

В. И. МАЛЮГИН, Н.В. ГРИНЬ

КОМПЬЮТЕРНЫЙ АНАЛИЗ ДАННЫХ И МОДЕЛИРОВАНИЕ НА IBM SPSS

учебно-методические указания
по лабораторному практикуму
«Компьютерный анализ данных и моделирование»

по курсу
«Многомерных статистический анализ данных в экономике»,
специальности: «Экономическая кибернетика»

МИНСК, БГУ
2013

ОГЛАВЛЕНИЕ

1. Описание используемых статистических данных.....	3
2. Предварительный статистический анализ данных.....	3
3. Корреляционный анализ данных.....	7
4. Факторный анализ. Метод главных компонент.....	8
5. Кластерный анализ данных.....	10
6. Дискриминантный анализ данных.....	11
7. Классификация на основе логит-модели множественного выбора.....	15
Литература.....	16
Приложение 1. Финансовые коэффициенты и формулы их расчета	17
Приложение 2. Границы цензурирования для финансовых коэффициентов.....	19
Приложение 3. Архив данных.....	20

1. ОПИСАНИЕ ИСПОЛЬЗУЕМЫХ СТАТИСТИЧЕСКИХ ДАННЫХ

Анализируемая выборка представляет собой статистические данные по нефинансовым предприятиям. Источником информации является бухгалтерская отчетность субъектов хозяйствования, представленная на отчетную дату (бухгалтерский баланс, отчет о прибылях и убытках).

С точки зрения отчетного периода имеются 2 типа данных:

- ♦ годовые данные (информация представляется за отчетный год)
- ♦ квартальные данные (информация представляется за отчетные периоды: 3 месяца, 6 месяцев, 9 месяцев, 12 месяцев).

С целью сокращения исходного пространства признаков и перехода к безразмерным показателям формируется множество финансовых коэффициентов. Для наиболее полного и многостороннего анализа финансового состояния предприятий были отобраны коэффициенты из 5 групп, характеризующие различные стороны финансово-хозяйственной деятельности предприятий. Далее приводится список финансовых коэффициентов, формулы их расчета даны в Приложении.

I. Показатели ликвидности

K1 – коэффициент текущей ликвидности;

K2 – коэффициент абсолютной ликвидности;

K3 – коэффициент ликвидности средств в обращении (коэффициент быстрой (срочной) ликвидности);

II. Коэффициенты финансовой устойчивости

K4 – коэффициент обеспеченности собственными оборотными средствами;

K5 – коэффициент обеспеченности финансовых обязательств активами;

K6 – качество дебиторской задолженности;

K7 – качество кредиторской задолженности;

K8 – коэффициент финансовой независимости;

K9 – коэффициент финансирования;

K10 – коэффициент мобильности средств;

III. Коэффициенты динамики развития

K11 – темпы роста выручки;

K12 – рост капитала;

IV. Коэффициенты деловой активности

K13 – коэффициент оборачиваемости активов;

K14 – коэффициент оборачиваемости кредиторской задолженности;

K15 – коэффициент оборачиваемости дебиторской задолженности;

K16 – коэффициент соотношения длительности периодов оборачиваемости дебиторской задолженности и кредиторской задолженности;

K17 – коэффициент оборачиваемости товарных запасов;

V. Коэффициенты рентабельности (эффективности деятельности)

K18 – коэффициент рентабельности продукции;

K19 – коэффициент рентабельности активов;

K20 – коэффициент рентабельности капитала.

2. ПРЕДВАРИТЕЛЬНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ

Цели и задачи предварительного статистического анализа рассматриваемых данных

Предварительный статистический анализ данных позволяет повести первичное исследование свойств рассматриваемой выборки данных. На данном этапе проводится

графический анализ данных, рассматриваются основные описательные статистики, исследуется закон распределения, анализируется присутствие аномальных наблюдений, корреляционные связи и т.д. На основании полученных результатов делается заключение о возможности и целесообразности применения в дальнейшем к рассматриваемым данным различных статистических методов и алгоритмов анализа.

2.1. Анализ распределений выборок на основе графиков и статистических тестов

ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Исследовать основные описательные статистики для каждого коэффициента (минимум, максимум, среднее, медиана, стандартное отклонение) с целью ознакомления с особенностями данных и их распределением.
2. Провести графический анализ данных с целью анализа их распределения на основе построения гистограмм.

ИЛЛЮСТРАТИВНЫЕ ПРИМЕРЫ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ АНАЛИЗА

1. В табл.1 приведена основная описательная статистика для коэффициентов $K1$ и $K4$. На рис. 1 – соответствующие гистограммы.

Таблица 1. Описательная статистика для коэффициентов $K1$, $K4$

Характеристика	$K1$	$K4$
Минимум	0,370	-
Максимум	14,970	0,910
Среднее	2,238	0,072
Медиана	1,543	0,163
Стд. отклонение	2,059	0,558

