Analysis and Modeling of Complex Data in Behavioural and Social Sciences

Book of short papers

JCS - CLADAG 12

September 3-4, 2012 - Anacapri, Italy

Analysis and Modeling of Complex Data in Behavioural and Social Sciences

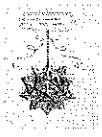
Book of Short Paper

Akinori Okada, Donatella Vicari, Giancarlo Ragozini
Editors

JCS – CLADAG 12

Joint meeting of the Japanese Classification Society and the Italian Classification and Data Analysis Group

Villa Orlandi – Anacapri, Italy
September 3-4, 2012


clelup

Prima edizione: agosto 2012

ISBN 978-88-6129-916-0

Printed by CLEUP sc
Coop. Libraria Editrice Università di Padova
Via G. Belzoni, 118/3 - Padova (Tel. 049/650261)
www.cleup.it

CLADAG Solicited Session 3
Analysis and Modelling of Relational Data in the Behavioural and Social Sciences 2
Organizers: Maria Rosaria D'Esposito, Susanna. Zaccarin

Social Network Analysis for Social Planning Evaluation
Rosaria LUMINO, Concetta SCOLORATO

Statistical Characterization of Virtual Water Trade Network
Alessandra PETRUCCI, Emilia ROCCO

Evaluating Cooperative Behaviours in Innovation Networks: a Prespecified Blockmodeling Analysis
Laura PROTA, Maria Prosperina VITALE

Research Institute Analysis Based on Collaborative Research Activities
Nagayoshi YAMASHITA, Yuan SUN, Kei KURAKAWA, Yasumasa BABA

Assessing Cooperation in Open Systems: an Empirical Test in Healthcare
Paola ZAPPA

A Multidisciplinary Approach to Validate a Friendship Psychometric Scale: a Study in an Italian Sample of Students
Emma ZAVARRONE, Sandra DE FRANCISCI EPIFANI

Contributed papers Sessions

Spatial Clustering of Earthquakes by Data Depth
Claudio AGOSTINELLI, Mario ROMANAZZI

Clustering by Considering the Local Density of Units
Vladimir BATAGELJ

The Aggregate Association Index and its Extensions
Eric J. BEH, Duy TRAN, Irene L. HUDSON, Linda MOORE

A Competing Risks Analysis of Italian University Careers
Matilde BINI, Margherita VELUCCHI

Scale Reliability Evaluation for *a-priori* Clustered Data
Giuseppe BOARI, Gabriele CANTALUPPI, Marta NAI RUSCONE

Teenagers and Mass Media in Messina: an Estimation of the Cumulative Proportional Odds Model

Zirilli A., Alibrandi A., Giacalone M., Ucchino R. B.[1]

Abstract The aim of this paper is to analyze the relationship between teenagers and media, in order to try to better understand the habits and to conduct analysis on social interactions with young people. From the methodological point of view, we estimated ordinal logistic regression model, to test the dependence of mass media influence and inspiration with respect to the preference of kind of television programs, the time spent on TV and on computer and the most used social network.

1 Introduction

The media are constantly evolving at a breakneck pace and mainly young people, with their flexibility of mind and their continuous curiosity, can better understand the characteristics and potential of such advances. These are precisely the reasons for the success of the two most visited sites in the world, according to the recent ranking by Google (July 2011): Facebook is mostly visited site and YouTube is second one coming after. As part of the Scientific Degree Plan, established by the Ministry of Italian University and Scientific Research Group, for the scholastic year 2010/2011, a

[1] Zirilli Agata, Department of Economical, Financial, Social, Environmental, Statistical and Territorial Sciences, Messina University, azirilli@unime.it

Alibrandi Angela, Department of Economical, Financial, Social, Environmental, Statistical and Territorial Sciences, Messina University, aalibrandi@unime.it

Giacalone Massimiliano, Faculty of Political Sciences, Calabria University, maxgiacit@yahoo.it

Ucchino Rosaria Barbara, Department of Economical, Financial, Social, Environmental, Statistical and Territorial Sciences, Messina University, ucchinor@unime.it

laboratory of "statistics for class" has been researched. To evaluate the influence of mass media in the life of the people, the Statistics degree course of Messina University, decided to plan an investigation to recognize the influence portion of the media in the life of Messina's teenagers. A questionnaire entitled "Perceptions of the influence exerted by mass media" was administered to a sample of 449 subjects, all students of secondary school in Messina province (42% males and 58% females).

2 The Cumulative Proportional Odds Model

2.1 Methodology

The *Cumulative Proportional Odds Model* is a model for ordinal data; it is obtained by using a *logit* link for the cumulate probabilities (Agresti, 2010; O'Connel, 2006). The model is an extension of the logistic regression model for dichotomous data (Kleimbaum and Mitchel, 2002) at categorical ordinal data. The most interesting point of this model is the observation of a particular score or less; for an ordinal variable with three modalities it's possible to define the following *odds*:

$$\Theta_1 = \text{prob. (score of 1) / prob. (score greater than 1)} \quad (1)$$

$$\Theta_2 = \text{prob. (score of 1 or 2) / prob. (score greater than 2)} \quad (2)$$

$$\Theta_3 = \text{prob. (score of 1, 2 o 3) / prob. (score greater than 3)} \quad (3)$$

The last category doesn't have an odds that is associated with it since the probability of scoring up to and including the last score is 1.

