# Распределения, используемые в математической статистике

## Распределения, используемые в математической статистике

Введем некоторые распределения случайных величин, которые основаны на нормальном распределении и понадобятся в дальнейшем.

## Распределения, используемые в математической статистике

Введем некоторые распределения случайных величин, которые основаны на нормальном распределении и понадобятся в дальнейшем.

### 1. **Распределение хи-квадрат (Пирсона)**

Пусть $X_1, X_2, ... X_n$ - $n$ независимых случайных величин, $X_i \sim N(0,1)$.

## Распределения, используемые в математической статистике

Введем некоторые распределения случайных величин, которые основаны на нормальном распределении и понадобятся в дальнейшем.

### 1. **Распределение хи-квадрат (Пирсона)**

Пусть $X_1, X_2, \ldots X_n$ - $n$ независимых случайных величин, $X_i \sim N(0,1)$. Тогда

$$K = X_1^2 + X_2^2 + \cdots + X_n^2$$

- хи-квадрат случайная величина с $n$ степенями свободы:

$$\boxed{K \sim \chi_n^2}$$

## Распределения, используемые в математической статистике

Введем некоторые распределения случайных величин, которые основаны на нормальном распределении и понадобятся в дальнейшем.

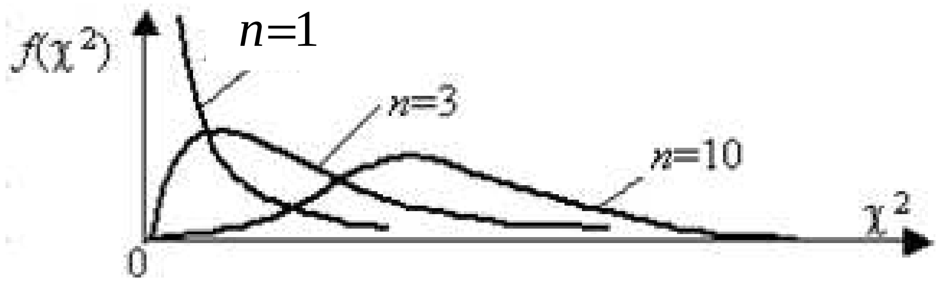
### 1. **Распределение хи-квадрат (Пирсона)**

Пусть $X_1, X_2, \ldots X_n$ - $n$ независимых случайных величин, $X_i \sim N(0,1)$. Тогда

$$K = X_1^2 + X_2^2 + \cdots + X_n^2$$

- хи-квадрат случайная величина с $n$ степенями свободы:

$$\boxed{K \sim \chi_n^2}$$




Плотность распределения хи-квадрат

# Основные свойства распределения хи-квадрат

## Основные свойства распределения хи-квадрат

1. Хи-квадрат распределение является частным случаем гамма-распределения: $\chi_n^2 = \Gamma(\frac{1}{2}, \frac{n}{2})$;

## Основные свойства распределения хи-квадрат

1. Хи-квадрат распределение является частным случаем гамма-распределения: $\chi_n^2 = \Gamma(\frac{1}{2}, \frac{n}{2})$;

2. При больших $n$ хи-квадрат распределение аппроксимируется нормальным: $\chi_n^2 \approx N(n, \sqrt{2n})$.

## Основные свойства распределения хи-квадрат

1. Хи-квадрат распределение является частным случаем гамма-распределения: $\chi_n^2 = \Gamma(\frac{1}{2}, \frac{n}{2})$;

2. При больших $n$ хи-квадрат распределение аппроксимируется нормальным: $\chi_n^2 \approx N(n, \sqrt{2n})$.

3. Пусть все $X_i \sim N(a, \sigma)$, $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ - выборочная дисперсия.

## Основные свойства распределения хи-квадрат

1. Хи-квадрат распределение является частным случаем гамма-распределения: $\chi_n^2 = \Gamma(\frac{1}{2}, \frac{n}{2})$;

2. При больших $n$ хи-квадрат распределение аппроксимируется нормальным: $\chi_n^2 \approx N(n, \sqrt{2n})$.

3. Пусть все $X_i \sim N(a, \sigma)$, $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ - выборочная дисперсия. Тогда величина

$$\boxed{\frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2}$$

## Основные свойства распределения хи-квадрат

1. Хи-квадрат распределение является частным случаем гамма-распределения: $\chi_n^2 = \Gamma(\frac{1}{2}, \frac{n}{2})$;

2. При больших $n$ хи-квадрат распределение аппроксимируется нормальным: $\chi_n^2 \approx N(n, \sqrt{2n})$.

3. Пусть все $X_i \sim N(a, \sigma)$, $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ - выборочная дисперсия. Тогда величина

$$\boxed{\frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2}$$

не зависит от $a, \sigma$

# Основные свойства распределения хи-квадрат

1. Хи-квадрат распределение является частным случаем гамма-распределения: $\chi_n^2 = \Gamma(\frac{1}{2}, \frac{n}{2})$;

2. При больших $n$ хи-квадрат распределение аппроксимируется нормальным: $\chi_n^2 \approx N(n, \sqrt{2n})$.

3. Пусть все $X_i \sim N(a, \sigma)$, $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ - выборочная дисперсия. Тогда величина

$$\boxed{\frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2}$$

не зависит от $a, \sigma$

Замечание. Из свойства 3 следует, что $\boxed{(n-1)\frac{\tilde{S}^2}{\sigma^2} \sim \chi_{n-1}^2}$,

где $\tilde{S}^2 = \frac{1}{n-1}\sum_{i}(X_i - \bar{X})^2$ - несмещенная оценка дисперсии.

## 2. Распределение Стьюдента

Пусть $Y, K$ - независимые случайные величины; $Y \sim N(0,1)$, $K \sim \chi_n^2$.

## 2. Распределение Стьюдента

Пусть $Y, K$ - независимые случайные величины; $Y \sim N(0,1)$, $K \sim \chi_n^2$. Обозначим

$$T = \frac{Y}{\sqrt{K/n}}.$$

## 2. Распределение Стьюдента

Пусть $Y, K$ - независимые случайные величины; $Y \sim N(0,1)$, $K \sim \chi_n^2$. Обозначим

$$T = \frac{Y}{\sqrt{K/n}}.$$

Распределение величины $T$ называется распределением Стьюдента с $n$ степенями свободы ($t$-распределением):

