

ASSESSMENT OF THE CREDITWORTHINESS OF COMMERCIAL BANK BORROWERS USING OF APPLIED STATISTICS

The article argues the feasibility and the possibility of complex use of such methods of applied statistical analysis in assessing the creditworthiness of bank borrowers, such as discriminant analysis, factor analysis and logistic regression. Applied research was carried out on the basis of real data of one of the largest commercial banks in Minsk, including six hundred borrowers — individuals for fourteen indicators. Automation of model calculations carried out using the software SPSS.

Keywords: creditworthiness; crediting probability; discriminant analysis; principal component analysis; logistic regression.

Г. О. Читая
доктор экономических наук, доцент
БГЭУ (Минск)

ОЦЕНКА КРЕДИТОСПОСОБНОСТИ ЗАЕМЩИКОВ КОММЕРЧЕСКОГО БАНКА МЕТОДАМИ ПРИКЛАДНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА

В статье аргументирована целесообразность и возможность комплексного использования таких методов прикладного статистического анализа в оценке кредитоспособности заемщиков банка, как дискриминантный анализ, факторный анализ и логистическая регрессия. Прикладное исследование проведено по реальным данным одного из крупнейших коммерческих банков г. Минска, включающим шестьсот заемщиков — физических лиц по четырнадцати показателям. Автоматизация модельных расчетов осуществлена с помощью программного продукта SPSS.

Ключевые слова: оценка кредитоспособности; вероятность получения кредита; дискриминантный анализ; главные компоненты; логистическая регрессия.

Вопросы обоснования состава комплекса методов многомерного статистического анализа и определения последовательности их применения в исследовании кредитоспособности заемщиков коммерческого банка обладают научной и практической ценностью. Многомерное моделирование может послужить дополнительным инструментом совершенствования используемых коммерческими банками методик оценки кредитоспособности.

Для оценки кредитоспособности заемщиков целесообразно использование методов дискриминантного, факторного и логистического анализа, которые в конечном счете приводят к обоснованию вероятностной меры выдачи кредита заемщику. Апробация рекомендуемого набора методов осуществляется на основе реальных данных одного из крупнейших коммерческих банков г. Минска. Выборочная совокупность включает шестьсот заемщиков — физических лиц по четырнадцати показателям, которым ранее выдавались кредиты и известна история выполнения ими обязательств перед банком. Автоматизацию модельных расчетов удобно проводить в рамках статистического пакета прикладных программ, в частности с применением программного продукта IBM SPSS.

Проведение исследования предполагает соблюдение требований к последовательной процедуре прикладного статистического анализа, включающей установление однородности обезличенного состава заемщиков банка по объему выданных кредитов и степени

неравномерности их распределения, снижение размерности признакового пространства методом главных компонент, выбор значимой дискриминантной функции для соотнесения заемщиков к «хорошим» и «плохим» клиентам и проверку состава двух выделенных категорий заемщиков построением модели логистической регрессии на основе сокращенного, статистически значимого числа исходных показателей (переменных). Так как автор оперировал данными о заемщиках, кредитная история которых была известна, им численными расчетами установлена упущенная выгода банка, образовавшаяся за счет невыданных кредитов.

1. Анализ неравномерности и однородности распределения выданных физическим лицам кредитов. Включать в статью выборочную совокупность обезличенных заемщиков в количестве 600 физических лиц по 14 показателям в силу ограниченности ее размеров нецелесообразно. С целью ознакомления с составом и размерностью данных и проведения их прикладного статистического анализа в табл. 1 содержатся данные по последним 50 заемщикам (551–600).

