# 9. Интервальные оценки параметров нормального распределения: математического ожидания, генеральной дисперсии, генеральной доли (3 отдельных вопроса).

**Интервальной оценкой** называется оценка, определяемая двумя числами — концами интервала, который с заданной точностью (надежностью) накрывает оцениваемый параметр.

## 1. Интервальная оценка математического ожидания

Для математического ожидания $a$ нормального распределения $N(a, \sigma)$ метод построения доверительного интервала зависит от того, известна ли генеральная дисперсия $\sigma^2$.

### Случай 1: Генеральное среднее квадратическое отклонение $\sigma$ известно

В этом случае используется статистика, имеющая **стандартное нормальное распределение** $N(0, 1)$.
Доверительный интервал строится по формуле:

$$\bar{x} - u_\gamma \frac{\sigma}{\sqrt{n}} < a < \bar{x} + u_\gamma \frac{\sigma}{\sqrt{n}}$$

Где:

- $\bar{x}$ — выборочное среднее.
- $n$ — объем выборки.
- $u_\gamma$ — квантиль, определяемый из уравнения $\Phi^{-1}(\gamma) = u_\gamma$ (где $\Phi^{-1}(\gamma)$ — обратная функция Лапласа, а $\gamma$ — доверительная вероятность).

а) $\sigma^2$ **- известная дисперсия генеральной совокупности.**

Будем строить симметричную оценку параметра $\theta = M(x) = a$. Нам необходимо найти такое значение $\Delta$, чтобы

$$P(\overline{x} - \Delta < a < \overline{x} + \Delta) = \gamma \quad \text{или} \quad P(|\overline{x} - a| < \Delta) = \gamma.$$

Нахождение $\Delta$ основано на том, что статистика

$$u = \frac{\overline{x} - a}{\sigma/\sqrt{n}} \sim N(0,1)$$

распределена по стандартному нормальному закону ($M(u) = 0$ и $D(u) = 1$).

Из условия $P(|u| < u_\gamma) = \gamma$ находим доверительный интервал.

$$\left|\frac{a - \overline{x}}{\sigma/\sqrt{n}}\right| < u_\gamma,$$

$$-u_\gamma < \frac{a - \overline{x}}{\sigma/\sqrt{n}} < u_\gamma,$$

$$-\frac{u_\gamma \cdot \sigma}{\sqrt{n}} < a - \overline{x} < \frac{u_\gamma \cdot \sigma}{\sqrt{n}},$$

$$\overline{x} - \frac{u_\gamma \cdot \sigma}{\sqrt{n}} < a < \overline{x} + \frac{u_\gamma \cdot \sigma}{\sqrt{n}},$$

Значение $u_\gamma$ - квантиль стандартного нормального распределения уровня $\gamma$ определяем по таблицам для функции Лапласа (см. Приложение 1) $\Phi^{-1}(\gamma) = u_\gamma$.

Схема 1:

$$\gamma \to u_\gamma \to \triangle = \frac{u_\gamma \sigma}{\sqrt{n}} \to \begin{cases} \overline{x} - \triangle - \text{нижняя граница} \\ \overline{x} + \triangle - \text{верхняя граница} \end{cases}$$

Получаем интервальную оценку для математического ожидания: $\overline{x} - \triangle < a < \overline{x} + \triangle$.

## Случай 2: Генеральное среднее квадратическое отклонение $\sigma$ неизвестно

Если $\sigma$ неизвестна, вместо нее используется **исправленное выборочное среднее квадратическое отклонение** $s$. Соответствующая статистика имеет **распределение Стьюдента** с числом степеней свободы $k = n - 1$.

$$s^2 = \frac{\Sigma(x_i - \overline{x})^2}{n-1}$$

Доверительный интервал строится по формуле:

$$\bar{x} - t_\gamma \frac{s}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{s}{\sqrt{n}}$$

Где $t_\gamma$ — критическая точка распределения Стьюдента, найденная по таблице для заданного уровня значимости $\alpha = 1 - \gamma$ и $k$ степеней свободы.

Попробуем сконструировать центральную статистику для построения интервальной оценки математического ожидания для этого случая. За основу возьмем точечную оценку математического ожидания — выборочное среднее. Подправим его, чтобы не было зависимости от параметра $\widehat{\theta} = a$.

