

Цель работы: построение гистограммы и эмпирической функции распределения.

Содержание работы

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

1 Вариант 21, выборка

$\vec{x} = ($
-14.34, -16.97, -14.09, -14.74, -16.69, -13.85, -15.55, -14.62, -13.30, -15.52,
-14.75, -16.51, -17.15, -16.87, -15.06, -13.60, -14.48, -14.71, -14.17, -13.88,
-14.55, -15.37, -14.81, -16.05, -17.06, -15.86, -15.12, -15.98, -14.16, -15.81,
-15.06, -16.19, -16.22, -16.19, -14.87, -15.62, -15.86, -15.25, -16.34, -14.44,
-14.72, -15.17, -15.24, -14.44, -15.93, -14.87, -16.53, -15.76, -15.12, -12.91,
-16.06, -16.06, -14.89, -15.57, -13.59, -16.84, -13.88, -14.33, -15.45, -16.58,
-16.05, -14.34, -13.55, -16.78, -14.15, -14.28, -14.40, -13.98, -16.23, -15.35,
-14.77, -15.61, -15.59, -15.64, -14.76, -17.18, -15.13, -15.01, -14.21, -13.91,
-16.55, -15.44, -14.03, -16.44, -15.57, -15.07, -16.28, -16.30, -15.74, -14.03,
-14.85, -15.73, -15.81, -14.42, -14.14, -15.14, -15.49, -16.42, -14.22, -14.20,
-17.17, -15.82, -14.96, -14.75, -14.98, -13.64, -14.00, -17.29, -14.51, -16.18,
-15.70, -15.07, -14.28, -14.55, -13.85, -15.36, -15.74, -14.61, -16.32, -15.34
)

2 Формулы для вычисления величин

Пусть $\vec{x} = (x_1, \dots, x_n)$ - выборка из генеральной совокупности X .

1. Максимальное значение выборки $M_{max} = \max(x_1, \dots, x_n)$
2. Минимальное значение выборки $M_{min} = \min(x_1, \dots, x_n)$
3. Размах выборки $R = M_{max} - M_{min}$
4. Выборочное среднее (математическое ожидание) $\hat{\mu}(\vec{x}) = \frac{1}{n} \sum_{i=1}^n x_i$
5. Состоятельная оценка дисперсии $S^2(\vec{x}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

3 Определение эмпирической плотности и гистограмм

Эмпирической плотностью (отвечающей выборке \vec{x}) называют функцию

$$\hat{f}(x) = \begin{cases} \frac{n_i}{n\Delta}, & x \in J_i, i = \overline{1; p} \\ 0, & \text{иначе} \end{cases}$$

где J_i – полуинтервал статистического ряда, n_i – количество элементов выборки, входящих в полуинтервал, n – количество элементов выборки.

Пусть \vec{x} – выборка из генеральной совокупности X . Если объем n этой выборки велик, то значения x_i группируют не только в статистический ряд, но и в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ (где $x_{(1)} = \min(x_1, \dots, x_n)$, $x_{(n)} = \max(x_1, \dots, x_n)$) делят на m равновеликих частей:

$$J_i = [a_i, a_{i+1}), i = \overline{1; m-1}$$

$$J_m = [a_m, a_{m+1}]$$

$$a_i = x_{(1)} + (i - 1) \cdot \Delta, i = \overline{1; m + 1}$$

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

Интервальным статистическим рядом называют таблицу

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

Здесь n_i – количество элементов выборки \vec{x} , которые $\in J_i$

Требуемый интервал $m = \lceil \log_2 n \rceil + 2$

Гистограмма – это график эмпирической плотности.

4 Определение эмпирической функции распределения

Пусть $\vec{x} = (x_1, \dots, x_n)$ – выборка из генеральной совокупности X . Обозначим $n(x, \vec{x})$ – число элементов вектора \vec{x} , которые имеют значения меньше x .

Эмпирической функцией распределения называют функцию $F_n : \mathbb{R} \rightarrow \mathbb{R}$, определенную условием $F_n(x) = \frac{n(x, \vec{x})}{n}$.

