

# Correspondence Analysis in the Banking System

Irina Ioniță, Mădălina Cărbureanu

Petroleum-Gas University of Ploiești, Informatics Department, 39 Bucharest Bd., Ploiesti, Romania  
e-mail: tirinelle@yahoo.com, mcarbureanu04@yahoo.com

## Abstract

*Lending has always been an issue of broad interest to banks wishing to maintain a balance in a world of continuous competition. Various tests and case studies presented in literature highlight the importance of knowing the problems that may arise in a loan department. These concerns with banking strategies aimed at developing the best possible decision on granting loans in accordance with restrictions imposed by the Central Bank, and by each branch separately. Identifying those elements which directly and significantly affect a decision in favor of granting a bank loan is as laborious as it is important. In this paper we use the SAS tool for analyzing the correspondences between variables that describe the granted loan process. The case study presented in this paper aims to highlight the correlations between different variables considered for the analysis of consumer loan problem.*

**Key words:** *correspondence analysis, loan approval, knowledge management, SAS, decision*

**JEL Classification:** *C44, D81*

## Introduction

Exploration of data provided by an organization during a significant period of time has become an ongoing concern, assuming engagement of an expert's team in a remarkable work for knowledge discovery process that can favorably influence the development of the target organization. Knowledge management has seen a considerable expansion in recent years in business, which involves better coordination of activities in organizations and increasing benefit with reduced costs.

In banking, lending has always been a topic of wide interest. The main objective of banks is to maintain a balance in a world of continuous competition and to make profit. Knowledge management is present in the banking sector as well, as confirmed by different analyses and case studies described in the literature. Knowing the problems that might arise within a loan department is important. These concerns in the banking system aim to develop the best possible decision making strategies as regards loan granting in accordance with the restrictions imposed by the Central Bank, and by each branch separately. Analyzing a loan application means taking into account many factors that influence to a greater or lower extent the final response received by the bank customer. Identifying those elements with a high degree of importance whose presence in the lending description model directly and significantly affects a favourable decision, is as laborious work as it is important.

In this paper we use the SAS tool for analyzing the correspondences between variables that describe the credit model. The case study presented aims to reveal correlations between different variables considered for the description of the analyzed problem.

## The Issue of Granting Loans

One high-risk area within a bank is the bank lending sector. The evolution of banking environment is fluctuating and directly influences the development of banks and leads them to improve their strategies to adapt to permanent changes imposed by economic crisis, focusing on knowledge management.

Credit banking involves risks, which, unchecked, can lead to cases of bank failure such as bankruptcy. The decision of banks to grant credit to the new clients implies accountability by guaranteeing the loan, and a reciprocal relationship between the client and bank, with the aim of returning the credit with its costs. If the first part of the hypothesis is confirmed, in many cases customers can not repay the borrowed money on time so there is a deficit in banking institution functioning. Knowledge management at this level requires better control over the factors influencing the decision to grant a loan and identifying those features with key role in decision making process. Lending decisions are directed to avoid risk. The need to develop objective decisions, to make a good classification of potential customers, to properly handle situations arising in the context of loan approval involves informatics. The adoption of modern methods in banking sector generates two problems (Basno 2002, pp.112):

- designing a system for quantifying and ranking the conditions and prerequisites of credit;
- implementing an information system and electronic data acquisition to ensure objective ranking of each application.

With a bank history, namely a database of bad loans, through the application of modern data mining techniques we can make predictions on the evolution of loans, formulate the criteria that determine an association between major risk or credit, achieve a profile of the good customer.

