МЕЖРЕГИОНАЛЬНАЯ
АКАДЕМИЯ УПРАВЛЕНИЯ ПЕРСОНАЛОМ

МАУП

Р. К. Чорней

ПРАКТИКУМ
по теории вероятностей и математической статистике

Учебное пособие
для студентов-иностранцев
вузовской подготовки

Киев
ДП «Издательский дом «Персонал»
2009

ББК 22.17я7
Ч-75

Рецензенты: *И. И. Юртин*, канд. физ.-мат. наук, проф. МАУП
Б. М. Ляшенко, д-р физ.-мат. наук, проф.
П. С. Кнопов, д-р физ.-мат. наук, проф.

Одобрено Ученым советом Межрегиональной Академии управления персоналом (протокол № 5 от 30.05.07)

Чорней, Р. К.
Ч-75 Практикум по теории вероятностей и математической статистике: Учеб. пособие для студ.-иностр. вузовской подготовки / Р. К. Чорней. — К. : ДП «Изд. дом «Персонал», 2009. — 336 с. : ил. — Библиогр.: с. 326–327.

ISBN 978-966-608-929-1

В предлагаемом практикуме представлены необходимые теоретические данные и формулы, решения типичных задач, задачи для самостоятельного решения с указаниями и ответами к ним. Значительное внимание уделено методам статистической обработки экспериментальных данных.

Для студентов-иностранцев нематематических специальностей высших учебных заведений, а также для всех, кто применяет теорию вероятностей и статистические методы при решении практических задач.

ББК 22.17я7

ISBN 978-966-608-929-1

Предисловие

«Практикум по теории вероятностей и математической статистике» — результат обработки семестрового курса лекций по теории вероятностей и математической статистике, прочитанных автором на протяжении нескольких лет в Межрегиональной Академии управления персоналом для студентов-иностранцев нематематических специальностей. Главная цель практикума — помочь как преподавателям, так и студентам на практических занятиях по дисциплине «Теория вероятностей и математическая статистика».

Практикум рассчитан прежде всего на студентов экономических специальностей, которые имеют математическую подготовку в объёме обычного курса высших экономических учебных заведений. Автор старался изложить предмет по возможности проще и нагляднее, не связывая себя рамками полной математической строгости. В этой связи не даётся «строго математического» определения вероятности, много положений подаются на интуитивно понятном уровне.

В начале каждого раздела даются ключевые слова в не совсем привычном виде. Автор умышленно избегал словосочетаний, разбивая их на отдельные слова. Например, словосочетание «непрерывная случайная величина» разбито на три слова «непрерывный», «случайный», «величина». Это сделано для того, чтобы студенты-иностранцы имели возможность напротив каждого ключевого слова записать карандашом перевод на родной язык.

С большой благодарностью автор примет от читателей любые пожелания, относящиеся к содержанию практикума, стилю изложения и характеру рассмотренных примеров.

Руслан Чорней
chorneir@i.com.ua

Часть I
Случайные события

Раздел 1. Случайные события. Определение вероятности

Ключевые слова

Благоприятный
Вероятность
Геометрический
Группа
Достоверный
Испытание
Классический
Невозможный
Несовместимый
Операция
Определение
Опыт
Относительный
Произведение
Пространство
Противоположный
Полный
Разность
Случайный
Событие
Совместимый
Сумма
Статистический
Частота
Элементарный

1.1. Случайные события

Обеспечение определённого комплекса условий называют *испытанием* или *опытом*, а возможный результат испытания — *событием*. Например, подбрасывание монеты — испытание, а выпадение «герба» или «номинала» — событие. События будем обозначать большими латинскими буквами: A, B, C.

Событие называют *случайным*, если оно может состояться или не состояться в данном испытании.

Достоверным называют событие, которое обязательно состоится в данном испытании.

Невозможным называют событие, которое точно не состоится в данном испытании.

Отметим, что любое событие связано с определённым испытанием.

Два события называют *совместимыми*, если появление одного из них не исключает появления другого в одном и том же испытании.

Два события называют *несовместимыми*, если они не могут выполняться одновременно в одном и том же испытании.

Попарно несовместимые случайные события $A_1, A_2, \ldots, A_n$ образуют *полную группу событий*, если вследствие испытания одно из них обязательно состоится. Например, события «выигрыш», «проигрыш» и «ничья» (для определённого игрока) образуют полную группу событий в испытании — игре в шахматы двух соперников.

Элементарными событиями в определённом испытании называют все возможные результаты этого испытания, которые нельзя разложить на более простые. Множество всех возможных элементарных событий ω называют *пространством элементарных событий*, которое обозначают Ω. Например, при подбрасывании игрального кубика пространство элементарных событий образуют события $\omega_i = \{\text{выпадет } i \text{ очков}\}$, $i = 1, 2, 3, 4, 5, 6$.

Элементарные события, при появлении которых происходит определённое событие, называют *благоприятными* для этого события. Например, при подбрасывании игрального кубика для события $A = \{\text{выпадет нечётное число очков}\}$ благоприятными являются элементарные события ω_1, ω_3, ω_5.

Каждое событие можно рассматривать как некоторое подмножество пространства элементарных событий в данном испытании. В частности, событие $A = \Omega$ — достоверное, а событие $B = \varnothing$ — невозможное.

Пример 1. Монету подбрасывают дважды. Для данного испытания описать пространство элементарных событий.

Решение. При двукратном подбрасывании монеты возможны четыре элементарных исхода:

$$(\mathrm{А}, \mathrm{А}); \quad (\mathrm{А}, \mathrm{Р}); \quad (\mathrm{Р}, \mathrm{А}); \quad (\mathrm{Р}, \mathrm{Р}),$$

где А — выпадение аверса (изображение «герба»); Р — выпадение реверса (изображение «номинала»). Очевидно, они образуют полную группу событий, поэтому

$$\Omega = \{(\mathrm{А}, \mathrm{А}); (\mathrm{А}, \mathrm{Р}); (\mathrm{Р}, \mathrm{А}); (\mathrm{Р}, \mathrm{Р})\} \text{ —}$$

пространство элементарных событий данного испытания. □

1.2. Операции над событиями

Суммой двух случайных событий A и B называют такое событие, которое состоит в появлении хотя бы одного из событий A или B. Эту операцию обозначают $A + B$ (или $A \cup B$).

Суммой n случайных событий $A_1, A_2, \ldots, A_n$ называют такое событие, которое состоит в появлении по крайней мере одного из этих событий (обозначается $\bigcup\limits_{i=1} n_i$).

Произведением двух случайных событий A и B называют такое событие, которое состоит в совместном появлении обоих событий A и B. Эту операцию обозначают $A \cdot B$ (или $A \cap B$).

Произведением n случайных событий $A_1, A_2, \ldots, A_n$ называют такое событие, которое состоит в совместном появлении всех этих событий (обозначается $\bigcap\limits_{i=1} n_i$).

Разностью двух случайных событий A и B называют событие, которое состоит в том, что происходит событие A и не происходит событие B. Эту операцию обозначают $A - B$ (или $A \setminus B$).

Событие $\overline{A}$ называют *противоположным* к событию A в данном испытании, если оно происходит тогда, когда не происходит событие A, т. е. $\overline{A} = \Omega - A$. Очевидно, что противоположные события несовместимы и образуют полную группу событий.

Пример 1. В ящике находятся шарики белого и чёрного цвета. Наугад из него вынимают один шарик. Событие $A = \{$вынут шарик белого цвета$\}$, событие $B = \{$вынут шарик чёрного цвета$\}$. Совместимы или несовместимы эти события?

Решение. Эти события несовместимы, так как появление события A исключает возможность появления события B, и наоборот. В данном испытании события A и B являются противоположными:

$$A = \overline{B}, \qquad B = \overline{A}. \qquad \square$$

Пример 2. Подбрасывают два игральных кубика. Пусть события $A_i = \{$выпадет i очков на первом кубике$\}$, $i = 1, 2, 3, 4, 5, 6$, $B_j = \{$выпадет j очков на втором кубике$\}$, $j = 1, 2, 3, 4, 5, 6$. Выразить через A_i, B_j такие события:

а) сумма очков на двух кубиках равняется пяти;

б) выпадет в сумме хотя бы десять очков;

в) выпадет в сумме не более трёх очков.

Решение. а) Пусть $C_1 = \{$сумма очков на двух кубиках равняется пяти$\}$. Это событие возможно лишь тогда, когда на первом кубике выпадет i очков, а на втором — j очков так, чтобы $i + j = 5$, т. е. $i = 1$, $j = 4$, или $i = 2$, $j = 3$, или $i = 3$, $j = 2$, или $i = 4$, $j = 1$. Итак,

$$C_1 = A_1 \cdot B_4 + A_2 \cdot B_3 + A_3 \cdot B_2 + A_4 \cdot B_1.$$

б) Обозначим $C_2 = \{$выпадет в сумме хотя бы десять очков$\}$. Событие C_2 состоится тогда, когда на двух кубиках в сумме выпадет или 10, или 11, или 12 очков, т. е. $i = 4$, $j = 6$, или $i = 5$, $j = 5$, или $i = 6$, $j = 4$, или $i = 6$, $j = 5$, или $i = 5$, $j = 6$, или $i = 6, j = 6$. Поэтому

$$C_2 = A_4 \cdot B_6 + A_5 \cdot B_5 + A_6 \cdot B_4 + A_6 \cdot B_5 + A_5 \cdot B_6 + A_6 \cdot B_6.$$

в) Пусть $C_3 = \{$выпадет в сумме не более трёх очков$\}$. Поскольку наименьшее количество очков, которое может выпасть на каждом кубике, равняется единице, то событие C_3 возможно лишь тогда, когда сумма очков на двух кубиках будет равняться или двум, или трём. Поэтому

$$C_3 = A_1 \cdot B_1 + A_1 \cdot B_2 + A_2 \cdot B_1. \qquad \square$$

Пример 3. Два стрелка стреляют в мишень по одному разу. Событие $A = \{$в мишень попал первый стрелок$\}$, событие $B = \{$в мишень попал второй стрелок$\}$. Выразить через A и B такие события: $C = \{$два попадания в мишень$\}$, $D = \{$ни одного попадания в мишень$\}$, $E = \{$хотя бы одно попадание в мишень$\}$, $F = \{$лишь одно попадание в мишень$\}$.

Решение. Пространство элементарных событий состоит из четырёх событий:

$$A \cdot B, \qquad \overline{A} \cdot B, \qquad A \cdot \overline{B}, \qquad \overline{A} \cdot \overline{B}.$$

Событие C состоится тогда, когда оба стрелка попадут в мишень. Поэтому оно является произведением двух событий A и B. Итак,

$$C = A \cdot B.$$

Событие D состоит в том, что в мишень не попадёт ни один стрелок, т. е. не попадёт ни первый ($\overline{A}$), ни второй ($\overline{B}$). Поэтому

$$D = \overline{A} \cdot \overline{B}.$$

Событие E состоится тогда, когда в мишень попадёт хотя бы один стрелок. Это может случиться тогда, когда или оба стрелка попадут в мишень, или первый попадёт, а второй не попадёт, или первый не попадёт, а второй попадёт. Поэтому

$$E = AB + \overline{A}B + A\overline{B},$$

т. е.

$$E = A + B.$$

Событие F состоит в том, что первый стрелок попадёт в мишень, а второй не попадёт или второй попадёт, а первый не попадёт. Поэтому

$$F = A\overline{B} + \overline{A}B. \qquad \square$$

1.3. Классическое определение вероятности

Пусть все элементарные исходы равновозможны.

Вероятность события A равняется отношению количества элементарных исходов, благоприятствующих этому событию, к количеству всех равновозможных элементарных исходов в данном испытании.

Вероятность события A обозначают $\mathbf{P}(A)$, поэтому по определению

$$\mathbf{P}(A) = \frac{m}{n}, \tag{1}$$

где m — количество элементарных событий, благоприятствующих событию A; n — количество всех элементарных событий в данном испытании.

Из классического определения вероятности вытекает, что

$$0 \leqslant \mathbf{P}(A) \leqslant 1,$$

причем $\mathbf{P}(A) = 0$, когда $A = \varnothing$ — невозможное событие, и $\mathbf{P}(A) = 1$, когда $A = \Omega$ — достоверное событие.

Пример 1. В урне находится 5 белых, 3 чёрных и 4 красных шара. Наугад вынимают один. Найти вероятность того, что наугад вынутый шар красный.

Решение. Пусть событие $A = \{$вынутый из урны шар красный$\}$. Общее количество шаров в урне — $5 + 3 + 4 = 12$, причём вынуть можно любой из них с одинаковой вероятностью. Поэтому в данном испытании есть 12 равновозможных исходов, т. е. $n = 12$. Количество событий, которые благоприятствуют событию A, определяется количеством красных шаров, т. е. $m = 4$. Итак, по определению (1) вероятность

$$\mathbf{P}(A) = \frac{m}{n} = \frac{4}{12} = \frac{1}{3}. \qquad \square$$

Пример 2. Найти вероятность того, что выбранное случайным образом двузначное число делится на:

а) 3;

б) 5.

Решение. В данном случае испытание состоит в том, что выбирают случайным образом двузначное число. Исходом такого испытания является одно из чисел от 10 до 99. Поскольку таких чисел 90, то $n = 90$.

а) Пусть событие $A = \{$выбранное двузначное число делится на $3\}$. Поскольку каждое третье с 90 двузначных чисел делится на 3, то благоприятными для события A являются 30 исходов, т. е. $m = 30$. Тогда по формуле (1) вероятность события A

$$\mathbf{P}(A) = \frac{30}{90} = \frac{1}{3}.$$

б) Пусть событие $B = \{$выбранное двузначное число делится на $5\}$. Общее количество исходов испытания, как и в предыдущем случае, $n = 90$. Определим количество чисел, которые делятся на 5. Очевидно, что таких чисел будет $m = 18$ (каждое пятое число делится на 5). Итак,

$$\mathbf{P}(B) = \frac{18}{90} = \frac{1}{5}. \qquad \square$$

Классическое определение вероятности предусматривает, что количество элементарных исходов конечное. Если множество всех элементарных исходов испытания бесконечное, применяют *геометрическое определение вероятности.*

1.4. Геометрическое определение вероятности

Пусть множество всех элементарных событий испытания бесконечно и образует некоторое множество Ω, все элементарные события равновозможны, причём событию A благоприятствуют те элементарные события, которые образуют множество $A \subseteq \Omega$. Тогда вероятность события A равняется отношению меры множества A к мере множества Ω, т. е.

$$\mathbf{P}(A) = \frac{m(A)}{m(\Omega)}. \tag{2}$$

Мерой множества на прямой, плоскости, в пространстве является соответственно длина, площадь, объём геометрической фигуры, которую образует это множество.

Пример 1. Два действительных числа случайным образом выбирают из интервала $[0; 5]$. Какая вероятность того, что:

а) сумма двух чисел меньше 4;

б) произведение двух чисел больше 5;

в) разность двух чисел меньше 2, а их произведение больше 3?

Решение. Обозначим через x первое число, выбранное случайным образом из интервала $[0; 5]$, а через y — второе число. Тогда $0 \leqslant x \leqslant 5$, $0 \leqslant y \leqslant 5$. Вследствие бесконечного количества таких действительных чисел надо воспользоваться определением геометрической вероятности. В этом случае множеством всех возможных исходов испытания является квадрат (рис. 1—3) со стороной 5, площадь которого равняется 25, т. е.

$$m(\Omega) = 25.$$

а) Пусть событие $A = \{$сумма двух чисел меньше 4$\}$. Тогда

$$A = \big\{(x; y)\colon x + y < 4\big\},$$

или

$$A = \{y < 4 - x\}.$$

Итак, элементарные события испытания, которые благоприятствуют событию A, образуют фигуру (заштрихованную на рис. 1), площадь которой

$$m(A) = \frac{1}{2} \cdot 4 \cdot 4 = 8.$$

Применяя формулу (2) для геометрической вероятности, получаем

$$\mathbf{P}(A) = \frac{m(A)}{m(\Omega)} = \frac{8}{25}.$$

б) Пусть событие $B = \{$произведение двух чисел больше 5$\}$. Тогда

$$B = \big\{(x; y)\colon xy > 5\big\},$$

или

$$B = \left\{(x; y)\colon y > \frac{5}{x}\right\}.$$

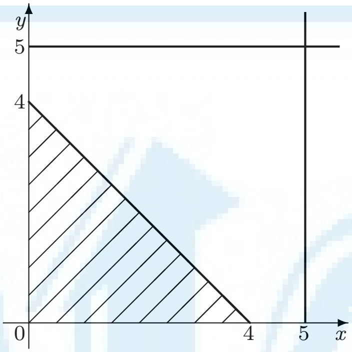


Рис. 1

Множеству всех событий, которым благоприятствует событие B, отвечает фигура ABC, заштрихованная на рис. 2, где линия AC — график функции $y = \dfrac{5}{x}$. Вычислим площадь этой фигуры с помощью определённого интеграла:

$$m(B) = \int\limits_1^5 \left(5 - \frac{5}{x}\right) dx = (5x - 5\ln x)\Big|_1^5 = 20 - 5\ln 5.$$

Итак,

$$\mathbf{P}(B) = \frac{m(B)}{m(\Omega)} = \frac{20 - 5\ln 5}{25} \approx 0{,}478.$$

в) Пусть событие $C = \{$разность двух чисел меньше 2, а их произведение больше 3$\}$, тогда

$$C = \big\{(x; y)\colon |x - y| < 2, xy > 3\big\},$$

откуда получаем неравенства

$$y < x + 2, \qquad y > x - 2, \qquad y > \frac{x}{3}.$$

Множество точек, координаты которых удовлетворяют указанным неравенствам, образует фигуру $ABCDE$ (рис. 3). Прямая AE

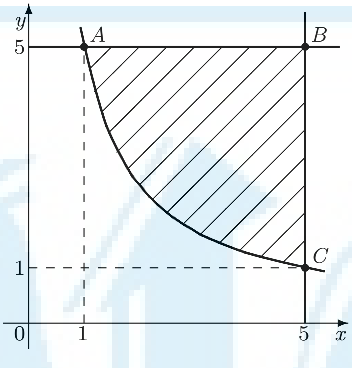


Рис. 2

задана уравнением $y = x + 2$, прямая CD — уравнением $y = x - 2$, а линия $MEDN$ — графиком функции $y = \dfrac{3}{x}$. Абсциссу точки M определяем из равенства $\dfrac{3}{x} = 5$, т. е. $x = \dfrac{3}{5}$. Ординату точки N находим из равенства $y = \dfrac{3}{5}$. Абсциссу точки E как точки пересечения двух линий $y = \dfrac{3}{x}$ и $y = x+2$ определяем из уравнения $x+2 = \dfrac{3}{x}$, откуда имеем $x = 1$. Аналогично находим абсциссу точки D как точки пересечения двух линий $y = x - 2$ и $y = \dfrac{3}{x}$, т. е. $x = 3$.

Для определения вероятности события C вычислим площадь области $ABCDE$. Очевидно, что

$$S_{ABCDE} = S_{MBN} - S_{MAE} - S_{CND} = S_{MBN} - 2S_{CND},$$

поскольку $S_{MAE} = S_{CND}$. Тогда

$$S_{MBN} = \int\limits_{\frac{3}{5}}^{5} \left(5 - \frac{3}{x}\right) dx = (5x - 3\ln x)\Big|_{\frac{3}{5}}^{5} =$$

$$= 25 - 3\ln 5 - \left(3 - 3\ln\frac{3}{5}\right) = 22 + 3\ln 3 - 6\ln 5;$$

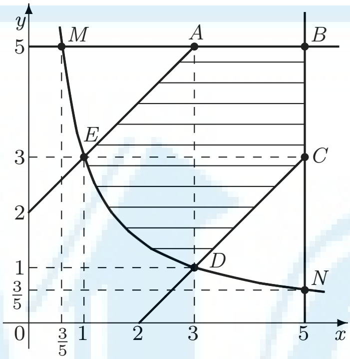


Рис. 3

$$S_{CND} = \int_3^5 \left(x - 2 - \frac{3}{x}\right) dx = \left(\frac{x^2}{2} - 2x - 3\ln x\right)\Big|_3^5 =$$

$$= \frac{25}{2} - 10 - 3\ln 5 - \left(\frac{9}{2} - 6 - 3\ln 3\right) = 4 + 3\ln 3 - 3\ln 5.$$

Итак,

$$m(C) = 22 + 3\ln 3 - 6\ln 5 - 8 - 6\ln 3 + 6\ln 5 = 14 - 3\ln 3,$$

а

$$\mathbf{P}(C) = \frac{14 - 3\ln 3}{25} \approx 0{,}428. \qquad \square$$

Пример 2. Два студента назначили встречу в определённом месте между тремя и четырьмя часами дня. Тот, кто придёт первым, ждёт другого в течение 15 мин, после чего покидает место встречи. Найти вероятность того, что встреча состоится.

Решение. Обозначим через x время прихода на место встречи первого студента, а через y — второго. Предположим также, что время, когда может состояться встреча, несущественно, т. е. студенты могут встретиться на протяжении одного часа. Тогда для x и y выполняются условия

$$0 \leqslant x \leqslant 1 \qquad \text{и} \qquad 0 \leqslant y \leqslant 1.$$

Пусть событие A состоит в том, что встреча состоялась. Это возможно лишь тогда, когда разность между временем прихода на место встречи первого и второго студентов меньше 15 мин, или $\dfrac{1}{4}$ ч, т. е.

$$|x - y| \leqslant \frac{1}{4}.$$

Отсюда получаем неравенства

$$y \leqslant x + \frac{1}{4} \qquad \text{и} \qquad y \geqslant x - \frac{1}{4}.$$

Множество точек, координаты которых удовлетворяют этим неравенствам, образует фигуру $ABCDOE$, изображённую на рис. 4.

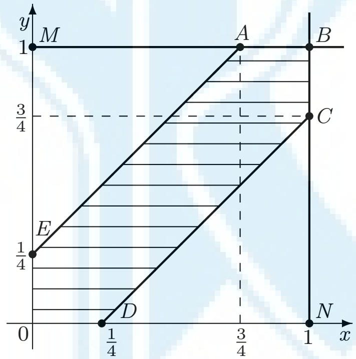


Рис. 4

Поскольку

$$S_{ABCDOE} = S_{MBNO} - 2S_{EMA},$$

причём

$$S_{MBNO} = 1, \qquad S_{EMA} = \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16},$$

то

$$S_{ABCDOE} = 1 - \frac{9}{16} = \frac{7}{16}.$$

Итак, вероятность того, что встреча состоится, $\mathbf{P}(A) = \dfrac{7}{16}$. $\square$

1.5. Статистическое определение вероятности

Поскольку классическое определение вероятности предусматривает, что все элементарные исходы испытания равновозможны, что трудно обосновать, то рассматривают ещё и *статистическое определение вероятности.*

Относительной частотой события A называют отношение количества испытаний, в которых событие A состоялось, к количеству всех проведенных испытаний. Относительную частоту события A обозначают $W(A)$. Тогда

$$W(A) = \frac{m}{n},$$

где m — количество испытаний, в которых состоялось событие A; n — количество всех проведенных испытаний.

Число, вокруг которого группируется значение частоты события A при большом количестве испытаний, называют вероятностью события A:

$$\mathbf{P}(A) = \lim_{n\to\infty} W(A).$$

Пример 1. При проверке готовой продукции было выявлено 5 бракованных единиц товара из 200 проверенных. Найти относительную частоту бракованных единиц товара.

Решение. Пусть событие A состоит в том, что выявлена бракованная единица товара. Тогда относительная частота события A

$$W(A) = \frac{m}{n} = \frac{5}{200} = 0{,}025. \qquad \square$$

Пример 2. При стрельбе по мишени было выявлено, что относительная частота попаданий равняется 0,85. Проведено 100 выстрелов. Сколько выстрелов были точны?

Решение. Пусть событие A состоит в том, что выстрел был точным. Тогда по формуле для относительной частоты события A получаем, что количество точных выстрелов

$$m = n \cdot W(A) = 100 \cdot 0{,}85 = 85. \qquad \square$$

Задачи к разделу 1

Задача 1. Два стрелка производят по одному выстрелу в мишень. Совместимы или несовместимы события A и B, если $A = \{$первый стрелок попал в мишень$\}$, $B = \{$второй стрелок попал в мишень$\}$.

Ответ. Совместимы.

Задача 2. Проверить, образуют ли полную группу такие события:

а) $A = \{$выпадение не менее трёх очков$\}$, $B = \{$выпадение не более трёх очков$\}$ в испытании — подбрасывании игрального кубика;

б) $A = \{$один промах$\}$, $B = \{$одно попадание$\}$, $C = \{$два попадания$\}$ в испытании, которое состоит в том, что два стрелка производят по одному выстрелу в мишень;

в) $A = \{$выпадение двух аверсов$\}$, $B = \{$выпадение хотя бы одного реверса$\}$ в испытании — подбрасывании двух монет.

Являются ли противоположными события A и B в каждом из этих испытаний?

Ответ. Во всех случаях события не образуют полной группы событий и не являются противоположными.

Задача 3. Описать пространство элементарных событий для испытания, которое состоит в подбрасывании игрального кубика.

Ответ. $\Omega = \{w_1; w_2; w_3; w_4; w_5; w_6\}$, где $w_i = \{$выпадет i очков$\}$.

Задача 4. Подбрасывают дважды игральный кубик. Пусть $A_i = \{$выпадет i очков при первом подбрасывании$\}$, $B_i = \{$выпадет i очков при втором подбрасывании$\}$. Выразить через A_i, B_i такие события:

$A = \{$оба раза выпадет парное количество очков$\}$,

$B = \{$сумма очков при двух подбрасываниях равняется 6$\}$,

$C = \{$сумма очков при двух подбрасываниях больше 8$\}$,

$D = \{$оба раза выпадет одинаковое количество очков$\}$.

Ответ. $A = \overline{A}_1 \cdot \overline{A}_3 \cdot \overline{A}_5 \cdot \overline{B}_1 \cdot \overline{B}_3 \cdot \overline{B}_5$, $B = A_1B_5 + A_2B_4 + A_3B_3 + A_4B_2 + A_5B_1$, $C = A_3B_6 + A_4B_5 + A_4B_6 + A_5B_4 + A_5B_5 + A_5B_6 + A_6B_3 + A_6B_4 + A_6B_5 + A_6B_6$, $D = A_1B_1 + A_2B_2 + A_3B_3 + A_4B_4 + A_5B_5 + A_6B_6$.

Задача 5. Студент на экзамене отвечает на билет, в котором три вопроса. Пусть $A_i = \{$студент ответил на i-й вопрос$\}$. Выразить через A_i такие события:

$A = \{$студент ответил по крайней мере на два вопроса$\}$,

$B = \{$студент не ответил ни на один вопрос$\}$,

$C = \{$студент ответил только на один вопрос$\}$.

Ответ. $A = A_1 \cdot A_2 + \overline{A}_1 \cdot A_2 \cdot A_3 + A_1 \cdot \overline{A}_2 \cdot A_3$, $B = \overline{A}_1 \cdot \overline{A}_2 \cdot \overline{A}_3$, $C = A_1 \cdot \overline{A}_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot A_2 \cdot \overline{A}_3 + \overline{A}_1 \cdot \overline{A}_2 \cdot A_3$.

Задача 6. Подбрасывают три монеты. Для данного испытания записать пространство элементарных событий и разложить на элементарные такие события:

$A = \{$на двух монетах выпадет «аверс»$\}$,

$B = \{$ни на одной монете не выпадет «реверс»$\}$,

$C = \{$хотя бы на одной монете выпадет «реверс»$\}$.

Ответ. $\Omega = \{(\text{А}, \text{А}, \text{А}); (\text{А}, \text{А}, \text{Р}); (\text{А}, \text{Р}, \text{А}); (\text{А}, \text{Р}, \text{Р}); (\text{Р}, \text{А}, \text{А}); (\text{Р}, \text{А}, \text{Р}); (\text{Р}, \text{Р}, \text{А}); (\text{Р}, \text{Р}, \text{Р})\}$, $A = \{(\text{А}, \text{А}, \text{Р}); (\text{А}, \text{Р}, \text{А}); (\text{Р}, \text{А}, \text{А})\}$, $B = \{(\text{А}, \text{А}, \text{А})\}$, $C = \{(\text{А}, \text{А}, \text{Р}); (\text{А}, \text{Р}, \text{А}); (\text{А}, \text{Р}, \text{Р}); (\text{Р}, \text{А}, \text{А}); (\text{Р}, \text{А}, \text{Р}); (\text{Р}, \text{Р}, \text{А}); (\text{Р}, \text{Р}, \text{Р})\}$.

Задача 7. Стрелок производит четыре выстрела по мишени. Пусть событие $A_i = \{$попадание в мишень при i-м выстреле$\}$. Выразить через A_i такие события:

$A = \{$три попадания$\}$,

$B = \{$хотя бы один промах$\}$,

$C = \{$не более одного попадания$\}$,

$D = \{$хотя бы одно попадание$\}$.

Ответ. $A = A_1 A_2 A_3 \overline{A}_4 + A_1 A_2 \overline{A}_3 A_4 + A_1 \overline{A}_2 A_3 A_4 + \overline{A}_1 A_2 A_3 A_4$, $B = \overline{A_1 A_2 A_3 A_4}$, $C = \overline{A}_2 \overline{A}_3 \overline{A}_4 + \overline{A}_1 A_2 \overline{A}_3 \overline{A}_4 + \overline{A}_1 \overline{A}_2 A_3 \overline{A}_4 + \overline{A}_1 \overline{A}_2 \overline{A}_3 A_4$, $D = \overline{\overline{A}_1 \overline{A}_2 \overline{A}_3 \overline{A}_4}$.

Задача 8. Подбрасывают четыре раза монету. Разложить на элементарные такие события:

$A = \{$два раза выпадет «аверс»$\}$,

$B = \{$выпадет не более одного «реверса»$\}$.

Ответ. $A = \{(\text{А}, \text{А}, \text{Р}, \text{Р}); (\text{А}, \text{Р}, \text{А}, \text{Р}); (\text{А}, \text{Р}, \text{Р}, \text{А}); (\text{Р}, \text{А}, \text{А}, \text{Р}); (\text{Р}, \text{А}, \text{Р}, \text{А}); (\text{Р}, \text{Р}, \text{А}, \text{А})\}$, $B = \{(\text{А}, \text{А}, \text{А}, \text{А}); (\text{А}, \text{А}, \text{А}, \text{Р}); (\text{А}, \text{А}, \text{Р}, \text{А}); (\text{А}, \text{Р}, \text{А}, \text{А}); (\text{Р}, \text{А}, \text{А}, \text{А})\}$.

Задача 9. В классе учится 10 парней и 12 девушек. Найти вероятность того, что наугад выбранный ученик парень.

Ответ. $\mathbf{P}(\text{«выбран парень»}) = \dfrac{5}{11}$.

Задача 10. Найти вероятность того, что при двух подбрасываниях монеты:

а) хотя бы один раз выпадет «реверс»;

б) ни разу не выпадет «реверс».

Ответ. а) $\mathbf{P} = \dfrac{3}{4}$; б) $\mathbf{P} = \dfrac{1}{4}$.

Задача 11. Определить вероятность того, что при подбрасывании игрального кубика выпадет:

а) чётное количество очков;

б) нечётное количество очков;

в) количество очков больше четырёх;

г) количество очков меньше десяти;

д) количество очков больше шести.

Ответ. а) $\mathbf{P} = \dfrac{1}{2}$; б) $\mathbf{P} = \dfrac{1}{2}$; в) $\mathbf{P} = \dfrac{1}{3}$; г) $\mathbf{P} = 1$; д) $\mathbf{P} = 0$.

Задача 12. В круг наугад брошена точка. Какая вероятность того, что она попадёт в квадрат, вписанный в этот круг?

Ответ. $\mathbf{P} = \dfrac{2}{\pi}$.

Задача 13. Абонент в течение часа ждёт телефонного звонка. Какая вероятность того, что ему позвонят на протяжении первых 15 мин?

Ответ. $\mathbf{P} = \dfrac{1}{4}$.

Задача 14. Студент и студентка договорились встретиться в определённом месте между 19 и 20 ч. Если студент приходит первым, он ждет студентку 30 мин и покидает место встречи. Если первой прийдёт студентка, она ждет студента 10 мин и покидает место встречи. Найти вероятность того, что встреча состоится в условленном месте.

Ответ. $\mathbf{P} = \dfrac{19}{36}$.

Задача 15. На плоскости проведены прямые на одинаковом расстоянии 15 см. На эту плоскость бросают монету диаметром 5 см.

Найти вероятность того, что монета пересечет одну из параллельных прямых.

Ответ. $\mathbf{P} = \dfrac{1}{3}$.

Задача 16. Два действительных числа p и q случайным образом выбирают из интервала $[-2; 2]$. Найти вероятность того, что квадратное уравнение $x_2 + px + q = 0$ имеет:

а) действительные корни;

б) действительные положительные корни;

в) один корень положительный, а другой отрицательный;

г) хотя бы один действительный корень из интервала $[-1; 1]$.

Ответ. а) $\mathbf{P} = \dfrac{7}{12}$; б) $\mathbf{P} = \dfrac{1}{24}$; в) $\mathbf{P} = \dfrac{1}{2}$; г) $\mathbf{P} = \dfrac{25}{48}$.

Задача 17. При проверке готовой продукции было выявлено 7 бракованных единиц товара из 140 проверенных. Найти относительную частоту бракованных единиц товара.

Ответ. $W = 0{,}04$.

Задача 18. При стрельбе по мишени было выявлено, что относительная частота попаданий равняется 0,9. Произведено 70 выстрелов. Сколько выстрелов были точны?

Ответ. $m = 63$.

Раздел 2. Элементы комбинаторики и их применение в теории вероятностей

Ключевые слова

Вероятность	Размещение
Вычисление	Событие
Комбинаторика	Соотношение
Перестановка	Сочетание

При определении вероятностей событий довольно часто нужно подсчитывать количество элементарных событий (благоприятных для некоторого события или всех возможных событий). В большинстве случаев это сопряжено со значительными трудностями, преодолеть которые помогает комбинаторика, изучающая способы подсчёта количества размещений, перестановок, сочетаний.

Прежде чем представить детали, напомним, что выражение $n!$ читается «эн-факториал» и означает произведение всех натуральных чисел до n:

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n,$$

причём считают, что $0! = 1$.

Размещениями из n элементов по m называют множества из m элементов, выбранных из n элементов, которые могут различаться между собой как составом элементов, так и их порядком. Например, размещениями из трёх элементов по два будут такие множества: $\{1;2\}$, $\{1;3\}$, $\{2;1\}$, $\{2;3\}$, $\{3;1\}$, $\{3;2\}$. Количество всех размещений из n элементов по m определяют по формуле

$$A_n^m = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-m+1) = \frac{n!}{(n-m)!}. \tag{3}$$

Перестановками из n элементов называют множества из n элементов, которые различаются лишь их порядком. Например, перестановками из трёх элементов будут такие множества: $\{1;2;3\}$,

$\{1;3;2\}$, $\{2;1;3\}$, $\{2;3;1\}$, $\{3;1;2\}$, $\{3;2;1\}$. Количество всех перестановок из n элементов определяют так:

$$P_n = \prod_{k=1}^{n} k = 1 \cdot 2 \cdot \ldots \cdot n = n! \tag{4}$$

Сочетаниями из n элементов по m называют множества из m элементов, выбранных из n элементов, которые различаются между собой только составом элементов. Например, сочетаниями из трёх элементов по два будут такие множества: $\{1;2\}$, $\{1;3\}$, $\{2;3\}$. Количество всех сочетаний из n элементов по m определяют по формуле

$$C_n^m = \frac{(n-m+1)\cdot(n-m+2)\cdot\ldots\cdot n}{m!} = \frac{n!}{m!\cdot(n-m)!}. \tag{5}$$

Между перечисленными понятиями существуют такие соотношения:

$$P_n = A_n^n = n!; \qquad C_n^m = \frac{A_n}{P_m}; \qquad C_n^m = C_n^{n-m};$$

$$C_n^0 = C_n^n = 1; \qquad C_n^1 = n; \qquad C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^n = 2^n.$$

Пример 1. Согласно учебному плану студенты на протяжении семестра изучают 10 дисциплин. На каждый день планируются 4 пары по разным дисциплинам. Сколькими способами можно составить расписание занятий на один день?

Решение. Все возможные расписания занятий на один день — это размещения из 10 элементов по 4, поскольку они могут различаться дисциплинами или порядком. Поэтому количество способов составления расписания на один день определяется по формуле (3), т. е.

$$A_{10}^4 = 10 \cdot 9 \cdot 8 \cdot 7 = 5040. \qquad \square$$

Пример 2. Сколько пятизначных чисел можно образовать с помощью цифр 1, 2, 3, 4, 5, если любая из них в числе встречается лишь один раз?

Решение. Разные пятизначные числа, образованные с помощью цифр 1, 2, 3, 4, 5, можно получить лишь перестановкой этих цифр в числе. Поэтому их количество определяется перестановкой из пяти элементов. Согласно формуле (4)

$$P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120. \qquad \square$$

Пример 3. В группе 15 студентов. Сколькими способами можно избрать:

а) студенческий совет в количестве трёх студентов;

б) председателя, заместителя и секретаря студенческого совета?

Решение. а) В этом случае студенческий совет из трёх студентов, выбранных из 15 студентов группы, различается лишь составом. Порядок выбранных студентов не имеет значения. Поэтому количество возможных выборов определяется количеством сочетаний из 15 элементов по 3, которое согласно формуле (5) равняется

$$C_{15}^3 = \frac{15 \cdot 14 \cdot 13}{1 \cdot 2 \cdot 3} = 455.$$

б) В этом случае студенческий совет из трёх студентов, выбранных из 15 студентов группы, различается не только составом, но и тем, кто будет председателем, заместителем и секретарем. Порядок выбранных студентов имеет значение. Поэтому количество возможных выборов определяется количеством размещений из 15 элементов по 3, которое согласно формуле (3) равняется

$$A_{15}^3 = 15 \cdot 14 \cdot 13 = 2730. \qquad \square$$

Пример 4. В урне находится 8 чёрных и 5 белых шаров. Наугад вынимают 4 шара. Какая вероятность того, что вынули:

а) 4 чёрных шара;

б) 2 чёрных и 2 белых шара?

Решение. а) Количество шаров в урне — 13. Определим, сколькими способами можно получить 4 шара из 13 (без возвращения их в урну). Поскольку при этом не имеет значения порядок вынутых шаров, количество сочетаний определяем по формуле (5), т. е. количество всех элементарных событий в данном испытании

$$n = C_{13}^4 = \frac{13 \cdot 12 \cdot 11 \cdot 10}{1 \cdot 2 \cdot 3 \cdot 4} = 715.$$

Пусть событие A состоит в том, что вынуто 4 чёрных шара. Эти шары можно вынуть из урны, в которой находится восемь чёрных шаров. Поэтому количество событий, благоприятствующих событию A, определяем по формуле (5), т. е.

$$m = C_8^4 = \frac{8 \cdot 7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3 \cdot 4} = 70.$$

Итак, вероятность того, что вынули 4 чёрных шара,

$$\mathbf{P}(A) = \frac{m}{n} = \frac{70}{715} = \frac{14}{143} \approx 0{,}098.$$

б) В этой задаче количество всех возможных событий такое же, как и в предыдущем пункте: $n = 715$. Пусть событие B состоит в том, что вынуто 2 чёрных и 2 белых шара. Определим, сколькими способами это можно сделать. Два чёрных шара можно вынуть из восьми шаров, которые находятся в урне,

$$m_1 = C_8^2 = \frac{8 \cdot 7}{1 \cdot 2} = 28$$

способами, а два белых — из пяти белых, которые находятся в урне,

$$m_2 = C_5^2 = \frac{5 \cdot 4}{1 \cdot 2} = 10$$

способами. Тогда 2 белых и 2 чёрных шара можно вынуть

$$m = m_1 \cdot m_2 = 28 \cdot 10 = 280$$

способами. Поэтому количество событий, благоприятствующих событию B, $m = 280$.

Итак, вероятность события B

$$\mathbf{P}(B) = \frac{280}{715} = \frac{56}{143} \approx 0{,}39. \qquad \square$$

Пример 5. Шестнадцать вариантов контрольной работы написаны на отдельных карточках и распределяются случайным образом среди 14 студентов, которые сидят в одном ряду. Каждый студент получает одну карточку. Найти вероятность того, что:

а) варианты 1 и 2 не будут использованы;

б) варианты 1 и 2 выдадут студентам, которые сидят рядом.

Решение. Имеем испытание — распределение 16 билетов среди 14 студентов. В этом случае события отличаются друг от друга не только номерами вариантов, которые распределяются среди студентов, но и порядком распределения. Поэтому такие соединения называют размещениями, а количество таких размещений определяется по формуле (3):

$$n = A_{16}^{14} = \frac{16!}{2!}.$$

а) Обозначим через A событие, которое состоит в том, что варианты 1 и 2 останутся нераспределёнными. Тогда другие 14 билетов будут распределены среди 14 студентов. Такие соединения называют перестановками, а их количество определяют по формуле (4):

$$m = P_{14} = 14!$$

Итак, применив классическую формулу вероятности (1), получим

$$\mathbf{P}(A) = \frac{m}{n} = \frac{14!2!}{16!} = \frac{1}{15 \cdot 8} \approx 0{,}008.$$

б) Пусть событие B состоит в том, что варианты 1 и 2 выданы студентам, которые сидят рядом. В ряду из 14 мест есть 13 пар соседних мест, причём в каждой паре варианты могут быть распределены двумя способами:

$$m_1 = 13 \cdot 2 = 26.$$

Другие 14 вариантов билетов распределяются между 12 студентами

$$m_2 = A_{14}^{12} = \frac{14!}{2!}$$

способами. Поэтому событию B благоприятствуют

$$m = m_1 \cdot m_2 = 26 \cdot \frac{14!}{2!}$$

событий.

Итак, вероятность события B

$$\mathbf{P}(B) = \frac{26 \cdot 14! \cdot 2!}{2! \cdot 16!} = \frac{13}{120} \approx 0{,}108. \qquad \square$$

Пример 6. Комплект состоит из восьми разных изделий, 3 из которых сто́ят по 4 грн, ещё 3 — по 5 грн и 2 остальных — по 3 грн. Найти вероятность того, что взятые наугад 2 изделия сто́ят 7 грн.

Решение. Выбор двух изделий из восьми является сочетанием, количество сочетаний

$$n = C_8 2 = \frac{8 \cdot 7}{2} = 28.$$

Это общее количество событий в испытании — выборе двух изделий из восьми.

Пусть событие A состоит в том, что стоимость двух выбранных изделий составляет 7 грн. Это возможно лишь тогда, когда одно изделие сто́ит 4 грн, а другое — 3 грн. Поскольку изделий стоимостью 4 грн в комплекте три, а изделий стоимостью 3 грн — два, то выбрать два изделия стоимостью 7 грн можно $m = 3 \cdot 2 = 6$ способами. Итак,

$$\mathbf{P}(A) = \frac{m}{n} = \frac{6}{28} \approx 0{,}214. \qquad \square$$

Пример 7. В группе 10 парней и 5 девушек, среди которых выбирают двух студентов для участия в конференции. Какая вероятность того, что:

а) выберут двух парней;

б) выберут парня и девушку?

Решение. В группе 15 студентов, среди которых 10 парней и 5 девушек. Выбор двух студентов из 15 является сочетанием, количество сочетаний

$$n = C_{15}^2 = \frac{15 \cdot 14}{2} = 105.$$

Это общее количество событий в испытании — выборе двух студентов из 15.

а) Пусть событие $A = \{$выбрали двух парней$\}$. Общее количество событий, которые благоприятствуют событию A, определяется количеством выборов двух парней из 10:

$$m = C_{10}^2 = \frac{10 \cdot 9}{2} = 45.$$

Итак,

$$\mathbf{P}(A) = \frac{m}{n} = \frac{45}{105} \approx 0{,}43.$$

б) Пусть событие $B = \{$выбрали парня и девушку$\}$. Это возможно

$$m = 10 \cdot 5 = 50$$

способами.

Следовательно,

$$\mathbf{P}(B) = \frac{m}{n} = \frac{50}{105} \approx 0{,}48. \qquad \square$$

Задачи к разделу 2

Задача 1. В студенческий совет института избрано 8 студентов. Сколькими способами можно избрать руководящую группу в составе председателя, заместителя и секретаря?

Ответ. 336 способов.

Задача 2. В библиотеку одновременно зашли четыре посетителя. Сколькими способами они могут образовать очередь?

Ответ. 24 способа.

Задача 3. В деканате находится 10 студенческих книжек. Шестеро студентов, зашедших в деканат, наугад берут по одной студенческой книжке. Какая вероятность того, что:

а) все студенты взяли свои студенческие книжки;

б) 4 студента взяли свои студенческие книжки?

Ответ. а) $\mathbf{P} = \dfrac{1}{148200}$; б) $\mathbf{P} = \dfrac{5}{2016} \approx 0{,}0025$.

Задача 4. Из партии в 40 изделий, среди которых 20 % брака, наугад вынимают 7 изделий. Найти вероятность того, что среди вынутых окажется 5 бракованных изделий.

Ответ. $\mathbf{P} = \dfrac{704816}{2330445} \approx 0{,}3$.

Задача 5. У туриста есть десять одинаковых консервных банок, среди которых три банки — с тушеным мясом, а семь — с рыбой. Во время ливня этикетки отклеились. Какая вероятность того, что две наугад открытые банки будут различаться содержимым?

Ответ. $\mathbf{P} = \dfrac{7}{15}$.

Задача 6. В гуманитарном классе 25 учеников, из которых семь интересуются математикой, десять — экономическими дисциплинами, остальные восемь — литературой. Найти вероятность того, что два наугад выбранных ученика будут интересоваться одной дисциплиной.

Ответ. $\mathbf{P} = \dfrac{47}{150} \approx 0{,}31$.

Задача 7. В вазе 5 роз розового, 7 красного и 3 белого цвета. Наугад вынимают две розы. Какая вероятность того, что они будут:

а) одного цвета;

б) разных цветов?

Ответ. а) $\mathbf{P} = \dfrac{34}{105} \approx 0{,}32$; б) $\mathbf{P} = \dfrac{71}{105} \approx 0{,}68$.

Задача 8. В корзине лежат 8 красных и 2 зелёных яблока. Найти вероятность того, что среди четырёх взятых яблок будут:

а) все красные;

б) 2 зелёных.

Ответ. а) $\mathbf{P} = \dfrac{1}{3}$; б) $\mathbf{P} = \dfrac{2}{15}$.

Задача 9. Четыре билета в театр разыгрывают 5 парней и 7 девушек. Найти вероятность того, что в театр пойдут 2 парня и 2 девушки.

Ответ. $\mathbf{P} = \dfrac{14}{33} \approx 0{,}4$.

Задача 10. Из десяти лотерейных билетов книжной лотереи, которые находятся в продаже, два выигрышные. Определить вероятность того, что среди купленных пяти билетов:

а) один выигрышный;

б) хотя бы один выигрышный.

Ответ. а) $\mathbf{P} = \dfrac{5}{9}$; б) $\mathbf{P} = \dfrac{7}{9}$.

Задача 11. Пять книжек, среди которых 2 учебника по математике, произвольно размещают на полке. Какая вероятность того, что эти 2 учебника будут стоять рядом?

Ответ. $\mathbf{P} = \dfrac{2}{5}$.

Задача 12. В урне 9 белых и 6 чёрных шара. Из урны вынимают наугад 5 шаров. Найти вероятность того, что три из них — белые, а два — чёрные.

Ответ. $\mathbf{P} = \dfrac{60}{143} \approx 0{,}42$.

Задача 13. На девяти карточках написаны буквы «о», «с», «к», «н», «э», «о», «и», «т», «м». Найти вероятность того, что, наугад выкладывая эти карточки, вы получаете слово «экономист».

Ответ. $\mathbf{P} = \dfrac{1}{181440}$.

Задача 14. Три стрелка независимо друг от друга стреляют по мишени. Вероятность попадания в мишень для первого равняется 0,7, для второго — 0,8, для третьего — 0,9. Какая вероятность того, что:

а) хотя бы один из них попадёт в мишень;

б) только двое попадут в мишень;

в) ни один не попадёт в мишень?

Ответ. а) $\mathbf{P} = 0{,}994$; б) $\mathbf{P} = 0{,}398$; в) $\mathbf{P} = 0{,}006$.

Задача 15. Из урны, в которой находится 8 цветных и 7 белых шаров, наугад вынимают один за другим два шара. Найти вероятность того, что они цветные, если:

а) первый шар возвращают в урну, после чего вынимают второй;

б) первый шар не возвращают в урну.

Ответ. а) $\mathbf{P} = \dfrac{64}{225} \approx 0{,}28$; б) $\mathbf{P} = \dfrac{4}{15}$.

Задача 16. В соревнованиях по легкой атлетике принимают участие 6 учеников восьмого класса, 7 — девятого и 8 — десятого. Найти вероятность того, что по жеребьёвке в первую пару бегунов попадут двое учеников из одного класса.

Ответ. $\mathbf{P} = \dfrac{32}{105} \approx 0{,}3$.

Раздел 3. Формулы сложения и умножения вероятностей

Ключевые слова

Вероятность	Событие
Зависимый	Сумма
Независимый	Умножение
Произведение	Условный
Сложение	Формула

Вероятность суммы двух произвольных случайных событий A и B равняется сумме их вероятностей минус вероятность их произведения, т. е.

$$\mathbf{P}(A+B)=\mathbf{P}(A)+\mathbf{P}(B)-\mathbf{P}(A\cdot B).$$

Если события A и B несовместимы, то $\mathbf{P}(A\cdot B)=0$. Тогда

$$\mathbf{P}(A+B)=\mathbf{P}(A)+\mathbf{P}(B).$$

Если случайные события $A_1, A_2, \ldots, A_n$ попарно несовместимы, то вероятность появления хотя бы одного из этих событий равняется сумме их вероятностей:

$$\mathbf{P}(A_1+A_2+\ldots+A_n)=\mathbf{P}(A_1)+\mathbf{P}(A_1)+\ldots+\mathbf{P}(A_n).$$

Сумма вероятностей случайных событий $A_1, A_2, \ldots, A_n$, образующих полную группу, равняется единице, т. е.

$$\mathbf{P}(A_1)+\mathbf{P}(A_2)+\ldots+\mathbf{P}(A_n)=1.$$

В частности, для противоположных событий A и $\overline{A}$ выполняется равенство

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

Случайные события A и B называются *зависимыми*, если вероятность появления одного из них зависит от появления или непоявления другого. В противном случае события A и B называются *независимыми*.

Вероятность события B, рассчитанную при условии появления события A, называют *условной вероятностью* события B (при условии появления события A) и обозначают $\mathbf{P}(B/A)$ или $\mathbf{P}_A(B)$.

Если события A и B независимы, то $\mathbf{P}(B/A) = \mathbf{P}(B)$, и наоборот, $\mathbf{P}(A/B) = \mathbf{P}(A)$.

Вероятность произведения двух случайных событий A и B равняется произведению вероятности одного из этих событий на условную вероятность другого события при условии, что первое событие состоялось, т. е.

$$\mathbf{P}(A \cdot B) = \mathbf{P}(A) \cdot \mathbf{P}(B/A) = \mathbf{P}(B) \cdot \mathbf{P}(A/B).$$

Если события независимы, то

$$\mathbf{P}(A \cdot B) = \mathbf{P}(A) \cdot \mathbf{P}(B).$$

Пусть есть n независимых случайных событий $A_1, A_2, \ldots, A_n$. Вероятность появления хотя бы одного из этих событий определяется по формуле

$$\mathbf{P}(A_1 + A_2 + \ldots + A_n) = 1 - \mathbf{P}(\overline{A}_1) \cdot \mathbf{P}(\overline{A}_2) \cdot \ldots \cdot \mathbf{P}(\overline{A}_n).$$

Пример 1. Мишень состоит из трех областей. Вероятность попадания стрелком в первую область мишени равняется 0,45, во вторую — 0,35, в третью — 0,15. Какая вероятность того, что при одном выстреле стрелок:

а) попадёт в первую или во вторую область;

б) не попадёт в мишень?

Решение. Введем обозначения:

событие $A_1 = \{$попадание в первую область$\}$;

событие $A_2 = \{$попадание во вторую область$\}$;

событие $A_3 = \{$попадание в третью область$\}$;

событие $A_4 = \{$непопадание в мишень$\}$.

Тогда

$$\mathbf{P}(A_1) = 0{,}45; \qquad \mathbf{P}(A_2) = 0{,}35; \qquad \mathbf{P}(A_3) = 0{,}15.$$

а) При одном выстреле события A_1, A_2, A_3 несовместимы. Поэтому

$$\mathbf{P}(A_1 + A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) = 0{,}45 + 0{,}35 = 0{,}8.$$

б) События A_1, A_2, A_3, A_4 попарно несовместимы и образуют полную группу случайных событий. Поэтому

$$\mathbf{P}(A_1) + \mathbf{P}(A_2) + \mathbf{P}(A_3) + \mathbf{P}(A_4) = 1.$$

Итак,

$$\mathbf{P}(A_4) = 1 - \mathbf{P}(A_1) - \mathbf{P}(A_2) - \mathbf{P}(A_3) =$$
$$= 1 - 0{,}45 - 0{,}35 - 0{,}15 = 0{,}05. \qquad \square$$

Пример 2. Для выполнения работы заказчик обратился к двум исполнителям. Вероятность того, что первый исполнитель выполнит заказ, равняется 0,8, а второй — 0,9. Найти вероятность того, что заказ будет выполнен.

Решение. Обозначим событие A = {заказ будет выполнен}; A_1 = {заказ выполнит первый исполнитель}; A_2 = {заказ выполнит второй исполнитель}. Тогда

$$A = A_1 + A_2.$$

Поскольку выполнение заказа первым исполнителем не исключает выполнения этого же заказа вторым, события совместимы. Тогда

$$\mathbf{P}(A) = \mathbf{P}(A_1 + A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cdot A_2).$$

Однако события A_1 и A_2 независимы, так как вероятность выполнения заказа первым исполнителем не зависит от того, выполнит ли этот заказ второй исполнитель. Поэтому

$$\mathbf{P}(A_1 \cdot A_2) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2).$$

Итак,

$$\mathbf{P}(A) = 0{,}8 + 0{,}9 - 0{,}8 \cdot 0{,}9 = 0{,}98. \qquad \square$$

Задачи к разделу 3

Задача 1. В первой урне 7 белых и 3 чёрных шара, во второй — 5 белых и 5 чёрных шаров, в третьей — 4 белых и 6 чёрных шаров. Из каждой урны наугад вынимают по одному шару. Найти вероятность того, что среди выбранных шаров окажется:

а) лишь один белый;

б) два белых;

в) три белых;

г) хотя бы один белый.

Ответ. а) $\mathbf{P} = 0{,}36$; б) $\mathbf{P} = 0{,}41$; в) $\mathbf{P} = 0{,}14$; г) $\mathbf{P} = 0{,}91$.

Задача 2. Найти вероятность того, что случайным образом выбранное двузначное число делится:

а) либо на 2, либо на 3;

б) на 2 и на 3.

Ответ. а) $\mathbf{P} = \dfrac{2}{3}$; б) $\mathbf{P} = \dfrac{1}{6}$.

Задача 3. Группу из 30 студентов, среди которых 10 отличников, случайным образом разбивают на две равные подгруппы. Найти вероятность того, что в каждой подгруппе будет по 5 отличников.

Ответ. $\mathbf{P} = \dfrac{1001}{3335} \approx 0{,}3$.

Задача 4. В группе 8 студентов мужского пола и 4 женского. Случайным образом группу разбивают на четыре равные части. Найти вероятность того, что в каждой части окажется одна студентка.

Ответ. $\mathbf{P} = \dfrac{9}{55} \approx 0{,}16$.

Задача 5. Комплект из 20 изделий содержит 40 % нестандартных. Дважды наугад из комплекта вынимают по 6 изделий. Найти вероятность того, что после этих двух выниманий в комплекте останутся лишь нестандартные изделия.

Ответ. $\mathbf{P} = \dfrac{1}{38760}$.

Задача 6. Первый стрелок может попасть в цель с вероятностью 0,8, второй — с вероятностью 0,9, а третий — с вероятностью 0,85. Какая вероятность того, что хотя бы один стрелок попадёт в цель?

Ответ. $\mathbf{P} = 0{,}997$.

Задача 7. Рабочий обслуживает одновременно три станка. Вероятность нарушения работы в течение часа для первого станка равняется 0,1, для второго — 0,15, для третьего — 0,2. Какая вероятность того, что:

а) все три станка будут работать в течение часа;

б) хотя бы один из них выйдет из строя?

Ответ. а) $\mathbf{P} = 0{,}712$; б) $\mathbf{P} = 0{,}329$.

Раздел 4. Формула полной вероятности. Формула Байеса

Ключевые слова

Вероятность	Полный
Гипотеза	Случайный
Группа	Событие
Несовместимый	Формула

Если случайное событие A может произойти только совместно с одним из несовместимых между собою событий $B_1, B_2, \ldots, B_n$, образующих полную группу, то вероятность появления события A определяется по *формуле полной вероятности*

$$\mathbf{P}(A) = \sum_{k=1}^{n} \mathbf{P}(B_k) \cdot \mathbf{P}(A/B_k). \tag{6}$$

В условиях теоремы неизвестно, с каким из несовместимых событий $B_1, B_2, \ldots, B_n$ состоится событие A. Поэтому появление любого из этих событий можно считать гипотезой, а $\mathbf{P}(B_k)$ — вероятностью k-й гипотезы.

Если в результате проведенного испытания состоялось событие A, то условная вероятность $\mathbf{P}(B_k/A)$ может не равняться $\mathbf{P}(B_k)$. Чтобы получить условную вероятность, используют *формулу Байеса*:

$$\mathbf{P}(B_k/A) = \frac{\mathbf{P}(B_k) \cdot \mathbf{P}(A/B_k)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B_k) \cdot \mathbf{P}(A/B_k)}{\sum\limits_{k=1}^{n} \mathbf{P}(B_k) \cdot \mathbf{P}(A/B_k)}. \tag{7}$$

Пример 1. В первом ящике 30 деталей, из которых 20 стандартных. Во втором — 15 деталей, из которых 10 стандартных. Из второго ящика берут наугад одну деталь и перекладывают её в первый ящик.

а) Найти вероятность того, что наугад взятая после этого деталь из первого ящика стандартная.

б) Пусть известно, что из первого ящика взята стандартная деталь. Найти вероятность того, что в первый ящик переложили:

- стандартную деталь;
- нестандартную деталь.

Решение. Введем такие события: $A = \{$из первого ящика взята стандартная деталь$\}$; $B_1 = \{$из второго ящика переложили в первый стандартную деталь$\}$; $B_2 = \{$из второго ящика переложили в первый нестандартную деталь$\}$.

а) По условию задачи из первого ящика деталь берут лишь после того, как в него положат деталь, взятую из второго ящика, т. е. событие A состоится лишь тогда, когда состоится событие B_1 или событие B_2. Эти события несовместимы, а событие A может состояться только совместно с одним из них. Поэтому для определения вероятности события A можно воспользоваться формулой полной вероятности (6):

$$\mathbf{P}(A) = \mathbf{P}(B_1) \cdot \mathbf{P}(A/B_1) + \mathbf{P}(B_2) \cdot \mathbf{P}(A/B_2).$$

Найдем нужные вероятности:

$$\mathbf{P}(B_1) = \frac{10}{15} = \frac{2}{3}; \qquad \mathbf{P}(B_2) = \frac{5}{15} = \frac{1}{3};$$
$$\mathbf{P}(A/B_1) = \frac{21}{31}; \qquad \mathbf{P}(A/B_2) = \frac{20}{31}.$$

Итак,

$$\mathbf{P}(A) = \frac{2}{3} \cdot \frac{21}{31} + \frac{1}{3} \cdot \frac{20}{31} = \frac{62}{93} \approx 0{,}67.$$

б) Если событие A уже состоялось, то условная вероятность того, что в первый ящик переложена стандартная деталь,

$$\mathbf{P}(B_1/A) = \frac{\mathbf{P}(B_1) \cdot \mathbf{P}(A/B_1)}{\mathbf{P}(A)} = \frac{\frac{2}{3} \cdot \frac{21}{31}}{\frac{62}{93}} = \frac{21}{31},$$

а если деталь нестандартная, то

$$\mathbf{P}(B_2/A) = \frac{\mathbf{P}(B_2)\cdot\mathbf{P}(A/B_2)}{\mathbf{P}(A)} = \frac{\frac{1}{3}\cdot\frac{20}{31}}{\frac{62}{93}} = \frac{10}{31}. \qquad \square$$

Пример 2. В урне 10 шаров — 3 белых и 7 чёрных. Наугад вынимают два шара. Найти вероятность того, что второй шар оказался чёрным, если:

а) первый шар возвращали в урну;

б) первый шар не возвращали в урну.

Решение. Пусть событие A_i состоит в том, что в i-й раз взят белый шар, а событие B_i — в i-й раз взят чёрный шар.

а) Если шар, взятый первым, возвращают в урну, то вероятность появления второго чёрного шара не зависит от того, какой шар взят первым. Поэтому

$$\mathbf{P}(B_2) = \frac{7}{10} = 0{,}7.$$

б) Если первый шар не возвращать, то вероятность появления второго чёрного шара уже зависит от того, какой шар был взят первым. Если первым взяли белый шар, то в урне осталось 2 белых и 7 чёрных шаров. Поэтому

$$\mathbf{P}(B_2/A_1) = \frac{7}{9}.$$

Если первым взяли чёрный шар, то

$$\mathbf{P}(B_2/B_1) = \frac{6}{9} = \frac{2}{3}.$$

Вероятность того, что первым взяли белый шар,

$$\mathbf{P}(A_1) = \frac{3}{10} = 0{,}3,$$

чёрный шар —

$$\mathbf{P}(B_1) = \frac{7}{10} = 0{,}7.$$

Поскольку события A_1 и B_1 несовместимы и образуют полную группу, то по формуле полной вероятности (6) получаем

$$\mathbf{P}(B_2) = \mathbf{P}(A_1) \cdot \mathbf{P}(B_2/A_1) + \mathbf{P}(B_1) \cdot \mathbf{P}(B_2/B_1) = \\ = 0{,}3 \cdot \frac{7}{9} + 0{,}7 \cdot \frac{2}{3} = 0{,}7.$$

□

Пример 3. Детали, изготовленные цехом завода, попадают для проверки их стандартности к одному из двух контролёров. Вероятность того, что деталь попадёт к первому контролёру, равняется 0,6, ко второму — 0,4. Вероятность того, что годная деталь будет признана стандартной первым контролёром, равняется 0,94, вторым — 0,98. Годная деталь при проверке признана стандартной. Найти вероятность того, что деталь проверял:

- первый контролёр;
- второй контролёр.

Решение. Обозначим такие события:

$A = \{$годная деталь признана стандартной$\}$;

$B_1 = \{$деталь проверял первый контролёр$\}$;

$B_2 = \{$деталь проверял второй контролёр$\}$.

Тогда в соответствии с условием задачи

$$\mathbf{P}(B_1) = 0{,}6; \qquad \mathbf{P}(B_2) = 0{,}4;$$
$$\mathbf{P}(A/B_1) = 0{,}94; \qquad \mathbf{P}(A/B_2) = 0{,}98.$$

Согласно формуле Байеса (7) имеем

$$\mathbf{P}(B_1/A) = \frac{\mathbf{P}(B_1) \cdot \mathbf{P}(A/B_1)}{\mathbf{P}(A)} = \frac{0{,}6 \cdot 0{,}94}{0{,}6 \cdot 0{,}94 + 0{,}4 \cdot 0{,}98} = \\ = \frac{141}{239} \approx 0{,}59;$$

$$\mathbf{P}(B_2/A) = \frac{\mathbf{P}(B_2) \cdot \mathbf{P}(A/B_2)}{\mathbf{P}(A)} = \frac{0{,}4 \cdot 0{,}98}{0{,}6 \cdot 0{,}94 + 0{,}4 \cdot 0{,}98} = \\ = \frac{98}{239} \approx 0{,}41.$$

□

Задачи к разделу 4

Задача 1. На двух автоматических линиях изготавливают одинаковые детали: на первой — 30 %, на второй — 70 %. Вероятность изготовления стандартной детали на первой линии равняется 0,9, на второй — 0,5. Все изготовленные на этих линиях детали поступают на склад.

а) Найти вероятность того, что наугад выбранная со склада деталь стандартная.

б) Наугад выбранная деталь, изготовленная на одной из линий, оказалась стандартной. Найти вероятность того, что она изготовлена на первой линии.

Ответ. а) $\mathbf{P} = 0{,}62$; б) $\mathbf{P} = \dfrac{27}{62} \approx 0{,}4$.

Задача 2. В первой урне 4 белых и 3 чёрных шара, а во второй — 3 белых и один чёрный шар. Из первой урны наугад вынули один шар и переложили его во вторую урну. Найти вероятность того, что после перекладывания наугад вынутый из второй урны шар будет белым.

Ответ. $\mathbf{P} = \dfrac{5}{7}$.

Задача 3. В урне 8 шаров — 5 белых и 3 чёрных. Наугад вынимают два шара. Найти вероятность того, что второй шар оказался чёрным, если:

а) первый шар возвращали в урну;

б) первый шар не возвращали в урну.

Ответ. а) $\mathbf{P} = \dfrac{3}{8}$; б) $\mathbf{P} = \dfrac{3}{8}$.

Задача 4. Из 30 студентов группы, которые пришли на экзамен, 8 подготовлены отлично, 10 — хорошо, 8 — удовлетворительно, а остальные — неудовлетворительно. Программа экзамена включает 50 вопросов. Билет содержит 3 вопроса. Студент, подготовленный отлично, знает все вопросы; хорошо — 40 вопросов; удовлетворительно — 25 вопросов и неудовлетворительно — 10 вопросов.

а) Найти вероятность того, что наугад вызванный студент ответит на все 3 вопроса билета.

б) Студент ответил на все вопросы. Найти вероятность того, что студент подготовлен:

- хорошо;
- удовлетворительно;
- неудовлетворительно.

Ответ. а) $\mathbf{P} = \dfrac{691}{1470} \approx 0{,}47$; б) $\mathbf{P}\{\text{хорошо}\} = \dfrac{247}{691} \approx 0{,}36$; $\mathbf{P}\{\text{удовлетворительно}\} = \dfrac{46}{691} \approx 0{,}07$; $\mathbf{P}\{\text{неудовлетворительно}\} = \dfrac{6}{691} \approx 0{,}01$.

Задача 5. В центр статистических исследований поступает информация из трёх пунктов: первого — 50 %, второго — 30 %, третьего — 20 % всей информации. Вероятность допущения ошибки при обработке статистических данных в первом пункте равняется 0,1, во втором — 0,05, в третьем — 0,15. Какая вероятность того, что полученная центром в данный момент информация правильна?

Ответ. $\mathbf{P} = 0{,}905$.

Задача 6. В одном классе 5 отличников, в втором — 3 отличника, а в третьем классе отличников нет. Из наугад взятого класса выбрали ученика. Найти вероятность того, что он отличник, если в каждом классе по 30 учеников.

Ответ. $\mathbf{P} = \dfrac{4}{45} \approx 0{,}09$.

Задача 7. Два экономиста заполняют документы, которые складывают в общую папку. Вероятность допустить ошибку для первого экономиста равняется 0,1, для второго — 0,2. Первый экономист заполнил 40 документов, второй — 60. Во время проверки наугад взятый из папки документ оказался с ошибкой. Найти вероятность того, что его заполнил первый экономист.

Ответ. $\mathbf{P} = 0{,}25$.

Задача 8. Рабочий обслуживает три станка. Вероятность изготовления бракованной детали на первом станке равняется 0,04, на втором — 0,01, на третьем — 0,07. Производительность работы первого станка втрое превышает производительность второго, а производительность третьего вдвое меньше производительности второго.

а) Найти вероятность того, что наугад выбранная деталь, изготовленная рабочим, бракованная.

б) Взятая наугад деталь оказалась бракованной. На каком станке, вероятнее всего, она изготовлена?

Ответ. а) $\mathbf{P} = \dfrac{11}{300} \approx 0{,}037$; б) на первом станке.

Задача 9. Вероятности того, что во время работы компьютера произойдет сбой в арифметическом устройстве, в оперативной памяти или в устройстве ввода соотносятся как $2 : 1 : 3$. Вероятность обнаружить сбой в этих устройствах равна соответственно 0,9; 0,75; 0,7. Найти вероятность обнаружения сбоя в работе компьютера.

Ответ. $\mathbf{P} = 0{,}775$.

Задача 10. В корзине 20 яблок зелёного и 15 красного цвета. Наугад берут три яблока из корзины и кладут вместо них три яблока красного цвета. Потом наугад вынимают одно яблоко. Какая вероятность того, что вынуто яблоко красного цвета?

Ответ. $\mathbf{P} \approx 0{,}4776$.

Раздел 5. Независимые повторные испытания. Формула Бернулли

Ключевые слова

Интегральный
Локальный
Предельный
Приближённый
Схема
Теорема
Формула
Функция

5.1. Формула Бернулли

Пусть проводится n независимых испытаний, причём вероятность появления события A в каждом испытании одна и та же, т. е. не зависит от его появления в других испытаниях. Такую серию повторных независимых испытаний называют *схемой Бернулли*. Примером повторных независимых испытаний является подбрасывание монеты определённое количество раз, стрельба по мишени и т. п.

Пусть случайное событие A может произойти в каждом испытании с одинаковой вероятностью $\mathbf{P}(A) = p$ или не произойти с вероятностью $q = 1 - p$. Вероятность того, что при n испытаниях событие A появится m раз, определяется по *формуле Бернулли*:

$$\mathbf{P}_n(m) = C_n^m \cdot p^m \cdot q^{n-m}. \tag{8}$$

Вероятность появления события A в n испытаниях m раз, где число m находится между числами k_1 и k_2, $0 \leqslant k_1 \leqslant k_2 \leqslant n$, определяют по формуле

$$\mathbf{P}_n(k_1 \leqslant m \leqslant k_2) = \mathbf{P}_n(k_1) + \mathbf{P}_n(k_1 + 1) + \ldots + \mathbf{P}_n(k_2).$$

Вероятность появления события A в n испытаниях хотя бы один раз

$$\mathbf{P}_n(1 \leqslant m \leqslant n) = 1 - q^n.$$

Наиболее вероятное количество m_0 появления события A в n испытаниях определяется из неравенств

$$n \cdot p - q \leqslant m_0 \leqslant n \cdot p + p.$$

Если вероятность появления события A в каждом испытании равняется p, то количество n испытаний, которые необходимо осуществить, чтобы с вероятностью $\mathbf{P}$ можно было утверждать, что событие A состоится хотя бы один раз, вычисляют по формуле

$$n > \frac{\ln(1-\mathbf{P})}{\ln(1-p)}.$$

Пример 1. Прибор составлен из 10 блоков, надежность каждого из них 0,8. Блоки могут выходить из строя независимо друг от друга. Найти вероятность того, что:

а) откажут два блока;

б) откажет хотя бы один блок;

в) откажут не менее двух блоков.

Найти наиболее вероятное количество блоков, которые выйдут из строя.

Решение. Пусть событие A — отказ блока. Тогда

$$\mathbf{P}(A) = p = 1 - 0{,}8 = 0{,}2.$$

Поэтому $q = 0{,}8$. По условию задачи $n = 10$. Применив формулу Бернулли (8) и следствия из нее, получим:

а) $\mathbf{P}_{10}(2) = C_{10}^2 p^2 q^8 = 45 \cdot (0{,}2)^2 \cdot (0{,}8)^8 \approx 0{,}302;$

б) $\mathbf{P}_{10}(1 \leqslant m \leqslant 10) = 1 - q^{10} = 1 - (0{,}8)^{10} \approx 0{,}893;$

в) $\mathbf{P}_{10}(2 \leqslant m \leqslant 10) = 1 - [\mathbf{P}_{10}(0) + \mathbf{P}_{10}(1)] =$

$= 1 - C_{10}^0 \cdot (0{,}2)^0 \cdot (0{,}8)^{10} - C_{10}^1 \cdot (0{,}2)^1 \cdot (0{,}8)^9 \approx$

$\approx 0{,}624.$

Наиболее вероятное количество блоков, которые выйдут из строя, найдём из неравенств

$$n \cdot p - q \leqslant m_0 \leqslant n \cdot p + p,$$

т. е.

$$10 \cdot 0{,}2 - 0{,}8 \leqslant m_0 \leqslant 10 \cdot 0{,}2 + 0{,}2.$$

Итак,

$$m_0 = 2. \qquad \square$$

Пример 2. Оператор за минуту отправляет 3 сообщения. Через сколько минут вероятность отправления оператором хотя бы одного ошибочного сообщения будет не менее 0,952, если вероятность ошибки оператора равняется 0,02?

Решение. Пусть событие A — ошибка оператора. Определим, сколько сообщений может отправить оператор, чтобы с вероятностью 0,952 можно было утверждать, что он сделает хотя бы одну ошибку, если вероятность ошибки оператора по условию $p = 0{,}02$:

$$n > \frac{\ln(1 - 0{,}952)}{\ln(1 - 0{,}02)} > 150.$$

Итак, оператор может отправить более 150 сообщений. $\square$

5.2. Предельные теоремы в схеме Бернулли

Для приближённого вычисления вероятности появления события A в n независимых испытаниях схемы Бернулли m раз при больших значениях n и маленьких p таких, что $np < 10$, целесообразно использовать *формулу Пуассона*:

$$\mathbf{P}_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}, \qquad \lambda = np.$$

Значения функции $\mathbf{P}_n(m)$ в формуле Пуассона приведены в прил. 3.

