Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

# ОТЧЕТ

по лабораторной работе №1.

по курсу «Математическая статистика.»

на тему: «Гистограмма и эмпирическая функция распределения.»

Студент ИУ7-62Б (Группа) ______ (Подпись, дата) Трунов А. Р. (И. О. Фамилия)

Преподаватель ______ (Подпись, дата) Влассов П. А. (И. О. Фамилия)

*2022 г.*

# Теоретическая часть

## Формулы для вычисления величин

1. **Максимальное значение выборки** находится по формуле

$$M_{\max} = \max\{x_1, .., x_n\}.$$

2. **Минимальное значение выборки** находится по формуле

$$M_{\min} = \min\{x_1, .., x_n\}.$$

3. **Размах выборки** находится по формуле

$$R = M_{\max} - M_{\min}.$$

4. **Выборочное среднее** находится по формуле

$$\hat{\mu}(\vec{x}) = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}.$$

5. **Исправленная выборочная дисперсия** находится по формуле

$$S^2(\vec{x}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

## Определение эмпирической функции распределения

Пусть $\vec{x} = (x_1, .., x_n)$ выборка из генеральной совокупности $X$. Обозначим $n(t, \vec{x})$ – число элементов вектора $\vec{x}$, которые имеют значения меньше, чем $t$.

**Эмпирической функцией распределения**, построенной по выборке $\vec{x}$, называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определённую правилом

$$F_n(t) = \frac{n(t, \vec{x})}{n}.$$

**<u>Замечания</u>:**

1. Функция $F_n$ может принимать значения из множества

$$\{0 = \frac{0}{n}, \frac{1}{n}, \frac{2}{n}, .., \frac{n-1}{n}, \frac{n}{n} = 1\}.$$

2. Функция $F_n$ – кусочно-постоянная функция, которая имеет скачки (разрывы первого рода) в точках $t = z_{(i)}$, где $z_{(i)}$ – все попарно различные значения элементов выборки $\vec{x}$, $i = \overline{1, m}$.

3. Если все элементы выборки $\vec{x}$ попарно различны, то

$$F_n(t) = \begin{cases} 0, \ t \leqslant x_{(1)} \\ \frac{1}{n}, \ x_{(1)} < t \leqslant x_{(2)} \\ \ldots \\ \frac{n-1}{n}, \ x_{(n-1)} < t \leqslant x_{(n)} \\ 1, \ x_{(n)} < t \end{cases}$$

4. Эмпирическая функция распределения обладает всеми свойствами функции распределения.

## Определение эмпирической плотности и гистограммы

Пусть $\vec{x}$ – выборка из генеральной совокупности $X$. Если объём $n$ этой выборки достаточно велик ($n > 50$), то элементы выборки группируют в интервальный статистический ряд.

Для этого отрезок $J = [x_{(1)}, x_{(n)}]$, где $x_{(1)} = min\{x_1, .., x_n\}$ и $x_{(n)} = max\{x_1, .., x_n\}$ , разбивают на $m$ равновеликих промежутков, ширина каждого из которых

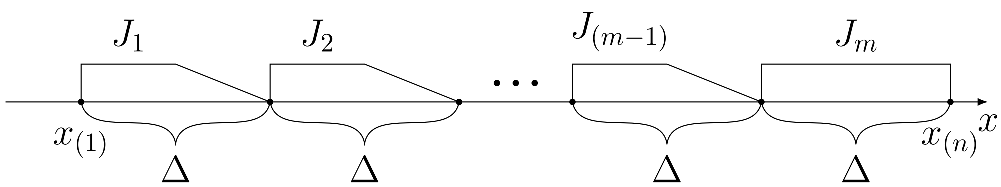
$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m}\text{, где } m = \lfloor \log_2 n \rfloor + 2.$$

Далее полагают:

$$J_i = [x_{(1)} + (i - 1)\Delta;\ x_{(1)} + i\Delta),\ i = \overline{1,\ m-1};$$

$$J_m = [x_{(1)} + (m - 1)\Delta;\ x_{(n)}].$$

Выше написанные формулы для промежутков $J_i$ и $J_m$ наглядно изображены на рисунке ниже.




**Интервальным статистическим рядом**, отвечающим выборке $\vec{x}$, называют таблицу вида

| | | | | |
|---|---|---|---|---|
| $J_1$ | $\cdots$ | $J_i$ | $\cdots$ | $J_m$ |
| $n_1$ | $\cdots$ | $n_i$ | $\cdots$ | $n_m$ |

Здесь $n_i$ – число элементов выборки $\vec{x}$, попавших в промежуток $J_i$, где $i = \overline{1,\ m}$.