$$\boxed{T \sim t_n}.$$

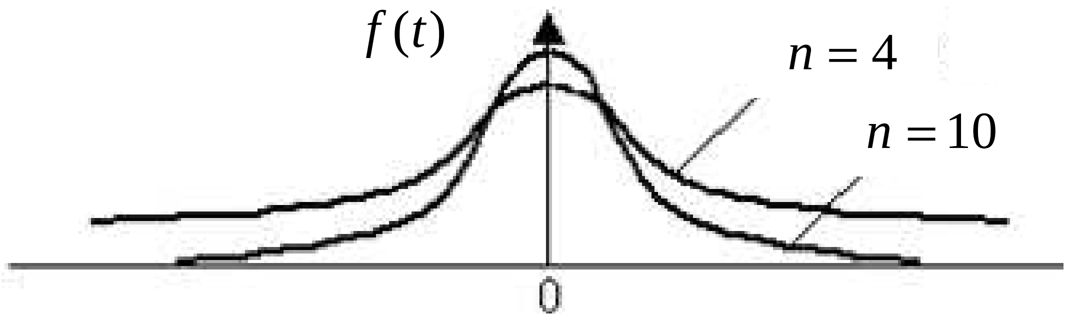
## 2. Распределение Стьюдента

Пусть $Y, K$ - независимые случайные величины; $Y \sim N(0,1)$, $K \sim \chi_n^2$. Обозначим

$$T = \frac{Y}{\sqrt{K/n}}.$$

Распределение величины $T$ называется распределением Стьюдента с $n$ степенями свободы ($t$-распределением):

$$\boxed{T \sim t_n}.$$




Плотность $t$-распределения

# Основные свойства распределения Стьюдента

## Основные свойства распределения Стьюдента

1. При увеличении $n$ распределение Стьюдента приближается к нормальному;

## Основные свойства распределения Стьюдента

1. При увеличении $n$ распределение Стьюдента приближается к нормальному;

2. График плотности $t$-распределения симметричен относительно вертикальной оси, причем $ET = 0$.

## Основные свойства распределения Стьюдента

1. При увеличении $n$ распределение Стьюдента приближается к нормальному;

2. График плотности $t$-распределения симметричен относительно вертикальной оси, причем $ET = 0$.

3. Пусть все $X_i \sim N(a, \sigma)$, $i = 1,..,n$, и независимы. Тогда случайная величина

$$\frac{\bar{X} - a}{S}\sqrt{n-1}$$

подчиняется распределению Стьюдента $t_{n-1}$

## Основные свойства распределения Стьюдента

1. При увеличении $n$ распределение Стьюдента приближается к нормальному;

2. График плотности $t$-распределения симметричен относительно вертикальной оси, причем $ET = 0$.

3. Пусть все $X_i \sim N(a, \sigma)$, $i = 1,..,n$, и независимы. Тогда случайная величина

$$\frac{\bar{X} - a}{S}\sqrt{n-1}$$

подчиняется распределению Стьюдента $t_{n-1}$ (распределение не зависит от $a, \sigma$).

## Основные свойства распределения Стьюдента

1. При увеличении $n$ распределение Стьюдента приближается к нормальному;

2. График плотности $t$-распределения симметричен относительно вертикальной оси, причем $ET = 0$.

3. Пусть все $X_i \sim N(a,\sigma)$, $i = 1,..,n$, и независимы. Тогда случайная величина

$$\frac{\bar{X} - a}{S}\sqrt{n-1}$$

подчиняется распределению Стьюдента $t_{n-1}$ (распределение не зависит от $a,\sigma$).

**Докажем 3.** Обозначим $Y = \dfrac{\bar{X} - a}{\sigma}\sqrt{n}$, $K = \dfrac{nS^2}{\sigma^2}$. Тогда $Y \sim N(0,1)$, $K \sim \chi^2_{n-1}$ по свойству хи-квадрат распределения.

$$\frac{\bar{X}-a}{S}\sqrt{n-1} = \underbrace{\frac{\bar{X}-a}{\sigma}\sqrt{n}}_{Y}\underbrace{\boxed{\frac{\sigma}{\sqrt{n}S}}}_{1/\sqrt{K}}\sqrt{n-1} = \frac{Y}{\sqrt{K/(n-1)}} \sim t_{n-1}$$

$$\frac{\bar{X}-a}{S}\sqrt{n-1}=\underbrace{\frac{\bar{X}-a}{\sigma}\sqrt{n}}_{Y}\,\underbrace{\boxed{\frac{\sigma}{\sqrt{n}S}}}_{1/\sqrt{K}}\sqrt{n-1}=\frac{Y}{\sqrt{K/(n-1)}}\sim t_{n-1}$$

- получили распределение Стьюдента. $\blacksquare$

$$\frac{\bar{X}-a}{S}\sqrt{n-1}=\underbrace{\frac{\bar{X}-a}{\sigma}\sqrt{n}}_{Y}\underbrace{\boxed{\frac{\sigma}{\sqrt{n}S}}}_{1/\sqrt{K}}\sqrt{n-1}=\frac{Y}{\sqrt{K/(n-1)}}\sim t_{n-1}$$

- получили распределение Стьюдента. $\blacksquare$

Замечание (3’). Аналогично можно показать, что

$$\frac{\bar{X}-a}{\tilde{S}}\sqrt{n}\sim t_{n-1},$$

где $\tilde{S}=\sqrt{\tilde{S}^2}$, $\tilde{S}^2$ - несмещенная оценка дисперсии.

## 3. $F$ – распределение (распределение Фишера)

Пусть $K_m, K_n$ - независимые случайные величины:

$K_m \sim \chi^2(m),\ \ K_n \sim \chi^2(n).$

## 3. $F$ – распределение (распределение Фишера)

Пусть $K_m, K_n$ - независимые случайные величины:
$K_m \sim \chi^2(m), \;\; K_n \sim \chi^2(n)$.

Тогда распределение случайной величины

$$F = \frac{K_m / m}{K_n / n}$$

называется $F$-распределением с $m$ и $n$ степенями свободы:

$$\boxed{F \sim F_{m,n}}.$$

## 3. $F$ – распределение (распределение Фишера)

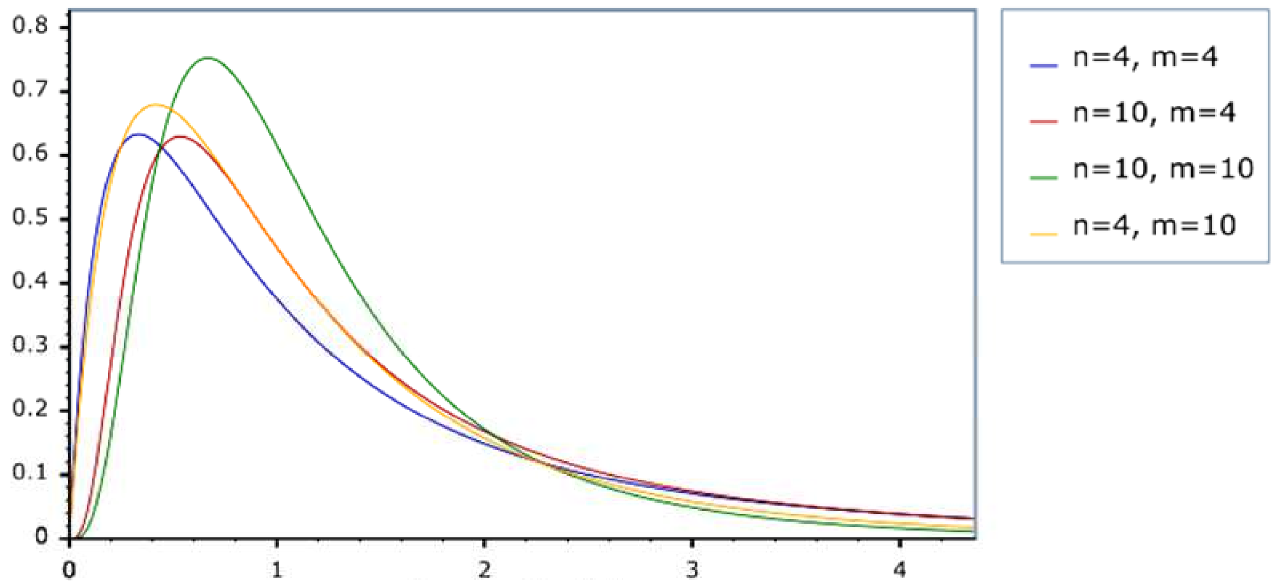
Пусть $K_m, K_n$ - независимые случайные величины:
$K_m \sim \chi^2(m), \ K_n \sim \chi^2(n)$.

