

Санкт-Петербургский
Политехнический университет Петра Великого

**Отчет по лабораторной работе
по дисциплине
"Математическая статистика"**

Студент:	Табакова Виктория Олеговна
Преподаватель:	Васильчук Владимир Юрьевич
Группа:	5030102/10401

Санкт-Петербург
2024

Содержание

1	Постановка задачи	2
2	Теоретическое обоснование	3
3	Описание работы	4
4	Результаты	5
5	Выводы	15

1 Постановка задачи

Имеется выборка некоторой СВ ξ в виде интервального статистического ряда (табл.). Требуется:

1. Построить гистограмму и график эмпирической функции распределения $F_n(x)$;
2. Вычислить выборочные: среднее, дисперсию, медиану, коэффициент вариации, коэффициент асимметрии, эксцесс;
3. Добавить искусственно к данным большую флуктуацию (порядка 1000). Как изменятся вычисленные параметры? Почему?

1.

Интервал	(21; 23)	(23; 25)	(25; 27)	(27; 29)	(29; 31)
Частота	30	70	65	30	5

2.

Интервал	(40; 45)	(45; 50)	(50; 55)	(55; 60)	(60; 65)
Частота	50	100	105	40	5

3.

Интервал	(100; 105)	(105; 110)	(110; 115)	(115; 120)	(120; 125)
Частота	45	105	100	40	10

4.

Интервал	(10; 15)	(15; 20)	(20; 25)	(25; 30)	(30; 35)
Частота	60	140	135	55	10

5.

Интервал	(80; 90)	(90; 100)	(100; 110)	(110; 120)	(120; 130)
Частота	80	165	170	65	20

6.

Интервал	(140; 145)	(145; 150)	(150; 155)	(155; 160)	(160; 165)
Частота	60	140	135	55	10

7.

Интервал	(170; 185)	(185; 200)	(200; 215)	(215; 230)	(230; 245)
Частота	80	165	170	65	20

8.

Интервал	(490; 495)	(495; 500)	(500; 505)	(505; 510)	(510; 515)
Частота	110	240	235	95	20

9.

Интервал	(130; 150)	(150; 170)	(170; 190)	(190; 210)	(210; 230)
Частота	95	200	205	80	20

10.

Интервал	(150; 175)	(175; 200)	(200; 225)	(225; 250)	(250; 275)
Частота	110	240	235	95	20

2 Теоретическое обоснование

- Выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

- Выборочная медиана

$$med\ x = \begin{cases} x_{(l+1)} & \text{при } n = 2l + 1 \\ \frac{x_{(l)} + x_{(l+1)}}{2} & \text{при } n = 2l \end{cases} \quad (2)$$

- Выборочная дисперсия

$$D(z) = \overline{z^2} - \bar{z}^2 \quad (3)$$

- Выборочное стандартное отклонение

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (4)$$

- Выборочный коэффициент вариации

$$CV = \sigma / \bar{x} \quad (5)$$

- Выборочный коэффициент асимметрии

$$A = \frac{n}{(n-1)(n-2)} \times \Sigma \left(\frac{(x_i - \bar{x})}{\sigma} \right)^3 \quad (6)$$

- Выборочный эксцесс

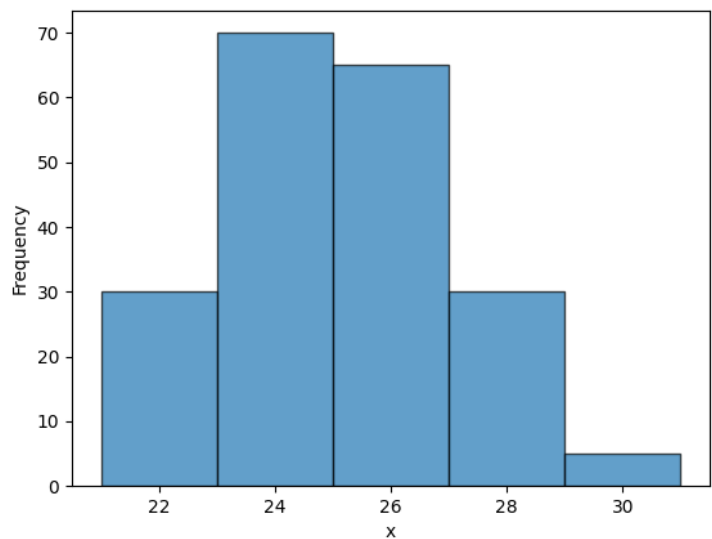
$$E = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \times \Sigma \left(\frac{(x_i - \mu)}{\sigma} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (7)$$

