Test Analysis Report

Detailed information regarding the statistical performance of an test
Overview

- Detailed information regarding the statistical performance of a test

Report Sections

- First section presents information in a tabular format
  - Stats including mean, standard deviation and reliability

- Second section: Distribution of tests scores using a frequency distribution table
  - lists all scores obtained by examinees
  - histogram graphically displays the range of scores
# Test Analysis Report

- First section presents information in a tabular format
  - Stats including mean, standard deviation and reliability

<table>
<thead>
<tr>
<th>Assessment description</th>
<th>Summative assessment on Sociology course</th>
<th>Assessment ID</th>
<th>7081124475921145812</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment name</td>
<td>Sociology 101 final exam</td>
<td>Assessment author</td>
<td>Prof. J. Bloggs</td>
</tr>
<tr>
<td>Assessment last modified</td>
<td>March 13, 2006</td>
<td>Report date &amp; time</td>
<td>December 31, 2005 - 12:00</td>
</tr>
</tbody>
</table>

**Filters**

Group name that contains all the letters of the ALPHA-bet

<table>
<thead>
<tr>
<th>Table of Test Statistics</th>
<th>Number of examinees</th>
<th>Number of items</th>
<th>Minimum score</th>
<th>Maximum score</th>
<th>Test reliability (Cronbach’s Alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35</td>
<td>50</td>
<td>12/50 (24%)</td>
<td>50/50 (100%)</td>
<td>0.831</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>Mode</td>
<td>Standard deviation</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>30/50 (60%)</td>
<td>28/50 (56%)</td>
<td>31/50 (62%)</td>
<td>5/50 (10%)</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Standard error of mean</td>
<td>Standard error of measurement</td>
<td>Skew</td>
<td>Kurtosis</td>
<td>1.599</td>
</tr>
</tbody>
</table>

* Reliability is most meaningful if all items cover the same subject area
Test Analysis Report

- Second section: Distribution of tests scores using a frequency distribution table
  - lists all scores obtained by examinees
  - histogram graphically displays the range of scores
Report Management and Distribution

- Distribute via Browser, PDF, Word, Excel, Email, or Print
- My Reports
Internal Consistency

- Questionmark uses Cronbach’s Alpha Test
- Reliability
- Reason it’s the gold standard:
  - Highly documented and very well researched (60+ years)
  - Can be used with a wide variety of question types (partial credit and right/wrong)
  - Most widely used method in high and medium stakes
Cronbach’s Alpha

- Internal consistency test reliability: A measure of how well the assessment questions relate to one another to measure the same topic.
- Crucial information for legal defensibility – is the test reliably measuring what was designed to measure?
- Speaks to the quality of assessment questions: Are they of high statistical quality?

\[
\frac{k}{k-1} \left( \frac{S_T^2 - \sum S_p^2}{S_T^2} \right)
\]
Other Measures

- Skew
- Mode
- Kurtosis
- Mean (arithmetic)
- Median
- Mode
- Standard Deviation
- Variance
- Standard Error of measurement
- Standard Error of the Mean
Skew

A measure of the symmetry of the distribution of scores

- Are scores are “pushed” or skewed to one side or the other?
- Ranges from about -2 to +2
Kurtosis

\[
\left\{ \frac{n(n + 1)}{n(n - 1)(n - 2)(n - 3)} \sum (z)^4 \right\} - \frac{3(n - 1)^2}{(n - 2)(n - 3)}
\]

- A measure of the symmetry of a distribution of scores
  - How peaked/pointed versus flat are the distribution of scores? What is happening at the tails?
  - Normal range from about -3 to +3
Mean (arithmetic)

- The most commonly used measure of central tendency (refers to the “middle” of a distribution of scores)
- Range of values depends on scores
Median

- Another measure of central tendency, less sensitive than the mean to outliers
- Range of values depends on scores
- Where 50% of participants obtained higher scores and 50% of participants obtained lower scores
Mode

- A third measure of central tendency, used a great deal in survey analysis
- The most common score in a distribution of scores
- Range of values depends on scores

Mode = 56%
Standard deviation

- The “spread” or variation of scores between participants
- Are the scores spread out (e.g., 0 to 100%) or clustered together (e.g., all scores between 55 and 62%)
- Range of values depends on scores

Rick’s test
score = 75%

Sally’s test
score = 83%

Mark’s test
score = 53%

Ella’s test
score = 91%

Standard deviation = 16.36%
Variance

- Another measure of variation
- The first step in calculating a standard deviation: The standard deviation is the square root of the variance
- Range of values depends on scores
- Used in some advanced calculations (e.g., Analysis of variance: ANOVA, Multiple analysis of variance: MANOVA)
Standard Error of measurement

\[ \text{observed test score} = \text{theoretical true score} + \text{error} \]

**Product knowledge test**

- Theoretical test score = 66.4%
- Theoretical test score = 63.7%
- Theoretical test score = 67.1%
- Theoretical test score = 65.8%
- Theoretical test score = 67.5%

Rick’s observed score = 66.1%

\[ s_x \sqrt{1 - r_{xx}} \]

Theoretical standard deviation = 1.26%

± 1.26% of error surrounding Rick’s observed score
Standard Error of the Mean

Sample mean = 56.78%
Sample standard deviation = 15.21%
Standard error of mean = 1.23%

Typically the population information is not known, but if you could see the information...

Population “true” mean = 57.81%
More information:

www.questionmark.com