Краснодарский университет МВД России

И. Н. Старостенко
А. А. Хромых

ЭЛЕМЕНТЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ

Учебное пособие

Краснодар
2020

УДК 519.2
ББК 22.172
С77

Одобрено
редакционно-издательским советом
Краснодарского университета
МВД России

Рецензенты:

В. Ю. Иванов, кандидат технических наук, доцент (Московский университет МВД России им. В.Я. Кикотя);

А. Г. Карпика, кандидат технических наук, доцент (Ростовский юридический институт МВД России).

Старостенко И. Н.

С77 Элементы статистической обработки данных : учебное пособие / И. Н. Старостенко, А. А. Хромых. – Краснодар : Краснодарский университет МВД России, 2020. – 74 с.

ISBN 978-5-9266-1613-9

Рассматриваются основы математической статистики, статистические показатели и средние статистические оценки параметров распределения, статистическая проверка гипотез.

Для профессорско-преподавательского состава, курсантов и слушателей образовательных организаций МВД России.

УДК 519.2
ББК 22.172

ISBN 978-5-9266-1613-9

Содержание

Введение

Математическая статистика – это раздел прикладной математики, в котором изучаются методы сбора, систематизации и обработки результатов наблюдений массовых случайных явлений для выявления существующих закономерностей.

Предметом математической статистики является изучение случайных величин по результатам наблюдений.

Основными задачами математической статистики являются:

1) представление полученных данных в удобном для дальнейшего анализа виде;

2) оценка интересующих характеристик случайной величины;

3) проверка статистических гипотез, т.е. решение вопроса согласования результатов оценивания с опытными данными.

Анализ статистических данных включает оценку вероятностей события, функции распределения вероятностей или плотности вероятностей, оценку параметров известного распределения, оценку связей между случайными величинами.

Математическая статистика опирается на теорию вероятностей и в свою очередь служит основой для разработки методов обработки и анализа статистических результатов в конкретных областях человеческой деятельности.

Условные обозначения:

▼ – условие теоремы;

○ – начало доказательства теоремы;

● – конец доказательства теоремы;

▲ – условие примера;

□ – начало решения примера;

■ – конец решения примера;

! – замечание.

1. Основы математической статистики

1.1. Генеральная и выборочная совокупности

Введем основные понятия математической статистики.

Совокупность всех подлежащих изучению объектов или возможных результатов всех мыслимых наблюдений, производимых в неизменных условиях над одним объектом, называется *генеральной совокупностью*.

Генеральная совокупность может быть *конечной* или *бесконечной* в зависимости от того, конечна или бесконечна совокупность составляющих ее объектов.

Выборочной совокупностью (выборкой) называется совокупность объектов, отобранных случайным образом из генеральной совокупности.

Для получения хороших оценок характеристик генеральной совокупности необходимо, чтобы выборка была *репрезентативной* (*представительной*), т.е. достаточно полно представлять изучаемые признаки генеральной совокупности.

Условием обеспечения репрезентативности выборки является соблюдение случайности отбора, т.е. все объекты генеральной совокупности должны иметь равные вероятности попасть в выборку.

Различают выборки с возвращением (*повторные*) и без возвращения (*бесповторные*). *Повторные выборки* – отобранный объект возвращается в генеральную совокупность перед извлечением следующего. *Бесповторные выборки* – не возвращается. На практике чаще используются бесповторные выборки.

Число N объектов генеральной совокупности и число n объектов выборочной совокупности называют *объемами* генеральной и выборочной совокупностей соответственно. При этом предполагают, что N значительно больше n.

Относительный показатель выборки – это отношение объема выборочной совокупности к объему генеральной совокупности.

1.2. Вариационные ряды

Операция расположения наблюдаемых значений в порядке возрастания называется *ранжированием статистических данных*.

После проведения операции ранжирования значения случайной величины объединяют в группы, то есть группируют так, чтобы в каждой отдельной группе значения случайной величины были бы одинаковыми. Каждое такое значение называется *вариантом.* Варианты обозначаются строчными буквами латинского алфавита с индексами, соответствующими порядковому номеру группы x_i , y_j ,….

Изменение значения варианта называется *варьированием*.

Последовательность вариантов, записанных в возрастающем порядке, называется *вариационным рядом*.

Число, которое показывает, сколько раз встречаются соответствующие значения вариантов в ряде наблюдений, называется *частотой* или *весом варианта* и обозначается n_i $(i = 1,2,...,k)$, где i – номер варианта, $\sum\limits_{i=1}^{k} n_i = n$ – объем выборки.

Отношение частоты данного варианта к общему числу наблюдений n называется *относительной частотой* или *частостью* (*долей*) соответствующего варианта и обозначается $\omega_i = \dfrac{n_i}{n}$ $(\sum\limits_{i=1}^{k} \omega_i = 1,\ i = 1,2,...,k)$, где k – число вариантов, n – объем выборки. Частость ω_i является статистической вероятностью появления варианта x_i. Она является аналогом вероятности p_i появления значения x_i случайной величины X.

Накопленная частота $n_i^{нак}$ показывает, сколько наблюдалось вариантов со значением признака, меньшим x. Отношение накопленной частоты $n_i^{нак}$ к общему числу наблюдений n называется *накопленной частостью* $\omega_i^{нак}$.

Дискретным статистическим рядом называется ранжированная совокупность вариантов x_i с соответствующими им частотами n_i или частостями ω_i.

Дискретный статистический ряд удобно записывать в виде таблицы:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

x_i	x_1	x_2	…	x_k
ω_i	ω_1	ω_2	…	ω_k

Основные характеристики дискретного статистического ряда:

1. *Размах варьирования* $R = x_{\max} - x_{\min}$.

2. *Мода* M_0 – вариант, имеющий наибольшую частоту.

3. *Медиана* M_e – вариант, приходящийся на середину ранжированной выборки.

Пусть n – объем выборки. Тогда:

1) если $n = 2k$, то есть ряд имеет четное число членов, то $M_e = \dfrac{x_k + x_{k+1}}{2}$;

2) если $n = 2k+1$, то есть ряд имеет нечетное число членов, то $M_e = x_{k+1}$.

▲ ***Пример***. Правоохранительными органами зарегистрированы преступления, совершенные подростками. Ниже приведены данные по преступлениям (в качестве исследуемого фактора выступает возраст): 14, 16, 12, 13, 14, 15, 14, 16, 14, 15, 15, 16, 14, 15, 13, 14, 15, 15, 16, 13.

Составить дискретный ряд распределения числа преступлений. Найти объем, размах, моду и медиану выборки.

□ ***Решение***. Произведем ранжирование исходных данных: 12, 13, 13, 13, 14, 14, 14, 14, 14, 14, 14, 15, 15, 15, 15, 15, 16, 16, 16, 16.

Объем выборки: $n = 20$.

Наименьшее значение варианты: $x_{\min} = 12$.

Наибольшее значение варианты: $x_{\max} = 16$.

Размах: $R = x_{\max} - x_{\min} = 16 - 12 = 4$.

Мода: $M_0 = 14$, так как вариант $x_i = 14$ встречается наибольшее количество раз $n_i = 7$.

Так как $n = 20$, то медиана $M_e = \dfrac{x_{10} + x_{11}}{2} = \dfrac{14+14}{2} = \dfrac{28}{2} = 14$.

Вариационный ряд:

x_i	12	13	14	15	16
n_i	1	3	7	5	4

■

1.3. Переход от дискретного к интервальному вариационному ряду

Если изучаемая случайная величина X является непрерывной или число ее значений достаточно велико, то составляют *интервальный статистический ряд.*

Весь интервал наблюдаемых значений разбивается на k частичных интервалов $[x_1, x_2)$, $[x_2, x_3)$,…, $[x_k, x_{k+1}]$ одинаковой длины h. Затем подсчитывают количество попаданий наблюдений в каждый интервал. Эти числа принимают за частоты n_i. Согласно формуле Стерджеса рекомендуемое число интервалов разбиений $k \approx 1 + \log_2 n$, а длины частичных интервалов $h = \dfrac{x_{\max} - x_{\min}}{1 + \log_2 n}$, n – объем выборки. Если шаг окажется дробным, то за длину интервала берут ближайшее целое число или ближайшую простую дробь.

За начало первого интервала рекомендуется брать величину $x_{нач} = x_{\min} - \dfrac{h}{2}$, а конец последнего должен удовлетворять условию $x_{кон} - h \le x_{\max} < x_{кон}$. Промежуточные интервалы получают, прибавляя к концу предыдущего интервала шаг h.

Сгруппированный ряд представляется в виде таблицы (*интервальный статистический ряд*):

$x_i - x_{i+1}$	$x_1 - x_2$	$x_2 - x_3$	…	$x_k - x_{k+1}$
n_i	n_1	n_2	…	n_k

▲ ***Пример***. Правоохранительными органами за отчетный период зарегистрировано следующее количество преступлений: 178, 160. 154, 183, 155, 153, 167, 186, 163, 155, 157, 175, 170, 166, 159, 173, 182, 167, 171, 169, 179, 165, 156, 179, 158, 171, 175, 173, 164, 172. Построить интервальный статистический ряд.

□ ***Решение***. Ранжируем полученные данные: 153, 154, 155, 155, 156, 157, 158, 159, 160, 163, 164, 165, 166, 167, 167, 169, 170, 171, 171, 172, 173, 173, 175, 175, 178, 179, 179, 182, 183, 186.

Объем выборки: $n=30$.

Наименьшее значение варианты: $x_{\min}=153$.

Наибольшее значение варианты: $x_{\max}=186$.

По формуле Стерджеса находим число интервалов разбиений: $k\approx 1+\log_2 30\approx 5{,}9069\approx 6$.

Длина частичных интервалов: $h=\dfrac{186-153}{1+\log_2 30}\approx\dfrac{33}{6}\approx 5{,}5\approx 6$.

Начальное значение: $x_{нач}=153-\dfrac{6}{2}=150$.

Составляем интервальный статистический ряд.

x_i-x_{i+1}	150–156	156–162	162–168	168–174	174–180	180–186
n_i	4	5	6	7	5	3

■

1.4. Переход от интервального к дискретному вариационному ряду

Иногда интервальный статистический ряд необходимо заменить дискретным. В этом случае серединное значение i-го интервала принимают за вариант x_i, а соответствующую интервальную частоту n_i – за частоту этого варианта.

▲ ***Пример***. Дан интервальный вариационный ряд:

x_i-x_{i+1}	150–156	156–162	162–168	168–174	174–180	180–186
n_i	4	5	6	7	5	3

Построить дискретный вариационный ряд.

□ ***Решение***. Найдем середины частичных интервалов, получим числа: 153, 159, 165, 171, 177, 183. Запишем дискретный вариационный ряд.

x_i	153	159	165	171	177	183
n_i	4	5	6	7	5	3

■

1.5. Эмпирическая функция распределения

Пусть получено статистическое распределение выборки и каждому варианту x_i из этой выборки поставлена в соответствие его частость ω_i.

Эмпирической функцией (*функцией распределения выборки*) называется относительная частота (частость) того, что признак (случайная величина X) примет значение, меньшее заданного x, т.е. $F_n(x) = w(X < x) = w_x^{нак} = \dfrac{n_x^{нак}}{n}$,

где n – объем выборки.

Из определения следует, что значения эмпирической функции распределения при каждом x являются случайными величинами. Она обладает теми же свойствами, что и обычная функция распределения вероятностей:

1. $0 \le F^*(x) \le 1$;
2. $F^*(x)$ – неубывающая функция;
3. $F^*(x)$ – непрерывна слева;
4. $F^*(x) = 0$, при $x \le x_{\min}$ и $F^*(x) = +\infty$, при $x > x_{\max}$.

▲ ***Пример***. Построить эмпирическую функцию распределения по данному вариационному ряду:

x_i	0	1	2	3	4	5
n_i	3	6	5	3	2	1

□***Решение***. Объем выборки: $n = 20$.

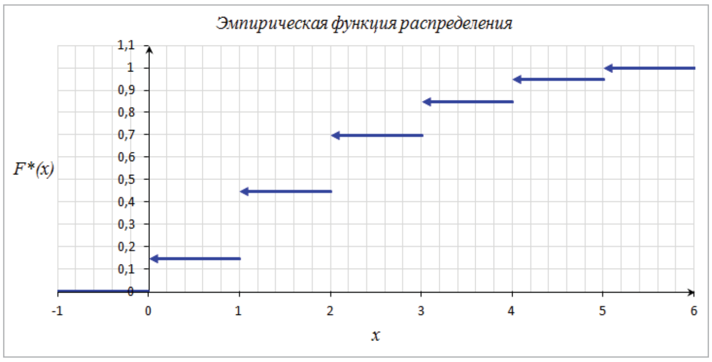
Составим таблицу, в которой найдем относительные частоты и накопленные относительные частоты.

x_i	0	1	2	3	4	5
n_i	3	6	5	3	2	1
ω_i	0,15	0,3	0,25	0,15	0,1	0,05

$\omega_i^{нак}$	0,15	0,45	0,7	0,85	0,95	1

На основе полученной таблицы запишем эмпирическую функцию распределения и построим ее график:

$$F^*(x)=\begin{cases}0, & при \quad x\le 0;\\ 0{,}15, & при \quad 0<x\le 1;\\ 0{,}45, & при \quad 1<x\le 2;\\ 0{,}7, & при \quad 2<x\le 3;\\ 0{,}85, & при \quad 3<x\le 4;\\ 0{,}95, & при \quad 4<x\le 5;\\ 1, & при \quad x>5.\end{cases}$$



■

1.6. Графическое представление статистических данных

Для графического изображения вариационных рядов наиболее часто применяются *полигон, гистограмма* и *кумулятивная кривая*.

Полигон частот или *относительных частот* служит для изображения дискретного вариационного ряда и представляет собой ломаную линию, отрезки которой соединяют точки с координатами $(x_i;n_i)$ или $(x_i;\omega_i)$, $i=1,2,...,k$. Полигон относительных частот является аналогом многоугольника распределения дискретной случайной величины в теории вероятностей.

Гистограмма частот или *относительных частот* служит только для изображения интервальных вариационных рядов и представляет собой ступенчатую фигуру, состоящую из прямоугольников с основаниями, равными частичным интервалам значений признака $h_i=x_{i+1}-x_i$, $i=1,2,...,k$, и высотами, равными частотам n_i или относительным частотам ω_i. Если соединить середины верхних оснований прямоугольников отрезками прямой, томожно получить полигон того же распределения.

Кумулятивная кривая (*кумулята*) – кривая накопленных частот (частостей). Для дискретного вариационного ряда кумулята представляет ломаную линию, соединяющую точки $\left(x_i;\, n_i^{нак}\right)$ или $\left(x_i;\, \omega_i^{нак}\right)$, $i=1,2,...,k$.

Для интервального вариационного ряда ломаная начинается с точки, абсцисса которой равна началу первого интервала, а ордината – накопленной частоте (частости), равной нулю. Другие точки этой ломаной соответствуют концам интервалов.

▲ ***Пример***. В результате некоторого эксперимента получены следующие данные: 2, 4, 9, 5, 4, 8, 3, 6, 7, 2, 8, 7, 5, 3, 7, 4, 4, 3, 4, 6.

Найти размах варьирования R, моду M_0, медиану M_e. Построить полигон и кумуляту частот и относительных частот.

□ ***Решение***. Объем выборки: $n=20$. Размах: $R=7$. Мода: $M_0=4$.

Медиана: $M_e=\dfrac{4+5}{2}=4{,}5$.

Ранжируем исходные данные: 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 6, 6, 7, 7, 7, 8, 8, 9.

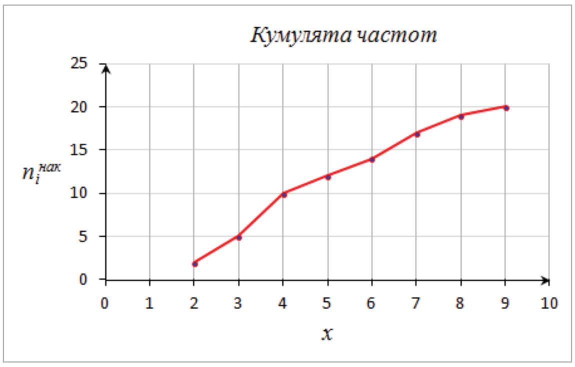
Составим дискретный вариационный ряд.

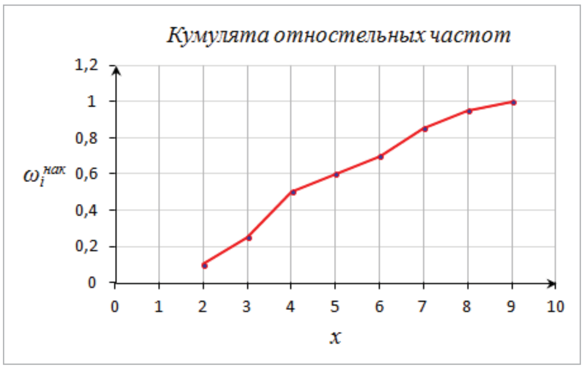
x_i	2	3	4	5	6	7	8	9
n_i	2	3	5	2	2	3	2	1

Найдем относительные частоты, накопленные частоты и накопленные относительные частоты. Составим таблицу:

x_i	2	3	4	5	6	7	8	9
n_i	2	3	5	2	2	3	2	1
ω_i	0,1	0,15	0,25	0,1	0,1	0,15	0,1	0,05
$n_i^{нак}$	2	5	10	12	14	17	19	20
$\omega_i^{нак}$	0,1	0,25	0,5	0,6	0,7	0,85	0,95	1

На основе полученной таблицы строим полигон частот и относительных частот, кумуляту частот и относительных частот.





