

Luca Scrucca

Curriculum Vitae

PERSONAL INFORMATION

Born: Perugia, September 27, 1969
Actual position: Associate Professor of Statistics
Address: Department of Economics
Università degli Studi di Perugia
Via A. Pascoli, 20
06123 Perugia (ITALY)
Tel. +39 – 075 – 5855229
Fax +39 – 075 – 5855950
E-mail: luca.scrucca@unipg.it
Web Page: <http://www.stat.unipg.it/luca>

EDUCATION

2000 **Ph.D.** on “Statistical and Mathematical Methods for Economic and Social Research” at the Department of Statistics, University of Perugia, Italy, with a doctoral thesis titled:

Graphics for Studying Logistic Regression Models

Advisor: Prof. Gianfranco Galmacci.

2000 **Master of Science**, Statistics, School of Statistics, University of Minnesota, U.S.A. (advisor Prof. Sanford Weisberg).

1995 **Graduated** with a first-class honors degree (110/110 cum laude) on Political Sciences (curriculum on Quantitative Analysis for Social Sciences), University of Perugia, with a thesis on Social Statistics titled:

Electoral polls: methodological aspects and an evaluation of the recent Italian experience

Advisor: Prof.ssa Lina Brunelli.

RESEARCH INTERESTS

Mixture-models for clustering and classification; Gaussian Finite Mixture modelling; Statistical computing; Genetic algorithms; Dimension reduction methods; Classification and clustering of high-dimensional data; DNA microarray data analysis; Regression graphics; Linear and generalized linear models; Spatial clustering.

PUBLICATIONS

Under review

1. Fop M., Murphy T. B., Scrucca L. (2017) Model-based clustering with sparse covariance matrices. Submitted to **Statistics and Computing**. Pre-print available at <https://arxiv.org/abs/1711.07748>.
2. Scrucca L. (2017) A transformation-based approach to Gaussian mixture density estimation for bounded data. Submitted to **Computational Statistics & Data Analysis**.
3. O'Hagan A., Murphy T. B., Gormley I. C. and Scrucca L. (2015) On Estimation of Parameter Uncertainty in Model-Based Clustering. Submitted to **Computational Statistics**.

Journal articles

1. Scrucca L. and Raftery A. E. (2017) clustvarsel: A Package Implementing Variable Selection for Model-based Clustering in R. To appear in **Journal of Statistical Software**. Pre-print available at <http://arxiv.org/abs/1411.0606>.
2. Chirieleison C. and Scrucca L. (2017) Event Sustainability And Transportation Policy: A Model-Based Cluster Analysis For A Cross-Comparison Of Hallmark Events. **Tourism Management Perspectives**, 24, pp. 72–85. <https://doi.org/10.1016/j.tmp.2017.07.020>
3. Chirieleison C. and Scrucca L. (2017) Shaping Students' Attitudes towards Business Ethics and Corporate Social Responsibility: Education Versus Personal Perspectives, **Education**, 7(5), pp. 83-95. <http://article.sapub.org/10.5923.j.edu.20170705.01.html>
4. Scrucca L. (2017) On some extensions to GA package: hybrid optimisation, parallelisation and islands evolution, **The R Journal**, 9/1, pp. 187–206. <https://journal.r-project.org/archive/2017/RJ-2017-008>
5. Parisi A., Scrucca L., Desiderio J., Gemini A., Guarino S., Ricci F., Cirocchi R., Palazzini G., D'Andrea V., Minelli L., Trastulli S. (2017) Robotic right hemicolectomy: Analysis of 108 consecutive procedures and multidimensional assessment of the learning curve, **Surgical Oncology**, 26/1, pp. 28–36.
6. Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models, **The R Journal**, 8/1, pp. 205–233. <https://journal.r-project.org/archive/2016-1/scrucca-fop-murphy-etal.pdf>
7. Scrucca L. (2016) Identifying connected components in Gaussian finite mixture models for clustering. **Computational Statistics & Data Analysis**. 93, pp. 5–17.
8. Sahdra B., Ciarrochi J., Parker P. and Scrucca L. (2016) Using genetic algorithms in a large nationally representative American sample to abbreviate the Multi-dimensional Experiential Avoidance Questionnaire. **Frontiers in Psychology**, 7/189, pp. 1–14.
9. Scrucca L. and Raftery A. E. (2015) Improved initialisation of model-based clustering using Gaussian hierarchical partitions. **Advances in Data Analysis**

and Classification, 9/4, pp. 447–460.

