|  |
| --- |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| Федеральное государственное автономное образовательное  учреждение высшего образования |
| **НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ «МИФИ»** |



***ИНСТИТУТ ФИНАНСОВЫХ ТЕХНОЛОГИЙ И ЭКОНОМИЧЕСКОЙ БЕЗОПАСНОСТИ***

КАФЕДРА «ФИНАНСОВЫЙ МОНИТОРИНГ»

ОТЧЕТ

по лабораторной работе №2

по курсу «Макростатистический анализ и прогнозирование»

Выполнил

студент группы С18-712

Степаненкова М.А.

Преподаватель: Домашова Д. В.

Москва, 2021

# Постановка задачи

Исходные данные:

* X1 - валовой региональный продукт (млн руб на 1 000 человек населения)
* X2 - инвестиции в основной капитал (в фактически действовавших ценах; млн руб на 1 000 человек населения)
* X3 - число предприятий и организаций (на конец года на 1 000 человек населения)
* X4 - уровень занятости для людей в возрасте 15-72 лет (процентов)
* X5 - коэффициенты миграционного прироста на 10 000 человек населения
* X6 - зарегистрировано преступлений особой тяжести (на 1 000 человек населения)
* X7 - зарегистрировано преступлений, связанных с незаконным оборотом наркотиков (на 1 000 человек населения)
* X8 - общие коэффициенты смертности (число умерших на 1000 человек населения)
* X9 - мощность амбулаторно-поликлинических организаций всего, тыс. посещений в смену (на конец года, на 1 000 человек населения)
* X10 - численность врачей всех специальностей, всего, тыс. человек (на конец года, на 1 000 человек населения)
* X11 - численность пенсионеров, тыс. человек (на 1 000 человек населения)

На основе предварительного анализа было выделено 6 групп регионов.

На основе 6 обучающих выборок из многомерных нормально распределенных генеральных совокупностей провести классификацию оставшихся субъектов РФ, не вошедших в обучающие выборки. Дать экономическую интерпретацию результатов классификации.

Исходные стандартизированные данные для анализа представлены в виде матрицы. Фрагмент таблицы с исходными данными в пакете Statistica представлен на рисунке 1.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 1 – Исходные данные.

# Преддискриминантный этап анализа

Была проведена оценка параметров распределения в классах:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Pooled Within-Groups Correlations (Данные для ЛР2 in Workbook4) | | | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | | |  | | --- | | X10 | | |  | | --- | | X11 | |
| |  | | --- | | X1 | | 1.00 | 0.91 | 0.38 | -0.44 | 0.14 | 0.36 | -0.03 | 0.47 | -0.17 | -0.16 | 0.27 |
| |  | | --- | | X2 | | 0.91 | 1.00 | 0.31 | -0.53 | 0.14 | 0.30 | -0.28 | 0.30 | -0.36 | -0.24 | 0.09 |
| |  | | --- | | X3 | | 0.38 | 0.31 | 1.00 | -0.36 | -0.13 | 0.18 | 0.21 | 0.39 | -0.43 | -0.25 | 0.48 |
| |  | | --- | | X4 | | -0.44 | -0.53 | -0.36 | 1.00 | -0.12 | 0.02 | 0.17 | -0.11 | 0.25 | 0.15 | -0.19 |
| |  | | --- | | X5 | | 0.14 | 0.14 | -0.13 | -0.12 | 1.00 | 0.09 | -0.01 | 0.19 | 0.12 | 0.17 | 0.26 |
| |  | | --- | | X6 | | 0.36 | 0.30 | 0.18 | 0.02 | 0.09 | 1.00 | 0.25 | 0.37 | -0.01 | 0.23 | 0.09 |
| |  | | --- | | X7 | | -0.03 | -0.28 | 0.21 | 0.17 | -0.01 | 0.25 | 1.00 | 0.31 | 0.31 | -0.02 | 0.25 |
| |  | | --- | | X8 | | 0.47 | 0.30 | 0.39 | -0.11 | 0.19 | 0.37 | 0.31 | 1.00 | -0.00 | -0.02 | 0.73 |
| |  | | --- | | X9 | | -0.17 | -0.36 | -0.43 | 0.25 | 0.12 | -0.01 | 0.31 | -0.00 | 1.00 | 0.57 | 0.24 |
| |  | | --- | | X10 | | -0.16 | -0.24 | -0.25 | 0.15 | 0.17 | 0.23 | -0.02 | -0.02 | 0.57 | 1.00 | 0.23 |
| |  | | --- | | X11 | | 0.27 | 0.09 | 0.48 | -0.19 | 0.26 | 0.09 | 0.25 | 0.73 | 0.24 | 0.23 | 1.00 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Pooled Within-Groups Covariances (Данные для ЛР2 in Workbook4) | | | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | | |  | | --- | | X10 | | |  | | --- | | X11 | |
| |  | | --- | | X1 | | 0.12 | 0.11 | 0.08 | -0.08 | 0.02 | 0.06 | -0.01 | 0.07 | -0.04 | -0.04 | 0.06 |
| |  | | --- | | X2 | | 0.11 | 0.13 | 0.06 | -0.11 | 0.02 | 0.05 | -0.05 | 0.05 | -0.09 | -0.06 | 0.02 |
| |  | | --- | | X3 | | 0.08 | 0.06 | 0.34 | -0.12 | -0.03 | 0.05 | 0.06 | 0.10 | -0.17 | -0.10 | 0.17 |
| |  | | --- | | X4 | | -0.08 | -0.11 | -0.12 | 0.32 | -0.03 | 0.00 | 0.05 | -0.03 | 0.09 | 0.06 | -0.07 |
| |  | | --- | | X5 | | 0.02 | 0.02 | -0.03 | -0.03 | 0.18 | 0.02 | -0.00 | 0.03 | 0.03 | 0.05 | 0.07 |
| |  | | --- | | X6 | | 0.06 | 0.05 | 0.05 | 0.00 | 0.02 | 0.23 | 0.06 | 0.07 | -0.00 | 0.08 | 0.03 |
| |  | | --- | | X7 | | -0.01 | -0.05 | 0.06 | 0.05 | -0.00 | 0.06 | 0.26 | 0.07 | 0.10 | -0.01 | 0.08 |
| |  | | --- | | X8 | | 0.07 | 0.05 | 0.10 | -0.03 | 0.03 | 0.07 | 0.07 | 0.17 | -0.00 | -0.01 | 0.18 |
| |  | | --- | | X9 | | -0.04 | -0.09 | -0.17 | 0.09 | 0.03 | -0.00 | 0.10 | -0.00 | 0.43 | 0.26 | 0.10 |
| |  | | --- | | X10 | | -0.04 | -0.06 | -0.10 | 0.06 | 0.05 | 0.08 | -0.01 | -0.01 | 0.26 | 0.49 | 0.10 |
| |  | | --- | | X11 | | 0.06 | 0.02 | 0.17 | -0.07 | 0.07 | 0.03 | 0.08 | 0.18 | 0.10 | 0.10 | 0.37 |