■

▲ ***Пример***. Дан интервальный вариационный ряд:

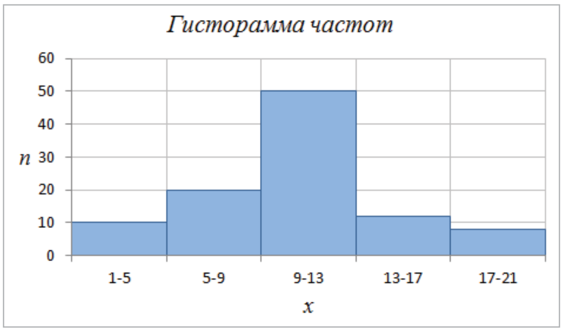
$x_i - x_{i+1}$	1 – 5	5 – 9	9 – 13	13 – 17	17 – 21
n_i	10	20	50	12	8

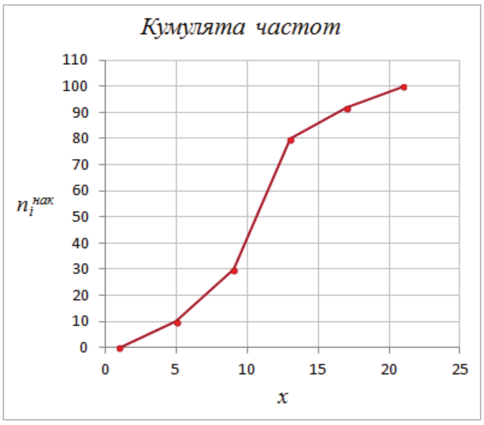
Построить гистограмму и кумуляту частот и относительных частот.

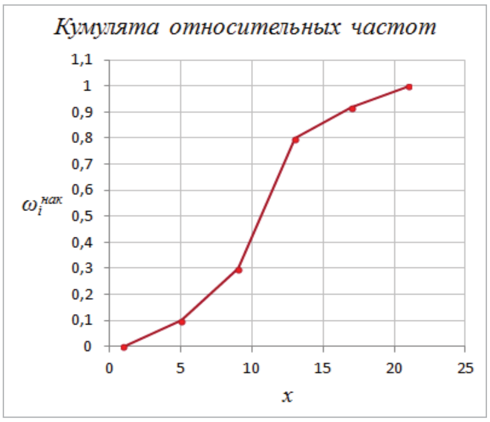
□ ***Решение***. Найдем относительные частоты, накопленные частоты и накопленные относительные частоты. Составим таблицу:

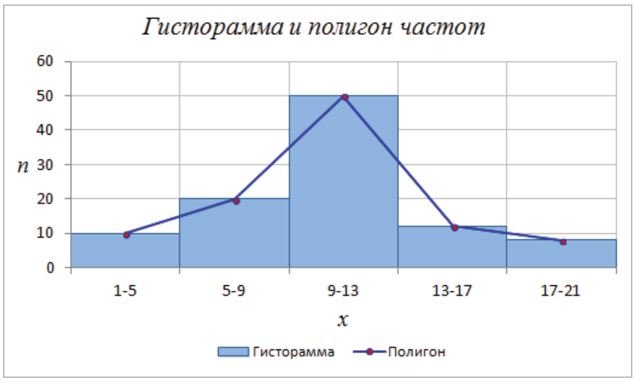
$x_i - x_{i+1}$	1 – 5	5 – 9	9 – 13	13 – 17	17 – 21
n_i	10	20	50	12	8
ω_i	0,1	0,2	0,5	0,12	0,08
$n_i^{нак}$	10	30	80	92	100
$\omega_i^{нак}$	0,1	0,3	0,8	0,92	1

На основе полученной таблицы строим гистограмму частот и относительных частот, кумуляту частот и относительных частот.









■

1.7. Числовые характеристики статистического распределения

Пусть дано статистическое распределение выборки объема n:

x_i	x_1	x_2	$\ldots$	x_m
n_i	n_1	n_2	$\ldots$	n_m

Выборочным средним $\overline{x_в}$ называется среднее арифметическое всех значений выборки: $\overline{x_в} = \frac{1}{n}\sum_{i=1}^{m} x_i n_i$.

Так как $\omega_i = \frac{n_i}{n}$, то для вычисления выборочной средней можно использовать формулу: $\overline{x_в} = \sum_{i=1}^{m} x_i \omega_i$.

В случае интервального вариационного ряда в качестве x_i берут середины интервалов, а n_i – соответствующие им частоты.

Выборочной дисперсией $D_в$ называется среднее арифметическое квадратов отклонений значений выборки от выборочного среднего $\overline{x_в}$: $D_в = \frac{1}{n}\sum_{i=1}^{m}\left(x_i - \overline{x_в}\right)^2 \cdot n_i$.

Выборочная дисперсия может быть вычислена по формуле: $D_в = \overline{x_в{}^2} - \left(\overline{x_в}\right)^2$, где $\overline{x_в{}^2} = \frac{1}{n}\sum_{i=1}^{m} x_i^2 n_i$.

Так как $\omega_i = \frac{n_i}{n}$, то для вычисления выборочной дисперсии можно использовать формулу: $D_в = \sum_{i=1}^{m}\left(x_i - \overline{x_в}\right)^2 \cdot \omega_i$.

Выборочное среднее квадратическое отклонение вычисляется по формуле: $\sigma_в = \sqrt{D_в}$.

Особенность $\sigma_в$ состоит в том, что оно измеряется в тех же единицах, что и данные выборки.

При решении практических задач используется *исправленная выборочная дисперсия* S^2: $S^2 = \frac{n}{n-1} D_в$.

! *Замечание*. Формулы для вычисления выборочной дисперсии и исправленной дисперсии отличаются только знаменателями. При достаточно большом объеме выборки n выборочная $D_в$ и исправленная S^2 дисперсии мало отличаются друг от друга, поэтому на практике исправленной дисперсией пользуются, если $n \le 30$.

Исправленное среднее квадратическое отклонение S определяется по формуле: $S = \sqrt{S^2}$.

Еще одним показателем вариационного ряда является безразмерная характеристика – *коэффициент вариации*, равный процентному отношению среднего квадратического отклонения к средней арифметической: $\widetilde{v} = \frac{\sigma_B(x)}{\bar{x}_B} \cdot 100\% \ (\bar{x} \ne 0)$.

Чем больше значение коэффициента вариации, тем относительно больший разброс и меньшая выравненность исследуемых значений. Если коэффициент вариации меньше 10%, то изменчивость вариационного ряда принято считать незначительной, от 10% до 20% относится к средней, больше 20% и меньше 33% к значительной и если коэффициент вариации превышает 33%, то это говорит о неоднородности информации и необходимости исключения самых больших и самых маленьких значений.

▲ ***Пример***. Дан статистический ряд распределения частот:

x_i	1	3	4	5	6
n_i	3	5	7	2	3

Вычислить среднюю выборочную ($\overline{x_в}$), выборочную дисперсию ($D_в$), выборочное среднее квадратическое отклонение ($\sigma_в$), исправленную дисперсию (S^2), исправленное среднее квадратическое отклонение (S).

□ ***Решение***. Объем выборки: $n=20$.

Выборочная средняя:

$$\overline{x_в}=\frac{1}{n}\sum_{i=1}^{m}x_i n_i=\frac{1\cdot3+3\cdot5+4\cdot7+5\cdot2+6\cdot3}{20}=3{,}7.$$

Выборочная дисперсия:

$$\overline{x_в^{\,2}}=\frac{1}{n}\sum_{i=1}^{m}x_i^2 n_i=\frac{1^2\cdot3+3^2\cdot5+4^2\cdot7+5^2\cdot2+6^2\cdot3}{20}=15{,}9.$$

$$D_в=\overline{x_в^{\,2}}-\left(\overline{x_в}\right)^2=15{,}9-(3{,}7)^2=2{,}21.$$

Выборочное среднее квадратическое отклонение:

$$\sigma_в=\sqrt{D_в}=\sqrt{2{,}21}\approx1{,}4866.$$

Исправленная дисперсия:

$$S^2=\frac{n}{n-1}D_в=\frac{20}{20-1}\cdot2{,}21\approx2{,}3263.$$

Исправленное среднее квадратическое отклонение:

$$S=\sqrt{S^2}=\sqrt{2{,}3263}\approx1{,}5252.$$ ■

1.8. Выборочные начальные и центральные моменты статистического распределения. Асимметрия. Эксцесс

Средняя выборочная и выборочная дисперсия являются частными случаями *моментов статистического распределения*.

Начальным выборочным моментом порядка k называется среднее арифметическое k-х степеней всех значений выборки: $\tilde{\nu}_k=\dfrac{\sum_{i=1}^{m}x_i^k\cdot n_i}{n}$ или

$$\tilde{\nu}_k=\sum_{i=1}^{m}x_i^k\cdot\omega_i.$$

Центральным выборочным моментом порядка k называется среднее арифметическое k-х степеней отклонений наблюдаемых значений выборки от выборочного среднего $\overline{x_в}$: $\widetilde{\mu}_k = \dfrac{\sum_{i=1}^{m}\left(x_i - \overline{x_в}\right)^k \cdot n_i}{n}$ или $\widetilde{\mu}_k = \sum_{i=1}^{m}\left(x_i - \overline{x_в}\right)^k \cdot \omega_i$.

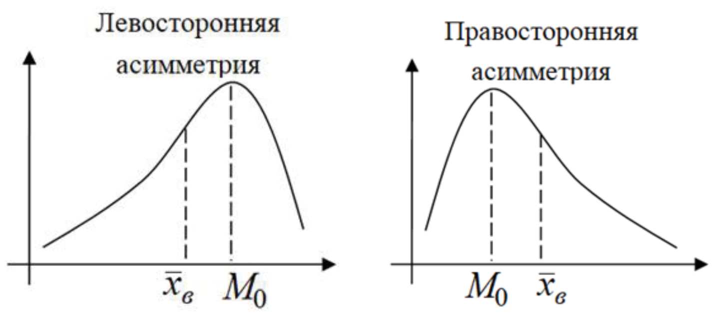
Выборочным коэффициентом асимметрии называется число $\widetilde{A}$, вычисляемое по формуле: $\widetilde{A} = \dfrac{\widetilde{\mu}_3}{\sigma_в^3}$.

Выборочный коэффициент асимметрии служит для характеристики асимметрии полигона вариационного ряда.

Если $\widetilde{A} = 0$, то распределение имеет симметричную форму, т.е. варианты равноудаленные от x, имеют одинаковую частоту (средняя арифметическая, мода и медиана совпадают). Если это равенство нарушается, то распределение ассиметрично.

Если $\widetilde{A} < 0$, то наблюдается положительная (правосторонняя) асимметрия, т.е. более пологий «спуск» полигона наблюдается слева.

Если $\widetilde{A} > 0$, то наблюдается отрицательная (левосторонняя) асимметрия, т.е. более пологий «спуск» полигона наблюдается справа.

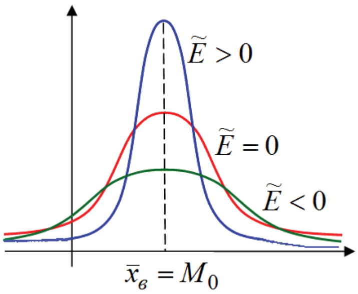


Выборочным коэффициентом эксцесса или *коэффициентом крутости* называется число $\widetilde{E}$, вычисляемое по формуле: $\widetilde{E} = \dfrac{\widetilde{\mu}_4}{\sigma_в^4} - 3$.

Выборочный коэффициент эксцесса служит для сравнения на «крутость» выборочного распределения с нормальным распределением.

Коэффициент эксцесса для случайной величины, распределенной по нормальному закону, равен нулю.

Поэтому за стандартное значение выборочного коэффициента эксцесса принимают $\widetilde{E}=0$. Если $\widetilde{E}>0$ ($\widetilde{E}<0$), то полигон вариационного ряда имеет более крутую (пологую) вершину по сравнению с нормальной кривой.



▲ ***Пример***. Вычислить коэффициент асимметрии и эксцесс по данным представленным в таблице:

№	Значения	Частота
1	94-100	3
2	100-106	7
3	106-112	11
4	112-118	20
5	118-124	28
6	124-130	19
7	130-136	10
8	136-142	2

□ ***Решение***. Объем выборки: $n=100$.

Составим дискретный вариационный ряд распределения частот:

№	x_i	n_i
1	97	3
2	103	7
3	109	11
4	115	20
5	121	28
6	127	19
7	133	10
8	139	2

Выборочная средняя:

$$\overline{x_в}=\frac{1}{n}\sum_{i=1}^{m}x_i n_i=\frac{1}{20}\cdot(97\cdot3+103\cdot7+109\cdot7+115\cdot20+121\cdot28+127\cdot19+$$

$$+133\cdot10+139\cdot2)=119{,}2.$$

Выборочная дисперсия:

$$\overline{x_в^{\,2}}=\frac{1}{n}\sum_{i=1}^{m}x_i^2 n_i=\frac{1}{20}\cdot(97^2\cdot3+103^2\cdot7+109^2\cdot7+115^2\cdot20+121^2\cdot28+$$

$$+127^2\cdot 19+133^2\cdot 10+139^2\cdot 2)=14296{,}1.$$

$$D_в=\overline{x_в^{\,2}}-\left(\overline{x_в}\right)^2=14296{,}1-(119{,}2)^2=87{,}48.$$

Выборочное среднее квадратическое отклонение:

$$\sigma_в=\sqrt{D_в}=\sqrt{87{,}48}\approx 9{,}353.$$

Центральный выборочный момент 3-го порядка:

$$\tilde{\mu}_3=\frac{1}{n}\sum_{i=1}^{m}\left(x_i-\overline{x_в}\right)^3\cdot n_i=\frac{1}{100}\cdot((97-119{,}2)^3\cdot 3+(103-119{,}2)^3\cdot 7+$$

$$+(109-119{,}2)^3\cdot 7+(115-119{,}2)^3\cdot 20+(121-119{,}2)^3\cdot 28+$$

$$+(127-119{,}2)^3\cdot 19+(133-119{,}2)^3\cdot 10+(139-119{,}2)^3\cdot 2)\approx -247{,}54$$

Центральный выборочный момент 4-го порядка:

$$\tilde{\mu}_4=\frac{1}{n}\sum_{i=1}^{m}\left(x_i-\overline{x_в}\right)^4\cdot n_i=\frac{1}{100}\cdot((97-119{,}2)^4\cdot 3+(103-119{,}2)^4\cdot 7+$$

$$+(109-119{,}2)^4\cdot 7+(115-119{,}2)^4\cdot 20+(121-119{,}2)^4\cdot 28+$$

$$+(127-119{,}2)^4\cdot 19+(133-119{,}2)^4\cdot 10+(139-119{,}2)^4\cdot 2)\approx 20767{,}8$$

Коэффициент асимметрии:

$$\tilde{A}=\frac{\tilde{\mu}_3}{\sigma_в^3}=\frac{-247{,}54}{9{,}353^3}=-0{,}3025.$$

Эксцесс:

$$\tilde{E}=\frac{\tilde{\mu}_4}{\sigma_в^4}-3=\frac{20767{,}8}{9{,}353^4}-3=-0{,}2862.$$

Так как коэффициент асимметрии $\tilde{A}$ отрицателен, то наблюдается левосторонняя асимметрия. Поскольку эксцесс $\tilde{E}$ отрицательный, то рассматриваемое распределение имеет пологую вершину по сравнению с нормальной кривой. ■

1.9. Применение электронных таблиц Excel для нахождения числовых характеристик выборочной совокупности

Для вычисления числовых характеристик выборочной совокупности удобно применять электронные таблицы Excel. Продемонстрируем это на конкретном примере.

▲ ***Пример***. Зарегистрированы данные о лицах, совершивших кражи:

Возраст	Число выявленных лиц, совершивших преступление
10-14	25
14-18	17
18-22	19
22-26	21
26-30	18
30-34	30
34-38	16
38-42	20
42-46	16
46-50	23

На основании этих данных необходимо вычислить объем выборки n, размах R, среднюю выборочную $\bar{x}_в$, выборочную дисперсию $D_в$, выборочное среднее квадратическое отклонение $\sigma_в$, исправленную дисперсию S^2, исправленное среднее квадратическое отклонение S, начальные ν_k ($k=1,2,3,4$) и центральные моменты μ_k ($k=1,2,3,4$), асимметрию $\widetilde{A}$, эксцесс $\widetilde{E}$.

