t(\gamma, n-1)s}{\sqrt{n}},\ \overline{x}_в + \frac{t(\gamma, n-1)s}{\sqrt{n}}\right)}. \qquad (9.2.15)$$

Пример 9.2.2. По выборке объема $n = 9$ из нормально распределенной генеральной совокупности найдены выборочные значения $\overline{x}_в = 1.5$ и $s = 2$. Построить интервальную оценку для математического ожидания с надежностью $\gamma = 0.95$.

Решение. Пользуясь табл. П2, находим величину $t(0.95, 9-1) = t(0.95, 8) = 2.306 \approx 2.31$. Это же значение $t(0.95, 8)$ можно получить используя функцию (9.2.10) с параметрами: СТЬЮДРАСПОБР $(0.05, 8)$. Далее определяем точность δ: $\delta = \frac{t(0.95, 9)S}{\sqrt{n}} = \frac{2.31}{3}S = 0.77S$, а интервальная оценка имеет границы $\left(\overline{X}_в - 0.77 \cdot S,\ \overline{X}_в + 0.77 \cdot S\right)$, которые зависят от двух случайных величин: $\overline{X}_в$ и S. Подставляя вместо S ее вычисленное значение $s = 2$, получаем интервал

$$\left(\overline{X}_в - 1.54,\ \overline{X}_в + 1.54\right).$$

Сравнивая эту оценку с интервальной оценкой примера 9.2.1, видим, что замена неизвестной величины σ вычисляемой величиной s приводит к уменьшению точности интервальной оценки и увеличению длины доверительного интервала. Подставив вместо случайной величины $\overline{X}_в$ ее конкретное значение $\overline{x}_в = 1.5$, получаем конкретное значение границ интервальной оценки для параметра a нормального распределения:

$$(1.5 - 1.54, 1.5 + 1.54) = (-0.04, 3.04). \quad ☻$$

9.3. Интервальная оценка для среднеквадратического отклонения нормального распределения

Пусть количественный признак X генеральной совокупности распределен нормально. Требуется по вычисленному выборочному среднеквадратическому отклонению s найти доверительный интервал для параметра σ с заданной доверительной вероятностью (надежностью) γ, т.е.

$$P\left(s-\delta<\sigma<s-\delta\right)=\gamma. \qquad (9.3.1)$$

Введя величину

$$\boxed{q_\gamma = \delta/s}, \qquad (9.3.2)$$

неравенство $s-\delta<\sigma<s-\delta$ можно представить как

$$s(1-q_\gamma)<\sigma<s(1+q_\gamma).$$

Учитывая, что $\sigma>0$, получаем окончательную запись – доверительный интервал для параметра σ с заданной доверительной вероятностью (надежностью) γ:

$$\boxed{\left(s\cdot\max(0,(1-q_\gamma)),\, s(1+q_\gamma)\right)}. \qquad (9.3.3)$$

Параметр q_γ зависит от надежности γ и объема выборки n, и его значения приведены в табл. П3.

Пример 9.3.1. Количественный признак X генеральной совокупности распределен нормально. По выборке объема $n = 50$ найдено «исправленное» среднеквадратическое отклонение $s = 0.8$. Найти доверительный интервал для среднеквадратического отклонения σ с надежностью $\gamma = 0.99$.

Решение. По табл. П3 для $\gamma = 0.99$, $n = 50$ находим $q_\gamma = 0.30$. Тогда искомый доверительный интервал имеет вид: $\left(0.8\cdot(1-0.30),(1+0.30)\cdot 0.8\right)$, или $\left(0.56, 1.04\right)$. ☻

Пример 9.3.2. Количественный признак X генеральной совокупности распределен нормально. По выборке объема $n = 5$ найдено «исправленное» среднее квадратическое отклонение $s = 0.2$. Найти доверительный интервал для среднеквадратического отклонения σ с доверительной вероятностью $\gamma = 0.95$

Решение. По табл. П3 найдем $q_\gamma = 1.37$. Так как $q_\gamma > 1$, то доверительный интервал имеет вид (см. (9.3.3)):

$\left(0.2 \cdot \max(0, -0.37)), 0.2(1+1.37)\right)$, или $\left(0,\ 0.474\right)$. ☻

Пример 9.3.3. Доверительный интервал для оценки генерального среднеквадратического отклонения σ нормально распределенной величины X имеет границы (α, 1.3). Какое значение имеет параметр α, если «исправленное» выборочное среднеквадратическое отклонение ($s = 0.7$)?

Решение. Если границы доверительного интервала для параметра σ имеют вид: (α, β), то $s = \dfrac{\alpha+\beta}{2}$. Следовательно, в нашем примере $s = \dfrac{\alpha+1.3}{2} = 0.7$. Решая это уравнение, находим $\alpha = 0.1$. ☻

Пример 9.3.4. Количественный признак X генеральной совокупности распределен по нормальному закону. Каким должен быть объем выборки n, чтобы с надежность $\gamma = 0.95$ доверительный интервал для среднеквадратического отклонения σ имел границы (0.2, 1.2)?

Решение. Определим значение s по формуле $s = \dfrac{\alpha+\beta}{2}$. В нашем примере $s = \dfrac{0.2+1.2}{2} = 0.7$. Из (9.3.3) следует, что границы доверительного интервала для среднеквадратического отклонения σ должны удовлетворять соотношениям:

$$\alpha = s(1-q_\gamma); \quad \beta = s(1+q_\gamma).$$

Найдем значение q_γ, например, из уравнения $1.2 = 0.7 \cdot (1+q_\gamma)$.

Очевидно, что q_γ = (1.2 / 0.7) – 1 = 0.71. По табл. П3 для столбца с надежностью γ = 0.95 и строки $q_\gamma = 0.71$ находим объем выборки $n = 9$. ☻

Замечание 9.3.1. Если в предыдущем примере увеличить надежность и положить γ = 0.99, то для q_γ = 0.71 по табл. П3 находим $n = 16$. Можно сделать вывод, что при заданной точности интервальной оценки (δ = 0.5) увеличить ее надежность (γ_1 = 0.95, γ_2 = 0.99) можно, увеличивая объем выборки (n_1 = 9, n_2 = 16). ●

Пример 9.3.5. Найти надежность интервальной оценки 0.632 < σ < 0.968, если признак X подчиняется нормальному распределению и по результатам выборочного обследования с объемом n = 50 выборочное значение s = 0.8.

Решение. Так как точность интервальной оценки

$$\delta = \frac{0.968 - 0.632}{2} = 0.168,$$

то по формуле (9.3.2) вычисляем: $q_\gamma = \dfrac{\delta}{s} = \dfrac{0.168}{0.8} = 0.21.$ По табл. П3 для строки, соответствующей n = 50 и значению $q_\gamma = 0.21$, находим столбец, соответствующий γ = 0.95. ☻

9.4. Интервальные оценки для дисперсии нормального распределения

Как и при построении интервальных оценок для математического ожидания, в данном случае также необходимо определить случайную величину, распределение которой было известно и включало оцениваемый параметр σ^2. Не приводя доказательства, отметим, что такой величиной может быть случайная величина $\dfrac{nD_в}{\sigma^2}$, распределенная по закону χ^2_{n-1} с $(n-1)$ степенями свободы.

Замечание 9.3.1. Распределение χ^2 (распределение Пирсона). Пусть $N_1, ..., N_n$ – независимые нормально распределенные случайные величины с параметрами (0, 1). Распределение случайной величины

$$\chi_n^2 = N_1^2 + N_2^2 + N_3^2 + ... + N_n^2$$

называется *распределением* χ_n^2 *с n степенями свободы, а сама величина* χ_n^2 *– случайной величиной* χ^2 *с n степенями свободы.*

Заметим, что количество степеней свободы n является единственным параметром распределения и значения χ^2 неотрицательны, т.е. $P(\chi_n^2 < 0) = 0$. Распределение χ_n^2 с n степенями свободы имеет следующие числовые характеристики: $M[\chi_n^2] = n$; $D[\chi_n^2] = 2n$. При числе степеней свободы $n > 30$ χ_n^2-распределение хорошо аппроксимируется нормальным распределением $N(a, \sigma)$ с параметрами $a = n$, $\sigma = \sqrt{2n}$.

На рис. 9.2 показаны плотности распределения $p(x)$ случайной величины χ_n^2 при $n = 2$, $n = 6$ и $n = 20$. Видно, что при увеличении n плотность $p(x)$ «приближается» к плотности нормального распределения ●

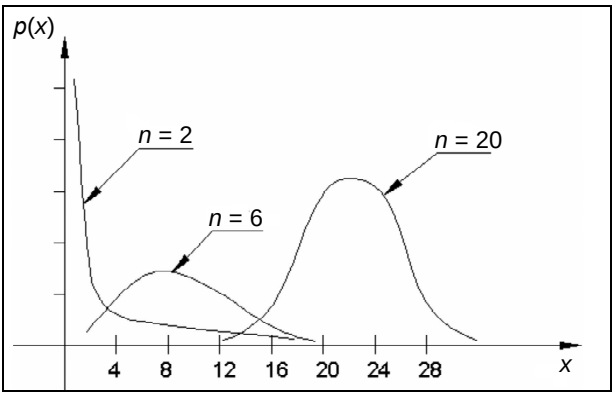


Рис. 9.2. Плотность распределения χ^2

Зададим надежность интервальной оценки γ, тогда имеет место следующее равенство:

$$P\left(\chi^2_{лев,\gamma} < \frac{nD_в}{\sigma^2} < \chi^2_{пр,\gamma}\right) = \gamma, \qquad (9.4.1)$$

в котором границами являются: $\chi^2_{лев,\gamma}$ – квантиль χ^2_{n-1}-распределения уровня $\alpha/2$, $\chi^2_{пр,\gamma}$ – уровня $1-\alpha/2$, где $\alpha = 1-\gamma$. Заметим, что квантилем уровня q некоторого распределения $p(x)$ называется такая величина x_q, для которой

$$\int_{-\infty}^{x_q} p(x)dx = q. \qquad (9.4.2)$$

Указанные квантили могут быть вычислены с использованием функции Excel ХИ2ОБР следующими выражениями:

$$\chi^2_{лев,\gamma} = \text{ХИ2ОБР}(1-\alpha/2\,;\ n-1); \qquad (9.4.3)$$

$$\chi^2_{пр,\gamma} = \text{ХИ2ОБР}(\alpha/2\,;\ n-1). \qquad (9.4.4)$$

Выполнив несложные преобразования неравенства $\chi^2_{лев,\gamma} < \frac{nD_в}{\sigma^2} < \chi^2_{пр,\gamma}$, получаем интервальную оценку для дисперсии σ^2

$$\boxed{\left(\frac{nD_в}{\chi^2_{пр,\gamma}}, \frac{nD_в}{\chi^2_{лев,\gamma}}\right)} \qquad (9.4.5)$$

надежности γ.

Так как $D_в = (n-1)S^2/n$, то $nD_в = (n-1)S^2$ и интервал

$$\boxed{\left(\frac{n-1}{\chi^2_{пр,\gamma}}S^2, \frac{n-1}{\chi^2_{лев,\gamma}}S^2\right)} \qquad (9.4.6)$$

также является интервальной оценкой для дисперсии σ^2 надежности γ.

Заметим, что границы интервалов (9.4.4), (9.4.5) являются случайными величинами (почему?), и с вероятностью γ можно утверждать, что интервалы (9.4.5), (9.4.6) накроют неизвестную дисперсию σ^2.

Пример 4.3. По выборке объема $n = 19$ из нормально распределенной генеральной совокупности вычислено значение дисперсии выборки $d_в = 1.5$. Построить интервальную оценку для параметра σ^2 надежности $\gamma = 0.96$.

Решение. Значения $\chi^2_{лев,\gamma}$, $\chi^2_{пр,\gamma}$ находим из условий:

$$P\left(\chi^2_{19} < \chi^2_{лев,\gamma}\right) = 0.02;\ P\left(\chi^2_{19} < \chi^2_{пр,\gamma}\right) = 0.98.$$

Эти условия означают, что $\chi^2_{лев,\gamma}$ есть квантиль χ^2-распределения с 18 степенями свободы уровня 0.02, а $\chi^2_{пр,\gamma}$ – квантиль уровня 0.98. По табл. П4 квантилей χ^2-распределения находим $\chi^2_{лев,\gamma} = 7.90$; $\chi^2_{пр,\gamma} = 32.34$. Эти же значения квантилей можно получить в Excel, обращаясь к функциям:

$\chi^2_{лев,0.96}$ = ХИ2ОБР(0.98; 18) = 7.90;

$\chi^2_{пр,0.96}$ = ХИ2ОБР(0.02; 18) = 32.34.

Интервальная оценка (9.4.5) принимает вид $(0.59D_в,\ 2.33D_в)$. Подставляя вычисленное значение $d_в = 1.5$ случайной величины $D_в$, получаем $0.89 < \sigma^2 < 3.488$. ☻

Вопросы и задачи для самопроверки

1. Когда возникает необходимость построения интервальной оценки?
2. Что называется интервальной оценкой?
3. Что такое надежность интервальной оценки?
4. Отличаются ли два понятия «надежность интервальной оценки» и «доверительная вероятность»?

