

Санкт-Петербургский Государственный Университет

Поляков Иван Михайлович

Отчёт по Лабораторной Работе № 1_1

Основы обработки статистических данных

Направление 01.04.02: «Прикладная математика и информатика»
Образовательная программа ВМ.5505.2021: «Математическое и информационное
обеспечение экономической деятельности»

Преподаватель:
доктор технических наук,
профессор Буре Владимир Мансурович

Санкт-Петербург
2022 г.

1 Постановка задачи

Пусть имеется выборка, смоделированная из двух выборок нормально распределённых случайных величин по 80 и 60 элементов соответственно. Необходимо вычислить выборочные характеристики, а также проанализировать полученную выборку на однородность.

Пусть имеется выборка из 140 значений нормально распределённой случайной величины. Необходимо вычислить корреляцию данной выборки.

1.1 Выборочные характеристики

Изначально было сгенерировано две выборки объёмом 60 и 80 значений из нормально распределённых генеральных совокупностей с разными случайными значениями параметров математического ожидания и среднеквадратического отклонения. Их этих двух выборок была сформирована обобщённая выборка (объёмом 140 значений), которая была упорядочена по возрастанию её значений с помощью сортировки.

Вручную были получены основные выборочные характеристики. Коэффициент вариации составляет 385%, что говорит о большом разбросе значений и, как следствие, о неоднородности выборки. Коэффициент асимметрии достаточно близок к нулю, что говорит о не большом отклонении от пика распределения (математического ожидания). Коэффициент эксцесса показывает, что данное распределение обладает более пологим пиком, чем эмпирическое.

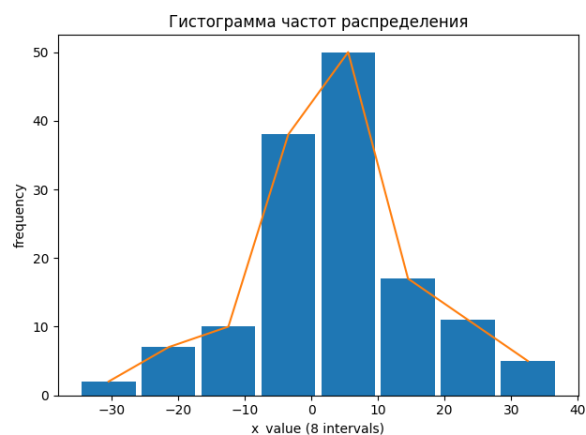
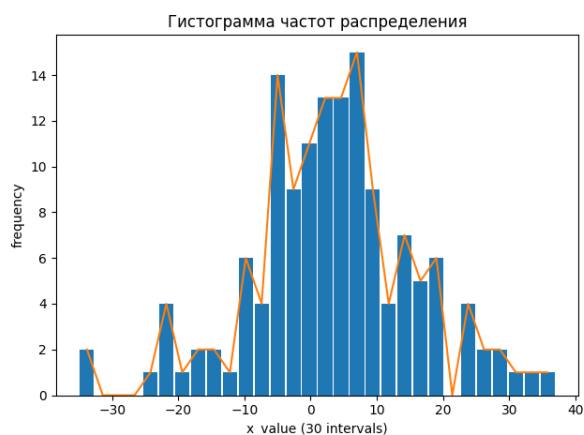
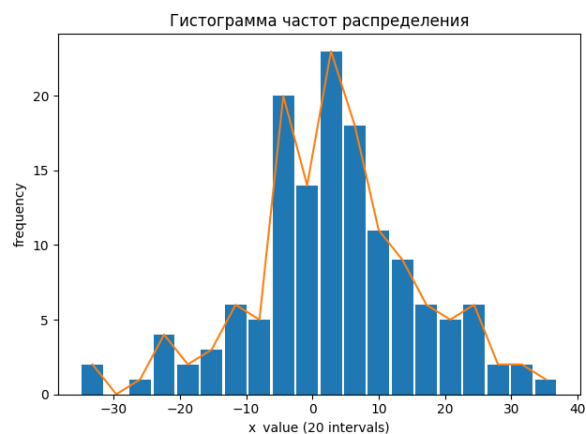
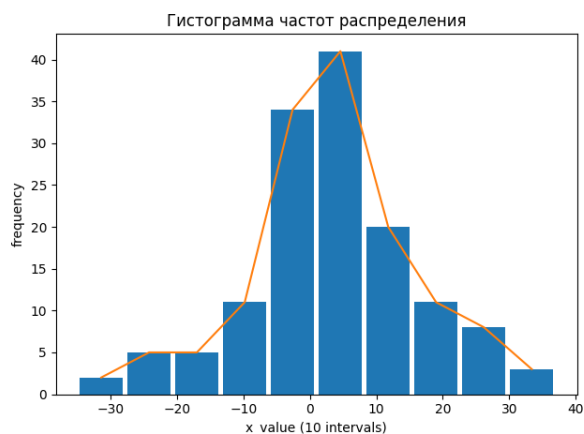
Сгенерированная выборка и полученные выборочные характеристики представлены в конце отчёта.