□ ***Решение***. На рабочем листе книги Excel составим дискретный вариационный ряд распределения частот.

	A	B	C
1	№	x_i	n_i
2	1	2	3
3	1	12	25
4	2	16	17
5	3	20	19
6	4	24	21
7	5	28	18
8	6	32	30
9	7	36	16
10	8	40	20
11	9	44	16
12	10	48	23

Для нахождения всех числовых характеристик выборочной совокупности дополним таблицу вспомогательными столбцами.

	A	B	C	D	E	F	G	H	I	J	K
1	№	x_i	n_i	w_i	x_iw_i	x_i^2	$x_i^2w_i$	x_i^3	$x_i^3w_i$	x_i^4	$x_i^4w_i$
2	1	2	3	4	5	6	7	8	9	10	11
3	1	12	25	0,121951	1,463415	144	17,56098	1728	210,7317	20736	2528,780488
4	2	16	17	0,082927	1,326829	256	21,22927	4096	339,6683	65536	5434,692683
5	3	20	19	0,092683	1,853659	400	37,07317	8000	741,4634	160000	14829,26829
6	4	24	21	0,102439	2,458537	576	59,00488	13824	1416,117	331776	33986,80976
7	5	28	18	0,087805	2,458537	784	68,83902	21952	1927,493	614656	53969,79512
8	6	32	30	0,146341	4,682927	1024	149,8537	32768	4795,317	1048576	153450,1463
9	7	36	16	0,078049	2,809756	1296	101,1512	46656	3641,444	1679616	131091,9805
10	8	40	20	0,097561	3,902439	1600	156,0976	64000	6243,902	2560000	249756,0976
11	9	44	16	0,078049	3,434146	1936	151,1024	85184	6648,507	3748096	292534,322
12	10	48	23	0,112195	5,385366	2304	258,4976	110592	12407,88	5308416	595578,3805
13	сумма		205	1	29,77561		1020,41		38372,53		1533160,273

	L	M	N	O	P	Q	R	S
1	x_i-$x_в$	$(x_i$-$x_в)w_i$	$(x_i$-$x_в)^2$	$(x_i$-$x_в)^2w_i$	$(x_i$-$x_в)^3$	$(x_i$-$x_в)^3w_i$	$(x_i$-$x_в)^4$	$(x_i$-$x_в)^4w_i$
2	12	13	14	15	16	17	18	19
3	-17,7756	-2,16776	315,9723	38,53321	-5616,6	-684,951	99838,5	12175,43
4	-13,7756	-1,14237	189,7674	15,73681	-2614,16	-216,784	36011,68	2986,334
5	-9,77561	-0,90603	95,56255	8,857016	-934,182	-86,5827	9132,2	846,399
6	-5,77561	-0,59165	33,35767	3,417127	-192,661	-19,736	1112,734	113,9874
7	-1,77561	-0,15591	3,15279	0,27683	-5,59812	-0,49154	9,940085	0,872788
8	2,22439	0,325521	4,947912	0,724085	11,00609	1,610647	24,48183	3,582707
9	6,22439	0,485806	38,74303	3,023847	241,1518	18,8216	1501,023	117,153
10	10,22439	0,997501	104,5382	10,19884	1068,839	104,277	10928,23	1066,168
11	14,22439	1,110196	202,3333	15,79187	2878,068	224,6297	40938,76	3195,22
12	18,22439	2,044688	332,1284	37,26319	6052,838	679,0988	110309,3	12376,16
13		0		133,8228		19,89203		32881,31

Легко заметить, что сумма столбца 3 соответствует объему выборки, т.е. $n = \sum_{i=1}^{k} n_i = 205.$

Сумма столбца 5 соответствует значению средней выборочной и начальному моменту первого порядка, т.е. $\bar{x}_в = \sum_{i=1}^{m} x_i \omega_i = 29{,}77561$ и

$$\nu_1 = \sum_{i=1}^{m} x_i \cdot \omega_i = 29{,}77561.$$

Сумма столбца 7 определяет начальный момент второго порядка, т.е.

$$\nu_2 = \sum_{i=1}^{m} x_i^2 \cdot \omega_i = 1020{,}41.$$

Сумма столбца 9 определяет начальный момент третьего порядка, т.е.

$$\nu_3 = \sum_{i=1}^{m} x_i^3 \cdot \omega_i = 38372{,}53.$$

Сумма столбца 11 соответствует значению начального момента четвертого порядка, т.е. $\nu_4 = \sum_{i=1}^{m} x_i^4 \cdot \omega_i = 153316($.

Сумма столбца 13 определяет значение центрального момента первого порядка, т.е. $\mu_1 = \sum_{i=1}^{m} \left(x_i - \overline{x_в}\right) \cdot \omega_i = 0$.

Сумма столбца 15 соответствует значению выборочную дисперсию и центральному моменту второго порядка, т.е. $D_в = \sum_{i=1}^{m} \left(x_i - \overline{x_в}\right)^2 \cdot \omega_i = 133{,}8228$ и

$\mu_2 = \sum_{i=1}^{m} \left(x_i - \overline{x_в}\right)^2 \cdot \omega_i = 133{,}8228.$

Сумма столбца 17 определяет значение центрального момента третьего порядка: $\mu_3 = \sum_{i=1}^{m} \left(x_i - \overline{x_в}\right)^3 \cdot \omega_i = 19{,}89203.$

Сумма столбца 19 соответствует значению центральный момент четвертого порядка: $\mu_4 = \sum_{i=1}^{m} \left(x_i - \overline{x_в}\right)^4 \cdot \omega_i = 3288{,}31.$

Остальные числовые характеристики не сложно найти по известным формулам:

– размах: $R = x_{\max} - x_{\min} = 48 - 12 = 36$;

– выборочное среднее квадратическое отклонение: $\sigma_в = \sqrt{D_в} = 11{,}5681\!\!\!\;8$;

– исправленную дисперсию: $S^2 = \dfrac{n}{n-1} D_в = 134{,}4788$;

– исправленное среднее квадратическое отклонение: $S = \sqrt{S^2} = 11{,}5965$;

– асимметрия: $\widetilde{A} = \dfrac{\widetilde{\mu}_3}{\sigma_в^3} = \dfrac{19{,}89203}{11{,}56818^3} = 0{,}01284 9$;

– эксцесс: $\widetilde{E} = \dfrac{\widetilde{\mu}_4}{\sigma_в^4} - 3 = \dfrac{3288{,}31}{11{,}56818^4} - 3 = -1{,}16393.$ ■

▲ ***Пример***. При статистическом обследовании семей получены сведения о возрасте детей, являющихся студентами ВУЗов:

Дети студенты ВУЗов по возрасту, лет (x)	Число семей, (n)
17	18
18	46
19	58
20	12
21	23
22	29
23	35
24	19

Необходимо рассчитать показатели вариации среднего возраста детей-студентов ВУЗов.

□ ***Решение***. Составим и заполним таблицу:

x_i	n_i	$x_i n_i$	x_i-x_{cp}	$(x_i-x_{cp})^2$	$(x_i-x_{cp})^2 n_i$
17	18	306	-3,242	10,508	189,1513
18	46	828	-2,242	5,0251	231,1532
19	58	1102	-1,242	1,5417	89,42069
20	12	240	-0,242	0,0584	0,700833
21	23	483	0,7583	0,5751	13,2266
22	29	638	1,7583	3,0917	89,66035
23	35	805	2,7583	7,6084	266,2941
24	19	456	3,7583	14,125	268,3763
Σ	240	4858			1147,983

Общее количество семей: $n = 240$.

1. Средний возраст детей-студентов определим по формуле средней выборочной: $\overline{x_в} = \frac{1}{n}\sum_{i=1}^{m} x_i n_i = \frac{4858}{240} = 20{,}24$ (года).

2. Вычислим отклонения индивидуального возраста студента от среднего $\left(x_i - \overline{x_в}\right)$, а также квадрат отклонений $\left(x_i - \overline{x_в}\right)^2$. Далее вычислим дисперсию и среднее квадратическое отклонение:

$$\sigma_в^2 = \frac{1}{n}\sum_{i=1}^{m}\left(x_i - \overline{x_в}\right)^2 n_i = \frac{1147{,}983}{240} = 4{,}78 \text{ (года)}; \ \sigma_в = \sqrt{\sigma_в^2} = 2{,}19 \text{ (года)}.$$

Таким образом, индивидуальное значение возраста студента в среднем отклоняется на 2,19 года.

4. Вычислим коэффициент вариации:

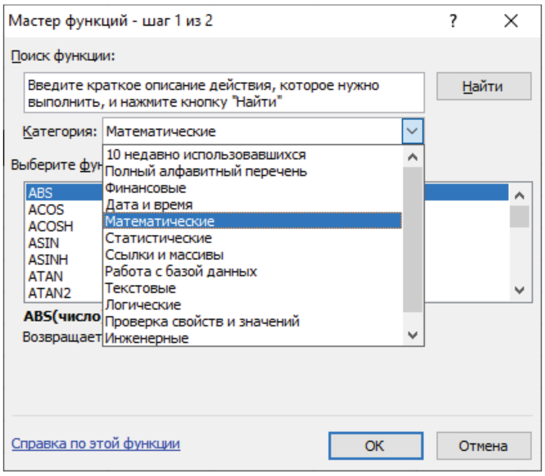
$$\tilde{v} = \frac{\sigma_в(x)}{\bar{x}_в} \cdot 100\% = \frac{2{,}19}{20{,}24} \cdot 100\% = 10{,}8\%.$$

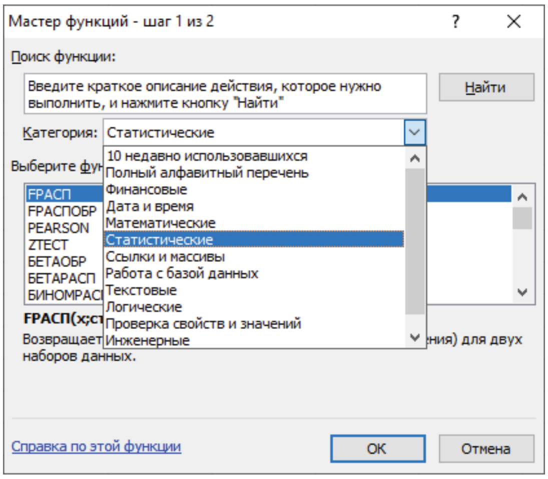
Полученное значение показывает, что в результате статистического наблюдения исследуемая статистическая совокупность распределения детей-студентов ВУЗов в семьях является однородной, т.к. $\tilde{v} = 10{,}8\% < 33\%$. ■

1.10. Вычисление числовых характеристик выборочной совокупности с помощью встроенных функций Excel

Важным инструментом обработки данных в электронных таблицах являются встроенные функции.

Основными разделами библиотеки табличного процессора Excel являются *Статистические* и *Математические функции*. Эти разделы предназначены для решения некоторых наиболее востребованных задач теории вероятностей и математической статистики.





Рассмотрим некоторые функции, применяемые для вычисления числовых характеристик выборочной совокупности:

ДИСП(число1*;число2; …)* –оценивает дисперсию по выборке;

КВАДРОТКЛ(число1*;число2;…)* – возвращает сумму квадратов отклонений точек данных от их среднего;

КОРЕНЬ(число) – возвращает положительное значение квадратного корня;

МАКС(число1*;число2; …)* – возвращает наибольшее значение в списке аргументов;

МЕДИАНА(число1*;число2;…)* – возвращает медиану заданных чисел;

МИН(число1*;число2; ...)* – возвращает наименьшее значение в списке аргументов;

МОДА(число1*;число2; ...)* – возвращает значение моды набора данных;

СРЗНАЧ(число1*; число2; ...)* – возвращает среднее арифметическое своих аргументов;

СТАНДОТКЛОН(число1*; число2; ...)* – оценивает стандартное отклонение по выборке;

СУММ(число1*;число2; ...)* –суммирует аргументы;

СУММПРОИЗВ(массив1;массив2;массив3; ...) – перемножает соответствующие элементы заданных массивов и возвращает сумму произведений;

СЧЕТ(значение1*; значение2; ...)* – подсчитывает количество чисел в списке аргументов;

ЧАСТОТА(массив_данных;массив_интервалов) – возвращает распределение частот в виде вертикального массива;

ЭКСЦЕСС(число1*;число2; ...)* – возвращает эксцесс множества данных.

▲ ***Пример***. В результате некоторого эксперимента получены следующие статистические данные: 183, 158, 171, 173, 158, 185, 168, 180, 168, 178, 169, 170, 183, 170, 173, 174, 178, 180, 180, 183, 155, 183, 178, 188, 190, 170.

Необходимо с помощью статистических функций Excel вычислить объем выборки n, размах R, моду M_o, медиану M_e, среднюю выборочную $\bar{x}_в$, выборочную дисперсию $D_в$, выборочное среднее квадратическое отклонение $\sigma_в$, исправленную дисперсию S^2, исправленное среднее квадратическое отклонение S, эксцесс $\widetilde{E}$.

□ ***Решение***. Запишем исходные данные на лист книги Excel и, применяя встроенные функции, вычислим все числовые характеристики.

	A	B
1	№	x_i
2	1	183
3	2	158
4	3	171
5	4	173
6	5	158
7	6	185
8	7	168
9	8	180
10	9	168
11	10	178
12	11	169
13	12	170
14	13	183
15	14	170
16	15	173
17	16	174
18	17	178
19	18	180
20	19	180
21	20	183
22	21	155
23	22	183
24	23	178
25	24	188
26	25	190
27	26	170

	D	E	F	G	H	I
1						
2	Объем выборки:				N=	=СЧЁТ(B2:B27)
3	Наименьшая варианта:				x_{min}=	=МИН(B2:B27)
4	Наибольшая варианта:				x_{max}=	=МАКС(B2:B27)
5	Размах выборки:				R=	=МАКС(B2:B27)-МИН(B2:B27)
6	Мода:				M_0=	=МОДА(B2:B27)
7	Медиана:				M_e=	=МЕДИАНА(B2:B27)
8	Средняя выборочная:				x_B=	=СРЗНАЧ(B2:B27)
9	Исправленная дисперсия:				S^2_B=	=ДИСП(B2:B27)
10	Исправленнаое ср. квадр. отклонение:				S_B=	=СТАНДОТКЛОН(B2:B27)
11	Выборочная дисперсия:				D_B=	=(I2-1)/I2*I9
12	Среднее квадратическое отклонение:				σ_B=	=КОРЕНЬ(I11)
13	Сумма всех вариант:				Σx_i=	=СУММ(B2:B27)
14	Эксцесс:				E_s=	=ЭКСЦЕСС(B2:B27)

В результате получим следующие данные:

	D	E	F	G	H	I
1						
2	Объем выборки:				N=	26
3	Наименьшая варианта:				x_{min}=	155
4	Наибольшая варианта:				x_{max}=	190
5	Размах выборки:				R=	35
6	Мода:				M_0=	183
7	Медиана:				M_e=	176
8	Средняя выборочная:				x_B=	174,8461538
9	Исправленная дисперсия:				S^2_B=	82,53538462
10	Исправленнаое ср. квадр. отклонение:				S_B=	9,084898712
11	Выборочная дисперсия:				D_B=	79,36094675
12	Среднее квадратическое отклонение:				σ_B=	8,908476118
13	Сумма всех вариант:				Σx_i=	4546
14	Эксцесс:				E_s=	-0,132939724

■

1.11. Обработка статистических данных с помощью надстройки Анализ данных табличного процессора Excel

Наиболее распространенными средствами анализа данных являются статистические процедуры *Пакета анализа*. Они обладают большими возможностями, чем статистические функции. С их помощью можно решать более сложные задачи обработки статистических данных и выполнять более тонкий анализ этих данных.

