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

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

Критерий Пирсона, или критерий , 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:

,

где – функция Лапласа

Если – 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 . 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

Количество интервалов

Ширина интервала

Статистическая оценка математического ожидания:

Исправленная выборочная дисперсия:

Статистическая оценка СКО:

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

Случайная величина :

Эта случайная величина распределена по закону Стьюдента с степенями свободы. Справедливо соотношение:

Доверительный интервал для оценки математического ожидания:

– выборочное среднее

– исправленное СКО

– определено из соответствующей таблицы

(по заданным значениям , )

Можно сделать вывод, что интервал с вероятностью (надежностью) содержит в себе истинное значение математического ожидания.

* Построим доверительный интервал для среднеквадратического отклонения:

Доверительный интервал для оценки СКО:

– исправленное СКО

– определено из соответствующей таблицы

(по заданным значениям , )

Можно сделать вывод, что интервал с вероятностью (надежностью) содержит в себе истинное значение среднеквадратического отклонения.

* Проверим гипотезу о нормальности заданного распределения с помощью критерия Пирсона

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

Распределение хи-квадрат зависит от числа степеней свободы , которое вычисляется как . По числу степеней свободы и уровню значимости вычисляется значение . Область принятия гипотезы определяется условием:

Найдем теоретические частоты. Вычисления представлены в табл. 2.

*Таблица 2*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
| 320.0 | 357.0 | 5 |  | -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 |  | 0.4495 | 0.5 | 0.0505 | 5.252 |

Вычислим наблюдаемое значение критерия . Результаты представлены в табл. 3.

*Таблица 3*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 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 |

Найдем по заданному уровню значимости и числу степеней свободы :

Сравним с наблюдаемым значением:

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 . Исходя из полученных результатов можно сделать вывод, что интервал с вероятностью (надежностью) содержит в себе истинное значение математического ожидания.

Были вычислены границы доверительного интервала для среднеквадратического отклонения. Определено, что интервал с вероятностью (надежностью) содержит в себе истинное значение среднеквадратического отклонения.

Была выполнена проверка гипотезы о нормальности заданного распределения с помощью критерия (Пирсона). Было выяснено, что , 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[ ]: