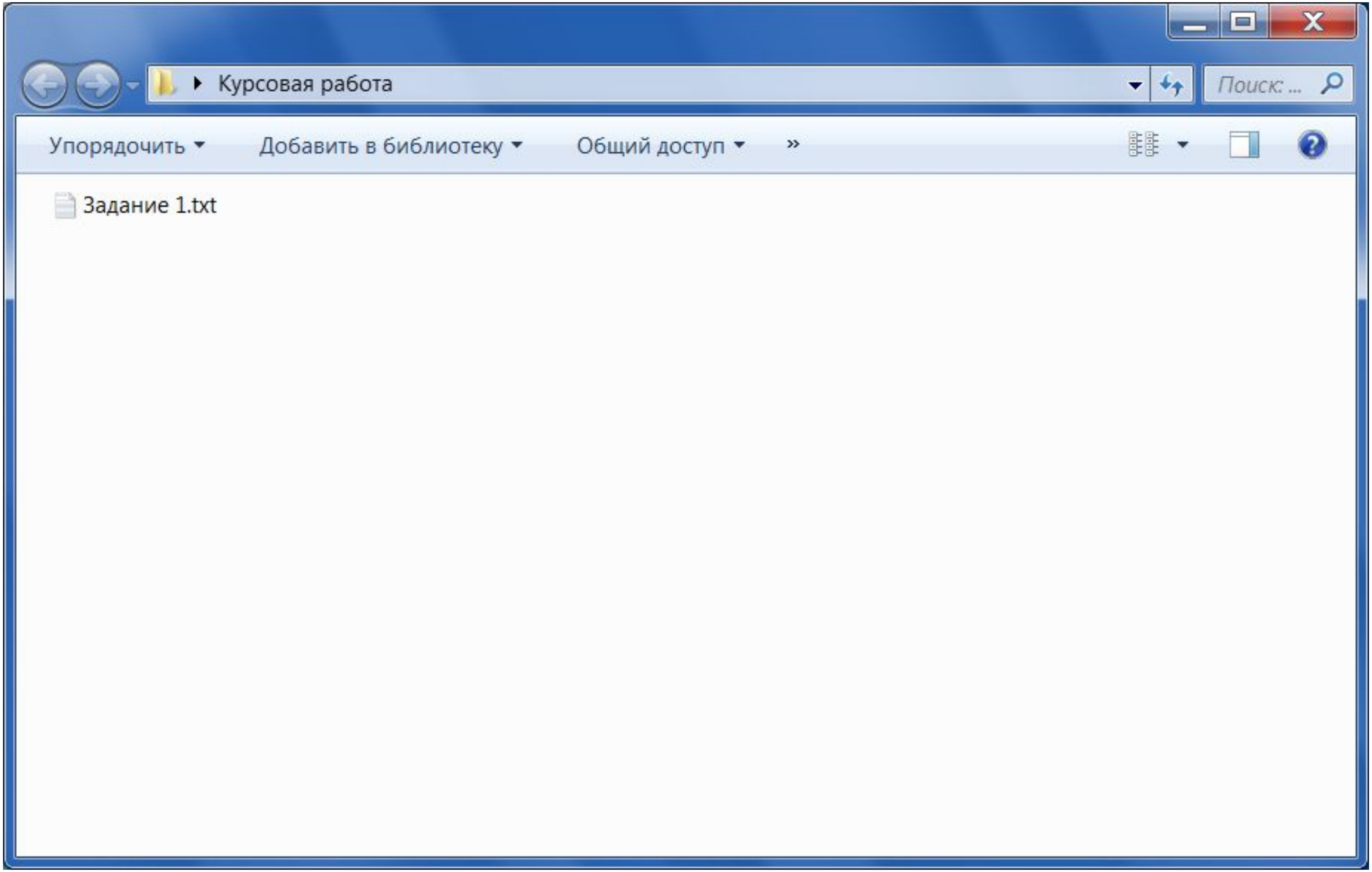
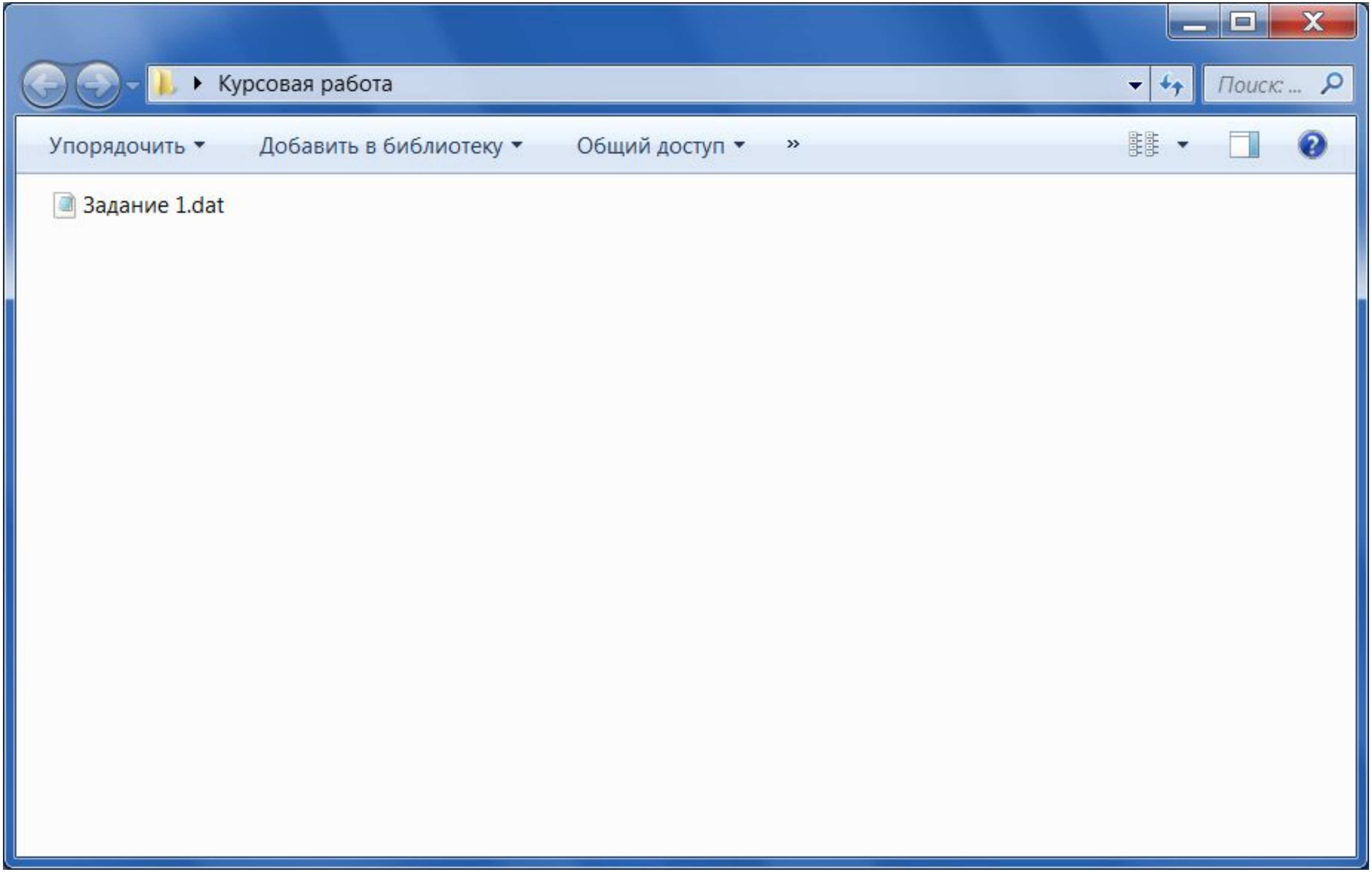
Тема 7
РЕШЕНИЕ ЗАДАЧ ОПИСАТЕЛЬНОЙ СТАТИСТИКИ В MATHCAD

