Structure of Mobility Phenomenon: Outcomes of an Exploratory Analysis with Techniques of Non-linear Multivariate Analysis

Pasquale Colonna[1]
Department of Highways and Transportation
Politecnico di Bari
colonna@poliba.it

Simona d'Amoja[2]
Department of Highways and Transportation
Politecnico di Bari
s.damoja@poliba.it

Achille Fonzone[3]
Department of Highways and Transportation
Politecnico di Bari
a.fonzone@poliba.it

[1] Colonna defined the outline of the research, coordinated the work and reviewed the paper.
[2] d'Amoja contributed to design and to draw up the questionnaire and coordinated the survey, taking part in the first stage of the work.
[3] Fonzone carried out the statistical analysis and the related elaborations and interpretations and wrote the paper.

Synopsis

The awareness of the negative effects of transport asks research some urgent questions about the conditions for a sustainable mobility. The answer to these questions is needed to plan and to carry out organizational and structural interventions on transport systems able to bring about a real improvement of life quality and then to actually contribute to development. The intrinsic complexity of the matter is made worse by globalization that widens the geographic and temporal boundaries of problems.

An effective compass to find the way in this intricate context seems to be the knowledge of the native human needs whose fulfilment is the minimal goal for every development policy. Recent studies seem to corroborate the idea, persistent in some secondary lines of research and confirmed by the common experience but neglected by the current scientific paradigm, that an intrinsic positive utility is associated to mobility or, in different words, that the demand for mobility has a non-derived component. According to these studies, mobility phenomenon would be governed by an unobserved Time Travel Budget, linked to such a positive utility. If the existence of this latent desired Travel Time Budget were confirmed, its introduction in planning models could permit a better effectiveness of sustainable mobility policies.

The report presents a pilot study, preliminary to a regional-level research on the latent desire for mobility. The study has involved a sample of 100 people who have filled in a questionnaire with 88 questions about several dimensions of the matter of mobility. The collected data have been analysed with multivariate analysis techniques based on the procedures of multidimensional scaling proposed by the Data Theory Scaling System Group, Leiden University.

The analysis provides a confirmation of the existence of the unobserved desired TTB, an initial rough estimate of which is 1 hour per day, and it shows up the usefulness of introducing factors linked to the non-derived dimension of transport demand in the models interpreting mobility phenomenon.

Structure of Mobility Phenomenon: Outcomes of an Exploratory Analysis with Techniques of Non-linear Multivariate Analysis

The awareness of the negative effects of transport asks research some urgent questions about the conditions for a sustainable mobility. The answer to these questions is needed to plan and to carry out organizational and structural interventions on transport systems able to bring about a real improvement of life quality and then to actually contribute to development (Colonna, 2003). The intrinsic complexity of the matter is made worse by globalization that widens geographic and temporal boundaries of problems.

An effective compass to find the way in this intricate context seems to be the knowledge of the native human needs whose fulfilment is the minimal goal for every development policy. The paradigm prevailing in transport and planning research regards mobility as a disutility or, in other words, a derived need, whose costs (in terms of money and time) are justified by the utility of taking part in other spatially separate activities (see, as just one example of a very large number of cases, Papacostas and Prevedouros, 1993). However, during the last three decades, several “dissenters” (Reichman, 1976; Jones, 1978; Hupkes, 1982; Marchetti, 1994, cited in Mokhtarian e al., 2001; Salomon and Mokhtarian, 1998; Mokhtarian and Salomon, 2001; Redmond and Mokhtarian, 2001) have highlighted that this axiom is partial, recognizing that a positive utility is associated to mobility. This utility could be the cause of an excess of mobility in comparison with the mobility strictly required to fulfil non-derived needs.

In particular, Mokhtarian and Salomon (2001) recognize three kinds of utility linked to travelling: the utility of the activities at destination, the utility of the activities carried out while travelling and the utility of travelling itself. To back their theory, the authors present the outcomes of a research interesting the San Francisco Bay Area, whose aim was to determine the relations between objective mobility, perceived mobility, relative desired mobility and travel liking and some social-demographic and personality characteristics of participants. Their final hypothesis is that mobility phenomenon is governed not by an observed Travel Time Budget (TTB, i.e. by the fact that the time each individual dedicates to travelling every day is constant over time and space. See, in favour, Zahavi and Talvittie, 1980; against, Goodwin, 1981), but by a latent desired TTB, depending on attitudes and demographic features of individuals and changing with purposes and modes of travelling. In other words, as also the common experience seems to suggest, the demand for mobility has a non-derived component.

The observation of the existence of this latent desire for mobility, the identification of the factors on which it depends and of the processes through which it changes into the actual consumption of mobility promise interesting developments towards the sustainability of transport: in fact, e. g., the policies of mobility reduction through the increase of transport costs could be more effective if applied in contexts in which the mobility desired by people is lower than the actual one; similarly, the adoption of means of transport slower but less environmentally dangerous could be favoured by a positive difference between desired and actual mobility.

The authors have begun a regional-level research aiming to measure this latent desire for mobility, the dynamics that generate it and the mechanisms that cause its transformation into consumed mobility. The paper reports the results of a pilot study, preliminary to drawing up the main questionnaire and to designing of the related sample, whose goal is sketching the outlines of the structure of mobility phenomenon and, in particular, assessing the utility of factors ascribable to the innate component of the mobility need in explaining the phenomenon itself.

The report is organized in this way: the section “Subject of the analysis” describes the questionnaire and the sample to which it has been submitted and it introduces the dimensions of mobility phenomenon analyzed in the following. “Multivariate analysis techniques applied” is an introduction to the statistical methods employed to extract information from the survey and it aims to help the reader in the interpretation of results. “Dynamics of mobility” presents the relations between the dimensions taken into consideration. In “Analysis of actual mobility” factors are studied influencing the daily consumed mobility, with the related hierarchy. The results are summarized in “Conclusion”.

SUBJECT OF THE ANALYSIS

Survey

Considering the exploratory character of the survey, a questionnaire has been arranged with a high number of questions (88), in order to register a wide range of psychological, cultural and social features and to explore travel behaviour, related to purpose and means, and attitudes towards mobility. A particular attention has been paid to verify the existence of the desired TTB. 80 out of the 88 questions have a closed format, with a number of options between 2 and 11; some questions accept multiple answers.

The questionnaire has been filled in by 100 arbitrarily chosen people in a self-administered way; the representativeness of the sample seems suitable for the explorative aim of the research and the full universality of the studied phenomenon. 51.0% of participants are women. Age ranges from 15 to 84 years, the average is 38.55 and the values are concentrated around two modes, 28 and 54. 28.0% of participants tick the “Student” option in the question about employment, 18.0% chose “Other”, 16.0% “White collar”; the less represented categories are “Manager” and “Retailer” (3.0%) and “Craftsman” (1.0%). Family income is “High” for 5.1% of the sample, “Upper-middle” for 24.2%, “Middle” for 58.6%, “Lower-middle” for 10.1% and “Low” for 2.0%. Figure 1 shows the age – sex pyramid of the sample, figure 2 the distribution of employments by class of income.