Пусть для данной выборки $\vec{x}$ построен интервальный статистический ряд $(J_i,\ n_i)$, где $i = \overline{1,\ m}$.

**Эмпирической плотностью распределения** (соответствующей выборке $\vec{x}$) называется функция

$$f_n(x) = \begin{cases} \frac{n_i}{n\Delta}\ , \text{если } x \in J_i,\ i = \overline{1,\ m} \\ 0, \quad \text{иначе} \end{cases}$$

<u>**Замечания:**</u>

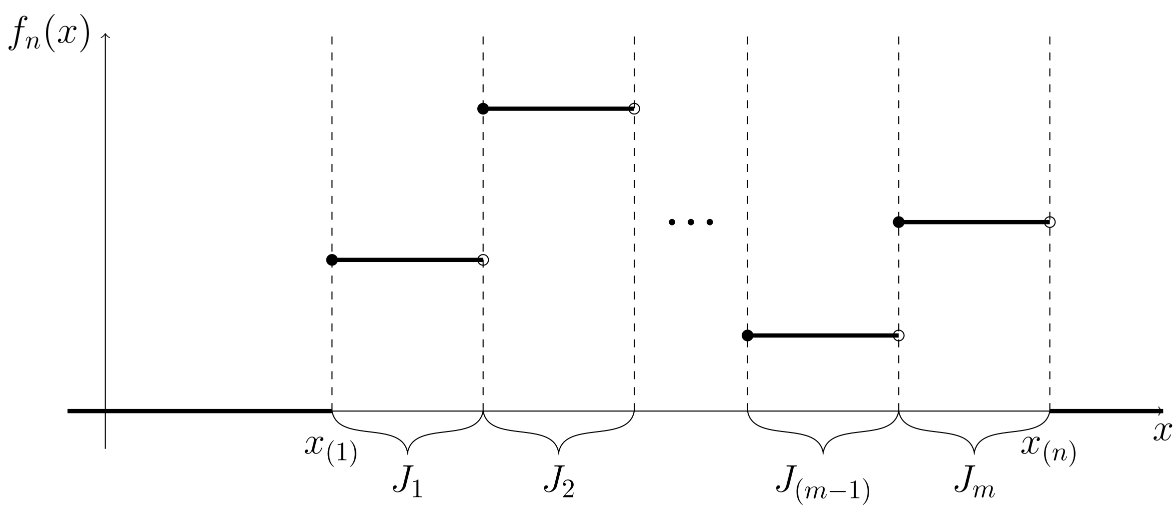
1. Эмпирическая плотность удовлетворяет условию нормировки и обладает всеми свойствами функции плотности.

$$\int\limits_{-\infty}^{+\infty} f_n(x)\,dx = \int\limits_{x_{(1)}}^{x_{(n)}} f_n(x)\,dx =$$

$$\sum_{i=1}^{m} \int\limits_{J_i} f_n(x)\,dx = \sum_{i=1}^{m} \left(\frac{n_i}{n\Delta} \cdot \Delta\right) =$$

$$\frac{1}{n}\sum_{i=1}^{m} n_i = \frac{n}{n} = 1.$$

2. Функция $f_n(x)$ – кусочно-постоянная функция.




3. Функция $f_n(x)$ – статистический аналог функции плотности распределения вероятностей.

$$\hat{f}_n(x) \xrightarrow[n \to \infty]{P} f(x) \;\Leftrightarrow\; f_n(x) \approx f(x), \text{ при } n \gg 1.$$

