Ярославский государственный университет им. П.Г. Демидова

**С. Б. Московский**

**А. Н. Сергеев**

# ОСНОВЫ СТАСТИЧЕСКОЙ ОБРАБОТКИ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

Министерство образования и науки Российской Федерации
Ярославский государственный университет им. П.Г.Демидова
Кафедра микроэлектроники и общей физики

**С. Б. Московский**
**А. Н. Сергеев**

# ОСНОВЫ СТАСТИЧЕСКОЙ ОБРАБОТКИ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

*Учебно-методическое пособие*

*Рекомендовано*
*Научно-методическим советом для университета для студентов, обучающихся по направлениям подготовки Электроника и наноэлектроника*

Ярославль
ЯрГУ
2018

УДК
ББК

*Рекомендовано*

*Редакционно-издательским советом университета*
*в качестве учебного издания. План 2018 года.*

*Рецензент*

*Кафедра микроэлектроники и общей физики*

*ЯрГУ им. П. Г. Демидова*

**Московский, Сергей Борисович.**

Основы статистической обработки результатов измерений: учебно-методическое пособие / С. Б. Московский, А. Н. Сергеев; Яросл. гос. ун-т им. П.Г. Демидова.- Ярославль: ЯрГУ, 2018. – с.

В данном учебно-методическом пособии рассматриваются числовые характеристики дискретных случайных величин (математическое ожидание, дисперсия); функции распределения и их свойства; оценка погрешности прямых и косвенных измерений; парная регрессия и корреляция. Приведен список заданий для самостоятельного решения.

Предназначено для студентов, обучающихся по направлению подготовки 11.03.04 Электроника и наноэлектроника (дисциплина в рупе "Статистическая обработка экспериментальных данных") очной формы обучения. Результаты освоения дисциплины Статистическая обработка экспериментальных данных в свою очередь используются при последующем изучении общефизических и профессиональных дисциплин, а также в научно-исследовательской работе студента на уровне бакалавриата и магистратуры.

УДК
ББК

Содержание.

## Введение.

Характерным для современного этапа развития естественных и технических наук является весьма широкое и плодотворное применение статистических методов во всех областях знания. Задача любой науки состоит в выявлении и исследовании закономерностей, которым подчиняются реальные процессы. Найденные закономерности имеют не только теоретическую ценность, они широко применяются на практике – в планировании, управлении и прогнозировании.

Математическая статистика– раздел математики, изучающий математические методы сбора, систематизации, обработки и интерпретации результатов наблюдений с целью выявления статистических закономерностей. Математическая статистика по наблюденным значениям (выборке) оценивает вероятности событий либо осуществляет проверку предположений (гипотез) относительно этих вероятностей.

Изучение вероятностных моделей дает возможность понять различные свойства случайных явлений на абстрактном и обобщенном уровне, не прибегая к эксперименту. В математической статистике, наоборот, исследование связано с конкретными данными и идет от практики (наблюдения) к гипотезе и ее проверке.

При большом числе наблюдений случайные воздействия в значительной мере погашаются (нейтрализуются) и получаемый результат оказывается практически неслучайным, предсказуемым. Это утверждение (принцип) и является базой для практического использования вероятностных и математико-статистических методов исследования. Цель указанных методов состоит в том, чтобы, минуя сложное (а зачастую и невозможное) исследование отдельного случайного явления, изучить закономерности массовых случайных явлений, прогнозировать их характеристики, влиять на ход этих явлений, контролировать их, ограничивать область действия случайности.

Результаты эксперимента для инженера-исследователя были и остаются главным критерием при решении практических задач и при проверкетеоретических гипотез. Однако при этом важно не только умело спланировать и поставить эксперимент, но и грамотно обработать его результаты.

Этому вопросу часто не уделяется должного внимания, и нередки случаи, когда результаты дорогостоящих экспериментов не подвергают даже простейшей обработке; при этом, как следствие, теряется огромное количество полезной информации.

Следует также подчеркнуть, что обработке экспериментальных данных с целью построения моделей «сложных систем» (эмпирических зависимостей) должна предшествовать предварительная обработка, содержание которой, в основном, состоит в отсеивании грубых погрешностей измерений ив проверке соответствия распределения результатов нормальному закону.

Следует помнить, что только после выполнения предварительной обработкиможно с наибольшей эффективностью, а главное корректно, использоватьболее сложные экспериментально-статистические методы, позволяющие получать математические модели даже таких процессов, строгое детерминированное описание которых вообще отсутствует.

## 1. Характеристики случайной величины

Естественный ответ на этот вопрос — “получение истинного значения физической величины” — вызывает следующий вопрос, а что такое “истинное значение”? Если исключить фундаментальные физические постоянные типа скорости света $c = (2.99792458 \pm 0.00000001)\cdot 10^{10}$ см/с, постоянной Планка $\hbar = (1.054589 \pm 0.000005)\cdot 10^{-27}$эрг · с, заряда электрона $e = (1.602189 \pm 0.000006)\cdot 10^{-19}$К и др., которые, вероятно, имеют точные значения (приведенный разброс связан с ограниченной точностью измерений), все остальные величины, в принципе, могут быть определены только как некоторое математическое ожидание $M(X)$.

Действительно, такая величина, как скорость радиоактивного распада атомов определенного сорта, определяется только как среднее значение вследствие случайности времени распада каждого отдельного атома, приводящей к флуктуациям числа распадов в единицу времени. Проводя измерения числа распадов в течение ограниченного времени измерения, мы можем вычислить только некоторое выборочное среднее, отличающееся от математического ожидания.

Другие измеряемые величины, такие как сопротивление провода, скорость звука в газе, показатель преломления стекла, также не могут иметь "истинного значения" вследствие флуктуаций во времени и пространстве плотности, температуры и других характеристик сред. Более того, даже такой простой вопрос, как измерение длины стержня, приводит к целой последовательности вопросов:

Постоянна ли длина стержня в процессе измерения?

Каково качество обработки торцов стержня, а если торец не прямой, то что считать за конец?

Если с некоторой точностью мы можем считать торец гладким, то параллелен ли ему второй торец? И т.д.

Таким образом, факторы, связанные со статистическими флуктуациями, а также с природным разбросом или распределением измеряемой величины по измеряемому объекту, ведут к разбросу численных значений измеряемой величины, получаемых в разных сериях измерений.

Но даже если максимально зафиксировать параметра среды так, чтобы флуктуации измеряемого параметра стали очень малы, имеются другие источники разброса, связанные с неточностями самих приборов, внешними помехами (вибрация, электрические наводки и т.п.), ошибками экспериментатора. Эти ошибки могут носить как статистический, так и систематический характер. В результате измеряемые в серии опытов значения физической величины лежат в определенном интервале. В этом случае говорят о распределении измеряемой величины. Рассмотрим все эти вопросы подробнее.

В табл.1 приведены результаты измерения роста студентов 1-го курса физического факультета, проведенные в 2017 г. на первой лекции по методам физического эксперимента.

*Таблица 1*

*Результаты измерений роста студентов 1-го курса*

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $x_i$ | 177 | 178 | 171 | 189 | 189 | 189 | 176 | 183 | 176 | 175 | 176 | 179 |
| $i$ | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| $x_i$ | 181 | 171 | 154 | 164 | 164 | 165 | 173 | 173 | 184 | 188 | 188 | 176 |
| $i$ | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
| $x_i$ | 174 | 174 | 176 | 177 | 177 | 178 | 176 | 173 | 182 | 177 | 159 | 178 |
| $i$ | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 |
| $x_i$ | 186 | 173 | 180 | 171 | 166 | 177 | 173 | 177 | 192 | 175 | 181 | 183 |
| $i$ | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
| $x_i$ | 159 | 166 | 181 | 160 | 179 | 175 | 185 | 186 | 170 | 172 | 168 | 182 |
| $i$ | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 |
| $x_i$ | 184 | 166 | 175 | 171 | 179 | 178 | 187 | 176 | 152 | 169 | 180 | 192 |
| $i$ | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 |
| $x_i$ | 180 | 178 | 184 | 187 | 175 | 183 | 179 | 171 | 171 | 177 | 159 | 179 |
| $i$ | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 |
| $x_i$ | 159 | 181 | 185 | 177 | 175 | 178 | 169 | 170 | 175 | 175 | 183 | 173 |
| $i$ | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 |
| $x_i$ | 178 | 183 | 177 | 186 | 190 | 185 | 170 | 178 | 169 | 165 | 182 | 186 |
| $i$ | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 |
| $x_i$ | 184 | 176 | 169 | 183 | 168 | 181 | 179 | 180 | 183 | 174 | 175 | 177 |
| $i$ | 121 | 122 | 123 | 124 | | | | | | | | |
| $x_i$ | 176 | 169 | 176 | 178 | | | | | | | | |

Отклонение измеренного значения x определяемой величины от истинного ее значения *M(X)* (здесь и далее под “истинным значением” понимается математическое ожидание величины) называется “погрешностью” или “ошибкой” (error) *i*-го измерения $\Delta x_i$. Термин “ошибка” в статистике не несет в себе эмоциональной негативной окраски, а имеет смысл разброса значений, присущего данному явлению или измерительному устройству (на это указывает и некоторое смысловое различие между словами error и mistake в английском языке). По-видимому, наиболее адекватным (по смыслу) переводом на русский язык было бы слово “отклонение”. Ошибки измерений возникают вследствие того, что измерительные приборы имеют ограниченную точность, не всегда можно учесть влияние неконтролируемых экспериментальных условий. Иногда сама природа изучаемого явления служит основной причиной разброса измерений. Ошибки измерений принято разделять на два типа — систематические и случайные.

Систематическими называют такие ошибки, которые вызываются факторами, действующими одинаковым образом при многократном повторении одних и тех же измерений. При измерении роста студентов на лекции систематической ошибкой было измерение роста в обуви.

Систематические ошибки, в принципе, могут быть устранены или учтены, хотя нахождение источников этих ошибок в конкретном опыте чрезвычайно сложное дело. Любой эксперимент не может гарантировать защищенность от наличия неучтенной систематической ошибки.

Случайные ошибки (отклонения) всегда присутствуют в эксперименте. При отсутствии систематических ошибок они служат причиной разброса результатов повторных измерений, как между собой, так и относительно истинного значения измеряемой величины. Природа случайных ошибок может быть различной. Например, при измерении длины стрежня, которая точно определена (с учетом сказанного выше), к случайным ошибкам приводит вибрация стержня, флуктуации нулевого положения измерительного прибора и т.п. В опытах по измерению скорости радиоактивного распада ядер сама измеряемая величина определена лишь статистически (как некоторое среднее значение), и флуктуации измеренного числа распадов в равные промежутки времени будут наблюдаться даже при идеально точной аппаратуре. Увеличивая число измерений, и используя формулы теории ошибок, можно получить достаточно точную оценку случайной ошибки.

Случайные ошибки последовательных измерений, как правило, независимы и характеризуются тем или иным законом распределения (см.ниже). Они обладают свойством концентрации, т.е. малые по абсолютной величине случайные ошибки встречаются чаще, чем большие.

Отметим еще раз, что разделение ошибок на случайные и систематические в достаточной мере условно. В зависимости от обстоятельств, один и тот же фактор может приводить либо к случайным, либо к систематическим ошибкам.

Следует особо выделить такой вид ошибки, как грубый просчет — промах. Под промахом понимается ошибка измерения, сделанная вследствие недосмотра экспериментатора, или вызванная неисправностями аппаратуры. Так, например, неправильно записанный отсчет, замыкание электрической цепи и т.п. являются промахами, которых следует по возможности избегать. Как правило, грубые ошибки легко обнаруживаются. Такие измерения следует отбрасывать, хотя при этом желательно определить причину данного промаха.

Систематические ошибки, присущие системе измерений, должны быть обнаружены и ликвидированы (или учтены при обработке данных). Оставшиеся необнаруженными систематические ошибки вносят в результаты измерений неизвестный сдвиг относительно истинного значения. Поскольку практически невозможно выявить все систематические ошибки, даже очень точные измерения разных авторов могут не совпадать. В примере с водяным конденсатором разброс значений постоянной Керра у разных авторов можно связать, в частности, с различием в методах очистки воды и наличием неконтролируемых примесей. Можно без преувеличения сказать, что мастерство экспериментатора состоит в искусстве обнаружения систематических ошибок.

Если вам удалось снизить до достаточно низкого уровня систематические ошибки, то точность измерений определяется теперь случайными погрешностями. Если же точность измерений определяется случайными ошибками, то она может быть подвергнута статистическому анализу. Обсудим далее, как следует оценивать точность значения некоторой средней величины $X$, полученной экспериментально.

Генеральной совокупностью называют полный набор всех возможных значений, которые может принимать случайная величина при бесконечном числе испытаний. Генеральную совокупность можно описать с помощью распределения вероятности появления в отдельном испытании некоторой конкретной величины $x_i$. Набор n значений величин $x_i$, полученный из генеральной совокупности в результате конечного числа испытаний $N$, называют выборкой объема $N$. Цель статистической обработки набора величин $x_i$ заключается в попытке описать характеристики генеральной совокупности как можно точнее по отдельной выборке.

Проведем серию из $N$ измерений и получим некоторый набор (см.табл. 1) значений $x_1,..., x_i,..., x_N$. Это — так называемая выборка. По оси абсцисс (рис.1) будем откладывать полученные в отдельных измерениях значения $x_i$ величины $x$. Разобьем ось $x$ на равные интервалы $\Delta$x и подсчитаем число измерений $\Delta n_k$, в результате которых получены значения x, лежащие в интервале $x_k \pm 1/2\Delta x$ (здесь $x_k$ — координата центра интервала на оси $x$). На каждом интервале построим прямоугольник высотой $\Delta n_k$ и шириной $\Delta x$ (точки, лежащие точно на границе интервала, будем всегда относить, например, в левый столбик). Диаграмму, полученную таким образом, называют гистограммой. Гистограммы могут быть построены как для непрерывных величин (например, скорость частиц), так и для дискретных (например, число радиоактивных распадов в секунду). Сумма высот прямоугольников гистограммы равна полному числу экспериментов в данной серии.