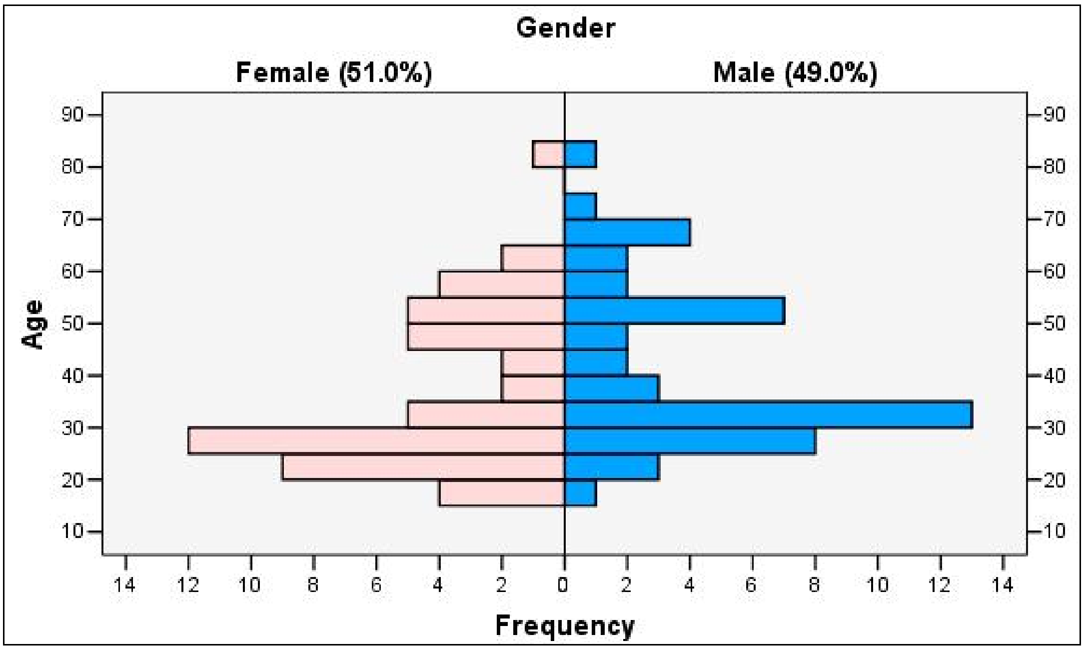

Figure 1: Age – Sex structure

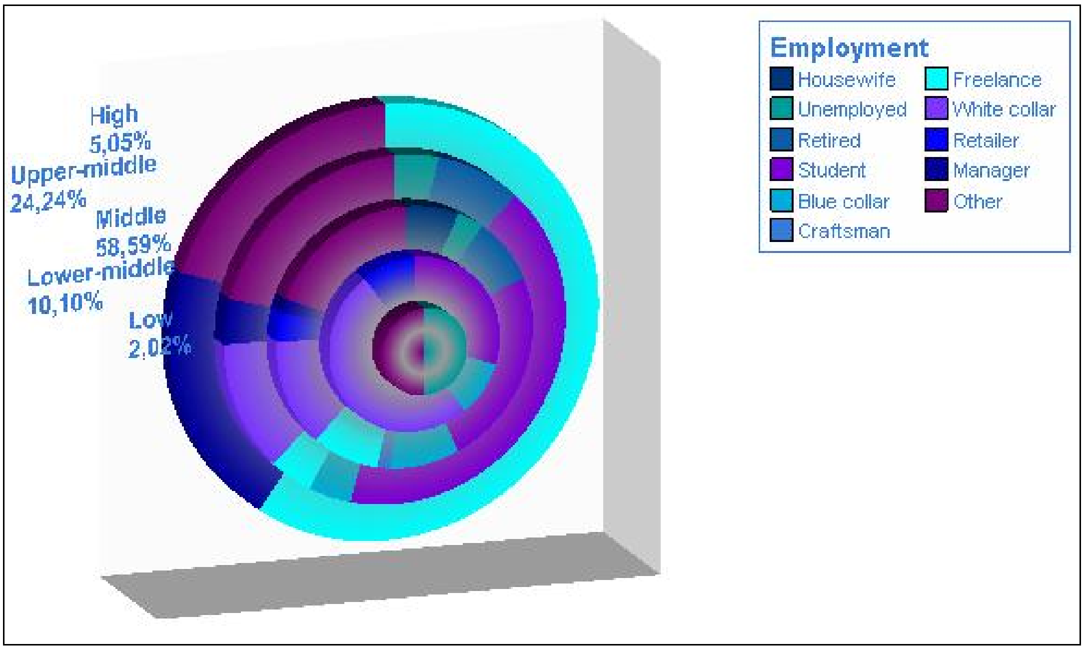

Figure 2: Kind of employment by family income

Questions analysed

In the paper, five aspects of mobility phenomenon re studied:

- Personal features of participants (9 variables), among which, in addition to those usually included in this kind of survey, 3 other possible factors of increment of the consumed mobility have been surveyed, i. e. optimism and so the expectation of a utility in the relationship with reality, desire for knowing the surrounding world (connected to newspaper reading) and multiplicity and vastness of interests (linked to internet use).
- Characteristics of residential locations, in particular the quality of public transport supply (2 variables).
- Pleasure in compulsory and non-compulsory travel (2 variables).
- Use of mobility, in terms of frequency and duration of some kinds of travel (4 variables).
- Desired mobility, in absolute terms as regards commuting, in relative terms as to total amount of travel (2 variables).

The questions chosen to describe these five dimensions are summed up in table 1. To simplify the analysis, the 11 original categories of the variable *Employment* have been gathered in 6 new attributes: Home-working (Housewife), Not employed (Unemployed, Retired), Student, Employed (Blue collar, White collar, Manager), Self-employed (Craftsman, Freelance, Retailer), Other.

Table 1 – Variables in the analysis

DIMENSION	VARIABLE	QUESTION	ORIGINAL CATEGORIES
PERSONAL FEATURES	*Gender*	What is your gender?	Female, Male
	Age	How old are you?	*Numeric variable*
	Marital Status	What is your marital status?	Single, Married
	Household	How many people are in your household?	*Numeric variable*
	Optimism	Are you optimist?	Yes, No
	Internet	Do you use internet?	Yes. No
	Newspaper	Do you read a newspaper?	Yes, No
	Employment	What is your employment?	Housewife, Unemployed, Retired, Student, Blue collar, Craftsman, Freelance, White collar, Retailer, Manager, Other
	Family Income	What annual income does your family have?	High, Upper-middle, Middle, Lower-middle, Low
TERRITORY CHARACTERISTICS	*Residential location*	In which kind of context do you live?	Metropolis, City, Town, Hamlet

DIMENSION	VARIABLE	QUESTION	ORIGINAL CATEGORIES
	Effective Means	Do you think that public transport is efficient in your place?	Yes, No
TRAVEL LIKING	*Employment Travel Liking*	How much do you like travelling for your employment?	Much, Fairly, A little, Not at all
	Recreation Travel Liking	How much do you like travelling for recreation?	Much, Fairly, A little, Not at all
ACTUAL MOBILITY	*Employment Travel Frequency*	How often do you travel for your employment?	Every day, At least once a week, At least once a month, Never
	Recreation Travel Frequency	How often do you travel for recreation?	Every day, At least once a week, At least once a month, Never
	Commuting Travel Time	How much time of your day do you spend for commuting?	Up to 30 minutes, 1 hour, 1 hour and 30 minutes, More
	Daily Travel Time	How much time do you travel every day to do what you do?	Less than 1 hour, A bit more than 1 hour, Between 1 and 2 hours, More than 2 hours
DESIRED MOBILITY	*Desired Commuting Travel Time*	What is the longest time you are willing to spend to reach your workplace?	Up to 30 minutes, 1 hour, More than 1 hour
	Relative Desired Travel Time	Would you like to spend a longer time for travelling?	Yes, No, I would like spend less time, Current time is ok

