PERCEPTION OF ROAD SAFETY AMONG YOUNG PEOPLE: THE PSSG PROJECT[1]

Pietro Iaquinta, Michelangelo Misuraca

1. A brief international and Italian frame about road safety

Road safety – and in general the culture of security – represents currently a key distinctive feature of socially developed countries.

European Programs 2007-2013 primarily focused on this very important aspect of social living, by including road safety in their agendas as a main target of major impact for the European global society. The COM(2001)-370 directive (*White Paper. European transport policy for 2010: time to decide*) implicitly assumed a duty for the whole Community of reducing the number of road victims. Despite these commitments, the main target of decreasing of 50% the number of fatalities by 2010 has in fact not been achieved, because of the divergences among the different EU members. Nevertheless, the reduction of road traffic mortality in the European countries has been confirmed as one of the priorities also in the programming plans for 2014-2020. The aim is to effectively cut down the average of 40.000 deaths that yearly stain with blood the roads of the whole Europe. The COM(2010)-389 document (*Towards a European road safety area: policy orientations on road safety 2011-2020*) already proposed to halve the number of victims in the EU28 area before 2020, trying to reach the goal of 15.000 victims, with a "zero deaths" target. The UN also decided to support this ambitious program and on May 11th 2011 launched its own 2011-2020 road safety initiative.

At a national level Italy traditionally paid attention to road safety, and promoted a *National Plan for Road Safety* (PNSS) in 1999. The PNSS was a well-defined program that included specific actions and intervention strategies, focusing in particular on the dissemination of road safety culture. With regard to the 2001 EU White Paper's target, Italy reduced the number of deaths by 42,4%. The reduction of overall number of victims – from 7.000 in 2001 to just 3.385 in 2013 – was in line with the EU27 average.

[1] Il presente articolo, pur essendo frutto del lavoro congiunto degli autori, è principalmente attribuibile a Pietro Iaquinta per i paragrafi 1, 2 e 3, e a Michelangelo Misuraca per i paragrafi 4, 5 e 6

Two important aspects, probably not sufficiently highlighted during the years in the several public awareness campaigns about safety, are the devastating effects on accidents' survivors and the impact on public opinion. Injuries are still a problem with massive social consequences. In 1972, the bleakest year for road accident in Italy, an overall number of 268.000 injured was recorded in addition to 11.000 deaths. As recently as 2013 despite a decrease in the number of deaths to 3.385 the number of accidents was 257.421 (ISTAT, 2014). The 2010 EU White Paper discussed the morbidity of road accidents in the EU and proposed to distinguish between *active* and *passive* safety on vehicles.

The data reported above mean around 9 deaths and more than 700 casualties per day, unacceptable figures from both an ethical and economical point of view. This is even more relevant considering that a substantial portion of people involved in accidents has between 20 and 44 years of age. Moreover, road accidents are the first cause of death in Italy in the class of age 15-24. The untimely death of so many young people touches a community not only at a family level but also affects the generational replacement. This is a real concern for countries such as Italy that are already suffering a progressive population aging.

2. Morbidity and mortality of road accidents: a wrong perception of safety?

A key issue about road safety is the erroneous opinions sometime induced by awareness campaigns. These should be in fact promoting the culture of security and preventing fatalities. However, the public perception – especially among youngsters – associates road accidents and the related danger of death only with respect to specific moments. For example, Saturday night is considered particularly dangerous, because of higher probability of assuming huge quantity of alcohol or drugs. In reality, this is not the only circumstance in which accident can occur. Other factors may also affect the abilities of drivers, such as for example the external context (road type and condition) and the actual performances of the vehicle. The Italian National Statistics Institute reported that accidents with casualties have a higher probability to take place on Tuesdays and Wednesdays, when the estimated volumes of traffic are over the average, with a pick of cases between 6 pm and 8 pm of Monday.

As stated above, the high percentage of young people involved into road accidents is due to a "juvenile unconsciousness". A lack of driving experiences can also cause a dramatic decrease in safety conditions. The yearly debate about the structure of the examination required for obtaining a driving license is fuelled by opposite sides. The main controversy is on the need, or not, to adapt the tests (and/or, in any case, the learning step) to driving standards more appropriate to the

current traffic conditions. These are very different from those of the immediate post-war period when the "modern" driving tests have been designed.

Being able to park a car perfectly in line, in reverse, or without bumping the rear wheel on the sidewalk, is certainly an interesting indication of knowledge and mastery of the vehicle. However, the exam does not verify how the driver would manage driving with potential risk factors, such as bad weather or at night time.

