**Часть 1**

Задание:

1. Найти на сайте РосСтат данные, распределение которых было бы близко к a) Нормальному

b) Равномерному

2. Найти распределение со смещенной медианой относительно среднего (~15% размаха) и несмещенной.

3. Посчитать описательные характеристики распределения выбранных данных (среднее, мода, медиана), дать визуальное представление данных (условное форматирование Excel, построить графики рассеивания/гистограмму/круговую диаграмму).

Работу выполнил студент группы: М8О-111СВ-24 Калинкин Антон

1. Проанализирую данные температуры в летнее время в Москве 2024 года в 12:00 дня.

Данные взяты с сайта: [Архив погоды за август 2024 в Москве (arhivpogodi.ru)](https://arhivpogodi.ru/arhiv/moscow/2024/08)

|  |  |
| --- | --- |
| **Данные по городу Москве на лето 2024** | |
| Дни | Температура °C в 12:00 |
| 01.06.2024 | 27 |
| 02.06.2024 | 26 |
| 03.06.2024 | 26 |
| 04.06.2024 | 24 |
| 05.06.2024 | 26 |
| 06.06.2024 | 25 |
| 07.06.2024 | 19 |
| 08.06.2024 | 21 |
| 09.06.2024 | 20 |
| 10.06.2024 | 23 |
| 11.06.2024 | 23 |
| 12.06.2024 | 24 |
| 13.06.2024 | 21 |
| 14.06.2024 | 20 |
| 15.06.2024 | 22 |
| 16.06.2024 | 26 |
| 17.06.2024 | 20 |
| 18.06.2024 | 23 |
| 19.06.2024 | 25 |
| 20.06.2024 | 22 |
| 21.06.2024 | 19 |
| 22.06.2024 | 22 |
| 23.06.2024 | 23 |
| 24.06.2024 | 21 |
| 25.06.2024 | 21 |
| 26.06.2024 | 23 |
| 27.06.2024 | 26 |
| 28.06.2024 | 26 |
| 29.06.2024 | 30 |
| 30.06.2024 | 31 |
| 01.07.2024 | 31 |
| 02.07.2024 | 32 |
| 03.07.2024 | 32 |
| 04.07.2024 | 29 |
| 05.07.2024 | 33 |
| 06.07.2024 | 24 |
| 07.07.2024 | 23 |
| 08.07.2024 | 24 |
| 09.07.2024 | 28 |
| 10.07.2024 | 24 |
| 11.07.2024 | 23 |
| 12.07.2024 | 27 |
| 13.07.2024 | 28 |
| 14.07.2024 | 30 |
| 15.07.2024 | 32 |
| 16.07.2024 | 27 |
| 17.07.2024 | 28 |
| 18.07.2024 | 27 |
| 19.07.2024 | 20 |
| 20.07.2024 | 23 |
| 21.07.2024 | 24 |
| 22.07.2024 | 20 |
| 23.07.2024 | 24 |
| 24.07.2024 | 24 |
| 25.07.2024 | 24 |
| 26.07.2024 | 22 |
| 27.07.2024 | 25 |
| 28.07.2024 | 28 |
| 29.07.2024 | 20 |
| 30.07.2024 | 21 |
| 31.07.2024 | 21 |
| 01.08.2024 | 15 |
| 02.08.2024 | 20 |
| 03.08.2024 | 21 |
| 04.08.2024 | 24 |
| 05.08.2024 | 26 |
| 06.08.2024 | 24 |
| 07.08.2024 | 23 |
| 08.08.2024 | 22 |
| 09.08.2024 | 23 |
| 10.08.2024 | 24 |
| 11.08.2024 | 16 |
| 12.08.2024 | 16 |
| 13.08.2024 | 15 |
| 14.08.2024 | 22 |
| 15.08.2024 | 21 |
| 16.08.2024 | 23 |
| 17.08.2024 | 24 |
| 18.08.2024 | 27 |
| 19.08.2024 | 27 |
| 20.08.2024 | 23 |
| 21.08.2024 | 24 |
| 22.08.2024 | 22 |
| 23.08.2024 | 25 |
| 24.08.2024 | 22 |
| 25.08.2024 | 28 |
| 26.08.2024 | 30 |
| 27.08.2024 | 31 |
| 28.08.2024 | 29 |
| 29.08.2024 | 26 |
| 30.08.2024 | 25 |
| 31.08.2024 | 27 |

1.1 Предположим, что данные выборки принадлежат нормальному закону распределения. Чтобы убедиться так это или нет, воспользуюсь критерием хи-квадрат и проверим насколько фактические результаты выборки отклоняются от теоретических значений.