Тогда распределение случайной величины

$$F = \frac{K_m / m}{K_n / n}$$

называется $F$-распределением с $m$ и $n$ степенями свободы:

$$\boxed{F \sim F_{m,n}}.$$




Плотность $F$-распределения

## Интервальное оценивание

Пусть $X \sim F(x, \theta)$, причём вид функции распределения $F(x, \theta)$ известен, а одномерный параметр $\theta$ - нет.

## Интервальное оценивание

Пусть $X \sim F(x, \theta)$, причём вид функции распределения $F(x,\theta)$ известен, а одномерный параметр $\theta$ - нет. Требуется по выборке указать такой интервал $[\hat{\theta}_1, \hat{\theta}_2]$, который с заданной вероятностью $\gamma$ «накрывает» неизвестный параметр $\theta$:

$$\boxed{P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \gamma}$$

## Интервальное оценивание

Пусть $X \sim F(x,\ \theta)$, причём вид функции распределения $F(x,\theta)$ известен, а одномерный параметр $\theta$ - нет. Требуется по выборке указать такой интервал $[\hat{\theta}_1,\ \hat{\theta}_2]$, который с заданной вероятностью $\gamma$ «накрывает» неизвестный параметр $\theta$:

$$\boxed{P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \gamma}$$

Интервал $[\hat{\theta}_1,\ \hat{\theta}_2]$, называется *доверительным*, а $\gamma$ – *доверительной вероятностью*.

## Интервальное оценивание

Пусть $X \sim F(x,\ \theta)$, причём вид функции распределения $F(x,\theta)$ известен, а одномерный параметр $\theta$ - нет. Требуется по выборке указать такой интервал $[\hat{\theta}_1,\ \hat{\theta}_2]$, который с заданной вероятностью $\gamma$ «накрывает» неизвестный параметр $\theta$:

$$\boxed{P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \gamma}$$

Интервал $[\hat{\theta}_1,\ \hat{\theta}_2]$, называется *доверительным*, а $\gamma$ – *доверительной вероятностью*.

Величина $\alpha = 1 - \gamma$ называется уровнем значимости.

## Интервальное оценивание

Пусть $X \sim F(x, \theta)$, причём вид функции распределения $F(x,\theta)$ известен, а одномерный параметр $\theta$ - нет. Требуется по выборке указать такой интервал $[\hat{\theta}_1, \hat{\theta}_2]$, который с заданной вероятностью $\gamma$ «накрывает» неизвестный параметр $\theta$:

$$\boxed{P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \gamma}$$

Интервал $[\hat{\theta}_1, \hat{\theta}_2]$, называется *доверительным*, а $\gamma$ – *доверительной вероятностью*.

Величина $\alpha = 1 - \gamma$ называется уровнем значимости.

Границы интервала – функции от выборки:

$\hat{\theta}_1 = \hat{\theta}_1(x_1, \ldots, x_n)$, $\hat{\theta}_2 = \hat{\theta}_2(x_1, \ldots, x_n)$,

т.е. являются случайными величинами.

Замечание 1. Для дискретных распределений может не существовать такого интервала $[\hat{\theta}_1,\ \hat{\theta}_2]$, что $\boxed{P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \gamma}$

Тогда знак равенства заменяется на неравенство

$$\boxed{P(\hat{\theta}_1 < \theta < \hat{\theta}_2) \geq \gamma}$$

Замечание 2. В случае вектора параметров определяют доверительный интервал для каждой компоненты вектора параметров.

Замечание 3. В общем случае доверительных интервалов может быть несколько, поэтому из них выбирают интервал минимальной длины.

Желательно иметь $\gamma$ близким к единице, а интервал – наименьшей длины. Величина $\gamma$ показывает *надёжность* оценки, длина интервала – её *точность*.

Желательно иметь $\gamma$ близким к единице, а интервал – наименьшей длины. Величина $\gamma$ показывает *надёжность* оценки, длина интервала – её *точность*.

Обычно задают $\gamma = 0.95, \quad \gamma = 0.99$. Интервал также называют $\gamma \cdot 100\%$ доверительным интервалом.

Желательно иметь $\gamma$ близким к единице, а интервал – наименьшей длины. Величина $\gamma$ показывает *надёжность* оценки, длина интервала – её *точность*.

Обычно задают $\gamma = 0.95$, $\gamma = 0.99$. Интервал также называют $\gamma \cdot 100\%$ доверительным интервалом.

Рассмотрим основные способы построения интервальных оцениванок.

## Построение доверительного интервала с использованием распределения точечной оценки параметров

Если имеется некоторая точечная оценка $T_n = T(\mathbf{X}_n)$ для параметра $\theta$ и известна ее функция распределения $F_T(t,\theta)$ непрерывная и монотонная по $\theta$, то доверительный интервал можно построить, основываясь на этой функции:

1. Вычисляем точечную оценку $T_n = T(\mathbf{X}_n)$.
2. Решаем относительно $\tilde{\theta}_1, \tilde{\theta}_2$ уравнения

$$F_T\left(T_n, \tilde{\theta}_1\right) = \frac{1-\gamma}{2}, \; F_T\left(T_n, \tilde{\theta}_2\right) = \frac{1+\gamma}{2}.$$

3. Определяем границы доверительного интервала:

$$T_1 = \min\left(\tilde{\theta}_1, \tilde{\theta}_2\right), \; T_2 = \max\left(\tilde{\theta}_1, \tilde{\theta}_2\right).$$

## Построение доверительного интервала с использованием центральной статистики

Определение. Статистика $G\left(\mathbf{X}_n,\theta\right)$ называется центральной статистикой, если

1) распределение $G\left(\mathbf{X}_n,\theta\right)$ не зависит от $\theta$, и

2) при любом фиксированном $\theta$ статистика $G\left(\mathbf{X}_n,\theta\right)$ непрерывна и строго монотонна по $\theta$.

