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 С19-702:** | Нашивочников А.В. |
| **Проверил:** | Домашова Д.В. |

**Оглавление**

**[Постановка задачи](#_Toc86011305)** [3](#_Toc86011305)

**[Порядок выполнения работы](#_Toc86011306)** [4](#_Toc86011306)

**[Преддискриминантный этап анализа](#_Toc86011307)** [4](#_Toc86011307)

**[Дискриминантный анализ](#_Toc86011308)** [9](#_Toc86011308)

# **Постановка задачи**

1. В данной лабораторной работе предметом исследования является состояние субъектов РФ, которое характеризуется следующими показателями:
2. X1 – брачность (число браков на 1000 человек)
3. X2 - средний размер назначенных пенсий
4. X3 – заболеваемость на 1000 человек населения
5. X4 - процент безработных мужчин
6. X5 - число посещений музея на 1000 человек населения
7. X6 - число населения на одну больничную койку
8. X7 – потребительские расходы в среднем на душу населения тыс. рублей
9. X8 – количество врачей на 1000 человек населения
10. X9 – число спортивных сооружений на 1000 человек
11. На основе предварительного экспертного анализа было выделено семь групп субъектов РФ. К первой группе отнесено 29 субъектов, ко второй - 13, к третьей – 7, к четвертой – 19, к пятой – 2, к шестой – 10, к седьмой – 5 субъекта РФ;
12. На основе семи обучающих выборок из многомерных нормально распределенных генеральных совокупностей с равными ковариационными матрицами необходимо провести классификацию оставшихся 55 субъектов РФ;
13. Дать экономическую интерпретацию результатов классификации.

# 

# **Порядок выполнения работы**

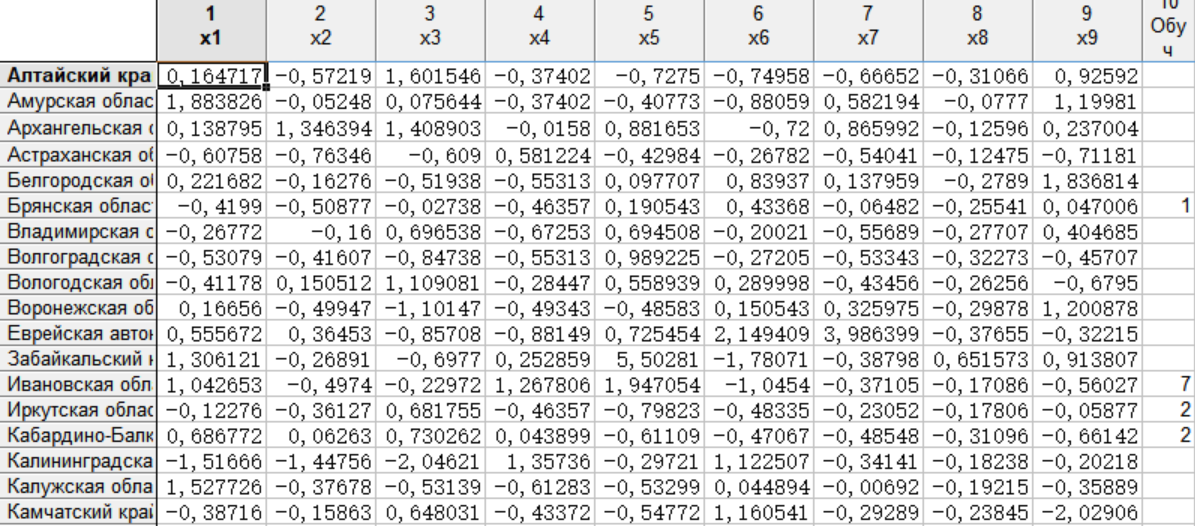
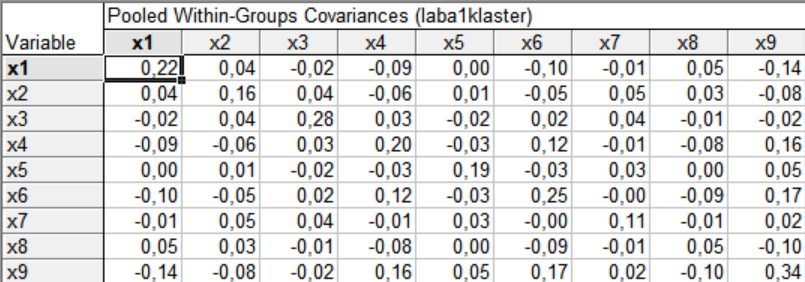


Таблица 1 – Исходные данные

# **Преддискриминантный этап анализа**



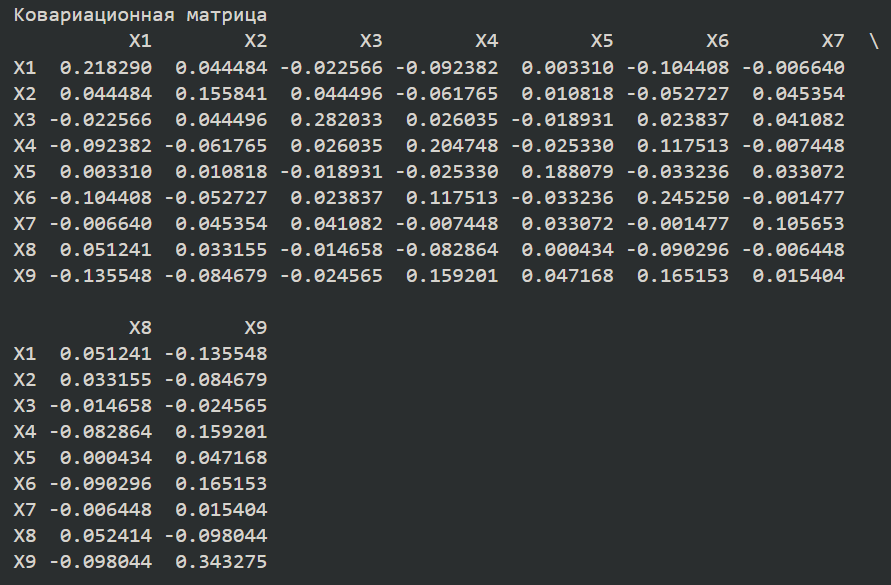
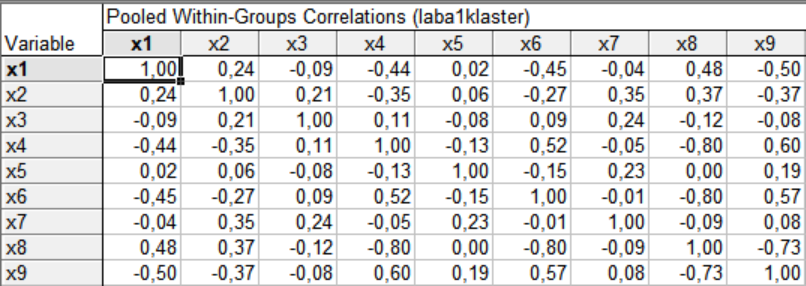
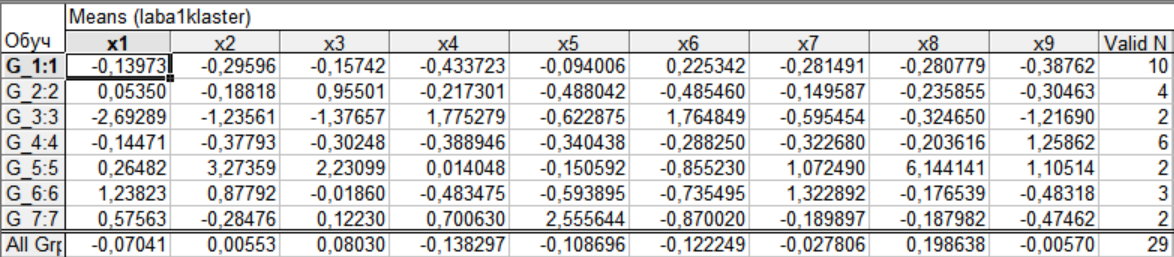


