

В результате эксперимента получены данные, записанные в виде таблицы:

56,5	47,3	23,1	38,6	92,5	50,9	74,9	65,7	47,5	83,9
11,8	70,1	57,1	39,9	54,7	70,9	47,4	28,1	39,1	76,2
32,3	92,1	20,7	48,6	87,1	66,3	45,8	41,4	56,9	22,6
45,8	58,4	53,4	51,4	11,6	30,9	31,4	37,4	65,8	19,3
45,3	74,4	21,2	25,7	56,7	20,3	48,3	60,1	46,2	64,1
15,1	47,7	12,7	92,6	29,5	52,0	60,2	32,1	74,5	54,2
36,1	47,2	26,1	65,3	42,0	50,1	72,1	56,4	25,1	75,1
83,8	38,7	81,2	65,1	87,4	35,3	92,4	85,6	83,5	20,5
76,3	69,4	41,6	35,9	29,7	80,9	49,9	59,5	83,4	76,5
24,4	55,9	74,2	27,3	76,7	29,9	69,1	30,1	65,4	18,4

Требуется:

1. Представить выборку в виде статистического ряда. Построить гистограмму и полигон частот.
2. Определить моду, медиану.
3. Определить несмещенные точечные оценки для математического ожидания, дисперсии, среднеквадратического отклонения.
4. Найти доверительный интервал для математического ожидания.
5. Приняв в качестве нулевой гипотезу H_0 : генеральная совокупность, из которой извлечена выборка, имеет нормальное распределение, проверить ее, пользуясь критерием Пирсона при уровне значимости $\alpha = 0,025$.

Для обработки данных в среде MathCad их нужно представить в виде вектора-столбца. Это можно сделать, по крайней мере двумя способами.
1-й способ - ввести вручную:

$$i := 0..99 \quad X_i :=$$

56.5
47.3

и так далее, через запятую, все 100 штук. Помните, что разделитель разрядов точка, а не запятая.

2 -й способ -- сохранить данные в виде текстового файла -- один столбец.

Как это сделать.

1. В ворде преобразовать таблицу в текст
2. Заменить знаки табуляции на конец строки (абзац)
3. Заменить запятую на точку.
4. Сохранить в виде текстового файла (у меня 56.prn)

Такие файлы MathCad читает и воспринимает как данные.

Сохраним MathCad`овский документ в папке с данными и вводим данные. Смотрите ниже.

$X := \text{READPRN}("56.prn")$ Вводим данные

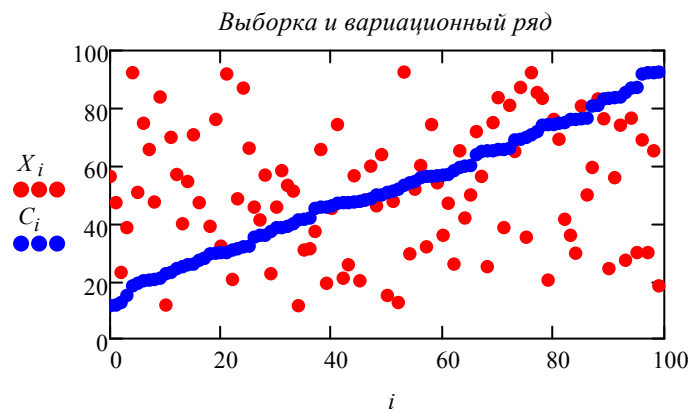
$N := \text{rows}(X) = 100$ Определяем количество элементов выборки

$i := 0..N - 1$ $C := \text{sort}(X)$ Задаём переменную для индексации массива и сортируем данные: получаем вариационный ряд

$$X^T =$$

	0	1	2	3	4	5	6	7	8	9
0	56.5	47.3	23.1	38.6	92.5	50.9	74.9	65.7	47.5	...