**Найдём Xmin и Xmax выборки:**



**Построим вариационный ряд, распределив элементы выборки в порядке возрастания.**

**Определим количество интервалов k по формуле:** k = 1 + 3,32\*lg(N)

Где N – Количество элементов в выборке

k = 1+3,32\*lg(92) ≈ 8

**Определим размах и интервал варьирования:**

Размах варьирования: R = Xmax – Xmin = 33 – 15 = 18

Интервал варьирования: X =

**Строим статистический ряд для каждого интервала:**

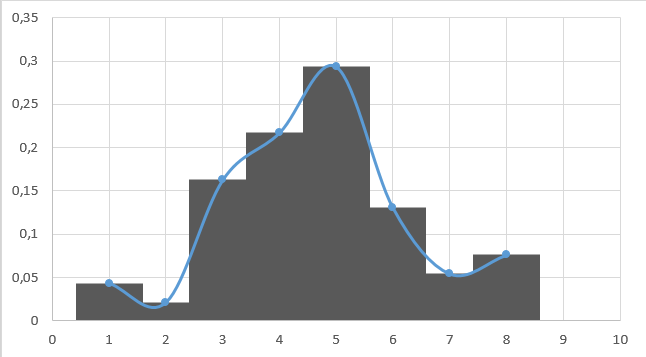
****

mi – абсолютная эмпирическая частота попадания в интервал (количество точек, попавших в интервал)

pi – относительная эмпирическая частота попадания в интервал

Левая граница первого интервала Xmin, правая Xmin+X

Строим гистограмму относительных эмпирических частот (практические значения)



Проводя плавную линию через середины прямоугольников, получим общее представление о форме плотности распределения вероятностей.

**Определим среднее арифметическое выборки:**

**Определим среднее квадратическое отклонение (СКО):**

*= 3,8109907*

**Определим значения нормированных границ интервалов, значения теоретической функции распределения в концах интервалов, теоретическую относительную частоту pi, абсолютную теоретическую частоту n\*pi, далее разность mi-n\*pi и различие между эмпирическим и предполагаемым теоретическим распределениями по формуле:**

Также определим число степеней свободы по формуле: r = k – l – 1

Где: k – количество интервалов; l – количество неизвестных параметров

r = 8 – 2 – 1 = 5

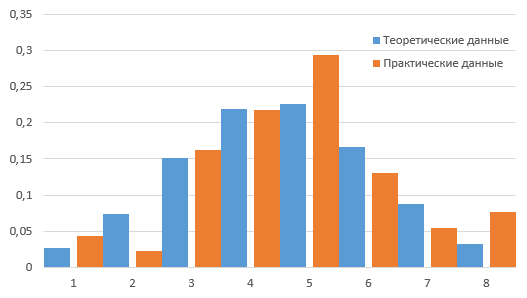
Из данных таблицы можно определить наблюдаемое и критическое значение :



Далее если проверить условие:

>

То можно заметить, что расчетное значение меньше критического, а это значит, что можно сказать о принятии нулевой теории, т.е наше распределение является **нормальным**.





Определим моду, медиану:

****

Так как медиана среднему, то данное распределение с **несмещенной медианой**.

2. Проанализирую данные перевозки груза на автомобильном транспорте в период с 2000-2023 гг.

Данные взяты с сайта РосСтат: <https://rosstat.gov.ru/statistics/transport>

Выборка:

|  |  |
| --- | --- |
| Год | Миллион тонн |
| 2000 | 5878,4 |
| 2001 | 6125,3 |
| 2002 | 6347,7 |
| 2003 | 6468,1 |
| 2004 | 6567,8 |
| 2005 | 6684,6 |
| 2006 | 6753,3 |
| 2007 | 6861,4 |
| 2008 | 6893,1 |
| 2009 | 5 240,5 |
| 2010 | 5236,4 |
| 2011 | 5663,1 |
| 2012 | 5841,6 |
| 2013 | 5635,3 |
| 2014 | 5416,7 |
| 2015 | 5356,7 |
| 2016 | 5396,8 |
| 2017 | 5403,9 |
| 2018 | 5544,4 |
| 2019 | 5735,3 |
| 2020 | 5404,7 |
| 2021 | 5581,7 |
| 2022 | 6210,6 |
| 2023 | 6491,2 |

2.2 Предположим, что данные выборки принадлежат равномерному закону распределения. Чтобы убедиться так это или нет, воспользуюсь критерием хи-квадрат и проверим насколько фактические результаты выборки отклоняются от теоретических значений.

