МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ «ЛЭТИ»
ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра алгоритмической математики

ОТЧЁТ

по практической работе №3
по дисциплине «Статистический анализ»
Тема: Обработка выборочных данных.
Нахождение интервальных оценок параметров распределения.
Проверка статистической гипотезы о нормальном законе распределения

Студент гр. 9372	____________________	Иванов Р. С.
Преподаватель	____________________	Сучков А. И.

Санкт-Петербург
2021

Цель работы

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

Основные теоретические положения

Определение 1. *Интервальной статистической оценкой* называется интервал (θ_1^*, θ_2^*), накрывающий с вероятностью γ истинное значение параметра θ, где γ называется *надежностью оценки*.

Алгоритм нахождения доверительного интервала для математического ожидания при неизвестном СКО

Будем искать доверительный интервал в виде $x_{\text{г}} \in (\overline{x}_{\text{в}} - \varepsilon, \overline{x}_{\text{в}} + \varepsilon)$. Тогда при наших условиях:

$$\varepsilon = \frac{t_\gamma \cdot s}{\sqrt{n}},$$

где s - несмещенная оценка $\sigma_{\text{в}}$; $t_\gamma = t(\gamma, n)$ - константа распределения Стьюдента.

Алгоритм нахождения доверительного интервала для оценки СКО

Возьмем $q = q(\gamma, n)$ - табличное значение. Тогда

$$\sigma_{\text{в}} \in \begin{cases} q < 1, & \text{(s - sq, s + sq)} \\ q > 1, & \text{(0, s + sq) ,} \end{cases}$$

где s - исправленное СКО.

Критерий χ^2 Пирсона. Алгоритм для нормального распределения.

Пусть m_i - *эмпирические частоты*, тогда назовем ***теоретическими частотами*** числа $m_i' = n \cdot p_i$, где n - объем выборки, p_i - точечная вероятность варианты x_i дискретной случайной величины.

Алгоритм действий

1. Вычислить точечные вероятности p_i. Для расчета попадания случайной величины X в интервал $(x_{i-1}, x_i]$, используем функцию *Лапласа*:

$$p_i = P\{x_{i-1} < X \leqslant x_i\} = \Phi\left(\frac{x_i - x_{\text{в}}}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - x_{\text{в}}}{\sigma}\right)$$

2. Вычислить выравнивающие(теоретические) частости m_i.

3. Найти статистику $\chi^2_{\text{набл}} = \sum\limits_{i=1}^{k} \frac{(m_i - m_i')^2}{m_i'}$

4. Определить число степеней свободы по формуле $l = k - 3$, где k - число частичных интервалов выборки.

5. По таблице найти критическую величину $\chi^2_{\alpha;l}$, где α - заданный уровень значимости.

6. Сделать выводы. Если $\chi^2_{\text{набл}} < \chi^2_{\alpha;l}$, то закон теоритического распределения не противоречит опытным данным.

Постановка задачи

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

Выполнение работы

1.

Найдем доверительный интервал математического ожидания

- при $\gamma = 0.95$:

$$\overline{x} = 50225$$
$$s = 39915$$
$$\varepsilon = 7373$$
$$x \in (42852, 57599),$$

- при $\gamma = 0.99$

$$\varepsilon = 9752$$
$$x \in (40473, 59977)$$

2.

Найдем границы доверительного интервала для СКО

- при $\gamma = 0.95$

$$q_\gamma = 0.143$$
$$\sigma_{\text{в}} \in (34207, 45623)$$

- при $\gamma = 0.99$

$$q_\gamma = 0.198$$
$$\sigma_{\text{в}} \in (32012, 47819)$$

3. Проверка гипотезы о нормальном распределении

Вычислим теоретические частоты:

i	1	2	3	4	5	6	7	8	9
p_i	0.2776	0.2264	0.2783	0.1456	0.0559	0.0138	0.002	0.0001	0.000001
m_i	32	26	32	16.74	6.43	1.59	0.25	0.0214	0.0016

Таблица 1 – Вычисления

Число свободы $l = k - 3 = 9 - 3 = 6$. Вычислим $\chi^2_{\text{набл}}$ и сравним его с табличным значением при $\alpha = 0.05$:

$$\chi^2_{\text{набл}} = 673$$
$$\chi^2_{\alpha;l} = 12.6$$

Значение $\chi^2_{\text{набл}}$ получилось таким большим из-за того что последний элемент m_i очень близок к 0, и при делении последнее слагаемое в $\chi^2_{\text{набл}}$ становится равным 620.

