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»

Лабораторная работа № 1
по дисциплине «Математическая статистика»

Тема Гистограмма и эмпирическая функция распределения

Студент Шахнович Дмитрий Сергеевич

Группа ИУ7-62Б

Вариант №20

Преподаватель Власов П. А.

Москва, 2025

1 Задание

Цель работы: построение гистограммы и эмпирической функции распределения.

Задачи:

1) Для выборки объёма n из генеральной совокупности X реализовать в виде программы на ЭВМ

— вычисление максимального значения M_{max} и минимального значения M_{min};

— размаха R выборки;

— вычисление оценок $\widehat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;

— группировку значений выборки в $m = \lfloor log_2 n \rfloor + 2$ интервала;

— построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\widehat{\mu}$ и дисперсией S^2;

— построение на другой координатной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\widehat{\mu}$ и дисперсией S^2.

2) Провести вычисления и построить графики для выборки из индивидуального варианта.

2 Математические сведения

2.1 Формулы для вычисления M_{max}, M_{min}, R, $\widehat{\mu}$, S^2

Пусть $\vec{X} = X_1, \ldots, X_n$ – случайная выборка из генеральной совокупности X, а $X_{(1)}, \ldots, X_{(n)}$ – вариационный ряд этой случайной выборки. Тогда формулы расчёта минимального и максимального значения имеют вид 2.1.

$$\begin{aligned} M_{max} = X_{(n)} \\ M_{min} = X_{(1)} \end{aligned} \tag{2.1}$$

Формула вычисления размаха R имеет вид 2.2.

$$R = M_{max} - M_{min} \tag{2.2}$$

В качестве оценок математического ожидания MX и дисперсии DX выступят выборочное среднее $\overline{X}$ и исправления выборочная дисперсия соответственно. Формулы их вычисления имеют вид 2.3 и 2.4.

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \tag{2.3}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{\mu})^2 \tag{2.4}$$

2.2 Определение эмпирической плотности и гистограммы

Пусть дана $\vec{x} = x_1, \ldots, x_n$ – выборка из генеральной совокупности X, и $x_{(1)}, \ldots, x_{(n)}$ – вариационный ряд этой выборки. При больших значениях n (n > 50) значения выборки можно сгруппировать. Для этого отрезок $J = [x_{(1)}; x_{(n)}]$ делят на $m - 1$ равномощных интервалов и 1 отрезок 2.5, при этом $m = \lfloor log_2(n) \rfloor + 2$

$$J_i = \begin{cases} [x_(1) + (i-1)\triangle; x_(1) + i\triangle) & i \in [1, m-1], \\ [x_(1) + (m-1)\triangle; x(n)] & i = m, \end{cases} \tag{2.5}$$

где $\triangle = \frac{|J|}{m} = \frac{x(n) - x(1)}{m}$.

На основе этого разбиение, формируется интервальный статический ряд – таблица вида 2.6, где n_i – количества элементов $v \in \vec{x}$, таких, что $v \in J_i, i \in [1, m]$.

$$\begin{aligned} J_1 \ldots J_i \ldots J_m \\ n_1 \ldots n_i \ldots n_m \end{aligned} \tag{2.6}$$

Итак, эмпирической плотностью распределения, отвечающей выборке $\vec{x}$, называется функция $f_n(x)$, такая что:

$$f_n(x) = \begin{cases} \frac{n_i}{n\triangle} & \exists i \in [1, m], x \in J_i, \\ 0 & x \notin J_i, \forall i \in [1, m]. \end{cases} \tag{2.7}$$

Гистограммой называется график эмпирической функцией плотности.

2.3 Определение эмпирической функции распределения

Пусть дана $\vec{x} = x_1, \ldots, x_n$ – выборка из генеральной совокупности X.

Эмпирической функцией распределения называется функция $F_n(x) = \frac{h(x,\vec{x})}{n}$, где $h(x, \vec{x})$ – количество элементов вектора $\vec{x}$, которые меньше x.

