HOW THE COMMUTERS MOVE: A STATISTICAL ANALYSIS BASED ON ITALIAN CENSUS DATA

Gabriella Schoier, Adriana Monte

1. Introduction

Human mobility and, in particular, commuting has an important role in understanding socio-economic systems (Goodman, 2013). Analyzing and modelling the networks formed by commuters has become a crucial requirement in studying areas dynamics and in helping decision-making. The analysis of commuting and its effects on housing, labor market, planning of infrastructures etc. can be viewed as a part of spatial interaction modelling in a spatial Data Mining context (Koperski et al., 1998).

In 1971 questions on mobility have been introduced in the Italian Census. From 1981 matrices for inter-municipal mobility referring to home to work/study flows have been built. In the 2011 Census important innovations have been introduced by example the use of sampling in order to produce estimates for traditionally exhaustive variables. Two questionnaires have been used: the long form questionnaire (L) and the short form (S) questionnaire.

In this study we use the matrix derived from the 2011 Italian Census to analyze mobility flows of Friuli Venezia Giulia region. Each record represents a single stratum individuate on the base of some variables such as sex, commuting reason (work/study), place of residence, place of work/study, used modes of transport, time of departure, time employed to reach work/study place.

The approach is based on spatial network theory. Moreover two step cluster analysis which allows to take into account flow typologies in a detailed way and different kind of variables is considered and used to individuate clusters of commuters who share common characteristics.

2. Mobility data from Census

In 1971 for first time questions on mobility have been introduced in the Italian Census, but only a sample has been processed by Istat. Since 1981 all the collected information have been used and matrices for inter-municipal mobility referring to

home to work/study flows have been built. In the subsequent Censuses the questionnaires contained a specific section for commuting (Salas-Olmedo et al., 2012).

In the 2011 Census important innovations have been introduced; in particular the use of sampling techniques for the collection of some socio-economic data. Two questionnaires have been prepared: the long form questionnaire (L) and the short form questionnaire (S). Referring to mobility, the S questionnaire collects these information: the condition of being a commuter (a person is defined commuter if he moves from habitual residence to work or study place every day and in the same day he comes back home), the reason of commuting (work or study), the habitual residence, the work or study location (municipality, foreign country). In the L questionnaire there are also three others information: mode of transport, time to arrive at work or school, time when the commuter departs from home. The sampling has concerned only the municipalities with 20,000 inhabitants or more and the provincial capital, in this case the S questionnaire has been distributed. The L questionnaire has been distributed to the remaining families and to the families of the municipalities with less than 20,000 inhabitants. The consequence of this procedure is that data collected using only L questionnaire inferred to all the population by an estimation procedure.

3. Mobility for study and work in Friuli Venezia Giulia

Friuli Venezia Giulia is a small Italian region where the resident population is approximately 1.2 million (see Table 1) and most of it (about 64%) lives in little towns, with less than 20,000 inhabitants.

Table 1 – *Resident population and commuters in Friuli Venezia Giulia at 2011 Census.*

Provinces	Resident population (R)	Commuters (P)				P/R %
		Total	In community	In family		
				study	work	
Gorizia	140,143	68,356	37	20,008	48,311	48.78
Pordenone	310,811	163,221	93	47,426	115,702	52.51
Trieste	232,601	114,967	147	33,319	81,501	49.43
Udine	535,430	270,895	169	77,831	192,895	50.59
Total FVG	1,218,985	617,439	446	178,584	438,409	50.65

Source: 2011 Census.

Mobility for study or work regards about the 50% of the population. The most used mode is car as in the previous Censuses (see Table 2). Some differences regards the municipality of Trieste, the main town (approximatively 200,000

inhabitants) of the region. Car is the most frequent mode in this town and in its province but the distribution of commuters among modes of transport is different from the others provinces. This is due not only to the dimension of the town but also to its geographical conformation (Istat, 2015).

Table 2 – *Percent distribution of commuters by modes of transport in Friuli Venezia Giulia at 1991, 2001 and 2011 Censuses.*

Mode of transport	1991	2001	2011
railway, tram	2.86	1.87	1.89
bus, work bus	18.52	13.79	13.73
car (as driver)	43.93	51.10	50.78
car (as passenger)	9.15	12.11	13.34
bicycle	4.84	5.24	3.52
others	20.71	15.89	16.75

Source: 1991, 2001 and 2011 Censuses.

