|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **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 предмету: «Математическая статистика»*

**Тема: Гистограмма и эмпирическая функция распределения**

*Вариант 25*

Студент: Юмаев Артур Русланович

Группа: ИУ7-65Б

# Цель и содержание работы

**Цель работы:** построение гистограммы и эмпирической функции распределения.

**Содержание работы:**

1. Для выборки объема из генеральной совокупности реализовать в виде программы на ЭВМ
   1. вычисление максимального значения и минимального значения
   2. размаха *R* выборки;
   3. вычисление и математического ожидания M*X* и дисперсии D*X*;
   4. группировку значений выборки в интервала;
   5. построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием и дисперсией ;
   6. построение на другой коордитной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием и дисперсией ;
2. Провести вычисления и построить графики для выборки из индивидуального варианта.

# Формулы для вычисления величин

Минимальное значение выборки:

, где – реализация случайной выборки.

Максимальное значение выборки:

, где – реализация случайной выборки.

Размах выборки:

Выборочным средним (выборочным математическим ожиданием) называется статистика:

Несмещенная оценка дисперсии:

# Эмпирическая плотность и гистограмма

При больших объемах выборки обычно производят группирование исходных данных следующим образом. Промежуток , содержащий все выборочные значения, разбивают на полуинтервалов , как правило, одинаковой длины и таких, что каждый из них, кроме последнего, содержит левую границу, а последний содержит обе границы, и подсчитывают число элементов выборки, попавших в -ый промежуток , а результаты представляют в виде слудующей таблицы 1, которую называют интервальным статистическим рядом.

количество полуинтервалов интервала

длина полуинтервала равная

количество элементов выборки в полуинтервале

количество элементов в выборке.

Таблица 1. Интервальный статистический ряд

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |

*Определение*. График функции , представляющий собой кусочно постоянную функцию называют ***гистограммой*** (рис. 1).