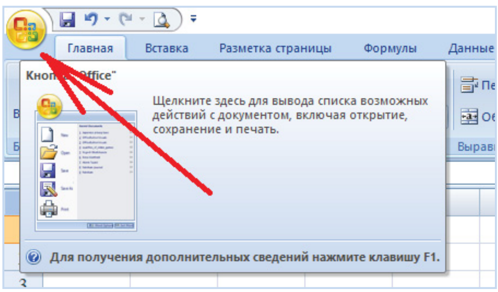
Для доступа к процедурам *Пакета анализа* необходимо выполнить следующие действия:

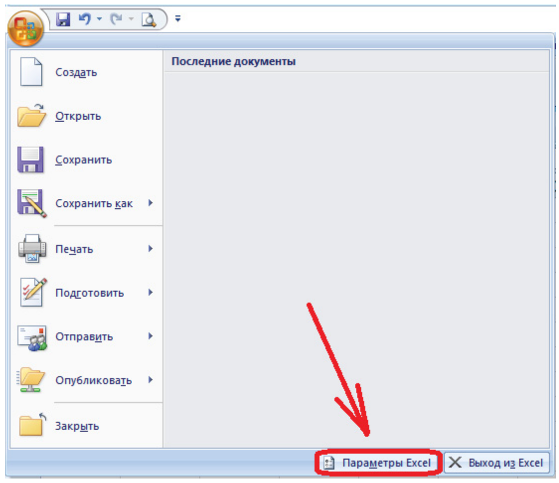
– зайти в параметры Excel;

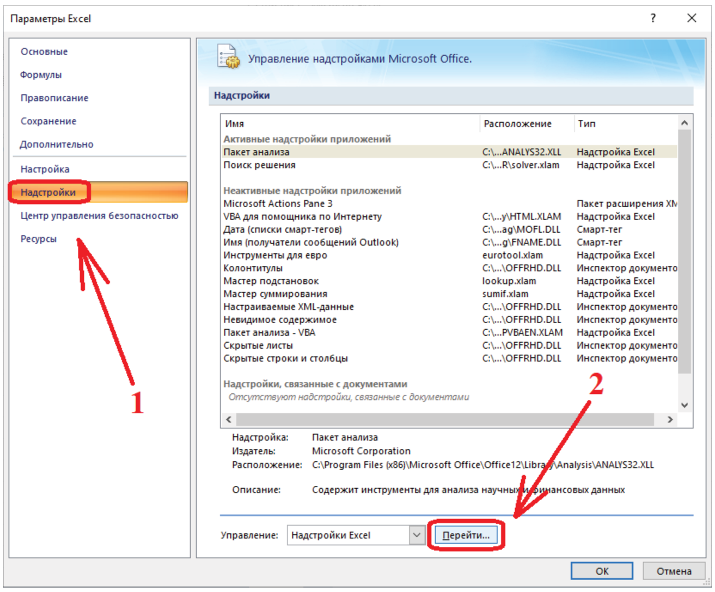
– перейти в надстройки;

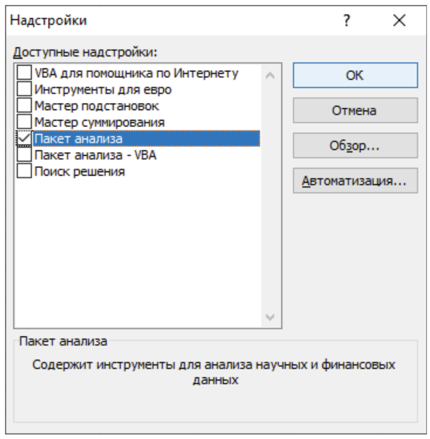
– установить флажок *Пакет анализа*;

– нажать OK.









▲ ***Пример***. В результате некоторого эксперимента получены следующие статистические данные:183, 158, 171, 173, 158, 185, 168, 180, 168, 178, 169, 170, 183, 170, 173, 174, 178, 180, 180, 183, 155, 183, 178, 188, 190, 170.

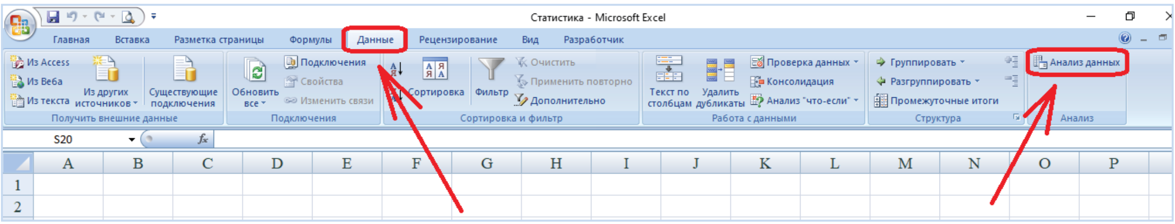
1) Исследовать статистические данные с помощью надстройки Анализ данных Excel.

2) Построить интервальный вариационный ряд распределения частот, разбив исходные данные на интервалы равной длины $h = 5$ и $x_{нач} = 155$.

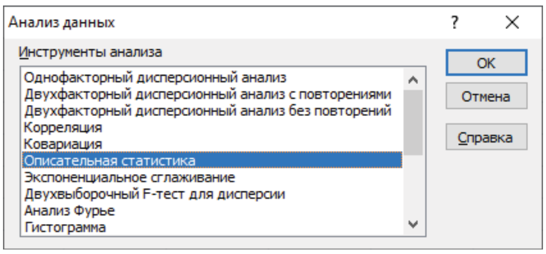
□ ***Решение***. 1) Запишем исходные данные на лист книги Excel.

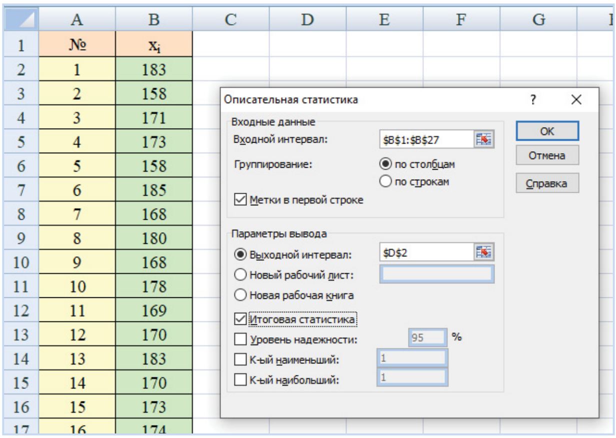
№	x_i
1	183
2	158
3	171
4	173
5	158
6	185
7	168
8	180
9	168
10	178
11	169
12	170
13	183
14	170
15	173
16	174
17	178
18	180
19	180
20	183
21	155
22	183
23	178
24	188
25	190
26	170

В меню *Данные* выбираем *Анализ данных*.



В появившемся окне выбираем *Описательная статистика* и заполняем диалоговое окно.





В качестве *Входной интервал* указываем столбец с исходными статистическими данными (включая название столбца); *Выходной интервал* – любая свободная ячейка рабочего листа; устанавливаем флажки – *Метки в первой строке* и *Итоговая статистика*. В результате получаем данные:

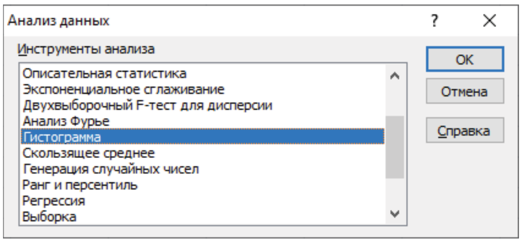
	A	B	C	D	E
1	№	x_i			
2	1	183		*xi*	
3	2	158			
4	3	171		Среднее	174,8461538
5	4	173		Стандартная ошибка	1,781695224
6	5	158		Медиана	176
7	6	185		Мода	183
8	7	168		Стандартное отклонение	9,084898712
9	8	180		Дисперсия выборки	82,53538462
10	9	168		Эксцесс	-0,132939724
11	10	178		Асимметричность	-0,529111747
12	11	169		Интервал	35
13	12	170		Минимум	155
14	13	183		Максимум	190
15	14	170		Сумма	4546
16	15	173		Счет	26

2) Построим интервальный вариационный ряд распределения частот, разбив исходные данные на интервалы равной длины $h=5$ и $x_{нач}=155$.

На рабочем листе книги Excel составим таблицу:

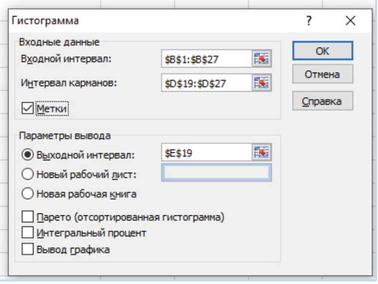
	D
19	Интервалы:
20	155
21	160
22	165
23	170
24	175
25	180
26	185
27	190

Для построения интервального вариационного ряда воспользуемся процедурой *Гистограмма* надстройки *Анализ данных*.



В качестве *Входной интервал* указываем столбец с исходными статистическими данными (включая название столбца); *Интервал карманов* – столбец «*Интервалы:*» (включая название столбца); *Выходной интервал* – любая свободная ячейка; устанавливаем флажок – *Метки*.

	A	B	C	D	E
13	12	170		Минимум	155
14	13	183		Максимум	190
15	14	170		Сумма	4546
16	15	173		Счет	26
17	16	174			
18	17	178			
19	18	180		Интервалы:	
20	19	180		155	
21	20	183		160	
22	21	155		165	
23	22	183		170	
24	23	178		175	
25	24	188		180	
26	25	190		185	
27	26	170		190	
28					



В результате получаем следующие значения:

	D	E	F
19	Интервалы:	*Интервалы:*	*Частота*
20	155	155	1
21	160	160	2
22	165	165	0
23	170	170	6
24	175	175	4
25	180	180	6
26	185	185	5
27	190	190	2
28		Еще	0

Таким образом, интервальный вариационный ряд имеет вид:

≤155	(155;160]	(160;165]	(165;170]	(170;175]	(175;180]	(180;185]	(185;190]
1	2	0	6	4	6	5	2

■

Задачи для самостоятельного решения

Задание 1. По данным МВД зарегистрировано преступлений, совершенных в районе несовершеннолетними в возрасте: 12 13 10 16 10 11 15 14 16 13 14 10 14 15 15 16 16 15 14 15 15 14 16 11 14 10 16 15 16 15 13 15 15 13 15 14 15 13 10 14 11 10 15 12 16.

На основании этих данных необходимо:

1. Построить дискретный ряд распределения частот и относительных частот.

2. Указать наименьшее значение варианты ($x_{наим.}$).

3. Указать наибольшее значение варианты ($x_{наиб.}$).

4. Найти объем выборки (n).

5. Вычислить размах выборки (R).

6. Изобразить графически ряд распределения частот и относительных частот.

7. Найти эмпирическую функцию распределения $F_n(x)$ и построить ее график.

Задание 2. Зарегистрированы данные о лицах, совершивших кражи:

Возраст	Число выявленных лиц, совершивших преступление
14-17	25
17-20	17
20-23	19
23-26	21

Возраст	Число выявленных лиц, совершивших преступление
26-29	18
29-32	30
32-35	16
35-38	19
38-41	15

На основании этих данных необходимо:

1. Построить дискретный ряд распределения частот и относительных частот.

2. Для дискретного вариационного ряда найти объем выборки n, размах варьирования R, несмещенные оценки (x_B, D_B, σ_B), медиану M_e, моду M_o, начальные ν_k и центральные $\mu_к$ моменты *к-го* порядка (k = 1, 2, 3, 4), коэффициент асимметрии A и эксцесс E.

3. Построить полигон и кумуляту частот.

Задание 3. Известны данные о стоимости основного капитала фирм в млн. руб.: 10,4 18,6 10,3 26,0 45,0 18,2 17,3 19,2 25,8 18,7 28,2 25,2 18,4 17,5 41,8 14,6 10,0 37,8 10,5 16,0 18,1 16,8 38,5 37,7 17,9 29,0 10,1 28,0 12,0 14,0 14,2 20,8 13,5 42,4 15,5 17,9 19,2 10,8 12,1 12,4 12,9 12,6 16,8 19,7 18,3 36,8 15,0 37,0 13,0 19,5.

Требуется показать распределение фирм по стоимости основного капитала. Графически изобразить интервальный ряд распределения частот.

2. Статистические оценки параметров распределения

2.1. Оценки параметров

Одной из задач статистики является *описание характера распределения* некоторого признака в генеральной совокупности на основании изучения этого признака по некоторой части совокупности (выборки), полученной в результате случайного отбора.

При этом распределение относительных частот в выборке рассматривается как *эмпирическое приближение* к теоретическому распределению вероятностей в генеральной совокупности.

Выяснение закона распределения по данным выборки составляет *главную проблему математической статистики*, так как на основании закона распределения изучаемого признака можно решать задачи по анализу и предсказанию результатов массового процесса.

На практике часто теоретический закон распределения случайной величины в генеральной совокупности известен (или построено его приближенное аналитическое представление), т.е. известно, что закон распределения принадлежит к тому или иному семейству (нормальный закон, закон Пуассона и т.д.), зависящему от одного или нескольких параметров.

Если бы точные значения параметров были известны, например, a и σ при нормальном законе, λ при законе Пуассона, тогда закон распределения был бы полностью определен. Поэтому именно для определения этих параметров и проводится само *статистическое исследование*.

Пусть изучается случайная величина X с законом распределения, зависящим от одного или нескольких параметров. Требуется по выборке $X_1, X_2, ..., X_n$, полученной в результате n наблюдений (опытов), оценить неизвестный параметр θ.

$X_1, X_2, ..., X_n$ – это случайные величины, где X_1 – результат первого наблюдения, X_2 – второго и т.д., причем случайные величины X_i, $i = 1,2,...,n$ имеют такое же распределение, что и случайная величина X .Конкретная выборка $x_1, x_2, ..., x_n$ – это значения (реализация) независимых случайных величин $X_1, X_2, ..., X_n$.

Статистической оценкой $\tilde{\theta}$ *параметра* θ *теоретического распределения* называют его приближенное значение, зависящее от данных выбора.

Оценка $\tilde{\theta}$ представляет собой значение некоторой функции результатов наблюдений над случайной величиной X: $\tilde{\theta} = \tilde{\theta}(X_1, X_2, ..., X_n)$.

Функцию результатов наблюдений называют *выборочной функцией* или *статистикой,* а ее значение в приближенном равенстве – *оценкой.*

Можно сказать, что оценка $\tilde{\theta}$ параметра θ– это *статистика*, которая близка к истинному значению θ.

Оценка $\tilde{\theta}$ является случайной величиной, так как является функцией независимых случайных величин $X_1, X_2, ..., X_n$. Если произвести другую выборку, то функция примет другое значение.

Если число опытов (наблюдений) невелико, то замена неизвестного параметра θ его оценкой $\tilde{\theta}$, например математического ожидания средним арифметическим, приводит к ошибке. Чем меньше число опытов, тем больше ошибка.

К оценке любого параметра предъявляется ряд требований *(свойства статистических оценок),* которым она должна удовлетворять, чтобы быть «близкой» к истинному значению параметра: *свойства несмещенности*, *состоятельности*, *эффективности.*

Оценка $\tilde{\theta}$ параметра θ называется *несмещенной*, если математическое ожидание оценки равно оцениваемому параметру, т.е. $M(\bar{\theta}) = \theta$.

В противном случае оценка называется *смещенной*.

Если $M(\tilde{\theta}) \to \theta$, то оценка $\tilde{\theta}$ называется *асимптотически несмещенной.*

Требование несмещенности оценки является особенно важным при малом числе наблюдений (опытов).

Пусть $\tilde{\theta}$ – статистическая оценка неизвестного параметра $\tilde{\theta}$ теоретического распределения. Допустим, что по выборке объема n найдена оценка $\tilde{\theta}_1$.

Повторим опыт, т.е. извлечем из генеральной совокупности другую выборку того же объема и по ее данным найдем оценку $\tilde{\theta}_2$. Повторяя опыт многократно, получим числа $\tilde{\theta}_1, \tilde{\theta}_2, ..., \tilde{\theta}_k$, которые будут различны между собой. Таким образом, оценку $\tilde{\theta}$ можно рассматривать как случайную величину, а числа $\tilde{\theta}_1, \tilde{\theta}_2, ..., \tilde{\theta}_k$ – как ее возможные значения.

Предположим, что оценка $\tilde{\theta}$ дает приближенное значение θ с избытком. В этом случае каждое найденное по данным выборок число $\tilde{\theta}_i$ ($i = 1, 2, ..., k$) больше истинного значения θ . Следовательно, и математическое ожидание (среднее значение) случайной величины θ больше, чем $\tilde{\theta}$, т.е. $M(\tilde{\theta}) > \theta$. Очевидно, что если $\tilde{\theta}$ дает оценку с недостатком, то выполняется условие $M(\tilde{\theta}) < \theta$.

Таким образом, использование статистической оценки, математическое ожидание которой не равно оцениваемому параметру, привело бы к систематическим (одного знака) ошибкам.

В связи с этим естественно потребовать, чтобы математическое ожидание оценки $\tilde{\theta}$ было равно оцениваемому параметру. Хотя соблюдение этого требования не устранит ошибок (одни значения $\tilde{\theta}$ больше, а другие меньше $\tilde{\theta}$), однако ошибки разных знаков будут встречаться одинаково часто. Иными словами, соблюдение требований $M(\tilde{\theta}) = \theta$ гарантирует от получения систематических ошибок.

Оценка $\tilde{\theta}$ параметра θ называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру, т.е. для любого $\delta > 0$ выполнено условие: $\lim\limits_{n \to \infty} P(|\tilde{\theta} - \theta| < \delta) = 1$.

Это означает, что с увеличением объема выборки оценка параметра θ приближается к истинному его значению.

Свойство состоятельности обязательно для любого правила оценивания (несостоятельные оценки не используются).

Состоятельность оценки $\tilde{\theta}$ часто может быть установлена с помощью следующей теоремы.

▼ ***Теорема 1.*** Если оценка $\tilde{\theta}$ параметра θ является несмещенной и $D(\theta) \to 0$ при $n \to \infty$, то $\tilde{\theta}$ – состоятельная оценка.

Несмещенная оценка $\tilde{\theta}$ параметра θ называется *эффективной*, если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра θ, т.е. оценка $\tilde{\theta}$ эффективна, если ее дисперсия минимальна.

