Research Methodology: Tools

Applied Data Analysis (with SPSS)

Lecture 02: Measurement Scales & Item Analysis

March 2011

Prof. Dr. Jürg Schwarz     juerg.schwarz@hslu.ch

<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aims</td>
</tr>
<tr>
<td>Questionnaire</td>
</tr>
<tr>
<td>Outline</td>
</tr>
<tr>
<td>Quality of a questionnaire: Objectivity, Reliability, Validity</td>
</tr>
<tr>
<td>Measurement Scales</td>
</tr>
<tr>
<td>Item Analysis</td>
</tr>
<tr>
<td>Reliability Analysis using SPSS</td>
</tr>
</tbody>
</table>
# Table of contents

## Aims
- Aims of the Lecture ............................................. 5

## Questionnaire
- Two examples of questionnaires ........................................... 6
  - Typical social science survey – European Social Survey (www.europeansocialsurvey.org) ............ 7

## Outline
- Quality of a questionnaire: Objectivity, Reliability, Validity .................................................. 9

## Measurement Scales
- Types of Scales .......................................................................................................................... 11
  - Nominal Scale ....................................................................................................................... 12
  - Ordinal Scale ........................................................................................................................ 13
  - Metric Scale .......................................................................................................................... 14
- Properties of Scales .................................................................................................................. 15
- Hierarchy of Scales ................................................................................................................... 16
- Summary: Type of Scales ......................................................................................................... 17

## Exercises: Scales
- Item batteries and constructs .................................................................................................. 20
- Example NHIS (National Health Interview Survey) (www.cdc.gov/nchs) ................................. 20

## Item Analysis
- Item Difficulty .......................................................................................................................... 22
  - Fictional example of a survey on taxes .................................................................................. 23
  - Item difficulty rules of thumb ............................................................................................... 24
- Item Discrimination .................................................................................................................. 25

## Reliability Analysis using SPSS
- Introduction ............................................................................................................................. 27
  - Batteries of Likert-type items ............................................................................................... 27
  - How to measure reliability ..................................................................................................... 27
- An Example ............................................................................................................................. 28
  - European Social Survey 2004 (Swiss data) ........................................................................ 28
  - SPSS Output .......................................................................................................................... 30
  - Interpretation of the results .................................................................................................. 31
Aims

Aims of the Lecture

You know that questionnaires contain – among other things – items and item batteries.
You know the quality features: objectivity, reliability, and validity.
You know that quality features are related hierarchically.
You know four types of measurement scales: nominal, ordinal, interval, and ratio.
You know that measurement scales are related hierarchically.
You know item difficulty.
You are able to calculate item difficulty with SPSS.
You know item discrimination.
You are able to calculate item discrimination with SPSS.
You know reliability, measured by Cronbach’s Alpha.
You are able to calculate Cronbach’s Alpha with SPSS.

Questionnaire

Two examples of questionnaires


Binary question

What is this person’s sex? Mark ONE box.

☐ Male
☐ Female

Multiple question

Is this person a CITIZEN of the United States?

☐ Yes, born in the United States → Skip to 15a
☐ Yes, born in Puerto Rico, Guam, the U.S. Virgin Islands, or Northern Mariana
☐ Yes, born abroad of American parent or parents
☐ Yes, a U.S. citizen by naturalization
☐ No, not a citizen of the United States
Typical social science survey – European Social Survey (www.europeansocialsurvey.org)

Single item of Likert scale type (R. Likert, 1903 - 1981, American psychologist)

<table>
<thead>
<tr>
<th>B29</th>
<th>STILL CARD 11</th>
<th>Still using this card, please say what you think overall about the state of health services in [country] nowadays?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Extremely bad</td>
<td>Agree strongly</td>
</tr>
<tr>
<td></td>
<td>00</td>
<td>01</td>
</tr>
</tbody>
</table>

Set of items ("item battery")

