Теория измерений и метрология

Тема: Погрешности и неопределенности измерений

С.В. Муравьев
e-mail: muravyov@tpu.ru

Томский политехнический университет

9

Лекция

Интервальные оценки истинного значения измеряемой величины

- Стандартное нормальное распределение и функция Лапласа
- Доверительный интервал и доверительные границы

Интервальные оценки

Точечные оценки не дают всей требуемой информации о погрешности. Необходимо также оценивать ***достоверность*** результата измерения путем указания на

- ***границы интервала***, в котором находится погрешность с заранее заданной вероятностью, либо
- ***функцию распределения*** погрешности.

При этом пользуются понятиями

- ***доверительного интервала и***
- ***доверительной вероятности***.

Теория ***интервального оценивания*** была разработана Р. Фишером и Е. Нейманом в **1925-1935** гг.

Ежи Нейман (Jerzy Neyman) 1894-1981

Рональд Эйлмер Фишер (Sir Ronald Aylmer Fisher) 1890-1962

Стандартное нормальное распределение и функция Лапласа

Пусть случайная величина x распределена ***нормально***, то есть

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$ и

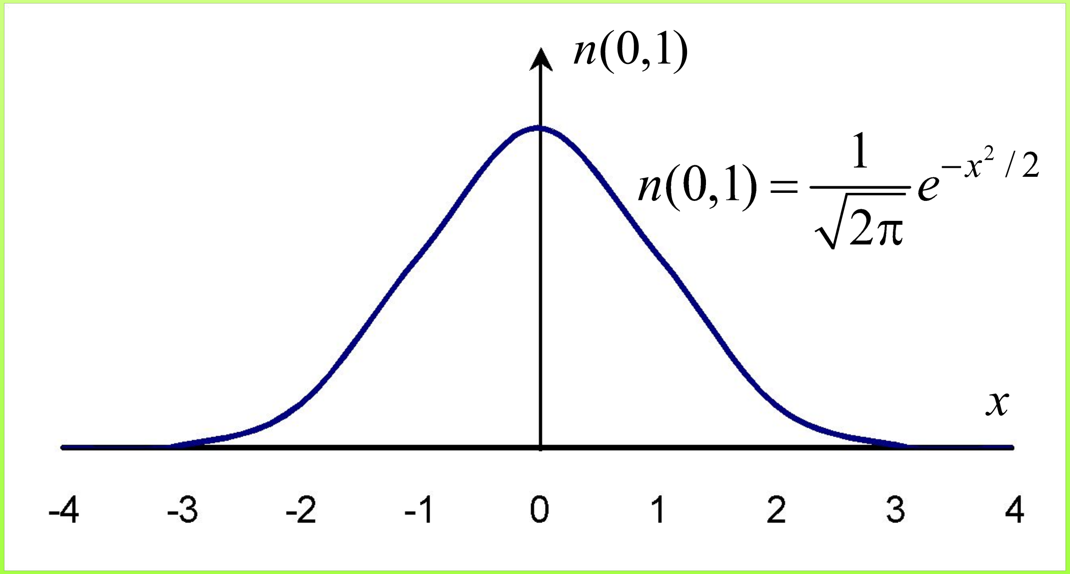
$$F(x) = \int_{-\infty}^{x} f(x)dx$$

Для нормально распределенной случайной величины введем обозначения:

$$f(x) = n(\mu, \sigma) \text{ и}$$
$$F(x) = N(\mu, \sigma)$$

При $\mu = 0$ и $\sigma = 1$ получается ***стандартное нормальное распределение*** $N(0,1)$

Функция $n(0,1)$ имеет вид, представленный на рисунке.



Для того, чтобы все возможные варианты нормального распределения свести к единому случаю $n(0,1)$, пользуются ***функцией Лапласа*** $\Phi(z)$.

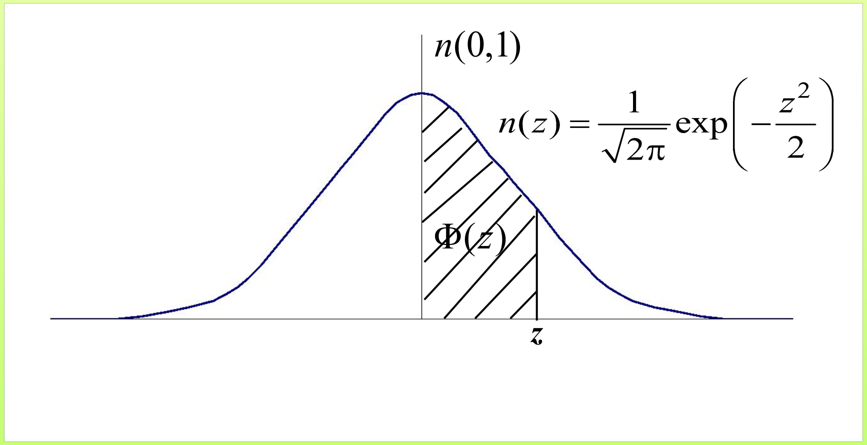
Пусть $z = \dfrac{x-\mu}{\sigma}$ – ***стандартная нормальная величина (квантильный коэффициент)***, тогда

