

External Validity

- Research Hypotheses, Findings & Validity
- Types of Research Validity
 - Measurement
 - Internal
 - External
 - Statistical conclusion
- Components of External Validity
 - Population
 - Setting
 - Task/Stimulus
- Participant Selection -- Population Validity

Bivariate RH:s, Research Designs and Validity...

- A RH: is a guess about the relationships between behaviors
- In order to test our RH: we have to decide on a research design, sample participants, collect data, statistically analyze those data and make a final conclusion about whether or not our results support our RH:
- When we are all done, we want our conclusion to be “valid”

Validity ... has lots of types, definitions & procedures but basically it means ... **Accuracy** or **Correctness**

Important to remember !!! No one study, no matter how well-done can ever be conclusive !! You must further apply the research loop -- replication and convergence are necessary before you can be sure about the final answer to your RH:

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/behaviors ?

Statistical Conclusion Validity

- have we reached the correct conclusion about whether or not there is a relationship between the variables/behaviors we are studying ?

How types of validity interrelate -- consider the “flow” of a study

the **research “design”** -- all the choices of how we will run the study

Internal validity

- control
- causal interpretability

External validity

- generalizability
- applicability

Measurement Validity

the **data** -- if we can't get an accurate measure of a behavior we can't study that behavior

Statistical Conclusion Validity

the data analysis -- we must decide whether or not the behaviors we are studying are related (and if so, how)

External Validity

Do the who, where, what & when of our study represent what we intended want to study?

Measurement Validity

Do the measures/data of our study represent the characteristics & behaviors we intended to study?

Internal Validity

Are there confounds or 3rd variables that interfere with the characteristic & behavior relationships we intend to study?

Statistical Conclusion Validity

Do our results represent the relationships between characteristics and behaviors that we intended to study?

- did we get non-representative results “by chance” ?
- did we get non-representative results because of external, measurement or internal validity flaws in our study?

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 ?

Some practice -- pick the parts of the design relating to each ...

Nice study you've found! It describes how 1960's college students decided whether or not to join a protest march against the college administration building during the Vietnam war ! That's interesting, but what does it tell me about which members of our Union will join the picket line outside the plant if we call a strike ?

| | |
|---------------------|---|
| Population validity | students vs. workers |
| Setting validity | college campus vs. industrial plant |
| Task/Stimulus | joining a protest march vs. picket line |
| Temporal/Social | 1960's vs. now |

Some more practice ...

I found an article that supports the use of physical punishment for children who don't follow instructions. Juvenile rats (21 days old) were placed on a wooden block on a shock grid. The animal received a shock whenever it stepped off the block. Most rats learned to stay on the block after only 2-3 shocks. We should apply this in school -- children who don't follow instructions should be paddled.

| | |
|---------------------|---|
| Population validity | juvenile rats vs. children |
| Setting validity | cage vs. schools |
| Task/Stimulus | passive avoidance vs. following instruction shocks vs. paddled |
| Temporal/Social | ????? |



While we have separate definitions for the components of external validity...Population, Setting, Task/Stim & Soc/Temp
...they sometimes get “intertwined” when applied to real research.

Population & Setting -- sometimes “where you’re studying”
changes “who you’re studying”

- a study of hospitalized with schiz. vs. out-patients with schiz
-- different setting -- but maybe also “different” schizophrenia

Setting & Task/Stim -- “where you are” may influence “what they
are doing” or the “stimuli used”

- a study of argument role playing in a lab vs. start of bar fights
-- different setting & maybe a different kind of argument

Population & Task/Stim -- sometimes you have to “adjust” tasks &
stimuli for who you are studying

- a study of elementary vs. high school math learning
-- different population & maybe very different kind of “math”

Here are two other related types of validity -- both of which can be
understood as specific combinations of certain elements of external
validity...

Cultural Validity -- different behaviors or relationships between
behaviors across cultures

- a culture is jointly defined by its members and location
- this can be expressed as a combination of population and
setting components of external validity

Ecological Validity – sometimes used as a synonym for external
validity

- the “ecology” of a study includes the elements that the
participant interacts with and within
- this can be expressed as a combination of setting and
task/stimulus components of external validity

Just a reminder – there is no “correct way” to run a study – no “critical experiment”!!!
We must use convergent operations to provide convincing evidence of external validity!

So, external validity is about the “generalizability” or
“applicability” of the results of a study.

- It’s important to distinguish generalizability from applicability!

Generalizability of a finding (broadly speaking) is whether or not
the results will hold for all (or at least most) combinations of the
elements of external validity.

