WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods in Regional Science

Vol. 6
No. 2
Summer
2011

ISSN 1842–4562

JAQM Advisory Board

Contents

QUANTITATIVE METHODS FOR IDENTIFICATION OF REGIONAL CLUSTERS IN ROMANIA[1]

Adriana Elena REVEIU[2]

PhD, Associate Professor, Department of Economic Informatics and Cybernetics
University of Economic Studies, Bucharest, Romania

E-mail: reveiua@ase.ro

Marian DARDALA[3]

PhD, University Professor, Department of Economic Informatics and Cybernetics
University of Economic Studies, Bucharest, Romania

E-mail: dardala@ase.ro

Abstract:

The main goal of this paper is to describe some methods, that make use of employment data and that allow to measure to what the companies are spatially proximate. Specifically, we will outline the most prominent spatial concentration quotients that have been suggested in the literature to analyse the degree to which companies of the same sector are proximate (spatial concentration). We apply the methods on the employment statistics available for Romania's counties.

Key words: *regional clusters; quantitative methods; location coeficient method; shift-share analysis; Gini's location quotient; Ellison and Glaeser's agglomeration index*

1. Regional Clusters

The economic activity is concentrated in space, and therefore there is an increased attention over the forces of agglomeration and the role of location in economic development. Porter [1][4] defines clusters as a group of interconnected companies and associated institutions, close from geographical point of view, working in a particular field and linked by common and complementary elements.

Because of the proximity among them, both in terms of geography and of activities, the clusters constituent enjoy the economic benefits of several types of positive location-specific externalities.

Knox [2] defines a spatial cluster as a geographically bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance. This is a useful operational definition, but there are very few situations when phenomena are expected to be distributed randomly in space. In most cases an implicit assumption in spatial cluster analysis is that the researcher has accounted for all the factors known, to influence the variable of study.

From a functional perspective, clusters are defined as networks of independent producers of powerful firms, including specialized suppliers, linked to each other in the value-added production channel. [3]

Spatial proximity has grown rapidly in importance, the cluster literature have made the distinction between industrial complexes and industrial clusters on spatial agglomeration of these industrial groups. Spatial proximity of the industrial activities interconnected assumed to influence the performance of these sectors and regional clusters on short and long term [4].

Clusters differ in many dimensions, such as: the type of products and services they produce, the location dynamics they are subject to, their stage of development, and the business environment that surrounds them.

Clusters can be **classified** by the type of product and services they provide. There are clusters in automotive, in financial services, in tourism, in a specific industrial area, and so on. Researches have pointed out how different locations play different roles. The development of clusters has discouraged many regions with no realistic chance of achieving a similar level of performance as the top level clusters.

From location point of view, the *local industries* are serving only local markets and are distributed across space approximately according to population size. They might be a kind of a cluster in a more narrow geographical sense, like a part of a city, due to the complementarities in attracting customers, but these effects are not strong enough to influence the development of clusters across regions.

On the other hand, the *natural resource dependent industries* serve global markets and are concentrated across space, in areas in which there are natural resources presented.

Finally, there are many industries that choose their location according to the quality of the cluster-specific business environment. There are the case of *traded industries* which serve markets in many regions and countries, and concentrate across various geographic locations. The cluster belongings to one industry is strong and its presence is a key part of the attractiveness of a specific location. Understanding the differences between these types of industries is important, because it affects the types of policies that are relevant to upgrade them.

From another point of view, clusters can be classified by the stage of development they have reached. The stage of development depends on two dimensions: on the quality of the external business environment the cluster operates in, and on the progress the cluster has made in mobilizing the potential of its business environment through active cooperation and other internal activities. [5] Researchers have looked at clusters in less developed economies as well as in less developed regions of advanced economies, such as rural regions or inner cities. Most of the theoretical literature suggests that clusters are a factor at every stage of economic development, but that in weaker environments clusters will tend to be weaker and more narrow as well. [6] Researchers have focused on the role of cultural factors, institutions, and individual leadership. There is strong view in the literature that cluster dynamics do not occur automatically, but that they depend on and can be reinforced by purposeful action. [6]

2. Quantitative Methods Used for Indentification of Regional Clusters

In the literature there are several ways of grouping the industries into clusters. To have a credible image of the cluster construction process, we can use different types of statistics and databases and various ways to collect information. Generaly, the choice of method for cluster representation depends on the type of cluster.

Location Coeficient Method is designed to group local industries into clusters, using regional data about employees. In order to identify the leading regions of a spatially concentrated industry, Kim (1995) and Hoover (1936), suggest to calculate for each locational unit in a given sample industries' employment shares, with respect to each industry's total employment in the aggregated locational unit.

$$Location\ quotient = \frac{\frac{n_{A,R}}{N_R}}{\frac{n_{AT}}{N_T}}$$

where: $n_{A,R}$ is the number of employees in industry **A,** in region **R**,

N_R is the whole number of employees, in the region **R**,

n_{AT} is the number of employees, in industry **A,** at the national level,

N_T is the whole number of employees, from national level.

A region is considered to be specialized in one industry if the location quotient calculated for that region is greater than or equal to 1.5.

The method is structured as following:

1. The target geographical area is divided into regions.
2. Identification of global industries, based on the location quotient calculate for each industry. Using this quotient, the industries from each regions, could be classified in three groups: local industries, global industries and dependent by the resources industries. If there are several regions specialized in an industry, the methodology assumes that the industry is oriented globally. An industry is considered to be globally or global oriented if it exports the products outside the region or country. These are very important industries for a region because they are promoting economic growth for other industries. Local industries are the industries without export outside the region or country. Dependent by resources industries are those for which the location is defined by the resources availability.
3. Location quotients are analysed to identify patterns of clustering. Clustering algorithm is used to browse the different ways of grouping the industries to identify the best solution for grouping industries, based on the location quotient. It is used as a cluster quotient when the same group of industries is over represented in some different regions.

The choice of regions, industries and group identification are parts of an iterative process. In each step can be made refinements, until the definition of clusters match the reality. To do this, the resulted clusters are verified by various qualitative assessments.

The method has been applied in many countries because it uses only employment data, which are relatively easy available.

The main shortcoming of the method is the large dependence by the regions bounds choosing. The choice of regions must be a priori to identify clusters.

To resolve the problem of choosing the size of regions used in location quotient method and to have a more flexible method for clusters mapping, the **Ripley's K method** can be used.

Ripley's K method consider clusters mapping like an optimization problem of the distances between the companies. In this situation, it is not necessary to choose the regions in advance because the method identifies the optimal size of each cluster without predetermined geographical boundaries.

The methodology consists in:

1. Designing the locations of all the companies in each industry and compute the distance between the companies for each industry. Geographical concentrations of each industry can now be compared to measure the performance and the distribution of all the employees. The comparison shows whether a particular industry has a local over representation and if it can be considered as it is globally oriented. Geographical concentrations are identified by optimizing the distances between the companies, which is the size of specialized areas. This issue solve the problem of predefined chooses of regions of the location quotient method.
2. The patterns relating to the location of global industries are evaluated using statistical features. A cluster algorithm tries to identify the locations for each industry, in order to identify systematic patterns of clustering among industries. Like in the location quotient method, the mapping is an iterative process to identify the best clustering corresponding to reality.

The main shortcoming of this method is a greater dependence on the details about the location of each company, data hardly available. More than this, in the case of Ripley's K method, the volume of calculations to be made is extremely high.

Shift-share analysis decomposes in factors the changes in value of an indicator, such as number of employees, income, added value and so on. Decomposition is done in three parts and expresses the effect of absolute change of the indicator and the effect of changes in its structure. The method uses the assumption that regional economic growth can be explained by a combined effect of three components: increasing at national level, growth in the structure of the branch and growth due to other factors, the local factors. The last so-called competitive component is rated as most important; it emphasizes the region's top branches. Mathematically, the decomposition can be expressed by the equation:

$$ZP = ZN + ZO + ZR$$

where: ZP = changes in share of the selected index,

ZN = changes of the selected index, at the national level,

ZO = changes in share of the branch structure of the selected index,

ZR = regional changes in share of the selected index.

Changes of the index value is compared for two time periods, not necessarily two consecutive years, but rather is recommended a longer period (3-5 years). To perform the computations, the available values of the index should cover a larger area and the region for the two selected years should be divided according to NACE specifications. The individual components of the equation are determined by the relations established according with the following equations:

$$ZN = \frac{CR^t}{CR^{t-n}}$$

$$ZO = \frac{CR_i^t}{CR_i^{t-n}} - ZN$$

$$ZR = \frac{R_i^t}{R_i^{t-n}} - \frac{CR_i^t}{CR_i^{t-1}}$$

where: CR^t – the average number of employees of the national economy, in year *t,*

CR^{t-n} – the average number of employees of the national economy, in year *t-n*,

CR_i^t – the average number of employees of branch *i*, in year *t*,

CR_i^{t-n} – the average number of employees of branch *i*, in year *t-n*,

R_i^t – the average number of employees, from the region, in branch *i*, in year *t*,

R_i^{t-n} – the average number of employees, from the region, in branch *i*, in year *t-n*,

n- the length of analysed perriod.

Ellison and Glaeser's agglomeration index

The index defines the share of total geographical concentration for the branch *i*.

$$G_{EG}^k = \frac{\sum_i \left(l_i^k - x_i\right)^2}{1 - \sum_k x_i^2},$$

where: $x_i = \frac{\sum_k z_i^k}{\sum_i \sum_k z_i^k}$ – is the share of region *k* in the employment of the whole industry.

$l_i^k = \frac{z_i^k}{Z_i}$, where:

l_i^k – the share of employment of region *k*, from branch *i,*

z_i^k – the number of employees in branch *i*, in region *k*,

Z_i – the number of employees in branch *i*, at the national level.

The index is based on the comparation of the shares of employees in the selected branch, in the region and in the whole manufacturing branch.

If the index values are less than 0, then the branch is dispersed across the whole territory and cannot be described as geographically concentrated.

For the index value in a range between 0 to 0.02, it is an insignificant, very weak geographical concentration of the branch. For an index between 0.02 and 0.05 it is a medium-strong geographical concentration, and above 0.05 of strong geographical concentration [7]. In order to determine industry concentration, the modified Herfindahl index must be determined for branch *i* through application of the equation: $H^i = \sum_j \left(z_j^i\right)^2$ where: z_j^i – share of employees in enterprise *j* in the whole share of employees in the branch *i*.

The aglomeration index calculed using the formula: $\gamma_{EG}^{i} = \frac{G_{EG}^{i} - H^{i}}{1 - H^{i}}$ express the degree of additional geographic concentration of the relevant industrial branch.

Ellison and Glaeser argue that there aren't savings in agglomerations where the territorial units there are equally attractive, to a certain branch. In such a situation, the gross geographic concentration is the same with the industry concentration express by the Herfindahl. The γ_{EG}^{i} index reflects the additional concentration of the branch, developed by the region's competitive margin.

Gini's location quotient

For an assessment of the overall spatial concentration of an industry compared to other industries, Krugman (1991) suggested to compute Gini's location quotient.
This method involves the following steps to determine the location quotient:

1. It determines the share of employees in a particular branch, in total employment at the national level, using the following equations: $l_i^n = \frac{z_i^n}{Z_i}$, where:

l_i^n – the share of employment in the branch *i*, of the region *n,*

z_i^n – the number of employees in industry *i*, from the region *n*,

Z_i – the number of employees in industry *i*, at the national level.

2. The regions must be descending order to ensure that: $l_i^1 \geq l_i^2 \geq \cdots \geq l_i^N$

The whole number of regions is equal with N.

3. It is necessary the cumulative share of the employees in the branch *i* and the cumulative share of the employment in the whole branch. The cumulative shares could be represented by so-called Lorenz curves. Gini's location quotient is represented by the sourface between the straight line and an angular quotient of 45° and Lorenz curve, and could be determined using the equation:
$GC = \frac{1}{2}\sum_{n=1}^{N}(u_{n-1} + u_n)g_n - \frac{1}{2}$ unde $GC \in [0; 0.5]$

The more geographically concentrated the branch of industry is, the higher the value of GC is. The maximum GC value is 0.5. On the contrary, the branch exhibiting the same spatial distribution as that of the entire industry will have a GC equal to 0.

where: $u_n = \sum_{n=1}^{N} l_i^n$ – cumulative amount of share of employees of branch *i*, in region *n*,

g_n – share of region *n* in the employment in the entire manufacturing industry.

3. Application of the methods employment data from Romania

We are going to demonstrate the usage of the presented methods on a statistical data set related to the employment, in all the Romania's counties, in 2009. The data source is the Romania's Statistical Yearbook, from 2010. In figure 1, we exemplify the results of applying the location quotient model in the mining and quarrying branch. There are 8 counties: Bacau, Mures, Hunedoara, Gorj, Valcea, Dambovita, Prahova and Teleorman, which are considered to be specialised in this branch.

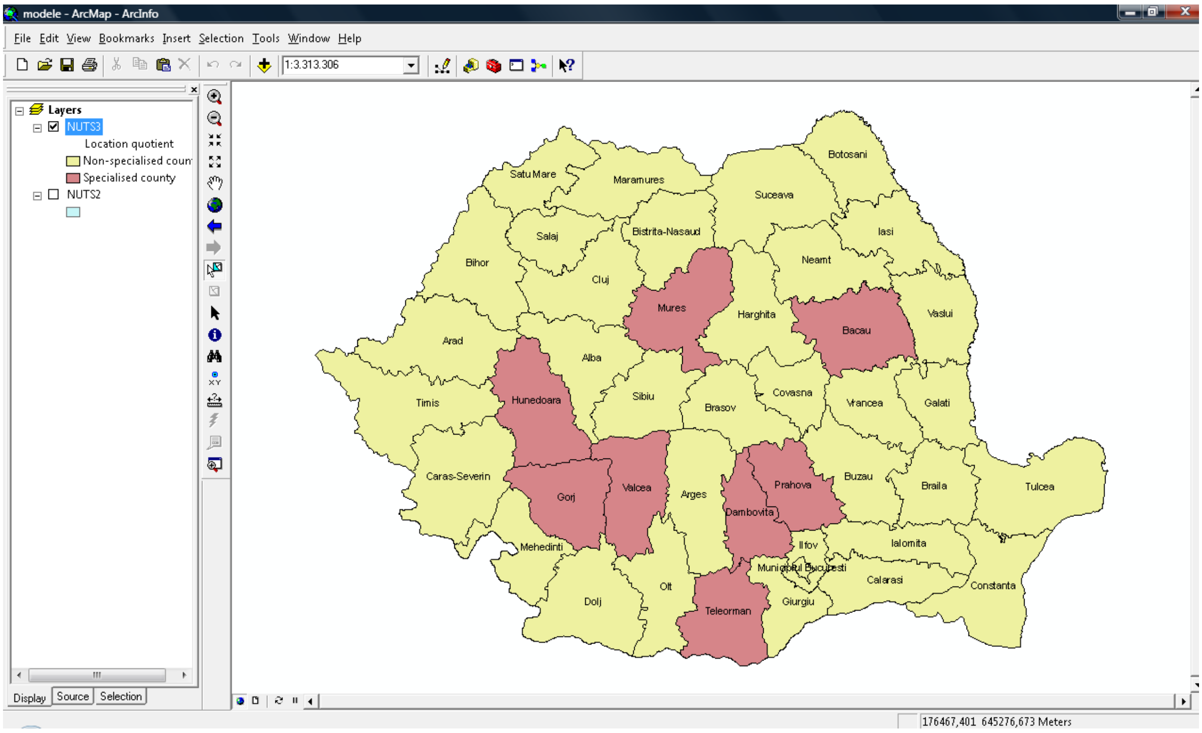


Figure 1. Location quotient method results- the mining and quarrying branch

The **shift-share analysis** used employment data from the Bucharest City and from Romania, in 2004 and 2009.

Table 1. The results of shift-share analysis applied on the Bucharest City employment data

	SS	ZN	ZR	ZP	ZO
Agriculture, forestry and fishing	0.24	1.02	-0.67	0.24	-0.11
Industry	0.77	1.02	-0.09	0.77	-0.16
Mining and quarrying	1.35	1.02	0.73	1.35	-0.40
Manufacturing	0.68	1.02	-0.15	0.68	-0.19
Electricity, gas, steam and air conditioning production and supply	0.74	1.02	0.19	0.74	-0.47
Construction	1.75	1.02	0.27	1.75	0.46
Wholesale and retail; repair of motor vehicles and motorcycles	1.35	1.02	0.15	1.35	0.18
Transport and storage	3.15	1.02	0.11	3.15	2.01
Hotels and restaurants	0.29	1.02	-0.01	0.29	-0.72
Financial intermediation and insurance	1.74	1.02	0.41	1.74	0.31
Real estate activities	0.10	1.02	-0.03	0.10	-0.89
Public administration and defence; social insurance of public sector	1.48	1.02	0.07	1.48	0.39
Education	0.93	1.02	-0.03	0.93	-0.06
Health and social assistance	1.20	1.02	0.10	1.20	0.08
Other service activities	0.59	1.02	-0.02	0.59	-0.41

The table 1 shows that the number of employees in Romania increased by 1.02% on average, during the period under consideration (ZN component). The ZO component compares the change in employment in the branch, to the average change in employment in the entire country. The most interesting is the ZR component, which compares the relative change in the number of staff in the branch and region to the relative change at the national level. The ZR component represents that section of the branch development in the region, which is explained away by regional factors, namely, the nature of local conditions for development of economic activity. From this point of view, good condition have been generated in Bucharest Region, during 2004 and 2009, for the development of the following branchs: *Mining and quarrying*, *Financial intermediation and insurance*, *Construction, Electricity, gas, steam and air conditioning production and supply, Wholesale and retail; repair of motor vehicles and motorcycles, Transport and storage,* and *Health and social assistance*.

Ellison and Glaeser's agglomeration index

The table 2 contains the values of Ellison and Glaeser's agglomeration index and the type of resulted geographial concentration, calculated based on employment statistical data available for the main branches from Romania.

Table 2. The Ellison and Glaeser's agglomeration index results

NACE	Ellison and Glaeser index value	Type of geographical concentration
Agriculture, forestry and fishing	0.01997	very weak
Industry	0.08482	Strong
Mining and quarrying	0.02170	medium-strong
Manufacturing	0.04904	medium-strong
Electricity, gas, steam and air conditioning production and supply	0.03698	medium-strong
Water supply; sewerage, waste management and decontamination activities	0.06339	Strong
Construction	0.03936	medium-strong
Wholesale and retail; repair of motor vehicles and motorcycles	0.04110	medium-strong
Transport and storage	0.05531	Strong
Hotels and restaurants	0.31619	Strong
Information and communication	0.19040	Strong
Financial intermediation and insurance	0.07051	Strong
Real estate activities	0.16910	Strong
Professional, scientific and technical activities	0.13824	Strong
Activities of administrative services and of support services	0.04569	medium-strong
Public administration and defence; social insurance of public sector	0.03474	medium-strong
Education	0.03675	medium-strong
Health and social assistance	0.10843	Strong
Shows, culture and recreation activities	0.07344	Strong
Other service activities	0.00000	very weak

Gini's location quotient

In the table 3 have been presented the Gini's location quotients. According to Gini's, the most concentrated industries are: *Information and communication*, *Financial intermediation and insurance, Professional, scientific and technical activities, Activities of* *administrative services and of support services* and *Shows, culture and recreation activities*. But the concentration level is very soft, the values being very small.

One of the main problem of the usage of Gini's location quotient is the fact that it does not control for industrial concentration. Gini's quotient consider an industry localized, if it is strongly concentrated in a limited number of geographical units.

