

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ВОЛГОГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

И. Э. Симонова, В. Д. Савельев
Л. С. Сагателова, А. Б. Симонов

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА:
МАТЕРИАЛЫ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ

Учебное пособие



Волгоград
2013

Рецензенты:

кафедра «Математические методы и информатика в экономике» ВолгГУ,
зав. кафедрой *Л. Ю. Богачкова*;
д-р физ.-мат. наук, профессор МГУ им. М. В. Ломоносова *В. Н. Тутубалин*;
д-р тех. наук, профессор ВГСПУ *Б. А. Жуков*

Печатается по решению редакционно-издательского совета
Волгоградского государственного технического университета

Симонова, И. Э.

Теория вероятностей и математическая статистика: материалы к самостоятельной работе : учеб. пособие / И. Э. Симонова, В. Д. Савельев, Л. С. Сагателова, А. Б. Симонов; ВолгГТУ. – Волгоград, 2013. – 80 с.

ISBN 978–5–9948–1251–8

Пособие содержит базовые сведения по вводному курсу теории вероятностей и математической статистики. Каждая тема сопровождается справочными таблицами. Основное внимание уделено темам «Случайные величины», «Статистический анализ одномерных случайных величин» и «Регрессионный анализ». Приведены варианты заданий к семестровым работам и образцы их выполнения. Показана возможность выполнения заданий по математической статистике на ППП Statgraphics и Excel.

Предназначено студентам направления 080100.62 «Экономика» всех форм обучения. Может быть полезно студентам других экономических, а также технических специальностей.

Ил. 12. Табл. 13. Библиогр.: 8 назв.

ISBN 978–5–9948–1251–8

© Волгоградский государственный
технический университет, 2013

ВВЕДЕНИЕ

Причинные связи явлений в законах природы и общества проявляются в двух основных формах – в виде детерминированных и статистических закономерностей.

В повседневной жизни, в бизнесе, в экономических исследованиях мы постоянно сталкиваемся с событиями и явлениями с неопределенными, неоднозначными исходами.

Теория вероятностей – математическая наука, изучающая закономерности случайных явлений. **Математическая статистика** дает методы анализа статистических данных, т. е. данных, характеризующих отдельные элементы массовых совокупностей случайных событий.

Настоящее пособие предназначено для самостоятельной работы студентов, изучающих курс «Теория вероятностей и математическая статистика», входящий в базовую часть блока «Б.2. Математический и естественно-научный блок» программы обучения бакалавров по направлению «Экономика».

При написании пособия авторы старались добиться максимальной доступности изложения, сохраняя необходимый уровень строгости. Ограниченность объема пособия привела к жесткому отбору материала. Основной упор сделан на разделах «Случайные величины», «Статистический анализ одномерных случайных величин» и «Регрессионный анализ», по которым представлены задания к семестровой работе и примеры их выполнения. Эти темы особенно важны для экономических и технических приложений, а также служат фундаментом прикладных статистических дисциплин – «Эконометрика» (для экономистов) и «Методы обработки результатов экспериментов» (для технических специальностей).

Предполагается, что студенты уже освоили тему «Случайные события». Необходимые справочные материалы по этой теме приведены в табл. П.Б.1. Тем самым не все темы курса отражены в пособии подробно, и оно не может заменить вузовские учебники.

Особое внимание в пособии уделено построению математических моделей случайных явлений и их использованию для описания, анализа и прогнозирования в реальных экономических задачах.

Стремясь приблизить материал к практическим задачам экономики, авторы подобрали соответствующие задания по математической статистике. Использовались данные, публикуемые журналом «Российская экономика: прогнозы и тенденции» [6], данные Госкомстата, а также экспериментальные данные, предоставленные техническими кафедрами ВолгГТУ. Показана возможность использования ППП Statgraphics и Excel при выполнении семестровых заданий.

1. ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ И ИХ ПРИМЕНЕНИЕ К РЕШЕНИЮ ЗАДАЧ

1.1. ОДНОМЕРНАЯ СЛУЧАЙНАЯ ВЕЛИЧИНА, ЕЕ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

В различных вероятностных экспериментах мы часто наблюдаем величины, так или иначе характеризующие исследуемое явление и принимающие значения, случайным образом меняющиеся от испытания к испытанию под воздействием случайных обстоятельств – «случайные величины» (СВ). Строгое определение случайной величины дается на основе понятия вероятностного пространства $(\Omega; F; P)$, где пространство элементарных событий Ω – произвольное множество, элементы которого называются элементарными случайными событиями, F – алгебра подмножеств множества Ω , называемых случайными событиями, F включает в себя множество Ω , все счетные объединения и пересечения и все возможные дополнения до Ω подмножеств из Ω , $P = P(A)$ – вероятностная функция на множествах $A \in F$, т. е. числовая функция, удовлетворяющая аксиомам:

$$A.1. P(A) \geq 0,$$

$$A.2. P(\Omega) = 1,$$

$$A.3. P\left(\sum_i A_i\right) = \sum_i P(A_i) \text{ при } A_i A_j = \emptyset, i \neq j.$$

Если каждому элементарному событию – «точке» из Ω – поставлена в соответствие точка на прямой (на плоскости или в n -мерном пространстве): $\omega \rightarrow X(\omega)$, то говорят об одномерных (двумерных или n -мерных) СВ.

Одномерной случайной величиной называется функция $X(\omega)$, определенная на пространстве Ω и для любого действительного x удовлетворяющая условию $\{\omega: X(\omega) \leq x\} \in F$.

Случайные величины обозначают обычно прописными буквами X, Y , а принимаемые ими значения – соответствующими строчными буквами. Часто при изучении случайных величин удобно вместо вероятностного пространства $(\Omega; F; P)$ использовать вероятностное пространство $(\Omega_x; F_x; P_x)$, порожденное этой случайной величиной. При этом вероятности P_x приписываются непосредственно значениям случайной величины, а не их прообразам в Ω .

События «случайная величина X приняла значение x » можно рассматривать как элементарные, и в качестве пространства Ω_x взять множество всех значений случайной величины X . Тогда σ -алгебра случайных собы-

тий F_x – это система подмножества Ω_x , порожденная множествами $(-\infty; x) \cap \Omega_x \subset R$. Осталось задать распределение вероятностей P_x на множествах из F_x , положив его равным в каждой точке x вероятности того, что случайная величина X примет значение, меньшее x : $F(x) = P(X < x) = P\{\omega: X(\omega) < x\}$. Это наиболее общий способ задать распределение вероятностей P_x . Функция $F(x)$ называется **функцией распределения** СВ X (ф. р.).

Свойства функции распределения приведены в табл. 1.1. Доказательство приведено в [1]. Справочные сведения по теме «Случайные события» представлены в табл. П.Б.1.

Иногда распределение вероятностей на F_x удобно задавать не с помощью функции распределения $F(x)$, а как-то иначе. Любое правило, устанавливающее связь между возможными значениями случайной величины X (точнее, множествами $A \in F_x$ и соответствующими им вероятностями $P(A)$), называется законом распределения или просто распределением случайной величины X .

1.2. ДИСКРЕТНЫЕ И НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Случайная величина X называется **дискретной**, если она принимает конечное или счетное множество значений $\Omega_x = \{x_1; x_2; x_3; \dots\}$. Для задания дискретной случайной величины необходимо знать не только ее значения, но и то, как часто, точнее, с какой вероятностью она принимает эти значения, т. е. знать $p_n = P(X = x_n)$. Поэтому дискретная случайная величина X задается в виде ряда распределения – таблицы вида

Значения x_i	x_1	x_2	...	x_n	...
Вероятности p_i	p_1	p_2	...	p_n	...

где $p_n = P\{X = x_n\}$, $\sum_{n=1}^{\infty} p_n = 1$. Это наиболее простая форма задания закона распределения. По вероятностям p_n функция распределения $F(x)$ легко находится из равенства $F(x) = P(X < x) = \sum_{i: x_i < x} P(X = x_i) = \sum_{i: x_i < x} p_i$. (1.1)

Основные свойства функций распределения представлены в табл. 1.1.

Случайные величины

Дискретная						Непрерывная	
1) Ряд распределения СВ X							
x_i	X_1	X_2	...	x_n	...		
p_i	p_1	p_2	...	p_n	...		
где $p_n=P\{X = x_n\}$, $\sum_{n=1}^{\infty} p_n = 1$.							
2) Функция распределения $F(x) = P(X < x)$							
$F(x) = \sum_{i: x_i < x} p_i$.						$F(x) = \int_{-\infty}^x f(y)dy$, где $f(y) \geq 0$ – плотность распределения.	
3) Свойства функции распределения: а) $0 \leq F(x) \leq 1$; б) $F(x)$ – неубывающая; в) $F(x)$ непрерывна слева; г) $F(-\infty) = 0$, $F(+\infty) = 1$.							
						3) свойства плотности распределения: а) $f(x) \geq 0$, б) $\int_{-\infty}^{\infty} f(x)dx = 1$; в) $f(x) = F'(x)$ в точках непре- рывности $f(x)$.	
5) Вероятность попадания в полуинтервал $[x_1; x_2)$							
$P(x_1 \leq X < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx$							
6) Числовые характеристики Математическое ожидание (среднее значение)							
$M(X) = \sum_{i=1}^n x_i p_i$.						$MX = \int_{-\infty}^{+\infty} xf(x)dx$	
Дисперсия $DX = M(X - MX)^2 = MX^2 - (MX)^2$							
$DX = \sum_i x_i^2 p_i - (MX)^2 = \sum_i (x_i - MX)^2 p_i$.						$DX = \int_{-\infty}^{\infty} x^2 f(x)dx - (MX)^2$	
Среднеквадратическое отклонение (мера рассеяния около среднего) $\sigma(X) = \sqrt{DX} \Leftrightarrow DX = \sigma^2$. Момент порядка ν: $m_\nu = MX^\nu$							

Пример 1.1. Три элемента вычислительного устройства работают независимо. Вероятность отказа первого элемента за время t равна $0,1$, второго – $0,2$, третьего – $0,3$. Определить вероятность того, что за время t откажут: 1) все три элемента; 2) хотя бы один элемент. Найти ряд распределения и функцию распределения случайной величины X , равной числу отказавших элементов, построить график функции распределения.

Решение. Случайная величина X может принимать только значения, равные $0; 1; 2; 3$, т.е. $\Omega_x = \{0; 1; 2; 3\}$. Чтобы вычислить вероятности p_i , с которыми случайная величина X принимает эти значения, составим пространство элементарных исходов. Пусть A_i, \bar{A}_i – события, состоящие в том, что i -й элемент проработает исправно или, соответственно, откажет в течение времени t . Тогда возможны следующие исходы:

$\Omega = \{\omega_1 = A_1 A_2 A_3; \omega_2 = \bar{A}_1 A_2 A_3; \omega_3 = A_1 \bar{A}_2 A_3; \omega_4 = A_1 A_2 \bar{A}_3; \omega_5 = \bar{A}_1 \bar{A}_2 A_3; \omega_6 = \bar{A}_1 A_2 \bar{A}_3; \omega_7 = A_1 \bar{A}_2 \bar{A}_3; \omega_8 = \bar{A}_1 \bar{A}_2 \bar{A}_3\}$. Т.к. отказы элементов взаимно независимы, то по теореме умножения вероятностей независимых событий (см. табл. П.3) находим:

$$P(\omega_1) = P(A_1 A_2 A_3) = P(A_1)P(A_2)P(A_3) = 0,9 \cdot 0,8 \cdot 0,7 = 0,504;$$

$$P(\omega_2) = P(\bar{A}_1 A_2 A_3) = P(\bar{A}_1)P(A_2)P(A_3) = 0,1 \cdot 0,8 \cdot 0,7 = 0,056;$$

$$P(\omega_3) = P(A_1 \bar{A}_2 A_3) = P(A_1)P(\bar{A}_2)P(A_3) = 0,9 \cdot 0,2 \cdot 0,7 = 0,126;$$

$$P(\omega_4) = P(A_1 A_2 \bar{A}_3) = P(A_1)P(A_2)P(\bar{A}_3) = 0,9 \cdot 0,8 \cdot 0,3 = 0,216;$$

$$P(\omega_5) = P(\bar{A}_1 \bar{A}_2 A_3) = P(\bar{A}_1)P(\bar{A}_2)P(A_3) = 0,1 \cdot 0,2 \cdot 0,7 = 0,014;$$

$$P(\omega_6) = P(\bar{A}_1 A_2 \bar{A}_3) = P(\bar{A}_1)P(A_2)P(\bar{A}_3) = 0,1 \cdot 0,9 \cdot 0,3 = 0,024;$$

$$P(\omega_7) = P(A_1 \bar{A}_2 \bar{A}_3) = P(A_1)P(\bar{A}_2)P(\bar{A}_3) = 0,9 \cdot 0,2 \cdot 0,3 = 0,054;$$

$$P(\omega_8) = P(\bar{A}_1 \bar{A}_2 \bar{A}_3) = P(\bar{A}_1)P(\bar{A}_2)P(\bar{A}_3) = 0,1 \cdot 0,2 \cdot 0,3 = 0,006.$$

Зная вероятности элементарных исходов $\omega_i, i = 1, 2, \dots, 8$, находим вероятности событий ($X = x_i$):

$$P(X = 0) = P\{\omega: X(\omega) = 0\} = P(\omega_1) = 0,504,$$

$$P(X = 1) = P\{\omega: X(\omega) = 1\} = P(\{\omega_2; \omega_3; \omega_4\}) = 0,398,$$

$$P(X = 2) = P\{\omega: X(\omega) = 2\} = P(\{\omega_5; \omega_6; \omega_7\}) = 0,092;$$

$$P(X = 3) = P\{\omega: X(\omega) = 3\} = P(\omega_8) = 0,006.$$

Ряд распределения случайной величины X имеет вид:

x_i	0	1	2	3
p_i	0,504	0,398	0,092	0,006

Контроль: $\sum_{i=1}^4 P\{X = x_i\} = 1$. Ф. р. $F(x)$ определяем по формуле (1.1).

Если $x \leq 0$, то $F(x) = 0$. Если $0 < x \leq 1$, то $F(x) = 0,504$.

Если $1 < x \leq 2$, то $F(x) = 0,902$.

Если $2 < x \leq 3$, то $F(x) = 0,994$. Если $x > 3$, то $F(x) = 1$.

График функции $F(x)$ – ступенчатая функция, постоянная на промежутках $(-\infty; x_1]$, $(x_i, x_{i+1}]$, $(x_4; +\infty)$, $i = 1, 2, 3$. Зная ряд распределения и функцию распределения, легко находим вероятность события A , состоящего в том, что за время t откажут все три элемента, и вероятность события B , состоящего в том, что откажет хотя бы один элемент. Действительно,

$$P(A) = P(X = 3) = 0,006; P(B) = P(X \geq 1) = 1 - P(X < 1) = 1 - F(1) = 0,496.$$

Вероятность того, что случайная величина X примет значение из промежутка $[\alpha; \beta)$, равна разности значений ее функции распределения на концах этого полуинтервала: $P(\alpha \leq X < \beta) = F(\beta) - F(\alpha)$.

Пример 1.2. Производится 3 независимых выстрела по мишени, и вероятность попадания при каждом выстреле равна 0,1. Требуется построить ряд распределения СВ X – числа попаданий в мишень после трех выстрелов.

Решение. Очевидно, что множество возможных значений рассматриваемой СВ X состоит из четырех элементов $\Omega_x = \{0; 1; 2; 3\}$. Условия примера сводятся к системе повторения независимых испытаний в одинаковых условиях, поэтому распределение вероятностей возможных значений $P(X = m)$ находится по формуле Бернулли, а закон распределения

вероятностей называется биномиальным законом распределения случайной величины X (табл. П.Б.1):

$$P(X = k) = C_n^k p^k q^{n-k} \quad (1.2)$$

В нашем примере $n = 3$. Последовательные вычисления по формуле (1.2) для $k = 0; 1; 2; 3$ позволяют получить следующий ряд распределения:

X_i	0	1	2	3
p_i	0,729	0,243	0,027	0,001

Контроль: $\sum_{i=1}^4 P\{X = x_i\} = 1.$

При большом числе испытаний по схеме Бернулли вычисления по формуле (1.2) становятся громоздкими, так как факториалы, входящие в формулу, становятся очень большими числами. В этом случае для вычисления вероятностей $P(X = k)$ следует использовать подходящие *асимптотические формулы*.

Если число испытаний велико ($n \geq 100$), а вероятность p появления события в каждом испытании мала ($n \cdot p < 0,1$), то используют предельную формулу Пуассона: $P_n(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}$, $\lambda = n \cdot p$, $k = 0, 1, 2, \dots$

Закон вероятностей, задаваемый последней формулой, называется *законом Пуассона*. Параметр λ равен среднему числу появления события в n испытаниях.

В случае, когда $n \geq 100$, $p, q > 0,2$ (оба параметра p и q в формуле заметно отличны от нуля), применяются локальная (1.3) и интегральная (1.4) формулы Муавра–Лапласа:

$$P_n(X = k) \approx \frac{1}{\sqrt{npq}} \varphi(x) \quad (1.3)$$

где аргумент x равен $x = \frac{k - np}{\sqrt{npq}}$, $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – табличная плотность стандартного нормального закона (четная функция: $\varphi(-x) = \varphi(x)$);

$$P_n(k_1 \leq X \leq k_2) = \Phi(x_2) - \Phi(x_1), \quad (1.4)$$

где $x_1 = \frac{k_1 - np}{\sqrt{npq}}$; $x_2 = \frac{k_2 - np}{\sqrt{npq}}$; $\Phi_1(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$; $\Phi_1(-x) = -\Phi_1(x)$.

$\Phi_1(x)$ – функция Лапласа (её значения см. в табл. П.А.1).

С помощью приведенных выше формул при $n \geq 100$ и при $npq \geq 10$ обычно удается получить искомые вероятности с точностью до трех-четырех знаков после запятой.

Подробнее нормальный закон распределения рассмотрен ниже в п. 1.4.

З а м е ч а н и е 1. В качестве условия применимости предельной формулы Пуассона при $n \geq 100$, следуя Е. С. Вентцель и Л. А. Овчарову, мы берем условие $np < 0,1$, [Вентцель, Е. С. Теория вероятностей и ее технические приложения, 2-е изд., стер. / Е. С. Вентцель, Л. А. Овчаров. – М.: Высшая школа, 2000. – 480 с. – § 10.2]. В этом случае $p^2 < 0,1 \cdot n^{-1} < 10^{-3}$, $np^2 < 10^{-4}$, т. е. выполнено условие $np^2 \ll 1$ (или $np^2 \rightarrow 0$), даваемое в некоторых других источниках.

З а м е ч а н и е 2. Необходимое условие применимости предельных формул Муавра–Лапласа в ряде источников дано в виде $npq \geq 10$. Нетрудно проверить, что при $n \geq 100$, $p, q > 0,2$ неравенство $npq \geq 10$ выполнено, так как наименьшее значение функции $g(p) = n \cdot p \cdot q = n \cdot p(1 - p)$ на отрезке $[0,2; 0,8]$ равно $100 \cdot 0,2 \cdot 0,8 = 16$.

Пример 1.3. При массовом производстве однотипных изделий вероятность брака при штамповке равна 0,02. Какова вероятность того, что из 100 наудачу взятых изделий 5 будут бракованными?

Р е ш е н и е. Здесь СВ X – число бракованных изделий среди 100 изделий – имеет биномиальное распределение, и вероятность, определяемая по формуле (1.2), составляет $P(X = 5) = C_{100}^5 0,02^5 \cdot 0,98^{95} = 0,0353$.

Непосредственный подсчет по этой формуле связан с техническими трудностями. Значительно проще задача решается приближенно по формуле Пуассона: $\lambda = n \cdot p = 100 \cdot 0,02 = 2$, $P(X = 5) \approx \frac{2^5 e^{-2}}{5!} = 0,0361$.

Пример 1.4. Вероятность появления события за период испытания равна 0,25. Найти вероятность того, что при проведении ста независимых испытаний событие появится а) ровно 20 раз; б) не менее 15 раз и не более 35 раз.

Решение. а) По условию $n = 100$; $k = 20$; $p = 0,25$; $q = 0,75$. Так как $n = 100$ – достаточно большое число, а p и q заметно отличны от 0 ($npq \geq 10$), то воспользуемся локальной формулой Муавра–Лапласа (1.4) Найдем значение аргумента функции:

$$x = \frac{k - np}{\sqrt{npq}}. \quad x = \frac{20 - 100 \cdot 0,25}{\sqrt{100 \cdot 0,25 \cdot 0,75}} = -1,15.$$

По таблице плотности нормального закона находим $\varphi(-1,15) = \varphi(1,15) = 0,206$.

$$\text{По формуле (1.3) находим } P(X = 20) = \frac{1}{\sqrt{100 \cdot 0,25 \cdot 0,75}} \cdot 0,206 = 0,05.$$

Для решения второго вопроса воспользуемся интегральной формулой Муавра–Лапласа (1.4). Подставим в эту формулу данные примера.

Определяя по табл. П.А.1 значение функции Лапласа $\Phi_1(0,15) = 0,4922$, получим: $P(15 \leq X \leq 35) = 2 \cdot 0,4922 = 0,9844$.

(Подробнее использование статистических таблиц обсуждается в п. 1.4.)

Второй наиболее важный класс случайных величин – класс **непрерывных** СВ – таких случайных величин, для которых их функция распределения в каждой точке x может быть представлена в виде:

$$F(X) = \int_{-\infty}^x f(y)dy, \quad (1.5)$$

где $f(y)$ – некоторая неотрицательная интегрируемая функция, называемая плотностью распределения вероятностей. Из определения (1.5) следует непрерывность ф. р. $F(x)$ для непрерывной СВ. Для построения математической модели эксперимента с непрерывной СВ можно задать на множестве возможных значений Ω_x либо ф. р. $F(X)$, либо плотность распределения $f(x)$. Вероятность попадания СВ X в полуинтервал $[x_1; x_2)$ можно вычислить по формуле:

$$P\{x_1 \leq X < x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx. \quad (1.6)$$

Из формулы (1.5), (1.6) и свойств плотности распределения, приведенных в табл. 1.1, следует, что множество возможных значений Ω_x непрерывной СВ несчетно, а вероятность каждого отдельного значения равна нулю. Поэтому по формуле (1.6) также вычисляются вероятности $P(x_1 < X < x_2)$, $P(x_1 < X \leq x_2)$, $P(x_1 \leq X \leq x_2)$.

Геометрически вероятность попадания случайной величины в интервал $(a; b)$ равна площади криволинейной трапеции, ограниченной плотностью распределения, осью Ox и прямыми $x = a$, $x = b$.

Пример 1.5. Непрерывная СВ X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x \leq 1, \\ A \ln x & \text{при } 1 < x < e, \\ 1 & \text{при } x \geq e. \end{cases}$$

Найти: 1) значение коэффициента A , 2) плотность распределения $f(x)$, 3) вероятность попадания случайной величины X в интервал $(1; 2)$.

Решение. 1) Коэффициент A находим, используя непрерывность функции $F(x)$: $1 = F(e) = \lim_{x \rightarrow e} F(x) = A \cdot \ln e = A$, откуда $A = 1$.

2) Чтобы найти плотность распределения $f(x)$, продифференцируем ф. р. $F(x)$ по x (см. третье свойство плотности распределения в табл. 1.1):

$$f(x) = F'(x) = \begin{cases} 0 & \text{при } x \leq 1, \\ \frac{1}{x} & \text{при } 1 < x < e, \\ 0 & \text{при } x \geq e. \end{cases}$$

3) Вероятность того, что случайная величина X примет значение из интервала $(1; 2)$, равна определенному интегралу от плотности распределения в пределах от 1 до 2: $P(1 < X < 2) = \int_1^2 \frac{1}{x} dx = \ln x \Big|_1^2 = \ln 2$.

1.3. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

Наиболее важными числовыми характеристиками случайных величин являются математическое ожидание (м. о.), дисперсия, среднеквадратическое отклонение и моменты различных порядков.

