Statistica & Applicazioni
Vol. III, n. speciale 1, 2005
pp. 45-52

Multilevel models for perceived quality analysis: the case of local public transport of Rome

Simone Borra[§]

Sara Chiavarini[§]

Summary: *The presence of heterogeneity among the 19 boroughs (Municipio) of Rome gives rise to a group effect on the evaluations expressed by citizens about perceived quality of local public transport. By fitting an ordered logit multilevel model with random intercept it is useful to explain the heterogeneity in terms of individual and contextual explanatory variables. The results show that citizens' opinion on local public transport quality depends not only on individual characteristics, but also on contextual indicators concerning demographic and environmental features of borough in which any citizen lives.*

Keywords: *customer satisfaction, perceived quality of transport, ordered logit multilevel model.*

1. Introduction

The aim of this paper is to apply a multilevel model to explain the perception of citizens about the local public transport (LPT) quality of Rome in terms of a set of individual and contextual explanatory variables. The available data come from the "Third Survey on the public local services quality in the city of Rome" conducted in 2002 by the *Agency for the control and the quality of public local services of Rome.*

The data set consists of a random sample of 4039 Roman citizens, stratified by gender, age group and boroughs (Municipio) in which Rome is divided.

The result of multi-stage sampling is a hierarchical structure of data: citizens are nested within the 19 boroughs of Rome. As expected, individual behaviours are influenced by the context in which citizens live. Therefore, the respondents' opinion on the LPT quality depends not only on individual characteristics but also on the characteristics of the place where they live.

In order to take into account the hierarchical structure of data we can use multilevel models which are able to describe relations between *individual*

[§] Dip. SEFEMEQ - Università degli Studi di Roma "Tor Vergata"- via Columbia 2, 00133 Rome (e-mail: borra@economia.uniroma2.it ; sachiavarini@yahoo.it).

and *contextual* levels of hierarchical population and to estimate the group effect on individual behaviours (A.S. Bryk, S.W. Raudenbush, 1992).

2. The model

The dependent variable studied in our model is the opinion on LPT quality, an ordinal response with five categories which represent the levels of citizens' satisfaction: 1=*very bad;* 2=*bad;* 3=*sufficient;* 4=*good;* 5=*very good.*

In order to consider hierarchical structure of data and ordinal nature of response variable, it is correct to specify a two-level ordered logit model with random intercept (Snijders T.A.B., Bosker R.J., 1999).

Let Y_{ij} denote the response of individual *i* within cluster *j*. It is possible to represent the model as a threshold model where an underlying latent continuous response Y_{ij}^{*} determines the observed response Y_{ij} in this way:

$$Y_{ij} = 1 \text{ if } Y_{ij}^{*} \le k_1;\ Y_{ij} = 2 \text{ if } k_1 < Y_{ij}^{*} \le k_2\ \ldots; Y_{ij} = 5 \text{ if } Y_{ij}^{*} \ge k_4$$

The latent variable is assumed as a multilevel random intercept model:

$$Y_{ij}^{*} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}'\mathbf{z}_{j} + u_{j} + \varepsilon_{ij} \qquad (1)$$

under the assumptions:

$u_j \sim \mathrm{N}\left(0, \tau^2\right);\ \varepsilon_{ij} \sim \text{logistic}\left(0, \pi^2/3\right); Cov\left(u_j, \varepsilon_{ij}\right).$

Moreover, $\mathbf{x}_{ij}$ is the vector of explanatory observed variables at the individual level; $\mathbf{z}_j$ is the vector of explanatory observed variables at the group level; $\boldsymbol{\beta}$ is the vector of fixed coefficients related to the individual variables; $\boldsymbol{\gamma}$ is the vector of fixed coefficients related to the group variables; u_j is a random variable representing the group *j* random effect; ε_{ij} is the first level stochastic component.

The relation between the opinion on local public transport quality and the observed variables can be represented as a cumulative probability:

$$Pr\left(Y_{ij} \le c \middle| u_j\right) = Pr\left(Y_{ij}^{*} \le k_c \middle| u_j\right) = \frac{exp\left(k_c - \boldsymbol{\beta}'\mathbf{x}_{ij} - \boldsymbol{\gamma}'\mathbf{z}_j\right)}{1 + exp\left(k_c - \boldsymbol{\beta}'\mathbf{x}_{ij} - \boldsymbol{\gamma}'\mathbf{z}_j\right)} \qquad (2)$$

where *c* is the number of response categories; k_c are the thresholds assumed to be constant among the groups. The Maximum Likelihood estimates of model parameters have been obtained by the GLLAMM procedure

implemented in STATA software (Rabe-Hesketh S. et al., 2001; Grilli L., Rampichini C., 2001).