Во-первых, центрируем: $\overline{x} - a$. Во-вторых, нормируем, поделив на среднее квадратическое отклонение

$$\sigma(\overline{x}) = \sqrt{D(\overline{x})} = \sqrt{s^2/n} = s/\sqrt{n}.$$

Обратим внимание, что используется исправленная дисперсия $s^2$ вместо $\sigma^2$.

Выполним некоторые преобразования, чтобы понять, каков закон распределения найденной статистики. Обозначим ее пока $K$:

$$K = \frac{\overline{x} - a}{s/\sqrt{n}} = \frac{\overline{x} - a}{(\sigma/\sqrt{n}) \cdot \frac{s}{\sigma}} = \frac{\overline{x} - a}{(\sigma/\sqrt{n})} \cdot \frac{1}{\frac{1}{\sigma}\sqrt{\frac{\Sigma(x_i - \overline{x})^2}{n-1}}} =$$

$$= \frac{\overline{x} - a}{(\sigma/\sqrt{n})} \cdot \frac{1}{\sqrt{\frac{1}{n-1}\frac{\Sigma(x_i - \overline{x})^2}{\sigma^2}}} = \frac{\overline{x} - a}{(\sigma/\sqrt{n})} \cdot \frac{1}{\sqrt{\frac{1}{n-1}\Sigma\left(\frac{x_i - \overline{x}}{\sigma^2}\right)^2}}$$

Первый из полученных множителей в точности равен статистике $u = \dfrac{\overline{x} - \theta}{\sigma/\sqrt{n}}$, имеет стандартное нормальное распределение $N(0;1)$.

Во втором множителе в знаменателе под корнем стоит сумма квадратов стандартных нормальных величин $\Sigma\left(\dfrac{x_i - \overline{x}}{\sigma^2}\right)^2$, которая имеет $\chi^2$ распределение (см. п.1.3) с числом степеней свободы $k = n - 1$, т.к. одна степень свободы «теряется» (вычислялась $s^2$).

В результате имеем:

$$u = \frac{\overline{x} - \theta}{\sigma/\sqrt{n}} \sim N(0;1), \qquad \Sigma\left(\frac{x_i - \overline{x}}{\sigma^2}\right)^2 \sim \chi^2(k = n-1).$$

Тога построенная нами центральная стаитистика

$$K = \frac{u}{\sqrt{\chi^2/k}} \sim t(k = n - 1)$$

имеет закон распределения Стьюдента с числом степеней свободы $k = n - 1$ (см. п.1.3).

Она и используется для построения интервальной оценки математического ожидания в случае неизвестной генеральной дисперсии

$$t_B = \frac{\overline{x} - a}{s/\sqrt{n}} \sim t_{\gamma,n-1}.$$

Из условия $P(|t_B| < t_\gamma) = \gamma$ находим доверительный интервал.

$$\left|\frac{a - \overline{x}}{S/\sqrt{n}}\right| < t_\gamma,$$

$$-t_\gamma < \frac{a - \overline{x}}{S/\sqrt{n}} < t_\gamma,$$

$$-\frac{t_\gamma \cdot S}{\sqrt{n}} < a - \overline{x} < \frac{t_\gamma \cdot S}{\sqrt{n}},$$

$$\overline{x} - \frac{t_\gamma \cdot S}{\sqrt{n}} < a < \overline{x} + \frac{t_\gamma \cdot S}{\sqrt{n}},$$

где $t_\gamma$ определяется по таблицам критических точек распределения Стьюдента (см. Приложение 2).

Статистические таблицы для распределения Стьюдента построены в зависимости от значения уровня значимости $\alpha$. Поэтому при построении интервальной оценки будем использовать квантили уровня $\alpha = 1 - \gamma$. Причем, следует обратить внимание, что надо брать квантили для двусторонней критической области.

Схема 2:

$$\left.\begin{array}{l} \alpha = 1 - \gamma \\ k = n - 1 \end{array}\right\} \to t_{\alpha,k} \to \triangle = \frac{t_{\alpha,k} \cdot s}{\sqrt{n}} \to \left\{\begin{array}{l} \overline{x} - \triangle - \text{нижняя граница} \\ \overline{x} + \triangle - \text{верхняя граница} \end{array}\right.$$

Получаем интервальную оценку для математического ожидания: $\overline{x} - \triangle < a < \overline{x} + \triangle$.