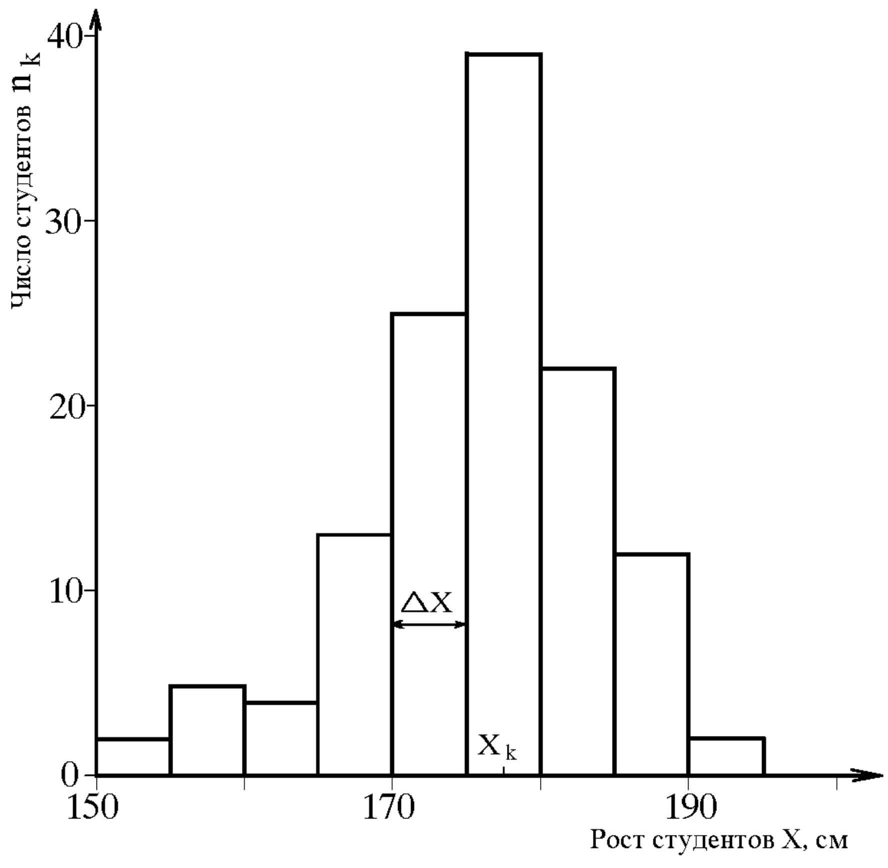


*Рис. 1. Гистограмма распределения роста студентов 1-го курса физического факультета*

Если мы имеем две выборки с разным полным числом событий, то амплитудные значения гистограмм, естественно, будут различны, и их трудно будет сравнивать. Поэтому более удобно представлять гистограмму в несколько ином виде. Если откладывать по оси ординат не $\Delta n_k$, а величину

$$\frac{1}{N}\frac{\Delta n_k}{\Delta x_k},$$

где $N = \sum n_k$. В этом случае произведение высоты на $\Delta x$, т.е. площадь каждого столбика, имеет размерность вероятности попадания результата отдельного измерения в данный интервал $\Delta x$, а суммарная площадь под всей гистограммой, как и должно, быть в этом случае, равна единице:

$$\sum \frac{1}{N}\frac{\Delta n_k}{\Delta x_k}\Delta x_k = 1 \qquad (1.1)$$

Теперь нетрудно сравнивать гистограммы с сильно отличающимся числом измерений.

Если число измерений $N$ достаточно велико, то ширину интервала можно сделать очень малой (при этом в каждом интервале будет еще достаточно много отсчетов). Тогда в пределе вместо гистограммы мы получим график типа показанного на рис. 2, на котором по оси ординат отложена величина $f(x)$, имеющая размерность $[x^{-1}]$ и пропорциональная доле числа отсчетов $n_k/n$, попадающей в каждый интервал. Такой график называют кривой распределения, а функцию $f(x)$ называют плотностью вероятности.

Смысл плотности вероятности заключается в том, что произведение $f(x)dx$ (см. рис. 2) дает долю полного числа отсчетов $n$, приходящуюся на интервал от x до $x+dx$ или, иными словами, вероятность того, что результат любого очередного измерения $x_i$ будет иметь значение, лежащее в указанном интервале. Очевидно, что вероятность получить при измерении хоть какое-нибудь значение равна единице.

Тогда функция $f(x)$ должна удовлетворять соотношению

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \qquad (1.2)$$

Отсюда видно, что размерность функции плотности вероятности $f(x)$ есть $[x]^{-1}$. Понятно, что бесконечные пределы интегрирования здесь имеют формальный смысл, а реально измеряемые величины должны быть ограничены некоторым разумным диапазоном.

В случае дискретной величины вместо $f(x)dx$ используют вероятность $p_i$, полученную в результате измерения значение $x_i$. Естественно, что

$$\sum_{i=1}^{N} p_i = 1 \qquad (1.3)$$

где $N$ — число интервалов дискретного распределения. Ниже в этом разделе мы будем говорить о непрерывных распределениях.

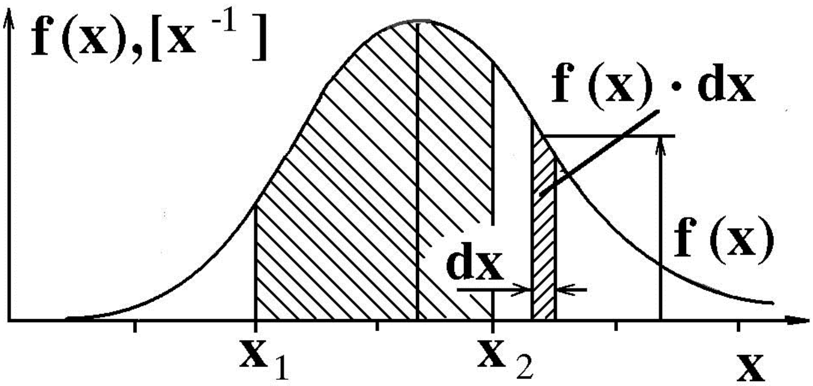

*Рис. 2.Пример кривой распределения*

Вероятность попадания измеряемой величины (в данном измерении) в интервал от $-\infty$ до $x$ называют функцией распределения или интегральной функцией распределения:

$$F(x) = \int_{-\infty}^{x} f(z)dz \qquad (1.4)$$

Понятно, что плотность вероятности $f(x)$ равна просто производной функции распределения $F(x)$ в точке$x$.

Если проинтегрировать плотность вероятности в пределах от $x_1$ до $x_2$(рис. 2), то полученная величина будет представлять вероятность Pтого, что результат отдельного измерения будет лежать в интервале $x_1, x_2$:

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1) \qquad (1.5)$$

Эта вероятность $P$, равная разности интегральных функций распределения в соответствующих точках, равна заштрихованной площади под кривой. При расширении пределов она стремится к единице. Смысл введения такой величины заключается в том, что при заданной функции распределения результат любого отдельного измерения с достоверностью $P$ даст величину, лежащую внутри указанного интервала. Это определение будет нам полезно при определении понятия доверительного интервала и доверительной вероятности.

Если функция распределения известна, то математическое ожидание измеряемой величины, которое было бы получено при бесконечном числе измерений, вычисляется с помощью выражения

$$M(x) = \int_{-\infty}^{+\infty} xf(x)dx \qquad (1.6)$$

а математическое ожидание функции $y(x)$ от непрерывной случайной величины $x$ есть

$$M(y(x)) = \overline{y(x)} = \int_{-\infty}^{+\infty} y(x)f(x)dx \qquad (1.7)$$

Здесь следует отметить, что в физической литературе очень часто функцией распределения называют плотность вероятности $f(x)$, подразумевая, что читатель понимает о чем идет речь.

Если в выражении (1.6) под интегралом стоит n-я степень x, то такая величина называется $n$-м моментом функции распределения. Функция распределения однозначно определяется всей совокупностью (вообще говоря, бесконечной) своих моментов. Наиболее часто используемые в теории ошибок и в физике распределения, как правило, определяются небольшим числом своих моментов.

Помимо математического ожидания, важнейшим из таких моментов является второй момент или дисперсия, которая определяется выражением

$$\sigma^2 = D(x) = M((x-\overline{x})^2) = \int_{-\infty}^{+\infty} (x-\overline{x})^2 f(x)dx \qquad (1.8)$$

Поскольку дисперсия имеет размерность квадрата случайной величины, а это не всегда удобно, то вводится среднее квадратическое отклонение $\sigma$, которое представляет собой положительный квадратный корень из дисперсии

$$\sigma = \sqrt{\sigma^2} \qquad (1.9)$$

## 2. Числовые характеристики дискретных случайных величин

Мы уже знаем, что закон распределения полностью характеризует случайную величину. Однако часто закон распределения неизвестен и приходится ограничиваться меньшими сведениями. Иногда даже выгоднее пользоваться числами, которые он описывают случайную величину суммарно; такие числа называют числовыми характеристиками случайной величины. К числу важных числовых характеристик относится математическое ожидание.

Математическое ожидание, как будет показано далее, приближенно равно среднему значению случайной величины. Для решения многих задач достаточно знать математическое ожидание. Например, если известно, что математическое ожидание числа выбиваемых очков у первого стрелка больше, чем у второго, то первый стрелок в среднем выбивает больше очков, чем второй и, следовательно, стреляет лучше второго. Хотя математическое ожидание дает о случайной величине значительно меньше сведений, чем закон ее распределения,— для решения задач, подобных приведенной и многих других, знание математического ожидания оказывается достаточным.

## 2.1. Математическое ожидание дискретной случайной величины

Математическим ожиданием дискретной случайной величины называют сумму произведении всех ее возможных значений на их вероятности.

Пусть случайная величина *X* может принимать только значения $x_1, x_2, ..., x_n$, вероятности которых соответственно равны $p_1, p_2, ..., p_n$. Тогда математическое ожидание *M(X)* случайной величины *X* определяется равенством

$$M(x) = x_1 p_1 + x_2 p_2 + ... + x_n p_n \tag{2.1}$$

Если дискретная случайная величина *X* принимает счетное множество возможных значений, то

$$M(x) = \sum_{i=1}^{\infty} x_i p_i \tag{2.2}$$

Причем математическое ожидание существует, если ряд в правой части равенства сходится абсолютно.

Математическое ожидание числа появлений события в одном испытании равно вероятности этого события. При большом количестве испытаний математическое ожидание приближенно равно среднему арифметическому наблюдаемых значений случайной величины.

Рассмотрим свойства математического ожидания:

*Свойство 1.* Математическое ожидание постоянной величины равно самой постоянной:

$$M(C) = C \tag{2.3}$$

*Доказательство.* Будем рассматривать постоянную С как дискретную случайную величину, которая имеет одно возможное значение С и принимает его с вероятностью p=1. Следовательно,

$$M(C) = C \cdot 1 = C.$$

*Свойство 2.* Постоянный множитель можно выносить за знак математического ожидания:

$$M(CX) = C \cdot M(X) \tag{2.4}$$

*Доказательство.* Пусть случайная величина *X* задана законом распределения вероятностей:

| $X$ | $x_1$ | $x_2$ | … | $x_n$ |
|---|---|---|---|---|
| $p$ | $p_1$ | $p_2$ | … | $p_n$ |

Учитывая, что вероятности возможных значении $CX$ равны вероятностям соответствующих возможных значений $X$ напишем закон распределения случайной величины $CX$:

| $CX$ | $Cx_1$ | $Cx_2$ | … | $Cx_n$ |
|---|---|---|---|---|
| $p$ | $p_1$ | $p_2$ | … | $p_n$ |

Математическое ожидание случайной величины $CX$:

$$M(CX) = Cx_1p_1 + Cx_2p_2 + ... + Cx_np_n =$$
$$= C(x_1p_1 + x_2p_2 + ... + x_np_n) = C \cdot M(X)$$

*Свойство 3.* Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий:

$$M(XY) = M(X)M(Y) \tag{2.5}$$

*Доказательство.* Пусть независимые случайные величины $X$ и $Y$ заданы своими законами распределения вероятностей (Мы ограничились малым числом возможных значений, чтобы упростить выкладки. В общем случае доказательство аналогичное):

| $X$ | $x_1$ | $x_2$ |
|---|---|---|
| $p$ | $p_1$ | $p_2$ |

| $Y$ | $y_1$ | $y_2$ |
|---|---|---|
| $g$ | $g_1$ | $g_2$ |

Составим все значения, которые может принимать случайная величина $XY$и напишем ее «закон распределения», для чего перемножим все возможные значения $X$ на каждое возможное значение $Y$; в итоге получим: $x_1y_1$, $x_2y_1$, $x_1y_2$, $x_2y_2$:

| $XY$ | $x_1y_1$ | $x_2y_1$ | $x_1y_2$ | $x_2y_2$ |
|---|---|---|---|---|
| $p$ | $p_1g_1$ | $p_2g_1$ | $p_1g_2$ | $p_2g_2$ |

Математическое ожидание равно сумме произведений всех возможных значении на их вероятности:

$$M(XY) = x_1y_1p_1g_1 + x_1y_2p_1g_2 + x_2y_1p_2g_1 + x_2y_2p_2g_2,$$

или

$$M(XY) = y_1g_1(x_1p_1 + x_2p_2) + y_2g_2(x_1p_1 + x_2p_2) =$$
$$= (x_1p_1 + x_2p_2)(y_1g_1 + y_2g_2) = M(X)M(Y)$$

*Следствие.* Математическое ожидание произведения нескольких взаимно независимых случайных величин равно произведению их математических ожиданий.

Например, для трех случайных величин имеем:

$$M(XYZ) = M(XY \cdot Z)M(XY)M(Z) = M(X)M(Y)M(Z) \quad (2.6)$$

*Свойство 4.* Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий слагаемых:

$$M(X+Y) = M(X) + M(Y) \quad (2.7)$$

*Доказательство*. Пусть случайные величины $X$ и $Y$ заданы следующими законами распределения:

| $X$ | $x_1$ | $x_2$ |
|---|---|---|
| $p$ | $p_1$ | $p_2$ |

| $Y$ | $y_1$ | $y_2$ |
|---|---|---|
| $g$ | $g_1$ | $g_2$ |

Составим все возможные значения величины $X+Y$, для чего к каждому возможному значению $X$ прибавим каждое возможное значение $Y$; получим $x_1+y_1$, $x_2+y_1$, $x_1+y_2$, $x_2+y_2$. Обозначим вероятности этих значений соответственно через $p_{11}$, $p_{21}$, $p_{12}$, $p_{22}$.

Математическое ожидание величины $X+Y$ равно сумме произведений возможных значений на их вероятности:

$$M(X+Y) = (x_1 + y_1)p_{11} + (x_2 + y_1)p_{21} + (x_1 + y_2)p_{12} +$$
$$+(x_2 + y_2)p_{22},$$

или

$$M(X+Y)=x_1(p_{11}+p_{12})+x_2(p_{21}+p_{22})+y_1(p_{11}+p_{21})+ \\ +y_2(p_{12}+p_{22}) \tag{2.8}$$

Докажем, что $p_{11}+p_{12}=p_1$. Событие, состоящее в том что $X$ примет значение $x_1$ (вероятность этого события равна $p_1$), влечет за собой событие, которое состоит в том, что $X+Y$ примет значение $x_1+y_1$ или $x_1+y_2$ (вероятность этого события но теореме сложения равна $p_{11}+p_{12}$) и обратно. Аналогично доказываются равенства

$$p_{21}+p_{22}=p_2,\ p_{11}+p_{21}=g_1 \text{ и } p_{12}+p_{22}=g_2.$$

Подставляя правые части этих равенств в соотношение (2.8), получим:

$$M(X+Y)=(x_1p_1+x_2p_2)+(y_1g_1+y_2g_2)=M(X)+M(Y).$$

*Следствие*. Математическое ожидание суммы нескольких случайных величин равно сумме математических ожиданий слагаемых.

Например, для трех слагаемых величин имеем:

$$M(X+Y+Z)=M[(X+Y)+Z]= \\ =M(X+Y)+M(Z)=M(X)+M(Y)+M(Z) \tag{2.9}$$

*Свойство 5*. Математическое ожидание М($X$) числа появлений события $A$ в $n$ независимых испытаниях равно произведению числа испытаний на вероятность появления события в каждом испытании:

$$M(X)=np \tag{2.10}$$

*Доказательство*. Будем рассматривать в качестве случайной величины $X$ — число наступлений события $A$ в $n$ независимых испытаниях. Очевидно, общее число $X$ появлений события $A$ в этих испытаниях складывается из чисел появлении события в отдельных испытаниях. Поэтому, если $X_1$— число появлении события в первом испытании, $X_2$ — во втором, …, $X_n$ — в $n$-м, то общее число появлений события $X=X_1+X_2+\ldots+X_n$.