In general the mode of transport is related to the reason of commuting. The commuters for work move by car as driver in 70.3%, meanwhile the commuters for study use other modes of transport: the most frequent modes are bus (33.2%), car as passenger (37.0%) and walking (16.5%).

As regards time to arrive at work the most of commuters travel up to 15 minutes. If we consider time to arrive up to 30 minutes, the percentage arises to 87.3% in the case of work and to 83.2% in the case of study (see Table 3).

Table 3 – *Percent distribution of commuters by time to arrive at work or study in Friuli Venezia Giulia at 2011 Census.*

Time to arrive	Work %	Study %
Up to 15 minutes	55.61	62.58
16 – 30	31.71	21.64
31 – 45	7.03	7.06
46 – 60	3.30	4.97
More than 60 minutes	2.35	3.76
Total	100.00	100.00

Source: 2011 Census.

4. Data organization and used methodology

In this study we use data of the matrix for inter-municipal mobility of 2011 Census. In order to identify commuter typologies we use only the L records as they

contain the information relate to mode of transport, time to arrive[1] and time of departure.

We use the data of commuters usually residents in Friuli Venezia Giulia extracted from the matrix of Italian municipalities. The L records refer to commuters living in family (the 99.9% of commuters in Friuli Venezia Giulia). Each record represents a typology of commuter who has the same value for each variable. There are 106,167 different typologies of commuters and 616,993 commuters. Variables are both qualitative and quantitative.

Table 4 – *Variables of the matrix for inter-municipal mobility of 2011 Census (L records).*

variables		modalities description
Province of residence	qualitative	Istat code
Municipality of residence	qualitative	Istat code
Sex	qualitative	male female
Mobility reason	qualitative	work study
Work/study place	qualitative	municipality of residence other municipality foreign country
Province of work/study	qualitative	Istat code
Municipality of residence	qualitative	Istat code
Foreign country of work/study	qualitative	Istat code
Mode of transport	qualitative	railway, tram, metro, city bus, bus, work or school bus, car (as a driver),car (as a passenger), motorcycle , bicycle, other mode, on foot
Time of departure from home	quantitative continuous	before 7.15 from 7.15 until 8.14 from 8.15 until 9.14 after 9.14
Time to arrive at work/study place	quantitative continuous	up to 15 minutes from 16 to 30 minutes from 31 to 60 minutes more than 60 minutes
Estimated number of commuters	quantitative	

[1] Furthermore time to arrive at work or study place is comprehensive of the time eventually used to accompany children at school

We know that a commuter might use more than one mode of transport in order to reach the work or study place[2]; but Census asks the commuter to reply only on the mode of transport used for the longer distance. The unit of our analysis is the commuter moreover we consider separately the commuters for work and the commuters for study because they have different characteristics for example the age which determines the mode of transport.

In order to obtain clusters of commuters we have applied SPSS two step cluster analysis as developed by Chiu et al. 2001. It consists in two steps:

Step 1: Pre-clustering of cases. A sequential approach is used to pre-cluster the cases. The aim is to compute a new data matrix with fewer cases for the next step; in order to reach this aim, the computed pre-clusters and their characteristics (cluster features) are used as new cases. The pre-clusters are defined as dense regions in the analyzed attribute space. The results may depend on the input order of cases therefore it is recommended to use random order.

Step 2: Clustering of cases. A model based hierarchical technique is applied. Similar to agglomerative hierarchical techniques, the pre-clusters are merged stepwise until all clusters are in one cluster. In contrast to agglomerative hierarchical techniques, an underlying statistical model is used. The model assumes that the continuous variables are within clusters independent normal distributed and the categorical variables are within clusters independent multinomial distributed. Two distance measures are available: euclidean distance and a log-likelihood distance in case of mixed types of attributes (Ming-Yi, 2010; Bacher et al., 2004).

In order to run two step cluster analysis we have used the same variables in both the matrices, in particular we use three qualitative variables (sex, work/study place, mode of transport) and two quantitative variables (time of departure and time to arrive). These last two variables have been recoded using the central value of each class. For time of departure the first class has been considered larger in order to contemplate the commuters departing very soon in the morning.