Таблица2–Оценка общей ковариационной матрицыТаблица 3 – Оценка общей корреляционной матрицы



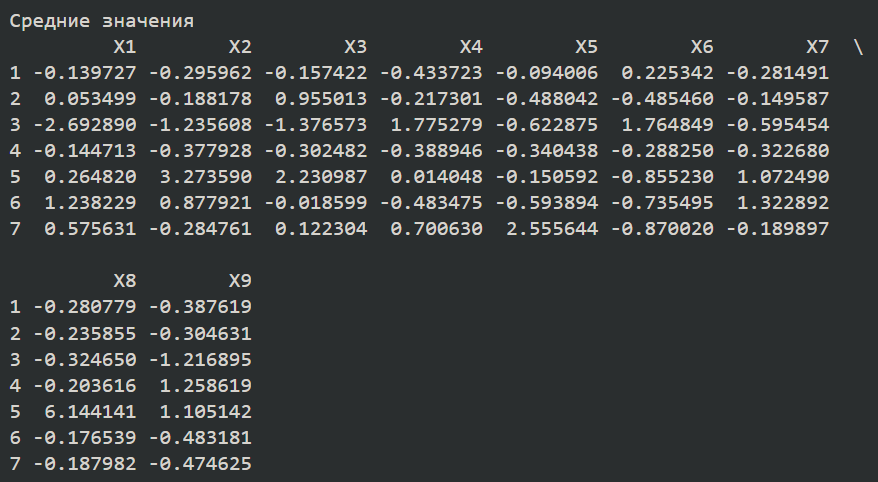
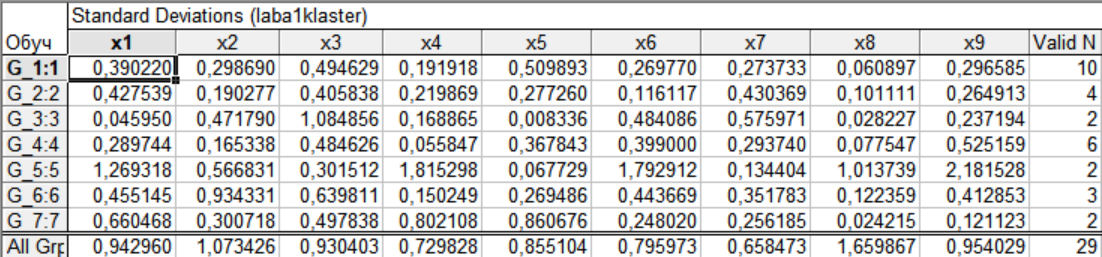


Таблица 4 – Оценки математических ожиданий признаков в классах

Таблица 5 –Оценки средних квадратических отклонений признаков в классах

Выводы по таблице 4:

Кластер 1.

Высокое среднее значение: -

Низкое среднее значение:-

Кластер 2.

Высокое среднее значение: заболеваемость на 1000 человек населенияX3

Низкое среднее значение:-

Кластер 3.

Высокое среднее значение: процент безработных мужчин X4, число населения на одну больничную койку X6

Низкое среднее значение: брачность X1, средний размер назначенных пенсий X2, заболеваемость на 1000 человек населения X3, потребительские расходы в среднем на душу населения X7

Кластер 4.

Высокое среднее значение: число спортивных сооружений на 1000 человек X9

Низкое среднее значение:-

Кластер 5.

Высокое среднее значение: средний размер назначенных пенсий X2, заболеваемость на 1000 человек населения X3,количество врачей на 1000 человек населения X8, числа спортивных сооружений на 1000 человек X9

Низкое среднее значение: число населения на одну больничную койку X6

Кластер 6.

Высокое среднее значение: брачность X1, средний размер назначенных пенсий X2, потребительские расходы в среднем на душу населения X7

Низкое среднее значение: процент безработных мужчин X4

Кластер 7.

Высокое среднее значение: процент безработных мужчин X4, число посещений музея на 1000 человек населения X5

Низкое среднее значение: число населения на одну больничную койку X6

Далее представлены оценки ковариационной и корреляционной матриц по объединенной выборке:

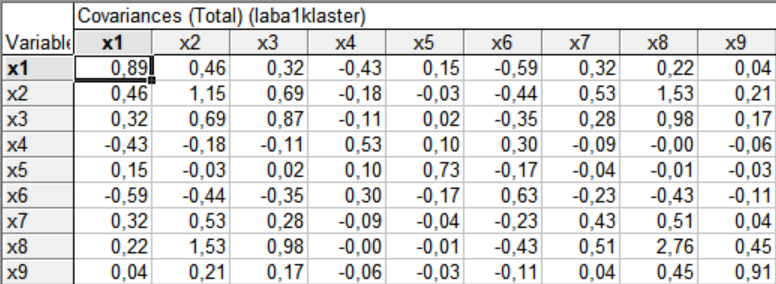


Таблица 6 – Оценка ковариационной матрицы по объединённой выборке

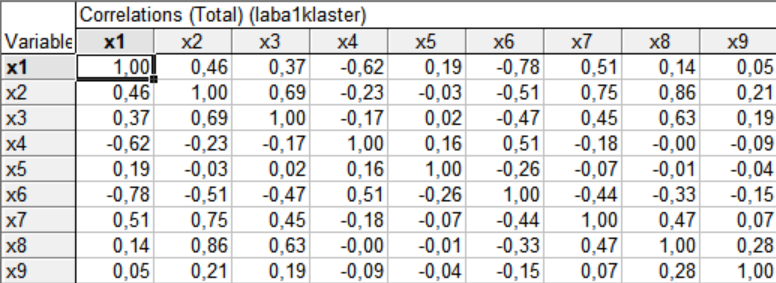


Таблица 7 – Оценка корреляционной матрицы по объединённой выборке

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 представлены в таблице 8:

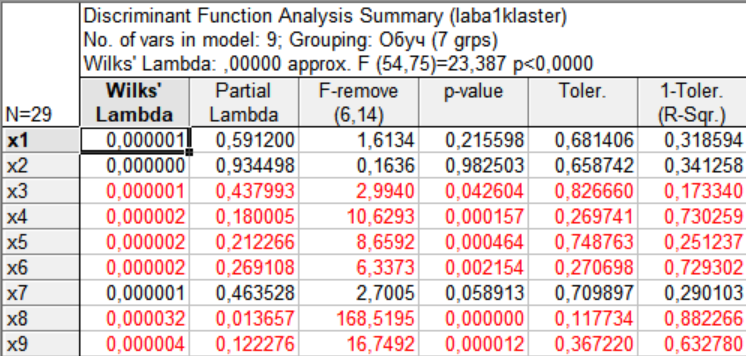


Таблица 8 – Результаты проверки гипотезы об отсутствии различий в групповых средних значениях по каждому признаку

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

Различие в групповых средних значениях всех признаков значимо (значимость нулевой гипотезы ), гипотеза отвергается только для признаков X2 средний размер назначенных пенсий , X1 брачность (число браков на 1000 человек) и X7 потребительские расходы в среднем на душу населения тыс. рублей

В подобной ситуации целесообразно было бы обратиться к процедурам пошагового дискриминантного анализа.

