Converting Research Questions into Statistical Hypotheses

One of the more difficult tasks that graduate students have in analyzing their data is the conversion of what they know they want to find out into hypotheses that can be tested statistically. This is because the logic of null hypothesis testing is poorly taught and poorly understood. The purpose of this paper is to briefly present that logic, and to describe how it relates to answering your research questions. To do so, we will describe (a) the principle of disconfirmation, (b) the relationship between the null and alternative hypotheses, and (c) the process of null hypothesis testing.

Consider the research question "Is ownership of a personal computer necessary for completing the requirements of a Ph.D.?" As a statement rather than a question, this would read: "Ownership of a personal computer is necessary for completing the requirements of a Ph.D." Now let's assume that you go about trying to confirm this statement by asking individuals who have recently received their Ph.D. if they owned a personal computer. You find person after person who responds that "Yes, I had a personal computer." How many people would you have to receive that response from to confirm, or prove, that the statement was true? The answer is that no matter how many people you find who respond affirmatively to that question, you will never be able to prove that statement. Now let's say that you find one person who says "No, I didn't own a personal computer while working on my Ph.D." Eureka, you have just *dis*proved your statement, with a *single* negative response. The point is that we can *disprove* statements, but we can not *prove* them. This is the principle of disconfirmation, and it forms the basis for scientific inquiry.

We use this principle in data analysis in the following way. First, we make a statement about what we really think is going on, such as “An extra hour of tutoring per week will help students pass their statistics class,” and we call this the alternative hypothesis. Then we derive a statement that is the opposite, such as “An extra hour of tutoring per week does *not* help students pass their statistics class,” and call this the null hypothesis. It is crucial that the two hypotheses account for all possible outcomes. In this case they do: Either the tutoring helps or it doesn’t help, there is no third possible outcome. Maybe the tutoring actually hurts, but that outcome is contained in the null hypothesis of “it doesn’t help.” No matter what result we find, exactly one of these two hypotheses must be correct, because we have designed them to be mutually exclusive and exhaustive.

Now, knowing that we can’t prove a hypothesis but can disprove it, we take the tact of attempting to disprove the null hypothesis. If we are successful then we have, in an admittedly backwards and somewhat convoluted manner, supported our real hypothesis, the alternative hypothesis. While you can’t prove that a statement or hypothesis is true, you can disprove that its opposite is true, thereby obtaining the desired result, provided that there are no possibilities other than your hypothesis and its opposite. It is really a rather ingenious system.

Going back to our example, the hypothesis that tutoring helps students pass their statistical class takes on the form of the alternative hypothesis, while the opposite, the hypothesis that it doesn’t help, takes on the form of the null hypothesis. We then test the null hypothesis statistically: If we reject the null, we have supported what we really think, the alternative hypothesis. If we fail to reject the null hypothesis, however, then we are in a somewhat ambiguous position: True, we did not find support for the alternative hypothesis, but this does not mean that it is necessarily incorrect, only that we did not show that it is correct. This may be due to a small sample size, a small effect size, or a variety of other factors. This is why the rejection of the null hypothesis is referred to as a *strong result*, while the failure to reject it is referred to as a *weak result*. This system is also the reason why we refer to not rejecting the null as ‘failure to reject’ rather than as ‘accepting’ or ‘finding support for’ it. We don’t *accept* it, we just *couldn’t reject it*, a rather weak conclusion. In our example, if we fail to reject the null, then we haven’t shown that tutoring helps, which is slightly different than showing that it doesn’t help. Nevertheless, we are forced to proceed as though it does not.

John C. Caruso

Director of Services, Dissertation Statistical Services

www.diss-stat.com