Математическое ожидание MX дискретной СВ X , принимающей значения x_i с вероятностями p_i , $i = 1, 2, \dots$, определяется как сумма ряда

$MX = \sum_{i=1}^{\infty} x_i p_i$ при условии, что ряд этот абсолютно сходится (в противном случае м.о. не существует). Если СВ X принимает лишь конечное число значений x_1, \dots, x_n , то

$$MX = \sum_{i=1}^n x_i p_i. \quad (1.7)$$

Математическое ожидание указывает среднее значение СВ X , около которого происходит разброс наблюдаемых значений СВ X .

Для непрерывной СВ суммирование в (1.7) заменяется интегрированием, и м.о. определяется как интеграл (если он абсолютно сходится):

$$MX = \int_{-\infty}^{\infty} xf(x)dx. \quad (1.8)$$

Дисперсией случайной величины X называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$DX = M(X - MX)^2 \quad (1.9)$$

Дисперсия характеризует рассеяние значений случайной величины около ее математического ожидания.

Пример 1.6. Найти математическое ожидание и дисперсию случайной величины X , определенной в примере 1.2.

Решение. Зная ряд распределения, MX и DX находим по формулам (1.7) и (1.9):

$$MX = 0 + 1 \cdot 0,398 + 2 \cdot 0,092 + 3 \cdot 0,006 = 0,6;$$

$$DX = 1 \cdot 0,398 + 2^2 \cdot 0,092 + 3^2 \cdot 0,006 - 0,6^2 = 0,46.$$

Пример 1.7. Физиками экспериментально доказано, что для радиоактивных веществ время жизни атома X - случайная величина, распределенная по показательному закону: $P(X < t) = 1 - e^{-\omega t}$ ($t \geq 0$), где ω - постоянная, называемая «вероятностью распада» и характеризующая данный тип атомов. Найти: 1) среднее время жизни атома t и разброс около среднего;

2) вероятность того, что атом распадется за время, меньшее t ; за время от t до $2t$; 3) доказать, что радиоактивные атомы «не стареют»: если известно, что атом не распался до момента τ , т. е. $X > \tau$, то вероятность прожить еще время, не меньшее t , определяется по тому же закону:

$$P(X > t + \tau | X > \tau) = e^{-\omega t} = P(X > t).$$

З а м е ч а н и е. Время ожидания в системах массового обслуживания обычно также распределено по показательному закону.

Решение. 1) Для нахождения числовых характеристик СВ сначала надо найти плотность распределения $f(x)$ как производную от ф. р. (при $t \geq 0$): $F(t) = P(X < t) = 1 - P(X \geq t) = 1 - e^{-\omega t} \Rightarrow f(t) = F'(t) = \omega e^{-\omega t}$.

Так как СВ X принимает только положительные значения, то при $t < 0$ $F(t) = 0$, $f(t) = 0$, и в формулах для MX и DX интегрирование фактически проводится по интервалу $(0; \infty)$:

$$\bar{t} = MX = \int_{-\infty}^{\infty} tf(t)dt = \int_0^{\infty} t\omega e^{-\omega t} dt = \frac{1}{\omega}, \quad (1.10)$$

$DX = \int_0^{\infty} t^2 \omega e^{-\omega t} dt - (MX)^2 = \frac{1}{\omega^2}$ (интегралы, входящие в формулы для MX и DX , студентам предоставляется возможность вычислить самостоятельно).

2) Вероятность того, что случайная величина X примет значение из полуинтервала $[a; b)$, есть $F(b) - F(a)$. Поэтому $P(X < \bar{t}) = 1 - e^{-\omega \bar{t}}$, $P(\bar{t} < X < 2\bar{t}) = F(2\bar{t}) - F(\bar{t})$. Заменяя в соответствии с (1.10) в этих формулах ω на $\frac{1}{\bar{t}}$, $\omega \cdot \bar{t}$ на 1, получаем: $P(X < \bar{t}) = 1 - e^{-1} = 1 - \frac{1}{e} \approx 0,63$,

$$P(\bar{t} < X < 2\bar{t}) = 1 - e^{-\frac{2}{\bar{t}} \bar{t}} - (1 - e^{-1}) \approx 0,23.$$

В соответствии с законом больших чисел это означает, что за время, меньшее \bar{t} , распадается около 63 % атомов, имевшихся в начальный момент, за время от \bar{t} до $2\bar{t}$ – 23 % атомов, за время, большее $2\bar{t}$ – 14 % атомов.

3) По определению условной вероятности $P(A|B) = \frac{P(AB)}{P(B)}$. Поэтому

$$P(X > t + \tau | X > \tau) = \frac{P(X > t + \tau | X > \tau)}{P(X > \tau)} = \frac{P(X > t + \tau)}{P(X > \tau)} = \frac{e^{-\omega(t+\tau)}}{e^{-\omega\tau}} = e^{-\omega t}.$$

Пример 1.8. СВ X имеет плотность распределения $f(x) = A - |x|$ при $|x| \leq 1$; 0 при $|x| > 1$, $f(x) = 0$. Требуется:

1) найти постоянный параметр A и ф.р. $F(x)$;

2) вычислить $P(X > 0,5)$; 3) найти MX и DX .

Решение. Представим $f(x)$ в следующем виде:

$$f(x) = \begin{cases} 0 & \text{при } x < -1, x > 1, \\ A + x & \text{при } -1 \leq x \leq 0, \\ A - x & \text{при } 0 \leq x \leq 1. \end{cases}$$

Для нахождения параметра A воспользуемся третьим свойством плотности распределения, согласно которому $\int_{-\infty}^{\infty} f(x) dx = \int_{-1}^1 (A - |x|) dx = 1$.

Отсюда находим $A = 1$. Значение параметра подставляем в формулу плотности распределения и находим функцию распределения по ее определению:

$$F(x) = \int_{-\infty}^x f(x) dx = \begin{cases} \int_{-\infty}^x 0 dt = 0 & \text{при } x < -1, \\ \int_{-\infty}^{-1} 0 dx + \int_{-1}^x (1+t) dt = \frac{1}{2}x^2 + x + \frac{1}{2} & \text{при } -1 \leq x \leq 0, \\ \int_{-\infty}^{-1} 0 dt + \int_{-1}^0 (1+t) dt + \int_0^x (1-t) dt = -\frac{1}{2}x^2 + x + \frac{1}{2} & \text{при } 0 \leq x \leq 1, \\ \int_{-\infty}^{-1} 2t + \int_{-1}^0 (1+t) dt + \int_0^1 (1-t) dt + \int_1^x 0 dt = 1 & \text{при } x > 1. \end{cases}$$

Вероятность $P(X > 0,5)$ можно вычислить по четвертому свойству плотности распределения:

$$P\left(X > \frac{1}{2}\right) = P\left(\frac{1}{2} < X < \infty\right) = \int_{\frac{1}{2}}^{\infty} f(x)dx = \int_{\frac{1}{2}}^1 f(x)dx = \int_{\frac{1}{2}}^1 (1-t)dt + \int_1^{\infty} 0dt = \frac{1}{8}.$$

Или, еще проще, через функцию распределения:

$$P\left(X > \frac{1}{2}\right) = P\left(\frac{1}{2} < X < 1\right) = F(1) - F(0.5) = 1 - \frac{7}{8} = \frac{1}{8}.$$

Математическое ожидание MX и дисперсия DX непрерывной СВ X определяются при известной плотности распределения по формулам из табл. 1.1:

$$MX = \int_{-\infty}^{\infty} xf(x)dx; \quad DX = \int_{-\infty}^{\infty} (x - MX)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - MX^2.$$

В том случае, когда возможные значения случайной величины X ограничены конечным отрезком $[a, b]$, пределами интегрирования в формулах для математического ожидания и дисперсии служат границы этого отрезка. Так как в нашем примере $f(x)$ и $x^2 f(x)$ – четные функции, а $x \cdot f(x)$ – нечетная, то по свойству интегралов от нечетной и от четной функций в симметричных пределах имеем:

$$MX = \int_{-1}^1 xf(x)dx = 0,$$

$$DX = \int_{-1}^1 x^2 f(x)dx - 0 = 2 \int_0^1 x^2 (1-x)dx = 2 \left(\frac{1}{3}x^3 - \frac{1}{4}x^4 \right) \Big|_0^1 = \frac{1}{6}.$$

1.4. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И РАСПРЕДЕЛЕНИЯ, СВЯЗАННЫЕ С НИМ (РАСПРЕДЕЛЕНИЯ χ^2 , СТЬЮДЕНТА, ФИШЕРА)

СВ X называется нормально распределенной СВ с параметрами a и σ , если ее плотность распределения равна $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-a)^2}{2\sigma^2}}$. Постоянные

a и σ ($\sigma > 0$) называются параметрами нормального распределения. Известно, что $MX=a$, $DX=\sigma^2$; $\sigma=\sqrt{DX}$. Для стандартного нормального закона $MX=0$, $DX=1$, а функция распределения имеет вид

$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$. Эта функция не является элементарной, и для ее вычисления используют специальные таблицы (в табл. П.1 приведены значения функции Лапласа $\Phi_1(x)$, связанные с $\Phi(x)$ формулой $\Phi(x) = \frac{1}{2} + \Phi_1(x)$).

Если СВ X имеет нормальное распределение с параметрами a и σ , то СВ $\xi = \frac{X - a}{\sigma}$ – нормальное распределение с параметрами $(0; 1)$, поэтому функция распределения СВ X $F_X(x)$ выражается через функцию распределения случайной величины ξ по формуле:

$$F_X = P(X < x) = P\left(\frac{X - a}{\sigma} < \frac{x - a}{\sigma}\right) = F_\xi\left(\frac{x - a}{\sigma}\right).$$

Рассмотрим три распределения, широко применяемые в статистике.

1. Распределение χ^2 -Пирсона. Пусть n случайных величин имеют нормальное распределение с параметрами $(0; 1)$: $X_1, \dots, X_n \sim N(0; 1)$, и независимы. Тогда $\chi_n^2 = X_1^2 + X_2^2 + \dots + X_n^2$ – случайная величина χ^2 с n степенями свободы.

Графики плотности распределения хи-квадрат в зависимости от числа степеней свободы приведены на рис. 1.1.

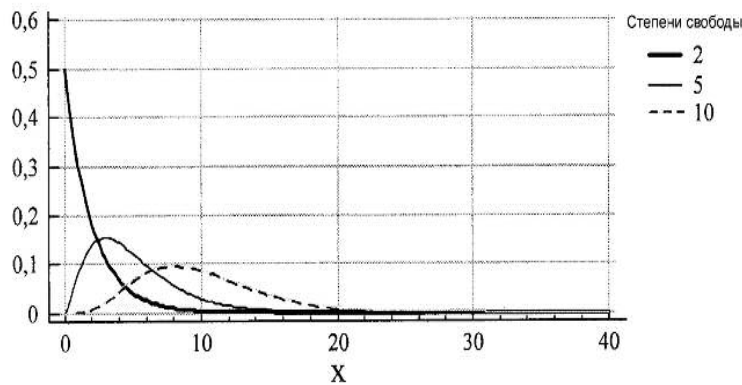


Рис. 1.1. Графики плотности распределения χ^2 -Пирсона в зависимости от числа степеней свободы

2. Распределение Стьюдента (с n степенями свободы)

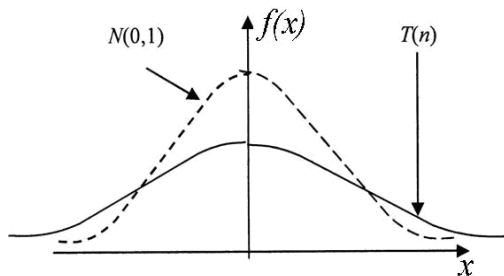


Рис. 1.2. Графики плотности распределения СВ-н $N(0; 1)$ и $T(n)$

При больших n ($n \approx 100$) распределение $T(n)$ близко к нормальному $N(0; 1)$.

Пусть Y, X_1, \dots, X_n – независимые, нормально распределенные СВ с параметрами $0, 1$.

Тогда
$$T(n) = \frac{Y}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}}.$$

Графики плотности распределения СВ-н $N(0; 1)$ и $T(n)$ приведены на рис. 1.2.

3. Распределение Фишера. Рассмотрим две независимые случайные величины, имеющие распределение χ_k^2 и χ_m^2 с k и m степенями свободы соответственно. Поделив каждую из них на их степени свободы, получим «фишерово отношение» с k и m степенями свободы: $F(k; m) = \frac{\chi_k^2 / k}{\chi_m^2 / m}$.

Использование статистических таблиц

Статистические таблицы составлены для стандартных распределений вероятностей, широко применяемых в статистике. С их помощью можно решать две задачи:

- 1) по заданной вероятности p находить квантиль x_p уровня p ;
- 2) для заданного числа x находить значение функции распределения $F(x)$, а также вероятность попадания в любой полуинтервал $[a, b)$ по формуле $P(a \leq X < b) = F(b) - F(a)$.

Покажем, как это делается для нормальной СВ X . В табл. П.А.1 приведены для $x > 0$ значения функции Лапласа $\Phi_1(x) = \int_0^x \varphi(t) dt$, где $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ – плотность стандартного нормального закона с параметрами $0, 1$ (рис. 1.3).

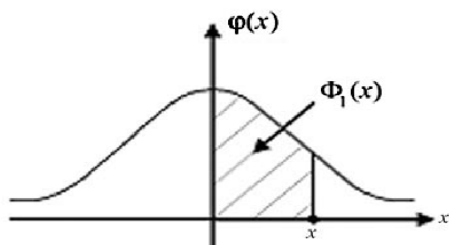


Рис. 1.3. Функции $\varphi(x)$ и $\Phi_1(x)$.

Функция распределения СВ X равна $\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{2} + \Phi_1(x)$,
то есть $\Phi(x) = 0,5 + \Phi_1(x)$.
Для $x < 0$ используем нечетность $\Phi_1(x)$: $\Phi_1(-x) = -\Phi_1(x)$.

Для произвольного интервала $(a; b)$

$$P(a < X < b) = P(a \leq X < b) = \Phi(b) - \Phi(a) = \Phi_1(b) - \Phi_1(a).$$

Пример 1.9. Найти вероятность попадания стандартной нормальной СВ в интервал $(-k, k)$ для $k = 1; 2; 3$.

Решение. Так как $\Phi_1(1) = 0,3413$, то $\Phi_1(-1) = -\Phi_1(1)$ и $P(-1 \leq x < 1) = 2\Phi_1(1) = 0,6826$.

Аналогично, $\Phi_1(2) = 0,47725$, $\Phi_1(3) = 0,49865$.

Поэтому $X \in (-2, 2)$ с вероятностью $2\Phi_1(2) = 0,9545$,

$X \in (-3, 3)$ с вероятностью $2\Phi_1(3) = 0,9973$.

Пусть теперь Y – нормальная СВ с параметрами m и σ^2 . Тогда нормированная случайная величина $X = (Y - m)/\sigma$ – нормальная СВ с параметрами $0, 1$. Найдем вероятность того, что $|Y - m| < 3\sigma$:

$$P(|Y - m| < 3\sigma) = P\left(\left|\frac{Y - m}{\sigma}\right| < 3\right) = P(|X| < 3) = 0,997.$$

Эта формула выражает **правило трех сигм** (3σ) для нормального закона: нормальная случайная величина отклоняется от своего среднего значения на величину, не превышающую 3σ , с вероятностью $0,997$.

Правило трех сигм позволяет оценивать изменчивость ряда экономических показателей, прогнозировать доходы страховых кампаний и т. п.

Пример 1.10. Найти симметричные интервалы, в которые значения нормальной СВ $X \sim N(0, 1)$ попадают с вероятностями $p_1 = 0,95$ и $p_2 = 0,99$.

Решение. Так как для $p = 0,95$, $p/2 = 0,475 = \Phi_1(x)$, то по табл. П.А.1 определяем $x = 1,96$, т. е. $X \in (-1,96; 1,96)$ с вероятностью $0,95$.

Аналогично, для $p = 0,99$ получаем:

$p/2 = 0,495 = \Phi_1(x)$, $x = 2,58$, т.е. $X \in (-2,58; 2,58)$ с вероятностью $0,99$.

1.5. ГЕНЕРАЛЬНАЯ И ВЫБОРОЧНАЯ СОВОКУПНОСТЬ. ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ

До возникновения теории вероятностей и математической статистики объектом исследования науки были явления или опыты, в которых результат полностью определяется влияющими на него факторами. Явления, для которых это выполняется, называются *детерминированными* или *закономерными*, а те, для которых не выполняется – *стохастическими*, *случайными*.

Объектом изучения теории вероятности и математической статистики являются случайные события и явления – те, которые при осуществлении заданной совокупности условий S могут либо произойти, либо не произойти. При этом изучается не отдельное событие, а те закономерности, которые проявляются в результате многократного проведения вероятностного эксперимента.

Во многих явлениях, в том числе экономических, присутствуют оба вида изменчивости – закономерная и случайная, и для нахождения закономерностей нам приходится отсеивать мешающие случайные факторы. Такие задачи важны для прикладных экономических исследований и рассмотрены ниже, в п.п. 1.6, 1.7, 1.9, 2.2.

Исходным понятием при исследовании статистических данных является план получения данных Π . Множество всех возможных при таком плане данных X_Π называется **генеральной совокупностью**. **Выборочной совокупностью** (выборкой) называется совокупность данных, которая наблюдалась в результате проведения эксперимента по плану Π . В статистике в разных ситуациях рассматриваются различные планы наблюдений.

Ниже предполагается, что наблюдения x_1, x_2, \dots, x_n образуют независимую выборку объема n из генеральной совокупности с функцией распределения $F(x)$, то есть значения x_1, x_2, \dots, x_n – реализации независимых одинаково распределенных случайных величин X_1, X_2, \dots, X_n с функцией распределения $F(x)$: $x_1 = X_1(\omega), x_2 = X_2(\omega), \dots, x_n = X_n(\omega)$.

Эти предположения достаточно хорошо согласуются с практикой, например, при исследовании выборки из генеральной совокупности большого объема или при наблюдении значений случайной величины, когда испытания проводятся в независимых условиях. Функция распределения $F(x)$ может зависеть от некоторого неизвестного параметра θ , возможно, многомерного: $F(x) = F(x; \theta), \theta = (\theta_1, \dots, \theta_k)$.

Возникает задача оценивания параметра θ по наблюдениям x_1, x_2, \dots, x_n .

Статистической оценкой параметра $\tilde{\theta}_n$ (статистикой) называется функция от результатов наблюдений: $\tilde{\theta}_n = \tilde{\theta}(x_1, \dots, x_n)$.

Так как наблюдения случайны, то оценка $\tilde{\theta}_n$ – случайная величина.

Перечислим основные свойства, характеризующие *качество оценки*.

1. Оценка $\tilde{\theta}_n$ называется **несмещенной оценкой** θ , если ее среднее значение равно оцениваемому параметру (отсутствует систематическое смещение): $M\tilde{\theta}_n = \theta$.

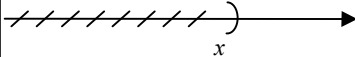
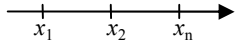
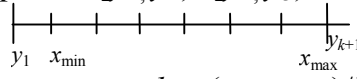
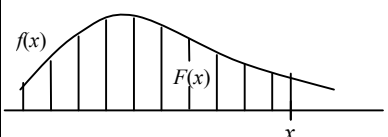
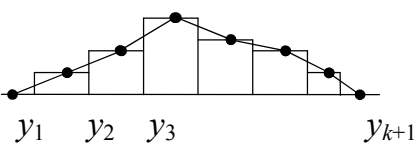
2. Несмещенная оценка $\tilde{\theta}_n$ называется **эффективной**, если она имеет наименьший возможный разброс (дисперсию) при данном n : $D\tilde{\theta}_n \rightarrow \min$.

3. Оценка $\tilde{\theta}_n$ называется **состоятельной оценкой** параметра θ , если с ростом числа наблюдений эта оценка приближается к оцениваемому параметру по вероятности: $\tilde{\theta}_n \xrightarrow{(p)} \theta$ при $n \rightarrow \infty$, т. е. $\lim_{n \rightarrow \infty} P\{|\tilde{\theta}_n - \theta| < \varepsilon\} = 1$ для любого $\varepsilon > 0$.

В табл. 1.2 приведены основные характеристики одномерных случайных величин, а также их оценки по наблюдениям (выборке) x_1, \dots, x_n .

Таблица 1.2

Статистический анализ одномерной случайной величины

Характеристики случайных величин	Их оценка по выборке x_1, x_2, \dots, x_n																											
<div>1. Функция распределения (ф. р.) $F(x) = P(X < x) = P(\omega: X(\omega) < x)$Свойства: 1. неубывающая; 2. $F(-\infty) = 0, F(\infty) = 1$; 3. непрерывна слева.</div>	<div>1. Эмпирическая ф. р. («накопленные частоты») $F_n(x) = \sum_{i: x_i < x} \frac{n_i}{n}$ n_i – частота наблюдения x_i в выборке, n_i/n – относительная частота (доля). Группировка выборки в k интервалов группировки $[y_1, y_2), [y_2, y_3), \dots [y_k, y_{k+1}]$  шаг группировки $h = (y_{k+1} - y_1)/k$</div>																											
<div>2. Ряд распределения <table border="1" data-bbox="205 1113 692 1207"><tr><td>Значения</td><td>x_1</td><td>x_2</td><td>\dots</td><td>x_n</td><td>\dots</td></tr><tr><td>Вероятности</td><td>p_1</td><td>p_2</td><td>\dots</td><td>p_n</td><td>\dots</td></tr></table>$p_n = P(X = x_n), \quad \sum_i p_i = 1$</div>	Значения	x_1	x_2	\dots	x_n	\dots	Вероятности	p_1	p_2	\dots	p_n	\dots	<div>2. Ряд распределения группированной выборки ($\sum_i n_i = n$) <table border="1" data-bbox="732 1151 1377 1341"><tr><td>Интервал</td><td>$[y_1, y_2)$</td><td>$[y_2, y_3)$</td><td>\dots</td><td>$[y_k, y_{k+1}]$</td></tr><tr><td>Частота n_i</td><td>n_1</td><td>n_2</td><td>\dots</td><td>n_k</td></tr><tr><td>Относительная частота</td><td>$\frac{n_1}{n}$</td><td>$\frac{n_2}{n}$</td><td>\dots</td><td>$\frac{n_k}{n}$</td></tr></table></div>	Интервал	$[y_1, y_2)$	$[y_2, y_3)$	\dots	$[y_k, y_{k+1}]$	Частота n_i	n_1	n_2	\dots	n_k	Относительная частота	$\frac{n_1}{n}$	$\frac{n_2}{n}$	\dots	$\frac{n_k}{n}$
Значения	x_1	x_2	\dots	x_n	\dots																							
Вероятности	p_1	p_2	\dots	p_n	\dots																							
Интервал	$[y_1, y_2)$	$[y_2, y_3)$	\dots	$[y_k, y_{k+1}]$																								
Частота n_i	n_1	n_2	\dots	n_k																								
Относительная частота	$\frac{n_1}{n}$	$\frac{n_2}{n}$	\dots	$\frac{n_k}{n}$																								
<div>3. Плотность распределения $F(x) = \int_{-\infty}^x f(x) dx, f(x) \geq 0.$Свойства: 1. $\int_{-\infty}^{\infty} f(x) dx = 1$ 2. $F'(x) = f(x)$ в точках непрерывности $f(x)$</div>	<div>3. Гистограмма и полигон относительных частот На каждом из интервалов группировки строим прямоугольник, площадь которого равна относительной частоте: $S_i = n_i / n, \quad H_i = n_i / (n \cdot h)$ Полигон – ломаная, соединяющая середины верхних сторон прямоугольников.</div>																											

Характеристики случайных величин	Их оценка по выборке x_1, x_2, \dots, x_n
4. Числовые характеристики СВ	4. Оценки числовых характеристик СВ
<p>1) Среднее значение (математическое ожидание) для дискретной СВ: $MX = x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n + \dots$; для непрерывной СВ: $MX = \int_{-\infty}^{\infty} x f(x) dx$.</p>	<p>1) Эмпирическое среднее $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ несмещенная, состоятельная, эффективная оценка математического ожидания (MX)</p>
<p>2) Дисперсия $DX = M(X - MX)^2 = MX^2 - (MX)^2$; для дискретной СВ: $DX = \sum_i (x_i - MX)^2 p_i = \sum_i x_i^2 p_i - (MX)^2$; для непрерывной СВ: $DX = \int_{-\infty}^{\infty} x^2 f(x) dx - (MX)^2$.</p>	<p>2) Выборочная дисперсия $\bar{\sigma}_x^2 = \frac{1}{n} \sum_i x_i^2 - (\bar{x})^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$; исправленная выборочная дисперсия $s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$ – несмещенная оценка дисперсии.</p>
<p>3) Среднеквадратическое отклонение $\sigma_x = \sqrt{DX}$.</p>	<p>3) Выборочное среднеквадратическое отклонение (мера разброса около среднего) $\tilde{\sigma}_x = \sqrt{\tilde{\sigma}_x^2}$.</p>
<p>4) Квантиль уровня p – такое значение X_p, что $F(x_p) \leq p$, $F_n(x_p + 0) \geq p$</p> 	<p>4) Эмпирическая квантиль уровня p $F_n(x_p) \leq p$ $F_n(x_p + 0) \geq p$</p>
<p>5) Медиана $x_{0,5}$ – квантиль уровня 0,5</p> 	<p>5) Эмпирическая медиана $F_n(x_{0,5}) \leq 0,5$ $F_n(x_{0,5} + 0) \geq 0,5$.</p>

$$\text{Рассмотрим эмпирическое среднее } \bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n), \quad (1.11)$$

являющееся оценкой для математического ожидания.