MULTIVARIATE ANALYSIS TECHNIQUES APPLIED

The exploratory statistical analysis of a phenomenon aims to obtain information and to build knowledge from surveying the values that some variables, supposed relevant for the phenomenon dynamic, take inside a sample extracted from the population subject of the study.
Therefore the study of complex phenomena, such as mobility, requires techniques to reduce dimensionality or, in different words, to reduce information redundancy inside data. In particular, to reach the goal of the present research three multivariate statistics methods have been employed: Principal Component Analysis (for the reduction of redundancy inside the data of personal features of participants), Non Linear Canonical Correlation Analysis (to recognize the relations between personal features, territory characteristics and variables concerning mobility) and Categorical Regression Analysis (to establish whether relationships between the use of mobility and the other considered dimensions exist or not). The first two techniques are developed on the idea that the variance of the observed variables can be explained by a small number of latent factors, to be pinpointed, which govern the phenomenon. The factors, real or virtual, are identified so as to minimize the loss of information inevitably bound up with the reduction of the dimensions describing the phenomenon. Regression analysis allows studying the importance of some predictors in explaining a target variable and evaluating the statistical significance of results (Fabbris, 1997).
In case of subjects concerning the psychological or social sphere of people, the quantities to be surveyed are very often of a qualitative nature and can be measured only at nominal (the categories of the variables represent non-ordered quantities) or ordinal (the categories have an intrinsic order) level. The possibility of applying the exploratory statistics techniques developed for numerically-scaled data to categorical variables is tied up to the assignment of numerical quantifications to the categories. In the present work, scaling procedures have been employed developed by the Data Theory Scaling System Group (DTSSG) of Leiden University, whose works are published under the nom de plume Albert Gifi. Through an iterative method known as Alternated Least Squares (ALS), these methods associate quantifications to categories which optimize the outcomes of the multivariate statistics techniques applied to study the data: e. g., in the case of Non Linear Canonical Correlation Analysis, such values are assigned to categories that maximize the correlations between the canonical variables (Gifi, 1990; Michailidis e De Leeuw, 1996).
In the following a short description of the applied techniques is provided, useful to evaluate results. In particular, a greater attention is paid to Non Linear Canonical Correlation Analysis, which is central for the description of mobility dynamics and, at the same time, it is the less common in the field of transport research (interesting instances, also for the interpretation of the obtainable outcomes, are Hensher and Golob, 1999 and Golob and Recker, 2004).

Principal Component Analysis (PCA)

The numerousness of variables in social statistics surveys makes the interpretation of relations between variables themselves in terms of simple and partial correlation coefficients very difficult: e. g., in the present case, to analyze the 19 variables chosen as representative, it would be necessary to work out and to interpret at least 171 simple correlation coefficients and 2907 coefficients of partial correlation with one control variable. Moreover the redundancies in data complicate calculations and interpretations. To face these problems, statistics research has developed a family of methods known as factorial analysis, to which PCA belongs.

A “component” is a linear combination of the observed variables. The PCA tries to reproduce the observed total (common and unique) variance through a set of principal (that is not correlated, orthogonal) components, the number of which is smaller than that of the variables themselves. Principal components are defined identifying the linear combination which extracts the maximum of variability from the variance-covariance matrix after that the variance explained by the already extracted dimensions has been eliminated. The process is usually stopped when the last extracted component does not give a relevant contribute to explaining the observed variance or in order to prevent the high number of components from making interpretation difficult. The meaning of components can be deduced analyzing the so called component loadings, that is to say the correlation coefficients between variables and components. The square of a component loading is the percentage of variance of the variable explained by the related factor.

Non Linear Canonical Correlation Analysis (NLCCA)

Pearson’s correlation coefficient, usually indicated with the letter r, measures the strength of the relationship between two numerical variables. r is the covariance of the standardized variables and, as a consequence, can take values from -1 to +1. It is positive when an increase of one variable corresponds to an increase in the other variable. The maximum of the range is reached when a linear relationship exists between variables. r^2, called coefficient of determination, is the percentage of the variance of one variable explained by the other.

Canonical Correlation Analysis (CCA) is a generalization of multiple regression analysis which allows evaluating many- to- many relationships. The first mention of the problem of regressions with more than one target variable can be found in Hotelling (1935), who proposes the replacement of a set of dependent variables by a single target variable, function of the observed dependent variables themselves, and, as a criterion, the maximization of the ratio of the latent (or canonical) variable variance explained by the set of the independent variables. Later Hotelling (1936) himself formulates again the problem in a more symmetric way and suggests studying the relations between two sets of variables after having removed the linear dependencies within each of the two sets. In this way each variable contributes to the extent that it provides independent information with respect to the others of its own set and it is linearly dependent from the variables of the other set. Practically, in order to asses the link between two sets of variables, the correlation between two canonical variables is worked out, each of which is a linear combination of the observed variables of one of the two sets. The weights through which the original variables contribute to the latent variable are defined so as to maximize the correlation between the latent variables themselves. Obviously it is possible to identify different couples of canonical variables for the same two sets of surveyed variables; each couple defines a dimension of the link between the two sets. In CCA component loadings are the correlation coefficients between the observed variables and the related canonical variables; also in CCA component loadings permit inferring the meaning of dimensions. CCA can be extended to cases with K > 2 sets making the definition of the canonical variables dependent on the maximization of a function of the correlation matrix *R,* of dimension K×K, which gathers the ½ K×(K-1) involved canonical correlations (Gittins, 1985).

The problem of the canonical correlation between groups including variables not measured on a numerical scale has been effectively solved by the already mentioned DTSSG with the ALS method and assuming the maximization of the largest eigenvalue of *R* as a criterion of optimization. This modus operandi is equivalent to maximizing the sum of the correlations between each canonical variable and a vector x of unknown coordinates (Gifi, 1990). Let N be the number of the observed instances (objects); J the number of the observed variables; I_j the number of the categories of the variable j (j = 1, …, J); J(1), …, J(K) a partition of the set of the indexes of the J variables; p the number of the dimensions to be analyzed. Set $G_j(i,t) = \begin{cases} 1 \text{ if the object i belongs to the category t} \\ 0 \text{ otherwise} \end{cases}$ $\left(i = 1,...,N; t = 1,...,I_j, j = 1,...,J \right)$ and called *X* the matrix of the object coordinates in the space R^p and Y_j the matrix containing the coordinates of the I_j categories of the variable j in the same R^p, the optimal representation of the N objects and of the $\sum_j I_j$ categories in a space with p dimensions is obtained by Gifi minimizing the loss function

$$\sigma(X;Y_1,...,Y_J)=K^{-1}\sum_{k=1}^{K}SSQ\left(X-\sum_{j\in J(k)}G_j\cdot Y_j\right)$$

(where SSQ is the sum of squares of the elements of the matrix), subject to

$$X'X=NI_p$$

$$u'X=0$$

(where I_p is the identity matrix of order p and u a vector with all 1). Imposing rank-one restrictions

$$Y_j=q_j\cdot\lambda_j'$$

on the multiple category quantifications (where q_j is a l_j-column vector of category quantifications and λ_j a p-column vector of weights, called component loadings), it is possible to obtain multidimensional solutions for the object scores with single quantifications of the categories of variables, also scaled at ordinal and numeric level. Gifi's algorithm, called OVERALS, is very effective in the cases where the goal of the analysis is finding out the structure of a set of qualitative data. To that end, some properties of the DTSSG solution are useful:

- The score of an object on a dimension is the average of the quantifications of the categories to which the object belongs. As a consequence, the distance between the points of R^p representing two objects depends on the similarity of the profiles of the points themselves.
- Atypical cases are far away from the origin of the space R^p, those with an average profile are more central.
- The point representing a category is the centroid of the objects belonging to the category.
- The points of the categories with high marginal frequency are close to the origin of the space R^p, whereas those of the categories to which few objects belong are more peripheral.