**Гистограммой** называется график эмпирической функции плотности.

# Практическая часть

## Текст программы

```
%Лабораторная работа №1. Вариант №22.

function main()
    pkg load statistics;
    %Генеральной совокупностью называется множество возможных значений
    %случайного величины X с учётом её закона распределения.

    %Создание выборки объёма n из генеральной совокупности X.
    X = [7.76 5.96 4.58 6.13 5.05 6.40 7.46 5.55 5.01 ...
         3.79 7.65 8.87 5.94 7.25 6.76 6.92 6.68 4.89 ...
         7.47 6.53 6.76 6.96 6.58 7.92 8.47 6.27 8.05 ...
         5.24 5.60 6.69 7.55 6.02 7.34 6.81 7.22 6.39 ...
         6.40 8.28 5.39 5.68 6.71 7.89 5.69 5.18 7.84 ...
         7.18 7.54 6.04 4.58 6.82 4.45 6.75 5.28 7.42 ...
         6.88 7.10 5.24 9.12 7.37 5.50 5.52 6.34 5.31 ...
         7.71 6.88 6.45 7.51 6.21 7.44 6.15 6.25 5.59 ...
         6.68 6.52 4.03 5.35 6.53 3.68 5.91 6.68 6.18 ...
         7.80 7.17 7.31 4.48 5.69 7.11 6.87 6.14 4.73 ...
         6.60 5.61 7.32 6.75 6.28 6.41 7.31 6.68 7.26 ...
         7.94 7.67 4.72 6.01 5.79 7.38 5.98 5.36 6.43 ...
         7.25 5.54 6.66 6.47 6.84 6.13 6.21 5.52 6.33 ...
         7.55 6.24 7.84];

    %Нахождения количества элементов в выборке.
    len_X = length(X);

    %Нахождение максимального и минимального элементов выборки X.
    M_max = max(X);
    M_min = min(X);
    fprintf('Нахождение максимального и минимального элементов выборки X.\nM_max = %f;\nM_min = %f;\n\n', M_max, M_min);

    %Нахождение размаха выборки X.
    %R = razmah(X);
```

```
    R = M_max - M_min;
    fprintf('Нахождение размаха выборки X.\nR = %f;\n\n', R);

    %Нахождение выборочного среднего выборки X.
    _mu_ = mean(X);
    %_mu_ = sum(X) / len_X;
    fprintf('Нахождение выборочного среднего выборки X.\nmu = %f;\n\n', _mu_);

    %Нахождение исправленной выборочной дисперсии выборки X.
    S_2 = var(X);
    %S_2 = sum(power(X - _mu_, 2)) / (len_X - 1);
    fprintf('Нахождение исправленной выборочной дисперсии выборки X.\nS_2 = %f;\n\n', S_2);

    %Группировка значений выборки в m = [log_2 n] + 2 интервала.
    [N, J, m] = IntervalStaticRow(X, len_X, M_min, M_max, R);

    %Построение на одной координатной плоскости гистограммы и графика функции плотности
    %распределения вероятностей нормальной случайной величины с математическим ожиданием
    %_mu_ и дисперсией S_2.
    MakeHistogram_1(X, N, J, m, len_X, R, M_min, M_max, _mu_, S_2);

    %Построение на другой координатной плоскости графика эмпирической функции распределения
    %и функции распределения нормальной случайной величины с математическим ожиданием
    %_mu_ и дисперсией S_2.
    MakeHistogram_2(X, len_X, R, M_min, M_max, _mu_, S_2);
end

function R = razmah(X)
    R = max(X) - min(X);
end

function [N, J, m] = IntervalStaticRow(X, len_X, M_min, M_max, R)
    %Нахождение m - количества требуемых интервалов.
```

```
    m = floor(log2(len_X)) + 2;

    %Создание вектора интервалов.
    J = M_min : (R / m) : M_max;

    %Нахождение количества элементов в каждом из интервалов.
    N = histc(X, J);
    len_N = length(N);
    N(len_N - 1) += N(len_N);
    N(len_N) = [];
    len_N = length(N);

    %Вывод интервального статистического ряда.
    fprintf('Интервальный статистический ряд выборки X (m = %d):\n', m);
    for i = 1 : (m - 1)
        fprintf('[ %f ; %f ) - %d;\n', J(i), J(i + 1), N(i));
    endfor
    fprintf('[ %f ; %f ] - %d;\n', J(len_N), J(len_N + 1), N(len_N));
end

function MakeHistogram_1(X, N, J, m, len_X, R, M_min, M_max, _mu_, S_2)
    %Построение окна.
    figure("numbertitle", "off", ...
            "name", "Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием _mu_ и дисперсией S_2.", ...
            "position", [10, 100, 1000, 850]);

    Ni = [0, N / (len_X * (R / m)), 0];
    Ji = [M_min, J];

    %Построение гистограммы.
    stairs(Ji, Ni, "linewidth", 1.2, "color", "b");

    hold on;

    x = (M_min : (R / (len_X - 1)) : M_max);
    y = normpdf(x, _mu_, sqrt(S_2));
```

```
    %Построение графика.
    plot(x, y, "color", "r", "linewidth", 1.2);

    title("Гистограмма и график функции плотности распределения вероятностей\nнормальной случайной величины с математическим ожиданием mu и дисперсией S^2.", ...
        "fontsize", 15);
    legend("Гистограмма", "График", "fontsize", 15);
    grid;

    hold off;
end

function Y = ecdf(X, I)
    ni = length(I);
    nx = length(X);

    Y = zeros(1, ni);
    j = 1;
    i = 1;
    while (i <= nx && j <= ni)
        if (X(i) >= I(j))
            Y(j) = i - 1;
            j = j + 1;
        else
            i = i + 1;
        endif
    endwhile

    while (j <= ni)
        Y(j) = i - 1;
        j = j + 1;
    endwhile

    for i = 1:ni
        Y(i) = Y(i) / nx;
    end

