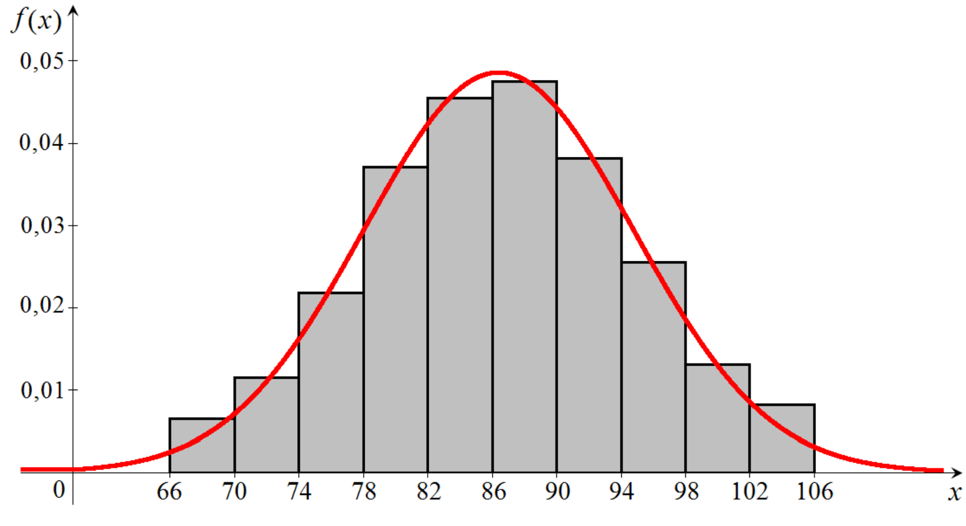
Д. К. Агишева, С. А. Зотова,
Т. А. Матвеева, В. Б. Светличная

Математическая статистика



ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
ВОЛЖСКИЙ ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ (ФИЛИАЛ)
ГОСУДАРСТВЕННОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ВОЛГОГРАДСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Д. К. Агишева, С. А. Зотова, Т. А. Матвеева, В. Б. Светличная

Математическая статистика

Учебное пособие

Допущено Научно-методическим советом по математике
Министерства образования и науки Российской Федерации
в качестве учебного пособия (справочника) для студентов техничес-
ких направлений и специальностей высших учебных заведений

Волгоград 2010

УДК 519.2

Рецензенты:

Кульков В. Г. – доктор физико-математических наук, профессор филиала ГОУВПО «Московский энергетический институт (технический университет)».

Тырымов А. А. – кандидат физико-математических наук, доцент кафедры «Прикладная математика» Волгоградского государственного технического университета.

Кобышев В. А. – кандидат физико-математических наук, доцент кафедры «Высшая математика» Волгоградского государственного технического университета.

Печатается по решению редакционно-издательского совета
Волгоградского государственного технического университета

Агишева, Д. К., Зотова, С. А., Матвеева, Т. А., Светличная, В. Б.

Математическая статистика: учеб. пособие /Д.К. Агишева, С.А. Зотова, Т.А. Матвеева, В.Б. Светличная; ВПИ (филиал) ВолгГТУ. – Волгоград, 2010. – 159 с.: ил.

ISBN 978-5-9948-0412-4

Учебное пособие содержит необходимый теоретический материал и большое количество примеров и задач, иллюстрирующих основные понятия по учебной дисциплине “Математическая статистика”. Разработаны варианты семестровых заданий.

Рассчитано на студентов различных форм обучения высших учебных заведений всех специальностей и направлений и лиц, использующих статистические методы при решении практических задач.

Табл. 16. Библиогр.: 12 названий

ISBN 978-5-9948-0412-4

ОГЛАВЛЕНИЕ

Введение 5
Глава 1. Элементы математической статистики 6
1.1. Предмет и задачи математической статистики 6
1.2. Генеральная совокупность и выборка 7
1.3. Статистическое распределение выборки 9
1.3.1. Дискретный случай 9
1.3.2. Непрерывный случай 11
1.4. Графическое изображение статистического ряда 14
1.5. Числовые характеристики выборки 19
1.5.1. Средние величины. Показатели вариации 19
1.5.2. Вычисление выборочных средней и дисперсии при больших (очень малых) значениях вариант 24
1.5.3. Эмпирические моменты 27
1.5.4. Асимметрия и эксцесс 27
1.6. Оценка неизвестных параметров 29
1.6.1. Понятие оценки параметров 29
1.6.2. Метод моментов 31
1.6.3. Метод наибольшего правдоподобия 33
1.6.4. Точечная оценка вероятности события 35
1.6.5. Точечная оценка математического ожидания 36
1.6.6. Точечная оценка дисперсии при неизвестном математическом ожидании 37
1.7. Интервальные оценки параметров распределения 39
1.7.1. Точность и надёжность оценок числовых характеристик 39
1.7.2. Односторонние доверительные интервалы 40
1.8. Некоторые распределения функций нормальных случайных величин 42
1.8.1. Распределение χ^2 (хи-квадрат) Пирсона 42
1.8.2. Распределение Стьюдента 43
1.8.3. Распределение Фишера-Снедекора 43
1.9. Доверительные интервалы параметров нормального распределения 44
1.9.1. Доверительный интервал для оценки вероятности события A 44
1.9.2. Доверительный интервал для оценки математического ожидания 45
1.9.3. Доверительный интервал для оценки среднего квадратического отклонения 48

1.10. Проверка статистических гипотез 49
1.10.1. Статистическая гипотеза 49
1.10.2. Задачи статистической проверки гипотез 50
1.10.3. Общая схема проверки статистических гипотез 51
1.10.4. Статистики сравнения точечных оценок неизвестных генеральных 54
1.11. Построение теоретического закона распределения случайной величины по опытным данным. Проверка гипотез о законе распределения 63
1.11.1. Критерий χ^2 Пирсона 65
Глава 2. Элементы теории корреляции 71
2.1. Понятие о корреляционной зависимости. Корреляционная таблица 71
2.2. Теснота корреляционной связи 75
2.3. Линейная регрессия 76
2.3.1. Проверка гипотезы о наличии линейной зависимости 76
2.3.2. Поиск уравнения линейной зависимости 77
2.3.3. Теснота линейной корреляционной связи 79
2.3.4. Показатели качества (адекватности) регрессии 80
2.4. Нелинейные корреляционные связи 81
2.4.1. Поиск уравнения параболической зависимости 81
2.4.2. Поиск уравнения гиперболической зависимости 81
Разбор типовых задач 86
Варианты семестровых заданий 108
Порядок отчётности. Вопросы к защите семестровой работы 148
Приложения 150
Библиографический список 158

Введение

Пособие «Математическая статистика» охватывает все основные разделы математической статистики и теории корреляции, встречающиеся при решении практических вопросов.

Предлагаемое пособие написано на основе лекций, прочитанных авторами в Волжском политехническом институте.

Руководство рассчитано на учащихся высших учебных заведений и может быть использовано как в процессе первоначального изучения математической статистики, так и для выработки практических навыков применения статистических методов исследования.

Цель предлагаемого пособия – помочь изучающим математическую статистику приобрести навыки применения её результатов к решению различных прикладных вопросов. Поэтому при подборе задач и методов их решения основное внимание было обращено не на формально математическую сторону статистики, а на её прикладное содержание и на умение решать конкретные задачи.

В каждой главе дана краткая сводка рабочих формул и схем, применение которых иллюстрируется решением типовых задач; составлены вопросы, по которым можно подготовиться к отчёту по теме «Математическая статистика». Преподаватель может использовать задачи для практических занятий и индивидуальных домашних заданий. Руководство содержит статистические таблицы, необходимые для решения задач.

При изложении материала авторы постарались сохранить сложившуюся терминологию и традиционные обозначения в формулировках задач и в их решениях.

Глава 1. Элементы математической статистики

1.1. Предмет и задачи математической статистики

Математической статистикой называется раздел математики, занимающийся разработкой методов получения, описания и обработки опытных данных с целью выявления и изучения закономерностей случайных массовых явлений для научных и практических выводов.

Математическая статистика наряду с теорией вероятностей изучает массовые случайные явления. Связующим звеном между ними выступают предельные теоремы теории вероятностей. При этом свойства реального процесса теория вероятностей выводит из математической модели, а математическая статистика определяет свойства модели, исходя из данных наблюдений.

Предметом математической статистики является изучение случайных величин (событий, процессов) по результатам наблюдений.

Задачи математической статистики:

- сбор, описание и упорядочение статистического материала (собрать данные и представить в удобном для обозрения и анализа виде);
- выбор и определение вида распределения для полученных в эксперименте наборов случайных величин;
- оценка, хотя бы приблизительная, параметров распределения (дать оценку неизвестной вероятности события, оценку неизвестной функции распределения и т. п.);
- проверка правдоподобия выдвигаемой гипотезы о соответствии статистического материала теоретическим выводам (например, выдвигается гипотеза, что случайная величина подчиняется нормальному закону или математическое ожидание, наблюдаемой случайной величины равно нулю, случайное событие обладает данной вероятностью и т. п.).

Полученные результаты исследования статистических данных методами математической статистики ***используются для принятия решения*** (в задачах планирования, прогнозирования, управления и организации производства, при контроле качества продукции, в медицинских обследованиях, …), т.е. для научных и практических выводов.

Говорят, что «математическая статистика – это теория принятия решений в условиях неопределённости».

1.2. Генеральная совокупность и выборка

Пусть требуется изучить множество однородных объектов относительно некоторого качественного признака, характеризующего эти объекты. Например, если имеется партия деталей, то качественным признаком может служить стандартность детали, а количественным – контролируемый размер детали.

Лучше всего провести сплошное обследование, т. е. изучить каждый объект статистической совокупности (например, всеобщая перепись населения). Но на практике сплошные исследования проводят крайне редко. К тому же, если совокупность содержит большое количество объектов или исследование объекта нарушает его целостность, проводить сплошное исследование бессмысленно (так, если необходимо знать глубину воронки при взрыве снарядов из опытной партии, то при этом всю партию снарядов уничтожать нецелесообразно). В таких случаях из всей совокупности случайно отбирают ограниченное число объектов для изучения.

Множество всех изучаемых объектов или возможных результатов всех мыслимых наблюдений некоторой случайной величины, которые могут быть получены в данных условиях, называется ***генеральной совокупностью***. Число N (конечное или бесконечное) объектов (наблюдений) в совокупности называется её ***объёмом***.

Множество случайно n отобранных объектов (измерений) случайной величины из генеральной совокупности называется ***выборочной совокупностью*** или просто ***выборкой***. Число n – ***объём выборки*** ($n << N$).

Для получения хороших оценок характеристик генеральной совокупности необходимо, чтобы объекты выборки правильно представляли изучаемые признаки, т.е. выборка должна быть ***репрезентативной (представительной)*** и основана на принципе случайного и равновероятностного отбора единиц. Считается, что выборка репрезентативна, если все объекты генеральной совокупности имеют одинаковую вероятность попасть в выборку, т. е. выбор проводится случайно. Например, для того чтобы оценить будущий урожай ещё несозревших яблок и исследовать их характеристики (массу, качество и пр.). Если вся выборка будет сделана с одного дерева, то она не будет репрезентативной. Репрезентативная выборка должна состоять из случайно отобранных плодов со случайно выбранных деревьев.

Различают следующие виды выборок:

- ***простая*** или ***собственно-случайная выборка***, при которой из генеральной совокупности случайным образом извлекают по одному объекту;
- ***механическая выборка***, когда элементы отбирают через определённый интервал, например каждая десятая деталь;
- ***типическая выборка*** характеризуется случайным отбором элементов из типических групп (например, мнение о прошедших выборах спрашивают у случайно отобранных людей, разделённых по признаку пола, возраста, социального статуса, …);
- ***серийная выборка***, в которую случайным образом отбираются не отдельные элементы, а целые группы совокупности подвергаются сплошному наблюдению (сдача ЕГЭ по математике выпускниками школ).

Используют два способа образования выборки:

- ***повторный отбор (с возвращением)***, когда случайно отобранный и уже обследованный объект, возвращается в общую совокупность и теоретически может быть повторно отобран;
- ***бесповторный отбор***, когда отобранный элемент не возвращается в общую совокупность.

На практике пользуются сочетанием вышеуказанных способов и видов отбора.

Важнейшей задачей выборочного метода является оценка параметров (математическое ожидание, дисперсия и т. д.) генеральной совокупности, что невозможно для больших значений N. Оцениваемые характеристики рассчитываются для выборки и объявляются точечными оценками характеристик всей совокупности. Чем больше $n \to N$, тем с большим основанием можно судить о свойствах генеральной совокупности согласно закону больших чисел.

1.3. Статистическое распределение выборки

1.3.1. Дискретный случай

Пусть для исследования некоторой случайной величины X из генеральной совокупности извлечена выборка, в которой значение x_1 наблюдалось m_1 раз, x_2 – m_2 раз, …, значение x_k – m_k раз, причём $m_1 + m_2 + \ldots + m_k = n$ – объём выборки. Значения x_1, x_2, …, x_k называются ***вариантами***.

Числа наблюдений m_1, m_2, …, m_k называются ***частотами*** или ***весами***. Отношения частот к объёму выборки $p_i^* = \dfrac{m_i}{n}$ называются ***относительными частотами*** или ***частостями***. Т. е. по определению $p_1^* = \dfrac{m_1}{n}$, $p_2^* = \dfrac{m_2}{n}$, …, $p_k^* = \dfrac{m_k}{n}$. При этом $p_1^* + p_2^* + \ldots + p_k^* = 1$.

Вся совокупность значений случайной величины X представляет собой первичный статистический материал, который необходимо обработать, прежде всего – упорядочить (расположение выборочных наблюдений по возрастанию называется ***ранжированием***).

Упорядоченный перечень вариант и соответствующих им частот или частостей называется ***статистическим распределением выборки*** или ***дискретным вариационным рядом***. Записывается такое распределение в виде таблицы.

x_i	x_1	x_2	$\dots$	x_k	Σ
m_i	m_1	m_2	$\dots$	m_k	$\sum_{i=1}^{k} m_i = n$
$p_i^* = \frac{m_i}{n}$	p_1^*	p_2^*	$\dots$	p_k^*	$\sum_{i=1}^{k} p_i^* = 1$

Статистическое распределение выборки является оценкой неизвестного распределения. В соответствии с теоремой Бернулли относительные частоты p_i^* сходятся к соответствующим вероятностям p_i при $n \to \infty$, т. е. $p_i^* \xrightarrow[n\to\infty]{P} p_i$. А это означает, что при большом объёме выборки, статистическое распределение мало отличается от истинного распределения.

Пример 1. В супермаркете проводились наблюдения над числом X покупателей, обратившихся в кассу за один час. Наблюдения в течение 30 часов (15 дней с 9 до 10 и с 10 до 11 часов) дали следующие результаты: 80, 90, 100, 110, 90, 60, 100, 110, 80, 60, 70, 100, 70, 100, 80, 90, 60, 100, 100, 110, 80, 90, 80, 110, 70, 80, 90, 80, 100, 100.

Число X является дискретной случайной величиной, а полученные данные представляют собой выборку из $n = 30$ наблюдений. Требуется составить ряд распределения частот (вариационный ряд).

☺ *Решение*. Сначала составим ранжированный ряд: 60, 60, 60, 70, 70, 70, 80, 80, 80, 80, 80, 80, 80, 90, 90, 90, 90, 90, 100, 100, 100, 100, 100, 100, 100, 100, 110, 110, 110, 110.

Получено шесть различных значений случайной величины (шесть вариант). Подсчитав частоту значений каждой варианты, составим таблицу 1, которая и будет представлять собой вариационный ряд.

Таблица 1

Число обращений в кассу x_i	60	70	80	90	100	110	Σ
Частота m_i	3	3	7	5	8	4	$n=\sum_{i=1}^{6} m_i = 30$
Относительная частота $p_i^* = \dfrac{m_i}{n}$	$\dfrac{3}{30}$	$\dfrac{3}{30}$	$\dfrac{7}{30}$	$\dfrac{5}{30}$	$\dfrac{8}{30}$	$\dfrac{4}{30}$	$\sum_{i=1}^{6} p_i^* = 1$

☻

1.3.2. Непрерывный случай

Если признак является непрерывным или число различных значений в выборке велико, вычислять частоту каждого из них не имеет большого смысла. В этом случае составляют ***интервальный вариационный ряд***. Весь промежуток измерения значений выборки $[x_{\min}; x_{\max}]$ (от минимального до максимального) разбивают на частичные интервалы (чаще одинаковой длины), т. е. производится группировка.

Число интервалов k следует брать не очень большим, чтобы после группировки ряд не был громоздким, и не очень малым, чтобы не потерять особенности распределения признака.

Число интервалов может быть определено по формуле Стерджеса

$$k \approx 1 + \log_2 n, \qquad (1)$$

где $\log_2 n \approx 3{,}322 \cdot \lg n$, значение k подбирается целым.

Такой способ определения числа интервалов (1) является лишь рекомендуемым, но не обязательным.

Длина интервала находится по формуле

$$h = \frac{x_{\max} - x_{\min}}{k}. \qquad (2)$$

За начало первого частичного интервала, как правило (но не обязательно), выбирается точка $x_0 = x_{\min} - \frac{h}{2}$.

В первую строку таблицы интервального ряда вписывают частичные промежутки $[x_0; x_1]$, $(x_1; x_2]$, …, $(x_{k-1}; x_k]$, имеющие одинаковую длину h, при этом весь интервал $[x_0; x_k]$ должен полностью покрывать все имеющиеся значения признака, т. е. $x_0 \le x_{\min}$, $x_{\max} \le x_k$. Во второй строке вписывают количество наблюдений m_i $(i = 1, 2, ..., k)$, попавших в каждый интервал. Таким образом, статистическое распределение примет вид:

$(x_{i-1}; x_i]$	$[x_0; x_1]$	$(x_1; x_2]$	…	$(x_{k-1}; x_k]$	Σ
m_i	m_1	m_2	…	m_k	$\sum_{i=1}^{k} m_i = n$
$p_i^* = \frac{m_i}{n}$	p_1^*	p_2^*	…	p_k^*	$\sum_{i=1}^{k} p_i^* = 1$

Пример 2. В таблице 2 приведена выборка результатов измерения роста 105 студентов (юношей). Измерения проводились с точностью до 1 см.

Таблица 2

155	170	185	180	188	152	173	178	178	168	185	172	170	183	175
173	170	183	175	180	175	193	178	183	180	197	178	181	187	168
174	179	184	183	178	180	178	163	166	178	175	182	190	167	170
178	183	170	178	181	173	168	185	175	170	155	169	186	179	189
156	174	179	179	169	186	174	171	184	175	193	178	184	180	196
175	181	188	168	179	178	183	184	178	181	177	163	166	178	175
183	190	167	170	178	183	170	178	182	173	168	186	176	171	188

Требуется составить интервальный вариационный ряд.

☺ *Решение*. Рост юношей есть случайная непрерывная величина.

По формуле (1) найдём количество интервалов при $n = 105$:

$$k \approx 1 + \log_2 n = 1 + \log_2 105 = 7{,}714 \approx 8.$$

Учитывая, что $x_{\min} = 152$, $x_{\max} = 196$, по формуле (2) находим длину частичного интервала: $h = \frac{x_{\max} - x_{\min}}{k} = \frac{192 - 156}{8} \approx 6$.

Примем $x_0 = x_{\min} - \frac{h}{2} = 152 - \frac{6}{2} = 152 - 3 = 149$.

Исходные данные разбиваем на 8 интервалов: $[149;155]$, $(155;161]$, $(161;167]$, $(167;173]$, $(173;179]$, $(179;185]$, $(185;191]$, $(191;197]$.

Подсчитав число студентов m_i, попавших в каждый из полученных промежутков, получим интервальный вариационный ряд (табл. 3).

Таблица 3

Рост x_i	149-155	155-161	161-167	167-173	173-179	179-185	185-191	191-197
Частота m_i	3	1	6	22	33	26	10	4
Частость $p_i^* = \frac{m_i}{n}$	0,03	0,01	0,06	0,21	0,31	0,25	0,09	0,04

Здесь $n = \sum_{i=1}^{8} m_i = 105$. ☻

При изучении вариационных рядов наряду с понятием частоты используется понятие ***накопленной частоты*** (обозначаем $m_i^{\text{нак}}$). Накопленная частота показывает, сколько наблюдалось вариантов со значением признака, меньшим чем *x*. Отношение накопленной частоты к общему числу наблюдений назовём ***накопленной частостью*** $p_i^{\text{нак}}$.

Накопленные частоты (частости) для каждого интервала находятся последовательным суммированием частот (частостей) всех предшествующих интервалов, включая данный.

1.4. Графическое изображение статистического ряда

Для наглядности статистический ряд можно изобразить графически в виде полигона, гистограммы и кумуляты.

Дискретный вариационный ряд графически изображается с помощью полигона. ***Полигоном частот*** называется ломаная линия, соединяющая точки дискретного ряда $(x_1;m_1)$, $(x_2;m_2)$, …, $(x_k;m_k)$; ***полигоном относительных частот*** называется ломаная линия, соединяющая точки $(x_1;p_1^*)$, $(x_2;p_2^*)$, …, $(x_k;p_k^*)$ (рис. 1). Варианты x_i откладываются по оси *Ox*, а соответствующие им частоты m_i (или частости p_i^*) – по оси *Oy*.

Заметим, что полигон, построенный по дискретному вариационному ряду, является выборочным аналогом многоугольника распределения дискретной случайной величины. В первом приближении полигон указывает на вид теоретического распределения.

Интервальный вариационный ряд графически изображают с помощью гистограммы. ***Гистограммой относительных частот*** называется ступенчатая фигура (рис. 2), состоящая из прямоугольников, основаниями которых являются частичные интервалы (длины $h_i = x_i - x_{i-1}$), а высоты равны отношению $f_i^* = \dfrac{p_i^*}{h}$ (либо иногда $f_i^* = p_i^*$).

Неотрицательное число f_i^* называется ***плотностью относительной частоты***.

Каждый прямоугольник гистограммы имеет площадь $S_i = h \cdot \dfrac{p_i^*}{h} = p_i^*$. Тогда площадь всей фигуры $S = \sum\limits_{i=1}^{k} S_i = \sum\limits_{i=1}^{k} p_i^* = 1$, также как и площадь под графиком истинной дифференциальной функции $f(x)$ теоретического распределения непрерывной случайной величины.

Таким образом, f_i^* – *статистический аналог дифференциальной функции* $f(x)$. *Гистограмма, являясь статистическим аналогом кривой* *распределения*, в первом приближении указывает на вид теоретического распределения непрерывной случайной величины.

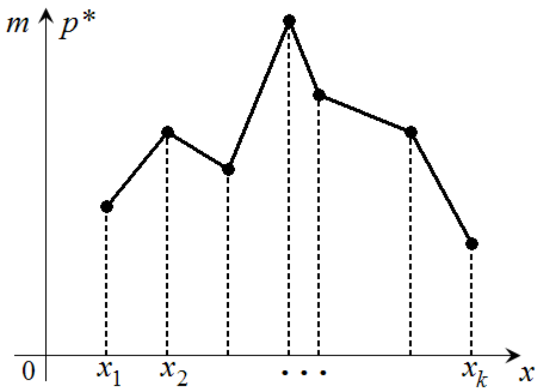


Рис. 1. Полигон частот (относительных частот)

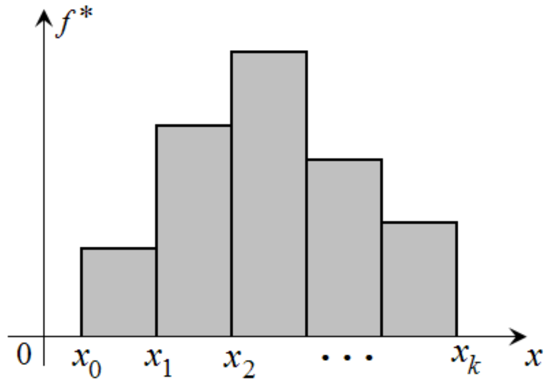


Рис. 2. Гистограмма относительных частот

Если соединить середины верхних оснований прямоугольников отрезками, то получим полигон того же распределения. Полигон относительных частот также указывает на вид теоретического распределения.

Кумулятой называется кривая накопленных частот (рис. 3), которая представляет собой ломаную линию, соединяющую точки $(x_i; m_i^{\text{нак}})$ или $(x_i; p_i^{\text{нак}})$.

В примере 2 для значения $x = 179$ имеем:

– накопленная частота равна сумме всех предшествующих частот

$$m_i^{\text{нак}} = 4 + 6 + 22 + 33 = 65,$$

– соответствующая накопленная относительная частота

$$p_i^{\text{нак}} = 0,03 + 0,01 + 0,06 + 0,21 + 0,31 = 0,62.$$

Весьма важным является понятие эмпирической функции распределения.

Эмпирической функцией распределения называется функция $F^*(x)$, задающая для каждого значения *x* относительную частоту события $\{X < x\}$. Следовательно, по определению

$$F^*(x) = \frac{m_x}{n}, \tag{3}$$

где m_x – число выборочных значений величины *X*, меньших *x*, а *n* – объём выборки.

Очевидно, что $F^*(x)$ удовлетворяет тем же условиям, что и истинная функция распределения $F(x)$, т. к. является её статистическим аналогом.

Значения $F^*(x)$ *на интервалах*:	*Свойства* $F^*(x)$:
$F^*(x)=\begin{cases} 0, & x \le x_1, \\ p_1^*, & x_1 < x \le x_2, \\ p_1^* + p_2^*, & x_2 < x \le x_3, \\ p_1^* + p_2^* + p_3^*, & x_3 < x \le x_4, \\ \ldots\ldots\ldots\ldots\ldots\ldots, & \\ 1, & x > x_k. \end{cases}$	• неубывающая; • неотрицательная; • область значений: $[0,1]$; • если x_1 – наименьшая варианта, то $(\forall x \le x_1)$: $F^*(x)=0$; • если x_k – наибольшая варианта, то $(\forall x > x_k)$: $F^*(x)=1$.

При увеличении числа наблюдений (опытов) относительная частота p_i^* события $\{X < x\}$ приближается к вероятности p_i этого события (теорема Бернулли). Таким образом, $F^*(x)$ *является статистическим аналогом истинной функции распределения* $F(x)$ *случайной величины.*

Имеет место **теорема (Гливенко).** Пусть $F(x)$ – теоретическая функция распределения случайной величины *X*, а $F^*(x)$ – эмпирическая. Тогда для любого $\varepsilon > 0$ верно, что $\lim\limits_{n\to\infty} P\Big\{\big|F^*(x)-F(x)\big| > \varepsilon\Big\} = 0$.

Примем теорему без доказательства.

Для дискретного вариационного ряда эмпирическая функция распределения представляет собой разрывную ступенчатую функцию (рис. 4.) по аналогии с функцией распределения дискретной случайной величины с той лишь разницей, что теперь по оси ординат вместо вероятностей p располагаются относительные частоты p^*.

Для интервального вариационного ряда имеем значения накопленных частот на концах интервалов. Поэтому для графического изображения этой функции в виде непрерывной линии целесообразно её доопределить, соединив точки графика, соответствующие концам интервалов, отрезками прямой. В результате полученная ломаная совпадёт с кумулятой (рис. 3).

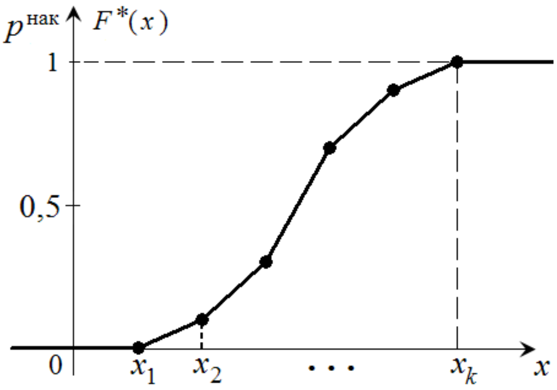


Рис. 3. Кумулята
относительных частот

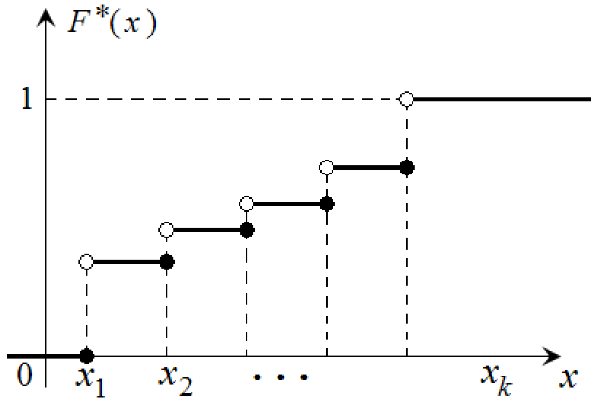


Рис. 4. Эмпирическая функция
дискретного вариационного ряда

Пример 3. Построить полигон относительных частот и эмпирическую функцию распределения, используя условие и результаты примера 1.

☺ *Решение.* Используя данные таблицы 1 для точек $(x_i; p_i^*)$, построим полигон (рис.5). Значения функции $F^*(x)$ найдем по формуле (3):

Имеем $F^*(x) = \dfrac{0}{30} = 0$ при $x \le 60$

(наблюдений меньше 60 нет);

$F^*(x) = \dfrac{3}{30}$ при $60 < x \le 70$;

$F^*(x) = \dfrac{3}{30} + \dfrac{3}{30} = \dfrac{6}{30}$ при $70 < x \le 80$

и т. д.

$$\Rightarrow F^*(x) = \begin{cases} 0, & \text{при } x \le 60, \\ \dfrac{3}{30}, & \text{при } 60 < x \le 70, \\ \dfrac{6}{30}, & \text{при } 70 < x \le 80, \\ \dfrac{13}{30}, & \text{при } 80 < x \le 90, \\ \dfrac{18}{30}, & \text{при } 90 < x \le 100, \\ \dfrac{26}{30}, & \text{при } 100 < x \le 110, \\ 1, & \text{при } x > 110. \end{cases}$$

График функции $F^*(x)$ представлен на рис. 6.

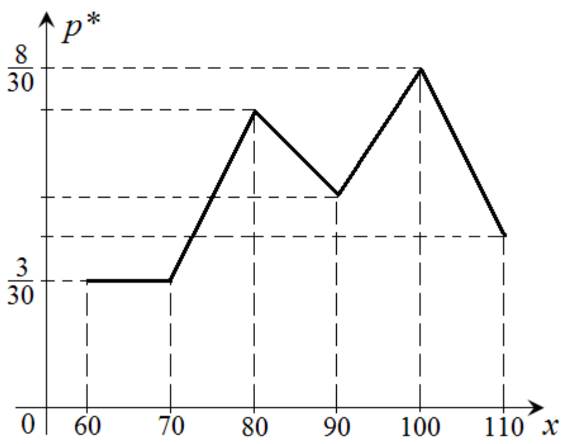


Рис. 5. Полигон
относительных частот

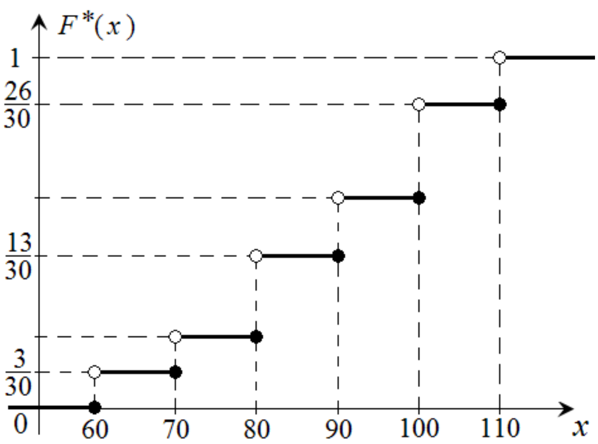


Рис. 6. Эмпирическая
функция распределения ☻

Пример 4. Построить гистограмму и эмпирическую функцию распределения, используя условие и результаты примера 2.

☺ *Решение.* В таблице 3 доопределим середины частичных интервалов $\tilde{x}_i = \dfrac{x_{i-1} + x_i}{2}$, $i = 1, 2, ..., k$. Запишем таблицу 4.

Таблица 4

Рост, $(x_{i-1}; x_i]$	149-155	155-161	161-167	167-173	173-179	179-185	185-191	191-197
Середина, $\tilde{x}_i$	152	158	164	170	176	182	188	194
Частота, m_i	3	1	6	22	33	26	10	4
Частость, p_i^*	0,03	0,01	0,06	0,21	0,31	0,25	0,09	0,04

Гистограмма относительных частот интервального ряда представляет собой ступенчатую фигуру (рис. 7), состоящую из прямоугольников с основаниями длиной $h = 6$ и высотами $f_i^* = \dfrac{p_i^*}{6}$, $i = 1, 2, ..., 8$.

Для интервального вариационного ряда эмпирическая функция распределения совпадает с кумулятой. Значения функции $F^*(x)$ найдем по формуле (3) и запишем в таблице:

x	$x \le 149$	155	161	167	173	179	185	191	$x \ge 197$
$F^*(x)$	0	0,03	0,04	0,10	0,31	0,62	0,87	0,96	1

Отметим на плоскости точки, соответствующие значениям $F^*(x)$, и соединим их отрезками прямых (рис. 8).

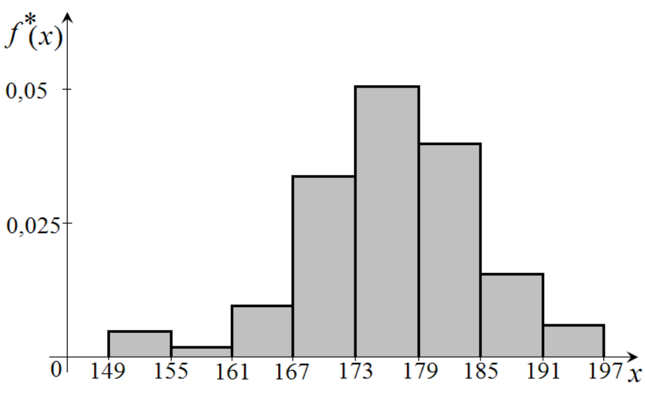


Рис. 7. Гистограмма относительных частот

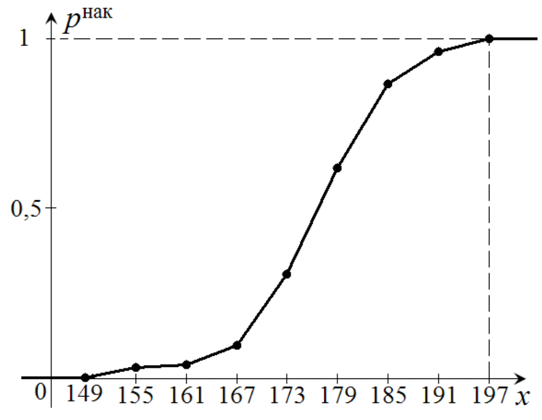


Рис. 8. Эмпирическая функция распределения ☻

1.5. Числовые характеристики выборки

1.5.1. Средние величины. Показатели вариации

Для выборки можно определить ряд числовых характеристик, аналогичных тем, которые в теории вероятностей определялись для случайных величин.

Пусть имеется выборка объёма n:

Варианта, x_i	x_1	x_2	…	x_k	$\sum$
Частота, m_i	m_1	m_2	…	m_k	$\sum_{i=1}^{k} m_i = n$
Частость, p_i^*	p_i^*	p_2^*	…	p_k^*	$\sum_{i=1}^{k} p_i^* = 1$

Выборочным средним называется среднее арифметическое всех значений выборочной совокупности (обозначения: $\overline{x}_в$, $\overline{x}$, $M^*[X]$, m_x^*).

Выборочное среднее является статистическим аналогом математического ожидания случайной величины.

Если все значения x_1, x_2, …, x_n признака выборки объёма n различны, то выборочное среднее определяется по формуле

$$\boxed{\overline{x}_в = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i}, \tag{4}$$

Если статистические данные сгруппированы (т. е. каждому значению варианты x_i соответствует определённая частота m_i), то формула (4) примет вид

$$\boxed{\overline{x}_в = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i m_i} = \sum_{i=1}^{k} x_i \cdot \frac{m_i}{n} \boxed{= \sum_{i=1}^{k} x_i p_i^*}. \tag{5}$$

В случае интервального ряда в качестве x_i берутся середины частичных интервалов, т. е. значения $\tilde{x}_i = \dfrac{x_{i-1} + x_i}{2}$, m_i – соответствующие им частоты, $i = 1, 2, ..., k$.

Выборочной дисперсией называется среднее арифметическое квадратов отклонений вариант от выборочной средней $\overline{x}_в$

$$\boxed{D_в = \frac{1}{n}\left((x_1-\overline{x}_в)^2+(x_2-\overline{x}_в)^2+\ldots+(x_n-\overline{x}_в)^2\right) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\overline{x}_в)^2}, \quad (6)$$

если все значения в выборке различны.

