

GLM Approach to Evaluate Effects of Measure Scales for Different Methods to Fill a Questionnaire

Un Approccio GLM per la Valutazione degli Effetti delle Scale di Misura con Differenti Metodi di Somministrazione del Questionario

Enrica Aureli and Domenica Fioredistella Iezzi
Dipartimento di Statistica, Probabilità e Statistica Applicata
Università "La Sapienza" Roma
enrica.aureli@uniroma1.it stella.iezzi@uniroma1.it

Abstract: In questo lavoro, si esamina la qualità della vita a Roma attraverso la percezione delle paure individuali e sociali dei cittadini nelle diverse zone della città. Si propone un approccio esplorativo e predittivo basato sull'analisi delle corrispondenze multiple e sul modello generale linearizzato.

Gli obiettivi sono: esplorare le paure individuali e sociali e cercarne le dimensioni latenti; comparare le scale di misura utilizzate nella raccolta dei dati e valutare se vi sono delle differenze significative dipendenti dalle tecniche di intervista sperimentate (CATI e autocompilato)

Keywords: Computer Automatic Telephone Interview (CATI), General Linear Model (GLM), Multiple Correspondence Analysis (MCA) and Self Interview

1. Introduction

In this paper, we examine the risk perception (Thurstone, 1928; Sjöberg, L. 2000). Recent studies (Giddens, 1999; Z. Bauman, 2000,) have shown that problems of post-modernity are latent or real risks. We investigate on perception of risk and we compare different scales of measure. In social study, risk perception is measured by ordinal scale mainly to verbal category (Lauwrence E., Hazelrigg and Melissa A. Hardy, 2000; Eijk, C. van der, 2001). The present paper aims to:

- i) compare different scales of measure and evaluate the influence of methodology used on the survey.
- ii) explore social and individual risk and identify latent dimensions;
- iii) verify if social latent risks depend on individual latent risks and if there are significant effects that come from method of survey.

2. Content of research

In a survey on quality of life in Rome, we have collected data on perception of individual and social risks by a questionnaire, that focuses attention on six topics: food, health, environment, traffic, criminality and relationship (see table 1). Respondents have been selected by stratified random sampling.

An identical questionnaire has been submitted to different groups of subject: one group contacted by telephone with a Computer Automatic Telephone Interview (CATI) and a other has filled in a Self Interview questionnaire.

Each area of the questionnaire presents a set of structured questions, with different numbers of questions for each section. We analyse individual and social risks by verbal and numerical scale, respectively. Verbal scale presents the following modality: “much”, “enough”, “few”, “none” and neutral modality; “I don’t know”; numerical scale is from 1 to 10.

Social and individual risks are developed in separate manners, but they interact among themselves. Individual risk is the feeling of anxiety and of anguish that occurs when one suffers damage; for example, victims of a theft, an aggression, a robbery or a violation. Those feelings derive from a perception of probability that realise the damage. Social fear is demonstrated as a form of preoccupation and the fear of possible aggressions in the living area. It is a feeling that does not produce anguish. It is demonstrated in the need to restart social order. Social risk concerns the whole community, while it exists in a reduced manner in the life of the individual.

Table 1 *Questions examined*

How high is the risk...	Individual risk perception	Social risk perception
Food	beef meat, genetic modification, mineral water in bottle, tap water	food in general
Health	health	health in general
Environment	pollution, noise, water, electromagnetic fields	environment in general
Traffic	by car, by motor bicycle, by bicycle, by train, by the underground, by bus, on foot	traffic in general
Crime	robberies, snatching, pick pocketing, intimidation in one’s own backyard, in the street and on the work place	crime in general
Relationship	relationship	relationship in general

3. The method

The analysis is developed in two steps:

1. Univariate;
2. Multivariate.

In the first step, we examine distribution of frequencies of items, we apply non parametric tests of Mann Whitney and Wilcoxon to evaluate existence of significant difference between CATI and Self Interview. We apply indices for ordinal data (Maravelakis P.E., Perakis, M., Psarakis S., and Panaretos J.; 2003), with the aim to measure the degree of concentration of the “positive” or “negative” answers in examined questions.

In the second step, we explore and individualise latent dimensions by multiple correspondences analysis (MCA) (Greenacre, 1995). The collected data has been organised into two data matrices: $\mathbf{A}=\{x_{ij} : i \in \mathbf{I}, j \in \mathbf{J}\}$, where x_{ij} is the value of j^{th} preference expressed by i^{th} respondent on social risks and $\mathbf{B}=\{x_{ij} : i \in \mathbf{I}, j \in \mathbf{J}\}$, where x_{ij} is the value of j^{th} preference expressed by i^{th} respondent on individual risks. Gender, age, level of education, marital status and profession have been used like illustrative

variables. MCA allows the construction of principal components, which optimally summarise the data, and enables the construction of graphical displays. Principal factors have been organised into two data matrices: $\mathbf{X} \equiv \{c_{ij} : i \in \mathbf{I}, j \in \mathbf{J}\}$, where c_{ij} is the value of j^{th} individual principal factor belongs to i^{th} respondent and $\mathbf{Y} \equiv \{s_{ij} : i \in \mathbf{I}, j \in \mathbf{J}\}$, where s_{ij} is social principal factor belongs to i^{th} respondent. A multiple general linear regression (MGLM) was tried in order to verify the dependence between social and individual latent risks and to evaluate if there were significant differences between CATI and Self compilation (Nelder, J. A e Wedderburn R. W. M, 1972; Goldstein, 1995, Hox, 1995; Snijder and Bosker, 1999). In this paper, the matrix $\mathbf{Y}$ contains predictor variable, the matrix $\mathbf{X}$ represent covariates and factor fixed is dummy variables “CATI or Self Interview”.