3 Описание работы

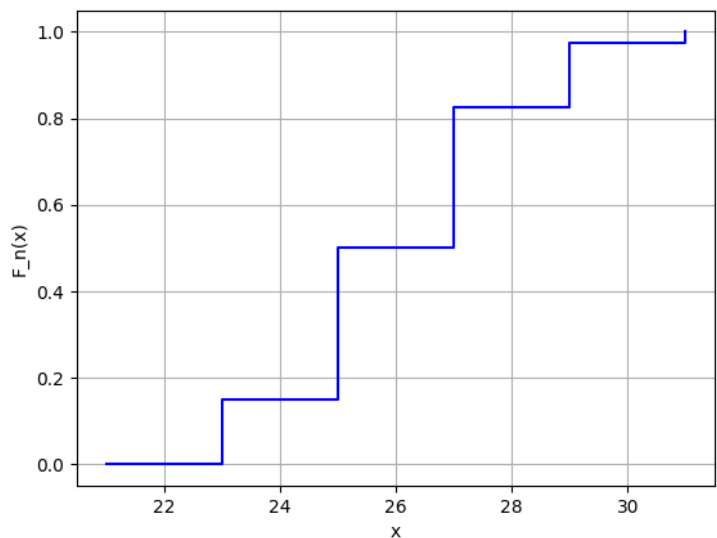
Лабораторные работы выполнены с использованием Python и его сторонних библиотек `numpy`, `pandas`, `matplotlib`.

Ссылка на GitHub репозиторий: <https://github.com/VictoriaTabakova/mathematical-statics>