Если статистические данные сгруппированы, то формула (6) примет вид

$$\boxed{D_в = \frac{1}{n}\cdot\sum_{i=1}^{k}(x_i-\overline{x}_в)^2\cdot m_i} = \sum_{i=1}^{k}(x_i-\overline{x}_в)^2\cdot\frac{m_i}{n} \boxed{=\sum_{i=1}^{k}(x_i-\overline{x}_в)^2\cdot p_i^*}. \quad (7)$$

В результате преобразований в (7), можно получить другую формулу для вычисления выборочной дисперсии:

$$D_в = \sum_{i=1}^{k}(x_i-\overline{x}_в)^2\cdot p_i^* = \sum_{i=1}^{k}\left(x_i^2-2\overline{x}_в\cdot x_i+\overline{x}_в^2\right)\cdot p_i^* =$$

$$= \sum_{i=1}^{k}x_i^2\cdot p_i^* - 2\overline{x}_в\sum_{i=1}^{k}x_i\cdot p_i^* + \overline{x}_в^2\sum_{i=1}^{k}p_i^* = \overline{x^2} - 2\overline{x}_в\cdot\overline{x}_в + \overline{x}_в^2\cdot 1 = \overline{x^2} - \overline{x}_в^2.$$

Таким образом, получаем

$$\boxed{D_в = \overline{x^2} - (\overline{x}_в)^2}. \quad (8)$$

Выборочная дисперсия является статистическим аналогом дисперсии теоретического распределения.

Выборочное среднее квадратическое отклонение определяется по формуле

$$\boxed{\sigma_в = \sqrt{D_в}}. \quad (9)$$

Выборочное среднее квадратическое отклонение является мерилом надёжности выборочной средней и измеряется в тех же величинах, что и изучаемый признак. Чем меньше выборочное квадратическое отклонение, тем лучше выборочная средняя отражает собой всю представляемую совокупность.

В качестве описательных характеристик вариационного ряда используются мода, медиана, размах и т. д.

Модой M_O^* вариационного ряда называется такое значение варианты, которой соответствует наибольшая частота.

В случае интервального вариационного ряда мода находится внутри частичного интервала $\ell_i = (x_{i-1}; x_i)$, которому соответствует наибольшая частота m_i.

На рис. 9 показан наглядный геометрический способ нахождения приближённого значения моды.

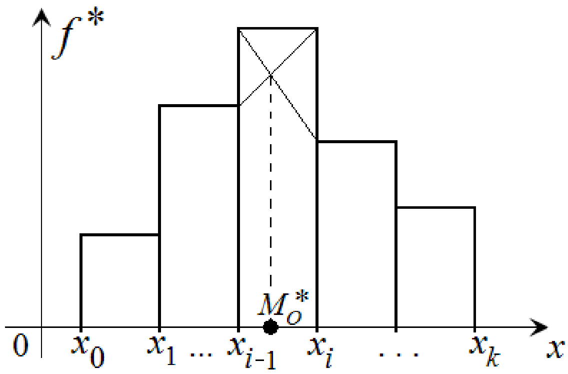


Рис. 9

Истинное значение моды $M_O^* \in (x_{i-1}; x_i)$ вычисляется по формуле линейной интерполяции:

$$\boxed{M_O^* = x_{i-1} + h_i \cdot \frac{p_i^* - p_{i-1}^*}{(p_i^* - p_{i-1}^*) + (p_i^* - p_{i+1}^*)}}, \qquad (10)$$

где h_i – длина частичного интервала ℓ_i,

p_{i-1}^* – частость, соответствующая предыдущему частичному интервалу ℓ_{i-1},

p_{i+1}^* – частость, соответствующая следующему частичному интервалу ℓ_{i+1}.

Особенность моды как меры центральной тенденции заключается в том, что она не изменяется при изменении крайних членов ряда, т. е. обладает определённой устойчивостью к вариации признака.

Медианой M_e^* называется значение признака, приходящееся на середину ранжированного ряда наблюдений.

Если вариационный ряд имеет нечётное число членов ($n = 2m+1$), то $M_e^* = x_{m+1}$, если число членов ряда чётно ($n = 2m$), то $M_e^* = \dfrac{x_m + x_{m+1}}{2}$.

В случае интервального ряда медиана принадлежит тому частичному интервалу $\ell_i = (x_{i-1}; x_i)$, для которого накопленная частота составляет половину или больше половины всей суммы частот, а предыдущая накопленная частота меньше половины всей суммы частот, при этом прямая $x = M_e^*$ делит площадь гистограммы пополам.

Медиана может быть приближённо найдена с помощью кумуляты (графика $F^*(x)$), как значение признака, для которого $p_i^{\text{нак}} = 1/2$.

На рис. 10 показан наглядный геометрический способ нахождения приближённого значения медианы.

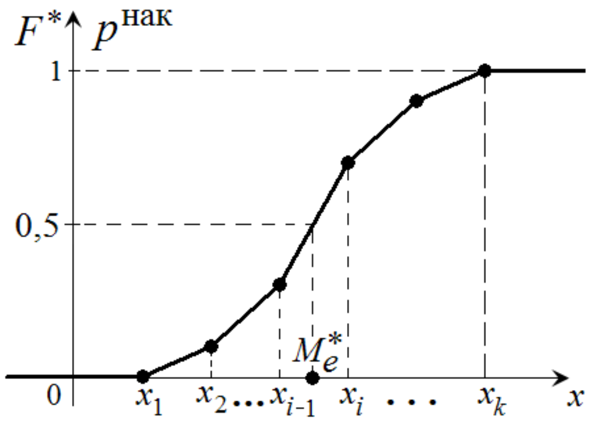


Рис. 10

Истинное значение медианы $M_e^* \in (x_{i-1}; x_i)$ интервального вариационного ряда вычисляется по формуле линейной интерполяции:

$$\boxed{M_e^* = x_{i-1} + \frac{h_i}{p_i^*} \cdot \left(0{,}5 - \sum_{j=1}^{i-1} \frac{h_j}{p_j^*} \right)}. \qquad (11)$$

Достоинство медианы как меры центральной тенденции заключается в том, что на неё не влияет изменение крайних членов вариационного ряда, если любой из них, меньший медианы, остаётся меньше её, больший медианы, продолжает быть больше её. Медиана предпочтительнее средней арифметической для ряда, у которого крайние варианты по сравнению с остальными оказались чрезмерно большими или малыми.

Для характеристики меры колеблемости изучаемого признака относительно выборочной средней используется ***коэффициент вариации***:

$$\boxed{V^* = \frac{\sigma_в}{\overline{x}_в} \cdot 100\,\%}. \qquad (12)$$

Коэффициент вариации служит для сравнения величин рассеяния по отношению к выборочной средней двух вариационных рядов: тот из рядов имеет большее рассеяние по отношению к выборочной средней, у которого коэффициент вариации больше.

В отличие от $\sigma_в$ коэффициент вариации является безразмерной величиной, поэтому он пригоден для сравнения рассеяний вариационных рядов, варианты которых имеют различную размерность, например, если варианты одного ряда выражены в сантиметрах, а другого – в граммах.

Если коэффициент вариации признака, принимающего только положительные значения, высок (например, более 100 %), то, как правило, это говорит о неоднородности значений признака.

Замечание. Выше предполагалось, что вариационный ряд составлен по данным выборки, поэтому все описанные характеристики называют ***выборочными***; если вариационный ряд составлен по данным генеральной совокупности, то характеристики называют ***генеральными***.

Пример 5. Данные о распределении 100 рабочих цеха по выработке в отчётном году (в процентах к предыдущему году) представлены в табл.5.

Таблица 5

Выработка в отчётном году в процентах к предыдущему, $(x_{i-1};x_i]$	Середина интервала, $\tilde{x}_i$ (%)	Количество рабочих, m_i (чел.)
94 – 100	97	3
100 – 106	103	7
106 – 112	109	11
112 – 118	115	20
118 – 124	121	28
124 – 130	127	19
130 – 136	133	10
136 – 142	139	2
Σ	**–**	**100**

Найти среднюю выработку по цеху. Вычислить дисперсию, среднее квадратическое отклонение, моду, медиану, коэффициент вариации.

☺ *Решение*. Вычисляем среднюю выборочную интервального вариационного ряда по формуле (5):

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k}\tilde{x}_i\, m_i = \frac{1}{100}(97\cdot 3 + 103\cdot 7 + \ldots + 139\cdot 2) = 119{,}2\,.$$

Вычислим выборочную дисперсию по формуле (8):

$$D_в = \overline{x^2} - (\overline{x}_в)^2 = \frac{1}{100}\left(97^2 \cdot 3 + 103^2 \cdot 7 + \ldots + 139^2 \cdot 2\right) - 119{,}2^2 = 87{,}48.$$

Тогда, согласно (9) $\sigma_в = \sqrt{D_в} = \sqrt{87{,}48} \approx 9{,}35$ (%).

Найдём моду, медиану и коэффициент вариации по формулам (10), (11) и (12) соответственно:

$$M_O^* = 118 + 6 \cdot \frac{0{,}28 - 0{,}20}{(0{,}28 - 0{,}20) + (0{,}28 - 0{,}19)} = 118 + 6 \cdot \frac{0{,}08}{0{,}17} \approx 120{,}82.$$

$$M_e^* = 118 + \frac{6}{0{,}28} \cdot \left(0{,}5 - (0{,}03 + +0{,}07 + 0{,}11 + 0{,}20)\right) \approx 119{,}93.$$

$$V^* = \frac{9{,}35}{119{,}2} \cdot 100\,\% \approx 7{,}84\,\%.$$ ☻

1.5.2. Вычисление выборочных средней и дисперсии при больших (очень малых) значениях вариант

Для выборочного среднего и выборочной дисперсии справедливы следующие ***свойства***:

1. *Если варианты увеличить* (*уменьшить*) *на одно и то же число C, то выборочное среднее также увеличится* (*уменьшится*) *на это число, а выборочная дисперсия останется неизменной.*

Тогда для $u_i = x_i + C$ имеем:

$$\overline{u}_в = \sum_{i=1}^{k} u_i \cdot p_i^* = \sum_{i=1}^{k} (x_i + C) \cdot p_i^* = \sum_{i=1}^{k} x_i \cdot p_i^* + \sum_{i=1}^{k} C \cdot p_i^* = \overline{x}_в + C\sum_{i=1}^{k} p_i^* = \overline{x}_в + C,$$

$$D_u = \overline{u^2} - (\overline{u}_в)^2 = \sum_{i=1}^{k} u_i^2 \cdot p_i^* - (\overline{u}_в)^2 = \sum_{i=1}^{k} (x_i + C)^2 \cdot p_i^* - (\overline{x}_в + C)^2 =$$

$$= \sum_{i=1}^{k} \left(x_i^2 + 2C \cdot x_i + C^2\right) \cdot p_i^* - \left(\overline{x}_в^2 + 2C \cdot \overline{x}_в + C^2\right) =$$

$$= \sum_{i=1}^{k} x_i^2 \cdot p_i^* + \sum_{i=1}^{k} 2C \cdot x_i \cdot p_i^* + \sum_{i=1}^{k} C^2 \cdot p_i^* - \left(\overline{x}_в^2 + 2C \cdot \overline{x}_в + C^2\right) =$$

$$= \overline{x^2} + 2C \cdot \overline{x}_в + C^2 - \left(\overline{x}_в^2 + 2C \cdot \overline{x}_в + C^2\right) = \overline{x^2} - (\overline{x}_в)^2 = D_x.$$

2. *Если варианты увеличить (уменьшить) в h раз, то выборочное среднее также увеличится (уменьшится) в h раз, а выборочная дисперсия увеличится (уменьшится) в h^2 раз.*

Действительно, если $u_i = h \cdot x_i$, то

$$\overline{u}_в = \sum_{i=1}^{k} u_i \cdot p_i^* = \sum_{i=1}^{k} h \cdot x_i \cdot p_i^* = h \cdot \sum_{i=1}^{k} x_i \cdot p_i^* = h \cdot \overline{x}_в,$$

$$D_u = \overline{u^2} - (\overline{u}_в)^2 = \sum_{i=1}^{k} u_i^2 \cdot p_i^* - (\overline{u}_в)^2 = \sum_{i=1}^{k} (h \cdot x_i)^2 \cdot p_i^* - (h \cdot \overline{x}_в)^2 =$$

$$= h^2 \sum_{i=1}^{k} x_i^2 \cdot p_i^* - h^2 (\overline{x}_в)^2 = h^2 \cdot \overline{x^2} - h^2 \cdot (\overline{x}_в)^2 = h^2 \cdot \left(\overline{x^2} - (\overline{x}_в)^2\right) = h^2 D_x.$$

Из свойств вытекает ***метод упрощённых вычислений***.

Вводим новые варианты $u_i = \dfrac{x_i - C}{h}$, где C и h – специально подобранные постоянные. Согласно свойствам получаем

$$\begin{cases} \overline{u}_в = \dfrac{\overline{x}_в - C}{h}, \\ D_u = \dfrac{1}{h^2} \cdot D_x, \end{cases} \Rightarrow \begin{cases} \overline{x}_в = h \cdot \overline{u}_в + C, \\ D_x = h^2 \cdot D_u = h^2 \left(\overline{u^2} - (\overline{u}_в)^2\right). \end{cases} \qquad (13)$$

Замечание. Выборочная дисперсия, вычисленная по сгруппированным данным интервального вариационного ряда (формула (7)), оказывается меньше дисперсии, найденной по несгруппированным результатам измерений (формула (6)), на величину, приблизительно равную $\dfrac{h^2}{12}$. Это следует учитывать при округлении значения дисперсии, сохраняя один сомнительный знак. Значение выборочной средней округляют при этом до единиц того разряда, который сохранен в значении выборочной дисперсии.

Пример 6. Вычислить упрощённым способом среднюю арифметическую и дисперсию распределения рабочих по выработке по данным табл. 5.

☺ *Решение*. Возьмём постоянную h, равную величине интервала, т. е. $h = 6$, а постоянную C, равную середине одного из двух серединных интервалов, например пятого, т. е. $C = 121$. Новые варианты теперь имеют вид $u_i = \frac{x_i - 121}{6}$. Дальнейшие вычисления представим в табл. 6.

Таблица 6

$(x_{i-1}; x_i]$	$\tilde{x}_i$	$u_i = \frac{x_i - 121}{6}$	m_i	$u_i m_i$	$u_i^2 m_i$
94 – 100	97	– 4	3	– 12	48
100 – 106	103	– 3	7	– 21	63
106 – 112	109	– 2	11	– 22	44
112 – 118	115	– 1	20	– 20	20
118 – 124	121	0	28	0	0
124 – 130	127	1	19	19	19
130 – 136	133	2	10	20	40
136 – 142	139	3	2	6	18
Σ		**–**	**100**	**– 30**	**252**

В итоговой строке табл. 6 имеем $\sum_{i=1}^{8} u_i m_i = -30$, $\sum_{i=1}^{8} u_i^2 m_i = 252$.

Тогда $\overline{u}_в = \frac{1}{100} \cdot (-30) = -0{,}3$, $\overline{u^2} = \frac{1}{100} \cdot 252 = 2{,}52$.

По формулам (13) находим:

$$\overline{x}_в = h \cdot \overline{u}_в + C = 6 \cdot (-0{,}3) + 121 = 119{,}2,$$

$$D_x = h^2 \left(\overline{u^2} - (\overline{u}_в)^2 \right) = 6^2 \left(2{,}52 - (-0{,}3)^2 \right) = 87{,}48.$$

Как видим, найденные значения совпадают с вычислениями, выполненными в примере 5. ☻

1.5.3. Эмпирические моменты

Для вычисления сводных характеристик выборок используют эмпирические моменты, аналогичные соответствующим теоретическим моментам. Выборочное среднее и дисперсия вариационного ряда являются частными случаями более общего понятия – моментов вариационного ряда.

Выборочный начальный момент k-го порядка определяется по формуле:

$$\boxed{\nu_k^* = \sum_{i=1}^{r} x_i^k p_i^*}. \qquad (14)$$

В частности: $\nu_1^* = \overline{x}_в$, т. е. *выборочное среднее является выборочным начальным моментом первого порядка*.

Выборочный центральный момент k-ого порядка находится по формуле:

$$\boxed{\mu_k^* = \sum_{i=1}^{r} (x_i - \overline{x}_в)^k \cdot p_i^*}. \qquad (15)$$

В частности: $\mu_1^* = 0$;

$\mu_2^* = D_в = \nu_2^* - (\nu_1^*)^2$, т. е. *выборочная дисперсия является выборочным центральным моментом 2-го порядка*;

$\mu_3^* = \nu_3^* - 3\nu_1^* \nu_2^* + 2\,(\nu_1^*)^3$;

$\mu_4^* = \nu_4^* - 4\nu_1^* \nu_4^* + 6(\nu_1^*)^2 \cdot \nu_2^* - 3(\nu_1^*)^4$.

Среди моментов высших порядков особое значение имеют центральные моменты 3-го и 4-го порядка.

1.5.4. Асимметрия и эксцесс

Нормальное распределение является одним из наиболее применяемых в математической статистике. Для оценки отклонения эмпирического распределения от нормального используют коэффициенты асимметрии и эксцесс, которые служат для сравнения полигона теоретических частостей вариационного ряда с функцией плотности нормального распределения.

Выборочным коэффициентом асимметрии **(*«скошенности»*)** называется величина

$$\boxed{a_s^* = \frac{\mu_3^*}{\sigma_в^3}}. \qquad (16)$$

Выборочный коэффициент эксцесса **(*«островершинности»*)** определяется по формуле

$$\boxed{\varepsilon_k^* = \frac{\mu_4^*}{\sigma_в^4} - 3}. \qquad (17)$$

Эксцесс характеризует крутизну подъёма кривой распределения по сравнению с нормальной кривой.

Для нормального распределения $a_s = \varepsilon_k = 0$.

При отклонении от нормального распределения:

$a_s^* < 0$	– если «длинная» и более пологая часть кривой распределения расположена слева от соответствующей моде точки на оси абсцисс (рис. 11).
$a_s^* > 0$	– если часть кривой расположена справа от моды (рис. 12).
$\varepsilon_k^* < 0$	– кривая имеет более низкую и пологую вершину (рис. 13).
$\varepsilon_k^* > 0$	– кривая имеет более высокую и острую вершину (рис. 13).

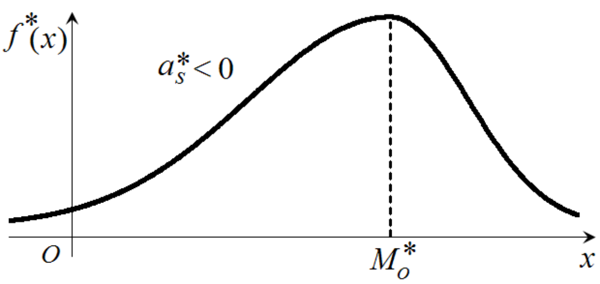


Рис. 11

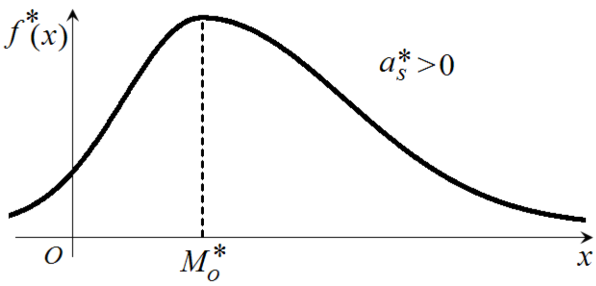


Рис. 12

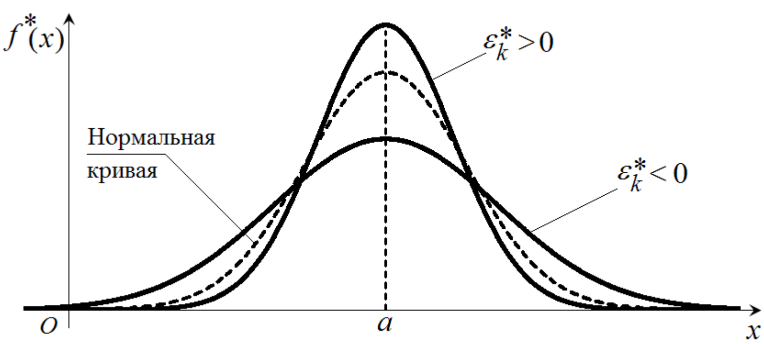


Рис. 13

Следует иметь в виду, что при использовании указанных характеристик сравнения опорными являются предположения об одинаковых величинах математического ожидания и дисперсии для нормального и теоретического распределений.

1.6. Оценка неизвестных параметров

1.6.1. Понятие оценки параметров

Пусть рассматривается случайная величина X с законом распределения, содержащим один или несколько параметров (например, параметр λ в распределении Пуассона или параметры a и σ для нормального закона распределения).

В результате n наблюдений получена выборка X_1, X_2, …, X_n, где X_i– результат измерения в i-ом опыте $(i=1,2,...,n)$, который можно рассматривать как случайную величину, имеющую закон распределения случайной величины X. Необходимо оценить параметр θ, связанный с законом распределения случайной величины X генеральной совокупности.

Приближённое значение неизвестного параметра θ назовём его ***оценкой*** и обозначим θ^*.

Любая оценка θ^*, вычисляемая на основе выборки, есть значение некоторой функции результатов наблюдений, т. е. $\theta^* = \theta^*\left(X_1, X_2, ..., X_n\right)$.

Сама оценка является случайной величиной, так как зависит от случайных величин X_1, X_2, …, X_n. Если произвести другую выборку, то функция примет, вообще говоря, другое значение.

Если число наблюдений (опытов) сравнительно невелико, то замена неизвестного параметра θ его оценкой θ^*, например, математического ожидания выборочным средним, приводит к ошибке. Эта ошибка в среднем тем больше, чем меньше число опытов.

Для того чтобы оценка неизвестного параметра давала хорошее приближение (была достаточно «близкой» к истинному значению параметра), она должна удовлетворять некоторым ***требованиям***.

1) Желательно, чтобы при использовании величины θ^* вместо неизвестного параметра θ мы не делали систематических ошибок ни в сторону завышения, ни в сторону занижения, т. е. чтобы выполнялось равенство

$$M[\theta^*]=\theta .$$

Если математическое ожидание оценки по всевозможным выборкам данного объёма равно истинному значению определяемого параметра, то оценка θ^* называется ***несмещённой***. Если $M[\theta^*]\to\theta$, то оценка θ^* называется ***асимптотически несмещённой***.

Требование несмещённости особенно важно при малом числе наблюдений.

2) Желательно, чтобы с увеличением числа опытов значения случайной величины θ^* концентрировались около θ всё более тесно, т. е. чтобы с ростом *n* точность оценки возрастала.

Оценка θ^* параметра θ называется ***состоятельной***, если она сходится по вероятности к оцениваемому параметру, т. е.

$$\theta^* \xrightarrow[n\to\infty]{P} \theta \text{ или } \lim_{n\to\infty} P\left(|\theta^*-\theta|<\varepsilon\right)=1,$$

где ε – сколь угодно малое положительное число.

Поскольку мерой рассеяния случайной величины вокруг её математического ожидания является дисперсия, то из неравенства Чебышева для удовлетворения этого требования достаточно, чтобы оценка была несмещённой и $\lim\limits_{n\to\infty} D[\theta^*]=0$.

Несостоятельные оценки не имеют практического смысла, поэтому *свойство состоятельности обязательно для любого правила оценивания*.

3) Несмещённая оценка θ^* параметра θ называется ***эффективной***, если она имеет наименьшую дисперсию среди всех возможных несмещённых оценок, вычисленных по выборкам одного и того же объёма.

Эффективность оценки определяется отношением $\mathrm{eff}\,\theta^* = \dfrac{D[\theta_э^*]}{D[\theta^*]}$, где $D[\theta_э^*]$ и $D[\theta^*]$ – дисперсии соответственно эффективной и данной оценок. Чем ближе $\mathrm{eff}\,\theta^*$ к 1, тем эффективнее оценка. Если $\mathrm{eff}\,\theta^* \to 1$ при $n \to \infty$, то оценка называется ***асимптотически эффективной***.

В качестве статистических оценок параметров желательно использовать оценки, удовлетворяющие одновременно требованиям несмещённости, состоятельности и эффективности. Все три свойства определяют оценку однозначно. Но на практике не всегда возможно выполнение всех перечисленных требований, поэтому приходится использовать оценки, не обладающие сразу тремя свойствами.

1.6.2. Метод моментов

Можно полагать, что *начальные и центральные эмпирические моменты являются состоятельными оценками соответствующих теоретических моментов*. На этом предположении базируется метод моментов, который основан на приравнивании теоретических моментов рассматриваемого распределения и соответствующих эмпирических моментов того же порядка. При этом различают случаи распределений с одним параметром и с двумя параметрами.

1. Оценка одного параметра. Пусть задана плотность распределения $f(x,\theta)$ с одним параметром. Согласно методу моментов приравниваем, например, соответствующие начальные моменты первого порядка, т.е. среднюю выборки $\bar{x}_в$ и математическое ожидание распределения $M[X]$. Здесь достаточно одного уравнения относительно этого параметра

$$M[X] = \bar{x}_в. \tag{18}$$

Поскольку математическое ожидание является функцией параметра θ

$$M[X] = \int_{-\infty}^{+\infty} x \cdot f(x, \theta)\, dx = \varphi(\theta), \tag{19}$$

соотношение (18) можно рассматривать как уравнение с одним неизвестным, которое определяет точечную оценку параметра θ, являющуюся функцией.

Пример 7. Методом моментов по выборке x_1, x_2, ..., x_n найти точечную оценку неизвестного параметра λ показательного распределения с известной функцией плотности распределения $f(x) = \lambda e^{-\lambda x}$ ($x \geq 0$).

☺ *Решение*. Формула (19) при помощи интегрирования по частям даёт: $M[X] = \frac{1}{\lambda}$. Далее из формулы (18) получаем, что $\lambda = \frac{1}{\bar{x}_в}$, т. е. искомая точечная оценка параметра λ показательного распределения равна обратной выборочной средней: $\lambda^* = \frac{1}{\bar{x}_в}$. ☻

2. Оценка двух параметров. Пусть задана функция плотности распределения $f(x, \theta_1, \theta_2)$. Для нахождения неизвестных параметров нужно иметь два уравнения, поэтому приравняем друг другу соответственно начальные теоретические и эмпирические моменты первого и второго порядков:

$$M[X] = \bar{x}_в,\ D[X] = D_в. \tag{20}$$

Поскольку $M[X]$ и $D[X]$ есть функции от θ_1 и θ_2, соотношения (20) определяют точечные оценки этих параметров как функции от выборки:

$$\theta_1^* = \theta_1(x_1, x_2, ..., x_n),\ \theta_2^* = \theta_2(x_1, x_2, ..., x_n).$$

Пример 8. Найти методом моментов по выборке x_1, x_2, ..., x_n точечные оценки неизвестных параметров a и σ нормального распределения с функцией плотности распределения вероятности

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

☺ *Решение*. Из курса теории вероятностей известно, что для нормального распределения $M[X]=a$, $D[X]=\sigma^2$. Используя формулы (20), получаем искомые точечные оценки параметров: $a^* = \bar{x}_в$, $\sigma^* = \sqrt{D_в}$. ☻

1.6.3. Метод наибольшего правдоподобия

Дискретные случайные величины. Пусть X – дискретная случайная величина, которая приняла значения x_1, x_2, …, x_n в результате n испытаний. Пусть известен закон распределения случайной величины X, но неизвестен определяющий его параметр θ. Требуется найти точечную оценку этого параметра.

Обозначим вероятность того, что в результате испытания величина X примет значение x_i, через $p(x_i, \theta)$.

Функция аргумента θ

$$L(x_1, x_2, ..., x_n, \theta) = p(x_1, \theta) \cdot p(x_2, \theta) \cdot ... \cdot p(x_n, \theta), \tag{21}$$

где x_1, x_2, …, x_n – фиксированные числа, называется ***функцией правдоподобия дискретной случайной величины*** X.

В качестве точечной оценки параметра θ принимается значение θ^*, при котором функция (21) достигает максимума. Такую оценку θ^* называют ***оценкой наибольшего правдоподобия***.

Функцию $\ln L(x_1, x_2, ..., x_n, \theta) = \ln[p(x_1, \theta) \cdot p(x_2, \theta) \cdot ... \cdot p(x_n, \theta)]$, называют ***логарифмической функцией правдоподобия***.

Точка максимума у обеих функций одна и та же, но вместо функции L, удобнее анализировать функцию $\ln L$.

Пример 9. Методом наибольшего правдоподобия найти оценку неизвестного параметра p биномиального распределения

$$P_n(m) = C_n^m p^m (1-p)^{n-m},$$

если в n_1 испытаниях событие A наступило m_1 раз, а в n_2 испытаниях – m_2 раз.

☺ *Решение*. Составим функцию правдоподобия (21), где $p=\theta$,

$$L=P_{n_1}(m_1)\cdot P_{n_2}(m_2)=C_{n_1}^{m_1}C_{n_2}^{m_2}\,p^{m_1+m_2}(1-p)^{n+m-m_1-m_2}.$$

Логарифмическая функция правдоподобия имеет вид:

$$\ln L=\ln\left(C_{n_1}^{m_1}C_{n_2}^{m_2}\right)+(m_1+m_2)\ln p+(n_1-m_1+n_2-m_2)\ln(1-p).$$

Затем находим производную по p и, приравнивая её к нулю, получаем

$$\frac{m_1+m_2}{p}-\frac{n_1-m_1+n_2-m_2}{1-p}=0.$$

Решение этого уравнения относительно p имеет вид: $p=\dfrac{m_1+m_2}{n_1+n_2}$.

Нетрудно убедиться, что вторая производная функции $\ln L<0$, т. е. полученное значение p является точкой максимума логарифмической функции правдоподобия, а, значит, эту величину нужно принять в качестве оценки наибольшего правдоподобия неизвестного параметра p^* биномиального распределения: $p^*=\dfrac{m_1+m_2}{n_1+n_2}$. ☻

Непрерывные случайные величины. Пусть X – непрерывная случайная величина, которая приняла значения x_1, x_2, …, x_n в результате n испытаний. Пусть известна функция распределения плотности $f(x,\theta)$, но неизвестен определяющий её параметр θ. Требуется найти точечную оценку этого параметра.

Функция аргумента θ

$$L(x_1,x_2,...,x_n,\theta)=f(x_1,\theta)\cdot f(x_2,\theta)\cdot...\cdot f(x_n,\theta), \tag{22}$$

где x_1, x_2, …, x_n – фиксированные числа, называется ***функцией правдоподобия непрерывной случайной величины X***.

Метод оценки наибольшего правдоподобия неизвестного параметра θ такой же, как и в предыдущем случае: поиск точки максимума функции L.

Если в известной функции распределения $f(x,\theta_1,\theta_2)$ неизвестны два параметра, то функция правдоподобия имеет вид

$$L(x_1,x_2,...,x_n,\theta_1,\theta_2)=f(x_1,\theta_1,\theta_2)\cdot f(x_2,\theta_1,\theta_2)\cdot...\cdot f(x_n,\theta_1,\theta_2),$$

и тогда необходимо найти точку максимума функции двух переменных.

Соответствующие логарифмические функции наибольшего правдоподобия вводятся так же, как и в случае дискретной случайной величины.

Пример 10. Методом наибольшего правдоподобия найти оценку неизвестного параметра λ показательного распределения из примера 7.

☺ *Решение*. Составим функцию правдоподобия (22). В нашем случае

$$\theta = \lambda \quad \Rightarrow \quad L = (\lambda e^{-\lambda x_1}) \cdot (\lambda e^{-\lambda x_2}) \cdot \ldots \cdot (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda(x_1 + x_2 + \ldots + x_n)}.$$

Отсюда найдём логарифмическую функцию правдоподобия

$$\ln L = n \ln \lambda - \lambda(x_1 + x_2 + \ldots + x_n).$$

Производную этой функции по λ приравниваем к нулю. Получаем

$$\lambda = \frac{n}{x_1 + x_2 + \ldots + x_n} = \frac{1}{\bar{x}_в},$$

что совпадает с результатом примера 7. ☻

1.6.4. Точечная оценка вероятности события

Пусть произошло n независимых опытов. Случайная величина X_i – индикатор события A в i-ом опыте: $X_i = \begin{cases} 1, & \text{если } A \text{ произошло,} \\ 0, & \text{если } A \text{ не произошло.} \end{cases}$

Тогда событие A произошло $X = \sum\limits_{i=1}^{n} X_i$ раз, и $(n - X)$ раз оно не произошло. При этом X является случайной величиной, имеющей тот же закон распределения, что и случайная величина X_i.

Рассмотрим относительную частоту $p^* = \frac{X}{n} = \frac{1}{n} \cdot \sum\limits_{i=1}^{n} X_i$ как оценку p – вероятности появления события. Так как опыты независимы, то случайная величина X распределена по биномиальному закону с параметрами n и p.

Учитывая, что $M[X]=np$, получим

$$M[p^*]=M\left[\frac{1}{n}\cdot\sum_{i=1}^{n}X_i\right]=\frac{1}{n}\cdot M\left[\sum_{i=1}^{n}X_i\right]=\frac{1}{n}\cdot M[X]=\frac{1}{n}\cdot np=p.$$

$\Rightarrow$ p^* – *несмещённая оценка* p.

По теореме Бернулли имеем $\lim\limits_{n\to\infty}P\left(|\,p-p^*\,|<\varepsilon\right)=1$

$\Rightarrow$ p^* – *состоятельная оценка* p.

Таким образом, *эмпирическая функция распределения выборки* $F^*(x)$, *являясь оценкой вероятности события* $\{X<x\}$, *есть несмещённая состоятельная оценка функции распределения* $F(x)$ *случайной величины X.*

Пример 11. Монету подбрасывают *n* раз. Вероятность выпадения герба при каждом подбрасывании равна *p*. В ходе опыта монета выпала гербом *m* раз. Показать несмещённость оценки $\theta^*=\dfrac{m}{n}$ вероятности $\theta=p$ выпадения герба в каждом опыте.

☺ *Решение*. Так как число успехов *m* имеет распределение Бернулли, то $M[m]=np$.

Следовательно, $M[\theta^*]=M\left[m/n\right]=\dfrac{1}{n}M[m]=\dfrac{1}{n}\cdot n\cdot p=p=\theta$.

Таким образом, оценка $\theta^*=\dfrac{m}{n}$ – *несмещённая*. ☻

1.6.5. Точечная оценка математического ожидания

Рассмотрим выборочную среднюю $\overline{x}_в$ как оценку генеральной средней $\overline{x}_г=M[X]$ – истинного значения распределения.

$$M[\overline{x}_в]=M\left[\frac{1}{n}\cdot\sum_{i=1}^{n}X_i\right]=\frac{1}{n}M\left[\sum_{i=1}^{n}X_i\right]=\frac{1}{n}\sum_{i=1}^{n}M[X_i]=\frac{1}{n}\sum_{i=1}^{n}M[X]=$$
$$=\frac{1}{n}\cdot n\cdot M[X]=M[X]=\overline{x}_г\Rightarrow$$

$\Rightarrow$ $\overline{x}_в$ – *несмещённая оценка* для $\overline{x}_г$.

По теореме Чебышева имеем $\lim\limits_{n\to\infty} P\left(\left|\frac{1}{n}\sum\limits_{i=1}^{n} X_i - x_г\right| < \varepsilon\right) = 1.$

$\Rightarrow$ $\overline{x}_в$ – *состоятельная оценка* для $\overline{x}_г$.

Можно показать, что при нормальном распределении случайной величины X эта оценка, т. е. $\overline{x}_в$, будет и эффективной.

На практике *во всех случаях в качестве оценки математического ожидания* $\overline{x}_г$ *используется среднее арифметическое, т. е.* $\overline{x}_в$.

1.6.6. Точечная оценка дисперсии при неизвестном математическом ожидании

Рассмотрим выборочную дисперсию $D_в$ как оценку генеральной дисперсии $D_г$ – истинного значения распределения.

Покажем, что $M[D_в] = \frac{n-1}{n} D[X]$, где $D[X]$ – дисперсия генеральной совокупности.

Действительно, $M[D_в] = M\left[\frac{1}{n}\cdot\sum\limits_{i=1}^{n}(X_i - \overline{x}_в)^2\right] = M\left[\overline{x^2} - (\overline{x}_в)^2\right] =$

$$= M\left[\frac{1}{n}\cdot\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\cdot\sum_{i=1}^{n} X_i\right)^2\right] = \frac{1}{n} M\left[\sum_{i=1}^{n} X_i^2\right] - \frac{1}{n^2} M\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] =$$

$$= \frac{1}{n} M\left[X_1^2 + X_2^2 + \ldots + X_n^2\right] - \frac{1}{n^2} M\left[(X_1 + X_2 + \ldots + X_n)^2\right] =$$

$$= \frac{1}{n}\left(M\left[X_1^2\right] + M\left[X_2^2\right] + \ldots + M\left[X_n^2\right]\right) -$$

$$- \frac{1}{n^2} M\left[\left(X_1^2 + X_2^2 + \ldots + X_n^2\right) + 2\cdot\underbrace{\left(X_1X_2 + X_1X_3 + X_2X_3 + \ldots + X_{n-1}X_n\right)}_{C_n^2}\right] =$$

$$= \frac{n-1}{n^2}\left(M\left[X_1^2\right] + M\left[X_2^2\right] + \ldots + M\left[X_n^2\right]\right) -$$

$$- \frac{2}{n^2}\left(M[X_1]M[X_2] + M[X_1]M[X_3] + M[X_2]M[X_3] + \ldots + M[X_{n-1}]M[X_n]\right) =$$

$$=\frac{n-1}{n^2}\underbrace{\left(M\left[X^2\right]+M\left[X^2\right]+\ldots+M\left[X^2\right]\right)}_{n}-$$

$$-\frac{2}{n^2}\left(M[X]\cdot M[X]+M[X]\cdot M[X]+M[X]\cdot M[X]+\ldots+M[X]\cdot M[X]\right)=$$

$$=\frac{n-1}{n^2}\cdot n\cdot M\left[X^2\right]-\frac{2}{n^2}\cdot\frac{n(n-1)}{2}\cdot(M[X])^2=\frac{n-1}{n}\cdot M\left[X^2\right]-\frac{n-1}{n}\cdot(M[X])^2=$$

$$=\frac{n-1}{n}\cdot\left(M\left[X^2\right]-(M[X])^2\right)=\frac{n-1}{n}\cdot D[X].$$

Так как $M[D_в]\neq D[X]$, то $D_в$ – *смещённая оценка* $D[X]$.

