

Reliability and Validity

Understanding the Meaning of Study Results

Evaluation of Studies

- Reliability refers to error of measurement. The more error, the poorer the measures.
- Validity refers the quality of inference based on the study.
- Internal validity concerns whether the intended IV cause a change in the DV, or something else is responsible.
- External validity is about generalizing the results to other situations.

Reliability (1)

- Aka consistency, repeatability. 3 Major forms:
- 1. Test-retest. Same measure, same people, later time. Compute the correlation.
- 2. Alternate forms. Two measures, same time, same people. Compute the correlation.

Reliability (2)

- 3. Internal Consistency. 1 measure, 1 time, 1 set of people. Cheapest, most efficient method. Considers items as mini-tests. Only works with multi-item measures.
- Ex1: Split half – cut test in two. Compute correlation. Correct to whole with Spearman Brown Prophecy Formula. Many halves.
- Ex2: Alpha – only one estimate. Theoretical estimate of ‘randomly parallel’ test. Expected correlation of another test made up of items just like these.

Reliability (3)

- Reliability is important because it tells about the amount of error in a measure [from 0 to 1].
- If measures are very unreliable (composed solely of error), then
 - The correlation of the measures with other things will be zero
 - Neither DV nor IV can show an effect in a significance test except by chance.
- Always report reliability estimates as part of your study. [Project 2.]

Internal Validity

- Did the *intended* IV change the DV or is there some other explanation?
- There are many threats to internal validity; they’ve been cataloged.
- **He Made The IRS MIDday.**
- **He – History** – meaningful events change people b/t repeated measures. Job attitudes, strike, job attitudes. Magic Johnson study. Team sports.

Internal Validity (2)

- **Made** – *Maturation* – changes in people during a study. People get bored, fatigued, etc. Ex: physical ability measures, long survey (proj 2?)
- **The** – *Testing*. Practice effect. Also reactance. Ex: motor pursuit, SAT; attitude.
- **Irs** – *Instrumentation*. Changes in measuring instrument, wear and tear. Ex: work sample test; computer breakdown. Also judges.

Internal Validity (3)

- **iRs** – *Regression* – people selected on the basis of test scores tend to have retest scores closer to the mean. Ex – remedial education; second year slump.
- **irS** – *Selection of group members*. Group composition unequal at start of study. Ex – volunteers for effects of beer on driving. Effects of exercise with more athletes in one group.

Internal Validity (4)

Midday – *Mortality*. People quit the study. Ex – smoking cessation, exercise.

mIdday – *Interaction with selection* – one group responds to something other than the IV. Ex – male & female attitudes on AIDS before and after Magic Johnson announcement (selection & history).

miDday – *Diffusion* – people tell each other the treatment. Ex – corporate training, visualization in sport; website for class.

Review

| | | |
|--|---|--|
| History- some event with repeated measures | Maturation – during study, get tired, hungry, bored | Testing- practice effect; reactance |
| Instrumentation Equipment probs, judges; double-blind | Regression- move to mean | Selection- unequal groups before study |
| Mortality- death & loss | *Interaction with selection- selection-history | Diffusion- imitation or spread |

He Made The IRS MIDday.
