

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра математического обеспечения и применения ЭВМ

ОТЧЕТ
по лабораторной работе №1
по дисциплине «Статистические методы обработки экспериментальных
данных»
Тема: Формирование и первичная обработка выборки. Ранжированный
и интервальный ряды

Студент гр. 7381

Павлов А.П.

Студент гр. 7381

Габов Е.С.

Преподаватель

Сучков А.И.

Санкт-Петербург

2021

Цель работы.

Ознакомление с основными правилами формирования выборки и подготовки выборочных данных к статистическому анализу.

Основные теоретические положения.

Статистический ряд – последовательность элементов выборки, расположенных в порядке их получения (наблюдения).

Ранжированный ряд – последовательность элементов выборки, расположенных в порядке возрастания их значений. Номер элемента ранжированного ряда в последовательности называется рангом.

Вариационный ряд – получается из ранжированного ряда в результате объединения одинаковых элементов. Элементы вариационного ряда называются вариантами.

Варианта – отдельные значения признака, по которому производится группировка.

Частота – число, показывающее, как часто встречается та или иная варианта. Сумма всех абсолютных частот равна общему числу наблюдений, относительных – единице.

Полигон частот – это один из способов графического представления плотности вероятности распределения выборки.

Гистограмма – это наглядное представление функции вероятности некоторой случайной величины, построенное по выборке. Гистограмма строится с помощью интервального ряда.

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$.

График $F^*(x)$ представляет собой лестничный график, длина каждой ступеньки которого равна длине соответствующего интервала, а высота – отношению накопленной частоты до середины этого интервала к объёму выборки, т.е.:

$$F^*(\tilde{x}_i) = \frac{m_i^{\text{нак}}}{N} = \frac{\sum_{j=1}^{i-1} m_j}{N}; i = 1; 2; \dots; k$$

Постановка задачи.

Осуществить формирование репрезентативной выборки заданного объема из имеющейся генеральной совокупности экспериментальных данных. Осуществить последовательное преобразование полученной выборки в ранжированный, вариационный и интервальный ряды. Применительно к интервальному ряду построить и отобразить графически полигон, гистограмму и эмпирическую функцию распределения для абсолютных и относительных частот. Полученные результаты содержательно проинтерпретировать.

Выполнение работы.

Для выполнения лабораторной работы выбран язык программирования Python. Код программы представлен в приложении А.

Произведена выборка из представленной генеральной совокупности экспериментальных данных размером 107 элементов. Результат представлен на рис. 1. Выборка сформирована из генеральной совокупности путем случайного выбора необходимого количества элементов.

Сформированная выборка:

```
29 40 66 370 40 22 36 58 32 45 26 19 52 489 24 77 14 27 22 84 72
16 18 30 237 12 76 277 27 17 20 38 22 66 30 35 24 23 21 27 21 20
7 32 134 71 38 66 32 72 397 636 105 60 132 92 45 16 36 51 35 318 54
36 70 26 111 62 22 132 18 368 172 269 14 116 67 248 45 138 33 144 11 86
62 64 33 465 50 326 31 16 40 25 32 367 18 13 38 49 18 44 8 61 75
22 40
```

Рисунок 1 – Сформированная выборка

Ранжированный ряд представлен на рис. 2.

Ранжированный ряд:

```
7 8 11 12 13 14 14 16 16 16 17 18 18 18 18 19 20 20 21 21 22
22 22 22 22 23 24 24 25 26 26 27 27 27 29 30 30 31 32 32 32 32
33 33 35 35 36 36 36 38 38 38 40 40 40 40 44 45 45 45 49 50 51
52 54 58 60 61 62 62 64 66 66 66 67 70 71 72 72 75 76 77 84 86
92 105 111 116 132 132 134 138 144 172 237 248 269 277 318 326 367 368 370 397 465
489 636
```

Рисунок 2 – Ранжированный ряд

Вариационный ряд представлен на рис.3.

Вариационный ряд:

```
7 ==> 1 8 ==> 1 11 ==> 1 12 ==> 1 13 ==> 1 14 ==> 2 16 ==> 3 17 ==> 1 18 ==> 4 19 ==> 1 20 ==> 2
21 ==> 2 22 ==> 5 23 ==> 1 24 ==> 2 25 ==> 1 26 ==> 2 27 ==> 3 29 ==> 1 30 ==> 2 31 ==> 1 32 ==> 4
33 ==> 2 35 ==> 2 36 ==> 3 38 ==> 3 40 ==> 4 44 ==> 1 45 ==> 3 49 ==> 1 50 ==> 1 51 ==> 1 52 ==> 1
54 ==> 1 58 ==> 1 60 ==> 1 61 ==> 1 62 ==> 2 64 ==> 1 66 ==> 3 67 ==> 1 70 ==> 1 71 ==> 1 72 ==> 2
75 ==> 1 76 ==> 1 77 ==> 1 84 ==> 1 86 ==> 1 92 ==> 1 105 ==> 1 111 ==> 1 116 ==> 1 132 ==> 2 134 ==> 1
138 ==> 1 144 ==> 1 172 ==> 1 237 ==> 1 248 ==> 1 269 ==> 1 277 ==> 1 318 ==> 1 326 ==> 1 367 ==> 1 368 ==> 1
370 ==> 1 397 ==> 1 465 ==> 1 489 ==> 1 636 ==> 1
```