Локальной функцией Лапласа называют функцию

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

Интегральной функцией Лапласа называют функцию

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt.$$

Значения функций $\varphi(x)$ и $\Phi(x)$ приведены соответственно в прил. 2 и 1.

Локальная теорема Муавра—Лапласа. Если в схеме Бернулли количество испытаний n достаточно большое, а вероятность p появления события A во всех испытаниях одинаковая и отличная от нуля и единицы, то вероятность появления m раз события A можно найти приближённо по формуле

$$\mathbf{P}_n(m) = \frac{1}{\sqrt{npq}}\varphi(x),$$

где $x = \dfrac{m - np}{\sqrt{npq}}$; $\varphi(x)$ — локальная функция Лапласа.

Интегральная теорема Муавра—Лапласа. Если в схеме Бернулли количество испытаний n достаточно большое, а вероятность p появления события A во всех испытаниях одинаковая и отличная от нуля и единицы, то вероятность появления события A не менее m_1 и не более m_2 раз можно найти по приближённой формуле

$$\mathbf{P}_n(m_1 \leqslant m \leqslant m_2) = \Phi(x_2) - \Phi(x_1),$$

де $x_2 = \dfrac{m_2 - np}{\sqrt{npq}}$, $x_1 = \dfrac{m_1 - np}{\sqrt{npq}}$, $\Phi(x)$ — интегральная функция Лапласа.

Теорема Бернулли. Если в n независимых испытаниях вероятность p появления события A одинаковая и событие A произойдет m раз, то для любого положительного числа $\varepsilon > 0$ выполняется равенство

$$\lim_{n\to\infty} \mathbf{P}\left(\left|\frac{m}{n} - p\right| \geqslant \varepsilon\right) = 0,$$

т. е. событие, для которого отклонение определяется формулой

$$\left|\frac{m}{n} - p\right| \geqslant \varepsilon$$

при больших значениях n, почти невозможно. Поэтому противоположное событие

$$\left|\frac{m}{n} - p\right| < \varepsilon$$

почти достоверно:

$$\lim_{n\to\infty} \mathbf{P}\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1.$$

Если $m_1 = np - \varepsilon n$, $m_2 = np + \varepsilon n$, то

$$x_2 = \frac{np + \varepsilon n - np}{\sqrt{npq}} = \varepsilon\sqrt{\frac{n}{pq}}; \qquad x_1 = \frac{np - \varepsilon n - np}{\sqrt{npq}} = -\varepsilon\sqrt{\frac{n}{pq}},$$

поэтому

$$\mathbf{P}\left(\left|\frac{m}{n} - p\right| \leqslant \varepsilon\right) = \mathbf{P}(np - \varepsilon n \leqslant m \leqslant np + \varepsilon n) = 2\Phi(\varepsilon)\sqrt{\frac{n}{pq}}.$$

Эти формулы целесообразно применять при условии $n > 100$, $npq > 20$.

Пример 1. Учебник напечатан тиражом 90000 экземпляров. Вероятность неправильного брошюрования учебника равняется 0,0001. Найти вероятность того, что тираж содержит 5 бракованных учебников.

Решение. Брошюрование учебника можно рассматривать как испытание, которое вкладывается в схему Бернулли. Количество испытаний n большое, а вероятность каждого испытания p незначительная. Поэтому в этом случае целесообразно применить формулу Пуассона.

В соответствии с условием задачи

$$n = 90000, \qquad p = 0{,}0001.$$

Итак, при $\lambda = np = 9$ из прил. 3 получаем

$$\mathbf{P}_{90000}(5) = \frac{9^5}{5!} \cdot e^{-9} \approx 0{,}0607. \qquad \square$$

Пример 2. Игральный кубик подбрасывают 800 раз. Какая вероятность того, что количество очков, кратное трём, появится 267 раз?

Решение. В этой задаче количество испытаний $n = 800$ довольно большое. Вероятность p того, что при подбрасывании кубика выпадет число, кратное трём, равняется $\frac{1}{3}$ и постоянная для всех испытаний, $q = 1 - p = \frac{2}{3}$. Поэтому для вычисления вероятности воспользуемся локальной теоремой Муавра—Лапласа. Для этого найдём x:

$$x = \frac{267 - 800 \cdot \frac{1}{3}}{\frac{40}{3}} = 0{,}025.$$

Итак,

$$\mathbf{P}_{800}(267) = \frac{3}{40} \cdot \varphi(0{,}025) = \frac{3}{40} \cdot 0{,}3988 \approx 0{,}03. \qquad \square$$

Пример 3. Игральный кубик подбрасывают 800 раз. Какая вероятность того, что количество очков, кратное трём, появится не менее 260 раз и не более 274 раз?

Решение. Для определения вероятности $\mathbf{P}_{800}(260 \leqslant m \leqslant 274)$ применим интегральную теорему Муавра—Лапласа. Определим x_1, x_2:

$$x_1 = \frac{m_1 - np}{\sqrt{npq}} = \frac{260 - 800 \cdot \frac{1}{3}}{\frac{40}{3}} = \frac{20}{40} = -0{,}5;$$

$$x_2 = \frac{m_2 - np}{\sqrt{npq}} = \frac{274 - 800 \cdot \frac{1}{3}}{\frac{40}{3}} = \frac{22}{40} = 0{,}55.$$

Из прил. 1 находим

$$\mathbf{P}_{800}(260 \leqslant m \leqslant 274) = \Phi(0{,}55) - \Phi(-0{,}5) = 0{,}2088 + 0{,}1915 = \\ = 0{,}4003. \qquad \square$$

Пример 4. Вероятность появления события в любом из 625 независимых испытаний равняется 0,8. Найти вероятность того, что относительная частота появления события отклоняется от вероятности по абсолютной величине не более чем на 0,04.

Решение. По условию задачи

$$n = 625; \qquad p = 0{,}8; \qquad q = 0{,}2; \qquad \varepsilon = 0{,}04.$$

Нужно найти $\mathbf{P}\left(\left|\dfrac{m}{n} - 0{,}8\right| \leqslant 0{,}04\right)$.

По теореме Бернулли имеем

$$\mathbf{P}\left(\left|\frac{m}{n} - 0{,}8\right| \leqslant 0{,}04\right) = 2\Phi\left(0{,}04\sqrt{\frac{625}{0{,}8\cdot 0{,}2}}\right) = 2\Phi(2{,}5) = \\ = 2\cdot 0{,}4938 = 0{,}9876. \qquad \square$$

Пример 5. Вероятность появления некоторого события в любом из независимых испытаний равняется 0,5. Найти количество испытаний n, при котором с вероятностью 0,7698 можно ожидать, что относительная частота появления события отклонится от его вероятности по абсолютной величине не более чем на 0,02.

Решение. По условию задачи

$$p = 0{,}5; \qquad q = 0{,}5; \qquad \varepsilon = 0{,}02.$$

Необходимо найти количество испытаний n, для которого

$$\mathbf{P}\left(\left|\frac{m}{n} - 0{,}8\right| \leqslant 0{,}02\right) = 0{,}7698.$$

По теореме Бернулли имеем

$$\mathbf{P}\left(\left|\frac{m}{n} - 0{,}8\right| \leqslant 0{,}02\right) = 2\Phi\left(0{,}02\sqrt{\frac{n}{0{,}5\cdot 0{,}5}}\right) = 0{,}7698.$$

Отсюда

$$\Phi(0{,}04\sqrt{n}) = 0{,}3843.$$

Определив с помощью прил. 1 аргумент интегральной функции Лапласа, при котором эта функция равняется 0,3843, получим уравнение

$$0{,}04\sqrt{n} = 1{,}2.$$

Итак,

$$n = \left(\frac{1{,}2}{0{,}04}\right)^2 = 900. \qquad \square$$

Задачи к разделу 5

Задача 1. Обзорную лекцию должны прослушать 100 студентов. Вероятность присутствия на этой лекции для каждого студента равняется 0,7. Найти вероятность того, что на лекцию придёт больше половины студентов.

Ответ. $\mathbf{P}_{100}(m > 50) \approx 0{,}99998$.

Задача 2. Вероятность своевременной реализации со склада одной пары обуви равняется 0,8. Найти вероятность того, что своевременно будет реализовано не менее 75 пар, если на склад завезено 100 пар обуви. Найти наиболее вероятное количество своевременно реализованных пар обуви.

Ответ. $\mathbf{P}_{100}(m \geqslant 75) \approx 0{,}8943$; $m_0 = 80$.

Задача 3. Из статистических данных известно, что вероятность заболеть гриппом во время эпидемии для каждого индивида равняется 0,1. Какая вероятность того, что из 100 проверенных лиц больными окажутся от 20 до 50?

Ответ. $\mathbf{P}_{100}(20 \leqslant m \leqslant 50) \approx 0{,}00043$.

Задача 4. В магазин зашло восемь покупателей. Вероятность того, что любой из них не уйдёт из магазина без покупки, равняется 0,4.

а) Найти вероятность того, что трое из них что-то купят.

б) Какая вероятность того, что ни один из них ничего не купит?

Ответ. а) $\mathbf{P}_8(3) \approx 0{,}2787$; б) $\mathbf{P}_8(0) \approx 0{,}0168$.

Задача 5. Найти вероятность того, что среди 100 прохожих будет не более 40 брюнетов, если около 30 % населення — брюнеты.

Ответ. $\mathbf{P}_{100}(m \leqslant 40) \approx 0{,}9855$.

Задача 6. В процессе производства вероятность дефектов в каждой партии продукции составляет 0,1. Какая вероятность того, что из десяти партий дефекты будут иметь меньше двух партий?

Ответ. $\mathbf{P}_{10}(m < 2) \approx 0{,}7361$.

Задача 7. Найти вероятность того, что среди 1000 новорожденных будет 480 девочек, если вероятность рождения мальчика равняется 0,515.

Ответ. $\mathbf{P}_{1000}(480) \approx 0{,}024$.

Задача 8. Предположим, что вероятности родиться в любой из дней года одинаковые. Найти вероятность того, что среди 500 учеников школы:

а) трое родились 8-го марта;

б) ни один не родился 1-го января.

Ответ. а) $\mathbf{P}_{500}(3) \approx 0{,}1089$; б) $\mathbf{P}_{500}(0) \approx 0{,}2541$.

Задача 9. Вероятность попадания в самолёт из винтовки при каждом выстреле равняется 0,001. Производится 3000 выстрелов. Найти вероятность того, что будет хотя бы одно попадание.

Ответ. $\mathbf{P}_{3000}(m \geqslant 1) \approx 0{,}9502$.

Задача 10. Среди пяти студентов проводится психологический тест на определение типа характера человека. Вероятность того, что для каждого из них будет правильно определён тип характера по результатам тестирования равняется 0,9. Найти вероятность того, что только для трех протестированных студентов будет правильно определён тип характера.

Ответ. $\mathbf{P}_5(3) = 0{,}0729$.

Задача 11. В полученной партии текстильных изделий 0,6 % брака. Какая вероятность при случайном отборе 1000 изделий обнаружить:

а) шесть бракованных изделий;

б) хотя бы одно бракованное изделие?

Ответ. а) $\mathbf{P}_{1000}(6) \approx 0{,}1606$; б) $\mathbf{P}_{1000}(m \geqslant 1) \approx 0{,}9975$.

Часть II
Случайные величины

Раздел 6. Дискретные случайные величины

Ключевые слова

Величина
Геометрический
Гипергеометрический
Дискретный
Закон
Многоугольник
Множество
Показательный
Равномерный
Распределение
Случайный
Счётный
Функция

Случайной величиной, связанной с данным вероятностным экспериментом, называется величина, которая при каждом проведении этого эксперимента получает определенное числовое значение, причём заранее неизвестно, какое именно. Случайные величины будем обозначать греческими буквами ξ, ψ, η, ... (прил. 10) или большими буквами латинского алфавита (прил. 9).

Множество называется *счётным*, если между его элементами и элементами множества натуральных чисел можно установить взаимно однозначное соответствие. Другими словами, счётное множество — это такое бесконечное множество, элементы которого можно пронумеровать с помощью множества натуральных чисел.

Случайная величина называется *дискретной*, если множество её возможных значений конечное или счётное.

Законом распределения случайной величины называется произвольное соответствие, которое устанавливает связь между возможными значениями случайной величины и соответствующими им вероятностями.

В случае дискретной случайной величины закон распределения удобнее всего описывать с помощью ряда распределения — таблицы, в которой приведены все возможные значения этой случайной величины и соответствующие им вероятности:

ξ	x_1	x_2	$\cdots$	x_n
$\mathbf{P}$	p_1	p_2	$\cdots$	p_n

причём $p_i = \mathbf{P}\{\xi = x_i\}$, $\sum\limits_{i=1}^{n} p_i = 1$.

Ломаная с вершинами в точках (x_i, p_i) называется *многоугольником распределения* (рис. 5).

Другой способ задания распределения случайной величины — указание её функции распределения.

Функцией распределения случайной величины ξ называется функция $F(x)$, значение которой равняется вероятности того, что случайная величина ξ принимает значение меньше x, т. е.

$$F(x) = \mathbf{P}\{\xi < x\}.$$

Функция распределения имеет такие свойства.

Свойство 1. $\lim\limits_{x\to+\infty} F(x) = 1$, $\lim\limits_{x\to-\infty} F(x) = 0$.

Свойство 2. $F(x)$ — неубывающая функция, т. е.

$$x_1 < x_2 \quad \Rightarrow \quad F(x_1) \leqslant F(x_2).$$

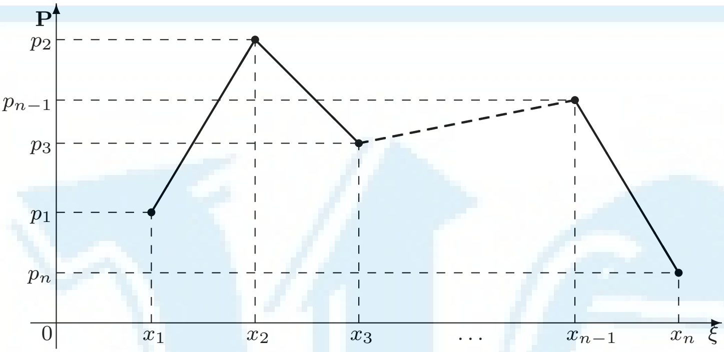


Рис. 5

Свойство 3. $F(x)$ — непрерывная слева функция, т. е. для произвольного $x \in \mathbb{R}$

$$\lim_{\varepsilon \to 0+} F(x - \varepsilon) = F(x).$$

Если случайная величина ξ дискретная, то

$$F(x) = \sum_{i:\ x_i < x} p_i, \tag{1}$$

где суммирование выполняется по всем индексам i, для которых выполняется условие $x_i < x$. Другими словами, по ряду распределения легко восстанавливается функция распределения. И наоборот, по функции распределения легко восстанавливается ряд распределения дискретной случайной величины: *множество значений случайной величины совпадает с множеством точек $\{x_i\}$ разрыва $F(x)$, а соответствующие вероятности p_i равняются величинам скачков функции распределения в точках x_i*, т. е.

$$p_i = \mathbf{P}\{\xi = x_i\} = \lim_{\varepsilon \to 0} F(x + \varepsilon) - F(x), \quad \varepsilon > 0.$$

Отметим, что с одним и тем же экспериментом может быть связано много разных случайных величин.

Пример 1. Эксперимент состоит в подбрасывании симметричного однородного игрального кубика.

Пусть случайная величина ξ равняется количеству очков, выпавших при одном подбрасывании, а случайная величина ψ — количество шестёрок при одном подбрасывании кубика.

Построить ряд распределения, многоугольник распределения, функцию распределения случайной величины:

а) ξ;

б) ψ.

Решение. а) Случайная величина ξ может принимать значения 1, 2, 3, 4, 5 и 6 с одинаковыми вероятностями $p = \dfrac{1}{6}$. Итак, ряд распределения случайной величины ξ имеет такой вид:

ξ	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

,

а многоугольник распределения изображен на рис. 6.

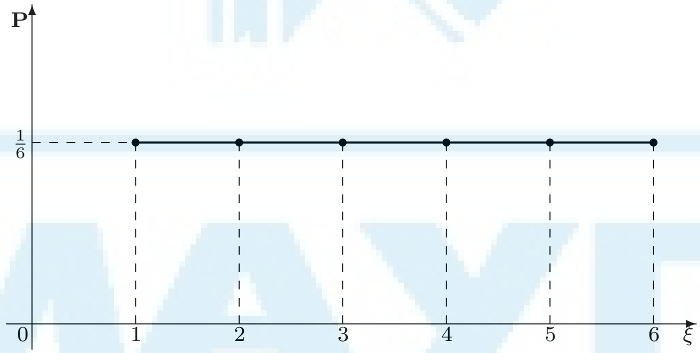


Рис. 6

Воспользовавшись формулой (1), получим

$$F(x) = \begin{cases} 0, & x \leqslant 1; \\ \dfrac{1}{6}, & 1 < x \leqslant 2; \\ \dfrac{2}{6}, & 2 < x \leqslant 3; \\ \dfrac{3}{6}, & 3 < x \leqslant 4; \\ \dfrac{4}{6}, & 4 < x \leqslant 5; \\ \dfrac{5}{6}, & 5 < x \leqslant 6; \\ 1, & x > 6. \end{cases}$$

График функции распределения $F(x)$ изображён на рис. 7.

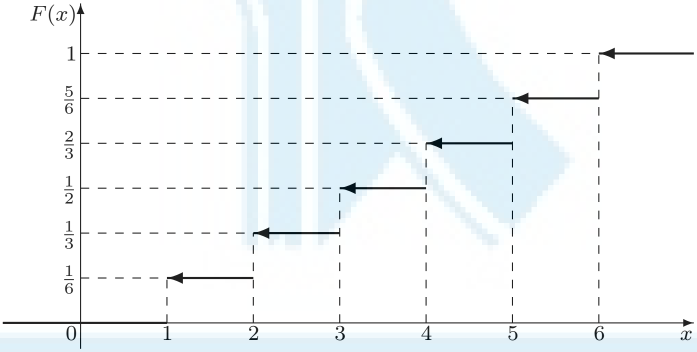


Рис. 7

б) Поскольку при одном подбрасывании игрального кубика шестёрка может или совсем не выпасть, или выпасть лишь один раз, случайная величина ψ принимает два значения: 0 и 1. Легко убедиться, что

$$\mathbf{P}\{\psi = 0\} = \frac{5}{6}; \qquad \mathbf{P}\{\psi = 1\} = \frac{1}{6}.$$

Ряд распределения случайной величины ψ имеет такой вид:

$$\begin{array}{c|c|c} \psi & 0 & 1 \\ \hline \mathbf{P} & \frac{5}{6} & \frac{1}{6} \end{array}.$$

Многоугольник распределения случайной величины ψ изображен на рис. 8.

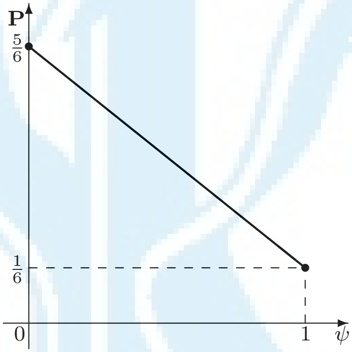


Рис. 8

Функция распределения случайной величины ψ

$$F(x) = \begin{cases} 0, & x \leqslant 0; \\ \dfrac{5}{6}, & 0 < x \leqslant 1; \\ 1, & x > 1. \end{cases}$$

График функции распределения изображён на рис. 9. □

Наиболее важными являются такие типы дискретных распределений: равномерный, биномиальный, показательный, геометрический, гипергеометрический.

Биномиальное распределение (распределение Бернулли) — это распределение случайной величины, возможные значения

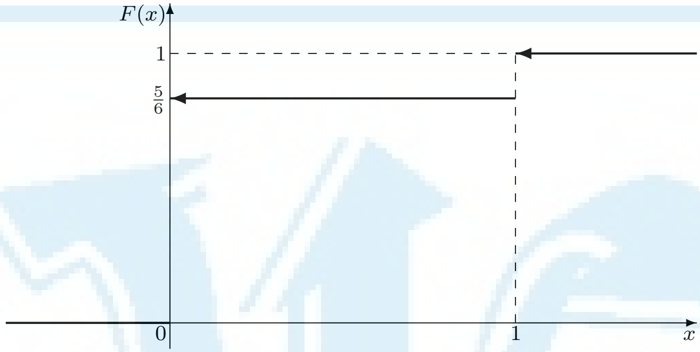


Рис. 9

$i \in \{0, 1, \ldots, n\}$ которой принимаются с вероятностями

$$p_i = C_n^i p^i (1-p)^{n-i},$$

где $p \in (0; 1)$.

Биномиальное распределение имеет случайная величина, которая равняется количеству «успехов» в серии из n независимых испытаний, в любом из которых «успех» происходит с вероятностью p.

Равномерное распределение — это распределение случайной величины, которая принимает n разных значений с одинаковыми вероятностями

$$p_i = \frac{1}{n}.$$

Показательное распределение (распределение Пуассона) — это распределение случайной величины, которая принимает значения $k \in \{0, 1, 2, \ldots, n\}$ с вероятностями

$$\mathbf{P}_n(k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

где $\lambda = np$, $n \to \infty$, $p \to 0$.

Геометрическое распределение — это распределение случайной величины, которая принимает значения $k \in \{1, 2, \ldots\}$ с вероятностями

$$p_k = pq^{k-1},$$

где $q = 1 - p$. Геометрическое распределение имеет случайная величина, которая равняется количеству попыток до первого «успеха» в серии независимых испытаний, в каждом из которых вероятность успеха равняется p.

Гипергеометрическое распределение — это распределение случайной величины, которая принимает значения $k \in \{0, 1, \ldots, m\}$ с вероятностями

$$p_k = \frac{C_n^k C_{N-n}^{m-k}}{C_N^m},$$

где $n \geqslant m$, $N \geqslant n$.

Гипергеометрическое распределение имеет такая случайная величина. Пусть в ящике находится N одинаковых по физическим свойствам шаров, среди которых S белых и $(N - S)$ чёрных. Из ящика наугад вынимают m шаров. Тогда случайная величина ξ, которая равняется количеству белых шаров среди m вынутых, имеет гипергеометрическое распределение (рис. 10).

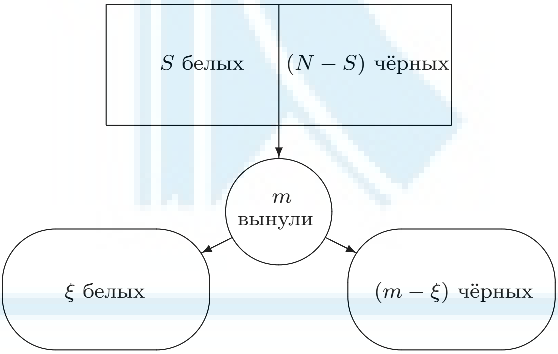


Рис. 10

Пример 2. Пусть случайная величина ξ равняется количеству номеров, угаданных игроком в лотереи «6 из 39». Построить ряд распределения случайной величины ξ. Найти значения функции распределения в точке $x = 3$.

Решение. Случайная величина ξ имеет гипергеометрическое распределение. Процесс розыгрыша можно смоделировать так: в лототроне (ящике) находится $N = 39$ одинаковых по физическим свойствам шаров, среди которых $S = 6$ «белых» (так можно называть шары, которые выпадут в результате розыгрыша) и $N - S = 33$ «чёрных». Допустим, что шары хорошо перемешаны. Поэтому можно считать, что из 39 шаров наугад вынимают $m = 6$ шт. Тогда ξ — это количество «белых» шаров среди шести выбранных (это номера, загаданные игроком).

Очевидно, случайная величина ξ принимает значения от 0 до 6, причём

$$\mathbf{P}\{\xi = i\} = \frac{C_6^i \cdot C_{33}^{6-i}}{C_{39}^6}.$$

Выполнив подсчеты для $i = 0, 1, 2, 3, 4, 5, 6$, получим:

ξ	0	1	2	3	4	5	6
$\mathbf{P}$	$\frac{1107568}{3262623}$	$\frac{1424016}{3262623}$	$\frac{613800}{3262623}$	$\frac{109120}{3262623}$	$\frac{7920}{3262623}$	$\frac{198}{3262623}$	$\frac{1}{3262623}$

Значение функции распределения $F(x)$ случайной величины ξ в точке $x = 3$

$$\begin{aligned} F(3) = \mathbf{P}\{\xi < 3\} &= \mathbf{P}\{\xi = 0 \text{ або } \xi = 1 \text{ або } \xi = 2\} = \\ &= \mathbf{P}\{\xi = 0\} + \mathbf{P}\{\xi = 1\} + \mathbf{P}\{\xi = 2\}. \end{aligned}$$

Другими словами, вероятность того, что игрок ничего не выиграет, заполнив один лотерейный билет, равняется

$$\frac{1107568}{3262623} + \frac{1424016}{3262623} + \frac{613800}{3262623} = \frac{3145384}{3262623} \approx 0{,}9641. \qquad \square$$

Задачи к разделу 6

Задача 1. Пусть ξ — случайная величина, которая равняется количеству мальчиков в семье с тремя детьми. Считая рождение мальчика и девочки равновероятными событиями, найти ряд распределения и вероятность того, что в семье будет больше мальчиков, чем девочек.

Ответ.

ξ	0	1	2	3
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

; $\mathbf{P}$(«мальчиков больше») $= \dfrac{1}{2}$.

Задача 2. Пусть ξ — случайная величина, которая равняется количеству выпадений грани «6» при двух подбрасываниях кубика. Найти ряд распределения, функцию распределения случайной величины ξ и построить её график.

Ответ.

ξ	0	1	2
P	$\frac{25}{36}$	$\frac{5}{18}$	$\frac{1}{36}$

; $$F_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ \dfrac{25}{36}, & 0 < x \leqslant 1; \\ \dfrac{35}{36}, & 1 < x \leqslant 2; \\ 1, & x > 2. \end{cases}$$

График функции распределения изображён на рис. 11.

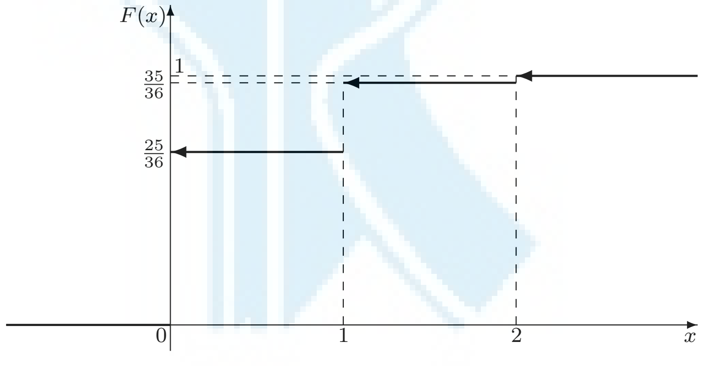


Рис. 11

Задача 3. Стрелок делает по одному выстрелу по четырём мишеням. Случайная величина ξ — количество попаданий. Считая, что вероятность попадания при одном выстреле равняется 0,7, найти ряд распределения случайной величины ξ и вероятность того, что попаданий будет больше, чем промахов.

Ответ.

ξ	0	1	2	3	4
P	0,0081	0,0756	0,2646	0,4116	0,2401

; $\mathbf{P}$(«попаданий больше») $= 0{,}6517$.

Задача 4. Два стрелка делают по одному независимому выстрелу в одну мишень. Вероятность попадания для первого стрелка равняется 0,8, а для второго — 0,5. Найти ряд распределения случайной величины ξ — количества попаданий в мишень, а также вероятность того, что количество попаданий будет равняться количеству промахов.

Ответ.

ξ	0	1	2
P	0,1	0,5	0,4

; $\mathbf{P}$(«количество попаданий равняется количеству промахов») $= 0{,}5$.

Задача 5. Студент подбрасывает монету до первого выпадения аверса. Случайная величина ξ — количество подбрасываний. Найти распределение случайной величины ξ. Какая вероятность того, что студент вынужден будет выполнить более 10 подбрасываний?

Ответ. $\mathbf{P}(\xi = k) = \dfrac{1}{2^k}$, $k = 1, 2, \ldots$; $\mathbf{P}(\xi > 10) = \dfrac{1}{1024}$.

Задача 6. Игральный кубик подбрасывают до первого появления пятёрки. Случайная величина ξ — количество подбрасываний кубика. Найти ряд распределения случайной величины ξ и наиболее вероятное количество подбрасываний.

Ответ. $\mathbf{P}(\xi = k) = \dfrac{1}{6} \cdot \left(\dfrac{5}{6}\right)^{k-1}$, $k = 1, 2, \ldots$; $m_0 = 1$.

Задача 7. Студент выучил 10 билетов из 25 и пытается сдать экзамен. Случайная величина ξ — количество попыток до успешной сдачи. Найти ряд распределения случайной величины ξ и вероятность того, что студента отчислят за академзадолженность (экзамен разрешается пересдавать дважды).

Ответ. $\mathbf{P}(\xi = k) = \dfrac{2}{5} \cdot \left(\dfrac{3}{5}\right)^{k-1}$, $k = 1, 2, \ldots$; $\mathbf{P}(\xi > 3) = 0{,}216$.

Задача 8. Игрок в казино ставит «на красное» до тех пор, пока не выиграет. Случайная величина ξ — количество попыток. Найти ряд распределения случайной величины ξ, наиболее вероятное количество попыток и то количество попыток x, при котором выполняется условие $\mathbf{P}\,(\xi \leqslant x) \geqslant 0{,}999$.

Указание. Считать, что рулетка имеет 37 полей (от 0 до 36), среди которых 18 красных.

Ответ. $\mathbf{P}(\xi = k) = \dfrac{18}{37} \cdot \left(\dfrac{19}{37}\right)^{k-1}$, $k = 1, 2, \ldots$; $m_0 = 1$; $x \geqslant 11$.

Задача 9. Студент знает 20 вопросов из 25. Наугад вынимает 5 вопросов. Случайная величина ξ — количество вопросов, на которые студент знает ответ. Найти ряд распределения случайной величины ξ и вероятность сдачи экзамена (экзамен сдан, если студент отвечает более чем на половину вопросов).

Ответ.

ξ	0	1	2	3	4	5
$\mathbf{P}$	$\frac{1}{3125}$	$\frac{4}{625}$	$\frac{32}{625}$	$\frac{128}{625}$	$\frac{256}{625}$	$\frac{1024}{3125}$

; $\mathbf{P}(\text{«сдаст»}) = \dfrac{2944}{3125} = 0{,}94208$.

Задача 10. В партии из 10 деталей 2 бракованные. Наугад берут 3 детали. Пусть ξ — случайная величина, которая равняется количеству бракованных деталей среди выбранных. Найти ряд распределения и функцию распределения случайной величины ξ.

Ответ.

ξ	0	1	2
$\mathbf{P}$	$\frac{7}{15}$	$\frac{7}{15}$	$\frac{1}{15}$

; $F_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ \dfrac{7}{15}, & 0 < x \leqslant 1; \\ \dfrac{14}{15}, & 1 < x \leqslant 2; \\ 1, & x > 2. \end{cases}$

Задача 11. Пусть ξ — случайная величина, которая равняется количеству угаданных номеров в лотереи «5 из 36». Найти ряд распределения случайной величины ξ и вероятность выиграть хотя бы что-нибудь (выигрыши начинаются с трёх угаданных номеров).

Ответ.

ξ	0	1	2	3	4	5
$\mathbf{P}$	0,45	0,42	0,12	0,01	0,0004	0,000003

;
$\mathbf{P}(\text{«выигрыш»}) \approx 0{,}013$.

Задача 12. После тестирования оказалось, что среди 15 студентов — один меланхолик, 5 флегматиков, 6 сангвиников и 3 холерика. Из этой группы наугад выбирают четырёх студентов. Пусть ξ — случайная величина, которая равняется количеству сангвиников среди выбранных студентов. Найти ряд распределения этой случайной величины.

Ответ.

ξ	0	1	2	3	4
P	0,47	0,41	0,11	0,01	0,0003

Задача 13. Магазин получил 10000 бутылок минеральной воды. Вероятность того, что при перевозке бутылка окажется разбитой, равняется 0,0004. Найти ряд распределения случайной величины ξ, которая характеризует количество разбитых бутылок, и вероятность того, что разбитых бутылок будет больше трёх.

Ответ. $\mathbf{P}(\xi = k) = \dfrac{4^k}{k!}e^{-4}$, $k = 1, 2, \ldots$; $\mathbf{P}(\xi \geqslant 4) \approx 0{,}5665$.

Задача 14. Учебник издан тиражом 10000 экземпляров. Вероятность того, что сброшюрованный экземпляр окажется некачественным, равняется 0,001. Найти ряд распределения случайной величины ξ, которая характеризует количество некачественно сброшюрованных экземпляров, и вероятность того, что бракованных экземпляров будет больше двух.

Ответ. $\mathbf{P}(\xi = k) = \dfrac{10^k}{k!}e^{-10}$, $k = 1, 2, \ldots$; $\mathbf{P}(\xi > 2) \approx 0{,}9972$.

Задача 15. При наборе текста машинистка набирает неправильный символ с вероятностью 0,001. Найти вероятность того, что при наборе страницы текста (3000 символов) будет сделано больше трёх ошибок.

Ответ. $\mathbf{P}_{3000}(\xi > 3) \approx 0{,}3528$.

Задача 16. Вероятность попадания в муху на лету из рогатки равняется 0,001. Производится 2000 выстрелов. Случайная величина ξ равняется количеству убитых мух. Построить ряд распределения случайной величины ξ и найти вероятность того, что будет убито меньше четырёх мух.

Ответ. $\mathbf{P}(\xi = k) = \dfrac{2^k}{k!}e^{-2}$, $k = 1, 2, \ldots$; $\mathbf{P}(\xi < 4) \approx 0{,}8571$.

Задача 17. Автоматическая телефонная станция обслуживает 10000 телефонных номеров. Вероятность того, что в течение 1 мин на АТС поступит вызов от абонента, равняется 0,0004. Найти ряд распределения случайной величины X, которая равняется количеству вызовов, поступивших на АТС в течение 1 мин, и вероятность того, что за это время поступит хотя бы один вызов.

Ответ. $\mathbf{P}(\xi = k) = \dfrac{4^k}{k!}e^{-4}$, $k = 1, 2, \ldots$; $\mathbf{P}(\xi \geqslant 1) \approx 0{,}9817$.

Раздел 7. Непрерывные случайные величины

Ключевые слова

Абсолютно	Практический
Величина	Промежуток
Дифференциальный	Равномерный
Достоверность	Распределение
Непрерывный	Сингулярный
Нормальный	Случайный
Плотность	Функция
Показательный	Экспоненциальный

Случайная величина называется *непрерывной*, если её функция распределения является непрерывной.

Если ξ — непрерывная случайная величина, то вероятность того, что ξ примет некоторое конкретное значение x, равняется 0 ($\mathbf{P}(\xi = x) = 0,\ x \in \mathbb{R}$). Поэтому в отличие от дискретной случайной величины распределение непрерывной случайной величины невозможно задать, указав значения, которые она принимает, и соответствующие им вероятности. Понятно также, что множество возможных значений непрерывной случайной величины несчётное.

Простейшим и наиболее распространённым способом задания распределения непрерывной случайной величины является указание её функции распределения $F_\xi(x)$. Тогда для любого промежутка $(a;b]$ можно вычислить вероятность

$$\mathbf{P}(\xi \in (a;b]) = F_\xi(b) - F_\xi(a).$$

Непрерывные случайные величины делятся на два класса: *сингулярные* и *абсолютно непрерывные*.

Случайная величина ξ называется *абсолютно непрерывной*, если существует такая функция $f_\xi(x)$, для которой выполняется равенство

$$F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt.$$

При этом функция $f_\xi(x)$ называется *плотностью распределения случайной величины* ξ (или *дифференциальной функцией распределения*).

Очевидно, что:

1. $f_\xi(x) \geqslant 0$;
2. $F'_\xi(x) = f_\xi(x)$ в каждой точке непрерывности функции $f_\xi(x)$;
3. $\int\limits_{-\infty}^{+\infty} f_\xi(x)dx = 1$;
4. $\mathbf{P}(\xi \in (a;b]) = F_\xi(b) - F_\xi(a) = \int\limits_{a}^{b} f_\xi(x)dx.$

Другими словами, вероятность того, что случайная величина ξ примет значение из промежутка $(a;b]$, численно равняется площади криволинейной трапеции, ограниченной графиком функции $y = f_\xi(x)$, осью абсцисс и прямыми $x = a$, $x = b$.

Сингулярные случайные величины, в отличие от абсолютно непрерывных, невозможно задать с помощью плотности распределения.

Наиболее важными являются такие типы абсолютно непрерывных распределений: равномерный, нормальный, экспоненциальный.

Распределение случайной величины ξ называется *равномерным* на отрезке $[a;b]$, если его плотность имеет вид

$$f_\xi(x) = \begin{cases} 0, & x \notin [a;b]; \\ \dfrac{1}{b-a}, & x \in [a;b]. \end{cases}$$

Функция распределения равномерно распределённой случайной величины имеет вид

$$F_\xi(x) = \begin{cases} 0, & x \leqslant a; \\ \dfrac{x-a}{b-a}, & a < x \leqslant b; \\ 1, & x > b. \end{cases}$$

Графики плотности и функции распределения равномерно распределённой случайной величины изображены соответственно на рис. 12 и 13.

Распределение случайной величины ξ называется *нормальным* с параметрами a, σ, если его плотность

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

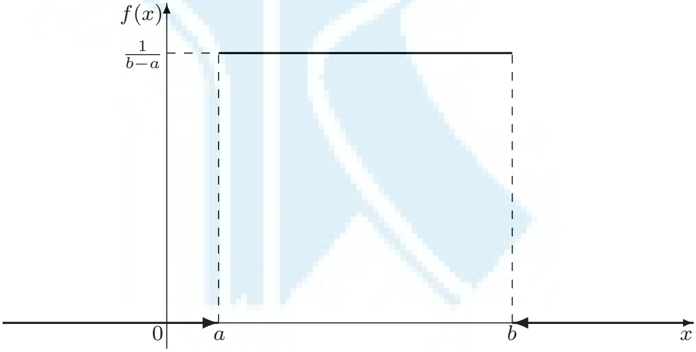


Рис. 12

Функция $y = f_\xi(x)$ быстро убывает при $x \to \pm\infty$. Площадь фигуры, ограниченной графиком функции $y = f_\xi(x)$ и осью Ox, равняется единице. Площади криволинейных трапеций над интервалами

$$[a-\sigma; a+\sigma], \qquad [a-2\sigma; a+2\sigma], \qquad [a-3\sigma; a+3\sigma]$$

равняются соответственно

$$0{,}6827, \qquad 0{,}9545, \qquad 0{,}9973.$$

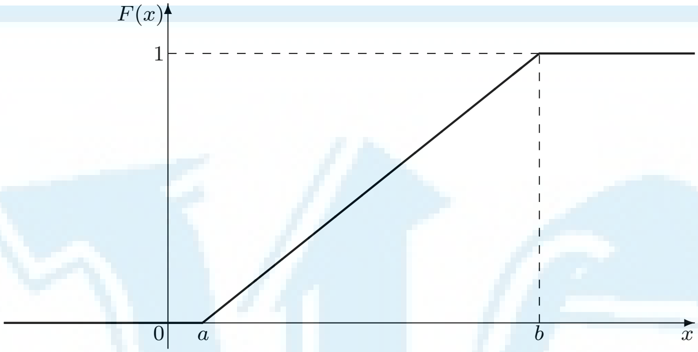


Рис. 13

Итак, вероятность того, что нормально распределённая случайная величина примет значение из отрезка

$$[a - 3\sigma; a + 3\sigma],$$

равняется 0,9973. В этом случае говорят, что нормально распределённая случайная величина ξ попадает в *«трёхсигмовый» промежуток* $[a - 3\sigma; a + 3\sigma]$ с *практической достоверностью*. На рис. 14 изображён график функции плотности нормального распределения с математическим ожиданием 0.

Функция распределения нормально распределённой случайной величины имеет вид

$$F_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}\, dt.$$

Эта функция не является элементарной. Поэтому её значения табулированы. Для удобства рассматривают функцию Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{x} e^{-\frac{t^2}{2}}\, dt.$$

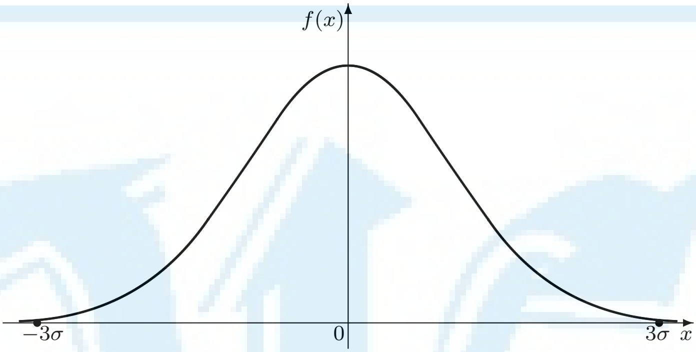


Рис. 14

Функции $F_\xi(x)$ и $\Phi(x)$ связаны соотношением

$$F_\xi(x) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right).$$

Таблицы значений функции $\Phi(x)$ приведены в прил. 1. Отметим, что $\Phi(x)$ имеет такие свойства.

Свойство 1. $\Phi(-x) = -\Phi(x)$.

Свойство 2. При $x > 5$ можно считать, что $\Phi(x) \approx 0,5$.

Вероятность того, что нормально распределённая случайная величина примет значение, которое принадлежит отрезку $[\alpha; \beta]$,

$$\mathbf{P}(\xi \in [\alpha; \beta]) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right).$$

Вероятность того, что нормально распределенная случайная величина примет значение из промежутка $[a-\delta; a+\delta]$,

$$\mathbf{P}(|\xi - a| \leqslant \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

Нормальное распределение играет чрезвычайно важную роль в теории вероятностей и её применениях. Эта роль в значительной степени объясняется центральной предельной теоремой. Если известно, что исследуемая случайная величина является суммой большого количества случайных слагаемых, каждое из которых мало влияет на всю сумму, то можно считать, что результирующая случайная величина имеет нормальное распределение. Например, случайная погрешность измерения имеет, как правило, нормальное распределение.

Распределение случайной величины ξ называется *экспоненциальным (показательным)* с параметром $\lambda > 0$, если его плотность имеет вид

$$f_\xi(x) = \begin{cases} 0, & x < 0; \\ \lambda e^{-\lambda x}, & x \geqslant 0. \end{cases}$$

График функции плотности экспоненциального закона изображён на рис. 15.

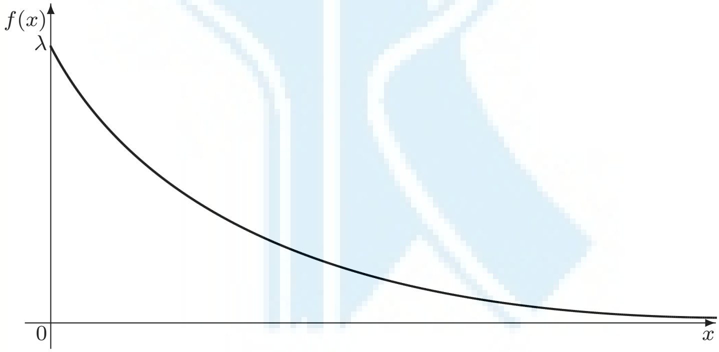


Рис. 15

Функция распределения случайной величины с экспоненциальным распределением имеет вид

$$F_\xi(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\lambda x}, & x \geqslant 0. \end{cases}$$

Если $\alpha > 0$ и $\beta > 0$, то

$$\mathbf{P}(\xi \in [\alpha; \beta]) = e^{-\lambda\alpha} - e^{-\lambda\beta}.$$

Если $\alpha \leqslant 0$ и $\beta > 0$, то

$$\mathbf{P}(\xi \in [\alpha; \beta]) = \mathbf{P}(\xi \in [0; \beta]) = 1 - e^{-\lambda\beta}.$$

Показательное распределение часто встречается в теории массового обслуживания (в частности, при описании работы телефонных станций).

Пример 1. Случайная величина ξ задана функцией распределения

$$F_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ \sqrt{x}, & 0 < x \leqslant \dfrac{1}{4}; \\ 8x^2, & \dfrac{1}{4} < x \leqslant \dfrac{\sqrt{2}}{4}; \\ 1, & x > \dfrac{\sqrt{2}}{4}. \end{cases}$$

Убедитесь, что величина ξ имеет плотность распределения вероятностей, и найдите её.

Решение. В точках

$$x = 0; \frac{1}{4}; \frac{\sqrt{2}}{4}$$

правосторонние и левосторонние границы функции $F_\xi(x)$ совпадают, из чего вытекает, что $F_\xi(x)$ непрерывная всюду. Производная $F'_\xi(x)$ непрерывная всюду, за исключением точек $x = 0; \frac{1}{4}; \frac{\sqrt{2}}{4}$. Следовательно, плотность распределения вероятностей $f_\xi(x)$ существует. Вычислим односторонние границы $F'_\xi(x)$ в точках $x = 0; \frac{1}{4}; \frac{\sqrt{2}}{4}$:

$$F'_\xi(0-0) = 0, \qquad F'_\xi(0+0) = \infty,$$

$$F'_\xi\left(\frac{1}{4} - 0\right) = 1, \qquad F'_\xi\left(\frac{1}{4} + 0\right) = 4,$$

$$F'_\xi\left(\frac{\sqrt{2}}{4} - 0\right) = 4\sqrt{2}, \qquad F'_\xi\left(\frac{\sqrt{2}}{4} + 0\right) = 0.$$

Итак, точки 0, $\dfrac{1}{4}$ и $\dfrac{\sqrt{2}}{4}$ являются точками разрыва для $F'_\xi(x)$.

Таким образом,

$$f_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ \dfrac{1}{2\sqrt{x}}, & 0 < x \leqslant \dfrac{1}{4}; \\ 16x, & \dfrac{1}{4} < x \leqslant \dfrac{\sqrt{2}}{4}; \\ 0, & x > \dfrac{\sqrt{2}}{4}. \end{cases}$$

Примечание. Точки 0, $\dfrac{1}{4}$ и $\dfrac{\sqrt{2}}{4}$ можно включить в любой соседний промежуток. □

Пример 2. Дана функция

$$f(x) = \begin{cases} 0, & x \leqslant 0; \\ ae^{-\alpha x}, & x > 0, \end{cases} \qquad \alpha > 0.$$

При каком значении постоянной a функция $f(x)$ является плотностью распределения вероятностей некоторой случайной величины ξ? Найти функцию распределения $F_\xi(x)$ величины ξ. Вычислить вероятность того, что случайная величина ξ попадёт в промежуток $[0; 1]$, двумя способами: с помощью плотности распределения вероятностей $f_\xi(x)$ и с помощью функции распределения $F_\xi(x)$.

Решение. Прежде всего постоянная a должна быть неотрицательной: $a \geqslant 0$. Для нахождения значения a запишем условие

$$\int\limits_{-\infty}^{+\infty} f_\xi(x)dx = 1$$

и преобразуем его:

$$\int\limits_{-\infty}^{\infty} f_\xi(x)dx = a\int\limits_{0}^{\infty} e^{-\alpha x}dx = -\frac{a}{\alpha}e^{-\alpha x}\Big|_0^\infty = \frac{a}{\alpha} = 1.$$

Итак, $a = \alpha$ и функция $f_\xi(x)$ имеет вид

$$f(x) = \begin{cases} 0, & x \leqslant 0; \\ \alpha e^{-\alpha x}, & x > 0, \end{cases} \qquad \alpha > 0. \tag{2}$$

Найдём функцию распределения $F_\xi(x)$ по формуле

$$F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt.$$

Если $x \leqslant 0$, то

$$F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt = \int\limits_{-\infty}^{x} 0dt = 0,$$

если $x > 0$, то

$$F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt(x) = \int\limits_{-\infty}^{0} 0dt + \int\limits_{0}^{x} \alpha e^{-\alpha t}dt = 1 - e^{-\alpha x}.$$

Итак,

$$F_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ 1 - \alpha e^{-\alpha x}, & x > 0, \end{cases} \qquad \alpha > 0. \tag{3}$$

По формуле

$$\mathbf{P}(x_1 \leqslant x \leqslant x_2) = \int\limits_{x_1}^{x_2} f_\xi(x)dx$$

вычислим вероятность $\mathbf{P}(0 \leqslant x \leqslant 1)$:

$$\mathbf{P}(0 \leqslant x \leqslant 1) = \int\limits_{0}^{1} f_\xi(x)dx = \int\limits_{0}^{1} \alpha e^{-\alpha x}dx = 1 - \frac{1}{e^{\alpha}}.$$

Формула
$$\mathbf{P}(x_1 \leqslant x \leqslant x_2) = F_\xi(x_2) - F_\xi(x_1)$$
даёт такой же результат:
$$\mathbf{P}(0 \leqslant x \leqslant 1) = F_\xi(1) - F_\xi(0) = 1 - \frac{1}{e^\alpha}.$$

Случайные величины с плотностью (2) или с функцией распределения (3) довольно часто встречаются на практике (показательный закон распределения). □

Пример 3. Случайная величина ξ имеет нормальное распределение с параметрами $a = 7$ и $\sigma = 2$. Сравнить вероятности попадания случайной величины ξ в промежутки $(3; 7]$ и $(-100; 1]$. Указать интервал, в который случайная величина ξ попадает с практической достоверностью.

Решение. Если случайная величина ξ имеет нормальное распределение, то
$$\mathbf{P}(\xi \in (\alpha; \beta]) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Поэтому
$$\mathbf{P}(\xi \in (\alpha; \beta]) = \Phi\left(\frac{7 - 7}{2}\right) - \Phi\left(\frac{3 - 7}{2}\right) = \Phi(0) - \Phi(-2).$$

По таблице значений функции $\Phi(x)$ (прил. 1) находим:
$$\Phi(0) = 0; \qquad \Phi(-2) = -\Phi(2) \approx -0{,}4772.$$

Поэтому
$$\mathbf{P}(\xi \in (3; 7]) \approx 0{,}4772.$$

Аналогично,
$$\mathbf{P}(\xi \in (-100; 1]) = \Phi\left(\frac{1 - 7}{2}\right) - \Phi\left(\frac{-100 - 7}{2}\right) =$$
$$= \Phi(-3) - \Phi(-53{,}5) = -\Phi(3) + \Phi(53{,}5).$$

Из прил. 1 получаем:

$$\Phi(3) \approx 0{,}49865.$$

Значение $\Phi(53,5)$ в прил. 1 найти невозможно. Поскольку $53{,}5 > 5$, то с большой точностью можно считать, что

$$\Phi(53{,}5) \approx 0{,}5.$$

Итак,

$$\mathbf{P}(\xi \in (-100; 1]) \approx -0{,}49865 + 0{,}5 = 0{,}00135.$$

Нормально распределённая случайная величина с практической достоверностью попадает в так называемый трёхсигмовый интервал

$$(a - 3\sigma; a + 3\sigma].$$

Следовательно,

$$\mathbf{P}(\xi \in (1; 12]) = 0{,}9973.$$

Предлагаем читателю самостоятельно убедиться в правильности последнего равенства. □

Задачи к разделу 7

Задача 1. Случайная величина X имеет функцию распределения

$$F(x) = \begin{cases} 0, & x \leqslant -4, \\ \dfrac{a(x+4)^2}{16}, & -4 < x \leqslant 0, \\ 1, & x > 0. \end{cases}$$

Найти параметр a, аналитическое выражение для плотности и вероятность попадания случайной величины X в интервал $(-2; 5)$.

Ответ. $a = 1$, $f(x) = \begin{cases} \dfrac{x+4}{8}, & x \in (-4;0); \\ 0, & x \notin (-4;0), \end{cases}$ $\mathbf{P}\{X \in (-2;5)\} = $
$= \dfrac{3}{4}$.

Задача 2. Случайная величина X имеет функцию распределения

$$F(x) = \begin{cases} 0, & x \leqslant 0, \\ a\sqrt{2x}, & 0 < x \leqslant 8, \\ 1, & x > 8. \end{cases}$$

Найти параметр a, аналитическое выражение для плотности и вероятность попадания случайной величины X в интервал $(1;6)$.

Ответ. $a = \dfrac{1}{4}$, $f(x) = \begin{cases} \dfrac{1}{4\sqrt{2x}}, & x \in (0;8); \\ 0, & x \notin (0;8), \end{cases}$ $\mathbf{P}\{X \in (1;6)\} = $
$= \dfrac{2\sqrt{3}-\sqrt{2}}{4} \approx 0,51$.

Задача 3. При каких значениях параметров a и b функция

$$F(x) = \begin{cases} 0, & x \leqslant a, \\ \sin x, & a < x \leqslant b, \\ 1, & x > b \end{cases}$$

будет функцией распределения некоторой случайной величины? Найти плотность данной случайной величины.

Ответ. $a = 2\pi k$, $b = \dfrac{\pi}{2} + 2\pi k$, $k \in \mathbb{Z}$. Если, например, $k = 0$, то
$f(x) = \begin{cases} 0, & x \notin \left(0; \dfrac{\pi}{2}\right); \\ \cos x, & x \in \left(0; \dfrac{\pi}{2}\right). \end{cases}$

Задача 4. При каких значениях параметров a и b функция

$$F(x) = \begin{cases} 0, & x \leqslant a, \\ \operatorname{tg} x, & a < x \leqslant b, \\ 1, & x > b \end{cases}$$

будет функцией распределения некоторой случайной величины? Найти плотность данной случайной величины.

Ответ. $a = \pi k$, $b = \dfrac{\pi}{4} + \pi k$, $k \in \mathbb{Z}$. Если, например, $k = 0$, то
$$f(x) = \begin{cases} 0, & x \notin \left(0; \dfrac{\pi}{4}\right); \\ \dfrac{1}{\cos^2 x}, & x \in \left(0; \dfrac{\pi}{4}\right). \end{cases}$$

Задача 5. Плотность случайной величины X задана графически (рис. 16).

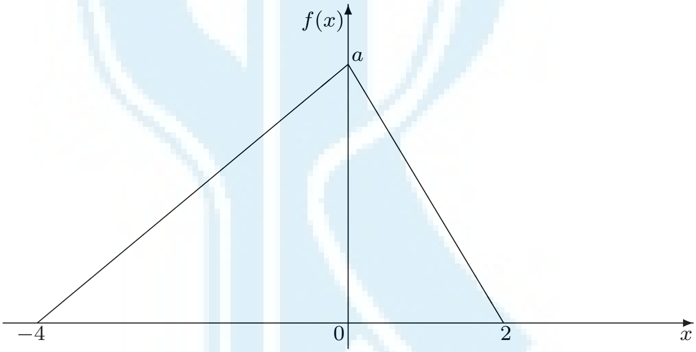


Рис. 16

Определить значения параметра a и аналитические выражения для $f(x)$ и $F(x)$. Найти
$$\mathbf{P}\{X \in [0;2]\}, \qquad \mathbf{P}\{X \in [-2;1]\}.$$

Ответ. $a = \dfrac{1}{3}$, $f(x) = \begin{cases} 0, & x \notin (-4;2); \\ \dfrac{1}{12}(x+4), & x \in [-4;0]; \\ \dfrac{1}{6}(2-x), & x \in [0;2], \end{cases}$

$$F(x) = \begin{cases} 0, & x \leqslant -4; \\ \dfrac{(x+4)^2}{24}, & -4 \leqslant x \leqslant 0; \\ 1 - \dfrac{(2-x)^2}{12}, & 0 \leqslant x \leqslant 2; \\ 1, & x \geqslant 2, \end{cases} \qquad \mathbf{P}\{X \in [0;2]\} = \frac{1}{3},$$

$\mathbf{P}\{X \in [-2;1]\} = \dfrac{3}{4}$.

Задача 6. Плотность случайной величины ξ задана графически (рис. 17).

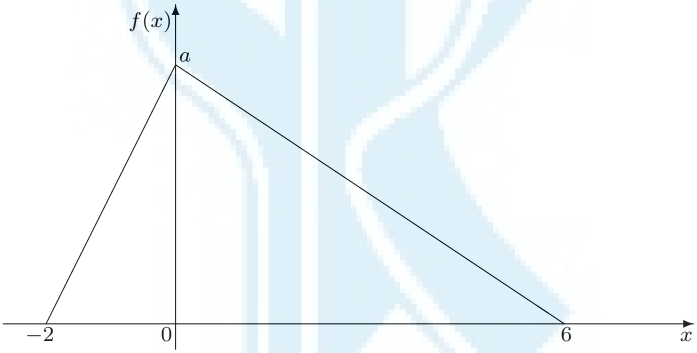


Рис. 17

Определить значения параметра a и аналитические выражения для $f(x)$ и $F(x)$. Найти

$$\mathbf{P}\{\xi \in [6;8]\}, \qquad \mathbf{P}\{\xi \in [-2;2]\}.$$

Ответ. $a = \dfrac{1}{4}$, $f(x) = \begin{cases} 0, & x \notin [-2;6]; \\ \dfrac{1}{8}(x+2), & x \in [-2;0]; \\ \dfrac{1}{24}(6-x), & x \in [0;6], \end{cases}$

$$F(x) = \begin{cases} 0, & x \leqslant -2; \\ \dfrac{(x+2)^2}{16}, & -2 < x \geqslant 0; \\ 1 - \dfrac{(6-x)^2}{48}, & 0 < x \geqslant 6; \\ 1, & x > 6, \end{cases} \qquad \mathbf{P}\{\xi \in [6; 8]\} = 0,$$

$\mathbf{P}\{\xi \in [-2; 2]\} = \dfrac{2}{3}$.

Задача 7. При каких значениях параметров a и b функция

$$f(x) = \begin{cases} 0, & x \leqslant a, \\ \sin x, & a < x \leqslant b, \\ 0, & x > b \end{cases}$$

может быть плотностью некоторой случайной величины X? Найти функцию распределения данной случайной величины и вероятность попадания этой случайной величины в промежуток $\left[\dfrac{a+b}{2}; b\right]$.

Ответ. Число a можно выбрать произвольно из промежутка $\left[2\pi k; \dfrac{\pi}{2} + 2\pi k\right]$, $k \in \mathbb{Z}$, а число b определить из условия $\cos a - \cos b = 1$.

Если $a = 0$, то $b = \dfrac{\pi}{2}$. Тогда $F(x) = \begin{cases} 0, & x \leqslant 0; \\ 1 - \cos x, & x \in \left[0; \dfrac{\pi}{2}\right]; \\ 1, & x \geqslant \dfrac{\pi}{2}, \end{cases}$

$\mathbf{P}\left\{X \in \left[\dfrac{\pi}{4}; \dfrac{\pi}{2}\right]\right\} = \dfrac{\sqrt{2}}{2}$.

Задача 8. При каких значениях параметров a и b функция

$$f(x) = \begin{cases} 0, & x \leqslant a, \\ \operatorname{tg} x, & a < x \leqslant b, \\ 0, & x > b \end{cases}$$

может быть плотностью некоторой случайной величины X? Найти функцию распределения данной случайной величины и вероятность попадания этой случайной величины в промежуток $\left[a; \dfrac{a+b}{2}\right]$.

Ответ. Число a можно выбрать произвольно из промежутка $\left[\pi k; \dfrac{\pi}{2} + \pi k\right]$, $k \in \mathbb{Z}$, а число b определить из условия $\dfrac{\cos a}{\cos b} = e$. Если $a = 0$, то $b = \arccos \dfrac{1}{e}$. Тогда $F(x) = \begin{cases} 0, & x \leqslant 0; \\ -\ln(\cos x), & x \in \left[0; \arccos \dfrac{1}{e}\right]; \\ 1, & x \geqslant \arccos \dfrac{1}{e}, \end{cases}$

$\mathbf{P}\left\{X \in \left[0; \dfrac{1}{2} \arccos \dfrac{1}{e}\right]\right\} = -\dfrac{1}{2} \ln\left(\dfrac{e+1}{2e}\right) \approx 0{,}19.$

Задача 9. Случайная величина ξ задана плотностью

$$f_\xi(x) = \begin{cases} 0, & x \notin \left[-\dfrac{\pi}{2}; \dfrac{\pi}{2}\right]; \\ a\cos x, & x \in \left[-\dfrac{\pi}{2}; \dfrac{\pi}{2}\right]. \end{cases}$$

Найти параметр a, функцию распределения $F_\xi(x)$ и вероятность попадания случайной величины ξ в интервал $\left(0; \dfrac{\pi}{4}\right]$.

Ответ. $a = \dfrac{1}{2}$; $F_\xi(x) = \begin{cases} 0, & x \leqslant -\dfrac{\pi}{2}; \\ \dfrac{1+\sin x}{2}, & -\dfrac{\pi}{2} < x \leqslant \dfrac{\pi}{2}; \\ 1, & x > \dfrac{\pi}{2}; \end{cases}$

$\mathbf{P}\left(\xi \in \left(0; \dfrac{\pi}{4}\right]\right) = \dfrac{\sqrt{2}}{4}.$

Задача 10. Случайная величина ξ задана плотностью

$$f_\xi(x) = \begin{cases} 0, & x \leqslant -2; \\ a, & -2 < x \leqslant 5; \\ 0, & x > 5. \end{cases}$$

Найти параметр a, функцию распределения $F_\xi(x)$ и вероятность попадания случайной величины ξ в интервал $(0; 3]$.

Ответ. $a = \dfrac{1}{7}$; $F_\xi(x) = \begin{cases} 0, & x \leqslant -2; \\ \dfrac{x+2}{7}, & -2 < x \leqslant 5; \\ 1, & x > 5; \end{cases}$ $\mathbf{P}(\xi \in (0;3]) = \dfrac{3}{7}$.

Задача 11. Случайная величина ξ имеет равномерный закон распределения на отрезке $[2;5]$. Найти аналитические выражения для плотности и функции распределения этой случайной величины. Построить их графики. Найти вероятность попадания случайной величины в интервал $(0;3]$.

Ответ. $f_\xi(x) = \begin{cases} 0, & x \notin [2;5]; \\ \dfrac{1}{3}, & x \in [2;5], \end{cases}$ $F_\xi(x) = \begin{cases} 0, & x \leqslant 2; \\ \dfrac{x-2}{3}, & 2 < x \leqslant 5; \\ 1, & x > 5. \end{cases}$

Графики плотности и функции распределения изображены соответственно на рис. 18 и 19. $\mathbf{P}(\xi \in (0;3]) = \dfrac{1}{3}$.

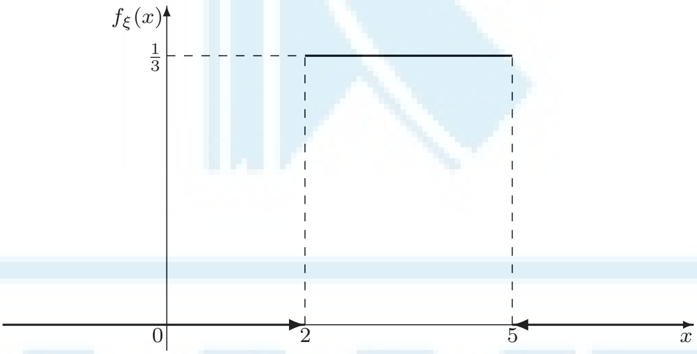


Рис. 18

Задача 12. Поезда метро идут с интервалом 2 мин. Считая, что время ξ ожидания поезда на остановке имеет равномерное распределение, найти аналитические выражения для плотности и функции

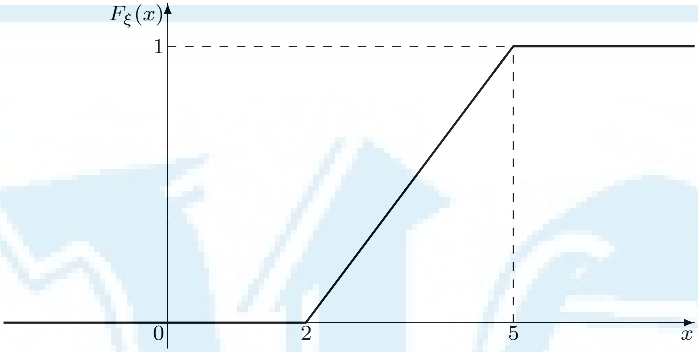


Рис. 19

распределения этой случайной величины. Построить их графики. Найти вероятность того, что время ожидания превысит 30 с.

Ответ. $f_\xi(x) = \begin{cases} 0, & x \notin [0;2]; \\ \dfrac{1}{2}, & x \in [0;2], \end{cases}$ $F_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ \dfrac{x}{2}, & 0 < x \leqslant 2; \\ 1, & x > 2. \end{cases}$ Графики плотности и функции распределения изображены соответственно на рис. 20 и 21. $\mathbf{P}(\xi > 30 \text{ с}) = \dfrac{3}{4}$.

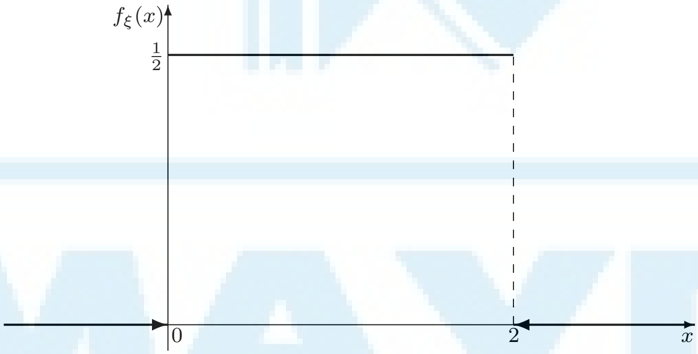


Рис. 20

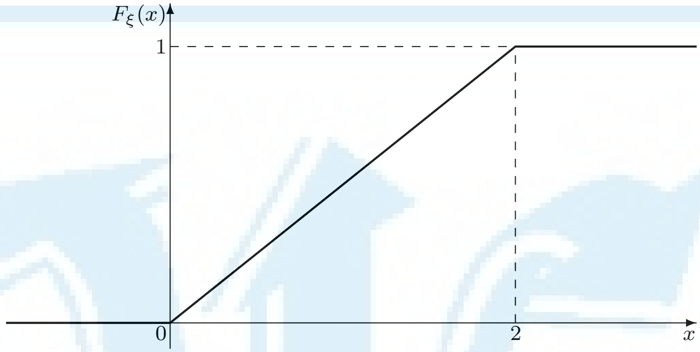


Рис. 21

Задача 13. Нормальный закон распределения случайной величины X задан функцией распределения

$$F(x) = \frac{1}{4\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t+1)^2}{32}} dt.$$

Построить график функции $f(x)$. Вычислить

$$\mathbf{P}\{-5 < X < 3\}, \qquad \mathbf{P}\{|X+1| < 12\}.$$

Ответ. $a = -1$; $\sigma = 4$. График $f(x)$ получается из графика плотности стандартного нормального распределения путём сжатия в 4 раза вдоль оси Oy и перенесения на 1 по левую сторону вдоль оси Ox (рис. 22). $\mathbf{P}\{-5 < X < 3\} = 2 \cdot \Phi(1) \approx 0{,}6318$; $\mathbf{P}\{|X+1| < 12\} = 2 \cdot \Phi(3) \approx 0{,}9973$.

Задача 14. Нормальный закон распределения случайной величины X задан функцией распределения

$$F(x) = \frac{1}{2\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t-3)^2}{8}} dt.$$

Построить график функции $f(x)$. Вычислить

$$\mathbf{P}\{1 < X < 5\}, \qquad \mathbf{P}\{|X-3| < 6\}.$$

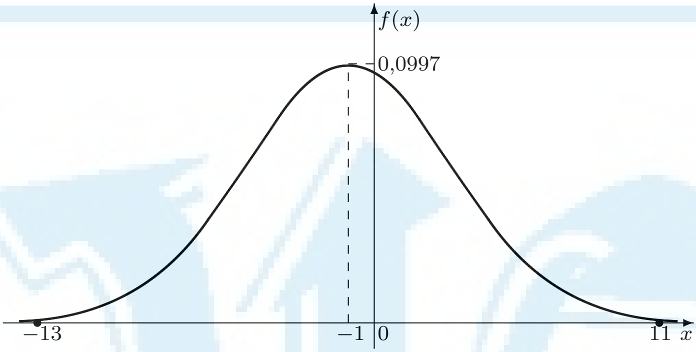


Рис. 22

Ответ. $a = 3$; $\sigma = 2$. График $f(x)$ получается из графика плотности стандартного нормального распределения путём сжатия в 2 раза вдоль оси Oy и перенесения на 3 по правую сторону вдоль оси Ox (рис. 23). $\mathbf{P}\{1 < X < 5\} \approx 0{,}6318$; $\mathbf{P}\{|X - 3| < 6\} \approx 0{,}9973$.

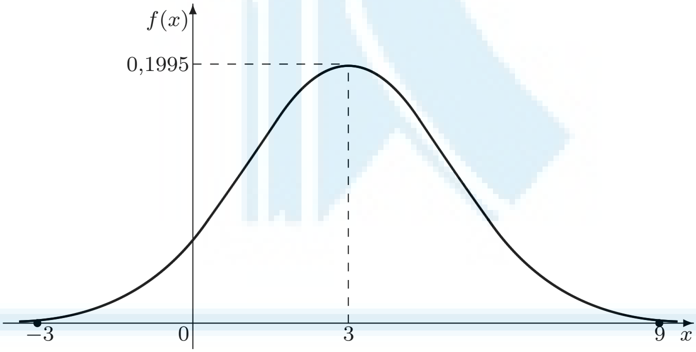


Рис. 23

Задача 15. Случайная величина ξ имеет нормальное распределение с параметрами a, σ. Записать выражение для плотности и функции распределения случайной величины ξ и найти вероятность того, что случайная величина ξ примет значение из промежутка $(\alpha, \beta]$, если:

а) $a = 0$, $\sigma = 1$, $\alpha = 0$, $\beta = 5$;
б) $a = 5$, $\sigma = 3$, $\alpha = 2$, $\beta = 8$;
в) $a = 9$, $\sigma = 2$, $\alpha = 3$, $\beta = 15$.

Ответ. а) $f_\xi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $F_\xi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$, $\mathbf{P}(\xi \in (0;5]) \approx 0{,}5$; б) $f_\xi(x) = \dfrac{1}{3\sqrt{2\pi}} e^{-\frac{(x-5)^2}{18}}$, $F_\xi(x) = \dfrac{1}{3\sqrt{2\pi}} \times$ $\times \displaystyle\int\limits_{-\infty}^{x} e^{-\frac{(t-5)^2}{18}} dt$, $\mathbf{P}(\xi \in (2;8]) \approx 0{,}6827$; в) $f_\xi(x) = \dfrac{1}{2\sqrt{2\pi}} e^{-\frac{(x-9)^2}{8}}$, $F_\xi(x) = \dfrac{1}{2\sqrt{2\pi}} \displaystyle\int\limits_{-\infty}^{x} e^{-\frac{(t-9)^2}{8}} dt$, $\mathbf{P}(\xi \in (3;15]) \approx 0{,}9973$.

Задача 16. Плотность нормально распределённой случайной величины ξ имеет вид $f_\xi(x) = c \cdot e^{-\frac{(x+3)^2}{8}}$. Найти параметры c, a и σ. Сравнить вероятности попадания случайной величины ξ в промежутки $(-9;3]$ и $(4;100]$.

Ответ. $c = \dfrac{1}{2\sqrt{2\pi}}$; $a = -3$; $\sigma = 2$; $\mathbf{P}(\xi \in (-9;3]) > \mathbf{P}(\xi \in$ $\in (4;100])$.

Задача 17. Случайная величина ξ имеет нормальное распределение с параметрами a, σ. В каждом из следующих пунктов найти интервал, в который попадает случайная величина ξ с практической достоверностью (с вероятностью 0,9973):
а) $a = 0$, $\sigma = 1$;
б) $a = 5$, $\sigma = 2$;
в) $a = 5$, $\sigma = 100$.

Ответ. а) $[-3;3]$; б) $[-1;11]$; в) $[-295;305]$.

Задача 18. Случайная величина ξ имеет показательное распределение с параметром $\lambda = 3$. Найти вероятность того, что случайная величина ξ примет значение из промежутка $[\alpha;\beta]$, если:
а) $\alpha = -2$, $\beta = -1$;
б) $\alpha = 2$, $\beta = 3$;
в) $\alpha = -2$, $\beta = 3$.

Ответ. а) $\mathbf{P}(\xi \in [-2;-1]) = 0$; б) $\mathbf{P}(\xi \in [-3;-1]) \approx 0{,}002355$; в) $\mathbf{P}(\xi \in [-2;3]) \approx 0{,}9999$.

Задача 19. Случайная величина ξ имеет экспоненциальное распределение. Вероятность того, что эта случайная величина примет значение из промежутка $[0;5]$, равняется 0,7. Найти вероятность того, что эта случайная величина примет значение из промежутка $[7;9]$.

Ответ. $\mathbf{P}(\xi \in [7;9]) \approx 0{,}0708$.

Раздел 8. Числовые характеристики случайных величин

Ключевые слова

Абсолютно
Величина
Дискретный
Дисперсия
Квадратический
Математический
Медиана
Мода
Непрерывный
Ожидание
Отклонение
Полимодальный
Распределение
Случайный
Средний
Унимодальный

К основным числовым характеристикам, которые описывают распределение случайной величины, относятся *математическое ожидание, дисперсия, среднее квадратическое отклонение, мода, медиана.*

Математическим ожиданием дискретной случайной величины ξ, которая принимает значения x_i с вероятностями p_i, называется число

$$\mathbf{M}(\xi) = \sum_i x_i p_i. \tag{4}$$

Если множество значений случайной величины ξ конечно, то математическое ожидание — обычная сумма произведений значений случайной величины на соответствующие вероятности. Если множество значений случайной величины ξ счётно, то будем требовать абсолютной сходимости ряда (4).

