Assignment: Two aspects of data quality

|  |  |
| --- | --- |
|  | **Unit 8** - Today’s assignment is about two aspects that say something about the quality of measurement instruments: reliability and validity. The exercises will cover the understanding of validity and reliability as well as their differences. |
|  | Measurement Validity and ReliabilityMeasurement reliabilityMeasurement validity |  | Operationalization; Indicator; Reliability; Random error; (Measurement) Validity (and its reverse: data collection bias); Construct validity; Content validity; Criterion-related validity  |
|  | Babbie, Ch. 5, pp. 145-152 |  | In the tutorial setting: most answers will be discussed in groups  |  | 90 minutes |
|  | Read this assignment carefully and answer the questions. Bring either a print or digital version to the tutorial / lecture. |

1. Validity and reliability both refer to the quality of the measurement instruments used to collect data, but to different types of errors. Which types of error are related to *reliability* and *validity* and what is the difference between these two types of errors?

**Random errors:** can be anything, they happen accidentally. You want an instrument to give you the same outcome over and over again under the same circumstances

**Systematic errors:** are systematic. You want to measure the correct construct.

2. Give an example for both types of error using the measurement of someone’s temperature.

Example of random errors: when measuring someone’s temperature, the thermometer seems to be “broken”; the temperature varies 3 degrees (Celsius) while repeating the measurement five times in two minutes.

Example of systematic errors: you want to measure someone’s temperature, but you are using a scale to do this. A scale gives an indication of someone’s weight and so the observation is not valid. Or, you measure the temperature of a person, but you keep your own hand close to the thermometer, which affects the temperature (it becomes closer to yours).

3. Also the word ‘bias’ is used. What is that?

Bias is just another word for invalidity.

4. We often talk about reliable and unreliable operationalizations, and about invalid and valid measures. Reliability and validity are thus ‘variables’ describing ‘operationaliziations’ (our units, you could say). Which level of measurement is used for the variables ‘reliability’ and ‘validity’?

Reliability and validity are not dichotomous variables. It is possible to compute values that indicate the degree of reliability / validity.

5. Below, you will find some examples about measurement instruments (=operationalizations). Indicate whether it is about the reliability or the validity of an instrument.

|  |  |  |
| --- | --- | --- |
| **Example** | **Is it about reliability?** | **Is it about validity?** |
| 1. A researcher constructed a new instrument to measure depression. The relation between someone’s outcome on this new instrument doesn’t seem to be related to older instruments that are measuring depression as well.  | No | Yes |
| 2. Jane searched the internet for a test to determine if she was a good match for her secret crush Mario. She filled in both their names and found out they almost perfectly were fit for each other: 96% She wanted to show a friend the outcome, so she called a friend and let her fill in the same names, on the same website. Unfortunately, now the website indicated it was a match for just 45%.  | Yes | Not explicitly, although I doubt whether the correct construct is measured |
| 3. A history teacher constructed an exam with 40 multiple choice questions about World War II. To examine the quality of the measurement, he looked at the relationship between student scores on the first 20 questions and their scores on the last 20 questions.  | Yes | No |
| 4. A group of researchers wanted to find out if more stress at work is related to a higher BMI. After the data collection process, they found out that the scale used to measure the weight had a defect and overestimated the weight of the participants. | No | Yes |
| 5. Four students code a set of articles from Reuters about the refugee crises. They count the number of articles per day and also code some aspects of the content of articles. They write down what they would code, but code independently. The student have very different conclusions about the content. Some argue that most articles about the problem, whereas others argue it is about solutions. | Yes | Maybe, it is not very clear what is exactly operationalized |

6. There are several ways to look at reliability: the procedure of stability and the procedure of consistency. Which of the examples in the previous exercise was an example of the procedure of stability? And which one was an example of consistency?

Stability: 2

Consistency: 3

Whether 5 is stability or consistency is not clear. Both can be argued. I think it is mainly about consistency.

