Министерство сельского хозяйства РФ
Департамент научно-технологической политики и образования
ФГОУ ВПО
Волгоградская государственная сельскохозяйственная академия
Кафедра высшей математики

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. РЕШЕНИЕ ЗАДАЧ ПО ТЕМЕ: «ТОЧЕЧНЫЕ И ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ»

Методическая разработка

Волгоград
ИПК «Нива»
2010

УДК 519.2(075.3)
ББК 22.17я72
К 67

*Рекомендовано к изданию кафедрой высшей математики ВГСХА
и учебно-методическим советом факультета электрификации с.х.*

Корниенко, В.С.
К 67 Математическая статистика. Решение задач по теме: «Точечные и интервальные оценки параметров генеральной совокупности». Методическая разработка [Текст] /В.С. Корниенко; Волгогр. гос. с.-х. акад. Волгоград, 2010. 32 с.

Приведены необходимые теоретические сведения. Решено (в «ручном» режиме и с помощью Mathcad) достаточное число задач на нахождение точечных и интервальных оценок параметров.
Для студентов специальности 110302 - «Электрификация и автоматизация сельского хозяйства».

УДК 519.2(0.75.3)
ББК 22.17я72

Содержание

1. Теоретические сведения

1.1. Предмет математической статистики. Генеральная и выборочная совокупность

Роль статистики обусловлена тем, что статистические представления являются важнейшей составляющей интеллектуального багажа современного человека. Они нужны и для повседневной жизни в современном цивилизованном обществе, и для продолжения образования практически во всех сферах человеческой деятельности, например, таких, как социология, экономика, право, медицина, демография.

На современном этапе развития общества, когда в нашу жизнь стремительно вошли референдумы и социологические опросы, кредиты и страховые полисы, разнообразные банковские начисления и т.п., становится очевидной актуальность включения в курс математики материала вероятностно-статистического характера.

Принципиально важным представляется то обстоятельство, что при решении задач математической статистики студентам приходится выполнять довольно много математических расчетов, часто прибегая к подсчетам на калькуляторе или компьютере. Таким образом, изучение этого раздела математики даст серьезный импульс для совершенствования вычислительных умений студентов, развития алгоритмического мышления.

Основная задача математической статистики – это разработка методов получения научно доказанных выводов о массовых явлениях и процессах на основе статистических данных, полученных в результате наблюдений и экспериментов.

Полученные таким образом выводы и заключения представляют собой утверждения об общих вероятностных характеристиках изучаемого процесса или явления, т.е. о вероятностях, законах распределения, математических ожиданиях, дисперсиях и т.д., а не относятся к отдельным испытаниям, из повторения которых складывается данное массовое явление.

В математической статистике рассматриваются две основные категории задач: оценивание и статистическая проверка гипотез.

Первая задача математической статистики подразделяется на точечное оценивание и интервальное оценивание параметров распределения.

Вторая задача математической статистики заключается в том, что делается предположение о законе распределения вероятностей некоторой случайной величины или о параметрах известного распределения вероятностей и осуществляется проверка данного предположения.

При проведении статистического исследования часто задача состоит в изучении группы однородных объектов, явлений или процессов относительно некоторого качественного или количественного признака, характеризующего эти объекты.

При решении данной задачи можно провести сплошное обследование, т.е. обследовать каждый из объектов данной совокупности относительно изучаемого признака. Однако на практике подобное обследование зачастую связано с большими материальными затратами, поэтому в большинстве случаев из всей совокупности объектов отбирается определенное количество объектов, которые подвергаются изучению.

Таким образом, бывают «сплошное исследование» и «выборочное исследование». Обратим внимание на причины, по которым сплошное исследование заменяется выборочным. Во-первых, проведение сплошного исследования часто требует больших организационных усилий и финансовых затрат. Так, перепись населения страны, которая является ярким примером сплошного исследования, связана с подготовкой разнообразной документации, выделением и инструктажем переписчиков, сбором информации, обработкой собранных сведений. Кроме того, в ряде случаев осуществить сплошное исследование просто невозможно, так как это привело бы к порче или уничтожению продукции. Например, при исследовании продолжительности горения партии электроламп, выпущенных заводом, невозможно проверить всю партию, поскольку это просто привело бы к ее уничтожению.

Генеральная совокупность – это совокупность объектов, явлений или процессов, из которых производится выборка.

Выборочная совокупность [***n*-выборка**] – это совокупность n случайно отобранных объектов из генеральной совокупности.

Объем совокупности – это число объектов генеральной совокупности.

Повторная выборка – это выборка, при которой отобранным случайным образом объект обязательно возвращается в генеральную совокупность перед отбором следующего объекта.

Бесповторная выборка – это выборка, при которой отобранный случайным образом объект больше в генеральную совокупность не возвращается.

Одним из основных требований к формированию выборочных совокупностей является требование **репрезентативности** [представительности] выборки, т.е. для характеристики по данным выборочной совокупности интересующего исследователей признака генеральной совокупности необходимо, чтобы единицы выборки в достаточной степени обладали этим признаком.

Пример. Определить, является ли репрезентативной выборка:

1) число автомобильных аварий в июне, если необходимо составить статистический отчет по авариям в городе за год;

2) городские жители при подсчете числа автомобилей на душу населения в стране;

3) люди в возрасте от 40 до 50 лет при выяснении рейтинга молодежной телепрограммы.

Р е ш е н и е. 1) Выборка не является репрезентативной. Летом нет снега и наледи на дорогах, а это одна из основных причин аварий.

2) Выборка не является репрезентативной. Понятно, что в городе машин намного больше, чем в сельских районах. Это необходимо учитывать.

3) Выборка не является репрезентативной. Люди в возрасте от 40 до 50 лет едва ли проявят интерес к программе, ориентированной на молодежную аудиторию. При использовании такой выборки рейтинг может сильно упасть, но это не отразит реального положения вещей.

Для формирования выборочной совокупности применяются различные способы отбора.

Статистические данные должны быть представлены так, чтобы ими можно было пользоваться.

1.2. Статистическое распределение выборки. Эмпирическая функция распределения

Пусть дана некоторая генеральная совокупность. Из данной генеральной совокупности извлечена выборка, причем в ходе отбора значение x_1 наблюдалось m_1 раз, x_2 - m_2 раз, x_k - m_k раз.

$$\sum_{i=1}^{k} m_i = n,$$

где n – объем выборки.

Наблюдаемые значения x_i, которые были получены в результате отбора, называются **вариантами**. Число m_i показывает, сколько раз данное значение x_i встречалось в ходе отбора, и называется **частотой варианты** x_i. Если записать последовательность вариант в возрастающем или убывающем порядке и соответствующие им частоты, то получится таблица, называемая **дискретным вариационным рядом:**

Значения	x_1	x_2	…	x_k
Частоты	m_1	m_2	…	m_k

Отношение частоты m_i к объему выборки n называется **относительной частотой:**

$$\frac{m_i}{n} = w_i. \qquad (1)$$

Статистическое распределение выборки – это последовательность вариант и соответствующих им частот или относительных частот, т.е. дискретный вариационный ряд как раз и характеризует статистическое распределение выборки.

Если промежуток между наименьшим и наибольшим значениями признака в выборке разбить на несколько интервалов одинаковой длины и каждому интервалу поставить в соответствие число выборочных значений признака, попавших в этот интервал, то получим **интервальный вариа-**

ционный ряд. Статистическое распределение выборки может быть также задано интервальным вариационным рядом.

Предположим, что статистическое распределение частот количественного признака X известно заранее. Введем следующие обозначения:

m_x - число наблюдений, при которых наблюдается значение признака, меньшее x;

n – объем выборки.

Относительная частота появления события $X < x$ равна $\frac{m_x}{n}$. В том случае, если значение признака x будет изменяться, то будет изменяться и относительная частота, следовательно, относительную частоту $\frac{m_x}{n}$ можно рассматривать как функцию от x. Данная функция называется эмпирической, потому что была получена эмпирическим путем (в результате исследований).

