

ГУАП

КАФЕДРА № 41

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ
ПРЕПОДАВАТЕЛЬ

<u>старший преподаватель</u>	<u></u>	<u>Е. К. Григорьев</u>
должн., уч. степень, звание	подпись, дата	инициалы, фамилия

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ № 5.

СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

по курсу: МОДЕЛИРОВАНИЕ.

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ ГР. №	<u>4217</u>	<u></u>	<u>У. А. Мазориев</u>
		подпись, дата	инициалы, фамилия

Санкт-Петербург 2025

Цель работы

Получить навыки обработки выборочных данных средствами Python.

Результат выполнения работы

Программный код, файл с расчетами в формате xlsl можно посмотреть, перейдя по ссылке на GitHub:

<https://github.com/Mrx112426/Modelirovanie/tree/main>

Ход выполнения работы

Индивидуальное задание: Вариант 10

1. Решение аналитическим способом

Для начала определяться нужно число интервалов для нашей выборки, наша выборка состоит из 50 элементов, в качестве формулы для вычисления оптимального количества интервалов было выбрано соотношение: $z = \sqrt{N}$, где N – количество элементов, т.е. $z = \sqrt{50}$ (рис. 1).

=ЦЕЛОЕ(КОРЕНЬ(B1))					
	D	E	F	G	H
			Кол-во интервалов z		7

Рисунок 1 – Количество интервалов

Далее вычисляется размах выборки, для этого вычисляется разница между максимальным и минимальным значением (рис. 2).

=МАКС(B8:B57)-МИН(B8:B57)					
	D	E	F	G	H
			Размах выборки R		15,81

Рисунок 2 – Размах выборки

Длина интервала вычисляется по формуле $R/z = 15.81/7=2.26$ (рис. 3).

=H3/H1					
	D	E	F	G	H
			Длина интервала		2,26

Рисунок 3 – Длина интервала

Далее формируется таблица с границами интервалов, средними значениями, частотой и относительной частотой (рис. 4):

1. Количество интервалов известно из предыдущих подсчетов
2. Для того чтобы вычислить границы берется минимальное значение – начало для первого интервала, чтобы найти его край нужно прибавить длину интервала, следующий интервал будет начинаться с конца предыдущего, причем если первый интервал $[-6,51; 4; 25)$, то следующий $[-4; 25; -1; 99)$, последний интервал включает две границы в себя
3. Для расчета середина интервала берется среднее значение между границами
4. Чтобы определить частоту нужно подсчитать количество значений, который входят в интервал
5. Для подсчета относительной частоты используется формула $1/N * w(i)$ где $w(i)$ – частота, а N – кол-во значений

Интервалы	Границы интервалов		Середина интервала	Частота	Относительная частота
1	-6,51	-4,25	-5,38	3	0,06
2	-4,25	-1,99	-3,12	5	0,1
3	-1,99	0,27	-0,86	19	0,38
4	0,27	2,52	1,39	12	0,24
5	2,52	4,78	3,65	7	0,14
6	4,78	7,04	5,91	3	0,06
7	7,04	9,30	8,17	1	0,02

Рисунок 4 – Таблица интервалов

После вычисляется значения эмпирической функции распределения (рис. 5)

		0	x	$(-\infty; -5,38)$
		0,06	x	$[-4,25; -1,99)$
		0,16	x	$[-1,99; 0,27)$
F_n(x)	=	0,54	x	$[0,27; 2,52)$
		0,78	x	$[2,52; 4,78)$
		0,92	x	$[4,78; 7,04)$
		0,98	x	$[7,04; 9,30]$
		1	x	$(9,30; \infty)$

Рисунок 5 – Значения ЭФР

Далее строится график гистограммы выборки (рис. 6).