Было произведено разбиение размаха выборки на конечное число непересекающихся интервалов. Число интервалов разбиения, согласно постановке задачи, производилось по следующим величинам:

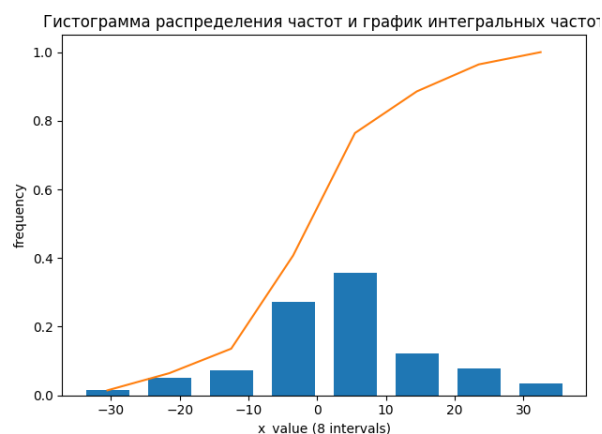
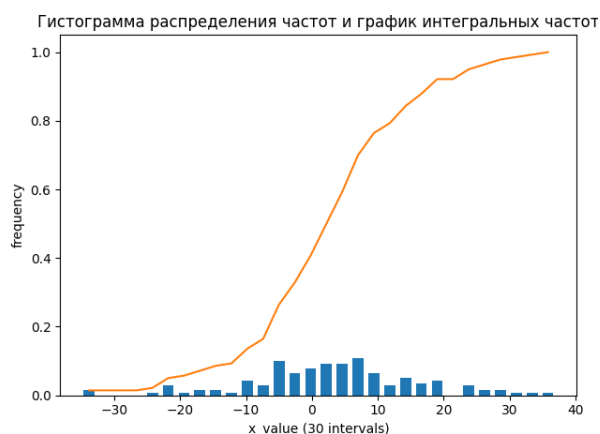
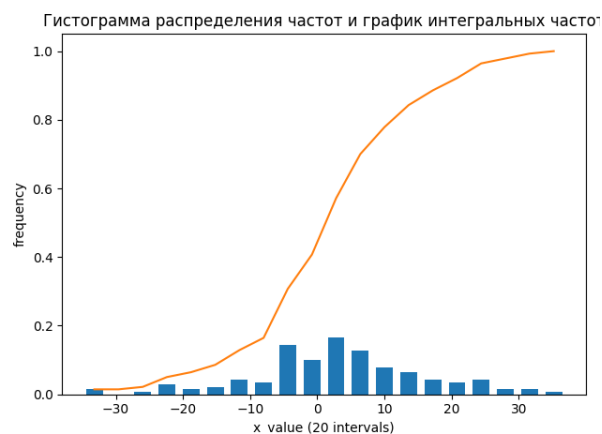
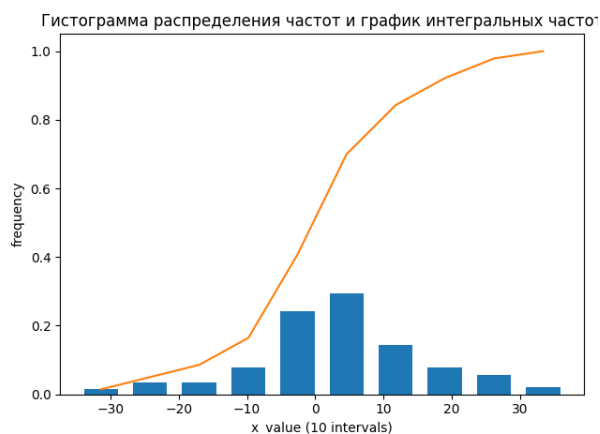
- $r = 10$;
- $r = 20$;
- $r = 30$;
- $r = [1 + 3.2 \lg n]$, где n - объём выборки.