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 представлены в табл. 9, табл.10 и табл. 11 соответственно.

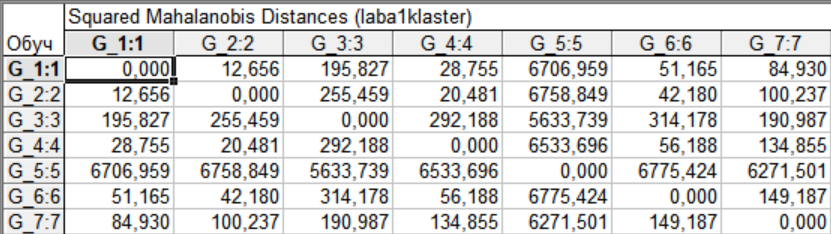


Таблица 9 – Квадрат расстояния между классами

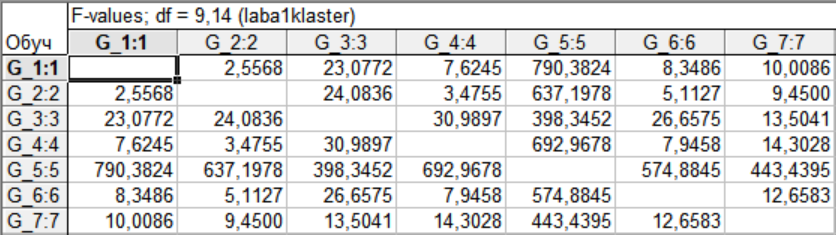


Таблица 10 – Наблюденное значение F-статистики

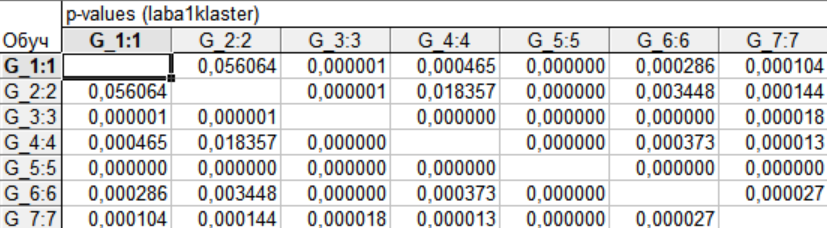


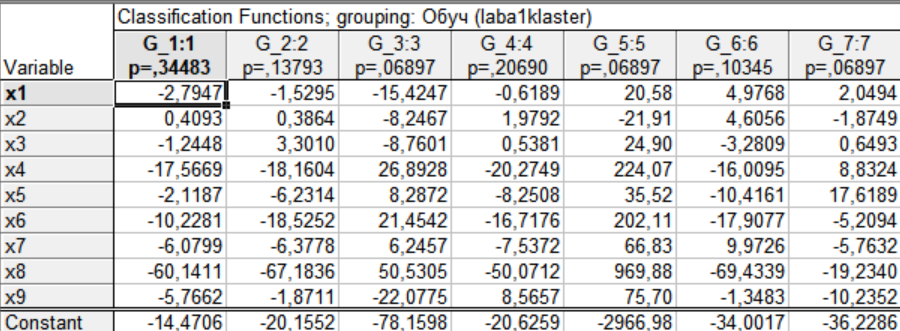
Таблица 11– Значимость нулевой гипотезы о равенстве нулю расстояния между классами

# **Дискриминантный анализ**

Рассчитаем коэффициенты линейных дискриминантных функций Фишера. Результаты представлены в табл. 12:



Рисунок 1 – Результаты дискриминантного анализа



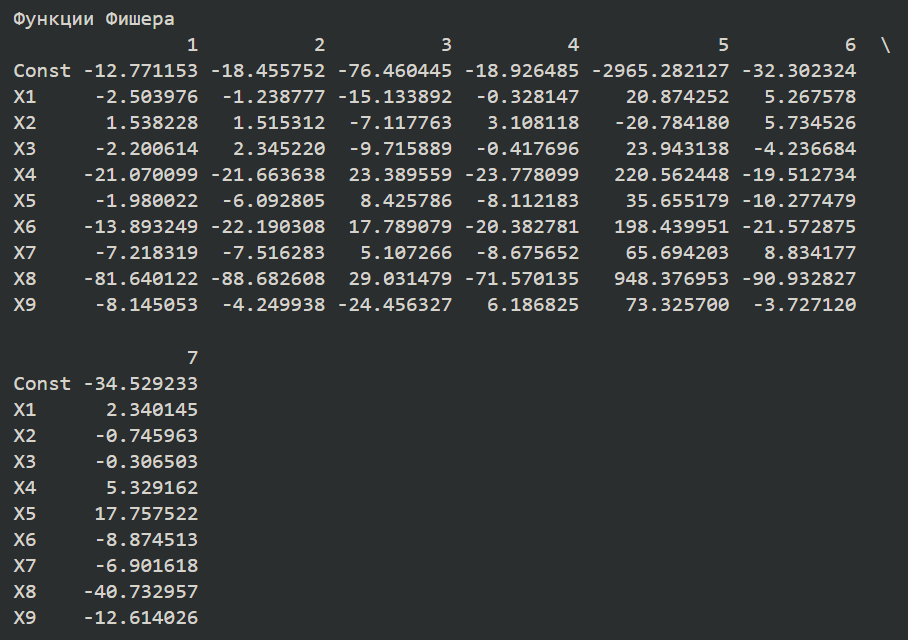
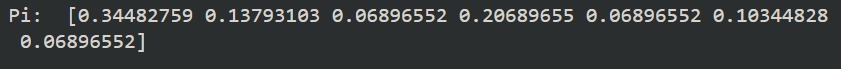


Таблица 12 – Коэффициенты линейных дискриминантных функций Фишера

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

.