Grant credit should be based on a set of rules, which clearly define the principles of lending. These principles and objectives of the credit must be dynamic and adaptable to changing economic environment and market specifics. Lending principles may be summarized by the six C's lending (Basno 2002, pp. 109):

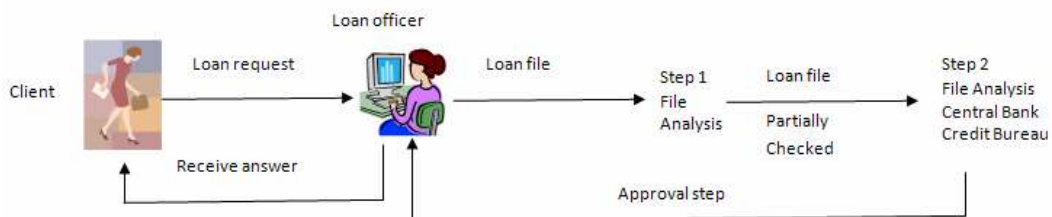
- Character - refers to honesty, integrity and credibility of the borrower, as it can and customer willingness to repay credit;
- Capital - indicates presence or absence of financial resources in the short or long, which can be used to settle the obligations if the immediate future availability is not enough;
- Capacity - refers to the presence or absence of current resources that can be used for the repayment due date;
- Collateral - refers to the presence or absence and amount of assets that can be used to settle obligations if payments are not made under a payment schedule or the client is in default,
- Conditions - represent the general economic environment and conditions (both the creditor and the debtor) that may affect the client's ability to repay the bank loan or credit.
- Compliance with legal norms.

The present work discusses about consumer loans. In the process of granting a loan there have been identified several stages, as follows:

- the customer requests a loan to a bank;
- the customer achieves an initial credit scoring made by a loan officer, as a result of querying to collect the preliminary data (amount requested, period, average monthly income, length of service, maintenance persons in etc.);

- in the favorable case, the client obtains a high score, and it must return with supporting documents to complete the file (certificate of pay, tax statement, the last utility bill for verifying the applicant etc.);
- the customer completes a credit application form;
- file analysis is made that includes checking the customer in the credit bureau database, verifying the accuracy of customer data etc.;
- expanding credit agreement and completion of an insurance certificate, if requested or required;
- approval phase;
- create an account and transfer money into the account created.

A schematic representation of the corresponding information flow process of granting a bank loan is shown in Figure 1.



**Fig. 1.** The grant loan process

Each bank reserves the right to establish minimum criteria for granting credit, such as minimum age, maximum age, minimum wage, age at last job etc. Analyzing existing data (stored in a database of customers), a bank shall establish a customer profile, which can be changed depending on the evolution in the banking environment. The loan department of a bank, establishing the pattern of lending, involves a complex activity. Banking experts have developed methods and techniques for assessing the creditworthiness of a credit applicant, resulting in credit scoring (Hand 1997, pp. 523, Lyn 2000, pp.149, Olteanu, 2003).

Credit scoring is to determine the creditworthiness of a bank's client, represented by calculating a score. In this case, are necessary:

- to establish a number of variables that characterize a customer in financial and non-financial terms;
- a system to implement aggregation variables.

Given the current situation of the Romanian banking system, banks have started to make some compromises on loans approval. For now, loan applicants would not be automatically denied from getting a loan only because they appear to have a debt to pay a loan. By implementing a scoring model, the Credit Bureau provides banks requesting a "note" for each customer, based on a set of variables that takes into account all available information in the database, not just negative.

The main indicators used to determine scoring are:

- payment history, or that if it was not outstanding;
- current rate, the total amount that has to be repaid in the form of loan and interest;
- ongoing credit types: mortgage, consumer, credit card etc.

Calculation methodology is based on international statistical model applied to data from Credit Bureau. Credit file analysis takes into account several factors whose values are provided in the credit application form. These factors may have a more pronounced or not significant influence on the calculation method of scoring. Thus, to obtain results as conclusive, the processing of

which will influence the decision to grant credit, it seeks to eliminate those factors with the minimum degree of influence.