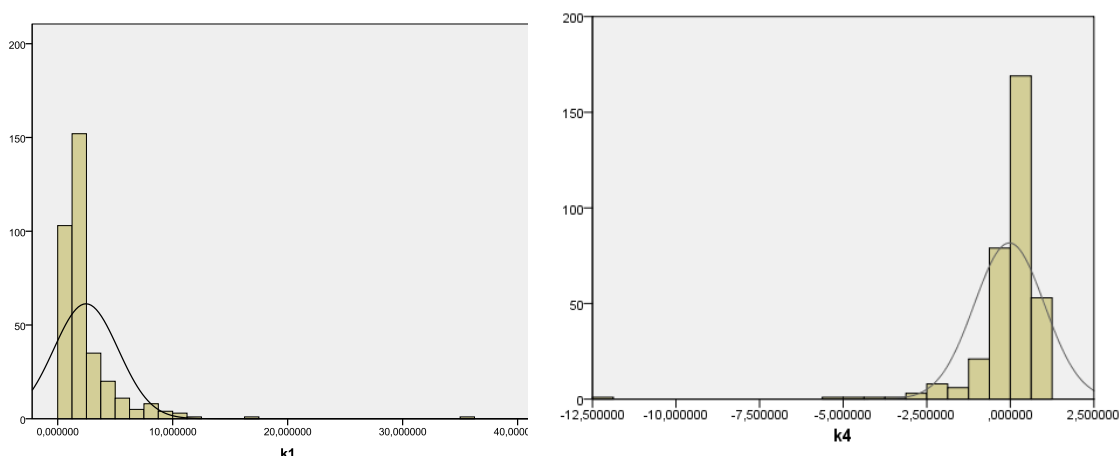


Рис. 1. Гистограммы распределений коэффициентов $K1$, $K4$

На основе анализа полученных результатов дать заключения относительно следующих свойств данных:

- 1) относительно области значений для анализируемых коэффициентов существенно отличаются;
- 2) относительно разброса значений для отдельных коэффициентов;

3) относительно близости средних значений и медиан рассматриваемых коэффициентов, для каких коэффициентов можно говорить об асимметричности распределений либо о наличии аномальных наблюдений в выборке;

4) относительно свойств распределения коэффициентов, в частности относительно нормальности распределения;

5) относительно наличия в выборках аномальных и экстремальных наблюдений.

2.2. Анализ аномальных наблюдений

Графический анализ данных. С целью выявления аномальных наблюдений может проводиться визуальный анализ данных. Для этого используются ящичные диаграммы – графики, компактно изображающие одномерные распределения вероятностей. Не входящие в статистически значимый диапазон точки (аномальные наблюдения) изображаются на графиках отдельно. В случае наличия большого числа аномальных наблюдений или ассиметричного распределения анализируемой переменной (что приводит к идентификации большого числа наблюдений как аномальных) данный метод использовать проблематично. В этом случае требуется использовать специальные алгоритмы, автоматизирующие поиск аномальных наблюдений.

Выявление экстремальных наблюдений. Для выявления так называемых «экстремальных» наблюдений в симметричных распределениях, как правило, не имеющих содержательной экономической интерпретации, может использоваться тест Хампеля, в соответствии с которым экстремальными считаются наблюдения x_* из выборки x_1, \dots, x_n , не попавшие в интервал

$$[x_{med} - 5.2 \cdot x_{mad}, x_{med} + 5.2 \cdot x_{mad}], \quad (1)$$

где x_{med} – медиана выборки $\{x_i\}$, а x_{mad} – медиана выборки $\{|x_i - x_{med}|\}$ ($i = \overline{1, n}$) абсолютных отклонений от x_{med} . Множитель, используемый в (1) может корректироваться таким образом, чтобы в результате использования данного условия экстремальными признавалось не более 5-8% наблюдений.

Альтернативным вариантом исключения экстремальных наблюдений может быть использование "*правила шести сигм*": наблюдение исключается из дальнейшего анализа, если его отклонение от выборочного среднего более чем в 6 раз превышает среднеквадратическое отклонение выборки.

ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Исследовать возможность применения указанных тестов для анализа экстремальных аномальных наблюдений в исследуемых данных. Проанализировать влияние ассиметричности распределений на результаты анализа.

2. Осуществить коррекцию границ интервалов для указанных тестов с учетом экономической интерпретации результатов анализа.

Цензурирование и нормировка данных. Следующим этапом обработки аномальных наблюдений является цензурирование выборки, то есть замена аномально больших и малых значений коэффициентов на экономически и статистически обоснованные максимальные и минимальные пороговые значения (K_{min} , K_{max}), задаваемые экспертным путем. В связи с этим возникает задача выбора для каждого коэффициента экономически обоснованных интервалов $[K_{min}; K_{max}]$, учитывающих особенности распределения значений коэффициентов в анализируемой выборке. Для этой цели проводится частотный и графический анализ распределения значений показателей по

заданным диапазонам. Значения рассматриваемых показателей, не попадающие в заданный диапазон $[K_{\min}; K_{\max}]$, подвергаются цензурированию.

Для возможности более удобной интерпретации результатов производится приведение значений всех рассчитанных финансовых коэффициентов к шкале $[0,1]$. При этом более высокие значения показателей соответствуют предприятиям с более высокой степенью кредитоспособности. С этой целью используются следующие формулы:

$$\tilde{K} = \frac{K - K_{\min}}{K_{\max} - K_{\min}} \quad \text{или} \quad \tilde{K} = \frac{K_{\max} - K}{K_{\max} - K_{\min}} \quad (2)$$

для случаев прямой и обратной зависимости кредитоспособности (платежеспособности) от рассматриваемого показателя соответственно.

ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Провести графический анализ данных с целью выявления аномальных наблюдений на основе построения ящичных диаграмм.
2. На основании условия (1) или «правила шести сигм» определить экстремальные наблюдения в выборке. Множитель в (1) подобрать таким образом, чтобы экстремальными признавалось не более 5-8% наблюдений
3. Провести цензурирование и нормировку данных.

ИЛЛЮСТРАТИВНЫЕ ПРИМЕРЫ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ АНАЛИЗА:

1. Ящичные диаграммы для отдельных коэффициентов K_9 – K_{12} в качестве примера приведены на рис. 2. Из данных графиков следует, что для отдельных коэффициентов достаточно много значений относится к аномальным. Исключение всех таких наблюдений приведет к значительному сокращению выборки.

2. Использование условия (1) с множителем равным 6 приводит к идентификации в качестве экстремальных 208 наблюдений, что составляет 8% от общего количества наблюдений. Информация по таким наблюдениям из выборки исключается.