Доцент кафедры ТАМ, к.т.н. Шипулин Леонид Викторович

Ввод исходных данных для расчета



Ввод исходных данных для расчета



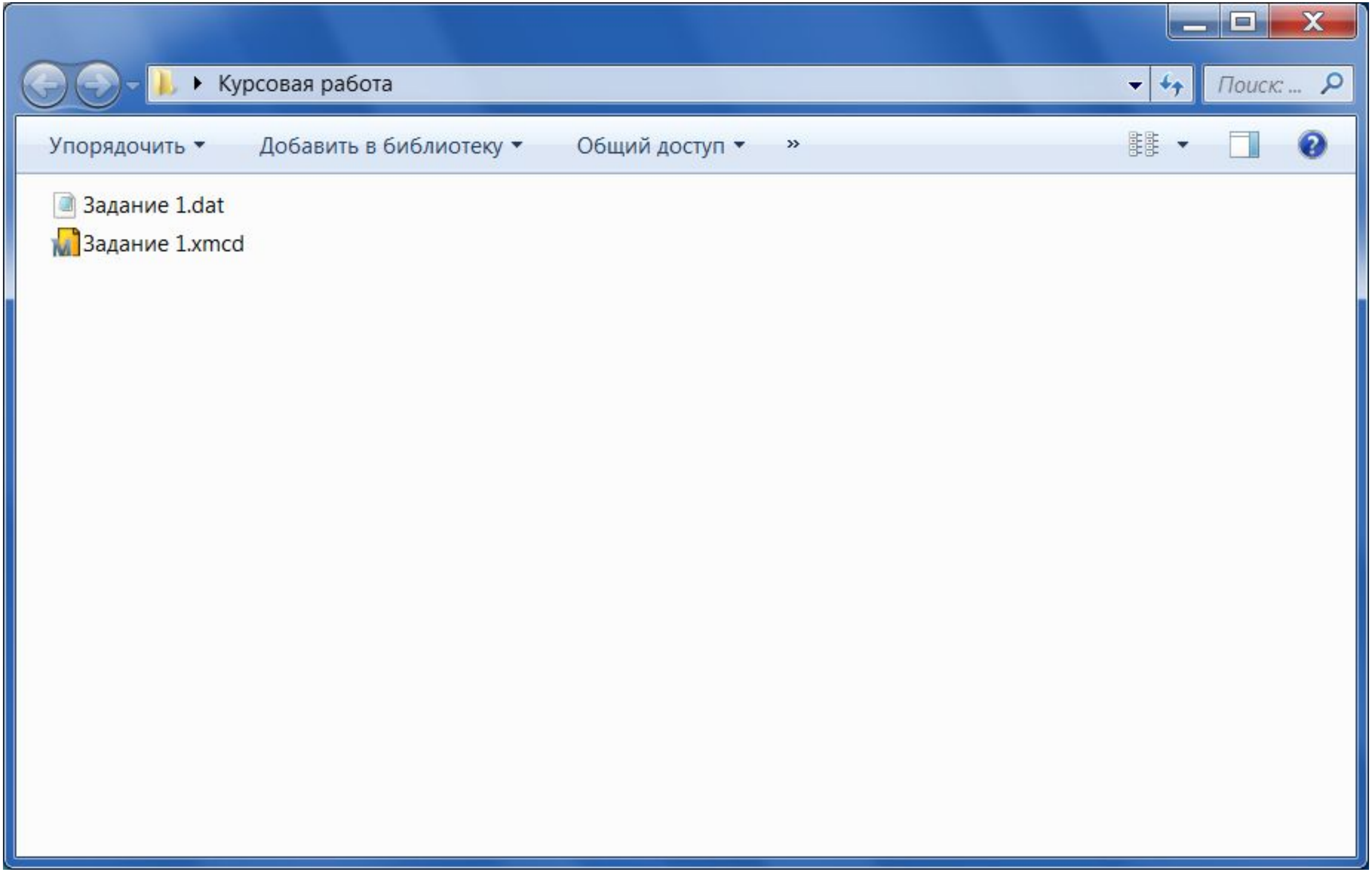
Ввод исходных данных для расчета

Задание 1.dat — Блокнот

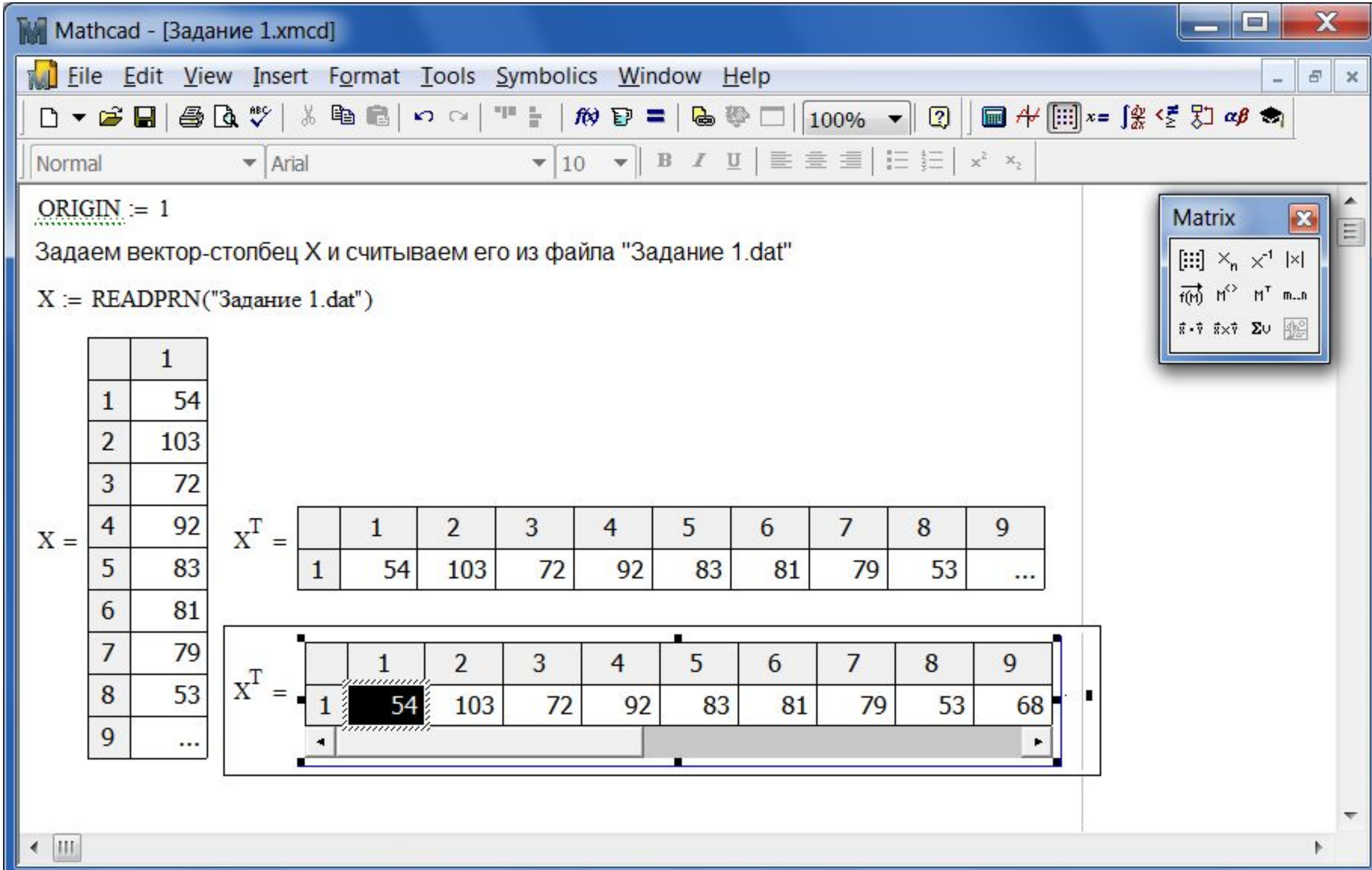
Файл Правка Формат Вид Справка

54
103
72
92
83
81
79
53
68
82
94
65
97
110
78
82
63
101
68
87

Ввод исходных данных для расчета



Считывание данных в вектор



ORIGIN := 1

Задаем вектор-столбец X и считываем его из файла "Задание 1.dat"

X := READPRN("Задание 1.dat")

X =

	1
1	54
2	103
3	72
4	92
5	83
6	81
7	79
8	53
9	...

