Лекция 4

Доверительные интервалы для параметров нормального распределения

Распределение Стьюдента

Напоминалка

Мы рассматривали случайную выборку

$$X_i \sim \text{i.i.d.}, \quad \mathrm{E}(X_i) = \mu, \quad \mathrm{D}(X_i) = \sigma^2.$$

Оценка для математического ожидания:

- выборочное среднее $$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Для дисперсии:

- выборочная дисперсия $$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2;$$
- скорректированная выборочная дисперсия $$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Что мы о них знаем:

$$\mathrm{E}(\bar{X}) = \mu, \qquad \mathrm{D}(\bar{X}) = \frac{\sigma^2}{n}.$$

$$\mathrm{E}(S^2) = \frac{n-1}{n}\sigma^2, \qquad \mathrm{E}(\hat{\sigma}^2) = \sigma^2.$$

Чего нам не хватает?

Оценка — случайная величина. Она, как правило, не совпадает с оцениваемым параметром. В большинстве практически важных задач:

$$P\left(\hat{\theta}=\theta\right)=0.$$

Увы.

Мы не можем получить по выборке точную оценку какой-либо характеристики генеральной совокупности.

А что можем?

Доверительные интервалы

Пусть имеется случайная выборка $X_1, \ldots, X_n$, по которой оценивается параметр θ.

Представим себе случайные величины L и U, ограничивающие параметр θ с некоторой вероятностью γ:

$$P(L<\theta<U)=\gamma, \quad \forall\,\theta\in\Theta.$$

область допустимых
значений параметра

Интервал (L, U) называется доверительным интервалом для параметра θ с доверительной вероятностью (уровнем доверия) γ.

Подступая с таким интервалом к оцениванию, мы будем знать, что с вероятностью θ мы найдём верные границы для нужного нам параметра.

Напоминалка: теорема Фишера

Пусть $X_1, \ldots, X_n$ независимы, $X_i \sim N(\mu, \sigma^2)$.

Тогда:

(1) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right);$

(2) $\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1};$

(3) $\bar{X}$ и $\hat{\sigma}^2$ независимы.

I. Доверительный интервал для среднего при известной дисперсии

Имеется выборка из нормального распределения:

$$X_1, \ldots, X_n \text{ независимы}, \quad X_i \sim N(\mu, \sigma^2).$$

хотим оценить известна

Задача: построить доверительный интервал для μ.

Уровень доверия: $1\text{-}\alpha$ (α - вероятность ошибки доверительного интервала).

Опираемся на выборочное среднее:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

I. Доверительный интервал для среднего при известной дисперсии

$$X_1, \ldots, X_n \text{ независимы,} \quad X_i \sim N(\mu, \sigma^2).$$

Задача: построить доверительный интервал для μ.

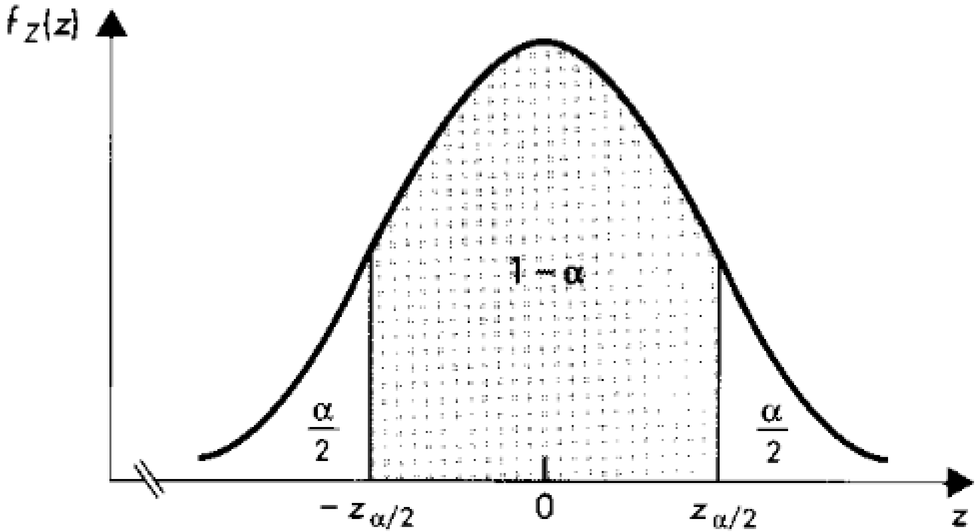
Уровень доверия: $1\text{-}\alpha$ (α - вероятность ошибки доверительного интервала).

Опираемся на выборочное среднее:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Возьмём такое число $z_{\frac{\alpha}{2}}$, что $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

Тогда $1 - \alpha = P(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}})$



картинка из Ньюболда

I. Доверительный интервал для среднего при известной дисперсии

$$X_1, \ldots, X_n \text{ независимы,} \quad X_i \sim N(\mu, \sigma^2).$$

Задача: построить доверительный интервал для μ.

Уровень доверия: $1\text{-}\alpha$ (α - вероятность ошибки доверительного интервала).

