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

ОТЧЕТ

по лабораторной работе №2
по дисциплине «Статистические методы обработки экспериментальных данных»
Тема: Обработка выборочных данных. Нахождение точечных оценок параметров распределения.

Студентка гр. 8382	____________________	Звегинцева Е.Н.
Студент гр. 8382	____________________	Мирончик П.Д.
Преподаватель	____________________	Середа А.-В.И.

Санкт-Петербург
2022

Цель работы.

Получение практических навыков нахождения точечных статистических оценок параметров распределения.

Основные теоретические положения.

Математическим ожиданием дискретной случайной величины называется сумма произведений ее возможных значений на соответствующие им вероятности:

$$M(X) = \sum_{i=1}^{n} x_i n_i$$

Дисперсией случайной величины называется математическое ожидание квадрата ее отклонения от ее математического ожидания:

$$D(X) = M\big(X - M(X)\big)^2$$

Среднеквадратическим отклонением случайной величины X (стандартом) называется квадратный корень из ее дисперсии:

$$\sigma = \sqrt{D(X)}$$

Асимметрией, или коэффициентом асимметрии, называется числовая характеристика, определяемая выражением:

$$A \quad = \frac{m_3}{S^3},$$

где m_3 – центральный эмпирический момент третьего порядка, S – исправленная выборочная дисперсия.

Центральным моментом порядка k случайной величины X называется математическое ожидание величины:

$$M\big(X - M(X)\big)^k = m_k.$$

Исправленная выборочная дисперсия определяется по формуле:

$$S^2 = \frac{N}{N-1} D_B,$$

где $D_B = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^2\, n_i$ – выборочная дисперсия.

Эксцессом называется численная характеристика случайной величины, которая определяется выражением:

$$E \quad = \frac{m_4}{S^4} - 3.$$

Для нормального закона $\frac{m_4}{S^4} = 3$. Отсюда следует, что для нормального закона $E = 0$. Эксцесс показывает, как быстро уменьшается плотность распределения вблизи её максимального значения.

Мода дискретной случайной величины – это наиболее вероятное значение этой случайной величины. Модой непрерывной случайной величины называется ее значение, при котором плотность вероятности максимальна.

$$M_O(X) = x_{M_O} + h\frac{(m_2 - m_1)}{(m_2 - m_1) + (m_2 - m_3)}$$

Медиана случайной величины X – это такое ее значение M_e, для которого выполнено равенство

$$\mathrm{P(X < Me) = P(X > Me)},$$

$$M_e(X) = x_{M_e} + h\frac{0{,}5n - SM_{e^{-1}}}{n_{M_e}}$$

Постановка задачи.

Для заданных выборочных данных вычислить с использованием метода моментов и условных вариант точечные статистические оценки математического ожидания, дисперсии, среднеквадратического отклонения, асимметрии и эксцесса исследуемой случайной величины. Полученные результаты содержательно проинтерпретировать.

Выполнение работы.

В данной работе используется интервальный ряд, полученный в первой лабораторной работе. Далее приведены значения, полученные при расчетах.

```
Интервальный ряд: k=7.833037916139762 (формула Стерджесса), h=29.23514509334261=29 (шаг)
Интервальный ряд:
Интервал        Абс. частоты    Отн. частоты
[321, 350)      2               0.01754
[350, 379)      10              0.08772
[379, 408)      10              0.08772
[408, 437)      13              0.11404
[437, 466)      36              0.31579
[466, 495)      19              0.16667
[495, 524)      13              0.11404
[524, 553)      11              0.09649
```

Рисунок 1 – Интервальный ряд, подготовленный с помощью лр1.

Размер выборки $n = 114$.

Далее вычислим среднее значение для каждого интервала и с помощью абсолютной частоты найдем условные варианты по формуле:

$$\tilde{x}_i = \frac{1}{h}(x_i - C),$$

где h – ширина интервала, C – середина интервала с наибольшей частотой.

Для нахождения условных моментов нужно вычислить величины $\tilde{x}_i^l n_i$. Результаты вычислений представлены в табл. 1.

Таблица 1

x_i	n_i	$\tilde{x}_i$	$\tilde{x}_i * n_i$	$\tilde{x}_i^2 * n_i$	$\tilde{x}_i^3 * n_i$	$\tilde{x}_i^4 * n_i$
335.5	2	-4	-8.0	32.0	-128.0	512.0
364.5	10	-3	-30.0	90.0	-270.0	810.0
393.5	10	-2	-20.0	40.0	-80.0	160.0
422.5	13	-1	-13.0	13.0	-13.0	13.0
451.5	36	0	0.0	0.0	0.0	0.0
480.5	19	1	19.0	19.0	19.0	19.0
509.5	13	2	26.0	52.0	104.0	208.0
538.5	11	3	33.0	99.0	297.0	891.0

Также вычислим величину $\widetilde{(x_i^4} + 1)\verb|^|4 * n_i$ для каждого интервала(табл.2) для проведения проверки.

