INTERACTING LATENT BUDGET ANALYSIS AND CORRESPONDENCE ANALYSIS TO ANALYSE BEAUTY SALON MANAGEMENT DATA

Enio Galinkin JELIHOVSCHI[1]
Rosália Rodrigues ALVES[2]
Fábio Mathias CORRÊA[1]

▪ *ABSTRACT: Latent budget analysis (LBA) and correspondence analysis (CA) are used interactively to analyse data about managment of beauty salons. The survey consists of basically two types of questions; the first identifies the profile of the owner manager, the second are questions referring to the degree of professionalism with respect to market and finances. We built a stacked table where the rows are the profile questions and the columns the market-financial ones. The above methods are used to analyse this table. CA is used at first to find the important answers with respect to their mass followed by a LBA, then the graph results of CA are used to describe the latent budgets and mixing parameters. Either two or three latent budgets divide salons between low and higher level of professionalism, as it was expected.*

▪ *KEYWORDS: Correspondence analysis; latent budget analysis; market surveys of beauty salons.*

1 Introduction

Latent budget analysis (LBA) is a method for the analysis of contingency tables, as well as correspondence analysis (CA). LBA is used to understand the relation between rows and columns of the table whenever the rows are explanatory and the columns are response variable. Both LBA and CA look at the same matrix.

[1]Universidade Estadual de Santa Cruz – UESC, Departamento de Ciências Exatas e Tecnológicas — DCET, Área de Estatística, Campus Soane Nazaré de Andrade, CEP: 45662-900, Ilhéus, BA, Brasil. E-mail: *eniojelihovs@gmail.com* / *fmcron@gmail.com*

[2]Universidade Federal de Lavras – UFLA, Departamento de Ciências Exatas – DEX, Caixa Postal 37, CEP: 37200-000, Lavras, MG, Brasil. E-mail: *rosaliaadm@gmail.com*

In CA it is called row profile matrix and in LBA matrix of conditional probability of the response given the explanatory variable, or compositional data. The LBA allows us to find which categories of the response are related to different groups of the explanatory categories. If the table has a product multinomial distribution we can understand the latent budget model (LBM) as explaining the relationship between the explanatory and the response variables assuming that conditioned on the latent variable they are independent. In that sense, the latent budgets, which are categories of a latent variable, are hidden values which explain the relationship between the explanatory and response variables. Both methods LBA and CA, reduce the dimensionality of the original problem, thus making it easier to understand its hidden relations.

In the empirical data set described above in the abstract, the profile questions are the explanatory variables whereas the market-financial are the response ones. The main quest of this article is to find a hidden (latent) variable which explains away the question of professionalism of the beauty salons managers. In other words, we want to find out whether it is possible to attribute the relationship between explanatory and response variable to another variable which is latent, that is, not observable. We will see that this variable can be called professionalism with two classes, the low and high level. They will be called **latent budgets**.

In this article we first use multiple correspondence analysis (MCA) to identify the important categories of the various questions of the survey with respect to their masses, that is the proportion of respondents who answered that category, and built a stacked matrix (Greenacre, 2007, chap.17), where the categories of the explanatory variable are the rows and the categories of the response variable the columns. This matrix is built by tabulating each explanatory question with each response question and then stacking them in a way such that the profile questions become the rows and profissionalism questions the columns. This matrix is used as the table for LBA.

The solution of LBM is in form of matrices one for the mixing parameters and the other for the latent components where the rows are either the explanatory or the response categories and columns are the corresponding latent components. McCutcheon (1998), use CA as an aid to Latent Class Analysis and **?** use CA also as an aid to LBA. The former use the matrix of latent class probabilities to plot a CA graph, the latter uses the matrix of the maximum likelihood estimates of the expected budgets to plot a CA graph and plot the latent budgets by projecting them as suplementary points in the CA space. In our case, we use CA to plot a slightly modified version of both the matrix of mixing parameters and the matrix of latent components. Those graphs are then used as a tool to the interpretation of the model. When the number of categories of either explanatory and response variables are large those graphs are indeed very helpful.

Graphical results are much nicer and easier to interpred than tables and the CA results are, in this regard, very representative and go further in the understanding of the many relations among the variables in the data set.

2 Data and methods

2.1 LBM concepts and definitions

LBM is a mixture model for compositional data. A row of a compositional data is called *composition* or a *budget* and its elements are the *components*. We will follow the notation of **?** and as he says: "by performing LBA we aproximate I observed budgets, which may represent, persons, groups or objects, by a small number of latent budgets, consisting of typical characteristics of the sample. The idea of the LBM was proposed by Goodman (1974), and elaborated by Clogg (1981) by interpreting a simple latent class model in asymetric way. Independently, de Leeuw and Van der Heijden (1988) introduced the model and named it LBM because they used it to analyse time budget data". See also de de Leeuw et al. (1990) and **?**.

The original contingency table is $N(I,J)$. Let us also define:

- row total: $n_{i+} = \sum_j n_{ij}$.
- column total: $n_{+j} = \sum_i n_{ij}$.
- total: $n = \sum_j \sum_i n_{ij}$.

