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"

Тема Гистограмма и эмпирическая функция распределения

Студент Хамзина Р. Р.

Группа ИУ7-63Б

Оценка (баллы) ______

Преподаватели Власов П. А.

Москва — 2022 г.

1 Задание

Цель работы: построение гистограммы и эмпирической функции распределения.

1. Для выборки объема n из генеральной совокупности X реализовать в виде программы на ЭВМ

   (a) вычисление максимального значения $M_{\max}$ и минимального значения $M_{\min}$;

   (b) размаха R выборки;

   (c) вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;

   (d) группировку значений выборки в $m = [\log_2 n] + 2$ интервала;

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

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

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

2 Теоретическая часть

Пусть дана выборка

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

объема n из генеральной совокупности X.

2.1 Формулы для вычисления величин

Максимальное и минимальное значения выборки

$$\begin{aligned} M_{\max} &= \max\{x_1, .., x_n\} \\ M_{\min} &= \min\{x_1, .., x_n\} \end{aligned} \tag{2.1}$$

Размах выборки

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

Оценки математического ожидания и дисперсии

Выборочное среднее:

$$\hat{\mu}(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.3}$$

Исправленная выборочная дисперсия:

$$S^2(\vec{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2, \tag{2.4}$$

где $\overline{x} = \hat{\mu}$

2.2 Определение эмпирической плотности и гистограммы

Пусть $\vec{x}$ — выборка из генеральной совокупности X. Если объем этой выборки n достаточно велик, то значения x_i группируют в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ разбивают на m равновеликих промежутков. Ширина каждого из них:

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m} \tag{2.5}$$

Полагают

$$J_i = [x_{(1)} + (i-1)\cdot\Delta, x_{(1)} + i\cdot\Delta), i = \overline{1; m-1} \tag{2.6}$$

$$J_m = [x_{(1)} + (m-1)\cdot\Delta, x_{(n)}] \tag{2.7}$$

Интервальным статистическим рядом, отвечающим выборке $\vec{x}$, называется таблица вида

J_1	J_2	...	J_m
n_1	n_2	...	n_m

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

Для выбора m используют формулу:

$$m = [\log_2 n] + 2, \tag{2.8}$$

где n — размер выборки.

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

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

$$f_n(x) = \begin{cases} \frac{n_i}{n\Delta}, \text{если } x \in J_i, \\ 0, \text{иначе.} \end{cases} \tag{2.9}$$

График эмпирической функции плотности называется **гистограммой**.

2.3 Определение эмпирической функции распределения

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

Эмпирической функцией распределения, построенной по выборке $\vec{x}$, называют функцию

$$F_n : R \to R, \tag{2.10}$$

определенную правилом

$$F_n(t) = \frac{n(t, \vec{x})}{n} \tag{2.11}$$

3 Практическая часть

3.1 Текст программы

```
function lab_01
    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);
    n = length(X);

    M_max = max(X);
    M_min = min(X);
    fprintf("\nМаксимальное значение = %.4f\n", M_max);
    fprintf("Минимальное значение = %.4f\n", M_min);

    R = M_max - M_min;
    fprintf("\nРазмах выборки = %.4f\n", R);

    mu = sum(X) / n;
    S2 = sum((X-mu).^2) / (n - 1);
    fprintf("\nОценка математического ожидания = %.4f\n", mu);
    fprintf("Оценка дисперсии = %.4f\n", S2);

    m = floor(log2(n)) + 2;
    fprintf("\nЧисло интервалов = %d\n\n", m);

    d = (X(n) - X(1)) / m;

    J = [];

    for i = 1 : m
      J(i, 1) = X(1) + (i - 1) * d;
      J(i, 2) = X(1) + i * d;
    end