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_0^z \exp\left(-\frac{z^2}{2}\right)dz.$$

Функция $\Phi(z)$ табулирована.

Пьер Симон де Лаплас
Pierre Simon de Laplace
(1749-1827)

Таблица значений функции Лапласа



z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0,1	0,040	1,1	0,364	2,1	0,482
0,2	0,079	1,2	0,385	2,2	0,486
0,3	0,118	1,3	0,403	2,3	0,489
0,4	0,155	1,4	0,419	2,4	0,492
0,5	0,192	1,5	0,433	2,5	0,494
0,6	0,226	1,6	0,445	2,6	0,495
0,7	0.258	1,7	0,455	2,7	0,497
0,8	0,288	1,8	0,464	2,8	0,497
0,9	0,316	1,9	0,471	2,9	0,498
1,0	0,341	2,0	0,477	3,0	0,499

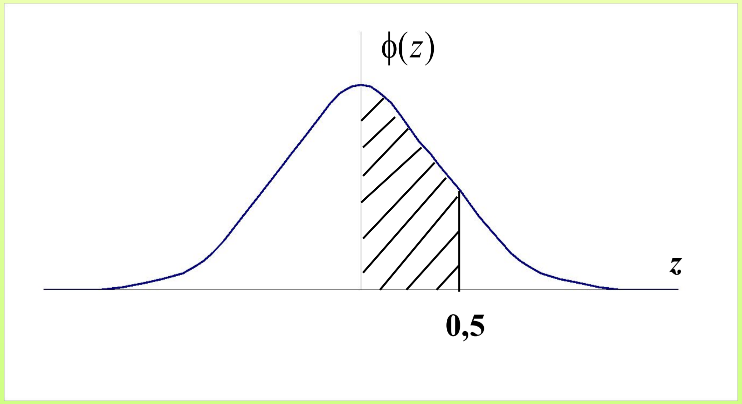
Свойства функции Лапласа:

- $\Phi(0) = 0$;
- $\Phi(-x) = -\Phi(x)$;
- $\Phi(\infty) = 0,5$;

При помощи этой таблицы можно непосредственно найти, например, следующие вероятности:

$$P(0 \le z \le 0,5) = \Phi(0,5) = 0,192$$

$$P(0 \le z \le 0,3) = \Phi(0,3) = 0,118$$



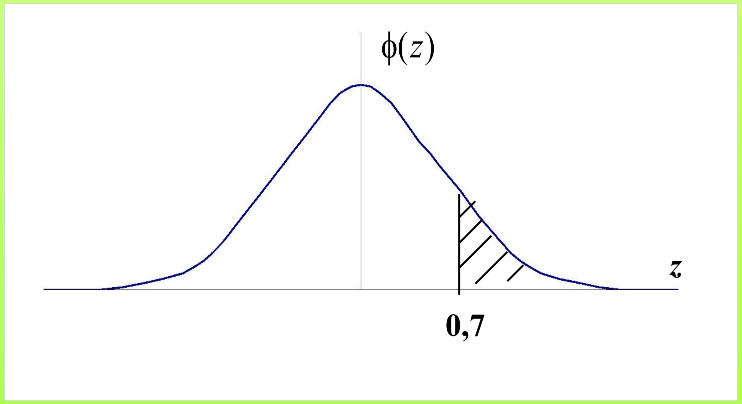
Из симметричности графика плотности распределения вероятности следует, что