На практике не всегда удается удовлетворить всем выше перечисленным требованиям (несмещенность, состоятельность, эффективность). Поэтому приходится использовать оценки, которые не обладают сразу всеми тремя свойствам.

2.2. Точечные оценки математического ожидания и дисперсии

Оценки статистических параметров подразделяются на *точечные* и *интервальные*. *Точечная оценка* параметра θ определяется одним числом $\tilde{\theta}$. При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, т.е. приводить к грубым ошибкам.

Интервальная оценка определяется двумя числами $\tilde{\theta}_1$ и $\tilde{\theta}_2$ – концами интервала, внутри которого содержится неизвестный параметр θ. Интервальные оценки позволяют установить *точность* и *надежность оценок*.

Пусть изучается случайная величина X с математическим ожиданием $a = M(X)$ и дисперсией $D(X)$. Оба параметра неизвестны.

Статистика, используемая в качестве приближенного значения неизвестного параметра генеральной совокупности, называется ее *точечной оценкой.* Т.е. точечная оценка характеристики генеральной совокупности – это число, определяемое по выборке.

▼ ***Теорема 2***. Пусть $X_1, X_2, ..., X_n$ – выборка, извлеченная из генеральной совокупности, и $M(X_i) = M(X) = a$, $D(X_i) = D(X)$. Тогда выборочное среднее $\bar{x}_в$ – *несмещенная* и *состоятельная* оценка математического ожидания $M(X)$.

Для нормально распределенной случайной величины X оценка $\bar{x}_в$ будет несмещенной, состоятельной и эффективной.

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

Рассмотрим равенство: $M(D_в)=\frac{n-1}{n}D(X)$.

Из данного равенства следует, что $M(D_в)\neq D(X)$, т.е. выборочная дисперсия является смещенной оценкой дисперсии $D(X)$.

Поэтому выборочную дисперсию исправляют, умножая ее на множитель $\left(\frac{n-1}{n}\right)$, получая формулу: $S^2=\frac{n-1}{n}D_в$.

▼ ***Теорема 3***. Пусть $X_1,X_2,...,X_n$ – выборка, извлеченная из генеральной совокупности, и $M(X_i)=M(X)=a$, $D(X_i)=D(X)$. Тогда исправленная выборочная дисперсия S^2 – *несмещенная* и *состоятельная* оценка дисперсии $D(X)$.

▼ ***Теорема 4***. Относительная частота появления события A в n независимых испытаниях является *несмещенной, состоятельной и эффективной* оценкой неизвестной вероятности $p=P(A)$ этого события(p – вероятность наступления события A в каждом испытании).

▼ ***Теорема 5***. Эмпирическая функция распределения выборки $F_n(x)$ является *несмещенной* и *состоятельной* оценкой функции распределения $F(x)$случайной величины X.

2.3. Понятие интервального оценивания параметров

Точечная оценка $\tilde{\theta}$ является лишь приближенным значением неизвестного параметра θ, и при выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра. Чтобы получить представление о точности и надежности оценки $\tilde{\theta}$ параметра θприменяют интервальные оценки.

Оценка неизвестного параметра называется *интервальной*, если она определяется парой чисел $\tilde{\theta}_1$ и $\tilde{\theta}_2$ – концами интервала, т.е. $(\tilde{\theta}_1;\tilde{\theta}_2)$.

Пусть, найденная по данным выборки статистическая характеристика $\tilde{\theta}$, служит оценкой неизвестного параметра θ. Будем считать θ постоянным числом (θ может быть и случайной величиной).

Очевидно, что чем меньше абсолютная величина разности $\left|\theta-\tilde{\theta}\right|$, тем точнее $\tilde{\theta}$ определяет параметр θ.

Другими словами, если $\delta>0$ и $\left|\theta-\tilde{\theta}\right|<\delta$, то чем меньше δ, тем оценка точнее. Таким образом, положительное число δ характеризует *точность оценки*.

Следует заметить, что статистические методы не позволяют категорически утверждать, что оценка $\tilde{\theta}$ удовлетворяет неравенству $\left|\theta-\tilde{\theta}\right|<\delta$, можно лишь говорить о вероятности γ, с которой это неравенство осуществляется.

Надежностью (*доверительной вероятностью*) оценки называют вероятность $\gamma=1-\alpha$ с которой выполняется неравенство $\left|\theta-\tilde{\theta}\right|<\delta$. Обычно надежность оценки задается наперед, причем в качестве γ берут число, близкое к единице, так чтобы событие с вероятностью γ можно было считать практически достоверным. Наиболее часто задают надежность, равную 0,95; 0,99 и 0,999.

Значение α – это уровень значимости, принимающее значения 0,05; 0,01; 0,001.

Пусть выполняется условие $P\left(\left|\theta-\tilde{\theta}\right|<\delta\right)=\gamma$ или $P\left(\tilde{\theta}-\delta<\theta<\tilde{\theta}+\delta\right)=\gamma$.

Т.е. вероятность того, что интервал $\left(\tilde{\theta}-\delta;\tilde{\theta}+\delta\right)$ содержит в себе неизвестный параметр θ, равна γ.

Доверительным интервалом называют интервал $\left(\tilde{\theta}-\delta;\tilde{\theta}+\delta\right)$, который покрывает неизвестный параметр с заданной надежностью γ.

Метод доверительных интервалов разработал американский статистик Ю. Нейман, исходя из идей английского статистика Р. Фишера.

2.4. Доверительные интервалы для оценки параметров нормального распределения

2.4.1. Доверительные интервалы для оценки математического ожидания при известной дисперсии

Предположим, что количественный признак X генеральной совокупности распределен нормально, причем среднее квадратическое отклонение σ этого распределения известно.

Необходимо оценить неизвестное математическое ожидание a по выборочной средней $\bar{x}_в$. Необходимо найти доверительные интервалы, покрывающие параметр a с надежностью γ.

Будем рассматривать выборочную среднюю $\bar{x}_в$ как случайную величину $\bar{X}$ ($\bar{x}_в$ изменяется от выборки к выборке) и выборочные значения признака $x_1, x_2, ..., x_n$ – как одинаково распределенные независимые случайные величины $X_1, X_2, ..., X_n$ (эти числа также изменяются от выборки к выборке).

Другими словами, математическое ожидание каждой из этих величин равно a и среднее квадратическое отклонение – σ.

Если случайная величина X распределена нормально, то выборочная средняя $\bar{x}_в$, найденная по независимым наблюдениям, также распределена нормально. Параметры распределения $\bar{X}$: $M(\bar{X}) = a$; $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.

Данные равенства получены из следующих рассуждений.

Рассмотрим n взаимно независимых случайных величин $X_1, X_2, ..., X_n$, которые имеют одинаковые распределения, а, следовательно, и одинаковые характеристики (математическое ожидание, дисперсию и др.). Наибольший интерес представляет изучение числовых характеристик среднего арифметического этих величин $\bar{X}$.

Рассмотрим следующие положения, которые устанавливают связь между числовыми характеристиками среднего арифметического и соответствующими характеристиками каждой отдельной величины.

1. Математическое ожидание среднего арифметического n одинаково распределенных взаимно независимых случайных величин равно математическому ожиданию a каждой из величин: $M(\bar{X})=a$.

2. Дисперсия среднего арифметического n одинаково распределенных взаимно независимых случайных величин в n раз меньше дисперсии D каждой из величин: $D(\bar{X})=\dfrac{D}{n}$.

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

Вернемся к задаче нахождения доверительных интервалов для оценки математического ожидания нормального распределения при известном σ (покрывающие параметр a с надежностью γ).

Пусть выполняется следующее соотношение: $P(|X-a|<\delta)=\gamma$.

Из теории случайных величин следует, что вероятность того, что отклонение случайной величины X, распределенной по нормальному закону, от математического ожидания a не превысит величину $\delta>0$ (по абсолютной величине), равна: $P(|X-a|<\delta)=2\Phi\left(\dfrac{\delta}{\sigma}\right)$, где $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$ – функция Лапласа.

Заменим в последнем выражении X на $\bar{X}$ и σ на $\sigma(\bar{X})$, тогда получим: $P\left(\left|\bar{X}-a\right|<\delta\right)=2\Phi\left(\dfrac{\delta}{\sigma(\bar{X})}\right)=2\Phi\left(\dfrac{\delta\sqrt{n}}{\sigma}\right)=2\Phi(t)$, где $t=\dfrac{\delta\sqrt{n}}{\sigma}$.

Из последнего выражения выразим δ, тогда получим, что $\delta=\dfrac{t\sigma}{\sqrt{n}}$.

Подставим полученное выражение в последнее соотношение, тогда: $P\left(\left|\bar{X}-a\right|<\dfrac{t\sigma}{\sqrt{n}}\right)=2\Phi(t)$.

Учитывая, что вероятность P задана и равна γ, то окончательно получим (выборочную среднюю вновь обозначим через $\bar{x}_в$): $P\left(\bar{x}_в-\dfrac{t\sigma}{\sqrt{n}}<a<\bar{x}_в+\dfrac{t\sigma}{\sqrt{n}}\right)=2\Phi(t)=\gamma$.

С надежностью γ можно утверждать, что доверительный интервал $\left(\bar{x}_в-\dfrac{t\sigma}{\sqrt{n}};\bar{x}_в+\dfrac{t\sigma}{\sqrt{n}}\right)$ покрывает неизвестный параметр a. Точность оценки равна $\delta=\dfrac{t\sigma}{\sqrt{n}}$.

Число t определяется из равенства $2\Phi(t)=\gamma$ или $\Phi(t)=\dfrac{\gamma}{2}$. По таблице функции Лапласа находим аргумент t, которому соответствует значение функции Лапласа, равное $\dfrac{\gamma}{2}$.

! *Замечание.* Оценку $\left|\bar{X}-a\right|<\dfrac{t\sigma}{\sqrt{n}}$ называют *классической*. Из формулы $\delta=\dfrac{t\sigma}{\sqrt{n}}$, определяющей точность классической оценки, можно сделать следующие выводы:

1. При *возрастании объема выборки* n число δ убывает и, следовательно, точность оценки *увеличивается*.

2. Увеличение надежности оценки $\Phi(t)=\dfrac{\gamma}{2}$ приводит к увеличению t ($\Phi(t)$ – возрастающая функция), следовательно, и к возрастанию δ.

Другими словами, увеличение надежности классической оценки влечет за собой уменьшение ее точности.

▲ ***Пример.*** Случайная величина X имеет нормальное распределение с известным средним квадратическим отклонением $\sigma = 3$. Найти доверительный интервал для оценки неизвестного математического ожидания a по выборочной средней $\bar{x}_в$, если объем выборки $n = 36$ и надежность оценки $\gamma = 0{,}95$.

□ ***Решение.*** Найдем t. Из соотношения $2\Phi(t) = 0{,}95$ получим, что $\Phi(t) = 0{,}475$. По таблице функции Лапласа (Приложение 1) находим значение $t = 1{,}96$.

Найдем точность оценки: $\delta = \frac{t\sigma}{\sqrt{n}} = \frac{1{,}96 \cdot 3}{\sqrt{36}} = 0{,}98$.

Тогда доверительный интервал примет вид: $(\bar{x}_в - 0{,}98; \bar{x}_в + 0{,}98)$.

Например, если $\bar{x}_в = 3$, то доверительный интервал имеет следующие границы: $(3 - 0{,}98; 3 + 0{,}98)$ или $(2{,}02; 3{,}98)$.

Таким образом, значение неизвестного параметра a, согласующиеся с данными выборки, удовлетворяют неравенству: $2{,}02 < a < 3{,}98$. ■

Ошибочным является утверждение, что $P(2{,}02 < a < 3{,}98) = 0{,}95$.

Действительно, так как a – постоянная величина, то либо она заключена в найденном интервале (тогда событие $2{,}02 < a < 3{,}98$ достоверно и его вероятность равна единице), либо в нем не заключена (в этом случае событие $2{,}02 < a < 3{,}98$ невозможно и его вероятность равна нулю).

Другими словами, доверительную вероятность не следует связывать с оцениваемым параметром; она связана лишь с границами доверительного интервала, которые изменяются от выборки к выборке.

Надежность $\gamma = 0{,}95$ указывает на то, что если произведено достаточно большое число выборок, то 95% из них определяет такие доверительные интервалы, в которых параметр действительно заключен, лишь в 5% случаев он может выйти за границы доверительного интервала.

! *Замечание.* Если требуется оценить математическое ожидание с наперед заданной точностью δ и надежностью γ, то минимальный объем выборки n, который обеспечит эту точность, вычисляют по формуле: $n = \dfrac{t^2\sigma^2}{\delta^2}$ (так как $\delta = \dfrac{t\sigma}{\sqrt{n}}$).

▲ ***Пример.*** Найти доверительный интервал для оценки с надежностью 0,99 неизвестного математического ожидания нормально распределенного признака X генеральной совокупности, если даны генеральное среднее квадратическое отклонение σ=4, выборочная средняя $\bar{x}_в = 10{,}2$ и объем выборки $n = 16$.

□ ***Решение.*** Чтобы найти доверительный интервал для оценки неизвестного математического ожидания нормально распределенного признака X генеральной совокупности воспользуемся формулой: $\left(\bar{x}_в - \dfrac{t\sigma}{\sqrt{n}}; \bar{x}_в + \dfrac{t\sigma}{\sqrt{n}}\right)$.

Из соотношения $2\Phi(t) = 0{,}99$ получим, что $\Phi(t) = 0{,}495$. По таблице функции Лапласа (Приложение 1) находим значение $t = 2{,}57$.

Подставляем все данные в формулу, получаем $\left(10{,}2 - \dfrac{2{,}57 \cdot 4}{\sqrt{16}}; 10{,}2 + \dfrac{2{,}57 \cdot 4}{\sqrt{16}}\right)$, $(7{,}63; 12{,}77)$.

Таким образом, с надежностью $\gamma = 0{,}99$ математическое ожидание нормально распределенного признака X генеральной совокупности заключено в доверительном интервале $7{,}63 < a < 12{,}77$. ■

▲ ***Пример.*** Найти минимальный объем выборки, при котором с надежностью 0,9942 точность оценки математического ожидания a генеральной совокупности по выборочной средней равна δ=0,7, если известно среднее квадратическое отклонение σ=1,2 нормально распределенной генеральной совокупности.

□ ***Решение.*** Чтобы найти минимальный объем выборки воспользуемся формулой $n=\frac{t^2\sigma^2}{\delta^2}$.

Из соотношения $2\Phi(t)=0{,}9942$ получим, что $\Phi(t)=0{,}4971$. По таблице функции Лапласа (Приложение 1) находим значение $t=2{,}76$.

Подставляем все данные в формулу, получаем: $n=\frac{(2{,}76)^2\cdot(1{,}3)^2}{(0{,}9)^2}=15{,}894\approx 16$.

Следовательно, минимальный объем выборки равен 16. ■

2.4.2. Доверительные интервалы для оценки математического ожидания при неизвестной дисперсии

Предположим, что количественный признак X генеральной совокупности распределен нормально, причем среднее квадратическое отклонение σ этого распределения неизвестно. Необходимо оценить неизвестное математическое ожидание a с помощью доверительных интервалов.

Найдем такое число δ, чтобы выполнялось соотношение: $P(|X-a|<\delta)=\gamma$.

Введем случайную величину, имеющая распределениеСтьюдента с $k=n-1$ степенями свободы: $T=\frac{(\overline{X}-a)}{\frac{S}{\sqrt{n}}}$, где S – исправленное среднее квадратическое отклонение; n – объем выборки; $\overline{X}$ – выборочная средняя,

$$S=\sqrt{\frac{n}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}.$$

Перейдем от случайной величины $\overline{X}$ к случайной величине T, получим: $P\left(\left|\frac{\overline{X}-a}{\frac{S}{\sqrt{n}}}\right|<\frac{\delta}{\frac{S}{\sqrt{n}}}\right)=\gamma$.

Следовательно, $P\left(|T| < \frac{\delta\sqrt{n}}{S}\right) = \gamma$ или $P(|T| < t_\gamma) = \gamma$, где $t_\gamma = \frac{\delta\sqrt{n}}{S}$, $\delta = t_\gamma \frac{S}{\sqrt{n}}$.

Таким образом, доверительный интервал принимает вид:

$$P\left(\overline{X} - t_\gamma \frac{S}{\sqrt{n}} < a < \overline{X} + t_\gamma \frac{S}{\sqrt{n}}\right) = \gamma.$$

Значит, интервал $\left(\overline{X} - t_\gamma \frac{S}{\sqrt{n}}; \overline{X} + t_\gamma \frac{S}{\sqrt{n}}\right)$ покрывает неизвестный параметр a с надежностью γ, т.е. является доверительным интервалом для неизвестного математического ожидания случайной величины X.

Значение t_γ находится по таблице распределения Стьюдента (Приложение 2)в зависимости от доверительной вероятности γ и числа степеней свободы $(n-1)$.

