**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.**

|  |  |  |
| --- | --- | --- |
| Студент гр. 8383 |  | Киреев К.А. |
| Студент гр. 8383 |  | Муковский Д.В. |
| 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:

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

Среднеквадратическим отклонением случайной величины Х называется квадратный корень из ее дисперсии:

Выборочная дисперсия определяется по формуле:

Исправленная выборочная дисперсияопределяется по формуле:

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

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

где – центральный эмпирический момент третьего порядка,  *–* исправленнаявыборочная дисперсия.

Эксцессом, или коэффициентом эксцесса, называется численная характеристика случайной величины, которая определяется выражением:

Мода дискретной случайной величины – это наиболее вероятное значение этой случайной величины.

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

**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**

Для интервального ряда, полученного в лабораторной работе №1 были найдены середины интервалов и накопленные частоты. Интервальный ряд представлен в таблице 1.

*Таблица 1*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Границы***  ***интервалов*** | ***Середины***  ***интервалов*** | ***Абсолютная***  ***частота*** | ***Относительная***  ***частота*** | ***Накопленная частота*** |
| [320, 357) | 338.5 | 5 | 0.048 | 0.048 |
| [357, 394) | 375.5 | 8 | 0.077 | 0.125 |
| [394, 431) | 412.5 | 23 | 0.221 | 0.346 |
| [431, 468) | 449.5 | 25 | 0.240 | 0.586 |
| [468, 505) | 486.5 | 24 | 0.231 | 0.817 |
| [505, 542) | 523.5 | 15 | 0.144 | 0.961 |
| [542, 576) | 559 | 4 | 0.039 | 1 |

Объем выборки

Количество интервалов

Ширина интервала

Условные варианты можно найти как , где – условный ноль.

Условные моменты k-го порядка:

Результаты вычислений представлены в табл. 2.

*Таблица 2*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| 338.5 | 0.048 | -3 | -0.144 | 0.432 | -1.296 | 3.888 | 0.768 |
| 375.5 | 0.077 | -2 | -0.154 | 0.308 | -0.616 | 1.232 | 0.077 |
| 412.5 | 0.221 | -1 | -0.221 | 0.221 | -0.221 | 0.221 | 0.0 |
| 449.5 | 0.240 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.24 |
| 486.5 | 0.231 | 1 | 0.231 | 0.231 | 0.231 | 0.231 | 3.696 |
| 523.5 | 0.144 | 2 | 0.288 | 0.576 | 1.152 | 2.304 | 11.664 |
| 559 | 0.039 | 3 | 0.117 | 0.351 | 1.053 | 3.159 | 9.984 |
|  |  |  | 0.117 | 2.119 | 0.303 | 11.035 | 26.429 |

Сумма элементов последнего столбца является контрольной суммой, и так как в данном случае во втором столбце записаны относительные частоты, должно быть выполнено равенство:

Эмпирические начальные и центральные моменты вычислены ниже:

Найдем выборочное среднее и дисперсию с помощью стандартных формул.

Статистическая оценка математического ожидания:

Статистическая оценка дисперсии:

Данная статистическая оценка является смещенной оценкой, поэтому вычислим исправленную оценку дисперсии:

Статистические оценки СКО:

Статистические оценки математического ожидания и дисперсии, вычисленные по стандартным формулам и с помощью условных вариант совпадают с небольшой погрешностью.

Статистические оценки коэффициентов асимметрии и эксцесса можно вычислить по формулам:

Центральные эмпирические моменты третьего и четвертого порядков были найдены выше.

Статистическая оценка коэффициента асимметрии:

Статистическая оценка коэффициента эксцесса:

Коэффициент асимметрии отрицателен, следовательно, в данном случае это левосторонняя асимметрия, которая характеризуется удлиненным левым хвостом, а также неравенством , но полученное значение незначительно и скос распределения небольшой.

Коэффициент эксцесса также отрицателен, следовательно, эмпирическое распределение является более низким и пологим относительно нормального распределения.