$$P(z \ge 0) = P(z \le 0) = 0,5.$$

Поэтому

$$P(z \ge 0,7) = 0,5 - P(0 \le z \le 0,7) =$$

$$= 0,5 - \Phi(0,7) = 0,5 - 0,258 = 0,242$$

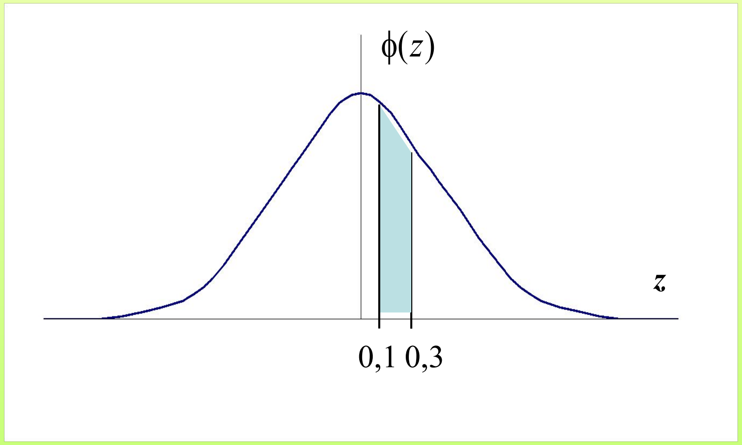


$$P(-0,4 \le z \le 0) = P(0 \le z \le 0,4) =$$

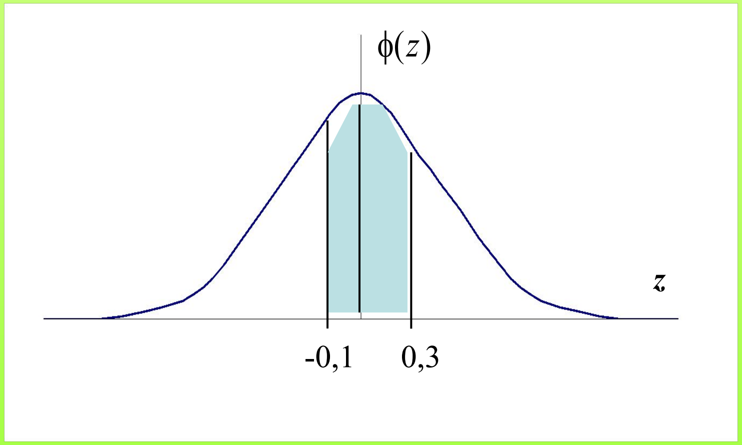
$$= \Phi(0,4) = 0,155$$

Другие случаи:

$$P(0{,}1 \le z \le 0{,}3) = P(-0{,}3 \le z \le -0{,}1) = \Phi(0{,}3) - \Phi(0{,}1) =$$
$$= 0{,}118 - 0{,}040 = 0{,}078$$



$$P(-0{,}1 \le z \le 0{,}3) = \Phi(0{,}3) + \Phi(0{,}1) =$$
$$= 0{,}118 + 0{,}040 = 0{,}158$$



Функция Лапласа выражает связь между ***любым*** нормальным и ***стандартным*** нормальным распределениями:

$$N(\mu, \sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Она позволяет вычислять ***вероятность принадлежности значения x случайной величины интервалу*** (a, b):

$$P(a < x < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Действительно, поскольку оба неравенства

$$a < x < b \qquad \text{и}$$

$$\frac{a-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{b-\mu}{\sigma}$$

выполняются (или не выполняются) всегда одновременно, то соответствующие вероятности равны друг другу, а именно:

$$P(a < x < b) = P\left(\frac{a-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) =$$

$$P\left(\frac{a-\mu}{\sigma} < z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Пример.

Вес пачек с сахаром, является нормально распределенной случайной величиной с математическим ожиданием 1 кг и стандартным отклонением 10 г. Какой процент пачек имеет вес в интервале от 990 до 1020 г?

Решение:

Вес пачки сахара является случайной величиной, распределенной по закону $N(1000, 10)$.

$$P(990 < x < 1020) = P\left(\frac{990-1000}{10} < z < \frac{1020-1000}{10}\right) =$$

$$= P(-1 < z < 2) = \Phi(1) + \Phi(2) = 0,341 + 0,477 = 0,818.$$

Т.о., 82 % пачек имеют вес в заданном интервале.

Представляет интерес частный случай, когда $a=-\varepsilon,\ b=\varepsilon$:

$$P(-\varepsilon < x < \varepsilon) = P(|\,x\,| < \varepsilon) = P\left(\frac{-\varepsilon-\mu}{\sigma} < z < \frac{\varepsilon-\mu}{\sigma}\right) = \\ = \Phi\left(\frac{\varepsilon-\mu}{\sigma}\right) - \Phi\left(\frac{-\varepsilon-\mu}{\sigma}\right) = \Phi\left(\frac{\varepsilon-\mu}{\sigma}\right) + \Phi\left(\frac{\varepsilon+\mu}{\sigma}\right). \qquad (*)$$

