

А.А. ХАЛАФЯН

кандидат технических наук, доцент, профессор кафедры прикладной математики Кубанского государственного университета

Е.Ю. ПЕЛИПЕНКО

*аспирант кафедры прикладной математики
Кубанского государственного университета*

**Дискриминантный анализ в определении
кредитоспособности предприятий**

Аннотация

В настоящее время актуальна проблема определения риска банкротства предприятия. Решение проблемы возможно путем использования статистических классификационных методов, которые требуют наличия обучающей выборки. В статье предложен метод определения величины риска банкротства любого предприятия, если известны определенные экономические показатели этого предприятия.

Abstract

Now days there are an urgency problem of enterprises non-payment risk. It is possible to solve a problem by using classification statistic methods, that based on instruction sample. In the present work the enterprises non-payment risk method is suggested. This one is based on financial indices patterns, that are known.

Ключевые слова: классификационный анализ, обучающая выборка, дискриминантный анализ, канонический анализ, рентабельность,

платежеспособность, функции классификации, матрица классификации, средние значения, диаграмма рассеяния, апостериорные вероятности.

Keywords: classification analysis, instraction sample, discriminant analysis, canonical analysis, profitability, solvency, classification functions, classification matrix, means, scatterplot, postrior probabilities.

Мировой кризис конца 2008 года значительно усугубил финансовое состояние многих кредитных организаций. Проблема выплат по долговым обязательствам клиентов стала одной из наиболее острых. Целью настоящей работы является разработка модуля анализа данных на языке STATISTICA VISUAL BASIC (SVB) в среде пакета Statistica, который позволит автоматизировать процесс классификации предприятий по уровню риска неплатежеспособности.

Для создания программного модуля за основу был взят дискриминантный анализ, который использует статистические методы классификации с обучением, т.е. когда исследователь обладает обучающей выборкой с известными результатами анализа. В качестве обучающей выборки мы использовали виртуальную клиентскую базу предприятий, создание которой было подробно описано в работе [1], где все множество виртуальных клиентов было сгенерировано по пяти группам неплатежеспособности:

- *1 группа* – предприятия очень высокого уровня риска;
- *2 группа* – предприятия, находящиеся в зоне высокого риска ;
- *3 группа* – предприятия, находящиеся в зоне среднего риска;
- *4 группа* – предприятия, находящиеся в зоне низкого риска;
- *5 группа* – предприятия, находящиеся в зоне очень низкого риска (кредитоспособные предприятия).

В качестве наиболее весомых экономических показателей были взяты 16 параметров предприятий, образующих четыре группы [1]:

1. Рентабельность:

R1 – общая рентабельность; **R2** – рентабельность активов; **R3** – рентабельность собственного капитала; **R4** – рентабельность продукции; **R5** – рентабельность оборотных активов.

2. Ликвидность и платежеспособность:

L1 – быстрый коэффициент ликвидности; **L2** – коэффициент покрытия запасов; **P1** – текущий коэффициент ликвидности.

3. Деловая активность:

A2 – оборачиваемость активов; **A4** – оборачиваемость кредиторской задолженности; **A5** – оборачиваемость дебиторской задолженности; **A6** – оборачиваемость запасов.

4. Финансовая устойчивость:

F1 – коэффициент финансовой зависимости; **F2** – коэффициент автономии собственных средств; **F3** – обеспеченность запасов собственными оборотными средствами; **F4** – индекс постоянного актива.

Используя модуль *Discriminat Analysis* программы Statistica был проведен дискриминантный анализ виртуальной клиентской базы предприятий. Из таблицы результатов дискриминантного анализа следует, что значение лямбды Уилкса (λ) равно 0,0000049. Это говорит о хорошей дискриминации, так как значение λ приняло значение близкое к 0. Чем λ ближе к 1, тем дискриминация хуже.

Исследование итоговой таблицы анализа данных (табл. 1) показало, что присутствие переменной R2 в процедуре дискриминации является наиболее желательным (соответствует наибольшее значение *Wilks Lambda*).

