Multiple Correspondences in E-Commerce

M^A BEATRIZ BERNÁBE LORANCA
Cuerpo Académico de Ingeniería de Software -Aplicaciones Estadísticas.
Facultad de Ciencias de la Computación, BUAP.
Calle 14 sur y Avenida San Claudio, Colonia San Manuel, Puebla.
MÉXICO.

LUIS ANTONIO OLSINA SANTOS

GIDIS, Departamento de Informática, UNLPam
Calle 9 y 110, (6360) General Pico, La Pampa.
ARGENTINA.

Abstract:- When an exploratory data analysis is performed where there are more than two qualitative variables, the application of univariate, bivariate and multivariate statistical techniques allows to successfully describe the data table. Particularly, the single correspondence technique gives important correlation and dimensionality reduction results, which help to give an objective interpretation of the data. In this paper we use the technique known as factor analysis of multiple correspondences which is a generalization of the single correspondence technique used to corroborate results.

We also use log-linear adjustment, with the purpose of continuing with the Principal Components and Cluster Analyses [3]. The binary variables under study are the result of the e-commerce sites' evaluation process for the quality attributes of the "Functionality" feature [8, 9]. This data is concentrated in a binary table of 49 sites and 17 attributes [3, 8].

Key-words:- E-Commerce, Log-Linear Adjustment, Multiple Correspondences, Multivariate Statistical.

1 Introduction

While the marked differences between physical and electronic commerce are evident, it is also known that many common features can be abstracted and incorporated when building a Web application for the *virtual* mode of commerce [7]. Therefore, we must resort to several tools and strategies for the development and evaluation of a software product of this magnitude. Specifically, and considering a general audience, it seems appropriate to ask, Which features and attributes must be taken into account at the design (and evaluation) stage of an e-commerce site or application in order to obtain a quality product with the potential to influence traffic and sales?, How to decide which evaluation process is adequate? Which strategy must be followed in the data collection and organization for the evaluation?, and ultimately, Which statistical techniques must be applied in order to effectively interpret the datasets stemming from the evaluation?

A starting point for this task comes from an Argentina's e-commerce sites quality study [8], where an evaluation of directly measurable *functionality* attributes was performed. Such attributes belong to *Product Information*, *Purchase Features*, *Client Customization*, *Promotion Policies* and *Searching Mechanisms* sub-features; all of them grouped under the *Functionality* feature. The attributes and sub-features form the quality requirements for an audience called "General visitor".

Accordingly, this study's main objective consists on continuing the statistical analysis of the binary data obtained by the former survey. During such survey, seventeen attributes grouped in the sub-features

already mentioned were observed from a sample of forty nine sites (in [10], about ninety attributes were considered for evaluation).

The experimental studies that are performed with a large number of attributes –i.e. explaining independent variables, follow data analysis methods and techniques that range from data description and summary up to exploratory analysis to identify potential relations among variables, or analysis to classify and categorize the objects under study.

In this paper we resort to the application of multivariate statistical techniques, in a way that allows us to keep describing the binary data table; its dimensionality is also reduced and a classification of quality attributes is given, this time from a multiple factor point of view. Previously, an analysis on datasets was conducted using Principal Components, Cluster and Single Correspondence Factor Analyses [3] techniques. Now, a Multiple Correspondence Factor Analysis is applied, and the established hypotheses of relationships and/or categorization of sites and variables are tested with a well-suited approach for this type of problems, called Log-linear adjustment. During this multivariate process the importance, relationships, and categorization of the qualitative attributes mentioned earlier are confirmed, among other aspects. Eventually, this leads us to draw conclusions about the relevance of the sub-features involved, but now from a perspective that enriches the results produced by the used statistical technique: i.e., the log-linear adjustment.

Regarding the number of evaluated sites the field study required a sample of 49 e-commerce sites from the most diverse categories. The sample was drawn from 100 e-commerce Argentinean sites [8]. The sites were selected randomly with the use of a random value table. In addition, an evaluation of 17 attributes referring to five categories of the *Functionality* feature in the e-commerce domain was used. Table A.1 of appendix A shows the listing.