Рисунок 2 – Оценки общих ковариационной и корреляционной матриц

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2а – Оценка общей ковариационной матрицы (Python)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Класс | Means (Данные для ЛР2 in Workbook4) | | | | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | | |  | | --- | | X10 | | |  | | --- | | X11 | | |  | | --- | | Valid N | |
| |  | | --- | | G\_1:1 | | -0.042792 | -0.170935 | -0.064096 | 0.329558 | -0.130280 | 0.200048 | 0.153414 | -0.06223 | 1.13193 | 0.216123 | -0.38313 | 6 |
| |  | | --- | | G\_2:2 | | -0.432293 | -0.321964 | -0.438649 | -0.926561 | -0.441928 | -0.693247 | -0.017935 | -1.14337 | -1.37211 | -0.202547 | -1.24707 | 4 |
| |  | | --- | | G\_3:3 | | -0.324600 | -0.294838 | -0.011138 | -0.329472 | -0.362841 | -0.294388 | -0.592730 | 0.76954 | -0.23016 | -0.435987 | 0.53594 | 11 |
| |  | | --- | | G\_4:4 | | 0.657218 | 0.446944 | 0.829185 | 1.412921 | -0.361384 | -0.138277 | 0.160637 | -0.65194 | 0.51422 | 1.992296 | 0.50438 | 4 |
| |  | | --- | | G\_5:5 | | 5.746277 | 5.904978 | 1.629687 | 1.931191 | -0.208078 | 0.554598 | -0.656485 | -2.06369 | -0.42315 | 0.736588 | -0.45732 | 2 |
| |  | | --- | | G\_6:6 | | -0.218005 | -0.217277 | -0.568876 | -0.335636 | -0.407910 | 0.638563 | 0.855718 | 0.25943 | -0.27729 | -0.214067 | -0.26452 | 6 |
| |  | | --- | | All Grps | | 0.219905 | 0.204163 | 0.027308 | 0.065065 | -0.328782 | -0.012833 | -0.036588 | -0.05032 | -0.05096 | 0.116626 | -0.05685 | 33 |

Рисунок 3 – Оценки математических ожиданий признаков в классах

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 3a – Оценки математических ожиданий признаков в классах (Python)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Класс | Standard Deviations (Данные для ЛР2 in Workbook4) | | | | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | | |  | | --- | | X10 | | |  | | --- | | X11 | | |  | | --- | | Valid N | |
| |  | | --- | | G\_1:1 | | 0.154145 | 0.113321 | 0.324404 | 0.359014 | 0.385652 | 0.485479 | 0.433934 | 0.372111 | 0.344698 | 0.646563 | 0.308817 | 6 |
| |  | | --- | | G\_2:2 | | 0.060284 | 0.061064 | 0.487585 | 0.820928 | 0.684936 | 0.435800 | 0.223035 | 0.283492 | 0.097828 | 0.382913 | 0.626583 | 4 |
| |  | | --- | | G\_3:3 | | 0.083902 | 0.106248 | 0.364423 | 0.318571 | 0.266184 | 0.470329 | 0.383767 | 0.328768 | 0.408982 | 0.510261 | 0.398636 | 11 |
| |  | | --- | | G\_4:4 | | 0.677439 | 0.651850 | 0.884837 | 0.341629 | 0.592343 | 0.587195 | 1.000469 | 0.696400 | 1.494576 | 1.525411 | 1.193684 | 4 |
| |  | | --- | | G\_5:5 | | 1.249714 | 1.411067 | 1.906338 | 1.685088 | 0.541051 | 0.663693 | 0.252486 | 0.960343 | 1.442918 | 0.411828 | 1.162595 | 2 |
| |  | | --- | | G\_6:6 | | 0.085843 | 0.048251 | 0.323902 | 0.577187 | 0.363092 | 0.369523 | 0.533475 | 0.242402 | 0.359476 | 0.431320 | 0.447181 | 6 |
| |  | | --- | | All Grps | | 1.495300 | 1.525371 | 0.787593 | 0.964962 | 0.404485 | 0.621653 | 0.715338 | 0.918387 | 0.957167 | 1.008366 | 0.826891 | 33 |

Рисунок 4 – Оценки средних квадратических отклонений признаков в классах

По данным рисунка 3 можно дать интерпретацию классам.

