Министерство образования и науки Российской Федерации

Севастопольский государственный университет

Кафедра ИС

Отчет

по лабораторной работе №2

«ДИСКРИМИНАНТНЫЙ АНАЛИЗ»

Выполнил:

ст. гр. ИTб-33д

Лисянский А. И.

Проверил:

Токарев А.И.

Севастополь

2015

**1.Цель работы**

Закрепить теоретические знания и приобрести практические навыки в проведении дискриминантного анализа по экспериментальным данным.

По результатам наблюдений за функционированием объектов получены экспериментальные данные. Требуется провести дискриминантный анализ этих данных.

**2.Ход работы**

**Дискриминантный анализ в Statistica**

Классический пример Фишера – анализа цветков ириса.

Задача состоит в том, чтобы по результатам измерения длины и ширины чашелистиков и лепестков цветков ириса отнести ирис к одному из трех типов: SETOSA, VERSICOL, VIRGINIC (150 цветков ириса, по 50 каждого типа.)

## Шаг 1. Нажмем кнопку Open Data (Открыть данные) и откроем файл данных. Следующий файл данных появится на экране (рисунок 1).

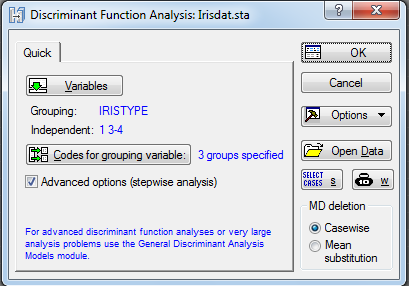
## 1_2.png

Рисунок 1 – Файл данных Irisdat.sta

## Шаг 2. Из Переключателя модулей Statistica откроем стартовую панель модуля Discriminate function analysis (Дискриминантный функциональный анализ) (рисунки 2, 3).

## 2_2.png

## Рисунок 2 – Запуск модуля Дискриминантный анализ

****

## Рисунок 3 – Стартовая панель модуля Дискриминантный анализ

## Шаг 3. Сделаем установки, как показано на рисунке 3. Для этого нажмите кнопку Variables (Переменные) и выберем переменные для анализа.

## В качестве Grouping variable (Группирующая переменная) выберем переменную Iristype (Тип ириса) (рисунок 4).

## В качестве Independent variables (Независимые переменные) выберем переменные Sepallen, Sepalwid, Petallen, Petalwid (Длина чашелистника, ширина чашелистника, длина лепестка, ширина лепестка) (рисунок 4). Нажмем кнопку ОK.

## Выполним установку Codes for grouping variable (Коды для группировки переменных) (рисунок 5). Нажмем кнопку ОK.

## 2_3.png

## Рисунок 4 – Установка переменных (Variables)

## 2_5.png

## Рисунок 5 – Установка кодов для группировки переменных

## Шаг 4. Нажмем кнопку ОK и откроем диалоговое окно Model Definition (Определение модели) (рисунок 6).

## 2_6.png

## Рисунок 6 – Окно определения модели дискриминантного анализа

## Шаг 5. Сделаем установки, как показано на рисунке 6. Нажмем кнопку OK и запустим вычислительную процедуру, реализующую пошаговый метод включения.

## Шаг 6. Просмотрим итоги в диалоговом окне Discriminant Function Analysis Results (Результаты дискриминантного анализа) (рисунок 7).

## 2_7.png

## Рисунок 7 – Окно результатов дискриминантного анализа данных

## Информационная часть окна сообщает, что использовано:

##  Stepwise analysis (Пошаговый анализ), Step 4 Final step (4 шага);

##  Number of variables in the model (Число переменных в модели): 4;

##  Last variable entered (Последняя включенная переменная): Sepallen, соответствующее значение статистики F-критерия (F(2, 144) = 4,72), уровень значимости р < 0,01;

##  Wilks lambda (Значение лямбды Уилкса): 0,023;

##  approx. F (8,288) = 199,14 (Приближенное значение F- статистики), связанной с лямбдой Уилкса;

##  р – уровень значимости F-критерия для значения 199,14;

##  значения статистики лямбда Уилкса лежат в интервале 0 – 1.

## Значения статистики Уилкса, лежащие около нуля, свидетельствуют о хорошей дискриминации. Значения статистики Уилкса, лежащие около единицы, свидетельствуют о плохой дискриминации.

## Иными словами, это можно выразить следующим образом: если значения лямбды Уилкса близки к нулю, то мощность дискриминации (мощность = 1 – вероятность ошибки) близка к 1, если лямбда Уилкса близка к единицы, то мощность близка к нулю.

## Шаг 7.Нажмем кнопку Variables in the model (Переменные, включенные в модель). На экране появится итоговая таблица анализа (рисунок 8).

