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»**

|  |  |
| --- | --- |
| **Выполнил студент группы С21-703:** | Монастырский М.О. |
| **Проверил:** | Домашова Д.В. |

Оглавление

[1. Постановка задачи 3](#_Toc180348315)

[2. Выполнение работы 4](#_Toc180348316)

[3. Преддискриминантный этап анализа 4](#_Toc180348317)

[4. Дискриминантный анализ 8](#_Toc180348318)

[5. Пошаговый анализ. Включение 11](#_Toc180348319)

[6. Пошаговый анализ. Исключение 13](#_Toc180348320)

[Приложение А «Результаты дискриминантного анадиза» 16](#_Toc180348321)

[Приложение Б «Результаты анализа с включением» 23](#_Toc180348322)

[Приложение В «Результаты анализа с исключением» 29](#_Toc180348323)

1. Постановка задачи

1. Выбор предмета исследования. В данной лабораторной работе предметом исследования является состояние субъектов РФ, которое характеризуется следующими показателями:

|  |  |
| --- | --- |
|  | Число дорожно-транспортных происшествий и пострадавших в них на 100 000 человек населения |
|  | Смертность населения старше трудоспособного возраста, на 100 000 человек населения соответствующего возраста |
|  | Продажа сильно алкогольной продукции населению(тысяч декалитров)/**на тыс населения** |
|  | Средняя Стоимость минимального (условного) набора потребительских товаров и услуг |
|  | Число спортивных сооружений/ **на тыс населения** |
|  | Доходы консолидированных бюджетов субъектов Российской Федерации / **на тыс населения** |
|  | Предварительно расследовано преступлений, совершенных в состоянии алкогольного опьянения/ **на тыс населения** |
|  | Среднедушевые доходы населения (в месяц), руб. |
|  | Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры на 10 000 человек населения, всего |

1. На основе предварительного экспертного анализа было выделено семь групп субъектов РФ. К первой группе отнесено 2 субъекта, ко второй – 2, к третьей – 2, к четвертой – 8, к пятой – 2, к шестой – 9, к седьмой – 5
2. На основе семи обучающих выборок из многомерных нормально распределенных генеральных совокупностей с равными ковариационными матрицами необходимо провести классификацию оставшихся 55 субъектов РФ;
3. Дать экономическую интерпретацию результатов классификации

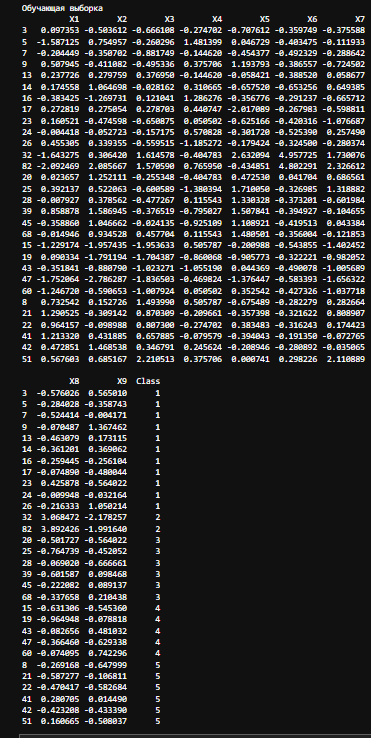
2. Выполнение работы

Фрагмент таблицы с исходным и данными для анализа в пакете Statistica представлен на рис. 1. В первых девяти столбцах введены наблюденные значения признаков, характеризующих демографическое состояние субъектов РФ, в десятом столбце – значения признака, указывающего на принадлежность к классу. Так для субъектов, относящихся по условию к первой обучающей выборке, в десятом столбце введена цифра 1, для субъектов, относящихся ко второй обучающей выборке – цифра 2 и т.д. Для субъектов, подлежащих классификации, значение признака «Обучающая выборка» не указывается.

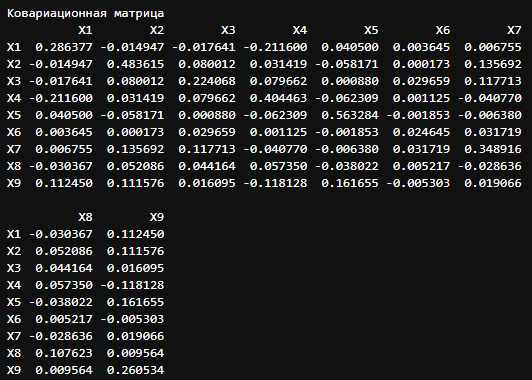
3. Преддискриминантный этап анализа

В рис. 2 представлена оценка общей ковариационной матрицы, в рис. 3 оценка общей корреляционной матрицы.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Pooled Within-Groups Correlations (Данные) | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | |
| |  | | --- | | X1 | | 1,00 | -0,09 | 0,29 | -0,43 | 0,26 | -0,09 | 0,04 | -0,10 | -0,35 |
| |  | | --- | | X2 | | -0,09 | 1,00 | -0,12 | 0,47 | -0,39 | -0,26 | -0,05 | 0,07 | 0,14 |
| |  | | --- | | X3 | | 0,29 | -0,12 | 1,00 | 0,24 | 0,11 | 0,01 | 0,05 | 0,00 | -0,09 |
| |  | | --- | | X4 | | -0,43 | 0,47 | 0,24 | 1,00 | -0,46 | -0,16 | -0,12 | 0,23 | 0,14 |
| |  | | --- | | X5 | | 0,26 | -0,39 | 0,11 | -0,46 | 1,00 | 0,13 | -0,08 | -0,02 | -0,00 |
| |  | | --- | | X6 | | -0,09 | -0,26 | 0,01 | -0,16 | 0,13 | 1,00 | -0,15 | 0,38 | -0,01 |
| |  | | --- | | X7 | | 0,04 | -0,05 | 0,05 | -0,12 | -0,08 | -0,15 | 1,00 | -0,27 | -0,09 |
| |  | | --- | | X8 | | -0,10 | 0,07 | 0,00 | 0,23 | -0,02 | 0,38 | -0,27 | 1,00 | 0,02 |
| |  | | --- | | X9 | | -0,35 | 0,14 | -0,09 | 0,14 | -0,00 | -0,01 | -0,09 | 0,02 | 1,00 |

