

Доверительный интервал для среднего

План:

- Точечные и интервальные оценки
- ДИ для среднего при известной дисперсии
- ДИ для среднего при неизвестной дисперсии

Точечная оценка (point estimate)

Точечной оценкой называется число, которое используется в качестве оценки параметра генеральной совокупности.

Например, среднее значение выборки является точечной оценкой среднего значения генеральной совокупности. Доля признака, рассчитанная по выборке, может рассматриваться как оценка доли признака в генеральной совокупности.



μ

Оценка Параметр

Ошибка оценки (estimation error)

- разность между оцениваемым параметром генеральной совокупности и оценкой, рассчитанной на основе выборки. Ошибка оценки обычно неизвестна, поскольку неизвестен параметр.

$$\text{Ошибка оценки} = \text{Параметр} - \text{Оценка}$$

Критерии точечных оценок

Несмещенность оценки означает, что ее математическое ожидание равно значению оцениваемого параметра генеральной совокупности.

Эффективность оценки означает, что статистика, используемая в качестве точечной оценки параметра генеральной совокупности имеет минимальную стандартную ошибку.

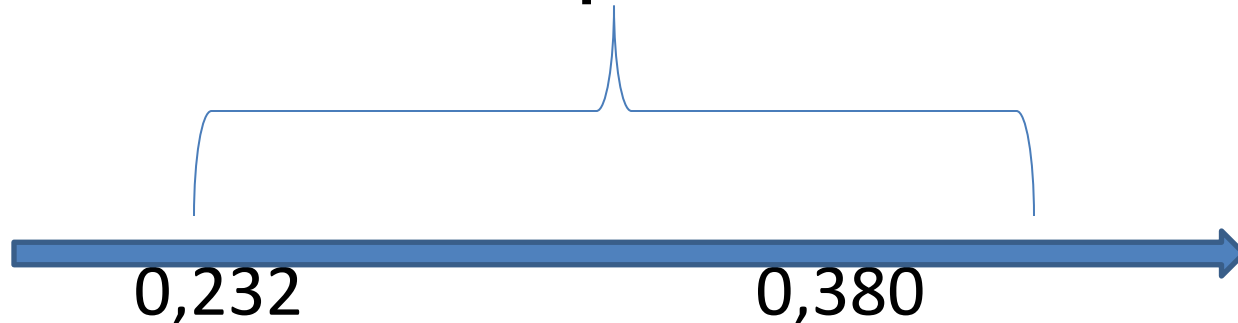
Состоятельность оценки означает, что по мере увеличения объема выборки ее значение приближается к значению оцениваемого параметра генеральной совокупности.

Доверительный интервал (confidence interval)

Доверительный интервал – вычисленный на основе выборки интервал значений признака, который с известной вероятностью содержит оцениваемый параметр генеральной совокупности.

«Мы на 95% уверены, что доля людей которым известна наша торговая марка находится где-то между 23,2% и 38,0%».

**«Параметр находится где-то здесь
с 95% вероятностью»**



Доверительная вероятность

Доверительная вероятность (или уровень доверия, confidence level) – это вероятность того, что доверительный интервал содержит значение оцениваемого параметра.