5. Даны две доверительные вероятности $\gamma_1 = 0.95$ и $\gamma_2 = 0.99$. Для какой вероятности длина доверительного интервала будет больше?
6. Какой вид имеет интервальная оценка для параметра a (математического ожидания нормального распределения) в случае, когда известен параметр σ?
7. Что называется точностью интервальной оценки?
8. В каком случае точность интервальной оценки для параметра a выше: когда известен или неизвестен параметр σ?
9. По выборке объема $n = 16$ найдено среднее значение $\overline{x}_в = 3.0$. Считая, что генеральная совокупность распределена по нормальному закону с $\sigma = 4$, построить три интервальных оценки для математического ожидания с надежностью $\gamma = 0.95, \gamma = 0.98, \gamma = 0.99$. Установить, как зависит длина доверительного интервала от надежности интервальной оценки γ.
10. По выборке объема $n = 16$ найдены выборочное среднее значение $\overline{x}_в = 3.0$ и выборочная дисперсия $d_в = 4$. Считая, что генеральная совокупность распределена по нормальному закону, построить три интервальных оценки для математического ожидания с надежностью $\gamma = 0.95, \gamma = 0.98, \gamma = 0.99$. Установить, как зависит длина доверительного интервала от надежности интервальной оценки γ. Сравнить длины построенных интервальных оценок с длинами интервалов задания 9.
11. Доверительный интервал для математического ожидания нормально распределенной случайной величины имеет вид $(\mu, 3.2)$. Каким должно быть значение μ, если $\overline{x}_в = 2.1$?

    Ответ: $\mu = 1$.
12. На контрольных испытаниях 16 осветительных ламп были определены: средняя продолжительность работы лампы $\overline{x}_в = 3000$ часов и $s = 20$ часов. Считая, что срок службы каждой лампы подчиняется нормальному распределению, определить с надежностью 0.9 доверительный интервал для математического ожидания.

Ответ: доверительный интервал имеет вид: (2991.235, 3008.765).

13. Вычислить с надежностью 0.98 интервальную оценку для среднеквадратического отклонения σ, если по выборке объемом $n = 17$ вычислена оценка $s^2 = 25$.
Ответ: интервальная оценка имеет вид (2.18, 7.82).

Тема 10. Проверка статистических гипотез

10.1. Понятие статистической гипотезы

Под ***статистической гипотезой*** понимается всякое высказывание или предположение о генеральной совокупности (случайной величине X), проверяемое по выборочной совокупности (по результатам наблюдений).

Статистическую гипотезу, содержащую утверждение о параметрах генеральной совокупности, называют ***параметрической***. Гипотезу, содержащую утверждение обо всем распределении случайной величины, называют ***непараметрической.***

Рассмотрим один пример параметрической статистической гипотезы.

Пример 10.1.1. Результаты многолетних статистических исследований показали, что для населения некоторого региона вероятность предрасположения к данному заболеванию R равна $p_0 = 0.1$. После строительства в этом регионе химического предприятия была проведена выборочная проверка населения. Из 1000 обследованных у 120 человек были обнаружены признаки заболевания R.

Можно ли утверждать, что а) полученные данные не противоречат предположению, что строительство не повлияло на уровень заболевания R; б) изменение экологической обстановки после строительства комбината повлияло на распространение заболевания R? ☻

Приведенный пример является типичной задачей проверки статистической гипотезы. В примере высказывание формулируется в терминах вероятности p_0 события A = {наличие у чело-

века заболевания $R\}$. Не располагая сведениями обо всей генеральной совокупности, высказанную гипотезу сопоставляют по определенным правилам с выборочными данными и делают вывод о том, можно принять гипотезу или нет. Эта процедура сопоставления называется *проверкой гипотезы*.

10.2. Основные этапы проверки гипотезы

Рассмотрим этапы проверки гипотезы и используемые при этом понятия.

Этап 1. Располагая выборочными данными и руководствуясь конкретными условиями рассматриваемой задачи, формулируют гипотезу H_0, которую называют *основной* или *нулевой*, и гипотезу H_1, *конкурирующую* с гипотезой H_0. Гипотезу H_1 называют также *альтернативной*.

Термин «конкурирующая» означает, что являются взаимоисключающими следующие два события:

- по выборке принимается решение о справедливости для генеральной совокупности гипотезы H_0;
- по выборке принимается решение о справедливости для генеральной совокупности гипотезы H_1.

Вернемся к примеру 10.1.1. Обозначим через A событие, состоящее в том, что случайно выбранный человек в данном регионе предрасположен к заболеванию R. До строительства химического предприятия вероятность события A была равна 0.1. В качестве гипотезы H_0 рассмотрим гипотезу о том, что после строительства химического предприятия вероятность события A не изменилась. Таким образом, если p_1 – вероятность события A после строительства предприятия, то в качестве нулевой (основной) гипотезы принимается

$$\boxed{H_0 : p_1 = p_0}.$$

Учитывая, что а) строительство комбината вряд ли улучшило экологическую обстановку в регионе; б) при выборке из 1000 человек у 120 человек обнаружено заболевание R, что соответ-

ствует относительной частоте $p^* = 120/1000 = 0.12 > 0.1$, в качестве альтернативной гипотезы примем

$$\boxed{H_1 : p_1 > p_0}.$$

Этап 2 . Задается вероятность α, которую называют *уровнем значимости*. Эта вероятность имеет следующий смысл.
Решение о том, можно ли считать высказывание H_0 справедливым для генеральной совокупности, принимается по выборочным данным, т.е. по ограниченному объему информации. Следовательно, это решение может быть ошибочным. При этом может иметь место ошибка двух родов:

- ***ошибка первого рода*** совершается при отклонении гипотезы H_0 (т.е. принимается альтернативная H_1), тогда как на самом деле гипотеза H_0 верна; вероятность такой ошибки обозначим $P(H_1 / H_0)$;
- ***ошибка второго рода*** совершается при принятии гипотезы H_0, тогда как на самом деле высказывание H_0 неверно и следовало бы принять гипотезу H_1; вероятность ошибки второго рода обозначим как $P(H_0 / H_1)$.

Тогда уровень значимости α определяет вероятность ошибки первого рода, т.е.

$$\boxed{\alpha = P(H_1 / H_0)}. \qquad (10.2.1)$$

Поэтому вероятность α задается малым числом, поскольку это вероятность ошибочного высказывания. При этом обычно используются стандартные значения: 0.05; 0.01; 0.005. Например, $\alpha = 0.05$ означает следующее: если гипотезу H_0 проверять по каждой из 100 выборок одинакового объема, то в среднем в 5 случаях из 100 совершим ошибку первого рода.

Обратим внимание на то, что в результате проверки гипотезы H_0 могут быть приняты *правильные решения* двух следующих видов:

- принимается гипотеза H_0 тогда, когда она верна (т.е. H_0 имеет место в генеральной совокупности); вероятность этого решения равна $P(H_0 / H_0) = 1 - \alpha$ (почему?);

- не принимается гипотеза H_0 (т.е. принимается гипотеза H_1) тогда, когда и на самом деле она неверна (т.е. справедлива гипотеза H_1), вероятность этого решения равна $P(H_1/H_1)=1-\beta$, где $\beta=P(H_0/H_1)$ – вероятность ошибки второго рода (почему?).

Этап 3. Определяют величину K такую, что а) ее значения зависят от выборочных данных $x_1, x_2, ..., x_n$, т.е. $K=K(x_1, x_2, ..., x_n)$; б) будучи величиной случайной (в силу случайности выборки $x_1, ..., x_n$), величина K подчиняется при выполнении гипотезы H_0 некоторому известному закону распределения; в) ее значения позволяют судить о расхождении гипотезы H_0 с выборочными данными. Величину K называют *критерием*.

Обратимся вновь к примеру 10.1.1. Пусть S_{1000} – количество обследуемых, предрасположенных к заболеванию R в выборке из 1000 человек. Если гипотеза H_0 верна, т.е. $p_1 = p_0 = 0.1$, то случайная величина S_{1000} распределена по биномиальному закону и ее числовые характеристики равны $M(S_{1000})=1000\cdot 0.1=100$, $D(S_{1000})=1000\cdot 0.1\cdot(1-0.1)=90$ (см. формулы (4.3.3), (4.4.3)). С другой стороны, ее распределение близко к нормальному. Поэтому случайная величина

$$K=\frac{S_{1000}-100}{\sqrt{90}}=\frac{S_{1000}-100}{9.487} \qquad (10.2.2)$$

распределена по закону, близкому к нормальному $N(0,1)$.

Заметим, что если вероятность события A возросла после строительства химического комбината, то случайная величина K преимущественно будет принимать положительные значения (почему?), и это может трактоваться в пользу принятия гипотезы H_1. Видно, что величина (10.2.2) удовлетворяет требованиям (а), (б), (в) и может быть принята при проверке гипотезы $H_0: p_1 = p_0$ при альтернативной $H_1: p_1 > p_0$.

Этап 4. В области всевозможных значений критерия K выделяют подобласть ω, называемую *критической областью*. Значения критерия, попавшие в критическую область, свидетельствуют о существенном расхождении выборки с гипотезой H_0. Поэтому руководствуются следующим правилом: если вычисленное по выборке значение критерия попадает в критическую область ω, то гипотеза H_0 отвергается и принимается альтернативная H_1. При этом следует помнить, что такое решение может быть ошибочным – на самом деле гипотеза H_0 может быть справедливой. Таким образом, ориентируясь на критическую область, можно совершить ошибку первого рода, вероятность которой задана заранее и равна α. Отсюда вытекает следующее требование к критической области ω:

Вероятность принятия критерием K значения из критической области ω при справедливости гипотезы H_0 должна быть равна вероятности ошибки первого рода α, т.е.

$$\boxed{P(K \in \omega) = \alpha}. \qquad (10.2.3)$$

Однако критическая область определяется этим равенством неоднозначно. Пусть $p_K(x)$ является плотностью распределения критерия K. Тогда нетрудно увидеть, что на оси X существует бесчисленное множество интервалов таких, что площади построенных на них криволинейных трапеций, ограниченных сверху кривой $p_K(x)$, равны α. Поэтому, кроме требования формулы (10.2.3), выдвигается следующее: критическая область ω должна быть расположена так, чтобы при заданной вероятности α – ошибки первого рода вероятность β – ошибки второго рода была минимальной.

Обычно этому требованию удовлетворяют три случая расположения критической области (в зависимости от вида нулевой и альтернативной гипотез, формы и распределения критерия K):

- ***правосторонняя критическая область*** (рис. 10.1*а*), состоящая из интервала $(x_{пр,\alpha}, +\infty)$, где точка $x_{пр,\alpha}$ определяется из условия

$$\boxed{P(K > x_{пр,\alpha}) = \alpha} \qquad (10.2.4)$$

и называется *правосторонней критической точкой*;

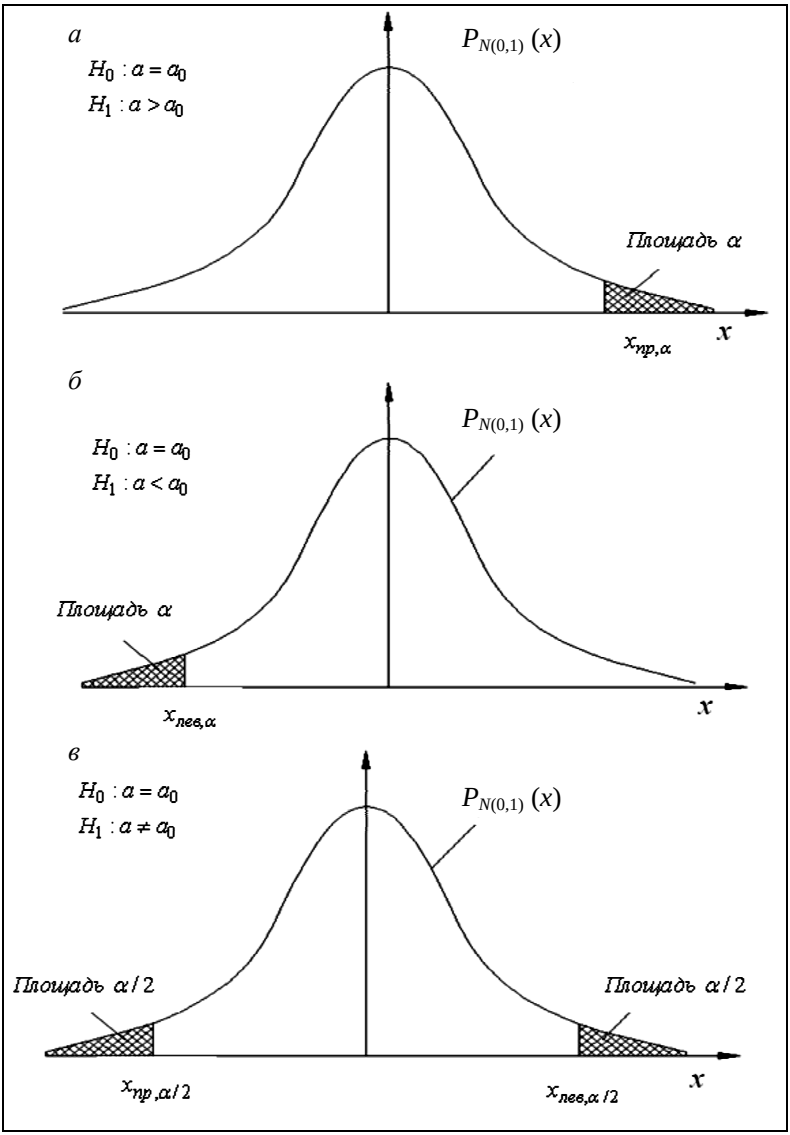


Рис. 10.1. Три вида критических областей при проверке статистических гипотез

- ***левосторонняя критическая область*** (см. рис. 10.1*б*) состоит из интервала $(-\infty, x_{лев,\alpha})$, где $x_{лев,\alpha}$ определяется из условия

$$\boxed{P(K < x_{лев,\alpha}) = \alpha} \qquad (10.2.5)$$

и называется *левосторонней критической точкой*;