Table 3. The Gini's location quotient results

NACE	Gini quotient
Information and communication	0.220
Financial intermediation and insurance	0.187
Professional, scientific and technical activities	0.181
Activities of administrative services and of support services	0.173
Shows, culture and recreation activities	0.165
Other service activities	0.151
Construction	0.150
Real estate activities	0.149
Hotels and restaurants	0.141
Wholesale and retail; repair of motor vehicles and motorcycles	0.138
Electricity, gas, steam and air conditioning production and supply	0.134
Transport and storage	0.134
Public administration and defence; social insurance of public sector	0.130
Health and social assistance	0.129
Education	0.127
Water supply; sewerage, waste management and decontamination activities	0.126
Industry	0.118
Manufacturing	0.118
Mining and quarrying	0.105
Agriculture, forestry and fishing	0.092

4. Conclusions

The range of quantitative methods used to identify potential clusters is very large. In this paper we tackle only a few. It is not possible to declare that a certain method is generally better or worse in comparison with other methods. The selection of a specific method depends on the type of cluster and the links between its members we seek to identify.

In practice, we find out the most widely used are the location quotients. They involve a rather undemanding method suitable for searching of local and regional clusters. The strong points of the method include the fact that the recalculations may generally use the available statistical resources. But, the location quotients cannot capture the mutual links between companies.

The shift-share analysis specify the branches that are successful in the region, in terms of the trends in employment. One disadvantage of the method is the fact that favourable results may be reached by branches where the share in overall employment in the region is largely negligible and where the region does not reveal any specialisation. On the contrary, for important branches, the results of the shift-share analysis may be ambiguous, where the given branch in the region is passing through a stage of growing at a slower speed than in the other parts of the country.

While identifying national clusters the usage of the locational Gini coefficient or Ellison and Glaeser's agglomeration index is recommended. These methods may be used to determine whether a certain branch is geographically concentrated on a national scale.

But, a simple concentration of a certain industry in the region does not necessarily mean that a cluster is present.

It is important to establish links between branches, also. That may be done by applying a broad range of methods with the objective to measure the importance of purchasing and sales flows. Based on the established links, the initial cluster map may be outlined.

References

1. Bertinelli, L. and Decrop, J. **Geographical Agglomeration: the Case of Belgian Manufacturing Industry**, in: Working Paper 14-02. Federal Planning Bureau, Brussels, 2002
2. Ellison, G. and Glaeser, E. **Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach**, Journal of Political Economy, 105, 1997
3. Enright, M. **Regional clusters and economic development: A research agenda, Business Networks: Prospects for Regional Development**, Walter de Gruyter, Berlin, 1996
4. Kim, S. **Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860-1987**, Quaterly Journal of Economics, 110, 1995
5. Knox, E.G. **Detection of clusters**, in Elliott, P. (ed.) "Methodology of enquiries into disease clustering", Small Area Health Statistics Unit, London, pp.17-20
6. Krugman, P. **Geography and trade,** MIT Press, Cambridge, 1991
7. Maskell, P. **Towards a Knowledge-based Theory of the Geographical Cluster**, Industrial and Corporate Change, 2001, pp. 919-941
8. Porter, M. **Clusters and the New Economics of Competition**, Harvard Business Review, 1998
9. Porter, M. **The Competitive Advantage of the Inner City, In: On Competition**, Boston, Harvard Business School Press, 1998
10. Roelandt, T. J.A. and den Hertog, P. **Boosting innovation; the cluster approach**, Paris, OECD, 1999

[1] **Acknowledgments**
This work was supported by the project "Performance and excellence in postdoctoral research in Romanian economics science domain", contract POSDRU/89/1.5/S/59184.

[2]Adriana REVEIU has graduated the Faculty of Cybernetics, Statistics and Economic Informatics. She holds a PhD diploma in Economic Cybernetics and Statistics. She is associate professor in Economic Informatics field and branches within Department of Economic Informatics at faculty of Cybernetics, Statistics and Economic Informatics from Academy of Economic Studies. She is the author and co-author of 10 books and over 50 articles in journal and proceedings of national and international conferences, symposiums, workshops in the fields of multimedia, communications, learning systems and data management.

[3] Marian DÂRDALĂ has graduated the Faculty of Cybernetics, Statistics and Economic Informatics. He holds a PhD diploma in Economic Cybernetics and Statistics. He is professor in Economic Informatics field and branches within Department of Economic Informatics at faculty of Cybernetics, Statistics and Economic Informatics from Academy of Economic Studies. He is the author and co-author of 15 books and over 50 articles in journal and proceedings of national and international conferences, symposiums, workshops in the fields of multimedia, human computer interaction - HCI, GIS, data bases and object oriented programming.

[4] **Codification of references in text:**

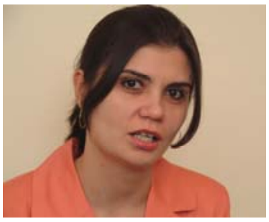
[1]	Porter, M. **Clusters and the New Economics of Competition**, Harvard Business Review, 1998
[2]	Knox, E.G. **Detection of clusters**, in Elliott, P. (ed.) "Methodology of enquiries into disease clustering", Small Area Health Statistics Unit, London, pp.17-20
[3]	Roelandt, T. J.A. and den Hertog, P. **Boosting innovation; the cluster approach**, Paris, OECD, 1999

[4]	Maskell, P. **Towards a Knowledge-based Theory of the Geographical Cluster**, Industrial and Corporate Change, 2001, pp. 919-941
[5]	Enright, M. **Regional clusters and economic development: A research agenda, Business Networks: Prospects for Regional Development**, Walter de Gruyter, Berlin, 1996
[6]	Porter, M. **The Competitive Advantage of the Inner City, In: On Competition**, Boston, Harvard Business School Press, 1998
[7]	Bertinelli, L. and Decrop, J. **Geographical Agglomeration: the Case of Belgian Manufacturing Industry**, in: Working Paper 14-02. Federal Planning Bureau, Brussels, 2002
[8]	Krugman, P. **Geography and trade,** MIT Press, Cambridge, 1991
[9]	Ellison, G. and Glaeser, E. **Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach**, Journal of Political Economy, 105, 1997
[10]	Kim, S. **Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860-1987**, Quaterly Journal of Economics, 110, 1995

SPECIALIZATION AND GEOGRAPHIC CONCENTRATION OF THE ECONOMIC ACTIVITIES IN THE ROMANIAN REGIONS[1]

Liliana Mihaela MOGA[2]
PhD, Postdoctoral Researcher,
University of Economic Studies, Bucharest, Romania
Associate Professor, Dunarea de Jos University of Galati, Romania

E-mail: liliana.moga@gmail.com



Daniela Luminita CONSTANTIN[3]
PhD, University Professor
University of Economic Studies, Bucharest, Romania

E-mail: danielaconstantin_2005@yahoo.com

Abstract: *The study of regional specialization and of concentrating the economic activities contributes to the identification of the place and role of each economic activity within the national economy and its growth potential. Thus, the possibility to emphasize the contribution brought by each economic activity to the development of each region is created. The aim of this paper is to verify relation between the evolution of the regional specialization and geographic concentration of economic activities in eight Romanian development regions. For this purpose, an empirical study of specialization and concentration was performed, at the level of the eight regions of Romania, before and after the moment of integration in the European Union.*

Key words: *specialization of economic activities; geographic concentration; regional; development; economic growth, European Union Integration*

1. Introduction

The specialty literature makes available a multitude of theories approached the problematic of regional specialization and concentration of the economic activities. From the analysis of the viewpoints expressed by the specialists of the regional development domain, the conclusion can be drawn that the definitions of regional specialization and geographic concentration of the industrial activities are based on the same production structures, reflecting the same reality (Aiginger, 1999). Through the definitions promoted Goschin et al. (2009) best highlights the correlations between the two concepts. Regional specialization expresses the territorial perspective and emphasizes the distribution of the economic activities, while the geographic concentration of an economic activity reflects its geographic distribution.

The scientific literature focused on the evolution of on location of economic activities and regional growth is not always congruent. While Aiginger (1999) supports the correlation of the analysis of regional specialization with the analysis regarding the concentration of economic activities, Dalum et al. (1998) claim that it is possible that regional specialization and the geographical concentration do not evolve in the same direction and it is probable that their evolution will happen at different speeds. A more radical viewpoint belongs to Rossi-Hansberg (2005), which states that regional specialization and geographical concentration can evolve in even different directions. Following a thorough analysis of the specialty literature, Hallet (2000) reached the conclusion that the consecrated theories are unable to offer comprehensive answers to the questions related to regional specialization, and so, it is expected that empirical studies will bring an addition of information.

Amiti (1997) conducted a study focused both on specialization and geographic concentration of the economic activities, having as purpose to determine whether specialization patterns are consistent with trade theories. The evolution of specialization and geographical concentration in European Union countries was analyzed between 1968 and 1990, using production data in current prices for 27 economic activities. He highlights that during 198 and 1990, Belgium, Denmark, Germany, Greece, Italy, and the Netherlands registered a significant increase of specialization when France, Spain and the UK registered a significant fall in specialization. There was a significant increase in specialization between 1980 and 1990 in all the studied countries. In terms of geographic concentration of industries, the study reveals that 17 out of 27 economic activities experienced an increase in geographical concentration. Six of them registered a fall in geographic concentration.

The aim of this paper is to verify the relation between the evolution of the regional specialization and geographic concentration of economic activities in eight Romanian development regions. According to Marelli's (2007) opinion, many economists agree that Krugman's hypothesis of a growing sectoral specialization is more realistic at the regional level than at the national one. The formulated reason is that, in most of the situations, the smaller the spatial units analyzed, the more specialized they are. This affirmation sustains the relevance of this study conducted at the regional level. The European Union integration would have as implications modifications in the location of the economic activities which is reflected in the evolution of the spatial concentration of the economic activities and in the regional concentration of some of them. This is why, the correlation between the evolution of regional specialization and geographic concentration is analyzed before and after the moment of integration in the European Union.

The measurement of the concentration of industrial branched and of the specialization of regions is performed by processing indicators calculates at different aggregation levels, selected depending on the aspects intended to emphasize. A complex system of indicators is developed by Hallet. He suggests the calculation of indicators for measuring concentration, clusters, centricity and the income index, on the basis of the gross added value, of the gross domestic product, and of the localization elements (Hallet, 2000).

The measurement of the concentration of industrial branches and of regions' specialization is performed by processing indicators calculate at different aggregation levels, selected depending on the aspects that the authors attempt to highlight. Thus, the authors of a study performed at the level of Romania, used the Gross Added Value (Herfindahl Index, Krugman index and the coefficient of structural changes) and the population occupation (Herfindahl Index and Krugman index), at the level of branch and region, in order to measure concentration and specialization (Goschin et al, 2009). The indicators systems that have as basis the population occupation on activities of the local economy and localization elements were developed through the study „Can Cluster Policies and Foreign Direct Investment Offer Viable Solutions to Underdeveloped Regions? Lessons that can be learnt by Romania's Eastern border regions from successful experiences of other transition countries" (Constantin et al., 2010).

In order to explore the main characteristics and the interaction between regional specialization and sectoral concentration in Romania, and to ensure relevance as high as possible for the research performed, this paper proposes a set of indicators for statistical measurement, verified at the level of the Romanian regions.

2. Methodology

Specialization and concentration could be evaluated using absolute and relative measures.

There are several indicators proposed in the existing literature. Following the review of the empirical studies, as well as the limitations due to the statistical data available at the level of Romania, a statistics was elaborated, based on the Herfindahl-Hirschman indexes and on the Krugman Dissimilarity Index. The indexes are computed for the region *i* and for the branch *j* of economic activity. The analysis of the absolute values of these indicators and their comparison to the values recorded at the national level, supply sufficient information to determine the place of each economic branch, and its ties to the other economic activities, at the level of each region in Romania, in view of determining the concentration and specialization of the economic activities.

The first statistical measure that was used within the empirical study is the Herfindahl-Hirschman Index[4], which is one of the indexes of concentration and specialization presented in the most of regional studies and assures an absolute measure. The Herfindahl-Hirschman Index is increasing with the degree of concentration or specialization, reaching its maximum of 1 when the branch of economic activity *j* is concentrated in one region or the region *i* is specialized in only one economic branch. The lowest level of concentration is reached when the ratio 1/n is the same for all regions that means they have equal shares in branch of economic activity *j*. The lowest specialization is reached when the ratio 1/m is the same for all the branches of economic activities that means they have equalled shares in region *i*. Herfindahl Index is sensitive to the number of observations, limiting direct comparisons (e.g. to countries having exactly the same number of regions).

The second indicator is Krugman Specialization Index[5]. This index was used in 1993 by Krugman, in order to compare the level of specialization between European Union and US (Marelli, 2007). The index is used to measure both concentration (K_j^C) and specialization (K_i^S). Krugman Index is a relative measure of specialization and concentration which is employed for comparing one branch of economic activity/region with the overall economy. Its values range from 0 that identifies identical territorial/sectoral structures, to 2 that characterizes totally different structures.

The statistical indicators are computed by processing the statistical information regarding the occupation of the population, on economic activities, and localization elements. The extent and the analytical character of the study are strictly determined by the data supplied by the National Statistics Institute. Thus, for the execution of this study:

- The data is collected at the national level;
- The period considered is of 15 years, between 1994 and 2009;
- The variables analyzed are connected to population occupation;
- The level of thoroughness of regional specialization was set depending on the degree of disaggregation of the statistical data, at the level of ten economic activities (Agriculture, Industry, Constructions, Commerce, Transports, storage and communications, Real estate transactions, Financial intermediations and other

services, Public administration and defense, Teaching, health and social assistance and Other activities of the national economy), for which data was supplied for the eight regions of Romania.

3. South East Region characteristics

Romania is the seventh largest among the European Union countries having almost 22 million of inhabitants. Romania formulated the request to join the European Union in 1995 and the accession negotiations begun in 2000. The accession was scheduled for 2007. Romania has experienced strong economic growth during the last years, as result of the efforts for the preparation of the access in the European Union. In spite of the positive economic evolution, the country is one of the poorest of the EU, with a GDP per capita positioned around 23 per cent of the EU-average in 2007 and 26 percent immediately after accession. The financial crisis period had as consequence the dropping back of the GDP below the level registered in 2007.

In the process of the EU accession, Romania implemented the NUTS system. It was drawn on the existing administrative territorial structure that consists in communes and towns which are grouped in counties. Once that the Law 151/1998 was adopted the territorial structure of the country was redesigned by creating a regional level, without juridical personality. The new regional focused structure was obtained by grouping the 41 Romanian counties which have some common boundaries. As result, were identified the following regions: North-West Region, North-East Region, South-East Region, South - Muntenia, Bucharest - Ilfov Region, South-West Oltenia Region and West Region. These regions are the equivalent of the NUTS II level of the European Union. The boundaries of the new regions are following the boundaries of the counties and of the Bucharest city. The reduction of interregional disparities is one of the major objectives of the regional development assumed by the Romanian governance. Supporting a balanced development and the catching-up of the better developed regions are some of the proposed solutions (Benedek & Horvath, 2008). An analysis specialization and geographic concentration of the economic activities in the Romanian regions could provide an image of the economic development of the regions which support the design of customised solution for the regional development. For this purpose, an empirical study of specialization and concentration was conducted. Its results provide information for the identification of the economic disparities between the eight Romanian regions, as it follows.

4. Specialization of economic activities at the level of Romania's regions

Analyzing the values taken by the Herfindahl-Hirschman Index in the period 1994 – 2009, presented in Appendix I, it is seen that the values have a descending trend, which signifies the fact that, at the level of the regions in Romania, the level of specialization in a certain economic activity has decreased. Comparing with the data from the period preceding the moment of Romania's accession to the European Union with the data from the period subsequent to the integration moment, respectively from 2004 to 2009, could be noticed a decrease in the level of specialization for all regions. The evolution of specialization at the level of the South-East Region falls within the general trend.

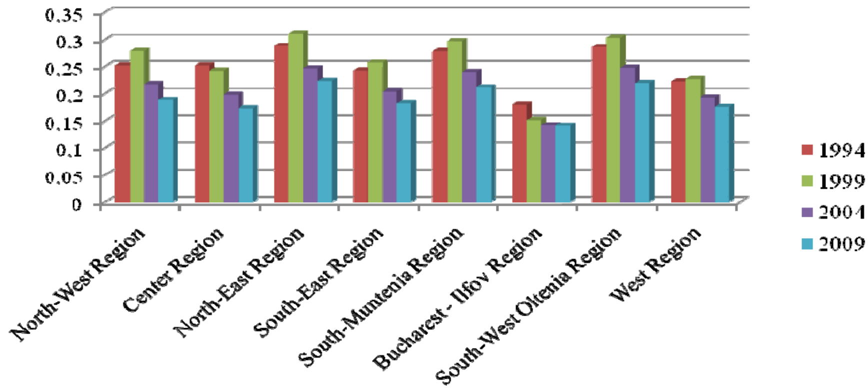


Figure 1. Statistical measures of specialization computed using employment data by Herfindahl-Hirschman Index

Throughout the entire period studies, the Bucharest - Ilfov Region remains the region with the lowest degree of specialization, while the North-East region is the region with the highest specialization degree. In year 2009, the South-East Region ranked fifth out of the total of eight regions, with an economy with a low degree of specialization. From the analysis of the economic evolution of the South-East Region, it can be seen a reduction of the industrial activities, through the decrease of the number of enterprises.

Analyzing the values calculated for the Krugman Specialization Index at the level of Romania, throughout the period analyzed, which are synthesized in Appendix II, there can be seen a tendency to reduce the values recorded by it. The South-East Region records one of the highest decreases. Analyzing the values registered after the moment of Romania's integration into the European Union, no particular trend can be identified, the evolutions going both ways. If the values recorded at the regional level in Romania are related to the EU15 average, based on regional employment data, which is below 0.150, could be observed that the majority are close to it. While the Bucharest Ilfov Region registers the highest deviations from the European average, the South-East Region constantly oscillates around it. Krugman Specialization Index shows that seven Romanian regions have a structure of economic activities performed within them which is close to that characterizing Romania at the country level.

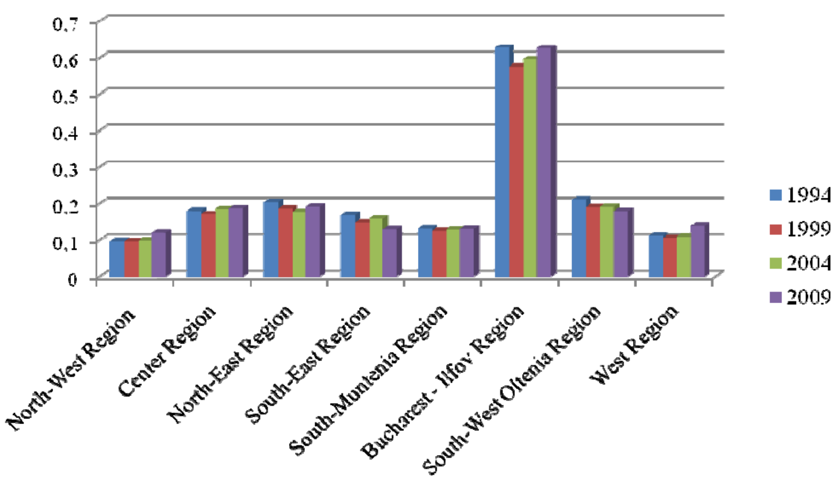


Figure 2. Statistical measures of specialization computed using employment data by Krugman Index

Geographic concentration of economic activities of Romania's regions

Extending the idea according that comparative advantage sustain nations tendency to become more specialized in sectors in which they have a comparative advantage, to the regional level of Romania, it is expected to find some Romanian regions specialized in some distinct economic activities. Analyzing the values in Appendix III, it is seen that, only after the moment of Romania's integration into the European Union, economic activities such as Real estate transactions, Financial intermediations, Transports and Constructions, register an increase of concentration.

An Index computed for concentration shows lower values than the specialization index and little variation in respect to the data employed. Analyzing the values taken by the Herfindahl-Hirschman Index in the period studied, presented in Table 3, it is seen that the values recorded by Agriculture, Transports, storage and communications, Real estate transactions, Financial intermediations and other services and Public administration and defense have a strong ascending trend. A positive evolution, marked also by decreases during certain periods of time, has constructions and commerce, as well as the activities generically grouped into the "Other" activities of the national economy.