Рисунок 3 – Вариационный ряд

Для построения интервального ряда требуется рассчитать количество различных групп. Требуемые расчёты представлены на рис. 4. Построенный интервальный ряд представлен на рис. 5.

```
Количество интервалов определенное по формуле Стерджесса = 7
Минимальное значение в выборке = 7
Максимальное значение в выборке = 636
Размах выборки = 629
```

Рисунок 4 – Расчёты для интервального ряда

Интервальный ряд

Интервал	абс. част.	относ. част
7 - 96	85	0.794392523364486
96 - 185	9	0.08411214953271028
185 - 274	3	0.028037383177570093
274 - 363	3	0.028037383177570093
363 - 452	4	0.037383177570093455
452 - 541	2	0.018691588785046728
541 - 630	1	0.009345794392523364

Рисунок 5 – Интервальный ряд с частотами

На рис. 6 представлены построенные полигон и гистограмма абсолютных частот.

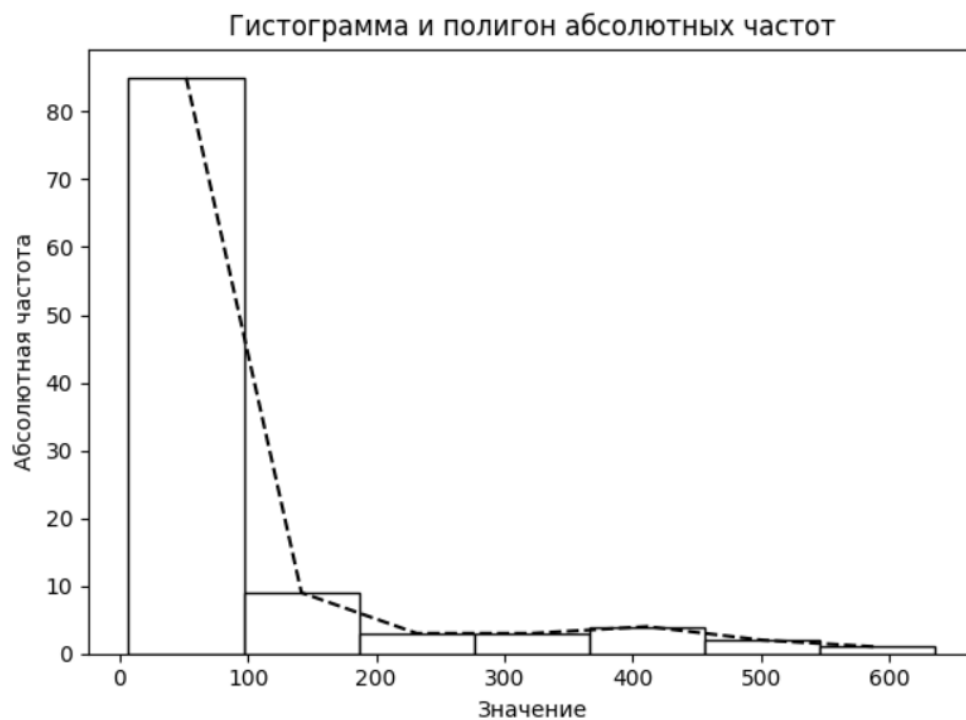


Рисунок 6 – Гистограмма и полигон абсолютных частот

На рис. 7 представлены построенные полигон и гистограмма относительных частот.