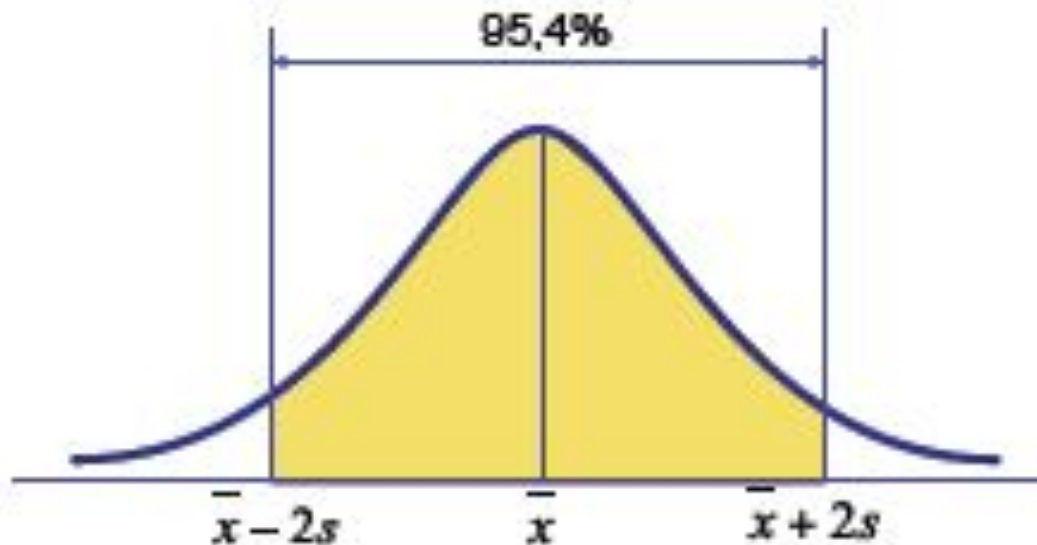
Доверительную вероятность принято устанавливать на уровнях 90%, 95% и 99%. Чем выше доверительная вероятность, тем более широкий и менее полезный интервал мы получим.

90% **95%** 99%

Используется наиболее часто

Для нормального распределения...

Значение нормально распределенного признака находится в пределах двух стандартных отклонений от среднего значения в 95,4% случаев.



Форма записи доверительного интервала

Вариант 1. «Мы на 95% уверены, что среднее значение роста студентов находится где-то между 165 и 175 см».

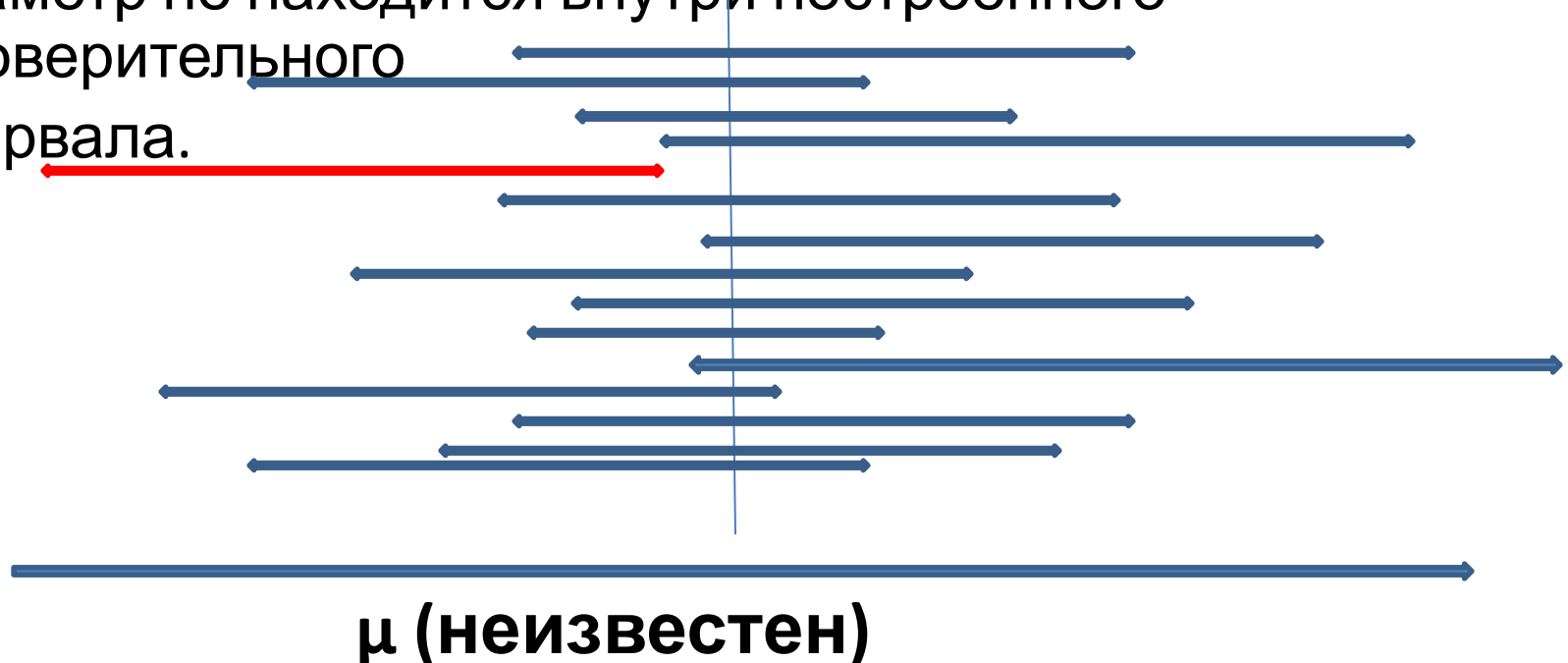
Вариант 2. Среднее значение μ генеральной совокупности находится в интервале от 165 до 175 с доверительной вероятностью 0,95.