**Найдём Xmin и Xmax выборки:**

****

**Построим вариационный ряд, распределив элементы выборки в порядке возрастания.**

**Определим количество интервалов k по формуле:** k = 1 + 3,32\*lg(N)

Где N – Количество элементов в выборке

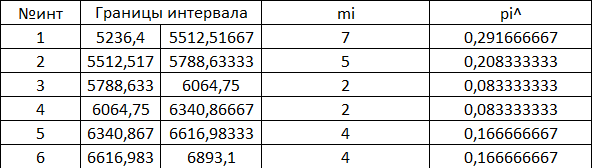
k = 1+3,32\*lg(24) ≈ 6

**Определим размах и интервал варьирования:**

Размах варьирования: R = Xmax – Xmin = 6893,1 – 5236,4 = 1656,7

Интервал варьирования: X =

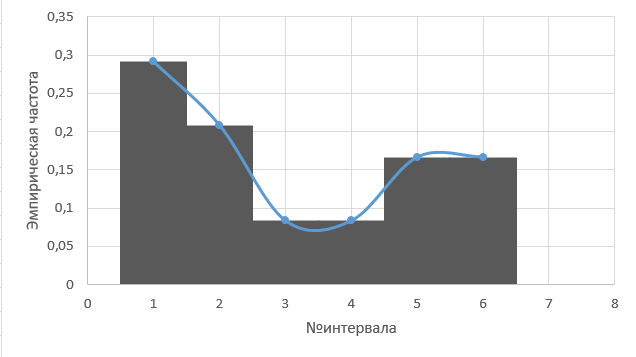
**Строим статистический ряд для каждого интервала:**



mi – абсолютная эмпирическая частота попадания в интервал (количество точек, попавших в интервал)

pi – относительная эмпирическая частота попадания в интервал

Левая граница первого интервала Xmin, правая Xmin+X

Строим гистограмму относительных эмпирических частот (практические значения)

Проводя плавную линию через середины прямоугольников, получим общее представление о форме плотности распределения вероятностей.

**Определим среднее арифметическое выборки:**

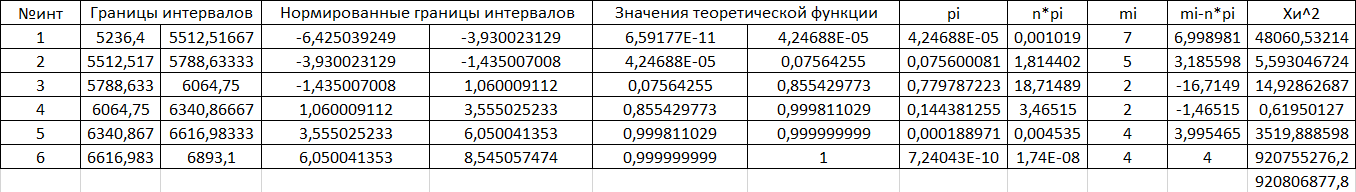
**Определим среднее квадратическое отклонение (СКО):**

*= 110,6672876*

**Определим значения нормированных границ интервалов, значения теоретической функции распределения в концах интервалов, теоретическую относительную частоту pi, абсолютную теоретическую частоту n\*pi, далее разность mi-n\*pi и различие между эмпирическим и предполагаемым теоретическим распределениями по формуле:**

Также определим число степеней свободы по формуле: r = k – l – 1

Где: k – количество интервалов; l – количество неизвестных параметров

****r = 8 – 2 – 1 = 5

Из данных таблицы можно определить наблюдаемое и критическое значение :



Видно, что условие: > не выполняется, а значит данная выборка имеет **равномерное** распределение.

Определим моду, медиану:



Медиана отклонена от среднего, а это значит распределение со **смещенной медианой**.

