Learning

Tuesday May 14, 2019 at 19:42

What should be learned? How should I evaluate your learning?

Importance of your ability to do arithmetic

On a scale from 0 to 10: 0

I could care less about your ability to do arithmetic. I will never ask you a question on homework or a test where you show that you can be a human calculator. Arithmetic is what computers do, and they do it much, much, much faster than us, and with zero errors. You will likely never, even one time in your career, need to do any arithmetic without at least access to a calculator on your phone.

That being said, I suspect in the past you may have taken even university level classes where you had to be a human calculator. I remember talking to one student a few years ago who told me that for his stat class he had to memorize the computational formula for a standard error for the t-statistic of a mean difference and then do those calculations on a test. Not only memorize a formula, but memorize a version of the formula optimized for calculations instead of showing the concept. The professor of that course and the school that hired him or her should be sued for incompetence.

Formula memorization

On a scale from 0 to 10: 1

Very little is gained from memorizing formulas. What good is memorizing $t = \frac{b_1 - 0}{s_b}$ if you do not know the meaning of t, b_1, s_b, or why the 0 is there?

Memorizing arbitrary formulas with no meaning to you, just a bunch of symbols that you memorize the night before a test and then forget, is a completely worthless exercise regarding your future career. Unfortunately, that strategy has likely worked for many of you in the past, reinforced by the way the material was taught, but that is not what this course is about.

Learning concepts

On a scale from 0 to 10: 8

But there are concepts, and concepts are what we want to learn. Here are two basic *constructions*:

A *hypothesis test* evaluates how many standard errors a sample value is from the corresponding hypothesized value, such as attained by a t-statistic.

A *confidence interval* of a population value is about two standard errors on either side of the corresponding sample statistic.

Two of the most important concepts in statistics in two simple sentences. Further, the sentences are completely general, applicable to a wide range of statistics and specific analyses.

If you learned concepts presented previously in this course, then those definitions of a HT and CI are straightforward definitions to understand. But if you never learned from a previous week (let alone from your stat class) that a standard error is just a reference to the standard deviation of a statistic, and why we care, than definitions such as those for a hypothesis test and confidence interval will befuddle you. And if you learned in a previous week that 95% of the values of a normal distribution are almost two standard deviations on either side of the population value, such as a mean or slope coefficient, now it all makes sense.

I want you to know the concept of, for example, a confidence interval if you are doing data analysis, which is the purpose of this course. So, on the homework and on the tests, show me – with the actual numbers given to you on the output – that you understand where the confidence interval comes from. Get the standard error from the output, multiple by two to get the margin of error, and the add and minus from the sample value. Write out the expression without doing arithmetic because the computer has done the arithmetic for you.

This is the meaning of *apply the definition*: State the definition illustrated with the specific numbers of the analysis. To minimize formula memorization, formulas are not going to be sufficient on tests. You will see this phrase *apply the definition* in many of the questions regarding computer output, so important to understand what it means.

Interpreting the output

On a scale from 0 to 10: 10

In the context of understanding a process and business decision making, the entire point of the analysis is to be able to interpret the results. That means “relaxed English” with zero, absolutely zero, statistical jargon. If you cannot communicate the results there was no point in even doing the analysis.

Instead of writing a formula for confidence interval, present the definition verbally and then fill in with the numbers from this particular analysis. Here is the *meaning* of a confidence interval:

meaning of *CI of a slope coefficient*: Range of plausible values at a specified level of probability that likely contains the true population slope coefficient, Beta or β.

From that meaning, here is the *interpretation*, which expands on the defintion of a slope coefficient as it applies to communicating the results in relaxed English:

interpretation of *95% CI of a slope coefficient*: We are 95% confident that as X increases one unit, on average Y increases [or decreases] anywhere from [lower bound] to [upper bound].

In any one analysis, replace X and Y with the specific variable names, replace unit with the actual unit from the data, and enter the acutal lower and upper bounds. In particular, do not use the term *slope coefficient* in your interpretation. Do *not* assume the people to whom you are presenting results understand that term.

Or, for a t-statistic, state that the t-statistic refers to the number of standard errors a sample statistic is from the hypothesized value, then fill in with the specific numbers. No formula given at all.