Finally, binary data criteria was adopted for the assessment, i.e., it is available (1) or *not available* (0). This should be interpreted as "satisfies completely the elementary quality requirement" (100%) or "does not satisfy the elementary quality requirement at all" (0%).

The rest of this article proceeds as follows. In the next two sections the multiple correspondence factor analysis and the log-linear adjustment over data from e-commerce sites are discussed. Finally, concluding remarks are drawn.

2 Multiple Correspondences Factoring

To begin our study we'll return to the first idea about the relation between qualitative variables from the e-commerce table (see Table 1), considering the p correlation coefficient.

Remember that this *correlational* number does not always implies *causality*. Moreover, p emerges from assuming the Null Hypothesis to be true.

Table 1 Correlation coefficients among variables

Variables			CC	p
IP-IMA	vs	IP- DES	.63	.000001
CC-ICAN	vs	IP-IMA	-.30	.032999
CC-CAR	vs	MB-RES	.49	.000321
CC-CAR	vs	PC-CUEN	.43	.001904
CC-CAR	vs	CC-TRANS	.46	.000827
CC-CAR	vs	CC-V-OFF	-.46	.000869
CC-CAR	vs	MB-GLOB	.29	.044327
CC-CAR	vs	CC-V-ON	.54	.000073
CC-V-OFF	vs	CC-V-ON	-.34	.017172
CC-V-OFF	vs	CC-IEC	.29	.044566
CC-IEC	vs	CC-TRANS	.29	.041356
CC-IEC	vs	CC-IP	.70	.000000
MB-RES	vs	PC-CUEN	.35	.013890
MB-RES	vs	CC-TRANS	.30	.035447

Table 1 contains the values for CC (Correlation Coefficients). They highlight the important relation between the Shopping Features, which are correlated between them and also with some others. The attribute "shopping cart" has a high positive correlation coefficient with another five attributes, while its correlation coefficient is negative with the attribute "off-line sales". On the other hand, the largest correlation coefficient appears between two attributes from the shopping features: payment information and shipping and cost information, which means that the more one appears, the more the other appears too. The same occurs with the "product information" attributes concerning basic description and product image. The full names of the variables that appear in the table can be found in Table A.1 of Appendix A.

Even after performing a single correspondence study with the binary table, it is necessary to corroborate with another complementary techniques the results produced about the important classification and relationship between sites and attributes.

The multiple correspondence analysis [1, 4, 5] allows us to study the relationship between the modes of all the qualitative features involved.

As a factor method, just like the principal components and single correspondence analyses, it is necessary to diagonalize certain matrix in order to

obtain the factors. In this case such matrix is the product of the transposed complete disjunctive matrix (in our case, the table of site's attributes) by itself, lets denote it with B. B has to be pre-multiplied by a diagonal matrix (D) which consists of the diagonal elements of B and zeros elsewhere and finally divided by the number of features or qualitative variables (Q).

The B table (Table A.2, appendix A) is known as *Burt Table* and is nothing more than a symmetric matrix formed by Q^2 blocks. The diagonal's blocks are diagonal tables that cross a variable with itself, the diagonal elements being the effective ones of each mode. The blocks outside the diagonal are the contingency tables obtained by crossing the features pair-wise and whose elements are the frequencies of association for the two corresponding modes. In the Burt's Table we can see the values for mass, quality, inertia and cosine which allow us to interpret in three dimensions the relationship between the variables under study [6].

The same concepts from simple correspondence are used in order to interpret the multiple correspondence analysis. The eigenvalues table (Table 2) shows that when we select three dimensions almost 42% of the phenomenon under study is explained. The first dimension explains 17.14% of the phenomenon's variance, the second explains 12.54% and the third 12.15%.

