Application of Correspondence Analysis and Related Methods in Evaluation of Knowledge and Skills of Young People

Józef Z. Dziechciarz, Wroclaw University of Economics
Alicja Grześkowiak, Wroclaw University of Economics
Agnieszka Stanimir, Wroclaw University of Economics

Available at: https://works.bepress.com/jozefz-dziechciarz/24/
International Conference on

CORRESPONDENCE ANALYSIS
AND RELATED METHODS

(CARME 2011)

Agrocampus Ouest
Rennes

8 – 11 February 2011

Scientific Committee

• Jörg Blasius (University of Bonn)
• Michael Greenacre (Universitat Pompeu Fabra)
• Jérôme Pagès (Agrocampus Rennes)

Local Organizing Committee

• Julie Josse
• Karine Bagory  • Alain Bernardeau
• Yuna Blum  • Anne Bourdeau
• Marine Cadoret  • Stéphane Crespel
• François Husson  • Gabriel Jalam
• Sébastien Lê  • Elisabeth Lenauld
• Vinciane Marchais  • Jérôme Pagès
Jean-Paul Benzécri, 1st April 2009, 
Symposium on Statistical Learning and Data Science, University Paris Dauphine 
(snapshot Jérôme Pagès).

Cover design 
*Graphism*: Michael Greenacre. The word cloud on the cover design is based on the words of all the abstracts of this CARME conference, so in a certain sense it is an “analysis” (most frequent words, and size related to frequency) 
*Design*: Yuna Blum
Avant-propos


Rapidement, une équipe l'entoura. Au premier rang de laquelle, Brigitte Escofer, élève de la première heure et contributrice décisive dès cette époque puisque c'est à ce moment qu'elle découvrit les relations de transition. Que l'on y songe un instant : en 1961, l'analyse en composantes principales existait depuis presque trente ans et ces relations, qui, *mutatis mutandis*, s'y appliquent aussi, étaient passées inaperçues ! Par la suite, pendant trente années, Brigitte a enrichi l'analyse des données par de nombreuses recherches importantes (rassemblées en 2003 dans un livre) et tout naturellement un congrès à Rennes sur l'analyse des correspondances ne peut qu'être aussi en son hommage.

Outre Brigitte Escofer, citons deux autres élèves de la première heure : Marie-Odile Lebeaux et Brigitte Le Roux -Brigitte nous parlera de la période rennaise. En septembre 1965, Jean-Paul Benzécri partit pour la capitale avec son équipe.

A la fin des années 60, Michel Kerbaol, ancien de l'équipe parisienne, revint sur Rennes, où, travaillant à l'INSERM, il introduisit l'analyse des correspondances dans le monde médical. C'est précisément Michel Kerbaol qui fit ce qui fut sans doute le premier cours d'analyse des données sur Rennes. C'était en 1971, à Agrocampus. Il suscita l'enthousiasme de plusieurs étudiants. L'un d'entre eux, votre serviteur, devenu enseignant l'année d'après, y développait, autour de l'analyse des correspondances, un enseignement d'abord, une recherche ensuite, solidement épaulé en cela par l'équipe fondateure, alors parisienne : c'est ainsi que, par ordre alphabétique, Pierre Cazes, Ludovic Lebart et Maurice Roux vinrent à plusieurs reprises nous apporter la bonne parole. Nous sommes enchantés de les accueillir à nouveau à l'occasion de ce congrès.

Aujourd'hui, l'analyse des données est une composante principale d'Agrocampus. En enseignement, tous les étudiants reçoivent un cours de base et les plus chanceux bénéficient, en seconde année de master, d'un cours approfondi. Ces derniers participent activement à ce congrès, sur le plan matériel en aidant à l'organisation et sur le plan scientifique en présentant, sous forme de posters, un travail personnel réalisé ces deux derniers mois au laboratoire de mathématiques appliquées. En recherche, plusieurs nouvelles méthodes, en particulier dans le domaine des tableaux multiples, ont été mises au point à Agrocampus. Elles sont rassemblées dans le logiciel libre FactoMiner, package R qui contient aussi les méthodes classiques. L'importance des package R n'est plus à démontrer : la première journée de ce congrès leur est dédiée.

A l'origine, l'analyse des correspondances fut mise au point pour... analyser des données. Et, fait singulier en statistique, sa diffusion doit beaucoup aux utilisateurs. De ce point de vue, ce congrès CARME est bien dans l'esprit originel : bon nombre de sessions ont comme dénominateur commun, et donc comme intitulé, un domaine d'application : parmi elles, une place prépondérante revient aux sciences sociales.