3. Descriptive analysis and results

The individual variables used in our model arise from responses of the questionnaire, submitted in the Third Survey, to get the opinion by citizens on local services quality of Rome. Contextual indicators, instead, come from administrative sources. The choice of local public transport as object of this research rises from realty of city of Rome, where public transport impact gets over citizens satisfaction, including life quality, Rome liveability and environmental issues.

From the model with only random intercept it is possible to calculate the intra-class correlation coefficient, a measure of a group effect, that leads to the value of 0.054. This means the 5.4% of the variability giving an opinion on LPT quality is attributable to the group heterogeneity. The hierarchical structure of data set gives rise to a positive correlation between citizens belonging to the same group.

The group heterogeneity can be observed through the following map of Rome, where we considered three new response modalities by aggregating "very bad" and "bad" into "negative opinion", "good" and "very good" into "positive opinion", and "sufficient opinion". Different colours indicate the most frequent opinion level observed in each borough.

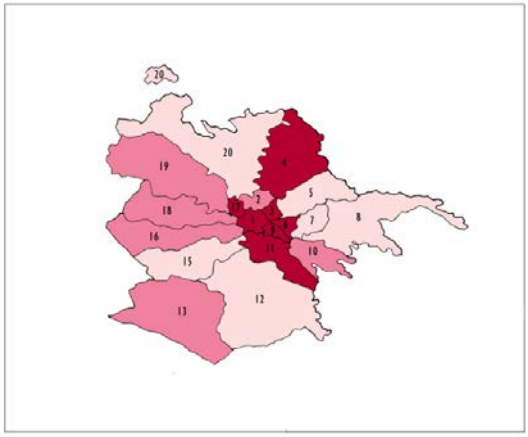


Figure 1. *Distribution, across boroughs, of the mode of opinion on local transport quality.*

Differences among citizens in terms of opinion on LPT quality can be explained by fitting the model including:

- individual variables, related to socio-demographic characteristics, frequency use of public and private transports and personal opinion concerning responsibility about transport inefficiency;
- contextual indicators, related to transport system, demographic and environmental characteristics for each borough.

Table 1 shows the estimated coefficients related to the selected model: positive coefficients indicate a positive relation between the explanatory variable and the probability of giving a favourable opinion on local transport's quality.

Table 1. *Estimated two-level ordered logit model with random intercept.*

Variable	**coefficient**	**p-value**	**OR**
Individual variables			
Age	- 0.0070	0.000	0.9930
Gender (*1=Female 0=Male*)	- 0.1645	0.008	0.8483
Occupational status (*1=unemployed, student, housewife, pensioner; 0=employed*)	- 0.1224	0.065 *0.000*	0.8847
Use of public transports (*1=yes, 0=No*)	0.2165	0.006	1.2417
Frequency use of private transports		*0.000*	
Monthly frequency use of private transports (*1=yes, 0=No*)	0.1912	0.075	1.2106
Weekly/daily frequency use of private transports (*1=yes, 0=No*)	- 0.2611	0.001	0.9324
Responsibility imputed to Atac-Metro-Trambus (*1=no to 4=much*)	- 0.1614	0.000	0.8509
Responsibility imputed to Mayor of Rome (*1=no to 4=much*)	- 0.1752	0.000	0.8392
Context variables			
Presence of one tube stop at least	0.4044	0.000	1.4984
Bus stop per km^2 road	- 0.0070	0.000	0.9930
Growth rate of population	- 0.2690	0.001	0.7641
Difference between public green areas per km^2 road and the mean of Rome	0.1242	0.106 *0.019*	1.1321
Threshold 1 (k_1)	-4.2126	0.270	
Threshold 2 (k_2)	-2.6083	0.264	
Threshold 3 (k_3)	-0.9716	0.261	
Threshold 4 (k_4)	1.3235	0.267	
Log-likelihood	-5165.08		
Second level variance τ^2	0.099		
Intra-class correlation coefficient ρ	0.029		

We can note all regression coefficients result significant. The value in italic corresponds to the *p-value* of LR test.

The effect of each variable produced on the response probability is given by the Odds Ratio (OR), which is calculated by $exp(\hat{\beta})$ and $exp(\hat{\gamma})$. Table 1, in the last column, shows OR values for each variable. If OR>1, the probability of giving a higher response modality increases as explanatory variable value increases.

