Joel H. Levine

Exceptions Are the Rule

An Inquiry into Methods in the Social Sciences

New Perspectives in Sociology

New Perspectives in Sociology

Charles Tilly and Scott McNall
Series Editors

Exceptions Are the Rule: An Inquiry into Methods in the Social Sciences, Joel H. Levine

FORTHCOMING IN 1993

Criminological Controversies, John Hagan, A. R. Gillis, and David Brownfield

Freudians and Feminists, Edith Kurzweil

New Perspectives in Sociology

Published in 1993 in the United States of America by Westview Press, Inc., 5500 Central Avenue, Boulder, Colorado 80301-2877, and in the United Kingdom by Westview Press, 36 Lonsdale Road, Summertown, Oxford OX2 7EW

A CIP catalog record for this book is available from the Library of Congress.
ISBN 0-8133-1646-4
ISBN 0-8133-1647-2 (pbk.)

Printed and bound in the United States of America

The paper used in this publication meets the requirements of the American National Standard for Permanence of Paper for Printed Library Materials Z39.48-1984.

10 9 8 7 6 5 4 3 2 1

Exceptions Are the Rule

Exceptions Are the Rule

An Inquiry into Methods in the Social Sciences

Joel H. Levine
Dartmouth College

Westview Press
Boulder • San Francisco • Oxford

We still carry the historical baggage of a Platonic heritage that seeks sharp essences and definite boundaries. . . . This Platonic heritage, with its emphasis on clear distinctions and separated immutable entities, leads us to view statistical measures of central tendency wrongly, indeed opposite to the appropriate interpretation in our actual world of variation, shadings, and continua. In short, we view means and medians as the hard "realities," and the variation that permits their calculation as a set of transient and imperfect measurements of this hidden essence. . . . But all evolutionary biologists know that variation itself is nature's only irreducible essence. Variation is the hard reality, not a set of imperfect measures for a central tendency. Means and medians are the abstractions.

— Stephen J. Gould, "The Median Isn't the Message"
Discover, June 1985

Students Should Ask

To Nicholas C. Mullins

with thanks for the contributions of Joel R. Reidenberg

First, what is a "student"? My pocket dictionary tells me a student is "one who attends school." Wrong! A student is one who actively attempts to acquire knowledge — not as a librarian acquires a book and not as one who *attends* school to acquire a degree — but for understanding. "Student" includes many of those "who attend school" as well as those who stand before them at the lectern.

That being said, there are questions I would like to whisper in the ear of students who attend statistical methods classes, particularly the good students, questions they should ask. Such mischief might enliven these classes. More important, it would alert the student to discriminate between ends and means, between what we wish to accomplish as a science and what we are able to accomplish, so far. The end is understanding, not warm sympathy that "understands" human pain, but understanding that is communicable, testable, valid, and offers real hope. The end is understanding that provides the base for affecting the human world, just as physics and biology provide the base for engineering and medicine.

That's the goal. Statistics is one means. I hesitate to criticize it because too many who criticize it reject statistics, and science and rationality as well. That's foolishness. That is unilateral disarmament in the face of the unknown, disavowing the best resources of the human mind. But all too often we fail to criticize statistical means with respect to scientific ends — what do they say? How do they help me understand? Answering that is not the statistician's job, nor is it the job

of a sociological "methods" course. But both should alert you to the questions.

I would have a student ask: "Professor! When you tell me that education predicts income, what does that tell me? How does it work?" The question would give your methods professor the all-important opportunity to pass the buck, reminding the student — one who learns — that that is a question for research, for theory, and ultimately, perhaps, for other mathematical models.

The two parts of this book divide most easily into part one, summary statistics, and part two, distributions — the stuff not usually described by summary statistics. Equally well I could divide it into part one, "Everything you thought you knew is wrong," and part two, "This is where I look for answers." That is flip, inaccurately critical, and overly optimistic, but it should make the point. The point is that "modern" methodology hasn't accomplished a whole lot, not when it is measured against the unknowns of sociological research. Do we yet have one precise testable statement we would care to call a "law"? So let's relax and try to do better; there is not a whole lot to lose by releasing the death grip of one-generation-old traditional methods.

Perhaps the difficulty in relaxing, playing with new questions, is that the advances of the last generation were hard won against real opposition, forcing "methodology" into a defensive posture. Quantitative social science was "jump started" after World War II by a generation of determined social scientists, mathematicians, statisticians, foundations, and, yes, federal bureaucrats, motivated, I like to think, by a feeling of great human need for a real social science. They won the battle. Students on both sides of the lectern can now handle mathematical reasoning. They won, but now the agenda is to get on with the analyses for which the older generation created us — even when we find we must gently revise or set aside the techniques and the style of that older generation. Sometimes they proffered their command of probabilities and significance and Greek symbols as stigmata proving they had established a science. They proved no such thing, but the rituals do seem to have done their job on the opposition. They won. Now, to work . . .

