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	_________________	Бабенко Н.С.
Студент гр. 8383	_________________	Сахаров В.М.
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.

Интервальной оценкой математического ожидания при неизвестном среднем квадратическом отклонении σ генеральной совокупности служит доверительный интервал:

$$\overline{x_в} - \frac{S}{\sqrt{n}} t_\gamma \le \alpha \le \overline{x_в} + \frac{S}{\sqrt{n}} t_\gamma,$$

$\overline{x_в}$ – статистическая оценка математического ожидания

S – исправленное СКВО

n – объём выборки

t_γ – из таблицы

Доверительный интервал для оценки СКВО:

$$S(1-q) \le \sigma \le S(1+q),$$

S – исправленное СКВО

q – из таблицы

Критерий Пирсона, или критерий χ^2(Хи-квадрат), 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 $\chi^2_{наб} \le \chi^2_{крит}$ - гипотеза принимается, иначе ($\chi^2_{наб} > \chi^2_{крит}$) – гипотезу отвергают.

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

При выполнении лабораторной работы №2 были получены выборочные данные, представленные в табл. 1.

Таблица 1

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

Объем выборки $N = 100$. Количество интервалов $k = 7$. Ширина интервала $h = 44$.

Статистическая оценка математического ожидания:

$$\bar{x}_в = \frac{1}{N}\sum_{i=1}^{k} x_i n_i = 465.26$$

Выборочная дисперсия:

$$D_в = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = 2948.352$$

Исправленная выборочная дисперсия:

$$s^2 = \frac{N}{N-1} D_в = \frac{100}{99} * 2948.352 = 2978.133$$

Статистическая оценка СКО:

$$s = \sqrt{s^2} = \sqrt{2978.133} = 54.572$$

1. Доверительный интервал для математического ожидания при неизвестном СКО

Вычислим доверительный интервал для оценки математического ожидания по формуле ниже:

$$\bar{x}_в - t_\gamma \frac{S}{\sqrt{n}} \le \alpha \le \bar{x}_в + t_\gamma \frac{S}{\sqrt{n}}, \text{где}$$

$\bar{x}_в$ – выборочное среднее

s – исправленное СКО

$t_\gamma = 2.627$ – из таблицы (при $\gamma = 0.99$, $N = 100$)

$$\bar{x}_в - t_\gamma \frac{s}{\sqrt{N}} = 465.26 - 2.627 * \frac{54.572}{10} = 450.92$$

$$\bar{x}_в + t_\gamma \frac{s}{\sqrt{N}} = 465.26 + 2.627 * \frac{54.572}{10} = 479.6$$

Доверительный интервал $(450.92;\ 479.6)$ покрывает истинное значение математического ожидания α с надежностью $\gamma = 0.99$.

2. Построим доверительный интервал для среднеквадратического отклонения.

Доверительный интервал для оценки СКО:

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

s – исправленное СКО

$q = 0.198$ –из таблицы (при $\gamma = 0.99$, $N = 100$)

$$s(1-q) = 54.572 * 0.802 = 43.767$$

$$s(1+q) = 54.572 * 1.198 = 66.377$$

Доверительный интервал $(43.767;\ 66.377)$ покрывает истинное значение среднеквадратического отклонения σ с надежностью $\gamma = 0.99$.

3. 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:

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

$$\chi^2_{\text{крит}} = \chi^2(\alpha, k) - \text{из таблицы}$$

Гипотеза H_0 принимается при условии:

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

Вычислим теоретические частоты. Вычисления представлены в табл. 2.

Таблица 2

$\mathbf{x_i}$	$\mathbf{x_{i+1}}$	$\mathbf{n_i}$	$\mathbf{z_i}$	$\mathbf{z_{i+1}}$	$\Phi(\mathbf{z_i})$	$\Phi(\mathbf{z_{i+1}})$	$\mathbf{p_i}$	$\mathbf{n_i}'$
321.0	365.0	4	$-\infty$	-1.84	-0.5	-0.4671	0.0329	3.29
365.0	409.0	9	-1.84	-1.03	-0.4671	-0.3485	0.1186	11.86
409.0	453.0	27	-1.03	-0.22	-0.3485	-0.0871	0.2614	26.14
453.0	497.0	35	-0.22	0.58	-0.0871	0.219	0.3061	30.61
497.0	541.0	17	0.58	1.39	0.219	0.4177	0.1987	19.87

