**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 |  | Бабенко Н.С. |
| Студент гр. 8383 |  | Сахаров В.М. |
| 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:

**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**

***Переменная***

Интервальный ряд из первой лабораторной работы для переменной и с посчитанными накопленными частотами представлен в таблице 1.

Таблица 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Границы**  **интервалов** | **Середины**  **интервалов** | **Абсолютная**  **частота** | **Относительная**  **частота** | **Накопленная частота** |
| [321, 365) | 343 | 4 | 0.04 | 0.04 |
| [365, 409) | 387 | 9 | 0.09 | 0.13 |
| [409, 453) | 431 | 27 | 0.27 | 0.4 |
| [453, 497) | 475 | 35 | 0.35 | 0.75 |
| [497, 541) | 519 | 17 | 0.17 | 0.92 |
| [541, 585) | 563 | 6 | 0.06 | 0.98 |
| [585, 623) | 604 | 2 | 0.02 | 1 |

Объем выборки

Условные варианты можно найти с помощью формулы:

Условные моменты k-го порядка можно найти по формуле:

Результаты вычислений представлены в табл. 2.

Таблица 2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| 343 | 0.04 | -3 | -0.12 | 0.36 | -1.08 | 3.24 | 0.64 |
| 387 | 0.09 | -2 | -0.18 | 0.36 | -0.72 | 1.44 | 0.09 |
| 431 | 0.27 | -1 | -0.27 | 0.27 | -0.27 | 0.27 | 0.0 |
| 475 | 0.35 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.35 |
| 519 | 0.17 | 1 | 0.17 | 0.17 | 0.17 | 0.17 | 2.72 |
| 563 | 0.06 | 2 | 0.12 | 0.24 | 0.48 | 0.96 | 4.86 |
| 604 | 0.02 | 3 | 0.06 | 0.18 | 0.54 | 1.62 | 5.12 |
|  |  |  | -0.22 | 1.58 | -0.88 | 7.7 | 13.78 |

Проверить вычисления можно с помощью последнего столбца:

Число совпадает с суммой элементов последнего столбца, следовательно вычисления правильные.

Был посчитан первый начальный эмпирический момент с помощью условных вариант, который обозначает выборочное среднее:

Также был посчитан второй центральный эмпирический момент с помощью условных вариант, который обозначает выборочную дисперсию:

Далее были найдены выборочное среднее и дисперсия с помощью стандартных формул.

Исправленная оценка дисперсии:

Были найдены статистические оценки СКО:

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

Были найдены статистические оценки коэффициентов асимметрии и эксцесса:

Статистическая оценка коэффициента асимметрии:

Статистическая оценка коэффициента эксцесса:

Коэффициент асимметрии положительный, следовательно, это правосторонняя асимметрия, и , но полученный коэффициент незначительный и скос распределения небольшой. Коэффициент эксцесса же отрицателен, следовательно, эмпирическое распределение является более низким и пологим относительно нормального распределения.

Вычислим моду и медиану заданного распределения для интервального ряда. Мода заданного распределения:

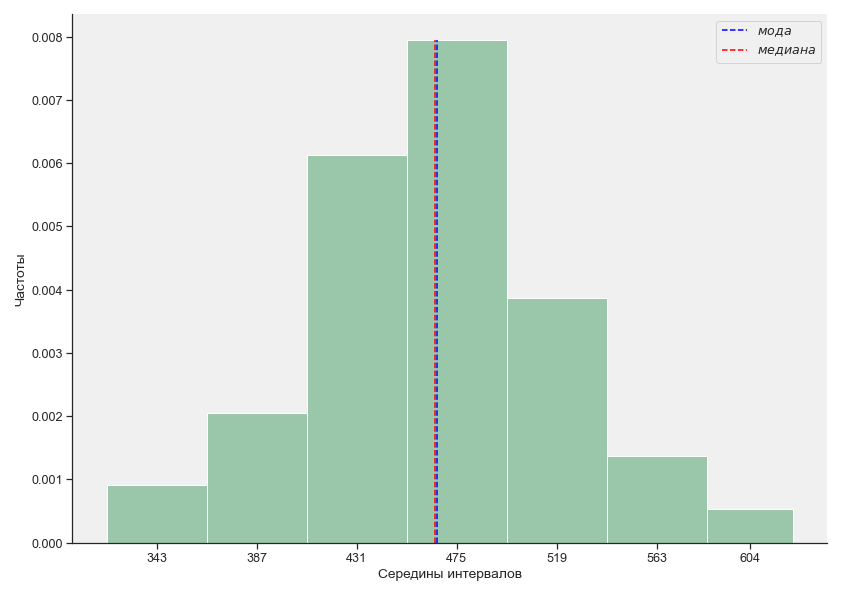
Медиана заданного распределения:

Рисунок 1 – Гистограмма относительных частот

На рисунке можно увидеть, что мода смещена относительно центра модального интервала в сторону левого интервала с большей частотой. Медиана также смещена левее, так как по левую сторону находится большее количество вариант.