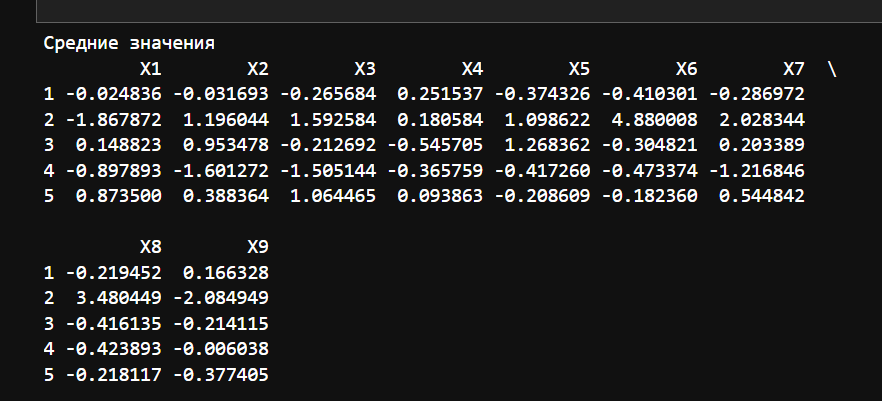


|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Pooled Within-Groups Covariances (Sheet1 in norm\_data) | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | |
| |  | | --- | | X1 | | 0,29 | -0,01 | -0,02 | -0,21 | 0,04 | 0,00 | 0,01 | -0,03 | 0,11 |
| |  | | --- | | X2 | | -0,01 | 0,48 | 0,08 | 0,03 | -0,06 | 0,00 | 0,14 | 0,05 | 0,11 |
| |  | | --- | | X3 | | -0,02 | 0,08 | 0,22 | 0,08 | 0,00 | 0,03 | 0,12 | 0,04 | 0,02 |
| |  | | --- | | X4 | | -0,21 | 0,03 | 0,08 | 0,40 | -0,06 | 0,00 | -0,04 | 0,06 | -0,12 |
| |  | | --- | | X5 | | 0,04 | -0,06 | 0,00 | -0,06 | 0,56 | -0,00 | -0,01 | -0,04 | 0,16 |
| |  | | --- | | X6 | | 0,00 | 0,00 | 0,03 | 0,00 | -0,00 | 0,02 | 0,03 | 0,01 | -0,01 |
| |  | | --- | | X7 | | 0,01 | 0,14 | 0,12 | -0,04 | -0,01 | 0,03 | 0,35 | -0,03 | 0,02 |
| |  | | --- | | X8 | | -0,03 | 0,05 | 0,04 | 0,06 | -0,04 | 0,01 | -0,03 | 0,11 | 0,01 |
| |  | | --- | | X9 | | 0,11 | 0,11 | 0,02 | -0,12 | 0,16 | -0,01 | 0,02 | 0,01 | 0,26 |



Далее рассчитываются оценки математических ожиданий и средних квадратических отклонений признаков в классах.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| обучающая выборка | Means (Sheet1 in norm\_data) | | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | | |  | | --- | | Valid N | |
| |  | | --- | | G\_1:1 | | -0,02484 | -0,03169 | -0,26568 | 0,251537 | -0,374326 | -0,410301 | -0,28697 | -0,219452 | 0,16633 | 11 |
| |  | | --- | | G\_2:2 | | -1,86787 | 1,19604 | 1,59258 | 0,180584 | 1,098622 | 4,880008 | 2,02834 | 3,480449 | -2,08495 | 2 |
| |  | | --- | | G\_3:3 | | 0,14882 | 0,95348 | -0,21269 | -0,545705 | 1,268362 | -0,304821 | 0,20339 | -0,416135 | -0,21412 | 6 |
| |  | | --- | | G\_4:4 | | -0,89789 | -1,60127 | -1,50514 | -0,365759 | -0,417260 | -0,473374 | -1,21685 | -0,423893 | -0,00604 | 5 |
| |  | | --- | | G\_5:5 | | 0,87350 | 0,38836 | 1,06447 | 0,093863 | -0,208609 | -0,182360 | 0,54484 | -0,218117 | -0,37741 | 6 |
| |  | | --- | | All Grps | | -0,07882 | 0,06961 | -0,07175 | -0,047059 | 0,078396 | -0,001442 | -0,02316 | -0,045935 | -0,19732 | 30 |



