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

Дисциплина Математическая статистика

Тема _Гистограмма и эмпирическая функция распределения_

Студент _Ильясов И. М._

Группа _ИУ7-63Б_

Вариант _9_

Оценка (баллы) _______________

Преподаватель _Власов П. А._

Москва, 2020 г.

Формулы для вычисления величин

Приведенные ниже формулы применяются для выборки из генеральной совокупности $\vec{x} = (x_1, \dots, x_n)$.

Ниже приведены формулы для вычисления следующих величин – M_{max}, M_{min}, размаха выборки R, оценки математического ожидания $\hat{\mu}$, оценки дисперсии S^2.

1. **Максимальное значение выборки M_{max}**

$$M_{max} = \max\{x_1, \dots, x_n\} \tag{1}$$

2. **Минимальное значение выборки M_{min}**

$$M_{min} = \min\{x_1, \dots, x_n\} \tag{2}$$

3. **Размах выборки R**

$$R = M_{max} - M_{min} \tag{3}$$

4. **Оценка математического ожидания выборки $\hat{\mu}$**

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \tag{4}$$

5. **Оценка дисперсии выборки S^2**

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{5}$$

Определение эмпирической плотности и гистограммы

Эмпирической плотностью называют функцию

$$\widehat{f_n}(x) = \begin{cases} \frac{n_i}{n*\Delta}, x \in J_i, i = \overline{1; m} \\ 0, \text{иначе} \end{cases}, \text{ где} \tag{6}$$

- m – количество полуинтервалов интервала $J = \left[x_{(1)}, x_{(n)}\right]$;
- $J_i, i = \overline{1; m}$ – полуинтервал из $J = \left[x_{(1)}, x_{(n)}\right]$, где

$$x_{(1)} = \min\{x_1, \dots, x_n\}, \qquad x_{(n)} = \max\{x_1, \dots, x_n\}; \tag{7}$$

при этом все полуинтервалы равновеликие и все, кроме последнего, не содержат правую границу, т.е.

$$J_i = \left[x_{(1)} + (i-1) * \Delta, x_{(i)} + i * \Delta\right), i = \overline{1, m-1}; \tag{8}$$

$$J_m = \left[x_{(1)} + (m-1) * \Delta, x_{(1)} + m * \Delta\right); \tag{9}$$

- Δ – длина полуинтервала $J_i, i = \overline{1, m}$ равная

$$\Delta = \frac{x_{(n)} - x_{(1)}}{m} = \frac{|J|}{m}; \tag{10}$$

- n_i – количество элементов выборки в полуинтервале $J_i, i = \overline{1; m}$;
- n – количество элементов в выборке.

В рамках данной лабораторной работы $m = [\log_2 n] + 2$.

Гистограммой называют график эмпирической плотности.

Определение эмпирической функции распределения

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

Эмпирической функцией распределения называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определенную условием