<table>
<thead>
<tr>
<th>CARD 12</th>
<th>Using this card, please say to what extent you agree or disagree with each of the following statements. READ OUT EACH STATEMENT AND CODE IN GRID</th>
</tr>
</thead>
<tbody>
<tr>
<td>B30</td>
<td>The government should take measures to reduce differences in income levels</td>
</tr>
<tr>
<td></td>
<td>Agree strongly</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>B31</td>
<td>Gay men and lesbians should be free to live their own life as they wish</td>
</tr>
<tr>
<td></td>
<td>Agree strongly</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>B32</td>
<td>Political parties that wish to overthrow democracy should be banned</td>
</tr>
<tr>
<td></td>
<td>Agree strongly</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>B33</td>
<td>Modern science can be relied on to solve our environmental problems</td>
</tr>
<tr>
<td></td>
<td>Agree strongly</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Outline

Questionnaire as an instrument in research

Questionnaires are used for survey research and also experiments. Questions – also called **items**, measure the value of attributes with a certain type of scale. There are four types of scales: nominal, ordinal, interval and ratio.

Rating scale items are used to measure the direction and intensity of attitudes. Batteries of rating scale items are used to measure constructs.

**Note**: "scale" is used as a general term for the way in which data is processed. An important example would be aggregated scales, which are made up of multiple individual measurements (e.g. sum scale of an item battery which represents the value of a construct).

How to measure, rate and improve the quality of a questionnaire?

Conduct item analysis
- Item difficulty
- Item discrimination

Conduct reliability analysis
- Cronbach’s Alpha
Quality of a questionnaire: Objectivity, Reliability, Validity

Objectivity
- Objectivity ("Objektivität") of an instrument is given, if it is independent of the tester in administration, and also if the results are independent of the method of calculation.
  
  **Example:** Two independent examiners use the same questionnaire in two statistically comparable samples. The instrument (questionnaire) is objective only if the values are statistically comparable (same mean, same variance in the limits of confidence intervals).

Reliability
- Reliability ("Zuverlässigkeit") is the degree to which an instrument (questionnaire) measures the same way each time it is used. Reliability is inversely related to random error.
  
  **Example:** The same questionnaire is used twice in one sample, for example, one month ago and today. The instrument (questionnaire) is reliable only if the values are statistically comparable (same mean, same variance in the limits of confidence intervals).

Validity
- Validity ("Gültigkeit") is the extent to which a test measures what it was intended to measure.
  
  **Example:** A questionnaire measures the level of a certain attitude of persons in a sample. The instrument (questionnaire) is valid only if the value is statistically comparable with the population (same mean, same variance in the limits of confidence intervals).

Hierarchy

Objectivity does not imply reliability.
Reliability does not imply validity. A reliable questionnaire measures consistently, but not necessarily what it is supposed to be measuring.

Comparison with target practice: reliability is consistency, validity is accuracy.

- Hits are distributed all over the target. Neither consistent nor accurate => Neither reliable nor valid
- Hits are concentrated in one area but not in the center of the target. Consistent but not accurate => Reliable but not valid
- Hits are concentrated and also located in the center of the target. Consistent and accurate => Reliable and valid
Measurement Scales

Types of Scales

Attributes of measurement objects can be measured with four different types of scales*.

Example: Survey of health-state:

<table>
<thead>
<tr>
<th>Measurement object</th>
<th>Sex</th>
<th>Attitude to health</th>
<th>Temperature</th>
<th>Net-Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute of Object</td>
<td>Male / Female</td>
<td>1 to 5</td>
<td>Real number</td>
<td>Real number</td>
</tr>
<tr>
<td>Value of Attribute</td>
<td>Nominal</td>
<td>Ordinal</td>
<td>Interval</td>
<td>Ratio</td>
</tr>
</tbody>
</table>


Nominal Scale

Composed of "names" (categories). Must be measured in statistical sense which means distinct. Names do not have any specific order.

Examples:
- Gender is either male or female
- Types of cancer treatment include surgery, radiation therapy and chemotherapy.

Associate numbers to a nominal scale by way of assigning an arbitrary code to each category.
Ordinal Scale

Consists of an ordering in the values of a measurement.

Examples:

- Disease severity measured in ordered values (none, mild, moderate, serious, critical).
- Self-perception of health ordered from very bad to very good on a 5-point Likert scale.

Associate numbers to ordinal scale by way of assigning an ordered code to each category.

<table>
<thead>
<tr>
<th>Disease severity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mild</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>moderate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>serious</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>critical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Self-perception of health ordered from very bad to very good on a 5-point Likert scale.