- ***двусторонняя критическая область*** (см. рис. 10.1*в*), состоящая из двух интервалов: $(-\infty, x_{лев,\alpha/2})$, $(x_{пр,\alpha/2}, +\infty)$, где точки $x_{лев,\alpha/2}$, $x_{пр,\alpha/2}$ определяются из условий

$$\boxed{P(K < x_{лев,\alpha/2}) = \alpha/2;\ P(K > x_{пр,\alpha/2}) = \alpha/2}. \qquad (10.2.6)$$

Вернемся к нашему примеру 10.1.1. Так как альтернативная гипотеза имеет вид $H_1 : p_1 > p_0$, то принимается правосторонняя критическая область (см. рис. 10.1*а*). При справедливости гипотезы H_0 критерий K, определяемый выражением (10.2.2), имеет нормальное распределение $N(0, 1)$, и следовательно, по таблице функции Лапласа $\Phi(x)$ (табл. П1) необходимо найти такое $x_{пр,\alpha}$, что $\Phi(x_{пр,\alpha}) = (1-\alpha) - 0.5 = 0.5 - \alpha$. Для нашего примера примем $\alpha = 0.005$. Тогда по табл. П1 находим значение $x_{пр,\alpha}$, для которого $\Phi(x_{пр,\alpha}) = 0.495$. Это значение равно 2.58.

Тогда вероятность того, что критерий K при справедливости гипотезы H_0 примет значение больше 2.58, равна

$$P(K > 2.58) = P(2.58 < N(0,1) < \infty) = \Phi(\infty) - \Phi(2.58) = 0.005,$$

т.е. вероятности α ошибки первого рода.

Этап 5. В формулу критерия K, который является функцией n случайных величин $X_1, X_2, ..., X_n$, подставляются выборочные значения $x_1, x_2, ..., x_n$ и подсчитывается числовое значение критерия $K_{наб}$.

Если $K_{наб}$ попадает в критическую область ω, то гипотеза H_0 отвергается и принимается гипотеза H_1. При этом можно допус-

тить ошибку первого рода с вероятностью α. Если $K_{наб}$ не попадает в критическую область, гипотеза H_0 не отвергается. Однако это не означает, *что H_0 является единственной подходящей гипотезой*: просто H_0 не противоречит результатам наблюдений; возможно, таким же свойством, наряду с H_0, могут обладать и другие гипотезы.

Вновь обратимся к нашему примеру. Напомним, что из обследованных 1000 человек признаки заболевания R были обнаружены у 120 человек, т.е. $S_{1000} = 120$. Подставляя это выборочное значение в формулу (10.2.2), получаем: $K_{наб} = \frac{120-100}{9.487} = 2.108$.

Правосторонняя критическая точка ранее была определена как $x_{пр,\alpha/2} = 2.58$. Так как 2.108 < 2.58, то можно принять гипотезу $H_0 : p_1 = p_0$, а полученные расхождения между теоретической вероятностью $p_0 = 0.1$ и наблюдаемой частностью 0.120 считать допустимыми на уровне значимости $\alpha = 0.005$.

Если бы количество человек с признаками заболевания R составило 130 (из 1000 обследованных), то $K_{наб} = \frac{130-100}{9.487} = 3.162$.

В этом случае случайная величина K приняла значение из критической области, т.е. произошло событие $K > x_{пр,\alpha}$, которое практически невозможно, если гипотеза H_0 справедлива. Поэтому следует отвергнуть гипотезу H_0 в пользу альтернативной гипотезы $H_1 : p_1 > p_0$.

10.3. Проверка гипотезы о значении математического ожидания нормального распределения

Полагаем, что X является случайной величиной, имеющей нормальное распределение с параметрами a и σ, т.е. $X = N(a, \sigma)$, причем числовое значение a неизвестно.

Дать точный ответ на вопрос, каково численное значение неизвестного параметра a, по выборочной совокупности, нельзя.

Поэтому поступают следующим образом. Полагая, что наблюдения $X_1, X_2, ..., X_n$ независимы, вычисляют значение выборочной оценки $\overline{X}_в$, которое дает приближенные представления об a. Затем приступают к проверке гипотез о числовых значениях неизвестного параметра a.

Проверка гипотезы о числовом значении математического ожидания при известной дисперсии. Предполагается, что $X = N(a, \sigma)$, причем значение математического ожидания a неизвестно, а числовое значение дисперсии σ^2 известно.
Выдвинем гипотезу H_0 о том, что неизвестный параметр a равен числу a_0. Возможны три случая: 1) параметр a равен числу a_1, которое больше числа a_0 (т.е. $a > a_0$); 2) параметр a равен числу a_1, которое не равно a_0 (т.е. $a \neq a_0$); 3) параметр a равен числу a_1, которое меньше a_0 (т.е. $a < a_0$). Для случаев 1, 2 рассмотрим этапы проверки гипотезы H_0, приведенные в п. 10.2.

Случай 1. Определим следующие этапы.
Этап 1. Сформулируем нулевую гипотезу

$$\boxed{H_0 : a = a_0} \qquad (10.3.1)$$

и альтернативную

$$\boxed{H_1 : a = a_1 > a_0}. \qquad (10.3.2)$$

Этап 2. Зададимся уровнем значимости α.

Этап 3 . В качестве критерия возьмем величину

$$\boxed{K = \frac{\overline{X}_в - a_0}{\sigma/\sqrt{n}}}, \qquad (10.3.3)$$

значение которой зависит от выборочных данных (почему?), является случайной величиной и при выполнении гипотезы (10.3.1) подчиняется нормальному распределению $N(0, 1)$, т.е.

$$K = \frac{\overline{X}_в - a_0}{\sigma/\sqrt{n}} = N(0,1). \qquad (10.3.4)$$

Этап 4 . Построим критическую область ω, т.е. область таких значений критерия K, при которых гипотеза H_0 отвергается. Если нулевая и альтернативная гипотезы имеют вид (10.3.1), (10.3.2) соответственно, а критерий (10.3.4) – вид $K = N(0,1)$, то критическая область будет правосторонней: ее образует интервал $(x_{пр,\alpha}, +\infty)$, где $x_{пр,\alpha}$ определяется из условия

$$\boxed{P(N(0,1) > x_{пр,\alpha}) = \alpha}. \qquad (10.3.5)$$

Остановимся на методике вычисления $x_{пр,\alpha}$ (которая будет использована в дальнейшем для других критических точек). Вероятность события $N(0,1) \le x_{пр,\alpha}$ можно представить как

$$\int_{-\infty}^{0} p_{N(0,1)}(x)dx + \int_{0}^{x_{пр,\alpha}} p_{N(0,1)}(x)dx = \tfrac{1}{2} + \Phi(x_{пр,\alpha}),$$

где $p_{N(0,1)}(x)$ – плотность нормального распределения $N(0, 1)$; $\Phi(x)$ – функция Лапласа (см. табл. П1). Следовательно, вероятность противоположного события $N(0,1) > x_{пр,\alpha}$ выражается в виде $1 - \left[\frac{1}{2} + \Phi(x_{пр,\alpha})\right] = \frac{1}{2} - \Phi(x_{пр,\alpha})$, и эта вероятность должна быть равна α. Таким образом, приходим к уравнению

$$\boxed{\Phi(x_{пр,\alpha}) = \frac{1}{2} - \alpha}. \qquad (10.3.6)$$

Воспользовавшись табл. П1, находим значение $x_{пр,\alpha}$, удовлетворяющее этому уравнению. Для вычисления $x_{пр,\alpha}$ можно использовать функцию Excel НОРМСТОБР следующим образом:

$$\boxed{x_{пр,\alpha} = \text{НОРМСТОБР}(1-\alpha)}.$$

Критическая область изображена на рис. 10.1*а*.

Этап 5. Используя вместо $X_1, X_2, ..., X_n$ конкретные значения выборочной совокупности $x_1, x_2, ..., x_n$, находим $\overline{x}_в$ (см. формулу (8.2.2)), а затем численное значение $K_{наб}$ критерия (10.3.3). Если $K_{наб} > x_{пр,\alpha}$, то гипотеза H_0 (10.3.1) отвергается и принимается гипотеза H_1 (10.3.2). Напомним, что, поступая таким образом, можем совершить ошибку первого рода. Вероятность такой ошибки равна α.

Случай 2. Выполним следующие этапы.

Этап 1. Сформулируем нулевую гипотезу

$$\boxed{H_0 : a = a_0} \qquad (10.3.7)$$

и альтернативную

$$\boxed{H_1 : a \neq a_0}. \qquad (10.3.8)$$

Этап 2. Зададимся уровнем значимости α.

Этап 3. В качестве критерия, как и в случае 1, возьмем величину (10.3.3), которая при справедливости гипотезы (10.3.7) удовлетворяет распределению $N(0, 1)$.

Этап 4. Если нулевая и альтернативная гипотезы имеют соответственно вид (10.3.7), (10.3.8), а критерий определяется выражением (10.3.4), то критическая область будет двусторонней: ее образуют интервалы $(-\infty, x_{лев,\alpha/2})$, $(x_{пр,\alpha/2}, +\infty)$, где критические точки $x_{пр,\alpha/2}$, $x_{лев,\alpha/2}$ находятся из условия (10.2.6), которое, учитывая (10.3.4), запишется так:

$$\boxed{P(N(0,1) < x_{лев,\alpha/2}) = \frac{\alpha}{2}; \quad P(N(0,1) > x_{пр,\alpha/2}) = \frac{\alpha}{2}}. \qquad (10.3.9)$$

Из рис. 10.1*в* видно, что

$$\boxed{\Phi(x_{пр,\alpha/2}) = \frac{(1-\alpha)}{2}}. \qquad (10.3.10)$$

Воспользовавшись табл. П1, находим решение этого уравнения $x_{пр,\alpha/2}$. В силу симметричности функции плотности распределения $N(0, 1)$ имеем

$$\boxed{x_{лев,\alpha/2} = -x_{пр,\alpha/2}}. \qquad (10.3.11)$$

Для вычисления $x_{пр,\alpha/2}$ можно использовать функцию Excel НОРМСТОБР следующим образом:

$$\boxed{x_{пр,\alpha/2} = \text{НОРМСТОБР}(1-\frac{\alpha}{2})}.$$

Критическая область изображена на рис. 10.1*в*.

Этап 5. Находим числовое значение $K_{наб}$ критерия (10.3.3). Если $K_{наб}$ попадает в интервал $(-\infty, x_{лев,\alpha/2})$ или $(x_{пр,\alpha/2}, +\infty)$, то гипотеза H_0 (10.3.7) отвергается и принимается альтернативная (10.3.8). Поступая таким образом, можно с вероятностью α допустить ошибку первого рода.

Пример 10.3.1. По результатам n = 9 замеров установлено, что среднее время изготовления детали $\overline{x}_в = 52$ с. Предполагая, что время изготовления подчиняется нормальному распределению с дисперсией $\sigma^2 = 9\,\text{с}^2$, дать ответы на следующие вопросы:

а) можно ли принять 50 с в качестве нормативного времени (математического ожидания) изготовления детали;

б) можно ли принять за норматив 51 с?

Уровень значимости принять равным α = 0.05.

Решение.

а) по условию задачи нулевая гипотеза H_0: a = 50 с. Так как $\overline{x}_в = 52$ с, то в качестве альтернативной возьмем гипотезу $H_1 : a > 50$ с, т.е. имеем случай 1 (см. формулы (10.3.1), (10.3.2)) при $a_0 = 50$ с. По изложенной схеме получаем $x_{пр,\alpha} = 1.65$. Подставляя в формулу (5.11) исходные данные $\overline{x}_в = 52$ с, $\sigma = 3$, $n = 9$,

получаем $K_{наб} = \frac{52-50}{3/\sqrt{9}} = 2.$ Так как число 2 попадает в критическую область $(1.65, \infty)$, то гипотеза $H_0 : a = 50$ с отвергается и принимается $H_1 : a > 50$ с;

б) здесь нулевая гипотеза $H_0 : a = 51$ с, альтернативная $H_1 : a > 51$ с. Снова имеет место случай 1 при $a_0 = 51$ с. Так как $K_{наб} = \frac{51-50}{3/\sqrt{9}} = 1$ не попадает в критическую область, то гипотеза $H_0 : a = 51$ с не отвергается, и в качестве норматива времени изготовления детали берем 51 с. ☻

Проверка гипотезы о числовом значении математического ожидания при неизвестной дисперсии. В этом случае за основу проверки гипотезы

$$H_0 : a = a_0, \tag{10.3.12}$$

где a_0 – заранее заданное число, положен критерий

$$\boxed{K = \frac{\overline{X}_в - a_0}{S/\sqrt{n}}}, \tag{10.3.13}$$

где $\overline{X}_в$, S – случайные величины, вычисляемые по формулам (8.2.1) и (8.2.15). Этот критерий при выполнении гипотезы (10.3.12) имеет t-распределение с числом степеней свободы $k = n - 1$, т.е.

$$\boxed{K = \frac{\overline{X}_в - a_0}{S/\sqrt{n}} = T_{n-1}}, \tag{10.3.14}$$

где T_{n-1} – случайная величина, подчиняющаяся распределению Стьюдента.