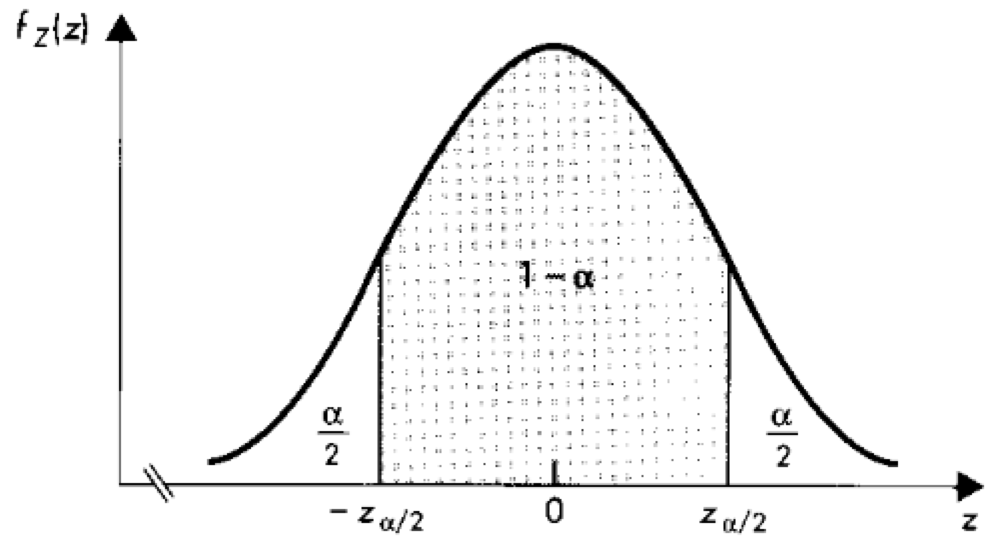
Опираемся на выборочное среднее:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Возьмём такое число $z_{\frac{\alpha}{2}}$, что $P\left(Z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$.

Тогда $1 - \alpha = P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) =$

$$= P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}}\right) =$$



картинка из Ньюболда

I. Доверительный интервал для среднего при известной дисперсии

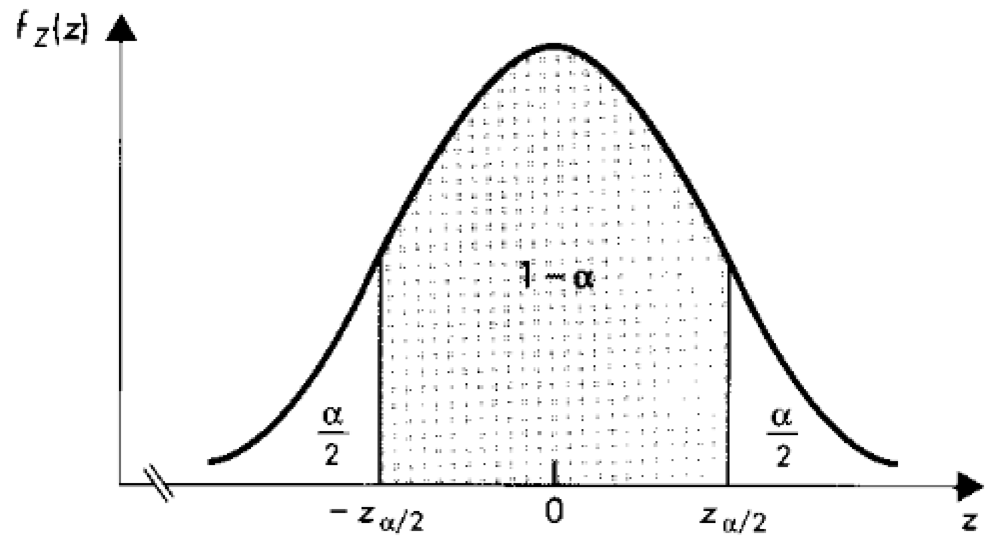
$$X_1, \ldots, X_n \text{ независимы}, \quad X_i \sim N(\mu, \sigma^2).$$

Задача: построить доверительный интервал для μ.
Уровень доверия: $1\text{-}\alpha$ (α - вероятность ошибки доверительного интервала).

Опираемся на выборочное среднее:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Возьмём такое число $z_{\frac{\alpha}{2}}$, что $P\left(Z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$.



картинка из Ньюболда

Тогда $1 - \alpha = P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) =$

$$= P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}}\right) =$$

$$= P\left(-z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) =$$

I. Доверительный интервал для среднего при известной дисперсии

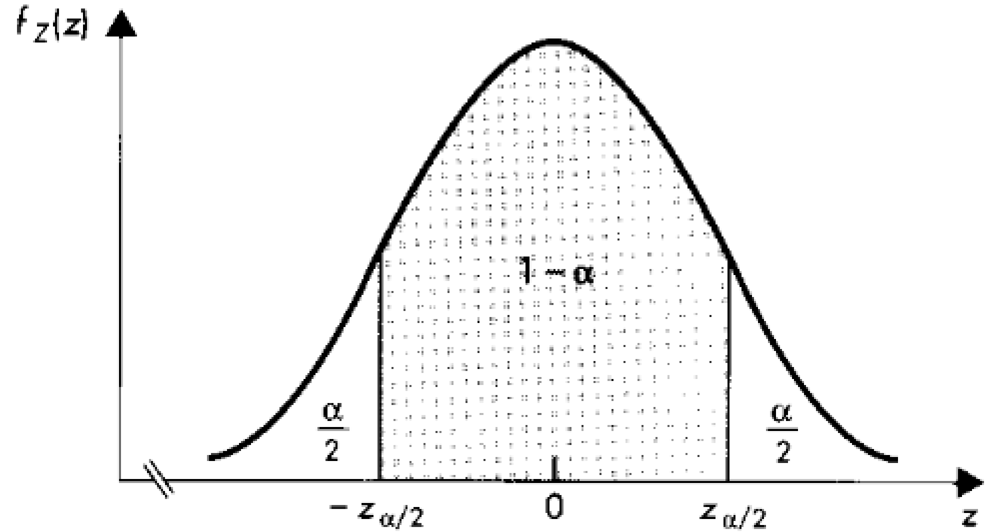
$$X_1,\ldots,X_n \text{ независимы, } \quad X_i \sim N(\mu,\sigma^2).$$

Задача: построить доверительный интервал для μ.
Уровень доверия: $1\text{-}\alpha$ (α - вероятность ошибки доверительного интервала).

