И.Р.Ентинзон

Теория ошибок

(Лекции)

Киев – 2023

С О Д Е Р Ж А Н И Е

1. Введение……3
2. Понятие вероятности……6
3. Случайные величины……7
4. Числовые параметры……9
5. Нормальный закон распределения……11
6. Задачи математической статистики……...14
7. Понятие оценки……17
8. Оценки параметров закона распределения $\widetilde{x}$, $\widetilde{\sigma}_X^2$, $\widetilde{\sigma}_X$ ..20
9. Центральная предельная теорема……..22
10. Задача 1. Найти оценку $\widetilde{x}$ при известной дисперсии σ_X^2 ……25
11. Задача 2. Найти оценку $\widetilde{\sigma}_X^2$ при известном математическом ожидании $\bar{x}$……28
12. Задача 3. Найти оценки $\widetilde{x}$, $\widetilde{\sigma}_X^2$ при неизвестных $\bar{x}$ и σ_X^2 ……33
13. Метод доверительных интервалов……37
14. Задача 1. Доверительный интервал для оценки $\widetilde{x}$ математического ожидания $\bar{x}$ при известной дисперсии σ_X^2 нормальной случайной величины X ……..39
15. Задача 2. Доверительный интервал для оценки $\widetilde{\sigma}_X^2$ дисперсии нормальной случайной величины X при известном математическом ожидании $\bar{x}$……42
16. Задача 3. Доверительные интервалы для оценок $\widetilde{x}$, $\widetilde{\sigma}_X^2$ нормальной случайной величины X при неизвестных $\bar{x}$ и σ_X^2 ……...48
17. Процедуры……..54
18. Численный пример……58
19. Метод наименьших квадратов…..……62

20. Литература……...66
21. Приложение……67

1. ВВЕДЕНИЕ

Повторяющиеся измерения любой величины всегда приводят к различным числовым значениям. Предположим, что существует истинное значение измеряемой величины. Тогда будем называть ошибкой измерения ε разность результата измерения x и истинного значения измеряемой величины a, т.е. положим

$$\varepsilon = x - a \tag{1.1}$$

В зависимости от причин, вызывающих ошибки измерения, их делят на случайные, систематические и грубые (промахи). Под *случайными ошибками* понимают ошибки, значения которых меняются от одного измерения к другому. Величина этих ошибок не может быть установлена до опыта. Их возникновение вызвано неточностью измерения (случайными ошибками измерительных приборов, случайными ошибками экспериментатора, неточным соблюдением методики измерения) и непостоянством самой измеряемой величины.

Отличие *систематических ошибок* от случайных состоит в том, что их значение остается постоянным при проведении серии однотипных измерений, и, следовательно, они могут быть исключены из окончательного результата, если их величина определена путем предварительного исследования методики измерения.

Очевидно, деление ошибок на систематические и случайные носит условный характер, так как в одних случаях одни и те же ошибки следует рассматривать как случайные, а в других – как систематические. Например, серия измерений, проведенных при одинаковой температуре, будет содержать систематическую ошибку, появившуюся вследствие того, что не учитывается (или учитывается неправильно) температурная поправка. Однако в серии измерений, проведенных при различных (не зафиксированных) температурах, те же ошибки будут носить случайный характер. Аналогично неучтенная ошибка эталона (ошибки градуировки) для данной лаборатории будет носить систематический характер, однако

при обработке результатов, полученных в разных лабораториях, эта ошибка должна рассматриваться как случайная.

Грубые ошибки (промахи) являются также случайными, однако их характер существенно отличается от характера случайных ошибок измерения. Это отличие связано с тем, что если случайные ошибки измерения возникают при исправно работающей аппаратуре и правильных действиях экспериментатора, то причиной возникновения грубых ошибок является неисправность измерительной техники или ошибки в работе экспериментатора. Если грубые ошибки значительны, то результаты резко выделяются среди других результатов и их следует отбросить. Если это не так, то также результаты можно отбрасывать, но только после применения специальных количественных критериев.

Если измеряется величина, непосредственно интересующая экспериментатора, то измерения называются *прямыми*. Если определяемая величина вычисляется с помощью заданной функциональной зависимости по величинам, полученным путем прямых измерений, то такие измерения называются *косвенными.* В том случае, когда несколько измеряемых величин связано между собой известной функциональной зависимостью (например, три угла треугольника; температура, объем и давление газа и т.п.), вследствие ошибок измерений получаемые величины, как правило, будут противоречить друг другу. Такие измерения называются *условными.*

При прямых, косвенных и условных измерениях ошибки могут быть систематическими, случайными и грубыми.

Любой реальный измерительный прибор, работающий в реальных ситуациях подвержен, как и сама ситуация (установка) неконтролируемому влиянию многих внешних и внутренних факторов. Поэтому измеряемая величина может принимать различные числовые значения.

Статистики шутят по этому поводу, что измеряемая величина α представляет собой функцию от бесконечного числа параметров Φ_i

$$\alpha = \alpha (\Phi_1 ,\Phi_2 ,\Phi_3 ,\Phi_4, \ldots,\Phi_{40}, \ldots,\Phi_{400} , \ldots \Phi^{\infty}) \quad . \qquad (1.2)$$

Теоретически влияют ∞ число параметров, практически 400. Из них

40 параметров экспериментатор может, если он всесторонне обдумает ситуацию, назвать. И только 4 параметра экспериментатор в состоянии контролировать (фиксировать) и то, в пределах точности измерительного прибора.

В этой шутке заключена главная особенность любого измерения – количество факторов, влияющих на измеряемую величину, огромно; их влияние неконтролируемо, случайно. В связи с этим измеряемую величину следует рассматривать как величину случайную. Истинную величину из эксперимента, как видно, получить нельзя и, поэтому, в статистике понятие "истинная величина" отсутствует.

Однако, "нет худа без добра" - факт бесконечного (или огромного) количества влияющих параметров оказался решающим при выработке объективных принципов обсуждения результатов прямых измерений. Рассмотрению этих принципов посвящены настоящие лекции.

Будем называть случайной величиной такую величину, которая в результате опыта может принимать различные числовые значения. Случайная величина - это возможность появится числу на экране, а конкретное число на экране – это реализация случайной величины. Принимать случайная величина может либо дискретные значения, либо любые значения в заданном интервале, т.е. она может быть дискретной или непрерывной случайной величиной. Далее будем рассматривать только непрерывные случайные величины.

В теории ошибок основным объектом изучения являются случайные величины. Поэтому основным математическим аппаратом этой теории является теория вероятностей и вытекающая из нее математическая статистика.

2. ПОНЯТИЕ ВЕРОЯТНОСТИ

В современной теории не делают попыток дать определение понятию вероятности; его считают основным (неопределимым) понятием, удовлетворяющим некоторым аксиомам, которые можно рассматривать как опытные факты:

1. Каждому случайному событию A постановлено в соответствии число

P (A), $0 \le P(A) \le 1$, которое называют вероятностью события A.

2. Вероятность достоверного события U равна 1;

$$P(U) = 1. \tag{2.1}$$

3. Аксиома аддитивности:

если A_1 и A_2 несовместимые случайные события, т.е. пересечение

$A_1 \cap A_2 = V$ (где V – невозможное событие), то

$$P(A_1 U A_2) = P(A_1) + P(A_2), \tag{2.2}$$

т.е. вероятность объединения несовместимых событий равна сумме вероятности этих событий. Отсюда, в частности, следует, что

$$P(U \cup V) = P(U) + P(V) = 1 + P(V) \le 1. \tag{2.3}$$

Поэтому P (V)=0, т.е. вероятность невозможного события равна нулю. Но не наоборот! Если вероятность некоторого события A равна 0,

P(A)=0, то это не означает, что событие A невозможно, A $\neq$ V !

3. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ.

Действительное переменное, которое в зависимости от исхода опыта, т.е. в зависимости от случая принимает различные значения, называется случайной величиной. Случайную величину обозначают большой буквой, например X. Функцией распределения (или законом распределения) случайной величиной X называется функция

$$F(x)=P\{X < x\}. \tag{3.1}$$

Эта функция полностью характеризует случайную величину.

Функцию $f(x)$ называют плотностью распределения или дифференциальной функцией распределения, если функцию распределения $F(x)$ можно представить в виде

$$F(x) = \int_{\infty}^{x} f(t)dt \tag{3.2}$$

Функция распределения $F(x)$ произвольной случайной величины обладает следующими свойствами:

$$1)\ \lim_{x \to +\infty} F(x) = 1, \quad \lim_{x \to -\infty} F(x) = 0, \tag{3.3}$$

заметим, что

$$\int_{-\infty}^{+\infty} f(t)dt = 1 \tag{3.4}$$

.

2) $F(x)$ монотонно не убывает, т.е. при $x_1 < x_2$ имеет место неравенство

$$F(x_1) \le F(x_2) \tag{3.5}$$

3) $F(x)$ непрерывна слева:

$$\lim_{x \to x_0 - 0} F(x) = F(x_0) \tag{3.6}$$

Из аксиом следует, что вероятность того, что случайная величина попадет в заданный полуоткрытый промежуток, равна

$$P\{a \le X < b\} = F(b) - F(a) = \int_a^b f(x)dx. \tag{3.7}$$

Заметим, что здесь $P(X = a) = 0$.

Однако отсюда не следует, что событие $X = a$ является невозможным!

4. ЧИСЛОВЫЕ ПАРАМЕТРЫ

Функции $F(x), f(x)$ известны далеко не всегда и во многих задачах оказывается возможным ограничится приближенной характеристикой плотности вероятности путем задания некоторых числовых параметров, дающих представление об общих свойствах этой функции. В качестве таких параметров часто удобно выбирать величины, называемые в теории вероятностей начальными и центральными моментами различных порядков случайной величины X .

Начальные моменты:

$$m_k = \int_{-\infty}^{+\infty} x^k f(x)dx; \tag{4.1}$$

где k =1,2…

Центральные моменты:

$$\mu_l = \int_{-\infty}^{+\infty} (x - m_1)^l f(x)dx; \tag{4.2}$$

где l =1,2…

Важнейшими среди этого бесконечного множества моментов являются два следующих:

первый начальный момент случайной величины X, имеющий также название математическое ожидание случайной величины X

$$m_1 = \bar{x} = M(X) = \int_{-\infty}^{+\infty} xf(x)dx \tag{4.3}$$

и второй центральный момент случайной величины X, называется также дисперсией случайной величины X -

$$\mu_2 = \sigma_X^2 = D(X) = \int_{-\infty}^{+\infty} (x - \bar{x})^2 f(x) dx \qquad (4.4)$$

$D(X)$ характеризует разброс случайной величины относительно своего математического ожидания.

Заметим, что, если c – постоянная, то

$$M(cX) = cM(X), \qquad (4.5)$$

$$D(cX) = c^2 D(X). \qquad (4.6)$$

5. НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ

Рассмотрим конкретную плотность вероятности - закон нормального распределения вероятности (нормальный закон, закон Гаусса). В дальнейшем будет показано, что все практически важные результаты получены с использованием нормального закона.

Итак, случайная величина X подчиняется нормальному закону, если ее плотность вероятности $f(x)$ может быть представлена в виде

$$f(x)=\frac{1}{\sigma_X\sqrt{2\pi}}e^{-\frac{(x-\bar{x})^2}{2\sigma_X^2}} \tag{5.1}$$

$-\infty < x < +\infty$

Можно показать, что эта функция сконструирована таким образом, чтобы проявлять все свойства плотности вероятности и имеет два параметра, а именно, математическое ожидание и дисперсию:

$$M(X)=\bar{x}, \qquad D(X)=\sigma_X^2 \quad . \tag{5.2}$$

Для функции распределения нормальной величины имеем

$$F(x)=\frac{1}{\sigma_X\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(x_1-\bar{x})^2}{2\sigma_X^2}}dx_1 \quad . \tag{5.3}$$

После некоторого преобразования приходим к виду.

$$F(x)=\frac{1}{2}\left[1+\Phi\left(\frac{x-\bar{x}}{\sigma}\right)\right], \tag{5.4}$$

где $\Phi\left(\frac{x-\bar{x}}{\sigma}\right)$ – интегральная функция Лапласа от аргумента $\frac{x-\bar{x}}{\sigma}$, которая определена формулой (Лаплас Пьер Симон, 1749-1827)

$$\Phi(Z)=\frac{2}{\sqrt{2\pi}}\int_{0}^{Z}e^{-\frac{t^2}{2}}dt$$

(5.5)

Эта функция нечетная,

$$\Phi(-Z)=-\Phi(Z), \tag{5.6}$$

и при $Z \to \infty$ она стремится к 1. Если нормальное распределение $F(x)$ центрировано ($\bar{x} = 0$) и нормировано ($\sigma_X^2 = 1$), то

$$\Phi(x) = 2F(x) - 1. \tag{5.7}$$

Заметим, если закон распределения случайной величины нормален, то задача сведется к поиску (по результатам эксперимента) двух параметров $\bar{x}$ и σ_X^2.

