STATISTICAL CONCLUSION VALIDITY

Definition

- The validity with which statements about the association of two variables can be made based on statistical tests.
- Addresses the question: Is there a relationship between the two variables?

Threats to Statistical Conclusion Validity

1. **Low statistical power**: Is a threat when sample sizes are too small or when alpha is set low. This is because low statistical power increases the likelihood of making a Type II error (accepting null when it is false).

2. **Violated assumptions of statistical tests**: Is a threat when the assumptions underlying statistical tests (e.g., normality) are not met. This is because some statistical tests cannot be meaningfully interpreted when assumptions are violated.

3. **Fishing and error rate problem**: Is a threat when researchers make numerous multiple comparisons (when conducting a large number of statistical tests). This is because the likelihood of making a Type I error (rejecting the null when it is true) increases the more comparisons a researcher makes.

4. **Reliability of measures**: Is a threat when measures of reliability for a scale are low. This is because an unreliable scale cannot be relied on for detecting true differences or changes (note that low reliability for research purposes is less than .7 to .8 but for decision making especially in clinical settings, you would want it to be higher).

5. **Reliability of treatment implementation**: Is a threat when treatment or instruction is not the same (and should be the same) for all participants. This is because unreliable treatment implementation will inflate error variance and consequently, decrease the chance that a true difference will be detected.

6. **Random irrelevancies in the experimental setting**: Is a threat when irrelevant non-treatment features of the experimental setting influence scores on the dependent variable (e.g., excessive noise for one group but not another). This is because random irrelevancies in the experimental setting inflate error variance.

7. **Random heterogeneity of respondents**: Is a threat when respondents in any treatment group differ on factors (not the treatment factors) that are also correlated with the major dependent variables. One impact of random heterogeneity of respondents may be that certain kinds of respondents will be more affected by a treatment than others (a matter of external validity). Error variance may also be inflated.
Advice to Enhance Statistical Conclusion Validity

1. Each person might be his/her own control.
2. Samples might be selected to be as homogeneous as possible.
3. Pre-test measures should be collected on the same scales that are used for measuring effect.
4. Matching might take place, before or after randomization, on variables that are correlated with the post-test.
5. Effects of other variables that are correlated with the post-test might be covaried out.
6. Reliability of dependent variable measures might be increased.
INTERNAL VALIDITY

Definition

• The validity with which statements can be made about whether there is a causal relationship from one variable to another in the form in which the variables were manipulated or measured.
• Addresses the question: Given that there is a relationship, are the changes in one variable caused by the other variable, or would the same relationship have been obtained in the absence of any treatment of any kind?

Threats to Internal Validity

1. **History:** Is a threat when an event occurs between pre-testing and post-testing, because the effect might be due to the event that takes place between pre-testing and post-testing, rather than due to the treatment.

2. **Maturation:** Is a threat when an observed effect might be due to the respondents’ growing older, wiser, more experienced, and the like, between pre-testing and post-testing and when this maturation is not the treatment of research interest.

3. **Testing:** Is a threat when participants are tested more than once in the same way, because the effect might be due to the number of times particular responses are measured, rather than due to the treatment.

4. **Instrumentation:** Is a threat when instrumentation changes between pre-testing and post-testing, because the effect might be due to a change in the measuring instrument and not to the treatment’s differential impact at each time interval. This can happen when human observers become more experienced between a pre-test and post-test or when a test shifts in metric at different points.

5. **Statistical regression:** Is a threat when respondents are classified into experimental groups, at say, the pretest, based on their pretest scores or correlates of pretest scores. This is a threat because extremely high or low scores at Time 1 are likely to regress (move) toward the population mean at Time 2.

6. **Selection:** Is a threat when two participant groups are reliably different on a characteristic that is not manipulated. This is because the effect of treatment on any group differences will be confounded with selection differences.

7. **Mortality:** Is a threat when the participants who drop out of a particular treatment are different from those who continue that treatment, because the experimental groups are then composed of different kinds of persons at post-test (a selection artifact).