Утверждение. Эмпирическое среднее является несмещенной, состоятельной, эффективной оценкой для математического ожидания.

Д о к а з а т е л ь с т в о. Используем свойства математического ожидания и дисперсии, известные из теории вероятностей.

1) Обозначим $MX=m$. Проверим, что \bar{x} – несмещенная оценка:

$$M\bar{x} = M \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} M [x_1 + \dots x_n] = \frac{1}{n} [Mx_1 + \dots Mx_n] = \frac{1}{n} \cdot m \cdot n = m = MX.$$

$$2) \text{ Дисперсия } D\bar{x} = D \left(\frac{1}{n} \sum_{i=1}^n x_i \right) = \frac{1}{n^2} (Dx_1 + \dots Dx_n) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n},$$

$\sqrt{D\bar{x}} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Известно, что это наименьшая из возможных дисперсий.

3) Состоятельность оценки \bar{x} следует из закона больших чисел, изучаемого в теории вероятностей.

Из этого утверждения делаем заключение:

Вывод о точности оценивания математического ожидания

Погрешность оценки \bar{x} прямо пропорциональна разбросу исходных данных и обратно пропорциональна корню из числа наблюдений (\sqrt{n}). С ростом числа наблюдений точность оценивания растет, но медленно (как \sqrt{n}). Например, при $n = 100$, $\sqrt{n} = 10$.

Несмещенной состоятельной оценкой дисперсии является исправленная выборочная дисперсия s^2 : $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$. (1.12)

При большом числе наблюдений статистика s^2 незначительно отличается от выборочной дисперсии $\tilde{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

1.6. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Любая точечная оценка $\tilde{\theta}$ параметра θ по выборке x_1, \dots, x_n – случайная величина, поэтому для конкретной выборки она может значительно отличаться от оцениваемого параметра θ (особенно для выборки малого объема). Поэтому наряду с точечными оценками используют и интервальные.

Пусть $\tilde{\theta}_1 = \tilde{\theta}_1(x_1, \dots, x_n)$, $\tilde{\theta}_2 = \tilde{\theta}_2(x_1, \dots, x_n)$ – две статистики, и $\tilde{\theta}_1 < \tilde{\theta}_2$.

Доверительным интервалом надежности p для параметра θ называется интервал $(\tilde{\theta}_1; \tilde{\theta}_2)$, содержащий оцениваемый параметр с доверительной вероятностью (надежностью) p : $P(\tilde{\theta}_1 < \theta < \tilde{\theta}_2) = p$.

Интервальная оценка тем точнее, чем меньше интервал $(\tilde{\theta}_1; \tilde{\theta}_2)$.

Пусть $\tilde{\theta}$ – точечная оценка θ . Если найдено такое число ε , что

$P(|\tilde{\theta} - \theta| < \varepsilon) = p$, то, заменив неравенство $|\tilde{\theta} - \theta| < \varepsilon$ эквивалентным неравенством $\tilde{\theta} - \varepsilon < \theta < \tilde{\theta} + \varepsilon$, получим для параметра θ доверительный интервал надежности p :

$$P(\tilde{\theta} - \varepsilon < \theta < \tilde{\theta} + \varepsilon) = p.$$

Оценка $\tilde{\theta}$ тем точнее, чем меньше ε (при заданной вероятности p). При построении доверительных интервалов используют квантили таких распределений, как нормальное распределение, распределение Стьюдента, хи-квадрат распределение и др. Значение квантилей определяют по статистическим таблицам (табл. П.А.1 и П.А.2).

Пример 1.11. По выборке x_1, \dots, x_n из нормально распределенной генеральной совокупности с неизвестными параметрами m и σ^2 построить доверительный интервал надежности p : 1) для математического ожидания m ; 2) для дисперсии σ^2 .

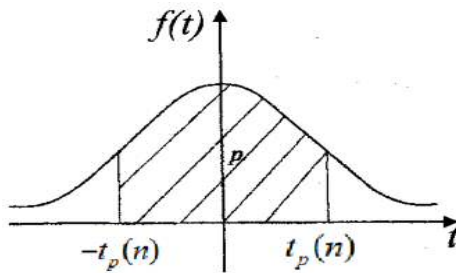


Рис. 1.4. Значение $t_p(n)$

Решение: 1) Пусть \bar{x} , s^2 определены формулами (1.11), (1.12). Известно, что СВ $\frac{\bar{x} - m}{s/\sqrt{n}} = T(n-1)$ имеет распределение Стьюдента с $n-1$ степенью свободы.

Пусть при заданных значениях надежности p и числа степеней свободы n , $t_p(n)$ – такое число (табл. П.А.2), что $P\{|T(n)| < t_p(n)\} = p$ (рис. 1.4). Выразив m из неравенства

$$\left| \frac{\bar{x} - m}{s/\sqrt{n}} \right| = |T(n-1)| < t_p(n-1),$$

получим доверительный интервал для математического ожидания m надежности p :

$$\bar{x} - \frac{s}{\sqrt{n}} t_p(n-1) < m < \bar{x} + \frac{s}{\sqrt{n}} t_p(n-1). \quad (1.13)$$

Известно, что СВ $\frac{(n-1)s^2}{\sigma^2} = \chi^2(n-1)$ имеет распределение χ^2 с $(n-1)$ степенью свободы. При заданных n и p определим квантили $\chi^2_{\alpha/2}(n)$ и $\chi^2_{1-\alpha/2}(n)$.

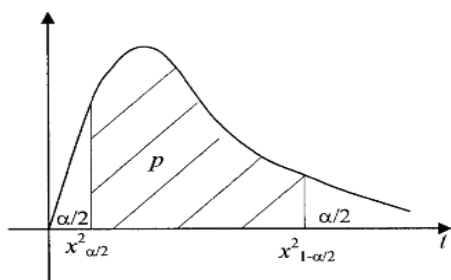


Рис. 1.5. Квантили распределения χ^2

Тогда $\chi^2_{\alpha/2}(n) < \chi^2(n-1) < \chi^2_{1-\alpha/2}(n)$ с вероятностью $1 - \alpha = p$ (рис. 1.5). Заменяя в этих неравенствах $\chi^2(n-1)$ на $(n-1) \cdot s^2 / \sigma^2$ и выразив из них σ^2 , получим доверительный интервал для σ^2 :

$$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} \quad (1.14)$$

с вероятностью $p = 1 - \alpha$.

1.7. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ: ОСНОВНЫЕ ПОНЯТИЯ

Часто в экономических задачах результаты статистических наблюдений используются для проверки предположений (гипотез) относительно некоторых свойств генеральной совокупности.

Основные типы задач, приводящих к проверке гипотез:

1. Определение вида закона распределения случайной величины X (нормальный, равномерный, показательный и т. д.).
2. Сравнение различных экономических явлений (или технологических процессов или методов обработки) по каким-либо показателям.
3. Исследование зависимости случайных величин.

Статистической гипотезой H_0 называется предположение относительно параметра или закона распределения случайной величины (возможно, многомерной).

Статистическая гипотеза называется **простой**, если она однозначно определяет случайную величину, сложной – в противном случае.

Гипотеза называется **параметрической**, если требуется проверить предположения относительно параметров распределения.

Примеры гипотез:

- 1) $H_0 : X \sim N(m; \sigma^2)$ – простая, непараметрическая;
- 2) $H_0 : MX > 0$ – сложная, параметрическая.

Основная гипотеза H_0 – та, которую мы собираемся проверять. Например, H_0 : «студент сдаст экзамен на отлично», или $H_0: MX = 0$.

Альтернативную гипотезу обозначают через H_1 . Альтернативные гипотезы могут быть разные, например, $H_1: MX > 0$; $H_1: MX \neq 0$; $H_1: MX < 0$.

Альтернативная гипотеза представляет собой то заключение, которое мы делаем, если экспериментальные данные указывают на ложность гипотезы H_0 .

Правило, по которому принимается решение о принятии или отклонении гипотезы, называется **критерием K** . Для построения критерия K подбирают такую статистику, которая характеризует отклонение эмпирических данных от гипотетических (соответствующих гипотезе). Обычно распределение этой статистики известно и не зависит от оцениваемых параметров. Зачастую она нормальная или имеет распределение Стьюдента, или χ^2 (хи-квадрат). Эта статистика $T = T(x_1, \dots, x_n)$ называется статистикой критерия K .

Затем выбирают достаточно малое положительное число α – уровень значимости, и по нему находят такое множество V_α значений статистики T , что вероятность попадания T в это множество при условии справедливости гипотезы H_0 равна α (обычно полагают $\alpha = 0,01$ или $\alpha = 0,05$):

$$P(T \in V_\alpha | H_0) = \alpha.$$

Множество V_α называется **критической областью**.

Критерий проверки гипотезы H_0 формулируется так:

если $T \in V_\alpha$, то гипотеза H_0 отклоняется;

если $T \notin V_\alpha$, то нет оснований для отклонения H_0 .

Если значение T попало в критическую область V_α , то говорят, что отклонение эмпирических данных от гипотетических, характеризуемое статистикой T , значимо. Если же $T \notin V_\alpha$, то у вас нет причин отвергнуть гипотезу H_0 (построение критической области показано в примере 1.12 и на рис. 1.6).

При проверке гипотез могут возникнуть *ошибки первого и второго рода*. Ошибка первого рода возникает, когда мы отвергаем истинную гипотезу, ошибка второго рода – когда принимаем ложную гипотезу.

Рассмотрим пример из юридической практики. За нулевую гипотезу принимают вариант, что подсудимый невиновен. Ошибка первого рода происходит, когда суд невиновного признает виновным. Ошибка второго рода происходит, когда виновного ошибочно оправдывают.

В соответствии с формулой $P(T \in V_\alpha | H_0) = \alpha$ вероятность ошибочного отклонения гипотезы H_0 в случае ее истинности равна α .

В качестве примера рассмотрим критерий Стьюдента проверки гипотезы о равенстве среднего значения выборки из нормальной совокупности заданному значению m_0 .

Пример 1.12. По независимым наблюдениям x_1, \dots, x_n из нормальной совокупности требуется проверить гипотезу $H_0: m = m_0$ при двусторонней альтернативе $H_1: m \neq m_0$ и заданном уровне значимости α .

Решение. Если дисперсия σ^2 неизвестна, то по наблюдениям рассчитывается ее оценка. В качестве статистики критерия берется t -статистика:

$$t = \frac{\bar{x} - m_0}{s / \sqrt{n}},$$

которая имеет распределение Стьюдента $T(n-1)$ с $n-1$ степенью свободы, если гипотеза H_0 верна.

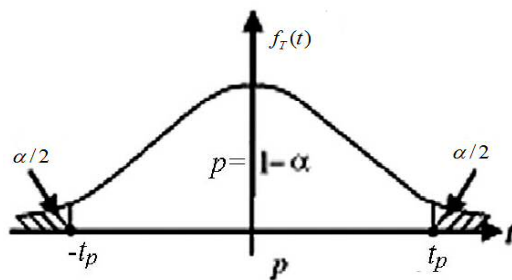


Рис. 1.6. Критические значения распределения Стьюдента

По табл. П.А.2, приведенной в приложении, определим: для заданной вероятности $p = 1 - \alpha$ и число степеней свободы $n - 1$ критическую точку $t_p(n - 1) = t_p$ – такое значение, что $P(|T(n - 1)| < t_p) = p = 1 - \alpha$ (рис. 1.6).

Тогда область $|T(n - 1)| \geq t_p(n - 1)$ – критическая область V_α , а *критерий Стьюдента* имеет вид: гипотеза H_0 не отклоняется на уровне значимости α , если

$$\left| \frac{\bar{x} - m_0}{s / \sqrt{n}} \right| < t_p, \quad (1.15)$$

и отклоняется в противном случае.

Задача проверки гипотез о параметрах распределения тесно связана с построением доверительных интервалов для параметров (см. п. 1.6). Действительно, соотношение (1.15) можно переписать так: гипотеза H_0 не отклоняется, если $m_0 \in (\bar{x} - t_p \cdot s / \sqrt{n}, \bar{x} + t_p \cdot s / \sqrt{n})$. Другими словами, если гипотетическое значение m_0 попадает в доверительный интервал надежности $p = 1 - \alpha$ для математического ожидания, то гипотеза H_0 не отклоняется, в противном случае – отклоняется.

1.8. ДВЕ ПЕРЕМЕННЫЕ: МЕРЫ ИЗМЕНЧИВОСТИ И СВЯЗИ

Ряд случайных величин, используемых в экономике, являются многомерными. *Двумерной СВ* называется пара случайных величин $(X, Y) = (X(\omega), Y(\omega))$.

Математическим ожиданием двумерной СВ (X, Y) называется вектор, составленный из математических ожиданий X и Y :

$$(MX, MY) = (m_x, m_y).$$

Ковариацией (covariation) случайных величин называется величина $\text{cov}(X, Y) = M(X \cdot Y) - MX \cdot MY$.

Для независимых случайных величин $M(XY) = MX \cdot MY$. Значит, $\text{cov}(X, Y) = 0$, то есть независимые СВ являются некоррелированными. Из независимости случайных величин следует их некоррелированность (равенство нулю их ковариации). Обратное неверно.

Коэффициентом корреляции (correlation) случайных величин X и Y называется нормированный коэффициент $r = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y}$.

Неверно утверждение, что некоррелированные величины независимы.

Свойства коэффициента корреляции:

1. Нормирован: $-1 \leq r \leq 1$.
 2. Если $Y = a + b \cdot X$, то есть X и Y – линейно зависимы, то при $b > 0$ $r = 1$, при $b < 0$ $r = -1$.
 3. При $r > 0$ Y в среднем возрастает, при $r < 0$ – в среднем убывает.
- Коэффициент корреляции является хорошей мерой линейной зависимости.

Ковариационной матрицей СВ (X, Y) называется матрица

$$\begin{pmatrix} DX & \text{cov}(X, Y) \\ \text{cov}(X, Y) & DY \end{pmatrix}.$$

Статистической оценкой ковариации и коэффициента корреляции по наблюдениям $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$ являются выборочная ковариация и выборочный коэффициент корреляции:

$$\widetilde{\text{cov}}(X, Y) = \frac{1}{n} \sum_i x_i y_i - \bar{x} \cdot \bar{y}, \quad \tilde{r}_{x,y} = \widetilde{\text{cov}}(X, Y) / (\tilde{\sigma}_x \cdot \tilde{\sigma}_y). \quad (1.16)$$

З а м е ч а н и е. Несмещенная оценка для ковариации рассчитывается по формуле $\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})$.

В качестве примера корреляционной зависимости приведем зависимость между уровнем образования и средней зарплатой: между ними существует положительная корреляция.

Пусть имеется n пар наблюдений над СВ (X, Y) : $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Изобразив их на графике, получим *корреляционное поле* (диаграмму рассеяния) (рис. 1.7).

Вид корреляционного поля зависит от значения коэффициента корреляции r (рис. 1.8):

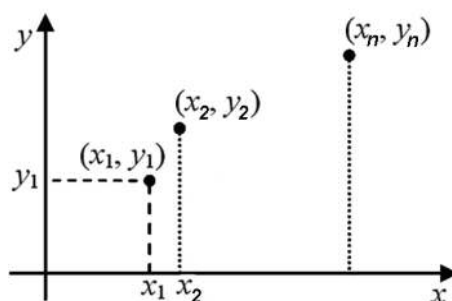


Рис. 1.7. Корреляционное поле

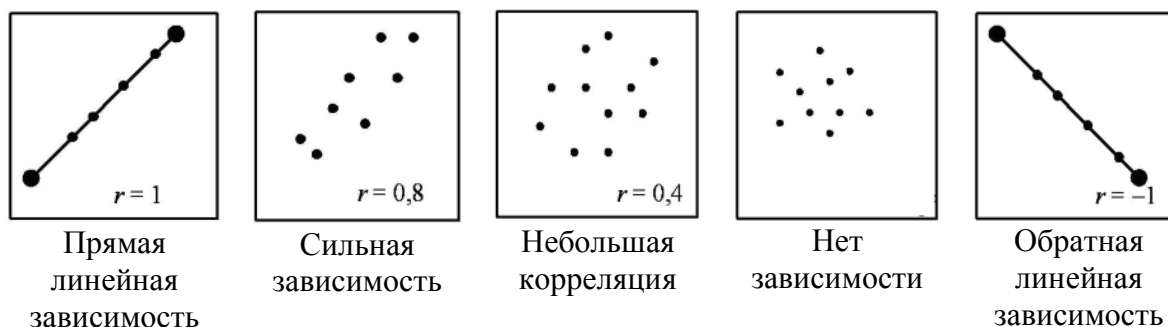


Рис. 1.8. Вид корреляционного поля в зависимости от коэффициента корреляции

1.9. ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ (МНК)

Регрессионный анализ – это анализ статистической зависимости СВ Y от X . СВ X может быть многомерной: $X = (X_1, \dots, X_k)$.

Различают два основных типа зависимости величин X и Y :

1. **Функциональная** зависимость: $Y = f(X)$. В этом случае поведение Y полностью определяется поведением X . Например, в формулу для дисперсии входит СВ $Y = X^2$: $DX = MX^2 - (MX)^2$.

2. **Статистическая** зависимость: $Y = f(X) + \varepsilon$, или $Y = f(X; \varepsilon)$, где ε – случайная компонента (не зависящая от X).

В парном регрессионном анализе предполагается, что величина X – одномерная. Предположим, что имеется n пар наблюдений

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Значения x_1, \dots, x_n величины X могут быть как случайными, так и неслучайными. Требуется оценить зависимость Y от X . Изобразив точки (x_i, y_i) , $i=1, \dots, n$, на графике, получим корреляционное поле (диаграмму рассеяния) (рис. 1.7).

Пример 1.13. Пусть исследуется зависимость расходов на личное потребление Y от располагаемого дохода X . Простейшая связь между Y и X – линейная: $Y = \alpha + \beta X$, где β ($0 < \beta < 1$) – некоторая постоянная, характеризующая в данном круге хозяйств их склонность к потреблению, связанную с традициями и привычками, а α – «автономное потребление». Однако если изобразить экспериментальные данные в виде точек на плоскости, то они не будут лежать на прямой $Y = \alpha + \beta X$, а будут образовывать «облако», вытянутое в некотором направлении. Поэтому **модель линейной регрессии** принимает форму $y_i = \alpha + \beta x_i + \varepsilon_i$, $i = 1, \dots, n$, и содержит две составляющие:

1) составляющую $\alpha + \beta \cdot X$, где X выступает как объясняющая (независимая) переменная;

2) случайную составляющую ε , отражающую совокупное влияние дополнительных факторов, не включенных в модель.

Оценивание зависимости проводится в два этапа:

I этап – определение вида зависимости с точностью до неизвестных параметров: $Y = f(X; \alpha_1, \dots, \alpha_k)$, где $\alpha_1, \alpha_2, \dots, \alpha_k$ – параметры.

II этап – нахождение оптимальных значений параметров.

Вид зависимости определяется на первом этапе на основе:

1) визуального исследования расположения наблюдений на корреляционном поле;

2) понимания экономической или физической сути задачи;

3) опыта предыдущих исследований.

Наиболее употребляемые в регрессионных моделях функции имеют вид:

- 1) $y = \alpha + \beta x$ – простая линейная регрессия;
- 2) $y = \alpha_0 + \alpha_1 x + \dots + \alpha_k x^k$ – полиномиальная регрессия;
- 3) $y = \alpha + \beta \cdot \ln(x)$ – логарифмическая регрессия;
- 4) $\ln(y) = \alpha + \beta x$ – экспоненциальная регрессия, и другие.

Замечание. Оптимальная оценка зависимости Y от X определяется уравнением *теоретической регрессии* $\varphi(x) = M(Y/X = x)$, изучаемой в курсе теории вероятностей. Эта функция минимизирует среднее значение квадрата отклонения Y от $\varphi(X)$: $M(Y - \varphi(X))^2 \rightarrow \min$.

Функция $y = \varphi(x)$ априори неизвестна, поэтому ищется ее оценка по наблюдениям – *эмпирическое уравнение регрессии*.

На втором этапе оценки коэффициентов определяются по **методу наименьших квадратов** (МНК).

Пусть $\tilde{y}_i = f(x_i; \alpha_1, \dots, \alpha_k)$ – расчетные значения, $\delta_i = y_i - \tilde{y}_i$ – погрешность (отклонение эмпирических значений от расчетных), $i = 1, 2, \dots, n$,

$$S^2 = \sum_{i=1}^n (y_i - \tilde{y}_i)^2 = \sum_{i=1}^n \delta_i^2 \text{ – сумма квадратов отклонений.}$$

В МНК наилучшими считаются те значения параметров, которые минимизируют сумму квадратов отклонений S^2 .

В случае линейной модели для минимизации функции

$$S^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 = S^2(a, b)$$

составляют и решают систему нормальных уравнений $\frac{\partial S^2}{\partial a} = 0, \frac{\partial S^2}{\partial b} = 0$.

Нетрудно проверить, что ее решением являются

$$b = \frac{\widetilde{\text{cov}}(X, Y)}{\tilde{\sigma}_x^2}, \quad a = \bar{y} - b\bar{x}, \quad (1.17)$$

где \bar{y}, \bar{x} – эмпирические средние СВ X и Y , $\widetilde{\text{cov}}(X, Y)$ – выборочная ковариация, $\tilde{\sigma}_x^2$ – выборочная дисперсия СВ X .

Итак, линейное **уравнение эмпирической регрессии** имеет вид:

$$y = a + bx, \quad (1.18)$$

где a и b рассчитываются, в соответствии с МНК, по формуле (1.17).

Соответствующая **теоретическая модель парной линейной регрессии**

$$Y = \alpha + \beta \cdot X + \varepsilon, y_i = \alpha + \beta \cdot x_i + \varepsilon_i, \quad i = 1, \dots, n.$$

Подчеркнем, что коэффициенты a и b , рассчитываемые по формулам (1.17), являются статистическими оценками коэффициентов модели α и β .

Причины наличия случайной составляющей в модели

1. Не включены некоторые объясняющие переменные, например, психологический фактор.

2. Агрегирование переменных (попытка объединить ряд микроэкономических факторов, чтобы каждый не учитывать отдельно). Пример – функция суммарного потребления.

3. Не совсем правильное описание структуры модели (чаще всего – наличие сдвига во времени).

4. Неправильный выбор функциональной зависимости.

5. Ошибки измерения (из-за наличия связи между переменными).

Основные формулы, относящиеся к парному регрессионному анализу, приведены в табл. 1.3.

Цель регрессионного анализа состоит в объяснении поведения зависимой переменной y . Разброс наблюдаемых значений y_1, \dots, y_n можно описать с помощью выборочной дисперсии $\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ или с помощью

суммы квадратов уклонений $S_{\text{общ}}^2 = \sum_{i=1}^n (y_i - \bar{y})^2$. В парном регрессионном анализе мы стараемся объяснить поведение Y с помощью его зависимости от X , представив y_i в виде:

$$y_i = \tilde{y}_i + \delta_i = a + bx_i + \delta_i,$$

где \tilde{y}_i – значение Y , спрогнозированное по значению x_i , а остаток δ_i есть разница между фактическим и спрогнозированным значением Y .

