МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное автономное образовательное учреждение высшего образования

«Омский государственный технический университет»

Факультет информационных технологий и компьютерных систем

Кафедра «Прикладная математика и фундаментальная информатика»

Расчётно-графическая работа

по дисциплине Теория вероятностей и математическая статистика

Студента	
	фамилия, имя, отчество полностью
Курс	2 Группа ФИТ-231
Направление	02.03.02 Фундаментальная информатика и информационные технологии
	код, наименование
Руководитель	доц., канд. техн. наук, доцент
	должность, ученая степень, звание
	Болдовская Т.Е.
	фамилия, инициалы
Выполнил	
	дата, подпись студента
баллы	
	дата, подпись руководителя

Омск-2025

Оглавление

Введение

Современная статистика представляет собой мощный инструментарий для анализа случайных величин, позволяющий на основе ограниченной выборки делать выводы о свойствах всей генеральной совокупности. Одной из фундаментальных задач математической статистики является проверка гипотез о законе распределения случайной величины, что имеет важное значение для последующего моделирования и прогнозирования.

В данной лабораторной работе проводится комплексное исследование выборочных данных, включающее:

- Первичную обработку и систематизацию данных через построение вариационного и интервального рядов
- Визуализацию распределения с использованием гистограммы, полигона частот и эмпирической функции распределения
- Расчет основных числовых характеристик (моментов распределения)
- Выдвижение и проверку гипотезы о виде теоретического распределения
- Оценку параметров предполагаемого распределения
- Проверку адекватности модели с помощью критерия согласия Пирсона
- Построение доверительных интервалов для параметров распределения

Выборка

В соответствии с выданным преподавателем вариантом в данной работе будут производиться расчёты над данной выборкой:

18,5; 19,0; 20,6; 18,4; 18,3; 19,2 18,5; 20,3; 18,6; 20,6; 18,5; 18,3; 19,1; 21,0; 18,6; 19,1; 18,4; 19,9; 18,7; 19,5; 18,4; 21,6; 18,9; 19,5; 20,1; 19,4; 19,0; 19,2; 19,7; 19,9; 19,5; 20,0; 18,4; 18,3; 19,6; 18,8; 23,1; 19,6; 18,5; 20,7; 18,7; 18,7; 22,8; 18,9; 20,2; 19,0; 19,2; 19,6; 18,9; 20,3; 21,0; 18,9; 20,3; 18,3; 19,5; 18,5;18,5; 18,6; 19,3; 18,6

Выборка в отсортированном виде представляет собой вариационный ряд:

18.3, 18.3, 18.3, 18.3, 18.4, 18.4, 18.4, 18.4, 18.5, 18.5, 18.5, 18.5, 18.5, 18.5, 18.6, 18.6, 18.6, 18.6, 18.7, 18.7, 18.7, 18.8, 18.9, 18.9, 18.9, 18.9, 19.0, 19.0, 19.0, 19.1, 19.1, 19.2, 19.2, 19.2, 19.3, 19.4, 19.5, 19.5, 19.5, 19.5, 19.6, 19.6, 19.6, 19.7, 19.9, 19.9, 20.0, 20.1, 20.2, 20.3, 20.3, 20.3, 20.6, 20.6, 20.7, 21.0, 21.0, 21.6, 22.8, 23.1

Интервальный статистический ряд

Для расчёта воспользуемся формулой Стёрджесса:

$$h = \frac{R}{m} = \frac{x_{max} - x_{min}}{1 + 3.322 * \log_{10} n}$$

Где h – длина частичного интервала, R – размах выборки, m – число интервалов, n – объём выборки.

$$R = x_{max} - x_{min} = 23.1 - 18.3 = 4.8$$

$$m = 1 + 3.322 * \log_{10} n = 1 + 3.322 * \log_{10} 60 = 6.907 \approx 7$$

$$h = \frac{R}{m} = \frac{4.8}{7} = 0.686$$

По данной формуле оптимальное количество интервалов – 7 с шагом 0.686. В качестве начальной точке возьмём минимальный элемент выборки.