Таблица 1. Фрагмент заемщиков банка по показателям их кредитоспособности

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
551	0	825	4	1975	1	39	1	6	0	1	17	100	1	0
552	0	1018	10	2572	4	54	1	2	1	1	21	45	1	0
553	1	427	12	889	2	49	0	15	1	1	9	150	1	1
554	1	356	12	2237	2	42	1	12	1	0	20	30	0	0
555	1	628	14	2789	3	33	1	12	0	1	10	80	1	0
556	0	465	9	4541	3	46	0	13	1	0	5	46	0	0
557	0	760	12	527	1	19	0	5	1	1	0,6	100	1	1
558	0	657	11	1287	1	39	0	17	1	0	17	45	0	0
559	0	1140	13	3623	3	54	1	9	0	1	10	150	1	1
560	1	671	1	4322	4	33	1	4	1	1	10	30	1	0
561	0	519	11	1706	5	19	0	18	1	1	0,6	10	1	1
562	1	965	4	3728	3	25	0	22	1	1	3	20	0	1
563	0	928	1	4669	2	32	1	5	0	1	10	50	1	1
564	0	875	4	3638	3	29	0	24	0	0	5	100	1	0
565	1	1048	15	4016	4	50	1	15	1	1	12	45	1	1
566	1	468	8	4052	2	28	0	23	0	0	5	150	0	0
567	0	766	3	3862	5	37	1	9	1	0	17	30	1	0
568	0	772	5	3090	1	36	1	16	0	1	10	80	1	0
569	1	540	15	1911	5	20	1	9	1	0	0,6	46	0	0
570	1	408	4	1414	3	23	0	20	1	1	0,6	1000	1	0
571	1	777	9	2333	1	20	1	18	0	0	0,6	200	1	0
572	0	954	12	3886	5	40	1	5	0	0	17	20	0	1
573	0	975	14	1784	3	48	0	11	1	1	16	15	0	0
574	0	997	8	4402	5	35	0	22	1	1	10	170	1	1

Окончание табл. 1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
575	1	1040	12	1758	5	38	0	15	1	1	17	50	1	0
576	1	972	6	1962	5	25	1	14	1	1	3	70	1	0
577	1	804	12	2233	2	47	0	14	1	1	22	65	1	0
578	1	725	7	4549	1	49	1	21	0	0	12	10	0	0
579	1	558	4	3029	4	26	1	24	1	1	3	20	0	0
580	1	475	6	4242	4	48	0	10	0	0	17	50	1	1
581	1	713	4	4761	4	42	1	22	1	1	18	100	1	0
582	0	588	6	2906	4	36	1	22	1	1	10	45	1	0
583	1	642	15	518	5	54	0	23	1	1	9	150	1	0
584	0	343	14	827	5	53	1	25	0	0	20	30	0	0
585	1	791	9	812	2	21	1	12	1	1	0,6	80	1	0
586	1	558	13	1659	5	41	1	10	1	0	5	46	0	0
587	0	949	14	698	2	46	0	7	1	1	6	100	1	0
588	1	990	8	3332	1	29	1	19	0	0	5	45	0	0
589	0	1100	12	4274	5	30	1	24	0	1	5	150	1	1
590	0	544	11	3600	4	53	0	13	1	1	11	30	1	0
591	0	1113	15	2233	2	44	1	16	1	1	13	100	1	1
592	0	741	1	2950	2	44	1	5	1	1	5	45	1	0
593	0	1079	1	3841	5	54	0	13	0	1	7	150	1	1
594	1	454	12	3838	4	48	0	23	1	0	10	30	0	0
595	0	929	5	4065	2	22	1	11	1	1	0,6	80	1	0
596	0	531	6	1126	3	24	1	4	1	0	1	46	0	0
597	0	1070	1	4777	4	28	1	25	0	0	5	1000	1	0
598	0	643	5	576	4	32	0	10	1	1	10	200	0	0
599	1	608	6	4816	3	54	1	5	1	0	9	15	1	1
600	0	466	3	817	5	36	0	13	1	1	10	25	1	1

Источники: составлено автором на основе [1].

x_1 — брался ли кредит (1 — да, 0 — нет);

x_2 — среднемесячный доход семьи заемщика, руб.;

x_3 — период погашения кредита, лет;

x_4 — размер рассматриваемого к выдаче кредита, руб.;

x_5 — состав семьи заемщика, чел.;

x_6 — возраст заемщика, лет;

x_7 — пол заемщика (1 — мужской, 0 — женский);

x_8 — срок проживания заемщика по последнему адресу, лет;

x_9 — наличие/отсутствие долгов по ранее взятому кредиту (1 — есть долг, 0 — нет долга);

x_{10} — наличие автомобиля в собственности (1 — есть, 0 — нет);

x_{11} — стаж работы, лет;

x_{12} — численность сотрудников в организации, в которой работает заемщик, чел.;
 x_{13} — наличие собственного жилья (1 — имеется, 0 — не имеется);
 x_{14} — вероятность погашения кредита.

Предварительно следует установить степень однородности данных. В качестве ключевого статистического признака среди 14 перечисленных выступает x_4 — размер рассматриваемого к выдаче кредита. Процедура установления однородности заемщиков по признаку размера выдаваемого кредита предполагает разбиение заемщиков на оптимальное число групп. Можно воспользоваться формулой Стерджесса [1].

$$n = 1 + [3,322 \cdot \lg N],$$

где n — число групп; N — общее число группируемых объектов (заемщиков).

