



Элементы математической статистики

Лекция №3

Интервальное оценивание

Пусть $\hat{\theta}_1(x_1, \dots, x_n)$ и $\hat{\theta}_2(x_1, \dots, x_n)$ – такие функции от выборочных значений x_1, \dots, x_n , что

$$P\{\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2\} \geq 1 - \alpha, \quad (3.1)$$

где α – небольшое число: $\alpha = 0,01; 0,05; 0,1$. Обычно величину $1 - \alpha$ обозначают через γ .

Интервал $(\hat{\theta}_1, \hat{\theta}_2)$ называется доверительным интервалом для параметра θ с коэффициентом доверия (надежности) $1 - \alpha$.

Рассмотрим несколько случаев построения доверительных интервалов, если выборка x_1, \dots, x_n взята из нормально распределенной генеральной совокупности.

Доверительный интервал для математического ожидания при известной дисперсии

А) При построении доверительного интервала, используется точечная оценка параметра μ – математического ожидания, т.е.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Из свойств математического ожидания и дисперсии следует, что

$$M\bar{x} = \mu; \quad D\bar{x} = \frac{\sigma^2}{n} \quad (3.2)$$


Поэтому, нормированная случайная величина

$$\frac{(\bar{x} - \mu)}{\sigma} \cdot \sqrt{n} \quad (3.3)$$

имеет нормированное нормальное распределение $N(0,1)$. Границы доверительного интервала будем находить из соотношения

$$P\{|\bar{x} - \mu| \leq \delta\} = 1 - \alpha, \quad (3.4)$$

которое, используя предыдущую формулу (3.3), преобразуем к виду


$$P\left\{\frac{|\bar{x} - \mu|}{\sigma} \sqrt{n} \leq \frac{\delta \sqrt{n}}{\sigma}\right\} = 1 - \alpha. \quad (3.5)$$

Отсюда получаем:

$$2\Phi\left(\frac{\delta \sqrt{n}}{\sigma}\right) = 1 - \alpha \quad (3.6)$$

Здесь $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ – функция Лапласа.

Из последнего соотношения (3.6) по таблице значений функции Лапласа (Таблица II Приложения), находим значение z_α :

$$z_\alpha = \frac{\delta \sqrt{n}}{\sigma}, \quad (3.7)$$

из которого определяем величину


$$\delta = z_\alpha \frac{\sigma}{\sqrt{n}}. \quad (3.8)$$

Обычно используются стандартные значения (Таблица 2)

Таблица 2.

α	0,01	0,05	0,1
z_α	2,576	1,96	1,64

Величина δ задает границы доверительного интервала, т.е. определяет точность интервальной оценки. Коэффициент доверия $1 - \alpha = \gamma$ имеет следующий смысл: если мы будем повторять выборку и для каждой из них находить доверительный интервал, то в среднем на 100 выборок доля тех интервалов, которые накроют оцениваемый параметр, составит $(1 - \alpha) \cdot 100\%$.



Итак, мы получили доверительный интервал для параметра μ при известной дисперсии σ^2 , как следует из (3.4):

$$|\bar{x} - \mu| \leq z_\alpha \frac{\sigma}{\sqrt{n}}$$

или в более принятой форме:


$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}} \quad (3.9)$$

Из анализа формулы (3.9) можно сделать следующие выводы:

а) при увеличении объема выборки n точность интервальной оценки увеличивается, так как величина $\delta = z_\alpha \frac{\sigma}{\sqrt{n}}$ уменьшается. При больших n хорошей оценкой для μ становится \bar{x} , т.е. точечная оценка;

б) в силу того, что функция $\Phi(x)$ является неубывающей, при увеличении надежности γ растет величина δ , т.е. уменьшается точность (интервал становится шире);

в) для фиксированных значений надежности $1 - \alpha = \gamma$ и точности δ из формулы (3.8) можно определить необходимый объем выборки, обеспечивающий заданное значение $1 - \alpha$ и δ . Следует запомнить, что при неизменном объеме выборки одновременно увеличивать точность и надежность оценки **нельзя**.