Table 2 Eigenvalues by dimension

	Singular Values	Eigen-Values	Perc. of Inertia	Cumulatv Percent	Chi Squares
1	.414	.171	17.14	17.1	227.2
2	.354	.125	12.54	29.7	166.2
3	.349	.122	12.15	41.8	161.1
4	.321	.103	10.33	52.2	136.9
5	.272	.074	7.39	59.6	98.0
6	.259	.067	6.70	66.2	88.8
7	.242	.059	5.86	72.1	77.6
8	.230	.053	5.31	77.4	70.3
9	.203	.041	4.12	81.5	54.6
10	.198	.039	3.91	85.4	51.9
11	.181	.033	3.26	88.7	43.2
12	.175	.031	3.07	91.8	40.7
13	.163	.027	2.66	94.4	35.2
14	.156	.024	2.42	96.9	32.1
15	.118	.014	1.38	98.2	18.3
16	.100	.010	1.00	99.2	13.3
17	.087	.008	.76	100.0	10.1

From the analysis of table A.2 (Appendix A) and the corresponding charts, it stands out that:

For the first dimension

1. If the attribute *Shopping Cart* is not present in a web-site, neither will be the *Basic product description, On-line sales, Restricted search mechanism*, nor *Secure Transaction* attributes. (Fig. 1)
2. When the attribute *Shopping Cart* is present, sites will also feature the *On-line sales, Secure transaction* and *Restricted search* features. (Fig. 1)

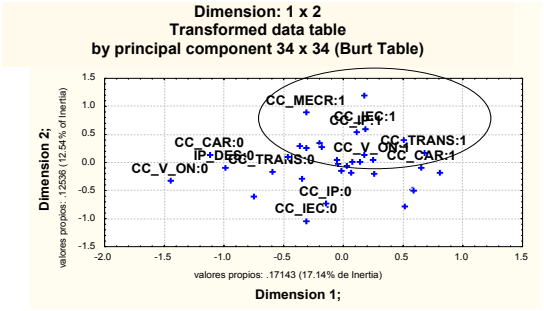

Figure 1 Dimension 1 x Dimension 2

For the second dimension

1. If sites have the *Shipping and costs information*, they generally have the *Payment information* and vice-versa (Fig. 2)

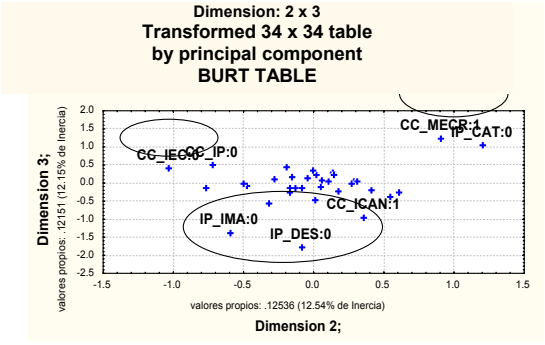

Figure 2 Dimension 2 x Dimension 3

For the third dimension

1. Sites that do not feature product information attributes (*Basic description* and *Product image*) have *Purchase cancellation information*.
2. Absence of the *Product information by catalog* attribute is related to the presence of *Quick buy Mechanism* (see Figure 3)

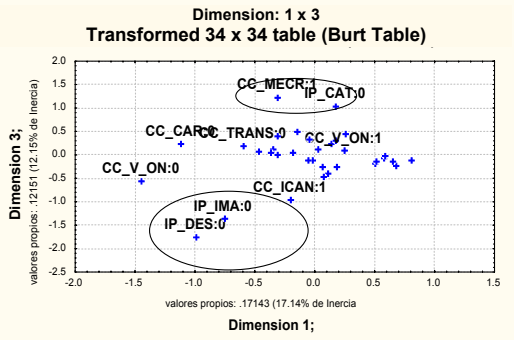

Figure 3 Dimension 1 x Dimension 3

The above interpretation is corroborated by the frequency charts in Figures 4 to 7.

Note that the tallest bar reflects the high frequency of sites where the *Shopping cart* attribute is present together with the *On-line sales* attribute. (Figure 4).