По рассматриваемому массиву данных получим

$$n = 1 + 3,322 \cdot \lg 500 = 1 + [3,322 \cdot 2,778] = 10.$$

Таким образом, число групп составит 10, а совокупность заемщиков будет разбиваться по размеру выдаваемого кредита интервалами с равной длиной, соответствующей 500 руб. (максимальный размер выдаваемого кредита, составивший 5000 руб., делится на число групп или интервалов).

Для расчета частоты попадания заемщиков в заданные, равные по длине интервалы использовалась встроенная MS Excel статистическая функция ЧАСТОТА. Для этого сначала столбик был преобразован с помощью MS Excel в вариационный ряд по принципу возрастания значений признака. Полученные результаты представлены в табл. 2.

Таблица 2. Результаты разбиения заемщиков на интервалы по признаку размера рассматриваемого к выдаче кредита и частот их попадания в интервалы

Группа	Интервал выдаваемых кредитов, руб.	Частота попадания заемщиков в интервалы
1	1–500	6
2	501–1000	42
3	1001–1500	42
4	1501–2000	54
5	2001–2500	54
6	2501–3000	54
7	3001–3500	72
8	3501–4000	72
9	4001–4500	78
10	4501–5000	126

Источники: составлено автором.

Воспользовавшись табл. 2, можно рассчитать коэффициент неравномерности распределения заемщиков Джини, чем меньше его значение, тем более однородна совокупность статистического массива заемщиков. С этой целью построим вспомогательную табл. 3, в которой представляются два параметра: накопленные доли групп в общем числе групп (p_i) и накопленные доли частот попадания заемщиков в интервалы в общем числе заемщиков (q_i). Коэффициент неравномерности распределения Джини рассчитывается по формуле [1]

$$G = \sum_{i=1}^{n-1} p_i \cdot q_{i+1} - \sum_{i=1}^{n-1} p_{i+1} \cdot q_i.$$

Таблица 3. Накопленные доли групп в общем числе групп (p_i) и накопленные доли частот попадания заемщиков в интервалы в общем числе заемщиков (q_i)

p_i	q_i
0,1	0,01
0,2	0,08
0,3	0,15
0,4	0,24
0,5	0,33
0,6	0,42
0,7	0,54
0,8	0,66
0,9	0,79
1,0	1,00

Источник: составлено автором.

Неравномерность распределения составляет 0,256 или 25,6 %. Отсюда следует, что массив данных заемщиков по выделенному признаку является достаточно однородным. Геометрическая интерпретация коэффициента Джини выражается с помощью кривой Лоренца (рис. 1). Отношение площади фигуры, заключенной между кривой и диагональю квадрата, к площади прямоугольного треугольника и есть коэффициент Джини.

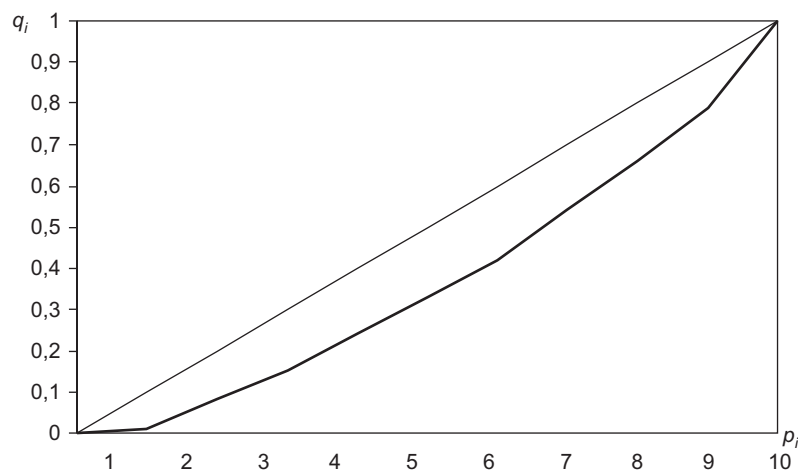


Рис. 1. Кривая Лоренца неравномерности распределения заемщиков коммерческого банка

Источник: разработано автором.