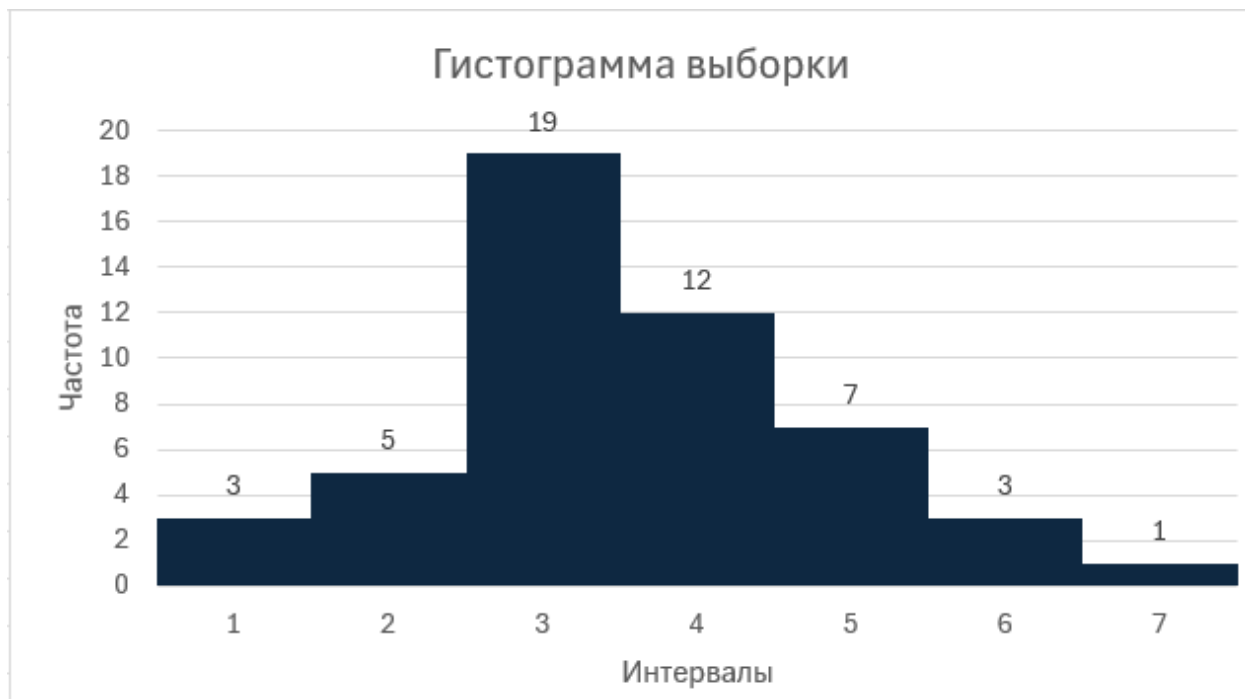


Рисунок 6 – График гистограммы выборки

После строится график полигона частот (рис. 7).

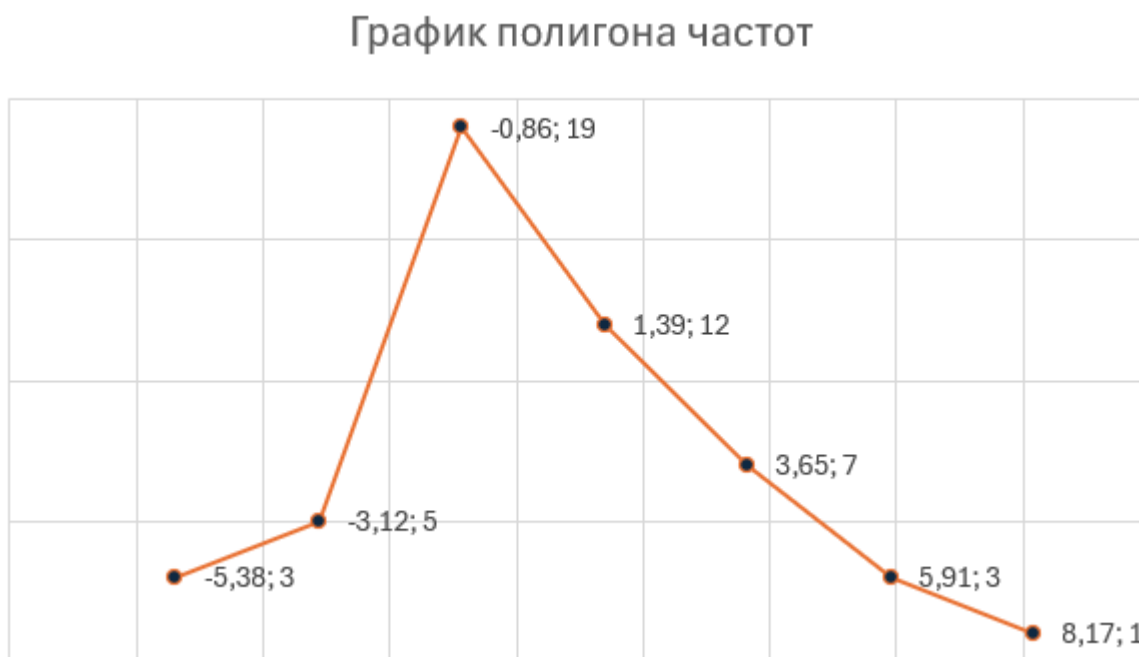


Рисунок 7 – График полигона частот

Далее строится график эмпирической функции распределения (рис. 8).