В качестве примера определим вероятность того, что нормальная случайная величина отклонится от своего математического ожидания по абсолютному значению не более, чем на σ, 2σ, 3σ,…., то есть попадет в интервал $(\bar{x} - k\sigma; \bar{x} + k\sigma)$.

$$P\{\bar{x} - k\sigma \le X \le \bar{x} + k\sigma\} = \frac{1}{2}\left[1 + \Phi\left(\frac{\bar{x} + k\sigma - \bar{x}}{\sigma}\right)\right] - \frac{1}{2}\left[1 + \Phi\left(\frac{\bar{x} - k\sigma - \bar{x}}{\sigma}\right)\right] = \Phi(k). \tag{5.8}$$

Мы получили ответ, зная только вид распределения вероятности и не зная конкретных $\bar{x}, \sigma_X^2$. Поэтому результат справедлив при любых $\bar{x}, \sigma_X^2$.

Значения функции $\Phi(k)$ табулированы; для k = I, 2, 3 и 4 она представлена в таблице 5.1:

Таблица 5.1

$k\sigma$	k	$P = \Phi(k)$

σ	1	0,6827
2σ	2	0,9545
3σ	3	0,9973
4σ	4	0,99994

Числа, приведенные в третьем столбце таблицы, равны вероятности того, что отклонение случайной нормальной величины от ее математического ожидания не превысят по абсолютной величине одного, двух и т.д. средних квадратических отклонений. Во многих задачах, где предполагается нормальный закон распределения ошибок, можно приближенно считать, что отклонения, превышающие 3σ настолько маловероятны, что с возможностью их возникновения (в этих задачах) можно не считаться. Это утверждение-рекомендация называется "правило 3σ". Значения функций Ф, f и F приведены в таблицах, см. Приложение.

6. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Математическая статистика рассматривает задачи, в которых требуется сформулировать какие-нибудь утверждения о свойствах множества элементов, называемого генеральной совокупностью, используя сведения о некоторых свойствах подмножества элементов из генеральной совокупности. В математической статистике таким подмножеством выступает случайная выборка n элементов из генеральной совокупности, причем выбор отдельных элементов производится независимо друг от друга.

Здесь мы рассмотрим обработку результатов прямых измерений, являющуюся основанием всей теории ошибок.

Результатом n прямых измерений одной и той же величины является серия чисел

$$x_1,\ x_2,\ \ldots\ldots,\ x_n. \tag{6.1}$$

Целью обработки полученного ряда чисел является поиск приближенного значения измеряемой величины с наибольшей точностью, оценка точности, а также, в некоторых случаях, определение вида закона распределения.

Эта серия чисел $x_1,\ x_2,\ \ldots\ldots,\ x_n$, полученная в результате n опытов, является случайной выборкой объема n из генеральной совокупности $\left\{x_i^k\right\}$ - множества всех возможных реализаций случайных величин $X_1, X_2, \quad , X_n$ (где $i = 1,2,...,n$, а $k \in \{1,2,...,m\}$).

$X_1 \qquad X_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots X_n.$

$x_1^{(1)} \qquad x_2^{(1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots x_n^{(1)}$

$$x_1^{(2)} \quad x_2^{(2)} \ldots\ldots\ldots\ldots\ldots\ldots x_n^{2} \tag{6.2}$$

$$x_1^{(3)} \quad x_2^{(3)} \ldots\ldots\ldots\ldots\ldots\ldots x_n^{(3)}$$

$$\cdot \qquad \cdot \qquad\qquad \cdot$$

$$\cdot \qquad \cdot \qquad\qquad \cdot$$

$$x_1^{(m)} \quad x_2^{(m)} \ldots\ldots\ldots\ldots\ldots\ldots x_n^{(m)}$$

где $m \to \infty$

Заметим, что законы распределения случайных величин X_1, X_2, ……., X_n. неизвестны, а также неизвестны параметры законов распределения этих случайных величин:

$$f_1(x), f_2(x), \ldots\ldots\ldots, f_n(x) \qquad \Rightarrow ? \tag{6.3}$$

$$\overline{x}_1, \overline{x}_2, \ldots\ldots\ldots\ldots\ldots\ldots \overline{x}_n \qquad \Rightarrow ? \tag{6.4}$$

$$\sigma^2_{X_1}, \sigma^2_{X_2} \ldots\ldots\ldots\ldots\ldots \sigma^2_{X_n} \qquad \Rightarrow ? \tag{6.5}$$

Неизвестен и разброс σ_X, который называется также и стандартным отклонением, и средним квадратическим (или квадратичным) отклонением.

Поскольку повторяющиеся измерения однотипны, то следует считать у всех случайных величин X_1, X_2, ……., X_n. одинаковыми закон распределения

$$f_1(x) = f_2(x) = \ldots\ldots\ldots\ldots = f_n(x) = f(x) \quad (6.6)$$

и параметры

$$\bar{x}_1 = \bar{x}_2 \ldots\ldots\ldots\ldots\ldots\ldots = \bar{x}_n = \bar{x} \quad (6.7)$$

$$\sigma^2_{X_1} = \sigma^2_{X_2} = \ldots\ldots\ldots = \sigma^2_{X_n} = \sigma^2_X \quad (6.8)$$

Поэтому будем рассматривать выборку $x_1, x_2, \ldots x_n$ как реализацию независимых случайных величин $X_1, X_2, \ldots\ldots, X_n$ с одинаковыми $f(x), \bar{x}, \sigma^2_X$ Тогда n-мерный случайный вектор ($X_1, X_2, ..., X_n$), в котором величины X_i независимы друг от друга и все имеют распределение $f(x)$, называется математической выборкой объема n. Каждая реализация
($x_1, x_2, \ldots x_n$) случайного вектора ($X_1, X_2, \ldots\ldots, X_n$) есть выборка.

7. ПОНЯТИЕ ОЦЕНКИ

Назовем приближенное значение оцениваемой величины ее оценкой.

Условимся обозначать оценку той же буквой, что и оцениваемую величину, но с волнистой чертой сверху:

a – числовой параметр закона распределения;

$\tilde{a}$ – оценка этого значения;

A – случайная величина, реализацией которой является оценка $\tilde{a}$.

Пример оценок числовых параметров:

$\tilde{x}$ – оценка математического ожидания $\bar{x}$ случайной величины X;

$\tilde{\sigma}_X^2$ – оценка дисперсии σ_X^2 случайной величины X;

$\tilde{\sigma}_X$ – оценка разброса σ_X случайной величины X.

Заметим, что числа

$$\tilde{x} \qquad \tilde{\sigma}_X^2 \qquad \tilde{\sigma}_X \tag{7.1}$$

представляют собой конкретные реализации новых случайных величин

$$\tilde{X} \qquad \tilde{\Sigma}_X^2 \qquad \tilde{\Sigma}_X \tag{7.2}$$

со своими плотностями функций распределения

$$f(\tilde{X}) \qquad f(\tilde{\Sigma}_X^2) \qquad f(\tilde{\Sigma}_X), \tag{7.3}$$

а также математическими ожиданиями

$$M(\tilde{X}) \qquad M(\tilde{\Sigma}_X^2) \qquad M(\tilde{\Sigma}_X) \tag{7.4}$$

и дисперсиями.

$$\sigma_{\tilde{X}}^2 \qquad \sigma_{\tilde{\Sigma}_X^2}^2 \qquad \sigma_{\tilde{\Sigma}_X}^2 \tag{7.5}$$

Новые случайные величины $\widetilde{\mathrm{X}}$, $\widetilde{\Sigma}^2_X$, $\widetilde{\Sigma}_X$ обычно определяют как функции математической выборки (X_1, X_2, ……., X_n).

Например, рассмотрим функцию математической выборки

$$\widetilde{X} = \frac{1}{n}\sum_{i}^{n} X_i \qquad (7.6)$$

Если математическая выборка реализуется, то функция выборки

$$\widetilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad (7.7)$$

называется точечной оценкой математического ожидания $\bar{x}$ случайной величины X.

Свойства точечных оценок.

1) Оценка должна быть *состоятельной*, т.е. дисперсия (разброс) случайной величины $\widetilde{\mathrm{A}}$ в пределе равна нулю,

$$\lim_{n\to\infty} D\left(\widetilde{A}\right) = 0 \qquad (7.8)$$

Тогда она обеспечивает малые отклонения оценки $\widetilde{a}$ от ее математического ожидания.

2) Оценка должна быть *несмещенной*, т.е.

$$M(\widetilde{\mathrm{A}}) = a, \qquad (7.9)$$

математическое ожидание случайной величины $\widetilde{\mathrm{A}}$ должно быть равно оцениваемому числовому параметру. Тогда несмещенность гарантирует отсутствие систематической ошибки: $\widetilde{a} \to a$.

В некоторых случаях довольствуются асимптотической несмещенностью,

$$\lim_{n\to\infty} M\left(\widetilde{a}\right) = \bar{a} \qquad (7.10)$$

Оценки такого типа содержат систематические ошибки, однако их абсолютная величина уменьшается с ростом объема выборки и асимптотически стремится к нулю.

3) Оценка должна быть *эффективной* – из двух оценок $\widetilde{a}_1$ и $\widetilde{a}_2$ одного параметра более эффективной считается та, дисперсия которой меньше. Она является наилучшей оценкой в том смысле, что лучше всего используют информацию, имеющуюся в выборке.

Для определения оценок применяются два основных метода: метод моментов и метод максимального правдоподобия. Здесь мы рассмотрим метод моментов, предложенный английским математиком Карлом Пирсоном (1857-1936). В методе моментов параметры Φ_1, Φ_2, ... закона распределения $f(x;\Phi_1,\Phi_2...)$ выражаются через начальные и центральные моменты исследуемой случайной величины. При этом в качестве их оценок принимаются такие средние значения соответствующих величин, которые удовлетворяют требованиям состоятельности, несмещенности и эффективности.

8. ОЦЕНКИ ПАРАМЕТРОВ ЗАКОНА РАСПРЕДЕЛЕНИЯ: $\tilde{x}, \tilde{\sigma}_X^2, \tilde{\sigma}_X$

В качестве оценки математического ожидания случайной величины X по выборке $x_1, x_2, \ldots, x_n$ принимают обычную среднюю величину

$$\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8.1}$$

$\tilde{x}$ является реализацией случайной величины

$$\tilde{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{8.2}$$

Целесообразность такого выбора была обоснована в середине девятнадцатого века Пафнутием Львовичем Чебышевым.

Теорема Чебышева.

Если $X_1, X_2 \ldots X_n$ есть независимые случайные величины с законами распределения, имеющими математические ожидания $\bar{x}_i$ и конечные дисперсии $\sigma_{X_i}^2 \left(\sigma_{X_i}^2 < \infty\right)$, а $\tilde{x}$ - среднее арифметическое выборки из этих величин (8.1), то для любого $\varepsilon > 0$ справедливо предельное соотношение

$$\lim_{n\to\infty} P\left\{\left|\tilde{x} - \bar{x}\right| \quad \varepsilon\right\} = 0 \tag{8.3}$$

,

где

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} \bar{x}_i. \tag{8.4}$$

Выражение (8.3) определяет стремление $\tilde{x} \to \bar{x}$ при $n \to \infty$ по вероятности.

Теорема показывает, что при достаточно большом числе слагаемых вероятность отклонения среднего арифметического значения случайных величин от среднего арифметического их математических ожиданий на

любую малую величину ε при увеличении n может быть достаточно малой.

Таким образом, при большом n средним арифметическим значением случайной величины можно уверенно пользоваться в качестве приближенной оценки математического ожидания (даже с различными дисперсиями, т.е. разными приборами и установками).

Это принципиальный результат, однако он не указывает путь получения наилучших оценок, удовлетворяющих критериям точечных оценок. Далее будет показано, что в ряде конкретных случаев среднее арифметическое выборки $\widetilde{x}$ является точечной оценкой математического ожидания $\bar{x}$ случайной величины X.