In order to apply the two step cluster method we expand the file so that a record represents a commuter not a stratum. Furthermore from the original matrix of Friuli Venezia Giulia we have derived two matrices, the former for the municipalities up to 20,000 inhabitants and the latter for the other six municipalities (Trieste, Gorizia, Pordenone, Udine, Monfalcone, Sacile). A preliminary analysis has

[2] Since 1993 every year Istat produces the sample survey on households: " Aspects of daily life " with a section dedicated to the daily commute done for work or study reasons, but the territorial level is the region, not the municipality.

revealed different behaviors of the commuters not only due to the size, but also to the different characteristics of the six larger towns.

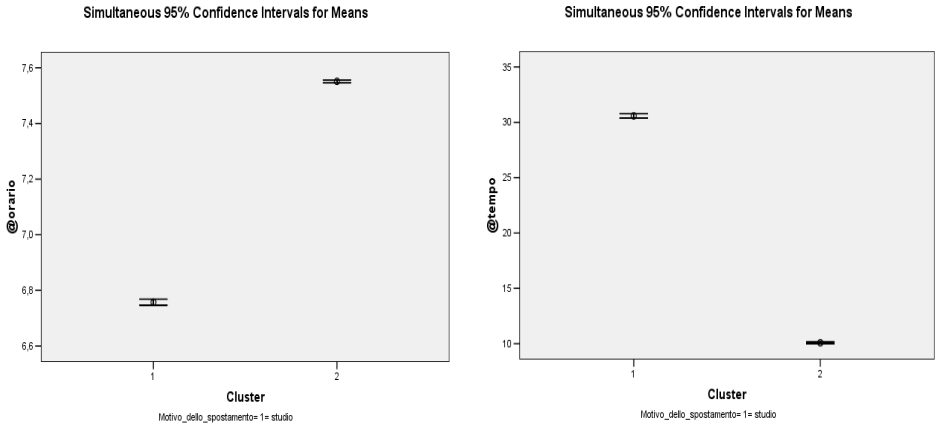
5. Results relating to study mobility

We divided the results for study mobility according to the number of inhabitants: under or above 20,000. The total amount of commuters is 178,584, the 65.4% belongs to municipalities under 20,000 inhabitants, the rest is distributed in the other municipalities (Trieste, Gorizia, Pordenone, Udine, Monfalcone, Sacile).

5.1. Study mobility in municipalities under 20,000 inhabitants

In this case we found two clusters, the former contains 44.3% of commuters, the letter the remaining 55.7%. Sex is identically distributed in both clusters. Next figure presents box-plots as regards time of departure from home (@orario) and time to arrive at studying place (@tempo). As one can see they are well differentiated.

Figure 1 – *Box-plots of time of departure from home and time to arrive at studying place .*



The modes of transport are different in the two clusters. In detail the commuter profiles in the two clusters are:

Cluster 1 - students who go to study in a municipality different from that of residence, who go out home early (on average before seven) and employ on average about thirty minutes to get to the place of study mainly using a bus, but also the car as a passenger.

Cluster 2 - commuters who study in the municipality of residence, who leave home about an hour after those in the Cluster 1. They employ just over 10 minutes to get to the place of study, mainly they use car as passengers.

5.2. Study mobility in municipalities over 20,000 inhabitants

In this analysis we consider the two most important municipalities of the Friuli Venezia Giulia i.e. Trieste and Udine. The former is the regional capital and that of larger dimension. The latter is smaller, but its province is the largest of the region in terms of number of municipalities, territorial extension and population.

Three clusters have been identified in both municipalities.

In Udine, for its territorial shape, commuters use more the bicycle than in Trieste. Almost everyone who moves out of municipality to study uses the city bus in Trieste (90.7%), while in Udine in addition to this, the car as passenger, is used.

As regards Trieste the clusters are formed by:

Cluster 1 – commuters who study inside the municipality, use mainly the car as a passenger or motorcycle leave home later and employ an average of twelve minutes to get to the place of study.

Cluster 2 – commuters who study in other municipalities and mainly use the city bus (90.7%), go out home before the other commuters and take about twenty five minutes to arrive at place of study.

Cluster 3 – commuters who study at Trieste, move on foot, go out later from home and use less time of the commuters of the other two clusters to reach the place of study (ten minutes an average).