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| обучающая выборка | Standard Deviations (Sheet1 in norm\_data) | | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | | |  | | --- | | Valid N | |
| |  | | --- | | G\_1:1 | | 0,579913 | 0,659350 | 0,418860 | 0,737262 | 0,757557 | 0,111595 | 0,488568 | 0,284395 | 0,624331 | 11 |
| |  | | --- | | G\_2:2 | | 0,317628 | 1,258118 | 0,031104 | 0,827834 | 2,168658 | 0,109909 | 0,421815 | 0,582623 | 0,131958 | 2 |
| |  | | --- | | G\_3:3 | | 0,421468 | 0,450811 | 0,383173 | 0,599529 | 0,438190 | 0,172708 | 0,685956 | 0,255948 | 0,388268 | 6 |
| |  | | --- | | G\_4:4 | | 0,747630 | 0,881423 | 0,455540 | 0,644855 | 0,709076 | 0,102857 | 0,299747 | 0,380086 | 0,608668 | 5 |
| |  | | --- | | G\_5:5 | | 0,338103 | 0,638171 | 0,675591 | 0,325637 | 0,365760 | 0,240023 | 0,830170 | 0,356904 | 0,269207 | 6 |
| |  | | --- | | All Grps | | 0,884731 | 1,082735 | 1,011888 | 0,677594 | 0,993723 | 1,338737 | 0,969784 | 1,010411 | 0,730458 | 30 |

По указанным данным можно дать экономическую интерпретацию:

Так по показателю X1 лидируют представители группы 4, самые плохие показатели имеет группа 2

По показателю X2 лидирует вторая группа, худшая группа по показателю – третья

По показателю X3 лидирует группа 5, хуже всех кластер номер 2

По показателю X4 лидирует кластер №2 и отстает группа 5

По показателю X5 в лидерах второй кластер, худшие показатели имеет 5

По показателю X6 лучший показатель имеет пятая группа, а худший четвертая

По показателю X7 лучшая группа – 5-я, худшая – 4-я

По показателю X8 лучшая группа – 2-я, худшая – три

По показателю X9 лучшая группа – 1-я, худшая - 2

Далее представлены оценки ковариационной и корреляционной матриц по объединенной выборке:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Covariances (Total) (Sheet1 in norm\_data) | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | |
| |  | | --- | | X1 | | 0,78 | 0,19 | 0,20 | -0,16 | -0,03 | -0,59 | 0,04 | -0,46 | 0,27 |
| |  | | --- | | X2 | | 0,19 | 1,17 | 0,67 | 0,04 | 0,39 | 0,46 | 0,71 | 0,36 | -0,13 |
| |  | | --- | | X3 | | 0,20 | 0,67 | 1,02 | 0,20 | 0,17 | 0,70 | 0,78 | 0,52 | -0,32 |
| |  | | --- | | X4 | | -0,16 | 0,04 | 0,20 | 0,46 | -0,19 | 0,08 | 0,03 | 0,14 | -0,10 |
| |  | | --- | | X5 | | -0,03 | 0,39 | 0,17 | -0,19 | 0,99 | 0,39 | 0,31 | 0,20 | -0,07 |
| |  | | --- | | X6 | | -0,59 | 0,46 | 0,70 | 0,08 | 0,39 | 1,79 | 0,82 | 1,28 | -0,70 |
| |  | | --- | | X7 | | 0,04 | 0,71 | 0,78 | 0,03 | 0,31 | 0,82 | 0,94 | 0,53 | -0,35 |
| |  | | --- | | X8 | | -0,46 | 0,36 | 0,52 | 0,14 | 0,20 | 1,28 | 0,53 | 1,02 | -0,48 |
| |  | | --- | | X9 | | 0,27 | -0,13 | -0,32 | -0,10 | -0,07 | -0,70 | -0,35 | -0,48 | 0,53 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Correlations (Total) (Sheet1 in norm\_data) | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | |
| |  | | --- | | X1 | | 1,00 | 0,19 | 0,22 | -0,26 | -0,03 | -0,50 | 0,04 | -0,52 | 0,42 |
| |  | | --- | | X2 | | 0,19 | 1,00 | 0,61 | 0,06 | 0,36 | 0,32 | 0,68 | 0,32 | -0,17 |
| |  | | --- | | X3 | | 0,22 | 0,61 | 1,00 | 0,29 | 0,17 | 0,52 | 0,79 | 0,51 | -0,43 |
| |  | | --- | | X4 | | -0,26 | 0,06 | 0,29 | 1,00 | -0,29 | 0,09 | 0,04 | 0,20 | -0,21 |
| |  | | --- | | X5 | | -0,03 | 0,36 | 0,17 | -0,29 | 1,00 | 0,29 | 0,32 | 0,20 | -0,09 |
| |  | | --- | | X6 | | -0,50 | 0,32 | 0,52 | 0,09 | 0,29 | 1,00 | 0,63 | 0,95 | -0,72 |
| |  | | --- | | X7 | | 0,04 | 0,68 | 0,79 | 0,04 | 0,32 | 0,63 | 1,00 | 0,54 | -0,49 |
| |  | | --- | | X8 | | -0,52 | 0,32 | 0,51 | 0,20 | 0,20 | 0,95 | 0,54 | 1,00 | -0,65 |
| |  | | --- | | X9 | | 0,42 | -0,17 | -0,43 | -0,21 | -0,09 | -0,72 | -0,49 | -0,65 | 1,00 |

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. Результаты проверки такой гипотезы по каждому отдельному признаку