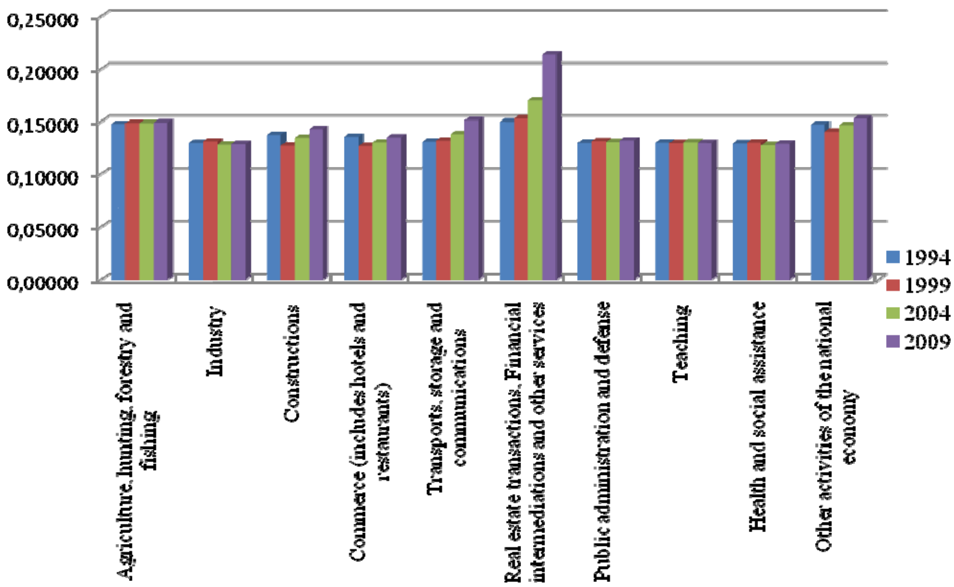


Figure 3. Statistical measures of concentration computed using employment data by Herfindahl-Hirschman Index

According to Morelli (2007), the analysis of the Krugman Specialization Index reveals a prevalent decreasing specialization across the European countries and regions. This result characterizes both old and the new European Countries. Despite an initially high heterogeneity, new integrated countries regions are becoming progressively more similar to the old Europe. A partial exception is given by the Polish regions. Analyzing the value registered by the Krugman Specialization Index for Romania, synthesized in Appendix IV, in the post-accession period one can see, still, slight increases of specialization, more evident for Real estate transactions, Financial intermediations and other services, Transports, storage and communications and Commerce. At the opposite pole are the social services.

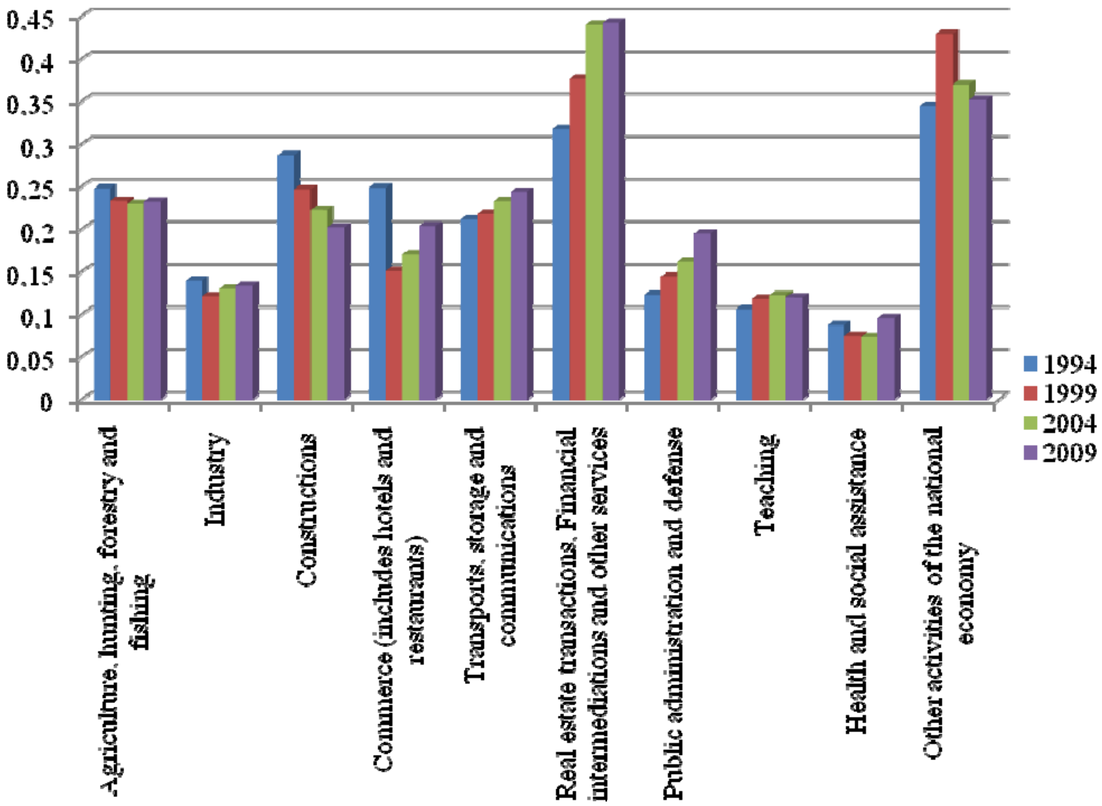


Figure 4. Statistical measures of concentration computed using employment data by Krugman Specialization Index

Conclusions

At the European Union level, through the Economic and Social Cohesion Politics, a balanced development is promoted, reducing regional disparities. As a result of this policy, Marrelli (2007) stated that at the country level there could be observed a convergence across countries. An opposite trend is exhibited by the EU10 group of new members; in particular, increasing regional disparities characterize the first stages of growth of individual countries.

From the analysis of the results obtained following the analysis of specialization and concentration at the level of the eight regions, it is seen that the level of Romania there are no major disparities, and the values of the indexes by means of which specialization and concentration are quantified are close to the averages recorded at the European Union level. The information providing by the existing studies that correlate the moment of Romania's integration into the European Union do not provide the framework for a comparison. This situation is because of the different NACE classifications used in the different studies for the computation of the statistical indicators and because of the periods of the studies which are not convergent. At the moment there is not available a study conducted at the EU 25 countries level that provides information from the pre-ascension and post-ascension periods of EU10 countries. On the basis of the concentration indices calculated for manufacturing branches in Bulgaria, Estonia, Hungary, Romania and Slovenia were grouped the industries according to the following characteristics: scale economies, technology level, and wages level (Traistaru et al., 2002). The manufacturing classification is according to the Eurostat NACE Rev1 (2 digit classification) for Estonia, Romania, and Slovenia. Employment data have been collected according to national classifications in Hungary and Bulgaria. For these two latter cases aggregations have been made to bring these classifications as close as possible to the NACE classification.

Analyzing the Herfindahl-Hirschman synthesized in the Appendix I, in the period 1994 – 2009 the descendent trend of regional specialization that is seen at the level of all the regions analyzed was interrupted in 1999 for most of the regions. This relative index indicates an continuous evolution of specialization only in the Center Region and in the Bucharest Ilfov Region. Concerning the results registered for the Krugman Index that are synthesized in Appendix II, computed for the same period, show a fluctuant evolution of regional specialization for all the Romanian regions. In what concerns the evolution of concentration, the data in Appendix III and Appendix IV shows that there is no constant trend, each separate activity having a specific evolution, with increases and decreases which cannot be classified within a particular tendency. This proves that in Romania, for a short period of time, which includes stages of economic development specific to passing to the market economy, the pre-accession and post-accession period, regional specialization and the geographical concentration do not evolve in the same direction which verifies the hypothesis postulated by Dalum et al. (1998).

If the analysis is focused solely on the interval 2004 - 2009, which symmetrically covers, both the pre-accession and the post-accession periods of Romania, the values registered by the Herfindahl-Hirschman Index maintain the descended trend recorded at the level of all the regions analyzed, while the Krugman relative index reflects a fluctuant evolution of the specialization. In what concerns the evolution of the geographical concentration, most economic activities analyzed present an ascending trend, even they are measured using absolute or relative indexes. The evolution of the two indexes for short time, which captures the two distinct moments in the integration evolution, verifies the radical hypothesis of Rossi-Hansberg (2005), which states that regional specialization and geographical concentration can evolve in even different directions.

References

1. Aiginger. K. **Do industrial structures converge? A survey on the empirical literature on specialization and concentration of industries,** WIFO-Working papers, 1999
2. Amiti, M. **Specialisation patterns in Europe,** CEPDP, 363. Centre for Economic Performance, London School of Economics and Political Science, London, UK, 1997
3. Benedek, J. and Horvath, R. **Chapter 12, Romania,** in: Baun, M, and Marek, D., (eds.) "EU regional policy after enlargement", Palgrave Macmillan, Basingstoke, pp. 226 – 247, 2008
4. Constantin, D.L., Bodea, C.N., Pauna, C.B., Goschin, Z., Dragusin, M., Stancu, I., and Popescu, O. **Can Cluster Policies and Foreign Direct Investment Offer Viable Solutions To Underdeveloped Regions? Lessons that can be learnt by Romania's Eastern border regions from successful experiences of other transition countries,** 2010 retrieved December 10, 2010, from http://www.cerge-ei.cz/pdf/gdn/RRCIX_39_paper_01.pdf

5. Goschin, Z., Constantin, D.L., Roman, M. and Ileanu, B. **Regional Specialisation and Geographic Concentration of Industries in Romania,** South-Eastern Europe Journal of Economics, Vol.7, No.1, 2009, pp. 61-76
6. Hallet, M. **Regional Specialization and Concentration in the EU. Economic papers,** 141, March, 2000
7. Krugman, P. **Lessons of Massachusetts for EMU,** in: Torres, F. and Giavazzi, F. (eds.), "Adjustment and Growth in the European Monetary Union", Cambridge University Press, Cambridge, 1993
8. Marelli, E. **Specialization and Convergence of European Regions,** The European Journal of Comparative Economics, no 2, 2006
9. Marelli, E. **The Integration Process of the European Regions, European Association for Comparative Economics Studies (EACES),** Paper presented at "9th Bi-Annual Conference: Development Strategies - A Comparative View", 2007
10. Molle, W. **The Regional Economic Structure of the European Union: an Analysis of Long-Term Developments,** Karin Peschel (ed.), Physica-Verlag: Heidelberg, 1996
11. Rossi-Hansberg, E. and Wright, L.J. **Urban Structure and Growth,** NBER Working Papers 11262, National Bureau of Economic Research, Inc., 2005
12. Traistaru, I., Nijkamp, P. and Longhi, S. **Regional Specialization And Concentration Of Industrial Activity In Accession Countries,** Working paper, Center for European Integration Studies, 2002

- * * * **Development strategy of the South-East Development Region,** 2010, retrieved January 20, 2011 from http://www.adrse.ro
- * * * **Regional Development Plan 2007-2013 of the South-East Development Region,** 2009, retrieved January 20, 2011, from http://www.adrse.ro
- * * * **Statistical yearbook. (1992-2009),** Statistical indicators for the occupation of the work force at the regional level and detailed for the South-East Region, Bucharest: National Institute of Statistics
- * * * **Territorial Audit of the South-East Region,** retrieved January 20, 2011, from http://www.adrse.ro
- * * * **The GDP of Romania,** January 23, 2010, retrieved March 20, 2011, from http://www.romania-central.com/the-gdp-of-romania/

---

[1] **Acknowledgements**: This work was co-financed from the European Social Fund through the Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/1.5/S/59184 „Performance and excellence in postdoctoral research in Romanian economics science domain".

[2]Liliana Mihaela Moga is Associate Professor in Economic Analysis and Financial Information Systems at Dunarea de Jos University of Galati and Postdoctoral Researcher in Regional Development at Bucharest Academy of Economic Studies. She is involved in numerous research projects. Also, she published numerous articles and books in Romania and abroad.

[3]Daniela-Luminita Constantin is Professor of Regional Economics at the Academy of Economic Studies of Bucharest and President of the Romanian Regional Science Association. She carried out numerous research stages abroad as Fulbright, JSPS, DAAD scholar and authored a large number of books and articles published in Romania and abroad.

[4] $H_j^C = \sum_{i=1}^{n}(g_{ij}^C)^2$ and $H_i^S = \sum_{j=1}^{m}(g_{ij}^S)^2$, where: $g_{ij}^C = \frac{X_{ij}}{\sum_{i=1}^{n} X_{ij}} = \frac{X_{ij}}{X_j}$ and $g_{ij}^S = \frac{X_{ij}}{\sum_{j=1}^{m} X_{ij}} = \frac{X_{ij}}{X_i}$

Explanations: i represents the region; j represents the branch of economic activity; X represents the employment; X_{ij} represents the employment in the branch of economic activity *j* in region *i*; X_j represents the employment in the branch of economic activity *j*; X_i represents the employment in region i; g_{ij}^C represents the share of region *i* in the total national value of the branch of economic activity *j*; g_{ij}^S represents the share of the branch of economic activity *j* in the total value of region *i*.

5 $K_j^C = \sum_{i=1}^{n} \left| g_{ij}^C - g_i \right|$ and $K_i^S = \sum_{j=1}^{m} \left| g_{ij}^S - g_j \right|$, where: $g_i = \frac{X_i}{X}$ and $g_j = \frac{X_j}{X}$

Appendix I: Statistical measures of specialization computed using employment data by Herfindahl-Hirschman Index

Year	1994	1999	2004	2009
North-West Region	0.25152	0.27795	0.21810	0.18786
Center Region	0.25150	0.24234	0.19776	0.17287
North-East Region	0.28845	0.31123	0.24645	0.22369
South-East Region	0.24289	0.25719	0.20322	0.18180
South-Muntenia Region	0.27780	0.29707	0.23999	0.21200
Bucharest - Ilfov Region	0.17919	0.15149	0.14191	0.14111
South-West Oltenia Region	0.28613	0.30307	0.24761	0.22022
West Region	0.22308	0.22713	0.19277	0.17549

Appendix II: Statistical measures of specialization computed using employment data by Krugman Index

Year	1994	1999	2004	2009
North-West Region	0.09605	0.09573	0.09779	0.11953
Center Region	0.18007	0.17109	0.18409	0.18668
North-East Region	0.20299	0.18572	0.17587	0.19103
South-East Region	0.16826	0.14735	0.15805	0.12934
South-Muntenia Region	0.13115	0.12466	0.12818	0.13013
Bucharest - Ilfov Region	0.62683	0.57467	0.59473	0.62537
South-West Oltenia Region	0.21025	0.18966	0.19030	0.17979
West Region	0.11185	0.10476	0.10942	0.13867

Appendix III: Statistical measures of concentration computed using employment data by Herfindahl-Hirschman Index

Year	1994	1999	2004	2009
Agriculture, hunting, forestry and fishing	0.14774	0.14865	0.14841	0.14905
Industry	0.12935	0.13083	0.12790	0.12829
Constructions	0.13722	0.12687	0.13446	0.14295
Commerce (includes hotels and restaurants)	0.13538	0.12655	0.12990	0.13474
Transports, storage and communications	0.13060	0.13153	0.13783	0.15129
Real estate transactions, Financial intermediations and other services	0.14981	0.15282	0.17033	0.21351
Public administration and defense	0.12952	0.13104	0.13025	0.13177
Teaching	0.12965	0.12931	0.13002	0.12934
Health and social assistance	0.12897	0.12978	0.12741	0.12864
Other activities of the national economy	0.14739	0.14028	0.14655	0.15273

Appendix IV: Statistical measures of concentration computed using employment data by Krugman Specialization Index

Year	1994	1999	2004	2009
Agriculture, hunting, forestry and fishing	0.24786	0.23293	0.22913	0.23169
Industry	0.13962	0.12130	0.12975	0.13337
Constructions	0.28678	0.24670	0.22276	0.20208
Commerce (includes hotels and restaurants)	0.24853	0.15068	0.16990	0.20377
Transports, storage and communications	0.21135	0.21797	0.23230	0.24404
Real estate transactions, Financial intermediations and other services	0.31753	0.37638	0.44016	0.44242
Public administration and defense	0.12287	0.14447	0.16225	0.19489
Teaching	0.10607	0.11837	0.12247	0.11983
Health and social assistance	0.08704	0,07425	0.07317	0.09538
Other activities of the national economy	0.34493	0.42946	0.36978	0.35151

THE VILLAGES' DEVELOPMENT LEVEL FROM DOBROGEA REGION[1]

Claudiu HERTELIU
PhD, Assistant Professor, Department of Statistics and Econometrics,
University of Economic Studies, Bucharest, Romania

E-mail: claudiu.herteliu@gmail.com, **Web page:** http://www.hertz.ase.ro

Adrian COSTEA
PhD, Associate Professor
University of Economic Studies, Bucharest, Romania

E-mail: acostea74@yahoo.com **Web page:** http://www.abo.fi/~acostea

Abstract: *This article intends to analyze the position of the villages in the historical region of Dobrogea from the development point of view. With the help of a methodology similar to Human Development Index (HDI) used by the UN, in Romania, in 2009 an inter-institution project (Sandu et al) elaborated the Village Development Index (VDI).*
Based on this statistic information it can be ascertained that the villages in Dobrogea have a development level superior to the national average. We then built some econometric models to establish the influence factors such as: ethnicity, coast area, delta area, which may have an influence on the development level.

Key words: *villages' development index; VDI; rural; Dobrogea; Romania*

General context

Dobrogea is a historical region well documented even since the oldest times. The ancient name of this region was Scythia Minor (visible in the map presented in Figure 1). With a tumultuous history starting with the roman domination (byzantine afterwards), continuing with brief periods of independence, periods with extended autonomy (from the Ottoman Empire), Bulgarian control, very short periods of Romanian domination(1388-289 and 1599-1601). After the Russian-Turkish war, after which the Romanian Kingdom regained its independence, the Berlin Congress established that the Northern part of Dobrogea be given to Romania, while the Southern[2] part to Bulgaria.

After the Peace Treaty in Bucharest (1913) which followed to the Second Balkan War, Romania takes South Dobrogea[3] which it keeps discontinuously[4] until 1940 when following the Treaty in Craiova, this territory goes under Bulgarian administration. In this article, we refer to the territory of Dobrogea[5] as a component part of Romania. This territory is composed of two counties: Constanta (in the Southern part) and Tulcea (in the Northern part).

Figure 1. Macedonia, Thracia, Illyria, Moesia and Dacia
Source of map: http://soltdm.com/geo/pubmaps/1.htm

Methodology and data sources

The statistic information at the base of this article is mainly taken from the results of the Population and households' census in 2002 (relative to the size of the villages and the structure of the population from the ethnical standpoint). A second data set, regarding the villages' development level, is the one supplied to readers and researchers interested by this field by Professor Dumitru Sandu – the initiator of the process (Sandu et al, 2009) of inter-institutional cooperation (NIS, Bucharest Univerity, Ministry of Finance, Ministry of Administration and Internal Affairs). The analyses are mainly descriptive without neglecting statistical validations with specific tests or building econometric models. From the software point of view, for the analysis of the data we used SPSS while the maps were elaborated using ArcGIS.

Data Analysis

The state of the facts shows that the villages of the two counties of Dobrogea (as can be seen in figure 2) are at a higher level relative to the national average. Taking into account that the development quintiles were built at national level, if Dobrogea had a status similar to the national one, all five categories should contain a equal number of villages (one fifth). Most of the villages in Dobrogea (32.3%) have a high level of development while they are quickly followed by the villages in the median area (25%). The other quintiles have weights as follows: 17.7% lower development, 14.6% the highest development and 10.4% the lowest development.