***Переменная***

Интервальный ряд из первой лабораторной работы для переменной и с посчитанными накопленными частотами представлен в таблице 3.

Таблица 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Границы**  **интервалов** | **Середины**  **интервалов** | **Абсолютная**  **частота** | **Относительная**  **частота** | **Накопленная частота** |
| [84.9, 100.9) | 92 | 6 | 0.06 | 0.06 |
| [100.9, 116.9) | 108 | 14 | 0.14 | 0.2 |
| [116.9, 132.9) | 124 | 32 | 0.32 | 0.52 |
| [132.9, 148.9) | 140 | 33 | 0.33 | 0.85 |
| [148.9, 164.9) | 156 | 9 | 0.09 | 0.94 |
| [164.9, 180.9) | 172 | 4 | 0.04 | 0.98 |
| [180.9, 195.7) | 188 | 2 | 0.02 | 1 |

Результаты вычислений условных моментов представлены в табл. 4.

Таблица 4

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| 92 | 0.06 | -3 | -0.18 | 0.54 | -1.62 | 4.86 | 0.96 |
| 108 | 0.14 | -2 | -0.28 | 0.56 | -1.12 | 2.24 | 0.14 |
| 124 | 0.32 | -1 | -0.32 | 0.32 | -0.32 | 0.32 | 0.0 |
| 140 | 0.33 | 0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.33 |
| 156 | 0.09 | 1 | 0.09 | 0.09 | 0.09 | 0.09 | 1.44 |
| 172 | 0.04 | 2 | 0.08 | 0.16 | 0.32 | 0.64 | 3.24 |
| 188 | 0.02 | 3 | 0.06 | 0.18 | 0.54 | 1.62 | 5.12 |
|  |  |  | -0.55 | 1.85 | -2.11 | 9.77 | 11.23 |

Проверим вычисления с помощью последнего столбца:

Число совпадает с суммой элементов последнего столбца, следовательно вычисления правильные.

Был посчитан первый начальный эмпирический момент с помощью условных вариант, который обозначает выборочное среднее:

Также был посчитан второй центральный эмпирический момент с помощью условных вариант, который обозначает выборочную дисперсию:

Далее были найдены выборочное среднее и дисперсия с помощью стандартных формул.

Исправленная оценка дисперсии:

Были найдены статистические оценки СКО:

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

Были найдены статистические оценки коэффициентов асимметрии и эксцесса:

Статистическая оценка коэффициента асимметрии:

Статистическая оценка коэффициента эксцесса:

Коэффициент асимметрии положительный, следовательно, это правосторонняя асимметрия, и , но полученный коэффициент незначительный и скос распределения небольшой. Коэффициент эксцесса же 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:

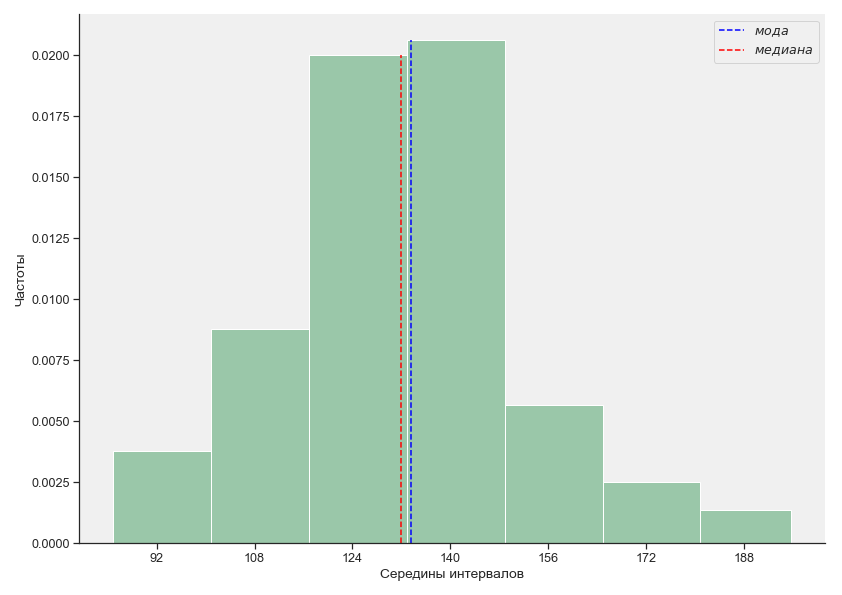


Рисунок 2 – Гистограмма относительных частот для переменной E

На рисунке можно увидеть, что мода смещена относительно центра модального интервала в сторону левого интервала с большей частотой. Медиана же смещена правее, так как по правую сторону находится большее количество вариант.

