Description:

Numbers for the Variation; Numbers for the Average

Data analysis is cumulative: *If* the data pass the first test that establishes whether or not it is worthwhile to continue, and *if* the stem and leaf begins to make sense of the data *then* it may be time to introduce numbers that summarize what has been observed in the stem and leaf.

The arithmetic by which we summarize data is very simple: It takes no more than a few minutes to master the arithmetic of an average. If anything, the arithmetic is too simple, causing analysts to compute the numbers, report them, and quickly move forward to something more mathematically difficulty and, presumably, more sophisticated. But my real job in these pages is to teach data analysis, not arithmetic, and to accomplish that I have to rivet your attention on two issues.

The first issue is variation: The whole idea of "average," assumes variation. If the income of the *American family* is $30,000, every family, that means one thing. If the *average* income of American families is $30,000, that means something else: It means some families have less that $30,000, some families have more than $30,000. And, in the middle, very few families will have exactly $30,000.00 — to the penny. Incomes vary. Do "typical" incomes vary between $29,995 and $30,005, or do typical incomes vary between $10,000 and $100,000: As data, as a message, the variation makes a big difference for any analysis of family incomes in the United States.

The second issue is strategy. I could simply list a couple of formulas with instructions for their use. And, in fact, I will do just that after I have established a context. But why use one technique or another? Is there a logic? Is there a consistency that logically binds one technique to another. There is of course. And that, the ;underlying logic of the statistical tools is based on strategy.