

Nonparametric Item Response Theory and Cluster Analysis to study patterns of Information and Communication Technologies Adoption in Brazil companies

Emerson Gomes dos Santos*

Paulista School of Policy, Economics and Business (Unifesp), Osasco, Brazil
emerson.gomes@unifesp.br

Renato de Castro Garcia

Institute of Economics (UNICAMP), Campinas, Brazil
renato.garcia.usp@gmail.com

Abstract

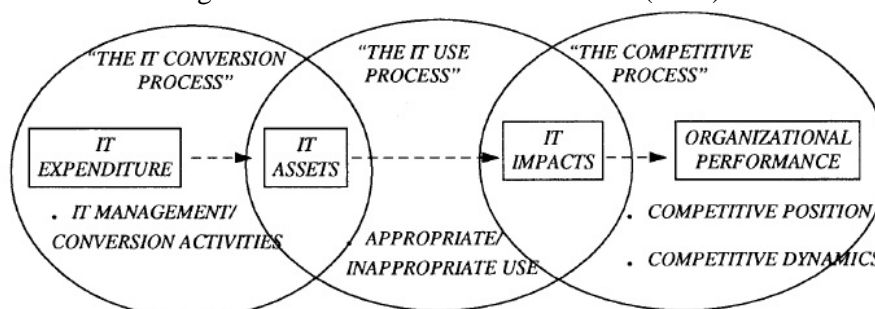
Information and Communication Technologies (ICTs) have assumed an increasingly important role in business development. The observation of the firm's characteristics can provide relevant elements to assess how technologies could increase the competitiveness of companies. In this study, nonparametric Item Response Theory was used in the survey questionnaire to define relevant dimensions to be used in a subsequent cluster analysis. After simulate many available methods and choose the one that best classified the units of analysis, this study describes how mapping of the ICT adoption in Brazilian companies and conclude that companies with Information Technology (IT) Expenditures are in a moment of ICT adoption relatively superior and possibly improve their competitiveness. The summary of this study motivates investments for the adoption and use of IT. The identification of factors that influence the performance may provide important information for the design of public policies aimed at support of the innovative activities in Brazilian companies.

Keywords: Cluster Analysis; Nonparametric Item Response Theory; Information and Communication Technologies.

1. Introduction

The observation of the firm's characteristics can provide relevant elements to assess how technologies could increase the competitiveness of companies. However, there are contradictions in results of the studies on Information and Communication Technologies (ICT) expenditures, summarized in productivity paradox or Solow paradox (1987). Organizations needed to justify why they spent more money on technology, although it is not always notice good results in your statistics. In this context, Soh and Markus (1995) proposed a framework that systematized concepts in the point of view: "how, when and why" Information Technology (IT) expenditures are converted into favorable organizational performance. Figure 1 shows the framework in three parts marked by circles: "The IT Conversion Process" related to IT Expenditure and IT Assets, "The IT Use Process" related to IT Assets to IT Impacts and, finally, "The Competitive Process" on IT Impacts and Organizational Performance.

Figure 1 – Soh and Markus Framework (1995)



The Soh and Markus framework was used for the definition of variables that represent a relatively complete picture of the characteristics of the studied companies. Therefore, this study describes how mapping of the ICT adoption in Brazilian companies from theoretical dimensions. The design of the variables include a calculated variable for IT expenditure and an application of nonparametric Item Response Theory (NIRT) in the questionnaire's items of a survey (CGI, 2011). In this aspect, the aim was to find ICT adoption patterns in Brazilian companies from IT expenditure and other selected variables. The hypothesis is that companies with IT expenditure would tend to have a better adoption of ICT.

2. Methods and Results

In order to study pattern of ICT adoption the data used come from the Survey on the Use of Information and Communication Technologies in Brazil (CGI, 2011). Data collection was conducted by telephone with a structured questionnaire (CATI - Computer-Assisted telephone Interviewing) on a sample of approximately 5600 companies with 10 or more employees.

The variable for IT Expenditure was calculated using Diegues (2010), who proposed a methodology related to how much payment (salary) were made to specific occupations related to software and service activities and reflects the previous spent of the companies (categorized in « yes » or « no »).

Regarding the variables and the theoretical dimensions - IT Management, IT Assets and IT Use - the idea was to select the questionnaire's items to create a coherent scale (Frame 1). All these items are obtained in « yes » or « no » questions, except connection speed. The IRT in its nonparametric version (NIRT) were applied and it were observed that the methodology used produces scales according to the theoretical model proposed by Soh and Markus (1995).