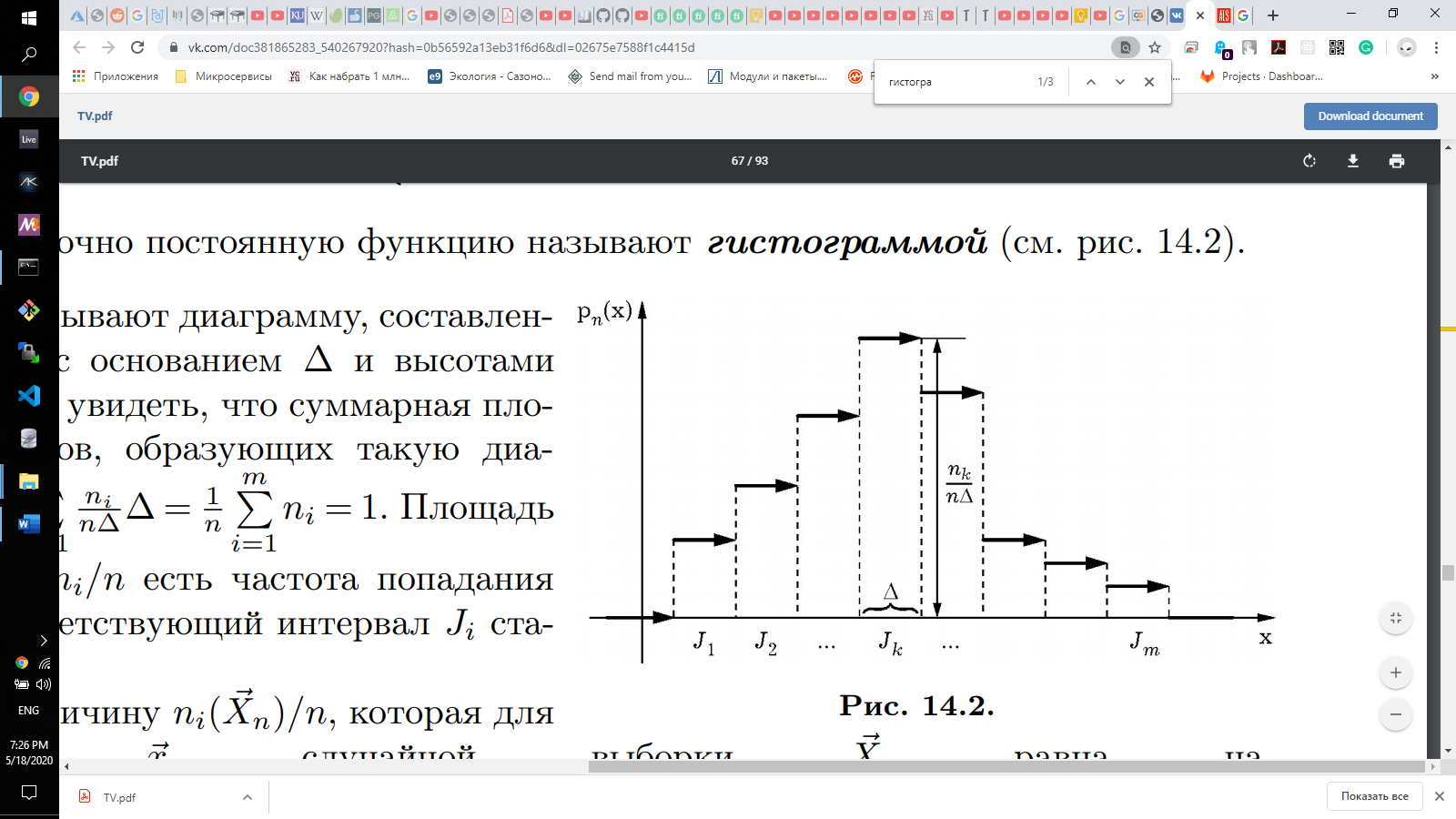


Рисунок 1. Гистограмма

Часто гистограммой называют диаграмму, составленную из прямоугольников с основанием и высотами Нетрудно увидеть, что суммарная площадь всех прямоугольников, образующих такую диаграмму, равна 1, так как:

Площадь каждого прямоугольника есть частота попадания элементов выборки в соответсвующий интервал статистического ряда.

Пусть

1. случайная выборка,
2. реализация случайной выборки,
3. количество элементов выборки , которые имеют значения меньше *x.*

*Определение.* Эмпирической функцией распределения называют функцию

, определенную условием

# Листинг

Листинг 1. Программа на Matlab к лабораторной работе №1

|  |
| --- |
| function lab1  clear  X=[-17.04,-18.29,-17.38,-18.11,-18.96,-17.65,-17.02,-17.22,-16.25,-17.44,-17.69,-17.61,-17.09,-17.19,-16.02,-17.56,-16.94,-17.29,-16.93,-16.61,-19.38,-17.53,-16.39,-17.89,-17.98,-17.04,-16.22,-19.09,-18.91,-17.77,-18.30,-17.44,-18.84,-16.39,-16.13,-18.37,-16.37,-16.70,-17.78,-17.03,-17.76,-17.87,-17.20,-18.44,-17.19,-17.75,-16.81,-17.97,-18.03,-16.87,-16.10,-19.16,-16.51,-18.39,-16.48,-18.08,-17.49,-18.89,-19.09,-17.96,-18.40,-16.96,-18.15,-18.71,-17.81,-17.86,-19.47,-17.86,-17.60,-17.30,-17.60,-17.71,-18.42,-16.88,-16.76,-18.00,-17.97,-16.83,-18.00,-18.08,-17.61,-17.02,-16.73,-17.64,-18.76,-17.68,-18.04,-16.45,-18.79,-18.03,-17.38,-15.27,-15.97,-17.41,-18.61,-18.00,-17.42,-17.77,-19.05,-16.16,-16.27,-18.00,-18.90,-17.05,-17.46,-17.49,-18.20,-17.59,-15.78,-18.88,-18.53,-17.39,-17.83,-18.17,-16.15,-17.66,-17.76,-18.32,-17.70,-17.56];    % Пункт a)  Mmax = max(X);  Mmin = min(X);    % Пункт б)  R = Mmax - Mmin;    % Пункт в)  mu = find\_mu(X);  S2 = find\_S2(X);    % Пункт г)  % Гистограмма  m = find\_m(X);  [J, count] = intervalization(X, m);  stairs([J(1) J], [0 count 0]);  % График 1  hold on;  Y1 = f(X, mu, S2);  plot(X, Y1, '.');  grid on;  legend('Гистограмма','Функция плотности распределения нормальной СВ');  hold off;  % График 2  figure;  empF(sort(X));  hold on;  F(sort(X), mu, getS(X), m, R);  grid on;  legend('Эмпирическая функция распределения','Функция распределения нормальной СВ');  hold off;   function [mu] = find\_mu(X)  mu = sum(X)/size(X,2);  end   function [S2] = find\_S2(X)  S2 = sum((X - find\_mu(X)) .\* (X - find\_mu(X)))/ (size(X,2) - 1);  end   function sigma = getSigmaSqr(X)  tempMu = find\_mu(X);  sigma = sum((X - tempMu) .\* (X - tempMu)) / size(X,2);  end   function Ssqr = getS(X)  n = size(X,2);  Ssqr = n/ (n - 1) \* getSigmaSqr(X);  end   function [m] = find\_m(X)  m = floor(log2(size(X, 2))) + 2;  end   function [J, count] = intervalization(X, m)  sortX = sort(X);  n = size(sortX,2);  delta = (sortX(end) - sortX(1)) / m;  J = sortX(1):delta:sortX(end);  count = zeros(1,m);   for i = 1:(size(J,2)-1)  for j = 1:n  if (sortX(j) >= J(i) && sortX(j) < J(i+1))  count(i) = count(i) + 1;  end  end  end   count(end) = count(end) + sum(sortX==(sortX(end)));   for i = 1:size(count,2)  count(i) = count(i)/(n \* delta);  end   end   function [Y] = f(X, MX, DX)  Y = normpdf(X, MX, sqrt(DX));  end   function empF(X)  [yy, xx] = ecdf(X);  stairs(xx, yy);  end    function F(X, MX, DX, m, R)  delta = R/m;  Xn = min(X):delta/20:max(X);  Y = 1/2 \* (1 + erf((Xn - MX) / sqrt(2\*DX)));  plot(Xn, Y, '--');  end end |

