

Министерство образования Республики Беларусь

Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет Информационных технологий и управления
Кафедра Интеллектуальных информационных технологий

Индивидуальная практическая работа №2
по дисциплине «Статистические основы индуктивного вывода»
на тему
**ДИСКРИМИНАНТНЫЙ АНАЛИЗ С ГОТОВЫМИ
КЛАСТЕРАМИ В MS EXCEL**

Выполнил:

Е. С. Колосовский

Студент группы
021703

Проверил:

А. А. Ефремов

Минск 2023

1 Оценка страны производителя автомобиля на основе дискриминативного анализа

1.1 Задачи

1. Подобрать в открытых источниках data set, состоящий из результативного признака (заданного бинарной переменной) и нескольких факторных признаков (не менее 3)
2. Выполнить дискриминантный анализ наблюдений из тестовой выборки, пользуясь методическими указаниями из примера ниже.
3. В отчёте представить: постановку задачи с описанием переменных (А), фрагмент таблицы с исходными данными (Б), расстояния до кластеров (В), оценку точности анализа через расчёт процента ошибок (Г).

1.2 Суть задачи

В качестве датасета будет использоваться <https://www.kaggle.com/code/thebrownviking20/cars-k-means-clustering-script/input> Известна следующая информация по автомобилям:

1. Страна производства
2. Расход топлива (коэффициент x_1)
3. Количество цилиндров (коэффициент x_2)
4. Объем двигателя (коэффициент x_3)

Показатели потенциального x_1, x_2, x_3 производителя являются следующими - 5.1 6.4 5.4 соответственно Требуется:

1. Требуется построить множественную дискриминантную модель и с ее помощью отнести автомобиль к одному из трёх классов;
2. Построить регрессионную дискриминантную модель, найти граничное значение и отнести автомобиль к одному из классов.

Информация об автомобилях приведена в следующей таблице.

Номер п/п	Расход топлива (x_1)	Цилиндров (x_2)	Объем двигателя (x_3)	Страна производитель(Z)
1	14	8	350	US
2	31,9	4	89	Europe
3	17	8	302	US
4	15	8	400	US
5	30,5	4	98	US
6	23	8	350	US
7	13	8	351	US
8	14	8	440	US
9	25,4	5	183	Europe
10	37,7	4	89	Japan
11	34	4	108	Japan
12	34,3	4	97	Europe
13	16	8	302	US
14	11	8	350	US
15	19,1	6	225	US
16	16,9	8	350	US
17	31,8	4	85	Japan
18	16	8	304	US
19	24	4	113	Japan
20	24	4	107	Europe
21	37,2	4	86	Japan

Таблица 1 – Характеристики автомобиля.

1.3 Выполнение

1.3.1 Пункт 1

Рассчитаем центроиды для каждого класса. $\overline{x}_1^0 = 5,006$

$$\overline{x}_2^0 = 5,936$$

$$\overline{x}_3^0 = 6,588$$

$$\sigma_{11} = 0,352489687 \quad \sigma_{12} = 0,516171147 \quad \sigma_{13} = 0,635879593$$

$$\overline{x}_1^1 = 3,418$$

$$\overline{x}_2^1 = 2,77$$

$$\overline{x}_3^1 = 2,974$$

$$\sigma_{21} = 0,3810244 \quad \sigma_{22} = 0,31379832 \quad \sigma_{23} = 0,32249664$$

$$\overline{x}_1^2 = 1,464$$

$$\overline{x}_2^2 = 4,26$$

$$\overline{x}_3^2 = 5,552$$

$$\sigma_{31} = 0,17351116 \quad \sigma_{32} = 0,469910977 \quad \sigma_{33} = 0,5518947$$

Для оценки коэффициентов проверки работоспособности модели используем её для тестовых данных и сравним результаты с их классами. Оценим близость вектора заданных параметров автомобиля к центроидам каждой из трех групп по формуле::

$$D(x, \overline{x}^s) = \sqrt{\left(\frac{x_1 - \overline{x}_1^s}{\sigma_1^s}\right)^2 + \left(\frac{x_2 - \overline{x}_2^s}{\sigma_2^s}\right)^2 + \left(\frac{x_3 - \overline{x}_3^s}{\sigma_3^s}\right)^2}$$

В результате наша модель правильно предсказала страну производителя в 67,4582159% случаев. Наш потенциальный автомобиль модель относит к первому классу, что соответствует действительности.