Задаваясь уровнем значимости α, построим критическую область для проверки гипотезы (10.3.12) при следующих альтернативных гипотезах.

Случай 1. Альтернативная гипотеза:

$$\boxed{H_1 : a > a_0}. \qquad (10.3.15)$$

Критическая область является правосторонней – ее образует интервал $(x_{пр,\alpha}, +\infty)$, где точка $x_{пр,\alpha}$ определяется из условия (10.2.5), которое с учетом (10.3.14) можно записать в виде:

$$\boxed{P(T_{n-1} > x_{пр,\alpha}) = \alpha}. \qquad (10.3.16)$$

В табл. П2 приведены значения $t(\gamma, n-1)$, определяемые соотношением $\int\limits_{-t(\gamma,n-1)}^{t(\gamma,n-1)} p_{T_{n-1}}(x)dx = \gamma$, где $n-1$ – число степеней свободы. Так как функция плотности *t*-распределения симметрична относительно нуля, то искомая точка $x_{пр,\alpha}$ определяется как

$$\boxed{x_{пр,\alpha} = t(1-2\alpha, n-1)}. \qquad (10.3.17)$$

Значение $x_{пр,\alpha}$ можно также вычислить, используя функцию Excel СТЬЮДРАСПОБР в виде:

$$\boxed{x_{пр,\alpha} = \text{СТЬЮДРАСПОБР}(2\alpha; n-1)}.$$

Подставив в (10.3.13) конкретные значения $\overline{X}_в$, *S*, получаем значение критерия $K_{наб}$. Если $K_{наб} > x_{пр,\alpha}$ (т.е. попадает в критическую область), то гипотеза (10.3.12) отвергается и принимается гипотеза (10.3.15). При этом возможна ошибка первого рода с вероятностью α.

Случай 2. Альтернативная гипотеза

$$\boxed{H_1 : a \neq a_0}. \qquad (10.3.18)$$

Критическая область состоит из двух интервалов

$$(-\infty, x_{лев,\alpha/2}),\ (x_{пр,\alpha/2}, +\infty),$$

где критические точки $x_{лев,\alpha/2}$, $x_{пр,\alpha/2}$ определяются из условий (10.2.7), которые с учетом (10.3.14) можно записать в виде:

$$\boxed{P(T_{n-1} < x_{лев,\alpha/2}) = \alpha/2;\ P(T_{n-1} > x_{пр,\alpha/2}) = \alpha/2}.$$

Обращаясь к табл. П2, находим

$$\boxed{x_{лев,\alpha/2} = -t(1-\alpha, n-1);\ x_{пр,\alpha/2} = t(1-\alpha, n-1)}. \quad (10.3.19)$$

Значение $x_{пр,\alpha/2}$ можно также вычислить, используя функцию Excel СТЬЮДРАСПОБР в виде:

$$\boxed{x_{пр,\alpha/2} = \text{СТЬЮДРАСПОБР}(\alpha; n-1)},$$

а затем $\boxed{x_{лев,\alpha/2} = -x_{пр,\alpha/2}}$.

Подставляя в формулу (10.3.13) конкретные значения величин $\overline{X}_в$, S, получаем значение критерия $K_{наб}$. Если $K_{наб}$ попадает в интервал $(-\infty, x_{лев,\alpha/2})$ или $(x_{пр,\alpha/2}, +\infty)$, то гипотеза H_0 (10.3.12) отвергается и принимается альтернативная гипотеза H_1 (10.3.18). Если $K_{наб} \in [x_{лев,\alpha/2}, x_{пр,\alpha/2}]$, то принимается основная гипотеза H_0 (10.3.12).

Пример 10.3.2. Хронометраж затрат времени на сборку узла машины n = 20 слесарей показал, что $\overline{x}_в = 77$ мин, а $s^2 = 4$ мин^2. В предположении о нормальности распределения решить вопрос: можно ли на уровне значимости $\alpha = 0.05$ считать 80 мин нормативом (математическим ожиданием) трудоемкости?

Решение. В качестве основной гипотезы принимается $H_0 : a = 80$ мин, в качестве альтернативной $H_1 : a \neq 80$ мин, т.е. имеем случай 2, при этом $a_0 = 80$. Используя формулу (10.3.19) и табл. П2 $(k = n-1 = 19)$, находим

$$x_{пр,\alpha/2} = 2.093;\quad x_{лев,\alpha/2} = -x_{пр,\alpha/2} = -2.093. \quad (10.3.20)$$

Значение $x_{пр,\alpha/2}$ можно вычислить, используя функцию СТЬЮДРАСПОБР:

$$x_{пр,\alpha/2} = \text{СТЬЮДРАСПОБР}(0.05;19) = 2.093.$$

По формуле (10.3.13) вычисляем $K_{наб} = (77-80)/(2\sqrt{2}) =$ $= -6.708$. Так как число –6.708 попадает в критическую область (конкретно в интервал $(-\infty, -2.093)$), то гипотеза $H_0 : a = 80$ мин отвергается. ☻

10.4. Проверка гипотезы о равенстве математических ожиданий двух нормальных распределений

Проверка гипотезы о равенстве математических ожиданий двух генеральных совокупностей имеет важное практическое значение. Действительно, иногда оказывается, что средний результат $\overline{x}_в$ одной серии наблюдений отличается от среднего результата $\overline{y}_в$ другой серии. Возникает вопрос: можно ли это различие объяснить случайной ошибкой экспериментов или оно неслучайно? Иначе говоря, можно ли считать, что результаты экспериментов представляют собой выборки из двух генеральных совокупностей с одинаковыми средними? Приведем точную формулировку задачи.

Пусть генеральные совокупности X и Y распределены по нормальному закону, причем их среднеквадратические отклонения известны и равны соответственно σ_X и σ_Y. Требуется по двум независимым выборкам $x_1, \ldots, x_n$ и $y_1, \ldots, y_m$ из генеральных совокупностей X и Y проверить гипотезу о равенстве генеральных средних, т.е. основная гипотеза имеет вид:

$$\boxed{H_0 : M(X) = M(Y)}. \qquad (10.4.1)$$

Построим критерий проверки этой гипотезы, основываясь на следующем соображении: так как приближенное представление о математическом ожидании дает выборочное среднее, то в

основе проверки гипотезы (10.4.1) должно лежать сравнение выборочных средних $\overline{X}_в, \overline{Y}_в$. Найдем закон распределения разности $(\overline{X}_в - \overline{Y}_в)$. Эта разность является случайной величиной, и если гипотеза H_0 (10.4.1) верна, то

$$M(\overline{X}_в - \overline{Y}_в) = M\left(\frac{X_1 + ... + X_n}{n} - \frac{Y_1 + ... + Y_m}{m}\right) = M(X) - M(Y) = 0.$$

Пользуясь свойствами дисперсии, получим

$$D(\overline{X}_в - \overline{Y}_в) = D\left(\frac{X_1 + ... + X_n}{n} - \frac{Y_1 + ... + Y_m}{m}\right) =$$
$$= \frac{nD(X)}{n^2} + \frac{mD(Y)}{m^2} = \frac{D(X)}{n} + \frac{D(Y)}{m} = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}. \qquad (10.4.2)$$

Так как случайная величина $\overline{X}_в - \overline{Y}_в$ является линейной комбинацией независимых нормально распределенных случайных величин $X_1, ..., X_n$, $Y_1, ..., Y_m$, то $\overline{X}_в - \overline{Y}_в$ распределена по нормальному закону с параметрами $a = 0$, $\sigma^2 = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$. В качестве критерия выберем пронормированную случайную величину $\overline{X}_в - \overline{Y}_в$, т.е.

$$\boxed{K = \frac{\overline{X}_в - \overline{Y}_в}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}}. \qquad (10.4.3)$$

Таким образом, если гипотеза (10.4.1) верна, случайная величина K имеет нормальное распределение $N(0, 1)$, т.е.

$$K = \frac{\overline{X}_в - \overline{Y}_в}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = N(0,1). \qquad (10.4.4)$$

Теперь зададимся уровнем значимости α и перейдем к построению критических областей и проверке гипотезы (10.4.1)

для двух видов альтернативной гипотезы H_1. Заметим, что вычисление критических точек критерия, распределенного по нормальному закону $N(0, 1)$, подробно рассматривалось в п. 10.3. Поэтому здесь ограничимся только определением соответствующих критических областей.

1. Альтернативная гипотеза имеет вид:

$$\boxed{H_1 : M(X) > M(Y)}. \qquad (10.4.5)$$

В этом случае критическая область есть интервал $(x_{пр,\alpha}, +\infty)$, где критическая точка $x_{пр,\alpha}$ определяется из условия $P(N(0,1) > x_{пр,\alpha}) = \alpha$ (см. п. 10.3). Для вычисления $x_{пр,\alpha}$ можно использовать функцию Excel НОРМСТОБР следующим образом:

$$\boxed{x_{пр,\alpha} = \text{НОРМСТОБР}(1-\alpha)}.$$

Критическая область приведена на рис. 10.1*а*. Подставляя в (10.4.3) вместо случайных величин $\overline{X}_в, \overline{Y}_в$ их выборочные значения $\overline{x}_в, \overline{y}_в$, найдем значение критерия $K_{наб}$. Если $K_{наб} > x_{пр,\alpha}$, то гипотезу H_0 (10.4.1) отвергаем и принимаем гипотезу H_1 (10.4.5). Поступая таким образом, можно допустить ошибку первого рода с вероятностью α.

Пример 10.4.1. По двум независимым выборкам, извлеченным из нормальных генеральных совокупностей, объемы которых равны $n = 12$ и $m = 8$, найдены средние значения $\overline{x}_в = 143$, $\overline{y}_в = 122$. Генеральные дисперсии известны: $\sigma_X^2 = D(X) = 36, \sigma_Y^2 = D(Y) = 8$. При уровне значимости $\alpha = 0.005$ нужно проверить гипотезу $H_0 : M(X) = M(Y)$ при конкурирующей гипотезе $H_1 : M(X) > M(Y)$.

Решение. Критическую точку $x_{пр,\alpha}$ находим по табл. П1 из условия $\Phi(x_{пр,\alpha}) = \frac{1}{2} - \alpha = 0.495$. Получаем $x_{пр,\alpha} = 2.58$. Наблюдаемое значение критерия равно:

$$K_{наб} = \frac{143-122}{\sqrt{\frac{36}{12}+\frac{8}{8}}} = \frac{21}{2} = 10.5.$$

Так как $K_{наб} > 2.58$, то гипотеза о равенстве генеральных средних отвергается на уровне значимости α = 0.05. ☻

2. Альтернативная гипотеза имеет вид:

$$\boxed{H_1 : M(x) \neq M(y)}. \qquad (10.4.6)$$

В этом случае наибольшая мощность критерия достигается при двусторонней критической области, состоящей из двух интервалов: $(-\infty, x_{лев,\alpha/2})$ и $(x_{пр,\alpha/2}, +\infty)$. Критические точки определяются из условия:

$$\boxed{P(N(0,1) < x_{лев,\alpha/2}) = \alpha/2;\ P(N(0,1) > x_{пр,\alpha/2}) = \alpha/2}.$$

Для вычисления $x_{пр,\alpha/2}$ можно использовать функцию Excel НОРМСТОБР следующим образом:

$$\boxed{x_{пр,\alpha/2} = \text{НОРМСТОБР}(1-\frac{\alpha}{2})}.$$

В силу симметрии плотности распределения $N(0, 1)$ относительно нуля $x_{лев,\alpha/2} = -x_{пр,\alpha/2}$.

Если числовое значение критерия $K_{наб}$, вычисленное по формуле (10.4.3), попадает в интервал $(-\infty, x_{лев,\alpha/2})$ или $(x_{пр,\alpha/2}, +\infty)$, то гипотеза H_0 отвергается и принимается гипотеза H_1 (10.4.6); если $x_{лев,\alpha/2} < K_{наб} < x_{пр,\alpha/2}$, то принимаем гипотезу H_0 (10.4.1).