▲ ***Пример.*** Количественный признак X генеральной совокупности распределен нормально. При выборке объема $n=16$ найдены выборочная средняя $\bar{x}_в = 20{,}2$и исправленное среднее квадратическое отклонение $s = 0{,}8$. Оценить неизвестное математического ожидание при помощи доверительного интервала с надежностью $\gamma = 0{,}95$.

□ ***Решение***. По условию имеем $\bar{x}_в = 20{,}2$, $s = 0{,}8$, $n = 16$, $\gamma = 0{,}95$.

По таблице (Приложение 2) для $\gamma = 0{,}95$ и $n = 16$найдем значение t_γ. Получим, $t_\gamma = t(0.95;16) = 2{,}13$.

Найдем доверительные границы интервала по формуле $\left(\overline{X} - t_\gamma \frac{S}{\sqrt{n}}; \overline{X} + t_\gamma \frac{S}{\sqrt{n}}\right)$.

Получаем, $\left(20{,}2 - 2{,}13 \cdot \frac{0{,}8}{\sqrt{16}}; 20{,}2 + 2{,}13 \cdot \frac{0{,}8}{\sqrt{16}}\right), (19{,}774; 20{,}626)$.

Следовательно, с надежностью $\gamma = 0{,}95$ неизвестный параметр a заключен в доверительном интервале $(19{,}774; 20{,}626)$. ■

▲ ***Пример.*** Имеется выборка объема $n=13$ из некоторой генеральной совокупности. Исследуемый количественный признак X этой совокупности распределен нормально. По этой выборке найдены выборочная средняя $\bar{x}_в=0{,}2$ и выборочная дисперсия $D_в=0{,}48$. Найти доверительный интервал, покрывающий параметр a с надежностью $\gamma=0{,}95$.

□ ***Решение***. По условию имеем $\bar{x}_в=0{,}2$, $D_в=0{,}48$, $n=13$, $\gamma=0{,}95$. По таблице (Приложение 2) для $\gamma=0{,}95$ и $n=13$ найдем значение t_γ. Получим, $t_\gamma=t(0{,}95;13)=2{,}18$.

Вычислим $S^2=\dfrac{n}{n-1}\cdot D_в=\dfrac{13}{12}\cdot 0{,}48=0{,}52$.

Найдем доверительные границы интервала по формуле $\left(\overline{X}-t_\gamma\dfrac{S}{\sqrt{n}};\overline{X}+t_\gamma\dfrac{S}{\sqrt{n}}\right)$.

Получаем, $\left(0{,}2-2{,}18\cdot\dfrac{0{,}52}{\sqrt{13}};0{,}2+2{,}18\cdot\dfrac{0{,}52}{\sqrt{13}}\right)$, $(-0{,}24;0{,}64)$.

Следовательно, с надежностью $\gamma=0{,}95$ неизвестный параметр a заключен в доверительном интервале $(-0{,}24;0{,}64)$. ■

При неограниченном возрастании объема выборки n распределение Стьюдента стремится к нормальному. Поэтому на практике при $n>30$ можно вместо распределения Стьюдента воспользоваться нормальным распределением.

Для малых выборок замена распределения нормальным приводит к грубым ошибкам, а именно к неоправданному сужению доверительного интервала, т.е. к повышению точности оценки.

Например, если $n=5$ и $\gamma=0{,}99$, то, применяя распределение Стьюдента, найдем $t_\gamma=4{,}6$, а используя функцию Лапласа, найдем $t_\gamma=2{,}58$.Таким образом, доверительный интервал в последнем случае окажется более узким, чем найденный по распределению Стьюдента.

▲ ***Пример.*** По данным девяти независимых равноточных (одинаковые дисперсии) измерений физической величины найдены среднее арифметическое результатов отдельных измерений $\bar{x}_в = 42{,}319$ и исправленное среднее квадратическое отклонение $s = 5$. Оценить истинное значение измеряемой величины с надежностью $\gamma = 0{,}95$.

□ ***Решение.*** Истинное значение измеряемой величины равно ее математическому ожиданию. Поэтому задача сводится к оценке математического ожидания (при неизвестном σ) при помощи доверительного интервала: $\left(\bar{X} - t_\gamma \frac{S}{\sqrt{n}}; \bar{X} + t_\gamma \frac{S}{\sqrt{n}}\right)$.

Для $\gamma = 0{,}95$ и $n = 9$, получаем $t_\gamma = 2{,}31$. Найдем доверительные интервалы: $\left(\bar{X} - t_\gamma \frac{S}{\sqrt{n}}; \bar{X} + t_\gamma \frac{S}{\sqrt{n}}\right)$, $(38{,}469; 46{,}169)$.

Таким образом, с надежностью $\gamma = 0{,}95$ истинное значение измеряемой величины заключено в доверительном интервале: $38{,}469 < a < 46{,}169$. ■

2.4.3. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения

Пусть количественный признак X генеральной совокупности распределен нормально. Требуется оценить неизвестное генеральное среднее квадратическое отклонение σ по исправленному выборочному среднему квадратическому отклонению s. Необходимо найти доверительные интервалы, покрывающие параметр σ с заданной надежностью γ.

По определению доверительного интервала имеем: $P(|\sigma - s| < \delta) = \gamma$.
Т.е. $P(s - \delta < \sigma < s + \delta) = \gamma$.

Для того, чтобы можно было пользоваться готовой таблицей, преобразуем двойное неравенство: $s\left(1 - \frac{\delta}{s}\right) < \sigma < s\left(1 + \frac{\delta}{s}\right)$.

Введем обозначение: $q = \frac{\delta}{s}$. Следовательно, $s(1-q) < \sigma < s(1+q)$.

! ***Замечание***. Типичная для подобных задач ситуация, когда $\delta < s$, т.е. $q < 1$. Считаем, что последнее выполняется, тогда: $\frac{1}{s(1+q)} < \frac{1}{\sigma} < \frac{1}{s(1-q)}$.

Умножим все части неравенства на $s\sqrt{n-1} > 0$, получим: $\frac{\sqrt{n-1}}{(1+q)} < \frac{s\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{(1-q)}$.

Введем случайную величину, которая распределена по закону χ^2 с $(n-1)$ степенями свободы. Это распределение не зависит от оцениваемого параметра σ, зависит лишь от объема выборки n: $\chi^2 = \frac{s^2(n-1)}{\sigma^2}$.

Следовательно, $\chi = \frac{s\sqrt{n-1}}{\sigma}$.

Получим окончательные формулы для доверительного интервала

$$\frac{s\sqrt{n-1}}{\chi_2} < \sigma < \frac{s\sqrt{n-1}}{\chi_1}, \quad \chi_1^2 = \chi^2\left(\frac{1+\gamma}{2}; n-1\right), \quad \chi_2^2 = \chi^2\left(\frac{1-\gamma}{2}; n-1\right).$$

Определяются по таблице «*хи -квадрат*» распределения.

При построении доверительного интервала для дисперсии достаточно возвести все члены последнего неравенства в квадрат.

▲ ***Пример.*** Пусть по выборке объема 20 найдено $s = 0{,}03$. Требуется с надежностью $\gamma = 0{,}9$ построить доверительный интервал для σ.

□ ***Решение.*** По таблице для $\gamma = 0{,}9$ и $n = 20$ находим:

$$\chi_1^2 = \chi^2\left(\frac{1+0{,}9}{2}; 20-1\right) = \chi^2(0{,}95; 19) = 10{,}1,$$

$$\chi_2^2 = \chi^2\left(\frac{1-0{,}9}{2}; 20-1\right) = \chi^2(0{,}05; 19) = 30{,}1.$$

Доверительный интервал:

$$\frac{s\sqrt{n-1}}{\chi_2} < \sigma < \frac{s\sqrt{n-1}}{\chi_1}, \quad \frac{0{,}03\sqrt{20-1}}{\sqrt{30{,}1}} < \sigma < \frac{0{,}03\sqrt{20-1}}{\sqrt{10{,}1}}, \quad 0{,}024 < \sigma < 0{,}041.$$

■

Оценка точности измерений

В теории ошибок принято точность измерений (точность прибора) характеризовать с помощью среднего квадратического отклонения σ случайных ошибок измерений. Для оценки σ используют «исправленное» среднее квадратическое отклонение s.

Поскольку обычно результаты измерений взаимно независимы, имеют одно и то же математическое ожидание (истинное значение измеряемой величины) и одинаковую дисперсию (в случае равноточных измерений), то выше изложенная теория применима для оценки точности измерений.

▲ ***Пример.*** По 15 равноточным измерениям найдено «исправленное» среднее квадратическое отклонение $s = 0{,}12$. Найти точность измерений с надежностью $\gamma = 0{,}99$.

□ ***Решение.*** Точность измерений характеризуется средним квадратическим отклонением σ случайных ошибок, поэтому задача сводится к отысканию доверительного интервала, покрывающего σ с заданной надежностью $\gamma = 0{,}99$.

$$\chi_1^2 = \chi^2\left(\frac{1+0{,}99}{2};15-1\right) = \chi^2(0{,}995;14) = 4{,}1,$$

$$\chi_2^2 = \chi^2\left(\frac{1-0{,}99}{2};14\right) = \chi^2(0{,}005;14) = 31{,}32.$$

Доверительный интервал

$$\frac{s\sqrt{n-1}}{\chi_2} < \sigma < \frac{s\sqrt{n-1}}{\chi_1},\quad \frac{0{,}12\sqrt{15-1}}{\sqrt{31{,}32}} < \sigma < \frac{0{,}12\sqrt{15-1}}{\sqrt{4{,}1}},\quad 0{,}08 < \sigma < 0{,}22.$$

■

Задачи для самостоятельного решения

Задание 1. В результате измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты (в мм): 3,1; 3,3; 3,7; 4,1; 4,5; 4,7; 5,3. Вычислить несмещенную оценку математического ожидания, смещенную оценку дисперсии, несмещенную оценку дисперсию.

Задание 2. Дан доверительный интервал (-0,26; 1,44) для оценки математического ожидания нормально распределенного количественного признака. Определить точечную оценку математического ожидания и точность этой оценки. Что произойдет с доверительным интервалом при уменьшении (увеличении) надежности (доверительной вероятности)? Что произойдет с доверительным интервалом при уменьшении объема?

Задание 3. Дан доверительный интервал (24,7; 26,9) для оценки математического ожидания нормально распределенного количественного признака при известном среднем квадратическом отклонении генеральной совокупности. Чему будет равен этот доверительный интервал при уменьшении объема выборки в четыре раза?

Задание 4. Найти доверительный интервал для оценки с надежностью 0,99 (0,98; 0,95; 0,9) неизвестного математического ожидания a нормально распределенного признака X генеральной совокупности, если известны генеральное среднее квадратическое отклонение σ=4, выборочная средняя $x_в$=12,3 и объем выборки n=18.

Задание 5. По данным 17 независимых равноточных измерений некоторой физической величины найдены средние арифметическое результатов измерений $x_в$=36,2 и исправленное среднее квадратическое отклонение s=7. Оценить истинное значение измеряемой величины с надежностью γ=0,95 (γ=0,99, γ=0,999).

Задание 6. Из генеральной совокупности X извлечена выборка:

x_i	0,2	0,3	0,6	0,9	1,1	1,2	1,9	2,3	2,5	3,0
n_i	3	2	1	2	2	1	2	4	2	1

Вычислить выборочную среднюю и справленное среднее квадратическое отклонение. Используя исправленное среднее квадратическое отклонение, оценить с надежностью γ=0,95 (γ=0,99, γ=0,999) математическое ожидание a нормально распределенного признака генеральной совокупности с помощью доверительного интервала.

Задание 7. Найти минимальный объем выборки, при котором с надежностью 0,9796 точность оценки математического ожидания a генеральной совокупности по выборочной средней равна $\delta=0,3$, если известно среднее квадратическое отклонение $\sigma=1,2$ нормально распределенной генеральной совокупности.

3. Статистическая проверка гипотез

3.1. Статистическая гипотеза. Ошибки первого и второго рода

Одна из часто встречающихся на практике задач, связанных с применением статистических методов, состоит в решении вопроса о том, должно ли на основании данной выборки быть принято или отвергнуто некоторое предположение (гипотеза) относительно генеральной совокупности (случайной величины).

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

Задача статистической проверки гипотез заключается в следующем: относительно некоторой генеральной совокупности высказывается та или иная гипотеза H. Из этой генеральной совокупности извлекается выборка. Требуется указать правило, при помощи которого можно было бы по выборке решить вопрос о том, следует ли отклонить гипотезу H или принять ее.

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известных распределений.

Гипотеза называется *параметрической*, если в ней содержится некоторое утверждение о значении параметра известного распределения. Гипотезу, в которой сформулированы предположения относительно вида распределения, называют *непараметрической*.

Например, статистическими являются гипотезы:

1) генеральная совокупность распределена по закону Пуассона – сделано предположение о виде неизвестного распределения;

2) дисперсии двух нормальных совокупностей равны между собой ($\sigma_1^2 = \sigma_2^2$) – сделано предположение о параметрах двух известных распределений;

3) математическое ожидание $M(X)$ нормально распределенной случайной величины X равно числу m_0.

Наряду с выдвинутой гипотезой рассматривают и противоположную (противоречащая) ей гипотезу. Если выдвинутая гипотеза будет отвергнута, то имеет место противоположная (противоречащая) гипотеза. По этой причине эти гипотезы целесообразно различать.

Нулевой (основной) называют выдвинутую гипотезу H_0.

Конкурирующей (альтернативной) называют гипотезу H_1 , которая противоречит нулевой, т.е. является логическим отрицанием гипотезы H_0.

Например, если нулевая гипотеза состоит в предположении, что математическое ожидание a нормального распределения равно 10, то конкурирующая гипотеза, может состоять в предположении, что $a \neq 10$, т.е. если $H_0 : a = 10$, то $H_1 : a \neq 10$.

Различают гипотезы *простые* и *сложные*. *Простой* называют гипотезу, содержащую только одно предположение. *Сложной* называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез.

Например, гипотеза H_0 ,состоящая в предположении, что математическое ожидание случайной величины X равно a_0, т.е. $M(X) = a_0$, является простой. В качестве альтернативной гипотезы можно рассматривать одну из следующих гипотез: $H_1 : M(X) > a_0$ (сложная), $H_1 : M(X) < a_0$ (сложная), $H_1 : M(X) \neq a_0$ (сложная) или $H_1 : M(X) \neq a_1$ (простая).

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость ее проверки. Правило, по которому решают о принятии или отклонении гипотезы H_0 (соответственно отклонении или принятии H_1), называют *статистическим критерием*.

В зависимости от вида распределения критерий обозначают через U (если распределен по нормальному закону), T (если распределен по закону Стьюдента), F (если распределен по закону Фишера-Снедекора) и т.д.

В целях общности обозначим статистический критерий через K. *Наблюдаемым значением критерия* называют значение критерия, вычисленное по выборкам, т.е. $K_{набл} = K(x_1, x_2, ..., x_n)$.

После выбора критерия множество его значений разделяются на два подмножества: область S отклонения нулевой гипотезы (*критическая область*) и область $\bar{S}$ принятия этой гипотезы (*область принятия гипотезы*).

Если значение x критерия попало в область S, то нулевая гипотеза H_0 отвергается и принимается альтернативная гипотеза H_1. Если точка x попала в область $\bar{S}$, то принимается гипотеза H_0, а H_1– отвергается.

При проверке гипотезы может быть принято неправильное решение, т.е. могут быть допущены ошибки двух родов.

Ошибка первого рода состоит в том, что отвергается нулевая гипотеза H_0, когда на самом деле она верна. Таким образом, будет принята альтернативная гипотеза H_1.

Вероятность ошибки первого рода обозначают через $\alpha : \alpha = P\{H_1 / H_0\}$, где $P\{H_1 / H_0\}$ – вероятность того, что будет принята гипотеза H_1, если на самом деле верна гипотеза H_0.

Чем меньше α, тем ниже вероятность отклонить верную гипотезу. Допустимую ошибку первого рода обычно задают заранее.

Число α называют *уровнем значимости*. Обычно для α используют стандартные значения: 0,1; 0,05; 0,01 и др.

Ошибка второго рода состоит в том, что отвергается альтернативная гипотеза H_1, когда на самом деле она верна. Таким образом, будет принята основная гипотеза H_0.

Вероятность ошибки второго рода обозначают через $\beta: \beta = P\{H_0 / H_1\}$,где $P\{H_0 / H_1\}$ – вероятность того, что будет принята гипотеза H_0, если на самом деле верна гипотеза H_1.

Гипотеза H_0	**Отвергается**	**Принимается**
верна	Ошибка 1 рода	Правильное решение
неверна	Правильное решение	Ошибка 2 рода

Правильные решения также подразделяют на два вида. Если будет принята гипотеза H_0, тогда как и на самом деле в генеральной совокупности она верна, то вероятность такого решения будет равна: $1-\alpha = P\{H_0 / H_0\}$.

Может быть принята гипотеза H_1, тогда как и на самом деле она верна. Вероятность решения будет равна: $1-\beta = P\{H_1 / H_1\}$.

Число $1-\beta$ называют *мощностью критерия*. Чем больше мощность критерия, тем вероятность ошибки второго рода ниже.