Таким образом, для каждого r были построены:

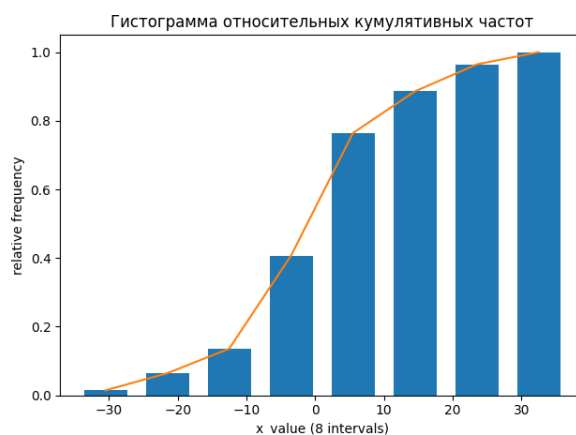
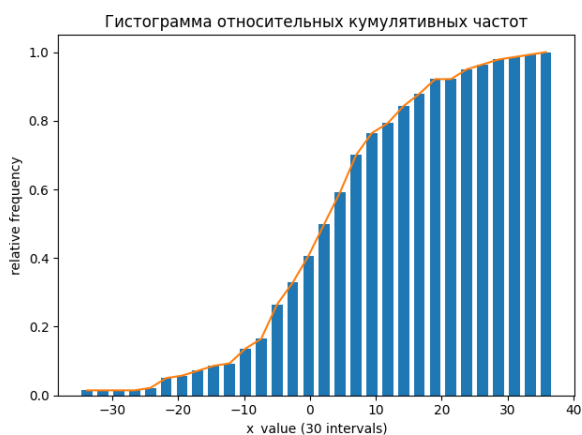
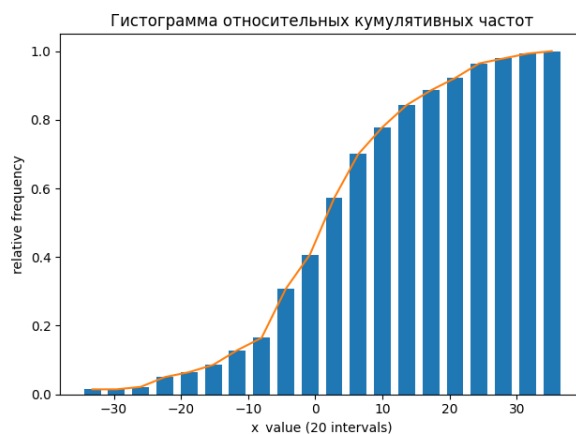
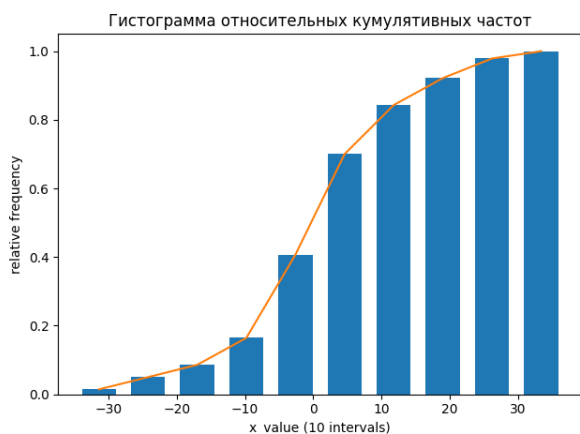
- Гистограмма частот распределения;