Регионы первого класса характеризуются серединными значениями среди величин всех признаков.

Регионы второго класса отличаются максимальным значением коэффициента миграционного прироста (Х5) и близким к минимальному значению общего коэффициента смертности (Х8). При этом остальные средние значения признаков принимают минимальное значение по сравнению с другими классами.

Регионы третьего класса характеризуются близким к минимальным значениям валового регионального продукта (Х1), инвестиций в основной капитал (Х2), количества преступлений, связанных с незаконным оборотом наркотиков (Х7), численности врачей всех специальностей (Х10). При этом значения общего коэффициента смертности (Х8) и численности пенсионеров (Х11) являются максимальными. Остальные средние значения признаков у объектов кластера принимают промежуточные (средние) значения.

Регионы четвертого класса характеризуются максимальным значением уровня занятости (Х4) и численности врачей всех специальностей (Х10). Численность пенсионеров (Х11) и количество крупных и средних организаций агропромышленного комплекса (Х3) близки к максимальным.

Регионы пятого класса характеризуются максимальными значениями валового регионального продукта (Х1), инвестиций в основной капитал (Х2), количества крупных и средних организаций агропромышленного комплекса (Х3), уровня занятости (Х4). Значения количества преступлений, связанных с незаконным оборотом наркотиков (Х7), общего коэффициента смертности (Х8), численности пенсионеров (Х11) минимальны.

Регионы шестого класса характеризуются минимальным значением количества крупных и средних организаций агропромышленного комплекса (Х3), коэффициента миграционного прироста (Х5). Близки к минимальным значения валового регионального продукта (Х1), инвестиций в основной капитал (Х2), уровня занятости (Х4), численности пенсионеров (Х11). Значения количества тяжких и особо тяжких преступлений (Х6) и количества преступлений, связанных с незаконным оборотом наркотиков (Х7) максимальны.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 5 – Результаты дискриминантного анализa

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=33 | Discriminant Function Analysis Summary (Данные для ЛР2 in Workbook4) No. of vars in model: 11; Grouping: Класс (6 grps) Wilks' Lambda: .00004 approx. F (55,82)=11.820 p<0.0000 | | | | | |
| |  | | --- | | Wilks' Lambda | | |  | | --- | | Partial Lambda | | |  | | --- | | F-remove (5,17) | | |  | | --- | | p-value | | |  | | --- | | Toler. | | |  | | --- | | 1-Toler. (R-Sqr.) | |
| |  | | --- | | X1 | | 0.000044 | 0.910699 | 0.333398 | 0.885732 | 0.088582 | 0.911418 |
| |  | | --- | | X2 | | 0.000050 | 0.804980 | 0.823709 | 0.549819 | 0.074931 | 0.925069 |
| |  | | --- | | X3 | | 0.000104 | 0.387252 | 5.379821 | 0.003809 | 0.248907 | 0.751093 |
| |  | | --- | | X4 | | 0.000090 | 0.446330 | 4.217683 | 0.011228 | 0.587372 | 0.412628 |
| |  | | --- | | X5 | | 0.000050 | 0.805027 | 0.823459 | 0.549975 | 0.739298 | 0.260702 |
| |  | | --- | | X6 | | 0.000049 | 0.822753 | 0.732470 | 0.609012 | 0.492786 | 0.507214 |
| |  | | --- | | X7 | | 0.000088 | 0.457099 | 4.038220 | 0.013423 | 0.417272 | 0.582728 |
| |  | | --- | | X8 | | 0.000077 | 0.520138 | 3.136723 | 0.034669 | 0.223797 | 0.776203 |
| |  | | --- | | X9 | | 0.000150 | 0.269025 | 9.238235 | 0.000216 | 0.264299 | 0.735701 |
| |  | | --- | | X10 | | 0.000055 | 0.729022 | 1.263785 | 0.324178 | 0.439571 | 0.560429 |
| |  | | --- | | X11 | | 0.000069 | 0.585434 | 2.407654 | 0.079781 | 0.156035 | 0.843965 |

Рисунок 6 – Результаты проверки гипотезы об отсутствии различий в групповых средних значениях по каждому признаку

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.

Различие в групповых средних значениях всех признаков значимо (значимость нулевой гипотезы p<0,0000), нулевая гипотеза принимается для признаков Х1 (значимость нулевой гипотезы p=0,886 > 0,05), Х2 (значимость нулевой гипотезы p=0,550 > 0,05), Х5 (значимость нулевой гипотезы p=0,550 > 0,05), Х6 (значимость нулевой гипотезы p=0,609 > 0,05), Х10 (значимость нулевой гипотезы p=0,324 > 0,05), Х11 (значимость нулевой гипотезы p=0,080 > 0,05). Таким образом, классы значимо различаются между собой по количеству крупных и средних организаций агропромышленного комплекса (Х3), уровню занятости для людей в возрасте 15 лет и старше (Х4), количеству преступлений, связанных с незаконным оборотом наркотиков (Х7), общему коэффициенту смертности (Х8), числу зарегистрированных заболеваний у пациентов с диагнозом, установленным впервые в жизни (Х9).

