

## Анализ данных

### Лабораторная работа 1: Построение вариационного ряда

Тема: "Вариационный ряд"

Цель: получить навыки установления статистических закономерностей, присущих массовым случайным явлениям средствами Excel;

Используемое оборудование: Персональный компьютер, Excel 2013, Microsoft Word 2013, online IDE Repl.it

#### 1) Постановка задачи:

1. Имеются данные о распределении 100 рабочих цеха по выработке в отчетном году (в процентах к предыдущему году). Всего  $n=100$  значений. (см. файл **zadanie1.docx** в приложении, внизу этой страницы)

|               |       |       |     |     |     |       |       |     |     |
|---------------|-------|-------|-----|-----|-----|-------|-------|-----|-----|
| Ряд признаков | 103.4 | 115.2 | 127 | 131 | ... | 102,3 | 114.5 | 118 | 127 |
|---------------|-------|-------|-----|-----|-----|-------|-------|-----|-----|

Необходимо построить вариационный ряд. Сгруппировать его и построить графические изображения вариационного ряда. Построить полигон (гистограмму), кумулянту и эмпирическую функцию распределения рабочих

#### 1) Математическая модель:

Частость:  $w_i = \frac{m_i}{n}$

Накопленная частота  $m_x = \sum_{x_i < x} m_i$

Накопленной частоты:  $w_x = \frac{m_x}{n} = \frac{1}{n} \sum_{x_i < x} m_i$

Количество интервалов:  $k = 1 + 1,4 \sqrt{n}$

Длина интервала:  $\Delta = x_{\max} - x_{\min} / k$

Начало 1-го интервала:  $x_{\text{нач}} = x_{\min} - \frac{k}{2}$

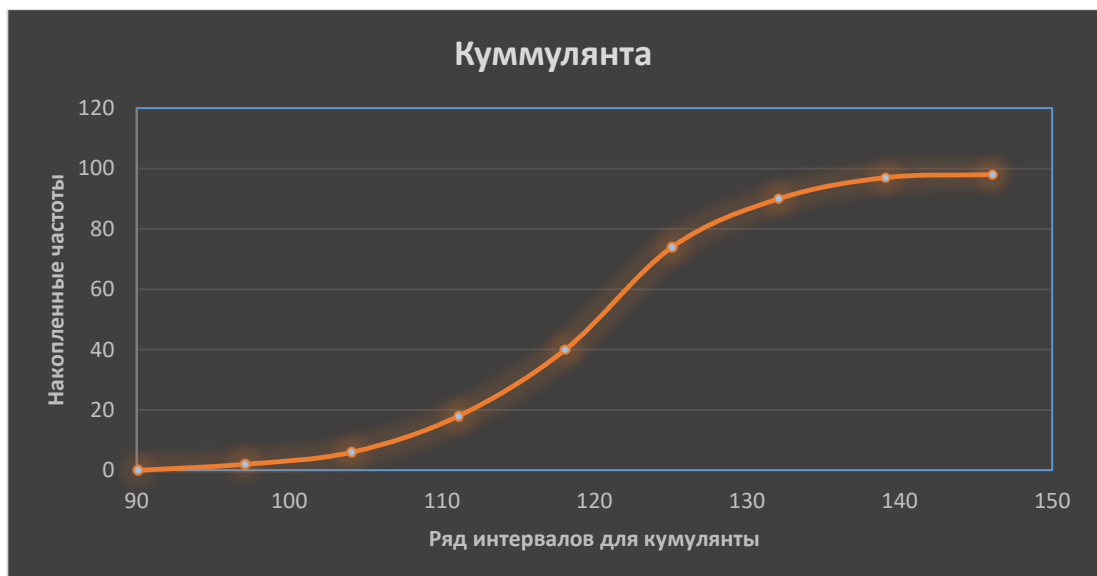
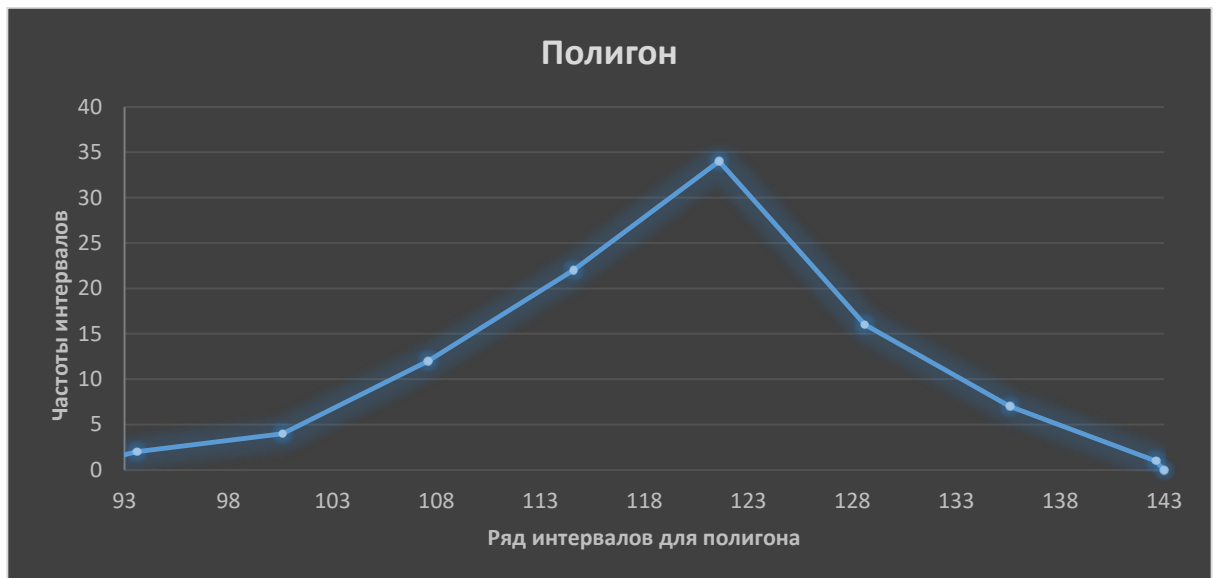
Эмпирическая функция распределения  $F_n(x) = \frac{m_x}{n}$