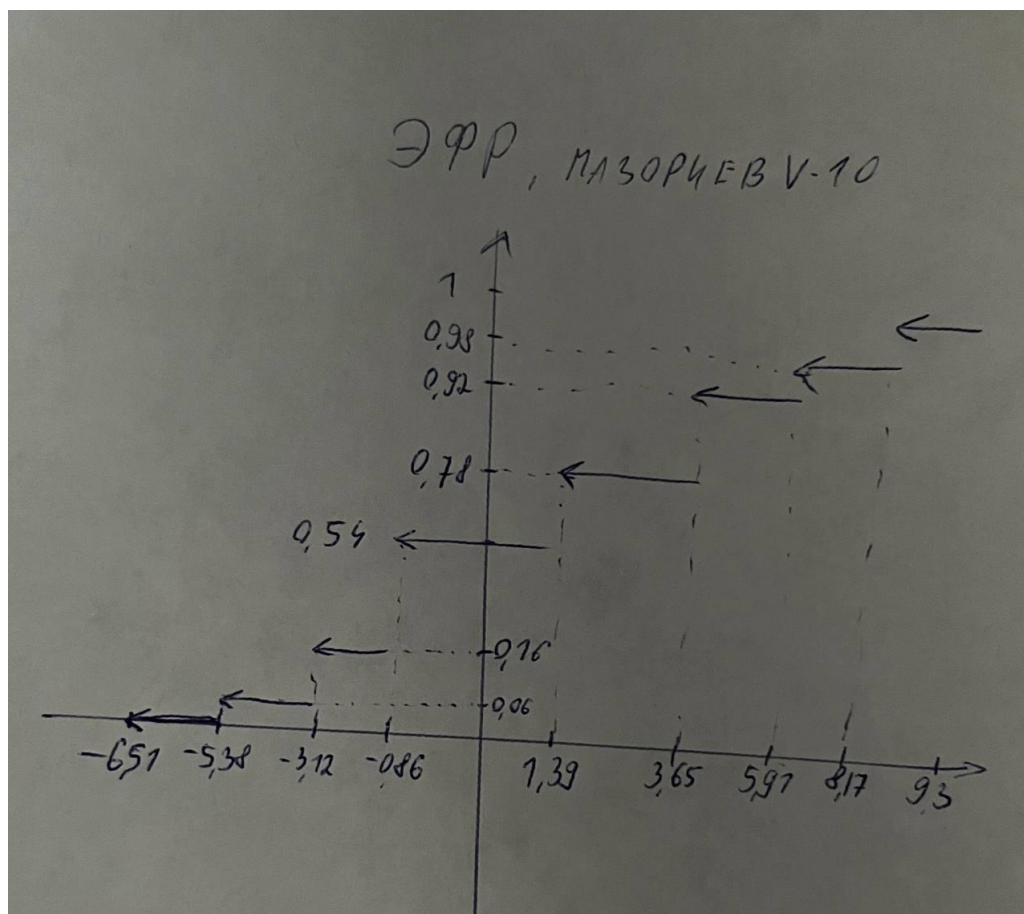


Рисунок 8 – График ЭФР

После считается выборочное среднее $M(x)$, медиана, мода (самое часто-встречающееся значения в выборке. Для подсчета медианы в выборки с четным N используется формула $(X_{\frac{N}{2}} + X_{\frac{N}{2}+1}) * \frac{1}{2}$, где $X_{\frac{N}{2}}$ – элемент в отсортированной выборке (рис. 9)

Выборочное среднее $M(x)$	0,41
Медиана	-0,19
Мода	-1,38

Рисунок 9 – Подсчитанные значения

Далее считается дисперсии выборки, кф асимметрии и эксцесса для их расчета используется формулы на рисунке 10.

$$D(x) = \frac{1}{N-1} \sum_{i=1}^N \left(x_i - M(x) \right)^2$$

$$A_s = \frac{m_3}{\sigma^3} = \frac{\sum_{i=1}^N \left(x_i - M(x) \right)^3}{N\sigma^3}$$

$$E_k = \frac{m_4}{\sigma^4} - 3 = \frac{\sum_{i=1}^N \left(x_i - M(x) \right)^4}{N\sigma^4} - 3$$

Рисунок 10 – Формулы для расчета

Оценка дисперсии выборки	9,37
Оценка кф асимметрии	0,42
Оценка кф эксцесса	0,21

Рисунок 11 – Результат расчета

В конечном счете получаются результаты представленные на рисунке 12.

Кол-во интервалов z	7	Выборочное среднее M(x)	0,41	Оценка дисперсии выборки	9,37
Размах выборки R	15,81	Медиана	-0,19	Оценка кф асимметрии	0,42
Длина интервала	2,26	Мода	-1,38	Оценка кф эксцесса	0,21

Рисунок 12 – Посчитанные значения аналитическим способом

1. Центральная тенденция

Среднее значение (0.41) близко к нулю, что говорит о сбалансированности положительных и отрицательных значений.

Медиана (-0.19) немного меньше среднего, что указывает на небольшую асимметрию в сторону меньших значений.

Мода (-1.38) показывает, что наиболее частое значение находится в отрицательной области.

Размах (15.81) показывает значительное расстояние между минимальным и максимальным значениями.

Дисперсия (9.37) и стандартное отклонение (3.06) указывают на достаточно большой разброс значений.

Коэффициент асимметрии (0.43) положительный, значит, распределение немного скошено вправо (есть больше крупных положительных выбросов).

Коэффициент эксцесса (0.21) положительный, что говорит о том, что выборка немного более "остроконечная", чем нормальное распределение (возможны частые выбросы).

Для наглядности визуализируются график QQ-Plot, который показывает, что распределение имеет небольшое отклонение от нормального, имеются выбросы (рис. 13).