Математическим ожиданием абсолютно непрерывной случайной величины ξ с плотностью $f(x)$ называется число

$$\mathbf{M}(\xi) = \int\limits_{-\infty}^{\infty} x f(x) dx, \tag{5}$$

при условии абсолютной сходимости несобственного интеграла (5).

Математическое ожидание имеет такие свойства.

Свойство 1. Если $\mathbf{P}\{\xi = C\} = 1$, то $\mathbf{M}(C) = C$.

Свойство 2. $\mathbf{M}(C \cdot \xi) = C \cdot \mathbf{M}(\xi)$.

Свойство 3. $\mathbf{M}(\xi_1 + \xi_2) = \mathbf{M}(\xi_1) + \mathbf{M}(\xi_2)$.

Свойство 4. Если ξ_1 и ξ_2 — независимые случайные величины, то

$$\mathbf{M}(\xi_1 \cdot \xi_2) = \mathbf{M}(\xi_1) \cdot \mathbf{M}(\xi_2).$$

Дисперсией случайной величины ξ называют математическое ожидание квадрата отклонения случайной величины от её математического ожидания:

$$\mathbf{D}(\xi) = \mathbf{M}[\xi - \mathbf{M}(\xi)]^2. \tag{6}$$

Формула (6) эквивалентна такой формуле:

$$\mathbf{D}(\xi) = \mathbf{M}(\xi^2) - [\mathbf{M}(\xi)]^2. \tag{7}$$

Для дискретной случайной величины формулы (6) и (7) приобретают соответственно вид

$$\mathbf{D}(\xi) = \sum_i (x_i - \mathbf{M}(\xi))^2 p_i, \tag{8}$$

$$\mathbf{D}(\xi) = \sum_i x_i^2 p_i - [\mathbf{M}(\xi)]^2. \tag{9}$$

Для абсолютно непрерывной случайной величины формулы (6) и (7) приобретают соответственно вид

$$\mathbf{D}(\xi) = \int_{-\infty}^{\infty} (x - \mathbf{M}(\xi))^2 f(x)dx, \tag{10}$$

$$\mathbf{D}(\xi) = \int_{-\infty}^{\infty} x^2 f(x)dx - [\mathbf{M}(\xi)]^2. \tag{11}$$

В определениях считается, что ряд (8) и несобственный интеграл (10) абсолютно совпадают.

Дисперсия имеет такие свойства.

Свойство 1. $\mathbf{D}(\xi) \geqslant 0$.

Свойство 2. Якщо $\mathbf{P}\{\xi = C\} = 1$, то $\mathbf{D}(C) = 0$.

Свойство 3. $\mathbf{D}(C \cdot \xi) = C^2 \cdot \mathbf{D}(\xi)$.

Свойство 4. Если ξ_1 и ξ_2 — независимые случайные величины, то

$$\mathbf{D}(\xi_1 + \xi_2) = \mathbf{D}(\xi_1) + \mathbf{D}(\xi_2).$$

Средним квадратическим отклонением случайной величины ξ называют число

$$\sigma(\xi) = \sqrt{\mathbf{D}(\xi)}.$$

Дисперсия и среднее квадратическое отклонение характеризуют степень рассеивания значений случайной величины вокруг её математического ожидания.

Модой $\mathbf{Mo}(\xi)$ дискретной случайной величины называется наиболее вероятное её значение.

Распределение называется *унимодальным*, если оно имеет единственную моду (биномиальное, показательное и т. п.).

Распределение называется *полимодальным*, если оно имеет более одной моды (равномерное и т. п.).

Модой абсолютно непрерывной случайной величины называется точка максимума плотности распределения.

Медианой случайной величины называется такое число $\mathbf{Me}(\xi)$, для которого выполняется условие

$$\mathbf{P}\{\xi < \mathbf{Me}(\xi)\} = \mathbf{P}\{\xi > \mathbf{Me}(\xi)\}.$$

Если ξ имеет дискретное распределение, то, вообще говоря, $\mathbf{Me}(\xi)$ определяется неоднозначно. Например, если ξ — количество очков, которые выпали на игральном кубике, то в качестве $\mathbf{Me}(\xi)$ можно взять произвольное число из интервала $(3; 4)$.

В табл. 1 приведены математические ожидания, дисперсии и средние квадратические отклонения наиболее распространённых распределений вероятностей.

Пример 1. Случайная величина X имеет такое распределение:

X	1	2	3
P	0,2	0,6	0,2

Найти числовые характеристики распределения случайной величины X.

Решение. Математическое ожидание случайной величины X вычислим по формуле

$$\mathbf{M}(X) = \sum_i x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 = \\ = 1 \cdot 0{,}2 + 3 \cdot 0{,}6 + 5 \cdot 0{,}2 = 3.$$

Таблица 1

Числовые характеристики наиболее распространённых вероятностных распределений

№ п/п	Название распределения	$\mathbf{M}(\xi)$	$\mathbf{D}(\xi)$	$\sigma(\xi)$
I. Дискретные распределения				
1	Равномерное с параметром n	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$	$\sqrt{\dfrac{n^2-1}{12}}$
2	Биномиальное с параметрами n и p	np	$np(1-p)$	$\sqrt{np(1-p)}$
3	Показательное с параметром λ	λ	λ	$\sqrt{\lambda}$
4	Геометрическое с параметром p	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{\sqrt{1-p}}{p}$
II. Непрерывные распределения				
5	Равномерное с параметрами a, b	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{b-a}{2\sqrt{3}}$
6	Нормальное с параметрами a, σ	a	σ^2	σ
7	Показательное с параметром a	$\dfrac{1}{a}$	$\dfrac{1}{a^2}$	$\dfrac{1}{a}$

Дисперсию случайной величины X найдём по формуле

$$\mathbf{D}(X) = \sum_i x_i^2 p_i - [\mathbf{M}(X)]^2 = x_1^2 p_1 + x_2^2 p_2 + x_3^2 p_3 - 3^2 = \\ = 1^2 \cdot 0{,}2 + 3^2 \cdot 0{,}6 + 5^2 \cdot 0{,}2 - 3^2 = 1{,}6.$$

Среднее квадратическое отклонение случайной величины X

$$\sigma(X) = \sqrt{\mathbf{D}(X)} = \sqrt{1{,}6} \approx 0{,}2649.$$

Мода и медиана равны:

$$\mathbf{Mo}(X) = \mathbf{Me}(X) = 3. \qquad \square$$

Пример 2. Пусть случайная величина ξ равняется количеству номеров, угаданных игроком в лотереи «6 из 39».

Найти математическое ожидание, дисперсию и среднее квадратическое отклонение распределения случайной величины ξ.

Решение. Ряд распределения данной случайной величины построен в примере 2 раздела 6:

ξ	0	1	2	3	4	5	6
P	$\frac{1107568}{3262623}$	$\frac{1424016}{3262623}$	$\frac{613800}{3262623}$	$\frac{109120}{3262623}$	$\frac{7920}{3262623}$	$\frac{198}{3262623}$	$\frac{1}{3262623}$

По формуле

$$\mathbf{M}(\xi) = \sum_i x_i p_i$$

вычислим математическое ожидание случайной величины ξ:

$$\mathbf{M}(\xi) = 0 \cdot \frac{1107568}{3262623} + 1 \cdot \frac{1424016}{3262623} + 2 \cdot \frac{613800}{3262623} + 3 \cdot \frac{109120}{3262623} + \\ + 4 \cdot \frac{7920}{3262623} + 5 \cdot \frac{198}{3262623} + 6 \cdot \frac{1}{3262623} = \\ = \frac{12}{13} < 1.$$

Итак, среднее количество угаданных номеров приблизительно равняется единице.

По формуле

$$\mathbf{D}(\xi) = \sum_i x_i^2 p_i - [\mathbf{M}(\xi)]^2$$

найдём дисперсию случайной величины ξ:

$$\mathbf{D}(\xi) = 0^2 \cdot \frac{1107568}{3262623} + 1^2 \cdot \frac{1424016}{3262623} + 2^2 \cdot \frac{613800}{3262623} + 3^2 \cdot \frac{109120}{3262623} + \\ + 4^2 \cdot \frac{7920}{3262623} + 5^2 \cdot \frac{198}{3262623} + 6^2 \cdot \frac{1}{3262623} - \left(\frac{12}{13}\right)^2 = \\ = \frac{2178}{3211} \approx 0{,}6783.$$

Среднее квадратическое отклонение случайной величины ξ

$$\sigma(\xi) = \sqrt{\mathbf{D}(\xi)} = \sqrt{\frac{2178}{3211}} \approx 0{,}8236. \qquad \square$$

Пример 3. Случайная величина X задана плотностью распределения

$$f(x) = \begin{cases} 0, & x \leqslant 0, \\ \dfrac{1}{2\sqrt{x}}, & 0 < x \leqslant \dfrac{1}{4}, \\ 16x, & \dfrac{1}{4} < x \leqslant \dfrac{\sqrt{2}}{4}, \\ 0, & x > \dfrac{\sqrt{2}}{4}. \end{cases}$$

Найти числовые характеристики данной случайной величины.

Решение. Математическое ожидание случайной величины X вычислим по формуле $\mathbf{M}(X) = \int\limits_{-\infty}^{\infty} xf(x)dx$:

$$\mathbf{M}(X) = \int\limits_{-\infty}^{0} x \cdot 0dx + \int\limits_{0}^{\frac{1}{4}} x \cdot \frac{1}{2\sqrt{x}}dx + \int\limits_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}} x \cdot 16xdx + \int\limits_{\frac{\sqrt{2}}{4}}^{\infty} x \cdot 0dx = \\ = \frac{1}{2}\int\limits_{0}^{\frac{1}{4}} \sqrt{x}dx + 16\int\limits_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}} x^2dx = \frac{1}{2} \cdot \frac{x^{\frac{3}{2}}}{\frac{3}{2}}\Bigg|_0^{\frac{1}{4}} + 16 \cdot \frac{x^3}{3}\Bigg|_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}} =$$

$$= \frac{\sqrt{2}}{6} - \frac{1}{24} \approx 0{,}1940.$$

По формуле

$$\mathbf{D}(X) = \int\limits_{-\infty}^{\infty} x^2 f(x)dx - [\mathbf{M}(\xi)]^2$$

найдём дисперсию случайной величины X:

$$\mathbf{D}(X) = \left(\int\limits_{-\infty}^{0} x^2 \cdot 0dx + \int\limits_{0}^{\frac{1}{4}} x^2 \cdot \frac{1}{2\sqrt{x}}dx + \int\limits_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}} x^2 \cdot 16xdx + \right.$$

$$\left. + \int\limits_{\frac{\sqrt{2}}{4}}^{\infty} x^2 \cdot 0dx \right) - \left(\frac{\sqrt{2}}{6} - \frac{1}{24} \right)^2 =$$

$$= \frac{1}{2} \int\limits_{0}^{\frac{1}{4}} x\sqrt{x}dx + 16 \int\limits_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}} x^3 dx - \left(\frac{\sqrt{2}}{6} - \frac{1}{24} \right)^2 =$$

$$= \frac{1}{2} \cdot \left. \frac{x^{\frac{5}{2}}}{\frac{5}{2}} \right|_0^{\frac{1}{4}} + 16 \cdot \left. \frac{x^4}{4} \right|_{\frac{1}{4}}^{\frac{\sqrt{2}}{4}} - \frac{11}{192} + \frac{\sqrt{2}}{72} =$$

$$= \frac{\sqrt{2}}{72} - \frac{1}{240} \approx 0{,}0155.$$

Среднее квадратическое отклонение случайной величины X

$$\sigma(\xi) = \sqrt{\mathbf{D}(\xi)} = \sqrt{\frac{\sqrt{2}}{72} - \frac{1}{240}} \approx 0{,}1244. \qquad \square$$

Задачи к разделу 8

Задача 1. Пусть ξ — случайная величина, которая равняется количеству очков, которые выпадут при подбрасывании кубика. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины ξ.

Ответ. $\mathbf{M}(\xi) = 3{,}5$; $\mathbf{D}(\xi) = \dfrac{35}{12}$; $\sigma(\xi) = \sqrt{\dfrac{35}{12}} \approx 1{,}71$.

Задача 2. В ящике лежит 5 пронумерованных шариков (номера изменяются от 1 до 5). Наугад вынимают шарик. Случайная величина ξ — номер этого шарика. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины ξ.

Ответ. $\mathbf{M}(\xi) = 3$; $\mathbf{D}(\xi) = 2$; $\sigma(\xi) = \sqrt{2} \approx 1{,}41$.

Задача 3. Симметричную монету подбрасывают 3 раза. Случайная величина X — количество «аверсов», которые при этом выпадали. Найти математическое ожидание и дисперсию случайной величины X.

Ответ. $\mathbf{M}(X) = \dfrac{3}{2}$; $\mathbf{D}(X) = \dfrac{3}{4}$.

Задача 4. Найти числовые характеристики случайной величины X, которая равняется количеству мальчиков в семье с тремя детьми (считая рождение мальчика и девочки равновероятными событиями).

Ответ. $\mathbf{M}(X) = \dfrac{3}{2}$; $\mathbf{D}(X) = \dfrac{3}{4}$; $\sigma(X) = \dfrac{\sqrt{3}}{2} \approx 0{,}87$.

Задача 5. Стрелок производит по одному выстрелу по четырём мишеням. Найти числовые характеристики случайной величины X, которая равняется количеству попаданий, считая, что вероятность попадания при одном выстреле 0,7.

Ответ. $\mathbf{M}(X) = 2{,}8$; $\mathbf{D}(X) = 0{,}84$; $\sigma(X) = \sqrt{0{,}84} \approx 0{,}92$.

Задача 6. Студент знает 20 вопросов из 25. Наугад он вынимает 5 вопросов. Найти числовые характеристики случайной величины X, которая равняется количеству вопросов, на которые студент знает ответ.

Указание. Воспользуйтесь числовыми характеристиками гипергеометрического распределения с параметрами m, n, N:

$$\mathbf{M}(X) = \frac{mn}{N}, \qquad \mathbf{D}(X) = \frac{mn(N-m)(N-n)}{N^2(N-1)},$$

$$\sigma(X) = \sqrt{\frac{mn(N-m)(N-n)}{N^2(N-1)}}.$$

Ответ. $\mathbf{M}(X) = 4$; $\mathbf{D}(X) = \frac{2}{3}$; $\sigma(X) = \sqrt{\frac{2}{3}} \approx 0{,}82$.

Задача 7. В ящике лежат шары — 7 зелёных и 3 жёлтых. Наугад выбирают шар, фиксируют его цвет и шар возвращают в ящик. Эксперимент повторяют 4 раза. Найти математическое ожидание и дисперсию случайной величины X, которая равняется количеству зелёных шаров, которые появились при вынимании.

Ответ. $\mathbf{M}(X) = 2{,}8$; $\mathbf{D}(X) = 0{,}84$; $\sigma(X) = \sqrt{0{,}84} \approx 0{,}92$.

Задача 8. В группе из 20 студентов три отличника. Случайно выбирают 4 студента. Найти числовые характеристики случайной величины ξ, которая равняется количеству отличников среди выбранных студентов.

Ответ. $\mathbf{M}(\xi) = 0{,}6$; $\mathbf{D}(\xi) = \frac{204}{475} \approx 0{,}43$; $\sigma(\xi) = \sqrt{\frac{204}{475}} \approx 0{,}66$.

Задача 9. Среди 10 деталей есть 2 бракованные. Для контроля наугад выбирают 3 детали. Найти числовые характеристики случайной величины ξ, которая равняется количеству бракованных деталей среди выбранных.

Ответ. $\mathbf{M}(\xi) = 0{,}6$; $\mathbf{D}(\xi) = \frac{28}{75} \approx 0{,}37$; $\sigma(\xi) = \sqrt{\frac{28}{75}} \approx 0{,}61$.

Задача 10. Найти числовые характеристики случайной величины Y, которая равняется количеству угаданных номеров в лотереи «5 из 36».

Ответ. $\mathbf{M}(Y) \approx 0{,}69$; $\mathbf{D}(Y) \approx 0{,}53$; $\sigma(Y) \approx 0{,}73$.

Задача 11. Кубик подбрасывают до тех пор, пока не выпадет шестёрка. X — случайная величина, которая равняется необходимому количеству подбрасываний. Найти числовые характеристики случайной величины X и

$$\mathbf{P}\left\{|X - \mathbf{M}(X)| < 3\sigma(X)\right\}.$$

Ответ. $\mathbf{M}(X) = 6$; $\mathbf{D}(X) = 30$; $\sigma(X) = \sqrt{30} \approx 5{,}48$; $\mathbf{P}\{|X - \mathbf{M}(X)| < 3\sigma(X)\} = \mathbf{P}\{1 < X < 22\} = 1 - \dfrac{1}{6^{22}} \approx 1$.

Задача 12. Игрок играет в рулетку до первого своего выигрыша, постоянно ставя на красное и утраивая ставку. X — случайная величина, которая равняется количества попыток, Y — случайная величина, которая равняется выигрышу игрока. Найти числовые характеристики случайных величин X и Y, считая, что игрок имеет неограниченное количество денег и размер ставки не ограничен правилами.

Ответ. $\mathbf{M}(X) = \dfrac{37}{18} \approx 2{,}06$; $\mathbf{D}(X) = \dfrac{703}{324} \approx 2{,}17$; $\sigma(X) = \sqrt{\dfrac{703}{324}} \approx 1{,}47$. Случайная величина $Y = a$, если $X = 1$, и $Y = 2a \cdot 3^{n-2} \cdot X$, если $X = n$, $n = 2, 3, \ldots$, где a — размер первой ставки; $\mathbf{M}(Y) = +\infty$.

Задача 13. Магазин получил 10000 бутылок минеральной воды. Вероятность того, что при перевозке бутылка окажется разбитой, равняется 0,0004. Найти числовые характеристики случайной величины X, которая равняется количеству разбитых бутылок.

Ответ. $\mathbf{M}(X) = 4$; $\mathbf{D}(X) = 4$; $\sigma(X) = 2$.

Задача 14. Учебник издан тиражом 10000 экземпляров. Вероятность того, что сброшюрованный экземпляр окажется некачественным, равняется 0,001. Найти числовые характеристики случайной величины, которая равняется количеству бракованных экземпляров.

Ответ. $\mathbf{M}(X) = 10$; $\mathbf{D}(X) = 10$; $\sigma(X) = \sqrt{10} \approx 3{,}16$.

Задача 15. Автоматическая телефонная станция обслуживает 10000 телефонных номеров. Вероятность того, что в течение 1 мин на АТС поступит вызов от абонента, равняется 0,0004. Найти числовые характеристики случайной величины X, которая равняется количеству вызовов, поступивших на АТС в течение 1 мин.

Ответ. $\mathbf{M}(X) = 4$; $\mathbf{D}(X) = 4$; $\sigma(X) = 2$.

Задача 16. Вероятность попадания в муху на лету из рогатки равняется 0,001. Производится 2000 выстрелов. Найти числовые характеристики случайной величины Y, которая равняется количеству убитых мух.

Ответ. $\mathbf{M}(Y) = 2$; $\mathbf{D}(Y) = 2$; $\sigma(Y) = \sqrt{2} \approx 1{,}41$.

Задача 17. Случайная величина ξ равномерно распределена на отрезке $[1; 11]$. Найти числовые характеристики случайной величины ξ.

Ответ. $\mathbf{M}(X) = 6$; $\mathbf{D}(X) = \dfrac{25}{3}$; $\sigma(X) = \sqrt{\dfrac{25}{3}} \approx 2{,}89$.

Задача 18. Случайная величина ξ равномерно распределена на $[7; a]$, причём плотность на этом промежутке равняется $\dfrac{1}{20}$. Указать значения параметра a и найти числовые характеристики случайной величины ξ.

Ответ. $a = 27$; $\mathbf{M}(\xi) = 17$; $\mathbf{D}(\xi) = \dfrac{100}{3}$; $\sigma(\xi) = \sqrt{\dfrac{100}{3}} \approx 5{,}77$.

Задача 19. Случайная величина ξ задана плотностью

$$f(x) = \begin{cases} 0, & x \leqslant 0, \\ \dfrac{2}{e^{2x}}, & x > 0. \end{cases}$$

Найти числовые характеристики случайной величины ξ.

Ответ. $\mathbf{M}(\xi) = \dfrac{1}{2}$; $\mathbf{D}(\xi) = \dfrac{1}{4}$; $\sigma(\xi) = \dfrac{1}{2}$.

Задача 20. Случайная величина ξ задана функцией распределения

$$F(x) = \begin{cases} 0, & x \leqslant 2, \\ (x-2)^2, & 2 < x \leqslant 3, \\ 1, & x > 3. \end{cases}$$

Найти числовые характеристики случайной величины ξ.

Ответ. $\mathbf{M}(\xi) = \dfrac{8}{3}$; $\mathbf{D}(\xi) = \dfrac{1}{18}$; $\sigma(\xi) = \dfrac{\sqrt{2}}{6} \approx 0{,}24$.

Задача 21. Плотность случайной величины X задана графически (рис. 24).

Найти значения параметра a и числовые характеристики случайной величины X.

Ответ. $a = \dfrac{2}{7}$; $\mathbf{M}(X) = -\dfrac{2}{3}$; $\mathbf{D}(X) = \dfrac{43}{18} \approx 2{,}39$; $\sigma(X) = \sqrt{\dfrac{43}{18}} \approx$ $\approx 1{,}55$.

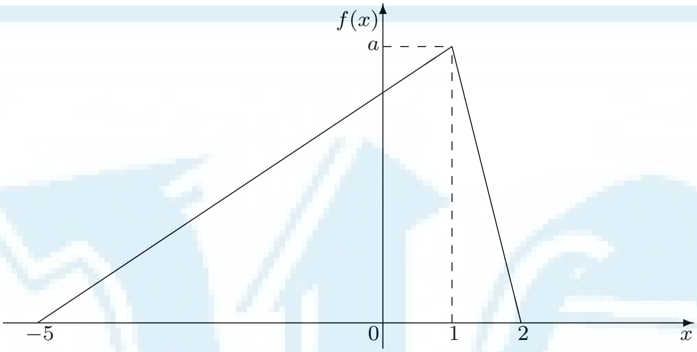


Рис. 24

Задача 22. В середине круга радиуса R случайным образом выбирают точку. Вероятность попадания точки в любую область, которая помещается в круге, пропорциональна площади этой области. Найти функцию распределения и дисперсию случайной величины ξ, которая равняется расстоянию от точки до центра круга.

Ответ. $F_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ \dfrac{x}{2}, & 0 < x \leqslant R; \\ 1, & x > R, \end{cases}$ $\mathbf{D}(\xi) = \dfrac{R^2}{12}$.

Задача 23. Коробки с шоколадом пакуются автоматически, их средняя масса равняется 1,06 кг. Считая, что масса коробок распределена по нормальному закону, найти стандартное отклонение, если 5 % коробок имеют массу меньше 1 кг.

Ответ. $\sigma = \dfrac{0{,}06}{2{,}58} \approx 0{,}023$.

Раздел 9. Системы двух случайных величин

Ключевые слова

Абсолютно
Величина
Дискретный
Закон
Ковариация
Корреляционный
Корреляция
Коэффициент
Момент
Независимый
Некоррелированный
Непрерывный
Плотность
Распределение
Случайный
Условный
Функция

Если на одном и том же пространстве элементарных событий Ω задано две случайные величины ξ и ψ, то считают, что задана случайная величина (ξ, ψ) (или случайный двумерный вектор).

При изучении системы случайных величин, вообще говоря, недостаточно знать информацию о каждой случайной составляющей. Необходимо учитывать еще и зависимость между ними.

Функцией распределения двумерной случайной величины (ξ, ψ) называется функция двух переменных

$$F(x, y) = \mathbf{P}\{\xi < x, \psi < y\}.$$

Функция распределения геометрически определяет вероятность попадания случайной точки (ξ, ψ) в бесконечный квадрант с вершиной в точке (x, y), который находится левее и ниже от нее.

Функция распределения двумерной случайной величины имеет такие свойства.

Свойство 1. $\lim\limits_{\substack{x\to+\infty\\ y\to+\infty}} F(x, y) = 1.$

Свойство 2. Для любой системы случайных величин

$$\lim_{x\to-\infty} F(x, y) = \lim_{y\to-\infty} F(x, y) = \lim_{\substack{x\to-\infty\\ y\to-\infty}} F(x, y) = 0.$$

Свойство 3. а) Если $x_1 < x_2$, то

$$F(x_1, y) \leqslant F(x_2, y)$$

для всех $y \in \mathbb{R}$.

б) Если $y_1 < y_2$, то

$$F(x, y_1) \leqslant F(x, y_2)$$

для всех $x \in \mathbb{R}$.

Свойство 4. Если случайная величина ξ имеет функцию распределения $F_\xi(x)$, а случайная величина ψ — функцию распределения $F_\psi(x)$, то

$$\lim_{x \to +\infty} F(x, y) = F_\psi(y), \qquad \lim_{y \to +\infty} F(x, y) = F_\xi(x).$$

Свойство 5. Вероятность попадания случайной точки (ξ, ψ) в прямоугольник со сторонами, параллельными координатным осям, вычисляется по формуле

$$\mathbf{P}\{a \leqslant \xi < b; c \leqslant \psi < d\} = F(b, d) - F(a, d) - F(b, c) + F(a, c).$$

Как и в одномерном случае, двумерные случайные величины можно разделить на два класса: *дискретные* и *непрерывные*.

Двумерная случайная величина (ξ, ψ) называется *дискретной*, если и случайная величина ξ, и случайная величина ψ являются дискретными случайными величинами. В этом случае случайную величину (ξ, ψ) удобно задавать с помощью таблицы:

$\psi \backslash \xi$	ξ_1	ξ_2	$\ldots$	ξ_n
ψ_1	p_{11}	p_{12}	$\ldots$	p_{1n}
ψ_2	p_{21}	p_{22}	$\ldots$	p_{2n}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
ψ_m	p_{m1}	p_{m2}	$\ldots$	p_{mn}

где p_{ij} — вероятности того, что одновременно будут выполняться равенства $\psi = \psi_i$, $\xi = \xi_j$. Очевидно, при этом

$$\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} = 1.$$

Для дискретной случайной величины (ξ, ψ) функция распределения приобретает вид суммы

$$F(x, y) = \sum_{(i,j):\,\begin{cases}\xi_j < x;\\ \psi_i < y.\end{cases}} p_{ij},$$

т. е. суммирование выполняется по всем наборам (i, j), для которых одновременно $\xi_j < x$ и $\psi_i < y$.

Пользуясь приведенной выше таблицей, легко определить распределения каждой случайной величины отдельно по формулам

$$\mathbf{P}\{\xi = \xi_j\} = p_{1j} + p_{2j} + \ldots + p_{mj}, \qquad j = 1, 2, \ldots, n;$$
$$\mathbf{P}\{\psi = \psi_i\} = p_{i1} + p_{i2} + \ldots + p_{in}, \qquad i = 1, 2, \ldots, m.$$

Двумерная случайная величина называется *непрерывной*, если вероятность попадания случайной точки (ξ, ψ) в произвольную фиксированную точку равняется нулю, т. е.

$$\mathbf{P}\{(\xi, \psi) = (x_0, y_0)\} = 0$$

для любых (x_0, y_0).

Случайная величина (ξ, ψ) называется *абсолютно непрерывной*, если существует такая функция $f(x, y)$, для которой выполняется равенство

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y)\,dx\,dy.$$

Функция $f(x, y)$ называется *плотностью* распределения двумерной случайной величины (ξ, ψ).

Плотность $f(x, y)$ имеет такие свойства.

Свойство 1. $f(x, y) = F''_{xy}(x, y)$.

Свойство 2. $f(x, y) \geqslant 0$.

Свойство 3. $\displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\,dx\,dy = 1.$

Свойство 4. Вероятность попадания случайной точки (ξ, ψ) в плоскую область $\mathcal{D}$

$$\mathbf{P}\{(\xi, \psi) \in \mathcal{D}\} = \iint\limits_{\mathcal{D}} f(x, y) dx dy.$$

Случайные величины ξ и ψ называются *независимыми*, если

$$F(x, y) = F_\xi(x) \cdot F_\psi(y).$$

Для абсолютно непрерывных случайных величин это условие равносильно условию

$$f(x, y) = f_\xi(x) \cdot f_\psi(y),$$

где $f_\xi(x)$ и $f_\psi(y)$ — плотность распределения случайных величин ξ и ψ соответственно.

Плотности $f_\xi(x)$ и $f_\psi(y)$ можно вычислить по таким формулам:

$$f_\xi(x) = \int\limits_{-\infty}^{+\infty} f(x, y) dy = F'_\xi(x);$$

$$f_\psi(y) = \int\limits_{-\infty}^{+\infty} f(x, y) dx = F'_\psi(y).$$

Если случайные величины ξ и ψ зависимы, то для нахождения закона распределения двумерной случайной величины (ξ, ψ) недостаточно знать законы распределения каждой из случайных величин ξ и ψ.

Условным законом распределения называется распределение одной случайной величины, найденный при условии, что другая случайная величина системы приняла некоторое фиксированное значение.

Условный закон можно задавать как функцией распределения $F(x/y)$, так и плотностью $f(x/y)$ (для абсолютно непрерывного случая), где

$$f(x/y) = \frac{f(x, y)}{f_\psi(y)}; \qquad f(y/x) = \frac{f(x, y)}{f_\xi(x)}.$$

Итак,

$$f(x,y) = f_\xi(x) \cdot f(y/x) = f_\psi(y) \cdot f(x/y).$$

Условная плотность обладает всеми свойствами плотности одномерного закона распределения.

Числовые характеристики условных законов распределения вычисляют так же, как и для безусловных, но используют условную плотность.

Для описания степени зависимости случайных величин ξ и ψ используют такие числовые характеристики, как *ковариация* и *коэффициент корреляции.*

Ковариацией (корреляционным моментом) системы случайных величин называют число

$$\operatorname{cov}(\xi,\psi) = \mathbf{M}\left[(\xi - \mathbf{M}\xi)(\psi - \mathbf{M}\psi)\right].$$

Если случайные величины ξ и ψ дискретные, то

$$\operatorname{cov}(\xi,\psi) = \sum_i \sum_j \left(x_i - \mathbf{M}\xi\right)\left(y_j - \mathbf{M}\psi\right) p_{ij}.$$

Если случайные величины ξ и ψ непрерывные, то

$$\operatorname{cov}(\xi,\psi) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} \left(x - \mathbf{M}\xi\right)\left(y - \mathbf{M}\psi\right) dxdy.$$

Коэффициентом корреляции системы случайных величин ξ и ψ называют число

$$r(\xi,\psi) = \frac{\operatorname{cov}(\xi,\psi)}{\sigma(\xi)\sigma(\psi)}.$$

Если ξ и ψ — независимые случайные величины, то

$$\operatorname{cov}(\xi,\psi) = r(\xi,\psi) = 0.$$

Обратное утверждение неправильное. Существуют зависимые случайные величины, для которых ковариация и коэффициент корреляции равняются нулю.

Для произвольных случайных величин

$$-1 \leqslant r(\xi, \psi) \leqslant 1.$$

Чем ближе $|r(\xi, \psi)|$ к единице, тим более сильный степень зависимости между ξ и ψ. Если $r(\xi, \psi) = 0$, то случайные величины называются *некоррелированными*.

Пример 1. Из коробки, в которой находятся три красных и два синих шара, наугад без возвращений вынимают последовательно шары до первого появления синего шара. Далее шары вынимают до первого появления красного шара.

Необходимо описать закон распределения вероятностей системы случайных величин (X, Y), где X — количество шаров, взятых из коробки до первого появления синего шара; Y — количество шаров, взятых из коробки до первого появления красного шара после того, как первый синий шар был вынут из коробки.

Составить отдельные законы распределения для случайных величин X и Y.

Решение. Случайная величина X может принимать значения 1, 2, 3, 4, а случайная величина Y — значения 0, 1, 2. Вычислим $p_{i,j}$, $i = \overline{1,4}$, $j = 0,1,2$ — вероятности того, что $X = i$, $Y = j$:

$$p_{11} = \mathbf{P}(X = 1, Y = 1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10};$$

$$p_{12} = \mathbf{P}(X = 1, Y = 2) = \frac{2}{5} \cdot \frac{1}{4} \cdot 1 = \frac{1}{10};$$

$$p_{21} = \mathbf{P}(X = 2, Y = 1) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = \frac{2}{10};$$

$$p_{22} = \mathbf{P}(X = 2, Y = 2) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} \cdot 1 = \frac{1}{10};$$

$$p_{31} = \mathbf{P}(X = 3, Y = 1) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{10};$$

$$p_{32} = \mathbf{P}(X = 3, Y = 2) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{10};$$

$$p_{41} = \mathbf{P}(X = 4, Y = 1) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} \cdot 0 = 0;$$

$$p_{42} = \mathbf{P}(X = 4, Y = 2) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} \cdot 0 = 0;$$

$$p_{10} = \mathbf{P}(X = 1, Y = 0) = 0;$$
$$p_{20} = \mathbf{P}(X = 2, Y = 0) = 0;$$
$$p_{30} = \mathbf{P}(X = 3, Y = 0) = 0;$$
$$p_{40} = \mathbf{P}(X = 4, Y = 0) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} \cdot 1 = \frac{1}{10}.$$

Полученные значения запишем в таблицу распределения вероятностей системы случайных величин (X, Y) (табл. 2).

Таблица 2

Y	X				$\sum_{i=1}^{4} p_{ij}$
	1	2	3	4	
0	0	0	0	$\frac{1}{10}$	$\frac{1}{10}$
1	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	0	$\frac{6}{10}$
2	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	0	$\frac{3}{10}$
$\sum_{j=1}^{3} p_{ij}$	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\sum_{i=1}^{4}\sum_{j=1}^{3} p_{ij} = 1$

Пользуясь табл. 2, довольно легко записать безусловные законы распределения для случайных величин X и Y:

X	1	2	3	4
$\mathbf{P}$	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$

;

Y	0	1	2
$\mathbf{P}$	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$

. □

Пример 2. Задана таблица распределения двумерной случайной величины (X, Y):

Y \ X	1	2	3
−1	0,2	0,1	0,3
0	0,15	0,15	0,1

Необходимо:

а) найти безусловные законы распределения двумерной случайной величины (X, Y);

б) найти условный закон распределения X при условии, что $Y = -1$;

в) найти условный закон распределения Y при условии, что $X = 3$;

г) выяснить, зависимы или нет случайные величины X и Y.

Решение. а) Вычислим вероятности $\mathbf{P}(X = i)$ $(i = 1, 2, 3)$ и $\mathbf{P}(Y = j)$ $(j = -1, 0)$:

$$\begin{aligned} \mathbf{P}(x = 1) &= 0{,}2 + 0{,}15 = 0{,}35; \\ \mathbf{P}(x = 2) &= 0{,}1 + 0{,}15 = 0{,}25; \\ \mathbf{P}(x = 3) &= 0{,}3 + 0{,}1 = 0{,}4; \\ \mathbf{P}(y = -1) &= 0{,}2 + 0{,}1 + 0{,}3 = 0{,}6; \\ \mathbf{P}(y = 0) &= 0{,}15 + 0{,}15 + 0{,}1 = 0{,}4. \end{aligned}$$

Запишем безусловные законы распределения X и Y:

X	1	2	3
$\mathbf{P}$	0,35	0,25	0,4

;

Y	-1	0
$\mathbf{P}$	0,6	0,4

.

б) Поскольку

$$\begin{aligned} \mathbf{P}(X = 1/Y = -1) &= \frac{0{,}2}{0{,}2 + 0{,}1 + 0{,}3} = \frac{1}{3}; \\ \mathbf{P}(X = 2/Y = -1) &= \frac{0{,}1}{0{,}2 + 0{,}1 + 0{,}3} = \frac{1}{6}; \\ \mathbf{P}(X = 3/Y = -1) &= \frac{0{,}3}{0{,}2 + 0{,}1 + 0{,}3} = \frac{1}{2}, \end{aligned}$$

условный закон распределения X при условии, что $Y = -1$, будет иметь такой вид:

X	1	2	3
$\mathbf{P}(X/Y = -1)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$

.

в) Поскольку

$$\mathbf{P}(Y=-1/X=3)=\frac{0{,}3}{0{,}3+0{,}1}=\frac{3}{4};$$
$$\mathbf{P}(Y=0/X=3)=\frac{0{,}1}{0{,}3+0{,}1}=\frac{1}{4},$$

условный закон распределения Y при условии, что $X=3$, будет иметь такой вид:

Y	-1	0
$\mathbf{P}(Y/X=3)$	$\frac{3}{4}$	$\frac{1}{4}$

г) Тот факт, что безусловный закон распределения величины X не совпадает с условным законом распределения этой величины, свидетельствует о том, что величины X и Y зависимы. $\square$

Пример 3. Система случайных величин (X, Y) имеет такое распределение вероятностей:

Y \ X	0	1
0	0,2	0,15
1	0,15	0,15
2	0,1	0,25

Найти:

а) математические ожидания $\mathbf{M}(X)$ и $\mathbf{M}(Y)$;

б) дисперсии $\mathbf{D}(X)$ и $\mathbf{D}(Y)$.

Решение. а) Чтобы определить математическое ожидание $\mathbf{M}(X)$, можно воспользоваться формулой

$$\mathbf{M}(X)=\sum_i\sum_j x_i p_{ij},$$

откуда

$$\mathbf{M}(X)=0\cdot 0{,}2+0\cdot 0{,}15+0\cdot 0{,}1+1\cdot 0{,}15+1\cdot 0{,}15+1\cdot 0{,}25=0{,}55.$$

Такой же результат получим, если запишем безусловный закон распределения случайной величины X:

X	0	1
$\mathbf{P}$	0,45	0,55

и дальше воспользуемся формулой

$$\mathbf{M}(X) = \sum_i x_i p_i,$$

откуда

$$\mathbf{M}(X) = 0 \cdot 0{,}45 + 1 \cdot 0{,}55 = 0{,}55.$$

Аналогично вычисляем математическое ожидание:

$$\mathbf{M}(Y) = 0 \cdot 0{,}2 + 0 \cdot 0{,}5 + 1 \cdot 0{,}15 + 1 \cdot 0{,}15 + 2 \cdot 0{,}1 + 2 \cdot 0{,}25 = 1,$$

или из безусловного закона распределения

Y	0	1	2
$\mathbf{P}$	0,35	0,3	0,35

получаем

$$\mathbf{M}(Y) = 0 \cdot 0{,}35 + 1 \cdot 0{,}3 + 2 \cdot 0{,}35 = 1.$$

б) Чтобы найти дисперсию $\mathbf{D}(X)$, можно воспользоваться формулой

$$\mathbf{D}(X) = \sum_i \sum_j (x_i - \mathbf{M}(X))^2 p_{ij},$$

откуда

$$\begin{aligned}\mathbf{D}(X) = (0 - 0{,}55)^2 \cdot 0{,}2 + (0 - 0{,}55)^2 \cdot 0{,}15 + (0 - 0{,}55)^2 \cdot 0{,}1 +\\ + (1 - 0{,}55)^2 \cdot 0{,}15 + (1 - 0{,}55)^2 \cdot 0{,}15 + (1 - 0{,}55)^2 \cdot 0{,}25 =\\ = 0{,}0605 + 0{,}045375 + 0{,}03025 + 0{,}030375 + 0{,}030375 +\\ + 0{,}050625 = 0{,}2475.\end{aligned}$$

Такой же результат получим, если воспользуемся безусловным законом распределения вероятностей случайной величины X и формулой

$$\mathbf{D}(X) = \sum_i (x_i - \mathbf{M}(X))^2 p_i,$$

откуда

$$\mathbf{D}(X) = (0 - 0{,}55)^2 \cdot 0{,}45 + (1 - 0{,}55)^2 \cdot 0{,}55 = \\ = 0{,}136125 + 0{,}111375 = 0{,}2475.$$

Аналогично вычисляется дисперсия $\mathbf{D}(Y)$. Воспользуемся безусловным законом распределения вероятностей случайной величины Y и формулой

$$\mathbf{D}(Y) = \sum_i (y_i - \mathbf{M}(Y))^2 p_i.$$

Получим

$$\mathbf{D}(Y) = (0 - 1)^2 \cdot 0{,}35 + (1 - 1)^2 \cdot 0{,}3 + (2 - 1)^2 \cdot 0{,}35 = \\ = 0{,}35 + 0 + 0{,}35 = 0{,}7.$$

□

Пример 4. По мишени производится один выстрел. Вероятность попадания равняется p. Рассматривается X — количество попаданий, Y — количество промахов.

Построить функцию распределения $F(x, y)$ двумерной случайной величины (X, Y).

Решение. Найдём вероятности:

$$\mathbf{P}(X = 0/Y = 0) = 0; \qquad \mathbf{P}(X = 1/Y = 0) = p;$$
$$\mathbf{P}(X = 0/Y = 1) = 1 - p; \qquad \mathbf{P}(X = 1/Y = 1) = 0.$$

Построим таблицу распределения вероятностей системы случайных величин (X, Y):

$Y \backslash X$	0	1
0	0	p
1	$1 - p$	0

Функцию распределения вероятностей системы случайных величин удобно представить в виде таблицы:

$F(x,y)$	$x \leqslant 0$	$0 < x \leqslant 1$	$x > 1$
$y \leqslant 0$	0	0	0
$0 < y \leqslant 1$	0	0	p
$y > 1$	0	$1 - p$	1

□

Пример 5. Система дискретных случайных величин (X,Y) задана таблицей распределения вероятностей (табл. 3).

Таблица 3

Y	X				$\sum\limits_{i=1}^{4} p_{ij}$
	-10	-8	-6	-4	
10	0,023	0,027	0,05	0,1	0,2
20	0,05	0,1	0,025	0,025	0,2
30	0,05	0,05	0,025	0,025	0,15
40	0,027	0,023	0,05	0,35	0,45
$\sum\limits_{j=1}^{4} p_{ij}$	0,15	0,2	0,15	0,5	$\sum\limits_{i=1}^{4}\sum\limits_{j=1}^{4} p_{ij} = 1$

Вычислить математические ожидания $\mathbf{M}(X)$ и $\mathbf{M}(Y)$, средние квадратические отклонения $\sigma(X)$ и $\sigma(Y)$, корреляционный момент $\mathrm{cov}(X,Y)$, коэффициент корреляции $r(X,Y)$, а также условные математические ожидания $\mathbf{M}(X/Y = 40)$ и $\mathbf{M}(Y/X = -8)$.

Решение. Построим безусловные законы распределения X и Y:

X	-10	-8	-6	-4
$\mathbf{P}$	0,15	0,2	0,15	0,5

;

Y	10	20	30	40
$\mathbf{P}$	0,2	0,2	0,15	0,45

.

Определим математические ожидания по формулам

$$\mathbf{M}(X) = \sum_{i=1}^{4} x_i p_i, \qquad \mathbf{M}(Y) = \sum_{i=1}^{4} y_i p_i.$$

Итак,

$$\mathbf{M}(X) = -1{,}5 - 1{,}6 - 0{,}9 - 2 = -5,$$
$$\mathbf{M}(Y) = 2 + 4 + 4{,}5 + 18 = 28{,}5.$$

Среднее квадратическое отклонение

$$\sigma(X) = \sqrt{\mathbf{D}(X)},$$

где $\mathbf{D}(X)$ — дисперсия случайной величины X:

$$\mathbf{D}(X) = (-10+5)^2 \cdot 0{,}15 + (-8+5)^2 \cdot 0{,}2 + (-6+5)^2 \cdot 0{,}15 + \\ + (-4+5)^2 \cdot 0{,}5 = 6{,}2.$$

Итак,

$$\sigma(X) = \sqrt{6{,}2} \approx 2{,}49.$$

Аналогично вычисляем $\mathbf{D}(Y)$ и $\sigma(Y)$:

$$\mathbf{D}(Y) = (10-28{,}5)^2 \cdot 0{,}2 + (20-28{,}5)^2 \cdot 0{,}2 + (30-28{,}5)^2 \cdot 0{,}15 + \\ + (40-28{,}5)^2 \cdot 0{,}45 = 136{,}4;$$

$$\sigma(Y) = \sqrt{136{,}4} \approx 11{,}68.$$

Корреляционный момент

$$\mathrm{cov}(X,Y) = \mathbf{M}(XY) - \mathbf{M}(X)\mathbf{M}(Y),$$

где

$$\begin{aligned}\mathbf{M}(XY) &= \sum_i \sum_j x_i y_j p_{ij} = \\ &= -2{,}3 - 80 \cdot 0{,}027 - 60 \cdot 0{,}05 - 40 \cdot 0{,}1 - \\ &\quad - 200 \cdot 0{,}05 - 160 \cdot 0{,}1 - 120 \cdot 0{,}025 - \\ &\quad - 80 \cdot 0{,}025 - 30 \cdot 0{,}05 - 240 \cdot 0{,}05 - \\ &\quad - 180 \cdot 0{,}025 - 120 \cdot 0{,}025 - 400 \cdot 0{,}027 - \\ &\quad - 320 \cdot 0{,}023 - 240 \cdot 0{,}05 - 160 \cdot 0{,}35 = \\ &= -149{,}62; \\ \mathbf{M}(X)\mathbf{M}(Y) &= -5 \cdot 28{,}5 = -142{,}5.\end{aligned}$$

Следовательно,

$$\mathrm{cov}(X,Y) = -149{,}62 - (-142{,}5) = -7{,}12.$$

Коэффициент корреляции

$$r(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Итак,

$$r(X,Y) = \frac{-7{,}12}{2{,}49 \cdot 11{,}68} = -0{,}24.$$

Поскольку $r(X,Y) \neq 0$, можем сделать вывод о коррелированности случайных величин X и Y.

Для вычисления $\mathbf{M}(X/Y=40)$ построим условный закон распределения для случайной величины X при $Y=40$. Применяя формулу

$$\mathbf{P}(X=x_i/Y=y_j) = \frac{\mathbf{P}(X=x_i, Y=y_j)}{\mathbf{P}(Y=y_j)},$$

получаем

X	-10	-8	-6	-4
$\mathbf{P}(X/Y=40)$	$\frac{27}{450}$	$\frac{23}{450}$	$\frac{50}{450}$	$\frac{50}{450}$

.

Таким образом,

$$\begin{aligned}\mathbf{M}(X/Y=40) &= \sum_{i=1}^{4} x_i \mathbf{P}(X=x_i/Y=40) = \\ &= \frac{1}{450}(-10\cdot 27 - 8\cdot 23 - 6\cdot 50 - 4\cdot 350) = \\ &= -\frac{2154}{450} \approx -4{,}79.\end{aligned}$$

Аналогично вычисляем $\mathbf{M}(Y/X=-8)$:

Y	10	20	30	40
$\mathbf{P}(Y/X=-8)$	$\frac{27}{200}$	$\frac{100}{200}$	$\frac{50}{200}$	$\frac{23}{200}$

.

$$\mathbf{M}(Y/X=-8) = \frac{1}{200}(270+2000+1500+920) = \frac{4690}{200} = 23{,}45. \qquad \square$$

Пример 6. Система случайных величин (X,Y) имеет плотность распределения вероятностей

$$f(x,y)=a(xy+y^2),$$

где $(x,y)\in\mathcal{D}=\{(x,y)\colon 0\leqslant x\leqslant 1;\ \ 0{,}5x\leqslant y\leqslant x\}$.

Найти значения константы a, математические ожидания $\mathbf{M}(X)$ и $\mathbf{M}(Y)$, средние квадратические отклонения $\sigma(X)$ и $\sigma(Y)$, корреляционный момент $\operatorname{cov}(X,Y)$ и коэффициент корреляции $r(X,Y)$.

Решение. Чтобы отыскать константу a, воспользуемся характеристическим свойством функции плотности распределения вероятностей $f(x,y)$:

$$\iint\limits_{\mathbb{R}^2} f(x,y)dxdy=1;$$

$$\iint\limits_{\mathcal{D}} a(xy+y^2)dxdy=a\iint\limits_{\mathcal{D}}(xy+y^2)dxdy=a\int\limits_0^1\int\limits_{0,5x}^{x}(xy+y^2)dxdy=$$

$$=a\int\limits_0^1\left(\int\limits_{0,5x}^{x}(xy+y^2)dy\right)dx=$$

$$=a\int\limits_0^1\left(\frac{xy^2}{2}+\frac{y^3}{3}\right)\Bigg|_{0,5x}^{x}dx=$$

$$=a\int\limits_0^1\left(\frac{x^3}{2}+\frac{x^3}{3}-\frac{0{,}25x^3}{2}-\frac{0{,}125x^3}{3}\right)dx=$$

$$=a\int\limits_0^1\frac{4x^3}{6}dx=\frac{4a}{6}\cdot\frac{x^4}{4}\Bigg|_0^1=\frac{a}{6},$$

откуда

$$\frac{a}{6}=1,$$

поэтому

$$a = 6.$$

Функцию $f(x, y)$ можно записать так:

$$f(x, y) = 6(xy + y^2).$$

Математическое ожидание $\mathbf{M}(X)$ вычисляем по формуле

$$\mathbf{M}(X) = \iint\limits_{\mathbb{R}^2} x f(x, y) dx dy.$$

Получаем

$$\mathbf{M}(X) = \int\limits_0^1 \int\limits_{0,5x}^x 6x(xy + y^2) dx dy = 6 \int\limits_0^1 \left(\int\limits_{0,5x}^x (x^2 y + xy^2) dy \right) dx =$$

$$= 6 \int\limits_0^1 \left(\frac{x^2 y^2}{2} + \frac{xy^3}{3} \right) \Bigg|_{0,5x}^x dx =$$

$$= 6 \int\limits_0^1 \left(\frac{x^4}{2} + \frac{x^4}{3} - \frac{0{,}25x^4}{2} - \frac{0{,}125x^4}{3} \right) dx =$$

$$= 6 \int\limits_0^1 \frac{4}{6} x^4 dx = 4 \cdot \frac{x^5}{5} \Bigg|_0^1 = \frac{4}{5} = 0{,}8.$$

Аналогично вычисляем $\mathbf{M}(Y)$:

$$\mathbf{M}(Y) = \iint\limits_{\mathbb{R}^2} y f(x, y) dx dy.$$

$$\mathbf{M}(Y) = \int\limits_0^1 \int\limits_{0,5x}^x 6y(xy + y^2) dx dy = 6 \int\limits_0^1 \left(\int\limits_{0,5x}^x (xy^2 + y^3) dy \right) dx =$$

$$= 6\int\limits_0^1 \left(\frac{xy^3}{3}+\frac{y^4}{4}\right)\Bigg|_{0,5x}^{x} dx =$$

$$= 6\int\limits_0^1 \left(\frac{x^4}{3}+\frac{x^4}{4}-\frac{0{,}125x^4}{3}-\frac{0{,}0625x^4}{4}\right) dx =$$

$$= 6\int\limits_0^1 \frac{6{,}3125}{12}x^4 dx = \frac{1}{2}\cdot 6{,}3125\cdot\frac{x^5}{5}\Bigg|_0^1 = 0{,}63125.$$

Среднее квадратическое отклонение

$$\sigma(X) = \sqrt{\mathbf{D}(X)},$$

где $\mathbf{D}(X)$ — дисперсия случайной величины X, которая вычисляется по формуле

$$\mathbf{D}(X) = \iint\limits_{\mathbb{R}^2} x^2 f(x,y)dxdy - (\mathbf{M}(X))^2 .$$

Итак,

$$\mathbf{D}(X) = 6\int\limits_0^1 \int\limits_{0,5x}^{x} x^2(xy+y^2)dxdy - 0{,}64 =$$

$$= 6\int\limits_0^1 \left(\frac{x^3y^2}{2}+\frac{x^2y^3}{3}\right)\Bigg|_{0,5x}^{x} dx - 0{,}64 =$$

$$= 6\int\limits_0^1 \left(\frac{x^5}{2}+\frac{x^5}{3}-\frac{0{,}25x^5}{2}-\frac{0{,}125x^5}{3}\right) dx - 0{,}64 =$$

$$= \frac{4x^6}{6}\Bigg|_0^1 - 0{,}64 = \frac{2}{3} - 0{,}64 = \frac{2}{75};$$

$$\sigma(X) = \sqrt{\frac{2}{75}} \approx 0{,}163.$$

Аналогично вычисляем $\sigma(Y)$:

$$\mathbf{D}(Y) = \iint\limits_{\mathbb{R}^2} y^2 f(x,y)dxdy - (\mathbf{M}(Y))^2 =$$

$$= 6\int\limits_0^1 \int\limits_{0,5x}^{x} y^2(xy+y^2)dxdy - 0{,}398 =$$

$$= 6\int\limits_0^1 \left(\frac{xy^4}{4} + \frac{y^5}{5}\right)\Bigg|_{0,5x}^{x} dx - 0{,}398 =$$

$$= 6\int\limits_0^1 \left(\frac{x^5}{4} + \frac{x^5}{5} - \frac{0{,}0625x^5}{4} - \frac{0{,}03125x^5}{5}\right) dx - 0{,}398 \approx$$

$$\approx 0{,}428;$$

$$\sigma(X) = \sqrt{0{,}428} \approx 0{,}654.$$

Корреляционный момент

$$\operatorname{cov}(X,Y) = \mathbf{M}[(X - \mathbf{M}(X))(Y - \mathbf{M}(Y))].$$

В интегральном изображении имеем

$$\operatorname{cov}(X,Y) = \iint\limits_{\mathbb{R}^2} xyf(x,y)dxdy - \mathbf{M}(X)\mathbf{M}(Y) =$$

$$= 6\int\limits_0^1 \int\limits_{0,5x}^{x} xy(xy+y^2)dxdy - 0{,}8 \cdot 0{,}63125 =$$

$$= 6\int\limits_0^1 \left(\frac{x^2y^3}{3} + \frac{xy^4}{4}\right)\Bigg|_{0,5x}^{x} dx - 0{,}505 =$$

$$= 6\int\limits_0^1 \left(\frac{x^5}{3} + \frac{x^5}{4} - \frac{0{,}125x^5}{3} - \frac{0{,}0625x^5}{4}\right) dx - 0{,}505 =$$

$$= \frac{6{,}3125}{12}x^6\Big|_0^1 - 0{,}505 \approx 0{,}021.$$

По формуле

$$r(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

вычислим коэффициент корреляции:

$$r(X,Y) = \frac{0{,}021}{0{,}163 \cdot 0{,}654} \approx 0{,}2.$$

Итак,

$$r(X,Y) = 0{,}2,$$

а это значит, что случайные величины X и Y коррелированы. □

Пример 7. Система случайных величин (X,Y) имеет равномерное распределение вероятностей в области

$$\mathcal{D} = \{(x,y)\colon x \in [0;1], y \in [0;1]\}.$$

За пределами области $\mathcal{D}$ плотность распределения вероятностей равняется нулю.

Записать функцию плотности распределения вероятностей $f(x,y)$, интегральную функцию распределения вероятностей $F(x,y)$.

Решение. Согласно условию задачи в общем виде дифференциальную функцию $f(x,y)$ можно записать так:

$$f(x,y) = \begin{cases} c, & (x,y) \in \mathcal{D}; \\ 0, & (x,y) \notin \mathcal{D}. \end{cases}$$

Согласно характеристическому свойству

$$\iint\limits_{\mathbb{R}^2} f(x,y)dxdy = 1$$

и с учетом того, что за пределами области $\mathcal{D}$ плотность распределения вероятностей равняется нулю, можно записать

$$\iint\limits_{\mathcal{D}} f(x,y)dxdy = 1.$$

Геометрическая интерпретация последнего равенства следующая: объём параллелепипеда, ограниченного снизу областью $\mathcal{D}$ и сверху графиком функции $f(x, y)$, равняется единице. Поскольку система случайных величин (X, Y) по условию задачи имеет равномерное распределение, геометрическую интерпретацию можно детализировать так: объем прямоугольного параллелепипеда, в основе которого лежит прямоугольник $OABC$ (рис. 25), равняется единице.

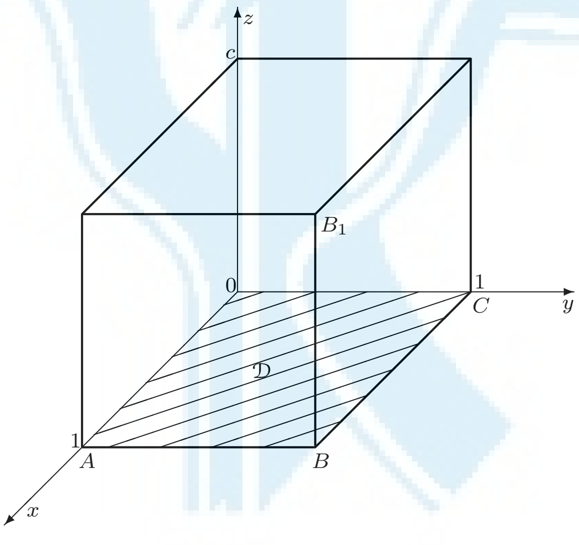


Рис. 25

Длина отрезка BB_1 — высоты прямоугольного параллелепипеда — равняется значению c в записи функции плотности $f(x, y)$.

Итак, очевидным есть способ отыскания c:

$$V_{\text{парал}} = 1.$$

Поскольку

$$V_{\text{парал}} = S_{\text{осн}} \cdot H,$$

где $S_{\text{осн}}$ — площадь основы параллелепипеда (прямоугольника $OABC$); H — высота параллелепипеда ($H = BB_1 = c$),

$$S_{OABC} = AO \cdot OC = 1 \cdot 1 = 1;$$
$$V_{\text{парал}} = 1 \cdot BB_1 = c.$$

Итак, $c = 1$.

Функцию $f(x, y)$ можно записать так:

$$f(x, y) = \begin{cases} 1, & (x, y) \in \mathcal{D}; \\ 0, & (x, y) \notin \mathcal{D}. \end{cases}$$

По определению функции распределения вероятности

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

С геометрической точки зрения это вероятность попадания системы случайных величин $(X; Y)$ в бесконечный квадрант с вершиной в точке $(x; y)$ (рис. 26).

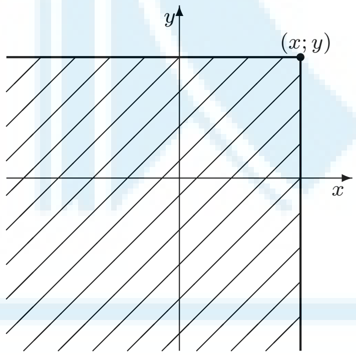


Рис. 26

Согласно условию задачи при $x \leqslant 0$ и $y \leqslant 0$

$$F(x, y) = 0,$$

поскольку при таком условии $f(x, y) = 0$.

Если $0 < x \leqslant 1$ и $0 < y \leqslant 1$, то для отыскания $F(x, y)$ необходимо найти объём прямоугольного параллелепипеда с основой $OA'B''$ (рис. 27).

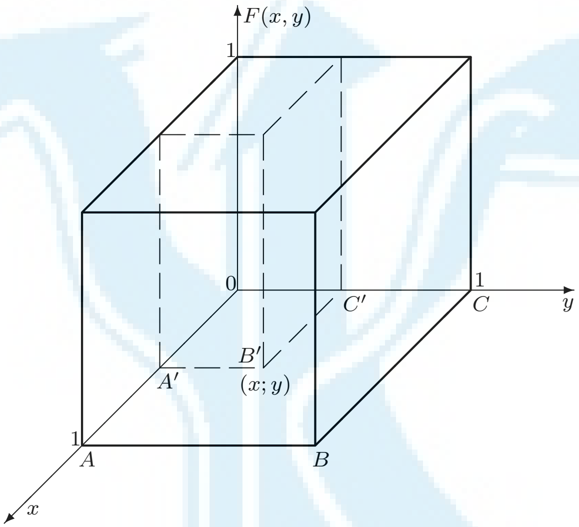


Рис. 27

Укажем, что т. B' имеет координаты $(x; y)$, по которым определим искомый объём. Очевидно, что площадь прямоугольника $OA'B''$

$$S_{OA'B''} = xy.$$

Учитывая, что высота параллелепипеда равняется единице, можно записать, что

$$F(x, y) = xy, \qquad \text{если} \quad 0 < x \leqslant 1 \quad \text{и} \quad 0 < y \leqslant 1.$$

Понятно, что в случае $0 < x \leqslant 1$ и $y > 1$ (рис. 28) имеем

$$F(x, y) = x,$$

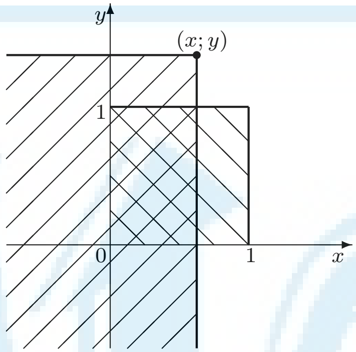


Рис. 28

а при $x > 1$ і $0 < y \leqslant 1$ (рис. 29)

$$F(x, y) = y.$$

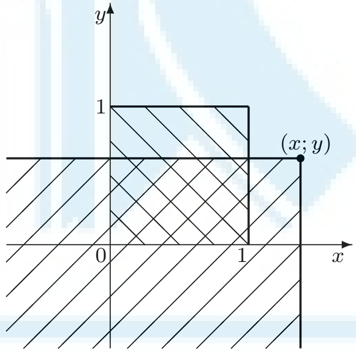


Рис. 29

Если $x > 1$ и $y > 1$ (рис. 30), то

$$F(x, y) = 1,$$

поскольку объём прямоугольного параллелепипеда с основой $OABC$ равняется единице.

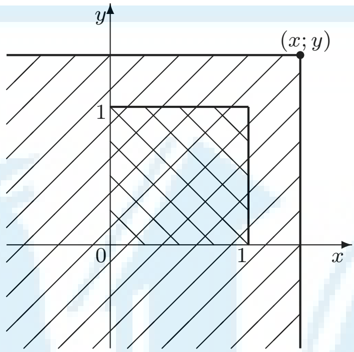


Рис. 30

Окончательно можно записать выражение для интегральной функции распределения вероятностей:

$$F(x,y)=\begin{cases}0, & x\leqslant 0 \quad \text{или} \quad y\leqslant 0;\\ x, & 0<x\leqslant 1 \quad \text{и} \quad y>1;\\ y, & x>1 \quad \text{и} \quad 0<y\leqslant 1;\\ xy, & 0<x\leqslant 1 \quad \text{и} \quad 0<y\leqslant 1;\\ 1, & x>1 \quad \text{и} \quad y>1.\end{cases}$$

Такой же результат можно получить, не прибегая к геометрическим выкладкам, а применяя лишь определение $F(x,y)$ и аппарат интегрального исчисления:

а) если $0<x\leqslant 1$ и $0<y\leqslant 1$, то

$$F(x,y)=\mathbf{P}(X<x,Y<y)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y}f(x,y)dxdy=$$

$$=\int\limits_{-\infty}^{0}\int\limits_{-\infty}^{0}0dxdy+\int\limits_{0}^{x}\int\limits_{0}^{y}1dxdy=\int\limits_{0}^{x}\int\limits_{0}^{y}dxdy=xy;$$

б) если $0 < x \leqslant 1$ и $y > 1$, то

$$F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x,y)dxdy = \int\limits_{-\infty}^{0}\int\limits_{-\infty}^{0} 0dxdy + \int\limits_{0}^{x}\int\limits_{0}^{y} f(x,y)dxdy =$$

$$= \int\limits_{0}^{x}\int\limits_{0}^{1} 1dxdy + \int\limits_{0}^{x}\int\limits_{1}^{y} 0dxdy = \int\limits_{0}^{x}\int\limits_{0}^{1} dxdy = x;$$

в) если $x > 1$ и $0 < y \leqslant 1$, то аналогично получаем

$$F(x,y) = \int\limits_{0}^{1}\int\limits_{0}^{y} dxdy = y;$$

г) если $x > 1$ и $y > 1$, то

$$F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x,y)dxdy =$$

$$= \int\limits_{-\infty}^{0}\int\limits_{-\infty}^{0} 0dxdy + \int\limits_{0}^{1}\int\limits_{0}^{1} 1dxdy + \int\limits_{1}^{x}\int\limits_{1}^{y} 0dxdy = 1,$$

что и подтверждает правильность предыдущего решения. □

Пример 8. Система случайных величин (X, Y) имеет равномерное распределение вероятностей в области

$$\mathcal{D} = \{(x,y)\colon |x| + |y| \leqslant 1\}.$$

За пределами области $\mathcal{D}$ плотность распределения вероятностей равняется нулю.

Записать выражения для дифференциальной функции распределения $f(x,y)$ и интегральной функции $F(x,y)$.

Решение. Область $\mathcal{D}$ является квадратом $ABCD$ с вершинами на осях координат (рис. 31):

$$A(1;0), \qquad B(0;1), \qquad C(-1;0), \qquad D(0;-1).$$

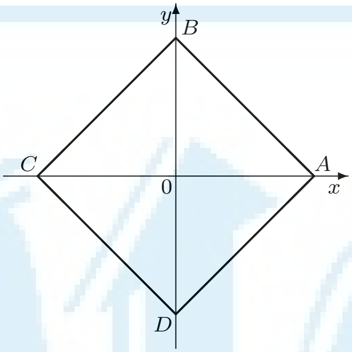


Рис. 31

Поскольку в границах области $\mathcal{D}$ распределение вероятностей равномерно, в общем виде функция плотности запишется так:

$$f(x,y)=\begin{cases} c, & (x,y)\in\mathcal{D};\\ 0, & (x,y)\notin\mathcal{D}.\end{cases}$$

Значение c определим, используя характеристическое свойство функции плотности $f(x,y)$, а именно

$$\iint\limits_{\mathbb{R}^2} f(x,y)dxdy=1.$$

Для данного примера можем записать

$$\iint\limits_{\mathcal{D}} c\,dxdy=1.$$

Учитывая равномерность распределения и геометрическую интерпретацию последнего интеграла, значение c найдем как длину высоты прямоугольного параллелепипеда, основой которого является квадрат $ABCD$:

$$V_{\text{парал}}=S_{ABCD}\cdot H,\qquad H=\frac{V_{\text{парал}}}{S_{ABCD}}.$$

Поскольку

$$S_{ABCD} = AB \cdot CD = \sqrt{2} \cdot \sqrt{2} = 2,$$

то

$$c = H = \frac{1}{2}.$$

Итак, функция $f(x, y)$ имеет вид

$$f(x, y) = \begin{cases} \dfrac{1}{2}, & (x, y) \in \mathcal{D}; \\ 0, & (x, y) \notin \mathcal{D}. \end{cases}$$

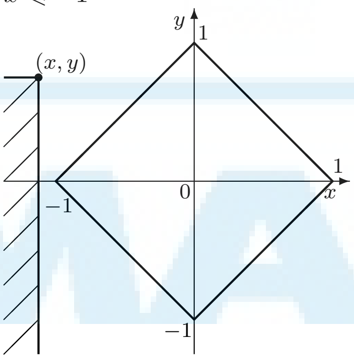
Функцию $F(x, y)$ определим, опираясь на определение интегральной функции распределения вероятностей

$$F(x, y) = \mathbf{P}(X < x, Y < y)$$

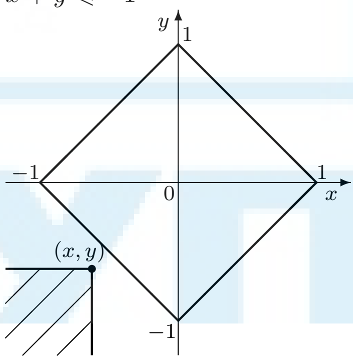
и геометрические соображения, связанные с двойными интегралами.

Чтобы получить выражение функции $F(x, y)$, необходимо рассмотреть такие варианты размещения точки $f(x, y)$ относительно области $\mathcal{D}$:

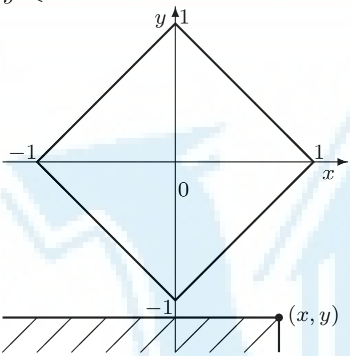
а) $x \leqslant -1$



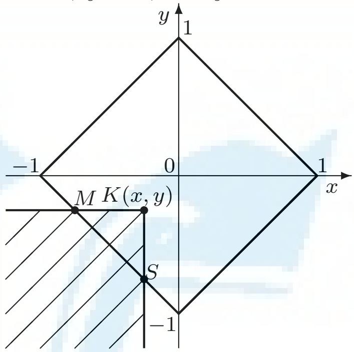
б) $x + y \leqslant -1$



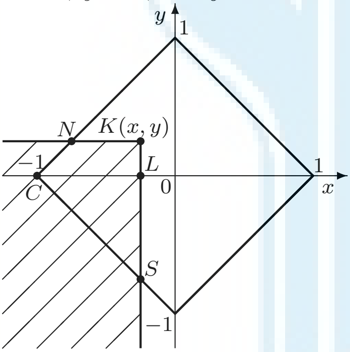
в) $y \leqslant -1$



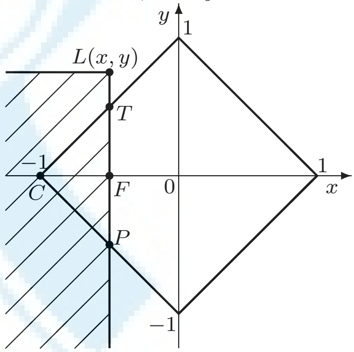
г) $x \leqslant 0,\ y \leqslant 0,\ x + y > -1$



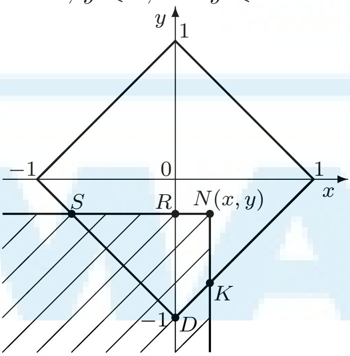
д) $x \leqslant 0,\ y > 0,\ x - y > -1$



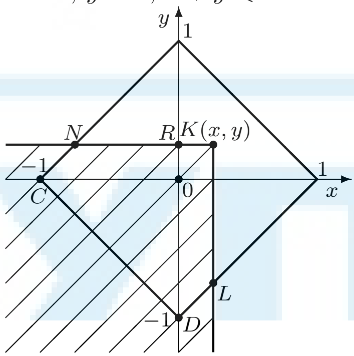
е) $-1 < x \leqslant 0,\ x - y \leqslant -1$



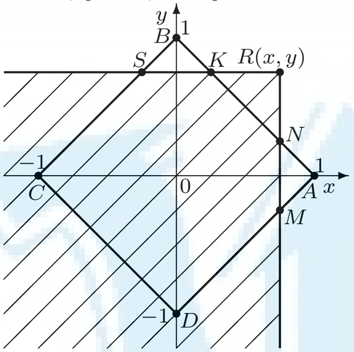
ё) $x > 0,\ y \leqslant 0,\ x - y \leqslant 1$



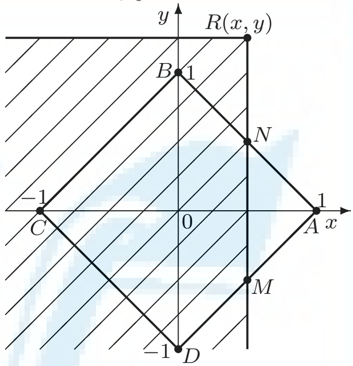
ж) $x > 0,\ y > 0,\ x + y \leqslant 1$



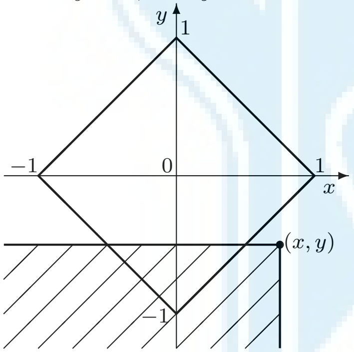
з) $x \leqslant 1,\ y \leqslant 1,\ x + y > 1$



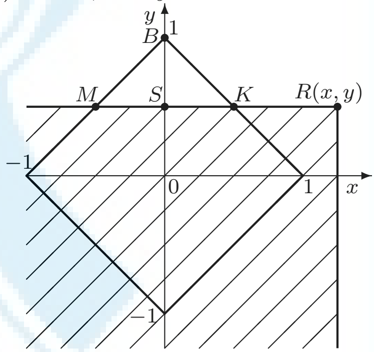
и) $0 < x \leqslant 1,\ y > 1$



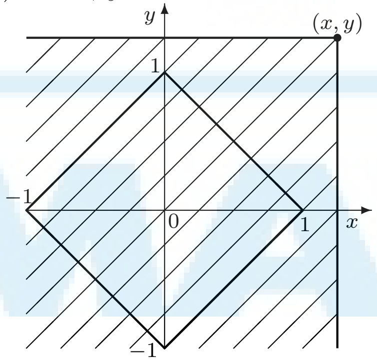
й) $-1 < y \leqslant 0,\ x - y > 1$



к) $x > 1,\ 0 < y \leqslant 1$



л) $x > 1,\ y > 1$



Очевидно, что в случаях а), б) и в)

$$F(x, y) = 0,$$

поскольку в соответствии с условием задачи вне границ области $\mathcal{D}$

$$f(x, y) = 0$$

и ни одна часть области $\mathcal{D}$ с бесконечным квадрантом не пересекается.