Ради интереса выводим данные в виде матрицы-строки



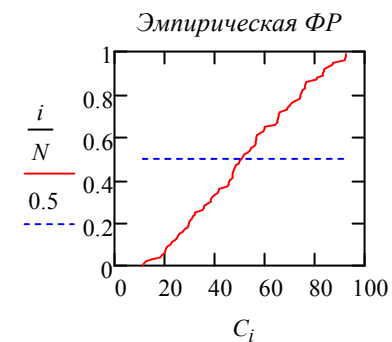
Медиану ищем как средний член вариационного ряда

$$Med := \frac{C_{49} + C_{50}}{2} = 50.5$$

Или из условия $F(Med) = 0.5$

$$Med := median(X) = 50.5$$

Мода $mode(X) = 45.8$



Выборочное среднее => $X_o := \frac{1}{N} \cdot \left(\sum_i X_i \right)$ $X_o = 51.718$ $mean(X) = 51.718$

Выборочная дисперсия => $D := \frac{1}{N-1} \cdot \left[\sum_i (X_i - X_o)^2 \right]$ $D = 484.934$ $var(X) = 480.085$

Исправленная дисперсия=> $D_o := \frac{N}{N-1} \cdot (D)$ $D_o = 489.833$ $Var(X) = 484.934$

Доверительный интервал для мат. ожидания на уровне довери $\gamma := 0.95$

Статистика $\frac{X_o - Mx}{\sqrt{D_o}} \cdot \sqrt{N}$ имеет распределение Стьюдента с $N - 1$ степенью свободы. Квантиль распределения Стьюдента на уровне $\frac{1 + \gamma}{2}$ равен $q := qt\left(\frac{1 + \gamma}{2}, N - 1\right) = 1.984$

Доверительные границы $\Delta := \sqrt{\frac{D_o}{N}} \cdot q = 4.392$ $X_L := X_o - \Delta = 47.326$ $X_R := X_o + \Delta = 56.11$

Для получения статистического ряда, гистограммы и полигона необходимо сгруппировать данные. Число интервалов группировки выбирается из условия: число точек, попавших в каждый интервал должно быть не меньше пяти.
При объёме выборки не более 200 можно пользоваться формулой Старджеса

$$\underline{K} := \text{ceil}(\log(N, 2)) \quad K = 7 \quad \leq \quad \text{Число интервалов группировки}$$

$$\underline{h} := \frac{\text{ceil}(\max(X)) - \text{trunc}(\min(X))}{K} \quad h = 11.714 \quad \leq \quad \text{ширина интервала}$$

Проиндексируем и зададим концы интервалов группировки. Для вывода на экран координат границ интервалов транспонируем соответствующий столбец с помощью команды Ctrl+1.

$$\underline{j} := 0..K \quad \underline{d}_j := \text{trunc}(\min(X)) + j \cdot h \quad \underline{d}^T = (11 \quad 22.714 \quad 34.429 \quad 46.143 \quad 57.857 \quad 69.571 \quad 81.286 \quad 93)$$

Для подсчёта "чисел заполнения" интервалов используем системную функцию *hist*

$$\underline{H} := \text{hist}(\underline{d}, X) \quad \underline{H}^T = (11 \quad 15 \quad 14 \quad 22 \quad 13 \quad 14 \quad 11) \quad \leq \quad \text{получаем наблюдаемый массив чисел заполнения}$$

$$\underline{l} := 0..K - 1 \quad \leq \quad \text{индекс для построения гистограммы и полигона} \quad \sum H = 100 \quad \text{Проверка}$$