3. Проведены цензурирование и нормировка данных.

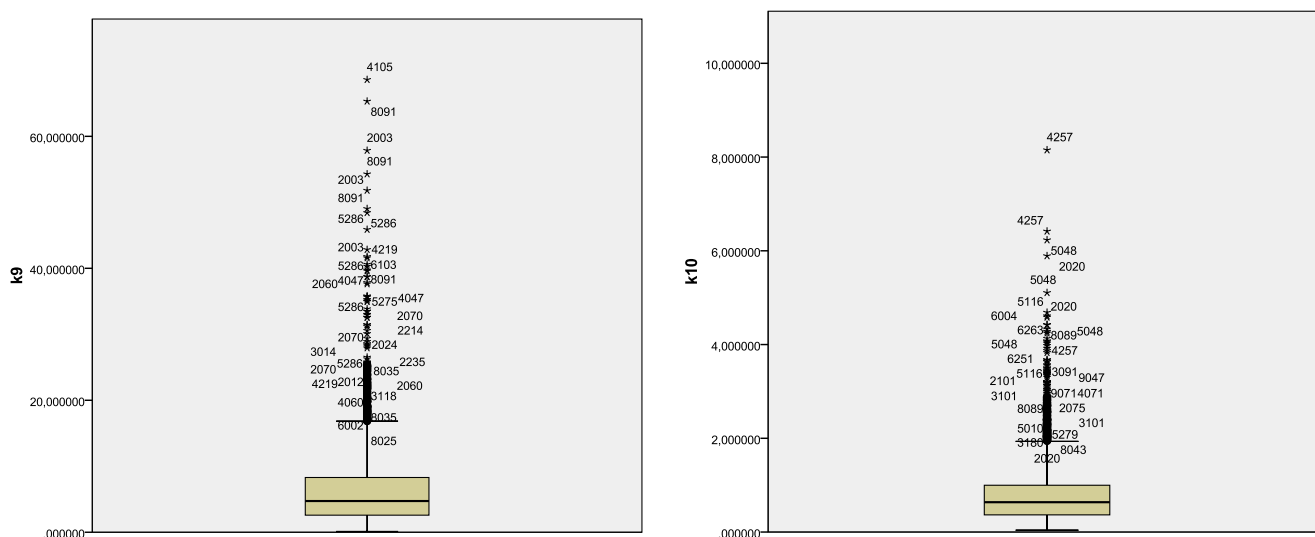


Рис. 2. Ящичная диаграмма для коэффициента K_9 и а K_{10}

3. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ДАННЫХ

Важным этапом статистического анализа данных является исследование корреляционных зависимостей между переменными, позволяющее выявить существующие взаимосвязи между переменными, их характер и тесноту. Особое внимание следует уделять анализу знаков парных коэффициентов корреляции для получения достоверных, имеющих содержательную экономическую интерпретацию выводов. Данный анализ целесообразно проводить после исключения аномальных наблюдений, так как их присутствие может исказить реально существующие зависимости.

ЗАДАЧИ ИССЛЕДОВАНИЯ:

1. Получить матрицу парных коэффициентов корреляции для всех имеющихся показателей.
2. Выявить переменные, имеющие наиболее тесные взаимосвязи, а также проанализировать знаки показателей корреляции.
3. Выделить переменные, которые возможно исключить из дальнейшего анализа в силу существования тесной связи с другой/ими переменной/ыми.

ИЛЛЮСТРАТИВНЫЕ ПРИМЕРЫ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ АНАЛИЗА:

1. В табл. 2 приведены коэффициенты парной корреляции и соответствующий им уровень значимости для коэффициентов *K1-K10*. Большинство анализируемых коэффициентов демонстрирует наличие статистически значимой взаимосвязи с остальными показателями. Наиболее сильная взаимосвязь наблюдается для показателей в рамках отдельных групп, характеризующих различные особенности деятельности предприятий.

Таблица 2. Коэффициенты парной корреляции для коэффициентов *K1–K10*

		K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
K1	<i>Коррел.</i>	1	0,74	0,88	0,48	-0,41	-0,09	-0,24	0,41	0,54	0,08
	<i>Знач.</i>		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
K2	<i>Коррел.</i>	0,74	1	0,82	0,34	-0,32	-0,11	-0,21	0,32	0,40	0,07
	<i>Знач.</i>	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
K3	<i>Коррел.</i>	0,88	0,82	1	0,42	-0,36	-0,09	-0,22	0,36	0,49	0,10
	<i>Знач.</i>	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00
K4	<i>Коррел.</i>	0,48	0,34	0,42	1	-0,57	-0,08	-0,25	0,57	0,32	0,22
	<i>Знач.</i>	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00
K5	<i>Коррел.</i>	-0,41	-0,32	-0,36	-0,57	1	0,03	0,17	-1	-0,57	0,43
	<i>Знач.</i>	0,00	0,00	0,00	0,00		0,19	0,00	0,00	0,00	0,00
K6	<i>Коррел.</i>	-0,09	-0,11	-0,09	-0,08	0,03	1	0,39	-0,03	-0,02	-0,09
	<i>Знач.</i>	0,00	0,00	0,00	0,00	0,19		0,00	0,23	0,28	0,00
K7	<i>Коррел.</i>	-0,24	-0,21	-0,22	-0,25	0,17	0,39	1	-0,17	-0,15	-0,08
	<i>Знач.</i>	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00
K8	<i>Коррел.</i>	0,41	0,32	0,36	0,57	-1	-0,03	-0,17	1	0,57	-0,43
	<i>Знач.</i>	0,00	0,00	0,00	0,00	0,00	0,23	0,00		0,00	0,00
K9	<i>Коррел.</i>	0,54	0,40	0,49	0,32	-0,57	-0,02	-0,15	0,57	1	-0,29
	<i>Знач.</i>	0,00	0,00	0,00	0,00	0,00	0,28	0,00	0,00		0,00
K10	<i>Коррел.</i>	0,08	0,07	0,10	0,22	0,43	-0,09	-0,08	-0,43	-0,29	1

В соответствии с полученными результатами из дальнейшего анализа целесообразно исключить коэффициент K_8 , так как для него существует корреляция близкая к -1 с коэффициентом K_5 . С целью принятия решения о дальнейшем использовании в анализе более тщательному рассмотрению следует подвергнуть коэффициенты ликвидности, так как в рамках этой группы корреляционные связи между переменными наиболее сильные.

