Post-lecture Questions II.4 – Inferential Statistics (***t***-tests)

Study Questions

What two aspects of a data-set determine how precise an estimate of the population mean can be made?

What are the two types of error that can be made when coming to a yes-or-no conclusion with regard to an experiment? [Try to say what the errors are in words, then give the technical labels.]

What factors cause each of the two types of error?

In what specific way are within-subject designs better than between-subject designs (from a purely statistical point of view)? Loosely speaking, why is this true?

Assume that you have conducted a two-condition experiment and found a significant difference. What do you need to worry about before taking your finding seriously?

Assume that you have conducted a two-condition experiment and found no significant difference. What do you need to worry about before taking your finding seriously?

Example multiple-choice questions (from last Spring):

1. The main advantage of within-subject designs over between-subject designs is that they _____ .
    (A) have more statistical power
    (B) make fewer false-alarm errors
    (C) are easier to analyze
    (D) have no statistical assumptions

2. Having lots of "noise" in the data (e.g., using a very unreliable measure) _____ .
    (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

The standard error (of the mean) is what determines the precision of the estimate of the population mean and standard error is determined by two things: the spread in the sample (namely, the sd of the sample) and the size of the sample (N).

You could conclude that the two population means are different when they really are the same, which is a "false-alarm" or Type-I error, or you could conclude that the two population means are the same when they really are different, which is a "miss" or Type-II error.

Type-I errors are caused by breaking the rules for the statistical procedure that was employed or plain old bad luck, since false-alarms occur 5% of the time, even when you obey all the rules. Type-II errors are caused by breaking stats rules (as above) or are due to low power. Low power is caused by (a) too much "noise" (variability) in the data, (b) a sample that was too small, or (c) the use of an insensitive design (i.e., a between-subject design).

Within-subject designs have more power than between-subject designs, which means that they make fewer Type-II errors. [Note: the two design types do not differ in the rate of Type-I errors.] Within-subject designs have more power because differences between subjects get "subtracted out" of the data when the difference scores are calculated. These individual differences remain in the data and act as a major source of noise when you use a between-subject design.

If you found a significant difference between the conditions, then you need to worry whether this is a false-alarm (Type-I) error, as opposed to there being a real difference in the sampled population. False alarms are [mostly] caused by two things: bad luck and/or violations of the assumptions of the analysis.

If you didn't find a significant difference between the conditions, then you need to worry whether this is a miss (Type-II) error, as opposed to there being no real difference in the sampled population. Misses are [mostly] caused by three things: bad luck and/or violations of the assumptions of the analysis and/or a lack of statistical power. A lack of power comes from any of three things: the data are too variable (i.e., noisy) for the size of the effect that you are trying to detect, you just didn't collect enough data for the size of the effect, and/or you used an insensitive design.

The answer to the first multiple-choice question is A: they have more power. B is definitely wrong, as both designs use the same α (.05) so they both make false alarms 5% of the time (when all assumptions are warranted); C is also not true, especially using modern stats packages; and D is outrageously not true, since all inferential statistical procedures make assumptions.

The answer to the second question is B: it increases Type-II (miss) errors only. Noisy data do not increase the rate of Type-I (false-alarm) errors, since the stats automatically adjust to the noise to maintain the 5% error rate, so A and C are both incorrect; that noise raises Type-II errors means that D can't be correct.