Furthermore, road accidents have also astonishing economic and social costs. Social security and insurance providers estimate in more than 1,6 million € per capita the overall cost of a death by accident. This includes factor such as the loss of productivity of the victim and the direct/indirect costs related to the fatality itself. In any case, depending on the used methods, values oscillate between 2% and 4% of the GDP, almost close on the levels usually considered on average for an extraordinary government budget.

The analysis of these data, also available from the Italian *National Institute for Insurance against Accidents at Work* (INAIL), provides a clear picture of the pressing need to study the dynamics of the events and, above all, the perception that the new generations have about the issue of road safety and prevention.

3. The PSSG project

Official statistics, both at an international (Eurostat, Euro NCAP) and national (Istat) level, as well as survey conducted by independent organisations, confirm that the majority of road accidents are essentially caused by a lack of knowledge about vehicle's behaviour. Moreover, there is often a non-compliance with the most basic traffic rules. The leading cause of accident in Italy in 2013 has been distracted driving, together with unsure and improper practices.

The *Perception of Road Safety among Young People* project (PSSG) was launched considering that a deeper knowledge about drivers' behaviours and their driving style can contribute in increasing the culture of road safety. Formerly agreed by INAIL and the *Community of Mediterranean Universities* (CUM), a non-governmental organisation recognised by UNESCO, the project has been designed considering that young people will be the main users of the road system on average for the next 60 years. Therefore the perceptions of youngsters about safety are fundamental to change traffic rules in the future. The project was also supported by the Traffic section of the Italian Police, the University of Calabria and the Provincial Government of Cosenza.

The project aimed at assessing the adequacy of the current driving course and the related examination procedures. The PSSG consists of four steps:

1) a CAWI survey is conducted, aiming at collecting data about the different driving behaviours and attitudes of 18-35 years old individuals;
2) among the respondents which have completed the questionnaire a sample of 200 units is selected, and a new questionnaire is administered in order to deeply measure the perception about road safety.
3) this sample participate in a practical training course of safe driving; the course is delivered by professional instructors selected by the Traffic section of the Italian Police and the Sport Committee of the Italian Automobile Club;
4) the sample is followed up, at least 3-4 months after the safe driving training, and the perception about safety is measured again with a questionnaire.

This paper discusses the first step of the study. The questionnaire administered in the starting step considers socio-demographic characteristics, behaviours and opinions about road safety. It consists of 29 closed-ended questions with multiple choices. Data are collected online with the support of a Computer Assisted Web Interviewing system. Technically the questionnaire is drawn with *Survey Monkey*, a well known survey platform that allows also a real-time supervision of the collection process.

At the present more than 1.000 questionnaire have been collected, and the subsequent steps are still in progress. Only data acquired before May 2015 are considered hereby. The interpretation of the results will be however useful for understanding road safety, having an exploratory standpoint on the phenomenon.

4. An integrated strategy for analysing and classifying questionnaire data

The analysis of complex systems of data – with both a huge number of observations and a considerable number of features – often implies substantial critical states. We are interested in how strongly and in which way the different variables are interrelated. This aim can require hundreds if not thousands cross-tabulations, and makes really challenging the study of a multifaceted phenomenon.

Sample surveys, where typically data are collected with questionnaires, are a very typical case. In this context, it becomes necessary to carry out a statistical treatment of data, aiming at extracting the maximum amount of information and highlighting the key aspects of the observed phenomena. The statistical techniques for analysing contingency tables do not lend themselves to the required level of detail, because they often leave unused most of the inherent (and significant) information. At the same time, it is necessary to consider a strategy that achieves a general assessment of the questionnaire without *a priori* choices, in an exploratory fashion, offering new working hypothesis and viewpoints.

A suitable approach that fulfil these latter research needs is the well-known *Multiple Correspondence Analysis* (MCA), originally developed by Benzécri (1973) and Lebart et al. (1984) in the frame of the so called French School of Data Analysis. This method – far from being merely a generalisation of *Correspondence Analysis* – is able to perform an appropriate reduction of dimensionality onto a $n \times p$ table, where n is the number of observations (e.g., the respondents) and p a set of categorical variables (e.g., the questions). The original variables are then replaced by linear combinations of themselves known as *latent variables* or *factors*. The functional connections between the quantities are transformed into geometric relations, providing in this way a planar representation that allows an overall view and an immediate visualisation, of both features and units.