Квадрат расстояния Махаланобиса между классами, наблюденное значение *F*-статистики и значимость нулевой гипотезы о равенстве нулю расстояния между классами представлены на рисунке 7.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Класс | Squared Mahalanobis Distances (Данные для ЛР2 in Workbook4) | | | | | |
| |  | | --- | | G\_1:1 | | |  | | --- | | G\_2:2 | | |  | | --- | | G\_3:3 | | |  | | --- | | G\_4:4 | | |  | | --- | | G\_5:5 | | |  | | --- | | G\_6:6 | |
| |  | | --- | | G\_1:1 | | 0.000 | 70.1903 | 26.972 | 86.7660 | 1012.136 | 44.599 |
| |  | | --- | | G\_2:2 | | 70.190 | 0.0000 | 61.688 | 95.6714 | 989.614 | 25.752 |
| |  | | --- | | G\_3:3 | | 26.972 | 61.6879 | 0.000 | 132.9389 | 1177.621 | 31.635 |
| |  | | --- | | G\_4:4 | | 86.766 | 95.6714 | 132.939 | 0.0000 | 595.223 | 112.848 |
| |  | | --- | | G\_5:5 | | 1012.136 | 989.6135 | 1177.621 | 595.2230 | 0.000 | 1057.730 |
| |  | | --- | | G\_6:6 | | 44.599 | 25.7516 | 31.635 | 112.8478 | 1057.730 | 0.000 |

Рисунок 7a - Квадрат расстояния Махаланобиса между классами

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Класс | F-values; df = 11,17 (Данные для ЛР2 in Workbook4) | | | | | |
| |  | | --- | | G\_1:1 | | |  | | --- | | G\_2:2 | | |  | | --- | | G\_3:3 | | |  | | --- | | G\_4:4 | | |  | | --- | | G\_5:5 | | |  | | --- | | G\_6:6 | |
| |  | | --- | | G\_1:1 | |  | 9.64230 | 5.9938 | 11.91937 | 86.9005 | 7.65845 |
| |  | | --- | | G\_2:2 | | 9.64230 |  | 10.3575 | 10.95229 | 75.5261 | 3.53759 |
| |  | | --- | | G\_3:3 | | 5.99382 | 10.35748 |  | 22.32060 | 114.0716 | 7.03009 |
| |  | | --- | | G\_4:4 | | 11.91937 | 10.95229 | 22.3206 |  | 45.4267 | 15.50232 |
| |  | | --- | | G\_5:5 | | 86.90054 | 75.52605 | 114.0716 | 45.42667 |  | 90.81522 |
| |  | | --- | | G\_6:6 | | 7.65845 | 3.53759 | 7.0301 | 15.50232 | 90.8152 |  |

Рисунок 7в - Наблюденное значение *F*-статистики

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Класс | p-values (Данные для ЛР2 in Workbook4) | | | | | |
| |  | | --- | | G\_1:1 | | |  | | --- | | G\_2:2 | | |  | | --- | | G\_3:3 | | |  | | --- | | G\_4:4 | | |  | | --- | | G\_5:5 | | |  | | --- | | G\_6:6 | |
| |  | | --- | | G\_1:1 | |  | 0.000027 | 0.000568 | 0.000006 | 0.000000 | 0.000124 |
| |  | | --- | | G\_2:2 | | 0.000027 |  | 0.000017 | 0.000011 | 0.000000 | 0.009746 |
| |  | | --- | | G\_3:3 | | 0.000568 | 0.000017 |  | 0.000000 | 0.000000 | 0.000214 |
| |  | | --- | | G\_4:4 | | 0.000006 | 0.000011 | 0.000000 |  | 0.000000 | 0.000001 |
| |  | | --- | | G\_5:5 | | 0.000000 | 0.000000 | 0.000000 | 0.000000 |  | 0.000000 |
| |  | | --- | | G\_6:6 | | 0.000124 | 0.009746 | 0.000214 | 0.000001 | 0.000000 |  |

Рисунок 7г – Результаты проверки гипотезы о равенстве нулю расстояния между классами

# Дискриминантный анализ

Рассчитываются коэффициенты линейных дискриминантных функций Фишера. Результаты представлены на рисунке 8.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 5 – Результаты дискриминантного анализа