## Построение доверительного интервала с использованием центральной статистики

С помощью центральной статистики можно построить доверительный интервал. Пусть $f_G\left(g\right)$ плотность распределения статистики $G\left(\mathbf{X}_n, \theta\right)$.

1. Найдем такие значения $g_1, g_2$,что

$$P\left\{g_1 < G\left(\mathbf{X}_n, \theta\right) < g_2\right\} = \int_{g_1}^{g_2} f_G\left(g\right) dg = \gamma.$$

2. Решим относительно $\tilde{T}_1, \tilde{T}_2$ уравнения

$$G\left(\mathbf{X}_n, \tilde{T}_1\right) = g_1, \; G\left(\mathbf{X}_n, \tilde{T}_2\right) = g_2.$$

3. Определяем границы доверительного интервала:

$$T_1 = \min\left\{\tilde{T}_1, \tilde{T}_2\right\}, \, T_2 = \max\left\{\tilde{T}_1, \tilde{T}_2\right\}$$

## Построение доверительного интервала с использованием центральной статистики

Для построения доверительного интервала с помощью центральной статистики основная проблема заключается в нахождении этой центральной статистики. Можно выделить класс моделей, для которых центральная статистика существует и имеет простой вид.

Пусть $F(x,\theta)$ – функция распределения наблюдаемой случайной величины, монотонная по параметру $\theta$.

Можно положить в качестве центральной статистики функцию $G(\mathbf{X}_n,\theta) = -\sum_{i=1}^{n} \ln F(X_i,\theta)$, которая подчинена гамма-распределению с параметром формы $n$.

Пример. Построить точный $\gamma$-доверительный интервал по выборке $X_1, ..., X_n$ для параметра $\theta$ экспоненциального распределения $f(x;\theta) = \frac{1}{\theta}\exp\left\{-\frac{x}{\theta}\right\}$, $x \geq 0$.

Решение:

Проверим монотонность функции распределения по параметру $\theta$: $F'_\theta(x;\theta) = \frac{x}{\theta^2}e^{-\frac{x}{\theta}} > 0$, следовательно,

$G(\mathbf{X}_n, \theta) = -\sum_{i=1}^{n}\ln F(X_i, \theta)$ - центральная статистика.

Границы $\gamma$-доверительного интервала $(T_1, T_2)$ определяются при численном решении уравнений:

$G\left(\mathbf{X}_n, T_1\right) = g_1$, $G\left(\mathbf{X}_n, T_2\right) = g_2$,

где $g_1$ и $g_2$ : $P\left\{g_1 < G\left(\mathbf{X}_n, \theta\right) < g_2\right\} = \int\limits_{g_1}^{g_2}\frac{x^{n-1}e^{-x}}{\Gamma(n)}dx = \gamma$.

**Рассмотрим нахождение доверительных интервалов для параметров нормального распределения $a, \sigma$**

| | $a$ известно | $a$ неизвестно |
|---|---|---|
| $\sigma$ известно | | 1 задача |
| $\sigma$ неизвестно | 2 задача | 3,4 задачи |

## Задача 1. Доверительный интервал для математического ожидания нормального закона при известной дисперсии

Пусть $X \sim N(a, \sigma)$, причём σ известно. Имеется выборка

$$(x_1, x_2, \ldots , x_n).$$

По свойству нормального распределения, выборочное среднее $\bar{X} \sim N\left(a, \sigma / \sqrt{n}\right)$, поэтому $\dfrac{\bar{X} - a}{\sigma / \sqrt{n}} \sim N(0,1)$.

Следовательно, $P\left(\left|\dfrac{\bar{X} - a}{\sigma / \sqrt{n}}\right| < \varepsilon\right) = 2\Phi_0(\varepsilon)$,

где $\Phi_0$ - функция Лапласа.

Равносильное неравенство: $P\left(\bar{X} - \dfrac{\sigma\varepsilon}{\sqrt{n}} < a < \bar{X} + \dfrac{\sigma\varepsilon}{\sqrt{n}}\right) = 2\Phi(\varepsilon)$.

Процедура построения доверительного интервала для параметра $a$:

a) Задаётся доверительная вероятность $\gamma$.

b) По $\gamma$ с помощью таблицы функции Лапласа находится ε из уравнения $2\Phi_0 = \gamma$ (или $\varepsilon = \Phi_0^{-1}(\gamma/2)$).

Доверительный интервал:

$$\left(\bar{X} - \Delta, \bar{X} + \Delta\right),$$

где $\Delta = \dfrac{\sigma\varepsilon}{\sqrt{n}}$ - называется предельной ошибкой.

Замечания.

1. Длина доверительного интервала сколь угодно мала при больших $n$: $\dfrac{2\sigma\varepsilon}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0.$

2. Увеличение доверительной вероятности $\gamma$ приводит к увеличению $\varepsilon$ и $\Delta$ (так как функция Лапласа – возрастающая), то есть к уменьшению точности.

3. По двум известным величинам из $\Delta$, $\varepsilon$ (или $\gamma$), $n$ можно найти третью.

Например, пусть требуется оценить математическое ожидание с наперед заданной предельной ошибкой $\Delta$ и доверительной вероятностью $\gamma$. Тогда можно определить минимальный объем выборки, который обеспечит такую точность и надежность:

$$\Delta = \frac{\sigma \varepsilon}{\sqrt{n}} \quad \Rightarrow \quad \boxed{n = \frac{\sigma^2 \left(\Phi_0^{-1}(\gamma/2)\right)^2}{\Delta^2}}.$$

**Пример.** Случайная величина $\xi$ имеет нормальное распределение с известным средним квадратическим отклонением $\sigma = 2$.
Найти 95% доверительный интервал для неизвестного математического ожидания, если объем выборки $n = 49$, $\overline{x} = 5$.

По таблице функции Лапласа: $\gamma = 0.95, \quad 2\Phi_0(\varepsilon) = 0.95 \;\Rightarrow$ $\varepsilon = 1.96$.

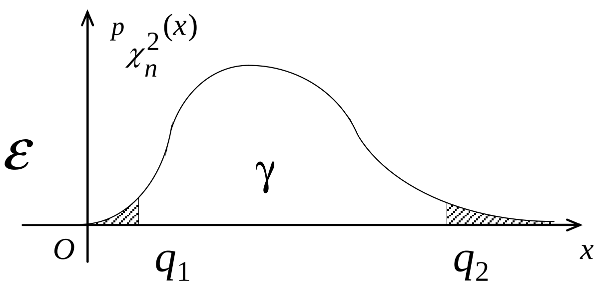
Предельная ошибка $\Delta = \dfrac{1.96 \cdot 2}{\sqrt{49}} = \dfrac{3.92}{7} = 0.56$.

Значит, с вероятностью $0.95$ интервал $(5 - 0.56, 5 + 0.56)$ содержит $EX$.