Эмпирическая функция распределения [**функция распределения *n*-выборки**] – это функция $F_n(x)$, которая для каждого значения x определяет относительную частоту события $X < x$:

$$F_n(x) = \frac{m_x}{n}. \qquad (2)$$

Следовательно, для того чтобы найти $F_n(x_5)$, необходимо число вариант, меньших x_5, разделить на объем выборки n:

$$F_n(x_5) = \frac{m_{x_5}}{n}.$$

Интегральная функция $F(x)$ распределения генеральной совокупности называется **теоретической функцией распределения** (в отличие от эмпирической функции распределения выборочной совокупности).

Они отличаются тем, что теоретическая функция $F(x)$ характеризует вероятность события $X < x$, а эмпирическая функция $F_n(x)$ характеризует относительную частоту события $X < x$.

В то же время $F_n(x)$ обладает всеми свойствами функции $F(x)$.

Если n-выборка случайной величины X задана в виде интервального вариационного ряда,

Интервалы	$[a_1, a_2)$	$[a_2, a_3)$	$\ldots$	$[a_k, a_{k+1}$
Частоты	m_1	m_2	$\ldots$	m_k

где $\sum_{i=1}^{k} m_i = n$, то в этом случае эмпирическая функция распределения $F_n(x)$ определяется только на концах интервала. В этом случае ее можно изобразить ломаной, проходящей через точки $(a_i; F_n(a_i))$, $i=\overline{1,k}$. При этом $F_n(a_1)=0$, $F_n(a_{k+1})=1$.

Пример 1. Пусть выборка случайной величины X задана дискретным вариационным рядом (табл. 1).

Таблица 1

x_i	37	38	39	40	41	42	43	44
m_i	1	3	5	8	12	9	5	2

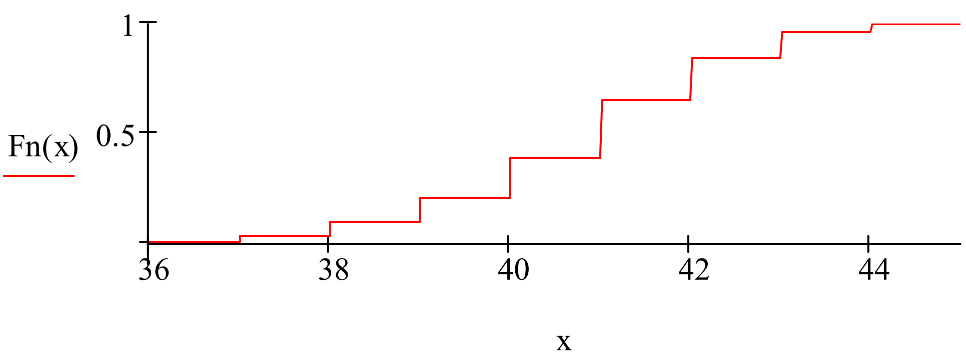
По данным этой табл. найдем накопленные частоты и накопленные относительные частоты (табл. 2).

Таблица 2

x_i	37	38	39	40	41	42	43	44	45
m_{x_i}	0	1	4	9	17	29	38	43	45
w_{x_i}	0	0,022	0,089	0,2	0,378	0,644	0,844	0,956	1

Определим эмпирическую функцию $F_n(x)$ и построим ее график.

$$\mathrm{Fn}(x) := \begin{cases} 0 \ \text{ if } \ x \le 37 \\ 0.022 \ \text{ if } \ 37 < x \le 38 \\ 0.089 \ \text{ if } \ 38 < x \le 39 \\ 0.2 \ \text{ if } \ 39 < x \le 40 \\ 0.378 \ \text{ if } \ 40 < x \le 41 \\ 0.644 \ \text{ if } \ 41 < x \le 42 \\ 0.844 \ \text{ if } \ 42 < x \le 43 \\ 0.956 \ \text{ if } \ 43 < x \le 44 \\ 1 \ \text{ if } \ x > 44 \end{cases}$$



Приведем другое решение этого примера в Mathcad.

$ORIGIN := 1$

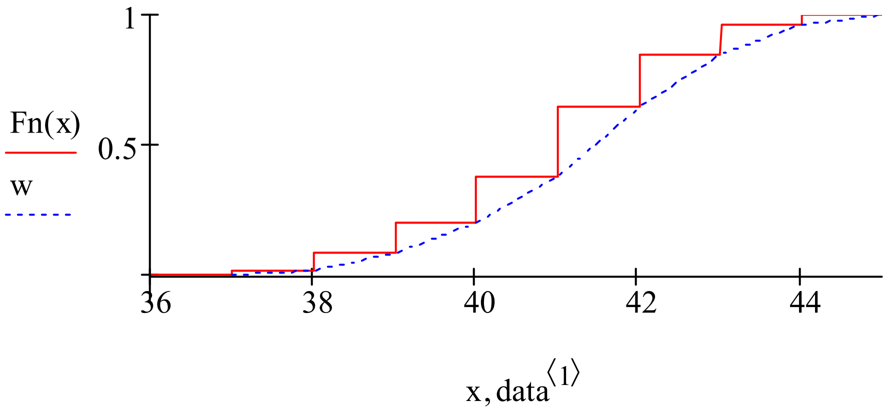
$$data := \begin{pmatrix} 37 & 38 & 39 & 40 & 41 & 42 & 43 & 44 & 45 \\ 1 & 3 & 5 & 8 & 12 & 9 & 5 & 2 & 45 \end{pmatrix} \qquad data := data^T$$

$$n := \sum data^{\langle 2 \rangle} - 45$$

$$k := rows(data) \qquad m := data^{\langle 2 \rangle} \qquad mx_1 := 0$$

$$s := 2..k \qquad mx_s := \sum_{i=1}^{s-1} m_{i,1} \qquad w := \frac{mx}{n}$$

$$mx = \begin{pmatrix} 0 \\ 1 \\ 4 \\ 9 \\ 17 \\ 29 \\ 38 \\ 43 \\ 45 \end{pmatrix} \qquad w = \begin{pmatrix} 0 \\ 0.022 \\ 0.089 \\ 0.2 \\ 0.378 \\ 0.644 \\ 0.844 \\ 0.956 \\ 1 \end{pmatrix} \qquad Fn(x) := \left| \begin{array}{ll} w_{1,1} & \text{if } \ x \le 37 \\ w_{2,1} & \text{if } \ 37 < x \le 38 \\ w_{3,1} & \text{if } \ 38 < x \le 39 \\ w_{4,1} & \text{if } \ 39 < x \le 40 \\ w_{5,1} & \text{if } \ 40 < x \le 41 \\ w_{6,1} & \text{if } \ 41 < x \le 42 \\ w_{7,1} & \text{if } \ 42 < x \le 43 \\ w_{8,1} & \text{if } \ 43 < x \le 44 \\ w_{9,1} & \text{if } \ x > 44 \end{array} \right.$$



Важно! Пунктирная кривая называется **кумулятивной кривой** [**кумулятой**].

Пример 2. Пусть выборка случайной величины X задана интервальным вариационным рядом (табл. 1).

Таблица 1

x_i	[-1,75; -1,25)	[-1,25; -0,75)	[-0,75; -0,25)	[-0,25; 0,25)
m_i	5	8	9	12
x_i	[0,25; 0,75)	[0,75; 1,25)	[1,25; 1,75)	
m_i	9	3	4	

Для построения эмпирической функции распределения $F_n(x)$ вычислим накопленные частоты w_{a_i} (табл. 2).

Таблица 2

a_i	-1,75	-1,25	-0,75	-0,25	0,25	0,75	1,25	1,75
w_{a_i}	0	0,1	0,26	0,44	0,68	0,86	0,92	1

Воспользовавшись Mathcad, определим эмпирическую функцию распределения $F_n(x)$ и построим ее график.