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Classification Functions; grouping: Класс (Данные для ЛР2 in Workbook4) | | | | | |
| |  | | --- | | G\_1:1 p=.18182 | | |  | | --- | | G\_2:2 p=.12121 | | |  | | --- | | G\_3:3 p=.33333 | | |  | | --- | | G\_4:4 p=.12121 | | |  | | --- | | G\_5:5 p=.06061 | | |  | | --- | | G\_6:6 p=.18182 | |
| |  | | --- | | X1 | | -1.2757 | 14.8142 | -3.3029 | 11.6573 | 29.157 | -2.8655 |
| |  | | --- | | X2 | | -2.8964 | -14.3181 | -9.1579 | 13.7503 | 98.920 | -0.5660 |
| |  | | --- | | X3 | | 8.9320 | -10.0504 | 0.8682 | 1.5913 | -11.885 | -9.9311 |
| |  | | --- | | X4 | | 0.0172 | -5.8070 | -3.7998 | 11.8517 | 41.111 | -4.6170 |
| |  | | --- | | X5 | | 2.3078 | -3.7474 | -2.0660 | -3.1680 | -10.864 | -5.2040 |
| |  | | --- | | X6 | | 0.1472 | -1.4875 | -0.1682 | -6.1200 | -14.225 | 4.0801 |
| |  | | --- | | X7 | | -4.8980 | 7.2150 | -6.8573 | 6.5482 | 29.881 | 6.5328 |
| |  | | --- | | X8 | | 8.4152 | -11.2468 | 14.4626 | -19.1717 | -81.972 | -0.4822 |
| |  | | --- | | X9 | | 9.4008 | -12.0927 | 2.1817 | -4.7799 | -13.695 | -5.9336 |
| |  | | --- | | X10 | | -1.2724 | 3.4151 | -1.6055 | 7.4514 | 10.246 | -0.0058 |
| |  | | --- | | X11 | | -10.5260 | 5.7728 | -4.1923 | 8.8225 | 40.760 | 4.5109 |
| |  | | --- | | Constant | | -8.1225 | -18.1672 | -10.5780 | -34.2358 | -477.959 | -11.0015 |

Рисунок 8 – Коэффициенты линейных дискриминантных функций Фишера

В первой строке таблицы, представленной на рисунке 8, приведены оценки априорных вероятностей, рассчитанные по первому варианту:

, , , , , .

Линейные дискриминантные функции Фишера имеют вид:

(1)

(2)

(3)

(4)

(5)

(6)

На основе рассчитанных классификационных функций (1)-(6) проводится повторная классификация объектов обучающих выборок.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | Classification Matrix (Данные для ЛР2 in Workbook4) Rows: Observed classifications Columns: Predicted classifications | | | | | | |
| |  | | --- | | Percent Correct | | |  | | --- | | G\_1:1 p=.18182 | | |  | | --- | | G\_2:2 p=.12121 | | |  | | --- | | G\_3:3 p=.33333 | | |  | | --- | | G\_4:4 p=.12121 | | |  | | --- | | G\_5:5 p=.06061 | | |  | | --- | | G\_6:6 p=.18182 | |
| |  | | --- | | G\_1:1 | | 100.0000 | 6 | 0 | 0 | 0 | 0 | 0 |
| |  | | --- | | G\_2:2 | | 100.0000 | 0 | 4 | 0 | 0 | 0 | 0 |
| |  | | --- | | G\_3:3 | | 100.0000 | 0 | 0 | 11 | 0 | 0 | 0 |
| |  | | --- | | G\_4:4 | | 100.0000 | 0 | 0 | 0 | 4 | 0 | 0 |
| |  | | --- | | G\_5:5 | | 100.0000 | 0 | 0 | 0 | 0 | 2 | 0 |
| |  | | --- | | G\_6:6 | | 100.0000 | 0 | 0 | 0 | 0 | 0 | 6 |
| |  | | --- | | Total | | 100.0000 | 6 | 4 | 11 | 4 | 2 | 6 |

Рисунок 9 – Результаты классификации объектов обучающих выборок

Как видно из рисунка 9, изменений в первоначальном составе классов не произошло. Процент корректной классификации составил 100%. Это свидетельствует о хорошей дискриминации объектов обучающих выборок на основе дискриминантных функций (1)-(6).

Результаты классификации представлены на рисунке 10. Неправильные классификации объектов обучающих выборок помечаются «звездочкой» (в рассматриваемом примере такие ситуации не встречаются).

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 10 – Результаты классификации с помощью функций (1)-(6)

Рассчитываются квадраты расстояния Махаланобиса от объектов до центров каждого из классов. Результаты расчетов представлены на рисунке 11. Объект следует отнести к тому классу, расстояние до которого наименьшее.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 11 – Расстояния до центров классов

Результаты расчетов апостериорных вероятностей представлены на рисунке 12. Объект следует отнести к тому классу, апостериорная вероятность для которого наибольшая.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 12 – Апостериорные вероятности классификации

На основании таблиц, представленных на рисунках 10-12, субъекты РФ, не вошедшие в обучающие выборки, можно классифицировать следующим образом:

|  |  |  |
| --- | --- | --- |
| Номер  класса | Кол-во  объектов в классе (не обучающих) | Состав класса |
| 1 | 15 | |  | | --- | | Белгородская область | | Город федерального значения Севастополь | | Калининградская область | | Калужская область | | Краснодарский край | | Ленинградская область | | Липецкая область | | Московская область | | Нижегородская область | | Новгородская область | | Новосибирская область | | Омская область | | Приморский край | | Республика Адыгея (Адыгея) | | Тюменская область (кроме Ханты-Мансийского автономного округа-Югры и Ямало-Ненецкого автономного округа) | |
| 2 | 7 | |  | | --- | | Астраханская область | | Республика Дагестан | | Республика Ингушетия | | Республика Северная Осетия-Алания | | Республика Тыва | | Ставропольский край | | Чеченская Республика | |
| 3 | 17 | |  | | --- | | Владимирская область | | Воронежская область | | Костромская область | | Курская область | | Орловская область | | Псковская область | | Республика Башкортостан | | Республика Крым | | Республика Марий Эл | | Республика Мордовия | | Ростовская область | | Рязанская область | | Самарская область | | Смоленская область | | Тульская область | | Ульяновская область | | Чувашская Республика - Чувашия | |
| 4 | 5 | |  | | --- | | Город Москва столица Российской Федерации город федерального значения | | Магаданская область | | Мурманская область | | Ханты-Мансийский автономный округ - Югра (Тюменская область) | | Чукотский автономный округ | |
| 6 | 8 | |  | | --- | | Еврейская автономная область | | Иркутская область | | Курганская область | | Республика Бурятия | | Республика Карелия | | Республика Коми | |