## Задача 2. Оценивание доверительного интервала для дисперсии при **известном** математическом ожидании

Пусть $X \sim N(a, \sigma)$, где $a$ известно. Случайные величины $\dfrac{X_i - a}{\sigma}$ независимы и подчиняются стандартному нормальному распределению $\Rightarrow \sum_{i=1}^{n} \left( \dfrac{X_i - a}{\sigma} \right)^2 \sim \chi_n^2$.

По таблице распределения $\chi^2$ для числа степеней свободы $n$ найдем квантили $q_1$ и $q_2$ такие, что $\chi_n^2(q_1) = \varepsilon / 2$, $\chi_n^2(q_2) = 1 - \varepsilon / 2$, где $\varepsilon = 1 - \gamma$ Тогда




$$P\left( q_1 < \sum_{i=1}^{n} \left( \frac{X_i - a}{\sigma} \right)^2 < q_2 \right) = \chi_n^2(q_2) - \chi_n^2(q_1) = 1 - \varepsilon$$

$$\Rightarrow P\left( \frac{\sum_{i=1}^{n} (X_i - a)^2}{q_2} < \sigma^2 < \frac{\sum_{i=1}^{n} (X_i - a)^2}{q_1} \right) = 1 - \varepsilon = \gamma.$$

**Пример.** Известно, что $X \sim N(4,\sigma)$. Результаты наблюдений: $\{3; 6; 2; 5; 4\}$. Найти оценку для $\sigma$; построить $95\%$ доверительный интервал.

**Решение.** $\varepsilon = 0.05$; по таблице $\chi^2$ для числа степеней свободы $5$ определим: $q_1 = 0.831$ и $q_2 = 12.832$.

$$\sum_{i=1}^{n}\left(X_i - a\right)^2 = (-1)^2 + 2^2 + (-2)^2 + 1^2 + 0^2 = 10.$$

Доверительный интервал для $\sigma^2$:

$$\left(\frac{10}{12.832} < \sigma^2 < \frac{10}{0.831}\right) \text{ или } \left(0.78 < \sigma^2 < 12.03\right) \Rightarrow \left(0.88 < \sigma < 3.46\right)$$

## Задача 3. Оценивание доверительного интервала для математического ожидания нормального распределения при **неизвестной** дисперсии

Пусть $X \sim N(a,\sigma)$, где $a,\sigma$ неизвестны. Будем искать интервал вида $\left(\bar{X}-\Delta, \bar{X}+\Delta\right)$, такой, что

$$P\left(|\bar{X}-a|<\Delta\right)=\gamma \Leftrightarrow P\left(\frac{\sqrt{n-1}}{S}\cdot|\bar{X}-a|<\frac{\sqrt{n-1}}{S}\cdot\Delta\right)=\gamma \Leftrightarrow$$

$$P\left(|T|<\frac{\sqrt{n-1}}{S}\Delta\right)=\gamma$$

по свойству распределения Стьюдента, где $T \sim t_{n-1}$.

По таблице распределения Стьюдента найдем для вероятности $\gamma$ и числа степеней свободы $n-1$ такое значение $t_{\gamma,n-1}$, что $P\left(|T|<t_{\gamma,n-1}\right)=\gamma$. Тогда

$$t_{\gamma,n-1}=\frac{\sqrt{n-1}}{S}\Delta \quad \Rightarrow \quad \boxed{\Delta=t_{\gamma,n-1}\cdot\frac{S}{\sqrt{n-1}}}.$$

Замечание. Эквивалентная формула: $\boxed{\Delta = t_{\gamma,n-1} \cdot \frac{\tilde{S}}{\sqrt{n}}}$.

**Пример.** Известно, что $X \sim N(a,\sigma)$. Результаты наблюдений: $\{3;6;2;5;4\}$. Найти оценку для $a$, если $\sigma$ неизвестно; построить $95\%$ доверительный интервал.

**Решение**.

$$\overline{x} = 4, \quad s^2 = \overline{x^2} - \left(\overline{x}\right)^2 = \frac{1}{5}\left(3^2 + 6^2 + 2^2 + 5^2 + 4^2\right) - 4^2 = 2.$$

значит оценка $a^* = 4$.

Оценка дисперсии $s^2 = 2$ .

Из таблицы Стьюдента $t_{0.95,4} = 2.78$.

Предельная ошибка $\Delta = t_{\gamma,n-1} \cdot \frac{s}{\sqrt{n-1}} = 2.78 \frac{\sqrt{2}}{\sqrt{4}} \approx 1.96$.

Интервал: $\left(4 - 1.96, 4 + 1.96\right) = \left(2.04, 5.96\right)$

Замечание. Существует три вида таблиц распределения Стьюдента:

1. Указываются значения $t$ непосредственно для $\gamma$.

| $f$ | $p$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **0.80** | **0.90** | **0.95** | **0.98** | **0.99** | **0.995** | **0.998** | **0.999** |
| **1** | 3.0770 | 6.3130 | 12.7060 | 31.8200 | 63.6560 | 127.6560 | 318.3060 | 636.6190 |
| **2** | 1.8850 | 2.9200 | 4.3020 | 6.9640 | 9.9240 | 14.0890 | 22.3270 | 31.5990 |
| **3** | 1.6377 | 2.35340 | 3.1820 | 4.5400 | 5.8400 | 7.4580 | 10.2140 | 12.9240 |
| **4** | 1.5332 | 2.13180 | 2.7760 | 3.7460 | 4.6040 | 5.5970 | 7.1730 | 8.6100 |
| **5** | 1.4759 | 2.01500 | 2.5700 | 3.6490 | 4.0321 | 4.7730 | 5.8930 | 6.8630 |

Замечание. Существует три вида таблиц распределения Стьюдента:

1. Указываются значения $t$ непосредственно для $\gamma$.

| *f* | *p* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **0.80** | **0.90** | **0.95** | **0.98** | **0.99** | **0.995** | **0.998** | **0.999** |
| **1** | 3.0770 | 6.3130 | 12.7060 | 31.8200 | 63.6560 | 127.6560 | 318.3060 | 636.6190 |
| **2** | 1.8850 | 2.9200 | 4.3020 | 6.9640 | 9.9240 | 14.0890 | 22.3270 | 31.5990 |
| **3** | 1.6377 | 2.35340 | 3.1820 | 4.5400 | 5.8400 | 7.4580 | 10.2140 | 12.9240 |
| **4** | 1.5332 | 2.13180 | 2.7760 | 3.7460 | 4.6040 | 5.5970 | 7.1730 | 8.6100 |
| **5** | 1.4759 | 2.01500 | 2.5700 | 3.6490 | 4.0321 | 4.7730 | 5.8930 | 6.8630 |

2. Указываются значения $t$ для $1-\gamma$ («двухсторонняя таблица»).