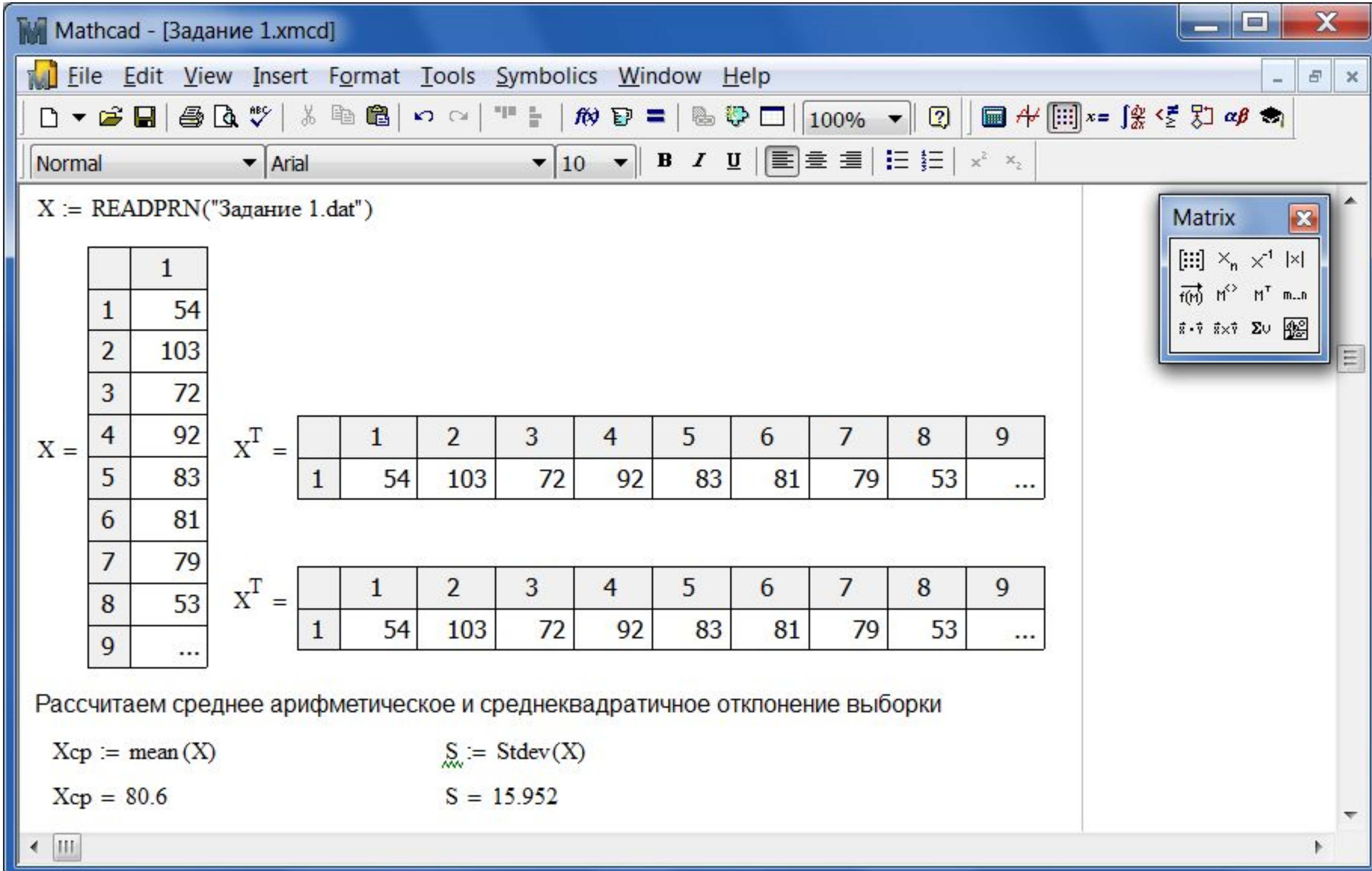
X^T =

	1	2	3	4	5	6	7	8	9
1	54	103	72	92	83	81	79	53	...

X^T =

	1	2	3	4	5	6	7	8	9
1	54	103	72	92	83	81	79	53	68

Среднее арифметическое. Среднеквадратичное отклонение выборки



$X =$

	1
1	54
2	103
3	72
4	92
5	83
6	81
7	79
8	53
9	...

$X^T =$

	1	2	3	4	5	6	7	8	9
1	54	103	72	92	83	81	79	53	...

$X^T =$

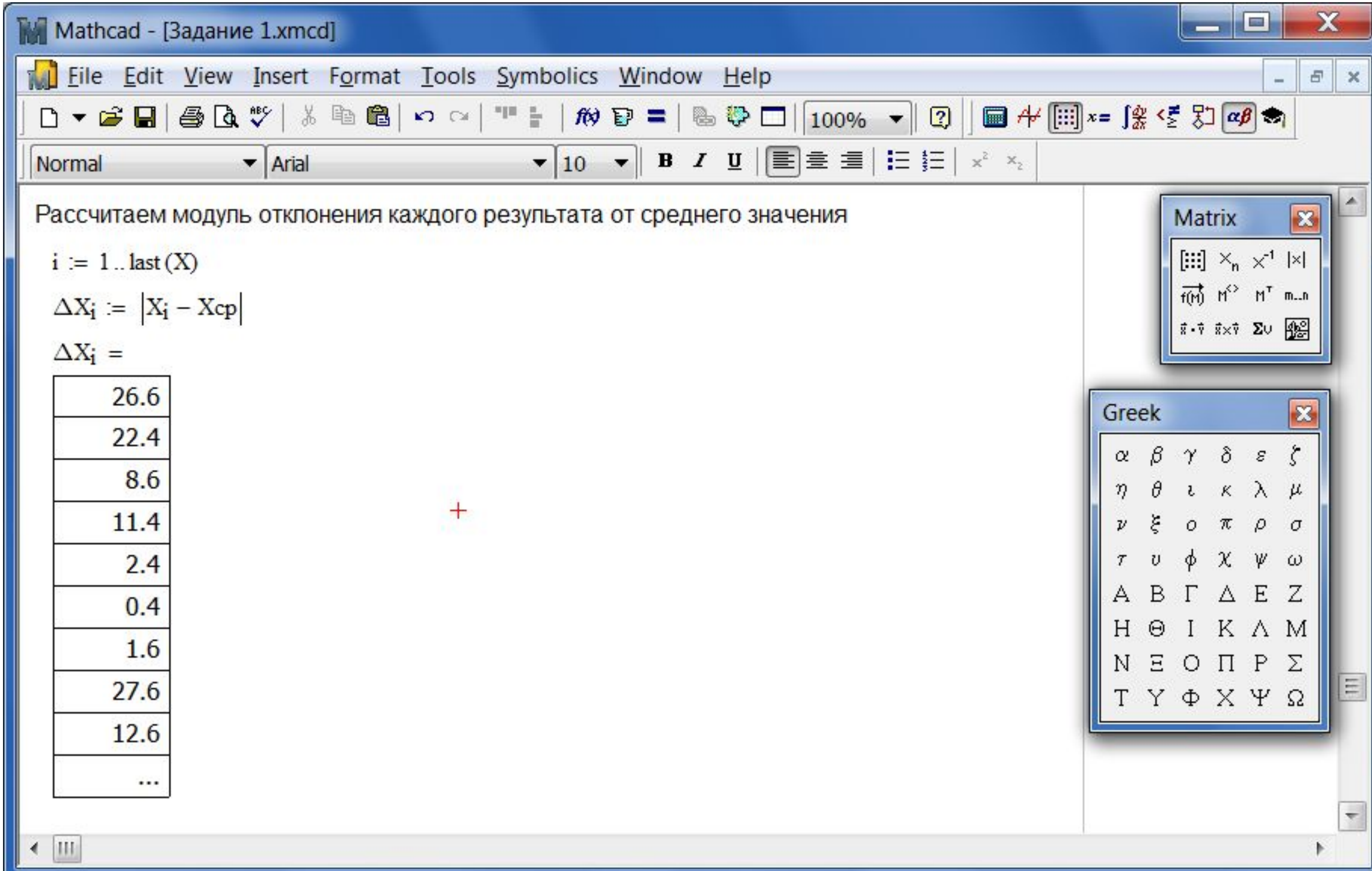
	1	2	3	4	5	6	7	8	9
1	54	103	72	92	83	81	79	53	...