Categorical Regression Analysis (CRA)

Multiple regression analysis allows studying the link between a dependent variable and a set of independent variables or predictors. In particular, it allows finding out to what extent the variance of the dependent variable is explained by the predictors at a certain level of statistical significance and what is the weight of each independent variable in forecasting the dependent.
Usual regressive models are of a linear kind, i.e. the relation between the dependent variable y and the n predictors is of the kind

$$y=c+\sum_{i=1}^{n}b_i\cdot x_i$$

where c is a constant and b_i the regression coefficients which represent the variation of y when x_i increases by one unit, all other variables being the same. Weights of independent variables can be expressed also by the regression coefficients for the standardized data, β_i. β_i is the increase of y when x_i increases by one standard deviation and all the other variables are unchanged. Note that β_i's express only the unique contribute of each independent variable to the explained variance, therefore they can underestimate the importance of variables with high common and low unique contributions. The quality of a model is measured by the proportion R^2 of the dependent variable variance explained, commonly and uniquely, by the independent variables. Augmenting the number of predictors inevitably leads to a larger R^2, but this can be due to casual variations of the added predictors rather than to an actual improvement of the predictive power of the model, hence the goodness of fit of the regression has to be assessed considering also adjusted R^2, obtained reducing R^2 in such a way as to take the number of cases and of predictors into account (Garson, 2004).
Categorical variables can be included in regression analyses either through the introduction of dummy variables, and so the construction of one model for each combination of the categories of the categorical variables, or considering the attribute numeric coding as scaled at an interval level, with the problem that this coding is absolutely arbitrary. Instead Gifi's CRA assigns every attribute such a numeric value that maximizes the squared correlation between the transformed dependent variable and the linear combination of the predictors, transformed too. This method can take non-linear relationships into account because it is possible to transform the variables at nominal, ordinal or numeric level; moreover, non-linear transformations reduce the dependence between predictors.

Software and hardware

Data have been analyzed with the package SPSS version 13.0; in particular, the analysis of the categorical data has been carried out with the module Categories 13.0 which implements the techniques of multidimensional scaling of the DTSSG: the version 1.1 of the algorithm CATPCA for PCA, the version 1.0 of

OVERALS for NLCCA and the version 2.1 of the algorithm CATREG for CRA. The software has run on a pc with an Intel ® Pentium ® 4 / 3 GHz processor and a 512 Mb RAM.

DYNAMICS OF MOBILITY

Redundancy reduction in the data of personal features

Before studying the structural relations of mobility phenomenon, it has been reckoned advisable to carry out a PCA of the 9 variables chosen to represent personal features of participants (tab. 1), in order to evaluate the existence and, in case, to remove information redundancies and so to simplify the interpretation of the NLCCA and of the CRA.
The scree test has been adopted as a criterion to estimate the minimum number of dimensions needed to represent the data set effectively. The curve of the eigenvalues of the analysis with 9 dimensions (fig. 3) shows an elbow around the third eigenvalue. This proves the usefulness of the factorial analysis to reduce the dimensionality of the studied data set and it indicates that it is necessary to use 2 dimensions to represent the information which can be extracted from the sample properly (Harman, 1976).

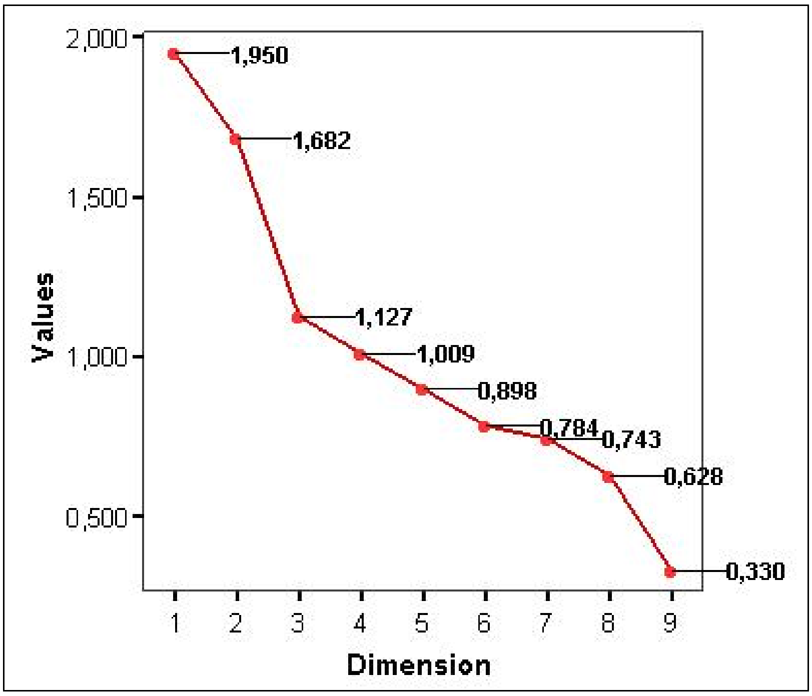

Figure 3: Scree plot

The 2 dimensions, identified with the variables scaled and grouped as shown in table 2 (in particular, note that the optimization of the analysis results leads to re-codifying the continuous variable *Age* in four categories and reducing the attributes of *Family Income* from the original five to three. In the following, reference is always made to the re-codified version of these variables), explain 50.2% of the total variance.

Table 2: Summary of the PCA

VARIABLE	SCALING LEVEL	CATEGORIES [QUANTIFICATION]*	EXPLAINED VARIANCE		
			DIM 1	DIM 2	TOT
Gender	Nominal	Female [-1.001], Male [0.999]	0.041	0.055	0.096
Age	Ordinal	15-23 [-1.538], 24-37 [-0.606], 39-53 [-0.993], 54-84 [-1.225]	0.863	0.006	0.869
Marital Status	Nominal	Single [-0.890], Married [-1.124]	0.819	0.003	0.822
Household	Ordinal	1-3 [-0.874], 4 [-0.447], 5-7 [1.531]	0.013	0.516	0.529
Optimism	Nominal	Yes [-0.346]. No [2.852]	0.004	0.294	0.298
Internet	Nominal	Yes [-0.611], No [-1.642]	0.566	0.070	0.636
Newspaper	Nominal	Yes [-0.764], No [1.325]	0.027	0.392	0.419

VARIABLE	SCALING LEVEL	CATEGORIES [QUANTIFICATION]*	EXPLAINED VARIANCE		
			DIM 1	DIM 2	TOT
Employment	Nominal	Home-working [1.297], Not employed [1.163], Student [-1.544], Employed [0.551], Self-employed [-0.212], Other [0.417]	0.606	0.094	0.700
Family Income	Ordinal	High – Upper-middle [-0.717], Middle [-0.186], Lower-middle – Low [2.631]	0.020	0.128	0.147
TOTAL			2.957	1.559	4.516

*Notice that, here and in table 3, equal quantifications of different categories of the same variable denote that the categories have been gathered in a single attribute

The graph of the component loadings (fig. 4) points out three groups of variables: the first one encompasses the variables *Internet, Marital Status*, *Age* and *Employment*, for which the vector component along the first dimension, positive in all the cases, is preponderant in comparison with that along the second; the second one includes *Household*, *Newspaper* and *Optimism* whose vectors are arranged mainly along the second component; the third one is made up of *Family Income* and *Gender* for which it is not possible to identify a prevailing direction and which play a secondary role in the explanation of the variance, as it is evident from the limited length of the corresponding vectors. The last result can be explained in different ways: first of all, it could be due to the loss of information intrinsic to the reduction of dimensionality of the data set, i. e., in other words, the relevance of the variables *Family Income* and *Gender*, which do not fit well in a 2 dimension solution, could increase in solutions with more dimensions. Moreover, it seems reasonable to hypothesize that the variable *Gender* is not discriminating, i. e. the other characteristics of the sample do not change when the gender changes. Such an interpretation does not appear adequate to explain the small importance of the variable *Family Income*, the effect of which could be partially captured by other variables, such as *Employment* and *Household*.