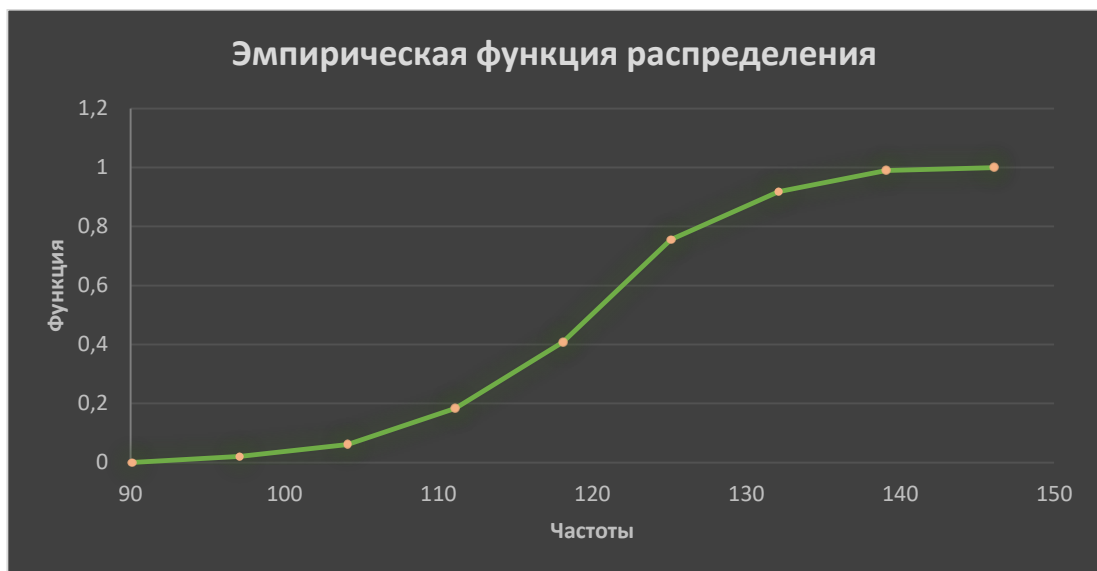
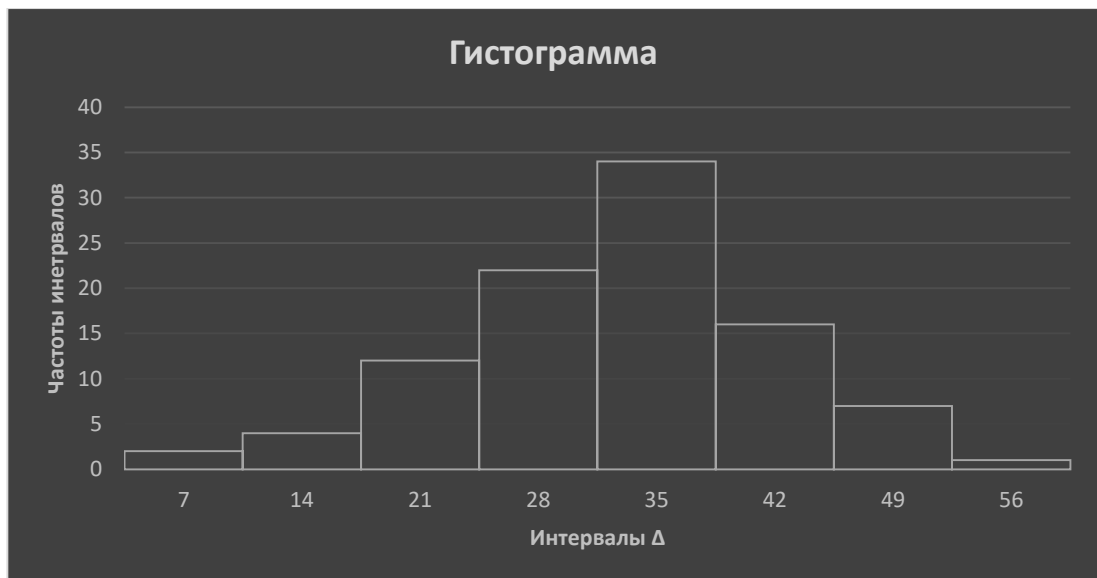
Где  $m_i$  – частоты,  $n$  – количество элементов в ряду,  $x_{\max}$  – максимальный элемент ряда,  $x_{\min}$  – минимальный элемент ряда.

