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 Сушина А.Д.

Группа ИУ7-61б

Оценка (баллы) _______________

Преподаватель Саркисян П.С.

Москва.
2020 г

Оглавление

1 Постановка задачи

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
   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 Аналитическая часть

2.1 Формулы для вычисления величин

Реализация случайной выборки:

$(x_{1,} x_{2,} \ldots , x_n)$ (1)

Максимальное значение выборки:

$M_{max} = max(x_{1,} x_{2,} \ldots x_n)$ (2)

Минимальное значение выборки:

$M_{min} = min(x_{1,} x_{2,} \ldots x_n)$ (3)

Размах выборки:

$R = M_{max} - M_{min}$ (4)

Оценка математического ожидания:

$\hat{\mu}(\vec{X}) = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ (5)

Несмещенная оценка дисперсии:

$S^2(\vec{X}) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ (6)

2.2 Определение эмпирической плотности и гистограммы

Определение 2.1 Эмпирической плотностью распределения случайной выборки $\vec{X}_n$ называют функцию

$f(x) = \begin{cases} \frac{n_i}{n * \Delta}, если\, x \in J_i \\ 0, иначе \end{cases}$ (7)

где

m — количество полуинтервалов интервала $J = [X_{(1)}; X_{(n)}]$;

n — количество элементов в выборке;

Δ — длина полуинтервала:

$\Delta = \frac{X_{(n)} - X_{(1)}}{m} = \frac{|J|}{m}$; (8)

n_i — число элементов выборки принадлежащих $J_i, i = \overline{1, m}$;

$J_i, i=\overline{1,m}$ — полуинтервал из $J=[X_{(1)}; X_{(n)}]$, где

$x_{(1)}=min\{x_1,\dots,x_n\}$, $x_{(n)}=max\{x_1,\dots,x_n\}$;(9)

при этом

J_i = [$J_i=X_{(min)}+(i-1)\Delta; X_{(min)}+i\Delta$), $i=\overline{1,m-1}$; (9)

$J_m=[X_{(min)}+(m-1)\Delta; X_{(min)}+m\Delta]$. (10)

Определение 2.2 График функции $f_n(x)$ называют гистограммой.

2.3 Определение эмпирической функции распределения

Определение 2.3 Эмперической функцией распределения называют функцию

$$F_n: R\rightarrow R, F_n(x)=\frac{n(x,\vec{x})}{n} \quad (11)$$

где $n(x,\vec{x})$ - количество элементов выборки, которые меньше x, а n — объем выборки.

Если все элементы этой выборки попарно различны, то

$$F_n(x)=\left\{\begin{array}{c} 0, x\leq x_{(1)} \\ \frac{i}{n}, x_{(i)}<x\leq X_{(i+1)} \\ 1, x>x_{(n)} \end{array}\right\} \quad (12)$$