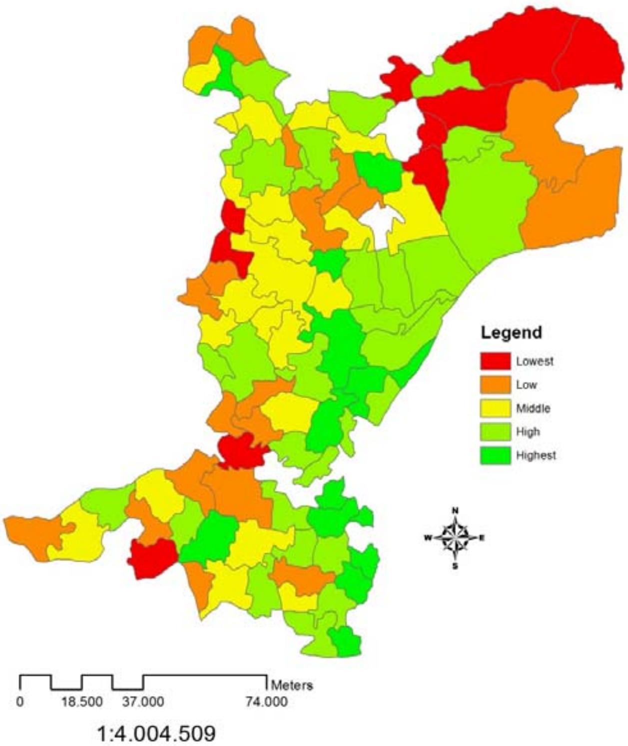


Figure 2. Villages' development level in Dobrogea region
(by quintiles designed at national level)

From the figure we can see that, in general, the villages found in the area of the Black Sea coast are found in the superior part of the classification, while the less developed villages are in the Danube Delta.

In figure 3 we present the distribution of the villages in Dobrogea by development level and number of ethnical groups present. We can state that there is a direct proportional connection between ethnic diversity and development level in the village.

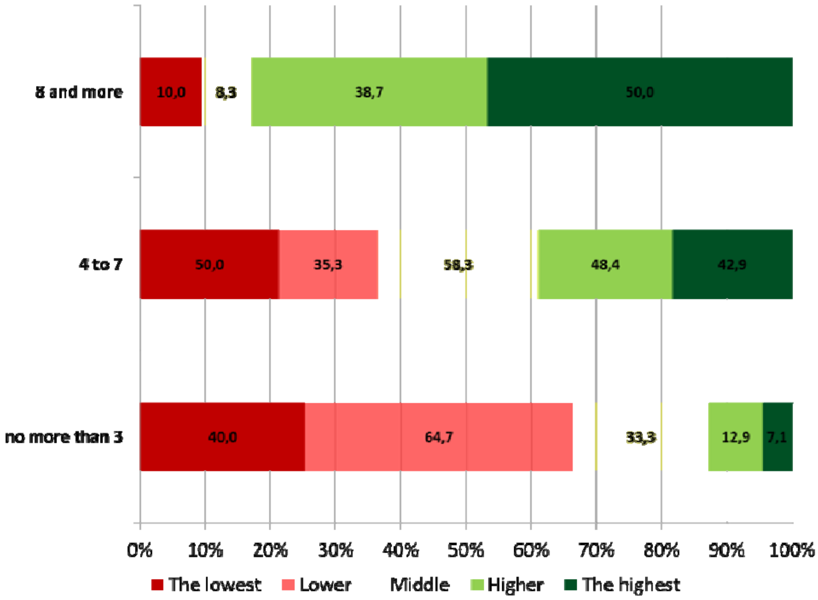


Figure 3. Villages' development level by number of ethnic groups

This directly proportional connection is statistically significant (after running the χ^2 test) with a probability higher than 99.99%. The Pearson's coefficient φ had a level of 0.55 while Crammer's V was 0.39. These values show a strong connection between the two variables.

The next step in our analysis was to build a regression model which highlights the factors which influence the villages' development level. We used the following elements as independent variables: the number of ethnic groups in a village, the weight of Turkish ethnics with two dummy variables which signal if the respective commune is in the Black Sea coast or in the Danube Delta. We chose a linear regression model and the results of parameters' estimation are presented in Table 1.

Table 1. Regression summary

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	42,919	2,215		19,373	,000
	Number of ethnic groups	1,474	,369	,329	3,989	,000
	Turks share	-28,335	17,953	-,126	-1,578	,118
	Coast village?	17,172	3,393	,421	5,061	,000
	Delta village?	-7,917	3,280	-,194	-2,413	,018

a. Dependent Variable: VDI - Village's Development Index

The model is valid (after running the F test) with a probability higher than 99.99% and explains the dependent variable to a degree of 42.7%. Except for the weight of the Turkish ethnics (which can be kept in the model with a probability of maximum 88.2%) all the other variables are statistically significant. As was shown in the one to one analysis (figure 3) ethnical diversity has a favorable impact on the dependent variable. The same direction is followed by the dummy variable which signals the villages placed in the Black Sea coast. An unfavorable influence is registered for the placing of the village in the Danube Delta area along with the weight of Turkish ethnics[6].

Conclusions

In conclusion, we can state that although the villages in Dobrogea are situated above the national level, the situation is not uniform. Thus, the villages found on the Black Sea coast or the ones with a stronger ethnical diversity have a better placement from the development level point of view. At the opposite end are the villages in the Danube Delta or the ones with a significant Turkish minority.

References

1. Ailenei, D. and Mosora, L.C. **Economics of Sustainable Development. Competitiveness and Economic Growth,** ECTAP, Vol. 18, no. 2/2011, pp. 5-12

2. Constantin, D. L., Goschin, Z. şi Drăguşin, M. **Ethnic Entrepreneurship as an Integrating Factor in Civil Society and a Gate to Religious Tolerance: A Spotlight on Turkish Entrepreneurs in Romania**, Journal for the Study of Religions and Ideologies, 7, 20, Summer 2008, pp. 49-79
3. Frunza, M. and Frunza, S. **Etica, superstitie si laicizarea spatiului public,** Journal for the Study of Religions and Ideologies, nr. 23, Summer 2009, pp. 13-35
4. Isaic-Maniu, Al. and Herteliu, C. **Ethnic and Religious Groups in Romania – Educational (Co)Incidences,** Journal for the Study of Religions and Ideologies, No. 12, Winter 2005, pp. 68-75
5. Rotariu, T (coord.), Bădescu, G., Culic, I., Mezei, E., Mureşan, C., **Metode statistice aplicate în ştiinţele sociale,** Ed. Polirom, Iaşi, 2000
6. Sandu, D. **IDC Database,** https://sites.google.com/site/dumitrusandu/bazededate, accessed in multiple occasions in September 2012
7. Sandu, D., Voineagu, V. and Panduru, F. **Dezvoltarea comunelor din Romania,** INS, SAS, July 2009
8. Smidt, C.E., Kellstedt, L.A. and Guth, J.L. (eds.) **The Oxford Handbook of Religion**
9. **and American Politics,** Oxford University Press, New York, 2009, pp. 3-42
10. * * * http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/local_administrative_units
11. * * * http://soltdm.com/geo/pubmaps/1.htm
12. * * * **Recensământul populaţiei şi al locuinţelor 18 martie 2002 – vol. IV,** Institutul Naţional de Statistică, Bucharest, 2004

---

[1] **Acknowledgements:** This work was co-financed from the European Social Fund through the Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/1.5/S/59184 „Performance and excellence in postdoctoral research in Romanian economics science domain".

[2] In the south of Danube and from Silitra, in a straight line up to the village of fishermen Ofidaki (today's Vama Veche)

[3] Provence also known under the name "Quadrialater"

[4] During the First World War, in 1969 Bulgaria (along with the German allies) reconquers the Quadrilater as well as an important part of Northern Dobrogea which it administers for two years until the ending of the Neuilly sur Seine Peace Treaty in 1919.

[5] Also called North Dobrogea

[6] This negative impact is probably given also by the poor instruction level registered by the feminine population in the Turkish ethnic group (see Isaic-Maniu and Herteliu, 2005)

MEASURING PASSENGER SATISFACTION: A STRATEGY BASED ON RASCH ANALYSIS AND THE ANOM

Michele GALLO
Department of Human and Social Science,
University of Naples L'Orientale, Naples, Italy

E-mail: mgallo@unior.it

Abstract: *Measuring passenger satisfaction presents several difficulties since customer satisfaction in the public transport sector is subject to different conditions which are different than those that affect other sectors. In this work, a strategy based on Rasch analysis and the Analysis of Means (ANOM) is proposed. This study is based on the idea that the Rasch rating scale model gives 'sufficient statistic' for an underlying unidimensional latent trait such as the satisfaction generated by local transport operators. Furthermore, the ability of passengers, measured by the rating scale model, is studied by means of ANOM decision charts to verify if there are different levels of satisfaction between the different groups of passengers.*

Key words: *customer satisfaction; rating scale model; analysis of means; metro system*

1. Introduction

Nowadays, people are more mobile and expect efficient, high quality public transportation services. In order to meet the increasing mobility demand, public transport companies have to tailor their services they supply to the wants and needs of their current or potential customers. An important source of information on quality assurance is the customer satisfaction survey, where customer satisfaction in the public transport sector is subject to conditions different than those concerning other sectors. In fact, satisfaction is not the only factor influencing the users' behavior since it is also influenced by a range of other factors, such as the accessibility to a certain model in a certain situation. Moreover, when local transport is considered to be the freedom of an individual to choose from different means of transportation (public or private), it is presumed, and the customer satisfaction becomes a vital concern for companies and organizations in their efforts to improve service quality, and retain the passenger's loyalty.

In the last three decades, more conceptual customer satisfaction models have been proposed in statistical literature. Customer satisfaction is a result of a latent complex information process summarized in a multiple-items questionnaire, in which one set of alternative responses are used for estimating probabilities of responses. For this reason, the analysis of multi-item data should be considered as the multidimensional nature of customer satisfaction and the different nature of the data (Gallo, 2007). However, the multi-item scale needs to measure only customer satisfaction. No more attributes or behavior need to be measured together but only one latent variable.

Many latent trait models could be used to measure customer satisfaction, but the Rasch models are distinguished from others by a fundamental statistical characteristic, viz., the subject's sum score is a 'sufficient statistic' for the underlying unidimensional latent trait (Wright and Linacre, 1989). The model is based on the simple idea that the passengers who have a high total score on an item are more satisfied than those passengers with low scores. Likewise, items that receive lower ratings are more difficult to endorse than items that receive higher ratings. This way, on a single continuum of interest, it is possible to clearly identify the items which are more difficult to generate satisfaction and the passengers who are more satisfied than others.

Generally, a customer satisfaction survey should be designed to collect the data in a less intrusive and idiosyncratic way, as much as possible. In a public transport sector a good way to submit a customer satisfaction questionnaire is on the platform or on the train. This is only possible when the questionnaire has not many items to measure customer satisfaction and a few additional items regarding the characteristics of passengers are given. These latter items could be used to identify the different levels of satisfaction within various groups of passengers.

To items ('station cleanness', 'train cleanness', 'passenger comfort', 'regularity of service', 'frequency of service', 'staff behavior', 'passenger information', 'safety', 'personal and financial security', 'escalators and elevators working') are used to measure the passenger satisfaction, where each item has four different levels (Likert scale). Other items (sex, age, profession, purpose of travel, day of interview, number of travel frequency in a week, intermodal transportation service used) would give additional information on the passengers.

The purpose of this work is to determine whether the questionnaire used is adequate to give a measure along the continuum of the underlying passenger satisfaction. Therefore, rating scale model is applied to improve the measurement tool. When a valid measure of passenger satisfaction is given, a graphical procedure like the Analysis of Means (ANOM) is used in order to understand the different levels of satisfaction between different groups of passengers.

2. Theory

2.1. Rating scale model

When all items present the same set of alternatives, it seems reasonable to expect that the relative difficulties of the steps between categories will not vary from item to item. For these kinds of questionnaires the rating scale (Andrich, 1978; Wright and Masters, 1982) is the more appropriate version of the Rasch models.

Rating scale model - within a probabilistic framework - converts ordinal raw-score data into an interval-based measure, the log-odd metric or logit. Let $P_{ij(m)}$ be passenger *i*'s probability of scoring *m* on item *j*, the rating scale model can be written as:

$$P_{ij(m)} = \frac{\exp\left(\xi_i - \delta_j - \gamma_m\right)}{\left[1 + \exp\left(\xi_i - \delta_j - \gamma_m\right)\right]} \tag{1}$$

where δ_j is the difficulty for item *j* to generate satisfaction, ξ_i is the attitude of *i*th passenger to be satisfied, and γ_m is the threshold parameter associated with the transition between response categories $m-1$ to m.

The logits measures are given by $\ln\left(P_{ij(m)}/\left(1-P_{ij(m)}\right)\right)$. For passenger, the logit indicates whether one passenger is more able than another to get satisfaction. For item, logit measures indicate whether one item is more difficult than another to generate satisfaction. And for rating scale categories, logit measures indicate whether one rating scale category is greater or less than another in degree (for example: does the 'satisfaction' category represents less satisfaction than the 'strong satisfaction' category).

This method is more flexible and it is independent from specific passenger and item distributional forms. Moreover the logit measure $\ln\left(P_{ij(m)}/\left(1-P_{ij(m)}\right)\right)$, of the items, passengers and rating scale categories, convert ordinal raw scores into linear interval measures. When the diagnostic analysis assures that the measures of passenger satisfaction are valid and reliable, they can be employed in a model that needs linear and normal distributed data like ANOM.

2.2. Analysis of Means

The phrase "analysis of means" was used for the first time by Ott (1967). And based on Bonferroni inequalities, he proposed ANOM as a multiple comparison procedure that could be used instead of, or as a follow up to, analysis of variance (ANOVA). However, after 1982 exact critical value for the main effects of ANOM in balanced designs were obtained (Nelson, 1982). Nowadays ANOM is proposed in many cases for experimental or non experimental data related to normally, binomial and Poisson distributed data (Nelson et al., 2005). In this paper, ANOM is useful when the desired outcome is to identify differences between groups and, in case of observational data, when a different number of observations is generally given for each group (one-factor unbalanced ANOM).

Let n_k be the number of observations into group *k* (*k* = 1,…,*K*) with μ_k being the mean for a *k*th group, the hypothesis to test is $H_0 : \mu_1 = \ldots = \mu_k = \ldots = \mu_K$ versus the alternative one that is different. Similar to the ANOVA, ANOM tests whether there are differences among the groups, but dissimilar to analysis of variance, when there are differences, it also indicates how groups differ by a decision chart.

If data is least approximately normally distributed and all the different groups have the same variance for obtaining upper and lower decision lines, the sample means

$$\bar{y}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} y_{ik}$$

and sample variances $s_k^2 = \dfrac{1}{n_k - 1}\sum_{i=1}^{n_k}\left(y_{ik} - \bar{y}_k\right)^2$ of each group can be used.

ANOM procedures for studies with unequal samples require to consider the decision lines

$$\bar{y} \pm cv\left(\alpha, K, n-K\right)\sqrt{MS_e}\sqrt{\frac{n-n_k}{nn_k}} \qquad (2)$$

where $\bar{y} = \frac{1}{K}\sum_{k=1}^{K}\bar{y}_k$ is the overall mean, $MS_e = \frac{1}{K}\sum_{k=1}^{K}s_k^2$ is the mean square error, and $cv(\alpha, K, n-K)$ is a critical value that depends on the level of significance desired α, the number of groups K, the degree of freedom for MS_e.

When the sample means for each groups are plotted between the decision lines given by (2) then there are not differences between groups on the level of significance α. Full theory behind ANOM (as multivariate negatively correlated singular *t* distribution, power curve etc.) was showed by Nelson (1985).

3. Application

MetroNapoli S.p.A., which manages all of the rail transport in the city of Naples, conducted a survey analysis on 2,473 passengers, according to stratified random sampling, to measure the passenger satisfaction of medium sized metro systems. The questionnaires were submitted by 10 different interviewers in the second week of October. Each item had four ordinal scales viz., 'strong dissatisfaction' / 'dissatisfaction' / 'satisfaction' / 'strong satisfaction'.

The analysis consisted of two parts. Firstly, as stated above, Rasch analysis focuses on the psychometric properties of the items, passengers, and rating scale categories. The WINSTEPS program (Linacre and Wright, 2000) was used in order to obtain Rasch measurements from these data. Secondly, with the goal of investigating whether the passenger's satisfaction was influenced by personal characteristics (profession, age, sex, purpose of travel), the logit measure of the passenger satisfaction were used into ANOM analysis. Moreover, other aspects of service (the day on which interview took place, weekly travel frequency, intermodal transportation service) were investigated. The ANOM was done by a simple routine developed in R software, but it is included as a standard option in more statistical software (included SPSS, SAS and MINITAB).

3.1. Rating scale model results

The WINSTEPS computer program was used to perform the rating scale model and the compatibility of the raw data with the Rasch measurement model, which was verified by more fit statistics. A statistics summary of the WINSTEPS program shows the reliability and separation for items and passengers. In this case, the reliability index observed is 0.99 for items and 0.78 for passengers, where the values range between 0 and 1. The estimates for items show the replicability placement of items across other passengers measuring the same construct index. Separation index, that is alternative to reliability, estimates the spread of items (15.15) and the spread of passengers (1.89) on the underlying latent trait.

The results for the rating scale analysis of the passenger satisfaction are shown in Figure 1. The vertical line represents the variable passenger satisfaction into a log-odds scale. Passengers are aligned to the left and represented by "#". The more satisfied are at the top. Items are aligned to the right, and the more difficult to generate satisfaction are at the top. It is verified that the distribution of passenger is normal and displayed in a higher position than item distribution. Therefore, passengers have more probability to get satisfaction from MetroNapolis' service.

```
        Persons - MAP - Items
                 <more>|<rare>
    5         .######  +
                       |
                       |
                       |
                       |
                   .#  |
    4                  +
                       |
                       |
                       |
                  .##  |
                    .  |
    3               . T+
                  .##  |
                       |
                    .  |
                 .###  |
                    .  |
    2            .###  +
                      S|
                .####  |
                       |
                .####  |
                .####  |
    1               .  +
        .############  |T
              .###### M|  Passenger information
                    .  |  Staff behavior
             .#######  |S
             .#######  |  Passenger comfort           Train cleanness
    0         .######  +M Personal and financial sec  Station cleanness
                .####  |
                       |S Frequency of service        Regularity of service
                 .### S|
                  .##  |  Security
                   ##  |T
   -1              .#  +
                   .#  |
                       |
                    .  |
                    . T|
                       |
   -2               .  +
                    .  |
                       |
                    .  |
                       |
                       |
   -3               .  +
                       |
                       |
                       |
                       |
                       |
   -4               .  +
                 <less>|<frequ>
```

Figure 1. Person-item map for passenger satisfaction

Note: Each "#" is 27 passengers, which are aligned to the left of the corresponding log-odds measure of satisfaction. The items are aligned to the right of the corresponding log-odds measure of difficulty to generate satisfaction.

More details for item measure are given in Tables 1. This table lists items in order of measure. 'Passenger information' is the attribute of service that has more difficulty to generate satisfaction followed by 'Staff behavior' and 'Train cleanliness'. The attributes that have less difficulty to generate satisfaction are 'Security' and 'Regularity of service'. Two types of fit statistics are given for each item. Ideally, for rating scale model the infit and outfit mean-square will be 1.0, but values included between 0.6 and 1.4 indicate that the deviation from expectation is acceptable (Bond e Fox, 2001). In particular, the 1.16 infit mean-square statistic for the item 'Passenger information' is the highest variation between observed data and the Rasch model predicted (16% more variation). The 'Train cleanliness' and 'Station cleanliness' items have 18% less variation in the observed response than the value that had been modeled. Similarly, outfit mean-square for the item 'Passenger information' has the highest variation (20%) and 'Station cleanliness' has 17% less variation in the observed data than the value modeled. Finally, the point-measure correlation is, for each item, a positive value that is included between 0.58 and 0.68, and these values show the absence of mis-scoring and anormal polarity.

