**Standard Error of Measurement (SEM)**

Reliability is conceived as a correlation between parallel forms:

In sample statistics notation:

Now we can solve for the standard error (standard deviation of the error scores)

=

Solving for Se, we get

This is the SEM, and since 1 – *rxx* is the proportion of error, this is essentially taking the proportion of error () of the standard deviation – it is the amount of error in the typical deviation of scores. We use this to interpret scores – it is the “sampling” error in a given score.

To estimate the SEM, we simply need the score standard deviation (*Sx*) and an estimate of reliability. Which estimate of reliability we need depends on how we plan to use scores.

The question now is about what sources of sampling error we care to consider – given over what conditions we hope to generalize scores. Typically, we want to generalize scores over items, over time, over contexts/conditions. Different forms of generalization introduce different forms of error and require different sources to be captured in our estimate of reliability.

**Estimating Reliability** (score consistency across…)

1. Stability
   1. Test-retest – ability to generalize over time or across testing periods – long range prediction
   2. When trait being measured is thought of as being stable
   3. Affected by practice and other psychological conditions
   4. To interpret, we need to know the time period between testing and relevant intervening experiences
   5. Estimate *r*x1x2 for time 1 and time 2
2. Equivalence
   1. Two forms are administered simultaneously
   2. Ability to generalize in terms of the items selected on a given form
   3. We can also generalize from the sample of items on the forms to the pool of items in a domain
   4. Classically, this is the parallel form correlation
   5. Estimate *r*xx’
3. Stability and equivalence
   1. Allows generalization over forms and over time
   2. Estimate *r*x1x’2
4. Internal consistency
   1. One form administered one time
   2. We want to generalize from the sample of items on the form to the population of items in the pool or domain
   3. Treats item variance as error variance – each item is considered an observation of the trait
   4. Split-half method – splitting the test in half and correlating the two halves
   5. KR-20 – for dichotomously scored items
   6. Coefficient alpha is a function of the average inter-item correlations
   7. Assumes parallel forms
   8. Estimate: 

Congeneric Reliability Coefficients

Alternatives for internal consistency estimates that meet the relaxed assumptions of the congeneric measurement model.

Summary

|  |  |  |
| --- | --- | --- |
|  | Forms | |
| Sessions | **1** | **2** |
| **1** | *Internal Consistency*  Sampling items—Heterogeneity  α; split-half; KR20  Congeneric estimates (Omega, ω) | Equivalence Sampling items  Alternate Form correlation (concurrent) |
| **2** | Stability Sampling time Test-Retest correlation | *Stability & Equivalence*  Sampling time and items  Alternate Form correlation (with delay) |