As expected, we find that Age, Gender and Occupational status influence negatively the response probability. Regarding use of transport means, we notice differences between public and private transport users. In particular, people who use private means more frequently (weekly/daily) present a lower probability of giving a favourable opinion, than people who use monthly private means. Finally, we observe response probability is influenced negatively by the responsibility imputed to Transport Companies and Mayor of Rome about service quality.

As regards context effect, we observe the probability of giving a favourable response is higher in borough where is located tube line (at least one tube stop). Whereas, increasing the number of bus stop per km^2 road decreases the probability of giving a favourable opinion. In fact, as a consequence, the waiting time and routing time can increase considerably. In borough with high growth rate of population, the probability of giving a favourable opinion decreases. These boroughs, as the number 8, 12 and 19, are located in large suburban areas of Rome, where it is more difficult to deliver public services than in central areas of Rome. The last context indicator is related to municipal green areas, as a proxy of environmental policy to save population health and to reduce air pollution. We observe, moreover, the probability of giving a favourable opinion is higher in boroughs of centre of Rome, as the number 1, 2, 3 and 17, where there are large public parks, large archaeological areas and limited traffic zones.

The second level variance τ^2 is decreased considerably (from 0.1829 to 0.099) and the intra-class correlation coefficient ρ is decreased too (from 0.0525 to 0.029).

4. Predicted Probabilities

Estimated model allows us to calculate the probabilities of giving an opinion, from very bad to very good. Through cumulative probabilities it is possible to make a comparison among the 19 boroughs. The following graph shows the municipal heterogeneity: for instance, the probability of giving a negative opinion is higher in borough 5, 8, 13 and 20 (see also light coloured area in Figure 1); the probability of giving a favourable opinion is higher in borough 1, 2, 3, 9 and 17 (see also dark coloured area in Figure 1).

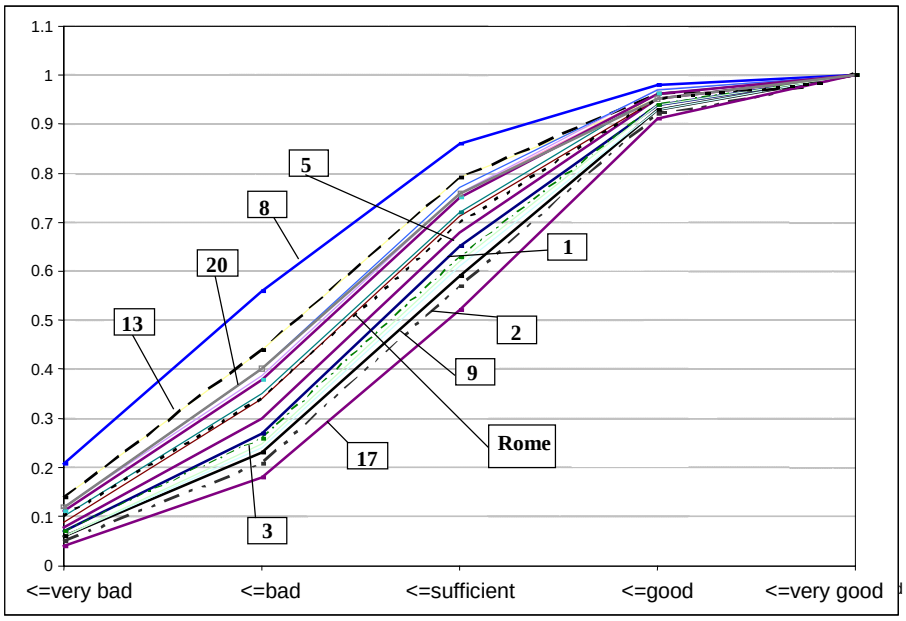


Figure 2. *Estimated cumulative probability functions of opinion on LPT quality by borough.*

5. The first and second level variability

In order to analyse variability both at individual and group level, it is necessary to isolate respectively the influence of contextual characteristics and the influence of individual ones.

For example, by using three different users it is possible to show how the response probability varies according to individual characteristics, like Age, Occupational status and Life habits in general.

The bar chart in Figure 3 points out the first level variability: taken a borough, the probability of giving a favourable opinion on LPT quality depends on individual characteristics. The third user, an old woman (65 years old) who takes neither public nor private means and imputes much responsibility both to Atac-Metro-Trambus and Mayor of Rome, presents a lower value of probability of giving a favourable opinion (0.16) than the first user (0.48), a young unemployed woman (20 years old), who takes public means and imputes responsibility neither to Atac-Metro-Trambus nor to Mayor of Rome. The second user, a 45 years old employed man who takes private means daily and imputes responsibility neither to Atac-Metro-Trambus nor to Mayor of Rome, presents an intermediate value (0.28).

