

Can You Tell Me the Industry Average?

by Dr. David Zubrow

**Director of Measurement and Analysis
Software Engineering Institute (SEI)**

“Can you tell me the industry average?”

I get a question like this probably once a week. It always sends me into a tirade. People often call because they need a number for a proposal or because they are evaluating a proposal. Or maybe they're setting targets for an improvement program. Whatever their purpose, I usually launch into an unwelcome lecture and it goes something like this:

“First, what industry are you referring to? I assume you would like data relevant to the type of work performed by your organization.” This may sound obvious, but an industry average will often include data from a wide variety of organizations performing a wide variety of services and producing a wide variety of products. Do the types of organizations contributing data to the average matter? I would think so. For instance, organizations making systems that must meet safety standards will have a different productivity rate than those who are not required to meet such standards. Would you want those organizations in your average?

Knowing details about the types of organizations that have contributed data to the average is just one kind of question that should be asked. Other questions include: “How are the data collected? Are they collected with a survey? If so, what projects are included in the data? Who submits the data for the organization and what sources do they draw upon? Are the data verified in some manner? Are the data normalized some how?” If the data are collected some other way, the questions still apply. Nevertheless, the answers to such questions would inform one as to the confidence, appropriateness, and risk associated with specific usages of the “industry average.”

A final issue revolves around “the average.” The average itself may or may not be the most likely or most probable value of a distribution. Hence, it may not be the right number to ask for.

It is important to understand the shape of the data's distribution. For example, the average for a narrow, bell-shaped distribution may be used with the expectation that future results would be similar. On the other hand, another set of data may have the same average but a very broad range. Therefore, using the average could carry greater risk. In a rectangular distribution, to give you another example, all values are equally likely and therefore the average may be as good a number to use as any other. But the magnitude of likely error could be great. In a

bimodal distribution, the average may be the least likely number to occur. Consequently, you may want to consider use of the median or mode.

My point here is simply to stimulate skepticism and inquisitiveness on the part of those seeking numbers to meet their needs. I am interested in creating a community of savvy and informed consumers of data. Unfortunately, there are little high-quality data easily available to answer the kinds of questions that need to be asked. This is understandable given the expense and effort that goes into developing quality data sets. Nevertheless, we must be informed about the quality and appropriateness of the data we are using to build and evaluate our proposals. The questions above should be able to provide a starting point for you in evaluating your own sources of data.