Значение *Partial Lambda* (частной лямбды) характеризует единичный вклад соответствующей переменной в разделительную силу модели. Чем меньше статистика, тем больше вклад в общую

дискриминацию. Из таблицы видно, что переменные R2, F1 и F2 дают значительный вклад, а переменные A2 и A4 вносят в общую дискриминацию меньший вклад. Поэтому на этой стадии исследования можно заключить, что показатель рентабельности активов (**R2**), коэффициент финансовой зависимости (**F1**) и коэффициент автономии собственных средств (**F2**) являются главными переменными, которые позволяют производить дискриминацию между различными группами предприятий.

Таблица 1. Итоговая таблица анализа данных

N=1000	Итоговая таблица дискриминантного анализа данных Число переменных: 16; Группирующие переменные: Классы(5 групп); Лямбда Уилкса: 0,00000 F(64,3838)=1300,3 p<0,0000					
	Лямбда Уилкса	Частная лямбда	Значение F-критерия	Уровень значимости F – критерия	Толерантность	Коэффициент множественной корреляции
L1	0,000006	0,827377	51,1165	0,000000	0,971886	0,028114
L2	0,000006	0,845276	44,8461	0,000000	0,981104	0,018896
P1	0,000006	0,840156	46,6124	0,000000	0,984020	0,015980
F1	0,000008	0,626187	146,2570	0,000000	0,989247	0,010753
F2	0,000007	0,664263	123,8299	0,000000	0,995379	0,004621
F3	0,000006	0,883961	32,1615	0,000000	0,966743	0,033257
F4	0,000006	0,877814	34,1025	0,000000	0,984351	0,015649
R1	0,000006	0,764040	75,6640	0,000000	0,979979	0,020021
R2	0,000013	0,366750	423,0309	0,000000	0,988742	0,011258
R3	0,000005	0,910658	24,0364	0,000000	0,974302	0,025698
R4	0,000006	0,893515	29,1980	0,000000	0,975140	0,024860
R5	0,000006	0,882433	32,6416	0,000000	0,971746	0,028254
A2	0,000005	0,934974	17,0395	0,000000	0,972912	0,027088
A4	0,000005	0,930492	18,3016	0,000000	0,981252	0,018748
A5	0,000006	0,861200	39,4867	0,000000	0,963720	0,036280
A6	0,000006	0,883622	32,2679	0,000000	0,972059	0,027941

Толерантность (*Toler*) является мерой избыточности переменной в модели (чем меньше ее значение, тем избыточнее переменная в

модели, тем меньшую дополнительную информацию несет переменная). Из таблицы видно, что значения толерантности всех показателей достаточно большие, значит можно говорить об их успешном включении в модель.

Для получения дальнейших результатов о природе дискриминации был проведен *канонический анализ*. Для оценки того, как шестнадцать переменных разделяют различные совокупности (группы предприятий), были вычислены четыре дискриминантные функции (число переменных (16) или число совокупностей (5) минус один, в зависимости от того, какое из чисел меньше).

Таблица результатов с пошаговым критерием для дискриминантных функций (канонических корней) (табл. 2) позволила определить какие из корней являются статистически значимыми при помощи последовательного применения критерия значимости.

Roots Removed	Chi-Square Tests with Successive Roots Removed (N)					
	Eigen-value	Canonicl R	Wilks' Lambda	Chi-Sqr.	df	p-level
0	887,2200	0,999437	0,000005	12079,27	64	0,00
1	24,3397	0,980070	0,004381	5368,13	45	0,00
2	6,6115	0,931998	0,111002	2172,93	28	0,00
3	0,1836	0,393839	0,844891	166,61	13	0,00