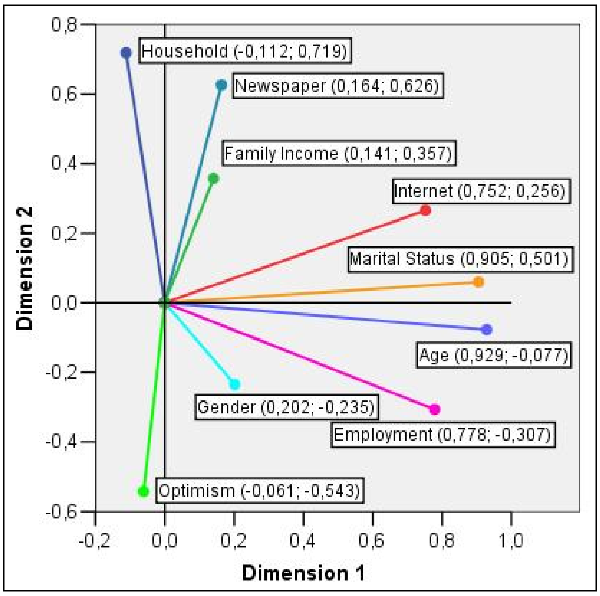

Figure 4: PCA component loadings

In the present work, PCA has been used not to characterize the dimensions discriminating the personal features of participants, but only with the purpose of reducing the number of variables in the following analyses; to that end, figure 4 suggests employing the variables *Age* and *Household* since they are the ones that saturate most respectively the first and the second dimension. Looking at the joint plot of category points (fig. 5) as for the relevant variables, it can be seen that *Age* categories “39-53” and “54-84” have positive scores on the first dimension and in this they are linked, as expected, to those who do not use internet, to married people and to those who are not student. In particular, as regards the variable *Employment*, scaled at nominal level, notice that the category “Student” is correctly associated to the category “15-23”, the categories “Not employed” and “Home-working” are mainly linked to the eldest brackets of participants, whereas working people (belonging to the categories “Self-employed”, “Employed” and “Other”) tend to have an average age, included between the superior range of the category “24-37” and the category “39-53”. On the second dimension, people belonging to large households (“5-7”) have a “Lower-middle – Middle” family income, are not inclined to read newspaper and are more optimist than those belonging to small households (“1-3”).

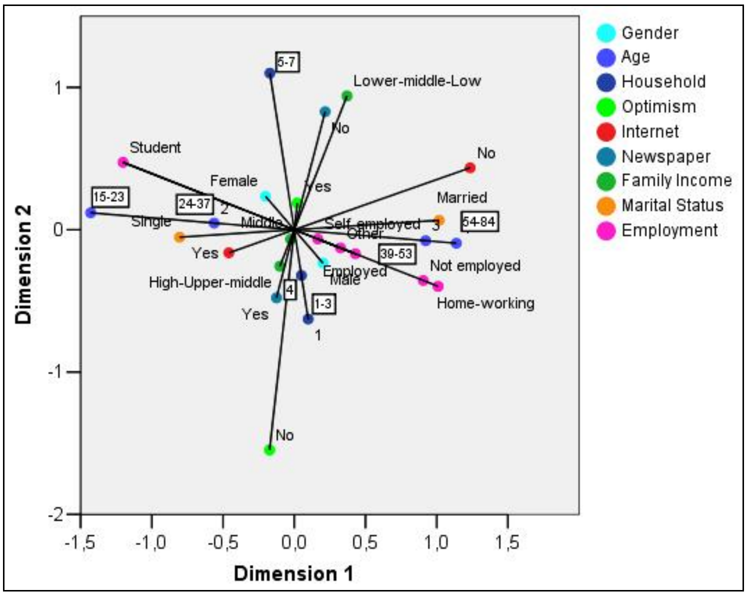

Figure 5: PCA joint plot of category points

Relations of mobility variables

12 variables have been considered in the NLCAA, as summarized in table 3. Since the aim of the analysis is just to detect the dynamics underpinning the tendency of the variables concerning TRAVEL LIKING, ACTUAL MOBILITY and DESIRED MOBILITY to increase or to decrease, it has deemed right to scale these variables on an ordinal level, i. e. so that the numeric quantifications assigned to categories by the algorithm OVERALS keep the original arrangement of categories. In the case of *Age*, *Household* and *Residential Location* it has been decided to allow free clusters of categories and so they are scaled like nominal variables, i. e. allowing OVERALS to assign quantifications whose order is not forced to respect that of original categories. For *Effective Means*, scaling level is not relevant because it has only two attributes.

Table 3: Summary of the NLCCA

DIMENSION	VARIABLE	SCALING LEVEL	CATEGORIES [QUANTIFICATION]*	SINGLE FIT	COMPONENT LOADINGS DIM 1	DIM 2	VECTOR LENGTH
PERSONAL FEATURES	*Age*	Nom.	15-23 [1.506], 24-37 [-0.138], 39-53 [-1.533], 54-84 [0.765]	0.578	0.636	-0.319	0.712
	Household	Nom.	1-3 [-1.439], 4 [0.289], 5-7 [1.027]	0.830	-0.615	-0.614	0.869
TERRITORY CHARACTERISTCS	*Residential Location*	Nom.	Metropolis [-1.629], City [-0.719], Town [0.438], Hamlet [2.236]	0.737	-0.770	-0.030	0.771
	Effective Means	Nom.	Yes [-1.465], No [0.705]	0.279	-0.230	-0.320	0.394
TRAVEL LIKING	*Employment Travel Liking*	Ord.	Much [-1.625], Fairly [-1.625], A little [0.691], Not at all [0.691]	0.137	-0.136	0.242	0.278
	Recreation Travel Liking	Ord.	Much [-0.31], Fairly [-0.31], A little [4.081], Not at all [no occurrence]	0.392	-0.565	0.037	0.566
ACTUAL MOBILITY	*Employment Travel Frequency*	Ord.	Every day [-1.290], At least once a week [0.856], At least once a month [0.856], Never [1.234]	0.072	0.240	-0.410	0.475
	Recreation Travel Frequency	Ord.	Every day [-1.297], At least once a week [-1.297], At least once a month [0.207], Never [1.880]	0.077	0.250	-0.066	0.259
	Commuting Travel Time	Ord.	Up to 30 minutes [-0.936], 1 hour [1.547], 1 hour and 30 minutes [1.547], More [1.547]	0.236	-0.172	0.646	0.669
	Daily Travel Time	Ord.	Less than 1 hour [-0.983], A bit more than 1 hour [1.306], Between 1 and 2 hours [1.306], More than 2 hours [1.306]	0.254	0.061	0.635	0.638
DESIRED MOBILITY	*Desired Commuting Travel Time*	Ord.	Up to 30 minutes [-0.744], 1 hour [1.419], More than 1 hour [1.419]	0.324	-0.137	0.550	0.567
	Relative Desired Travel Time	Ord.	No, I would like o spend less time [-.905], Current time is ok [1.206], Yes [1.206]	0.339	0.188	-0.549	0.580

