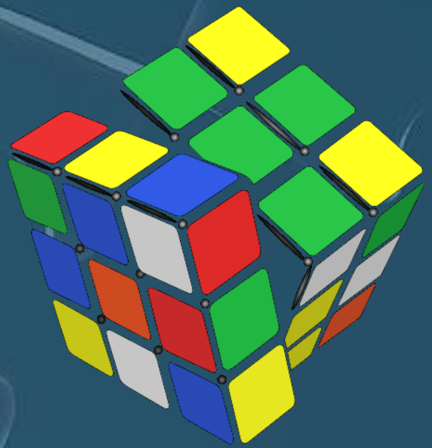
ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

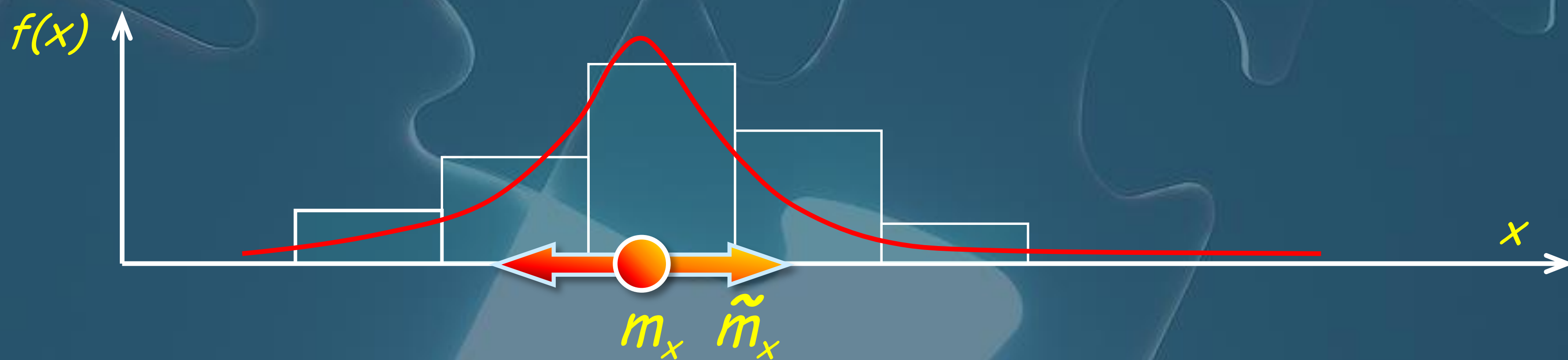


{ интервальные оценки параметров - некоторые распределения СВ связанные с нормальным распределением - доверительный интервал для выборочного среднего при известной дисперсии - доверительный интервал для дисперсии при известном математическом ожидании - доверительный интервал для дисперсии при неизвестном среднем - доверительный интервал для среднего при неизвестной дисперсии }

Интервальные оценки параметров

- Статистика $\tilde{a}_i$, используемая в приближенном равенстве $\tilde{a}_i = a_i$ называется **точечной оценкой неизвестного параметра по выборке**.

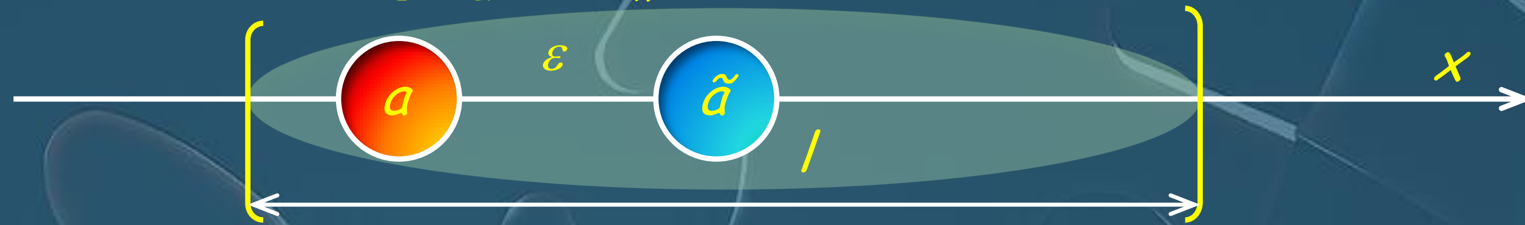
Пример: $\tilde{m}_x = \frac{1}{n}\sum_{i=1}^{n} x_i \Longleftrightarrow \tilde{D}_x = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \tilde{m}_x^2\right)$



Точечные оценки $\tilde{a}_i$ не совпадают (за исключение редких случаев) с истинным значением неизвестных параметров a_i .