4 Результаты

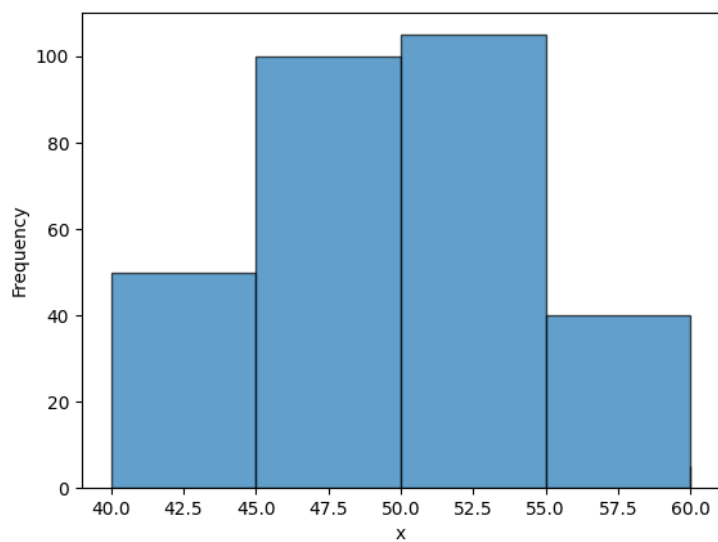


(а) Гистограмма

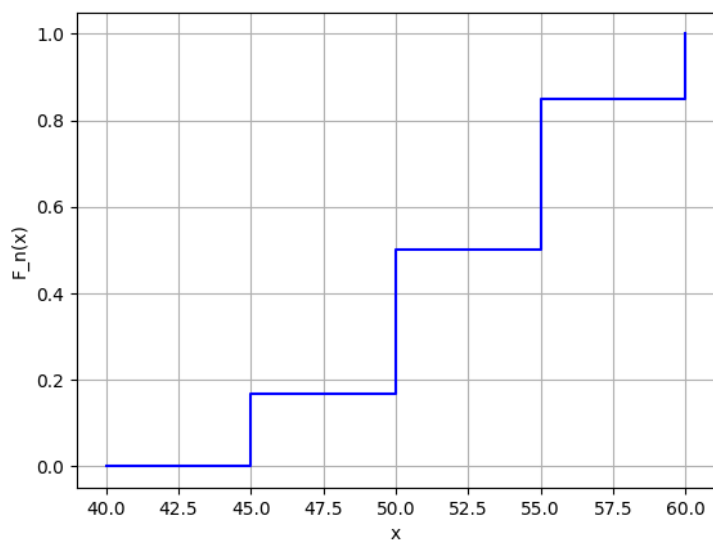


(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	25.1	25.0	3.99	0.08	0.24	-0.51
С флуктуациями	837.52	1000.0	132004.83	0.43	-1.79	1.2

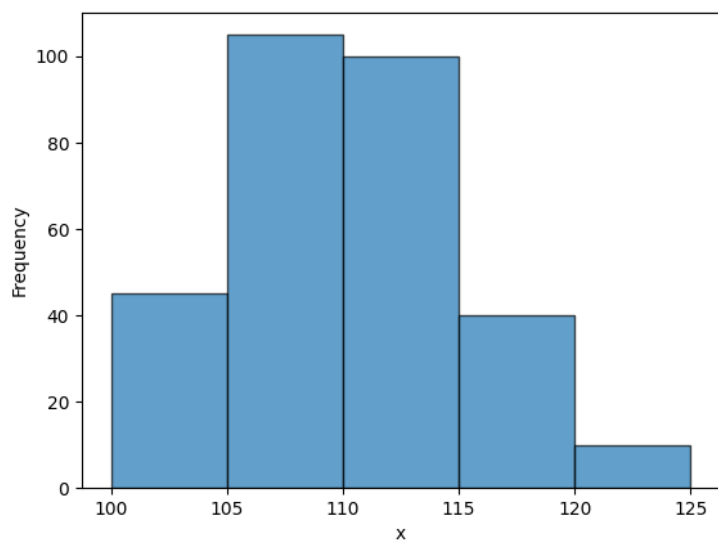


(a) Гистограмма

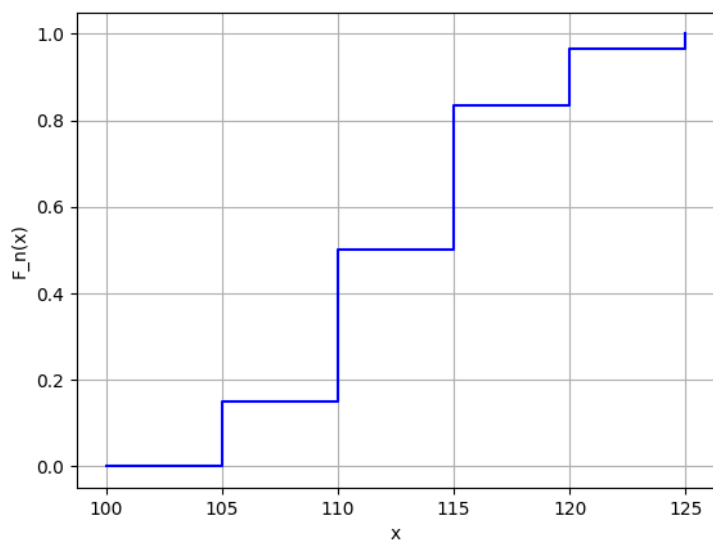


(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	49.96	50.0	22.81	0.1	0.05	-0.8
С флуктуациями	780.76	1000.0	160226.42	0.51	-1.28	-0.37