10. Cencetti C., de Rosa P., Fredduzzi A., Minelli A., and Scrucca L. (2015) A statistical test for drainage network recognition using MeanStreamDrop analysis. **Geomatics, Natural Hazards and Risk**, 5–7, pp. 534–553.
11. Scrucca L. (2014) Graphical tools for model-based mixture discriminant analysis. **Advances in Data Analysis and Classification**, 8/2, pp. 147–165.
12. Morris K., McNicholas P., and Scrucca L. (2013) Dimension reduction for model-based clustering via mixtures of multivariate t-distributions. **Advances in Data Analysis and Classification**, 7/3, pp. 321–338.
13. Scrucca L. (2013) GA: A Package for Genetic Algorithms in R. **Journal of Statistical Software**, 5/4, pp. 1–37. <http://www.jstatsoft.org/v53/i04>
14. Chirieleison C., Montrone A., Scrucca L. (2013) Measuring the impact of a profit oriented event on tourism: the Eurochocolate festival in Perugia, Italy. **Tourism Economics**, 19/6, pp. 1411–1428.
15. Chirieleison C., Montrone A., Scrucca L. (2012) L'impatto degli eventi sul territorio: il caso Eurochocolate. **AUR&S**, 7-8/12, pp. 133–156.
16. Scrucca L. (2011) Model-based SIR for dimension reduction. **Computational Statistics & Data Analysis**, 55/11, pp. 3010–3026.
17. Scrucca L. (2010) Dimension Reduction for Model-Based Clustering. **Statistics and Computing**, 20 (4) pp. 471–484.
18. Scrucca L., Santucci A., Aversa F. (2010) Regression Modeling of Competing Risk Using R: An In Depth Guide for Clinicians. **Bone Marrow Transplantation**, 45, pp. 1388–1395.
19. Scrucca L. (2007) Class prediction and gene selection for DNA microarrays using sliced inverse regression. **Computational Statistics & Data Analysis**, 52, pp. 438–451.
20. Scrucca L., Santucci A., Aversa F. (2007) Competing risks analysis using R: an easy guide for clinicians. **Bone Marrow Transplantation** (2007) 40, 381–387.
21. Scrucca L. (2004) L'identificazione dei distretti industriali e delle aree di sviluppo economico omogenee e/o integrate. **Rivista dell'Agenzia Umbria Ricerche**, 1, pp. 71–90.
22. Scrucca L. (2004) qcc: an R package for quality control charting and statistical process control. **R News**, 4/1, June 2004, 11-17. <http://cran.r-project.org/doc/Rnews>
23. Scrucca L., Weisberg S. (2004) A simulation study to investigate the behavior of the log-density ratio under normality. **Communications in Statistics (simulation and computation)**, 33/1, 159–178.
24. De Angelis M., Scrucca L., Leandri M., Mincigrucci S., Bistoni S., Bovi M., Calabrese G., Pippi R., Parretti D., Grilli P., Colorio P., Fattorini M., Flamini O., Pacetti E., Travaglini A., Santeusano F. (2003) Prevalence of carotid stenosis in type 2 diabetic patients asymptomatic for cerebrovascular disease. **Diabetes, Nutrition & Metabolism**, 16/1, 48–55.
25. Scrucca L. (2002) Graphics for studying logistic regression models. **Statistical**

Methods & Applications, 11/3, 371–394.