Applicability of a finding is whether or not the results will hold for
a particular combination of the elements of external validity for
which we’d like to use the results

So, generalizability is much more demanding than applicability.
Much more convergent research is required to support claims of
generalizability than claims of applicability.

Also, generalizability is more difficult to evaluate because it
requires a deeper knowledge of the extent to which population,
setting and task/stimulus differences will influence research
findings – the more psychology you know the better...


Generalizability is a property of the study -- but
“applicability is in the eye of the applier”

Approaches to “defending” limited external validity of a study...

De-emphasize external validity (emphasize Internal Validity)

- if the main focus of the study is causal interpretability (internal validity), you might make a large number of choices each of which hinders the generalizability of the results
- common among theoretical researchers -- but doesn't help the applications folks (& why we have “applied research”)

Eschew external validity (emphasize focused applicability)

- basically the argument is that this study used exactly the pop, setting, task, stimulus, etc. that was of interest to the researcher
- common among applied researchers
 - “my research exactly matches my application; what's to generalize?”
 - “my research exactly matches my application; generalization to your application is your problem!” 

Participant Selection / Sampling

- “Who will be in the study?”
- goal is to have a sample that represents the target population
- related type of validity is External Validity -- Population
- Note -- participant selection (sampling) has nothing to do with the causal interpretability (internal validity) of the study results -- only the “Population” component of External Validity !!!!!

Stages of Selection/Sampling

Target Population – defining people/animals we want to study

Sampling Frame – “best list” we can get of population members

Selected Sample – sampling frame members who are selected to participate in the research

Data Sample – participants from whom useful data are collected

Identify each -- telling the number, if possible...

For our study of UNL students we collected complete data from 72 of the 100 students that were selected from a data file of all UNL undergraduates

| | | | |
|----------------|------------------|-----------------|--------------|
| population | UNL students | selected sample | 100 students |
| sampling frame | registrar's list | data sample | 72 students |

Comments on sampling ?? Poor purposive sampling frame used ...
“UNL students” vs. “UNL undergrads”

For our study of California voters, we obtained the names of all registered voters in that state, selected 2000 and collected data from 1214.

| | | | |
|----------------|--------------------|-----------------|-------------|
| population | Calif. voters | selected sample | 2000 voters |
| sampling frame | list of reg.voters | data sample | 1214 voters |

Selection/Sampling Procedures

Psychologists have devised many different ways of “acquiring” participants, but all involve three choices...

- Population Sampling Frame vs. Purposive Sampling Frame
- Researcher selected vs Self-selected
- Simple Sampling vs. Stratified Sampling

... any form of participant sampling/selection can be identified as one of the (eight) combinations of these three choices

In an important sense -- all participants are “volunteers”

- participants must be invited with full knowledge of any risks incurred through their participation
- they might refuse to participate when invited
- they might start to participate but later withdraw -- called attrition, drop-out or “experimental mortality”



“Kinds” of Selection/Sampling

Population Sampling Frame vs. Purposive Sampling Frame

- a “sampling frame” is the list of members of the target population the researcher starts with
- sometimes it isn’t a paper list, but a way of contacting everybody

A “population” sampling frame includes the **entire** population

- consider how unlikely this is ...

A “purposive” sampling frame includes a **subset** of the entire population that is deemed “representative” of the entire population

- using Intro Psyc students to represent “college students” because many different majors & ages take it
- using Lincoln citizens to represent “Americans”
- 10-15 “market test cities”
- ⌚ nearly all sampling is purposive -- getting full population list is difficult/impossible, expensive, and not necessarily better than a properly chosen purposive list

Some practice - which are “complete pop” and which “purposive”

Start by identifying the sampling frame and the population

- | | |
|---|---------------------|
| • students drawn from this class to represent university students | Purposive |
| • all the students from this class to represent this class | Complete population |
| • students drawn from this class to represent all psychology students | Purposive |
| • 200 students from Psyc181 sampled to represent Psyc181 | Purposive |
| • 350 students from Psyc181 sampled to represent introductory psyc students | Purposive |
| • all the students from psyc181 sampled to represent UNL students | Purposive |

Researcher selected vs. Self-selected

Researcher selected -- potential participants from the sampling frame are selected by the researcher (almost always randomly), individually contacted and requested to participate in the research.

- the selection might be from an actual list -- e.g., registered voters
- or done in “real time” -- e.g., randomly determining whether or not to approach each customer emerging from a store
- sometimes called “probabilistic” sampling

How is this done ??? Two common ways ...

- Sampling frame (list) is cut into strips with each name, put into a box and the desired number of folks drawn
- Each member of sampling frame given a number and numbers are drawn at random (computer, random # table, etc.)