Если $\mu = 0$, то

$$P(-\varepsilon < x < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

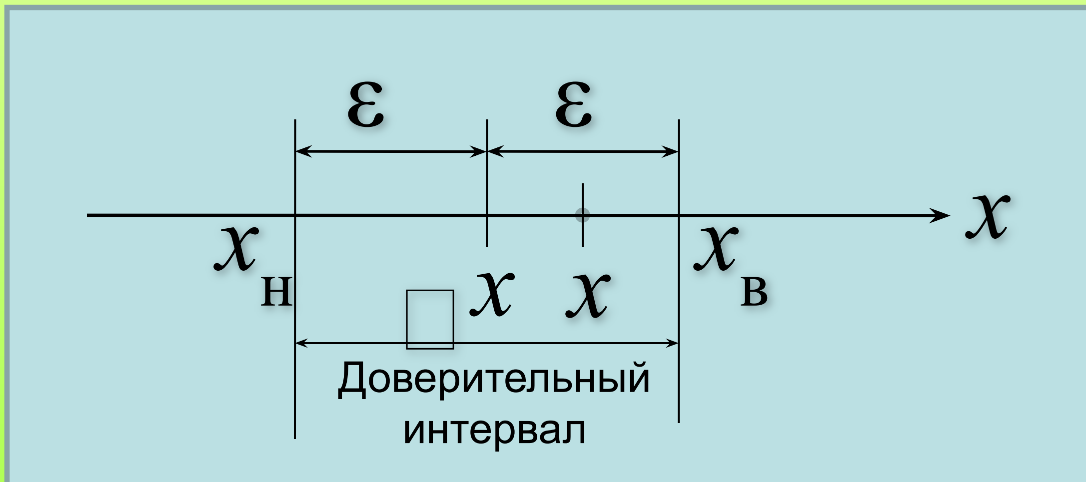
Доверительная вероятность, доверительный интервал и доверительные границы

Доверительный интервал – это интервал, между ***границами*** $x_{н}$ и $x_{в}$ которого с заданной ***доверительной вероятностью*** $P_{д}$ лежит истинное значение x,

т.е. $P(x_{н} < x < x_{в}) = P_{д}$ или

$P(\bar{x} - \varepsilon < x < \bar{x} + \varepsilon) = P_{д}$, или в ***терминах погрешности***

$$P(-\varepsilon < x - \bar{x} < \varepsilon) = P(-\varepsilon < \Delta < \varepsilon) = P(|\Delta| < \varepsilon) = P_{д} \qquad (*)$$



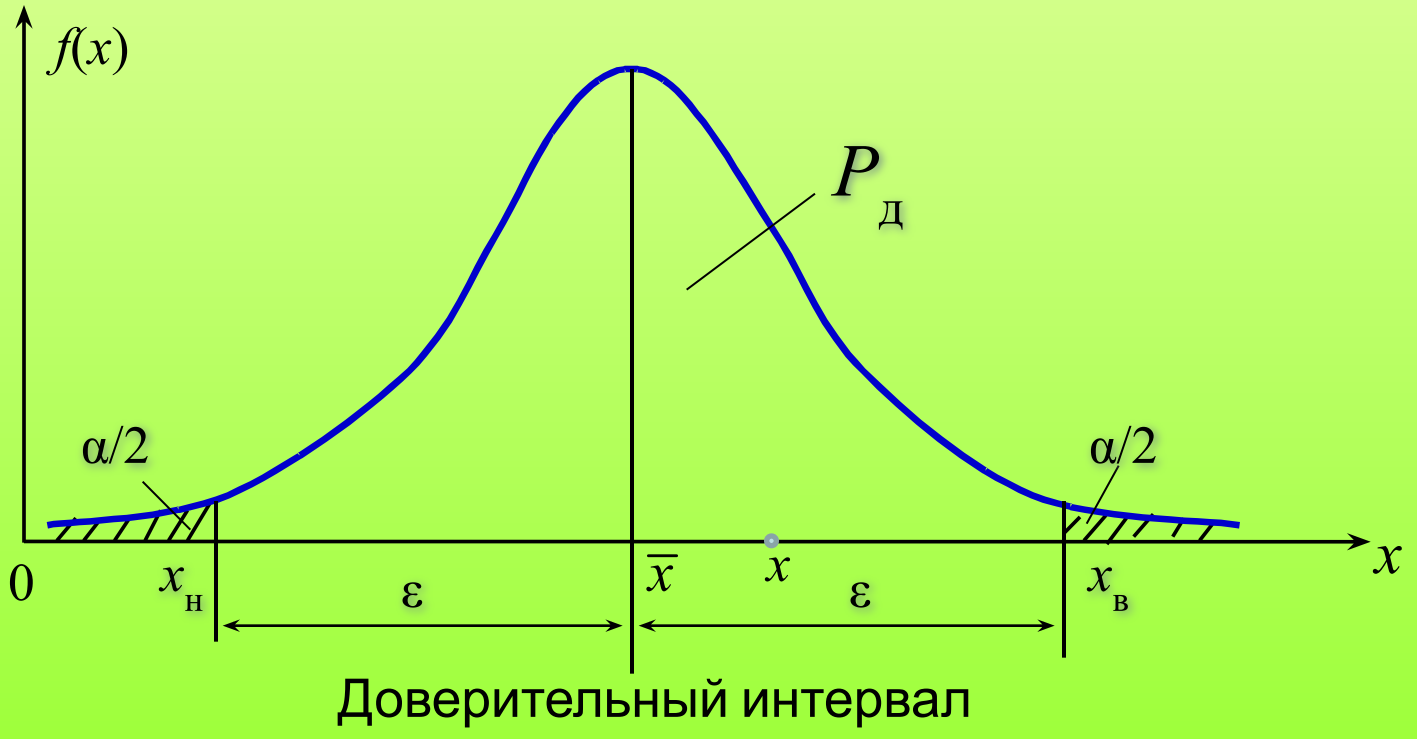
Т.о., доверительный интервал для ***результата измерения*** $\bar{x}$ является доверительным интервалом и для его ***погрешности*** Δ.