Опираемся на выборочное среднее:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Возьмём такое число $z_{\frac{\alpha}{2}}$, что $P\left(Z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$.



картинка из Ньюболда

Тогда $1-\alpha = P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) =$

$$= P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}}\right) =$$

$$= P\left(-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(-\bar{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) =$$

I. Доверительный интервал для среднего при известной дисперсии

$$X_1,\ldots,X_n \text{ независимы, } \quad X_i \sim N(\mu,\sigma^2).$$

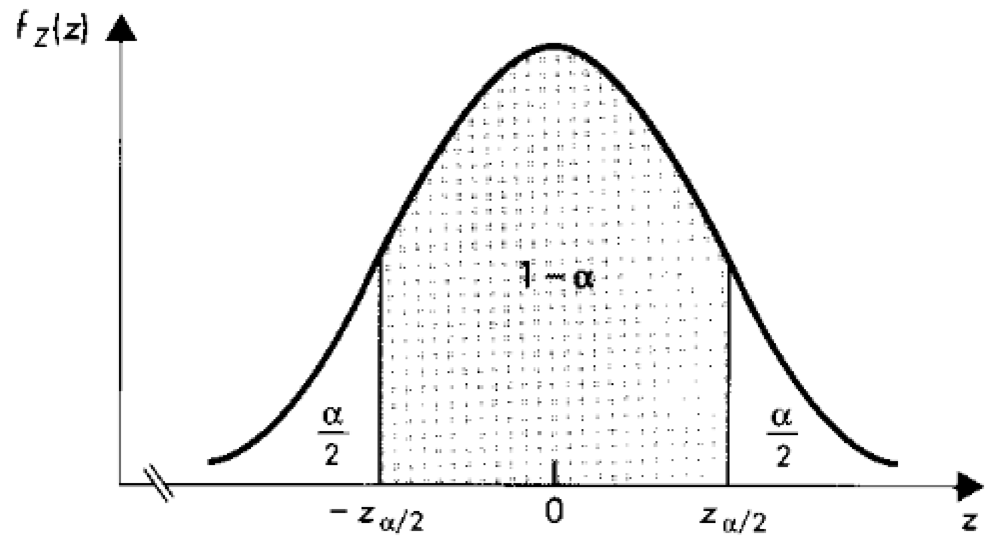
Задача: построить доверительный интервал для μ.

Уровень доверия: $1\text{-}\alpha$ (α - вероятность ошибки доверительного интервала).

Опираемся на выборочное среднее:

$$\bar{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right) \quad \Rightarrow \quad Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Возьмём такое число $z_{\frac{\alpha}{2}}$, что $P\left(Z>z_{\frac{\alpha}{2}}\right)=\frac{\alpha}{2}$.



картинка из Ньюболда

Тогда $1-\alpha = P\left(-z_{\frac{\alpha}{2}}<Z<z_{\frac{\alpha}{2}}\right)=$

$$=P\left(-z_{\frac{\alpha}{2}}<\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<z_{\frac{\alpha}{2}}\right)=$$

$$=P\left(-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<\bar{X}-\mu<z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)=$$

$$=P\left(-\bar{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<-\mu<-\bar{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)=P\left(\bar{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right).$$

Вывод

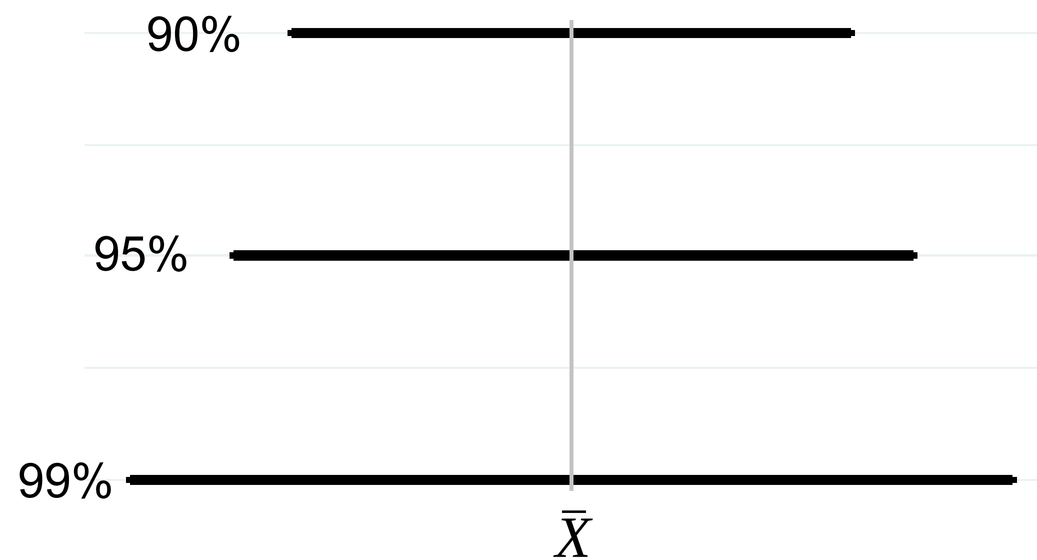
$$\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Доверительный интервал для среднего с уровнем доверия $(1-\alpha)$ в случае

- нормальной генеральной совокупности,
- известной дисперсии.

Торг: точность против надёжности

Интервалы для среднего с разным уровнем доверия:



Пример

Геодезист измеряет расстояние до удалённого объекта. Вот результаты четырёх независимых измерений в метрах:

262.2 263.2 262.0 264.5

Ошибки измерения нормально распределены с нулевым математическим ожиданием и стандартным отклонением 1.2 м. Рассчитайте 90% доверительный интервал для измеряемого расстояния.