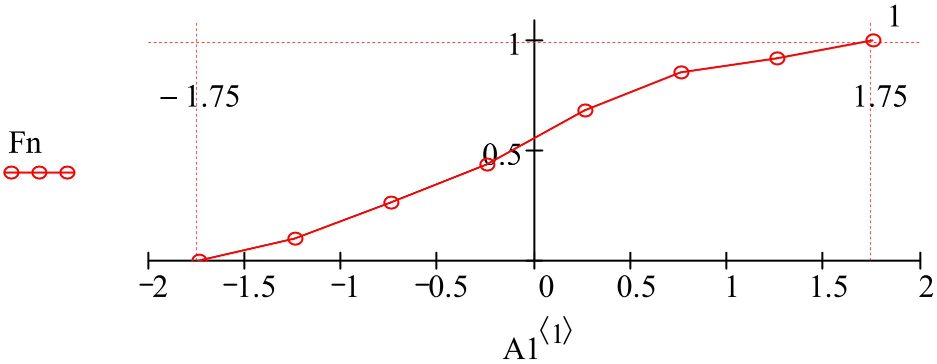
ORIGIN := 1

$$A1 := \begin{pmatrix} -1.75 & -1.25 & 5 \\ -1.25 & -0.75 & 8 \\ -0.75 & -0.25 & 9 \\ -0.25 & 0.25 & 12 \end{pmatrix} \qquad A2 := \begin{pmatrix} 0.25 & 0.75 & 9 \\ 0.75 & 1.25 & 3 \\ 1.25 & 1.75 & 4 \end{pmatrix} \qquad A := \mathrm{stack}(A1, A2)$$

$$a := (1.75 \quad 1.75 \quad 0) \qquad A1 := \mathrm{stack}(A, a)$$

$$k := \mathrm{rows}(A1) \qquad m := A1^{\langle 3 \rangle} \qquad n := \sum m$$

$$Fn_1 := 0 \qquad s := 2..k \qquad Fn_s := \frac{1}{n} \cdot \sum_{i=1}^{s-1} m_{i,1} \qquad Fn = \begin{pmatrix} 0 \\ 0.1 \\ 0.26 \\ 0.44 \\ 0.68 \\ 0.86 \\ 0.92 \\ 1 \end{pmatrix}$$



1.3. Статистические оценки параметров распределения. Требования к статистическим оценкам

Большинство практических задач, которые решает статистика, состоит в оценивании некоторого количественного признака генеральной совокупности. Предположим, что исследователю удалось установить, какому именно закону распределения подчиняется изучаемый количественный признак. В этом случае необходимо оценить параметры, которыми определяется предполагаемое распределение. Например, если удалось установить, что количественный признак подчиняется показательному закону распределения вероятностей, тогда необходимо оценить параметр λ, которым определяется данное распределение.

Предположим, что имеются данные выборки, например, значения количественного признака x_1, x_2, …, x_n, полученные в результате n наблюдений. Будем рассматривать x_1, x_2, …, x_n как независимые случайные величины X_1, X_2, …, X_n.

Статистическая оценка неизвестного параметра теоретического распределения – это функция от наблюдаемых случайных величин.

Таким образом, определить статистическую оценку неизвестного параметра теоретического распределения, значит, определить функцию от наблюдаемых случайных величин X_1, X_2, …, X_n, которая дает приближенное значение оцениваемого параметра.

Для того чтобы статистические оценки $\overline{\theta}_i$ можно было бы принять за оценки параметров θ_i, необходимо и достаточно, чтобы оценки $\overline{\theta}_i$ удовлетворяли трем статистическим свойствам: несмещенности, состоятельности и эффективности.

$\overline{\theta}_i$ называется **несмещенной оценкой** для параметра θ_i, если ее выборочное математическое ожидание равно оцениваемому параметру генеральной совокупности.

$$M\left(\overline{\theta}_i\right) = \theta_i; \quad M\left(\overline{\theta}_i\right) - \theta_i = \varphi_i,$$

где φ_i - это смещение оценки.

Смещенная оценка – это оценка параметра, чье математическое ожидание не равно оцениваемому параметру, т.е.

$$M\left(\overline{\theta}_i\right) \neq \theta_i \text{ или } \varphi_i \neq 0.$$

$\overline{\theta}_i$ является **состоятельной оценкой** для параметра θ_i, если она удовлетворяет закону больших чисел. Закон больших чисел гласит о том, что с увеличением выборки значение оценки $\overline{\theta}_i$ стремится к значению параметра θ_i генеральной совокупности:

$$P\left(\left|\overline{\theta}_i - \theta_i\right| < \varepsilon\right) \to 1 \text{ при } n \to \infty.$$

Это же условие можно записать с помощью теоремы ***Бернулли:***

$$\bar{\theta}_i \xrightarrow{P} \theta_i \text{ при } n \to \infty,$$

т.е. значение оценки $\bar{\theta}_i$ сходится по вероятности к значению параметра θ_i генеральной совокупности при условии, что объем выборки n стремится к бесконечности.

Для определения состоятельности оценки достаточно выполнения двух условий:

1) $\varphi_i = 0$ или $\varphi_i \to 0$ при $n \to \infty$ - смещение оценки равно нулю или стремится к нему при объеме выборки, стремящемся к бесконечности;

2) $D(\bar{\theta}_i) \to 0$ при $n \to \infty$ - дисперсия оценки параметра θ_i стремится к нулю при объеме выборки, стремящемся к бесконечности.

$\bar{\theta}_i$ является **эффективной оценкой** для параметра θ_i, если статистическая оценка $\bar{\theta}_i$ имеет наименьшую возможную дисперсию при заданном объеме выборки n .

1.4. Точечные оценки параметров распределения

Рассмотрим повторную выборку $x = (x_1 \quad x_2 \quad \ldots \quad x_n)$ значений генеральной совокупности X. Пусть $M(X) = a$, $D(X) = \sigma^2$ генеральные средняя и дисперсия совокупности. В качестве оценок для a и σ рассмотрим **выборочную среднюю** [среднюю арифметическую выборки]

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (\text{в Mathcad } \bar{x} = mean(x)) \tag{1}$$

и **выборочную дисперсию**

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \quad (\text{в Mathcad } S^2 = \mathrm{var}(x)). \tag{2}$$

Доказано, что $\bar{x}$ является несмещенной, состоятельной и эффективной оценкой для a, причем

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}. \tag{3}$$

Доказано, что $M\left(S^2\right) = \frac{n-1}{n} D(X)$, т.е. оценка S^2 является смещенной. На практике, чтобы избавиться от этого недостатка, для оценки неизвестной дисперсии генеральной совокупности пользуются **исправленной несмещенной оценкой**

$$\overline{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \quad (\text{в Mathcad } \overline{S^2} = Var(x)). \tag{4}$$

Тем не менее, из закона больших чисел следует, что как оценка S^2, так и $\overline{S^2}$ являются состоятельными оценками для σ^2.

Для бесповторной выборки оценки (1), (4) так же являются несмещенными и состоятельными, а дисперсия $\sigma_{\bar{x}}^2$ равна

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right), \tag{5}$$

где N - объем генеральной совокупности. При $N \to \infty$ бесповторная выборка неотличима от повторной и формула (5) переходит в формулу (3).

Пусть генеральная совокупность содержит M элементов, обладающих некоторым признаком A. Тогда **генеральной долей признака** A называется величина

$$p = \frac{M}{N}. \tag{6}$$

Для доли p несмещенной и состоятельной оценкой будет **выборочная доля**

$$w = \frac{m}{n}, \tag{7}$$

где m – число элементов выборки, обладающих признаком A.

Дисперсия σ_w^2 в случае повторной выборки равна

$$\sigma_w^2 = \frac{pq}{n}, \tag{8}$$

а в случае бесповторной выборки

$$\sigma_w^2 = \frac{pq}{n}\left(1 - \frac{n}{N}\right), \tag{9}$$

где $q = 1 - p$.

Если n намного меньше N, то повторная выборка практически не отличается от бесповторной и формулы (8) и (9) дают одинаковый результат. Если же $n = N$, то объем выборки равен объему генеральной совокупности и выборочная доля равна генеральной, тогда $\sigma_w^2 = 0$.