2.1.1 The latent budget model

The compositional data matrix P is formed by dividing the raw data by their corresponding row total and let us call the observed components $p_{i|j}(i = 1,\ldots,I; j = 1,\ldots,J)$ then, $p_{i|j} = \frac{n_{ij}}{n_{i+}}$, $p_{i+} = \frac{n_{i+}}{n}$ and $p_{+j} = \frac{n_{+j}}{n}$.

Each row vector p_i of P is called observed budget and is aproximated by the expected budget π_i which is a mixture of $K(K \leqq min(I,J))$ latent budgets.

The row vectors π_i $(i = 1,\ldots,I)$ form the expected matrix Π which has a lower rank and, in LBM, approximates P.

The latent budgets are represented by β_k $(k = 1,\ldots,K)$ and the model is written as

$\pi_i = \alpha_{1|i}\beta_1 + \ldots + \alpha_{k|i}\beta_k + \ldots + \alpha_{K|i}\beta_K$, where $\alpha_{k|i}$ are the mixing parameters.

The elements of Π are $\pi_{j|i}$ and are called *expected components*.

The elements of β_k are $\beta_{j|k}$ called *latent components*. In scalar notation, $\pi_{j|i} = \sum_{k=1}^{K} \alpha_{k|i}\beta_{j|k}$, and in matrix notation $\Pi = AB^t$ where Π is an IxJ matrix whose rows are the expected budgets. A is an IxK matrix of mixing parameters and B is a JxK matrix whose columns are the latent budgets. LBM(K) is then the latent budget model with K latent budgets. Similar to the observed components, the parameters of LBM are subject to the sum constraints $\sum_{j=1}^{J} \pi_{j|i} = \sum_{k=1}^{K} \alpha_{k|i} = \sum_{j=1}^{J} \beta_{j|k} = 1$ and the nonnegativity constraints $0 \leqq \pi_{j|i}, \alpha_{k|i}, \beta_{j|k} \leqq 1$. In this way, all parameters are proportions what further facilitates the the interpretation of the model.

Quoting **?** "The latent budgets can be characterized by being compared to the latent budgets of LBM(1). LBM(1) is the independence model with $\alpha_{1|i} = 1$ and $\beta_{j|1} = p_{+j}$, in this case $\pi_i = \beta_1$. Hence, if latent component $\beta_{j|k} \geqq p_{+j}$, then

β_k is characterized by the j-th category. On the other hand, if $\beta_{j|k} \leqq p_{+j}$, then the j-th category is of lesser importance. The relative importance of each latent budget, in terms of how much of the expected data they account for, is expressed by the budget proportions

$$\pi_k = \sum_i p_{i+}\alpha_{k|i} \tag{1}$$

π_k also denotes the probability of latent budget k when there is no information about the level of the row variable. To understand how the expected budgets are construted from the latent budgets, we must compare the mixing parameters to π_k. If $\alpha_{k|i} \geqq \pi_k$ then the expected budget π_i is characterized by more than average by latent budget β_k, otherwise, if $\alpha_{k|i} \leqq \pi_k$ then the expected budget π_i is characterized by less than average by latent budget β_k. In practice, the mixture model interpretation is easier to carry out when we first characterize the latent budgets and then interpret the expected budgets in terms of them."

LBA can be either confirmatory or exploratory depending on the original table being a result of a product multinomial distribution or not. The product multinomial distribution occurs when the n_{i+} observations of the i-th row $(n_{i1}, \ldots, n_{iJ})$ are independent, all with probability distribution $(\pi_{1|i}, \ldots, \pi_{J|i})$, and the I rows are also independent (Agresti, 1990). In our case we can only use LBA as a exploratory analysis, since the stacked matrix used does not have independent rows; each respondent answers every question therefore, there is no independence among the questions.

Compositional data that follows the product multinomial sampling scheme may be estimated by maximum likelihood estimation method (MLE), on the other hand, if the data cannot be assumed to follow that scheme above it is not recommended to use the MLE. Following **?** we opted to use weighted least squares estimators since it is a distribution free method, that is, one which does not assume any probability configuration on the data.

2.1.2 Correspondence analysis – CA

CA is a dimension reduction method for data analysis of multivariate categorical data. After performing a chi-squared test for independency which resulted in rejection of the indepence hipothesis, CA graphical results show what are the relation among the rows and columns of a contingency table, namely which rows and columns relate to each other and also which rows (columns) can be grouped toghether. CA started out 60 years ago when statisticians, including Sir Ronald Fisher, began to work on the interpretation of the correlaton between rows and columns of a contingency table (Fisher, 1936). However, the graphical application emerged in France with the work of Jean Paul Benzécri and his students (Murtagh, 2005; Greenacre, 2007). CA is a result of an optimization process where an space of 2 or 3 dimensions is found using SVD(Singular Value Decomposition).

The row and column profiles are then projected into that space and then plotted in a 2 or 3 dimensional graphic. Basically the PC(Principal Components, or Principal Axis) are found and become the basis of the low dimensional space. CA is summed up in Jelihovschi and Ferraz (2010), see also (Murtagh, 2005; Greenacre, 2007). CA can also be viewed as a special case of Principal Components Analysis applied either to the rows or columns of a contingency table where the chi-squared metrics is used instead of the usual euclidean metrics. Finally, while the usual chi-squared test shows only that there exist or not a relation between rows and columns, CA shows how those variable are related. CA can also be used with more than two variables. In this case, it is called Multiple Correspondence Analysis, MCA.