Линейные дискриминантные функции Фишера имеют вид:

-14,4706 + (-2,7947) x₁ + 0,4093 x₂ + (-1,2448) x₃ + (-17,5669) x₄ + (-2,1187) x₅ + (-10,2281) x₆ + (-6,0799) x₇ + (-60,1411) x₈ + (-5,7662) x₉;

;

-78,1598 + (-15,4247) x₁ + (-8,2467) x₂ + (-8,7601) x₃ + 26,8928 x₄ + 8,2872 x₅ + 21,4542 x₆ + 6,2457 x₇ + 50,5305 x₈ + (-22,0775) x₉;

;

;

;

;

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

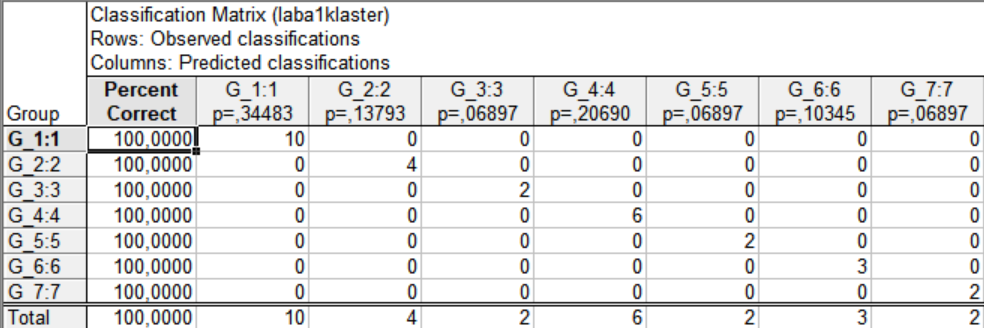


Таблица 13 – Результаты классификации объектов обучающих выборок

Вывод:

Как видно из таблицы 13, изменений в первоначальном составе классов не произошло: к первому классу по-прежнему относятся 10 объектов, ко второму – 4, к третьему – 2, к четвертому – 6, к пятому – 2, к шестому – 3, к седьмому – 2. Процент корректной классификации составил 100%. Это свидетельствует о хорошей дискриминации объектов обучающих выборок на основе дискриминантных функций (1) - (7).

Неправильные классификации объектов обучающих выборок помечаются «звездочкой» (в рассматриваемом примере такие ситуации не встречаются).

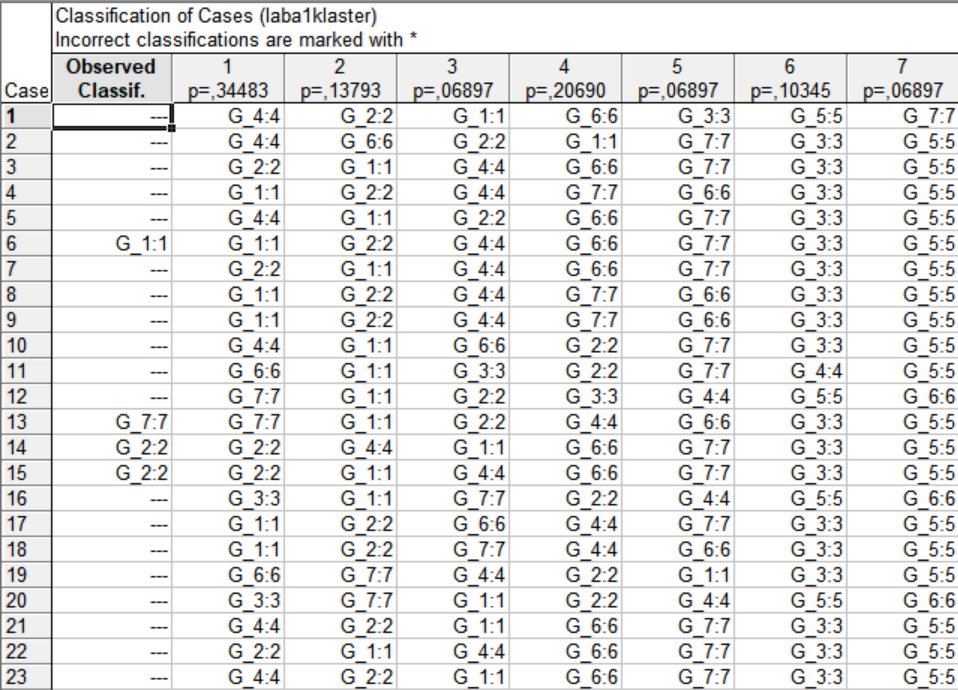
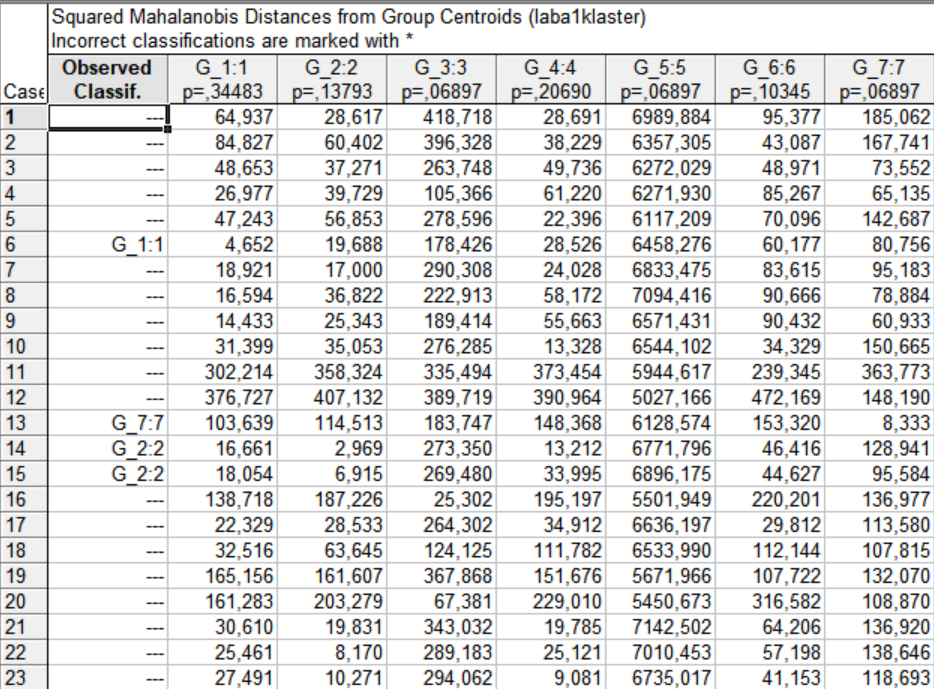


Таблица 14 – Результаты классификации с помощью функций (1) – (7)

Далее рассчитываются квадраты расстояния Махаланобиса от объектов до центров каждого из классов. Результаты расчетов представлены на таблице 15. Объект следует отнести к тому классу, расстояние до которого наименьшее. Так, например, первый регион следует отнести к пятому классу, поскольку расстояние от этого объекта до центра пятого класса наименьшее.