2. Снижение размерности исходного признакового пространства и интерпретация новых переменных. Методы стохастического факторного анализа позволяют, с одной стороны, уменьшить количество исходных характеристик (переменных), не снижая их информативности, а с другой стороны, установить интегральные качественные характеристики [2]. В этих целях применяется метод главных компонент, основная идея кото-

рого состоит в использовании показателя общей дисперсии, нормированной к единице (или 100 %). Главные компоненты представляют собой комплексные факторы (математические величины), формируемые в виде линейной комбинации исходных признаков-показателей. Каждая компонента объясняет долю дисперсии в общей вариации исходных показателей. В качестве математической величины, позволяющей установить долю дисперсии по главной компоненте, выступает собственное значение матрицы коэффициентов парной корреляции исходных переменных [3]. Сумма собственных чисел корреляционной матрицы равна следу матрицы (в корреляционной матрице главная диагональ содержит только единицы). Доли собственных значений и определяют доли дисперсий, обусловленных отдельными компонентами. Число главных компонент равно числу исходных переменных, однако лишь часть главных компонент берет на себя подавляющую суммарную долю дисперсии. Следовательно, количество значимых компонент окажется меньше, чем это необходимо. В свою очередь главные компоненты количественно отражают скрытые явления, поддающиеся экономической интерпретации [4].

Собственные значения корреляционной матрицы показателей кредитоспособности и объясненные ими доли общей дисперсии содержатся в табл. 4.

Таблица 4. Собственные значения корреляционной матрицы и объясненные ими доли общей дисперсии

Компонента	Собственное значение		
	Итого	% дисперсии	Кумулятивный %
1	2,059	14,706	14,706
2	1,665	11,893	26,599
3	1,566	11,183	37,782
4	1,440	10,288	48,070
5	1,290	9,214	57,284
6	1,140	8,144	65,427
7	,902	6,445	71,872
8	,858	6,125	77,997
9	,754	5,385	83,382
10	,613	4,376	87,758
11	,563	4,022	91,780
12	,525	3,746	95,527
13	,357	2,550	98,076
14	,269	1,924	100,000

Источник: составлено автором.

Из табл. 4 следует, что, например, на первую главную компоненту приходится около 15 % общей дисперсии. В свою очередь около 72 % общей дисперсии объясняют 6 главных компонент. Поэтому целесообразно сократить общее число компонент. Об этом свидетельствует график собственных значений (рис. 2), согласно которому для дальнейшей интерпретации остаются собственные числа со значениями больше единицы.

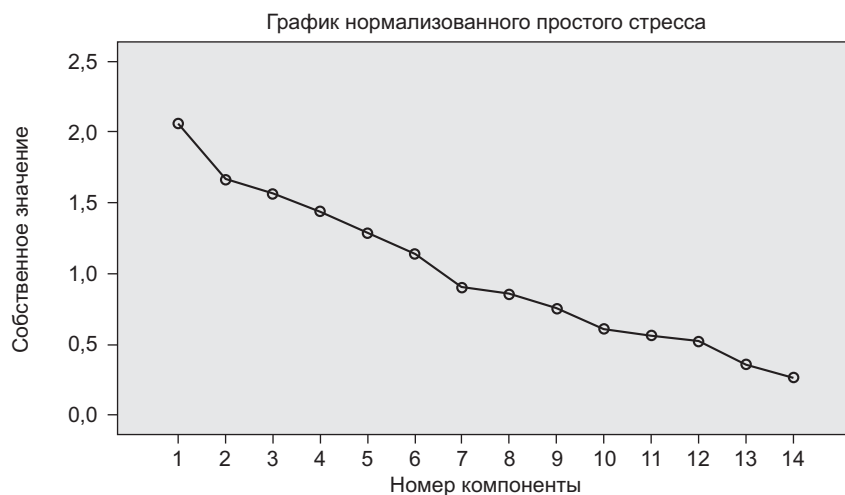


Рис. 2. График формирования собственных значений

Источники: разработано автором.

Для содержательной интерпретации главных компонент следует прибегнуть к данным матрицы факторных нагрузок, причем в соответствии с алгоритмом реализации метода главных компонент осуществляется ее вращение (получается так называемая матрица повернутых компонент) (табл. 5).