3 Dataset

The data set is the result of a survey of 47 beauty salons located at the city of Lavras, Brazil, consisting of two types of questions; the first identifies the profile of the owner manager (explanatory variable), the second are questions referring to the degree of professionalism with respect to planing, market and finances (response variable). The categories of each question of the second type is a number from 0 to 4 where 0 denotes a low level of professionalism and 4 high level of professionalism in managing her(his) business. For example, question MA4 shown below has levels 0 - none, 1 - around 25%, 2 - around 50%, 3 - around 75%, 4 - close to 100%. Number 5 was described as non apply response.

Profile questions

- Partnership(pt) 0, 1, 2 or 3 partners.
- Employees(em) 0, 1, 2, 3 or 6 employees.
- Place(pl) 1-owned, 2-rented.
- Sex(sx) 1-male, 2-female.
- Date of Birth(br) 1-after 1980, 2-between 1970 and 80, 3-before 1970.
- Marital Status(ms) 1-single, 2-married, 3-divorced, 4-other.
- Education(ed) 1-primary, 2-junior high, 3-senior high incomplete, 4-senior high complete, 5-college incomplete, 6-college complete.
- Family Size(fs) 1, 2, 3, 4, 5, 6.
- Type of Customers(tc) 1-male, 2-female, 3-both.
- Years of Experience(ye) 1- < 5; 2- $\geqq 5$ and < 10; 3- $\geqq 10$.
- Number of Customers(nc) 1- < 100; 2- $\geqq 100$ and < 200; 3- $\geqq 200$ and < 400; 4- $\geqq 400$.
- Reason to open the business(rb) 1-habilities and previous experience; 2-entrepreneurship; 3-employ oneself.

- Influence to open the business(ib) 1-family tradition; 2-help of relative-friend; 3-peer influence; 4-none of the above.

Planing

- PA1 What is the dependence of the owner to function properly? Varying from 0 - null then 1,2,3 till 4 - only functions when the owner is present.

- PA2 What are your plans towards next year? 0 - only a dream; 1 - vague goals; 2 - clear short term goals; 3 - clear short goals not precise long term goals; 4 - clear short and long term goals.

Marketing variables

- MA1 Your business tries to systematically assess the customer satisfaction and use that as a basis for management decisions.

- MA2 Your business offer more than the usual services.

- MA3 Your business is focused to further customer loyalty.

- MA4 What is the proportion, among current customers, of those who are customers for more than 6 months.

- MB1 Your business offers more services than when it began.

- MB2 How is your business quality perceived as compared to the competition?

- MB3 How is your business range of services perceived as compared to the competition?

- MC1 What is your business level of prices perceived as compared to the competition?

- MD1 Your business location is perceived as appropriate to the target market.

- ME1 Your business uses formal media to advertise itself.

- ME2 Your business selects the media which is suitable to its target customers.

Financial variables

- F1 Your business clearly separates the owner bills from the business bills.

- F2 Your owners withdrawal are planned and controlled in advance.

- F3 Your business pays for its purchases in installments.

- F4 Your business knows today whether it will be able to pay its short-term bills of 60 days.

- F5 Your business uses short-term cash-flow analysis to plan for its short-term bills.
- F6 Your business has formal control of the monthly amount it makes from its services.
- F7 Your business uses either credit card, checkbook payment or loans, to finance its needs for working capital.
- F8 Your business uses specific credit to finance its needs for capital.
- F9 The company demonstrates knowledge to properly assess the costs of products used in services and costs of renting and taxes.
- F10 Your business clearly identifies the need for working capital.
- F11 Your business lays down the price of services in a systematic way.
- F12 The company calculates the interest on contracted loans.

The categories of the variables which have a very low mass found after running a MCA on the 2 matrices, the matrix of profiles and the matrix of market- finance variables, were taken out of further computations. Mass is defined as the marginal total of a row or a column of a table divided by the grand total of the table. This is done since variables of low mass, either have low influence on the results or, if they have a high inertia, are considered to be outliers, and have a strong influence in the graph. Also, decreasing the numbers of total categories helps the interpretation of the results. When defining the stacked matrix to be used in the LBM the low mass is taken at less than 9 permill. Furthermore, the number of categories (levels) in the matrix of market-finance variables was changed and so the levels were redefined:

- $0, 1, 5$ became 1.
- 2 remained 2.
- $3, 4$ became 3.

This also helps the interpretation by decreasing the number of categories to low level, middle and high level. The level number 5 (non apply) was interpreted as a low level response since by not willing to answer it is assumed the owner-manager did not use it. The reason for doing this is that for every question there is a level interpreted as "not doing it at all", so that, if the manager answers that it does not apply, he(she) is implying a level 0.

Anyhow, the impact is not very important to the analysis since the number of level 5 answers are 0 in 15 questions, 1 in 5 questions, 2 in 1 question, 8 and 9 answers 1 question each, and ME2 with 26 answers and F12 with 33 answers. Since most of beauty salons are small business and those 2 questions are the most difficult for them to practice and understand their usefullness, it understandable that they answer non-apply to those questions.