The factors and the factorial maps can be read in terms of percentage of explained inertia, a measure of variability that expresses the amount of original information represented in the reduced space obtained by MCA.

The parameters of MCA are estimated by pooling the data across units, under the implicit assumption that all the observations come from a single, homogenous group. However, it seems more realistic to assume that units come from heterogeneous groups, so that they are different with respect to their attitudes and behaviours, and the other characteristics of interest. The presence of groups depends clearly on the association structure over the data, but often MCA visualisations are not really intuitive and easy to read. It is not uncommon that the low-dimensional data are often difficult to interpret so that the resultant segments become difficult to characterise.

To cope with these issues, the French School proposed a two-step sequential approach sometime called *tandem analysis* (Arabie et al., 1996): after performing a factorial method for reducing the original variables, a clustering method onto the new variables is carried out. This strategy has the advantage of working on variables that are orthogonal and ordered with respect to the borrowed information.

Nevertheless, the choice of the factorial method is an important and critical phase because it may affect the final results, but taking into account this caveat an improvement of the overall quality of clustering will be achieved.

A satisfying way for obtaining non-overlapping clusters, is to consider a hierarchical CA, in which different levels of aggregation are investigated at the same time (Gordon, 1999). In practical applications, the agglomerative methods are the most used as they allow to construct a hierarchy of partitions with a significantly reduced computational cost. A partition with a higher aggregation index means that the distance between the two closest clusters is large, so they are well separated. Cutting the tree at a level corresponding to a significant "jump" in the index level leads to a good partition.

5. Data structure and main results

The online questionnaire used in the first step of the PSSG project was published on the website <http://www.liberastatistica.it> between the end of March and June 2015, and advertised both on the official website of University of Calabria and on the main social networks, aiming at reaching mainly young people.

The collected data have been pre-treated. By recoding the original variables, 21 new ones have been considered (Table 1). Even if the language used in the survey was Italian here in the follow the labels have been translated in English.

Table 1 – *Active variables.*

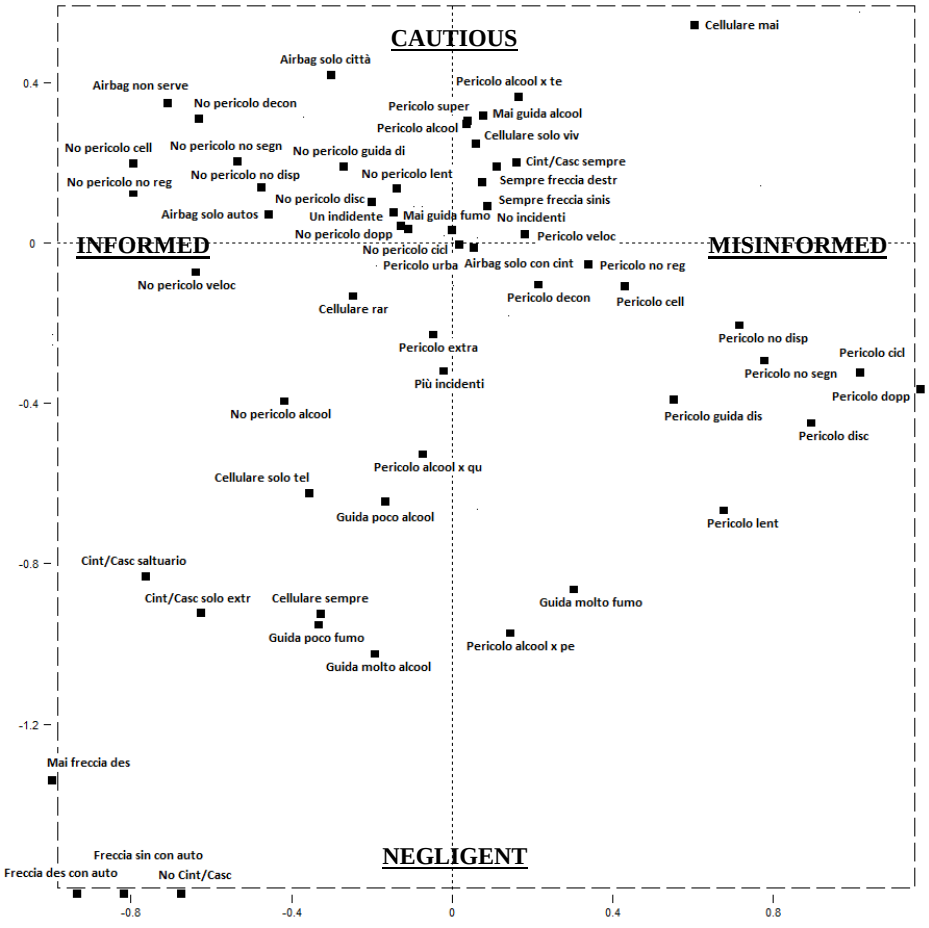
ID	Variable	No. of features
D_08	Use of seatbelt/helmet	4
D_09	Usefulness of Airbag	4
D_10	Use of signal on left turn	3
D_11	Use of signal on right turn	3
D_12	Use of mobile phone	5
D_13a	Driving after drinking	3
D_13b	Driving after smoking marijuana	3
D_14	Involvement in accidents	3
D_15	Danger of roads	3
D_16a	Danger in exceeding speed limit	2
D_16b	Danger of slow driving	2
D_16c	Danger in not being focused on driving	2
D_16d	Danger of double/triple parking	2
D_16e	Danger of using mobile and drive	2
D_16f	Danger of discussing with passengers	2
D_16g	Danger in not using vehicle lighting system	2
D_16h	Danger in not using vehicle security devices	2
D_16i	Danger in not respecting traffic rules	2
D_16m	Danger of a sly drive style	2
D_16n	Danger of cyclists	2
D_17	Danger in assuming alcohol before driving	5

