

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.8382

Нечепуренко Н.А.

Студент гр.8382

Терехов А.Е.

Преподаватель

Середа А.-В.И.

Санкт-Петербург

2022

Цели работы.

Получение практических навыков нахождения точечных статистических оценок параметров распределения.

Постановка задачи.

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

Порядок выполнения работы.

1. Для интервального ряда, полученного в лабораторной работе №1 найти середины интервалов, а также накопленные частоты. Результаты занести в таблицу.
2. Для полученных вариантов вычислить условные варианты. Результаты занести в таблицу.
3. Вычислить условные эмпирические моменты M_i^* через условные варианты. С помощью условных эмпирических моментов вычислить центральные эмпирические моменты μ_i^* . Полученные результаты занести в таблицу.
4. Вычислить выборочные среднее и дисперсию с помощью стандартной формулы и с помощью условных вариантов. Убедиться, что результаты совпадают.
5. Найти статистическую оценку коэффициентов асимметрии и эксцесса. Сделать выводы.

6. Дополнительное необязательное задание: для интервального ряда вычислить моду и медиану заданного распределения. Сделать выводы.

Основные теоретические положения.

Статистической оценкой Θ^* неизвестного параметра теоретического распределения Θ называется функция от наблюдаемых значений случайной величины:

$$\Theta^* = f(x_1, x_2, \dots, x_n)$$

Статистические оценки, определяемые одним числом, называются *точечными*.

Оценка называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру Θ при любом объеме выборки n .

Начальным эмпирическим моментом k -того порядка называется среднее значение k -х степеней элементов вариационного или интервального ряда:

$$\overline{M}_k = \frac{1}{N} \sum n_j x_j^k$$

Центральным эмпирическим моментом k -того порядка называется среднее значение k -х степеней разностей $x_j - \bar{x}$ для вариационного или интервального ряда.

$$\overline{m}_k = \frac{1}{N} \sum n_j (x_j - \bar{x})^k$$

Для упрощения вычислений используют условные моменты k -того порядка:

$$\overline{M}_k^* = \frac{1}{N} \sum n_j \left(\frac{x_j - C}{h} \right)^k = \frac{1}{N} \sum n_j u_j^k$$

где C – середина интервала, принятого за условный ноль, h – длина интервала.

Центральные эмпирические моменты связаны с условными следующими соотношениями:

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

Статистические оценки асимметрии и эксцесса вычисляются по формулам:

$$\begin{aligned}\bar{A}_s &= \frac{\bar{m}_3}{s^3} \\ \bar{E}_s &= \frac{\bar{m}_4}{s^3} - 3\end{aligned}$$

где s^2 – несмещенная оценка дисперсии.

Выполнение работы.

Для выполнения работы был выбран язык Python3 и среда Jupyter Notebook с сервисом Google Colab.

В лабораторной работе №1 были получены интервальные ряды для мужского и женского роста. Вычислим середины интервалов, а также накопленные частоты (см. табл. 1 и 2).

Таблица 1 – Интервальный ряд роста мужчин

Интервал	Середина интервала	Абсолютная частота	Накопленная частота
[172.15 173.81)	172.98	16	16
[173.81 175.47)	174.64	26	42
[175.47 177.13)	176.3	25	67

[177.13 178.8)	177.96	15	82
[178.8 180.46)	179.63	10	92
[180.46 182.12)	181.29	14	106
[182.12 183.78]	182.95	4	110

Таблица 2 – Интервальный ряд роста женщин

Интервал	Середина интервала	Абсолютная частота	Накопленная частота
[158.29 160.01)	159.15	8	8
[160.01 161.74)	160.88	24	32
[161.74 163.46)	162.6	26	58
[163.46 165.19)	164.32	16	74
[165.19 166.91)	166.05	19	93
[166.91 168.64)	167.77	11	104
[168.64 170.36]	169.5	6	110

Для вычисления условных вариантов в качестве условного нуля возьмем середину 4 интервала как для мужчин, так и для женщин.

Построим таблицы вычисления условных вариантов согласно рассмотренной в лекции, только будем использовать абсолютные частоты. Результаты приведены в таблицах 3 и 4.

