

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

ОТЧЕТ
по лабораторной работе №3
по дисциплине «Статистические методы обработки экспериментальных
данных»

Тема: Обработка выборочных данных. Нахождение интервальных
оценок параметров распределения. Проверка статистической гипотезы
о нормальном распределении.

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 вычисления интервальных статистических оценок параметров распределения выборочных данных и проверки «справедливости» статистических гипотез.

Постановка задачи.

Для заданной надежности определить (на основании выборочных данных и результатов выполнения практической работы №2) границы доверительных интервалов для математического ожидания и среднеквадратичного отклонения случайной величины. Проверить гипотезу о нормальном распределении исследуемой случайной величины с помощью критерия Пирсона χ^2 .
Дать содержательную интерпретацию полученным результатам.

Порядок выполнения работы.

1. Вычислить точность и доверительный интервал для математического ожидания при неизвестном среднеквадратичном отклонении при заданном объёме выборки для доверительной точности $\gamma \in \{0.95, 0.99\}$. Сделать выводы.
2. Для вычисления границ доверительного интервала для среднеквадратичного отклонения определить значение q при заданных γ и n . Построить доверительные интервалы, сделать выводы.
3. Проверить гипотезу о нормальности заданного распределения с помощью критерия χ^2 (Пирсона). Для этого необходимо найти теоретические частоты и вычислить наблюдаемое значение критерия. Далее по заданному уровню значимости $\alpha = 0.05$ и числу степеней свободы найти критическую точку и сравнить с наблюдаемым значением. Сделать

ВЫВОДЫ.

Основные теоретические положения.

Интервальной оценкой математического ожидания по выборочной средней \bar{x}_B при неизвестном среднеквадратическом отклонении σ генеральной совокупности служит доверительный интервал:

$$\bar{x}_B - \frac{t_\gamma s}{\sqrt{N}} < a < \bar{x}_B + \frac{t_\gamma s}{\sqrt{N}}$$

Интервальной оценкой среднеквадратического отклонения σ по исправленной выборочной дисперсии служит доверительный интервал:

$$\frac{\sqrt{2N}}{\sqrt{2N-3} - t_\gamma} s < \sigma < \frac{\sqrt{2N}}{\sqrt{2N-3} + t_\gamma} s$$

Критерий Пирсона – применяют для проверки гипотезы о соответствии эмпирического распределения предполагаемому теоретическому распределению $F(x)$.

Рассмотрим *z-преобразование*

$$z = \frac{x - a}{\sigma}$$

Полученные величины используем в функции Лапласа

$$\Phi(z) = \frac{1}{2\pi} \int_0^z \exp\left(-\frac{t^2}{2}\right) dt$$

С помощью рассмотренной функции можем вычислить теоретические частоты p_i :

$$p_i = \Phi(z_{i+1}) - \Phi(z_i)$$

$$n'_i = p_i N$$

Затем можно вычислить наблюдаемое значение критерия и сравнить с табличным значением, тем самым, принять или отклонить гипотезу о нормальном характере распределения.

$$\chi^2_{\text{набл}} = \sum_{i=1}^K \frac{(n_i - n'_i)^2}{n'_i}$$

Выполнение работы.

В предыдущей работы были получены следующие данные. *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

Средний выборочный рост мужчин составил 176.83, женщин – 163.73. Исправленные СКО равны соответственно 2.86 и 2.74.

Зададим $\gamma = 0.95$, тогда табличное значение t_γ равно 1.9819. По формулам выше найдем интервальные оценки для выборочных средних и СКО.

$$176.29 < a_{\text{м}} < 177.37$$

$$163.21 < a_{\text{ж}} < 164.25$$

$$2.54 < \sigma_{\text{м}} < 3.33$$

$$2.43 < \sigma_{\text{ж}} < 3.19$$

Проверим гипотезу о нормальном распределении исследуемых величин с помощью критерия Пирсона χ^2 (см. табл. 3 и 4).

Таблица 3 – Теоретические частоты роста мужчин

x_i	x_{i+1}	n_i	p_i	n'_i
$-\infty$	173.81	16	0.14	16.03
173.81	175.47	26	0.17	18.89
175.47	177.13	25	0.22	24.69
177.13	178.8	15	0.21	23.26
178.8	180.46	10	0.14	15.80
180.46	182.12	14	0.07	7.73
182.12	∞	4	0.03	3.56

Таблица 4 – Теоретические частоты роста женщин

x_i	x_{i+1}	n_i	p_i	n'_i
$-\infty$	160.01	8	0.08	9.61
160.01	161.74	24	0.14	16.04
161.74	163.46	26	0.22	24.97
163.46	165.19	16	0.24	26.53
165.19	166.91	19	0.17	19.23
166.91	168.64	11	0.08	9.51
168.64	∞	6	0.03	4.07

В качестве критического значения будем использовать $\chi^2_{\text{крит}} = 9.5$ при уровне значимости $\alpha = 0.05$ и 4 степенях свободы.

Для мужского роста значение критерия равно 12.86, для женского – 9.59. Исходя из этого, гипотеза о нормальном характере распределения отклоняется для обеих величин $\chi^2_{\text{набл}} > \chi^2_{\text{крит}}$.

На самом деле, подобный результат совершенно ожидаем, ведь в построении выборке участвовали 110 «самых высоких» стран, таким образом был потерян «левый хвост».

Выводы.

В результате выполнения работы были получены границы интервальных оценок для среднего роста мужчин и женщин, а также границы интервалов выборочного СКО.

Также была проверена гипотеза о нормальном характере распределения рассматриваемых величин. Значение критерия Пирсона оказалось выше критического значения, поэтому гипотеза была отклонена для обеих величин.

ПРИЛОЖЕНИЕ А. ИСХОДНЫЙ КОД ПРОГРАММЫ.

```
from scipy.special import erf

def lab3(data, x_avg, s):
    t = 1.981967489688474
    di_avg = [x_avg - (t * s) / sqrt(SIZE), x_avg + (t * s) /
               sqrt(SIZE)]
    print(f'DI: {di_avg}')
    di_s = [sqrt(2 * SIZE) / (sqrt(2 * SIZE - 3) + t) * s, sqrt(2
        * SIZE) / (sqrt(2 * SIZE - 3) - t) * s]
    print(f'DI: {di_s}')

    z = []
    z.append(-inf)
    for i in range(N_INT - 1):
        z.append((data[i][1] - x_avg) / s)
    z.append(inf)
    print(z)

    phi_fn = lambda x: erf(x / 2 ** 0.5) / 2
    phi = []
    p = []
    for i in range(N_INT):
        phi.append(phi_fn(z[i]))
        p.append(phi_fn(z[i + 1]) - phi_fn(z[i]))
    phi.append(phi_fn(z[7]))
    print(phi)
    print(p)
    p = pd.Series(p)
    n_ = p * SIZE
    print(f'N_: {n_}')
```

```
n = pd.Series([data[i][3] for i in range(N_INT)])  
chi2 = sum((n - n_) ** 2 / n_)  
print(f'chi2: {chi2}')
```



```
lab3(male_hist, x_avg_m, s_m)  
lab3(female_hist, x_avg_f, s_f)
```