```

```
end

function MakeHistogram_2(X, len_X, R, M_min, M_max, _mu_, S_2)
    %Построение окна.
    figure("numbertitle", "off", ...
           "name", "График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием _mu_ и дисперсией S_2.", ...
           "position", [10, 100, 1000, 850]);

    %Построение графиков.
    I = (M_min : (R / (len_X - 1)) : M_max);
    Y = ecdf(sort(X), I);
    stairs(I, Y, "linewidth", 1.2, "color", "b");

    hold on;

    x = (M_min: (R / (len_X - 1)) : M_max);
    y = normcdf(x, _mu_, sqrt(S_2));
    plot(x, y, "color", "r", "linewidth", 1.2);

    title("График эмпирической функции распределения и функции распределения\nнормальной случайной величины с математическим ожиданием mu и дисперсией S^2.", ...
          "fontsize", 15);
    legend("Функция №1", "Функция №2", "fontsize", 15, "location", "northwest");
    grid;

    hold off;
end
```

## Результаты расчётов для выборки из индивидуального варианта – 22 вариант

Ниже представлен результат работы программы, описанной выше.

```
>> main

Нахождение максимального и минимального элементов выборки X.
M_max = 9.120000;
M_min = 3.680000;

Нахождение размаха выборки X.
R = 5.440000;

Нахождение выборочного среднего выборки X.
mu = 6.459583;

Нахождение исправленной выборочной дисперсии выборки X.
S_2 = 1.101315;

Интервальный статистический ряд выборки X (m = 8):
[ 3.680000 ; 4.360000 ) - 3;
[ 4.360000 ; 5.040000 ) - 8;
[ 5.040000 ; 5.720000 ) - 20;
[ 5.720000 ; 6.400000 ) - 22;
[ 6.400000 ; 7.080000 ) - 30;
[ 7.080000 ; 7.760000 ) - 25;
[ 7.760000 ; 8.440000 ) - 9;
[ 8.440000 ; 9.120000 ] - 3;
```

Рисунок 1 – Вывод информации в консоль при запуске программы.