<table>
<thead>
<tr>
<th>Self-perception</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>very bad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>poor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>good</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very good</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Metric Scale

Reflects characteristics which can be measured exactly in terms of quantity.

Examples:

- Clinical measurements, such as body size, weight, blood pressure.
- Socio-economic status (SES) measurements, such as age, salary, commuting distance.

Associate numbers to metric scale by assigning the value of measurement itself.
## Properties of Scales

<table>
<thead>
<tr>
<th>Level</th>
<th>Determination of ...</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>equality or unequal</td>
<td>Mode</td>
</tr>
<tr>
<td>Ordinal</td>
<td>greater, equal or less</td>
<td>Median</td>
</tr>
<tr>
<td>Interval</td>
<td>equality of differences</td>
<td>Arithmetic mean</td>
</tr>
<tr>
<td>Ratio</td>
<td>equality of ratios</td>
<td>Geometric mean</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th>Possible transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>one-to-one substitution</td>
</tr>
<tr>
<td>Ordinal</td>
<td>monotonic increasing</td>
</tr>
<tr>
<td>Interval</td>
<td>positiv linear</td>
</tr>
<tr>
<td>Ratio</td>
<td>positiv proportional</td>
</tr>
</tbody>
</table>

## Hierarchy of Scales

Nominal scales are the "lowest" and ratio scales are the "highest".

<table>
<thead>
<tr>
<th>Scale may ...</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>not be treated as ordinal, interval or ratio. Male = 0, Female = 1. There is no possibility to order. Only few transformations possible. E.g. add 1: 0 -&gt; 1, 1 -&gt; 2 =&gt; Male = 1, Female = 2.</td>
</tr>
<tr>
<td>Ordinal</td>
<td>be treated as nominal but not as interval or ratio. Self-perception of health ordered from very bad (= 1) to very good (= 5). Nor is very good five times better than very bad, neither has very good a distance 4 to very bad.</td>
</tr>
<tr>
<td>Interval</td>
<td>be treated as ordinal or nominal but not as ratio. Temperature: Difference between 5° and 10° is 5°. Difference between 20° and 25° is also 5°. Differences can be compared. But, 10° is not twice 5°. Compare with Fahrenheit! 10° C = 50° F, 5° C = 41° F</td>
</tr>
<tr>
<td>Ratio</td>
<td>be treated as ordinal, nominal or interval. Salary: $8000.- is twice $4000.- There is a &quot;natural&quot; zero in this scale: $0.- Any transformations possible, e.g. division by 1000.</td>
</tr>
</tbody>
</table>
Summary: Type of Scales

Statistical analysis assume that the variables have specific levels of measurement. Variables that are measured nominal or ordinal are also called categorical variables. Exact measurements on metric scale are preferable. They have the highest level.

Why does it matter whether a variable is categorical or metric? For example, it would not make sense to compute an average gender. In short, an average requires a variable to be metric.

Sometimes variables are "in between" ordinal and metric. Example: likert scale with "strongly agree", "agree", "neutral", "disagree" and "strongly disagree". If it is not clear that the intervals between each of these five values are the same, then it is not a metric variable, but an ordinal variable. In order to calculate statistics, it is often assumed that the intervals are equally spaced.

Many circumstances lead to grouping of metric data into categories – e.g. the salary. Such ordinal categories are sometimes easier to comprehend than exact metric measurements. In this process, however, valuable exact information is lost.

Exercises: Scales

1. Read "Summary: Type of Scales" on slide 17 above.

2. Which type of scale?

<table>
<thead>
<tr>
<th>License number of a car in Switzerland (e.g. ZH 441 010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where do you live?</td>
</tr>
<tr>
<td>north  south  east  west</td>
</tr>
<tr>
<td>Size of T-shirt (XS, S, M, L, XL, XXL)</td>
</tr>
<tr>
<td>Please mark one box ☐ per question</td>
</tr>
<tr>
<td>Compared with the health of others in my age, my health is</td>
</tr>
<tr>
<td>very bad  1  2  3  4  5  very good</td>
</tr>
<tr>
<td>How much did you spend on food this week? ______ $</td>
</tr>
</tbody>
</table>