Таблица 5. Матрица повернутых компонент

	Компонента					
	1	2	3	4	5	6
x_1	-,255	,020	,318	,627	-,081	,302
x_2	,229	,020	-,128	-,230	,750	-,072
x_3	-,280	,050	-,243	,096	,164	,667
x_4	-,100	,034	-,072	,006	,455	-,598
x_5	,145	,124	-,070	,739	,054	-,084
x_6	-,024	,902	,018	,120	-,068	,027
x_7	-,125	-,057	,086	,186	,699	,054
x_8	-,044	,167	,711	,009	-,172	-,148
x_9	,357	-,027	,014	-,114	,002	,604
x_{10}	,820	-,103	-,139	-,023	-,016	,164
x_{11}	-,038	,902	,006	,002	,039	-,029
x_{12}	,250	-,297	,640	,048	,238	-,019
x_{13}	,791	,030	,171	,149	,043	-,075
y	,242	-,067	-,516	,548	-,007	-,228

Источники: составлено автором.

Столбцы матрицы повернутых компонент соответствуют значениям коэффициентов корреляции с главными компонентами. Чем ближе коэффициент по модулю к единице, тем выше значимость исходного показателя в формировании соответствующей главной компоненты [3]. Это обстоятельство позволяет объяснить их содержание. Например, сумма квадратов первого столбца равна первому собственному значению, т.е. 2,059. В свою очередь в этой сумме квадратов, характеризующей дисперсию первой главной компоненты, вызванной вариабельностью исходных признаков, весомая доля приходится на переменные x_{10}, x_{13} $((0,82^2 + 0,791^2) / 2,049 = 0,633$, т.е. около 63 %).

Вращение векторов-столбцов факторных нагрузок осуществлялось по критерию Вари-макс. Процедура вращения в SPSS сошлась за 9 итераций. В соответствии с данными табл. 5 экономическое содержание главных компонент, рассматриваемые как качественные интегральные характеристики кредитоспособности заемщиков, представлено в табл. 6.

Таблица 6. Экономическое содержание главных компонент

Главная компонента	Объясняющие исходные переменные	Экономическая интерпретация
F_1	x_{10}, x_{13}	Имущество заемщика
F_2	x_6, x_{11}	Возраст со стажем работы
F_3	x_8, x_{12}, y	Безопасность возврата кредита для банка
F_4	x_1, x_5, y	Кредитная история заемщика
F_5	x_2, x_7, x_4	Доход по половой принадлежности заемщика
F_6	x_3, x_4, x_9	Срочность кредита

Источники: составлено автором.

Основной вывод, следующий из проведенного анализа состоит в том, что для банка подтверждаются важные качества заемщиков, на основе которых принимается решение о выдаче кредита (см. табл. 6). К таким качествам относятся: имущество заемщика, возраст, безопасность возврата кредита для банка, кредитная история заемщика, наличие дохода у заемщика с учетом половой принадлежности и соблюдение срочности кредита.

3. Дискриминантный анализ заемщиков коммерческого банка. Используемые в анализе выборочные данные о заемщиках содержат информацию о принятом банком решении выдавать или не выдавать кредит. Возникает вопрос: насколько можно верить оценкам банка о кредитоспособности заемщиков? В соответствии с данными табл. 1 в столбце y (вероятность погашения кредита) представлены оценки банка своих предполагаемых заемщиков на основе 13 показателей (x_1, x_2, \dots, x_{13}). Другими словами, это оценки банка. Для того чтобы проверить справедливость банковских оценок целесообразно применение метода дискриминантного анализа. Разобьем весь массив заемщиков на большую и маленькую части. Пусть большая часть содержит первых 550 заемщиков, а малая — оставшихся 50. Вторая часть будет соответствовать так называемой экзаменуемой совокупности, а на большей части будет применяться метод линейного дискриминантного анализа как на обучающей совокупности. 550 заемщиков представим из двух выборок: первая выборка заемщиков, возвращающих кредит («хорошие»), вторая выборка — невозвращающие кредит заемщики («плохие»). Соответственно при проведении дискриминантного анализа дискриминантная функция или переменная выбора (по терминологии SPSS) будет строиться по принципу: заемщикам, попадающим в хорошую выборку присваивается 1, а в плохую — 2. Экзаменуемой совокупности заемщиков (по номерам 551–600) можно присвоить любое число, например 9.

Дело в том, что дискриминантный анализ представляет собой классификацию объектов путем их соотнесения (дискриминации) к заданным обучающим выборкам.

В результате получена дискриминантная функция с пошаговым отбором исходных дискриминантных переменных (табл. 7).