1) Решение:

- Воспользуемся встроенными инструментами форматирования таблиц и данных.
- Воспользуемся встроенными формулами и ссылками Excel для расчётов формул.
- Воспользуемся встроенными инструментами Excel для построения необходимых графиков.

1) Результат:





## 2) Постановка задачи:

**2.** В таблице 2 дан дискретный ряд. В ней приведены данные о распределении 50-ти рабочих цеха по тарифному разряду.

|                                       |   |   |   |   |    |   |
|---------------------------------------|---|---|---|---|----|---|
| Тарифный разряд $x_i$<br>(варианты)   | 1 | 2 | 3 | 4 | 5  | 6 |
| Частота (количество<br>рабочих) $n_i$ | 2 | 3 | 6 | 8 | 22 | 9 |

Построить полигон (гистограмму), кумулянту и эмпирическую функцию распределения рабочих

2) Математическая модель:

Частость:  $w_i = \frac{m_i}{n}$

Накопленная частота  $m_x = \sum_{x_i < x} m_i$

Накопленной частоты:  $w_x = \frac{m_x}{n} = \frac{1}{n} \sum_{x_i < x} m_i$

Количество интервалов:  $k = 1 + 1,4 \sqrt{n}$

Длина интервала:  $\Delta = (x_{\max} - x_{\min}) / k$

Начало 1-го интервала:  $x_{\text{нач}} = x_{\min} - \frac{k}{2}$

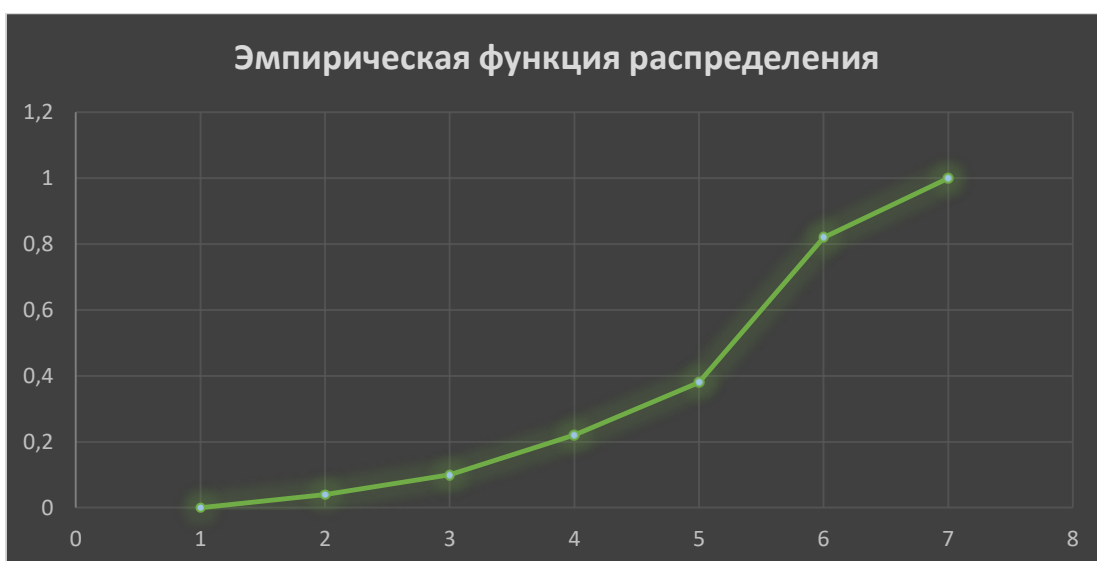
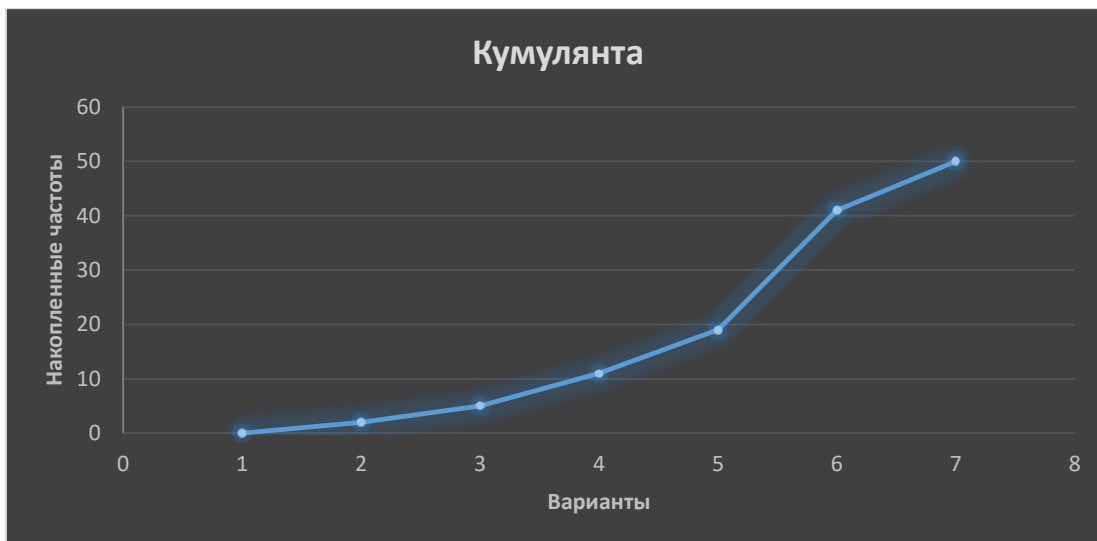
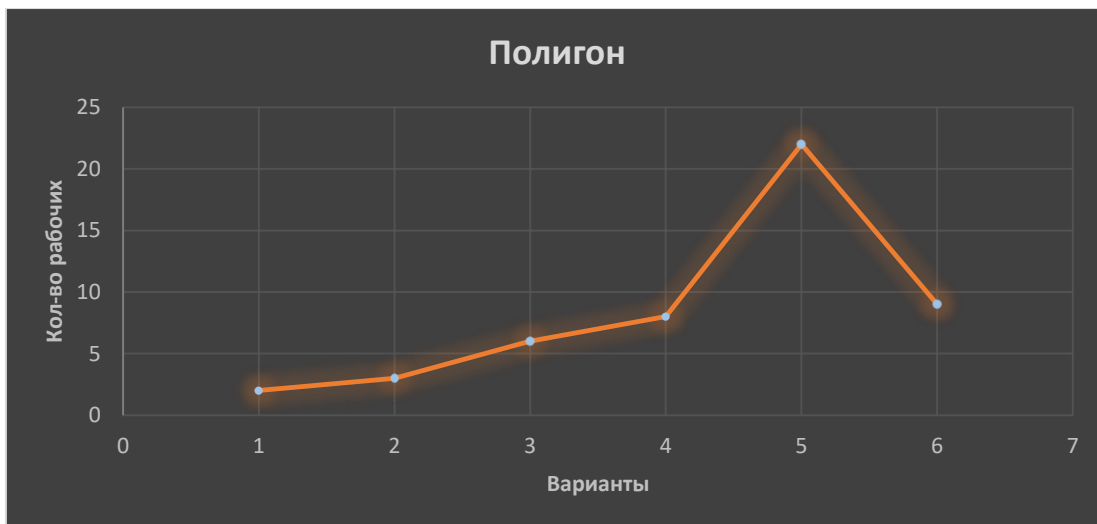
Эмпирическая функция распределения  $F_n(x) = w_x = \frac{m_x}{n}$

Где  $m_i$  – частоты,  $n$  – количество элементов в ряду,  $x_{\max}$  – максимальный элемент ряда,  $x_{\min}$  – минимальный элемент ряда.

2) Решение:

- Воспользуемся встроенными инструментами форматирования таблиц и данных.
- Воспользуемся встроенными формулами и ссылками Excel для расчётов формул.
- Воспользуемся встроенными инструментами Excel для построения необходимых графиков.

2) Результат:



### 3) Постановка задачи:

3. В файле zadanie2.docx (в приложении внизу страницы) содержатся выборочные данные. Постройте вариационный ряд и его графические изображения (гистограмму, полигон, кумулянту и эмпирическую функцию распределения).