Наиболее часто вероятность $P_д$ берется равной **0,9; 0,95** или **0,99**.

Вероятность $P_д$ связана с ***уровнем значимости*** α:

$$P_д = 1 - \alpha.$$

- ***Доверительная вероятность*** показывает, какая часть (или какой процент) генеральной совокупности находится внутри границ заданного интервала.
- ***Уровень значимости*** показывает, что большие по абсолютной величине погрешности наблюдаются с вероятностью $\alpha = 1 - P_д$.



Пусть $y_1, y_2, \ldots, y_n$ – независимая выборка из нормального распределения $N(\mu,\sigma)$, где σ – известное СКО результатов измерений. Тогда случайная величина $\bar{X}$ имеет нормальное распределение $N(\bar{x}, \sigma/\sqrt{n})$.

Верхнюю и нижнюю доверительные границы можно тогда найти из уравнения:

$$P_{\text{д}} = P(\bar{x} - \varepsilon < x < \bar{x} + \varepsilon) = \Phi\left(\frac{\bar{x} + \varepsilon - \bar{x}}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\bar{x} - \varepsilon - \bar{x}}{\sigma/\sqrt{n}}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right),$$

где $z = \dfrac{\varepsilon\sqrt{n}}{\sigma}$ – аргумент функции Лапласа (квантильный множитель), отвечающий вероятности $\dfrac{P_{\text{д}}}{2}$.

С учетом того, что обычно вместо σ используется S, доверительная граница определяется по формуле:

$$\varepsilon = z\frac{S}{\sqrt{n}}.$$

Таким образом, доверительный интервал описывается аналитически следующим образом:

$$\boxed{\overline{x} - z\frac{S}{\sqrt{n}} < x < \overline{x} + z\frac{S}{\sqrt{n}}}.$$

Он соответствует доверительной вероятности:

$$\boxed{P(\overline{x} - z\frac{S}{\sqrt{n}} < x < \overline{x} + z\frac{S}{\sqrt{n}}) = P_{\text{д}} = 2\Phi(z)}.$$

Значение ***доверительной границы***, определяющей ***половину длины*** доверительного интервала, оценивается по формуле:

$$\boxed{\varepsilon = z\frac{S}{\sqrt{n}} = zS(\overline{x})}.$$

Пример

Произведено 50 измерений постоянного сопротивления R. Определить доверительный интервал для матожидания значения постоянного сопротивления. Закон распределения R нормальный с параметрами

$\overline{R}$ = 590 Ом, S = 90 Ом при доверительной вероятности P = 0,9.

Доверительный интервал определяется по формуле

$$P(\overline{x} - z\frac{S}{\sqrt{n}} < x < \overline{x} + z\frac{S}{\sqrt{n}}) = P_{\text{д}} = 2\Phi(z)$$

Отсюда $\Phi(z)$ = 0,45. Из таблицы функции Лапласа находим z = 1,65. Следовательно, доверительный интервал запишется в виде

$$590 - 1,65 \cdot 90/\sqrt{50} < R < 590 + 1,65 \cdot 90/\sqrt{50}$$

или 590 – 21 < R < 590 + 21, ε = 21 Ом. Окончательно имеем

569 Ом < R < 611 Ом или

R = (590 ± 21) Ом.

