

Doing It Ourselves: The *SPSS Manual* as Sociology's Most Influential Recent Book

Barry Wellman

**Centre for Urban and Community Studies, University of Toronto
Toronto, Canada M5S 1A1 wellman@chass.utoronto.ca**

April, 1998

*Pp. 71-78 in Required Reading: Sociology's Most Influential Recent Books
edited by Daniel Clawson. Amherst, University of Massachusetts Press, 1998.*

Acknowledgements

My thanks to David Armor, Arthur Couch, Michael Schwartz and Philip Stone who taught me to use the computer for both quantitative and qualitative analysis at Harvard's Department of Social Relations in the mid-1960s, to Bev Meyrowitz/Bev Wellman who spent many hours at Harvard and Toronto helping me keypunch and consoling me when the runs did not work after twelve-hour turnaround, and to Joanne Daciuk who recently shared her memories of *SPSS* with me.

The Calcutta Revelation

This volume's presentation of influential books for sociology is mostly concerned with weighty tomes discussing world-historic subjects. By contrast, I believe that the most influential books have been those that have empowered sociologists through precept or example. That is why I propose the *SPSS manual* (Nie, Bent and Hull 1970) as our most influential book, for it was the *SPSS* statistical package that in the early 1970s revolutionized how sociology was done.¹

It was only in Calcutta in 1986 that I fully appreciated how empowering the *SPSS* revolution has been. Social networks had brought me to Calcutta. I had been corresponding with Suraj Bandyopadhyay's research group at the Indian Statistical Institute. As I was going to be in India for the World Congress of Sociology, why not come to the ISI and work together?

At a small workshop, I showed my techniques for analyzing personal community networks by means of linking information about the characteristics of network members and their ties (Wellman 1992). "No problem!" I announced in breezy AmeriCanuck, "*SPSS* or *SAS* can do this easily."

The Indian sociologists looked stricken, a *Maalox™ cum Imodium™* moment. "We do not have *SPSS* or *SAS*," they said wistfully. Because of strained relations with the United States, American computer technology or software were not readily available in India.² Instead, the ISI had a Russian-made mainframe computer with a Russian operating system and statistical package. Only a few initiates at the computer center knew how to work the system: a Marxian *cum* Weberian instance of the over-bureaucratization of statist societies. To do one statistical analysis on the computer, ISI members had to queue for weeks and plead for the attention of the priestly group serving the mainframe. If they got a command wrong or wanted to do another analysis, they went to the back of the queue again. Hence, only very crucial or large analyses were done on the computer. The researchers did most analyses by hand or by calculator, with the help of assistants.

The *SPSS* Revolution

¹Early users will recall that the original blue *SPSS* manual (and subsequent maroon manual, 1975) also were fairly weighty tomes. Fortunately, there was only a single volume at that time. Today, there are too many.

²A *rapprochement* between the U.S. and India has changed this situation. Whatever the new dependency relationship this entailed, American software is now available.

My realization was the unlikely product of the wisdom of Norman Nie (the “father” of *SPSS*) and Harold Garfinkel (the maven of making the taken-for-granted visible). Jointly, their spirits guided my appreciation of how user-friendly statistical packages had enabled social scientists to become masters of our own analyses. Such packages have enabled most of us to do our own computer-based statistical analyses instead of being forced to rely upon high priests of the Great Machine. We no longer have to queue and beg an expert who possess the rare knowledge of how to get a user-unfriendly statistical package to work. Instead, we are now routinely able to do complex analyses from a number of perspectives instead of just talking about complex processes but measuring simple relationships. We can easily check a variety of alternative causes and correlates, and we can take endless alternative views of our data. We revel in our post-P(e)arsonian ability to try dozens of complex procedures, to view things in ten different ways, to obsessively clean and re-weight data, to transform the intractable, to hunt down pesky residuals, and to apply once-obscure statistical tests with exotic-sounding names. It is an orgy of do-it-yourself quantitative analysis.