Frame 1 – Coefficient H for scales and items

Dimension	Variable	Coefficient H and (Standard Deviation)	
		Scale	Item
IT Management	Purchase	Escala 5 - 0.373 (0.054)	0.373 (0.054)
	Sale		0.373 (0.054)
	Introduced software	Escala 6 - 0.171 (0.026)	0.183 (0.031)
	Remote Access		0.150 (0.026)
	Developed software		0.175 (0.033)
	External Supplier	Escala 7 - 0.103 (0.027)	0.103 (0.027)
	Connection Speed		0.103 (0.027)
IT Assets	LAN (Wired Network)	Escala 2 - 0.427 (0.039)	0.486 (0.059)
	LAN (Wireless Network)		0.430 (0.045)
	Website		0.356 (0.024)
	Intranet	Escala 3 - 0.447 (0.050)	0.447 (0.050)
	Extranet		0.447 (0.050)
	ERP	Escala 4 - 0.407 (0.035)	0.407 (0.035)
	CRM		0.407 (0.035)
IT Use	IT Department	Escala 1 - 0.532 (0.021)	0.421 (0.024)
	Use of Open Source Software		0.560 (0.022)
	Use of Software by license		0.594 (0.023)
	Developed Software for use		0.546 (0.027)
	Use of Freeware Operating System		0.410 (0.031)

The coefficient H, also called Loewinger's H coefficients of scalability by Loewinger (1948) was used to check the validity of the model assumptions and reliability of the scale. From data provided by ICT Enterprises survey was found a strong scale for IT Use ($H \geq 0.5$); three reasonable scales to IT Assets ($0.4 \leq H < 0.5$) and three weak scales to IT Management ($H \leq 0.4$).

Table 1 presents a test of association in the first circle from Soh and Markus framework: "The IT Conversion Process" from IT Expenditure to IT Assets also considering IT Management variables. All variables are not independent of the IT, the only exception was the "external supplier" variable.

Table 1 – Chi-Square Test for IT Expenditures and variables

		IT Expenditure				Chi-Square Test
		No		Yes		p-value
Introduced Software	No	1510	80%	376	20%	< 0,0001
	Yes	938	66%	494	34%	
External Supplier	No	928	76%	296	24%	0.121
	Yes	1665	73%	603	27%	
Remote Access	No	1340	83%	278	17%	< 0,0001
	Yes	1263	67%	621	33%	
Connection Speed	1	218	84%	41	16%	< 0,0001
	2	658	79%	172	21%	
	3	1034	69%	461	31%	
	4	311	67%	153	33%	
	5	55	83%	11	17%	
	6	226	81%	53	19%	
Purchase	No	945	80%	235	20%	< 0,0001
	Yes	1624	72%	639	28%	
Sale	No	2278	75%	741	25%	0.0004
	Yes	309	68%	147	32%	
Developed software	No	1801	83%	381	17%	< 0,0001
	Yes	799	61%	516	39%	
LAN (Wired Network)	No	285	94%	19	6%	< 0,0001
	Yes	2320	73%	880	28%	
LAN (Wireless Network)	No	788	90%	90	10%	< 0,0001
	Yes	1816	69%	809	31%	
Intranet	No	1504	85%	274	15%	< 0,0001
	Yes	1106	64%	626	36%	
Extranet	No	1693	81%	389	19%	< 0,0001
	Yes	884	64%	506	36%	
ERP	No	1522	86%	250	14%	< 0,0001
	Yes	989	61%	642	39%	
CRM	No	1830	78%	502	22%	< 0,0001
	Yes	724	65%	390	35%	
Website	No	1019	91%	102	9%	< 0,0001
	Yes	1578	66%	798	34%	

For an initial mapping of the ICT adoption of the Brazilian companies was used Cluster Analysis, a multivariate analysis that following some steps and aims to form homogeneous groups of units. Therefore, the technique was applied to form, validate and interpret groups. In this study was used the ClusterSim package of software R, wherewith it is possible to simulate many available methods and compare them to choose the one that best classified the units of analysis.

All variables from dimensions – IT Expenditure, IT Assets, IT Management and IT Use - were considered for cluster analysis because they are dichotomous variables except connection speed. Therefore, all variables can have the same treatment in the application of clustering technique and directed the choice of similarity criteria. The similarity measure chosen to compare and group companies was the simple coefficient as called Cohen's Kappa coefficient.