3 Технологическая часть

3.1 Текст программы

Листинг 1. Текст программы lab1.m

```
function lab1()
    clear all;
    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);

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

    fprintf('Mmin = %s\n', num2str(Mmin));
    fprintf('Mmax = %s\n', num2str(Mmax));

    R = Mmax - Mmin;
    fprintf('R = %s\n', num2str(R));

    MU = getMU(X);
    fprintf('MU = %s\n', num2str(MU));

    Ssqr = getSsqr(X);
    fprintf('S^2 = %s\n', num2str(Ssqr));

    m = getNumberOfIntervals(X);
    fprintf('m = %s\n', num2str(m))

    createGroup(X);
    hold on;
    distributionDensity(X, MU, Ssqr, m);

    figure;
    empiricF(X);
    hold on;
    distribution(X, MU, Ssqr, m);
end

function mu = getMU(X)
    n = length(X);
```

```
    mu = sum(X)/n;
end

function Ssqr = getSsqr(X)
    n = length(X);
    MX = getMU(X);
    Ssqr = sum((X - MX).^2) / (n-1);
end

function m = getNumberOfIntervals(X)
    m = floor(log2(length(X)) + 2);
end

function createGroup(X)
    n = length(X);
    m = getNumberOfIntervals(X);

    intervals = zeros(1, m+1);
    numCount = zeros(1, m+1);
    Delta = (max(X) - min(X)) / m;
    fprintf('Delta = %s\n', num2str(Delta));

    for i = 0: m
        intervals(i+1) = X(1) + Delta * i;
    end

    j = 1;
    count = 0;
    for i = 1:n
        if (X(i) >= intervals(j+1))
            j = j + 1;
        end
        numCount(j) = numCount(j) + 1;
        count = count + 1;
    end

    for i = 1:m-1
        fprintf('[%5.2f; %5.2f) ', intervals (i), intervals(i+1));
    end
    fprintf('[%5.2f, %5.2f]\n', intervals(m), intervals(m+1));

    for i = 1:m
        fprintf('%8d       ', numCount(i));
    end
    fprintf('\n\n');

    graphBuf = numCount(1:m+1);
    for i = 1:m+1
        graphBuf(i) = numCount(i) / (n*Delta);
    end

    stairs(intervals, graphBuf),grid;
```

```
end

function distributionDensity(X, MX, DX, m)
    R = X(end) - X(1);
    delta = R/m;
    Sigma = sqrt(DX);

    Xn = (MX - R): delta/50 :(MX + R);
    Y = normpdf(Xn, MX, Sigma);
    plot(Xn, Y), grid;
end

function distribution(X, MX, DX, m)
    R = X(end) - X(1);
    delta = R/m;

    Xn = (MX - R): delta :(MX + R);
    Y = 1/2 * (1 + erf((Xn - MX) / sqrt(2*DX)));
    plot(Xn, Y, 'r'), grid;
end

function empiricF(X)
    [yy, xx] = ecdf(X);

    stairs(xx, yy), grid;
end
```

4 Экспериментальная часть

4.1 Результаты расчетов для выборки из индивидуального варианта

M_{min}	11.74
M_{max}	18.25
R	6.51
$\hat{\mu}(\vec{X}_n)$	14.3492
$S^2(\vec{X}_n)$	1.2776
m	8
Δ	0.81375

Интервальная группировка значений выборки при m = 8:

[11.74; 12.55)	[12.55; 13.37)	[13.37; 14.18)	[14.18; 15.00)
4	21	27	37
[15.00; 15.81)	[15.81; 16.62)	[16.62; 17.44)	[17.44, 18.25]
18	10	2	1

```
Mmin = 11.74
Mmax = 18.25
R = 6.51
MU = 14.3492
S^2 = 1.2776
m = 8
Delta = 0.81375
[11.74; 12.55) [12.55; 13.37) [13.37; 14.18) [14.18; 15.00) [15.00; 15.81) [15.81; 16.62) [16.62; 17.44) [17.44, 18.25]
      4              21             27             37             18             10              2              1
```

Рис 1. Результат работы программы

Построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2 :