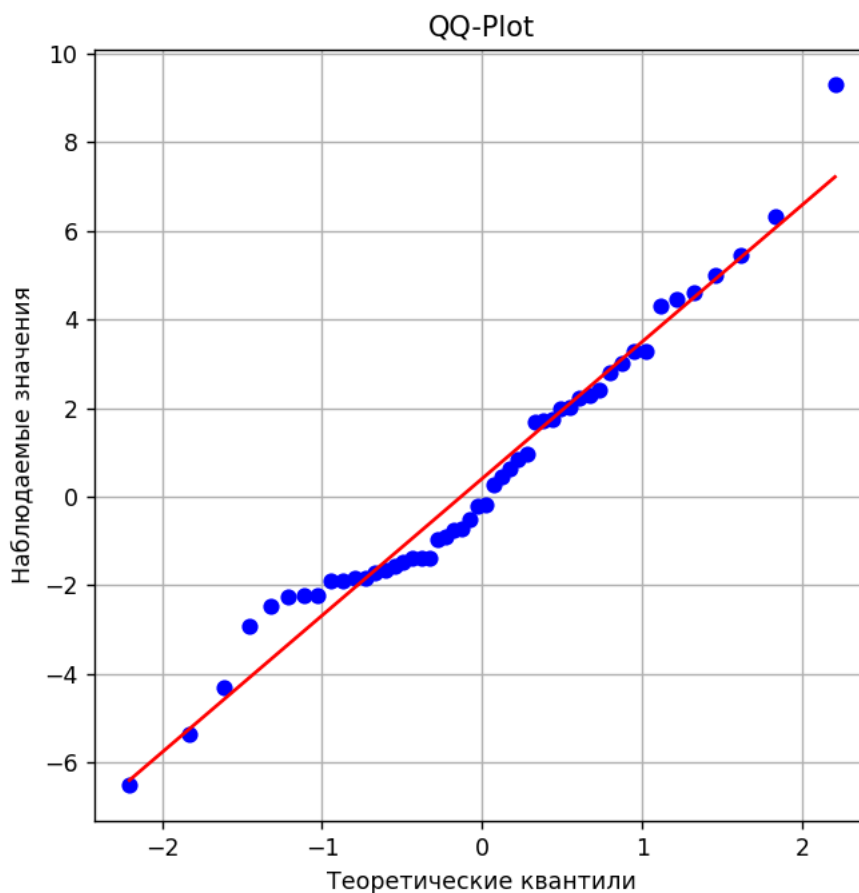


Рисунок 13 – График QQ-plot

2. Решение через Python

Далее пишется скрипт, который считает метрики посчитанные в 1 части, для проверки правильности подсчетов скрипт проверяется на выборки для 1 части, результаты приведены на рисунках 14–17.

