Validity & Research Designs

- Research Hypotheses & Validity Review
- Research Design Review
- Internal Validity Review

Types of Research Hypotheses

Attributive Hypothesis -- a construct (phenomena, behavior, etc.) exists

- an operational definition of the construct
- a system to measure the construct
- demonstration that the construct can be differentiated from other (related) constructs

Associative Hypothesis -- two constructs are related (i.e., knowing the value of one provides information about the value of the other)

- demonstration of a statistical relationship between the variables used to measure the constructs
- specific statistical analysis is not important, as long as it is appropriate to the data and the expression of the research hypothesis

Causal Hypothesis -- the value of one construct influences (causes, produces, etc.) the value of the other construct

- temporal precedence -- operation of IV comes before measurement of DV
- no alternative hypotheses (no design flaws, confounds, alternative explanations of the results, etc.)
- statistical relationship between IV and DV

The types of RH: are “hierarchically arranged” !!

- Posing a causal hypothesis assumes the associative hypothesis that the IV and DV are related -- “if two things aren’t related then one can cause the other”
- Posing an associative hypothesis <u>assumes</u> support for the attributive hypothesis of each construct/variable-- “unmeasureable things can’t be statistically analyzed”

Types of Validity

Measurement Validity

- do our variables/data accurately represent the characteristics & behaviors we intend to study ?

External Validity

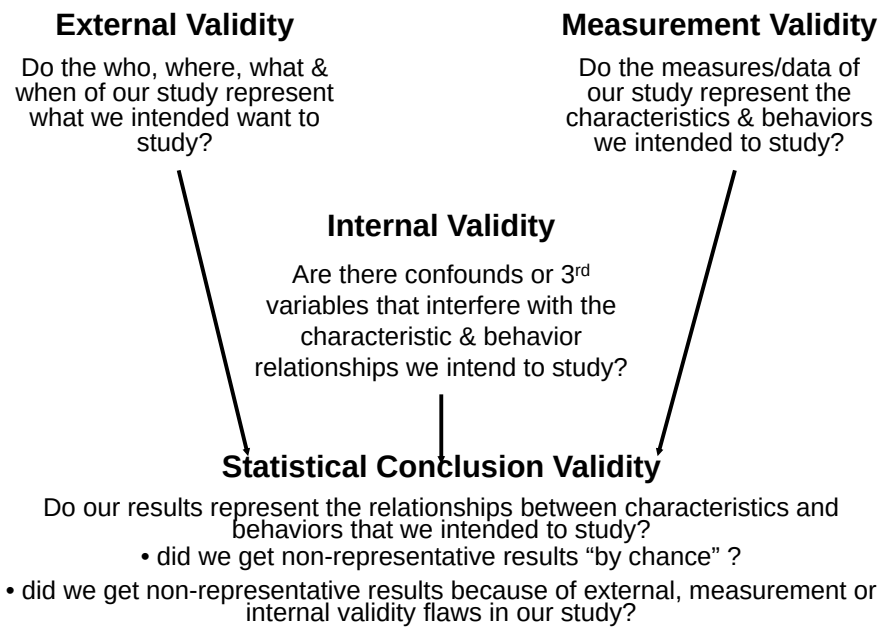
- to what extent can our results can be accurately generalized to other participants, situations, activities, and times ?

Internal Validity

- is it correct to give a causal interpretation to the relationship we found between the variables ?

Statistical Conclusion Validity

- have we reached the correct conclusion about the relationships among the variables we are studying ?



Components of External Validity

Whether we are testing attributive, associative, or causal research hypotheses, we should be concerned about the generalizability of the research results

Population

- Will the results generalize to other persons or animals ?
  - Will a study of college students generalize to your target population of "consumers" ?
  - Will a study of chronically depressed patients transfer to a those who are acutely depressed ?
  - Will a study of captive bred turtles generalize to wild-caught turtles ?

Setting

- Will the findings apply to other settings ?
  - Will a laboratory study generalize to what happens in the classroom ?
  - Will a study in a psychiatric hospital generalize to an out-patient clinic?
  - Will a laboratory study generalize to retail stores?

Components of External Validity, cont.

Task/Stimuli

- Will the results generalize to other tasks or stimuli ?
- Usually the participant is “doing something” that directly or indirectly generates the behavior that is being measured
  - Will a “lever pressing” task tell us anything about “compliment seeking” ?
  - What do I learn about “consumer decision making” from a study that asks participants to select the best “widgit” ?
  - Will research using visual illusions inform us about the perception of everyday objects ?

Societal/Temporal changes

- Will the findings continue to apply
  - Will a study conducted in 1965 generalize to today ?
  - Will a study conducted today still be useful 10 years from now ? … 5 years from now ?