Пример

Геодезист измеряет расстояние до удалённого объекта. Вот результаты четырёх независимых измерений в метрах:

262.2 263.2 262.0 264.5

Ошибки измерения нормально распределены с нулевым математическим ожиданием и стандартным отклонением 1.2 м. Рассчитайте 90% доверительный интервал для измеряемого расстояния.

Решение. Применяем доверительный интервал $\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$.

Рассчитываем среднее: $\bar{X} = \dfrac{262.2 + 263.2 + 262.0 + 264.6}{4} = 263.$

осторожно, двойственное обозначение!

По таблице стандартного нормального распределения $z_{\frac{\alpha}{2}} = z_{\frac{0.1}{2}} = 1.645.$

Нижняя граница доверительного интервала: $263 - \dfrac{1.645 \times 1.2}{\sqrt{4}} = 262.013.$

Верхняя граница доверительного интервала: $263 + \dfrac{1.645 \times 1.2}{\sqrt{4}} = 263.987.$

Вывод: расстояние находится в пределах примерно от 262 до 264 метров.

Откуда нам в жизни знать дисперсию?!

II. Доверительный интервал для дисперсии

$X_1, \dots, X_n$ независимы, $X_i \sim N(\mu, \sigma^2)$.

Хотим получить доверительный интервал для σ^2 с уровнем доверия $(1\text{-}\alpha)$.

Опираемся на оценку $\hat{\sigma}^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$

По т. Фишера $\dfrac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}.$

II. Доверительный интервал для дисперсии

$X_1, \dots, X_n$ независимы, $X_i \sim N(\mu, \sigma^2)$.

Хотим получить доверительный интервал для σ^2 с уровнем доверия $(1\text{-}\alpha)$.

Опираемся на оценку $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

По т. Фишера $\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$.

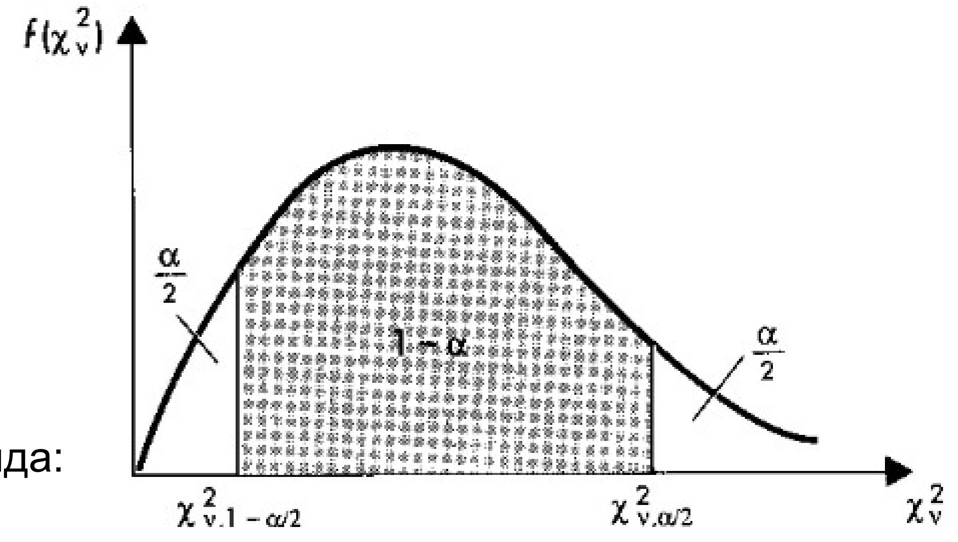
Возьмём такое число $\chi^2_{n-1,\frac{\alpha}{2}}$, что $P\left(Y > \chi^2_{n-1,\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$, где $Y \sim \chi^2_{n-1}$.

$\chi^2_{n-1,\frac{\alpha}{2}}$ - квантиль распределения хи-квадрат с $(n\text{-}1)$ степенью свободы порядка $1-\frac{\alpha}{2}$.

И ещё нам пригодится значение $\chi^2_{n-1,1-\frac{\alpha}{2}}$ (аналогично определяется).

$$1-\alpha = P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} < Y < \chi^2_{n-1,\frac{\alpha}{2}}\right)$$

Картиночка из Ньюболда:



II. Доверительный интервал для дисперсии

$X_1, \ldots, X_n$ независимы, $X_i \sim N(\mu, \sigma^2)$.

Хотим получить доверительный интервал для σ^2 с уровнем доверия $(1\text{-}\alpha)$.

Опираемся на оценку $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

По т. Фишера $\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$.

Возьмём такое число $\chi^2_{n-1,\frac{\alpha}{2}}$, что $P\left(Y > \chi^2_{n-1,\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$, где $Y \sim \chi^2_{n-1}$.

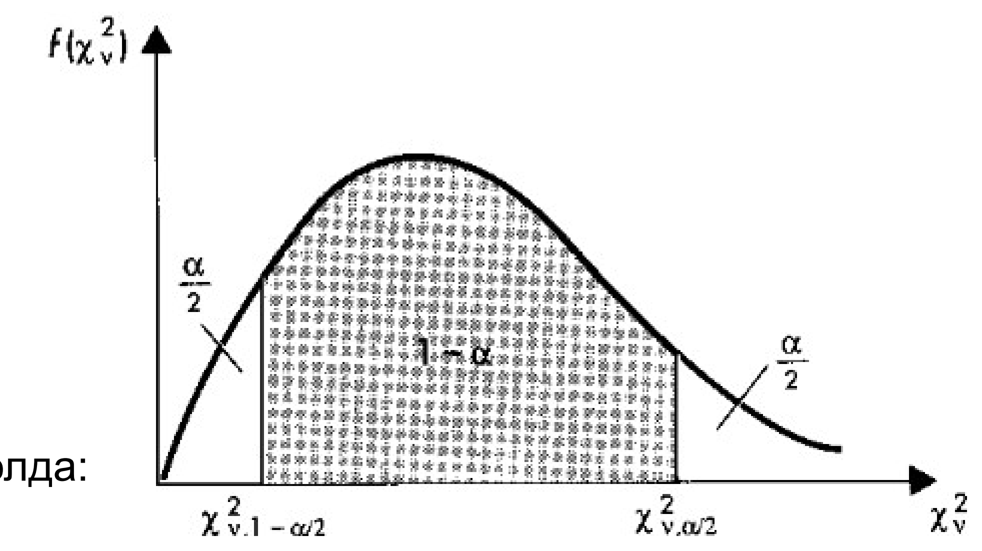
$\chi^2_{n-1,\frac{\alpha}{2}}$ - квантиль распределения хи-квадрат с $(n\text{-}1)$ степенью свободы порядка $1 - \frac{\alpha}{2}$.