26. Scrucca L. (2001) Nonparametric kernel smoothing methods. The sm library for Xlisp-Stat. **Journal of Statistical Software**, 6. <http://www.jstatsoft.org/v06/i07>
27. Scrucca L. (2001) A review and computer code for assessing the structural dimension of a regression model: uncorrelated 2D views. **Computational Statistics & Data Analysis**, 36/2, pp. 163–177.
28. Scrucca L. (2000) Assessing multivariate normality through interactive dynamic graphics. **Quaderni di Statistica**, 2, pp. 221–240.
29. Scrucca (1996) I sondaggi elettorali nella recente esperienza italiana. **Induzioni**, n. 12, pp. 59–81.

Book Chapters

1. Scrucca L. (2016) Genetic algorithms for subset selection in model-based clustering. In **Unsupervised Learning Algorithms**, Celebi M. E. and Aydin K. (editors), Springer International Publishing, pp. 55–70.
2. Chirieleison C., Scrucca L. (2016) CSR education in Italy. In **Social Responsibility Education Across Europe. A Comparative Approach**, Turker D., Altuntas Vural C., Idowu S. O. (editors), Springer International Publishing, pp. 139–159.
3. Scrucca L., Bar-Hen A. (2013) A model-based dimension reduction approach to classification of gene expression data. In **Advances in Theoretical and Applied Statistics**, Torelli N., Pesarin F., Bar-Hen A. (editors), Springer, pp. 221–230.
4. Scrucca L. (2011) A Geometric Approach to Subset Selection and Sparse Sufficient Dimension Reduction. In **New Perspectives in Statistical Modeling and Data Analysis**, editors Ingrassia S., Rocci R., Vichi M., Springer-Verlag, Berlin Heidelberg, pp. 569–576.
5. Bartolucci F., Scrucca L. (2010) Point Estimation Methods with Applications to Item Response Theory Models. In **International Encyclopedia of Education**, Penelope Peterson, Eva Baker, Barry McGaw (editors), 3rd edition, Oxford: Elsevier, Vol. 7, pp. 366–373.
6. Scrucca L. (2010) Visualization of model-based clustering structures. In **Data Analysis and Classification**, editors Palumbo F., Lauro C., Greenacre M., Berlin, Springer-Verlag, pp. 67–75.
7. Scrucca L. (2006) Regularized sliced inverse regression with applications in classification. In **Data Analysis, Classification and the Forward Search**, editors Zani S., Cerioli A., Riani M., Vichi M., Berlin, Springer-Verlag, pp. 59–66.
8. Scrucca L. (2005) La metodologia utilizzata per l'identificazione di aggregazioni sub-regionali omogenee. In *Verso una rilettura dei sistemi locali dell'Umbria: prime applicazioni di una nuova metodologia di clusterizzazione*, **Quaderni PRASSI - Agenzia Umbria Ricerche**, Cap. 2.
9. Scrucca L. (2003) Aree di sviluppo omogenee e/o integrate. Approcci e metodologie. **Paper dell'Agenzia Umbria Ricerche**, Collana: sviluppo e locale, n. 8, pp. 1–47.