As regards Udine the clusters are formed by:

Cluster 1 – commuters who study also in other municipalities and mainly use the city bus (69.6%) or the car as a passenger (17.3%). They leave home early and employ a longer time than others to reach the place of study.

Cluster 2 – commuters who study at Udine, and use the car as a passenger, leave home half an hour later than commuters of Cluster 1 and take about ten minutes to arrive at study place.

Cluster 3 – commuters who study at Udine and go mainly on foot or by bicycle to the place of study. They leave home a little after persons belonging to Cluster 2 and employ on average a few minutes to go to the place of study.

6. Results relating to work mobility

Also the results for work mobility are divided according to the dimension of municipalities: under or above 20,000. The total amount of commuters is 438,409, the 65.4% belongs to municipalities under 20,000 inhabitants, the rest is distributed among the other municipalities (Trieste, Gorizia, Pordenone, Udine, Monfalcone, Sacile).

6.1. Work mobility in municipalities under 20,000 inhabitants

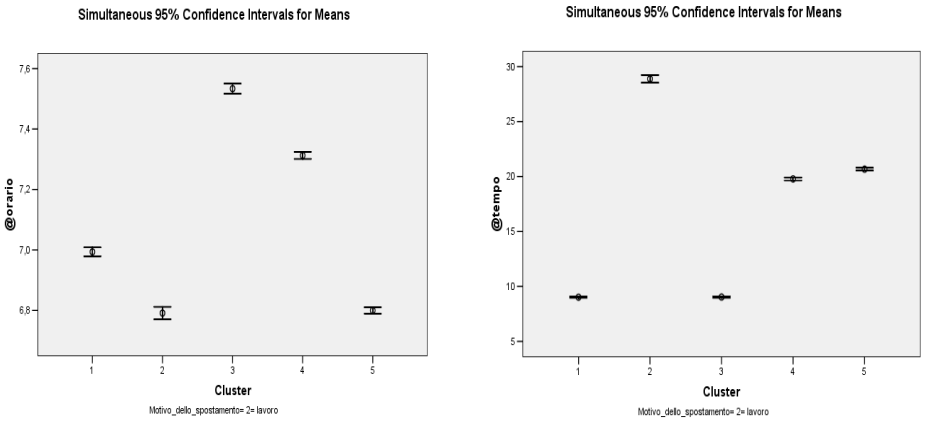
In this case we found five clusters, the distribution of commuters is presented in Table 5. The largest cluster is Cluster 5, followed by Cluster 4.

Table 5 – *Distribution of commuters for work in the clusters.*

Cluster	N	%
Cluster 1	48,012	16.75
Cluster 2	28,373	9.90
Cluster 3	37,932	13.23
Cluster 4	76,649	26.73
Cluster 5	95,742	33.39
Total	286,708	100.00

In cluster 1 and cluster 5 there are only males while in cluster 3 and cluster 4 there are only females. Next figure presents box-plots as regards time of departure from home (@orario) and time to arrive at work (@tempo). As one can see there is a good differentiation among the five box-plots for both variables.

Figure 2 – *Box-plots of time of departure from home and time to arrive at work.*



The modes of transport differ in the five clusters. For example commuters in Cluster 4 and Cluster 5 use exclusively car as drivers, while commuters in Cluster 2 use all the other modes, only a few of them drive the car to go to work. In detail the commuter profiles in the five clusters are:

Cluster 1 – Males who work in the same municipality of residence, use mostly their own car (61.8%), but move on foot (20.7%) and by bicycle (11.6 %), go out of home at around 7 a.m. and take about 10 minutes to get to the place of work.

Cluster 3 – Females who work in the same municipality of residence and move likewise persons belonging to Cluster 1 but come out much later from home.

Cluster 4 – Females who work in municipalities other than that of residence, they use car, they go out from home a little before females of Cluster 3 and employ on average twenty minutes to reach workplace.

Cluster 5 – this cluster differs from Cluster 4 for the fact that it is formed by males going out home much earlier than females in their same condition.

Cluster 2 – it is formed by males and females who work primarily outside the municipality of residence, have in common that they do not to use the car but use all other types of vehicles (including car as a passenger). They tend to go out soon, before commuters of any other cluster.

