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

ФАКУЛЬТЕТ «Информатика и системы управления»
КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

Лабораторная работа № 1

Дисциплина	Математическая статистика.
Тема	Гистограмма и эмпирическая функция распределения.
Студент	Сиденко А.Г.
Группа	ИУ7-63Б
Оценка (баллы)	
Преподаватель	Власов П.А.

Москва, 2020 г.

1 Формулы для вычисления величин

$$\vec{x} = (x_1, ..., x_n)$$

1. **Максимальное значение выборки**

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

2. **Минимальное значение выборки**

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

3. **Размах выборки**

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

4. **Выборочное среднее (математическое ожидание)**

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

5. **Состоятельная оценка дисперсии**

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

где $\overline{x} = \hat{\mu}$

2 Определение эмпирической плотности и гистограммы

Эмпирической плотностью (отвечающей выборке $\vec{x}$) называют функцию

$$\hat{f}_n(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1;p} \\ 0, \text{ иначе} \end{cases}$$

где (J_i, n_i) – интервальный статистический ряд

Пусть $\vec{x}$ – выборка из генеральной совокупности X. Если объем n этой выборки велик, то значения x_i группируют не только в статистический ряд, но и в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ (где $x_{(1)} = \min\{x_1, .., x_n\}$, $x_{(n)} = \max\{x_1, .., x_n\}$) делят на p равновеликих частей:

$$J_i = [a_i, a_{i+1}), i = \overline{1; p-1}$$

$$J_p = [a_p, a_{p+1}]$$

$$a_i = x_{(1)} + (i-1) \cdot \Delta, i = \overline{1; p+1}$$

$$\Delta = \frac{|J|}{p} = \frac{x_{(n)} - x_{(1)}}{p}$$

Интервальным статистическим рядом называют таблицу

J_1	...	J_i	...	J_p
n_1	...	n_i	...	n_p

Здесь n_i – количество элементов выборки $\vec{x}$, которые $\in J_i$

В нашем случае $p = m = [\log_2 n] + 2$

Гистограммой называют график эмпирической плотности.

3 Определение эмпирической функции распределения

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

Эмпирической функцией распределения называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определенную условием $F_n(x) = \frac{n(x,\vec{x})}{n}$.

4 Текст программы

```
function lab1()
    % Выборка объема 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];
    % Максимальное значение
    Mmax = max(X);
    % Минимальное значение
    Mmin = min(X);
    % Размах выборки
    R = Mmax - Mmin;
    % Выборочное среднее
    mu = mean(X);
    % Состоятельная оценка дисперсии
    s2 = var(X);

    % Вывод полученных ранее значений
    fprintf('Mmax = %f\n', Mmax);
    fprintf('Mmin = %f\n', Mmin);
    fprintf('R = %f\n', R);
    fprintf('mu = %f\n', mu);
    fprintf('S2 = %f\n', s2);
    % Построить интервальный ряд
    [count, edges, m] = groupInterval(X);

    % Построение гистограммы
    plotHistogram(X, count, edges, m);
    % Построение на одной координатной плоскости
    hold on;
    % График функции плотности распределения вероятностей нормальной
    % случайной величины
    fn = @(x, mu, s2) normpdf(x, mu, s2);
    plotGraph(fn, mu, s2, Mmin, Mmax, 0.1);

    % Новая координатная плоскость
    figure;
    % График эмпирической функции распределения
    plotEmpiricalF(X);
    % Построение на одной координатной плоскости
    hold on;
    % График функции распределения нормальной случайной величины
    Fn = @(x, mu, s2) normcdf(x, mu, s2);
    plotGraph(Fn, mu, s2, Mmin, Mmax, 0.1);
end

% Функция для группировки значений выборки
function [count, edges, m] = groupInterval(X)
    % Нахождение количества интервалов
    m = floor(log2(length(X))) + 2;
    % С помощью функции histcounts разбиваем выборку на m интервалов от
    % минимума до максимума. Возвращаем интервалы и количество элементов
    % в каждом из них
    [count, edges] = histcounts(X, m, 'BinLimits', [min(X), max(X)]);
```

```
    lenC = length(count);

    % Вывод интервалов и количества элементов
    fprintf('\nИнтервальный ряд для m = %d \n', m);
    for i = 1 : (lenC - 1)
        fprintf('[%f,%f) - %d\n', edges(i), edges(i + 1), count(i));
    end
    fprintf('[%f,%f] - %d\n', edges(lenC), edges(lenC + 1), count(lenC));
end

% Функция для отрисовки гистограммы
function plotHistogram(X, count, edges, m)
    % Построение гистограммы
    h = histogram();
    % Задаем интервалы
    h.BinEdges = edges;
    % Задаем значение в каждом интервале (эмпирическую плотность)
    h.BinCounts = count / length(X) / ((max(X) - min(X)) / m);
    h.LineWidth = 2;
    h.DisplayStyle = 'stairs';
end

% Функция для отрисовки графиков func, с математическим ожиданием mu
% и дисперсией s2, от min до max с шагом step
function plotGraph(func, mu, s2, min, max, step)
    x = min : step : max;
    y = func(x, mu, s2);
    plot(x, y, 'LineWidth', 2);
end

% График эмпирической функции распределения
function plotEmpiricalF(X)
    % Поиск уникальных элементов
    u = unique(X);
    % Подсчет количества каждого из уникальных элементов
    count = histcounts(X, u);
    % Подсчет количества элементов, меньших текущего уникального элемента
    for i = 2 : (length(count))
        count(i) = count(i) + count(i - 1);
    end
    count = [0 count];
    % Отрисовка графика
    stairs(u, count / length(X), 'LineWidth', 2);
end
```

