

**Кислицына Валентина Васильевна,**

*к.э.н., доцент,*

*ФГБОУ ВО «Вятский государственный университет»*

*г. Киров, Россия*

## **ДИСКРИМИНАНТНЫЙ АНАЛИЗ КАК ИНСТРУМЕНТ КЛАССИФИКАЦИИ СУБЪЕКТОВ ПРЕДПРИНИМАТЕЛЬСТВА**

**Аннотация.** Моделирование, как универсальный научный инструмент, приобретает все большее значение для современной системы менеджмента. В последнее время довольно широко стали применяться статистические методы анализа данных, в частности дискриминантный анализ, основной задачей которого является выявление дискриминирующих переменных и определение принадлежности объектов анализа к той или иной группе. Спектр применения дискриминантного анализа довольно широк, в частности он позволяет классифицировать субъекты по основным характеристикам их хозяйственно-финансовой деятельности.

**Ключевые слова:** дискриминантный анализ, модель, Statistica, квадраты Махаланобиса, апостериорные вероятности.

***Kislitsyna Valentina Vasilievna,***

*candidate of economic Sciences, Associate professor,*

*Vyatka state University,*

*Kirov, Russia*

## **DISCRIMINANT ANALYSIS AS A TOOL CLASSIFICATION OF BUSINESS ENTITIES**

**Abstract.** Modeling, as a versatile research tool, is becoming increasingly important for the modern system of management. Recently quite widely applied statistical methods of data analysis, in particular discriminant analysis, whose main task is the identification of discrimination variables and the identification of objects of the analysis to one group or another. The range of applications of discriminant analysis is quite broad, in particular it allows to classify business entities according to the main characteristics of their economic and financial activities.

**Key words:** discriminant analysis model, Statistica, squares Mahalanobis, a posteriori probability.

Необходимость моделирования в системе как традиционного, так и инновационного менеджмента обусловлена неоднозначностью многих

управленческих процессов, сложностью организационных структур, слабой формализацией поступающей информации, нестабильностью условий деятельности объектов и субъектов управления. Использование моделей позволяет определить и уточнить характеристики объектов и рационализировать способы принятия управленческих решений. Моделирование, как универсальный научный инструмент, приобретает все большее значение в проектно-аналитической деятельности, в том числе и в сфере бизнеса, что позволяет прорабатывать и просчитывать решения, направленные на развитие предприятий.

В последнее время довольно широко в данной сфере стали применять статистические методы анализа данных, используемые в том случаях, когда необходимо получить обоснованные суждения о группе объектов, характеризующихся определенной неоднородностью показателей. Одним из таких методов является дискриминантный анализ, который так же, как и кластерный анализ, относится к методам многомерной классификации и используется для принятия решения о том, какие переменные различают (дискриминируют) анализируемые совокупности, но при этом базируется на иных предпосылках. Основное отличие заключается в том, что в ходе дискриминантного анализа формулируется правило, по которому новые единицы совокупности относятся к одному из уже существующих множеств (классов). Основанием для отнесения каждой единицы совокупности к определенному множеству служит величина дискриминантной функции, рассчитанная по соответствующим значениям дискриминантных переменных [1].

Таким образом, основная идея дискриминантного анализа заключается в том, чтобы определить, отличаются ли совокупности по среднему какой-либо переменной; затем переменная/комбинация переменных используется для того, чтобы определить принадлежность отдельных объектов к определенной группе. Аппарат дискриминантного анализа разрабатывался многими учеными-специалистами, начиная с конца 50-х годов XX в. Дискриминантным анализом,

как и другими методами многомерной статистики, занимались П.Ч. Махаланобис, Р. Фишер, Г. Хотеллинг и другие ученые.

В нашем случае дискриминантный анализ проводился на примере предприятий-субъектов малого предпринимательства, разделенных на классы в зависимости от существующих тенденций изменения финансового положения: 1-ый класс соответствовал тенденции улучшения, 2-ой – тенденции нормального развития, 3-ий – тенденции ухудшения (таблица 1).

