Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский Томский политехнический Университет»



Инженерная школа ядерных технологий

Отделение математики и математической физики

01.03.02 Прикладная математика и информатика

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Лабораторная работа № 5

**Вариант - 6**

по дисциплине:

**Математическая статистика**

|  |  |  |  |
| --- | --- | --- | --- |
| **Выполнил:** |  | | |
| студент группы | 0В21 | Дзебан А.А. |  |
|  |  |  |  |
| **Проверил:** | Шинкеев М.Л. | | |
|  |  |  |  |
|  |  |  |

Томск – 2024

Исследовать влияние на среднее значение результирующего показателя факторов A и B и их взаимодействия, используя многофакторный дисперсионный анализа на основе линейных моделей с разложением сумм квадратов по типу, соответствующему варианту задания: для вариантов 1, 4, 7, … (остаток от деления на 3 номера варианта равен 1) - использовать I тип разложения; для вариантов 2, 5, 8, … (остаток от деления на 3 номера варианта равен 2) - использовать II тип разложения; для вариантов 3, 6, 9, … (остаток от деления на 3 номера варианта равен 0) - использовать III тип разложения.

В случае, если эффект фактора (эффект взаимодействия факторов) значим, провести анализ значимости попарных различий средних значений отклика, используя один из методов апостериорного множественного сравнения средних.

**Ход работы:**

Исходные данные:

|  |  |  |
| --- | --- | --- |
| Вариант | 6 |  |
| Фактор A | Фактор B | Отклик |
| 1 | 1 | 3,33 |
| 1 | 1 | 2,75 |
| 1 | 1 | 3,32 |
| 1 | 1 | 1,44 |
| 1 | 1 | 4,41 |
| 1 | 2 | 5,64 |
| 1 | 2 | 6,96 |
| 1 | 2 | 8,15 |
| 2 | 1 | 6,99 |
| 2 | 1 | 5,6 |
| 2 | 1 | 5,02 |
| 2 | 1 | 5,01 |
| 2 | 2 | 3,24 |
| 2 | 2 | 4,1 |
| 2 | 2 | 5,12 |
| 2 | 2 | 3,16 |
| 2 | 2 | 3,84 |
| 2 | 2 | 3,82 |
| 3 | 1 | 6,51 |
| 3 | 1 | 6,25 |
| 3 | 1 | 6,61 |
| 3 | 1 | 7,17 |
| 3 | 1 | 6,92 |
| 3 | 1 | 4,1 |
| 3 | 2 | 8,21 |
| 3 | 2 | 7,83 |
| 3 | 2 | 5,36 |
| 3 | 2 | 7,86 |

**Используем 3 тип разложения:**

Результаты многофакторной модели ANOVA:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Cумм кв | df | F | p-val |
| A | 30.241873 | 2.0 | 13.712723 | 0.000136 |
| B | 28.033333 | 1.0 | 25.422588 | 0.000048 |
| A:B | 33.697931 | 2.0 | 15.279821 | 0.000069 |

Исходя из значений p-val оба фактора по отдельности и их взаимодействие являются значимыми.

Тогда воспользуемся методом Тьюки для попарного сравнения средних значений отклика

Для параметра А:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Ур 1** | **Ур 2** | **Разн ср** | **p** | **Нижн ДИ** | **Верхн ДИ** | **Отклон** |
| 1 | 2 | 0,09 | 0,9921 | -1,7785 | 1,9585 | ЛОЖЬ |
| 1 | 3 | 2,182 | 0,0198 | 0,3135 | 4,0505 | ИСТИНА |
| 2 | 3 | 2,092 | 0,0177 | 0,3304 | 3,8536 | ИСТИНА |

Делаем вывод о равенстве средних значений отклика для уровней 1,2 и делаем выводы о том, что средние значения отклика значимо различаются для уровней 2,3 и 1,3

Аналогично для параметра B:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Группа 1** | **Группа 2** | **Ср разн** | **p-val** | **Нижн ДУ** | **Верх ДУ** | **Отклон** |
| 1 | 2 | 0,609 | 0,3935 | -0,8337 | 2,0517 | ЛОЖЬ |

Делаем вывод о равенстве средних значений отклика для уровней 1,2.