4. ФАКТОРНЫЙ АНАЛИЗ. МЕТОД ГЛАВНЫХ КОМПОНЕНТ

ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Применить факторный анализ (метод главных компонент) к совокупности анализируемых данных. Проанализировать целесообразность использования данного метода анализа на основе специальных критериев и тестовых статистик.
2. На основании результатов факторного анализа определить наиболее значимые коэффициенты, сохраняющие наибольший процент дисперсии исходных показателей.
3. Рассчитать интегральный показатель кредитоспособности как взвешенную сумму полученных главных компонент. В качестве весов использовать соответствующие величины процентов объясненной дисперсии.

ИЛЛЮСТРАТИВНЫЕ ПРИМЕРЫ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ АНАЛИЗА

Целесообразность выполнения факторного анализа определяется наличием существенной корреляции между используемыми переменными. Для оценки целесообразности применения факторного анализа для конкретной выборки данных используются специальные критерии и тестовые статистики, которые позволяют исследовать требуемые свойства выборочной корреляционной матрицы. В рамках проводимого анализа используются: *критерий сферичности Бартлетта* и *статистика Кайзера–Мейера–Олкина* (КМО).

Критерий сферичности Бартлетта проверяет нулевую гипотезу об отсутствии корреляций между переменными в генеральной совокупности. Статистика КМО позволяет проверить, насколько корреляцию между парами переменных можно объяснить другими переменными (факторами).

В соответствии с табл. 3, нулевая гипотеза о том, что корреляционная матрица является единичной, отклоняется критерием сферичности Бартлетта. Приближенное значение статистики равно 2926,418, она является значимой на уровне 0,05. Значение статистики КМО, равное 0,665, свидетельствует о приемлемом качестве выборки для применения факторного анализа. Пороговым в данном случае обычно считается значение 0,6. Следовательно, использование факторного анализа применительно к данной выборке является целесообразным.

Таблица 3. Анализ пригодности выборочной корреляционной матрицы

Статистика Кайзера–Мейера–Олкина		0,665
Критерий сферичности Бартлетта	Статистика χ^2	2926,418
	p -значение	0,000

Согласно табл. 4, в совокупности выделенные факторы объясняют 73,133% дисперсии исходных показателей, в частности, первый фактор до вращения объясняет 30,781% дисперсии, второй – 16,207%, третий – 10,464%, четвертый – 8,416%, пятый – 7,264%. Вращение факторов (кватримакс) приводит к более равномерному распределению объясненной дисперсии между факторами и к увеличению удельного веса каждого из них (за исключением первого фактора). Таким образом, на долю первого фактора после вращения приходится 27,547% объясненной дисперсии, второго – 13,357%, третьего – 12,214%, четвертого – 10,103%, пятого – 9,912%.

Таблица 4. Собственные значения и объясненная дисперсия

№ фактора	Начальные факторы до вращения			Факторы после вращения		
	Собственные значения	% объясненной дисперсии	Кумулятивный % объясненной дисперсии	Собственные значения	% объясненной дисперсии	Кумулятивный % объясненной дисперсии
1.	4,309	30,781	30,781	3,857	27,547	27,547
2.	2,269	16,207	46,988	1,870	13,357	40,904
3.	1,465	10,464	57,453	1,710	12,214	53,118
4.	1,178	8,416	65,868	1,414	10,103	63,220
5.	1,017	7,264	73,133	1,388	9,912	73,133

Для интерпретации факторов используется анализ факторных нагрузок до и после вращения факторов, приведенные для используемой выборки в табл. 5 и 6 соответственно.

Таблица 5. Факторные нагрузки до вращения

Коэффициенты	Факторы				
	F1	F2	F3	F4	F5
K1 n	0,777				
K2 n	0,740				
K4 n	0,829				
K5 n	0,708	-0,531			
K6 n				0,820	
K7 n	0,519			0,497	
K9 n	0,590	-0,555			
K10 n		0,706	-0,467		
K11 n					0,648
K13 n		0,689			
K14 n	0,618				
K15 n			0,742		
K18 n	0,578				0,444
K19 n	0,688	0,519			

Таблицы 5, 6 содержат элементы, которые представляют собой коэффициенты корреляции для отдельных показателей и построенных факторов (элементы, величины которых не превосходят 0,4 – опущены). Как видно из таблицы 5, наиболее тесная корреляция с первой главной компонентой, сохраняющей наибольший процент

дисперсии исходных показателей, существует со следующими коэффициентами: $K1_n$, $K2_n$, $K4_n$, $K5_n$, $K7_n$, $K9_n$, $K14_n$, $K18_n$, $K19_n$. Отметим также, что все исходные показатели имеют достоверную связь с одной из шести компонент, при этом, если первые две компоненты тесно коррелируют с множеством рассматриваемых коэффициентов, то для второй – пятой компонент таких коэффициентов может быть выявлено меньше.

Вращение факторов позволяет достичь более простой структуры факторных нагрузок для исходных переменных, что может несколько упростить интерпретацию полученных факторов.