## 2_8.png

## Рисунок 8 – Итоговая таблица анализа данных

## Шаг 8. Просмотрим разделение групп на графике. Для этого инициируем кнопку Perform Canonical analysis (Канонический анализ). В появившемся диалоговом окне Canonical Analysis (канонический анализ) нажмем кнопку Scatterplot of canonical scores (Диаграмма рассеяния канонических значений). На экране появится следующий график (рисунок 9).

## 2_9.png

## Рисунок 9 – Разделение трех типов ириса

## Шаг 9. Просмотрим функции классификации. В диалоговом окне Результаты дискриминантного анализа нажмем кнопку Classification functions (Функции классификации) (рисунок 10).

## 2_10.png

## Рисунок 10 – Функции классификации, построенные пошаговым Forward stepwise (методом вперед)

## С помощью этих функций можем вычислить классификационные значения (метки) для вновь наблюдаемых цветков по формулам:

## SETOSA = –16,43\*Sl+23,69\*Sw–17,4\*Pl+23,54\*Pw–86,31;

## VERSICOL = 5,21\*Sl+7,07\*Sw–6,43\*Pl+15,70\*Pw–72,85;

## VIRGINIC = 12,76\*Sl+3,69\*Sw–21,08\*Pl+12,5\*Pw–104,37,

## где:

##  Sl – Sepallen;

##  Pl – Petallen;

##  Pw – Petalwid.

## Пусть у нас есть новый цветок со значениями: Sepallen, Sepalwid, Petallen, Petalwid.

## К какому типу ириса его отнести? Формально следует подставить эти значения в приведенные выше формулы и вычислить классификационные значения SETOSA, VERSICOL, VIRGINIC.

## Новый цветок относится к тому классу, для которого классификационное значение максимально.

## Конечно, построенные классификационные функции могут быть определены в электронных таблицах как формулы, и для каждого добавленного случая по ним могут быть вычислены классификационные метки. Таким образом, каждый новый объект автоматически относится к определенному классу.

## Шаг 10. Нажмем кнопку Squared Mahalanobis distance (Квадрат расстояния Махаланобиса) и увидим таблицу с квадратами расстояния Махаланобиса от точек (случаев) до центров групп (рисунок 11).

## 2_11.png

## Рисунок 11 – Расстояния Махаланобиса для данных

## Случай относится к группе, до которой расстояние Махаланобиса минимально.

## Шаг 11. Апостериорные вероятности.

## Рассмотрим группу опций внизу диалогового окна Результаты дискриминантного анализа: A priori classifications probabilities (Априорные вероятности классификации). До анализа мы задаем для каждого объекта (в данном примере цветка) вероятность, с какой он принадлежит к определенному классу. После того как анализ выполнен, пересчитаем эти вероятности и получим апостериорные вероятности классификации. Нажав кнопку Posterior probabilities (Апостериорные вероятности), мы увидим таблицу с апостериорными вероятностями принадлежности объекта к определенному классу (рисунок 12).

## 2_12.png

## Рисунок 12 – Таблица апостериорных вероятностей

## Интерпретация данной таблицы очень проста. В первом столбце указан тип ириса для каждого случая. Во втором, третьем, четвертом столбцах даны апостериорные вероятности отнесения каждого цветка к определенному типу.

## Цветок относится к группе с максимальной апостериорной вероятностью.

## Знаком \* отмечаются неправильно классифицированные при использовании данного правила случаи (5, 9, 12). В исходной таблице (рисунок 1) необходимо внести изменения, а затем заново выполнить процедуру дискриминантного анализа.

## Шаг 12. Классификация новых случаев.

## Не закрывая диалога Результаты дискриминантного анализа, добавим в таблицу исходных данных новый случай (151 строка на рисунке 13).

## 2_13.png

## Рисунок 13 – Новое наблюдение в данных

## Шаг 13. Запустим процедуру дискриминантного анализа. Для того чтобы понять, к какому классу относится этот объект, нажмем кнопку Posterior probabilities (Апостериорные вероятности), мы увидим ту же таблицу с апостериорными вероятностями, к которой будет добавлена строка (рисунок 14).

## 2_14.png

## Рисунок 14 – Классификация нового наблюдения

## Итак, новое наблюдение с вероятностью 0,999 можно отнести к типу SETOSA.

**Выводы**

В ходе лабораторной работы Закрепили теоретические знания и приобрели практические навыки в проведении дискриминантного анализа по экспериментальным данным.

По результатам наблюдений за функционированием объектов получены экспериментальные данные. провели дискриминантный анализ этих данных по классическому примеру Фишера – анализа цветков ириса.

Задача состояла в том, чтобы по результатам измерения длины и ширины чашелистиков и лепестков цветков ириса отнесли ирис к одному из трех типов: SETOSA, VERSICOL, VIRGINIC (150 цветков ириса, по 50 каждого типа.)