Смещение оценки произошло потому, что в формуле выборочной дисперсии $D_в=\frac{1}{n}\cdot\sum_{i=1}^{n}\left(X_i-\bar{x}_в\right)^2$ отклонение X_i отсчитывается не от истинного математического ожидания, а от его статистического аналога $\bar{x}_в$.

Чтобы исправить этот недостаток, вводят ***исправленную выборочную дисперсию***

$$\boxed{s^2=\frac{n}{n-1}\cdot D_в}, \qquad (23)$$

где s – ***исправленное выборочное среднее квадратическое отклонение***.

Так как $\lim\limits_{n\to\infty}s^2=\lim\limits_{n\to\infty}\frac{n}{n-1}\cdot D_в=\lim\limits_{n\to\infty}\frac{n}{n-1}\cdot\lim\limits_{n\to\infty}D_в=1\cdot D[X]=D[X]$, то получаем, что $D_в$ – *состоятельная оценка* $D[X]$.

Замечание. *При больших значениях n разница между $D_в$ и s^2 невелика и они практически равны, поэтому оценку s^2 чаще используют для оценки дисперсии при малых выборках, обычно при $n\leq 30$.*

Пример 12. Вычислить исправленную дисперсию и исправленное среднее квадратическое отклонение по данным примера 5.

☺ *Решение*. По результатам примера 5 имеем $D_в=87{,}48$, $\sigma_в=9{,}35$.

По формуле (23): $s^2=\frac{100}{99}\cdot 87{,}48\approx 88{,}36$, тогда $s=\sqrt{88{,}36}\approx 9{,}4$. ☻

1.7. Интервальные оценки параметров распределения

1.7.1. Точность и надёжность оценок числовых характеристик

Точечная оценка неизвестного параметра θ часто бывает достаточной для практических выводов в качестве первоначальных результатов обработки наблюдений. Однако если есть необходимость в более детальном анализе, то надо оценить насколько истинное значение параметра расходится с точечной оценкой этого значения. Вопрос о точности особенно существенен для выборок небольшого объёма, так как между θ и θ^* может быть большое расхождение. Кроме того, при решении практических задач часто необходимо определить и надёжность этих оценок. В этом случае возникает задача о приближении параметра θ не одним числом, а некоторым интервалом (θ_1^*, θ_2^*).

Интервал выбирается таким образом, чтобы вероятность включения в этот интервал параметра θ была достаточно велика (близка к единице). Говоря более строго, это означает, что вероятность выполнения двойного неравенства $\theta_1^* < \theta < \theta_2^*$ не меньше заданного числа γ.

Интервал (θ_1^*, θ_2^*), накрывающий с вероятностью γ истинное значение параметра θ, называется ***доверительным интервалом***, а вероятность γ – ***надёжностью оценки*** или ***доверительной вероятностью***.

Чаще всего (но не обязательно) доверительный интервал выбирается симметричным относительно несмещённой точечной оценки θ^*, т. е. выбирается интервал вида $(\theta^* - \varepsilon, \theta^* + \varepsilon)$ такой, что

$$P\left\{ \theta \in (\theta^* - \varepsilon, \theta^* + \varepsilon) \right\} = P\left\{ |\theta - \theta^*| < \varepsilon \right\} = \gamma. \qquad (24)$$

Число $\varepsilon > 0$ называется ***точностью оценки***: чем меньше разность $|\theta - \theta^*|$, тем точнее оценка.

Величина γ для формулы (24) выбирается заранее, её выбор зависит от конкретно решаемой задачи. Например, степень доверия авиапассажира к надёжности самолёта, очевидно, должна быть выше степени доверия покупателя к надёжности телевизора.

Надёжность γ принято выбирать равной 0,9; 0,95; 0,99 или 0,999.

Ещё раз подчеркнём, что интервальная оценка зависит не только от имеющихся данных, но и от требуемой надёжности. Так, например, при $\gamma = 0{,}95$ на 95 % можно быть уверенным, что доверительный интервал покрывает параметр θ, при $\gamma = 0{,}99$ на 99 % и т.д. Это значит, что если сделать много выборок, то для 95 % из них (если $\gamma = 0{,}95$) вычисленные доверительные интервалы действительно покроют θ.

Приведём неформальный пример, поясняющий различие точечной и интервальной оценок.

Когда о каком-либо человеке говорят: “Ему примерно 38 лет”, это ни что иное, как точечная оценка возраста. Когда же говорят: “Ему лет 35-40”, это интервальная оценка, доверительный интервал при этом (35; 40). Надёжность оценки при этом в явном виде не указывается, но предполагается довольно близкой к единице.

Иногда можно слышать и высказывания такого рода: “Ему на вид лет 35-40, по крайней мере, не больше 45 лет”. Очевидно, что доверительный интервал (35; 45) имеет бóльшую доверительную вероятность, чем интервал (35; 40). Однако интервальная оценка (35; 40) более информативна, чем интервальная оценка (35; 45).

1.7.2. Односторонние доверительные интервалы

Если интерес представляет ситуация, когда важно сравнение только с одним критическим значением, то используют односторонние доверительные интервалы: для заданного ***уровня доверия (надёжности)*** γ строят двусторонний доверительный интервал, который затем расширяют за счёт одной из его границ.

Для двустороннего доверительного интервала имеем:

$$\gamma = P\left(|\theta - \theta^*| < \varepsilon \right) = \int_{\theta^* - \varepsilon}^{\theta^* + \varepsilon} f(x)\,dx.$$

С геометрической точки зрения: *доверительная вероятность* γ *численно равна площади заштрихованной области под графиком дифференциальной функции* $f(x)$, *вычисленной на интервале* $(\theta^* - \varepsilon, \theta^* + \varepsilon)$ (рис. 14).

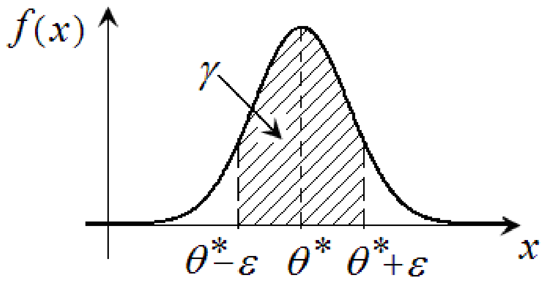


Рис. 14

Для одностороннего доверительного интервала справедливо

$$p = \int_{-\infty}^{\theta^*+\varepsilon} f(x)\,dx = \int_{\theta^*-\varepsilon}^{+\infty} f(x)\,dx = \gamma + \frac{1-\gamma}{2} = \frac{1+\gamma}{2} = \gamma'.$$

В результате получаем односторонние интервалы $(-\infty; u_p)$ и $(u_p; +\infty)$ с большей гарантией γ' (рис. 15, 16). Таким образом «односторонний» подход позволяет вдвое снизить ошибку $\alpha = 1 - \gamma$.

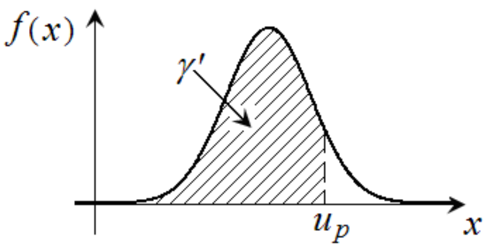


Рис. 15

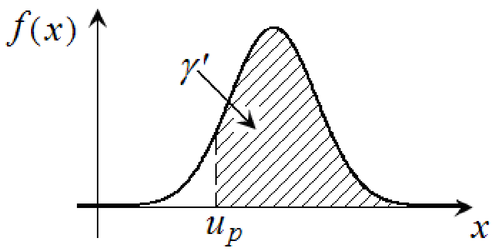


Рис. 16

Значение u_p, для которого выполняется $p = P\{x < u_p\} = \int_{-\infty}^{u_p} f(x)\,dx$, называется ***квантилью***.

Нахождение квантили u_p заключается в выборе такого значения, чтобы площадь заштрихованной области была равна p.

В статистике для обработки результатов эксперимента широко используют законы распределения Пирсона, Стьюдента, Фишера-Снедекора.

Квантили этих распределений табулированы (см. приложения).

1.8. Некоторые распределения функций нормальных случайных величин

1.8.1. Распределение χ^2 (хи-квадрат) Пирсона

Распределением χ^2 ***Пирсона*** с k степенями свободы называется распределение суммы квадратов нормально распределённых независимых случайных величин X_1, X_2, …, X_k с параметрами $a=0$, $\sigma=1$:

$$\chi^2(k)=X_1^2+X_2^2+\ldots+X_k^2.$$

Плотность этого распределения определяется формулой

$$f_{\chi_k^2}(x)=\begin{cases}\dfrac{x^{\frac{k}{2}-1}\cdot e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\cdot\Gamma(k/2)}, & \text{при } x>0,\\ 0, & \text{при } x\le 0,\end{cases}$$

где $\Gamma(x)=\int\limits_0^{+\infty}t^{x-1}e^{-t}dt$ – гамма-функция Эйлера $\left(\Gamma(x)=(x-1)!\right.$ для натуральных значений $x\left.\right)$.

Распределение χ^2 определяется только одним параметром – числом степеней свободы k. Графики функции $f_{\chi_k^2}(x)$ для различных значений k представлены на рис. 17. С увеличением числа степеней свободы k ($k\to\infty$) распределение χ^2 приближается к нормальному закону распределения (при $k>30$ различий практически нет).

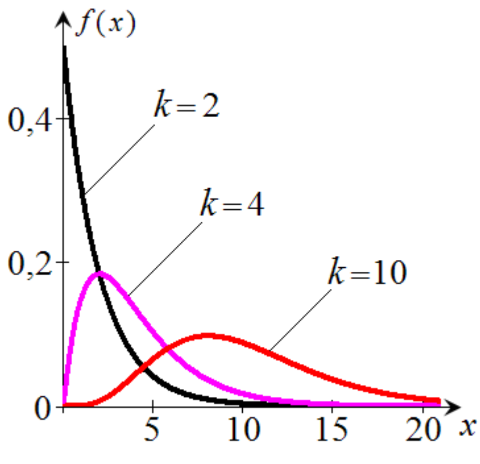


Рис. 17

Числовые характеристики распределения χ^2:

$$M[\chi^2]=k,\ D[\chi^2]=2k,\ a_s=\sqrt{\frac{8}{k}},\ \varepsilon_k=\frac{12}{k}.$$

На практике, как правило, используют не плотность вероятности, а квантили распределения χ_k^2.

1.8.2. Распределение Стьюдента

Пусть $X,\ X_1,\ X_2,\ \ldots,\ X_k$ – независимые случайные величины, имеющие стандартное нормальное распределение с параметрами $a=0$, $\sigma=1$. ***Распределением Стьюдента*** (или *t-распределением*) с k степенями свободы называется распределение отношения

$$T=\frac{X}{\sqrt{\frac{1}{k}(X_1^2+X_2^2+\ldots+X_k^2)}}=\frac{X}{\sqrt{\frac{\chi^2}{k}}}.$$

Плотность этого распределения определяется по формуле

$$f(x)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\cdot\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}},\quad -\infty<x<+\infty.$$

С увеличением значений k распределение Стьюдента достаточно быстро приближается к нормальному распределению (рис. 18).

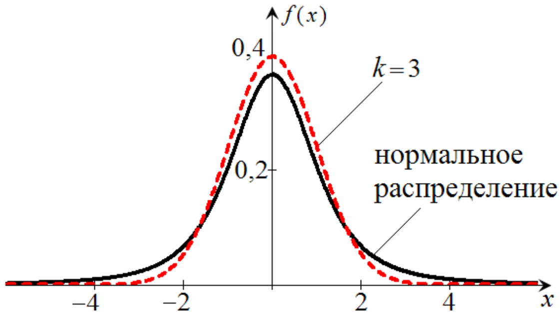


Рис. 18

Числовые характеристики распределения Стьюдента:

$$M[T]=0,\ D[T]=\frac{k}{k-2}\ (k>2),\ a_s=0,\ \varepsilon_k=\frac{6}{k-4}.$$

1.8.3. Распределение Фишера-Снедекора

F-распределением Фишера-Снедекора со степенями свободы k_1 и k_2 называется распределение отношения

$$F=\frac{\chi^2(k_1)/k_1}{\chi^2(k_2)/k_2}.$$

Плотность этого распределения вычисляется по формуле

$$f(x)=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\cdot\Gamma\left(\frac{k_2}{2}\right)}\cdot\frac{k_1^{\frac{k_1}{2}}\cdot k_2^{\frac{k_2}{2}}\cdot x^{\frac{k_1}{2}-1}}{(k_1x+k_2)^{\frac{k_1+k_2}{2}}},\quad x\ge 0,$$

График плотности F-распределения при $k_1 = 10$ и $k_2 = 15$ представлен на рис. 19.

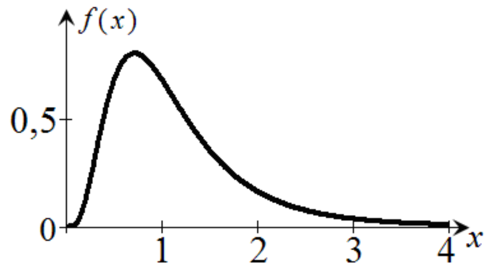


Рис. 19

Числовые характеристики распределения Фишера-Снедекора:

$$M[F] = \frac{k_2}{k_2 - 2}\ (k_2 > 2),\quad D[F] = \frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}\ (k_2 > 4).$$

1.9. Доверительные интервалы параметров нормального распределения

1.9.1. Доверительный интервал для оценки вероятности события A: $\boxed{p \in (p^* - \varepsilon; p^* + \varepsilon)}$

<u>I случай</u> (*неизвестен объём генеральной совокупности N*).

Пусть событие A наступило m раз в n испытаниях (n – объём выборки). Тогда $p^* = \frac{m}{n}$ – точечная оценка вероятности p наступления события в одном испытании.

По следствию из центральной предельной теоремы, имеем:

$$\gamma = P\{|p - p^*| < \varepsilon\} = 2\Phi\left(\varepsilon \cdot \sqrt{\frac{n}{pq}}\right) \quad \Rightarrow \quad t = \varepsilon \cdot \sqrt{\frac{n}{pq}} \quad \Rightarrow \quad \varepsilon = t \cdot \sqrt{\frac{pq}{n}},$$

где t – аргумент функции Лапласа $\Phi(t) = \frac{\gamma}{2}$.

Так как $p \approx p^*$, $q = 1 - p \approx 1 - p^*$, то

$$\boxed{\varepsilon = t \cdot \sqrt{\frac{p^* \cdot (1 - p^*)}{n}}}. \tag{25}$$

<u>II случай</u> (*известен объём генеральной совокупности N*).

$$\boxed{\varepsilon = t\cdot\sqrt{\frac{p^*\cdot(1-p^*)}{n}\cdot\left(1-\frac{n}{N}\right)}}. \qquad (26)$$

Отсюда можно получить формулу для нахождения объёма выборки

$$\boxed{n = \frac{p^*\cdot(1-p^*)}{\left(\frac{\varepsilon}{t}\right)^2 + \frac{p^*\cdot(1-p^*)}{N}}}. \qquad (27)$$

Замечание. Данные формулы справедливы при всех n, для которых справедливо неравенство $npq \approx np^*\cdot(1-p^*) > 9$.

1.9.2. Доверительный интервал для оценки математического ожидания $\bar{x}_г$: $\boxed{\bar{x}_г \in (\bar{x}_в - \varepsilon;\ \bar{x}_в + \varepsilon)}$

<u>I случай</u> (*известно среднее квадратическое отклонение* $\sigma_г$).

Определим случайную величину $X = \sum_{i=1}^{n} X_i$. Её среднее выборочное значение $\bar{x}_в = \frac{1}{n}\sum_{i=1}^{n} X_i$ представляет собой сумму сравнительно большого числа n независимых величин и, согласно центральной предельной теореме, имеет распределение, близкое к нормальному. При этом:

$M[\bar{x}_в] = \bar{x}_г$ (т. к. $\bar{x}_в$ – несмещённая оценка $\bar{x}_г$),

$$D[\bar{x}_в] = D\left[\frac{1}{n}\cdot\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\cdot\sum_{i=1}^{n} D[X_i] = \frac{1}{n^2}\cdot n\cdot D_г = \frac{D_г}{n} \quad \Rightarrow \quad \sigma_в = \frac{\sigma_г}{\sqrt{n}}.$$

Тогда $\gamma = P\left(\left|\bar{x}_г - \bar{x}_в\right| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma_г}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma_в}\right) \quad \Rightarrow \quad t = \frac{\varepsilon\sqrt{n}}{\sigma_в}$.

Отсюда получаем

$$\boxed{\varepsilon = \frac{t\cdot\sigma_в}{\sqrt{n}}}, \qquad (28)$$

где t – аргумент функции Лапласа $\Phi(t) = \frac{\gamma}{2}$ (приложение 2).

Из формулы (28) выразим величину *n*, тогда

$$\boxed{n=\left(\frac{t\cdot\sigma_{в}}{\varepsilon}\right)^{2}}. \tag{29}$$

Полученное выражение позволяет оценить, каков должен быть объём выборки, чтобы точность оценки $\left(\overline{x}_{г}\approx\overline{x}_{в}\right)$ не превосходила заданного значения ε с заданным уровнем доверия γ.

Пример 13. Произведено пять независимых наблюдений над случайной величиной $X\sim N(a,20)$. Результаты наблюдений таковы: $x_1=-25$, $x_2=34$, $x_3=-20$, $x_4=10$, $x_5=21$. Найти оценку для $a=M[X]$, а также построить для него 95%-й доверительный интервал.

☺ *Решение*. Находим $\overline{x}_{в}=\frac{1}{5}\cdot(-25+34-20+10+21)=4$.

Т. к. $\gamma=0{,}95$, то $\Phi(t)=\frac{\gamma}{2}=0{,}475$. Тогда по таблице приложения 2 находим $t=1{,}96$.

Значит по формуле (28) имеем: $\varepsilon=\frac{t\cdot\sigma_{в}}{\sqrt{n}}=\frac{1{,}96\cdot 20}{\sqrt{5}}\approx 17{,}5$.

Таким образом, доверительный интервал для $a=M[X]$ таков: $(4-17{,}5;4+17{,}5)$, т. е. $(-13{,}5;21{,}5)$. ☻

II случай (*неизвестны* $\sigma_{г}$ *и N – объём генеральной совокупности*).

В этом случае

$$\boxed{\varepsilon=\frac{t_{\gamma}\cdot s}{\sqrt{n}}}, \tag{30}$$

где s – несмещённая оценка $\sigma_{г}$; значение $t_{\gamma}=t(\gamma,n)$ − табулировано распределением Стьюдента (таблица приложения 6).

Выражая *n* из формулы (30), получим

$$\boxed{n=\left(\frac{t_{\gamma}\cdot s}{\varepsilon}\right)^{2}}. \tag{31}$$

С помощью формулы (31) можно найти объём выборки, необходимый для определения точности оценки $(\bar{x}_г \approx \bar{x}_в)$, которая не превосходит заданного значения ε с заданным уровнем доверия γ.

Пример 14. По условию примера 13, считая, что случайная величина $X \sim N(a,\sigma)$, построить для неизвестного $M[X]=a$ доверительный интервал при $\gamma = 0{,}95$.

☺*Решение*. Ранее вычислено $\bar{x}_в = 4$. Находим s:

$$s^2 = \frac{1}{4}\cdot\left((-25-4)^2+(34-4)^2+(-20-4)^2+(10-4)^2+(21-4)^2\right) = 660{,}5,$$

тогда $s \approx 25{,}7$.

По таблице приложения 6 для $\gamma = 0{,}95$ и $n = 5$ находим $t_\gamma = 2{,}78$.

Значит по формуле (30) имеем: $\varepsilon = \dfrac{t_\gamma \cdot s}{\sqrt{n}} = \dfrac{2{,}78\cdot 25{,}7}{\sqrt{5}} \approx 31{,}9$.

Таким образом, получаем доверительный интервал $(-27{,}9; 35{,}9)$. ☻

__***III случай***__ (*неизвестно* $\sigma_г$, *известно* N).

Величина ε вычисляется по формуле

$$\boxed{\varepsilon = \frac{t_\gamma \cdot s}{\sqrt{n}}\sqrt{\left(1-\frac{n}{N}\right)}}. \tag{32}$$

Используя (32), можно найти объём выборки

$$\boxed{n = \frac{N\cdot t_\gamma^2 \cdot s^2}{t_\gamma^2\cdot s^2 + N\cdot\varepsilon^2}}, \tag{33}$$

необходимый для определения точности оценки $(\bar{x}_г \approx \bar{x}_в)$, которая не превосходит заданного значения ε с заданным уровнем доверия γ.

1.9.3. Доверительный интервал для оценки генерального среднего квадратического отклонения

Пусть для выборки объёма n задана надёжность, с которой нужно оценить отклонение найденного исправленного квадратического отклонения s от истинного $\sigma_г$: $\gamma = P\left(|\sigma_г - s| < \varepsilon \right)$.

Тогда $|\sigma_г - s| < \varepsilon \;\Rightarrow\; s - \varepsilon < \sigma_г < s + \varepsilon \;\Rightarrow\; s \cdot \left(1 - \dfrac{\varepsilon}{s}\right) < \sigma_г < s \cdot \left(1 + \dfrac{\varepsilon}{s}\right)$.

Введём $q = \dfrac{\varepsilon}{s} = q(\gamma; n)$, которое табулировано (приложение 7).

Учитывая, что $\sigma_г \geq 0$, получим:

если $q < 1$,	то $\sigma_г \in (s - sq;\; s + sq)$,	(34)
если $q > 1$,	то $\sigma_г \in (0;\; s + sq)$.	(35)

Пример 15. Количественный признак генеральной совокупности распределён нормально. По выборке объёма $n = 25$ найдено исправленное среднее квадратическое отклонение $s = 0{,}8$. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение $\sigma_г$ с надёжностью $0{,}95$.

☺ *Решение*. По таблице приложения 7 при данных значениях:

$$\gamma = 0{,}95,\; n = 25,$$

находим $q = 0{,}32$.

Т. к. $q = 0{,}32 < 1$, то доверительный интервал следует искать по формуле (30):

$$0{,}8 - 0{,}8 \cdot 0{,}32 < \sigma_г < 0{,}8 + 0{,}8 \cdot 0{,}32$$

Таким образом, получаем доверительный интервал $(0{,}544; 1{,}056)$. ☻

1.10. Проверка статистических гипотез

1.10.1. Статистическая гипотеза

Обычно в практических задачах не встречаются случайные величины, распределения которых точно соответствовали бы теоретическим распределениям. Последние представляют собой математические модели реальных распределений. Подбор таких моделей и анализ их адекватности моделируемым случайным величинам являются одной из основных задач математической статистики. Эта задача в свою очередь сводится к проверке предположений (гипотез) о виде модели распределения и её параметрах.

Например, верно ли, что новое лекарство эффективнее, чем применявшееся ранее, что новые методы обучения лучше, чем старые и т. д.

Статистической гипотезой (или просто ***гипотезой***) называется любое предположение о генеральной совокупности, проверяемое по выборке.

Например, гипотезами являются: предположение о виде неизвестного распределения, о параметрах известных распределений, об отношениях между случайными величинами и т. д.

Если гипотеза содержит некоторое утверждение о параметрах распределения случайной величины (когда сам закон распределения считается известным), то она называется ***параметрической***, и ***непараметрической*** – в иных случаях.

Нулевой (***основной***) гипотезой H_0 называется предположение, которое выдвигается изначально, пока наблюдения не заставят признать обратное.

Альтернативной (***конкурирующей***) гипотезой H_1 называется гипотеза, которая противоречит нулевой гипотезе H_0 и которую принимают, если отвергнута основная гипотеза.

Гипотезы бывают ***простые*** (содержащие только одно предположение) и ***сложные*** (состоящие из конечного или бесконечного числа простых гипотез).

Например, гипотеза H_0, состоящая в предположении, что математическое ожидание нормального распределения $a = 5$, является простой, тогда в качестве альтернативной гипотезы можно рассматривать одну из следующих: H_1: $a > 5$ (сложная), H_2: $a < 5$ (сложная), H_3: $a \neq 5$ (сложная).

1.10.2. Задачи статистической проверки гипотез

Процедура сопоставления высказанного предположения (гипотезы) с выборочными данными называется ***проверкой гипотез***.

***Задачи статистической проверки гипотез*:**

- Относительно некоторой генеральной совокупности высказывается та или иная гипотеза *H*.
- Из этой генеральной совокупности извлекается выборка.
- Необходимо указать правило, с помощью которого можно было по выборке ответить на вопрос о том, следует ли отклонить гипотезу *H* или принять её.

Отметим, что статистическими методами ***гипотезу можно только опровергнуть или не опровергнуть***, но не доказать.

Например, для проверки утверждения автора, что в рукописи ошибок нет (гипотеза *H*), рецензент прочёл (изучил) несколько страниц рукописи. Если он обнаружил хотя бы одну ошибку, то гипотеза *H* отвергается, в противном случае – не отвергается, тогда говорят, что результат проверки согласуется с гипотезой, хотя ошибка могла иметь место и в непроверенных страницах.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость её проверки.

1.10.3. Общая схема проверки статистических гипотез

Имея две гипотезы H_0 и H_1, необходимо на основе выборочных данных либо принять основную гипотезу H_0, либо конкурирующую H_1.

Правило, по которому принимается решение принять или отклонить гипотезу H_0 (или H_1), называется ***статистическим критерием*** (или просто ***критерием***) проверки гипотезы H_0.

Статистикой (или ***тестом***) критерия называют случайную величину τ, которая служит для проверки статистических гипотез.

Приведём **схему проверки статистических гипотез:**

1. Для основной гипотезы H_0 формулируется альтернативная гипотеза H_1.

2. Выбирается ***уровень значимости*** проверки – малое число $\alpha > 0$.

3. Рассматриваются теоретические выборки значений случайных величин, о которых сформулирована гипотеза H_0, и выбирается (формируется) случайная величина τ. Значения и распределение τ (обычно из перечисленных ниже: *U* – нормальное распределение, χ^2 – распределение Пирсона, *T* – Стьюдента, *F* – Фишера-Снедекора) полностью определяются по выборкам при предположении о верности гипотезы H_0.

4. На числовой оси задают интервал *D*, такой, что вероятность попадания случайной величины τ в этот интервал: $P(\tau \in D) = 1 - \alpha$.

Интервал *D* называется ***областью принятия гипотезы*** H_0, а оставшаяся область числовой оси – ***критической областью*** (величина $\tau = \tau_{кр}$ – ***критическое значение теста проверки***).

Различают три типа критических областей. Критическая область определяется с учётом гипотез:

H_0	H_1	*Критическая область D*	
$\theta = \tilde{\theta}$	$\theta < \tilde{\theta}$	$(-\infty; -\tau_{кр}]$	– *левосторонняя* (рис. 20)
	$\theta \neq \tilde{\theta}$	$(-\infty; -\tau_{кр}] \cup [\tau_{кр}; +\infty)$	– *двусторонняя* (рис. 21)
	$\theta > \tilde{\theta}$	$[\tau_{кр}; +\infty)$	– *правосторонняя* (рис. 22)

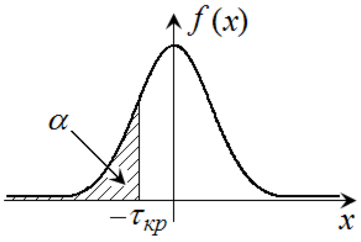


Рис. 20

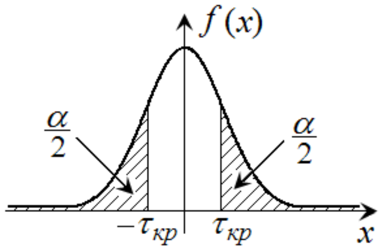


Рис. 21

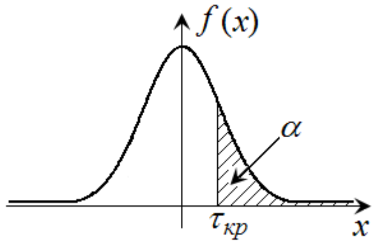


Рис. 22

Соответственно интервалам критерий проверки называется ***правосторонним***, ***двусторонним*** или ***левосторонним***.

5. По реализациям анализируемых выборок вычисляется конкретное (наблюдаемое) значение теста τ (обозначим его $\tau = \tau_{набл}$) и проверяется выполнение условия $P(\tau \in D) = 1 - \alpha$:

а) если оно выполняется (например, $\tau_{набл} < \tau_{кр}$ для правосторонней области), то гипотеза H_0 принимается в том смысле, что она не противоречит опытным данным и нет оснований её отвергнуть;

б) если условие не выполняется ($\tau_{набл} > \tau_{кр}$ для правосторонней области), то полагается, что гипотеза H_0 неверна и её отвергают.

Для каждого критерия имеются соответствующие таблицы, по которым и находят критическое значение, удовлетворяющее приведённым выше соотношениям.

Принцип принятия статистической гипотезы не даёт логического доказательства её верности или неверности. Принятие гипотезы H_0 в сравнении с альтернативной H_1 не означает, что мы уверены в абсолютной правильности H_0, или, что высказанное в гипотезе H_0 утверждение является наилучшим, единственно подходящим. Просто гипотеза H_0 не противоречит имеющимся у нас выборочным данным. Таким же свойством наряду с H_0 могут обладать и другие гипотезы. Более того, возможно, что при увеличении объёма выборки *n* или при испытании H_0 против другой альтернативной гипотезы H_2 гипотеза H_0 будет отвергнута.

Таким образом, *принятие гипотезы* H_0 *следует расценивать не как раз и навсегда установленный, абсолютно верный содержащийся в ней факт, а лишь как достаточно правдоподобное, не противоречащее опыту утверждение*.

Из представленной схемы следует, что при проверке гипотезы H_0 может быть принято неправильное решение, т. е. могут быть допущены ошибки двух видов:

ошибка* I *рода	***ошибка* II *рода***
Отвергается основная (нулевая) гипотеза, хотя она верна.	Отвергается конкурирующая гипотеза, хотя она верна.
Вероятность ошибки $P(H_1 \mid H_0) = \alpha$, α – ***уровень значимости критерия*** (обычно $\alpha = 0{,}05$; 0,01; 0,005; 0,001).	Вероятность ошибки $P(H_0 \mid H_1) = \beta$ (величина β, как правило, заранее неизвестна)
Вероятность принять верную (нулевую) гипотезу $P(H_0 \mid H_0) = 1 - \alpha$.	Вероятность принять верную (конкурирующую) гипотезу $P(H_1 \mid H_1) = 1 - \beta$, $(1 - \beta)$ – ***мощность критерия***.

Последствия ошибок 1-го и 2-го рода могут быть абсолютно различными: в одних случаях надо минимизировать α, а в других – β. Так, применительно к радиолокации говорят, что α – вероятность пропустить сигнал, β – вероятность ложной тревоги. Применительно к производству, к торговле можно сказать, что α – риск поставщика (т. е. забраковка по всей партии изделий, удовлетворяющих стандарту), β – риск потребителя (т. е. приём по выборке всей партии изделий, не удовлетворяющих стандарту). Применительно к судебной системе, ошибка 1-го рода приводит к оправданию виновного, ошибка 2-го – осуждение невиновного.

Следует отметить, что одновременное уменьшение ошибок 1-го и 2-го рода возможно лишь при увеличении объёма выборок. Поэтому обычно при заданном уровне значимости α отыскивается критерий с наибольшей мощностью.

1.10.4. Статистики сравнения точечных оценок неизвестных генеральных

1) Проверка гипотез для одной выборки

Пусть генеральная совокупность X распределена по нормальному закону.

Генеральная средняя a хотя и неизвестна, но есть основания предполагать, что она равна предполагаемому значению a_0.

Например, если X – совокупность размеров x_i партии деталей, производящихся станком автоматической линии, то можно предположить, что генеральная средняя a этих размеров равна проектному размеру a_0.

Для того, чтобы проверить правильность настройки этого станка, очевидно надо убедиться в том, что среднее значение параметра у производимых на нём изделий будет соответствовать номиналу.

Таким образом, необходимо проверить гипотезу $H_0 : a = a_0$ против альтернативной:

$$H_1 : a \neq a_0, \text{ или } H_2 : a < a_0, \text{ или } H_3 : a > a_0.$$

Если различие окажется незначимым, то станок обеспечивает в среднем проектный размер; если различие значимое, то станок требует наладки.

При произвольной настройке станка может возникнуть необходимость проверки гипотезы о том, что точность изготовления изделий по данному параметру, задаваемая дисперсией σ^2, равна заданной величине σ_0^2 $\left(H_0 : \sigma^2 = \sigma_0^2\right)$, или например, того, что доля бракованных изделий производимых станком, равна заданной величине p_0 $(H_0 : p = p_0)$ и т. д.

Выдвигаемые гипотезы и соответствующие критерии проверки гипотез о числовых значениях параметров нормального закона приведены в табл. 7.

Таблица 7

H_0	Статистика критерия	H_1	Область принятия H_0
$a = a_0$ $\sigma^2 = \sigma_г^2$ известно	$U = \dfrac{\bar{x} - a_0}{\sigma}\sqrt{n}$	$a \neq a_0$	$\lvert U \rvert < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \dfrac{\alpha}{2}$
		$a < a_0$	$U > -u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
		$a > a_0$	$U < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
$a = a_0$ $\sigma^2 = \sigma_г^2$ неизвестно	$T = \dfrac{\bar{x} - a_0}{s}\sqrt{n}$	$a \neq a_0$	$\lvert T \rvert < t_{кр}$, $t_{кр} = t_{\alpha, n-1}$ для двусторонней области
		$a < a_0$	$T > -t_{кр}$, $t_{кр} = t_{\alpha, n-1}$ для односторонней области
		$a > a_0$	$T < t_{кр}$, $t_{кр} = t_{\alpha, n-1}$ для односторонней области
$\sigma^2 = \sigma_0^2$ a неизвестно	$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$	$\sigma^2 \neq \sigma_0^2$	$\chi^2_{1-\frac{\alpha}{2}; n-1} < \chi^2 < \chi^2_{\frac{\alpha}{2}; n-1}$
		$\sigma^2 < \sigma_0^2$	$\chi^2 > \chi^2_{1-\alpha; n-1}$
		$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{\alpha; n-1}$
$p = p_0$ достаточно большие n, $np_0 > 5$, $nq_0 > 5$, $q_0 = 1 - p_0$	$U = \dfrac{p^* - p_0}{\sqrt{p_0 q_0}}\sqrt{n}$, $p^* = \dfrac{m}{n}$	$p \neq p_0$	$\lvert U \rvert < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \dfrac{\alpha}{2}$
		$p < p_0$	$U > -u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
		$p > p_0$	$U < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$

Примечание. Критические значения статистик на уровне значимости α определяют по соответствующим таблицам приложения.

Пример 16. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 5$ извлечена выборка объёма

$n = 100$, и по ней найдено выборочное среднее 26,5. Требуется на уровне значимости 0,05 проверить гипотезу $H_0 : a = a_0 = 25$ против альтернативной гипотезы $H_1 : a \neq a_0$. Изменится ли результат, если изменить альтернативную гипотезу на $H_1 : a > a_0$?

☺ *Решение*. Найдём значение статистики критерия (табл. 7)

$$U = \frac{\bar{x} - a_0}{\sigma}\sqrt{n} = \frac{26{,}5 - 25}{5}\sqrt{100} = 3.$$

При проверке гипотезы $H_1 : a \neq 25$ по таблице приложения 2 из соотношения $\Phi(u_{кр}) = 0{,}5 - \frac{0{,}05}{2} = 0{,}475$ находим $u_{кр} = 1{,}96$.

Т. к. $|U| > u_{кр}$, то основная гипотеза отвергается.

При проверке гипотезы $H_1 : a > a_0$ из соотношения $\Phi(u_{кр}) = 0{,}5 - 0{,}05 = 0{,}45$ находим (таблица приложения 1) $u_{кр} = 1{,}65$, значит $U > u_{кр}$. Следовательно, основная гипотеза отвергается.