Интервальные оценки параметров

- Всегда имеется некоторая погрешность при замене неизвестного параметра его оценкой, т.е. $|\tilde{a}(x_1, x_2, \ldots, x_n) - a| < \varepsilon$:



$$P(|\tilde{a} - a| < \varepsilon) = P(-\varepsilon < \tilde{a} - a < \varepsilon) = P(\tilde{a} - \varepsilon < a < \tilde{a} + \varepsilon)$$

Если эта вероятность близка к единице $P(\tilde{a} - \varepsilon < a < \tilde{a} + \varepsilon) = \gamma$
то диапазон практически возможных значений ошибки, возникающей при замене $\tilde{a}$ на a равен $\pm\varepsilon$.

Чем меньше будет ε , тем точнее оценка $\tilde{a}$.

- Вероятность того, что интервал $(\tilde{a} - \varepsilon; \tilde{a} + \varepsilon)$ со случайными границами накроет неизвестный параметр a , равна $1 - 2\varepsilon = \gamma$.
Эта вероятность называется *доверительной вероятностью.*

Доверительный интервал

- *Доверительным интервалом уровня γ для параметра $\tilde{a}$ выборки $X = (x_1, x_2, \ldots, x_n)$ из генеральной совокупности $F(x, a)$ называется интервал $(\tilde{a}^{(1)}, \tilde{a}^{(2)})$ со случайными границами, такой что:*

$$P_a(\tilde{a}^{(1)}(x) < a < \tilde{a}^{(2)}(x)) \geq \gamma$$

Число γ называется доверительным уровнем интервала. Оно характеризует надежность этого интервала. Увеличивая длину интервала, мы увеличиваем надежность. Но при этом уменьшается точность оценки.

$$L = 2\varepsilon = M(\tilde{a}^{(2)} - \tilde{a}^{(1)})$$

Найти оптимальное решение по всем объектам как правило невозможно. Один из способов: задаться надежностью (обычно это число близкое единице: 0.9; 0.95; 0.99) и затем попытаться найти из всех интервалов уровня γ такой , у которого длина L будет наименьшей, то есть оценка будет наиболее точной.

Распределение Пирсона

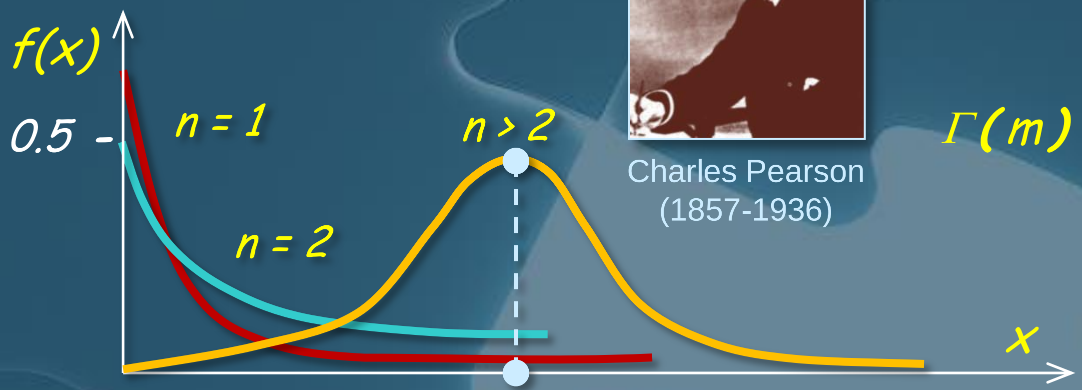
Пусть $\xi_1, \xi_2, \ldots, \xi_n$ есть независимые СВ, имеющие стандартное нормальное распределение. Распределение суммы квадратов этих величин называется распределением Хи-квадрат с n степенями свободы.