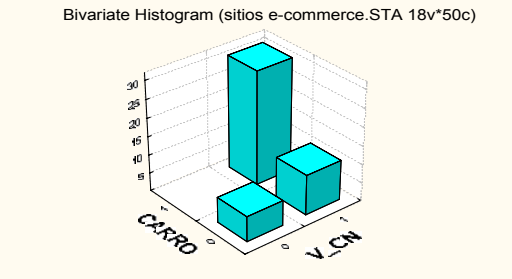

Figure 4 Shopping Cart and On-line Sales

The next frequency chart represents the crossing of the *Shopping cart* and *Secure Transaction attribute*, it also shows the high frequency of sites which feature both attributes simultaneously (Figure 5)

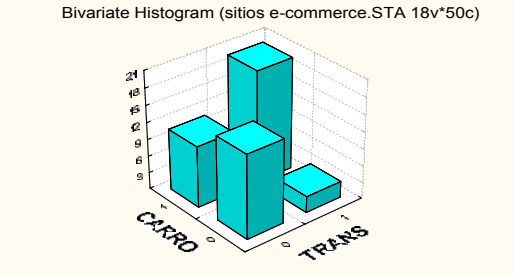

Figure 5 Shopping Cart and Secure Trans

The third frequency chart (Figure 6) shows the crossing of the *Shopping cart* again, this time with the *Quick buy mechanism*. It can be seen that there is a high frequency of sites which do not have the second when they have the first.

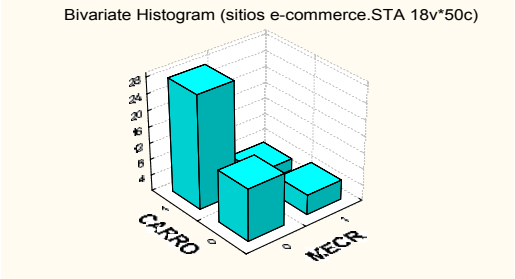

Figure 6 Shopping Cart and Quick Buy Mech.

Finally, the fourth and last frequency chart shows the high frequency of web-sites that simultaneously have the *Shipping and costs information* and *Payment information.* (See Figure 7)

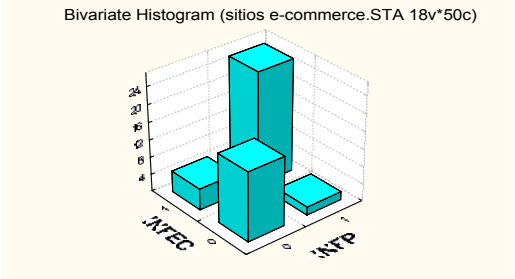

Figure 7 Shipping & Costs and Payment Info.

The different statistical approaches used so far have brought to light the extraordinary importance the "Shopping Features" attributes have, independently of the other (also important) attributes such as "Product Information", essential for any e-commerce site and the "Promotion Policies".

Under the assumption that some of the attributes considered among the "Shopping Features" group are highly related and that these relationships have to be cleared up, we proceed to perform the confirmatory study of this hypothesis.

3 Log-linear Adjustment

Given that many of the "Shopping Features" attributes appear as important attributes in the web-site quality study, we decided to adjust some log-linear model that allowed us to point out relationships among these attributes. In this sense there is the inconvenience of having only 49 web-sites, while the number of shopping features is 8 and the cross-tabulations can't be analyzed since numerical problems of different kind appear.

After exploring different tables it was possible to pick the following one, which also presents important

attributes, according to the exploratory studies formerly explained:

CC-TRANS (Secure Transaction)
CC-CAR (Shopping Cart)
CC-IP (Payment Information)
CC-ICAN (Cancel Purchase Information)
CC- MECR (Quick buy Mechanism)

The model with all the second-order interactions achieves a good adjustment, since the Pearson's χ^2 statistic has 11.07 with 16 degrees of freedom, with a value of 0.8048 for p. The cross-tabulations of two entries and the corresponding "odds" ratios are shown in Table 3.

As it is known, the Odds ratio (likelihood ratio) in the two-entry tables, that we'll denote by θ, may take the following values:

- $\theta = 1$ When the variables are independent
- $1 < \theta$ When sites from row 1 are *more likely* to fall into column 1 than sites from row 2.
- $0 < \theta < 1$ When sites from row 1 are *less likely* to fall into column 1 than sites from row 2.