По третьему свойству математического ожидания,

$$M(X)=\sum_{i=1}^{N}M(X_i) \qquad (2.11)$$

Каждое из слагаемых правой части равенства есть математическое ожидание числа появлений события в одном испытании: $M(X_1)$ — в первом, $M(X_2)$ — во втором и т. д. Так как математическое ожидание числа появлений события в одном испытании равно вероятности события, то $M(X_1)=M(X_2)=M(X_n)=p$. Подставляя в правую часть равенства (2.11) вместо каждого слагаемого p, получим

$$M(X)=np$$

*Свойство 6.* Математическое ожидание отклонения равно нулю:

$$M[X-M(X)]=0 \qquad (2.12)$$

*Доказательство*. Пользуясь свойствами математического ожидания (математическое ожидание разности равно разности математических ожиданий, математическое ожидание постоянной равно самой постоянной) и приняв во внимание, что $M(X)$ есть постоянная величина, имеем:

$$M[X-M(X)]=M(X)-M[M(X)]=M(X)-M(X)=0 \qquad (2.13)$$

## 2.2. Дисперсия дискретной случайной величины

На практике часто требуется оценить рассеяние возможных значений случайной величины вокруг ее среднего значения. Например, в артиллерии важно знать, насколько кучно лягут снаряды вблизи цели, которая должна быть поражена. На первый взгляд может показаться, что для оценки рассеяния проще всего вычислить все возможные значения отклонения случайной величины и затем найти их среднее значение. Однако такой путь ничего не даст, так как среднее значение отклонения, т. е. $M[X-M(X)]$, для любой случайной величины равно пулю. Это объясняется тем, что одни возможные отклонения положительны, а другие отрицательны; в результате их взаимного погашения среднее значение отклонения равно нулю. Эти соображения говорят о целесообразности заменить возможные отклонения их абсолютными значениями или их квадратами. Так и поступают на деле. Правда, в случае, когда возможные отклонения заменяют их абсолютными значениями, приходится оперировать с абсолютными величинами, что приводит иногда к серьезным затруднениям. Поэтому чаще всего идут по другому пути, т. е. вычисляют среднее значение квадрата отклонения, которое и называют дисперсией.

Дисперсией (рассеянием) дискретной случайной величины называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D(X) = M[X - M(X)]^2 \tag{2.14}$$

Пусть случайная величина $X$ задана законом распределения вероятностей:

| $X$ | $x_1$ | $x_2$ | … | $x_n$ |
|---|---|---|---|---|
| $p$ | $p_1$ | $p_2$ | … | $p_n$ |

Тогда квадрат отклонения имеет следующий закон распределения:

| $M[X - M(X)]^2$ | $M[x_1 - M(X)]^2$ | $M[x_2 - M(X)]^2$ | … | $M[x_N - M(X)]^2$ |
|---|---|---|---|---|
| $p$ | $p_1$ | $p_2$ | … | $p_n$ |

По определению дисперсия равна

$$D(X) = M[X - M(X)]^2 = [x_1 - M(X)]^2 p_1 +$$

$$+[x_2 - M(X)]^2 p_2 + \ldots + [x_n - M(X)]^2 p_n$$

Таким образом, для того, чтобы найти дисперсию, достаточно вычислить сумму произведений возможных значений квадрата отклонения на их вероятности.

Для вычисления дисперсии часто бывает удобно пользоваться следующей теоремой.

*Теорема*. Дисперсия равна разности между математическим ожиданием квадрата случайной величины $X$ и квадратом ее математического ожидания:

$$D(X)=M(X^2)-[M(X)]^2 \qquad (2.15)$$

*Доказательство*. Математическое ожидание *M(X)* есть постоянная величина, следовательно, *2M(X)* и $M^2(X)$ есть также постоянные величины. Приняв это во внимание и пользуясь свойствами математического ожидания (постоянный множитель можно вынести за знак математического ожидания, математическое ожидание суммы равно сумме математических ожиданий слагаемых), упростим формулу, выражающую определение дисперсии:

$$D(X)=M[X-M(X)]^2=M[X^2-2X\cdot M(X)+M^2(X)]=$$

$$=M(X^2)-2M(X)M(X)+M^2(X)=M(X^2)-M^2(X)$$

Рассмотрим свойства дисперсии:

*Свойство 1*. Дисперсия постоянной величины *C* равна нулю:

$$D(C)=0 \qquad (2.16)$$

*Доказательство*. По определению дисперсии имеем

$$D(C)=M[C-M(C)]^2.$$

Пользуясь первым свойством математического ожидания (математическое ожидание постоянной равно самой постоянной), получим

$$D(C)=M[C-C]^2=M(0)=0$$

Свойство становится ясным, если учесть, что постоянная величина сохраняет одно и то же значение и рассеяния, конечно, не имеет.

*Свойство 2.* Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

$$D(CX)=C^2D(X) \qquad (2.17)$$

*Доказательство*. По определению дисперсии имеем

$$D(CX)=M[CX-M(CX)]^2$$

Пользуясь вторым свойством математического ожидания (постоянный множитель можно выносить за знак математического ожидания) получим

$$D(CX)=M[CX-M(CX)]^2=M\{C^2[X-M(X)]^2\}=$$

$$=C^2M[X-M(X)]^2=C^2D(X)$$

Свойство становится ясным, если принять во внимание, что при $|C|>1$ величина $CX$ имеет возможные значения (по абсолютной величине) большие, чем величина $X$. Отсюда следует, что эти значения рассеяны вокруг математического ожидании $M(CX)$ больше, чем возможные значения $X$ вокруг $M(X)$, т. е. $D(CX)>D(X)$. Напротив, если $0<|C|<1$, то $D(CX)<D(X)$.

*Свойство 3*. Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин:

$$D(X+Y)=D(X)+D(Y) \tag{2.18}$$

*Доказательство*. По формуле для вычисления дисперсии имеем

$$D(X+Y)=M[(X+Y)^2]-[M(X+Y)]^2$$

Раскрыв скобки и пользуясь свойствами математического ожидания суммы нескольких величин и произведения двух независимых случайных величин, получим

$$D(X+Y)=M[X^2+2XY+Y^2]-[M(X)+M(Y)]^2=$$

$$=M(X^2)+2M(X)M(Y)+M(Y^2)-M^2(X)-$$

$$-2M(X)M(Y)-M^2(Y)=[M(X^2)-M^2(X)]+$$

$$+[M(Y^2)-M^2(Y)]=D(X)+D(Y)$$

*Следствие 1*. Дисперсия суммы нескольких взаимно независимых случайных величин равна сумме дисперсий этих величин.

Например, для трех слагаемых имеем

$$\begin{gathered}D(X+Y+Z)=D([X+Y]+Z)=D(X+Y)+D(Z)=\\=D(X)+D(Y)+D(Z)\end{gathered} \tag{2.19}$$

*Следствие 2*. Дисперсия суммы постоянной величины и случайной равна дисперсии случайной величины:

$$D(C+X)=D(X) \qquad (2.20)$$

*Доказательство*. Величины $C$ и $X$ независимы, поэтому по третьему свойству

$$D(C+X)=D(C)+D(X)$$

А в силу первого свойства $D(C)=0$.

Свойство становится понятным, если учесть, что величины $X$ и $X+C$отличаются лишь началом отсчета и, значит, рассеяны вокруг своих математических ожиданий одинаково.

*Свойство 4*. Дисперсия разности двух независимых случайных величин равна сумме их дисперсий:

$$D(X-Y)=D(X)+D(Y) \qquad (2.21)$$

*Доказательство*. В силу третьего свойства

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

По второму свойству

$$D(X-Y)=D(X)+(-1)^2 D(Y)=D(X)+D(Y)$$

*Свойство 5*. Дисперсия числа появлений события А в $n$независимых испытаниях, в каждом из которых вероятность появления события постоянна, равна произведению числа испытаний на вероятности появления и непоявления события в одном испытании:

$$D(X)=npq \qquad (2.22)$$

*Доказательство*. Рассмотрим случайную величину $X$ — число появлении события А в $n$ независимых испытаниях. Очевидно, общее число появлений события в этих испытаниях равно сумме появлений события в отдельных испытаниях:

$$X=X_1+X_2+...+X_n$$

где $X_1$— число наступлении события в первом испытании, $X_2$ — во втором, ..., $X_n$ — в $n$-ом.

Величины $X_1, X_2, \ldots, X_n$взаимно независимы, так как исход каждого испытания не зависит от исходов остальных, поэтому мы вправе воспользоваться следствием 1 свойства 3:

$$D(X) = D(X_1) + D(X_2) + \ldots + D(X_n) \qquad (2.23)$$

Вычислим дисперсию $X_1$ по формуле

$$D(X_1) = M(X_1^2) - [M(X_1)]^2 \qquad (2.24)$$

Величина $X_1$ есть число появлений события $A$ в первом испытании, поэтому $M(X_1)=p$.

Найдем математическое ожидание величины $X_1^2$, которая может принимать только два значения, а именно $1^2$ с вероятностью $p$ и $0^2$ с вероятностью $q$:

$$M(X_1^2) = 1^2 p + 0^2 q.$$

Подставляя найденные результаты в соотношение (2.24), имеем

$$D(X_1) = p - p^2 = p(1-p) = pq$$

Очевидно, дисперсия каждой из остальных случайных величин также равна $pq$. Заменив каждое слагаемое правой части (2.23) через $pq$, окончательно получим

$$D(X) = npq$$

Для оценки рассеяния возможных значений случайной величины вокруг ее среднего значения кроме дисперсии служат некоторые другие характеристики. К их числу относится среднее квадратическое отклонение.

Средним квадратическим отклонением случайной величины $X$ называют квадратный корень из дисперсии:

$$\sigma(X) = \sqrt{D(X)} \qquad (2.25)$$

Легко показать, что дисперсия имеет размерность равную квадрату размерности случайной величины. Так как среднее квадратическое отклонение равно квадратному корню из дисперсии, то размерность $\sigma(X)$ совпадет с размерностью $X$. Поэтому в тех случаях, когда желательно, чтобы оценка рассеяния имела размерность случайной величины, вычисляют среднее квадратическое отклонение, а не дисперсию.

Пусть известны средние квадратические отклонения нескольких взаимно независимых случайных величин. Как найти среднее квадратическое отклонение суммы этих величин? Ответ па этот вопрос дает следующая теорема.

*Теорема*. Среднее квадратическое отклонение суммы конечного числа взаимно независимых случайных величин равно квадратному корню из суммы квадратов средних квадратических отклонений этих величин:

$$\sigma(X_1 + X_2 + ... + X_n) = \sqrt{\sigma^2(X_1) + \sigma^2(X_2) + ... + \sigma^2(X_n)} \quad (2.26)$$

*Доказательство*. Обозначим через $X$ сумму рассматриваемых взаимно независимых величин:

$$X = X_1 + X_2 + ... + X_n$$

Так как дисперсия суммы нескольких взаимно независимых случайных величин равна сумме дисперсий слагаемых, то

$$D(X) = D(X_1) + D(X_2) + ... + D(X_n)$$

Отсюда

$$\sqrt{D(X)} = \sqrt{D(X_1) + D(X_2) + ... + D(X_n)}$$

или окончательно

$$\sigma(X_1 + X_2 + ... + X_n) = \sqrt{\sigma^2(X_1) + \sigma^2(X_2) + ... + \sigma^2(X_n)}$$

## 2.3. Одинаково распределенные взаимно независимые случайные величины

По закону распределения можно найти числовые характеристики случайной величины. Отсюда следует, что если несколько случайных величин имеют одинаковые распределения, то их числовые характеристики одинаковы.

Рассмотрим n взаимно независимых случайных величин $X_1$, $X_2$, .... $X_n$, которые имеют одинаковые распределения, а следовательно, и одинаковые характеристики (математическое ожидание, дисперсию и др.). Наибольший интерес представляет изучение числовых характеристик среднего арифметического этих величии, чем мы и займемся в настоящем параграфе.

Обозначим среднее арифметическое рассматриваемых случайных величин через *X*:

$$\bar{X} = \left(X_1 + X_2 + \ldots + X_n\right) / n \tag{2.27}$$

Следующие ниже три положения устанавливают связь между числовыми характеристиками среднего арифметического *X* и соответствующими характеристиками каждой отдельной величины.

1. Математическое ожидание среднего арифметического одинаково распределенных взаимно независимых случайных величин равно математическому ожиданию а каждой из ветчин:

$$M(\bar{X}) = a \tag{2.28}$$

*Доказательство*. Пользуясь свойствами математического ожидания (постоянный множитель можно вынести за знак математического ожидания; математическое ожидание суммы равно сумме математических ожидании слагаемых), имеем:

$$M(\bar{X}) = M\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) =$$

$$= \frac{M(X_1) + M(X_2) + \ldots + M(X_n)}{n}.$$

Приняв во внимание, что математическое ожидание каждой из величин по условию равно *a*, получим

$$M(\bar{X}) = \frac{na}{n} = a$$

2. Дисперсия среднего арифметического *n* одинаково распределенных взаимно независимых случайных величин в *n* раз меньше дисперсии *D* каждой из величин:

$$D(\bar{X}) = \frac{D}{n} \tag{2.29}$$

*Доказательство*. Пользуясь свойствами дисперсии (постоянный множитель можно вынести за знак дисперсии, возведя его в квадрат; дисперсия суммы независимых ветчин равна сумме дисперсий слагаемых), имеем

$$D(\bar{X}) = D\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) =$$

$$= \frac{D(X_1) + D(X_2) + ... + D(X_n)}{n^2}.$$

Приняв во внимание, что дисперсия каждой из величин по условию равна *D*, получим:

$$D(\bar{X}) = \frac{nD}{n^2} = \frac{D}{n}$$

3. Среднее квадратическое отклонение среднего арифметического *n* одинаково распределенных взаимно независимых случайных величин в $\sqrt{n}$ раз меньше среднего квадратического отклонения $\sigma$ каждой из величин:

$$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} \qquad (2.30)$$

*Доказательство*. Так как $D(\bar{X}) = D / n$, то среднее квадратическое отклонение $\bar{X}$ равно

$$\sigma(\bar{X}) = \sqrt{D(\bar{X})} = \sqrt{D / n} = \sigma / \sqrt{n}$$

Общий вывод из формул (2.29) и (2.30): вспоминая, что дисперсия и среднее квадратическое отклонение служат мерами рассеяния случайной величины, заключаем, что среднее арифметическое достаточно большого числа взаимно независимых случайных величин имеет значительно меньшее рассеяние, чем каждая отдельная величина.

## 3. Функции распределения и их свойства

### 3.1. Биномиальное распределение.

Дискретная случайная величина *X* имеет биномиальный закон распределения с параметрами *n* и *p*, если она принимает значения 0, 1, 2, …, *m*, …, *n* с вероятностями

$$P(X = m) = C_n^m p^m q^{n-m}, \qquad (3.1)$$

где $0<p<1$, $q=1-p$.

Биномиальный закон распределения представляет собой закон распределения числа $X=m$ наступлений событий $A$ в $n$ независимых испытаниях, в каждом из которых оно может произойти с одной и той же вероятностью $p$.