Table 1. Items statistics

Item	Model		Infit	Outfit	Ptmea	Exact	Match
	Measure	S.E.	MnSq	MnSq	Corr.	Obs%	Exp%
Passenger information	.63	.03	1.16	1.20	.63	49.0	50.9
Staff behavior	.45	.03	1.08	1.10	.65	52.4	52.5
Train cleanliness	.17	.03	.82	.84	.68	60.2	54.3
Passenger comfort	.10	.03	.89	.93	.64	61.2	55.3
Station cleanliness	.08	.03	.82	.83	.67	61.6	55.3
Personal and financial security	-0.2	.03	1.11	1.10	.62	54.9	55.6
Frequency of service	-.34	.03	1.05	1.09	.59	56.8	57.7
Regularity of service	-.39	.03	1.07	1.07	.58	56.6	57.8
Security	-.67	.03	.99	.96	.59	61.8	58.8

Note: 'Measure' is the estimate for the item difficulty to generate satisfaction. 'S.E.' is the standard error of the estimate. 'Infit MnSq' and 'Outfit MnSq' are infit and outfit mean-square statistic, respectively. 'Ptmea Corr' is the point measure correlation.

Category frequency counts and percentage for the rating scale is shown in Table 2. Similarly to the mean-square infit and outfit, these fit statistics have only small deviation from expectation. The highest deviation for infit and outfit mean-square is given from category 1 with 1.5 and 1.09 respectively.

Table 2. Summary of category structure

Rating scale category (Score)	Category			Infit MnSq	Outfit MnSq
	Count	Percentage	Measure		
Strong dissatisfaction (1)	1,444	7%	(-2.83)	1.05	1.09
Dissatisfaction (2)	4,621	22%	-1.02	1.02	1.06
Satisfaction (3)	9,736	47%	.85	.93	.92
Strong satisfaction (4)	4,838	23%	(3.11)	.98	.99

Note: 'Measure' is the estimate for each category. 'Infit MnSq' and 'Outfit MnSq' are infit and outfit mean-square statistic respectively.

MetroNapoli S.p.A. is interested to investigate whether the satisfaction depends on the personal characteristics of passengers (profession, age, purpose of travel, and sex). With this goal, ANOM decision chart was built. They have a centerline, located at the overall mean, and upper and lower decision limit (see equation 2). The group means are plotted, and those that fall beyond the decision lines are significantly different from the overall mean. In Figure 2.a the passengers were stratified into eight professional categories (viz., 'freelancer', 'businessman', 'worker', 'employee', 'unemployed', 'housewife', 'student', 'others') and the mean of passenger satisfaction (in logit measure) for each group was plotted in decision chart. Similarly, the passengers were stratified into age groups: 'less than 25', '25 – 40', '40 – 60, 'more than 60' (Figure 2.b), and according to the purpose of travel - 'work', 'school', 'college/university', 'personal practice', 'tourism', 'others' – (Figure 2.c). Finally, the sex of passengers was considered (Figure 2.d). At a 5% level of significance, only two professional categories have a level of satisfaction which is statistically different from the overall mean satisfaction ('student' and 'housewife'). The first can be satisfied with more difficulty, while to satisfy the housewives is easier. In the Figure 2.b it is possible to observe that there is a trend between passenger satisfaction and age groups, when the age of passengers increases is more simply observe passenger satisfied. Regarding the passengers that use MetroNapoli's service, the 'tourism' category is the most satisfied of the overall passengers. Significantly more satisfied of the overall mean are the passengers that use this service for 'personal practice', while the passengers that would use MetroNapoli's service to go to 'college/university' are significantly less satisfied. Finally, there is no difference in term of satisfaction between the 'male' and 'female' passengers.

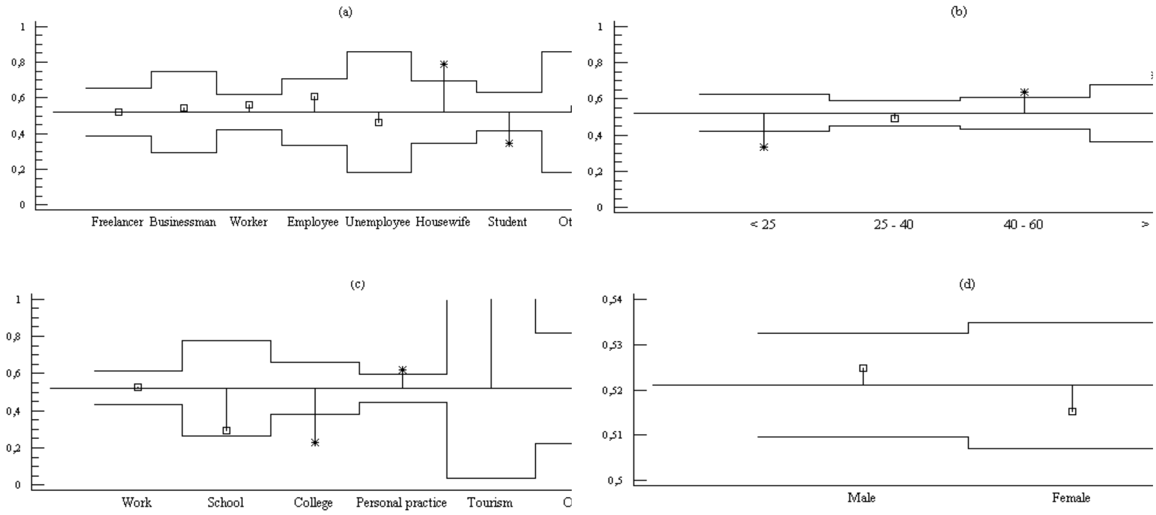


Figure 2. ANOM decision chart for personal characteristics of passengers – Profession (a), Age (b), Reason of travel (c), and Sex (d)

The decision chart helps to verify whether some characteristics of the travel influence the passenger satisfaction. For this reason, in Figure 3.a - 3.c passenger satisfaction considers the day when the interview was carried out, the frequency of travel per week ('3 or less', '4 – 6', '6 – 10', 'more than 10'), and modes of transportation that the passengers had used before arriving at the station where interview was held (intermodal passenger transportation) - 'None', 'Line 1', 'Line 6', 'Cable railway', 'Line 2', 'Railway', 'Bus', 'Cumana', 'MetroCampania', 'Circumvesuviana', and 'Car' are the categories considered. ANOM decision chart shows how, at a 5% level of significance, only Mondays and Fridays have a level of passenger satisfaction different from the overall mean. Respectively, Monday is the day of week which is harder to generate satisfaction, while Friday is easier (Figure 3.a). Regarding the relationship between the passenger satisfaction and the weekly frequency of travel, it is possible to observe that only when the number of times the users travel weekly is 'more than 10' the passenger satisfaction is statistically lower than the overall mean (Figure 3.b). Figure 3.c shows only a category that gives a level of satisfaction different from the overall mean. Passengers who had used the MetroNapoli's service after the Circunvesiana's service are statistically more satisfied than others.

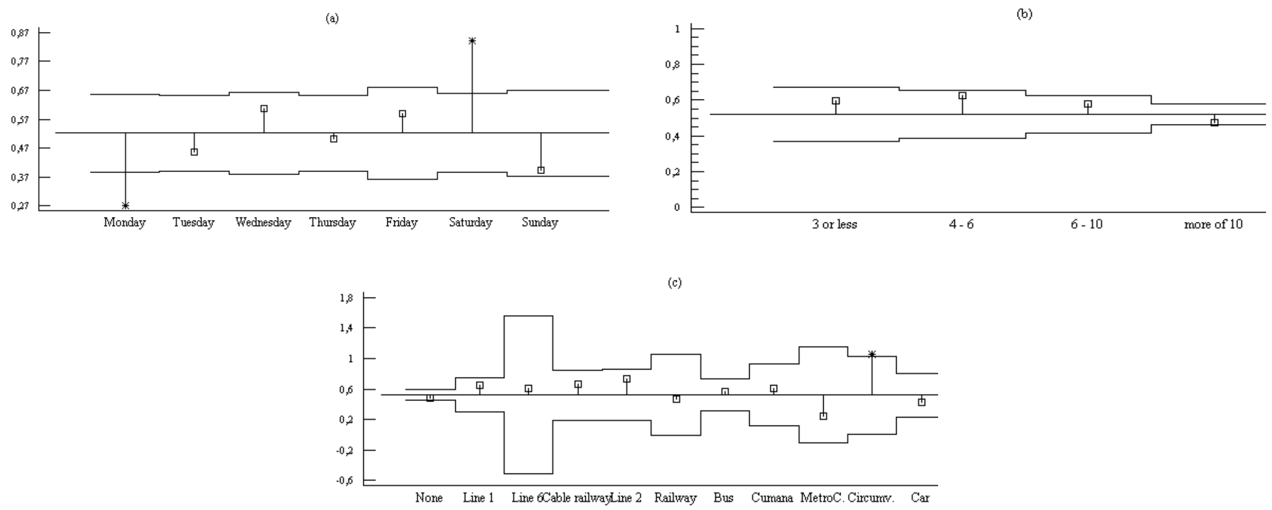


Figure 3. ANOM decision chart for characteristics of travel – Day of interview (a), Frequency of travel per week (b), and Intermodal passenger transportation (c).

4. Conclusions and discussion

The principal purpose of this paper has been to show how the Rasch analysis could be linked with other statistical methods to extract more information of data. Here, Rasch analysis was used to measure the passenger satisfaction of MetroNapoli S.p.A., and then, to get more fine information, the statistical relationship between the linear measurement of passenger satisfaction and some personal passenger information was studied. In this way ANOM decision chart is a tool that well integrated Rasch analysis.

The second aim of the work is show how, in the same case, ANOM is a good alternative to the most popular ANOVA to compare a group of means. In fact, it offers two clear advantages over ANOVA. First of all, it is more intuitive and provides an easily understood graphical result, which clearly indicates the means that are different from the overall mean. Finally, it sheds light on the nature of the differences among the groups. Moreover, in the same case, ANOM is able to give evidences on differences between the groups which can be seen in the ANOVA table. In fact, Figure 3.c shows that the passengers who had used the Circumvesuviana's service were more satisfied than the overall mean. In this case, and generally, when only one (or few) category is very different from the overall mean, ANOVA does not reject the null hypothesis (Table 3).

Table 3. ANOVA

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	20,3728	10	2,03728	1,97	0,0331
Within groups	2323,68	2243	1,03597		
Total (Corr.)	2344,05	2253			

References

1. Andrich, D. **A rating formulation for ordered response categories,** Psychometrika, 43, 1978, pp. 561-574
2. Bond, T. G. and Fox, C. M. **Applying the Rasch model: Fundamental measurement in the human sciences,** London: Erlbaum, 2001
3. Gallo, M. **The scaling problems in service quality evaluation,** Metodološki Zvezki, 2; 2007, pp. 165-176
4. Linacre, J. M. and Wright, B. D. **A user's guide to WINSTEPS: Rasch model computer program,** Chicago: MESA Press, 2000
5. Nelson, P. R. **Exact critical point for the analysis of means,** Comunications in Statistics: Theory and Methods, 11(6), 1982, pp. 699-709
6. Nelson, P. R., Wludyka, P. S. and Copeland, K. A. F. **The Analysis of Means: A graphical method for comparing means, rates, and proportions,** Philadelphia: Siam, 2005
7. Nelson, P. R. **Power curves for the Analysis of Means,** Technometrics, 27, 1985, pp. 65-73
8. Ott, E. R. **Analysis of Means – A Graphical Procedure,** Industrial Quality Control, 24, 1967, pp. 101-109. Reprinted in Journal of Quality Technology, 15, 1983, pp. 10-18
9. Wright, B. D. and Masters, G. N. **Rating scale analysis,** Chicago: MESA Press, 1982
10. Wright, B.D. and Linacre, J.M. **Observations are always ordinal; measurements, however, must be interval,** Chicago, IL: MESA Psychometric Laboratory, 1989

A PERMUTATION APPROACH TO EXAMINE THE SATISFACTION OF THE ITALIAN POPULATION TOWARDS TRANSPORT SERVICE[1]

Angela ALIBRANDI
Researcher of Statistics for Experimental and Technological Research (SECS-S/02),
Department of Economical, Financial, Social, Environmental,
Statistical and Territorial Sciences,
Messina University, Messina, Italy

E-mail: aalibrandi@unime.it

Massimiliano GIACALONE
Adjunct Professor of Statistics (SECS-S/01),
Department of Economics and Statistics,
Calabria University, Arcavacata (Rende), Cosenza, Italy

E-mail: maxgiacit@yahoo.it

Abstract: *When public transport system represents a primary need for citizens, the analysis of users' satisfaction is of the utmost importance to offer and obtain an efficient service. It is clear that a customer will be "satisfied", if his expectations are met and will be" disappointed", if his needs are ignored. In the transport field, the formulation and the definition of organizational and operational criteria are essential requisites to improve service quality. Restoring and improving modes and procedures will certainly guarantee an increasing efficiency, but the evaluation of customer satisfaction has been gaining more and more importance in order to achieve such a goal: different transport companies have been able to set such quality standards not only thanks to their own abilities, but also by taking into account specific service needs, directly expressed by customers through an adequate monitoring process.*
With this paper our aim is to make a study that analizes the satisfaction of the Italian population using transport service. With a particular reference to "Buses," "Coaches" and "Trains", we are going to evaluate the proportion of satisfied public transport users according to ISTAT indicators (frequency, punctuality and seats availability) both in regional and in geographical divisions. The methodology used is based on permutation tests.

Key words: *Permutation Approach; Transport Service; Italian Population; ISTAT Indicators*

Introduction

This work focuses on citizens' satisfaction towards the quality of transport services with reference to the great importance covered by this kind of service in the economy of every country.

According to the current priorities on the European Union agenda, high quality of transport services has become a goal to achieve (the White Paper "European transport policy until 2010: time to decide " shows the member countries the direction to follow). Each country should ensure safe and efficient transport services and high quality standards to all the European citizens, compatibly with sustainable development policies.

In Italy the current transportation system has many critical points and does not fulfill users' needs in some areas. The inadequacy of rail transport, the lack of structural and technical coaches and the slow routes cause problems to commuters and students, especially at the opening of each school year and mainly in the South: service inefficiency reaches unbearable situations for the doubling of requirements due to the higher presence of students. Freight transport in cities and road problems certainly need a comprehensive political approach. In addition to that, the extremely high number of operators seems to be impeding efficiency and affecting the entire industrial system. Surveys and measurements as well as the monitoring and the evaluation of satisfaction level, achieved by transport service users, acquire a fundamental importance, because of the increasing attention towards citizens and their numerous rights.

From a methodological point of view, some scientific contributions related to users' satisfaction towards transport services have been provided in literature. In particular, Baltes [2][2] points out to specific service elements referring to bus rapid transit system, emphasizing the importance that customers occupy. In Washingston et al. [19] the most common statistical and econometric methods are applied to analyze transportation data; in Lawson and Montgomery [13] a logistic regression analysis is proposed, with reference to customer satisfaction data in transport service; in Morfoulaki et al. [14] a multinomial logistic model was developed and estimated to provide some understanding of the factors contributing to the overall satisfaction level of customers within public transport service; in Litman [12] some developing indicators for a sustainable transport planning are individualized; in Eboli and Mazzulla [6], [7] a tool for measuring customer satisfaction in public transport is proposed; in particular, a structural equation model is formulated to explore the impact of the relation between global customer satisfaction and service quality attributes.

The purpose of our paper is to compare the proportion of satisfied Italian users towards some ISTAT indicators of transport services by geographical divisions and by different modes. Since this phenomenon shows an elevated variability on the Italian territory, we are going to perform our analysis making a comparison between territorial divisions and, subsequently, among transport modes. With reference to the most used transport modes, we are going to analyze the proportion of satisfied users particularly according to frequency of routes, punctuality and seats availability and comparing different territorial areas.

Data and methods

The data

Our data are percentages, related to users in 2007 on the basis of three ISTAT indicators (*frequency*, *punctuality* and *seats availability*), three typologies of transport modes (*Buses, Coaches* and *Trains*), and three different geographical divisions; they are taken by the ISTAT Italian Statistical Yearbook in the "Transports and telecommunications" section [10]. In particolar, we referred to table 19.14 of the above-mentioned volume, entitled *"Persone di 14 anni e oltre che utilizzano i vari mezzi di trasporto (utenza), soddisfatte per frequenza delle corse, puntualità, posto a sedere, per regione e ripartizione geografica - Anno 2007"*.

As reported in the volume, in year 2007 passenger traffic (to place of study or work) regarded more than 32 million people.

Frequency, punctuality and seats availability are the indicators chosen by ISTAT to detect the population's satisfaction towards the quality of public transport [10].

The original contribution of our work is not based on the choice of appropriate indicators to assess the quality of transport, but it concerns the statistical methodology with which we make comparisons among different territorial divisions and public transport modes.

In the figures 1-3, the radar plots show users' satisfaction according to the frequency, the punctuality and the *seats availability* of the considered transport modes in three different territorial divisions.

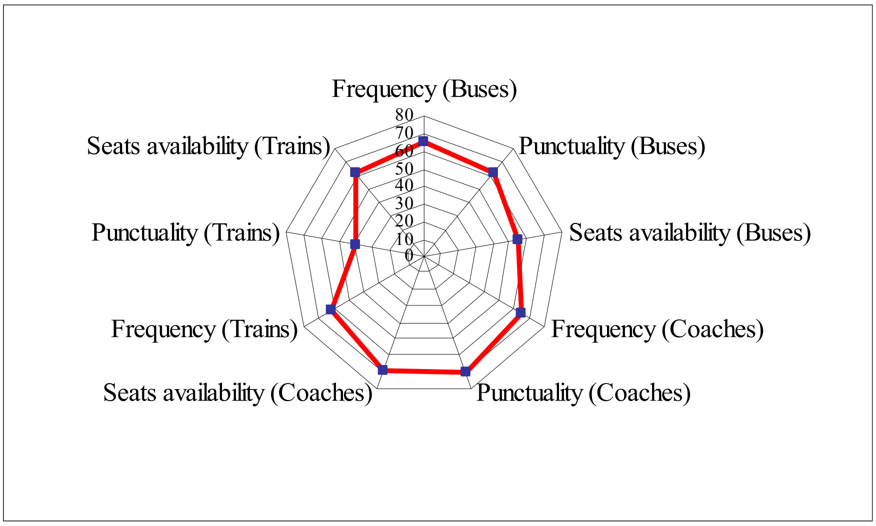


Figure 1. Users' satisfaction according to indicators and transport modes – North division

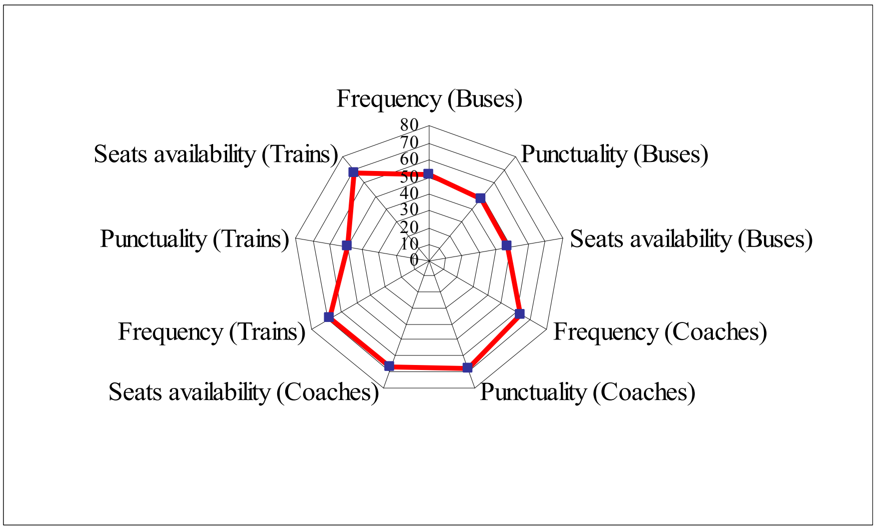


Figure 2. Users' satisfaction according to indicators and transport modes – Center division

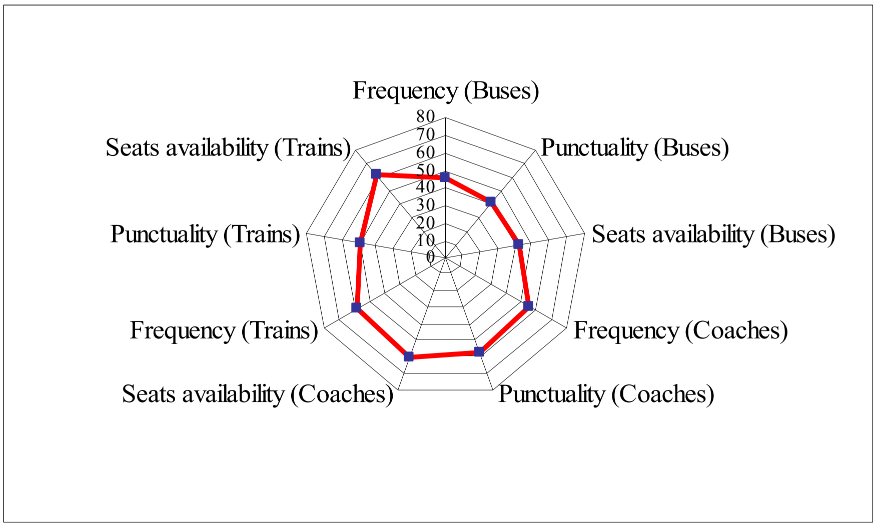


Figure 3. Users' satisfaction according to indicators and transport modes – South division

By examining Fig.1, we can see that, in the North, the lowest satisfaction is for train punctuality; the highest satisfaction is for punctuality and seats availability on coaches. In the Center (Fig.2), a low satisfaction for train punctuality persists; moreover, we can see low satisfaction regarding punctuality and seats availability on buses. By examining the South division, we can affirm that the satisfaction for punctuality and seats availability on buses is even lower, revealing the existence of serious problems related to urban transport service.