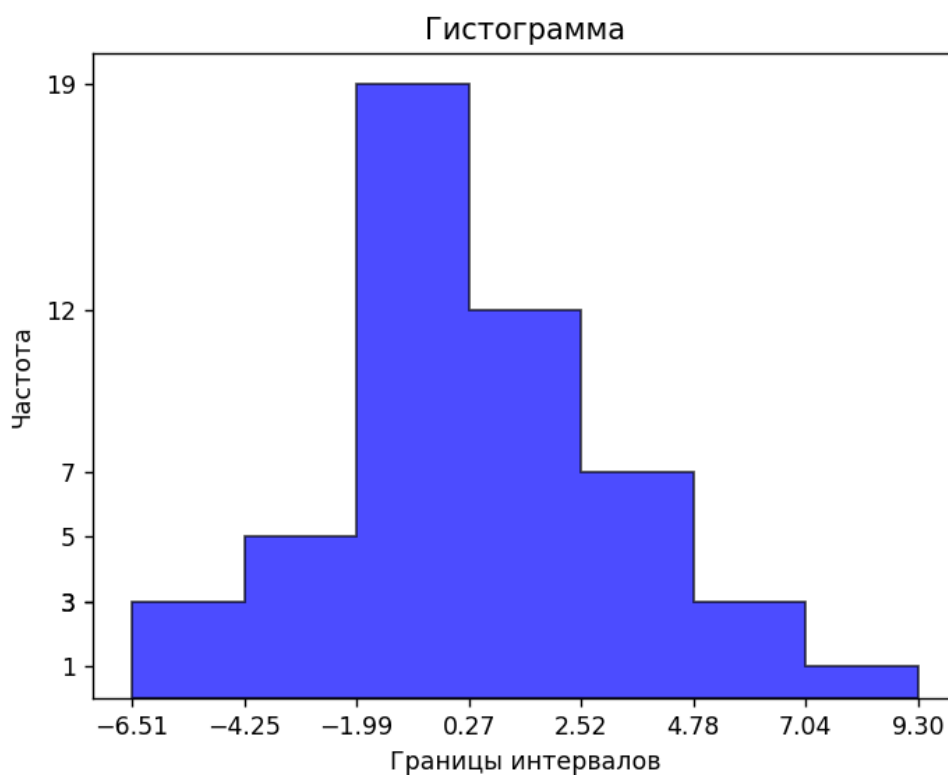


Рисунок 14 – Гистограмма выборки для 1 части

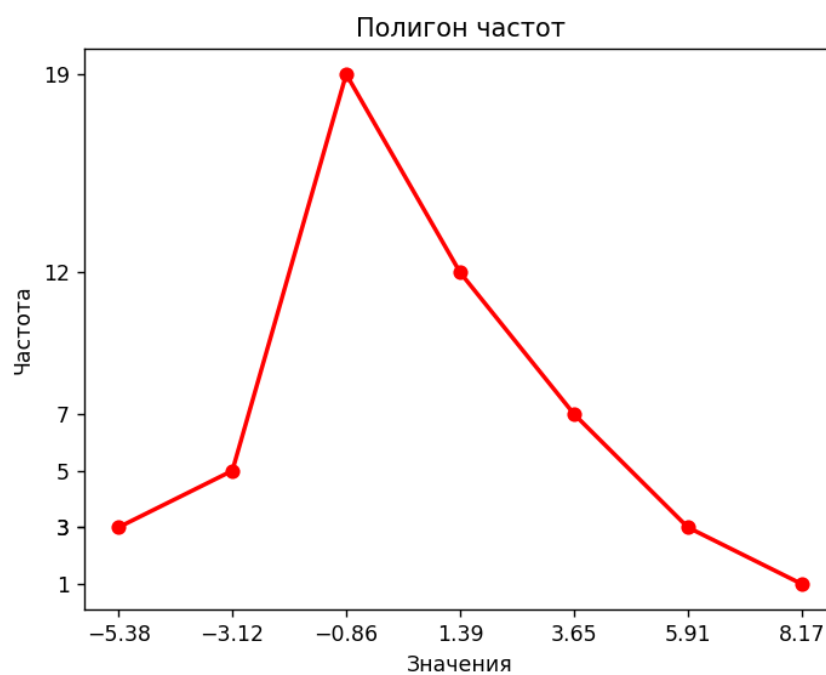


Рисунок 15 – Гистограмма выборки для 1 части

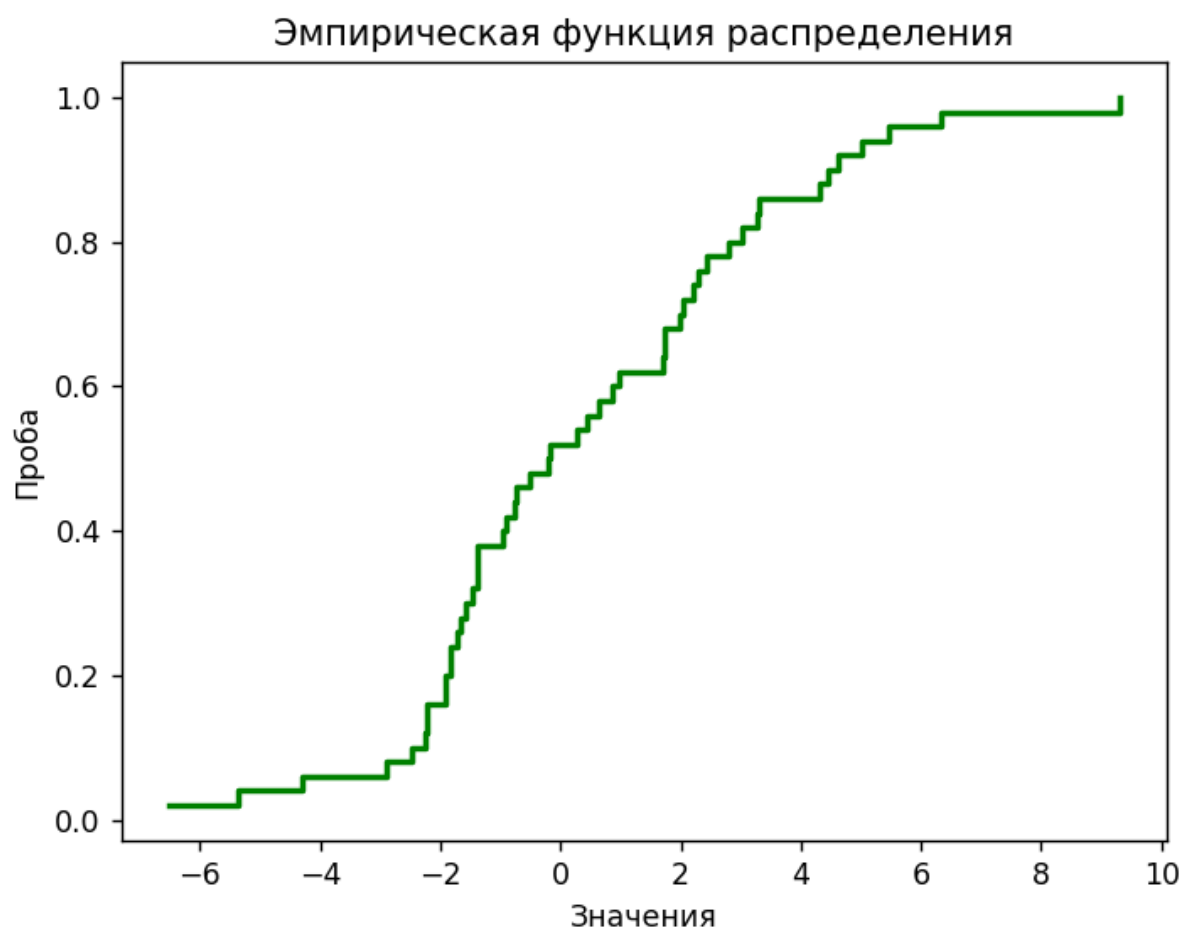


Рисунок 16 – ЭФР для 1 части

```

Количество интервалов: 7
Размах 15.81
Длина интервала 2.2585714285714285
Выборочное среднее: 0.41200000000000003
Выборочная дисперсия: 9.369522448979591
Среднеквадратическое отклонение: 3.0609675674498074
Мода: -1.38
Медиана: -0.19
Коэффициент асимметрии: 0.4296208556059158
Коэффициент эксцесса (без коррекции): 0.21344981495423854

```

Рисунок 17 – Результаты подсчетов для 1 части

Значения подсчитанные вручную совпадают со значениями в Python.

Далее подсчитываются метрики и строятся графики для 2 части (рис. 17 – рис. 20)

