

R Fortunes of Jan de Leeuw: Collected Wisdom

July 7, 2016

¹ The term alternating least squares was apparently first used by De Leeuw. . .

—*J. de Leeuw*

Convergence Rate of Alternating Least Squares Algorithms, Unpublished note (2007)

² Making an inventory and classifying software necessarily reflects my bias, which is in academia [and] in research and development. Thus, I am not a user, and I don't care (much) about sales. If I say "student, I tend to mean "graduate student.

—*J. de Leeuw*

Presentation on the State of Statistical Software (2003)

³ Although some people refuse to admit it, the XLISP-STAT system is dead.

—*J. de Leeuw*

Presentation on the State of Statistical Software (2003)

⁴ [We] assume that there is a legitimate academic discipline called Statistics. This is, by no means, uncontroversial. Many scientists feel that they do not need statisticians to analyze their data, and many university administrators think that statistics is just an undergraduate course that students take to satisfy the general quantitative requirements.

—*J. de Leeuw*

Statistics and the Sciences. In I. Borg and P.P. Mohler, editors, Trends and Perspectives in Empirical Social Research (1994)

⁵ Statistics is defined as the science of building and evaluating tools for data analysis. The word tools is chosen on purpose here. It indicates that statistics is close to engineering, and in some instances perhaps even close to carpentry.

—*J. de Leeuw*

Statistics and the Sciences. In I. Borg and P.P. Mohler, editors, Trends and Perspectives in Empirical Social Research (1994)

⁶ Science is, presumably, cumulative. This means that we all stand, to use Newton's beautiful phrase, on the shoulders of giants. It also means, fortunately, that we stand on top of a lot of miscellaneous stuff put together by thousands of midgets. If we want to study a scientific problem we do this in the historical content, and we do not start from scratch. This is one of the peculiar things about the social sciences. They do not seem to accumulate knowledge, there are very few giants, and every once in a while the midgets destroy the heaps.

—*J. de Leeuw*

Statistics and the Sciences. In I. Borg and P.P. Mohler, editors, Trends and Perspectives in Empirical Social Research (1994)

⁷ It is very important to remember this analogy with tools: if you go to the statistician, you are going to buy a tool. Some people will try to sell you a tool which is far too expensive and elaborate for your purpose, and no matter where you go there will always be commercials. Bayesian commercials, frequentist commercials. Many statisticians will try to convince you that what you have been doing in the past is incorrect, that using tools other than the ones offered by them is irresponsible, or even incoherent. But these are all commercials, and they should be evaluated as such.

—*J. de Leeuw*

In J.J. Hox and J. De Jong-Gierveld, editors, *Operationalization and Research Strategy* (1990)

⁸ Mathematical theories are no more than maps of nature, indeed the connection between the equations and the physical reality is even more remote than that between the map and the country (which at least both exist in the physical world).

—*J. de Leeuw*

In J.J. Hox and J. De Jong-Gierveld, editors, *Operationalization and Research Strategy* (1990)

⁹ There are many interesting methodological problems with modeling observations as random variables, let alone sequences of random variables, but we ignore these problems in this paper. There are other interesting philosophical problems dealing with the notion of truth. We ignore those as well.

—*J. de Leeuw*

The Truth, Unpublished note (2004)

¹⁰ No matter where you go, there you are.

—*Buckaroo Banzai* (Quoted in Jan's email footer)
(2016)

¹¹ Many nights on the road and not dead yet — the end of autumn.

—*Basho* (Quoted in Jan's email footer)
(2008)

¹² That psychometricians have other types of confidence intervals than statisticians is nonsense.

—*J. de Leeuw* (On how to interpret psychometrician's confidence intervals)

Forum on design, sample-size, and power analysis, UCLA Statistical Consulting Center (September 2009)

¹³ The UCLA Electronic Statistics Textbook, from now on UCLA-EST, is an attempt to write a statistics textbook which is 1. freely available to everyone on the Internet; 2. independent of the level of the student, i.e., useful at the undergraduate, graduate, and postdoc level; 3. interactive, using graphics and demos; 4. complete, i.e., it covers most of statistical theory as traditionally taught.

—*J. de Leeuw*

The UCLA Electronic Statistics Textbook. Preprint Series 201, UCLA Department of Statistics (1995)

¹⁴ From the point of view of availability, the textbook is free and available to anybody on the net (with suitable hardware and software). Specifically, this means that you have to have an internet connection, a WWW browser (preferably graphic, preferably Netscape), and the necessary helpers on your client machine (in particular, Xlisp-Stat). Given the enormous investment of resources that is required, the Textbook is also open-ended in time. It will for ever be under construction. How quickly it proceeds will depend on support, and on the stability of tools and languages on the WWW. One thing has not changed: the major problem is still the writing of (browsable) text, plain words, and the unifying ideas and constructs behind the project.

