

# Gaston Sanchez, PhD

gaston.stat@gmail.com  
gastonsanchez.com  
github.com/gastonstat  
linkedin.com/in/sanchezgaston

## EDUCATION

**Universitat Politecnica de Catalunya (UPC)**, Spain.

**PhD in Statistics** - Department of Statistics and Operations Research, July 2009.

Dissertation: *PATHMOX Approach: Segmentation Trees in Partial Least Squares Path Modeling*

**MS in Statistics** - Department of Statistics and Operations Research, June 2006.

**Instituto Tecnologico Autonomo de Mexico (ITAM)**, Mexico.

**BS in Actuarial Science** - Department of Actuary, June 2002.

Thesis: *Cluster Analysis with Web Mining Applications*.

## CERTIFICATES

**Universitat Oberta de Catalunya (UOC)**, Spain.

Certificate in Decision Support Systems, December 2007.

**Instituto Tecnologico Autonomo de Mexico (ITAM)**, Mexico.

Certificate in Brokerage Finance, August 2003.

**Instituto Tecnologico Autonomo de Mexico (ITAM)**, Mexico.

Certificate in Credit and Banking, July 2003.

## PUBLICATIONS

Sanchez G (2015) **The Saga of PLS**, (online text)

<http://sagaofpls.github.io>

Aluja T, Lamberti G, Sanchez G (2014) **Modeling with Heterogeneity**, in *Advances in Latent Variables: Methods, Models and Applications*, Carpita M, Brentari E, Qannari EM, (Eds), Springer.

Aluja T, Lamberti G, Sanchez G (2013) **Extending the Pathmax Approach to detect which constructs differentiate segments**, in *New Perspectives in Partial Least Squares and Related Methods*, Esposito, Russolillo, Abdi, Chin, Trinchera (Eds), Springer.

Lohmueller et al (2013) **Whole-Exome Sequencing of 2000 Danish Individuals and the Role of Rare Coding Variants in Type 2 Diabetes**. *The American Journal of Human Genetics*, 93 (6), 1072-1086.

Qin et al (2012) **A Metagenome-wide Association Study of Gut Microbiota in Type 2 Diabetes**. *Nature*, 490 (7418), 55-60.

Sanchez G (2013) **PLS Path Modeling with R** (ebook)

[http://www.gastonsanchez.com/PLS\\_Path\\_Modeling\\_with\\_R.pdf](http://www.gastonsanchez.com/PLS_Path_Modeling_with_R.pdf)

Sanchez G (2013) **Handling and Processing Strings in R** (ebook)

[http://www.gastonsanchez.com/Handling\\_and\\_Processing\\_Strings\\_in\\_R.pdf](http://www.gastonsanchez.com/Handling_and_Processing_Strings_in_R.pdf)

## CONFERENCES AND TALKS

Sanchez G. **Promoting Open Science**. Universidad Juarez Autonoma de Tabasco, Mexico, June 10, 2016.

Sanchez G. **From PLS Modeling to Multiblock Dilemmas**. McGill University, Montreal, July 6, 2015.

Hanafi M, Sanchez G. **Thinking Multiblock**. *PLS 2014 8th International Conference on Partial Least Squares and Related Methods*. CNAM, Paris. May 26–28, 2014.

Sanchez G. **Creating Arc Diagrams with R**. *Bay Area UseR Group Meetup*, Mountain View, CA. April 8, 2014.

Sanchez G. **STATS Software as seen from my experience in academia**. Applied Mathematics Department, Agrocampus-Ouest, Rennes (France). November 19, 2013.

Sanchez G. **Tips for Presentations in Data Analysis**. Oniris – National College of Veterinary, Food Science and Engineering, Nantes, France. September 18, 2013.

Aluja T, Lamberti G, Sanchez G. **Modeling with Heterogeneity**. *SIS 2013 Statistical Conference*, University of Brescia, Italy. June 19–21, 2013.

Sanchez G. **Data Analysis Visually Enforced**. Oniris – National College of Veterinary, Food Science and Engineering, Nantes, France. April 18, 2013.

Sanchez G. **Genetic Association Analysis**. Nielsen Group, University of California Berkeley. June 2012.