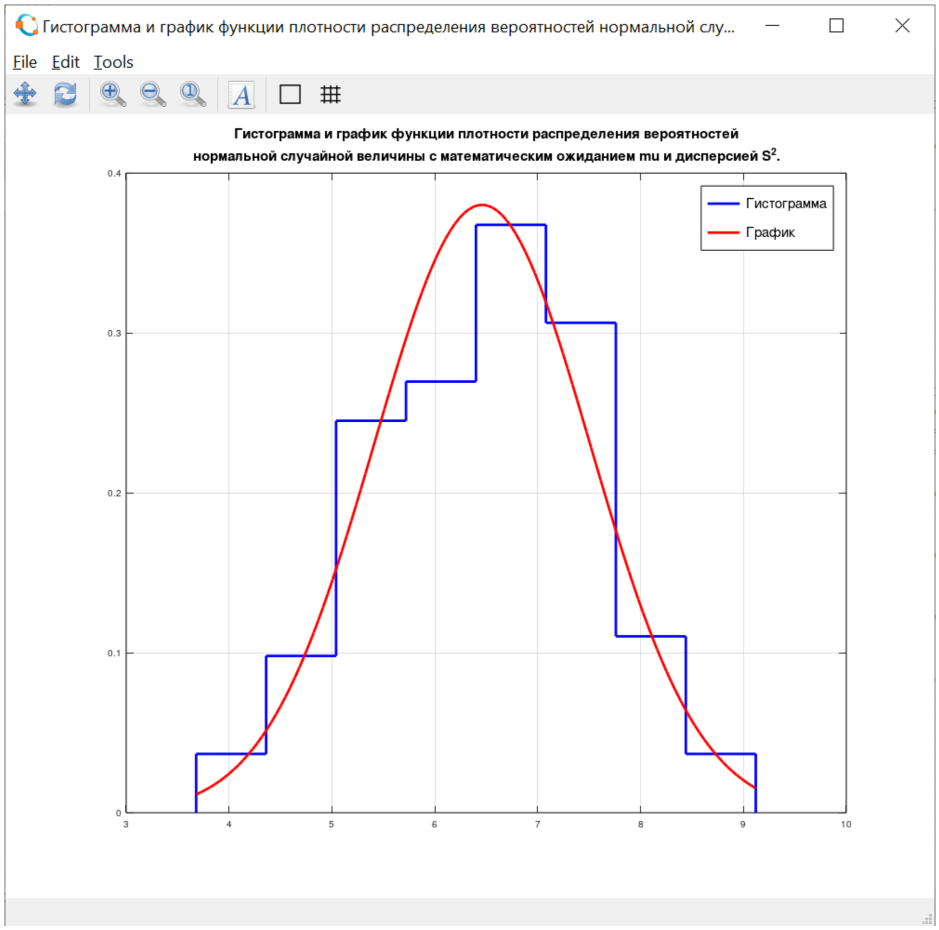



Рисунок 2 – Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией $S^2$.

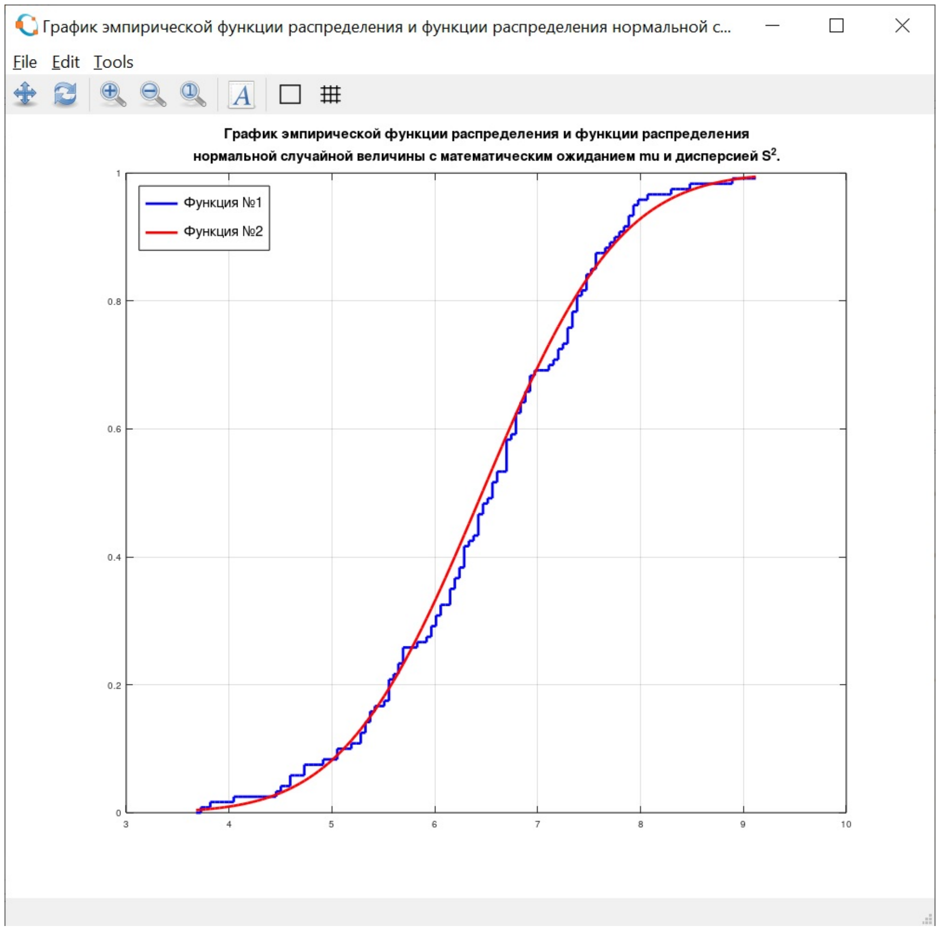



Рисунок 3 – График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией $S^2$.