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. 8383	____________________	Киреев К.А.
Студент гр. 8383	____________________	Муковский Д.В.
Преподаватель	____________________	Середа А.-В.И.

Санкт-Петербург
2022

Цель работы

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

Основные теоретические положения

Доверительным называют интервал, который с заданной надежностью γ покрывает заданный параметр. Доверительный интервал для оценки математического ожидания при неизвестном СКО, который покрывает неизвестное значение параметра a с надежностью γ можно построить как:

$$\left(\overline{x_{\text{в}}} - t_{\gamma}\frac{s}{\sqrt{N}}, \overline{x_{\text{в}}} + t_{\gamma}\frac{s}{\sqrt{N}}\right)$$

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

$$s(1-q) \leq \sigma \leq s(1-q),$$

Критерий Пирсона, или критерий χ^2, применяют для проверки гипотезы о соответствии эмпирического распределения предполагаемому теоретическому распределению. Метод позволяет оценить статистическую значимость различий двух или нескольких относительных показателей.

Теоретические частоты вычисляются по формуле:

$$n_i^{'} = p_i * N,$$

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

где $\Phi(z_i)$ – функция Лапласа

Если $\chi^2_{\text{наб}} \leq \chi^2_{\text{крит}}$ – гипотеза принимается, иначе – гипотеза отвергается.

Постановка задачи

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

Выполнение работы

Выборочные данные лабораторной работы №2 представлены в табл. 1.

Таблица 1

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

Объем выборки $N = 104$

Количество интервалов $k = 1 + 3.31 * \lg N = 7$

Ширина интервала $h = \frac{x_{max} - x_{min}}{k} = \frac{576-320}{7} = 37$

Статистическая оценка математического ожидания:

$$\overline{x_{\text{в}}} = \frac{1}{N}\sum_{i=1}^{k} x_i n_i = 453.71$$

Исправленная выборочная дисперсия:

$$s^2 = \frac{N}{N-1} D_{\text{в}} = \frac{104}{103} * 2865.5 = 2893.32$$

Статистическая оценка СКО:

$$s = \sqrt{s^2} = \sqrt{2893.32} = 53.79$$

- Вычислим точность и доверительный интервал для математического ожидания при неизвестном СКО для доверительной точности γ

  Случайная величина t:

$$t = \frac{\overline{x_{\text{в}}} - a}{s/\sqrt{N}}$$

Эта случайная величина распределена по закону Стьюдента с $k = N - 1$ степенями свободы. Справедливо соотношение:

$$P\left(\left|\frac{\overline{x_{\text{в}}} - a}{s/\sqrt{N}}\right| < t_\gamma\right) = 2\int_0^{t_\gamma} S(t, N)dt = \gamma$$

$$P\left(\overline{x_{\text{в}}} - t_\gamma\, s/\sqrt{N} < a < \overline{x_{\text{в}}} + t_\gamma\, s/\sqrt{N}\right) = \gamma$$

Доверительный интервал для оценки математического ожидания:

$$\left(\overline{x_{\text{в}}} - t_\gamma \frac{s}{\sqrt{N}}, \overline{x_{\text{в}}} + t_\gamma \frac{s}{\sqrt{N}}\right), \text{где}$$

$\overline{x_{\text{в}}}$ – выборочное среднее

s – исправленное СКО

$t_\gamma = 1.984$ – определено из соответствующей таблицы

(по заданным значениям $\gamma = 0.95$, $N = 104$)

$$\overline{x_{\text{в}}} - t_\gamma \frac{s}{\sqrt{N}} = 453.71 - 1.984 * \frac{53.79}{\sqrt{104}} = 443.25$$

$$\overline{x_{\text{в}}} + t_\gamma \frac{s}{\sqrt{N}} = 453.71 + 1.984 * \frac{53.79}{\sqrt{104}} = 464.17$$

Можно сделать вывод, что интервал $(443.25;\ 464.17)$ с вероятностью (надежностью) $\gamma = 0.95$ содержит в себе истинное значение математического ожидания.