# Результат работы программы

Таблица 1. Интервальная группировка значений выборки при m = 8

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| [-19.470,  -18.945) | [-18.945,  -18.419) | [-18.410,  -17.895) | [-17.895,  -17.369) | [17.369,  -16.845) | [-16.845,  -16.320) | [-16.320,  -15.795) | [15.795,  -15.270] |
| 7 | 12 | 23 | 37 | 18 | 12 | 9 | 2 |

# Графики

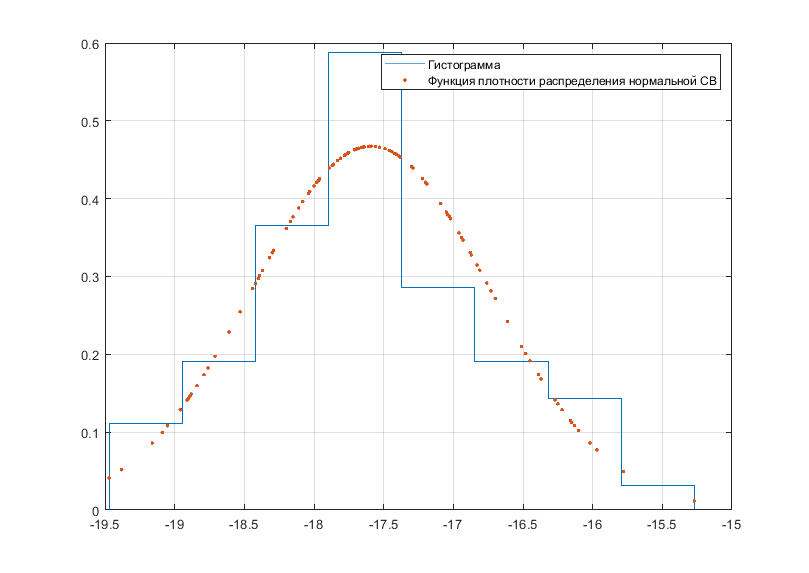


Рисунок 2. Гистограмма и график функции плотности распределения нормальной случайной

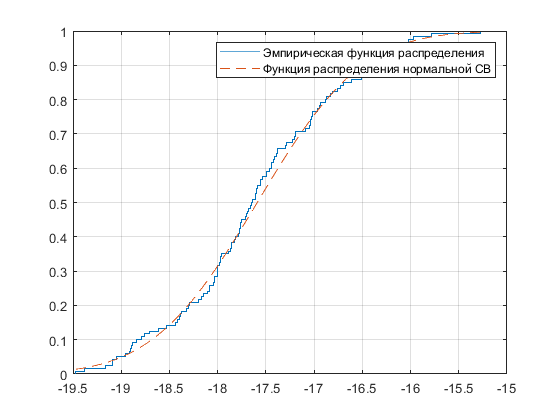


Рисунок 3. График эмпирической функции распределения и функции распределения нормального распределения