Оценивание качества приближения экспериментальных данных уравнением регрессии называется **оцениванием качества модели**.

Остаток δ_i – та часть y_i , которую мы не можем объяснить с помощью уравнения регрессии, наличие δ_i связано с действием прочих неучтенных или случайных факторов (соответствующий рисунок см. в табл. 1.3).

Итак, общую сумму квадратов отклонений можно разложить на две части – объясненную уравнением регрессии и необъясненную:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \quad (1.19)$$

общий разброс
(общая сумма квадратов)

сумма квадратов, объясненная регрессией

остаточная сумма квадратов

или $S_{\text{общ}}^2 = S_{\text{рег}}^2 + S_{\text{ост.}}^2$. Это тождество называется **основным тождеством регрессионного анализа**.

Таблица 1.3

Линейная регрессия и метод наименьших квадратов

Название	Формулы
Корреляционное поле	
Эмпирическое уравнение регрессии	$\tilde{y} = a + bx$
Погрешности	$\delta_i = y_i - \tilde{y}_i$, где $\tilde{y}_i = a + bx_i$, $i = 1, 2, \dots, n$
Сумма квадратов уклонений	$S^2 = \sum_i (y_i - \tilde{y}_i)^2 = \sum_i (y_i - a - bx_i)^2 \rightarrow \min$
Коэффициенты	$a = \bar{y} - b\bar{x}$, $b = \widetilde{\text{cov}}(X; Y) / \tilde{\sigma}_x^2$
Выборочная дисперсия	$\tilde{\sigma}_x^2 = \frac{1}{n} \sum x_i^2 - (\bar{x})^2$
Выборочные коэффициенты ковариации, корреляции	$\widetilde{\text{cov}}(X; Y) = \frac{1}{n} \sum x_i y_i - \bar{x} \cdot \bar{y}$ $\tilde{r} = \widetilde{\text{cov}}(X; Y) / \tilde{\sigma}_x \tilde{\sigma}_y$
Коэффициент детерминации (доля объясненной дисперсии)	$R^2 = 1 - \frac{\sum (y_i - \tilde{y}_i)^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (\tilde{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$
Основное тождество регрессионного анализа	$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \tilde{y}_i)^2$ $S_{\text{общ}}^2 = S_{\text{регр}}^2 + S_{\text{ост}}^2$
Статистика Фишера	$F_{\text{расч}} = \frac{R^2}{1 - R^2} \cdot (n - 2) = \frac{S_{\text{регр}}^2}{S_{\text{ост}}^2} \cdot (n - 2)$

Если бы прочие факторы не влияли на результат, то все экспериментальные точки лежали бы на прямой, остаточная сумма $S_{\text{ост}}^2$ равнялась бы нулю, а сумма квадратов, объясненная регрессией, совпадала бы с общей

суммой квадратов. Если же фактор X не оказывает влияние на Y , то $b = 0$, линия регрессии горизонтальна и $\tilde{y}_i = \bar{y}$, то есть все экспериментальные точки лежат на прямой. Тогда остаточная сумма равна нулю, а общая сумма квадратов совпадает с объясненной регрессией суммой. Так как на практике точки корреляционного поля рассеяны вдоль линии регрессии, то пригодность уравнения регрессии для прогноза зависит от того, какая доля вариации Y приходится на объясненную вариацию. Показателем этого служит коэффициент детерминации

$$R^2 = 1 - \frac{S_{ост}^2}{S_{общ}^2} = \frac{S_{рег}^2}{S_{общ}^2}. \quad (1.20)$$

Заметим, что $0 \leq R^2 \leq 1$. Коэффициент детерминации равен доле дисперсии Y , объясненной регрессией, в общей дисперсии Y .

В случае парной линейной регрессии нетрудно доказать, что коэффициент детерминации равен квадрату выборочного коэффициента корреляции: $R^2 = \tilde{r}_{x,y}^2$. Качество линейной регрессионной модели тем выше, чем ближе к 1 выборочный коэффициент детерминации.

Графическое представление результатов регрессионного анализа включает подобранную линию регрессии и две доверительные полосы (при заданной доверительной вероятности) – для среднего значения СВ Y и индивидуальных наблюдений. Пример приведен ниже, на рис. 2.4.

Почему при оценивании параметров используется МНК (метод наименьших квадратов)? Потому что этот метод дает несмещенные, эффективные оценки параметров α и β в случае выполнения предположений Гаусса–Маркова о случайной составляющей.

Условия Гаусса–Маркова:

1. Математическое ожидание случайной компоненты в каждом наблюдении равно нулю: $M\varepsilon_i = 0, i = 1, 2, \dots, n$.

Это, по сути, означает, что ε не имеет систематического смещения (колебания – около нуля).

2. Дисперсия случайной составляющей для всех наблюдений постоянна:

$$D\varepsilon_1 = D\varepsilon_2 = \dots = D\varepsilon_n = \sigma^2.$$

3. Нет систематической связи между случайной составляющей в двух различных наблюдениях: $cov(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$.

4. Случайная составляющая ε меняется независимо от объясняющей переменной X .

Дополнительное условие – предположение о нормальности случайной составляющей: $\varepsilon_i \sim N(0, \sigma^2)$.

Предположение о нормальности основывается на центральной предельной теореме теории вероятности, в соответствии с которой случайные величины, являющиеся суммой большого числа других случайных факторов, ни один из которых не является доминирующим, имеют приближенно нормальное распределение (даже если отдельные факторы его не имели).

Случайный член определяется несколькими факторами, которые не входят в явной форме в уравнение регрессии. Поэтому даже если ничего не известно о распределении факторов, мы имеем основание предположить, что ε имеет нормальное распределение. О силе воздействия случайных погрешностей в регрессионной модели говорит величина остаточной дисперсии σ^2 . Априори она неизвестна, ее несмещенной оценкой является *выборочная остаточная дисперсия*

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n \delta_i^2 \quad (1.21)$$

Качество модели. Коэффициент детерминации

Сначала остановимся на качестве оценок параметров уравнения регрессии, полученных по МНК (формулы (1.17)). В соответствии с теоремой Гаусса–Маркова эти оценки являются наилучшими линейными несмещенными оценками.

Теорема Гаусса–Маркова

При выполнении условий Гаусса–Маркова оценки a и b коэффициентов α и β , рассчитанные по МНК, имеют наименьшую дисперсию в классе всех линейных несмещенных оценок, причем

$$b = \beta + \varepsilon_b, \quad Mb = \beta, \quad a = \alpha + \varepsilon_a, \quad Ma = \alpha, \\ \sigma_b = \frac{\sigma}{\sqrt{n}\sigma_x}, \quad \sigma_a = \frac{\sigma}{\sqrt{n}} \sqrt{1 + \frac{\bar{x}^2}{\tilde{\sigma}_x^2}}. \quad (1.22)$$

Если выполнено предположение о нормальности, то $b \sim N(\beta, \sigma_b)$, $a \sim N(\alpha, \sigma_a)$.

На основании анализа формул (1.22) можно сделать вывод о точности оценивания коэффициентов регрессии. Покажем, как это сделать для параметра β . Точность оценки b определяется величиной $\sigma_b = \frac{\sigma}{\sqrt{n}\sigma_x}$.

Выводы о точности: 1) с ростом числа экспериментов точность увеличивается, но медленно, как \sqrt{n} ;

2) чем больше фактор случайности модели σ , тем точность оценок меньше;

3) чем меньше σ_x , тем хуже точность.

Аналогичные выводы можно сделать о точности оценивания параметра α .

1.10. ПРОВЕРКА ГИПОТЕЗ, ОТНОСЯЩИХСЯ К УРАВНЕНИЮ РЕГРЕССИИ

Есть два способа начать статистическое исследование – с теоретического выдвижения гипотезы или с эмпирического анализа. Можно сначала сформулировать гипотезу, а затем использовать результаты эксперимента для ее проверки. Это приведет к задаче проверки статистической гипотезы.

С другой стороны, можно сначала провести эксперимент и затем определить, какие из теоретических гипотез соответствуют экспериментальным данным. При оценивании параметров распределений это приведет к построению доверительных интервалов.

Первый этап при проверке гипотез – выдвижение нулевой гипотезы H_0 на основе каких-то гипотетических соображений.

Пример. Пусть X – темп инфляции, вызванной ростом заработной платы, а Y – темп общей инфляции. Можно предположить, что между ними имеется линейная зависимость

$$Y = \alpha + \beta X + \varepsilon. \quad (1.23)$$

Далее, можно выдвинуть гипотезу H_0 о том, что темп общей инфляции равен темпу инфляции, вызванной ростом зарплат, что соответствует случаю $\beta = 1$. Альтернативная гипотеза $H_1: \beta \neq 1$.

Другой пример нулевой гипотезы – гипотеза об отсутствии связи между величинами X и Y , что соответствует значению $\beta = 0$ в модели (1.23).

Второй этап – процедура проверки выдвинутой гипотезы.

Покажем, как проверяется гипотеза H_0 о равенстве коэффициента β конкретному значению β_0 .

Итак, пусть $H_0: \beta = \beta_0$, альтернативная гипотеза $H_1: \beta \neq \beta_0$.

Проверку гипотезы проведем в предположении, что выполнены четыре условия Гаусса–Маркова и дополнительное условие о нормальности погрешностей: $\varepsilon_i \sim N(0, \sigma^2)$. Рассмотрим два случая. В первом случае величина σ^2 известна. Значит, (см. п.1.9) оценка b имеет нормальное распределение с параметрами β_0 и σ_b , где $\sigma_b = \frac{\sigma}{\sqrt{n}\sigma_x}$. В качестве статистики критерия берется величина $U = (b - \beta_0) / \sigma_b$, имеющая нормальное распределение с параметрами 0 и 1: $U = (b - \beta_0) / \sigma_b \sim N(0, 1)$.

Выберем уровень значимости α (обычно $\alpha = 0,01$ или $\alpha = 0,05$) и по таблице нормального закона (табл. П.А.1) определим квантиль

$U_{1-0,5\alpha}$ уровня $1 - \alpha/2$. Например, если $\alpha = 0,05$, то

$1 - \alpha/2 = 0,975$, $U_{0,975} = 1,96$. Получаем область принятия гипотезы на уровне значимости α (рис. 1.9): $|U| = \left| \frac{b - \beta_0}{\sigma_b} \right| < U_{1-0,5\alpha}$, то есть гипотеза H_0

принимается, если $\beta_0 - \Delta < b < \beta_0 + \Delta$, где $\Delta = \sigma_b \cdot U_{1-0,5\alpha}$

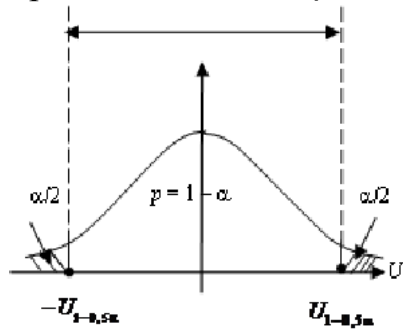


Рис. 1.9. Квантили нормального распределения

Второй случай – когда стандартное отклонение σ погрешностей неизвестно. В этом случае используется несмещенная оценка для σ^2 , которая равна

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n \delta_i^2. \text{ Оценкой для } \sigma_b \text{ является}$$

$$\tilde{\sigma}_b = s / \sqrt{n \tilde{\sigma}_x^2}. \quad (1.24)$$

Соответствующая статистика критерия (t -статистика) $t_b = \frac{b - \beta_0}{\tilde{\sigma}_b}$ имеет распределение Стьюдента с $n - 2$ степенью свободы. Поэтому область принятия гипотезы H_0 имеет вид:

$$\beta_0 - \Delta < b < \beta_0 + \Delta,$$

где $\Delta = \tilde{\sigma}_b \cdot t_p(n - 2)$; $p = 1 - \alpha$, $t_p(n - 2)$ – критическая точка уровня p распределения Стьюдента с $n - 2$ степенями свободы (табл. П.А.2).

По критерию Стьюдента проверяется гипотеза о значимом отличии от нуля коэффициентов уравнения теоретической регрессии α и β . Значимость уравнения регрессии в целом оценивается с помощью критерия Фишера. Для его построения используется основное тождество регрессионного анализа [см. (1.9)]: $S_{\text{общ}}^2 = S_{\text{регр}}^2 + S_{\text{ост}}^2$.

Любая сумма квадратов отклонений связана с числом степеней свободы (*d. f. – degrees of freedom*), определяемым через число единиц в выборке n , и число определяемых по выборке констант. Так как при вычислении $S_{\text{общ}}^2$ среднее рассчитывается по выборке, то теряется одна степень свободы, общее число степеней свободы равно $n - 1$.

В модели линейной регрессии

$$\tilde{y}_i = a + bx_i = \bar{y} - b\bar{x} + bx_i = \bar{y} - b(x_i - \bar{x}), i = 1, \dots, n,$$

поэтому

$$S_{\text{регр}}^2 = \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2 = b^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

и зависит только от одного параметра b , то есть $S_{\text{регр}}^2$ имеет одну степень свободы. Число степеней свободы для $S_{\text{ост}}^2$ равно $(n - 1) - 1 = n - 2$.

Разделив каждую сумму квадратов на соответствующее ей число степеней свободы, получим дисперсию, приходящуюся на одну степень свободы.

F-отношение (*F-ratio*) равно отношению факторной дисперсии к остаточной:

$$F = \frac{S_{\text{факт}}^2/k}{S_{\text{ост}}^2/(n-k-1)} = \frac{D_{\text{факт}}}{D_{\text{ост}}} \quad (1.25)$$

где n – число наблюдений, k – число параметров при переменных (как отмечено выше, в линейной модели $k = 1$; в параболической модели $k = 2$ и т. д.).

Критерий Фишера используется для проверки гипотезы о равенстве факторной и остаточной дисперсий $H_0: D_{\text{факт}} = D_{\text{ост}}$, что соответствует гипотезе об отсутствии влияния переменной X на Y .

По заданному уровню значимости α определяется табличное значение $F_{\text{табл}}$ распределения Фишера $F(k, n-k-1)$ с $k, n-k-1$ степенями свободы.

Если $F_{\text{табл}} > F_{\text{факт}}$, то считается, что факторная дисперсия значимо не отличается от остаточной, и делается вывод об отсутствии влияния фактора X на Y . Если $F_{\text{табл}} < F_{\text{факт}}$, что бывает, когда факторная дисперсия превышает остаточную в несколько раз, то гипотеза о случайности различий факторной и остаточной дисперсий отклоняется и признается статистическая значимость уравнения регрессии.

Замечание 1. Поделив в (1.25) числитель и знаменатель на $S_{\text{общ}}^2$, получим формулу, связывающую значения F и R^2 :

$$F = \frac{S_{\text{факт}}^2/S_{\text{общ}}^2}{(S_{\text{общ}}^2 - S_{\text{факт}}^2)/S_{\text{общ}}^2} \cdot \frac{n-k-1}{k} = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{k}. \quad (1.26)$$

Замечание 2. Известно, что в модели парной линейной регрессии t -статистики для коэффициента корреляции r и коэффициента b равны: $t_r = t_b$. Это означает, что проверка гипотезы о значимости отличия от нуля коэффициентов регрессии b равносильна проверке гипотезы о значимом отличии от нуля коэффициента корреляции.

Кроме того, значение t -статистики Стьюдента для коэффициента b связано со значением статистики Фишера: $t_b^2 = F$.

II. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

2.1. СЕМЕСТРОВАЯ РАБОТА ПО ТЕМЕ «СТАТИСТИЧЕСКИЙ АНАЛИЗ ОДНОМЕРНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН»

Задание 1

По заданному варианту выборочной совокупности независимых измерений случайной величины X (СВ X) (предварительно удалив резко выделяющиеся наблюдения):

1. Составить интервальный статистический ряд распределения относительных частот, построить гистограмму и полигон относительных частот.

2. Найти эмпирическую функцию распределения и построить ее график.

3. Вычислить точечные оценки для математического ожидания, дисперсии, коэффициентов асимметрии и эксцесса.

4. Исходя из общих представлений о механизме образования СВ X , а также по виду гистограммы и полигона относительных частот и вычисленным числовым характеристикам, выдвинуть гипотезу о виде закона распределения СВ X ; записать плотность распределения вероятностей и функцию распределения для выдвинутого гипотетического закона, заменяя параметры закона вычисленными для них оценками.

5. Вычислить интервальные оценки для математического ожидания и дисперсии, соответствующие доверительным вероятностям $p = 0,95$ и $p = 0,99$.

При выполнении работы *с использованием статистических ППП*

6.1. По критерию согласия хи-квадрат Пирсона проверить соответствие выборочного распределения гипотетическому закону для уровня значимости $\alpha = 0,05$.

6.2. Вычислить оценки моды, медианы, эксцесса, выборочных квартилей.

З а м е ч а н и е. Варианты экспериментальных данных к этому заданию приведены в разделе 3.2, задача 1.

Пример выполнения задания

Исходные данные:

X – число клиентов, посетивших отделение сбербанка в течение 100 дней (табл. 2.1).

Таблица 2.1

220	232	243	215	235	255	211	247	249	225
230	256	241	234	233	260	248	231	235	250
228	237	222	252	236	216	244	223	243	240
252	238	231	224	239	243	241	247	226	242
258	221	245	222	251	235	252	242	234	214
233	223	245	265	246	231	253	236	225	240
241	253	243	213	233	241	254	234	264	247
242	237	255	235	232	218	243	222	235	262
245	225	243	257	231	242	259	233	270	261
233	259	241	233	258	228	242	243	232	232

1. Сгруппируем исходные данные, то есть разобьем их на k частичных интервалов $(y_i, y_{i+1}]$, $i = 1, 2, \dots, k$, и посчитаем частоту попадания наблюдаемых значений в частичные интервалы.

При выборе числа k частичных интервалов руководствуются обычно следующими правилами:

- при $n \geq 200$ выбираем от 10 до 20 интервалов; при $n = 100$ лучше брать пять-семь интервалов;
- выбранные k интервалов должна охватывать всю область данных;
- интервалы не должны перекрываться; первый интервал включает левую границу;
- выбрать интервалы желательно одинаковой ширины h и так, чтобы границы были целыми числами или несложными дробями.

Для нашего примера: $x_{\max} - x_{\min} = 270 - 211 = 59$.

Увеличим диапазон до отрезка $[210; 270]$ длины 60. Этот отрезок удобно разбить на 6 интервалов длины $h = 10$ (табл. 2.1). За начало первого интервала берем целое число 210. Результаты промежуточных расчетов приведены в табл. 2.2.

Таблица 2.2

№ п/п	частичные интервалы $(y_i; y_{i+1}]$	середины интервалов x_i^*	частоты n_i	относительные частоты $\frac{n_i}{n}$	накоплен- ные частоты
1	[210; 220]	215	7	0,07	0,07
2	(220; 230]	225	14	0,14	0,21
3	(230; 240]	235	30	0,3	0,51
4	(240; 250]	245	28	0,28	0,79
5	(250; 260]	255	16	0,16	0,95
6	(260; 270]	265	5	0,05	1
Σ			100	1	

З а м е ч а н и е. В расчетах использован ППП Statgraphics, в соответствии с которым группировка проведена в интервалы вида $(y_i; y_{i+1}]$ (а не $[y_i; y_{i+1})$), как в ряде других источников).

На основании найденных значений относительных частот и накопленных относительных частот построим графики гистограммы и полигона относительных частот (рис. 2.1 и 2.2). Высоты H_i столбиков гистограммы рассчитаны по формуле $H_i = \frac{n_i h}{n} = \frac{n_i}{10}$.

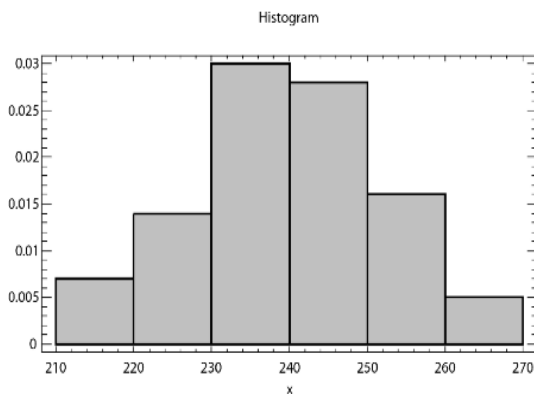


Рис. 2.1. График гистограммы

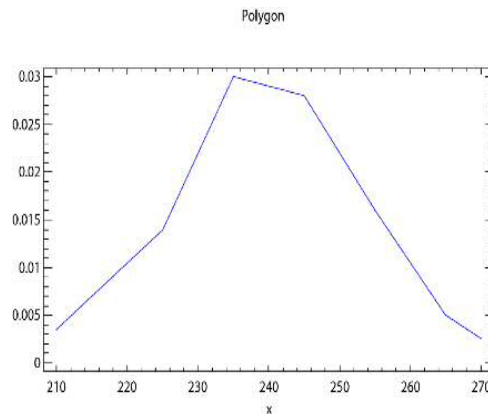


Рис. 2.2. График полигона

2. Эмпирическая функция распределения СВ X определяет для каждого значения x относительную частоту события $(X < x)$: $F_n(x) = \frac{n_x}{n}$, где n_x – число значений СВ X в выборке, меньших x ; n – объем выборки.

Значения эмпирической функции распределения находятся как накопленные частоты. График $F_n(x)$ представлен на рис. 2.3.

Аналитически эмпирическая функция записывается в виде:

$$F(x) = \begin{cases} 0,07 & \text{при } 210 < x \leq 220, \\ 0,21 & \text{при } 220 < x \leq 230, \\ 0,51 & \text{при } 230 < x \leq 240, \\ 0,79 & \text{при } 240 < x \leq 250, \\ 0,95 & \text{при } 250 < x \leq 260, \\ 1 & \text{при } 260 < x \leq 270. \end{cases}$$

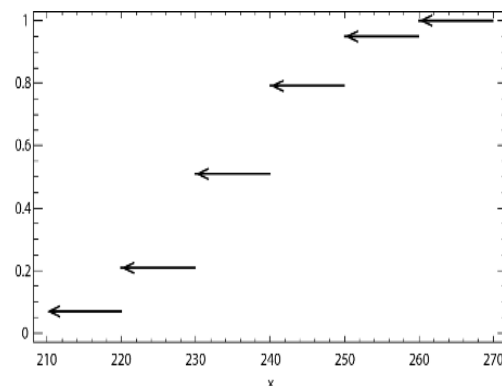


Рис. 2.3. График эмпирической функции распределения

3. Вычислим оценки числовых характеристик распределения. Несмещенной и состоятельной оценкой начального момента первого порядка (математического ожидания) является выборочное среднее $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Выборочные центральные моменты μ_m порядка m вычисляются по формуле:

$$\tilde{\mu}_m = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^m, \quad (2.1)$$

где x_i^* – середина i -го частичного интервала. Проводя вычисления с помощью ППП Statgraphics, получаем значение $\bar{x} = 239,23$. Вместо выборочной дисперсии $\tilde{\sigma}_x^2$ программой выдаются *несмещенная* оценка дисперсии $s^2 = 168,7$, и $s = 12,99$. Как указано в табл. 1.2, $\mu^2 = \tilde{\sigma}^2 = \frac{n-1}{n} \cdot s^2$, что позволяет рассчитать значение $\tilde{\sigma}_x^2 = 167,01$. При большом числе наблюдений ($n \geq 100$) оценки \bar{x} и $\tilde{\mu}_m$ могут быть приближенно вычислены по группированным данным:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i^* n_i, \quad \tilde{\mu}_m = \frac{1}{n} \sum_{i=1}^k (x_i^* - \bar{x})^m \cdot n_i \quad (2.2)$$

Для нашего примера

$$\bar{x} \approx (1/100) \cdot [215 \cdot 7 + 225 \cdot 14 + 235 \cdot 30 + 245 \cdot 28 + 255 \cdot 1 + 265 \cdot 5] = 239,7,$$

$$\tilde{\sigma}_x^2 \approx (1/100) \cdot [(215 - 239,7)^2 \cdot 7 + (225 - 239,7)^2 \cdot 14 + \dots + (265 - 239,7)^2] = 149,1.$$

$$\text{Тогда } s^2 = \frac{n}{n-1} \tilde{\sigma}_x^2 = 150,9.$$

Оценкой разброса будет выборочное среднеквадратическое отклонение $s = 12,27$. Видим, что оценки среднего и среднеквадратического отклонения, найденные по группированной выборке, мало отличаются от рассчитанных на ЭВМ значений $\bar{x} = 239,23$ и $s = 12,99$.