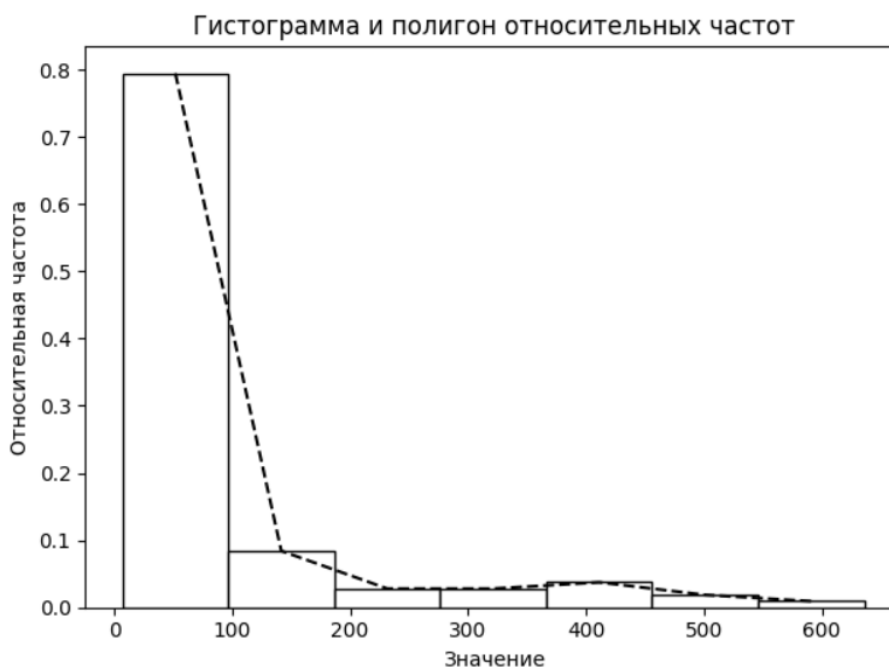


Рисунок 7 – Гистограмма и полигон относительных частот

Для интервального ряда абсолютных частот построена эмпирическая функция распределения. Построенная эмпирическая функция распределения представлена на рис. 8.

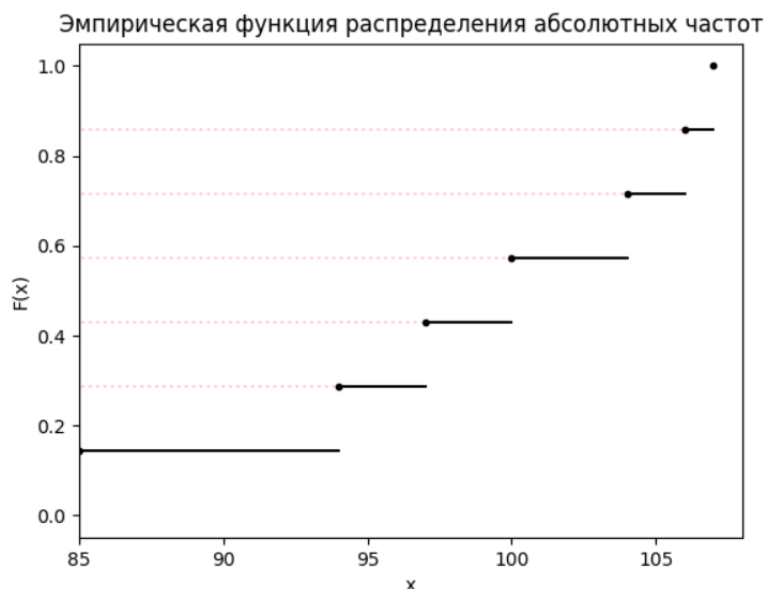


Рисунок 8 – Эмпирическая функция распределения для интервального ряда абсолютных частот

Для интервального ряда относительных частот построена эмпирическая функция распределения. Построенная эмпирическая функция распределения представлена на рис. 9.

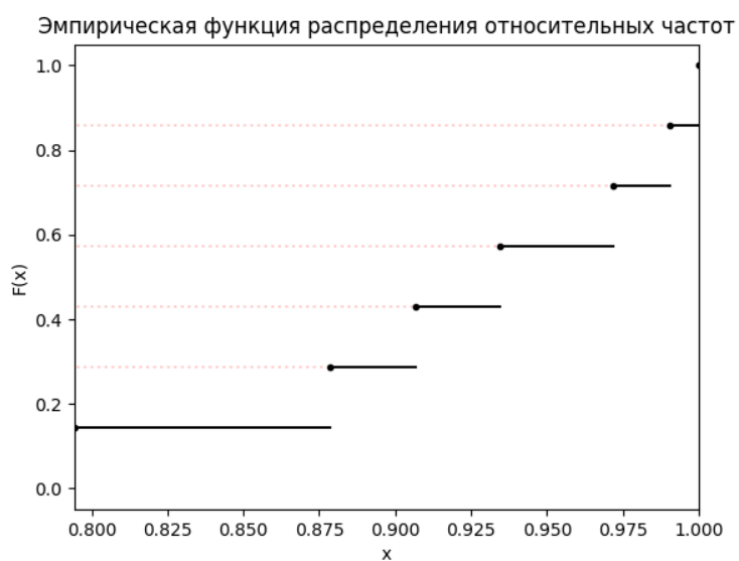


Рисунок 9 – Эмпирическая функция распределения для интервального ряда относительных частот

Выводы.

В ходе выполнения лабораторной работы был повторён и применён на практике материал с правилами формирования случайной выборки и подготовкой выборочных данных к статистическому анализу. В практической части были сделаны следующие этапы работы: формирование выборки, преобразование в ранжированный, вариационный и интервальный ряды, построение по интервальному ряду гистограмм, полигонов частот и эмпирической функции распределения.

