# Министерство науки и высшего образования Российской Федерации

## Федеральное государственное бюджетное образовательное учреждение высшего образования

## «Новосибирский государственный технический университет»



## Кафедра теоретической и прикладной информатики

### Лабораторная работа №1 по дисциплине «Основы теории машинного обучения»

**ОЦЕНКА ПАРАМЕТРОВ И ПРОВЕРКА ГИПОТЕЗ В ЛИНЕЙНЫХ**

**МОДЕЛЯХ С КАЧЕСТВЕННЫМИ ФАКТОРАМИ**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Факультет: | ПМИ |  |  |
| Группа: | ПМИМ-01 |  |  |
| Студенты: | Ершов П.К.  Малышкина Е.Д.  Слободчикова А.Э. | | |
| Вариант: | 4 |  |  |
| Преподаватель: | Попов А.А. |  |  |
|  |  |  |  |  |

Новосибирск

2021

1. **Задание**
2. По имеющимся данным (см. варианты заданий) сформировать матрицу наблюдений X, постулировать модель дисперсионного анализа с главными эффектами (без взаимодействий уровней факторов);
3. Провести редукцию модели к модели полного ранга, определить базис ФДО;
4. По методу МНК- оценивания провести оценивание ФДО в редуцированной модели. Проверить гипотезы о незначимости различий в эффектах уровней для каждого фактора и фактора в целом;
5. Отчет должен содержать постановочную часть, решения по редукции модели, компьютерный листинг, результаты расчетов по проверке гипотез, статистические выводы.

**Данные варианта:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровни фактора 1 | Уровни фактора 2 | | | |
| *B1* | *B2* | *B3* | *B4* |
| *A1* | 7,14  6,9 | 4,11  3,95 | 8,13  7,97 | 4,07  3,99 |
| *A2* | 3,1  2,91 | 0,0  0,01 | 4,03  3,98 | 0,02  0,0 |
| *A3* | 7,09  6,92 | 4,11  3,96 | 8,12  8,01 | 4,07  4,0 |

1. **Ход работы**

В ячейках таблиц располагаются значения отклика, полученные в двух параллельных наблюдениях. Пересечение соответствующих строки и столбца определяет условия эксперимента. Таким образом, количество наблюдений – 2, количество уровней фактора 1 – 3, количество уровней фактора 2 – 4.

1. Для построения модели вида

где – наблюдаемый в эксперименте отклик, измеряемый в количественной шкале;

– матрица наблюдения, порождаемая факторами

– вектор неизвестных параметров, подлежащих оцениванию из эксперимента;

– аддитивная случайная составляющая модели наблюдения;

необходимо сформировать матрицу наблюдений :

Затем вычислим – среднее значение измеряемой величины *y* по -ой подгруппе, где :

|  |  |  |  |
| --- | --- | --- | --- |
|  | Фактор 2 |  |  |
|  | Фактор 1 | B1 | B2 | B3 | B4 | Среднее значение по строкам А | |
| y1 | A1 | 7,14 | 4,11 | 8,13 | 4,07 | 5,78 |
| y2 |  | 6,90 | 3,95 | 7,97 | 3,99 |  |
| y1 | A2 | 3,10 | 0,00 | 4,03 | 0,02 | 1,76 |
| y2 |  | 2,91 | 0,01 | 3,98 | 0,00 |  |
| y1 | A3 | 7,09 | 4,11 | 8,12 | 4,07 | 5,79 | Общее сред. по A |
| y2 |  | 6,92 | 3,96 | 8,01 | 4,00 |  | 4,44 |
| Среднее значение  по столбцам B | | 5,68 | 2,69 | 6,71 | 2,69 |  |
|  | | 4,44 |  |
|  | Генеральное среднее | | 4,44 |

Затем вычислим генеральное среднее по следующим формулам:

( 4 \* 5,78 + 4 \* 1,76 + 4 \* 5,79 + 3 \* 5,68 + 3 \* 2,69 + 3 \* 6,71 + 3 \* 2,69) = 4,44.

Таким образом, можем представить модель дисперсионного анализа в виде:

где – эффект -го уровня первого фактора (),– эффект -го уровня второго фактора (), – ошибка эксперимента*;*

*.*

1. Редуцирование построенной модели к модели полного ранга можно проводить через факторизацию матрицы , где – матрица полного строчного ранга:

Где матрица задает базис ФДО и имеет вид , – единичная матрица размера *r*, .

В модели в связи с ее внутренним дефектом ранга несмещенно будут оцениваться только линейно независимых функций, допускающих оценку (ФДО), образующих базис ФДО.

