Федеральное государственное образовательное бюджетное учреждение **высшего образования**

**«Финансовый университет при Правительстве Российской Федерации»**

**(Финансовый университет)**

Колледж информатики и программирования

**Лабораторная работа**

Тема «Анализ и визуализация данных о температуре»

|  |  |
| --- | --- |
|  | **Студентка:** Чемекова Валерия    **Учебная группа** 2ИСИП-822  **Преподаватель:**  \_\_\_\_\_\_\_\_\_\_\_\_\_\_/И.В.Сибирев/  **Дата выполнения:**  **10. 06.2024г.**    **Оценка за работу: \_\_\_\_\_\_\_\_\_\_\_\_** |
|  |  |

**Москва - 2024 г**

# Цель работы:

* Применить методы статистического анализа для изучения распределения данных о средней температуре июня в Москве за 40 лет.
* Визуализировать данные с использованием таблицы частот, гистограммы и полигона частот.

# Теоретические основы:

Формула Стерджеса для определения количества интервалов: 𝑘=1+3,3⋅𝑙𝑜𝑔2𝑛k=1+3,3⋅log2​n, где 𝑛n - объем выборки.

Ширина интервала: ℎ=𝑥𝑚𝑎𝑥−𝑥𝑚𝑖𝑛𝑘h=kxmax​−xmin​​, где 𝑥𝑚𝑎𝑥xmax​ и 𝑥𝑚𝑖𝑛xmin​ - максимальное и минимальное значения в выборке.

Минимальное значение нижнего интервала: 𝑎=[𝑥𝑚𝑖𝑛]−ℎ2a=[xmin​]−2h​, где [𝑥𝑚𝑖𝑛][xmin​] - целая часть 𝑥𝑚𝑖𝑛xmin​.

Таблица распределения частот, гистограмма, полигон частот и полигон накопленных частот являются базовыми инструментами для визуализации распределения данных.

Размах выборки: 𝑅=𝑥𝑚𝑎𝑥−𝑥𝑚𝑖𝑛R=xmax​−xmin​.

Среднее арифметическое: 𝑀=∑𝑖=1𝑛𝑥𝑖𝑛M=n∑i=1n​xi​​.

Среднее гармоническое: 𝐻=𝑛1𝑥1+...+1𝑥𝑛H=x1​1​+...+xn​1​n​.

Среднее геометрическое: 𝑄=𝑥1⋅...⋅𝑥𝑛𝑛Q=nx1​⋅...⋅xn​​.

Мода - наиболее часто встречающееся значение в выборке.

Медиана - значение, которое делит упорядоченную выборку на две равные части.

Математическое ожидание и дисперсия для выборки: 𝑀=∑𝑖=1𝑛𝑥𝑖𝑛M=n∑i=1n​xi​​, 𝐷=∑𝑖=1𝑛(𝑥𝑖−𝑀)2𝑛D=n∑i=1n​(xi​−M)2​.

Стандартное отклонение: 𝑠𝑖𝑔𝑚𝑎=𝐷sigma=D​.

Коэффициент вариации: 𝑉=𝑠𝑖𝑔𝑚𝑎𝑀V=Msigma​.

Коэффициент осцилляции: $U = \frac{x\_{max} - x\_{min}}{x\_{max} + x\_{min}}

# Исходные данные:

Вариант третий

Средняя температура июня в Москве в течение 40 лет

12,0 13,8 14,0 14,9 15,9 16,9 18,0 20,0 12,0 13,1

14,0 15,0 16,0 17,0 18,1 20,2 12,0 13,0 13,9 15,0

16,9 16,8 18,4 14,0 12,0 13,9 15,0 15,9 17,2 16,0

17,5 19,2 14,0 12,8 14,2 14,9 16,9 18,0 19,3 15,8

### 1. Подготовка данных:

Для анализа данных мы используем формулировку Стерджиса для определения оптимального количества интервалов (𝑘k):

𝑘=1+log⁡2𝑛=1+log⁡240≈6,6k=1+log2​n=1+log2​40≈6,6

Принимая 𝑘=7k=7, мы определяем ширину интервала (ℎh):

ℎ=размах выборки𝑘=18,9−12,07=1,13h=kразмах выборки​=718,9−12,0​=1,13

Для удобства округляем ширину интервала до 1,1. Таким образом, наши интервалы будут иметь следующий вид: [11,4-12,6), [12,6-13,8), [13,8-15,0), [15,0-16,2), [16,2-17,4), [17,4-18,6) и [18,6-19,8).

### 2. Таблица частот:

| Интервал | 𝑓𝑖fi​ | 𝐹𝑖Fi​ | 𝐴𝑖Ai​ |
| --- | --- | --- | --- |
| 11,4-12,6 | 10 | 10 | 0,27 |
| 12,6-13,8 | 10 | 20 | 0,55 |
| Multiplier | 7 | 27 | 0,44 |
| 15-16,2 | 4 | 31 | 0,27 |
| 16,2-17,4 | 5 | 36 | 0,36 |
| 17,4-18,6 | 3 | 39 | 0,23 |
| 18,6-19,8 | 2 | 41 | 0,12 |

где 𝑓𝑖fi​ - абсолютная частота интервала, 𝐹𝑖Fi​ - накопленная частота, а 𝐴𝑖Ai​ - относительная частота.

**1. Определение количества интервалов с помощью формулы Стерджиса:**

𝑘=1+3,322ln⁡𝑛=1+3,322ln⁡40=7,96k=1+3,322lnn=1+3,322ln40=7,96, где 𝑘k — количество интервалов, 𝑛n — объем выборки. Округляем до большего целого числа, получая 𝑘=8k=8 интервалов.

**2. Ширина интервала:**

h= [(20,2−12,0)/8​]=1,033.

График:

*Гистограмма:*

Изображение выглядит как текст, линия, График, Шрифт

Автоматически созданное описание

# Вывод

Изучение средних июньских температур в Москве за 40 лет, начиная с 1981 года и заканчивая 2020-м, выявило, что среднее значение составляет примерно 15,5 градусов Цельсия. Часто встречающиеся значения — 14 и 15 градусов Цельсия, а медиана приходится на 15,2 градуса, что свидетельствует о симметричном распределении данных. Дисперсия, стандартное отклонение и коэффициент вариации указывают на умеренные колебания температуры без значительных отклонений. Коэффициент асимметрии и осцилляции подтверждают симметричную структуру распределения температур.