2.01. Compared with the health of others in my age, my health is very bad or very good.

b. How many people, including this person, usually rode to work in the car, truck, or van LAST WEEK?
- Drove alone
- 2 people
- 3 people
- 4 people
- 5 or 6 people
- 7 or more people
5. What is the highest level of education that you have completed? (please tick the highest level you have completed)

- 1. primary school
- 2. some secondary school
- 3. completed high school
- 4. some additional training (apprenticeship, TAFE courses etc.)
- 5. undergraduate university
- 6. postgraduate university

Please write a number between 1 and 5 next to each item below. Put a 1 next to the item that is MOST important to you in selecting an on-line university course. Put a 5 next to the item that is LEAST important. Please use each number only ONCE.

- a. Availability of instructor for assistance.
- b. Tuition cost for the course.
- c. Ability to work in groups with other students.
- d. Quality and quantity of instructor feedback.
- e. Number of students enrolled.

**Item batteries and constructs**

Example NHIS (National Health Interview Survey) ([www.cdc.gov/nchs](http://www.cdc.gov/nchs))

The psychological term "distress" is a **construct** which is composed of 6 items.

The distress battery is used to examine non-specific distress in the U.S. adult population.

<table>
<thead>
<tr>
<th>Item</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(7)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>So sad that nothing could cheer you up?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Nervous?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Restless or fidgety?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Hopeless?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>That everything was an effort?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Worthless?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
Realization in the questionnaire:

<table>
<thead>
<tr>
<th>ALL</th>
<th>MOST</th>
<th>SOME</th>
<th>A LITTLE</th>
<th>NONE</th>
<th>REFUSED</th>
<th>DON'T KNOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(7)</td>
<td>(9)</td>
</tr>
</tbody>
</table>

Total distress score: Average of all six items

Distress scoring:

- 1.00 – 1.99 High Distress
- 2.00 – 2.99 High-Moderate Distress
- 3.00 – 3.99 Low-Moderate Distress
- 4.00 – 5.00 Low Distress

---

Item Analysis

Item Difficulty

The term "item difficulty" was established in the field of testing people with questionnaires.

Misleading: If the item difficulty is low, the item (question) is difficult.

Item difficulty is also called p-value.

Binary scored items (Yes/No)

Item difficulty is the proportion $p$ of people who "correctly" answered the item (that is with "yes").

"Correct Answer" $\rightarrow$ score = 1   "Incorrect Answer" $\rightarrow$ score = 0

Continous scored items (Rating scale)

Item difficulty is the arithmetic mean of an item.

Calculation of arithmetic means can also be part of the "RELIABILITY"-Command of SPSS.
Fictional example of a survey on taxes

As an example, wording of item 4: “Do you think taxes should be reduced? Yes or No?”

<table>
<thead>
<tr>
<th>Proband</th>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>Item 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>n</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>n</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>n</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>n</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

ny = number of "yes"-responds
nr = number of responding respondents
nt = number of total responds
p = item difficulty (p = ny/nr)

Item 4 is too "easy" and there is no discrimination. This item should be removed.
Item 5 seems to be either too "difficult" or not understandable. This item should also be removed.

Item 1 has mean difficulty. Item 2 is relatively difficult. Item 3 is relatively easy.

Item difficulty rules of thumb

High p-values => item is easy  Low p-values => item is difficult

(Since this is misleading, some researcher tried to change the name item difficulty to item popularity, but this did not happen yet.)

If p-value = 1.0 (or 0.0), everyone answered question correctly (or incorrectly) and there will be no variability in item scores, then item need revision or should be removed.

If p-value is too low, then item is too difficult and need revision.

There should be a mixture of difficulty in items on test.

In order to obtain maximum spread of scores it is best to use items with moderate difficulties.

Once difficulty of items is known, they are usually sorted from easiest to hardest on test.

In the fictional example items should be arranged according to their difficulty:

Item 3 –> 1 –> 2
Item Discrimination

**Definition**

Item discrimination is the correlation between an item and the entire item battery.

Can range in value from -1 to 1.

Since often dealing with dichotomous items (yes/no, 1/0), this correlation is usually either a biserial or point-biserial correlation.