Для наглядности и удобства пользования сведем данные о точечных оценках параметров генеральной совокупности в табл. 1.

Т а б л и ц а 1. Точечные оценки параметров генеральной свовокупности

Параметр	Оценка	Дисперсия для повторной выборки	Дисперсия для бесповторной выборки
a	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\frac{\sigma^2}{n}$	$\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)$
σ^2	$\overline{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$	-	-
p	$\frac{m}{n}$	$\frac{p(1-p)}{n}$	$\frac{p(1-p)}{n}\left(1 - \frac{n}{N}\right)$

Важно! При больших n ($n > 30$) неизвестные параметры в формулах для дисперсии можно заменить на их выборочные значения без особой потери точности.

1.5. Метод максимального правдоподобия

Рассмотрим один из методов получения точечных оценок параметров генеральной совокупности. Этот метод основан на понятии функции правдоподобия.

Для непрерывной случайной величины *функцией правдоподобия* называется плотность вероятности совместного появления результатов выборки $x = (x_1 \quad x_2 \quad \dots \quad x_n)$:

$$L(x_1, x_2, ..., x_n, \theta) = P(x_1, \theta) \cdot P(x_2, \theta) \cdot ... \cdot P(x_n, \theta). \tag{1}$$

Для дискретной случайной величины в этой формуле $P(x_i, \theta)$ означает вероятность появления значений x_i. В качестве оценки неизвестного параметра θ принимается такое значение $\overline{\theta}_n$, которое максимизирует функцию L.

Функция L является *мерой правдоподобия* полученных наблюдений x_1, x_2, …, x_n. Поэтому естественно выбирать оценку $\overline{\theta}_n$ так, чтобы имеющиеся наблюдения были наиболее правдоподобны.

Необходимым условием максимума функции L по θ является равенство

$$\frac{\partial L}{\partial \theta} = 0. \tag{2}$$

Так как максимумы функций L и $\ln(L)$ достигаются в одной и той же точке, то часто вместо уравнения (2) рассматривают уравнение

$$\frac{\partial Ln(L)}{\partial \theta} = 0. \tag{3}$$

Если оцениваемых параметров θ^1, θ^2, …, θ^k несколько, т.е.

$$L = L(x_1, x_{2,} ..., x_n, \theta^1, ..., \theta^k),$$

то для определения их оценок $\overline{\theta_n^1}$, …, $\overline{\theta_n^k}$ решают систему уравнений

$$\frac{\partial L}{\partial \theta^i} = 0, \quad i = \overline{1,k}. \tag{4}$$

Пример 1. Найдем оценку для вероятности P наступления события A по данному числу *m* появления этого события в *n* испытаниях.

В этом случае функция правдоподобия L равна

$$L = C_n^m P^m (1-P)^{n-m}.$$

Тогда

$$\ln(L) = \ln\left(C_n^m\right) + m\ln(P) + (n-m)\ln(1-P).$$

Уравнение для определения оценки:

$$\frac{\partial \ln(L)}{\partial P} = 0 \Rightarrow \frac{m}{P} - \frac{n-m}{1-P} = 0 \Rightarrow P = \frac{m}{n} = w.$$

Значит, оценкой методом правдоподобия вероятности наступления события будет его относительная частота

$$\overline{P}_n = \frac{m}{n}.$$

Пример 2. Пусть $x = (x_1 \quad x_2 \quad \dots \quad x_n)$ - выборка из нормального распределения $N(a,\sigma)$. Определить параметры a и σ методом наибольшего правдоподобия.

Р е ш е н и е. Составим функцию правдоподобия:

$$L = \frac{1}{\left(\sqrt{2\pi}\right)^n \sigma^n} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i - a)^2 \right).$$

Тогда

$$\ln(L) = \frac{-n}{2} \ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - a)^2 .$$

Рассмотрим уравнения для определения максимума функции $\ln(L)$.

$$\frac{\partial \ln(L)}{\partial a} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - a) = 0 ,$$

$$\frac{\partial \ln(L)}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - a)^2 = 0 .$$

Из первого уравнения находим

$$\sum_{i=1}^{n} x_i - na = 0 \;\Rightarrow\; a = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x} .$$

Из второго уравнения –

$$\sigma^2 n = \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 \;\Rightarrow\; \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 = S^2 .$$

Таким образом,

$$\overline{a} = \overline{x} , \quad \overline{\sigma^2} = S^2 .$$

1.6. Основные статистические распределения

Для решения статистических задач используются специальные распределения случайных величин.

Наиболее важным в статистических исследованиях является нормальное распределение. Рассмотрим несколько случайных величин, сконструированных на основе нормального распределения, которые наиболее часто встречаются в математической статистике.

1) Распределение χ_n^2 (хи-квадрат с n степенями свободы).

Если $\xi_i \in N(0,1)$, $i = \overline{1,n}$, - независимые стандартные нормальные случайные величины, то говорят, что случайная величина

$$\chi_n^2 = \sum_{i=1}^{n} \xi_i^2 \qquad (1)$$

имеет **распределение хи-квадрат с n степенями свободы**.

2) Распределение Стьюдента t_n.

Пусть ξ_0, ξ_1, $\xi_2, \dots$, ξ_n - независимые стандартные нормальные случайные величины. Тогда случайная величина

$$t_n = \frac{\xi_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i^2}} \qquad (2)$$

имеет **распределение Стьюдента с n степенями свободы**.

3) Распределение Фишера.

Пусть

$$\xi_i \in N(0,1), \quad i = \overline{1,m}; \qquad \eta_j \in N(0,1), \quad j = \overline{1,n},$$

независимые случайные величины. Тогда случайная величина

$$F_{m,n} = \frac{\frac{1}{m}\sum_{i=1}^{m}\xi_i^2}{\frac{1}{n}\sum_{j=1}^{n}\eta_j^2} \qquad (3)$$

распределена по закону Фишера со степенями свободы *m*, *n*.

Распределения случайных величин χ_n^2, t_n, $F_{m,n}$ имеются в Mathcad.

1.7. Интервальное оценивание

Точечная оценка параметра – это оценка, которая характеризуется одним конкретным числом (например, математическим ожиданием, дисперсией, средним квадратичным отклонением и т.д.). Точечные оценки параметров генеральной совокупности могут быть приняты в качестве ориентировочных, первоначальных результатов обработки выборочных данных. Их основной недостаток заключается в том, что неизвестно, с какой точностью оценивается параметр. Если для выборок большого объема точность обычно бывает достаточной (при условии несмещенности, эффективности и состоятельности оценок), то для выборок небольшого объема вопрос точности становится очень важным. По этой причине при небольшом объеме выборки следует пользоваться интервальными оценками.

Интервальные оценки позволяют построить с заданной вероятностью интервал, в котором находится оцениваемый параметр генеральной совокупности. Таким образом, интервальные оценки характеризуются двумя числами – концами интервалов.

Предположим, что статистическая характеристика $\bar{\theta}$, рассчитанная по данным выборки, является выборочной оценкой неизвестного параметра θ генеральной совокупности, причем θ - это постоянное число. Тем точнее оценка $\bar{\theta}$ характеризует параметр θ, чем меньше абсолютная величина разности $|\theta - \bar{\theta}|$. Если $|\theta - \bar{\theta}| < \Delta$, где $\Delta > 0$, следовательно, чем меньше Δ, тем точнее оценка. Таким образом, **точность оценки** $\bar{\theta}$ параметра θ характеризует положительное число Δ.

Важно! Нельзя утверждать, что оценка $\bar{\theta}$ абсолютно удовлетворяет неравенству $|\theta - \bar{\theta}| < \Delta$. Однако можно задать вероятность γ, с которой это неравенство осуществляется.

Надежность [**доверительная вероятность**] оценки θ по $\overline{\theta}$ - это вероятность γ, с которой осуществляется неравенство $|\theta - \overline{\theta}| < \Delta$. Принято надежность оценки задавать перед процессом оценивания параметра генеральной совокупности. В качестве вероятности γ берут число, близкое к единице (например, 0,95; 0,99; 0,999).