Ряд распределения биномиального закона имеет вид:

| $x_i$ | 0 | 1 | 2 | … | $m$ | … | $n$ |
|---|---|---|---|---|---|---|---|
| $p_i$ | $q$ | $C_n^1 pq^{n-1}$ | $C_n^2 p^2 q^{n-2}$ | ... | $C_n^m p^m q^{n-m}$ | ... | $p$ |

Очевидно, что определение биномиального закона корректно, так как основное свойство ряда распределения $\sum_{i=1}^{n} p_i = 1$ выполнено, ибо $\sum_{i=1}^{n} p_i$ есть не что иное, как сумма всех членов разложения бинома Ньютона:

$$q + C_n^1 pq^{n-1} + C_n^2 p^2 q^{n-2} + ... + C_n^m p^m q^{n-m} + ... + p = (p+q)^n = 1$$

(отсюда и название закона — биномиальный).

Математическое ожидание: $M(X)=np$.

Дисперсия: $D(X)=npq$.

Биномиальный закон распределения широко используется в теории и практике статистического контроля качества продукции, при описании функционирования систем массового обслуживания, при моделировании цен активов, в теории стрельбы и т.д. (рис.3).

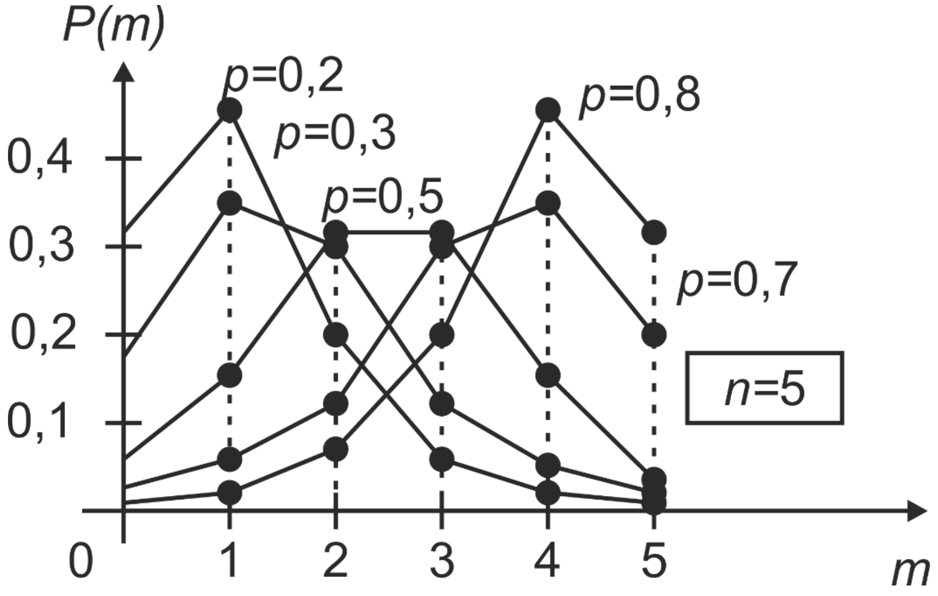


*Рис. 3Многоугольник (полигон) распределения случайной величины X, имеющей биномиальный закон распределения с параметрами n=5, p=0,2;0,3;0,5;0,7;0,8*

## 3.2. Распределение Пуассона.

Дискретная случайная величина *X* имеет закон распределения Пуассона с параметром $\lambda > 0$, если она принимает значения 0, 1, 2, …, *m*, … (бесконечное, но счетное множество значений) с вероятностями

$$P(X = m) = \frac{\lambda^m e^{-\lambda}}{m!} = P_m(\lambda), \tag{3.2}$$

Ряд распределения Пуассона имеет вид:

| $x_i$ | 0 | 1 | 2 | … | m | … |
|---|---|---|---|---|---|---|
| $p_i$ | $e^{-\lambda}$ | $\lambda e^{-\lambda}$ | $\frac{\lambda^2 e^{-\lambda}}{2!}$ | … | $\frac{\lambda^m e^{-\lambda}}{m!}$ | … |

Очевидно, что определение закона Пуассона корректно, так как основное свойство ряда распределения $\sum_{i=1}^{n} p_i = 1$ выполнено, ибо сумма ряда

$$\sum_{i=1}^{n} p_i = e^{-\lambda} + \lambda e^{-\lambda} + \ldots + \frac{\lambda^m e^{-\lambda}}{m!} + \ldots =$$

$$= e^{-\lambda}(1 + \lambda + \ldots + \frac{\lambda^m}{m!} + \ldots) = e^{-\lambda} e^{\lambda} = 1$$

Математическое ожидание:M(X)=λ.

Дисперсия:D(X)=λ.

При достаточно больших $n$ ($n \to \infty$) и малых значениях $p$ ($p \to 0$) при условии, что произведение $np$ – постоянная величина ($np \to \lambda = const$), закон распределения Пуассона является предельным случаем биномиального закона (закон массовых и редких событий). Кроме этого, по закону Пуассона распределены число сбоев на автоматической линии, число отказов сложной системы в «нормальном режиме», число «требований на обслуживание», поступивших в единицу времени в системах массового обслуживания и др. (рис.4).

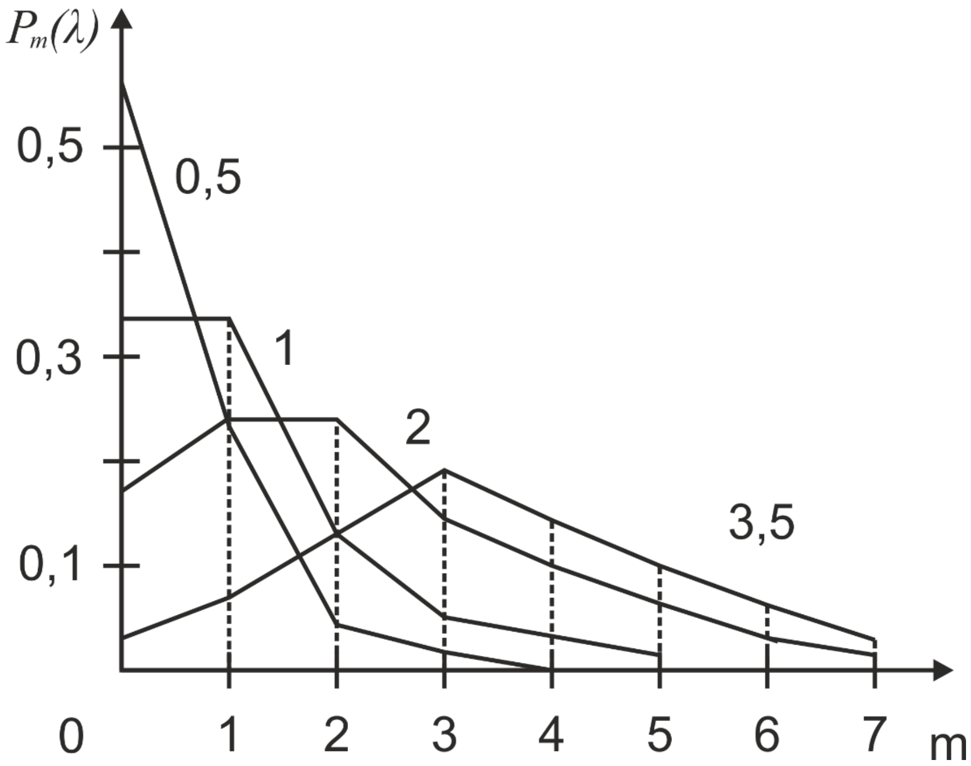


*Рис. 4 Многоугольник (полигон) распределения случайной величины, распределенной по закону Пуассона$P(X=m)=P_m(\lambda)$ параметрами $\lambda=0,5$, $\lambda=1$, $\lambda=2$, $\lambda=3,5$*

### 3.3. Равномерный закон распределения.

Непрерывная случайная величина *X* имеет равномерный закон распределения на отрезке *[a, b]*, если ее плотность вероятности постоянна на этом отрезке и равна нулю вне его, т.е.

$$\phi(x)=\begin{cases}\dfrac{1}{b-a}, \text{при } a\le x\le b\\ 0 \quad\quad, \text{при } x<a, x>b\end{cases} \tag{3.3}$$

Функция распределения случайной величины *X*, распределенной по равномерному закону, есть

$$F(x)=\begin{cases}0 \quad\quad, \text{при } x<a,\\ \dfrac{x-a}{b-a}, \text{при } a\le x\le b,\\ 1 \quad\quad, \text{при } x>b\end{cases} \tag{3.4}$$

Кривая распределения *φ(x)* и график функции распределения *F(x)* случайной величины *X* приведены на рис. 5.

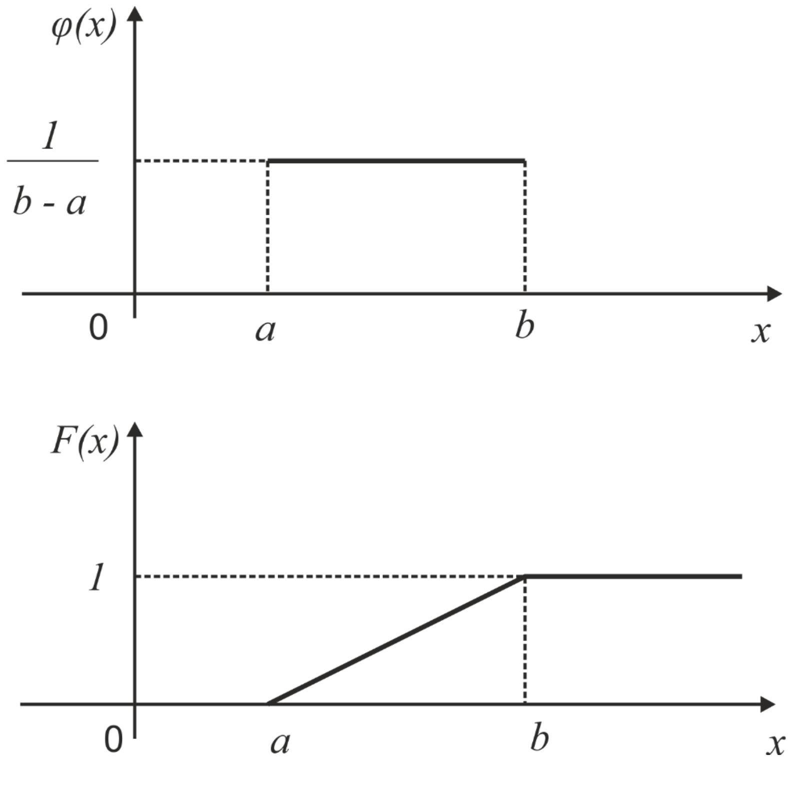


*Рис. 5 Графики плотности вероятности φ(x) и функции распределения F(x) случайной величины, распределенной по равномерному закону*

Математическое ожидание: $M(X)=\dfrac{a+b}{2}$.

Дисперсия: $D(X)=\dfrac{(b-a)^2}{12}$.

Равномерный закон распределения используется при анализе ошибок округления при проведении числовых расчетов, в ряде задач массового обслуживания, при статистическом моделировании наблюдений, подчиненных заданному распределению.

## 3.4. Показательный (экспоненциальный) закон распределения.

Непрерывная случайная величина X имеет показательный (экспоненциальный) закон распределения с параметром λ >0, если ее плотность вероятности имеет вид:

$$\phi(x)=\begin{cases}\lambda e^{-\lambda x}, \text{ при } x\ge 0,\\ 0 \quad\quad , \text{ при } x<0\end{cases} \tag{3.5}$$

Кривая распределения φ(x) и график функции распределения F(x) случайной величины X приведены на рис. 6.

Математическое ожидание: $M(X)=1/\lambda$.

Дисперсия: $D(X)=1/\lambda^2$.

Показательный закон распределения играет большую роль в теории массового обслуживания и теории надежности.

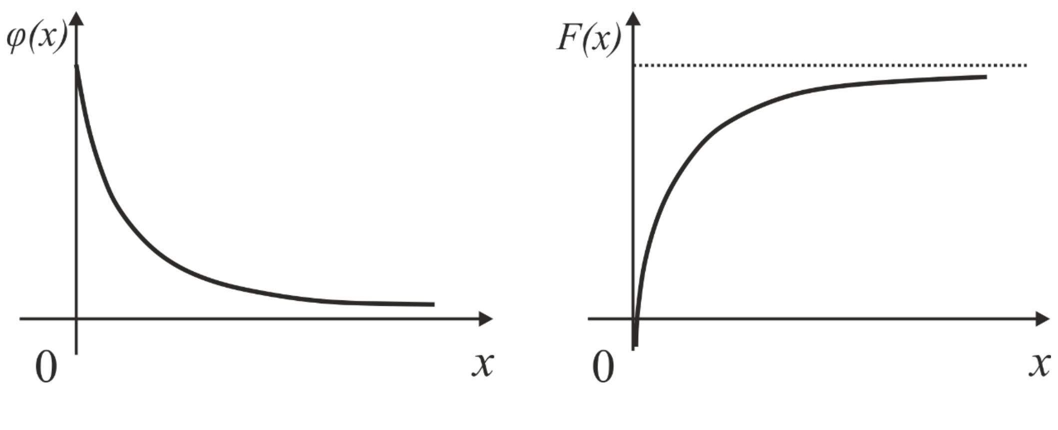


*Рис. 6 Кривая распределения φ(x) и график функции распределения F(x) случайной величины X*

## 3.5. Нормальный закон распределения.

Непрерывная случайная величина X имеет нормальный закон распределения с параметрами а и $\sigma^2$, если ее плотность вероятности имеет вид:

$$\phi_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad (3.6)$$

Обозначение: $X \sim N(a;\sigma^2)$.

Математическое ожидание: $M(X) = a$.

Дисперсия: $D(X) = \sigma^2$.

Кривую нормального закона распределения называют нормальной или гауссовой кривой (рис.7). При увеличении σ ордината максимума кривой уменьшается, но так как площадь под любой кривой распределения должна оставаться равной единице, то кривая становится более плоской, растягиваясь вдоль оси абсцисс; при уменьшении σ, напротив, нормальная кривая вытягивается вверх, одновременно сжимаясь с боков. Таким образом, параметр а (он же математическое ожидание) характеризует положение центра, а параметр σ (он же дисперсия) — форму нормальной кривой.

Функция распределения нормально распределенной случайной величины имеет вид:

$$F_N(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt \qquad (3.7)$$

Вероятность попадания случайной величины X, распределенной по нормальному закону, в интервале [$x_1$, $x_2$], равна

$$P(\alpha \le X \le \beta) = \frac{1}{2}\left[\Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)\right] \qquad (3.8)$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ – функция Лапласа.

Нормальный закон распределения занимает центральное место в теории и практике вероятностно-статистических методов.

Он является предельным законом, к которому приближаются многие другие законы распределения. Центральная предельная теорема теории вероятностей утверждает, что сумма очень большого числа случайных величин, влияние каждой из которых близко к 0, имеет распределение, близкое к нормальному.