В качестве оценки дисперсии случайной величины X, когда математическое ожидание $\bar{x}$ известно, выбирают выражение

$$\widetilde{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \quad , \tag{8.5}$$

для которого доказано также стремление $\widetilde{\sigma}_X^2 \to \sigma_X^2$ при $n \to \infty$ по вероятности, т.е. для любого $\varepsilon > 0$

$$\lim_{n \to \infty} P\left\{\left|\widetilde{\sigma}_X^2 - \sigma_X^2\right| \quad \varepsilon\right\} = 0 . \tag{8.6}$$

При этом оценка среднего квадратического отклонения

$$\widetilde{\sigma}_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2} , \tag{8.7}$$

будет также состоятельной, однако уже не несмещенной, а только асимптотически несмещенной (несмещенной в пределе при $n \to \infty$).

9. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Закон распределения суммы независимых случайных величин (при достаточно общих условиях) с ростом числа слагаемых n, независимо от вида законов распределения слагаемых случайных величин, стремится к нормальному закону. Впервые теорема такого типа была доказана в 1730 году английским математиком (французом по происхождению) Абрахамом Муавром (1667-1754). В настоящее время имеется ряд теорем, устанавливающих предельную нормальность сумм независимых слагаемых, отличающихся друг от друга дополнительными условиями, налагаемыми на законы распределения слагаемых.

Независимо от математической формулировки физический смысл этих условий сводится к тому, что ни одно из слагаемых суммы не должно иметь дисперсию, соизмеримую с дисперсией всей суммы при увеличении числа слагаемых, т.е. ни одно слагаемое не должно доминировать в сумме. Это означает, что случайная величина ξ распределена нормально, если она является суммой независимых случайных величин при $n >> 1$ или $n \to \infty$

$$\xi = \xi_0 + \xi_1 + \xi_2 + ... + \xi_n \tag{9.1}$$

с дисперсией $$\sigma_\xi^2 = \sigma_{\xi_0}^2 + \sigma_{\xi_1}^2 + \sigma_{\xi_2}^2 + ... + \sigma_{\xi_n}^2, \tag{9.2}$$

причем $$\sigma_\xi^2 >> \sigma_{\xi_0}^2 \sim \sigma_{\xi_1}^2 \sim \sigma_{\xi_2}^2 \sim \sigma_{\xi_3}^2 ... \sigma_{\xi_n}^2 \tag{9.3}$$

Для оценки математического ожидания при $n \to \infty$ это условие соблюдено.

Действительно

$$\widetilde{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + ...\frac{1}{n}X_n \quad , \tag{9.4}$$

$$\sigma^2_{\widetilde{X}} = \sigma^2_{\frac{1}{n}X_1} + \sigma^2_{\frac{1}{n}X_2} + ... + \sigma^2_{\frac{1}{n}X_n} \ , \tag{9.5}$$

$$\sigma^2_{\widetilde{X}} = \frac{1}{n^2}\sigma^2_{X_1} + \frac{1}{n^2}\sigma^2_{X_2} + ... + \frac{1}{n^2}\sigma^2_{X_n} \ , \tag{9.6}$$

$$\sigma^2_{\widetilde{X}} = \frac{1}{n^2}\sigma^2_{X} + \frac{1}{n^2}\sigma^2_{X} + ... + \frac{1}{n^2}\sigma^2_{X} \ , \tag{9.7}$$

$$\sigma^2_{\widetilde{X}} = \frac{1}{n}\sigma^2_{X}; \tag{9.8}$$

причем . $$\sigma^2_{\frac{X}{n}} = \frac{1}{n^2}\sigma^2_{X} \tag{9.9}$$

Поэтому

$$D(\widetilde{X}) >> D\left(\frac{1}{n}X_i\right), \tag{9.10}$$

что и требовалось доказать.

Посмотрим на случайную величину X как на сумму

$$X = \bar{x} + \xi_1 + \xi_2 + \xi_3 + ...\xi_n, \tag{9.11}$$

где ξ_i - случайные величины, вызванные влиянием различных независимых факторов, которые не удалось фиксировать абсолютно точно.

Поскольку число факторов можно считать очень большим, а дисперсии – сравнимыми, то имеем

$$\sigma_X^2 = \sigma_X^2 + \sigma_{\xi_1}^2 + \sigma_{\xi_2}^2 + ... + \sigma_{\xi_n}^2 = 0 + \sigma_{\xi_1}^2 + \sigma_{\xi_2}^2 + ... + \sigma_{\xi_n}^2, \quad (9.12)$$

а, значит, можно считать случайную величину X - распределенной по нормальному закону.

При применении предельных теорем к различным конкретным случаям основное условие этих теорем, согласно которому число слагаемых n должно стремиться к бесконечности, не выполняется, так как n остается конечным, а иногда и не особенно большим числом. Однако предельные свойства сумм обычно проявляются уже при сравнительно небольшом n, и хотя предельные теоремы не дают ответа на вопрос, начиная с какого n их результатами можно пользоваться, практика показывает, что во многих случаях эти свойства можно считать имеющими место уже при сравнительно малых n ($n \sim 8 \div 10$).

Именно поэтому нормальный закон распределения случайной величины X играет определяющую роль в исследовании экспериментальных результатов. В дальнейшем мы используем этот факт при анализе той или иной статистической процедуры.

Далее рассмотрим три основные задачи математической статистики, которые возникают при обработке результатов эксперимента. Во всех этих задачах предполагается, что выборка дана, однако в разных задачах могут быть известны дополнительные данные о случайной величине X, а обработка полученных данных будет иметь различные цели.

10. ЗАДАЧА 1. НАЙТИ ОЦЕНКУ $\tilde{x}$ ПРИ ИЗВЕСТНОЙ ДИСПЕРСИИ σ_X^2

Найти оценку математического ожидания $\bar{x}$ случайной величины X и определить дисперсию оценки, если дана выборка $x_1, x_2, \ldots\ldots, x_n$ и известно, что дисперсия случайной величины X равна σ_X^2 . Эта задача возникает в том случае, когда производится n-кратное измерение искомой величины, причем измерительный метод свободен от систематических ошибок, а его точность характеризуется известной дисперсией

$X_1 \quad X_2 \quad X_3 \quad \ldots\ldots. X_n$ случайные величины

$\bar{x}, \sigma_X^2 \quad \bar{x}, \sigma_X^2 \quad \bar{x}, \sigma_X^2 \ldots\ldots. \bar{x}, \sigma_X^2$ параметры (10.1)

$x_1 \quad x_2 \quad x_3 \quad \ldots \ldots \ldots x_n$ выборка

1. В качестве оценки $\tilde{x}$ математического ожидания $\bar{x}$ случайных

величин $X_1, X_2, \ldots X_n$ возьмем среднее арифметическое

$$\tilde{x} = \frac{1}{n} \sum_{i=n}^{n} x_i . \tag{10.2}$$

$\tilde{x}$ по построению является реализацией случайной величины

$$\tilde{X} = \frac{1}{n} X_1 + \frac{1}{n} X_2 + \ldots\ldots\ldots + \frac{1}{n} X_n \tag{10.3}$$

2. Вычислим дисперсию случайной величины $\widetilde{X}$, учтя свойство дисперсии

$$D(c\xi) = c^2 D(\xi), \tag{10.4}$$

$$D(\widetilde{X}) = \sigma^2_{\widetilde{X}} = D(\frac{1}{n} X_1) + D(\frac{1}{n} X_2) + ...D(\frac{1}{n} X_n) =$$
$$= \frac{1}{n^2} D(X_1) + \frac{1}{n^2} D(X_2) + ... + \frac{1}{n^2} D(X_n) =$$
$$= \frac{1}{n^2}\sigma^2_X + \frac{1}{n^2}\sigma^2_X + ... + \frac{1}{n^2}\sigma^2_X = \frac{1}{n}\sigma^2_X \tag{10.5}$$

Оценка (10.2) является состоятельной и несмещенной. Действительно, устремляя $n \to \infty$, на основании (10.5) находим

$$\lim_{n\to\infty} D(\widetilde{X}) = \lim_{n\to\infty} \frac{1}{n}\sigma^2_X = 0, \tag{10.6}$$

т.е. $\widetilde{x}$ является состоятельной оценкой математического ожидания. Найдя математическое ожидание обеих частей равенства (10.2), получим

$$M(\widetilde{X}) = M\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n} n\,\bar{x} = \bar{x}, \tag{10.7}$$

т.е. оценка $\widetilde{x}$ является несмещенной.

Как видно из соотношения (10.5), для получения более эффективных оценок следует увеличивать объем выборки n.

Если случайная величина X является нормальной, то $\widetilde{X}$, будучи суммой нормальных случайных величин, является также нормальной, и, следовательно, плотность вероятности этой оценки определяется формулой Гаусса

$$f(\widetilde{x}) = \frac{1}{\sigma_{\widetilde{x}}\sqrt{2\pi}} e^{-\frac{(\widetilde{x}-\bar{x})^2}{2\sigma^2_{\widetilde{x}}}} = \frac{\sqrt{n}}{\sigma_x\sqrt{2\pi}} e^{-\frac{n(\widetilde{x}-\bar{x})^2}{2\sigma^2_x}}$$

(10.8)

с параметрами $\bar{x}, \sigma_X^2 / n.$.

Если закон распределения X не является нормальным, но число n слагаемых достаточно велико, то формулу (10.8) для функции $f(\tilde{x})$ можно рассматривать как приближенную, опираясь на центральную предельную теорему. Точность приближения при этом зависит от вида закона распределения случайной величины X и объема выборки n.

11. ЗАДАЧА 2. НАЙТИ ОЦЕНКУ $\widetilde{\sigma}_X^2$ ПРИ ИЗВЕСТНОМ МАТЕМАТИЧЕСКОМ ОЖИДАНИИ $\bar{x}$

Найти оценку дисперсии случайной величины X, если даны выборка $x_1, x_2, \ldots x_n$ и ее математическое ожидание $\bar{x}$.

Эта задача соответствует обработке результатов измерений эталона, произведенных для определения точности измерительного метода. Предполагается, что условия измерения не меняются, а систематическая ошибка или отсутствует или остается неизменной при всех измерениях.

X_1 X_2 X_3 X_4

случайные величины

$\widetilde{x}, \sigma_X^2$ $\widetilde{x}, \sigma_X^2$ $\widetilde{x}, \sigma_X^2$ $\widetilde{x}, \sigma_X^2$

параметры

x_1 x_2 x_3 x_n выборка

По определению, как уже известно, дисперсия случайной величины X_i есть математическое ожидание случайной величины $(X_i - \bar{x})^2$, которая в i- м опыте приняла значение $(x_i - \bar{x})^2$,

$(X_1 - \bar{x})^2$ $(X_2 - \bar{x})^2$ $(X_3 - \bar{x})^2$ $(X_n - \bar{x})^2$ случайные величины

$(x_1 - \bar{x})^2$ $(x_2 - \bar{x})^2$ $(x_3 - \bar{x})^2$ $(x_n - \bar{x})^2$ выборка.

Для оценки дисперсии, т.е. математического ожидания случайной величины $(X_i - \bar{x})^2$, возьмем среднее арифметическое

$$\widetilde{D}(X) = \widetilde{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (11.1)$$

Эта оценка является несмещенной, так как, находя математическое ожидание обеих частей последнего равенства, имеем

$$M(\widetilde{\Sigma}_X^2) = M\left[\frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{x})^2\right] = \frac{1}{n}\sum_{i=1}^{n} M\left[(X_i - \bar{x})^2\right] = \frac{1}{n} n\,\sigma_X^2 = \sigma_X^2 \qquad (11.2)$$

Можно показать, что она также является состоятельной:

$$\lim_{n \to \infty} D(\widetilde{\Sigma}_X^2) = 0 \qquad (11.3)$$

если четвертый центральный момент случайной величины X существует.

В частном случае, когда случайная величина X нормальная, дисперсия случайной величины $\widetilde{\Sigma}_X^2$ принимает вид

$$D(\widetilde{\Sigma}_X^2) = \frac{2}{n}\sigma_X^4 . \qquad (11.4)$$

Таким образом, для нормальной случайной величины оценка $\widetilde{\sigma}_X^2$, полученная методом моментов, является состоятельной, а ее эффективность определяется простой формулой (11.4).