В обоих случаях результат одинаков. ☻

Пример 17. По выборке объёма $n = 16$, извлечённой из нормальной генеральной совокупности, найдены $\bar{x} = 12{,}4$ и $s = 1{,}2$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : a = 11{,}8$ при конкурирующей гипотезе $H_1 : a \neq 11{,}8$.

☺ *Решение*. Найдём наблюдаемое значение статистики критерия (табл. 7)

$$T = \frac{\bar{x} - a_0}{s}\sqrt{n} = \frac{12{,}4 - 11{,}8}{1{,}2}\sqrt{16} = 2.$$

Поскольку конкурирующая гипотеза имеет вид $a \neq a_0$, то искомая критическая область двусторонняя. Из таблицы критических точек распределения Стьюдента (таблица приложения 4) найдём по уровню значимости

$\alpha = 0{,}05$ и числу степеней свободы $k = n - 1 = 15$ критическую точку $t_{кр} = t_{кр}(0{,}05;15) = 2{,}13$.

Т. к. $|T| < t_{кр}$, то нет оснований отвергнуть нулевую гипотезу. ☻

Пример 18. Точность работы станка-автомата проверяется по дисперсии размеров изделий, которая не должна превышать $\sigma_0^2 = 0{,}01$ (мм^2). По выборке из 25 изделий получена исправленная выборочная дисперсия $s^2 = 0{,}02$ (мм^2). На уровне значимости 0,05 проверить, обеспечивает ли станок необходимую точность?

☺ *Решение*. Найдём значение статистики критерия (табл. 7)

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{24 \cdot 0{,}02}{0{,}01} = 48.$$

По таблице приложения 3 находим критическую точку распределения χ^2-квадрат: $\chi^2_{0{,}05;24} = 36{,}4$. Т. к. $48 > 36{,}4$, то основная гипотеза отвергается. Следовательно, станок не обеспечивает необходимой точности. ☻

2) Проверка гипотез для двух независимых выборок

Пусть имеются две независимые нормально распределённые выборки $x_1, x_2, \ldots, x_{n_1}$ и $y_1, y_2, \ldots, y_{n_2}$ с параметрами (a_x, σ_x^2) и (a_y, σ_y^2) соответственно. Обычно ставится задача проверки их однородности, т. е. равенства обоих параметров, либо следует проверить равенство параметров по отдельности.

Сравнение средних двух совокупностей имеет важное практическое значение. На практике часто встречается случай, когда средний результат одной серии экспериментов отличается от среднего результата другой серии. При этом возникает вопрос, можно ли объяснять обнаруженное расхождение средних неизбежными случайными ошибками эксперимента или оно вызвано некоторыми закономерностями. В промышленности задача сравнения средних часто возникает при выборочном контроле качества изделий, изготовленных на разных установках или при различных технологических режимах, в финансовом анализе – при сопоставлении уровня доходности различных активов и т. д.

Гипотеза о *равенстве средних при известных дисперсиях* проверяется обычно в случае больших выборок (объёмом порядка сотен), когда оценки дисперсий можно принять за их точные значения.

Гипотеза о *равенстве средних при неизвестных дисперсиях* требует вначале проверки гипотезы о равенстве дисперсий двух выборок.

Гипотезы о дисперсиях возникают достаточно часто, так как дисперсия характеризует такие исключительно важные показатели, как точность машин, приборов, технологических процессов, степень однородности совокупностей, риск, связанный с отклонением доходности активов от ожидаемого уровня, и т. д.

Сравнение долей признака в двух совокупностях – достаточно часто встречающаяся на практике задача. Например, если выборочная доля признака в одной совокупности отличается от такой же доли в другой совокупности, то указывает ли это на то, что наличие признака в одной совокупности действительно вероятнее, или полученное расхождение долей является случайным?

Сформулируем задачу. Имеются две совокупности *X* и *Y*, генеральные доли признака в которых равны соответственно p_x и p_y. Необходимо проверить нулевую гипотезу о равенстве генеральных долей. Для проверки гипотезы H_0 из этих совокупностей взяты две независимые выборки достаточно большого объёма n_1 и n_2. Выборочные доли признака равны соответственно $p_1^* = \dfrac{m_1}{n_1}$ и $p_2^* = \dfrac{m_2}{n_2}$, где m_1 и m_2 – соответственно число элементов первой и второй выборок, обладающих данным признаком.

Выдвигаемые гипотезы и соответствующие критерии проверки гипотез представлены в табл. 8.

Таблица 8

H_0	Статистика критерия	H_1	Область принятия H_0
$a_x = a_y$ σ_x^2 и σ_y^2 известны	$U = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n_1} + \dfrac{\sigma_y^2}{n_2}}}$	$a_x \neq a_y$	$\lvert U \rvert < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \dfrac{\alpha}{2}$
		$a_x < a_y$	$U > -u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
		$a_x > a_y$	$U < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
$a_x = a_y$ σ_x^2 и σ_y^2 неизвестны, но равны	$T = \dfrac{\bar{x} - \bar{y}}{s \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$, $s = \sqrt{\dfrac{s_x^2 \cdot (n_1 - 1) + s_y^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}}$	$a \neq a_0$	$\lvert T \rvert < t_{кр}$, $t_{кр} = t_{\alpha, n_1 + n_2 - 1}$ для двусторонней области
		$a < a_0$	$T > -t_{кр}$, $t_{кр} = t_{\alpha, n_1 + n_2 - 1}$ для односторонней области
		$a > a_0$	$T < t_{кр}$, $t_{кр} = t_{\alpha, n_1 + n_2 - 1}$ для односторонней области
$\sigma_x^2 = \sigma_y^2$ a_x и a_y неизвестны	$F = \dfrac{s_{\max}^2}{s_{\min}^2}$	$\sigma_x^2 \neq \sigma_y^2$	$F < F_{кр}$, $F_{кр} = F_{\alpha/2,\, n_1 - 1,\, n_2 - 1}$
		$\sigma_x^2 > \sigma_y^2$	$F < F_{кр}$, $F_{кр} = F_{\alpha,\, n_1 - 1,\, n_2 - 1}$
$p_x = p_y$ n_1 и n_2 достаточно большие	$U = \dfrac{p_1^* - p_2^*}{\sqrt{p^*(1 - p^*)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, $p_1^* = \dfrac{m_1}{n_1}$, $p_2^* = \dfrac{m_2}{n_2}$, $p^* = \dfrac{m_1 + m_2}{n_1 + n_2}$	$p_x \neq p_y$	$\lvert U \rvert < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \dfrac{\alpha}{2}$
		$p_x < p_y$	$U > -u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$
		$p_x > p_y$	$U < u_{кр}$, $\Phi(u_{кр}) = 0{,}5 - \alpha$

Пример 19. Для проверки эффективности новой технологии отобраны две группы рабочих: в первой группе численностью $n_1 = 50$ чел., где применялась новая технология, выборочная средняя выработка составила $\bar{x} = 85$ (изделий), во второй группе численностью $n_2 = 70$ чел. выбо-

рочная средняя – $\bar{y}=78$ (изделий). Предварительно установлено, что дисперсии выработки в группах равны соответственно $\sigma_x^2=100$ и $\sigma_y^2=74$. На уровне значимости $\alpha=0{,}05$ выяснить влияние новой технологии на среднюю производительность.

☺*Решение*. Проверяемая гипотеза H_0: $a_x=a_y$, т. е. средние выработки рабочих одинаковы по новой и старой технологиям. В качестве конкурирующей гипотезы можно взять H_1: $a_x>a_y$ или H_2: $a_x\neq a_y$ (в данной задаче более естественна гипотеза H_1, т. к. её справедливость означает эффективность применения новой технологии).

Находим фактическое значение статистики критерия (табл. 8)

$$U=\frac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n_1}+\dfrac{\sigma_y^2}{n_2}}}=\frac{85-78}{\sqrt{\dfrac{100}{50}+\dfrac{74}{70}}}=4.$$

При альтернативной гипотезе H_1 по таблице приложения 2 из соотношения $\Phi(u_{кр})=0{,}5-0{,}05=0{,}45$ найдём критическое значение $u_{кр}=1{,}64$.

При конкурирующей гипотезе H_2 найдём критическое значение из соотношения $\Phi(u_{кр})=0{,}5-\dfrac{0{,}05}{2}=0{,}475$, тогда $u_{кр}=1{,}96$.

Т. к. $U>u_{кр}$ при любой из взятых конкурирующих гипотез, то гипотеза H_0 отвергается. Т. е. на 5%-ном уровне значимости можно сделать вывод, что новая технология позволяет повысить среднюю выработку рабочих. ☻

Пример 20. Реклама утверждает, что из двух типов пластиковых карт «Русский экспресс» и «Супер-понт» богатые люди предпочитают первый. С целью проверки этого утверждения были обследованы среднемесячные платежи $n_1=16$ обладателей «Русского экспресса» и $n_2=11$ обладателей «Супер-понта». При этом выяснилось, что платежи по картам «Русский экспресс» составляют в среднем 563 долл. с исправленным средним квадратическим отклонением 178 долл., а по картам «Супер-понт» – в среднем 485 долл. с исправленным средним квадратическим отклонением 196 долл.

Предварительный анализ законов распределения месячных расходов как среди обладателей «Русского экспресса», так и среди обладателей «Супер-понта» показал, что они достаточно хорошо описываются нормальным приближением.

Проверить утверждение рекламы на уровне значимости 10 %.

☺ *Решение*. В этом случае следует проверить гипотезу о средних при неизвестных дисперсиях (объёмы выборок малы). Поэтому, прежде всего, необходимо проверить гипотезу о равенстве дисперсий. Имеем (табл. 8):

$$F = \frac{s_{\max}^2}{s_{\min}^2} = \frac{196^2}{178^2} = \frac{38416}{31684} = 1{,}21.$$

Из таблицы критических значений Фишера-Снедекора (приложение 5) по уровню значимости $\alpha/2 = 0{,}05$ и числам степеней свободы $k_1 = n_{\max} - 1 = 10$ и $k_2 = n_{\min} - 1 = 15$ ($n_{\max}$ и $n_{\min}$ соответствуют $s_{\max}^2$ и $s_{\min}^2$) находим критическую точку $F_{кр} = 2{,}55$. Поскольку $1{,}21 < 2{,}55$, принимаем гипотезу о равенстве дисперсий двух выборок.

Теперь можно воспользоваться критерием Стьюдента для проверки гипотезы о равенстве средних. Имеем

$$s = \sqrt{\frac{s_x^2 \cdot (n_1 - 1) + s_y^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}} = \sqrt{\frac{38416 \cdot 10 + 31684 \cdot 15}{11 + 16 - 2}} = 185{,}4.$$

Вычисление статистики критерия даёт

$$T = \frac{\overline{x} - \overline{y}}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{563 - 485}{185{,}4\sqrt{\frac{1}{11} + \frac{1}{16}}} = 1{,}07.$$

По таблице критических точек распределения Стьюдента (приложение 4) для односторонней области по уровню значимости $\alpha = 0{,}1$ и числу степеней свободы 25 находим $t_{кр} = 1{,}32$.

Поскольку $T < t_{кр}$, то принимается основная гипотеза (о равенстве средних). Таким образом, утверждение рекламы не подтверждается имеющимися данными. ☻

Пример 21. В партии из 500 деталей, изготовленных первым станком-автоматом, оказалось 60 нестандартных, из 600 деталей второго станка – 42 нестандартных. На уровне значимости $\alpha = 0{,}01$ проверить нулевую гипотезу $H_0 : p_x = p_y$ о равенстве вероятностей изготовления нестандартной детали обоими станками против конкурирующей гипотезы $H_1 : p_x \neq p_y$.

☺ *Решение*. По условию имеем:

$$p_1^* = \frac{60}{500} = 0{,}12,\ p_2^* = \frac{42}{600} = 0{,}07,\ p^* = \frac{60+42}{500+600} = 0{,}09.$$

Находим значение статистики критерия (табл. 8)

$$U = \frac{p_1^* - p_2^*}{\sqrt{p^*(1-p^*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0{,}12 - 0{,}07}{\sqrt{0{,}09 \cdot 0{,}91 \cdot \left(\frac{1}{500} + \frac{1}{600}\right)}} = 2{,}85.$$

Критическую точку находим из соотношения $\Phi(u_{кр}) = 0{,}495$, откуда $u_{кр} = 2{,}57$ (таблица приложения 2). Так как $|U| > u_{кр}$, то гипотеза H_0 отвергается. Т. е. вероятности изготовления нестандартных деталей на двух станках различны. ☻

1.11. Построение теоретического закона распределения случайной величины по опытным данным. Проверка гипотез о законе распределения

Одной из важнейших задач математической статистики является *установление теоретического закона распределения случайной величины*, характеризующей изучаемый признак по опытному (эмпирическому) распределению, представляющему вариационный ряд. Для решения этой задачи *необходимо определить вид и параметры закона распределения*.

Гипотеза о виде закона распределения может быть выдвинута исходя из теоретических предпосылок, опыта аналогичных предшествующих исследований и, наконец, на основании графического изображения эмпирического распределения.

Параметры распределения, как правило, неизвестны, поэтому их заменяют наилучшими оценками по выборке ($\bar{x}$, $D_в$, $\sigma_в$ и т.д.).

Распределением, играющим основную роль в теории статистического оценивания, является нормальное распределение. Приведём его основные характеристики (табл. 9), рассмотренные в курсе теории вероятностей.

Таблица 9

Нормальное распределение ($a=\bar{x}_в$, $\sigma=\sigma_в$ или $\sigma=s$)		
Функция плотности	Функция распределения	Вероятность попадания в интервал
$f(x)=\frac{1}{\sigma}\cdot\varphi(t)$, где $t=\frac{x-a}{\sigma}$, $\varphi(t)$ – функция Гаусса	$F(x)=0{,}5+\Phi(t)$, где $t=\frac{x-a}{\sigma}$, $\Phi(t)$ – функция Лапласа	$p_i = P\{x_{i-1} < x < x_i\} = \frac{x_i-x_{i-1}}{\sigma}\cdot\varphi\left(\frac{\tilde{x}_i-a}{\sigma}\right) = \Phi\left(\frac{x_i-a}{\sigma}\right)-\Phi\left(\frac{x_{i-1}-a}{\sigma}\right)$, где $i=1,2,\ldots,k$.

Графики функций $\varphi(t)$ (рис. 23) и $\Phi(t)$ (рис. 24):

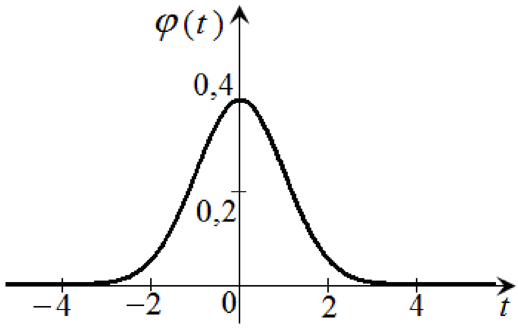


Рис. 23. Функция Гаусса

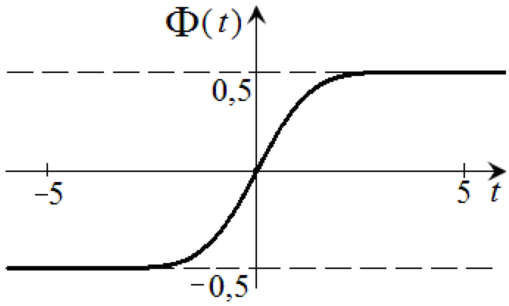


Рис. 24. Функция Лапласа

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \text{ (приложение 1)} \qquad \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt \text{ (приложение 2)}$$

Широкое использование в статистических выводах нормального распределения имеет как эмпирическое, так и теоретическое обоснования. Многочисленные примеры построения гистограмм и сглаживание их непрерывными кривыми для экспериментальных данных самой различной природы показывают, что во многих случаях нормальное распределение является довольно точным представлением таких данных. Применимость нормального распределения обосновывается и центральной предельной теоремой.

Как бы хорошо ни был выбран теоретический закон распределения, неизбежны расхождения между эмпирическим и теоретическим распределениями. Естественно, возникает вопрос: эти расхождения объясняются только случайными обстоятельствами, связанными с небольшим числом наблюдений, или они являются существенными и связаны с тем, что теоретический закон распределения подобран неудачно. Для ответа на этот вопрос используем специально подобранную величину – критерий согласия.

Критерием согласия называют статистический критерий проверки гипотезы о предполагаемом законе неизвестного распределения. Он используется для проверки согласия предполагаемого вида распределения с опытными данными на основании выборки.

Пусть необходимо проверить гипотезу H_0 о том, что рассматриваемая случайная величина X подчиняется определённому закону распределения. Для проверки гипотезы H_0 выбирают некоторую случайную величину τ, характеризующую степень расхождения теоретического и эмпирического распределений, закон распределения которой при достаточно больших n известен и практически не зависит от закона распределения случайной величины X.

Зная закон распределения τ, можно найти такое критическое значение $\tau = \tau_{кр}$, что если гипотеза H_0 верна, то вероятность того, что τ приняла значение больше чем τ_α: $P(\tau > \tau_\alpha) = \alpha$ – мала, где α – уровень значимости критерия.

Если наблюдаемое в опыте значение τ окажется больше критического: $\tau = \tau_{набл} > \tau_{кр}$ (т. е. попадёт в критическую область), то в соответствии с принципом практической уверенности это означает, что такие большие значения практически τ невозможны и противоречат гипотезе H_0. В этом случае нулевую гипотезу отвергают.

Если $\tau = \tau_{набл} < \tau_{кр}$, то расхождение между эмпирическим и теоретическим распределениями несущественно и гипотезу H_0 можно считать правдоподобной или, по крайней мере, не противоречащей опытным данным.

1.11.1. Критерий χ^2 Пирсона

Существуют различные критерии согласия: Пирсона, Колмогорова, Фишера, Смирнова и др. Критерий согласия Пирсона – наиболее часто употребляемый критерий для проверки простой гипотезы о законе распределения.

Эмпирические частоты – это частоты m_i, наблюдаемые в эксперименте.

Выравнивающие (теоретические) частоты – это частоты, которые находятся по формуле $m_i' = n \cdot p_i$, где n – объём выборки, p_i – точечная вероятность варианты x_i дискретной случайной величины или интервальная вероятность для варианты $x \in (x_{i-1}; x_i]$ непрерывной случайной величины.

Алгоритм действий:

1. Выбрать закон распределения случайной величины.

2. По соответствующей формуле вычислить точечные (или интервальные) вероятности p_i.

3. Вычислить выравнивающие частоты $m_i' = n \cdot p_i$, где n – объём выборки.

4. Найти статистику $\chi^2_{набл} = \sum\limits_{i=1}^{k} \dfrac{(m_i - m_i')^2}{m_i'}$.

5. Определить число степеней свободы $l = k - r - 1$, где k – число частичных интервалов выборки; r – число параметров дифференциальной функции распределения. Выражения для нахождения числа степеней свободы известных законов распределения представлены в табл. 10.

Таблица 10

Закон распределения	Число степеней свободы
Биномиальный закон	$l = k - 1$, если p_A известно
	$l = k - 2$, если p_A неизвестно
Закон распределения Пуассона	$l = k - 2$
Равномерный закон	$l = k - 3$
Показательный закон	$l = k - 2$
Нормальный закон	$l = k - 3$

6. По таблице приложения 3 необходимо найти критическую величину

$\chi^2_{кр} = \chi^2_{\alpha; l}$, где α – заданный уровень значимости.

7. *Если* $\chi^2_{набл} < \chi^2_{кр}$, *то закон теоретического распределения не противоречит опытным данным, нет оснований отвергнуть гипотезу о выбранном законе распределения. В противном случае выдвинутая гипотеза отвергается.*

Пример 22. Для эмпирического распределения рабочих цеха по выработке по данным табл. 5 на уровне значимости $\alpha = 0{,}05$ выдвинуть гипотезу о распределении генеральной совокупности и проверить её с помощью критерия χ^2 Пирсона.

☺ *Решение.* По виду гистограммы распределения рабочих по выработке (рис. 25) можно предположить нормальный закон распределения признака.

Параметры a и σ^2 нормального закона распределения, являющиеся соответственно математическим ожиданием и дисперсией случайной величины X, неизвестны. Поэтому заменяем их «наилучшими» оценками по выборке – несмещёнными и состоятельными оценками соответственно выборочной средней $\overline{x}_в$ и исправленной выборочной дисперсией s^2. Т. к. число наблюдений $n = 100$ достаточно велико, то вместо s^2 можно взять σ^2. В примере 5 были вычислены $\overline{x} = 119{,}2$ (%), $\sigma = 9{,}35$ (%).

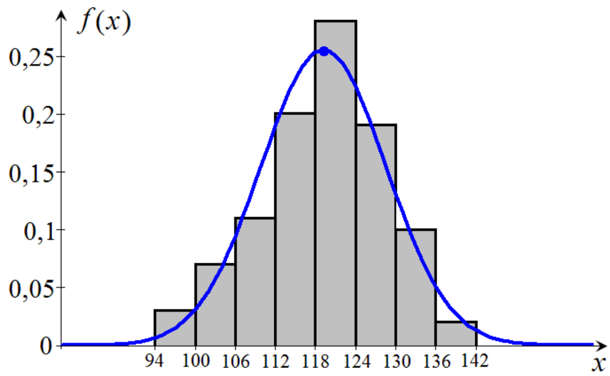


Рис. 25

Сформулируем основную гипотезу H_0: «Случайная величина X – выработка рабочих цеха – распределена нормально с параметрами $a = 119{,}2$, $\sigma = 9{,}35$, т. е. $X \sim N(119{,}2; 9{,}35)$». Альтернативная гипотеза H_1: «Случайная величина X не распределена по нормальному закону».

Число наблюдений в крайних интервалах (табл. 5) меньше 5, поэтому объединим их с соседними.

$(x_{i-1}, x_i]$	94-106	106-112	112-118	118-124	124-130	130-142	**Σ**
m_i	10	11	20	28	19	12	**100**

Для расчёта вероятностей p_i попадания случайной величины X в интервал $(x_{i-1}, x_i]$ используем функцию Лапласа в соответствии со свойствами нормального распределения (табл. 9):

$$p_i = P\{x_{i-1} \le X \le x_i\} = \Phi\left(\frac{x_i - a}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - a}{\sigma}\right).$$

Найдём значения p_i ($i = 1, 2, ..., 6$).

Т. к. случайная величина $X \sim N(a;\sigma)$ определена на интервале $(-\infty, +\infty)$, то крайние промежутки в ряде распределения заменяем, соответственно на $(-\infty, 106]$ и $(130, +\infty)$. Тогда, используя значения функции Лапласа (таблица приложения 1), получаем:

$$p_1 = P\{-\infty \le X \le 106\} = \Phi\left(\frac{106 - 119,2}{9,35}\right) - \Phi(-\infty) =$$

$$= \Phi(-1,41) - \Phi(-\infty) = -0,4207 + 0,5 = 0,0793.$$

$$p_2 = P\{106 \le X \le 112\} = \Phi\left(\frac{112 - 119,2}{9,35}\right) - \Phi\left(\frac{106 - 119,2}{9,35}\right) =$$

$$= \Phi(-0,77) - \Phi(-1,41) = -0,2794 + 0,4207 = 0,1413.$$

$$p_3 = P\{112 \le X \le 118\} = \Phi\left(\frac{118 - 119,2}{9,35}\right) - \Phi\left(\frac{112 - 119,2}{9,35}\right) =$$

$$= \Phi(-0,13) - \Phi(-0,77) = -0,0517 + 0,2794 = 0,2277.$$

$$p_4 = P\{118 \le X \le 124\} = \Phi\left(\frac{124 - 119,2}{9,35}\right) - \Phi\left(\frac{118 - 119,2}{9,35}\right) =$$

$$= \Phi(0,51) - \Phi(-0,13) = 0,1950 + 0,0517 = 0,2467.$$

$$p_5 = P\{124 \le X \le 130\} = \Phi\left(\frac{130 - 119,2}{9,35}\right) - \Phi\left(\frac{124 - 119,2}{9,35}\right) =$$

$$= \Phi(1,16) - \Phi(0,51) = 0,3770 - 0,1950 = 0,1820.$$

$$p_6 = P\{130 \le X \le +\infty\} = \Phi(+\infty) - \Phi\left(\frac{130 - 119,2}{9,35}\right) =$$

$$= \Phi(+\infty) - \Phi(1,16) = 0,5 - 0,3770 = 0,1230.$$

Для определения статистики χ^2 удобно составить таблицу (табл. 11)

Таблица 11

i	$(x_{i-1}, x_i]$	m_i	p_i	$m'_i = n \cdot p_i$	$(m_i - m'_i)^2$	$\frac{(m_i - m'_i)^2}{m'_i}$
1	$(-\infty, 106]$	10	0,079	7,9	4,41	0,558
2	$(106, 112]$	11	0,141	14,1	9,61	0,682
3	$(112, 118]$	20	0,228	22,8	7,84	0,344
4	$(118, 124]$	28	0,247	24,7	10,89	0,441
5	$(124, 130]$	19	0,182	18,2	0,64	0,035
6	$(130, +\infty)$	12	0,123	12,3	0,09	0,007
Σ		**100**	**1**	**100**	**–**	**2,067**

Итак, фактически наблюдаемое значение статистики $\chi^2_{набл} = 2{,}067$.

Число интервалов $k = 6$, тогда согласно табл. 10, для нормального закона распределения число степеней свободы $l = k - 3 = 3$. Соответствующее критическое значение статистики χ^2 по таблице приложения 3 $\chi^2_{кр} = \chi^2_{0,05;3} = 7{,}82$. Т. к. $\chi^2_{набл} < \chi^2_{кр}$, то гипотеза о выбранном теоретическом нормальном законе $N(119{,}2; 9{,}35)$ не противоречит опытным данным. Значит, нет оснований отвергнуть проверяемую гипотезу. ☻

Замечание. Для графического изображения эмпирического и выравнивающего его теоретического нормального распределений необходимо использовать одинаковый для двух распределений масштаб по оси ординат.

Пример 23. Для эмпирического распределения, заданного таблицей

Варианта, x_i	70	80	90	100	Σ
Частота, m_i	9	8	8	5	$n = 30$

на уровне значимости $\alpha = 0{,}05$ проверить гипотезу о нормальном распределении генеральной совокупности с помощью критерия χ^2 Пирсона.

☺ *Решение*. Сформулируем основную гипотезу H_0: «Случайная величина X – распределена нормально», тогда альтернативная гипотеза H_1: «Случайная величина X не распределена по нормальному закону».

Вычислим точечные оценки параметров a и σ:

$$a = \bar{x}_в = \frac{1}{30}\cdot(70\cdot 9 + 80\cdot 8 + 90\cdot 8 + 100\cdot 5) = 83,$$

$$\sigma_в^2 = \frac{1}{30}\cdot\left[(70-83)^2\cdot 9 + (80-83)^2\cdot 8 + (90-83)^2\cdot 8 + (100-83)^2\cdot 5\right] = 114,333.$$

Т. к. объём выборки невелик, то перейдём к исправленной дисперсии $s^2 = \frac{30}{29}\cdot\sigma_в^2$, тогда полагаем $\sigma = \sqrt{s^2} = 10,875$. Таким образом, имеем нормальный закон распределения $N(83;10,875)$.

Выравнивающие частоты m_i' найдём по формуле (табл. 9)

$$m_i' = \frac{nh}{\sigma}\cdot\varphi(t_i),$$

где $n = 30$, $h = 10$ – разность между двумя соседними вариантами, $t_i = \frac{x_i - a}{\sigma}$ и $\varphi(t)$ – функция Гаусса (см. табл. приложения 1).

Заполним вспомогательную таблицу (табл. 12).

Таблица 12

i	x_i	m_i	$t_i = \frac{x_i - a}{\sigma}$	$\varphi(t_i)$	$m_i' = \frac{nh}{\sigma}\cdot\varphi(t_i)$	$(m_i - m_i')^2$	$\frac{(m_i - m_i')^2}{m_i'}$
1	70	9	– 1,20	0,1942	5,35	13,32	2,45
2	80	8	– 0,28	0,3836	10,58	6,66	0,63
3	90	8	0,64	0,3251	8,97	0,94	0,10
4	100	5	1,56	0,1182	3,26	3,03	0,93
Σ		**30**	**–**	**–**	**28,47**	**–**	**4,11**

Итак, фактически наблюдаемое значение статистики $\chi^2_{набл} = 4,11$.

Т. к. число наблюдений равно 4, то согласно табл. 12, для нормального закона распределения число степеней свободы $k = 4 - 3 = 1$. По таблице приложения 3 имеем $\chi^2_{кр} = \chi^2_{0,05;1} = 3,8$. Так как $\chi^2_{набл} > \chi^2_{кр}$, то гипотеза о выбранном теоретическом нормальном законе $N(83;10,875)$ противоречит опытным данным, значит, принимаем альтернативную гипотезу: «Случайная величина X не распределена по нормальному закону». ☻

Глава 2. Элементы теории корреляции

2.1. Понятие о корреляционной зависимости. Корреляционная таблица

Часто приходится иметь дело с более сложной зависимостью, чем функциональная (когда каждому значению одной переменной соответствует вполне определённое единственное значение другой). Такова, например, связь между осадками и урожаем, связь между толщиной снегового покрова зимой и объёмом стока последующего половодья, связь производительности труда на предприятии от его энерговооружённости и т.п. Здесь каждому значению одной величины соответствует множество возможных значений другой величины. Подобного рода зависимости относятся к корреляционным зависимостям.

Статистической называют зависимость, при которой изменение одной из величин влечёт изменение распределения другой. Если при этом изменение одной величины приводит к изменению среднего значения другой, то статистическую зависимость называют ***корреляционной***.

Пусть, например, Y – урожай зерна, X – количество удобрений. С одинаковых по площади участков земли при равных количествах внесённых удобрений снимают различный урожай, т. е. Y не является функцией от X. Это объясняется влиянием случайных факторов (осадки, температура воздуха и пр.). Вместе с тем, как показывает опыт, средний урожай является функцией от количества удобрений, т. е. Y связан с X корреляционной зависимостью.

Таким образом, из представленных зависимостей наиболее общей выступает статистическая зависимость. Каждая корреляционная зависимость является статистической, но не каждая статистическая зависимость является корреляционной. Функциональная зависимость представляет собой частный случай корреляционной.

При большом числе наблюдений одно и то же значение x может встретиться m_x раз, одно и то же значение y – m_y раз, одна и та же пара чисел (x, y) может наблюдаться m_{xy} раз. Поэтому данные наблюдений группируют, т. е. подсчитывают частоты m_x, m_y, m_{xy}. Все сгруппированные данные записывают в виде таблицы 13, которую называют корреляционной.

Таблица 13

X	Y				m_{x_i}
	y_1	y_2	$\ldots$	y_ℓ	
x_1	$m_{x_1 y_1}$	$m_{x_1 y_2}$	$\ldots$	$m_{x_1 y_\ell}$	$m_{x_1} = \sum_{j=1}^{\ell} m_{x_1 y_j}$
x_2	$m_{x_2 y_1}$	$m_{x_2 y_2}$	$\ldots$	$m_{x_2 y_\ell}$	$m_{x_2} = \sum_{j=1}^{\ell} m_{x_2 y_j}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_k	$m_{x_k y_1}$	$m_{x_k y_2}$	$\ldots$	$m_{x_k y_\ell}$	$m_{x_k} = \sum_{j=1}^{\ell} m_{x_k y_j}$
m_{y_j}	$m_{y1} = \sum_{i=1}^{k} m_{x_i y_1}$	$m_{y2} = \sum_{i=1}^{k} m_{x_i y_2}$	$\ldots$	$m_{y\ell} = \sum_{i=1}^{k} m_{x_i y_\ell}$	$\sum_{i=1}^{k} m_{x_i} = \sum_{j=1}^{\ell} m_{y_j} = n$

где m_{x_i} – частота появления варианты x_i,

m_{y_j} – частота появления варианты y_j,

$m_{x_i y}$ – частота появления варианты x_i при заданном значении y,

$m_{x y_j}$ – частота появления варианты y_j при заданном значении x,

n – объём выборки.

Определим некоторые первичные понятия корреляционной зависимости в табл. 14.

Таблица 14

Случайная величина *X*	**Случайная величина *Y***
Общее среднее	
– это среднее арифметическое всех значений случайной величины *X*: $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot m_{x_i}$	– это среднее арифметическое всех значений случайной величины *Y*: $\overline{y} = \frac{1}{n}\sum_{j=1}^{\ell} y_j \cdot m_{y_j}$
Условное среднее	
− это среднее арифметическое тех значений случайной величины *X*, которые соответствуют значению случайной величины $Y = y$: $\overline{x}_y = \frac{\sum_{i=1}^{k} x_i \cdot m_{x_i y}}{\sum_{i=1}^{k} m_{x_i y}}$	− это среднее арифметическое тех значений случайной величины *Y*, которые соответствуют значению случайной величины $X = x$: $\overline{y}_x = \frac{\sum_{j=1}^{\ell} y_j \cdot m_{x y_j}}{\sum_{j=1}^{\ell} m_{x y_j}}$
Межгрупповое среднее	
$\overline{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{\ell} x_i \cdot y_j \cdot m_{x_i y_j}$	
Общее среднее квадратическое отклонение	
$\sigma_X^2 = \overline{x^2} - (\overline{x})^2$	$\sigma_Y^2 = \overline{y^2} - (\overline{y})^2$
Межгрупповая дисперсия	
$\delta_X^2 = \frac{1}{n}\sum_{j=1}^{\ell} (\overline{x}_{y_j} - \overline{x})^2 \cdot m_{y_j}$	$\delta_Y^2 = \frac{1}{n}\sum_{i=1}^{k} (\overline{y}_{x_i} - \overline{y})^2 \cdot m_{x_i}$

Пример 24. Найти числовые характеристики случайных величин X и Y по элементам данной таблицы:

$x \backslash y$	5	6	m_{x_i}
2	2	1	**3**
3	3	4	**7**
m_{y_j}	**5**	**5**	**10**

☺ *Решение*. Найдём числовые характеристики по соответствующим формулам табл. 14.

$$\overline{x} = \frac{\sum_{i=1}^{2} x_i \cdot m_{x_i}}{n} = \frac{2\cdot 3 + 3\cdot 7}{10} = 2{,}7.$$

$$\overline{y} = \frac{\sum_{j=1}^{2} y_j \cdot m_{y_j}}{n} = \frac{5\cdot 5 + 6\cdot 5}{10} = 5{,}5.$$

$$\overline{x^2} = \frac{\sum_{i=1}^{2} x_i^2 \cdot m_{x_i}}{n} = \frac{2^2\cdot 3 + 3^2\cdot 7}{10} = 7{,}5.$$

$$\overline{y^2} = \frac{\sum_{j=1}^{2} y_j^2 \cdot m_{y_j}}{n} = \frac{5^2\cdot 5 + 6^2\cdot 5}{10} = 30{,}5.$$

$$\sigma_X^2 = \overline{x^2} - (\overline{x})^2 = 7{,}5 - 2{,}7^2 = 0{,}21.$$

$$\sigma_Y^2 = \overline{y^2} - (\overline{y})^2 = 30{,}5 - 5{,}5^2 = 0{,}25.$$

$$\overline{x}_{y=5} = \frac{2\cdot 2 + 3\cdot 3}{5} = \frac{13}{5} = 2{,}6.$$

$$\overline{y}_{x=2} = \frac{5\cdot 2 + 6\cdot 1}{3} = \frac{16}{3}.$$

$$\overline{x}_{y=6} = \frac{2\cdot 1 + 3\cdot 4}{5} = \frac{14}{5} = 2{,}8.$$

$$\overline{y}_{x=3} = \frac{5\cdot 3 + 6\cdot 4}{7} = \frac{39}{7}.$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{2} x_i \cdot y_j \cdot m_{x_i y_j} = \frac{(5\cdot 2)\cdot 2 + (5\cdot 3)\cdot 3 + (6\cdot 2)\cdot 1 + (6\cdot 3)\cdot 4}{10} = 14{,}9.$$

$$\delta_X^2 = \frac{1}{n}\sum_{j=1}^{2}(\overline{x}_{y_j} - \overline{x})^2 \cdot m_{y_j} = \frac{5\cdot\left(2{,}6 - 2{,}7\right)^2 + 5\cdot\left(2{,}8 - 2{,}7\right)^2}{10} = 0{,}01.$$

$$\delta_Y^2 = \frac{1}{n}\sum_{i=1}^{2}(\overline{y}_{x_i} - \overline{y})^2 \cdot m_{x_i} = \frac{3\cdot\left(\frac{16}{3} - 5{,}5\right)^2 + 7\cdot\left(\frac{39}{7} - 5{,}5\right)^2}{10} = 0{,}054.$$ ☻

Вычисленные групповые средние $\bar{y}_x$ (и $\bar{x}_y$) можно изобразить графически в виде ломаной, называемой *эмпирической линией регрессии* *Y* по *X* (*X* по *Y*).

Уравнение, определяющее корреляционную зависимость $\bar{y}_x = f(x)$, называют ***уравнением регрессии*** случайной величины Y на случайную величину X .