From the values calculated in the table the next relationships arise:

- Web-sites that do not have shopping carts are 27.27 times more likely to not having secure transactions compared to the ones that do.
- The web-sites that do not have quick buy mechanism are 0.28 times less likely to not having shopping carts compared to the ones that do have quick buy mechanism.
- Web-sites without quick buy mechanism are 0.32 less likely to have secure transactions than the ones that have it.
- Web-sites without payment information are 1.87 (i. e. almost twice) more likely to not having secure transactions compared to the ones that have payment information.

Table 3 Attributes' Cross-tabulations

CC- TRANS			
CC-CAR	0	1	Total
0	15	3	18
1	11	20	31
Total	26	23	49
Odds ratio = 27.27			

CC- TRANS			
CC-IP	0	1	Total
0	13	8	21
1	13	15	28
Total	26	23	49
Odds ratio = 1.87			

CC- TRANS			
CC-ICAN	0	1	Total
0	23	20	43
1	3	3	6
Total	26	23	49
Odds ratio = 1.15			

CC- TRANS			
CC-MERC	0	1	Total
0	20	21	41
1	6	2	8
Total	26	23	49
Odds ratio = 0.32			

CC- CAR			
CC-IP	0	1	Total
0	8	13	21
1	10	18	28
Total	18	31	49
Odds ratio = 1.11			

CC- CAR			
CC-ICAN	0	1	Total
0	16	27	43
1	2	4	6
Total	18	31	49
Odds ratio = 1.18			

CC- CAR			
CC-MERC	0	1	Total
0	13	28	41
1	5	3	8
Total	18	31	49
Odds ratio = 0.28			

CC- IP			
CC-ICAN	0	1	Total
0	20	23	43
1	1	5	6
Total	21	28	49
Odds ratio = 4.35			

CC- IP			
CC-MECR	0	1	Total
0	17	24	41
1	4	4	8
Total	21	28	49
Odds ratio = 0.71			

CC- ICAN			
CC-MECR	0	1	Total
0	36	5	41
1	7	1	8
Total	43	6	49
Odds ratio = 1.028			

4 Conclusions

After an exhaustive analysis of the data table for 17 functionality attributes of 49 web-sites, in the sense of a site having or not having each attribute, we can conclude:

All selected attributes for the quality study of the "functionality" aspect of web-sites are relevant and even though, logically, the "Product Information" attributes should be considered essential and are therefore available in a high proportion of the sites, we consider that the "Shopping Features" attributes stand out because of their importance in these site's quality. Such attributes are:

On-line sales, Secure transaction, Shopping cart, Payment information, Purchase Cancellation information and Quick buy mechanisms. The relationships between these attributes have been described in prior analyses [2, 3, 9].

The "Promotion Policies" attributes follow in order of importance.

1. Particularly, and returning to the log-linear adjustment results, we can state that those web-sites that feature shopping carts should consider having secure transactions.

2. On the other hand, the e-commerce sites without quick buy mechanism are unlikely to lack shopping carts. That is, even though the probability ratios obtained by the log-linear adjustments are low for this pair of variables, it is necessary to pay special attention to them when an e-commerce site is being designed/built. In other words, it is not completely necessary to include a shopping cart if a quick buy mechanism is present, but it is highly advisable to have both. We draw this conclusion from the fact that the probability of finding an e-commerce site without both of these attributes is low, but not zero.
3. It is unusual not to pair quick buy mechanisms with secure transactions. The probability of finding a site which lacks both attributes is low. Given this, we recommend that when building an e-commerce site and including a quick buy mechanism, there must be a secure transaction mechanism.
4. If a web-site does not have cancellation information then it is more than four times more likely to also lack payment information than having it. This means that there is a four times higher probability of finding sites without both attributes than finding a site that does not have the attributes mentioned in conclusions 2 and 3.
5. The payment information attributes can not be exclusive.

To put into practice these recommendations of pairing some attributes with others requires a redesign of one of the sites with the observed restriction, and then a comparison with of the "new and improved" with the original, which implies an expensive marketing study. As attractive as it sounds, the cost is not only a function of the research work associated, but also of the high economic and time investment.