Вариант 3. При помощи формулы:

$$P(165 < \mu < 175) = 0,95$$

Зависимость от выборки

Доверительные интервалы, построенные для 15 различных выборок, различны. Только для пятой выборки оцениваемый параметр не находится внутри построенного доверительного интервала.



Описание проблемы

случай: σ известна или $n \geq 30$

Цель. Оценить среднее для генеральной совокупности, имеющей нормальный закон распределения с параметрами μ , σ .

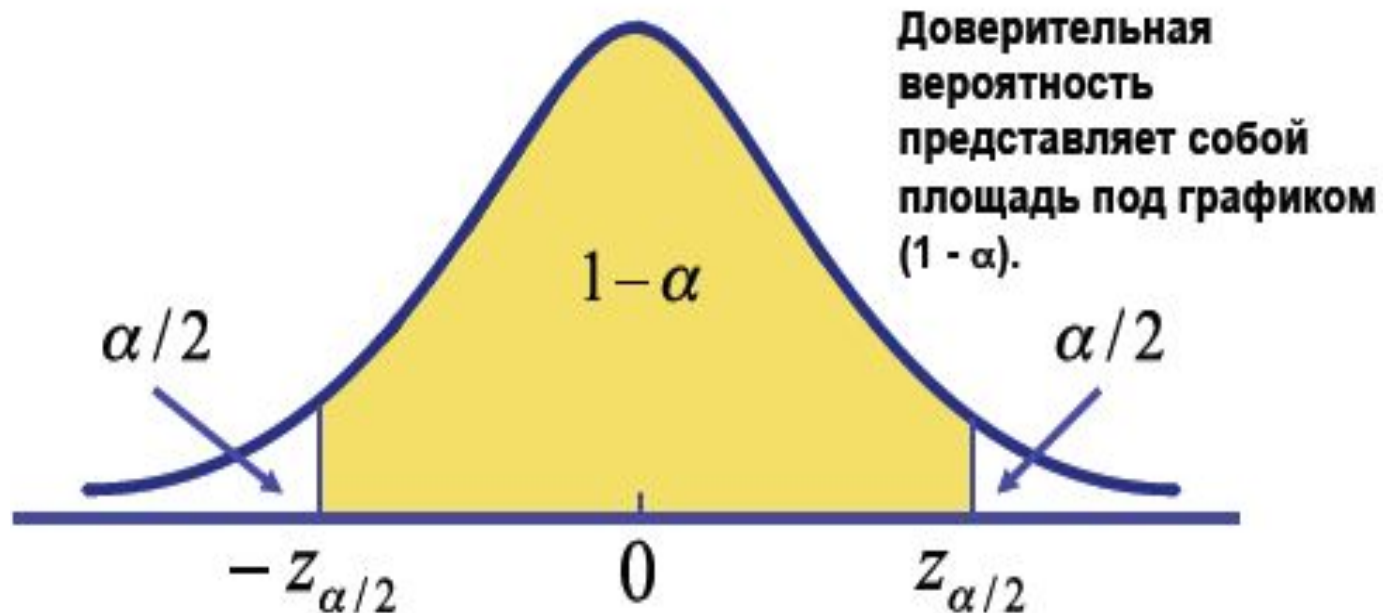
Что мы имеем. Имеем случайную выборку объема n из генеральной совокупности. Стандартное отклонение σ предполагается **известным** или объем выборки $n \geq 30$.