The aggregation process may cause the lost of features that may play an important role in analyzing credit file, and by clearing may be considered insignificant factors that may lead to the formation of a false positive image of the credit applicant, involving major risks. The question which arises is to optimize the aggregation process and to consider the most significant factors. In the next section of the paper we present a descriptive analysis method using SAS software.

## SAS Correspondence Analysis

To achieve a good analysis of a great amount of data, it is necessary the usage of some powerful statistical tools. Such an example is SAS, because it offers very powerful instruments for data management, a large gamut of statistical analysis procedures and a batch of graphical procedures, some of these being presented in paper number three from references (Cărbureanu 2008, pp. 57). The statistical data used in the proposed application was procured from the address mentioned at point one from references and it describes the recordings of a bank about the customers who asked for a credit. The data was processed in a number of steps, into an Excel file, being finally codified for the application in SAS correspondence analysis.

The case study objective is to examine if there can be any correspondence between each of the variables: *property* (estates, cars etc.), *loan\_history* ( history of payments), *period* (the loan period), *purpose* (education, training etc.), *loan\_amount* (the loan value), *years\_employed* (the years on service), *residence* (the years of residency), *age*, *house* (property or rented house etc.), *job* (occupation), *children* (maintenance children in) and the variable *approve* (the answer for the loan request ), considered to be the target variable. To reach the proposed objective we will use the correspondence analysis implemented in SAS with the help of *corresp* procedure (SAS).

The correspondence analysis achieves a simple correspondence analysis, which can be used to analyze the frequencies of data and the associations between two or more nominal variables. (10). It can be applied when we have at disposal contingency tables of  $(k_{ij})$  form where  $k_{ij}$  represents the observations' frequencies (Colonic 2007, pp.109).

A primary study consists in the correspondence analysis for the *property* (estates, cars, etc.) and *approval* (loan approval) factors. The possible values (categories) for *property* variable are integer values between 0 and 3, with the following significance:

- 0 – *unknown* (it is unknown the fact that the respective person has any property);
- 1 – *car* ( the person owns a car, personal property);
- 2 – *real estate* (the person owns real estates and lands);
- 3 – *savings* (the person savings).

The possible values (categories) for *approve* variable are integer values between 0 and 2, with the following meaning:

- 0 – *no* (the loan is not granted);
- 1 – *yes* (the loan is granted);
- 2 – *waiting* (the person who asked the loan is on a waiting list).

Applying the *corresp* procedure it is generated a contingency table, which indicates the absolute frequencies of the persons number who own properties (car, real estate, savings, unknown) for each possible case (the loan is not approved, is approved, waiting list), as well as their sums on lines and columns.

**Table 1.** Contingency Table

	no	waiting	yes	Sum
car	4	0	8	12
real estate	1	2	6	9
savings	1	0	4	5
unknown	3	0	1	4
Sum	9	2	19	30

The correspondence analysis allows the points representation (the contingency table column and line vectors) in spaces of a reduced dimension. The viewing of these points on the same graphic makes possible the achievement of some interpretations regarding the association level between the categories of these two variables. Those two extracted dimensions (Table 2) cumulate 100% percent from *chi-square* statistics, fact that shows a very good representation in two-dimensional space, as it can be observed in Table 3.