Таблица 7. Коэффициенты классифицирующей функции

Значимая переменная	Y выбора	
	Выборка 1	Выборка 2
x_5	1,930	1,453
x_8	,156	,202
x_9	2,785	3,427
x_{11}	,225	,277
x_{12}	–,001	,002
x_{13}	2,665	1,570
(Константа)	–8,471	–8,001

Источник: составлено автором.

Значимость дискриминантной функции и нормальность распределения заемщиков в двух обучающих выборках подтверждаются критерием Лямбда Уилкса и значимостью Хи-квадрат критерия [5]. Кроме того, пошаговый отбор переменных привел к исключению семи исходных переменных. Наиболее значимыми оказались: x_5 , x_8 , x_9 , x_{11} , x_{12} , x_{13} . Действительно речь идет о составе семьи заемщика; надежности клиента в смысле длительности проживания по последнему адресу; кредитной истории; масштабов организации, в которой работает заемщик; наличии собственного жилья. Они в основном соответствуют содержанию выявленных в подразделе 2 главными компонентам.

В табл. 8 сосредоточены результаты дискриминантного анализа. В первой колонке указаны номера последних 50 заемщиков от 551 до 600. Во второй колонке — возможное решение банка выдавать (1) или не выдавать кредит (0) (переменная x_{14}). Третья колонка содержит информацию о том, к какой выборке следует отнести заемщика, четвертая и пятая колонки показывают вероятность принадлежности к «хорошим» или «плохим» клиентам. В последней колонке знаком «звездочка» отмечены клиенты, по которым банк выдал ошибочную оценку. В соответствии с используемой методикой дискриминантного анализа граничной вероятностью дискриминации объектов выступает число 0,5. Например, по клиенту № 564 вероятность оказаться в «хорошей» группе составляет 0,5095, а в плохой — 0,4905. Поэтому клиент соотнесен к первой выборке. Согласно данным последней колонки табл. 8 расчеты позволяют выявить ошибочные оценки банка в выдаче кредита. Общее количество ошибочных оценок по последним 50 заемщикам составило 21. Причем если принимать во внимание наши рекомендации по 14 клиентам, банк заготовил решение не выдавать кредит, тогда как их следовало кредитовать. В оставшихся 7 случаях банку не следовало принимать решение о кредитовании клиентов.

Таблица 8. Результаты дискриминантного анализа

Заемщики	Выдавать/не выдавать, x_{14}	Принадлежность к выборке	Вероятность «хороший» (1)	Вероятность «плохой» (2)	Ошибки банка
1	2	3	4	5	6
551	0	2	0,45658	0,54342	
552	0	1	0,69142	0,30858	*
553	1	2	0,37967	0,62033	*

1	2	3	4	5	6
554	0	2	0,11774	0,88226	
555	0	1	0,72257	0,27743	*
556	0	2	0,24242	0,75758	
557	1	2	0,43013	0,56987	
558	0	2	0,05532	0,94468	*
559	1	1	0,76572	0,23428	
560	0	1	0,78131	0,21869	*
561	1	1	0,72479	0,27521	
562	1	2	0,35934	0,64066	*
563	1	1	0,68977	0,31023	
564	0	1	0,50954	0,49046	*
565	1	1	0,72628	0,27372	
566	0	2	0,18236	0,81764	
567	0	1	0,63288	0,36712	*
568	0	2	0,43264	0,56736	
569	0	1	0,63301	0,36699	*
570	0	2	0,0807	0,9193	
571	0	2	0,43978	0,56022	
572	1	1	0,64975	0,35025	*
573	0	2	0,28651	0,71349	
574	1	1	0,54176	0,45824	
575	0	1	0,70713	0,29287	*
576	0	1	0,85682	0,14318	*
577	0	2	0,29251	0,70749	
578	0	2	0,2155	0,7845	
579	0	2	0,43111	0,56889	
580	1	1	0,68291	0,31709	
581	0	2	0,45328	0,54672	
582	0	1	0,51425	0,48575	*
583	0	1	0,67261	0,32739	*
584	0	2	0,33923	0,66077	
585	0	1	0,61934	0,38066	*
586	0	1	0,63224	0,36776	*
587	0	1	0,55154	0,44846	*
588	0	2	0,29539	0,70461	
589	1	1	0,79294	0,20706	
590	0	1	0,5975	0,4025	*
591	1	2	0,37988	0,62012	*
592	0	1	0,62951	0,37049	*

Окончание табл. 8

1	2	3	4	5	6
593	1	1	0,87901	0,12099	
594	0	2	0,25598	0,74402	*
595	0	1	0,54472	0,45528	
596	0	2	0,41839	0,58161	
597	0	2	0,14081	0,85919	
598	0	2	0,29953	0,70047	
599	1	1	0,68314	0,31686	
600	1	1	0,70201	0,29799	

Источники: составлено автором.