A competitive alternative for predicting results from the above recommendations is to use a data mining approach, which may give good results for economic models, but the data involved has to "have a history". For this reason, the binary table we have is not of much help and we should consider another data collecting process, adequate for the data mining approach.

On the other hand, we intend to apply pattern recognition approaches, with or without training which can be applied to binary data. [11]

Appendix A

Table A.1 Functionality and Content-oriented E-Commerce Attributes and Variables.

1	Product Information	
1.1	Basic Product Description	IP-DES
1.2	Product Image	IP-IMA
1.3	Catalog	IP-CAT
2	Shopping Features	
2.1	On-line Sales	CC-V-ON
2.2	Off-line Sales	CC-V-OFF
2.3	Secure Transaction	CC-TRANS
2.4	Shopping Cart	CC-CAR
2.5	Shipping and Cost Information	CC-IEC
2.6	Payment Information	CC-IP
2.7	Purchase Cancellation Information	CC-ICAN
2.8	Quick Buy Mechanism	CC-MERC
3	Client Customization	
3.1	Subscription	PC-SUS
3-2	Customized Account	PC-CUEN
4	Promotion Policies	
4.1	Promotion at Sale	PP-PROV
4.2	Promotion by prizes	PP-PROVR
5	Searching Mechanisms	
5.1	Global	MB-GLO
5.2	Restricted	MB-RES

Table A.2 Column coordinates and inertia contributions (Burt Table).

	Row Num	Coo D1	Coo D2	Coor D3	Mass	Quality	Rel Inertia	Iner D1	Cos² D1	Iner D2	Cos² D2	Inert D3	Cos² D3
IP_DES:0	1	-.99	-.09	-1.76	.007	.570	.052	.04	.13	.000	.000	.180	.430
IP_DES:1	2	.14	.01	.25	.052	.570	.007	.006	.137	.000	.001	.026	.430
IP_IMA:0	3	-.76	-.60	-1.36	.011	.625	.048	.036	.129	.031	.080	.165	.416
IP_IMA:1	4	.17	.13	.31	.048	.625	.011	.008	.129	.007	.080	.037	.416
IP_CAT:0	5	.181	.20	1.05	.006	.293	.053	.001	.004	.069	.165	.054	.125
IP_CAT:1	6	-.02	-.14	-.12	.053	.293	.006	.000	.004	.008	.165	.006	.125
CCVON:0	7	-1.45	-.32	-.55	.008	.418	.050	.103	.350	.007	.017	.021	.051
CCVON:1	8	.24	.05	.09	.050	.418	.008	.017	.350	.001	.017	.004	.051
CCVOFF:0	9	.50	.41	-.18	.024	.312	.035	.036	.175	.032	.116	.006	.022
CCVOFF:1	10	-.35	-.28	.12	.035	.312	.024	.024	.175	.022	.116	.004	.022
CTRA.:0	11	-.60	-.15	.19	.031	.473	.028	.065	.406	.006	.027	.009	.041
CTRA.:1	12	.68	.17	-.22	.028	.473	.031	.074	.406	.007	.027	.011	.041
CCCAR:0	13	-1.12	.15	.24	.022	.779	.037	.159	.734	.004	.012	.010	.033
CCCAR:1	14	.65	-.08	-.14	.037	.779	.022	.092	.734	.002	.012	.006	.033
CC_IEC:0	15	-.31	-1.04	.42	.022	.782	.037	.012	.056	.185	.623	.031	.102
CC_IEC:1	16	.18	.60	-.24	.037	.782	.022	.007	.056	.107	.623	.018	.102
CC_IP:0	17	-.15	-.72	.51	.025	.603	.034	.003	.016	.105	.392	.054	.196
CC_IP:1	18	.11	.54	-.38	.034	.603	.025	.002	.016	.079	.392	.041	.196
CCICA:0	19	.03	-.05	.13	.052	.151	.007	.000	.006	.001	.018	.008	.127
CCICA:1	20	-.21	.36	-.95	.007	.151	.052	.002	.006	.007	.018	.054	.127
CCME:0	21	.06	-.18	-.24	.049	.474	.010	.001	.019	.012	.159	.023	.296
CCME:1	22	-.31	.90	1.23	.010	.474	.049	.006	.019	.062	.159	.120	.296
PCSUS:0	23	-.18	.28	.05	.043	.315	.016	.008	.093	.027	.216	.001	.006
PCSUS:1	24	.51	-.77	-.13	.016	.315	.043	.024	.093	.074	.216	.002	.006
PCCUE:0	25	-.37	.31	.04	.036	.363	.023	.028	.212	.027	.147	.001	.003
PCCUE:1	26	.58	-.48	-.07	.023	.363	.036	.045	.212	.042	.147	.001	.003
PPPRV:0	27	-.05	-.01	.35	.034	.164	.025	.001	.003	.000	.000	.033	.161
PPPRV:1	28	.07	.01	-.46	.025	.164	.034	.001	.003	.000	.000	.044	.161
PPPPR:0	29	-.06	.04	-.10	.048	.071	.011	.001	.015	.001	.009	.004	.048
PPPPR:1	30	.25	-.20	.46	.011	.071	.048	.004	.015	.003	.009	.019	.048
MBGO:0	31	-.31	.27	.00	.038	.317	.020	.022	.184	.022	.133	.000	.000
MBGO:1	32	.59	-.50	-.01	.020	.317	.038	.041	.184	.041	.133	.000	.000
MBRE:0	33	-.47	.10	.07	.037	.401	.022	.047	.376	.003	.017	.001	.008
MBRS:1	34	.80	-.17	-.12	.022	.401	.037	.082	.376	.005	.017	.002	.008