Предположим, что γ – это вероятность того, что $|\theta - \overline{\theta}| < \Delta$, т.е.

$$P\left(\left|\theta - \overline{\theta}\right| < \Delta\right) = \gamma .$$

Заменим неравенство $|\theta - \overline{\theta}| < \Delta$ равносильным ему двойным неравенством $-\Delta < \theta - \overline{\theta} < \Delta$ или $\overline{\theta} - \Delta < \theta < \overline{\theta} + \Delta$.

В результате имеем

$$P\left(\overline{\theta} - \Delta < \theta < \overline{\theta} + \Delta\right) = \gamma .$$

Вероятностный смысл данного отношения таков: вероятность того, что интервал $\left(\overline{\theta} - \Delta; \overline{\theta} + \Delta\right)$ заключает в себе [покроет] неизвестный параметр θ, равна γ.

Доверительный интервал – это интервал $\left(\overline{\theta} - \Delta; \overline{\theta} + \Delta\right)$, который покрывает оцениваемый параметр θ с заданной надежностью γ.

Границы $\overline{\theta} - \Delta$ и $\overline{\theta} + \Delta$ называются **доверительными границами** интервала. Они определяются на основе выборочных данных и являются функциями от случайных величин x_1, x_2, …, x_n, и, следовательно, сами являются случайными величинами. Поэтому доверительный интервал $\left(\overline{\theta} - \Delta; \overline{\theta} + \Delta\right)$ тоже можно охарактеризовать как случайный, т.е. он может покрывать параметр θ или нет. Именно в таком смысле следует понимать случайное событие, состоящее в том, что доверительный интервал $\left(\overline{\theta} - \Delta; \overline{\theta} + \Delta\right)$ покрывает оцениваемый параметр θ генеральной совокупности.

1.7.1. Доверительные интервалы для генеральной средней и генеральной доли признака

Пусть $x = (x_1, x_2, ..., x_n)$- n-выборка из генеральной совокупности объема N, $\overline{x} = mean(x)$ - выборочное среднее, $\overline{S^2} = \overline{D} = Var(x)$ - выборочная (несмещенная) дисперсия, $\overline{S} = \sqrt{\overline{S^2}} = \overline{\sigma} = \sqrt{Var(x)}$ - выборочное среднее квадратичное отклонение, $w = \dfrac{m}{n}$ - выборочная доля признака.

Доверительный интервал уровня надежности γ для генеральной средней a имеет вид:

$$\overline{x} - \Delta < a < \overline{x} + \Delta , \qquad (1)$$

где Δ - предельная ошибка, зависящая от доверительной вероятности γ.

При $n > 30$ для повторной выборки

$$\Delta = t\frac{\overline{S}}{\sqrt{n}}, \tag{2}$$

а для бесповторной выборки

$$\Delta = t\frac{\overline{S}}{\sqrt{n}}\sqrt{1-\frac{n}{N}}. \tag{3}$$

Здесь t определяется из условия

$$\Phi(t) = \gamma, \tag{4}$$

где $\Phi(t)$ - функция Лапласа. В Mathcad

$$t = qnorm\left(1-\frac{\alpha}{2},0,1\right).$$

Если $n \le 30$, то доверительный интервал для a строится только для нормальной генеральной совокупности и для повторной выборки

$$\Delta = t_{n-1}\frac{\overline{S}}{\sqrt{n}}. \tag{5}$$

Здесь t_{n-1} определяется из условия

$$P(|\,\xi\,| < t_{n-1}) = \gamma, \tag{6}$$

где случайная величина ξ имеет распределение Стьюдента с $n-1$ степенями свободы и t_{n-1} находится по таблицам распределения Стьюдента.

В Mathcad

$$t_{n-1} = qt\left(1-\frac{\alpha}{2},n-1\right).$$

Для бесповторной выборки

$$\Delta = t_{n-1}\frac{\overline{S}}{\sqrt{n}}\left(1-\frac{n}{N}\right). \tag{7}$$

Доверительный интервал для генеральной доли P равен $(w-\Delta, w+\Delta)$, т.е.

$$w-\Delta < P < w+\Delta, \tag{8}$$

где при $n > 30$ для повторной выборки

$$\Delta = t\sqrt{\frac{w(1-w)}{n}}, \tag{9}$$

а для бесповторной выборки

$$\Delta = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}, \tag{10}$$

где t определяется равенством (4).

При этом должны выполняться условия: $nw \ge 5$ и $n(1-w) \ge 5$, где $w = \frac{m}{n}$ - выборочная доля объектов, обладающая определенным свойством.

Важно! При $n \le 30$ рассматриваются только выборки из нормальной совокупности. Для повторной выборки Δ определяется по формуле (9), а для бесповторной – по формуле (10).

Для удобства пользования приведенные выше формулы сведем в таблицу:

Параметр	**Оценка**	**Предельная ошибка выборки**			
		Повторная выборка		**Бесповторная выборка**	
		$n > 30$	$n \le 30$	$n > 30$	$n \le 30$
a	$\overline{x} = mean(x)$	$\Delta = t\dfrac{\overline{S}}{\sqrt{n}}$	$\Delta = t_{n-1}\dfrac{\overline{S}}{\sqrt{n}}$	$\Delta = t\dfrac{\overline{S}}{\sqrt{n}}\sqrt{1-\dfrac{n}{N}}$	$\Delta = t_{n-1}\dfrac{\overline{S}}{\sqrt{n}}\sqrt{1-\dfrac{n}{N}}$
P	$w = \dfrac{m}{n}$	$\Delta = t\sqrt{\dfrac{w(1-w)}{n}}$		$\Delta = t\sqrt{\dfrac{w(1-w)}{n}\left(1-\dfrac{n}{N}\right)}$	
		$t = qnorm\left(1-\dfrac{\alpha}{2},0,1\right)$ $\quad t_{n-1} = qt\left(1-\dfrac{\alpha}{2},n-1\right)$ $\quad \overline{S} = \sqrt{\overline{S^2}}$			

1.7.2. Доверительный интервал для дисперсии

Если $x = (x_1 \quad x_2 \quad \dots \quad x_n)$- n-выборка из нормальной совокупности $N(a,\sigma)$, где a и σ неизвестны, то статистика

$$\psi = \frac{n\overline{S^2}}{\sigma^2}$$

имеет распределение χ^2_{n-1} (хи-квадрат с $n-1$ степенями свободы). Тогда доверительный интервал для σ^2 можно найти из соотношения

$$P\left(\frac{n\overline{S^2}}{z_2} < \sigma^2 < \frac{n\overline{S^2}}{z_1}\right) = \gamma, \tag{11}$$

где z_1 и z_2 определяются из условия

$$P\left(z_1 < \chi^2_{n-1} < z_2\right) = \gamma .$$

Обычно z_1 и z_2 выбирают таким образом, чтобы

$$P\left(\chi^2_{n-} < z_1\right) = P\left(\chi^2_{n-1} > z_2\right) = \frac{1-\gamma}{2} = \frac{\alpha}{2},$$

тогда

$$P\left(\chi^2_{n-1} > z_1\right) = \frac{1+\gamma}{2} = 1-\frac{\alpha}{2}; \quad P\left(\chi^2_{n-1} > z_2\right) = \frac{1-\gamma}{2} = \frac{\alpha}{2}. \tag{12}$$

В Mathcad

$$z_1 = qchisq\left(1-\frac{\alpha}{2}, n-1\right); \quad z_2 = qchisq\left(\frac{\alpha}{2}, n-1\right).$$

1.8. Объем выборки

При решении статистических задач часто требуется определить необходимый объем выборки для достижения требуемой надежности доверительного интервала. Объем выборки n определяется уровнем надежности γ и предельной ошибкой Δ. Из приведенных выше формул найдем объемы выборок для различных случаев и занесем полученные формулы в таблицу:

Параметр	**Объем выборки**	
	Повторная выборка	**Бесповторная выборка**
a	$n = \dfrac{t^2 \overline{S^2}}{\Delta^2}$	$n = \dfrac{Nt^2 \overline{S^2}}{N\Delta^2 + t^2 \overline{S^2}}$
P	$n = \dfrac{w(1-w)t^2}{\Delta^2}$	$n = \dfrac{Nt^2 w(1-w)}{N\Delta^2 + t^2 w(1-w)}$

Замечание. Если $N \to \infty$ (бесконечный объем генеральной совокупности), то бесповторная выборка не отличается от повторной и формулы для бесповторной выборки в таблицах подпункта 1.7.1 и п. 1.8 переходят в формулы для повторной выборки.