### 3) Математическая модель:

Частость:  $w_i = \frac{m_i}{n}$

Накопленная частота  $m_x = \sum_{x_i < x} m_i$

Накопленной частоты:  $w_x = \frac{m_x}{n} = \frac{1}{n} \sum_{x_i < x} m_i$

Количество интервалов:  $k = 1 + 1,4 \sqrt{n}$

Длина интервала:  $\Delta = x_{\max} - x_{\min} / k$

Начало 1-го интервала:  $x_{\text{нач}} = x_{\min} - \frac{k}{2}$

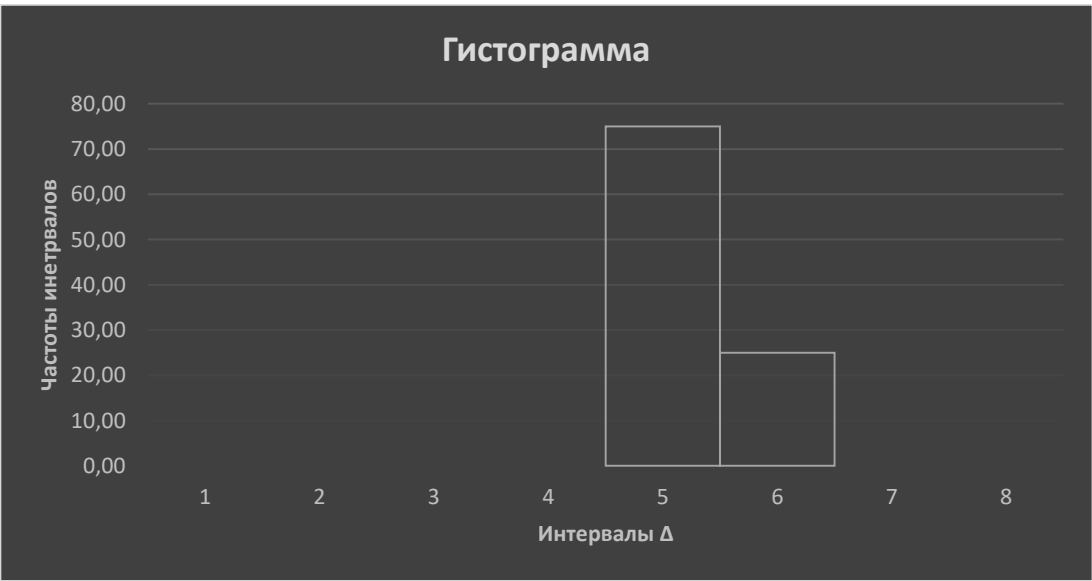
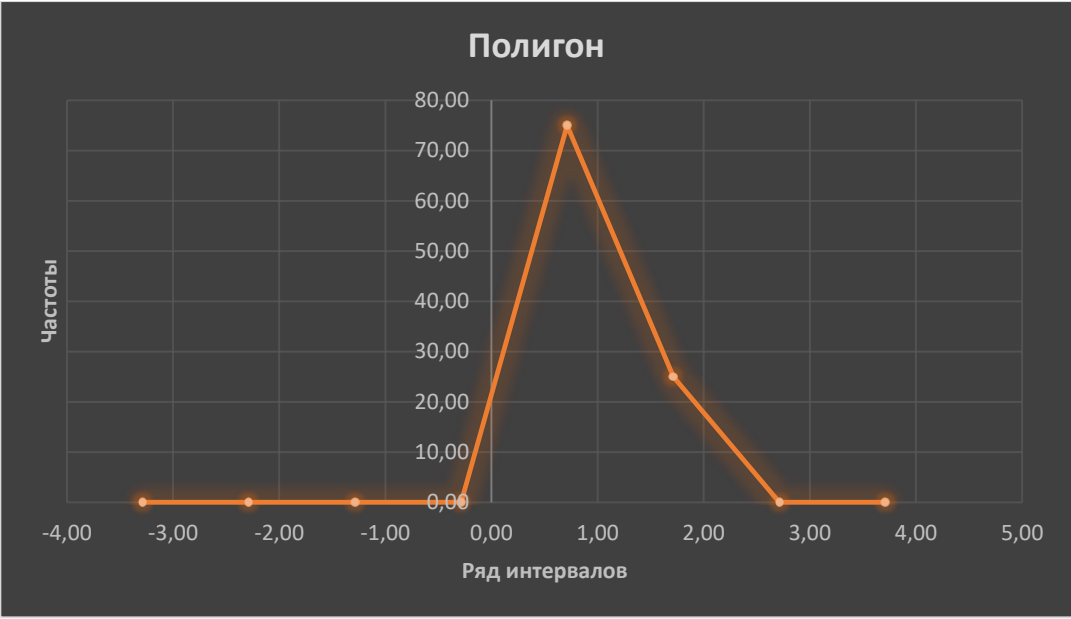
Эмпирическая функция распределения  $F_n(x) = w_x = \frac{m_x}{n}$

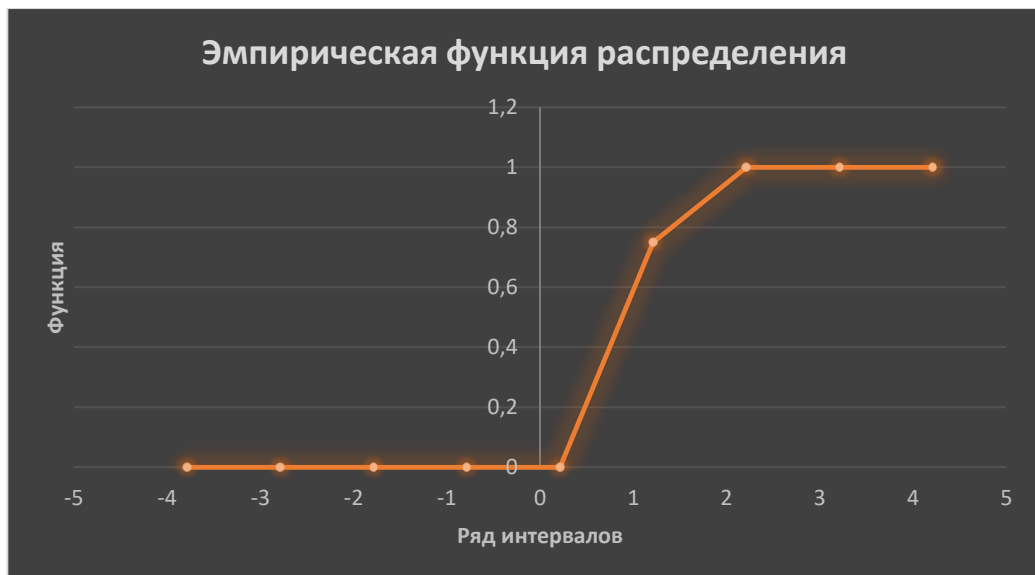
Где  $m_i$  – частоты,  $n$  – количество элементов в ряду,  $x_{\max}$  – максимальный элемент ряда,  $x_{\min}$  – минимальный элемент ряда.