Вычислим моду и медиану заданного распределения для интервального ряда.

Мода заданного распределения:

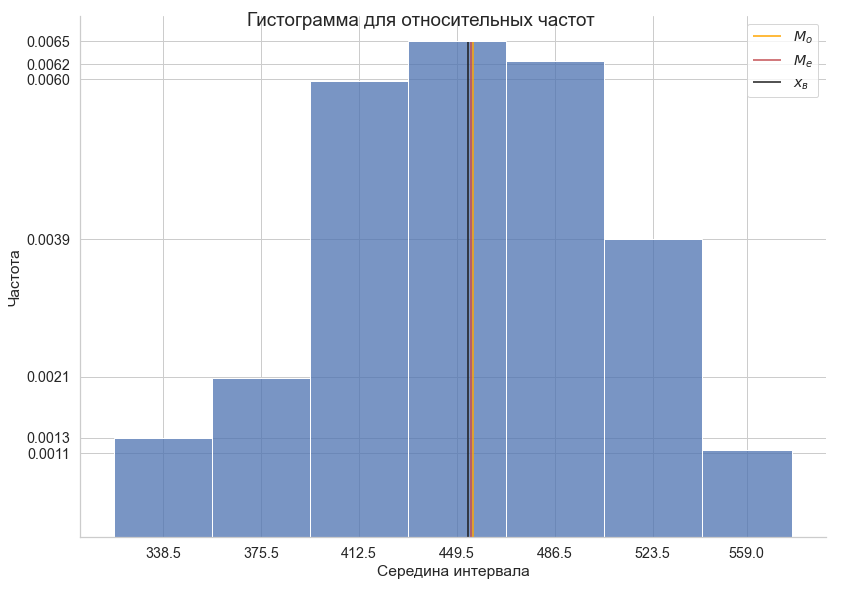
Медиана заданного распределения:

Рисунок 1 – Гистограмма относительных частот с отмеченными модой,

медианой и выборочным средним

Видно, что мода смещена относительно центра модального интервала в сторону правого интервала с большей частотой. Медиана также смещена правее, так как по правую сторону находится большее количество вариант.

В целом можно заметить, что медиана, мода и выборочное среднее примерно равны, поэтому можно предположить, что анализируемая переменная имеет примерно нормальное распределение.

**Выводы**

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

Для интервального ряда из лабораторной работы №1 были найдены середины интервалов и накопленные частоты, далее для полученных вариант были вычислены условные варианты. Были вычислены условные эмпирические моменты через условные варианты, и с их помощью вычислены начальные и центральные эмпирические моменты. Корректность вычислений была проверена контрольной суммой, которая дала понять, что вычисления были верны.

Были посчитаны выборочное среднее и дисперсия с помощью стандартных формул и с помощью условных вариант. Статистические оценки, вычисленные по стандартным формулам и с помощью условных вариант совпали.

Была найдена статистическая оценка коэффициентов асимметрии и эксцесса. Коэффициент асимметрии получился отрицательным, то есть — это левосторонняя асимметрия, которая характеризуется удлиненным левым хвостом, а также неравенством , но полученное значение незначительно и скос распределения небольшой. Коэффициент эксцесса также отрицателен, следовательно, эмпирическое распределение является более низким и пологим относительно нормального распределения. Данные наблюдения также можно увидеть на рисунке 1.

Для интервального ряда была вычислена мода и медиана. Мода оказалась смещена относительно центра модального интервала в сторону правого интервала с большей частотой. Медиана также смещена правее, так как по правую сторону находится большее количество вариант.

