Week 3

Christopher David Desjardins

January 30, 2015

Validity

Validity

- What is validity?
  - An indicator of how well the test measures the latent construct(s) it purports to.
  - A determination of the appropriateness of the test scores for specific uses/users
  - Validity of the test for a given purpose, at a given time, for a given population
  - You are a lawyer presenting evidence to a judge to make the case for the validity of your instrument

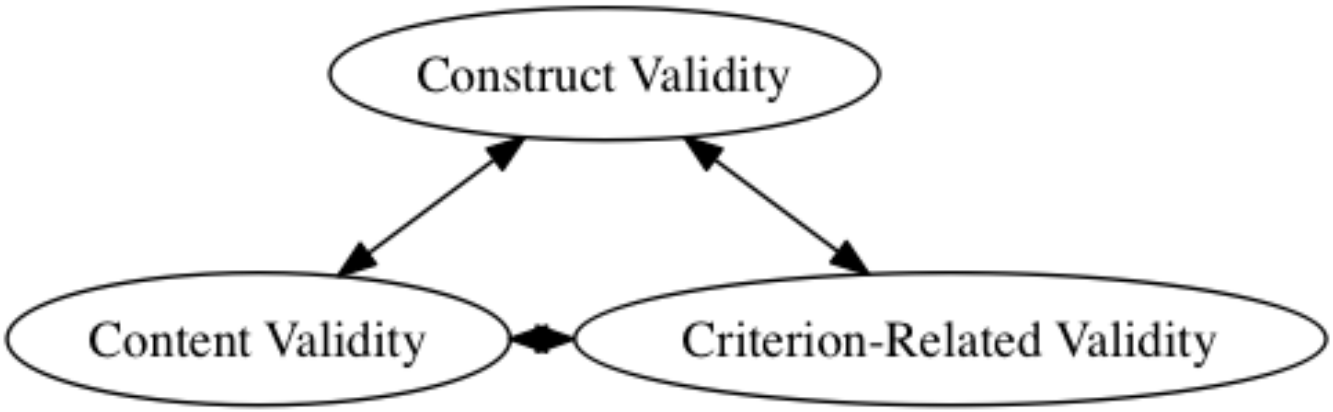
SAT

- The SAT is a standardized test measuring mathematics, reading, and writing
- Typically administered to 15, 16, and 17 year olds (sophomores, juniors, and seniors) in the USA
- Purpose to measure "college success"
- Schools within a city, within a state, and across states in the USA are quite diverse
- Would this test be valid for Iceland?
- Would this be appropriate for HÍ , HR, or UNAK?

Making the SAT valid for Iceland

- Could administer the test as it is or alter the test and conduct a local validation study
- Should translate it to Icelandic
- Update it to reflect Icelandic curriculum
- Age appropriate
- Is it for university-studies or menntaskóli?
- Anything else?