```

```
N = zeros(m);

for i = 1 : n
  for j = 1 : m
    if X(i) >= J(j, 1) && X(i) < J(j, 2)
      N(j)++;
    end
  end
end

N(m)++;

for i = 1 : m - 1
  fprintf("[%.4f,%.4f) - %d\n", J(i, 1), J(i, 2), N(i));
end

fprintf("[%.4f,%.4f] - %d\n", J(m, 1), J(m, 2), N(m));

f = [];

for i = 1 : m
  f(i, 1) = (J(i, 1) + J(i, 2)) / 2;
  f(i, 2) = N(i) / (n * d);
end

args = X(1):1e-4:X(n);
sigma = sqrt(S2);
f_normal = normpdf(args, mu, sigma);

bar(f(:,1), f(:,2), 1, "b");
hold on;
plot(args, f_normal, 'g', 'LineWidth', 2);
grid;

F = [];

F(1, 2) = 0;
F(1, 1) = X(1) - 1;

for i = 2 : (n + 1)
  cnt = 0;
  for j = 1 : n
    if X(j) <= X(i - 1)
      cnt++;
    end
  end

  F(i, 1) = X(i - 1);
```

```
        F(i, 2) = cnt / n;
    end

    F(n + 2, 2) = 1;
    F(n + 2, 1) = X(n) + 1;

    F_normal = normcdf(args, mu, sigma);

    figure;
    stairs(F(:,1), F(:,2), "b");
    hold on;
    plot(args, F_normal, "r", 'LineWidth', 2);
    grid;
endfunction
```

3.2 Результаты расчетов для выборки из индивидуального варианта

```
Максимальное значение = 18.2500
Минимальное значение = 11.7400

Размах выборки = 6.5100

Оценка математического ожидания = 14.3492
Оценка дисперсии = 1.2776

Число интервалов = 8

[11.7400,12.5538) - 4
[12.5538,13.3675) - 21
[13.3675,14.1813) - 27
[14.1813,14.9950) - 37
[14.9950,15.8087) - 18
[15.8087,16.6225) - 10
[16.6225,17.4363) - 2
[17.4363,18.2500] - 1
```

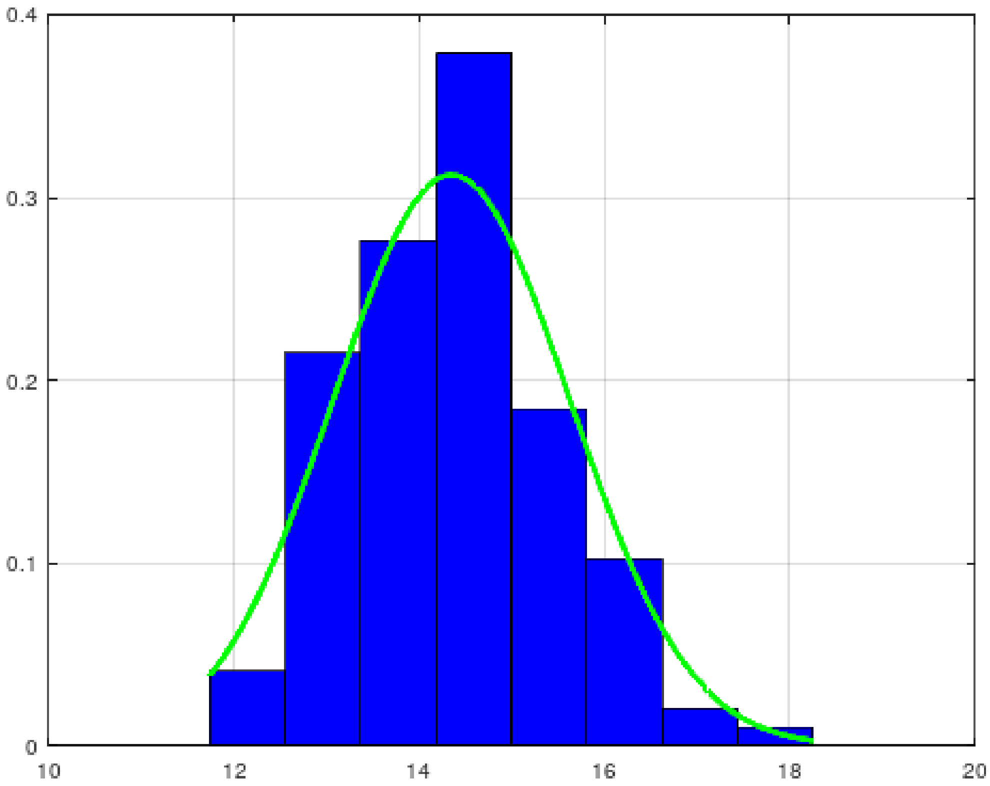

Рисунок 3.1 – Гистограмма и график функции плотности распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