Рассчитаем среднее арифметическое и среднеквадратичное отклонение выборки

$Xcp := mean(X)$ $S := Stdev(X)$

$Xcp = 80.6$ $S = 15.952$

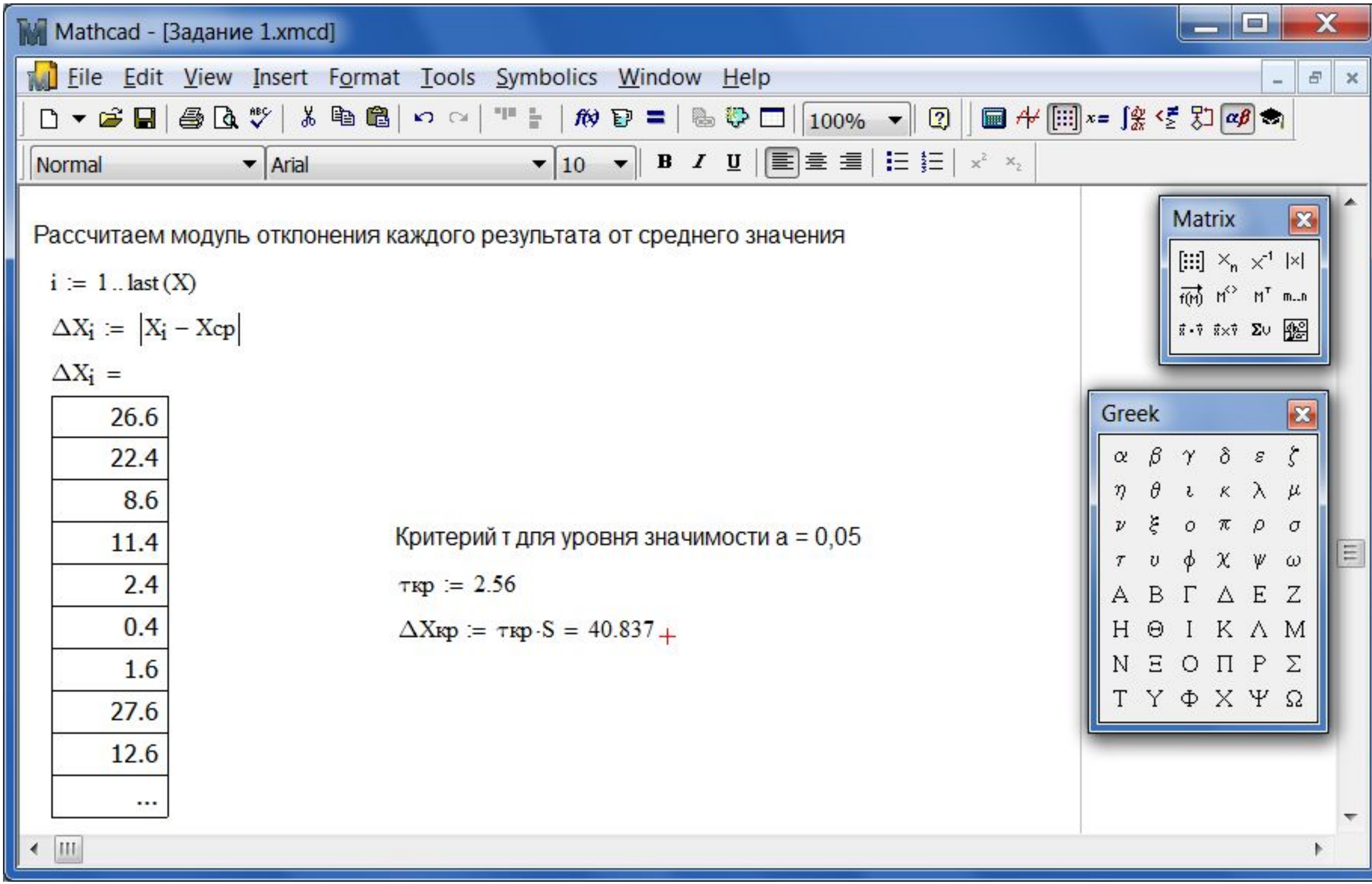
Отклонение результатов от среднего значения



Выбор значения τ-критерия (для α=0,05)

Количество испытаний n	$\tau_{крит}$	Количество испытаний n	$\tau_{крит}$	Количество испытаний n	$\tau_{крит}$
3	1.15	11	2.23	19	2.53
4	1.46	12	2.29	20	2.56
5	1.67	13	2.33	21	2.58
6	1.82	14	2.37	22	2.60
7	1.94	15	2.41	23	2.62
8	2.03	16	2.44	24	2.64
9	2.11	17	2.48	25	2.66
10	2.18	18	2.50		

τ-критерий и критическое отклонение величины



Проверка на критическое отклонение случайной величины от среднеарифметического

Просматриваем вектор ΔX_i. Если некоторое значение превышает критическое значение $\Delta X_{кр}$, то это значение из первоначально заданной выборки следует выкинуть (в текстовом файле .dat) и пересчитать все заново, сменив значение τ-критерия для уровня значимости α = 0,05.