Рассмотрим проверку ***гипотезы о равенстве математических ожиданий двух произвольных распределений по выборкам большого объема.*** Пусть $x_1,...,x_n$ – выборка из генеральной совокупности X, а $y_1,...,y_m$ – выборка из генеральной

совокупности Y, причем объемы выборок n и m достаточно большие (не менее 30 элементов в каждой). Распределение генеральных совокупностей нам неизвестно, но недостаток этой информации компенсируется большими объемами выборок. Согласно центральной предельной теореме, случайная величина $\overline{X}_в - \overline{Y}_в$ распределена по закону, близкому к нормальному. Если гипотеза $H_0 : M(X) = M(Y)$ верна, то $M(\overline{X}_в - \overline{Y}_в) = 0$. Как и ранее, $D(\overline{X}_в - \overline{Y}_в) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$, однако σ_X^2, σ_Y^2 неизвестны. Но при выборках большого объема случайные величины $D_{вх}$ (выборочная дисперсия X) и $D_{ву}$ (выборочная дисперсия Y) являются достаточно хорошими оценками для $D(x)$ и $D(y)$. Поэтому случайная величина

$$\boxed{K = \frac{\overline{X}_в - \overline{Y}_в}{\sqrt{\frac{D_{вх}}{n} + \frac{D_{ву}}{m}}}} \qquad (10.4.7)$$

распределена по закону, близкому к нормальному $N(0, 1)$, и может быть принята в качестве критерия. Тогда построение критических областей для двух видов конкурирующих гипотез осуществляется так же, как и для случая известных дисперсий.

Пример 10.4.2. По двум независимым выборкам объемов $n = 120$, $m = 150$ найдены значения выборочных дисперсий $d_{вх} = 1.2$ и $d_{ву} = 4.5$, а также средние значения $\overline{x}_в = 30$, $\overline{y}_в = 28.3$. При уровне значимости $\alpha = 0.05$ проверить гипотезу $H_0 : M(X) = M(Y)$ при конкурирующей $H_1 : M(X) \neq M(Y)$.

Решение. Вычислим наблюдаемое значение критерия K:

$$K_{наб} = \frac{\overline{X}_в - \overline{Y}_в}{\sqrt{\frac{d_{вх}}{n} + \frac{d_{ву}}{m}}} = \frac{30 - 28.3}{\sqrt{\frac{1.2}{120} + \frac{4.5}{150}}} = 8.5.$$

Правую границу $x_{пр,\alpha/2}$ двусторонней критической области $(x_{пр,\alpha/2},+\infty)$ найдем из условия $\Phi(x_{пр,\alpha/2}) = (1-\alpha)/2$ = 0,475. Получаем $x_{пр,\alpha/2} = 1.96$, $x_{лев,\alpha/2} = -1.96$. Так как $K_{наб} > x_{пр,\alpha/2}$, гипотеза о равенстве генеральных средних на уровне значимости $\alpha = 0.05$ отвергается. ☻

10.5. Проверка гипотезы о числовом значении дисперсии нормального распределения

Полагаем, что X является случайной величиной, имеющей нормальное распределение $N(a,\sigma)$ причем числовое значение дисперсии σ^2 неизвестно. Исправленная оценка дисперсии $S^2 = \sum_{i=1}^{n}(X_i - X_в)^2/(n-1)$ дает приближенное представление о σ^2. Используя эту оценку, проверим гипотезу

$$\boxed{H_0 : \sigma^2 = \sigma_0^2}, \qquad (10.5.1)$$

где σ_0^2 – заранее заданное число. В качестве критерия возьмем случайную величину

$$\boxed{K = \frac{(n-1)S^2}{\sigma_0^2}}. \qquad (10.5.2)$$

При выполнении гипотезы (10.5.1) эта величина подчиняется χ^2-распределению с числом степеней свободы $k = n-1$, т.е.

$$K = \frac{(n-1)S^2}{\sigma_0^2} = \chi^2_{n-1}. \qquad (10.5.3)$$

Зададимся уровнем значимости α и перейдем к построению критических областей для проверки гипотезы H_0 (10.5.1) при следующих двух альтернативных гипотезах H_1.

Случай 1. В качестве альтернативной гипотезы примем

$$\boxed{H_1 : \sigma^2 > \sigma_0^2}. \qquad (10.5.4)$$

Критическая область является правосторонней и определяется интервалом $(x_{пр,\alpha}, +\infty)$, где критическая точка $x_{пр,\alpha}$ находится из условия:

$$\boxed{P(\chi^2_{n-1} > x_{пр,\alpha}) = \alpha}. \qquad (10.5.5)$$

В табл. П4 приведены квантили $\chi^2(\gamma, k)$, определяемые соотношением

$$P\left(\chi_k^2 < \chi^2(\gamma, k)\right) = \gamma = 1 - \alpha .$$

Следовательно, искомая критическая точка $x_{пр,\alpha}$ находится как

$$x_{пр,\alpha} = \chi^2(1-\alpha, n-1) . \qquad (10.5.6)$$

Для вычисления $x_{пр,\alpha}$ можно обратиться к функции ХИ2ОБР:

$$\boxed{x_{пр,\alpha} = \text{ХИ2ОБР}(1-\alpha ; n-1)}. \qquad (10.5.7)$$

Подставив в (10.5.2) конкретные значения S^2, σ_0^2, находим $K_{наб}$. Если $K_{наб} > x_{пр,\alpha}$, то гипотеза H_0 (10.5.1) отвергается и принимается гипотеза H_1 (10.5.4) с вероятностью ошибки первого рода равной α.

Случай 2. В качестве альтернативной гипотезы примем

$$\boxed{H_1 : \sigma^2 \neq \sigma_0^2}. \qquad (10.5.8)$$

В этом случае критическая область состоит из двух интервалов $(0, x_{лев,\alpha/2})$ и $(x_{пр,\alpha/2}, +\infty)$, где критические точки $x_{лев,\alpha/2}$, $x_{пр,\alpha/2}$ определяются из условий:

$$\boxed{P(\chi^2_{n-1} < x_{лев,\alpha/2}) = \alpha/2 ;\ P(\chi^2_{n-1} > x_{пр,\alpha/2}) = \alpha/2}.$$

Обращаясь к табл. П4, находим

$$x_{лев,\alpha/2} = \chi^2(\alpha/2, n-1)\,;\ x_{пр,\alpha/2} = \chi^2(1-\alpha/2, n-1)\,.$$

Для вычисления критических точек $x_{лев,\alpha/2}$ $x_{пр,\alpha/2}$ можно обратиться к функции ХИ2ОБР:

$$\boxed{x_{лев,\alpha/2} = \text{ХИ2ОБР}(1-\alpha/2\,; n-1)}; \qquad (10.5.9)$$

$$\boxed{x_{пр,\alpha/2} = \text{ХИ2ОБР}(\alpha/2\,; n-1)}. \qquad (10.5.10)$$

Если значение $K_{наб}$, вычисленное по формуле (10.5.2), попадает в один из интервалов $(0, x_{лев,\alpha/2})$ или $(x_{пр,\alpha/2}, \infty)$, то гипотеза H_0 отвергается и принимается гипотеза H_1 (10.5.8). В противном случае нет оснований отвергнуть гипотезу H_0 (10.5.1).

Пример 10.5.1. Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделия. По выборке из 25 деталей вычислена $s^2 = 0.25$. При уровне значимости α = 0.05 проверить гипотезу $H_0 : \sigma^2 = 0.15$.

Решение. За альтернативную примем гипотезу $H_1 : \sigma^2 > 0.15$, т.е. имеем случай 1. По табл. П4 находим $x_{пр,0.05} = \chi^2(0.95, 24) = 36.4$, следовательно, критическая область $(36.4, \infty)$. Значение $x_{пр,0.05}$ можно вычислить, используя функцию ХИ2ОБР (см. формулу (10.5.7)):

$$x_{пр,0.05} = \text{ХИ2ОБР}(1-0.05\,; 24) = 36.415.$$

По формуле (5.26) находим

$$K_{наб} = (25-1)0.25/0.15 = 40.$$

Так как $K_{наб}$ попадает в критическую область, гипотезу H_0 отвергаем в пользу альтернативной гипотезы. ☻

10.6. Проверка гипотезы о законе распределения с применением критерия согласия Пирсона

Нередко в приложениях математической статистики фигурируют задачи, в которых закон распределения генеральной совокупности заранее неизвестен, но есть основания предположить, выдвинуть гипотезу, что он имеет определенный вид. Проверка гипотезы о предполагаемом законе неизвестного распределения производится так же, как и проверка гипотезы о параметрах распределения, т.е. при помощи специально подобранной случайной величины – *критерия согласия* (или критерия χ^2) К. Пирсона. С этой целью сравниваются эмпирические m_i (наблюдаемые) и теоретические n_i (вычисленные в предположении о правильности гипотезы) частоты.

Критерий согласия Пирсона. Пусть дана выборка ($x_1, x_2, ..., x_n$), извлеченная из генеральной совокупности случайной величины X. Требуется проверить гипотезу о том, что X подчиняется некоторому закону распределения $F_0(x)$:

$$\boxed{H_0 : F(x) = F_0(x)}. \qquad (10.6.1)$$

Здесь $F(x)$ – неизвестная функция распределения случайной величины X, а $F_0(x)$ – известная (предполагаемая) функция распределения. Альтернативная гипотеза:

$$\boxed{H_1 : F(x) \neq F_0(x)}. \qquad (10.6.2)$$

При этом можно выделить два случая: 1) параметры $\theta_1, ..., \theta_l$ распределения X известны; 2) параметры $\theta_1, ..., \theta_l$ неизвестны. Проверку нулевой гипотезы, когда параметры $\theta_1, ..., \theta_l$ неизвестны (более сложного случая), проводят по схеме.

1. По выборочным данным находят оценки параметров $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_l$, которые принимают за неизвестные параметры.

2. Область возможных значений величины X разбивают на L интервалов: $\Delta_1, \Delta_2, ..., \Delta_L$ (не обязательно одинаковой вели-

чины). Если X может принимать значения всей действительной оси, то начало первого интервала надо положить равным $(-\infty)$, а конец последнего – $(+\infty)$. Далее находят частоты $m_1, m_2, ..., m_L$ $(\sum_{i=1}^{L} m_i = n)$ в каждом интервале, т.е. строят интервальный вариационный ряд.

3. Находят вероятности p_i того, что X принимает значения из интервала Δ_i, при справедливости гипотетического закона $F_0(x)$:

$$p_i = P(z_i < X < z_{i+1}) = F_0(z_{i+1}) - F_0(z_i),$$

где z_i и z_{i+1} – левая и правая границы интервала Δ_i. При этом должно выполняться $\sum_{i=1}^{L} p_i = 1$.

4. Если H_0 верна, то выборочную частоту m_i можно рассматривать как число появления события, которое в каждом из n проведенных испытаний появляется с заданной вероятностью p_i. Следовательно, математическое ожидание $M(m_i) = np_i$ и $\sigma = \sqrt{np_i(1-p_i)}$. Зная теоретические частоты np_i для каждого интервала Δ_i, можно сравнить их с соответствующими эмпирическими частотами m_i. В качестве меры различия между m_i и np_i вычисляют величину критерия

$$\boxed{K = \sum_{i=1}^{L} \frac{(m_i - np_i)^2}{np_i}}. \qquad (10.6.3)$$

Критерий (10.6.3) в условиях справедливости H_0 распределен по закону «хи-квадрат» – χ_k^2 с числом степеней свободы $k = L - l - 1$, где l – число параметров гипотетического закона распределения $F_0(x)$ (см. замечание 10.3.1). Критерий (10.6.3) носит название «критерий Пирсона».

5. Далее задаемся уровнем значимости α и, зная распределение критерия K, строим правостороннюю критическую об-

ласть. Это будет область вида $(x_{пр,\alpha}, +\infty)$. Критическая точка $x_{ї\partial,\alpha}$ находится из условия $P(\chi_k^2 > x_{пр,\alpha}) = \alpha$. В табл. П4 приведены значения $\chi^2_{\gamma,k}$, удовлетворяющие условию $P(\chi_k^2 < \chi^2_{\gamma,k}) = \gamma$. Следовательно,

$$x_{пр,\alpha} = \chi^2_{1-\alpha,k}\,. \qquad (10.6.4)$$

Значение $x_{пр,\alpha}$ можно вычислить с помощью функции ХИ2ОБР обращением:

$$\boxed{x_{пр,\alpha} = \text{ХИ2ОБР}(\alpha; k)}. \qquad (10.6.5)$$

Если наблюдаемое значение критерия (10.6.3)

$$K_{наб} < x_{пр,\alpha}\,, \qquad (10.6.6)$$

то считают, что выборочные данные согласуются с гипотетическим распределением $F_0(x, \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_l)$. В противном случае H_0 отвергается в пользу альтернативной гипотезы H_1.

Замечание 10.6.1. Величина критерия K из формулы (10.5.3) практически распределена по закону χ_k^2, если для каждого Δ_i $np_i \geq 10$. Если это условие не выполняется для некоторых интервалов, то рекомендуется эти интервалы объединить с соседними. На практике часто выдвигают требование, чтобы все m_i были не меньше 4. ●

Замечание 10.6.2. Если в качестве гипотетического закона используется нормальный закон $N(\overline{x}_в, S)$, то число неизвестных параметров $m = 2$, значит число степеней свободы

$$\boxed{k = L - 2 - 1 = L - 3}. \quad ● \qquad (10.6.7)$$

Замечание 10.6.3. Выбор гипотетического закона распределения $F_0(x)$ осуществляется, исходя из вида гистограммы распределения и сравнения ее с кривой плотности вероятности, соответствующей $F_0(x)$. ●

Пример 10.6.1. По выборке объема $n = 144$ (табл. 10.1) составлен группированный статистический ряд.

