Introduction

- What is measurement?
  - _____________ consideration
  - _____________ consideration
- Reliability
- Error
  - _____________
  - Non-___________
Reliability

- Degree of:
  1. dependability,
  2. consistency, or
  3. ____________

- ...of ____________ on a measure

\[ X_{\text{obtained}} = \text{______} + X_{\text{error}} \]

Where:
- \( X_{\text{obtained}} \) = obtained score for a person on a measure
- \( \text{______} \) = true score on the measure, that is, actual amount of attribute measured that person really possesses
- \( X_{\text{error}} \) = error score on the measure that is assumed to represent random fluctuations or chance factors
Relationship Between Errors of Measurement and Reliability of a Measure for Hypothetical Obtained and True Scores

Things to Consider Before Selecting Reliability Method

- ____________________ of assessment?
- Will data hold for future?
- Are scores ____________________?
- Do raters agree?
- What role does _________ play in rating?
Methods of Assessing Reliability

- Test - __________
- Parallel forms
- Internal __________
- Inter-rater

Test-___________ Reliability

- Administer measure __________
- Correlate two sets of scores
- Pearson product-moment __________ coefficient
Illustration of Design for Estimating Test-retest Reliability

<table>
<thead>
<tr>
<th>Job Applicant</th>
<th>Test Scores Time 1 (t1)</th>
<th>Retest Scores Time 2 (t2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Rank</td>
</tr>
<tr>
<td>J. S. Friedman</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>J. A. Fukai</td>
<td>87</td>
<td>2</td>
</tr>
<tr>
<td>B. Y. Woodward</td>
<td>80</td>
<td>3</td>
</tr>
<tr>
<td>T. A. Hinata</td>
<td>70</td>
<td>4</td>
</tr>
<tr>
<td>A. C. Zimiski</td>
<td>56</td>
<td>5</td>
</tr>
</tbody>
</table>

**NOTE:**
- Test-retest reliability for Time 1—Time 2 (Case A) = 0.94
- Test-retest reliability for Time 1—Time 2 (Case B) = 0.07

Parallel Forms Reliability

- **AKA:**
  - Equivalent forms reliability
  - ____________ forms reliability

- Consistency with which an attribute is measured

- Using ____________ of a measure
Procedures for Developing Parallel Forms Test

Internal __________ Reliability

- Extent to which all parts of a measure
  - items or questions
- Are similar ________________
- Variety of methods
  - Split __________
  - KR-20
  - ________________ alpha
Development of Odd-even (Day) Split-half Reliability Computed for a Job Performance Criterion Measure

Prophecy Formula

\[
\hat{r}_{te} = \frac{nr_{12}}{1 + (n - 1)r_{12}}
\]

*Where:*
- \( \hat{r}_{te} = \) the corrected split-half reliability coefficient for the total selection measure
- \( n = \) number of times the test was increased in length
- \( r_{12} = \) the correlation between Parts 1 and 2 of the selection measure
Data Used in Computing
K-R 20 Reliability Coefficient

<table>
<thead>
<tr>
<th>Applicant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiley Boyles</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Tammy Allen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Beryl Hesketh</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Sidney Craft</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>David Speed</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Cheri Ostroff</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Number Correctly Answering Item
4 1 3 5 4 3 4 4

NOTE: 0 = incorrect response; 1 = correct response.

Formula

\[ r_{tt} = \frac{k}{k-1} \left( \frac{\sum p_i (1 - p_i)}{\sigma_y^2} \right) \]

Where:
- \( k \) = number of items on the test
- \( p_i \) = proportion of examinees getting each item(i) correct
- 1-\( p_i \) = proportion of examinees getting each item(i) incorrect
- \( \sigma_y^2 \) = variance of examiner's total test scores
## Applicant Data in Computing
### Coefficient Alpha (\(\alpha\))

<table>
<thead>
<tr>
<th>Items Used to Assess Applicant</th>
<th>Case 1 Applicants</th>
<th>Case 2 Applicants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness—A Personality Trait</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1. Dependability</td>
<td>4 1 4 2 4 3 5 2</td>
<td>1 4 4 5 3 2 1 2</td>
</tr>
<tr>
<td>Item 2. Organized</td>
<td>5 1 4 2 4 4 5 2</td>
<td>2 3 1 4 3 3 1 5</td>
</tr>
<tr>
<td>Item 3. Hardworking</td>
<td>5 1 4 2 5 3 5 3</td>
<td>4 1 3 3 4 4 3 4</td>
</tr>
<tr>
<td>Item 4. Persistence</td>
<td>4 1 5 3 5 2 5 3</td>
<td>2 5 5 2 1 5 3 3</td>
</tr>
<tr>
<td>Total Score</td>
<td>18 4 17 9 18 12 20 10</td>
<td>9 13 13 14 11 14 8 14</td>
</tr>
</tbody>
</table>

**NOTE:** Applicants rate their behavior using the following rating scale:
1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, and 5 = Strongly Agree

Case 1 coefficient alpha reliability = 0.83
Case 2 coefficient alpha reliability = 0.40

## Cronbach’s

\[
\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum \sigma_{i}^2}{\sigma_{y}^2} \right)
\]

Where:
- \(k\) = number of items on the selection measure
- \(\sigma_{i}^2\) = variance of respondents' scores on each item(i) on the measure
- \(\sigma_{y}^2\) = variance of respondents' scores on the measure
Interrater Reliability

- Sources of measurement error
  - ________ is being rated (e.g., employee behavior)
  - __________ the rating (rater characteristics)

- Measure of agreement between two or more raters
  - Interrater Agreement
  - ________ class Correlation
  - ________ class Correlation

Research Design for Computing Intraclass Correlation to Assess Interrater Reliability of Employment Interviewers

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. J. Mitra</td>
<td>9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>N. E. Harris</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>R. C. Davis</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>T. M. Zuckerman</td>
<td>7</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

NOTE: The numbers represent hypothetical ratings of interviewees given by each interviewer.
Factors Influencing the Reliability of a Measure

- _________ of estimating reliability
- Individual differences among respondents
- _________ of a measure
- Test question ____________

Factors (cont’d)

- _________ of a measure’s content
- Response format
- ______________ of a measure
Relation among Test Length, Probability of Measuring True Score, and Test Reliability

Illustration of Relation Between Test Question Difficulty and Test Discriminability

**Case 1:**
Item Difficulty = 0.10
10% Get Test Question Correct

**Case 2:**
Item Difficulty = 0.50
50% Get Test Question Correct
Standard Error of

- Estimated error in _______________ score on the measure
- Calculating standard error of _______________

\[ \sigma_{\text{meas}} = \sigma_x \sqrt{1 - r_{xx}} \]

where
- \( \sigma_{\text{meas}} \) = the standard error of measurement for measure X
- \( \sigma_x \) = the standard deviation of obtained scores on measure X
- \( r_{xx} \) = the reliability of measure X.

Uses of the SEM

1. Shows that scores are approximation represented by _______________ of scores on measure

2. Aids decision making in which only _______ is involved

3. Can determine whether scores for individuals _______________ from one another

4. Helps establish confidence in scores obtained from different groups of respondents
Guidelines for Interpreting Individual Scores Using SEM

- Difference between two individuals’ scores should not be considered _________ unless difference is at least twice the SEM of measure

- Before difference between scores of same individual on two different measures should be treated as significantly different, the difference should be greater than _______ _________ of either measure

That’s all folks!

BrianMillerPhD@gmail.com