Таблица 2. Результаты с пошаговым критерием для канонических корней

Удалено корней	Chi квадрат – критерий последовательности удаления корней					
	Собственные значения		лямбда Уилкса	Chi-квадрат	Число степеней свободы	Уровень значимости
0	887,2200	0,999437	0,000005	12079,27	64	0,00
1	24,3397	0,980070	0,004381	5368,13	45	0,00
2	6,6115	0,931998	0,111002	2172,93	28	0,00
3	0,1836	0,393839	0,844891	166,61	13	0,00

Первая строка таблицы дает критерий значимости для всех корней. Так как уровень значимости p меньше чем 0,05, то делаем

вывод, что есть хотя бы один канонический корень, который является статистически значимым. Вторая строка характеризует значимость корней, оставшихся после удаления первого корня. Так как p меньше чем 0,05, среди оставшихся корней есть статистически значимые. Аналогично рассуждая, получаем, что все корни статистически значимы.

Стандартизованные коэффициенты дискриминантных функций характеризуют направление и вклад переменных (экономических показателей) в значения дискриминантных функций. Из табл.3 видно, что в дискриминантную функцию 1 (Root 1) наибольший вклад вносят переменные L3, P1; в дискриминантную функцию 2 (Root 2) – R2, F2; в дискриминантную функцию 3 (Root 3) – R2, F1; в дискриминантную функцию 4 (Root 4) – R1, R2.

Также в таблице приведены собственные значения (*Eigenval*) для каждой дискриминантной функции и кумулятивная доля объясненной дисперсии (*Cum.Prop.*), накопленной каждой функцией. Из таблицы следует, что функция 1 ответственна за 96,61% объясненной дисперсии, т.е. 96,61% всей дискриминирующей мощности определяется этой функцией. Поэтому эта функция наиболее «важна». Функция 2 ответственна за 2,65% объясненной дисперсии. Функция 3 ответственна за 0,72% объясненной дисперсии. Функция 4 ответственна за 0,02% объясненной дисперсии.

Таблица 3. Стандартизованные коэффициенты дискриминантных функций

Переменные	Стандартизованные коэффициенты для канонических переменных			
	Корень 1	Корень 2	Корень 3	Корень 4
L1	-0,2904	0,14070	-0,292229	0,035224
L2	-0,3280	-0,04160	-0,226685	0,159778

P1	-0,3273	0,11009	-0,196083	0,251\9101
F1	-0,2704	-0,30603	-0,476079	-0,340887
F2	-0,2411	0,46264	0,264315	-0,289232
F3	-0,2810	0,09769	0,185699	0,114463
F4	-0,3214	-0,0954	-0,104048	0,006158
R1	-0,2834	0,12333	-0,309402	-0,636344
R2	-0,2572	-0,50746	0,595168	-0,353335
R3	-0,2237	-0,20049	-0,059691	0, 008761
R4	-0,2004	-0,15121	-0,221389	-0,171159
R5	-0,2743	-0,13161	-0,133987	0, 296246
A2	-0,2008	-0,13216	0,100914	0,077736
A4	-0,1011	-0,24877	-0,027481	0,058407
A5	-0,1894	-0,33209	-0,050374	0, 016576
A6	-0,1487	-0,29310	-0,074540	0, 257743
Собственные значения	887,2200	24,33970	6,611511	0,183585
Кумулятивная доля объясненной дисперсии	0,9661	0,99260	0,999800	1,000000

Таблица средних значений для дискриминантных функций (табл. 4) позволила определить группы, лучше всего идентифицируемые конкретной дискриминантной функцией. Функция 1 идентифицирует в основном группы 1 и 5, т.е. «ОТКАЗ» и «КРЕДИТ», так как им соответствуют наибольшие значения функции 1. Функция 2 – группы 4 и 5, т.е. «НИЗКИЙ РИСК» и «КРЕДИТ». Функция 3 – группы 3 и 4, т.е. «НИЗКИЙ РИСК» и «СРЕДНИЙ РИСК». Функция 4 – группы 1 и 2, т.е. «ОТКАЗ» и «ВЫСОКИЙ РИСК».