$$F_n(x) = \frac{n(x, \vec{x})}{n}. \tag{11}$$

Текст программы

```
function lab1()
    % Выборка объема n из генеральной совокупности X
    X = [-8.47,-7.45,-7.12,-8.30,-8.15,-6.01,-5.20,-7.38,-6.76,-9.18,-6.00,-8.08, ...
         -7.96,-8.34,-6.82,-8.46,-8.07,-7.04,-7.24,-8.16,-8.20,-8.27,-7.79,-7.37, ...
         -7.02,-7.13,-6.99,-7.65,-8.18,-6.71,-8.41,-6.71,-7.04,-9.15,-7.74,-10.11, ...
         -8.20,-7.07,-7.63,-8.99,-6.62,-6.23,-7.13,-6.41,-7.06,-7.72,-8.44,-8.85, ...
         -8.02,-6.98,-6.08,-7.20,-7.48,-7.82,-9.19,-8.31,-7.95,-7.97,-6.66,-6.59, ...
         -9.10,-7.87,-9.02,-8.77,-7.62,-9.44,-8.05,-7.60,-7.33,-6.94,-8.51,-7.39, ...
         -6.44,-8.88,-8.21,-7.66,-6.91,-8.39,-7.37,-7.26,-6.04,-7.58,-7.28,-7.02, ...
         -7.10,-7.33,-8.63,-8.21,-7.12,-8.11,-9.03,-8.11,-8.79,-9.22,-7.32,-5.97, ...
         -7.26,-6.39,-7.64,-8.38,-7.67,-7.70,-7.70,-8.95,-6.25,-8.09,-7.85,-8.10, ...
         -7.73,-6.78,-7.78,-8.20,-8.88,-8.51,-7.45,-7.14,-6.63,-7.38,-7.72,-6.25];


    % Поиск минимального значения выборки
    min_value = CountMin(X);

    % Поиск максимального значения выборки
    max_value = CountMax(X);

    % Поиск размаха выборки
```

```
    delta = max_value - min_value;

    % Поиск оценки мат. ожидания выборки
    m = mean(X);

    % Поиск оценки дисперсии выборки
    d = CountD(X);

    % Вывод найденных параметров
    fprintf("Min = %.3f\n", min_value);
    fprintf("Max = %.3f\n", max_value);
    fprintf("R    = %.3f\n", delta);
    fprintf("M    = %.3f\n", m);
    fprintf("D    = %.3f\n", d);

    % Группировка значений по интервалам
    CountInters(X);

    % Отрисовка графиков
    DrawHist(X);
    DrawDistribution(X);
end

% Функция для вычисления оценки дисперсии выборки
function d = CountD(X)
    d = sum((X - mean(X)) .^ 2) / (length(X) - 1);
    return
end

% Функция для группировки значений в m интервалов
% и вывода на экран количества элементов в каждом интервале
function CountInters(X)
    % Поиск количества интервалов
    m = floor(log2(length(X))) + 2;

    % Разбиение выборки на m интервалов и поиск количества элементов в
    % каждом из интервалов
    step = (max(X) - min(X)) / m;
    dots = CountMin(X) : step : CountMax(X);

    fprintf("%d intervals:\n", m);

    for i = 1:(length(dots) - 1)
        count = 0;
        for x = X
            % Последний интервал включает в себя крайнее правое значение
            if (i == length(dots) - 1) && (x >= dots(i)) && (x <= dots(i + 1))
                count = count + 1;
            % Остальные интервалы включают только слева
            elseif (x >= dots(i)) && (x < dots(i + 1))
                count = count + 1;
            end
        end

        OutputString = "[%.3f; %.3f) -> %d\n";

        if (i == length(dots) - 1)
            OutputString = "[%.3f; %.3f] -> %d\n";
        end

        fprintf(OutputString, dots(i), dots(i + 1), count);
    end
end
% Функция для отрисовки гистограммы и графика функции плотности
% распределения
function DrawHist(X)
    x = sort(X);
    m = floor(log2(length(X))) + 2;

    subplot(2, 1, 1);
```

```
    % Гистограмма
    histogram(X, m, "Normalization", "pdf", "BinLimits", [CountMin(X), CountMax(X)]);
    hold on;

    % График функции плотности распределения
    f_1 = normpdf(x, mean(x), sqrt(CountD(x)));
    p_1 = plot(x, f_1);
    p_1.LineWidth = 2;
    hold off;
    legend(["Гистограмма", "Функция плотности распределения"], "Location", "northwest");
end

% Функция для отрисовки графиков функции распределения
function DrawDistribution(X)
    x = sort(X);

    subplot(2, 1, 2);

    % Эмпирическая функция распределения
    histogram(X, length(X), "Normalization", "cdf", "BinLimits", [CountMin(X),
                                                               CountMax(X)]);
    hold on;

    % Функция распределения нормальной случайной величины
    f_2 = normcdf(x, mean(x), sqrt(CountD(x)));
    p_2 = plot(x, f_2);
    p_2.LineWidth = 2;
    hold off;
    legend(["Эмпирическая функция ", "Функция распределения нормальной случайной величины"],
                                                        "Location", "northwest");
end

% Функция для вычисления максимального значения выборки
function max_value = CountMax(X)
    max_value = X(1);
    for i = 1:(length(X))
        if (max_value < X(i))
            max_value = X(i)
        end
    end
end

% Функция для вычисления минимального значения выборки
function min_value = CountMin(X)
    min_value = X(1);
    for i = 1:(length(X))
        if (min_value > X(i))
            min_value = X(i)
        end
    end
end
```

Результаты расчетов для выборки из индивидуального варианта (вариант №9)

1. Максимальное значение выборки $M_{max} = -5.2$
2. Минимальное значение выборки $M_{min} = -10.11$
3. Размах выборки $R = 4.91$
4. Оценка математического ожидания выборки $\hat{\mu} = -7.661$
5. Оценка дисперсии выборки $S^2 = 0.778$
6. Группировка значений выборки в $p = m = [\log_2 n] + 2$

```
8 intervals:
[-10.110; -9.496) -> 1
[-9.496; -8.883) -> 10
[-8.883; -8.269) -> 18
[-8.269; -7.655) -> 32
[-7.655; -7.041) -> 30
[-7.041; -6.428) -> 18
[-6.428; -5.814) -> 10
[-5.814; -5.200] -> 1
```

7. Гистограмма, график функции плотности распределения вероятностей, график эмпирической функции распределения, график функции распределения нормальной случайной величины