I focus on *SPSS* because it was the first widely implemented, easy-to-use, reasonably comprehensive package. To be sure, there were earlier ones such as *BMD* (whose company has recently been bought by *SPSS*), *Osiris* (at the University of Michigan) and *Data-Text* (which I helped code a bit at Harvard in the mid-1960s). Indeed, I use *SAS* now because it is more network-analysis-friendly. But *SPSS* came to the social sciences well before *SAS*, and it was *SPSS* that first proliferated in the developed world’s computer centers. It was *SPSS* that made the revolution and which remains the most user-friendly of the main packages. Undergraduates at the University of Toronto now use it soon after starting their first statistics course, easily doing the kinds of analyses that senior Indian scholars had been blocked from doing a decade earlier,

I do not want to commit the “presentist” fallacy of asserting that no statistical analyses were done before statistical packages were available. Soon after coming as a new graduate student to Harvard in 1963, I was taken to the basement of Emerson Hall to view with awe the very counter-sorter that Samuel Stouffer had used. In my mind’s eye, I could see the statistical analyses of *The American Solider* (1950) dropping into the sorter pockets in front of me. I was even initiated into the priesthood and learned how to rewire this very same machine all by myself. Of course, folks like Sam Stouffer and Paul Lazarsfeld did wonderful work using counter-sorters, but they were rare giants. In those pre-computer days, most people could do only limited analyses using counter-sorters and other IBM machines. Now, we do not have to be giants. We can be ordinary people, using statistical packages to play with data and examine hundreds of analytic possibilities.

The Costs of Revolution

Every revolution has its victims. This one has had several, brought forth by *SPSS*’s triumph. Not only did *SPSS* become a tool for empowerment, it fostered a world view:

1. SPSS has made statistical analysis so easy that theory and common sense have sometimes fallen by the wayside. I have no doubt that Lazarsfeld and Stouffer would have

happily embraced statistical packages, but I am also confident that they would have urged the careful specification of variables and relationships beforehand. With statistical packages and multivariate routines, it is easy to pour in a heap of variables into the regression and stir wildly to see what sticks to what. Many spurious and silly things have come out of such stews.

2. SPSS tilted the sociological playing field so much toward statistical analysis that other modes of inquiry, such as field work, have become neglected. Because quantitative analysis became much easier than the qualitative analysis of texts, it became much more popular. Moreover, it became the only prevailing orthodoxy in almost all major research universities. Inevitably, students came to believe that quantitative procedures were the best, and perhaps the only, road to the truth. Until recently, there have been no easy ways to do computer-based analyses of field situations. When I used *The General Inquirer* in the 1960s, (Stone, et al. 1969), I needed six months of preparatory work and produced only frequency counts of concept categories. *Data-Text* never implemented the latter part of its name, and remained only a statistical package until it was supplanted by the more widely-usable *SPSS* in the 1970s.. The result has been that field work has remained hard and imprecise, generally requiring lots of transcribing and hand sorting of notes. A cursory look at sociology journals and books shows that the balance has swung much more to statistical analysis since statistical packages proliferated in the 1970s.

Many qualitative analysts suspect a quantitative plot against them. Be that as it may, it is clear that the current orthodoxy of statistical analyses is related to their greater ease of use and the greater availability of quantitatively-coded data sets. It is only very recently that microcomputer textual analysis packages such as *Nud.ist* have arrived to help those working with transcripts and other texts (Miles and Huberman 1994; Weitzman and Miles 1995). *Nud.ist* 4.0 now links with *SPSS*, and the *SPSS* organization introduced *TextSmart* 1.0 in 1998 to capture the meaning of open-ended survey responses. We are starting to have the kinds of powerful hybrid analyses that *Data-Text* promised thirty years ago.

3. The proliferation of statistical packages led to survey research perspectives dominating sociological research. *SPSS* was born and raised at the National Opinion Research Center, a survey shop *par excellence*. Surveys are almost always based on a random sample whose very essence is that each individual must be treated as a separate unit of analysis or else, the sample would be biased. Yet as soon as the unit of analysis becomes the discrete individual, crucial information is lost about the structure of the social system in which this person is embedded. The very assumption of statistical independence, which makes standard statistical analyses so powerful detaches individuals from social structures and forces analysts to treat them as parts of a disconnected mass. This “methodological individualism” (Coleman 1958:28) has shifted analyses away from looking directly at social structures and social processes to efforts that try to infer structure and processes from the cross-classified, aggregated characteristics of analytically-disconnected individuals. Each record — which usually means each individual — is treated as a separate entity consisting of variables measuring discrete social characteristics (e.g., age, socioeconomic status, attitudes). Yet aggregating each person’s (or organization’s) characteristics independently obscures or destroys structural information. “Individuals do not act

randomly with respect to one another. They form attachments to certain persons, they group together in cliques, they establish institutions” (Coleman 1964: 88).