Ce n'est pas un hasard : les sciences sociales sont à la fois un domaine qui nous concerne tous et

Au moment où ces lignes sont écrites, en janvier 2011, d’un certain point de vue rien ne s’est encore passé. Et pourtant, que d’énergie a été dépensée pour que ce congrès puisse avoir lieu ! Je ne saurais trop remercier :

- la direction d’Agro campus, en la personne de Grégoire Thomas, Directeur général, qui a toujours soutenu la statistique en tant que discipline à part entière (et non uniquement au service des autres, ce qui ne va pas nécessairement de soi dans une institution centrée sur un domaine d’application) ;

- le pack logistique d’Agro campus : Anne Bourdeau (questions administratives), Stéphane Crespel (questions pratiques), Vinciane Marchais (inscriptions en ligne), Alain Bernardeau (questions administratives non euclidiennes) ; sans eux, rien n’est possible.

- le laboratoire de mathématiques appliquées : Karine Bagory, Marine Cadoret, David Causeur, Thibaut Dutrion, Magalie Houée, François Husson, Julie Josse, Maela Kloareg, Sébastien Lê, Elisabeth Lenaault, Marie Verbanck ; là est le cœur du cœur ;

- plusieurs amis au sein d’Agro campus : Gabriel Jalam (films), Yuna Blum (design) ; pour les " plus " indispensables.

Jérôme Pagès
14 janvier 2011

Brigitte Escofier (cliché Jean-Pierre Escofier).
Foreword

In 1960, Jean-Paul Benzécri was appointed Professor in Rennes. In 1963, he gave the first lecture on Correspondence Analysis. In the meantime, this singular approach to data analysis was born: this conference celebrates its fiftieth birthday.

Quickly, a team surrounded him. At the forefront was Brigitte Escofier, one of the first students and to become a decisive contributor. It is at this time that she discovered the transition formulas. Let’s think a moment about this: in 1961, principal component analysis has existed for almost thirty years and these formulas, mutatis mutandis applicable to PCA, were unnoticed! Subsequently, during thirty years, Brigitte enriched data analysis by many important researches (gathered in a book in 2003); naturally, a conference in Rennes on correspondence analysis is necessarily a tribute to her too.

In addition to Brigitte Escofier, let’s remember two other students of that first period: Marie-Odile Lebeaux and Brigitte Le Roux - Brigitte will talk about the "Rennes period" at our conference. In September 1965, Jean-Paul Benzécri went to Paris with his team.

In the late ’60s, Michel Kerbaol, formerly member of the Paris team, returned to Rennes; working in INSERM, he introduced correspondence analysis in the medical world. It is precisely Michel Kerbaol who gave what was probably the first data analysis lecture in Rennes, at Agrocampus in 1971. He aroused the enthusiasm of several students. One of them, yours truly, became assistant-professor one year later: he developed, around correspondence analysis, first a course then research, securely helped by the founders, who were working then in Paris. So, alphabetically, Pierre Cazes, Ludovic Lebart and Maurice Roux came several times to Rennes to bring us the good word. We are delighted to welcome them again in this Conference.

Today, data analysis is a principal component of Agrocampus. All the students attend a basic course and the luckiest ones, in the second year of their master degree, a comprehensive course. These last students actively participate in this Conference, practically by helping organization, and scientifically, by the way of posters, fruit of their personal work done these last two months in the Applied Mathematics department. Several new methods, in particular in the field of multiple tables, were developed at Agrocampus. They are gathered in the free software FactoMiner, an R package that also contains the classical methods. The importance of the R packages is now evident: the first day of this conference is devoted to them.

Originally, correspondence analysis was developed for ... analysing data! And its dissemination owes much to users. From this point of view, this CARME Conference is in the original spirit: many sessions have, as a common denominator, and therefore as a title, an application domain, and among them a prominent place is devoted to the social sciences.

It is not a coincidence: the social sciences are both an area which concerns all of us and a marvellous field for correspondence analysis. In fact, in 1991, CARME was initiated by Walter Kristof, at this time Professor at the Institute of Sociology of Hamburg, with Jörg Blasius and Michael Greenacres organizing the first meeting in 1991 in Cologne. Because of the success of this conference, Michael and Jörg have assured the continuity of the CARME Conferences, which have taken place three
times in Cologne (Correspondence Analysis in the Social Sciences 1991, Visualization of Categorical Data 1995, Large Scale Data Analysis 1999), then in Barcelona (CARME 2003) and Rotterdam (CARME 2007). So we also celebrate 20 years of CARME at our conference in Rennes.