Таблица 3 – Условные варианты роста мужчин

\bar{x}_i	u	n	nu	nu^2	nu^3	nu^4	$n(u+1)^4$
172.98	-3	16	-48	144	-432	1296	256
174.64	-2	26	-52	104	-208	416	26
176.30	-1	25	-25	25	-25	25	0

177.96	0	15	0	0	0	0	15
179.62	1	10	10	10	10	10	160
181.28	2	14	28	56	112	224	1134
182.94	3	4	12	36	108	324	1024
Σ		110	-75	375	-435	2295	2615
M_k^*			-0.68	3.41	-3.95	20.86	-

Выполним проверку по последнему столбцу: $2295 - 1740 + 2250 - 300 + 110 = 2615$ – верно.

Таблица 4 – Условные варианты роста женщин

\bar{x}_i	u	n	nu	nu^2	nu^3	nu^4	$n(u+1)^4$
159.15	-3	8	-24	72	-216	648	128
160.87	-2	24	-48	96	-192	384	24
162.60	-1	26	-26	26	-26	26	0
164.32	0	16	0	0	0	0	16
166.04	1	19	19	19	19	19	304
167.77	2	11	22	44	88	176	891
169.49	3	6	18	54	162	486	1536
Σ		110	-39	311	-165	1739	2899
M_k^*			-0.35	2.83	-1.5	15.81	

Выполним проверку по последнему столбцу: $1739 - 660 + 1866 - 156 + 110 = 2899$ – верно.

Обозначим эмпирические моменты как μ_k (в лекциях и формулах в разделе основные теоретические положения эмпирические моменты обозначают

лись как \overline{m}_k). Вычислим данные величины по формулам выше, получаем:

$$\mu_1^M = 176.83$$

$$\mu_2^M = 8.12$$

$$\mu_3^M = 10.93$$

$$\mu_4^M = 144.30$$

$$\mu_1^Ж = 163.73$$

$$\mu_2^Ж = 7.45$$

$$\mu_3^Ж = 6.50$$

$$\mu_4^Ж = 120.13$$

Вычислим среднее и дисперсию интервальных рядов с помощью стандартных формул:

$$\begin{aligned}\bar{x}_B &= \frac{1}{n} \sum n_i \bar{x}_i \\ D &= \frac{1}{n} \sum n_i * (\bar{x}_i - \bar{x}_B)^2\end{aligned}$$

Результаты совпали.

Вычислим несмещенную оценку дисперсии по формуле:

$$s^2 = \frac{n}{n-1} D$$

Получаем следующие значения:

$$s_M^2 = 8.20$$

$$s_{Ж}^2 = 7.52$$

По приведенным в разделе Основные теоретические положения форму-

лам вычислим оценку коэффициентов асимметрии и эксцесса:

$$\overline{A}_M = 0.46$$

$$\overline{E}_M = 2.14$$

$$\overline{A}_Ж = 0.31$$

$$\overline{E}_Ж = 2.12$$

Выборки роста мужчин и женщин имеют правостороннюю асимметрию, что вполне ожидаемо, так как были отобраны 110 стран с наиболее высоким населением в среднем. Так как значение коэффициента эксцесса положительно, кривые распределений островершинные.

Выводы.

В результате выполнения работы были найдены точечные статистические оценки параметров распределения для заданной выборки. С помощью метода условных эмпирических моментов были вычислены центральные эмпирические моменты 1-4 порядка. Результаты были сравнены с результатами стандартных формул для 1 и 2 порядка. Также были получены оценки коэффициентов асимметрии и эксцесса.

