МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Национальный исследовательский ядерный университет «МИФИ»

(НИЯУ МИФИ)

Институт Финансовых Технологий и Экономической Безопасности

Кафедра Финансового Мониторинга

Лабораторная работа №2:

По курсу «Макростатический анализ и прогнозирование»

Работу выполнил: студент группы С18-712: Луканов А. В.

Проверила: Домашова Д. В.

Москва 2021

# 1. Постановка задачи

Субъект РФ характеризуются следующими признаками:

X1 - ОБЩИЕ КОЭФФИЦИЕНТЫ СМЕРТНОСТИ (число умерших на 1000 человек населения);

Х2 – СРЕДНЕДУШЕВЫЕ ДЕНЕЖНЫЕ ДОХОДЫ НАСЕЛЕНИЯ (в месяц; рублей);

Х3 – ОБЩАЯ ПЛОЩАДЬ ЖИЛЫХ ПОМЕЩЕНИЙ, ПРИХОДЯЩАЯСЯ В СРЕДНЕМ НА ОДНОГО ЖИТЕЛЯ (на конец года; квадратных метров);

Х4 – Уровень занятости для людей в возрасте 15-72 лет(процентов)

Х5 – коэффициенты миграционного прироста на 10 000 человек населения;

Х6 – зарегистрировано преступлений, связанных с незаконным оборотом наркотиков (на 1000 человек);

Х7 – Зарегистрировано преступлений экономической направленности на 1000 человек;

Х8 – ЧИСЛО ПРЕДПРИЯТИЙ И ОРГАНИЗАЦИЙ (на конец года) на 1000 человек;

Х9 – РАСХОДЫ НА ОХРАНУ ОКРУЖАЮЩЕЙ СРЕДЫ(млн руб) на 1000 чел;

Х10 - ВЫБРОСЫ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ В АТМОСФЕРНЫЙ ВОЗДУХ,   
ОТХОДЯЩИХ ОТ СТАЦИОНАРНЫХ ИСТОЧНИКОВ тысяч тонн;

На основе предварительного экспертного анализа было выделено 7 групп субъектов. К первой группе отнесено 3 субъекта, ко второй – 2, к третьей – 7, к четвертой - 5, к пятой – 8, к шестой 2 и к седьмой - 2.

На основе шести обучающих выборок из многомерных нормально распределенных генеральных совокупностей с равными ковариационными матрицами провести классификацию оставшихся 56 субъектов. Дать экономическую интерпретацию результатов классификации.

# 2. Преддискриминантный анализ.

Первым делом стандартизируем исходные данные, чтобы привести их к одному масштабу.

Если исходные признаки, по которым производится классификация объектов, имеют разные единицы измерения, то необходимо перейти к стандартизованным переменным одним из следующих способов:

; ; ; ; ; ,

где  - исходное значение j-го признака на *i*-ом объекте наблюдения;

 - нормированное значение исходного *j*-го признака на *i*-ом объекте наблюдения;

 - среднее значение *j*-го признака;

 - выборочное среднеквадратическое отклонение *j*-го признака;

 - максимальное значение *j*-го признака;

 - минимальное значение *j*-го признака.

Были рассчитаны оценки общих для 6 классов ковариационной и корреляционной матриц. Матрицы представлены на Рисунках 1.1 и 1.2.

Изображение выглядит как текст, кроссворд

Автоматически созданное описание

Изображение выглядит как текст, кроссворд, стена, белый

Автоматически созданное описание

Рисунок 1.1 – Оценки общих ковариационной и корреляционной матриц (Statistica).

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 1.2 – Оценки общих ковариационной матрицы (Python)

Также были рассчитаны оценки математических ожиданий и средних квадратических отклонений признаков в классах. Результаты представлены на Рисунках 2.1, 2.2 и 3.1, 3.2 соответственно.

