Poster presentation

Assessment of e-government maturity in Portuguese municipalities using regression and clustering approaches

Marco Costa[12], Gonçalo Paiva Dias[13]

[1] Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, Apartado 473, 3754-909 Águeda, Portugal
[2] Centro de Matemática e Aplicações Fundamentais da Universidade de Lisboa
marco@ua.pt
[3] Unidade de Investigação em Governança, Competitividade e Políticas Públicas da Universidade de Aveiro
gpd@ua.pt

Abstract: In order to evaluate the development of websites of the 308 Portuguese municipalities in this work it was performed an analysis using regression models and clustering techniques. That analysis allowed recognizing a group of socioeconomic variables that are significant to characterize homogenous groups of municipalities in what concerns e-government maturity.

Keywords: regression model; clustering; e-government; public administration.

1 Introduction

Public attention to performance analysis in the public sector has grown considerably in recent decades (Heinrich, 2008) and, particularly, in recent years, in the area of e-government. Statistical analysis on the performance assessment of e-government issues is also very recently. Most of these works apply linear regression models and correlation analysis as Mitra and Gupta (2008) or Kumar and Best (2006). To investigate the 'demand' side of e-government, Gauld et al. (2010) applied the multiple logistic regression. Principal components analysis (PCA) was applied in an study about citizens' attitudes towards e-government and e-governance in United Kingdom by Kolsaker and Lee-Kelley (2008). This paper focuses on the evaluation of website maturity of the 308 Portugal's municipalities regarding the features they offer. A combination of multivariate techniques, as regression models and clustering procedures, allows recognizing a group of variables that are significant to characterize some homogeneous groups of municipalities identified by cluster analysis.

2 Methodology

The websites of the 308 municipalities were classified according to the features available, namely, into three dimensions: information online (*Info*), online services (*Serv*) and online participation (*Particip*). Each component was classified according to an evaluation grid translating into a score from an ordinal scale (0-4 points). Thus, each municipality is characterized by a vector with three scores (*Info*, *Serv*, *Partic*). Moreover, a large set of variables was collected (19 variables almost all the National Institute of Statistics of Portugal, INE) that includes variables related with demographic characteristics, economic development, education levels, participation in government modernization programs, etc. Firstly, it was performed an exploratory analysis of variables and outliers were identified for some of them. From a global point of view, the analysis focuses on the sum of the three variables collected, i.e., in a new variable *Total* that indicates a global measure of the maturity of a website. The preliminary analysis indicated possible quadratic relations between some independent variables and *Total* that were considered in modelling procedure. An ordinary least squares (OLS) multiple regression model with backward procedure was fitted to identify a restrict group of exogenous variables that describes significantly the global indicator *Total* as a dependent variable, removing the independent variable with largest p-value (since more than 5%). The final model was validated by verification of the usual assumptions. In a supplementary analysis, a clustering procedure was performed to identify homogenous groups taking into account the websites' scores (*Info*, *Serv*, *Partic*). The clusters analysis considered the squared euclidian distance as disparity measure and the average linkage to hierarchical clustering process. The choice of number of clusters is performed analyzing clusters distance and R-squared criterion. Finally, the solution obtained in the clusters analysis was interpreted through an analysis of the variable's statistics.

3 Regression analysis

Table 1 summarizes the results of the final regression model

$$Total_i = \beta_0 + \beta_1 IRS_i + \beta_2 S1EdInitial_i + \beta_3 Digital_i + \beta_4 Population100_i +$$

$$+ \beta_5 PercUrbPop_i + \beta_6 MTT_i \beta_7 MTT_i^2 + \beta_8 MTV_i + \beta_9 EHR_i^2 + \epsilon_i$$

with $i = 1, 2, ..., 308$. It is possible to identify that the demand of digital services, as the online submission of tax forms (*IRS*), is significant to the municipality's score as well as the participation of the municipalities in the modernization program *Simplex* 2008/09 (*S1EdInitial*). Variables *Population100* (number of residents, unity=100 000) and *%UrbPop* (% of population residing in an urban area) represent a demographic characterization. The economic development of municipality is represented by tax