Уравнение, определяющее корреляционную зависимость $\bar{x}_y = g(y)$, называют ***уравнением регрессии*** случайной величины X на случайную величину Y .

Если обе функции регрессии $f(x)$ и $g(y)$ линейны, то корреляционную зависимость называют ***линейной***.

2.2. Теснота корреляционной связи

На практике часто бывает важно знать, существует ли зависимость между наблюдаемыми величинами, и если да, то какая она – сильная или слабая, положительная или отрицательная?

Корреляционным отношением случайной величины *Y* на случайную величину *X* (или *X* на *Y*) называется отношение межгруппового среднего квадратического отклонения случайной величины *Y* (или *X*) к общему среднему квадратическому отклонению этой же случайной величины:

$$\boxed{\eta_{YX} = \frac{\delta_Y}{\sigma_Y}} \quad \text{или} \quad \boxed{\eta_{XY} = \frac{\delta_X}{\sigma_X}}. \qquad (36)$$

Корреляционное отношение $\eta \in [0;1]$ служит для **оценки тесноты** (наличия) корреляционной связи между случайными величинами *X* и *Y*.
В зависимости от величины η возможны следующие ситуации (табл. 15):

Таблица 15

Значения η	Зависимость случайных величин X и Y
$\eta = 0$	Случайные величины X и Y не связаны корреляционной зависимостью.
$0 < \eta < 0{,}3$	Корреляционная зависимость практически отсутствует.
$0{,}3 < \eta < 0{,}5$	Слабая корреляционная зависимость.
$0{,}5 < \eta < 0{,}7$	Умеренная корреляционная зависимость.
$0{,}7 < \eta < 1$	Сильная корреляционная зависимость.
$\eta = 1$	Случайные величины X и Y связаны функциональной зависимостью.

2.3. Линейная регрессия

2.3.1. Проверка гипотезы о наличии линейной зависимости

Пусть некоторая двумерная генеральная совокупность распределена нормально и из неё извлечена выборка объёма n, для которой найден ***выборочный коэффициент корреляции***

$$\boxed{r_в = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_X \cdot \sigma_Y}}. \quad (37)$$

Тогда проверяют **нулевую гипотезу** H_0: «$r_г = 0$ − линейная зависимость в генеральной совокупности отсутствует» при выдвигаемой **альтернативной гипотезе** H_1: «$r_г \neq 0$ − линейная зависимость присутствует».

Алгоритм:

1. Вычислить статистику $T_{набл} = \dfrac{r_в \cdot \sqrt{n-2}}{\sqrt{1-r_в^2}}$. (38)

2. Определить критическое значение распределения Стьюдента (табл. приложения 4) $t_{кр} = t(\alpha; k = n-2)$. Критической областью при этом является двусторонняя область $D = (-\infty; -t_{кр}) \bigcup (t_{кр}; +\infty)$.

3. Сделать вывод:

$T_{набл} \in D \Rightarrow$ гипотезу H_0 отвергают, т. е. имеется линейная зависимость;

$T_{набл} \notin D \Rightarrow$ гипотезу H_0 принимают, т. е. линейной зависимости нет.

2.3.2. Поиск уравнения линейной зависимости

Пусть в результате n испытаний получены выборочные пары чисел $(x_1;y_1)$, $(x_2;y_2)$, …, $(x_n;y_n)$, которые располагаются на плоскости вдоль некоторой прямой линии $y_x = k\cdot x + b$.

Угловой коэффициент k уравнения $y_x = k\cdot x + b$ называется ***выборочным коэффициентом регрессии*** *Y* на *X* и обозначают $k = \rho_{YX}$.

Неизвестные параметры k и b находят по методу наименьших квадратов.

Составим функцию $F(k,b) = \sum_{i=1}^{n}\left(y_{x_i} - y_i\right)^2 = \sum_{i=1}^{n}(k\cdot x_i + b - y_i)^2$ и решим систему относительно равенства нулю её частных производных:

$$\begin{cases} \dfrac{\partial F}{\partial k} = 0, \\ \dfrac{\partial F}{\partial b} = 0. \end{cases} \Rightarrow \begin{cases} 2\cdot\sum\limits_{i=1}^{n}(k\cdot x_i + b - y_i)\cdot x_i = 0, \\ 2\cdot\sum\limits_{i=1}^{n}(k\cdot x_i + b - y_i)\cdot 1 = 0. \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} k\cdot\sum\limits_{i=1}^{n}x_i^2 + b\cdot\sum\limits_{i=1}^{n}x_i = \sum\limits_{i=1}^{n}x_i y_i, \\ k\cdot\sum\limits_{i=1}^{n}x_i + b\cdot n = \sum\limits_{i=1}^{n}y_i. \end{cases} \Rightarrow \begin{cases} k\cdot\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i^2 + b\cdot\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i = \dfrac{1}{n}\sum\limits_{i=1}^{n}x_i y_i, \\ k\cdot\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i + b = \dfrac{1}{n}\sum\limits_{i=1}^{n}y_i. \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} k\cdot\overline{x^2} + b\cdot\overline{x} = \overline{xy}, \\ k\cdot\overline{x} + b = \overline{y}. \end{cases} \Rightarrow \begin{cases} k\cdot\overline{x^2} + (\overline{y} - k\cdot\overline{x})\cdot\overline{x} = \overline{xy}, \\ b = \overline{y} - k\cdot\overline{x}. \end{cases} \Rightarrow$$

$$\Rightarrow \quad \rho_{YX} = k = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - (\overline{x})^2} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\sigma_X^2}.$$

Таким образом, уравнение прямой линии регрессии примет вид

$$\boxed{y_x - \overline{y} = \rho_{YX}\cdot(x - \overline{x})}.$$

Аналогично можно получить уравнение прямой линии регрессии *X* на *Y*, которое имеет вид

$$\boxed{x_y - \overline{x} = \rho_{XY}\cdot(y - \overline{y})}.$$

Преобразуем выборочные коэффициенты регрессии:

$$\rho_{YX}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_X^2}\ \Bigg|\ \cdot\frac{\sigma_X}{\sigma_Y}\quad\Rightarrow\quad\frac{\rho_{YX}\cdot\sigma_X}{\sigma_Y}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_Y\cdot\sigma_X}=r_в;$$

$$\rho_{XY}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_Y^2}\ \Bigg|\ \cdot\frac{\sigma_Y}{\sigma_X}\quad\Rightarrow\quad\frac{\rho_{XY}\cdot\sigma_Y}{\sigma_X}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_Y\cdot\sigma_X}=r_в\ .$$

Получим систему: $\begin{cases}\dfrac{\rho_{YX}\cdot\sigma_X}{\sigma_Y}=r_в\,,\\[2mm] \dfrac{\rho_{XY}\cdot\sigma_Y}{\sigma_X}=r_в\,.\end{cases}\quad\Rightarrow\quad\begin{cases}\rho_{YX}=\dfrac{r_в\cdot\sigma_Y}{\sigma_X},\\[2mm] \rho_{XY}=\dfrac{r_в\cdot\sigma_X}{\sigma_Y}.\end{cases}$

В итоге уравнения прямых линий регрессии принимают вид:

$$\boxed{y_x-\overline{y}=\frac{r_в\cdot\sigma_Y}{\sigma_X}\cdot(x-\overline{x})}\ \text{или}\ \boxed{x_y-\overline{x}=\frac{r_в\cdot\sigma_X}{\sigma_Y}\cdot(y-\overline{y})},\qquad(39)$$

где $r_в=\dfrac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_X\cdot\sigma_Y}$.

Взаимное расположение прямых линий регрессии на плоскости изображено на рис. 26.

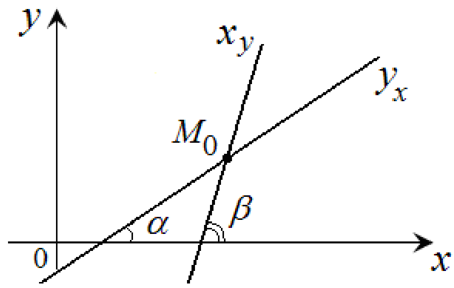


Рис. 26

Здесь $M_0(\overline{x};\overline{y})$ – точка пересечения двух линий регрессии.

Для регрессии *Y* на *X*: $k_1=\operatorname{tg}\alpha=\dfrac{r_в\cdot\sigma_Y}{\sigma_X}$.

Для регрессии *X* на *Y*: $k_2=\operatorname{tg}\beta=\dfrac{\sigma_Y}{r_в\cdot\sigma_X}$.

2.3.3. Теснота линейной корреляционной связи

Выборочный коэффициент корреляции $r_в$ является характеристикой линейности уравнения регрессии, т. к. определяет тесноту зависимости между случайными величинами X и Y.

При этом $\begin{cases} \dfrac{\rho_{YX} \cdot \sigma_X}{\sigma_Y} = r_в, \\ \dfrac{\rho_{XY} \cdot \sigma_Y}{\sigma_X} = r_в, \end{cases} \Rightarrow \quad r_в^2 = \rho_{XY} \cdot \rho_{YX}.$

Выборочным коэффициентом корреляции переменных X и Y, между которыми предполагается линейная корреляционная связь, называется среднее геометрическое их коэффициентов регрессии и имеющее знак последних:

$$r_в = \pm\sqrt{\rho_{XY} \cdot \rho_{YX}}, \quad |r_в| \le \eta \le 1.$$

В зависимости от принимаемых значений величины $r_в$ возможны следующие случаи (табл. 16):

Таблица 16

Значения $r_в$	**Зависимость случайных величин** X **и** Y
$\lvert r_в \rvert = \eta$	Точная линейная корреляционная зависимость.
$0 < r_в < 1$	Прямая связь линейной зависимости (возрастающая функция).
$-1 < r_в < 0$	Обратная связь линейной зависимости (убывающая функция).
$r_в = 0$	Отсутствует линейная зависимость (графики уравнений регрессий взаимно перпендикулярны).
$0 < \lvert r_в \rvert < 0{,}3$	Линейная зависимость практически отсутствует.
$0{,}3 < \lvert r_в \rvert < 0{,}5$	Слабая линейная зависимость.
$0{,}5 < \lvert r_в \rvert < 0{,}7$	Умеренная линейная зависимость.
$0{,}7 \le \lvert r_в \rvert < 1$	Сильная линейная зависимость.
$\lvert r_в \rvert = 1$	Функциональная линейная зависимость (графики уравнений регрессий совпадают).

2.3.4. Показатели качества (адекватности) регрессии

Качество модели регрессии связывается с адекватностью модели наблюдаемым данным и проводится на основе анализа остатков (разностей) $\varepsilon_{y_i} = y_i - y_{x_i}$ (или $\varepsilon_{x_j} = x_j - x_{y_j}$), которые представляют собой отклонения фактических значений зависимой переменной от значений данной переменной, вычисленных по уравнению регрессии.

Если $\varepsilon_i = 0$, $i = \overline{1,m}$ ($\varepsilon_j = 0$, $j = \overline{1,n}$), то для всех наблюдений фактические значения зависимой переменной совпадают с расчётными значениями. Графически это означает, что теоретическая линия регрессии проходит через все точки корреляционного поля. Следовательно, результативный признак полностью обусловлен влиянием фактора.

На практике, как правило, происходит некоторое рассеивание точек корреляционного поля относительно теоретической линии регрессии. Величина этих отклонений является основой расчёта показателей качества (адекватности) уравнения регрессии.

Рассчитываются следующие показатели качества линейной регрессии:

1. Теоретический коэффициент детерминации

$$\boxed{R^2 = r_в^2}. \tag{40}$$

2. Средняя квадратическая ошибка S_ε уравнения регрессии представляет собой среднее квадратическое отклонение наблюдаемых значений результативного признака от теоретических значений, рассчитанных по модели. Величина S_ε вычисляется по формуле

$$\boxed{S_{\varepsilon_y} = \sqrt{\frac{1}{m-2}\sum_{i=1}^{m}\varepsilon_{y_i}^2}} \text{ или } \boxed{S_{\varepsilon_x} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\varepsilon_{x_j}^2}}. \tag{41}$$

3. Средняя ошибка аппроксимации (приближения) A равна:

$$\boxed{A_y = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\varepsilon_{y_i}}{y_i}\right|} \text{ или } \boxed{A_x = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\varepsilon_{x_i}}{x_i}\right|}. \tag{42}$$

Чем меньше рассеяние эмпирических точек от теоретических, тем меньше средняя ошибка аппроксимации. Ошибка аппроксимации меньше 7% свидетельствует о хорошем качестве модели.

2.4. Нелинейные корреляционные связи

2.4.1. Поиск уравнения параболической зависимости

Пусть в результате n испытаний получены выборочные пары чисел $(x_1;y_1)$, $(x_2;y_2)$, …, $(x_n;y_n)$, которые располагаются на плоскости вдоль некоторой кривой линии регрессии вида $y_x = a\,x^2 + b\,x + c$.

Неизвестные параметры a, b и c находятся по методу наименьших квадратов. Необходимо составить функцию:

$$F(a,b,c) = \sum_{i=1}^{n}(y_{x_i} - y_i)^2 = \sum_{i=1}^{n}(a\,x_i^2 + b\,x_i + c - y_i)^2$$

и решить систему относительно равенства нулю её частных производных, откуда определяются неизвестные a, b и c

$$\begin{cases} a\cdot\sum\limits_{i=1}^{n}x_i^4 + b\cdot\sum\limits_{i=1}^{n}x_i^3 + c\cdot\sum\limits_{i=1}^{n}x_i^2 = \sum\limits_{i=1}^{n}x_i^2\,y_i, \\ a\cdot\sum\limits_{i=1}^{n}x_i^3 + b\cdot\sum\limits_{i=1}^{n}x_i^2 + c\cdot\sum\limits_{i=1}^{n}x_i = \sum\limits_{i=1}^{n}x_i\,y_i, \\ a\cdot\sum\limits_{i=1}^{n}x_i^2 + b\cdot\sum\limits_{i=1}^{n}x_i + c\cdot n = \sum\limits_{i=1}^{n}y_i. \end{cases} \tag{43}$$

2.4.2. Поиск уравнения гиперболической зависимости

В результате n испытаний получены выборочные пары чисел $(x_1;y_1)$, $(x_2;y_2)$, …, $(x_n;y_n)$, которые располагаются на плоскости вдоль некоторой кривой линии регрессии вида $y_x = \dfrac{k}{x} + b$.

Неизвестные параметры k и b найдём по методу наименьших квадратов. Необходимо составить функцию:

$$F(k,b) = \sum_{i=1}^{n}(y_{x_i} - y_i)^2 = \sum_{i=1}^{n}\left(\frac{k}{x_i} + b - y_i\right)^2,$$

и решить систему относительно равенства нулю её частных производных, откуда определяются неизвестные k и b

$$\begin{cases} k \cdot \sum\limits_{i=1}^{n} \dfrac{1}{x_i^2} + b \cdot \sum\limits_{i=1}^{n} \dfrac{1}{x_i} = \sum\limits_{i=1}^{n} \dfrac{y_i}{x_i}, \\ k \cdot \sum\limits_{i=1}^{n} \dfrac{1}{x_i} \; + b \cdot n \qquad = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

Замечание. Аналогично можно рассуждать и для других видов корреляционных связей.

Пример 25. Дана двумерная выборка:

$x \backslash y$	**48**	**67**	**86**	m_{x_i}
0	1	–	–	**1**
1	2	29	–	**31**
2	–	2	30	**32**
3	–	29	6	**35**
4	1	–	–	**1**
m_{y_j}	**4**	**60**	**36**	**100**

По данным таблицы:

1. вычислить корреляционное отношение η_{YX};

2. вычислить коэффициент корреляции $r_в$;

3. выдвинуть гипотезу о наличии или отсутствии линейной зависимости на уровне значимости $\alpha = 0{,}01$;

4. найти соответствующее уравнение регрессии;

5. вычислить теоретические значения условной средней y_x (по уравнению регрессии в точках x_i). Результаты представить в виде таблицы:

x_i	x_1	x_2	$\cdots$	x_k
$\bar{y}_{x=x_i}$	$\bar{y}_1$	$\bar{y}_2$	$\cdots$	$\bar{y}_k$
y_{x_i}	y_1	y_2	$\cdots$	y_k

6. на одном чертеже построить линию регрессии и вычисленные средние значения $\bar{y}_x$.

☺ *Решение*. Для проведения расчётов заполним таблицы:

x	m_x	xm_x	x^2m_x
0	1	0	0
1	31	31	31
2	32	64	128
3	35	105	315
4	1	4	16
Σ	**100**	**204**	**490**

$\overline{x} = \dfrac{204}{100} = 2{,}04$.

$\overline{x^2} = \dfrac{490}{100} = 4{,}9$.

$D_x = \overline{x^2} - \overline{x}^2 = 4{,}9 - 2{,}04^2 = 0{,}7384$.

$\sigma_x = \sqrt{D_x} = \sqrt{0{,}7384} = 0{,}859$.

y	m_y	ym_y	y^2m_y
48	4	192	9216
67	60	4020	269340
86	36	3096	266256
Σ	**100**	**7308**	**544812**

$\overline{y} = \dfrac{7308}{100} = 73{,}08$.

$\overline{y^2} = \dfrac{544812}{100} = 5448{,}12$.

$D_y = \overline{y^2} - \overline{y}^2 = 5448{,}12 - 73{,}08^2 = 107{,}4336$.

$\sigma_y = \sqrt{D_y} = \sqrt{107{,}4336} = 10{,}365$.

$(x;y)$	(0;48)	(1;48)	(1;67)	(2;67)	(2;86)	(3;67)	(3;86)	(4;48)	**Σ**
m_{xy}	1	2	29	2	30	29	6	1	**100**
xym_{xy}	0	96	1943	268	5160	5829	1548	192	**15036**

Следовательно, $\overline{xy} = \dfrac{15036}{100} = 150{,}36$.

Найдём условные средние значения $\overline{y}_x$ и межгрупповую дисперсию:

$\overline{y}_{x=0} = \dfrac{48\cdot 1}{1} = 48$.

$\overline{y}_{x=1} = \dfrac{48\cdot 2 + 67\cdot 29}{31} = 65{,}8$.

$\overline{y}_{x=2} = \dfrac{67\cdot 2 + 86\cdot 30}{32} = 84{,}8$.

$\overline{y}_{x=3} = \dfrac{67\cdot 29 + 86\cdot 6}{35} = 70{,}3$.

$\overline{y}_{x=4} = \dfrac{48\cdot 1}{1} = 48$.

$\Rightarrow$

$$\delta_y^2 = \frac{1}{100}\Big((48-73{,}08)^2\cdot 1 + (65{,}8-73{,}08)^2\cdot 31 + (84{,}8-73{,}08)^2\cdot 32 + (70{,}3-73{,}08)^2\cdot 35 + (48-73{,}08)^2\cdot 1\Big) = 75{,}669.$$

$\delta_y = \sqrt{75{,}669} = 8{,}699$.

1. Вычислим корреляционное отношение по формуле (36):

$$\eta_{YX} = \frac{\delta_y}{\sigma_y} = \frac{8,699}{10,365} = 0,839,$$

По табл. 15 определяем, что *имеется сильная корреляционная зависимость.*

2. Найдём коэффициент корреляции по формуле (37)

$$r_в = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y} = \frac{150,36 - 2,04 \cdot 73,08}{0,859 \cdot 10,365} = 0,143,$$

По табл. 16 определяем, что *линейная зависимость практически отсутствует.*

3. Проверим гипотезу $H_0 : r_г = 0$ (линейная зависимость отсутствует) при альтернативе $H_1 : r_г \neq 0$ (имеется линейная зависимость).

По формуле (38) вычислим значение статистики:

$$T_{набл} = \frac{r_в \cdot \sqrt{n-2}}{\sqrt{1 - r_в^2}} = \frac{0,143 \cdot \sqrt{100-2}}{\sqrt{1 - 0,143^2}} = 1,43.$$

$$\alpha = 0,01,\ l = 100 - 2 = 98 \Rightarrow t_{кр} = t_{кр}(\alpha; l) = 2,62.$$

Т. к. $|T_{набл}| < t_{кр}$, то гипотезу H_1 отвергаем и принимаем гипотезу H_0 о наличии нелинейной корреляции в генеральной совокупности.

4. По виду расположения условных средних значений $\overline{y}_x$ на плоскости (рис. 27) предполагаем квадратическую зависимость. Составим уравнение нелинейной параболической регрессии: $y_x = ax^2 + bx + c$.

Заполним вспомогательную таблицу для вычисления коэффициентов:

x	$\overline{y}_x$	m_x	xm_x	x^2m_x	x^3m_x	x^4m_x	$\overline{y}_x m_x$	$x\overline{y}_x m_x$	$x^2\overline{y}_x m_x$
0	48	1	0	0	0	0	48	0	0
1	65,8	31	31	31	31	31	2039,8	2039,8	2039,8
2	84,8	32	64	128	256	512	2713,6	5427,2	10854,4
3	70,3	35	105	315	945	2835	2460,5	7381,5	22144,5
4	48	1	4	16	64	256	48	192	768
Σ	**–**	**100**	**204**	**490**	**1296**	**3634**	**7309,9**	**15040,5**	**35806,7**

Система (43) принимает вид

$$\begin{cases} a\cdot\sum\limits_{i=1}^{n}x_i^{4} + b\cdot\sum\limits_{i=1}^{n}x_i^{3} + c\cdot\sum\limits_{i=1}^{n}x_i^{2} = \sum\limits_{i=1}^{n}x_i^{2}\,y_i, \\ a\cdot\sum\limits_{i=1}^{n}x_i^{3} + b\cdot\sum\limits_{i=1}^{n}x_i^{2} + c\cdot\sum\limits_{i=1}^{n}x_i = \sum\limits_{i=1}^{n}x_i\,y_i, \\ a\cdot\sum\limits_{i=1}^{n}x_i^{2} + b\cdot\sum\limits_{i=1}^{n}x_i + c\cdot n = \sum\limits_{i=1}^{n}y_i. \end{cases}$$

Подставляя значения из итоговой строки предыдущей таблицы в систему, получим:

$$\begin{cases} 490a + 204b + 100c = 7309{,}9, \\ 1296a + 490b + 204c = 15040{,}5, \\ 3634a + 1296b + 490c = 35806{,}7. \end{cases}$$

Решая систему, получим: $a = -12{,}2$, $b = 50{,}7$, $c = 29{,}5$.

Таким образом, искомое уравнение регрессии имеет вид

$$y_x = -12{,}2x^2 + 50{,}7x + 29{,}5.$$

5. Заполним таблицу:

X	0	1	2	3
$\bar{y}_x$	48	65,8	84,8	70,3
y_x	29,5	68	82,1	71,8

6. Построим на одном чертеже график параболической регрессии

$$y_x = -12{,}2x^2 + 50{,}7x + 29{,}5$$

и нанесём экспериментальные данные $\bar{y}_x$ (рис. 27).

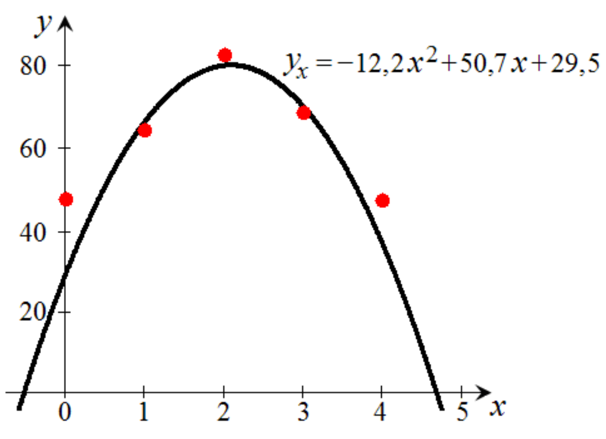


Рис. 27

☻

Разбор типовых задач

***Задание* № 1.** В результате опыта получена выборочная совокупность:

88	104	91	97	77	103	86	79	86	100
82	68	71	87	89	89	81	81	70	79
84	91	87	83	90	69	83	96	79	94
93	86	81	83	84	92	93	85	84	88
77	85	93	85	87	100	76	79	90	91
84	74	76	75	93	103	80	96	72	95
81	102	75	80	90	85	82	77	94	102
87	95	99	83	80	93	90	79	93	105
95	85	84	90	93	95	98	88	79	91
86	88	93	80	88	88	90	68	89	90

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0{,}05$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже: *а*) полигон относительных частот p_i^* и кривую распределения p_i; *б*) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0{,}99$.

☺ *Решение.*

1. Разобьем всю вариацию объёмом $n = 100$ на $k = 10$ частичных интервалов равной длины и посчитаем частоты попадания наблюдаемых значений в частичные интервалы.

Длину интервала находим по формуле $h = \dfrac{x_{\max} - x_{\min}}{k} = \dfrac{105 - 68}{10} \approx 4$.

За начало первого интервала примем $x_0 = x_{\min} - \dfrac{h}{2} = 68 - \dfrac{4}{2} = 66$. Получим последовательность интервалов: [66; 70], (70; 74], … , (102; 106].

Составим вариационный ряд частот и относительных частот:

i	интервал $(x_{i-1}; x_i]$	середина интервала $\tilde{x}_i$	частота m_i	относительная частота $p_i^* = \dfrac{m_i}{n}$
1	[66; 70]	68	4	0,04
2	(70; 74]	72	3	0,03
3	(74; 78]	76	7	0,07
4	(78; 82]	80	16	0,16
5	(82; 86]	84	18	0,18
6	(86; 90]	88	20	0,20
7	(90; 94]	92	15	0,15
8	(94; 98]	96	7	0,07
9	(98; 102]	100	6	0,06
10	(102; 106]	104	4	0,04
Σ	–	–	**100**	**1**

Отметим, что $\sum_{i=1}^{10} m_i = n = 100$ – объём выборки; $\sum_{i=1}^{10} p_i^* = \sum_{i=1}^{10} \dfrac{m_i}{n} = \dfrac{n}{n} = 1$.

Статистическое распределение выборки является оценкой неизвестного распределения. В частности, *относительные частоты p_i^* являются статистическими аналогами вероятностей полной группы несовместных событий.*

2. Вторым этапом обработки статистических данных является построение полигона, гистограммы относительных частот и эмпирической функции распределения.

а) Полигон относительных частот вариационного ряда – ломаная линия, соединяющая точки $(\tilde{x}_i; p_i^*)$ (рис. 28).

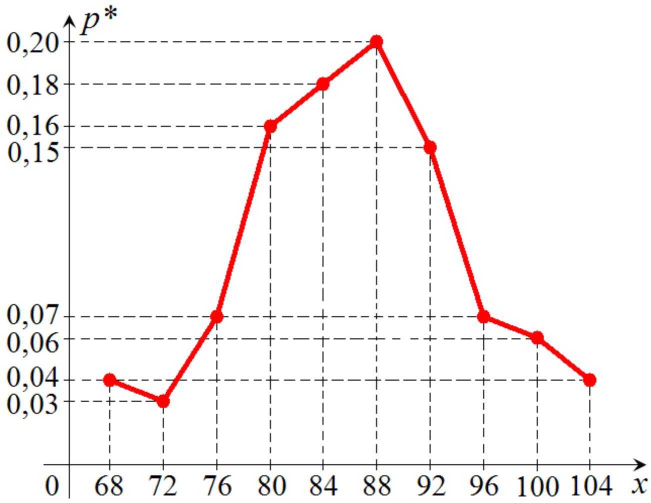


Рис. 28

Полигон относительных частот является статистическим аналогом многоугольника распределения дискретной случайной величины X.

б) Гистограмма относительных частот изображается только для интервального ряда и имеет вид ступенчатой фигуры. На каждом частичном интервале строим прямоугольник высотой $\dfrac{p_i^*}{h}$ (рис. 29).

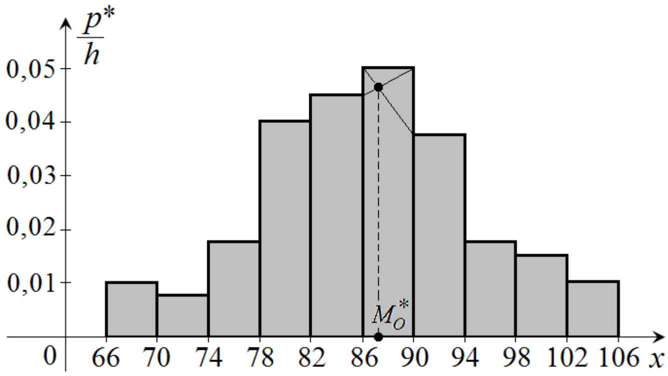


Рис. 29

Гистограмма относительных частот является статистическим аналогом дифференциальной функции распределения (плотности) $f(x)$ *непрерывной случайной величины X.*

в) График эмпирической функции распределения $F^*(x) = \sum\limits_{x_i<x} \dfrac{m_i}{n}$ непрерывной случайной величины X совпадает с кумулятой (графиком накопленных частот).

Отметим на плоскости точки, соответствующие значениям функции $F^*(x)$ на концах интервалов, и соединим их отрезками прямых (рис. 30).

x	$x \le 66$	70	74	78	82	86	90	94	98	102	$x \ge 106$
$F^*(x)$	0	0,04	0,07	0,14	0,30	0,48	0,68	0,83	0,90	0,96	1

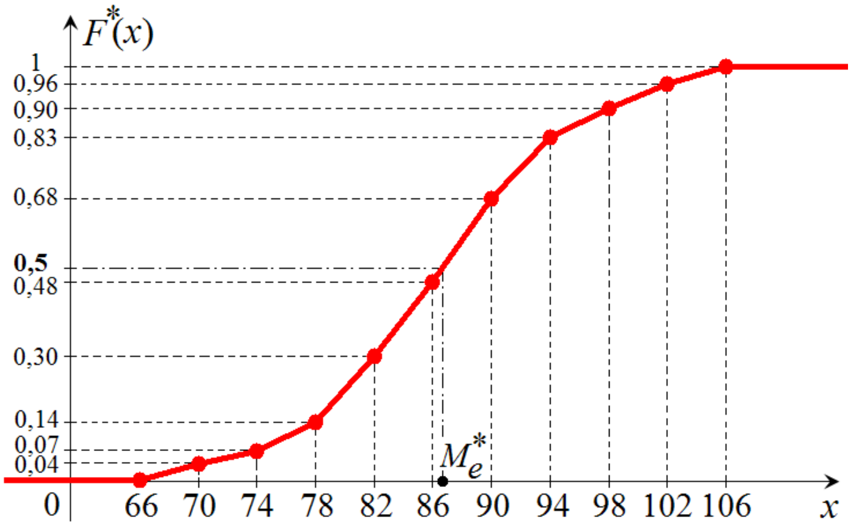


Рис. 30

Эмпирическая функция распределения $F^*(x)$ *является статистическим аналогом интегральной функции распределения* $F(x)$ *случайной величины* X.

3. Найдем числовые характеристики выборки.

Выборочные характеристики – это функции наблюдений, приближённо оценивающие соответствующие числовые характеристики случайной величины.

1) Мода M_o^* находится внутри интервала, для которого соответствующая частота максимальна. В нашем случае $M_o^* \in (86;\, 90]$, при этом $p^*_{\max} = p^*_6 = 0,20$ $(i = 6)$.

Моду приближённо можно определить на чертеже гистограммы (см. рис. 29) или вычислить по формуле (10), тогда получим

$$M_o^* = 86 + 4 \cdot \frac{0,20 - 0,18}{(0,20 - 0,18) + (0,20 - 0,15)} = 87,14.$$

2) Медиана M_e^* интервального вариационного ряда принадлежит тому частотному интервалу, для которого накопленная частота составляет половину или больше половины всей суммы частот, а предыдущая накопленная частота меньше половины всей суммы частот. Геометрически прямая $x = M_e^*$ делит площадь гистограммы пополам.

Медиана может быть приближённо найдена на чертеже графика $F^*(x)$ (рис. 30), как значение признака, для которого $F^*(M_e^*) = 0,5$. Для данного вариационного ряда $M_e^* \in (86;\ 90]$.

Значение M_e^* вычисляем по формуле (11):

$$M_e^* = 86 + 4 \cdot \frac{0,5 - (0,04 + 0,03 + 0,07 + 0,16 + 0,18)}{0,20} = 86,4\,.$$

3) Для нахождения выборочной средней $\overline{x}_в$, выборочной дисперсии $D_в$, выборочного среднего квадратического отклонения $\sigma_в$, коэффициента вариации V^*, асимметрии a_s^* и эксцесса ε_k^* вариационного ряда (*статистические аналоги соответствующих числовых характеристик случайной величины*) заполним вспомогательную таблицу.

i	$\tilde{x}_i$	m_i	p_i^*	$\tilde{x}_i\ p_i^*$	$\tilde{x}_i^2\ p_i^*$	$\tilde{x}_i^3\ p_i^*$	$\tilde{x}_i^4\ p_i^*$
1	68	4	0,04	2,72	184,96	12577,28	855255,04
2	72	3	0,03	2,16	155,52	11197,44	806215,68
3	76	7	0,07	5,32	404,32	30728,32	2335352,32
4	80	16	0,16	12,80	1024,00	81920,00	6553600,00
5	84	18	0,18	15,12	1270,08	106686,72	8961684,48
6	88	20	0,2	17,60	1548,80	136294,40	11993907,20
7	92	15	0,15	13,80	1269,60	116803,20	10745894,40
8	96	7	0,07	6,72	645,12	61931,52	5945425,92
9	100	6	0,06	6,00	600,00	60000,00	6000000,00
10	104	4	0,04	4,16	432,64	44994,56	4679434,24
Σ	–	**100**	**1**	**86,4**	**7535,04**	**663133,44**	**58876769,28**

Находим выборочное среднее по формуле (5):

$$\overline{x}_в = M^*[X] = \sum_{i=1}^{10} \tilde{x}_i\ p_i^* = 86,4\,;$$

выборочную дисперсию по формуле (8):

$$D_в = \sum_{i=1}^{10} \tilde{x}_i^2\ p_i^* - \overline{x}_в{}^2 = 7535,04 - (86,4)^2 = 70,08\,;$$

выборочное среднее квадратическое отклонение (9): $\sigma_в = \sqrt{D_в} = 8,37$;
исправленную выборочную дисперсию (23):

$$s^2 = \frac{n}{n-1} \cdot D_в = \frac{100}{99} \cdot 70,08 = 70,79 ;$$

исправленное выборочное среднее квадратическое отклонение:

$$s = \sqrt{s^2} = 8,41 .$$

Т. к. число наблюдений $n = 100$ достаточно велико, то вместо s^2 можно использовать неисправленную выборочную дисперсию $\sigma_в^2$.

4) Для характеристики меры колеблемости изучаемого признака относительно выборочной средней вычислим коэффициент вариации по формуле (12):

$$V = \frac{\sigma_в}{m_в} \cdot 100\,\% = \frac{8,37}{86,4} \cdot 100\,\% = 9,69\,\% ,$$

что говорит об однородности значений признака.

Найдём начальные (14) и центральные (15) моменты, используя результаты вспомогательной таблицы:

$$\nu_1^* = \sum_{i=1}^{10} \tilde{x}_i\, p_i^* = \overline{x}_в = 86,4; \qquad \nu_2^* = \sum_{i=1}^{10} \tilde{x}_i^2\, p_i^* = 7535,04;$$

$$\nu_3^* = \sum_{i=1}^{10} \tilde{x}_i^3\, p_i^* = 663133,44; \qquad \nu_4^* = \sum_{i=1}^{10} \tilde{x}_i^4\, p_i^* = 58876769,28.$$

$$\mu_3^* = 2(\nu_1^*)^3 - 3\nu_1^*\nu_2^* + \nu_3^* = 2 \cdot (86,4)^2 - 3 \cdot 86,4 \cdot 7535,04 + 663133,44 = -3,84;$$

$$\mu_4^* = \nu_4^* - 4\nu_1^*\nu_3^* + 6(\nu_1^*)^2\nu_2^* - 3(\nu_1^*)^4 = 13602,202 .$$

Выборочный коэффициент асимметрии вычислим по формуле (16):

$$a_s^* = \frac{\mu_3^*}{\sigma_в^3} = \frac{-3,84}{(8,37)^3} = -0,007 ;$$

выборочный коэффициент эксцесса по формуле (17):

$$\varepsilon_k^* = \frac{\mu_4^*}{\sigma_в^4} - 3 = \frac{13602,202}{(8,37)^4} - 3 = -0,23.$$

4. Точечной оценкой математического ожидания a является средняя выборочная $\overline{x}_в$, тогда полагаем $a = \overline{x}_в = 86,4$; точечной оценкой генерального среднего квадратического отклонения σ является исправленное выборочное среднее квадратическое отклонение, т. е. $\sigma = s = 8,41$.