The total number of questionnaires collected at the end of April was 606 in all, but only 465 complete questionnaire has been selected for the analysis. Socio-demographic variables have been used as supplementary ones. In this sample, around one in three respondents are male (37,85%), and there is a majority of people in the class of age 24-29 (41,51%). Moreover, more than a half travels on average up to 1.000 Km per month (58,50%), and four in five (82,37%) has only a driving license for driving cars.

5.1. Factorial Analysis

In order to analyse and graphically represent the latent association structure, the table cross-tabulating the 465 respondents and the 21 variables were considered. The first factorial plane (Figure 1) explains about the 67% of the total inertia. On this map is it possible to see on the first axis – from left to right – an opposition between the misinformed respondents and the informed ones (e.g., with respect to the traffic rules and the safest driving practices). On the second axis, it is possible to see instead – from the bottom to the top side – an opposition between respondents with a negligent and a cautious driving approach.

Figure 1 – *MCA: most significant active features on the first factorial plane.*

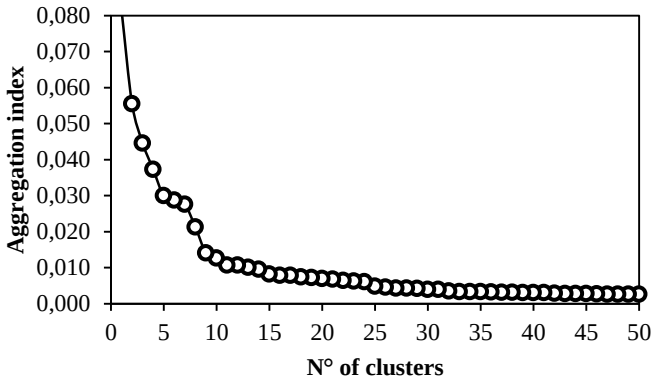


It would be possible to read the map in terms of quadrants, but in the frame of tandem analysis, it is more appropriate to perform clustering on the MCA factors.

5.2. Cluster Analysis

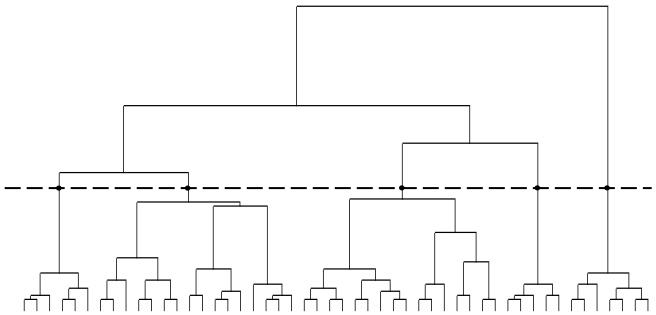
The clustering procedures performed is a hierarchical cluster analysis based on the Ward criterion for aggregating the units at the different steps. At each step of the criterion leads to aggregate together those groups for which there is a lower increase of deviance within the groups or, alternatively, the greatest decrease of deviance between the groups. The algorithm used in the clustering procedure is the nearest-neighbours chain (Benzécri, 1982).