At the time these lines are written, in January 2011, from a certain point of view nothing has happened. But what an amount of energy has been spent so that this Conference can take place! It’s nice to thank:

• the direction of Agrocampus, in the person of Grégoire Thomas, Director General, who always supports statistics as a discipline in its own right (and not only in the service of others, which is not necessarily evident in an institution centred on an application field).

• Agrocampus logistic pack: Anne Bourdeau (administrative matters), Stéphane Crespel (practical issues), Vinciane Marchais (online registration), Alain Bernardeau (non-Euclidean administrative matters); without them, nothing is possible.

• Applied Mathematics department: Karine Bagory, Marine Cadoret, David Causeur, Thibaut Dutrion, Magalie Houée, François Husson, Julie Josse, Sébastien Lê, Elisabeth Lenauld, Marie Verbanck; the heart of the heart.

• several friends in Agrocampus: Gabriel Jalam (movies), Yuna Blum (design); the indispensable "plus" factor.

Jérôme Pagès
14 January 2011

Michael Greenacre, Walter Kristof (who initiated the original CARME conference in 1991) and Jörg Blasius.
Application of Correspondence Analysis and Related Methods in Evaluation of Knowledge and Skills of Young People

Prof. Jozef Dziechciarz¹, dr Alicja Grześkowiak¹, dr Agnieszka Stanimir¹

1. Wroclaw University of Economics, Poland
   * jozef.dziechciarz@ue.wroc.pl, alicja.grzeskowiak@ue.wroc.pl, agnieszka.stanimir@ue.wroc.pl

Keywords: multiway correspondence analysis, multidimensional statistical analysis, knowledge and skills of young people

The analysis of knowledge and skills of a young person is an extremely important task in the educational process. The direct effect is the possibility to support the creation and orientation of educational and professional development paths. It is also crucial to properly identify relationships between the level of knowledge and various aspects of life (demographic, social, economic, etc.) from the perspective of both authorities constituting educational policy and teachers. With comprehensive information, teachers are able to help young people with the choice concerning further education and students can gain reliable information about their perspectives. This paper attempts to analyze the level of knowledge and skills of young people at regional level (Lower Silesia) and global level (Europe).

Variables describing the skills and competences as well as socio-economic factors are often nominal or ordinal. It is therefore natural to apply correspondence analysis and related techniques to identify associations between categories or relationships between variables.

References


Application of Correspondence Analysis and Related Methods to Evaluation of Knowledge and Skills of Students

prof. Jozef Dziechciarz
dr Alicja Grzeskowiak
dr Agnieszka Stanimir
Education Policy Background

Main goal of education reform (Poland, 1999):

“the improvement of quality in formal educational system and its adjustment to requirement of the knowledge-base economy, assuring the access to lifelong learning system for the entire society”

Source: Perspektywy gospodarki opartej o wiedzę w Polsce; World Bank Report, KBN, 2004

Principles of the reform:

• new structure of the education system, ranging from nursery school to doctoral studies
• an independent assessment, external examination system
• supervision separated from management
• new system of financing
• new professional status of the teaching staff linked with the promotion paths
Education Reform and PISA (Programme for International Students Assessment) Surveys

- 1999 – PISA trial version of survey, introduction of Education Reform in Poland
- 2000 – PISA first survey, participating Polish students approaching the end of compulsory education in the old system (very good baseline for comparisons)
- 2009 – fourth PISA survey; after ten years after the introduction of the reform Poland’s participation was summarized as:

“on average, the reform was associated with significant improvements. Poland improved its score in mathematics by 0.25 of a standard deviation, in reading, by 0.28 of a standard deviation, and in science, by 0.16 of a standard deviation”

Source: The Impact of the 1999 Education Reform in Poland, OECD Education Working Paper no.49, OECD, 2010
Scope of Research

• PISA results evaluation was published in 2009 but the newest information taken into consideration was from 2006

• the need for evaluation based on the up-to-date results

• national exam results from 2010 analyzed, similar competences as PISA test regarded

• special attention paid to differences according to gender, place of residence and to changes in time
Gymnasium as a Part of Education System in Poland

- Age of students: 14-16
- General education
- Compulsory school
- No specialization paths

Source: The Education and Assessment System in Poland, Central Examination Board, 1999
Regional Aspect of the Study