Таблица 1 – Показатели деятельности предприятий анализируемой выборки

Предприятие	Класс	Коэффициент финансового риска	Коэффициент абсолютной ликвидности	Коэффициент рентабельности и капитала	Коэффициент рентабельности и текущих затрат	Коэффициент рентабельности и продаж
А	1	0,65	0,37	0,26	0,30	0,35
Б	2	0,83	0,30	0,12	0,17	0,19
В	3	1,00	0,22	0,06	0,08	0,07
Г	1	0,55	0,42	0,21	0,28	0,28
Д	3	1,00	0,12	0,06	0,07	0,12
Е	2	0,92	0,32	0,14	0,22	0,22
Ж	3	0,05	0,25	0,04	0,03	0,09
З	1	1,00	0,38	0,17	0,33	0,37
И	2	0,87	0,31	0,10	0,15	0,17
К	2	0,90	0,30	0,14	0,19	0,20

Тестируемая выборка включает пять предприятий той же отрасли (таблица 2). Нашей задачей является отнесение предприятий тестируемой выборки к определенному классу по аналогии с предприятиями анализируемой выборки.

Таблица 2 – Показатели деятельности предприятий тестируемой выборки

Предприятие	Класс	Коэффициент финансового риска	Коэффициент абсолютной ликвидности	Коэффициент рентабельности и капитала	Коэффициент рентабельности и текущих затрат	Коэффициент рентабельности и продаж
Л	-	0,75	0,46	0,19	0,31	0,34
М	-	0,82	0,31	0,12	0,23	0,17
Н	-	0,93	0,30	0,14	0,18	0,22
О	-	0,63	0,54	0,22	0,29	0,38
П	-	1,00	0,22	0,09	0,07	0,12

Для проведения дискриминантного анализа воспользуемся пакетом прикладных программ Statistica с предварительной стандартизацией данных (вкладка «Данные – Стандартизировать»). Выберем стандартный метод анализа, когда все переменные (предикторы) одновременно включаются в формируемую модель. Установим показатель «Толерантность» – степень нескоррелированности признака со всеми остальными, определяющий количество информационных переменных, на значении 0,01 в связи с небольшим количеством переменных, используемых для анализа. Принимаем число шагов, заданное программой, равным пяти. Вывод результатов отражен в таблице 3.

Таблица 3 – Итоги анализа дискриминантных функций (значения в модели)

Итоги анализа дискриминантн. функций (Таблица данных 1): Шаг 5, Переменных в модели – 5; Группир.: Класс (3 гр.) Лямбда Уилкса: 0,00894 прилб. F (10,6)=5,7461 p< 0,0221						
	Уилкса- Лямбда	Частная- Лямбда	F-искл.- (2,3)	p-уров.	Толер.	1-толер.- (R-кв.)
Коэффициент рентабельности текущих затрат	0,011846	0,754604	0,487797	0,655509	0,080782	0,919218
Коэффициент финансового риска	0,015577	0,573875	1,113809	0,434737	0,080395	0,919605
Коэффициент абсолютной ликвидности	0,023840	0,374954	2,500490	0,229598	0,111372	0,888628
Коэффициент рентабельности продаж	0,016201	0,551765	1,218550	0,409855	0,127934	0,872066
Коэффициент рентабельности капитала	0,014576	0,613278	0,945872	0,480271	0,496735	0,503265

Величина лямбды Уилкса, значения которой лежат в пределах от 0 (абсолютно точная классификация) до 1 (абсолютно ошибочная классификация), равная в нашем случае 0,00894, свидетельствует о хорошей дискриминации, то есть приемлемом разделении исследуемых объектов по выбранным признакам/дискриминантным переменным. О достоверности модели свидетельствует также величина значимости критерия  $p < 0,0221$ , поскольку модель считается достоверной при принятом в большинстве статистических программ  $p < 0,05$ . Коэффициент детерминации имеет значение больше 0,75 для всех переменных, включенных в модель, кроме

«Коэффициента рентабельности капитала» (0,503). Но, исходя из общих показателей достоверности модели, принимаем решение оставить данную переменную.

Определим принадлежность исследуемых объектов к группам с помощью кнопки «Классификация». Матрица классификации (таблица 4) выводит число наблюдений в каждой группе. В данном случае правильная классификация объектов составляет 100% по каждой группе, что позволяет сделать вывод об отсутствии неверных значений и приемлемости модели для дальнейшего анализа.

Таблица 4 – Матрица классификации

Матрица классификации (Таблица данных 1) Строки: наблюдаемые классы Столбцы: предсказанные классы				
	Процент - правиль.	G_1:1 - p=,30000	G_2:2 - p=,40000	G_3:3 - p=,30000
G_1:1	100,0000	3	0	0
G_2:2	100,0000	0	4	0
G_3:3	100,0000	0	0	3
Всего	100,0000	3	4	3

Об этом же свидетельствуют данные таблицы, выводящей квадраты расстояний Махаланобиса, которые являются мерой расстояний между анализируемыми группами (таблица 5).