The methodology

The non-normality in the distribution of data does not guarantee valid asymptotic results, so analysis have been performed through a non parametric approach [18], using the Non Parametric Combination (NPC) test based on the permutation test [16], [5]. Permutation tests represent an effective solution for problems concerning the verifying of multidimensional hypotheses that are difficult to face in a parametric context. In comparison to the classical approach, the NPC test is characterized by several advantages: it does not require normality and homoschedasticity assumptions, it can include any type of variable [17], [11], it is reliable also in case of lacking data, it is efficient in presence of low sampling size [4], it resolves multivariate problems without needing to specify the structure of dependence among variables, it allows stratified analyses and resolves problems when observation numbers are less than variable numbers [8], [3]. The same methodology was used by Alibrandi and Zirilli [1] to analyze the customer satisfaction of the Italian population towards some aspects of socio-economic and relational life.

We consider that two or more (k) variables are observed on a set of n statistical units, gathered into two or more (C) groups, defined by a classification criterion (a treatment). The purpose of this procedure is to verify if there are any statistically significant differences among the various profiles of answer-variables of the C compared groups. We suppose that an appropriate k-dimensional distribution exists; the null hypothesis establishes equality in the distribution of the k-dimensional distribution among the C groups, against the alternative one, where at least one strict inequality is satisfied:

$$H_0 : [X_1 \overset{d}{=} \ldots \overset{d}{=} X_C]$$

that can be also expressed with

$$H_0 : [\bigcap_{i=1}^{k} X_{1i} \overset{d}{=} \ldots \overset{d}{=} X_{ci}] = [\bigcap_{i=1}^{k} H_{0i}] \quad (1)$$

$$H_1 : \bigcup_{i=1}^{k} [X_{1i} \overset{d}{\neq} \ldots \overset{d}{\neq} X_{Ci}] = [\bigcup_{i=1}^{k} H_{1i}] \quad (2)$$

where *i* represents the stratification index, referred to each stratum. Through the mentioned procedure it is preliminarily possible to define a set of *k* one-dimensional permutation tests, denominated partial tests, where the marginal contribution of every answer-variable can be examined with the comparison among groups. The partial tests are non-parametrically combined through CMC (Conditional Monte Carlo) procedure in combined tests, using a proper combination function (generally Fisher, Tippett or Liptak); these tests globally verify the existence of differences among the various distributions of the groups. If the analysis is stratified, it is possible to determine a specific test that combines the tests obtained by each stratification; it allows to draw evaluations on the possible differences among the groups in relation with all the examined variables and strata.

We can assume that, without loss of generality, partial tests acquire real values and they are marginally correct, consistent and significant for great values. The NPC test (based on Conditional Monte Carlo resampling) is carried out in the following way:

1. the value of the k-variated statistic is calculated on observations;
2. for every resampling conditioned by the observed data, we calculate the vector of the permuted statistics;
3. for each partial test and resampling, the transformation in rank is performed;
4. p-values related to the partial tests are calculated;
5. the combined resampling value is calculated using Fisher's combination function;
6. the observed value of the second order combined test and its p-value are calculated.
7. if p-value is minor than α, the H_0 hypothesis is rejected at a fixed significance level.

Results

Statistical comparisons between territorial divisions

We applied the Non Parametric Combination test in order to evaluate the existence of possible significant differences among the three geographical divisions (North, Center and South) the Italian territory can be divided in. The observations per region are considered as replications. The hypotheses system is expressed as follows:

$$H_{0i} : \left\{ Frequency_{1i} \overset{d}{=} Frequency_{2i} \right\} \cap \ldots \cap \left\{ Seats\ availab._{1i} \overset{d}{=} Seats\ availab._{2i} \right\} \quad (3)$$

$$H_{1i} : \left\{ Frequency_{1i} \overset{d}{\neq} Frequency_{2i} \right\} \cup \ldots \cup \left\{ Seats\ availab._{1i} \overset{d}{\neq} Seats\ availab._{2i} \right\} \quad (4)$$

where 1 and 2 represent the geographical division (North, Center and South) and the stratification index i (i=1,…,3) is referred to the examined transport services (buses, coaches, trains).

Table 1 shows the p-values of the partial and combined tests, calculated by Tippett's combination function [9]. The results underline an articulated territorial situation: the population residing in the North and in the Center has more elevated levels of satisfaction than the population residing in the South towards the frequency of the routes and the punctuality of "Buses" and "Coaches". The p-values associated to these comparisons are statistically significant, with a directionality of the comparisons major for the North and the Center than for the South (see the p-value reported in bold in Table 1). The satisfaction towards the railway service appears, instead, homogeneous on the entire territory, since no statistically significant differences exist in the comparisons among the territorial divisions.

Table 1. p-value of partial and combined test for comparison between divisions.

		Frequency	Punctuality	Seats availability		Comb.
NORTH VS CENTER	BUSES	0.661	0.837	0.800	→	0.850
	COACHES	0.904	0.749	0.870	→	0.957
	TRAINS	0.237	0.288	0.448	→	0.297
		↓	↓	↓		↓
	Combined	0.689	0.750	0.882	→	0.827
CENTER VS SOUTH	BUSES	**0.026 (>)**	**0.035 (>)**	0.130	→	0.042
	COACHES	**0.046 (>)**	**0.031 (>)**	0.243	→	0.054
	TRAINS	0.841	0.600	0.257	→	0.252
		↓	↓	↓		↓
	Combined	0.071	0.047	0.137	→	0.061
NORTH VS SOUTH	BUSES	**0.002 (>)**	**0.003 (>)**	**0.046 (>)**	→	0.003
	COACHES	**0.019(>)**	**0.008 (>)**	0.060	→	0.018
	TRAINS	0.074	0.628	0.068	→	0.082
		↓	↓	↓		↓
	Combined	0.006	0.008	0.096	→	0.010

Comparisons among transport modes

The Non Parametric Combination Test has also been applied in order to evaluate the existence of possible significant differences between transport modes (Buses, Coaches, Trains), considered two by two. The hypotheses system is expressed as follows:

$$H_{0i}: \left\{ Frequency_{1i} \overset{d}{=} Frequency_{2i} \right\} \cap \ldots \cap \left\{ Seats\ availab._{1i} \overset{d}{=} Seats\ availab._{2i} \right\} \quad (5)$$

$$H_{1i}: \left\{ Frequency_{1i} \overset{d}{\neq} Frequency_{2i} \right\} \cup \ldots \cup \left\{ Seats\ availab._{1i} \overset{d}{\neq} Seats\ availab._{2i} \right\} \quad (6)$$

where 1 and 2 represent the examined transport services (Buses vs Coaches, Coaches vs Trains, Coaches vs Trains) and the stratification index i=1,…,3 is referred to the territorial division in exam (North, Center, South). The tab. 2 shows the p-values of the partial and combined tests.

Table 2. p-value of partial and combined test for comparisons between divisions.

		Frequency	Punctuality	Seats availability		Comb.
BUSES VS COACHES	NORTH	0.977	0.168	**0.030 (<)**	→	0.065
	CENTER	0.880	0.331	0.291	→	0.491
	SOUTH	0.250	0.257	0.062	→	0.102
		↓	↓	↓		↓
	Combined	0.577	0.161	0.083	→	0.177
COACHES VS TRAINS	NORTH	0.131	**0.002 (>)**	**0.029 (>)**	→	0.001
	CENTER	0.114	**0.007 (>)**	0.097	→	0.007
	SOUTH	0.771	**0.042 (>)**	0.332	→	0.092
		↓	↓	↓		↓
	Combined	0.304	0.001	0.080	→	0.001
BUS VS TRAIN	NORTH	0.139	**0.001 (>)**	0.423	→	0.001
	CENTER	0.203	**0.045 (>)**	0.943	→	0.120
	SOUTH	0.368	0.908	0.165	→	0.274
		↓	↓	↓		↓
	Combined	0.357	0.001	0.409	→	0.001

The results underline that, mostly in the North and in the Centre, there is a greater satisfaction of the Italian population towards coaches, in terms of punctuality and seats availability, in comparison to the other modes in examination (see the p-value reported in bold in Table 2). Trains seem to be the less preferred transport mode among the examined ones.

The problem of performance evaluation of a transport service has significant aspects of complexity. Evaluation process must necessarily reflect the points of view of the different concerned parts: the company that runs the service and the users directly or indirectly involved with the transport service. In particular, the users' point of view is mainly influenced by the criteria of regularity (frequency of service and punctuality) and comfort (seats availability) and is related to the use of the service (ISTAT indicators). Focusing on these indicators, the authors have applied the NPC test to compare the proportion of satisfied users towards transport service according to three different transport modes (buses, coaches, trains) in three geographical divisions (North, Center, South).

Looking at the results, we can affirm that the satisfaction level of the Italian population towards the considered aspects of the transport service is a territorially divided reality: in the North and in the Center, the factor with the lowest satisfaction is the punctuality of trains; moreover, both in the Center and in the South, there exists low satisfaction regarding punctuality and seats availability on buses, revealing the existence of problems related to urban transport. Examining the results of our analysis, performed by a permutation approach in order to observe potential differences among different territorial divisions, we are able to underline that the population of the North and of the Center has more elevated levels of satisfaction regarding the frequency of the routes and the punctuality of Coaches. The satisfaction towards railway service appears, instead, homogeneous on the entire territory, since no statistically significant difference exists. The comparison between transport modes shows that, especially in the North and in the Center, there is a higher proportion of satisfied Italian users of coaches (in terms of punctuality and seats availability) than of other transport modes. Trains seem to be, among the examined ones, the less preferred.

Final remarks

In this article a permutation approach has been proposed to compare users' satisfaction in different territorial divisions (North, Center and South) with reference to three transport modes (trains, buses and coaches) and to three ISTAT indicators (frequency, punctuality and seats availability). Although NPC methodology is well-known and widely applied in several fields of research, presently there aren't many practical applications in the public transport sector, and specifically in order to examine satisfaction degree. In our paper we have applied this methodology on the basis of the non-normality of data distribution and for the flexibility of this procedure, in particular, it does not require normality and homoschedasticity assumptions, it can draw any type of variable (also percentages), it can resolve various problems [15] without needing to specify the structure of dependence among variables and it can allow stratified analyses. With the results obtained, we are to able to affirm that in the North there is the highest share of positive reviews; in the South and, partially, in the Center, we observed the largest percentage of negative reviews, probably due to the lower standard of living and to the minor efficiency of transport service. Even though data have been obtained by reliable existing sources (ISTAT yearbook), a limitation of this paper could be that these data consist of percentages. Furthermore, the analysis according to territorial divisions gives quite generalized results. To have more detailed results a comparison of users' satisfaction among regions within each division could be made. In spite of its limitation, this study could be a starting point for more exhaustive researches: as a future development, it could also perform a more focused analysis by means of a sample survey that allows to identify the users' satisfaction predictors, more related to service quality in public transport. In our country, the planning of transport services does not pay adequate attention to users' necessities and to their implicit and explicit expectations; in the reality of many public and private transport companies, quality systems are implemented without a real customer involvement; in this context a survey on users' satisfaction level is, instead, a useful tool that could overcome these gaps, allowing to monitor and measure performance quality.

Bibliography

1. Alibrandi A. and Zirilli A. **Un approccio multidimensionale per la valutazione del livello di soddisfazione della popolazione italiana su alcuni aspetti della vita socio-economica e relazionale**, Atti della Riunione Scientifica. Valutazione e Customer Satisfaction per la Qualità dei Servizi. Tor Vergata. 12-13 Aprile 2007. (pp. 1-4). ROMA, Centro Stampa Università, 2007
2. Baltes, M.R., **The importance customers place on specific service elements of bus rapid transit**, Journal of Public Transportation, 6, 4, 2003, pp. 1-19
3. Basso D., Chiarandini M. and Salmaso L. **Synchronized permutation tests in I×J designs**, Journal of Statistical Planning and Inference, 137, 2007, pp.2564-2578

4. Brombin C. and Salmaso L. **Multi-aspect permutation tests in shape analysis with small sample size**, Computational Statistics & Data Analysis (doi: 10.1016/j.csda.2009.05.010), 53,12, 2009, pp. 3921-3931
5. Corain L. and Salmaso L., **Multivariate and Multistrata Nonparametric Tests: the NPC method**, Journal of Modern Applied Statistical Methods, 3, 2, 2004, pp. 443-461
6. Eboli L. and Mazzulla G. **Service quality attributes affecting customer satisfaction for bus transit**, Journal of Public Transportation, 10, 3, 2007, pp. 21-34
7. Eboli, L. and Mazzulla, G. **La misura della qualità dei servizi di trasporto collettivo. Strumenti, metodi, modelli**. Aracne editrice, Roma, 2008
8. Finos L. and Salmaso L. **Weighted methods controlling the multiplicity when the number of variables is much higher than the number of observations**, Journal of Nonparametric Statistics, 18, 2, 2006, pp. 245-261
9. Finos L., Pesarin F. and Salmaso, L. **Confronti multipli tramite metodi di permutazione**, Statistica Applicata, 15, 2, 2003, pp. 275-300
10. Härdle, W. and Simar, L. **Applied Multivariate Statistical Analysis**, Springer, 2004
11. ISTAT, **Trasporti e telecomunicazioni**, in Annuario Statistico Italiano Anno 2007, pp. 477-511
12. Klingenberg, B., Solari, A., Salmaso, L. and Pesarin, F. **Testing marginal homogeneity against stochastic order in multivariate ordinal data**, Biometrics (DOI: 10.1111/j.1541-0420.2008.01067.x) 65, 2008, pp. 452-462
13. Lawson, C. and Montgomery, D.C. **Logistic regression analysis of customer satisfaction data**, Quality and Reliability Engineering International, 22, 2006, pp. 971-984
14. Litman T. **Developing indicators for comprehensive and sustainable transport planning**, Transportation Research Board (www.trb.org), 2017, 2007, pp. 10-15
15. Morfoulaki, M., Tyrinopoulos, Y. and Aifadopoulou, G. **Estimation of satisfied customers in public transport systems: a new methodological approach**, Journal of the Transportation Research Forum, 46, 1, 2007
16. Pesarin, F. and Salmaso, L. **Permutation Tests for Univariate and Multivariate Ordered Categorical Data**, Austrian Journal of Statistics, 35, 2&3, 2006, pp. 315–324
17. Pesarin, F., **Multivariate Permutation Test, with application in biostatistics,** Chichester, John Wiley & Sons, 2001
18. Sprent, P. and Smeeton, N.C. **Applied nonparametric statistical methods,** Chapman&Hall, 2001
19. Washington, S. P., Karlaftis, M. G. and Mannering, F. L. **Statistical and econometric methods for transportation data analysis,** Florida, Chapman&Hall/CRC Press, 2003

[1] **Acknowledgement**
The authors thank Prof. Luigi D'Ambra ("Federico II University" of Naples) for his substantial scientific support and his helpful advice.

[2] **Codification of references in text:**

[1]	Alibrandi A. and Zirilli A. **Un approccio multidimensionale per la valutazione del livello di soddisfazione della popolazione italiana su alcuni aspetti della vita socio-economica e relazionale**, Atti della Riunione Scientifica. Valutazione e Customer Satisfaction per la Qualità dei Servizi. Tor Vergata. 12-13 Aprile 2007. (pp. 1-4). ROMA, Centro Stampa Università, 2007
[2]	Baltes, M.R., **The importance customers place on specific service elements of bus rapid transit**, Journal of Public Transportation, 6, 4, 2003, pp. 1-19
[3]	Basso D., Chiarandini M. and Salmaso L. **Synchronized permutation tests in I×J designs**, Journal of Statistical Planning and Inference, 137, 2007, pp.2564-2578
[4]	Brombin C. and Salmaso L. **Multi-aspect permutation tests in shape analysis with small sample size**, Computational Statistics & Data Analysis (doi: 10.1016/j.csda.2009.05.010), 53,12, 2009, pp. 3921-3931
[5]	Corain L. and Salmaso L., **Multivariate and Multistrata Nonparametric Tests: the NPC method**, Journal of Modern Applied Statistical Methods, 3, 2, 2004, pp. 443-461
[6]	Eboli L. and Mazzulla G. **Service quality attributes affecting customer satisfaction for bus transit**, Journal of Public Transportation, 10, 3, 2007, pp. 21-34

[7]	Eboli, L. and Mazzulla, G. **La misura della qualità dei servizi di trasporto collettivo. Strumenti, metodi, modelli**. Aracne editrice, Roma, 2008
[8]	Finos L. and Salmaso L. **Weighted methods controlling the multiplicity when the number of variables is much higher than the number of observations**, Journal of Nonparametric Statistics, 18, 2, 2006, pp. 245-261
[9]	Finos L., Pesarin F. and Salmaso, L. **Confronti multipli tramite metodi di permutazione**, Statistica Applicata, 15, 2, 2003, pp. 275-300
[10]	ISTAT, **Trasporti e telecomunicazioni**, in Annuario Statistico Italiano Anno 2007, pp. 477-511
[11]	Klingenberg, B., Solari, A., Salmaso, L. and Pesarin, F. **Testing marginal homogeneity against stochastic order in multivariate ordinal data**, Biometrics (DOI: 10.1111/j.1541-0420.2008.01067.x) 65, 2008, pp. 452-462
[12]	Litman T. **Developing indicators for comprehensive and sustainable transport planning**, Transportation Research Board (www.trb.org), 2017, 2007, pp. 10-15
[13]	Lawson, C. and Montgomery, D.C. **Logistic regression analysis of customer satisfaction data**, Quality and Reliability Engineering International, 22, 2006, pp. 971-984
[14]	Morfoulaki, M., Tyrinopoulos, Y. and Aifadopoulou, G. **Estimation of satisfied customers in public transport systems: a new methodological approach**, Journal of the Transportation Research Forum, 46, 1, 2007
[15]	Härdle, W. and Simar, L. **Applied Multivariate Statistical Analysis**, Springer, 2004
[16]	Pesarin, F., **Multivariate Permutation Test, with application in biostatistics,** Chichester, John Wiley & Sons, 2001
[17]	Pesarin, F. and Salmaso, L. **Permutation Tests for Univariate and Multivariate Ordered Categorical Data**, Austrian Journal of Statistics, 35, 2&3, 2006, pp. 315–324
[18]	Sprent, P. and Smeeton, N.C. **Applied nonparametric statistical methods,** Chapman&Hall, 2001
[19]	Washington, S. P., Karlaftis, M. G. and Mannering, F. L. **Statistical and econometric methods for transportation data analysis,** Florida, Chapman&Hall/CRC Press, 2003

CLUSTER ANALYSIS – A STANDARD SETTING TECHNIQUE IN MEASUREMENT AND TESTING

Muhammad Naveed KHALID
University of Cambridge ESOL Examinations, UK

E-mail: Khalid.M@cambridgeesol.org

Abstract: *Standard setting plays an important role in educational and psychological testing. This paper is focused on standard setting using 'cluster analysis' technique. Cluster analysis is a statistical procedure for forming homogenous groups of subjects (examinees). It explores the process of doing cluster analysis and its types are – K-Means and Hierarchical clustering. In the hierarchical cluster analysis, all objects are initially being considered to be a unique cluster. The analysis proceeds sequentially by merging clusters together one step at a time until all objects are merged into a single cluster. In the K-Means cluster analysis, the number of clusters into which the objects which will be portioned is specified initially. The K-means algorithm then establishes the centers of each cluster which are represented by a vector of means (called the cluster centroid) corresponding to the variables used to cluster subjects. The procedure was applied to an achievement test in science. A five cluster solution best separated the examinees according to their proficiency skills. The study concludes that cluster analysis has an edge over other techniques in regard to reducing subjectivity based on expert ratings of items and applicability to performance-based assessments. It does not remove subjectivity from the standard setting process, but does provide subject-matter experts and test developers with a quantitative method for determining different groups of test takers.*

Key words: *Cluster analysis; Standard Setting; K-Means Clustering; Hierarchical Clustering*

1. Introduction

Standard setting is an important and perennial problem in educational and psychological testing. It plays a significant role in teaching profession in selecting the most competent examinees for various purposes. It has also become very important because of the legal and political implications of having crude selection criteria based on non objective standard setting. Over the years many methods and techniques have been evolved for standard setting. Many of these techniques share some common features with each other; others differ to a large or small extent. One of such techniques for developing standard setting in science is 'cluster analysis'. This technique is aimed at clustering examinees with similar profiles such that the task of standard setting becomes easier. This technique provides standard setters with options and perspectives than other techniques do not. It is more embedded in statistical procedures than most other techniques which make use of subjective judgments. Though cluster analysis is a technique that can be reinforced by using external validity criteria, it is also possible to execute the full process using statistical procedures alone without resorting to external subjective opinion. This makes it a powerful tool for standard setters.