Для того чтобы записать выражение $F(x, y)$ в случае г), необходимо через координаты точки K выразить объём прямой призмы с основой MKS и высотой $\frac{1}{2}$:

$$V_{\text{призмы}} = S_{MKS} \cdot H.$$

Уравнение прямой MS имеет вид $-x-y = 1$, или $y = -x-1$. Если (x, y) — координаты точки K, то длину отрезка KS можно найти как разность ординаты точек S и K, т. е.

$$KS = y - (-x - 1) = x + y + 1.$$

Поскольку $MK = KS$, то

$$S_{MKS} = \frac{1}{2} \cdot MK \cdot KS = \frac{(x + y + 1)^2}{2}.$$

Итак,

$$V_{\text{призмы}} = \frac{(x + y + 1)^2}{2} \cdot \frac{1}{2} = \frac{(x + y + 1)^2}{4}.$$

Окончательно для случая г) функция $F(x, y)$ имеет вид

$$F(x, y) = \frac{(x + y + 1)^2}{4}.$$

В случае д) необходимо вычислить объём прямой призмы с основой $KNCS$ и высотой $\frac{1}{2}$, выразив его через координаты (x, y) точки

K. Для этого разобьём фигуру $KNCS$ на две: $LKNC$ и CLS. Длина катетов $LS = LC$ равнобедренного треугольника

$$LC = x - (-1) = x + 1.$$

Итак,

$$S_{CLB} = \frac{1}{2}(x+1)(x+1) = \frac{(x+1)^2}{2}.$$

Фигура $CLKN$ является трапецией с основами

$$CL = x + 1, \qquad KN = x - y + 1$$

(использованы уравнения прямой CB: $-x + y = 1$ и тот факт, что точка N принадлежит прямой CB) и высотой $KL = y$. Учитывая указанное, имеем

$$S_{CLKN} = \frac{x+1+x-y+1}{2} \cdot y = \frac{2x-y+2}{2} \cdot y;$$

$$V_{\text{призмы}} = \frac{1}{2} \cdot \left(\frac{(x+1)^2}{2} + \frac{2xy - y^2 + 2y}{2} \right) = \frac{1}{4}\left((x+1)^2 + 2y(x+1) - y^2\right) = \frac{1}{4}\left((x+y+1)^2 - 2y^2\right).$$

Итак, для случая д) функция $F(x,y)$ имеет вид

$$F(x,y) = \frac{1}{4}\left((x+y+1)^2 - 2y^2\right).$$

Рассматривая случай е), необходимо найти объём призмы с основой CPT. При этом искомый объём нужно выразить через координаты (x,y) точки L. Поскольку

$$S_{CPT} = \frac{1}{2}CF \cdot PT, \qquad CF = x+1, \qquad TP = 2TF = 2CF = 2(x+1),$$

площадь треугольника CPT

$$S_{CPT} = \frac{1}{2}(x+1) \cdot 2(x+1) = (x+1)^2.$$

Итак,

$$V_{\text{призмы}} = \frac{1}{2}(x+1)^2.$$

Выражение $F(x,y)$ для случая е) будет таким:

$$F(x,y) = \frac{(x+1)^2}{2}.$$

В случае ё) ситуация аналогичная случаю д). Площадь фигуры $KNSD$ вычисляется как сумма площадей фигур $KNRD$ и DRS. При этом используются координаты точки $N(x,y)$, уравнение прямой AD: $x-y=1$, факт принадлежности точки K прямой AD и то, что SDR — равнобедренный треугольник, а $DRNK$ — трапеция.

$$S_{DRS} = \frac{1}{2}\cdot DR\cdot RS, \qquad DR = RS, \qquad DR = y+1.$$

Итак,

$$S_{DRS} = \frac{(y+1)^2}{2}.$$

$$S_{DRNK} = \frac{NK+RD}{2}\cdot NR, \quad RD = y+1, \quad NK = y-x+1, \quad NR = x.$$

Итак,

$$S_{DRNK} = \frac{y-x+1+y+1}{2}\cdot x = \frac{2y+2-x}{2}\cdot x;$$

$$V_{\text{призмы}} = \frac{1}{2}\cdot\left(\frac{(y+1)^2}{2} + \frac{2y+2-x}{2}\cdot x\right) =$$

$$= \frac{1}{4}\left((y+1)^2 + 2xy + 2x - x^2\right) = \frac{1}{4}\left((x+y+1)^2 - 2x^2\right).$$

Окончательно для случая ё) выражение функции $F(x,y)$ можно записать так:

$$F(x,y) = \frac{1}{4}\left((x+y+1)^2 - 2x^2\right).$$

В случае ж) необходимо вычислить объём прямой призмы с основой $LKNCD$ через координаты (x,y) точки K. Для этого найдём площадь фигуры $LKNCD$, разбив её на три части: $LKRD$, $ORNC$ и OCD:

$$S_{LKRD} = \frac{y+1+y-x+1}{2} \cdot x = \frac{2y+2-x}{2} \cdot x,$$
$$S_{ORNC} = \frac{1-y+1}{2} \cdot y = \frac{2-y}{2} \cdot y,$$
$$S_{OCD} = \frac{1}{2} \cdot 1 \cdot 1 = \frac{1}{2};$$
$$S_{LKNCD} = \frac{2y+2-x}{2} \cdot x + \frac{2-y}{2} \cdot y + \frac{1}{2} =$$
$$= \frac{1}{2}(2xy + 2x - x^2 + 2y - y^2 + 1) = x + y - \frac{(x-y)^2}{2} + \frac{1}{2}.$$

Объём прямой призмы с основой $LKNCD$

$$V_{\text{призмы}} = \frac{1}{2} S_{LKNCD}.$$

Следовательно, в случае ж)

$$F(x,y) = \frac{x+y}{2} - \frac{(x-y)^2}{4} + \frac{1}{4}.$$

Для того чтобы записать выражение функции $F(x,y)$ в случае з), необходимо выразить площадь фигуры $MNKSCD$ через координаты (x,y) точки R. Указанную площадь вычислим как разность площадей прямоугольника $ABCD$ и двух треугольников AMN и KBS:

$$S_{ABCD} = 2,$$
$$S_{ANM} = \frac{1}{2} \cdot (1-x) \cdot 2(1-x) = (1-x)^2,$$
$$S_{KBS} = \frac{1}{2} \cdot (1-y) \cdot 2(1-y) = (1-y)^2;$$
$$S_{MNKSCD} = 2 - (1-x)^2 - (1-y)^2.$$

Объём прямой призмы с основой $MNKSCD$

$$V_{\text{призмы}} = \frac{1}{2} \cdot S_{MNKSCD}.$$

Итак, в случае з)

$$F(x,y) = \frac{1}{2}(2-(1-x)^2-(1-y)^2) = x+y-\frac{1}{2}(x^2+y^2).$$

В случае и) необходимо вычислить объём прямой призмы с основой $NBCDM$ через координаты (x,y) точки R. Для этого найдём площадь фигуры $NBCDM$ как разность площадей квадрата $ABCD$ и треугольника ANM:

$$\begin{aligned} S_{ABCD} &= 2, \\ S_{ANM} &= (1-x)^2; \\ S_{NBCDM} &= 2-(1-x)^2. \end{aligned}$$

Объём прямой призмы с основой $NBCDM$

$$V_{\text{призмы}} = \frac{1}{2}S_{NBCDM}.$$

Итак, в случае и)

$$F(x,y) = 1-\frac{(1-x)^2}{2}.$$

Аналогичная случаю е) ситуация в случае й), лишь x и y меняются ролями:

$$F(x,y) = \frac{(y+1)^2}{2}.$$

Аналогичная случаю и) ситуация в случае к), лишь x и y меняются ролями:

$$F(x,y) = 1-\frac{(1-y)^2}{2}.$$

Очевидно, что для случая л)

$$x > 1, \qquad y > 1$$

функция распределения $F(x,y)$ равняется единице. □

Пример 9. Функция плотности распределения вероятностей системы случайных величин (X, Y) имеет вид

$$f(x,y) = ae^{-2x^2+xy-y^2}$$

при $x \in \mathbb{R}$ и $y \in \mathbb{R}$.

Определить константу a. Найти $f(x)$, $f(y)$, $f(x/y)$, $f(y/x)$. Вычислить математические ожидания $\mathbf{M}(X)$, $\mathbf{M}(Y)$, коэффициент корреляции $r(X, Y)$. Выяснить, коррелированы ли случайные величины X и Y.

Решение. Значение константы a вычислим, используя характеристическое свойство функции плотности:

$$\iint\limits_{\mathbb{R}^2} f(x,y)dxdy = 1.$$

В нашем случае

$$\iint\limits_{\mathbb{R}^2} f(x,y)dxdy = \iint\limits_{\mathbb{R}^2} ae^{-2x^2+xy-y^2}dxdy = a\iint\limits_{\mathbb{R}^2} e^{-2x^2+xy-y^2}dxdy =$$

$$= a\iint\limits_{\mathbb{R}^2} e^{-2x^2} \cdot e^{-(y^2-xy)}dxdy =$$

$$= a\iint\limits_{\mathbb{R}^2} e^{-2x^2} \cdot e^{-(y^2-2\cdot\frac{1}{2}xy+\frac{x^2}{4}-\frac{x^2}{4})}dxdy =$$

$$= a\iint\limits_{\mathbb{R}^2} e^{-2x^2} \cdot e^{-\left(y-\frac{1}{2}x\right)^2+\frac{x^2}{4}}dxdy =$$

$$= a\iint\limits_{\mathbb{R}^2} e^{-2x^2} \cdot e^{\frac{x^2}{4}} \cdot e^{-\left(y-\frac{x}{2}\right)^2}dxdy =$$

$$= a\int\limits_{-\infty}^{+\infty} e^{-\frac{7}{4}x^2}dx \cdot \underbrace{\int\limits_{-\infty}^{+\infty} e^{-\left(y-\frac{x}{2}\right)^2}d\left(y-\frac{x}{2}\right)}_{=\sqrt{\pi}\text{ как интеграл Пуассона}} =$$

$$= a\sqrt{\pi}\int\limits_{-\infty}^{+\infty} e^{-\frac{7}{4}x^2}dx =$$

$$= a\sqrt{\pi}\cdot\frac{2}{\sqrt{7}}\int\limits_{-\infty}^{+\infty} e^{-\left(\frac{\sqrt{7}x}{2}\right)^2} d\left(\frac{\sqrt{7}x}{2}\right) =$$

$$= \frac{2a\sqrt{\pi}}{\sqrt{7}}\int\limits_{-\infty}^{+\infty} e^{-u^2}du = \frac{2a\sqrt{\pi}}{\sqrt{7}}\sqrt{\pi} = \frac{2a\pi}{\sqrt{7}}.$$

Итак,

$$a\cdot\frac{2\pi}{\sqrt{7}} = 1,$$

откуда

$$a = \frac{\sqrt{7}}{2\pi}.$$

Функция плотности имеет вид

$$f(x,y) = \frac{\sqrt{7}}{2\pi}e^{-2x^2+xy-y^2}, \qquad x\in\mathbb{R}, \quad y\in\mathbb{R}.$$

Для того чтобы найти функцию плотности распределения вероятностей случайной величины X, воспользуемся формулой

$$f(x) = \int\limits_{-\infty}^{+\infty} f(x,y)dy.$$

Следовательно,

$$f(x) = \int\limits_{-\infty}^{+\infty}\frac{\sqrt{7}}{2\pi}e^{-2x^2+xy-y^2}dy = \frac{\sqrt{7}}{2\pi}e^{-2x^2}\int\limits_{-\infty}^{+\infty} e^{-\left(y^2-2y\cdot\frac{1}{2}x+\frac{x^2}{4}-\frac{x^2}{4}\right)}dy =$$

$$= \frac{\sqrt{7}}{2\pi}e^{-2x^2}\int\limits_{-\infty}^{+\infty} e^{-\left(y-\frac{x}{2}\right)^2}\cdot e^{\frac{x^2}{4}}dy = \frac{\sqrt{7}}{2\pi}e^{-2x^2}\cdot e\frac{x^2}{4}\int\limits_{-\infty}^{+\infty} e^{-u^2}du =$$

$$= \frac{\sqrt{7}}{2\pi} e^{-\frac{7x^2}{4}} \sqrt{\pi} = \frac{\sqrt{7}}{2\sqrt{\pi}} e^{-\frac{7x^2}{4}}.$$

Аналогично определяем функцию $f(y)$:

$$f(y) = \int\limits_{-\infty}^{+\infty} f(x,y)dx = \frac{\sqrt{7}}{2\pi} \int\limits_{-\infty}^{+\infty} e^{-2x^2+xy-y^2} dx =$$

$$= \frac{\sqrt{7}}{2\pi} \int\limits_{-\infty}^{+\infty} e^{-2x^2+xy} \cdot e^{-y^2} dx = \frac{\sqrt{7}}{2\pi} e^{-y^2} \int\limits_{-\infty}^{+\infty} e^{-\left(-2x^2-xy\right)} dx =$$

$$= \frac{\sqrt{7}}{2\pi} e^{-y^2} \int\limits_{-\infty}^{+\infty} e^{-\left(-\sqrt{2}x - \frac{y}{2\sqrt{2}}\right)^2 + \frac{y^2}{8}} dx =$$

$$= \frac{\sqrt{7}}{2\pi} e^{-y^2} \int\limits_{-\infty}^{+\infty} e^{-\left(-\sqrt{2}x - \frac{y}{2\sqrt{2}}\right)^2} \cdot e^{\frac{y^2}{8}} dx =$$

$$= \frac{\sqrt{7}}{2\pi} e^{-\frac{7y^2}{8}} \cdot \frac{1}{\sqrt{2}} \int\limits_{-\infty}^{+\infty} e^{-\left(-\sqrt{2}x - \frac{y}{2\sqrt{2}}\right)^2} d\left(-\sqrt{2}x - \frac{y}{2\sqrt{2}}\right) =$$

$$= \frac{\sqrt{7}}{2\sqrt{2}\pi} e^{-\frac{7y^2}{8}} \cdot \sqrt{\pi} = \frac{\sqrt{7}}{2\sqrt{2\pi}} e^{-\frac{7y^2}{8}}.$$

Для определения $f(x/y)$ используем формулу

$$f(x/y) = \frac{f(x,y)}{f(y)}.$$

Итак,

$$f(x/y) = \frac{\frac{\sqrt{7}}{2\pi} e^{-2x^2+xy-y^2}}{\frac{\sqrt{7}}{2\sqrt{2\pi}} e^{-\frac{7y^2}{8}}} = \sqrt{\frac{2}{\pi}}\, e^{-2x^2+xy-\frac{y^2}{8}}.$$

Аналогично найдем $f(y/x)$:

$$f(y/x) = \frac{f(x,y)}{f(x)} = \frac{\frac{\sqrt{7}}{2\pi}e^{-2x^2+xy-y^2}}{\frac{\sqrt{7}}{2\sqrt{\pi}}e^{-\frac{7x^2}{4}}} = \frac{1}{\sqrt{\pi}}e^{-\frac{x^2}{4}+xy-y^2}.$$

Вычислим математическое ожидание $\mathbf{M}(X)$:

$$\mathbf{M}(X) = \int\limits_{-\infty}^{+\infty} xf(x)dx = \int\limits_{-\infty}^{+\infty} x\frac{\sqrt{7}}{2\sqrt{\pi}}e^{-\frac{7x^2}{4}}dx = \frac{\sqrt{7}}{2\sqrt{\pi}}\int\limits_{-\infty}^{+\infty} xe^{-\frac{7x^2}{4}}dx =$$

$$= \frac{\sqrt{7}}{2\sqrt{\pi}}\int\limits_{-\infty}^{+\infty} xe^{-\left(\frac{\sqrt{7}x}{2}\right)^2}dx =$$

$$= \frac{\sqrt{7}}{2\sqrt{\pi}}\left(\int\limits_{-\infty}^{0} xe^{-\left(\frac{\sqrt{7}x}{2}\right)^2}dx + \int\limits_{0}^{+\infty} xe^{-\left(\frac{\sqrt{7}x}{2}\right)^2}dx\right) =$$

$$= \frac{\sqrt{7}}{2\sqrt{\pi}}\left(\lim_{A\to-\infty}\int\limits_{A}^{0} xe^{-\left(\frac{\sqrt{7}x}{2}\right)^2}dx + \lim_{A\to+\infty}\int\limits_{0}^{A} xe^{-\left(\frac{\sqrt{7}x}{2}\right)^2}dx\right) = 0.$$

Аналогично определяем $\mathbf{M}(Y)$:

$$\mathbf{M}(Y) = \int\limits_{-\infty}^{+\infty} yf(y)dy = \int\limits_{-\infty}^{+\infty} y\frac{\sqrt{7}}{2\sqrt{2\pi}}e^{-\frac{7y^2}{8}}dy = \frac{\sqrt{7}}{2\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} ye^{-\frac{7y^2}{8}}dy = 0.$$

Итак,

$$\mathbf{M}(X) = \mathbf{M}(Y) = 0.$$

Коэффициент корреляции

$$r(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Найдем ковариацию $\operatorname{cov}(X, Y)$:

$$\operatorname{cov}(X,Y) = \iint\limits_{\mathbb{R}^2} (x - \mathbf{M}(X))(y - \mathbf{M}(Y)) f(x,y)\,dxdy =$$

$$= \iint\limits_{\mathbb{R}^2} xy \cdot \frac{\sqrt{7}}{2\pi} e^{-2x^2+xy-y^2}\,dxdy =$$

$$= \frac{\sqrt{7}}{2\pi} \iint\limits_{\mathbb{R}^2} xe^{-\frac{7}{4}x^2} \cdot ye^{-\left(y-\frac{1}{2}x\right)^2}\,dxdy =$$

$$= \frac{\sqrt{7}}{2\pi} \int\limits_{-\infty}^{+\infty} xe^{-\frac{7}{4}x^2}\,dx \cdot \int\limits_{-\infty}^{+\infty} ye^{-\left(y-\frac{x}{2}\right)^2}\,dy = \frac{\sqrt{7}}{2\pi} \cdot 0 = 0,$$

т. е.

$$r(X,Y) = 0,$$

что означает некоррелированность случайных величин X и Y. $\square$

Задачи к разделу 9

Задача 1. Система случайных величин (X, Y) имеет равномерное распределение вероятностей в области $\mathcal{D}$. Вне этой области плотность распределения вероятностей системы случайных величин равняется нулю.

Записать аналитическое выражение для функции распределения вероятностей $F(x, y)$ и функции плотности $f(x, y)$. Вычислить $\mathbf{M}(X)$, $\mathbf{M}(Y)$, если

$$\mathcal{D} = \{(x,y)\colon x \in [-1;1], y \in [-1;1]\}.$$

Ответ. $$F(x,y)=\begin{cases}0, & x\leqslant -1 \text{ или } y\leqslant -1;\\ \dfrac{1}{4}(x+1)(y+1), & (x,y)\in\mathcal{D};\\ \dfrac{1}{2}(y+1), & x>1 \text{ и } -1<y\leqslant 1;\\ \dfrac{1}{2}(x+1), & y>1 \text{ и } -1<x\leqslant 1;\\ 1, & x>1 \text{ и } y>1,\end{cases}$$

$f(x,y)=\begin{cases}\dfrac{1}{4}, & (x,y)\in\mathcal{D};\\ 0, & (x,y)\notin\mathcal{D},\end{cases}$ $\mathbf{M}(X)=0$, $\mathbf{M}(Y)=0$.

Задача 2. Система дискретных случайных величин задана таблицей распределения (табл. 4).

Таблица 4

Y	X				$\sum\limits_{i=1}^{4} p_{ij}$
	5	10	15	20	
−10	0,012	0,038	0,2	0,1	0,35
−8	0,038	0,012	0,05	0,05	0,15
−6	0,05	0,05	0,012	0,038	0,15
−4	0,1	0,2	0,038	0,012	0,35
$\sum\limits_{j=1}^{4} p_{ij}$	0,2	0,3	0,3	0,2	$\sum\limits_{i=1}^{4}\sum\limits_{j=1}^{4} p_{ij}=1$

Вычислить математические ожидания $\mathbf{M}(X)$, $\mathbf{M}(Y)$, средние квадратические отклонения $\sigma(X)$, $\sigma(Y)$, коэффициент корреляции $r(X,Y)$ и условные математические ожидания $\mathbf{M}(X/Y=-8)$, $\mathbf{M}(Y/X=5)$.

Ответ. $\mathbf{M}(X)=12{,}5$; $\mathbf{M}(Y)=-7$; $\sigma(X)=5{,}12$; $\sigma(Y)=2{,}57$; $r(X,Y)=-0{,}513$; $\mathbf{M}(X/Y=-8)=13{,}73$; $\mathbf{M}(Y/X=5)=-5{,}62$.

Задача 3. Плотность распределения вероятностей системы случайных величин (X,Y) имеет вид

$$f(x,y)=ae^{-4x^2-6xy-10y^2},\qquad x\in\mathbb{R},\quad y\in\mathbb{R}.$$

Найти a, $\mathbf{M}(X)$, $\mathbf{M}(Y)$, $\sigma(X)$, $\sigma(Y)$, $r(X,Y)$.

Ответ. $a = \dfrac{\sqrt{31}}{\pi}$, $\mathbf{M}(X) = 0$, $\mathbf{M}(Y) = 0$, $\sigma(X) = \dfrac{5}{31}$, $\sigma(Y) = \dfrac{2}{31}$, $r(X, Y) = 0$.

Раздел 10. Функции случайных величин

Ключевые слова

Величина	Случайный
Вероятностный	Функция
Полный	Эксперимент
Прообраз	

Пусть ξ — случайная величина, связанная с некоторым вероятностным экспериментом, $y = \varphi(\xi)$ — числовая функция, которая удовлетворяет таким требованиям:

- функция $y = \varphi(x)$ определена для всех значений, которые принимает случайная величина ξ;
- для произвольного значения $y_0 \in \mathbb{R}$ можно вычислить вероятность $\mathbf{P}(\xi \in \varphi^{-1}(y_0))$, где

$$\varphi^{-1}(y_0) = \{x\colon \varphi(x) = y_0\},$$

т. е. $\varphi^{-1}(y_0)$ — полный прообраз функции φ в точке y_0.

Тогда $\eta = \varphi(\xi)$ — случайная величина, которая принимает значение в зависимости от того, какое значения приняла случайная величина ξ. Если в результате эксперимента случайная величина ξ принимает значение ξ_0, то случайная величина η принимает значение $\eta_0 = \varphi(\xi_0)$.

Если ξ — дискретная случайная величина с рядом распределения

ξ	x_1	x_2	$\cdots$	x_n
$\mathbf{P}$	p_1	p_2	$\cdots$	p_n

то случайная величина $\eta = \varphi(\xi)$ — дискретная случайная величина с рядом распределения

η	y_1	y_2	$\cdots$	y_n
$\mathbf{P}$	p_1	p_2	$\cdots$	p_n

где $y_1 = \varphi(x_1)$, $y_2 = \varphi(x_2)$, …, $y_n = \varphi(x_n)$.

Если некоторые значения y_i и y_j $(i \neq j)$ равны между собой, то в ряд распределения записывают это значение лишь один раз, а вероятность, которая ему отвечает, равняется $p_i + p_j$.

Если ξ — непрерывная случайная величина, то случайная величина $\eta = \varphi(\xi)$ может:

а) быть дискретной случайной величиной (если $y = \varphi(x)$ — ступенчатая функция, т. е. разрывная функция, которая принимает не более чем счётное количество значений);

б) быть непрерывной случайной величиной, если $y = \varphi(x)$ — непрерывная функция без интервалов постоянства;

в) не быть ни дискретной, ни непрерывной случайной величиной (такие случайные величины называются случайными величинами смешанного типа).

В случае а) случайная величина η принимает значения y_1, y_2, ... с вероятностями

$$q_1 = \mathbf{P}(\xi \in \varphi^{-1}(y_1)), \quad q_2 = \mathbf{P}(\xi \in \varphi^{-1}(y_2)), \ldots .$$

В случае б) для произвольного интервала $(a; b)$ мы должны определить полный прообраз

$$\varphi^{-1}((a; b)) = \{x \colon \varphi(x) \in (a; b)\}$$

и вероятность

$$\mathbf{P}(\eta \in (a; b)) = \mathbf{P}(\xi \in \varphi^{-1}((a, b))).$$

Функция распределения случайной величины η имеет вид

$$G_\eta(y) = \mathbf{P}(\eta < y) = \mathbf{P}(\xi \in \varphi^{-1}((-\infty; y))) = \int\limits_{D_y} f_\xi(x) dx,$$

где

$$D_y = \varphi^{-1}((-\infty; y)) = \{x \colon \varphi(x) < y\},$$

$f_\xi(x)$ — плотность распределения вероятностей случайной величины ξ.

Если ξ — абсолютно непрерывная случайная величина с плотностью $f_\xi(x)$ и $y = \varphi(x)$ — дифференцированная строго монотонная функция с обратной функцией $x = \psi(y)$, то плотность распределения $g(y)$ случайной величины η определяется из равенства

$$g_\eta(y) = f_\xi(\psi(y)) \cdot |\psi'(y)|.$$

При этом функция распределения

$$G_\eta(y) = \int\limits_{-\infty}^{y} g_\eta(u)du.$$

Пример 1. Случайная величина ξ задана рядом распределения

ξ	-2	-1	0	1	2	3
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Найти ряды распределения таких случайных величин:
а) $\eta = 2\xi$;
б) $\zeta = \xi^2$.

Решение. а) Найдём возможные значения случайной величины $\eta = 2\xi$:

$$\eta_1 = 2 \cdot (-2) = -4; \qquad \eta_2 = 2 \cdot (-1) = -2; \qquad \eta_3 = 2 \cdot 0 = 0;$$
$$\eta_4 = 2 \cdot 1 = 2; \qquad \eta_5 = 2 \cdot 2 = 4; \qquad \eta_6 = 2 \cdot 3 = 6.$$

Разным значениям случайной величины ξ отвечают разные значения случайной величины η. Итак, ряд распределения случайной величины η имеет вид

η	-4	-2	0	2	4	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

б) Найдём возможные значения случайной величины $\zeta = \xi^2$:

$$\zeta_1 = (-2)^2 = 4; \qquad \zeta_2 = (-1)^2 = 1; \qquad \zeta_3 = 0^2 = 0;$$
$$\zeta_4 = 1^2 = 1; \qquad \zeta_5 = 2^2 = 4; \qquad \zeta_6 = 3^2 = 9.$$

Поскольку

$$\zeta_1 = \zeta_5, \qquad \zeta_2 = \zeta_4,$$

то ряд распределения случайной величины η имеет такой вид:

η	0	1	4	9
P	$\frac{1}{6}$	$\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$	$\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$	$\frac{1}{6}$

. □

Пример 2. Случайная величина ξ имеет равномерное распределение на отрезке $[0;9]$. Найти закон распределения случайной величины $\eta = \varphi(\xi)$, если:

а) $\varphi(x) = \begin{cases} 0, & x \leqslant 1; \\ 5, & 1 < x \leqslant 6; \\ 7, & 6 < x \leqslant 9; \\ 2002, & x > 9; \end{cases}$

б) $\varphi(x) = \sqrt{x}$;

в) $\varphi(x) = \begin{cases} 0, & x \leqslant 0; \\ x^2, & 0 < x \leqslant 2; \\ 4, & 2 < x \leqslant 6; \\ x-2, & x > 6. \end{cases}$

Решение. а) Поскольку множество значений функции $y = \varphi(x)$ состоит лишь из четырёх чисел, то случайная величина η дискретная с возможными значениями 0, 5, 7, 2002 и вероятностями, которые им отвечают:

$$\mathbf{P}(\eta = 0) = \mathbf{P}(\xi \in \varphi^{-1}(0)) = \mathbf{P}(\xi \in (-\infty; 1]);$$
$$\mathbf{P}(\eta = 5) = \mathbf{P}(\xi \in \varphi^{-1}(5)) = \mathbf{P}(\xi \in (1; 6]);$$
$$\mathbf{P}(\eta = 7) = \mathbf{P}(\xi \in \varphi^{-1}(7)) = \mathbf{P}(\xi \in (6; 9]);$$
$$\mathbf{P}(\eta = 2002) = \mathbf{P}(\xi \in \varphi^{-1}(2002)) = \mathbf{P}(\xi \in (9; +\infty)).$$

Поскольку случайная величина ξ имеет равномерное распределение на $[0;9]$, то её функция распределения

$$F_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ \dfrac{x}{9}, & 0 < x \leqslant 9; \\ 1, & x > 9. \end{cases}$$

Отсюда получаем:

$$\mathbf{P}(\xi \in (-\infty; 1]) = F_\xi(1) - F_\xi(-\infty) = \frac{1}{9} - 0 = \frac{1}{9};$$
$$\mathbf{P}(\xi \in (1; 6]) = F_\xi(6) - F_\xi(1) = \frac{6}{9} - \frac{1}{9} = \frac{5}{9};$$
$$\mathbf{P}(\xi \in (6; 9]) = F_\xi(9) - F_\xi(6) = \frac{9}{9} - \frac{6}{9} = \frac{3}{9} = \frac{1}{3};$$
$$\mathbf{P}(\xi \in (9; +\infty)) = F_\xi(+\infty) - F_\xi(9) = 1 - \frac{9}{9} = 0.$$

Итак, ряд распределения случайной величины η имеет вид

η	0	5	7
$\mathbf{P}$	$\frac{1}{9}$	$\frac{5}{9}$	$\frac{1}{3}$

б) Поскольку $0 \leqslant \xi \leqslant 9$, то $0 \leqslant \sqrt{\xi} \leqslant 3$. Функция $y = \sqrt{x}$ на интервале $[0;9]$ является строго возрастающей дифференцированной, поэтому случайная величина $\eta = \varphi(\xi)$ имеет непрерывное распределение, причем плотность $g_\eta(y)$ этого распределения удовлетворяет равенству

$$g_\eta(y) = f_\xi(\psi(y)) \cdot |\psi'(y)|,$$

где $f_\xi(\cdot)$ — плотность распределения вероятностей случайной величины ξ:

$$f_\xi(x) = \begin{cases} 0, & x \notin [0;9]; \\ \dfrac{1}{9}, & x \in [0;9], \end{cases}$$

$\psi(\cdot)$ — функция, обратная к функции $y = \sqrt{x}$, т. е.

$$\psi(y) = y^2; \qquad \psi'(y) = 2y \quad (\text{при } 0 \leqslant y \leqslant 3).$$

Тогда

$$g_\eta(y) = \begin{cases} 0, & y \notin [0;3]; \\ \dfrac{1}{9} \cdot 2y = \dfrac{2}{9}y, & y \in [0;3]. \end{cases}$$

Функция распределения $G_\eta(y)$ случайной величины η имеет вид

$$G_\eta(y) = \int\limits_{-\infty}^{y} g(u)du = \begin{cases} 0, & y \leqslant 0; \\ \dfrac{y^2}{9}, & 0 < y \leqslant 3; \\ 1, & y > 3. \end{cases}$$

в) Функция $y = \varphi(x)$ является непрерывной и неубывающей, график которой изображён на рис. 32.

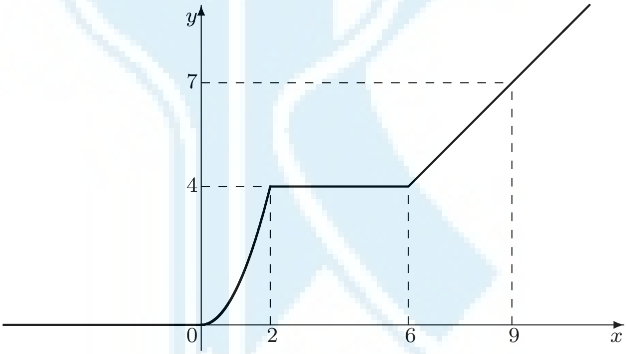


Рис. 32

Образом отрезка $[0;9]$ при отображении $y = \varphi(x)$ является отрезок $[0;7]$. Поскольку ξ распределена на $[0;9]$, то случайная величина η распределена на $[0;7]$. Если $y_0 \neq 4$ и $y_0 \in [0;7]$, то прообразом точки y_0 является единственная точка $x_0 \in [0;9]$. Поэтому

$$\mathbf{P}(\eta = y_0) = \mathbf{P}(\xi = x_0 = \varphi^{-1}(y_0)) = 0, \qquad y_0 \neq 4.$$

Если $y_0 = 4$, то $\varphi^{-1}(4) = [2;6]$. Поэтому

$$\mathbf{P}(\eta = 4) = \mathbf{P}(\xi \in [2;6]) = F_\xi(6) - F_\xi(2) = \frac{6}{9} - \frac{2}{9} = \frac{4}{9}.$$

Если $y \in [0; 4)$, то

$$y = \varphi(x) = x^2 \quad \text{і} \quad x = \psi(y) = \sqrt{y} \quad (0 \leqslant y < 4), \qquad |\psi'(y)| = \frac{1}{2\sqrt{y}}.$$

Поскольку

$$g_\eta(y) = f_\xi(\psi(y)) \cdot |\psi'(y)|,$$

где

$$f_\xi(x) = \begin{cases} 0, & x \notin [0; 9]; \\ \dfrac{1}{9}, & x \in [0; 9], \end{cases}$$

то

$$g_\eta(y) = \frac{1}{9} \cdot \frac{1}{2\sqrt{y}} = \frac{1}{18\sqrt{y}}$$

при $y \in [0; 4)$.

Если $y \in (4; 7]$, то $y = \varphi(x) = x - 2$. Тогда

$$x = \psi(y) = y + 2, \qquad |\psi'(y)| = 1$$

и

$$g_\eta(y) = f_\xi(\psi(y)) \cdot |\psi'(y)| = \frac{1}{9} \cdot 1 = \frac{1}{9}.$$

Итак,

$$g_\eta(y) = \begin{cases} 0, & y \notin [0; 7]; \\ \dfrac{1}{18\sqrt{y}}, & y \in [0; 4); \\ \text{не существует}, & y = 4; \\ \dfrac{1}{9}, & y \in (4; 7], \end{cases}$$

$$
G_\eta(y) = \begin{cases} 0, & y \leqslant 0; \\ \int\limits_0^y \dfrac{du}{18\sqrt{u}}, & 0 < y \leqslant 4; \\ \int\limits_0^4 \dfrac{du}{18\sqrt{u}} + \mathbf{P}(y=4) + \int\limits_4^y \dfrac{1}{9}du, & 4 < y \leqslant 7; \\ 1, & y > 7; \end{cases} =
$$

$$
= \begin{cases} 0, & y \leqslant 0; \\ \dfrac{\sqrt{y}}{9}, & 0 < y \leqslant 4; \\ \dfrac{y+2}{9}, & 4 < y \leqslant 7; \\ 1, & y > 7. \end{cases}
$$

График функции $G_\eta(y)$ изображён на рис. 33. □

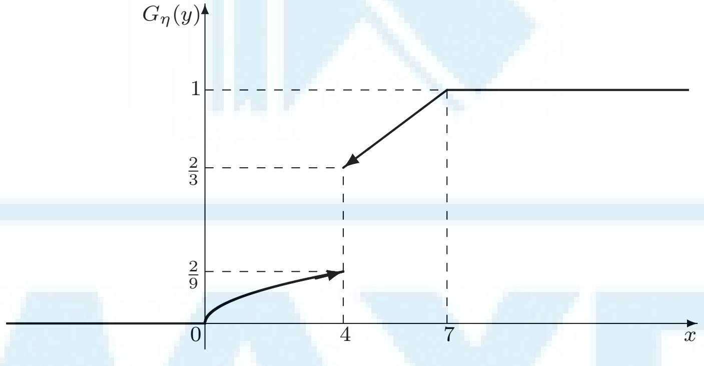


Рис. 33

Задачи к разделу 10

Задача 1. Случайная величина ξ имеет такой ряд распределения:

ξ	1	2	3	4
P	0,1	0,2	0,3	0,4

Построить ряд распределения случайной величины η, если:

а) $\eta = \xi + 5$;

б) $\eta = |\xi - 2|$.

Ответ. а)

η	6	7	8	9
P	0,1	0,2	0,3	0,4

б)

η	0	1	2
P	0,2	0,4	0,4

Задача 2. Случайная величина ξ имеет такой ряд распределения:

ξ	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π
P	0,2	0,2	0,2	0,2	0,2

Построить ряд распределения случайной величины X, если:

а) $X = \sin(2\xi)$;

б) $X = \dfrac{4\xi}{\pi}$.

Ответ. а)

X	-1	0	1
P	0,2	0,6	0,2

б)

X	0	1	2	3	4
P	0,2	0,2	0,2	0,2	0,2

Задача 3. Случайная величина X имеет такой ряд распределения:

X	-3	-2	-1	0	1	2
P	0,25	0,05	0,2	0,1	0,3	0,1

Построить ряд распределения случайной величины Y, если:

а) $Y = X^3$;

б) $Y = |X|$.

Ответ. а)

Y	-27	-8	-1	0	1	8
P	0,25	0,05	0,2	0,1	0,3	0,1

б)

Y	0	1	2	3
P	0,1	0,5	0,15	0,25

Задача 4. Случайная величина X имеет такой ряд распределения:

X	-9	-4	-1	0	1	4	9	16
P	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Построить ряд распределения случайной величины Y, если:
а) $Y = \sqrt{|X|}$;
б) $Y = 2X + 1$.

Ответ. а)

Y	0	1	2	3	4
P	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$

б)

Y	-17	-7	-1	1	3	9	19	33
P	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Задача 5. Случайная величина ξ равномерно распределена на отрезке $[0; 4]$. Найти распределение случайной величины $X = [\xi]$, где $[a]$ означает целую часть числа a.

Ответ.

X	0	1	2	3
P	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Задача 6. Случайная величина ξ равномерно распределена на отрезке $[-5; 5]$. Найти распределение случайной величины $X = \varphi(\xi)$, если

$$\varphi(x) = \begin{cases} 3, & x \leqslant 0; \\ 7, & 0 < x \leqslant 2; \\ 11, & x > 2. \end{cases}$$

Ответ.

X	3	7	11
P	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{3}{10}$

Задача 7. Случайная величина ξ равномерно распределена на отрезке $[0; 2]$. Найти плотность и функцию распределения случайной величины $X = \xi^2$.

Ответ. $f(x) = \begin{cases} \dfrac{1}{4\sqrt{x}}, & x \in [0; 4]; \\ 0, & x \notin [0; 4], \end{cases}$ $F(x) = \begin{cases} 0, & x \leqslant 0; \\ \dfrac{\sqrt{x}}{2}, & 0 < x \leqslant 4; \\ 1, & x > 4. \end{cases}$

Задача 8. Случайная величина ξ равномерно распределена на отрезке $[1; 10]$. Найти плотность и функцию распределения случайной величины $X = \sqrt{\xi - 1}$.

Ответ. $f(x) = \begin{cases} \dfrac{2x}{9}, & x \in [0; 3]; \\ 0, & x \notin [0; 3], \end{cases}$ $F(x) = \begin{cases} 0, & x \leqslant 0; \\ \dfrac{x^2}{9}, & 0 < x \leqslant 3; \\ 1, & x > 3. \end{cases}$

Задача 9. Случайная величина ξ имеет такую плотность распределения:

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

Найти плотность распределения случайной величины $X = \dfrac{1}{\xi}$.

Ответ. $f(x) = \dfrac{1}{\pi(1 + x^2)}$.

Задача 10. Случайная величина ξ имеет такую плотность распределения:

$$f(x) = \begin{cases} xe^{-\frac{x^2}{2}}, & x > 0; \\ 0, & x \leqslant 0. \end{cases}$$

Найти плотность распределения случайной величины $X = e^{-\xi^2}$.

Ответ. $f(x) = \begin{cases} \dfrac{1}{2\sqrt{x}}, & x \in [0; 1]; \\ 0, & x \notin [0; 1]. \end{cases}$

Задача 11. Случайная величина ξ имеет такую плотность распределения:

$$f(x) = \begin{cases} \sin x, & x \in \left[0; \dfrac{\pi}{2}\right]; \\ 0, & x \notin \left[0; \dfrac{\pi}{2}\right]. \end{cases}$$

Найти плотность распределения случайной величины $X = \xi^4$.

Ответ. $f(x) = \begin{cases} \dfrac{\sin\sqrt[4]{x}}{4\sqrt[4]{x^3}}, & x \in \left[0; \dfrac{\pi^4}{16}\right]; \\ 0, & x \notin \left[0; \dfrac{\pi^4}{16}\right]. \end{cases}$

Задача 12. Случайная величина ξ имеет такую плотность распределения:

$$f(x) = \begin{cases} \dfrac{1}{2}\cos x, & x \in \left[-\dfrac{\pi}{2}; \dfrac{\pi}{2}\right]; \\ 0, & x \notin \left[-\dfrac{\pi}{2}; \dfrac{\pi}{2}\right]. \end{cases}$$

Найти плотность распределения случайной величины $X = \xi^3$.

Ответ. $f(x) = \begin{cases} \dfrac{\cos\sqrt[3]{x}}{6\sqrt[3]{x^2}}, & x \in \left[-\dfrac{\pi^3}{8}; 0\right) \bigcup \left(0; \dfrac{\pi^3}{8}\right]; \\ 0, & x \notin \left[-\dfrac{\pi^3}{8}; 0\right) \bigcup \left(0; \dfrac{\pi^3}{8}\right]. \end{cases}$

Задача 13. Случайная величина ξ равномерно распределена на отрезке $[-10; 10]$.

Найти закон распределения случайной величины $X = \varphi(\xi)$, если

$$\varphi(x) = \begin{cases} 0, & x \leqslant 2; \\ x - 2, & 2 < x \leqslant 9; \\ 7, & x > 9. \end{cases}$$

Ответ. Распределение случайной величины X является смесью дискретного и непрерывного распределений.

$F(x) = \begin{cases} 0, & x \leqslant 0; \\ 0{,}6 + 0{,}05x, & 0 < x \leqslant 7; \\ 1, & x > 7. \end{cases}$

Задача 14. Случайная величина ξ равномерно распределена на отрезке $[0; 5]$.

Найти закон распределения случайной величины $X = \varphi(\xi)$, если

$$\varphi(x) = \begin{cases} 3, & x \in [-2\pi; 2\pi]; \\ \ln\left(\sqrt{\cos x + 3}\right), & x \notin [-2\pi; 2\pi]. \end{cases}$$

Ответ. Случайная величина X имеет вырожденное распределение:

$$\mathbf{P}\{X = 3\} = 1.$$

Раздел 11. Предельные теоремы теории вероятностей

Ключевые слова

Большой
Вероятность
Закон
Неравенство
Предельный
Случайный
Средний
Стойкость
Теорема
Теория
Центральный
Число
Явление

Предельные теоремы теории вероятностей можно условно разбить на два класса: законы больших чисел и центральные предельные теоремы.

Законы больших чисел математически описывают стойкость средних значений массовых случайных явлений. Исторически первым законом больших чисел была следующая теорема.

Теорема 1 (теорема Я. Бернулли). Пусть k — количество появлений события A в серии из n независимых испытаний, в каждом из которых событие A происходит с вероятностью p. Тогда для произвольного $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{k}{n} - p\right| < \varepsilon\right\} = 1.$$

Другими словами, с вероятностью, сколь угодно близкой к единице, можно утверждать, что при довольно большом количестве независимых испытаний статистическая частота $\dfrac{k}{n}$ появления события A как угодно мало отличается от вероятности появления события A в одном испытании.

Эта теорема, в частности, объясняет, почему при многократном подбрасывании симметричной монеты количество гербов составляет приблизительно половину от общего количества подбрасываний.

Обобщением теоремы 1 является следующая теорема.

Теорема 2 (теорема П. Л. Чебишева). Пусть ξ_i — последовательность попарно независимых случайных величин с конечными математическими ожиданиями $\mathbf{M}(\xi_i)$ и ограниченными дисперсиями $\mathbf{D}(\xi_i) \leqslant c_0$. Тогда для произвольного $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}(\xi_i)\right| < \varepsilon\right\} = 1.$$

В обеих теоремах существенно использовано неравенство Чебишева, которое, впрочем, имеет и самостоятельную ценность.

Лемма 1 (неравенство Чебишева). Пусть ξ — произвольная случайная величина с дисперсией $\mathbf{D}(\xi)$ и математическим ожиданием $\mathbf{M}(\xi)$. Тогда для произвольного $\varepsilon > 0$

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < \varepsilon\} \geqslant 1 - \frac{\mathbf{D}(\xi)}{\varepsilon^2}.$$

Неравенство Чебишева можно записать также в форме

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| \geqslant \varepsilon\} < \frac{\mathbf{D}(\xi)}{\varepsilon^2}.$$

Кроме законов больших чисел, описанных в теоремах 1 и 2, наблюдается еще одно довольно интересное явление, которое заключается в том, что при большом количестве случайных слагаемых, каждое из которых вносит лишь небольшой вклад в общую сумму, распределение каждого из случайных слагаемых не влияет на суммарный результат. Более точное утверждение сформулировано в следующей теореме.

Теорема 3 (центральная предельная теорема). Пусть ξ_i — одинаково распределённые случайные величины с математическим ожиданием $\mathbf{M}(\xi_i) = a$ и дисперсией $\mathbf{D}(\xi_i) = \sigma^2$. Тогда при большом значении n распределение суммы $\xi_1 + \xi_2 + \ldots + \xi_n = \xi$ близко к нормальному распределению.

Если результирующая случайная величина ξ имеет математическое ожидание $\mathbf{M}(\xi)$ и дисперсию $\mathbf{D}(\xi)$, то из теоремы 3 вытекает, что

$$\mathbf{P}\{a \leqslant \xi \leqslant b\} \approx \Phi\left(\frac{b - \mathbf{M}(\xi)}{\sqrt{\mathbf{D}(\xi)}}\right) - \Phi\left(\frac{a - \mathbf{M}(\xi)}{\sqrt{\mathbf{D}(\xi)}}\right),$$

следовательно,

$$\mathbf{P}\big\{\mathbf{M}(\xi) - \delta \leqslant \xi \leqslant \mathbf{M}(\xi) + \delta\big\} \approx 2\Phi\left(\frac{\delta}{\sqrt{\mathbf{D}(\xi)}}\right).$$

Из последних равенств вытекает, что вероятность p события можно оценить, пользуясь экспериментальными данными. Действительно, если в серии из n независимых испытаний исследуемое событие A происходит k раз, то статистическая частота появления события A равняется $\frac{k}{n}$. Эта частота является случайной величиной ξ, для которой математическое ожидание

$$\mathbf{M}(\xi) = p,$$

а дисперсия

$$\mathbf{D}(\xi) = \frac{p(1-p)}{n}.$$

Тогда при достаточно больших значениях n $(n \geqslant 30)$ выполняется равенство

$$\mathbf{P}\left\{\frac{k}{n} - \delta < p < \frac{k}{n} + \delta\right\} \approx 2\Phi\left(\frac{\delta}{\sqrt{\mathbf{D}(\xi)}}\right) = 2\Phi\left(\delta\sqrt{\frac{n}{p(1-p)}}\right).$$

Пример 1. Используя неравенство Чебишева, найти вероятность события A, которое заключается в том, что случайная величина X примет значение, которое будет отличаться от математического ожидания $\mathbf{M}(X)$ на величину, не превышающую утроенного среднего квадратического отклонения. Изменится ли ответ, если известно, что случайная величина X имеет нормальное распределение?

Решение. Согласно неравенству Чебишева

$$\mathbf{P}(A) = \mathbf{P}\{|X - \mathbf{M}(X)| < 3\sigma(X)\} \geqslant 1 - \frac{\mathbf{D}(X)}{(3\sigma(X))^2} =$$

$$= 1 - \frac{\mathbf{D}(X)}{9\mathbf{D}(X)} = \frac{8}{9}.$$

Итак,

$$\mathbf{P}(A) \geqslant \frac{8}{9} \approx 0{,}89.$$

Если X имеет нормальное распределение, то

$$\mathbf{P}(A) = \mathbf{P}\{|X - \mathbf{M}(X)| < 3\sigma(X)\} = 2\Phi\left(\frac{3\sigma(X)}{\sigma(X)}\right) = 2\Phi(3) =$$
$$= 0{,}9973. \qquad \square$$

Пример 2. При производстве дискет брак составляет 1 %. Сколько дискет нужно отобрать для проверки качества, чтобы с вероятностью 0,95 можно было утверждать, что в случайной выборке дискет процент бракованных отличается от 1 % не более чем на 0,5 %?

Решение. Количество бракованных дискет является случайной величиной

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n,$$

где $\xi_1, \xi_2, \ldots, \xi_n$ — независимые одинаково распределённые случайные величины; ξ_i — случайная величина, которая равняется количеству бракованных дискет при изготовлении одной дискеты, т. е. ξ_i может принимать значение или 0, или 1 с вероятностью соответственно 0,99 и 0,01.

Если n — достаточно большое число, то согласно центральной предельной теореме распределение случайной величины ξ близко к нормальному. Поэтому

$$\mathbf{P}\left\{\left|\frac{k}{n} - p\right| < \delta\right\} \approx 2\Phi\left(\frac{\delta}{\sqrt{D(\xi)}}\right) = 2\Phi\left(\delta\sqrt{\frac{n}{p(1-p)}}\right),$$

где $\dfrac{k}{n}$ — частота брака; $p = 0{,}01$; $\delta = 0{,}005$; n — неизвестное количество дискет.

Число n нужно выбрать таким, чтобы

$$\mathbf{P}\left\{\left|\frac{k}{n} - 0{,}01\right| < 0{,}005\right\} \approx 2\Phi\left(0{,}005\sqrt{\frac{n}{0{,}01 \cdot 0{,}99}}\right) = 0{,}95.$$

Другими словами,

$$\Phi\left(0{,}5\sqrt{\frac{n}{99}}\right) = 0{,}475.$$

По таблице значений функции Лапласа (прил. 1) находим значение аргумента x такое, что $\Phi(x) = 0{,}475$:

$$x = 1{,}96.$$

Решив уравнение

$$0{,}5\sqrt{\frac{n}{99}} = 1{,}96,$$

получим

$$n \geqslant 99 \cdot \left(\frac{1{,}96}{0{,}5}\right)^2 \geqslant 1522. \qquad \square$$

Пример 3. При анонимном тестировании оказалось, что 10 % работников предприятия совсем не употребляют спиртного. Случайная величина X — количество людей, которые совсем не употребляют спиртного в случайной выборке из 900 рабочих. Указать границы, в которые попадает случайная величина X с вероятностью 0,9.

Решение. Согласно центральной предельной теореме можно считать, что случайная величина X имеет распределение, близкое к нормальному. По условию

$$p = 0{,}1; \qquad n = 900.$$

Тогда

$$\mathbf{P}\left\{\left|\frac{X}{900} - 0{,}1\right| < \delta\right\} \approx 2\Phi\left(\delta\sqrt{\frac{900}{0{,}1 \cdot 0{,}9}}\right) = 0{,}9.$$

Итак,

$$\Phi\left(100\delta\right) = 0{,}45.$$

По таблице значений функции Лапласа находим значение аргумента x такое, что $\Phi(x) = 0{,}45$:

$$x = 1{,}65.$$

Отсюда $100\delta = 1{,}65$ и $\delta = 0{,}0165$.

Итак, с вероятностью 0,9 будет выполняться условие

$$\mathbf{P}\left\{\left|\frac{X}{900} - 0{,}1\right| < 0{,}0165\right\} \geqslant 0{,}9,$$

что равносильно тому, что случайная величина X попадёт в интервал $(75; 105)$ с вероятностью 0,9. □

Пример 4. Пусть x — случайно выбранное число из отрезка $[0; 1]$,

$$x = \alpha_1(x)\alpha_2(x)\ldots\alpha_k(x)\ldots\ -$$

десятичное представление числа x, $\alpha_k(x) \in \{0, 1, \ldots, 9\}$. Пусть $N_7(x, k)$ — количество цифр «7», которые встретились в десятичном представлении числа x до k-го места включительно. Найти вероятность того, что частота цифры «7», которая равняется

$$\lim_{k\to\infty} \frac{N_7(x,k)}{k},$$

будет отличаться от 0,1 не более чем на ε.

Решение. $\xi = N_7(x, k)$ — случайная величина, которая является суммой

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n,$$

где $\xi_i = 0$, если $\alpha_i(x) \neq 7$, и $\xi_i = 1$, если $\alpha_i(x) = 7$. Другими словами, ξ_i имеет такое распределение:

ξ_i	0	1
$\mathbf{P}$	0,9	0,1

Легко увидеть, что

$$\mathbf{M}(\xi_i) = 0{,}1; \qquad \mathbf{M}(\xi_i^2) = 0{,}1.$$

Поэтому

$$\mathbf{D}(\xi_i) = \mathbf{M}(\xi_i^2) - \mathbf{M}^2(\xi_i) = 0{,}1 - 0{,}01 = 0{,}09.$$

Поскольку случайные величины ξ_i независимы и имеют ограниченные дисперсии и математические ожидания, то выполняется теорема Чебишева:

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}(\xi_i)\right| < \varepsilon\right\} = \\ = \lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{N_7(x,k)}{n} - 0{,}1\right| < \varepsilon\right\} = 1.$$

Итак, для произвольного $\varepsilon > 0$ с вероятностью 1 частота цифры «7» в десятичном представлении случайно выбранного числа x отличается от 0,1 не более чем на ε. □

Задачи к разделу 11

Задача 1. Используя неравенство Чебишева, оценить вероятность того, что

$$|\xi - \mathbf{M}(\xi)| < 0{,}3,$$

если $\mathbf{D}(\xi) = 0{,}0025$. Оценить вероятность этого же события, если известно, что ξ имеет нормальное распределение.

Ответ. Согласно неравенству Чебишева

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 0{,}3\} \geqslant 0{,}97.$$

Если предположить, что ξ имеет нормальное распределение, то

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 0{,}3\} \approx 1 \quad \text{(с точностью до 0,0001).}$$

Задача 2. Используя неравенство Чебишева, оценить вероятность того, что

$$|\xi - \mathbf{M}(\xi)| > 0{,}1,$$

если $\sigma(\xi) = 0{,}4$. Оценить вероятность этого же события, если известно, что ξ имеет нормальное распределение.

Ответ. Согласно неравенству Чебишева

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| > 0{,}1\} < 16.$$

Если предположить, что ξ имеет нормальное распределение, то

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| > 0{,}1\} \approx 0{,}8026.$$

Задача 3. Масса деталей, изготавливаемых на станке, является случайной величиной, среднее значение которой (математическое ожидание) равняется 1,2 кг. Дисперсия этой величины равняется 0,012. Используя неравенство Чебишева, оценить вероятность того, что:

а) отклонение массы детали от её среднего значения по абсолютной величине не превысит 0,2;

б) масса детали примет значение от 1,18 до 1,22.

Ответ. а) $\mathbf{P} \geqslant 0{,}7$; б) $\mathbf{P} \geqslant 0$.

Задача 4. Прибор состоит из 10 элементов, которые работают независимо. Вероятность отказа каждого элемента за время t равняется 0,05. Используя неравенство Чебишева, оценить вероятность того, что абсолютная величина разности между количеством элементов, которые отказали, и средним количеством (математическим ожиданием) отказов за время t будет меньше 2.

Ответ. $\mathbf{P} \geqslant 0{,}88$.

Задача 5. Из 1000 изделий наугад отобраны 200. После проверки среди них выявлено 15 бракованных. Приняв долю бракованных изделий как вероятность изготовления брака, оценить вероятность того, что во всей партии бракованных изделий будет не больше 10 % и не меньше 5 %.

Ответ. $\mathbf{P} \geqslant 0{,}889$.

Задача 6. В стране N проживает 50 млн людей. ВИЧ-тестирование прошли 1000 случайно отобранных граждан, среди которых оказалось 20 инфицированных. Приняв долю инфицированных в исследованной группе как вероятность того, что случайно отобранный гражданин страны N инфицирован, оценить вероятность того, что в данной стране процент ВИЧ-инфицированных составляет не больше 3 % и не меньше 1 %.

Ответ. $\mathbf{P} \geqslant 0{,}99996$.

Задача 7. Сколько изделий необходимо отобрать для проверки качества продукции, чтобы с вероятностью 0,95 можно было утверждать, что доля бракованных деталей отличается от вероятности $p = 0{,}1$ выпуска бракованной детали не более чем на 0,02?

Ответ. $n \geqslant 865$.

Задача 8. При производстве полиэтиленовых пакетов брак составляет 5 %. Сколько изделий нужно отобрать для проверки качества продукции, чтобы с вероятностью 0,9 можно было утверждать, что доля бракованных пакетов составляет от 4 до 6 %?

Ответ. $n \geqslant 1294$.

Задача 9. Случайная величина ξ имеет такой закон распределения:

ξ	3	5
P	0,4	0,6

Используя неравенство Чебишева, оценить вероятность того, что:

а) $|\xi - \mathbf{M}(\xi)| < 0{,}2$;

б) $|\xi - \mathbf{M}(\xi)| < 2$.

Вычислить точные вероятности этих событий.

Ответ. а) Согласно неравенству Чебишева

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 0{,}2\} \geqslant 0.$$

Действительно,

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 0{,}2\} = 0.$$

б) Согласно неравенству Чебишева

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 2\} \geqslant 0{,}76.$$

Действительно,

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 2\} = 1.$$

Задача 10. Случайная величина ξ имеет такой закон распределения:

ξ	1	5
P	0,7	0,3

Используя неравенство Чебишева, оценить вероятность того, что:
а) $|\xi - \mathbf{M}(\xi)| < 1$;
б) $|\xi - \mathbf{M}(\xi)| < 2$.
Вычислить точные вероятности этих событий.

Ответ. а) Согласно неравенству Чебишева

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 1\} \geqslant 0.$$

Действительно,

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 1\} = 0.$$

б) Согласно неравенству Чебишева

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 2\} \geqslant 0{,}16.$$

Действительно,

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| < 2\} = 0{,}7.$$

Задача 11. Пусть x — случайно выбранное число из отрезка $[0;1]$,

$$x = \alpha_1(x)\alpha_2(x)\ldots\alpha_k(x)\ldots \text{ —}$$

десятичное представление числа x, $\alpha_k(x) \in \{0,1,\ldots,9\}$. Пусть $N_5(x,k)$ — количество цифр «5», которые встретились в десятичном представлении числа x до k-го места включительно. Найти вероятность того, что частота цифры «5», которая равняется

$$\lim_{k\to\infty} \frac{N_5(x,k)}{k},$$

будет находиться в пределах от 0,0999 до 0,1001.

Ответ. $\mathbf{P} = 1$.

Задача 12. Пусть

$$x = \alpha_1(x)\alpha_2(x)\ldots\alpha_k(x)\ldots \text{ —}$$

двоичное представление числа $x \in [0;1]$, $\alpha_k(x) \in \{0,1\}$, $N_0(x,k)$ — количество нулей, которые встретились в двоичном представлении числа x до k-го места включительно. Найти вероятность того, что частота цифры «0», которая равняется

$$\lim_{k\to\infty}\frac{N_0(x,k)}{k},$$

будет находиться в пределах от 0,49999 до 0,50001.

Ответ. $\mathbf{P} = 1$.

Часть III

Математическая статистика

Раздел 12. Элементы математической статистики. Выборочный метод

Ключевые слова

Варианта
Вариационный
Вероятность
Величина
Выборка
Выборочный
Гистограмма
Дисперсия
Интервал
Интервальный
Квадратический
Медиана
Медианный
Метод
Мода
Модальный
Оптимальный
Отклонение

Относительный	Средний
Плотность	Статистический
Полигон	Сумма
Поправка	Точечный
Произведение	Функция
Размах	Частный
Распределение	Частота
Ряд	Эмпирический

12.1. Выборочный метод

Пусть для изучения количественного (дискретного или непрерывного) признака X из генеральной совокупности получена выборка $x_1, x_2, \ldots, x_n$ объёма n.

Если записать элементы выборки в порядке возрастания, получим *вариационный ряд*.

Наблюдаемые **разные** значения x_i признака X называют *вариантами*, количество значений одной варианты в выборке — её *частотой* n_i (сумма частот всех вариант равняется объёму выборки), отношение частоты к объёму выборки — *относительной частотой* или *эмпирической вероятностью* $w_i = \dfrac{n_i}{n}$ (сумма относительных частот всех вариант равняется единице).

Статистическим распределением выборки называют перечень вариант вариационного ряда с соответствующими им частотами и/или относительными частотами. Статистическое распределение также задают в виде последовательности замкнутых справа полуинтервалов и соответствующих им частот и/или относительных частот (частотой интервала считают сумму частот всех вариант из данного интервала).

Чтобы подчеркнуть указанные различия в первом случае говорят о *точечном*, а во втором — об *интервальном* статистическом распределении выборки.

Пример 1. Во время исследования количественного признака X из генеральной совокупности была получена выборка

$$4, 3, 6, 4, 7, 2, 5, 1, 2, 5, 4, 4, 3, 5, 6, 3, 4, 1, 3, 4.$$

Найти объём выборки, построить вариационный ряд выборки и её статистическое распределение.

Решение. Поскольку выборка состоит из 20 значений, то объём выборки $n = 20$.

Построим вариационный ряд выборки:

$$1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7.$$

В данной выборке всего семь разных значений, т. е. вариант:

$$1, 2, 3, 4, 5, 6, 7.$$

Найдем их частоты:

$$n_1 = 2; \quad n_2 = 2; \quad n_3 = 4; \quad n_4 = 6; \quad n_5 = 3; \quad n_6 = 2; \quad n_7 = 1.$$

Запишем искомое статистическое распределение:

x_i	1	2	3	4	5	6	7
n_i	2	2	4	6	3	2	1

. □

Пример 2. Выборка задана распределением частот:

x_i	3	5	7	10	15
n_i	2	4	7	4	3

.

Найти распределение относительных частот.

Решение. Вычислим объём выборки:

$$n = 2 + 4 + 7 + 4 + 3 = 20.$$

Определим относительные частоты:

$$w_1 = \frac{2}{20} = 0{,}1; \qquad w_2 = \frac{4}{20} = 0{,}2; \qquad w_3 = \frac{7}{20} = 0{,}35;$$

$$w_4 = \frac{4}{20} = 0{,}2; \qquad w_5 = \frac{3}{20} = 0{,}15.$$

Итак, искомое распределение относительных частот имеет вид

x_i	3	5	7	10	15
w_i	0,1	0,2	0,35	0,2	0,15

.

Контроль: $0{,}1 + 0{,}2 + 0{,}35 + 0{,}2 + 0{,}15 = 1$. □

Эмпирической функцией распределения (функцией распределения выборки) называют функцию

$$F^{\star}(x) = \frac{n_x}{n},$$

где n_x — количество элементов выборки, меньших x (т. е. сумма частот всех вариант, меньших x); n — объём выборки.

Рассмотрим свойства эмпирической функции распределения.

Свойство 1. Эмпирическая функция распределения $F^{\star}(x)$ неотрицательная и не больше единицы при любых значениях x:

$$F^{\star}(x) \in [0; 1].$$

Свойство 2. Эмпирическая функция распределения $F^{\star}(x)$ неубывающая:

$$x < y \quad \Rightarrow \quad F^{\star}(x) \leqslant F^{\star}(y).$$

Свойство 3. Если x_1 — наименьшая варианта, а x_n — наибольшая, то эмпирическая функция распределения $F^{\star}(x) = 0$ при $x \leqslant x_1$ и $F^{\star}(x) = 1$ при $x > x_n$.

Пример 3. Найти эмпирическую функцию распределения по данному распределению выборки:

x_i	3	5	7	10	15
n_i	2	4	7	4	3

Решение. Определим объём выборки:

$$n = 2 + 4 + 7 + 4 + 3 = 20.$$

Поскольку наименьшая варианта равняется трём,

$$F^{\star}(x) = 0$$

при всех $x \leqslant 3$.

Значение $X < 5$, а именно $x_1 = 3$, наблюдалось дважды, поэтому

$$F^{\star}(x) = \frac{2}{20} = 0{,}1$$

при $3 < x \leqslant 5$.

Значение $X < 7$, а именно $x_1 = 3$ и $x_2 = 5$, наблюдалось $2 + 4 = 6$ раз, поэтому

$$F^\star(x) = \frac{6}{20} = 0{,}3$$

при $5 < x \leqslant 7$.

Значение $X < 10$, а именно $x_1 = 3$, $x_2 = 5$ и $x_3 = 7$, наблюдалось $2 + 4 + 7 = 13$ раз, поэтому

$$F^\star(x) = \frac{13}{20} = 0{,}65$$

при $7 < x \leqslant 10$.

Значение $X < 15$, а именно $x_1 = 3$, $x_2 = 5$, $x_3 = 7$ и $x_4 = 10$, наблюдалось $2 + 4 + 7 + 4 = 17$ раз, поэтому

$$F^\star(x) = \frac{17}{20} = 0{,}85$$

при $10 < x \leqslant 15$.

Поскольку $x_5 = 15$ — наибольшая варианта, то

$$F^\star(x) = 1$$

при $x > 15$.

Итак, запишем искомую эмпирическую функцию:

$$F^\star(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 3; \\ 0{,}1 & \text{при} \quad 3 < x \leqslant 5; \\ 0{,}3 & \text{при} \quad 5 < x \leqslant 7; \\ 0{,}65 & \text{при} \quad 7 < x \leqslant 10; \\ 0{,}85 & \text{при} \quad 10 < x \leqslant 15; \\ 1 & \text{при} \quad x > 15. \end{cases}$$

График этой функции изображён на рис. 34. □

Полигоном частот называют ломаную, отрезки которой соединяют точки

$$(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k),$$

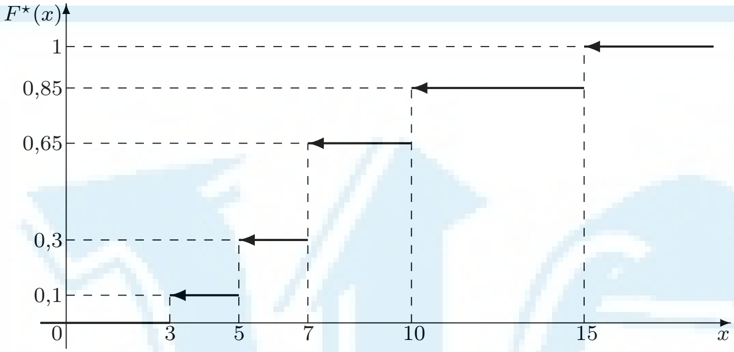


Рис. 34

где x_i — варианты выборки; n_i — соответствующие частоты, $i = 1, 2, \ldots$.

Полигоном относительных частот называют ломаную, отрезки которой соединяют точки

$$(x_1, w_1), (x_2, w_2), \ldots, (x_k, w_k),$$

где x_i — варианты выборки; w_i — соответствующие эмпирические вероятности, $i = 1, 2, \ldots$.

При непрерывном распределении признака X в случае большого количества наблюдений весь интервал, в котором размещены наблюдаемые значения признака, как правило, разбивают на несколько частных интервалов одинаковой длины h и находят n_i — сумму частот вариант, которые попали в i-й интервал. Для выбора *оптимальной величины интервала* рекомендуют использовать формулу

$$h \approx \frac{x_{\max} - x_{\min}}{1 + 3{,}2 \cdot \lg n},$$

где $x_{\max}$, $x_{\min}$ — соответственно наибольшее и наименьшее значение в выборке; n — объём выборки.

Если задано интервальное статистическое распределение выборки, то для построения полигона частот или эмпирических вероятностей по данным выборки соединяют точки, абсциссами которых являются значения середин частных интервалов, а ординатами — соответствующие им значения частот или эмпирических вероятностей.

Гистограммой частот называют ступенчатую фигуру, которая состоит из прямоугольников, основами которых являются частные интервалы длины h, а высоты равняются отношению $\frac{n_i}{h}$ (*плотность частоты*). Площадь i-го частного прямоугольника

$$h \cdot \frac{n_i}{h} = n_i,$$

т. е. сумма частот вариант, которые попали в i-й частный интервал. Площадь гистограммы частот равняется сумме всех частот, т. е. объёму выборки n.

Гистограммой относительных частот называют ступенчатую фигуру, которая состоит из прямоугольников, основами которых являются частные интервалы длины h, а высоты равняются отношению $\frac{w_i}{h}$ (*плотность относительной частоты*). Площадь i-го частного прямоугольника

$$h \cdot \frac{w_i}{h} = w_i,$$

т. е. равняется сумме относительных частот вариант, попавших в i-й интервал. Площадь гистограммы относительных частот равняется сумме всех относительных частот, т. е. единице.

Если задано точечное статистическое распределение выборки, то для построения гистограммы частот или эмпирических вероятностей по данным выборки в качестве частных интервалов берут интервалы, серединами которых являются значения соответствующих им вариант.

Пример 4. Построить полигон частот по заданному распределению выборки:

x_i	1	2	3	4	5	6	7
n_i	5	1	8	3	2	2	6

Решение. Отложим на оси абсцисс значения вариант x_i, а на оси ординат — значения соответствующих им частот n_i. Последовательно соединяя между собой точки (x_i, n_i) отрезками, получаем полигон частот (рис. 35). □

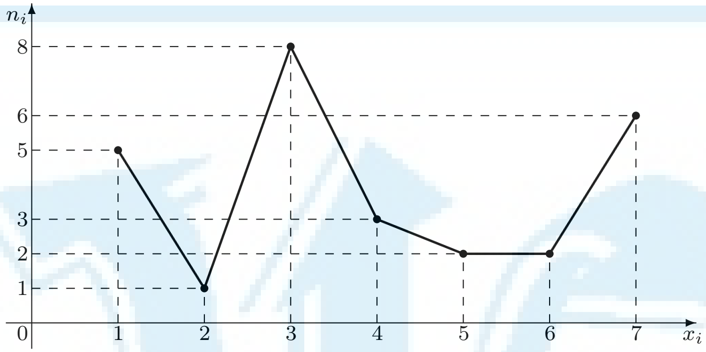


Рис. 35

Пример 5. Выборка задана интервальным распределением частот:

$(x_i;x_{i+1}]$	$(1;3]$	$(3;5]$	$(5;7]$	$(7;9]$	$(9;11]$
n_i	13	9	5	16	7

Построить полигон относительных частот.

Решение. Найдём объём выборки:

$$n = 13 + 9 + 5 + 16 + 7 = 50.$$

Для построения интервального статистического распределения относительных частот определим относительные частоты:

$$w_1 = \frac{n_1}{n} = \frac{13}{50} = 0{,}26; \qquad w_2 = \frac{n_2}{n} = \frac{9}{50} = 0{,}18;$$

$$w_3 = \frac{n_3}{n} = \frac{5}{50} = 0{,}1; \qquad w_4 = \frac{n_4}{n} = \frac{16}{50} = 0{,}32;$$

$$w_5 = \frac{n_5}{n} = \frac{7}{50} = 0{,}14.$$

Построим интервальное статистическое распределение относительных частот:

$(x_i;x_{i+1}]$	$(1;3]$	$(3;5]$	$(5;7]$	$(7;9]$	$(9;11]$
w_i	0,26	0,18	0,1	0,32	0,14

Найдём середины частных интервалов:

$$x_1^{\text{сер}} = \frac{x_1 + x_2}{2} = \frac{1+3}{2} = 2; \qquad x_2^{\text{сер}} = \frac{x_2 + x_3}{2} = \frac{3+5}{2} = 4;$$

$$x_3^{\text{сер}} = \frac{x_3 + x_4}{2} = \frac{5+7}{2} = 6; \qquad x_4^{\text{сер}} = \frac{x_4 + x_5}{2} = \frac{7+9}{2} = 8;$$

$$x_5^{\text{сер}} = \frac{x_5 + x_6}{2} = \frac{9+11}{2} = 10.$$

Отложим на оси абсцисс значения середин частных интервалов $x_i^{\text{сер}}$, а на оси ординат — значения соответствующих им относительных частот w_i. Последовательно соединяя между собой точки $(x_i^{\text{сер}}, w_i)$ отрезками, получаем полигон относительных частот (рис. 36). □

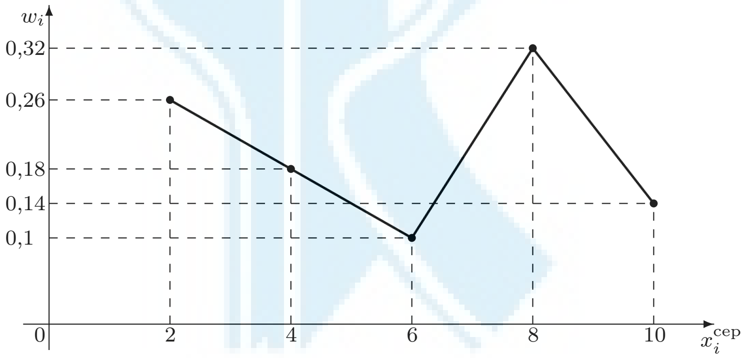


Рис. 36

Пример 6. Выборка задана интервальным распределением частот:

$(x_i; x_{i+1}]$	$(1; 2]$	$(2; 3]$	$(3; 4]$	$(4; 5]$	$(5; 6]$	$(6; 7]$	$(7; 8]$
n_i	19	9	12	14	7	22	17

Построить гистограмму частот.