Таблица 5 – Квадраты расстояний Махаланобиса

Квадраты расстояний Махаланобиса до центров (Таблица данных 1) Неправильные классификации отмечены *				
	Наблюд. - Класс.	G_1:1 - p=,30000	G_2:2 - p=,40000	G_3:3 - p=,30000
А	G_1:1	4,6064	98,0694	356,0675
Б	G_2:2	83,6436	0,3725	77,2657
В	G_3:3	288,8162	70,7135	4,4963
Г	G_1:1	4,6155	54,8954	269,1108
Д	G_3:3	361,1724	114,9205	4,4744
Е	G_2:2	58,9919	3,3242	105,4183
Ж	G_3:3	307,1145	87,3460	4,6373
З	G_1:1	4,5138	84,9189	332,0525
И	G_2:2	91,2979	3,5510	83,2512
К	G_2:2	72,5869	0,4086	87,5505

Мы видим, например, что для предприятия А, отнесенного к первому классу развития, квадрат Махаланобиса, равный 4,6064, наименьший, следовательно, классификация, проведенная по построенной модели, является верной. То же самое можно сказать о других наблюдениях. Аналогичные результаты дает таблица апостериорных, то есть полученных после опыта, вероятностей (таблица 6).

Таблица 6 – Апостериорные вероятности

Апостериорные вероятности (Таблица данных 1)				
Неправильные классификации отмечены *				
	Наблюд. - Класс.	G_1:1 - p=,30000	G_2:2 - p=,40000	G_3:3 - p=,30000
А	G_1:1	1,000000	0,000000	0,000000
Б	G_2:2	0,000000	1,000000	0,000000
В	G_3:3	0,000000	0,000000	1,000000
Г	G_1:1	1,000000	0,000000	0,000000
Д	G_3:3	0,000000	0,000000	1,000000
Е	G_2:2	0,000000	1,000000	0,000000
Ж	G_3:3	0,000000	0,000000	1,000000
З	G_1:1	1,000000	0,000000	0,000000
И	G_2:2	0,000000	1,000000	0,000000
К	G_2:2	0,000000	1,000000	0,000000

Для визуальной оценки статистических различий между выбранными классами проведем канонический анализ. Результаты разделения по группам отражены графически – см. рисунок.

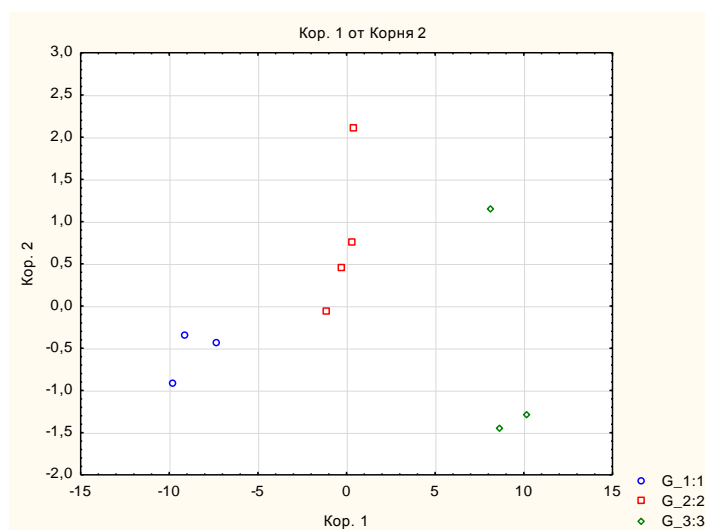


Рисунок 1 – Диаграмма рассеяния канонических значений

Рисунок показывает, что наблюдения (предприятия), отнесенные к одному и тому же классу, локализованы в определенных плоскостях диаграммы с наблюдаемым разбросом значений, что также свидетельствует о значимом различии между группами. Наиболее различаются между собой первый и третий классы, соответствующие тенденциям улучшения и ухудшения состояния предприятий соответственно. Таким образом, анализ построенной модели дает основание сделать вывод о возможности ее использования для классификации объектов проверяемой совокупности. Для этого воспользуемся расчетом апостериорных вероятностей – таблица 7.