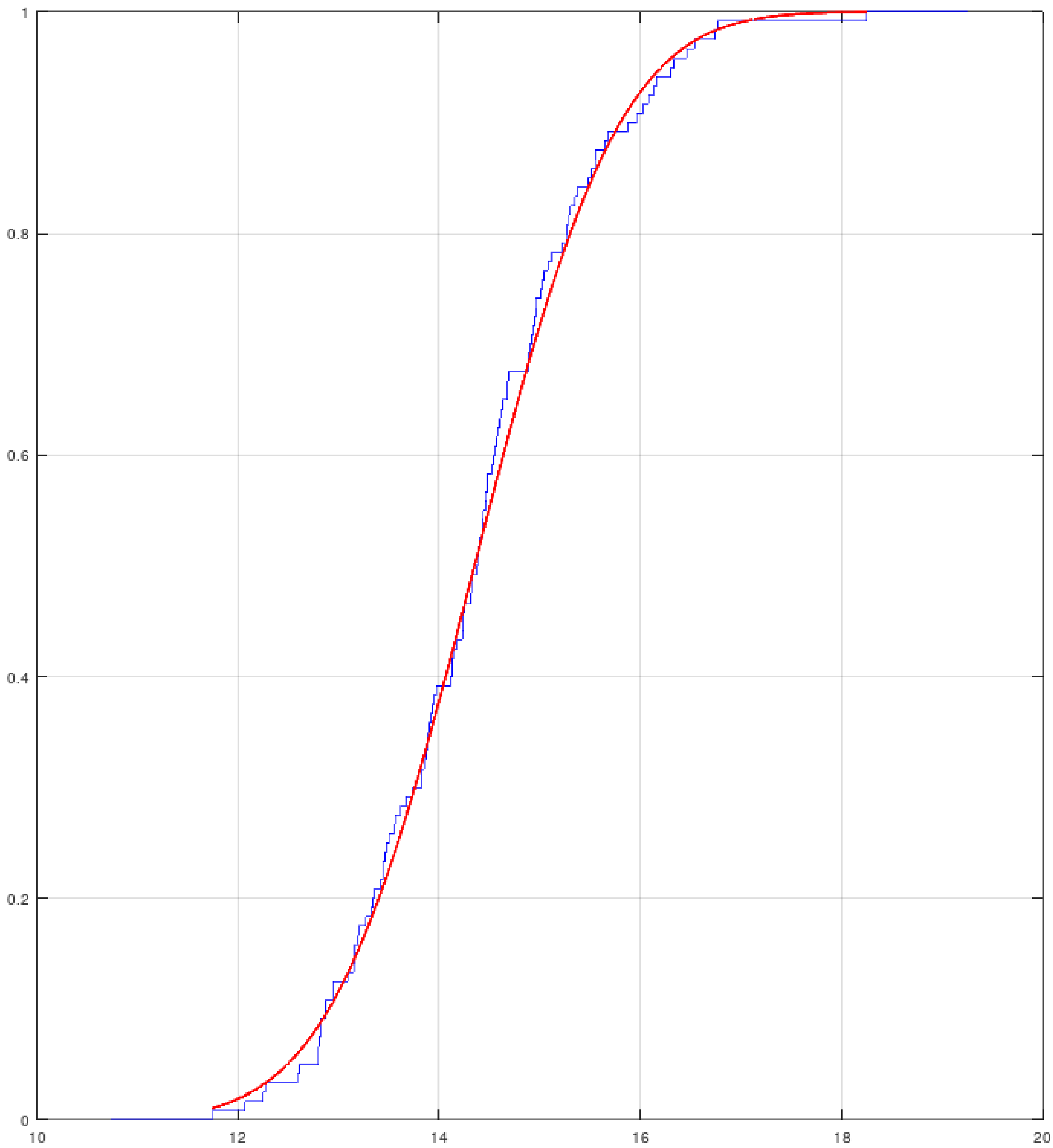

Рисунок 3.2 – График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2