5 Результаты расчетов для выборки из индивидуального варианта (вариант 22)

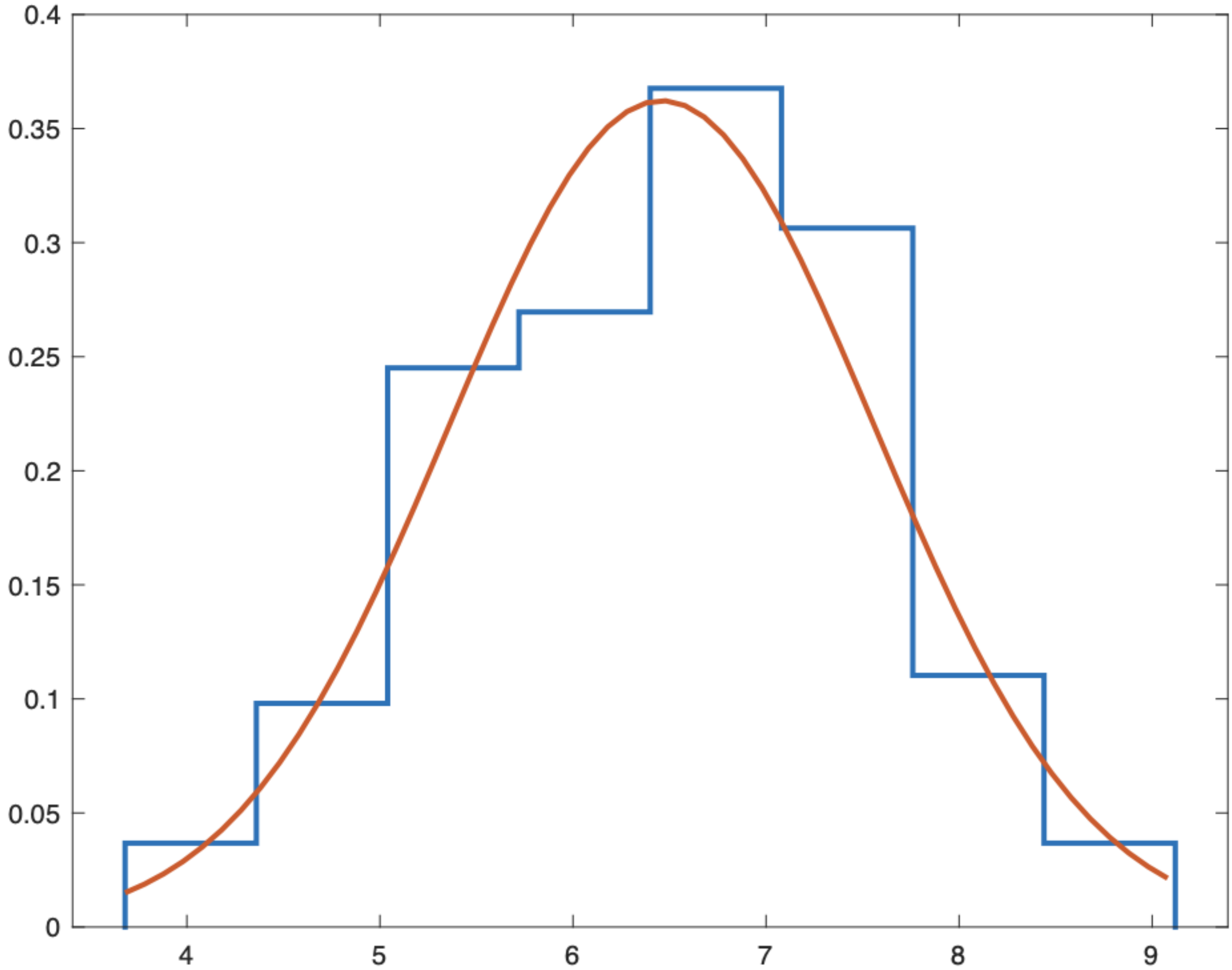
1. Максимальное значение выборки

2. Минимальное значение выборки
3. Размах выборки
4. Выборочное среднее (математическое ожидание)
5. Состоятельная оценка дисперсии
6. Группировка значений выборки в $m = [log_2 n] + 2$ интервала

```
Mmax = 9.120000
Mmin = 3.680000
R = 5.440000
mu = 6.459583
S2 = 1.101315

Интервальный ряд для 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
```

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



0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
4
5
6
7
8
9

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