**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 оценок параметров распределения.

Проверка статистической гипотезы о нормальном законе распределения.

|  |  |  |
| --- | --- | --- |
| Студент гр. 6381 |  | Шарыпина Д.В. |
| Преподаватель |  | Фиалковский М.С. |

Санкт-Петербург

2020

**Цель 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 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.

**Основные теоретические положения.**

Оценка неизвестного параметра называется *интервальной*, если она определяется двумя числами – концами интервала.

Интервал , накрывающий с вероятностью истинное значение параметра , называется *доверительным интервалом*, а вероятность – *надежностью* оценки или *доверительной вероятностью*.

Число характеризует точность оценки: чем меньше разность , тем точнее оценка.

*Доверительные интервалы для параметров нормального распределения.*

Доверительный интервал заданной надежности для математического ожидания имеет следующий вид:

где – оценка математического ожидания, – исправленная оценка среднеквадратического отклонения, – объем выборки, – квантиль распределения Стьюдента уровня .

Доверительный интервал заданной надежности для среднеквадратического отклонения имеет следующий вид:

где – определяется по таблице по заданным и .

Критерий согласия Пирсона (критерий согласия ) позволяет проверить гипотезу о том, что случайная величина подчиняется определенному закону распределения. Данный критерий основан на сравнении теоретической плотности распределения и гистограммы.

Для того, чтобы проверить гипотезу о том, что выборка взяла из генеральной совокупности , имеющей распределение, заданное функцией распределения необходимо множество значений разбить на интервалы:

Для каждого интервала необходимо вычислить наблюдаемые частоты , а также теоретические (ожидаемые) частоты При этом интервалы должны быть сформированы таким образом, чтобы ожидаемые частоты удовлетворяли условию .

Наблюдаемые частоты равны количеству элементов выборки, попавших в соответствующий -ый интервал. Теоретические частоты вычисляются по формуле , где – объем выборки, а - вероятность попадания в -ый интервал, вычисляемая следующим образом:

где

а , где - граница интервала, – оценка математического ожидания, – исправленная оценка среднеквадратического отклонения.

Статистика критерия имеет вид:

Статистика критерия имеет распределение c степенями свободы, где – число интервалов, а – число неизвестных параметров распределения. В случае для нормального распределения .

Для проверки гипотезы необходимо сравнить значение статистики критерия, полученное по выборке с критическим значением для степенями свободы и на уровне значимости .

В случае, если , то гипотеза отвергается в пользу гипотезы , в противном случае у нас не достаточно оснований, чтобы отвергнуть нулевую гипотезу.

**Ход работы.**

*Доверительный интервал для математического ожидания.*

Найдем границы доверительного интервала для математического ожидания случайной величины, имеющей нормальное распределение, для случая неизвестной дисперсии.

Воспользуемся формулой:

В предыдущей лабораторной работе для выборки размера были получены следующие значения:

По таблице для и получим

Тогда получим доверительный интервал для математического ожидания:

*Доверительный интервал для среднеквадратического отклонения.*

Найдем границы доверительного интервала для СКО.

Для этого по таблице найдем для и . Получим .

Тогда используя формулу

получим следующий вид доверительного интервала для СКО:

*Проверка гипотезы с помощью критерия Пирсона ( критерия).*

Проверим гипотезу о согласии с нормальным распределением с математическим ожиданием 455.17 и среднеквадратическим отклонением 54.3. Для этого воспользуемся критерием для проверки сложной гипотезы.

Сначала сформулируем гипотезы и .а

выборка взята из нормального распределения с параметрами и .

:выборка взята не из нормального распределения с параметрами и .

Для проверки гипотезы с помощью критерия сформируем интервалы таким образом, чтобы ожидаемые частоты были не менее

Если брать интервалы, совпадающие с интервальным рядом, то получим, что для самого правого и самого левого интервалов не выполняется условие . Поэтому объединим крайние интервалы слева и справа, тогда получим следующие 5 интервалов и промежуточные значения для каждого из интервалов:

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

Статистика критерия, посчитанная по выборке

Критическое значение статистики критерия на уровне значимости и степеней свободы получаем .