Пример

Погрешность измерения напряжения Δ распределена по нормальному закону, причем систематическая погрешность Δ_c равна 30 мВ, а СКО составляет 50 мВ. Найти вероятность того, что результат измерения отличается от истинного значения не более чем на 120 мВ.

Решение. Воспользуемся соотношением (*) на слайде 45:

$$P(|\Delta| < \varepsilon) = \Phi\left(\frac{\varepsilon - \mu}{\sigma}\right) + \Phi\left(\frac{\varepsilon + \mu}{\sigma}\right)$$

Поскольку число измерений n не указано, считаем, что дано СКО оценки матожидания $S(\overline{x}) = 50$ мВ. Подставляя в формулу значения $\varepsilon = 120$ мВ и $\Delta_c = 30$ мВ (в качестве матожидания), получаем

$$P = \Phi\left(\frac{120 - 30}{50}\right) + \Phi\left(\frac{120 + 30}{50}\right) = \Phi(1{,}8) + \Phi(3) = 0{,}464 + 0{,}499 = 0{,}963.$$

Пример (продолжение)

Предположим теперь, что систематическая погрешность отсутствует, скажем, были предприняты меры для того, чтобы $\Delta_c = 0$. Тогда получаем

$$P(|\Delta| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)$$

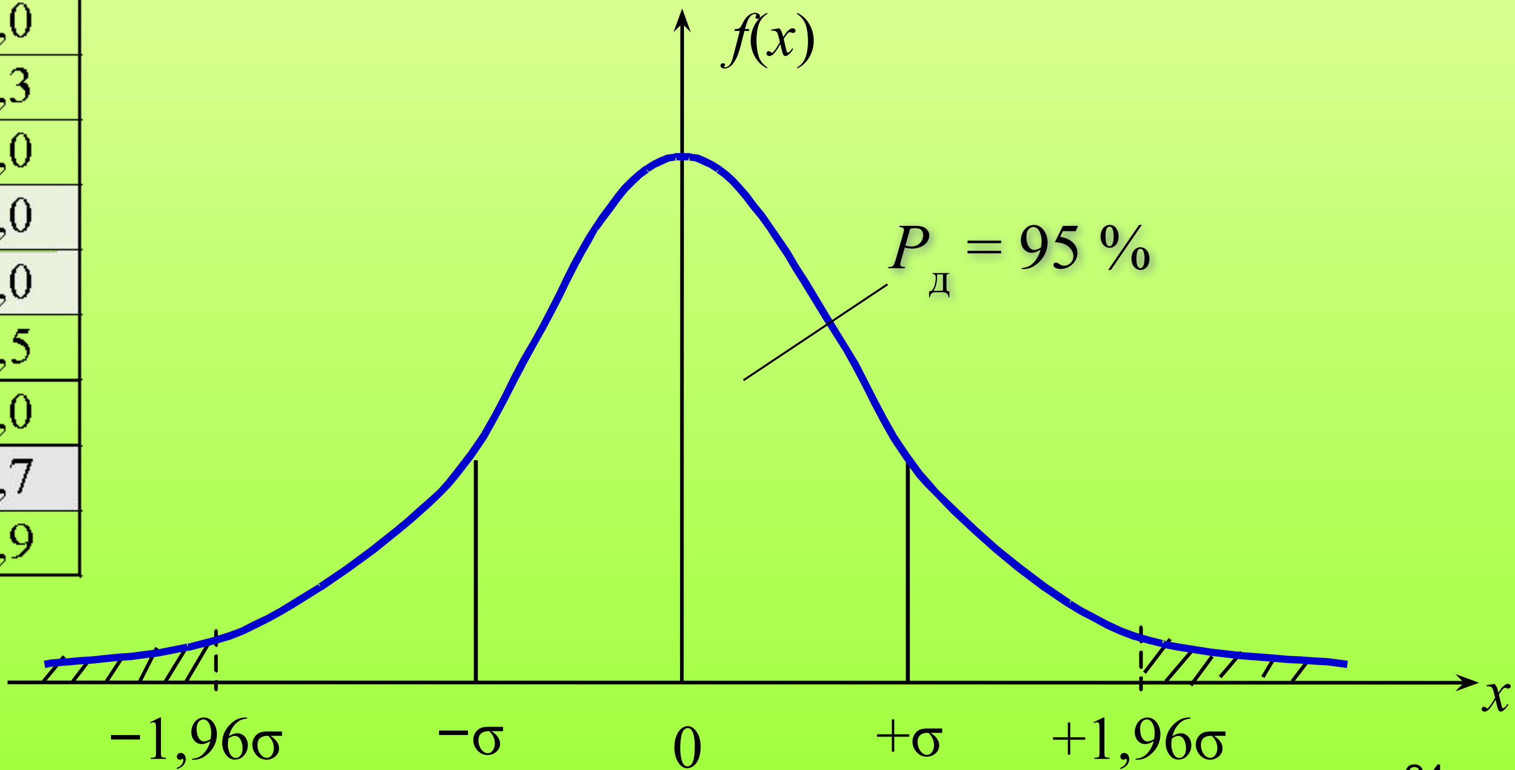
и искомая вероятность $P = 2 \cdot \Phi\left(\frac{120}{50}\right) = 2 \cdot \Phi(2{,}4) = 2 \cdot 0{,}492 = 0{,}984.$