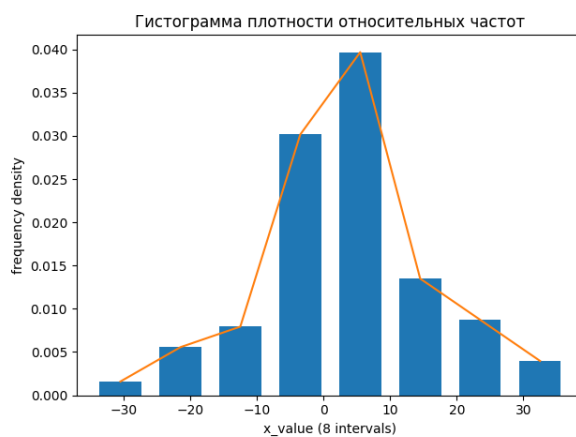
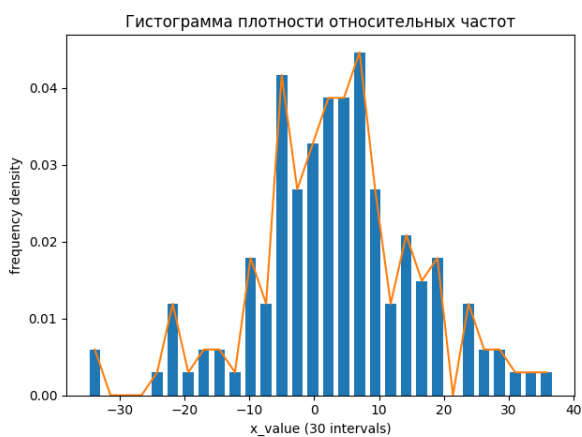
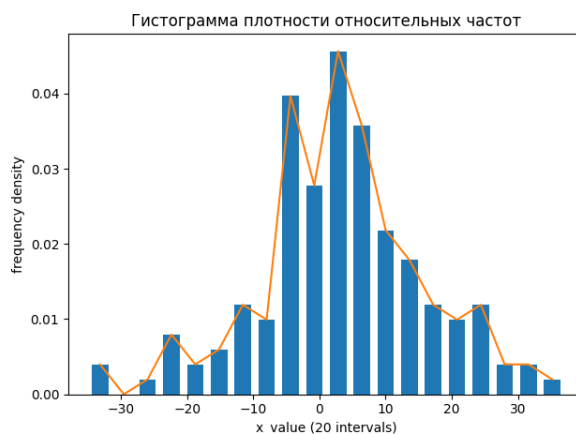
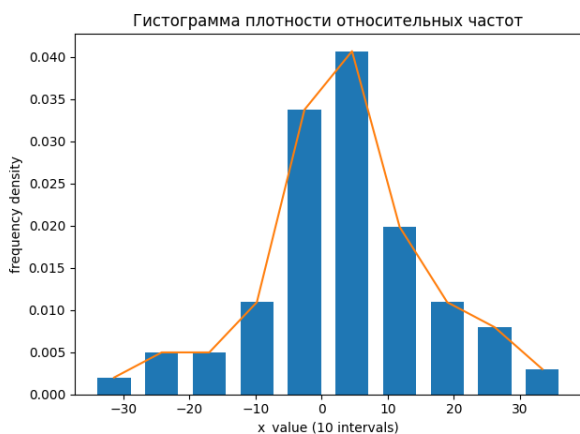
- Гистограмма распределения частот и график интегральных частот;