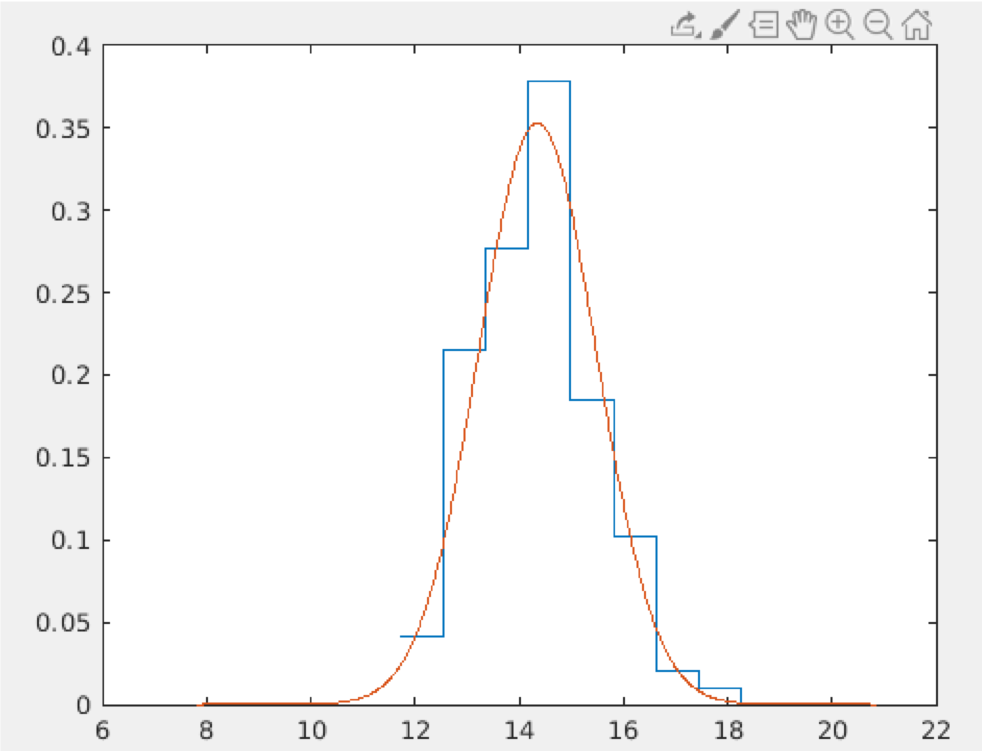

График 1

Построение на другой координатной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2 :

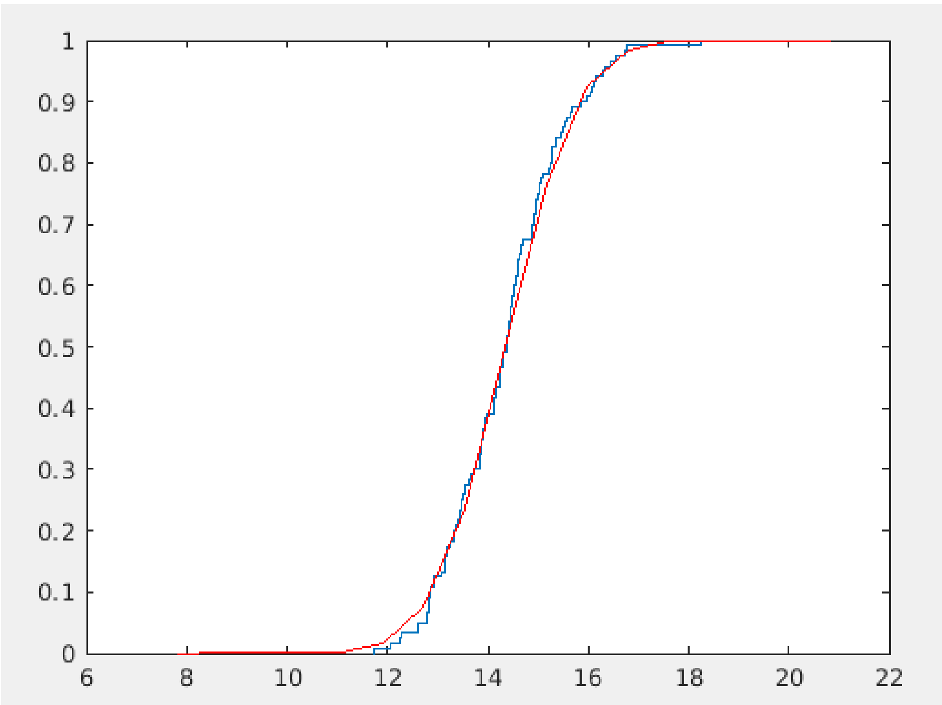

График 2