Нетрудно заметить, что для нормального закона распределения погрешностей при одинаковом доверительном интервале доверительная вероятность больше в том случае, когда $\Delta_c \to 0$. Это условие может быть обеспечено, например, внесением соответствующей поправки в результат измерения, если значение Δ_c известно.

Для нормального закона значение квантильного коэффициента z показывает ***число средних квадратических отклонений,*** которое следует отложить вправо и влево от центра рассеивания, чтобы вероятность попадания в полученный интервал была равна $P_д$.

Значения z для различных значений доверительной вероятности $P_д$ можно свести в таблицу:

$\varepsilon = \pm z\sigma$	$P_д$, %
0,675σ	50,0
1σ	68,3
1,282σ	80,0
1,645σ	90,0
1,96σ	95,0
2σ	95,5
2.58σ	99,0
3σ	99,7
3,29σ	99,9



Распределение Стьюдента

Коэффициент Стьюдента

Все, что говорилось выше, ***справедливо для большого числа измерений*** ($n > 30$), когда оценку СКО можно определить достаточно точно (случай известной дисперсии).

Если ***число измерений мало*** ($n \leq 30$), то для оценки границ доверительного интервала используют формулу (случай неизвестной дисперсии):

$$\varepsilon = t\sigma ,$$

где ***t – коэффициент Стьюдента***, зависящий от значения доверительной вероятности и числа наблюдений.

Коэффициент Стьюдента t – это отношение отклонения среднего от истинного значения к СКО среднего:

$$t = \frac{\bar{x} - E(\bar{x})}{S(\bar{x})} = \frac{\bar{x} - x}{S(\bar{x})} = \sqrt{n}\,\frac{\bar{x} - x}{S}, \quad \text{где } S(\bar{x}) = S_{\bar{x}} = \frac{S}{\sqrt{n}}.$$

Вероятность того, что коэффициент Стьюдента примет некоторое значение в интервале $(-t, t)$:

$$P\left(-t < \frac{\bar{x} - x}{S(\bar{x})} < t\right) = P\left(|\bar{x} - x| < \frac{tS}{\sqrt{n}}\right) = 2\int_0^t S(t,v)dt$$

где $S(t,v)$ – ***распределение Стьюдента*** (У. Госсет, 1908),

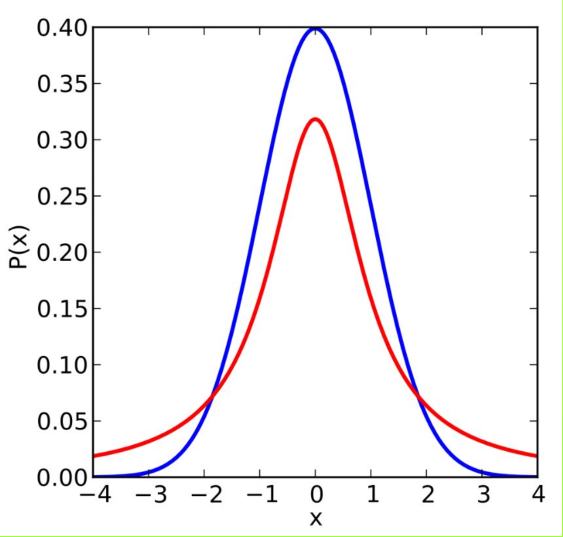
$v = n - 1$ – ***число степеней свободы***.

Значения коэффициентов Стьюдента табулированы.

При увеличении числа наблюдений, распределение Стьюдента $S(t,v)$ приближается к нормальному распределению и при $n > 30$ практически с ним совпадает.