- Построим доверительный интервал для среднеквадратического отклонения:

$$P(s - \delta < \sigma < s + \delta) = \gamma$$

$$P(s(1 - \delta/s) < \sigma < s(1 + \delta/s)) = \gamma$$

$$q = \delta/s$$

Доверительный интервал для оценки СКО:

$$s(1-q) < \sigma < s(1+q), \text{где}$$

s – исправленное СКО

$q = 0.141$ – определено из соответствующей таблицы

(по заданным значениям $\gamma = 0.95$, $N = 104$)

$$s(1-q) = 53.79 * 0.859 = 46.206$$

$$s(1+q) = 53.79 * 1.141 = 61.374$$

Можно сделать вывод, что интервал $(46.206; 61.374)$ с вероятностью (надежностью) $\gamma = 0.95$ содержит в себе истинное значение среднеквадратического отклонения.

- Проверим гипотезу о нормальности заданного распределения с помощью критерия Пирсона χ^2

Гипотеза H_0 – выборочные данные представляют значения случайной величины, распределённой по нормальному закону распределения. Согласно критерию Пирсона, вычисляется наблюдаемое значение случайной величины χ^2:

$$\chi^2_{\text{набл}} = \sum_1^K \frac{(n_i - n_i')^2}{n_i{}'}$$

Распределение хи-квадрат зависит от числа степеней свободы k, которое вычисляется как $k = K - 3$. По числу степеней свободы и уровню значимости вычисляется значение $\chi^2_{\text{крит}} = \chi^2(\alpha, k)$. Область принятия гипотезы H_0 определяется условием:

$$\chi^2_{\text{набл}} \le \chi^2_{\text{крит}}$$

Найдем теоретические частоты. Вычисления представлены в табл. 2.

Таблица 2

x_i	x_{i+1}	n_i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	p_i	$n_i{}'$
320.0	357.0	5	$-\infty$	-1.8	-0.5	-0.4641	0.0359	3.7336
357.0	394.0	8	-1.8	-1.11	-0.4641	-0.3665	0.0976	10.1504
394.0	431.0	23	-1.11	-0.42	-0.3665	-0.1628	0.2037	21.1848
431.0	468.0	25	-0.42	0.27	-0.1628	0.1064	0.2692	27.9968
468.0	505.0	24	0.27	0.95	0.1064	0.3289	0.2225	23.14
505.0	542.0	15	0.95	1.64	0.3289	0.4495	0.1206	12.5424
542.0	576.0	4	1.64	$+\infty$	0.4495	0.5	0.0505	5.252

Вычислим наблюдаемое значение критерия $\chi^2_{\text{набл}}$. Результаты представлены в табл. 3.

Таблица 3

n_i	$n_i{}'$	$n_i - n_i'$	$(n_i - n_i')^2$	$(n_i - n_i')^2/n_i'$
5	3.7336	1.2664	1.6038	0.4296
8	10.1504	-2.1504	4.6242	0.4556
23	21.1848	1.8152	3.295	0.1555
25	27.9968	-2.9968	8.9808	0.3208
24	23.14	0.86	0.7396	0.032
15	12.5424	2.4576	6.0398	0.4816
4	5.252	-1.252	1.5675	0.2985

$$\chi^2_{\text{набл}} = \sum_{1}^{K} \frac{(n_i - n_i')^2}{n_i{}'} = 2.1736$$

Найдем $\chi^2_{\text{крит}}$ по заданному уровню значимости $\alpha = 0.05$ и числу степеней свободы $k = K - 3 = 4$:

$$\chi^2_{\text{крит}} = 9.5$$

Сравним $\chi^2_{\text{крит}}$ с наблюдаемым значением:

$$\chi^2_{\text{набл}} = 2.1736$$

$$\chi^2_{\text{крит}} = 9.5$$

$$\chi^2_{\text{набл}} \le \chi^2_{\text{крит}}$$

Из полученных результатов можно сделать вывод, что выдвинутая нулевая гипотеза принимается, то есть выборочные данные позволяют предположить, что случайная величина распределена по нормальному закону распределения.