—*J. de Leeuw*.

Presentation on the UCLA Statistics Server and Textbook (2000)

¹⁵ Code snippets. Perhaps. This is an old plan, but we (the board) cannot decide on the content and format. Ultimately this will only happen if one person or a small group of persons decides to start this, lobby for it, and manage it. In this respect it is the same as anything else in the world. I am not going to pull this one.

—*J. de Leeuw* (On whether to start publishing Code Snippets)

Presentation on the Journal of Statistical Software (2003)

¹⁶ JSS is trying to develop it's Code Snippets section. We have some snippets lined up and one published in the latest volume. . . If you have small chunks of code of obvious relevance to statistical computing (need not be in R) consider submitting it – why keep it to yourself? Just a matter of pasting a minimal amount of TeX into our templates and diving into our breathtakingly efficient review process.

—*J. de Leeuw* (In a call for Code Snippets)

R-help (May 2006)

¹⁷ I am not going to comment on The Bell Curve, because I did not read it. The considerations are the same as for most books I buy or read: one looks at reviews written by people whose opinion one values (in this case people like Goldberger, Gould, Hauser, Winship, Manski, Fienberg). These reviews are uniformly negative. As far as geneticists are concerned, I am told (again by people who know far more about this than I do) that there is no respectable geneticist who takes the M&H arguments seriously for even a single second. I have read basically everything on the first wave debate around 1920, and a lot on the second wave debate around 1980, I don't have the time and energy to go through a third wave of this scientific calvinism (JBS Haldane's term). Although of course obviously people should take it seriously and debunk it, in the same way as creationist science or orgone boxes or Yuri Geller should be taken seriously.

—*J. de Leeuw* (regarding The Bell Curve Wars)

Stat-I mailing list (November 1995)

¹⁸ Academic statisticians limited their conversations with psychometricians to snotty remarks about factor analysis, while psychometricians joked about the gothic irrelevances erected in the Annals of Mathematical Statistics.

—*J. de Leeuw* (Introduction to the special volume Psychometrics in R)

Journal of Statistical Software (May 2007)

¹⁹ What I see in [Big Data and Cloud Computing] is a lot of computer science, a lot of fads, a lot of ad-hoc work, and not much of a general rational approach. That may be unavoidable.

—*J. de Leeuw* (In response to a question on the effects of big data and cloud computing on statistics)

Interview with Ajay Ohri on decisionstats.com (July 2014)

²⁰ I decided in 1963 to major in psychology, mainly because I wanted to discover big truths about myself. About a year later I discovered that psychology and philosophy do not produce big truths, and that my self was not a very interesting object of study anyway. I switched to physics for a while, and minored in math, but by that time I already had a research assistant job, was developing software, and was not interested any more in going to lectures and doing experiments. In a sense I dropped out. It worked out fairly well, but it sometimes gives rise to impostor syndrome.

—*J. de Leeuw* (In response to a question on failure)
Interview with Ajay Ohri on decisionstats.com (July 2014)

²¹ Well, if one is not brilliant, one has to be a workaholic. It's hard on the family. I decided around 1975 that my main calling was to gather and organize groups of researchers with varying skills and interests – and not to publish as much as possible. That helped.

—*J. de Leeuw* (In response to a question on work habits)
Interview with Ajay Ohri on decisionstats.com (July 2014)

²² I started in 1968 with PL/I. Card decks had to be flown to Paris to be compiled and executed on the IBM/360 mainframes. Around the same time APL came up and satisfied my personal development needs, although of course APL code was difficult to communicate. It was even difficult to understand your own code after a week. We had APL symbol balls on the Selectrix typewriters and APL symbols on the character terminals. The basic model was there – you develop in an interpreted language (APL) and then for production you use a compiled language (FORTRAN). Over the years APL was replaced by XLISP and then by R. Fortran was largely replaced by C, I never switched to C++ or Java. We discouraged our students to use SAS or SPSS or MATLAB. UCLA Statistics promoted XLISP-STAT for quite a long time, but eventually we had to give it up. . . Of course the WWW came up in the early nineties and we used a lot of CGI and PHP to write instructional software for browsers. . . Generally, there has never been an computational environment like R – so integrated with statistical practice and development, and so enormous, accessible and democratic.

—*J. de Leeuw* (In response to a question on five decades of change in statistics)
Interview with Ajay Ohri on decisionstats.com (July 2014)