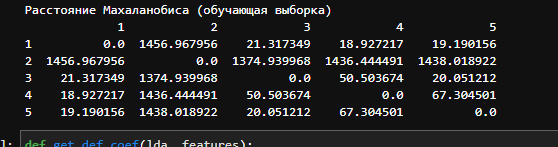
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=30 | Discriminant Function Analysis Summary (Sheet1 in norm\_data) No. of vars in model: 9; Grouping: обучающая выборка (5 grps) Wilks' Lambda: ,00009 approx. F (36,65)=20,256 p<0,0000 | | | | | |
| |  | | --- | | Wilks' Lambda | | |  | | --- | | Partial Lambda | | |  | | --- | | F-remove (4,17) | | |  | | --- | | p-value | | |  | | --- | | Toler. | | |  | | --- | | 1-Toler. (R-Sqr.) | |
| |  | | --- | | X1 | | 0,000123 | 0,701010 | 1,81269 | 0,172914 | 0,632126 | 0,367874 |
| |  | | --- | | X2 | | 0,000171 | 0,505038 | 4,16522 | 0,015662 | 0,608887 | 0,391113 |
| |  | | --- | | X3 | | 0,000137 | 0,633030 | 2,46374 | 0,084513 | 0,522853 | 0,477147 |
| |  | | --- | | X4 | | 0,000109 | 0,790827 | 1,12412 | 0,377938 | 0,521394 | 0,478606 |
| |  | | --- | | X5 | | 0,000145 | 0,594321 | 2,90102 | 0,053379 | 0,666666 | 0,333334 |
| |  | | --- | | X6 | | 0,000668 | 0,129421 | 28,58847 | 0,000000 | 0,696954 | 0,303046 |
| |  | | --- | | X7 | | 0,000110 | 0,787490 | 1,14690 | 0,368317 | 0,473899 | 0,526101 |
| |  | | --- | | X8 | | 0,000109 | 0,791578 | 1,11902 | 0,380124 | 0,575166 | 0,424834 |
| |  | | --- | | X9 | | 0,000206 | 0,419783 | 5,87427 | 0,003706 | 0,580657 | 0,419343 |

Значение статистики Уилкса принадлежит интервалу от 0 до 1. Значение близкое к нулю (0,00001) свидетельствует о хорошей дискриминации.

Гипотеза принимается только для признаков X2;X6;X9

Далее исследуется различие между классами с использованием расстояния Махаланобиса. Квадрат расстояния Махаланосиба между классами, наблюденное значение F-статистики и значимость нулевой гипотезы о равенстве нулю расстояния между классами

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| обучающая выборка | Squared Mahalanobis Distances (Sheet1 in norm\_data) | | | | |
| |  | | --- | | G\_1:1 | | |  | | --- | | G\_2:2 | | |  | | --- | | G\_3:3 | | |  | | --- | | G\_4:4 | | |  | | --- | | G\_5:5 | |
| |  | | --- | | G\_1:1 | | 0,000 | 2109,474 | 27,006 | 19,438 | 26,147 |
| |  | | --- | | G\_2:2 | | 2109,474 | 0,000 | 1915,553 | 2003,434 | 1974,660 |
| |  | | --- | | G\_3:3 | | 27,006 | 1915,553 | 0,000 | 42,400 | 35,304 |
| |  | | --- | | G\_4:4 | | 19,438 | 2003,434 | 42,400 | 0,000 | 64,286 |
| |  | | --- | | G\_5:5 | | 26,147 | 1974,660 | 35,304 | 64,286 | 0,000 |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| обучающая выборка | F-values; df = 9,17 (Sheet1 in norm\_data) | | | | |
| |  | | --- | | G\_1:1 | | |  | | --- | | G\_2:2 | | |  | | --- | | G\_3:3 | | |  | | --- | | G\_4:4 | | |  | | --- | | G\_5:5 | |
| |  | | --- | | G\_1:1 | |  | 269,7242 | 7,9218 | 5,0485 | 7,6698 |
| |  | | --- | | G\_2:2 | | 269,7242 |  | 217,0960 | 216,2437 | 223,7948 |
| |  | | --- | | G\_3:3 | | 7,9218 | 217,0960 |  | 8,7370 | 8,0021 |
| |  | | --- | | G\_4:4 | | 5,0485 | 216,2437 | 8,7370 |  | 13,2469 |
| |  | | --- | | G\_5:5 | | 7,6698 | 223,7948 | 8,0021 | 13,2469 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 0,000000 | 0,000148 | 0,002058 | 0,000181 |
| 0,000000 |  | 0,000000 | 0,000000 | 0,000000 |
| 0,000148 | 0,000000 |  | 0,000079 | 0,000139 |
| 0,002058 | 0,000000 | 0,000079 |  | 0,000005 |
| 0,000181 | 0,000000 | 0,000139 | 0,000005 |  |

4. Дискриминантный анализ

Рассчитаем коэффициенты линейных дискриминантных функций Фишера. Результаты представлены в таблице 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Classification Functions; grouping: обучающая выборка (Sheet1 in norm\_data) | | | | |
| |  | | --- | | G\_1:1 p=,36667 | | |  | | --- | | G\_2:2 p=,06667 | | |  | | --- | | G\_3:3 p=,20000 | | |  | | --- | | G\_4:4 p=,16667 | | |  | | --- | | G\_5:5 p=,20000 | |
| |  | | --- | | X1 | | -1,7656 | -4,797 | -3,4912 | -7,2170 | 5,5433 |
| |  | | --- | | X2 | | -1,0217 | 11,553 | 6,4248 | -4,1702 | 1,1240 |
| |  | | --- | | X3 | | 0,6508 | -17,656 | -2,0389 | -3,5717 | 8,5398 |
| |  | | --- | | X4 | | 3,0864 | -28,648 | -2,9870 | -2,0490 | 2,0975 |
| |  | | --- | | X5 | | -2,1021 | 16,429 | 3,7602 | -0,6119 | -3,3705 |
| |  | | --- | | X6 | | -25,8005 | 306,134 | -16,7621 | -18,3788 | -16,4817 |
| |  | | --- | | X7 | | 4,4649 | -40,625 | 1,8667 | 2,0100 | -0,2084 |
| |  | | --- | | X8 | | -9,5598 | 65,775 | -3,1783 | -3,6069 | -5,7703 |
| |  | | --- | | X9 | | 7,7128 | -54,722 | -3,7504 | 5,0958 | -1,6318 |
| |  | | --- | | Constant | | -9,8530 | -883,752 | -11,6364 | -15,4372 | -11,6271 |