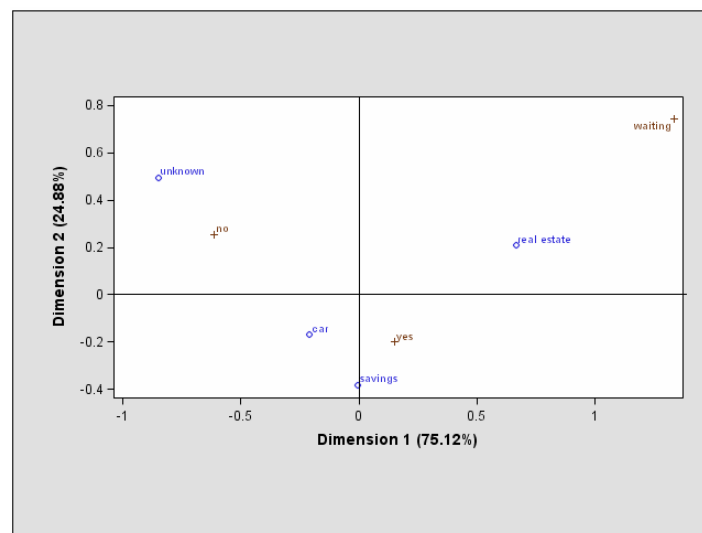
**Table 2.** The extracted dimensions

	Dim1	Dim2g
car	-0.2105	-0.1668
real estate	0.6623	0.2121
savings	-0.0051	-0.3793
unknown	-0.8523	0.4972

**Table 3.** Chi-Square statistic

Singular value	Principal Inertia	Chi-Square	Percent	Cumulative Percent
0.49618	0.24619	7.38572	75.12	75.12
0.28558	0.08155	2.44664	24.88	100.00
Total	0.32775	9.83236	100.00	

Figure 2 represents the categories for *property* and *approval* variables reported at these two extracted dimensions.

**Fig. 2.** The variables categories representation

For axis one (dimension one), the points which determine this axis, namely the points which are most distant from centroid (the point of (0, 0) coordinates) are on the one side the *unknown* and *no* categories, and on the other side the *real estate* and waiting categories.

It can be seen that dimension one highlights the association between the state of unawareness regarding the respective person's properties (*unknown*) and the refusal of the solicited loan (*no*). At the same time, it can be observed the association between the ownership of real estates and lands (*real estate*) and the writing on a waiting list of the persons who had asked for a loan (*waiting*).

For the second dimension, we can remark the association between the financial status of the person who wants the loan (*savings*) and the loan approval (*yes*).

Using the same reasoning as fore mentioned, the following results presented in Table 4 and 5 have been obtained:

**Table 4.** Variables association

Variables association	DIM 1		DIM 2	
	property	approve	property	approve
property-approve	unknown	no	savings	yes
	real estate	waiting	-	-
loan_history-approve	<b>loan_history</b>	<b>approve</b>	<b>loan_history</b>	<b>approve</b>
	ok_at_this_bank	yes	past_delays	waiting
	critical	no	-	-
period-approve	<b>period</b>	<b>approve</b>	<b>period</b>	<b>approve</b>
	long	no	medium	yes
	short	waiting	-	-
purpose-approve	<b>purpose</b>	<b>approve</b>	<b>purpose</b>	<b>approve</b>
	education, television	yes	furniture	waiting
	business	no	-	-

**Table 5.** Variables association

Variables association	DIM 1		DIM 2	
loan_amount-approve	<b>loan_amount</b>	<b>approve</b>	<b>loan_amount</b>	<b>approve</b>
	medium	yes	small	waiting
	big	no	-	-
years_employed-approve	<b>years_employed</b>	<b>approve</b>	<b>years_employed</b>	<b>Approve</b>
	<7	yes	<1	waiting
	unemployed	no	-	-
residence-approve	<b>residence</b>	<b>approve</b>	<b>residence</b>	<b>approve</b>
	<=15	waiting	>15	yes
	<10	no	-	-
age-approve	<b>age</b>	<b>approve</b>	<b>age</b>	<b>approve</b>
	<30	no	>40	yes
	<=40	waiting	-	-
house-approve	<b>house</b>	<b>approve</b>	<b>house</b>	<b>approve</b>
	free	no	own	yes
	rent	waiting	-	-

Table 5 (cont.)

job-approve	<b>job</b>	<b>approve</b>	<b>job</b>	<b>approve</b>
	skilled	yes	management	no
children-approve	unskilled	waiting	-	-
	<b>children</b>	<b>approve</b>	<b>children</b>	<b>approve</b>
	1child	no	no child	yes
	2children	waiting	-	-

From the analyzed data for the presented case study it can be generated a batch of rules which describe the favourable and unfavourable situations regarding the bank response to the client application. The utility of this batch of rules can consist in the development of an expert system whose goal is to evaluate the banking loan approval conditions for personal needs.