	Интервал	Фактические частоты	Относительные частоты
1	[18.3, 18.99)	26	0.433
2	[18.99, 19.67)	17	0.283
3	[19.67, 20.36)	9	0.15
4	[20.36, 21.04)	5	0.083
5	[21.04, 21.73)	1	0.0167
6	[21.73, 22.42)	0	0
7	[22.42, 23.10)	2	0.0333

Полигон и гистограмма относительных частот

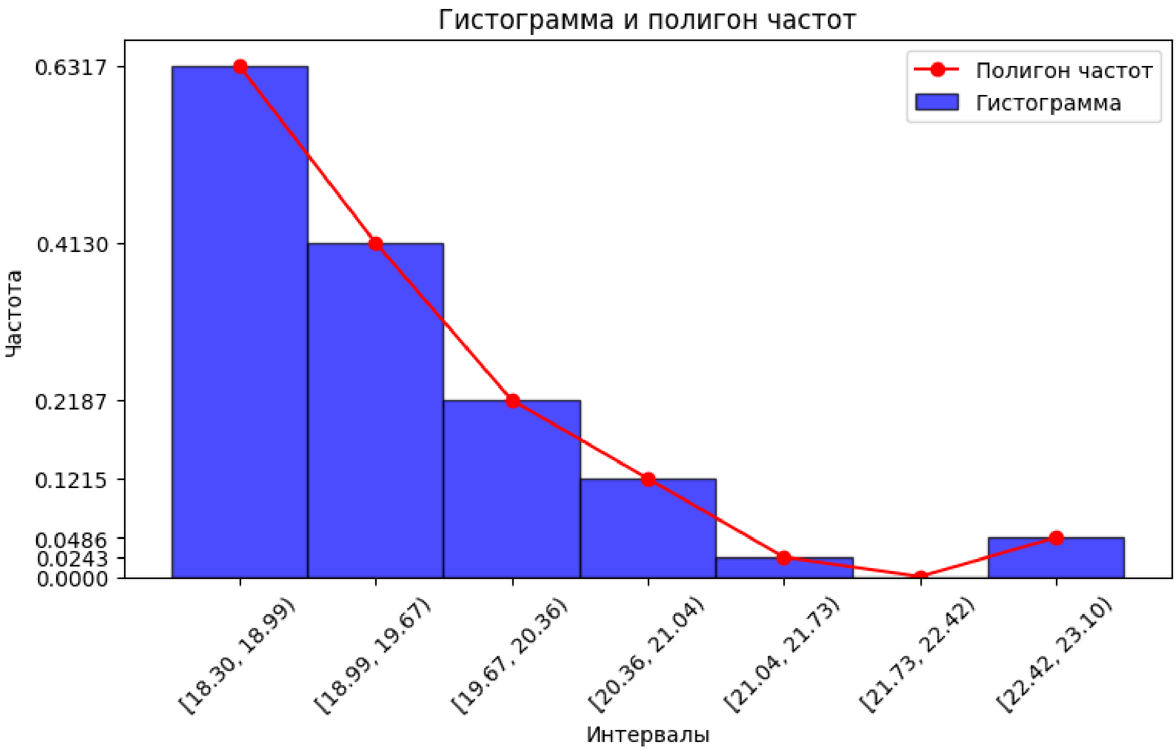


График эмпирической функции распределения

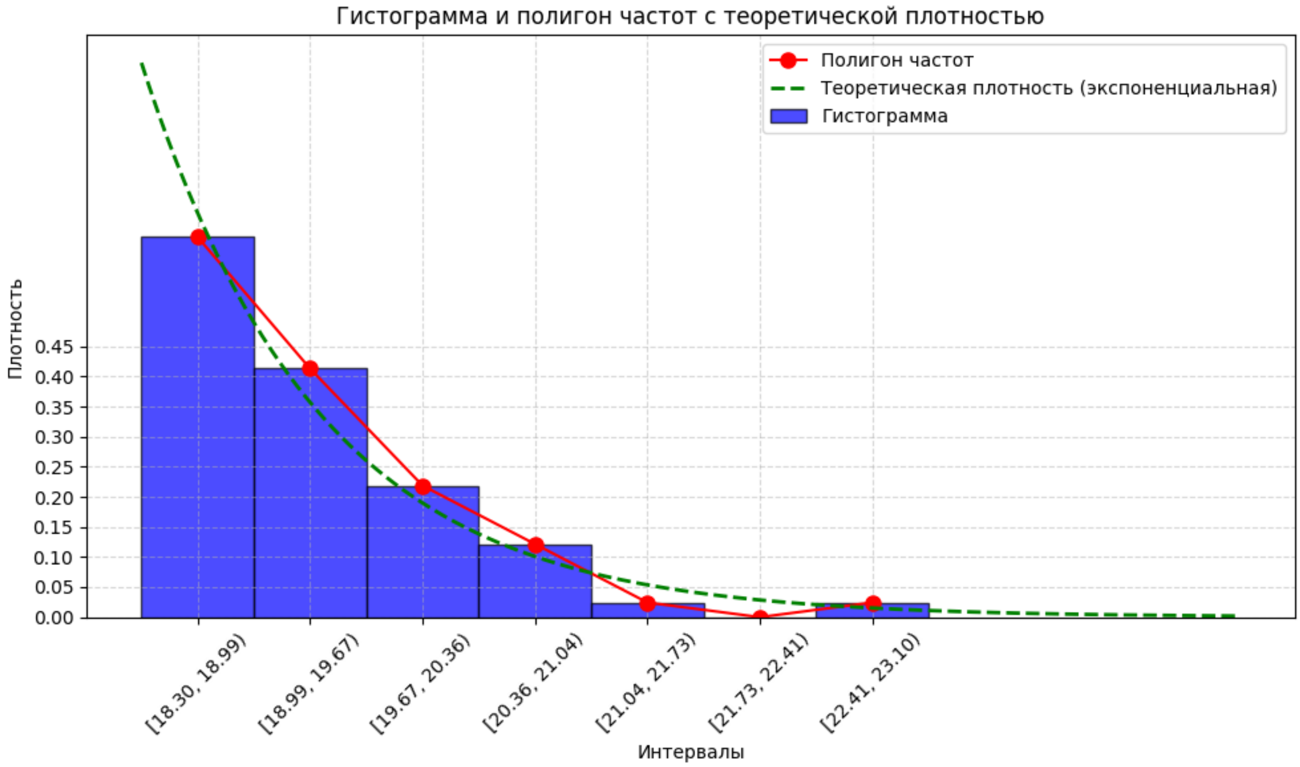
Числовые характеристики

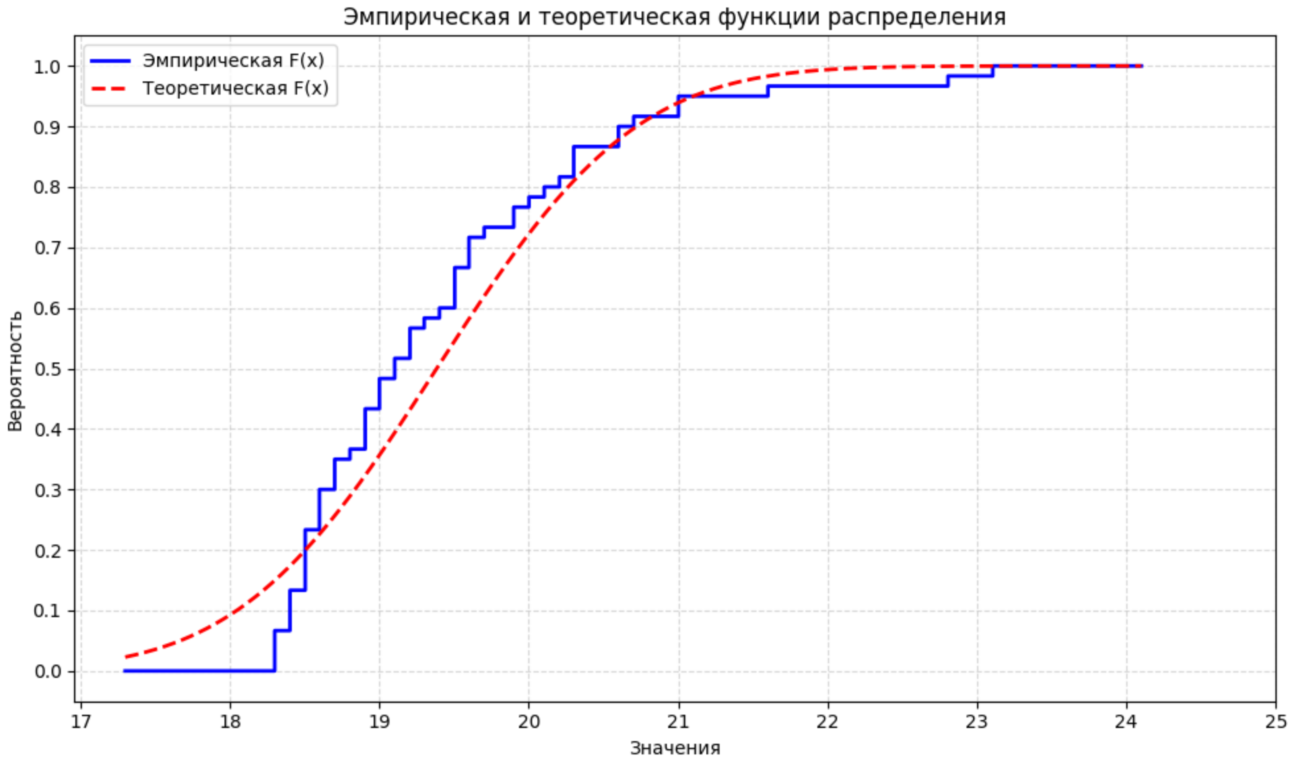
В этом разделе будут рассчитаны основные числовые характеристики выборки.