Components of Internal Validity

Initial Equivalence

- Prior to manipulation of the causal variable, participants in the different conditions are the same (on the average) on all measured/subject variables

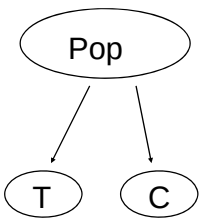
Ongoing Equivalence

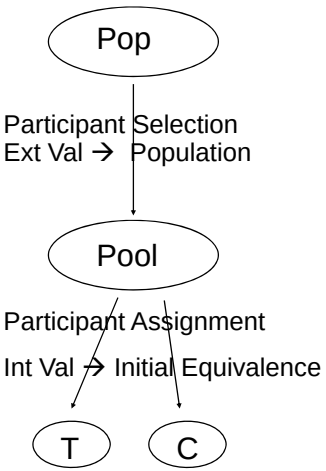
- during manipulation of the causal variable, completion of the task, and measurement of the effect variable, participants in the different conditions are the same (on the average) on all manipulated/procedural variables except the causal variable.

Separating “Selection” & “Assignment”

A common representation of the participant acquisition process is shown below.

Folks are randomly chosen from the pop and placed into one of 2 groups.





What usually happens is shown above: First participants are selected into a “pool” and then are assigned into groups. Different aspects of validity are influenced by each step!!!

Research Designs

True Experiments

If "well-done," can be used to test causal RH: -- alternative hyp. are ruled out because there are no confounds !!!

True Experiment

- ⌘ random assignment of individual participants by researcher before IV manip (helps eliminate confounds)
- ⌘ treatment/manipulation performed by researcher (helps eliminate confounds)
- ⌘ good control of procedural variables during task completion & DV measurement (helps eliminate confounds)

Non-Experiments

No version can be used to test causal RH: -- can't rule out alternative hyp. Because there are confounds !!

What designs "go with" what types of RH: ????

Quasi-Experiment

- ⌘ no random assignment of individuals (no confound control)
- ⌘ treatment/manipulation performed by researcher
- ⌘ poor or no control of procedural variables during task, etc. (no confound control)

Natural Groups Design also called Concomitant Measures or Correlational Design

- ⌘ no random assignment of ind. (no confound control)
- ⌘ no treatment manipulation performed by researcher (all variables are measured) -- a comparison among participants already in groups (no confound control)
- ⌘ no control of procedural variables during task, etc. (no confound control)

Basic Statistical Designs - BG vs. WG

Between Subjects (Between Groups)

- each subject completes 1 of the IV conditions
- different groups each complete 1 of the IV conditions

Within-subjects (Within-groups, Repeated Measures)

- each subject completes all of the IV conditions
- one group of subjects completes all of the IV conditions

"Design Language"

- For both Between & Within designs, we refer to the IV and the DV
  - Typically the IV ("causal variable") is qualitative
  - Typically the DV ("effect variable") is quantitative

"SPSS Language"

Between Groups Designs

- the IV is the "grouping variable" -- which IV condition each subject was in
- the DV is the "response variable" and tells each participant's score on the DV

Within-groups Designs

- there is no IV -- each variable is referred to as a DV
  - there is one DV score for each IV condition
  - each DV score tells the participant's score in that IV condition

ANOVA

Between Groups (Independent Samples, etc.)

- H0: Populations represented by the IV conditions have the same mean DV.
- degrees of freedom (df) numerator = 1, denominator = N - 2
- Range of values 0 to ∞
- Reject Ho: If $F_{obtained} > F_{critical}$ or If $p < .05$

Within-groups (Dependent Samples, etc.)

- H0: Populations represented by the IV conditions have the same mean DV.
- degrees of freedom (df) numerator = 1, denominator = N - 1
- Range of values 0 to ∞
- Reject Ho: If $F_{obtained} > F_{critical}$ or If $p < .05$

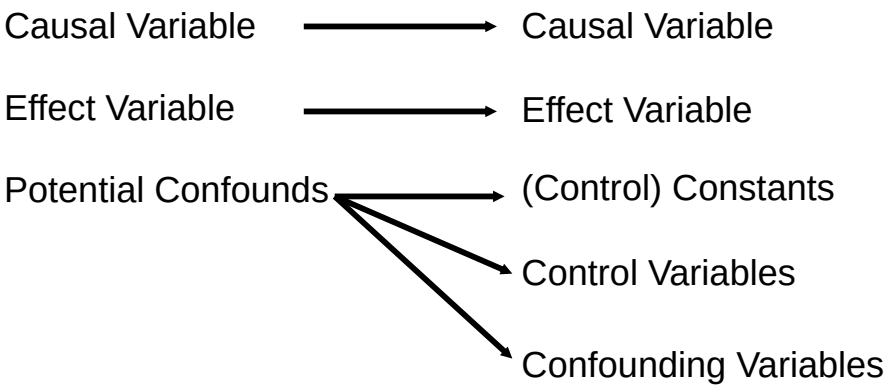
Research Designs & Internal Validity

Putting this all together -- here's a summary of the four types of designs we'll be working with ...