В первой строке таблицы приведены оценки априорных вероятностей, рассчитанные по первому варианту:

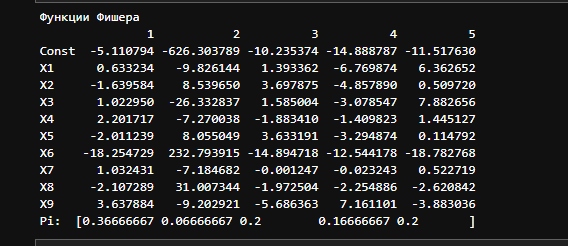
Линейные дискриминантные функции Фишера имеют вид:

*;*

;

;

;



На основе рассчитанных классификационных функций проводится повторная классификация объектов обучающих выборок.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group | Classification Matrix (Sheet1 in norm\_data) Rows: Observed classifications Columns: Predicted classifications | | | | | |
| |  | | --- | | Percent Correct | | |  | | --- | | G\_1:1 p=,36667 | | |  | | --- | | G\_2:2 p=,06667 | | |  | | --- | | G\_3:3 p=,20000 | | |  | | --- | | G\_4:4 p=,16667 | | |  | | --- | | G\_5:5 p=,20000 | |
| |  | | --- | | G\_1:1 | | 100,0000 | 11 | 0 | 0 | 0 | 0 |
| |  | | --- | | G\_2:2 | | 100,0000 | 0 | 2 | 0 | 0 | 0 |
| |  | | --- | | G\_3:3 | | 100,0000 | 0 | 0 | 6 | 0 | 0 |
| |  | | --- | | G\_4:4 | | 100,0000 | 0 | 0 | 0 | 5 | 0 |
| |  | | --- | | G\_5:5 | | 100,0000 | 0 | 0 | 0 | 0 | 6 |
| |  | | --- | | Total | | 100,0000 | 11 | 2 | 6 | 5 | 6 |

Далее производится Классификация всех объектов (Приложение А)

В первой графе таблицы указаны номера объектов; во второй графе – номера обучающих выборок (прочерками отмечаются объекты, не вошедшие в обучающие выборки); в третьем-десятом столбцах – результаты классификации с помощью дискриминантных функций (1) – (8).

Далее рассчитываются квадраты расстояния Махаланобиса (Приложение А) от объектов до центров каждого из классов. Объект следует отнести к тому классу, расстояние до которого наименьшее.

Также рассчитываются апостериорные вероятности классификации. (Приложение А) Объект следует отнести к тому классу, апостериорная вероятность для которого наибольшая.

|  |  |  |
| --- | --- | --- |
| **Номер кластера** | **Количество объектов** | **Состав класса** |
| {S41} | 32 | Астраханская Область, Брянская Область, Волгоградская Область, Воронежская Область, Ивановская Область, Иркутская Область, Калининградская Область, Калужская Область, Краснодарский Край, Красноярский Край, Курская Область, Нижегородская Область, Новосибирская Область, Омская Область, Оренбургская Область, Орловская Область, Пермский Край, Республика Крым, Республика Марий-Эл, Республика Мордовия, Республика Саха, Республика Татарстан, Рязанская Область, Самарская Область, Саратовская Область, Свердловская Область, Томская Область, Тульская Область, Удмуртская Республика, Ульяновская Область, Челябинская Область, Ярославская Область. |
| {S42} | 3 | Ненецкий Автономный Округ, Чукотский Автономный Округ, Ямало-ненецкий Автономный округ |
| {S43} | 17 | Алтайский Край, Белгородская Область, Кемеровская Область, Курганская Область, Липецкая Область, Новгородская Область, Пензенская Область, Республика Башкортостан, Республика Бурятия, Республика Калмыкия, Республика Тыва, Республика Хакассия, Смоленская Область, Тамбовская Область, Тверская Область, Тюменская Область, Чувашская республика |
| {S44} | 14 | Москва, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Московская область, Республика Адыгея, Республика Дагестан, Республика Ингушетия, Республика Северная Осетия – Алания, Ростовская Область, Санкт-Петербург, Севастополь, Ставропольский край, Ханты-Мансийский Автономный округ, Чеченская Республика |
| {S25} | 19 | Амурская Область, Архангельская Область, Владимирская Область, Вологодская Область, Еврейская Автономная Область, Забайкальский Край, Камчатский Край, Кировская Область, Костромская Область, Ленинградская Область, Магаданская Область, Мурманская Область, Приморский Край, Псковская Область, Республика Алтай, Республика Карелия, Республика Коми, Сахалинская Область, Хабаровский Край. |

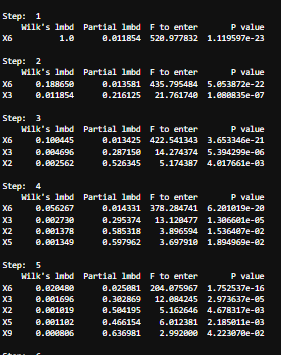
Точность:100,00%

Выборка была составлена по методу полной связи, совпадение классификации с исходной составило .