# 2. Интервальная оценка генеральной дисперсии

Для построения доверительного интервала генеральной дисперсии $\sigma^2$ используется статистика, имеющая **распределение $\chi^2$ (хи-квадрат) Пирсона**.

## Случай 1: Математическое ожидание $a$ известно

Используется статистика $\chi^2 = \frac{ns_*^2}{\sigma^2}$ с $k = n$ степенями свободы, где $s_*^2 = \frac{\sum(x_i-a)^2}{n}$.
Доверительный интервал:

$$\frac{ns_*^2}{\chi_2^2} < \sigma^2 < \frac{ns_*^2}{\chi_1^2}$$

Где $\chi_1^2$ и $\chi_2^2$ — квантили распределения $\chi^2$.

Схема 3:

$$\gamma, n \rightarrow \left.\begin{array}{l} \alpha_1 = \dfrac{1+\gamma}{2} = 1 - \dfrac{\alpha}{2} \\ \alpha_2 = \dfrac{1-\gamma}{2} = \dfrac{\alpha}{2} \end{array}\right\} \rightarrow \left\{\begin{array}{l} \chi_1^2(\alpha_1, n) \\ \\ \chi_2^2(\alpha_2, n) \end{array}\right. \rightarrow \left\{\begin{array}{l} \sigma_2^2 = \dfrac{ns_*^2}{\chi_1^2} \text{ — правая граница} \\ \sigma_1^2 = \dfrac{ns_*^2}{\chi_2^2} \text{ — левая граница} \end{array}\right.$$

Получаем интервальную оценку для генеральной дисперсии:

$$\sigma_1^2 < \sigma^2 < \sigma_2^2$$

## Случай 2: Математическое ожидание $a$ неизвестно

В этом случае используется статистика $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ с числом степеней свободы $k = n - 1$.
Формула доверительного интервала:

$$\frac{(n-1)s^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_1^2}$$

Где:

- $s^2$ — исправленная выборочная дисперсия.
- $\chi_1^2 = \chi^2(1 - \alpha/2; n - 1)$ и $\chi_2^2 = \chi^2(\alpha/2; n - 1)$ — квантили распределения хи-квадрат, определяемые по таблицам.

$$\widehat{\sigma}^2 = \frac{\Sigma(x_i - \overline{x})^2}{n} \quad \text{или} \quad s^2 = \frac{\Sigma(x_i - a)^2}{n-1}$$

Статистика

$$\chi^2_{\text{В}} = \frac{\overset{(n-1)}{\cancel{n}}\widehat{\sigma}^2}{\sigma^2} \quad \text{или} \quad \chi^2_{\text{В}} = \frac{\overset{(n-1)}{\cancel{n}}s^2}{\sigma^2}$$



распределена по закону Пирсона - $\chi^2$ с $k = n - 1$ степенями свободы (одна степень свободы теряется на определение среднего арифметического).

Из $P(\chi^2_1 < \chi^2_{\text{В}} < \chi^2_2) = \gamma$ так же, как и в предыдущем случае, находим доверительный интервал для генеральной дисперсии.

Алгоритм построения интервальной оценки практически такой же, как и в случае (а) с заменой $n$ на $n - 1$ и $s^2_*$ на выборочную дисперсию $\widehat{\sigma}^2$ или исправленную выборочную дисперсию $s^2$.

Схема 4:

$$\left.\begin{matrix} \gamma \\ k = n-1 \end{matrix}\right\} \rightarrow \left.\begin{matrix} \alpha_1 = 1 - \dfrac{\alpha}{2} \\ \alpha_2 = \dfrac{\alpha}{2} \end{matrix}\right\} \rightarrow \left\{\begin{matrix} \chi^2_1(\alpha_1, n-1) \\ \\ \chi^2_2(\alpha_2, n-1) \end{matrix}\right. \rightarrow \left\{\begin{matrix} \sigma^2_2 = \dfrac{\overset{(n-1)}{\cancel{n}}s^2}{\chi^2_1} - \text{правая гр.} \\ \sigma^2_1 = \dfrac{\overset{(n-1)}{\cancel{n}}s^2}{\chi^2_2} - \text{левая гр.} \end{matrix}\right.$$

Интервальная оценка генеральной дисперсии запишется: $\sigma^2_1 < \sigma^2 < \sigma^2_2$.