The GLMs go therefore beyond allowing linear transformations or linear combinations of subordinate multiple variables. The method used in order to find the coefficient β is the same of multiple regression in different sets of regression coefficients are separately found for the m different subordinate variables. Its formulation becomes:

$$\begin{aligned} \mathbf{Y} \mathbf{M} &= \mathbf{X} \beta + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 \mathbf{I}) \end{aligned} \quad (1)$$

where $\mathbf{M}$ is a matrix ($m \times s$) of coefficients that define s linear transformations of the subordinate variables. ; ε is a vector ($n \times 1$) error assumed like uncorrelated, with mean zero and variance constant σ^2 ; β is a vector ($p \times 1$) of unknown parameters and estimator $\mathbf{B}$ of β is obtained with Ordinary Least Squares (OLS).

The normal equation becomes $\mathbf{X} \mathbf{T} \mathbf{X} \mathbf{B} = \mathbf{X} \mathbf{T} \mathbf{Y} \mathbf{M}$ and the solution to this equation is given by $\mathbf{B} = (\mathbf{X} \mathbf{T} \mathbf{X})^{-1} \mathbf{X} \mathbf{T} \mathbf{Y} \mathbf{M}$. GLM gives the opportunity advantage to find a solution for normal equation also if variables of matrix $\mathbf{X}$ are not linear independent and does not exist inverse of $\mathbf{X} \mathbf{T} \mathbf{X}$. It is used as a balanced model to fixed effect and only the intercept is variable, while coefficient of regression β and the variance within group are considered constant.

4. Principal result and conclusion

Univariate analysis showed that the mode of questions on risk perceptions of food and health are not different, while for questions about the traffic, crime and relationship were different in respect to the survey. On the basis of the first analysis we cannot affirm that there is a systematic effect dependent on the survey. To verify statistical significance applied non parametric tests of Mann-Whitney and Wilcoxon. Those tests in a social risk perception show a significant difference between Cati and Self Interview for three of six items examined (traffic, crime and relationship); in individual risk perception, nine on twenty-two items (genetic modification, mineral water in bottle, health, noise, traffic by motorcycle, by bicycle, by train, by bus, by the underground, on feet and crime in the street). Indices show that in CATI method the respondents give fast answers, often not reflecting; in Self Interview the answers are more concentrated in central values. An interesting result is given regarding the relationship questions where there is a high concentration of negative answers in the Self Interview, compare to the CATI results.

In second step, we have found that in social domain, latent factors have resulted 2: health, traffic and crime; 2) food and relationships.; in individual domain, latent factors are 3: 1) traffic and criminality; 2) food, health and environment 3) relationship. Social latent factors and individual latent factors produce significant results, but the interaction of dummy (CATI and Self Interview) is not significant (see table 2).

Table 2 *Parameter Estimates*

Dependent Variable	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Social latent factor 1	Individual latent factor 1	0,502	0,026	19,580	0,000	0,451	0,552
	Individual latent factor 2	-0,145	0,027	-5,384	0,000	-0,198	-0,092
	Individual latent factor 3	0,071	0,026	2,677	0,008	0,019	0,123
	CATI	-0,084	0,044	-1,886	0,060	-0,170	0,003
	Self Interview	0,007	0,036	0,203	0,839	-0,064	0,079
Social latent factor 2	Individual latent factor 1	-0,197	0,028	-7,007	0,000	-0,253	-0,142
	Individual latent factor 2	-0,290	0,030	-9,795	0,000	-0,348	-0,232
	Individual latent factor 3	0,064	0,029	2,189	0,029	0,007	0,121
	CATI	0,051	0,049	1,041	0,298	-0,045	0,146
	Self Interview	-0,007	0,040	-0,166	0,868	-0,085	0,072

References

- AA.VV.(2001) *Ambiente urbano e naturale: la percezione del rischio a Roma*, Academic Press.
- Bauman Z. (2000). *La società dell'incertezza*. Il Mulino ed., Bologna.
- Beck U. (1999). *Che cos'è la globalizzazione*, Carocci ed., Roma.
- Eijk, C. van der (2001). Measuring Agreement in Ordered Rating Scales. *Quality and Quantity*, 35: 325-341.
- Giddens A (1993) *The consequences of Modernity* by Polity Press, Stanford, CA.
- Greenacre, M. (1984), *Theory and applications of correspondence analysis*, London, J.J. Hox (1995) *Applied Multilevel Analysis*, TT- TT-Publikaties, Amsterdam.
- Maravelakis P. E., Perakis M., Psarakis S., Panaretos J. (2003) .The Use of Indices in Survey , *Quality and Quantity* (37), Kluwer Academic Publishers, Netherlands.
- Nelder, J. A e Wedderburn R. W. M (1972). Generalized Linear Models, *Journal of the Royal Statistical Society*, A 135, pp. 370-84.
- Tom Snijders and Roel Bosker (1999). *Multilevel Analysis*, SAGE Publications.
- Lauwrence E., Hazelrigg and Melissa A. Hardy (2000). Scaling the semantics of satisfaction, *Social Indicators research* 49: 149-180.
- Sjöberg, L. (2000). The Methodology of Risk Perception Research. *Quality and Quantity*, 34: 407-418.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology* 33: 529-554.