Выводы

В ходе выполнения лабораторной работы был вычислен доверительный интервал для математического ожидания при неизвестном СКО с доверительной точностью $\gamma = 0.95$. Исходя из полученных результатов можно сделать вывод, что интервал $(443.25;\ 464.17)$ с вероятностью (надежностью) $\gamma = 0.95$ содержит в себе истинное значение математического ожидания.

Были вычислены границы доверительного интервала для среднеквадратического отклонения. Определено, что интервал $(46.206;\ 61.374)$ с вероятностью (надежностью) $\gamma = 0.95$ содержит в себе истинное значение среднеквадратического отклонения.

Была выполнена проверка гипотезы о нормальности заданного распределения с помощью критерия χ^2 (Пирсона). Было выяснено, что $\chi^2_{\text{набл}} \le \chi^2_{\text{крит}}$, следовательно, выдвинутая нулевая гипотеза принимается, то есть выборочные данные позволяют предположить, что случайная величина распределена по нормальному закону распределения.

ПРИЛОЖЕНИЕ А

ИСХОДНЫЙ КОД

```
#!/usr/bin/env python
# coding: utf-8

# In[127]:


import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import scipy
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
sns.set_theme(style="whitegrid", palette='deep', context='notebook', font_scale=1.3)

# ## Переменная $\nu$

# In[68]:


int_row = pd.read_csv('c:/Users/gandh/dev/unv/smoed/me/data/interval.csv')
N = int_row['af'].sum()
h = 37
N

# In[52]:


xv = (np.dot(int_row['avg_inter'], int_row['af'])/N).round(2)
dv = (np.dot((int_row['avg_inter']-xv)**2, int_row['af'])/N)
s = np.sqrt(dv*(N/(N-1))).round(2)

# In[53]:
```

```
k = N-1
gamma = 0.95
tg = 1.984

# In[54]:

di_a = (xv-tg*s/np.sqrt(N), xv+tg*s/np.sqrt(N))
xv
di_a

# In[58]:

q = 0.141
di_s = (s*(1-q), s*(1+q))
s
di_s

# In[241]:

alpha = 0.05

# In[242]:

df = int_row.copy().drop(['avg_inter', 'inter', 'rf'], axis=1)
df['xi'] = int_row['avg_inter']-h/2
df['xi+1'] = int_row['avg_inter']+h/2
df = df[['xi', 'xi+1', 'af']]
df = df.rename(columns={'af': 'ni'})
df.iloc[6, 0], df.iloc[6, 1] = 542, 576
df['zi'] = np.round((df['xi']-xv)/s, 2)
df['zi+1'] = np.round((df['xi+1']-xv)/s, 2)
df.loc[0, 'zi'], df.loc[6, 'zi+1'] = -np.inf, np.inf
```

```
# In[258]:

df['F(zi)'] = np.array([-5000,-4641,-3665,-1628,1064,3289,4495])/10000
df['F(zi+1)'] = np.array([-4641,-3665,-1628,1064,3289,4495,5000])/10000
df['pi'] = np.round(df['F(zi+1)'] - df['F(zi)'], 4)
df['ni*'] = np.round(df['pi']*N, 4)
df.to_csv('data/data1.csv', index=False)
df

# In[261]:

k = len(df)-3
(k, alpha)
hi_crit = 9.5
hi_nabl = np.dot((df['ni']-df['ni*'])**2, 1/df['ni*']).round(4)
(hi_nabl, hi_crit)
'True' if hi_nabl <= hi_crit else 'False'

# In[ ]:
```