Post-lecture Questions II.3 – Inferential Statistics (***t***-tests)

Study Questions

What two numerical values are created in the first step of point estimation? [Alternatively: in the first step of point estimation, two values are estimated … what are they?] What are these values (i.e., how are they calculated) when the point being estimated is the population's mean?

How and why is a paired-samples ***t***-test (aka a within-subjects ***t***-test) pretty much the same as point estimation?

Why do you have to use a different procedure for conducting an independent-samples ***t***-test (aka a between-subjects ***t***-test)?

What, in general happens during the second step of either type of ***t***-test?

What arbitrary number, that must be agreed on in advance, is involved in the second step? [If you can give both the label and a description of what it means, that would be great.] What value for this number do we use in psychology?

What is the label we give to concluding that the two population means are different when they really are the same? What is the label we give to concluding that the two population means are the same when they really are different? [Note: there are actually two labels for each; one is technical and arbitrary and the other is much more useful, since it's in words. Knowing both is best.]

What things cause us to make each of the two types of error? [In other words, what causes us to lose statistical conclusion validity?][Note: one type of error has two sources and the other type of error has the same two sources plus one more.]

Example multiple-choice questions (from last Spring):

1. Assuming that all statistical rules are being obeyed, the probability of making a **Type-I error** is _____ in psychology.
   (A) zero
   (B) .05 or 5%
   (C) .95 or 95%
   (D) almost always unknown

2. Violating one or more assumptions of the analysis (i.e., breaking a statistical rule) _____ .
   (A) increases Type-I errors only
   (B) increases Type-II errors only
   (C) increases both types of statistical-conclusion error
   (D) has no effect on the chances of making a statistical-conclusion error

Answers to Study Questions

In the first step of point estimation, you calculate a best guess for what you're interested in, plus an estimate of how wrong you might be. The best guess for the mean of the population is the same as the mean from the sample; the estimate of how wrong you might be – which is called the "standard error of the mean" – is the sample's standard deviation divided by the square-root of the number of subjects in the sample.

Paired-sample ***t***-tests are the same as point estimation because you can (and should) think about each subject as providing only one piece of data: the difference between the two conditions. For example, for each subject in the lighting/memory experiment – assuming that it was run within-subjects – you subtract the number of items recalled in the dim-room condition from the number recalled in the very-bright-room condition. Then, conduct a ***t***-test on these difference scores. If the mean of the difference scores can be shown to be different from zero, then lighting condition had an effect on memory. [Note: what this really means is that you do some extra pre-processing work when you have a within-subjects design: you perform the subtraction separately for each of the subjects, such that the data that come out of preprocessing is just a single set of difference scores.]

You have to use a different procedure for an independent-samples ***t***-test because you can't calculate a difference score for each subject, because each subject was only run in one of the two conditions. Therefore, in the first step of an independent-samples ***t***-test, you first calculate the mean for each of the two conditions separately and then subtract one from the other to get the best guess for the difference between them. As above, you also get an estimate of how wrong you might be concerning this difference.

In general, the second step of a ***t***-test converts what you calculated in the first step to a simple yes-or-no answer to the question: "are these two means significantly different?" This translates to your conclusion as to whether the two means are different in the sampling population.

In order to convert the values from the first step into a yes-or-no answer, we need some kind of cut-off or threshold for chance. The label use for this cut-off is α (i.e., alpha) which is also known as "risk." The value of α is the probability that we will conclude that there is a difference between the two population means when there really isn't a difference. We set this value to .05 (i.e., 5%) in psychology.

Concluding that the two population means are different when they are really the same is called a "Type-I" or "false-alarm" error. Concluding that the population means are the same when they really are different is a "Type-II" or "miss" error.

Type-I errors are caused by violating the assumptions of ***t***-tests and/or bad luck. For example, if your data are nowhere close to being normally distributed, then you are probably violating the assumption that the sampling distribution for the mean is normal. Note, however, that even if you obey all the rules (and your data fit the assumptions), when you come to the conclusion that the population means from the two conditions are different, there's still a 5% chance of being wrong. Type-II errors are also caused by violating the assumptions of ***t***-tests and/or bad luck, but they are also caused by having too much "noise" in your data and/or too small of a sample for the level of noise. Because we never know for sure how much noise there will be in our data, nor how big the difference between conditions really is, we can't actually say, precisely, how often we make these errors. We try to keep the "power" of the experiment at 80% or more (i.e., we try to keep β under .20), but this isn't under our direct control.

The answer to the first multiple-choice question is B. The answer to the second is C.