2. Задачи

Задача 1. Из 1500 деталей отобрано 250, распределение которых по размеру задано в таблице

Размер детали	7,8-8,0	8,0-8,2	8,2-8,4	8,4-8,6	8,6-8,8	8,8-9,0
Количество деталей	5	20	80	95	40	10

Найти точечные оценки $\bar{x}$, $\overline{S^2}$ для среднего и несмещенной дисперсии, а также дисперсию $\sigma^2_{\bar{x}}$ оценки $\bar{x}$ для повторного и бесповторного отбора.

Р е ш е н и е.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{250}(7{,}9\cdot 5 + 8{,}1\cdot 20 + 8{,}3\cdot 80 + 8{,}5\cdot 95 + 8{,}7\cdot 40 + 8{,}9\cdot 10) = 8{,}44;$$

$$\overline{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{1}{249}[(7{,}9-8{,}44)^2\cdot 5 + (8{,}1-8{,}44)^2\cdot 20 +$$

$$(8{,}3-8{,}44)^2\cdot 80 + (8{,}5-8{,}44)^2\cdot 95 + (8{,}7-8{,}44)^2\cdot 40 + (8{,}9-8{,}44)^2\cdot 10] = 0{,}042.$$

Для повторной выборки

$$\sigma^2_{\bar{x}} = \frac{\overline{S^2}}{n} = \frac{0{,}042}{250} = 0{,}00016;$$

для бесповторной

$$\sigma^2_{\bar{x}} = \frac{\overline{S^2}}{n}\left(1-\frac{n}{N}\right) = 0{,}00016\left(1-\frac{250}{1500}\right) = 0{,}00013.$$

Mathcad-контроль:

ORIGIN := 1

$$xv := \begin{pmatrix} 7.9 & 8.1 & 8.3 & 8.5 & 8.7 & 8.9 \\ 5 & 20 & 80 & 95 & 40 & 10 \end{pmatrix} \qquad xv := xv^{T} \qquad n := \sum xv^{\langle 2\rangle}$$

$$xs := \frac{xv^{\langle 1\rangle}\cdot xv^{\langle 2\rangle}}{n} \qquad xs = 8.44$$

$$S2 := \frac{1}{n-1}\cdot xv^{\langle 2\rangle}\cdot\left(xv^{\langle 1\rangle} - xs\right)^2 \qquad S2 = 0.042$$

$$\sigma 12 := \frac{S2}{n} \qquad \sigma 12 = 1.687\times 10^{-4}$$

$$N := 1500 \qquad \sigma 22 := \frac{S2}{n}\cdot\left(1-\frac{n}{N}\right) \qquad \sigma 22 = 1.406\times 10^{-4}$$

Задача 2. Выборочно обследовали партию кирпича, поступившего на стройку. Из 100 проб в 12 случаях кирпич оказался бракованным. Найти точечную оценку $\overline{p}$ доли бракованного кирпича и $\sigma_{\overline{p}}$.

Р е ш е н и е. Согласно формуле (8) п. 1.4 имеем

$$m = 12, \quad n = 100, \quad \overline{p} = \frac{m}{n} = 0{,}12.$$

$$\sigma_{\overline{p}}^2 = \frac{\overline{p}\left(1 - \overline{p}\right)}{n} = \frac{0{,}12 \cdot 0{,}88}{100} = 0{,}001056; \quad \sigma_{\overline{p}} = \sqrt{\sigma_{\overline{p}}^2} = 0{,}031.$$

Mathcad-контроль:

$m := 12 \qquad n := 100 \qquad p1 := \frac{m}{n} \qquad \sigma p12 := \frac{p1 \cdot (1 - p1)}{n}$

$\sigma p12 = 1.056 \times 10^{-3} \qquad \sigma p1 := \sqrt{\sigma p12} \qquad \sigma p1 = 0.032$

Задача 3. Для определения среднего процентного содержания белка в зернах пшеницы было отобрано 625 зерен, обследование которых показало, что выборочное среднее равно $\overline{x} = 16{,}8$, а выборочная несмещенная дисперсия $\overline{S^2} = 4$. Чему равна предельная ошибка выборки? Доверительная вероятность $\gamma = 98{,}8$ %.

Р е ш е н и е. Считаем, что объем генеральной совокупности бесконечен, тогда выборку можно считать повторной. Так как $\gamma = 0{,}988$, то по таблицам функции $\Phi(x)$ определим t из условия $\Phi(t) = 0{,}988$; $t = 2{,}5$. Тогда (см. таблицу подпункта 1.7.1) предельная ошибка выборки равна

$$\Delta = t\frac{\overline{S}}{\sqrt{n}} = 2{,}5 \cdot \frac{2}{\sqrt{625}} = 0{,}2.$$

Mathcad-контроль:

$\gamma := 0.988 \qquad t := \mathrm{qnorm}(\gamma, 0, 1) \qquad t = 2.257$

$S2 := 4 \qquad S := \sqrt{S2} \qquad n := 625 \qquad \Delta := t \cdot \frac{S}{\sqrt{n}} \qquad \Delta = 0.181$

Задача 4. Выборочное среднее квадратичное отклонение десяти измерений некоторой величины равно 10 см. Найти предельную ошибку выборки. Доверительная вероятность $\gamma = 60$ %.

Р е ш е н и е. Здесь $n = 10 < 30$ и выборка повторная. По таблицам распределения Стьюдента находим для $\gamma = 0{,}6$ $\quad t_9 = 0{,}9$. Согласно таблице подпункта 1.7.1

$$\Delta = \frac{t_9 \overline{S}}{\sqrt{n}} = \frac{0{,}9 \cdot 10}{\sqrt{10}} = 2{,}846.$$

Mathcad-контроль:

$$n := 10 \qquad \gamma := 0.6 \qquad S := 10 \qquad \alpha := 1 - \gamma \qquad t := qt\left(1 - \frac{\alpha}{2}, n - 1\right)$$

$$t = 0.883 \qquad \Delta := t\cdot\frac{S}{\sqrt{n}} \qquad \Delta = 2.794$$

Задача 5. Среди стандартных изделий одной фабрики в среднем 15 % относится ко второму сорту. С какой вероятностью можно утверждать, что процент p изделий второго сорта среди 1000 стандартных изделий данной фабрики отличается от 15 % не более чем на 2 %?

Р е ш е н и е. Здесь $n = 1000$, $w = 0{,}15$, $\Delta = 0{,}02$. Из формулы (9) подпункта 1.7.1 получаем

$$t = \Delta \cdot \sqrt{\frac{n}{w(1-w)}} = 0{,}2 \cdot \sqrt{\frac{1000}{0{,}15 \cdot 0{,}85}} = 1{,}771.$$

Тогда

$$P(|\, p - 0{,}15 \,| \le 0{,}02) = \Phi(1{,}77) = 0{,}96233.$$

Mathcad-контроль:

$$n := 1000 \qquad w := 0.15 \qquad \Delta := 0.02 \qquad t := \Delta\cdot\sqrt{\frac{n}{w\cdot(1-w)}}$$

$$t = 1.771 \qquad P := pnorm(t, 0, 1) \qquad P = 0.962$$

Задача 6. Из партии в 5000 электрических ламп было отобрано 300 по схеме бесповторной выборки. Средняя продолжительность горения ламп в выборке оказалась равной 1450 часам, а дисперсия $\overline{S^2} = 4000$. Найти доверительный интервал для среднего срока горения лампы. Доверительная вероятность $\gamma = 99{,}96$ %.