Если по выборке требуется определить не оценку $\widetilde{\sigma}_X^2$ дисперсии случайной величины X, а оценку $\widetilde{\sigma}_X$ среднего квадратического отклонения случайной величины X, то в качестве ее берем

$$\widetilde{\sigma}_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad (11.5)$$

Эта оценка $\lim_{n\to\infty} D\left(\widetilde{\Sigma}_X\right) = 0$ состоятельна ,но уже не несмещенная, а только асимптотически несмещенная (несмещенная в пределе при $n \to \infty$),

$$\lim_{n\to\infty} M(\widetilde{\Sigma}_X) = \sigma_X \qquad (11.6)$$

Если случайная величина X нормальная, то, как следует из теории, для получения несмещенной оценки среднего квадратического отклонения необходимо ввести поправочный множитель k_{n+1} и переписать формулу (11.5) в виде

$$\widetilde{\sigma}_X = k_{n+1}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \approx \sqrt{\frac{1}{n-0.45}\sum_{i=1}^{n}(x_i - \bar{x})^2}. \qquad (11.7)$$

Поправочный коэффициент k_{n+1} определяется формулой

$$k_{n+1} = \frac{\sqrt{n}\,\Gamma(\frac{n}{2})}{\sqrt{2}\,\Gamma(\frac{n+1}{2})}, \qquad (11.8)$$

где гамма-функция при n целом положительном равна

$$\Gamma(n) = (n-1)! \qquad (11.9)$$

В табл. 11.1 приводятся значения коэффициента k_n, которые показывают, что приближение этого коэффициента к единице с ростом n происходит довольно быстро.

Значения коэффициента k_n Таблица 11.1

n	3	4	5	6	7	10	15	25	**50**
k_n	1,128	1,085	1,064	1,051	1,042	1,028	1,018	1,010	**1,005**

Необходимость введения в формуле (11.7) поправочного коэффициента показывает, что оценка параметра, полученная путем нелинейных операций (в нашем случае – это извлечение квадратного корня), уже перестает быть несмещенной. Этот результат является прямым следствием того, что математическое ожидание функции случайной величины равно этой функции от математического ожидания аргумента только для линейной функциональной зависимости, т.е.

$$M(f(X)) = f(\bar{x}); \qquad (11.10)$$

когда $f(X) = a + bX$.

12. ЗАДАЧА 3. НАЙТИ ОЦЕНКИ $\widetilde{x}$ и $\widetilde{\sigma}_X^2$ ПРИ НЕИЗВЕСТНЫХ $\bar{x}$ и $\bar{\sigma}_X^2$

Найти оценки математического ожидания $\bar{x}$ и дисперсии $\bar{\sigma}_X^2$ случайной величины X по выборке $x_1, x_2, \ldots, x_n$

Эта задача весьма типична для физических измерений и возникает в том случае, когда не известны ни измеряемая величина, ни точность применяемого метода измерения.

Оценка математического ожидания $\widetilde{x}$ и в этом случае в соответствии с методом моментов определяется формулой

$$\widetilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (12.1)$$

Она, как было показано в задаче 1, состоятельная, несмещенная и эффективная. Однако при вычислении оценки дисперсии возникает осложнение – в данном случае нельзя непосредственно применять формулу (11.1), так как в нее входит неизвестная в данной задаче величина $\bar{x}$. Естественно заменить в этой формуле величину $\bar{x}$ ее оценкой $\widetilde{x}$, т.е. воспользоваться формулой

$$\widetilde{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \widetilde{x})^2, \qquad (12.2)$$

однако при этом выясняется, что полученная таким образом оценка дисперсии перестает быть несмещенной. Происходит это потому, что одна и та же выборка использована дважды: для нахождения оценки математического ожидания и нахождения оценки дисперсии. Поэтому в формуле (12.2) слагаемые уже не являются независимыми, поскольку между ними устанавливается связь вследствие наличия соотношения (12.1). В результате один элемент выборки как бы теряется. И для получения несмещенной оценки дисперсии необходимо производить

усреднение величины $(x_i - \tilde{x})^2$ уже не между n, а между $(n-1)$ значениями, т.е. вместо формулы (12.2) использовать формулу

$$\tilde{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \tilde{x})^2. \qquad (12.3)$$

Можно показать, что эта оценка является несмещенной, т.е.

$$M(\tilde{\Sigma}_X^2) = \sigma_X^2. \qquad (12.4)$$

Если четвертый центральный момент случайной величины X существует, то оценка (12.3) является состоятельной (в частности, она состоятельна для нормальной случайной величины X) т.е.

$$\lim_{n\to\infty} D(\tilde{\Sigma}_X^2) = 0 \qquad (12.5)$$

Корень квадратный из оценки дисперсии (12.3) дает оценку среднего квадратического отклонения; оценка будет только асимптотически несмещенной.

Для выборки из нормальной генеральной совокупности (т.е. в том случае, когда случайная величина X нормальная) несмещенная оценка может быть получена, если ввести корректирующий множитель, положив

$$\tilde{\sigma}_X = k_n \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \tilde{x})^2}, \qquad (12.6)$$

где k_n определяется формулой (11.8).

Однако в данном случае вместо k_{n+1} должен быть взят множитель k_n, т.е. с индексом, на единицу меньше, чем в формуле (11.7) (один элемент выборки как бы теряется).

Рассмотрим параметры закона распределения случайной величины $\widetilde{X}$: $M(\widetilde{X}), D(\widetilde{X})$ и $\sigma_{\widetilde{X}}$.

Поскольку $\widetilde{x}$ в этой задаче определяется, той же формулой, что и в предыдущей задаче, то по-прежнему

$$M(\widetilde{X}) = \overline{x} = M(X); \tag{12.7}$$

$$D(\widetilde{X}) = \sigma^2_{\widetilde{X}} = \frac{1}{n}\sigma^2_X. \tag{12.8}$$

Следовательно, для оценки дисперсии $\sigma^2_{\widetilde{X}}$ получим формулу

$$\widetilde{D}(\widetilde{X}) = \widetilde{\sigma}^2_{\widetilde{X}} = \frac{1}{n}\widetilde{\sigma}^2_X, \tag{12.9}$$

где $\widetilde{\sigma}^2_X$ дается формулой (12.3)

Поскольку соотношение между $\widetilde{\sigma}^2_{\widetilde{X}}$ и $\widetilde{\sigma}^2_X$ линейное, оценка $\widetilde{D}(\widetilde{X})$ будет состоятельной и несмещенной. При переходе от оценки $\widetilde{\sigma}^2_{\widetilde{X}}$ к оценке $\widetilde{\sigma}_{\widetilde{X}}$ вследствие нелинейности преобразования несмещенность теряется, однако для нормальной величины X и в данном случае можно получить несмещенную оценку, если воспользоваться несмещенной оценкой (12.6) для среднего квадратического отклонения $\widetilde{\sigma}_{\widetilde{X}}$, т.е. положить

$$\widetilde{\sigma}_{\widetilde{X}} = \frac{1}{\sqrt{n}}\widetilde{\sigma}_X \tag{12.10}$$

Таким образом, оценка дисперсии случайной величины X при неизвестном математическом ожидании отличается от ее оценки при известном $\overline{x}$ тем, что в ряде формул объем выборки n заменяется на (n-1).

Эта замена, являющаяся несущественной при большом n, может привести к заметному увеличению оценки дисперсии при небольшом объеме выборки.

В заключение, перечислим ряд вычисляемых оценок, характеризующих случайные величины X и $\tilde{X}$:

$\tilde{X}$ - средняя арифметическая случайная величина по математической выборке объема n;

$\tilde{x}$ - оценка математического ожидания $\bar{x}$ случайной величины X;

$\tilde{\sigma}_X^2$ - оценка дисперсии σ_X^2 случайной величины X;

$\tilde{\sigma}_X$ - оценка среднего квадратического отклонения случайной величины X;

$\sigma_{\tilde{X}}^2$ - дисперсия случайной величины $\tilde{X}$ - оценки математического ожидания $\bar{x}$ случайной величины X;

$\tilde{\sigma}_{\tilde{X}}^2$ - оценка дисперсии случайной величины $\tilde{X}$- оценки математического ожидания $\bar{x}$ случайной величины X;

$\tilde{\sigma}_{\tilde{X}}$- оценка среднего квадратического отклонения случайной величины $\tilde{X}$ - оценки математического ожидания $\tilde{x}$ случайной величины X.

13. МЕТОД ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ

Оценки параметров закона распределения $f(x)$, получаемые по случайным значениям элементов выборки, являются случайными величинами, которые следует рассматривать как приближенные значения этих параметров. Исчерпывающей характеристикой точности этого приближения являются законы распределения оценок. Однако такая характеристика точности оценок неудобна, так как она малонаглядна и достаточно сложна. Поэтому в математической статистике для характеристики точности оценок широкое распростарнение получил метод, называемый методом доверительных интервалов. Сущность метода заключается в следующем. Предположим, требуется охарактеризовать точность оценки $\tilde{a}$ параметра a : $(\bar{x}, \sigma_X^2, \sigma_X)$. Выберем некоторую большую вероятность α считая ее настолько близкой к 1, чтобы возможностью появления событий, вероятность которых равна (1-α), можно было бы практически не считаться. В химии чаще всего выбирают α = 0,95. Определим такой интервал $(a-\varepsilon, a+\varepsilon)$, вероятность попадания в который случайной величины $\tilde{A}$ равнялась бы α:

$$\alpha = P\left\{ \left| a - \tilde{A} \right| \le \varepsilon \right\} = \int_{a-\varepsilon}^{a+\varepsilon} f(\tilde{a})\ d\tilde{a}, \qquad (13.1)$$

где $f(\tilde{a})$ - плотность вероятности случайной величины $\tilde{A}$. Интервал $(a-\varepsilon, a+\varepsilon)$ называется доверительным интервалом, соответствующим доверительной вероятности α.

Смысл этого подхода состоит в том, что с вероятностью α гарантируется ошибка в определении параметра a, не превосходящая по абсолютной величине $\varepsilon \equiv \varepsilon_\alpha$, которую называют погрешностью измерения или шириной α- доверительного интервала.

Значение доверительной вероятности α выбирают с таким расчетом, чтобы в более ответственных случаях иметь большую гарантию того, что ошибка не выйдет за доверительный интервал, а в менее

ответственных случаях – меньшую вероятность. Известная неопределенность в выборе величины α не особенно существенно сказывается на величине ε, так как начиная с некоторого значения α, величина ε начинает расти сравнительно медленно.

В некоторых случаях целесообразно задать ширину доверительного интервала (величину наибольшей по модулю ошибки ε) и определить соответствующую этой величине доверительную вероятность. В этом случае получаем возможность установить процент случаев, в которых будет наблюдаться ошибка, не превосходящая некоторого значения, предоставляющего определенный интерес в данной задаче.

Независимо от того, является исходной величиной доверительная вероятность или доверительный интервал, для определения связи α и ε необходимо явно выразить вероятность $P\left\{\left|a-\widetilde{A}\right|\le\varepsilon\right\}$ как функцию ε.

Для нахождения указанной вероятности необходимо знать закон распределения оценки $\widetilde{A}$, который существенно зависит от вида закона распределения случайной величины X, выборка значений которой обрабатывается и который обычно неизвестен. Эта задача в различных ситуациях решается различно. Основные результаты, полученные для доверительных интервалов, относятся к выборкам из нормальной генеральной совокупности.

Положим приближенно, в соответствии с выводами центральных предельных теорем (раздел 9), закон распределения ошибок измерения нормальным. И в этом приближении далее рассмотрим вычисление доверительных интервалов для нескольких оценок, наиболее часто встречающихся в практике.

Заметим, что процедуры вычисления доверительных интервалов и соответствующая форма представления экспериментальных данных узаконены государственными стандартами всех развитых стран.