7 The micro lecture covers three of ways to assess validity: content validity, criterion-related validity and construct validity. There is also something called “face validity”. Please look up this term. What’s the definition of face validity. Why (do you think) this kind of validity is not discussed in the micro lecture?

Definition according to Babbie (The Practice of Social Research, 14th ed.), p.149: “That quality of an indicator that makes it seem a reasonable measure of some variable.”

Face validity is not mentioned in the lecture, probably because face validity merely means ‘seems reasonable to the researcher’, but the researcher is implicitly using other criteria to arrive at that conclusion.

8. Some people think it is difficult to remember the difference between content validity, criterion-related validity and construct validity. This might be due to the fact that they all begin with a “C”. In spite of this similarity, there are clear differences between the three methods of looking at validity. Please indicate for all the following examples which of the validities is mentioned.

|  |  |
| --- | --- |
| **Example** | **Method** |
| 1. A school wants to hire a new science teacher. All candidates are asked to take a test, so the candidate with the highest score can be picked. After evaluation of the test, the board sees that most of the test questions were about physics, but the teacher is also expected to teach biology and chemistry. | Content validity |
| 2. Years ago, researcher tried to measure the IQ of their participants by measuring their head circumference. Later, other researchers found out that head circumference is not related to the IQ of a person according to IQ-tests.  | Construct validity |
| 3. Some tests at primary school try to predict the ability of children in secondary school.  | Criterion-related validity |
| 4. In a study, a new survey is used for measuring social anxiety. As social anxiety, normally is strongly related to social avoidance, an existing survey for measuring social avoidance is also used in the study. After the data collection, the researchers look at the relationship between the outcome of the two surveys.  | Construct validity |
| 5. The outcomes of a self-report measure on conflicts at work are compared to the actual numbers of fights people have at work. | Criterion-related validity |

9. A researcher wants to find out whether a survey she constructed is a reliable and valid instrument for measuring whether someone is a cyber-bully or not. Try to think of a way to look at the reliability and a way to look at the validity of the instrument.

9.1. Reliability:

All questions are measuring cyber-bullying. If there are a lot of questions, you could use the procedure of consistency in which you compare the answers on the questions with each other.

9.2. Validity:

The nicest way to see whether you are really measuring cyber-bullying or not, you could compare the outcomes on the survey with the actual behaviour of the participants. You could, for example, monitor the online behaviour for a month, to get an indication of the behaviour (criterion related validity). But other methods are available too. For example, check whether the survey contains questions about all aspects of cyberbullying (content validity).

10. Take a look at the following multiple choice question. You don’t have to answer this multiple choice question. We’d like you to estimate how many of your fellow students will answer the next question correctly instead. Write down your answer in percentages:

“I think …….. % of my fellow students will give a correct answer to this multiple choice question.”

There is no right or wrong here.

We’ll collect the answers to q10 in the lecture / tutorial and then we’ll show you what it has to do with reliability and / or validity.

*Given*

*A researcher sees that the score of one child on a test that measures the reading ability is three times as high as the score of another child on the same test.*

*Question*

*What is the level of measurement of the variable?*

*A. Nominal*

*B. Ordinal*

*C. Interval*

*D. Ratio*

\*\* we want to operationalize the ‘difficulty of a question’ If students give very different answers, that is an indication of the unreliability of that operationalization. If the actual percentage of students answering that question incorrectly is similar to the average of the estimates, that average may be a considered a (criterion) valid indication of the difficulty of the question.

\*11. One approach to reliability is ‘consistency’. Using the dataset used in the context of unit 5. How would you estimate the consistency (reliability) of the index ‘external efficacy’? First answer the question without actually using SPSS.

Check the extent to which the items of that index are actually correlated.

You can do that by correlate variables.

A more formal procedure, which gives you one single measure of consistency is looking at cronbachs alpha.