Решение. Найдём длину интервалов:

$$h = x_{i+1} - x_i \equiv 1.$$

Определим плотность частоты:

$$n_1^h = \frac{n_1}{h} = \frac{19}{1} = 19; \quad n_2^h = \frac{n_2}{h} = \frac{9}{1} = 9; \quad n_3^h = \frac{n_3}{h} = \frac{12}{1} = 12;$$

$$n_4^h = \frac{n_4}{h} = \frac{14}{1} = 14; \quad n_5^h = \frac{n_5}{h} = \frac{7}{1} = 7; \quad n_6^h = \frac{n_6}{h} = \frac{22}{1} = 22;$$

$$n_7^h = \frac{n_7}{h} = \frac{17}{1} = 17.$$

Построим на оси абсцисс заданные частные интервалы длины $h = 1$. Проведём над этими интервалами отрезки, параллельные оси абсцисс и расположенные от нее на расстояниях, которые равняются соответствующим плотностям частоты n_i^h. Концы этих отрезков соединим с осью абсцисс отрезками, параллельными оси ординат. Полученная гистограмма частот изображена на рис. 37. □

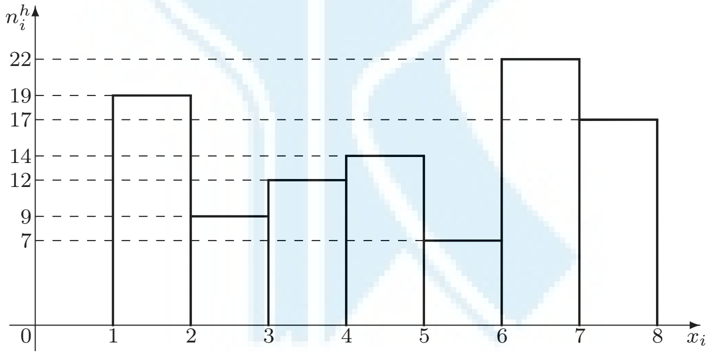


Рис. 37

Пример 7. Выборка задана распределением частот:

x_i	2	4	6	8	10	12	14
n_i	2	3	5	1	4	2	3

.

Построить гистограмму относительных частот.

Решение. Чтобы построить гистограмму относительных частот, данное дискретное статистическое распределение частот нужно превратить в интервальное статистическое распределение относительных частот (эмпирических вероятностей) и найти плотность этих относительных частот. Для этого сначала определим объём выборки и эмпирические вероятности:

$$n = 2 + 3 + 5 + 1 + 4 + 2 + 3 = 20,$$

$$w_1 = \frac{n_1}{n} = \frac{2}{20} = 0{,}1; \qquad w_2 = \frac{n_2}{n} = \frac{3}{20} = 0{,}15;$$

$$w_3 = \frac{n_3}{n} = \frac{5}{20} = 0{,}25; \qquad w_4 = \frac{n_4}{n} = \frac{1}{20} = 0{,}05;$$

$$w_5 = \frac{n_5}{n} = \frac{4}{20} = 0{,}2; \qquad w_6 = \frac{n_6}{n} = \frac{2}{20} = 0{,}1;$$

$$w_7 = \frac{n_7}{n} = \frac{3}{20} = 0{,}15.$$

Найдём длину частных интервалов, эти частные интервалы и плотность относительных частот. Частные интервалы определим из условия, что заданные в дискретном статистическом распределении варианты должны быть серединами частных интервалов.

Итак, длина частных интервалов

$$h = x_{i+1} - x_i \equiv 2,$$

искомое интервальное статистическое распределение эмпирических вероятностей имеет вид

$(x_i; x_{i+1}]$	$(1;3]$	$(3;5]$	$(5;7]$	$(7;9]$	$(9;11]$	$(11;13]$	$(13;15]$
w_i	0,1	0,15	0,25	0,05	0,2	0,1	0,15

,

а плотности эмпирических вероятностей такие:

$$w_1^h = \frac{w_1}{h} = \frac{0{,}1}{2} = 0{,}05; \quad w_2^h = \frac{w_2}{h} = \frac{0{,}15}{2} = 0{,}075;$$

$$w_3^h = \frac{w_3}{h} = \frac{0{,}25}{2} = 0{,}125; \quad w_4^h = \frac{w_4}{h} = \frac{0{,}05}{2} = 0{,}025;$$

$$w_5^h = \frac{w_5}{h} = \frac{0{,}2}{2} = 0{,}1; \quad w_6^h = \frac{w_6}{h} = \frac{0{,}1}{2} = 0{,}05;$$

$$w_7^h = \frac{w_7}{h} = \frac{0{,}15}{2} = 0{,}075.$$

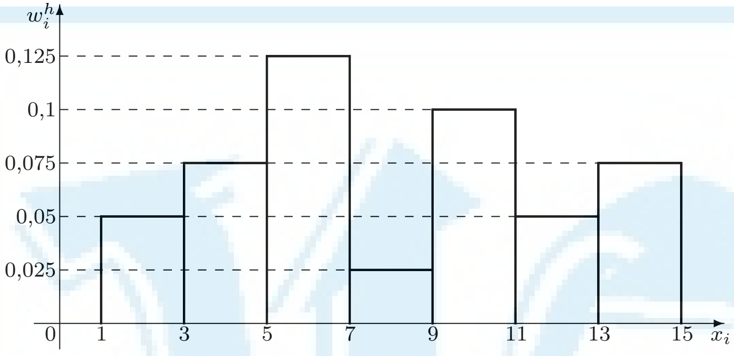


Рис. 38

Теперь, применив теоретические выкладки предыдущего примера, можно легко построить искомую гистограмму относительных частот (рис. 38). □

Пример 8. Дана выборка: 0,1; 0,4; 0,23; 0,12; 0,35; 0,46; 0,11; 0,04; 0,51; 0,27; 0,31; 0,34; 0,09; 0,18; 0,49; 0,33; 0,3; 0,22; 0,14; 0,5; 0,41; 0,25; 0,48; 0,32; 0,29; 0,31; 0,31; 0,46; 0,44; 0,38; 0,39; 0,13; 0,47; 0,4; 0,53; 0,37; 0,16; 0,44; 0,39; 0,27; 0,25; 0,46; 0,2; 0,11; 0,32; 0,41; 0,48; 0,22; 0,35; 0,52.

По данным выборки построить интервальное статистическое распределение, полигон и гистограмму частот.

Решение. Найдём объём выборки: $n = 50$. Поскольку наибольшее значение $x_{\max} = 0,53$, а наименьшее — $x_{\min} = 0,04$, оптимальная величина частного интервала

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}2 \cdot \lg n} = \frac{0{,}53 - 0{,}04}{1 + 3{,}2 \cdot \lg 50} \approx \frac{0{,}49}{1 + 3{,}2 \cdot 1{,}69897} \approx 0{,}076.$$

Частные интервалы начнём находить, начиная со срединного интервала, согласно условию, что он должен размещаться на одинаковом расстоянии от концов вариационного ряда (т. е. от 0,04 и 0,53). Итак, середина срединного частного интервала

$$x_{\text{сер}}^{\text{сер}} = \frac{x_{\max} + x_{\min}}{2} = \frac{0{,}53 + 0{,}04}{2} = 0{,}285.$$

Поскольку оптимальная величина частного интервала $h = 0{,}076$, то срединный интервал

$$
\begin{aligned}
(x_4; x_5] &= \left(x^{\text{сер}}_{\text{сер}} - \frac{h}{2}; x^{\text{сер}}_{\text{сер}} + \frac{h}{2}\right] = \\
&= \left(0{,}285 - \frac{0{,}076}{2}; 0{,}285 + \frac{0{,}076}{2}\right] = \\
&= (0{,}247; 0{,}323].
\end{aligned}
$$

Теперь легко найти остальные частные интервалы (сначала будем двигаться в направлении от срединного частного интервала к первому, а потом — от срединного к последнему):

$$
\begin{aligned}
(x_3; x_4] &= (0{,}247 - 0{,}076; 0{,}247] = (0{,}171; 0{,}247],\\
(x_2; x_3] &= (0{,}171 - 0{,}076; 0{,}171] = (0{,}095; 0{,}171],\\
(x_1; x_2] &= (0{,}095 - 0{,}076; 0{,}095] = (0{,}019; 0{,}095],\\
(x_5; x_6] &= (0{,}323; 0{,}323 + 0{,}076] = (0{,}323; 0{,}399],\\
(x_6; x_7] &= (0{,}399; 0{,}399 + 0{,}076] = (0{,}399; 0{,}475],\\
(x_7; x_8] &= (0{,}475; 0{,}475 + 0{,}076] = (0{,}475; 0{,}551].
\end{aligned}
$$

Определив частоты частных интервалов, получим интервальное статистическое распределение выборки:

Номер интервала i	Частный интервал $(x_i, x_{i+1}]$	Сумма частот вариант интервала n_i
1	(0,019; 0,095]	2
2	(0,095; 0,171]	7
3	(0,171; 0,247]	5
4	(0,247; 0,323]	11
5	(0,323; 0,399]	8
6	(0,399; 0,475]	10
7	(0,475; 0,551]	7

Контроль: $n = 2 + 7 + 5 + 11 + 8 + 10 + 7 = 50$.

Чтобы построить полигон частот, найдём значения середин частных интервалов:

$$x_1^{\text{сер}} = \frac{x_1 + x_2}{2} = \frac{0{,}019 + 0{,}095}{2} = 0{,}057;$$
$$x_2^{\text{сер}} = \frac{x_2 + x_3}{2} = \frac{0{,}095 + 0{,}171}{2} = 0{,}133;$$
$$x_3^{\text{сер}} = \frac{x_3 + x_4}{2} = \frac{0{,}171 + 0{,}247}{2} = 0{,}209;$$
$$x_4^{\text{сер}} = \frac{x_4 + x_5}{2} = \frac{0{,}247 + 0{,}323}{2} = 0{,}285;$$
$$x_5^{\text{сер}} = \frac{x_5 + x_6}{2} = \frac{0{,}323 + 0{,}399}{2} = 0{,}361;$$
$$x_6^{\text{сер}} = \frac{x_6 + x_7}{2} = \frac{0{,}399 + 0{,}475}{2} = 0{,}437;$$
$$x_7^{\text{сер}} = \frac{x_7 + x_8}{2} = \frac{0{,}475 + 0{,}551}{2} = 0{,}513.$$

Гистограмму частот построим, определив плотность частоты:

$$n_1^h = \frac{n_1}{h} = \frac{2}{0{,}076} \approx 26; \qquad n_2^h = \frac{n_2}{h} = \frac{7}{0{,}076} \approx 92;$$
$$n_3^h = \frac{n_3}{h} = \frac{5}{0{,}076} \approx 66; \qquad n_4^h = \frac{n_4}{h} = \frac{11}{0{,}076} \approx 145;$$
$$n_5^h = \frac{n_5}{h} = \frac{8}{0{,}076} \approx 105; \qquad n_6^h = \frac{n_6}{h} = \frac{10}{0{,}076} \approx 132;$$
$$n_7^h = \frac{n_7}{h} = \frac{7}{0{,}076} \approx 92.$$

Отложив на осях координат все необходимые точки, легко построить полигон (рис. 39) и гистограмму частот (рис. 40). □

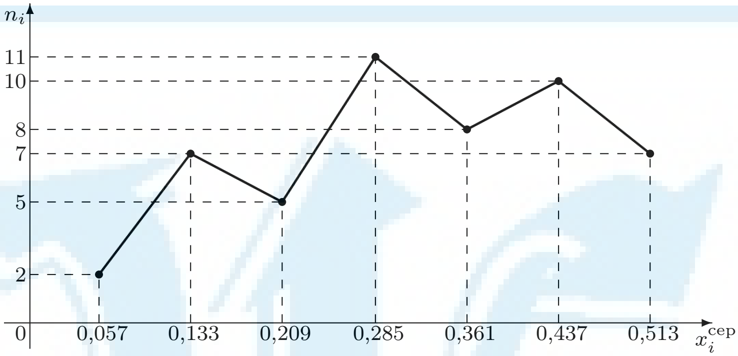


Рис. 39

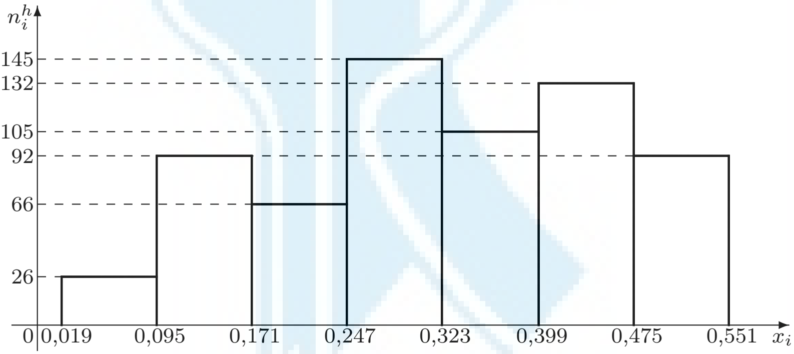


Рис. 40

12.2. Числовые характеристики выборки

Среднее арифметическое значение выборки называется *выборочным средним* $\overline{x}_{\text{в}}$ и вычисляется по формуле

$$\overline{x}_{\text{в}} = \frac{\sum\limits_{i=1}^{k} x_i n_i}{n} \quad \text{или} \quad \overline{x}_{\text{в}} = \sum_{i=1}^{k} x_i w_i,$$

где x_i — значение i-й варианты; n_i — частота i-й варианты; n — объём выборки; k — количество вариант в выборке; w_i — относительная частота i-й варианты.

Средний квадрат отклонения значений элементов выборки от выборочного среднего называется *выборочной дисперсией* $D_\text{в}$ и вычисляется по формулам

$$\mathbf{D}_\text{в} = \frac{\sum\limits_{i=1}^{k}(x_i - \overline{x}_\text{в})^2 n_i}{n} \quad \text{или} \quad \mathbf{D}_\text{в} = \sum_{i=1}^{k}(x_i - \overline{x}_\text{в})^2 w_i.$$

После преобразований формулы для нахождения выборочной дисперсии несколько упрощаются:

$$\mathbf{D}_\text{в} = \frac{\sum\limits_{i=1}^{k} x_i^2}{n} - \overline{x}_\text{в}^2 \quad \text{или} \quad \mathbf{D}_\text{в} = \sum_{i=1}^{k} x_i^2 w_i - \overline{x}_\text{в}^2,$$

т. е. выборочная дисперсия равняется разности среднего квадрата элементов выборки и квадрата выборочного среднего.

Квадратный корень из выборочной дисперсии $\sigma = \sqrt{\mathbf{D}_\text{в}}$ называется *средним квадратическим отклонением выборки*.

Пример 1. Задано статистическое распределение выборки:

x_i	1	3	4	7	10	12	15
n_i	5	2	12	7	4	3	2

Найти выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение выборки.

Решение. Вычислим объём выборки:

$$n = 5 + 2 + 12 + 7 + 4 + 3 + 2 = 35.$$

Найдём соответственно выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение выборки:

$$\overline{x}_{\text{в}} = \frac{\sum\limits_{i=1}^{7} x_i n_i}{n} = \frac{1 \cdot 5 + 3 \cdot 2 + 4 \cdot 12 + 7 \cdot 7 + 10 \cdot 4 + 12 \cdot 3 + 15 \cdot 2}{35} =$$

$$= \frac{214}{35} \approx 6{,}1;$$

$$\mathbf{D}_{\text{в}} = \frac{\sum\limits_{i=1}^{7} (x_i - \overline{x}_{\text{в}})^2 n_i}{n} = \frac{\sum\limits_{i=1}^{7} x_i^2 n_i}{n} - \overline{x}_{\text{в}}^2 =$$

$$= \frac{1^2 \cdot 5 + 3^2 \cdot 2 + 4^2 \cdot 12 + 7^2 \cdot 7 + 10^2 \cdot 4 + 12^2 \cdot 3 + 15^2 \cdot 2}{35} -$$

$$- \left(\frac{214}{35}\right)^2 = \frac{18604}{1225} \approx 15{,}2;$$

$$\sigma = \sqrt{\mathbf{D}_{\text{в}}} = \sqrt{\frac{18604}{1225}} \approx 3{,}9.$$

□

Если задано интервальное статистическое распределение выборки, то выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение выборки ищут с помощью такого статистического распределения: вариантами считаются середины частных интервалов, а частоты или относительные частоты остаются теми же.

Пример 2. Задано интервальное статистическое распределение выборки:

$(x_i; x_{i+1}]$	$(0; 2]$	$(2; 4]$	$(4; 6]$	$(6; 8]$	$(8; 10]$	$(10; 12]$	$(12; 14]$
w_i	0,1	0,2	0,3	0,1	0,1	0,1	0,1

Найти выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение выборки.

Решение. Сначала превратим данное интервальное статистическое распределение выборки в точечное, определив середины частных интервалов:

$$x_1 = \frac{x_1 + x_2}{2} = \frac{0+2}{2} = 1; \qquad x_2 = \frac{x_2 + x_3}{2} = \frac{2+4}{2} = 3;$$
$$x_3 = \frac{x_3 + x_4}{2} = \frac{4+6}{2} = 5; \qquad x_4 = \frac{x_4 + x_5}{2} = \frac{6+8}{2} = 7;$$
$$x_5 = \frac{x_5 + x_6}{2} = \frac{8+10}{2} = 9; \qquad x_6 = \frac{x_6 + x_7}{2} = \frac{10+12}{2} = 11;$$
$$x_7 = \frac{x_7 + x_8}{2} = \frac{12+14}{2} = 13.$$

Итак, мы получили статистическое распределение выборки:

x_i	1	3	5	7	9	11	13
w_i	0,1	0,2	0,3	0,1	0,1	0,1	0,1

.

Найдём соответственно выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение выборки:

$$\overline{x}_{\text{в}} = \sum_{i=1}^{7} x_i w_i = 1 \cdot 0{,}1 + 3 \cdot 0{,}2 + 5 \cdot 0{,}3 + 7 \cdot 0{,}1 + 9 \cdot 0{,}1 + {}$$
$$+\ 11 \cdot 0{,}1 + 13 \cdot 0{,}1 = 6{,}2;$$
$$\mathbf{D}_{\text{в}} = \sum_{i=1}^{7} (x_i - \overline{x}_{\text{в}})^2 w_i = \sum_{i=1}^{7} x_i^2 w_i - \overline{x}_{\text{в}}^2 = 1^2 \cdot 0{,}1 + 3^2 \cdot 0{,}2 + {}$$
$$+\ 5^2 \cdot 0{,}3 + 7^2 \cdot 0{,}1 + 9^2 \cdot 0{,}1 + 11^2 \cdot 0{,}1 + 13^2 \cdot 0{,}1 - 6{,}2^2 = {}$$
$$= 12{,}96;$$
$$\sigma = \sqrt{\mathbf{D}_{\text{в}}} = \sqrt{12{,}96} = 3{,}6.$$

□

Медианой (**Me**) называется значение среднего элемента вариационного ряда. Если объём выборки $n = 2m+1$ непарный, то медианой будет значение элемента вариационного ряда с номером $m+1$:

$$\mathbf{Me} = x_{m+1}.$$

Если объём выборки $n = 2m$ парный, то медианой будет среднее значение элементов вариационного ряда с номерами m и $m+1$:

$$\mathbf{Me} = \frac{x_m + x_{m+1}}{2}.$$

Если задано интервальное статистическое распределение выборки, то сначала находят *медианный частный интервал*, т. е. первый частный интервал, для которого сумма частот всех предыдущих частных интервалов включительно с данным превышает половину объёма выборки. В этом случае медиану находят по формуле

$$\mathbf{Me} = x_{\mathbf{Me}}^{\min} + h_{\mathbf{Me}} \frac{\frac{1}{2}n - \sum\limits_{i=1}^{\mathbf{Me}-1} n_i}{n_{\mathbf{Me}}},$$

де $x_{\mathbf{Me}}^{\min}$ — начало медианного частного интервала; $h_{\mathbf{Me}}$ — длина медианного частного интервала; n — объём выборки; $n_{\mathbf{Me}}$ — частота медианного частного интервала; $\mathbf{Me} - 1$ — номер предыдущего медианному частного интервала.

Медиана имеет такое свойство: сумма абсолютных величин отклонений элементов выборки от медианы меньше, чем от любой другой величины:

$$\sum_{i=1}^{n} |x_i - \mathbf{Me}| < \sum_{i=1}^{n} |x_i - a|, \qquad a \neq \mathbf{Me}.$$

Пример 3. На одном из отрезков железной дороги планируется организовать остановку пассажирского поезда. Распределение населенных пунктов с численностью их населения приведено в таблице.

На каком километре железной дороги расположен населенный пункт, км	10	12	15	25	28	30	33
Численность населения, тыс. чел.	5	2	3	10	1	4	6

На каком километре железной дороги нужно расположить эту остановку, чтобы суммарное расстояние, которое будут покрывать потенциальные пассажиры до этой остановки, было наименьшим.

Решение. Поскольку согласно свойству медианы сумма абсолютных величин отклонений элементов выборки от медианы меньше, чем от любой другой величины, для решения примера нужно найти медиану.

Сначала определим объём выборки:

$$n = 5 + 2 + 3 + 10 + 1 + 4 + 6 = 31.$$

Итак, серединой (средним членом) вариационного ряда будет элемент под номером 16: $\mathbf{Me} = x_{16}$. Поскольку вариационный ряд можно записать в виде

$$\underbrace{10, \ldots, 10}_{5\text{ раз}}, 12, 12, 15, 15, 15, \underbrace{25, \ldots, 25}_{10\text{ раз}}, 28, \underbrace{30, \ldots, 30}_{4\text{ раза}}, \underbrace{33, \ldots, 33}_{6\text{ раз}}$$

легко увидеть, что $x_{16} = 25$, т. е. остановку следует расположить на 25-м километре железной дороги. □

Модой ($\mathbf{Mo}$) называется варианта с наибольшей частотой. В случае интервального статистического распределения выборки с одинаковыми по длине частными интервалами *модальный частный интервал* определяется по наибольшей частоте, а при разных по длине частных интервалах — по наибольшей плотности $\max\limits_i \dfrac{n_i}{h_i}$, где n_i, h_i — соответственно частота и длина i-го частного интервала. В этом случае моду находят по формуле

$$\mathbf{Mo} = x_{\mathbf{Mo}}^{\min} + h_{\mathbf{Mo}} \frac{n_{\mathbf{Mo}} - n_{\mathbf{Mo}-1}}{2n_{\mathbf{Mo}} - n_{\mathbf{Mo}-1} - n_{\mathbf{Mo}+1}},$$

где $x_{\mathbf{Mo}}^{\min}$ — начало модального частного интервала; $h_{\mathbf{Mo}}$ — длина модального частного интервала; $n_{\mathbf{Mo}}$ — частота модального частного интервала; $n_{\mathbf{Mo}-1}$ — частота предыдущего к модальному частного интервала; $n_{\mathbf{Mo}+1}$ — частота следующего за модальным частного интервала.

Вариационным размахом называется разность между наибольшим и наименьшим значениями выборки:

$$\mathbf{R} = x_{\max} - x_{\min}.$$

Пример 4. Задано интервальное статистическое распределение выборки:

$(x_i, x_{i+1}]$	$(10; 15]$	$(15; 20]$	$(20; 25]$	$(25; 30]$	$(30; 35]$	$(35; 40]$	$(40; 45]$
n_i	7	4	5	1	12	3	18

Найти медиану, моду и вариационный размах.

Решение. Определим объём выборки:

$$n = 7 + 4 + 5 + 1 + 12 + 3 + 18 = 50.$$

Медианным частным интервалом будет пятый интервал, поскольку это первый интервал, для которого сумма частот всех предыдущих частных интервалов включительно с данным превышает половину объёма выборки:

$$7 + 4 + 5 + 1 + 12 = 29 > \frac{n}{2} = \frac{50}{2} = 25.$$

Модальным частным интервалом будет последний интервал, поскольку он имеет наибольшую частоту.

Найдем начало медианного частного интервала $x_{\mathbf{Me}}^{\min}$, длину медианного частного интервала $h_{\mathbf{Me}}$, номер предыдущего к медианному частного интервала $\mathbf{Me} - 1$, частоту медианного частного интервала $n_{\mathbf{Me}}$, начало модального частного интервала $x_{\mathbf{Mo}}^{\min}$, длину модального частного интервала $h_{\mathbf{Mo}}$, частоту модального частного интервала $n_{\mathbf{Mo}}$, частоту предыдущего к модальному частного интервала $n_{\mathbf{Mo}-1}$, частоту следующего за модальным частного интервала $n_{\mathbf{Mo}+1}$, наибольшее и наименьшее значения выборки:

$$x_{\mathbf{Me}}^{\min} = 30; \qquad h_{\mathbf{Me}} = 35 - 30 = 5; \qquad \mathbf{Me} - 1 = 4;$$
$$n_{\mathbf{Me}} = 12; \qquad x_{\mathbf{Mo}}^{\min} = 40; \qquad h_{\mathbf{Mo}} = 45 - 40 = 5;$$
$$n_{\mathbf{Mo}} = 18; \qquad n_{\mathbf{Mo}-1} = 3; \qquad n_{\mathbf{Mo}+1} = 0;$$
$$x_{\max} = 45; \qquad x_{\min} = 10.$$

Искомые медиана, мода и вариационный размах соответственно будут такие:

$$\mathbf{Me} = x_{\mathbf{Me}}^{\min} + h_{\mathbf{Me}} \frac{\frac{1}{2}n - \sum\limits_{i=1}^{\mathbf{Me}-1} n_i}{n_{\mathbf{Me}}} = 30 + 5 \cdot \frac{\frac{1}{2} \cdot 50 - \sum\limits_{i=1}^{4} n_i}{12} =$$

$$= \frac{100}{3} \approx 33{,}3;$$

$$\mathbf{Mo} = x_{\mathbf{Mo}}^{\min} + h_{\mathbf{Mo}} \frac{n_{\mathbf{Mo}} - n_{\mathbf{Mo}-1}}{2n_{\mathbf{Mo}} - n_{\mathbf{Mo}-1} - n_{\mathbf{Mo}+1}} = 40 + 5 \cdot \frac{18-3}{2 \cdot 18 - 3 - 0} =$$

$$= \frac{465}{11} \approx 42{,}34;$$

$$\mathbf{R} = x_{\max} - x_{\min} = 45 - 10 = 35.$$

□

12.3. Метод произведений вычисления выборочного среднего и выборочной дисперсии

Пусть выборка задана в виде распределения равноотстоящих вариант и соответствующих им частот. В этом случае выборочное среднее и выборочную дисперсию удобно находить *методом произведений* по формулам

$$\overline{x}_{\text{в}} = \mathbf{M}_1^{\star} h + C, \qquad \mathbf{D}_{\text{в}} = \left[\mathbf{M}_2^{\star} - (\mathbf{M}_1^{\star})^2\right] h^2,$$

где $\mathbf{M}_1^{\star} = = \dfrac{\sum\limits_i n_i u_i}{n}$ — условный момент первого порядка; $u_i = \dfrac{x_i - C}{h}$ — условная варианта; h — длина шага (т. е. расстояние между соседними вариантами); C — ложный нуль (значение варианты, расположенной приблизительно посредине вариационного ряда); $\mathbf{M}_2^{\star} = \dfrac{\sum\limits_i n_i u_i^2}{n}$ — условный момент второго порядка.

Пример 1. Найти методом произведений выборочное среднее и выборочную дисперсию для такого распределения выборки объёма

$n = 100$:

x_i	24	28	32	36	40	44	48
n_i	6	13	20	45	11	3	2

Решение. Составим расчётную табл. 5, а именно:

1) запишем значения вариант в первый столбец;

2) запишем значения частот во второй столбец, сумму частот (объём выборки) — в нижней его клеточке;

3) в качестве ложного нуля C выберем значение четвертой варианты $x_4 = 36$, которая имеет наибольшую частоту (в качестве C можно взять значение любой варианты, расположенной приблизительно посредине столбца), в клеточке третьего столбца, которая отвечает ложному нулю, запишем 0, над ним последовательно записываем -1, -2, -3, а под ним — 1, 2, 3;

4) произведения частот n_i на значения соответствующих условных вариант u_i запишем в четвертый столбец, отдельно найдём сумму всех чисел четвертого столбца, которую поместим в нижнюю клеточку четвертого столбца;

5) произведения частот n_i на значения квадратов соответствующих условных вариант u_i^2, т. е. $n_i u_i^2$, запишем в пятый столбец (удобнее перемножить числа каждой строки третьего и четвертого столбцов: $u_i \cdot n_i u_i = n_i u_i^2$), отдельно найдём сумму всех чисел пятого столбца, которую поместим в нижнюю клеточку пятого столбца;

6) произведения частот n_i на значения квадратов соответствующих условных вариант u_i, увеличенных на единицу, т. е. $n_i(u_i + 1)^2$, запишем в шестой столбец, отдельно найдём сумму всех чисел шестого столбца, которую поместим в нижнюю клеточку шестого столбца.

Итак, мы получили расчётную табл. 5.

Для контроля вычислений воспользуемся тождеством

$$\sum_i n_i(u_i + 1)^2 \equiv \sum_i n_i u_i^2 + 2\sum_i n_i u_i + n.$$

Таблица 5

1	2	3	4	5	6
x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i(u_i+1)^2$
24	6	-3	-18	54	24
28	13	-2	-26	52	13
32	20	-1	-20	20	0
36	45	0	0	0	45
40	11	1	11	11	44
44	3	2	6	12	27
48	2	3	6	18	32
	$n=100$		$\sum\limits_i n_i u_i = -41$	$\sum\limits_i n_i u_i^2 = 167$	$\sum\limits_i n_i(u_i+1)^2 = 185$

Контроль:

$$\sum_i n_i(u_i+1)^2 = 185;$$

$$\sum_i n_i u_i^2 + 2\sum_i n_i u_i + n = 167 + 2\cdot(-41) + 100 = 185.$$

Совпадение контрольных сумм свидетельствует о правильности вычислений.

Определим условные моменты первого и второго порядка:

$$\mathbf{M}_1^\star = \frac{\sum_i n_i u_i}{n} = \frac{-41}{100} = -0{,}41; \qquad \mathbf{M}_2^\star = \frac{\sum_i n_i u_i^2}{n} = \frac{167}{100} = 1{,}67.$$

Найдём шаг (разность между любыми двумя соседними вариантами):

$$h = 28 - 24 = 4.$$

Определим искомые выборочные среднее и дисперсию, учитывая, что ложный нуль (значение варианты, которая имеет наибольшую частоту) $C = 36$:

$$\overline{x}_\text{в} = \mathbf{M}_1^\star h + C = -0{,}41\cdot 4 + 36 = 34{,}36;$$

$$\mathbf{D}_\text{в} = \left[\mathbf{M}_2^\star - (\mathbf{M}_1^\star)^2\right] h^2 = \left[1{,}67 - (-0{,}41)^2\right]\cdot 4^2 = 24{,}0304. \qquad \square$$

Если варианты выборки не являются равноотстоящими, то интервал, в котором размещены все варианты выборки, делят на несколько одинаковых по длине частных интервалов (каждый из них должен содержать не менее 8–10 вариант). Затем находят середины частных интервалов, которые и образуют последовательность равноотстоящих вариант. В качестве частоты каждой середины частного интервала берут сумму частот вариант, которые попали в соответствующий частный интервал.

При вычислении выборочной дисперсии с целью уменьшения погрешности, обусловленной группированием (особенно при малом количестве интервалов), используют *поправку Шеппарда*, а именно вычисляют дисперсию по формуле

$$\mathbf{D}'_{\text{в}} = \mathbf{D}_{\text{в}} - \frac{1}{12}h^2.$$

Пример 2. Найти методом произведений выборочное среднее и выборочную дисперсию для такого распределения выборки объёма $n = 100$:

x_i	2	8	9	13	15	18	20	21	24	27
n_i	8	1	15	13	9	3	4	10	22	15

Решение. Разобьём интервал $[2; 27]$ на пять частных интервалов длиной $h = 5$:

$$[2; 7], \quad (7; 12], \quad (12; 17], \quad (17; 22], \quad (22; 27].$$

Новыми вариантами будут середины этих частных интервалов:

$$y_1 = \frac{2+7}{2} = 4{,}5; \qquad y_2 = \frac{7+12}{2} = 9{,}5; \qquad y_3 = \frac{12+17}{2} = 14{,}5;$$

$$y_4 = \frac{17+22}{2} = 19{,}5; \qquad y_5 = \frac{22+27}{2} = 24{,}5.$$

В качестве частот m_i вариант y_i возьмём суммы частот вариант, которые попали в соответствующий i-й интервал:

$$m_1 = n_1 = 8; \quad m_2 = n_2 + n_3 = 1 + 15 = 16;$$

$$m_3 = n_4 + n_5 = 13 + 9 = 22; \quad m_4 = n_6 + n_7 + n_8 = 3 + 4 + 10 = 17;$$

$$m_5 = n_9 + n_{10} = 22 + 15 = 37.$$

Итак, мы получили такое статистическое распределение равноотстоящих вариант:

y_i	4,5	9,5	14,5	19,5	24,5
m_i	8	16	22	17	37

Как и в примере 1, составим расчётную табл. 6.

Таблица 6

1	2	3	4	5	6
y_i	m_i	u_i	$m_i u_i$	$m_i u_i^2$	$m_i(u_i+1)^2$
4,5	8	−2	−16	32	8
9,5	16	−1	−16	16	0
14,5	22	0	0	0	22
19,5	17	1	17	17	68
24,5	37	2	74	148	333
	$n = 100$		$\sum\limits_i m_i u_i = 59$	$\sum\limits_i m_i u_i^2 = 213$	$\sum\limits_i m_i(u_i+1)^2 = 431$

Контроль:

$$\sum_i m_i(u_i+1)^2 = 431;$$

$$\sum_i m_i u_i^2 + 2\sum_i m_i u_i + n = 213 + 2 \cdot 59 + 100 = 431.$$

Вычислим условные моменты первого и второго порядка:

$$\mathbf{M}_1^\star = \frac{\sum_i m_i u_i}{n} = \frac{59}{100} = 0{,}59; \qquad \mathbf{M}_2^\star = \frac{\sum_i m_i u_i^2}{n} = \frac{213}{100} = 2{,}13.$$

Методом произведений найдём выборочные среднее и дисперсию, учитывая, что ложный нуль $C = 14{,}5$:

$$\overline{x}_\text{в} = \mathbf{M}_1^\star h + C = 0{,}59 \cdot 5 + 14{,}5 = 17{,}45;$$

$$\mathbf{D}_\text{в} = \left[\mathbf{M}_2^\star - (\mathbf{M}_1^\star)^2\right] h^2 = \left[2{,}13 - 0{,}59^2\right] \cdot 5^2 = 44{,}5475.$$

Поскольку количество частных интервалов мало (пять), воспользуемся поправкой Шеппарда:

$$\mathbf{D}'_{\text{в}} = \mathbf{D}_{\text{в}} - \frac{1}{12}h^2 = 44{,}5475 - \frac{5^2}{12} \approx 42{,}46. \qquad \square$$

12.4. Метод сумм вычисления выборочного среднего и выборочной дисперсии

Пусть выборка задана в виде распределения равноотстоящих вариант и соответствующих им частот. В этом случае выборочное среднее и выборочную дисперсию можно вычислить по формулам (см. подразд. 12.3)

$$\overline{x}_{\text{в}} = \mathbf{M}_1^{\star}h + C, \qquad \mathbf{D}_{\text{в}} = \left[\mathbf{M}_2^{\star} - (\mathbf{M}_1^{\star})^2\right] h^2.$$

При использовании *метода сумм* условные моменты первого и второго порядка определяют по формулам

$$\mathbf{M}_1^{\star} = \frac{d_1}{n}, \qquad \mathbf{M}_2^{\star} = \frac{s_1 + 2s_2}{n},$$

где $d_1 = a_1 - b_1$, $s_1 = a_1 + b_1$, $s_2 = a_2 + b_2$.

Итак, нужно найти значения a_1, a_2, b_1, b_2. Как практически рассчитать эти числа, показано в примере 1.

Пример 1. Найти методом сумм выборочное среднее и выборочную дисперсию по заданному статистическому распределению выборки объёма $n = 100$:

x_i	52	58	64	70	76	82	88	94	100	106
n_i	4	8	10	13	18	21	14	9	2	1

Решение. Составим расчётную табл. 7, а именно:

1) запишем значения вариант в первый столбец;

2) запишем значения частот во второй столбец, сумму частот (объём выборки) — в нижней его клеточке;

3) в качестве ложного нуля C выберем значение шестой варианты $x_6 = 82$, которая имеет наибольшую частоту (в качестве C можно взять значение любой варианты, расположенной приблизительно посредине столбца), в клеточках строки, которая отвечает ложному нулю, запишем нули, в четвертом столбце над и под нулём — еще по одному нулю;

4) в пустых клеточках третьего столбца, которые расположены над нулём (кроме верхней), сверху вниз запишем последовательно накопленные частоты:

$$4, \quad 4+8=12, \quad 12+10=22, \quad 22+13=35, \quad 35+18=53.$$

Сложив все накопленные частоты, получим число $b_1 = 126$, которое запишем в верхнюю клеточку третьего столбца. В пустых клеточках третьего столбца, размещенных под нулём (кроме нижней), снизу вверх запишем последовательно накопленные частоты:

$$1, \quad 1+2=3, \quad 3+9=12, \quad 12+14=26.$$

Сложив все накопленные частоты, получим число $a_1 = 42$, которое запишем в нижнюю клеточку третьего столбца;

5) аналогично заполним четвёртый столбец, причём накапливаются частоты третьего столбца. Сумму накопленных частот, расположенных над нулём, обозначим b_2 и запишем в верхнюю клеточку четвёртого столбца. Сумму накопленных частот, расположенных под нулём, обозначим a_2 и запишем в нижнюю клеточку четвёртого столбца.

Итак, мы получили расчётную табл. 7.

Найдём d_1, s_1, s_2:

$$d_1 = a_1 - b_1 = 42 - 126 = -84;$$
$$s_1 = a_1 + b_1 = 42 + 126 = 168;$$
$$s_2 = a_2 + b_2 = 21 + 131 = 152.$$

Вычислим условные моменты первого и второго порядка:

$$\mathbf{M}_1^\star = \frac{d_1}{n} = \frac{-84}{100} = -0{,}84;$$
$$\mathbf{M}_2^\star = \frac{s_1 + 2s_2}{n} = \frac{168 + 2 \cdot 152}{100} = 4{,}72.$$

Методом произведений найдём выборочные среднее и дисперсию, учитывая, что ложный ноль $C = 82$, а шаг (расстояние между двумя соседними вариантами) $h = 6$; получим

$$\overline{x}_{\text{в}} = \mathbf{M}_1^\star h + C = -0{,}84 \cdot 6 + 82 = 76{,}96$$

и

$$\mathbf{D}_{\text{в}} = \left[\mathbf{M}_2^\star - (\mathbf{M}_1^\star)^2\right] h^2 = \left[4{,}72 - (-0{,}84)^2\right] \cdot 6^2 = 144{,}5184. \qquad \square$$

Таблица 7

1	2	3	4
x_i	n_i	$b_1 = 126$	$b_2 = 131$
52	4	4	4
58	8	12	16
64	10	22	38
70	13	35	73
76	18	53	0
82	21	0	0
88	14	26	0
94	9	12	16
100	2	3	4
106	1	1	1
	$n = 100$	$a_1 = 42$	$a_2 = 21$

Задачи к разделу 12

Задача 1. При исследовании количественного признака X из генеральной совокупности была получена выборка

$$5, 7, 4, 6, 5, 5, 5, 7, 5, 6, 6, 5, 5, 6, 4, 6, 5, 6, 4, 5.$$

Найти объём выборки, построить вариационный ряд выборки и её статистическое распределение.

Ответ. $n = 20$. Вариационный ряд выборки: 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 7, 7. Статистическое распределение:

x_i	4	5	6	7
n_i	3	9	6	2

.

Задача 2. Выборка задана распределением частот:

x_i	-3	-1	0	4	8	13	17
n_i	1	3	4	7	5	8	12

Найти распределение относительных частот.

Ответ.

x_i	-3	-1	0	4	8	13	17
w_i	0,025	0,075	0,1	0,175	0,125	0,2	0,3

Задача 3. Выборка задана интервальным распределением частот:

x_i	$(0; 2]$	$(2; 4]$	$(4; 6]$	$(6; 8]$	$(8; 10]$
n_i	5	14	16	11	4

Найти распределение относительных частот.

Ответ.

x_i	$(0; 2]$	$(2; 4]$	$(4; 6]$	$(6; 8]$	$(8; 10]$
w_i	0,1	0,28	0,32	0,22	0,08

Задача 4. Найти эмпирическую функцию распределения для заданного распределения выборки:

x_i	1	4	8	13	19	26	34
n_i	5	14	17	22	25	11	6

Ответ. $$F^\star(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 1; \\ 0{,}05 & \text{при} \quad 1 < x \leqslant 4; \\ 0{,}19 & \text{при} \quad 4 < x \leqslant 8; \\ 0{,}36 & \text{при} \quad 8 < x \leqslant 13; \\ 0{,}58 & \text{при} \quad 13 < x \leqslant 19; \\ 0{,}83 & \text{при} \quad 19 < x \leqslant 26; \\ 0{,}94 & \text{при} \quad 26 < x \leqslant 34; \\ 1 & \text{при} \quad x > 34. \end{cases}$$

Задача 5. Построить полигон частот для заданного распределения выборки:

x_i	3	5	8	14	16	20	22
n_i	32	41	18	56	23	12	47

Ответ. Искомый полигон частот изображён на рис. 41.

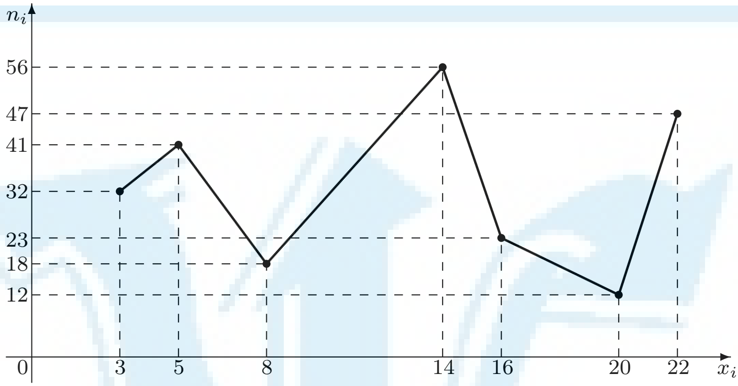


Рис. 41

Задача 6. Построить полигон относительных частот для заданного распределения выборки:

x_i	1	7	11	12	18	26	31
w_i	0,15	0,2	0,1	0,15	0,1	0,05	0,25

Указание. Строить так, будто вместо относительных частот указаны частоты.

Ответ. Искомый полигон эмпирических вероятностей изображён на рис. 42.

Задача 7. Выборка задана интервальным распределением частот:

$(x_i; x_{i+1}]$	$(5; 10]$	$(10; 15]$	$(15; 20]$	$(20; 25]$	$(25; 30]$	$(30; 35]$	$(35; 40]$
n_i	34	12	23	18	42	30	7

Построить гистограмму частот.

Ответ. Искомая гистограмма частот изображена на рис. 43.

Задача 8. Выборка задана распределением частот:

x_i	5	11	17	23	29	35	41
n_i	26	54	33	67	39	21	10

Построить гистограмму относительных частот.

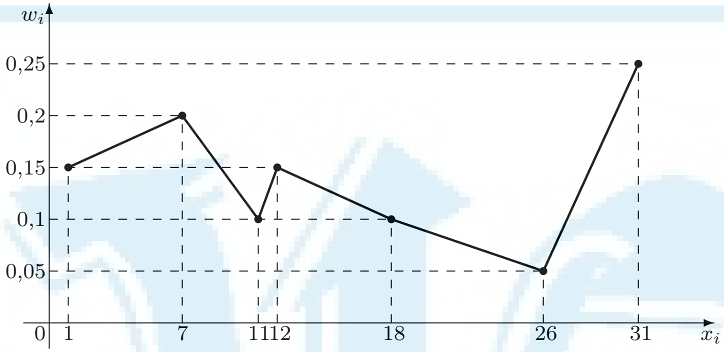


Рис. 42

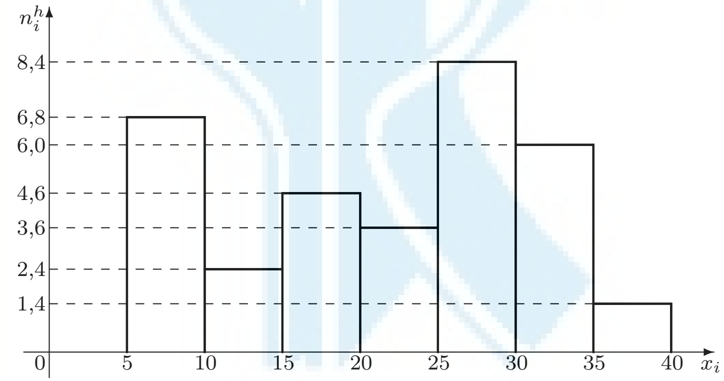


Рис. 43

Ответ. Искомая гистограмма эмпирических вероятностей изображена на рис. 44.

Задача 9. Дана выборка: 15; 18; 11; 12; 35; 46; 10; 41; 23; 27; 31; 34; 9; 8; 15; 13; 17; 22; 21; 44; 19; 14; 30; 25; 10; 14; 11; 18; 48; 20; 12; 32; 7; 13; 15; 41; 23; 19; 24; 28; 12; 9; 33; 17; 16; 8; 29; 15; 11; 40.

По данным выборки построить интервальное статистическое распределение, полигон и гистограмму частот.

Ответ. Интервальное распределение выборки такое:

Номер интервала i	Частный интервал $(x_i, x_{i+1}]$	Сумма частот вариант интервала n_i
1	(5,1; 11,5]	11
2	(11,5; 17,9]	13
3	(17,9; 24,3]	10
4	(24,3; 30,7]	5
5	(30,7; 37,1]	5
6	(37,1; 43,5]	3
7	(43,5; 49,9]	3

Искомые полигон и гистограмма частоты изображены соответственно на рис. 45 и 46.

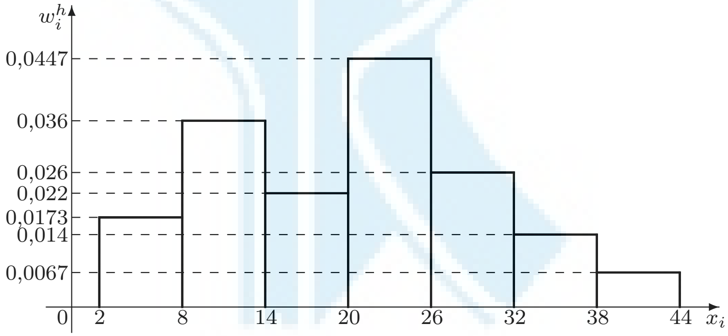


Рис. 44

Задача 10. Задано статистическое распределение выборки:

x_i	2	3	5	8	11	12	18
n_i	2	7	15	32	21	14	9

Найти выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение выборки.

Ответ. $\overline{x}_{\text{в}} = 9{,}17$; $\mathbf{D}_{\text{в}} = 15{,}5811$; $\sigma \approx 3{,}95$.

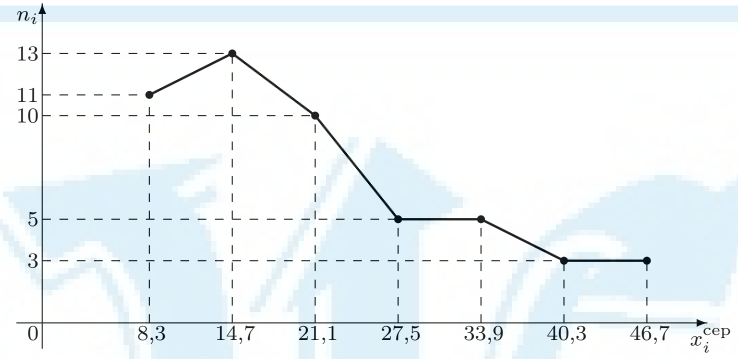


Рис. 45

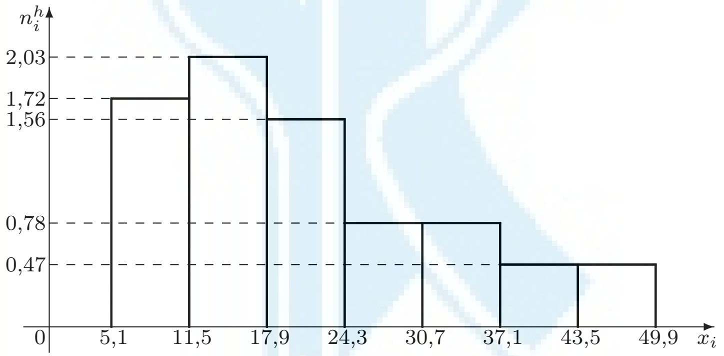


Рис. 46

Задача 11. Задано интервальное статистическое распределение выборки:

$(x_i; x_{i+1}]$	$(1;3]$	$(3;5]$	$(5;7]$	$(7;9]$	$(9;11]$	$(11;13]$	$(13;15]$
w_i	0,15	0,05	0,2	0,1	0,05	0,25	0,2

Найти выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение выборки.

Ответ. $\overline{x}_{\text{в}} = 8{,}8$; $\mathbf{D}_{\text{в}} = 17{,}76$; $\sigma \approx 4{,}214$.

Задача 12. На одном из отрезков автострады планируется обустроить остановку автобуса. Распределение населённых пунктов с численностью их населения приведено в таблице.

На каком километре автострады расположен населённый пункт, км	3	6	8	14	20	22	25
Численность населения, тыс. чел.	4	3	5	6	2	3	1

На каком километре автострады нужно расположить эту остановку, чтобы суммарное расстояние, которое будут покрывать потенциальные пассажиры до этой остановки, было наименьшим.

Ответ. На 11-м километре.

Задача 13. Задано интервальное статистическое распределение выборки:

$(x_i, x_{i+1}]$	$(4; 8]$	$(8; 12]$	$(12; 16]$	$(16; 20]$	$(20; 24]$	$(24; 28]$	$(28; 32]$
n_i	3	6	6	10	14	5	1

Найти медиану, моду и вариационный размах.

Ответ. $\mathbf{Me} = 19$; $\mathbf{Mo} = \dfrac{276}{13} \approx 21{,}23$; $\mathbf{R} = 28$.

Задача 14. Найти методом произведений выборочное среднее и выборочную дисперсию для такого распределения выборки объёма $n = 100$:

x_i	40	45	50	55	60	65	70
n_i	12	18	20	24	14	7	5

Указание. В качестве ложного нуля выбрать значение четвёртой варианты: $C = x_4 = 55$.

Ответ. $\overline{x}_{\text{в}} = 52{,}55$; $\mathbf{D}_{\text{в}} = 65{,}7475$.

Задача 15. Найти методом произведений выборочное среднее и выборочную дисперсию для такого распределения выборки объёма $n = 100$:

x_i	5	7	10	11	14	15	18	20	22	25
n_i	6	8	11	18	22	11	7	7	5	5

Указание. Разбить весь интервал на пять частных интервалов, в качестве ложного нуля выбрать значение середины третьего интервала. При определении выборочной дисперсии учесть поправку Шеппарда.

Ответ. $\overline{x}_{\text{в}} = 14{,}08$; $\mathbf{D}_{\text{в}} \approx 20{,}06$.

Задача 16. Найти методом сумм выборочное среднее и выборочную дисперсию для заданного статистического распределения выборки объёма $n = 100$:

x_i	42	49	56	63	70	77	84	91	98	105
n_i	2	5	8	19	26	14	10	8	5	3

Указание. В качестве ложного нуля выбрать значение пятой варианты: $C = = x_5 = 70$.

Ответ. $\overline{x}_{\text{в}} = 72{,}45$; $\mathbf{D}_{\text{в}} = 194{,}4075$.

Раздел 13. Статистические оценки параметров распределения

Ключевые слова

Величина
Вероятность
Выборочный
Генеральный
Дискретный
Дисперсия
Доверительный
Интервал
Интервальный
Исправленный
Квадратический
Логарифмический
Максимальный
Математический
Метод
Момент
Наибольший
Непрерывный
Несмещённый
Ожидание
Отклонение
Оценка
Поправка
Правдоподобие
Случайный
Смещённый
Средний
Статистический
Точечный
Уравнение
Функция

13.1. Точечные оценки

Статистической оценкой $\Theta^\star$ неизвестного параметра Θ теоретического распределения называют функцию $f(X_1, X_2, \ldots, X_n)$ от наблюдаемых случайных величин $X_1, X_2, \ldots, X_n$.

Точечной называют статистическую оценку, которая определяется одним единственным числом $\Theta^\star = f(x_1, x_2, \ldots, x_n)$, где $x_1, x_2, \ldots, x_n$ — результаты n наблюдений над количественным признаком случайной величины X (выборка).

Несмещённой называют точечную оценку $\Theta^\star$, математическое ожидание которой равняется оцениваемому параметру Θ при любом объёме выборки:

$$\mathbf{M}(\Theta^\star) = \Theta.$$

Смещённой называют точечную оценку $\Theta^\star$, математическое ожидание которой отличается от оцениваемого параметра Θ:

$$\mathbf{M}(\Theta^\star) \neq \Theta.$$

Несмещённой оценкой генерального среднего (математического ожидания) является выборочное среднее

$$\overline{x}_\text{в} = \frac{\sum\limits_{i=1}^{k} n_i x_i}{n},$$

где n_i — частота варианты x_i; x_i — варианта выборки; $n = \sum\limits_{i=1}^{k} n_i$ — объём выборки.

Замечание 1. Если варианты x_i выборки являются очень большими или очень маленькими (близкими к нулю) числами, то для упрощения расчётов целесообразно вычесть (в случае больших отрицательных чисел — прибавить) от каждой варианты одно и то же число C (в качестве C можно выбрать любое число, расположенное приблизительно посредине вариационного ряда), потом поделить (в случае близких к нулю чисел — умножить) на одно и то же число b (в качестве b можно выбрать наибольшее общее кратное), т. е. перейти к условным вариантам

$$u_i = \frac{x_i \pm C}{b} \qquad (u_i = x_i b).$$

Тогда

$$\overline{x}_\text{в} = b\frac{\sum\limits_{i=1}^{k} n_i u_i}{n} \mp C = b\overline{u}_\text{в} \mp C \qquad \left(\overline{x}_\text{в} = \frac{\sum\limits_{i=1}^{k} n_i u_i}{bn} = \frac{\overline{u}_\text{в}}{b}\right).$$

Смещённой оценкой генеральной дисперсии является выборочная дисперсия

$$\mathbf{D}_\text{в} = \frac{\sum\limits_{i=1}^{k} n_i (x_i - \overline{x}_\text{в})^2}{n}.$$

Эта оценка смещённая, поскольку

$$\mathbf{M}(\mathbf{D}_{\text{в}}) = \frac{n-1}{n}\mathbf{D}_{\text{г}}.$$

Для вычисления выборочной дисперсии можно воспользоваться более удобной формулой

$$\mathbf{D}_{\text{в}} = \overline{x^2}_{\text{в}} - \overline{x}_{\text{в}}{}^2 = \frac{\sum\limits_{i=1}^{k} n_i x_i^2}{n} - \frac{\left(\sum\limits_{i=1}^{k} n_i x_i\right)^2}{n^2}.$$

Замечание 2. Если варианты x_i выборки являются очень большими или очень маленькими (близкими к нулю) числами, то для упрощения расчётов целесообразно вычесть (в случае больших отрицательных чисел — прибавить) от каждой варианты одно и то же число C (в качестве C можно выбрать любое число, расположенное приблизительно посредине вариационного ряда), потом поделить (в случае близких к нулю чисел — умножить) на одно и то же число b (в качестве b можно выбрать наибольшее общее кратное), т. е. перейти к условным вариантам

$$u_i = \frac{x_i \pm C}{b} \qquad (u_i = x_i b).$$

Тогда

$$\mathbf{D}_{\text{в}} = b^2 \left(\frac{1}{n} \sum_{i=1}^{k} n_i u_i^2 - \frac{1}{n^2} \left(\sum_{i=1}^{k} n_i u_i \right)^2 \right) = b^2 (\overline{u^2}_{\text{в}} - \overline{u}_{\text{в}}{}^2)$$

$$\left(\mathbf{D}_{\text{в}} = \frac{\sum\limits_{i=1}^{k} n_i u_i^2}{b^2 n} - \frac{\left(\sum\limits_{i=1}^{k} n_i u_i\right)^2}{(bn)^2} = \frac{\overline{u^2}_{\text{в}} - \overline{u}_{\text{в}}{}^2}{b} \right).$$

Несмещённой оценкой генеральной дисперсии является *исправленная выборочная дисперсия*

$$s^2 = \frac{n}{n-1}\mathbf{D}_{\text{в}}.$$

Поправку $\dfrac{n}{n-1}$ называют *поправкой Бесселя*.

Пример 1. Из генеральной совокупности получена некоторая выборка объёма $n = 100$:

x_i	50	80	90	110	120	140	170
n_i	12	7	6	6	15	24	30

Найти несмещённую оценку генерального среднего.

Решение. Несмещённой оценкой генерального среднего $\overline{x}_{\text{г}}$ является выборочное среднее $\overline{x}_{\text{в}}$. Поскольку варианты выборки являются большими числами, перейдём к условным вариантам

$$u_i = \frac{x_i - 110}{10}$$

(в качестве C выбрано значение четвёртой варианты $x_4 = 110$, а в качестве b — число 10, поскольку все варианты кратны десяти):

$$u_1 = \frac{x_1 - 110}{10} = \frac{50 - 110}{10} = -6; \quad u_2 = \frac{x_2 - 110}{10} = \frac{80 - 110}{10} = -3;$$
$$u_3 = \frac{x_3 - 110}{10} = \frac{90 - 110}{10} = -2; \quad u_4 = \frac{x_4 - 110}{10} = \frac{110 - 110}{10} = 0;$$
$$u_5 = \frac{x_5 - 110}{10} = \frac{120 - 110}{10} = 1; \quad u_6 = \frac{x_6 - 110}{10} = \frac{140 - 110}{10} = 3;$$
$$u_7 = \frac{x_7 - 110}{10} = \frac{170 - 110}{10} = 6.$$

Теперь легко найти выборочное среднее:

$$\overline{x}_{\text{в}} = b \cdot \frac{\sum\limits_{i=1}^{7} n_i u_i}{n} + C = 10 \cdot \frac{\sum\limits_{i=1}^{7} n_i u_i}{100} + 110 =$$
$$= \frac{12 \cdot (-6) + 7 \cdot (-3) + 6 \cdot (-2) + 6 \cdot 0 + 15 \cdot 1 + 24 \cdot 3 + 30 \cdot 6}{10} +$$
$$+ 110 = \frac{-72 + (-21) + (-12) + 0 + 15 + 72 + 180}{10} + 110 =$$
$$= \frac{162}{10} + 110 = 126{,}2.$$

□

Пример 2. Из генеральной совокупности получена некоторая выборка объёма $n = 100$:

x_i	0,002	0,004	0,006	0,008	0,01	0,012	0,014
n_i	7	29	35	12	9	5	3

Найти смещённую оценку генеральной дисперсии.

Решение. Смещённой оценкой генеральной дисперсии $\mathbf{D}_{\text{г}}$ является выборочная дисперсия $\mathbf{D}_{\text{в}}$. Поскольку варианты выборки являются малыми близкими к нулю числами, перейдём к условным вариантам

$$u_i = x_i \cdot 1000$$

(в качестве b выбрано число 1000, поскольку в таком случае мы получим целые числа):

$$u_1 = x_1 \cdot 1000 = 0{,}002 \cdot 1000 = 2; \quad u_2 = x_1 \cdot 1000 = 0{,}004 \cdot 1000 = 4;$$
$$u_3 = x_1 \cdot 1000 = 0{,}006 \cdot 1000 = 6; \quad u_4 = x_1 \cdot 1000 = 0{,}008 \cdot 1000 = 8;$$
$$u_5 = x_1 \cdot 1000 = 0{,}01 \cdot 1000 = 10; \quad u_6 = x_1 \cdot 1000 = 0{,}012 \cdot 1000 = 12;$$
$$u_7 = x_1 \cdot 1000 = 0{,}014 \cdot 1000 = 14.$$

Теперь легко найти выборочную дисперсию:

$$\mathbf{D}_{\text{в}} = \frac{\sum\limits_{i=1}^{7} n_i u_i^2}{b^2 n} - \frac{\left(\sum\limits_{i=1}^{7} n_i u_i\right)^2}{(bn)^2} =$$
$$= \frac{7 \cdot 2^2 + 29 \cdot 4^2 + 35 \cdot 6^2 + 12 \cdot 8^2 + 9 \cdot 10^2 + 5 \cdot 12^2 + 3 \cdot 14^2}{1000^2 \cdot 100} -$$
$$- \left[\frac{7 \cdot 2 + 29 \cdot 4 + 35 \cdot 6 + 12 \cdot 8 + 9 \cdot 10 + 5 \cdot 12 + 3 \cdot 14}{1000 \cdot 100}\right]^2 =$$
$$= \frac{28 + 232 + 1260 + 768 + 900 + 720 + 588}{100000000} -$$
$$- \left[\frac{14 + 116 + 210 + 96 + 90 + 60 + 42}{100000}\right]^2 =$$
$$= 4496 \cdot 10^{-8} - 394384 \cdot 10^{-10} = 5{,}5216 \cdot 10^{-6}. \qquad \square$$

Пример 3. По данным выборки объёма $n = 21$ найдена выборочная дисперсия $\mathbf{D}_\text{в} = 5$.

Найти несмещённую оценку генеральной дисперсии.

Решение. Несмещённой оценкой генеральной дисперсии является исправленная дисперсия:

$$s^2 = \frac{n}{n-1}\mathbf{D}_\text{в} = \frac{21}{21-1} \cdot 5 = 5{,}25. \qquad \square$$

Пример 4. В результате статистических исследований случайной величины X получена такая выборка: 47, 45, 46, 46, 46, 45, 47, 44, 46, 45, 45, 46, 46, 44, 46, 48, 46, 46, 45, 46, 44, 46, 45, 47, 46, 46, 47, 46, 46, 48, 44, 46, 45, 46, 45, 46, 44, 47, 46, 46, 45, 47, 48, 44, 46, 46, 45, 46, 47, 45.

Найти несмещённые оценки генерального среднего и генеральной дисперсии.

Решение. Найдём объём выборки: $n = 50$. Построим статистическое распределение выборки:

x_i	44	45	46	47	48
n_i	6	11	23	7	3

Контроль: $n = 6 + 11 + 23 + 7 + 3 = 50$.

Несмещённой оценкой генерального среднего является выборочное среднее

$$\overline{x}_\text{в} = 46 + \frac{-2 \cdot 6 + (-1) \cdot 11 + 0 \cdot 23 + 1 \cdot 7 + 2 \cdot 3}{50} = 46 - \frac{10}{50} = 45{,}8.$$

Чтобы найти несмещённую оценку генеральной дисперсии — исправленную выборочную дисперсию, определим выборочную дисперсию и умножим её на поправку Бесселя:

$$\mathbf{D}_\text{в} = \frac{(-2)^2\cdot 6+(-1)^2\cdot 11+0^2\cdot 23+1^2\cdot 7+2^2\cdot 3}{50} - \left[\frac{(-2)\cdot 6+(-1)\cdot 11+0\cdot 23+1\cdot 7+2\cdot 3}{50}\right]^2 = \frac{54}{50} - \left[\frac{-10}{50}\right]^2 = 1{,}04;$$

$$s^2 = \frac{50}{49}\cdot 1{,}04 = \frac{52}{49} \approx 1{,}06.$$

Итак, несмещёнными оценками генерального среднего и генеральной дисперсии являются

$$\overline{x}^\star_\text{г} = 45{,}8; \quad \mathbf{D}^\star_\text{г} \approx 1{,}06.$$ □

13.2. Метод моментов

Для того чтобы найти точечные оценки, используют *метод моментов*, при котором для вычисления неизвестных параметров заданного распределения приравнивают соответствующие теоретические и эмпирические моменты.

Если распределение определяется одним параметром, то для нахождения его оценки приравнивают математическое ожидание к выборочному среднему:

$$\mathbf{M}(X) = \overline{x}_\text{в},$$

а потом из этого уравнения определяют искомую точечную оценку неизвестного параметра.

Если распределение определяется двумя параметрами, то их точечные оценки находят из системы уравнений

$$\begin{cases} \mathbf{M}(X) = \overline{x}_\text{в}, \\ \mathbf{D}(X) = \mathbf{D}_\text{в}. \end{cases}$$

Левыми частями этих уравнений являются математическое ожидание и дисперсия, которые приравнивают соответственно к выборочному среднему и выборочной дисперсии.

Пример 1. Случайная величина X (количество бракованных деталей в партии товара) распределена по закону Пуассона с параметром λ. В результате статистических исследований получено такое статистическое распределение количества бракованных деталей в $n = 1000$ партиях товара:

Количество бракованных деталей	0	1	2	3	4	5	6
Количество партий товара	505	284	131	63	12	3	2

Найти методом моментов точечную оценку неизвестного параметра λ распределения Пуассона.

Решение. Поскольку распределение Пуассона зависит лишь от одного параметра, то приравниваем математическое ожидание и выборочное среднее. Математическое ожидание распределения Пуассона равняется его параметру:

$$\mathbf{M}(X) = \lambda,$$

а выборочное среднее

$$\overline{x}_{\text{в}} = \frac{0 \cdot 505 + 1 \cdot 284 + 2 \cdot 131 + 3 \cdot 63 + 4 \cdot 12 + 5 \cdot 3 + 6 \cdot 2}{1000} = 0{,}81.$$

Итак, точечная оценка неизвестного параметра λ распределения Пуассона

$$\lambda^{\star} = 0{,}81. \qquad \square$$

Пример 2. Случайная величина X (рост взрослого человека) распределена по нормальному распределению с параметрами a, σ. В результате статистических исследований получено такое статистическое распределение роста взрослых людей для $n = 1000$ лиц:

Рост, см	Количество лиц
(145; 155]	24
(155; 165]	112
(165; 175]	263
(175; 185]	322
(185; 195]	202
(195; 205]	66
(205; 215]	11

Найти методом моментов точечную оценку неизвестных параметров a, σ нормального распределения.

Решение. Превратим интервальное статистическое распределение в точечное, выбрав в качестве вариант середины частных интервалов:

Рост, см	150	160	170	180	190	200	210
Количество лиц	24	112	263	322	202	66	11

.

Поскольку нормальный закон распределения зависит от двух параметров, нужно найти выборочное среднее и выборочную дисперсию, а затем приравнять их соответственно к математическому ожиданию и дисперсии.

Перейдём к условным вариантам

$$u_i = \frac{x_i - 180}{10}$$

(x_i — рост человека):

$$u_1 = \frac{150-180}{10} = -3; \qquad u_2 = \frac{160-180}{10} = -2;$$
$$u_3 = \frac{170-180}{10} = -1; \qquad u_4 = \frac{180-180}{10} = 0;$$
$$u_5 = \frac{190-180}{10} = 1; \qquad u_6 = \frac{200-180}{10} = 2;$$
$$u_7 = \frac{210-180}{10} = 3.$$

Найдём выборочное среднее и выборочную дисперсию:

$$\overline{u}_{\text{в}} = \frac{-3\cdot 24 - 2\cdot 112 - 1\cdot 263 + 0\cdot 322 + 1\cdot 202 + 2\cdot 66 + 3\cdot 11}{1000} =$$
$$= -0{,}192;$$
$$\overline{x}_{\text{в}} = 10\cdot\overline{u}_{\text{в}} + 180 = 178{,}08;$$
$$\mathbf{D}_{\text{в}} = 10^2\cdot\Big([(-3)^2\cdot 24 + (-2)^2\cdot 112 + (-1)^2\cdot 263 + 0^2\cdot 322 +$$
$$+ 1^2\cdot 202 + 2^2\cdot 66 + 3^2\cdot 11]/1000 - [-0{,}192]^2\Big) = 145{,}5136.$$

Поскольку параметр a нормального закона распределения является математическим ожиданием, а параметр σ — средним квадратическим отклонением, то оценками этих параметров являются

$$a^\star = \overline{x}_{\text{в}} = 178{,}08; \qquad \sigma^\star = \sqrt{\mathbf{D}_{\text{в}}} = \sqrt{145{,}5136} \approx 12{,}0629. \qquad \square$$

13.3. Метод наибольшего правдоподобия

Метод наибольшего (максимального) правдоподобия заключается в нахождении максимума функции одного или нескольких оцениваемых параметров.

Предположим, что X — дискретная случайная величина с известным законом распределения, однако нам неизвестен параметр Θ, которым определяется этот закон распределения. По данным выборки $x_1, x_2, \ldots, x_n$, полученной в результате наблюдений над случайной величиной X, необходимо найти точечную оценку $\Theta^\star = \Theta^\star(x_1, x_2, \ldots, x_n)$ параметра Θ.

Функцией правдоподобия дискретной случайной величины называют функцию аргумента Θ

$$L(x_1, x_2, \ldots, x_n; \Theta) = p(x_1; \Theta) \cdot p(x_2; \Theta) \cdot \ldots \cdot p(x_n; \Theta),$$

где $p(x_i; \Theta)$ — вероятность того, что в результате испытания случайная величина X примет значение x_i.

Оценкой наибольшего или максимального правдоподобия параметра Θ называют такое его значение $\Theta^\star$, при котором функция правдоподобия достигает своего максимума.

Логарифмической функцией правдоподобия называют функцию $\ln L$.

Поскольку функции L и $\ln L$ достигают своего максимума при одном и том же значении аргумента Θ, довольно часто удобнее находить максимум функции $\ln L$, а не L.

Если случайная величина X непрерывна, то известной считается плотность распределения вероятностей $f(x)$, а неизвестным — параметр, от которого зависит эта плотность.

Функцией правдоподобия непрерывной случайной величины называют функцию аргумента Θ

$$L(x_1, x_2, \ldots, x_n; \Theta) = f(x_1; \Theta) \cdot f(x_2; \Theta) \cdot \ldots \cdot f(x_n; \Theta).$$

Оценку наибольшего правдоподобия неизвестного параметра распределения непрерывной случайной величины ищут так же, как и в случае дискретной случайной величины, а именно:

1) определяют производную $\dfrac{d\ln L}{d\Theta}$ (или $\dfrac{dL}{d\Theta}$);

2) находят корни $\Theta_i^\star$ уравнения $\dfrac{d\ln L}{d\Theta} = 0$ (или $\dfrac{dL}{d\Theta}$=0). Эти уравнения называют *уравнениями правдоподобия*;

3) определяют вторую производную $\dfrac{d^2\ln L}{d\Theta^2}$ (или $\dfrac{d^2 L}{d\Theta^2}$). Корень $\Theta_i^\star$ уравнения правдоподобия, для которого вторая производная отрицательная, берут в качестве оценки $\Theta^\star$ наибольшего правдоподобия параметра Θ.

Если параметр Θ двумерный, т. е. $\Theta = (\Theta_1, \Theta_2)$, то для отыскания максимума функции правдоподобия составляют и решают систему уравнений

$$\begin{cases} \dfrac{\partial \ln L}{\partial \Theta_1} = 0, \\ \dfrac{\partial \ln L}{\partial \Theta_2} = 0 \end{cases} \quad \text{или} \quad \begin{cases} \dfrac{\partial L}{\partial \Theta_1} = 0, \\ \dfrac{\partial L}{\partial \Theta_2} = 0. \end{cases}$$

Пример 1. Случайная величина X (количество битой стеклянной посуды в одной упаковке) распределена по закону Пуассона с неизвестным параметром λ. В результате статистических исследований получено такое эмпирическое распределение количества битой стеклянной посуды в $n = 1000$ упаковках:

x_i	0	1	2	3	4	5	6
n_i	554	324	98	19	3	1	1

Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра λ распределения Пуассона.

Решение. Поскольку случайная величина X распределена по закону Пуассона, то функция правдоподобия имеет вид

$$L = \left(\frac{\lambda^0 e^{-\lambda}}{0!}\right)^{554} \left(\frac{\lambda^1 e^{-\lambda}}{1!}\right)^{324} \left(\frac{\lambda^2 e^{-\lambda}}{2!}\right)^{98} \left(\frac{\lambda^3 e^{-\lambda}}{3!}\right)^{19} \times$$
$$\times \left(\frac{\lambda^4 e^{-\lambda}}{4!}\right)^{3} \left(\frac{\lambda^5 e^{-\lambda}}{5!}\right)^{1} \left(\frac{\lambda^6 e^{-\lambda}}{6!}\right)^{1} = \frac{\lambda^{600} e^{-1000\lambda}}{2^{133}\cdot 3^{25}\cdot 5^2}.$$

Запишем логарифмическую функцию правдоподобия:

$$\ln L = 600\ln\lambda - 1000\lambda - \ln(2^{133}\cdot 3^{25}\cdot 5^2).$$

Определим производную логарифмической функции распределения по λ:

$$\frac{d\ln L}{d\lambda} = \frac{600}{\lambda} - 1000.$$

Приравняв её к нулю, найдём единственный корень уравнения правдоподобия:

$$\lambda^\star = 0{,}6.$$

Поскольку вторая производная логарифмической функции правдоподобия

$$\frac{d^2\ln L}{d\lambda^2} = -\frac{600}{\lambda^2} < 0$$

всегда отрицательная, точечной оценкой максимального правдоподобия параметра λ распределения Пуассона будет

$$\lambda^\star = 0{,}6. \qquad \square$$

Пример 2. Случайная величина X (длина детали) распределена по нормальному закону с неизвестными параметрами a, σ. В результате статистических исследований получено такое эмпирическое распределение длины $n = 1000$ деталей:

Длина детали, см	5,1	5,2	5,3	5,4	5,5	5,6	5,7
Количество деталей	7	78	289	392	195	36	3

Найти методом наибольшего правдоподобия точечную оценку неизвестных параметров a, σ нормального распределения.