**Выводы**

В ходе выполнения данной лабораторной работы были получены практические навыки нахождения точечных статистических оценок параметров распределения. Для интервального ряда для обеих переменных из лабораторной работы №1 были вычислены условные эмпирические моменты через условные варианты. Была проведена корректность вычислений через контрольную сумму, вычисления оказались верны для обеих переменных. 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 для обеих переменных.

Были найдены коэффициенты асимметрии и эксцесса. Для обеих переменных коэффициент асимметрии получился положительным (правосторонняя асимметрия), то есть присутствует удлиненный правый хвост и , но полученное значение незначительно и скос распределения небольшой. Коэффициент эксцесса для обеих переменных получился уже отрицательным, следовательно, эмпирическое распределение является более низким и пологим относительно нормального распределения.

Для интервальных рядов была вычислена мода и медиана. Мода для обеих переменных оказалась смещена относительно центра модального интервала в сторону левого интервала с большей частотой. Для переменной : медиана также смещена левее, так как по левую сторону находится большее количество вариант, для переменной : медиана смещена правее, так как уже 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

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

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

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

n = 100

h = 44

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

int\_row['cum\_sum'] = np.round(np.cumsum(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

usl\_mom\_f = usl\_mom.append(np.round(usl\_mom.sum(), 3), ignore\_index=True)

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

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

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

'True' if checker == usl\_mom\_f.loc[7, ['nu4+']][0] else 'False'

checker

M1 = moms[0]\*h+475

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)

(M1, m2, m3, m4)

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)

int\_mean

int\_var

s

std\_s

std\_var

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

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

As, Ex

raw\_mode = 453+h\*(8/26)

raw\_median = 453+(((0.5\*n)-40)/35)\*h

raw\_mode

raw\_median

int\_mean

sns.set\_theme(palette='crest', font\_scale=1.15)

sns.set\_style('ticks', {"axes.facecolor": ".94"})

ax = sns.displot(data=original, x='nu', bins=np.array([321, 365, 409, 453, 497, 541, 585, 623]),

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

plt.vlines(raw\_mode, 0, int\_row.loc[3, 'rf']/h, colors='b', linestyles='--', label='$мода$')

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

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

ax.set\_axis\_labels('Середины интервалов', 'Частоты')

ax.set(xticks=int\_row['avg\_inter'])

plt.legend()

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

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

var\_row = pd.read\_csv('c:/Users/gandh/dev/unv/smoed/NB/lab1/data2/data4.csv')

var\_row.to\_csv('data/var\_row2.csv', index=False)

n = 100

h = 16

int\_row = pd.read\_csv('c:/Users/gandh/dev/unv/smoed/NB/data/interval2.csv')

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

int\_row.to\_csv('data/int\_row2.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

usl\_mom\_f = usl\_mom.append(np.round(usl\_mom.sum(), 3), ignore\_index=True)

usl\_mom\_f.to\_csv('data/usl\_mom2.csv', index=False)

usl\_mom\_f

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

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

'True' if checker == usl\_mom\_f.loc[7, ['nu4+']][0] else 'False'

checker

M1 = moms[0]\*h+140

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)

M1, m2, m3, m4

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)

int\_mean

int\_var

s

std\_s

std\_var

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

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

As, Ex

original.mean()

raw\_mode = 132.9+h\*(1/25)

raw\_median = 116.9+(((0.5\*n)-20)/32)\*h

raw\_mode

raw\_median

int\_mean

sns.set\_theme(palette='crest', font\_scale=1.15)

sns.set\_style('ticks', {"axes.facecolor": ".94"})

ax = sns.displot(data=original, x='E', bins=np.array([84.9, 100.9, 116.9, 132.9, 148.9, 164.9, 180.9, 195.7]),

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

plt.vlines(raw\_mode, 0, int\_row.loc[3, 'rf']/h, colors='b', linestyles='--', label='$мода$')

plt.vlines(raw\_median, 0, int\_row.loc[2, 'rf']/h, colors='r', linestyles='--', label='$медиана$')

# plt.vlines(int\_mean, 0, int\_row.loc[2, 'rf']/h, colors='k', linestyles='--', label='$x\_в$')

ax.set\_axis\_labels('Середины интервалов', 'Частоты')

ax.set(xticks=int\_row['avg\_inter'])

plt.legend()

plt.savefig('pics/2.png')