И ещё нам пригодится значение $\chi^2_{n-1,1-\frac{\alpha}{2}}$ (аналогично определяется).

$$1 - \alpha = P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} < Y < \chi^2_{n-1,\frac{\alpha}{2}}\right) =$$

$$= P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} < \frac{(n-1)\hat{\sigma}^2}{\sigma^2} < \chi^2_{n-1,\frac{\alpha}{2}}\right) =$$

Картиночка из Ньюболда:



II. Доверительный интервал для дисперсии

$$X_1, \dots, X_n \text{ независимы}, \quad X_i \sim N(\mu, \sigma^2).$$

Хотим получить доверительный интервал для σ^2 с уровнем доверия $(1-\alpha)$.

Опираемся на оценку $\hat{\sigma}^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$

По т. Фишера $\dfrac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}.$

Возьмём такое число $\chi^2_{n-1,\frac{\alpha}{2}}$, что $P\left(Y > \chi^2_{n-1,\frac{\alpha}{2}}\right) = \dfrac{\alpha}{2}$, где $Y \sim \chi^2_{n-1}.$

$\chi^2_{n-1,\frac{\alpha}{2}}$ - квантиль распределения хи-квадрат с $(n$-$1)$ степенью свободы порядка $1-\dfrac{\alpha}{2}$.

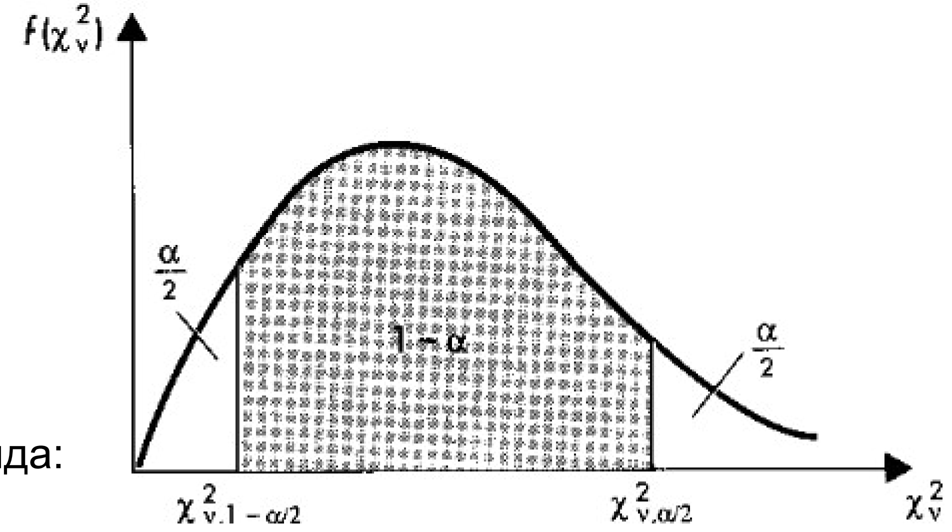
И ещё нам пригодится значение $\chi^2_{n-1,1-\frac{\alpha}{2}}$ (аналогично определяется).

$$1-\alpha = P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} < Y < \chi^2_{n-1,\frac{\alpha}{2}}\right) =$$

$$= P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} < \frac{(n-1)\hat{\sigma}^2}{\sigma^2} < \chi^2_{n-1,\frac{\alpha}{2}}\right) =$$

$$= P\left(\frac{1}{\chi^2_{n-1,\frac{\alpha}{2}}} < \frac{\sigma^2}{(n-1)\hat{\sigma}^2} < \frac{1}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right) =$$

Картиночка из Ньюболда:



II. Доверительный интервал для дисперсии

$X_1, \dots, X_n$ независимы, $X_i \sim N(\mu, \sigma^2)$.

Хотим получить доверительный интервал для σ^2 с уровнем доверия $(1\text{-}\alpha)$.

Опираемся на оценку $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

По т. Фишера $\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$.

Возьмём такое число $\chi^2_{n-1,\frac{\alpha}{2}}$, что $P(Y > \chi^2_{n-1,\frac{\alpha}{2}}) = \frac{\alpha}{2}$, где $Y \sim \chi^2_{n-1}$.