Таблица 2

$\widetilde{(x_i^4} + 1)\verb|^|4 * n_i$
162,0
160,0
10,0
0,0
36,0
304,0
1053,0
2816,0

$$\sum(\tilde{x}_i \ + 1)^4 * n_i = 4541$$

$$\sum \tilde{x}_i^4 * n_i + 4 * \sum \tilde{x}_i^3 * n_i + 6 * \sum \tilde{x}_i^2 * n_i + 4 * \sum \tilde{x}_i * n_i + \sum n_i =$$

$$= 114 + 4 * 7 + 6 * 345 + 4 * (-71) + 2613 = 4541$$

Как мы видим, значения совпали, значит все вычислено верно.

Далее мы находим условные моменты, по формуле:

$$\widetilde{M_l^*} \quad = \frac{1}{N} \sum_{i=1}^{k} \tilde{x}_i^l n_i$$

Условный эмпирический момент 1 порядка: 0.06140350877192982

Условный эмпирический момент 2 порядка: 3.026315789473684

Условный эмпирический момент 3 порядка: -0.6228070175438597

Условный эмпирический момент 4 порядка: 22.92105263157895

Используя вычисленные условные моменты, вычислим центральные эмпирические моменты и начальный эмпирический момент 1го порядка(выборочное среднее), по формулам:

$$\bar{x} = \bar{M}_1^* h + C$$
$$\bar{m}_2 = \left(\bar{M}_2^* - \left(\bar{M}_1^*\right)^2\right) h^2$$
$$\bar{m}_3 = \left(\bar{M}_3^* - 3\bar{M}_2^*\bar{M}_1^* + 2\left(\bar{M}_1^*\right)^3\right) h^3$$
$$\bar{m}_4 = \left(\bar{M}_4^* - 4\bar{M}_3^*\bar{M}_1^* + 6\bar{M}_2^*\left(\bar{M}_1^*\right)^2 - 3\left(\bar{M}_1^*\right)^4\right) h^4$$

Начальный эмпирический момент 1го порядка: 453.280701754386

Центральный эмпирический момент 2го порядка: 2541.960680209295

Центральный эмпирический момент 3го порядка: -28774.708304309555

Центральный эмпирический момент 4го порядка: 16368209.866476053

Вычислим выборочное среднее и дисперсию, используя стандартные формулы:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{k} x_i n_i = 453.280701754386 = \bar{x}$$

$$D_B = \frac{1}{N}\sum_{i=1}^{k} (x_i - \bar{x})^2 n_i = 2541.9606802092953 = \overline{m_2}$$

Результаты совпали с вычисленными с помощью условных моментов, следовательно формулы с условными моментами использованы корректно.

Найдем исправленную выборочную дисперсию и несмещенную оценку стандартного отклонения:

$$S^2 = \frac{N}{N-1} D_B = 2564.4559$$

$$S = \sqrt{S^2} = 50.64046$$

Далее вычислим асимметрию:

$$A = \frac{m_3}{S^3} = -0.183768$$

Вычислим эксцесс:

$$E = \frac{m_4}{S^4} - 3 = -1.060545$$

Можно увидеть, что основной центр тяжести графика немного смещен вправо, в связи с маленьким по модулю коэффициентом асимметрии. По эксцессу можно сделать вывод, что пик распределения случайной величины не такой острый, как в случае нормального распределения. Это можно наблюдать на рис.2

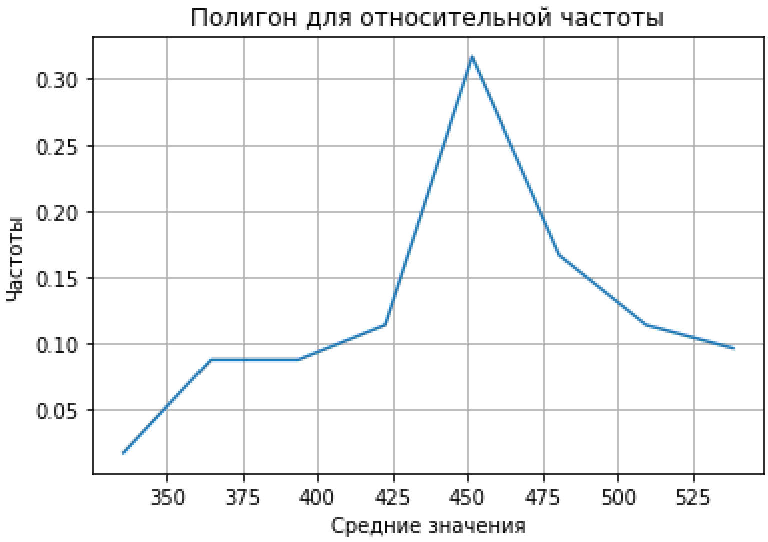

Рисунок 2 – полигон для относительной частоты из лр1

Далее найдем моду вариационного ряда по формуле:

$$M_O(X) = x_{M_O} + h\frac{(m_2 - m_1)}{(m_2 - m_1) + (m_2 - m_3)}$$

$$M_O = 453.675$$

Далее найдем медиану вариационного ряда по формуле:

$$M_e(X) = x_{M_e} + h\frac{0{,}5n - SM_{e^{-1}}}{n_{M_e}}$$

$$M_e = 454.722$$

Выводы.

