Measurement Strategies

I. Reliability
A. Classical Testing Theory
1. Reliability = The Consistency of a Measure (Does it always measure the same thing)
2. Classical Testing Theory
   2 Assumptions
      a. All measures suffer from inaccuracies (Unreliability)
         - random vs. systematic
      b. More Measurements = More Reliability

A. Classical Testing Theory
3. Modeling Reliability
   - $x_i = t_i + e_i$
   - Where: $x_i$ = Subject Score
     $t_i$ = True Score
     $e_i$ = Error
   - More Generally
     $\sigma^2_x = \sigma^2_{true\ score} + \sigma^2_{error}$
A. Classical Testing Theory

3. Estimating Reliability

- Reliability Coefficient – ratio of true score variance to total variance

\[
\text{Reliability} = r_{xx'} = \frac{\sigma_t^2}{\sigma_x^2} = \frac{\sigma_t^2 - \sigma_e^2}{\sigma_x^2}
\]

- If \( \sigma_x = \sigma_{\text{error}} \), Then Reliability = ?
- If \( \sigma_{\text{error}} = 0 \), Then Reliability = ?
- If \( \sigma_x = \sigma_{\text{true score}} \), Then Reliability = ?
- \( 1 - r_{xx'} \) = ratio of error variance to total variance
B. Estimating Reliability

1. Multiple Presentations of a Test
   a. Test-Retest Reliability
      - Give Same Test at two Time points \((X_1 \text{ & } X_2)\)
      - Correlate Results – High Positive Correlations = Good Reliability
      - Moderate Reliability = .50 across 1-3 months
      - Practice Effects?
      - State vs. Trait Constructs?
   b. Parallel Forms
      - Present 2 Conceptually Identical Forms
      - Items differ
      - Control Practice Effects
      - Non-equivalence vs. unreliability?
B. Estimating Reliability
1. Multiple Presentations of a Test
c. Reliability Coefficient = $r_{xx'}$

\[ r_{xx'} = \frac{\sigma^2_x}{\sigma^2_x} \]

Numerator = Covariance Between Time 1 and Time 2 (or Form 1 and Form 2)
Denominator = pooled variance for T1 and T2 (Form 1 and Form 2)
B. Estimating Reliability

2. Internal Consistency – Item Covariation
   - Single Test with Multiple Items

a. Split Half Reliability
   - Split test into equivalent halves
     - e.g. odd/even, random split, first $\frac{1}{2}$ - second $\frac{1}{2}$. (essentially Parallel Forms)
     - Correlate one half with other half

\[
R_{\text{sh}} = \frac{\sum X_oX_e - \frac{1}{n}(\sum X_o)(\sum X_e)}{\sqrt{\sum X_o^2 - \frac{(\sum X_o)^2}{n}} \sqrt{\sum X_e^2 - \frac{(\sum X_e)^2}{n}}}
\]

KISS Attitudes Test - KAT
Please carefully read the statements below and indicate the degree to which you agree with each statement using the following scale:

1) Strongly disagree, 2) disagree, 3) Neither Agree nor Disagree, 4) Agree, 5) Strongly Agree

1. Fire Breathing is the quintessence of cool.
2. Men in make-up make me squeal with delight.
3. I often wish I was “evil incarnate.”
4. Long tongues make me squeal with delight.
B. Estimating Reliability

2. Internal Consistency – Item Covariation

a. Split-Half Reliability

- Problems
  - Assumes that all items are of equal difficulty for performance tests
  - Reliability for same data can be different depending on how you half the test
  - Under-estimates reliability – estimate based on only \( \frac{1}{2} \) of actual test length
    - Longer tests are typically more reliable
    - Dilutes the impact of bad items

- Spearman-Brown Prophesy Formula
  estimates the reliability for whole test based on the correlation for \( \frac{1}{2} \) the test

\[
r_{xx'} = \frac{2r_{oe}}{1 + r_{oe}}
\]

Where \( r_{oe} = \) correlation between odd and even items
Where \( r_{xx'} = \) estimated reliability for full test.
B. Estimating Reliability
2. Internal Consistency – Item Covariation
a. Split-Half Reliability
   - Spearman Brown Prophesy Formula
   - General Form – estimates reliability ($r_{xx'}$) of a test that is $n$ times longer.