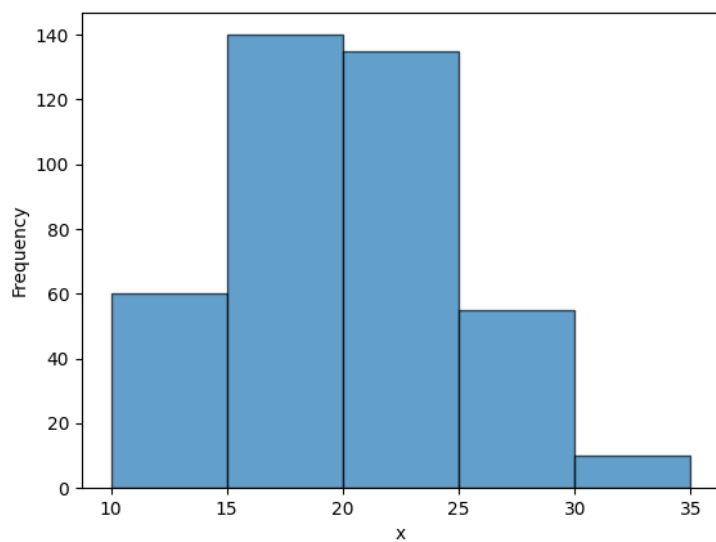


(a) Гистограмма

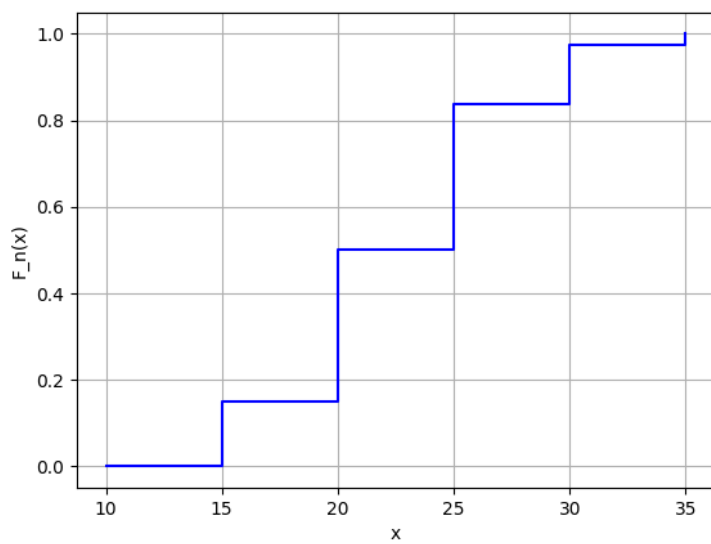


(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	110.25	110.0	25.35	0.05	0.3	-0.37
С флуктуациями	794.67	1000.0	140536.34	0.47	-1.28	-0.37