Figure 2 – *Scree-plot with the number of clusters and the aggregation level.*



In this case, a partition with 5 clusters seems to be the optimal solution, as confirmed by the scree-plot (Figure 2). This solution is confirmed looking at the dendrogram (Figure 3)

Figure 3 – *Dendrogram representation with the partition in 5 clusters.*



The five clusters of the chosen partition can be described by considering for each one the most relevant features. To decide if a given feature *j* is a relevant characteristic of a cluster *k*, it is necessary to verify if it is significantly more present in *k* than in the whole sample. To deal with this problem a statistical significance test can be performed: the null hypothesis assumes an equal proportion of *j* in *k* and in the sample (against an unusually high proportion among the individuals in *k* with respect to the sample). By considering a hypergeometric random variable *N*, number of feature *j* observed in *k*, it is possible to calculate:

$$p_k(j) = \mathrm{P}\{N \geq n_{jk} \mid H_0\} \tag{1}$$

The higher is the quantity n_{jk} the lower is the hypergeometric probability, so that the null hypothesis is more doubtfully true. It can be read in an easier way if we consider the value, known as *test-value*, assumed by a Gaussian variable for the same probability $p_k(j)$ of a hypergeometric distribution:

$$test-value = \frac{n_{jk} - n_k \dfrac{n_j}{n}}{\sqrt{n_k \dfrac{n-n_k}{n-1}\dfrac{n_j}{n}\left(1-\dfrac{n_j}{n}\right)}} \tag{2}$$

The higher is the value the more the feature is characteristic. It has empirically shown that |test-value| > 2 suggest relevant features. For a proper reading of the following tables, it is necessary to remember that the different features have to be considered ever in terms of logical disjunctions.

Table 2 – *CLUSTER 1/5: Irresponsible Drivers (rel.: 25,6%).*

Variable	Feature	% of feature in the cluster	% of feature in the sample	Test-value
Danger in using mobile phone and driving	No dangerous	78,15	35,27	11,17
Danger in not using vehicle lighting system	No dangerous	95,80	59,35	10,26
Danger in not using vehicle security devices	No dangerous	94,12	60,22	9,46
Danger in not respecting traffic rules	No dangerous	61,34	30,11	8,26
Danger in not being focused on driving	No dangerous	53,78	25,59	7,73

To the first cluster (Table 2), namely *Irresponsible Drivers*, belongs one-fourth of the respondents. They are characterised by a very low perception about danger. The majority of people in this cluster considers not dangerous the use of mobile

devices (78,15%), as well as they consider safe not following the basic traffic rules (96% thinks is not danger not using lighting system). The age does not characterise particularly this group, because the three class of age are quite equally represented.

Table 3 – *CLUSTER 2/5: Haphazard Drivers (rel.: 15,1%).*

Variable	Feature	% of feature in the cluster	% of feature in the sample	Test-value
Use of mobile phone	Ever	31,43	6,67	7,23
Use of mobile phone	Only for calls	42,86	12,26	7,16
Use of seatbelt/helmet	Only suburban roads	30,00	6,24	7,12
Driving after drinking	After few alcohol	58,57	29,03	5,51
Danger in exceeding the set speed limit	No dangerous	40,00	22,37	3,51

In the second cluster (Table 3), namely *Haphazard Drivers*, falls the 15,1% of the sample. The respondents belonging to this cluster have a subjective perception of safety (30% uses seatbelt only on suburban roads; 58,57% judges possible to drive after drinking small quantities of alcohol). It is interesting to note that around two in three in this group travel on average for more than 500 KM per months.

Table 4 – *CLUSTER 3/5: Cautious Drivers (rel.: 40,4%).*

Variable	Feature	% of feature in the cluster	% of feature in the sample	Test-value
Danger in using mobile phone and driving	Dangerous	89,36	64,73	9,51
Danger in exceeding speed limit	Dangerous	95,21	77,63	7,96
Use of seatbelt/helmet	Ever	96,81	81,72	7,44
Danger in not using vehicle security devices	Dangerous	54,79	39,78	5,35
Use of signal on right turn	Ever	99,47	92,04	5,33