### 3) Решение:

- Воспользуемся встроенными инструментами форматирования таблиц и данных.
- Воспользуемся встроенными формулами и ссылками Excel для расчётов формул.
- Воспользуемся встроенными инструментами Excel для построения необходимых графиков

### 3) Результат:





4) Постановка задачи:

**4.** Имеются выборочные данные по результатам экзамена по химии учащихся школы. Построить интервальный вариационный ряд и его графическое представление (гистограмму, полигон, кумулянту и эмпирическую функцию распределения).

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 4 | 3 | 3 | 2 | 5 | 2 | 3 | 3 | 4 |
| 3 | 4 | 4 | 2 | 5 | 2 | 3 | 3 | 4 | 4 |
| 3 | 3 | 4 | 4 | 2 | 5 | 5 | 2 | 3 | 3 |

4) Математическая модель:

Частость:  $w_i = \frac{m_i}{n}$

Накопленная частота  $m_x = \sum_{x_i < x} m_i$

Накопленной частоты:  $w_x = \frac{m_x}{n} = \frac{1}{n} \sum_{x_i < x} m_i$

Количество интервалов:  $k = 1 + 1,41 \sqrt{n}$

Длина интервала:  $\Delta = (x_{\max} - x_{\min}) / k$

Начало 1-го интервала:  $x_{\text{нач}} = x_{\min} - \frac{k}{2}$



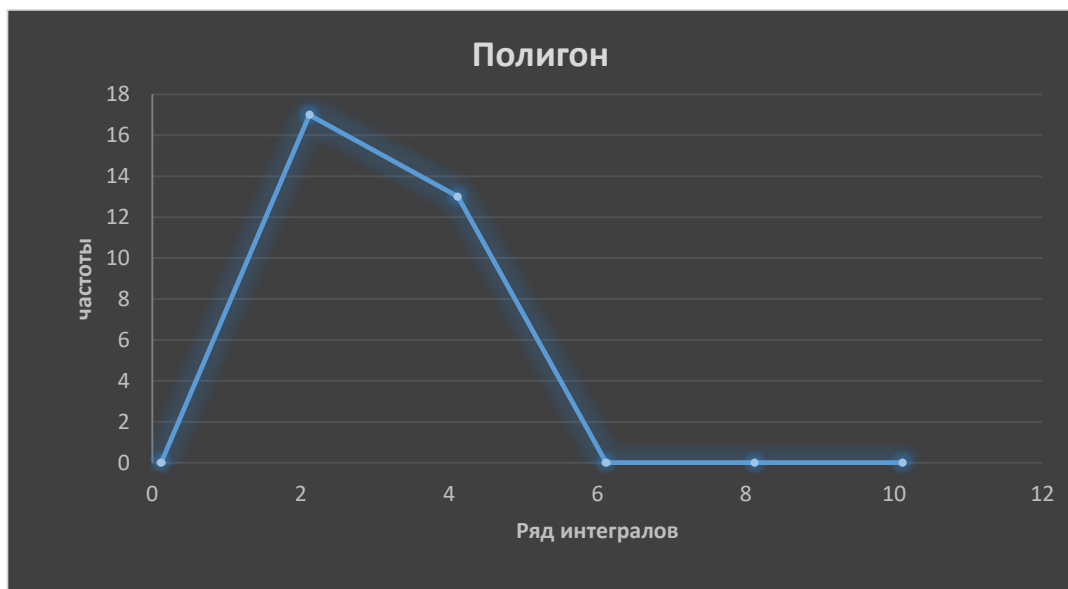
Эмпирическая функция распределения  $F_n(x) = \frac{m_x}{n}$

