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. 7381	____________________	Алясова А.Н.
Студент гр. 7381	____________________	Кортев Ю.В.
Преподаватель	____________________	Середа А.-В.И.

Санкт-Петербург

2021

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:

$$A_s = \frac{m_3}{S^3},$$

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:

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

Для нормального закона $\frac{m_4}{S^4} = 3$. Отсюда следует, что для нормального закона $E = 0$. Смысл термина «эксцесс» состоит в том, что он 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.

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

где x_{M_O} – начало модального интервала, h – длина частичного интервала (шаг), m_1 – частота предмодального интервала, m_2 – частота модального интервала, m_3 – частота послемодального интервала.

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

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

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

где x_{M_e} – начало медианного интервала, h – длина частичного интервала (шаг), n – объем совокупности, $SM_{e^{-1}}$ – накопленная частота интервала, предшествующая медианному, n_{M_e} – частота медианного интервала.

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 ходе выполнения лабораторной работы №1 был получен интервальный ряд, представленный в табл. 1.

Таблица 1

Интервал	**Абсолютная частота**	**Относительная частота**
[320, 354)	9	0.07692
[354,388)	4	0.03419
[388,422)	27	0.23077
[422,456)	25	0.21368
[456,490)	24	0.20513
[490,524)	17	0.14530
[524,558)	7	0.05983
[558,592)	3	0.02564
[592,593]	1	0.00855

Количество интервалов определено по формуле Стерджесса:

$$k = 1 + 3.322 * \log(n) = 8,$$

где n – объем выборки.

Ширина интервала:

$$h \approx 34.$$

Размер выборки:

$$n = 117.$$

Найдем условные моменты по формуле:

$$\widetilde{M}_l = \frac{1}{N}\sum_{i=1}^{k} \tilde{x}_i^l n_i,$$

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

где h – длина интервала, $C = x_5$ – ложный ноль.

Результаты вычислений представлены в табл. 2.

Таблица 2

x_i	n_i	$\tilde{x}_i$	$\tilde{x}_i * n_i$	$\tilde{x}_i^2 * n_i$	$\tilde{x}_i^3 * n_i$	$\tilde{x}_i^4 * n_i$	$(\tilde{x}_i^4 + 1)^4 * n_i$
337	9	-4	-36	144	-576	2304	729
371	4	-3	-12	36	-108	324	64
405	27	-2	-54	108	-216	432	27
439	25	-1	-25	25	-25	25	0
473	24	0	0.0	0.0	0.0	0.0	24,00
507	17	1	17	17	17	17	272
541	7	2	14	28	56	112	567
575	3	3	9	27	81	243	768
609	1	4	4	16	64	256	625
$\sum =$ 4257	$\sum =$ 117	$\sum =$ 0	$\sum =$ -83	$\sum =$ 401	$\sum =$ -707	$\sum =$ 3713	$\sum =$ 3076
Условные моменты:			-0.7094	3.4274	-6.0427	31.7350	

Проверим правильность вычислений:

$$\sum \tilde{x}_i^4 * n_i + 4 * \sum \tilde{x}_i^3 * n_i + 6 * \sum \tilde{x}_i^2 * n_i + 4 * \sum \tilde{x}_i * n_i + \sum n_i = 3076$$

Вычислим статистические оценки математического ожидания:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{k} x_i n_i = 448.8803$$

Вычислим статистические оценки дисперсии:

$$D_B = \frac{1}{N} \sum_{i=1}^{k} (x_i - \bar{x})^2 \, n_i = 3380.2592$$

Отсюда следует, что среднеквадратическое отклонение:

$$\sigma = \sqrt{D_{\text{В}}} = \sqrt{3380.2592} = 58,1400$$

Найдем исправленную выборочную дисперсию:

$$S^2 = \frac{N}{N-1} D_B = \frac{117}{116} * 3380.2592 = 3409,3994$$

$$S = \sqrt{S^2} = \sqrt{3409,3994} = 58,3901$$

Для вычисления ассиметрии и эксцесса найдем центральные эмпирические моменты третьего и четвертого порядка:

$$m_3 = \left(\widetilde{M}_3 - 3\widetilde{M}_2\widetilde{M}_1 + 2\widetilde{M}_1^3\right) * h^3 = 21120.2315$$

$$m_4 = \left(\widetilde{M}_4 - 4\widetilde{M}_3\widetilde{M}_1 + 6\widetilde{M}_2 * \widetilde{M}_1^2 - 3\widetilde{M}_1^4\right) * h^4 = 32308933.7863$$

Вычислим ассиметрию:

$$As = \frac{m_3}{S^3} = \frac{21120.2315}{58{,}3901^3} = 0.1061$$

Вычислим эксцесс:

$$\mathrm{Ex} = \frac{m_4}{S^4} - 3 = -0.2205$$

Далее найдем моду вариационного ряда по формуле:

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

$$M_O = 388 + 34 * (27 - 4)/((27 - 4) + (27 - 25)) = 419.28$$

Далее найдем медиану вариационного ряда по формуле:

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

$$M_e = 456 + 34 * (0.5 * 117 - 65)/24 = 446.7917$$

Выводы.

В ходе выполнения данной лабораторной работы были получены практические навыки нахождения точечных статистических оценок параметров распределения.