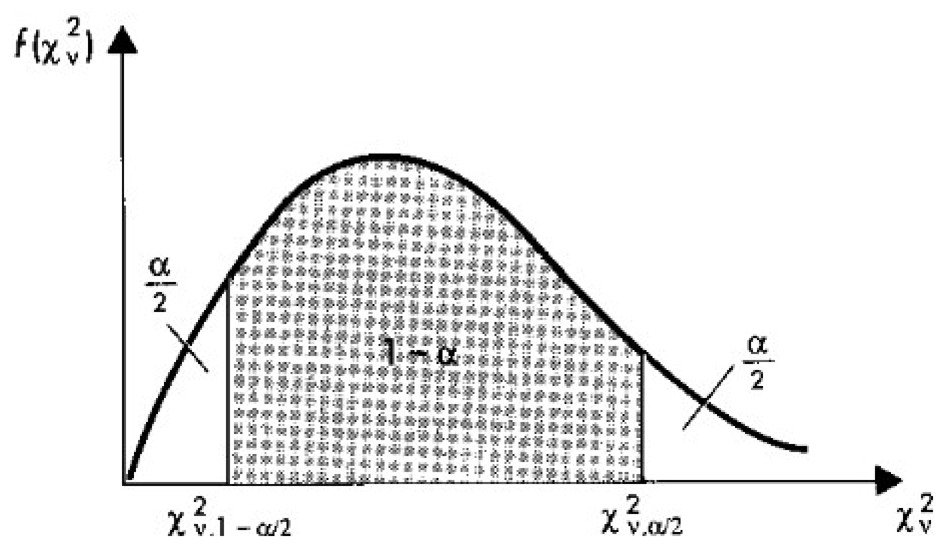
И ещё нам пригодится значение $\chi^2_{n-1,1-\frac{\alpha}{2}}$ (аналогично определяется).

$$1-\alpha = P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} < Y < \chi^2_{n-1,\frac{\alpha}{2}}\right) =$$

$$= P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} < \frac{(n-1)\hat{\sigma}^2}{\sigma^2} < \chi^2_{n-1,\frac{\alpha}{2}}\right) =$$

$$= P\left(\frac{1}{\chi^2_{n-1,\frac{\alpha}{2}}} < \frac{\sigma^2}{(n-1)\hat{\sigma}^2} < \frac{1}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right) =$$

$$= P\left(\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right).$$



Вывод

$$\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}$$

Доверительный интервал для дисперсии с уровнем доверия $(1-\alpha)$

в случае нормальной генеральной совокупности.

Пример

Крот Борислав, обследовав 8 случайно отобранных образцов свёклы на очень-очень большом огороде Владислава Валерьевича, измерил вес корнеплодов в граммах:

157 201 194 197 204 152 199 143.

Предположив, что вес распределён нормально, рассчитайте 95% доверительный интервал для дисперсии веса свёклы в огороде.

Пример

Крот Борислав, обследовав 8 случайно отобранных образцов свёклы на очень-очень большом огороде Владислава Валерьевича, измерил вес корнеплодов в граммах:

157 201 194 197 204 152 199 143.

Предположив, что вес распределён нормально, рассчитайте 95% доверительный интервал для дисперсии веса свёклы в огороде.

Решение. Оценим среднее и дисперсию по выборке:

$$\bar{X} = \frac{157 + 201 + 194 + 197 + 204 + 152 + 199 + 143}{8} = 180.875.$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{1}{7}\sum_{i=1}^{8}\left(X_i - 180.875\right)^2 = 648.411.$$

Квантили: $\chi^2_{7,\frac{0.05}{2}} = 16.013\,;\quad \chi^2_{7,1-\frac{0.05}{2}} = 1.670.$

Нижняя граница 95% интервала: $\dfrac{(n-1)\hat{\sigma}^2}{\chi_{n-1,\frac{\alpha}{2}}} = \dfrac{(8-1)\times 648.411}{16.013} = 283.450.$

Верхняя граница 95% интервала: $\dfrac{(n-1)\hat{\sigma}^2}{\chi_{n-1,1-\frac{\alpha}{2}}} = \dfrac{(8-1)\times 648.411}{1.670} = 2717.890.$

Пример

Крот Борислав, обследовав 8 случайно отобранных образцов свёклы на очень-очень большом огороде Владислава Валерьевича, измерил вес корнеплодов в граммах:

157 201 194 197 204 152 199 143.

Предположив, что вес распределён нормально, рассчитайте 95% доверительный интервал для дисперсии веса свёклы в огороде.

Ответ. Доверительный интервал для дисперсии:

$$283.450 < \sigma^2 < 2717.890\,.$$

Можно взять корень и получить интервал для стандартного отклонения:

$$16.836 < \sigma < 52.133\,.$$

А зачем была сделана оговорка про большой-большой огород?

Распределение Стьюдента
(t-распределение)

Пусть с.в. $Z \sim N(0,1)$, $Y \sim \chi_k^2$, Z и Y независимы;

$$U = \frac{Z}{\sqrt{\frac{1}{k} Y}}.$$

Распределение величины U называется распределением Стьюдента с k степенями свободы.

Обозначается $U \sim t_k$.