Conferences and workshops

1. Scrucca L. (2017) Boosting Gaussian mixture models for classification. **Working Group on Model-Based Clustering**. Università degli Studi di Perugia, Perugia, July 17–21, 2017.
2. Scrucca L. (2016) On some recent improvements to mixture modelling estimation: Gaussian hierarchical partitions based on data transformations and evolutionary EM. **Working Group on Model-Based Clustering**. Université Paris Descartes, Paris, July 17–23, 2016.
3. Scrucca L. (2016) Poisson change-point models estimated by genetic algorithms. In **SIS2016, 48th Scientific Meeting of the Italian Statistical Society**. Salerno, 8–10 June, 2016.
4. Scrucca L. (2015) A quick tour of new functionalities in MCLUST version 5. **Working Group on Model-Based Clustering**. University of Washington, Seattle, July 19–25, 2015.
5. Galarini R., Moretti S., Romanelli S., Scrucca L. (2015) Informatics tools to manage validation data of multiresidue analytical methods. MASSA 2015, Alghero (Italy), June 2015.
6. Scrucca L. (2014) Model-based density estimation for bounded data. **Working Group on Model-Based Clustering**. University College Dublin, Dublin, July 20–26, 2014.
7. Scrucca L., Raftery A. E. (2014) Improved initialisation of model-based clustering using a Gaussian hierarchical partition. **MBC2 – Workshop on Model Based Clustering and Classification**. Catania, 3–5 September, 2014.
8. Scrucca L. (2013) Identifying connected components in model-based clustering. **Working Group on Model-Based Clustering**. Università di Bologna, July 21–27, 2013.
9. Scrucca L. (2013) On the implementation of a parallel algorithm for variable selection in model-based clustering. **CLADAG 2013. 9th Meeting of the Classification and Data Analysis Group. Book of Abstracts** (editors Tommaso Minerva, Isabella Morlini, Francesco Palumbo). Cleup. 9th Meeting of the CLassification and Data Analysis Group of the Italian Statistical Society, Modena, 18–20 September 2013.
10. Scrucca L. (2012) MCLUST version 4 session. **Working Group on Model-Based Clustering**. University of Guelph, July 16–20, 2012.
11. Scrucca L. (2012) Graphical tools for model-based mixture discriminant analysis. **MBC2 – Workshop on Model Based Clustering and Classification**, Catania, 6–7 September 2012
12. Scrucca L. (2011) Genetic algorithms for subset selection in model-based clustering. **Working Group on Model-Based Clustering**. University of Glasgow, Glasgow, July 17–23, 2011.
13. Scrucca L. (2011) Classification on a dimension reduced subspace. **CLADAG 2011, Book of Abstracts** (editors Paola Cerchiello, Claudia Tarantola), Pavia University Press, p. 47. 8th Scientific Meeting of the CLassification and Data Analysis Group of the Italian Statistical Society, Pavia, 7–9 September 2011.

14. Scrucca L. (2010) Some recent advances on dimension reduction for high-dimensional data. In **GfKL-CLADAG 2010, Book of Abstracts**, pp. 99–100. Joint meeting GfKI-CLADAG 2010, 8–10 September 2010.
15. Scrucca L., Bar-Hen A. (2010) A model-based dimension reduction approach to classification of gene expression data. In **SIS2010, Atti della XLV Riunione Scientifica della Società Italiana di Statistica**, CLEUP. Padova, 16–18 June, 2010.
16. Scrucca L. (2010) Genetic algorithms for subset selection in model-based clustering. In **SIS2010, Atti della XLV Riunione Scientifica della Società Italiana di Statistica** CLEUP. Padova, 16-18 June, 2010.
17. G. Saluti, Galarini R., Scrucca L. (2010) Multivariate Charts for Quality Control of Multiresidue Analytical Methods. In *Sixth International Symposium on Hormone and Veterinary Drug Residue Analysis*, Ghent (Belgium) 1-4 June 2010.
18. Scrucca L. (2009) A Finite Mixture Model Approach to Sliced Inverse Regression. **Working Group on Model-Based Clustering**. Institut Henri Poincaré, Paris, July 19–25, 2009.
19. Scrucca L. (2009) A Geometric Approach to Subset Selection and Sparse Sufficient Dimension Reduction. In **CLADAG 2009, Book of Short Papers** (editors Ingrassia S. e Rocci R.), Padova, CLEUP, pp. 235–238. VII Riunione Scientifica del Gruppo di Classificazione e Analisi dei Dati della Società Italiana di Statistica, Catania, 9-11 September 2009.
20. Scrucca L. (2008) A mixture model approach to dimension reduction in regression (Un approccio basato sui modelli mistura per la riduzione della dimensionalità nell'ambito della regressione). In **SIS2008, Atti della XLIV Riunione Scientifica della Società Italiana di Statistica**, pp. 233–240. Arcavacata di Rende, 25-27 June, 2008.
21. Scrucca L. (2007) Dimension reduction and visualization for model-based clustering. **Working Group on Model-Based Clustering**. Trinity College Dublin, Dublin, July 16–20, 2007.
22. Scrucca L. (2007) Visualization of model-based clustering structures. In **CLADAG 2007, Book of Short papers**, Lauro C. e Palumbo F. (editors), pp. 451–454. VI Riunione Scientifica del Gruppo di Classificazione e Analisi dei Dati della Società Italiana di Statistica, Macerata 12-14 September 2007.
23. Scrucca L. (2006) Dimension reduction and visualization for model-based clustering. In **COMPSTAT 2006, Book of abstracts**, 17th Symposium of IASC on Computational Statistics, Rome, 28 August - 1 September 2006.
24. Piersanti A., Galarini R., Tavoloni T., Scrucca L. (2006) Polychlorobiphenyl (18 congeners) in mussels from middle adriatic sea. In **Proceedings of 26th International Symposium on Halogenated Persistent Organic Pollutants - DIOXIN 2006**, Oslo, 21-25 August 2006, pp. 1951–1954. Organohalogen Compounds Vol. 68.
25. Scrucca L. (2005) Regularized sliced inverse regression with applications in classification. In **CLADAG 2005, Book of Short papers** (editors Zani S. e Cerioli A.), pp. 509–512. V Riunione Scientifica del Gruppo di Classificazione e