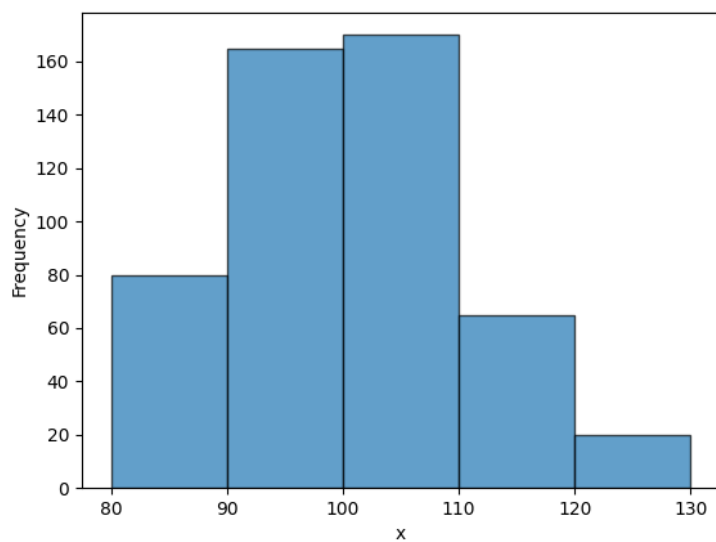


(a) Гистограмма

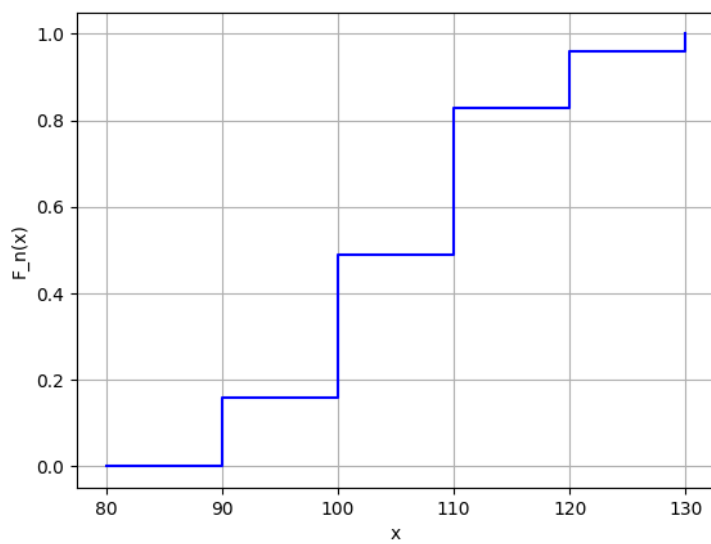


(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	110.25	110.0	25.35	0.05	0.3	-0.37
С флуктуациями	794.67	1000.0	140536.34	0.47	-1.28	-0.37

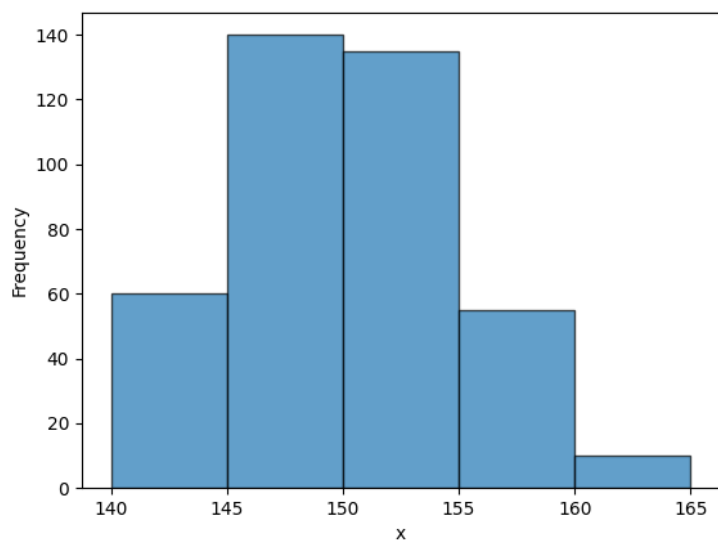


(a) Гистограмма

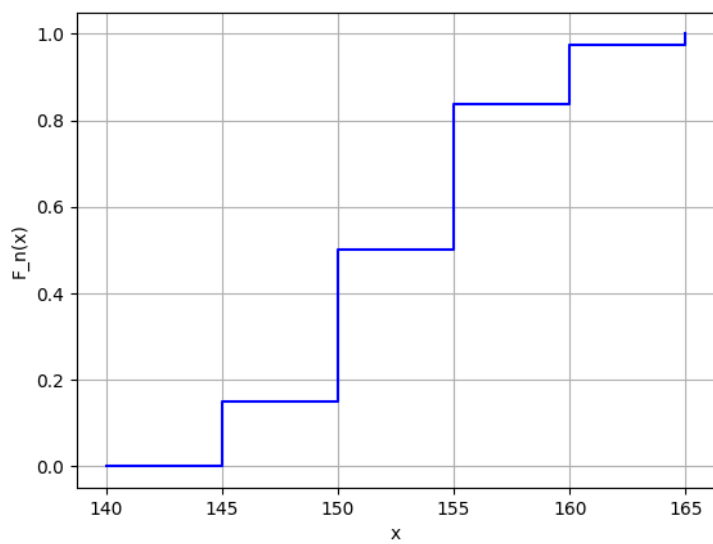


(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	100.6	105.0	106.64	0.1	0.3	-0.39
С флуктуациями	700.2	1000.0	179795.63	0.61	-0.71	-1.5