ПРИЛОЖЕНИЕ А

ИСХОДНЫЙ КОД РАЗРАБОТАННОЙ ПРОГРАММЫ

```
import csv
import random
import math
from collections import Counter
import matplotlib.pyplot as plt
import numpy as np
from statsmodels.distributions.empirical_distribution import ECDF

random_seed = 79
data_elem_size = 107

def getDataFromFile(attribureNum):
    with open('machine.data', newline='') as csvfile:
        spamreader = csv.reader(csvfile, delimiter=',')
        data = []
        for row in spamreader:
            data.append(int(row[attribureNum]))
    return data

def generateSample(data):
    random.seed(random_seed)
    data = random.sample(data, data_elem_size)
    return data

def printData(data):
    count = 0
    for i in data:
        count = count + 1
        if(count > 20):
            count = 0
            print(i)
        else:
            print(i, end = " "),
    print()

data = getDataFromFile(8)
data = generateSample(data)
print("Сформированная выборка:")
printData(data)
print("Ранжированный ряд:")
printData(sorted(data))

countedData = Counter(sorted(data))
count = 0
print('Варианционный ряд:')
for i in sorted(set(data)):
    count = count + 1
```



```

    if (count > 10):
        count = 0
        print(f'{i:3} ==> {countedData[i]:1}')
    else:
        print(f'{i:3} ==> {countedData[i]:1}', end=" ")
print()

intervalNum = int(1 + 3.322 * math.log10(data_elem_size))
print('Количество интервалов определенное по формуле Стерджесса = ', intervalNum)
print('Минимальное значение в выборке = ', min(data))
print('Максимальное значение в выборке = ', max(data))
dataRange = max(data) - min(data)
intervalSize = int(dataRange / intervalNum)
print('Размах выборки = ', max(data) - min(data))

isInBucket = lambda x: min(int((abs(x) - min(data)) / dataRange * intervalNum), intervalNum-1)
borders = [(min(data) + dataRange/intervalNum*i, min(data) + dataRange/intervalNum*(i+1)) for i in range(intervalNum)]
buckets = [[] for i in range(intervalNum)]
for value in data:
    buckets[isInBucket(value)].append(value)

print('Интервальный ряд')
print('  Интервал    абс. част.    относ. част')
for i in range(0, intervalNum):
    print(f'{(min(data) + intervalSize*i):3} - {min(data) + intervalSize*(i+1):3}'
          f'{len(buckets[i]):10} {len(buckets[i])/len(data):25}')

fig, ax = plt.subplots()
ax.hist(data, bins=intervalNum, density=False, edgecolor='black', facecolor='white')
center_of_borders = [(border[0] + border[1])/2 for border in borders]
y = [len(bucket) for bucket in buckets]
ax.plot(center_of_borders, y, '--k')
ax.set_xlabel('Значение')
ax.set_ylabel('Абсолютная частота')
ax.set_title('Гистограмма и полигон абсолютных частот')
fig.tight_layout()
plt.show()

fig, ax = plt.subplots()
ax.hist(data, intervalNum, weights=np.ones(len(data)) / len(data), density=False, edgecolor='black', facecolor='white')
center_of_borders = [(border[0] + border[1])/2 for border in borders]
y = [len(bucket)/data_elem_size for bucket in buckets]
ax.plot(center_of_borders, y, '--k')
ax.set_xlabel('Значение')
ax.set_ylabel('Относительная частота')

```

```

ax.set_title('Гистограмма и полигон относительных частот')
fig.tight_layout()
plt.show()

maxBucketLen = 0
for bucket in buckets:
    if(maxBucketLen < len(bucket)/len(data)):
        maxBucketLen = len(bucket)/len(data)

absValues = []
sum = 0
print(sum)
for bucket in buckets:
    sum = sum + len(bucket)/len(data)
    absValues.append(sum)

ecdf = ECDF(absValues)
fig, ax = plt.subplots()
ax.set_xlabel('x')
ax.set_ylabel('F(x)')
ax.set_title('Эмпирическая функция распределения относительных
частот')
ax.axis(xmin=maxBucketLen, xmax=max(ecdf.x))
ax.axis(ymin=-0.05, ymax=1.05)

for i in range(len(ecdf.x)-1):
    xs = [0, ecdf.x[i]]
    ys = [ecdf.y[i]] * 2
    ax.plot(xs, ys, 'r:', alpha=0.2)
    ax.plot(ecdf.x[i], ecdf.y[i], "k.") # точки
    xs = [ecdf.x[i], ecdf.x[i+1]]
    ys = [ecdf.y[i]] * 2
    ax.plot(xs, ys, 'k-')
ax.plot(ecdf.x[-1], ecdf.y[-1], "k.")
plt.show()

```