Model fit to data is 0.82 out of a theoretically obtainable maximum of 2; the first extracted dimension explains 54.4% of this fit. Table 3 reports also the values of the single fit that is the variance of category coordinates. It can be seen that the variables which discriminate most the objects are, in descending order, *Household*, *Residential Location* and *Age*, that is participants are more homogeneous with respect to the attitudes concerning mobility than as regards the social, demographic and psychological features. This suggests that phenomena exist which homogenize involved people from the point of view of mobility: it would seem to recognize a confirmation of the geographic-temporal approach of Hägerstrand (1970) who emphasizes the crucial influence of space-time limitations on human activities “against” the factors linked to motivations and choices (which are the basis of the approach of Chapin, 1974). Looking closely at the group of the variables

concerning travelling, one can note that *Recreation Travel Liking* is more discriminating than *Employment Travel Liking*; this is due to the rarity of those who do not like to travel for recreation; they tend to form a separate group, of the "sedentary" persons, and, being, in a certain sense, outliers, are worthy of further investigations. Duration of daily journeys (*Commuting Travel Time* and *Daily Travel Time*) distinguishes clearly more than frequency of journeys (*Employment Travel Frequency* and *Recreation Travel Frequency*); the very low resolving power of the frequency variables appears to suggest that the studied factors are not able to interpret the travel rate on a scale greater than the daily one. Finally, one can note that the profiles of participants with different answers to the two questions about the Desired mobility (*Desired Commuting Travel Time* and *Relative Desired Travel Time*) are quite different and, therefore, that this kind of desired (daily) mobility is actually related to the variables introduced into the model.
A hierarchical cluster analysis of the lengths of the vectors of component loadings (tab. 3) has suggested the existence of two classes of variables, that can be interpreted in terms of significance: *Effective Means*, *Employment Travel Liking*, *Employment Travel Frequency* and *Recreation Travel Frequency* (length ≤ 0.475) have a limited weight in the model, i. e. their contribution to the explanation of the mobility dynamics is not very important in this analysis and it is ignored in the following. From the graph of the loading component vectors (fig. 6) three groups of relevant variables emerge: that made up of Travel liking and Territory characteristics (notice that OVERALS quantifies *Residential Location* like an ordinal variable, assigning the highest score to "Metropolis") whose vectors are quasi-parallel to the first dimension axis; the second one grouping the variables concerning Actual mobility and Desired mobility, for which the loadings on the second dimension are greater, in modulus, than those on the first dimension; finally, the third set representing the Personal features, which are discussed separately because quantified as nominal. Analysing component loadings it is possible to recognize affinities among the variables: in fact, perpendicular vectors characterize statistically independent (barely correlated) variables, whereas parallel vectors with similar length indicate similar behaviours of variables. Bearing these considerations in mind, some interesting indications on the structure of mobility phenomenon can be deduced.

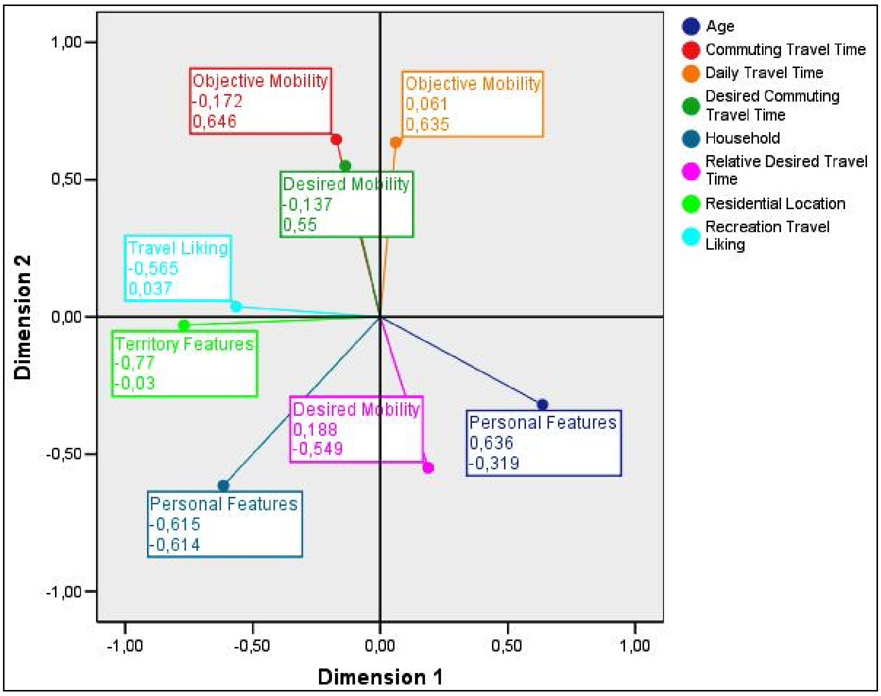

Figure 6: NLCCA component loadings

Relations between variables concerning mobility
The two variables of Actual mobility, *Commuting Travel Time* and *Daily Travel Time*, are correlated, as expected; this points to the crucial role played by the derived mobility linked to employment in determining personal consumption of mobility. *Desired Commuting Travel Time* is perfectly in line with *Commuting Travel Time*: at first glance, that could be interpreted as an evidence of the influence of personal inclinations to mobility on localization choices, but, given that the residential localization is generally chosen by households and not by individuals, it is seems more reasonable to think that habit determines a sort of "addiction" and in this way it influences desire. The other variable of the second group, *Relative Desired Travel Time*, is inversely correlated with the first three. To be more precise, those who dedicate up to 30 minutes to commuting and who travel less than 1 hour per day in all tend to be satisfied with their mobility or to desire a greater one, whereas the others are inclined to wish a smaller mobility: it would seem that the desired daily

mobility is around 1 hour per day (notice that this threshold is similar to the observed TTB frequently surveyed in literature, for instance in Zahavi and Talvittie, 1980. However, the extremely qualitative character of the variable *Relative Desired Travel Time* does not allow drawing significant deductions from this coincidence). The inverse correlation between *Desired Commuting Travel Time* and *Relative Desired Travel Time*, i. e. the fact that people less inclined to devote time to commuting are the ones who tend to wish a greater mobility, seems to stress that journeys with different purposes could be not substitutable. Vectors show up that the variables of the second group are almost independent from *Recreation Travel Liking*, outcome explained by the fact that OVERALS distinguishes only two attributes of people as regards pleasure in recreation travel, "sedentary" and "non-sedentary", and that the former is distinctive of nearly all participants, apart from any other characteristic.

Relations between variables concerning mobility and personal and territory characteristics

The variables representing the PERSONAL FEATURES, *Age* and *Household*, have been scaled at nominal level. Figure 7 seems to suggest that the analysis distinguishes people belonging to the most "active" groups of population (in fact, people from 24 to 53 years old can be considered active, also from the point of view of employment) from the least "active" ones (students, also university students, and elderly), arranging the related centroids along a line from the 4th to the 2nd quadrant.

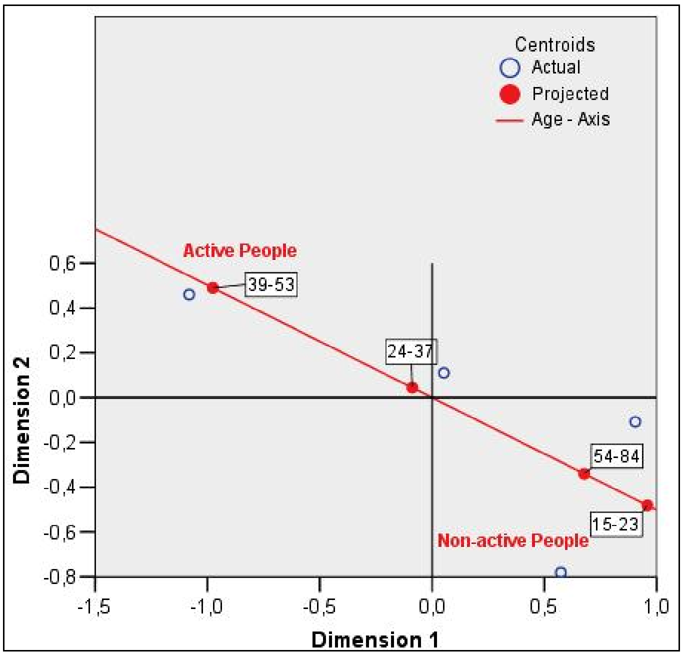

Figure 7: Age centroids

As to households (fig. 8), iterative implementation of Gifi's algorithm discriminates three homogeneous clusters, whose centroids place themselves along the bisecting line of the 1st and the 3rd quadrant according to an ordinal arrangement from the smallest households (1-3 people), upper right quadrant, to the most numerous ones (5-7), lower left quadrant.