Таблица 7 – Апостериорные вероятности для выборки проверяемых объектов

Апостериорные вероятности (Таблица данных 1)				
Неправильные классификации отмечены *				
	Наблюд. - Класс.	G_1:1 - p=,30000	G_2:2 - p=,40000	G_3:3 - p=,30000
...	...	...	...	...
К	G_2:2	0,000000	1,000000	0,000000
Л	---	0,272623	0,727377	0,000000
М	---	0,000045	0,999955	0,000000
Н	---	0,003325	0,996675	0,000000
О	---	0,936852	0,063148	0,000000
П	---	0,000133	0,999867	0,000000

Данные таблицы апостериорных вероятностей, выполненных для новых наблюдений, позволяют сделать вывод об их принадлежности к определенному классу. Так, предприятия Л, М, Н и П отнесены ко второму, предприятие О – к первому классу развития. Аналогичные выводы получаем после расчета квадратов расстояний Махаланобиса, где для предприятий Л, М, Н и П наименьшее расстояние наблюдается для второй, для О – для первой группы (таблица 8).

Таблица 8 – Квадраты расстояний Махаланобиса для выборки проверяемых объектов

Квадраты расстояний Махаланобиса до центров (Таблица данных 1)				
Неправильные классификации отмечены *				
	Наблюд. - Класс.	G_1:1 - p=,30000	G_2:2 - p=,40000	G_3:3 - p=,30000
...	...	...	...	...

К	G_2:2	72,5869	0,4086	87,5505
Л	---	31,5099	30,1226	207,1188
М	---	37,8025	18,3572	174,5328
Н	---	39,5051	28,6743	197,3724
О	---	31,2625	37,2319	223,2656
П	---	46,0591	28,7855	192,1027

Отнесение предприятий тестируемой выборки к той или иной группе развития позволяет разработать управленческие решения для каждого из них. Для предприятия О основной задачей является постоянный мониторинг конкурентной ситуации и оценка уровня собственной конкурентоспособности. Предприятиям второй группы необходимо усилить существующие конкурентные преимущества (Л – специализация, М – низкий уровень цен и высокая степень использования производственных мощностей, Н и П – выгодное месторасположение) или осуществить поиск новых с тем, чтобы обеспечить себе устойчивое положение на рынке в краткосрочной перспективе и стратегическое развитие.

Выполнение дискриминантного анализа может сопровождаться рядом проблем. Во-первых, возникает вопрос о выборе дискриминантных переменных и их количестве. Использование большего набора переменных позволяет выбрать наиболее значимые из них, но усложняет процесс сбора и обработки необходимой для этого информации. Во-вторых, необходимо определиться с выбором метода анализа. При использовании стандартного метода может оказаться, что отдельные переменные являются малозначимыми, что, в целом, ухудшает качество построенной модели. Пошаговый анализ позволяет включить в исследование те переменные, которые наилучшим образом разделяют совокупности между собой, но возникает проблема определения параметров для проведения анализа, в частности, значения F–включить/исключить для переменных и выбора их статистической значимости для получения максимальной дискриминации – меры вклада переменных в предсказание членства в совокупности. В-третьих, если множества в анализируемой выборке расположены близко друг от друга, то есть характеризуются схожестью по используемым переменным, то программа



может вывести неверную классификацию для новых наблюдений. Ошибка будет тем больше, чем сильнее отличаются значения переменных тестируемой выборки от значений анализируемой выборки. С другой стороны, если имеется большая изменчивость значений переменных в совокупности с высокими средними в отдельных переменных, то выявленные средние значения не будут являться надежными при проведении анализа. В-четвертых, дискриминантный анализ, как и многие многомерные методы, основан на построении линейных комбинаций, и предполагается, что анализируемые переменные представляют выборку из многомерного нормального распределения.

Но, в целом, дискриминантный анализ как совокупность связанных статистических процедур, позволяет ответить на такие основные вопросы, как: «Возможно ли отличить одну совокупность наблюдений от другой?», «Насколько хорошо выбранные переменные позволяют провести дискриминацию?» и «Какие переменные являются наиболее значимыми?» с тем, чтобы достичь основной цели – группировать наблюдения тестируемой выборки или предсказать будущие события на основании имеющихся данных [2].

#### *СПИСОК ЛИТЕРАТУРЫ*

- 1. Методы дискриминантного анализа [Электронный ресурс]. – Режим доступа: <http://mirznanii.com/a/264966/metody-diskriminantnogo-analiza> (дата обращения 06.04.2017).*
- 2. Основы дискриминантного анализ [Электронный ресурс]. – Режим доступа: [http://masters.donntu.org/2005/kita/kapustina/library/discr\\_an.htm](http://masters.donntu.org/2005/kita/kapustina/library/discr_an.htm) (дата обращения 06.04.2017).*