Таблица 4. Средние значения канонических переменных

Группы	Средние значения канонических переменных
---------------	---

	Корень 1	Корень 2	Корень 3	Корень 4
Отказ	36,3172	-3,65463	-2,57269	-0,416201
Высокий риск	24,3966	-0,88211	1,06452	0,755099
Средний риск	3,8340	2,84699	4,24179	-0,408506
Низкий риск	-18,0988	7,82423	-2,67280	0,052658
Кредит	-46,4491	-613447	-0,06081	0,016950

Судить о результатах разделения программой наблюдений по группам, удобней по диаграмме рассеяния (рис.1). На диаграмме видно, что наблюдения (предприятия), принадлежащие одинаковым группам, локализованы в определенных областях плоскости. При этом, группы ОТКАЗ и КРЕДИТ наиболее отдалены друг от друга, так как показателям в этих группах соответствуют большие значения корня 1 (Root 1), т.е. дискриминантная функция 1 наилучшим образом дискриминирует показатели между этими группами.

Матрица классификации (табл. 5), которая содержит информацию о количестве и проценте корректно классифицированных наблюдений в каждой группе, показала, что исходные (строки матрицы) и предсказанные (столбцы матрицы) классы полностью совпали. Это также означает, что дискриминация была проведена успешно.

В табл. 6 приведен фрагмент таблицы апостериорных вероятностей, в которой отображена вероятность принадлежности каждого предприятия к группам риска банкротства. Клиент приписывается к той группе риска неплатежеспособности, которой соответствует наибольшая апостериорная вероятность классификации.

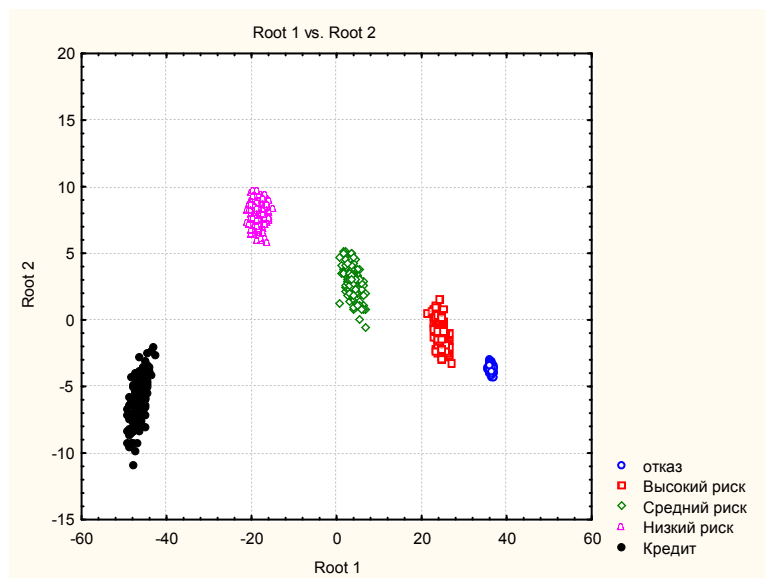


Рис. 1. Диаграмма рассеяния канонических значений для пар значений дискриминантных функций 1 и 2

Таблица 5. Матрица классификации

Группы	Матрица классификаций					
	Стрки: Исходная классификация					
	Столбцы: Предсказанная классификация					
	% совпадений	Отказ	Высокий риск	Средний риск	Низкий риск	Кредит
Отказ	100,0000	200	0	0	0	0
Высокий риск	100,0000	0	200	0	0	0
Средний риск	100,0000	0	0	200	0	0
Низкий риск	100,0000	0	0	0	200	0
Кредит	100,0000	0	0	0	0	200
Итого	100,0000	200	200	200	200	200

Таблица 6. Значения апостериорных вероятностей

Случай	Апостериорные вероятности					
	Исходная классификация	Отказ p=0,2000	Высокий риск p=0,2000	Средний риск p=0,2000	Низкий риск p=0,2000	Кредит p=0,2000
1	Отказ	1,000000	0,000000	0,000000	0,000000	0,000000