а) Вид гистограммы относительных частот напоминает график плотности функции $f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x-a)^2}{2 \cdot \sigma^2}}$ нормального распределения непрерывной случайной величины X. Построим на одном чертеже с гистограммой относительных частот p_i^* (рис. 29) её теоретический аналог

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 8,41} \cdot e^{-\frac{(x-86,4)^2}{2 \cdot 8,41^2}} \text{ (рис. 31).}$$

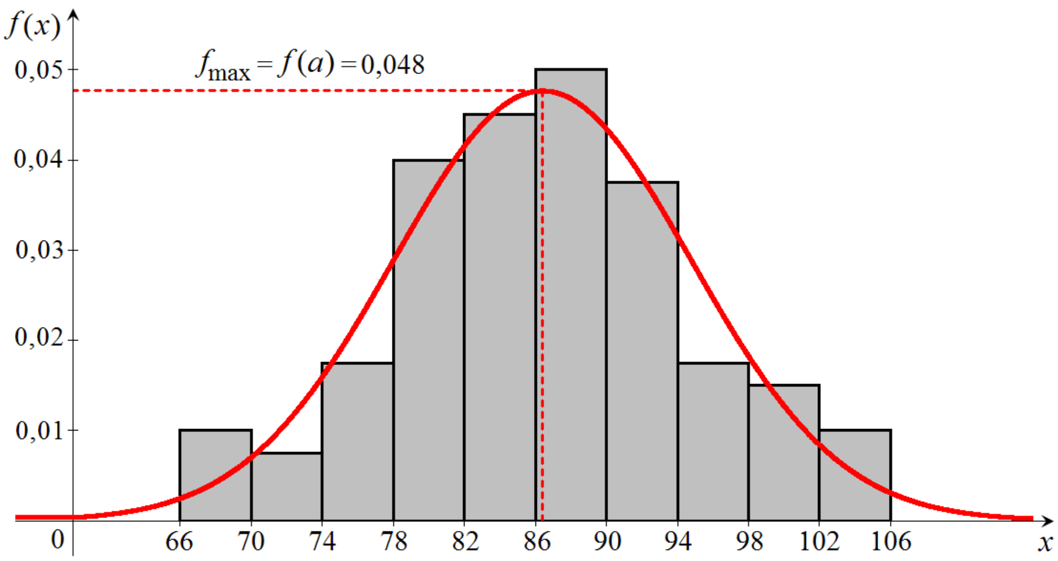


Рис. 31

б) Вид эмпирической функции распределения $F^*(x)$ напоминает интегральную функцию $F(x) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right)$ нормального распределения. Построим на одном чертеже с эмпирической функцией $F^*(x)$ (рис. 30) её теоретический аналог $F(x) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right)$, где $a = 86,4$, $\sigma = 8,41$ (рис. 32).

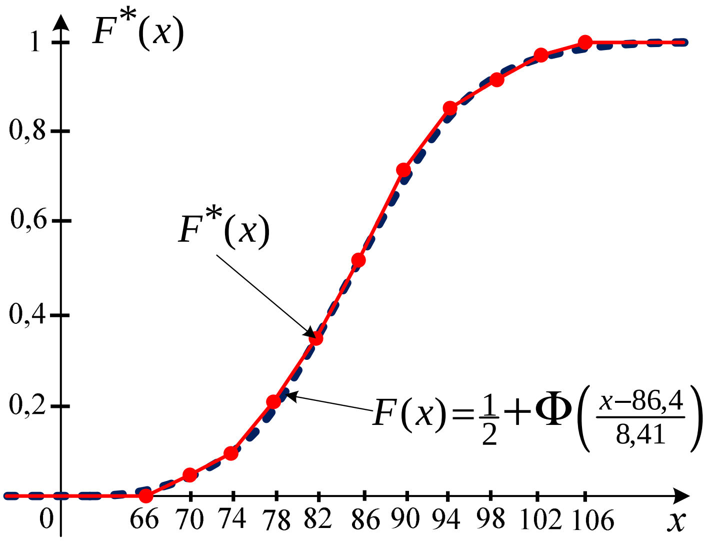


Рис. 32

5. По виду гистограммы и функции $F^*(x)$ **выдвигаем основную (нулевую) гипотезу** H_0: «Генеральная совокупность распределена по нормальному закону с параметрами $a=86,4$, $\sigma=8,41$» и **альтернативную гипотезу** H_1: «Генеральная совокупность не распределена по нормальному закону».

6. Проверим выполнение правила «трёх сигм»:

$$P\{|X-a|\le 3\sigma\}=0,9973,$$

которое требует, чтобы в 99,73 % значения случайной величины, распределенной по нормальному закону, попадали на отрезок $[a-3\sigma; a+3\sigma]$.

В нашем случае: $a-3\sigma=61,17$, $a+3\sigma=111,63$, отсюда получаем, что интервал опытных данных $[66;106]\subset[61,17;111,63]$. Таким образом, найденный промежуток полностью накрыл наши статистические значения.

Т. к. условие правила «трёх сигм» выполняется, то *есть основание предполагать, что изучаемое распределение является нормальным.*

7. Проверим соответствие выдвинутой гипотезы H_0 опытным данным. Для этого необходимо вычислить теоретические вероятности p_i и выравнивающие частоты $m_i'=np_i$.

Необходимым условием применения критерия Пирсона является наличие в каждом из интервалов не менее 5 наблюдений (т. е. $m_i\ge 5$).

Т. к. число наблюдений в крайних интервалах меньше 5, то объединим их с соседними. Получим следующий ряд распределения:

$(x_{i-1}; x_i]$	[64; 74]	(74; 78]	(78; 82]	(82; 86]	(86; 90]	(90; 94]	(94; 98]	(98; 106]
m_i	7	7	16	18	20	15	7	10

Найдём интервальные вероятности p_i. Т. к. случайная величина определена на интервале $(-\infty, +\infty)$, то крайние промежутки в ряде распределения заменяем, соответственно на $(-\infty; 74]$ и $(98; +\infty)$ (рис. 33).

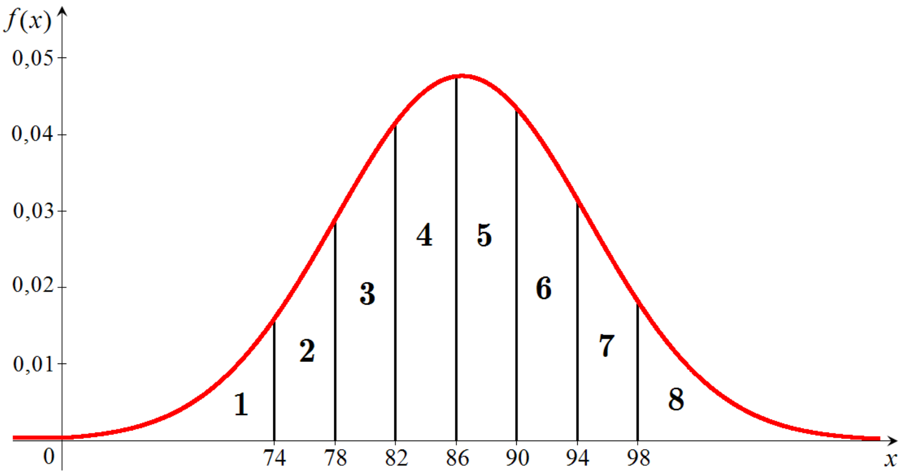


Рис. 33

Искомые вероятности вычисляем по формуле

$$p_i = P\{x_{i-1} \le X \le x_i\} = F(x_i) - F(x_{i-1}) = \Phi\left(\frac{x_i - a}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - a}{\sigma}\right).$$

Используя таблицу приложения 2 и свойства функции Лапласа, находим:

$$p_1 = P\{-\infty < X \le 74\} = F(74) - F(-\infty) = \Phi\left(\frac{74 - 86,4}{8,41}\right) - \Phi(-\infty) =$$

$$= \Phi(-1,47) + \Phi(\infty) = -\Phi(1,47) + 0,5 = -0,4292 + 0,5 = 0,0708;$$

$$p_2 = P\{74 \le X \le 78\} = \Phi\left(\frac{78 - 86,4}{8,41}\right) - \Phi\left(\frac{74 - 86,4}{8,41}\right) = \Phi(-1,00) - \Phi(-1,47) =$$

$$= -\Phi(1,00) + \Phi(1,95) = -0,3438 + 0,4292 = 0,0854;$$

$$p_3 = P\{78 \le X \le 82\} = \Phi\left(\frac{82 - 86,4}{8,41}\right) - \Phi\left(\frac{78 - 86,4}{8,41}\right) = \Phi(-0,52) - \Phi(-1,00) =$$

$$= -\Phi(0,52) + \Phi(1,00) = -0,1985 + 0,3438 = 0,1453;$$

$$p_4 = P\{82 \le X \le 86\} = \Phi\left(\frac{86-86,4}{8,41}\right) - \Phi\left(\frac{82-86,4}{8,41}\right) = \Phi(-0,05) - \Phi(-0,52) =$$

$$= -\Phi(0,05) + \Phi(0,52) = -0,0199 + 0,1985 = 0,1786;$$

$$p_5 = P\{86 \le X \le 90\} = \Phi\left(\frac{90-86,4}{8,41}\right) - \Phi\left(\frac{86-86,4}{8,41}\right) = \Phi(0,43) - \Phi(-0,05) =$$

$$= \Phi(0,43) + \Phi(0,05) = 0,1664 + 0,0199 = 0,1863;$$

$$p_6 = P\{90 \le X \le 94\} = \Phi\left(\frac{94-86,4}{8,41}\right) - \Phi\left(\frac{90-86,4}{8,41}\right) = \Phi(0,90) - \Phi(0,43) =$$

$$= 0,3159 - 0,1664 = 0,1495;$$

$$p_7 = P\{94 \le X \le 98\} = \Phi\left(\frac{98-86,4}{8,41}\right) - \Phi\left(\frac{94-86,4}{8,41}\right) = \Phi(1,38) - \Phi(0,90) =$$

$$= 0,4162 - 0,3159 = 0,1003;$$

$$p_8 = P\{98 \le X < +\infty\} = \Phi(+\infty) - \Phi\left(\frac{98-86,4}{8,41}\right) = \Phi(\infty) - \Phi(1,38) =$$

$$= 0,5 - 0,4162 = 0,0838.$$

Для дальнейших расчётов заполним вспомогательную таблицу:

i	интервал $(x_{i-1}; x_i]$	частота m_i	теоретическая частота $m_i' = np_i$	$\frac{(m_i - m_i')^2}{m_i'}$
1	$(-\infty; 74]$	7	7,08	0,0009
2	(74; 78]	7	8,54	0,2777
3	(78; 82]	16	14,53	0,1487
4	(82; 86]	18	17,86	0,0011
5	(86; 90]	20	18,63	0,1007
6	(90; 94]	15	14,95	0,0002
7	(94; 98]	7	10,03	0,9153
8	$(98; +\infty)$	10	8,38	0,3132
Σ	–	**100**	**100**	**1,7578**

Наблюдаемое значение критерия согласия Пирсона (итоговая строка таблицы) $\chi^2_{набл} = \sum_{i=1}^{8} \frac{(m_i - m_i')^2}{m_i'} = 1,7578$.

По таблице приложения 3 критических точек распределения χ^2, по заданному уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = 8 - 3 = 5$ найдём критическое значение $\chi^2_{кр}(\alpha; k) = \chi^2_{кр}(0{,}05; 5) = 11{,}1$.

Т. к. $\chi^2_{набл} < \chi^2_{кр}$, то нет оснований отвергнуть проверяемую нулевую гипотезу. Т. е. принимаем предположение, что **статистические данные распределены по нормальному закону с параметрами** $a = 86{,}4$ и $\sigma = 8{,}41$.

8. *а*) Построим на одном чертеже полигон эмпирических относительных частот p_i^* и кривую распределения вероятностей p_i (рис. 34).

– кривая p_i^*.

– кривая p_i.

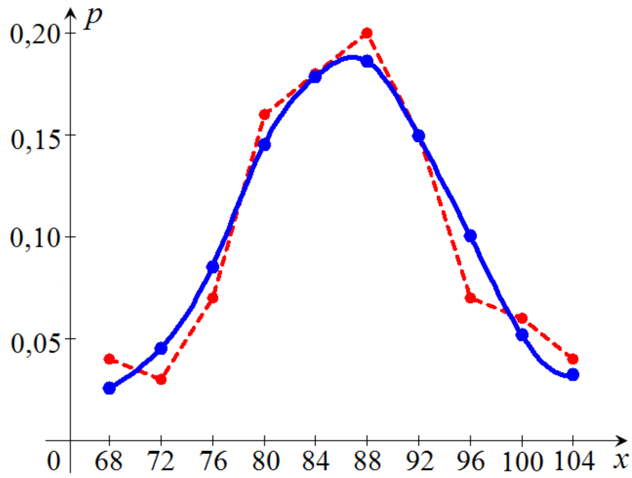


Рис. 34

Полигон относительных частот p_i^* представляет данные, взятые по выборке. Кривая p_i выравнивает эмпирические данные, тем самым приближая распределение генеральной совокупности к нормальному.

б) Построим на одном чертеже гистограмму теоретических вероятностей (относительных частот) p_i и график функции $f(x)$ (рис. 35).

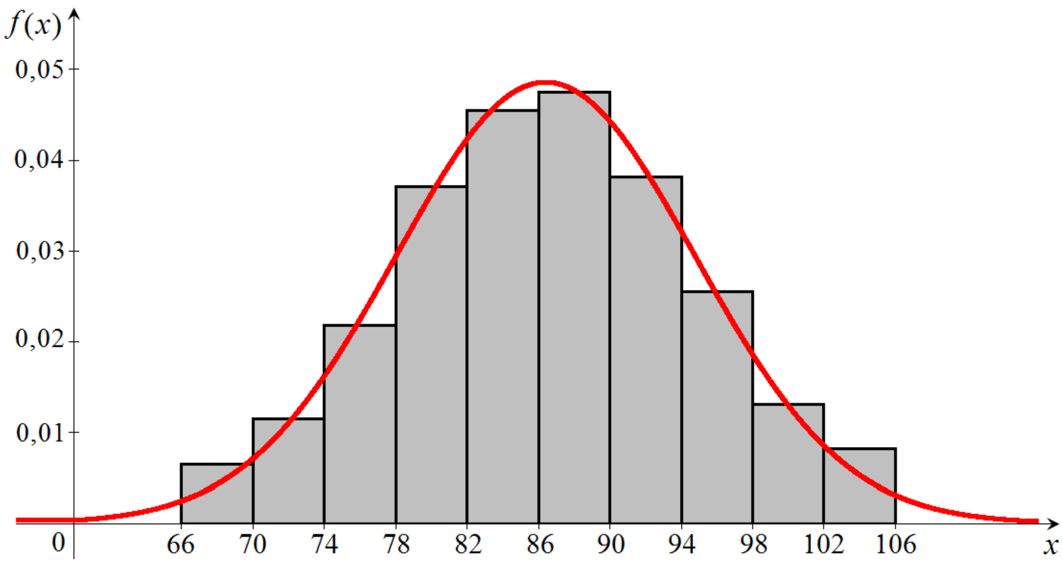


Рис. 35

Т. к. $a_s^* < 0$, то «длинная» часть кривой теоретического распределения расположена несколько левее от $M_o^* = 87{,}14$, при этом кривая является более пологой ($\varepsilon_k^* < 0$), т. е. менее островершинной, чем идеально нормальная кривая.

9. Т. к. $\sigma_г$ неизвестно, то доверительный интервал для генеральной средней $x_г$ имеет вид

$$\overline{x}_в - \varepsilon < \overline{x}_г < \overline{x}_в + \varepsilon,$$

где величина ε определяется по формуле (30): $\varepsilon = \dfrac{t_\gamma \cdot s}{\sqrt{n}}$.

Значение $t_\gamma = t(\gamma, n)$ находим в таблице приложения 6 по заданному уровню надёжности $\gamma = 0{,}99$ и $n = 100$: $t(0{,}99;100) = 2{,}627$.

Тогда $\varepsilon = \dfrac{2{,}627}{\sqrt{100}} \cdot 8{,}41 = 2{,}21$.

Таким образом, получаем доверительный интервал для $\overline{x}_г$:

$$86{,}4 - 2{,}21 < \overline{x}_г < 86{,}4 + 2{,}21 \quad \Rightarrow \quad 84{,}19 < \overline{x}_г < 88{,}61.$$

Это означает, что в 99 % случаев истинное значение генеральной средней $\overline{x}_г$ находится в промежутке $(84{,}19; 88{,}61)$.

Доверительный интервал для генерального среднего квадратического отклонения определяется по формуле (34):

$$s - sq < \sigma_г < s + sq.$$

По таблице приложения 7 величина $q(\gamma, n) = q(0{,}99;100) = 0{,}198 < 1$. Отсюда получаем:

$$8{,}41 - 8{,}41 \cdot 0{,}198 < \sigma_г < 8{,}41 + 8{,}41 \cdot 0{,}198 \quad \Rightarrow \quad 6{,}74 < \sigma_г < 10{,}08.$$

Это означает, что в 99 % истинное значение генерального среднего квадратического отклонения $\sigma_г$ находится промежутке (6,74; 10,08).

☻

Задание **№ 2.** При обследовании 2000 тепличных хозяйств было отобрано 110 теплиц. Распределение их по объёму совокупных ежегодных продаж (ден. ед.) приведено в таблице:

Размер объёма совокупных ежегодных продаж, ден. ед.	менее 500	500-1000	1000-1500	1500-2000	2000-2500	Всего
Число теплиц	8	20	52	18	12	110

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что случайная величина X – размер объёма совокупных ежегодных продаж распределена по нормальному закону.

2. Найти:

а) вероятность того, что средний объём продаж во всех тепличных хозяйствах отличается от среднего объёма продаж в выборке не более чем на 100 ден. ед. (по абсолютной величине);

б) границы, в которых с вероятностью 0,97 заключена доля теплиц, объём продаж которых не более 1000 ден. ед.;

в) каким должен быть объём выборки, чтобы те же границы для доли теплиц, объём продаж которых не более 1000 ден. ед., можно было гарантировать с вероятностью 0,999?

☺ *Решение.*

1. По условию $N = 2000$, $n = 110$. Найдём середины интервалов.

$\tilde{x}_i$	250	750	1250	1750	2250	Всего
m_i	8	25	47	18	12	110

Найдём числовые характеристики выборки (формулы (5), (8), (23)):

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{5}\tilde{x}_i m_i = \frac{1}{110}(250\cdot 8 + 750\cdot 25 + 1250\cdot 47 + 1750\cdot 18 + 250\cdot 12) = 1255;$$

$$D_в = \frac{1}{n}\sum_{i=1}^{5}\tilde{x}_i^2 m_i - \overline{x}_в^2 = 279524{,}7934,\ \sigma_в = \sqrt{D_в} = 528{,}701;$$

$$s = \sqrt{\frac{n}{n-1}}\cdot\sigma_в = 531{,}121.$$

Используя критерий согласия Пирсона, при уровне значимости $\alpha = 0{,}05$ проверим гипотезу H_0: о нормальном распределении случайной величины X с параметрами $a = \overline{x}_в = 1255$ и $\sigma = s = 531{,}121$ при альтернативной гипотезе H_1: «Случайная величина X не распределена по нормальному закону».

Вычислим вероятности p_i попадания случайной величины X в заданные интервалы с помощью функции Лапласа (табл. приложения 2):

$$p_i = \Phi\left(\frac{x_i - a}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - a}{\sigma}\right).$$

$p_1 = P\{-\infty < x \le 500\} = \Phi(-1,42) - \Phi(-\infty) = -0,4222 + 0,5 = 0,0778$;

$p_2 = P\{500 \le x < 1000\} = \Phi(-0,48) - \Phi(-1,42) = -0,1844 + 0,4222 = 0,2378$;

$p_3 = P\{1000 \le x < 1500\} = \Phi(0,46) - \Phi(-0,48) = 0,1772 + 0,1844 = 0,3616$;

$p_4 = P\{1500 \le x < 2000\} = \Phi(1,40) - \Phi(0,46) = 0,4192 - 0,1772 = 0,2420$;

$p_5 = P\{x \ge 2000\} = \Phi(+\infty) - \Phi(1,40) = 0,5 - 0,4192 = 0,0808$.

Для проведения расчётов заполним вспомогательную таблицу:

i	интервал $(x_{i-1}; x_i]$	частота m_i	теоретическая частота $m_i' = np_i$	$\frac{(m_i - m_i')^2}{m_i'}$
1	менее 500	8	7,78	0,0364
2	500-1000	25	23,78	0,0513
3	1000-1500	47	36,16	1,3120
4	1500-2000	18	24,20	2,7913
5	2000-2500	12	8,08	1,0896
Σ	–	**110**	**110**	**5,2806**

Наблюдаемое значение критерия согласия: $\chi^2_{набл} = \sum_{i=1}^{5} \frac{(m_i - m_i')^2}{m_i'} = 5,2806$.

По таблице приложения 3 по заданному уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $l = 5 - 3 = 2$ найдём критическое значение $\chi^2_{кр}(\alpha; l) = \chi^2_{кр}(0,05; 2) = 5,99$.

Т. к. $\chi^2_{набл} < \chi^2_{кр}$, то нулевая гипотеза о нормальном распределении принимается как не противоречащая опытным данным.

2. *а*) Для вычисления искомой вероятности применим формулу $P\{|\overline{x}_г - \overline{x}_в| \le \varepsilon\} = 2\Phi(t)$, где $\varepsilon = 100$, t – аргумент функции Лапласа, который в случае неизвестного $\sigma_г$ и известного объёма генеральной совокупности N, определяется по формуле (35).

Имеем $t = \dfrac{\varepsilon}{\sqrt{\dfrac{s^2}{n}\cdot\left(1-\dfrac{n}{N}\right)}} = \dfrac{100}{\sqrt{\dfrac{282089{,}241}{110}\cdot\left(1-\dfrac{110}{2000}\right)}} = \dfrac{100}{49{,}2281} = 2{,}03\cdot$

Тогда $P\{|\overline{x}_г - \overline{x}_в| \le 100\} = 2\,\Phi(2{,}03) = 2\cdot 0{,}4788 = 0{,}9576$.

б) По исходной таблице найдём долю теплиц с объёмом продаж не более 1000 ден. ед.: $p^* = \dfrac{m}{n} = \dfrac{8+25}{110} = \dfrac{33}{110} = 0{,}3$.

Доля $p^* = 0{,}3$ с вероятностью 0,97 попадает в некоторый интервал $(p^* - \varepsilon, p^* + \varepsilon)$, где значение ε определяется по формуле (29).

Т. к. $\gamma = 0{,}97 = P\{|p - p^*| < \varepsilon\} = 2\,\Phi(t)$, то $\Phi(t) = \dfrac{0{,}97}{2} = 0{,}485$.

По таблице приложения 2 находим $t = 2{,}17$.

Тогда $\varepsilon = t\sqrt{\dfrac{p^*(1-p^*)}{n}\left(1-\dfrac{n}{N}\right)} = 2{,}17\cdot\sqrt{\dfrac{33}{110}\left(1-\dfrac{33}{110}\right)\cdot\dfrac{1}{110}\left(1-\dfrac{110}{2000}\right)} = 0{,}0922$.

Окончательно находим доверительные границы:

$$0{,}3 - 0{,}0922 < p < 0{,}3 + 0{,}0922;$$

$$0{,}2078 < p < 0{,}3922.$$

в) Объём выборки определяем по формуле (27): $n = \dfrac{p^*(1-p^*)}{\left(\dfrac{\varepsilon}{t}\right)^2 + \dfrac{p^*(1-p^*)}{N}}$,

где $\varepsilon = 0{,}095$ (см. п. *б*), т. к. по условию задачи границы те же).

Для доверительной вероятности $2\Phi(t) = 0{,}999$ находим по таблице приложения 2 значение аргумента $t = 3{,}3$.

Имеем $n = \dfrac{0{,}3\cdot 0{,}7}{\left(\dfrac{0{,}0922}{3{,}3}\right)^2 + \dfrac{0{,}3\cdot 0{,}7}{2000}} = 237{,}2615$.

Окончательно $n = 237$. ☻

Задание **№ 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 40$ является математическим ожиданием нормально распределённой случайной величины при 5% уровне значимости для двусторонней и односторонней критической области, если в результате обработки выборки объёма $n = 10$ получено выборочное среднее $\overline{x}_в = 38$, а несмещённое среднее квадратичное отклонение равно $s = 3,6$.

☺ *Решение.*

По условию $a_0 = 40$, $\overline{x}_в = 38$, $\alpha = 0,05$, $n = 10$, $s = 3,6$, неисправленное выборочное квадратическое отклонение $\sigma_в = \sqrt{\dfrac{n}{n-1}} \cdot s = 3,795$.

Проверяем нулевую гипотезу: «Заданное значение $a_0 = 40$ является математическим ожиданием нормально распределенной случайной величины с уровнем значимости $\alpha = 0,05$».

1) Для двусторонней критической области:

основная гипотеза $H_0 : a = a_0 = 40$;   альтернативная гипотеза $H_1 : a \neq 40$.

По выборочным данным вычисляем наблюдаемое значение статистического критерия (табл. 7): $T_{набл} = \dfrac{(\overline{x}_в - a_0) \cdot \sqrt{n}}{s} = -1,76$.

Критическая точка $t_{кр} = t_{\alpha, n-1} = t_{0,05;9} = 2,26$ определяется по таблице приложения 4 (двусторонняя критическая область).

Т. к. $|T_{набл}| < t_{кр}$, то принимаем основную гипотезу «$a = 40$».

2) Для левосторонней критической области:

основная гипотеза $H_0 : a = a_0 = 40$;   альтернативная гипотеза $H_1 : a < 40$.

Критическая точка $t_{кр} = t_{\alpha, n-1} = t_{0,05;9} = 1,83$ определяется по таблице приложения 4 (односторонняя критическая область).

Т. к. $|T_{набл}| < t_{кр}$, то принимаем основную гипотезу «$a = 40$».

3) Для правосторонней критической области:

основная гипотеза $H_0 : a = a_0 = 40$;   альтернативная гипотеза $H_1 : a > 40$.

Критическая точка $t_{кр} = t_{\alpha, n-1} = t_{0,05;9} = 1,83$ определяется по таблице приложения 4 (односторонняя критическая область).

Т. к. $|T_{набл}| < t_{кр}$, то принимаем основную гипотезу «$a = 40$». ☻

***Задание* № 4.** При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин X и Y (при альтернативной гипотезе H$_1$: $\sigma_x^2 \neq \sigma_y^2$) на основе выборочных данных.

X		Y	
x_i	m_{x_i}	y_i	m_{y_i}
12,1	1	12,2	1
12,5	2	12,4	8
12,7	4	12,5	1
13,0	2	12,7	2
13,2	2	13,0	1

☺ *Решение*. Проверим гипотезу $H_0 : D_x = D_y$ при альтернативной гипотезе $H_1 : D_x \neq D_y$.

Определим выборочные средние и выборочные дисперсии случайных величин X и Y, используя соответствующие ряды:

x_i	12,1	12,5	12,7	13,0	13,2
m_{x_i}	1	2	4	2	2

y_i	12,2	12,4	12,5	12,7	13,0
m_{y_i}	1	8	1	2	1

$$n_x = \sum_{i=1}^{5} m_{x_i} = 11. \qquad n_y = \sum_{i=1}^{5} m_{y_i} = 13.$$

$$\bar{x} = \frac{1}{n_x}\sum_{i=1}^{5} x_i m_{x_i} = 12{,}75. \qquad \bar{y} = \frac{1}{n_y}\sum_{i=1}^{5} y_i m_{y_i} = 12{,}48.$$

$$D_x = \frac{1}{n_x}\sum_{i=1}^{5} x_i^2 m_{x_i} - \bar{x}^2 = 0{,}215. \qquad D_y = \frac{1}{n_y}\sum_{i=1}^{5} y_i^2 m_{y_i} - \bar{y}^2 = 0{,}153.$$

Исправленные выборочные дисперсии:

$$s_x^2 = \frac{n_x}{n_x - 1} D_x = 0{,}237; \qquad s_y^2 = \frac{n_y}{n_y - 1} D_y = 2{,}192.$$

Найдём значение критерия Фишера-Снедекора (табл. 8):

$$F_{набл} = \frac{s_{\max}^2}{s_{\min}^2} = \frac{s_y^2}{s_x^2} = 2,626,$$

здесь рассматривается отношение бóльшей дисперсии к меньшей.

По условию конкурирующая гипотеза имеет вид $D_x \neq D_y$, поэтому критическая область – двухсторонняя. По таблице приложения 5, по уровню значимости, вдвое меньше заданного, т. е. при $\alpha/2 = 0{,}1/2 = 0{,}05$, и числам степеней свободы $k_1 = n_y - 1 = 12$ и $k_2 = n_x - 1 = 10$ находим критическое значение распределения Фишера-Снедекора $F_{кр}(0,05;12;10) = 2,91$.

Т. к. $F_{набл} < F_{кр}$, то принимаем нулевую гипотезу о равенстве дисперсий. ☻

***Задание* № 5.** Распределение 100 предприятий отрасли по объёму выпускаемой продукции *X* (тыс. единиц) и её себестоимости *Y* (руб.) представлено в таблице:

$x \backslash y$	**48**	**58**	**68**	**78**	m_{x_i}
17	–	–	–	1	**1**
18	–	–	29	1	**30**
19	–	29	10	–	**39**
20	21	6	–	–	**27**
21	3	–	–	–	**3**
m_{y_j}	**24**	**35**	**39**	**2**	**100**

Необходимо:

1. Отметить на координатной плоскости точки $(x_i; y_i)$ данной выборки.

2. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$.

3. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:

а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;

б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;

в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

☺ *Решение.*

1. Отметим на координатной плоскости точки данной выборки:

$(17;78)$, $(18;68)$, $(18;78)$, $(19;58)$,

$(19;68)$, $(20;48)$, $(20;58)$, $(21;48)$.

Заметим, что точки располагаются вдоль некоторой прямой линии (рис. 36).

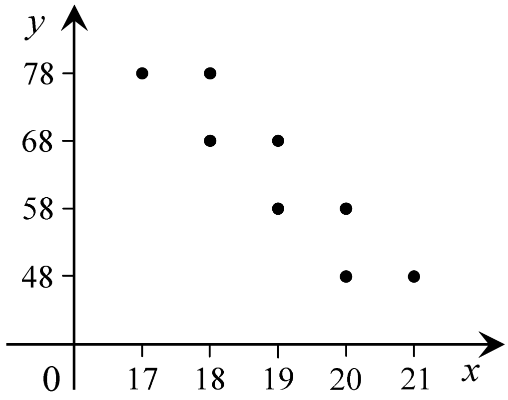


Рис. 36

2. Для проведения расчётов заполним вспомогательные таблицы:

x	m_x	xm_x	x^2m_x
17	1	17	289
18	30	540	9720
19	39	741	14079
20	27	540	10800
21	3	63	1323
Σ	**100**	**1901**	**36211**

$\overline{x} = \dfrac{1901}{100} = 19{,}01$;

$\overline{x^2} = \dfrac{36211}{100} = 362{,}11$;

$\sigma_x^2 = \overline{x^2} - \overline{x}^2 = 362{,}11 - 19{,}01^2 = 0{,}7299$;

$\sigma_x = \sqrt{D_x} = \sqrt{0{,}7299} = 0{,}854$.

y	m_y	ym_y	y^2m_y
48	24	1152	55296
58	35	2030	117740
68	39	2652	180336
78	2	156	12168
Σ	**100**	**5990**	**365540**

$\overline{y} = \dfrac{5990}{100} = 59{,}9$;

$\overline{y^2} = \dfrac{365540}{100} = 3655{,}4$;

$\sigma_y^2 = \overline{y^2} - \overline{y}^2 = 3655{,}4 - 59{,}9^2 = 67{,}39$;

$\sigma_y = \sqrt{D_y} = \sqrt{67{,}39} = 8{,}209$.

x	y	m_{xy}	xym_{xy}
17	78	1	1326
18	68	29	35496
18	78	1	1404
19	58	29	31958
19	68	10	12920
20	48	21	20160
20	58	6	6960
21	48	3	3024
Σ		**100**	**113248**

$\overline{xy} = \dfrac{113248}{100} = 1132{,}48$.

Найдём условные средние значения $\bar{x}_y$ и $\bar{y}_x$ и межгрупповые дисперсии:

$$\bar{x}_{y=48} = \frac{20\cdot 21 + 21\cdot 3}{24} = 20{,}1.$$

$$\bar{x}_{y=58} = \frac{19\cdot 29 + 20\cdot 6}{35} = 19{,}2.$$

$$\bar{x}_{y=68} = \frac{18\cdot 29 + 19\cdot 10}{39} = 18{,}3.$$

$$\bar{x}_{y=78} = \frac{17\cdot 1 + 18\cdot 1}{2} = 17{,}5.$$

$\Rightarrow$

$$\delta_x^2 = \frac{1}{100}\Big((20{,}1-19{,}01)^2\cdot 24 + (19{,}2-19{,}01)^2\cdot 35 + (18{,}3-19{,}01)^2\cdot 39 + (17{,}5-19{,}01)^2\cdot 2\Big) = 0{,}54.$$

$$\delta_x = \sqrt{0{,}54} = 0{,}73.$$

$$\bar{y}_{x=17} = \frac{78\cdot 1}{1} = 78.$$

$$\bar{y}_{x=18} = \frac{68\cdot 29 + 78\cdot 1}{30} = 68{,}3.$$

$$\bar{y}_{x=19} = \frac{58\cdot 29 + 68\cdot 10}{39} = 60{,}6.$$

$$\bar{y}_{x=20} = \frac{48\cdot 21 + 58\cdot 6}{27} = 50{,}2.$$

$$\bar{y}_{x=21} = \frac{48\cdot 3}{3} = 48.$$

$\Rightarrow$

$$\delta_y^2 = \frac{1}{100}\Big((78-59{,}9)^2\cdot 1 + (68{,}3-59{,}9)^2\cdot 30 + (60{,}6-59{,}9)^2\cdot 39 + (50{,}2-59{,}9)^2\cdot 27 + (48-59{,}9)^2\cdot 3\Big) = 54{,}3.$$

$$\delta_y = \sqrt{54{,}3} = 7{,}37.$$

3. Найдём корреляционные отношения по формулам (32):

$$\eta_{XY} = \frac{\delta_x}{\sigma_x} = \frac{0{,}73}{0{,}854} = 0{,}85 \text{ и } \eta_{YX} = \frac{\delta_y}{\sigma_y} = \frac{7{,}37}{8{,}209} = 0{,}898.$$

Следовательно, между случайными величинами X и Y имеется сильная (тесная) корреляционная зависимость (табл. 15).

а) Найдём коэффициент корреляции по формуле (33):

$$r_в = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x\cdot\sigma_y} = \frac{1132{,}48 - 19{,}01\cdot 59{,}9}{0{,}854\cdot 8{,}209} = -0{,}89,$$

Значит, имеется сильная убывающая линейная зависимость (табл. 16).

Проверим гипотезу $H_0 : r_г = 0$ (линейная зависимость отсутствует) при альтернативе $H_1 : r_г \neq 0$ (имеется линейная зависимость).

По формуле (34) вычислим значение статистики

$$T_{набл} = \frac{r_в \cdot \sqrt{n-2}}{\sqrt{1-r_в^2}} = \frac{-0{,}89 \cdot \sqrt{100-2}}{\sqrt{1-(-0{,}89)^2}} = -19{,}32.$$

По условию $\alpha = 0{,}05$, $k = 100 - 2 = 98$, тогда по таблице приложения 4 находим критическое значение $t_{кр} = t_{кр}(0{,}05;98) = 1{,}98$.

Т. к. $|T_{набл}| > t_{кр}$, то нулевую гипотезу отвергаем и принимаем гипотезу H_1 о наличии линейной корреляции.

б) Составим уравнения линейной регрессии по формулам (35):

$y_x - \bar{y} = \rho_{YX}(x - \bar{x})$	$x_y - \bar{x} = \rho_{XY}(y - \bar{y})$
$\rho_{YX} = \frac{\sigma_y}{\sigma_x} \cdot r_в = \frac{8{,}209}{0{,}854} \cdot (-0{,}89) = -8{,}6$	$\rho_{XY} = \frac{\sigma_x}{\sigma_y} \cdot r_в = \frac{0{,}854}{8{,}209} \cdot (-0{,}89) = -0{,}1$
$y_x - 59{,}9 = -8{,}6 \cdot (x - 19{,}01)$	$x_y - 19{,}01 = -0{,}1(y - 59{,}9)$
$y_x = -8{,}6x + 223{,}4$	$x_y = -0{,}1y + 25$ (или $y = -10x + 250$)
Полученное уравнение показывает, что при увеличении объёма выпуска X на 1 тыс. единиц себестоимость Y уменьшается в среднем на 8,6 (руб.).	Полученное уравнение показывает, что для уменьшения себестоимости Y на 1 руб. необходимо в среднем увеличить объём выпуска X на 0,1 тыс. единиц.

Построим на одном чертеже (рис. 37) графики прямых регрессии:

$$y_x = -8{,}6x + 223{,}4 \text{ и } x_y = -0{,}1y + 25$$

и отметим экспериментальные данные (условные средние $\bar{x}_y$ и $\bar{y}_x$).