Р е ш е н и е. Для $\gamma = 0{,}9996$ по таблицам находим $\Phi(t) = 0{,}9996$; $t = 3{,}57$. При $n > 30$ для бесповторной выборки (см. таблицу подпункта 1.7.1)

$$\Delta = \frac{3{,}57 \cdot \sqrt{4000}}{\sqrt{300}} \sqrt{1 - \frac{300}{5000}} \approx 12{,}997 = 13.$$

Тогда искомый доверительный интервал

$$1437 < a < 1463.$$

Mathcad-контроль:

$$\gamma := 0.9996 \qquad \alpha := 1 - \gamma \qquad t := qnorm\left(1 - \frac{\alpha}{2}, 0, 1\right) \qquad t = 3.54$$

$$S2 := 4000 \qquad n := 300 \qquad N := 50000 \qquad a := 1450 \qquad S := \sqrt{S2}$$

$$\Delta := t\cdot\frac{S}{\sqrt{n}}\cdot\left(1 - \frac{n}{N}\right) \qquad \Delta = 12.849$$

$a - \Delta = 1.437 \times 10^3$ $\qquad$ $a + \Delta = 1.463 \times 10^3$

Задача 7. Службой контроля проверен расход энергии в течение месяца в 10 квартирах 70-квартирного дома, в результате чего были получены значения (кВт·ч): 125, 78, 102, 140, 90, 45, 50, 125, 115, 112.

Определить доверительный интервал для оценки среднего расхода электроэнергии в доме. Доверительная вероятность $\gamma = 95$ %.

Р е ш е н и е. Считаем выборку бесповторной, тогда (при $n = 10 < 30$) предельная ошибка выборки по формуле (7) подпункта 1.7.1 равна

$$\Delta = t_{n-1} \frac{\overline{S}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}.$$

Здесь $n = 10$, $N = 70$. При $\gamma = 0{,}95$ получим $t_9 = 2{,}26$.

Находим

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = 98{,}2$$

и

$$\overline{S^2} = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 = 1033{,}29; \quad \overline{S} = \sqrt{\overline{S^2}} = 32{,}145.$$

Тогда

$$\Delta = 2{,}26 \cdot \frac{32{,}145}{\sqrt{10}} \sqrt{1 - \frac{10}{70}} = 21{,}27.$$

Получили доверительный интервал

$$\overline{x} - \Delta < a < \overline{x} + \Delta \;\Rightarrow\; 76{,}93 < a < 119{,}47.$$

Mathcad-контроль:

$x := (\,125 \;\; 78 \;\; 102 \;\; 140 \;\; 90 \;\; 45 \;\; 50 \;\; 125 \;\; 115 \;\; 112\,)$ $\qquad$ $n := \mathrm{cols}(x)$

$N := 70$ $\qquad$ $\gamma := 0.95$ $\qquad$ $\alpha := 1 - \gamma$ $\qquad$ $t := \mathrm{qt}\left(1 - \dfrac{\alpha}{2}, n - 1\right)$

$S := \sqrt{\mathrm{Var}(x)}$ $\qquad$ $S = 32.145$ $\qquad$ $\Delta := t \cdot \dfrac{S}{\sqrt{n}} \cdot \sqrt{1 - \dfrac{n}{N}}$ $\qquad$ $xs := \mathrm{mean}(x)$

$xs = 98.2$ $\qquad$ $xs - \Delta = 76.911$ $\qquad$ $xs + \Delta = 119.489$

Задача 8. В партии, содержащей 5000 изделий, проверено 400. Среди них оказалось 300 изделий высшего сорта. Найти доверительный интервал для доли изделий высшего сорта в случаях повторной и бесповторной выборок. Доверительная вероятность $\gamma = 95$ %.

Р е ш е н и е. Для $\gamma = 0{,}95$ определим t из условия $\Phi(t) = \gamma$. Получим $t = 1{,}96$. При $m = 300$, $n = 400$, $N = 5000$, найдем оценку для доли p, $w = \dfrac{m}{n} = \dfrac{3}{4}$.

а) Для случая повторной выборки предельная ошибка равна

$$\Delta = t\cdot\sqrt{\frac{w(1-w)}{n}} = 1{,}96\cdot\sqrt{\frac{0{,}75\cdot 0{,}25}{400}} = 0{,}0424\,.$$

Тогда доверительный интервал

$$0{,}75 - 0{,}0424 < p < 0{,}75 + 0{,}0424 \;\Rightarrow\; 0{,}7076 < p < 0{,}7924\,.$$

б) Для бесповторной выборки (см. формулу (10) подпункта 1.7.1)

$$\Delta = t\cdot\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} = 0{,}0424\cdot\sqrt{1-\frac{400}{5000}} = 0{,}0407$$

и доверительный интервал

$$0{,}7093 < p < 0{,}7907\,.$$

Mathcad-контроль:

$$m := 300 \qquad n := 400 \qquad N := 5000 \qquad w := \frac{m}{n} \qquad \gamma := 0.95$$

$$\alpha := 1-\gamma \qquad t := \mathrm{qnorm}\left(1-\frac{\alpha}{2},0,1\right) \qquad t = 1.96$$

$$\Delta := t\cdot\sqrt{\frac{w\cdot(1-w)}{n}} \qquad \Delta = 0.042 \qquad w-\Delta = 0.708 \qquad w+\Delta = 0.792$$

$$\Delta := \Delta\cdot\sqrt{1-\frac{n}{N}} \qquad \Delta = 0.041$$

Задача 9. Найти объемы повторной и бесповторной выборок из 10000 банок консервов для определения доли банок, не соответствующих стандарту. Предполагается, что предельная ошибка выборки не превосходит 0,05. Доверительная вероятность $\gamma = 99{,}95$ %.

Р е ш е н и е. По условию задачи $N = 10000$, $\Delta = 0{,}05$, $\gamma = 0{,}9995$.

Из условия $\Phi(t) = 0{,}9995$ находим $t = 3{,}5$.

Для повторной выборки получим (см. таблицу п. 1.8)

$$n = \frac{w(1-w)t^2}{\Delta^2}\,.$$

Так как w неизвестно, то выберем его таким, чтобы значение выражения $w(1-w)$ было максимальным, т.е. положим $w = 0{,}5$. Тогда завышенное значение *n* будет равно $n = 4900\cdot 0{,}5\cdot 0{,}5 = 1225$.

Для бесповторной выборки (см. таблицу п. 1.8)

$$n = \frac{Nt^2 w(1-w)}{Nt^2 + t^2 w(1-w)}\,.$$

Очевидно, и в этом случае наибольшее значение выражения $w(1-w)$ соответствует максимальному *n*. Положим $w = 0{,}5$; тогда

$$n = \frac{10000\cdot 3{,}5^2\cdot 0{,}5\cdot 0{,}5}{10000\cdot 0{,}05^2 + 3{,}5^2\cdot 0{,}5\cdot 0{,}5} = 1081\,.$$

Mathcad-контроль:

$N := 10000 \qquad \Delta := 0.05 \qquad \gamma := 0.9995 \qquad \alpha := 1 - \gamma$

$t := \mathrm{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \qquad t = 3.481 \qquad n(w) := \frac{w\cdot(1-w)\cdot t^2}{\Delta^2}$

$w := 0.5 \qquad n(w) = 1.212 \times 10^3$

$n(w) := \frac{N\cdot t^2\cdot w\cdot(1-w)}{N\cdot\Delta^2 + t^2\cdot w\cdot(1-w)} \qquad n(w) = 1.081 \times 10^3$

Задача 10. Объем генеральной совокупности равен $N = 1000$. Каков должен быть объем выборки, чтобы предельная ошибка Δ выборки была равна $0{,}1\sigma$? Рассмотреть случаи повторной и бесповторной выборок. Доверительная вероятность $\gamma = 95$ %.