- south-western part of the country (two administrative territories)
  - area: 29 thousand km² (9.1% of total)
  - population: 3.9 million (10.2% of total)
  - number of students taking the exam: 40.9 thousand (9.2% of total)
External Exams Idea

Standardized Procedures

Nationwide comparability of results

Examination papers coded to provide objective assessment

Not relevant for graduation, relevant for further enrolment

CARME Conference, Rennes, 8-11 February 2011
Gymnasium Exam - Division

Humanities
- Polish language
- History
- Civic education
- Arts

max 50 points

Science
- Mathematics
- Physics
- Astronomy
- Chemistry
- Biology
- Geography

max 50 points
Evaluated Competences Areas

**Humanities**
- Reading and interpretation of texts
- Creation of own text

**Science**
- Finding and using information
- Application of terms and procedures
- Identifying and describing facts, relationships and dependences
- Application of integrated knowledge and skills to solve problems
General Remarks on Results (2010)

The bar chart shows the distribution of points scored in science and humanities. The median score for science is 22, and for humanities, it is 31. The chart indicates that a significant number of students in both fields scored around the median values.

Gender

- Female: 49.97%
- Male: 50.03%

Place of Residence by Commune Type

- Urban (cities more than 80 th inhabitants): 21%
- Urban (towns less than 80 th inhabitants): 25%
- Rural: 19%
- Urban-rural: 35%

- Urban commune: limited to the administrative town/city borders
- Rural commune: the territory does not comprise a town
- Urban-rural: includes a town and some villages
Multivariate CA Based on Burt’s Matrix

• results obtained in points were coded as excellent, good, moderate and weak

• two analyses for two administrative territories (Lower Silesia Region and Opolskie Region) were conducted

• quality of display in two-dimensional space was equal to 26% and 30%, respectively (to regions)
Multivariate CA Based on Burt’s Matrix - Results

Lower Silesia Region

Opolskie Region

Excellent results in big cities

Girls better results than boys in both science and humanities

Weak results associated neither with gender nor with commune type

As the analysis based on Burt’s matrix did not indicated the exact diversification among the gymnasium students the data were analyzed with application of multiway contingency tables.
Analysis Based on Multiway Contingency Tables

- *higher-order associations* were applied between commune type and gender and between knowledge area and exam score.

- Data table used in this analysis is 20x8, dimensionality of this table is 7.

- Percentage of explained inertia in the map is 96%.
Analysis Based on Multiway Contingency Tables – General Results

Boys from the capital of the region excellent in science

Girls from big cities excellent in humanities

Boys from rural and urban-rural communes moderate in humanities

No particular categories associated with weak results in humanities

No particular categories associated with weak results in science

Boys from other big cities good in science

Boys from the capital of the region excellent in science

CARME Conference, Rennes, 8-11 February 2011
Analysis Based on Multiway Contingency Tables: Lower Silesia Region – General Overview

CARME Conference, Rennes, 8-11 February 2011
Comparison of Results Obtained in 2003 and 2010

From better results in the capital of the region to weaker results in rural communes.

Various changes in science results:
- Cities: better results in 2010
- Urban-rural communes: weaker results in 2010
- Rural communes: weaker results in 2010

No particular differences in humanities results in 2003 and 2010

Weaker results in rural communes.
Comparison of Results Obtained in 2003 and 2010

Obtained better results than boys in both areas regardless of the place of residence.

Bigger dispersion of results according to gender was observed in rural and urban-rural communes.
Other Methods: Application of Factor Analysis

- problem – identification of common factors in the competences areas
- design - variants of analysis (issues): students divided according to gender and place of residence; each issue analyzed separately
- result - the competences areas produce one common factor in all variants
Further Research Directions

• PISA results from other European countries

• Exploring results in connection with information from background questionnaires concerning:
  ✓ educational carrier (educational aspirations, school changing, studying interruptions)
  ✓ ICT familiarity (availability of ICT, attitude and use of computers)
  ✓ parents characteristics (educational status, economical status, attitude to reading, perception of and involvement in child’s school, child’s school choice criteria)
Conclusions

- the application of correspondence analysis to the study of knowledge and skills of students allowed to identify the associations among analyzed characteristics (gender, commune type, competences areas, results of exam)
- the application of the multiway contingency table also allowed to identify the relationships in any configuration of variables as well as improves the quality of presentation in low dimensional space
- there are further research directions as taking into consideration results from more years or extending the study to other countries and other characteristics
Thank you very much for your attention