Remember ...

- the **purpose is a representative sample** -- using a random sample is just a technique to achieve representation
- random selection doesn't guarantee the sample will be a good representation of population (though we act like it does)
- random selection tends to give better representation the larger the sample

Researcher selected vs. Self-selected, cont.

Self-selected -- **all** potential participants from the sampling frame are informed about the “opportunity” to participate in the research and invited to contact the researcher if they wish to volunteer.

- Assumes that the volunteers will be a “representative sample” of the target population
- This representativeness can be compromised if ...
 - the entire target population is not notified
 - if there is “uneven” motivation to volunteer across the population (e.g., a small payment for participation would lead to differential representation of those who do and don't find that amount “motivating”)

Some practice... identify “Researcher-selected” vs “Self-selected”

- | | |
|--|---------------------|
| • 40 folks are selected from the Lancaster County voter registration rolls and each contacted to participate | Researcher-selected |
| • Research announcements & invitations are mailed to all 12,234 on the Tali County voter registration rolls | Self-selected |
| • Psyc 181 research participation website was used to recruit 100 participants | Self-selected |
| • Harris labs selected 30 folks who had previously been research participants and who had indicated their interest in further participation to be part of their latest study | Researcher-selected |
| • Using the Psyc 181 grade roster, 200 research participants were selected. | Researcher-selected |
| • Advertisement for Harris Labs research that requires non-smokers aged 21-39 printed in local newspaper | Self-selected |

“Kinds” of Selection/Sampling, cont.

Simple Sampling vs. Stratified Sampling

In “**simple**” sampling **every member** of the sampling frame has an **equal probability** of being in the study

- every name on list has the same probability of being chosen
- every volunteer participant completes the study

“**Stratified**” sampling is a bit more involved ...

- first we have to **divide the sampling frame into “strata”** using one or more variables (e.g., age, gender, job)
- members **within each strata** have an **equal probability** of being in the study
- usually done to ensure representation of smaller segments or strata of the population
 - select 50 each of “Psyc majors” and “non-majors” from 181 rosters
 - have separate sign-up sheets for “majors” and “non-majors”

Some Practice -- is each an example of “simple” or “stratified” sampling ???

- We chose 40 African-Americans, 40 Asian-Americans, 40 European-Americans, and 40 Hispanic-Americans from the rolls of the Multicultural Club stratified
- We chose 100 folks from the Registrar’s student list simple
- Our participants were the first 40 folks who responded to the research participation advertisement simple
- After we’d had 50 male and 35 female volunteers, we changed the sign-up sheet to read “females only” stratified
- (Careful!) Our intention was that the 200 students selected from the Psyc 181 course roster would be 70% from the College of Arts & Sciences and 30% from other colleges. simple
- We sorted the Psyc 181 course roster into those from the College of Arts & Sciences vs. other colleges; then we chose 70 of the former and 30 of the latter stratified

So, there are **8** combinations of ways we obtain our participants...

| | Population sampling frame | | Purposive sampling frame | |
|---------------------|---------------------------|---------------------|--------------------------|---------------------|
| | Simple sampling | Stratified sampling | Simple sampling | Stratified sampling |
| Researcher-selected | * | * | ^ | ^ |
| Self-selected | | | + | + |

* what “random sampling” means in textbooks

^ how “random sampling” is usually done (e.g., Gallup polls)

+ how “participant selection” is usually done in empirical research

Time for practice... identify each as 1) complete or purposive sampling frame ...
2) researcher- or self-selected ...
3) simple or stratified ... sampling

We chose 40 African-Americans, 40 Asian-Americans, 40 European-Americans, and 40 Hispanic-Americans from the rolls of the Multicultural Club to gather information about opinions of members of the Multicultural Club.

complete
rshr-selected
stratified

We chose 160 members from the rolls of the Multicultural Club to gather information about opinions of students at UNL.

purposive
rshr-selected
simple

We posted two notices on the Multicultural Club bulletin board about a "forum" we were hosting to gather information about the opinions of college students, one for men and one for women.

purposive
self-selected
stratified

Putting together the "Stages" and "Procedures" of Sampling to provide a complete description of "from whom the data are obtained"

Target Population – defining people/animals we want to study

Sampling Frame – "best list/access" of population we can get
• Population Sampling Frame vs. Purposive Sampling Frame

Selected Sample – members of the sampling frame who are selected or intend to participate in the research
• Researcher selected vs. Group invitation/Self-selected
• Simple Sampling vs. Stratified Sampling

Data Sample – participants from whom useful data are collected
• Attrition