$$\mathrm{E}(U) = 0 \quad \text{при} \quad k > 1,$$

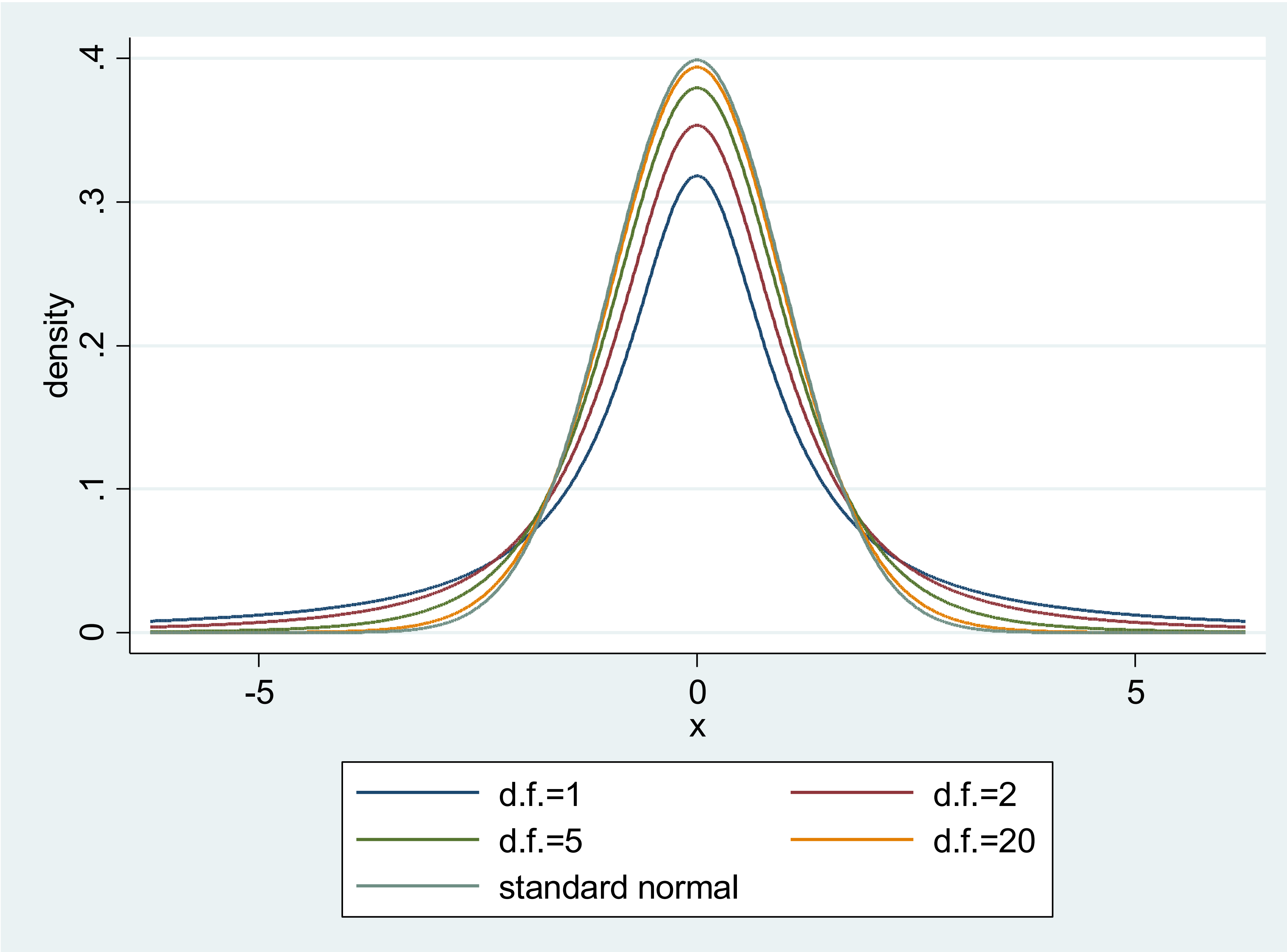
$$\mathrm{D}(U) = \frac{k}{k-2} \quad \text{при} \quad k > 2.$$

Плотность симметрична относительно нуля.

При больших k похоже на $N(0,1)$.

Стьюдент
(настоящее имя — Уильям Сили Госсет),
главный пивовар Гиннесс.

Функции плотности распределения Стьюдента
и стандартного нормального распределения



Следствие из теоремы Фишера

При выполнении предпосылок т. Фишера

$$\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}.$$

Доказательство:

$$\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} = \frac{\left(\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)}{\left(\dfrac{\hat{\sigma}/\sqrt{n}}{\sigma/\sqrt{n}}\right)} = \frac{\left(\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)}{\sqrt{\dfrac{\hat{\sigma}^2}{\sigma^2}}} = \frac{\left(\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)}{\sqrt{\dfrac{1}{n-1}\dfrac{(n-1)\hat{\sigma}^2}{\sigma^2}}}$$

Числитель распределён $N(0,1)$, знаменатель — $\sqrt{\dfrac{1}{n-1}\chi^2_{n-1}}$.
Числитель и знаменатель независимы.

Получается то же соотношение, что и в определении распределения Стьюдента:

$$U = \frac{Z}{\sqrt{\dfrac{1}{k}Y}}.$$

III. Доверительный интервал для среднего при неизвестной дисперсии

Имеется выборка из нормального распределения:

$$X_1, \ldots, X_n \text{ независимы}, \quad X_i \sim N\left(\mu, \sigma^2\right).$$

не знаем

Задача: построить доверительный интервал для μ с уровнем доверия 1-α.

Опираемся на выборочное среднее и оценку для дисперсии:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right); \qquad U = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}.$$

III. Доверительный интервал для среднего при неизвестной дисперсии

$$X_1,\dots,X_n \text{ независимы}, \quad X_i \sim N(\mu,\sigma^2).$$

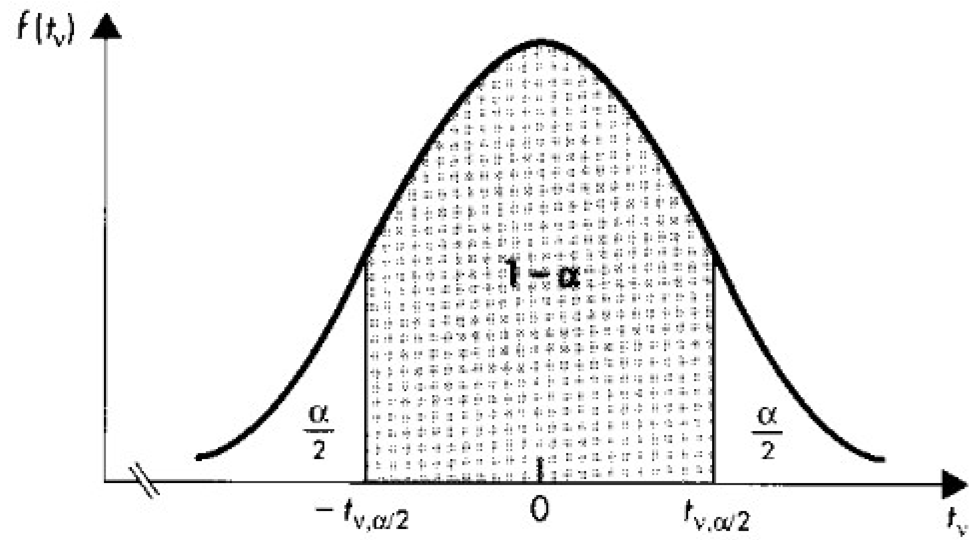
Задача: построить доверительный интервал для μ с уровнем доверия 1-α.