В таком случае , это означает, что у нас нет оснований отвергнуть гипотезу .

**Выводы.**

В ходе выполнения лабораторной работы были получены практические навыки нахождения интервальных статистических оценок для параметров нормального распределения – математического ожидания и среднеквадратического отклонения.

В результате выполнения работы была получена интервальная оценка математического ожидания генеральной совокупности, покрывающая его с 95% доверительной вероятностью: .

Также в результате работы была получена интервальная оценка среднеквадратического отклонения генеральной совокупности, покрывающая его с 95% доверительной вероятностью: .

В ходе выполнения лабораторной работы была проверена гипотеза согласия с нормальным распределением. Гипотеза проверена с использованием критерия на уровне значимости . В результате было получено, что у нас нет оснований отклонять гипотезу о том, что выборка получена генеральной совокупности, имеющей нормальное распределение.

**ПРИЛОЖЕНИЕ А**

**ИСХОДНЫЙ КОД**

#!/usr/bin/env python

# coding: utf-8

# ## Лабораторная работа 3

from scipy.stats import t

from math import sqrt

# доверительный интервал для МО

mean = 455.17

sd = 54.3

n = 94

left = mean + t.ppf(0.025, df=n-1)\*sd/sqrt(n)

right = mean + t.ppf(0.975, df=n-1)\*sd/sqrt(n)

print('{} <= a <= {}'.format(left, right))

# доверительный интервал для СКО

gamma = 0.95

q = 0.151

left = sd\*(1-q)

right = sd\*(1+q)

print('{} <= sd <= {}'.format(left, right))

# ---------------

# границы интервалов в интервальном ряду

bins = [320]

h = 39 # шаг

for i in range(1, 8):

bins.append(bins[i-1] + h)

# стандартизуем границы интервалов

z = [(e-mean)/(sd) for e in bins]

# объединим два интервала справа и два слева

# назначим самой левой и самой правой границам значение inf

z[0] = -float('inf')

z[-1] = float('inf')

z.pop(1)

z.pop(-2)

print('Z = {}'.format(z))

# частоты по интервальному ряду

vals = [4, 11, 16, 30, 21, 9, 3]

# частоты после объединения интервалов

vals = [13, 16, 30, 21, 12]

# объем выборки

n = sum(vals)

from scipy.special import erf

# определим функцию Лапласа

Ф = lambda x: erf(x/2\*\*0.5)/2

# посчитаем вероятности попадания в интервалы

p = []

for i in range(1, len(z)):

p.append(Ф(z[i])-Ф(z[i-1]))

print('Вероятности попадания в интервалы = {}'.format(p))

# теоретические частоты по интервалам

n\_theory = list(map(lambda x: x\*n, p))

print('Ожидаемые частоты = {}'.format(n\_theory))

# наблюдаемые частоты

n\_observed = vals

print('Наблюдаемые частоты = {}'.format(n\_observed))

# посчитаем статистику критерия Хи-квадрат

res = []

for i in range(len(n\_observed)):

res.append((n\_observed[i] - n\_theory[i])\*\*2/n\_theory[i])

print(res)

chi2\_observed = sum(res)

print('Значение статистики критерия: {}'.format(chi2\_observed))

from scipy.stats import chi2

k = 5 # число интервалов

l = 2 # число параметров

alpha = 0.05 # уровень значимости

# критическое значение статистики критерия

chi2\_criterion = chi2.ppf(1-alpha, df=k-l-1)

print('Критическое значение: {}'.format(chi2\_criterion))

# принимаем решение

if chi2\_observed > chi2\_criterion:

print('Отвергаем гипотезу H0')

else:

print('Не можем отвергнуть гипотезу H0')

Список литературы

1. Лекции по математической статистике Computer Science Center.
2. Лагутин М. Б. Наглядная математическая статистика. (Том 2, стр. 174) — М.: П-центр, 2003.
3. Севастьянов Б. А. Курс теории вероятностей и математической статистики. — Москва-Ижевск: Институт компьютерных исследований, 2004.