Т.о., доверительный интервал для $n < 30$ определяется по формуле:

$$P\left(\bar{x} - \frac{tS}{\sqrt{n}} < x < \bar{x} + \frac{tS}{\sqrt{n}}\right) = P_{д}$$



Нормальное распределение (синяя) и распределение Стьюдента (красная) при $v = 1$

Коэффициенты Стьюдента

$v = n - 1$	α, %				$v = n - 1$	α, %			
	10	**5**	**1**	**0,1**		**10**	**5**	**1**	**0,1**
1	6,31	12,71	63,66	636,6	18	1,73	2,10	2,88	3,92
2	2,92	4,30	9,92	31,60	19	1,73	2,09	2,86	3,88
3	2.35	3,18	5,84	12,92	20	1,73	2,09	2,85	3,85
4	2,13	2,78	4,60	8,61	21	1,72	2,08	2,83	3,82
5	2,02	2,57	4,03	6,87	22	1,72	2,07	2,82	3,79
6	1,94	2,45	3,71	5,96	23	1,71	2,07	2,81	3,77
7	1,90	2,37	3,50	5,41	24	1,71	2,06	2,80	3,75
8	1,86	2,31	3,36	5,04	25	1,71	2,06	2,79	3,73
9	1,83	2,26	3,25	4,78	26	1,71	2,06	2,78	3,71
10	1,81	2,23	3,17	4,59	27	1,70	2,05	2,77	3,69
11	1,80	2,20	3,11	4,44	28	1,70	2,05	2,76	3,67
12	1,78	2,18	3,05	4,32	29	1,70	2,05	2,76	3,66
13	1,77	2,16	3,01	4,22	30	1,70	2,04	2,75	3,65
14	1,76	2,14	2,98	4,14	40	1,68	2,02	2,70	3,55
15	1,75	2,13	2,95	4,07	60	1,67	2,00	2,66	3,46
16	1,75	2,12	2,92	4,02	120	1,66	1,98	2,62	3,37
17	1,74	2,11	2,90	3,97	∞	1,65	1,96	2,58	3,29
$P_д$	**0,90**	**0,95**	**0,99**	**0,999**	$P_д$	**0,90**	**0,95**	**0,99**	**0,999**

Пример.

Произведено 10 измерений сопротивления резистора с результатами:

Номер измерения	1	2	3	4	5	6	7	8	9	10
Сопротивление, кОм	0,4	0,5	0,3	0,6	0,7	0,5	0,8	0,4	0,6	0,5

Полагая закон распределения полученных результатов нормальным, вычислить доверительный интервал для сопротивления резистора при $P_д$ = 0,9.

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 0{,}53 \text{ кОм}; \quad S = \sqrt{\frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})^2} = 0{,}148 \text{ кОм}.$$

По таблице значений коэффициентов Студента определяем для $P_д$ = 0,9 и v = 9 значение t = 1,83 и вычисляем доверительные границы:

$$\varepsilon = \pm 1{,}83\,\frac{0{,}148}{\sqrt{10}} \approx \pm 0{,}09 \text{ кОм}.$$

Доверительный интервал: (0,53 ± 0,09) кОм.

Пример.

Определение удельных магнитных потерь для различных образцов одной партии электротехнической стали дало следующие результаты:
1,21; 1,17; 1,18; 1,13; 1,19; 1,14; 1,20 и 1,18 Вт/кг.

Считая, что систематическая погрешность отсутствует, а случайная распределена по нормальному закону, определить доверительный интервал при значениях доверительной вероятности 0,9 и 0,95.

Оценки среднего арифметического значения и СКО результатов измерений равны соответственно 1,18 и 0,0278 Вт/кг.

Доверительный интервал для случая известной дисперсии:

$$P(1,18 - z\frac{0,0278}{\sqrt{8}} < x < 1,18 + z\frac{0,0278}{\sqrt{8}}) = 2\Phi(z) = 0,9$$

Из таблицы Лапласа $z_{0,9} = 1,65$ и $z_{0,95} = 1,96$. Соответствующие доверительные интервалы: $(1,180 \pm 0,016)$ и $(1,180 \pm 0,019)$ Вт/кг.

Определим теперь доверительный интервал на основе распределения Стьюдента:

$$P(|\bar{x}-x|<\frac{t\cdot 0,0278}{\sqrt{8}})=0,9$$

По таблице коэффициентов Стьюдента находим $t_{0,9}$ = 1,9 и $t_{0,95}$ = 2,37.

Отсюда доверительные интервалы соответственно равны

(1,180 ± 0,019) и (1,180 ± 0,023) Вт/кг.

Видим, что коэффициенты Студента дают более осторожную оценку доверительных границ по сравнению с квантильными коэффициентами z.

Спасибо за внимание!