In papers it is sometimes called $r_{pbis}$

**Calculating item discrimination with SPSS**

Since item discrimination is the correlation between item score and total score of an item battery it can be calculated by using "standard" correlation analysis.

Item discrimination is also part of the "RELIABILITY"-Command

In the SPSS "RELIABILITY"-Command it is called **corrected item-total correlation** because the total of the battery is NOT the sum of all item scores, but the sum of item scores WITHOUT including the item in question.

**Example NHIS (National Health Interview Survey)**

<table>
<thead>
<tr>
<th>Item</th>
<th>Correlation*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAD</td>
<td>0.74</td>
</tr>
<tr>
<td>NERVOUS</td>
<td>0.73</td>
</tr>
<tr>
<td>RESTLESS</td>
<td>0.70</td>
</tr>
<tr>
<td>HOPELESS</td>
<td>0.78</td>
</tr>
<tr>
<td>EFFORT</td>
<td>0.71</td>
</tr>
<tr>
<td>WORTHLS</td>
<td>0.73</td>
</tr>
</tbody>
</table>

During the past 30 days, how often did you feel…

- So Sad that nothing could cheer you up? □  C
- Nervous? □ C
- Restless or fidgety? □ C
- Hopeless? □ C
- That everything was an effort? □ C
- Worthless? □ C

All items of the distress battery have relatively high values of item discrimination. They represent the whole battery each individually.

As a general rule, consider dropping or revising items with discriminations **lower than .30**.

Positive values close to 1 are desirable.

If negative, check scaling direction and if it is correct, drop or revise item.

*Corrected Item-Total Correlation. Calculated with SPSS' "RELIABILITY". See next slides.
Reliability Analysis using SPSS

Introduction

Batteries of Likert-type items
- Construct score by simply adding the items together
- Be sure that they are scaled in the same direction

How to measure reliability
- The most common approach to assessing the reliability of an item battery is to use a measure of "internal consistency."
- Roughly, an item battery is internally consistent if all of its items are strongly correlated.
- A high average correlation among the items suggests that they are all measuring the same construct.
- Cronbach's alpha coefficient is widely used to assess reliability. Alpha is a positive function of the average correlation between items in an item battery. Alpha is a positive function of the number of items in the item battery.
  => The higher the average correlation, the lower the error components of items
  => The more items, the greater the likelihood that errors will cancel out.

www.ats.ucla.edu/stat/spss/faq/alpha.html (Date of access: March, 2011)

An Example

European Social Survey 2004 (Swiss data)

Dataset & Syntax: ESS2CH.sav & ESS2CH.sps

Variables in SPSS: mdlswgt, mdhair, mdmemo, mdhappy, mdsexlf
RELIABILITY
/VARIABLES=mdhair mdhappy mdiswgt mdmemo mdsexlf
/SCALE('ALL VARIABLES') ALL
/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE
/SUMMARY=TOTAL.

SPSS Output

Case Processing Summary

<table>
<thead>
<tr>
<th>Cases</th>
<th>Valid</th>
<th>Excluded</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2093</td>
<td>48</td>
<td>2141</td>
</tr>
<tr>
<td>%</td>
<td>97.8</td>
<td>2.2</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Reliability Statistics

<table>
<thead>
<tr>
<th>Cronbach's Alpha</th>
<th>N of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>.803</td>
<td>5</td>
</tr>
</tbody>
</table>

Item Statistics

Since **item difficulty** is the arithmetic mean, you find the difficulties in the "Mean" column.

Item-Total-Statistics

Since **item discrimination** is the correlation between item score and total score, you find the discrimination in the "Corrected Item-Total Correlation" column.

Value of Cronbach's alpha = .803
**Interpretation of the results**

Item difficulties vary between 3.21 and 3.96 which is a reasonable mixture of moderate difficulty. The medium difficulty is 3 which is the center of the 1 to 5 Likert scale.

Cronbach’s alpha = 0.803 is relatively high. The most common rule of thumb is that alpha should exceed .80. In practice, scales with lower reliabilities are often used.

Item discrimination vary between .520 and .632 which reasonable higher than 0.30. There is no reason to drop any of the items. All of them represent the whole battery.

**Notes:**