Изображение выглядит как текст, кроссворд, окно

Автоматически созданное описание

Рисунок 2.1 – оценки математических ожиданий признаков в классах (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2.2 – оценки математических ожиданий признаков в классах (Python)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 3 – Оценки средних квадратических отклонений признаков в классах (Statistica)

Были рассчитаны оценки ковариационной и корреляционной матриц для объединенной выборки. Матрицы представлена на Рисунке 4.

Изображение выглядит как текст, кроссворд, стена, белый

Автоматически созданное описание

Изображение выглядит как текст, кроссворд, стена

Автоматически созданное описание

Рисунок 4 – оценки ковариационной и корреляционной матриц по объединенной выборке

После этого был проведен дискриминантный анализ, результаты которого представлены на рисунке 5.

Изображение выглядит как текст, кроссворд

Автоматически созданное описание

Рисунок 5 – результаты дискриминантного анализа

В информационной части формы представлены наблюденное значение статистики Уилкса, приближенное значение F-критерия и значимость нулевой гипотезы об отсутствии различий в групповых средних значениях всех признаков.

Значение статистики Уилкса принадлежит интервалу от 0 до 1. Значение статистики близкое к нулю свидетельствует о хорошей дискриминации, значение статистики близкое к единице – о плохой дискриминации классов. Так как значение статистики Уилкса = 0,00004 близко к нулю, можно сделать вывод о хорошей дискриминации.

Несмотря на то, что различие в групповых средних значениях всех признаков значимо (p <0.05), нулевая гипотеза отвергается только для признаков X1, X2, X5. Таким образом, классы значимо различаются между собой по общему коэффициенту смертности (Х1), среднедушевым доходам населения (Х2) и миграционному приросту (Х5). В подобной ситуации целесообразно было бы обратиться к процедурам пошагового дискриминантного анализа.

Исследуется различие между классами с использованием расстояния Махаланобиса. Квадрат расстояния Махаланобиса между классами, наблюденное значение F-статистики и значимость нулевой гипотезы о равенстве нулю расстояния между классами представлены на Рисунке 6.

Изображение выглядит как текст, кроссворд, стена, внутренний

Автоматически созданное описаниеИзображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст, кроссворд

Автоматически созданное описание

Изображение выглядит как текст, кроссворд

Автоматически созданное описание

Рисунок 6 – Результаты проверки гипотезы о равенстве нулю расстояния между классами

# 3. Дискриминантный анализ

Проведем дискриминантный анализ данных по регионам России. Результаты представлены на рисунках 7 и 8.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 7 – Результаты дискриминантного анализа (страница Classification)



Изображение выглядит как текст, кроссворд

Автоматически созданное описание

Рисунок 8.1 – Коэффициенты линейных дискриминантных функций Фишера(Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 8.2 – Коэффициенты линейных дискриминантных функций Фишера(Python)

В первой строке таблицы, представленной на рисунке 16, приведены оценки априорных вероятностей, рассчитанные по первому варианту: , , , , , ,.



Линейные дискриминантные функции Фишера имеют вид:

(1)

(2)

(3)

(4)

(5)

(6)

(6)

На основе рассчитанных классификационных функций (1), (2), (3), (4), (5), (6), (7) проведена повторная классификация объектов обучающих выборок, представленная на рисунке 9.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 9 – Результаты классификации объектов обучающих выборок

Как видно из рисунка 9, изменений в первоначальном составе классов не произошло: к первому классу по-прежнему относятся 3 объекта, ко второму – 2 и т.д. Процент корректной классификации составил 100%. Это свидетельствует о хорошей дискриминации объектов обучающих выборок на основе дискриминантных функций (1), (2), (3), (4), (5), (6), (7).

Получим полную картину классификации. Результаты представлены на рисунке 10.