Замечание. Существует три вида таблиц распределения Стьюдента:

1. Указываются значения $t$ непосредственно для $\gamma$.

| $f$ | $p$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **0.80** | **0.90** | **0.95** | **0.98** | **0.99** | **0.995** | **0.998** | **0.999** |
| **1** | 3.0770 | 6.3130 | 12.7060 | 31.8200 | 63.6560 | 127.6560 | 318.3060 | 636.6190 |
| **2** | 1.8850 | 2.9200 | 4.3020 | 6.9640 | 9.9240 | 14.0890 | 22.3270 | 31.5990 |
| **3** | 1.6377 | 2.35340 | 3.1820 | 4.5400 | 5.8400 | 7.4580 | 10.2140 | 12.9240 |
| **4** | 1.5332 | 2.13180 | 2.7760 | 3.7460 | 4.6040 | 5.5970 | 7.1730 | 8.6100 |
| **5** | 1.4759 | 2.01500 | 2.5700 | 3.6490 | 4.0321 | 4.7730 | 5.8930 | 6.8630 |

2. Указываются значения $t$ для $1-\gamma$ («двухсторонння таблица»).

3. Указываются значения $t$ для $\frac{1-\gamma}{2}$ («односторонняя таблица»).

## Задача 4. Оценивание доверительного интервала для дисперсии при **неизвестном** математическом ожидании

Пусть $X \sim N(a,\sigma)$, где $a$ неизвестно. По свойству хи - квадрат распределения

$$\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1} \quad \Rightarrow$$

из таблицы распределения $\chi^2$ для числа степеней свободы $n-1$ найдем $q_1$ и $q_2$: $\chi^2_{n-1}(q_1) = \varepsilon/2$, $\chi^2_{n-1}(q_2) = 1-\varepsilon/2$ где $\varepsilon = 1-\gamma$. Тогда

$$P\left(q_1 < \frac{nS^2}{\sigma^2} < q_2\right) = \chi^2_{n-1}(q_2) - \chi^2_{n-1}(q_1) = 1-\varepsilon$$

$$\Downarrow$$

$$P\left(\frac{nS^2}{q_2} < \sigma^2 < \frac{nS^2}{q_1}\right) = 1-\varepsilon = \gamma.$$

**Пример.**

Известно, что $X \sim N(a,\sigma)$. Результаты наблюдений: $\{3;6;2;5;4\}$, $a,\sigma$ неизвестны. Определить $95\%$ доверительный интервал для $\sigma$.

**Решение.** $\varepsilon = 0.05$. Из таблицы распределения $\chi^2$ для числа степеней свободы $4$ найдем: $q_1 = 0.484$ и $q_2 = 11.143$.

$$\overline{x} = 4,\quad s^2 = \overline{x^2} - \left(\overline{x}\right)^2 = \frac{1}{5}\left(3^2 + 6^2 + 2^2 + 5^2 + 4^2\right) - 4^2 = 2.$$

Интервал: $\left(\dfrac{5\cdot 2}{11.143} < \sigma^2 < \dfrac{5\cdot 2}{0.484}\right)$. или $\left(0,897 < \sigma^2 < 20,661\right)$

$\Rightarrow\ \left(0,95 < \sigma < 4.55\right)$.

## Интервальное оценивание вероятности события

Пусть требуется дать интервальную оценку вероятности $p$ некоторого события $A$, которое произошло $m$ раз в $n$ независимых экспериментах. Распределение величины $m$ является биномиальным $Bin(n, p)$.

Точечная оценка максимального правдоподобия есть $p^* = \dfrac{m}{n}$, т.е. частота события.

По интегральной теореме Муавра-Лапласа:

$$P\left(a < \frac{m - np}{\sqrt{npq}} < b\right) \approx \Phi_0(b) - \Phi_0(a),$$

где $\Phi_0$ - функция Лапласа. Обозначим $a = -\varepsilon,\ \ b = \varepsilon$, тогда

$$P\left(\left|\frac{m - np}{\sqrt{npq}}\right| < \varepsilon\right) \approx \Phi_0(\varepsilon) - \Phi_0(-\varepsilon) = 2\Phi_0(\varepsilon). \quad (1)$$

Пусть $\gamma = 2\Phi_0(\varepsilon)$, тогда $\varepsilon = \Phi_0^{-1}\left(\gamma/2\right)$.

Равносильное (1) неравенство:

$$P(m^2 - 2mnp + n^2p^2 \le \varepsilon^2 npq) \approx \gamma.$$

Заменим $q = 1 - p$:

$$P(p^2\underbrace{(n^2 + \varepsilon^2 n)}_{c_1} - p\underbrace{(2mn + \varepsilon^2 n)}_{c_2} + m^2 \le 0) \approx \gamma$$

Кривая $y = c_1 p^2 + c_2 p + m$ - парабола; существуют два корня $p_1 < p_2$:

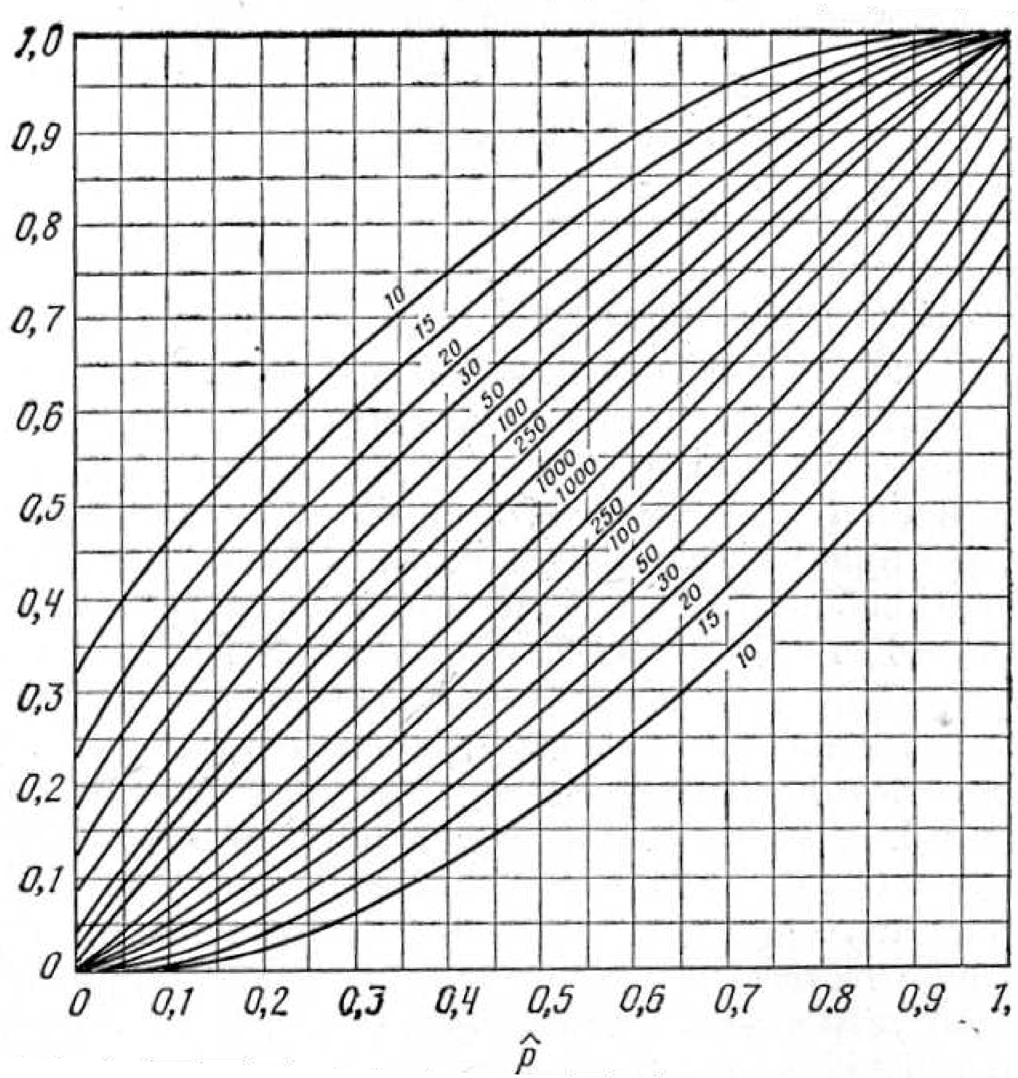
$$p_{1,2} = \frac{1}{(1 + \varepsilon^2/n)}\left(p^* + \frac{\varepsilon^2}{2n} \mp \varepsilon\sqrt{\frac{p^*(1-p^*)}{n} + \left(\frac{\varepsilon}{2n}\right)^2}\right).$$