Cluster analysis, unlike many other methods, also provides standard setters with multiple possibilities for standard setting and can give additional insights into the multidimensional competencies of examinees. Though cluster analysis may be used on its own for the purpose of standard setting it is perhaps better suited to enter the realm of standard setting as a supporting tool for other standard setting methods, like the Reckase charts[1] and in the future it may be better to expand efforts on consolidating such a position for it. Availability of methods like Cluster Analysis and Reckase Charts as supporting tools can give different perspectives into test taker performance which can assist in strengthening the process of standard setting. Thus the cluster analysis technique is a promising research avenue that can elevate the science of standard setting to the next level, either on its own or at-least in conjunction with other methods as a valuable support tool as discussed in this paper.

1.1. Concept of Cluster Analysis

Cluster Analysis is a statistical procedure for forming groups of similar objects. It finds a broad range of application in many fields apart from standard setting exercises in the field of education. For example, in medicine, cluster analysis is used to identify diseases and their stages: by examining patients who are diagnosed as depressed, one can find if there are several distinct sub-groups of patients with different types of depression. In marketing, cluster analysis is used to identify people with similar buying habits; by examining these characteristics one may be able to target future marketing strategies more efficiently. In the field of education cluster analysis is a relatively new technique for standards setting purposes. It is currently still being developed and seems to hold a lot of promise for the future.

Traditional standard setting-setting methods have been criticized due to reliance on untested subjective judgment, lack of demonstrated reliability and lack of external validation. Cluster Analysis builds on the strengths of other standard setting methods and addresses some of their weaknesses. In particular the method includes replication and the use of external evidence of validity while relying less on subjective judgment.

1.2. Current Standard Setting Methods

Jaeger (1989; 1995) classified standard setting methods as either *test* centered or *examinee* centered. Test centered methods involve the use of expert panelists to scrutinize items comprising the test and to make judgments regarding the probable levels of performance that borderline or marginally proficient test takers will exhibit on the items. The most popular test-centered method is the Angoff method and its modifications. For items that are scored dichotomously, the panelists in an Angoff study estimates the probability that a borderline examinee will answer an item correctly. For items scored polytomously, the panelists estimate the expected score of a borderline candidate on the item. Cutscores are set on the test score scale by summing the item probabilities or expected item scores for each judge and then averaging these sums across panelists or taking the median score. There are two primary criticisms of test-centered methods for standard setting. First, the cognitive task presented to the panelists is complex, and it is difficult to provide evidence that they understand the task or complete it as desired (Angoff, 1988; Cizek, 1996). Part of this difficulty results from the notion of a borderline test taker; it may be difficult for panelists to clearly envision the knowledge and skills characterizing this test taker and to compare these levels of knowledge and skills to those required for success on numerous test items. A second criticism of test-centered procedures is that the resulting passing standard may change if a different group of panelists is used (Angoff, 1988; Cizek, 2001). Though some change is acceptable because of the element of subjectivity and the fact that there are no golden standards, it is a good exercise to limit the subjective element in the standard setting process. Significant discrepancies due to subjective opinions of different panelists can be a serious threat to the defensibility of cut-scores.

Examinee-centered standard-setting methods use subject-matter experts to evaluate examinees rather than items. One such approach, the borderline group method, uses experts to select a group of test takers who are considered marginally proficient (i.e., who possess just enough knowledge and skills to be classified into a particular category). The median test score for this borderline group is then used as the relevant cut-score. To implement the borderline group method, the borderline test takers must be selected using criteria other than test performance. This requirement poses problems because there is no direct way to determine borderline proficiency. Thus, the same types of false-positive and false-negative classification errors associated with standard setting in general apply to the assignment of test takers to the borderline group. In addition, the cut-scores derived in this fashion would fluctuate directly with the (likely to be unknown) sampling variability over potential borderline groups.

Another popular examinee-centered method is the contrasting groups' method. In this method, experts select two groups of test takers, one considered to be above the relevant standard and one considered to be below this standard. The test scores that result in the fewest false-positive (e.g. passing a below-standard student) and false-negative (e.g. failing an above-standard student) misclassifications are selected as the passing score. Though it is easier to identify above standard and below standard groups than to identify borderline groups, in many cases identification of contrasting groups is not easy to validate. The resources required for identifying and testing above and below standard students are much larger compared with the borderline method. Overall, both methods share many short-comings like unknown sampling variability across examinee groups, classification errors in assigning examinees to groups, and practical constraints.

A review of traditional standard-setting methods reveals that, although each method has theoretical appeal, all are subject to significant limitations. Several researchers have suggested guidelines or standards for evaluating standard-setting studies (Kane, 1994b; Van der Linden, 1994; Cizek, 1996) Furthermore, it suggests that standard-setting studies can be improved by: a) including replications of the procedure to evaluate the consistency of the resulting standards; b) incorporating validity checks on the resulting standard (e.g. convergent validity with external criteria); and c) using more than one standard-setting method.

1.3. Process of Cluster Analysis

The standard setting problem is essentially a classification problem (Sireci, 1995). When standards are set on a test, the purpose is to classify each test taker into one or more groups, such that test takers with abilities close to each other should be separated from those test takers with abilities that are different such that all test takers can be classified into categories or groups with similar ability levels. Cluster analysis does exactly that, i.e. groups test takers into homogeneous clusters with respect to the proficiency measured. Each cluster is comprised examinees that are highly similar in proficiency. These clusters can then be ordered in a manner congruent with the groupings defined by the standard-setting problem.

Cluster analysis can force discrete decisions on a continuous scale. When cut-scores are used to classify test takers into one or more groups, score differences among examinees within each group are typically inconsequential. When standards are set on tests, the fundamental scaling problem is not how to best order examinees along a continuous scale but how to best partition test takers into the desired number of (discrete) groups motivated by the testing purpose. However the strength of cluster analysis can also be its shortcoming. The reason is that clustering procedures cluster the data regardless of whether truly different groups of examinees are present or not. Secondly, because it focuses on analysis of test response data, no standards can be set higher or lower than the test takers actually perform. The procedure derives standards based on what specific groups of test takers have done, rather than according to what they should have done. Although this limitation is serious theoretically; it is unlikely that a test would be constructed so far above or below examinee performance levels that no test takers would exhibit expected standards of performance. In this regard, we would like to quote some remarks of researchers who have analyzed cluster analysis results:

> *We applied the procedure to a state-wide mathematics proficiency test .The standards developed from cluster analysis were compared with those established at the local level and with those derived from a more traditional borderline and contrasting groups analysis. We observed relative congruence across the local cut score and those derived using cluster analysis, and we observed similar correlations among the resulting proficiency groupings and course grades. The results of the more traditional borderline and contrasting groups analysis were less favorable. We conclude that cluster analysis appears useful for helping set standards on educational tests (Sireci, 1999).*

1.4. Types of Cluster Analysis

There are two sorts of cluster analysis that can be used to form clusters. The first is called hierarchical cluster analysis and the second is called the K-Means cluster analysis. In hierarchical cluster analysis, all objects are initially being considered to be unique clusters. The analysis proceeds sequentially by merging clusters together one step at a time until all objects are merged into a single cluster. A "N-cluster" solution is, however of no practical use. The work for the standard setter is to determine the cluster solution in between these two extremes at which truly different clusters are merged together. The cluster solution preceding that point represents the best clustering of the data. The standard setter can make use of both internal and external criteria to help determine the optimal clustering solution. A severe limitation of this form of clustering is that once test takers are merged into a cluster, they are stuck for the remainder of the analysis, even if a rearrangement of test takers across clusters may improve the solution. Also this method is not suitable for large data sets due to the extremely large number of within and cross cluster comparisons that need to be made at each stage of analysis. However, in most educational standard setting exercises the goal is not to uncover the "true" cluster structure of the data but to identify the optimal partitioning of the examinee population that best corresponds to a stated number of groupings. Thus when the number of clusters in which examinees are to be partitioned is known at the start as in most educational instances the K-means clustering can be used. However, experienced examiners can use the hierarchical clustering as a preliminary to K-means clustering to have an estimate of how many real clusters there are that may then be specified in the K-means analysis.

In hierarchical clustering, clusters are formed by grouping cases into bigger and bigger clusters until all clusters are members of a single cluster. Before the analysis begins all cases are considered separate clusters: there are as many clusters as there are cases. At the first step, two of the cases are combined into a single cluster. At the second step either a third case is added to the existing cluster of two cases or two other cases are merged into a new cluster. At every step, either individual cases are added to the existing cluster or two new cases are merger into a new cluster. However once a cluster is formed it cannot be split. There are many criteria for deciding which cases or clusters should be combined at each step. A common method is the single linkage method; the first two cases combined are those that have the smallest distance between them. The distance between the new cluster and individual cases is then computed as the minimum distance between an individual case and a case in the cluster. The distance between cases that have not been joined do not change. At every step, the distance between two clusters is the distance between their two closest points. Another commonly used technique is called the complete linkage or the furthest neighbor technique. In this method, the distance between two clusters is calculated as the distance between their two furthest points. Yet another method is the average linkage between groups method, often called UPGMA which defines the distance between two clusters as the average of the distances between all pairs of cases in which one member of the pair is from each of the clusters. This differs from the other linkage methods in that it uses information about all pairs of distances, not just nearest or the furthest. Another method is the centroid method which calculates the distance between two clusters as the distances between their sums for all of the variables. In the centroid method, the centroid of a merged cluster is a weighted combination of the centroids of the two individual clusters, where the weights are proportional to the size of the clusters. In the median method the two clusters being combined are weighted equally in the computation of a centroid, regardless of the number of cases in each. This allows small groups to have equal effect on the characterization of larger clusters into which they are merged. When similarity measures are used, the criterion for combining is reversed, i.e. the clusters with large similarity based measures are merged.

Once the distance matrix between all cases and clusters has been calculated the actual formation of clusters commences which can be seen on an *icicle plot* or a *dendogram*. Both are graphical representations of the output. Commonly an *icicle plot* is used. An *icicle plot* is a graphical representation in which the clustering steps are shown on the vertical axis against the cases being clustered on the horizontal axis. The number of clustering steps is equal to the number of cases and at each step one case or cluster is combined with another case or cluster. Thus in step 1, there are as many clusters as cases and at every step the number of individual cases reduces by 1 until in the last step all cases have been merged into one cluster. The challenge for the examiner is to identify how many real clusters are there based on the results of the hierarchical analysis shown on the *icicle plot* or the *dendogram* etc.

In K-means clustering, the number of clusters into which the objects which will be portioned is specified initially .The K-means algorithm then establishes the centers of each cluster which are represented by a vector of means (called the cluster centroid) corresponding to the variables used to cluster test takers. For example, if test takers are being clustered based on their performance on four different sections of a test, the four means on each test section determine the centroid of a cluster, where the means are calculated using only those test takers in that cluster. The number of means constituting each centroid is equal to the number of variable used to cluster the objects. The number is denoted by K, hence the name K-means clustering. This type of scaling has two obvious differences from traditional psychometric scaling. First, the distance among test takers is not determined from a single mean but rather from a vector of means. Second, instead of test takers being placed on a continuous scale, they are placed into one of a discrete number of clusters. These clusters can be used to inform the standard-setting process by relating the examinee clusters to the proficiency groupings invoked by the standard setting and test development processes.

The typical K-means algorithm begins by searching through the data to find the Q test takers that are most different from one another with respect to the clustering variables e.g. sub-scores on the test, where Q represents the number of clusters specified in advance by the researcher. At this point, the K scores for these test takers are used as cluster centroids.

The K-means algorithm is iterative: each test taker is assigned to a cluster by computing the distance between the test taker and each cluster centroid and assigned to the cluster whose centroid it is closest to. Once all test takers have been assigned to the initial clusters, the cluster means are recomputed as an average of all cluster members and the clustering exercise is repeated. Some test takers are placed in a different cluster after every iteration. The iterations carry on until there is no test-taker movement across clusters. At this point the clusters are said to have stabilized and iterations finish. The resulting clusters are the final clusters; their membership represents the result of the clustering exercise.

1.5. Basic Steps in Cluster Analysis

Three main decisions need to be made in order to perform a cluster analysis on a set of data. The first is the selection of variables. This is a very crucial step. If important variables are excluded, poor or misleading findings may result. The variables chosen should be such that they cover the whole range of important factors that cause similarities or dis-similarities between the items. There are at-least three options for selecting the variables to be used for clustering the test takers: 1) use all individual items comprising the test, or 2) use orthogonal factor scores obtained from item level factor analysis, or 3) use sub scores derived from items comprising the major area of the test. The second decision is to look into 'how alike are the cases'? In cluster analysis, items are clustered on the basis of their nearness or closeness to each other. The nearness or closeness is measured in terms of their distance from each other. A commonly used index for distance between items is the either the Euclidean distance or squared Euclidean distance, which is the sum of the squared differences over all of the variables.

Euclidean distance $(x,y) = \{\sum_i (x_i - y_i)^2\}^{½}$

Squared Euclidean distance $(x,y) = \sum_i (x_i - y_i)^2$

The third decision is regarding the criteria for combining clusters. There are many criteria for deciding which clusters or cases should be combined. All criteria are based on a matrix of either distances or similarities between pairs of cases. Often it is sum of Euclidean distances of the items from the vector of means of the clusters (centroids) which determine the placement of an item in any cluster.

2. Methodology

In order to render a judgment on whether cluster analysis should be used or not, we first intended to carry out a practical test of cluster analysis on a set of pre-marked data in order to discuss the results of cluster analysis in light of real evidence. Unfortunately, as we were unable to get hold of real score cards where grades of test takers were listed next to their test marks. It would have been interesting to compare the grades suggested by cluster analysis with the actual grades of the test takers. Nevertheless we would like to demonstrate the results of a cluster analysis carried out on a set of non-graded data and explain the result of the clustering exercise and relevant statistical information.

2.1. Data

We have done cluster analysis on the data set which consisted of 60 dichotomous items marked 1 or 0 and two polytomous items. The sample size for this study was 3000.

2.2. Defining Variables

The first step in the analysis was to define the variables. Given the large number of items comprising the test and the unknown possibility of inter-correlation among the content areas, we decided to use the method based on content areas sub-scores. Sub-scores for each of the content areas defined in the test were used as the input variables for cluster analysis.

On the basis of the test data available, it seemed best to partition the test into five content areas. The two polytomous items were left as they were but we decided to group the 60 dichotomous items into three groups of 20 items each as the data file suggested that there test content consisting of the dichotomous part consisted of three different sort of test areas of 20 questions each. The sub-scores for students in the dichotomous area were computed by summing their item scores within each content area. So we ended up with five variables: three for the 3 sub-sections of the dichotomous part and two for the 2 polytomous items. The next step was to decide if we wanted to standardize the content area sub scores prior to clustering to account for differences in the raw score scales due to any differences in the number of items in the content area .The number of items in each content area of the dichotomous section were equal i.e. 20 but the raw scores scale in the 2 polytomous items were lower. They were marked on a scale of 10 which was half of the scale in other sub-categories i.e. 20. For analysis we assumed that each content area was equally important and was supposed to have an equal bearing on the final grade. Thus it was needed to rescale the content area sub scores and bring them at par with each other so that they have an equal effect on the measurement of distances during cluster formation. It was decided to transform the polytomous items scale by doubling all the item scores in the polytomous content area to bring it at par with the dichotomous content areas scale. Thus each content area was now represented on a scale of 1 to 20.

The plan was: a) to perform a Hierarchical Cluster Analysis on the data file; b) to perform a K-means Analysis on the data file; and c) compare results of a K-Means and Hierarchical Analysis and suggest ways for improvement.

3. Discussion of Results

3.1. Hierarchical Analysis

Hierarchical cluster analysis was performed on the data set. The analysis suggested that a minimum of five clusters should be used for grouping the examinees. A large difference was found in the 'coefficients' column in the attached agglomeration schedule between a four cluster solution and a five cluster solution. The column labeled 'co-efficients' represents the distance between two combining clusters. By examining these values we got an idea about how unlike the clusters being combined are: small co-efficients indicate that fairly homogenous clusters are being merged while large co-effecients indicate that clusters containing quite dissimilar members are being combined. These coefficients can be used as guidance in deciding how many clusters are needed to represent the data. It is best to stop further clustering as soon as the increase between two adjacent steps becomes large. In our case, there was a significant increase of around 34 between the five cluster solution steps.

3.2. K-Means Analysis

The researchers decided to select a number of clusters suggested by the hierarchical cluster analysis which suggested that at-least 5 different clusters should be there. Thus all test takers in the K-means analysis were grouped in each of the 5 levels which they are closest to.

The cluster centroids for each cluster were determined by the K-means algorithm. It selected the N number of students (where N is the number of specified clusters) whose scores were most different from each other. After that using the Euclidean distance formula, the K-means algorithm placed the rest of the students in their respective clusters after calculating their distances from the K-means centroids. The process was iterative and carried on until there was no shifting of test takers across clusters, i.e. stability was achieved. For the given data set in the SPSS file and the number of clusters specified as 5, the results of the K-means clustering can be seen in the SPSS output file shown in Table I.