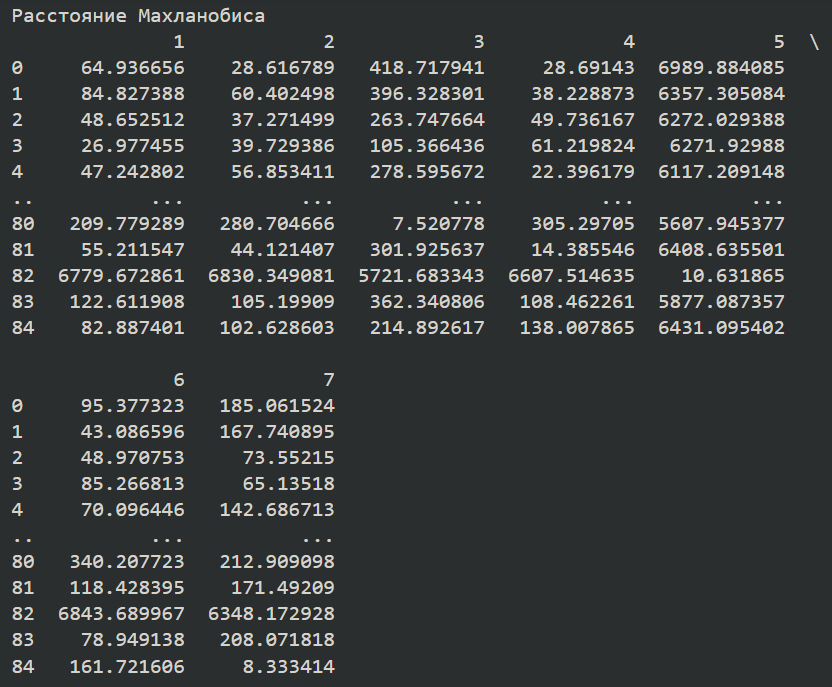
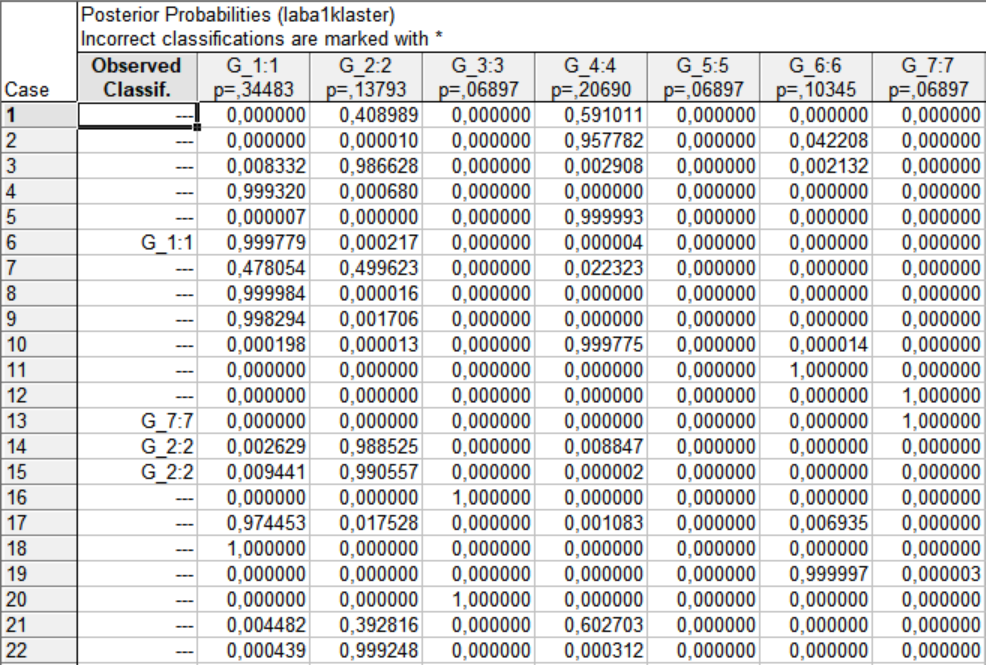


Таблица 15 – Расстояния до центров классов

Также рассчитываются апостериорные вероятности классификации. Результаты представлены в таблице 16. Объект следует отнести к тому классу, апостериорная вероятность для которого наибольшая.



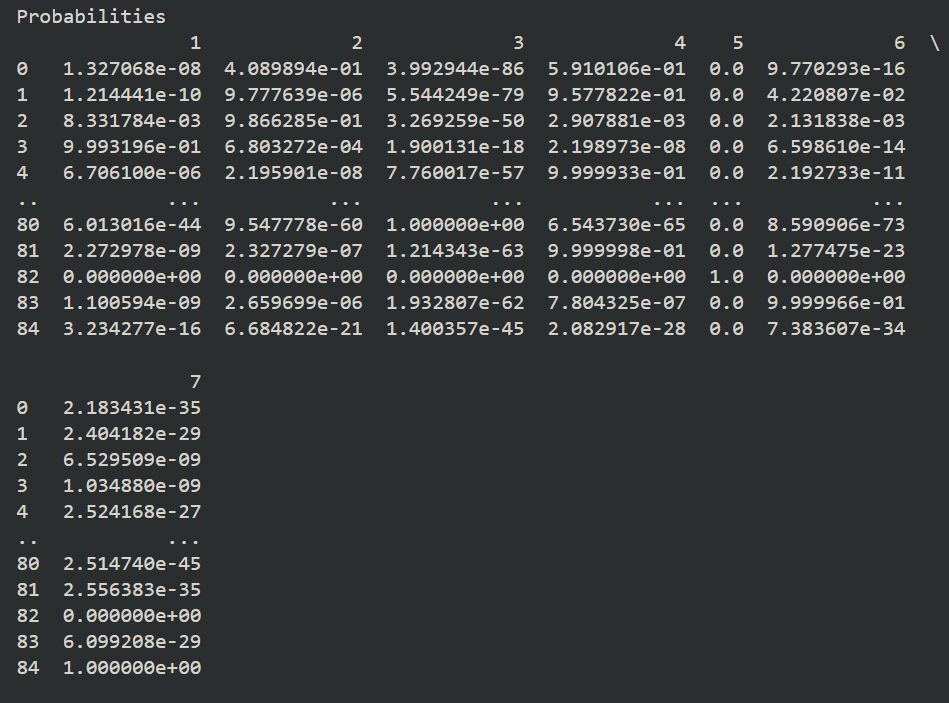


Таблица 16 – Апостериорные вероятности классификации

На основании таблиц, представленных на рисунках 7-9, субъекты, не вошедшие в обучающие выборки, можно классифицировать следующим образом:

|  |  |  |
| --- | --- | --- |
| **Номер**  **кластера** | **Кол-во**  **объектов** | **Состав класса** |
| кластер 1 {S11} | 22 | Брянская область Волгоградская область Кабардино-Балкарская Республика Калужская область Красноярский край Ленинградская область Липецкая область Московская область Пермский край Псковская область Республика Крым Республика Марий Эл Республика Хакасия Ростовская область Рязанская область Самарская область Саратовская область Севастополь Ставропольский край Удмуртская Республика Ульяновская область Челябинская область |
| кластер2 {S12} | 14 | Алтайский край Архангельская область без автономного округа Владимирская область Вологодская область Иркутская область Камчатский край Костромская область Новосибирская область Орловская область Республика Бурятия Республика Карелия Республика Коми Свердловская область Томская область |
| кластер3  {S13} | 10 | Астраханская область Калининградская область Кемеровская область Приморский край Республика Адыгея Республика Дагестан Республика Ингушетия Республика Калмыкия Республика Северная Осетия – Алания Чеченская Республика |
| кластер4 {S14} | 18 | Амурская область Белгородская область Воронежская область Кировская область Краснодарский край Курская область Магаданская область Нижегородская область Новгородская область Оренбургская область Пензенская область Республика Башкортостан Республика Тыва Смоленская область Тамбовская область Тверская область Тюменская область без автономных округов Чувашская Республика |
| кластер5 {S15} | 2 | Ненецкий автономный округ Чукотский автономный округ |
| кластер6 {S16} | 12 | Еврейская автономная область Карачаево-Черкесская Республика Курганская область Москва Мурманская область Республика Саха (Якутия) Санкт-Петербург Сахалинская область Тульская область Хабаровский край Ханты-Мансийский автономный округ – Югра Ямало-Ненецкий автономный округ |
| кластер7 {S17} | 7 | Забайкальский край Ивановская область Омская область Республика Алтай Республика Мордовия Республика Татарстан Ярославская область |