Тогда

$P(p_1 < p < p_2) = \gamma$ - доверительный интервал для $p$.

Замечание. Интервал может быть несимметричным относительно $p^*$

Номограмма для определения доверительных интервалов в зависимости от частоты и объема выборки.




$$\gamma = 0.95$$

При больших $n$ можно пренебречь величинами с порядком малости $n^{-1}$ :

$$p_{1,2} \approx p^* \mp \varepsilon \sqrt{\frac{p^*(1-p^*)}{n}}.$$

Таким образом с вероятностью $\gamma$

$$p \in [p^* - \Delta, p^* + \Delta],$$

где

$$\Delta = \varepsilon \sqrt{\frac{p^*(1-p^*)}{n}}, \quad \varepsilon = \Phi_0^{-1}\left(\gamma/2\right).$$

**Пример.** Из 200 опрошенных 55% - сторонники некоторой партии. Определить интервал, в котором с вероятностью 0.95 находится процент ее сторонников среди всех жителей:

**Пример.** Из 200 опрошенных 55% - сторонники некоторой партии. Определить интервал, в котором с вероятностью 0.95 находится процент ее сторонников среди всех жителей:

доля по генеральной совокупности = вероятность;

$$\gamma = 0.95 \quad \varepsilon = 1.96,$$

$$\Delta = 1,96 \cdot \sqrt{\frac{0,55 \cdot 0,45}{200}} = 0,07.$$

**Пример.** Из 200 опрошенных 55% - сторонники некоторой партии. Определить интервал, в котором с вероятностью 0.95 находится процент ее сторонников среди всех жителей:

доля по генеральной совокупности = вероятность;

$$\gamma = 0.95 \quad \varepsilon = 1.96,$$

$$\Delta = 1,96 \cdot \sqrt{\frac{0,55 \cdot 0,45}{200}} = 0,07.$$

Вывод: с вероятностью $0.95$ общая доля сторонников партии находится в интервале от 48% до 62%.

**Пример.** Из 200 опрошенных 55% - сторонники некоторой партии. Определить интервал, в котором с вероятностью 0.95 находится процент ее сторонников среди всех жителей:

доля по генеральной совокупности = вероятность;

$$\gamma = 0.95 \quad \varepsilon = 1.96,$$

$$\Delta = 1,96 \cdot \sqrt{\frac{0,55 \cdot 0,45}{200}} = 0,07.$$

Вывод: с вероятностью $0.95$ общая доля сторонников партии находится в интервале от 48% до 62%.

Замечание. Для бесповторной выборки используется поправка:

$$\Delta = \varepsilon \cdot \sqrt{\frac{p^*(1-p^*)}{n}}\sqrt{1-\frac{n}{N}}.$$

# Интервальное оценивание среднего значения для генеральной совокупности

## Интервальное оценивание среднего значения для генеральной совокупности

Пусть генеральная совокупность $\Gamma$ состоит из $N$ единиц (объектов), а случайная выборка $v$ состоит из $n$ единиц.

## Интервальное оценивание среднего значения для генеральной совокупности

Пусть генеральная совокупность $\Gamma$ состоит из $N$ единиц (объектов), а случайная выборка $v$ состоит из $n$ единиц.

Требуется оценить среднее значение некоторого показателя $X$ для генеральной совокупности объектов:

$$EX = \frac{1}{N}\sum_{i\in\Gamma} X_i .$$

## Интервальное оценивание среднего значения для генеральной совокупности

Пусть генеральная совокупность $\Gamma$ состоит из $N$ единиц (объектов), а случайная выборка $v$ состоит из $n$ единиц. Требуется оценить среднее значение некоторого показателя $X$ для генеральной совокупности объектов:

$$EX = \frac{1}{N}\sum_{i\in\Gamma} X_i .$$

Рассмотрим независимую повторную выборку, для которой распределение наблюдений $X_i$ не меняется при $i = 1,\ldots,n$.

## Интервальное оценивание среднего значения для генеральной совокупности

Пусть генеральная совокупность $\Gamma$ состоит из $N$ единиц (объектов), а случайная выборка $v$ состоит из $n$ единиц. Требуется оценить среднее значение некоторого показателя $X$ для генеральной совокупности объектов:

$$EX = \frac{1}{N} \sum_{i \in \Gamma} X_i .$$

Рассмотрим независимую повторную выборку, для которой распределение наблюдений $X_i$ не меняется при $i = 1, ..., n$.

Центральная предельная теорема $\Rightarrow$ выборочное среднее $\bar{X}$ случайной выборки $\{X_1, ..., X_n\}$ при больших $n$ имеет распределение, близкое к нормальному, для произвольного распределения $X$:

$$\overline{X} \sim N(a, \sigma_{\overline{X}}),$$

где $a = EX$ (математическое ожидание $EX$ неизвестно),

$\sigma^2_{\overline{X}}$ - дисперсия выборочного среднего,

$$\sigma^2_{\overline{X}} = \sigma^2 / n,$$

$\sigma^2 = DX$ - неизвестная дисперсия случайной величины $X$.

$$\overline{X} \sim N(a, \sigma_{\overline{X}}),$$

где $a = EX$ (математическое ожидание $EX$ неизвестно),
$\sigma^2_{\overline{X}}$ - дисперсия выборочного среднего,

$$\sigma^2_{\overline{X}} = \sigma^2 / n,$$

$\sigma^2 = DX$ - неизвестная дисперсия случайной величины $X$.
Оценки параметров:

$$a^* = \overline{X}, \quad \tilde{S}_{\overline{X}} = \sqrt{\tilde{S}^2_{\overline{X}}},$$

где $\tilde{S}^2_{\overline{X}} = \dfrac{\tilde{S}^2}{n}, \quad \tilde{S}^2 = \dfrac{1}{n-1}\sum_i \left(X_i - \overline{X}\right)^2$ - несмещенная оценка дисперсии $DX$.