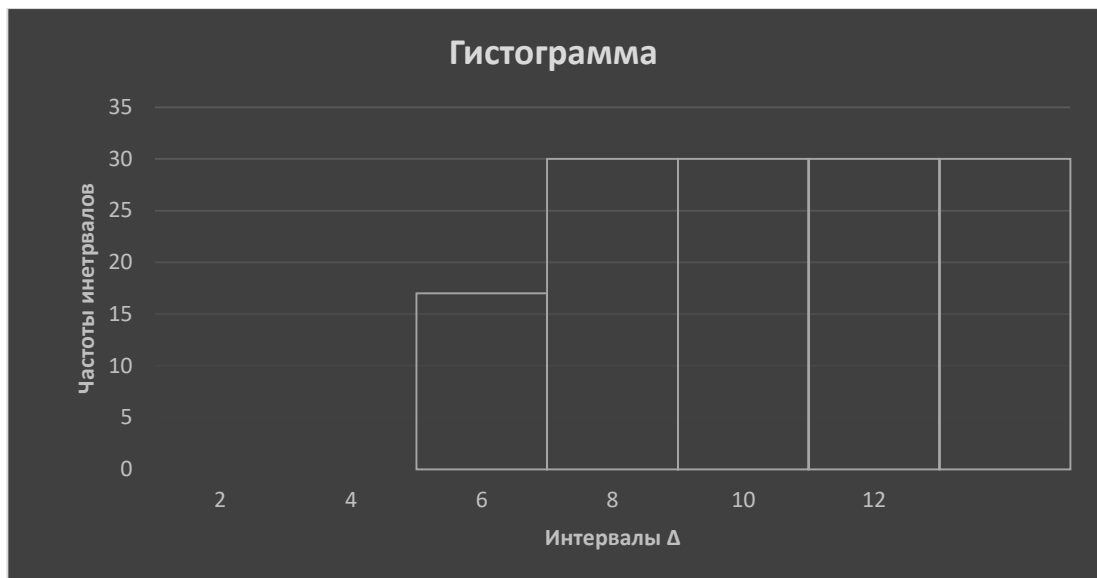
Где  $m_i$  – частоты,  $n$  – количество элементов в ряду,  $x_{\max}$  – максимальный элемент ряда,  $x_{\min}$  – минимальный элемент ряда.

4) Решение:

- Воспользуемся встроенными инструментами форматирования таблиц и данных.
- Воспользуемся встроенными формулами и ссылками Excel для расчётов формул.
- Воспользуемся встроенными инструментами Excel для построения необходимых графиков.

4) Результат:





5) Постановка задачи:

**5.** В ходе спортивных соревнований были получены результаты бега 30 спортсменов. Эти результаты образуют следующий ряд:

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 18 | 10 | 17 | 13 | 15 | 15 | 14 | 17 | 20 | 19 |
| 15 | 15 | 14 | 13 | 16 | 16 | 12 | 11 | 13 | 14 |

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 19 | 20 | 15 | 16 | 15 | 16 | 14 | 16 | 13 | 12 |
|----|----|----|----|----|----|----|----|----|----|

Построить интервальный вариационный ряд и его графическое представление (гистограмму, полигон, кумулянту и эмпирическую функцию распределения).

5) Математическая модель:

Частость:  $w_i = \frac{m_i}{n}$

Накопленная частота  $m_x = \sum_{x_i < x} m_i$

Накопленной частоты:  $w_x = \frac{m_x}{n} = \frac{1}{n} \sum_{x_i < x} m_i$

Количество интервалов:  $k = 1 + 1,414 \sqrt{n}$

Длина интервала:  $\Delta = (x_{\max} - x_{\min}) / k$

Начало 1-го интервала:  $x_{\text{нач}} = x_{\min} - \frac{k}{2}$

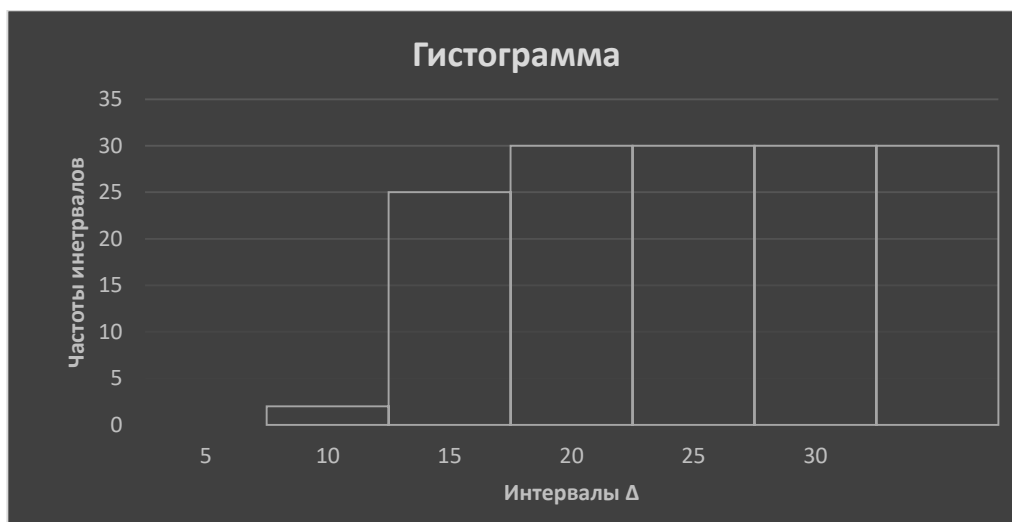
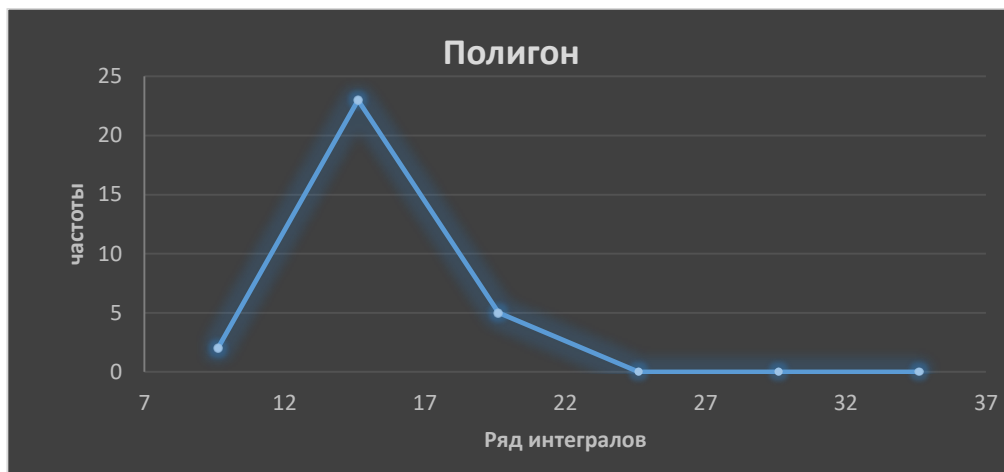
Эмпирическая функция распределения  $F_n(x) = w_x = \frac{m_x}{n}$