Вероятности ошибок 1-го и 2-го рода определяются выбором критической области. При фиксированном объеме выборки уменьшение ошибки 1-го рода α неизбежно приводит к увеличению ошибки 2-го рода β и наоборот.

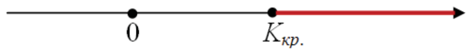
Лишь при увеличении объема выборки возможно одновременное уменьшение вероятностей α и β.

Поскольку критерий K – одномерная случайная величина, то все ее возможные значения принадлежат некоторому интервалу, поэтому критическую область и область принятия гипотезы можно *разделить точками*.

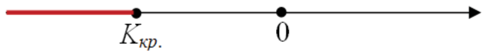
Критическими точками $K_{кр.}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Различают *одностороннюю (правостороннюю и левостороннюю)* и *двустороннюю* критические области.

Правосторонней называют критическую область, определяемую неравенством $K > K_{кр.}$, при этом $P(K > K_{кр.}) = \alpha$.



Левосторонней называют критическую область, определяемую неравенством $K < K_{кр.}$, при этом $P(K < K_{кр.}) = \alpha$.



Двусторонней называют критическую область, определяемую неравенствами $K < K_{кр1.}$ и $K > K_{кр2.}$, при этом

$$P(K > K_{кр2.}) = P(K < K_{кр1.}) = \frac{\alpha}{2}.$$

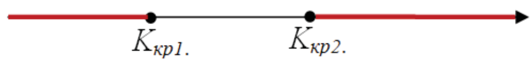


Схема проверки статистической гипотезы

1. Формулируем основную и альтернативную гипотезы: H_0 и H_1. Определяем уровень значимости α.

2. Выбираем критерий K и находим область отклонения S.

3. Делаем выборки из генеральной совокупности (проводим ряд наблюдений). Исходя из выборки, вычисляем наблюдаемое значение случайной величины K, т.е. $K_{набл}$.

4. Если $K_{набл} \in S$, то гипотезу H_0 отвергаем в пользу конкурирующей H_1 как несогласующуюся с выборкой (при этом совершаем ошибку с вероятностью α).Если $K_{набл} \in \overline{S}$, то гипотезу H_0 принимаем как согласующуюся с выборкой (при этом совершаем ошибку с вероятностью β).

Рассмотрим способы проверки некоторых наиболее часто встречающихся гипотез.

3.2. Проверка гипотезы о равенстве генеральной средней нормальной совокупности заданному числовому значению

Пусть генеральная совокупность X распределена нормально, причем имеются основания предполагать, что генеральная средняя этой совокупности $\bar{x}$ равна некоторому значению a.

Предположим, что *дисперсия генеральной совокупности* $D = \sigma^2$ *известна* (например, может быть найдена теоретически или вычислена по выборке большого объема).

Кроме того, по произведенной выборке объема n найдена выборочная средняя $\bar{x}$. Требуется по выборочной средней при заданном уровне значимости α проверить нулевую гипотезу $H_0 : \bar{x} = a$.

В качестве критерия проверки нулевой гипотезы принимают случайную величину: $U = \dfrac{(\bar{x} - a)\sqrt{n}}{\sigma}$, которая имеет нормальное распределение.

Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0 : \bar{x} = a$ нужно вычислить наблюдаемое значение критерия: $U_{набл} = \dfrac{(\bar{x}_в - a)\sqrt{n}}{\sigma}$.

1. При конкурирующей гипотезе $H_1 : \bar{x} \neq a$ критическую точку $U_{кр.}$ находим по таблице функции Лапласа из условия $\Phi(U_{кр.}) = \dfrac{1-\alpha}{2}$.

Если $|U_{набл}| < U_{кр.}$, то нет оснований отвергнуть нулевую гипотезу H_0. В противном случае нулевую гипотезу отвергают.

2. При конкурирующей гипотезе $H_1 : \bar{x} > a$ критическую точку $U_{кр.}$ правосторонней критической области находим по таблице функции Лапласа из условия $\Phi(U_{кр.}) = \dfrac{1-2\alpha}{2}$.

Если $U_{набл} < U_{кр.}$, то нет оснований отвергнуть нулевую гипотезу H_0. В противном случае нулевую гипотезу отвергают.

3. При конкурирующей гипотезе $H_1 : \bar{x} < a$ критическую точку $U_{кр.}$ левосторонней критической области находим по таблице функции Лапласа из условия $\Phi(U_{кр.}) = \dfrac{1-2\alpha}{2}$.

Если $U_{набл} > -U_{кр.}$, то нет оснований отвергнуть нулевую гипотезу H_0. В противном случае нулевую гипотезу отвергают.

▲ ***Пример.*** Из нормальной генеральной совокупности с известной дисперсией $\sigma^2 = 25$ извлечена выборка объема $n = 36$ и по ней найдено $\bar{x} = 17$. При уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу $H_0 : a = 20$ при конкурирующей гипотезе а) $H_1 : a \neq 20$, б) $H_1 : a < 20$.

□ ***Решение***. а) Вычисляем значение статистики:

$$|U_{набл}| = \frac{|\bar{x}_в - a|\sqrt{n}}{\sigma} = \frac{|17-20| \cdot 6}{5} = 3{,}6.$$

Так как критерий двусторонний, то $U_{кр.}$ находим из соотношения

$$\Phi(U_{кр.}) = \frac{1-\alpha}{2} = 0{,}475.$$ По таблице Лапласа находим $U_{кр.} = 1{,}96$.

Так как $|U_{набл}| > U_{кр.}$, то гипотеза $H_0 : a = 20$ – отклоняется.

б) Вычисляем значение статистики:

$$U_{набл} = \frac{(\bar{x}_в - a)\sqrt{n}}{\sigma} = \frac{(17-20) \cdot 6}{5} = -3{,}6\,.$$

Так как критерий левосторонний, то $U_{кр.}$ находим из соотношения

$$\Phi(U_{кр.}) = \frac{1-2\alpha}{2} = 0{,}45.$$ По таблице Лапласа находим $U_{кр.} = 1{,}64$.

Так как $U_{набл} < -U_{кр.}$ $(-3{,}6 < -1{,}64)$, то гипотеза $H_0 : a = 20$ отклоняется в пользу гипотезы $H_1 : a < 20$. ■

Предположим теперь, что *дисперсия генеральной совокупности* $D = \sigma^2$ *неизвестна*, а известна только ее исправленная выборочная оценка s^2.

При выборке небольшого объема (менее 30 наблюдений) нахождение критической точки по таблице функции Лапласа может привести к существенной погрешности.

В таком случае в качестве критерия проверки нулевой гипотезы принимают случайную величину: $T=\frac{(\bar{x}_в - a)\sqrt{n}}{s}$, которая имеет распределение Стьюдента с $(n-1)$ степенями свободы.

Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0 : \bar{x} = a$ нужно вычислить наблюдаемое значение критерия: $T_{набл} = \frac{(\bar{x}_в - a)\sqrt{n}}{s}$.

1. При конкурирующей гипотезе $H_1 : \bar{x} \neq a$ критическую точку $T_{кр.}(\alpha, n-1)$ находим по таблице критических точек распределения Стьюдента при $(n-1)$ степенях свободы и вероятности α.

Если $|T_{набл}| < T_{кр.}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу H_0 – отвергают.

2. При конкурирующей гипотезе $H_1 : \bar{x} > a$ критическую точку $T_{кр.}(2\alpha, n-1)$ правосторонней критической области находим по таблице критических точек распределения Стьюдента при $(n-1)$ степенях свободы и вероятности 2α.

Если $T_{набл} < T_{кр.}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу H_0 – отвергают.

3. При конкурирующей гипотезе $H_1 : \bar{x} < a$ критическую точку $T_{кр.}(2\alpha, n-1)$ левосторонней критической области находим по таблице критических точек распределения Стьюдента при $(n-1)$ степенях свободы и вероятности 2α.

Если $T_{набл} > -T_{кр.}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу H_0 – отвергают.

3.3. Проверка гипотезы о равенстве генеральной дисперсии нормальной совокупности заданному числовому значению

Пусть генеральная совокупность X распределена нормально, причем имеются основания предполагать, что генеральная дисперсия этой совокупности равна некоторому заданному значению σ_0^2.

Пусть по выборке объема n получена исправленная выборочная дисперсия s^2. Требуется по исправленной выборочной дисперсии при заданном уровне значимости α проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2$.

Для проверки гипотезы H_0 используют случайную величину $\dfrac{(n-1)s^2}{\sigma^2}$, которая имеет распределение χ^2 с $(n-1)$ степенями свободы.

Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2$, нужно вычислить наблюдаемое значение критерия: $\chi^2_{набл} = \dfrac{(n-1)s^2}{\sigma^2}$.

1. При конкурирующей гипотезе $H_1 : \sigma^2 \neq \sigma_0^2$ критическая область является двусторонней. Критические точки находим по таблице критических точек распределения χ^2 с $(n-1)$ степенями свободы: левую критическую точку при вероятности $1-\alpha/2$, а правую критическую точку при вероятности $\alpha/2$.

Таким образом, получим точки $\chi^2_{лев.кр}\left(1-\dfrac{\alpha}{2}, n-1\right)$, $\chi^2_{пр.кр}\left(\dfrac{\alpha}{2}, n-1\right)$.

Если $\chi^2_{лев.кр} < \chi^2_{набл} < \chi^2_{пр.кр}$,то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

2. При конкурирующей гипотезе $H_1 : \sigma^2 > \sigma_0^2$ критическую точку $\chi^2_{кр}(\alpha, n-1)$ правосторонней критической области находим по таблице критических точек распределения χ^2 с $(n-1)$ степенями свободы при

вероятности α.

Если $\chi^2_{набл} < \chi^2_{кр}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

3. При конкурирующей гипотезе $H_1 : \sigma^2 < \sigma_0^2$ критическую точку $\chi^2_{кр}(1-\alpha, n-1)$ левосторонней критической области находим по таблице критических точек распределения χ^2 с $(n-1)$ степенями свободы при вероятности $1-\alpha$.

Если $\chi^2_{набл} > \chi^2_{кр}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

▲ ***Пример.*** Из нормально распределенной совокупности извлечена выборка объема $n = 21$, для которой вычислено значение $s^2 = 10{,}3$. Требуется при уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу $H_0 : \sigma^2 = 9$ при альтернативной гипотезе $H_1 : \sigma^2 > 9$.

□ ***Решение***. Вычисляем значение статистики:

$$\chi^2_{набл} = \frac{(n-1)s^2}{\sigma^2} = \frac{20 \cdot 10{,}3}{9} \approx 22{,}9.$$

Критерий правосторонний, поэтому находим критическую точку $\chi^2_{кр}(\alpha, n-1) = 31{,}4$.

Так как $\chi^2_{набл} < \chi^2_{кр}$ $(22{,}9 < 31{,}4)$, то нет оснований отвергнуть нулевую гипотезу. ■

3.4. Проверка гипотезы о равенстве двух средних нормальных генеральных совокупностей

Пусть генеральные совокупности X_1 и X_2 имеют нормальное распределение, причем генеральные средние этих совокупностей $\bar{x}_1$ и $\bar{x}_2$ неизвестны.

По произведенным выборкам объемов n_1 и n_2 найдены выборочные средние $\bar{x}_{в1}$ и $\bar{x}_{в2}$.

Предполагаем, что дисперсии обеих генеральных совокупностей известны и равны $\sigma^2{}_1$ и $\sigma^2{}_2$ соответственно. Требуется при заданном уровне значимости α проверить нулевую гипотезу $H_0 : \bar{x}_1 = \bar{x}_2$.

В качестве критерия проверки нулевой гипотезы принимают случайную величину: $U = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}}$,

которая имеет нормальное распределение.

Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0 : \bar{x}_1 = \bar{x}_2$, нужно вычислить наблюдаемое значение критерия: $U_{набл} = \dfrac{\bar{x}_{в1} - \bar{x}_{в2}}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}}$.

1. При конкурирующей гипотезе $H_1 : \bar{x}_1 \neq \bar{x}_2$ критическую точку $U_{кр.}$ находим по таблице функции Лапласа из условия $\Phi(U_{кр.}) = \dfrac{1-\alpha}{2}$

Критическая область в этом случае является двусторонней.

Если $|U_{набл}| < U_{кр.}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

2. При конкурирующей гипотезе $H_1 : \bar{x}_1 > \bar{x}_2$ критическую точку $U_{кр.}$ правосторонней критической области находим по таблице функции Лапласа из условия $\Phi(U_{кр.}) = \dfrac{1-2\alpha}{2}$.

Если $U_{набл} < U_{кр.}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

3. При конкурирующей гипотезе $H_1 : \bar{x}_1 < \bar{x}_2$ критическую точку $U_{кр.}$ левосторонней критической области находим по таблице функции Лапласа из

условия $\Phi(U_{кр.}) = \dfrac{1-2\alpha}{2}$.

Если $U_{набл} > -U_{кр.}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

Приведенное выше правило можно также применять, если генеральные совокупности не имеют нормального распределения и их дисперсии неизвестны, а выборки имеют большой объем (не менее 30каждая) и независимы. В этом случае наблюдаемое значение критерия: $U_{набл} = \dfrac{\bar{x}_{в1} - \bar{x}_{в2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$.

Предположим теперь, что дисперсии обеих генеральных совокупностей неизвестны, а известны только их исправленные выборочные оценки s_1^2 и s_2^2, а выборки имеют небольшой объем (меньше 30). Предполагается, что дисперсии двух генеральных совокупностей одинаковы.

Если же нет оснований считать дисперсии одинаковыми, то, прежде чем сравнивать средние, следует проверить гипотезу о равенстве генеральных дисперсий, пользуясь критерием Фишера-Снедекора.

В этом случае в качестве критерия проверки нулевой гипотезы принимают случайную величину: $T = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$, где

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2},$$

имеющую распределение Стьюдента с $n_1 + n_2 - 2$ степенями свободы.

Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0: \bar{x}_1 = \bar{x}_2$, нужно вычислить наблюдаемое значение критерия: $T_{набл} = \dfrac{\bar{x}_{в1} - \bar{x}_{в2}}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$.

1. При конкурирующей гипотезе $H_1 : \bar{x}_1 \neq \bar{x}_2$ критическую точку $T_{кр.}(\alpha,\ n_1 + n_2 - 2)$ находим по таблице критических точек распределения Стьюдента при $n_1 + n_2 - 2$ степенях свободы и вероятности α.

Если $|T_{набл}| < T_{кр.}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

2. При конкурирующей гипотезе $H_1 : \bar{x}_1 > \bar{x}_2$ критическую точку $T_{кр.}(2\alpha,\ n_1 + n_2 - 2)$ правосторонней критической области находим по таблице критических точек распределения Стьюдента при $n_1 + n_2 - 2$ степенях свободы и вероятности 2α.

Если $T_{набл} < T_{кр.}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

3. При конкурирующей гипотезе $H_1 : \bar{x}_1 < \bar{x}_2$ критическую точку $T_{кр.}(2\alpha,\ n_1 + n_2 - 2)$левосторонней критической области находим по таблице критических точек распределения Стьюдента при $n_1 + n_2 - 2$ степенях свободы и вероятности 2α.

Если $T_{набл} > -T_{кр.}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

▲ ***Пример.*** Для двух независимых выборок объемов $n_1 = 6$ и $n_2 = 8$ из нормально распределенных совокупностей известны значения выборочных средних и исправленных выборочных дисперсий: $\bar{x}_1 = 6{,}58$, $\bar{x}_2 = 8{,}56$, $s_1^2 = 1{,}32$, $s_2^2 = 1{,}1$.

Требуется при уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу $H_0 : \bar{x}_1 = \bar{x}_2$ при альтернативной гипотезе $H_1 : \bar{x}_1 \neq \bar{x}_2$.

□ ***Решение***. Вычисляем значение статистики:

$$T_{набл} = \frac{\bar{x}_{в1} - \bar{x}_{в2}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},\ s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 1{,}19,\ T_{набл} = \frac{6{,}58 - 8{,}56}{1{,}1\sqrt{\frac{1}{6} + \frac{1}{8}}} = -3{,}36.$$

Значение $T_{кр.}$ находим по таблице критических точек распределения Стьюдента при $n_1+n_2-2=12$ степенях свободы и вероятности $\alpha=0{,}05$: $T_{кр.}=2{,}179$. Так как $|T_{набл}|>T_{кр.}$, то гипотеза H_0 – отвергается. ■

3.5. Проверка гипотез о законе распределения

Во многих случаях закон распределения изучаемой случайно величины неизвестен, но есть основания предположить, что он имеет вполне определенный вид: нормальный, биномиальный или какой-либо другой.

Пусть необходимо проверить гипотезу H_0 о том, что случайная величина X подчиняется определенному закону распределения, заданному функцией распределения $F_0(x)$ т.е. $H_0: F_x(x)=F_0(x)$. Под альтернативной гипотезой H_1 будем понимать в данном случае то, что просто не выполнена основная $H_1: F_x(x)\neq F_0(x)$.