TABLE 1. Regression results.

variable	$\hat{\beta}$	std. error	p-value
Intercept	3.024	0.526	.000
IRS	.021	.009	.015
S1EdInitial	1.326	0.579	.023
Digital	.413	.179	.021
Population100	−1.764	.613	.004
%UrbPop	1.036	.377	.006
MTT	.049	.022	.025
MTT^2	−.001	.000	.011
MTV	.938	.442	.035
EHR^2	−.001	.000	.019
$R^2 = .224$			

variables as MTT (municipal tax on transfers of property, in millions of euros) and MTV (municipal tax on vehicles, in millions of euros). Last variable ERH (expenditure on human resources) incorporates one factor related with staff dimension, namely the expenditure on human resources. The assumptions of normality of errors was verified with the Kolmogorov-Smirnov test (Statistic=.038; p-value=.200).

4 Clustering analysis

A clustering procedure was performed to identify homogenous groups of municipalities considering the vectors of the initial three variables. It was considered the square euclidian distance as measure of disparity and the average linkage approach in the hierarchical clustering procedure. As there are 308 objects, it is very difficult to choose the number of clusters based on dendogram because the graphic is huge. Therefore, two approaches were implemented to support this choice:

- R-squared criterion (greater than 80%), $R^2 = \frac{\sum\sum n_{ij}(X_{ij}-\bar{X}_{ij})^2}{\sum\sum\sum(X_{ijk}-\bar{X})^2}$ performed with the support of the usual ANOVA one-way;
- distance between clusters obtained in the agglomeration process.

Combining these approaches, seven clusters were adopted. Figure 1 shows, in a geographical view, the classification of the 308 municipalities according to the solution of clusters procedure. Attending to statistics of variables in each cluster, presented in Table 2, the clusters analysis allowed the identification ofseven main profiles of websites with different median scores in the three assessment components.

TABLE 2. Characterization of clusters solution (medians for ordinal variables and averages to quantitatives.

Cluster	Info	Serv	Partic	Total	IRS	%UrbPop	MTV
1 (6.2%)	3	3	2	8	69.2	59.5	.81
2 (6.8%)	3	3	1	7	67.1	40.7	.50
3 (11.6%)	3	1	2	6	62.7	32.1	.30
4 (32.1%)	3	1	1	5	61.0	32.0	.34
5 (14.9%)	2	1	1	4	60.3	32.7	.28
6 (17.9%)	3	1	0	4	61.9	23.8	.20
7 (10.4%)	2	0	0	2	56.6	12.5	.09

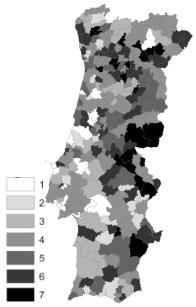

FIGURE 1. Representation of clusters analysis solution in Portugal's map.

References

Gauld R., Gold, S., Horsburgh S. (2010). Do they want it? Do they use it? The ?Demand-Side? of e-Government in Australia and New Zealand. *Gov Inform Q*, **27**, 177-186.

Heinrich, C.J. (2008). Advancing public sector performance analysis *Appl Stoch Model Bus Ind*, **24**, 373-389.

Kolsaker, A., Lee-Kelley, L. (2008). Citizens' attitudes towards e-government and e-governance: A UK study. *International Journal of Public Sector Management*, **21(7)**, 723-738.

Kumar, R., Best, M.L. (2006). Impact and Sustainability of E-Government Services in Developing Countries: Lessons Learned from Tamil Nadu, India. *Inform Soc*, **22**, 1-12.

Mitra, R.K., Gupta, M.P. (2008). A contextual perspective of performance assessment in eGovernment: A study of Indian Police Administration. *Gov Inform Q*, **25**, 278-302.