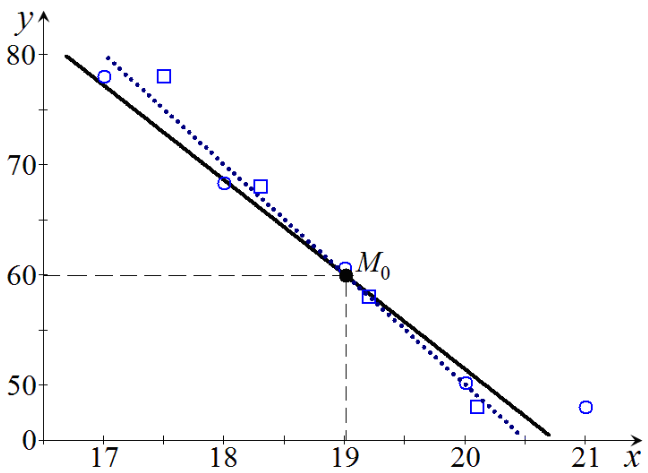


Рис. 37

Обозначения:

– прямая $x_y = -0{,}1y + 25$

– прямая $y_x = -8{,}6x + 223{,}4$

– условные средние $\bar{x}_y$

– условные средние $\bar{y}_x$

– $M_0(\bar{x};\bar{y})$ – точка пересечения прямых регрессии, где $\bar{x} = 19{,}01$, $\bar{y} = 59{,}9$

в) Вычислим показатели качества найденной модели регрессии.

Заполним вспомогательные таблицы:

i	x_i	$\overline{y}_i$	$y_{x_i} = -8,6x_i + 223,4$	$\varepsilon_{y_i} = \overline{y}_i - y_{x_i}$	$\varepsilon_{y_i}^2$	$\dfrac{\varepsilon_{y_i}}{\overline{y}_i}$
1	17	78	77,2	0,8	0,64	0,010
2	18	68,3	68,6	– 0,3	0,09	0,004
3	19	60,6	60	0,6	0,36	0,006
4	20	50,2	51,4	– 1,2	1,44	0,010
5	21	48	42,8	5,2	27,04	0,108
Σ	**–**	**–**	**–**		**29,57**	**0,138**

j	y_j	$\overline{x}_j$	$x_{y_j} = -0,1y_j + 25$	$\varepsilon_{x_j} = \overline{x}_j - x_{y_j}$	$\varepsilon_{x_j}^2$	$\dfrac{\varepsilon_{x_j}}{\overline{x}_j}$
1	48	20,1	20,2	– 0,1	0,01	0,005
2	58	19,2	19,2	0	0	0
3	68	18,3	18,2	0,1	0,01	0,005
4	78	17,5	17,2	0,3	0,09	0,017
Σ	**–**	**–**	**–**	**–**	**0,11**	**0,027**

Найдём теоретический коэффициент детерминации по формуле (40):

$$R^2 = r_в^2 = (-0,89)^2 = 0,7921.$$

Таким образом, $79,21\%$ *вариации себестоимости продукции объясняется уравнением линейной регрессии, остальные* $20,79\%$ *вариации себестоимости обусловлены влиянием не учтённых в модели факторов*.

Найдём средние квадратические ошибки по формулам (41):

$$S_{\varepsilon_y} = \sqrt{\frac{1}{5-2}\sum_{i=1}^{5}\varepsilon_{y_i}^2} = \sqrt{\frac{29,57}{3}} = 3,14 \text{ и } S_{\varepsilon_x} = \sqrt{\frac{1}{4-2}\sum_{i=1}^{4}\varepsilon_{x_i}^2} = \sqrt{\frac{0,11}{2}} = 0,23.$$

Поскольку $S_{\varepsilon_y} < \sigma_y$ и $S_{\varepsilon_x} < \sigma_x$, то *найденные модели линейной регрессии целесообразно использовать.*

Найдём средние ошибки аппроксимации по формулам (42):

$$A_y = \frac{1}{5}\sum_{i=1}^{5}\left|\frac{\varepsilon_{y_i}}{\overline{y}_i}\right| = \frac{0,138}{5} = 0,0276 \text{ и } A_x = \frac{1}{4}\sum_{i=1}^{4}\left|\frac{\varepsilon_{x_j}}{\overline{x}_j}\right| = \frac{0,027}{4} = 0,00675.$$

Средние ошибки составляют 2,76% и 0,675%, *что свидетельствует о незначительных погрешностях моделей.*

Варианты семестровых заданий

ВАРИАНТ № 1

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

47	45	56	44	47	50	62	63	38	56	38	46	52	54	55	52	51	56	45	42
39	56	61	45	50	41	53	46	60	42	55	44	34	39	35	32	49	58	60	55
40	45	53	62	50	38	44	51	53	62	51	58	41	52	38	65	42	72	60	49
49	67	48	46	63	42	55	51	49	49	39	49	46	67	58	41	45	50	71	39
44	62	34	50	38	56	61	53	43	66	54	51	54	64	48	48	57	46	65	55

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,01$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,95$.

***Задание* № 2.** Для определения среднего возраста 1000 учителей города было отобрано 100 учителей. Полученное распределение приведено в таблице:

Возраст, лет	25-30	30-35	35-40	40-45	45-50	50-55	Итого
Число учителей	10	16	29	22	15	8	100

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что случайная величина X – возраст учителей в данном городе распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:
а) вероятность того, что в данном городе доля учителей, возраст которых более 30 лет, отличается от доли таких учителей в выборке не более чем на 5 % (по абсолютной величине);
б) границы, в которых с вероятностью 0,95 заключён средний возраст учителей города;
в) каким должен быть объём выборки, чтобы те же границы среднего возраста можно было гарантировать с вероятностью 0,99?

Задание **№ 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 10$ является математическим ожиданием нормально распределённой случайной величины при H_1: $a \neq 10$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 10$ получено выборочное среднее $\overline{x}_в = 12$, а несмещённое среднее квадратическое отклонение $s = 1$.

Задание **№ 4.** При уровне значимости $\alpha = 0,01$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 < \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
37	1	38	4
38	3	39	3
40	4	40	2
41	3	41	2
42	2	43	1

Задание **№ 5.** Распределение 100 предприятий по сумме отчислений в пенсионный фонд X (тыс. руб.) и на социальное страхование работников Y (тыс. руб.) представлено в таблице:

x \ *y*	100-200	200-300	300-400	400-500	500-600	Итого:
50-150	5	3	–	–	–	**8**
150-250	7	8	–	–	–	**15**
250-350	–	8	13	5	–	**26**
350-450	–	4	10	8	6	**28**
450-550	–	–	9	6	8	**23**
Итого:	**12**	**23**	**32**	**19**	**14**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 2

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

42	40	35	29	58	46	39	46	45	38	53	31	39	53	50	21	57	25	30	47
23	56	42	39	46	45	45	47	25	48	56	22	40	41	41	33	45	34	43	37
29	41	36	31	35	63	56	42	40	48	39	33	39	20	37	34	33	38	40	61
45	45	43	39	26	14	36	55	48	34	43	44	37	37	31	58	39	47	56	41
57	31	27	46	42	43	47	44	46	37	34	45	30	46	37	33	28	56	42	29

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,01$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,99$.

***Задание* № 2.** Из 500 рабочих, обслуживающих цех производства окиси этилена, было отобрано 100 человек для контрольной проверки коэффициента использования рабочего времени в отчётном году по сравнению с предыдущим. Были получены следующие данные:

Коэффициент использования рабочего времени	0,8-0,84	0,84-0,88	0,88-0,92	0,92-0,96	0,96-1	Итого
Число рабочих	12	17	33	27	11	100

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,025$ проверить гипотезу о том, что случайная величина *X* – средний коэффициент использования рабочего времени распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:
а) вероятность того, что средний коэффициент использования рабочего времени в отчётном году отличается от идеального среднего коэффициента не более чем на 0,01 (по абсолютной величине);
б) границы, в которых с вероятностью 0,9606 заключен средний коэффициент использования рабочего времени одного рабочего этого цеха;
в) каким должен быть объём выборки, чтобы с вероятностью 0,7373 можно было утверждать, что доля рабочих в цехе, коэффициент использования рабочего времени которых будет менее 0,88, отличалась от доли таких рабочих в выборке не более чем на 0,02 (по абсолютной величине)?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 20$ является математическим ожиданием нормально распределённой случайной величины при H_1: $a \neq 20$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 18$ получено выборочное среднее $\overline{x}_в = 22$ и известно генеральное среднее квадратическое отклонение $\sigma = 4$.

***Задание* № 4.** При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H_1: $\sigma_x^2 \neq \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
142	3	140	5
145	2	146	3
146	2	147	2
148	3	151	2

***Задание* № 5.** Распределение 100 предприятий общественного питания по количеству мест в обеденном зале *X* и пропускной способности *Y* (человек) представлено в таблице:

x \ *y*	500-600	600-700	700-800	800-900	900-1000	Итого:
30-50	2	–	–	–	–	**2**
50-70	4	3	–	–	–	**7**
70-90	–	6	8	7	1	**22**
90-110	–	–	20	10	5	**35**
110-130	–	–	–	4	30	**34**
Итого:	**6**	**9**	**28**	**21**	**36**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0{,}05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 3

Задание **№ 1.** В результате опыта была получена выборочная совокупность.

73	85	108	55	61	95	93	67	86	96	69	87	90	75	92	93	95	94	87	75
77	59	98	68	94	78	86	84	82	87	70	67	74	70	80	75	71	89	85	68
70	75	101	89	61	101	91	70	73	86	75	79	79	68	74	87	79	72	67	89
92	69	86	96	70	69	69	85	81	84	92	95	61	65	83	69	92	74	68	55
69	87	79	74	75	73	81	68	81	75	81	78	95	76	72	82	74	69	75	67

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.
2. По сгруппированным данным построить:
а) полигон относительных частот;
б) гистограмму относительных частот;
в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.
4. Построить:
а) на чертеже гистограммы её теоретический аналог $f(x)$;
б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;
5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.
6. Проверить выполнение правила «трёх сигм».
7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,01$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.
8. Построить на одном чертеже:
а) полигон относительных частот p_i^* и кривую распределения p_i;
б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.
9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,999$.

Задание **№ 2.** Из партии, содержащей 10000 деталей, было отобрано 500 деталей, распределение которых по длине дано в таблице:

Длина, мм	30-32	32-34	34-36	36-38	38-40	Итого
Число деталей	10	40	270	150	30	500

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,025$ проверить гипотезу о том, что случайная величина X – длина детали распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.
2. Найти:
а) вероятность того, что средняя длина деталей во всей партии отличается в выборке не более чем на 0,1 мм (по абсолютной величине);

б) границы, в которых с вероятностью 0,95 заключена доля деталей с длиной менее 36 мм во всей партии;
в) каким должен быть объём выборки, чтобы с вероятностью 0,99 можно было гарантировать те же границы для доли деталей с длиной менее 36 мм во всей партии?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 20$ является математическим ожиданием нормально распределённой случайной величины при $H_1:\ a < 20$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 25$ получено выборочное среднее $\bar{x}_в = 18$, а несмещённое среднее квадратическое отклонение равно $s = 4$.

***Задание* № 4.** При уровне значимости $\alpha = 0,05$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 > \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
39	4	75	1
43	2	80	2
45	3	84	3
47	3	91	4
51	1	94	2

***Задание* № 5.** Распределение месячных расходов на командировки *X* (тыс. руб.) и сумму *Y* (млн. руб.), на которые заключены договора о поставке продукции, представлено в таблице:

x \ *y*	15-17	17-19	19-21	21-23	23-25	Итого:
50-70	7	5	4	–	–	**16**
70-90	8	7	9	–	–	**24**
90-110	–	10	8	6	–	**24**
110-130	–	–	6	8	9	**23**
130-150	–	–	–	5	8	**13**
Итого:	**15**	**22**	**27**	**19**	**17**	**100**

Необходимо:
1. Вычислить групповые средние $\bar{x}_{y_j}$ и $\bar{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 4

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

60	71	62	57	81	55	59	47	75	56	61	60	63	65	59	61	65	58	76	49
65	64	59	76	58	52	70	77	67	50	65	53	56	64	55	77	51	61	73	64
45	53	45	58	57	60	48	71	33	65	50	80	58	67	71	51	51	49	66	63
67	60	67	61	58	36	75	47	68	63	77	75	62	75	70	75	66	53	63	60
68	67	55	75	71	59	77	58	65	57	55	28	74	71	47	73	40	45	37	66

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:
- *а*) полигон относительных частот;
- *б*) гистограмму относительных частот;
- *в*) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:
- *а*) на чертеже гистограммы её теоретический аналог $f(x)$;
- *б*) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,025$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:
- *а*) полигон относительных частот p_i^* и кривую распределения p_i;
- *б*) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,95$.

***Задание* № 2.** Для определения средней суммы вкладов в сберегательной кассе, имеющей 4000 вкладчиков, проведено обследование 200 вкладчиков, которое дало следующие результаты:

Сумма вклада	100-200	200-300	300-400	400-500	500-600	Итого
Число вкладчиков	4	36	88	60	12	200

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что случайная величина X – сумма вкладов распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:

а) вероятность того, что доля всех вкладчиков, сумма вкладов которых составляет не менее 400 тыс. руб., отличается от их доли в выборке не более чем на 2 тыс. (по абсолютной величине);
б) границы, в которых с вероятностью 0,95 заключена средняя сумма вкладов всех вкладчиков;
в) каким должен быть объём выборки, чтобы те же границы для средней суммы вкладов можно было гарантировать с вероятностью 0,99?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 40$ является математическим ожиданием нормально распределённой случайной величины при $H_1:\ a < 40$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 12$ получено выборочное среднее $\overline{x}_в = 44$ и известно генеральное среднее квадратическое отклонение равно $\sigma = 3$.

***Задание* № 4.** При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе $H_1: \sigma_x^2 \ne \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
3,5	1	3,6	1
3,7	3	3,7	5
3,9	5	3,8	2
4,0	3	4,4	1
4,1	1	4,2	1

***Задание* № 5.** Распределение 100 предприятий одного типа по цене на производимые товары *X* (тыс. руб./ед. продукции) и количеству реализуемых товаров *Y* (тыс. руб.), представлено в таблице:

x \ *y*	15-25	25-35	35-45	45-55	55-65	Итого:
4-8	–	–	3	9	7	**19**
8-12	–	–	4	5	9	**18**
12-16	–	10	6	3	–	**19**
16-20	7	12	8	–	–	**27**
20-24	9	8	–	–	–	**17**
Итого:	**16**	**30**	**21**	**17**	**16**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0{,}05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 5

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

21	13	28	19	20	14	24	23	18	15	32	14	15	20	16	17	14	22	21	24
7	22	17	17	26	22	21	21	14	23	24	18	25	18	20	21	20	22	7	31
18	14	22	17	5	20	20	11	17	19	19	3	15	16	19	7	25	13	20	15
16	12	19	16	16	22	21	7	14	21	20	26	17	14	14	14	10	26	12	9
12	11	15	19	13	15	2	6	21	9	23	16	16	21	11	14	19	19	28	12

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,025$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,99$.

***Задание* № 2.** Данные о продолжительности 150 телефонных разговоров городской АТС приведены в таблице:

Продолжительность телефонных разговоров (мин.)	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	Итого
Число разговоров	10	20	30	36	27	15	7	5	150

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,01$ проверить гипотезу о том, что случайная величина X – продолжительность телефонного разговора распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:

а) найти вероятность того, что доля телефонных разговоров, продолжительность которых превышает 9 минут по данным выборки, отличается от доли таких же

разговоров в генеральной совокупности не более чем на 0,05 (по абсолютной величине);
б) границы, в которых с вероятностью 0,99 заключена средняя продолжительность телефонных разговоров городской АТС;
в) каким должен быть объём выборки, чтобы те же границы для средней продолжительности телефонных разговоров можно было гарантировать с вероятностью 0,95?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 58$ является математическим ожиданием нормально распределённой случайной величины при при H_1: $a > 58$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 29$ получено выборочное среднее $\overline{x}_в = 56$, а несмещённое среднее квадратическое отклонение равно $s = 4$.

***Задание* № 4.** При уровне значимости $\alpha = 0,05$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 < \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
9	4	9	2
10	3	10	2
11	3	11	4
12	2	13	1
14	1	14	3

***Задание* № 5.** Распределение 100 страховых договоров по сумме *X* (ден. ед.) и количеству выплат по страховым обязательствам *Y* (единиц), представлено в таблице:

x \ *y*	0-10	10-20	20-30	30-40	40-50	Итого:
1-2	–	–	10	9	8	**27**
2-3	–	–	8	6	10	**24**
3-4	–	8	4	5	–	**17**
4-5	6	11	–	–	–	**17**
5-6	8	7	–	–	–	**15**
Итого:	**14**	**26**	**22**	**20**	**18**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 6

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

43	34	32	33	29	39	40	37	38	38	29	28	35	29	37	34	30	30	37	40
26	33	38	26	30	32	28	44	30	35	20	34	28	34	24	26	34	37	36	39
29	39	31	33	32	28	37	32	37	32	35	39	30	31	33	38	39	20	37	28
34	37	33	25	43	32	36	35	33	28	35	31	25	38	41	37	24	33	42	30
28	34	36	26	33	32	35	43	33	32	33	31	29	33	27	44	33	24	29	41

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0{,}025$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0{,}999$.

***Задание* № 2.** Было отобрано 100 студентов из 500 и получены следующие данные о времени решения задачи по теории вероятностей:

Время (мин.)	5-8	8-11	11-14	14-17	17-20	Итого
Количество студентов	6	18	52	17	7	100

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что случайная величина X – время решения задачи студентом распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:

а) вероятность того, что среднее время решения задачи в выборке отличается от среднего времени решения задачи во всей генеральной совокупности не более чем на 1 минуту (по абсолютной величине);

б) границы, в которых с вероятностью 0,95 заключена доля студентов, решавших задачу не более 11 минут;
в) число студентов, которое нужно отобрать в выборку, чтобы то же отклонение для доли студентов, решавших задачу не более 11 минут, гарантировать с вероятностью 0,999?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 60$ является математическим ожиданием нормально распределённой случайной величины при H_1: $a > 60$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 36$ получено выборочное среднее $\overline{x}_в = 64$ и известно генеральное среднее квадратическое отклонение равно $\sigma = 6$.

***Задание* № 4.** При уровне значимости $\alpha = 0,01$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 > \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
6,1	2	5,8	3
6,5	3	6,0	4
6,6	3	6,2	2
7,0	2	6,3	2
7,4	2	6,8	1

***Задание* № 5.** Распределение 100 объектов основных средств предприятия по первоначальной стоимости объекта *X* (млн. руб.) и по годовой норме амортизационных отчислений *Y* (%), представлено в таблице:

x \ *y*	0-10	10-20	20-30	30-40	40-50	Итого:
5-15	–	–	–	6	4	**10**
15-25	–	–	2	7	6	**15**
25-35	–	–	3	6	5	**14**
35-45	10	18	4	–	–	**32**
45-55	12	13	4	–	–	**29**
Итого:	**22**	**31**	**13**	**19**	**15**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 7

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

78	73	66	74	63	59	62	84	60	71	63	62	76	81	57	64	69	58	55	77
56	75	67	50	66	57	53	71	65	79	70	63	59	68	67	59	62	54	68	75
58	59	65	58	73	59	64	61	68	66	68	57	69	73	69	56	74	63	79	54
75	57	69	71	58	60	62	62	77	73	62	69	66	73	75	71	65	51	55	56
59	67	71	63	73	59	60	59	73	70	66	57	71	63	52	67	58	72	66	83

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,05$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,95$.

***Задание* № 2.** Результаты 5%-ого статистического контроля производственного процесса автомата, производящего иглы, даны в таблице:

Длина, мм	31,0-31,4	31,4-31,8	31,8-32,2	32,2-32,6	32,6-33,0	Итого
Количество игл	25	50	150	200	75	500

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,01$ проверить гипотезу о том, что случайная величина *X* – длина иглы распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:

а) вероятность того, что средняя длина игл во всей совокупности отличается от средней длины игл в выборке не более чем на 0,01 мм (по абсолютной величине);

б) границы, в которых с вероятностью 0,95 заключена доля игл во всей совокупности, длина которых не менее 32,2 мм;
в) при каком объёме выборки те же границы для доли игл можно гарантировать с вероятностью 0,99?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 70$ является математическим ожиданием нормально распределённой случайной величины при при $H_1: \ a \neq 70$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 41$ получено выборочное среднее $\overline{x}_в = 66$, а несмещённое среднее квадратическое отклонение равно $s = 8$.

***Задание* № 4.** При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H_1: $\sigma_x^2 \neq \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
20	1	18	2
22	4	19	3
23	2	20	4
24	2	22	2
26	4	23	1

***Задание* № 5.** Распределение 100 фабрик, выпускающих плёнку для теплиц, по производственным мощностям *X* (тыс. м в год) и себестоимости 1 м плёнки *Y* (руб.) представлено в таблице:

x \ *y*	8-8,5	8,5-9	9-9,5	9,5-10	10-10,5	Итого:
10-20	–	–	–	4	8	**12**
20-30	–	–	7	13	7	**27**
30-40	–	1	6	20	–	**27**
40-50	6	7	8	–	–	**21**
50-60	3	3	7	–	–	**13**
Итого:	**9**	**11**	**28**	**37**	**15**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0{,}05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 8

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

28	31	34	36	21	26	37	35	33	28	37	47	32	37	22	22	26	35	26	47
33	33	34	40	11	36	36	32	26	38	32	44	47	26	49	32	36	35	49	30
51	28	42	30	40	46	28	44	14	27	30	18	28	41	34	35	48	36	43	14
40	22	41	45	27	37	26	42	45	21	22	26	33	36	29	20	34	39	29	37
46	35	43	49	24	44	24	52	23	31	54	60	36	31	38	37	47	22	40	36

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,05$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,99$.

***Задание* № 2.** Получены данные об урожайности гречихи:

Урожайность, ц/га	11-12	12-13	13-14	14-15	15-16	16-17	Всего
Число, га	94	186	230	226	176	88	1000

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,025$ проверить гипотезу о том, что случайная величина X – урожайность гречихи распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:

а) вероятность того, что средняя урожайность на всём массиве отличается от средней выборочной не более чем на 0,08 ц/га (по абсолютной величине);

б) границы, в которых с вероятностью 0,95 заключена доля гектаров с урожайностью не менее 14 ц;

в) каким должен быть объём выборки, чтобы с вероятностью 0,99 гарантировать те же границы для доли гектаров с урожайностью не менее 14 ц/га?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 20$ является математическим ожиданием нормально распределённой случайной величины при H_1: $a \neq 20$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 27$ получено выборочное среднее $\bar{x}_в = 18$ и известно генеральное среднее квадратическое отклонение равно $\sigma = 1$.

***Задание* № 4.** При уровне значимости $\alpha = 0,01$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 < \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
0,2	1	0,4	3
0,4	4	0,5	4
0,8	2	0,9	2
1,0	5	1,2	1
1,2	1	1,4	2

***Задание* № 5.** Распределение 100 сосен по диаметру ствола *X* (см) и высоте *Y* (м), представлено в таблице:

x \ *y*	5-15	15-25	25-35	35-45	45-55	Итого:
15-25	8	2	–	–	–	**10**
25-35	7	16	9	–	–	**32**
35-45	2	8	12	2	–	**21**
45-55	–	6	12	4	–	**22**
55-65	–	2	4	5	1	**12**
Итого:	**17**	**31**	**37**	**11**	**1**	**100**

Необходимо:

1. Вычислить групповые средние $\bar{x}_{y_j}$ и $\bar{y}_{x_i}$ и построить эмпирические линии регрессии.

2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:

а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;

б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;

в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 9

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

32	46	29	7	32	40	36	37	41	44	36	37	49	37	41	23	29	23	25	39
29	45	31	14	29	31	36	53	38	37	35	28	29	44	43	44	20	31	32	32
31	45	42	38	35	29	39	33	40	26	42	43	36	39	36	32	50	37	23	49
26	43	34	30	46	65	41	16	24	27	58	34	47	36	39	24	46	30	8	22
19	26	37	24	38	33	38	22	25	27	44	30	45	28	38	26	24	33	49	40

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.
2. По сгруппированным данным построить:
а) полигон относительных частот;
б) гистограмму относительных частот;
в) график эмпирической функции распределения.
3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.
4. Построить:
а) на чертеже гистограммы её теоретический аналог $f(x)$;
б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;
5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.
6. Проверить выполнение правила «трёх сигм».
7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,05$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.
8. Построить на одном чертеже:
а) полигон относительных частот p_i^* и кривую распределения p_i;
б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.
9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,999$.

***Задание* № 2.** На основе выборочных данных были получены сведения о затратах времени рабочими на изготовление одной детали:

Время изготовления одной детали, мин.	4,0-4,5	4,5-5,0	5,0-5,5	5,5-6,0	6,0-6,5	6,5-7,0	Итого
Число рабочих	5	21	53	57	20	4	160

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,01$ проверить гипотезу о том, что случайная величина X – время изготовления одной детали рабочим распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:
а) вероятность того, что доля рабочих, затрачивающих на изготовление одной детали от 5 до 6,5 минут, отличается от выборочной доли таких рабочих не более чем на 0,0175 (по абсолютной величине);
б) границы, в которых с вероятностью 0,95 заключено среднее время, затрачиваемое на изготовление одной детали всеми рабочими, если их число 2000;
в) каким должен быть объём выборки, чтобы те же границы для среднего времени можно было гарантировать с вероятностью 0,99?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 50$ является математическим ожиданием нормально распределённой случайной величины при H_1: $a < 50$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 19$ получено выборочное среднее $\overline{x}_в = 48$, а несмещённое среднее квадратическое отклонение равно $s = 2$.

***Задание* № 4.** При уровне значимости $\alpha = 0,05$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 > \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
31	1	85	1
33	2	88	2
34	3	95	4
38	5	97	2
42	2	100	3

***Задание* № 5.** Распределение 100 образцов материала по процентному содержанию синтетической добавки *X* (%) и предельному напряжению на разрыв *Y* (Н/м^2) представлено в таблице:

x \ *y*	10-15	15-20	20-25	25-30	30-35	35-40	Итого:
20-30	2	4	–	–	–	–	**6**
30-40	–	6	3	–	–	–	**9**
40-50	–	–	6	35	4	–	**45**
50-60	–	–	2	8	6	–	**16**
60-70	–	–	–	14	7	3	**24**
Итого:	**2**	**10**	**11**	**57**	**17**	**3**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 10

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

22	21	26	20	22	18	28	23	26	29	18	21	24	25	25	21	23	24	20	25
19	25	28	21	23	27	24	19	19	21	25	21	16	19	17	27	23	15	25	27
19	21	24	28	23	24	21	18	28	23	23	26	19	24	18	27	20	29	23	32
23	30	22	21	29	22	25	20	23	23	18	23	21	30	26	32	21	19	18	23
20	28	16	23	18	20	27	25	30	24	25	23	25	29	22	29	26	22	25	22

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,01$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,95$.

***Задание* № 2.** Результаты обследования 500 человек из группы мигрирующего населения по их возрасту приведены в таблице:

Возраст мигрирующего населения, лет	до 30	30-40	40-50	50-60	свыше 60	Итого
Количество мигрантов	40	170	180	80	30	500

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,025$ проверить гипотезу о том, что случайная величина X – возраст мигрирующего населения распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:
а) вероятность того, что доля мигрантов в возрасте 40 лет в выборке отличается от доли их во всей генеральной совокупности не более чем на 0,05 лет (по абсолютной величине);
б) границы, в которых с вероятностью 0,999 заключен средний возраст всего мигрирующего населения, если объём генеральной совокупности велик по сравнению с объёмом выборки;
в) каким должен быть объём выборки, чтобы те же границы среднего возраста можно было гарантировать с вероятностью 0,95?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 30$ является математическим ожиданием нормально распределённой случайной величины при H_1: $a < 30$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 11$ получено выборочное среднее $\overline{x}_в = 34$ и известно генеральное среднее квадратическое отклонение равно $\sigma = 4$.

***Задание* № 4.** При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе $H_1: \sigma_x^2 \neq \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
15	1	20	4
17	3	22	2
20	2	23	2
21	4	25	3
25	3	26	1

***Задание* № 5.** Распределение 100 посреднических предприятий по транспортным издержкам *X* (%) и издержкам при хранении продукции на складе *Y* (%) представлено в таблице:

x \ *y*	0-8	8-16	16-24	24-32	32-40	Итого:
0-10	36	6	–	–	–	**42**
10-20	7	12	8	–	–	**27**
20-30	1	6	7	2	–	**16**
30-40	–	1	4	6	1	**12**
40-50	–	–	1	1	1	**3**
Итого:	**44**	**25**	**20**	**9**	**2**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0{,}05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 11

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

42	31	40	53	46	41	44	45	36	47	27	49	41	36	43	40	44	51	51	47
50	42	42	31	32	41	43	52	40	33	48	31	34	45	43	33	41	49	21	34
35	27	44	28	31	38	37	39	34	29	32	37	42	35	42	44	38	24	45	42
43	38	45	42	43	36	38	43	41	42	42	42	38	33	40	44	38	43	53	42
49	42	45	30	34	41	43	43	33	49	32	40	33	33	39	46	40	36	46	34

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0{,}01$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0{,}99$.

***Задание* № 2.** Данные о продолжительности 100 телефонных разговоров городской АТС приведены в таблице:

Продолжительность телефонных разговоров (мин.)	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	Итого
Число разговоров	10	25	40	50	35	20	15	5	200

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что случайная величина X – продолжительность телефонного разговора распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:

а) найти вероятность того, что доля телефонных разговоров, продолжительность которых превышает 8 минут по данным выборки, отличается от доли таких же разговоров в генеральной совокупности не более чем на 0,05 (по абсолютной величине);

б) границы, в которых с вероятностью 0,95 заключена средняя продолжительность телефонных разговоров городской АТС;

в) каким должен быть объём выборки, чтобы те же границы для средней продолжительности телефонных разговоров можно было гарантировать с вероятностью 0,999?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 40$ является математическим ожиданием нормально распределённой случайной величины при при $H_1: \ a > 40$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 17$ получено выборочное среднее $\overline{x}_в = 36$, а несмещённое среднее квадратическое отклонение равно $s = 1$.

***Задание* № 4.** При уровне значимости $\alpha = 0,05$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H_1: $\sigma_x^2 < \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
27	2	28	5
29	4	29	3
32	5	30	3
33	2	32	1

***Задание* № 5.** Распределение 100 рабочих по стажу работы *X* (лет) и производительности труда *Y* (детал./ч) представлено в таблице:

x \ *y*	5-10	10-15	15-20	20-25	25-30	Итого:
1-3	11	4	–	–	–	**15**
3-5	9	8	3	–	–	**20**
5-7	6	9	9	7	1	**32**
7-9	–	2	9	9	2	**22**
9-11	–	–	4	4	3	**11**
Итого:	**26**	**23**	**25**	**20**	**6**	**100**

Необходимо:

1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.

2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:

а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;

б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;

в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 12

Задание **№ 1.** В результате опыта была получена выборочная совокупность.

48	40	46	53	41	37	58	48	51	51	50	39	61	52	52	48	48	45	43	51
49	54	45	50	55	45	31	44	41	52	51	42	59	54	44	61	51	51	46	46
50	56	40	49	67	42	51	49	55	54	63	72	44	56	59	60	56	50	60	55
49	66	44	56	46	50	42	51	56	53	55	59	50	47	55	50	62	44	62	46
55	58	62	57	59	45	47	52	32	49	70	53	66	52	46	38	48	52	47	49

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,01$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,999$.

Задание **№ 2.** Для определения ситуации, сложившейся в с/х секторе экономики США, Департаментом сельского хозяйства из 2000 фермерских хозяйств штата Аризона было отобрано 110 ферм. Распределение их по размеру объёма совокупных ежегодных продаж (долл.) приведено в таблице:

Размер объёма совокупных ежегодных продаж, долл.	Менее 500	500-1000	1000-1500	1500-2000	2000-2500	Всего
Число ферм	8	24	30	28	20	110

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,025$ проверить гипотезу о том, что случайная величина X – размер объёма совокупных ежегодных продаж распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:

а) вероятность того, что средний объём продаж во всех фермерских хозяйствах отличается от среднего объёма продаж в выборке не более чем на 100$ (по абсолютной величине);

б) границы, в которых с вероятностью 0,97 заключена доля ферм, объём продаж которых не более 1000$;

в) каким должен быть объём выборки, чтобы те же границы для доли ферм, объём продаж которых не более 1000$, можно было гарантировать с вероятностью 0,999?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 90$ является математическим ожиданием нормально распределённой случайной величины при H_1: $a > 90$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 22$ получено выборочное среднее $\bar{x}_в = 88$ и известно генеральное среднее квадратическое отклонение равно $\sigma = 6$.

***Задание* № 4.** При уровне значимости $\alpha = 0,01$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H_1: $\sigma_x^2 > \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
82	2	–10	4
83	2	–9	4
85	3	–6	1
90	3	–3	3

***Задание* № 5.** Распределение 100 спортсменов по росту *X* (см) и по весу *Y* (кг) представлено в таблице:

x \ *y*	60-70	70-80	80-90	90-100	100-110	Итого:
165-170	5	3	2	–	–	**10**
170-175	29	14	5	–	–	**48**
175-180	5	11	2	–	–	**18**
180-185	–	3	8	1	–	**12**
185-190	–	–	–	6	6	**12**
Итого:	**39**	**31**	**17**	**7**	**6**	**100**

Необходимо:

1. Вычислить групповые средние $\bar{x}_{y_j}$ и $\bar{y}_{x_i}$ и построить эмпирические линии регрессии.

2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:

а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;

б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;

в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 13

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

89	94	81	87	83	94	89	92	89	87	88	82	86	84	84	94	93	84	90	83
87	96	89	80	94	89	86	92	92	87	89	86	90	79	81	88	81	93	88	87
89	82	91	81	83	87	91	85	97	96	98	88	83	92	91	88	87	89	90	79
85	95	93	90	89	79	86	83	81	84	89	92	78	95	92	95	86	92	92	95
89	83	97	92	87	84	84	86	87	87	87	88	97	91	88	90	88	94	90	84

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,025$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,95$.

***Задание* № 2.** Из 1000 шахт было отобрано 100. Их распределение по годовой добыче угля приведено в таблице:

Годовая добыча, тыс. т	20-30	30-40	40-50	50-60	60-70	Итого
Число шахт	11	23	31	22	13	100

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что случайная величина X – годовая добыча угля в шахте местного значения распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:
а) вероятность того, что средняя годовая добыча всех шахт отличается от средней годовой добычи шахт в выборке не более чем на 2 тыс. т. (по абсолютной величине);
б) границы, в которых с вероятностью 0,95 заключена доля всех шахт с годовой добычей менее 40 тыс. т.;
в) каким должен быть объём выборки, чтобы те же границы для доли таких шахт можно было гарантировать с вероятностью 0,999?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 86$ является математическим ожиданием нормально распределённой случайной величины при при H_1: $a \neq 86$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 33$ получено выборочное среднее $\overline{x}_в = 84$, а несмещённое среднее квадратическое отклонение равно $s = 5$.

***Задание* № 4.** При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 \neq \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
51	1	15	3
53	3	18	5
55	4	20	2
56	3	23	2
59	1	27	1

***Задание* № 5.** Распределение 100 однотипных предприятий по степени механизации и автоматизации производственных процессов *X* (%) и себестоимости единицы продукции *Y* (руб.) представлено в таблице:

x \ *y*	2-7	7-12	12-17	17-22	22-27	Итого:
10-20	–	–	2	4	6	**12**
20-30	–	–	2	8	6	**16**
30-40	–	8	14	2	–	**24**
40-50	4	14	10	–	–	**28**
50-60	12	8	–	–	–	**20**
Итого:	**16**	**30**	**28**	**14**	**12**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0{,}05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 14

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

22	30	7	17	10	29	21	27	21	17	19	8	16	13	12	31	28	11	23	10
17	34	21	5	30	21	16	27	26	17	21	16	22	36	6	19	7	27	19	17
22	8	24	6	10	18	25	14	35	33	37	19	10	26	24	18	17	22	23	3
14	33	28	23	21	3	15	10	6	12	20	27	1	33	27	31	16	26	27	32
21	11	36	26	18	12	12	15	17	18	17	19	35	25	19	23	18	31	23	12

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,025$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,99$.

***Задание* № 2.** Из 1000 предприятий отрасли были проверены 100 предприятий. Получено следующее распределение предприятий по величине пени за несвоевременную уплату налогов (в млн. руб.):

Величина пени, млн. руб.	3-5	5-7	7-9	9-11	11-13	13-15	15-17	Итого
Число предприятий	4	6	20	40	20	4	6	100

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,01$ проверить гипотезу о том, что случайная величина X – величина пени за несвоевременную уплату налогов предприятий распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:

а) вероятность того, что доля предприятий во всей отрасли, у которых величина пени не менее 13 млн. руб., отличается от выборочной доли таких предприятий не более чем на 0,1 (по абсолютной величине);

б) границы, в которых с вероятностью 0,997 заключена средняя величина пени за несвоевременную уплату налогов предприятий во всей отрасли;

в) каким должен быть объём выборки, чтобы те же границы можно было гарантировать с вероятностью 0,999?

Задание **№ 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 80$ является математическим ожиданием нормально распределённой случайной величины при H_1: $a \neq 80$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 28$ получено выборочное среднее $\overline{x}_в = 78$ и известно генеральное среднее квадратическое отклонение равно $\sigma = 4$.