14. ЗАДАЧА 1. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ОЦЕНКИ $\tilde{x}$ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ $\bar{x}$ ПРИ ИЗВЕСТНОЙ ДИСПЕРСИИ σ_X^2 НОРМАЛЬНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ X

Случайная величина X – нормальная случайная величина, значит и оценка

$$\tilde{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{14.1}$$

являющаяся суммой нормальных случайных величин, подчинена нормальному закону с математическим ожиданием $\bar{x}$ и дисперсией $\sigma_{\tilde{X}}^2 = \sigma_X^2 / n$, т.е. плотность вероятности случайной величины $\tilde{X}$ записывается

$$f(\tilde{x}) = \frac{1}{\sigma_{\tilde{X}}\sqrt{2\pi}} e^{-\frac{(\tilde{x}-\bar{x})^2}{2\sigma_{\tilde{X}}^2}}. \tag{14.2}$$

Следовательно

$$P\left\{\left|\tilde{X}-\bar{x}\right| \le \varepsilon\right\} = P\left\{\bar{x}-\varepsilon \le \tilde{X} \le \bar{x}+\varepsilon\right\} =$$
$$= \int_{\bar{x}-\varepsilon}^{\bar{x}+\varepsilon} f(\tilde{x})d\tilde{x} = \frac{1}{\sigma_{\tilde{X}}\sqrt{2\pi}} \int_{\bar{x}-\varepsilon}^{\bar{x}+\varepsilon} e^{-\frac{(\tilde{x}-\bar{x})^2}{2\sigma_{\tilde{X}}^2}} d\tilde{x}. \tag{14.3}$$

Сделав в последнем интеграле замену переменных и положив

$$t = \frac{(\tilde{x}-\bar{x})}{\sigma_{\tilde{X}}}, \tag{14.4}$$

получим

$$P\left\{\left|\tilde{X}-\bar{x}\right|\le\varepsilon\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\varepsilon/\sigma_{\tilde{X}}}^{\varepsilon/\sigma_{\tilde{X}}}e^{-\frac{t^2}{2}}dt=\Phi\left(\frac{\varepsilon}{\sigma_{\tilde{X}}}\right),\qquad(14.5)$$

где Ф (Z) –функция Лапласа. Окончательно находим связь между доверительным интервалом ε и доверительной вероятностью α:

$$\alpha=\Phi(\varepsilon/\sigma_{\tilde{X}}),\qquad(14.6)$$

где $\sigma_{\tilde{X}}=\sigma_X/\sqrt{n}$.

Теперь по заданной вероятности α и таблице функции Лапласа Ф(Z) можно получить соответствующий аргумент Z. Учитывая связь $\varepsilon/\sigma_{\tilde{X}}=\mathrm{Z}$, получим $\varepsilon_\alpha=\mathrm{Z}/\sigma_{\tilde{X}}$. Окончательный результат записывают в виде

$$\bar{x}=\tilde{x}\pm\varepsilon_\alpha\qquad\text{или}\qquad\tilde{x}-\varepsilon_\alpha\le\bar{x}\le\tilde{x}+\varepsilon_\alpha\qquad(14.7)$$

и говорят, что случайный интервал $(\tilde{X}-\varepsilon_\alpha,\tilde{X}+\varepsilon_\alpha)$ с вероятностью α “покрывает” математическое ожидание $\bar{x}$ случайной величины X, а реализация этого случайного интервала (14.7) приблизительно с вероятностью α содержит математическое ожидание $\bar{x}$.

Если взять целые значения аргумента функции Лапласа, положив, учитывая связь $\mathrm{Z}=\varepsilon/\sigma_{\tilde{X}}$,

$$\varepsilon=\sigma_{\tilde{X}},2\sigma_{\tilde{X}},3\sigma_{\tilde{X}},4\sigma_{\tilde{X}},\qquad(14.8)$$

то для доверительной вероятности получим соответственно (табл.5.1)

$$\alpha=0{,}683;\ 0{,}954;\ 0{,}9973;\ 0{,}99994.\qquad(14.9)$$

Обычно считают, что вероятность 0,9973 практически равна I и поэтому часто окончательный результат записывают в виде

$$\bar{x} = \tilde{x} \pm 3\sigma_{\tilde{X}} = \tilde{x} \pm 3\left(\frac{1}{\sqrt{n}}\sigma_X\right) \quad (14.10)$$

и говорят что ошибка “не может выйти за три сигмы”.Не следует забывать, что “правило 3σ” довольно искусственное (может быть “правило 4σ”) и что эта вероятность получена для нормального закона распределения случайной величины X.

Отметим также, что резкое повышение доверительной вероятности при переходе от “правила 3σ” ($\alpha = 0{,}9973$) к “правилу 4 σ” ($\alpha = 0{,}99994$) увеличивает величину доверительного интервала только на 1/3.

15. ЗАДАЧА 2. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ОЦЕНКИ $\tilde{\sigma}_x^2$ ДИСПЕРСИИ НОРМАЛЬНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ X ПРИ ИЗВЕСТНОМ МАТЕМАТИЧЕСКОМ ОЖИДАНИИ $\bar{x}$

Случайная величина X - нормальная случайная величина. Было показано (раздел 11), что в соответствии с методом моментов в качестве оценки дисперсии и оценки среднего квадратического отклонения целесообразно брать величины,

$$\tilde{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \tag{15.1}$$

$$\tilde{\sigma}_X = k_n \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{15.2}$$

являющиеся реализациями соответствующих случайных величин $\tilde{\Sigma}_X^2, \tilde{\Sigma}_X$, распределение которых неизвестно.

Задача состоит в том, чтобы получить явное выражение для доверительной вероятности

$$\alpha = P\left\{\left|\tilde{\Sigma}_X^2 - \sigma_X^2\right| \le \varepsilon\right\} \tag{15.3}$$

и тем самым установить зависимость ширины доверительного интервала ε_α от доверительной вероятности α.

Введем для этой цели безразмерную случайную величину

$$\xi = n\frac{\widetilde{\Sigma}_X^2}{\sigma_X^2} \tag{15.4}$$

и новую ширину

$$\delta = \frac{n\varepsilon}{\sigma_X^2} \tag{15.5}$$

Исходное неравенство $\left|\widetilde{\Sigma}_X^2 - \sigma_X^2\right| \le \varepsilon$ умножим на $\frac{n}{\sigma_X^2}$ и получим эквивалентное неравенство

$$\left|\xi - n\right| \le \delta \tag{15.6}$$

с новой безразмерной случайной величиной ξ и шириной δ. Следовательно, для доверительной вероятности α имеем

$$\alpha = P\left(\left|\xi - n\right| \le \delta\right) = P(n - \delta \le \xi \le n + \delta)\,. \tag{15.7}$$

Учитывая новые обозначения и формулу для оценки дисперсии (15.1), получим другое выражение для случайной величины ξ:

$$\xi = n\frac{\widetilde{\Sigma}_X^2}{\sigma_X^2} = \frac{n}{\sigma_X^2}\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{x}\right)^2 = \sum_{i=1}^{n}\frac{\left(X_i - \bar{x}\right)^2}{\sigma_X^2} \tag{15.8}$$

Введем случайную величину $Y_i = \dfrac{X_i - \bar{x}}{\sigma_X}$, тогда

$$\xi = \sum_{i=1}^{n}\frac{\left(X_i - \bar{x}\right)^2}{\sigma_X^2} = \sum_{i=1}^{n} Y_i^2 \quad . \tag{15.9}$$

Случайные величины $\dfrac{X_i - \bar{x}}{\sigma_X}$ являются нормальными по определению, центрированными (их математические ожидания равны 0) и нормированными (их дисперсии равны 1). Покажем это.

$$M\left(\frac{X_i - \bar{x}}{\sigma_X}\right) = \frac{M(X_i) - \bar{x}}{\sigma_X} = \frac{\bar{x} - \bar{x}}{\sigma_X} = 0 \qquad (15.10)$$

$$D\left(\frac{X_i - x}{\sigma_X}\right) = \int_{-\infty}^{+\infty} \left(\frac{X_i - \bar{x}}{\sigma_X} - M\left(\frac{X_i - \bar{x}}{\sigma_X}\right)\right)^2 f(x)dx = \qquad (15.11)$$
$$= \frac{1}{\sigma_X^2} \int_{-\infty}^{+\infty} (X_i - \bar{x})^2 f(x)dx = \frac{1}{\sigma_X^2}\sigma_X^2 = 1$$

Следовательно, мы получили новую случайную величину ξ в виде суммы нормальных случайных величин Y_i, центрированных и нормированных (15.9). Такая случайная величина ξ изучена, она подчиняется так называемому "закону распределения χ^2" с n степенями свободы, для которого имеются подробные таблицы (см. Приложение).
Таким образом, для доверительной вероятности α получим:

$$\alpha = \int_{n-\delta}^{n+\delta} f_n(\xi)d\xi \qquad (15.12)$$

Эта формула не содержит параметров распределения случайной величины X. Поэтому, рассчитав значения интеграла (15.12) для различных δ и n, можно составить таблицу значений δ как функцию от α и n. Такие таблицы можно найти в справочниках.

Таблица 15.1

n\1 – α	0,4	0,3	0,2	0,1	0,05	0,04
3 4 5		.	. . δ	.		

6 … 30						

В таблице 15.1 по n и 1-α находим δ. Тогда окончательно для доверительного интервала оценки дисперсии имеем

$$P\{\left|\widetilde{\Sigma}^2_X - \sigma^2_X\right| \le \frac{\delta\sigma^2_X}{n}\} = \alpha, \quad \text{т.е.} \tag{15.13}$$

$$P\{\frac{\widetilde{\Sigma}^2_X}{1+\frac{\delta}{n}} \le \sigma^2_X \le \frac{\widetilde{\Sigma}^2_X}{1-\frac{\delta}{n}}\} = \alpha, \tag{15.14}$$

Результат представляют в виде интервала $(\frac{\widetilde{\sigma}^2_X}{1+\delta_\alpha/n}, \frac{\widetilde{\sigma}^2_X}{1-\delta_\alpha/n})$, который приблизительно с вероятностью α содержит дисперсию σ^2_X случайной величины X.

Плотность распределения χ^2 несимметрична, поэтому при определении доверительного интервала по формуле (15.12) вероятности выхода ошибки за доверительный интервал в меньшую или большую стороны не одинаковы. В некоторых случаях целесообразно добиться равенства этих вероятностей. Это можно сделать, расположив доверительный интервал несимметрично относительно математического ожидания случайной величины, выбрав границы δ_1 и δ_2 таким образом, чтобы вероятность выхода за интервал влево, $\chi^2 \le \delta_1$, или вправо, χ^2 δ_2, была бы равна $\frac{1}{2}(1-\alpha)$.Отправной точкой является тот факт, что случайная величина $\frac{n\widetilde{\Sigma}^2_X}{\sigma^2_X} = \xi$ удовлетворяет χ^2–распределению с n степенями свободы. Из таблиц по заданному α и m=n степеням свободы определяют два числа δ_1 и δ_2 такие, что

$$P(\chi^2 < \delta_1) = 1 - P(\chi^2 > \delta_1) = \frac{1-\alpha}{2},\ P(\chi^2 > \delta_2) = \frac{1-\alpha}{2}, \qquad (15.15)$$

$$\delta_1 = \chi^2_{(1+\alpha)/2}, \qquad \delta_2 = \chi^2_{(1-\alpha)/2}.$$

где $(1+\alpha)/2$ и $(1-\alpha)/2$ соответствующие уровни значимости.

Например, для α=0.92 и m=19 получаем, что $\delta_1 = \chi^2_{0.99} = 7.6$ и $\delta_2 = \chi^2_{0.01} = 36.2$.

Числа δ_1 и δ_2 определены так, что

$$P(\delta_1 < \frac{n\widetilde{\Sigma}^2_X}{\sigma^2_X} < \delta_2) = \alpha \qquad (15.16)$$

или

$$P(\frac{n\widetilde{\Sigma}^2_X}{\delta_2} < \sigma^2_X < \frac{n\widetilde{\Sigma}^2_X}{\delta_1}) = \alpha. \qquad (15.17)$$

Таким образом, $(\frac{n\widetilde{\Sigma}^2_X}{\delta_2}, \frac{n\widetilde{\Sigma}^2_X}{\delta_1})$ есть доверительная оценка σ^2_X с мерой надежности α.

Окончательно результат записывают в виде α–доверительного интервала для σ^2_X ,

$$\frac{n\widetilde{\sigma}^2_X}{\delta_2} < \sigma^2_X < \frac{n\widetilde{\sigma}^2_X}{\delta_1} \qquad (15.18)$$

Аналогично для среднеквадратичного отклонения σ_X ,

$$\sqrt{\frac{n}{\delta_2}}\widetilde{\sigma}_X < \sigma_X < \sqrt{\frac{n}{\delta_1}}\widetilde{\sigma}_X. \qquad (15.19)$$

При этом говорят, что приблизительно с вероятностью α в соответствующих интервалах (15.18) и (15.19) находятся соответствующие параметры σ^2_X и σ_X.

16. ЗАДАЧА 3. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ ОЦЕНОК $\widetilde{x}, \widetilde{\sigma}_X^2$ (МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ И ДИСПЕРСИИ) НОРМАЛЬНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ X ПРИ НЕИЗВЕСТНЫХ $\bar{x}$ и σ_X^2

Задача состоит в том, чтобы по выборке $x_1, x_2, \ldots x_n$ определить доверительные интервалы для оценок $\widetilde{x}, \widetilde{\sigma}_X^2$ неизвестных параметров $\bar{x}$ и σ_X^2 нормально распределенной случайной величины X.

$$1.\quad P\{\left|\widetilde{X} - \bar{x}\right| \le \varepsilon\} = \alpha, \tag{16.1}$$

$$2.\quad P\{\left|\widetilde{\Sigma}_X^2 - \sigma_X^2\right| \le \varepsilon\} = \alpha. \tag{16.2}$$

Этот случай наиболее распространен в экспериментах. Но σ_X^2 неизвестно и невозможно вычислить выражение (16.1) и $\bar{x}$ неизвестно и невозможно вычислить выражение (16.2), как это было в 1-ой и во 2-ой задачах.