Aluja T, Lamberti G, Sanchez G. **Extending the Pathmox Approach to detect which constructs differentiate segments**, in *Proceedings of the PLS'12 International Symposium on Partial Least Squares and Related Methods*, Houston, Texas, USA. May 19–22, 2012.

Sanchez G, Aluja T. **A Simulation Study of Pathmox with Non-Normal Data**. 6th International Conference on PLS and Related Methods. Beijing, China. September 4–7, 2009.

Aluja T, Sanchez G. **Path Modeling Segmentation Segmentation Trees: A Simulation Study of Pathmox Sensitivity with Non-Normal Data**. *57th Session of the International Statistical Institute*. Durban, South Africa. August 16–22, 2009.

Sanchez G, Aluja T. **A Simulation Study of Pathmox Sensitivity**. *Eurisbis'09 European Regional Meeting of the International Society of Business and Industrial Statistics*. Cagliari, Italy. May 30–June 3, 2009.

Sanchez G, Aluja T. **A Simulation Study of Pathmox (PLS Path Modeling Segmentation Trees) Sensitivity**, *PLS'07 International Symposium on Partial Least Squares and Related Methods*, Matforsk, Aas, Norway. September 5–7, 2007.

Sanchez G, Aluja T. **A Simulation Study of Pathmox Sensitivity with Normal Data**, *Research Workshop on PLS Developments*, ESSEC Business School, Paris, France. May 15, 2007.

Sanchez G, Aluja T. **PLS Path Modeling Segmentation in Job Motivational Analysis**. *KNEMO'06 Knowledge Extraction and Modeling*. Anacapri, Italy. September 4–6, 2006.

## TEACHING AND ADVISING EXPERIENCE

**Lecturer, University of California Berkeley, CA.** (Jun 2015 – present)

- Stat 2: Introduction to Statistics
- Stat 133: Concepts in Computing with Data.
- Stat 259: Reproducible and Collaborative Statistical Data Science
- Stat 375: Professional Preparation – Teaching of Probability and Statistics

**Adjunct Professor, Berkeley City College, CA.** (Jan 2015 – present)

- Math 13: Introduction to Statistics and Probability.
- Math 253: Pre-Algebra.

**Oniris – National College of Veterinary, Food Science and Engineering, Nantes, France.** ( Apr 2013 – Mar 2014)

- Lectured Master's course on Latent Variable Path Models.
- Taught workshop on Creating R packages.
- Taught workshops on statistical methods using R.
- Mentored PhD and masters students on statistical programming.

- Reviewed dissertations of PhD students in the Department of Sensometrics and Chemometrics.

#### **Development of curricula for teaching Data Science with R**

(Available in personal website [www.gastonsanchez.com](http://www.gastonsanchez.com))

- Manipulating Data Tables in R.
- Getting Data from the Web with R.
- Creating R Packages in Windows.
- Handling and Processing Strings in R.

#### **Nielsen Group, University of California Berkeley, CA. (Summer 2012)**

- Seminar on Statistical Methods for Genetic Association Analysis.

#### **Program on Reproductive Health and the Environment, UC San Francisco (Fall 2011)**

- Workshops on R for data manipulation, graphics, and basic statistical analysis.

#### **Department of Statistics and Operations Research. Universitat Politècnica de Catalunya. (2006 – 2008)**

- Tutoring R and statistical programming for graduate students.

## **RESEARCH EXPERIENCE**

#### **Insight Data Science, Palo Alto, CA. (Jun 2014 – Sep 2014)**

##### **Data Science Fellow**

- Created Web app “[www.skincareprods.info](http://www.skincareprods.info)” to score personal care products based on health risk.
- Scraped data from the Good Guide and EWG using R, stored database in MySQL.
- Applied Regression Trees with cross-validation for computing health scores of personal care products.
- Front end developed with Python (Flask). Visualization with D3js.

#### **Oniris – National College of Veterinary, Food Science and Engineering, Nantes, France. (Mar 2013 – Mar 2014)**

##### **Postdoctoral Researcher**

- Built a prototype toolbox in R for multiblock data analysis methods.
- Lead seminars, training workshops, and tutoring on multivariate methods and data analysis with R.
- Created R packages `tester`, `turner`, and `matrixkit` for simplifying development of other packages.

