

Preface to Faculty

I propose a test and a question: Collect the data that describe the gross national product of the United States, in sequence, for the last one hundred years. Plot the data on graph paper. And then ask someone to interpret it. That's the test. And the question is — How many people will be able to interpret it? I am not sure of the answer to that question, but I know that most of my colleagues — who teach undergraduates in both the natural and social sciences — are none too sanguine about the probable results for their own students.

And I submit that this is a serious problem. If I were to state that the gross domestic product of the United States had increased an average of 3% per year during the last 100 years, most listeners would claim to understand what I had said. If I were to state that the population of the United States had increased by 2% per year, during the last 200 years, most people would claim to have understood the meaning of the statement. But they do not: If an educated person can not compute such an average, 3% per year, or 2% per year, directly from the data then, in truth, that person does not understand the average — not what the average means, not really. And that's not good — not for the educa-

tion of scientists, not as preparation for business, not for policy, nor for education in the liberal arts.

I'm afraid to speculate what this means for the general public. I know it means that when I stand up to argue about tax *rates* in the forum of my town meeting, I'm kidding myself. I know it means that public debate about the rate of increase of U.S. health care costs vis à vis the *rate* of inflation — whatever that debate is about — will not be won, lost or, perhaps, even influenced very much by facts, not in the public domain. And, for education, the unhappy truth is that the problem is not going to be solved by pabulumizing the curriculum. The worlds we try to understand and control through data, the worlds, of science, and policy, and business, are not going to simplify themselves in order to accommodate the deficiencies of the analyst. Rather, we are going to have to bring our students up to the demands of the data. To do otherwise is not only wrong but an act of hubris. If, for example, the relation between a person's years of education and a person's income is not linear, then it does no good to describe it as linear anyway "because" linear analysis is easier. That's not an approximation or simplification. That's wrong. And, very likely, all the basic statistical machinery called upon to support that description and lend to it an air of

verisimilitude — invoking Gaussian probabilities, confidence intervals, significance tests, and so forth — will not set right what was wrong to begin with.

It's easy enough to offer broad generalities to explain-away the difficulties encountered by students — something about the decline of Western Civilization, or the decay of the scientific subculture, or, at the least, the failure of our educational systems. But these generalizations do not hold up. The problem and the solution lie elsewhere: In college we teach the elite, not the average. We speak to students who have made the cut of standardized tests. And we, the faculty, are not asking these students for technical skills they do not have: They all know what a straight line is. They all know how to draw a graph. Almost every undergraduate comes equipped with some knowledge of algebra. They have had an encounter with logarithms and most, or many, of our students have had an introduction to the calculus.

And, in truth, the mathematics and interpretative skills we, the faculty, have come to avoid in introductory data analysis and statistics classes are technically simpler than the mathematics we teach: For example, logarithmic relations and their interpretation, which we usually omit from data analysis, are technically simpler than

the basics of probability, Gaussian distributions, significance tests, Chi-squares and F-ratios which we teach.

So what's wrong? The problem is that students have trouble using the tools they have acquired, carrying their abstract knowledge across the line from abstraction to use: The problem is connecting the math they know to what our social science colleagues call the "substance" behind the methods, to what our physical science colleagues call the "physical intuition" — connecting the math to the data.

I offer these comments to prepare you for what follows: *Rules* is neither more nor less difficult than standard approaches to "methods." It is different and. It is based on a different diagnosis of the problems of teaching and learning data analysis. If I am right, or to the extent that I am right, faculties can not solve our students' problem with data analysis by sending them back to the mathematics department for additional preparation in mathematics. I certainly encourage additional training in mathematics, as much as the students can get, but it's not going to solve the problem of connecting abstractions to reality. That's the problem and the way to solve it is to lead the students back and forth between the two cultures, between math and data, between equations and interpretations, passing back and forth,

repeatedly and redundantly, until the path is well worn. That is something we in the sciences must do for our own students — it is not the business of mathematicians.