Решение. Плотность распределения вероятностей случайной величины X

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

Запишем функцию правдоподобия:

$$\begin{aligned}L=&\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5{,}1-a)^2}{2\sigma^2}}\right)^7\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5{,}2-a)^2}{2\sigma^2}}\right)^{78}\times\\
&\times\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5{,}3-a)^2}{2\sigma^2}}\right)^{289}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5{,}4-a)^2}{2\sigma^2}}\right)^{392}\times\\
&\times\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5{,}5-a)^2}{2\sigma^2}}\right)^{195}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5{,}6-a)^2}{2\sigma^2}}\right)^{36}\times\\
&\times\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(5{,}7-a)^2}{2\sigma^2}}\right)^{3}=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{1000}\times\\
&\times\exp\left\{[-7\cdot(5{,}1-a)^2+78\cdot(5{,}2-a)^2+289\cdot(5{,}3-a)^2+\right.\\
&+392\cdot(5{,}4-a)^2+195\cdot(5{,}5-a)^2+36\cdot(5{,}6-a)^2+\\
&\left.+3\cdot(5{,}7-a)^2]/[2\sigma^2]\right\}=\\
=&\left(2\pi\sigma^2\right)^{-500}\cdot e^{-\frac{1000a^2-10762a+28965{,}1}{2\sigma^2}}.\end{aligned}$$

Запишем логарифмическую функцию правдоподобия:

$$\ln L=-500\cdot\ln 2\pi\sigma^2-\frac{1000a^2-10762a+28965{,}1}{2\sigma^2}.$$

Определим частные производные логарифмической функции распределения по a и σ^2:

$$\frac{d\ln L}{da}=-\frac{2000a-10762}{2\sigma^2};$$

$$\frac{d\ln L}{d\sigma^2}=-\frac{500}{\sigma^2}+\frac{1000a^2-10762a+28965{,}1}{2\sigma^4}.$$

Приравняв их к нулю, найдём решение системы уравнений:

$$\begin{cases} -\dfrac{2000a - 10762}{2\sigma^2} = 0; \\ -\dfrac{500}{\sigma^2} + \dfrac{1000a^2 - 10762a + 28965{,}1}{2\sigma^4} = 0. \end{cases}$$

Решением будет:

$$a = 5{,}381; \qquad \sigma^2 = 0{,}009939.$$

Итак, точечными оценками максимального правдоподобия параметров a, σ нормального распределения являются

$$a^\star = 5{,}381; \qquad \sigma^\star = \sqrt{0{,}009939} \approx 0{,}1. \qquad \square$$

13.4. Интервальные оценки

Интервальной называют оценку, которая определяется числовым интервалом.

Доверительным называют интервал (Θ_1, Θ_2), в который с заданной надежностью α (вероятностью, близкой к единице) попадает оцениваемый параметр Θ:

$$\mathbf{P}(\Theta_1 < \Theta < \Theta_2) = \alpha, \quad \alpha \to 1.$$

1. *Интервальной оценкой с надежностью α математического ожидания a нормально распределенной случайной величины X по выборочному среднему $\overline{x}_\text{в}$ при известном среднем квадратическом отклонении σ* называют доверительный интервал

$$\overline{x}_\text{в} - t\frac{\sigma}{\sqrt{n}} < a < \overline{x}_\text{в} + t\frac{\sigma}{\sqrt{n}},$$

где $t\dfrac{\sigma}{\sqrt{n}} = \delta$ — точность оценки; t — значение аргумента функции Лапласа $\Phi(t)$ (прил. 1), при котором $\Phi(t) = \dfrac{\alpha}{2}$; n — объём выборки;

при неизвестном σ (и объёме выборки $n < 30$)

$$\overline{x}_{\text{в}} - t_\alpha \frac{s}{\sqrt{n}} < a < \overline{x}_{\text{в}} + t_\alpha \frac{s}{\sqrt{n}},$$

где t_α находят по прил. 4 по заданным n и α; s — «исправленное» выборочное среднее квадратическое отклонение.

2. *Интервальной оценкой с надежностью* α *среднего квадратического отклонения* σ *нормально распределенной случайной величины* X *по «исправленному» выборочному среднему квадратическому отклонению* s называют доверительный интервал

$$\begin{aligned} s(1-q) < \sigma < s(1+q) &\qquad \text{при} \qquad q < 1; \\ 0 < \sigma < s(1+q) &\qquad \text{при} \qquad q > 1, \end{aligned}$$

где q находят по прил. 5 по заданным n и α.

3. *Интервальной оценкой с надежностью* α *неизвестной вероятности* p *биномиального распределения по относительной частоте* w называют доверительный интервал

$$p_1 < p < p_2,$$

где

$$p_1 = \frac{n}{t^2+n}\left(w + \frac{t^2}{2n} - t\sqrt{\frac{w(1-w)}{n} + \left(\frac{t}{2n}\right)^2}\right);$$

$$p_2 = \frac{n}{t^2+n}\left(w + \frac{t^2}{2n} + t\sqrt{\frac{w(1-w)}{n} + \left(\frac{t}{2n}\right)^2}\right);$$

n — общее количество испытаний; t — значение аргумента функции Лапласа (прил. 1), при котором $\Phi(t) = \dfrac{\alpha}{2}$; $w = \dfrac{m}{n}$ — относительная частота; m — количество появлений события.

Замечание 1. При больших значениях n (несколько сотен) можно принять в качестве границ доверительного интервала такие значения:

$$p_1 = w - t\sqrt{\frac{w(1-w)}{n}}; \qquad p_2 = w + t\sqrt{\frac{w(1-w)}{n}}.$$

Пример 1. Найти доверительный интервал с надежностью 0,95 для оценки неизвестного математического ожидания a нормально распределенной случайной величины X, если дисперсия этой случайной величины $\sigma^2 = 16$, выборочное среднее $\overline{x}_{\text{в}} = 15$, а объём выборки $n = 25$.

Решение. Нужно определить доверительный интервал

$$\overline{x}_{\text{в}} - t\frac{\sigma}{\sqrt{n}} < a < \overline{x}_{\text{в}} + t\frac{\sigma}{\sqrt{n}}.$$

Для этого вычислим значения t и σ. Из соотношения

$$\Phi(t) = \frac{\alpha}{2} = \frac{0{,}95}{2} = 0{,}475$$

по прил. 1 находим

$$t = 1{,}96.$$

Среднее квадратическое отклонение

$$\sigma = \sqrt{\sigma^2} = \sqrt{16} = 4.$$

Подставив все значения, получим искомый доверительный интервал:

$$15 - 1{,}96 \cdot \frac{4}{\sqrt{25}} < a < 15 + 1{,}96 \cdot \frac{4}{\sqrt{25}},$$

т. е.

$$13{,}432 < a < 16{,}568. \qquad \square$$

Пример 2. С конвейера поступают электрические лампы. Каким должен быть минимальный размер партии электроламп для того, чтобы с надежностью 0,99 точность оценки математического ожидания a случайной величины X, которая характеризует продолжительность горения лампы, по выборочному среднему составляла $\varepsilon = 1$ ч, если известно среднее квадратическое отклонение случайной величины X: $\sigma = 3$ ч?

Считается, что случайная величина X имеет нормальный закон распределения.

Решение. Поскольку доверительный интервал для математического ожидания a случайной величины X вычисляется по формуле

$$\overline{x}_{\text{в}} - t\frac{\sigma}{\sqrt{n}} < a < \overline{x}_{\text{в}} + t\frac{\sigma}{\sqrt{n}},$$

точность ε оценки определяется так:

$$\varepsilon = t\frac{\sigma}{\sqrt{n}}.$$

Отсюда получаем формулу для вычисления минимального объёма выборки, который обеспечивает заданную точность оценивания:

$$n = \min\left\{n\colon n > t^2\frac{\sigma^2}{\varepsilon^2}\right\}.$$

Найдём t из соотношения

$$\Phi(t) = \frac{0{,}99}{2} = 0{,}495$$

по прил. 1:

$$t = 2{,}58.$$

Итак, минимальный объём выборки

$$n = \min\left\{n\colon n > 2{,}582 \cdot \frac{32}{12} = 23{,}22\right\} = 24. \qquad \square$$

Пример 3. В результате статистических исследований случайной величины X получена выборка объёма $n = 25$ с таким статистическим распределением:

x_i	0	1	2	3	4	5	6
n_i	1	3	4	6	5	4	2

Найти с надёжностью $\alpha = 0{,}99$ интервальную оценку математического ожидания a случайной величины X по выборочному среднему.

Считается, что случайная величина X нормально распределена.

Решение. Прежде чем находить доверительный интервал для математического ожидания a случайной величины X по формуле

$$\overline{x}_{\text{в}} - t_\alpha \frac{s}{\sqrt{n}} < a < \overline{x}_{\text{в}} + t_\alpha \frac{s}{\sqrt{n}},$$

необходимо определить выборочное среднее $\overline{x}_{\text{в}}$, «исправленное» выборочное среднее квадратическое отклонение s и t_α (прил. 4):

$$\overline{x}_{\text{в}} = \frac{\sum\limits_{i=1}^{7} n_i x_i}{n} = \frac{0\cdot 1 + 1\cdot 3 + 2\cdot 4 + 3\cdot 6 + 4\cdot 5 + 5\cdot 4 + 6\cdot 2}{25} = 3{,}24;$$

$$s = \sqrt{\frac{n}{n-1}\mathbf{D}_{\text{в}}} = \sqrt{\frac{n}{n-1}\left(\frac{\sum\limits_{i=1}^{7} n_i x_i^2}{n} - [\overline{x}_{\text{в}}]^2\right)} =$$

$$= \left[\frac{25}{25-1}\left(\frac{0^2\cdot 1 + 1^2\cdot 3 + 2^2\cdot 4 + 3^2\cdot 6 + 4^2\cdot 5 + 5^2\cdot 4 + 6^2\cdot 2}{25} - \right.\right.$$

$$\left.\left. - 3{,}24^2\right)\right]^{\frac{1}{2}} = \sqrt{\frac{25}{24}\cdot 2{,}5024} \approx 1{,}6145;$$

$$t_\alpha = t_\alpha(n,\alpha) = t_{0,99}(25; 0{,}99) = 2{,}797.$$

Искомый доверительный интервал

$$3{,}24 - 2{,}797\cdot\frac{1{,}6145}{\sqrt{25}} < a < 3{,}24 + 2{,}797\cdot\frac{1{,}6145}{\sqrt{25}},$$

т. е.

$$2{,}337 < a < 4{,}143. \qquad \square$$

Пример 4. По данным выборки объёма $n = 20$ определено «исправленное» выборочное среднее квадратическое отклонение $s = 2$ нормально распределенной случайной величини X.

Найти доверительный интервал для среднего квадратического отклонения σ случайной величины X с надёжностью $\alpha = 0{,}95$.

Решение. Сначала найдём по прил. 5 значение q:

$$q = q(n, \alpha) = q(20; 0{,}95) = 0{,}37.$$

Поскольку $q < 1$, доверительный интервал для среднего квадратического отклонения σ случайной величины X с надёжностью $\alpha = 0{,}95$ вычисляем по формуле

$$s(1 - q) < \sigma < s(1 + q),$$

т. е.

$$1{,}26 < \sigma < 2{,}74. \qquad \square$$

Пример 5. В результате статистических исследований случайной величины X получена выборка объёма $n = 15$ с таким статистическим распределением:

x_i	1	2	3	4	5
n_i	1	4	6	3	1

Найти с надёжностью $\alpha = 0{,}999$ интервальную оценку среднего квадратического отклонения σ случайной величины X.

Считается, что случайная величина X распределена по нормальному закону.

Решение. Сначала найдём по прил. 5 значение q и вычислим «исправленное» выборочное среднее квадратическое отклонение s:

$$q = q(n, \alpha) = q(15; 0{,}999) = 1{,}15;$$

$$s = \sqrt{\frac{n}{n-1}\mathbf{D}_{\text{в}}} = \sqrt{\frac{n}{n-1}\left(\frac{\sum\limits_{i=1}^{5} n_i x_i^2}{n} - \frac{\left(\sum\limits_{i=1}^{5} n_i x_i\right)^2}{n^2}\right)} =$$

$$= \sqrt{\frac{15}{14}\left(\frac{144}{15} - \frac{44^2}{15^2}\right)} = \sqrt{\frac{16}{15}} \approx 1{,}0328.$$

Поскольку $q > 1$, доверительный интервал для среднего квадратического отклонения σ случайной величины X с надёжностью $\alpha = 0{,}999$ вычисляем по формуле

$$0 < \sigma < s(1+q),$$

т. е.

$$0 < \sigma < 2{,}22. \qquad \square$$

Пример 6. Проводятся независимые испытания с одинаковой, но неизвестной вероятностью p успеха.

Найти доверительный интервал для оценки вероятности p с надёжностью $\alpha = 0{,}99$, если из 50 испытаний успешными были 12.

Решение. Для того чтобы найти доверительный интервал для оценки вероятности p, необходимо определить границы этого интервала по формулам

$$p_1 = \frac{n}{t^2+n}\left(w + \frac{t^2}{2n} - t\sqrt{\frac{w(1-w)}{n} + \left(\frac{t}{2n}\right)^2}\right);$$

$$p_2 = \frac{n}{t^2+n}\left(w + \frac{t^2}{2n} + t\sqrt{\frac{w(1-w)}{n} + \left(\frac{t}{2n}\right)^2}\right).$$

Найдём значение t из соотношения

$$\Phi(t) = \frac{\alpha}{2} = 0{,}495,$$

воспользовавшись прил. 1:

$$t = 2{,}58;$$

относительная частота

$$w = \frac{12}{50} = 0{,}24.$$

Объём выборки $n = 50$. Использовав эти значения, определим концы доверительного интервала:

$$p_1 = \frac{50}{2{,}58^2 + 50}\left[0{,}24 + \frac{2{,}58^2}{2 \cdot 50} - \right.$$
$$\left. - 2{,}58\sqrt{\frac{0{,}24(1 - 0{,}24)}{50} + \left(\frac{2{,}58}{2 \cdot 50}\right)^2}\,\right] \approx 0{,}12;$$
$$p_2 = \frac{50}{2{,}58^2 + 50}\left[0{,}24 + \frac{2{,}58^2}{2 \cdot 50} + \right.$$
$$\left. + 2{,}58\sqrt{\frac{0{,}24(1 - 0{,}24)}{50} + \left(\frac{2{,}58}{2 \cdot 50}\right)^2}\,\right] \approx 0{,}42.$$

Итак, искомый доверительный интервал такой:

$$0{,}12 < p < 0{,}42. \qquad \square$$

Пример 7. Во время заливания кипящей воды в стакан оказалось, что из 1000 стаканов треснули или разбились 24.

Найти с надёжностью $\alpha = 0{,}95$ интервальную оценку вероятности повреждения стакана при заливании в него кипящей воды.

Решение. Определим относительную частоту повреждения стакана:

$$w = \frac{24}{1000} = 0{,}024.$$

Найдём значение t из соотношения

$$\Phi(t) = \frac{\alpha}{2} = \frac{0{,}95}{2} = 0{,}475,$$

воспользовавшись прил. 1:

$$t = 1{,}96.$$

Учитывая, что объём выборки довольно большой ($n = 1000$), для вычисления значений концов доверительного интервала оценки вероятности повреждения стакана используем формулы:

$$p_1 = w - t\sqrt{\frac{w(1-w)}{n}} = 0{,}024 - 1{,}96\sqrt{\frac{0{,}024(1-0{,}024)}{1000}} \approx 0{,}0145;$$

$$p_2 = w + t\sqrt{\frac{w(1-w)}{n}} = 0{,}024 + 1{,}96\sqrt{\frac{0{,}024(1-0{,}024)}{1000}} \approx 0{,}0335.$$

Итак, искомый доверительный интервал такой:

$$0{,}0145 < p < 0{,}0335. \qquad \square$$

Задачи к разделу 13

Задача 1. Из генеральной совокупности получена такая выборка объёма $n = 100$:

x_i	0,03	0,04	0,05	0,06	0,07	0,08	0,09
n_i	15	10	9	4	12	21	29

Найти несмещённую оценку генерального среднего.

Указание. Перейдите к условным вариантам $u_i = 100x_i - 6$.

Ответ. $\overline{x}_\text{в} = 0{,}0667$.

Задача 2. Из генеральной совокупности получена такая выборка объёма $n = 100$:

x_i	260	300	340	380	420	460	500
n_i	8	30	36	11	8	4	3

Найти смещённую оценку генеральной дисперсии.

Указание. Перейдите к условным вариантам $u_i = \dfrac{x_i - 380}{40}$.

Ответ. $\mathbf{D}_\text{в} = 3020$.

Задача 3. По данным выборки объёма $n = 51$ определена выборочная дисперсия $\mathbf{D}_\text{в} = 15$.

Найти несмещённую оценку генеральной дисперсии.

Ответ. $s^2 = 15{,}3$.

Задача 4. В результате статистических исследований случайной величины X получена такая выборка: 0,01; 0,03; 0,02; 0,02; 0,03; 0,04; 0,01; 0,05; 0,04; 0,03; 0,03; 0,03; 0,02; 0,03; 0,01; 0,04; 0,03; 0,02; 0,05; 0,03; 0,03; 0,04; 0,02; 0,02; 0,01; 0,03; 0,02; 0,03; 0,04; 0,02; 0,01; 0,03; 0,03; 0,03; 0,05; 0,04; 0,02; 0,02; 0,03; 0,03; 0,05; 0,03; 0,02; 0,03; 0,01; 0,03; 0,04; 0,02; 0,03; 0,03.

Найти несмещённые оценки генерального среднего и генеральной дисперсии.

Указание. Перейдите к условным вариантам $u_i = 100x_i - 3$.

Ответ. $\overline{x}_{\text{в}} = 0{,}0282$, $s^2 = \dfrac{2869}{24500000} \approx 0{,}0001171$.

Задача 5. Случайная величина X (количество найденных самородков одним старателем за один день) распределена по закону Пуассона с параметром λ. В результате статистических исследований получено такое статистическое распределение количества найденных самородков $n = 1000$ старателями:

Количество найденных самородков	0	1	2	3	4	5	6
Количество старателей	369	367	183	62	15	3	1

Найти методом моментов точечную оценку неизвестного параметра λ распределения Пуассона.

Ответ. $\lambda^\star = 1$.

Задача 6. Случайная величина X (количество бракованных единиц товара) распределена по закону Пуассона с неизвестным параметром λ. В результате статистических исследований получено такое эмпирическое распределение количества брака в $n = 1000$ единицах товара:

x_i	0	1	2	3	4	5	6
n_i	82	208	261	217	136	68	28

Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра λ распределения Пуассона.

Ответ. $\lambda^\star = 2{,}433$.

Задача 7. Случайная величина X (масса тела взрослого человека) распределена по нормальному закону с параметрами a, σ. В результате статистических исследований получено такое статистическое распределение массы тела взрослых людей для $n = 1000$ лиц.

Масса тела, кг	Количество людей
(45; 55]	57
(55; 65]	136
(65; 75]	223
(75; 85]	249
(85; 95]	191
(95; 105]	100
(105; 115]	36
(115; 125]	8

Найти методом моментов точечную оценку неизвестных параметров a, σ нормального распределения.

Указание. Превратить интервальное статистическое распределение в точечное и перейти к условным вариантам $u_i = \dfrac{x_i - 80}{10}$ (x_i — масса тела человека).

Ответ. $a^\star = 78{,}65$; $\sigma^\star \approx 15{,}18$.

Задача 8. Случайная величина X (дальность полёта ракеты) распределена по нормальному закону с неизвестными параметрами a, σ. В результате статистических исследований получено такое эмпирическое распределение дальности полёта $n = 1000$ ракет:

Дальность полёта, км	32	33	34	35	36	37	38
Количество ракет	7	78	289	392	195	36	3

Найти методом наибольшего правдоподобия точечную оценку неизвестных параметров a, σ нормального распределения.

Ответ. $a^\star = 35{,}3$, $\sigma^\star \approx 0{,}991$.

Задача 9. Найти доверительный интервал с надёжностью 0,999 для оценки неизвестного математического ожидания a нормально распределённой случайной величины X, если дисперсия этой случайной величины $\sigma^2 = 25$, выборочное среднее $\overline{x}_{\text{в}} = 3$, а объём выборки $n = 36$.

Ответ. $0{,}25 < a < 5{,}75$.

Задача 10. На одном и том же станке изготовляются детали. Каким должно быть минимальное количество деталей для того, чтобы с надёжностью 0,95 точность оценки математического ожидания a случайной величины X, которая характеризует длину детали, по выборочному среднему составляла $\varepsilon = 3$ мм, если известно среднее квадратическое отклонение случайной величины X: $\sigma = 5$ мм?

Считается, что случайная величина X имеет нормальный закон распределения.

Ответ. $n = 11$.

Задача 11. В результате статистических исследований случайной величины X получена выборка объёма $n = 50$ с таким статистическим распределением:

x_i	-3	-2	-1	0	1	2	3
n_i	2	4	8	11	11	9	5

Найти с надёжностью $\alpha = 0{,}999$ интервальную оценку математического ожидания a случайной величины X по выборочному среднему.

Считается, что случайная величина X нормально распределена.

Ответ. $-0{,}349 < a < 1{,}229$.

Задача 12. По данным выборки объёма $n = 25$ найдено «исправленное» выборочное среднее квадратическое отклонение $s = 1$ нормально распределённой случайной величини X.

Найти доверительный интервал для среднего квадратического отклонения σ случайной величины X с надёжностью $\alpha = 0{,}99$.

Ответ. $0{,}27 < \sigma < 1{,}73$.

Задача 13. В результате статистических исследований случайной величины X получена выборка объёма $n = 10$ с таким статистическим распределением:

x_i	-1	0	1	2
n_i	1	3	4	2

Найти с надёжностью $\alpha = 0{,}99$ интервальную оценку среднего квадратического отклонения σ случайной величины X.

Считается, что случайная величина X распределена по нормальному закону.

Ответ. $0 < \sigma < 1{,}872$.

Задача 14. Проводятся независимые испытания с одинаковой, но неизвестной вероятностью p успеха.

Найти доверительный интервал для оценки вероятности p с надёжностью $\alpha = 0{,}95$, если из 100 испытаний успешными были 27.

Ответ. $0{,}193 < p < 0{,}364$.

Задача 15. Во время испытания стали на прочность оказалось, что из 1000 стальных прутьев не выдержали испытания 57.

Найти с надёжностью $\alpha = 0{,}999$ интервальную оценку вероятности того, что стальной прут не пройдёт испытания на прочность.

Ответ. $0{,}0328 < p < 0{,}0812$.

Раздел 14. Элементы теории регрессии и корреляции

Ключевые слова

Выборочный	Линия
Корреляционный	Параболический
Корреляция	Прямой
Кривой	Отношение
Криволинейный	Регрессия
Линейный	Уравнение

14.1. Уравнение прямой линии регрессии. Линейная корреляция

Выборочным уравнением прямой линии регрессии Y на X (X и Y — наблюдаемые случайные величины) называют уравнение

$$\overline{y}_x - \overline{y} = r_{xy}\frac{\sigma_y}{\sigma_x}(x - \overline{x}),$$

где $\overline{y}_x$ — условное среднее; $\overline{x}$ и $\overline{y}$ — выборочные средние случайных величин X и Y соответственно; σ_x и σ_y — выборочные средние квадратические отклонения; r_{xy} — выборочный коэффициент корреляции, причём

$$r_{xy} = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} w_{ij}x_i y_j - \overline{x}\cdot\overline{y}}{\sigma_x\sigma_y};$$

$$\overline{x}_y - \overline{x} = r_{xy}\frac{\sigma_x}{\sigma_y}(y - \overline{y}),$$

w_{ij} — эмпирическая вероятность появления значения (x_i, y_j); n и m — количество наблюдаемых вариант случайных величин X и Y соответственно.

Если обе линии регрессии Y на X и X на Y — прямые, корреляцию называют *линейной*.

Если данные наблюдений над случайными величинами X и Y заданы корреляционной таблицей с равноотстоящими вариантами, целесообразно перейти к условным вариантам:

$$u_i = \frac{x_i - C_1}{h_1}; \quad v_j = \frac{y_j - C_2}{h_2},$$

где C_1 и C_2 — «ложный нуль» вариант случайных величин X и Y соответственно; h_1 и h_2 — шаг вариант случайных величин X и Y ($h_1 = x_{i+1} - x_i$, $h_2 = y_{j+1} - y_j$) соответственно.

В этом случае выборочный коэффициент корреляции

$$r_{xy} = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} w_{ij}u_i v_j - \overline{u}\cdot\overline{v}}{\sigma_u \sigma_v}.$$

Величины $\overline{u}$, $\overline{v}$, σ_u, σ_v находят по формулам

$$\overline{u} = \sum_{i=1}^{n} w_i u_i; \qquad w_i = \sum_{k=1}^{m} w_{ik}; \qquad \sigma_u = \sqrt{\overline{u2} - \overline{u}2};$$

$$\overline{v} = \sum_{j=1}^{m} w_j v_j; \qquad w_j = \sum_{k=1}^{n} w_{kj}; \qquad \sigma_v = \sqrt{\overline{v2} - \overline{v}2},$$

а при большом количестве данных — методом произведений.

Зная эти значения, можно перейти обратно к величинам, которые входят в уравнение регрессии, с помощью формул

$$\overline{x} = \overline{u}h_1 + C_1; \qquad \sigma_x = \sigma_u h_1;$$

$$\overline{y} = \overline{v}h_2 + C_2; \qquad \sigma_y = \sigma_v h_2.$$

Пример 1. Найти выборочное уравнение прямой линии регрессии Y на X по данным корреляционной таблицы (табл. 8).

Решение. Преобразуем корреляционную таблицу:

- введём условные варианты $u_i = \dfrac{x_i - 65}{10}$, $v_j = \dfrac{y_j - 18}{3}$;
- все частоты заменим эмпирическими вероятностями.

Таблица 8

X	Y					n_x
	12	15	18	21	24	
45	8	12	—	—	—	20
55	1	9	11	—	—	21
65	—	3	15	14	—	32
75	—	—	4	10	8	22
85	—	—	—	1	4	5
n_y	9	24	30	25	12	$n = 100$

Итак, получена табл. 9.

Таблица 9

U	V					w_u
	-2	-1	0	1	2	
-2	0,08	0,12	—	—	—	0,2
-1	0,01	0,09	0,11	—	—	0,21
0	—	0,03	0,15	0,14	—	0,32
1	—	—	0,04	0,1	0,08	0,22
2	—	—	—	0,01	0,04	0,05
w_v	0,09	0,24	0,3	0,25	0,12	$w = 1$

Найдём $\overline{u}$, $\overline{v}$:

$$\overline{u} = \sum_{i=1}^{5} w_i u_i =$$

$$= 0{,}2 \cdot (-2) + 0{,}21 \cdot (-1) + 0{,}32 \cdot 0 + 0{,}22 \cdot 1 + 0{,}05 \cdot 2 = -0{,}29;$$

$$\overline{v} = \sum_{j=1}^{5} w_i v_i =$$

$$= 0{,}09 \cdot (-2) + 0{,}24 \cdot (-1) + 0{,}3 \cdot 0 + 0{,}25 \cdot 1 + 0{,}12 \cdot 2 = 0{,}07.$$

Вычислим $\overline{u^2}$, $\overline{v^2}$:

$$\overline{u^2} = \sum_{i=1}^{5} w_i u_i^2 = 0{,}2 \cdot (-2)^2 + 0{,}21 \cdot (-1)^2 + 0{,}32 \cdot 0^2 + 0{,}22 \cdot 1^2 +$$

$$+\,0{,}05\cdot 2^2 = 1{,}43;$$

$$\overline{v^2} = \sum_{j=1}^{5} w_i v_i^2 = 0{,}09\cdot(-2)^2 + 0{,}24\cdot(-1)^2 + 0{,}3\cdot 0^2 + 0{,}25\cdot 1^2 + {}$$
$$+\,0{,}12\cdot 2^2 = 1{,}33.$$

Найдём σ_u, σ_v:

$$\sigma_u = \sqrt{\overline{u^2} - \overline{u}^2} = \sqrt{1{,}43 - (-0{,}29)^2} \approx 1{,}16;$$
$$\sigma_v = \sqrt{\overline{v^2} - \overline{v}^2} = \sqrt{1{,}33 - 0{,}07^2} \approx 1{,}15.$$

Определим $\sum\limits_{i=1}^{5}\sum\limits_{j=1}^{5} w_{ij}u_iv_j$:

$$\begin{aligned}\sum_{i=1}^{5}\sum_{j=1}^{5} w_{ij}u_iv_j &= 0{,}08\cdot(-2)\cdot(-2) + 0{,}12\cdot(-2)\cdot(-1) + {}\\ &\quad + 0{,}01\cdot(-1)\cdot(-2) + 0{,}09\cdot(-1)\cdot(-1) + {}\\ &\quad + 0{,}11\cdot(-1)\cdot 0 + 0{,}03\cdot 0\cdot(-1) + 0{,}15\cdot 0\cdot 0 + {}\\ &\quad + 0{,}14\cdot 0\cdot 1 + 0{,}04\cdot 1\cdot 0 + 0{,}1\cdot 1\cdot 1 + {}\\ &\quad + 0{,}08\cdot 1\cdot 2 + 0{,}01\cdot 2\cdot 1 + 0{,}04\cdot 2\cdot 2 = \\ &= 1{,}11.\end{aligned}$$

Рассчитаем выборочный коэффициент корреляции:

$$r_{xy} = r_{uv} = \frac{\sum\limits_{i=1}^{5}\sum\limits_{j=1}^{5} w_{ij}u_iv_j - \overline{u}\cdot\overline{v}}{\sigma_u\sigma_v} = \frac{1{,}11 - (-0{,}29)\cdot 0{,}07}{1{,}16\cdot 1{,}15} \approx 0{,}85.$$

Определим $\overline{x}$, $\overline{y}$:

$$\overline{x} = \overline{u}\cdot 10 + 65 = -0{,}29\cdot 10 + 65 = 62{,}1;$$
$$\overline{y} = \overline{v}\cdot 3 + 18 = 0{,}07\cdot 3 + 18 = 18{,}21.$$

Вычислим σ_x, σ_y:

$$\sigma_x = 10 \cdot \sigma_u \approx 10 \cdot 1{,}16 = 11{,}6;$$
$$\sigma_y = 3 \cdot \sigma_v \approx 3 \cdot 1{,}15 = 3{,}45.$$

Подставив все значения в выборочное уравнение линейной регрессии, получим

$$\overline{y}_x - 18{,}21 = 0{,}85 \cdot \frac{3{,}45}{11{,}6}(x - 62{,}1),$$

т. е.

$$\overline{y}_x = 0{,}25x + 2{,}56. \qquad \square$$

14.2. Уравнение параболической регрессии. Параболическая корреляция

Выборочным уравнением кривой линии регрессии Y на X (X и Y — наблюдаемые случайные величины) называют уравнение

$$\overline{y}_x = f(x),$$

где $f(x)$ — некоторая криволинейная функция. В этом случае корреляцию называют *криволинейной*.

Частным случаем выборочного уравнения кривой линии регрессии является *выборочное уравнение параболической регрессии* Y на X:

$$\overline{y}_x = Ax^2 + Bx + C.$$

Неизвестные параметры A, B, C уравнения параболической регрессии находят из системы уравнений:

$$\begin{cases} A\sum_{i=1}^{m} n_{x_i}x_i^4 + B\sum_{i=1}^{m} n_{x_i}x_i^3 + C\sum_{i=1}^{m} n_{x_i}x_i^2 = \sum_{i=1}^{m} n_{x_i}\overline{y}_{x_i}x_i^2, \\ A\sum_{i=1}^{m} n_{x_i}x_i^3 + B\sum_{i=1}^{m} n_{x_i}x_i^2 + C\sum_{i=1}^{m} n_{x_i}x_i = \sum_{i=1}^{m} n_{x_i}\overline{y}_{x_i}x_i, \\ A\sum_{i=1}^{m} n_{x_i}x_i^2 + B\sum_{i=1}^{m} n_{x_i}x_i + Cn = \sum_{i=1}^{m} n_{x_i}\overline{y}_{x_i}, \end{cases} \qquad (1)$$

где n_{x_i} — частота варианты x_i; $\overline{y}_{x_i} = \dfrac{\sum\limits_{j=1}^{k} n_{x_i y_j} y_j}{n_{x_i}}$ — условное среднее; k — количество наблюдаемых вариант случайной величины Y; $n_{x_i y_j}$ — частота варианты (x_i, y_j); n — объём выборки; m — количество наблюдаемых вариант случайной величины X.

В этом случае говорят о *параболической корреляции второго порядка*.

Для оценки силы корреляции Y на X находят *выборочное корреляционное отношение* — отношение межгруппового выборочного среднего квадратического отклонения к общему выборочному среднему квадратическому отклонению:

$$\eta_{yx} = \frac{\sigma_{\overline{y}_x}}{\sigma_y},$$

где

$$\sigma_{\overline{y}_x} = \sqrt{D_{\text{межгр}}} = \sqrt{\frac{\sum\limits_{i=1}^{n} n_{x_i} (\overline{y}_{x_i} - \overline{y})^2}{n}};$$

$$\sigma_y = \sqrt{D_{\text{в}}} = \sqrt{\frac{\sum\limits_{j=1}^{k} n_{y_j} (y_j - \overline{y})^2}{n}}.$$

Аналогично определяют уравнение параболической регрессии X на Y и выборочное корреляционное отношение η_{xy} X к Y.

Пример 1. Найти выборочное уравнение параболической регрессии

$$\overline{y}_x = Ax^2 + Bx + C$$

по данным корреляционной таблицы (табл. 10).

Оценить силу корреляционной связи с помощью выборочного корреляционного отношения.

Решение. Составим расчетную таблицу (табл. 11).

Таблица 10

X	Y					n_x
	2	3	5	7	8	
12	5	—	—	2	6	13
14	1	8	—	12	3	24
17	—	6	11	10	1	28
18	—	4	15	8	—	27
20	—	—	7	1	—	8
n_y	6	18	33	33	10	$n = 100$

Таблица 11

X	n_{x_i}	$\overline{y}_{x_i}$	$n_{x_i}x_i$	$n_{x_i}x_i^2$	$n_{x_i}x_i^3$	$n_{x_i}x_i^4$	$n_{x_i}\overline{y}_{x_i}$	$n_{x_i}\overline{y}_{x_i}x_i$	$n_{x_i}\overline{y}_{x_i}x_i^2$
12	13	5,538	156	1872	22464	269568	72	864	10368
14	24	5,583	336	4704	65856	921984	134	1876	26264
17	28	5,393	476	8092	137564	2338588	151	2567	43639
18	27	5,296	486	8748	157464	2834352	143	2574	46332
20	8	5,25	160	3200	64000	1280000	42	840	16800
$\sum$	100		1614	26616	447348	7644492	542	8721	143403

Подставив числа из последней строки табл. 11 в систему уравнений (1), получим систему уравнений относительно неизвестных коэффициентов A, B, C:

$$\begin{cases} 7644492\,A + 447348\,B + 26616\,C = 143403, \\ 447348\,A + 26616\,B + 1614\,C = 8721, \\ 26616\,A + 1614\,B + 100\,C = 542. \end{cases}$$

Решив эту систему уравнений, получим

$$A \approx -0{,}00434; \qquad B \approx 0{,}08868; \qquad C \approx 5{,}1434,$$

т. е. уравнение параболической регрессии $\overline{y}_x = Ax^2 + Bx + C$ имеет вид

$$\overline{y}_x = -0{,}00434x^2 + 0{,}08868x + 5{,}1434.$$

Для того чтобы найти выборочное корреляционное отношение η_{yx}, определим сначала выборочное среднее $\overline{y}$, выборочное среднее квадратическое отклонение σ_y и межгрупповое среднее квадратическое отклонение $\sigma_{\overline{y}_x}$:

$$\overline{y} = \frac{\sum\limits_{j=1}^{5} n_{y_j} y_j}{n} = \frac{2 \cdot 6 + 3 \cdot 18 + 5 \cdot 33 + 7 \cdot 33 + 8 \cdot 10}{100} = 5{,}42;$$

$$\sigma_y = \sqrt{\frac{\sum\limits_{j=1}^{k} n_{y_j} y_j^2}{n} - \overline{y}^2} = \\ = \sqrt{\frac{2^2 \cdot 6 + 3^2 \cdot 18 + 5^2 \cdot 33 + 7^2 \cdot 33 + 8^2 \cdot 10}{100} - 5{,}42^2} \approx 1{,}82;$$

$$\sigma_{\overline{y}_x} = \sqrt{\frac{\sum\limits_{i=1}^{n} n_{x_i} (\overline{y}_{x_i} - \overline{y})^2}{n}} = \left(\frac{13(5{,}54 - 5{,}42)^2 + 24(5{,}58 - 5{,}42)^2}{100} + \right. \\ \left. + \frac{28(5{,}39 - 5{,}42)^2 + 27(5{,}3 - 5{,}42)^2 + 8(5{,}25 - 5{,}42)^2}{100} \right)^{\frac{1}{2}} \approx \\ \approx 0{,}122.$$

Итак, искомое выборочное корреляционное отношение

$$\eta_{yx} = \frac{\sigma_{\overline{y}_x}}{\sigma_y} = \frac{0{,}122}{1{,}82} \approx 0{,}067. \qquad \square$$

Задачи к разделу 14

Задача 1. Найти выборочное уравнение прямой линии регрессии Y на X по данным корреляционной таблицы (табл. 12).

Ответ. $\overline{y}_x = -1{,}39x + 48{,}68$.

Задача 2. Найти выборочное уравнение параболической регрессии

$$\overline{y}_x = Ax^2 + Bx + C$$

по данным корреляционной таблицы (табл. 13).

Таблица 12

X	Y					n_x
	27	31	35	39	43	
5	—	—	—	3	5	8
7	—	—	2	8	4	14
9	—	3	18	7	—	28
11	1	6	11	3	—	21
13	10	13	6	—	—	29
n_y	11	22	37	21	9	$n = 100$

Таблица 13

X	Y					n_x
	12	15	17	18	20	
3	—	—	15	—	—	15
4	—	12	3	15	—	30
6	9	3	—	5	11	28
8	7	1	—	2	9	19
9	3	—	—	—	5	8
n_y	19	16	18	22	25	$n = 100$

Оценить силу корреляционной связи с помощью выборочного корреляционного отношения.

Ответ. $\overline{y}_x = 0{,}05x^2 - 0{,}62x + 18{,}39$; $\eta_{yx} \approx 0{,}0632$.

Раздел 15. Статистическая проверка статистических гипотез

Ключевые слова

Альтернативный
Второй
Граница
Гипотеза
Двухсторонний
Допустимый
Значение
Значимость
Конкурирующий
Критерий
Критический
Левосторонний
Наблюдаемый
Нулевой
Мощность
Область
Основной
Ошибка
Первый
Правосторонний
Принцип
Принятие
Проверка
Простой
Равенство
Род
Сложный
Статистический
Точка
Уровень
Эмпирический

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известных распределений.

Нулевой (основной) гипотезой H_0 называют выдвинутую гипотезу.

Альтернативной (конкурирующей) гипотезой H_1 называют противоположную к нулевой гипотезу.

Простая гипотеза состоит из одного предположения.

Сложная гипотеза состоит из конечного или бесконечного количества предположений.

Ошибка первого рода заключается в отклонении правильной нулевой гипотезы вследствие её проверки.

Уровень значимости α — это вероятность ошибки первого рода.

Ошибка второго рода заключается в принятии неправильной нулевой гипотезы вследствие её проверки. Вероятность ошибки второго рода обозначают β.

Статистическим критерием (или просто *критерием*) называют случайную величину K, которая используется для проверки нулевой гипотезы.

Наблюдаемым (эмпирическим) значением $K_{\text{набл}}$ является значение критерия, вычисленное по данным выборки.

Критическая область — это совокупность значений критерия, при которых отклоняют нулевую гипотезу.

Областью принятия гипотезы (областью допустимых значений) является совокупность значений критерия, при которых принимают нулевую гипотезу.

Основной принцип проверки статистических гипотез такой: если наблюдаемое значение критерия принадлежит к критической области, нулевую гипотезу отклоняют; если наблюдаемое значение критерия принадлежит к области принятия гипотезы, гипотезу принимают.

Критическими точками (границами) $k_{\text{кр}}$ называют такие, которые отделяют критическую область от области принятия гипотезы.

Правосторонняя критическая область определяется неравенством

$$K > k_{\text{кр}}.$$

Левосторонняя критическая область определяется неравенством

$$K < k_{\text{кр}}.$$

Двухсторонняя критическая область определяется неравенством

$$K < k_1, \qquad K > k_2,$$

где $k_1 < k_2$. В частности, если критические точки симметричны относительно нуля, то двухсторонняя область определяется неравенствами

$$K < -k_{\text{кр}}, \quad K > k_{\text{кр}} \qquad (\text{считается, что } k_{\text{кр}} > 0)$$

или равносильным неравенством

$$|K| > k_{\text{кр}}.$$

Для получения критической области задают уровень значимости α и находят критические точки из таких соотношений:

а) для правосторонней критической области

$$\mathbf{P}(K > k_{\text{кр}}) = \alpha \qquad (k_{\text{кр}} > 0);$$

б) для левосторонней критической области

$$\mathbf{P}(K < k_{\text{кр}}) = \alpha \qquad (k_{\text{кр}} < 0);$$

в) для двухсторонней симметричной критической области

$$\mathbf{P}(K < -k_{\text{кр}}) = \frac{\alpha}{2}, \quad \mathbf{P}(K > k_{\text{кр}}) = \frac{\alpha}{2} \qquad (k_{\text{кр}} > 0).$$

Мощностью критерия называют вероятность попадания критерия в критическую область при условии, что правильной является конкурирующая гипотеза. Другими словами, мощность критерия — это вероятность того, что нулевая гипотеза будет отклонена, если верна конкурирующая гипотеза.

15.1. Проверка равенства выборочного среднего гипотетическому генеральному среднему

Пусть из нормальной генеральной совокупности с известной дисперсией σ^2 получена выборка объёма n.

Правило 1. Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0\colon a = a_0$ о равенстве генерального среднего a нормальной совокупности с известной дисперсией σ^2 гипотетическому (предполагаемому) значению a_0 при конкурирующей гипотезе $H_1\colon a \neq a_0$, нужно вычислить наблюдаемое значение критерия

$$U_{\text{набл}} = \frac{(\overline{x}_{\text{в}} - a_0)\sqrt{n}}{\sigma}$$

и по таблице функции Лапласа (прил. 1) найти критическую точку $u_{\text{кр}}$ двухсторонней критической области из равенства

$$\Phi(u_{\text{кр}}) = \frac{1-\alpha}{2}.$$

Если $|U_{\text{набл}}| < u_{\text{кр}}$, нет оснований отклонять нулевую гипотезу. Если $|U_{\text{набл}}| > u_{\text{кр}}$, нулевую гипотезу отклоняют.

Правило 2. При конкурирующей гипотезе $H_1\colon a > a_0$ критическую точку правосторонней критической области находят из равенства

$$\Phi(u_{\text{кр}}) = \frac{1}{2} - \alpha.$$

Если $U_{\text{набл}} < u_{\text{кр}}$, нет оснований отклонять нулевую гипотезу. Если $U_{\text{набл}} > u_{\text{кр}}$, нулевую гипотезу отклоняют.

Правило 3. При конкурирующей гипотезе $H_1\colon a < a_0$ сначала находят «вспомогательную» критическую точку $u_{\text{кр}}$ по правилу 2, а потом считают границей левосторонней критической области

$$u'_{\text{кр}} = -u_{\text{кр}}.$$

Если $U_{\text{набл}} > -u_{\text{кр}}$, нет оснований отклонять нулевую гипотезу. Если $U_{\text{набл}} < -u_{\text{кр}}$, нулевую гипотезу отклоняют.

Мощность критерия проверки основной гипотезы $H_0\colon a = a_0$ о равенстве генерального среднего a гипотетическому значению a_0 при известном среднем квадратическом отклонении σ находят в зависимости от вида альтернативной гипотезы.

При альтернативной гипотезе $H_1\colon a > a_0$ для гипотетического значения генерального среднего $a = a_1 > a_0$ мощность правостороннего критерия

$$1 - \beta = \frac{1}{2} - \Phi(u_{\text{кр}} - \lambda),$$

где $u_{\text{кр}}$ находят из равенства $\Phi(u_{\text{кр}}) = \dfrac{1}{2} - \alpha$, $\lambda = \dfrac{(a_1 - a_0)\sqrt{n}}{\sigma}$.

При разных значениях a_1 функция мощности одностороннего критерия

$$\pi_1(a_1) = \frac{1}{2} - \Phi(u_{\text{кр}} - \lambda).$$

При альтернативной гипотезе $H_1\colon a \neq a_0$ для гипотетического значения генерального среднего $a = a_1$ мощность двухстороннего критерия

$$1 - \beta = 1 - \big[\Phi(u_{\text{кр}} - \lambda) + \Phi(u_{\text{кр}} + \lambda)\big],$$

где $u_{\text{кр}}$ находят из равенства $\Phi(u_{\text{кр}}) = \dfrac{1-\alpha}{2}$, $\lambda = \dfrac{(a_1 - a_0)\sqrt{n}}{\sigma}$.

При разных значениях a_1 функция мощности одностороннего критерия

$$\pi_1(a_1) = 1 - \big[\Phi(u_{\text{кр}} - \lambda) + \Phi(u_{\text{кр}} + \lambda)\big].$$

Пример 1. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 4$ получена выборка объёма $n = 100$, по которой найдено выборочное среднее $\overline{x}_{\text{в}} = 29$. Необходимо при уровне значимости 0,01 проверить нулевую гипотезу $H_0\colon a = a_0 = 30$ при конкурирующей гипотезе $H_1\colon a \neq 30$.

Решение. Найдём сначала наблюдаемое значение критерия

$$U_{\text{набл}} = \frac{(\overline{x}_{\text{в}} - a_0)\sqrt{n}}{\sigma} = \frac{(29 - 30)\sqrt{100}}{4} = -2{,}5.$$

Поскольку по условию конкурирующая гипотеза имеет вид $H_1\colon a \neq a_0$, критическая область — двухсторонняя.

Вычислим критическую точку из равенства

$$\Phi(u_{\text{кр}}) = \frac{1-\alpha}{2} = \frac{1-0{,}01}{2} = 0{,}495.$$

По таблице функции Лапласа (прил. 1) находим критическую точку:

$$u_{\text{кр}} \approx 2{,}58.$$

Поскольку $|U_{\text{набл}}| < u_{\text{кр}}$, оснований отклонять основную гипотезу нет. Другими словами, выборочное и гипотетическое генеральное средние различаются незначаще. □

Пример 2. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 4{,}8$ получена выборка объёма $n = 144$, по которой найдено выборочное среднее $\overline{x}_{\text{в}} = 16$.

Необходимо при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : a = a_0 = 15$ при конкурирующей гипотезе:

а) $H_1 : a \neq 15$;

б) $H_1 : a > 15$;

в) $H_1 : a < 15$.

Кроме того, необходимо найти мощности правостороннего и двухстороннего критериев.

Решение. Сначала вычислим наблюдаемое значение критерия

$$U_{\text{набл}} = \frac{(\overline{x}_{\text{в}} - a_0)\sqrt{n}}{\sigma} = \frac{(16 - 15)\sqrt{144}}{4{,}8} = 2{,}5.$$

а) Воспользуемся правилом 1. Найдём критическую точку из равенства

$$\Phi(u_{\text{кр}}) = \frac{1 - \alpha}{2} = \frac{1 - 0{,}05}{2} = 0{,}475.$$

По таблице функции Лапласа (прил. 1) определим критическую точку:

$$u_{\text{кр}} \approx 1{,}96.$$

Поскольку $|U_{\text{набл}}| > u_{\text{кр}}$, основную гипотезу отклоняем. Другими словами, выборочное и гипотетическое генеральное средние различаются значаще.

б) Воспользуемся правилом 2. Найдём критическую точку из равенства

$$\Phi(u_{\text{кр}}) = \frac{1}{2} - \alpha = \frac{1}{2} - 0{,}05 = 0{,}45.$$

По таблице функции Лапласа (прил. 1) определим критическую точку:

$$u_{\text{кр}} \approx 1{,}64.$$

Поскольку $U_{\text{набл}} > u_{\text{кр}}$, основную гипотезу отклоняем. Другими словами, выборочное и гипотетическое генеральное средние различаются значаще.

в) Воспользуемся правилом 3. Критическая точка будет такой же, как и в пункте б), но с противоположным знаком:

$$u_{\text{кр}} \approx -1{,}64.$$

Поскольку $U_{\text{набл}} > u_{\text{кр}}$, оснований отклонять основную гипотезу нет. Другими словами, выборочное и гипотетическое генеральное средние различаются незначаще.

Теперь найдём мощности правостороннего и двухстороннего критериев. Напомним, что критические точки в этих случаях разные и равняются соответственно 1,64 и 1,96.

Определим параметр λ, который входит в оба уравнения для определения мощности критериев:

$$\lambda = \frac{(a_1 - a_0)\sqrt{n}}{\sigma} = \frac{(16 - 15)\sqrt{144}}{4{,}8} = 2{,}5.$$

Итак, мощности соответственно правостороннего и двухстороннего критериев будут такие:

$$\begin{aligned}
\pi_1^{\text{прав}}(16) &= \frac{1}{2} - \Phi(u_{\text{кр}} - \lambda) = \frac{1}{2} - \Phi(1{,}64 - 2{,}5) = \\
&= \frac{1}{2} + \Phi(0{,}86) = \frac{1}{2} + 0{,}3051 = 0{,}8051; \\
\pi_1^{\text{двухст}}(16) &= 1 - \big[\Phi(u_{\text{кр}} - \lambda) + \Phi(u_{\text{кр}} + \lambda)\big] = \\
&= 1 - \big[\Phi(1{,}96 - 2{,}5) + \Phi(1{,}96 + 2{,}5)\big] = \\
&= 1 - \big[-\Phi(0{,}54) + \Phi(4{,}46)\big] \approx \\
&\approx 1 + 0{,}2054 - 0{,}5 = 0{,}7054.
\end{aligned}$$

Другими словами, вероятности того, что нулевая гипотеза будет отклонена, если правильной будет конкурирующая гипотеза, равняются 0,8051 и 0,7054 соответственно для правостороннего и двухстороннего критериев. □

Если дисперсия генеральной совокупности неизвестна, в качестве критерия проверки основной гипотезы берут случайную величину

$$T = \frac{(\overline{x} - a_0)\sqrt{n}}{s},$$

где $s = \sqrt{\dfrac{n\sum\limits_i n_i x_i 2 - \Big(\sum\limits_i n_i x_i\Big)2}{n(n-1)}}$ — «исправленное» выборочное среднее квадратическое отклонение. Величина T имеет распределение Стьюдента с $k = n - 1$ степенями свободы.

Правило 1. Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0\colon a = a_0$ о равенстве неизвестного генерального среднего a нормальной совокупности с неизвестной дисперсией гипотетическому (предполагаемому) значению a_0 при конкурирующей гипотезе $H_1\colon a \neq a_0$, необходимо вычислить наблюдаемое значение критерия

$$T_{\text{набл}} = \frac{(\overline{x}_{\text{в}} - a_0)\sqrt{n}}{s}$$

и по таблице критических точек распределения Стьюдента (прил. 6) при заданном уровне значимости α, размещённом в верхней части таблицы, и числе степеней свободы $k = n - 1$ найти критическую точку $t_{\text{двухст. кр}}(\alpha; k)$.

Если $|T_{\text{набл}}| < t_{\text{двухст. кр}}$, нет оснований отклонять нулевую гипотезу. Если $|T_{\text{набл}}| > t_{\text{двухст. кр}}$, нулевую гипотезу отклоняют.

Правило 2. При конкурирующей гипотезе $H_1\colon a > a_0$ при уровне значимости α, размещённом в нижней части таблицы прил. 6, и числе степеней свободы $k = n - 1$ находят критическую точку правосторонней критической области $t_{\text{прав. кр}}(\alpha; k)$.

Если $T_{\text{набл}} < t_{\text{прав. кр}}$, нет оснований отклонять нулевую гипотезу. Если $T_{\text{набл}} > t_{\text{прав. кр}}$, нулевую гипотезу отклоняют.

Правило 3. При конкурирующей гипотезе $H_1\colon a < a_0$ сначала находят так называемую «вспомогательную» критическую точку (по правилу 2) $t_{\text{прав. кр}}(\alpha; k)$ и считают границей левосторонней критической области $t_{\text{лев. кр}} = -t_{\text{прав. кр}}$.

Если $T_{\text{набл}} > -t_{\text{прав. кр}}$, нет оснований отклонять нулевую гипотезу. Если $T_{\text{набл}} < -t_{\text{прав. кр}}$, нулевую гипотезу отклоняют.

Пример 3. Для выборки объёма $n = 25$, полученной из нормальной генеральной совокупности, найдено выборочное среднее $\overline{x} = 43$ и «исправленное» выборочное среднее квадратическое отклонение $s = 4$. Необходимо при уровне значимости $\alpha = 0{,}01$ проверить основную гипотезу $H_0\colon a = a_0 = 45$ при альтернативной гипотезе $H_1\colon a \neq 45$.

Решение. Вычислим сначала наблюдаемое значение критерия

$$T_{\text{набл}} = \frac{(\overline{x} - a_0)\sqrt{n}}{s} = \frac{(43 - 45)\sqrt{25}}{4} = 2{,}5.$$

Поскольку альтернативная гипотеза имеет вид $H_1\colon a \neq a_0$, то критическая область — двухсторонняя.

По таблице критических точек распределения Стьюдента (прил. 6) при уровне значимости $\alpha = 0{,}01$ и числе степеней свободы

$$k = n - 1 = 25 - 1 = 24$$

находим критическую точку

$$t_{\text{двухст. кр}}(0{,}01; 24) = 2{,}80.$$

Поскольку $|T_{\text{набл}}| < t_{\text{двухст. кр}}$, оснований отклонять основную гипотезу нет. Другими словами, выборочное и гипотетическое генеральное средние различаются незначаще. □

Пример 4. Контрольная масса буханки хлеба, которая выпекается на заводе, должны составлять $a = a_0 = 1250$ г. Контрольные измерения 20 случайно отобранных буханок хлеба дали такие результаты:

Масса, г	1240	1245	1250	1255	1260
Количество буханок	3	5	6	4	2

Необходимо при уровне значимости 0,001 проверить основную гипотезу $H_0\colon a = a_0 = 1250$ при конкурирующей гипотезе:

а) $H_1 : a \neq a_0$;

б) $H_1 : a > a_0$;

в) $H_1 : a < a_0$.

Решение. Вычислим сначала выборочные среднее $\overline{x}$ и «исправленное» среднее квадратическое отклонение s:

$$\overline{x} = \frac{1240 \cdot 3 + 1245 \cdot 5 + 1250 \cdot 6 + 1255 \cdot 4 + 1260 \cdot 2}{20} = 1249{,}25;$$

$$s = \left(\frac{1240^2 \cdot 3 + 1245^2 \cdot 5 + 1250^2 \cdot 6 + 1255^2 \cdot 4 + 1260^2 \cdot 2}{19} - \right.$$

$$\left. - \frac{20}{19} \cdot 1249{,}25\right)^{\frac{1}{2}} \approx 6{,}13.$$

Найдём наблюдаемое значение критерия:

$$T_{\text{набл}} = \frac{(\overline{x} - a_0)\sqrt{n}}{s} = \frac{(1249{,}25 - 1250)\sqrt{20}}{6{,}13} \approx -0{,}547.$$

а) Воспользуемся правилом 1. По прил. 6 при уровне значимости $\alpha = 0{,}001$ и числе степеней свободы

$$k = n - 1 = 20 - 1 = 19$$

найдём критическую точку распределения Стьюдента:

$$t_{\text{двухст. кр}}(0{,}001; 19) = 3{,}88.$$

Поскольку $|T_{\text{набл}}| < t_{\text{двухст. кр}}$, оснований отклонять основную гипотезу нет. Другими словами, выборочное и гипотетическое генеральное средние различаются незначаще.

б) Воспользуемся правилом 2. По прил. 6 при уровне значимости $\alpha = 0{,}001$ и числе степеней свободы

$$k = n - 1 = 20 - 1 = 19$$

найдём критическую точку распределения Стьюдента:

$$t_{\text{прав. кр}}(0{,}001; 19) = 3{,}58.$$

Поскольку $T_{\text{набл}} < t_{\text{прав. кр}}$, оснований отклонять основную гипотезу нет. Другими словами, выборочное и гипотетическое генеральное средние различаются незначаще.

в) Воспользуемся правилом 3. Критическая точка будет такой же, как и в пункте б), но с противоположным знаком:

$$t_{\text{лев. кр}}(0{,}001; 19) = -3{,}88.$$

Поскольку $T_{\text{набл}} > t_{\text{лев. кр}}$, оснований отклонять основную гипотезу нет. Другими словами, выборочное и гипотетическое генеральное средние различаются незначаще. □

15.2. Проверка равенства «исправленной» выборочной дисперсии генеральной дисперсии

Пусть из нормальной генеральной совокупности получена выборка объёма n, для которой найдена «исправленная» выборочная дисперсия s^2.

Правило 1. Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0\colon \sigma^2 = \sigma_0^2$ о равенстве неизвестной генеральной дисперсии σ^2 гипотетическому (предполагаемому) значению σ_0^2 при конкурирующей гипотезе $H_1\colon \sigma^2 > \sigma_0^2$, необходимо вычислить наблюдаемое значение критерия

$$\chi^2_{\text{набл}} = \frac{(n-1)s^2}{\sigma_0^2}$$

и по таблице критических точек распределения χ^2 (прил. 7), при заданном уровне значимости α и числе степеней свободы $k = n - 1$ найти критическую точку $\chi^2_{\text{кр}}(\alpha; k)$.

Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, нет оснований отклонять нулевую гипотезу. Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, нулевую гипотезу отклоняют.

Правило 2. При конкурирующей гипотезе $H_1\colon \sigma^2 \neq \sigma_0^2$ находят левую $\chi^2_{\text{лев. кр}}(1-\alpha/2; k)$ и правую $\chi^2_{\text{прав. кр}}(\alpha/2; k)$ критические точки.

Если $\chi^2_{\text{лев. кр}} < \chi^2_{\text{набл}} < \chi^2_{\text{прав. кр}}$, нет оснований отклонять нулевую гипотезу. Если $\chi^2_{\text{набл}} < \chi^2_{\text{лев. кр}}$ или $\chi^2_{\text{набл}} > \chi^2_{\text{прав. кр}}$, нулевую гипотезу отклоняют.

Правило 3. При конкурирующей гипотезе $H_1: \sigma^2 < \sigma_0^2$ находят критическую точку $\chi^2_{\text{кр}}(1-\alpha; k)$.

Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}(1-\alpha; k)$, нет оснований отклонять нулевую гипотезу. Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}(1-\alpha; k)$, нулевую гипотезу отклоняют.

Замечание 1. Если число степеней свободы $k > 30$, критическую точку $\chi^2_{\text{кр}}(\alpha; k)$ можно найти с помощью равенства Уилсона—Гильферти:

$$\chi^2_{\text{кр}}(\alpha; k) = k\left(1 - \frac{2}{9k} + z_\alpha\sqrt{\frac{2}{9k}}\right)^3,$$

где z_α определяют, используя таблицу интегральной функции Лапласа (прил. 1), из равенства

$$\Phi(z_\alpha) = \frac{1}{2} - \alpha.$$

Пример 1. Из нормальной генеральной совокупности получена выборка объёма $n = 25$, для которой найдена выборочная «исправленная» дисперсия $s^2 = 12{,}3$. Проверить при уровне значимости $\alpha = 0{,}01$ основную гипотезу $H_0: \sigma^2 = \sigma_0^2 = 12$, если альтернативная гипотеза $H_1: \sigma_0^2 > 12$.

Решение. Найдём наблюдаемое значение критерия:

$$\chi^2_{\text{набл}} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(25-1)\cdot 12{,}3}{12} = 24{,}6.$$

Поскольку альтернативная гипотеза имеет вид $H_1: \sigma_0^2 > 12$, критическая область правосторонняя (правило 1). По прил. 7 при уровне значимости $\alpha = 0{,}01$ и числе степеней свободы

$$k = n - 1 = 25 - 1 = 24$$

находим критическую точку:

$$\chi^2_{\text{кр}}(0{,}01; 24) = 42{,}98.$$

Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, нет оснований отклонять основную гипотезу о равенстве генеральной дисперсии гипотетическому значению

$\sigma_0^2 \leqslant 12$. Другими словами, разница между «исправленной» выборочной дисперсией $s^2 = 12{,}3$ и гипотетической генеральной дисперсией $\sigma_0^2 \leqslant 12$ незначащая. □

Пример 2. Из нормальной генеральной совокупности получена выборка объёма $n = 20$, для которой найдена выборочная «исправленная» дисперсия $s^2 = 2{,}7$. Проверить при уровне значимости $\alpha = 0{,}05$ основную гипотезу $H_0\colon \sigma^2 = \sigma_0^2 = 3$, если альтернативная гипотеза:

а) $H_1\colon \sigma_0^2 > 3$;
б) $H_1\colon \sigma_0^2 \neq 3$;
в) $H_1\colon \sigma_0^2 < 3$.

Решение. Найдём наблюдаемое значение критерия:

$$\chi^2_{\text{набл}} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(20-1)\cdot 2{,}7}{3} = 17{,}1.$$

а) Воспользуемся правилом 1. По прил. 7 при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы

$$k = n - 1 = 20 - 1 = 19$$

находим критическую точку

$$\chi^2_{\text{кр}}(0{,}05; 19) = 30{,}14.$$

Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, нет оснований отклонять основную гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2 \leqslant 3$. Другими словами, разница между «исправленной» выборочной дисперсией $s^2 = 2{,}7$ и гипотетической генеральной дисперсией $\sigma_0^2 \leqslant 3$ незначащая.

б) Воспользуемся правилом 2. По прил. 7 при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы

$$k = n - 1 = 20 - 1 = 19$$

находим критические точки:

$$\chi^2_{\text{лев. кр}}(1 - \alpha/2; k) = \chi^2_{\text{лев. кр}}(0{,}975; 19) = 8{,}91;$$
$$\chi^2_{\text{прав. кр}}(\alpha/2; k) = \chi^2_{\text{прав. кр}}(0{,}025; 19) = 32{,}85.$$

Поскольку $\chi^2_{\text{лев. кр}} < \chi^2_{\text{набл}} < \chi^2_{\text{прав. кр}}$, нет оснований отклонять основную гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2 = 3$. Другими словами, разница между «исправленной» выборочной дисперсией $s^2 = 2{,}7$ и гипотетической генеральной дисперсией $\sigma_0^2 = 3$ незначащая.

в) Воспользуемся правилом 3. По прил. 7 при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы

$$k = n - 1 = 20 - 1 = 19$$

находим критическую точку:

$$\chi^2_{\text{кр}}(1 - \alpha; k) = \chi^2_{\text{кр}}(0{,}95; 19) = 10{,}12.$$

Поскольку $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}(1 - \alpha; k)$, нет оснований отклонять основную гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2 \geqslant 3$. Другими словами, разница между «исправленной» выборочной дисперсией $s^2 = 2{,}7$ и гипотетической генеральной дисперсией $\sigma_0^2 \geqslant 3$ незначащая. □

Пример 3. Из нормальной генеральной совокупности получена выборка объёма $n = 100$:

x_i	5	6	7	8	9	10	11
n_i	9	15	19	20	17	13	7

.

Необходимо проверить при уровне значимости $\alpha = 0{,}01$ основную гипотезу $H_0\colon \sigma^2 = \sigma_0^2 = 4$, если альтернативная гипотеза $H_1\colon \sigma_0^2 > 4$.

Решение. Сначала по данным выборки найдём «исправленную» выборочную дисперсию s^2:

$$\overline{x} = \frac{5 \cdot 9 + 6 \cdot 15 + 7 \cdot 19 + 8 \cdot 20 + 9 \cdot 17 + 10 \cdot 13 + 11 \cdot 7}{100} = 7{,}88;$$

$$s^2 = \frac{5^2 \cdot 9 + 6^2 \cdot 15 + 7^2 \cdot 19 + 8^2 \cdot 20 + 9^2 \cdot 17 + 10^2 \cdot 132 + 11^2 \cdot 7}{100} -$$

$$- 7{,}88^2 = 2{,}9056.$$

Вычислим наблюдаемое значение критерия:

$$\chi^2_{\text{набл}} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(100-1) \cdot 2{,}9056}{4} = 71{,}9136.$$

Поскольку альтернативная гипотеза имеет вид $H_1\colon \sigma_0^2 > 12$, критическая область правосторонняя (правило 1).

Найдём число степеней свободы k:

$$k = n - 1 = 100 - 1 = 99.$$

Поскольку число степеней свободы $k > 30$, критическую точку $\chi^2_{\text{кр}}(\alpha; k)$ найдём из равенства Уилсона—Гильферти:

$$\chi^2_{\text{кр}}(\alpha; k) = k\left(1 - \frac{2}{9k} + z_\alpha\sqrt{\frac{2}{9k}}\right)^3,$$

где z_α определим из равенства

$$\Phi(z_\alpha) = \frac{1}{2} - \alpha = \frac{1}{2} - 0{,}01 = 0{,}49,$$

используя таблицу функции Лапласа (прил. 1):

$$z_\alpha \approx 2{,}33.$$

Итак,

$$\chi^2_{\text{кр}}(0{,}01; 99) = 99\left(1 - \frac{2}{9\cdot 99} + 2{,}33\sqrt{\frac{2}{9\cdot 99}}\right)^3 = 113{,}0187.$$

Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, нет оснований отклонять основную гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2 \leqslant 4$. Другими словами, разница между «исправленной» выборочной дисперсией $s^2 \approx 2{,}9$ и гипотетической генеральной дисперсией $\sigma_0^2 \leqslant 4$ незначащая. □

15.3. Проверка равенства относительной частоты гипотетической вероятности

Пусть при достаточно большом количестве n независимых испытаний, в каждом из которых вероятность p появления события постоянна, но неизвестна, найдена относительная частота $\dfrac{m}{n}$. Необходимо при заданном уровне значимости α проверить основную гипотезу, которая заключается в том, что неизвестная вероятность p равняется гипотетической вероятности p_0.

Правило 1. Для того чтобы при заданном уровне значимости α проверить основную гипотезу $H_0\colon p = p_0$ о равенстве неизвестной вероятности p гипотетической вероятности p_0 при конкурирующей гипотезе $H_1\colon p \neq p_0$, необходимо вычислить наблюдаемое значение критерия

$$U_{\text{набл}} = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}}$$

и по таблице функции Лапласа (прил. 1) найти критическую точку $u_{\text{кр}}$ из равенства

$$\Phi(u_{\text{кр}}) = \frac{1-\alpha}{2}.$$

Если $|U_{\text{набл}}| < u_{\text{кр}}$, нет оснований отклонять основную гипотезу. Если $|U_{\text{набл}}| > u_{\text{кр}}$, основную гипотезу отклоняют.

Правило 2. При альтернативной гипотезе $H_1\colon p > p_0$ критическую точку правосторонней критической области находят из равенства

$$\Phi(u_{\text{кр}}) = \frac{1}{2} - \alpha.$$

Если $U_{\text{набл}} < u_{\text{кр}}$, нет оснований отклонять основную гипотезу. Если $U_{\text{набл}} > u_{\text{кр}}$, основную гипотезу отклоняют.

Правило 3. При альтернативной гипотезе $H_1\colon p < p_0$ сначала находят «вспомогательную» критическую точку $u_{\text{кр}}$ по правилу 2, а потом считают границей левосторонней критической области

$$u'_{\text{кр}} = -u_{\text{кр}}.$$

Если $U_{\text{набл}} > -u_{\text{кр}}$, нет оснований отклонять основную гипотезу. Если $U_{\text{набл}} < -u_{\text{кр}}$, основную гипотезу отклоняют.

Замечание 1. Удовлетворительные результаты получаются в случае выполнения неравенства $np_0q_0 > 9$.

Пример 1. Для некоторого события в результате 100 независимых испытаний была найдена относительная частота $\dfrac{m}{n} = 0{,}34$. При уровне значимости 0,05 необходимо проверить основную гипотезу $H_0\colon p = p_0 = 0{,}3$ при конкурирующей гипотезе $H_1\colon p \neq 0{,}3$.

Решение. Сначала найдём q_0:

$$q_0 = 1 - p_0 = 1 - 0{,}3 = 0{,}7,$$

т. е.

$$np_0q_0 = 100 \cdot 0{,}3 \cdot 0{,}7 = 21 > 9.$$

Вычислим наблюдаемое значение критерия:

$$U_{\text{набл}} = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0q_0}} = \frac{(0{,}34 - 0{,}3)\sqrt{100}}{\sqrt{0{,}3 \cdot 0{,}7}} \approx 0{,}87.$$

Поскольку по условию альтернативная гипотеза имеет вид $H_1\colon p \neq p_0$, критическая область двухсторонняя (правило 1).

Определим критическую точку $u_{\text{кр}}$ из равенства

$$\Phi(u_{\text{кр}}) = \frac{1-\alpha}{2} = \frac{1-0{,}05}{2} = 0{,}475.$$

По прил. 1 находим

$$u_{\text{кр}} \approx 1{,}96.$$

Поскольку $|U_{\text{набл}}| < u_{\text{кр}}$, нет оснований отклонять основную гипотезу. Другими словами, наблюдаемая относительная частота $\dfrac{m}{n} = 0{,}34$ незначаще отличается от гипотетической вероятности $p_0 = 0{,}3$. □

Пример 2. Для некоторого события в результате 400 независимых испытаний была найдена относительная частота $\dfrac{m}{n} = 0{,}44$. При уровне значимости 0,01 необходимо проверить основную гипотезу $H_0\colon p = p_0 = 0{,}5$ при конкурирующей гипотезе:

а) $H_1\colon p > 0{,}5$;

б) $H_1\colon p < 0{,}5$.

Решение. Сначала найдём q_0:

$$q_0 = 1 - p_0 = 1 - 0{,}5 = 0{,}5,$$

т. е.

$$np_0q_0 = 400 \cdot 0{,}5 \cdot 0{,}5 = 100 > 9.$$

Вычислим наблюдаемое значение критерия:

$$U_{\text{набл}} = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0q_0}} = \frac{(0{,}44 - 0{,}5)\sqrt{400}}{\sqrt{0{,}5 \cdot 0{,}5}} \approx -2{,}4.$$

а) Применим правило 2. Определим критическую точку $u_{\text{кр}}$ из равенства

$$\Phi(u_{\text{кр}}) = \frac{1}{2} - \alpha = \frac{1}{2} - 0{,}01 = 0{,}49.$$

По прил. 1 находим

$$u_{\text{кр}} \approx 2{,}33.$$

Поскольку $U_{\text{набл}} < u_{\text{кр}}$, нет оснований отклонять основную гипотезу. Другими словами, наблюдаемая относительная частота $\dfrac{m}{n} = 0{,}47$ незначаще отличается от гипотетической вероятности $p_0 \leqslant 0{,}5$.

б) Применим правило 3. Используя данные пункта а), находим критическую точку

$$u_{\text{кр}} \approx -2{,}33.$$

Поскольку $U_{\text{набл}} < u_{\text{кр}}$, основная гипотеза отклоняется. Другими словами, наблюдаемая относительная частота $\dfrac{m}{n} = 0{,}47$ значаще отличается от гипотетической вероятности $p_0 \geqslant 0{,}5$. □

15.4. Проверка гипотезы о нормальном распределении по критерию Пирсона

Пусть статистическое распределение выборки задано в виде последовательности равноотстоящих вариант и соответствующих им частот:

x_i	x_1	x_2	$\cdots$	x_N
n_i	n_1	n_2	$\cdots$	n_N

.

Необходимо с помощью *критерия Пирсона* проверить гипотезу о нормальном распределении генеральной совокупности.

Правило 1. Для того чтобы при заданном уровне значимости α проверить гипотезу о нормальном распределении генеральной совокупности, необходимо:

1) вычислить выборочное среднее $\overline{x}_{\text{в}}$ и выборочное среднее квадратическое отклонение $\sigma_{\text{в}}$;

2) определить теоретические частоты

$$n'_i = \frac{nh}{\sigma_{\text{в}}} \cdot \varphi(u_i),$$

где n — объём выборки; h — шаг (разность между двумя соседними вариантами); $\varphi(u)$ — дифференциальная функция Лапласа (прил. 2);

$$u_i = \frac{x_i - \overline{x}_{\text{в}}}{\sigma_{\text{в}}};$$

3) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона. Для этого:

а) составляют расчётную таблицу (табл. 15), по которой находят наблюдаемое значение критерия Пирсона

$$\chi^2_{\text{набл}} = \sum_i \frac{(n_i - n'_i)^2}{n'_i};$$

б) по таблице критических точек распределения χ^2 (прил. 7) при заданном уровне значимости α и числе степеней свободы $k = s - 3$ (s — количество вариант выборки) находят критическую точку $\chi^2_{\text{кр}}(\alpha; k)$ правосторонней критической области.

Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, нет оснований отклонять гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначаще (случайно). Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, гипотезу отклоняют. Другими словами, эмпирические и теоретические частоты различаются значаще.