#### **Nielsen Group. University of California, Berkeley. (Jun 2011 – Mar 2013)**

##### **Postdoctoral Researcher**

- Researched statistical tests used in genetic association studies, and created the R package `AssotesteR`.
- Developed the R packages `colortools`, `arcDiagram` (arc diagrams) and `DiscriMiner` (discriminant analysis).
- Continued as guest researcher between March 2014 and May 2014.

#### **Biology Scholars Program (BSP). University of California, Berkeley. (Feb 2011 – Nov 2012)**

##### **Research Analyst II**

- Established the roadmap for studying the scope of the Biology Scholars Program between 1994 – 2010.
- Carried out cluster analysis for segmenting BSP members to target specific audiences and needs.
- Conducted analysis to measure the impact of the program on its members. Applied statistical tests in R to compare the difference of the program members against students outside the program.

#### **Program for Reproductive Health & the Environment (PRHE), UC San Francisco. (Aug 2011 – May 2012)**

##### **Research Analyst II**

- Collaborated with PRHE’s scientists on bibliometric studies on Bisphenol A (BPA) datasets.
- Analyzed data from the biomonitoring study “Chemicals in Our Bodies”. Analysis in R with statistical tests to identify threshold levels for different chemicals (metals and PBDA’s).

#### **Laboratory of Information Analysis and Modeling, UPC, Spain. (Jul 2005 – Jul 2009)**

##### **Statistical Programmer**

- Research on advanced methods for structural equation modeling and segmentation analysis.
- Created R package “plsrm” for partial least squares analysis (most popular PLS open source package).
- Carried out a customer satisfaction analysis for clients of Sabadell Bank. Analysis performed in R using structural equation models. Provided recommendations to increase Perceived Brand Image and Customer Loyalty.
- Built path models for the Barcelona School of Informatics Alumni Satisfaction Survey to evaluate perceptions between educational experience and job experience. Findings used to modify the syllabus of Computer Science degrees.

**Department of Operations Research, IESE Business School, Barcelona, Spain. (Feb 2007 – Nov 2007)**

#### **Research Assistant**

- Analyzed data from the International Manufacturing Strategy Survey (IMSS) comprising three separate surveys (1997, 2001, 2005) from more than 1700 companies around the world.
- Drafting analysis with tables and graphs for Prof. Adrian Done’s book “Global Trends”.

## **PROFESSIONAL EXPERIENCE**

**Swimming Instructor** – *Canyon Pool, El Sobrante, CA. (Spring 2011)*

- Instructor for kids and adults basic swimming and water safety.

**Rope Techniques Instructor** – *Mexican Alpine Club (CAM), Mexico. (Jan 2011)*

- Taught courses on theory and practice of rope techniques levels I and II.
- Rope Essentials, Descending, Ascending, Hauling, Highlines, Self-rescue.

**Outdoors Skills and Rock Climbing** – *Mexican Alpine Club (CAM), Mexico. (Jan – Apr 2010)*

- Head instructor of Rock Climbing.
- Responsible for the Rock Climbing sessions of the Mountaineering Course LX.

**Climbing and Mountaineering** – *Mexican Alpine Club (CAM), Mexico. (Sep 2000 – Aug 2004)*

- Instructor of Outdoor Activities, Rock Climbing and Mountaineering.
- Oversaw 8 semesters of courses at CAM

**Junior Analyst** – *National Banking and Securities Commission (CNBV), Mexico. (May 2002 – Aug 2004)*

- Supervised and audited credit unions in Mexico.
- Visual Basic (Excel) programming for data collection utilities of projects.
- Management of the Credit Unions Supervision department’s datasets: data entry, data dictionary, combining data.

## **SKILLS**

### **Software**

- R (expert), Python (basic), SQL (basic), HTML, CSS, Jekyll, LaTeX, Markdown

### **Languages**

- Spanish (native), English (fluent), Catalan (fluent), French (conversational), Italian (conversational)