2	Отказ	1,000000	0,000000	0,000000	0,000000	0,000000
3	Отказ	1,000000	0,000000	0,000000	0,000000	0,000000
4	Отказ	1,000000	0,000000	0,000000	0,000000	0,000000
5	Отказ	1,000000	0,000000	0,000000	0,000000	0,000000

При помощи языка STATISTICA VISUAL BASIC (SVB), был создан программный код модуля анализа данных, который по введенным значениям финансовых показателей предприятия определяет его в одну из пяти рассматриваемых групп уровня риска неплатежеспособности исходя из анализа таблицы значений апостериорных вероятностей. Для удобства использования модуля было создано пользовательское диалоговое окно ввода данных (рис.3) и окно вывода (рис.4), содержащее результат классификации.

L1	2,12	R2	0,546
L3	510,347	R3	11,227
P1	1,857	R4	21,352
F1	0,576	R5	11,341
F2	0,952	A2	13,472
F3	2,345	A4	3,423
F4	0,321	A5	4,237
R1	23,949	A6	8,487

OK

Рис.2. Окно ввода данных

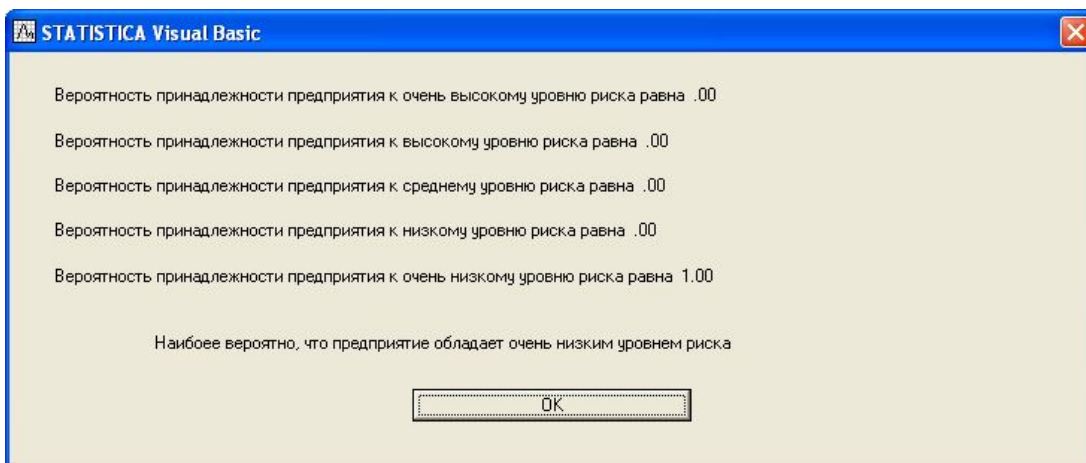


Рис. 3. Окно вывода результатов классификации

Таким образом дискриминантный анализ позволил создать программный продукт, который на основе виртуальной клиентской базы классифицирует предприятия по пяти группам риска невозврата кредита. Созданный программный модуль автоматизирует процесс классификации предприятий по уровню кредитоспособности, тем самым значительно сокращая затраты времени, предупреждая ошибки и обеспечивая простоту использования статистического метода. Используя данный программный продукт можно оперативно прогнозировать уровень риска неплатежеспособности реально существующего предприятия. От пользователя не требуется специальных знаний в области современных технологий анализа данных, так как программа имеет простой интерфейс, состоящий из окон ввода и вывода данных.

Список литературы

1. Шевченко И.В., Халафян А.А., Васильева Е.Ю. «Создание виртуальной клиентской базы для анализа кредитоспособности российских предприятий»// *Финансы и кредит*, №1, 2010, с.13-18.
2. Недосекин А.О., Бессонов Д.Н., Лукашев А.В. «Сводный финансовый анализ российских предприятий за 2000-2003 г.г.»