Требуется. Построить доверительный интервал для среднего:

$$x - E < \mu < x + E$$

Метод

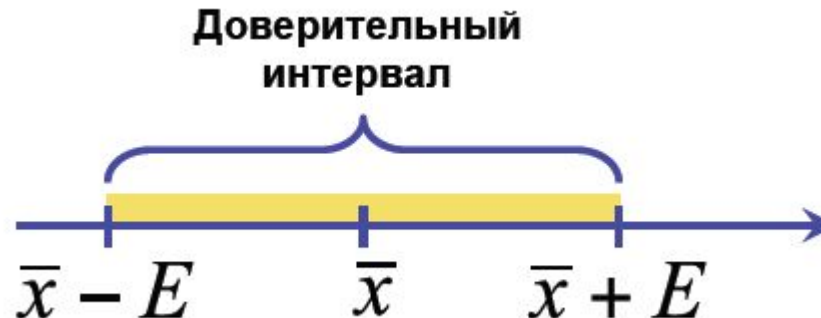
1. В качестве точечной оценки среднего генеральной совокупности рассматриваем выборочное среднее.
2. При построении доверительного интервала основываемся на свойствах нормального закона. Для нахождения z-значений используем таблицы



Доверительный интервал

Среднее генеральной совокупности, имеющей нормальный закон распределения, с доверительной вероятностью $1-\alpha$ находится в доверительном интервале:

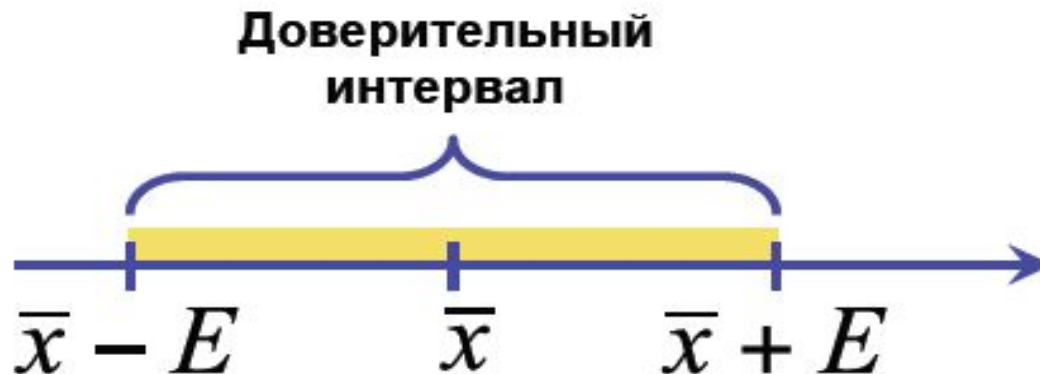
$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



Точность интервальной оценки

Точность интервальной оценки находится по формуле:

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



Последовательность действий

Шаг 1. По выборке вычислить выборочное среднее.

Шаг 2. По таблице нормального закона найти z-значение для

доверительной вероятности $1 - \alpha$.

Шаг 3. Вычислить формуле: $E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ абсолютной оценки по

Шаг 4. Подставить полученные значения в формулу для

доверительного интервала: $\bar{x} - E < \mu < \bar{x} + E$

Шаг 5. Написать ответ.