После применения процедуры вращения с первой главной компонентой наиболее тесно коррелируют показатели ликвидности и рентабельности. Со второй компонентой – коэффициенты финансовой устойчивости. С третьей компонентой – показатели деловой активности, с четвертой – коэффициенты динамики развития и рентабельности. Показатели, характеризующие качество кредиторской и дебиторской задолженности – с пятой компонентой.

Таблица 6. Факторные нагрузки после вращения

Коэффициенты	Факторы				
	F1	F2	F3	F4	F5
K1 n	0,884				
K2 n	0,771				
K4 n	0,897				
K5 n	0,710	-0,555			
K6 n					0,878
K7 n					0,661
K9 n	0,558	-0,623			
K10 n		0,897			
K11 n				0,780	
K13 n		0,418	0,768		
K14 n	0,431		0,569		
K15 n			0,756		
K18 n	0,509			0,657	
K19 n	0,528			0,549	

5. КЛАСТЕРНЫЙ АНАЛИЗ ДАННЫХ

ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Осуществить классификацию наблюдений с использованием алгоритма кластерного анализа (« k -средних»). Количество классов задать равным четырём.
2. Пронумеровать полученные классы в соответствии со средним значением интегрального показателя кредитоспособности для наблюдений отдельного класса: классу с наиболее высоким уровнем кредитоспособности соответствует более высокое среднее значение интегрального показателя кредитоспособности.

ИЛЛЮСТРАТИВНЫЕ ПРИМЕРЫ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ АНАЛИЗА

Информация, представленная в табл. 7, отражает распределение наблюдений по кластерам. Самыми малочисленными оказались первый и третий классы, многочисленными – второй и четвертый.

Анализ описательной статистики для интегрального показателя в разрезе выделенных кластеров (табл.8) свидетельствует в пользу хорошей разделимости классов: среднее значение, медиана, минимум и максимум интегрального показателя монотонно убывают с увеличением номера класса.

Таблица 7. Распределение наблюдений по кластерам

Номер кластера	Количество наблюдений	Относительная частота
1	383	0,159
2	753	0,313
3	715	0,297
4	555	0,231

Таблица 8. Описательная статистика для интегральных показателей

Кластер	Число наблюдений	Среднее	Медиана	Минимум	Максимум	Стандартное отклонение
1	383	0,5770	0,6113	0,07	1,01	0,22647
2	753	0,2211	0,1904	-0,24	0,85	0,23765
3	715	-0,1497	-0,1241	-0,70	0,30	0,22900
4	555	-0,3683	-0,3797	-0,87	0,28	0,22591

В табл. 9 приведены координаты центров кластеров в пространстве финансовых коэффициентов.

Таблица 9. Координаты центров полученных кластеров

Коэффициент (координата)	Кластер			
	1	2	3	4
K1_n	0,7147	0,1940	0,1828	0,0891
K2_n	0,7272	0,0930	0,0589	0,0340
K4_n	0,6393	0,1338	0,1308	0,0519
K5_n	0,8586	0,6283	0,6814	0,4381
K6_n	0,3596	0,6618	0,1679	0,2273
K7_n	0,8725	0,8257	0,2432	0,3303
K9_n	0,9570	0,8072	0,9145	0,2728
K10_n	0,3360	0,3080	0,1605	0,4828
K11_n	0,4418	0,5031	0,4030	0,4536
K13_n	0,3582	0,3476	0,1272	0,2628
K14_n	0,9331	0,8643	0,5865	0,4355
K15_n	0,9701	0,9531	0,9261	0,9084
K18_n	0,3390	0,2306	0,1486	0,1380
K19_n	0,6338	0,3982	0,1636	0,2103

6. ДИСКРИМИНАНТНЫЙ АНАЛИЗ ДАННЫХ

ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Осуществить кластерный анализ данных в пространстве финансовых коэффициентов и рассчитать вектор классификаций для выборки (см. раздел 5).

2. Исследовать возможность прогнозирования класса кредитоспособности новых наблюдений с использованием линейного дискриминантного анализа.

С помощью алгоритма классификации на основе линейного дискриминантного анализа классифицировать:

а) исходную выборку (режим переклассификации с вычислением условных и безусловных вероятностей ошибок);

б) выборку предварительно исключенных наблюдений из различных классов (по 10%) (режим экзамена с вычислением условных и безусловных вероятностей ошибок).

ИЛЛЮСТРАТИВНЫЕ ПРИМЕРЫ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ АНАЛИЗА

В табл. 10 приведены результаты переклассификации обучающей выборки: предсказанное и исходное распределение предприятий по четырем классам платежеспособности.

Таблица 10. Результаты переклассификации обучающей выборки

	Номер класса (рейтинг)	Предсказанная принадлежность к классу				Всего наблюдений
		1	2	3	4	
Частота	1	345	33	4	1	383
	2	3	714	24	12	753
	3	6	27	632	50	715
	4	3	8	9	535	555
%	1	90,1	8,6	1,0	0,3	100,0
	2	0,4	94,8	3,2	1,6	100,0
	3	0,8	3,8	88,4	7,0	100,0
	4	0,5	1,4	1,6	96,4	100,0

Оценки вероятностей ошибок при переклассификации обучающей выборки содержит табл. 11. Согласно этой информации 92,5% исходных наблюдений классифицированы верно. Наиболее высокая точность достигнута при классификации предприятий, относящихся к четвертому классу платежеспособности (94,3%), наименьшая – предприятий из третьего класса. Это может быть связано, например, с пересеканием третьего класса платежеспособности со смежными классами.