$$r_{xx'} = \frac{nr_{oe}^{*}}{1 + r_{oe}^{*}(n - 1)}$$

B. Estimating Reliability
2. Internal Consistency – Item Covariation
a. Split-Half Reliability
   - Spearman Brown Prophesy Formula
   - Special Case: Estimates the number of times longer ($n$) a test will need to be to reach a desired level of reliability ($r_{xx'}$).

$$n = \frac{r_{xx'}(1 - r_{oe}^{*})}{r_{oe}^{*}(1 - r_{oe}^{*})}$$
B. Estimating Reliability
2. Internal Consistency – Item Covariation
b. Chronbach’s Alpha
- Average intercorrelation between all of the items in a scale.
- All items should be responded to consistently. Minimum of .70
- Represents all possible split half reliabilities

B. Estimating Reliability
2. Internal Consistency – Item Covariation
b. Cronbach’s Alpha
- Item Intercorrelations
\[ r_{ij} = \frac{\sum (X_i - \bar{X}_i)(X_j - \bar{X}_j)}{\sqrt{\sum (X_i - \bar{X}_i)^2} \sqrt{\sum (X_j - \bar{X}_j)^2}} \]
- KAT Example
  - \( r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34} \)
B. Estimating Reliability

2. Internal Consistency – Item Covariation
   b. Cronbach’s Alpha
      - Conceptually
      \[ \alpha = r_{ij} = \frac{\sum r_{ij}}{K(K-1)/2} \]
      - More Precisely
      \[ \alpha = \frac{K(\bar{r}_{ij})}{1 + (K - 1)\bar{r}_{ij}} \]

3. Internal Consistency – Error Estimation
   a. Kuder-Richardson Formula 21
      - Used for performance tests – (correct/incorrect item scoring)
      - Very useful when item level data is not available
      - Assumes all items are of equal difficulty
      - Tends to be a conservative estimate of reliability
B. Estimating Reliability

3. Internal Consistency – Error Estimation

a. Kuder-Richardson Formula 21

\[ KR21 = \left( \frac{K}{K - 1} \right) \left( 1 - \frac{\bar{X}(K - \bar{X})}{K(\sigma^2)} \right) \]

Where

- \( K = \# \) of items
- \( \sigma^2 = \) variance of scores

\[ \sigma^2 = \frac{\sum X^2 - (\sum X)^2/N}{N} \]

---

**KISS Information Standard Skills Exam (KISSE) Questions.**

<table>
<thead>
<tr>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What band did Gene Simmons and Paul Stanley quit to form KISS?</td>
</tr>
<tr>
<td>2. In what state did KISS start their career?</td>
</tr>
<tr>
<td>3. What instrument does Ace Frehley play?</td>
</tr>
<tr>
<td>4. Which Original Band Member wore cat face makeup?</td>
</tr>
<tr>
<td>5. On what album did KISS first appear without makeup?</td>
</tr>
<tr>
<td>6. Which band member appeared on National Public Radio and claimed to have slept with over 5000 women?</td>
</tr>
</tbody>
</table>

**Answers:**

1. Wicked Lester
2. New York
3. Lead Guitar
4. Peter Criss; later replaced by Eric Car [died Nov. 1991, and replaced by Eric Singer]
5. Lick It Up
6. Gene Simmons
B. Estimating Reliability

3. Internal Consistency – Error Estimation

b. Kuder-Richardson Formula 20

- Used for performance tests – (correct/incorrect item scoring)
- Must have item level data
- Does not require items of equally difficulty
- Is analogous to Alpha (with respect to interpretation)

\[ r_{kr.20} = \left( \frac{k}{k - 1} \right) \left( 1 - \frac{\sum pq}{\sigma^2} \right) \]

- \( p \) = (# respondents getting it correct)/n
- \( q \) = (# respondents getting it wrong)/n : Or (1-p)
- \( \Sigma pq \) = pq summed across all items
- \( \sigma^2 \) = variance for the total test scores.
- \( k \) = number of items.