Проведем дискриминантный анализ данных по регионам России с использованием языка программирования Python.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 13 - коэффициенты линейных дискриминантных функций Фишера – Python



Рисунок 14 – оценки априорных вероятностей – Python

На рисунке 14 приведены оценки априорных вероятностей:

, , , , , .

Линейные дискриминантные функции Фишера имеют вид:

Регионы России, не вошедшие в обучающие выборки, можно классифицировать следующим образом:

|  |  |
| --- | --- |
| **Субъект Российской Федерации** | **Метод Уорда (Python)** |
| Амурская область | 2 |
| Архангельская область (кроме Ненецкого автономного округа) | 4 |
| Астраханская область | 2 |
| Белгородская область | 4 |
| Владимирская область | 4 |
| Воронежская область | 4 |
| Город Москва столица Российской Федерации город федерального значения | 2 |
| Город федерального значения Севастополь | 1 |
| Еврейская автономная область | 5 |
| Иркутская область | 5 |
| Калининградская область | 2 |
| Калужская область | 2 |
| Костромская область | 4 |
| Краснодарский край | 4 |
| Курганская область | 5 |
| Курская область | 4 |
| Ленинградская область | 1 |
| Липецкая область | 4 |
| Магаданская область | 2 |
| Московская область | 2 |
| Мурманская область | 2 |
| Нижегородская область | 2 |
| Новгородская область | 2 |
| Новосибирская область | 2 |
| Омская область | 2 |
| Орловская область | 4 |
| Приморский край | 5 |
| Псковская область | 5 |
| Республика Адыгея (Адыгея) | 1 |
| Республика Башкортостан | 4 |
| Республика Бурятия | 5 |
| Республика Дагестан | 1 |
| Республика Ингушетия | 1 |
| Республика Карелия | 4 |
| Республика Коми | 2 |
| Республика Крым | 4 |
| Республика Марий Эл | 4 |
| Республика Мордовия | 4 |
| Республика Северная Осетия-Алания | 5 |
| Республика Тыва | 5 |
| Ростовская область | 4 |
| Рязанская область | 4 |
| Самарская область | 4 |
| Смоленская область | 4 |
| Ставропольский край | 5 |
| Тульская область | 4 |
| Тюменская область (кроме Ханты-Мансийского автономного округа-Югры и Ямало-Ненецкого автономного округа) | 2 |
| Ульяновская область | 4 |
| Ханты-Мансийский автономный округ - Югра (Тюменская область) | 2 |
| Чеченская Республика | 1 |
| Чувашская Республика - Чувашия | 4 |
| Чукотский автономный округ | 6 |

# Пошаговый дискриминантный анализ

При пошаговом включении программа будет вводить переменные в модель постепенно, одну за другой, каждый раз выбирая переменную, вносящую наибольший вклад в дискриминацию.

На каждом шаге можно посмотреть переменные, которые не включены/включены в модель:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=33 | Variables currently not in the model (Данные для ЛР2 in Workbook4) Df for all F-tests: 5,27 | | | | | |
| |  | | --- | | Wilks' Lambda | | |  | | --- | | Partial Lambda | | |  | | --- | | F to enter | | |  | | --- | | p-value | | |  | | --- | | Toler. | | |  | | --- | | 1-Toler. (R-Sqr.) | |
| |  | | --- | | X1 | | 0.044382 | 0.044382 | 116.2712 | 0.000000 | 1.000000 | 0.00 |
| |  | | --- | | X2 | | 0.046548 | 0.046548 | 110.6103 | 0.000000 | 1.000000 | 0.00 |
| |  | | --- | | X3 | | 0.457184 | 0.457184 | 6.4114 | 0.000478 | 1.000000 | 0.00 |
| |  | | --- | | X4 | | 0.286490 | 0.286490 | 13.4489 | 0.000001 | 1.000000 | 0.00 |
| |  | | --- | | X5 | | 0.929071 | 0.929071 | 0.4123 | 0.836019 | 1.000000 | 0.00 |
| |  | | --- | | X6 | | 0.494718 | 0.494718 | 5.5153 | 0.001264 | 1.000000 | 0.00 |
| |  | | --- | | X7 | | 0.430729 | 0.430729 | 7.1369 | 0.000228 | 1.000000 | 0.00 |
| |  | | --- | | X8 | | 0.173594 | 0.173594 | 25.7070 | 0.000000 | 1.000000 | 0.00 |
| |  | | --- | | X9 | | 0.399928 | 0.399928 | 8.1024 | 0.000090 | 1.000000 | 0.00 |
| |  | | --- | | X10 | | 0.406119 | 0.406119 | 7.8966 | 0.000109 | 1.000000 | 0.00 |
| |  | | --- | | X11 | | 0.451093 | 0.451093 | 6.5709 | 0.000405 | 1.000000 | 0.00 |