- $$\chi_n^2 = \xi_1^2 + \xi_2^2 + \ldots + \xi_n^2$$

Плотность распределения

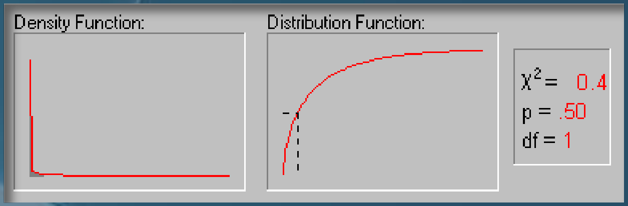
$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

$$\Gamma(m) = \int_0^{\infty} e^{-t} t^{m-1}\,dt$$ *Гамма функция Эйлера*



Max в точке $x = n - 2$

Charles Pearson
(1857-1936)



Квантили распределение Пирсона

$$F_{\chi_n^2}(s) = \int_0^s p_{\chi_n^2}(x)dx$$

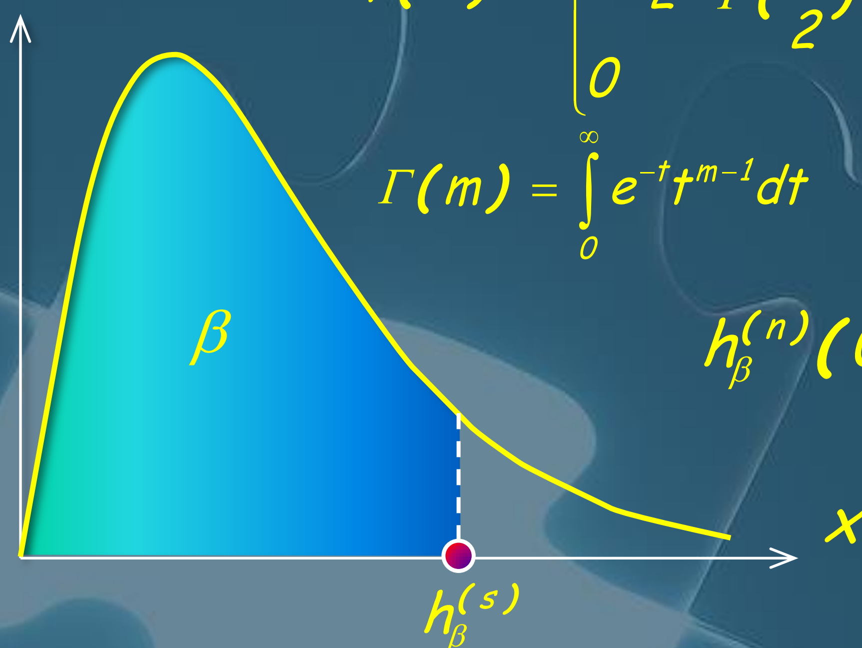
Плотность распределения Пирсона

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$\Gamma(m) = \int_0^\infty e^{-t} t^{m-1} dt$$

Гамма функция Эйлера

$$p_{\chi_n^2}(x)$$



$$F_{\chi_n^2}(h_\beta^{(n)}) = \beta$$

$$h_\beta^{(n)} (0 < \beta < 1)$$

Распределение Стьюдента

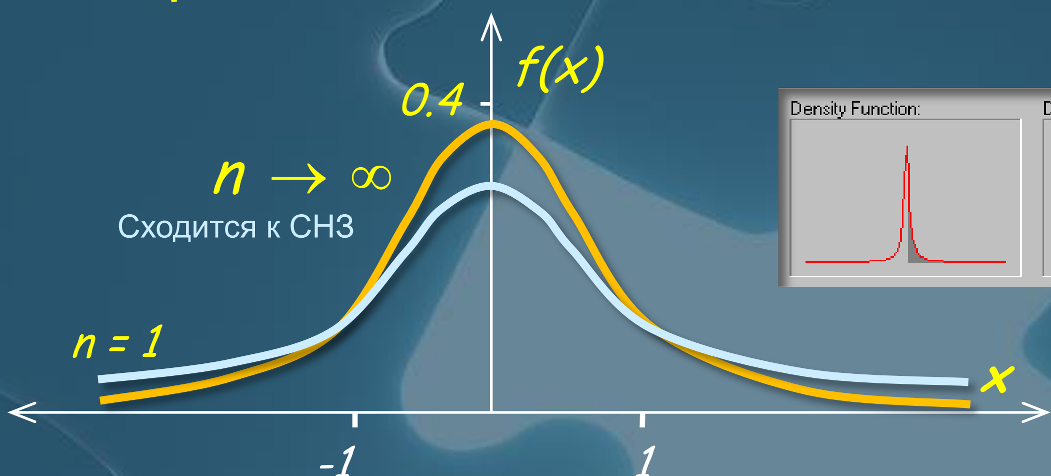
Пусть $\xi_1, \xi_2, \ldots, \xi_n, \xi_0$ есть независимые СВ, со стандартным нормальным распределением. Распределением Стьюдента с n степенями свободы называют распределение