Для характеристики асимметрии и эксцесса используют центральные моменты третьего и четвертого порядков $\tilde{\mu}_3$ и $\tilde{\mu}_4$. Для этого вычисляют безразмерные величины $\tilde{A}_x = \frac{\tilde{\mu}_3}{\tilde{\sigma}_x^3}$ – выборочный коэффициент асимметрии,

и $\tilde{\mathcal{E}}_x = \frac{\tilde{\mu}_4}{\tilde{\sigma}_x^4} - 3$ – эксцесс или коэффициент крутости. Значения коэффициентов, рассчитанные на ППП Statgraphics, равны $\tilde{A} = 0,22$, $\tilde{\mathcal{E}} = -0,81$.

При проведении расчетов моментов $\tilde{\mu}_3$ и $\tilde{\mu}_4$ вручную вместо формул (2.1) можно использовать формулы (2.2).

Заметим, что для нормального закона коэффициенты асимметрии и эксцесса равны нулю. О значимости асимметрии и эксцесса судят на основании сравнения полученных значений \tilde{A} и $\tilde{\mathcal{E}}$ с величинами среднеквадратических погрешностей их определения, которые вычисляются по формулам

$$\sigma_A = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}}, \quad \sigma_{\mathcal{E}} = \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}.$$

При $n = 100$, $\sigma_A = 0,24$, $\sigma_{\mathcal{E}} = 0,45$. Как видно, в рассматриваемом примере коэффициенты \tilde{A} и $\tilde{\mathcal{E}}$ по абсолютной величине сравнимы с погрешностями их определения (не выходят за рамки отрезка $[-2; 2]$), и поэтому не могут считаться значимыми.

4. Для наших экспериментальных данных среднее число клиентов, обслуживаемых отделением Сбербанка за день, оказалось приближенно равным 239. Можно предположить, что отклонение от среднего X является случайной величиной, порожденной совокупным действием большого числа факторов. К ним можно отнести различия в днях недели, погодные условия, близость к «критическим» срокам оплаты различных платежей (коммунальных услуг, налогов и т. п.), число работающих операторов и др. Поэтому СВ X можно представить в виде суммы ряда, вообще говоря, случайных факторов, которые можно считать малыми и независимыми (или слабо зависимыми): $X = X_1 + X_2 + \dots + X_r$.

Такое представление соответствует условиям центральной предельной теоремы. Поэтому распределение СВ X можно представить как сумму двух слагаемых: среднего числа обслуживаемых клиентов, и случайных колебаний, распределенных по нормальному закону.

Экспериментальные данные подтверждают это предположение. Вид гистограммы и полигона относительных частот напоминает нормальную кривую распределения; выборочные коэффициенты асимметрии и эксцесса по величине сравнимы с погрешностями их определения. Их отличие от нуля, вероятно, обусловлено ограниченным объемом выборки.

Поэтому гипотетическим распределением числа посещений банка следует считать нормальный закон со средним $m = \bar{x} = 239,23$ и дисперсией

$$\sigma^2 = s^2 = 168,7, \quad \sigma = 12,99.$$

При этом плотность распределения и функция распределения имеют

$$\text{вид: } f(x) = \frac{1}{12,99\sqrt{2\pi}} \exp\left[-\frac{(x - 239,23)^2}{2 \cdot 168,7}\right], \quad F(x) = \int_{-\infty}^x f(t)dt \quad (2.3)$$

5. Найдем доверительные интервалы для математического ожидания и среднеквадратического отклонения.

По заданной выборочной совокупности объема $n = 100$ были вычислены оценки $\bar{x} = 239,23$ и $s = 12,99$. В п. 4 обосновано предположение о том, что наша выборка извлечена из нормальной совокупности (оно проверяется в п.6).

Для нормального закона доверительный интервал для математического ожидания вычислим по формуле (1.13), подставив в нее найденные значения \bar{x} и s :

1) для $p = 0,95$ и $n = 100$ с вероятностью $0,95$

$$236,65 < m < 241,81$$

2) для $p = 0,99$ и $n = 100$ с вероятностью $0,99$

$$235,65 < m < 242,64.$$

Доверительный интервал для дисперсии найдем по формуле (1.14):

$$130 < \sigma^2 < 227,7 \text{ с вероятностью } p = 0,95.$$

$$120,11 < \sigma^2 < 251,2 \text{ с вероятностью } p = 0,99.$$

З а м е ч а н и е. Для других законов распределения при большом числе испытаний формулы (1.13), (1.14) используются как приближенные.

6. Проверка гипотезы о нормальном законе распределения СВ X проводится по критерию согласия хи-квадрат.

Процедура проверки осуществлялась с использованием ППП Statgraphics по следующей схеме:

а) Рассматриваем гипотетическую случайную величину, распределенную по нормальному закону со средним $m = \bar{x} = 239,23$ и среднеквадратическим отклонением $\sigma = s = 12,99$. Для нее вычисляем теоретические вероятности p_i , $i = 1, \dots, 6$, попадания в промежутки $(y_i, y_{i+1}]$. При этом полагаем, что $y_1 = -\infty$ и $y_7 = \infty$, так как по определению нормальная СВ X может принимать значения на всей числовой оси.

б) Вычисляем теоретические вероятности попадания СВ X в частичные интервалы по формуле: $p_i = F(y_{i+1}) - F(y_i)$, где $F(y)$ – функция распределения гипотетического закона, определяемая по формуле (2.3).

в) Определяем теоретические (гипотетические) частоты $n \cdot p_i$ (табл. 2.3). При этом если для некоторых интервалов $n \cdot p_i \leq 5$, то их объединяют с соседними так, чтобы гипотетические частоты стали больше пяти для каждого интервала.

Таблица 2.3.

Эмпирические частоты n_i	7	14	30	28	16	5
Теоретические частоты $n \cdot p_i$	6,8	16,9	28,7	27,4	14,8	5,4

г) Вычисляем, используя данные табл. 2.3, расчетное значение выборочной статистики χ^2 , равное нормированному отклонению эмпирических частот от теоретических: $\chi^2_{расч.} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$,

где k – число частичных интервалов изменения СВ X .

$$\chi^2_{расч.} = \left[\frac{0,2^2}{6,8} + \frac{2,9^2}{16,9} + \frac{1,3^2}{28,7} + \frac{0,6^2}{27,4} + \frac{1,2^2}{14,8} + \frac{0,4^2}{5,4} \right] = 0,69.$$

д) Определяем r – число степеней свободы χ^2 – распределения как число частичных интервалов k минус число s наложенных связей или условий. В нашем примере $r = 6 - 2 - 1 = 3$.

е) Для заданного уровня значимости α и числа степеней свободы r по таблице χ^2 -распределения определяем квантиль, являющийся границей критической области.

ж) Если значение $\chi^2_{расч.} < \chi^2_{табл.}$, то гипотеза о соответствии эмпирических данных гипотетическому закону принимается на уровне значимости α , в противном случае – отвергается. В нашем примере (вычислено на Statgraphics) $\chi^2_{расч.} = 0,69$. Для уровня значимости $\alpha = 0,05$ и числа степеней свободы $r = 3$ $\chi^2_{табл.} = 7,8$. Так как $\chi^2_{расч.} < \chi^2_{табл.}$, то делаем

вывод: гипотеза о том, СВ X распределена по нормальному закону, задаваемому формулами (2.3), принимается на уровне значимости $\alpha = 0,05$.

6.2. Рассчитанные на ППП Statgraphics значения статистик равны: медиана = 240, нижняя квартиль = 231,5, верхняя квартиль = 247,0, мода = 243.

2.2. СЕМЕСТРОВАЯ РАБОТА ПО ТЕМЕ «СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН»

Задание 2

1. По заданному варианту экспериментальных данных (x_i, y_i) , $i = 1, 2, \dots, n$ построить корреляционное поле и по визуальной оценке расположения точек на нем сделать предположение о виде зависимости Y от X . Отдельно рассмотреть резко выделяющиеся наблюдения.

2. Вычислить оценки числовых характеристик величин X и Y : эмпирические средние \bar{x}, \bar{y} , эмпирические дисперсии $\tilde{\sigma}_x, \tilde{\sigma}_y$, выборочную ковариацию $\text{cov}(X, Y)$ и выборочный коэффициент корреляции \tilde{r} .

3. Методом наименьших квадратов найти оценки коэффициентов a и b уравнения эмпирической регрессии. Записать уравнение эмпирической регрессии $y = a + b \cdot x$.

4. Предсказать значение y^* для заданного x^* : $y^* = a + b \cdot x^*$.

Вычислить $\tilde{y}_1 = a + bx_1$, погрешность $\delta_1 = y_1 - \tilde{y}_1$ и относительную погрешность δ_1/y_1 .

5. Построить прямую эмпирической регрессии $y = a + b \cdot x$ по точкам (x_1, y_1) и (x^*, y^*) на корреляционном поле.

6. Оценить качество модели: вычислить коэффициент детерминации R^2 ; определить значимость уравнения регрессии по критерию Фишера на уровне значимости $\alpha = 0,05$

При выполнении задания с использованием статистических пакетов Statgraphics, SPSS, Statistika и др. *дополнительно* выполнить задания:

7. а) вычислить среднеквадратические ошибки определения коэффициентов a и b , определить значимость коэффициентов по критерию

Стьюдента на уровне значимости $\alpha = 0,05$;

б) построить доверительный интервал для прогноза y^* и доверительную полосу для среднего значения СВ Y , соответствующие доверительной вероятности 0,95.

З а м е ч а н и е. Варианты экспериментальных данных к этому заданию приведены в разделе 3.2, задача 2.

Образец выполнения

Исходные данные.

Пусть X – личный доход в США (млрд долларов, в ценах 2000–2011 гг.), Y – персональный расход на потребление в 2000–2011 гг. (табл. 2.4)

Данные взяты на сайте <https://datamarket.com/data/set/1fsm/us-economic-statistics-disposable-personal-income-seasonally-adjusted-annual-rate#!>

<https://datamarket.com/data/set/1fs6/us-economic-statistics-personal-consumption-expenditures#!>

Таблица 2.4

Зависимость расходов на потребление от дохода

Годы	Личный доход (X)	Персональный расход на потребление (Y), млрд долларов	Годы	Личный доход (X)	Персональный расход на потребление (Y), млрд долларов
2000	7327,1	6830,3	2007	10123	9772,2
2001	7648,4	7151	2008	11024,5	10035,5
2002	8009,6	7439,1	2009	10788,8	9866,1
2003	8376,1	7804,1	2010	11179,6	10245,5
2004	8887,7	8272,7	2011	11553,2	10724,4
2005	9277,6	8803,5			
2006	9915,6	9300,9		$x^*=12000$	

1. По заданному варианту экспериментальных данных построим корреляционное поле (x_i, y_i) , $i = 1, 2, \dots, 12$ (см. рис. 2.3). На основании визуального исследования выдвинем гипотезу о линейной зависимости Y от X :

$$Y = \alpha + \beta X + \varepsilon.$$

2. Числовые характеристики случайных величин найдем по формулам, приведенным в табл. 1.2.

$$\bar{x} = (1/n)\sum x_i = (1/12)(7327,1 + 7648,4 + \dots + 11553,2) = 9534,4;$$

$$\bar{y} = (1/n)\sum y_i = (1/12)(6830,3 + 7151 + \dots + 10724,4) = 8853,8;$$

$$\begin{aligned}\tilde{\sigma}_x^2 &= (1/12)\sum x_i^2 - \bar{x}^2 = (1/12)((7327,1)^2 + \dots + (11553,2)^2) - (9534,4)^2 = \\ &= 2008664,8; \quad \tilde{\sigma}_x = \sqrt{2008664,8} = 1417,2;\end{aligned}$$

$$\begin{aligned}\tilde{\sigma}_y^2 &= (1/12)\sum y_i^2 - \bar{y}^2 = (1/12)((6830,3)^2 + \dots + (10724,4)^2) - (8853,8)^2 = \\ &= 1611909,8; \quad \tilde{\sigma}_y = \sqrt{1611909,8} = 1269,6;\end{aligned}$$

$$\begin{aligned}\widetilde{\text{cov}}(X, Y) &= (1/n)\sum x_i y_i - \bar{x} \bar{y} = (1/12)(6830,3 \cdot 7327,1 + \dots + 10724,4 \cdot \\ &11553,2) - 8853,8 \cdot 9534,4 = 1794293,6\end{aligned}$$

$$\tilde{r} = \widetilde{\text{cov}}(X, Y) / (\tilde{\sigma}_x \tilde{\sigma}_y) = 1794293,6 / (1269,6 \cdot 1417,2) = 0,99.$$

3. Оценки коэффициентов уравнения регрессии находим по формулам (1.17): $b = \widetilde{\text{cov}}(X, Y) / \tilde{\sigma}_x^2 = 1794293,6 / 2008664,8 = 0,893$

$$a = \bar{y} - b \cdot \bar{x} = 8853,8 - 0,893 \cdot 9534,4 = 339,5$$

Уравнение регрессии имеет вид

$$y = a + b \cdot x = 339,5 + 0,893 \cdot x. \quad (2.4)$$

4. Рассчитаем по уравнению регрессии значение \tilde{y}_1 и прогнозируемое значение y^* :

$$\tilde{y}_1 = a + b \cdot x_1 = 339,5 + 0,893 \cdot 7327,1 = 6882,6,$$

$$y^* = a + b \cdot x^* = 339,5 + 0,893 \cdot 12000 = 11055,5.$$

Погрешность $\delta_1 = y_1 - \tilde{y}_1 = 6830,3 - 6882,6 = -52,3$. Относительная погрешность вычисления y_1 равна 0,7 %: $\delta_1 / y_1 \approx -0,007$.

5. Построим линию регрессии $y = a + b \cdot x = 339,5 + 0,893 \cdot x$ по двум точкам (x_1, y_1) и (x^*, y^*) (рис. 2.4).

6. Коэффициент детерминации $r^2 = R^2 = (0,99)^2 = 0,98$, что говорит о высоком качестве модели. Действительно, 98 % изменчивости расходов Y объясняется влиянием фактора X , остальные 2 % – влиянием других факторов.

7. Критерий проверки гипотезы о значимости уравнения регрессии по критерию Фишера описан в п.1.10. Расчетное значение статистики

Фишера определяем по формуле (1.26), учитывая, что число факторов модели $k=1$: $F_{расч} = 0,98 \cdot 10 / (1 - 0,98) = 495$.

8. Критическое (табличное) значение статистики Фишера определим по таблице (например, [4], табл. А.3) для заданного уровня значимости $\alpha = 0,05$ и числа степеней свободы $k = 1$ и $n - k - 1 = 12 - 1 - 1 = 10$: $F_{табл}(1, 10) = 4,96$. Так как $F_{расч} > F_{табл}$, то гипотеза H_0 об отсутствии влияния переменной X на Y отклоняется, а уравнение регрессии (2.4) признается значимым на уровне значимости $\alpha = 0,05$.

Решение с помощью ППП Excel

При использовании в расчетах пакета Excel сначала в первые две колонки файла вводятся исходные данные. Далее можно в соседних колонках вычислить квадраты значений x_i^2 , y_i^2 и произведения значений двух столбцов $x_i \cdot y_i$, $i = 1, \dots, n$. Проводя суммирование по каждому из полученных пяти столбцов, получим значения $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum y_i^2$, $\sum x_i y_i$, по которым непосредственно рассчитываются числовые характеристики из заданий 2 и 3. Фрагмент расчетов для нашего примера приведен в табл. 2.5.

Таблица 2.5

n	X	y	x^2	y^2	xy
1	7327,1	6830,3	53686394	46652998	50046291,1
...
12	11553,1	10724,4	133476430	115012755	123901138
Итого	106245,8	112177,8	960020211	1087794120	1034519572

По найденным значениям сумм затем, как это сделано выше, проводятся все остальные расчеты.

Гораздо более эффективно задачу о нахождении уравнения линейной регрессии можно решить, используя встроенную функцию **ЛИНЕЙН.**, определяющую параметры линейной регрессии и ряд дополнительных статистик. Для обращения к этой функции:

1) открыть файл, содержащий анализируемые данные (два столбца), и выделить в нем область пустых ячеек 5×2 для вывода результатов анализа;

2) в главном меню выбрать пункт **Вставка/Функция**, а в нем функцию **Статистические/ЛИНЕЙН.**, и заполнить диалоговое окно ввода переменных (введя информацию о значениях X и Y и две константы 1 и 1);

3) вывести дополнительные статистики можно, нажав клавиши F2, затем Ctrl + Shift + Enter.

Решение с помощью ППП Statgraphics

1. Создайте файл, и в нем занесите ваши данные (x_i, y_i) , $i = 1, \dots, n$, в колонки X и Y соответственно.

2. Для нахождения числовых характеристик случайных величин X и Y откройте окно описательного анализа функций многих переменных **Describe\Numeric Data\One Variable Analysis**. В нем нажмите кнопку **Tabular Options** и в диалоговом окне отметьте пункты **Summary Statistics** и **Correlation**. Результат появится в виде таблицы, фрагмент которой представлен в табл. 2.6.

Таблица 2.6.

Summary statistics

	y	x
Count	12	12
Average	8853,8	9509,3
Variance	$2,1491 \cdot 10^6$	$1,758 \cdot 10^6$
Standard deviation	1325,9	1479,76

Приведенные значения для средних x , y (Average), несмещенных оценок дисперсий σ_x^2 , σ_y^2 (Variance) и для r (Correlation Coefficient, табл. 2.6) близки к найденным в приведенном выше образце выполнения. Их различие объясняется ограниченной точностью вычислений при расчетах на калькуляторе.

3. Для визуального определения вида зависимости случайной величины Y от X постройте корреляционное поле. Для этого нажмите на находящуюся на панели инструментов кнопку Scatterplot.

4. Так как точки корреляционного поля группируются вдоль прямой, то используем модель простой линейной регрессии. Выберем пункт меню Relate\Simple Regression. В появившемся окне ввода переменных сначала введите зависимую переменную y , а затем независимую переменную x .

5. На экране появятся результаты в виде табл. 2.7. В первом разделе таблицы приведены оценки параметров a (Intercept) и b (Slope) и среднеквадратичные ошибки их определения (Standard Error).

Таблица 2.7

Regression Analysis - Linear model: $Y = a + b \cdot X$

Dependent variable: y

Independent variable: x

Standard Parameter	T Estimate	Error	Statistic	P-Value
Intercept	307,1	307,205	1,0	0,34
Slope	0,899	0,032	28,12	0,00

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1,90966E7	1	1,90966E7	790,86	0,0000
Residual	241467,0	10	24146,7		
Total (Corr.)	1,9338E7	11			

Correlation Coefficient = 0,9937

R-squared = 98,75 percent

При проведении статистических исследований на ППП используется значение P -Value, равное вероятности того, что статистика критерия (T , F и др.) примет значение, большее или равное расчетному. Так как P -Value для t -статистики коэффициента b равно нулю, то по критерию Стьюдента определяем, что он значимо отличен от нуля. Значение t -статистики для

коэффициента a равно 1, а $P\text{-Value} = 0,34$. Это говорит о том, что коэффициент a в модели плохо обусловлен.

Второй раздел таблицы (Analysis of Variance) служит для анализа качества модели. Значение коэффициента детерминации R^2 ($R\text{-squared}$) показывает, что 98,75 % изменчивости Y объясняется моделью. Так как значение F -статистики ($F\text{-Ratio}$) велико, а $P\text{-Value} = 0$, то существует статистически значимая зависимость Y от X (на уровне значимости $\alpha = 0,01$), описываемая уравнением (2.4).

6. Для графика линии регрессии и доверительной полосы нажмем кнопку Graphical Options и выберем процедуру Plot of Fitted Model. Результат представлен на рис. 2.4.

7.

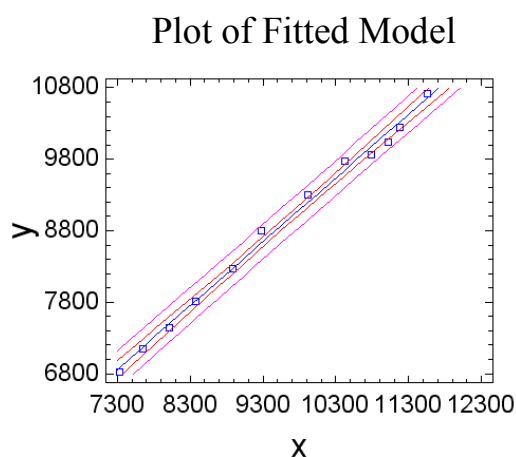


Рис. 2.4

7. Для расчета предсказанных значений, например, значения $y^* = a + b \cdot x^*$, нажмем кнопку Tabular Options и выберем процедуру **Forecasts**. Чтобы изменить установки анализа, например, ввести нужное значение x^* , по правой кнопке мыши выберем пункт

Pane Options\Forecasting Options. Для $x^* = 12000$ выводится прогноз $y^* = 11056,7$ и 95 % доверительный интервал для прогноза $(10921,5; 11191,9)$.

III. ЗАДАНИЯ К СЕМЕСТРОВЫМ РАБОТАМ

3.1. ВАРИАНТЫ ИНДИВИДУАЛЬНЫХ ЗАДАНИЙ ПО ТЕМЕ «СЛУЧАЙНЫЕ ВЕЛИЧИНЫ»

Задача № 1

Вариант 1.1. Имеются три базы с независимым снабжением. Вероятность отсутствия на каждой из баз нужного товара равна $0,1$; $0,2$; $0,3$ соответственно. Предприниматель решил закупить этот товар. Составить закон распределения числа баз X , на которых в данный момент этот товар отсутствует. Найти MX , DX , вероятность того, что X больше 1.

Вариант 1.2. В партии из 25 кожаных курток 5 имеют скрытый дефект. Покупают три куртки. Найти закон распределения числа X дефектных курток среди купленных, а также математическое ожидание и дисперсию X и вероятность того, что X окажется больше 1.

Вариант 1.3. Три покупателя независимо друг от друга могут сделать по одной покупке в магазине. Вероятность того, что покупку сделает первый покупатель, равна $0,9$, второй – $0,8$, третий – $0,6$. X – число покупок, сделанных этими покупателями. Найти ряд и функцию распределения для X , а также MX , DX и вероятность того, что X не меньше 2.

Вариант 1.4. Человек, имея 4 ключа, хочет открыть дверь. При этом он подбирает ключи случайно. Найти закон распределения, математическое ожидание и дисперсию числа испытаний X при условии, что испробованный ключ устраняется (только один ключ подходит к двери). Построить функцию распределения. Определить вероятность того, что число испытаний будет не менее двух.

Вариант 1.5. На пути движения автомашины 3 светофора, каждый из них либо разрешает, либо запрещает дальнейшее движение с вероятностью $0,5$. Найти закон распределения, математическое ожидание и дисперсию числа пройденных автомашиной светофоров до первой остановки. Построить функцию распределения. Определить вероятность того, что число пройденных светофоров будет не менее двух.

Вариант 1.6. Бросают три игральных кубика. Найти ряд и функцию распределения числа выпавших «пятерок» X , а также MX и DX и вероятность того, что X больше 1.

Вариант 1.7. Известно, что в партии из 20 телефонных аппаратов имеется 2 неисправных. Из партии выбрано 3 аппарата. Найти закон распределения, математическое ожидание и дисперсию числа неисправных аппаратов среди отобранных. Построить функцию распределения. Определить вероятность того, что число неисправных аппаратов среди отобранных будет не более двух.