Характеристика	Формула	Значение
Выборочное среднее	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	19.38
Исправленная выборочная дисперсия	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	1.09
Мода	$m_o^* = x_o + h * \frac{n_{m_o} - n_{m_o-1}}{2n_{m_o} - n_{m_o-1} - n_{m_o+1}}$	18.81
Медиана	$m_e^* = x_e + h * \frac{\frac{n}{2} - (n_1 + \ldots + n_{m_e-1})}{n_{m_e}}$	19.15
Эксцесс	$\hat{E} = \frac{m_4}{S^4} - 3$	3
Коэффициент асимметрии	$\hat{A} = \frac{m_3}{S^3}$	1.59

Гипотеза о распределении

На основе анализа гистограммы и эмпирической функции распределения предполагаем, что генеральная совокупность имеет экспоненциальное распределение с параметрами, близкими к выборочному среднему ≈ 19.38 и исправленной дисперсии ≈ 1.09. Положительные значения асимметрии 1.59 и эксцесса 3 указывают на отклонение от симметрии.





Проверка правила трёх сигм

Из расчетов выше, среднее выборочное равняется 19.38, сигма равна $\sqrt{1.09} = 1.0441$. Тогда диапазон для проверки равняется [16.25, 22.52]. В интервал попадает 58 значений из 60. Правило выполняется приближённо

Проверка критерием Пирсона

Для теста возьмём интервальный статистический ряд. Рассчитаем теоретические частоты при помощи формулы

$$p_i = P(a_i < X < a_{i+1}) = \Phi\left(\frac{a_{i+1} - 19.38}{1.0441}\right) - \Phi\left(\frac{a_i - 19.38}{1.0441}\right)$$

Тогда:

Интервал	$\Phi\left(\frac{a_{i+1} - 19.38}{1.0441}\right) - \Phi\left(\frac{a_i - 19.38}{1.0441}\right)$	np_i
[18.3, 18.99)	$\Phi\left(\frac{18.99 - 19.38}{1.0441}\right) - \Phi\left(\frac{18.3 - 19.38}{1.0441}\right) = \Phi(-0.37) - \Phi(-1.03)$ $= -0.144 + 0.348 = 0.204$	12,24
[18.99, 19.67)	$\Phi\left(\frac{19.67 - 19.38}{1.0441}\right) - \Phi\left(\frac{18.99 - 19.38}{1.0441}\right) = \Phi(0.27) - \Phi(-0.37)$ $=$ $= 0.106 + 0.144 = 0.251$	15,06
[19.67, 20.36)	$\Phi\left(\frac{20.36 - 19.38}{1.0441}\right) - \Phi\left(\frac{19.67 - 19.38}{1.0441}\right) = \Phi(0.93) - \Phi(0.27) =$ $= 0.324 - 0.106 = 0.217$	13,02
[20.36, 21.04)	$\Phi\left(\frac{21.04 - 19.38}{1.0441}\right) - \Phi\left(\frac{20.36 - 19.38}{1.0441}\right) = \Phi(1.58) - \Phi(0.93) =$ $= 0.443 - 0.324 = 0.119$	7.14
[21.04, 21.73)	$\Phi\left(\frac{21.73 - 19.38}{1.0441}\right) - \Phi\left(\frac{21.04 - 19.38}{1.0441}\right) = \Phi(2.25) - \Phi(1.58) =$ $= 0.488 - 0.443 = 0.045$	2.7
[21.73, 22.42)	$\Phi\left(\frac{22.42 - 19.38}{1.0441}\right) - \Phi\left(\frac{21.73 - 19.38}{1.0441}\right) = \Phi(2.9) - \Phi(2.25) =$ $= 0.498 - 0.488 = 0.010$	0.6
[22.42, 23.10)	$\Phi\left(\frac{23.1 - 19.38}{1.0441}\right) - \Phi\left(\frac{22.42 - 19.38}{1.0441}\right) = \Phi(3.6) - \Phi(2.9) =$ $= 0.5 - 0.498 = 0.002$	0.12