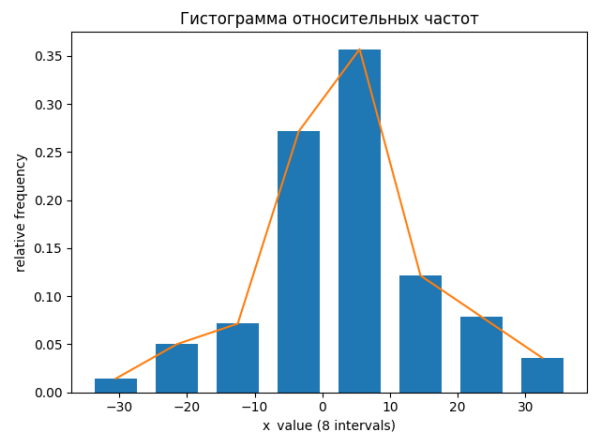
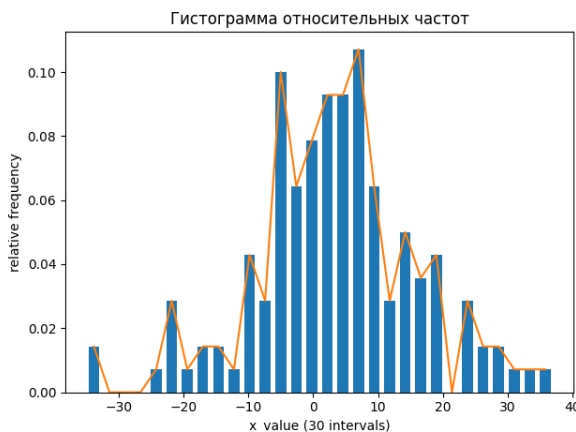
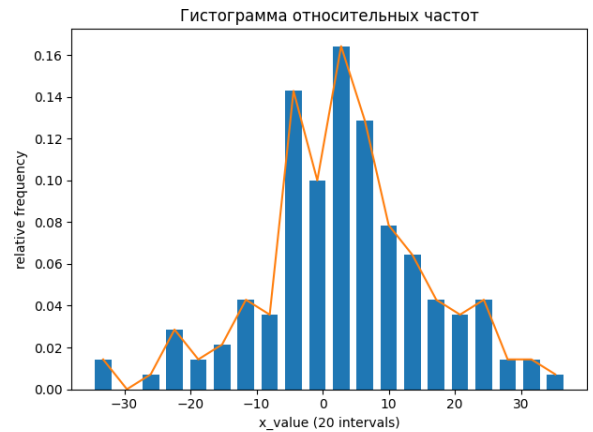
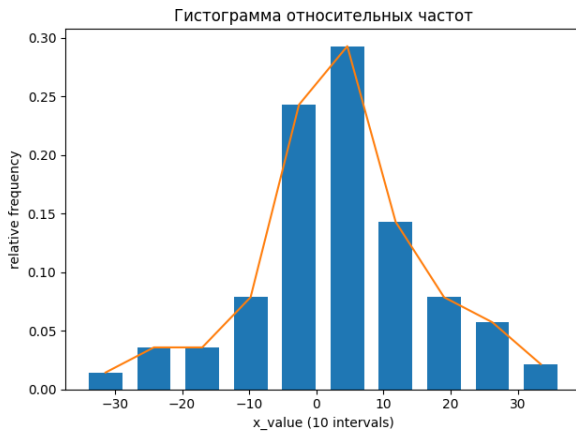
- Гистограмма относительных кумулятивных частот;



- Гистограмма плотности относительных частот;



- Гистограмма относительных частот.



При малых интервалах разбиения ($r = 8, r = 10$) нельзя утверждать о неоднородности выборки, однако при увеличении числа интервалов становится явно видно 2 пика, которые и могут послужить аргументом для утверждения неоднородности выборки. Исходя из описательной статистики, можно также сделать вывод о наличии существенной неоднородности в исходных данных.

Таким образом, в ходе работы были вычислены основные статистические характеристики, построены гистограммы, полигоны плотности и кумулятивные частоты и на основе всего вышесказанного сделаны выводы.