Пусть задана доверительная вероятность $\gamma$. Так как

$$\frac{\bar{X} - a}{\tilde{S}}\sqrt{n} \sim t_{n-1}$$

по свойству 3’ распределения Стьюдента,

Пусть задана доверительная вероятность $\gamma$. Так как

$$\frac{\overline{X}-a}{\tilde{S}}\sqrt{n} \sim t_{n-1}$$

по свойству 3' распределения Стьюдента,

то $\gamma \cdot 100\%$ доверительный интервал для $EX$:

$$\left(\overline{X}-\Delta, \overline{X}+\Delta\right),$$

Пусть задана доверительная вероятность $\gamma$. Так как

$$\frac{\overline{X} - a}{\tilde{S}}\sqrt{n} \sim t_{n-1}$$

по свойству 3' распределения Стьюдента,
то $\gamma \cdot 100\%$ доверительный интервал для $EX$ :

$$\left(\overline{X} - \Delta, \overline{X} + \Delta\right),$$

где

$$\Delta = t_{\gamma,n-1} \cdot \frac{\tilde{S}}{\sqrt{n}} = t_{\gamma,n-1} \cdot S_{\overline{X}},$$

Пусть задана доверительная вероятность $\gamma$. Так как

$$\frac{\bar{X} - a}{\tilde{S}}\sqrt{n} \sim t_{n-1}$$

по свойству 3' распределения Стьюдента,
то $\gamma \cdot 100\%$ доверительный интервал для $EX$:

$$\left(\bar{X} - \Delta, \bar{X} + \Delta\right),$$

где

$$\Delta = t_{\gamma,n-1} \cdot \frac{\tilde{S}}{\sqrt{n}} = t_{\gamma,n-1} \cdot S_{\overline{X}},$$

$t_{\gamma,n-1}$ - определяется по таблице распределения Стьюдента.

Пусть задана доверительная вероятность $\gamma$. Так как

$$\frac{\overline{X} - a}{\tilde{S}}\sqrt{n} \sim t_{n-1}$$

по свойству 3' распределения Стьюдента,

то $\gamma \cdot 100\%$ доверительный интервал для $EX$:

$$\left(\overline{X} - \Delta, \overline{X} + \Delta\right),$$

где

$$\Delta = t_{\gamma,n-1} \cdot \frac{\tilde{S}}{\sqrt{n}} = t_{\gamma,n-1} \cdot S_{\overline{X}},$$

$t_{\gamma,n-1}$ - определяется по таблице распределения Стьюдента.

$S_{\overline{X}}$ - называют стандартной ошибкой среднего.

**Пример**. Выборка из большой партии электроламп содержит 100 ламп. Среднее время горения лампы из выборки равно 1000 часов.

**Пример**. Выборка из большой партии электроламп содержит 100 ламп. Среднее время горения лампы из выборки равно 1000 часов. Найдем 99% доверительный интервал для среднего времени горения лампы всей партии при условии, что среднее квадратическое отклонение времени горения лампы в выборке равно 30 часам:

**Пример**. Выборка из большой партии электроламп содержит 100 ламп. Среднее время горения лампы из выборки равно 1000 часов. Найдем 99% доверительный интервал для среднего времени горения лампы всей партии при условии, что среднее квадратическое отклонение времени горения лампы в выборке равно 30 часам:

$$n = 100,\ \overline{x} = 1000,\ \tilde{s} = 30,\ \gamma = 0.99 \ \Rightarrow$$

$$t_{0.99,99} = 2.63,\ S_{\overline{X}} = \frac{30}{\sqrt{100}} = 3,\ \Delta = 2.63 \cdot 3 = 7.89.$$

**Пример**. Выборка из большой партии электроламп содержит 100 ламп. Среднее время горения лампы из выборки равно 1000 часов. Найдем 99% доверительный интервал для среднего времени горения лампы всей партии при условии, что среднее квадратическое отклонение времени горения лампы в выборке равно 30 часам:

$$n = 100,\ \bar{x} = 1000,\ \tilde{s} = 30,\ \gamma = 0.99 \ \Rightarrow$$

$$t_{0.99,99} = 2.63,\ S_{\bar{X}} = \frac{30}{\sqrt{100}} = 3,\ \Delta = 2.63 \cdot 3 = 7.89.$$

Значит, $99\%$ доверительный интервал для среднего по всей партии: $(992.11,\ 1007.89)$.

Замечание 1. При больших $n$ (обычно при $n > 60$) для нахождения $t_{\gamma,n-1}$ можно воспользоваться аппроксимацией нормальным распределением (например, $t_{0.95,n-1} \approx 1.96$ при $n > 60$).

Замечание 1. При больших $n$ (обычно при $n > 60$) для нахождения $t_{\gamma,n-1}$ можно воспользоваться аппроксимацией нормальным распределением (например, $t_{0.95,n-1} \approx 1.96$ при $n > 60$).

Замечание 2. В случае бесповторной выборки, когда проводится отбор из генеральной совокупности объема $N$, необходимо учитывать поправку:

Замечание 1. При больших $n$ (обычно при $n > 60$) для нахождения $t_{\gamma,n-1}$ можно воспользоваться аппроксимацией нормальным распределением (например, $t_{0.95,n-1} \approx 1.96$ при $n > 60$).

Замечание 2. В случае бесповторной выборки, когда проводится отбор из генеральной совокупности объема $N$, необходимо учитывать поправку:

$$\Delta = t_{\gamma.n-1} \cdot \frac{\tilde{S}}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}},$$

где $\frac{n}{N}$ - доля отбора. Обычно поправка используется, если доля отбора выше $0.05$.

## Асимптотические доверительные интервалы

ОМП при достаточно общих условиях являются асимптотически нормальными, следовательно

$$P\left\{\left|\hat{\theta}_n - \theta\right|\sqrt{ni\left(\hat{\theta}_n\right)} \le c_\gamma\right\} \to \Phi\left(c_\gamma\right) - \Phi\left(-c_\gamma\right) = 2\Phi\left(c_\gamma\right) - 1 = \gamma$$

где $\Phi\left(x\right)$ – ф. р. стандартного нормального закона,

$i(\theta)$ – информационное количество Фишера, $\hat{\theta}_n$ – ОМП.

Тогда $\left(\hat{\theta}_n - \dfrac{c_\gamma}{\sqrt{ni\left(\hat{\theta}_n\right)}} ; \hat{\theta}_n + \dfrac{c_\gamma}{\sqrt{ni\left(\hat{\theta}_n\right)}}\right)$ – асимптотически кратчайший $\gamma$-доверительный интервал для $\theta$.

где $c_\gamma = \Phi^{-1}\left(\dfrac{\gamma + 1}{2}\right)$