Для проверки гипотезы о распределении случайной величины X проведем выборку, которую оформим в виде статистического ряда:

x_i	x_1	x_2	…	x_m
n_i	n_1	n_2	…	n_m

Требуется сделать заключение: согласуются ли результаты наблюдений с высказанным предположением. Для этого используем *критерий согласия*.

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

Существуют различные критерии согласия: Пирсона, Колмогорова, Фишера, Смирнова и др.

Критерий Пирсона (χ^2)

Критерий согласия Пирсона – наиболее часто употребляемый критерий для проверки простой гипотезы о законе распределения.

Для проверки гипотезы H_0 поступают следующим образом.

Разбивают всю область значений случайной величины X на m интервалов Δ_1, Δ_2 ,..., Δ_m . Затем, используя формулу $P\{\alpha \le X \le \beta\} = F_0(\beta) - F_0(\alpha)$, подсчитывают вероятности p_i ($i = 1,2,...,m$) попадания случайной величины X (т.е. наблюдения) в интервал Δ_i.

Теоретическое число значений случайной величины X, попавших в интервал Δ_i, рассчитывается по формуле np_i. Таким образом, имеем статистический ряд распределения случайной величины X и теоретический ряд распределения:

Δ_1	Δ_2	…	Δ_m
$n'_1 = np_1$	$n'_2 = np_2$	…	$n'_m = np_m$

Если эмпирические частоты (n_i) сильно отличаются от теоретических ($n'_i = np_i$), то проверяемую гипотезу H_0 следует отвергнуть; в противном случае – принять.

В качестве меры расхождения между эмпирическими и теоретическими частотами Пирсон предложил величину: $\chi^2 = \sum_{i=1}^{m} \frac{n_i^2}{np_i} - n$.

Согласно теореме Пирсона, при $n \to \infty$ статистика имеет χ^2 - распределение с $k = m - r - 1$ степенями свободы, где m – число групп (интервалов) выборки, r – число параметров предполагаемого распределения.

В частности, если предполагаемое распределение нормально, то оценивают два параметра (a и σ), поэтому число степеней свободы $k = m - 3$.

Правило применения критерия χ^2 сводится к следующему:

1. Вычисляем $\bar{x}_1$ – выборочное значение статистики критерия.

2. Выбрав уровень значимости α критерия, по таблице χ^2 распределения находим критическую точку.

3. Если $\chi^2_{набл} \le \chi^2_{\alpha,k}$, то гипотеза H_0 не противоречит опытным данным. Если $\chi^2_{набл} > \chi^2_{\alpha,k}$, то гипотеза H_0 – отвергается.

Необходимым условием применения критерия Пирсона является наличие в каждом из интервалов не менее 5 наблюдений. Если в отдельных интервалах их меньше, то число интервалов надо уменьшить путем объединения (укрупнения) соседних интервалов.

▲ ***Пример.*** Обследовано 100 обработанных деталей. Отклонения от заданного размера приведены в таблице:

$[x_i, x_{i+1})$	$[-3,-2)$	$[-2,-1)$	$[-1,0)$	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, 4)$	$[4, 5)$
n_i	3	10	15	24	25	13	7	3

Проверить при уровне значимости $\alpha = 0{,}01$ гипотезу H_0 о том, что отклонения от проектного размера подчиняются нормальному закону распределения.

□ ***Решение***. Число наблюдений в крайних интервалах меньше 5.Для выполнения условия применения критерия Пирсона объединим их с соседними. Получим следующий ряд распределения:

$[x_i, x_{i+1})$	$[-3,-1)$	$[-1,0)$	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, 5)$
n_i	13	15	24	25	13	10

Случайную величину (отклонение) обозначим через X . Для вычисления вероятностей необходимо вычислить параметры, определяющие нормальный закон распределения (a и σ). Их оценки вычислим по выборке:

$$\bar{x}_1 = \frac{1}{100}(-2 \cdot 13 + (-0{,}5) \cdot 15 + \ldots + 4 \cdot 10) = 0{,}885 \approx 0{,}9,$$

$$D_в = \frac{1}{100}(4 \cdot 13 + 0{,}25 \cdot 15 + \ldots + 16 \cdot 10) - (0{,}885)^2 \approx 2{,}81,$$

$$\sigma \approx 1{,}7.$$

Находим вероятности p_i . Так как случайная величина $X \sim N(a, \sigma)$ определена на всей числовой оси $(-\infty, +\infty)$, то крайние интервалы в ряде распределения заменяем, соответственно, на $(-\infty, -1)$ и $(3, +\infty)$. Тогда

$$p_1 = p\{-\infty < X < -1\} = \left\langle p(x_1 \le X \le x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right) \right\rangle =$$

$$=\Phi_0\left(\frac{-1-0{,}9}{1{,}7}\right)-\Phi_0(-\infty)=\left\langle \Phi_0(-x)=-\Phi_0(x);\Phi_0(\infty)=\frac{1}{2}\right\rangle=\Phi_0(\infty)-\Phi_0(1{,}12)=$$

$$=\frac{1}{2}-\Phi_0(1{,}12)=0{,}1314.$$

Рассуждая аналогичным образом, получаем

$p_2=0{,}1667,\ p_3=0{,}2258,\ p_4=0{,}2183,\ p_5=0{,}1503,\ p_6=0{,}1075.$

$[x_i, x_{i+1})$	$[-3,-1)$	$[-1,0)$	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, 5)$
n_i	13	15	24	25	13	10
$n_1'=np_1$	13,14	16,67	22,58	21,83	15,03	10,75

Вычисляем $\chi^2_{набл}$:

$$\chi^2_{набл}=\sum_{i=1}^{6}\frac{n_i^2}{np_i}-n=\left(\frac{13^2}{13{,}14}+\frac{15^2}{16{,}67}+\ldots+\frac{10^2}{10{,}75}\right)-100\approx 1{,}045.$$

Найдем число степеней свободы. По выборке рассчитаны два параметра, следовательно, $r=2$. Количество интервалов $m=6$. Значит, $k=6-2-1=3$.

При $\alpha=0{,}01$ и $k=3$, по таблице χ^2 распределения находим $\chi^2_{0{,}01;3}=11{,}3$.

Итак, $\chi^2_{набл}<\chi^2_{\alpha,k}$, то нет оснований отвергнуть проверяемую гипотезу.

■

Задачи для самостоятельного решения

Задание 1. а) Из нормальной генеральной совокупности с известным средним квадратическим отклонением σ=40 извлечена выборка объема n=64 и по ней найдена выборочная средняя x_B=136,5. Требуется при уровне значимости 0,01 проверить нулевую гипотезу H_0: $a=a_0$=130 при конкурирующей гипотезе H_1: $a\neq$130. б) Решить эту задачу при конкурирующей гипотезе H_1: a>130. в) Решить эту задачу при конкурирующей гипотезе H_1: a<130.

Задание 2. а) По выборке объема n=16, извлеченной из нормальной генеральной совокупности, найдены выборочная средняя x_B=118,2 и «исправленное» среднее квадратическое отклонение s=3,6. Требуется при уровне значимости 0,05 проверить нулевую гипотезу H_0: $a=a_0=120$ при конкурирующей гипотезе H_1: $a\neq120$. б) Решить эту задачу при конкурирующей гипотезе H_1: $a>120$. в) Решить эту задачу при конкурирующей гипотезе H_1: $a<120$.

Задание 3. Получены следующие результаты:

x_i	34,8	34,9	35	35,1	35,3
n_i	2	3	4	6	5

Выборочная средняя $x_B=35$. Требуется вычислить исправленное среднее квадратическое отклонение и при уровне значимости 0,05 проверить нулевую гипотезу H_0: $a=a_0=35$ при конкурирующей гипотезе H_1: $a\neq35$.

Задание 4. Менеджер кредитного отдела нефтяной компании хотел бы выяснить, является ли среднемесячный баланс владельцев кредитных карточек равным 76 у.е. Аудитор случайным образом отобрал 100 счетов и нашел, что среднемесячный баланс владельцев составил 83,4 у.е. с выборочным исправленным отклонением, равным 23,65 у.е. Определить на 5% уровне значимости, может ли этот аудитор утверждать, что средний баланс отличен от 76 у.е.

Задание 5. Средний доход фирмы в день составлял 1030 единиц. После реорганизации выборочный средний доход в день за 30 рабочих дней составил 1060 единиц с исправленным средним квадратическим отклонением 90 единиц. Можно ли утверждать (уровень значимости 5%), что реорганизация привела к увеличению среднего дохода?

Задание 6. Средний дневной объем продаж в магазине составлял 510 единиц. После реорганизации выборочный средний дневной объем продаж за 25 рабочих дней составил 525 единиц с исправленным средним квадратическим отклонением 40 единиц. Можно ли утверждать (уровень значимости 10%), что реорганизация привела к увеличению среднего дохода?

Литература

1. Астафьев, Е.Р. Основы теории вероятностей и математической статистики [Текст]: учеб.-метод. пособие / Е.Р. Астафьев, Е.В. Михайленко. – Краснодар: Краснодарская академия МВД России, 2004.

2. Выск, Н.Д. Теория вероятностей и математическая статистика [Текст]: учеб.пособие / Н.Д. Выск. – М.: МАТИ-РГТУ им. К.Э. Циолковского, 2011.

3. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и статистике [Текст]: учеб. пособие для вузов / В.Е. Гмурман. – М.: Высшая школа, 2002.

4. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст]: учеб.пособие / В.Е. Гмурман. – 12-е изд. – М.: Юрайт, 2013.

5. Долгова, В.Н. Статистика [Текст]: учебник и практикум для бакалавров / В.Н. Долгова, Т.Ю. Медведева. – 2-е изд., перераб. и доп. – М.: Юрайт, 2014.\Кремер Н.Ш. Теория вероятностей и математическая статистика [Текст]: учеб.пособие для вузов / Н.Ш. Кремер. – 2-е изд., переб. и доп. – М.: ЮНИТИ-ДАНА, 2004.

6. Математическая статистика [Текст]: Методические разработки и контрольные задания / Авт.-сост.: С.В. Плотникова. Тамбов: Изд-во Тамб. гос. техн. ун-та, 2005.

7. Михайленко, Е.В. Теория вероятностей и математическая статистика [Текст]: учеб.пособие / Е.В. Михайленко, М.А. Жукова, Ф.Ф. Бараненко. – Краснодар: КрУ МВД России, 2011.

8. Письменный, Д.Т. Конспект лекций по теории вероятностей, математической статистике и случайным процессам [Текст] / Д.Т. Письменный. – 3-е изд. – М.: Айрис-пресс, 2008.

9. Статистика [Текст]: учебник / ред. И.И. Елисеева. – М.: Юрайт, 2012.

ПРИЛОЖЕНИЕ 1

ТАБЛИЦА ЗНАЧЕНИЙ ФУНКЦИИ $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^X e^{-\frac{z^2}{2}}dz$

0,00	0,0000	0,41	0,1591	0,82	0,2939	1,23	0,3907
0,01	0,0040	0,42	0,1628	0,83	0,2967	1,24	0,3925
0,02	0,0080	0,43	0,1664	0,84	0,2995	1,25	0,3944
0,03	0,0120	0,44	0,1700	0,85	0,3023	1,26	0,3962
0,04	0,0160	0,45	0,1736	0,86	0,3051	1,27	0,3980
0,05	0,0199	0,46	0,1772	0,87	0,3078	1,28	0,3997
0,06	0,0239	0,47	0,1808	0,88	0,3106	1,29	0,4015
0,07	0,0279	0,48	0,1844	0,89	0,3133	1,30	0,4032
0,08	0,0319	0,49	0,1879	0,90	0,3159	1,31	0,4049
0,09	0,0359	0,50	0,1915	0,91	0,3186	1,32	0,4066
0,10	0,0398	0,51	0,1950	0,92	0,3212	1,33	0,4082
0,11	0,0438	0,52	0,1985	0,93	0,3238	1,34	0,4099
0,12	0,0478	0,53	0,2019	0,94	0,3264	1,35	0,4115
0,13	0,0517	0,54	0,2054	0,95	0,3289	1,36	0,4131
0,14	0,0557	0,55	0,2088	0,96	0,3315	1,37	0,4147
0,15	0,0596	0,56	0,2123	0,97	0,3340	1,38	0,4162
0,16	0,0636	0,57	0,2157	0,98	0,3365	1,39	0,4177
0,17	0,0675	0,58	0,2190	0,99	0,3389	1,40	0,4192
0,18	0,0714	0,59	0,2224	1,00	0,3413	1,41	0,4207
0,19	0,0753	0,60	0,2257	1,01	0,3438	1,42	0,4222
0,20	0,0793	0,61	0,2291	1,02	0,3461	1,43	0,4236
0,21	0,0832	0,62	0,2324	1,03	0,3485	1,44	0,4251
0,22	0,0871	0,63	0,2357	1,04	0,3508	1,45	0,4265
0,23	0,0910	0,64	0,2389	1,05	0,3531	1,46	0,4279
0,24	0,0948	0,65	0,2422	1,06	0,3554	1,47	0,4292
0,25	0,0987	0,66	0,2454	1,07	0,3577	1,48	0,4306
0,26	0,1026	0,67	0,2486	1,08	0,3599	1,49	0,4319
0,27	0,1064	0,68	0,2517	1,09	0,3621	1,50	0,4332
0,28	0,1103	0,69	0,2549	1,10	0,3643	1,51	0,4345
0,29	0,1141	0,70	0,2580	1,11	0,3665	1,52	0,4357
0,30	0,1179	0,71	0,2611	1,12	0,3686	1,53	0,4370
0,31	0,1217	0,72	0,2642	1,13	0,3708	1,54	0,4382
0,32	0,1255	0,73	0,2673	1,14	0,3729	1,55	0,4394
0,33	0,1293	0,74	0,2703	1,15	0,3749	1,56	0,4406
0,34	0,1331	0,75	0,2734	1,16	0,3770	1,57	0,4418
0,35	0,1368	0,76	0,2764	1,17	0,3790	1,58	0,4429
0,36	0,1406	0,77	0,2794	1,18	0,3810	1,59	0,4441
0,37	0,1443	0,78	0,2823	1,19	0,3830	1,60	0,4452
0,38	0,1480	0,79	0,2852	1,20	0,3849	1,61	0,4463
0,39	0,1517	0,80	0,2881	1,21	0,3869	1,62	0,4474
0,40	0,1554	0,81	0,2910	1,22	0,3883	1,63	0,4484

Продолжение приложения 1

1,64	0,4495	1,88	0,4699	2,24	0,4875	2,72	0,4967
1,65	0,4505	1,89	0,4706	2,26	0,4881	2,74	0,4969
1,66	0,4515	1,90	0,4713	2,28	0,4887	2,76	0,4971
1,67	0,4525	1,91	0,4719	2,30	0,4893	2,78	0,4973
1,68	0,4535	1,92	0,4726	2,32	0,4898	2,80	0,4974
1,69	0,4545	1,93	0,4732	2,34	0,4904	2,82	0,4976
1,70	0,4554	1,94	0,4738	2,36	0,4909	2,84	0,4977
1,71	0,4564	1,95	0,4744	2,38	0,4913	2,86	0,4979
1,72	0,4573	1,96	0,4750	2,40	0,4918	2,88	0,4980
1,73	0,4582	1,97	0,4756	2,42	0,4922	2,90	0,4981
1,74	0,4591	1,98	0,4761	2,44	0,4927	2,92	0,4982
1,75	0,4599	1,99	0,4767	2,46	0,4931	2,94	0,4984
1,76	0,4608	2,00	0,4772	2,48	0,4934	2,96	0,4985
1,77	0,4616	2,02	0,4783	2,50	0,4938	2,98	0,4986
1,78	0,4625	2,04	0,4793	2,52	0,4941	3,00	0,49865
1,79	0,4633	2,06	0,4803	2,54	0,4945	3,20	0,49931
1,80	0,4641	2,08	0,4812	2,56	0,4948	3,40	0,49966
1,81	0,4649	2,10	0,4821	2,58	0,4951	3,50	0,49975
1,82	0,4656	2,12	0,4830	2,60	0,4953	3,60	0,499841
1,83	0,4664	2,14	0,4838	2,62	0,4956	3,80	0,499928
1,84	0,4671	2,16	0,4846	2,64	0,4959	4,00	0,499968
1,85	0,4678	2,18	0,4854	2,66	0,4961	4,50	0,499997
1,86	0,4686	2,20	0,4861	2,68	0,4963	5,00	0,499997
1,87	0,4693	2,22	0,4868	2,70	0,4965		

ПРИЛОЖЕНИЕ 2

ТАБЛИЦА ЗНАЧЕНИЙ $t_\gamma = t(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		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				

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	11,05
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

Учебное издание

Старостенко Игорь Николаевич
Хромых Анна Алексеевна

ЭЛЕМЕНТЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ

Учебное пособие

В авторской редакции

ISBN 978-5-9266-1613-9

Подписано в печать 04.08.2020. Формат 60x84 1/16.
Усл. печ. л. 4,3. Тираж 100 экз. Заказ 72.

Краснодарский университет МВД России.
350005, г. Краснодар, ул. Ярославская, 128.