All of the odds are of the form:

$$\Theta_j = \text{prob (score} \leq j) \text{ / prob(score} > j) \quad (4)$$

We can also write the equation as:

$$\Theta_j = \text{prob (score} \leq j) \text{ / (1- prob (score} \leq j) \quad (5)$$

since the probability of a score greater than j is 1 minus probability of a score is less than or equal to j. The ordinal logistic model for a single independent variable is :

$$ln(\Theta_j) = \alpha_j - \beta X \quad (6)$$

where $j=1,\ldots, j-1$, with j categories number. Larger coefficients indicate an association with lager scores. For a continuous variable, a positive coefficient indicates that since the values of the variable increase, the likelihood of higher scores increases. A negative coefficient indicates that lower scores are more similar and close each other. An association with higher scores shows smaller cumulative probabilities for lower scores, since they are less close to occur (Soliani, 2004). Each *logit* has its own term α_j, but the same coefficient β. That means that the effect of the independent variable is the same for different *logit* function. The ordinal logistic model is based on the assumption that includes a continuous latent variable and that the ordinal observed result derived from discretization of a underlying continuous variable. The estimation of the model was carried out using the Polytomous Universal Model (PLUM) procedure of SPSS (Norušis Marija, 2009) which provides estimates of models for ordinal data, in particularly for the Generalized Linear Models (Mc Cullagh and Nelder, 1989).

2.2 *Application*

In form of an analytical view, three statistical models were estimated in order to recognize the dependence of the mass media influence (ordinal variable expressed by 4 ordered categories: 1 = nothing; 2 = low; 3 = average; 4 = high) by:

- the kind of TV program: cartoon, cultural, films, entertainment; reality; sport;
- the time spent on TV viewing and computer use;
- the kind of the most used social networks (Twitter, Facebook and YouTube).

Results are shown in the following tables 1-3:

Table 1: *Cumulative Proportional Odds Model for the kind of television programme*

VARIABLES	COEFF.	S.E.	P-VALUE
Constant 1	-3.332	0.307	0.000
Constant 2	-1.215	0.237	0.000
Constant 3	-0.217	0.229	0.342
Cartoons	0.153	0.198	0.439
Cultural	-0.715	0.204	**0.000**
Film	0.565	0.547	0.302
Entertainment	0.173	0.201	0.391
Reality	0.467	0.203	**0.022**
Sport	-0.365	0.195	0.061
-2Log-Likelihood=290.3; Chi Square=20.7; **p=0.002**			

Table 2: *Cumulative Proportional Odds Model for time spent on television and on pc*

VARIABLES	COEFF.	S.E.	P-VALUE
Constant 1	-2.144	0.384	0.000
Constant 2	0.041	0.342	0.904
Constant 3	1.037	0.346	0.003
Time spent on TV	-0.034	0.037	0.364
Time spent on computer	0.319	0.130	**0.014**
-2Log-Likelihood=222.2; Chi Square=7.37; **p=0.025**			

Table 3: *Cumulative Proportional Odds Model for the kind of social network*

VARIABLES	COEFF.	S.E.	P-VALUE
Constant 1	-2.762	0.280	0.000
Constant 2	-0.747	0.220	0.001
Constant 3	0.269	0.217	0.215
Twitter	0.127	0.195	0.517
Facebook	0.422	0.385	0.273
YouTube	1.231	0.506	**0.015**
-2Log-Likelihood=370.8; Chi Square=17.6; **p=0.041**			

For the three applications of the model, we estimated Pearson test and Deviance tests; neither of them showed significance, denoting an adequate goodness of fit.

Furthermore, we estimated the Cox and Snell test, the Nagelkerke test and the McFadden test. In all three cases they provided information about a discrete degree of model fit to the observed data. Examining the first model and considering all the programs showed by the TV, the teenagers are more likely to follow the Reality, which exert a significant influence in the life of themselves. For cultural programs, however, we find a significance, but, unlike the reality, the coefficient is negative, denoting the perception that those programs minimally affect the choices of young people. The second model shows that people who spend more time in front of the PC is more influenced than who mostly watch television. Moreover, as showed in the third model, we note that, among the mainly used social networks, the first place is occupied by YouTube, rather than Facebook; this result is in contrast to what happens at the national level.

3 Final Remarks

The three applications of the *Cumulative Proportional Odds Model* showed that among the television programs, the cultural ones and the reality are remarkable, although with different influence. The teenagers perceive that cultural programs have minimal influence on their choices; on the contrary, the reality seem to represent for young people an attractive lifestyle to imitate. The Messina's teenagers who spend more time in front of the PC is more influenced than who mostly watch television. YouTube represents the social network mostly used by teenagers in Messina; in addition, examining the questionnaire responses, we noticed that the factors related to the success of such video platform derive mainly from watching videos of others and listening to music; these activities exclude all other possible functions of social network that would require more technical skills. Enchantment of the mass media affects mostly young people: they are "bombarded" by advertising and models, which fascinate them and influence their choices. Many youth behaviors, such as fashions and habits, are due to the mass media and to the messages they express. The media have a wide social legitimacy by the political and economic power and, for this, the viewer is led to trust according to mechanisms different from those used in normal life: in the media there is no direct and personal communication.

References

1. Agresti A. (2010), Analysis of Ordinal Categorical Data, 2 edition, Hoboken , NJ: Wiley.
2. Kleinbaum D.G., Mitchel K. (2002), Logistic Regression: a self learning text, Second edition, Springer-Verlag, New York.
3. Mc Cullagh P, Nelder J.A. (1989), Generalised Linear Models, Second edition, Chapman and Hall/CRC, USA.
4. Norušis Marija J. (2009), PASW Statistic 18. Statistical Procedures Companion, Pearson Education.
5. O'Connel A.A. (2006), *Logistic regression models for ordinal response variables.* Thousand Oaks: Sage.
6. Soliani L. (2004), Manuale di Statistica per la ricerca e la professione. Statistica univariata e bivariata parametrica e non, nelle discipline ambientali e biologiche, Uninova, Parma.