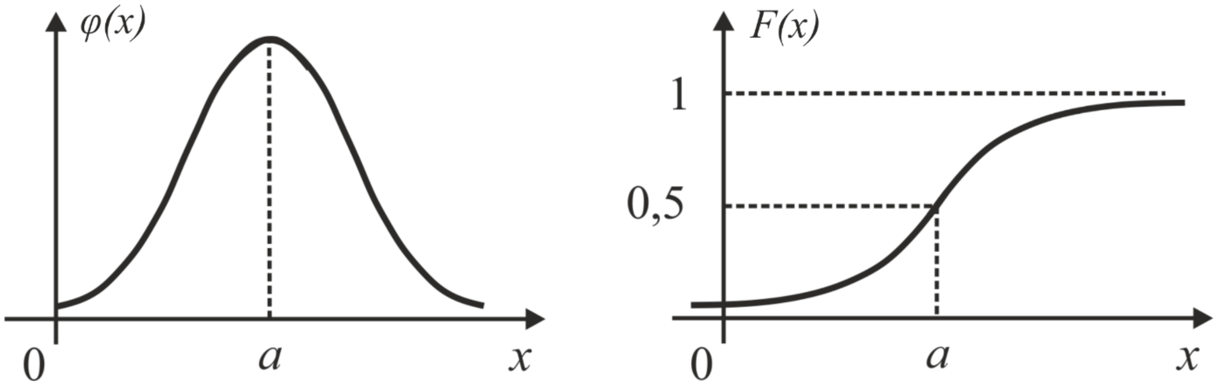


*Рис. 7 Нормальная кривая $\phi_N(x)$ с параметрами а и$\sigma^2$, т.е. N(a;$\sigma^2$), и график функции распределения случайной величины X, имеющей нормальный закон*

### 3.6. Распределение хи-квадрат ($\chi^2$).

Распределением хи-квадрат ($\chi^2$) с $k$ степенями свободы называется распределение суммы квадратов $k$ независимых случайных величин, распределенных по стандартному нормальному закону, т.е.

$$\chi^2 = Z_1^2 + \ldots + Z_k^2 \tag{3.9}$$

где$Z_i$имеют нормальное распределение$N(0,1)$.

Плотность вероятности $\chi^2$-распределения имеет вид:

$$\phi(x) = \begin{cases} \dfrac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma\left(\dfrac{k}{2}\right)}, \text{ при } x \geq 0, \\ 0 \qquad\qquad\quad , \text{ при } x < 0 \end{cases} \tag{3.10}$$

где $\Gamma(y) = \int\limits_0^{+\infty} e^{-t} t^{y-1} dt$ — гамма-функция Эйлера (для целых положительных значений $\Gamma(y) = (y-1)!$. Кривые $\chi^2$ для различных значений числа степеней свободы к приведены на рис. 8.

В математической статистике распределение хи-квадрат используется для построения интервальных оценок и статистических критериев.

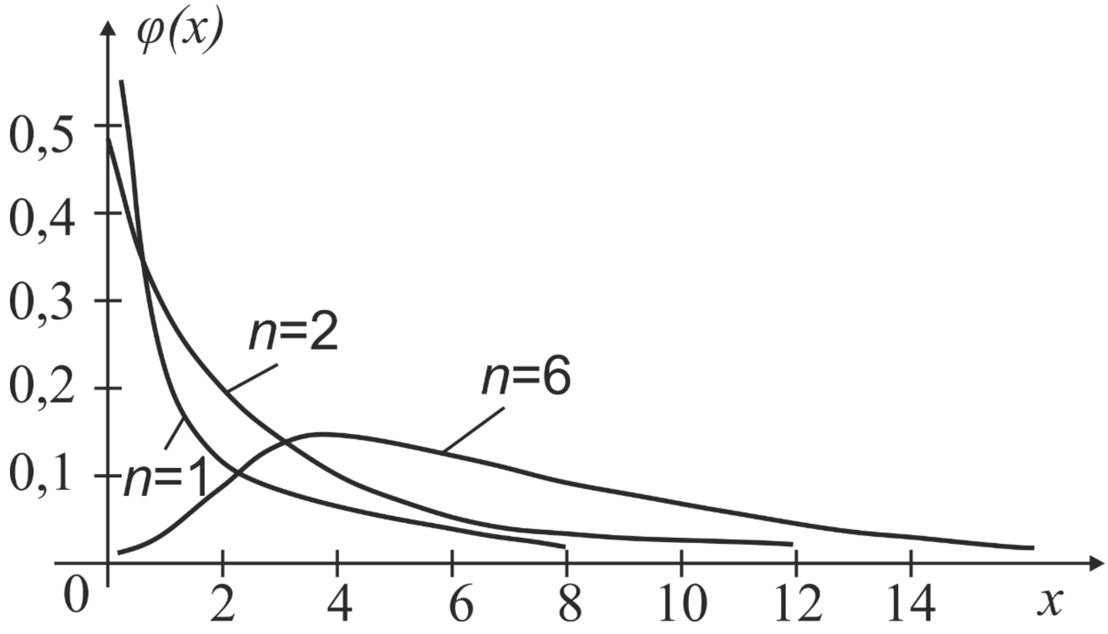


*Рис. 8 Кривые $\chi^2$-распределения для различных значений числа степеней свободы*

## 3.7. Распределение Стьюдента (t-распределение).

Распределением Стьюдента (или $t$-распределением) называется распределение случайной величины

$$t = \frac{Z}{\sqrt{\frac{1}{k}\chi^2}} \qquad (3.11)$$

где $Z$ – случайная величина, распределенная по стандартному нормальному закону, т.е. $N(0,1)$; $\chi^2$ – независимая от $Z$ случайная величина, имеющая $\chi^2$-распределение с $k$ степенями свободы.

Математическое ожидание: $M(t) = 0, k > 1$.

Дисперсия: $D(t) = \dfrac{k}{k-2}, k > 2$.

Плотность вероятности распределения Стьюдента имеет вид:

$$\phi(x)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}} \tag{3.12}$$

На рис. 9 показана кривая распределения Стьюдента. Как и стандартная нормальная кривая, кривая $t$-распределения симметрична относительно оси ординат, но по сравнению с нормальной более пологая.

Распределение Стьюдента применяется в статистике для построения доверительных интервалов и тестирования гипотез, касающихся неизвестного среднего статистической выборки из нормального распределения.

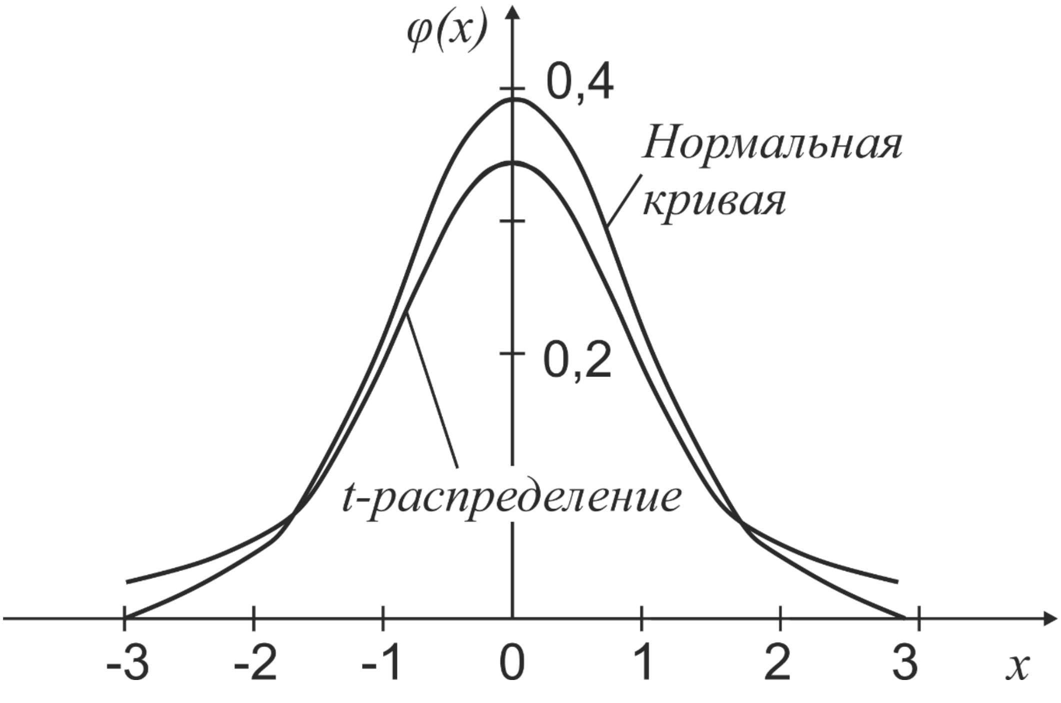


*Рис. 9 Распределение Стьюдента и нормальное распределение N(0;1)*

## 3.8. Распределение Фишера.

Распределением Фишера (или F-распределением) называется распределение случайной величины

$$F = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}, \tag{3.13}$$

где $\chi^2(k_1)$ и $\chi^2(k_2)$ – независимые случайные величины, имеющие распределение хи-квадрат со степенями свободы $k_1$ и $k_2$ соответственно.

На рис. 10 показаны кривые $F$-распределения при некоторых значениях числа степеней свободы $k_1$и $k_2$. Распределение Фишера используют при проверке гипотез об адекватности модели в регрессионном анализе, о равенстве дисперсий и др.

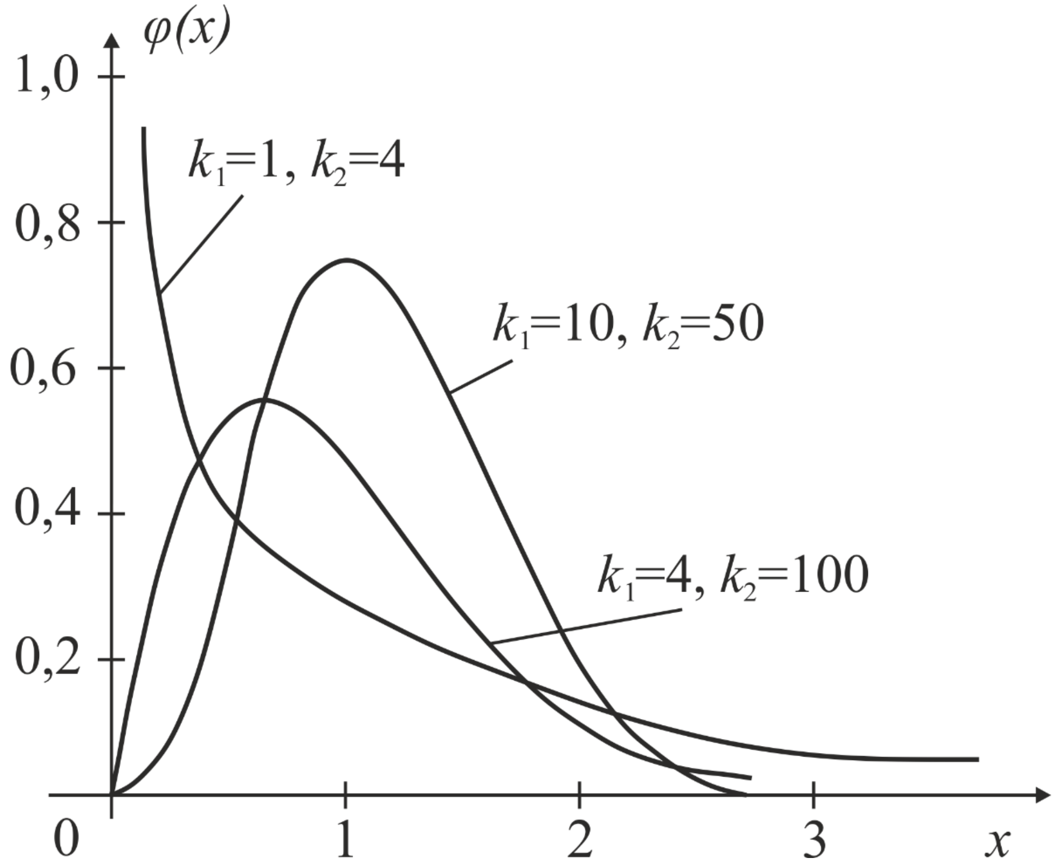


*Рис. 10. Кривые F-распределения при некоторых значениях числа степеней свободы*

# 4. Оценка погрешности прямых и косвенных измерений

## 4.1. Случайная ошибка и ее описание

Как было отмечено ранее, ошибки делятся на случайные и систематические. Есть еще промахи, но о них мы уже говорили. Кроме того, измерения бывают однократные и многократные, прямые и косвенные и т.д. Как уже отмечалось выше, все результаты измерения содержат случайную погрешность, а значит, при обработке таких результатов необходимо использовать аппарат и методы теории вероятности. Поскольку теория вероятности описывает характеристики генеральной совокупности, то ее результаты непосредственно применимы при обработке результатов эксперимента, только если мы имеем дело с генеральной совокупностью. Но мы в реальной практике всегда имеем дело с выборкой, т.е. с конечным числом измерений. Для описания таких процессов существует наука, которая называется математической статистикой. Суть ее в том, что она описывает поведение случайной величины и позволяет по выборке судить о генеральной совокупности и о ее параметрах. Но проблема состоит в том, что мы априори не знаем функцию распределения ошибок измерения даже в простейшем случае прямого измерения. Мы должны одновременно решать две взаимосвязанные задачи — определять "истинное" значение измеряемой величины и вычислять погрешность. Следует хорошо усвоить, что и измеренное нами значение и наши оценки являются величинами случайными, причем относительно характера распределения этих величин мы можем выдвигать более или менее обоснованные гипотезы, которые, вообще говоря, подлежат проверке в том же или другом эксперименте. Имеются как минимум две гипотезы, на которых чаще всего базируется теория ошибок:

a. Погрешность непосредственно измеряемой величины аддитивна, т.е. измеренное значение $x_i=X+\delta x$, где $X$ — "истинное" значение, а $\delta x$ — погрешность данного измерения (случайная величина). Это допущение настолько очевидно, что зачастую его переносят и на те области, в которых оно несправедливо. Ключевым словом в допущении является непосредственно измеряемая — если же речь идет о результате косвенного измерения, то очевидно, что относительно него мы не можем сделать такого вывода.

b. Предположение о характере функции распределения случайной величины — погрешности измерения. Это прежде всего гипотеза о нулевом математическом ожидании погрешности и гипотеза о конечности дисперсии этого распределения. Как правило, хотя и не всегда, сюда же добавляется предположение о нормальном (гауссовом) виде функции распределения.

Если первая гипотеза представляется почти очевидной, то использование второй чаще всего обосновывается центральной предельной теоремой, суть которой состоит в том, что распределение суммы случайных независимых величин, имеющих близкие распределения, стремится при увеличении их числа к распределению Гаусса. В реальности, при проведении экспериментов таких случайных составляющих у погрешности достаточно много, так что при рассмотрении случайных погрешностей, как правило, есть основания считать распределение гауссовым.

