Обучающая выборка:





Т.к. значение лямды Уилкса равное 0,00867 близко к нулю, а по значению статистики F-критерия равного 9,7372 с уровнем значимости p<0,0021 (т.е. нулевая гипотеза о том, что наблюдения принадлежат одному классу отвергается), то дискриминантный анализ возможен и данная классификация обучающих выборок корректна.   
  
Значение толерантности чем ближе к нулю, тем переменная менее информативная и если ее значение меньше порогового значения (обычно < 0,01), то ее исключают из дискриминантной модели. Заметим, что толерантность выше у переменных **Расх\_на\_здрав** и **Смертность**, следовательно эти переменные более информативны.

С другой стороны, самое большое значение лямба уилкса у переменной **ВНП \_на\_ душу\_ насел**, а после нее у переменной **Кол\_чел\_на\_1врача**, значит следует оставить эти переменные в процедуре дискриминации. Значение частной лямбды наоборот самое маленькое у переменной **ВНП \_на\_ душу\_ насел**, значит эта переменная несет наибольший вклад в общую дискриминацию. А после нее наибольший вклад в общую дискриминацию несет переменная **Кол\_чел\_на\_1врача.** Поэтому на данном этапе дискриминации можно заключить, что уровень ВВП и количество человек на 1 врача являются главными переменными (они выделены красным), которые позволяют производить дискриминацию по уровню здравохранения стран.

Для получения дальнейших результатов о природе дискриминации проведем канонический анализ.

Таблица результатов с пошаговым критерием для дискриминантных функций (канонических корней):

Первый столбец-количество удаленных корней (дискриминантных функций). Первая строка дает критерий значимости для всех корней. Т.к. уровень значимости p меньше чем 0,05, то можно сделать вывод, что есть хотя бы один канонический корень, который является статистически значимым. Вторая строка характеризует значимость значимость корней, оставшихся после удаления первого корня. Т.к. p больше чем 0,05, то это означает, что нет канонических корней, которые являлись бы статистически значимыми. Т.е. один корень (дискриминантная функция) статистически значима.

Стандартизированные переменные дискриминантных функций характеризуют направление и вклад переменных в значения дискриминантных функций.

*Стандартизованные коэффициенты полезно применять для уменьшения размерности исходного признакового пространства переменных. Если абсолютная величина коэффициента для данной переменной для всех дискриминантных функций мала, то эту переменную можно исключить, тем самым сократив число переменных.*



Из таблицы видно, что в дискриминантную функцию1 (Кор. 1) наибольший вклад вносят переменные **Кол\_чел\_на\_1врача** и **ВВП \_на\_ душу\_ насел**;

в дискриминантную функцию2 (Кор. 2) наибольший вклад вносит переменная **Расх\_на\_здрав** .

В таблице также приведены собственные значения для каждой дискриминантной функции и кумулятивная доля объясненной дисперсиию накопленной каждой функцией. Из таблицы следует, что функция 1 ответственна за 99,9 % объясненной дисперсии, т.е. 99,9 % всей дискриминирующей мощности объясняется этой функцией. Поэтому она самая важная. И только за 0,1% объясненной дисперсииответственна функция 2.



Данная таблица средних для дискриминантных функций позволяет определить группы, лучше всего идентифицируемые конкретной дискриминантной функцией. Функция идентифицирует в основном группы "низкий" и "высокий", т.к. им соответствуют наибольшие значения функции 1. Функция 2 - группы "удовлетворительный" и "высокий" уровни здравохранения.

Результаты разделения наблюдений по группам рассмотрим на диаграмме рассеивания.



При этом группы "низкий" и "высокий" наиболее отдалены друг от друга, т.к. показателям в этих группах соответствуют большие значения корня 1 (Кор.1), т.е. дискриминантная функция 1 наилучшим образом дискриминирует показатели между этими группами.



Из классификационной матрицы можно сделать вывод, что объекты (страны) были правильно отнесены экспертным способом к выделенным группам.

Если вдруг страны неправильно отнесены к соответствующим группам, то это можно увидеть в таблице классификации случаев:



Как видно из таблицы нет объектов помеченных \*, т.е. нет некорректно

отнесенных стран к соответствующим группам. Это подтверждает и ниже приведенная таблица квадратов растояний Махаланобиса до центров тяжести соответствующих групп. Отнесение всех объектов (стран) в определенную группу развития здравохранения является правильным (корректным), т.к. расстояния Махаланобиса от объектов до центров групп, куда они были отнесены экспертом, значительно ниже, чем от них до центров других групп, более того апостериорные вероятности попадания в свои группы выше критических значений.



Т.к. наблюдения обучающих выборок правильно отнесены к соответствующим группам, то на их основе можно проводить классификацию тех объектов (стран-Украина и Молдавия), которые подлежат группировке.

Проведем классификацию объектов (стран-Украина и Молдавия) двумя способами и сравним результаты:

1 способ: на основе апостериорных вероятностях принадлежности объекта к i-ой группе, т.е. страна приписывается к той группе уровня здравохранения, которой соответствует наибольшая апостериорная вероятность классификации. Отметим, что прежде чем вычислять апостериорные вероятности необходимо добавить в таблицу скорректированных данных новые случаи.



Получим:



Из таблицы видно, что Молдавия отнесена к группе "низкий", а Украина к группе "удовлетворительный" уровень здравохранения, т.к. соответствующие апостериорные вероятности классификации наибольшие.

2 способ: на основе классификационных функций для каждого класса.



На основе данных в таблице выпишем классификационные функции:

*высокий*=-1526,02+2,35\**Кол\_чел\_на\_1врача* +1,98\**Расх\_на\_здрав* +0,20\**ВВП \_на\_ душу\_ насел*+46,93\**Смертность*

*удовлетворительный*=-921,497+1,834\**Кол\_чел\_на\_1врача* +1,718\**Расх\_на\_здрав* +0,153\**ВВП \_на\_ душу\_ насел*+36,637\**Смертность*

*низкий*=-576,414+1,455\**Кол\_чел\_на\_1врача* +1,455\**Расх\_на\_здрав* +0,116\**ВВП \_на\_ душу\_ насел*+29,066\**Смертность*

Новые случаи будут относится к тому классу, для которого классификационное значение будет максимальное.

Для Молдавии и Украины, как видно из ниже приведенных расчетов Молдавия отнесена к группе "низкий", а Украина к группе "удовлетворительный" уровень здравохранения, т.к. значения соответствующих классификационных функций наибольшие.

|  |  |  |
| --- | --- | --- |
|  | Молдавия | Украина |
| Кол\_чел\_на\_1врача | 251 | 224 |
| Расх\_на\_здрав | 143 | 131 |
| ВВП\_на\_душу\_насел | 2500 | 3850 |
| Смертность | 12,6 | 16,4 |
| Конст-та | -576,414 | -1526,02 |
| низкий | **654,2842168** | 865,1819691 |
| высокий | 446,1991559 | 811,6443113 |
| удовлетвор. | 627,6967153 | **902,8690472** |

Как видно, результаты классификации для объектов-стран Молдавия и Украина двумя способами совпадают.