В ходе выполнения данной лабораторной работы были выполнены вычисления точечных оценок мат.ожидания, дисперсии, коэффициентов асимметрии и эксцесса методом моментов. Корректность используемого метода моментов была проверена с помощью сравнения значений, полученных с помощью условных вариант и стандартных формул. Были вычислены мода и медиана распределения. Также были сделаны выводы о небольшом сдвиге в правую сторону из-за отрицательной величины асимметрии, а также более пологом пике из-за отрицательной величины эксцесса.

ПРИЛОЖЕНИЕ А

ИСХОДНЫЙ КОД

```
from lr1 import spaces, spaces_abs_freq, data, h
import pandas as pd
import numpy as np
from math import sqrt

df = pd.DataFrame()
df['Средние значения'] = list(
    map(lambda space: (space[0] + space[1]) / 2, spaces))
df['Частоты'] = spaces_abs_freq
C = df.iloc[4, 0]
df['Условные варианты'] = df['Средние значения'].apply(lambda x: (x - C)
/ h)

# условные эмпирические моменты, M с чертой и звездочкой
moments = []
for i in range(1, 5):
    col = 'nu{}'.format(i)
    df[col] = df.iloc[:, 1:3].apply(lambda x: x[0] * x[1] ** i, axis=1)
    moments.append(df[col].sum() / len(data))

df['Проверка'] = df.iloc[:, 1:3].apply(lambda x: x[0] * ((x[1]+1)**4),
axis=1)
print(df)

for i, m in enumerate(moments):
    print("Условный эмпирический момент {} порядка: {}".format(i + 1, m))

sum1 = df['nu4'].sum() + df['nu3'].sum()*4 + df['nu2'].sum()*6 +
df['nu1'].sum()*4 + df['Частоты'].sum()
sum2 = df['Проверка'].sum()
print("Проверка:", sum1, sum2, "если сошлось - все ок")

print(" --- Вычисляем через условные моменты:")
start_moment_1_usl = moments[0]*h + C
print('Начальный эмпирический момент 1го порядка: ', start_moment_1_usl)

central_moment_2_usl = (moments[1] - moments[0]**2)*(h**2)
print('Центральный эмпирический момент 2го порядка: ', central_mo-
ment_2_usl)

central_moment_3_usl = (moments[2] - 3*moments[1]
```

```
                        * moments[0] + 2*(moments[0]**3))*(h**3)
print('Центральный  эмпирический  момент  3го  порядка:  ',  central_mo-
ment_3_usl)

central_moment_4_usl = (moments[3] - 4*moments[2]*moments[0] +
                        6*moments[1]*(moments[0]**2)       -       3*(mo-
ments[0]**4))*(h**4)
print('Центральный  эмпирический  момент  4го  порядка:  ',  central_mo-
ment_4_usl)

print(' --- Вычисляем по стандартной формуле:')
start_moment_1_emp = df.iloc[:, :2].apply(
    lambda x: x[0]*x[1], axis=1).sum() / len(data)
print('Начальный эмпирический момент 1го порядка: ', start_moment_1_emp)

central_moment_2_emp = df.iloc[:, :2].apply(lambda x: (
    (x[0] - start_moment_1_emp)**2)*x[1], axis=1).sum() / len(data)
print('Центральный  эмпирический  момент  2го  порядка:  ',  central_mo-
ment_2_emp)

print(" --- ")
s = np.sqrt((len(df)/(len(df)-1)) * central_moment_2_emp)
asim = central_moment_3_usl / (s**3)
print('Асимметрия: ', asim)

ecs = central_moment_4_usl / (s**4) - 3
print('Эксцесс: ', ecs)

space_moda_index = np.argmax(spaces_abs_freq)
m2, m1, m3 = spaces_abs_freq[space_moda_index], 0, 0
if space_moda_index > 0:
    m1 = spaces_abs_freq[space_moda_index - 1]
if space_moda_index < len(spaces_abs_freq):
    m3 = spaces_abs_freq[space_moda_index + 1]
moda = spaces[space_moda_index][0] + h * ((m2 - m1) / (2 * m2 - m1 - m3))
print("Мода:", moda)

median_index = 4
median_lower_freqs = np.sum(spaces_abs_freq[:median_index])
median = \
    spaces[median_index][0] \
    + h * ((0.5 * len(data) - median_lower_freqs) /
           spaces_abs_freq[median_index])
print("Медиана:", median)
```

```
print(" --- ")
x_middle = df.iloc[:, 0:2].apply(lambda x: x[0] * x[1], axis=1).sum() /
len(data)
print("Мат. ожидание:", x_middle)

disp = df.iloc[:, 0:2].apply(lambda x: (x[0] - x_middle)**2 * x[1],
axis=1).sum() / len(data)
print("Дисперсия:", disp)

sko = sqrt(disp)
print("СКО:", sko)

N = len(data)
S2 = N/(N-1)*disp
print("Исправленная выборочная дисперсия: S^2={}, СКО S=sqrt(S^2)={}".for-
mat(S2, sqrt(S2)))
```