Table 2 presents fifteen simulated cluster analysis of four or five groups and five methods (single, complete, average, k-means and McQuitty). The first column shows the ordering by validation method Silhouette. Silhouette is a validity measure for quality of group formation, which values close to 1 means a good allocation and negative values a misallocation. Thus, the grouping chosen for further analysis was the average of four groups. The distribution companies in these four groups resulting in 1225 in group 1, 1567 in group 2, 57 in group 3 and 15 in group 4.

Table 2 – Simulation for Cluster Analysis with methods *single, complete, average, k-means e mcquitty*.

	Number of Clusters	Method of Clustering	Validity measure (Silhouette)
1	4	Average	0.155
2	5	Average	0.125
3	4	K-means	0.125
4	4	Complete	0.115
5	6	Average	0.111
6	5	K-means	0.106
7	6	K-means	0.103
8	5	Mcquitty	0.088
9	4	Mcquitty	0.088
10	6	mcquitty	0.083
11	5	complete	0.079
12	6	complete	0.073
13	4	single	-0.023
14	5	single	-0.035
15	6	single	-0.064

3. Discussion and Conclusions

To summarize the information about the formed clusters the following simplification was adopted and present in frame 2. The proportion of companies with the presence of a given characteristic was categorized as low, medium or high, when the proportion of companies in the cluster was less than or equal to 40%, between 40% and 60% and equal to or greater than 60%, respectively. In this definition the most important are not the limits, but how to visualize and analyze how each group has different proportions of companies with the characteristics used in the clustering.

Frame 2 – Summary of the groups (clusters) by proportions of companies with the characteristics

		Groups (Cluster)			
		1	2	3	4
IT Expenditures		low	high	medium	Medium
IT Management	Introduced Software	low	high	low	low
	External Supplier	high	high	medium	low*
	Remote Access	low	high	medium	low
	Purchase	medium	high	low	medium
	Sale	low	low	low	low
	Developed software	low	medium	low	High
	LAN (Wired Network)	high	high	high	High
IT Assets	LAN (Wireless Network)	medium	high	high	High
	Intranet	low	high	low	high
	Extranet	low	medium	medium	low
	ERP	low	high	high	high
	CRM	low	medium	low	medium
	Website	medium	high	medium	high
IT Use	Use of Open Source Software	medium	medium	high	low
	Use of Software by license	high	high	medium	low
	Developed Software for use	low	medium	medium	high*
	Use of Freeware	low	high	high	medium
	Operating System	low	high	high	medium
	IT Department	low	high	high	high

One of the practical contributions of this study concerns the implementation of a cluster analysis that brought new information about the Brazilian companies, through the creation of four groups of companies that presents similar behaviors. Thereby contributes to the presentation of an initial mapping of the adoption of ICTs in Brazilian companies.

About the formed groups, it is concluded that Brazilian companies can be categorized according to the adoption of ICT and the companies with IT Expenditures (especially group 2) are in a moment of ICT adoption relatively superior to that did not have this type of expenditure (group 1).

The results indicate that companies with strategies that consider IT Expenditures have better adoption of technologies and possibly improve their competitiveness. The summary of this study directs to create a typology of national companies on IT and motivates investments for the adoption and use of IT. The identification of factors that influence the performance may provide important information for the design of public policies aimed at support of the innovative activities in Brazilian companies to increase competitiveness in different economic sectors and consequently its capacity for innovation.



References

Castells, M. (1999). *A sociedade em rede*. São Paulo: Paz e terra.

Diegues, A. C. (2010). *Atividades de Software no Brasil: Dinâmica Concorrencial, Política Industrial e Desenvolvimento*. Tese (doutorado), Programa de Pós-Graduação em Ciências Econômicas - UNICAMP, São Paulo.

Gutierrez, G. C. (2000). *Estimação das escalas dos construtos capital social, capital cultural e capital econômico e análise do efeito escola nos dados de Peru-PISA 2000*. Dissertação (Mestrado), Programa de Pós-Graduação em Engenharia Elétrica - PUC-Rio, Rio de Janeiro.

Hair, J. F.; Anderson, R. E.; Tatham R. L.; Black, W. C. (1998). *Multivariate data analysis*, 5.Ed., Prentice Hall, Upper Sanddle River, New Jersey, Chapter 11- Structural Equation Modeling.

Nelson, R.; Winter, S. (1982). *An evolutionary theory of economic change*. Harvard Un. Press.

Soh C.; Markus, M. L. (1995). *How it creates business value: a process theory synthesis*. In *Proceedings of the Sixteenth International Conference on Information Systems*. Amsterdam.

Solow, R. (1987). *We'd better watch out*, *New York Times Book Review*, pp. 36.