Р е ш е н и е. При $\gamma = 0{,}5$ получим $t = 1{,}96$. Для повторной выборки (заменяя $\overline{S}$ на σ) получим (см. таблицу п. 1.8)

$$n = \frac{t^2\overline{S^2}}{\Delta^2} = \frac{1{,}96^2\cdot\sigma^2}{(0{,}1\sigma)^2} = \frac{1{,}96^2}{0{,}01} = 384\,.$$

Для бесповторной выборки

$$n = \frac{Nt^2\overline{S^2}}{N\Delta^2 + t^2\overline{S^2}} = \frac{1000\cdot 1{,}96^2\cdot\sigma^2}{1000\cdot(0{,}1\sigma)^2 + 1{,}96^2\sigma^2} = \frac{1000\cdot 1{,}96^2}{1000\cdot 0{,}01 + 1{,}96^2} = 277\,.$$

Mathcad-контроль:

$N := 1000 \qquad \gamma := 0.95 \qquad t := 1.96 \qquad \Delta(\sigma) := 0.1\cdot\sigma$

$n(\sigma) := \frac{t^2\cdot\sigma^2}{\Delta(\sigma)^2} \qquad n(\sigma) \to 384.16000000000000000$

$n(\sigma) := \frac{N\cdot t^2\cdot\sigma^2}{N\cdot\Delta(\sigma)^2 + t^2\cdot\sigma^2} \qquad n(\sigma) \to 277.54016876661657612$

Задача 11. Признак X генеральной совокупности распределен нормально. Имеется выборка, данные которой приведены в таблице:

x_i	0,1	0,2	0,3	0,4	0,5
m_i	2	4	7	6	1

Найти доверительный интервал, накрывающий среднее квадратичное отклонение с доверительной вероятностью 0,99.

Р е ш е н и е. Вычислим

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{k}x_i m_i=\frac{1}{20}\left(0{,}1\cdot 2+0{,}2\cdot 4+0{,}3\cdot 7+0{,}4\cdot 6+0{,}5\cdot 1\right)=0{,}3;$$

$$\overline{S^2}=\frac{1}{19}[(0{,}1-0{,}3)^2\cdot 2+(0{,}2-0{,}3)^2\cdot 4+(0{,}4-0{,}3)^2\cdot 6+(0{,}5-0{,}3)^2]=0{,}01158.$$

При $n=20$, $\gamma=0{,}99$, используя формулы (12) подпункта 1.7.2, по таблицам распределения χ^2_{19} находим $z_1=6{,}8$; $z_2=38{,}5$. Тогда из формулы (11) подпункта 1.7.2 получим

$$\frac{n\overline{S^2}}{z_2}=\frac{20\cdot 0{,}01158}{38{,}5}=0{,}006;\quad \frac{n\overline{S^2}}{z_1}=\frac{20\cdot 0{,}01158}{6{,}8}=0{,}034.$$

Доверительный интервал для дисперсии будет

$$0{,}006<\sigma^2<0{,}034,$$

а для среднего квадратичного отклонения

$$0{,}077<\sigma<0{,}184.$$

Mathcad-контроль:

ORIGIN := 1

$$\mathrm{xv}:=\begin{pmatrix}0.1 & 0.2 & 0.3 & 0.4 & 0.5\\ 2 & 4 & 7 & 6 & 1\end{pmatrix}\qquad \mathrm{xv}:=\mathrm{xv}^{T}\qquad \mathrm{n}:=\sum \mathrm{xv}^{\langle 2\rangle}$$

$$\mathrm{xs}:=\frac{\mathrm{xv}^{\langle 1\rangle}\cdot \mathrm{xv}^{\langle 2\rangle}}{\mathrm{n}}\qquad \mathrm{xs}=0.3$$

$$\mathrm{S2}:=\frac{1}{\mathrm{n}-1}\cdot \mathrm{xv}^{\langle 2\rangle}\cdot\left(\mathrm{xv}^{\langle 1\rangle}-\mathrm{xs}\right)^2\qquad \mathrm{S2}=0.012$$

$$\gamma:=0.99\qquad \alpha:=1-\gamma$$

$$\mathrm{z1}:=\mathrm{qchisq}\left(1-\frac{\alpha}{2},\mathrm{n}-1\right)\qquad \mathrm{z2}:=\mathrm{qchisq}\left(\frac{\alpha}{2},\mathrm{n}-1\right)$$

$$\mathrm{z1}=38.582\qquad \mathrm{z2}=6.844$$

$$\frac{\mathrm{n}\cdot \mathrm{S2}}{\mathrm{z1}}=6.002\times 10^{-3}\qquad \frac{\mathrm{n}\cdot \mathrm{S2}}{\mathrm{z2}}=0.034$$

$$\sqrt{\frac{\mathrm{n}\cdot \mathrm{S2}}{\mathrm{z1}}}=0.077\qquad \sqrt{\frac{\mathrm{n}\cdot \mathrm{S2}}{\mathrm{z2}}}=0.184$$

Задача 12. Автомат, работающий со стандартным отклонением $S=5$ г, фасует чай в пачки. Проведена случайная выборка объемом $n=30$ пачек. Средний вес пачки в выборке 101 г. Найти доверительный интервал для среднего веса пачки чая в генеральной совокупности. Считать выборку повторной. Доверительная вероятность $\gamma = 95$ %.

Р е ш е н и е в Mathcad.

$\gamma := 0.95 \qquad \alpha := 1 - \gamma \qquad t := \text{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \qquad t = 1.96$

$S := 5 \qquad n := 30 \qquad \Delta := t \cdot \frac{S}{\sqrt{n}} \qquad \Delta = 1.789 \qquad xs := 101$

$xs - \Delta = 99.211 \qquad xs + \Delta = 102.789$

Задача 13. (Решить самостоятельно). Автомат фасует чай в пачки. Проведена случайная выборка объемом $n = 30$ пачек. Средний вес пачки чая в выборке 101 г, выборочное стандартное отклонение $S = 4$ г. Определить доверительный интервал для среднего веса пачки чая в генеральной совокупности. Доверительная вероятность $\gamma = 95$ %.

Ответ: искомый интервал (99,133; 102,867).

Задача 14. Проведена выборка объема $n = 2000$ шт. 150 из них оказались бракованными. Найти доверительный интервал доли бракованных изделий в генеральной совокупности. Доверительная вероятность $\gamma = 95$ %.

Р е ш е н и е. В данном случае $n = 2000$, $m = 150$. Тогда

$$w = \frac{m}{n} = \frac{150}{2000} = 0{,}075 \;\Rightarrow\; nw = m = 150 > 5;$$

$$n(1 - w) = 2000 \cdot (1 - 0{,}075) = 1850 > 5.$$

Оба условия выполнены.

$\gamma = 0{,}95 \;\Rightarrow\; \alpha = 1 - \gamma = 0{,}05$. Значит $\frac{\alpha}{2} = 0{,}025$; поэтому $t = 1{,}96$.

Тогда $\Delta = t \cdot \sqrt{\frac{w(1-w)}{n}} = 0{,}012$. Отсюда

$$w - \Delta = 0{,}075 - 0{,}012 = 0{,}063; \quad w + \Delta = 0{,}075 + 0{,}012 = 0{,}087,$$

т.е. искомый интервал (0,063; 0,087).

Mathcad-контроль:

$\gamma := 0.95 \qquad \alpha := 1 - \gamma \qquad t := \text{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \qquad t = 1.96$

$n := 2000 \qquad m := 150 \qquad w := \frac{m}{n}$

$\Delta := t \cdot \sqrt{\frac{w \cdot (1 - w)}{n}} \qquad \Delta = 0.012 \qquad w - \Delta = 0.063 \qquad w + \Delta = 0.087$

Задача 15. (Решить самостоятельно). Проведена выборка объема $n = 1000$ шт. 120 из них оказались бракованными. Найти доверительный интервал доли бракованных изделий в генеральной совокупности. Доверительная вероятность $\gamma = 99$ %.

Ответ: (0,094; 0,146).