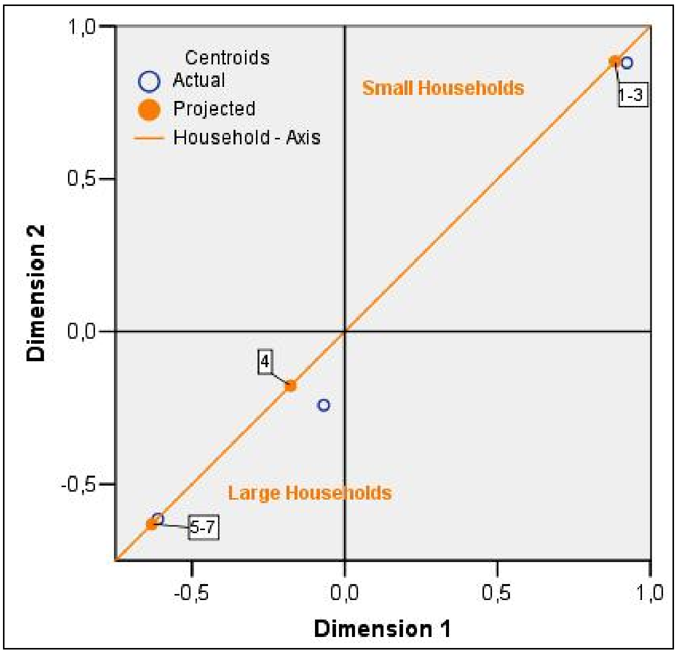

Figure 8: Household centroids

Figure 9 shows that the highest levels of actual mobility, and then also the desire for travelling less, are linked to people belonging to small households and active segments of population, while students and elderly people being members of numerous households travel less and are satisfied with the time spent for mobility.

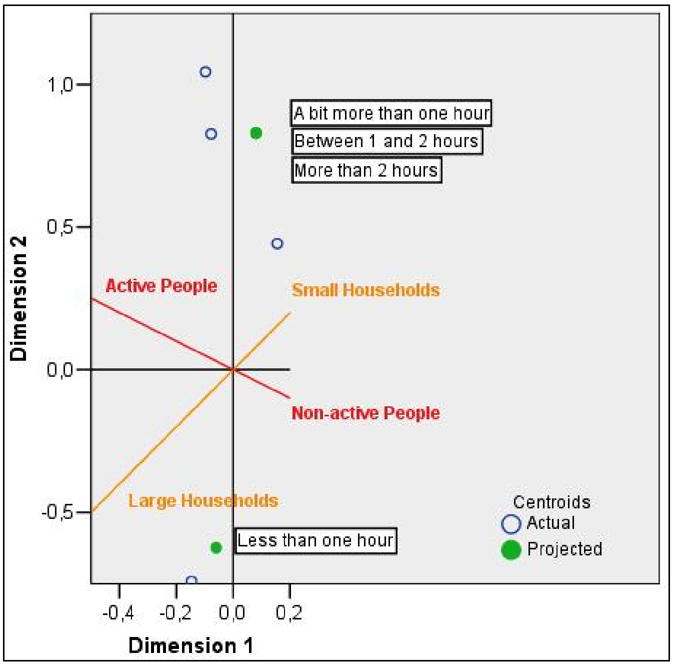

Figure 9: Daily Travel Time centroids

The correlation between territory characteristics of residential location and actual and desired mobility appears to be small. In the case of the ACTUAL MOBILITY, the outcome is rather singular (e. g. see the results of the CRA in the following section) and it could be probably ascribed to the exiguity of the sample and to the excessive distance between the attributes of the questions about mobility, which prevents the algorithm from discriminating participants effectively with respect to this characteristic: e. g., in the case of *Daily Travel Time*, OVERALS assembles all those responding “A bit more than 1 hour”, “Between 1 and 2 hours” and “More than 2 hours” in a single group. The outcome for *Relative Desired Travel Time* could be interpreted as another indication of the existence of the desired mobility, peculiar of the person and not depending on the need for travelling to carry out primary activities, which, instead, is linked to the territory context. Finally, a close association is present between *Residential Location* and *Recreation Travel Liking*, in particular between who does not like to travel for fun and who lives in a hamlet. This would seem to point to a sort of need for “escaping” of people living in chaotic environments; however the validity of this statement is heavily limited by the exiguity of the number of “sedentary” persons.

ANALYSIS OF ACTUAL MOBILITY

It has been believed useful to carry out a regression analysis of the variable *Daily Travel Time* to look for confirmations or refutations of the importance of dimensions such as TRAVEL LIKING and DESIRED MOBILITY in determining the real mobility.
The predictors introduced in the model (summarized in tab. 4) are the same of the NLCCA in order to allow a comparison with the outcomes inferred in the previous section. Given the nature of the variables taken into consideration, an ordinal scale has been used for all variables so as to obtain an eloquent interpretation of results. 70 cases have been analyzed. Transformed predictors explain 52.7% of the variance of the transformed dependent variable (adjusted R^2 is 0.347); the test of significance of R^2 proves that the model is significantly different from that with all predictor coefficients equal to 0.
The tolerance reported in the table is the percentage of the variance of the considered variable not explained by the other independent variables; variables with tolerance near to zero do not improve very much model predictive power and cause instability of results. In the present case, stability of results appears to be sufficiently assured by the upper-middle values of tolerances. The last remark together with the observation that all the values of Pratt’s index (described in a deeper way in the following) are positive is a sign of the absence of multicollinearity between predictors. Adopting a confidence level of 10%, it is statistically proved the influence of 6 variables out of 11 (highlighted in the table). Combining the signs of the standardized coefficients with the quantifications of the attributes of variables, one can observe a positive relation between daily time dedicated to travel and frequency of employment travel (*Employment Travel Frequency*), time spent for commuting (*Commuting Travel Time*), territory importance of residential location (*Residential Location*), class of age (*Age*) and pleasure in travelling for recreation (*Recreation Travel Liking*), whereas *Daily Travel Time* decreases when the desire of travelling more (*Relative Desired Travel Time*) increases. All the relationships have the expected sign.

Table 4 – Summary of the CRA

	LEAST QUANTIFICATION CATEGORY	TOLERANCE	SIGNIFICANCE	β	PRATT’S INDEX
Daily Travel Time	Less than 1 hour	*Dependent Variable*			
Employment Travel Frequency	Every day	0.659	0.006	-0.280	0.194
Relative Desired Travel Time	No	0.784	0.022	-0.262	0.189
Commuting Travel Time	Up to 30 minutes	0.650	0.074	0.197	0.172
Residential Location	Metropolis	0.777	0.001	-0.299	0.169
Age	15-23	0.684	0.005	0.310	0.129
Recreation Travel Liking	Much	0.804	0.009	-0.243	0.054
Desired Commuting Travel Time	Up to 30 minutes	0.634	0.630	0.059	0.035
Recreation Travel Frequency	Every day	0.767	0.374	-0.100	0.023
Household	1-3	0.839	0.698	-0.067	0.016
Effective Means	Yes	0.762	0.200	-0.143	0.011
Employment Travel Liking	Much	0.821	0.276	0.125	0.009

To assess predictor importance in explaining the target variable, Pratt’s index has been used which measures the ratio of total variance R^2 due to each predictor (Pratt, 1987). Moreover, Pratt’s index provides also information about the kind of mechanism through which a predictor influences the dependent variable: it can be demonstrated that suppressor variables, characterized by a very small Pratt’s index and a coefficient similar to those of the most important variables, have an indirect influence on the predicted variable, i. e. they have an effect on the dependent variable through the influence on the other independent variables (Thomas and al., 1998). To evaluate the importance of such suppressors, Thomas and al. (1998) suggest working out the ratio $\frac{R^2 - R^2_{NS}}{R^2}$, where R^2 is the variance explained by the whole model and R^2_{NS} that explained by the model with only the non-suppressor variables.