The softwares used are LBA (?), Rpackage ca (Greenacre and Nenadic, 2010) which belong to R environment for statistical computing (?), and as a graphical user interface (gui) to R we used Tinn-R (Faria et al., 2008).

4 Results

Thirty one profile questions categories were used as levels of the explanatory variable. They are:

Partnership 0, 1, 2	Employees 0, 1	Place 1, 2
Sex 1, 2	Date of Birth 2, 3	Marital Status 1, 2, 4
Education 4, 6	Family Size 2, 3, 4	Type of Customers 1, 3
Years of Experience 1, 2, 3	Number of Customers 1, 2, 3	
Reason to open the business 1, 2	Influence to open the business 1, 4	

Forty five response questions categories were used as levels of the response variable. They are:
PA13, PA21, MA11, MA12, MA13, MA21, MA22, MA31, MA32, MA42, MA43, MB11, MB12, MB13, MB22, MB23, MB31, MB32, MC11, MC12, MD13, ME11, ME21, F11, F13, F21, F23, F31, F33, F41, F42, F43, F51, F53, F61, F63, F71, F73, F81, F91, F93, F101, F111, F113, F121. The first 3 letter(s) and digits denote the question and the last digit denotes the level of the question.

Table 1 is the result of LBM(2) and has two matrices in it; one is A, the matrix of the mixing parameters with the vector π_k at the bottom and the other is B the matrix of latent components with the result of LBM(1) by its side.

4.1 Choice of the number of latent budgets

First, for the determination of number of budgets we use the weighted residual sum of squares (wRSS) between the observed and expected components since the model is distribution free. The decrease in wRSS shows the improvement by adding one more latent budget. This is compared to the improvement required per budget; if it is greater than the required, it means that the budget added is an actual improvement. We can see in Table 2 that 2, 3 or 4 budgets represent a real improvement from the previous one. However, another guideline for choosing a model is that the results should be interpretable. As we can see below, the best choice is 2 budgets, since a third budget will prove to be useless. Therefore, we will try the fitting of only 2 and 3 latent budgets. ? in chapter 7, when analysing real data, also uses interpretability as a parameter to choose the number of latent budgets in the model.

For any number of budgets K there maybe many solutions, latent commponents and mixing parameters, which give the same expected budgets, therefore, there is a need to identify one solution which gives the greatest separation among the latent bugets, thus helping the interpretation of the model. The way to do that is by maximizing the sum of chi-squared distances among the latent budgets.

Table 1 - Mixing parameters and latent components of the LBM(2) solution

mixing parameters			latent budgets			LBM(1)	latent budgets			LBM(1)
	LB1	LB2		LB1	LB2	LB1		LB1	LB2	LB1
pt0	0.378	0.622	PA13	0.029	0.040	0.034	F51	0.015	0.033	0.022
pt1	0.949	0.051	PA21	0.047	0.034	0.041	F53	0.019	0.008	0.014
pt2	0.872	0.128	MA11	0.017	0.014	0.016	F61	0.000	0.032	0.013
em0	0.526	0.474	MA12	0.011	0.021	0.015	F63	0.043	0.005	0.027
em1	0.852	0.148	MA13	0.018	0.008	0.014	F71	0.012	0.038	0.023
pl1	0.588	0.412	MA21	0.036	0.027	0.032	F73	0.033	0.005	0.021
pl2	0.606	0.394	MA22	0.009	0.011	0.010	F81	0.046	0.042	0.044
sx1	0.503	0.497	MA31	0.024	0.034	0.028	F91	0.007	0.045	0.023
sx2	0.631	0.369	MA32	0.017	0.003	0.011	F93	0.026	0.003	0.016
br2	0.487	0.513	MA42	0.010	0.014	0.012	F101	0.023	0.038	0.029
br3	0.432	0.568	MA43	0.028	0.032	0.030	F111	0.015	0.035	0.023
ms1	0.997	0.003	MB11	0.023	0.026	0.024	F113	0.018	0.006	0.013
ms2	0.565	0.435	MB12	0.015	0.004	0.011	F121	0.046	0.043	0.044
ms4	0.345	0.655	MB13	0.009	0.014	0.011				
ed4	0.617	0.383	MB22	0.025	0.020	0.023				
ed6	0.627	0.373	MB23	0.019	0.008	0.014				
fs2	0.650	0.350	MB31	0.006	0.029	0.016				
fs3	0.550	0.450	MB32	0.035	0.010	0.025				
fs4	0.471	0.529	MC11	0.018	0.037	0.026				
tc1	0.450	0.550	MC12	0.022	0.007	0.016				
tc3	0.607	0.393	MD13	0.030	0.035	0.032				
ye1	0.875	0.125	ME11	0.038	0.037	0.038				
ye2	0.752	0.248	ME21	0.040	0.036	0.039				
ye3	0.393	0.607	F11	0.006	0.041	0.020				
nc1	0.175	0.825	F13	0.035	0.001	0.021				
nc2	0.475	0.525	F21	0.012	0.045	0.025				
nc3	0.718	0.282	F23	0.031	0.000	0.018				
rb1	0.748	0.252	F31	0.025	0.015	0.021				
rb2	0.315	0.685	F33	0.017	0.022	0.019				
ib2	0.724	0.276	F41	0.003	0.029	0.014				
ib4	0.605	0.395	F42	0.025	0.007	0.017				
			F43	0.018	0.008	0.014				
π_k	0.58	0.419								