# 3. Интервальная оценка генеральной доли

Точечной оценкой генеральной доли $p$ является выборочная доля $\hat{p} = \frac{m}{n}$.

## Случай 1: Большой объем выборки ($np > 10$)

Для построения интервала используется нормальная аппроксимация. При $n \geq 100$ границы интервала определяются приближенными формулами:

$$p_1 = \hat{p} - u_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$p_2 = \hat{p} + u_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

При меньших объемах выборки ($n < 100$) используются более точные формулы, основанные на решении квадратного уравнения относительно $p$:

$$p_{1,2} = \frac{\hat{p} + \frac{u_\gamma^2}{2n} \mp u_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{u_\gamma^2}{4n^2}}}{1 + \frac{u_\gamma^2}{n}}$$

Разрешив неравенство, стоящее в скобках относительно $p$ (для $p$ получается квадратное уравнение), получим

$$p_1 = \frac{\widehat{p} + u_\gamma^2/(2n) - u_\gamma\sqrt{\widehat{p}(1-\widehat{p})/n + u_\gamma^2/(4n^2)}}{1 + u_\gamma^2/n} \tag{4.6}$$

$$p_2 = \frac{\widehat{p} + u_\gamma^2/(2n) + u_\gamma\sqrt{\widehat{p}(1-\widehat{p})/n + u_\gamma^2/(4n^2)}}{1 + u_\gamma^2/n} \tag{4.7}$$

Схема 5:

$$\gamma \to u_\gamma \to \begin{cases} p_1 & \text{вычисляется по формуле (4.6),} \\ p_2 & \text{вычисляется по формуле (4.7).} \end{cases}$$

При $n \gg 100$ используются приближенные формулы:

$$p_1 \approx \widehat{p} - u_\gamma\sqrt{\widehat{p}(1-\widehat{p})/n}, \tag{4.8}$$

$$p_2 \approx \widehat{p} + u_\gamma\sqrt{\widehat{p}(1-\widehat{p})/n}. \tag{4.9}$$



Схема 6:

$$\gamma \to u_\gamma \to \triangle = u_\gamma\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \to \begin{cases} p_1 = \widehat{p} - \triangle & (4.8) \\ p_2 = \widehat{p} + \triangle & (4.9) \end{cases}$$

## Случай 2: Малый объем выборки

## Когда число испытаний невелико, границы $p_1$ и $p_2$ определяются как решения нелинейных уравнений

## на основе биномиального распределения или находятся по специальным таблицам.

б) Малое число испытаний Бернулли.

Границы интервала $p_1$ и $p_2$ являются решениями нелинейных уравнений:

$$\sum_{x=0}^{m-1} \cancel{=} C_n^x p_1^x (1-p_1)^{n-x} = \frac{1+\gamma}{2}, \quad \sum_{x=0}^{m} \cancel{=} C_n^x p_2^x (1-p_2)^{n-x} = \frac{1-\gamma}{2}.$$

Существуют специальные таблицы для нахождения чисел $p_1$ и $p_2$ (Приложение 5).

Схема 7:

$$\left.\begin{array}{l}\gamma \\ n \\ n-m\end{array}\right\} \to p_1 \quad \text{и} \quad p_2 \quad \text{определяем по таблице}$$

**Аналогия:**

Интервальная оценка — это своего рода «сеть». Если точечная оценка — это попытка угадать точный вес рыбы (одно число), то интервальная оценка — это заявление, что вес рыбы находится в диапазоне от 500 до 600 граммов.

• Доверительная вероятность ($\gamma$) — это прочность сети (насколько мы уверены, что рыба внутри).

• Дисперсия и объем выборки определяют размер сети: чем больше неопределенность (дисперсия) и меньше попыток (выборка), тем шире должна быть «сеть» (интервал), чтобы гарантированно поймать результат.