Пример 1. При измерении уровня шума вырубочного прессы ПВГ-18 были получены следующие значения (дБ): 121,7; 117; 132,4; 117,9; 103,5 ($n = 5$). Считая дисперсию известной и равной $\sigma^2 = 26$, найти доверительный интервал для математического ожидания уровня шума с надежностью $\gamma = 0,95$ ($\alpha = 1 - \gamma = 0,05$).

Решение. Вычислим: $\bar{x} = \frac{121,7 + 117 + 132,4 + 117,9 + 103,5}{5} = 118,5.$

Для $\alpha = 0,05$ соответствующее значение $z_\alpha = 1,96$:

$$\frac{\delta\sqrt{n}}{\sigma} = 1,96 \quad \sqrt{n} = 2,24 \quad \sigma = 5,1$$

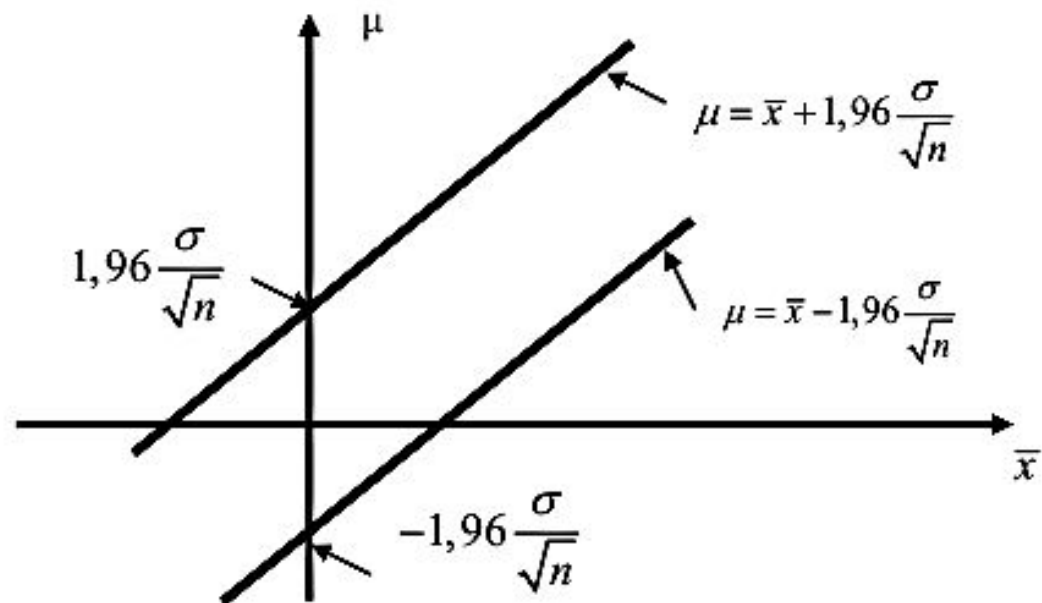
$$\delta = z_\alpha \frac{\sigma}{\sqrt{n}} \quad \delta = 1,96 \frac{5,1}{2,24} = 4,9$$

Окончательно получаем:

$$118,5 - 4,9 \leq \mu \leq 118,5 + 4,9$$

$$113,6 \leq \mu \leq 123,4$$

Б) Доверительный интервал для данных значений σ и n можно представить с помощью доверительной полосы графически



Границы задаются уравнениями:

$$\mu = \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}}$$

$$\mu = \bar{x} - 1,96 \frac{\sigma}{\sqrt{n}}$$

Для любого вычисленного значения \bar{x} по выборке данного объема n , используя рисунок, можно определить границы доверительного интервала для μ . При росте n границы доверительной полосы будут стремиться к линии $\mu = \bar{x}$. Рассмотренный доверительный интервал симметричен относительно \bar{x} . Кроме того, вероятность превзойти левую, либо правую границу интервала одинакова и равна $\frac{\alpha}{2} = 0,025$.

В) Для той же надежности γ можно построить и другой интервал, несимметричный. Вероятность того, что оцениваемый параметр лежит вне интервала, равна $\alpha = 1 - \gamma = 0,05$, но вероятность того, что он лежит левее левой границы равна, например, 0,01, а вероятность того, что он лежит правее правой границы, равна 0,04 ($0,01 + 0,04 = 0,05$).