Table 2 - Goodness of fit of LBM for K = 1, 2, 3

number of latent budgets	wRSS	actual improvement	improvement required	fit improved?
1	0.051	NA	NA	NA
2	0.033	0.018	0.0011	yes
3	0.025	0.008	0.0011	yes
4	0.02	0.005	0.0011	yes

We are first going to use the matrices A and B shown in Table 1. Table 1 has all the results used to interpret the model, however, as it was pointed out in the introduction, instead of doing the analysis on the table as it is usually done, we will make use of the CA graphs, which, due the large number of both variable levels, make it easier and more elegant to analise the results. The graphs used in Figure 1 and Figure 2 are made from a slight modification of those matrices as described below.

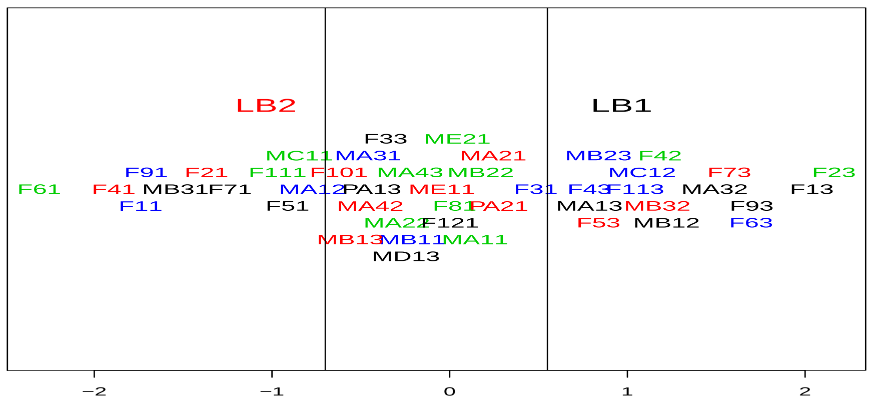

Figure 1 - CA graphics, latent components K = 2.

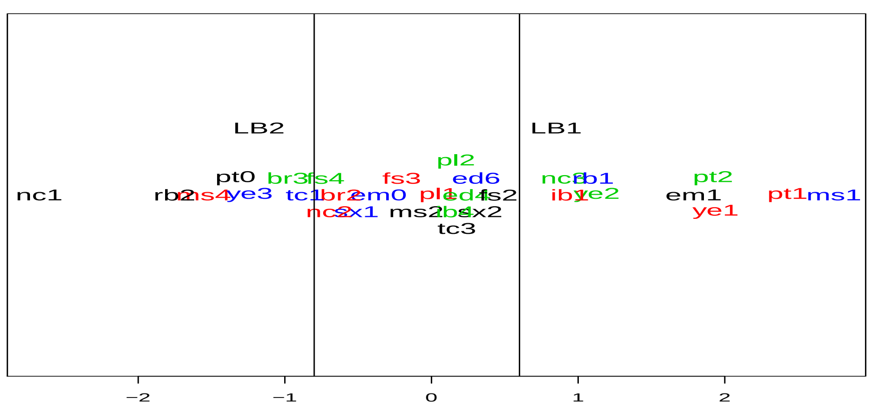

Figure 2 - CA graphics, mixing parameters K = 2.

4.2 Correspondence analysis plots

The characterizations of the latent budgets and mixing parameters are made by comparing the first with the LBM(1) results p_{+j} and the later with π_k, and this is done directly on the matrices A and B. Nevertheless when both matrices have a large number of lines it becomes a slow and difficult process so that, instead of using those matrices A and B we will use CA graphs on slightly modified versions of them. This is done in order to make the origin of the CA graphs become the projections of the vectors p_{+j} and π_k. As CA projects both the rows and columns of matrix A on the same graph and rows and columns of matrix B on the same graph, where the latent budgets are the columns on both matrices, important levels of both matrices relating to the budgets stands out almost automatically.

The matrix A has its columns elements multiplied respectively by the elements of the vector p_{i+} for $i = 1, \ldots, 31$. By using equation (1), we can see that the column totals of the transformed matrix become π_k and, since $\sum_{k=1}^{K} \alpha_{k|i} = 1$ the row totals amount to p_{i+}. The total of the transformed matrix is 1, therefore, both the row and column totals of this matrix remain unchanged when divided by the total $n = 1$. Those marginal totals become the origin of the CA graph. Moreover, the row profiles of this transformed matrix are the original row profile values of the matrix A, because $\sum_{k=1}^{K} \alpha_{k|i} = 1$.

The matrix B has its columns elements multiplied respectively by the elements of the vector $\frac{p_{+j}}{\sum_k \beta_{j|1}}$ for $j = 1, \ldots, 45$. In this case the row totals become p_{i+}. The row profiles become $\frac{\beta_{j|k}}{\sum_k \beta_{j|1}}$ which are the same as the row profiles of the matrix B, and since the total of this matrix is 1 the row totals of the transformed matrix remain p_{i+}.