В выборке, приведенной в примере, каждое значение вектора ΔX_i меньше, чем критическое $\Delta X_{кр}$. Поэтому никакие значения выкидывать не будем. Иными словами, **резко выделяющиеся значения** отсутствуют.

Можно приступать к дальнейшему исследованию выборки.

$\Delta X_i =$

26.6
22.4
8.6
11.4
2.4
0.4
1.6
27.6
12.6
1.4
13.4
15.6
16.4
29.4
2.6
1.4
17.6
20.4
12.6
6.4

Параметры выборки

Выборочное среднее значение $\mathrm{mean}(X) = 80.6$

Смещенная дисперсия $\mathrm{var}(X) = 241.74$

Несмещенная дисперсия $\mathrm{Var}(X) = 254.463$

Стандартное отклонение на основе смещенной дисперсии $\mathrm{stdev}(X) = 15.548$

Стандартное отклонение на основе несмещенной дисперсии $\mathrm{Stdev}(X) = 15.952$

Коэффициент асимметрии $\mathrm{kurt}(X) = -0.653$

Коэффициент эксцесса $\mathrm{skew}(X) = -4.46 \times 10^{-4}$

Построение гистограммы

Количество интервалов $N := 8$

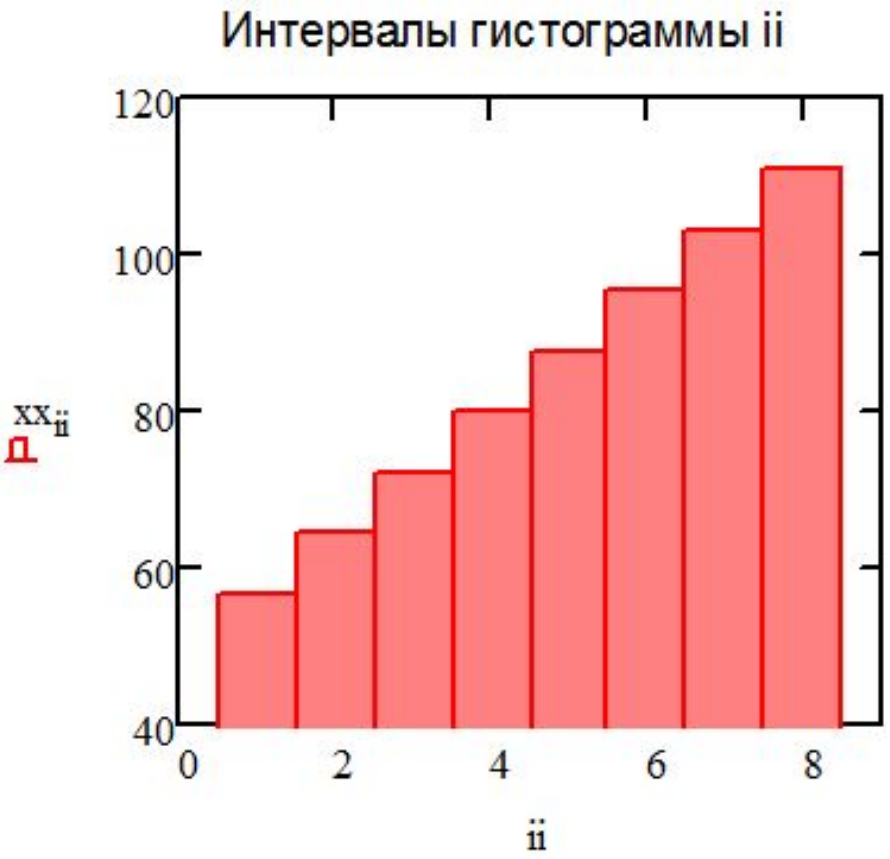
Автоматическое построение гистограммы

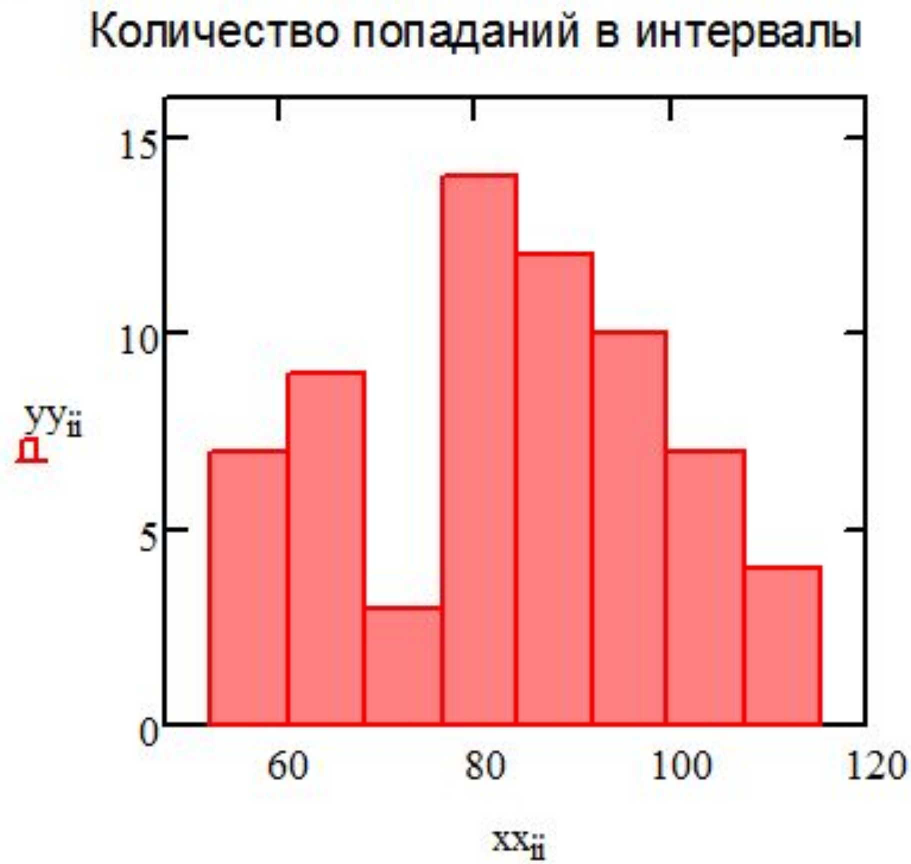
$$H := \mathrm{histogram}(N, X)$$

$$H = \begin{pmatrix} 56.562 & 2 \\ 63.687 & 2 \\ 70.813 & 3 \\ 77.938 & 3 \\ 85.063 & 4 \\ 92.188 & 2 \\ 99.313 & 2 \\ 106.438 & 2 \end{pmatrix}$$

Построение гистограммы

$ii := 1..N \qquad xx := H^{\langle 1 \rangle} \qquad yy := H^{\langle 2 \rangle}$

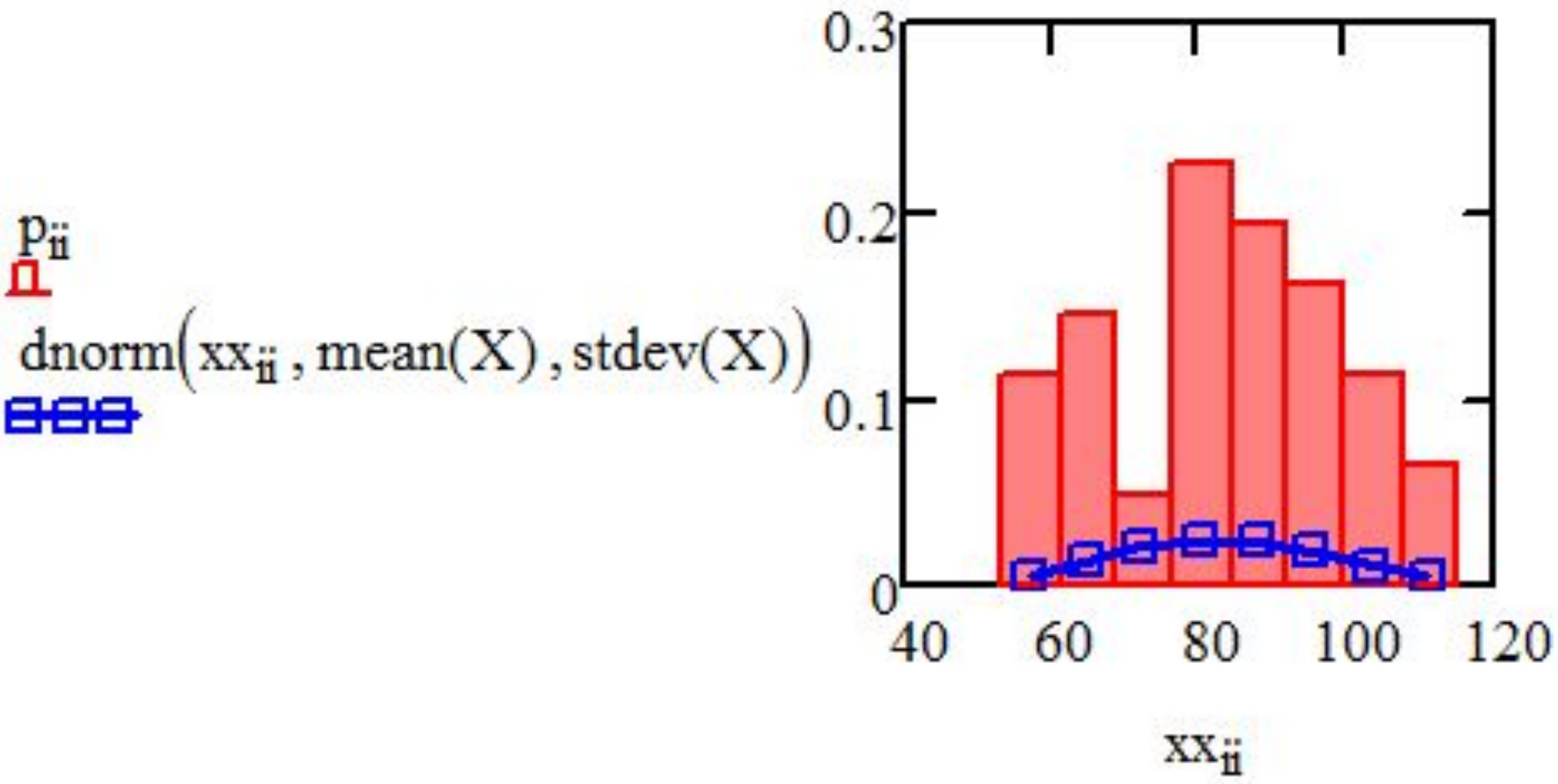




Построение гистограммы

$$p := \frac{yy}{N \cdot \left(xx_2 - xx_1\right)}$$

нормированные высоты столбцов



Гипотеза о нормальном распределении

Выдвигается гипотеза о том, что теоретический закон распределения случайной величины имеет вид нормального распределения. Воспользуемся критерием Пирсена для того, чтобы проверить это.