В этом случае вместо соотношения (3.5) будем использовать

$$P\left\{z_1 \leq \frac{(\bar{x} - \mu)}{\sigma} \sqrt{n} \leq z_2\right\} = 0,95, \quad (3.10)$$

а значения z_1 и z_2 найдем из соотношений:

$$P\left\{\frac{(\bar{x} - \mu)}{\sigma} \sqrt{n} \leq z_1\right\} = F(z_1) = 0,01,$$

$$P\left\{\frac{(\bar{x} - \mu)}{\sigma} \sqrt{n} \leq z_2\right\} = F(z_2) = 0,96,$$

где $F(x)$ – функции распределения $N(0,1)$. Поскольку существует очевидное соотношение: $F(x) = 0,5 + \Phi(x)$, (значения $\Phi(x)$ берутся из *Таблицы II* Приложения), получаем, что

$$z_1 = -2,3 \quad (F(-2,3) = 0,0107),$$

$$z_2 = 1,75 \quad (F(1,75) = 0,9599).$$

Так как $z_1 \neq z_2$, то интервал несимметричен относительно \bar{x} . Следует отметить, что ширина симметричного интервала для $\gamma = 0,95$ равна $3,92 \cdot \frac{\sigma}{\sqrt{n}}$, а для несимметричного интервала $(z_2 - z_1) \cdot \frac{\sigma}{\sqrt{n}} = 4,05 \cdot \frac{\sigma}{\sqrt{n}}$, т.е. при прочих равных условиях симметричный интервал уже, а, значит, оценка точнее.

Доверительный интервал для математического ожидания при неизвестной дисперсии

В этом случае, помимо точечной оценки для μ , будем еще использовать точечную оценку для дисперсии:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$


Для построения доверительного интервала используется величина

$$\frac{(\bar{x} - \mu)}{s} \sqrt{n},$$

имеющая распределение Стьюдента с $(n-1)$ степенями свободы.

Доверительный интервал будем находить из соотношения (3.11):

$$P \left\{ \frac{|\bar{x} - \mu|}{s} \sqrt{n} \leq \frac{\delta \sqrt{n}}{s} \right\} = 1 - \alpha. \quad (3.11)$$



Теперь для определения величины δ воспользуемся таблицами распределения Стьюдента (*Таблица III* Приложения)

откуда

$$\frac{\delta\sqrt{n}}{s} = t_{\alpha, n-1}; \quad \delta = t_{\alpha, n-1} \cdot \frac{s}{\sqrt{n}}$$

и, следовательно, доверительный интервал для μ запишется в следующем виде:

$$\bar{x} - t_{\alpha, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha, n-1} \frac{s}{\sqrt{n}} \quad (3.12)$$

ЗАМЕЧАНИЕ. Распределение Стьюдента при $n > 50$ близко к стандартному нормальному распределению. Поэтому, при $n > 50$ для определения величины δ можно пользоваться не значением $t_{\alpha, n-1}$, а соответствующим значением z_{α} . Например, при $n = 60$:

α	0,1	0,05	0,01
z_{α}	1,64	1,96	2,576
$t_{\alpha, 59}$	1,67	2,00	2,66

Выводы а) – в) остаются в силе и для соотношения (3.12).

Пример 2. При замере освещенности в одной из лабораторий были получены следующие значения в лк. 356,4; 353,3; 354,3; 350,5; 357,2. Найти доверительные границы для математического ожидания уровня освещенности при коэффициенте доверия $\gamma = 0,95$ ($n = 5$).

Решение. $\bar{x} = 354,9$; $n = 5$; $\gamma = 0,95$; $\alpha = 0,05$; $\sqrt{5} = 2,236$;

$$s^2 = \frac{1}{4} \sum_{i=1}^5 (x_i - 354,9)^2 = 6,86; \quad s = 2,62.$$

Для $\gamma = 0,95$ и четырех степеней свободы по таблицам распределения Стьюдента находим $t_{\alpha, n-1} = 2,776$. Следовательно, доверительный интервал, определяемый по формуле (3.12), запишется в виде:

$$354,9 - 2,776 \frac{2,62}{2,236} \leq \mu \leq 354,9 + 2,776 \frac{2,62}{2,236}$$

или

$$352,5 \leq \mu \leq 357,3.$$

Доверительный интервал для дисперсии

При построении доверительного интервала для дисперсии воспользуемся

тем, что величина $\frac{(n-1)s^2}{\sigma^2}$ принадлежит распределению Пирсона (χ^2) с $(n-1)$ степенями свободы.