References:

[1] Anderson ,T.W, 1984, "An Introduction to Multivariate Statistical Analysis", 2nd Edition, Wiley.

[2] Bernábe L., B.; 2003, “Evaluación Estadística de -Datos Cualitativos en Comercio Electrónico”, Tesis Master, Universidad Iberoamericana -l Golfo Centro, Puebla, México.

[3] Bernábe L., B, Olsina, L, 2003 **“**Análisis Factorial de Correspondencias Simples en el procesamiento de datos cualitativos sobre funcionalidad en comercio electrónico”, Novena Conferencia de Ingeniería Eléctrica 2003, CINVESTAV-IPN.

[4] Chatfield, C.; Collins, A.J., 1991, "Introduction to Multivariate Analysis", Ed. Chapman & Hall.

[5] Dillon, W.R. & Goldstein, M., 1984, "Multivariate Analysis: Methods and Applications", Wiley, New York.

[6] Dixon,W.J , 1990, "BMDP Statistical Software Manual", Vol I, II., Dixon,W.J Eds, University of California Press, Berkeley, California.

[7] Kalakota, R.; Whisnton, A.B.,1997, “Electronic Commerce: A Manager’s Guide”, Addison-Wesley.

[8] Lafuente, G.H.; Oliveto, J.; Olsina, L.; 2000, “Requerimientos de Calidad en Sitios de E-commerce” Proceed. JUCSE 00, Nuevas Tendencias en Ingeniería de Software, Universidad Católica de Santiago del Estero, Arg., ISBN 950-31-0045-3

[9]. Loranca, M.B. & Olsina, L.; 2003, “Técnicas Estadísticas para el Análisis de la Calidad de Sitios Web “, 6o Workshop Iberoamericano de Ingeniería de Requisitos y Ambientes Software (IDEAS 2003), Asunción Paraguay.

[10]. Olsina, L.; Lafuente, G.J.; Rossi, G.; 2000, “E-commerce Site Evaluation: a Case Study”, Lecture Notes in Computer Science 1875, Proc. 1st International Conference on Electronic Commerce and Web Technology (ECWeb 2000) , Springer-Verlag, London-Greenwich, UK, pp. 239-252.

[11]. Ruíz Shulcloper. J., Alba, E. y Lazo Cortés, M. "Reconocimiento de Patrones", (ediciones I, II, II y IV). México, D.F. (1994). Diplomado de Titulación. Benemérita Universidad Autónoma de Puebla.