Английский статистик Стьюдент (В. Госсет, 1876-1939) нашел выход – он построил новую случайную величину T,

$$\mathrm{T} = \frac{(\widetilde{X} - \bar{x})}{\sqrt{\widetilde{\Sigma}_{\widetilde{X}}^2}} = \frac{(\widetilde{X} - \bar{x})\sqrt{n}}{\sqrt{\widetilde{\Sigma}_X^2}}, \tag{16.3}$$

учтя связь $$\widetilde{\Sigma}_{\widetilde{X}}^2 = \frac{1}{n}\widetilde{\Sigma}_X^2. \tag{16.4}$$

При этом оказалось, что T имеет закон распределения, не зависящий от параметров $\bar{x}$ и закона распределения нормальной случайной величины X. Действительно, заменив оценки $\widetilde{X}$ и $\widetilde{\Sigma}_X^2$ их выражениями, полученными ранее:

$$\tilde{X} = \frac{1}{n}\sum_{i=1}^{n} x_i; \qquad \tilde{\Sigma}^2_X = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \tilde{x})^2, \tag{16.5}$$

получим

$$T = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(X_i - \bar{x}\right)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \tilde{x}\right)^2}} \tag{16.6}$$

Проведя ряд линейных преобразований над суммами в этом выражении, можно прийти к другой записи случайной величины T, но теперь уже через новые нормальные случайные величины

$$\xi_0, \xi_1, \ldots \xi_{n-1}, \tag{16.7}$$

которые составлены из данных случайных величин. При этом, каждая из них центрирована $(M(\xi_i) = 0)$ и нормирована $(D(\xi_i) = 1)$.

Новая запись имеет вид

$$T = \frac{\sqrt{n-1}\xi_0}{\sqrt{\sum_{i=1}^{n-1}\xi_i^2}} \tag{16.8}$$

Как видно, новая запись случайной величины T не содержит параметров закона распределения случайной величины X и, следовательно, закон распределения случайной величины T зависит только от объема выборки n.

Методами теории вероятности Стьюдент получил плотность вероятности этой случайной величины T, носящей название "Закон распределении Стьюдента с (n- I) степенью свободы",

$$f_{n-1}(t)=\frac{\Gamma\left(\frac{n}{2}\right)\left(1+\frac{t^2}{n-2}\right)^{-\frac{n}{2}}}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)}, \tag{16.9}$$

где

$$\Gamma(n)=(n-1)! \tag{16.10}$$

для всех натуральных чисел. Теперь получим доверительный интервал для $\overline{X}$, выполнив очевидные преобразования,

$$\begin{aligned}\alpha &= P\left\{\left|\widetilde{X}-\overline{x}\right|\le\varepsilon\right\}=P\left\{\left|\frac{\widetilde{X}-\overline{x}}{\sqrt{\widetilde{\Sigma}^2_{\widetilde{X}}}}\right|\le\frac{\varepsilon}{\sqrt{\widetilde{\sigma}^2_{\widetilde{X}}}}\right\}= \\ &= P\left\{|\mathrm{T}|\le t_\alpha\right\}=P\left\{-t_\alpha\le\mathrm{T}\le t_\alpha\right\}=\int_{-t_\alpha}^{t_\alpha} t_{n-1}(t)dt=2S(t_\alpha)-1\end{aligned} \tag{16.11}$$

где функция $S(t_\alpha)=P\{T<t_\alpha\}$ табулирована (см. Приложение) и называется функцией распределения закона Стьюдента, а

$$t_{\alpha,n-1}=\frac{\varepsilon}{\sqrt{\widetilde{\sigma}^2_{\widetilde{x}}}}=\frac{\sqrt{n}\varepsilon}{\sqrt{\widetilde{\sigma}^2_{x}}}. \tag{16.12}$$

Следовательно, на основании таблиц закона распределения Стьюдента по заданному α и числу степеней свободы (n-1) можно отыскать число $t_{\alpha,n-1}$, для которого справедливо равенство

$$P(-t_{\alpha,n-1}\le\frac{\widetilde{X}-\overline{x}}{\sqrt{\widetilde{\Sigma}^2_X}}\sqrt{n}\le t_{\alpha,n})=\alpha \tag{16.13}$$

или

$$P(\widetilde{X}-\frac{\sqrt{\widetilde{\Sigma}^2_X}}{\sqrt{n}}t_{\alpha,n-1}\le\overline{x}\le\widetilde{X}+\frac{\sqrt{\widetilde{\Sigma}^2_X}}{\sqrt{n}}t_{\alpha,n-1})=\alpha. \tag{16.14}$$

Таким образом, случайный интервал $(\widetilde{x} - \frac{\sqrt{\widetilde{\Sigma}_X^2}}{\sqrt{n}} t_{\alpha,n-1}; \widetilde{x} + \frac{\sqrt{\widetilde{\Sigma}_X^2}}{\sqrt{n}} t_{\alpha,n-1})$ есть доверительная оценка с мерой надежности (с вероятностью) $P = \alpha$. При этом говорят что конкретный интервал $(\widetilde{x} - \frac{\sqrt{\widetilde{\sigma}_X^2}}{\sqrt{n}} t_{\alpha,n-1}; \widetilde{x} + \frac{\sqrt{\widetilde{\sigma}_X^2}}{\sqrt{n}} t_{\alpha,n-1})$ приблизительно с вероятностью α содержит математическое ожидание $\bar{x}$.

В химии принято представлять результаты для α = 0,95, т.е. принят 95%-ный доверительный интервал. Величина $(1-\alpha)$ называется уровнем значимости, иногда в таблицах указывается именно она.

Для определения доверительного интервала для оценки $\widetilde{\sigma}_X^2$ дисперсии случайной величины X построим новую случайную величину ξ.

$$\xi = \frac{\widetilde{\Sigma}_X^2}{\sigma_X^2}(n-1) \qquad (16.15)$$

которая, как можно строго показать, подчиняется закону распределения χ^2 (в данном случае с m=(n–I) степенью свободы) $f_{n-1}(\xi)$.

Здесь можно действовать аналогично задаче №2. Из таблиц по заданному α и m=n–1 степеням свободы определяют два числа δ_1 и δ_2 такие,что

$$P(\chi^2 < \delta_1) = 1 - P(\chi^2 > \delta_1) = \frac{1-\alpha}{2}, \qquad P(\chi^2 > \delta_2) = \frac{1-\alpha}{2},$$

$$\delta_1 = \chi^2_{(1+\alpha)/2}, \qquad \delta_2 = \chi^2_{(1-\alpha)/2}. \qquad (16.16)$$

где $(1+\alpha)/2$ и $(1-\alpha)/2$ соответствующие уровни значимости.

Например, для α=0.98 и m=18 получаем, что $\delta_1 = \chi^2_{0.99} = 7.01$ и $\delta_2 = \chi^2_{0.01} = 34.8$.

Числа δ_1 и δ_2 определены так, что

$$P(\delta_1 < \frac{(n-1)\widetilde{\Sigma}^2_X}{\sigma^2_X} < \delta_2) = \alpha \qquad (16.17)$$

или

$$P(\frac{(n-1)\widetilde{\Sigma}^2_X}{\delta_2} < \sigma^2_X < \frac{(n-1)\widetilde{\Sigma}^2_X}{\delta_1}) = \alpha . \qquad (16.18)$$

Таким образом, $(\frac{(n-1)\widetilde{\Sigma}^2_X}{\delta_2}, \frac{(n-1)\widetilde{\Sigma}^2_X}{\delta_1})$ есть доверительная оценка σ^2_X с мерой надежности α.

Окончательно результат записывают в виде α–доверительного интервала для σ^2_X ,

$$\frac{(n-1)\widetilde{\sigma}^2_X}{\delta_2} < \sigma^2_X < \frac{(n-1)\widetilde{\sigma}^2_X}{\delta_1} \qquad (16.19)$$

Аналогично для среднеквадратичного отклонения σ_X ,

$$\sqrt{\frac{n-1}{\delta_2}}\widetilde{\sigma}_X < \sigma_X < \sqrt{\frac{n-1}{\delta_1}}\widetilde{\sigma}_X , \qquad (16.20)$$

Где $\widetilde{\sigma}_X$ – вычисляют по формуле (12.6) или приближенно полагают равной $\sqrt{\widetilde{\sigma}^2_X}$. При этом говорят, что приблизительно с вероятностью α в соответствующих интервалах (16.19) и (16.20) находятся соответствующие параметры σ^2_X и σ_X .

17. ПРОЦЕДУРЫ

Задача № 1

1. $x_1, x_2, \ldots x_n$
2. $x = \frac{1}{n}\sum_{i=1}^{n} x_i$
3. $\sigma_X = \frac{1}{\sqrt{n}}\sigma_X$
4. Задаем $\alpha = 0{,}95$; ищем в таблицах функции Лапласа по α значениям аргумента $Z_\alpha : \alpha = \Phi(Z)$

$\alpha=\Phi(Z)$			0,95	
Z			Z_α	

Функция Лапласа

5. Вычисляем погрешность ε_α при заданной доверительной вероятности α:

$$\varepsilon_\alpha = Z_\alpha \cdot \sigma_{\bar{X}} = Z_\alpha \frac{1}{\sqrt{n}}\sigma_X$$

6. Представляем результат:

$\tilde{x} - \varepsilon_\alpha \le \bar{x} \le \tilde{x} + \varepsilon_\alpha$ $\qquad \bar{x} = \tilde{x} \pm \varepsilon_\alpha$

Задача № 2.

1. $x_1, x_2, \ldots x_n$
2. $\tilde{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$; $\qquad \tilde{\sigma}_X = k_{n+1}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
3. По заданному $\alpha = 0{,}95$ и m=n в таблицах χ^2-

распределения находим $\delta_1 = \chi^2_{(1+\alpha)/2} = \chi^2_{0.975}$ и $\delta_2 = \chi^2_{(1-\alpha)/2} = \chi^2_{0.025}$.

α / m=n-1	…	Для $(1+\alpha)/2$	Для $(1-\alpha)/2$	…	…
…					
…		δ_1	δ_2		
…					

4. Представление результата для σ_X^2:

$$\frac{n}{\delta_2}\tilde{\sigma}_X^2 \le \sigma_X^2 \le \frac{n}{\delta_1}\tilde{\sigma}_X^2, \text{ при } \alpha = 0{,}95$$

5. Представление результата для σ_x:

$$\sqrt{\frac{n}{\delta_2}}\tilde{\sigma}_X \le \sigma_X \le \sqrt{\frac{n}{\delta_1}}\tilde{\sigma}_X, \text{ при } \alpha = 0{,}95$$

Задача № 3

1. $x_1, x_2, \ldots..x_n$
2. $\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \tilde{\sigma}_X^2 = \frac{1}{n-1}\sum_{i-1}^{n}\left(x_i - \tilde{x}\right)^2;$

$$\tilde{\sigma}_X = k_n\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \tilde{x}\right)^2}$$

3. По таблицам Стьюдента по заданному α=0.95 и числу степеней свободы m=(n-1) находим значение $t_{\alpha,n-1}$:

1- α		0.1	0.05	0.02	

m=n-1					
…					
…			$t_{1-\alpha,m}$		

4. Вычисляем α-доверительную погрешность ε_α по формуле :

$$\varepsilon_\alpha = \frac{\sqrt{\widetilde{\sigma}_X^2}}{\sqrt{n}} t_{\alpha,n-1};$$

Если встречается обозначение $S_0 = \sqrt{\widetilde{\sigma}_X^2}$, тогда

$$\varepsilon_\alpha = t_{\alpha,n-1} \cdot S_0 / \sqrt{n}$$

5. Представление результата:

а) $\widetilde{x} - \varepsilon_\alpha \le \bar{x} \le \widetilde{x} + \varepsilon_\alpha$; $\quad \bar{x} = \widetilde{x} \pm \varepsilon_\alpha$; $\quad \bar{x} = \widetilde{x} \pm t_{\alpha,n-1} S_0 / \sqrt{n}$

б) или иначе: $\widetilde{x}, S_0, n.$

6. По заданному $\alpha = 0{,}95$ и числу степеней m=(n-1) в таблицах χ^2-распределения находим $\delta_1 = \chi^2_{(1+\alpha)/2} = \chi^2_{0.975}$ и $\delta_2 = \chi^2_{(1-\alpha)/2} = \chi^2_{0.025}$

α / m=n-1	…	Для $(1+\alpha)/2$	Для $(1-\alpha)/2$	…	…
…					
…		δ_1	δ_2		
…					

7. Представление результата для σ_X^2:

$$\frac{n-1}{\delta_2}\widetilde{\sigma}_X^2 \le \sigma_X^2 \le \frac{n-1}{\delta_1}\widetilde{\sigma}_X^2, \qquad \text{при } \alpha = 0{,}95$$

8. Представление результата для σ_X:

$$\sqrt{\frac{n-1}{\delta_2}}\widetilde{\sigma}_X \le \sigma_X \le \sqrt{\frac{n-1}{\delta_1}}\widetilde{\sigma}_X \qquad \text{при } \alpha = 0{,}95$$

18. ЧИСЛЕННЫЙ ПРИМЕР

Задача №2.