ПРИЛОЖЕНИЕ А. ИСХОДНЫЙ КОД ПРОГРАММЫ.

```
from math import log10, sqrt

import pandas as pd

SIZE = 110
N_INT_FLOAT = 1 + 3.322 * log10(SIZE)
N_INT = int(N_INT_FLOAT)

data = pd.read_csv('https://raw.githubusercontent.com/
    nechepurenkoN/spbetu2022_stats_methods/master/cropped_sample.
    csv')
male = data.iloc[:SIZE, 1]
female = data.iloc[:SIZE, 2]

sorted_male = male.sort_values()
sorted_female = female.sort_values()

male_int_width = (sorted_male.iloc[-1] - sorted_male.iloc[0]) /
    N_INT
female_int_width = (sorted_female.iloc[-1] - sorted_female.iloc
    [0]) / N_INT

def intervals(series, width):
    beg = []
    end = []
    avg = []
    count1 = []
    count2 = []
    for i_interval in range(N_INT):
        b = series.min() + width * i_interval
```

```

    beg.append(b)
    e = series.min() + width * (i_interval + 1)
    end.append(e)
    avg.append((b + e) / 2)
    count1.append(len(series[series.between(b, e)]))
    count2.append(len(series[series.between(b, e)]) / SIZE)
    print(*list(zip([round(num, 3) for num in beg], [round(num,
        3) for num in end])), sep="\t")
    print(*[round(num, 3) for num in avg], sep="\t")
    print(*count1, sep="\t")
    print(*[round(num, 3) for num in count2], sep="\t")
    return list(zip(beg, end, avg, count1, count2, pd.Series(
        count1).cumsum()))

male_hist = intervals(sorted_male, male_int_width)
print('MALE INTERVAL SERIES')
for i in range(N_INT):
    print(
        f'[{round(male_hist[i][0], 2)} {round(male_hist[i][1], 2)}
        }) & {round(male_hist[i][2], 2)} & {round(male_hist[i]
        ][3], 2)} & {round(male_hist[i][5], 2)} \\\hline')

female_hist = intervals(sorted_female, female_int_width)
print('FEMALE INTERVAL SERIES')
for i in range(N_INT):
    print(
        f'[{round(female_hist[i][0], 2)} {round(female_hist[i]
        ][1], 2)} & {round(female_hist[i][2], 2)} & {round(
        female_hist[i][3], 2)} & {round(female_hist[i][5], 2)}
        \\\hline')

```

```

def cond_moments(data, h):
    x = pd.Series([d[2] for d in data])
    u = pd.Series(list(range(-3, 4)))
    n = pd.Series([d[3] for d in data])
    un = u * n
    u2n = u ** 2 * n
    u3n = u ** 3 * n
    u4n = u ** 4 * n
    u14n = (u + 1) ** 4 * n
    for i in range(N_INT):
        print(f'{x[i]} & {u[i]} & {n[i]} & {un[i]} & {u2n[i]} & {
            u3n[i]} & {u4n[i]} & {u14n[i]} \\\\'hline')
    print()
    print(f'{n.sum()} & {un.sum()} & {u2n.sum()} & {u3n.sum()} &
        {u4n.sum()} & {u14n.sum()} \\\\'hline')
    print()
    M1 = un.sum() / SIZE
    M2 = u2n.sum() / SIZE
    M3 = u3n.sum() / SIZE
    M4 = u4n.sum() / SIZE
    print(f'{round(M1, 2)} & {round(M2, 2)} & {round(M3, 2)} & {
        round(M4, 2)} & \\\\'hline')
    print()
    printПРОВЕРКА(": ", u4n.sum(), 4 * u3n.sum(), 6 * u2n.sum(),
        4 * un.sum(), n.sum())
    printПРОВЕРКА(": ", u4n.sum() + 4 * u3n.sum() + 6 * u2n.sum()
        + 4 * un.sum() + n.sum())

    printЭмпирические(" моменты")
    mu1 = M1 * h + data[3][2]
    mu2 = (M2 - M1 ** 2) * h ** 2

```

```

mu3 = (M3 - 3 * M2 * M1 + 2 * M1 ** 3) * h ** 3
mu4 = (M4 - 4 * M3 * M1 + 6 * M2 * M1 ** 2 - 3 * M1 ** 4) * h
      ** 4

print(f'mu1 = {mu1}')
print(f'mu2 = {mu2}')
print(f'mu3 = {mu3}')
print(f'mu4 = {mu4}')
x_avg = M1 * h + x[N_INT // 2]
print(f'x_avg = {x_avg}')
print(f'D_v = {mu2}')
print(f'sigma = {sqrt(mu2)}')
S2 = SIZE / (SIZE - 1) * mu2
print(f'S2 = {S2}')
s = sqrt(S2)
print(f'S = {s}')
As = mu3 / s ** 3
print(f'As = {As}')
Ex = mu4 / s ** 4
print(f'Ex = {Ex}')
print(f'M0 = {data[0][1] + (n[1] - n[0]) / ((n[1] - n[0]) + (
    n[1] - n[2])) * h}')
print(f'Me = {data[0][1] + (0.5 * SIZE) / n[1]}')

print('MALE COND MOMENTS')
cond_moments(male_hist, male_int_width)
print('FEMALE COND MOMENTS')
cond_moments(female_hist, male_int_width)

```