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»

«Гистограмма и эмпирическая функции распределения»

Студент Маслова Марина Дмитриевна

Группа ИУ7-63Б

Оценка (баллы) ______

Преподаватель Власов Павел Александрович

2022 г.

1 Задание

Цель работы: построение гистограммы и эмпирической функции распределения.

Содержание работы

1. Для выборки объема n из генеральной совокупности X реализовать в виде программы на ЭВМ
   а) вычисление максимального значения M_{max} и минимального значения M_{min};
   б) размаха R выборки;
   в) вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;
   г) группировку значений выборки в $m = [\log_2 n] + 2$ интервала;
   д) построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
   е) 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.

Содержание отчета

1. формулы для вычисления величин M_{max}, M_{min}, R, $\hat{\mu}$, S^2;
2. определение эмпирической плотности и гистограммы;
3. определение эмпирической функции распределения;
4. текст программы;
5. результаты расчетов для выборки1 из индивидуального варианта.

2 Теоретическая часть

2.1 Формулы для вычисления величин

Пусть $\vec{x} = (x_1, ..., x_n)$ — выборка объема n из генеральной совокупности X.

Максимальное и минимальное значения выборки:

$$M_{max} = x_{(n)} = \max\{x_1, ..., x_n\}, \tag{2.1}$$

$$M_{min} = x_{(1)} = \min\{x_1, ..., x_n\}, \tag{2.2}$$

где $x_{(1)}, x_{(n)}$ — крайние члены вариационного ряда, отвечающего выборке $\vec{x}$.

Размах выборки:

$$R = M_{max} - M_{min}. \tag{2.3}$$

Оценка математического ожидания — выборочное среднее:

$$\hat{\mu}(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{2.4}$$

Оценка дисперсии — исправленная выборочная дисперсия:

$$S^2(\vec{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2, \tag{2.5}$$

где $\overline{x} = \hat{\mu}$.

2.2 Определение эмпирической плотности и гистограммы

Если объем выборки достаточно велик ($n > 50$), то элементы выборки группируют в так называемый интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ разбивают на m равновеликих промежутков. Ширина каждого из них:

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m}. \tag{2.6}$$

Далее полагают:

$$\begin{aligned} J_i &= [x_{(1)} + (i-1)\Delta; x_{(1)} + i\Delta], \quad i = \overline{1, m-1}, \\ J_m &= [x_{(1)} + (m-1)\Delta; x_{(n)}]. \end{aligned} \tag{2.7}$$

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

J_1	...	J_i	...	J_m
n_1	...	n_i	...	n_m

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

Количество интервалов определяется формулой:

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

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

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

$$f(x) = \begin{cases} \frac{n_i}{n\Delta}, & x \in J_i, \ i = \overline{1, m}, \\ 0, & \text{иначе.} \end{cases} \tag{2.9}$$

Определение. 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.

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

$$F_n(t) = \frac{n(t, \vec{x})}{n}. \tag{2.10}$$

3 Практическая часть

3.1 Текст программы

```
pkg load statistics

x = [11.89,9.60,9.29,10.06,9.50,8.93,9.58,6.81,8.69,9.62,...
     9.01,10.59,10.50,11.53,9.94,8.84,8.91,6.90,9.76,7.09,...
     11.29,11.25,10.84,10.76,7.42,8.49,10.10,8.79,11.87,8.77,...
     9.43,12.41,9.75,8.53,9.72,9.45,7.20,9.23,8.93,9.15,...
     10.19,9.57,11.09,9.97,8.81,10.73,9.57,8.53,9.21,10.08,...
     9.10,11.03,10.10,9.47,9.72,9.60,8.21,7.78,10.21,8.99,...
     9.14,8.60,9.14,10.95,9.33,9.98,9.09,10.35,8.61,9.35,...
     10.04,7.85,9.64,9.99,9.65,10.89,9.08,8.60,7.56,9.27,...
     10.33,10.09,8.51,9.86,9.24,9.63,8.67,8.85,11.57,9.85,...
     9.27,9.69,10.90,8.84,11.10,8.19,9.26,9.93,10.15,8.42,...
     9.36,9.93,9.11,9.07,7.21,8.22,9.08,8.88,8.71,9.93,...
     12.04,10.41,10.80,7.17,9.00,9.46,10.42,10.43,8.38,9.01];

n = length(x);

fprintf('Задание а\n');

Mmax = max(x);
Mmin = min(x);

fprintf('M_max = %f\n', Mmax);
fprintf('M_min = %f\n\n', Mmin);

fprintf('Задание б\n');

R = Mmax - Mmin;

fprintf('R = %f\n\n', R)

fprintf('Задание в\n');

mu = sum(x) / n;
sqrS = sum((x - mu) .^ 2) / (n - 1);

fprintf('mu = %f\n', mu);
fprintf('S^2 = %f\n\n', sqrS);

fprintf('Задание г\n');

m = floor(log2(n)) + 2;

```

```
Delta = (Mmax - Mmin) / m;

limit_points = linspace(Mmin, Mmax, m + 1);

fprintf('m = %d\n', m);
fprintf('delta = %f\n', Delta);

nums = zeros(m);

for i = 1:m
  left = limit_points(i);
  right = limit_points(i + 1);

  for j = 1:n
    if x(j) >= left && x(j) < right
      nums(i)++;
    endif
  endfor

endfor

nums(m)++;

for i = 1:m
  fprintf('Interval [%f, %f%c (%d elements).\n', limit_points(i),...
         limit_points(i + 1), ifelse(i == m, ']', ')'), nums(i));
endfor

sigma = sqrt(sqrS);

x_vals = (Mmin:1e-3:Mmax);
f_theor = normpdf(x_vals, mu, sigma);

h = figure();
set(h,'numbertitle','off')
set(h, 'name', 'Задание д');
hold on;
plot(x_vals, f_theor, 'LineWidth', 10);

F_theor = normcdf(x_vals, mu, sigma);

g = figure();
set(g,'numbertitle','off')
set(g, 'name', 'Задание е');
plot(x_vals, F_theor, 'LineWidth', 10);
```

3.2 Результаты расчетов для выборки из индивидуального варианта

```
Задание а
M_max = 12.410000
M_min = 6.810000

Задание б
R = 5.600000

Задание в
mu = 9.487167
S^2 = 1.217306

Задание г
m = 8
delta = 0.700000
Interval [6.810000, 7.510000) (7 elements).
Interval [7.510000, 8.210000) (4 elements).
Interval [8.210000, 8.910000) (21 elements).
Interval [8.910000, 9.610000) (36 elements).
Interval [9.610000, 10.310000) (27 elements).
Interval [10.310000, 11.010000) (14 elements).
Interval [11.010000, 11.710000) (7 elements).
Interval [11.710000, 12.410000] (4 elements).
```