5 Листинг

Листинг 1: Реализация

```

1 function main()
2     X=[-5.05,-5.74,-6.39,-5.01,-4.94,-6.32,-4.73,-5.44,-4.79,-5.40,...
3         -4.50,-3.43,-5.21,-5.22,-5.07,-5.51,-4.45,-5.24,-6.50,-4.99,...
4         -5.42,-3.30,-5.06,-5.38,-6.48,-3.31,-5.56,-6.50,-5.22,-5.68,...
5         -4.80,-6.67,-4.95,-5.32,-5.68,-6.32,-3.72,-4.59,-6.33,-5.03,...
6         -4.49,-4.80,-6.04,-6.21,-3.60,-3.93,-5.89,-5.29,-7.41,-3.73,...
7         -6.61,-4.34,-5.99,-5.24,-4.08,-4.68,-5.38,-6.38,-4.66,-3.67,...
8         -4.61,-4.54,-4.51,-5.43,-6.47,-5.31,-4.30,-6.32,-5.82,-3.44,...
9         -5.92,-4.76,-4.45,-3.52,-4.91,-5.65,-5.02,-5.00,-5.26,-4.98,...
10        -6.16,-6.21,-4.42,-6.20,-5.84,-5.58,-5.34,-5.21,-5.78,-7.80,...
11        -5.21,-4.79,-4.53,-4.78,-6.39,-7.04,-4.82,-5.53,-3.52,-6.24,...
12        -3.58,-5.01,-5.79,-4.80,-6.04,-5.15,-7.03,-4.71,-4.38,-5.77,...
13        -4.05,-5.76,-5.86,-6.45,-4.81,-5.68,-7.48,-3.97,-5.16,-3.48];
14
15 % 1)

```

```

16 % а) Максимальное и минимальное значения
17 Mmax = max(X);
18 Mmin = min(X);
19 fprintf("\na) Mmax (максимальное значение) = %f; Mmin (минимальное
    значение) = %f ", Mmax, Mmin);
20
21 % б) Размах
22 R = Mmax - Mmin;
23 fprintf("\nb) R (размах) = %f ", R);
24
25 % в) Оценки
26 mu = sum(X) / length(X);
27 s2 = sum((X - mean(X)).^2) / (length(X) - 1);
28 fprintf("\nv)mu (оценка математического ожидания) = %f; s^2 (оценк
    а дисперсии) = %f ", mu, s2);
29
30 % г) Группировка значений выборки
31 % Нахождение количества интервалов
32 m = floor(log2(length(X))) + 2;
33 fprintf("\ng)Группировка значений выборки в m = [log2 n] + 2 интер
    вала: m = %f\n", m);
34
35 % Разбиение выборки на интервалы от min до max, с помощью
    BinLimits
36 % объединяем только те значения, которые находятся в интервале от
    мин
37 % до макс
38 [counts, edges] = histcounts(X, m, 'BinLimits', [min(X), max(X)]);
39
40 for i = 1: length(counts)
41     fprintf("[%f : %f] - %d\n", edges(i), edges(i + 1), counts(i))
42     ;
43 end
44
45 % д) Построение гистограммы
46 hist = histogram();
47 hist.BinEdges = edges;
48 hist.BinCounts = counts / length(X) / ((max(X) - min(X)) / m);
49
50 hold on; % Продолжаем работать с той же системой
51
52 % График функции плотности распределения вероятностей нормальной сл
    учайно величины
53 delta = R/m;
54 sigma = sqrt(s2);
55 Xn = min(X):delta/20:max(X);
56 Y = normpdf(Xn, mu, sigma);
57 plot(Xn, Y, 'blue');
58
59 % е) График распределения сормальной случайной величины
60 figure;
[yy, xx] = ecdf(X);

```

```

61 stairs(xx, yy);
62
63 hold on;
64
65 % График эмпирической функции распределения нормальной случайной в
    еличины
66 delta = R/m;
67 Xn = min(X):delta/20:max(X);
68 Y = 1/2 * (1 + erf((Xn - mu) / sqrt(2*s2)));
69 plot(Xn, Y, 'black');
70
71 end

```

6 Результаты расчетов для выборки варианта 21

1. Максимальное значение $M_{max} =$,
2. Минимальное значение выборки $M_{min} =$,
3. Размах выборки $R =$,
4. Оценки: $\hat{\nu} =$, $S^2 =$,
5. Группировка значений выборки в $m = [\log_2 n] + 2$ интервала предоставлена на рисунке 1.

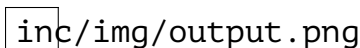
 inc/img/output.png

Рис. 1: Поток вывода программы.

6. Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2 .

 inc/img/outputGraph.png

Рис. 2: Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2 .

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

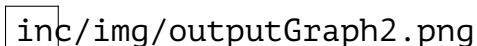
 inc/img/outputGraph2.png

Рис. 3: График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2 .