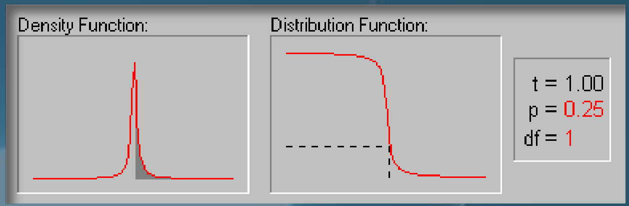
$$t_n = \frac{\xi_0}{\sqrt{\dfrac{\xi_1^2 + \xi_2^2 + \ldots + \xi_n^2}{n}}}$$

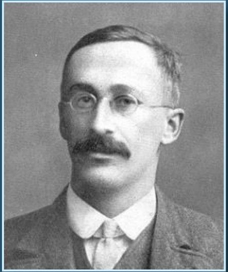
Плотность распределения

$$f_t(x) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

$$\Gamma(m) = \int_0^\infty e^{-t} t^{m-1} dt$$







William Gosset
(1876 – 1937)

Распределение Фишера

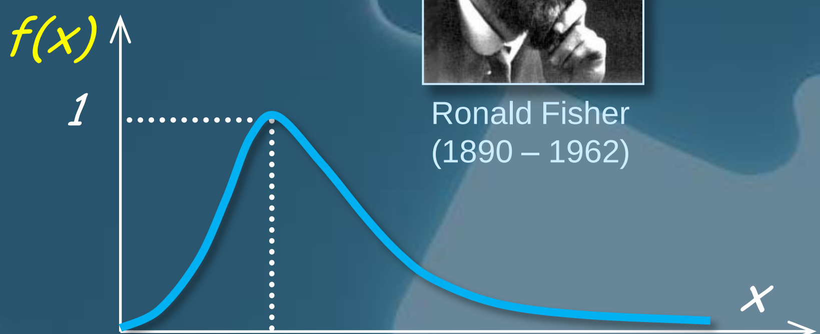
Пусть даны случайные величины $\xi_1, \xi_2, \dots, \xi_n, \xi_{n+1}, \xi_{n+2}, \dots, \xi_{n+m}$ - независимые СВ, имеющие стандартное нормальное распределение и величины, имеющие распределение хи-квадрат с n и m степенями свободы соответственно. *Распределением Фишера со степенями свободы* (n, m) – называется распределение $F_{n,m}$

- $$F_{n,m} = \frac{m\,\chi_n^2}{n\,\chi_m^2}$$

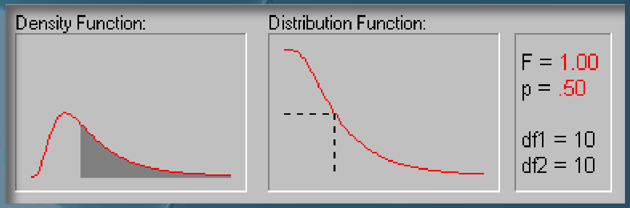
Ronald Fisher
(1890 – 1962)

Плотность распределения

$$f(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}}, \quad x \ge 0$$



$x = (n-2)m/n(m+2)$



Доверительный интервал для среднего при известной дисперсии

Предположим, что параметр m неизвестен, а дисперсия σ^2 известна.