Значения середин интервалов:

$$x := H^{\langle 1 \rangle}$$

Значения функции распределения:

$$\phi_{ii} := \frac{1}{stdev(X)\sqrt{2\cdot\pi}}\cdot e^{\frac{-\left(x_{ii}-mean(X)\right)^2}{stdev(X)^2\cdot 2}}$$

Шаг между серединами интервалов:

$h21 := x_2 - x_1 = 7.75$ $\quad$ $h32 := x_3 - x_2 = 7.75$

$$h := h21 = 7.75$$

Гипотеза о нормальном распределении

Значения теоретических частот:

$$n0_{ii} := \mathrm{last}(X) \cdot h \cdot \phi_{ii}$$

Значения фактических частот:

$$n := H^{\langle 2 \rangle}$$

Вычисление слагаемых для критерия χ^2:

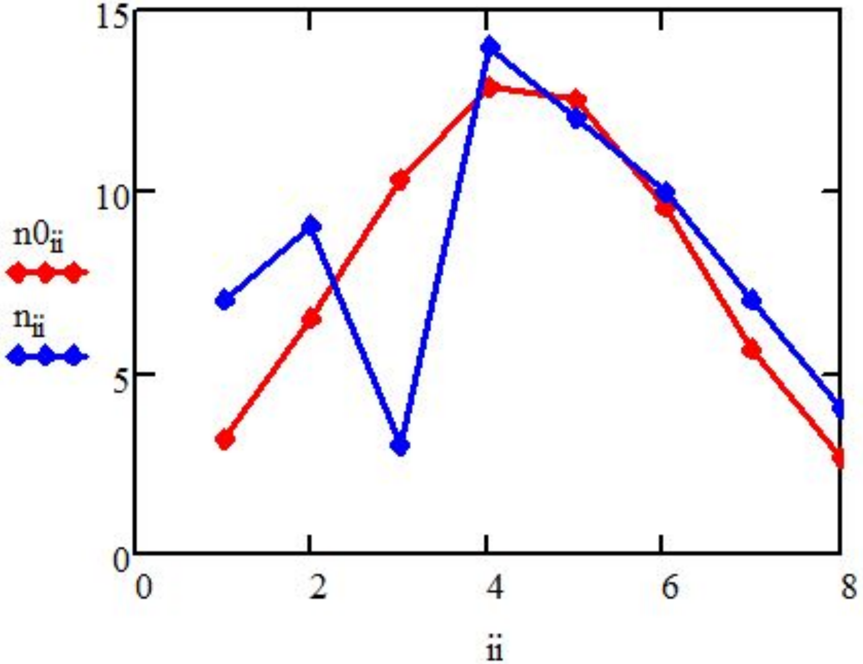
$$\chi_{ii} := \frac{\left(n_{ii} - n0_{ii}\right)^2}{n0_{ii}}$$

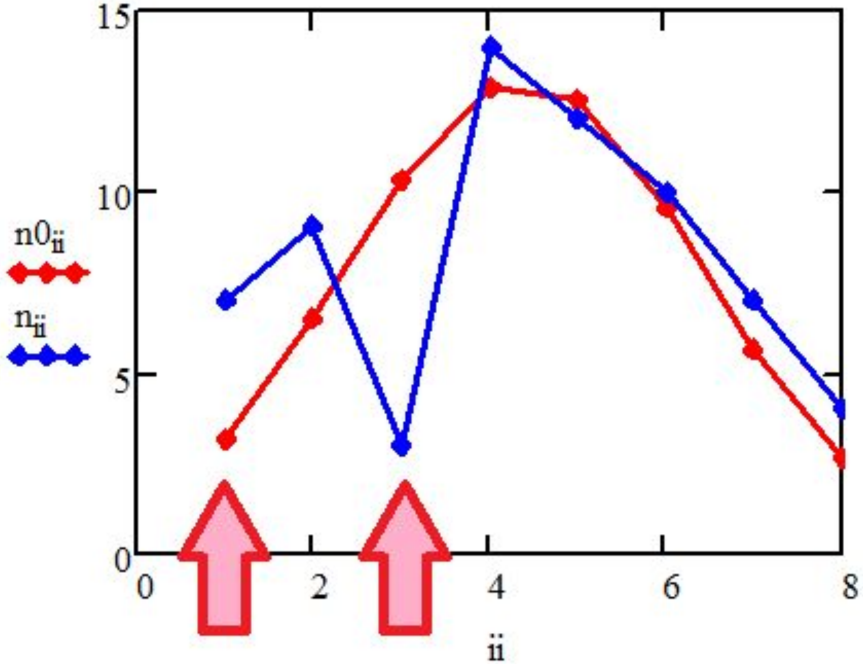
Гипотеза о нормальном распределении

Вывод результатов:

$x_{ii} =$	$\phi_{ii} =$	$n0_{ii} =$	$n_{ii} =$	$\chi_{ii} =$
56.875	$6.131 \cdot 10^{-3}$	3.136	7	4.762
64.625	0.013	6.443	9	1.015
72.375	0.02	10.32	3	5.192
80.125	0.025	12.888	14	0.096
87.875	0.025	12.548	12	0.024
95.625	0.019	9.524	10	0.024
103.375	0.011	5.636	7	0.33
111.125	$5.083 \cdot 10^{-3}$	2.6	4	0.754

Гипотеза о нормальном распределении





Гипотеза о нормальном распределении

Расчет критерия χ^2:

$$\chi 2n := \sum_{k=1}^{N} \chi_k = 12.197$$

Табличное значение χ^2 для количества степеней свободы К = 8 - 1 = 7 и для уровня значимости $\alpha = 0.05$:

$$\chi 2t := 14.1$$

Поскольку $\chi^2 n = 12.2$ меньше, чем $\chi^2 n$, то нулевая гипотеза о нормальном распределении принимается для данного уровня значимости.

Доверительный интервал генеральной выборки

Построение доверительного интервала для математического ожидания генеральной совокупности

Критерий Стьюдента для количества степеней свободы К = 7 и вероятности ошибки $\alpha = 0.05$:

$$t := 2.37$$

Максимальное отклонение от среднего в доверительном интервале:

$$\varepsilon := \frac{t \cdot \mathrm{stdev}(X)}{\sqrt{N}} = 13.015$$

Границы доверительного коридора:

$$\left(\mathrm{mean}(X) - \varepsilon \quad \mathrm{mean}(X) + \varepsilon \right) = \left(70.152 \quad 96.182 \right)$$

С вероятностью 0.95 можно утверждать, что среднее значение при выборке большего объема не выйдет за пределы интервала (70.152   96.182).