541.0	585.0	6	1.39	2.19	0.4177	0.4858	0.0681	6.81
585.0	623.0	2	2.19	$+\infty$	0.4858	0.5	0.0142	1.42

Вычислим наблюдаемое значение критерия $\chi^2_{\text{набл}}$ с помощью полученных частот по формуле ниже. Отдельные вычисления представлены в табл. 3.

$$\chi^2_{\text{наб}} = \sum_{i=1}^{7} \frac{(n_i - n_i^{'})^2}{n_i^{'}}$$

Таблица 3

$\mathbf{n_i}$	$\mathbf{n_i}'$	$\mathbf{n_i - n_i'}$	$\mathbf{(n_i - n_i')^2}$	$\mathbf{(n_i - n_i')^2/n_i'}$
4	3.29	0.71	0.5041	0.1532
9	11.86	-2.86	8.1796	0.6897
27	26.14	0.86	0.7396	0.0283
35	30.61	4.39	19.2721	0.6296
17	19.87	-2.87	8.2369	0.4145
6	6.81	-0.81	0.6561	0.0963
2	1.42	0.58	0.3364	0.2369

$$\chi^2_{\text{набл}} = 2.2485$$

Определим табличное значение $\chi^2_{\text{крит}}$ при $\alpha = 0.05$ и $k = 7 - 3 = 4$:

$$\chi^2_{\text{крит}} = 9.5$$

Сравним полученные значения:

$$\chi^2_{\text{набл}} = 2.2485 \leq \chi^2_{\text{крит}} = 9.5$$

Из полученных результатов можно сделать вывод, что нулевая гипотеза принимается, то есть можно 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 $\gamma = 0.99$. Исходя из полученных результатов можно сделать вывод, что доверительный интервал $(450.92;\ 479.6)$ покрывает истинное значение математического ожидания α с надежностью $\gamma = 0.99$. Вычислен доверительный интервал для среднеквадратического отклонения. Можно сделать вывод, что доверительный интервал $(43.767;\ 66.377)$ покрывает истинное значение среднеквадратического отклонения σ с надежностью $\gamma = 0.99$.

Выполнена проверка гипотезы о нормальности заданного распределения с помощью критерия χ^2 (Пирсона). Определено, что $\chi^2_{\text{набл}} \leq \chi^2_{\text{крит}}$, следовательно, нулевая гипотеза принимается, то есть можно предположить, что случайная величина распределена по нормальному закону распределения.

ПРИЛОЖЕНИЕ А
ИСХОДНЫЙ КОД

```
import numpy as np
import pandas as pd
import scipy
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
int_row = pd.read_csv('c:/Users/gandh/dev/unv/smoed/NB/data/inter-
val.csv')
N = int_row['af'].sum()
h = 44
N
int_row
xv = (np.dot(int_row['avg_inter'], int_row['af'])/N).round(3)
dv = (np.dot((int_row['avg_inter']-xv)**2,
int_row['af'])/N).round(3)
s = np.sqrt(dv*(N/(N-1))).round(3)
xv, dv, (dv*(N/(N-1))).round(3), s
gamma = 0.99
tg = 2.627
di_a = np.round((xv-tg*s/np.sqrt(N), xv+tg*s/np.sqrt(N)), 2)
xv
di_a
q = 0.198
di_s = np.round((s*(1-q), s*(1+q)), 3)
s
di_s
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] = 585, 623
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
df['F(zi)'] = np.array([-5000,-4671,-3485,-871,2190,4177,4858])/10000
df['F(zi+1)'] = np.array([-4671,-3485,-871,2190,4177,4858,5000])/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
df_nabl = pd.DataFrame()
df_nabl['ni'], df_nabl['ni*'] = df['ni'], df['ni*']
df_nabl['-'] = np.round(df_nabl['ni']-df_nabl['ni*'], 4)
df_nabl['-2'] = np.round(df_nabl['-']**2, 4)
df_nabl['-2/'] = np.round(df_nabl['-2']/df_nabl['ni*'], 4)
df_nabl.to_csv('data/data2.csv', index=False)
hi_nabl = df_nabl['-2/'].sum().round(4)
df_nabl
alpha = 0.05
k = len(df)-3
(k, alpha)
hi_crit = 9.5
(hi_nabl, hi_crit)
'True' if hi_nabl <= hi_crit else 'False'
```