При обработке результатов измерения величины, которая имеет определенное значение, но в результате влияния различных случайных факторов измеряется нами с некоторой случайной ошибкой, возникает задача — используя конечный набор эмпирических данных$x_i$, полученных в выборке из n измерений, найти “наилучшее” значение оценки $x$ “точного” значения измеряемой величины $X$ и определить точность наших измерений. В курсе математической статистики доказывается, что наилучшей оценкой величины $X$, непосредственно измеряемой и удовлетворяющей пунктам 1-2, является среднее значение выборки

$$\hat{x} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (4.1)$$

Поскольку мы предположили, что распределение ошибок (отклонений измеренных значений от истинного $X - x_i$) гауссово, то для оценки величины этой погрешности хорошо бы знать дисперсию $\sigma^2$ этого распределения. Вероятность того, что случайная величина $x_i$ отклонится от среднего значения не более, чем на величину среднего квадратического отклонения $\sigma$, равна 0.683, на $2\sigma$ — 0.955, на $3\sigma$— 0.997. Наилучшей оценкой среднего квадратического отклонения является величина

$$\hat{\sigma} = S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{4.2}$$

Хотя оценка величины погрешности одного измерения $x_i$ и представляет интерес, тем не менее гораздо важнее знать с какой точностью значение $\hat{x}$, найденное нами из некоторой выборки, соответствует истинному значению искомой величины $X$. Поскольку $\hat{x}$ найдено из ограниченного числа измерений, то повторяющиеся серии измерений давали бы нам новые значения $\bar{x}$. То есть, эмпирические средние тоже являются случайной величиной и их поведение можно описать некоторой функцией распределения относительно величины математического ожидания со своей дисперсией $S_x$. Теория показывает, что если $\bar{x}$ определено из $n$ измерений, то

$$S_x = \frac{\sigma}{\sqrt{n}} \tag{4.3}$$

Это выражение, однако, невозможно использовать, т.к. σ нам неизвестна. Вместо σ в выражении(4.3) используется оценка среднего квадратического отклонения, а полученную величину $S_x$ мы далее будем называть стандартной ошибкой. Выражение для нее

$$S_x = \frac{S_n}{\sqrt{n}} \tag{4.4}$$

содержит только измеряемые экспериментально величины. Легко видеть, что стандартная ошибка $S_x$, в отличие от стандартного отклонения $S_n$, уменьшается с ростом числа измерений приблизительно как $\frac{1}{\sqrt{n}}$, т.е. точность результата возрастает с ростом числа испытаний. Естественно, однако, что нет смысла увеличивать количество измерений после того, как величина $S_x$ станет сравнима с величиной систематической ошибки (например, величиной ошибки измерительного прибора).

Поскольку мы оценили дисперсию отклонения эмпирического среднего $\bar{x}$ от истинного значения величины $X$, то, используя распределение Гаусса для этой величины, можно определить вероятность того, что это отклонение лежит в диапазоне $\pm u$. Однако чаще представляет интерес обратная задача: при выбранном уровне вероятности $P$ найти величину доверительного интервала. В идеализированном случае, когда $\sigma$ точно известна, можно задать необходимую доверительную вероятность. Мы можем утверждать, что с вероятностью $P$ отклонение измеренного значения $\bar{x}$ от $X$ не превышает следующей величины:

$$\left|\bar{x} - X\right| < t(P)\frac{\sigma}{\sqrt{n}} \qquad (4.5)$$

В реальном случае дисперсия генеральной совокупности неизвестна, и поэтому кроме выборочного среднего $\bar{x}$ используется выборочная оценка дисперсии $S_n^2$. В теории показывается, что случайная величина

$$t = \frac{\bar{x} - X}{S_x}$$

подчиняется распределению Стьюдента. Доверительная оценка при выбранном уровне статистической достоверности $P$в этом случае принимает вид

$$\left|\bar{x} - X\right| < t(P,\nu)S_x \qquad (4.6)$$

где$\nu=n-1$.

Функция $t(P, \nu)$, называемая далее коэффициентом Стьюдента, зависит не только от $P$, но и от числа измерений ($\nu=n-1$). Таблицы значений этого коэффициента для нескольких уровней доверительной вероятности (называемых еще уровнями надежности) $P$ и различных $n$ приведены в табл. 2.

*Таблица 2*

*Величины t для различных значений доверительного уровня P*

| n-1 | P=68.3% | P=95% | P=99% | P=99.73% |
|---|---|---|---|---|
| 2 | 1.32 | 4.7 | 9.9 | 19.2 |
| 3 | 1.2 | 3.18 | 5.8 | 9.2 |
| 4 | 1.15 | 2.78 | 4.6 | 6.6 |
| 5 | 1.11 | 2.57 | 4 | 5.5 |
| 6 | 1.09 | 2.45 | 3.7 | 4.9 |
| 7 | 1.08 | 2.37 | 3.5 | 4.5 |
| 8 | 1.07 | 2.31 | 3.4 | 4.3 |
| 9 | 1.06 | 2.26 | 3.2 | 4.1 |
| 10 | 1.05 | 2.23 | 3.2 | 4 |
| 15 | 1.03 | 2.13 | 3 | 3.6 |
| 20 | 1.03 | 2.09 | 2.8 | 3.4 |
| 30 | 1.02 | 2.04 | 2.8 | 3.3 |
| 50 | 1.01 | 2.01 | 2.7 | 3.2 |
| 100 | 1 | 1.98 | 2.6 | 3.1 |
| 200 | 1 | 1.97 | 2.6 | 3 |
| ∞ | 1 | 1.96 | 2.58 | 3 |

## 4.2. Правила обработки прямых и косвенных измерений.

При проведении прямого измерения некоторой величины необходимо:

a. Провести многократные измерения при одних и тех же условиях и записать их в таблицу.
b. Рассчитать среднее значение по формуле (4.1):
c. Вычислить оценку дисперсии

$$\hat{\sigma}^2 = S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

d. Вычислить среднеквадратичную ошибку среднего

$$S_x = \frac{S_n}{\sqrt{n}}$$

e. Задавшись требуемым уровнем доверительной вероятности $P$, определить по табл. 2 коэффициент Стьюдента $t(P, n-1)$ и модуль доверительного интервала

$$\Delta x = t(P, \nu) S_x \tag{4.7}$$

f. Округлив соответствующие результаты, записать ответ в виде $X = \overline{x} \pm \Delta x$ при доверительной вероятности $P$.

В большинстве экспериментов измеряется не непосредственно величина, подлежащая определению, а другая величина или ряд величин, зависящих от интересующей нас величины тем или иным образом. При таких измерениях, которые называются косвенными, необходимо уметь вычислять ошибку результата. Когда интересующая величина у является функцией измеряемых величин $x_1, x_2, \ldots, x_k$

$$y = f(x_1, x_2, \ldots, x_k) \tag{4.8}$$

а для случайных ошибок, измеряемых непосредственно, справедливы описанные выше допущения, необходимо действовать следующим образом:

a. Оценка математического ожидания $\hat{y}$ функции от нескольких измеренных переменных вычисляется по формуле

$$\hat{y} = \overline{y} = f(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k) \tag{4.9}$$

b. Оценка среднего квадратического отклонения вычисляется способом, основанным на дифференциальном исчислении. Если среднеквадратичные ошибки $S_{x_1}, S_{x_2}, \ldots S_{x_k}$ величин $x_1, x_2, \ldots, x_k$, определенные по данным измерений малы по сравнению с измеряемыми величинами, то

$$S_y = \sqrt{\left(\frac{\partial f}{\partial x_1} S_{x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2} S_{x_2}\right)^2 + \ldots + \left(\frac{\partial f}{\partial x_k} S_{x_k}\right)^2} \tag{4.10}$$

Где $\partial f / \partial x_i$ — частная производная функции $f$, вычисленная при значении переменных, соответствующих средним значениям $x_1, x_2, \ldots, x_k$.

Для определения доверительного интервала используется та же процедура, что и при определении доверительного интервала для непосредственно измеряемой величины.

Сравнивая слагаемые в подкоренном выражении, легко определить ошибки какой из измеряемых величин вносят основной вклад в суммарную ошибку. Предварительный анализ выражения для $S_y$ полезен для того, чтобы при планировании постановки эксперимента попытаться добиться повышенной точности измерения величины, вносящей основной вклад в ошибку, а также понять, какие измерения следует производить тщательно, а на какие не следует тратить больших усилий.

## 5. Парная регрессия и корреляция

Парная регрессия представляет собой регрессию между двумя переменными – $y$ и $x$, т. е. модель вида:

$$y = f\left(x\right) \tag{5.1}$$

где $y$ – зависимая переменная (результирующий показатель); $x$ – независимая, или объясняющая, переменная (фактор-аргумент). Знак «^» означает, что между переменными $x$ и $y$ нет строгой функциональной зависимости, поэтому в каждом отдельном случае величина $y$ складывается из двух слагаемых:

$$y = y_x + \varepsilon \tag{5.2}$$

где $y$ – фактическое значение результирующего показателя; $\hat{y}_x$ – теоретическое значение результирующего показателя, найденное исходя из уравнения регрессии; $\varepsilon$ – случайная величина, характеризующая отклонения реального значения результирующего показателя от теоретического, найденного по уравнению регрессии.

Случайная величина $\varepsilon$ называется также возмущением. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

От правильно выбранной спецификации модели зависит величина случайных ошибок: они тем меньше, чем в большей мере теоретические значения результирующего признака $\hat{y}_x$, подходят к фактическим данным $y$. К ошибкам спецификации относятся неправильный выбор той или иной математической функции для $\hat{y}_x$ и недоучет в уравнении регрессии какого-либо существенного фактора, т. е. использование парной регрессии вместо множественной.

Для получения хорошего результата обычно исключают из совокупности единицы с аномальными значениями исследуемых признаков.

Использование временной информации также представляет собой выборку из всего множества хронологических дат. Изменив временной интервал, можно получить другие результаты регрессии.

Наибольшую опасность в практическом использовании методов регрессии представляют ошибки измерения. Если ошибки спецификации можно уменьшить, изменяя форму модели (вид математической формулы), а ошибки выборки – увеличивая объем исходных данных, то ошибки измерения практически сводят на нет все усилия по количественной оценке связи между признаками.

Предполагая, что ошибки измерения сведены к минимуму, основное внимание в исследованиях уделяется ошибкам спецификации модели.

В парной регрессии выбор вида математической функции $\hat{y} = f(x)$ может быть осуществлен тремя методами:

a. графическим;
b. аналитическим, т.е. исходя из теории изучаемой взаимосвязи;
c. экспериментальным.

При изучении зависимости между двумя признаками графический метод подбора вида уравнения регрессии достаточно нагляден. Он основан на анализе корреляционного поля. Основные типы кривых, используемые при количественной оценке связей, представлены на рис. 11:

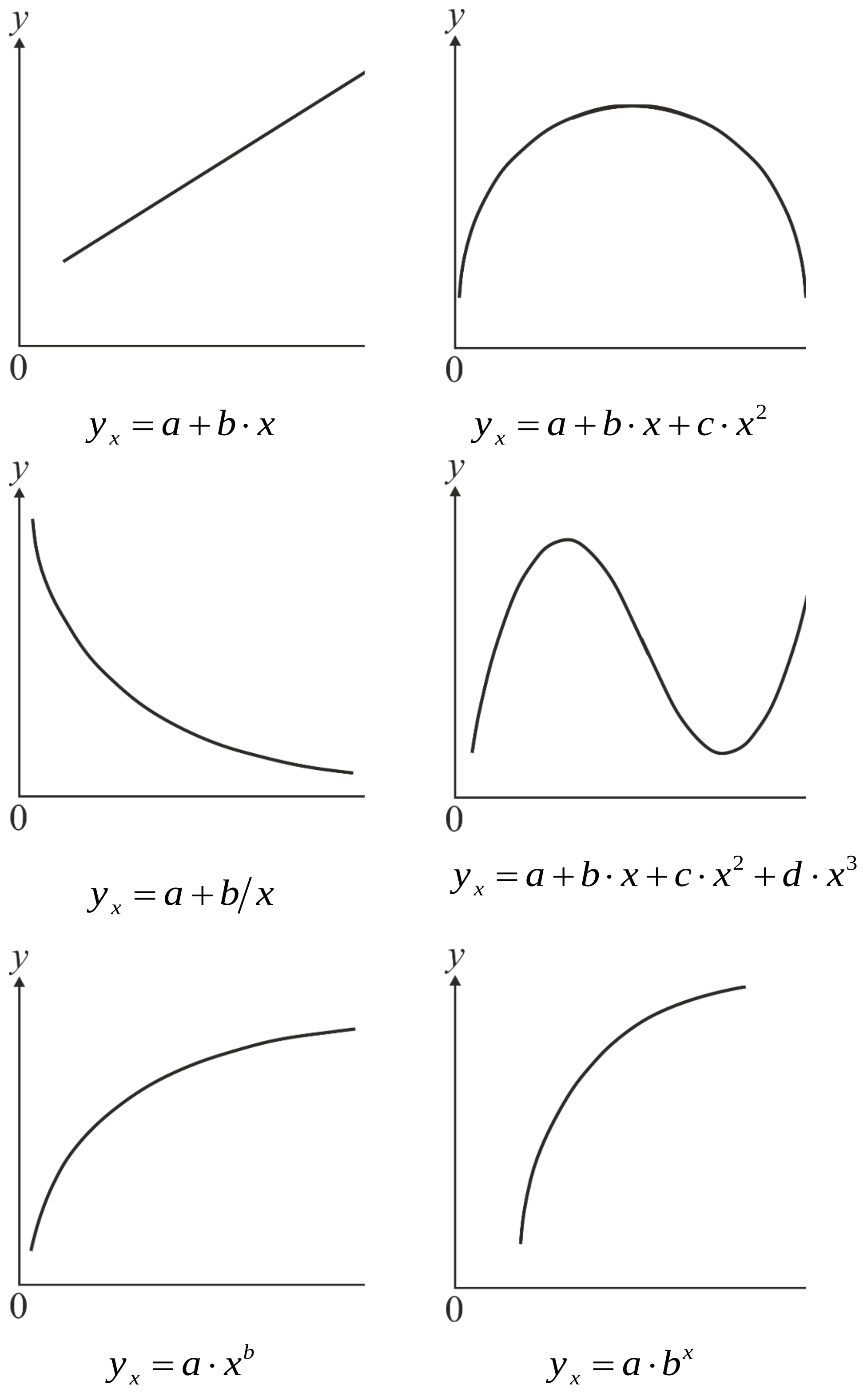


*Рис. 11. Основные типы кривых, используемые при количественной оценке связей между двумя переменными.*

Особый интерес представляет аналитический метод выбора типа уравнения регрессии. Он основан на изучении материальной природы связи исследуемых признаков.

При обработке информации на компьютере выбор вида уравнения регрессии обычно осуществляется экспериментальным методом, т. е. путем сравнения величины остаточной дисперсии $\sigma^2_{ост}$, рассчитанной при разных моделях.

Если уравнение регрессии проходит через все точки корреляционного поля, что возможно только при функциональной связи, когда все точки лежат на линии регрессии $\hat{y}_x = f(x)$, то фактические значения результативного показателя совпадают с теоретическими $\hat{y}_x = y$, т.е. они полностью обусловлены влиянием фактора *x*. В этом случае остаточная дисперсия $\sigma^2_{ост} = 0$.

В практических исследованиях, как правило, имеет место некоторое рассеяние точек относительно линии регрессии. Оно обусловлено влиянием прочих, не учитываемых в уравнении регрессии, факторов. Т.е. имеют место отклонения фактических данных от теоретических $y - \hat{y}_x$. Величина этих отклонений и лежит в основе расчета остаточной дисперсии:

$$\sigma^2_{\text{ост}} = \frac{1}{n}\sum\left(y - y_x\right)^2 \tag{5.3}$$

Чем меньше величина остаточной дисперсии, тем меньше влияние не учитываемых в уравнении регрессии факторов и тем лучше уравнение регрессии подходит к исходным данным.