И тогда теоретические частоты будут равны [12,24, 15,06, 13,02, 7,14, 2,7, 0.6, 0.12]. Фактические же наблюдаемые частоты - [26, 17, 9, 5, 1, 0, 2].

Теоретические частоты np_i должны быть больше 5, поэтому объединим 4, 5, 6 и 7 интервалы:

Интервал	n_i	np_i
[18.3, 18.99)	26	12,24
[18.99, 19.67)	17	15,06
[19.67, 20.36)	9	13,02
[20.36, 23.10)	8	10,56

Рассчитаем статистики хи-квадрат: $\chi^2 = \sum \frac{(n_i - np_i)^2}{np_i} \approx 17.6$

Определим число степеней свободы $df = k - r - 1 = k - 3 = 4 - 3 = 1.$

Для уровня значимости $\alpha = 0.05$ и $df = 1$: $\chi^2_{\text{крит}} = 3.841.$

$\chi^2 = 17.6 > 3.841 \Rightarrow$ Гипотеза о нормальном распределении отвергается

Доверительные интервалы

Для генерального среднего:

$$I_a = \left[\bar{X} - \frac{S}{\sqrt{n-1}} t_{n-1,1-\alpha}, \bar{X} + \frac{S}{\sqrt{n-1}} t_{n-1,1-\alpha}\right]$$

$S = \sqrt{1.09} = 1.044$

Степень свободы: 59

Критическое значение $t_{n-1,1-\alpha/2}$ при $\alpha = 0.05$:

$t_{59,0.975} = 2.001$

$$I_a = \left[19.38 - \frac{1.044}{\sqrt{59}} 2.001, 19.38 + \frac{1.044}{\sqrt{59}} 2.001\right] =$$

$$= [19.38 - 0.272, 19.38 + 0.272] = [19.11, 19.65]$$

Теперь для генерального среднеквадратичного отклонения

$$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}}, \quad \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}}\right)$$

Нижняя граница:

$$\sqrt{\frac{59 * 1.09}{\chi^2_{0.025,59}}} = \sqrt{\frac{59 * 1.09}{82.12}} \approx \sqrt{0.783} \approx 0.885$$

Верхняя граница:

$$\sqrt{\frac{59 * 1.09}{\chi^2_{0.975,59}}} = \sqrt{\frac{59 * 1.09}{39.66}} \approx \sqrt{1.622} \approx 1.273$$

Доверительный интервал:

$$[0.885;\ 1.273]$$

Заключение

Проведенное статистическое исследование позволило получить полную характеристику исследуемой выборки и сделать обоснованные выводы о законе ее распределения. На первом этапе были систематизированы исходные данные через построение вариационного и интервального рядов, что дало наглядное представление о структуре выборки.

Графические методы анализа (гистограмма, полигон частот, эмпирическая функция распределения) в сочетании с расчетом числовых характеристик позволили выдвинуть гипотезу о возможном нормальном распределении генеральной совокупности.

Дополнительный анализ показал, что:

- Правило "трех сигм" выполняется приближенно
- Доверительные интервалы для параметров распределения составили:
  - Для математического ожидания: $[19.11044; 19.64956]$
  - Для среднеквадратического отклонения: [0.878; 1.260]