1.2 Корреляция

Были сформированы 3 выборки следующим образом:

1. выборка со значениями равномерно распределённой случайной величины из отрезка $[0, 1]$;
2. выборка, полученная из предыдущей путём нахождения квантилей нормального распределения со случайными значениями математического ожидания и стандартного отклонения;
3. выборка, сформированная с помощью стандартных средств Python.

В ходе работы были вычислены выборочные коэффициенты корреляции для всех трёх полученных выборок, была проверена статистическая значимость каждого коэффициента на уровне значимости 5% с помощью t_r -статистики с числом степеней свободы $n - 2$:

$$t_r = \frac{r}{\sqrt{1 - r^2} \sqrt{n - 2}} \quad (1)$$

Также были вычислены выборочные ковариации, средние квадратические отклонения для выборок и построена корреляционная матрица 3×3 . Результаты данной и предыдущей работ представлены ниже в качестве сформированного вывода в консоль Python.

Из полученных значений видно, что между первой и второй выборкой имеется сильная корреляция (равна почти 1), тогда как выборки 2, 3 и 1, 3 корреляционных связей почти не имеют.

Выборочные характеристики

Параметры для формирования выборок:

```
a_1      : 6
a_2      : 3
sigma_1   : 13
sigma_2   : 11
```

Выборка:

```
[[-35 -33 -25 -22 -22 -21 -21]
 [-19 -18 -16 -15 -14 -13 -11]
 [-11 -10 -10 -10 -9 -8 -7]
 [ -7 -7 -6 -6 -6 -6 -5]
 [ -5 -5 -4 -4 -4 -4 -4]
 [ -4 -4 -3 -3 -3 -3 -3]
 [ -3 -2 -2 -2 -1 -1 -1]
 [ -1  0  0  0  0  0  0]
 [  0  1  1  2  2  2  2]
 [  2  2  2  2  3  3  3]
 [  4  4  4  4  4  4  4]
 [  4  4  4  5  5  5  6]
 [  6  6  7  7  7  7  8]
 [  8  8  8  8  8  8  8]
 [  9  9  9  9  9  9  9]
 [  9  9 11 11 12 12 13]
 [ 13 13 14 14 14 15 16]
 [ 16 16 16 17 18 19 19]
 [ 19 20 20 23 23 24 24]
 [ 26 26 28 29 31 33 37]]
```

```
Максимум          : 37.00
Минимум           : -35.00
Число наблюдений   : 140.00
Среднее значение    : 3.26
Медиана            : 3.50
Мода               : 4.00
Размах             : 72.00
Несмещённая выборочная дисперсия : 157.46
Смещённая выборочная дисперсия    : 156.33
Несмещённое стандартное отклонение : 12.55
Смещённое стандартное отклонение    : 12.50
Среднее абсолютное отклонение        : 9.43
Эксцесс                             : 0.66
Ассиметрия                           : -0.13
Коэффициент вариации                  : 3.85
Ошибка выборки                        : 1.06
```

Корреляция

Параметры для формирования выборок:

a : 2

sigma: 8

Выборочные коэффициенты корреляции:

r1 = 0.973

r2 = -0.043

r3 = -0.019

Статистическая значимость вычисленных парных коэффициентов корреляции:

1. Гипотеза о равенстве нулю корреляции отклоняется на уровне значимости 0.05, так как $|49.969| > 1.960$
2. Гипотеза о равенстве нулю корреляции принимается на уровне значимости 0.05, так как $|-0.501| \leq 1.960$
3. Гипотеза о равенстве нулю корреляции принимается на уровне значимости 0.05, так как $|-0.228| \leq 1.960$

Выборочные ковариации для пар выборок:

cov(1, 2) = 2.289

cov(2, 3) = -0.703

cov(3, 1) = -0.011

Выборочные средние квадратические отклонения:

sd1 = 0.286

sd2 = 8.223

sd3 = 2.006

Корреляционная матрица 3x3:

```
[[ 1.    0.97 -0.02]
 [ 0.97  1.   -0.04]
 [-0.02 -0.04  1.   ]]
```