4. Оценка кредитоспособности заемщиков с применением модели логистической регрессии. На основании выборочных данных рассчитаны вероятности получения или отказа в получении (дефолта) кредита по всем шестистам заемщикам. В качестве переменных, включаемых в модель логистической регрессии, использованы переменные, значимость которых была подтверждена в подразделе 3, т.е. $x_5, x_8, x_9, x_{11}, x_{12}, x_{13}$. Здесь выводятся значения ключевых характеристик логистической модели по последним 50 клиентам (551–600).

Согласно данным табл. 9 R квадрат Кокса и Снелла, R квадрат Нэйджелкерка являются мерами определенности. Они также как и при линейной регрессии указывают на ту часть дисперсии, которую можно объяснить с помощью логистической регрессии. Мера определенности по Коксу и Снеллу имеет тот недостаток, что значение равное 1 является теоретически не достижимым; этот недостаток устранен благодаря модификации данной меры по методу Нэйджелкерка. Часть дисперсии, объяснимой с помощью логистической регрессии, составляет 87 %, что говорит о хорошем качестве сформированной модели. Оценки коэффициентов логистической регрессии представлены в табл. 10.

Таблица 9. Сводка для модели

Шаг	Логарифм правдоподобия	R квадрат Кокса и Снелла	R квадрат Нэйджелкерка
1	631,675	,794	,870

Источники: составлено автором.

Таблица 10. Оценки коэффициентов при переменных в уравнении

		B	Стд. Ошибка	Вальд	ст. св.	Знач.	Exp(B)
Шаг 1 ^a	x_5	,483	,074	42,979	1	,000	1,620
	x_8	-,043	,011	14,091	1	,000	,958
	x_9	-,745	,209	12,771	1	,000	,475
	x_{11}	-,051	,017	9,064	1	,003	,951
	x_{12}	-,005	,001	9,961	1	,002	,995
	x_{13}	1,301	,233	31,288	1	,000	3,672
	Константа	-1,225	,399	9,409	1	,002	,294

Примечание: ^a — переменные, включенные на шаге 1: $x_5, x_8, x_9, x_{11}, x_{12}, x_{13}$.

Источники: составлено автором.

В соответствии с колонкой В табл. 10 детерминированная составляющая модели логистической регрессии примет вид

$$P_i = \frac{e^{-1,255+0,483x_5-0,043x_8-0,745x_9-0,051x_{11}-0,005x_{12}+1,301x_{13}}}{1 + e^{-1,255+0,483x_5-0,043x_8-0,745x_9-0,051x_{12}+1,301x_{13}}}, \quad i = \overline{1,600}.$$

Далее сформирована классификационная таблица, в которой наблюдаемые показатели принадлежности к группе (1 = «платежеспособный», 0 = «неплатежеспособный») противопоставляются предсказанным на основе расчетов по полученной логистической модели. Из табл. 11 можно сделать вывод о том, что из общего числа «неплатежеспособных» клиентов, равного 402, были признаны таковыми 364 (в таких случаях говорят о «строго положительных» результатах). Остальных 38 называют «ложно отрицательными»; они были признаны «платежеспособными», хотя являются «неплатежеспособными». Из общего числа «платежеспособных», равного 198, были признаны таковыми только 99 («строго отрицательные»), 99 признаны «неплатежеспособными», хотя они и являются «платежеспособными» («ложно положительные»). В общем, правильно были распознаны 463 случая из 600, это составляет 77,2 %, т.е. для 77,2 % объектов результаты прогноза оказались верными.

На основании данных табл. 11 можно рассчитать средний эффект от внедрения разработанного проекта по определенным соображениям. По оценкам банка 137 заемщиков не должны получить кредиты, что по проведенным модельным оценкам неправильно. Следовательно, не кредитуя их, банк упускает выгоду. Поскольку средняя процентная ставка по кредитам для физических лиц на период формирования выборочных данных составляла 25 % годовых, а размер кредита в среднем на одного заемщика 2544 руб./заемщик, упущенная выгода будет равна: $2544 \cdot 0,25 \cdot 137 = 87\,132$ руб. Другими словами, при использовании аргументированных нами модельными расчетами оценок кредитоспособности заемщиков банк мог дополнительно заработать 87 132 руб.

