**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 покрывает заданный параметр.

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

,

– статистическая оценка математического ожидания

– исправленное СКВО

– объём выборки

– из таблицы

Доверительный интервал для оценки СКВО:

– исправленное СКВО

­– из таблицы

Критерий Пирсона, или критерий (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:

,

где – функция Лапласа

Если - гипотеза принимается, иначе () – 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 . 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 . Количество интервалов . Ширина интервала .

Статистическая оценка математического ожидания:

Выборочная дисперсия:

Исправленная выборочная дисперсия:

Статистическая оценка СКО:

1. Доверительный интервал для математического ожидания при неизвестном СКО

Вычислим доверительный интервал для оценки математического ожидания по формуле ниже:

– выборочное среднее

– исправленное СКО

– из таблицы (при , )

Доверительный интервал покрывает истинное значение математического ожидания с надежностью .

1. Построим доверительный интервал для среднеквадратического отклонения.

Доверительный интервал для оценки СКО:

– исправленное СКО

–из таблицы (при , )

Доверительный интервал покрывает истинное значение среднеквадратического отклонения с надежностью .

1. 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:

Гипотеза принимается при условии:

Вычислим теоретические частоты. Вычисления представлены в табл. 2.

Таблица 2

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
| 321.0 | 365.0 | 4 |  | -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 |  | 0.4858 | 0.5 | 0.0142 | 1.42 |

Вычислим наблюдаемое значение критерия с помощью полученных частот по формуле ниже. Отдельные вычисления представлены в табл. 3.

Таблица 3

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

Определим табличное значение при и :

Сравним полученные значения:

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 . Исходя из полученных результатов можно сделать вывод, что доверительный интервал покрывает истинное значение математического ожидания с надежностью . Вычислен доверительный интервал для среднеквадратического отклонения. Можно сделать вывод, что доверительный интервал покрывает истинное значение среднеквадратического отклонения с надежностью .

Выполнена проверка гипотезы о нормальности заданного распределения с помощью критерия (Пирсона). Определено, что , 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/interval.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'