Важное замечание

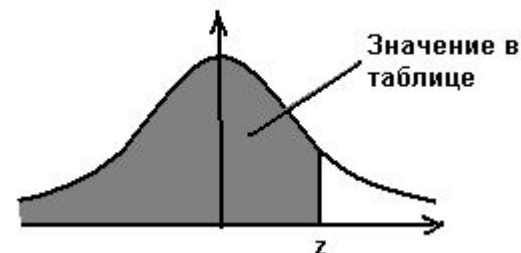
Если значение σ неизвестно и при этом объем выборки $n \geq 30$,
тогда вместо σ используем выборочное стандартное отклонение s :

$$E = z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Использование таблицы

Таблица А-2. Стандартное нормальное распределение

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890



Z-значение

Площадь

1,645 0,9500

1,96 0,9750

2,575 0,9950

Самые используемые z-значения

Z-значение	Площадь	Доверительная вероятность
1,645	0,9500	0,90 или 90%
1,96	0,9750	0,95 или 95%
2,575	0,9950	0,99 или 99%


$$P\left(\bar{x} - 1,65 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1,65 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0,9$$

$$P\left(\bar{x} - 1,96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1,96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

$$P\left(\bar{x} - 2,58 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2,58 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0,99$$

Пример

Ректор университета хочет узнать, каков средний возраст студентов, обучающихся в настоящее время.

Из предыдущих исследований известно, что стандартное отклонение равно 2 годам. Сделана выборка из 50 студентов и вычислено среднее. Оно оказалось равно 20,3 года.

Найти 95%-ый доверительный интервал для генерального среднего.

Решение

Шаг 1. По выборке вычислено выборочное среднее 20,3.

Шаг 2. Доверительная вероятность 95% соответствует z-значению 1,96.

Шаг 3. Вычислим точность интервальной оценки по формуле:

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1,96 \frac{2}{\sqrt{50}} = 0,55$$

Шаг 4. Подставим полученные значения в формулу для доверительного интервала: $20,3 - 0,55 < \mu < 20,3 + 0,55$

Шаг 5. Напишем ответ: $19,75 < \mu < 20,85$

Объем выборки для оценки среднего

Формула для нахождения точности оценки:

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Выражаем объем выборки:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

Если известны E , σ и доверительная вероятность, то по этой формуле подсчитывается **минимальный объем выборки**, который необходим для построения интервальной оценки.

Пример

Декан просит преподавателя по статистике оценить средний возраст студентов факультета.

Какого размера выборка необходима?

Преподаватель статистики считает, что оценка должна быть сделана с точностью до 1 года и с вероятностью 99%.

Из ранее проведенного исследования известно, что стандартное отклонение возраста – 2 года.

Решение

Для $\alpha = 1 - 0,99 = 0,01$ z-значение равно 2,58. По условию $E = 1$, $\sigma = 2$.

Подставим в формулу:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2 = \left[\frac{2,58 \cdot 2}{1} \right]^2 = 26,63 \approx 27$$

Ответ. Чтобы быть на 99% уверенным, что полученная оценка отличается от точного значения среднего возраста не больше чем на 1 год, преподавателю нужна выборка как минимум в 27 человек.

Описание проблемы

σ неизвестно и $n \leq 30$

Цель. Оценить среднее для генеральной совокупности, имеющей нормальный закон распределения с параметрами μ , σ .

Что мы имеем. Имеем случайную выборку объема n из генеральной совокупности. Стандартное отклонение σ неизвестно и объем выборки $n \leq 30$.

Требуется. Построить доверительный интервал для среднего:

$$\bar{x} - E < \mu < \bar{x} + E$$

Отличие метода

При построении доверительного интервала вместо нормального распределения используем распределение Стьюдента.

Для нахождения t -значений будем использовать таблицы распределения Стьюдента.

Число степеней свободы

Число степеней свободы – это количество значений, которые могут свободно изменяться после того, как по выборке было вычислено значение статистики.

Например, пусть известно, что среднее для выборки из пяти значений оказалось равно 10. Тогда четыре из пяти значений могут изменяться, а пятое всегда определено, поскольку сумма пяти есть 50. Число степеней свободы в этом случае: $5 - 1 = 4$.

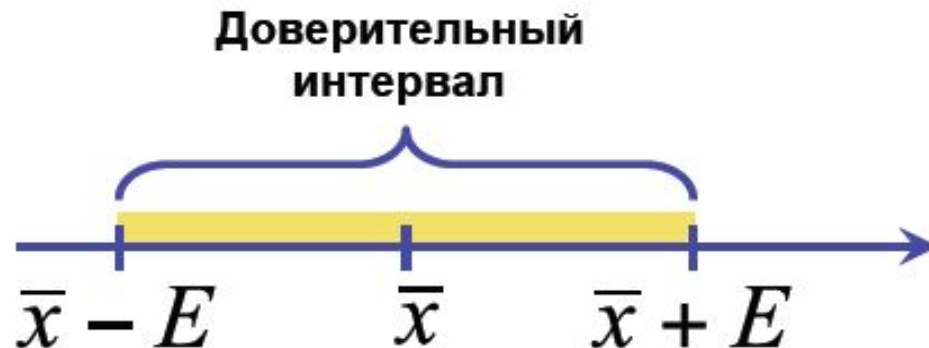
Обозначение: df (degrees of freedom).

Нахождение. Число степеней свободы при построении доверительного интервала для среднего: $df = n - 1$.

Доверительный интервал

Среднее генеральной совокупности, имеющей нормальный закон распределения с доверительной вероятностью $1-\alpha$ находится в доверительном интервале.

$$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$



Последовательность действий

- Шаг 1.** По выборке вычислить выборочное среднее и стандартное отклонение.
- Шаг 2.** По таблице t-распределения найти t-значение для доверительной вероятности $1 - \alpha$ и числа степеней свободы $df = n - 1$.
- Шаг 3.** Вычислить точность интервальной оценки по формуле:

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$$

- Шаг 4.** Подставить полученные значения в формулу для доверительного интервала:

$$\bar{x} - E < \mu < \bar{x} + E$$

- Шаг 5.** Написать ответ.

Использование таблицы

Таблица А-3. Распределение Стьюдента

Df	Односторонняя область				
	0,005	0,01	0,025	0,05	0,1
	Двусторонняя область				
	0,01	0,02	0,05	0,1	0,2
1	63,656	31,821	12,706	6,314	3,078
2	9,925	6,965	4,303	2,920	1,886
3	5,841	4,541	3,182	2,353	1,638
4	4,604	3,747	2,776	2,132	1,533
5	4,032	3,365	2,571	2,015	1,476
6	3,707	3,143	2,477	1,943	1,440
7	3,499	2,998	2,365	1,895	1,415
8	3,355	2,896	2,306	1,860	1,397
9	3,250	2,821	2,262	1,833	1,383
10	3,169	2,764	2,228	1,812	1,372
11	3,106	2,718	2,201	1,796	1,363
12	3,055	2,681	2,179	1,782	1,356
13	3,012	2,650	2,160	1,771	1,350
14	2,977	2,624	2,145	1,761	1,345
15	2,947	2,602	2,131	1,753	1,341
16	2,921	2,583	2,119	1,746	1,337
17	2,898	2,567	2,109	1,740	1,333
18	2,878	2,552	2,101	1,734	1,330
19	2,861	2,539	2,093	1,729	1,328
20	2,845	2,528	2,086	1,725	1,325
21	2,831	2,518	2,080	1,721	1,323

В заголовке таблицы пользуемся значениями для двусторонней области.

Например, при объеме выборки 20 число степеней свободы $df = 19$, если доверительная вероятность выбрана на уровне 95%, тогда t -значение равно 2,093.

Задача

У 20 студентов, сдававших выпускной экзамен, сердце билось в среднем со скоростью 96 ударов в минуту.

Стандартное отклонение выборки было равно 5 ударам в минуту.

Найти 95%-ый доверительный интервал для генерального среднего.

Решение

Шаг 1. По выборке вычислено выборочное среднее 96 и стандартное отклонение 5.

Шаг 2. Доверительная вероятность 95% и количество степеней свободы $df = 20 - 1 = 19$ соответствуют t-значению 2,093.

Шаг 3. Вычислим точность интервальной оценки:

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}} = 2,093 \frac{5}{\sqrt{20}} = 2,34$$

Шаг 4. Подставим полученные значения в формулу для доверительного интервала:

$$96 - 2,34 < \mu < 96 + 2,34$$

Шаг 5. Напишем ответ:

$$93,66 < \mu < 98,34$$