Table 4 reports independent variables in order of importance (Pratt's index) in the original model. Making reference to the 6 ones for which the results are statistically significant, one can notice that one belongs to the set of PERSONAL FEATURES, one to TERRITORY CHARACTERISTICS, two to ACTUAL MOBILITY and one each to the dimensions whose influence was to evaluate, DESIRED MOBILITY (*Relative Desired Travel Time*, the second for importance) and TRAVEL LIKING (*Recreation Travel Liking*, the sixth one). *Recreation Travel Liking* has the characteristics of a suppressor variable: R^2_{NS}, worked out with all (significant and non-significant) the variables, is 0.476, from which an importance index for the predictor is drawn of 0.097; this confirms the level of relative importance of the suppressor variable.

CONCLUSION

The necessity that the fulfilment of transport demand contributes to sustainable development urges that mobility phenomenon should be studied more in depth. To that end, the hypothesis of the existence of a latent desired TTB linked to an innate need for mobility appears to deserve a special attention.

The report presents the results of the preliminary stage of a research on such a non-observed desire for mobility. The survey has involved a sample of 100 people, who have filled in a questionnaire of 88 questions. The answers allow deducing interesting indications about the structure of mobility phenomenon and, in particular, on the contribution of factors linked to the non-derived component of mobility demand to the actually consumed mobility.

5 mobility-linked dimensions have been introduced in the NLCCA, described by 19 variables: demographic, social and personality features of respondents (represented by two variables, *Age* and *Household*, chosen among the 9 ones initially considered relevant through a PCA); territory characteristics of residential locations; travel liking; actual and desired mobility. The analysis highlights that

- Phenomena exist which homogenize people from the point of view of mobility. That can be interpreted as a confirmation of the correctness of some intuitions of Hägerstrand's space-time approach.
- It is reasonable to admit the existence of a desired mobility, the value of which is around 1 hour per day. Such a value is similar to the observed TTB found in many studies.
- The highest levels of actual mobility, associated to the desire for travelling less, are basically typical of persons belonging to active segments of population and to small households.

A regression analysis of the variable *Daily Travel Time*, a measure of ACTUAL MOBILITY, confirms the usefulness of considering DESIRED MOBILITY and TRAVEL LIKING, aspects linked to an innate need for travelling, in the explanation of the behaviour concerning mobility.

ACKNOWLEDGEMENTS

Authors wish to thank engineer Margherita Moretti for the contribution she provided in carrying out the interviews while doing her dissertation.

REFERENCES

CHAPIN, F. S. (1974), *Human Activity Patterns in the City: Things People Do in Time and Space*, Wiley, London, UK.

COLONNA, P. (2003), "Mobility, Roads, Development and Quality of Life", in *Proceeding of XXII PIARC Road World Congress*, Durban, South Africa.

FABBRIS, L. (1997), *Statistica multivariata – analisi esplorativa dei dati*, McGraw-Hill, Milano, Italy.

GARSON, G. D. (2004), *PA 765 Statnotes: An Online Textbook*, Available from: http://www2.chass.ncsu.edu/garson/pa765/statnote.htm.

GIFI, A. (1990), *Nonlinear Multivariate Analysis*, Wiley, Chichester, UK.

GITTINS, R. (1985), *Canonical Analysis; a Review with Applications in Ecology*, Springer-Verlag, Berlin, Germany.

GOLOB, T. F., RECKER, W. W. (2004), "A method for relating type of crash to traffic flow characteristics on urban freeways", *Transportation Research Part A*, 38, pp. 53-80.

GOODWIN, P. B. (1981), "The usefulness of travel budgets", *Transportation Research A*, 15, pp. 97-106.

HÄGERSTRAND, T. (1970), "What about people in regional science?", *Papers of the Regional Science Association*, 24, pp. 7-21.

HARMAN, H. H. (1976), *Modern Factor Analysis*, University of Chicago Press, Chicago, IL, USA.
HENSHER, D. A., GOLOB, T. F. (1999), "Searching for Policy Priorities in the Formulation of a Freight Transport Strategy: An Analysis of Freight Industry Attitudes", *Transportation Research Part F*, 35, pp. 241-267.
HOTELLING, H. (1935), "The most Predictable Criterion", *Journal of Educational Psychology*, 26, pp. 139-142.
HOTELLING, H. (1936), "Relations between two Sets of Variables", *Biometrika*, 28, pp. 321-377.
HUPKES, G. (1982), "The law of constant travel time and trip-rates", *Futures*, 14, pp. 38-46.
JONES, P. M. (1978), "Destination choice and travel attributes" in HENSHER, D., DALVI, Q, (eds.), *Determinants of Travel Choice*, Praeger, New York, NY, USA, pp. 266-311.
MARCHETTI, C. (1994), "Anthropological invariants in travel behaviour", *Technological Forecasting and Social Change*, 47, pp. 75-88.
MEULMAN, J. J., HEISER, W. J. (2004), *SPSS Categories ® 13,0*, SPSS Inc, Chicago, IL, USA.
MICHAILIDIS, G., DE LEEUW, J. (1996), "The Gifi System of Descriptive Multivariate Analysis", *UCLA Statistical Series Preprints*, 204.
MOKHTARIAN, P. L., SALOMON, I. (2001), "How derived is the demand for travel? Some conceptual and measurement considerations", *Transportation Research A*, 35, 8, pp. 695-719.
MOKHTARIAN, P. L., SALOMON, I., REDMOND, S. R. (2001), "Understanding the Demand for Travel: It's Not Purely 'Derived'", *Innovation*, 14, 4, 2001, pp. 355-380.
PAPACOSTAS, C. S., PREVEDOUROS, P. D. (1993), *Transportation Engineering and Planning, 2nd ed,*, Prentice Hall, Englewood Cliffs, NJ, USA.
PRATT, J. W. (1987), "Dividing the indivisible: Using simple symmetry to partition variance explained", in PUKKILA, T., PUNTANEN, S. (eds.), *Proceedings of the Second International Conference in Statistics*, University of Tampere, Tampere, Finland, pp. 245-260.
REICHMAN, S. (1976), "Travel adjustments and life styles – A behavioural approach", in STOPHER, P. R., MEYBURG, A. H. (eds.), *Behavioral Travel-demand Models*, D, C, Heath, Lexington, MA, USA.
REDMOND, L. S., MOKHTARIAN, P. L. (2001), "The positive utility of the commute: modelling ideal commute time and relative desired commute amount", *Transportation*, 28, 2, pp. 179-205.
SALOMON, I., MOKHTARIAN, P. L. (1998), "What happens when mobility-inclined market segments face accessibility-enhancing policies?", *Transportation Research D*, 3, 3, pp. 129-140.
THOMAS, D. R., HUGHES, E., ZUMBO, B. D. (1998), "On variable importance in linear regression", *Social Indicators Research*, 45, pp. 253-275.
ZAHAVI, Y., TALVITIE, A. (1980), "Regularities in Travel Time and Money Expenditures", *Transportation Research Record*, 750, pp. 13-19.