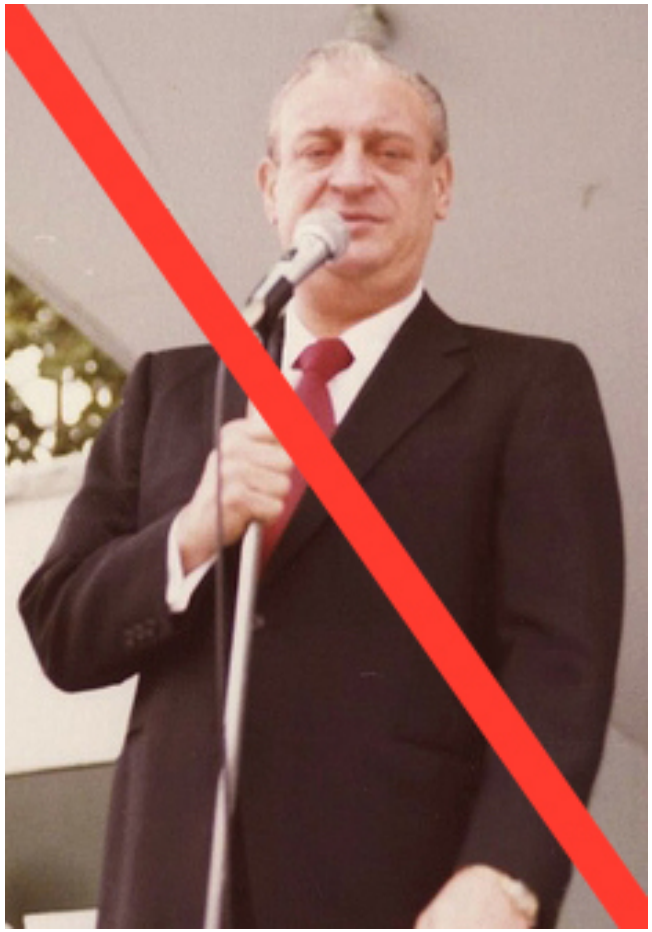
Types of Validity



Overview of Validity

- Content - Evaluation of subjects, topic, or content covered by the items in the test
- Criterion-Related - Evaluation relationship of scores obtained on the test to scores on other instruments measuring the same construct
- Construct - Evaluation relationship of scores obtained on the test to scores on other instruments measuring the same construct AND understanding how it fits within the theoretical framework of the latent construct

Face Validity is NOT Validity



source

Content Validity

- How adequately the test represents the latent construct of interest
- Do the items throughly and completely tap into the latent construct?
- How can we be sure I am teaching the entire domain of psychological testing?
- Create a test blueprint
  - What could be conceivably measured and in what proportion

Assessing Content Validity

- Assume you are giving an instrument to measure aggressive behavior in children
- How can we assume this is measuring the construct of aggression?
  - Experts assess whether each item is essential to the definition of aggression
  - $CVR = \frac{n_e - (N/2)}{N/2}$
  - Where n_e is number say "essential" and N is number of experts
  - Want this larger than chance (Table 6-1)

CVR in *R*

- "Does your child bite other children?"
- 20 experts, 17 say "essential"

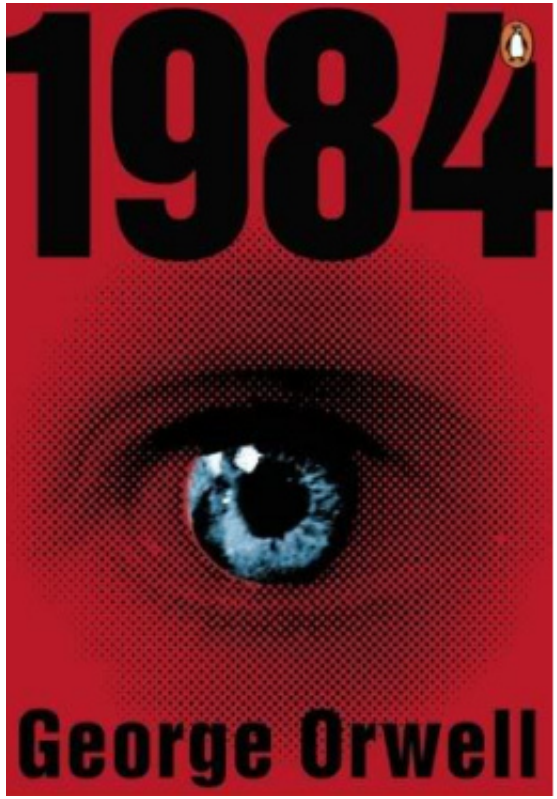
```
CVR <- function(e, N){
        (e - (N / 2))/(N / 2)
        }
# 17 say essential out of 20
CVR(17, 20)

## [1] 0.7
```

- $0.7 > 0.42$

Assessing Content Validity

BUT ... expert judgement!!!

"Who controls the past controls the future; who controls the present controls the past."



source

Criterion-Related Validity

- What the test score tells you about where a person falls on the underlying construct being measured w.r.t a criterion
- A criterion is a benchmark or standard used for comparison
- Score high on an instrument measuring depression, *but do you really have depression*?
- Show no symptoms of depression, instrument is irrelevant and invalid

Measuring Depression

- Predict whether someone is receiving counseling services based on Beck Depression Inventory
- Find out BDI was used to determine whether someone should receive services
- What is wrong with this?