1. Пошаговый анализ. Включение

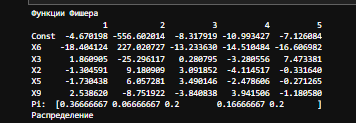
Аналогично проведем пошаговый дискриминантный анализ с пошаговым включением

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=30 | Discriminant Function Analysis Summary (Sheet1 in norm\_data) Step 5, N of vars in model: 5; Grouping: обучающая выборка (5 grps) Wilks' Lambda: ,00025 approx. F (20,70)=39,369 p<0,0000 | | | | | |
| |  | | --- | | Wilks' Lambda | | |  | | --- | | Partial Lambda | | |  | | --- | | F-remove (4,21) | | |  | | --- | | p-value | | |  | | --- | | Toler. | | |  | | --- | | 1-Toler. (R-Sqr.) | |
| |  | | --- | | X6 | | 0,011325 | 0,022310 | 230,0695 | 0,000000 | 0,854642 | 0,145358 |
| |  | | --- | | X3 | | 0,000866 | 0,291742 | 12,7454 | 0,000020 | 0,732451 | 0,267549 |
| |  | | --- | | X2 | | 0,000613 | 0,412291 | 7,4837 | 0,000653 | 0,714734 | 0,285266 |
| |  | | --- | | X9 | | 0,000517 | 0,489116 | 5,4837 | 0,003488 | 0,768478 | 0,231522 |
| |  | | --- | | X5 | | 0,000502 | 0,502994 | 5,1875 | 0,004572 | 0,852176 | 0,147824 |



Рассчитаем коэффициенты линейных дискриминантных функций Фишера:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Classification Functions; grouping: обучающая выборка (Sheet1 in norm\_data) | | | | |
| |  | | --- | | G\_1:1 p=,36667 | | |  | | --- | | G\_2:2 p=,06667 | | |  | | --- | | G\_3:3 p=,20000 | | |  | | --- | | G\_4:4 p=,16667 | | |  | | --- | | G\_5:5 p=,20000 | |
| |  | | --- | | X6 | | -18,7804 | 245,063 | -13,1879 | -14,9438 | -15,8776 |
| |  | | --- | | X3 | | 0,9760 | -24,547 | -2,5762 | -3,7587 | 7,3805 |
| |  | | --- | | X2 | | -0,8990 | 7,269 | 5,8473 | -4,7915 | 1,0092 |
| |  | | --- | | X9 | | 4,0383 | -23,686 | -4,5844 | 4,0767 | -2,6537 |
| |  | | --- | | X5 | | -1,0805 | 8,156 | 4,3469 | -0,7427 | -1,6112 |
| |  | | --- | | Constant | | -5,6338 | -614,636 | -9,9285 | -12,1364 | -7,8501 |



В первой строке таблицы приведены оценки априорных вероятностей, рассчитанные по первому варианту:

Линейные дискриминантные функции Фишера имеют вид:

*;*

*;*

*;*

*;*

*;*

На основе рассчитанных классификационных функций проводится повторная классификация объектов обучающих выборок.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group | Classification Matrix (Sheet1 in norm\_data) Rows: Observed classifications Columns: Predicted classifications | | | | | |
| |  | | --- | | Percent Correct | | |  | | --- | | G\_1:1 p=,36667 | | |  | | --- | | G\_2:2 p=,06667 | | |  | | --- | | G\_3:3 p=,20000 | | |  | | --- | | G\_4:4 p=,16667 | | |  | | --- | | G\_5:5 p=,20000 | |
| |  | | --- | | G\_1:1 | | 100,0000 | 11 | 0 | 0 | 0 | 0 |
| |  | | --- | | G\_2:2 | | 100,0000 | 0 | 2 | 0 | 0 | 0 |
| |  | | --- | | G\_3:3 | | 100,0000 | 0 | 0 | 6 | 0 | 0 |
| |  | | --- | | G\_4:4 | | 60,0000 | 2 | 0 | 0 | 3 | 0 |
| |  | | --- | | G\_5:5 | | 83,3333 | 0 | 0 | 1 | 0 | 5 |
| |  | | --- | | Total | | 90,0000 | 13 | 2 | 7 | 3 | 5 |

Как видно из таблицы, изменения в первоначальном составе классов произошли: К первому классу относится 13 объектов, ко второму – 2, К третьему 7, к четвертому 3, к пятому 5, при этом 2 объекта из 4 класса перескочили во второй, объект из 5 класса перескочил в третий,