Замечание 1. Малочисленные частоты выборки ($n_i < 5$) следует объединять; в этом случае соответствующие им теоретические частоты также нужно сложить. Если происходило объединение частот, то при определении числа степеней свободы по формуле $k = s - 3$ следует в качестве s взять количество вариант выборки, которые остались после объединения частот.

Пример 1. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о нормальном распределении генеральной совокупности X, если из этой совокупности получена такая выборка объёма $n = 100$:

x_i	-5	-3	-1	1	3	5	7	9
n_i	5	9	13	18	21	18	10	6

Решение. Сначала вычислим выборочное среднее $\overline{x}_{\text{в}}$ и среднее квадратическое отклонение выборки $\sigma_{\text{в}}$:

$$\overline{x}_{\text{в}} = \frac{-5\cdot 5 - 3\cdot 9 - 1\cdot 13 + 1\cdot 18 + 3\cdot 21 + 5\cdot 18 + 7\cdot 10 + 9\cdot 6}{100} = $$

$$= 2{,}3;$$

$$\sigma_{\text{в}} = \Big(\big[5^2\cdot 5 + 3^2\cdot 9 + 1\cdot 13 + 1\cdot 18 + 3^2\cdot 21 + 5^2\cdot 18 + 7^2\cdot 10 +$$

$$+ 9^2\cdot 6\big]/100 - 2{,}32\Big)^{\frac{1}{2}} \approx 3{,}64.$$

Определим теоретические частоты, учитывая, что $n = 100$, $h = 2$, $\sigma_{\text{в}} = 3{,}64$, по формуле

$$n'_i = \frac{nh}{\sigma_{\text{в}}}\cdot\varphi\left(\frac{x_i - \overline{x}_{\text{в}}}{\sigma_{\text{в}}}\right).$$

Для этого составим расчётную табл. 14 (значение дифференциальной функции Лапласа $\varphi(u)$ приведено в прил. 2).

Сравним эмпирические и теоретические частоты. Для этого вычислим наблюдаемое значение критерия

$$\chi^2_{\text{набл}} = \sum_i \frac{(n_i - n'_i)^2}{n'_i}$$

с помощью расчётной табл. 15.

Таблица 14

i	x_i	$u_i = \dfrac{x_i - \overline{x}_в}{\sigma_в}$	$\varphi(u_i)$	$n_i' = \dfrac{nh}{\sigma_в} \cdot \varphi(u_i)$
1	−5	−2,01	0,0529	2,91
2	−3	−1,46	0,1374	7,56
3	−1	−0,91	0,2637	14,50
4	1	−0,36	0,3739	20,56
5	3	0,19	0,3918	21,54
6	5	0,74	0,3034	16,68
7	7	1,29	0,1736	9,55
8	9	1,84	0,0734	4,04

Таблица 15

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\dfrac{(n_i - n_i')^2}{n_i'}$
1	5	2,91	2,09	4,37	1,50
2	9	7,56	1,44	2,07	0,27
3	13	14,50	−1,50	2,25	0,16
4	18	20,56	−2,56	6,55	0,32
5	21	21,54	−0,54	0,29	0,01
6	18	16,68	1,32	1,74	0,10
7	10	9,55	0,45	0,20	0,02
8	6	4,04	1,96	3,84	0,95
$\sum$	100				$\chi^2_{набл} = 3{,}33$

Из табл. 15 находим

$$\chi^2_{набл} = 3{,}33.$$

По таблице критических точек распределения χ^2 (прил. 7) при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы

$$k = s - 3 = 8 - 3 = 5$$

находим критическую точку правосторонней критической области:

$$\chi^2_{кр}(0{,}05; 5) = 11{,}07.$$

Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, гипотезу о нормальном распределении не отклоняем. Другими словами, эмпирические и теоретические частоты различаются незначаще. □

Пусть статистическое распределение выборки задано в виде последовательности интервалов и соответствующих им частот:

$(x_i; x_{i+1}]$	$(x_1; x_2]$	$(x_2; x_3]$	$\cdots$	$(x_N; x_{N+1}]$
n_i	n_1	n_2	$\cdots$	n_N

Необходимо с помощью *критерия Пирсона* проверить гипотезу о нормальном распределении генеральной совокупности.

Правило 2. Для того чтобы при заданном уровне значимости α проверить гипотезу о нормальном распределении генеральной совокупности, необходимо:

1) вычислить выборочное среднее $\overline{x}_{\text{в}}$ и выборочное среднее квадратическое отклонение $\sigma_{\text{в}}$, причём в качестве варианты $x_i^\star$ берут среднее арифметическое концов интервала:

$$x_i^\star = \frac{x_i + x_{i+1}}{2};$$

2) нормировать исследуемую случайную величину X, т. е. перейти к случайной величине $Z = \dfrac{X - \overline{x^\star}}{\sigma^\star}$, и вычислить концы интервалов: $z_i = \dfrac{x_i - \overline{x^\star}}{\sigma^\star}$, причём наименьшее значение Z, т. е. z_1, считают равным $-\infty$, а наибольшее, т. е. z_{N+1}, — равным ∞;

3) вычислить теоретические частоты:

$$n_i' = n \cdot \mathbf{P}_i,$$

где n — объём выборки; $\mathbf{P}_i = \Phi(z_{i+1}) - \Phi(z_i)$ — вероятности попадания X в интервалы $(x_i; x_{i+1}]$; $\Phi(z)$ — функция Лапласа (прил. 1);

4) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона. Для этого:

а) составляют расчётную таблицу (см. табл. 15), по которой находят наблюдаемое значение критерия Пирсона

$$\chi^2_{\text{набл}} = \sum_i \frac{(n_i - n_i')^2}{n_i'};$$

б) по таблице критических точек распределения χ^2 (прил. 7) при заданном уровне значимости α и числе степеней свободы $k = s - 3$ (s — количество интервалов выборки) находят критическую точку $\chi^2_{\text{кр}}(\alpha; k)$ правосторонней критической области.

Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, нет оснований отклонять гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначаще (случайно). Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, гипотезу отклоняют. Другими словами, эмпирические и теоретические частоты различаются значаще.

Замечание 2. Интервалы, которые содержат малочисленные частоты выборки ($n_i < 5$), следует объединить, а частоты этих интервалов сложить. Если происходило объединение интервалов, при определении числа степеней свободы по формуле $k = s - 3$ следует в качестве s взять количество интервалов выборки, которые остались после объединения интервалов.

Пример 2. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о нормальном распределении генеральной совокупности X, если из этой совокупности получена такая выборка объёма $n = 100$:

$(x_i; x_{i+1}]$	$(3; 7]$	$(7; 11]$	$(11; 15]$	$(15; 19]$	$(19; 23]$	$(23; 27]$	$(27; 31]$
n_i	6	16	19	17	15	14	13

Решение. Сначала вычислим выборочное среднее $\overline{x}_{\text{в}}$ и среднее квадратическое отклонение выборки $\sigma_{\text{в}}$. Для этого преобразуем интервальное статистическое распределение выборки в точечное:

x_i	5	9	13	17	21	25	29
n_i	6	16	19	17	15	14	13

Итак,

$$\overline{x}_{\text{в}} = \frac{5 \cdot 6 - 9 \cdot 16 - 13 \cdot 19 + 17 \cdot 17 + 21 \cdot 15 + 25 \cdot 14 + 29 \cdot 13}{100} = $$

$$= 17{,}52;$$

$$\sigma_{\text{в}} = \Big(\big[5^2 \cdot 6 + 9^2 \cdot 16 + 13^2 \cdot 19 + 17^2 \cdot 17 + 21^2 \cdot 15 + 25^2 \cdot 14 +$$

$$+ 29^2 \cdot 13\big] /100 - 17{,}52^2\Big)^{\frac{1}{2}} \approx 7{,}19.$$

Найдём нормированные интервалы $(z_i; z_{i+1}]$, учитывая, что выборочное среднее $\overline{x}_в = 17{,}52$ и среднее квадратическое отклонение выборки $\sigma_в = 7{,}19$. Для этого составим расчётную табл. 16 (левый конец первого интервала положим равным $-\infty$, а правый конец последнего интервала — равным ∞).

Таблица 16

i	x_i	x_{i+1}	$x_i - \overline{x}_в$	$x_{i+1} - \overline{x}_в$	$z_i = \dfrac{x_i - \overline{x}_в}{\sigma_в}$	$z_{i+1} = \dfrac{x_{i+1} - \overline{x}_в}{\sigma_в}$
1	3	7	−14,5	−10,5	$-\infty$	−1,46
2	7	11	−10,5	−6,5	−1,46	−0,90
3	11	15	−6,5	−2,5	−0,90	−0,35
4	15	19	−2,5	1,5	−0,35	0,21
5	19	23	1,5	5,5	0,21	0,76
6	23	27	5,5	9,5	0,76	1,32
7	27	31	9,5	13,5	1,32	∞

Определим теоретические вероятности $\mathbf{P}_i$ и теоретические частоты:

$$n_i' = n \cdot \mathbf{P}_i.$$

Для этого составим расчётную табл. 17.

Таблица 17

	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$\mathbf{P}_i = \Phi(z_{i+1}) - \Phi(z_{i+1})$	$n_i' = n \cdot \mathbf{P}_i$
1	$-\infty$	−1,48	−0,5000	−0,4306	0,0694	6,94
2	−1,48	−0,92	−0,4306	−0,3212	0,1094	10,94
3	−0,92	−0,35	−0,3212	−0,1368	0,1844	18,44
4	−0,35	0,21	−0,1368	0,0832	0,2200	22,00
5	0,21	0,78	0,0832	0,2823	0,1991	19,91
6	0,78	1,34	0,2823	0,4099	0,1276	12,76
7	1,34	∞	0,4099	0,5000	0,0901	9,01
$\sum$					1	100

Сравним эмпирические и теоретические частоты, используя критерий Пирсона. Для этого вычислим сначала наблюдаемое значение критерия Пирсона с помощью расчётной табл. 18. Столбцы 7 и 8 этой таблицы введены для контроля вычислений по формуле

$$\chi^2_{\text{набл}} = \sum_i \frac{n_i^2}{n_i'} - n.$$

Таблица 18

1	2	3	4	5	6	7	8
i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$	n_i^2	$\frac{n_i^2}{n_i'}$
1	6	6,94	−0,94	0,8836	0,1273	36	5,1873
2	16	10,94	5,06	25,6036	2,3404	256	23,4004
3	19	18,44	0,56	0,3136	0,0170	361	19,5770
4	17	22,00	−5,00	25,0000	1,1364	289	13,1364
5	15	19,91	−4,91	24,1081	1,2109	225	11,3009
6	14	12,76	1,24	1,5376	0,1205	196	15,3605
7	13	9,01	3,99	15,9201	1,7669	169	18,7569
$\sum$	100	100			$\chi^2_{\text{набл}} = 6{,}7194$		106,7194

Контроль: Поскольку

$$\sum_i \frac{n_i^2}{n_i'} - n = 106{,}7194 - 100 = 6{,}7194 = \chi^2_{\text{набл}},$$

вычисление проведено правильно.

По таблице критических точек распределения χ^2 (прил. 7) при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы

$$k = s - 3 = 7 - 3 = 4$$

найдём критическую точку правосторонней критической области:

$$\chi^2_{\text{кр}}(0{,}05; 4) = 9{,}49.$$

Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, гипотезу о нормальном распределении не отклоняем. Другими словами, эмпирические и теоретические частоты различаются незначаще. □

15.5. Проверка гипотезы о равномерном распределении

Пусть задано интервальное статистическое распределение выборки:

$(x_i; x_{i+1}]$	$(x_0; x_1]$	$(x_1; x_2]$	$\dots$	$(x_{N-1}; x_N]$
n_i	n_0	n_1	$\dots$	n_{N-1}

причём n — объём выборки.

Необходимо с помощью критерия Пирсона проверить гипотезу о равномерном распределении генеральной совокупности.

Правило 1. Для того чтобы проверить гипотезу о равномерном распределении исследуемой случайной величины X, т. е. о плотности X:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a; b], \\ 0, & x \notin [a; b], \end{cases}$$

необходимо:

1) оценить параметры a и b — концы интервала, в котором наблюдались возможные значения X, по формулам ($a^\star$ и $b^\star$ — оценки параметров)

$$a^\star = \overline{x}_\text{в} - \sqrt{3}\sigma_\text{в}, \qquad b^\star = \overline{x}_\text{в} + \sqrt{3}\sigma_\text{в};$$

2) найти плотность вероятности предполагаемого распределения:

$$f(x) = \begin{cases} \dfrac{1}{b^\star - a^\star}, & x \in [a^\star; b^\star], \\ 0, & x \notin [a^\star; b^\star]; \end{cases}$$

3) определить теоретические частоты:

$$n_1' = n\mathbf{P}_1 = n\frac{1}{b^\star - a^\star}(x_1 - a^\star);$$

$$n_i' = n\frac{1}{b^\star - a^\star}(x_i - x_{i-1}), \quad i = 2, 3, \dots, N-1;$$

$$n_N' = n\frac{1}{b^\star - a^\star}(b^\star - x_{N-1});$$

4) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, взяв число степеней свободы $k = s - 3$, где s — количество интервалов в выборке.

Пример 1. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о равномерном распределении генеральной совокупности X, если из этой совокупности получена такая выборка объёма $n = 200$:

$(x_i; x_{i+1}]$	$(-4; -2]$	$(-2; 0]$	$(0; 2]$	$(2; 4]$	$(4; 6]$	$(6; 8]$	$(8; 10]$
n_i	28	27	34	23	25	31	32

Решение. Сначала вычислим выборочное среднее $\overline{x}_{\text{в}}$ и среднее квадратическое отклонение выборки $\sigma_{\text{в}}$. Для этого преобразуем интервальное статистическое распределение выборки в точечное:

x_i	-3	-1	1	3	5	7	9
n_i	28	27	34	23	25	31	32

Итак,

$$\overline{x}_{\text{в}} = \frac{-3\cdot 28 - 1\cdot 27 + 1\cdot 34 + 3\cdot 23 + 5\cdot 25 + 7\cdot 31 + 9\cdot 32}{200} = $$
$$= 3{,}11;$$
$$\sigma_{\text{в}} = \left(\frac{3^2\cdot 28 + 1\cdot 27 + 1\cdot 34 + 3^2\cdot 23 + 5^2\cdot 25 + 7^2\cdot 31 + 9^2\cdot 32}{200} - \right.$$
$$\left. - 3{,}11^2\right)^{\frac{1}{2}} \approx 4{,}075.$$

Найдём оценки параметров a и b равномерного распределения:

$$a^\star = \overline{x}_{\text{в}} - \sqrt{3}\sigma_{\text{в}} \approx 3{,}11 - \sqrt{3}\cdot 4{,}075 \approx -3{,}95;$$
$$b^\star = \overline{x}_{\text{в}} + \sqrt{3}\sigma_{\text{в}} \approx 3{,}11 + \sqrt{3}\cdot 4{,}075 \approx 10{,}17.$$

Определим плотность предполагаемого равномерного распределения:

$$f(x) = \begin{cases} \dfrac{1}{b^\star - a^\star}, & x \in [a^\star; b^\star], \\ 0, & x \notin [a^\star; b^\star] \end{cases} =$$

$$= \begin{cases} \dfrac{1}{10{,}17 + 3{,}95}, & x \in [-3{,}95; 10{,}17], \\ 0, & x \notin [-3{,}95; 10{,}17] \end{cases} \approx$$

$$\approx \begin{cases} 0{,}07, & x \in [-3{,}95; 10{,}17], \\ 0, & x \notin [-3{,}95; 10{,}17]. \end{cases}$$

Вычислим теоретические частоты:

$$n'_1 = n \cdot \frac{1}{b^\star - a^\star} \cdot (x_1 - a^\star) = 200 \cdot 0{,}07 \cdot (-2 + 3{,}95) = 27{,}3;$$

$$n'_2 = n \cdot \frac{1}{b^\star - a^\star} \cdot (x_2 - x_1) = 200 \cdot 0{,}07 \cdot (0 + 2) = 28;$$

$$n'_3 = \dots = n'_6 = n'_2 = 28;$$

$$n'_7 = n \cdot \frac{1}{b^\star - a^\star} \cdot (b^\star - x_6) = 200 \cdot 0{,}07 \cdot (10{,}17 - 8) = 30{,}38.$$

Сравним эмпирические и теоретические частоты, используя критерий Пирсона с помощью расчетной табл. 19.

Таблица 19

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\dfrac{(n_i - n'_i)^2}{n'_i}$
1	28	27,3	−0,7	0,49	0,02
2	27	28	1	1	0,04
3	34	28	−6	36	1,29
4	23	28	5	25	0,89
5	25	28	3	9	0,32
6	31	28	−3	9	0,32
7	32	30,38	−1,62	2,62	0,09
$\sum$	200				$\chi^2_{\text{набл}} = 2{,}97$

Для этого вычислим сначала наблюдаемое значение критерия Пирсона. Из расчётной табл. 19 получаем

$$\chi^2_{\text{набл}} = 2{,}97.$$

По таблице критических точек распределения χ^2 (прил. 7) при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы

$$k = s - 3 = 7 - 3 = 4$$

находим критическую точку правосторонней критической области:

$$\chi^2_{\text{кр}}(0{,}05; 4) = 9{,}49.$$

Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, гипотезу о равномерном распределении не отклоняем. Другими словами, эмпирические и теоретические частоты различаются незначаще. □

15.6. Проверка гипотезы о показательном распределении

Пусть задано интервальное статистическое распределение выборки:

$(x_i; x_{i+1}]$	$(x_1; x_2]$	$(x_2; x_3]$	$\cdots$	$(x_N; x_{N+1}]$
n_i	n_1	n_2	$\cdots$	n_N

(n — объём выборки).

Необходимо с помощью критерия Пирсона проверить гипотезу о показательном распределении генеральной совокупности.

Правило 1. Для того чтобы при уровне значимости α проверить гипотезу о показательном распределении исследуемой случайной величины X, необходимо:

1) найти выборочное среднее $\overline{x}_{\text{в}}$, выбрав в качестве вариант середины интервалов

$$x_i^\star = \frac{x_i + x_{i+1}}{2};$$

2) взять в качестве оценки параметра λ показательного распределения величину, обратную выборочному среднему:

$$\lambda^\star = \frac{1}{\overline{x}_{\text{в}}};$$

3) найти вероятности попадания случайной величины X в частные интервалы $(x_i; x_{i+1}]$ по формуле

$$\mathbf{P}_i = \mathbf{P}(x_i < X < x_{i+1}) = e^{-\lambda^\star x_i} - e^{-\lambda^\star x_{i+1}};$$

4) вычислить теоретические частоты:

$$n_i' = n \cdot \mathbf{P}_i;$$

5) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, взяв число степеней свободы $k = s - 2$, где s — количество интервалов в выборке.

Замечание 1. Интервалы, которые содержат малочисленные частоты выборки ($n_i < 5$), следует объединить, а частоты этих интервалов сложить. Если происходило объединение интервалов, при определении числа степеней свободы по формуле $k = s - 2$ следует в качестве s взять количество интервалов выборки, которые остались после объединения интервалов.

Пример 1. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о показательном распределении генеральной совокупности X, если из этой совокупности получена такая выборка объёма $n = 200$:

$(x_i; x_{i+1}]$	$(0; 6]$	$(6; 12]$	$(12; 18]$	$(18; 24]$	$(24; 30]$	$(30; 36]$	$(36; 42]$
n_i	115	51	18	9	4	2	1

Решение. Поскольку мы имеем малочисленные частоты, то предварительно объединим их. Получим такое распределение выборки:

$(x_i; x_{i+1}]$	$(0; 6]$	$(6; 12]$	$(12; 18]$	$(18; 24]$	$(24; 42]$
n_i	115	51	18	9	7

Чтобы найти выборочное среднее $\overline{x}_{\text{в}}$, преобразуем интервальное статистическое распределение в точечное:

x_i	3	9	15	21	33
n_i	115	51	18	9	7

Итак,

$$\overline{x}_{\text{в}} = \frac{3 \cdot 115 + 9 \cdot 51 + 15 \cdot 18 + 21 \cdot 9 + 33 \cdot 7}{200} = 7{,}47.$$

Найдём оценку параметра предполагаемого показательного распределения:

$$\lambda^{\star} = \frac{1}{\overline{x}_{\text{в}}} = \frac{1}{7{,}47} \approx 0{,}13.$$

Другими словами, плотность предполагаемого показательного распределения имеет вид

$$f(x) = \begin{cases} 0, & x < 0; \\ 0{,}13 \cdot e^{-0{,}13x}, & x \geqslant 0. \end{cases}$$

Вычислим вероятности попадания случайной величины X в каждый из интервалов по формуле

$$\mathbf{P}_i = \mathbf{P}(x_i < X < x_{i+1}) = e^{-\lambda^{\star} x_i} - e^{-\lambda^{\star} x_{i+1}}.$$

Итак,

$$\mathbf{P}_1 = \mathbf{P}(0 < X < 6) = e^{-0{,}13 \cdot 0} - e^{-0{,}13 \cdot 6} \approx 0{,}5565;$$
$$\mathbf{P}_2 = \mathbf{P}(6 < X < 12) = e^{-0{,}13 \cdot 6} - e^{-0{,}13 \cdot 12} \approx 0{,}2468;$$
$$\mathbf{P}_3 = \mathbf{P}(12 < X < 18) = e^{-0{,}13 \cdot 12} - e^{-0{,}13 \cdot 18} \approx 0{,}1095;$$
$$\mathbf{P}_4 = \mathbf{P}(18 < X < 24) = e^{-0{,}13 \cdot 18} - e^{-0{,}13 \cdot 24} \approx 0{,}0486;$$
$$\mathbf{P}_5 = \mathbf{P}(24 < X < 42) = e^{-0{,}13 \cdot 24} - e^{-0{,}13 \cdot 42} \approx 0{,}0353.$$

Найдём теоретические частоты:

$$n'_1 = n \cdot \mathbf{P}_1 = 200 \cdot 0{,}5565 = 111{,}3;$$
$$n'_2 = n \cdot \mathbf{P}_2 = 200 \cdot 0{,}2468 = 49{,}36;$$
$$n'_3 = n \cdot \mathbf{P}_3 = 200 \cdot 0{,}1095 = 21{,}9;$$
$$n'_4 = n \cdot \mathbf{P}_4 = 200 \cdot 0{,}0486 = 9{,}72;$$
$$n'_5 = n \cdot \mathbf{P}_5 = 200 \cdot 0{,}0353 = 7{,}06.$$

Сравним эмпирические и теоретические частоты, используя критерий Пирсона. Для этого сначала вычислим наблюдаемое значение критерия Пирсона с помощью расчётной табл. 20.

Таблица 20

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\dfrac{(n_i - n'_i)^2}{n'_i}$
1	115	111,3	−3,7	13,69	0,1230
2	51	49,36	−1,64	2,6896	0,0545
3	18	21,9	3,9	15,21	0,6945
4	9	9,72	0,72	0,5184	0,0533
5	7	7,06	0,06	0,0036	0,0005
$\sum$	200				$\chi^2_{\text{набл}} = 0{,}9258$

Из расчётной табл. 20 получаем

$$\chi^2_{\text{набл}} = 0{,}9258.$$

По таблице критических точек распределения χ^2 (прил. 7) при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы

$$k = s - 2 = 5 - 2 = 3$$

находим критическую точку правосторонней критической области:

$$\chi^2_{\text{кр}}(0{,}05; 3) = 7{,}81.$$

Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, гипотезу о показательном распределении не отклоняем. Другими словами, эмпирические и теоретические частоты различаются незначаще. □

15.7. Проверка гипотезы о биномиальном распределении

Пусть проведено n опытов. Каждый опыт состоит из N независимых испытаний, в каждом из которых вероятность появления некоторого события A одна и та же. Регистрируется количество появлений события A в каждом опыте. В итоге получено статистическое распределение дискретной случайной величины X, которая характеризует количество появлений события A (в первой строке приведено количество появлений события A в одном опыте, а во второй — частота n_i, т. е. количество опытов, в которых зарегистрировано x_i появлений события A):

x_i	0	1	$\cdots$	N
n_i	n_0	n_1	$\cdots$	n_N

Необходимо с помощью критерия Пирсона проверить гипотезу о биномиальном законе распределения дискретной случайной величины X.

Правило 1. Для того чтобы при уровне значимости α проверить гипотезу о биномиальном распределении дискретной случайной величины X, необходимо:

1) найти по формуле Бернулли

$$\mathbf{P}_N(i) = C_N^i p^i (1-p)^{N-i}$$

вероятности $\mathbf{P}_N(i)$ появления события A ровно i раз в N испытаниях ($i = 0, 1, \ldots, s$, где s — максимальное количество наблюдаемых появлений события A в одном опыте, т. е. $s \leqslant N$);

2) определить теоретические частоты:

$$n_i' = n \cdot \mathbf{P}_N(i),$$

где n — количество опытов;

3) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, положив число степеней свободы $k = s - 1$ (считается, что вероятность p появления события A задана, т. е. она не

оценивалась по выборке и не объединялись малочисленные частоты).

Если вероятность p была оценена по выборке, то $k = s - 2$. Если, кроме того, были объединены малочисленные частоты, то s — количество вариант выборки, которые остались после объединения частот.

Пример 1. Над событием A, вероятность появления которого равняется 0,3, проведено $n = 100$ независимых испытаний, каждое из которых состояло из $N = 7$ опытов. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о биномиальном распределении случайной величины X (количество появлений события A), если получена такая выборка:

x_i	0	1	2	3	4	5	6	7
n_i	4	23	31	23	11	5	2	1

Решение. Учитывая, что

$$p = 0{,}3; \qquad q = 1 - p = 1 - 0{,}3 = 0{,}7,$$

по формуле Бернулли

$$\mathbf{P}_N(i) = C_N^i p^i q^{N-i}$$

вычислим вероятности $\mathbf{P}_N(i)$:

$$\mathbf{P}_7(0) = C_7^0 0{,}3^0 0{,}7^7 \approx 0{,}0824; \qquad \mathbf{P}_7(1) = C_7^1 0{,}3^1 0{,}7^6 \approx 0{,}2471;$$
$$\mathbf{P}_7(2) = C_7^2 0{,}3^2 0{,}7^5 \approx 0{,}3177; \qquad \mathbf{P}_7(3) = C_7^3 0{,}3^3 0{,}7^4 \approx 0{,}2269;$$
$$\mathbf{P}_7(4) = C_7^4 0{,}3^4 0{,}7^3 \approx 0{,}0972; \qquad \mathbf{P}_7(5) = C_7^5 0{,}3^5 0{,}7^2 \approx 0{,}0250;$$
$$\mathbf{P}_7(6) = C_7^6 0{,}3^6 0{,}7^1 \approx 0{,}0036; \qquad \mathbf{P}_7(7) = C_7^7 0{,}3^7 0{,}7^0 \approx 0{,}0002.$$

Найдём теоретические частоты по формуле $n_i' = n \cdot \mathbf{P}_N(i)$:

$$n_0' = 100 \cdot \mathbf{P}_7(0) = 8{,}24; \qquad n_1' = 100 \cdot \mathbf{P}_7(1) = 24{,}71;$$
$$n_2' = 100 \cdot \mathbf{P}_7(2) = 31{,}77; \qquad n_3' = 100 \cdot \mathbf{P}_7(3) = 22{,}69;$$
$$n_4' = 100 \cdot \mathbf{P}_7(4) = 9{,}72; \qquad n_5' = 100 \cdot \mathbf{P}_7(5) = 2{,}5;$$
$$n_6' = 100 \cdot \mathbf{P}_7(6) = 0{,}36; \qquad n_7' = 100 \cdot \mathbf{P}_7(7) = 0{,}02.$$

Поскольку частоты $n_0 = 4$, $n_6 = 2$ и $n_7 = 1$ малочисленные (меньше пяти), объединим их с другими частотами, а именно:

$$n_1 = 23 + 4 = 27; \qquad n_5 = 5 + 2 + 1 = 8.$$

В качестве теоретических частот, которые отвечают объединённым частотам, возьмём сумму соответствующих теоретических частот:

$$n_1' = 24{,}71 + 8{,}24 = 32{,}95; \qquad n_5' = 2{,}5 + 0{,}36 + 0{,}02 = 2{,}88.$$

Сравним эмпирические и теоретические частоты с помощью критерия Пирсона. Для этого составим расчётную табл. 21.

Таблица 21

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\dfrac{(n_i - n_i')^2}{n_i'}$
1	27	32,95	5,95	35,4025	1,0744
2	31	31,77	0,77	0,5929	0,0187
3	23	22,69	−0,31	0,0961	0,0042
4	11	9,72	−1,28	1,6384	0,1686
5	8	2,88	−5,12	26,2144	9,1022
$\sum$	100				$\chi^2_{\text{набл}} = 10{,}3681$

Из расчётной табл. 21 получаем

$$\chi^2_{\text{набл}} = 10{,}3681.$$

По таблице критических точек распределения χ^2 (прил. 7) при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы

$$k = s - 1 = 5 - 1 = 4$$

находим критическую точку правосторонней критической области:

$$\chi^2_{\text{кр}}(0{,}05; 4) = 9{,}49.$$

Поскольку $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, гипотезу о биномиальном распределении отклоняем. $\square$

15.8. Проверка гипотезы о распределении Пуассона

Пусть задано точечное статистическое распределение выборки. Необходимо с помощью критерия Пирсона проверить гипотезу о распределении генеральной совокупности по закону Пуассона.

Правило 1. Для того чтобы при уровне значимости α проверить гипотезу о том, что исследуемая случайная величина X распределена по закону Пуассона, необходимо:

1) найти по заданному статистическому распределению выборочное среднее $\overline{x}_{\text{в}}$;

2) взять в качестве оценки параметра λ распределения Пуассона выборочное среднее $\lambda^{\star} = \overline{x}_{\text{в}}$;

3) найти по формуле Пуассона (или в готовой таблице прил. 3) вероятности $\mathbf{P}_i$ появления ровно i событий в n испытаниях ($i = 1, 2, \ldots, r$, где r — максимальное количество наблюдаемых событий; n — объём выборки);

4) определить теоретические частоты

$$n'_i = n \cdot \mathbf{P}_i;$$

5) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, взяв число степеней свободы $k = s - 2$, где s — количество вариант выборки (если проводилось объединение малочисленных частот в одну группу, то s — количество вариант, которые остались после объединения частот).

Пример 1. В $n = 1000$ проверках партий товара регистрировалось количество x_i некачественной продукции, вследствие чего было получено такое статистическое распределение количества x_i брака в n_i партиях товара:

x_i	0	1	2	3	4	5	6	7
n_i	242	349	234	107	43	21	3	1

Необходимо при уровне значимости 0,05 проверить гипотезу о том, что количество бракованной продукции X распределено по закону Пуассона.

Решение. Сначала найдём выборочное среднее:

$$\overline{x}_{\text{в}} = \frac{0 \cdot 242 + 1 \cdot 349 + 2 \cdot 234 + 3 \cdot 107 + 4 \cdot 43 + 5 \cdot 21 + 6 \cdot 3 + 7 \cdot 1}{1000} =$$

$= 1{,}44.$

Возьмем в качестве оценки параметра λ распределения Пуассона выборочное среднее:

$$\lambda^\star = \overline{x}_{\text{в}} = 1{,}44.$$

Итак, предполагаемый закон Пуассона имеет вид

$$\mathbf{P}_{1000}(i) = 1{,}44^i \cdot \frac{e^{-1{,}44}}{i!}.$$

Положив $i = 0, 1, \ldots, 7$, вычислим вероятности $\mathbf{P}_i = \mathbf{P}_{1000}(i)$:

$$\mathbf{P}_0 = 1{,}44^0 \cdot \frac{e^{-1{,}44}}{0!} \approx 0{,}2369; \qquad \mathbf{P}_1 = 1{,}44^1 \cdot \frac{e^{-1{,}44}}{1!} \approx 0{,}3412;$$
$$\mathbf{P}_2 = 1{,}44^2 \cdot \frac{e^{-1{,}44}}{2!} \approx 0{,}2456; \qquad \mathbf{P}_3 = 1{,}44^3 \cdot \frac{e^{-1{,}44}}{3!} \approx 0{,}1179;$$
$$\mathbf{P}_4 = 1{,}44^4 \cdot \frac{e^{-1{,}44}}{4!} \approx 0{,}0424; \qquad \mathbf{P}_5 = 1{,}44^5 \cdot \frac{e^{-1{,}44}}{5!} \approx 0{,}0122;$$
$$\mathbf{P}_6 = 1{,}44^6 \cdot \frac{e^{-1{,}44}}{6!} \approx 0{,}0029; \qquad \mathbf{P}_7 = 1{,}44^7 \cdot \frac{e^{-1{,}44}}{7!} \approx 0{,}0006.$$

Найдём теоретические частоты $n'_i = n \cdot \mathbf{P}_i$:

$$n'_0 = 1000 \cdot \mathbf{P}_0 = 236{,}9; \qquad n'_1 = 1000 \cdot \mathbf{P}_1 = 341{,}2;$$
$$n'_2 = 1000 \cdot \mathbf{P}_2 = 245{,}6; \qquad n'_3 = 1000 \cdot \mathbf{P}_3 = 117{,}9;$$
$$n'_4 = 1000 \cdot \mathbf{P}_4 = 42{,}4; \qquad n'_5 = 1000 \cdot \mathbf{P}_5 = 12{,}2;$$
$$n'_6 = 1000 \cdot \mathbf{P}_6 = 2{,}9; \qquad n'_7 = 1000 \cdot \mathbf{P}_7 = 0{,}6.$$

Поскольку частоты $n_6 = 3$ и $n_7 = 1$ малочисленные (меньше пяти), объединим их с частотой n_5, а именно

$$n_5 = 21 + 3 + 1 = 25.$$

В качестве теоретической частоты, которая отвечает объединённой частоте, возьмём сумму соответствующих теоретических частот:

$$n'_5 = 12{,}2 + 2{,}9 + 0{,}6 = 15{,}7.$$

Сравним эмпирические и теоретические частоты с помощью критерия Пирсона. Для этого составим расчётную табл. 22.

Таблица 22

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\dfrac{(n_i - n'_i)^2}{n'_i}$
0	242	236,9	−5,1	26,01	0,1098
1	349	341,2	−7,8	60,84	0,1783
2	234	245,6	11,6	134,56	0,5479
3	107	117,9	10,9	118,81	1,0077
4	43	42,4	−0,6	0,36	0,0085
5	25	15,7	−9,3	86,49	5,5089
$\sum$	1000				$\chi^2_{\text{набл}} = 7{,}3611$

Из расчётной табл. 22 получаем

$$\chi^2_{\text{набл}} = 7{,}3611.$$

По таблице критических точек распределения χ^2 (прил. 7) при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы

$$k = s - 1 = 6 - 2 = 4$$

находим критическую точку правосторонней критической области:

$$\chi^2_{\text{кр}}(0{,}05; 4) = 9{,}49.$$

Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, нет оснований отклонять гипотезу о распределении случайной величины X (количество бракованного товара в партии) по закону Пуассона. $\square$

Задачи к разделу 15

Задача 1. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 3{,}2$ получена некоторая выборка объёма $n = 64$, по которой найдено выборочное среднее $\overline{x}_{\text{в}} = 13$. Необходимо при уровне значимости 0,05 проверить нулевую гипотезу $H_0\colon a = a_0 = 12$ при конкурирующей гипотезе $H_1\colon a \neq 12$.

Ответ. Основная гипотеза отклоняется.

Задача 2. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 3{,}3$ получена выборка объёма $n = 121$, по которой найдено выборочное среднее $\overline{x}_{\text{в}} = 10{,}6$. Необходимо при уровне значимости 0,01 проверить нулевую гипотезу $H_0\colon a = 10$ при конкурирующей гипотезе:

а) $H_1\colon a \neq 10$;

б) $H_1\colon a > 10$;

в) $H_1\colon a < 10$.

Кроме того, найти мощности правостороннего и двухстороннего критериев.

Ответ. Основные гипотезы не отклоняются во всех трех случаях. Мощности правостороннего и двухстороннего критериев равняются соответственно 0,3721 и 0,2824.

Задача 3. Для выборки объёма $n = 20$, полученной из нормальной генеральной совокупности, найдено выборочное среднее $\overline{x} = 37$ и «исправленное» выборочное среднее квадратическое отклонение $s = 4{,}2$. Необходимо при уровне значимости $\alpha = 0{,}05$ проверить основную гипотезу $H_0\colon a = a_0 = 35$ при альтернативной гипотезе $H_1\colon a \neq 35$.

Ответ. Основная гипотеза отклоняется.

Задача 4. Контрольная длина некоторой детали должна составлять $a = a_0 = 12$ мм. Контрольные измерения 25 случайно отобранных деталей дали такие результаты:

Длина, мм	11	11,5	12	12,5	13
Количество деталей	6	8	7	3	1

Необходимо при уровне значимости 0,01 проверить основную гипотезу $H_0\colon a = a_0 = 12$ при конкурирующей гипотезе:

а) $H_1: a \neq a_0$;

б) $H_1: a > a_0$;

в) $H_1: a < a_0$.

Ответ. а) и б) Основная гипотеза не отклоняется; в) основная гипотеза отклоняется.

Задача 5. Из нормальной генеральной совокупности получена выборка объёма $n = 31$, для которой найдена выборочная «исправленная» дисперсия $s^2 = 44$. Проверить при уровне значимости $\alpha = 0{,}01$ основную гипотезу $H_0\colon \sigma^2 = \sigma_0^2 = 25$, если альтернативная гипотеза $H_1: \sigma_0^2 > 25$.

Ответ. Основная гипотеза отклоняется.

Задача 6. Из нормальной генеральной совокупности получена выборка объёма $n = 26$, для которой найдена выборочная «исправленная» дисперсия $s^2 = 3{,}2$. Проверить при уровне значимости $\alpha = 0{,}05$ основную гипотезу $H_0\colon \sigma^2 = \sigma_0^2 = 2$, если альтернативная гипотеза:

а) $H_1: \sigma_0^2 > 2$;

б) $H_1: \sigma_0^2 \neq 2$;

в) $H_1: \sigma_0^2 < 2$.

Ответ. а) Основная гипотеза отклоняется; б) и в) основная гипотеза не отклоняется.

Задача 7. Из нормальной генеральной совокупности получена выборка объёма $n = 200$:

x_i	-1	0	1	2	3	4	5
n_i	17	24	30	34	35	33	27

Необходимо проверить при уровне значимости $\alpha = 0{,}01$ основную гипотезу $H_0\colon \sigma^2 = \sigma_0^2 = 3$, если альтернативная гипотеза $H_1: \sigma_0^2 > 3$.

Ответ. Основная гипотеза отклоняется.

Задача 8. Для некоторого события в результате 100 независимых испытаний была найдена относительная частота $\dfrac{m}{n} = 0{,}34$. При уровне значимости 0,05 проверить основную гипотезу $H_0\colon p = p_0 = 0{,}25$ при конкурирующей гипотезе $H_1: p \neq 0{,}25$.

Ответ. Основная гипотеза отклоняется.

Задача 9. Для некоторого события в результате 256 независимых испытаний была найдена относительная частота $\dfrac{m}{n} = 0{,}375$. При уровне значимости 0,01 проверить основную гипотезу $H_0: p = p_0 = 0{,}4$ при конкурирующей гипотезе:

а) $H_1: p > 0{,}4$;

б) $H_1: p < 0{,}4$.

Ответ. Основная гипотеза не отклоняется в обеих случаях.

Задача 10. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о нормальном распределении генеральной совокупности X, если из этой совокупности получена такая выборка объёма $n = 200$:

x_i	15	16	17	18	19	20	21	22
n_i	8	28	31	41	33	28	24	7

Ответ. Гипотеза о нормальном распределении отклоняется.

Задача 11. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о нормальном распределении генеральной совокупности X, если из этой совокупности получена такая выборка объёма $n = 200$:

$(x_i; x_{i+1}]$	$(0;3]$	$(3;6]$	$(6;9]$	$(9;12]$	$(12;15]$	$(15;18]$	$(18;21]$
n_i	15	25	45	51	34	22	8

Ответ. Гипотеза о нормальном распределении отклоняется.

Задача 12. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о равномерном распределении генеральной совокупности X, если из этой совокупности получена такая выборка объёма $n = 100$:

$(x_i; x_{i+1}]$	$(4;7]$	$(7;10]$	$(10;13]$	$(13;16]$	$(16;19]$	$(19;22]$	$(22;25]$
n_i	12	9	21	8	11	19	20

Ответ. Гипотеза о равномерном распределении отклоняется.

Задача 13. Над событием A, вероятность появления которого равняется 0,4, проведено $n = 200$ независимых испытаний, каждое из которых состояло из $N = 10$ опытов. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о биномиальном распределении случайной величины X (количество появлений события A), если получена такая выборка:

x_i	0	1	2	3	4	5	6	7	8	9
n_i	2	12	31	42	39	35	27	5	4	2

Указание. Следует объединить малочисленные частоты: n_0 — с n_1, а n_8, n_9 и n_{10} — с n_7.

Ответ. Гипотеза о биномиальном распределении не отклоняется.

Задача 14. В $n = 1000$ проверках партий товара регистрировалось количество x_i некачественной продукции, вследствие чего было получено такое статистическое распределение количества x_i брака в n_i партиях товара:

x_i	0	1	2	3	4	5	6	7
n_i	427	363	154	41	9	3	2	1

Необходимо при уровне значимости 0,05 проверить гипотезу о том, что количество бракованной продукции X распределено по закону Пуассона.

Указание. Следует объединить малочисленные частоты n_6 и n_7 с n_5.

Ответ. Гипотеза о распределении Пуассона не отклоняется.

Раздел 16. Элементы дисперсионного анализа

Ключевые слова

Анализ
Выборка
Двухфакторный
Дисперсионный
Индивидуальный
Несвязный
Общий
Однофакторный
Остаточный
Связный
Сумма
Факторный

16.1. Однофакторный дисперсионный анализ для несвязных выборок

Пусть на количественный нормально распределенный признак X действует некоторый фактор F, имеющий p постоянных уровней $F_1, F_2, \ldots, F_p$. На каждом уровне осуществлено по q испытаний **над разными выборками**. Результаты наблюдений — числа x_{ij}, где i — номер испытания, $i = 1, 2, \ldots, q$; j — номер уровня фактора, $j = 1, 2, \ldots, p$, — записывают в виде таблицы (табл. 23).

Таблица 23

Номер испытания i	Уровень фактора			
	F_1	F_2	$\cdots$	F_p
1	x_{11}	x_{12}	$\cdots$	x_{1p}
2	x_{21}	x_{22}	$\cdots$	x_{2p}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
q	x_{q1}	x_{q2}	$\cdots$	x_{qp}
Групповое среднее $\overline{x}_{\text{гр}}$	$\overline{x}_{\text{гр. }1}$	$\overline{x}_{\text{гр. }2}$	$\cdots$	$\overline{x}_{\text{гр. }p}$

Задача заключается в следующем: при уровне значимости α проверить основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы.

Чтобы решить эту задачу, находят:

а) *общую сумму* квадратов отклонений наблюдаемых значений признака от общего выборочного среднего:

$$S_{\text{общ}} = \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij} - \overline{x}_{\text{в}})^2;$$

б) *факторную сумму* квадратов отклонений групповых средних от общего выборочного среднего (характеризует рассеяние «между группами»):

$$S_{\text{факт}} = q\sum_{j=1}^{p}(\overline{x}_{\text{гр. } j} - \overline{x}_{\text{в}})^2;$$

в) *остаточную сумму* квадратов отклонений наблюдаемых значений группы от своего группового среднего (характеризует рассеяние «внутри группы»):

$$S_{\text{ост}} = \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij} - \overline{x}_{\text{гр. } j})^2.$$

Практически остаточную сумму находят по формуле

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}}.$$

Общую и факторную суммы удобнее вычислять по таким формулам:

$$S_{\text{общ}} = \sum_{j=1}^{p} P_j - \frac{\left(\sum\limits_{j=1}^{p} R_j\right)^2}{pq};$$

$$S_{\text{факт}} = \frac{\sum\limits_{j=1}^{p} R_j^2}{q} - \frac{\left(\sum\limits_{j=1}^{p} R_j\right)^2}{pq},$$

где $P_j = \sum\limits_{i=1}^{q} x_{ij}^2$ — сумма квадратов наблюдаемых значений признака на уровне F_j; $R_j = \sum\limits_{i=1}^{q} x_{ij}$ — сумма наблюдаемых значений признака на уровне F_j.

Если наблюдаемые значения признака сравнительно большие, то для упрощения вычислений вычитают от каждого наблюдаемого значения одно и то же число C, которое приблизительно равняется общему среднему. Если уменьшенные значения обозначить

$$y_{ij} = x_{ij} - C,$$

то

$$S_{\text{общ}} = \sum_{j=1}^{p} Q_j - \frac{\left(\sum\limits_{j=1}^{p} T_j\right)^2}{pq};$$

$$S_{\text{факт}} = \frac{\sum\limits_{j=1}^{p} T_j^2}{q} - \frac{\left(\sum\limits_{j=1}^{p} T_j\right)^2}{pq},$$

где $Q_j = \sum\limits_{i=1}^{q} y_{ij}^2$ — сумма квадратов уменьшенных значений признака на уровне F_j; $T_j = \sum\limits_{i=1}^{q} y_{ij}$ — сумма уменьшенных значений признака на уровне F_j.

Поделив факторную и остаточную суммы на соответствующее число степеней свободы

$$k_{\text{факт}} = p - 1, \qquad k_{\text{ост}} = p(q-1),$$

находят факторную и остаточную дисперсии:

$$s^2_{\text{факт}} = \frac{S_{\text{факт}}}{p-1}; \qquad s^2_{\text{ост}} = \frac{S_{\text{ост}}}{p(q-1)}.$$

Факторную и остаточную дисперсии сравнивают по *критерию Фишера—Снедекора*:

$$F_{\text{набл}} = \frac{s^2_{\text{факт}}}{s^2_{\text{ост}}}.$$

Далее по таблице критических точек распределения Фишера—Снедекора (прил. 8) при заданном уровне значимости α и числе степеней свободы

$$k_1 = k_{\text{факт}}, \qquad k_2 = k_{\text{ост}}$$

находят критическую точку $F_{\text{кр}}(\alpha; k_1, k_2)$.

Если $F_{\text{набл}} < F_{\text{кр}}$, различие групповых средних считают незначащим. Если $F_{\text{набл}} > F_{\text{кр}}$, то различие считают значащим.

Замечание 1. Однофакторный дисперсионный анализ требует не менее трех уровней фактора и не менее двух наблюдений на каждом уровне, причём для всех уровней количество наблюдений должно быть одинаковым.

Замечание 2. Результирующий признак должен быть нормально распределён в исследуемой выборке.

Замечание 3. Если факторная дисперсия окажется меньше остаточной, сразу можно делать вывод о том, что нет оснований отвергать основную гипотезу. Другими словами, в данном случае дальнейшие вычисления излишни.

Замечание 4. Если наблюдаемые значения x_{ij} кратны некоторому числу k (например, десятичные дроби кратны числу 10^{-d}, где d — количество цифр после запятой), целесообразно перейти к новым значениям:

$$y_{ij} = \frac{x_{ij}}{k}.$$

При этом факторная и остаточная дисперсии уменьшатся в k^2 раз, но их отношение не изменится.

Пример 1. Проведено по семь испытаний на каждом из четырёх уровней фактора над разными выборками. Методом дисперсионного анализа при уровне значимости 0,05 проверить основную гипотезу о равенстве групповых средних. Предполагается, что выборки получены из нормальных совокупностей с одинаковыми дисперсиями. Результаты испытания приведены в табл. 24.

Решение. Найдём общее среднее:

$$\begin{aligned}\overline{x}_{\text{в}} = (&24 + 26 + 21 + 25 + 25 + 27 + 22 + 26 + 26 + 28 + 24 + 28 + \\ &+ 29 + 29 + 25 + 28 + 31 + 30 + 27 + 29 + 32 + 32 + 29 + 30 + \\ &+ 33 + 35 + 30 + 33)/(7 \cdot 4) = 28.\end{aligned}$$

Для упрощения расчётов вычтем от каждого наблюдаемого значения x_{ij} общее среднее $\overline{x}_{\text{в}} = 28$, т. е. перейдём к уменьшенным

Таблица 24

Номер испытания i	Уровень фактора			
	F_1	F_2	F_3	F_4
1	24	26	21	25
2	25	27	22	26
3	26	28	24	28
4	29	29	25	28
5	31	30	27	29
6	32	32	29	30
7	33	35	30	33
Групповое среднее $\overline{x}_{\text{гр}}$	28,57143	29,57143	25,42857	28,42857

величинам:

$$y_{ij} = x_{ij} - 28.$$

Составим расчётную табл. 25.

Таблица 25

Номер испытания i	Уровень фактора								Итоговый столбик
	F_1		F_2		F_3		F_4		
	y_{i1}	y_{i1}^2	y_{i2}	y_{i2}^2	y_{i3}	y_{i3}^2	y_{i4}	y_{i4}^2	
1	−4	16	−2	4	−7	49	−3	9	
2	−3	9	−1	1	−6	36	−2	4	
3	−2	4	0	0	−4	16	0	0	
4	1	1	1	1	−3	9	0	0	
5	3	9	2	4	−1	1	1	1	
6	4	16	4	16	1	1	2	4	
7	5	25	7	49	2	4	5	25	
$Q_j = \sum\limits_{i=1}^{7} y_{ij}^2$		80		75		116		43	$\sum\limits_{j=1}^{4} Q_j = 314$
$T_j = \sum\limits_{i=1}^{7} y_{ij}$	4		11		−18		3		$\sum\limits_{j=1}^{4} T_j = 0$
T_j^2	16		121		324		9		$\sum\limits_{j=1}^{4} T_j^2 = 470$

Используя итоговый столбик табл. 25, находим общую и факторную суммы квадратов отклонений, учитывая, что количество уровней фактора $p = 4$, количество испытаний на каждом уровне $q = 7$:

$$S_{\text{общ}} = \sum_{j=1}^{p} Q_j - \frac{\left(\sum\limits_{j=1}^{p} T_j\right)^2}{pq} = 314 - 0 = 314;$$

$$S_{\text{факт}} = \frac{\sum\limits_{j=1}^{p} T_j^2}{q} - \frac{\left(\sum\limits_{j=1}^{p} T_j\right)^2}{pq} = \frac{470}{7} - 0 \approx 67{,}14.$$

Найдём остаточную сумму квадратов отклонений:

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}} = 314 - 67{,}14 = 246{,}86.$$

Определим количество степеней свободы:

$$k_{\text{факт}} = p - 1 = 4 - 1 = 3, \qquad k_{\text{ост}} = p(q-1) = 4 \cdot (7-1) = 24.$$

Вычислим факторную и остаточную дисперсии:

$$s^2_{\text{факт}} = \frac{S_{\text{факт}}}{p-1} = \frac{67{,}14}{3} \approx 22{,}38;$$

$$s^2_{\text{ост}} = \frac{S_{\text{ост}}}{p(q-1)} = \frac{246{,}86}{24} \approx 10{,}29.$$

Сравним факторную и остаточную дисперсии с помощью критерия Фишера—Снедекора. Для этого сначала определим наблюдаемое значение критерия:

$$F_{\text{набл}} = \frac{s^2_{\text{факт}}}{s^2_{\text{ост}}} = \frac{22{,}38}{10{,}29} \approx 2{,}175.$$

Учитывая, что число степеней свободы числителя $k_1 = 3$, а знаменателя — $k_2 = 24$, при уровне значимости $\alpha = 0{,}05$ по прил. 8 находим критическую точку:

$$F_{\text{кр}}(0{,}05; 3; 24) = 3{,}01.$$

Поскольку $F_{\text{набл}} < F_{\text{кр}}$, основную гипотезу о равенстве групповых средних не отвергаем. Другими словами, групповые средние различаются незначаще. □

16.2. Однофакторный дисперсионный анализ для связных выборок

Пусть на количественный нормально распределенный признак X действует фактор F, который имеет p постоянных уровней $F_1, F_2, \ldots, F_p$. На каждом уровне проведены испытания **над одной и той же выборкой**, которая состоит из q элементов. Результаты наблюдений — числа x_{ij}, где i — номер испытания, $i = 1, 2, \ldots, q$, j — номер уровня фактора F, $j = 1, 2, \ldots, p$, — записывают в виде таблицы (табл. 26).

Таблица 26

Номер испытания i	Уровень фактора				Индивидуальные средние $\overline{x}_i$
	F_1	F_2	$\cdots$	F_p	
1	x_{11}	x_{12}	$\cdots$	x_{1p}	$\overline{x}_1$
2	x_{21}	x_{22}	$\cdots$	x_{2p}	$\overline{x}_2$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
q	x_{q1}	x_{q2}	$\cdots$	x_{qp}	$\overline{x}_q$
Групповое среднее $\overline{x}_{\text{гр}}$	$\overline{x}_{\text{гр. }1}$	$\overline{x}_{\text{гр. }2}$	$\cdots$	$\overline{x}_{\text{гр. }p}$	

В этом случае возникают две задачи:

1) при уровне значимости α проверить основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестные, но одинаковые;

2) на уровне значимости α проверить основную гипотезу о равенстве индивидуальных средних при условии, что индивидуальные генеральные дисперсии хотя и неизвестные, но одинаковые.

Чтобы решить эти задачи, находят:

а) *общую сумму* квадратов отклонений наблюдаемых значений признака от общего выборочного среднего:

$$S_{\text{общ}} = \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij} - \overline{x}_{\text{в}})^2;$$

б) *факторную сумму* квадратов отклонений групповых средних от общего выборочного среднего (характеризует рассеяние «между группами»):

$$S_{\text{факт}} = q\sum_{j=1}^{p}(\overline{x}_{\text{гр. } j} - \overline{x}_{\text{в}})^2;$$

в) *индивидуальную сумму* квадратов отклонений индивидуальных средних от общего выборочного среднего (характеризует рассеяние «между индивидами»):

$$S_{\text{инд}} = p\sum_{i=1}^{q}(\overline{x}_i - \overline{x}_{\text{в}})^2;$$

г) *остаточную сумму* по формуле

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}} - S_{\text{инд}}.$$

Поделив факторную, индивидуальную и остаточную суммы на соответствующее число степеней свободы

$$k_{\text{факт}} = p - 1, \qquad k_{\text{инд}} = q - 1, \qquad k_{\text{ост}} = (p-1)(q-1),$$

находят факторную, индивидуальную и остаточную дисперсии:

$$s^2_{\text{факт}} = \frac{S_{\text{факт}}}{p-1}; \qquad s^2_{\text{инд}} = \frac{S_{\text{инд}}}{q-1}; \qquad s^2_{\text{ост}} = \frac{S_{\text{ост}}}{(p-1)(q-1)}.$$

Сравнивают факторную и индивидуальную дисперсии с остаточной дисперсией с помощью *критерия Фишера—Снедекора*:

$$F^{\text{факт}}_{\text{набл}} = \frac{s^2_{\text{факт}}}{s^2_{\text{ост}}}, \qquad F^{\text{инд}}_{\text{набл}} = \frac{s^2_{\text{инд}}}{s^2_{\text{ост}}}.$$

Далее по таблице критических точек распределения Фишера—Снедекора (прил. 8) при заданном уровне значимости α и числе степеней свободы ($k_{\text{факт}}$ — числитель, $k_{\text{инд}}$ — числитель, $k_{\text{ост}}$ — знаменатель) находят критические точки $F^{\text{факт}}_{\text{кр}}(\alpha; k_{\text{факт}}, k_{\text{ост}})$ и $F^{\text{инд}}_{\text{кр}}(\alpha; k_{\text{инд}}, k_{\text{ост}})$.

Если $F_{\text{набл}}^{\text{факт}} < F_{\text{кр}}^{\text{факт}}$, различие групповых средних считают незначащим. Если $F_{\text{набл}}^{\text{факт}} > F_{\text{кр}}^{\text{факт}}$, различие групповых средних считают значащим.

Если $F_{\text{набл}}^{\text{инд}} < F_{\text{кр}}^{\text{инд}}$, различие индивидуальных средних считают незначащим. Если $F_{\text{набл}}^{\text{инд}} > F_{\text{кр}}^{\text{инд}}$, различие индивидуальных средних считают значащим.

Замечание 1. Однофакторный дисперсионный анализ для связных выборок требует не менее трёх уровней фактора и не менее двух элементов выборки, признаки которых измеряются на всех уровнях фактора.

Замечание 2. Результирующий признак должен быть нормально распределён в исследуемой выборке.

Замечание 3. Если факторная или индивидуальная дисперсия окажется меньше остаточной, сразу можно делать вывод о том, что нет оснований отвергать основную гипотезу. Другими словами, в таких случаях дальнейшие вычисления излишни.

Замечание 4. Если наблюдаемые значения x_{ij} кратны некоторому числу k (например, десятичные дроби кратны числу 10^{-d}, где d — количество цифр после запятой), целесообразно перейти к новым значениям:

$$y_{ij} = \frac{x_{ij}}{k}.$$

При этом факторная, индивидуальная и остаточная дисперсии уменьшатся в k^2 раз, но их отношение не изменится.

Пример 1. Проведены испытания на каждом из четырёх уровней фактора над одной и той же выборкой из шести элементов. Методом дисперсионного анализа при уровне значимости 0,01 проверить основную гипотезу о равенстве групповых средних и основную гипотезу о равенстве индивидуальных средних. Предполагается, что результирующий признак нормально распределён в исследуемой выборке. Результаты испытания приведены в табл. 27.

Решение. Найдём общее среднее:

$$\overline{x}_{\text{в}} = (5 + 11 + 7 + 13 + 3 + 9 + 8 + 12 + 5 + 12 + 5 + 14 + 6 + 11 + \\ + 6 + 17 + 4 + 8 + 5 + 11 + 7 + 15 + 5 + 17)/(6 \cdot 4) = 9.$$

Таблица 27

Номер испытания i	Уровень фактора				Индивидуальные средние $\overline{x}_i$
	F_1	F_2	F_3	F_4	
1	5	11	7	13	9
2	3	9	8	12	8
3	5	12	5	14	9
4	6	11	6	17	10
5	4	8	5	11	7
6	7	15	5	17	11
Групповое среднее $\overline{x}_{\text{гр}}$	5	11	6	14	

Для вычисления общей суммы квадратов отклонений наблюдаемых значений признаков от общего выборочного среднего составим расчётную табл. 28, в которую вместо значений x_{ij} запишем значения

$$y_{ij} = (x_{ij} - \overline{x}_{\text{в}})^2.$$

Таблица 28

Номер испытания i	Уровень фактора				$(\overline{x}_i - \overline{x}_{\text{в}})^2$
	F_1	F_2	F_3	F_4	
1	16	4	4	16	0
2	36	0	1	9	1
3	16	9	16	25	0
4	9	4	9	64	1
5	25	1	16	4	4
6	4	36	16	64	4
$(\overline{x}_{\text{гр}} - \overline{x}_{\text{в}})^2$	16	4	9	25	$S_{\text{общ}} = 404$

По табл. 28 определим:

$$S_{\text{общ}} = 404;$$

$$S_{\text{факт}} = q\sum_{j=1}^{p}(\overline{x}_{\text{гр.}\ j} - \overline{x}_{\text{в}})^2 = 6\cdot(16+4+9+25) = 324;$$

$$S_{\text{инд}} = p\sum_{i=1}^{q}(\overline{x}_i - \overline{x}_{\text{в}})^2 = 4\cdot(0+1+0+1+4+4) = 40.$$

Остаточная сумма квадратов отклонений

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}} - S_{\text{инд}} = 404 - 324 - 40 = 40.$$

Найдём число степеней свободы:

$$\begin{aligned} k_{\text{факт}} &= p - 1 = 4 - 1 = 3, \\ k_{\text{инд}} &= q - 1 = 6 - 1 = 5, \\ k_{\text{ост}} &= (p-1)(q-1) = 3 \cdot 5 = 15. \end{aligned}$$

Найдём факторную, индивидуальную и остаточную дисперсии:

$$\begin{aligned} s^2_{\text{факт}} &= \frac{S_{\text{факт}}}{p-1} = \frac{324}{3} = 108; \\ s^2_{\text{инд}} &= \frac{S_{\text{инд}}}{q-1} = \frac{40}{5} = 8; \\ s^2_{\text{ост}} &= \frac{S_{\text{ост}}}{(p-1)(q-1)} = \frac{40}{15} = \frac{8}{3}. \end{aligned}$$

Сравним факторную и индивидуальную дисперсии с остаточной дисперсией по критерию Фишера—Снедекора:

$$\begin{aligned} F^{\text{факт}}_{\text{набл}} &= \frac{s^2_{\text{факт}}}{s^2_{\text{ост}}} = \frac{108 \cdot 3}{8} = 40{,}5; \\ F^{\text{инд}}_{\text{набл}} &= \frac{s^2_{\text{инд}}}{s^2_{\text{ост}}} = \frac{8 \cdot 3}{8} = 3. \end{aligned}$$

Далее по таблице критических точек распределения Фишера—Снедекора (прил. 8) при заданном уровне значимости $\alpha = 0{,}01$ и числе степеней свободы ($k_{\text{факт}}$ — числитель, $k_{\text{инд}}$ — числитель, $k_{\text{ост}}$ — знаменатель) находим критические точки:

$$\begin{aligned} F^{\text{факт}}_{\text{кр}}(0{,}01; 3; 15) &= 5{,}42; \\ F^{\text{инд}}_{\text{кр}}(0{,}01; 5; 15) &= 4{,}56. \end{aligned}$$

Поскольку $F^{\text{факт}}_{\text{набл}} > F^{\text{факт}}_{\text{кр}}$, основную гипотезу о равенстве групповых средних отвергаем. Другими словами, различие групповых средних значаще.

Поскольку $F_{\text{набл}}^{\text{инд}} < F_{\text{кр}}^{\text{инд}}$, основную гипотезу о равенстве индивидуальных средних не отвергаем. Другими словами, различие индивидуальных средних незначаще. □

16.3. Двухфакторный дисперсионный анализ для несвязных выборок

Пусть на количественный нормально распределенный признак X действуют два фактора F и G, которые имеют соответственно p и g постоянных уровней $F_1, F_2, \ldots, F_p$ и $G_1, G_2, \ldots, G_g$. На каждой паре $(F_i; G_j)$ уровней осуществлено по q испытаний **над разными выборками**. Результаты испытаний — числа (x_{ijk}, где i — номер испытания, $i = 1, 2, \ldots, q$, j — номер уровня фактора F, $j = 1, 2, \ldots, p$, k — номер уровня фактора G, $k = 1, 2, \ldots, g$) — записывают в виде таблицы (табл. 29).

Таблица 29

Уровни фактора G	Уровни фактора F				Общая сумма
	F_1	F_2	$\cdots$	F_p	
G_1	$x_{111}, x_{211}, \ldots, x_{q11}$	$x_{121}, x_{221}, \ldots, x_{q21}$	$\cdots$	$x_{1p1}, x_{2p1}, \ldots, x_{qp1}$	S_{G_1}
Сумма	S_{11}	S_{21}	$\cdots$	S_{p1}	
G_2	$x_{112}, x_{212}, \ldots, x_{q12}$	$x_{122}, x_{222}, \ldots, x_{q22}$	$\cdots$	$x_{1p2}, x_{2p2}, \ldots, x_{qp2}$	S_{G_2}
Сумма	S_{12}	S_{22}	$\cdots$	S_{p2}	
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
G_g	$x_{11g}, x_{21g}, \ldots, x_{q1g}$	$x_{12g}, x_{22g}, \ldots, x_{q2g}$	$\cdots$	$x_{1pg}, x_{2pg}, \ldots, x_{qpg}$	S_{G_g}
Сумма	S_{1g}	S_{2g}	$\cdots$	S_{pg}	
Общая сумма	S_{F_1}	S_{F_2}	$\cdots$	S_{F_p}	S

В табл. 29 введены такие обозначения:

- $S_{jk} = \sum\limits_{i=1}^{q} x_{ijk}$ — сумма значений признака для выборки, которая подвергалась исследованию на уровне F_j фактора F и уровне G_k фактора G, $j = 1, 2, \ldots, p$, $k = 1, 2, \ldots, g$;

- $S_{F_j} = \sum\limits_{k=1}^{g} S_{jk}$ — сумма значений признака для выборок, которые подвергались исследованию на уровне F_j фактора F, $j = 1, 2, \ldots, p$;
- $S_{G_k} = \sum\limits_{j=1}^{p} S_{jk}$ — сумма значений признака для выборок, которые подвергались исследованию на уровне G_k фактора G, $k = 1, 2, \ldots, g$;
- $S = \sum\limits_{j=1}^{p} S_{F_j} = \sum\limits_{k=1}^{g} S_{G_k}$ — сумма значений признака для всех выборок, которые подвергались исследованию.

Кроме этих обозначений введём обозначения для средних:

$$\begin{aligned}
\overline{x}_{\text{в}} &= \frac{1}{qpg} S; \\
\overline{x}^{F}_{\text{гр. } j} &= \frac{1}{qg} S_{F_j}, \qquad j = 1, 2, \ldots, p; \\
\overline{x}^{G}_{\text{гр. } k} &= \frac{1}{qp} S_{G_k}, \qquad k = 1, 2, \ldots, g; \\
\overline{x}^{FG}_{\text{гр. } jk} &= \frac{1}{q} S_{jk}, \qquad j = 1, 2, \ldots, p, \quad k = 1, 2, \ldots, g.
\end{aligned}$$

В этом случае возникают три задачи:

1) для фактора F при уровне значимости α проверить основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;

2) для фактора G при уровне значимости α проверить основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;

3) при уровне значимости α проверить основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы для фактора F при разных градациях фактора G, и наоборот.

Чтобы решить эти задачи, находят:

а) *общую сумму* квадратов отклонений наблюдаемых значений признака от общего выборочного среднего:

$$S_{\text{общ}} = \sum_{k=1}^{g}\sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ijk} - \overline{x}_{\text{в}})^2;$$

б) *факторную сумму* квадратов отклонений групповых средних от общего выборочного среднего для фактора F:

$$S_{\text{факт}}^{F} = qg\sum_{j=1}^{p}(\overline{x}_{\text{гр. }j}^{F} - \overline{x}_{\text{в}})^2;$$

в) *факторную сумму* квадратов отклонений групповых средних от общего выборочного среднего для фактора G:

$$S_{\text{факт}}^{G} = qp\sum_{k=1}^{g}(\overline{x}_{\text{гр. }k}^{G} - \overline{x}_{\text{в}})^2;$$

г) *факторную сумму* квадратов отклонений групповых средних от общего выборочного среднего для факторов F и G:

$$S_{\text{факт}}^{FG} = q\sum_{k=1}^{g}\sum_{j=1}^{p}(\overline{x}_{\text{гр. }jk}^{FG} - \overline{x}_{\text{в}})^2 - S_{\text{факт}}^{F} - S_{\text{факт}}^{G};$$

д) *остаточную сумму*:

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}}^{F} - S_{\text{факт}}^{G} - S_{\text{факт}}^{FG}.$$

Поделив факторные и остаточную суммы на соответствующее число степеней свободы

$$k_{\text{факт}}^{F} = p - 1; \qquad k_{\text{факт}}^{G} = g - 1;$$
$$k_{\text{факт}}^{FG} = (p-1)(g-1); \qquad k_{\text{ост}} = pg(q-1),$$

находят факторные и остаточную дисперсии:

$$\left(s_{\text{факт}}^{F}\right)^2 = \frac{S_{\text{факт}}^{F}}{p-1}; \qquad \left(s_{\text{факт}}^{G}\right)^2 = \frac{S_{\text{факт}}^{G}}{g-1};$$
$$\left(s_{\text{факт}}^{FG}\right)^2 = \frac{S_{\text{факт}}^{FG}}{(p-1)(g-1)}; \qquad s_{\text{ост}}^2 = \frac{S_{\text{ост}}}{pg(q-1)}.$$

Сравнивают факторные дисперсии с остаточной дисперсией с помощью *критерия Фишера—Снедекора*:

$$F^F_{\text{набл}} = \frac{\left(s^F_{\text{факт}}\right)^2}{s^2_{\text{ост}}}, \qquad F^G_{\text{набл}} = \frac{\left(s^G_{\text{факт}}\right)^2}{s^2_{\text{ост}}}, \qquad F^{FG}_{\text{набл}} = \frac{\left(s^{FG}_{\text{факт}}\right)^2}{s^2_{\text{ост}}}.$$

Далее по таблице критических точек распределения Фишера—Снедекора (прил. 8) при заданном уровне значимости α и числе степеней свободы (k_1 — число степеней свободы, которое отвечает большей дисперсии, k_2 — число степеней свободы, которое отвечает меньшей дисперсии) находят критические точки $F^F_{\text{кр}}(\alpha; k_1, k_2)$, $F^G_{\text{кр}}(\alpha; k_1, k_2)$ и $F^{FG}_{\text{кр}}(\alpha; k_1, k_2)$.

Если $F^F_{\text{набл}} < F^F_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F считают незначащим. Если $F^F_{\text{набл}} > F^F_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F считают значащим.

Если $F^G_{\text{набл}} < F^G_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора G считают незначащим. Если $F^G_{\text{набл}} > F^G_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора G считают значащим.

Если $F^{FG}_{\text{набл}} < F^{FG}_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F при разных градациях фактора G считают незначащим. Если $F^{FG}_{\text{набл}} > F^{FG}_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F при разных градациях фактора G считают значащим.

Замечание 1. Двухфакторный дисперсионный анализ для несвязных выборок требует не менее двух уровней каждого фактора и не менее двух наблюдений на каждой паре уровней, причём для всех пар уровней количество наблюдений должно быть одинаковым.

Замечание 2. Результирующий признак должен быть нормально распределён в исследуемой выборке.

Замечание 3. Если некоторая факторная дисперсия окажется меньше остаточной, сразу можно делать вывод о том, что нет оснований отвергать основную гипотезу. Другими словами, в таком случае для данного фактора дальнейшие вычисления излишни.

Замечание 4. Если наблюдаемые значения x_{ij} кратны некоторому числу k (например, десятичные дроби кратны числу 10^{-d}, где d — количество цифр после запятой), целесообразно перейти к новым значениям:

$$y_{ij} = \frac{x_{ij}}{k}.$$

При этом факторные и остаточная дисперсии уменьшатся в k^2 раз, но их отношение не изменится.

Пример 1. В результате проведения исследований над разными выборками по двум факторам, каждый из которых имеет два уровня, получены некоторые данные.

Методом двухфакторного дисперсионного анализа при уровне значимости 0,01 проверить:

- для фактора F основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;
- для фактора G основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;
- основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы для фактора F при разных градациях фактора G, и наоборот.

Предполагается, что результирующий признак нормально распределён в исследуемой выборке. Результаты испытания приведены в табл. 30.