6.2. Work mobility in municipalities over 20,000 inhabitants

As in the case of study mobility, in this analysis we consider the two most important municipalities, Trieste and Udine. In this case we have identified three clusters for Trieste and two for Udine.

As regards Trieste the cluster are formed by:

Cluster 1 – females working at Trieste, mainly using city bus, but going on foot and by motorcycles too. They leave home later than the others commuters and employ on average sixteen minutes to reach the work place.

Cluster 2 – males and females working in other municipalities and mainly using private cars. They leave home before the other commuters and use on average about twenty minutes to get to the place of work (56.8% are males).

Cluster 3 – males working in Trieste, moving by motorcycle, but also by bus or on foot. They leave home on average at 7.15 and use less time than the commuters of the other two clusters to reach the place of study.

As regards Udine the clusters are formed by:

Cluster 1 – predominantly male (63.2 %) who work out Udine and use primarily car, partly railways. They leave home soon and take longer than others for the trip.

Cluster 2 – these commuters work at Udine and use mainly the car, but also move by bike or on foot, partly they use city bus. They leave home half an hour after commuters of Cluster 1.

7. Conclusion

The methodology used for analysis i.e. the two step cluster analysis made it possible to obtain well defined groups of commuters in the region Friuli Venezia Giulia.

In the municipalities under the 20,000 inhabitants two distinct clusters are highlighted in the case of commuting to study and five in the case of commuting to work. Variables used discriminate well between clusters in both cases.

Among municipalities with at least 20,000 inhabitants we presented the results for the two main municipalities: Trieste and Udine.

By considering the results of this analysis it would be interesting strengthening incentives to use public transport by improving the supply, increasing the use of less polluting vehicles (electric cars) or car sharing.

References

BACHER J., K. WENZING M., VOGLER. M. 2004. SPSS Two Cluster A First Evaluation, *Universitat Erlangen-Nurnberg*, pp. 1-20, www.statisticalinnovations.com/products/twostep.pdf_cited July, 2015.

BORRELLI F., CARBONETTI G., DE FELICI. L., SOLARI F. 2012. Metodologie di stima per piccole aree applicabili a variabili di censimento, *Istat Working Papers*, No. 3.

CHIU T., FANG D., CHEN J., WANG Y., JERIS C. 2001. A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment.

In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2001*, pp. 263–268.

GOODMAN A. 2013. Walking, Cycling and Driving to Work in the English and Welsh 2011 Census: Trends, Socio-Economic Patterning and Relevance to Travel Behaviour in General, *PLoS ONE* 8(8) http://journals.plos.org/plosone/ article?id=10.1371/journal.pone.0071790 cited July 2015

ISTAT 2015. *Transport and road accidents*. Serie Storiche, archivio della Statistica Italiana. http://timeseries.istat.it/fileadmin/allegati/Trasporti/ testi_inglese/17_Travel_to_school__university_or_workplace.pdf cited July 2015

KOPERSKY K., HAN J., ADHIKARY J. 1998. Mining Knowledge in Geographical Data. \\$ftp://ftp.fas.sfu.ca/pubcs/han/pdf/geo_survey98.pdf.$ cited July 2015.

MING-YI S., JAR-WEN J., LIEN-FU L. 2010. A Two-Step Method for Clustering Mixed Categorical and Numeric Data, *Tamkang Journal of Science and Engineering*, Vol. 13, No. 1, pp. 11-19.

SALAS-OLMEDO M. H., NOGUES S. 2012. Analysis of commuting needs using graph theory and Census data: A comparison between two medium-sized cities in the UK, *Applied Geography*, Vol. 35, pp. 132-141.

SUMMARY

How the Commuters move: a Statistical Analysis based on Italian Census Data

The matrix of commuting, built from Census data, provides information on daily movements from the municipalities of residence to work or study place. The territorial detail considered in the matrix (available on the Istat website) is the municipality inside of which commuters are stratified on the basis of some relevant variables. In the present paper, using such matrices, we try to identify types of commuters that differ in terms of commuting according to the dimension of the municipalities: below and above 20,000 inhabitants. The methodology used is the two step cluster analysis.

Gabriella SCHOIER, University of Trieste, gabriella.schoier@deams.units.it
Adriana MONTE, University of Trieste, adriana.monte@deams.units.it