Таблица 10.1

X	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8
m_i	16	17	19	16	24	19	17	16

Проверить на уровне значимости α = 0.05 гипотезу о равномерности распределения генеральной совокупности на отрезке [0, 8].

Решение. Нулевая гипотеза имеет вид:

$$H_0 : p_X(x) = p(x) = \begin{cases} \dfrac{1}{8-0}, & 0 \le x \le 8; \\ 0 & \text{для остальных } x. \end{cases} \qquad (10.6.8)$$

Вычислим вероятность попадания случайной величины X на каждый интервал:

$$p_i = \int_i^{i+1} \frac{1}{8} dx = \frac{1}{8}(i - i + 1) = \frac{1}{8}, \qquad i = 1, 2, ..., 8.$$

Поэтому $np_i = \frac{1}{8} \cdot 144 = 18$ при любом i. Так как $np_i \ge 10$, то нет необходимости объединять несколько интервалов. Результаты дальнейших вычислений сведены в табл. 10.2.

Таким образом, числовое значение $K_{наб} = 2.90$. Для заданного уровня значимости $\alpha = 0.05$ находим $\gamma = 1 - \alpha = 0.95$, $x_{пр,\alpha} = \chi^2_{0.95,7} = 14.067 \approx 14.1$. Так как $K_{наб} < x_{пр,\alpha}$, то гипотеза H_0 (10.6.8) принимается. ☻

Обычной является ситуация, когда предполагается лишь, что распределение генеральной совокупности принадлежит некоторому классу распределений. Например, генеральная совокупность распределена нормально. В этой гипотезе не оговорены значения параметров a и σ. Отличие в применении критерия χ^2 в этом случае от ранее рассмотренного примера 10.6.1 со-

стоит в том, что нет возможности сразу вычислить значения вероятностей, так как параметры a и σ неизвестны.

Таблица 10.2

Номер интервала	m_i	np_i	$m_i - np_i$	$\frac{(m_i - np_i)^2}{np_i}$
1	16	18	–2	0.22
2	17	18	–1	0.06
3	19	18	1	0.06
4	16	18	–2	0.22
5	24	18	6	2.00
6	19	18	1	0.06
7	17	18	–1	0.06
8	16	18	–2	0.22
$\sum$	144	144	0	2.9

Поэтому вначале находят оценки неизвестных параметров. Например, для оценки параметра a, как известно, можно использовать случайную величину $\overline{X}_в$ и заменить a ее выборочной оценкой, т.е. $\hat{a} = \overline{x}_в$. В качестве оценки параметра σ^2 можно выбрать исправленную дисперсию S^2 и заменить σ^2 ее значением s^2. Таким образом, получаем плотность распределения:

$$p(x) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{(x-\overline{x}_в)^2}{2s^2}}. \qquad (10.6.9)$$

В качестве критерия также принимается случайная величина (10.6.3). Если гипотеза H_0 справедлива, то критерий имеет χ^2-распределение с k степенями свободы. Однако количество степеней свободы критерия подсчитывается по формуле

$k = L - 3$ (см. замечание 10.6.2). Вероятность p_i попадания случайной величины $X = N(\overline{x}_в, s)$ в интервал $[z_i, z_{i+1}]$, $i = 1, ..., L+1$ (L – количество интервалов) находится с помощью функции Лапласа:

$$\boxed{p_i = P(z_i < N(\overline{x}_в, s) < z_{i+1}) = \Phi\left(\frac{z_{i+1} - \overline{x}_в}{s}\right) - \Phi\left(\frac{z_i - \overline{x}_в}{s}\right)}. \quad (10.6.10)$$

Пример 10.6.2. Группированный вариационный ряд частот занесен в столбцы 2 и 3 табл. 10.3. Количество интервалов $L = 10$, узлы z_i ($i = 1, ..., 11$) вычисляются по формуле $z_i = -20 + (i-1) \cdot 5$. Номера интервалов приведены в столбце 1. По выборке объема $n = 200$ найдено $\overline{x}_в = 4.30$, $s^2 = 94.26$, $s = 9.71$. При уровне значимости $\alpha = 0.02$ проверить гипотезу о нормальности распределения генеральной совокупности.

Решение. Так как $p_i = \Phi\left(\frac{z_{i+1} - \overline{x}_в}{s}\right) - \Phi\left(\frac{z_i - \overline{x}_в}{s}\right)$, $i = 1, ..., 10$, то в графе 4 вычислены значения $\frac{z_{i+1} - \overline{x}_в}{s}$. При этом левая граница первого интервала z_1 заменена на $-\infty$, а правая граница последнего интервала z_{L+1} – на $+\infty$. Поэтому

$$\Phi\left(\frac{z_1 - \overline{x}_в}{s}\right) = \Phi(-\infty) = -\Phi(\infty) = -0.5; \quad \Phi\left(\frac{z_{L+1} - \overline{x}_в}{s}\right) = \Phi(\infty) = 0.5.$$

В графе 6 вычислены вероятности p_i по формуле (10.6.10), в графе 7 – математические ожидания np_i, а в столбце 8 – взвешенные отклонения $\frac{(m_i - np_i)^2}{np_i}$. Для примера вычислим вероятность p_1 попадания нормально распределенной величины $N(4.30, 7.91)$ в интервал $[z_1, z_2]$ по формуле:

$$p_1 = \Phi\left(\frac{z_2 - \overline{x}_в}{s}\right) - \Phi\left(\frac{z_1 - \overline{x}_в}{s}\right) = \Phi\left(\frac{z_2 - \overline{x}_в}{s}\right) - \Phi(-\infty) =$$

$$= \Phi\left(\frac{-15 - 4.30}{9.71}\right) - \Phi(-\infty) = \Phi(-1.99) - \Phi(-\infty) =$$

$$= -0.4767 + 0.5 = 0.23.$$

Вероятности для других интервалов вычисляются аналогичным образом.

Так как для 9-го и 10-го интервалов $np_9 = 7.2 < 10$ и $np_{10} = 3.32 < 10$, то эти интервалы объединяем. Для полученного «объединенного» интервала $[z_9, z_{11}] = [20, 30]$ вероятность $p_{об}$ попадания в него случайной величины $N(4.30, 7.91)$ вычисляется как

$$p_{об} = \Phi\left(\frac{z_{11} - \overline{x}_в}{s}\right) - \Phi\left(\frac{z_9 - \overline{x}_в}{s}\right) = \Phi(\infty) - \Phi\left(\frac{20 - 4.30}{9.71}\right) =$$

$$= \Phi(\infty) - \Phi(1.62) = 0.5 - 0.4474 = 0.0526.$$

Вычислим величину $np_{об} = 0.0526 \cdot 200 = 10.52 > 10$ (столбец 7).

Числовое значение критерия $K_{наб} = 7.19$ (столбец 8). По табл. ПЗ при $\gamma = 1 - \alpha = 0.98$ и $k = 9 - 2 - 1 = 6$ находим $\chi^2_{0.98,6} = 15.033 \approx 15.0$, $x_{пр,\alpha} = 15.0$.

Так как $K_{наб} < 15.0$, то гипотеза H_0 о нормальности распределения генеральной совокупности принимается на уровне значимости $\alpha = 0.02$.

Для графической иллюстрации правильности принятия гипотезы H_0 на рис. 10.2 показан фрагмент документа Excel, в котором приведены: середины интервалов $z_i^* = \frac{z_i + z_{i+1}}{2}$, $i = 1, ..., L = 10$ (столбец B); частоты m_i (столбец C); высота гистограммы относительных частот $y_i = \frac{m_i}{200 \cdot 5}$ (столбец F); высота

гистограммы вероятностей $y_i^* = p_i / 5$ (столбец G), где число 5 – длина отрезка в основании прямоугольников гистограммы.

Таблица 10.3

Номер интервала	Границы интервалов	m_i	$\frac{z_{i+1} - \overline{x}_{\hat{a}}}{s}$	$\Phi\left(\frac{z_{i+1} - \overline{x}_{\hat{a}}}{s}\right)$	p_i	np_i	$\frac{(m_i - np_i)^2}{np_i}$
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1	[–20, –15]	7	–1.99	–0.4767	0.023	4.66	1.18
2	[–15, –10]	11	–1.47	–0.4292	0.047	9.50	0.24
3	[–10, –5]	15	–0.96	–0.331	0.097	19.54	1.05
4	[–5, 0]	24	–0.44	–0.1700	0.161	32.30	2.13
5	[0, 5]	49	+0.07	0.0279	0.197	39.58	2.24
6	[5, 10]	41	0.59	0.222	0.194	38.90	0.11
7	[10, 15]	26	1.10	0.364	0.141	28.38	0.20
8	[15, 20]	17	1.62	0.4474	0.083	16.62	0.01
9	[20, 25]	7	2.13	0.4834	0.0526	10.52	0.03
10	[25, 30]	3	$+\infty$	0.5	–	–	–
$\sum$	–	***200***	–	–	***1***	***200.0***	***7.19***

На этом же рисунке нанесены столбчатые диаграммы, соответствующие: ряд 2 – значения y_i; ряд 3 – значения y_i^*. Видно достаточно хорошее совпадение контуров и высот этих гистограмм. ☻

	B	C	F	G
4	-17,5	7	0,007	0,0046
5	-12,5	11	0,011	0,0094
6	-7,5	15	0,015	0,0194
7	-2,5	24	0,024	0,0324
8	2,5	49	0,049	0,0396
9	7,5	41	0,041	0,0388
10	12,5	26	0,026	0,0282
11	17,5	17	0,017	0,0168
12	22,5	7	0,007	0,0074
13	27,5	3	0,003	0,0034
14			*0,2*	*0,2*

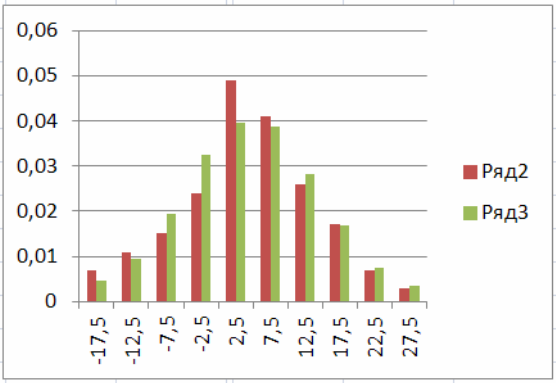


Рис. 10.2. Гистограммы эмпирических и теоретических вероятностей

10.7. Проверка статистических гипотез в Excel

В табличном процессоре Excel определены несколько функций и режимов работы модуля ***Пакета анализа***, которые можно использовать для проверки статистических гипотез, рассмотренных выше. Для вызова этого модуля необходимо:

- в табличном процессоре Excel 2003 обратиться к пункту **Сервис** строки меню Excel, а затем щелкнуть на команду *Анализ данных*;
- в табличном процессоре Excel 2007 или Excel 2010 обратиться к пункту **Данные** строки меню Excel, а затем щелкнуть на команду *Анализ данных*.

Проверка гипотезы о равенстве математических ожиданий двух нормальных распределений с известными дисперсиями. Изучаются две нормально распределенные случайные величины $X \sim N(a_X, \sigma_X)$, $Y \sim N(a_Y, \sigma_Y)$. Числовые значения дисперсий σ_X^2, σ_Y^2 известны. Проверяется основная гипотеза – $H_0: M(X) = M(Y)$.

Для проверки этой гипотезы используется режим работы ***Двухвыборочный z-тест для средних***, который задается в диалоговом окне модуля ***Пакета анализа,*** показанного на рис. 10.3.

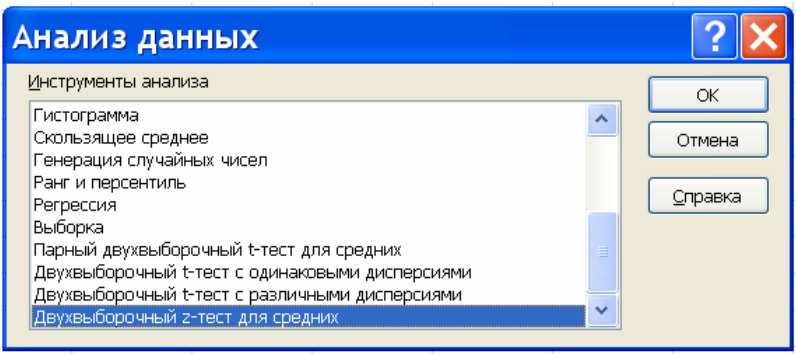


Рис. 10.3. Диалоговое окно модуля ***Пакет анализ***

В диалоговом окне этого режима (рис. 10.4) задаются следующие параметры:

Интервал переменной 1 – адреса ячеек, содержащих выборочные значения величины *X*.

Интервал переменной 2 – адреса ячеек, содержащих выборочные значения величины *Y*.