Решение. Для данной задачи значения параметров такие:

$$q = 4; \qquad p = 2; \qquad g = 2.$$

По данным табл. 30 найдём $\overline{x}_{\text{в}}$; $\overline{x}^F_{\text{гр. } j}$, $j = 1, 2$; $\overline{x}^G_{\text{гр. } k}$, $k = 1, 2$;

Таблица 30

Уровни фактора G	Уровни фактора F								Общая сумма
	F_1				F_2				
G_1	4	3	5	4	2	4	3	2	$S_{G_1} = 27$
Сумма	$S_{11} = 16$				$S_{21} = 11$				
G_2	4	5	6	4	4	5	5	5	$S_{G_2} = 38$
Сумма	$S_{12} = 19$				$S_{22} = 19$				
Общая сумма	$S_{F_1} = 35$				$S_{F_2} = 30$				$S = 65$

$\overline{x}^{FG}_{\text{гр. } jk}$, $j = 1, 2$, $k = 1, 2$:

$$\overline{x}_{\text{в}} = \frac{1}{qpg} S = \frac{1}{4 \cdot 2 \cdot 2} \cdot 65 = 4{,}0625;$$

$$\overline{x}^{F}_{\text{гр. } 1} = \frac{1}{qg} S_{F_1} = \frac{1}{4 \cdot 2} \cdot 35 = 4{,}375;$$

$$\overline{x}^{F}_{\text{гр. } 2} = \frac{1}{qg} S_{F_2} = \frac{1}{4 \cdot 2} \cdot 30 = 3{,}75;$$

$$\overline{x}^{G}_{\text{гр. } 1} = \frac{1}{qp} S_{G_1} = \frac{1}{4 \cdot 2} \cdot 27 = 3{,}375;$$

$$\overline{x}^{G}_{\text{гр. } 2} = \frac{1}{qp} S_{G_1} = \frac{1}{4 \cdot 2} \cdot 38 = 4{,}75;$$

$$\overline{x}^{FG}_{\text{гр. } 11} = \frac{1}{q} S_{11} = \frac{1}{4} \cdot 16 = 4;$$

$$\overline{x}^{FG}_{\text{гр. } 12} = \frac{1}{q} S_{12} = \frac{1}{4} \cdot 19 = 4{,}75;$$

$$\overline{x}^{FG}_{\text{гр. } 21} = \frac{1}{q} S_{21} = \frac{1}{4} \cdot 11 = 2{,}75;$$

$$\overline{x}^{FG}_{\text{гр.}22} = \frac{1}{q} S_{22} = \frac{1}{4} \cdot 19 = 4{,}75.$$

Вычислим:

а) общую сумму квадратов отклонений наблюдаемых значений признака от общего выборочного среднего:

$$
\begin{aligned}
S_{\text{общ}} = \sum_{k=1}^{g}\sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ijk} - \overline{x}_{\text{в}})^2 &= (4 - 4{,}0625)^2 + (3 - 4{,}0625)^2 + \\
&+ (5 - 4{,}0625)^2 + (4 - 4{,}0625)^2 + (2 - 4{,}0625)^2 + \\
&+ (4 - 4{,}0625)^2 + (3 - 4{,}0625)^2 + (2 - 4{,}0625)^2 + \\
&+ (4 - 4{,}0625)^2 + (5 - 4{,}0625)^2 + (6 - 4{,}0625)^2 + \\
&+ (4 - 4{,}0625)^2 + (4 - 4{,}0625)^2 + (5 - 4{,}0625)^2 + \\
&+ (5 - 4{,}0625)^2 + (5 - 4{,}0625)^2 = \\
&= 18{,}9375;
\end{aligned}
$$

б) факторную сумму квадратов отклонений групповых средних от общего выборочного среднего для фактора F:

$$
\begin{aligned}
S^F_{\text{факт}} &= qg\sum_{j=1}^{p}(\overline{x}^F_{\text{гр. } j} - \overline{x}_{\text{в}})^2 = \\
&= 4 \cdot 2 \cdot \left[(4{,}375 - 4{,}0625)^2 + (3{,}75 - 4{,}0625)^2\right] = \\
&= 1{,}5625;
\end{aligned}
$$

в) факторную сумму квадратов отклонений групповых средних от общего выборочного среднего для фактора G:

$$
\begin{aligned}
S^G_{\text{факт}} &= qp\sum_{k=1}^{g}(\overline{x}^G_{\text{гр. } k} - \overline{x}_{\text{в}})^2 = \\
&= 4 \cdot 2 \cdot \left[(3{,}375 - 4{,}0625)^2 + (4{,}75 - 4{,}0625)^2\right] = \\
&= 7{,}5625;
\end{aligned}
$$

г) факторную сумму квадратов отклонений групповых средних от общего выборочного среднего для факторов F и G:

$$S_{\text{факт}}^{FG} = q\sum_{k=1}^{g}\sum_{j=1}^{p}(\overline{x}_{\text{гр. }jk}^{FG} - \overline{x}_{\text{в}})^2 - S_{\text{факт}}^{F} - S_{\text{факт}}^{G} =$$

$$= 4 \cdot \left[(4 - 4{,}0625)^2 + (2{,}75 - 4{,}0625)^2 + (4{,}75 - 4{,}0625)^2 + \right.$$

$$\left. + (4{,}75 - 4{,}0625)^2\right] - 1{,}5625 - 7{,}5625 =$$

$$= 1{,}5625;$$

д) остаточную сумму:

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}}^{F} - S_{\text{факт}}^{G} - S_{\text{факт}}^{FG} =$$

$$= 18{,}9375 - 1{,}5625 - 7{,}5625 - 1{,}5625 =$$

$$= 8{,}25.$$

Поделив факторные и остаточную суммы на соответствующее число степеней свободы

$$k_{\text{факт}}^{F} = p - 1 = 2 - 1 = 1;$$

$$k_{\text{факт}}^{G} = g - 1 = 2 - 1 = 1;$$

$$k_{\text{факт}}^{FG} = (p-1)(g-1) = 1 \cdot 1 = 1;$$

$$k_{\text{ост}} = pg(q-1) = 2 \cdot 2 \cdot (4-1) = 12,$$

найдём факторные и остаточную дисперсии:

$$\left(s_{\text{факт}}^{F}\right)^2 = \frac{S_{\text{факт}}^{F}}{p-1} = \frac{1{,}5625}{1} = 1{,}5625;$$

$$\left(s_{\text{факт}}^{G}\right)^2 = \frac{S_{\text{факт}}^{G}}{g-1} = \frac{7{,}5625}{1} = 7{,}5625;$$

$$\left(s_{\text{факт}}^{FG}\right)^2 = \frac{S_{\text{факт}}^{FG}}{(p-1)(g-1)} = \frac{1{,}5625}{1} = 1{,}5625;$$

$$s_{\text{ост}}^2 = \frac{S_{\text{ост}}}{pg(q-1)} = \frac{8{,}25}{12} = 0{,}6875.$$

Сравним факторные дисперсии с остаточной дисперсией по критерию Фишера—Снедекора:

$$F^F_{\text{набл}} = \frac{\left(s^F_{\text{факт}}\right)2}{s^2_{\text{ост}}} = \frac{1{,}5625}{0{,}6875} \approx 2{,}27;$$

$$F^G_{\text{набл}} = \frac{\left(s^G_{\text{факт}}\right)2}{s^2_{\text{ост}}} = \frac{7{,}5625}{0{,}6875} = 11;$$

$$F^{FG}_{\text{набл}} = \frac{\left(s^{FG}_{\text{факт}}\right)2}{s^2_{\text{ост}}} = \frac{1{,}5625}{0{,}6875} \approx 2{,}27.$$

Далее по таблице критических точек распределения Фишера—Снедекора (прил. 8) при заданном уровне значимости $\alpha = 0{,}01$ и числе степеней свободы найдём критические точки:

$$F^F_{\text{кр}}(0{,}01; 1; 12) = 9{,}33;$$

$$F^G_{\text{кр}}(0{,}01; 1; 12) = 9{,}33;$$

$$F^{FG}_{\text{кр}}(0{,}01; 1; 12) = 9{,}33.$$

Поскольку $F^F_{\text{набл}} < F^F_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F считаем незначащим.

Поскольку $F^G_{\text{набл}} > F^G_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора G считаем значащим.

Поскольку $F^{FG}_{\text{набл}} < F^{FG}_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F при разных градациях фактора G считаем незначащим. □

16.4. Двухфакторный дисперсионный анализ для связных выборок

Пусть на количественный нормально распределенный признак X действуют два фактора F и G, которые имеют соответственно p и g постоянных уровней $F_1, F_2, \ldots, F_p$ и $G_1, G_2, \ldots, G_g$. На каждой паре $(F_i; G_j)$ уровней осуществлено по q испытаний **над одной и той же**

выборкой. Результаты наблюдений — числа (x_{ijk}, где i — номер испытания, $i = 1, 2, \ldots, q$, j — номер уровня фактора F, $j = 1, 2, \ldots, p$, k — номер уровня фактора G, $k = 1, 2, \ldots, g$) — записывают в виде табл. 31.

Таблица 31

Уровни фактора F	Уровни фактора G	Номер элемента выборки i				Общая сумма
		1	2	$\cdots$	q	
F_1	G_1	x_{111}	x_{211}	$\cdots$	x_{q11}	S_{11}
	G_2	x_{112}	x_{212}	$\cdots$	x_{q12}	S_{12}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	G_g	x_{11g}	x_{21g}	$\cdots$	x_{q1g}	S_{1g}
	Индивидуальные суммы по F_1	S_{1F_1}	S_{2F_1}	$\cdots$	S_{q_1}	S_{F_1}
F_2	G_1	x_{121}	x_{221}	$\cdots$	x_{q21}	S_{21}
	G_2	x_{122}	x_{222}	$\cdots$	x_{q22}	S_{22}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	G_g	x_{12g}	x_{22g}	$\cdots$	x_{q2g}	S_{2g}
	Индивидуальные суммы по F_2	S_{1F_2}	S_{2F_2}	$\cdots$	S_{q_2}	S_{F_2}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
F_p	G_1	x_{1p1}	x_{2p1}	$\cdots$	x_{qp1}	S_{p1}
	G_2	x_{1p2}	x_{2p2}	$\cdots$	x_{qp2}	S_{p2}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	G_g	x_{1pg}	x_{2pg}	$\cdots$	x_{qpg}	S_{pg}
	Индивидуальные суммы по F_p	S_{1F_p}	S_{2F_p}	$\cdots$	S_{q_p}	S_{F_p}
Индивидуальные суммы		S_1	S_2	$\cdots$	S_q	S

Введём обозначения, которые потребуются для дальнейших вычислений (некоторые из них приведены в табл. 31):

$$S_{jk} = \sum_{i=1}^{q} x_{ijk}, \qquad j = 1, 2, \ldots, p, \quad k = 1, 2, \ldots, g;$$

$$S_{iF_j} = \sum_{k=1}^{g} x_{ijk}, \qquad i = 1, 2, \ldots, q, \quad j = 1, 2, \ldots, p;$$

$$S_{iG_k} = \sum_{j=1}^{p} x_{ijk}, \qquad i = 1, 2, \ldots, q, \quad k = 1, 2, \ldots, g;$$

$$S_i = \sum_{j=1}^{p} S_{iF_j} = \sum_{k=1}^{g} S_{iG_k}, \qquad i = 1, 2, \ldots, q;$$

$$S_{F_j} = \sum_{i=1}^{q} S_{iF_j} = \sum_{k=1}^{g} S_{jk}, \qquad j = 1, 2, \ldots, p;$$

$$S_{G_k} = \sum_{i=1}^{q} S_{iG_k} = \sum_{j=1}^{p} S_{jk}, \qquad k = 1, 2, \ldots, g;$$

$$S = \sum_{i=1}^{q} S_i = \sum_{j=1}^{p} S_{F_j} = \sum_{k=1}^{g} S_{G_k}.$$

В этом случае возникают четыре задачи:

1) для фактора F при уровне значимости α проверить основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;

2) для фактора G при уровне значимости α проверить основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;

3) при уровне значимости α проверить основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы для фактора F при разных градациях фактора G, и наоборот;

4) при уровне значимости α проверить основную гипотезу о равенстве индивидуальных средних при условии, что индивидуальные генеральные дисперсии хотя и неизвестны, но одинаковы.

Чтобы решить эти задачи, находят такие суммы квадратов отклонений:

$$S_{\text{общ}} = \sum_{k=1}^{g} \sum_{j=1}^{p} \sum_{i=1}^{q} x_{ijk}^2 - \frac{1}{qpg} S^2;$$

$$S_{\text{инд}} = \frac{1}{pg}\sum_{i=1}^{q} S_i^2 - \frac{1}{qpg}S^2;$$

$$S_{\text{факт}}^F = \frac{1}{qg}\sum_{j=1}^{p} S_{F_j}^2 - \frac{1}{qpg}S^2;$$

$$S_{\text{факт}}^G = \frac{1}{qp}\sum_{k=1}^{g} S_{G_k}^2 - \frac{1}{qpg}S^2;$$

$$S_{\text{инд}}^F = \frac{1}{g}\sum_{j=1}^{p}\sum_{i=1}^{q} S_{iF_j}^2 - \frac{1}{qpg}S^2 - S_{\text{инд}} - S_{\text{факт}}^F;$$

$$S_{\text{инд}}^G = \frac{1}{p}\sum_{k=1}^{g}\sum_{i=1}^{q} S_{iG_k}^2 - \frac{1}{qpg}S^2 - S_{\text{инд}} - S_{\text{факт}}^G;$$

$$S_{\text{факт}}^{FG} = \frac{1}{q}\sum_{k=1}^{g}\sum_{j=1}^{p} S_{jk}^2 - \frac{1}{qpg}S^2 - S_{\text{факт}}^F - S_{\text{факт}}^G;$$

$$S_{\text{инд}}^{FG} = S_{\text{общ}} - S_{\text{инд}} - S_{\text{факт}}^F - S_{\text{факт}}^G - S_{\text{инд}}^F - S_{\text{инд}}^G - S_{\text{факт}}^{FG}.$$

Поделив вычисленные суммы на соответствующее число степеней свободы

$$\begin{aligned} k_{\text{инд}} &= q-1; & k_{\text{инд}}^F &= (q-1)(p-1);\\ k_{\text{факт}}^F &= p-1; & k_{\text{инд}}^G &= (q-1)(g-1);\\ k_{\text{факт}}^G &= g-1; & k_{\text{факт}}^{FG} &= (p-1)(g-1);\\ k_{\text{общ}} &= qpg-1; & k_{\text{инд}}^{FG} &= (q-1)(p-1)(g-1), \end{aligned}$$

находят дисперсии:

$$\begin{aligned} s_{\text{инд}}^2 &= \frac{S_{\text{инд}}}{q-1}; & \left(s_{\text{инд}}^F\right)^2 &= \frac{S_{\text{инд}}^F}{(q-1)(p-1)};\\ \left(s_{\text{факт}}^F\right)^2 &= \frac{S_{\text{факт}}^F}{p-1}; & \left(s_{\text{инд}}^G\right)^2 &= \frac{S_{\text{инд}}^G}{(q-1)(g-1)};\\ \left(s_{\text{факт}}^G\right)^2 &= \frac{S_{\text{факт}}^G}{g-1}; & \left(s_{\text{факт}}^{FG}\right)^2 &= \frac{S_{\text{факт}}^{FG}}{(p-1)(g-1)}; \end{aligned}$$

$$s^2_{\text{общ}} = \frac{S_{\text{общ}}}{qpg-1}; \qquad \left(s^{FG}_{\text{инд}}\right)^2 = \frac{S^{FG}_{\text{инд}}}{(q-1)(p-1)(g-1)}.$$

Сравнивают дисперсии с помощью *критерия Фишера—Снедекора*:

$$F^F_{\text{набл}} = \frac{\left(s^F_{\text{факт}}\right)^2}{\left(s^F_{\text{инд}}\right)^2}; \qquad F^G_{\text{набл}} = \frac{\left(s^G_{\text{факт}}\right)^2}{\left(s^G_{\text{инд}}\right)^2};$$

$$F^{FG}_{\text{набл}} = \frac{\left(s^{FG}_{\text{факт}}\right)^2}{\left(s^{FG}_{\text{инд}}\right)^2}; \qquad F^{\text{инд}}_{\text{набл}} = \frac{s^2_{\text{инд}}}{\left(s^{FG}_{\text{инд}}\right)^2}.$$

Далее по таблице критических точек распределения Фишера—Снедекора (прил. 8) при заданном уровне значимости α и числе степеней свободы находят критические точки:

$$F^F_{\text{кр}}(\alpha; k^F_{\text{факт}}, k^F_{\text{инд}}); \qquad F^G_{\text{кр}}(\alpha; k^G_{\text{факт}}, k^G_{\text{инд}});$$

$$F^{FG}_{\text{кр}}(\alpha; k^{FG}_{\text{факт}}, k^{FG}_{\text{инд}}); \qquad F^{\text{инд}}_{\text{кр}}(\alpha; k_{\text{инд}}, k^{FG}_{\text{инд}}).$$

Если $F^F_{\text{набл}} < F^F_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F считают незначащим. Если $F^F_{\text{набл}} > F^F_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F считают значащим.

Если $F^G_{\text{набл}} < F^G_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора G считают незначащим. Если $F^G_{\text{набл}} > F^G_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора G считают значащим.

Если $F^{FG}_{\text{набл}} < F^{FG}_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F при разных градациях фактора G считают незначащим. Если $F^{FG}_{\text{набл}} > F^{FG}_{\text{кр}}$, различие групповых средних квадратических отклонений для фактора F при разных градациях фактора G считают значащим.

Если $F^{\text{инд}}_{\text{набл}} < F^{\text{инд}}_{\text{кр}}$, различие индивидуальных средних квадратических отклонений при разных градациях факторов F и G считают незначащим. Если $F^{\text{инд}}_{\text{набл}} > F^{\text{инд}}_{\text{кр}}$, различие индивидуальных средних квадратических отклонений при разных градациях факторов F и G считают значащим.

Замечание 1. Двухфакторный дисперсионный анализ для связных выборок требует не менее двух уровней каждого фактора и не менее двух элементов выборки, признаки которых измеряются на всех уровнях всех пар факторов.

Замечание 2. Результирующий признак должен быть нормально распределён в исследуемой выборке.

Замечание 3. Если наблюдаемые значения x_{ij} кратны некоторому числу k (например, десятичные дроби кратны числу 10^{-d}, где d — количество цифр после запятой), целесообразно перейти к новым значениям:

$$y_{ij} = \frac{x_{ij}}{k}.$$

При этом факторные и остаточная дисперсии уменьшатся в k^2 раз, но их отношение не изменится.

Пример 1. В результате проведения исследований над одной и той же выборкой по двум факторам, каждый из которых имеет два уровня, получены некоторые данные.

Методом двухфакторного дисперсионного анализа при уровне значимости 0,01 проверить:

- для фактора F основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;
- для фактора G основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;
- основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы для фактора F при разных градациях фактора G, и наоборот;
- основную гипотезу о равенстве индивидуальных средних при условии, что индивидуальные генеральные дисперсии хотя и неизвестны, но одинаковы.

Предполагается, что результирующий признак нормально распределён в исследуемой выборке. Результаты испытания приведены в табл. 32.

Таблица 32

Уровни фактора F	Уровни фактора G	Номер элемента выборки i				Общая сумма
		1	2	3	4	
F_1	G_1	17	13	16	13	$S_{11} = 59$
	G_2	11	12	16	10	$S_{12} = 49$
	Индивидуальные суммы по F_1	$S_{1F_1} = 28$	$S_{2F_1} = 25$	$S_{3F_1} = 32$	$S_{4F_1} = 23$	$S_{F_1} = 108$
F_2	G_1	19	17	18	14	$S_{21} = 68$
	G_2	14	11	10	12	$S_{22} = 47$
	Индивидуальные суммы по F_2	$S_{1F_2} = 33$	$S_{2F_2} = 28$	$S_{3F_2} = 28$	$S_{4F_2} = 26$	$S_{F_2} = 115$
Индивидуальные суммы		$S_1 = 61$	$S_2 = 53$	$S_3 = 60$	$S_4 = 49$	$S = 223$

Решение. Для данной задачи значения параметров q, p, g такие:

$$q = 4; \qquad p = 2; \qquad g = 2.$$

Найдём значение сумм S_{i_k}, $i = 1, 2, 3, 4$, $k = 1, 2$; S_{G_k}, $k = 1, 2$, не приведенных в табл. 32:

$$S_{1G_1} = \sum_{j=1} 2x_{1j1} = 17 + 19 = 36;$$

$$S_{2G_1} = \sum_{j=1} 2x_{2j1} = 13 + 17 = 30;$$

$$S_{3G_1} = \sum_{j=1} 2x_{3j1} = 16 + 18 = 34;$$

$$S_{4G_1} = \sum_{j=1} 2x_{4j1} = 13 + 14 = 27;$$

$$S_{1G_2} = \sum_{j=1} 2x_{1j2} = 11 + 14 = 25;$$

$$S_{2G_2} = \sum_{j=1} 2x_{2j2} = 12 + 11 = 23;$$

$$S_{3G_2} = \sum_{j=1} 2x_{3j2} = 16 + 10 = 26;$$

$$S_{4G_2} = \sum_{j=1} 2x_{4j2} = 10 + 12 = 22;$$

$$S_{G_1} = \sum_{i=1} 4S_{iG_1} = 36 + 30 + 34 + 27 = 127;$$

$$S_{G_2} = \sum_{i=1} 4S_{iG_2} = 25 + 23 + 26 + 22 = 96.$$

Вычислим суммы квадратов отклонений:

$$\begin{aligned} S_{\text{общ}} = \sum_{k=1}^{g}\sum_{j=1}^{p}\sum_{i=1}^{q} x_{ijk}^2 - \frac{1}{qpg}S^2 = 17^2 + 13^2 + 16^2 + 13^2 + 11^2 + \\ + 12^2 + 16^2 + 10^2 + 19^2 + 17^2 + 18^2 + 14^2 + 14^2 + 11^2 + \\ + 10^2 + 12^2 - \frac{1}{4\cdot 2\cdot 2}\cdot 223^2 = \\ = 126{,}9375; \end{aligned}$$

$$\begin{aligned} S_{\text{инд}} &= \frac{1}{pg}\sum_{i=1}^{q} S_i^2 - \frac{1}{qpg}S^2 = \\ &= \frac{1}{2\cdot 2}\cdot\left(61^2 + 53^2 + 60^2 + 49^2\right) - \frac{1}{4\cdot 2\cdot 2}\cdot 223^2 = \\ &= 24{,}6875; \end{aligned}$$

$$\begin{aligned} S_{\text{факт}}^F &= \frac{1}{qg}\sum_{j=1}^{p} S_{F_j}^2 - \frac{1}{qpg}S^2 = \frac{1}{4\cdot 2}\cdot\left(108^2 + 115^2\right) - \frac{1}{4\cdot 2\cdot 2}\cdot 223^2 = \\ &= 3{,}0625; \end{aligned}$$

$$\begin{aligned} S_{\text{факт}}^G &= \frac{1}{qp}\sum_{k=1}^{g} S_{G_k}^2 - \frac{1}{qpg}S^2 = \frac{1}{4\cdot 2}\cdot\left(127^2 + 96^2\right) - \frac{1}{4\cdot 2\cdot 2}\cdot 223^2 = \\ &= 60{,}0625; \end{aligned}$$

$$\begin{aligned} S_{\text{инд}}^F &= \frac{1}{g}\sum_{j=1}^{p}\sum_{i=1}^{q} S_{i_j}^2 - \frac{1}{qpg}S^2 - S_{\text{инд}} - S_{\text{факт}}^F = \\ &= \frac{1}{2}\left(28^2 + 25^2 + 32^2 + 23^2 + 33^2 + 28^2 + 28^2 + 26^2\right) - \end{aligned}$$

$$-\frac{1}{4\cdot 2\cdot 2}\cdot 223^2 - 24{,}6875 - 3{,}0625 =$$
$$= 11{,}6875;$$

$$S_{\text{инд}}^{G} = \frac{1}{p}\sum_{k=1}^{g}\sum_{i=1}^{q} S_{i_k}^2 - \frac{1}{qpg}S^2 - S_{\text{инд}} - S_{\text{факт}}^{G} =$$
$$= \frac{1}{2}\cdot\left(36^2 + 30^2 + 34^2 + 27^2 + 25^2 + 23^2 + 26^2 + 22^2\right) -$$
$$-\frac{1}{4\cdot 2\cdot 2}\cdot 223^2 - 24{,}6875 - 60{,}0625 =$$
$$= 4{,}6875;$$

$$S_{\text{факт}}^{FG} = \frac{1}{q}\sum_{k=1}^{g}\sum_{j=1}^{p} S_{jk}^2 - \frac{1}{qpg}S^2 - S_{\text{факт}}^{F} - S_{\text{факт}}^{G} =$$
$$= \frac{1}{4}\cdot\left(59^2 + 49^2 + 68^2 + 47^2\right) -$$
$$-\frac{1}{4\cdot 2\cdot 2}\cdot 223^2 - 24{,}6875 - 3{,}0625 - 60{,}0625 =$$
$$= 7{,}5625;$$

$$S_{\text{инд}}^{FG} = S_{\text{общ}} - S_{\text{инд}} - S_{\text{факт}}^{F} - S_{\text{факт}}^{G} - S_{\text{инд}}^{F} - S_{\text{инд}}^{G} - S_{\text{факт}}^{FG} =$$
$$= 126{,}9375 - 24{,}6875 - 3{,}0625 - 60{,}0625 -$$
$$- 11{,}6875 - 4{,}6875 - 7{,}5625 =$$
$$= 15{,}1875.$$

Поделив вычисленные суммы на соответствующее число степеней свободы

$$k_{\text{инд}} = q - 1 = 4 - 1 = 3;$$
$$k_{\text{инд}}^{F} = (q-1)(p-1) = (4-1)(2-1) = 3;$$
$$k_{\text{факт}}^{F} = p - 1 = 2 - 1 = 1;$$
$$k_{\text{инд}}^{G} = (q-1)(g-1) = (4-1)(2-1) = 3;$$
$$k_{\text{факт}}^{G} = g - 1 = 2 - 1 = 1;$$
$$k_{\text{факт}}^{FG} = (p-1)(g-1) = (2-1)(2-1) = 1;$$

$$k_{\text{общ}} = qpg - 1 = 4 \cdot 2 \cdot 2 - 1 = 15;$$

$$k_{\text{инд}}^{FG} = (q-1)(p-1)(g-1) = (4-1)(2-1)(2-1) = 3,$$

найдем дисперсии:

$$s_{\text{инд}}^2 = \frac{S_{\text{инд}}}{q-1} = \frac{24{,}6875}{3} \approx 8{,}23;$$

$$\left(s_{\text{инд}}^F\right)^2 = \frac{S_{\text{инд}}^F}{(q-1)(p-1)} = \frac{11{,}6875}{3} \approx 3{,}90;$$

$$\left(s_{\text{факт}}^F\right)^2 = \frac{S_{\text{факт}}^F}{p-1} = \frac{3{,}0625}{1} = 3{,}0625;$$

$$\left(s_{\text{инд}}^G\right)^2 = \frac{S_{\text{инд}}^G}{(q-1)(g-1)} = \frac{4{,}6875}{3} \approx 1{,}5625;$$

$$\left(s_{\text{факт}}^G\right)^2 = \frac{S_{\text{факт}}^G}{g-1} = \frac{60{,}0625}{1} = 60{,}0625;$$

$$\left(s_{\text{факт}}^{FG}\right)^2 = \frac{S_{\text{факт}}^{FG}}{(p-1)(g-1)} = \frac{7{,}5625}{1} \approx 7{,}5625;$$

$$s_{\text{общ}}^2 = \frac{S_{\text{общ}}}{qpg-1} = \frac{126{,}9375}{15} = 8{,}4625;$$

$$\left(s_{\text{инд}}^{FG}\right)^2 = \frac{S_{\text{инд}}^{FG}}{(q-1)(p-1)(g-1)} = \frac{15{,}1875}{3} = 5{,}0625.$$

Сравним дисперсии по критерию Фишера—Снедекора:

$$F_{\text{набл}}^F = \frac{\left(s_{\text{факт}}^F\right)^2}{\left(s_{\text{инд}}^F\right)^2} = \frac{3{,}0625}{3{,}90} \approx 0{,}79;$$

$$F_{\text{набл}}^G = \frac{\left(s_{\text{факт}}^G\right)^2}{\left(s_{\text{инд}}^G\right)^2} = \frac{60{,}0625}{1{,}5625} = 38{,}44;$$

$$F_{\text{набл}}^{FG} = \frac{\left(s_{\text{факт}}^{FG}\right)^2}{\left(s_{\text{инд}}^{FG}\right)^2} = \frac{7{,}5625}{5{,}0625} \approx 1{,}49;$$

$$F_{\text{набл}}^{\text{инд}} = \frac{s_{\text{инд}}^2}{\left(s_{\text{инд}}^{FG}\right)^2} = \frac{8{,}23}{5{,}0625} \approx 1{,}63.$$

Далее по таблице критических точек распределения Фишера—Снедекора (прил. 8) при заданном уровне значимости $\alpha = 0{,}01$ и числе степеней свободы найдём критические точки:

$$F_{\text{кр}}^{F}(0{,}01; 1; 3) = 34{,}12;$$
$$F_{\text{кр}}^{G}(0{,}01; 1; 3) = 34{,}12;$$
$$F_{\text{кр}}^{FG}(0{,}01; 1; 3) = 34{,}12;$$
$$F_{\text{кр}}^{\text{инд}}(0{,}01; 3; 3) = 29{,}46.$$

Поскольку $F_{\text{набл}}^{F} < F_{\text{кр}}^{F}$, различие групповых средних квадратических отклонений для фактора F считаем незначимым.

Поскольку $F_{\text{набл}}^{G} > F_{\text{кр}}^{G}$, различие групповых средних квадратических отклонений для фактора G считаем значимым.

Поскольку $F_{\text{набл}}^{FG} < F_{\text{кр}}^{FG}$, различие групповых средних квадратических отклонений для фактора F при разных градациях фактора G считаем незначимым.

Поскольку $F_{\text{набл}}^{\text{инд}} < F_{\text{кр}}^{\text{инд}}$, различие групповых индивидуальных средних квадратических отклонений при разных градациях факторов F и G считаем незначимым. □

Задачи к разделу 16

Задача 1. Проведено по семь испытаний на каждом из четырёх уровней фактора над разными выборками. Методом дисперсионного анализа при уровне значимости 0,01 проверить основную гипотезу о равенстве групповых средних. Предполагается, что выборки получены из нормальных совокупностей с одинаковыми дисперсиями. Результаты испытания приведены в табл. 33.

Ответ. Основная гипотеза не отклоняется.

Задача 2. Проведены испытания на каждом из четырёх уровней фактора над одной и той же выборкой из шести элементов. Методом дисперсионного анализа при уровне значимости 0,05 проверить основную гипотезу о равенстве групповых средних и основную гипотезу о равенстве индивидуальных средних. Предполагается, что

Таблица 33

Номер испытания i	Уровень фактора			
	F_1	F_2	F_3	F_4
1	12	11	10	13
2	15	12	13	15
3	16	13	14	17
4	19	15	16	19
5	20	16	18	22
6	21	18	20	23
7	23	20	21	24
Групповое среднее $\overline{x}_{\text{гр}}$	18	15	16	19

результирующий признак нормально распределён в исследуемой выборке. Результаты испытания приведены в табл. 34.

Таблица 34

Номер испытания i	Уровень фактора				Индивидуальные средние $\overline{x}_i$
	F_1	F_2	F_3	F_4	
1	22	23	22	21	22
2	25	24	22	24	23,75
3	19	20	22	25	21,5
4	14	17	20	24	18,75
5	20	21	21	21	20,75
6	14	15	19	23	17,75
Групповое среднее $\overline{x}_{\text{гр}}$	19	20	21	23	

Ответ. Основная гипотеза о равенстве групповых средних не отвергается. Основная гипотеза о равенстве индивидуальных средних отвергается.

Задача 3. В результате проведения исследований над разными выборками по двум факторам, каждый из которых имеет по два уровня, получены некоторые данные.

Методом двухфакторного дисперсионного анализа при уровне значимости 0,05 проверить:

- для фактора F основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;

• для фактора G основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;

• основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы для фактора F при разных градациях фактора G, и наоборот.

Предполагается, что результирующий признак нормально распределён в исследуемой выборке. Результаты испытания приведены в табл. 35.

Таблица 35

Уровни фактора G	Уровни фактора F								Общая сумма
	F_1				F_2				
G_1	7	5	6	4	5	8	7	6	$S_{G_1} = 48$
Сумма	$S_{11} = 22$				$S_{21} = 26$				
G_2	4	3	5	4	7	5	6	5	$S_{G_2} = 39$
Сумма	$S_{12} = 16$				$S_{22} = 23$				
Общая сумма	$S_{F_1} = 38$				$S_{F_2} = 49$				$S = 87$

Ответ. Основная гипотеза о равенстве групповых средних для фактора F отвергается. В других случаях основная гипотеза о равенстве групповых средних не отвергается.

Задача 4. В результате проведения исследований над одной и той же выборкой по двум факторам, каждый из которых имеет по два уровня, получены некоторые данные.

Методом двухфакторного дисперсионного анализа при уровне значимости 0,05 проверить:

• для фактора F основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;

• для фактора G основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы;

• основную гипотезу о равенстве групповых средних при условии, что групповые генеральные дисперсии хотя и неизвестны, но одинаковы для фактора F при разных градациях фактора G, и наоборот;

• основную гипотезу о равенстве индивидуальных средних при условии, что индивидуальные генеральные дисперсии хотя и неизвестны, но одинаковы.

Предполагается, что результирующий признак нормально распределён в исследуемой выборке. Результаты испытания приведены в табл. 36.

Таблица 36

Уровни фактора F	Уровни фактора G	Номер элемента выборки i				Общая сумма
		1	2	3	4	
F_1	G_1	37	42	31	25	$S_{11} = 135$
	G_2	32	39	28	19	$S_{12} = 118$
	Индивидуальные суммы по F_1	$S_{1F_1} = 69$	$S_{2F_1} = 81$	$S_{3F_1} = 59$	$S_{4F_1} = 44$	$S_{F_1} = 253$
F_2	G_1	45	47	35	26	$S_{21} = 153$
	G_2	33	35	29	22	$S_{22} = 119$
	Индивидуальные суммы по F_2	$S_{1F_2} = 78$	$S_{2F_2} = 82$	$S_{3F_2} = 64$	$S_{4F_2} = 48$	$S_{F_2} = 272$
Индивидуальные суммы		$S_1 = 147$	$S_2 = 163$	$S_3 = 123$	$S_4 = 92$	$S = 525$

Ответ. Основная гипотеза о равенстве групповых средних для фактора F и фактора F при разных градациях фактора G не отвергается. В других случаях основная гипотеза отвергается.

Приложения

Приложение 1

Таблица значений функции $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x \mathrm{e}^{-\frac{z^2}{2}}\,\mathrm{d}z$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,22	0,0871	0,44	0,1700	0,66	0,2454	0,88	0,3106
0,01	0,0040	0,23	0,0910	0,45	0,1736	0,67	0,2486	0,89	0,3133
0,02	0,0080	0,24	0,0948	0,46	0,1772	0,68	0,2517	0,90	0,3159
0,03	0,0120	0,25	0,0987	0,47	0,1808	0,69	0,2549	0,91	0,3186
0,04	0,0160	0,26	0,1026	0,48	0,1844	0,70	0,2580	0,92	0,3212
0,05	0,0199	0,27	0,1064	0,49	0,1879	0,71	0,2611	0,93	0,3238
0,06	0,0239	0,28	0,1103	0,50	0,1915	0,72	0,2642	0,94	0,3264
0,07	0,0279	0,29	0,1141	0,51	0,1950	0,73	0,2673	0,95	0,3289
0,08	0,0319	0,30	0,1179	0,52	0,1985	0,74	0,2704	0,96	0,3315
0,09	0,0359	0,31	0,1217	0,53	0,2019	0,75	0,2734	0,97	0,3340
0,10	0,0398	0,32	0,1255	0,54	0,2054	0,76	0,2764	0,98	0,3365
0,11	0,0438	0,33	0,1293	0,55	0,2088	0,77	0,2794	0,99	0,3389
0,12	0,0478	0,34	0,1331	0,56	0,2123	0,78	0,2823	1,00	0,3413
0,13	0,0517	0,35	0,1368	0,57	0,2157	0,79	0,2852	1,01	0,3438
0,14	0,0557	0,36	0,1406	0,58	0,2190	0,80	0,2881	1,02	0,3461
0,15	0,0596	0,37	0,1443	0,59	0,2224	0,81	0,2910	1,03	0,3485
0,16	0,0636	0,38	0,1480	0,60	0,2257	0,82	0,2939	1,04	0,3508
0,17	0,0675	0,39	0,1517	0,61	0,2291	0,83	0,2967	1,05	0,3531
0,18	0,0714	0,40	0,1554	0,62	0,2324	0,84	0,2995	1,06	0,3554
0,19	0,0753	0,41	0,1591	0,63	0,2357	0,85	0,3023	1,07	0,3577
0,20	0,0793	0,42	0,1628	0,64	0,2389	0,86	0,3051	1,08	0,3599
0,21	0,0832	0,43	0,1664	0,65	0,2422	0,87	0,3078	1,09	0,3621

Окончание прил. 1

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,10	0,3643	1,47	0,4292	1,84	0,4671	2,21	0,4864	2,66	0,4961
1,11	0,3665	1,48	0,4306	1,85	0,4678	2,22	0,4868	2,68	0,4963
1,12	0,3686	1,49	0,4319	1,86	0,4686	2,23	0,4871	2,70	0,4965
1,13	0,3708	1,50	0,4332	1,87	0,4693	2,24	0,4875	2,72	0,4967
1,14	0,3729	1,51	0,4345	1,88	0,4699	2,25	0,4878	2,74	0,4969
1,15	0,3749	1,52	0,4357	1,89	0,4706	2,26	0,4881	2,76	0,4971
1,16	0,3770	1,53	0,4370	1,90	0,4713	2,27	0,4884	2,78	0,4973
1,17	0,3790	1,54	0,4382	1,91	0,4719	2,28	0,4887	2,80	0,4974
1,18	0,3810	1,55	0,4394	1,92	0,4726	2,29	0,4890	2,82	0,4976
1,19	0,3830	1,56	0,4406	1,93	0,4732	2,30	0,4893	2,84	0,4977
1,20	0,3849	1,57	0,4418	1,94	0,4738	2,31	0,4896	2,86	0,4979
1,21	0,3869	1,58	0,4429	1,95	0,4744	2,32	0,4898	2,88	0,4980
1,22	0,3888	1,59	0,4441	1,96	0,4750	2,33	0,4901	2,90	0,4981
1,23	0,3907	1,60	0,4452	1,97	0,4756	2,34	0,4904	2,92	0,4982
1,24	0,3925	1,61	0,4463	1,98	0,4761	2,35	0,4906	2,94	0,4984
1,25	0,3944	1,62	0,4474	1,99	0,4767	2,36	0,4909	2,96	0,4985
1,26	0,3962	1,63	0,4484	2,00	0,4772	2,37	0,4911	2,98	0,4986
1,27	0,3980	1,64	0,4495	2,01	0,4778	2,38	0,4913	3,00	0,4987
1,28	0,3997	1,65	0,4505	2,02	0,4783	2,39	0,4916	3,05	0,4989
1,29	0,4015	1,66	0,4515	2,03	0,4788	2,40	0,4918	3,10	0,49903
1,30	0,4032	1,67	0,4525	2,04	0,4793	2,41	0,4920	3,15	0,49918
1,31	0,4049	1,68	0,4535	2,05	0,4798	2,42	0,4922	3,20	0,49931
1,32	0,4066	1,69	0,4545	2,06	0,4803	2,43	0,4925	3,25	0,49942
1,33	0,4082	1,70	0,4554	2,07	0,4808	2,44	0,4927	3,30	0,49952
1,34	0,4099	1,71	0,4564	2,08	0,4812	2,45	0,4929	3,35	0,49960
1,35	0,4115	1,72	0,4573	2,09	0,4817	2,46	0,4931	3,40	0,49966
1,36	0,4131	1,73	0,4582	2,10	0,4821	2,47	0,4932	3,50	0,49977
1,37	0,4147	1,74	0,4591	2,11	0,4826	2,48	0,4934	3,60	0,49984
1,38	0,4162	1,75	0,4599	2,12	0,4830	2,49	0,4936	3,70	0,49989
1,39	0,4177	1,76	0,4608	2,13	0,4834	2,50	0,4938	3,80	0,499928
1,40	0,4192	1,77	0,4616	2,14	0,4838	2,52	0,4941	3,90	0,499952
1,41	0,4207	1,78	0,4625	2,15	0,4842	2,54	0,4945	4,00	0,499968
1,42	0,4222	1,79	0,4633	2,16	0,4846	2,56	0,4948	4,20	0,499987
1,43	0,4236	1,80	0,4641	2,17	0,4850	2,58	0,4951	4,40	0,4999946
1,44	0,4251	1,81	0,4649	2,18	0,4854	2,60	0,4953	4,60	0,4999979
1,45	0,4265	1,82	0,4656	2,19	0,4857	2,62	0,4956	4,80	0,4999992
1,46	0,4279	1,83	0,4664	2,20	0,4861	2,64	0,4959	5,00	0,4999997

Приложение 2

Таблица значений функции $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013

Окончание прил. 2

x	0	1	2	3	4	5	6	7	8	9
3,4	0,0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001
4,0	0001	0001	0001	0001	0001	0001	0001	0001	0001	0001
4,1	0001	0001	0001	0001	0001	0001	0001	0001	0001	0001
4,2	0001	0001	0001	0001	0000	0000	0000	0000	0000	0000

Приложение 3

Таблица значений функции Пуассона $P_n(m) = \dfrac{\lambda^m}{m!}\mathrm{e}^{-\lambda}$

m \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
0	0,9048	8187	7408	6703	6065	5488	4966	4493	4066	3679
1	0905	1637	2222	2681	3033	3293	3476	3595	3659	3679
2	0045	0164	0333	0536	0758	0988	1217	1438	1647	1839
3	0002	0011	0033	0072	0126	0198	0284	0383	0494	0613
4	0000	0001	0003	0007	0016	0030	0050	0077	0111	0153
5	0000	0000	0000	0001	0002	0004	0007	0012	0020	0031
6	0000	0000	0000	0000	0000	0000	0001	0002	0003	0005

m \ λ	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2
0	0,3329	3012	2725	2466	2231	2019	1827	1653	1496	1353
1	3662	3614	3543	3452	3347	3230	3106	2975	2842	2707
2	2014	2169	2303	2417	2510	2584	2640	2678	2700	2707
3	0738	0867	0998	1128	1255	1378	1496	1607	1710	1804
4	0203	0260	0324	0395	0471	0551	0636	0723	0812	0902
5	0045	0062	0084	0111	0141	0176	0216	0260	0309	0361
6	0008	0012	0018	0026	0035	0047	0061	0078	0098	0120
7	0001	0002	0003	0005	0008	0011	0015	0020	0027	0034
8	0000	0000	0001	0001	0001	0002	0003	0005	0006	0009
9	0000	0000	0000	0000	0000	0000	0001	0001	0001	0002

Окончание прил. 3

m \ λ	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3
0	0,1225	1108	1003	0907	0821	0743	0672	0608	0550	0498
1	2572	2438	2306	2177	2052	1931	1815	1703	1596	1494
2	2700	2681	2652	2613	2565	2510	2450	2384	2314	2240
3	1890	1966	2033	2090	2138	2176	2205	2225	2237	2240
4	0992	1082	1169	1254	1336	1414	1488	1557	1622	1680
5	0417	0476	0538	0602	0668	0735	0804	0872	0940	1008
6	0146	0174	0206	0241	0278	0319	0362	0407	0455	0504
7	0044	0055	0068	0083	0099	0118	0139	0163	0188	0216
8	0011	0015	0019	0025	0031	0038	0047	0057	0068	0081
9	0003	0004	0005	0007	0009	0011	0014	0018	0022	0027
10	0001	0001	0001	0002	0002	0003	0004	0005	0006	0008
11	0000	0000	0000	0000	0000	0001	0001	0001	0002	0002

m \ λ	3	3,5	4	4,5	5	6	7	8	9	10
0	0,0498	0302	0183	0111	0067	0025	0009	0003	0001	0000
1	1494	1057	0733	0500	0337	0149	0064	0027	0011	0005
2	2240	1850	1465	1125	0842	0446	0223	0107	0050	0023
3	2240	2158	1954	1687	1404	0892	0521	0286	0150	0076
4	1680	1888	1954	1898	1755	1339	0912	0573	0337	0189
5	1008	1322	1563	1708	1755	1606	1277	0916	0607	0378
6	0504	0771	1042	1281	1462	1606	1490	1221	0911	0631
7	0216	0385	0595	0824	1044	1377	1490	1396	1171	0901
8	0081	0169	0298	0463	0653	1033	1304	1396	1318	1126
9	0027	0066	0132	0232	0363	0688	1014	1241	1318	1251
10	0008	0023	0053	0104	0181	0413	0710	0993	1186	1251
11	0002	0007	0019	0043	0082	0225	0452	0722	0970	1137
12	0001	0002	0006	0016	0034	0113	0263	0481	0728	0948
13	0000	0001	0002	0006	0013	0052	0142	0296	0504	0729
14	0000	0000	0001	0002	0005	0022	0071	0169	0324	0521
15	0000	0000	0000	0001	0002	0009	0033	0090	0194	0347
16	0000	0000	0000	0000	0000	0003	0014	0045	0109	0217
17	0000	0000	0000	0000	0000	0001	0006	0021	0058	0128
18	0000	0000	0000	0000	0000	0000	0002	0009	0029	0071
19	0000	0000	0000	0000	0000	0000	0001	0004	0014	0037
20	0000	0000	0000	0000	0000	0000	0000	0002	0006	0019
21	0000	0000	0000	0000	0000	0000	0000	0001	0003	0009
22	0000	0000	0000	0000	0000	0000	0000	0000	0001	0004

Приложение 4

Таблица значений функции $t_\alpha = t(\alpha, n)$

n \ α	0,95	0,99	0,999	n \ α	0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,991	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Приложение 5

Таблица значений функции $q = q(\alpha, n)$

n \ α	0,95	0,99	0,999	n \ α	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	1,62	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	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	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

Приложение 6

Критические точки распределения Стьюдента

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,71	31,82	63,66	318,29	636,58
2	2,92	4,30	6,96	9,92	22,33	31,60
3	2,35	3,18	4,54	5,84	10,21	12,92
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,02	2,57	3,36	4,03	5,89	6,87
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,41
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,02	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,97
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,72	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,50	3,79
24	1,71	2,06	2,49	2,80	3,47	3,75
25	1,71	2,06	2,49	2,79	3,45	3,73
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)					

Приложение 7

Критические точки распределения χ^2

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63	5,02	3,84	0,0039	0,00098	0,00016
2	9,21	7,38	5,99	0,103	0,051	0,020
3	11,34	9,35	7,81	0,352	0,216	0,115
4	13,28	11,14	9,49	0,711	0,484	0,297
5	15,09	12,83	11,07	1,15	0,831	0,554
6	16,81	14,45	12,59	1,64	1,24	0,872
7	18,48	16,01	14,07	2,17	1,69	1,24
8	20,09	17,53	15,51	2,73	2,18	1,65
9	21,67	19,02	16,92	3,33	2,70	2,09
10	23,21	20,48	18,31	3,94	3,25	2,56
11	24,73	21,92	19,68	4,57	3,82	3,05
12	26,22	23,34	21,03	5,23	4,40	3,57
13	27,69	24,74	22,36	5,89	5,01	4,11
14	29,14	26,12	23,68	6,57	5,63	4,66
15	30,58	27,49	25,00	7,26	6,26	5,23
16	32,00	28,85	26,30	7,96	6,91	5,81
17	33,41	30,19	27,59	8,67	7,56	6,41
18	34,81	31,53	28,87	9,39	8,23	7,01
19	36,19	32,85	30,14	10,12	8,91	7,63
20	37,57	34,17	31,41	10,85	9,59	8,26
21	38,93	35,48	32,67	11,59	10,28	8,90
22	40,29	36,78	33,92	12,34	10,98	9,54
23	41,64	38,08	35,17	13,09	11,69	10,20
24	42,98	39,36	36,42	13,85	12,40	10,86
25	44,31	40,65	37,65	14,61	13,12	11,52
26	45,64	41,92	38,89	15,38	13,84	12,20
27	46,96	43,19	40,11	16,15	14,57	12,88
28	48,28	44,46	41,34	16,93	15,31	13,56
29	49,59	45,72	42,56	17,71	16,05	14,26
30	50,89	46,98	43,77	18,49	16,79	14,95

Приложение 8

Критические точки распределения F Фишера—Снедекора

Уровень значимости $\alpha = 0{,}01$

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6083	6107
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,41	99,42
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,96	9,89
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,06	2,99
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84
31	7,53	5,36	4,48	3,99	3,67	3,45	3,28	3,15	3,04	2,96	2,88	2,82
32	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02	2,93	2,86	2,80
33	7,47	5,31	4,44	3,95	3,63	3,41	3,24	3,11	3,00	2,91	2,84	2,78
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98	2,89	2,82	2,76

Продолжение прил. 8

Уровень значимости $\alpha = 0{,}01$												
$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88	2,80	2,74
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95	2,86	2,79	2,72
37	7,37	5,23	4,36	3,87	3,56	3,33	3,17	3,04	2,93	2,84	2,77	2,71
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92	2,83	2,75	2,69
39	7,33	5,19	4,33	3,84	3,53	3,30	3,14	3,01	2,90	2,81	2,74	2,68
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66
41	7,30	5,16	4,30	3,81	3,50	3,28	3,11	2,98	2,87	2,79	2,71	2,65
42	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97	2,86	2,78	2,70	2,64
43	7,26	5,14	4,27	3,79	3,48	3,25	3,09	2,96	2,85	2,76	2,69	2,63
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	2,75	2,68	2,62
45	7,23	5,11	4,25	3,77	3,45	3,23	3,07	2,94	2,83	2,74	2,67	2,61
46	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93	2,82	2,73	2,66	2,60
47	7,21	5,09	4,23	3,75	3,43	3,21	3,05	2,92	2,81	2,72	2,65	2,59
48	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	2,71	2,64	2,58
49	7,18	5,07	4,21	3,73	3,42	3,19	3,03	2,90	2,79	2,71	2,63	2,57
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,63	2,56
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,59	2,53
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50
65	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,80	2,69	2,61	2,53	2,47
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45
75	6,99	4,90	4,05	3,58	3,27	3,05	2,89	2,76	2,65	2,57	2,49	2,43
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42
85	6,94	4,86	4,02	3,55	3,24	3,02	2,86	2,73	2,62	2,54	2,46	2,40
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39
95	6,91	4,84	3,99	3,52	3,22	3,00	2,83	2,70	2,60	2,51	2,44	2,38
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,66	2,55	2,47	2,39	2,33
150	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,37	2,31
175	6,78	4,73	3,90	3,43	3,12	2,91	2,74	2,61	2,51	2,42	2,35	2,29
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38	2,31	2,24
400	6,70	4,66	3,83	3,37	3,06	2,85	2,68	2,56	2,45	2,37	2,29	2,23
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,28	2,22
700	6,67	4,64	3,81	3,35	3,04	2,83	2,66	2,54	2,43	2,35	2,27	2,21
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27	2,20
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25	2,18

Продолжение прил. 8

Уровень значимости $\alpha = 0{,}01$												
$k_2 \backslash k_1$	14	16	20	24	30	40	50	75	100	200	500	∞
1	6143	6170	6209	6234	6260	6286	6302	6324	6334	6350	6360	6366
2	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,48	99,49	99,49	99,50	99,50
3	26,92	26,83	26,69	26,60	26,50	26,41	26,35	26,28	26,24	26,18	26,15	26,13
4	14,25	14,15	14,02	13,93	13,84	13,75	13,69	13,61	13,58	13,52	13,49	13,46
5	9,77	9,68	9,55	9,47	9,38	9,29	9,24	9,17	9,13	9,08	9,04	9,02
6	7,60	7,52	7,40	7,31	7,23	7,14	7,09	7,02	6,99	6,93	6,90	6,88
7	6,36	6,28	6,16	6,07	5,99	5,91	5,86	5,79	5,75	5,70	5,67	5,65
8	5,56	5,48	5,36	5,28	5,20	5,12	5,07	5,00	4,96	4,91	4,88	4,86
9	5,01	4,92	4,81	4,73	4,65	4,57	4,52	4,45	4,41	4,36	4,33	4,31
10	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91
11	4,29	4,21	4,10	4,02	3,94	3,86	3,81	3,74	3,71	3,66	3,62	3,60
12	4,05	3,97	3,86	3,78	3,70	3,62	3,57	3,50	3,47	3,41	3,38	3,36
13	3,86	3,78	3,66	3,59	3,51	3,43	3,38	3,31	3,27	3,22	3,19	3,17
14	3,70	3,62	3,51	3,43	3,35	3,27	3,22	3,15	3,11	3,06	3,03	3,00
15	3,56	3,49	3,37	3,29	3,21	3,13	3,08	3,01	2,98	2,92	2,89	2,87
16	3,45	3,37	3,26	3,18	3,10	3,02	2,97	2,90	2,86	2,81	2,78	2,75
17	3,35	3,27	3,16	3,08	3,00	2,92	2,87	2,80	2,76	2,71	2,68	2,65
18	3,27	3,19	3,08	3,00	2,92	2,84	2,78	2,71	2,68	2,62	2,59	2,57
19	3,19	3,12	3,00	2,92	2,84	2,76	2,71	2,64	2,60	2,55	2,51	2,49
20	3,13	3,05	2,94	2,86	2,78	2,69	2,64	2,57	2,54	2,48	2,44	2,42
21	3,07	2,99	2,88	2,80	2,72	2,64	2,58	2,51	2,48	2,42	2,38	2,36
22	3,02	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,36	2,33	2,31
23	2,97	2,89	2,78	2,70	2,62	2,54	2,48	2,41	2,37	2,32	2,28	2,26
24	2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,37	2,33	2,27	2,24	2,21
25	2,89	2,81	2,70	2,62	2,54	2,45	2,40	2,33	2,29	2,23	2,19	2,17
26	2,86	2,78	2,66	2,58	2,50	2,42	2,36	2,29	2,25	2,19	2,16	2,13
27	2,82	2,75	2,63	2,55	2,47	2,38	2,33	2,26	2,22	2,16	2,12	2,10
28	2,79	2,72	2,60	2,52	2,44	2,35	2,30	2,23	2,19	2,13	2,09	2,06
29	2,77	2,69	2,57	2,49	2,41	2,33	2,27	2,20	2,16	2,10	2,06	2,03
30	2,74	2,66	2,55	2,47	2,39	2,30	2,25	2,17	2,13	2,07	2,03	2,01
31	2,72	2,64	2,52	2,45	2,36	2,27	2,22	2,14	2,11	2,04	2,01	1,98
32	2,70	2,62	2,50	2,42	2,34	2,25	2,20	2,12	2,08	2,02	1,98	1,96
33	2,68	2,60	2,48	2,40	2,32	2,23	2,18	2,10	2,06	2,00	1,96	1,93
34	2,66	2,58	2,46	2,38	2,30	2,21	2,16	2,08	2,04	1,98	1,94	1,91
35	2,64	2,56	2,44	2,36	2,28	2,19	2,14	2,06	2,02	1,96	1,92	1,89
36	2,62	2,54	2,43	2,35	2,26	2,18	2,12	2,04	2,00	1,94	1,90	1,87

Продолжение прил. 8

Уровень значимости $\alpha = 0{,}01$												
$k_2 \backslash k_1$	14	16	20	24	30	40	50	75	100	200	500	∞
37	2,61	2,53	2,41	2,33	2,25	2,16	2,10	2,03	1,98	1,92	1,88	1,85
38	2,59	2,51	2,40	2,32	2,23	2,14	2,09	2,01	1,97	1,90	1,86	1,84
39	2,58	2,50	2,38	2,30	2,22	2,13	2,07	1,99	1,95	1,89	1,85	1,82
40	2,56	2,48	2,37	2,29	2,20	2,11	2,06	1,98	1,94	1,87	1,83	1,80
41	2,55	2,47	2,36	2,28	2,19	2,10	2,04	1,97	1,92	1,86	1,82	1,79
42	2,54	2,46	2,34	2,26	2,18	2,09	2,03	1,95	1,91	1,85	1,80	1,78
43	2,53	2,45	2,33	2,25	2,17	2,08	2,02	1,94	1,90	1,83	1,79	1,76
44	2,52	2,44	2,32	2,24	2,15	2,07	2,01	1,93	1,89	1,82	1,78	1,75
45	2,51	2,43	2,31	2,23	2,14	2,05	2,00	1,92	1,88	1,81	1,77	1,74
46	2,50	2,42	2,30	2,22	2,13	2,04	1,99	1,91	1,86	1,80	1,76	1,73
47	2,49	2,41	2,29	2,21	2,12	2,03	1,98	1,90	1,85	1,79	1,74	1,71
48	2,48	2,40	2,28	2,20	2,12	2,02	1,97	1,89	1,84	1,78	1,73	1,70
49	2,47	2,39	2,27	2,19	2,11	2,02	1,96	1,88	1,83	1,77	1,72	1,69
50	2,46	2,38	2,27	2,18	2,10	2,01	1,95	1,87	1,82	1,76	1,71	1,68
55	2,42	2,34	2,23	2,15	2,06	1,97	1,91	1,83	1,78	1,71	1,67	1,64
60	2,39	2,31	2,20	2,12	2,03	1,94	1,88	1,79	1,75	1,68	1,63	1,60
65	2,37	2,29	2,17	2,09	2,00	1,91	1,85	1,77	1,72	1,65	1,60	1,57
70	2,35	2,27	2,15	2,07	1,98	1,89	1,83	1,74	1,70	1,62	1,57	1,54
75	2,33	2,25	2,13	2,05	1,96	1,87	1,81	1,72	1,67	1,60	1,55	1,52
80	2,31	2,23	2,12	2,03	1,94	1,85	1,79	1,70	1,65	1,58	1,53	1,49
85	2,30	2,22	2,10	2,02	1,93	1,83	1,77	1,69	1,64	1,56	1,51	1,47
90	2,29	2,21	2,09	2,00	1,92	1,82	1,76	1,67	1,62	1,55	1,49	1,46
95	2,28	2,20	2,08	1,99	1,90	1,81	1,75	1,66	1,61	1,53	1,48	1,44
100	2,27	2,19	2,07	1,98	1,89	1,80	1,74	1,65	1,60	1,52	1,47	1,43
125	2,23	2,15	2,03	1,94	1,85	1,76	1,69	1,60	1,55	1,47	1,41	1,37
150	2,20	2,12	2,00	1,92	1,83	1,73	1,66	1,57	1,52	1,43	1,38	1,33
175	2,19	2,10	1,98	1,90	1,81	1,71	1,64	1,55	1,50	1,41	1,35	1,30
200	2,17	2,09	1,97	1,89	1,79	1,69	1,63	1,53	1,48	1,39	1,33	1,28
300	2,14	2,06	1,94	1,85	1,76	1,66	1,59	1,50	1,44	1,35	1,28	1,22
400	2,13	2,05	1,92	1,84	1,75	1,64	1,58	1,48	1,42	1,32	1,25	1,19
500	2,12	2,04	1,92	1,83	1,74	1,63	1,57	1,47	1,41	1,31	1,23	1,16
600	2,11	2,03	1,91	1,82	1,73	1,63	1,56	1,46	1,40	1,30	1,22	1,15
700	2,11	2,03	1,90	1,82	1,72	1,62	1,55	1,45	1,39	1,29	1,21	1,14
800	2,10	2,02	1,90	1,81	1,72	1,62	1,55	1,45	1,39	1,29	1,20	1,13
1000	2,10	2,02	1,90	1,81	1,72	1,61	1,54	1,44	1,38	1,28	1,19	1,11
∞	2,08	2,00	1,88	1,79	1,70	1,59	1,52	1,42	1,36	1,25	1,15	1,00

Продолжение прил. 8

Уровень значимости $\alpha = 0{,}05$												
$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	161	199	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09
31	4,16	3,30	2,91	2,68	2,52	2,41	2,32	2,25	2,20	2,15	2,11	2,08
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07
33	4,14	3,28	2,89	2,66	2,50	2,39	2,30	2,23	2,18	2,13	2,09	2,06
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,07	2,04
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03

Продолжение прил. 8

Уровень значимости $\alpha = 0{,}05$												
$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
37	4,11	3,25	2,86	2,63	2,47	2,36	2,27	2,20	2,14	2,10	2,06	2,02
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02
39	4,09	3,24	2,85	2,61	2,46	2,34	2,26	2,19	2,13	2,08	2,04	2,01
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00
41	4,08	3,23	2,83	2,60	2,44	2,33	2,24	2,17	2,12	2,07	2,03	2,00
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,03	1,99
43	4,07	3,21	2,82	2,59	2,43	2,32	2,23	2,16	2,11	2,06	2,02	1,99
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	2,01	1,97
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	2,00	1,97
47	4,05	3,20	2,80	2,57	2,41	2,30	2,21	2,14	2,09	2,04	2,00	1,96
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96
49	4,04	3,19	2,79	2,56	2,40	2,29	2,20	2,13	2,08	2,03	1,99	1,96
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95
55	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	2,01	1,97	1,93
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,03	1,98	1,94	1,90
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89
75	3,97	3,12	2,73	2,49	2,34	2,22	2,13	2,06	2,01	1,96	1,92	1,88
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88
85	3,95	3,10	2,71	2,48	2,32	2,21	2,12	2,05	1,99	1,94	1,90	1,87
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86
95	3,94	3,09	2,70	2,47	2,31	2,20	2,11	2,04	1,98	1,93	1,89	1,86
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,87	1,83
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82
175	3,90	3,05	2,66	2,42	2,27	2,15	2,06	1,99	1,93	1,89	1,84	1,81
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,82	1,78
400	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90	1,85	1,81	1,78
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77
600	3,86	3,01	2,62	2,39	2,23	2,11	2,02	1,95	1,90	1,85	1,80	1,77
700	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,80	1,77
800	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,80	1,76
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75

Продолжение прил. 8

$k_2 \backslash k_1$	14	16	20	24	30	40	50	75	100	200	500	∞
	Уровень значимости $\alpha = 0{,}05$											
1	245	246	248	249	250	251	252	253	253	254	254	254
2	19,42	19,43	19,45	19,45	19,46	19,47	19,48	19,48	19,49	19,49	19,49	19,50
3	8,71	8,69	8,66	8,64	8,62	8,59	8,58	8,56	8,55	8,54	8,53	8,53
4	5,87	5,84	5,80	5,77	5,75	5,72	5,70	5,68	5,66	5,65	5,64	5,63
5	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,41	4,39	4,37	4,37
6	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,73	3,71	3,69	3,68	3,67
7	3,53	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,27	3,25	3,24	3,23
8	3,24	3,20	3,15	3,12	3,08	3,04	3,02	2,99	2,97	2,95	2,94	2,93
9	3,03	2,99	2,94	2,90	2,86	2,83	2,80	2,77	2,76	2,73	2,72	2,71
10	2,86	2,83	2,77	2,74	2,70	2,66	2,64	2,60	2,59	2,56	2,55	2,54
11	2,74	2,70	2,65	2,61	2,57	2,53	2,51	2,47	2,46	2,43	2,42	2,40
12	2,64	2,60	2,54	2,51	2,47	2,43	2,40	2,37	2,35	2,32	2,31	2,30
13	2,55	2,51	2,46	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,22	2,21
14	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
15	2,42	2,38	2,33	2,29	2,25	2,20	2,18	2,14	2,12	2,10	2,08	2,07
16	2,37	2,33	2,28	2,24	2,19	2,15	2,12	2,09	2,07	2,04	2,02	2,01
17	2,33	2,29	2,23	2,19	2,15	2,10	2,08	2,04	2,02	1,99	1,97	1,96
18	2,29	2,25	2,19	2,15	2,11	2,06	2,04	2,00	1,98	1,95	1,93	1,92
19	2,26	2,21	2,16	2,11	2,07	2,03	2,00	1,96	1,94	1,91	1,89	1,88
20	2,22	2,18	2,12	2,08	2,04	1,99	1,97	1,93	1,91	1,88	1,86	1,84
21	2,20	2,16	2,10	2,05	2,01	1,96	1,94	1,90	1,88	1,84	1,83	1,81
22	2,17	2,13	2,07	2,03	1,98	1,94	1,91	1,87	1,85	1,82	1,80	1,78
23	2,15	2,11	2,05	2,01	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76
24	2,13	2,09	2,03	1,98	1,94	1,89	1,86	1,82	1,80	1,77	1,75	1,73
25	2,11	2,07	2,01	1,96	1,92	1,87	1,84	1,80	1,78	1,75	1,73	1,71
26	2,09	2,05	1,99	1,95	1,90	1,85	1,82	1,78	1,76	1,73	1,71	1,69
27	2,08	2,04	1,97	1,93	1,88	1,84	1,81	1,76	1,74	1,71	1,69	1,67
28	2,06	2,02	1,96	1,91	1,87	1,82	1,79	1,75	1,73	1,69	1,67	1,65
29	2,05	2,01	1,94	1,90	1,85	1,81	1,77	1,73	1,71	1,67	1,65	1,64
30	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,70	1,66	1,64	1,62
31	2,03	1,98	1,92	1,88	1,83	1,78	1,75	1,70	1,68	1,65	1,62	1,61
32	2,01	1,97	1,91	1,86	1,82	1,77	1,74	1,69	1,67	1,63	1,61	1,59
33	2,00	1,96	1,90	1,85	1,81	1,76	1,72	1,68	1,66	1,62	1,60	1,58
34	1,99	1,95	1,89	1,84	1,80	1,75	1,71	1,67	1,65	1,61	1,59	1,57
35	1,99	1,94	1,88	1,83	1,79	1,74	1,70	1,66	1,63	1,60	1,57	1,56
36	1,98	1,93	1,87	1,82	1,78	1,73	1,69	1,65	1,62	1,59	1,56	1,55

Окончание прил. 8

Уровень значимости $\alpha = 0{,}05$												
$k_2 \backslash k_1$	14	16	20	24	30	40	50	75	100	200	500	∞
37	1,97	1,93	1,86	1,82	1,77	1,72	1,68	1,64	1,62	1,58	1,55	1,54
38	1,96	1,92	1,85	1,81	1,76	1,71	1,68	1,63	1,61	1,57	1,54	1,53
39	1,95	1,91	1,85	1,80	1,75	1,70	1,67	1,62	1,60	1,56	1,53	1,52
40	1,95	1,90	1,84	1,79	1,74	1,69	1,66	1,61	1,59	1,55	1,53	1,51
41	1,94	1,90	1,83	1,79	1,74	1,69	1,65	1,61	1,58	1,54	1,52	1,50
42	1,94	1,89	1,83	1,78	1,73	1,68	1,65	1,60	1,57	1,53	1,51	1,49
43	1,93	1,89	1,82	1,77	1,72	1,67	1,64	1,59	1,57	1,53	1,50	1,48
44	1,92	1,88	1,81	1,77	1,72	1,67	1,63	1,59	1,56	1,52	1,49	1,48
45	1,92	1,87	1,81	1,76	1,71	1,66	1,63	1,58	1,55	1,51	1,49	1,47
46	1,91	1,87	1,80	1,76	1,71	1,65	1,62	1,57	1,55	1,51	1,48	1,46
47	1,91	1,86	1,80	1,75	1,70	1,65	1,61	1,57	1,54	1,50	1,47	1,46
48	1,90	1,86	1,79	1,75	1,70	1,64	1,61	1,56	1,54	1,49	1,47	1,45
49	1,90	1,85	1,79	1,74	1,69	1,64	1,60	1,56	1,53	1,49	1,46	1,44
50	1,89	1,85	1,78	1,74	1,69	1,63	1,60	1,55	1,52	1,48	1,46	1,44
55	1,88	1,83	1,76	1,72	1,67	1,61	1,58	1,53	1,50	1,46	1,43	1,41
60	1,86	1,82	1,75	1,70	1,65	1,59	1,56	1,51	1,48	1,44	1,41	1,39
65	1,85	1,80	1,73	1,69	1,63	1,58	1,54	1,49	1,46	1,42	1,39	1,37
70	1,84	1,79	1,72	1,67	1,62	1,57	1,53	1,48	1,45	1,40	1,37	1,35
75	1,83	1,78	1,71	1,66	1,61	1,55	1,52	1,47	1,44	1,39	1,36	1,34
80	1,82	1,77	1,70	1,65	1,60	1,54	1,51	1,45	1,43	1,38	1,35	1,32
85	1,81	1,76	1,70	1,65	1,59	1,54	1,50	1,45	1,42	1,37	1,34	1,31
90	1,80	1,76	1,69	1,64	1,59	1,53	1,49	1,44	1,41	1,36	1,33	1,30
95	1,80	1,75	1,68	1,63	1,58	1,52	1,48	1,43	1,40	1,35	1,32	1,29
100	1,79	1,75	1,68	1,63	1,57	1,52	1,48	1,42	1,39	1,34	1,31	1,28
125	1,77	1,73	1,66	1,60	1,55	1,49	1,45	1,40	1,36	1,31	1,27	1,25
150	1,76	1,71	1,64	1,59	1,54	1,48	1,44	1,38	1,34	1,29	1,25	1,22
175	1,75	1,70	1,63	1,58	1,52	1,46	1,42	1,36	1,33	1,27	1,23	1,20
200	1,74	1,69	1,62	1,57	1,52	1,46	1,41	1,35	1,32	1,26	1,22	1,19
300	1,72	1,68	1,61	1,55	1,50	1,43	1,39	1,33	1,30	1,23	1,19	1,15
400	1,72	1,67	1,60	1,54	1,49	1,42	1,38	1,32	1,28	1,22	1,17	1,13
500	1,71	1,66	1,59	1,54	1,48	1,42	1,38	1,31	1,28	1,21	1,16	1,11
600	1,71	1,66	1,59	1,54	1,48	1,41	1,37	1,31	1,27	1,20	1,15	1,10
700	1,71	1,66	1,59	1,53	1,48	1,41	1,37	1,30	1,27	1,20	1,15	1,09
800	1,70	1,66	1,58	1,53	1,47	1,41	1,37	1,30	1,26	1,20	1,14	1,09
1000	1,70	1,65	1,58	1,53	1,47	1,41	1,36	1,30	1,26	1,19	1,13	1,08
∞	1,69	1,64	1,57	1,52	1,46	1,39	1,35	1,28	1,24	1,17	1,11	1,00

Приложение 9

Латинский алфавит

A	a	а	N	n	эн
B	b	бэ	O	o	о
C	c	це	P	p	пэ
D	d	дэ	Q	q	ку
E	e	е	R	r	эр
F	f	еф	S	s	эс
G	g	гэ (жэ)	T	t	тэ
H	h	ха (аш)	U	u	у
I	i	и	V	v	вэ
J	j	йот (жи)	W	w	дубль-вэ
K	k	ка	X	x	икс
L	l	эль	Y	y	игрек
M	m	эм	Z	z	зэт

Приложение 10

Греческий алфавит

A	α	альфа	N	ν	ню (ни)
B	β	бета	Ξ	ξ	кси
Γ	γ	гамма	O	o	омикрон
Δ	δ	дельта	Π	π	пи
E	ε	эпсилон	P	ρ	ро
Z	ζ	дзета	Σ	σ	сигма
H	η	эта	T	τ	тау
Θ	θ, ϑ	тета	Υ	υ	юпсилон (ипсилон)
I	ι	йота	Φ	φ	фи
K	$\varkappa$	каппа	X	χ	хи
Λ	λ	ламбда	Ψ	ψ	пси
M	μ	мю (ми)	Ω	ω	омега

Список использованной и рекомендуемой литературы

Основная

1. *Вентцель Е. С.* Теория вероятностей. — М.: Физматгиз, 1963.
2. *Гмурман В. Е.* Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие для студ. вузов. — Изд. 6-е, доп. — М.: Высш. шк., 2002. — 405 с.
3. *Гмурман В. Е.* Теория вероятностей и математическая статистика. — М.: Высш. шк., 1999.
4. *Гнеденко Б. В., Хинчин А. Я.* Элементарное введение в теорию вероятностей. — М.: Наука, 1976. — 168 с.
5. *Горбань С. Ф., Снижко Н. В.* Теория вероятностей и математическая статистика. — К.: МАУП, 1999. — 168 с.
6. *Жлуктенко В. І., Наконечний С. І.* Практикум з курсу «Теорія ймовірностей і математична статистика». — К.: КІНГ, 1991.
7. *Жлуктенко В. І., Наконечний С. І.* Теорія ймовірностей із елементами математичної статистики. — К.: НМК ВО, 1991.
8. *Кремер Н. Ш.* Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2000.

Дополнительная

9. *Бешелев С. Д., Гурвич Ф. Г.* Математико-статистические методы экспертных оценок. — М.: Статистика, 1980. — 263 с.

10. *Вайнберг Дж., Шуменер Дж.* Статистика: Пер. с англ. — М.: Статистика, 1979. — 389 с.
11. *Гнеденко Б. В.* Курс теории вероятностей. — М.: Физматгиз, 1961.
12. *Громыко Г. Л.* Статистика. — М.: Изд-во МГУ, 1981. — С. 3–166.
13. *Гурский Е. М.* Теория вероятностей с элементами математической статистики. — М.: Высш. шк., 1971.
14. *Карасев А. И.* Теория вероятности и математическая статистика. — М.: Статистика, 1977.
15. *Кимбл Г.* Как правильно пользоваться статистикой: Пер. с англ. — М.: Финансы и статистика, 1982. — 294 с.
16. *Лбов Г. С.* Методы обработки разнотипных экспериментальных данных. — Новосибирск: Наука, 1981.
17. *Математическая* статистика: Учебник. — М.: Высш. шк., 1981. — 371 с.
18. *Мюллер П., Нойман, Шторм Р.* Таблицы по математической статистике: Пер. с нем. — М.: Финансы и статистика, 1982. — 178 с.
19. *Окунь Я.* Факторный анализ: Пер. с польск. — М.: Статистика, 1974.
20. *Сидоренко Е. В.* Методы математической обработки в психологии. — СПб.: ООО «Речь», 2001. — 350 с.
21. *Справочник* по теории вероятностей и математической статистике. — К.: Наук. думка, 1978. — 582 с.
22. *Турчин В. М.* Математична статистика: Навч. посіб. — К.: Вид. центр «Академія», 1999. — 240 с.
23. *Хастинг Н., Пикон Дж.* Справочник по статистическим распределениям: Пер. с англ. — М.: Статистика, 1980. — 95 с.
24. *Чистяков В. П.* Курс теории вероятностей. — М., 1978.
25. *Шефтель З. Г.* Теорія ймовірностей. — К., 1994.

Предметный указатель

Оглавление

У пропонованому практикумі подано необхідні теоретичні дані і формули, розв'язки типових задач, задачі для самостійного розв'язування із вказівками та відповідями до них. Значну увагу приділено методам статистичної обробки експериментальних даних.

Для студентів-іноземців нематематичних спеціальностей вищих навчальних закладів, а також для всіх, хто застосовує теорію ймовірностей і статистичні методи при розв'язуванні практичних задач.

Навчальне видання

Чорней Руслан Костянтинович

ПРАКТИКУМ
з теорії ймовірностей
і математичної статистики

Навчальний посібник
для студентів-іноземців
вузівської підготовки
(Рос. мовою)

Відповідальний редактор *С. Г. Рогузько*
Редактор *Л. В. Логвиненко*
Коректор *Т. К. Валицька*
Комп'ютерне верстання *Р. К. Чорней*
Оформлення обкладинки *С. В. Фадєєв*

Підп. до друку 15.05.08. Формат $60\times84/_{16}$. Папір офсетний. Друк офсетний.
Ум. друк. арк. 19,53. Обл.-вид. арк. 17,2. Наклад 700 пр.

Міжрегіональна Академія управління персоналом (МАУП)
03039 Київ-39, вул. Фрометівська, 2, МАУП

ДП «Видавничий дім «Персонал»
03039 Київ-39, просп. Червонозоряний, 119, літ. ХХ

Свідоцтво про внесення до Державного реєстру
суб'єктів видавничої справи ДК № 3262 від 26.08.2008