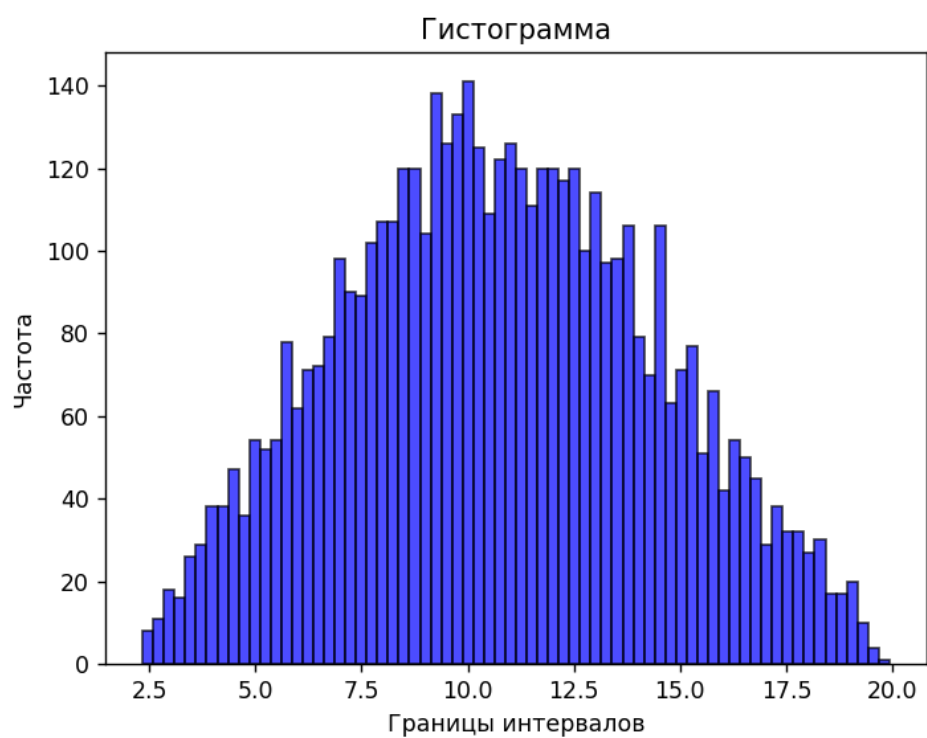


Рисунок 18 – Гистограмма выборки для 2 части

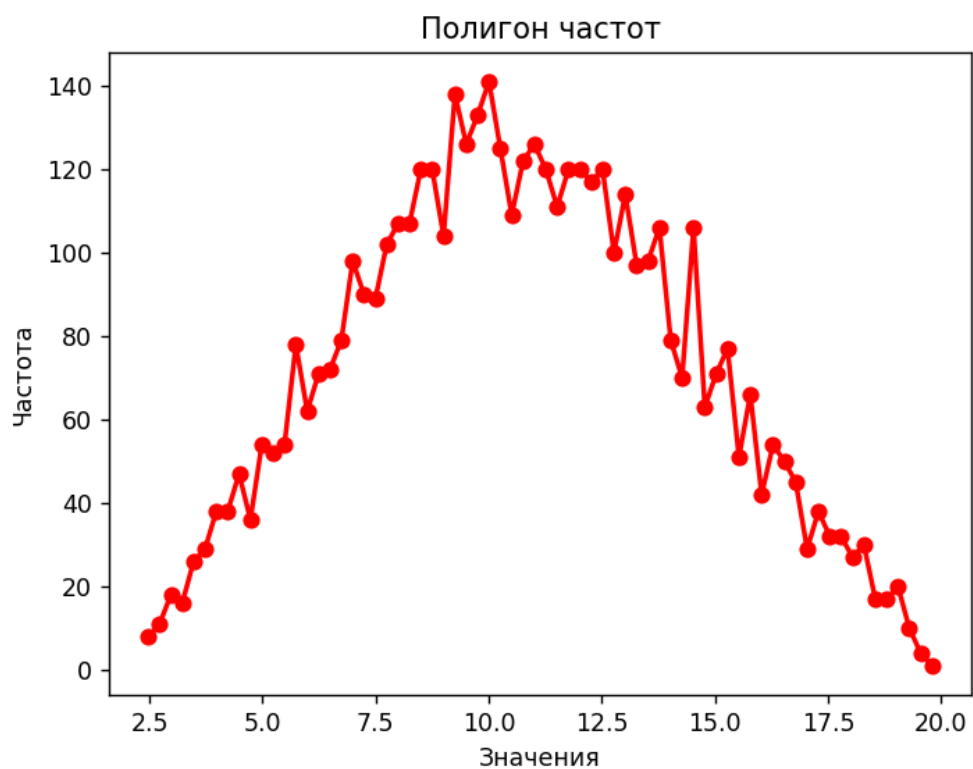


Рисунок 19 – Гистограмма выборки для 2 части

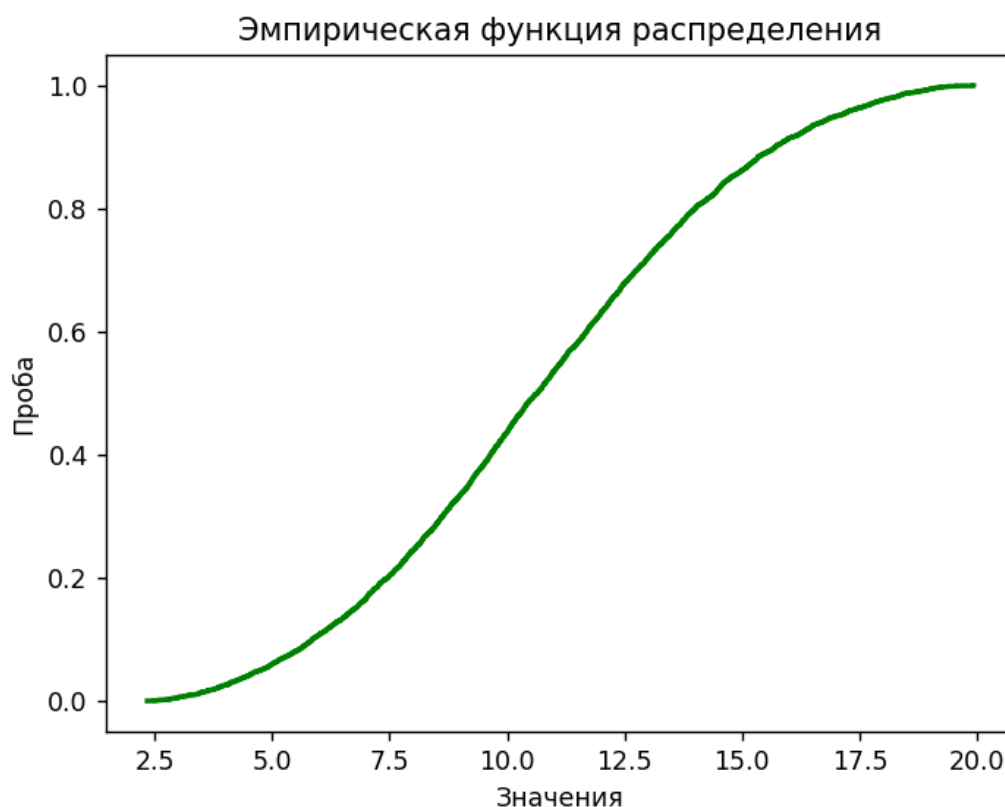


Рисунок 20 – ЭФР для 1 части

```

Количество интервалов: 70
Размах 17.58565640518573
Длина интервала 0.2512236629312247
Выборочное среднее: 10.732245847296818
Выборочная дисперсия: 13.37588733045089
Среднеквадратическое отклонение: 3.657306020891729
Мода: 2.34424316589137
Медиана: 10.64298492195425
Коэффициент асимметрии: 0.08706783512362702
Коэффициент эксцесса (без коррекции): -0.608756878966727

```

Рисунок 21 – Результаты подсчетов для 2 части

Среднее (10.73) и медиана (10.64) очень близки, что говорит об относительной симметричности распределения.

Мода (2.34) сильно отличается от среднего и медианы, что может свидетельствовать о наличии второго пика или скошенности распределения.

Размах (17.59) указывает на значительный диапазон значений.

Дисперсия (13.38) и стандартное отклонение (3.66) показывают, что данные довольно широко разбросаны вокруг среднего.