Table 5 – *CLUSTER 4/5: Frightened Drivers (rel.: 12,9%).*

Variable	Feature	% of feature in the cluster	% of feature in the sample	Test-value
Danger in double/triple parking	Dangerous	48,33	7,53	9,96
Danger in not using vehicle lighting system	Dangerous	95,00	40,65	9,46
Danger in arguing with the other passengers	Dangerous	66,67	18,28	8,98
Danger in not using vehicle security devices	Dangerous	81,67	39,78	6,99
Danger for the presence of cyclists	Dangerous	36,67	9,68	6,19

The third cluster observed (Table 4), namely *Cautious Drivers*, consists of 40,4% respondents. These individuals believe that is unsafe to use mobile phones (around 65% of the overall units). At the same time, nearly all of them consider dangerous the speediness (around 95% of people belonging to the cluster), or declare to use ever safety equipment like seatbelt (around 95% of people belonging to this group).

To the fourth cluster (Table 5), namely *Frightened Drivers*, belongs the 12,9% of respondents. These drivers have a high perception and sensitivity to danger. The majority (95%) considers risky not using devices like lights or turn signals, more than a half thinks is dangerous to argue with the other passengers, and one in three consider dangerous also the presence of cyclists on the same roads of cars. A quite high percentage also thinks that double (or triple) parking is dangerous.

Table 6 – *CLUSTER 5/5: Sly Drivers (rel.: 6%).*

Variable	Feature	% of feature in the cluster	% of feature in the sample	Test-value
Use of signal on left turn	Signal on left turn with other vehicles	82,14	4,95	12,18
Use of signal on right turn	Signal on right turn with other vehicles	89,29	7,31	11,47
Use of seatbelt/helmet	Sometime	39,29	11,83	3,71
Danger in assuming alcohol before driving	Dangerous only for some individuals	57,14	31,40	2,72
Danger in not using vehicle lighting system	No danger	82,14	59,35	2,42

The last cluster (Table 6), namely *Sly Drivers*, includes a restricted number of individuals (8%). In the group 4 in 5 admit to use turn signals only if there are other vehicles nearby. More than a half believes that danger in driving after drinking depends only on the quantity assumed. Less than 3% of the women in the sample belongs to this cluster, where among men percentage is slightly over 10%.

6. Remarks and future directions

The clusters obtained from the MCA factors seems to balance between the percentages of responsible and not responsible drivers. Nevertheless, it would be necessary to validate results with a wider sample of young drivers. On the other hand, the well-known problems related to sample representativeness in Web surveys should be considered by re-weighting the collected data, because of the biases related to the self-selection of individuals participating into the survey and

their peculiar characteristics. Overall, the analysis presented in this paper can be seen as a pilot step for the following phases of the PSSG project.

From a statistical point of view the tandem approach is very useful even if it could not be as a black box for solving some difficulties in reading and interpret factorial analysis. The choose of MCA and Cluster Analysis is very common and already used in several fields, but some alternatives can be considered in order to have more robust results, e.g. by using constrained analysis or integrating the two steps of tandem analysis in one simultaneous analysis (e.g., Hwang et al., 2005).

References

ARABIE P., HUBERT L.J., DE SOETE G. 1996. *Clustering and Classification*. River Edge (NJ): World Scientific.

BENZÉCRI J.P. 1973. *L'analyse des données*. Vol. 2, Paris: Dunod.

BENZÉCRI J.P. 1982. Construction d'une classification ascendante hiérarchique par la recherché en chaîne des voisins réciproques. *Les Cahiers de l'Analyse des Données*, Vol. 2, No. 7, pp. 209-218.

GORDON A.D. 1999. *Classification*. Chapman & Hall/CRC.

HWANG H., YANG B., TAKANE Y. 2005. A simultaneous approach to constrained Multiple Correspondence Analysis and Cluster Analysis for market segmentation. *Asia Pacific Advances in Consumer Research*, Vol. 6, pp. 197-199

ISTAT. 2014. *Incidenti stradali in Italia 2013*.

LEBART L., MORINEAU A., WARWICK M.K. 1984. *Multivariate Descriptive Statistical Analysis*. Wiley & Sons.

SUMMARY

Perception of road safety among young people: the PSSG project

The PSSG Project aims at investigating the perception of youth about road safety. From the data obtained by the questionnaire administered at the first stage of the project, we try to profile young drivers on the base of their opinions, in the frame of Multivariate Statistics, with particular reference to Multiple Correspondence Analysis and Cluster Analysis.

Pietro IAQUINTA, DiScAG – UniCal, pietro.iaquinta@unical.it
Michelangelo MISURACA, DiScAG – UniCal, michelangelo.misuraca@unical.it