Вариант 1.8. Производятся последовательно независимые испытания 3 приборов на надежность. Каждый следующий прибор испытывается только в том случае, если предыдущий оказался надежным, вероятность выдержать испытание для каждого из них равна $0,8$. Найти закон распределения, математическое ожидание и дисперсию числа приборов, прошедших испытания. Построить функцию распределения. Определить вероятность того, что будут испытаны хотя бы два прибора.

Вариант 1.9. В партии из 4 изделий одно бракованное. Чтобы его обнаружить, выбирают наугад одно изделие за другим и каждое вынуженное проверяют. Найти закон распределения, математическое ожидание и дисперсию числа проверенных изделий. Построить функцию распределения. Определить вероятность того, что хотя бы 2 изделия будут проверены.

Вариант 1.10. В партии 4 детали первого сорта и 3 детали второго сорта. Наудачу, одна за другой без возвращения в партию, отбираются детали до тех пор, пока деталь не окажется первосортной. Найти закон распределения, математическое ожидание и дисперсию числа отобранных при этом деталей второго сорта. Построить функцию распределения. Определить вероятность того, что деталей первого сорта будет отобрано не менее двух.

Вариант 1.11. Устройство состоит из трех элементов. Отказы элементов за некоторое время T независимы, а их вероятности равны соответственно $0,1$; $0,2$; $0,25$. Найти закон распределения, математическое ожидание и дисперсию числа отказавших за время T элементов. Построить функцию распределения. Определить вероятность того, что число отказавших элементов будет не менее двух.

Вариант 1.12. В партии из 100 деталей находится две бракованные детали. Из партии наудачу отбирается 10 деталей. Найти закон распределения, математическое ожидание и дисперсию числа бракованных изделий среди отобранных. Построить функцию распределения. Определить вероятность того, что среди отобранных деталей будет хотя бы одна бракованная.

Вариант 1.13. Баскетболист бросает мяч в корзину до первого попадания, но делает не более 4 бросков. Найти закон распределения, математическое ожидание и дисперсию числа бросков, если вероятность попадания в корзину равна $0,8$ при каждом броске. Построить функцию распределения. Определить вероятность того, что число бросков будет не менее 3.

Вариант 1.14. Среди 12 лампочек имеются две дефектные. Лампочки ввинчиваются в патрон и включается ток; при включении тока дефектная лампочка сразу же перегорает, после чего заменяется новой. Эта процедура повторяется до тех пор, пока лампочка не будет гореть. Найти закон распределения, математическое ожидание и дисперсию числа испробованных лампочек. Построить функцию распределения. Определить вероятность того, что будет испробовано более трех лампочек.

Вариант 1.15. Устройство состоит из трех элементов. Отказы элементов за некоторое время T независимы, а их вероятности равны соответственно $0,1$; $0,2$ и $0,25$. Найти закон распределения, математическое ожидание и дисперсию числа не отказавших элементов. Построить функцию распределения. Определить вероятность того, что не отказавших элементов будет не менее двух.

Вариант 1.16. В партии из 21 детали 7 деталей второго сорта, остальные первого. Отобраны случайным образом 3 детали. Найти закон распределения, математическое ожидание и дисперсию числа деталей второго сорта. Построить функцию распределения. Определить вероятность того, что в выборку попадет хотя бы одна деталь второго сорта.

Вариант 1.17. В конверте 12 карточек, среди которых 4 разыскиваемых. Наудачу конверт отбирают 3 карточки. Найти закон распределения, математическое ожидание и дисперсию числа разыскиваемых карточек среди отобранных. Ответ дать с точностью до $0,001$. Построить функцию распределения. Определить вероятность того, что среди отобранных есть хотя одна разыскиваемая карточка.

Вариант 1.18. Имеется 3 заготовки для одной и той же детали. Вероятность изготовления стандартной детали из каждой заготовки равна $0,2$. Найти закон распределения, математическое ожидание и дисперсию числа заготовок, оставшихся при изготовлении одной стандартной детали. Построить функцию распределения. Определить вероятность того, что число оставшихся заготовок не менее двух.

Вариант 1.19. Испытываемая аппаратура содержит 3 элемента. Отказы элементов за некоторое время T независимы, а их вероятности равны соответственно $0,05$; $0,1$ и $0,2$. Найти закон распределения, математическое ожидание и дисперсию числа отказавших за время T элементов. Построить функцию распределения. Определить вероятность того, что число отказавших элементов будет не более двух.

Вариант 1.20. В урне 3 белых и 4 черных шаров. Из урны наудачу, один за другим, без возвращения, извлекают шары до тех пор, пока не появится черный шар. Найти закон распределения, математическое ожидание и дисперсию числа появившихся при извлечении белых шаров. Построить функцию распределения. Определить вероятность того, что в выборке будет хотя бы 2 черных шара.

Вариант 1.21. Экзаменатор задает студенту дополнительные вопросы, но не более 3 вопросов. Вероятность того, что студент ответит на каждый из вопросов равна $0,8$. Преподаватель прекращает экзамен, как только студент не знает ответа на заданный вопрос. Найти закон распределения, математическое ожидание и дисперсию числа заданных вопросов. Построить функцию распределения. Определить вероятность того, что число заданных вопросов будет не менее двух.

Вариант 1.22. В поступившей на сборку партии находится 7 деталей первого сорта и 3 детали второго сорта. Наудачу, одно за другим, без возврата в партию, отбираются детали до тех пор, пока не появится деталь первого сорта. Найти закон распределения, математическое ожидание и дисперсию числа отобранных при этом деталей второго сорта. Построить функцию распределения. Определить вероятность того, что число отобранных деталей второго сорта будет не менее двух.

Вариант 1.23. В партии из 12 деталей 7 деталей второго сорта, остальные первого. Отобраны случайным образом 3 детали. Найти закон распределения, математическое ожидание и дисперсию числа деталей первого сорта в выборке. Построить функцию распределения. Определить вероятность того, что число деталей первого сорта в выборке будет не менее трех.

Вариант 1.24. На участке имеется три одинаковых станка, коэффициент использования которых по времени составляет 0,8. Найти закон распределения, математическое ожидание и дисперсию числа работающих станков. Построить функцию распределения. Определить вероятность того, что в данный момент времени будет простаивать (не работать) хотя бы один станок.

Вариант 1.25. В партии из 7 деталей имеется 5 деталей первого сорта. Наудачу отобраны 3 детали. Найти закон распределения, математическое ожидание и дисперсию числа деталей первого сорта среди отобранных. Построить функцию распределения. Определить вероятность того, что число деталей первого сорта будет не менее двух.

Задача № 2

Вариант 2.1. Вероятность того, что при составлении бухгалтерского баланса допущена ошибка, равна 0,3. Аудитору на заключение представлено n балансов предприятия. X – число положительных заключений на проверяемые балансы.

1) Построить ряд и функцию распределения X , если $n = 4$. Найти математическое ожидание и дисперсию X .

2) Найти вероятность того, что из 100 проверенных балансов ровно 4 содержат ошибки.

Вариант 2.2. Вероятность того, что аудитор допустит ошибку при проверке бухгалтерского баланса, равна 0,5. Аудитору на заключение представлено n балансов. X – число правильных заключений на проверяемые балансы.

1) Построить ряд и функцию распределения X , если $n = 3$. Найти математическое ожидание и среднеквадратическое отклонение X .

2) Найти вероятность того, что среди 100 сделанных аудитором заключений не менее 95 правильных.

Вариант 2.3. С завода поступило n партий измерительных приборов, по 20 приборов в каждой партии, из которых k приборов имеют знак качества. Наудачу отбираются по одному прибору из каждой партии.

1) Построить ряд и функцию распределения числа приборов со знаком качества среди отобранных, если $n = 4$ и $k = 3$. Вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что среди отобранных будет хотя бы один прибор со знаком качества, если $n = 40$, а $k = 1$.

Вариант 2.4. На конвейере задействовано n независимо работающих роботов, каждый из которых имеет надежность (вероятность безотказной заботы за время T), равную p .

1) Построить ряд и функцию распределения числа отказавших роботов среди четырех, закрепленных за механиком, если $p = 0,75$. Вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что за время T откажет не менее 3-х роботов, $n = 120$, а $p = 0,95$.

Вариант 2.5. За смену в среднем p процентов станков в автоматической линии, состоящей из n однотипных станков, требуют наладки.

1) Построить ряд и функцию распределения числа станков требующих наладки в течение смены, если $p = 40 \%$ и $n = 6$; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что наладок будет не менее 4 и не более 6, если $n = 50$, а $p = 4 \%$.

Вариант 2.6. В среднем 91 знаков из 100 передаются по каналу связи без искажений.

1) Построить ряд и функцию распределения числа искаженных (неправильных) знаков в сообщении, состоящем из четырех знаков; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что в сообщении, состоящем из 100 знаков, будет ровно 6 неизвестных знаков.

Вариант 2.7. В некотором цехе брак составляет $p \%$ изготовленных изделий.

1) Построить ряд и функцию распределения числа бракованных изделий среди четырех изделий, выбранных наудачу, если $p = 10 \%$. Вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Для случайной выборки в 1000 изделий и $p = 0,2 \%$ оценить вероятность того, что в выборке окажется ровно 5 бракованных изделий.

Вариант 2.8. На некотором предприятии k рабочих из общего n рабочих не имеют среднего образования. Требуется:

1) Построить ряд и функцию распределения числа рабочих, не имеющих среднего образования, среди b человек, отобранных наудачу, если на предприятии $n = 1000$, а $k = 250$; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Для предприятия, у которого $n = 1000$, а $k = 400$, оценить вероятность того, что среди наудачу отобранных рабочих 100 рабочих окажется не более 5, не имеющих среднего образования.

Вариант 2.9. Рабочий обслуживает линию, состоящую из n однотипных станков. Вероятность того, что каждый станок потребует внимания рабочего в течение часа равна p .

1) Построить ряд и функцию распределения числа станков, требующих в течение часа внимания рабочего, если $n = 4$, $p = 0,45$.

2) Оценить вероятность того, что за 1 час таких станков будет не более пяти, если $n = 100$, а $p = 0,025$.

Вариант 2.10. Завод выпускает в среднем 20 % изделий со знаком качества. В ОТК для проверки изделия поступают партиями по 5 штук.

1) Построить ряд и функцию распределения числа партий, содержащих 2 или 3 изделия со знаком качества, если проверено 4 партии изделий; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что среди 1000 партий, прошедших контроль, будет 5 партий, в каждой из которых окажется 4 изделия со знаком качества.

Вариант 2.11. В техническом устройстве n независимо работающих элементов, каждый из которых за время T отказывает с вероятностью p .

1) Построить ряд и функцию распределения числа отказавших элементов, если $n = 4$, а $p = 0,2$. Вычислить математическое ожидание и дисперсию рассматриваемой величины.

2) Оценить вероятность того, что при $n = 200$ и $p = 0,015$ откажет ровно 5 элементов.

Вариант 2.12. Имеется n станков с автоматическим приводом, которые включаются в работу независимо один от другого в случайные моменты времени так, что каждый из них в среднем работает $p\%$ всего рабочего времени.

1) Построить ряд и функцию распределения числа одновременно работающих станков в цехе, где $n = 5$, а $p = 80\%$, вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что на предприятии, у которого $n = 150$ и $p = 96\%$, в произвольно взятый момент времени не будут работать ровно 6 станков.

Варианты 2.13. На не отлаженной технологической линии брак составляет 20 % изготавливаемых изделий.

1) Построить ряд и функцию распределения числа бракованных изделий среди пяти изделий, выбранных наудачу; вычислить математическое ожидание и дисперсию рассматриваемой величины.

2) Для случайной выборки в 100 изделий найти вероятность того, что в выборке окажутся: а) ровно 10 бракованных изделий; б) от 0 до 10 бракованных изделий.

Вариант 2.14. В двух случаях из пяти радиолампа исправно работает дольше установленного срока.

1) Построить ряд и функцию распределения числа радиоламп, работающих дольше установленного срока, среди четырех радиоламп, взятых наудачу из большой партии; вычислить математическое ожидание и дисперсию рассматриваемой величины.

2) Найти вероятность того, что из 150 взятых наудачу радиоламп число таких, которые работают больше установленного срока, окажется:

а) ровно 40; б) меньше, чем 50.

Вариант 2.15. Электрическая цепь из n последовательно соединенных лампочек работает при повышенном напряжении в сети. Вероятность того, что лампочка перегорит, для всех n лампочек одинакова и в этих условиях равна 0,4.

1) Построить ряд и функцию распределения числа перегоревших лампочек в цепи из четырех лампочек, вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что при разрыве цепи из двухсот лампочек окажется перегоревших лампочек:

а) ровно половина; б) от 75 до 85.

Вариант 2.16. Вероятность обрыва нити на каждом из веретен ткацкого станка в течение времени t равна p .

1) Построить ряд и функцию распределения числа обрывов нити в течение одного часа у пяти веретен, если $t = 1$ час, вероятность $p = 0,18$. Вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что при обслуживании 1000 веретен будет два обрыва в течение 1 минуты, если для этого времени $p = 0,003$.

Вариант 2.17. Отдел технического контроля проверяет детали на стандартность. Вероятность того, что отдел признает деталь нестандартной, равна 0,2.

1) Построить ряд и функцию распределения числа нестандартных деталей среди пяти проверенных.

2) Оценить вероятность того, что в партии из 400 деталей окажется нестандартных:

а) ровно половина; б) от 75 до 85.

Вариант 2.18. При автоматической штамповке деталей 60 % продукции выпускается высшим сортом.

1) Построить ряд и функцию распределения числа деталей высшего сорта среди 5 деталей, взятых наудачу; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что из 800 деталей, изготовленных за смену, не менее 500 будут детали высшего сорта.

Вариант 2.19. Вероятность появления некоторого события при одном опыте 0,5.

1) Построить ряд и функцию распределения числа появления события при четырех опытах; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что при 100 опытах событие появится не менее 50 раз.

Вариант 2.20. В цехе имеется n станков, одинаковой мощности работающих независимо друг от друга в одинаковом режиме, при котором их привод оказывается включенным в течение 0,8 всего рабочего времени.

1) Построить ряд и функцию распределения числа включенных станков в произвольно взятый момент времени, если $n = 4$; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что в цехе, имеющем 100 станков, в произвольно взятый момент времени окажутся включенными:

а) 75 станков; б) от 70 до 86 станков.

Вариант 2.21. Вероятность выхода из строя за время t одного конденсатора равна 0,2.

1) Построить ряд и функцию распределения числа конденсаторов, вышедших из строя за время t , если на приборе 4 конденсатора; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что в устройстве, имеющем 100 конденсаторов, за время t выйдут из строя:

а) не менее 20 конденсаторов; б) менее 28.

Вариант 2.22. При изготовлении отливок получается 20 % дефектных.

1) Построить ряд и функцию распределения числа стандартных отливок из пяти изготовленных; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Определить, сколько необходимо запланировать отливок к изготовлению, чтобы с вероятностью не менее 0,95 была обеспечена программа выпуска изделий, для выполнения которой необходимо 100 стандартных отливок.

Вариант 2.23. 80 % изготовленных заводом электроламп выдерживают гарантийный срок службы.

1) Построить ряд и функцию распределения числа электроламп, выдерживающих гарантийный срок, среди четырех купленных электроламп; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что в партии из 500 электроламп число выдержавших гарантийный срок службы находится в пределах 440–480.

Вариант 2.24. Партия, состоящая из 200 однотипных радиоламп, содержит 80 радиоламп с истекшим сроком службы.

1) Построить ряд и функцию распределения числа радиоламп с истекшим сроком службы среди пяти радиоламп, взятых из партии наудачу; вычислить математическое ожидание и дисперсию рассматриваемой величины.

2) Определить какое количество радиоламп необходимо взять из партии, чтобы среди них с вероятностью 0,95 было не менее 25 радиоламп с истекшим сроком службы.

Вариант 2.29. Автоматизированная технологическая линия производит 50 % изделий высшего сорта.

1) Построить ряд и функцию распределения числа изделий высшего сорта среди четырех наудачу взятых изделий; вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Определить, сколько необходимо изготовить изделий, чтобы с вероятностью 0,997 в их числе было не менее 500 изделий высшего сорта.

Вариант 2.25. Вероятность изготовления деталей с заданными точностными характеристиками из стандартной заготовки равна p .

1) Построить ряд и функцию распределения числа бракованных изделий среди четырех изделий, изготовленных рабочим, для которого $p = 0,7$, вычислить математическое ожидание и дисперсию рассматриваемой случайной величины.

2) Оценить вероятность того, что среди 100 изготовленных деталей на станке-автомате, для которого $p = 0,97$, окажется не более двух бракованных.

Задача № 3

В вариантах 3.1–3.25 непрерывная СВ X задана функцией распределения $F(x)$. Найти:

- а) значения коэффициентов A и B ,
- б) плотность распределения $f(x)$,
- в) вероятность того, что СВ X примет значение в интервале (x_1, x_2) ,
- г) математическое ожидание и дисперсию СВ X ,
- д) построить графики $F(x)$ и $f(x)$.

Таблица 3.1

№	Функция распределения	x_1	x_2
3.1	$F(x) = \begin{cases} 0 & \text{при } x < 0, \\ A + B(2x^2 + 2x + 1)e^{-2x} & \text{при } x \geq 0 \end{cases}$	$x_1 = 1$	$x_2 = 2$
3.2	$F(x) = \begin{cases} 0 & \text{при } x \leq -\frac{\pi}{2}, \\ A + B\sin x & \text{при } -\frac{\pi}{2} < x < \frac{\pi}{2}, \\ 1 & \text{при } x \geq \frac{\pi}{2} \end{cases}$	$x_1 = -\frac{\pi}{4}$	$x_2 = \frac{\pi}{4}$
3.3	$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ A(B - \cos x) & \text{при } 0 \leq x \leq \pi, \\ 1 & \text{при } x \geq \pi \end{cases}$	$x_1 = \frac{\pi}{3}$	$x_2 = \frac{\pi}{2}$
3.4	$F(x) = \begin{cases} 0 & \text{при } x < -\frac{\pi}{4}, \\ A + B\sin x & \text{при } -\frac{\pi}{4} \leq x \leq \frac{\pi}{4}, \\ 1 & \text{при } x > \frac{\pi}{4} \end{cases}$	$x_1 = -\frac{\pi}{4}$	$x_2 = \frac{\pi}{4}$
3.5	$F(x) = \begin{cases} 0 & \text{при } x < 0, \\ A + Bx^2(x + 3) & \text{при } 0 \leq x \leq 1, \\ 1 & \text{при } x > 4 \end{cases}$	$x_1 = \frac{2}{3}$	$x_2 = 1$
3.6	$F(x) = \begin{cases} 0 & \text{при } x < 1, \\ A + B(4 - x)^2 & \text{при } 1 \leq x \leq 3, \\ 1 & \text{при } x > 3 \end{cases}$	$x_1 = 2$	$x_2 = 4$
3.7	$F(x) = \begin{cases} 0 & \text{при } x < 1, \\ A(x - B)^2 & \text{при } 1 \leq x \leq 3, \\ 1 & \text{при } x > 3 \end{cases}$	$x_1 = 1,5$	$x_2 = 2$
3.8	$F(x) = \begin{cases} 0 & \text{при } x < 2, \\ A + Bx^{-3} & \text{при } x \geq 2 \end{cases}$	$x_1 = 4$	$x_2 = 5$
3.9	$F(x) = \begin{cases} 0 & \text{при } x < 0, \\ A + Bx^3 & \text{при } 0 \leq x \leq 3, \\ 1 & \text{при } x > 3 \end{cases}$	$x_1 = 1$	$x_2 = 2$
3.10	$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ A\sin^2 \frac{x}{2} + B & \text{при } 0 < x \leq \pi, \\ 1 & \text{при } x > \pi \end{cases}$	$x_1 = -\pi$	$x_2 = \frac{\pi}{2}$
3.11	$F(x) = \begin{cases} 0 & \text{при } x \leq -1, \\ 0,5(Ax + 1)^2 & \text{при } -1 < x \leq 0, \\ 1 - 0,5(B - x)^2 & \text{при } 0 < x \leq 1, \\ 1 & \text{при } x > 1 \end{cases}$	$x_1 = -\frac{1}{2}$	$x_2 = \frac{1}{2}$
3.12	$F(x) = \begin{cases} 0 & \text{при } x < 0, \\ A + Be^{-\frac{x}{T}} & \text{при } x \geq 0 \end{cases}$	$x_1 = 0$	$x_2 = T$

№	Функция распределения	x_1	x_2
3.13	$F(x) = \begin{cases} 0 & \text{при } x < -2, \\ A + B \arccos \frac{x}{2} & \text{при } -2 \leq x \leq 2, \\ 1 & \text{при } x > 2 \end{cases}$	$x_1 = -3$	$x_2 = 1$
3.14	$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ A + B \cos x & \text{при } 0 < x < \frac{\pi}{2}, \\ 1 & \text{при } x \geq \frac{\pi}{2} \end{cases}$	$x_1 = \frac{\pi}{3}$	$x_2 = \frac{\pi}{2}$
3.15	$F(x) = \begin{cases} 0 & \text{при } x \leq 1, \\ A + Bx^{-0.5} & \text{при } 1 < x < 16, \\ 1 & \text{при } x \geq 16 \end{cases}$	$x_1 = 4$	$x_2 = 9$
3.16	$F(x) = \begin{cases} 0 & \text{при } x \leq 1, \\ Ax(x+B) & \text{при } 1 < x < 2, \\ 1 & \text{при } x \geq 2 \end{cases}$	$x_1 = 1,5$	$x_2 = 2$
3.17	$F(x) = \begin{cases} 0 & \text{при } x \leq -2, \\ \frac{1}{4A}(x^2 + 4x + B) & \text{при } -2 < x < 0, \\ \frac{1}{4A}(-x^2 + 4x + B) & \text{при } 0 < x < 2, \\ 1 & \text{при } x \geq 2 \end{cases}$	$x_1 = -1$	$x_2 = 2$
3.18	$F(x) = \begin{cases} 0 & \text{при } x \leq -1, \\ A + B \arcsin x & \text{при } -1 < x < 1, \\ 1 & \text{при } x \geq 1 \end{cases}$	$x_1 = -\frac{1}{2}$	$x_2 = \frac{1}{2}$
3.19	$F(x) = \begin{cases} 0 & \text{при } x \leq -\frac{\pi}{12}, \\ 0,5 \sin Ax + B & \text{при } -\frac{\pi}{12} < x \leq \frac{\pi}{12}, \\ 1 & \text{при } x > \frac{\pi}{12} \end{cases}$	$x_1 = \frac{\pi}{36}$	$x_2 = \frac{\pi}{18}$
3.20	$F(x) = \begin{cases} Ax - \frac{x^2}{3} + B & \text{при } 0 < x \leq 1, \\ 1 & \text{при } x > 1 \end{cases}$	$x_1 = 0$	$x_2 = 0,5$
3.21	$F(x) = \begin{cases} 0 & \text{при } x < 1, \\ A + \frac{B}{x^3} & \text{при } x \geq 1 \end{cases}$	$x_1 = 2$	$x_2 = 4$
3.23	$F(x) = \begin{cases} 0 & \text{при } x \leq -1, \\ A \left(\arcsin \frac{x}{2} + B \right) & \text{при } x < 1, \\ 1 & \text{при } x \geq 1 \end{cases}$	$x_1 = 0$	$x_2 = 1$
3.23	$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ A + B \sin x \cos x & \text{при } 0 < x < \frac{\pi}{4}, \\ 1 & \text{при } x > \frac{\pi}{4} \end{cases}$	$x_1 = 0$	$x_2 = \frac{\pi}{6}$

Окончание табл 3.1

№	Функция распределения	x_1	x_2
3.24	$F(x) = \begin{cases} 0 & \text{при } x \leq -2, \\ A + B \arcsin \frac{x}{2} & \text{при } x < 2, \\ 1 & \text{при } x \geq 2 \end{cases}$	$x_1 = -1$	$x_2 = 1$
3.25	$F(x) = \begin{cases} 0 & \text{при } x \leq -\frac{\pi}{4}, \\ A(B + \sin 2x) & \text{при } x < \frac{\pi}{4}, \\ 1 & \text{при } x \geq \frac{\pi}{4} \end{cases}$	$x_1 = 0$	$x_2 = \frac{\pi}{4}$

Задача № 4

В вариантах 4.1–4.25 СВ X задана плотностью распределения. Найти:

- значение коэффициента A ,
- функцию распределения $F(x)$,
- вероятность того, что СВ X примет значение в интервале (x_1, x_2) ,
- вероятность того, что СВ X в n независимых испытаниях, проводимых в одинаковых условиях, ни разу не попадет в интервал (x_1, x_2) ,
- математическое ожидание и среднеквадратичное отклонение СВ X ,
- построить графики $F(x)$ и $f(x)$.