Таблица 11. Таблица классификации^а

		Предсказанные			
		x_{14}		Процент корректных	
Наблюдаемые		,00	1,00		
Шаг 1	x_{14}	,00	364	38	90,5
		1,00	99	99	50,0
	Общий процент				

Примечание: ^а — разделяющее значение = 0,5.

Источники: составлено автором.

В результате проведенного исследования получен ряд выводов.

1. Многомерное статистическое моделирование может выступить инструментом совершенствования используемых коммерческими банками методик оценки кредитоспособности заемщиков. Для этого целесообразно использование методов дискриминантного и факторного анализа, а также логистической регрессии, которые в конечном счете приводят к получению вероятностной меры выдачи кредита и служат количественному обоснованию банковских решений.

2. Предварительный аналитический обзор выборочной совокупности данных о заемщиках предполагает установление степени однородности данных, что открывает возможности применения методов прикладного статистического анализа. В частности,

низкое значение коэффициента Джини с графической иллюстрацией кривой Лоренца подтверждает однородность выборки, включающей 600 заемщиков по показателю объема выданных кредитов.

3. Метод главных компонент, численная реализация которого осуществлено в SPSS, позволил снизить количество исходных 14 признаков до шести. Основным вывод, следующий из проведенного анализа состоит в том, что для банка подтверждаются важные качества заемщиков, на основе которых принимается решение о выдаче кредита. К таким качествам относятся имущество заемщика, возраст, безопасность возврата кредита для банка, кредитная история заемщика, наличие дохода у заемщика с учетом половой принадлежности и соблюдение срочности кредита.

4. Проверку банковских оценок заемщиков целесообразно проводить с применением дискриминантного анализа. Для этого всю выборочную совокупность из 600 заемщиков следует разбить на опытную и экзаменующую группы, соответственно включив в каждую из них 550 и 50 заемщиков. Опытную группу необходимо классифицировать из двух обучающих выборок: «хорошие» и «плохие» заемщики. Проведенный дискриминантный анализ показал, что по экзаменующей группе из 50 заемщиков по 21 клиенту банком принимались неправильные решения.

5. Модель логистической регрессии следует использовать для оценки правильности решений банка о выдаче займов. Согласно модельным расчетам правильными были распознаны 463 случая из 600, что составляет 77,2 %. Это подтверждается рассчитанными вероятностями целесообразности и нецелесообразности выдачи кредитов заемщикам.

Источники

1. Казаринова, С. Е. Теория статистики. Справочные материалы, практические и контрольные задания, тесты / С. Е. Казаринова. — М. : МГУ, 2009. — 490 с.

Kazarinova, S. E. Theory of Statistics. Reference materials, practical and control tasks, tests / S. E. Kazarinova. — Moscow : MSU, 2009. — 490 p.

2. Бююль, А. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : пер. с нем. / А. Бююль, П. Цефель. — СПб. : ДиаСофтЮП, 2005. — 608 с.

Byul, A. SPSS: the art of information processing. Analysis of statistical data and the restoration of hidden patterns : transl. from Germ. / A. Byul, P. Tsefel. — St Petersburg : DiaSoftUP, 2005. — 608 p.

3. Айвазян, С. А. Прикладная статистика и основы эконометрики : учебник для вузов / С. А. Айвазян, В. С. Мхитарян. — М. : ЮНИТИ-ДАНА, 1998. — 1022 с.

Aivazyan, S. A. Applied Statistics and Essentials of econometrics : textbook for high schools / S. A. Aivazyan, V. S. Mkhitarian. — Moscow : UNITY-DANA, 1998. — 1022 p.

4. Читая, Г. О. Факторный анализ промышленного развития макрорегионов России / Г. О. Читая // Вопр. статистики. — 2006. — № 2. — С. 19–28.

Chitaya, G. O. Factor analysis of the industrial development of macroregions of Russia / G. O. Chitaya // Questions of statistics. — 2006. — № 2. — P. 19–28.

5. Матальцкий, М. А. Теория вероятностей и математическая статистика : учебник / М. А. Матальцкий, Г. А. Хацкевич. — Минск : Выш. шк., 2017. — 591 с.

Matalytsky, M. A. Probability theory and mathematical statistics : textbook / M. A. Matalytsky, G. A. Khackevich. — Minsk : Higher School, 2017. — 591 p.

Статья поступила в редакцию 20.12.2018 г.