3 Текст разработанной программы

Листинг 3.1 — Программа лабораторной работы

```
%Исходная выборка и её преобразование в вариационный ряд
X = [14.90,14.40,13.56,15.55,13.97,16.33,14.37,13.46,15.51,
    14.69,13.41,14.24,15.65,14.54,13.55,13.15,14.32,15.04,
    13.27,14.60,13.83,13.93,14.11,14.15,15.48,15.96,14.46,
    13.87,13.67,15.30,13.95,16.08,18.25,14.93,15.37,14.38,
    15.56,13.92,14.23,12.80,13.16,13.89,14.24,13.90,12.82,
    13.20,13.89,13.50,13.44,16.13,14.68,15.27,13.35,13.62,
    16.16,16.46,13.83,14.13,15.68,15.22,12.59,12.94,13.09,
    16.54,14.61,14.63,14.17,13.34,16.74,16.30,13.74,15.02,
    14.96,15.87,16.03,12.87,14.32,14.48,14.57,14.43,12.61,
    14.52,15.29,12.07,14.58,11.74,14.97,14.31,12.94,12.82,
    14.13,14.48,12.25,14.39,15.08,12.87,14.25,15.12,15.35,
    12.27,14.43,13.85,13.16,16.77,14.47,14.89,14.95,14.55,
    12.80,15.26,13.32,14.92,13.44,13.48,12.81,15.01,13.19,
    14.68,14.44,14.89];
X = sort(X);

%а) Вычисление максимального значения Mmax и минимального значения
    Mmin
Xmax = X(end);
Xmin = X(1);

fprintf("Минимальное значение выборки: %f.\n", Xmin);
fprintf("Максимальное значение выборки: %f.\n", Xmax);

%б) Вычисление размаха R выборки
R = Xmax - Xmin;
fprintf("Размах выборки: %f.\n", R);

%в) Вычисление оценок математического ожидания MX и дисперсии DX
mx = sum(X) / length(X); % выборочное среднее
dx = sum((X - mx).^2) / (length(X) - 1); % несмещённая оценка дисперс
    ии
fprintf("Математическое ожидание X: %f.\n", mx);
fprintf("Дисперсия X: %f.\n", dx);

%г) Группировка значений выборки в m = [log2 n] + 2 интервала
m = floor(log2(length(X))) + 2;
```

```
delta = R / m;

borders = Xmin;

for i=2:(m+1)
borders(i) = borders(i - 1) + delta;
end
function statSeries = CreateSeries(m, borders, X)
statSeries = zeros(1, m);
function AddToSeries(val)
for i=1:(m - 1)
if (val >= borders(i)) && (val < borders(i+1))
statSeries(i) = statSeries(i) + 1;
end
end
% Для учёта элементов, равных максимальному
if (val >= borders(m)) && (val <= borders(m + 1))
statSeries(m) = statSeries(m) + 1;
end
end
arrayfun(@AddToSeries, X)
end
statSeries = CreateSeries(m, borders, X);
check = sum(statSeries) == length(X);
fprintf("Сумма интервального статистического ряда равна объему выборк
   и? : %d.\n", check);
out = sprintf("Интервальный статистический ряд: \n");
for i=1:m
out = [out, sprintf("[%f;%f) : %d\n", borders(i), borders(i+1),
   statSeries(i))];
end
disp(join(out, ''))

%д) Построение на одной координатной плоскости гистограммы и графика
   функции плотности распределения вероятностей нормальной случайной
   величины с математическим ожиданием MX и дисперсией DX
empirDist = statSeries / (length(X) * delta);
empirDist
x = Xmin:R/1000:Xmax;
y = normpdf(x, mx, dx);
figure;
```

```
hold on
grid on
set(gca, 'ytick', sort(empirDist))
plot(x, y);
histogram(X, borders, 'Normalization', 'pdf', 'FaceAlpha', 0.1, '
   EdgeAlpha', 0.5);
hold off

%e) построение на другой координатной плоскости графика эмпирической
   функции распределения и функции распределения нормальной случайной
    величины с математическим ожиданием MX и дисперсией DX
empirFunc = zeros(1, m);
empirFunc(1) = empirDist(1) * delta;
for i=2:m
empirFunc(i) = empirDist(i) * delta + empirFunc(i-1);
end
empirFunc

x = Xmin:R/1000:Xmax;
y = normcdf(x, mx, dx);
figure;
hold on;
grid on;
plot(x, y);

[uniqX, ~, ind] = unique(X);
cntX = accumarray(ind, 1);
cumcntX = cumsum(cntX);
edX = cumcntX / length(X);
stairs([-inf; uniqX(:)], [0; edX(:)], 'LineWidth', 1.5);


hold off;
```