Of course, sociologists try to infer something about social systems from the multivariate analysis of individual-level data. But analysts taking this approach can only study social structure indirectly by organizing and summarizing numerous individual covariations. Analysts are forced to neglect social properties that are more than the sum of individual acts and concentrate on the attributes that discrete individuals possess. They cannot directly study flows of information or other resources; discover clusters, cleavages or overlapping networks; or reveal underlying role structures. One partial solution is to use social network analytic programs, such as *UCINet*, to analyze structure, but such programs inherently lack the statistical firepower of mainstream packages such as *SPSS*. Another partial solution is to use hierarchical linear modeling, available in the *HLM* package for multivariate analysis.

Collateral Possibilities

Despite such caveats, the statistical-package revolution has been good for sociology, expanding our scope and empowering our efforts. What we started has spread to other fields, so much so that I was startled to read in an article about market-research “data mining” (talk about being atheoretical!) in *PC AI* (a magazine for computer scientists interested in artificial intelligence): “I suspect that undergraduates new to other disciplines (economics, business, and sociology, for example) could use other features of *SPSS* to great advantage (Schmuller 1996: 31). When I emailed him, the statistician-author was unaware that sociologists had ever heard of *SPSS*, and he was astonished to learn that we had been there and done that more than twenty-five years ago.

I do not want to be snobbish: The attention of the marketing mavens is a possible augury for statistical packages becoming integral parts of office suites. This may lead to future creative linkages of statistical analysis packages with other tools, such as those for graphical or textual analysis. Already, the development of statistical packages (including *SPSS* and *SAS*) for microcomputers has liberated sociologists from their thirty years of dependence on central computing centers.

While the *SPSS Manual* (and those for other statistical package)s have been the most influential books for sociology, there have other empowering books. Among my other candidates for consideration would be:

--Hubert Blalock's *Social Statistics* (first edition, 1960), an exemplar of the sociologically-minded statistics books which clearly told recent generations of sociologists what to do with *SPSS*.

--Blalock's *Causal Models in the Social Sciences* (1971), and Peter Blau & Otis Dudley Duncan's *American Occupational Structure* (1967). In tandem, they disseminated the lore of nuanced multivariate analysis.

--S.D. Berkowitz's *An Introduction to Structural Analysis* (1982) for its integrated presentation of social network analysis.

--Barney Glaser and Anselm Strauss' *The Discovery of Grounded Theory* (1967) which showed how to do qualitative analysis systematically.

--Charles Tilly's *The Vendée* (1964) which showed how to use a sociological perspective to interview the past.

And as an article-writing book reader, I would love to see a compilation of influential journal articles, for each substantive field as well as for sociology in general.

Choosing influential books is a nice game because the fun is more in the playing than in the outcome. I have taken a stand here in favor of empowering tools as the most influential sociological development in recent decades. Which is more important? The findings or the tools that enabled us to make them -- and many more? Which should we celebrate more? Copernicus' 16th-century hypothesis of the solar system or Galileo's 17th-century invention of the telescope that enabled scholars to understand it clearly? It is an unresolvable dialectic between knowing what to look for and knowing how to find something. But if pressed, I would vote for the toolmakers because they give us the eyes to see things.

References

- Blalock, Hubert. *Social Statistics*. New York: McGraw-Hill.
- Blalock, Hubert. 1971. *Causal Models in the Social Sciences*. Chicago: Aldine Atherton.
- Blau, Peter and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: Free Press.
- Coleman, James. 1958. "Relational Analysis." *Human Organization* 17: 28-36.
- Coleman, James. 1964. *Introduction to Mathematical Sociology*. New York: Free Press.
- Glaser, Barney and Anselm Strauss. 1967. *The Discovery of Grounded Theory*. New York: Aldine.
- Miles, Matthew and A. Michael Huberman. 1994. *Qualitative Data Analysis*. 2d ed. Thousand Oaks, CA: Sage.
- Nie, Norman, Dale Bent and C. Hadlai Hull. 1970. *SPSS: Statistical Package for the Social Sciences*. New York: McGraw-Hill.
- Schmuller, Joseph. 1996. "Statistical Modeling Takes Off: SPSS Advanced Statistics 6.1." *PC AI* 10 (Sept.): 28-31.
- Stone, Phillip J., Dexter Dunphy, Michael Smith and Daniel Ogilvie 1969. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press.
- Tilly, Charles. 1964. *The Vendée: A Sociological Analysis of the Counterrevolution of 1793*. Cambridge: Harvard University Press.
- Weitzman, Eben and Matthew Miles. 1995. *Computer Programs for Qualitative Data Analysis*. Thousand Oaks, CA: Sage.
- Wellman, Barry. 1992. "How to Use SAS to Study Egocentric Networks." *Cultural Anthropology Methods* 4(2): 6-12.