Опираемся на выборочное среднее и оценку для дисперсии:

$$\bar{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right); \qquad U = \frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}.$$

Возьмём такое число $t_{n-1,\frac{\alpha}{2}}$, что $P\left(U > t_{n-1,\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$.

картинка из Ньюболда



Тогда $1-\alpha = P\left(-t_{n-1,\frac{\alpha}{2}} < U < t_{n-1,\frac{\alpha}{2}}\right) =$

$$= P\left(-t_{n-1,\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} < t_{n-1,\frac{\alpha}{2}}\right) =$$

III. Доверительный интервал для среднего при неизвестной дисперсии

$X_1, \ldots, X_n$ независимы, $X_i \sim N(\mu, \sigma^2)$.

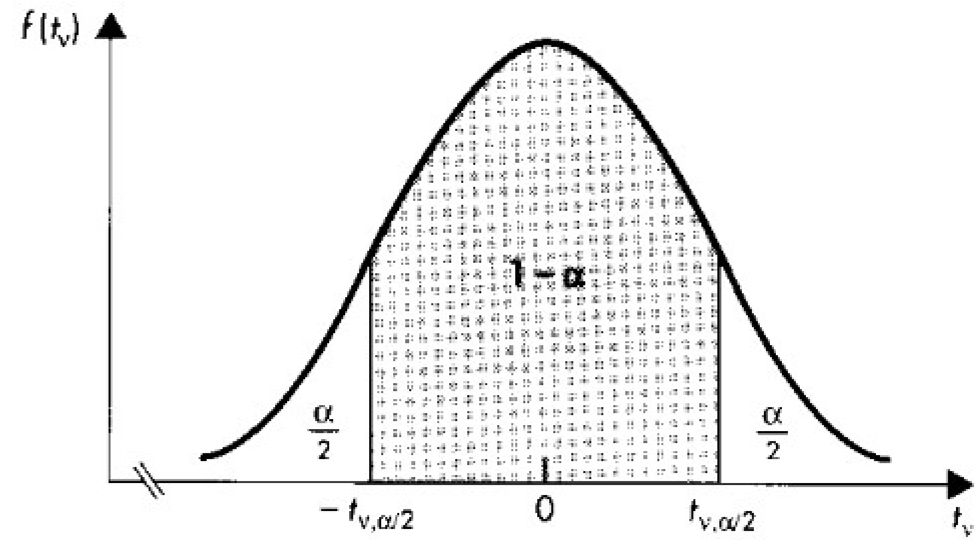
Задача: построить доверительный интервал для μ с уровнем доверия 1-α.

Опираемся на выборочное среднее и оценку для дисперсии:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right); \qquad U = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}.$$

Возьмём такое число $t_{n-1,\frac{\alpha}{2}}$, что $P\left(U > t_{n-1,\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$.

картинка из Ньюболда



Тогда $1 - \alpha = P\left(-t_{n-1,\frac{\alpha}{2}} < U < t_{n-1,\frac{\alpha}{2}}\right) =$

$$= P\left(-t_{n-1,\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} < t_{n-1,\frac{\alpha}{2}}\right) =$$

$$= P\left(-t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} < \bar{X} - \mu < t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}\right) =$$

$$= P\left(-\bar{X} - t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} < -\mu < -\bar{X} + t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}\right) = P\left(\bar{X} - t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}\right).$$

Вывод

$$\bar{X} - t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}$$

Доверительный интервал для среднего с уровнем доверия $(1-\alpha)$ в случае

- нормальной генеральной совокупности,
- неизвестной дисперсии.

Как он соотносится с интервалом при известной дисперсии?

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

Пример

Вернёмся к кроту Бориславу и его замерам:

157 201 194 197 204 152 199 143.

Теперь рассчитаем 95% доверительный интервал для среднего веса свёклы в огороде, предполагая, что вес нормально распределён.

Решение. В прошлый раз мы уже рассчитали среднее и дисперсию по выборке:

$$\bar{X} = 180.875, \quad \hat{\sigma}^2 = 648.411.$$

Выборочное стандартное отклонение: $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{648.411} = 25.464.$

Квантиль порядка 97.5%: $t_{7, \frac{0.05}{2}} = 2.365.$

Нижняя граница 95% интервала: $180.875 - \dfrac{2.365 \times 25.464}{\sqrt{8}} = 159.583.$

Верхняя граница 95% интервала: $180.875 + \dfrac{2.365 \times 25.464}{\sqrt{8}} = 202.167.$

Ответ. Средний вес свёклы в граммах лежит в пределах от 159.583 до 202.167.

Вывод и пища для размышлений

Итак, мы рассчитали 95% доверительный интервал для μ - среднего веса свёклы во всём огороде:

$$159.583 < \mu < 202.167\,.$$

Это не означает, что с вероятностью 95% средний вес плода лежит в указанных пределах.

А что это означает?

Следующая лекция

Асимптотический доверительный интервал для доли

Определение объёма выборки