**ПРИЛОЖЕНИЕ А**

**ИСХОДНЫЙ КОД**

# %%

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from IPython.core.interactiveshell import InteractiveShell

InteractiveShell.ast\_node\_interactivity = "all"

pd.set\_option('display.max\_columns', None)

pd.set\_option('display.max\_rows', None)

# %%

original = pd.read\_csv('c:/Users/gandh/dev/unv/smoed/me/lab1/data/data2.csv')

var\_row = pd.read\_csv('c:/Users/gandh/dev/unv/smoed/me/lab1/data/data4.csv')

var\_row.to\_csv('data/var\_row.csv', index=False)

n = 104

h = 37

# %%

int\_row = pd.read\_csv('c:/Users/gandh/dev/unv/smoed/me/data/interval.csv')

int\_row['cum\_sum'] = np.round(np.cumsum(int\_row['rf']), 3)

int\_row['rf'] = np.round(int\_row['rf'], 3)

int\_row.to\_csv('data/int\_row.csv', index=False)

# %%

usl\_mom = int\_row.copy()

usl\_mom = usl\_mom.iloc[:, [1,3]]

usl\_mom['u'] = np.arange(-3,4,1)

usl\_mom['nu'] = usl\_mom['rf']\*usl\_mom['u']

usl\_mom['nu2'] = usl\_mom['rf']\*pow(usl\_mom['u'], 2)

usl\_mom['nu3'] = usl\_mom['rf']\*pow(usl\_mom['u'], 3)

usl\_mom['nu4'] = usl\_mom['rf']\*pow(usl\_mom['u'], 4)

usl\_mom['nu4+'] = usl\_mom['rf']\*pow(usl\_mom['u']+1, 4)

# %%

usl\_mom\_f = usl\_mom.append(usl\_mom.sum(), ignore\_index=True)

usl\_mom\_f.to\_csv('data/usl\_mom.csv', index=False)

# %%

moms = usl\_mom\_f.iloc[7, [3,4,5,6]]

moms[3]+4\*moms[2]+6\*moms[1]+4\*moms[0]+1

# %%

int\_mean = (int\_row['avg\_inter']\*int\_row['af']).sum()/n

int\_var = (((int\_row['avg\_inter']-int\_mean)\*\*2)\*int\_row['af']).sum()/n

s = int\_var\*(n/(n-1))

std\_s = np.sqrt(s)

std\_var = np.sqrt(int\_var)

std\_s

std\_var

# %%

np.mean(original, axis=0)

np.std(original, axis=0)

np.var(original, axis=0)\*(n/(n-1))

# %%

M1 = moms[0]\*h+449.5

m2 = (moms[1] - pow(moms[0],2))\*pow(h,2)

m3 = (moms[2] - 3\*moms[1]\*moms[0] + 2\*pow(moms[0],3))\*pow(h,3)

m4 = (moms[3] - 4\*moms[2]\*moms[0] + 6\*moms[1]\*pow(moms[0],2) - 3\*pow(moms[0],4))\*pow(h,4)

# %%

As = m3/(pow(s, 3))

Ex = (m4/(pow(s, 4))) - 3

# %%

As, Ex

# %%

original.mean()

np.asarray(original.mode())

original.median()

# %%

raw\_mode = 431+h\*(2/3)

raw\_median = 431+(((0.5\*n)-36)/25)\*h

raw\_mode

raw\_median

int\_mean

# %%

original.head()

# %%

sns.set\_theme(style="whitegrid", palette='deep', context='notebook', font\_scale=1.3)

ax = sns.displot(data=original, x='nu', bins=np.array([320, 357, 394, 431, 468, 505, 542, 576]),

kind='hist', height=8.27, aspect=11.7/8.27, stat='density')

plt.vlines(raw\_mode, 0, int\_row.loc[3, 'rf']/h, colors='orange', linestyles='solid', label='$M\_o$')

plt.vlines(raw\_median, 0, int\_row.loc[3, 'rf']/h, colors='r', linestyles='solid', label='$M\_e$')

plt.vlines(int\_mean, 0, int\_row.loc[3, 'rf']/h, colors='k', linestyles='solid', label='$x\_в$')

ax.set\_axis\_labels('Середина интервала', 'Частота')

ax.set(xticks=int\_row['avg\_inter'], yticks=round((int\_row['rf']/h), 4))

ax.fig.suptitle('Гистограмма для относительных частот')

plt.legend()

plt.savefig('pics/1.png')

# %%