4 Результаты расчётов

Границы выборки: $M_{max} = 18.25$, $M_{min} = 11.74$

Размах выборки: $R = 6.51$

Оценка математического ожидания: $\widehat{\mu} = 14.349167$

Оценка дисперсии: $S^2 = 1.277621$

Интервальный статистический ряд:

[11.740000;12.553750) : 4

[12.553750;13.367500) : 21

[13.367500;14.181250) : 27

[14.181250;14.995000) : 37

[14.995000;15.808750) : 18

[15.808750;16.622500) : 10

[16.622500;17.436250) : 2

[17.436250;18.250000] : 1

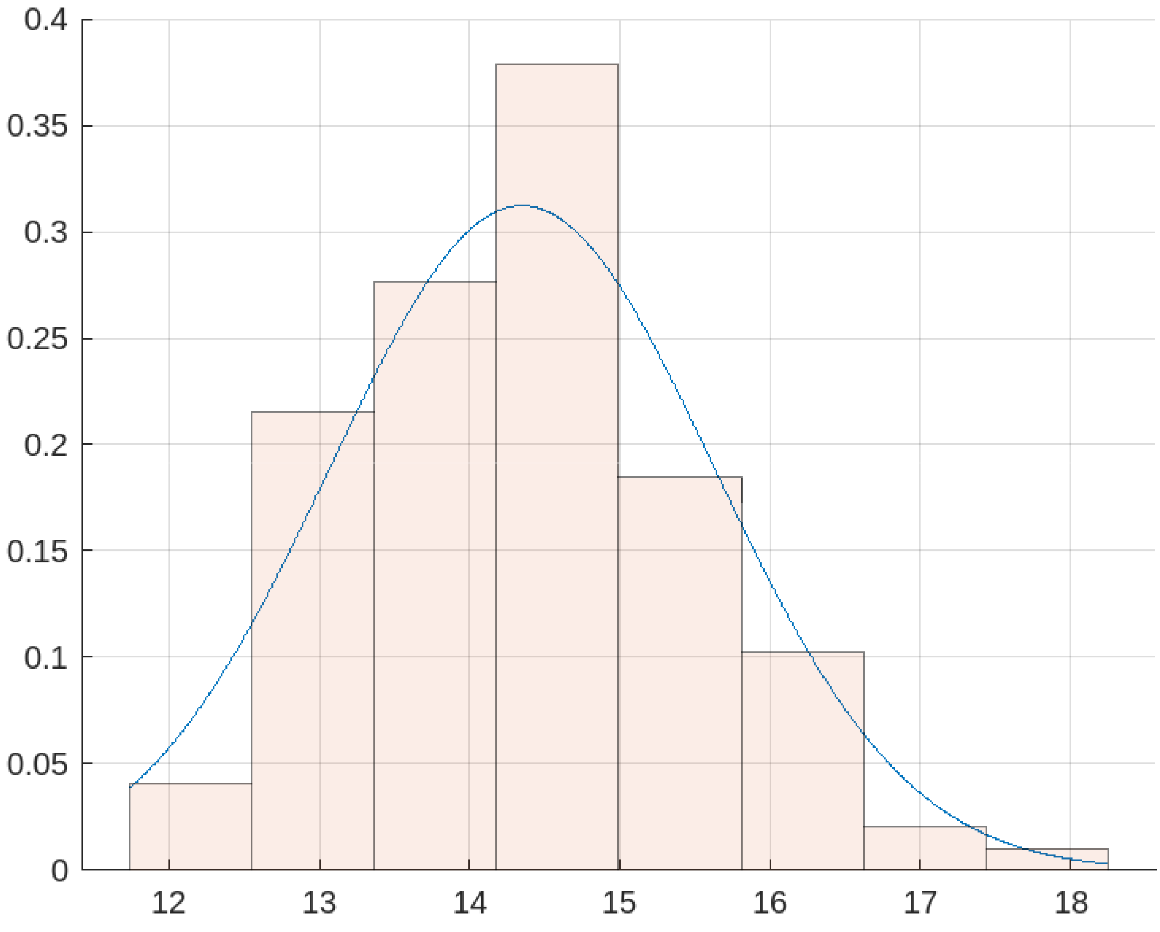

Рисунок 4.1 — Гистограмма выборки и функция плотности распределения вероятности нормальной величины с параметрами $\widehat{\mu}$ и S^2