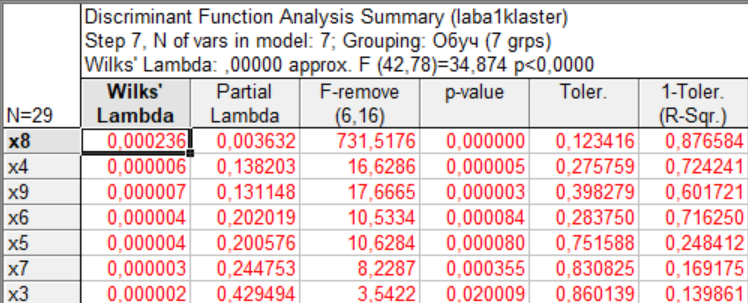
Таблица 17 – Классификация субъектов

Точность: 100%

Выборка была составлена по методу Уорда, совпадение классификации с исходной составило 82,35%.

# **Пошаговый дискриминантный анализ. Пошаговое включение**

Аналогично проведем пошаговый дискриминантный анализ с пошаговым включением. Результаты процедуры представлены в таблице 18:



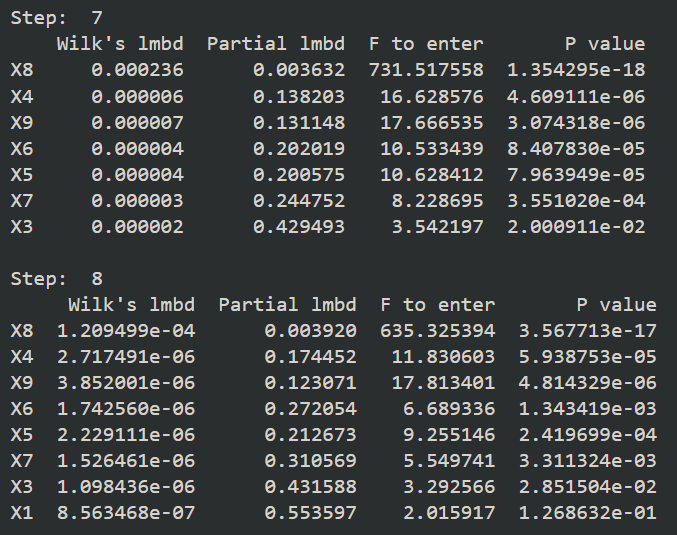
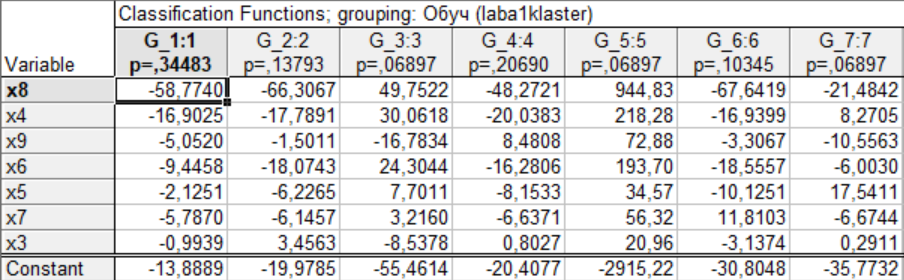


Таблица 18 – Результаты пошагового дискриминантного анализа

Рассчитаем коэффициенты линейных дискриминантных функций Фишера. Результаты представлены в таблице 19:



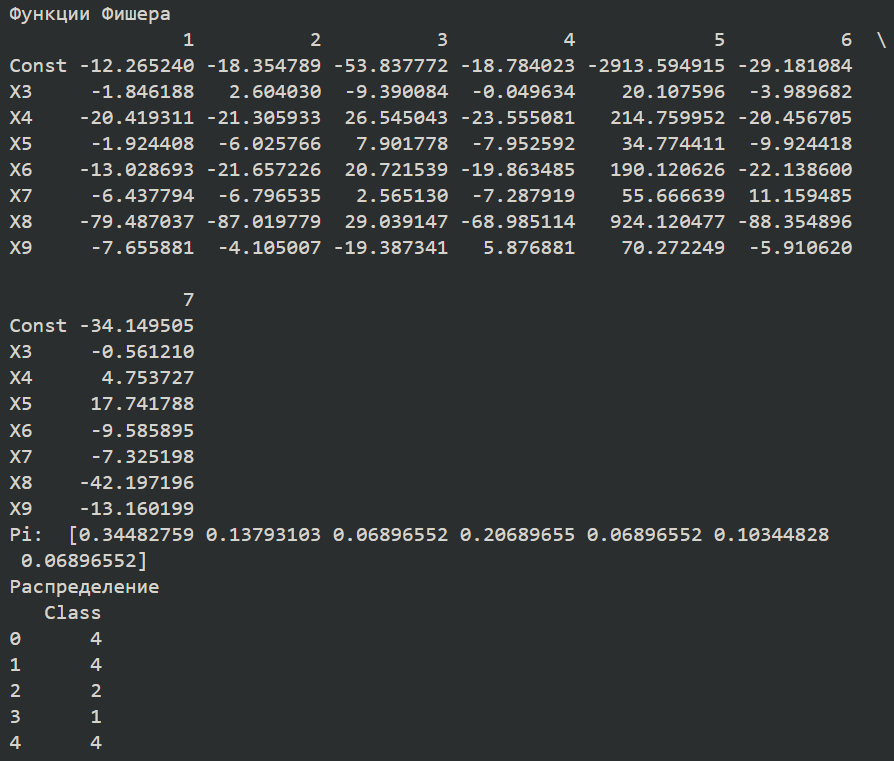


Таблица 19 – Коэффициенты линейных дискриминантных функций Фишера

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.