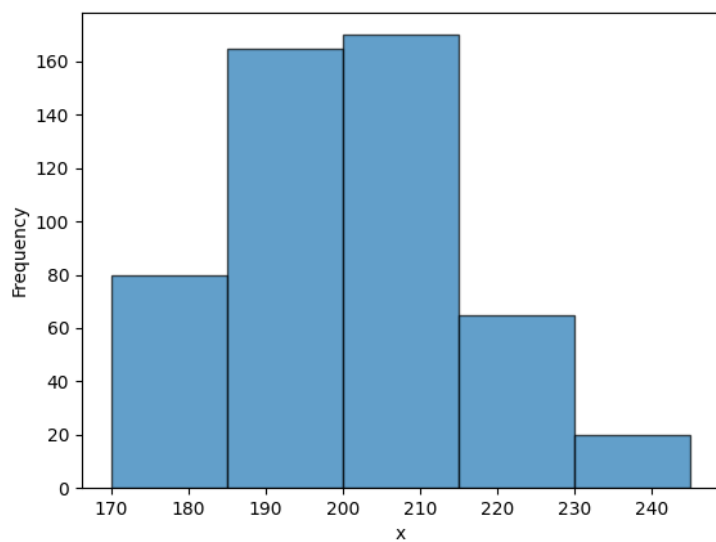


(a) Гистограмма

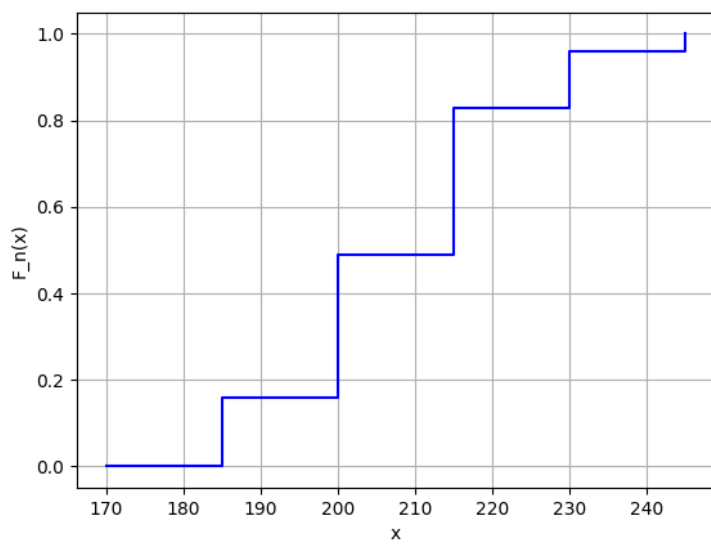


(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	150.19	150.0	24.34	0.03	0.25	-0.44
С флуктуациями	757.2	1000.0	147390.89	0.51	-0.95	-1.1

