Workshop

Biplots
in Practice

INSTRUCTORS:

Prof. Michael Greenacre
Professor of Statistics
Universitat Pompeu Fabra, Barcelona

Prof. Jan Graffelman
Associate Professor of Statistics
Universitat Politècnica de Catalunya

October 4-6, 2010 • 9:00-18:00
Fundación BBVA
Palacio del Marqués de Salamanca
Paseo de Recoletos, 10
28001 MADRID

The **BBVA Foundation** promotes and disseminates knowledge for the benefit of society. It centers its activity on the promotion of scientific research and advanced training, focusing especially on the analysis of emerging issues in five strategic areas: Environment, Biomedicine and Health, Economy and Society, Basic Sciences and Technology, and Arts and Humanities. The BBVA Foundation designs, develops and finances research projects in these areas; facilitates advanced, specialist training through grants, courses, seminars and workshops; organizes award schemes for researchers and professionals whose work has contributed to the advancement of knowledge; and communicates and disseminates new knowledge through publications, debates and lectures. The **BBVA Foundation** is guided in its work by the five values of scientific objectivity, innovation, independence, transparency and a commitment to excellence.

Biplots in Practice

The biplot is one of the most useful and versatile methods for visualizing and interpreting multivariate data. Whenever research observations are recorded in the form of a rectangular table, such as a table stored in a spreadsheet or in a database, a biplot is usually possible to display the data in a graphical representation to facilitate understanding of the data and to reduce its complexity. The biplot extends the idea of a simple scatterplot of two variables to the case of many variables, with the objective of visualizing the maximum possible amount of information in the data. The biplot takes its name from the fact that it visualizes the rows and the columns of the data table in a common space.

This workshop explains how the biplot is defined in many different areas of multivariate analysis, notably regression, generalized linear modelling, principal component analysis, log-ratio analysis, various forms of correspondence analysis and discriminant analysis. In each situation the interpretation of the biplot is similar, but adapted to the different data types and objectives of the analysis. The orientation of the workshop is towards understanding how the biplot functions and how it is applied in practice. Because of the universality and wide application of this methodology, the workshop is aimed at a wide audience of researchers in all disciplines, as well as statisticians and computer scientists interested in this subject. During the workshop applications will be presented in many different fields of the social and natural sciences, including three detailed case studies where the biplot reveals structure in large complex data sets in genomics (where thousands of variables are commonly encountered), in social survey research (where many categorical variables are studied simultaneously) and ecological research (where relationships between two sets of variables are investigated).

Computers will be provided for use during the practical sessions and participants are urged to bring their own data sets. The free software program **R** will be used throughout the course, and will be introduced during the first practical, but participants are expected to have a reasonable introductory knowledge of this program (for example, writing and editing programs, importing data, installing packages, and producing graphical output). The **R** program as well as abundant documentation can be freely downloaded at www.r-project.org. There are many introductory texts on **R** as well as free online tutorials, for example: faculty.washington.edu/tlumley/Rcourse/R-fundamentals.pdf. Participants are also expected to have a working knowledge of Excel and of introductory statistics, up to linear regression. Participants will receive a copy of Michael Greenacre's new book *Biplots in Practice*.

```
women.ca <- ca(wom...
plot(women.ca, labels=0)
women.F <- women.ca$rowcoord %*% diag(women...
women.G <- women.ca$colcoord %*% diag(women...
text(women.F, labels=women.ca$rownames, pos=...
text(women.G, labels=women.ca$colnames, pos...
text(max(women.G[,1]), 0, "0.0571 (82.1%)"
text(0, max(women.G[,2]), "0.0030 (4.4%)"
...(women.ca, map="rowgreen", mass=c(F...
```

PROGRAM

October 4

9:00-10:30	Biplots – the basic idea • Regression biplots
10:30-11:00	Coffee break
11:00-12:30	Generalized linear model biplots • Multidimensional scaling (MDS) biplots
12:30-13:00	Coffee break
13:00-14:30	Reduced-dimension biplots • Principal component biplots
14:30-16:00	Lunch
16:00-18:00	Supervised practical

October 5

9:00-10:30	Log-ratio biplots • Correspondence analysis biplots
10:30-11:00	Coffee break
11:00-12:30	Multiple correspondence analysis biplots: between and within variables
12:30-13:00	Coffee break
13:00-14:30	Discriminant analysis biplots • Constrained biplots and triplots
14:30-16:00	Lunch
16:00-18:00	Supervised practical

October 6

9:00-10:30	Case study 1 (Biomedicine): Comparing cancer types according to gene expression arrays
10:30-11:00	Coffee break
11:00-12:30	Case study 2 (Socio-economics): Positioning the "middle" category in survey research
12:30-13:00	Coffee break
13:00-14:30	Case study 3 (Ecology): The relationship between fish morphology and diet
14:30-16:00	Lunch
16:00-18:00	Supervised practical with participants' own data

INSTRUCTORS

Prof. Michael Greenacre

Michael Greenacre is Professor of Statistics at the Universitat Pompeu Fabra in Barcelona and research collaborator with the BBVA Foundation. Apart from these workshops at the Foundation, he has given courses on multivariate analysis to marine biologists at the Polar Environmental Centre in Tromsø, Norway, as well as the Norwegian College of Fishery Science and groups in Italy, Iceland, the UK and South Africa. He is a member of the ARCTOS network for research on Arctic ecology, and is presently statistical adviser to a comprehensive research project on the ecology of the Barents Sea. His research interests are in the analysis of large data sets in the social and environmental sciences, and he has authored or co-edited seven books and numerous journal articles on correspondence analysis and data visualization. The BBVA Foundation has published the Spanish translation of his book on correspondence analysis, entitled *La Práctica del Análisis de Correspondencias,* and recently his book *Biplots in Practice* on which this workshop is based.

Prof. Jan Graffelman

Jan Graffelman holds a doctorate in statistics from the Universitat Politècnica de Catalunya (UPC), and a doctorandus degree in biology from the Rijksuniversiteit Groningen (The Netherlands). He holds a post as Associate Professor in the Department of Statistics and Operations Research at the UPC, with teaching responsibilities in mathematics, statistics and engineering. His fields of research concern biometry, multivariate analysis, statistical genetics and statistical computing.