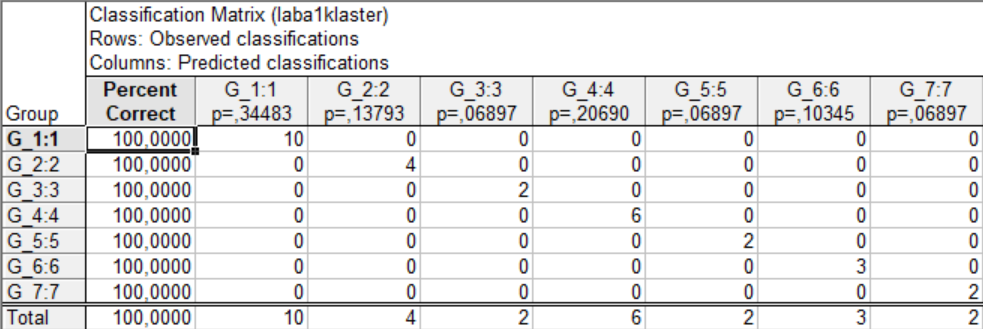


Таблица 20 – Результаты классификации объектов обучающих выборок

ВЫВОД

Как видно из таблицы, изменений в первоначальном составе классов не произошло: к первому классу по-прежнему относятся 10 объектов, ко второму – 4, к третьему – 2, к четвертому – 6, к пятому – 2, к шестому – 3, к седьмому – 2. Процент корректной классификации составил 100%. Это свидетельствует о хорошей дискриминации объектов обучающих выборок на основе дискриминантных функций (1) - (7).

Неправильные классификации объектов обучающих выборок помечаются «звездочкой» (в рассматриваемом примере такие ситуации не встречаются).

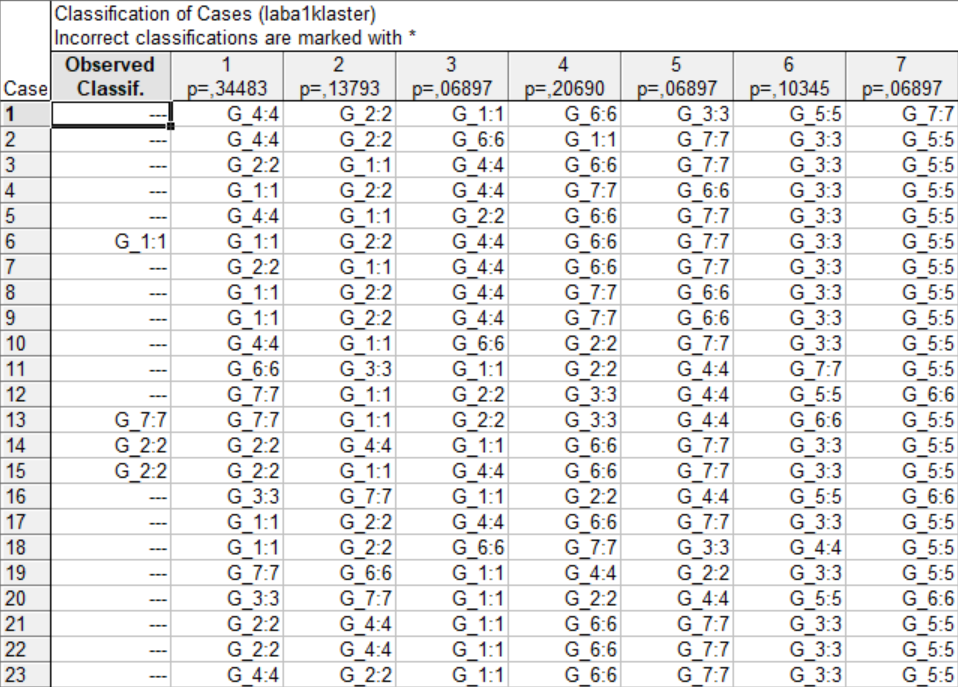


Таблица 21 – Результаты классификации с помощью функций (1) – (7)

В первой графе таблицы указаны номера объектов; во второй графе – номера обучающих выборок (прочерками отмечаются объекты, не вошедшие в обучающие выборки); во третьих-восьмых столбцах – результаты классификации с помощью дискриминантных функций (1) – (7).

Далее рассчитываются квадраты расстояния Махаланобиса от объектов до центров каждого из классов. Результаты расчетов представлены на рис. 9. Объект следует отнести к тому классу, расстояние до которого наименьшее.

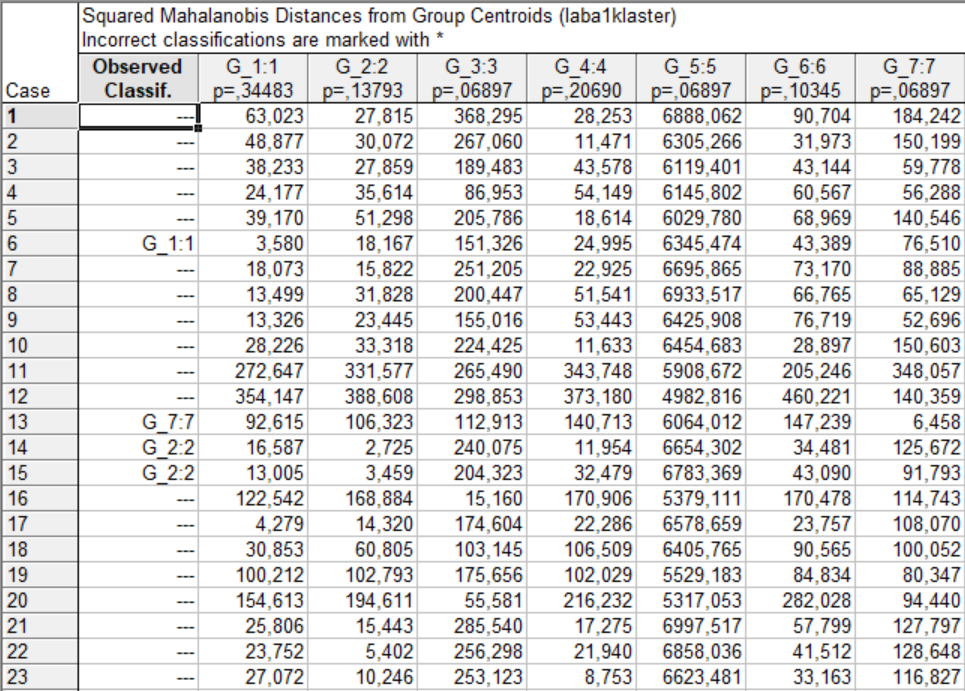
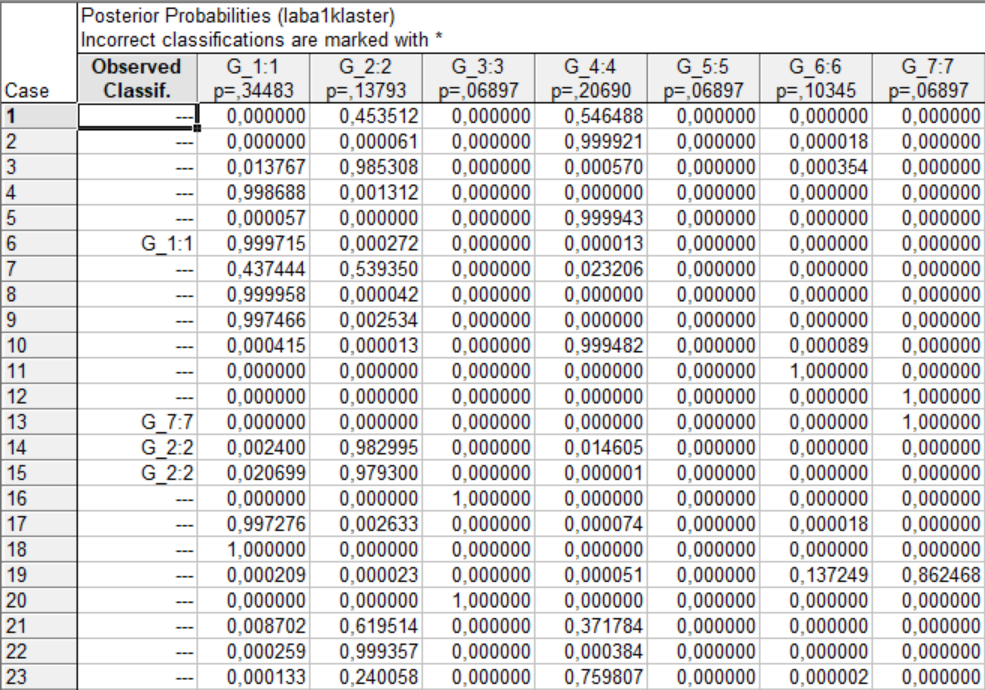