Figure 1 shows the CA plot of the latent components and $K = 2$. The graph has just one dimension since the matrix has 2 columns and the maximum number of dimensions in CA is $\min(I-1, J-1)$, in this case the the proportion of inertia explained by the CA is 100%.

Proportion of inertia measures the amount of information from the data matrix captured by the graphs. Whenever there is a reduction of dimensionality some information is lost, on the other hand, if the graphs have the same dimension of the profile matrix than no information is lost. The latent components are laid in more than one line because otherwise it would be impossible to distiguish among them.

The graph of Figure 1 is divided in 3 parts. The left is connected to the second latent budget that is the LB2 has greater proportions in those components and all belong to level 1 which indicates the LB2 can be called low level of profissionalism. Also, most of those components questions are of financial type what indicates that the low level of profissionalism is most frequent at the financial situations.

The right one is connected to the first latent budget, that is LB1 has greater proportions in those components and all belong to level 2 or 3 which indicates the LB1 can be called high level of profissionalism. In this case LB1 is formed by market and financial situations.

The middle part contains all budget components related do neither one of the two latent budgets, that is neither LB1 nor LB2 have strong greater proportions in any of those components that is, those latent components are close to the LBM(1) results.

Figure 2 shows the CA plot of the mixing parameters and $K = 2$. As before, the graph has just one dimension and explains 100% of the inertia.

The graph of Figure 2 is divided in in 3 parts. The left, which is strongly connected do the LB2 shows that beauty salon with just one owner manager, no partner, older one born before 1970, marital status not specified, only male customers (those are just barber shop), more than 10 years of experience (since they are older), less than 100 customers, opened the business because of entrepreneurship are predominantly of low level of profissionalism. In this case entrepreneurship must be understood as of very low level where people call themselves entrepeneurs but do not follow any of the rules of good entrepreneurship. The right one is strongly connected to LB1. They are: single, have one or two partners, less than 5 years of experience or between 5 to 10 years of experience with one employee, have between 200 and 400 customers, follows a family tradition and relied on habilities and previous experience to open the business. The mixing parameters in the middle part have their values approximately close to π_k.

Figure 3 shows the CA plot of the latent components and $K = 3$. The graph has two dimensions since the matrix has 3 columns and the maximum number of dimensions is 2, so that, it explains 100% of the inertia. The first principal factor of the CA explains 74.7% of the inertia and the second principal factor explains 25.3%. The financial budget components are lying along the first principal axis and the marketing budget components along the second principal axis. This shows that the financial budget components are more important in determining the latent budgets. This was somewhat hidden in the one dimensional graph of 2 latent budgets.

LB3 is the low level of profissionalism, LB1 is the high level of profissionalism being characterized mostly by financial variables. LB2 is a mixture of both, being characterized by the budget components located in the middle of the map what shows that it is not a necessary latent budget, that is, to be well interpreted, only 2 latent budgets are needed in the model.

Figure 4 shows the CA plot of the mixing parameters for $K = 3$. As before, the graph has two dimensions and explains 100% of the inertia. LB1 and LB3 are both at the same side of the second principal axis. This shows that the second principal factor separates between LB1, LB3 and LB2 and the first axis separates between LB1 and LB3.

The graph also shows that the mixing parameters that explains the low and high level of profissionalism are almost the same as those found for K = 2.

When the number of categories of both the response and explanatory variables is small it is not too difficult to interpret the latent budgets by looking straight at the latent components table and at the mixing parameters table. On the other hand, if the numbers categories is large it might become very hard to find the right number of latent budget and even harder to interpret them. The use of CA graph, as above, make it much easier to perform that task.

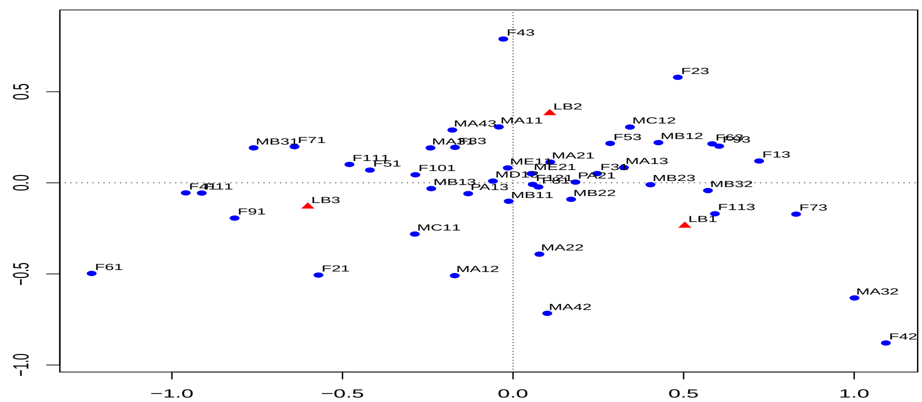

Figure 3 - CA graphics, latent components K = 3.