Рисунок 10 – Результаты классификации с помощью функций (1), (2), (3), (4), (5), (6)

Рассчитаем квадраты расстояния Махалонобиса от объектов до центров каждого из классов. Результаты представлены на рисунках 11.1 и 11.2.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 11.1 – квадраты расстояний до центров кластеров (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 11.2 – квадраты расстояний до центров кластеров (Python)

Рассчитаем апостериорные вероятности классификации. Результаты представлены на рисунках 12.1 и 12.2.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 12.1 – апостериорные вероятности классификации (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 12.2 – апостериорные вероятности классификации (Python)

На основании таблиц, представленных на рисунках 10-12, предприятия, не вошедшие в обучающие выборки, можно классифицировать следующим образом (Таблица 1):

Таблица 1

|  |  |
| --- | --- |
| **Название субъекта РФ** | **Класс** |
| Республика Бурятия | 1 |
| Республика Алтай |
| Республика Калмыкия |
| Кабардино-Балкарская  Республика |
| Карачаево-Черкесская Республика |
| Республика Дагестан |
| Чеченская Республика |
| Республика Ингушетия |
| Республика Тыва |
| г. Севастополь | 2 |
| Калининградская область |
| Московская область |
| Ленинградская область |
| Республика Адыгея |
| Республика Башкортостан | 3 |
| Томская область |
| Ростовская область |
| Пермский край |
| Краснодарский край |
| Республика Татарстан |
| Иркутская область |
| Архангельская область |
| Волгоградская область |
| Республика Карелия |
| Хабаровский край |
| Свердловская область |
| Новосибирская область |
| Республика Коми |
| Кемеровская область |
| Красноярский край |
| Республика Крым | 4 |
| Ставропольский край |
| Оренбургская область |
| Удмуртская Республика |
| Омская область |
| Челябинская область |
| Республика Хакасия |
| Астраханская область |
| Приморский край |
| Забайкальский край |
| Амурская область |
| Курганская область |
| Еврейская автономная область |
| Республика Северная  Осетия – Алания |
| Самарская область | 5 |
| Республика Марий Эл |
| Нижегородская область |
| Алтайский край |
| Чувашская Республика |
| Республика Мордовия |
| Костромская область |
| Кировская область |
| Ульяновская область |
| Воронежская область |
| Ярославская область |
| Саратовская область |
| Смоленская область |
| Калужская область |
| Ивановская область |
| Курская область |
| Владимирская область |
| Брянская область |
| Тульская область |
| Тамбовская область |
| Липецкая область |
| Вологодская область |
| Рязанская область |
| Орловская область |
| Пензенская область |
| Новгородская область |
| Белгородская область |
| Псковская область |
| Тверская область |
| Сахалинская область | 6 |
| Камчатский край |
| г. Санкт-Петербург |
| Магаданская область |
| г. Москва |
| Ненецкий автономный округ | 7 |
| Чукотский автономный округ |
| Республика Саха (Якутия) |
| Ханты-Мансийский автономный округ – Югра |
| Мурманская область |
| Тюменская область |
| Ямало-Ненецкий автономный  округ |

# 4. Пошаговый дискриминантный анализ.

**Пошаговый дискриминантный анализ с включением**

Проведем пошаговый дискриминантный анализ с включением.

Изображение выглядит как текст, кроссворд

Автоматически созданное описание

Рисунок 13 – переменные вне модели после шага 0 (Statistica)

Как видно из таблицы переменных вне модели, наибольшее значение величины F to enter принадлежит переменной Х2, соответственно, она первой будет введена в модель.

Изображение выглядит как текст, стол

Автоматически созданное описание

Рисунок 14.1 – результаты дискриминантного анализа на шаге 1 (Statistica)

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 14.2 – результаты дискриминантного анализа на шаге 1 (Python)

Внешне дискриминация между шестью группами регионов высоко значима.

Изображение выглядит как текст, кроссворд

Автоматически созданное описание

Рисунок 15 – переменные вне модели после шага 1 (Statistica)

Как видно из таблицы переменных вне модели, наибольшее значение величины F to enter принадлежит переменной Х1, соответственно, она следующей будет введена в модель.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 16.1 – результаты дискриминантного анализа на шаге 2 (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 16.2 – результаты дискриминантного анализа на шаге 2 (Python)

Внешне дискриминация между семью группами регионов высоко значима. Обе переменные Х2 и Х1 высоко значимы.