Задание **№ 4.** При уровне значимости $\alpha = 0,01$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H_1: $\sigma_x^2 < \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
12	1	44	4
15	3	46	2
18	2	47	2
19	2	50	1
23	4	52	3

Задание **№ 5.** Распределение 100 однотипных предприятий по производственным мощностям *X* (тыс. ед. прод./день) и себестоимости единицы продукции *Y* (руб.) представлено в таблице:

x \ *y*	2,5-4,0	4,0-5,5	5,5-7,0	7,0-8,5	8,5-10	Итого:
4-5	–	–	–	7	15	**22**
5-6	–	–	7	8	10	**25**
6-7	–	–	9	10	4	**23**
7-8	3	9	7	–	–	**19**
8-9	5	6	–	–	–	**11**
Итого:	**8**	**15**	**23**	**25**	**29**	**100**

Необходимо:

1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.

2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:

а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;

б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;

в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 15

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

71	76	79	73	78	75	77	72	80	70	69	71	81	67	78	75	76	65	75	77
76	74	75	71	76	79	80	79	72	78	76	70	79	76	78	78	76	79	77	71
68	72	70	84	80	81	76	76	70	80	73	78	74	69	75	75	75	74	77	78
70	75	72	73	86	74	69	76	76	81	61	81	73	76	73	74	70	74	81	71
77	76	79	73	79	77	75	81	73	73	67	64	73	85	81	72	76	71	70	73

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\overline{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить:

а) на чертеже гистограммы её теоретический аналог $f(x)$;

б) на чертеже эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,025$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,999$.

***Задание* № 2.** Из 200 работников фирмы было отобрано 50 человек для получения статистических данных о пребывании работников данной фирмы на больничном листе в течение года. Полученные данные представлены в таблице:

Количество дней, проведённых на больничном листе	1-3	3-5	5-7	7-9	9-11	11-13	Итого
Число работников	2	8	12	20	5	3	50

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что случайная величина X – количество дней на больничном листе распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:
а) вероятность того, что доля работников, пребывавших на больничном листе не менее 9 дней, во всей совокупности, отличается от выборочной доли таких работников не более чем на 0,05 (по абсолютной величине);
б) границы, в которых с вероятностью 0,9973 заключено среднее количество дней пребывания на больничном листе одного работника фирмы;
в) каким должен быть объём выборки, чтобы те же границы для среднего количества дней пребывания на больничном листе можно было гарантировать с вероятностью 0,999?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 60$ является математическим ожиданием нормально распределённой случайной величины при при $H_1: \ a < 60$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 26$ получено выборочное среднее $\overline{x}_в = 66$, а несмещённое среднее квадратическое отклонение равно $s = 5$.

***Задание* № 4.** При уровне значимости $\alpha = 0,05$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 > \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
–8	3	10	4
–5	2	14	2
–3	3	15	3
–1	4	18	2

***Задание* № 5.** Зависимость сопротивления резисторов *X* (Ом) от содержания примесей *Y* (%) представлена в таблице:

x \ *y*	0,5-1,0	1,0-1,5	1,5-2,0	2,0-2,5	2,5-3,0	Итого:
0,40-0,45	2	4	1	–	–	**7**
0,45-0,50	6	11	–	–	–	**17**
0,50-0,55	–	5	25	10	–	**40**
0,55-0,60	–	–	11	10	2	**23**
0,60-0,65	–	–	–	1	12	**13**
Итого:	**8**	**20**	**37**	**21**	**14**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 16

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

7	5	4	5	10	4	13	7	7	4	7	3	7	6	5	2	8	7	7	8
7	7	12	1	7	4	8	8	9	8	14	10	8	7	7	6	7	5	14	4
8	10	7	7	8	1	6	5	5	4	3	6	5	6	7	5	4	6	4	6
4	6	11	7	5	4	5	8	10	10	1	6	5	10	13	7	6	7	8	3
11	4	7	10	9	6	3	3	10	6	10	4	11	5	6	10	8	4	13	6

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить на чертеже:

а) гистограммы её теоретический аналог $f(x)$;

б) эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,05$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,95$.

***Задание* № 2.** Было отобрано 100 из 1000 предприятий города. Получено следующее распределение предприятий по суммам уплаченного штрафа за неуплату налогов (ден. ед.):

Сумма штрафов, ден. ед.	0-500	500-1000	1000-1500	1500-2000	2000-2500	2500-3000	Всего
Число предприятий	6	18	28	35	9	4	100

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,01$ проверить гипотезу о том, что случайная величина X – сумма штрафа предприятий распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:
а) вероятность того, что средняя сумма штрафа всех предприятий отличается от средней суммы штрафа предприятий в выборке не более чем на 75 ден. ед. (по абсолютной величине);
б) границы, в которых с вероятностью 0,9545 заключена доля предприятий, заплативших штраф не более 1000 ден. ед.;
в) каким должен быть объём выборки, чтобы те же границы для доли предприятий, заплативших штраф не более 1000 ден. ед., можно было гарантировать с вероятностью 0,999?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 100$ является математическим ожиданием нормально распределённой случайной величины при $H_1: \ a < 100$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 30$ получено выборочное среднее $\bar{x}_в = 96$ и известно генеральное среднее квадратическое отклонение равно $\sigma = 6$.

***Задание* № 4.** При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе $H_1: \sigma_x^2 \neq \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
42	2	84	4
45	4	87	2
46	5	92	5
50	1	96	1

***Задание* № 5.** Распределение 100 изделий по стоимости готового изделия *X* (тыс. руб.) и стоимости сырья *Y* (тыс. руб.) представлено в таблице:

x \ *y*	0,5-5,5	5,5-10,5	10,5-15,5	15,5-20,5	20,5-25,5	Итого:
5-15	7	11	–	–	–	**18**
15-25	–	5	19	3	–	**27**
25-35	–	–	15	15	2	**32**
35-45	–	–	5	6	4	**15**
45-55	–	–	–	1	7	**8**
Итого:	**7**	**16**	**39**	**25**	**13**	**100**

Необходимо:
1. Вычислить групповые средние $\bar{x}_{y_j}$ и $\bar{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0{,}05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 17

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

67	62	63	63	72	68	55	61	60	62	61	61	66	71	62	64	68	76	56	62
64	64	52	60	64	60	57	75	57	63	59	60	57	65	60	71	62	57	62	62
66	61	64	68	63	67	52	71	70	57	53	61	61	74	62	56	55	65	57	62
54	55	64	70	62	59	64	66	68	55	58	62	68	75	61	63	64	65	64	58
65	62	72	53	67	61	62	60	58	60	50	63	62	68	69	58	58	52	56	57

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить на чертеже:

а) гистограммы её теоретический аналог $f(x)$;

б) эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,05$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,99$.

***Задание* № 2.** Для контролирования отклонения диаметра деталей у 15000 изготовленных на токарном станке с ЧПУ валах ротора электродвигателя было проверено 150 валов. В результате получено следующее распределение положительных отклонений размера диаметра вала (в микронах) от номинального размера:

Отклонение, мкн.	24-28	28-32	32-36	36-40	40-44	44-48	48-52	Всего
Число валов	1	13	27	41	32	23	13	150

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,025$ проверить гипотезу о том, что случайная величина X – отклонение диаметра вала распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:
а) вероятность того, что доля отклонение диаметра вала от номинального, не превышающего 36 мкн в генеральной совокупности, отличается от выборочной доли таких отклонений не более чем на 0,04 (по абсолютной величине);
б) границы, в которых с вероятностью 0,9596 заключено среднее отклонение размера диаметра вала от номинального размера в генеральной совокупности;
в) каким должен быть объём выборки, чтобы те же границы для среднего отклонения размера диаметра вала можно было гарантировать с вероятностью 0,999?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 80$ является математическим ожиданием нормально распределённой случайной величины при при H_1: $a > 80$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 24$ получено выборочное среднее $\overline{x}_в = 78$, а несмещённое среднее квадратическое отклонение равно $s = 4$.

***Задание* № 4.** При уровне значимости $\alpha = 0,05$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 < \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
30	1	30	1
32	5	31	4
33	4	32	5
34	2	34	2
36	1	35	1

***Задание* № 5.** Распределение 100 рабочих завода по стажу работы *X* (лет) и затратам времени на обработку одной детали *Y* (мин.) представлено в таблице:

x \ *y*	17-19	19-21	21-23	23-25	25-27	Итого:
0-5	–	–	–	3	5	**8**
5-10	–	–	2	10	3	**15**
10-15	2	6	4	5	–	**17**
15-20	15	8	7	–	–	**30**
20-25	10	11	9	–	–	**30**
Итого:	**27**	**25**	**22**	**18**	**8**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 18

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

25	32	42	40	35	34	24	30	39	19	16	23	31	23	31	22	22	19	23	32
46	36	34	20	30	29	39	36	24	36	31	33	22	44	17	28	27	44	39	38
9	43	28	34	52	38	32	34	19	24	34	26	38	36	33	51	24	6	25	21
32	13	34	41	17	22	35	36	44	12	40	14	42	37	42	29	24	22	25	34
27	41	31	36	29	40	30	29	46	11	46	21	24	42	33	41	22	26	39	38

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.
2. По сгруппированным данным построить:
а) полигон относительных частот;
б) гистограмму относительных частот;
в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.
4. Построить на чертеже:
а) гистограммы её теоретический аналог $f(x)$;
б) эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.
6. Проверить выполнение правила «трёх сигм».
7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,05$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.
8. Построить на одном чертеже:
а) полигон относительных частот p_i^* и кривую распределения p_i;
б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,999$.

***Задание* № 2.** Было отобрано 100 деталей из 1500 и получены следующие данные об их весе:

Вес, г	42-46	46-50	50-54	54-58	58-62	62-66	Всего
Число деталей	10	14	26	28	12	10	100

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,01$ проверить гипотезу о том, что случайная величина *X* – вес деталей распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.
2. Найти:
а) каким должен быть объём выборки, чтобы с вероятностью 0,9545 доля деталей, не менее 58 г, во всей совокупности отличалась от выборочной доли таких деталей не более чем на 0,05 (по абсолютной величине)?

б) границы, в которых с вероятностью 0,95 заключен средний вес деталей во всей совокупности;

в) каким должен быть объём выборки, чтобы те же границы для среднего веса деталей можно было гарантировать с вероятностью 0,999?

Задание **№ 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 80$ является математическим ожиданием нормально распределённой случайной величины при H_1: $a > 80$ на 5% уровне значимости, если в результате обработки выборки объёма $n = 19$ получено выборочное среднее $\overline{x}_в = 84$ и известно генеральное среднее квадратическое отклонение равно $\sigma = 3$.

Задание **№ 4.** При уровне значимости $\alpha = 0,01$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 > \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
42	2	44	6
44	1	45	2
48	3	46	1
50	4	51	2
53	2	55	2

Задание **№ 5.** Распределение 100 рабочих по стажу работы *X* (лет) и по выполнению нормы *Y* (%) представлено в таблице:

x \ *y*	80-90	90-100	100-110	110-120	120-130	Итого:
1-4	11	4	–	–	–	**15**
4-7	9	7	3	–	1	**20**
7-10	6	9	9	7	1	**32**
10-13	–	2	10	8	2	**22**
13-16	–	–	3	5	3	**11**
Итого:	**26**	**22**	**25**	**20**	**7**	**100**

Необходимо:

1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.

2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:

а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;

б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;

в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 19

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

22	27	38	13	16	32	31	19	28	32	35	28	30	23	30	31	32	31	28	23
23	15	33	19	31	24	27	27	26	28	20	19	22	20	25	22	21	29	27	20
20	23	20	29	16	35	30	20	22	28	23	25	25	19	22	28	25	21	19	29
30	20	28	32	20	20	20	27	25	27	30	32	16	18	26	20	31	22	20	14
20	28	24	22	23	22	25	19	26	23	25	24	32	23	21	26	22	20	23	19

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить на чертеже:

а) гистограммы её теоретический аналог $f(x)$;

б) эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,01$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,95$.

***Задание* № 2.** Результаты выборочного 10% обследования сверхмарафонцев 100-километрового пробега на чемпионате мира в Токио в 1994 г. по скорости спортсменов представлены в таблице:

Скорость, км/ч	0-10	10-11	11-12	12-13	13-14	14-15	15-16	более 16	Итого
Число спортсменов	4	8	12	26	24	16	9	1	100

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что случайная величина X – скорость спортсмена распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:
а) вероятность того, что средняя скорость всех спортсменов отличается от средней выборочной не более чем на 0,25 км/час (по абсолютной величине);
б) границы, в которых с вероятностью 0,9973 заключена доля спортсменов, скорость которых не менее 14 км/час;
в) каким должен быть объём выборки, чтобы те же границы для доли спортсменов, скорость которых не менее 14 км/час, можно было гарантировать с вероятностью 0,999?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 50$ является математическим ожиданием нормально распределённой случайной величины при при $H_1: \ a \neq 50$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 32$ получено выборочное среднее $\overline{x}_в = 48$, а несмещённое среднее квадратическое отклонение равно $s = 2$.

***Задание* № 4.** При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе H$_1$: $\sigma_x^2 \neq \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
31	1	29	4
35	4	32	3
40	2	33	2
42	3	35	3
44	2	39	1

***Задание* № 5.** Проведена проверка работы 100 машинисток, печатающих сложные тексты. В таблице дана зависимость числа опечаток *X* (шт.) от стажа работы *Y* (мес.) на единицу печатной продукции:

x \ *y*	0-18	18-36	36-54	54-72	72-90	Итого:
0-2	–	–	–	–	4	**4**
2-4	–	–	–	4	6	**10**
4-6	–	–	6	10	8	**24**
6-8	3	4	10	20	–	**37**
8-10	6	7	12	–	–	**25**
Итого:	**9**	**11**	**28**	**34**	**18**	**100**

Необходимо:
1. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$ и построить эмпирические линии регрессии.
2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:
а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0{,}05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;
б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;
в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

ВАРИАНТ № 20

***Задание* № 1.** В результате опыта была получена выборочная совокупность.

73	69	84	68	71	59	87	74	70	62	78	86	82	60	97	68	72	72	74	80
78	64	75	63	78	53	67	61	59	79	63	59	62	55	65	76	60	77	67	70
66	61	78	61	86	70	64	67	87	92	73	70	69	65	73	85	61	75	69	73
68	69	81	88	69	64	44	74	80	80	74	73	58	51	81	61	75	72	79	77
43	75	76	96	64	64	83	85	76	63	58	72	68	65	79	71	90	70	74	77

1. По данной таблице составить интервальный вариационный ряд, разбив всю вариацию на 8-10 интервалов.

2. По сгруппированным данным построить:

а) полигон относительных частот;

б) гистограмму относительных частот;

в) график эмпирической функции распределения.

3. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

4. Построить на чертеже:

а) гистограммы её теоретический аналог $f(x)$;

б) эмпирической функции $F^*(x)$ её теоретический аналог $F(x)$;

5. По виду гистограммы и эмпирической функции распределения выборки выдвинуть гипотезу о распределении генеральной совокупности.

6. Проверить выполнение правила «трёх сигм».

7. Применив критерий согласия Пирсона χ^2 с заданным уровнем значимости $\alpha = 0,01$, окончательно принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

8. Построить на одном чертеже:

а) полигон относительных частот p_i^* и кривую распределения p_i;

б) гистограмму теоретических вероятностей (относительных частот) p_i и $f(x)$. Сравнить график p_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

9. Найти доверительные интервалы для генеральной средней и генерального среднего квадратического отклонения по уровню надёжности $\gamma = 0,99$.

***Задание* № 2.** Выборочное обследование качества нити на крепость дало следующие результаты:

Крепость нити, г	0-47,5	47,5-52,5	52,5-57,5	57,5-62,5	62,5-67,5	67,5-72,5	Итого
Число нитей	2	7	20	45	21	5	100

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0,025$ проверить гипотезу о том, что случайная величина X – крепость нити распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

2. Найти:

а) вероятность того, что доля нитей с крепостью не менее 62,5 во всей партии отличается от выборочной доли не более чем на 0,05 (по абсолютной величине);

б) границы, в которых с вероятностью 0,8904 заключена средняя крепость нити во всей партии, если количество нити очень велико;

в) каким должен быть объём выборки, чтобы те же границы для средней крепости нити можно было гарантировать с вероятностью 0,999?

***Задание* № 3.** Проверить нулевую гипотезу о том, что заданное значение $a_0 = 60$ является математическим ожиданием нормально распределённой случайной величины при $H_1:\ a \neq 60$ на 10% уровне значимости, если в результате обработки выборки объёма $n = 17$ получено выборочное среднее $\bar{x}_в = 54$ и известно генеральное среднее квадратическое отклонение равно $\sigma = 2$.

***Задание* № 4.** При уровне значимости $\alpha = 0,01$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* (при альтернативной гипотезе $H_1: \sigma_x^2 \neq \sigma_y^2$) на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
61	1	60	4
62	4	63	3
64	2	64	2
67	4	68	1
68	2	70	3

***Задание* № 5.** Распределение 100 предприятий химической промышленности по фондовооружённости *X* (млн руб.) и энерговооружённости труда *Y* (кВт·ч) представлено в таблице:

x \ *y*	0-1,5	1,5-3,0	3,0-4,5	4,5-6,0	6,0-7,5	Итого:
0-5	4	10	7			**21**
5-10	6	11	8			**25**
10-15			9	15	8	**32**
15-20			4	5	10	**19**
20-25				2	1	**3**
Итого:	**10**	**21**	**28**	**22**	**19**	**100**

Необходимо:

1. Вычислить групповые средние $\bar{x}_{y_j}$ и $\bar{y}_{x_i}$ и построить эмпирические линии регрессии.

2. Предполагая, что между переменными *X* и *Y* существует линейная корреляционная зависимость:

а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0,05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;

б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;

в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

Порядок отчётности. Вопросы к защите семестровой работы

Защита типового расчёта (семестровой работы) проводится после проверки преподавателем выполненных заданий.

Во время защиты студент должен:

- ✓ **уметь объяснить выполнение любого из заданий 1-5,**
- ✓ **правильно отвечать на следующие вопросы.**

1. Чем занимается математическая статистика? Каковы её основные задачи?

2. Что называется генеральной совокупностью, выборочной совокупностью?

3. Что называется статистическим законом распределения случайной величины?

4. Что называется эмпирической функцией распределения, в чём состоит различие между эмпирической и теоретической (интегральной) функцией распределения?

5. Дайте понятие точечной и интервальной оценок неизвестных параметров распределения.

6. Какая оценка параметра называется состоятельной, несмещённой, эффективной? Почему желательно, чтобы оценка была состоятельной, несмещённой?

7. В чём заключается сущность метода моментов получения точечной оценки параметров?

8. В чём заключается сущность метода наибольшего правдоподобия получения точечной оценки параметров?

9. Что называется доверительным интервалом, доверительной вероятностью?

10. Как строится доверительный интервал для математического ожидания случайной величины, распределённой по нормальному закону?

11. Как строится доверительный интервал для среднего квадратического отклонения случайной величины, распределённой по нормальному закону?

12. Дайте определение статистической гипотезы.

13. Что называется параметрической статистической гипотезой, непараметрической статистической гипотезой?

14. Приведите примеры нулевой, альтернативной, простой и сложной гипотез. Объясните, в чём состоит принцип проверки нулевых гипотез с помощью статистических критериев значимости.

15. Что называется ошибкой первого рода, ошибкой второго рода, уровнем значимости?

16. Как находятся критические точки (квантили) статистических критериев значимости в случае двусторонней критической области, в случае левосторонней критической области, в случае правосторонней критической области?

17. Что называется критерием согласия?

18. Являются ли критерии согласия статистическими критериями значимости?

19. Дайте общую схему проверки гипотезы о виде функции распределения с помощью критерия согласия Пирсона.

20. Укажите достоинства и недостатки критерия согласия Пирсона.

21. На основании каких признаков или критериев можно сделать предварительный выбор закона распределения?

22. Дайте понятие корреляционной и статистической зависимости между значениями случайных величин X и Y.

23. Что можно оценить с помощью корреляционного отношения.

24. Какие виды корреляционной зависимости чаще всего используются на практике? Каким методом может быть получено уравнение регрессии?

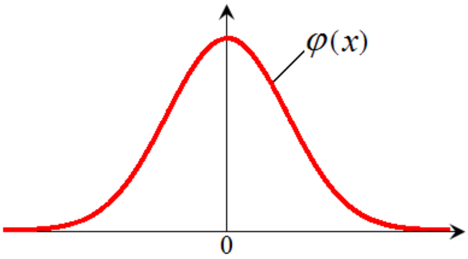
25. Что характеризует выборочный коэффициент корреляции?

ПРИЛОЖЕНИЯ

Приложение 1

Таблица значений функции Гаусса

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



	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551

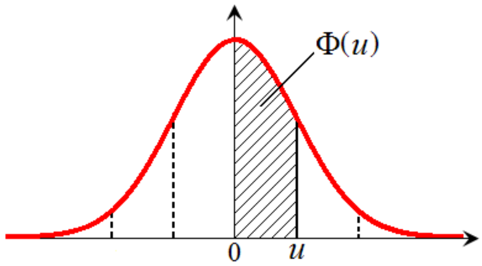
Продолжение прилож. 1

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
4,5	0,0000160									
5,0	0,0000015									

Приложение 2

Таблица значений функции Лапласа

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

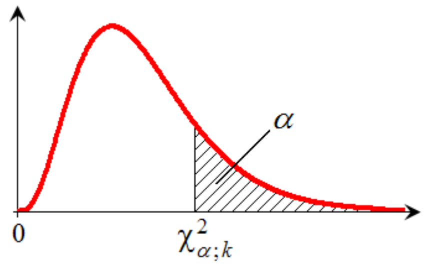


x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	**0,26**	0,1026	**0,52**	0,1985	**0,78**	0,2823
0,01	0,0040	**0,27**	0,1064	**0,53**	0,2019	**0,79**	0,2852
0,02	0,0080	**0,28**	0,1103	**0,54**	0,2054	**0,80**	0,2881
0,03	0,0120	**0,29**	0,1141	**0,55**	0,2088	**0,81**	0,2910
0,04	0,0160	**0,30**	0,1179	**0,56**	0,2123	**0,82**	0,2939
0,05	0,0199	**0,31**	0,1217	**0,57**	0,2157	**0,83**	0,2967
0,06	0,0239	**0,32**	0,1255	**0,58**	0,2190	**0,84**	0,2995
0,07	0,0279	**0,33**	0,1293	**0,59**	0,2224	**0,85**	0,3023
0,08	0,0319	**0,34**	0,1331	**0,60**	0,2257	**0,86**	0,3051
0,09	0,0359	**0,35**	0,1368	**0,61**	0,2291	**0,87**	0,3078
0,10	0,0398	**0,36**	0,1406	**0,62**	0,2324	**0,88**	0,3106
0,11	0,0438	**0,37**	0,1443	**0,63**	0,2357	**0,89**	0,3133
0,12	0,0478	**0,38**	0,1480	**0,64**	0,2389	**0,90**	0,3159
0,13	0,0517	**0,39**	0,1517	**0,65**	0,2422	**0,91**	0,3186
0,14	0,0557	**0,40**	0,1554	**0,66**	0,2454	**0,92**	0,3212
0,15	0,0596	**0,41**	0,1591	**0,67**	0,2486	**0,93**	0,3238
0,16	0,0636	**0,42**	0,1628	**0,68**	0,2517	**0,94**	0,3264
0,17	0,0675	**0,43**	0 1664	**0,69**	0,2549	**0,95**	0,3289
0,18	0,0714	**0,44**	0,1700	**0,70**	0,2580	**0,96**	0,3315
0,19	0,0753	**0,45**	0,1736	**0,71**	0,2611	**0,97**	0,3340
0,20	0,0793	**0,46**	0,1772	**0,72**	0,2642	**0,98**	0,3365
0,21	0,0832	**0,47**	0,1808	**0,73**	0,2673	**0,99**	0,3389
0,22	0,0871	**0,48**	0,1844	**0,74**	0,2703	**1,00**	0,3413
0,23	0,0910	**0,49**	0,1879	**0,75**	0,2734	**1,01**	0,3438
0,24	0,0948	**0,50**	0,1915	**0,76**	0,2764	**1,02**	0,3461
0,25	0,0987	**0,51**	0,1950	**0,77**	0,2794	**1,03**	0,3485

Продолжение прилож. 2

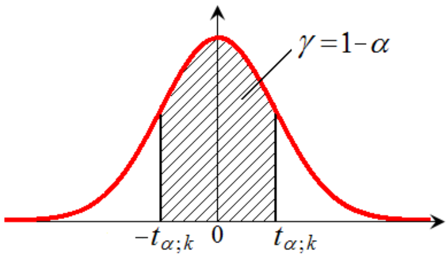
x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,04	0,3508	**1,43**	0,4236	**1,82**	0,4656	**2,42**	0,4922
1,05	0,3531	**1,44**	0,4251	**1,83**	0,4664	**2,44**	0,4927
1,06	0,3554	**1,45**	0,4265	**1,84**	0,4671	**2,46**	0,4931
1,07	0,3577	**1,46**	0,4279	**1,85**	0,4678	**2,48**	0,4934
1,08	0,3599	**1,47**	0,4292	**1,86**	0,4686	**2,50**	0,4938
1,09	0,3621	**1,48**	0,4306	**1,87**	0,4693	**2,52**	0,4941
1,10	0,3643	**1,49**	0,4319	**1,88**	0,4699	**2,54**	0,4945
1,11	0,3665	**1,50**	0,4332	**1,89**	0,4706	**2,56**	0,4948
1,12	0,3686	**1,51**	0,4345	**1,90**	0,4713	**2,58**	0,4951
1,13	0,3708	**1,52**	0,4357	**1,91**	0,4719	**2,60**	0,4953
1,14	0,3729	**1,53**	0,4370	**1,92**	0,4726	**2,62**	0,4956
1,15	0,3749	**1,54**	0,4382	**1,93**	0,4732	**2,64**	0,4959
1,16	0,3770	**1,55**	0,4394	**1,94**	0,4738	**2,66**	0,4961
1,17	0,3790	**1,56**	0,4406	**1,95**	0,4744	**2,68**	0,4963
1,18	0,3810	**1,57**	0,4418	**1,96**	0,4750	**2,70**	0,4965
1,19	0,3830	**1,58**	0,4429	**1,97**	0,4756	**2,72**	0,4967
1,20	0,3849	**1,59**	0,4441	**1,98**	0,4761	**2,74**	0,4969
1,21	0,3869	**1,60**	0,4452	**1,99**	0,4767	**2,76**	0,4971
1,22	0,3883	**1,61**	0,4463	**2,00**	0,4772	**2,78**	0,4973
1,23	0,3907	**1,62**	0,4474	**2,02**	0,4783	**2,80**	0,4974
1,24	0,3925	**1,63**	0,4484	**2,04**	0,4793	**2,82**	0,4976
1,25	0,3944	**1,64**	0,4495	**2,06**	0,4803	**2,84**	0,4977
1,26	0,3962	**1,65**	0,4505	**2,08**	0,4812	**2,86**	0,4979
1,27	0,3980	**1,66**	0,4515	**2,10**	0,4821	**2,88**	0,4980
1,28	0,3997	**1,67**	0,4525	**2,12**	0,4830	**2,90**	0,4981
1,29	0,4015	**1,68**	0,4535	**2,14**	0,4838	**2,92**	0,4982
1,30	0,4032	**1,69**	0,4545	**2,16**	0,4846	**2,94**	6,4984
1,31	0,4049	**1,70**	0,4554	**2,18**	0,4854	**2,96**	0,4985
1,32	0,4066	**1,71**	0,4564	**2,20**	0,4861	**2,98**	0,4986
1,33	0,4082	**1,72**	0,4573	**2,22**	0,4868	**3,00**	0,49865
1,34	0,4099	**1,73**	0,4582	**2,24**	0,4875	**3,20**	0,49931
1,35	0,4115	**1,74**	0,4591	**2,26**	0,4881	**3,40**	0,49966
1,36	0,4131	**1,75**	0,4599	**2,28**	0,4887	**3,60**	0,499841
1,37	0,4147	**1,76**	0,4608	**2,30**	0,4893	**3,80**	0,499928
1,38	0,4162	**1,77**	0,4616	**2,32**	0,4898	**4,00**	0,499968
1,39	0,4177	**1,78**	0,4625	**2,34**	0,4904	**4,50**	0,499997
1,40	0,4192	**1,79**	0,4633	**2,36**	0,4909	**5,00**	0,5
1,41	0,4207	**1,80**	0,4641	**2,38**	0,4913	**> 5**	0,5
1,42	0,4222	**1,81**	0,4649	**2,40**	0,4918	**∞**	0,5

Критические точки распределения χ^2

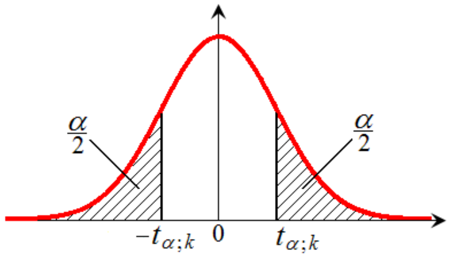


Число степеней свободы k	Уровень значимости α					
	0,01	**0,025**	**0,05**	**0,95**	**0,975**	**0,89**
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Критические точки $t_{\alpha;k}$ распределения Стьюдента



Двусторонняя критическая область



Односторонние критические области

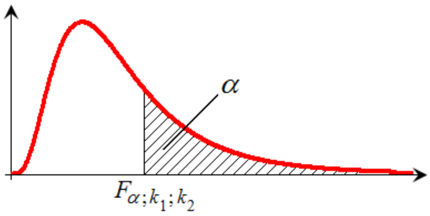
Число степеней свободы k	**Уровень значимости α (двусторонняя критическая область)**						
	0,2	**0,1**	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
1	3,08	6,31	12,7	31,82	63,7	318,3	637,0
2	1,89	2,92	4,30	6,97	9,92	22,33	31,6
3	1,64	2,35	3,18	4,54	5,84	10,22	12,9
4	1,53	2,13	2,78	3,75	4,60	7,17	8,61
5	1,48	2,01	2,57	3,37	4,03	5,89	6,86
6	1,44	1,94	2,45	3,14	3,71	5,21	5,96
7	1,42	1,89	2,36	3,00	3,50	4,79	5,40
8	1,40	1,86	2,31	2,90	3,36	4,50	5,04
9	1,38	1,83	2,26	2,82	3,25	4,30	4,78
10	1,37	1,81	2,23	2,76	3,17	4,14	4,59
11	1,36	1,80	2,20	2,72	3,11	4,03	4,44
12	1,36	1,78	2,18	2,68	3,05	3,93	4,32
13	1,35	1,77	2,16	2,65	3,01	3,85	4,22
14	1,34	1,76	2,14	2,62	2,98	3,79	4,14
15	1,34	1,75	2,13	2,60	2,95	3,73	4,07
16	1,34	1,75	2,12	2,58	2,92	3,69	4,01
17	1,33	1,74	2,11	2,57	2,90	3,65	3,96
18	1,33	1,73	2,10	2,55	2,88	3,61	3,92
19	1,33	1,73	2,09	2,54	2,86	3,58	3,88
20	1,33	1,73	2,09	2,53	2,85	3,55	3,85
21	1,32	1,72	2,08	2,52	2,83	3,53	3,82
22	1,32	1,72	2,07	2,51	2,82	3,51	3,79
23	1,32	1,71	2,07	2,50	2,81	3,49	3,77
24	1,32	1,71	2,06	2,49	2,80	3,47	3,74
25	1,32	1,71	2,06	2,49	2,79	3,45	3,72
26	1,32	1,71	2,06	2,48	2,78	3,44	3,71
27	1,31	1,71	2,05	2,47	2,77	3,42	3,69
28	1,31	1,70	2,05	2,46	2,76	3,40	3,66
29	1,31	1,70	2,05	2,46	2,76	3,40	3,66
30	1,31	1,70	2,04	2,46	2,75	3,39	3,65
40	1,30	1,68	2,02	2,42	2,70	3,31	3,55
60	1,30	1,67	2,00	2,39	2,66	3,23	3,46
120	1,29	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,28	1,64	1,96	2,33	2,58	3,09	3,29
Число степеней свободы k	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
	Уровень значимости α (односторонняя критическая область)						

Приложение 5

Критические точки $F_{\alpha;k_1;k_2}$ распределения Фишера-Снедекора

(k_1 – число степеней свободы бóльшей дисперсии,

k_2 – число степеней свободы мéньшей дисперсии)



Уровень значимости $\alpha = 0,01$

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Уровень значимости $\alpha = 0,05$

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Приложение 6

Таблица значений $t_\gamma = t(\gamma, n)$

n \ γ	0,95	0,99	0,999	n \ γ	0,95	0,99	0,999
5	2,78	4,60	8,61	**20**	2,093	2,861	3,883
6	2,57	4,03	6,86	**25**	2,064	2,797	3,745
7	2,45	3,71	5,96	**30**	2,045	2,756	3,659
8	2,37	3,50	5,41	**35**	2,032	2,720	3,600
9	2,31	3,36	5,04	**40**	2,023	2,708	3,558
10	2,26	3,25	4,78	**45**	2,016	2,692	3,527
11	2,23	3,17	4,59	**50**	2,009	2,679	3,502
12	2,20	3,11	4,44	**60**	2,001	2,662	3,464
13	2,18	3,06	4,32	**70**	1,996	2,649	3,439
14	2,16	3,01	4,22	**80**	1,001	2,640	3,418
15	2,15	2,98	4,14	**90**	1,987	2,633	3,403
16	2,13	2,95	4,07	**100**	1,984	2,627	3,392
17	2,12	2,92	4,02	**120**	1,980	2,617	3,374
18	2,11	2,90	3,97	**∞**	1,960	2,576	3,291
19	2,10	2,88	3,92				

Приложение 7

Таблица значений $q = q(\gamma, n)$

n \ γ	0,95	0,99	0,999	n \ γ	0,95	0,99	0,999
5	1,37	2,67	5,64	**20**	0,37	0,58	0,88
6	1,09	2,01	3,88	**25**	0,32	0,49	0,73
7	0,92	1,62	2,98	**30**	0,28	0,43	0,63
8	0,80	1,38	2,42	**35**	0,26	0,38	0,56
9	0,71	1,20	2,06	**40**	0,24	0,35	0,50
10	0,65	1,08	1,80	**45**	0,22	0,32	0,46
11	0,59	0,98	1,60	**50**	0,21	0,30	0,43
12	0,55	0,90	1,45	**60**	0,188	0,269	0,38
13	0,52	0,83	1,33	**70**	0,174	0,245	0,34
14	0,48	0,78	1,23	**80**	0,161	0,226	0,31
15	0,46	0,73	1,15	**90**	0,151	0,211	0,29
16	0,44	0,70	1,07	**100**	0,143	0,198	0,27
17	0,42	0,66	1,01	**150**	0,115	0,160	0,211
18	0,40	0,63	0,96	**200**	0,099	0,136	0,185
19	0,39	0,60	0,92	**250**	0,089	0,120	0,162

Библиографический список

1. *Баврин И. И., Матросов В. Л.* Общий курс высшей математики. – М.: Просвещение, 1995.
2. *Бронштейн И. Н., Семендяев К. А.* Справочник по математике для инженеров и учащихся втузов. – М.: Наука, 1981.
3. *Гмурман В. Е.* Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 1998.
4. *Гмурман В. Е.* Теория вероятностей и математическая статистика. – М.: Высшая школа, 1998.
5. *Гурский Е. И.* Сборник задач по теории вероятностей и математической статистике. – М.: Высшая школа, 1984.
6. *Красс М. С.* Математика в экономике. Математические методы и модели. – М.: Финансы и статистика, 2007.
7. *Кремер Н. Ш.* Теория вероятностей и математическая статистика. – М.: ЮНИТИ-ДАНА, 2007.
8. Математика. Большой энциклопедический словарь / Под редакцией Ю. В. Прохорова. – М.: Большая Российская энциклопедия, 1998.
9. *Письменный Д. Т.* Конспект лекций по теории вероятностей и математической статистике. – М.: Айрис-пресс, 2004.
10. Сборник задач по математике для ВТУЗов. Теория вероятностей и математическая статистика / Под редакцией А. В. Ефимова. – М.: Наука, 1990.
11. *Фадеева Л. Н., Жуков Ю. В., Лебедев А. В.* Математика для экономистов: теория вероятностей и математическая статистика. Задачи и упражнения. – М.: Эксмо, 2006.
12. *Шикин Е. В., Чхартишвили А. Г.* Математические методы и модели в управлении. – М.: Дело, 2000.

Учебное пособие

Джамиля Калимулловна **Агишева**
Светлана Александровна **Зотова**
Татьяна Александровна **Матвеева**
Виктория Борисовна **Светличная**

Математическая статистика

Редактор Е. М. Марносова
Темплан 2010, поз. № 5 В

Подписано в печать 15.02.2010 г. Формат 60×84 1/16. Бумага офсетная.
Гарнитура Times. Печать офсетная. Усл. печ. л. 9,06.
Тираж 150. экз. Заказ .

Волгоградский государственный технический университет
400131 Волгоград, пр. Ленина, 28, корп. 1.

Отпечатано в типографии ВолгГТУ
400131, г. Волгоград, пр. Ленина, 28, корп. 7.