Таблица 3.2.

№	Плотность распределения	x_1	x_2	n
4.1	$f(x) = \begin{cases} 0 & \text{при } x \leq 0 \text{ и } x \geq \pi, \\ A \sin x & \text{при } 0 < x < \pi \end{cases}$	$x_1 = 0$	$x_2 = 2$	$n = 2$
4.2	$f(x) = \frac{A}{(1+x^2)}$	$x_1 = -1$	$x_2 = 1$	$n = 2$
4.3	$f(x) = \begin{cases} Ax^2 + Bx + 6 & \text{при } 2 \leq x \leq 4, \\ 0 & \text{при } x > 2 \text{ или } x < 4 \end{cases}$	$x_1 = 1$	$x_2 = 3$	$n = 4$
4.4	$f(x) = \begin{cases} 0 & \text{при } x < 0 \text{ и } x \geq 4, \\ \frac{2}{A} \left(1 - \frac{x}{A}\right) & \text{при } 0 \leq x \leq 4 \end{cases}$	$x_1 = -1$	$x_2 = 0$	$n = 2$
4.5	$f(x) = \begin{cases} 0 & \text{при } x \leq -1 \text{ и } x \geq 1, \\ A(1-x^2)^{0,5} & \text{при } -1 < x < 1 \end{cases}$	$x_1 = -0,5$	$x_2 = 0,5$	$n = 2$
4.6	$f(x) = \begin{cases} A(2-x) & \text{при } 0 \leq x \leq 1, \\ 0 & \text{при } x < 0 \text{ и } x > 1 \end{cases}$	$x_1 = 0$	$x_2 = \frac{1}{2}$	$n = 2$
4.7	$f(x) = \begin{cases} \frac{A}{2} \sqrt{4-x^2} & \text{при } -2 < x < 2, \\ 0 & \text{при } x \leq -2 \text{ и } x \geq 2 \end{cases}$	$x_1 = -1$	$x_2 = 1$	$n = 2$
4.8	$f(x) = \begin{cases} Ax^{-4} & \text{при } x \geq 1, \\ 0 & \text{при } x < 1 \end{cases}$	$x_1 = 2$	$x_2 = 4$	$n = 4$

№	Плотность распределения	x_1	x_2	n
4.9	$f(x) = \begin{cases} A \cos 2x & \text{при } 0 < x < \frac{\pi}{4}, \\ 0 & \text{при } x \leq 0 \text{ и } x \geq \frac{\pi}{4} \end{cases}$	$x_1 = 0$	$x_2 = \frac{\pi}{6}$	$n = 2$
4.10	$f(x) = \begin{cases} Ax(x+2) & \text{при } 0 < x \leq 1, \\ 0 & \text{при } x \leq 0 \text{ и } x > 1 \end{cases}$	$x_1 = \frac{2}{3}$	$x_2 = 1$	$n = 2$
4.11	$f(x) = \begin{cases} \frac{A}{\sqrt{4-x^2}} & \text{при } -2 \leq x \leq 2, \\ 0 & \text{при } x < -2 \text{ и } x > 2 \end{cases}$	$x_1 = -1$	$x_2 = 1$	$n = 2$
4.12	$f(x) = \begin{cases} A \cos x & \text{при } -\frac{\pi}{2} < x < \frac{\pi}{2}, \\ 0 & \text{при } x \leq -\frac{\pi}{2} \text{ и } x \geq \frac{\pi}{2} \end{cases}$	$x_1 = -\frac{\pi}{4}$	$x_2 = \frac{\pi}{4}$	$n = 2$
4.13	$f(x) = \begin{cases} A \cos^2 x & \text{при } -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}, \\ 0 & \text{при } x < -\frac{\pi}{2} \text{ и } x > \frac{\pi}{2} \end{cases}$	$x_1 = 0$	$x_2 = \frac{\pi}{4}$	$n = 2$
4.14	$f(x) = \begin{cases} A(4-x) & \text{при } 0 \leq x \leq 4, \\ 0 & \text{при } x < 0 \text{ и } x > 4 \end{cases}$	$x_1 = 2$	$x_2 = 4$	$n = 4$
4.15	$f(x) = \begin{cases} A\left(1 - \frac{ x }{2}\right) & \text{при } -2 < x < 2, \\ 0 & \text{при } x \leq -2 \text{ и } x \geq 2 \end{cases}$	$x_1 = -1$	$x_2 = 2$	$n = 2$
4.16	$f(x) = Ae^{- x-2 } \text{ при } -\infty < x < \infty$	$x_1 = 1$	$x_2 = 3$	$n = 2$
4.17	$f(x) = \begin{cases} 0 & \text{при } x \leq 0 \text{ и } x \geq \pi, \\ A \sin x & \text{при } 0 < x < \pi \end{cases}$	$x_1 = \frac{\pi}{3}$	$x_2 = \frac{\pi}{2}$	$n = 4$
4.18	$f(x) = \begin{cases} A \cos x & \text{при } x < \frac{\pi}{4}, \\ 0 & \text{при } x > \frac{\pi}{4} \end{cases}$	$x_1 = -\frac{\pi}{6}$	$x_2 = \frac{\pi}{6}$	$n = 2$
4.19	$f(x) = \begin{cases} A(x-1) & \text{при } 1 \leq x \leq 3, \\ 0 & \text{при } x < 1 \text{ и } x > 3 \end{cases}$	$x_1 = 1,5$	$x_2 = 2$	$n = 2$
4.20	$f(x) = \begin{cases} Ax^{-4} & \text{при } x \geq 2, \\ 0 & \text{при } x < 2 \end{cases}$	$x_1 = 2$	$x_2 = 3$	$n = 2$
4.21	$f(x) = \begin{cases} Ax^2 & \text{при } 0 < x < 3, \\ 0 & \text{при } x \leq 0 \text{ и } x > 3 \end{cases}$	$x_1 = 1$	$x_2 = 2$	$n = 2$
4.22	$f(x) = \begin{cases} A(x+1) & \text{при } -1 < x \leq 0, \\ A(1-x) & \text{при } 0 < x \leq 1, \\ 0 & \text{при } x \leq -1 \text{ и } x > 1 \end{cases}$	$x_1 = 0,5$	$x_2 = 0,5$	$n = 2$
4.23	$f(x) = \begin{cases} Ax^{\frac{5}{2}} & \text{при } x \geq 1, \\ 0 & \text{при } x < 1 \end{cases}$	$x_1 = 1$	$x_2 = 4$	$n = 4$
4.24	$f(x) = \begin{cases} 0 & \text{при } x \leq -1, \\ x^2 & \text{при } -1 \leq x < 0, \\ Ax & \text{при } 0 \leq x < 1, \\ 0 & \text{при } x \geq 1 \end{cases}$	$x_1 = 0,5$	$x_2 = 0,5$	$n = 3$
4.25	$f(x) = \begin{cases} 0 & \text{при } x \leq 1 \text{ и } x > 16, \\ Ax^{-\frac{3}{2}} & \text{при } 1 < x < 16 \end{cases}$	$x_1 = 4$	$x_2 = 9$	$n = 4$

3.2. ВАРИАНТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Задание 1: «Статистический анализ одномерной случайной величины»

Вариант 1. Города России с численностью от 100 до 500 тыс. чел. (перепись 2010 г.)

473	453	431	415	414	409	408	404	403	398	361	356	348	345	343
333	326	325	323	317	314	312	311	307	306	301	297	280	273	271
269	268	263	263	257	251	248	246	241	240	240	240	239	235	234
233	221	218	215	214	210	210	207	203	202	200	199	192	188	188
183	181	179	178	175	175	174	173	172	170	169	165	159	158	156
156	155	151	147	146	145	144	144	142	139	137	121	133	131	131
128	126	124	124	123	122	121	120	120	119	118	116	116	111	111
110	110	110	109	109	109	108	108	108	106	104	104	103	103	102
102	102	102	101	101	101	100								

Вариант 2. Средняя продолжительность жизни населения в развивающихся странах, данные на 2010 г.

76,76	72,62	75,28	75,46	76,41	74,89	65,96	53,01	72,76	77,80
73,94	74,61	61,85	69,20	70,58	72,53	72,96	72,73	68,47	73,15
72,49	73,21	56,56	79,07	72,23	74,44	73,79	71,59	76,47	61,62
74,45	73,42	76,11	76,17	75,30	72,59	71,52	73,77	70,07	69,54
72,07	73,54	74,15							

Вариант 3. Данные о ВВП на душу населения (\$) в развивающихся странах в 2010 г. (с доходом не выше 4000 \$).

620	2180	1020	380	600	390	1490	950	840	330
710	1850	5120	520	400	1090	490	1140	710	410
1220	3880	1310	420	520	970	1350	1270	820	860
1810	460	290	1080	1250	1340	690	1760	900	980
1380	1280	1400	1020	920	900	460	740	2130	

Вариант 4. Данные о росте студенток 1 курса ФЭУ ВолгГТУ (см)

165	156	165	165	175	174	163	172	166	168
165	168	168	170	170	170	167	167	170	164
169	168	165	168	167	167	175	168	168	170
171	162	165	159	164	167	168	165	165	167
163	163	165	165	167	170	166	168	158	168
160	165	167	167	168	168	160	155	157	160
169	169	168	163	165	163	165	169	168	170
163	168	157	170	166	163	166	157	158	163
173	165	170	174	171	159	167	164	164	168
162	167	164	168	172	169	167	170	156	162

Вариант 5. Данные о месяце рождения студентов 1 курса факультета ФЭУ ВолгГТУ

8	10	6	5	4	6	8	6	6	9
8	7	11	10	10	1	1	4	2	11
9	11	1	4	12	4	3	3	7	5
9	11	7	10	1	12	4	10	9	12
1	9	11	8	3	7	5	11	7	12
1	12	4	8	12	5	1	12	9	1
10	3	6	9	3	5	1	6	9	4
9	11	8	8	12	6	7	1	11	9
7	8	5	9	2	5	12	7	8	12
6	3	8	6	12	7	3	10	3	3

Вариант 6. Данные о продолжительности в часах горения электроламп.

741	815	837	680	737	821	892	842	794	819
726	824	834	715	894	832	757	809	764	870
824	769	824	752	787	835	650	837	820	731
800	830	710	893	785	847	752	801	779	862
813	805	770	185	742	806	919	783	847	793
771	782	794	842	780	763	798	764	758	852
780	762	776	743	792	692	804	810	841	777
844	720	804	764	789	752	840	754	879	865
765	810	90	729	819	722	842	950	815	831
799	862	755	805	838	883	837	772	870	857

Вариант 7. Данные о промежутках времени между поступлением заявок в одно из подразделений службы быта в часах представлены в таблице (округлены до 0,1 часа).

4,2	16,1	9,0	13,0	11,3	8,2	5,3	18,0	1,7	7,5
5,3	7,8	8,5	0,5	2,0	0,1	1,2	1,7	3,0	6,4
17,2	13,4	1,2	1,8	8,7	1,4	7,5	9,0	3,2	5,8
15,6	11,2	3,0	7,2	11,8	2,5	5,2	2,3	7,6	3,4
5,9	3,1	0,2	4,2	2,4	3,4	13,0	0,1	10,1	4,2
2,2	7,5	3,5	5,7	8,3	1,8	7,8	3,4	5,5	1,4
4,8	3,8	1,8	1,4	1,7	4,0	20,0	7,5	11,6	7,5
2,2	9,1	1,4	3,2	2,0	1,9	19,0	6,0	3,1	5,2
3,5	0,9	3,5	4,4	1,2	2,5	5,2	3,6	3,8	6,4
0,6	5,1	1,8	3,2	3,2	12,0	6,5	1,6	1,8	7,4

Вариант 8. Результаты испытания предела прочности партии стальной проволоки приведены в таблице.

167	147	169	187	151	161	156	135	177	198
176	189	163	145	162	164	142	158	177	178
157	179	188	167	154	178	152	168	153	188
186	168	158	163	197	162	159	164	178	154
161	165	164	201	161	157	179	163	177	153
189	178	159	199	157	154	188	149	169	164
161	177	167	155	162	186	166	203	156	175
155	179	189	165	160	151	163	176	185	158
163	164	153	163	159	162	187	168	158	164
164	179	209	180	179	187	166	158	164	152

Вариант 9. Результаты химического анализа проб угля для определения содержания золы в процентах приведены в таблице.

8	14	14	20	11	10	12	14	12	24	13	16	6	19	17	3	17	7
16	14	14	17	7	14	14	13	14	15	21	16	17	13	10	13	20	17
15	13	16	17	20	9	18	11	10	8	13	15	14	8	17	14	13	19
17	18	23	14	19	21	14	19	20	14	9	15	16	21	11	14	13	19
17	11	18	22	16	19	12	14	27	13	12	16	20	10	16	16	10	11
13	11	14	15	16	23	17	12	18	15								

Вариант 10. Данные об ошибке при округлении с избытком представлены в таблице в условных единицах. Сами данные округлены с точностью до одной десятой единицы.

1,0	0,2	0,2	0,5	0,1	0,5	0,4	0,1	0,4	0,2	0,6	0,9	0,3	1,0	0,1
0,7	0,7	0,3	0,8	0,7	0,8	0,9	0,7	0,7	0,6	0,4	0,1	0,5	0,5	0,5
0,2	0,3	0,4	0,3	0,4	0,3	0,4	0,3	0,1	0,6	0,6	0,9	0,5	0,8	0,1
0,3	0,7	0,8	0,4	1,0	1,0	0,4	0,5	0,5	0,6	0,8	0,8	0,7	0,6	0,2
0,1	0,2	0,7	0,5	0,1	0,7	1,0	0,9	1,0	1,0	0,3	0,9	0,8	0,6	0,9
0,9	0,9	0,4	0,8	0,2	0,6	0,9	0,1	0,2	0,5	0,7	0,3	0,7	0,6	0,9
0,3	0,4	0,5	0,6	0,4	0,4	0,5	0,6	0,3	0,8					

Вариант 11. Данные о дневной выручке магазина (в тыс. руб.) за 100 дней представлены в таблице.

201,5	187,0	196,5	193,5	198,9	175,0	197,2	219,0	214,0	224,4
200,7	201,9	217,5	211,2	201,5	203,8	218,7	197,2	208,3	191,0
182,5	202,7	192,5	207,0	183,9	208,5	204,0	213,3	209,0	219,0
227,9	197,0	189,0	203,2	194,8	204,5	223,5	198,8	192,4	212,8
208,5	199,3	214,0	216,0	206,0	190,5	205,5	198,0	213,8	207,1
196,5	202,0	213,4	202,1	212,7	218,0	213,1	199,5	203,1	200,0
201,5	216,5	208,4	196,0	209,0	222,0	214,5	206,6	195,5	207,8
208,4	213,1	209,0	221,0	213,0	209,9	212,5	188,5	208,8	216,4
209,0	191,5	202,0	202,5	202,0	216,5	209,3	214,0	229,0	207,4
191,0	208,7	195,1	235,0	199,4	211,0	187,5	202,9	215,9	203,5

Вариант 12. В таблице приведены данные о месячной зарплате рабочих одного из предприятий, в тыс. руб.

22,1	23,3	18,0	21,5	23,5	26,0	20,1	23,4	21,1	23,7	20,0	25,4
24,5	20,7	24,3	22,7	23,1	25,6	25,4	27,3	24,3	25,3	26,1	23,3
25,1	21,0	24,5	25,0	22,3	22,3	26,5	25,5	23,9	19,5	25,0	24,5
24,4	21,3	25,7	24,3	22,5	24,2	22,4	23,8	24,1	26,1	24,8	27,5
28,2	23,5	26,4	28,0	24,8	25,1	21,2	24,7	19,8	23,2	23,3	23,1
23,1	22,0	24,5	25,5	21,9	26,2	25,1	25,0	21,5	22,8	23,7	22,9
22,1	24,4	28,4	24,5	26,5	23,2	24,8	22,1	24,2	22,6	24,7	23,9
25,2											

Вариант 13. В таблице приведены данные о промежутках времени между поступлениями заказов на международном переговорном пункте.

21	83	32	1	2	100	41	3	1	33	11	3	22	8	71
9	77	12	1	3	55	42	29	87	4	52	12	50	2	20
19	34	25	4	14	3	7	18	43	42	24	1	2	110	26
12	92	51	13	2	25	11	3	45	5	64	5	16	7	22
75	5	24	1	47	13	3	23	2	37	14	35	35	25	27
3	15	54	13	44	12	3	15	67	10	39	22	1	13	2
28	3	12	62	17	12	53								

Вариант 14. Данные о процентном содержании меди в сплаве по результатам 100 проб приведены в таблице.

16,06	15,92	16,01	16,05	15,80	15,97	16,18	16,26	16,22	16,20
16,03	16,31	15,92	16,17	16,07	15,99	16,06	16,07	15,98	16,08
16,09	16,21	16,05	15,87	16,03	16,21	16,23	16,04	15,94	16,03
15,96	16,09	16,24	16,10	15,97	16,16	16,11	16,12	16,14	16,11
16,03	16,03	16,05	16,22	15,97	16,27	16,16	16,32	16,21	16,14
16,23	16,17	16,00	16,13	16,10	15,98	16,12	16,07	16,06	16,11
16,02	15,89	16,29	16,14	15,93	16,17	16,16	15,99	16,15	16,13
16,14	16,05	16,34	16,13	16,13	15,91	16,18	16,06	16,18	16,07
16,09	16,15	16,02	16,08	16,40	16,03	16,16	15,89	16,04	16,15
16,22	16,08	16,09	16,17	16,10	16,11	16,12	16,08	16,13	16,10

Вариант 15. В таблице приведены результаты испытания крепости в граммах нитей.

325	341	285	302	275	284	281	295	220	303	330	286
248	286	247	347	307	286	285	319	299	293	310	331
261	304	337	305	318	270	285	317	247	301	277	281
263	280	333	359	325	318	280	271	318	324	287	272
278	337	348	305	265	287	328	317	232	255	308	275
270	301	317	325	305	333	268	319	274	339	324	355
291	350	307	290	308	259	315	273	308	330	315	273
293	272	292	321	291	297	380	312	325	296	263	305
328	307	295	271								

Вариант 16. Данные о длине заготовок после их первоначальной обработки приведены в таблице в миллиметрах.

1151	1158	1152	1155	1160	1151	1154	1156	1160	1151	1153	1155
1154	1156	1151	1156	1151	1154	1153	1157	1154	1154	1152	1154
1155	1152	1153	1156	1157	1155	1155	1153	1157	1158	1156	1158
1159	1156	1159	1156	1160	1153	1152	1156	1151	1157	1154	1158
1158	1160	1154	1159	1153	1157	1158	1157	1159	1155	1159	1158
1153	1151	1152	1154	1160	1155	1151	1159	1155	1158	1152	1153
1159	1155	1160	1158	1159	1152	1157	1156	1160	1151	1157	1154
1155	1157	1160	1152	1159	1159	1153	1159	1154	1158	1160	1158
1157	1156	1151	1160								

Вариант 17. Данные о яйценоскости 100 кур-несушек приведены в таблице в штуках.

200	210	223	217	201	220	215	220	213	232	211	203
222	218	238	215	229	205	224	216	223	207	193	195
205	230	170	185	207	209	221	214	198	227	210	213
225	224	237	204	187	192	211	184	197	214	203	208
216	199	217	222	209	221	219	190	205	192	215	183
208	209	194	211	197	201	195	214	213	224	212	208
235	195	207	211	193	202	201	238	215	227	207	218
210	194	205	188	195	204	198	203	199	213	198	225
250	195	212	226								

Вариант 18 Данные о посещаемости библиотеки за 100 дней приведены в таблице.

254	242	237	265	277	257	269	222	247	271	278	252	256	255	263
270	282	257	253	250	272	244	259	258	274	266	238	265	262	245
253	260	261	246	263	265	248	269	280	262	267	243	273	257	244
264	274	236	256	245	255	287	267	253	268	258	275	264	255	247
265	275	255	235	276	263	286	256	264	269	277	259	254	240	257
244	265	291	257	247	267	279	265	264	253	255	281	257	263	302
263	281	280	255	250	264	254	265	274	254					

Вариант 19. Данные о промежутках времени (мин.) между двумя очередными заявками на вызов скорой помощи приведены в таблице.

15	1	50	2	11	1	37	1	31	2	1	38	7	3	3	2	21	6	7	17
19	3	4	32	2	5	33	8	4	5	42	5	26	35	1	9	22	6	3	16
15	11	1	20	2	28	11	9	23	12	16	3	5	30	12	5	10	13	3	2
19	8	6	7	10	1	1	24	13	49	10	25	12	29	14	4	6	35	2	6
5	31	7	8	17	9	6	7	18	1	14	4	7	3	18	5	6	1	5	2

Вариант 20. Данные о пробеге такси в км за 100 смен приведены в таблице.

304	313	344	324	353	325	342	333	317	334	325	306	347	326	346
337	329	312	340	350	320	317	339	333	343	354	324	349	325	326
319	337	339	316	348	323	357	341	347	330	346	343	314	337	349
335	324	302	333	315	339	361	337	316	344	323	324	345	326	337
318	360	323	334	350	325	346	332	305	327	345	340	320	332	319
329	333	307	325	334	359	307	327	347	336	362	325	317	337	328
310	334	325	339	330	337	323	336	321	335	312				

Вариант 21. Объем вкладов в тыс. рублей в отделении банка за 60 месяцев представлен в таблице.

44,00	49,32	48,53	51,82	45,24	51,13	52,35	53,92	50,43	46,07
48,64	49,24	52,60	53,17	51,32	48,99	46,58	51,57	52,48	47,02
52,00	51,82	46,13	50,37	52,74	53,20	51,95	47,18	49,48	48,31
52,10	51,93	46,25	50,24	48,43	49,32	51,01	48,72	49,25	48,65
50,13	49,18	50,10	49,07	56,00	50,93	49,88	47,50	54,32	50,34
50,50	47,35	50,74	49,85	47,70	49,92	50,94	48,31	49,08	47,64

Вариант 22. Данные о времени ожидания пассажиром поезда метро в секундах представлены в таблице.

11	69	101	68	1	108	2	82	62	3	115	14	79	95	15
23	80	115	24	26	27	110	29	31	33	40	42	49	46	64
25	107	83	103	63	4	116	11	93	91	61	16	86	17	18
62	48	107	71	64	7	12	118	6	92	5	73	19	31	92
41	77	65	8	93	9	100	13	98	70	21	20	108	34	35
41	47	99	78	44	85	45	84	112	120	81	43	107	110	38
51	50	119	95	67	37	109	22	36	39					

Вариант 23. В таблице приведены данные об убойном весе 100 бычков.

473	503	427	501	550	537	410	509	486	459	530	541	517	481	525
514	495	435	538	478	497	519	458	525	475	508	518	555	535	497
471	542	485	507	465	523	436	524	470	510	512	499	478	525	534
527	560	487	537	508	469	503	488	488	442	533	481	527	493	498
525	471	524	495	476	449	501	580	521	487	498	499	565	505	500
473	530	486	540	513	570	512	513	484	420	498	523	452	481	469
525	488	538	522	483	510	590	529	482	513					

Вариант 24. Результаты контроля величины смещения полушарий в микронах у шариков для подшипников представлены в таблице.