Угадывание на расстоянии длины 30-и сантиметровой линейки разными людьми.

$$\bar{x} = 30; \sigma_X^2 = ?$$

1.

X_1	X_2	X_3	X_4	X_5
42	17	33	45	25

см n=5

2.

$$\widetilde{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) = \frac{1}{5}\left\{(42-30)^2 + (17-30)^2 + (33-30)^2 + (45-30)^2 + (25-30)^2\right\} =$$

$$= 114{,}4 \cong 114$$

$$\widetilde{\sigma}_X = k_{n+1}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = k_{5+1}\sqrt{\widetilde{\sigma}_X^2} = 1{,}0506 \cdot \sqrt{114{,}4} = 1{,}0506 \cdot 10{,}696 =$$

$$= 11{,}237 \cong 11$$

3. По $\alpha = 0{,}95$ и m=$n = 5$ находим $\delta_1 == \chi^2_{0.975} = 0.831$ и $\delta_2 == \chi^2_{0.025} = 12.8$

4. Представление результата для σ_X^2:

$$\frac{n}{\delta_2}\widetilde{\sigma}_X^2 \le \sigma_X^2 \le \frac{n}{\delta_1}\widetilde{\sigma}_X^2 \quad \Rightarrow \quad \frac{5}{12.8}114 \le \sigma_X^2 \le \frac{5}{0.831}114$$

$$44.5 \le \sigma_X^2 \le 685 \quad \text{при}$$

$$\alpha = 0{,}95$$

5. Представление результата для σ_X:

$$\sqrt{\frac{n}{\delta_2}}\widetilde{\sigma}_X \le \sigma_X \le \sqrt{\frac{n}{\delta_1}}\widetilde{\sigma}_X \quad \Rightarrow \quad \sqrt{44.5} \le \sigma_X \le \sqrt{685}$$

$$6.7 \le \sigma_X \le 26.2 \quad \text{при}$$

$$\alpha = 0{,}95$$

Задача №1.

1. Выборка та же, что и в задаче 2. Однако здесь $\sigma_X = 11; \bar{x} = ?$
2. $\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(42 + 17 + 33 + 45 + 25) = 32{,}4 \cong 32$
3. $\sigma_{\tilde{X}} = \frac{1}{\sqrt{n}}\sigma_X = \frac{1}{\sqrt{5}}11 = 4{,}92 \cong 5$
4. По $\alpha = 0{,}95;\ z_{0{,}95} = 1{,}96$
5. Вычисление погрешности $\varepsilon_{0{,}95}$:

   $\varepsilon_{0{,}95} = Z \cdot \sigma_{\tilde{X}} = 1{,}96 \cdot 4{,}92 = 9{,}64 \cong 10$
6. Представление результата:

   $22 \le \bar{x} \le 42;\quad \bar{x} = 32 \pm 10$ при $\alpha = 0{,}95$
7. Использование « правил 3σ»:

   $\tilde{x} \pm 3\sigma_{\tilde{x}} = 32 \pm 15$ с вероятностью $\alpha = 0{,}9973$

Задача № 3

1. Выборка та же, что и в задачах №1,2, но здесь $\bar{x} = ?$ и $\sigma_X^2 = ?$
2. $\tilde{x} = 32$ (вычислено в задаче №2)

$$\tilde{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \tilde{x})^2 = \frac{1}{5-1}\{(42-32)^2 + (17-32)^2 + (33-32)^2 + (45-32)^2 + (25-$$

$$\tilde{\sigma}_X^2 = 136$$

$$\widetilde{\sigma}_X = k_n \sqrt{\widetilde{\sigma}_X^2} = k_5 \sqrt{136} = 1{,}064 \cdot 11{,}7 = 12{,}45 \cong 12$$

$$\widetilde{\sigma}_X = 12$$

3. По $\alpha = 0{,}95$; n-1=4 (число степеней свободы) $\Rightarrow t = 2{,}77$

4. $\varepsilon_{0{,}95} = \dfrac{\sqrt{\widetilde{\sigma}_X^2}}{\sqrt{n}} t_{\alpha, n-1} = \dfrac{11{,}7}{5} 2{,}77 = 14{,}49 \cong 14$, $S_0 = \sqrt{\widetilde{\sigma}_X^2} = 11{,}7 \cong 12$

5. Представление результата $\bar{x}$:

а) $18 \le \bar{x} \le 46$; при α=0.95

б) $\widetilde{x} = 32$; $S_0 = 11{,}7$; n=5 при α=0.95

6. По m=n-1=4 и $\alpha = 0{,}95$ находим $\delta_1 == \chi^2_{0.975} = 0.484$, $\delta_2 == \chi^2_{0.025} = 11.1$

7. Представление результата для σ_X^2:

$$\frac{n-1}{\delta_2}\widetilde{\sigma}_X^2 \le \sigma_X^2 \le \frac{n-1}{\delta_1}\widetilde{\sigma}_X^2 \quad \Rightarrow$$

$$\frac{4}{11.1}136 \le \sigma_X^2 \le \frac{4}{0.484}136$$

$49 \le \sigma_X^2 \le 1124$ при $\alpha = 0{,}95$

8. Представление результата для σ_X:

$$\sqrt{\frac{n-1}{\delta_2}}\tilde{\sigma}_X \le \sigma_X \le \sqrt{\frac{n-1}{\delta_1}}\tilde{\sigma}_X \quad \Rightarrow \quad \sqrt{49} \le \sigma_X \le \sqrt{1124}$$

$$7 \le \sigma_X \le 33 \quad \text{при}$$

$$\alpha = 0{,}95$$

19. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Если между измеряемыми величинами существует функциональная зависимость, вид которой известен с точностью до численных значений некоторых параметров, то для определения оценок этих параметров применяется метод наименьших квадратов.

Рассмотрим этот метод на примере двух величин: y и x, связанных функциональным соотношением

$$y = \varphi(x; a_1, a_2, ..., a_m), \tag{1}$$

где вид функции φ известен, а постоянные $a_l (l = 1,2,..., m)$ подлежат определению из опыта. Предположим, что для нахождения этих постоянных при n значениях x_i аргумента x произведено n измерений значений функции $y_i (i = 1,2,..., n)$. Так как измерения сопровождаются ошибками, то получим не n пар чисел (y_i, x_i), а значения этих чисел, искаженные ошибками измерения. Обозначим результаты измерения, соответствующие числам (y_i, x_i), через (η_i, ξ_i) соответственно. Если бы величины (η_i, ξ_i) были свободны от ошибок, то каждая пара этих чисел удовлетворяла бы соотношению (1), т.е. выполнялись бы следующие равенства

$$\eta_i = \varphi(\xi_i; a_1, a_2, ..., a_m), i = 1,2,..., n \ . \tag{2}$$

В действительности из-за ошибок измерения соотношения (1) выполняется не строго (не выполняется), поэтому их называют "условными уравнениями".

Назовем разность между левой и правой частями условного $i-$ го уравнения через ε_i и будем называть её "невязкой $i-$ го измерения ",

$$\varepsilon_i = \eta_i - \varphi(\xi_i; a_1, a_2, ..., a_m), i = 1,2,..., n \ . \tag{3}$$

Если $n < m$ то число условных уравнений меньше общего числа неизвестных параметров a_j и оценки этих параметров определить невозможно. Если $n = m$, то число условных уравнений равно числу неизвестных параметров и задача нахождения их оценок сводится к задаче решения системы уравнений. При $n > m$ возникает задача

использования этих уравнений для получения наилучших оценок параметров a_j. Практически эта задача возникает в тех случаях, когда увеличением числа измерений пытаются скомпенсировать неизбежные ошибки каждого измерения.

В методе наименьших квадратов постоянные $a_l (l = 1,2,...,m)$ определяются так, чтобы сумма квадратов невязок обращалась в минимум, т.е. чтобы выполнялось условие

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \min. \qquad (4)$$

Сумма S является функцией неизвестных параметров a_j, поэтому из условия (4) следует обращение в нуль всех первых производных от S по a_j, т.е. выполнение равенств

$$\frac{\partial S}{\partial a_j} = 0, j = 1,2,...,m. \qquad (5)$$

Система уравнений (5), называемая "системой нормальных уравнений", содержит m уравнений, т.е. столько же, сколько неизвестных параметров a_j. Поэтому решение $a_1, a_2, ..., a_m$ существует!

Метод наименьших квадратов в общем случае не обеспечивает получения наиболее эффективных оценок, однако его применение оправдывается как простотой вычислительной схемы, так и тем, что в ряде случаев этот метод действительно даёт наилучшие оценки.

В науке и технике широко распространены при описании экспериментальных данных аппроксимирующие уравнения вида

$$Y = a_0 x_0 + a_1 x_1 + ... + a_l x_l = \sum_{q}^{l} a_q x_q, \qquad (6)$$

где y – изучаемая функция от переменных $x_0, x_1 ... x_l$.

Остановимся на двух первых случаях, когда измеряется постоянная величина $Y = a_0$, $(x_0 = 1, a_1 = a_2 = \ldots = 0)$. и когда имеем линейное уравнение $Y = a_0 + a_1 x_1$, $(x_0 = 1, a_2 = a_3 = \ldots = 0)$.

Первый случай разобран выше – это задача № 3 :

постоянная величина, $Y = a_0$,

1) $$\tilde{y} \pm t_{0.05} \cdot S_0 / \sqrt{n} \qquad (7)$$

где $$\tilde{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \tilde{\sigma}_X^2 = S_0^2 = \frac{1}{n-1}\sum_{i-1}^{n}\left(y_i - \tilde{y}\right)^2;$$

$t_{0.05}$ – коэффициент Стьюдента для $f = n-1$ степени свободы и для уровня значимости 0.05 (т.е. для 95% – ой вероятности).

2) Однако, если не стандартизовать 95% – ый доверительный интервал, как меру для выражения погрешности, то результат можно представлять в виде тройки чисел :

$$\tilde{y}, S_0^2, n. \qquad (8)$$

Второй случай – это случай линейного уравнения $Y = a + bx$, сглаживающего по условию (4) экспериментальные данные $\{y_i, x_i\}$. Можно показать, что это уравнение эквивалентно уравнению

$$Y = \tilde{y} + b(x - \tilde{x}), \qquad (9)$$

где

$$\tilde{y} = \frac{1}{n}\sum_{i=1}^{n} y_i; \quad \tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i;$$

$$b = \frac{1}{B}\sum_{i=1}^{n} y_i (x_i - \tilde{x}), \quad B = \sum_{i=1}^{n} (x_i - \tilde{x})^2,$$

(заметим, что здесь $a = \widetilde{y} - b\widetilde{x}$). Поэтому, результат измерений следует представлять или как "95%–ую доверительную дорогу" на плоскости xOy :

$$Y = \widetilde{y} + b(x - \widetilde{x}) \pm t_{0.05} \cdot S_0 \cdot \sqrt{\frac{1}{n} + \frac{(x - \widetilde{x})^2}{B}} \qquad (10)$$

или как шестерку чисел:

$$\widetilde{y}, n, \widetilde{x}, b, B, S_0^2, \qquad (11)$$

где $S_0^2 = \frac{1}{n-2} \sum_{i-1}^{n} \left(y_i - Y_i\right)^2$.

(Число степеней свободы f в рассматриваемом случае равно $f = n - 2$).