Доверительный интервал будем находить из следующего соотношения:

$$P\left\{|s^2 - \sigma^2| \leq \delta\right\} = 1 - \alpha = \gamma, \quad (3.13)$$

которое с помощью тождественных преобразований сводится к виду:

$$P\left\{\underline{x}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \bar{x}^2\right\} = \gamma. \quad (3.14)$$

Здесь \underline{x}^2 – нижняя граница доверительного интервала для случайной величины $\frac{(n-1)s^2}{\sigma^2}$; \bar{x}^2 – верхняя граница.

Значение \underline{x}^2 и \bar{x}^2 находим из таблиц распределения χ^2 (Таблица III Приложения) чаще всего из следующих соотношений:

$$P\left\{\chi_{n-1}^2 \geq \bar{x}^2\right\} = \frac{\alpha}{2}; \quad P\left\{\chi_{n-1}^2 \geq \underline{x}^2\right\} = 1 - \frac{\alpha}{2} \quad (3.15)$$

Найденные значения \underline{x}^2 и \bar{x}^2 подставляем в выражение

$$\underline{x}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \bar{x}^2, \quad (3.16)$$

откуда находим искомый доверительный интервал для σ^2 :

$$\frac{(n-1) \cdot s^2}{\bar{x}^2} \leq \sigma^2 \leq \frac{(n-1) \cdot s^2}{\underline{x}^2} \quad (3.17)$$

При $n > 30$ распределение χ_n^2 близко к $N(n, 2n)$. Поэтому в тех случаях, когда объем выборки $n > 30$, можно пользоваться для нахождения границ доверительного интервала стандартными значениями z_α нормального распределения, предварительно сделав нормировку случайной величины

$$\chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma^2}.$$

Итак,

$$\frac{(n-1)s^2}{\sigma^2} \in N(n-1; 2(n-1))$$


Нормируем случайную величину $\frac{(n-1)s^2}{\sigma^2}$:

$$\frac{\frac{(n-1)s^2}{\sigma^2} - (n-1)}{\sqrt{2(n-1)}} \in N(0;1)$$

Для данного значения надежности $1 - \alpha$ находим значение z_α .

Тогда

$$P \left\{ \left| \frac{\frac{(n-1)s^2}{\sigma^2} - (n-1)}{\sqrt{2(n-1)}} \right| \leq z_\alpha \right\} = 1 - \alpha$$



Путем тождественных преобразований выражения, стоящего в фигурных скобках, получаем доверительный интервал для σ^2 :

$$\frac{(n-1) \cdot s^2}{(n-1) + z_{\alpha} \sqrt{2(n-1)}} \leq \sigma^2 \leq \frac{(n-1) \cdot s^2}{(n-1) - z_{\alpha} \sqrt{2(n-1)}}. \quad (3.18)$$

Пример 3. Используя данные предыдущего примера, построить доверительный интервал для σ^2 .

Решение. $n = 5$; $\gamma = 0,95$; $s^2 = 6,86$.

$$P\{\chi_4^2 \geq \bar{x}^2\} = 0,025; \quad P\{\chi_4^2 \geq \underline{x}^2\} = 0,975$$

Воспользовавшись таблицей, находим (*Таблица III* Приложения)

$$\underline{x}^2 = 0,484; \quad \bar{x}^2 = 11,1$$

$$\frac{4 \cdot 6,86}{11,1} \leq \sigma^2 \leq \frac{4 \cdot 6,86}{0,484}$$

$$2,47 \leq \sigma^2 \leq 57,1$$

Доверительный интервал получился очень широким, что, безусловно, объясняется очень маленьким объемом выборки.