	True Experiment • w/ "proper" RA/CB - init eqiv • manip of IV by researcher	Non-experiment • no or poor RA/CB • may have IV manip
Between Groups (dif parts. in each IV condition)	Results <u>might</u> be causally interpreted -- if good ongoing equivalence	Results <u>can not</u> be causally interpreted
Within-Groups (each part. in all IV conditions)	Results <u>might</u> be causally interpreted -- if good ongoing equivalence	Results <u>can not</u> be causally interpreted

So, to summarize ...

Before the study begins... After the study is over ...



We must take control of the potential confounds, so that they become controls and not confounds, if we are going to be able to causally interpret our research results.

So, we have to be able to discriminate between these three things:
Constants vs. Control variables vs Confounding variables

So, we can tell these apart based on who is and isn't "different" !!!

Kind of thing...	Differences among individual people?	Differences among groups (on average)?
constant	no	no
Control variable	Yes (makes it a variable)	No (makes it a control)
Confounding variable	Yes (making it variable)	Yes (Making it a confound)

Attrition – also known as drop-out, data loss, response refusal, & experimental mortality – we're gonna have 2 kinds

Attrition endangers the population portion of externa validity

- After carefully obtaining a representative sample, some people drop out
- that attrition could render the sample non-representative
- E.g., the study is "harder" so participant drop out – producing a sample that doesn't represent the motivation of the pop

Differential Attrition endangers initial equivalence part of internal validity (of subject variables)

• random assignment is intended to produce initial equivalence of subject variables – so that the groups (IV conditions) have equivalent means on all subject variables

• e.g., If one condition is "harder" and so more participants drop out of that condition, producing a "motivation" difference between the two conditions.

So, "attrition" works much like "self assignment" to trash initial equivalence

Both involve a non-random determination of who provides data for what condition of the study!

Imagine a study that involves a "standard treatment" and an "experimental treatment"…

• random assignment would be used to ensure that the participants in the two groups are equivalent

• self-assignment is likely to produce non-equivalence (different "kinds" of folks likely to elect the different treatments)

• attrition (i.e., rejecting the randomly assigned condition) is similarly likely to produce non-equivalence (different "kinds" of folks likely to remain in the different treatments)

"Counterbalancing failure" causes Initial Equivalence problems!

Huh?? "counterbalancing failure" happens "during the procedure" – so why isn't it ongoing equivalence!

Initial equivalence – before manipulation of the IV, participants in the conditions are equivalent (on average) on all subject variables.

Ongoing equivalence – during manipulation of the IV, completion of the task, and measurement of the DV, participants in the conditions are equivalent (on average) on all procedural variables (except for the IV).

Two things to notice:

- Initial equivalence is about subject variables
- Ongoing is not "during the study" but during manip of IV, task, measurement of DV – for a particular condition

Keep going…

In a WG design, we are counting on counterbalancing to ensure that the set of participants are "equivalent on average on all subject variables" just before they begin the manipulation of the Control condition, as they are just before they begin the manipulation of the Treatment condition.

If the participants are different on any subject variables before they begin manip of the Control condition than they are before manipulation of the Treatment condition – then those difference on those subject variables is an Initial Equivalence problem!

So, if we don't counter balance (or if counterbalancing fails) then the differences produced are subject variable differences – a problem of Initial Equivalence

Let's say we are studying weightlifting…

We want to see if, for some particular movement, there is a difference in the number of reps a person can do with a "wide grip" vs. a "narrow grip". We decide to use a WG design.

We have everybody do the wide grip trial first and then immediately do the narrow grip trial. See a problem?

What is the problem? The variable of concern is "fatigue" – a subject variable! Participants are more fatigued just before the second narrow grip condition than just before the first wide grip condition!

Doing the study with this improper counterbalancing leads to a "fatigue diference" between the conditions – a subject variable problem → an initial equivalence problem! ;)

There is always "just one more thing" ...

Sometimes there is no counterbalancing in a Within-groups design, but there can still be causal interpretation…

- A good example is when the IV is "amount of practice" with "10 practice" and a "50 practice" conditions.
  - There is no way a person can be in the 50 practice condition, and then be in the 10 practice condition
- Under these conditions (called a "seriated IV"), what matters is whether or not we can maintain "ongoing equivalence" so that the only reason for a change in performance would be the increased practice
  - The length of time involved is usually a very important consideration
  - Whether the study is conducted in the laboratory or the field is also important

Which result would you be more comfortable giving a causal interpretation?

- When we gave folks an initial test, 10 practice and then the test again, we found that at their performance went up!
- When we gave folks an initial assessment, 6 months of once-a-week therapy and then the assessment again, their depression went down!