Коэффициент асимметрии близок к нулю, что означает почти симметричное распределение.

Коэффициент эксцесса (-0.61) меньше 0, что говорит о более "плоском" распределении по сравнению с нормальным (распределение имеет менее выраженный пик и более толстые хвосты).

На рисунке 22 показан график QQ-plot для второй выборки.

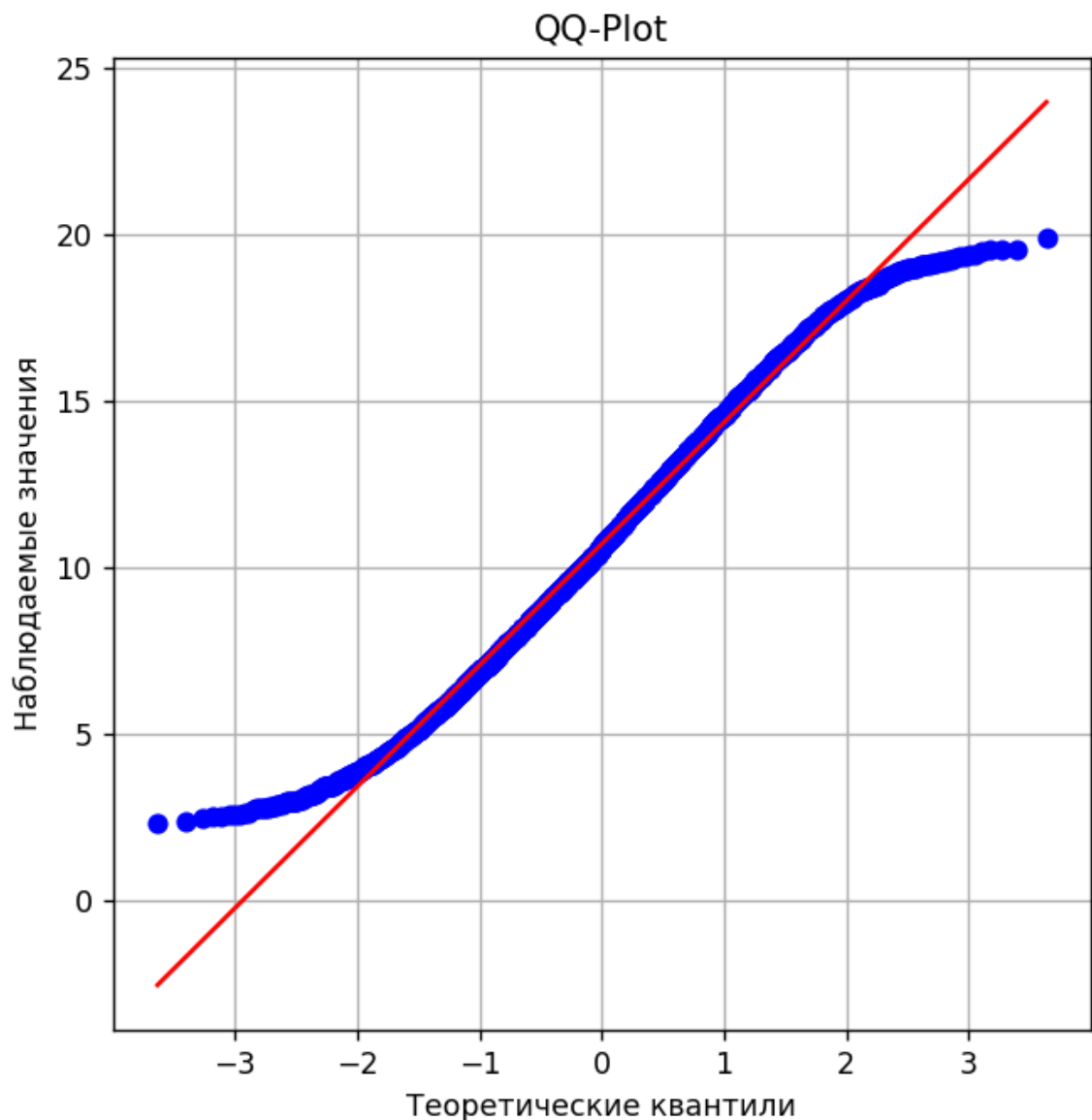


Рисунок 22 – График QQ-Plot для второй части

Вывод

В ходе работы были получены навыки анализа выборок, их графического отображения и вычисления метрик для определения типа распределения. Первая выборка имеет положительный коэффициент асимметрии (0.43) и небольшой положительный эксцесс (0.21), что указывает на скошенность вправо и остроконечность распределения. QQ-Plot показывает отклонения от нормального распределения, особенно в хвостах. Первая выборка имеет логнормальное распределение.

Вторая выборка характеризуется почти нулевой асимметрией (0.087) и отрицательным эксцессом (-0.61), что говорит о симметричности и более «плоском» распределении по сравнению с нормальным. График QQ-plot опроверг гипотезу о нормальности, распределение треугольное.