Таблица 11. Оценки вероятностей ошибок при переклассификации обучающей выборки

Класс (рейтинг) k	Оценка априорной вероятности класса $\hat{\pi}_k$	Оценка условной вероятности ошибки \hat{P}_k
1	0,159	0,083
2	0,313	0,066
3	0,297	0,113

4	0,231	0,057
Безусловная вероятность ошибки P		0,075 (7,5%)

В табл. 12 приведены коэффициенты классифицирующих функций d_i , $i = 1, \dots, 4$, использующихся для предсказания класса, которому принадлежит новое наблюдения. Правило классификации следующее: наблюдение принадлежит классу кредитоспособности i , для классифицирующей функции которого d_i получено наибольшее значение.

Таблица 12. Коэффициенты классифицирующих функций

	d_1	d_2	d_3	d_4
k1 n	20,854	11,276	9,721	10,644
k2 n	16,013	-1,489	-0,757	-1,146
k4 n	-20,986	-31,872	-32,366	-31,512
k5 n	44,539	50,489	51,714	52,646
k6 n	3,794	4,981	1,657	2,585
k7 n	6,223	7,020	-1,369	4,168
k9 n	6,075	6,219	7,848	-6,820
k10 n	18,017	20,580	21,796	26,737
k11 n	8,932	9,726	9,030	10,898
k13 n	2,206	1,089	0,439	-1,975
k14 n	2,333	5,097	-1,796	1,999
k15 n	52,184	49,621	51,907	50,497
k19 n	-3,622	-6,434	-7,793	-7,803
Константа	-65,343	-54,348	-47,589	-45,811

В таблице 13 приведены результаты классификации экзаменационной выборки предварительно исключенных наблюдений из различных классов (по 10%).

Таблица 13. Результаты классификации экзаменационной выборки

	Номер класса (рейтинг)	Предсказанная принадлежность к группе				Итого
		1	2	3	4	
Частота	1	317	23	3	1	344
	2	3	636	25	12	676
	3	4	24	574	42	644
	4	4	7	10	478	499
	Тестовая выборка	36	85	65	57	243
%	1	92,2	6,7	0,9	0,3	100,0
	2	0,4	94,1	3,7	1,8	100,0
	3	0,6	3,7	89,1	6,5	100,0
	4	0,8	1,4	2,0	95,8	100,0
	Тестовая выборка	14,8	35,0	26,7	23,5	100,0

Оценки вероятностей ошибок при классификации экзаменационной выборки (табл. 14) в основном немного выше аналогичных оценок для переклассификации исходной выборки за исключением вероятности \hat{P}_2 . В целом, 91,8% наблюдений экзаменационной выборки классифицированы верно.

Таблица 14. Оценки вероятностей ошибок при классификации экзаменационной выборки

Класс (рейтинг) k	Оценка априорной вероятности класса $\hat{\pi}_k$	Оценка условной вероятности ошибки \hat{P}_k
1	0,159	0,128
2	0,313	0,026
3	0,297	0,127
4	0,231	0,071
Безусловная вероятность ошибки P		0,082 (8,2%)

В табл. 15 приведены коэффициенты классифицирующих функций d_i , $i = 1, \dots, 4$, полученные при реализации дискриминантного анализа к сокращенной обучающей выборке за счет выделения экзаменационной выборки.

Таблица 15. Коэффициенты классифицирующих функций

	d_1	d_2	d_3	d_4
k1 n	12,192	3,294	1,381	2,114
k2 n	17,559	-0,688	-0,320	-0,188
k4 n	-19,720	-31,865	-32,733	-31,784
k5 n	52,726	58,791	61,022	61,099
k6 n	3,229	4,421	1,141	2,127
k7 n	8,144	8,763	0,615	5,886
k9 n	0,755	1,044	2,336	-11,745
k10 n	14,350	17,189	18,144	23,357
k13 n	4,527	2,304	0,991	-0,395
k14 n	10,639	13,239	7,008	10,556
k18 n	4,184	1,053	-0,431	1,396
k19 n	-2,134	-3,375	-4,052	-5,091
Константа	-41,890	-31,912	-23,893	-22,443

В табл. 16 систематизирована доля несовпадений прогнозных рейтингов от исходных для различных классов. Процент несовпадений с отклонением на $|k| \leq 1$ классов (не более, чем на один класс) при использовании дискриминантного анализа составляет 98,578% и 97,943% для режима переклассификации и режима экзамена соответственно.

Таблица 16. Отклонения прогнозных рейтингов от «истинных» для различных условий классификации

Алгоритм	Доля несовпадений с отклонением на k классов, в %							
	k							
	-3	-2	-1	0	1	2	3	$ k \leq 1$
Режим переклассификации	0,04	0,67	4,45	92,52	1,61	0,58	0,13	98,58
Режим экзамена	0,0	0,41	4,12	91,77	2,06	1,65	0,0	97,94

7. КЛАССИФИКАЦИЯ НА ОСНОВЕ ЛОГИТ-МОДЕЛИ МНОЖЕСТВЕННОГО ВЫБОРА

ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Исследовать возможность прогнозирования класса кредитоспособности новых наблюдений на основе логит-модели множественного выбора.
2. Сравнить результаты классификации на основе логит-модели множественного выбора с результатами классификации на основе дискриминантного анализа.

ИЛЛЮСТРАТИВНЫЕ ПРИМЕРЫ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ АНАЛИЗА

В табл. 17 приведены оценки коэффициентов логит-модели множественного выбора, а также оценки пороговых значений. В итоговую модель вошло 7 финансовых коэффициентов. Все параметры модели являются статистически значимыми на уровне 0.05. Гипотеза об адекватности модели в целом не отклоняется на основании критерия отношения правдоподобия (P -значение для LR -статистики менее 0.05), значение статистического критерия $pseudo R^2$ равно 0.423. Информационные статистики Акаике и Шварца равны: $AIC=1.59$, $SC=1.62$. Знаки коэффициентов в построенной модели согласуются с априорными представлениями. В соответствии с правилами проводимой нормировки преобразованные коэффициенты имеют интерпретацию: чем больше, тем лучше финансовое состояние предприятия. В то же время наиболее благоприятным является минимальное значение рейтинга, поэтому ожидаемым было получение отрицательных знаков при оценках коэффициентов модели (обратная зависимость между номером рейтинга и финансовыми показателями после нормировки).