Где  $m_i$  – частоты,  $n$  – количество элементов в ряду,  $x_{\max}$  – максимальный элемент ряда,  $x_{\min}$  – минимальный элемент ряда.

5) Решение:

- Воспользуемся встроенными инструментами форматирования таблиц и данных.
- Воспользуемся встроенными формулами и ссылками Excel для расчётов формул.
- Воспользуемся встроенными инструментами Excel для построения необходимых графиков.

## 5) Результат:





6) Постановка задачи:

Взять одну из задач. Построить вариационный ряд, рассчитать веса и необходимые величины, используя язык программирования. (Выбранная задача: №4)

6) Математическая модель:

Частость:  $w_i = \frac{m_i}{n}$

Накопленная частота  $m_x = \sum_{x_i < x} m_i$

Накопленной частоты:  $w_x = \frac{m_x}{n} = \frac{1}{n} \sum_{x_i < x} m_i$

Количество интервалов:  $k = 1 + 1,414 \sqrt{n}$

Длина интервала:  $\Delta = (x_{\max} - x_{\min}) / k$

Начало 1-го интервала:  $x_{\text{нач}} = x_{\min} - \frac{\Delta}{2}$

Эмпирическая функция распределения  $F_n(x) = w_x = \frac{m_x}{n}$

Где  $m_i$  – частоты,  $n$  – количество элементов в ряду,  $x_{\max}$  – максимальный элемент ряда,  $x_{\min}$  – минимальный элемент ряда.

6) Решение:

- Создадим массив, который наполним элементами из задачи №4 в том же порядке.
- С помощью возможностей Python, отсортируем список.
- Произведём вычисление частот определённых значений ряда.
- Проведём вычисление величин, необходимых для построения и обработки вариационного ряда.
- Создадим массивы, для сохранения данных о границах интервалов, интервалах, частотах и накопленных частотах.
- Оформим и выведем нужные значения на экран.

[Ссылка на код, расположенный в Repl.it](#)



6) Результат:

Изначальный ряд

[4, 4, 3, 3, 2, 5, 2, 3, 3, 4, 3, 4, 4, 2, 5, 2, 3, 3, 4, 4, 3, 3, 4, 4, 2, 5, 5, 2, 3, 3]

Ряд, обработанный по возрастанию

[2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5]

Количество оценок 5

4

Количество оценок 4

9

Количество оценок 3

11

Количество оценок 2

6

Количество элементов в ряду

30

Приближённое значение  $k$

6

Максимальное значение в ряду

5

Минимальное значение в ряду

2

$x$  начальное

-0.8808381671635086

Длина интервала

1.5

Первый интервал

$[-0.8808381671635086, 0.6191618328364914]$

Частота

0

Накопленная частость

0.0

Второй интервал

$[0.6191618328364914, 2.1191618328364914]$

Частота

6

Накопленная частость

0.0

Третий интервал

$[2.1191618328364914, 3.6191618328364914]$

Частота

11

Накопленная частость

0.2



Четвёртый интервал

[3.6191618328364914, 5.119161832836491]

Частота

13

Накопленная частость

0.5666666666666667

Пятый интервал

[5.119161832836491, 6.619161832836491]

Частота

0

Накопленная частость

1.0

Шестой интервал

[6.619161832836491, 8.119161832836491]

Частота

0

Накопленная частость

1.0

□

**Вывод:**

В итоге, нам удалось получить необходимые навыки установления статистических закономерностей, присущих массовым случайным явлениям, используя средства Excel и языка программирования Python.