Joel H. Levine

Contents

EPILOGUE

Part One
Summary Numbers

1

Introduction: Numbers for Religion and Politics

Philosophy is written in that vast book which stands forever open before our eyes, I mean the universe; but it cannot be read until we have learnt the language and become familiar with the characters in which it is written. It is written in mathematical language, and the letters are triangles, circles and other geometrical figures, without which it is humanly impossible to comprehend a single word.

— Galileo Galilei, *Il Saggiatore*[1]

How do you put numbers on things like religion and politics? Social science cannot yet describe political position with numbers and equations. It cannot yet make a statement like "Senator X is 2.5 units to the left of Senator Y and moving left at a velocity of one unit per year." How do you put numbers on such things and describe them by the equations of simple mathematics? Such questions are a torment to the practicing social scientist. We deal with messy variables, variables like politics, religion, and race. We deal with roles like "parent" and "child," with organizations that may be "hierarchical" or "democratic"; we deal with grand things like "social order" and "revolution." The variables that social scientists must deal with and the theories we construct seem, at times, generically different from the physical continua and the measurable intervals of conventional physics.

These chapters attempt to put numbers and equations on such things and, more broadly, to introduce concepts that get the job done. To

1. Cited by James R. Newman in "Commentary on Galileo Galilei," *The World of Mathematics,* Volume 2, edited by James R. Newman, Simon and Schuster, New York, 1956, p. 731.

do that I have to accomplish two things: First, I have to create doubt. These days social scientists trust numbers and equations, and that's not altogether a good thing. Everyone learns the methodological canon — learns to fit lines to data, learns a little about correlation, and learns to use a few elementary statistical techniques. Beyond that, it is supposed that "methods" can be relegated to specialists while the rest of us focus our attention on the substance and content of the science. Not so.

The reasons for doubt are plain to every new student. Take something simple like the generalization "Education pays off" — literally, with dollars. That's true. On the average the well-educated are, in fact, better off economically than the rest. But when you look at the facts behind that statement, when you look at the data and the equations, they are a shock. What you expect to see is a line: Go out and interview a few thousand people. Ask them about their educations. Ask them about their incomes. Plot their data on an ordinary piece of graph paper and you expect to see a line — so many years, so many dollars. More years, more dollars. But that's not what you get. Not at all. What you see when you graph the data is a cloud, not a line. The exceptions, the deviations from the rule, are more than the occasional millionaire who dropped out of pre-school. The exceptions are large and frequent. That's the general case in social science: One of the pioneers in sociological theory, now retired, used to preface some of his more risky assertions about social systems by saying, "It would not be altogether incorrect to state . . . ," and then he would proceed with his generalization of the day. The phrase should be applied to our methods: "It would not be altogether incorrect to state . . . that more highly educated people have higher incomes, on the average." Exceptions are the rule. The genius of the methodological canon is that it is able to extract central tendencies from the midst of such variation. When you translate textbook methods from the textbooks to practice, with real data, they don't tell us very much about the data. Specialists in "methodology" already know this and are amazed at the confidence, and non-involvement, that other people show for their work. But because people trust the basics and, trusting them, set them aside, I have to begin with doubt. I have to bring out the canon, part of it, and argue in language as plain as I can muster that the old standbys don't work, not easily, not well. There is work to be done.

That's my first task, creating doubt — not about mathematics or science, but about the basic methodological canon. The second task is, of course, to provide answers, or progress — where I can. How do you describe the messy data of social science with numbers and equations? To give you a preview, imagine two variables, height and weight. Height and weight are not the kind of thing social scientists worry about, but they are the kind of thing we practice on, so bear with me: I collect data describing the heights and weights of a few thousand people and then organize the data in a table in which columns specify the height and rows specify the weight. The table shows me how many six-footers weighed 200 pounds, how many weighed 210 pounds, how many weighed 220 pounds, and so forth. Very good — I'm ready to analyze the relation between height and weight. But before I begin let me make the problem more interesting: It turns out that my research assistant (that's me) lost part of the data. In fact I lost all the labels. I still have all the counts, five people of this kind, ten people of that, but I no longer know which people were six foot five and which ones were five foot six or how much they weighed. Artificially, my well-behaved data have become messy, and here's the challenge: Can I still analyze height and weight? There is an order to these variables, but I don't know it. There is a scale to these variables, but it's lost.

