Intro to Stats

Reliability & Validity

Reliability

- The extent to which a scale measures something consistently
- Any measurement is an observed score
- An observed score is the sum of an unknown true score and an error score
- Less error = observed score is closer to true score (more reliable)

Test-retest reliability

- Extent to which a test is reliable over time
- Calculate the Pearson coefficient between two time points for each person

Parallel forms reliability

- Extent to which two forms of a test are equivalent
- Calculate the Pearson coefficient between the two forms of the test

Internal consistency reliability

- Extent to which items are consistent with one another and represent one dimension
- Correlation between individual scores and the total score
- Chronbach's alpha (α)

Chronbach's alpha

$$\alpha = \left(\frac{k}{k-1}\right)\left(\frac{s^2_y - \Sigma s^2_i}{s^2_y}\right)$$

k = number of items

S^2_y = variance associated with observed score

$\Sigma\, s^2_i$ = sum of all variances for each item

Interrater reliability

- Agreement between two raters

$$ir = \frac{\text{\# of agreements}}{\text{\# of possible agreements}}$$

Validity

- The extent to which the scale measures what it intends to measure

Content Validity

- Items sample the universe of items for a construct
- Can ask an expert (or several) whether items seem representative

Criterion Validity

- Scale relates to other measures or behaviors in ways that would be expected
- Concurrent or predictive

Construct validity

- Scale measures the underlying construct as intended
- Relation to the behaviors that the construct represents