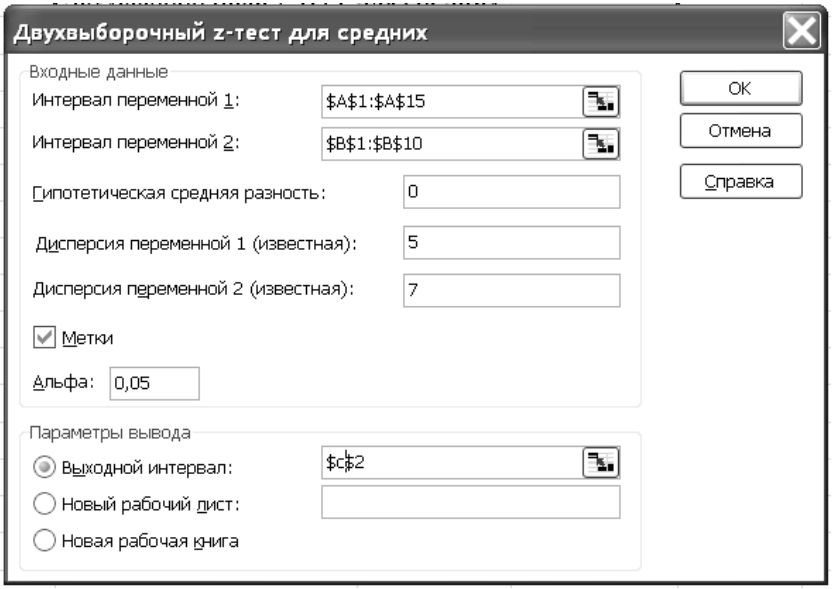


Рис. 10.4. Задание параметров режима
Двухвыборочный z-тест для средних

Гипотетическая средняя разность задает число, равное предполагаемой разности математических ожиданий $a_X - a_Y$ (при проверке гипотезы о равенстве математических ожиданий задается 0).

Дисперсия переменной 1 (известная) – вводится известное значение σ_x^2.

Дисперсия переменной 2 (известная) – вводится известное значение σ_y^2.

Метки – включается, если первая строка содержит заголовки столбцов.

Альфа – задается уровень значимости.

Выходной интервал / новый рабочий лист / новая рабочая книга – указывается, куда выводятся результаты вычислений. При включении *Выходной интервал* вводится адрес ячейки, начиная с которой выводятся результаты, которые оформлены в виде таблицы (рис. 10.5).

Двухвыборочный z-тест для средних		
	Автомат1	*Автомат 2*
Среднее	181,979	185,03
Известная дисперсия	5,000	7,00
Наблюдения	14,000	9,00
Гипотетическая разность средних	0,000	
z	-2,867	
P(Z<=z) одностороннее	0,002	
z критическое одностороннее	1,645	
P(Z<=z) двухстороннее	0,004	
z критическое двухстороннее	1,960	

Рис. 10.5. Результаты работы режима
Двухвыборочный z-тест для средних

Результаты работы представлены таблицей, приведенной на рис. 10.5. Буквой *z* обозначается наблюдаемое значение критерия. Дальше выводятся значения:

- z критическое одностороннее – значение $x_{пр,\alpha}$ при построении односторонней критической области (альтернативная гипотеза имеет вид $H_1 : M(X) > M(Y)$);
- z критическое двухстороннее – значение $x_{пр,\alpha/2}$ при построении двусторонней критической области (альтернативная гипотеза имеет вид $H_1 : M(X) \neq M(Y)$).

Остальные вычисленные величины приводятся в таблице с понятными названиями. Если выполняется неравенство

$$\boxed{|z| > z_{кр}}, \qquad (10.7.1)$$

то гипотеза H_0 отвергается и принимается альтернативная гипотеза H_1.

Пример 10.7.1. Выборочные данные о диаметре валиков (мм), изготовленных автоматом 1 и автоматом 2, приведены в столбцах *A* и *B* документа Excel, показанного на рис. 10.6. Предварительным анализом установлено, что размер валиков, изготовленных каждым автоматом, имеет нормальное распределение с дисперсиями $\sigma_X^2 = 5\,\text{мм}^2$ (автоматом 1) и $\sigma_Y^2 = 7\,\text{мм}^2$ (автоматом 2).

Необходимо проверить нулевую гипотезу H_0: $a_X = a_Y$ при альтернативной гипотезе $H_1 : a_X \neq a_Y$.

Решение. Обратимся к режиму ***Двухвыборочный z-тест для средних*** и в появившемся диалоговом окне зададим необходимые параметры (см. рис. 10.4), а затем щелкнем ОК. Результаты работы режима показаны на рис. 10.5. Заметим, что при альтернативной гипотезе $H_1 : a_X \neq a_Y$ критическая область является двусторонней.

Величина z является расчетным значением критерия (10.4.3) $K_{наб} = z = -2.867$, и это значение попадает в двустороннюю критическую область $(-\infty, -1.96] \cup [1.96, \infty)$. Действительно, $|K_{наб}| > |z_{кр}| = 1.96$. Поэтому нулевая гипотеза отвергается с

уровнем значимости $\alpha = 0.05$, и принимается альтернативная гипотеза $a_X \neq a_Y$. ☻

	A	B
1	Автомат1	Автомат 2
2	182,3	185,3
3	183,0	185,6
4	181,8	184,8
5	181,4	186,2
6	181,8	185,8
7	181,6	184,0
8	183,2	185,2
9	182,4	184,2
10	182,5	184,2
11	179,7	
12	179,9	
13	181,9	
14	182,8	
15	183,4	
16	Среднее	Среднее
17	***182,0***	***185,0***

Рис 10.6. Исходные данные к примеру 10.7.1

Проверка гипотезы о равенстве математических ожиданий двух нормальных распределений с неизвестными, но равными дисперсиями. Изучаются две нормально распределенные случайные величины $X \sim N(a_X, \sigma_X)$, $Y \sim N(a_Y, \sigma_Y)$. Известно, что дисперсии равны, но не известны, т.е. $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. Необходимо проверить статистическую гипотезу H_0: $a_X = a_Y$ при различных альтернативных гипотезах.

Для проверки этой гипотезы используется режим ***Двухвыборочный t-тест с одинаковыми дисперсиями***. В диалоговом окне этого режима задаются следующие параметры (рис. 10.7):

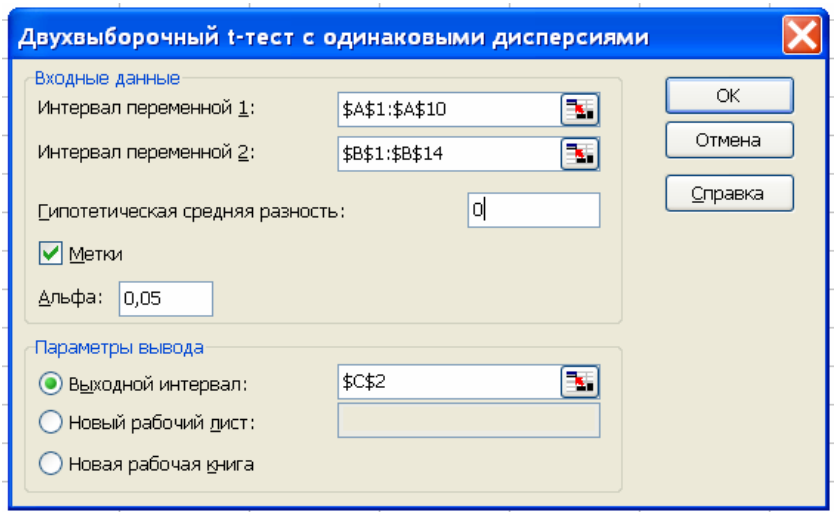


Рис. 10.7. Задание параметров режима
Двухвыборочный t-тест с одинаковыми дисперсиями

Интервал переменной 1 – адреса ячеек, содержащих выборочные значения величины *X*.

Интервал переменной 2 – адреса ячеек, содержащих выборочные значения величины *Y*.

Гипотетическая средняя разность задает число, равное предполагаемой разности математических ожиданий $a_X - a_Y$ (при проверке гипотезы $a_X = a_Y$ задается 0).

Метки – включается, если первая строка содержит заголовки столбцов.

Альфа – задает уровень значимости α.

Выходной интервал / новый рабочий лист / новая рабочая книга – указывается, куда выводятся результаты вычислений.

При включении *Выходной интервал* вводится адрес ячейки, начиная с которой выводятся результаты (рис. 10.8).

Двухвыборочный t-тест с одинаковыми дисперсиями		
	Старая технология	***Новая технология***
Среднее	307,11	304,92
Дисперсия	1,61	2,24
Наблюдения	9,00	13,00
Объединенная дисперсия	1,99	
Гипотетическая разность средних	0,00	
df	20,00	
t-статистика	3,58	
P(T<=t) одностороннее	0,00094	
t критическое одностороннее	1,72	
P(T<=t) двухстороннее	0,00189	
t критическое двухстороннее	2,09	

Рис. 10.8. Результаты работы режима
Двухвыборочный t-тест с одинаковыми дисперсиями

Результаты работы представлены таблицей, приведенной на рис. 10.8. Как *t*-статистика обозначается наблюдаемое значение критерия. Дальше выводятся значения:

- *t* критическое одностороннее – значение $x_{пр,\alpha}$ при построении односторонней критической области (альтернативная гипотеза имеет вид $H_1 : M(X) > M(Y)$);
- *t* критическое двустороннее – значение $x_{пр,\alpha/2}$ при построении двусторонней критической области (альтернативная гипотеза имеет вид $H_1 : M(X) \neq M(Y)$).

Остальные вычисленные величины приводятся в таблице с понятными названиями. Если выполняется неравенство

$$|t| > t_{кр}, \qquad (10.7.2)$$

то гипотеза H_0 отвергается, и принимается альтернативная гипотеза H_1.

Пример 10.7.2. Выборочные данные о расходе сырья при производстве продукции по старой и новой технологиям приведены в столбцах A, B документа Excel, показанного на рис. 10.9. При предположении, что расход сырья по старой и новой технологиям распределен по нормальному закону и имеет одинаковую дисперсию, проверить статистическую гипотезу $a_X = a_Y$ при уровне значимости α = 0.05.

	A	B
1	Старая технология	Новая технология
2	308	308
3	308	304
4	307	306
5	308	306
6	304	306
7	307	304
8	307	304
9	308	306
10	307	306
11		304
12		303
13		304
14		303

Рис. 10.9. Исходные данные к примеру 10.7.2

Решение. Обратимся к режиму ***Двухвыборочный t с одинаковыми дисперсиями*** и в появившемся диалоговом окне зададим необходимые параметры (см. рис. 10.5), а затем щелкнем ОК. Результаты работы режима показаны на рис. 10.6. Величина t-статистика является наблюдаемым значением критерия (5.46): $K_{наб} = 3.58$. Это значение попадает в критическую область $(-\infty, -2.09] \cup [2.09, \infty)$. Действительно, $|K_{наб}| > |t_{кр}| = 2.09$. Следовательно, ***нулевая гипотеза*** $a_X = a_Y$ ***отвергается*** с уровнем значимости 0.05, и принимается альтернативная гипотеза $a_X \neq a_Y$.

Вопросы и задания для самопроверки

1. Что понимается под статистической гипотезой?
2. Перечислить этапы проверки статистических гипотез.
3. Дать определение ошибки первого и второго рода.
4. Как связана величина уровня значимости с границами критической области?
5. Какова связь между выбором вида альтернативной гипотезы и типом критической области?
6. Если сформулированы гипотезы: основная H_0, альтернативная H_1 – и известно, что $P(H_0 / H_0) = 0.9$, то чему равна вероятность ошибки первого рода при проверке гипотез?
   Ответ: $P(H_1 / H_0) = 1 - \gamma = \alpha = 0.1$.
7. Если уровень значимости при проверке гипотезы равен 0.05, то каков уровень доверия при этой проверке?
   Ответ: $P(H_0 / H_0) = 1 - \alpha = \gamma = 1 - 0.05 = 0.95$.
8. Как называется вероятность принять альтернативную гипотезу, если она верна?
   Ответ: Мощность критерия.
9. Если мощность критерия равна 0.8, то какова вероятность ошибки второго рода?
   Ответ: $P(H_0 / H_1) = \beta = 1 - 0.8 = 0.2$.
10. Если уровень значимости при проверке гипотезы равен 0.2, то какова вероятность того, что наблюдаемое значение двустороннего критерия «попадет» в критическую область?
    Ответ: $\frac{\alpha}{2} = 0.1$.
11. Какие альтернативные гипотезы могут быть сформулированы при проверке основной гипотезы о равенстве математического ожиданий?
12. Привести виды критериев, используемых в задачах о проверке статистических гипотез.
13. По двум независимым выборкам объемов $n = 120$, $m = 150$ найдены значения выборочных дисперсий $d_{вх} = 1.2$ и $d_{ву} = 4.5$, а также средние значения $\overline{x}_в = 30$, $\overline{y}_в = 28.3$.

При уровне значимости $\alpha = 0.05$ проверить гипотезу $H_0 : M(X) = M(Y)$ при конкурирующей гипотезе $H_1 : M(X) \neq M(Y)$.