Вычислим :

В матрицу будут входить и следующие уровни 1 фактора: А1, А2, 2 фактора: B1, B2, B3. В матрицу – уровни А3 и B4 факторов 1 и 2 соответственно.

Таким образом:

Значит, примет следующий вид:

Известно, что матрица имеет ранг и ненулевые ее строки образуют базис ФДО, в нашем случае:

Таким образом, матрица задает вид базиса ФДО. Базис ФДО составят следующие параметрические функции:

1. Чтобы найти оценки для ФДО достаточно применить теорему Гаусса-Маркова для построенной модели: .

Тогда:

6.90, 3.95, 7.97, 3.99, 2.91, 0.01, 3.98, 0, 6.92, 3.96, 8.01, 4.00

7.02, 4.03, 8.05, 4.03, 3.01, 0.01, 4.01, 0.01, 7.01, 4.04, 8.07, 4.04

Найдем оценки для 2 параллельных исследований:

Таким образом,

*.*

Были получены оценки для , уровней А1, А2 и B1, B2, B3. Оценки параметров, соответствующие исключенным из модели линейно зависимым столбцам (регрессорам) – А3 и B4, автоматически считаются равными нулю.

Проверим гипотезы о незначимости различий в эффектах уровней для каждого фактора:

– гипотеза не отвергается,

– гипотеза отвергается,

– гипотеза отвергается.

– гипотеза отвергается,

– гипотеза не отвергается,

– гипотеза отвергается,

– гипотеза отвергается,

– гипотеза отвергается.

Исходя из результатов проверок гипотез, проверим гипотезы о незначимости факторов в целом:

– гипотеза отвергается,

– гипотеза отвергается.

1. **Статистические выводы**

В ходе проведённой работы было установлено, что оба фактора являются значимыми. В тоже время, была установленная не значимость эффектов первого и третьего уровня у первого фактора и не значимость эффектов второго и четвёртого уровня у второго фактора.

1. **Код**

**import** numpy **as** p  
**from** numpy **import** matrix  
**from** numpy **import** linalg  
  
X1 = matrix([[4.44, 1, 0, 1, 0, 0],  
 [4.44, 1, 0, 0, 1, 0],  
 [4.44, 1, 0, 0, 0, 1],  
 [4.44, 1, 0, 0, 0, 0],  
 [4.44, 0, 1, 1, 0, 0],  
 [4.44, 0, 1, 0, 1, 0],  
 [4.44, 0, 1, 0, 0, 1],  
 [4.44, 0, 1, 0, 0, 0],  
 [4.44, 0, 0, 1, 0, 0],  
 [4.44, 0, 0, 0, 1, 0],  
 [4.44, 0, 0, 0, 0, 1],  
 [4.44, 0, 0, 0, 0, 0]])  
  
print(X1)  
X2 = matrix([[0, 0],  
 [0, 0],  
 [0, 0],  
 [0, 1],  
 [0, 0],  
 [0, 0],  
 [0, 0],  
 [0, 1],  
 [1, 0],  
 [1, 0],  
 [1, 0],  
 [1, 1]])  
print(X2)  
  
y1 = matrix([7.14, 4.11, 8.13, 4.07, 3.10, 0, 4.03, 0.02, 7.09, 4.11, 8.12, 4.07])  
print(y1)  
y2 = matrix([6.90, 3.95, 7.97, 3.99, 2.91, 0.01, 3.98, 0, 6.92, 3.96, 8.01, 4.00])  
print(y2)  
y3 = matrix([7.02, 4.03, 8.05, 4.03, 3.01, 0.01, 4.01, 0.01, 7.01, 4.04, 8.07, 4.04])  
  
A\_ = (X1.T \* X1).I \* X1.T \* X2  
print(**"A\_:"**)  
A\_l = A\_.shape  
**for** i **in** range(A\_l[0]):  
 **for** j **in** range(A\_l[1]):  
 print(**"%.4f"** % A\_[i, j], end=**' '**)  
 print(**"\n"**)  
  
print(**"Проверка:"**)  
Pr = X1 \* A\_  
print(Pr)  
  
print(**"Оценки тетта1:"**)  
theta1 = (X1.T \* X1).I \* X1.T \* y1.T  
**for** i **in** range(theta1.shape[0]):  
 print(**"%.2f"** % theta1[i])  
  
print(**"Оценки тетта2:"**)  
theta2 = (X1.T \* X1).I \* X1.T \* y2.T  
**for** i **in** range(theta2.shape[0]):  
 print(**"%.2f"** % theta2[i])  
  
print(**"Оценки тетта:"**)  
theta = (X1.T \* X1).I \* X1.T \* y3.T  
**for** i **in** range(theta.shape[0]):  
 print(**"%.2f"** % theta[i])