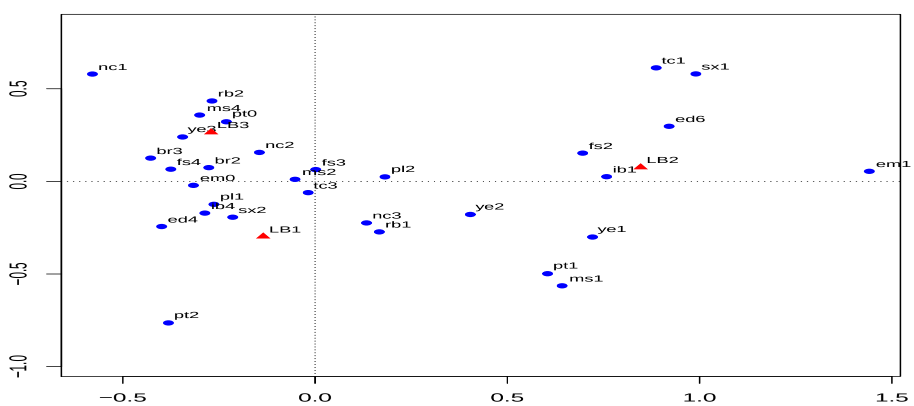

Figure 4 - CA graphics, mixing parameters K = 3.

We could also look at some of the results of MCA, Figure 5 and 6, applied to the original table of marketing and financial questions, ajusted to the 3 levels of answers as done before. The marketing and financial questions were separated and only the relevant questions were used. Again, high and low level of profissionalism were as well separated in the graphs.

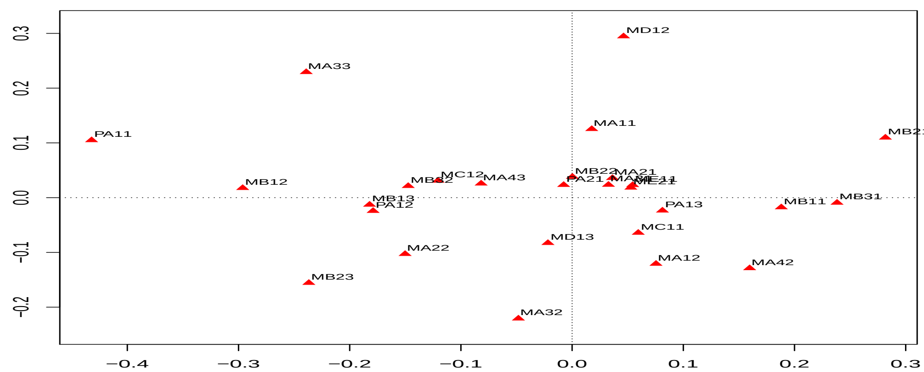

Figure 5 - MCA, only marketing questions.

Figure 5 shows the MCA graph of marketing questions. On the left side of the first axis only categories 2 or 3 are found with the exception of the planning variable PA11. The right side almost all are the category 1 is found. Figure 6 shows the same for financial questions with sides inverted, right is high level of profissionalism and the left low level. This division between high and low level of profissionalism is much more outstanding with financial questions as already shown in Figure 3.

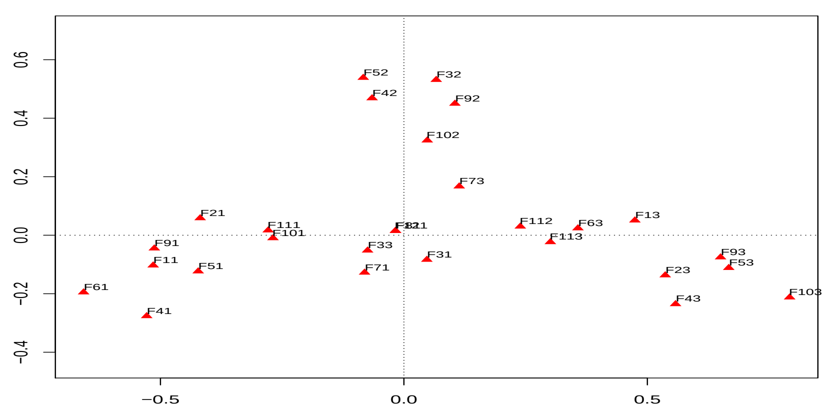

Figure 6 - MCA, only financial questions.

Conclusions

LBA is a technique, derived from latent class analysis, used when the data consists of one explanatory and one response variable, and the question of interest is how the expected budgets can be composed of a smaller amount of latent budgets. On the other hand, CA visualizes how row profiles can be explained by continuous axes, which can be interpreted as latent traits which are features of the data set that characterize the axes or principal components, this is shown in figures 5 and 6. Since the row profiles are equivalent to the expected budgets, the difference between LBA and CA could be summarized as the choice between a trait or an explanation of the latent budgets which amount to the same type of information. The analysis in this article gives us the benefits of both models.