Forms of C-R Validity

- Concurrent Validity
  - Instrument provides the same "scores" as an already validated measure
  - Instruments must be administered at the same time (or nearly so)
  - Example?
- Predictive Validity
  - How well an instrument predicts some criterion in the future
  - SAT should measure "college success"
  - So it should be highly correlated with?
- validity coefficient: an "appropriate" measure of association

Validity Coefficient

- In summary, everything that affects the correlation coefficient!
- Range restriction from attrition in a study or self-selection
- Make sure testtakers are relevant in the validation study and cover the scope of the test!
- Read the test manual and make sure test is appropriate for your testtakers
- Coefficient should be high enough to matter

Incremental Validity

- Want to predict final grade in first math class in college.
- Add most important predictor first (maybe SAT math score if in the USA)
- Then add additional variables, incrementally, and see what each predictor adds
- This is akin to stepwise regression in multiple regression
- This is unwise because of inflation of type I error (the probability of incorrectly rejecting a null hypothesis when you should have retained it)

Construct Validity

- Evidence supporting that the test measures the underlying construct and that it can spread testtakers along that construct
- A test maker has theories about the construct, it's definition, structure, and relationship to other constructs and has theories about how their test relates to other tests
- All forms of validity are really subsumed within construct validity

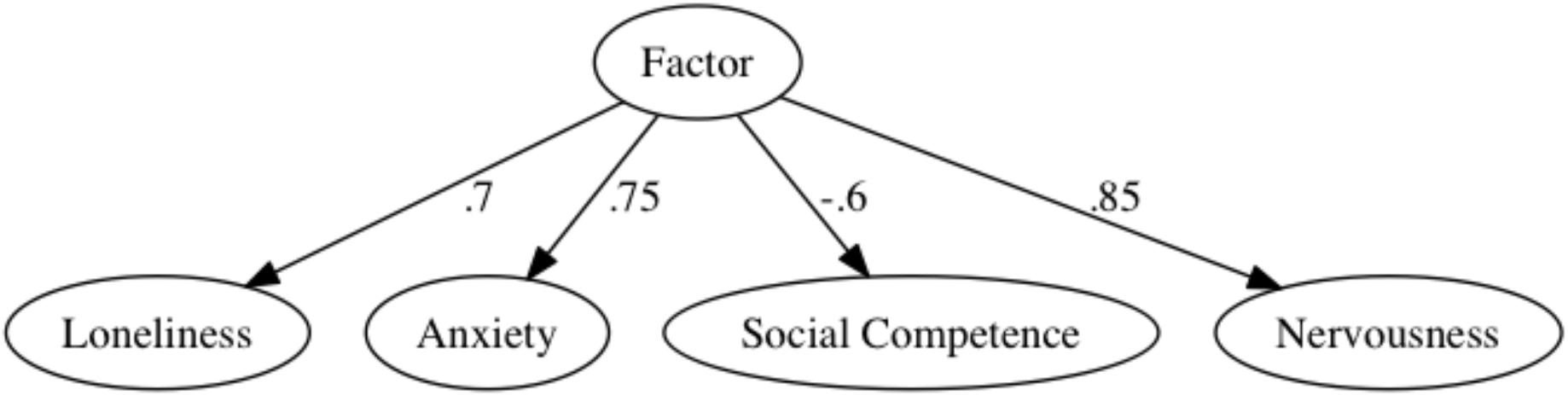
Types of Construct Validity

- Homogeniety
  - Structure of a test should be homogeneous if it is measuring a single construct
  - Responses to test items should be positively correlated with total score on the test
  - Items that are not need to be removed or rewritten
- Change with age and pre/post
  - Testtakers taking a test on algebra *should* score higher if they are older
  - Students getting tutored in algebra between a pre and post test should score higher on the post test

Types of Construct Validity

- Groups higher on the construct should have higher scores
  - Administer a test measuring tendency toward violent behavior
  - Higher scores on test: General population or prison inmates for assault and battery

Factor Analysis



- What should we call this factor?
- If Nervousness is our new instrument to measure the factor, how well does it do?
- What does it mean that social competence is negatively correlated with our factor?

Test Bias and Fairness

- Test bias - degree to which a test systematically favors one group or another
  - Can test for this statistically using logistic regression model
  - Known as differential item functioning
- Test fairness - the degree to which a test is fair and used in an equitable way
  - Administer a test to a group not involved in the validation sample
  - Maybe some groups of people are just different?