Table I. Initial Cluster Centers

	Cluster				
	1	2	3	4	5
VAR0001S	12.00	16.00	4.00	4.00	12.00
VAR0002S	12.00	14.00	2.00	2.00	4.00
VAR00003	13.00	15.00	16.00	5.00	4.00
VAR00004	6.00	20.00	19.00	5.00	14.00
VAR00005	14.00	19.00	14.00	4.00	10.00

Table II. Iteration History

Iteration	Change in Cluster Centers				
	1	2	3	4	5
1	6.942	7.296	6.462	5.910	5.605
2	.807	1.139	.780	1.691	.570
3	.371	.000	.209	.476	.834
4	.172	.199	.306	.313	.594
5	.395	.237	.000	.200	.187
6	.202	.176	.000	.294	.419
7	.138	.000	.000	.172	.000
8	.122	.000	.000	.202	.142
9	.120	.000	.000	.000	.129
10	.000	.000	.000	.000	.000

Tables I-IV represent the output of a K-Means clustering exercise for the data set specified before. The numbers of clusters specified were five. Table I shows the initial cluster centers selected by the K-Means algorithm. Table II shows the iteration history from which it can be seen that after 10 iterations the all clusters were stabilized into the final form. A convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any cluster is .000. The current iteration is 10. The minimum distance between initial centers is 13.638.

Table III. Final Cluster Centers

	Cluster				
	1	2	3	4	5
VAR0001S	9.16	13.28	7.53	6.28	11.19
VAR0002S	5.96	7.54	3.84	4.61	5.43
VAR00003	12.33	15.72	14.39	7.06	9.55
VAR00004	10.29	16.33	15.39	9.36	14.38
VAR00005	12.93	17.44	16.34	8.94	12.29

Table III shows the final cluster centers and Table IV shows the distance between the final cluster centers. Table V shows the cluster membership. As can be seen from the table the cluster membership is fairly even i.e. the examinees are fairly evenly spread across the five clusters, which is a desirable feature of an exam.

Table IV. Distance between Final Cluster Centers

Cluster	1	2	3	4	5
1		9.370	7.004	7.397	5.417
2	9.370		7.115	15.918	8.788
3	7.004	7.115		12.130	7.545
4	7.397	15.918	12.130		8.208
5	5.417	8.788	7.545	8.208	

Table V. Number of Cases in each Cluster

Cluster	1	45.000
	2	39.000
	3	38.000
	4	36.000
	5	42.000
Valid		200.000
Missing		.000

Clusters can now be ordered into a hierarchal order by content experts if certain content areas are to be given priority over others or simply by summing the means of each final cluster centroid and then placing them in ascending order according to their net total scores with the highest number representing the highest cluster. After the clusters have been aligned in a hierarchical order, the cut scores can then be set. One way could be to set the mean scores of clusters, i.e. cluster centroids as the cut-off scores. Another way could be to identify the overlapping regions between clusters and then take the mean score of the overlapping regions to be the cut scores. Yet another way that could better determine the middle point of the overlapping region would be to take the median score of the overlapping region as the cut-score. The median method reduces the effects of any large variances in test scores of individual test takers on the whole group of test takers in the region under study. With this particular method borderline students can be better identified. Border-line students would be those who lie in the overlapping regions and would barely pass or fail depending upon their position with respect to the mean score of the over lapping region. It would be interesting to see how much variance exists between taking the median cluster scores or median of overlapping regions as the cut scores?

It is also possible to carry out other statistical procedures on the cluster items to determine the variance of variables within and across different clusters. Using this, we can observe how student response to certain item sets i.e. the variables varies across clusters. A high ratio of inter-cluster vs. intra-cluster would mean the variable varying significantly across clusters. This can give an insight into how clusters differ from each other. To do this a one way ANOVA is done on the data set as shown in the table VI.

Table VI. Inter-cluster and Intra-cluster differences through ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
VAR0001S	301.075	4	5.810	195	51.819	.000
VAR0002S	76.121	4	6.418	195	11.861	.000
VAR00003	472.557	4	5.576	195	84.752	.000
VAR00004	383.855	4	5.288	195	72.587	.000
VAR00005	433.098	4	4.294	195	100.852	.000

A high value of F ratio between and within cluster and a low significance value implies that the variables vary significantly across clusters. As can be seen from Table VI, all variables vary across the clusters, with variable 2 varying the least and variable 5 varying the most.

3.3. Analysis of Hierarchical vs. K-Means Clustering Results

The comparative analysis if Hierarchical and K-Means showed that the results of a hierarchical cluster analysis for a 5 cluster solution compared with a K-means clustering solution for 5 clusters. The numbers below represent the case membership for 5 hierarchical clusters. Those numbers highlighted in bold represent those members whose cluster has changed in a subsequent K-means analysis. Those which are not highlighted represent those cases which remain in the same cluster in both hierarchical and K-means analysis.

Membership of Cluster 1:

1,2,**3,4,5**,**6**,8,**9,11**,14,15,**16**,18,19,**20**,22,**23**,**24,26**,27,**28**,**29**,30,33,**36,37**,**38**,39,**41,42**,43,44,**47,48**,**51,52,53**,54,**60,66,67,71**,72,**73,75,76,79**,80,82,85,89,92,**93**,**97,99**,102,**103**,**105**,106,108,112,**113**,114,**115,116,121**,122,**123,125**,128,130,131,**132**,133,**136,138**,145,**147,154,155,156,158**,159,**160,161,167**,169,**172,174,177,178,180**,**184,185**,187,**188,192,194**,**196**,197,**199,200**.

Remarks: Total members in hierarchical solution: 102.

Total number of members in K-means solution: 45

Number of common members: 36

Membership Cluster 2:
7,10,**12**,13,17,21,32,**34**,35,40,45,46,**49,50**,55,**56**,58,**59**,61,**62**,63,**64**,65,69,**74,78**,**83**,84,**86**,90,**91,95,96**,98,**100**,101,104,107,**111,117,118**,**119**,120,**124**,**126**,127,129,**134**,135,137,**139,142**,**143,144,146,148**,**150**,152,**153**,162,163,**164**,165,**166,168,170**,171,**173,175**,**176,179,181**,**182**,186,189,**190,191**,195,**198**.
Remarks: Total members in hierarchical solution: 79
Total number of members in K-means solution: 39
Number of common members: 34

Membership cluster 3:
25,31,57,68,70,87,140,151,193.(all cases have become members of cluster 4 of K-means)
Remarks: Total members in hierarchical solution: 9
Total number of members in K-means solution: 38
Number of common members: 0

Membership Cluster 4:
77,81,94,110,141,183.
Remarks: Total members in hierarchical solution: 6
Total number of members in K-means solution: 36
Number of common members: 0

Membership Cluster 5:
109,149. (Both cases have become members of cluster 2 in K-Means)
Remarks: Total members in hierarchical solution: 2
Total number of members in K-means solution: 42
Number of common members: 0

Overall Comparison between Hierarchical and K-Means:
Total number of cases in test = 200.
Number of cases falling in common clusters = 45 +39 = 84
Number of cases falling in different clusters = 200-84= 116.

Thus more than 50% of the cases in our test are apportioned into unlike clusters when put through a Hierarchical and K-means analysis subsequently. This suggests that there are significant discrepancies between the results of a K-means and Hierarchical analysis, even when the number of clusters for a K-means analysis is chosen after looking at the results of an initial Hierarchical analysis as explained before. Thus there still exists the need to find ways to bring the results of a Hierarchical analysis closer to a K-means analysis to give more legitimacy to the cluster analysis technique as a whole. One way could be to do a K-means clustering after every step in a Hierarchical cluster analysis. This way it would be possible to transfer cases across clusters if the need arises and the difference between a K-means outcome and a hierarchical outcome ought to be reduced. However this would require a more complex clustering algorithm which is not available for the time being.

4. Conclusions

There is no perfect method for setting standards on educational tests. However, the cluster analysis procedure can provide additional information that can be useful for helping set standards. If test data are available, cluster analysis can be used to help select potential borderline, proficient, below proficient, and other groups of examinees that are typically selected using only expert judgment. Thus, the performance of examinees in specific clusters can be compared to those identified using subjective judgment only. Thus, such analyses could be valuable in helping evaluate the results of both test-centered and (other) examinee-centered methods.

Though the clustering approach does not remove subjectivity from the standard-setting process, it does provide subject-matter experts and test developers with a quantitative method for determining different groups of test takers. A potentially desirable feature of the cluster analysis approach is that it provides different options for setting cut-scores. For example, the interval of overlap between examinees in adjacent clusters could be used to select a cut-score interval rather than a specific cut-score. Such an interval provides flexibility to policymakers who must consider politics, resources, and other factors when deciding where to set a cut-score. Similarly, comparing cut-scores resulting from cluster-defined contrasting and borderline groups allows for the evaluation of competing cut-scores. Thus, clustering procedures can provide a set of potential cut-scores, the elements of which can be further evaluated by content experts, psychometricians, and other relevant constituencies who may inform policy decisions.

An attractive feature of the clustering approach is the absence of a unidimensionality requirement. An interesting observation by Sireci (1995) is that by clustering examinees, groups of test takers with relative strengths and weaknesses across the different content areas may be observed, even when factor analysis of the test data indicates the test is measuring a unidimensional construct. Thus, cluster or factor analysis of examinees rather than of items may provide new insights regarding test dimensionality.

However there are two areas where attention will have to be given for the sake of validity of cluster analysis. These are: a) evaluation of the stability of the cluster solution across samples, and b) external validation of the solutions. These two evaluations are necessary to ensure the cluster solutions are stable and meaningful rather than artifactual. Future applications with larger sample sizes should consider replicating the analyses over several samples. For instance one way could be to use cross tabulation, in which the available data is divided into two sets and a clustering model is evolved that is compatible with the score distributions in the first set and then that very same particular clustering model is applied to the other data set to see if it also fits that nicely.

Future research should also explore other methods for deriving cut-scores from cluster analysis solutions. For example, given a score interval that seems to best separate clusters differing in proficiency, the score within this interval associated with the greatest test information (i.e., lowest conditional standard error of measurement) may be chosen as the cut-score. Thus, clustering approaches should be combined with emerging approaches for scaling and setting standards on educational tests to produce optimal results. In addition, the generalizability of the clustering approach needs to be further investigated with different types of tests and score distributions.

References

1. Angoff, W. H. **Proposals for theoretical and applied development,** Applied Measurement in Education, 1(3), 1988, pp. 215-222
2. Cizek, G. J. **An NCME instructional module on setting passing scores,** Educational Measurement: Issues and Practice, 15(2), 1996, pp. 20-31
3. Cizek, G. J. **Setting performance standards: Concepts, methods and perspectives,** Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers, 2001
4. Jaeger, R. M. **Certification of student competence,** in: Linn, R.L. (ed.) "Educational Measurement (3rd ed.)", New York: Macmillan, 1989, pp. 485-514
5. Jaeger, R. M. **Setting performance standards through two stage judgmental policy capturing,** Applied Measurement in Education, 8, 1995, pp. 15-40
6. Kane, M. **Validating the performance standards associated with passing scores,** Review of Educational Research, 64 (31), 1994
7. Sireci, S. G. **Using cluster analysis to facilitate standard setting,** Applied Measurement in Education, 12(3), 1999, pp. 301-325
8. Sireci, S. G. **Using cluster analysis to solve the problem of standard setting,** Paper presented at the meeting of the "American Psychological Association", New York, 1995
9. Van der Linden, W. J. **Internationalization in educational measurement,** Educational Measurement: Issues and Practice, 13, 4, 1994

a. * * * **Statistical Package for Social Sciences (2010),** SPSS for Windows (Version 18.0) [Computer software].

10. http://www.statsoft.com/textbook/stcluan.html, accessed in December 2010.
11. http://www.psychstat.smsu.edu/MultiBook/mlt04m.html, accessed in January 2011.
12. http://149.170.199.144/multivar/hc.htm, accessed in January 2011.

[1] Whereas the Reckase charts provide information about the probability of an examine with a certain test mark scoring correctly on a certain item in the test, cluster analysis groups alike students.

DEPENDENT-ALPHA CALCULATOR: TESTING THE DIFFERENCES BETWEEN DEPENDENT COEFFICIENTS ALPHA

Sabry M. ABD-EL-FATTAH
Department of Educational Psychology,
Faculty of Education,
Minia University, Minia, Egypt

E-mail: sabryrahma@hotmail.com

Hala K. HASSAN
Department of Educational Psychology,
Faculty of Education,
Minia University, Minia, Egypt

E-mail: halamaklad2000@yahoo.com

Abstract: *Coefficient alpha (α) was first introduced by Lee J. Cronbach in 1951 and since then it continues to serve as a valuable index of reliability within different areas of research. According to the Social Sciences Citation Index, between 1951 and 2010, Cronbach's seminar article (Cronbach, 1951) was cited 6,912 times by other published articles and numerous other publications often cite secondary sources in support of the use of coefficient alpha.*

Key words: *Dependent-Alpha Calculator; Dependent Coefficients Alpha*

Although most statistical packages offer computation of coefficient alpha, no widely available package incorporates tests comparing two coefficients alpha in dependent samples as may arise when testing the equality of alpha across time or when testing the equality of alpha for two test scores within the same sample. For a dependent group analysis, a feasible research question might be ''Does reducing test length or changing mode of item presentation significantly impact internal consistency?'' In this case, the null and alternative hypotheses are H0: $\alpha_{dif} = 0$ and H1: $\alpha_{dif} \neq 0$ where $\alpha_{dif} = \alpha 1 - \alpha 2$, and α1 and α2 are alpha coefficients for two different test scores in the same population.

According to Feldt, Wodruf, and Salih (1987), the methodology for the case of dependent statistics was first developed for H_0: $\alpha_{dif} = 0$ (H_0: $\zeta_1 = \zeta_2$). Feldt (1980) recommended the following test statistic

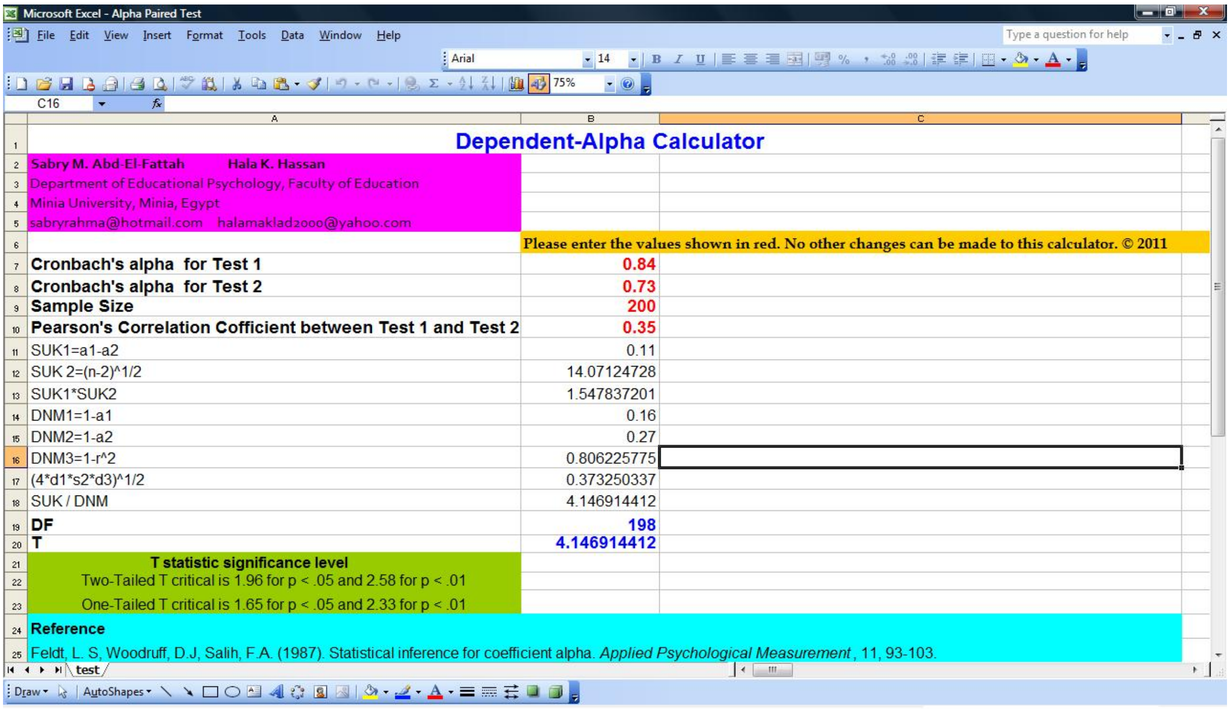
$$t = \frac{(\hat{\zeta}_1 - \hat{\zeta}_2)(N - 2)^{1/2}}{[4(1 - \hat{\zeta}_1)(1 - \hat{\zeta}_2)(1 - \hat{\rho}^2)]^{1/2}} \qquad (DF = N - 2)$$

Where $\hat{\zeta}_1$ denotes coefficient alpha for the first test, $\hat{\zeta}_2$ denotes coefficient alpha for the second test, $\hat{\rho}^2$ denotes the squared Pearson's correlation coefficient between the two total-test scores for the sample.

Although a FORTRAN program was developed for this purpose and made available by Lautenschlager (1989; Merino & Lautenschlager, 2003), it is not very accessible to users. The new Dependent-Alpha Calculator provides a user-friendly interface based on Microsoft Excel for testing hypotheses in line with the formula presented in Feldt et al. (1987) to allow tests for differences among a user-defined set of coefficient alpha values for dependent samples.

Input and Output

The user is queried by the Dependent-Alpha Calculator for coefficients alpha for both tests, sample size, and the Pearson's correlation coefficient between both tests. The Dependent-Alpha Calculator responds by calculating parts of the equation suggested by Feldt et al. (1987). The output includes the *t* test statistic and degrees of freedom. There is a built-in note in the Dependent-Alpha Calculator to help users determine the significance level of *t* tests statistic at both .05 and .01 significance levels.


Microsoft Excel - Alpha Paired Test

	A	B
1	Dependent-Alpha Calculator	
2	Sabry M. Abd-El-Fattah Hala K. Hassan	
3	Department of Educational Psychology, Faculty of Education	
4	Minia University, Minia, Egypt	
5	sabryrahma@hotmail.com halamaklad2000@yahoo.com	
6		Please enter the values shown in red. No other changes can be made to this calculator. © 2011
7	Cronbach's alpha for Test 1	0.84
8	Cronbach's alpha for Test 2	0.73
9	Sample Size	200
10	Pearson's Correlation Cofficient between Test 1 and Test 2	0.35
11	SUK1=a1-a2	0.11
12	SUK 2=(n-2)^1/2	14.07124728
13	SUK1*SUK2	1.547837201
14	DNM1=1-a1	0.16
15	DNM2=1-a2	0.27
16	DNM3=1-r^2	0.806225775
17	(4*d1*s2*d3)^1/2	0.373250337
18	SUK / DNM	4.146914412
19	DF	198
20	T	4.146914412
21	T statistic significance level	
22	Two-Tailed T critical is 1.96 for p < .05 and 2.58 for p < .01	
23	One-Tailed T critical is 1.65 for p < .05 and 2.33 for p < .01	
24	Reference	
25	Feldt, L. S, Woodruff, D.J, Salih, F.A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11, 93-103.	

Availability

The program is available-by contacting either author- as a Microsoft Excel file.

References

1. Cronbach, L. J. **Coefficient alpha and the internal structure of tests,** Psychometrika, 16, 1951, pp. 297-334
2. Feldt, L. S. **A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample,** Psychometrika, 45, 1980, pp. 99-105
3. Feldt, L. S., Woodruff, D. J., and Salih, F. A. **Statistical inference for coefficient alpha,** Applied Psychological Measurement, 11, 1987, pp. 93-103
4. Lautenschlager, G. J. **ALPHATST: Testing for differences in values of coefficient alpha,** Applied Psychological Measurement, 13, 284, 1989
5. Merino, C. and Lautenschlager, G. **Comparacio´n estadı´stica de la confiabilidad alfa de Cronbach: Aplicaciones en la medicio´n educacional y psicolo´gica,** Revista de Psicologı´a, 12(2), 2003, pp. 127-136