При вычислении условных моментов была сделана проверка, которая показала, что данные моменты были посчитаны верно.

Так как полученное значение эксцесса $\mathrm{Ex} = -0.2205 < 0$, то можно сделать вывод, плотность закона распределения случайной величины уменьшается медленно вблизи её моды, о чем свидетельствует рис. 1, полученный в лабораторной работе №1.

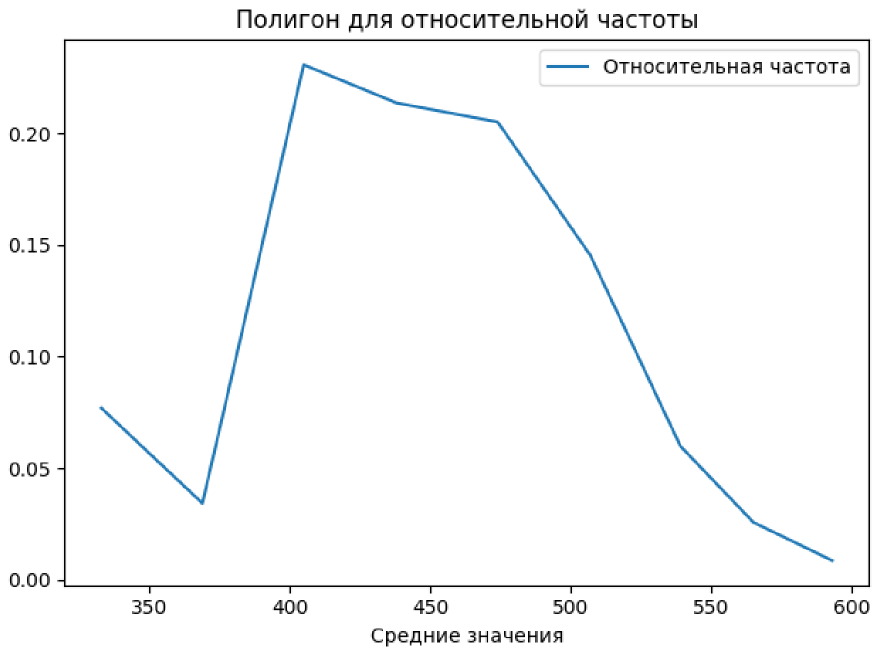

Рисунок 1

Из полученного значения коэффициента симметрии $As = 0.1061$ можно сделать вывод, что мода немного смещена влево относительно середины распределения, так как $As > 0$, но при этом находится достаточно близко к центру, так как значение As близко к 0, о чем свидетельствует рис. 1.

ПРИЛОЖЕНИЕ А
ИСХОДНЫЙ КОД

```
from lab1 import means, nums, h, df, low_range, up_range
import pandas as pd
import numpy as np

inter_df = pd.DataFrame({'Средние значения': means, 'Частота': nums}, dtype=np.int64)
C = inter_df.iloc[4, 0]
inter_df['Условные варианты'] = inter_df['Средние значения'].apply(lambda x: (x - C) / h)

moments = []
for num_of_moment in range(1, 5):
    col = 'Условный момент {}'.format(num_of_moment)
    inter_df[col] = inter_df.iloc[:, 1:3].apply(lambda x: x[0]*(x[1]**num_of_moment), axis=1)
    moments.append(inter_df[col].sum() / len(df))
    print(col, ': ', moments[-1])

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

inter_df.to_csv('Таблица.csv', index=0)

start_moment_1_usl = moments[0]*h + C
print('Начальный условный момент 1го порядка: ', start_moment_1_usl)
central_moment_2_usl = (moments[1] - moments[0]**2)*(h**2)
print('Центральный условный момент 2го порядка: ', central_moment_2_usl)
central_moment_3_usl = (moments[2] - 3*moments[1]*moments[0] + 2*(moments[0]**3))*(h**3)
print('Центральный условный момент 3го порядка: ', central_moment_3_usl)
central_moment_4_usl = (moments[3] - 4*moments[2]*moments[0] + 6*moments[1]*(moments[0]**2) - 3*(moments[0]**4))*(h**4)
print('Центральный условный момент 3го порядка: ', central_moment_4_usl)

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

```
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('Эксцесса: ', ecs)

max_low_val = 0
max_num = 0
max_low_val_prev = 0
max_low_val_next = 0

sum_median_prev_nums = 0
for i, (n, l, u) in enumerate(list(zip(nums, low_range, up_range))):
    if n > max_num:
        max_num = n
        max_low_val = l

        try:
            max_low_val_prev = nums[i-1]
            max_low_val_next = nums[i+1]
        except Exception:
            max_low_val_prev = 0
            max_low_val_next = 0

    if i < int(len(nums)/2):
        sum_median_prev_nums += n

moda = max_low_val + h * ((max_num - max_low_val_prev)/(2*max_num - max_low_val_prev - max_low_val_next))
median_int = int(len(nums)/2)
median_num = nums[median_int]
x_0_median = low_range[median_int]

median = x_0_median + h*((0.5 * sum(nums) - sum_median_prev_nums) / median_num)

print('Мода: ', moda)
print('Медиана: ', median)
```