Рисунок 13 - Переменные вне модели (при пошаговом включении на шаге 0)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=33 | Discriminant Function Analysis Summary (Данные для ЛР2 in Workbook4) Step 7, N of vars in model: 7; Grouping: Класс (6 grps) Wilks' Lambda: .00011 approx. F (35,90)=20.171 p<0.0000 | | | | | |
| |  | | --- | | Wilks' Lambda | | |  | | --- | | Partial Lambda | | |  | | --- | | F-remove (5,21) | | |  | | --- | | p-value | | |  | | --- | | Toler. | | |  | | --- | | 1-Toler. (R-Sqr.) | |
| |  | | --- | | X1 | | 0.001623 | 0.066239 | 59.20686 | 0.000000 | 0.536722 | 0.463278 |
| |  | | --- | | X8 | | 0.000572 | 0.188091 | 18.12966 | 0.000001 | 0.272754 | 0.727246 |
| |  | | --- | | X4 | | 0.000239 | 0.450051 | 5.13228 | 0.003135 | 0.675495 | 0.324505 |
| |  | | --- | | X7 | | 0.000373 | 0.288399 | 10.36316 | 0.000041 | 0.587352 | 0.412648 |
| |  | | --- | | X11 | | 0.000277 | 0.388890 | 6.59996 | 0.000774 | 0.217507 | 0.782494 |
| |  | | --- | | X9 | | 0.000438 | 0.245570 | 12.90310 | 0.000008 | 0.350094 | 0.649906 |
| |  | | --- | | X3 | | 0.000257 | 0.417728 | 5.85439 | 0.001542 | 0.310086 | 0.689914 |

Рисунок 14 - Результаты дискриминантного анализа (Переменные в модели)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Classification Functions; grouping: Класс (Данные для ЛР2 in Workbook4) | | | | | |
| |  | | --- | | G\_1:1 p=.18182 | | |  | | --- | | G\_2:2 p=.12121 | | |  | | --- | | G\_3:3 p=.33333 | | |  | | --- | | G\_4:4 p=.12121 | | |  | | --- | | G\_5:5 p=.06061 | | |  | | --- | | G\_6:6 p=.18182 | |
| |  | | --- | | X1 | | -2.99971 | -0.0670 | -11.9132 | 19.8956 | 108.368 | -3.38230 |
| |  | | --- | | X8 | | 8.16727 | -12.1014 | 15.0227 | -23.3127 | -88.544 | 3.16035 |
| |  | | --- | | X4 | | -0.01003 | -4.3150 | -3.0897 | 11.0877 | 33.952 | -3.86881 |
| |  | | --- | | X7 | | -3.84828 | 7.1493 | -5.5837 | 1.5815 | 12.393 | 6.33862 |
| |  | | --- | | X11 | | -9.69595 | 6.1887 | -5.2706 | 11.4442 | 39.923 | 0.26106 |
| |  | | --- | | X9 | | 8.41259 | -8.3496 | 2.4239 | -1.7198 | -15.274 | -4.19581 |
| |  | | --- | | X3 | | 7.90477 | -7.9481 | 2.1955 | 1.6098 | -12.378 | -6.58759 |
| |  | | --- | | Constant | | -7.58328 | -14.5905 | -9.2727 | -27.3188 | -418.257 | -8.26562 |

Рисунок 15 – Функции классификации, построенные пошаговым методом Forward stepwise

В первой строке таблицы, представленной на рисунке 15, приведены оценки априорных вероятностей, рассчитанные по первому варианту: , , , , , . Линейные дискриминантные функции Фишера имеют вид:

(1’)

(2’)

(3’)

(4’)

(5’)

(6’)

На основе рассчитанных классификационных функций (1`)-(6`) проводится повторная классификация объектов обучающих выборок.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | Classification Matrix (Данные для ЛР2 in Workbook4) Rows: Observed classifications Columns: Predicted classifications | | | | | | |
| |  | | --- | | Percent Correct | | |  | | --- | | G\_1:1 p=.18182 | | |  | | --- | | G\_2:2 p=.12121 | | |  | | --- | | G\_3:3 p=.33333 | | |  | | --- | | G\_4:4 p=.12121 | | |  | | --- | | G\_5:5 p=.06061 | | |  | | --- | | G\_6:6 p=.18182 | |
| |  | | --- | | G\_1:1 | | 100.0000 | 6 | 0 | 0 | 0 | 0 | 0 |
| |  | | --- | | G\_2:2 | | 100.0000 | 0 | 4 | 0 | 0 | 0 | 0 |
| |  | | --- | | G\_3:3 | | 100.0000 | 0 | 0 | 11 | 0 | 0 | 0 |
| |  | | --- | | G\_4:4 | | 100.0000 | 0 | 0 | 0 | 4 | 0 | 0 |
| |  | | --- | | G\_5:5 | | 100.0000 | 0 | 0 | 0 | 0 | 2 | 0 |
| |  | | --- | | G\_6:6 | | 100.0000 | 0 | 0 | 0 | 0 | 0 | 6 |
| |  | | --- | | Total | | 100.0000 | 6 | 4 | 11 | 4 | 2 | 6 |

Рисунок 16 – Результаты классификации объектов обучающих выборок

Как видно из рисунка 16, изменений в первоначальном составе классов не произошло. Процент корректной классификации составил 100%. Это свидетельствует о хорошей дискриминации объектов обучающих выборок на основе дискриминантных функций (1`)-(6`).