Таблица 17. Результаты оценки модели множественного выбора

Переменная	Коэффициент	z-статистика	P-значение
K1_n	-5.911561	-17.95596	0.0000
K6_n	-0.415496	-2.631979	0.0085
K7_n	-1.479929	-9.188005	0.0000
K9_n	-4.486302	-16.57195	0.0000
K10_n	-1.971785	8.271702	0.0000
K13_n	-1.384013	-5.780518	0.0000
K18_n	-1.330547	-4.508724	0.0000
Пороговые значения			
C(1)	-9.185464	-26.52656	0.0000
C(2)	-5.824439	-20.24824	0.0000
C(3)	-2.935452	-12.89076	0.0000

Табл. 18 представляет собой классификационную таблицу, отражающую распределение наблюдаемых и прогнозных рейтингов. В соответствии с полученными результатами доля совпадений по отдельным классам колеблется от 49.42% до 75.36%. Наиболее точным оказывается классификация наблюдений, которым соответствуют рейтинги, равные 1 и 2. Плохая разделимость 3 и 4 классов кредитоспособности приводит к более низким показателям точности классификации для наблюдений из этих классов.

Таблица 18. Классификационная таблица, %

Наблюдаемые значения	Прогнозные значения			
	$\hat{y} = 1$	$\hat{y} = 2$	$\hat{y} = 3$	$\hat{y} = 4$
$y = 1$	73.63014	24.31507	2.054795	0.000000
$y = 2$	6.004141	75.36232	18.63354	0.000000
$y = 3$	0.464037	25.29002	49.41995	25.98608
$y = 4$	0.000000	1.208459	39.27492	59.51662

В целом, доля точных прогнозов находится на уровне 63.44 % (табл. 19), доля прогнозов с отклонением не более чем на один рейтинговый класс от истинного рейтинга – более 99%. Использование дискриминантного анализа позволяет получить большее количество точных прогнозов. При этом доля прогнозов с отклонением не более чем на один рейтинговый класс от истинного рейтинга в случае дискриминантного анализа немного ниже.

Таблица 19. Ошибки прогнозов, %

-3	-2	-1	0	1	2	3	$ \Delta \leq 1$
0	0.39	16.14	63.44	19.65	0.39	0	99.22

ЛИТЕРАТУРА

1. Харин Ю.С., Малюгин В.И., Абрамович М.С. Математические и компьютерные основы статистического моделирования анализа данных. Минск: БГУ, 2008. – 455 с
2. Харин Ю.С., Жук Е.Е. Математическая и прикладная статистика. Минск, БГУ, 2006. – 279 с.
3. Айвазян С.А. и др. Прикладная статистика. Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983. – 561 с.
4. Айвазян С.А. и др. Прикладная статистика. Исследование зависимостей. М.: Финансы и статистика, 1985. – 487 с.
5. Айвазян С.А. и др. Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989. – 607 с.
6. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. М.: Финансы и статистика, 1998. – 352 с.
7. Магнус, Я.Р. Эконометрика. Начальный курс / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – Мн.: Дело, 2004. – 676 с.
8. Наследов А. SPSS: компьютерный анализ данных в психологии и социальных науках. СПб.: Питер, 2005. – 416 с.
9. Бююль, А. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.
10. Пациорковский В.В., Пациорковская В.В. SPSS для социологов. М.: ИСЭПН РАН 2005. – 433 с.
11. Савицкая, Г.В. Экономический анализ: Учеб. / Г.В. Савицкая. – 10 – е изд. – М.: Новое знание, 2008. – 640 с.

Приложение 1. Финансовые коэффициенты и формулы их расчета

Обозначения переменных в формулах расчета представлены в табл. П.1.

Таблица П.1. Обозначения исходных финансовых показателей

V1	Внеоборотные активы (№1, 190)
V7	Финансовые вложения (№1, 270)
V8	Денежные средства (№1, 260)
V9	Общая величина активов (№1, 300)
V10	Капитал и резервы (№1, 490)
V11	Резервы предстоящих расходов (№1, 640)
V14	Долгосрочные обязательства (№1, 590)
V17	Кредиторская задолженность, всего (№1, 620+630)
V18	из неё просроченная (в т.ч. просроченная свыше 3 мес.)
V19	Дебиторская задолженность, всего (№6-ф, 101)
V20	из неё просроченная (№6-ф, 101) (в т.ч. просроченная свыше 3 мес.)
V21	Выручка от реализации товаров, продукции, работ, услуг (№5-ф, 01)
V25	Прибыль (убыток) от реализации (№5-ф, 01)
V26	Прибыль (убыток) от налогообложения (№5-ф, 19+20+21)
V49	Краткосрочные обязательства (№1, 690)

Каждому коэффициенту соответствует степень важности, заданная экспертом, которая означает уровень значимости в сравнении с другими коэффициентами (измеряется от 1 до 5). Чем выше степень важности, тем более значимым является коэффициент. Ниже приведены рассчитываемые коэффициенты, сгруппированные по экономическому смыслу.

Группа 1. Коэффициенты ликвидности характеризуют уровень возможности своевременно оплачивать обязательства перед кредиторами, оценивают степень покрытия обязательств ликвидными активами.

1. $K1$ – коэффициент текущей ликвидности.

$$\text{Формула расчета: } K1 = \frac{V9 - V1}{V49 - V1}.$$

Степень важности: 3.