Считают, что число наблюдений должно в 7-8 раз превышать число рассчитываемых параметров при переменной *x*. Это означает, что искать линейную регрессию, имея менее 7 наблюдений, вообще не имеет смысла. Если вид функции усложняется, то требуется увеличение объема наблюдений, ибо каждый параметр при *x* должен рассчитываться хотя бы по 7 наблюдениям. Таким образом, если в качестве модели выбирают параболу второй степени $y_x = a + b \cdot x + c \cdot x^2$, то требуется объем информации уже не менее 14 наблюдений.

Рассмотрим простейшую модель парной регрессии – линейную регрессию. Линейная регрессия находит широкое применение в эконометрике ввиду четкой экономической интерпретации ее параметров.

Линейная регрессия сводится к нахождению уравнения вида

$$y_x = a + b \cdot x \text{ или } y = a + b \cdot x + \varepsilon \tag{5.4}$$

Уравнение вида $y_x = a + b \cdot x$ позволяет по заданным значениям фактора $x$ находить теоретические значения результирующего показателя, подставляя в него фактические значения фактора $x$.

Построение линейной регрессии сводится к оценке ее параметров – $a$ и $b$. Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров $a$ и $b$, при которых сумма квадратов отклонений фактических значений результирующего показателя $y$ от теоретических $\hat{y}_x$ минимальна:

$$\sum_{i=1}^{n}\left(y_i - y_{x_i}\right)^2 = \sum_{i=1}^{n}\varepsilon_i^2 \quad \rightarrow \quad \min \tag{5.5}$$

Т.е. из всего множества линий линия регрессии на графике выбирается так, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была бы минимальной (рис. 12):


*Рис. 12 Линия регрессии с минимальной дисперсией остатков.*

$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum\left(y - a - b\cdot x\right) = 0; \\ \dfrac{\partial S}{\partial b} = -2\sum x\left(y - a - b\cdot x\right) = 0. \end{cases} \tag{5.6}$$

После несложных преобразований, получим следующую систему линейных уравнений для оценки параметров $a$ и $b$:

$$\begin{cases} a\cdot n \quad + b\cdot\sum x \; = \sum y; \\ a\cdot\sum x + b\cdot\sum x^2 = \sum x\cdot y. \end{cases} \tag{5.7}$$

Решая систему уравнений (5.7), найдем искомые оценки параметров $a$ и $b$:

$$a = \bar{y} - b\bar{x},\; b = \frac{\operatorname{cov}\left(x,\, y\right)}{\sigma_x^2} = \frac{\overline{yx} - \bar{y}\bar{x}}{\overline{x^2} - \bar{x}^2}, \tag{5.8}$$

где $\operatorname{cov}\left(x, y\right) = \overline{yx} - \overline{y}\overline{x}$ – ковариация признаков $x$ и $y$, $\sigma_x^2 = \overline{x^2} - \overline{x}^2$ – дисперсия признака $x$ и

$$\overline{x} = \frac{1}{n}\sum x,\; \overline{y} = \frac{1}{n}\sum y,\; \overline{yx} = \frac{1}{n}\sum yx,\; \overline{x^2} = \frac{1}{n}\sum x^2. \tag{5.9}$$

Параметр $b$ называется коэффициентом регрессии. Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

Формально $a$ – значение $y$ при $x = 0$. Если фактор $x$ не может иметь нулевого значения, то тогда трактовка свободного члена $a$ не имеет смысла, т.е. параметр $a$ может не иметь физического содержания.

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает линейный коэффициент корреляции $r_{xy}$, который можно рассчитать по следующим формулам:

$$r_{xy} = \frac{\operatorname{cov}\left(x,\, y\right)}{\sigma_x\cdot\sigma_y} = b\cdot\frac{\sigma_x}{\sigma_y}. \tag{5.10}$$

где $\sigma_y^2 = \frac{1}{n}\sum\left(y - \bar{y}\right)^2 = \overline{y^2} - \bar{y}^2$ – дисперсия признака $y$.

Линейный коэффициент корреляции находится в пределах: $-1 \le r_{xy} \le 1$. Чем ближе абсолютное значение $r_{xy}$ к единице, тем сильнее линейная связь между факторами (при $r_{xy} = \pm 1$ имеем строгую функциональную зависимость). Однако близость абсолютной величины линейного коэффициента корреляции к нулю еще не означает отсутствия связи между признаками. При другой (нелинейной) спецификации модели связь между признаками может оказаться достаточно тесной.

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции $r_{xy}^2$, называемый коэффициентом детерминации. Коэффициент детерминации характеризует долю дисперсии результирующего показателя $y$, объясняемую регрессией, в общей дисперсии результирующего показателя:

$$r_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2} \qquad (5.11)$$

где $\sigma_{\text{ост}}^2 = \frac{1}{n}\sum\left(y - y_x\right)^2$ – остаточная дисперсия.

Соответственно величина $1 - r_{xy}^2$ характеризует долю дисперсии $y$, вызванную влиянием остальных, не учтенных в модели, факторов.

После того как найдено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и отдельных его параметров.

Проверить значимость уравнения регрессии – означает установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации:

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y - y_x}{y}\right| \cdot 100\% \qquad (5.12)$$

Средняя ошибка аппроксимации не должна превышать 8–10%.

Оценка значимости уравнения регрессии в целом производится на основе F-критерия Фишера, которому предшествует дисперсионный анализ. В математической статистике дисперсионный анализ рассматривается как самостоятельный инструмент статистического анализа.

Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной $y$ от среднего значения $\bar{y}$ раскладывается на две части – «объясненную» и «необъясненную»:

$$\sum(y - \bar{y})^2 = \sum\left(y_x - \bar{y}\right)^2 + \sum\left(y - y_x\right)^2 \qquad (5.13)$$

где $\sum(y - \bar{y})^2$ – общая сумма квадратов отклонений; $\sum\left(y_x - \bar{y}\right)^2$ – сумма квадратов отклонений, объясненная регрессией (или факторная сумма квадратов отклонений); $\sum\left(y - y_x\right)^2$ – остаточная сумма квадратов отклонений, характеризующая влияние неучтенных в модели факторов.

Схема дисперсионного анализа имеет вид, представленный в таблице 3 ($n$ – число наблюдений, $m$ – число параметров при переменной $x$).

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину $F$-критерия Фишера:

$$F = \frac{S^2_{\text{факт}}}{S^2_{\text{ост}}} \qquad (5.14)$$

*Таблица 3*

*Схема дисперсионного анализа*

| Компоненты дисперсии | Сумма квадратов | Число степеней свободы | Дисперсия на одну степень свободы |
|---|---|---|---|
| Общая | $\sum(y-\bar{y})^2$ | $n-1$ | $S^2_{\text{общ}} = \dfrac{\sum(y-\bar{y})^2}{n-1}$ |
| Факторная | $\sum(y_x-\bar{y})^2$ | $m$ | $S^2_{\text{факт}} = \dfrac{\sum(y_x-\bar{y})^2}{m}$ |
| Остаточная | $\sum(y-y_x)^2$ | $n-m-1$ | $S^2_{\text{ост}} = \dfrac{\sum(y-y_x)^2}{n-m-1}$ |

Фактическое значение $F$-критерия Фишера (5.14) сравнивается с табличным значением $F_{\text{табл}}(\alpha;\, k_1;\, k_2)$ при уровне значимости $\alpha$ и степенях свободы $k_1 = m$ и $k_2 = n-m-1$. При этом, если фактическое значение $F$-критерия больше табличного, то признается статистическая значимость уравнения в целом.

Для парной линейной регрессии $m = 1$, поэтому

$$F = \frac{S^2_{\text{факт}}}{S^2_{\text{ост}}} = \frac{\sum(y_x-\bar{y})^2}{\sum(y-y_x)^2}\cdot(n-2) \qquad (5.15)$$

Величина $F$-критерия связана с коэффициентом детерминации $r^2_{xy}$, и ее можно рассчитать по следующей формуле:

$$F = \frac{r^2_{xy}}{1-r^2_{xy}}\cdot(n-2) \qquad (5.16)$$

В парной линейной регрессии оценивается значимость не только уравнения в целом, но и отдельных его параметров. С этой целью по каждому из параметров определяется его стандартная ошибка: $m_b$ и $m_a$.

Стандартная ошибка коэффициента регрессии определяется по формуле:

$$m_b = \sqrt{\frac{S_{\text{ост}}^2}{\sum\left(x - \bar{x}\right)^2}} = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}} \qquad (5.17)$$

где $S_{\text{ост}}^2 = \dfrac{\sum\left(y - y_x\right)^2}{n-2}$ – остаточная дисперсия на одну степень свободы.

Величина стандартной ошибки совместно с t-распределением Стьюдента при $n-2$ степенях свободы применяется для проверки существенности коэффициента регрессии и для расчета его доверительного интервала.

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т.е. определяется фактическое значение $t$-критерия Стьюдента: $t_b = \dfrac{b}{m_b}$ которое затем сравнивается с табличным значением при определенном уровне значимости $\alpha$ и числе степеней свободы $\left(n-2\right)$. Доверительный интервал для коэффициента регрессии определяется как $b \pm t_{\text{табл}} m_b$. Поскольку знак коэффициента регрессии указывает на рост результативного показателя $y$ при увеличении фактора $x$ $(b > 0)$, уменьшение результативного показателя при увеличении признака-фактора $(b < 0)$ или его независимость от независимой переменной $(b = 0)$, то границы доверительного интервала для коэффициента регрессии не должны содержать противоречивых результатов, например, $-1{,}2 \le b \le 0{,}8$. Такого рода запись указывает, что истинное значение коэффициента регрессии одновременно содержит положительные и отрицательные величины и даже ноль, чего не может быть.

Стандартная ошибка параметра $a$ определяется по формуле:

$$m_a = \sqrt{S_{\text{ост}}^2 \frac{\sum x^2}{n \cdot \sum\left(x - \bar{x}\right)^2}} = S_{\text{ост}} \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n} \qquad (5.18)$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии. Вычисляется $t$-критерий: $t_a = \dfrac{a}{m_a}$, его величина сравнивается с табличным значением при $n-2$ степенях свободы.

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции $m_r$:

$$m_r = \sqrt{\frac{1-r^2}{n-2}} \tag{5.19}$$

Фактическое значение $t$-критерия Стьюдента определяется как $t_r = \dfrac{r}{m_r}$.

Существует связь между $t$-критерием Стьюдента и $F$-критерием Фишера:

$$t_b = t_r = \sqrt{F} \tag{5.20}$$

В прогнозных расчетах по уравнению регрессии определяется предсказываемое $y_p$ значение как точечный прогноз $y_x$ при $x = x_p$, т.е. путем подстановки в уравнение регрессии $y_x = a + b \cdot x$ прогнозного значения $x$. Точечный прогноз дополняется расчетом доверительного интервала прогнозного значения $y_p$:

$$y_p - \Delta_{y_p} \le y_p^* \le y_p + \Delta_{y_p},$$

где $\Delta_{y_p} = m_{y_p} \cdot t_{\text{табл}}$, а $m_{y_p}$ – средняя ошибка точечного прогноза:

$$m_{y_p} = S_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{n \cdot \sigma_x^2}} \tag{5.21}$$

## 6. Задачи

### 6.1. Случайные величины

**6.1.1.** Вероятность поражения вирусным заболеванием куста земляники равна 0,2. Составить закон распределения числа кустов земляники, зараженных вирусом, из четырех посаженных кустов.

**6.1.2.** Стрелок ведет стрельбу по цели с вероятностью попадания при каждом выстреле 0,2. За каждое попадание он получает 5 очков, а в случае промаха очков ему не начисляют. Составить закон распределения числа очков, полученных стрелком за 3 выстрела, и вычислить математическое ожидание этой случайной величины.

**6.1.3.** В рекламных целях торговая фирма вкладывает в каждую десятую единицу товара денежный приз размером 1 тыс. руб. Составить закон распределения случайной величины — размера выигрыша при пяти сделанных покупках. Найти математическое ожидание и дисперсию этой случайной величины.

**6.1.4.** Клиенты банка, не связанные друг с другом, не возвращают кредиты в срок с вероятностью 0,1. Составить закон распределения числа возвращенных в срок кредитов из 5 выданных. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение этой случайной величины.

**6.1.5.** Контрольная работа состоит из трех вопросов. На каждый вопрос приведено 4 ответа, один из которых правильный. Составить закон распределения числа правильных ответов при простом угадывании. Найти математическое ожидание и дисперсию этой случайной величины.

**6.1.6.** В среднем по 10% договоров страховая компания выплачивает страховые суммы в связи с наступлением страхового случая. Составить закон распределения числа таких договоров среди наудачу выбранных четырех. Вычислить математическое ожидание и дисперсию этой случайной величины.

**6.1.7.** В билете три задачи. Вероятность правильного решения первой задачи равна 0,9, второй — 0,8, третьей — 0,7. Составить закон распределения числа правильно решенных задач в билете и вычислить математическое ожидание и дисперсию этой случайной величины.

**6.1.8.** Вероятность попадания в цель при одном выстреле равна 0,8 и уменьшается с каждым выстрелом на 0,1. Составить закон распределения числа попаданий в цель, если сделано три выстрела. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение этой случайной величины.

**6.1.9.** Произведено два выстрела в мишень. Вероятность попадания в мишень первым стрелком равна 0,8, вторым — 0,7, Составить закон распределения числа попаданий в мишень. Найти математическое ожидание, дисперсию и функцию распределения этой случайной величины и построить ее график. (Каждый стрелок делает по одному выстрелу.)

**6.1.10.** Найти закон распределения числа пакетов трех акций, по которым владельцем будет получен доход, если вероятность получения дохода по каждому из них равна соответственно 0,5, 0,6, 0,7. Найти математическое ожидание и дисперсию данной случайной величины, построить функцию распределения.

**6.1.11.** Дан ряд распределения случайной величины

| $x_i$ | 2 | 4 |
|---|---|---|
| $p_i$ | $p_1$ | $p_2$ |

Найти функцию распределения этой случайной величины, если ее математическое ожидание равно 3,4, а дисперсия равна 0,84.

**6.1.12.** Из пяти гвоздик две белые. Составить закон распределения и найти функцию распределения случайной величины, выражающей число белых гвоздик среди двух одновременно взятых.

**6.1.13.** Из 10 телевизоров на выставке 4 оказались фирмы «Сони». Наудачу для осмотра выбрано 3. Составить закон распределения числа телевизоров фирмы «Сони» среди 3 отобранных.

**6.1.14.** Среди 15 собранных агрегатов 6 нуждаются в дополнительной смазке. Составить закон распределения числа агрегатов, нуждающихся в дополнительной смазке, среди пяти наудачу отобранных из общего числа.

**6.1.15.** В магазине продаются 5 отечественных и 3 импортных телевизора. Составить закон распределения случайной величины — числа импортных из четырех наудачу выбранных телевизоров. Найти функцию распределения этой случайной величины и построить ее график.

**6.1.16.** Вероятность того, что в библиотеке необходимая студенту книга свободна, равна 0,3. Составить закон распределения числа библиотек, которые посетит студент, если в городе 4 библиотеки. Найти математическое ожидание и дисперсию этой случайной величины.

**6.1.17.** Экзаменатор задает студенту вопросы, пока тот правильно отвечает. Как только число правильных ответов достигнет четырех либо студент ответит неправильно, экзаменатор прекращает задавать вопросы. Вероятность правильного ответа на один вопрос равна 2/3. Составить закон распределения числа заданных студенту вопросов.