Таблица 22 – Расстояния до центров классов

Также рассчитываются апостериорные вероятности классификации. Результаты представлены на рис. 10. Объект следует отнести к тому классу, апостериорная вероятность для которого наибольшая.



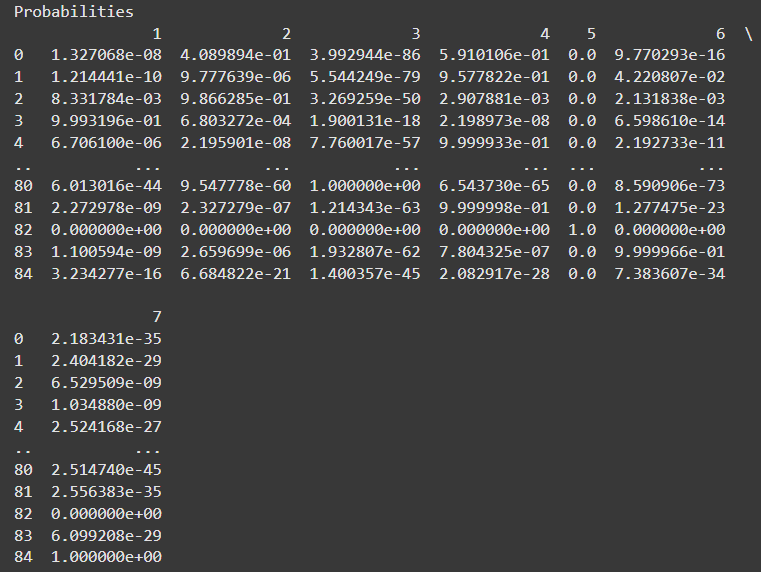


Таблица 23 – Апостериорные вероятности классификации

На основании таблиц, представленных на рисунках 21-23, субъекты, не вошедшие в обучающие выборки, можно классифицировать следующим образом:

|  |  |  |
| --- | --- | --- |
| **Номер**  **кластера** | **Кол-во**  **объектов** | **Состав класса** |
| кластер 1 {S11} | 24 | Астраханская область Брянская область Волгоградская область Вологодская область Калужская область Камчатский край Курганская область Ленинградская область Липецкая область Новосибирская область Омская область Пермский край Республика Бурятия Республика Крым Республика Мордовия Республика Татарстан Рязанская область Самарская область Свердловская область Ставропольский край Тульская область Удмуртская Республика Ульяновская область Челябинская область |
| кластер2 {S12} | 13 | Архангельская область без автономного округа Владимирская область Иркутская область Кабардино-Балкарская Республика Кировская область Костромская область Нижегородская область Орловская область Псковская область Республика Карелия Республика Коми Саратовская область Томская область |
| кластер3  {S13} | 12 | Калининградская область Кемеровская область Москва Республика Адыгея Республика Алтай Республика Дагестан Республика Ингушетия Республика Калмыкия Республика Северная Осетия – Алания Республика Тыва Севастополь Чеченская Республика |
| кластер4 {S14} | 17 | Алтайский край Амурская область Белгородская область Воронежская область Краснодарский край Курская область Магаданская область Новгородская область Оренбургская область Пензенская область Республика Башкортостан Республика Марий Эл Смоленская область Тамбовская область Тверская область Тюменская область без автономных округов Чувашская Республика |
| кластер5 {S15} | 2 | Ненецкий автономный округ Чукотский автономный округ |
| кластер6 {S16} | 12 | Еврейская автономная область Красноярский край Московская область Мурманская область Приморский край Республика Саха (Якутия) Ростовская область Санкт-Петербург Сахалинская область Хабаровский край Ханты-Мансийский автономный округ – Югра Ямало-Ненецкий автономный округ |
| кластер7 {S17} | 5 | Забайкальский край Ивановская область Карачаево-Черкесская Республика Республика Хакасия Ярославская область |

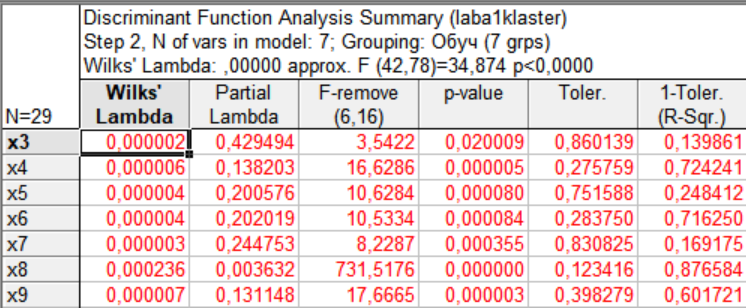
Таблица 24 – Классификация субъектов

Точность: 100%

Выборка была составлена по k-средних, совпадение классификации с исходной составило: 82,35%.

# **Пошаговый дискриминантный анализ. Пошаговое исключение**

Аналогично проведем пошаговый дискриминантный анализ с пошаговым исключением. Результаты процедуры представлены в таблица 16:



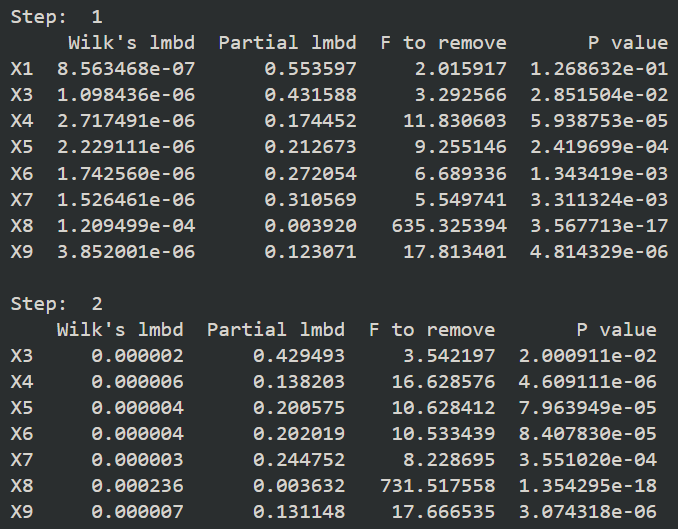


Таблица 24 – Результаты пошагового дискриминантного анализа

Результат соответствует результату с пошаговым включением.