2. $K2$ – коэффициент абсолютной ликвидности.

$$\text{Формула расчета: } K2 = \frac{V7 + V8}{V49 - V11}.$$

Степень важности: 3.

3. $K3$ – коэффициент ликвидности средств в обращении.

$$\text{Формула расчета: } K3 = \frac{V5 + V13 + V7 + V8}{V49 - V11}.$$

Степень важности: 3.

Группа 2. Коэффициенты финансовой устойчивости определяют степень зависимости компании от внешних источников финансирования. Показатели этой группы характеризуют структуру баланса, качество кредиторской и дебиторской задолженности.

4. $K4$ – коэффициент обеспеченности собственными оборотными средствами.

$$\text{Формула расчета: } K4 = \frac{V10 + V11 - V1}{V9 - V1}.$$

Степень важности: 5.

5. $K5$ – коэффициент обеспеченности финансовых обязательств активами.

Формула расчета: $K5 = \frac{V14 + V49 - V11}{V9}$.

Степень важности: 5.

6. $K6$ – качество дебиторской задолженности.

Формула расчета: $K6 = \frac{V20}{V19}$.

Степень важности: 3.

7. $K7$ – качество кредиторской задолженности.

Формула расчета: $K7 = \frac{V18}{V17}$.

Степень важности: 3.

8. $K8$ – коэффициент финансовой независимости.

Формула расчета: $K8 = \frac{V10 + V11}{V9}$.

Степень важности: 4.

9. $K9$ – коэффициент финансирования.

Формула расчета: $K9 = \frac{V10 + V11}{V17}$.

Степень важности: 2.

10. $K10$ – коэффициент мобильности средств.

Формула расчета: $K10 = \frac{V9 - V1}{V1}$.

Степень важности: 2.

Группа 3. Коэффициенты динамики развития показывают динамику объемных показателей. В условиях роста индексов цен, рост объемных показателей является необходимым условием стабильного финансового состояния.

11. $K11$ – темп роста выручки.

Формула расчета:

$$K11 = \frac{(V21 - V17) \text{ текущего года}}{(V21 - V17) \text{ аналогичный период прошлого года}}.$$

Степень важности: 5.

12. $K12$ – темп роста капитала. Формула расчета:

$$K12 = \frac{(V10 + V11) \text{ текущего года}}{(V10 + V11) \text{ аналогичный период прошлого года}}.$$

Степень важности: 3.

Группа 4. Коэффициенты деловой активности показывают сроки делового цикла предприятия и длительность нахождения средств в обороте. Скорость оборота средств, то есть превращение их в денежную форму, оказывает непосредственное влияние на платежеспособность предприятия.

13. $K13$ – коэффициент оборачиваемости активов.

Формула расчета: $K13 = \frac{V21 \text{ в годовом измерении}}{V9}$.

Степень важности: 4.

14. *K14* – коэффициент оборачиваемости кредиторской задолженности.

$$\text{Формула расчета: } K14 = \frac{V21 \text{ в годовом измерении}}{V17}.$$

Степень важности: 3.

15. *K15* – коэффициент оборачиваемости дебиторской задолженности.

$$\text{Формула расчета: } K15 = \frac{V21 \text{ в годовом измерении}}{V19}.$$

Степень важности: 3.

16. *K16* – коэффициент соотношения длительности периодов оборачиваемости дебиторской и кредиторской задолженности.

$$\text{Формула расчета: } K16 = \frac{K15}{K14}.$$

Степень важности: 2.

17. *K17* – коэффициент оборачиваемости товарных запасов.

$$\text{Формула расчета: } K17 = \frac{V21 \text{ в годовом измерении}}{V6}.$$

Степень важности: 3.

Группа 5. Коэффициенты рентабельности характеризует эффективность деятельности предприятия. Рентабельность обобщающий показатель деятельности предприятия с точки зрения соотношения затрат и результатов.

18. *K18* – коэффициент рентабельности продаж.

$$\text{Формула расчета: } K18 = \frac{V25}{V21} \cdot 100\%.$$

Степень важности: 4.

19. *K19* – коэффициент рентабельности активов.

$$\text{Формула расчета: } K19 = \frac{V26 \text{ в годовом измерении}}{V9} \cdot 100\%.$$

Степень важности: 3.

20. *K20* – коэффициент рентабельности капитала.

$$\text{Формула расчета: } K20 = \frac{V27 \text{ в годовом измерении}}{V10 + V11} \cdot 100\%.$$

Степень важности: 3.

Приложение 2. Границы интервалов цензурирования для финансовых коэффициентов

Пороговые значения для коэффициентов K1-K19, используемые при цензурировании, приведены в табл. П.2.

Таблица П.2. Границы интервалов цензурирования коэффициентов

K	K1	K2	K4	K5	K6	K7	K9	K10	K11	K13	K14	K15	K18	K19
K_{min}	1	0,05	0,15	0	0	0	0,8	0,3	1,0	0,5	1,0	0,0	0,0	0,0
K_{max}	5	0,8	1	1	0,3	0,3	5	2,0	10	3	10	5	30	20

Приложение 3. Архив данных

Для выполнения заданий компьютерного практикума используются следующие выборки годовых и квартальных данных, представленных в виде файлов данных *Excel* и *SPSS*.

- 1) годовые данные (2005-2011 гг.) до исключения аномальных наблюдений, нормировки и цензурирования (*Annual 2005-2011*, файл *START*);
- 2) годовые данные (2005-2011 гг.) после исключения аномальных наблюдений, нормировки и цензурирования (*Annual 2005-2011*, файл *CENSOR*);
- 3) квартальные данные (2008-2011 гг.) после исключения аномальных наблюдений, нормировки и цензурирования (*Quarter 2008-2011*, файл *CENSOR*).