Нормируем массив чисел, приводя его размерность к размерности плотности

$$\underline{HIST} := \frac{\underline{H}}{N \cdot h}$$

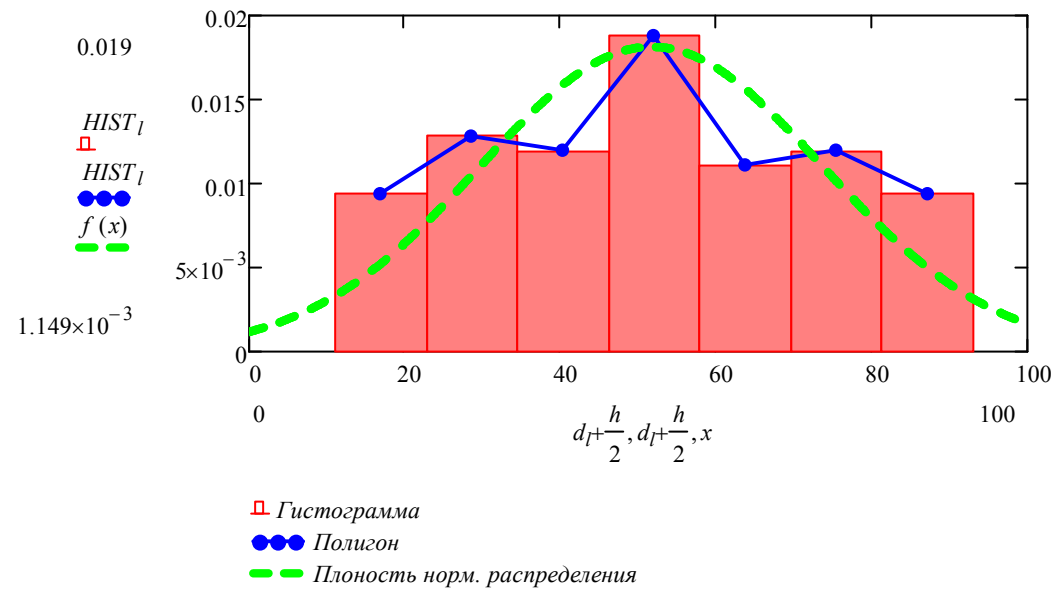
Рассчитаем числовые характеристики распределения

$$\underline{a} := \text{mean}(X) \quad a = 51.718 \quad \underline{s} := \text{Stdev}(X) \quad s = 22.021$$

Предполагаем, что это параметры нормального распределения

$$\underline{x} := 0, 0.1 .. 100$$

$$f(x) := \frac{1}{s \cdot \sqrt{2 \cdot \pi}} \cdot \exp \left[-\frac{(x - a)^2}{2 \cdot s^2} \right]$$



Проверка нормальности распределения с помощью критерия Пирсона

Нулевая гипотеза H_0 : выборка описывается нормальным распределением

Подсчитываем теоретические (ожидаемые) числа заполнения интервалов группировки

$F(x) := \text{pnorm}(x, a, s)$ <= теоретическая функция распределения

$d_0 := -\infty$ Бесконечность => Ctrl+Shift+z $d_K := \infty$ Полагаем границы области значений бесконечными

$n_l := (F(d_{l+1}) - F(d_l)) \cdot N$ <= Теоретические числа заполнения

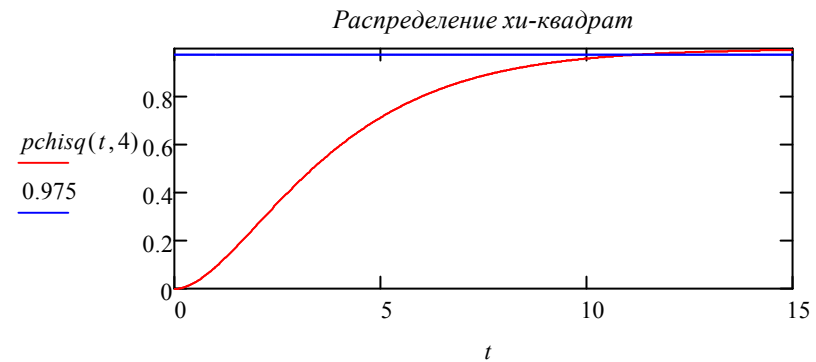
$$n^T = (9.391 \ 12.228 \ 18.388 \ 20.973 \ 18.145 \ 11.907 \ 8.969)$$

Подсчитаем значение статистики критерия хи-квадрат

$$\chi := \sum_l \frac{(H_l - n_l)^2}{n_l} \quad \chi = 4.288$$

Критическая точка:
квантиль хи-квадрат распределения
на уровне 0.975 при числе степеней свободы
 $K - 3 = 4$

$$qChi := qchisq(0.975, 7 - 3) \quad qChi = 11.143$$



$$\chi < qChi = 1$$

Значение статистики критерия попадает в область принятия гипотезы. Поэтому у нас нет оснований отвергать нулевую гипотезу