8. **Interaction with selection:** Is a threat when other threats interact with selection (e.g., selection-maturation). This is because interaction with selection can produce effects that appear as (but aren’t) treatment effects.
9. Ambiguity about the direction of causal influence: Is a threat when it is unclear whether A caused B or whether B caused A. This is a threat in many types of correlational studies, but not in most (but not all) experiments.

10. Diffusion or imitation of treatment: Is a threat when various experimental (and control) groups can communicate with each other. This is because respondents in one treatment group may learn the information intended for others and in the process eliminate group differences.

11. Compensatory equalization of treatments: Is a threat when a control group is compensated (usually by a third party) for not receiving goods or services that are believed to be desirable and that the experimental group received, to equalize treatment across groups.

12. Compensatory rivalry by respondents receiving less desirable treatments: Is a threat when the assignment of persons to groups is made public and this public knowledge generates competition between groups. Control group may outperform treatment group as a result!

13. Resentful demoralization of respondents receiving less desirable treatments: Is a threat when not all treatments are equally desirable and participants in the less desirable group become upset; participants who are resentful of being in the less desirable group may perform (poorly) or behave differently than they would otherwise.

**Advice to Enhance Internal Validity**

1. Random assignment to groups.
CONSTRUCT VALIDITY OF CAUSES OR EFFECTS

Definition

- Refers to the approximate validity with which we can make generalizations about higher-order constructs from research operations.
- Addresses the question: Given that the relationship is plausibly causal and is reasonably known to be from one variable to another, what are the particular cause and effect constructs involved in the relationship?

Threats to Construct Validity

1. Inadequate preoperational explication of constructs: Is a threat when preoperational explication of constructs is not based on the results of a conceptual analysis of the features of a construct.

2. Hypothesis-guessing within experimental conditions: Is a threat when participants guess the hypothesis (incorrectly or correctly) and then act on the basis of what they perceive the experimental hypothesis to be. (*Hawthorne effects*)

3. Experimenter expectancies: Is a threat when what the experimenter expects to happen biases the data. When this occurs, it will not be clear whether the causal treatment is the treatment-as-labeled or the expectations of the persons who deliver the treatments to respondents. (*Rosenthal Effects*)

4. Confounding constructs and levels of constructs: Is a threat when researchers do not specify sufficient levels of measurement for a construct (e.g., drug effect). When insufficient levels are specified (e.g., drug dosage), an effect that occurs (e.g., intoxication), albeit only when the sufficient construct level is specified (e.g., a dosage of at least three beers per person of a particular weight), will not occur.

5. Interaction of different treatments: Is a threat when participants receive more than one treatment, because it will be unclear which treatment is responsible for the effect.

6. Interaction of testing and treatment: Is a threat when theoretically irrelevant testing conditions influence treatment effectiveness. When testing conditions and treatment interact, generalizability is questionable. This can happen when participants are pre-tested.

Advice to Enhance Construct Validity

1. Define clearly the constructs of interest.

2. Find or develop valid measures or manipulations that operationalize your defined constructs.
EXTERNAL VALIDITY

Definition

- Refers to the approximate validity with which conclusions are drawn about the generalizability of a causal relationship to and across populations of persons, settings, and times.
- Addresses the question: Given that there is probably a causal relationship from construct A to construct B, how generalizable is this relationship across persons, settings, and times?

Threats to External Validity

1. Interaction of selection and treatment: Is a threat when the treatment only works with the people used in the study. To evaluate whether this is the case, you should ask yourself, can the results be generalized to other people beyond social class, race, age, geography, sex, or personality groups?

2. Interaction of setting and treatment: Is a threat when the treatment only works in the setting used in the study. To evaluate whether this is the case, you should ask yourself, can the results be generalized to other settings such as military camps, university campuses, offices, factories, malls etc.?

3. Interaction of history and treatment: Is a threat when the treatment only works for a group of people with particular experiences. To evaluate whether this is the case, you should ask yourself, can the results be generalized across time? e.g., does it matter if the day that participants were tested was the same day that Trudeau died?

Advice to Enhance External Validity

1. Random sampling of participants.