Результаты классификации представлены на рисунке 16. Неправильные классификации объектов обучающих выборок помечаются «звездочкой» (в рассматриваемом примере такие ситуации не встречаются).

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 16 – Результаты классификации с помощью функций (1’)-(6’)

Рассчитываются квадраты расстояния Махаланобиса от объектов до центров каждого из классов. Результаты расчетов представлены на рисунке 17. Объект следует отнести к тому классу, расстояние до которого наименьшее.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 17 – Расстояния до центров классов

Результаты расчетов апостериорных вероятностей представлены на рисунке 18. Объект следует отнести к тому классу, апостериорная вероятность для которого наибольшая.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 18 – Апостериорные вероятности классификации

На основании рисунков 16-18, субъекты РФ, не вошедшие в обучающие выборки, можно классифицировать следующим образом:

|  |  |  |
| --- | --- | --- |
| Номер  класса | Кол-во  объектов в классе (не обучающих) | Состав класса |
| 1 | 13 | |  | | --- | | Белгородская область | | Калининградская область | | Калужская область | | Краснодарский край | | Ленинградская область | | Липецкая область | | Московская область | | Нижегородская область | | Новгородская область | | Новосибирская область | | Омская область | | Приморский край | | Тюменская область (кроме Ханты-Мансийского автономного округа-Югры и Ямало-Ненецкого автономного округа) | |
| 2 | 6 | |  | | --- | | Республика Дагестан | | Республика Ингушетия | | Республика Коми | | Республика Северная Осетия-Алания | | Республика Тыва | | Чеченская Республика | |
| 3 | 20 | |  | | --- | | Город федерального значения Севастополь | | Иркутская область | | Костромская область | | Курская область | | Орловская область | | Псковская область | | Республика Адыгея (Адыгея) | | Республика Башкортостан | | Республика Крым | | Республика Марий Эл | | Республика Мордовия | | Ростовская область | | Рязанская область | | Смоленская область | | Тульская область | | Ульяновская область | | Чувашская Республика - Чувашия | |
| 4 | 5 | |  | | --- | | Магаданская область | | Мурманская область | | Ханты-Мансийский автономный округ - Югра (Тюменская область) | | Чукотский автономный округ | |
| 6 | 8 | |  | | --- | | Еврейская автономная область | | Курганская область | | Республика Бурятия | | Республика Карелия | | Самарская область | | Ставропольский край | |

Проведем пошаговый дискриминантный анализ данных по регионам России с использованием языка программирования Python.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 8 - Результаты пошагового дискриминантного анализа на 1 шаге

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 9 - Результаты пошагового дискриминантного анализа на 2 шаге

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 10 - Результаты пошагового дискриминантного анализа на последнем (7) шаге

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 11 - коэффициенты линейных дискриминантных функций Фишера и оценки априорных вероятностей

Оценки априорных вероятностей:

, , , , , .

(1’)

(2’)

(3’)

(4’)

(5’)

(6’)

Регионы России, не вошедшие в обучающие выборки, можно классифицировать следующим образом:

|  |  |
| --- | --- |
| **Субъект Российской Федерации** | **Номер класса (пошаговый дискр анализ с включением) Python** |
| Амурская область | 6 |
| Архангельская область (кроме Ненецкого автономного округа) | 3 |
| Астраханская область | 6 |
| Белгородская область | 1 |
| Владимирская область | 3 |
| Воронежская область | 3 |
| Город Москва столица Российской Федерации город федерального значения | 4 |
| Город федерального значения Севастополь | 3 |
| Еврейская автономная область | 6 |
| Иркутская область | 3 |
| Калининградская область | 1 |
| Калужская область | 1 |
| Костромская область | 3 |
| Краснодарский край | 1 |
| Курганская область | 6 |
| Курская область | 3 |
| Ленинградская область | 1 |
| Липецкая область | 1 |
| Магаданская область | 4 |
| Московская область | 1 |
| Мурманская область | 4 |
| Нижегородская область | 1 |
| Новгородская область | 1 |
| Новосибирская область | 1 |
| Омская область | 1 |
| Орловская область | 3 |
| Приморский край | 1 |
| Псковская область | 3 |
| Республика Адыгея (Адыгея) | 3 |
| Республика Башкортостан | 3 |
| Республика Бурятия | 6 |
| Республика Дагестан | 2 |
| Республика Ингушетия | 2 |
| Республика Карелия | 6 |
| Республика Коми | 2 |
| Республика Крым | 3 |
| Республика Марий Эл | 3 |
| Республика Мордовия | 3 |
| Республика Северная Осетия-Алания | 2 |
| Республика Тыва | 2 |
| Ростовская область | 3 |
| Рязанская область | 3 |
| Самарская область | 6 |
| Смоленская область | 3 |
| Ставропольский край | 6 |
| Тульская область | 3 |
| Тюменская область (кроме Ханты-Мансийского автономного округа-Югры и Ямало-Ненецкого автономного округа) | 1 |
| Ульяновская область | 3 |
| Ханты-Мансийский автономный округ - Югра (Тюменская область) | 4 |
| Чеченская Республика | 2 |
| Чувашская Республика - Чувашия | 3 |
| Чукотский автономный округ | 4 |