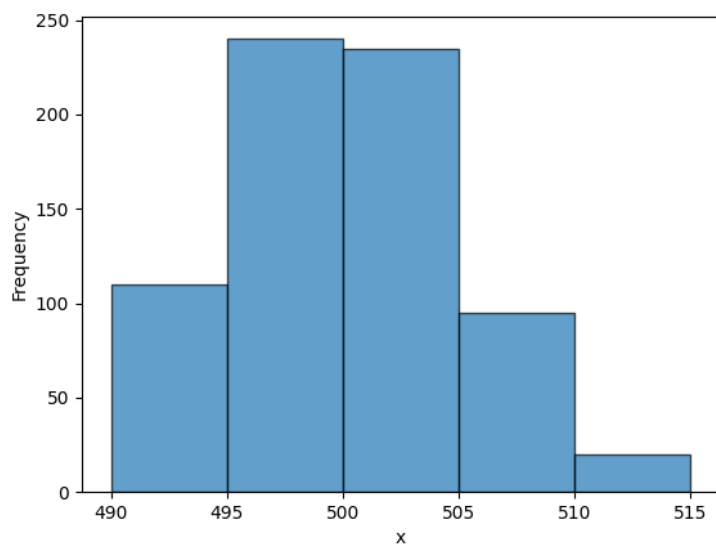


(a) Гистограмма

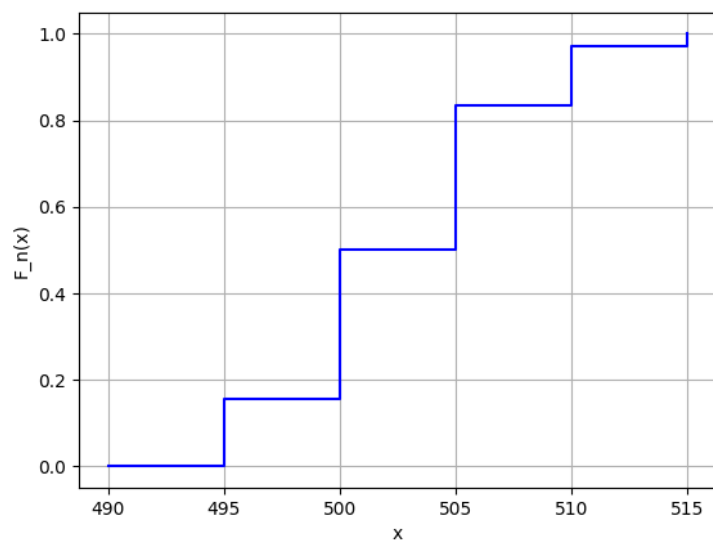


(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	200.9	207.5	239.94	0.08	0.3	-0.39
С флуктуациями	733.63	1000.0	141982.38	0.51	-0.71	-1.49

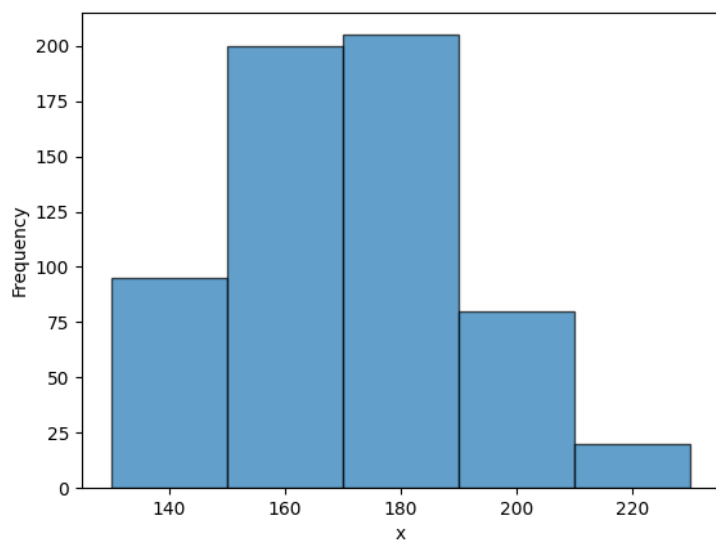


(a) Гистограмма

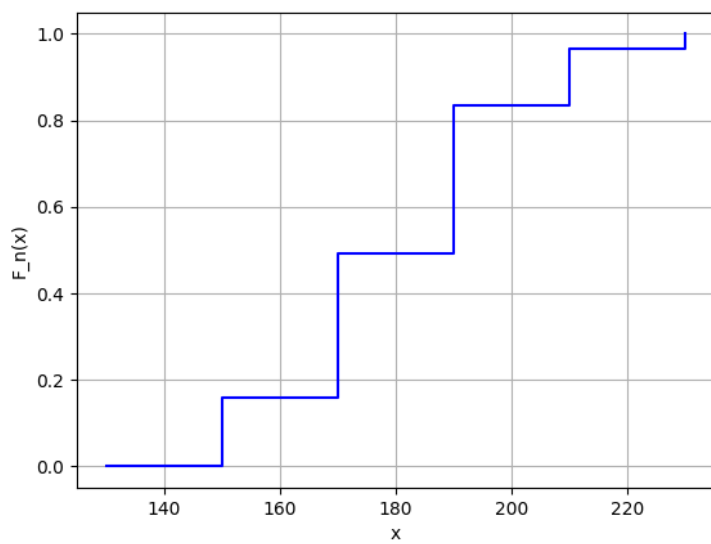


(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	500.18	500.0	25.15	0.01	0.26	-0.43
С флуктуациями	794.19	1000.0	60520.74	0.31	-0.36	-1.87