Для взаимодействия параметров A:B – будем оценивать равность средних значений отклика для каждой комбинации уровней факторов.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Группа 1** | **Группа 2** | **Ср разн** | **p-val** | **Нижн ДУ** | **Верх ДУ** | **Отклон** |
| 1\_1 | 1\_2 | 3,8667 | 0,0006 | 1,4777 | 6,2556 | ИСТИНА |
| 1\_1 | 2\_1 | 2,605 | 0,0139 | 0,4106 | 4,7994 | ИСТИНА |
| 1\_1 | 2\_2 | 0,83 | 0,7791 | -1,1508 | 2,8108 | ЛОЖЬ |
| 1\_1 | 3\_1 | 3,21 | 0,0006 | 1,2292 | 5,1908 | ИСТИНА |
| 1\_1 | 3\_2 | 4,265 | 0,0001 | 2,0706 | 6,4594 | ИСТИНА |
| 1\_2 | 2\_1 | -1,2617 | 0,6235 | -3,7601 | 1,2367 | ЛОЖЬ |
| 1\_2 | 2\_2 | -3,0367 | 0,0057 | -5,3497 | -0,7236 | ИСТИНА |
| 1\_2 | 3\_1 | -0,6567 | 0,9463 | -2,9697 | 1,6564 | ЛОЖЬ |
| 1\_2 | 3\_2 | 0,3983 | 0,9958 | -2,1001 | 2,8967 | ЛОЖЬ |
| 2\_1 | 2\_2 | -1,775 | 0,1345 | -3,8865 | 0,3365 | ЛОЖЬ |
| 2\_1 | 3\_1 | 0,605 | 0,9443 | -1,5065 | 2,7165 | ЛОЖЬ |
| 2\_1 | 3\_2 | 1,66 | 0,2619 | -0,6531 | 3,9731 | ЛОЖЬ |
| 2\_2 | 3\_1 | 2,38 | 0,0083 | 0,4914 | 4,2686 | ИСТИНА |
| 2\_2 | 3\_2 | 3,435 | 0,0006 | 1,3235 | 5,5465 | ИСТИНА |
| 3\_1 | 3\_2 | 1,055 | 0,6336 | -1,0565 | 3,1665 | ЛОЖЬ |

Здесь нулевая гипотеза о равенстве средних принимается для взаимодействий:

**1\_1** vs **2\_2**, **1\_2** vs **2\_1**, **1\_2** vs **3\_1**, **1\_2** vs **3\_2**, **2\_1** vs **2\_2**, **2\_1**vs **3\_1**, **2\_1** vs **3\_2**, **3\_1** vs **3\_2**

**Вывод**: выяснилось, что факторы A и B значимо влияют на отклик, а их взаимодействие усиливает этот эффект. Попарные сравнения позволили выявить конкретные комбинации уровней факторов, которые обеспечивают наилучшие (наихудшие) результаты.

Приложение:

# %%

import pandas as pd

import statsmodels.api as sm

from statsmodels.formula.api import ols

from statsmodels.stats.multicomp import pairwise\_tukeyhsd

from scipy.stats import shapiro, levene

import matplotlib.pyplot as plt

# %%

# Создаем DataFrame из данных

data = pd.DataFrame({

    'A': [1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3],

    'B': [1, 1, 1, 1, 1, 2, 2, 2, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2],

    'Отклик': [3.33, 2.75, 3.32, 1.44, 4.41, 5.64, 6.96, 8.15, 6.99, 5.60, 5.02, 5.01, 3.24, 4.10, 5.12, 3.16, 3.84, 3.82, 6.51, 6.25, 6.61, 7.17, 6.92, 4.10, 8.21, 7.83, 5.36, 7.86]

})

# %%

# Построение линейной модели с взаимодействием

model = ols('Отклик ~ C(A) + C(B) + C(A):C(B)', data=data).fit()

# Многофакторный ANOVA с III типом разложения сумм квадратов

anova\_table = sm.stats.anova\_lm(model, typ=3)

print("Многофакторный ANOVA (тип III):")

print(anova\_table)

# %%

# Проверка значимости факторов и взаимодействия

if anova\_table['PR(>F)']['C(A)'] < 0.05:

    print("\nФактор A значим. Проводим попарное сравнение для A:")

    tukey\_a = pairwise\_tukeyhsd(endog=data['Отклик'], groups=data['A'], alpha=0.05)

    print(tukey\_a)

if anova\_table['PR(>F)']['C(B)'] < 0.05:

    print("\nФактор B значим. Проводим попарное сравнение для B:")

    tukey\_b = pairwise\_tukeyhsd(endog=data['Отклик'], groups=data['B'], alpha=0.05)

    print(tukey\_b)

if anova\_table['PR(>F)']['C(A):C(B)'] < 0.05:

    print("\nВзаимодействие A и B значимо. Проводим попарное сравнение для взаимодействия:")

    data['A\_B'] = data['A'].astype(str) + "\_" + data['B'].astype(str)

    tukey\_interaction = pairwise\_tukeyhsd(endog=data['Отклик'], groups=data['A\_B'], alpha=0.05)

    print(tukey\_interaction)