Общая точность 90,00%

На основании таблиц, представленных в приложении Б, субъекты можно классифицировать следующим образом:

|  |  |  |
| --- | --- | --- |
| Класс | Количество субъектов | Наименование субъектов |
| 1 | 35 | Астраханская область  Белгородская область  Волгоградская область  Воронежская область  Ивановская область  Иркутская область  Калининградская область  Краснодарский край  Красноярский край  Курская область  Нижегородская область  Новосибирская область  Омская область  Орловская область  Республика Адыгея (Адыгея)  Республика Бурятия  Республика Калмыкия  Республика Крым  Республика Марий Эл  Республика Мордовия  Республика Саха (Якутия)  Республика Татарстан (Татарстан)  Ростовская область  Рязанская область  Самарская область  Санкт-Петербург  Саратовская область  Свердловская область  Севастополь  Томская область  Тульская область  Удмуртская Республика  Ульяновская область  Челябинская область  Чувашская Республика — Чувашия |
| 2 | 3 | Ненецкий автономный округ  Чукотский автономный округ  Ямало-Ненецкий автономный округ |
| 3 | 20 | Алтайский край  Амурская область  Брянская область  Еврейская автономная область  Забайкальский край  Кемеровская область — Кузбасс  Курганская область  Липецкая область  Новгородская область  Оренбургская область  Пензенская область  Псковская область  Республика Алтай  Республика Башкортостан  Республика Тыва  Республика Хакасия  Смоленская область  Тамбовская область  Тверская область  Тюменская область без автономных округов |
| 4 | 8 | г. Москва  Кабардино-Балкарская Республика  Карачаево-Черкесская Республика  Республика Дагестан  Республика Ингушетия  Республика Северная Осетия — Алания  Ставропольский край  Чеченская Республика |
| 5 | 19 | Архангельская область  Владимирская область  Вологодская область  Калужская область  Камчатский край  Кировская область  Костромская область  Ленинградская область  Магаданская область  Московская область  Мурманская область  Пермский край  Приморский край  Республика Карелия  Республика Коми  Сахалинская область  Хабаровский край  Ханты-Мансийский автономный округ — Югра  Ярославская область |

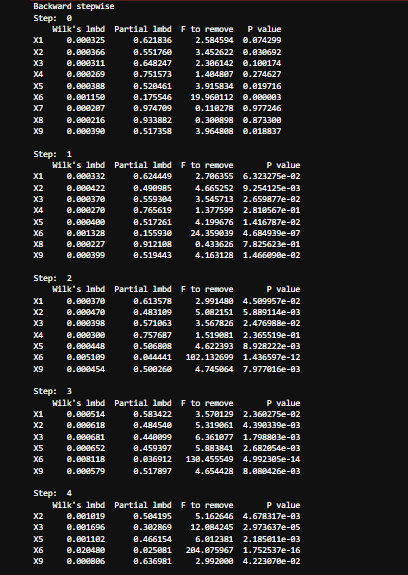
Обучающая Выборка была составлена по методу полной связи, совпадение классификации с исходной составило .

Результаты анализа представлены в приложении Б

1. Пошаговый анализ. Исключение

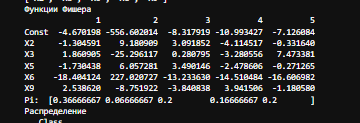
Аналогично проведем пошаговый дискриминантный анализ с пошаговым исключением

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=30 | Discriminant Function Analysis Summary (Sheet1 in norm\_data) Step 4, N of vars in model: 5; Grouping: обучающая выборка (5 grps) Wilks' Lambda: ,00025 approx. F (20,70)=39,369 p<0,0000 | | | | | |
| |  | | --- | | Wilks' Lambda | | |  | | --- | | Partial Lambda | | |  | | --- | | F-remove (4,21) | | |  | | --- | | p-value | | |  | | --- | | Toler. | | |  | | --- | | 1-Toler. (R-Sqr.) | |
| |  | | --- | | X2 | | 0,000613 | 0,412291 | 7,4837 | 0,000653 | 0,714734 | 0,285266 |
| |  | | --- | | X3 | | 0,000866 | 0,291742 | 12,7454 | 0,000020 | 0,732451 | 0,267549 |
| |  | | --- | | X5 | | 0,000502 | 0,502994 | 5,1875 | 0,004572 | 0,852176 | 0,147824 |
| |  | | --- | | X6 | | 0,011325 | 0,022310 | 230,0695 | 0,000000 | 0,854642 | 0,145358 |
| |  | | --- | | X9 | | 0,000517 | 0,489116 | 5,4837 | 0,003488 | 0,768478 | 0,231522 |



Рассчитаем коэффициенты линейных дискриминантных функций Фишера:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Classification Functions; grouping: обучающая выборка (Sheet1 in norm\_data) | | | | |
| |  | | --- | | G\_1:1 p=,36667 | | |  | | --- | | G\_2:2 p=,06667 | | |  | | --- | | G\_3:3 p=,20000 | | |  | | --- | | G\_4:4 p=,16667 | | |  | | --- | | G\_5:5 p=,20000 | |
| |  | | --- | | X2 | | -0,8990 | 7,269 | 5,8473 | -4,7915 | 1,0092 |
| |  | | --- | | X3 | | 0,9760 | -24,547 | -2,5762 | -3,7587 | 7,3805 |
| |  | | --- | | X5 | | -1,0805 | 8,156 | 4,3469 | -0,7427 | -1,6112 |
| |  | | --- | | X6 | | -18,7804 | 245,063 | -13,1879 | -14,9438 | -15,8776 |
| |  | | --- | | X9 | | 4,0383 | -23,686 | -4,5844 | 4,0767 | -2,6537 |
| |  | | --- | | Constant | | -5,6338 | -614,636 | -9,9285 | -12,1364 | -7,8501 |