Заметим, что величина $t_{0.05}$ – двусторонний критерий Стьюдента при уровне значимости 0.05 зависит от f – степени свободы и при $f \geq 8 - 12$ мало отличается от 2. На этом основана рекомендация проводить измерений столько, чтобы f была больше $8 - 12$, тогда

$$t_{0.05} \cdot S_0 \approx 2S_0 . \qquad (12)$$

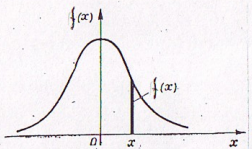
20. Литература

1. Свешников А.А. Основы теории ошибок.- Л., Изд-во Ленинград. ун-та, 1972.
2. Вентцель Е.С. Теория вероятностей. М., Наука, 1964.
3. Крамер Г. Математические методы статистики. М., Мир, 1975.
4. Линник Ю.В. Метод наименьших квадратов и основы теории обработки наблюдений.- Физматгиз, 1962.
5. Задачник по теории вероятностей, математической статистике и теории случайных функций. Под ред. А.А.Свешникова. М., Наука. 1970

21. Приложение

На страницах 74, 75 буквой α обозначен уровень значимости, который в тексте обозначен как $1 - \alpha$.

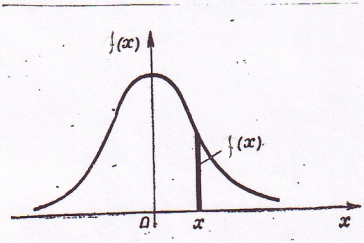
В литературе можно найти таблицы построенные и с помощью доверительной вероятности α, и с помощью уровня значимости $1 - \alpha$.



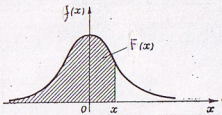
Плотность распределения $\varphi(x)$ нормированного нормального распределения

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	.39894	.39892	.39886	.39876	.39862	.39844	.39822	.39797	.39767	.39733
0,1	.39695	.39654	.39608	.39559	.39505	.39448	.39387	.39322	.39253	.39181
0,2	.39104	.39024	.38940	.38853	.38762	.38667	.38568	.38466	.38361	.38251
0,3	.38139	.38023	.37903	.37780	.37654	.37524	.37391	.37255	.37115	.36973
0,4	.36827	.36678	.36526	.36371	.36213	.36053	.35889	.35723	.35553	.35381
0,5	.35207	.35029	.34849	.34667	.34482	.34294	.34105	.33912	.33718	.33521
0,6	.33322	.33121	.32918	.32713	.32506	.32297	.32086	.31874	.31659	.31443
0,7	.31225	.31006	.30785	.30563	.30339	.30114	.29887	.29659	.29431	.29200
0,8	.28969	.28737	.28504	.28269	.28034	.27798	.27562	.27324	.27086	.26848
0,9	.26609	.26369	.26129	.25888	.25647	.25406	.25164	.24923	.24681	.24439
1,0	.24197	.23955	.23713	.23471	.23230	.22988	.22747	.22506	.22265	.22025
1,1	.21785	.21546	.21307	.21069	.20831	.20594	.20357	.20121	.19886	.19652
1,2	.19419	.19186	.18954	.18724	.18494	.18265	.18037	.17810	.17585	.17360
1,3	.17137	.16915	.16694	.16474	.16256	.16038	.15822	.15608	.15395	.15183
1,4	.14937	.14764	.14556	.14350	.14146	.13943	.13742	.13542	.13344	.13147
1,5	.12952	.12758	.12566	.12376	.12188	.12051	.11816	.11632	.11450	.11270
1,6	.11092	.10915	.10741	.10567	.10396	.10226	.10059	.09893	.09728	.09566
1,7	.09405	.09246	.09089	.08933	.08780	.08628	.08478	.08329	.08183	.08038
1,8	.07895	.07754	.07614	.07477	.07341	.07206	.07074	.06943	.06814	.06687
1,9	.06562	.06438	.06316	.06195	.06077	.05959	.05844	.05730	.05618	.05508
2,0	.05399	.05292	.05186	.05082	.04980	.04879	.04780	.04682	.04586	.04491
2,1	.04398	.04307	.04217	.04128	.04041	.03955	.03871	.03788	.03706	.03626
2,2	.03547	.03470	.03394	.03319	.03246	.03174	.03103	.03034	.02965	.02898
2,3	.02833	.02768	.02705	.02643	.02582	.02522	.02463	.02406	.02349	.02294
2,4	.02239	.02186	.02134	.02083	.02033	.01984	.01936	.01889	.01842	.01797
2,5	.01753	.01709	.01667	.01625	.01585	.01545	.01506	.01468	.01431	.01394
2,6	.01358	.01323	.01289	.01256	.01223	.01191	.01160	.01130	.01100	.01071
2,7	.01042	.01014	.00987	.00961	.00935	.00909	.00885	.00861	.00837	.00814
2,8	.00792	.00770	.00748	.00727	.00707	.00687	.00668	.00649	.00631	.00613
2,9	.00595	.00578	.00562	.00545	.00530	.00514	.00499	.00485	.00471	.00457
	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3,0	.00443	.00327	.00238	.00172	.00123	.00087	.00061	.00042	.00029	.00020

47

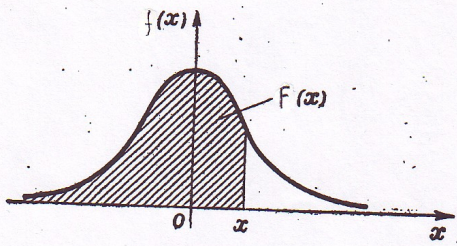


$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}$$



Функция распределения $F(x)$ нормированного нормального распределения

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	.500000	.503989	.507978	.511966	.515953	.519938	.523922	.527903	.531881	.535856
0,1	.539828	.543795	.547758	.551717	.555670	.559618	.563560	.567495	.571424	.575345
0,2	.579260	.583166	.587064	.590954	.594835	.598706	.602568	.606420	.610261	.614092
0,3	.617911	.621720	.625516	.629300	.633072	.636831	.640576	.644309	.648027	.651732
0,4	.655422	.659097	.662757	.666402	.670031	.673645	.677242	.680822	.684386	.687933
0,5	.691462	.694974	.698468	.702944	.705402	.708840	.712260	.715661	.719043	.722405
0,6	.725747	.729069	.732371	.735653	.738914	.742154	.745373	.748571	.751748	.754903
0,7	.758036	.761148	.764238	.767305	.770350	.773373	.776373	.779350	.782305	.785236
0,8	.788145	.791030	.793892	.796731	.799546	.802338	.805106	.807850	.810570	.813267
0,9	.815940	.818589	.821214	.823814	.826391	.873944	.831472	.833977	.836457	.838913
1,0	.841345	.843752	.846136	.848495	.850830	.853141	.855428	.857690	.859929	.862143
1,1	.864334	.866500	.868643	.870762	.872857	.874928	.876976	.879000	.881000	.882977
1,2	.884930	.886861	.888768	.890651	.892512	.894350	.896165	.897958	.899727	.901475
1,3	.903200	.904902	.906582	.908241	.909877	.911492	.913085	.914656	.916207	.917736
1,4	.919243	.920730	.922196	.923642	.925066	.926471	.927855	.929219	.930563	.931889
1,5	.933193	.934478	.935744	.936992	.938220	.939429	.940620	.941792	.942947	.944083
1,6	.945201	.946301	.947384	.948449	.949497	.950528	.951543	.952540	.953521	.954486
1,7	.955434	.956367	.957284	.958185	.959070	.959941	.960796	.961636	.962462	.963273
1,8	.964070	.964852	.965620	.966375	.967116	.967843	.968557	.969258	.969946	.970621
1,9	.971283	.971933	.972571	.973197	.973810	.974412	.975002	.975581	.976148	.976704
2,0	.977250	.977784	.978308	.978822	.979325	.979818	.980301	.980774	.981237	.981691
2,1	.982136	.982571	.982997	.983414	.983823	.984222	.984614	.984997	.985371	.985738
2,2	.986097	.986447	.986791	.987126	.987454	.987776	.988089	.988396	.988696	.988989
2,3	.989276	.989556	.989830	.990097	.990358	.990613	.990862	.991106	.991344	.991576
2,4	.991802	.992024	.992240	.992451	.992656	.992857	.993053	.993244	.993431	.993613
2,5	.993790	.993963	.994132	.994297	.994457	.994614	.994766	.994915	.995060	.995201
2,6	.995339	.995473	.995604	.995731	.995855	.995975	.996093	.996207	.996319	.996427
2,7	.996533	.996636	.996736	.996833	.996928	.997020	.997110	.997197	.997282	.997365
2,8	.997445	.997523	.997599	.997673	.997744	.997814	.997882	.997948	.998012	.998074
2,9	.998134	.998193	.998250	.998305	.998359	.998411	.998462	.998511	.998559	.998605
	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3,0	.998650	.999032	.999313	.999517	.999663	.999767	.999841	.999892	.999928	.999952



$$F(x) = \frac{1}{2}\left[1 + \Phi\left(\frac{x-\bar{x}}{\sigma}\right)\right],$$

$\Phi\left(\frac{x-\bar{x}}{\sigma}\right)$ – интегральная функция Лапласа,

$$\Phi(z) = \frac{2}{\sqrt{2\pi}} \int_0^z e^{-\frac{t^2}{2}} dt .$$

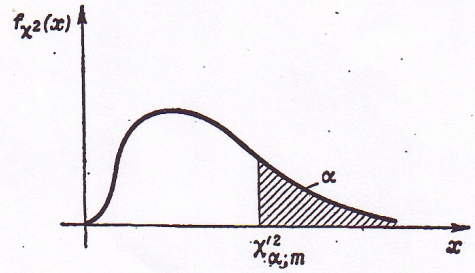
50

Значение интегральной функции Лапласа $\Phi(z)=\frac{2}{\sqrt{2\pi}}\int_0^z e^{-z^2/2}dz$

z	Φ (z)	z	Φ (z)
0	0	2.50	0.9876
0.25	0.1974	2.75	0.9940
0.50	0.3829	3.00	0.9973
0.75	0.5468	3.25	0.9988
1.00	0.6827	3.50	0.9995
1.25	0.7887	3.75	0.9998
1.50	0.8664	4.00	0.999937
1.75	0.9199	4.25	0.999965
2.00	0.9545	4.50	0.999993
2.25	0.9756	4.75	0.999997

Значения $\chi^2_{\alpha;m}$ для различных α и m

Число степеней свободы m	Уровень значимости α					
	0,99	0,975	0,95	0,05	0,025	0,01
1	0,00016	0,00098	0,0039	3,8	5,0	6,6
2	0,020	0,051	0,103	6,0	7,4	9,2
3	0,115	0,216	0,352	7,8	9,4	11,3
4	0,297	0,484	0,711	9,5	11,1	13,3
5	0,554	0 831	1,15	11,1	12,8	15,1
6	0,872	1,24	1,64	12,6	14,4	16,8
7	1,24	1,69	2,17	14,1	16,0	18,5
8	1,65	2,18	2,73	15,5	17,5	20,1
9	2,09	2,70	3,33	16,9	19,0	21,7
10	2,56	3,25	3,94	18,3	20,5	23,2
11	3,05	3,82	4,57	19,7	21,9	24,7
12	3,57	4,40	5,23	21,0	23,3	26,2
13	4,11	5,01	5,89	22,4	24,7	27,7
14	4,66	5,63	6,57	23,7	26,1	29,1
15	5,23	6,26	7,26	25,0	27,5	30,6
16	5,81	6,91	7,96	26,3	28,8	32,0
17	6,41	7,56	8,67	27,6	30,2	33,4
18	7,01	8,23	9,39	28,9	31,5	34,8
19	7,63	8,91	10,1	30,1	32,9	36,2
20	8,26	9,59	10,9	31,4	34,2	37,6
21	8,90	10,3	11,6	32,7	35,5	38,9
22	9,54	11,0	12,3	33,9	36,8	40,3
23	10,2	11,7	13,1	35,2	38,1	41,6
24	10,9	12,4	13,8	36,4	39,4	43,0
25	11,5	13,1	14,6	37,7	40,6	44,3
26	12,2	13,8	15,4	38,9	41,9	45,6
27	12,9	14,6	16,2	40,1	43,2	47,0
28	13,6	15.3	16,9	41,3	44,5	48,3
29	14,3	16,0	17,7	42,6	45,7	49,6
30	15,0	16,8	18,5	43,8	47,0	50,9



Значения $t_{\alpha;\,m}$ и $t'_{\alpha;\,m}$ для различных α и m

Число степеней свободы m	Уровень значимости α (двухстороннее ограничение)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (одностороннее ограничение)					