The future research will aim to design an expert system which will assist the decisional process from the credit bank department, involving the analysis of other types of banking credits.

## Conclusions

The subject discussed in this paper highlights the utility of descriptive analysis, in fact the utility of correspondence analysis in the banking sector. The product used for the application is the SAS statistical tool. The results obtained for the proposed case study provide insight into the conditions for granting a consumer loan.

Expert systems can be applied in banking on various plans in determining the decision regarding the credit, in financial analysis relating to the loan applicant, in optimizing the loan portfolio, on the banking market place. An expert system, based on a history of bank loans can be used in the pre-selection phase, preparing the decision, offering an objective support for decision making. In most cases, the final decision belongs to human decision-makers. Future work will focus on developing such an expert system to assist decision making in the banking sector in order to minimize credit risk.

## References

1. \*\*\* Correspondence analysis, <http://support.sas.com/documentation/cdl/en/imlsug/62558/HTML/default/ugmultca.htm> [accessed 15 February 2010].
2. \*\*\* SAS, <http://www.sas.com/technologies/analytics/statistics/stat/> [accessed 10 March 2010].
3. Basno, C., Dardac, N., *Management bancar*, Editura Economică, București, 2002.
4. Cărbureanu, M., An Application in Social Domain using SAS, *Petroleum-Gas University of Ploiesti Bulletin, Mathematics-Informatics-Physics Series*, Vol. LX, No. 1, 2008.
5. Chuang, C.L., Lin, R.H., Constructing a reassigning credit scoring model, *Expert System with Application: An International Journal*, Vol. 36, Issue 2, 2009, pp. 1685-1694, <http://www.sciencedirect.com/science>.
6. Clocotici, V., *Introducere în statistica multivariată*, <http://profs.info.uaic.ro/~lavinia/SUPPORTURI%20DE%20CURS%20IDD/SEMESTRUL%20I/StatIDD.pdf> [accessed 10 April 2010].
7. Hand, D.J., Heney, W.E., Statistical Classification Methods in Consumer Client Scoring: A Review, *J. R. Statist. Soc.*, 160, 1997, pp. 523-541.
8. Lyn, C. T., A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers, *International Journal of Forecasting*, vol. 16, Issue 2, 2000, pp. 149-172, <http://www.sciencedirect.com/science>.
9. Olteanu, A., Olteanu, F.M., Badea, L., *Management bancar. Caracteristici, strategii, studii de caz*, Editura Dareco, București, 2003.
10. <http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/> [accessed 15 April 2010].

## O aplicație a analizei corespondențelor în domeniul bancar

### Rezumat

*Acordarea creditelor a constituit dintotdeauna un subiect de larg interes pentru băncile care doresc să-și mențină echilibrul într-o lume a unei competiții continue. În literatura de specialitate se întâlnesc diverse analize, studii de caz ce evidențiază importanța cunoașterii problemelor ce pot apărea în cadrul unui departament de credite. Aceste preocupări în domeniul bancar au ca scop elaborarea unor strategii decizionale cât mai bune privind acordarea de credite, în concordanță cu restricțiile impuse de Banca Centrală, precum și de fiecare sucursală în parte. Identificarea acelor elemente care influențează în mod direct și semnificativ o decizie favorabilă de acordare de credit bancar este pe cât de laborioasă, pe atât de importantă. În acest articol se utilizează instrumentul SAS pentru analiza corespondențelor dintre variabilele cu rol decisiv în procesul de acordare a creditelor. Studiul de caz prezentat are ca scop evidențierea corelațiilor ce se pot stabili între diferitele variabile luate în considerare pentru descrierea problemei analizate.*