It should be pointed out that if the number of latent budgets is large it may happen that the proportion of inertia explained by the 2 axes is low and therefore the CA might not give all the information about LBA. Usually the number latent budgets used in practice is small so that the models can be easily and reliably interpreted

Acknowledgments

We are deeply indebted to Orson Galvão *galvao@uesc.br* for his solution on how run Lba.exe. Lba was written in Borland PASCAL7.0 on an AT-386 PC and runs under MS-DOS but do not run on new and fast processors. Mr. Galvão found the SLOWDWON software written by Bret Johnson *bretjohn@juno.com* what made it possible to run Lba.exe on any new computer.

JELIHOVSCHI, E. G.; ALVES, R. R.; CORRÊA, F. M. Análise de gerenciamento de salão de beleza usando uma interação de análise de provisão(budget) latente e análise de correspondência. *Rev. Bras. Biom.,* São Paulo, v.29, n.4, p.657-673, 2011.

- RESUMO: Análise de provisão(budget) latente (LBA) e análise de correspondência (CA) são usadas de uma forma interativa com a finalidade de analisar dados referentes ao gerenciamento de salões de beleza. A pesquisa consiste basicamente em dois tios de questões; o primeiro identifica o perfil do proprietário gerente, e o segundo são questões que identificam o grau de profissionalização do proprietário no que se refere a questões de ordem mercadológicas e financeiras. Foi feita uma matriz empilhada na qual as linhas são os perfis e as colunhas as questões mercadológicas e financeiras. Os métodos citados foram usados para analisar esta matriz. Primeiramente usamos CA com a finalidade de encontrar as categorias mais importantes com relação à sua massa seguida depois pela LBA. Depois os resultados gráficos de CA são usados para descrever os budgets latentes e os parametros de composição. Tanto dois como três budgets latentes dividem os salões entre baixo e alto nível de profissionalização.

- PALAVRAS-CHAVE: Análise de correspondência(CA); análise de provisão(budget) latente(LBA); pesquisa mercadológica e financeira de salões de beleza.

References

AGRESTI, A. *Categorical data analysis.* New York: Wiley, 1990. 558p.

CLOGG, C. C. Latent structure models of mobility. *Am. J. Sociol.*, Chicago, v.86, n.4, p.836-868, 1981.

DE LEEUW, J.; VAN DER HEIJDEN, P. G. M. The analysis of time-budgets with a latent time-budget model. In: DIDAY, E. (Ed.). *Data analysis and informatics V.* Amsterdam: North-Holland, 1988. p.159–166.

DE LEEUW, J.; VAN DER HEIJDEN, P. G. M.; VERBOON, P. A latent time budget model. *Stat. Neerl.*, Oxford, v.44, n.1, p.1–21, 1990.

FARIA, J. C.; GROSJEAN, P.; JELIHOVSCHI, E. *Tinn-R - GUI/Editor for R language and environment statistical computing.* Disponível em: $< http : //sourceforge.net/projects/tinn - r, 2008 >$.

FISHER, R. A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, London, v.7, p.179-188, 1936.

GOODMAN, L. A. The analysis of systems of qualitative variables when some of the variables are unobservable. A modified latent structure approach. *Am. J. Sociol.*, Chicago, v.79, n.5, p.1179-1259, 1974.

GREENACRE, M. *Correspondence analisys in practice.* 2.ed. Boca Raton, FL: Chapman Hall/CRC, 2007. 296p.

GREENACRE, M.; NENADIC, O. *CA: Simple, multiple and joint correspondence analysis.* R package version 0.33. $< http : //CRAN.R - project.org/package = ca.2010 >$.

JELIHOVSCHI, E. G.; FERRAZ, M. I. F. Análise do conjunto dos candidatos ao vestibular da UESC no ano de 2008 usando análise de correspondência. *Rev. Bras. Biom.*, São Paulo, v.28, n.4, p.117-136, 2010.

McCUTCHEON, A. L. Correspondence analysis used complementary to latent class analysis in comparative social research. In: GREENACRE, M.; BLASIUS, J. (Ed.). *Visualization of categorical data.* London: Academic Press, 1998. p.477–488.

MURTAGH, F. *Correspondence analysis and data coding with Java and R.* Londres: Chapman Hall/CRC, 2005. 230p.

R DEVELOPMENT CORE TEAM. *R*: a language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna: Disponível em: $< http : //www.R - project.org, 2010 >$.

UNESCO. $< http : //www.unesco.org/webworld/idams/advguide/Chapt6_5.htm >$.

VAN DER ARK, L. A. *Contributions to latent budget analysis*. 1999. 217f. Thesis (Doctoral) – University of Utrecht, Utrecht, 1999.

VAN DER ARK, L. A. *Latent budget analysis for two way tables.* Disponível em: $< http://come.to/lba/software.html, 1999 >$.

VAN DER ARK, L. A.; VAN DER HEIJDEN, P. G. M. Graphical display of latent budget analysis and latent class analysis, with special reference to correspondence analysis. In: GREENACRE, M.; BLASIUS, J. (Ed.) *Visualization of categorical data*. London: Academic Press, 1998. p.489–508.

VAN DER HEIJDEN, P. G. M; MOOIJAART, A.; DE LEEUW, J. Constrained latent budget analysis. In: CLOGG, C. C. (Ed.). *Sociological methodology*. v.22. Cambridge: Basil Blackwell, 1992. p.279–320.

Received in 20.08.2011.

Approved after revised in 12.04.2012.