$$G(x_1, x_2, \ldots, x_n, m) = \frac{x_1 + x_2 + \ldots + x_n - nm}{\sigma\sqrt{n}} \qquad F_G(t) = \Phi(t)$$

$$u_\varepsilon \to \frac{1}{\sqrt{2\pi}} \int_{u_\varepsilon}^{\infty} e^{-\frac{t^2}{2}} dt = \varepsilon \iff \Phi(\infty) - \Phi(u_\varepsilon) = \varepsilon \iff \frac{1}{2} - \Phi(u_\varepsilon) = \varepsilon$$

$$P(|\tilde{m} - m| < \varepsilon) = 2\Phi\left(\frac{\varepsilon_\gamma}{\sigma_{\tilde{m}}}\right) - 1 \qquad \sigma_{\tilde{m}} = \sqrt{\frac{D}{n}}$$

$$2\Phi\left(\frac{\varepsilon_\gamma}{\sigma_{\tilde{m}}}\right) - 1 = \gamma \qquad \varepsilon_\gamma = \sigma_{\tilde{m}} \arg\Phi\left(\frac{1+\gamma}{2}\right) \qquad I_\gamma = (\tilde{m} - \varepsilon_\gamma; \tilde{m} + \varepsilon_\gamma)$$

Используется таблица $t_\gamma = \arg\Phi\left(\frac{1+\gamma}{2}\right) \qquad I_\gamma = (\tilde{m} - t_\gamma\sigma_{\tilde{m}}; \tilde{m} + t_\gamma\sigma_{\tilde{m}})$

$$\tilde{m} - \frac{\sigma}{\sqrt{n}} t_\gamma < m < \tilde{m} + \frac{\sigma}{\sqrt{n}} t_\gamma$$

Доверительный интервал для дисперсии при известном математическом ожидании

$$G(x_1, x_2, .., x_n, \sigma^2) = \frac{\sum_{i=1}^{n}(x_i - m)^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{x_i - m}{\sigma}\right)^2$$

Функция $G(x_{норм.})$ имеет хи-квадрат распределение с n – степенями свободы, не зависящее от неизвестного параметра σ^2

Обозначая $h^{(n)}_{\varepsilon}$- квантили этого распределения и фиксируя $\varepsilon_1 + \varepsilon_2 = 1 - \gamma$ приходим к неравенству, выполняемому с вероятностью γ

$$h^{(n)}_{\varepsilon_1} < \frac{\sum_{i=1}^{n}(x_i - m)^2}{\sigma^2} < h^{(n)}_{1-\varepsilon_2}$$

$$\frac{\sum_{i=1}^{n}(x_i - m)^2}{h^{(n)}_{1-\varepsilon_2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(x_i - m)^2}{h^{(n)}_{\varepsilon_1}}$$

Доверительный интервал для дисперсии при неизвестном среднем

Оба параметра неизвестны. m – мешающий параметр. Функция $G\,(x_{норм.})$ имеет хи-квадрат распределение с $(n-1)$ – степенями свободы, не зависящее от неизвестного параметра σ^2

$$G(x_1, x_2, \ldots, x_n; m, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \tilde{m})^2$$

Обозначая $h^{(n-1)}_{\varepsilon}$- квантили этого распределения и фиксируя $\varepsilon_1 + \varepsilon_2 = 1 - \gamma$ приходим к неравенству, выполняемому с вероятностью γ

$$\frac{\sum_{i=1}^{n}(x_i - \tilde{m})^2}{h^{(n-1)}_{1-\varepsilon_2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(x_i - \tilde{m})^2}{h^{(n-1)}_{\varepsilon_1}}$$

$$\frac{(n-1)s^2}{h^{(n-1)}_{1-\varepsilon_2}} < \sigma^2 < \frac{(n-1)s^2}{h^{(n-1)}_{\varepsilon_1}}$$

Доверительный интервал для среднего при неизвестной дисперсии

Выбирается функция G имеющая распределение Стьюдента с $(n\text{-}1)$ – степенями свободы.

$$G(x_1,x_2,\ldots,x_n;\, m,\sigma^2) = \frac{\sqrt{n}(\tilde{m}-m)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\tilde{m})^2}} = \frac{\sqrt{n}(\tilde{m}-m)}{s}$$