**6.1.18.** Торговый агент имеет 5 телефонных номеров потенциальных покупателей и звонит им до тех пор, пока не получит заказ на покупку товара. Вероятность того, что потенциальный покупатель сделает заказ, равна 0,4. Составить закон распределения числа телефонных разговоров, которые предстоит провести агенту. Найти математическое ожидание и дисперсию этой случайной величины.

**6.1.19.** Каждый поступающий в институт должен сдать 3 экзамена. Вероятность успешной сдачи первого экзамена 0,9, второго —- 0,8, третьего — 0,7. Следующий экзамен поступающий сдает только в случае успешной сдачи предыдущего. Составить закон распределения числа экзаменов, сдававшихся поступающим в институт. Найти математическое ожидание этой случайной величины.

**6.1.20.** Охотник, имеющий 4 патрона, стреляет по дичи до первого попадания или до израсходования всех патронов. Вероятность попадания при первом выстреле равна 0,6, при каждом последующем — уменьшается на 0,1. Необходимо: а) составить закон распределения числа патронов, израсходованных охотником; б) найти математическое ожидание и дисперсию этой случайной величины.

**6.1.21.** Из поступивших в ремонт 10 часов 7 нуждаются в общей чистке механизма. Часы не рассортированы по виду ремонта. Мастер, желая найти часы, нуждающиеся в чистке, рассматривает их поочередно и, найдя такие часы, прекращает дальнейший просмотр. Составить закон распределения числа просмотренных часов. Найти математическое ожидание и дисперсию этой случайной величины.

**6.1.22.** Имеются 4 ключа, из которых только один подходит к замку. Составить закон распределения числа попыток открывания замка, если испробованный ключ в последующих попытках не участвует. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение этой случайной величины.

**6.1.23.** Абонент забыл последнюю цифру нужного ему номера телефона, однако помнит, что она нечетная. Составить закон распределения числа сделанных им наборов номера телефона до попадания на нужный номер, если последнюю цифру он набирает наудачу, а набранную цифру в дальнейшем не набирает. Найти математическое ожидание и функцию распределения этой случайной величины.

**6.1.24.** Дана функция распределения случайной величины X

$$F(x)=\begin{cases}0 & \text{при } x\le 1,\\ 0{,}3 & \text{при } 1<x\le 2,\\ 0{,}7 & \text{при } 2<x\le 3,\\ 1 & \text{при } 3>x.\end{cases}$$

Найти: а) ряд распределения; б) *M(X)* и *D(X)*; в) построить многоугольник распределения и график *F(x)*.

**6.1.25.** Даны законы распределения двух независимых случайных величин

| $x_i$ | 0 | 1 | 3 |
|---|---|---|---|
| $p_i$ | 0,2 | 0,5 | ? |

| $y_i$ | 2 | 3 |
|---|---|---|
| $p_i$ | 0,4 | ? |

Найти вероятности, с которыми случайные величины принимают значение 3, а затем составить закон распределения случайной величины *3X-2Y* и проверить выполнение свойств математических ожиданий и дисперсий.

**6.1.26.** На двух автоматических станках производятся одинаковые изделия. Даны законы распределения числа бракованных изделий, производимых в течение смены на каждом из них:

а) для первого

| $x_i$ | 0 | 1 | 2 |
|---|---|---|---|
| $p_i$ | 0,1 | 0,6 | 0,3 |

б) для второго

| $y_i$ | 0 | 2 |
|---|---|---|
| $p_i$ | 0,5 | 0,5 |

Необходимо: а) составить закон распределения числа производимых в течение смены бракованных изделий обоими станками; б) проверить свойство математического ожидания суммы случайных величин.

**6.1.27.** Одна из случайных величин задана законом распределения

| $x_i$ | -1 | 0 | 1 |
|---|---|---|---|
| $p_i$ | 0,1 | 0,8 | 0,1 |

а другая имеет биномиальное распределение с параметрами $n = 2$, $p = 0,6$. Составить закон распределения их суммы и найти математическое ожидание этой случайной величины.

**6.1.28.** Случайные величины X и Y независимы и имеют один и тот же закон распределения:

| *Значение* | 1 | 2 | 4 |
|---|---|---|---|
| *Вероятность* | 0,2 | 0,3 | 0,5 |

Составить закон распределения случайных величин 2X и X+Y. Убедиться в том, что $2X \neq X+Y$, но $M(2X)= M(X+Y)$.

**6.1.29.** По данным примера 6.1.28 убедиться в том, что $X^2 \neq XY$. Проверить равенство $M(XY) = [M(X)]^2$.

**6.1.30.** Два стрелка сделали по два выстрела по мишени. Вероятность попадания в мишень для первого стрелка равна 0,6, для второго — 0,7. Необходимо: а) составить закон распределения общего числа попаданий; б) найти математическое ожидание и дисперсию этой случайной величины.

**6.1.31.** Пусть X, Y, Z — случайные величины: X — выручка фирмы, У— ее затраты, Z= X-Y — прибыль. Найти распределение прибыли Z, если затраты и выручка независимы и заданы распределениями:

| $x_i$ | 3 | 4 | 5 |
|---|---|---|---|
| $p_i$ | 1/3 | 1/3 | 1/3 |

| $y_j$ | 1 | 2 |
|---|---|---|
| $p_j$ | 0,5 | 0,5 |

**6.1.32.** Пусть X — выручка фирмы в долларах. Найти распределение выручки в рублях Z=X·Y в пересчете по курсу доллара Y, если выручка X не зависит от курса Y, а распределения X и Y имеют вид

| $x_i$ | 1000 | 2000 |
|---|---|---|
| $p_i$ | 0,7 | 0,3 |

| $y_j$ | 25 | 27 |
|---|---|---|
| $p_j$ | 0,4 | 0,6 |

**6.1.33.** Сделано два высокорисковых вклада: 10 тыс. руб. в компанию А и 15 тыс. руб. — в компанию В. Компания А обещает 50% годовых, но может «лопнуть» с вероятностью 0,2. Компания В обещает 40% годовых, но может «лопнуть» с вероятностью 0,15. Составить закон распределения случайной величины — общей суммы прибыли (убытка), полученной от двух компаний через год, и найти ее математическое ожидание.

**6.1.34.**Дискретная случайная величина X задана рядом распределения

| $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $p_i$ | 0,2 | 0,3 | 0,3 | 0,1 | 0,1 |

Найти условную вероятность события X<5 при условии, что X>2.

**6.1.35.** Случайные величины $X_1$, $X_2$ независимы и имеют одинаковое распределение

| $x_i$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p_i$ | 0,25 | 0,25 | 0,25 | 0,25 |

Найти вероятность события $X_1+X_2>2$

**6.1.36.** Распределение дискретной случайной величины X задано формулой $p(X= k) = Ck^2$, где k = 1, 2, 3, 4, 5. Найти: а) константу C; б) вероятность события $|X— 2|\leq 1$.

**6.1.37.** Распределение дискретной случайной величины X определяется формулой $P(X=k)=C/2^k$, к=0, 1, 2,... . Найти: а) константу C; б) вероятность $P(X \leq 3)$.

**6.1.38.** Случайная величина X, сосредоточенная на интервале [—1; 3], задана функцией распределения $F(x)=\frac{1}{4}x+\frac{1}{4}$. Найти вероятность попадания случайной величины X в интервал [0; 2]. Построить график функции F(x).

**6.1.39.** Случайная величина X, сосредоточенная на интервале (1; 4), задана квадратичной функцией распределения $F(x)=ax^2+bx+c$, имеющей максимум при x=4. Найти параметры a, b, c и вычислить вероятность попадания случайной величины X интервал [2; 3].

**6.1.40.** Случайная величина X задана функцией распределения

$$F(x)=\begin{cases} 0 \ \ , \text{при } x \le 0, \\ x^2, \text{при } 0<x \le 1, \\ 1 \ \ , \text{при } x > 1. \end{cases}$$

Найти: а) плотность вероятности φ(x); б) математическое ожидание M(X); в) дисперсию D(X); г) вероятности P(X=0,5), P(X<0,5), P(0,5≤X≤1); д) построить графики φ(x) и F(x) и показать на них математическое ожидание M(X) и вероятности, найденные в пункте г).

## 6.2. Основные законы распределения

**6.2.1.** Вероятность выигрыша по облигации займа за все время его действия равна 0,1. Составить закон распределения числа выигравших облигаций среди приобретенных 19. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение и моду этой случайной величины.

**6.2.2.** По данным примера 6.2.1 найти математическое ожидание, дисперсию и среднее квадратическое отклонение доли (частоты) выигравших облигаций среди приобретенных.

**6.2.3.** Составить функцию распределения случайной величины, имеющей биномиальный закон распределения с параметрами n и p.

**6.2.4.** Устройство состоит из 1000 элементов, работающих независимо один от другого. Вероятность отказа любого элемента в течение времени t равна 0,002. Необходимо: а) составить закон распределения отказавших за время t элементов; б) найти математическое ожидание и дисперсию этой случайной величины; в) определить вероятность того, что за время t откажет хотя бы один элемент.

**6.2.5.** Вероятность поражения цели равна 0,05. Производится стрельба по цели до первого попадания. Необходимо: а) составить закон распределения числа сделанных выстрелов; б) найти математическое ожидание и дисперсию этой случайной величины; в) определить вероятность того, что для поражения цели потребуется не менее 5 выстрелов.

**6.2.6.** В магазине имеются 20 телевизоров, из них 7 имеют дефекты. Необходимо: а) составить закон распределения числа телевизоров с дефектами среди выбранных наудачу пяти; б) найти математическое ожидание и дисперсию этой случайной величины; в) определить вероятность того, что среди выбранных нет телевизоров с дефектами.

**6.2.7.** Цена деления шкалы измерительного прибора равна 0,2. Показания прибора округляют до ближайшего целого числа. Полагая, что при отсчете ошибка округления распределена по равномерному закону, найти: 1) математическое ожидание, дисперсию и среднее квадратическое отклонение этой случайной величины; 2) вероятность того, что ошибка округления: а) меньше 0,04; б) больше 0,05.

**6.2.8.** Среднее время безотказной работы прибора равно 80 ч. Полагая, что время безотказной работы прибора имеет показательный закон распределения, найти: а) выражение его плотности вероятности и функции распределения; б) вероятность того, что в течение 100 ч. прибор не выйдет из строя.

**6.2.9.** Текущая цена акции может быть смоделирована с помощью нормального закона распределения с математическим ожиданием 15 ден. ед. и средним квадратическим отклонением 0,2 ден. ед. Найти вероятность того, что цена акции: а) не выше 15,3 ден. ед.; б) не ниже 15,4 ден. ед.; в) от 14,9 до 15,3 ден. ед.

**6.2.10.** Цена некой ценной бумаги нормально распределена. В течение последнего года 20% рабочих дней она была ниже 88 ден. ед., а 75% — выше 90 ден. ед. Найти: а) математическое ожидание и среднее квадратическое отклонение цены ценной бумаги; б) вероятность того, что в день покупки цена будет заключена в пределах от 83 до 96 ден. ед.; в) с надежностью 0,95 определить максимальное отклонение цены ценной бумаги от среднего (прогнозного) значения (по абсолютной величине).

**6.2.11.** Коробки с конфетами упаковываются автоматически. Их средняя масса равна 540 г. Известно, что масса коробок с конфетами имеет нормальное распределение, а 5% коробок имеют массу, меньшую 500 г. Каков процент коробок, масса которых: а) менее 470 г; б) от 500 до 550 г; в) более 550 г; г) отличается от средней не более, чем на 30 г (по абсолютной величине)?

**6.2.12.** Случайная величина X имеет нормальное распределение с математическим ожиданием а = 25. Вероятность попадания Хв интервал (10; 15) равна 0,09. Чему равна вероятность попадания X в интервал: а) (35;40); б) (30;35)?

**6.2.13.** Нормально распределенная случайная величина имеет следующую функцию распределения: F(x)=0,5+0,5Ф(x-1). Из какого интервала (1; 2) или (2; 6) она примет значение с большей вероятностью?

**6.2.14.** 20%-ная точка нормально распределенной случайной величины равна 50, а 40%-ная точка равна 35. Найти вероятность того, что случайная величина примет значение в интервале (25;45).

**6.2.15.** Известно, что нормально распределенная случайная величина принимает значение: а) меньшее 248 с вероятностью 0,975; б) большее 279 с вероятностью 0,005. Найти функцию распределения случайной величины X.

**6.2.16.** Случайная величина X распределена по нормальному закону с нулевым математическим ожиданием. Вероятность попадания этой случайной величины на отрезок от -1 до +1 равна 0,5. Найти выражения плотности вероятности и функции распределения случайной величины X

**6.2.17.** Имеется случайная величина X, распределенная по нормальному закону с математическим ожиданием a и дисперсией $\sigma^2$. Требуется приближенно заменить нормальный закон распределения равномерным законом в интервале (α; β); границы α, β подобрать так, чтобы сохранить неизменными математическое ожидание и дисперсию случайной величины X.

**6.2.18.** Случайная величина X распределена по нормальному закону с математическим ожиданием a=0. При каком значении среднего квадратического отклонения a вероятность попадания случайной величины X в интервал (1;2) достигает максимума?

**6.2.19.** Время ремонта телевизора распределено по показательному закону с математическим ожиданием, равным 0,5 ч. Некто сдает в ремонт два телевизора, которые одновременно начинают ремонтировать, и ждет, когда будет отремонтирован один из них. После этого с готовым телевизором он уходит. Найти закон распределения времени: а) потраченного клиентом; б) которое должен потратить клиент, если он хочет забрать сразу два телевизора.

## 6.3. Парная регрессия и корреляция

**6.3.1.** Изучается зависимость результативного признака ***y*** от фактора ***x***.

1. Построить линейную модель парной регрессии (рассчитать параметры линейного уравнения парной регрессии).

2. Оценить модель, определив:

--- компоненты дисперсии,

— коэффициент корреляции,

— коэффициент детерминации,

— среднюю ошибку аппроксимации,

— $F$-критерий Фишера.

3. Дать интерпретацию рассчитанных характеристик.

4.Сделать оценку статистической значимости коэф. регрессии и корреляции (рассчитать случайные ошибки параметров линейной регрессии и коэффициента корреляции, рассчитать t- критерий стьюдента и доверительные интервалы каждого из показателей)

5. На графике отобразить исходные данные и линию регрессии.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***x*** | 72 | 52 | 73 | 74 | 76 | 79 | 54 | 68 | 73 | 64 | 55 | 71 | 82 | 90 | 92 |
| ***y*** | 121 | 84 | 119 | 117 | 129 | 128 | 102 | 111 | 112 | 98 | 103 | 115 | 134 | 150 | 152 |

## РЕКОМЕНДОВАННАЯ ЛИТЕРАТУРА


1. Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. —М.: Наука, 1983.
2. Брандт, Зигмунд. Анализ данных. Статистические и вычислительные методы для научных работников и инженеров: Пер. с англ. : Учебное пособие/ З. Брандт ; пер. : О. И. Волкова ; ред. пер. : Е. В. Чепурин. — М.: Мир, 2003. - 686 с. (20)
3. Кендал М., Стьюарт А. Теория распределений. — М.: Наука, 1966.
4. Корн Г., Корн Т. Справочник по математике (для научных работников и инженеров). — М.: Наука, 1974.
5. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений.— М.: Наука,1968.