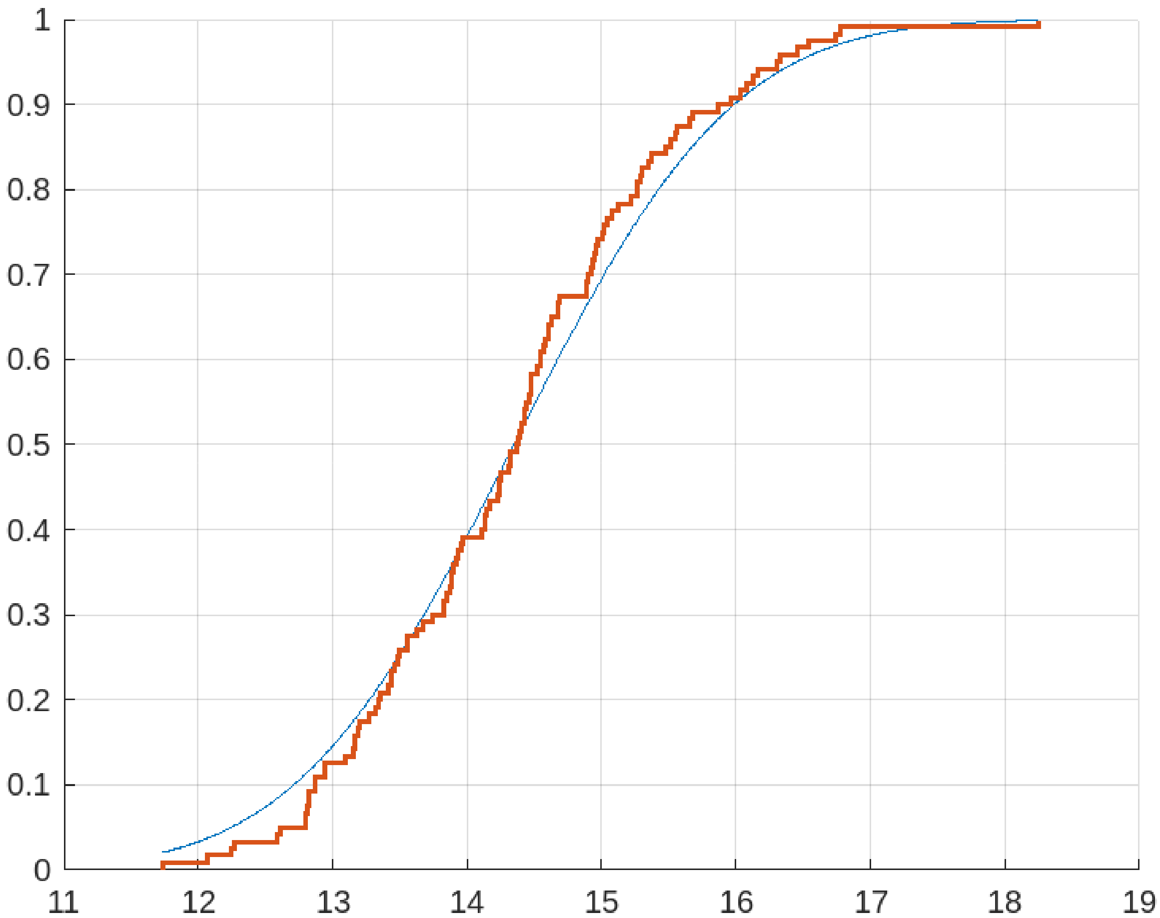

Рисунок 4.2 — Эмпирическая функция распределения и функция распределения вероятности нормальной величины с параметрами $\widehat{\mu}$ и S^2