Обозначая $t^{(n)}_{\varepsilon}$- квантили этого распределения и фиксируя $\varepsilon_1+\varepsilon_2=1-\gamma$ приходим к неравенству, выполняемому с вероятностью γ

$$\tilde{m}-\frac{s}{\sqrt{n}}t^{(n-1)}_{(1+\gamma)/2} < m < \tilde{m}+\frac{s}{\sqrt{n}}t^{(n-1)}_{(1+\gamma)/2}$$

Пример

@ Найти доверительный интервал для среднего значения генеральной совокупности при больших объемах выборки $(n > 30)$

Границы интервалов	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44	44 – 46	
Частоты m_j	2	3	30	40	20	5	$n = 100$

По выборочным данным находим выборочные среднее арифметическое для m и стандартное отклонение S

Это точечные оценки

$$\tilde{m}_x = \frac{1}{100}(35 \cdot 2 + 37 \cdot 3 + 39 \cdot 30 + 41 \cdot 40 + 43 \cdot 20 + 45 \cdot 5) = 40.76$$

$$s = \sqrt{\tilde{D}_x} = \sqrt{\frac{100}{99}\left(\frac{1}{100}(35^2 \cdot 2 + 37^2 \cdot 3 + 39^2 \cdot 30 + 41^2 \cdot 40 + 43^2 \cdot 20 + 45^2 \cdot 5) - (40.76)^2\right)} = 2$$

Пример

@

Задаемся доверительной вероятностью : $\gamma = 0{,}95$.

Находим значение t_φ , соответствующее заданной доверительной вероятности $t_{0,05} = 1{,}96$.

$$\tilde{m}_x - 1.96\frac{s}{\sqrt{n}} \le m \le \tilde{m}_x + 1.96\frac{s}{\sqrt{n}}$$

$$40.368 \le m \le 41.152$$

Пример

@ Для контроля качества в 40 пробах стали GS50 определялось содержание углерода x (%C) и прочность на разрыв z (Н/мм). Данные оформлены в виде таблицы чисел:

$$x_i : x_1, x_1, x_1, \dots\dots\dots\dots\dots\dots\dots\dots\dots . \, x_{40}$$

$$z_i : z_1, z_1, z_1, \dots\dots\dots\dots\dots\dots\dots\dots\dots . \, z_{40}$$

X: *0.3, 0.33, 0.37, 0.36, 0.31, 0.29, 0.34, 0.39, 0.37, 0.38, 0.35, 0.32, 0.39, 0.3, 0.32, 0.32, 0.38, 0.37, 0.38, 0.33, 0.37, 0.33, 0.34, 0.33, 0.3, 0.34, 0.36, 0.33, 0.34, 0.36, 0.29, 0.3, 0.33, 0.32, 0.32, 0.38, 0.37, 0.34, 0.35, 0.36*

$X = X(x_1, x_2, \dots, x_{40})$ – *выборка объемом* $n = 40$

Z: *589, 614, 612, 572, 548, 537, 574, 570, 540, 575, 535, 593, 582, 538, 566, 562, 601, 587, 587, 614, 602, 544, 545, 562, 576, 596, 605, 575, 570, 550, 572, 555, 555, 518, 539, 557, 558, 587, 580, 560*

$Z = Z(z_1, z_2, \dots, z_{40})$ – *выборка объемом* $n = 40$

Пример

@ Найти доверительные интервалы для m_x и m_z, теоретических значений содержания углерода и прочности на разрыв стали GS50.

Напомним, что объем каждой из выборок : $n = 40$.

Зафиксируем доверительную вероятность, близкую к единице : $\gamma = 0.95$.

По таблице распределения Стьюдента определим приближенно: $t_{0.975}^{(39)} \approx 2.02$

$$\tilde{m}_x \pm \frac{s_x}{\sqrt{40}} t_{0.975}^{(39)} = 0.3415 \pm \frac{0.289}{6.32} \cdot 2.02 \approx 0.3415 \pm 0.0092$$

$$\tilde{m}_z \pm \frac{s_z}{\sqrt{40}} t_{0.975}^{(39)} = 570.05 \pm \frac{24.14}{6.32} \cdot 2.02 \approx 570.05 \pm 7.71$$

$$0.332 < m_x < 0.351$$

$$562.2 < m_z < 577.8$$