



Introduction to Data Analysis: Rules of Evidence

Preface

Every professor wants every student to believe that the subject being taught is exciting, beautiful, and the most worthy calling a human being could possibly follow. And I of course, want students to feel that way about data analysis. But, in a sense, I have to insist — because if you approach data analysis mechanically, as a matter of ritual and routine, then you can not do it well. A computer, or a human acting like one, can not be programmed to use insight, to have hunches, good or bad, and to track the hunches down. And these things are absolutely essential to data analysis, every bit as necessary as solid logic and precision. Data analysis requires logic and clear thinking, but also style, intuition, eccentricity, inspired guesses, and hot pursuit of good ideas — that may or may not turn out to be correct.

Data analysis is, in part, a body of techniques for dealing with numbers: The numbers can be averages — the average income of a population. The numbers can be counts — the number of protozoa in a drop of pond water. The numbers can be temperatures or velocities — there is no end to the list. But

there is a defining characteristic that separates data analysis from mathematics: In data analysis the numbers carry a message from the real world to us. And we look to the data in order to figure out how that world works.

That's what's special about data analysis: Put "two of this" and "two of that" together in a test tube and — reality governs: You'd better check to see just what it is you've got. Granted that "two plus two" will always be "four," but that is a statement in mathematics where "two" and "plus" and "four" are reasonably well-defined abstractions dwelling in the human brain. By contrast, in data analysis the "input" is from the real world and the "output" is tested against the real world. However compelling the assumptions, however logical the conclusion, however precise the mathematics and convincing the theory, the results must be tested against the real world. It is arguable that there are more general kinds of data, data that are not numerical. But, in *Rules*, we, the co-authors, are going to talk about the analysis of data, numerical data, and basic strategies for deciphering the message that they carry.