На основе рассчитанных классификационных функций проводится повторная классификация объектов обучающих выборок.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group | Classification Matrix (Sheet1 in norm\_data) Rows: Observed classifications Columns: Predicted classifications | | | | | |
| |  | | --- | | Percent Correct | | |  | | --- | | G\_1:1 p=,36667 | | |  | | --- | | G\_2:2 p=,06667 | | |  | | --- | | G\_3:3 p=,20000 | | |  | | --- | | G\_4:4 p=,16667 | | |  | | --- | | G\_5:5 p=,20000 | |
| |  | | --- | | G\_1:1 | | 100,0000 | 11 | 0 | 0 | 0 | 0 |
| |  | | --- | | G\_2:2 | | 100,0000 | 0 | 2 | 0 | 0 | 0 |
| |  | | --- | | G\_3:3 | | 100,0000 | 0 | 0 | 6 | 0 | 0 |
| |  | | --- | | G\_4:4 | | 60,0000 | 2 | 0 | 0 | 3 | 0 |
| |  | | --- | | G\_5:5 | | 83,3333 | 0 | 0 | 1 | 0 | 5 |
| |  | | --- | | Total | | 90,0000 | 13 | 2 | 7 | 3 | 5 |

Как видно из таблицы, изменения в первоначальном составе классов произошли: К первому классу относится 13 объектов, ко второму – 2, К третьему 7, к четвертому 3, к пятому 5, при этом 2 объекта из 4 класса перескочили во второй, объект из 5 класса перескочил в третий,

Общая точность 90,00%

|  |  |  |
| --- | --- | --- |
| Класс | Количество субъектов | Наименование субъектов |
| 1 | 35 | Астраханская область  Белгородская область  Волгоградская область  Воронежская область  Ивановская область  Иркутская область  Калининградская область  Краснодарский край  Красноярский край  Курская область  Нижегородская область  Новосибирская область  Омская область  Орловская область  Республика Адыгея (Адыгея)  Республика Бурятия  Республика Калмыкия  Республика Крым  Республика Марий Эл  Республика Мордовия  Республика Саха (Якутия)  Республика Татарстан (Татарстан)  Ростовская область  Рязанская область  Самарская область  Санкт-Петербург  Саратовская область  Свердловская область  Севастополь  Томская область  Тульская область  Удмуртская Республика  Ульяновская область  Челябинская область  Чувашская Республика — Чувашия |
| 2 | 3 | Ненецкий автономный округ  Чукотский автономный округ  Ямало-Ненецкий автономный округ |
| 3 | 20 | Алтайский край  Амурская область  Брянская область  Еврейская автономная область  Забайкальский край  Кемеровская область — Кузбасс  Курганская область  Липецкая область  Новгородская область  Оренбургская область  Пензенская область  Псковская область  Республика Алтай  Республика Башкортостан  Республика Тыва  Республика Хакасия  Смоленская область  Тамбовская область  Тверская область  Тюменская область без автономных округов |
| 4 | 8 | г. Москва  Кабардино-Балкарская Республика  Карачаево-Черкесская Республика  Республика Дагестан  Республика Ингушетия  Республика Северная Осетия — Алания  Ставропольский край  Чеченская Республика |
| 5 | 19 | Архангельская область  Владимирская область  Вологодская область  Калужская область  Камчатский край  Кировская область  Костромская область  Ленинградская область  Магаданская область  Московская область  Мурманская область  Пермский край  Приморский край  Республика Карелия  Республика Коми  Сахалинская область  Хабаровский край  Ханты-Мансийский автономный округ — Югра  Ярославская область |
| Класс | Количество субъектов | Наименование субъектов |
| 1 | 2 | Ненецкий Автономный Округ, Чукотский Автономный округ; |
| 2 | 3 | Санкт-Петербург, Москва, Томская Область |
| 3 | 4 | Кабардино-балкарская Республика, Республика Дагестан, Республика Ингушетия, Чеченская Республика |
| 4 | 26 | Архангельская Область, Владимирская Область, Вологодская Область, Еврейская Автономная область, Ивановская область, Калужская область, Кировская Область, Костромская Область, Ленинградская Область, Московская Область, Мурманская Область, Пермский край, приморский край, Псковская Область, Республика Алтай, Республика Бурятия, Республика Карелия, Республика Коми, Республика Марий-Эл, Сахалинская Область, Тульская Область, Удмуртская Республика. Хабаровский Край, Чувашская Республика, Ханты-мансийский Автономный Округ, Ярославская Область |
| 5 | 5 | Забайкальский Край, Камчатский Край, Магаданская Область, Республика Тыва, Ямало-ненецкий Автономный Округ |
| 6 | 34 | Алтайский Край, Астраханская Область, Белгородская Область, Брянская Область, Волгоградская Область, Воронежская Область, Иркутская Область, Калининградская Область, Карачаево-Черкесская Республика, Краснодарский Край, Красноярский Край, Курская Область, Нижегородская Область, Новосибирская Область, Омская Область, Оренбургская Область, Орловская Область, Республика Адыгея, Республика Калмыкия, Республика Крым, Республика Мордовия, Республика Саха, Республика Северная Осетия – Алания, Республика Татарстан, Республика Хакассия, Ростовская Область, Рязанская Область, Самарская Область, Саратовская Область, Свердловская Область, Севастополь, Ставропольский Край, Ульяновская Область, Челябинская Область. |
| 7 | 11 | Амурская Область, Кемеровская Область, Курганская Область, Липецкая Область, Новгородская Область, Пензенская Область, Республика Башкортостан, Смоленская Область, Тамбовская Область, Тверская Облсть, Тюменская Область |

Результаты анализа в приложении В

Обучающая Выборка была составлена по методу полной связи, совпадение классификации с исходной составило .