Analisi dei Dati della Società Italiana di Statistica, Parma 6-8 giugno 2005.

26. Scrucca L. (2004) Class prediction and gene selection for DNA microarrays using sliced inverse regression. In **Proceedings of the 19th International Workshop on Statistical Modelling (IWSM 2004)**, Firenze, 4-8 July 2004, pp. 259–263.
27. Scrucca L. (2003) On a test procedure for assessing a parametric approximation to a nonlinear term in semiparametric models. In **Proceedings ISI 2003 - International Statistical Institute** (13-20 August 2003, Berlin, Germany), Contributed Papers, Vol. LX, Book 2, pp. 397–398.
28. Scrucca L. (2002) Robust sliced inverse regression and outliers detection. In **Euroworkshop on Statistical Modelling Model Building and Evaluation**, Bernried, Germany, Oct 31-Nov 3 2002.
29. Porzio G.C., Scrucca L. (2002) Visualizing the correlation coefficient (Il coefficiente di correlazione: una sua rappresentazione grafica). In **SIS2002, Atti della XLI Riunione Scientifica della Società Italiana di Statistica**, Milano, 5-7 Giugno, 2002.
30. Attygalle D., Francis B., Scrucca L. (2002) Model coverage, forward searching and multiple outlier detection. In **17th International Workshop on Statistical Modelling**. Crete, Greece, 8-12 July, 2002.
31. Ranalli M.G., Scrucca L. (2002) Understanding influential observations in regression estimation of finite population means using graphs. In **COMPSTAT 2002, Book of abstracts**, 15th Symposium of IASC on Computational Statistics, Berlin, Germany, 24-28 August 2002.
32. Scrucca L. (2001) Visualizing and assessing partial correlation coefficients through added variable plots. In **Atti del Convegno S.Co. 2001, Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione**, Bressanone, 24-26 Settembre 2001, a cura di C. Provasi, Cleup Editrice, Padova.
33. Scrucca L. (1999) Dynamic probability plot for assessing multivariate normality. In **Atti del Convegno S.Co. 1999, Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione**, Venezia, 27-29 Settembre 1999, pp. 122–129.

Reports

1. Scrucca L., Raftery A. E. (2014) clustvarsel: A Package Implementing Variable Selection for Model-based Clustering in R. **Technical Report, Department of Statistics, University of Washington**, No. 629, pp. 1–27.
2. Fraley C., Raftery A. E., Murphy T. B., Scrucca L. (2012) mclust version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. **Technical Report, Department of Statistics, University of Washington**, No. 597, pp. 1–58
3. Scrucca L. (2006) Subset selection in dimension reduction methods. **Quaderni del Dipartimento di Economia, Finanza e Statistica**, Università degli Studi di Perugia, n. 23, Settembre 2006.
4. Scrucca L. (2005) Clustering multivariate spatial data based on local measures of spatial autocorrelation. An application to the labour market of Umbria. **Quaderni**

del Dipartimento di Economia, Finanza e Statistica, Università degli Studi di Perugia, n. 20, Ottobre 2005.