14. Хронометраж затрат времени на сборку узла машины $n = 20$ слесарей показал, что среднее время сборки $\overline{x}_в = 77$ мин, а $s^2 = 4$. В предположении о нормальности распределения затрат времени на сборку решить вопрос о том, можно ли принять 80 мин нормативом (математическим ожиданием) трудоемкости на уровне значимости $\alpha = 0.01$.
15. Какая основная идея положена в основу критерия согласия Пирсона?
16. Используя критерий Пирсона, при уровне значимости 0.01 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с заданным эмпирическим распределением выборки.

x_i	2	4	6	8	10	12	14	16	18	20	22
n_i	2	4	6	10	18	20	16	11	7	5	1

17. При уровне значимости 0.025 проверить гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты.

Эмпирические частоты	6	15	38	74	108	85	32	14
Теоретические частоты	5	14	42	80	99	75	37	15

ЗАКЛЮЧЕНИЕ

В математических исследованиях прикладных задач важную роль играют методы и подходы, изучаемые в теории вероятностей и математической статистике.

В данном учебном пособии в разделе «Теория вероятностей» изложены основные понятия и определения теории вероятностей, важнейшие числовые (математическое ожидание, дисперсия) характеристики дискретных случайных величин и распределенные (функция распределения, плотность вероятности) характеристики непрерывных случайных величин. Большое внимание уделено исследованию корреляционного взаимодействия случайных величин. Эти вопросы также рассматриваются в учебниках [1, 3–6].

В разделе «Математическая статистика» рассмотрены: обработка статистических данных, построение выборочного коэффициента корреляции, получение точечных и интервальных оценок неизвестных параметров, статистическая проверка статистических гипотез. Эти вопросы также рассматриваются в учебниках [2–4, 6].

Для более глубокого изучения задач и методов теории вероятностей и математической статистики рекомендуется обратиться к учебникам [7, 8], приведенным в библиографическом списке.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Воскобойников Ю. Е. Теория вероятностей : учеб. пособие / Ю. Е. Воскобойников, Е. И. Тимошенко ; Новосибирский гос. архитектур.-строит. ун-т. – Новосибирск, 2003. – 134 с.
2. Воскобойников Ю. Е. Математическая статистика (с примерами в Excel) : учеб. пособие / Ю. Е. Воскобойников, Е. И. Тимошенко ; Новосибирский гос. архитектур.-строит. ун-т. – Новосибирск, 2006. – 156 с.
3. Гмурман В. Е. Теория вероятностей и математическая статистика : учеб. пособие для вузов / В. Е. Гмурман. – 10-е изд., стер. – М. : Высшая школа, 2008. – 480 с.
4. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособие для вузов / В. Е. Гмурман. – 5-е изд., стер. – М. : Высшая школа, 2008. – 405 с.
5. Вентцель Е. С. Задачи и упражнения по теории вероятностей / Е. С. Вентцель, Л. А. Овчаров. – М. : Высшая школа, 2000. – 325 с.
6. Смирнов Н. В. Курс теории вероятностей и математической статистики / Н. В. Смирнов, И. В. Дунин-Барковский. – М. : Наука, 1969. – 653 с.
7. Боровков А. А. Теория вероятностей / А. А. Боровков. – М. : Наука, 1976. – 354 с.
8. Боровков А. А. Математическая статистика / А. А. Боровков. – М. : Наука, 1984. – 472 с.

ПРИЛОЖЕНИЯ

Таблица П1

Значения функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-z^2/2)dz$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.0000	0.33	0.1293	0.66	0.2454	0.99	0.3389
0.02	0.0080	0.35	0.1368	0.68	0.2517	1.01	0.3438
0.03	0.0120	0.36	0.1406	0.69	0.2549	1.02	0.3461
0.04	0.0160	0.37	0.443	0.70	0.2580	1.03	0.3485
0.05	0.0199	0.38	0.1480	0.71	0.2611	1.04	0.3508
0.06	0.0239	0.39	0.1517	0.72	0.2642	1.05	0.3531
0.07	0.0279	0.40	0.1554	0.73	0.2673	1.06	0.3554
0.08	0.0319	0.41	0.1591	0.74	0.2703	1.07	0.3577
0.09	0.0359	0.42	0.1628	0.75	0.2734	1.08	0.3599
0.10	0.0398	0.43	0.1664	0.76	0.2764	1.09	0.3621
0.11	0.0438	0.44	0.1700	0.77	0.2794	1.10	0.3643
0.12	0.0478	0.45	0.1736	0.78	0.2823	1.11	0.3665
0.13	0.0517	0.46	0.1772	0.79	0.2852	1.12	0.3686
0.14	0.0557	0.47	0.1808	0.80	0.2881	1.13	0.3708
0.15	0.0596	0.48	0.1844	0.81	0.2910	1.14	0.3729
0.16	0.0636	0.49	0.1879	0.82	0.2939	1.15	0.3749
0.17	0.0675	0.50	0.1915	0.83	0.2967	1.16	0.3770
0.18	0.0714	0.51	0.1950	0.84	0.2995	1.17	0.3790
0.19	0.0753	0.52	0.1985	0.85	0.3023	1.18	0.3810
0.20	0.0793	0.53	0.2019	0.86	0.3051	1.19	0.3830
0.21	0.0832	0.54	0.2054	0.87	0.3078	1.20	0.3849
0.22	0.0871	0.55	0.2088	0.88	0.3106	1.21	0.3869
0.23	0.0910	0.56	0.2123	0.89	0.3133	1.22	0.3883
0.24	0.0948	0.57	0.2157	0.90	0.3159	1.23	0.3907
0.26	0.1026	0.59	0.2224	0.92	0.3212	1.25	0.3944
0.27	0.1064	0.60	0.2257	0.93	0.3238	1.26	0.3962
0.28	0.1103	0.61	0.2291	0.94	0.3264	1.27	0.3980
0.29	0.1141	0.62	0.2324	0.95	0.3289	1.28	0.3997
0.30	0.1179	0.63	0.2357	0.96	0.3315	1.29	0.4015
0.32	0.1225	0.65	0.2422	0.98	0.3365	1.31	0.4049

Окончание табл. П1

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1.32	0.4066	1.69	0.4545	2.12	0.4830	2.86	0.4979
1.33	0.4082	1.70	0.4554	2.14	0.4838	2.88	0.4980
1.34	0.4099	1.71	0.4564	2.16	0.4846	2.90	0.4981
1.35	0.4115	1.72	0.4573	2.18	0.4854	2.92	0.4982
1.36	0.4131	1.73	0.4582	2.20	0.4861	2.94	0.4984
1.37	0.4137	1.74	0.4591	2.22	0.4868	2.96	0.4985
1.38	0.4162	1.75	0.4599	2.24	0.4875	2.98	0.4986
1.39	0.4177	1.76	0.4608	2.26	0.4881	3.00	0.49865
1.40	0.4192	1.77	0.4616	2.28	0.4887	3.20	0.49931
1.41	0.4207	1.78	0.4625	2.30	0.4893	3.40	0.49966
1.42	0.4222	1.79	0.4633	2.32	0.4898	3.60	0.499841
1.43	0.4236	1.80	0.4641	2.34	0.4904	3.80	0.499928
1.44	0.4251	1.81	0.4649	2.36	0.4909	4.00	0.499968
1.45	0.4265	1.82	0.4656	2.38	0.4913	4.50	0.499997
1.46	0.4279	1.83	0.4664	2.40	0.4918	5.00	0.499997
1.47	0.4292	1.84	0.4671	2.42	0.4922		
1.49	0.4319	1.86	0.4686	2.46	0.4931		
1.50	0.4332	1.87	0.4693	2.48	0.4934		
1.51	0.4345	1.88	0.4699	2.50	0.4938		
1.52	0.4357	1.89	0.4706	2.52	0.4938		
1.53	0.4370	1.90	0.4713	2.54	0.4945		
1.54	0.4382	1.91	0.4719	2.56	0.4948		
1.55	0.4394	1.92	0.4726	2.58	0.4951		
1.56	0.4406	1.93	0.4732	2.60	0.4953		
1.57	0.4418	1.94	0.4738	2.62	0.4956		
1.58	0.4429	1.95	0.4744	2.64	0.4959		
1.59	0.4441	1.96	0.4750	2.66	0.4961		
1.60	0.4452	1.97	0.4756	2.68	0.4961		
1.62	0.4474	1.99	0.4767	2.72	0.4965		
1.63	0.4484	2.00	0.4772	2.74	0.4967		
1.64	0.4495	2.02	0.4783	2.76	0.4971		
1.65	0.4505	2.04	0.4793	2.78	0.4973		
1.66	0.4515	2.06	0.4803	2.80	0.4974		
1.68	0.4535	2.10	0.4821	2.84	0.4977		

Таблица П2

Таблица значений $t(\gamma, k)$, определяемых выражением $P(|T_k| < t(\gamma, k)) = \gamma,$ где k – число степеней свободы

k \ γ	0.95	0.99	0.999	k \ γ	0.95	0.99	0.999
4	2.78	4.6	8.61	19	2.093	2.861	3.883
5	2.57	4.03	6.86	24	2.064	2.797	3.745
6	2.45	3.71	5.96	29	2.045	2.756	3.659
7	2.37	3.50	5.41	34	2.032	2.720	3.600
8	2.31	3.36	5.04	39	2.023	2.070	3.558
9	2.26	3.25	4.78	44	2.016	2.692	3.527
10	2.23	3.17	4.59	49	2.009	2.679	3.502
11	2.20	3.11	4.44	59	2.001	2.662	3.464
12	2.18	3.06	4.32	69	1.996	2.649	3.439
13	2.16	3.01	4.22	79	1.991	2.640	3.418
14	2.15	2.98	4.14	89	1.987	2.633	3.403
15	2.13	2.95	4.07	99	1.984	2.627	3.339
16	2.12	2.92	4.02	119	1.980	2.617	3.374
17	2.11	2.90	3.97	∞	1.960	2.576	3.291
18	2.10	2.88	3.92				

Таблица П3

Таблица значений q_γ

(n – объем выборки)

n	γ			n	γ		
	0.95	0.99	0.999		0.95	0.99	0.999
5	1.37	2.67	5.64	20	0.37	0.58	0.88
6	1.09	2.01	3.88	25	0.32	0.49	0.73
7	0.92	162	2.98	30	0.28	0.43	0.63
8	0.80	1.38	2.42	35	0.26	0.38	0.56
9	0.71	1.20	2.06	40	0.24	0.35	0.50
10	0.65	1.08	1.80	45	0.22	0.32	0.46
11	0.59	0.98	1.60	50	0.21	0.30	0.43
12	0.55	0.90	1.45	60	0.188	0.269	0.38
13	0.52	0.83	0.33	70	0.174	0.245	0.34
14	0.48	0.78	1.23	80	0.161	0.266	0.31
15	0.46	0.73	1.15	90	0.151	0.211	0.29
16	0.44	0.70	1.07	100	0.143	0.198	0.27
17	0.42	0.66	1.01	150	0.115	0.160	0.211
18	0.40	0.63	0.96	200	0.099	0.136	0.185
19	0.39	0.60	0.92	250	0.089	0.120	0.162

Таблица П4

Таблица значений квантилей χ_k^2-распределения, определяемых соотношением $P(\chi_k^2 < \chi_{\chi,k}^2) = \gamma$

k \ γ	0.02	0.05	0.1	0.9	0.95	0.98
1	0.006	0.004	0.016	2.7	3.8	5.4
2	0.040	0.103	0.211	4.6	6.0	7.8
3	0.185	0.352	0.584	6.3	7.8	9.8
4	0.43	0.71	1.06	7.8	9.5	11.7
5	0.75	1.14	1.61	9.2	11.1	13.4
6	1.13	1.63	2.20	10.6	12.6	15.0
7	1.56	2.17	2.83	12.0	14.1	16.6
8	2.03	2.73	3.49	13.4	15.5	18.2
9	2.53	3.32	4.17	14.7	16.9	19.7
10	3.06	3.94	4.86	16.0	18.3	21.2
12	4.2	5.2	6.3	18.5	21.0	24.1
14	5.4	6.6	7.8	21.1	23.7	26.9
16	6.6	8.0	9.3	23.5	26.3	29.6
18	7.9	9.4	10.9	26.0	28.9	32.3
20	9.2	10.9	12.4	28.4	31.4	35.0
22	10.6	12.3	14.0	30.8	33.9	37.7
24	12.0	138	15.7	33.2	36.4	40.3
26	13.4	15.4	17.3	35.6	38.9	42.9
28	14.8	16.9	18.9	37.9	41.3	45.4
30	16.3	18.5	20.6	40.3	43.8	48.0

Учебное издание

Воскобойников Юрий Евгеньевич
Баланчук Татьяна Тимофеевна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА (с примерами в Excel)

Учебное пособие

ISBN 978-5-7795-0632-8

Темплан 2013 г.

Редактор Э.Е. Полякова

Санитарно-эпидемиологическое заключение
№ 54.НС.05.953.П.006252.06.06 от 26.06.2006 г.
Подписано к печати 28.08.2013. Формат 60×84 1/16 д.л.
Гарнитура Таймс. Бумага офсетная. Ризография.
Объём 11,6 уч.-изд.л.; 12,75 п.л. Тираж 160 экз. Заказ №

Новосибирский государственный
архитектурно-строительный университет (Сибстрин)
630008, Новосибирск, ул. Ленинградская, 113

Отпечатано мастерской оперативной полиграфии
НГАСУ (Сибстрин)