28	5	62	59	13	51	45	38	15	43	44	75	36	16	37	25	52	35	26	32
24	27	52	38	7	28	20	29	45	36	17	30	57	29	60	42	53	26	19	40
44	43	21	49	37	8	46	37	18	36	16	54	55	31	26	63	41	27	21	33
42	48	22	38	17	47	19	55	39	9	32	33	27	49	39	28	40	29	30	23
23	38	41	22	20	34	29	6	34	35	25	33	50	30	31	12	32	24	31	18

Вариант 25. Данные о промежутках времени между поступлениями заказов на такси в минутах (округлены с избытком с точностью до 1 мин)

7	6	1	8	1	2	4	1	2	4	3	6	1	1	2	3	4	3	3	1
7	5	1	9	3	8	4	3	1	4	1	2	6	1	1	4	7	3	1	10
2	4	4	3	2	6	2	2	7	2	5	2	2	7	3	4	1	6	10	2
3	3	5	2	1	1	5	2	1	3	1	10	2	5	1	3	1	2	4	3
1	2	1	4	2	1	1	3	1	1	2	3	8	1	2	4	5	1	1	2

Вариант 26. Отклонения от номинала предела текучести стали марки 35ГС (X , кг/мм²)

-3,99	-2,58	-1,98	-1,17	-0,66	-0,17	0,26	0,79	1,33	2,50
-3,29	-2,52	-1,90	-1,13	-0,63	-0,15	0,38	0,85	1,48	2,63
-4,76	-2,47	-1,84	-1,08	-0,58	-0,09	0,42	0,86	1,53	2,75
-3,37	-2,41	-1,74	-1,03	-0,53	-0,03	0,49	0,92	1,63	2,85
-4,35	-2,33	-1,62	-0,99	-0,47	0,00	0,50	0,95	1,77	2,93
-3,00	-2,28	-1,53	-0,93	-0,44	0,04	0,52	0,97	1,93	2,99
-2,92	-2,21	-1,47	-0,86	-0,37	0,08	0,57	1,00	2,02	3,09
-2,84	-2,14	-1,39	-0,83	-0,34	0,12	0,63	1,08	2,12	3,92
-2,79	-2,08	-1,31	-0,79	-0,25	0,17	0,68	1,14	2,21	4,55
-2,67	-2,03	-1,23	-0,72	-0,23	0,23	0,74	1,25	2,38	6,08

Вариант 27. Отклонения от номинала диаметра метрической резьбы на болтах (X , мкм).

-2,7	-2,5	-2,7	-3,2	-3,5	-3,2	-3,5	-3,2	-3,6	-3,7	-4,0	1,3	-1,3
-6,0	-5,8	-0,5	-5,5	5,9	6,0	5,2	5,0	-5,0	-5,3	4,9	4,7	4,5
4,3	-4,8	-4,5	-4,8	-4,3	-4,3	-9,9	-9,2	-8,3	-7,5	-7,0	-6,1	4,0
3,8	4,0	3,5	3,3	3,5	3,1	2,5	2,3	2,7	2,3	2,0	2,1	2,0
-3,8	11,0	-9,5	-8,7	-8,0	-7,5	-7,0	-6,5	0,0	0,1	0,0	0,2	0,4
0,5	0,4	0,7	0,8	1,0	0,8	1,1	1,3	1,4	1,5	1,7	1,8	-2,0
-2,1	-2,0	-1,8	-1,7	-1,6	-1,5	-1,4	-1,2	-1,0	-0,7	-1,0	-0,7	-0,5
-0,5	-0,4	-0,3	-0,2	-0,2	-0,1	-0,1	-2,3	-2,5				

Вариант 28. Содержание меди в мельхиоровых отливках (X , %)

66,0	66,1	66,0	66,1	66,3	70,0	70,1	70,0	70,3	70,0
70,4	64,3	68,0	63,1	68,0	68,1	68,3	68,4	68,3	68,0
72,0	72,2	72,4	66,5	66,5	66,7	65,7	70,5	72,0	70,6
70,5	70,8	70,9	64,5	64,7	64,8	60,5	68,6	68,5	68,7
68,5	68,8	68,9	68,8	72,5	72,7	67,0	67,1	67,0	67,2
71,0	71,1	71,0	71,2	71,0	71,3	65,0	65,2	67,0	67,1
67,0	67,2	67,0	67,3	67,4	67,3	73,0	73,2	67,5	67,6
67,5	67,7	67,8	67,7	67,9	71,5	71,7	71,5	71,8	71,9
65,5	65,5	69,5	69,6	69,5	69,7	69,5	69,6	69,6	69,9
69,8	69,9	69,8	73,5	73,8	66,0	66,8	66,5	66,9	67,3

Вариант 29. В таблице представлены данные о годовом удое молока в литрах от ста дойных коров.

3530	3480	3115	3417	3540	3957	3550	3315	3620	3840
3200	3719	3780	3167	3418	3570	3275	3580	3470	3620
3510	3227	3680	3820	3725	3375	3492	3520	3790	3695
3457	3675	3810	3455	3515	3258	3525	3769	3382	3530
3880	3760	3910	3770	3650	4050	3660	3980	3590	3670
3575	3393	3565	3860	3410	4025	3720	3630	3831	3610
3620	3430	3625	3324	3630	3857	3564	3610	3819	3670
3680	3351	3770	3755	3434	3515	3883	3635	3615	3714
3450	3660	3315	3524	3515	4200	3670	3464	3290	3520
3780	3312	3620	3730	3850	3780	3814	3452	3520	3630

Вариант 30. Данные об урожайности пшеницы в ц\га.

15,2	16,1	23,1	20,4	21,3	22,2	18,3	23,1	23,9	17,4
26,1	17,0	25,4	20,8	26,9	28,3	25,2	22,5	18,1	18,2
21,2	19,0	25,9	20,3	24,5	21,8	27,9	15,7	26,7	19,2
24,1	21,2	22,7	20,6	24,0	21,1	23,3	20,4	24,2	21,2
17,3	22,3	20,1	19,3	24,4	28,1	22,6	22,7	18,2	21,5
19,2	23,3	17,3	23,4	21,5	22,7	17,5	24,0	19,1	21,2
19,4	28,5	16,2	21,3	22,5	19,3	23,7	20,3	22,1	19,9
24,5	23,9	29,5	24,2	20,1	24,7	23,8	19,4	20,2	21,3
22,3	24,2	18,4	23,4	19,5	26,6	22,7	24,8	17,2	25,9
19,7	23,1	21,4	22,5	25,2	21,3	26,4	19,2	22,5	22,1

Задание 2. «Исследование зависимости случайных величин»

В вариантах 1–9 проанализировать динамику изменения экономических показателей, характеризующих состояние Российской экономики за год. В вариантах 10–30 проанализировать взаимосвязь между показателями

Таблица 3.3

№	Переменные величины	Номера наблюдений												X
		1	2	3	4	5	6	7	8	9	10	11	12	
1	Месяц (2011г.), х	1	2	3	4	5	6	7	8	9	10	11	–	15
	Инвестиции в основной капитал, у, % от января 2011г.	97,9	99,6	99,7	102,2	107,4	104,9	107,9	106,5	108,5	108,6	108,0	–	
2	Месяц (2011г.), х	1	2	3	4	5	6	7	8	9	10	11	12	15
	Объем промышленного производства, у, % от января 2011г.	110,2	108,4	109,8	110,4	112,6	109,7	105,9	107,0	106,2	104,9	106,9	107,1	

№	Переменные величины	Номера наблюдений												X
		1	2	3	4	5	6	7	8	9	10	11	12	
3	Месяц(2010г), х	1	2	3	4	5	6	7	8	9	10	11	12	15
	Оборот розничной торговли, у, % от января 2010г.	103,6	101,1	111,2	112,4	115,8	117,4	121,5	124,0	122,9	127,7	128,1	155,4	
4	Месяц(2011г.), х	1	2	3	4	5	6	7	8	9	10	11	12	15
	Экспорт, у, млн долл.	31,0	38,9	43,1	45,6	44,1	43,6	41,8	44,2	43,3	45,7	47,2	50,2	
5	Месяц (2011г.), х	1	2	3	4	5	6	7	8	9	10	11	12	15
	Импорт, у, млн долл.	14,7	20,5	25,8	25,9	26,9	26,2	26,0	28,8	26,1	27,8	27,9	28,4	
6	Месяц (2011г.), х	1	2	3	4	5	6	7	8	9	10	11	12	15
	Индекс потребительских цен, у, % к предыдущему месяцу	100,5	100,9	100,6	100,3	100,5	100,4	100,4	100,6	100,8	100,5	100,8	100,9	
7	Месяц (2011г.), х	1	2	3	4	5	6	7	8	9	10	11	12	15
	Индекс цен производителей, у, % к предыдущему месяцу	102,1	103,3	101,4	102,0	101,1	97,7	99,8	104,6	99,3	101,7	101,6	101,8	
8	Месяц (2011г.), х	1	2	3	4	5	6	7	8	9	10	11	12	15
	Реальные денежные доходы, у, % к январю 2010г	124,2	100,5	104,1	103,2	107,6	104,6	108,8	96,0	123,4	107,4	110,1	113,0	
9	Месяц (2011г), х	1	2	3	4	5	6	7	8	9	10	11	12	15
	Реальная заработная плата, у, % к январю 2010г	109,3	109,8	113,6	111,4	112,2	108,4	110,1	110,4	111,0	111,9	112,5	115,1	
10	Температура, х, °С	-10	-10	-10	-5	0	0	0	5	5	10	10	10	5
	Растворимость закиси азота в двуокиси азота, у, % от веса	1,28	1,33	1,52	1,21	1,27	1,11	1,04	0,82	0,82	0,65	0,54	0,63	
11	Температура рабочего конца, х, °С	50	100	150	200	250	300	350	400	450	500	550	600	300
	ТермоЭДС хромель-копелевой термопары, у, мВ	1,5	6,8	10,7	14,5	18,6	22,8	27,0	31,6	35,8	40,0	44,4	48,7	
12	Температура стержня, х, °С	20	50	100	150	200	250	300	350	400	450	500	750	200
	Предел прочности на разрыв дюралевых стержней, у, кг/мм	37	36	35	34	33	32	29	25	24	23	22	12	

№	Переменные величины	Номера наблюдений												X
		1	2	3	4	5	6	7	8	9	10	11	12	
13	Содержание хлорида кальция CaCl_2 в воде, x , %	0	1	3	5	6	8	10	12	13	15	17	20	10
	Растворимость хлорида бария BaCl_2 при 70°C , y , % от веса	32	29	26	25	24	30	17	17	14	11	8	5	
14	Масса подвешенного груза, x , г	50	70	100	125	150	175	200	225	250	275	300	325	200
	Растяжение мерной ленты, y , мм	0,05	0,052	0,120	0,160	0,170	0,23	0,27	0,33	0,34	0,43	0,46	0,48	
15	Температура стержня, x , $^\circ\text{C}$	20	50	100	150	200	250	300	350	400	450	500	750	300
	Предел прочности на разрыв медных стержней, y , кг/мм	21,0	20,0	18,8	18,6	16,8	15,2	14,6	14,0	13,6	12,1	10,0	6,0	
16	Температура, x , $^\circ\text{C}$	0	5	10	15	20	25	30	35	40	45	50	–	20
	ЭДС гальванического элемента (катод – кадмий, электр. – ртуть), y , В	1,0191	1,0188	1,0187	1,0186	1,0183	1,0181	1,0180	1,0177	1,0174	1,0173	1,0171	–	
17	Температура горячего спая, x , $^\circ\text{C}$	200	250	300	350	400	450	500	550	600	650	700	–	400
	ТермоЭДС радиевой платиновой термопары, y , мВ	1,44	2,21	2,33	2,91	3,50	4,02	4,10	4,72	5,63	5,70	6,21	–	
18	Температура, x , $^\circ\text{C}$	20	50	100	150	200	250	300	350	400	450	500	750	300
	Предел текучести образцов дюралю, y , кг/мм	24,0	23,6	23,0	22,0	21,3	20,0	28,4	17,6	16,8	16,0	14,7	8,6	
19	Температура, x , $^\circ\text{C}$	20	50	100	150	200	250	300	350	400	450	500	750	300
	Предел текучести образцов стали, y , кг/мм	22,0	21,5	21,0	19,8	19,1	18,2	17,5	16,8	16,5	16,0	15,1	12,0	
20	Температура горячего спая, x , $^\circ\text{C}$	200	250	300	350	400	450	500	550	600	650	–	–	400
	ТермоЭДС радиовой платиновой термопары, y , мВ	0,12	0,20	0,30	0,60	0,57	0,90	0,95	1,00	1,40	1,66	–	–	

№	Переменные величины	Номера наблюдений												X
		1	2	3	4	5	6	7	8	9	10	11	12	
21	Скорость резания стали 2Х17Н2, х, м/мин	8,5	8,5	8,5	9,0	9,0	9,0	9,5	9,5	9,5	10,0	10,0	10,0	9,0
	Срок службы вставных резцов из твёрдого сплава, у, °С	2402	2254	1926	1978	1561	1380	1389	1290	1024	902	781	546	
22	Температура горячего спая, х, °С	100	150	200	250	300	350	400	450	500	550	600	650	400
	ТермоЭДС никелевой хромоникелевой пары, у, мВ	4,07	5,82	8,12	10,72	12,22	15,80	16,32	19,40	20,62	21,98	24,87	27,42	
23	Температура в автоклаве, х, °С	120	123	125	127	130	136	138	139	142	145	146	148	130
	Усадка синтетической нити после термообработки, у, %	3,41	3,85	4,03	4,52	4,44	5,10	5,72	5,64	5,82	6,44	6,57	6,45	
24	Предел текучести различных марок стали, х, (кг/мм)	30	40	50	60	70	80	90	100	110	120	130	140	100
	Предел прочности, у, (кг/мм)	44	50	63	75	80	90	98	114	120	127	140	153	
25	Стоимость основных фондов регионов России, 2009 г. х, млрд руб.	263	300	350	436	492	529	586	664	1116	1331	—	—	500
	ВРП регионов России, у, млрд руб.	117	87	126	161	214	132	304	197	377	556	—	—	
26	Температура горячего спая, х, °С	100	150	200	250	300	350	400	450	500	—	—	—	300
	ТермоЭДС константно-медной термопары, у, мВ	4,1	6,9	8,8	11,0	14,1	18,0	19,9	22,1	26,0	—	—	—	
27	Чистота алюмосиликат катализатора при крекинге нефти, х, %	98,8	98,9	99,0	99,1	99,2	99,3	99,4	99,5	99,6	99,7	99,8	—	99
	Октановое число бензина, у, %	83,9	83,6	83,4	84,3	83,5	83,8	84,2	84,8	84,2	83,8	84,9	—	

№	Переменные величины	Номера наблюдений												X
		1	2	3	4	5	6	7	8	9	10	11	12	
28	Температура горячего спая, x , °C	200	250	300	350	400	450	500	550	600	650	700	–	400
	ТермоЭДС константа хромоникелевой термодпары, y , мВ	11,08	15,50	19,09	22,00	26,48	31,60	34,18	37,44	41,95	46,82	50,02	–	
29	Предел прочности алюминиевых сплавов, x , кг/мм	32	34	36	38	40	42	44	46	48	50	52	–	40
	Предел выносливости, y , кг/мм	13,5	14,4	14,5	15,0	16,0	16,4	16,1	17,0	18,4	18,1	18,9	–	
30	Предел прочности образцов стали, x , кг/мм	50	60	70	80	90	100	110	120	130	140	150	160	100
	Предел выносливости, y , кг/мм	29,8	30,6	33,8	36,9	38,0	42,7	45,0	46,8	48,2	52,3	53,0	55,8	

БИБЛИОГРАФИЧЕСКИЙ СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

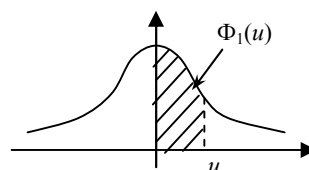
1. *Боровков, А. А.* Теория вероятностей и математическая статистика / А. А. Боровков. – М : URSS, 2009. – 652 с.
2. *Гмурман, В. Е.* Теория вероятностей и математическая статистика / В. Е. Гмурман. – 12-е изд. – М. : Высшее образование, 2007. – 470 с.
3. *Данко, П. Е.* Высшая математика в упражнениях и задачах. В 2 т. Т. 2 / П. Е. Данко, А. Г. Попов, Т. Я. Кожевникова. – 6-е изд. – М. : ОНИКС 21 век, 2006. – 416 с.
4. *Доугерти, К.* Введение в эконометрику / К. Доугерти. – М. : Инфра-М. – 2004. – 420 с
5. Кремер. Н. Ш. Теория вероятностей и математическая статистика / Н. Ш. Кремер. – 3-е изд. – М. : ЮНИТИ, 2002.
6. Российская экономика. Прогнозы и тенденции. – М. : Изд. дом ГУ ВШЭ. 2012. – № 4
7. *Симонова, И. Э.* Основы эконометрики : учеб. пособие (гриф УМО) / И. Э. Симонова, В. Ф. Казак, Б. В. Симонов; ВолгГТУ. – Волгоград, 2008. – 78 с.
8. *Чистяков, В. П.* Курс теории вероятностей : учебник для вузов [электронный ресурс] / В. П. Чистяков. – М. : Дрофа, 2007 – 245 с.

ПРИЛОЖЕНИЯ

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

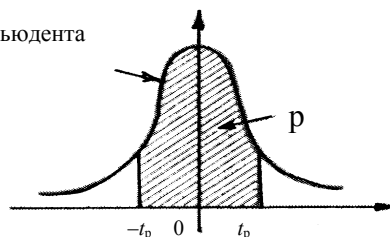
Таблица П.А.1

Функция Лапласа $\Phi_1(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-x^2/2} dx$



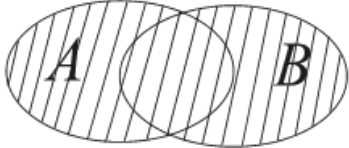
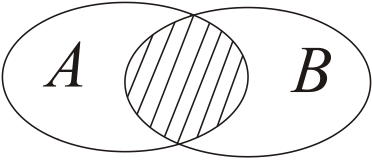
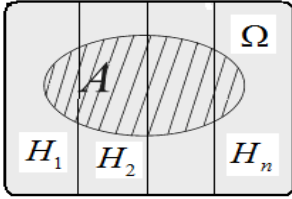
u	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0476	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0967	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1403	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2550	2611	2642	2673	2703	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	3413	3437	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3859	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4275	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4606	4616	4625	4630
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	47725	47778	47831	47872	47932	47961	48030	48077	46134	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	18537	48574
2,2	48610	48645	48679	48713	43746	48778	48809	48840	43870	48899
2,3	48928	48956	48963	49010	49036	49061	49086	49111	49135	49158
2,4	49180	492.2	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	48192	49506	40520
2,6	49534	49547	49560	49573	49585	495977	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49735	49801	40807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3	0,49865	49903	49931	49952	49966	49977	49984	49989	49993	49995
4,0	0,499968									
4,5	0,499997									
5,0	0,49999997									

Плотность распределения Стьюдента

Значения $t_p(k)$, определяемые уравнением $P = (|T(k)| < t_p(k)) = p$

$\begin{smallmatrix} p \\ k \end{smallmatrix}$	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,925	31,598
3	765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	741	941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	727	920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	718	906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	711	896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	706	889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	703	883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	700	879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	697	876	1,088	1,363	1,796	2,201	2,718	3,106	4,487
12	695	873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	694	870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	692	868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	691	866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	690	865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	689	863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	688	862	1,067	1,330	1,734	2,103	2,552	2,878	3,922
19	688	861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	687	860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	686	859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	686	858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	685	858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	685	857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	684	856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	064	856	1,056	1,315	1,706	2,056	2,479	2,779	3,707
27	684	855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	683	855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	683	854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	683	854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	681	851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	679	848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	677	845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	674	842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Случайные события

Сумма событий $A + B$ или	Произведение событий $A \cdot B$ и	Формулы полной вероятности Байеса	Схема Бернулли
<p>1. Для несовместных событий $P(A + B) = P(A) + P(B)$</p> <p>2. Для любых событий</p>  <p>$P(A + B) = P(A) + P(B) - P(AB)$</p> <p>3. $A = \{\text{хотя бы одно из } A_1, \dots, A_n\}$ \bar{A} - {ни одного} $P(A) = 1 - P(\bar{A}) = 1 - P(\bar{A}_1 \dots \bar{A}_n)$ Для независимых событий $P(A) = 1 - P(\bar{A}_1) \dots P(\bar{A}_n)$</p>	 <p>1. Условная вероятность $P(A B) = \frac{P(AB)}{P(B)}$</p> <p>2. А и В независимы, если $P(B A) = P(B)$</p> <p>3. Для независимых событий $P(AB) = P(A)P(B)$</p> <p>4. $P(A_1 A_2 \dots A_n) = P(A_1)P(A_2 A_1) \times$ $\times P(A_3 A_1 A_2) \dots P(A_n A_1 A_2 \dots A_{n-1})$</p>	<p>1. H_1, H_2, \dots, H_n – гипотезы (попарно несовместны, A – только с одним из H_i)</p>  <p>$A = AH_1 + AH_2 + \dots + AH_n = \sum_{i=1}^n AH_i$</p> <p>2. $P(A) = \sum_{i=1}^n P(H_i)P(A H_i)$ - формула полной вероятности</p> <p>3. $P(H_i A) = \frac{P(H_i)P(A H_i)}{\sum_{j=1}^n P(H_j)P(A H_j)}$ - формула Байеса</p>	<p>1. а) n – взаимно-независимых испытаний, б) в каждом из них – A или \bar{A}, в) $P(A) = p, P(\bar{A}) = 1 - p = q$</p> <p>2. X – число успехов в n испытаниях Бернулли $P(X = k) = C_n^k p^k q^{n-k}, C_n^k = \frac{n!}{k!(n-k)!}$</p> <p>3. При $n \geq 100, np < 0,1$ используют формулу Пуассона $P_n(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \lambda = np$</p> <p>4. При $npq \geq 10$ используют предельные формулы Муавра–Лапласа $P(X = k) \approx \frac{\varphi(x)}{\sqrt{npq}}, x = \frac{k - np}{\sqrt{npq}},$ $P(k_1 \leq X \leq k_2) \approx \Phi_1(x_2) - \Phi_1(x_1)$ $\Phi_1(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt, x_i = \frac{k_i - np}{\sqrt{npq}}, i = 1,$ $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \Phi_1(x) = \int_0^x \varphi(t) dt$</p>

ОГЛАВЛЕНИЕ

Введение.....	3
I. Основные теоретические положения и их применение к решению задач.....	4
1.1. Одномерная случайная величина, ее функция распределения.....	4
1.2. Дискретные и непрерывные случайные величины.....	5
1.3. Числовые характеристики случайных величин.....	11
1.4. Нормальное распределение и распределения, связанные с ним (распределения χ^2 , Стьюдента, Фишера).....	14
1.5. Генеральная и выборочная совокупность. Точечные оценки параметров распределения.....	17
1.6. Интервальные оценки параметров распределения.....	21
1.7. Проверка статистических гипотез: основные понятия.....	23
1.8. Две переменные: меры изменчивости и связи.....	25
1.9. Парный регрессионный анализ. Метод наименьших квадратов.....	27
1.10. Проверка гипотез, относящихся к уравнению регрессии.....	33
II. Методические материалы для самостоятельной работы студентов.....	36
2.1. Семестровая работа по теме «Статистический анализ одномерных случайных величин».....	36
2.2. Семестровая работа по теме «Статистическое исследование зависимости случайных величин».....	42
III. Задания к семестровым работам.....	49
3.1. Варианты индивидуальных заданий по теме «Случайные величины»...	49
3.2. Варианты экспериментальных данных.....	62
Библиографический список использованной литературы.....	74
Приложения.....	75

Учебное издание

Ирина Эдуардовна **Симонова**
Владимир Дмитриевич **Савельев**
Лиана Сергеевна **Сагателова**
Алексей Борисович **Симонов**

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА:
МАТЕРИАЛЫ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ**

Учебное пособие

Редактор *Л. И. Громова*

Темплан 2013 г. (учебники и учебные пособия). Поз. № 58.
Подписано в печать 30.08.2013 г. Формат 60×84 1/8. Бумага газетная.
Гарнитура Times. Печать офсетная. Усл. печ. л. 4,65. Уч.-изд. л. 4,60.
Тираж 100 экз. Заказ .

Волгоградский государственный технический университет.
400005, г. Волгоград, пр. Ленина, 28, корп. 1.

Отпечатано в типографии ИУНЛ ВолгГТУ
400005, г. Волгоград, пр. Ленина, 28, корп. 7.