TEACHING

Statistics, Department of Economics, Terni, University of Perugia, 2016/17.

Statistical Data Mining, Department of Economics, Perugia, University of Perugia, 2016/17.

Mixture Models for Model Based Clustering and Classification, Ph.D. course, Doctorate in Mathematics, Computer Science and Statistics (UniFI, UniPG, INdAM), 2014/15 – 2016/17.

Statistical Computing, Department of Economics, Perugia, University of Perugia, 2014/15 – 2015/16.

Statistics, Faculty of Economics, Assisi, University of Perugia, 2011/12 – 2013/14.

Statistical Methods for Economics and Finance, Faculty of Economics, University of Perugia, 2010/11.

Statistical Quality Control, Faculty of Economics, University of Perugia, 2001/02 – 2010/11.

Statistics, Faculty of Economics, University of Urbino “Carlo Bo”, 2007/08 – 2009/10.

Statistical Methods for Market Research, Faculty of Economics, University of Urbino “Carlo Bo”, 2005/06 – 2006/07.

Teaching assistant for *Statistics I*, Faculty of Economics, University of Perugia, 2004/05 – 2007/08.

Teaching assistant for *Statistics II*, Faculty of Economics, University of Perugia, 2004/05 – 2007/08.

Laboratory of Statistics and Informatics II, Faculty of Economics, University of Perugia, 2003/04 – 2007/08.

Laboratory of Statistics and Informatics IV, Faculty of Economics, University of Perugia, 2004/05.

Short Course *Introduction to Xlisp-Stat*, Centre for Applied Statistics, Lancaster University, Lancaster (UK) 10–11 October 2001.

Teaching assistant, School of Statistics, University of Minnesota (USA), 1998–1999.

Software

In-depth knowledge of the programming environments for statistical computing and graphics *R/S-Plus* and *Xlisp-Stat*. Discrete knowledge of software packages *STATA*, *SPSS*, *Minitab*, *SAS*.

I contributed to the *R* Open-Source project as developer and/or maintainer of the following packages:

- `mclust`: model-based clustering, classification and density estimation via Gaussian mixture modeling;
- `clustvarsel`: variable selection for model-based clustering;
- `GA`: flexible general-purpose set of tools for optimization using genetic algorithms;
- `GAabbreviate`: create abbreviated forms of lengthy questionnaires (or other measures) using genetic algorithms;
- `msir`: model-based sliced inverse regression;
- `qcc`: quality control charts;
- `dispmod`: modelling dispersion in GLM;
- `forward`: forward search approach to robust analysis in linear and generalized linear regression models.
- other functions for *R* developed during my research activities are available on my web page.

Scientific societies and other information

Member of the Italian Statistical Society (SIS).

Member of the CLADAG (*Classification & Data Analysis Group*), research unit on "Classification and Data Analysis" of SIS.

Member of the *Working Group on Model-Based Clustering* directed by Prof. Adrian Raftery, University of Washington, Seattle.

Member of the Academic board of the PhD in "Mathematics, Computer Science and Statistics", University of Florence, University of Perugia, INdAM, Italy, 2013-present.

Member of the *Scientific and Institutional Committee* Umbria Agency for Research (AUR).

Member of the Academic board of the PhD in "Mathematical and Statistical Methods for Economics and Social Sciences", University of Perugia, Italy, 2004-2013

Member of the European Board of IASC (*International Association for Statistical Computing*), 2002–2006.

Referee activities for the following journals and conferences: *Statistics and Computing*, *Journal of Statistical Software*, *Computational Statistics & Data Analysis*, *Statistical Analysis & Data Mining*, *Journal of Computational and Graphical Statistics*, *R Journal*, *Journal of Classification*, *Journal of Statistical Computing and Simulation*, *Studies in Nonlinear Dynamics & Econometrics*, *Information Sciences*, *Statistical Methods & Applications*, *CLADAG*, *COMPSTAT*, *AISTATS*.

Perugia, November 30, 2017