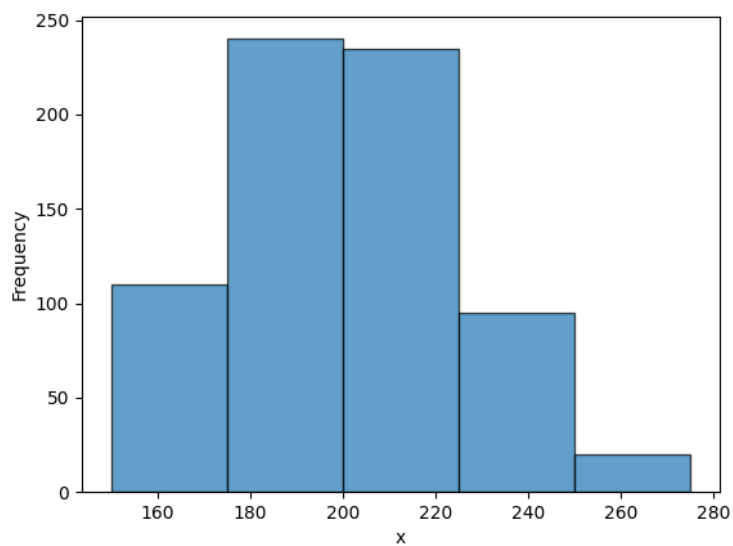


(a) Гистограмма

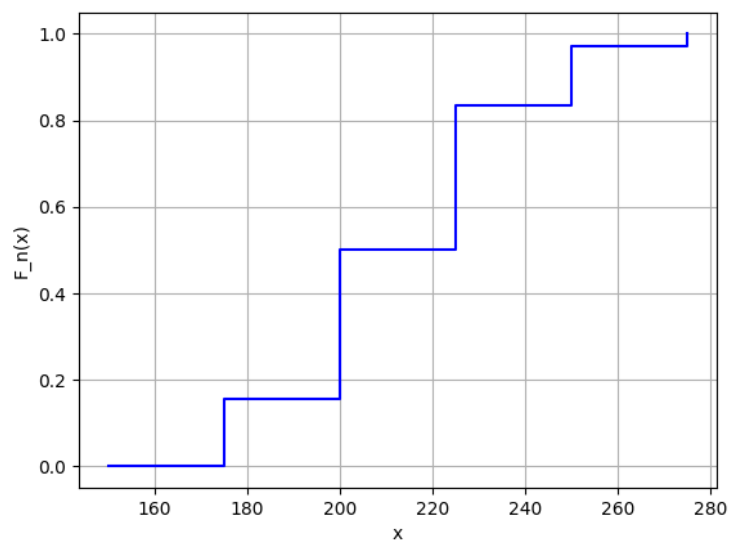


(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	171.0	180.0	412.33	0.12	0.27	-0.41
С флуктуациями	689.12	1000.0	161226.73	0.58	-0.52	-1.73



(a) Гистограмма



(b) Эмпирическая функция распределения

	\bar{x}	$medx$	$D(x)$	CV	A	E
Без флуктуаций	200.89	200.0	628.67	0.12	0.26	-0.43
С флуктуациями	670.96	1000.0	154930.34	0.59	-0.36	-1.86

5 Выводы

В ходе проведения лабораторной работы вы обнаружите, что выборочные значения, такие как среднее значение, дисперсия, медиана, коэффициент вариации, коэффициент асимметрии и эксцесс, дают важное понимание свойств вашего распределения данных. Гистограмма и график эмпирической функции распределения $F_n(x)$ также являются полезными инструментами для визуализации распределения ваших данных.

Добавление искусственной флуктуации к данным, особенно большой (порядка 1000), приводит к значительным изменениям во всех ваших статистических показателях. Среднее значение и дисперсия, вероятно, увеличатся значительно, так как они чувствительны к отклонениям и изменениям в данных. Коэффициент вариации также увеличится, отражая увеличившуюся вариативность данных. Коэффициенты асимметрии и эксцесса также могут значительно измениться, изменяя форму распределения.

Эти изменения объясняются тем, что введенная флуктуация является значительной по отношению к текущим данным, что приводит к резкому увеличению разброса данных. Это подчеркивает важность устойчивых и высококачественных данных при проведении статистического анализа, поскольку аномальные значения могут сильно исказить наши оценки и выводы.