Изображение выглядит как текст, кроссворд

Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 17.1 – результаты дискриминантного анализа на шаге 5 (Statistica)

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 17.2 – результаты дискриминантного анализа на шаге 5 (Python)

Значимыми в данной модели являются переменный Х1, Х2, Х4, Х5, X6 и Х9. Точность классификация равна 100%. Рассмотрим функции классификации (Рисунки 18.1 и 18.2)

Изображение выглядит как текст, кроссворд, стена, внутренний

Автоматически созданное описание

Рисунок 18.1 - Функции классификации, построенные пошаговым методом Forward stepwise (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 18.2 - Функции классификации, построенные пошаговым методом Forward stepwise (Python)

С помощью этих функций можно вычислить классификационные значения (метки) для вновь рассматриваемых регионов по формулам:

(1)

(2)

+0,01 (3)

(4)

(5)

(6)

(7)

Получим полную картину классификации. Результаты представлены на рисунке 19.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 19 – Результаты классификации с помощью функций (1), (2), (3), (4), (5), (6)

Видим, что присутствуют примеры неправильной классификации объектов обучающих выборок. Рассчитаем квадраты расстояния Махалонобиса от объектов до центров каждого из классов. Результаты представлены на рисунках 20.1 и 20.2.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 20 – квадраты расстояний до центров кластеров (Statistica)

Рассчитаем апостериорные вероятности классификации. Результаты представлены на рисунках 21.1 и 22.2.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 21 – апостериорные вероятности классификации (Statistica)

На основании таблиц, представленных на рисунках 19-21, предприятия, не вошедшие в обучающие выборки, можно классифицировать следующим образом (Таблица 2):

Таблица 2

|  |  |
| --- | --- |
| **Название субъекта РФ** | **Класс** |
| Республика Бурятия | 1 |
| Республика Алтай |
| Республика Калмыкия |
| Кабардино-Балкарская  Республика |
| Республика Дагестан |
| Чеченская Республика |
| Республика Ингушетия |
| Республика Тыва |
| Республика Крым |
| Республика Северная  Осетия – Алания |
| г. Севастополь | 2 |
| Калининградская область |
| Ленинградская область |
| Республика Адыгея |
| Московская область | 3 |
| Республика Башкортостан |
| Томская область |
| Ростовская область |
| Пермский край |
| Краснодарский край |
| Республика Татарстан |
| Иркутская область |
| Республика Карелия |
| Новосибирская область |
| Ставропольский край |
| Нижегородская область |
| Алтайский край |
| Чувашская Республика |
| Челябинская область |
| г. Санкт-Петербург |
| Мурманская область |
| Республика Коми | 4 |
| Кемеровская область |
| Оренбургская область |
| Удмуртская Республика |
| Омская область |
| Республика Хакасия |
| Астраханская область |
| Приморский край |
| Амурская область |
| Курганская область |
| Еврейская автономная область |
| Карачаево-Черкесская Рес-публика | 5 |
| Архангельская область |
| Волгоградская область |
| Хабаровский край |
| Свердловская область |
| Самарская область |
| Республика Марий Эл |
| Республика Мордовия |
| Костромская область |
| Кировская область |
| Ульяновская область |
| Воронежская область |
| Ярославская область |
| Саратовская область |
| Смоленская область |
| Калужская область |
| Ивановская область |
| Курская область |
| Владимирская область |
| Брянская область |
| Тульская область |
| Тамбовская область |
| Липецкая область |
| Вологодская область |
| Рязанская область |
| Орловская область |
| Пензенская область |
| Новгородская область |
| Белгородская область |
| Псковская область |
| Тверская область |
| Забайкальский край | 6 |
| Сахалинская область |
| Камчатский край |
| Магаданская область |
| Чукотский автономный округ |
| г. Москва | 7 |
| Ненецкий автономный округ |
| Красноярский край |
| Республика Саха (Якутия) |
| Ханты-Мансийский автономный округ – Югра |
| Тюменская область |
| Ямало-Ненецкий автономный  округ |