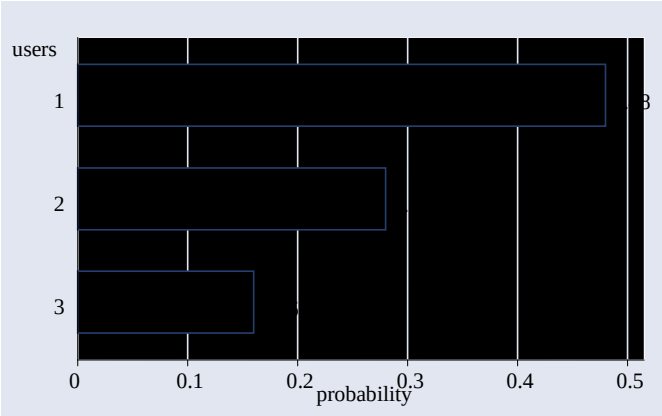
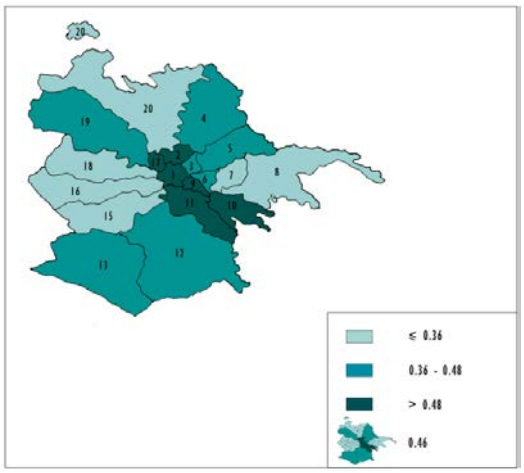


Figure 3: *Probability of giving a favourable opinion on LPT quality by kind of user.*

The map in Figure 4 shows the group effect: chosen a user the probability of giving a positive opinion (from "sufficient" to "very good") on local transport's quality depends on the contextual indicators of borough in which he/she lives.



For instance, a young unemployed woman (20 years old) presents a probability greater than 0.48 of giving a positive opinion if she lives in borough 1, 2, 9, 10, 11 and 17.

Figure 4: *Probability of giving a positive opinion on LPT quality by borough.*

From model estimates it is possible to observe customer satisfaction according to frequency use of public and private means. From sample data it results 81% of citizens uses the bus as public transport mean; 60% uses tube; instead 91% uses the car as private transport mean. Since 68% of sample uses conjointly public and private means, we report Table 2 with the estimated effect on the response probability according to the different combination of means use. We can note a negative effect on the satisfaction for people who take only private means very frequently, whereas a positive

effect for people who take both public means and private means sometimes a month.

Table 2. *Estimated effect by different use of transport means.*

Transport means	Weekly/daily private means	Public means + daily priv. means	Monthly private means	Public means	Public means + monthly private means
Effect	-0.26	-0.04	+0.19	+0.22	+0.41

6. Conclusions

The ordered logistic multilevel model is useful to analyse the dependence of customer satisfaction on a set of explanatory variables, in presence of a hierarchical structure of data. In this paper we applied the model to explain the citizens' opinion on public transport quality of Rome, taking into account the influence of residential context. As a consequence, it is correct to consider the needs of people living in a metropolitan environment like Rome, because in this way it is possible to find critical areas and uncomfortable situations for them to make interventions of quality control and quality improvement. We find that differences in demographic development, in territorial morphology and in transport system, among districts are specific relevant contextual aspects able to explain perceived quality about public transport.

References

Agenzia per il controllo e la qualità dei servizi pubblici locali del Comune di Roma (2003), *Relazione annuale sullo stato dei servizi pubblici locali e sull'attività svolta*, gennaio, Roma.

Bryk A.S., Raudenbush S.W. (1992), *Hierarchical Linear Models*: *Applications and Data Analysis Method*, Sage Publ., Newbury Park, CA.

Grilli L., Rampichini C. (2001), Modelli ordinali a componenti di varianza per la valutazione della didattica universitaria, in *Convegno intermedio della SIS, Processi e Metodi statistici di valutazione.* Roma 4-6 giugno, Univ. di Tor Vergata, 29-32.

Rabe-Hesketh S., Pickels A., Skrondal A. (2001), GLLAMM *Manual*, Technical Report 2001/01, Dep. of Biostatistics and Computing, Institute of Psychiatry, King's College, University of London.

Snijders T.A.B., Bosker R.B. (1999), *An introduction to basic and advanced multilevel modeling*, Sage Publ., London.