Выводы

В процессе работы мы познакомились с алгоритмами вычисления доверительных интервалов для математического ожидания и среднеквадратичного отклонения. Научились работать с функцией *Лапласа*, проверять гипотезы о выбранном законе распределения для выборки с помощью критерия Пирсона.

По найденному значению $\chi^2_{\text{набл}}$ можно смело отвергнуть гипотезу о нормальном распределении выборки.

ПРИЛОЖЕНИЕ А

```
# import numpy as np
# from nltk.probability import FreqDist
# import matplotlib.pyplot as plt
import math
import csv
# import scipy
import random
# from openpyxl import Workbook
# from openpyxl.drawing.image import Image

sample = []

with open(’Price-Mileage.csv’) as csv_file:  # Читаем выборку из файла 0 работы
    spam_reader = csv.reader(csv_file, quotechar=’|’)
    for row in spam_reader:
        x, y = row[0].split(’;’)
        if y.isdigit(): sample.append(int(y))

sample.sort()  # Используя встроенную функцию сортировки получаем ранжированный ряд

R = sample[len(sample) - 1] - sample[0]  # Размах
print(”R =”, R)

k = round(1 + math.log2(len(sample)))  # Число интерваловФормула( Стёрджеса)
print(”k = ”, k)

h = round(R / k)  # Длина интервала
print(”h =”, h)
```

```python

k += 1  # Иначе интервалы не покроют выборку

x0 = sample[0] - h / 2  # Начало первого частичного интервала
if x0 < 0:
    x0 = 0

print(”x0 =”, x0)

interval = []
x = x0

for i in range(k):
    interval.append([(x, x + h), 0, 0])
    x += h

# Получаем интервальный ряд
for i in sample:
    for j in range(k):
        if interval[j][0][0] < i <= interval[j][0][1]:
            interval[j][1] += 1
            break

for i in interval:
    i[2] = i[1] / len(sample)

print(interval)

middle_int = []

# Вычисляем серидины интервалов
for i in range(len(interval)):
    middle_int.append(interval[i][0][0] + h / 2)
```

```
n = len(sample)
Xs = 0  # Выброчное среднее по x
Dsx = 0  # Дисперсия по x
s = 0  # СКО по x
for i in range(n):
    Xs += sample[i]
Xs = Xs / n
for i in range(n):
    Dsx += (sample[i] - Xs) ** 2
Dsx = Dsx / n
d = math.sqrt(Dsx)
s = math.sqrt(Dsx) * n / (n - 1)
TY1 = 1.981       # t гамма для 95% n = 115
TY2 = 2.62        # t гамма для 99% n = 115
D1 = TY1*s/math.sqrt(n)
D2 = TY2*s/math.sqrt(n)
LB = Xs - D1
LR = Xs + D1
LB2 = Xs - D2
LB2 = Xs + D2
Q1 = 0.143       # q для n=100 95%
Q2 = 0.198       # q для n=100 99%
SLB1 = s - s*Q1
SRB1 = s + s*Q1
SLB2 = s - s*Q2
SRB2 = s + s*Q2
```

```python

# Пункт 3

laplas_arg = []
for i in range(8):
    laplas_arg.append((interval[i][0][1]-Xs)/d)

laplas = [-0.5, -0.2224, 0.0040, 0.2823, 0.4279, 0.4838, 0.4976, 0.4998, 0.499986, 0.5]
p = []        # Исправленные частоты
for i in range(k):
    p.append(laplas[i + 1] - laplas[i])

m = []        # m'
for i in range(k):
    m.append(p[i]*115)

m2 = []        # (m-m')/m'
for i in range(k-1):
    m2.append(((interval[i][1]-m[i])**2)/m[i])

ans = sum(m2)        # Последнее слагаемое в сумме равно 620, за счёт деления 1 на число близкое к 0
Xi2 = 12.6
```

Листинг 1 – Исходный код программы