Surprisingly, I can come very close to a normal analysis, even without the labels, and the trick is to use the variation. Just looking at the data I can probably spot which rows are similar to each other. The numbers of six-footers weighing 200 pounds, 210 pounds, and 220 pounds are undoubtedly similar to the numbers for people who are five foot ten. And the numbers for those who are five foot ten are surely similar to the numbers for people who are five foot eight. Just from the numbers, I can probably reconstruct the correct order by height. And again, just from the variation, I can probably reconstruct the order by weight. So even without the labels I can probably get the rows and columns of the table back in order.

That's just the beginning. I started with unknown intervals and reconstructed the order. If I take the next step and quantify "similarity" with an equation, then I can estimate that "five foot ten" is halfway between "six foot" and "five foot eight." Without knowing X and Y I could use the equation for similarity to estimate the scale and then

graph the data. I could tell you whether or not the relation between X and Y was linear and I could estimate the strength of the correlation. And what of the equation that I used to quantify "similarity"? If I'm careful, then that too can be tested.

Not bad, considering that I lost the measures of height and weight. But who cares about height and weight? I don't, except as an example. Now for the payoff: Suppose I lied to you. "Really" my unlabeled table was not about height and weight at all. Instead, it was a table describing political preference and religion, messy variables. If I had told you that, in advance, then you wouldn't have tried to recover the unknown intervals. But you did: You put the categories in order and estimated their intervals, including a scale for X and Y. You verified that the pattern was linear and estimated the correlation. *A priori,* political preference and religion are not supposed to behave that way, but they did in this hypothetical example. If such a thing happened, it would be best to believe, or consider, the evidence of your eyes and doubt *a priori* ideas about what can and cannot be done with the variables of social science. And such things do, in fact, happen. The details are different. The equation changes from problem to problem. And the "fit" is always a matter of degree. But it happens.

You will have to trust me on that, for the moment. But now let me re-state my opening question: Is there something different about social science? Is it limited to classification and counting, classifying people as "Democrats" and counting their number? I think not. But there is a difference among the sciences, and it lies in practical matters, like: How do you get started? How do you get your science off the ground? In physical science the first step seems obvious, at least in hindsight. You could start with familiar things that you could see and touch. You could assign numbers to things that were accessible to the senses — to length, to mass, and, with more difficulty, to time — and then get on with the labor of genius that was eventually able to discover "elementary" physical law. In social science, by contrast, the most familiar objects are the ones we trust least. The more our culture tells us about freedom and democracy, class, wealth, and country, the less objective is our knowledge of these things. That's a real problem.

But it is not fundamental. Even in physical science the appearance of simplicity, the appearance that it is a science of things you can see and

touch, is an illusion. Remember the conundrums of elementary physics: A *unit of force* accelerates a *unit of mass* up to a *unit of velocity* in a *unit of time*. And, of course, what is a unit of mass? A *unit of mass* is the mass that the *unit of force* accelerated. The verbal tangle of the riddle tells us that it's all circular — and it's supposed to be. Nothing in the physical theory exists in isolation. Neither force, nor mass, nor velocity can be measured apart from the equations that bind one to the others. This is what Carlsson called "intrinsic measurement" and what Campbell called "derived measurement": Everything fits together and is defined together, using the physical laws that describe their relation.[2] Modern students can think of objects as having size and mass and temperature, all of which can be weighed and measured independently. That's wrong, but in physical science you can get away with it. And that, I think, is the difference between social science and physical science. In social science you have to think right, from the beginning: The illusion of simplicity is not allowed and you have to figure out the equations while you are still trying to figure out the measure of the things that are subject to the equations. That is the special burden of the social sciences.

That's tricky, but science does it all the time. And that trick liberates social science from the tyranny of the obvious: *Intrinsically*, using measures derived from the problem at hand, no variable can be taken at its obvious value: Neither sex, nor age, nor income, nor education. We are, for example, rarely interested in "Sex" as a categorical statement about a person's physical endowments, not as social scientists. *Intrinsically*, "Sex" can only be measured by the equations that link it to other variables. *Intrinsically*, it can be entirely correct to measure the difference between the sexes in terms of intervals, observing that there is no difference between sexes with respect to behavior X, or that there is a great difference between sexes with respect to behavior Y, or that there is a changing difference between sexes with respect to behavior Z. We are interested in sex differences with respect to income, attitude, education, religion, and politics by which the variable is linked to other variables and thus, intrinsically, even a variable like "Sex" can be measured in

---

2. Gösta Carlsson, *Social Mobility and Class Structure*, Lund, Sweden, 1958, page 141, and Norman Robert Campbell, *What Is Science*, Dover Publications, Inc., 1921. Excerpts from Campbell's work are reprinted in Newman, ed., *op. cit.*, pp. 1797-1829.