

Reliability & Validity of Measurements

Learning Objectives

Introduction

Reliability

Validity

Key Takeaways

Dr. Abhishek Singh
abhishek@nith.ac.in

Department of Mathematics, NIT Hamirpur, (H.P.)



June 26, 2019

Outline



RMDA, 2019

Learning Objectives

Introduction

Reliability

Validity

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

Learning Objectives



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ Define reliability, including the different types and how they are assessed.
- ▶ Define validity, including the different types and how they are assessed.
- ▶ Describe the kinds of evidence that would be relevant to assessing the reliability and validity of a particular measure.
- ▶ Hands on practice in SPSS software.



- ▶ Measurement involves assigning scores to individuals so that they represent some characteristic of the individuals.
- ▶ Consider construct like intelligence, self-esteem, depression, or working memory capacity.
- ▶ How do researchers know that the scores actually represent the characteristic?



- ▶ Conduct research using the measure to confirm that the scores make sense.
- ▶ Collect data & demonstrate that they work.
- ▶ How to demonstrate?
- ▶ Reliability & Validity.



- ▶ Reliability refers to the consistency of a measure.
- ▶ There are three types of consistency:
 1. Test-Retest Reliability (Over Time)
 2. Internal Consistency (Across Items)
 3. Inter-Rater Reliability (Across Different Researchers)

Test-Retest Reliability



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ A construct to be measured is assumed to be consistent over time.
- ▶ Then the scores we obtain should also be consistent over time.
- ▶ This consistency over time is measured through Test-Retest Reliability.

Example

- ▶ A person who is highly intelligent today will be highly intelligent next week.
- ▶ This means that any good measure of intelligence should produce roughly the same scores for this individual next week as it does today.

Computation: Test-Retest Reliability



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ It is calculated using the measure on a group of people at one time, using it again on the same group of people at a later time.
- ▶ Then looking at **test-retest correlation** between the two sets of scores.
- ▶ This is done by graphing the data in a scatter-plot and computing Pearson's correlation coefficient.

Computation: Test-Retest Reliability



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

Scatter-diagram between two sets of scores of several university students on the Rosenberg Self-Esteem Scale, administered two times, a week apart.

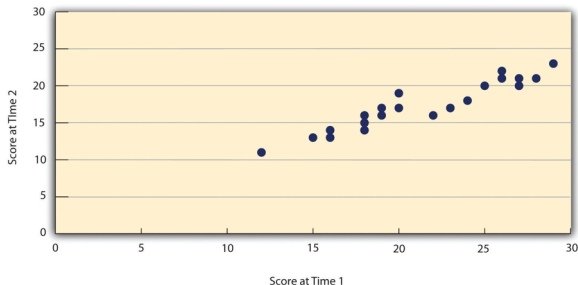


Figure: Scatter-diagram between two sets of scores

Computation: Test-Retest Reliability



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ Pearson's r for these data is $+0.95$.
- ▶ This means that Rosenberg Self-Esteem Scale is a reliable scale in terms of test-retest reliability.
- ▶ In general, a test-retest **correlation of $+0.80$ or greater** is considered to indicate good reliability.

Internal Consistency



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ It is the consistency of people's responses across the items on a multiple-item measure.
- ▶ In general, all the items on such measures are supposed to reflect the same underlying construct.
- ▶ Therefore, people's scores on those items should be correlated with each other.

Example

- ▶ On the Rosenberg Self-Esteem Scale, people who agree that they are a person of worth should tend to agree that that they have a number of good qualities.
- ▶ This is what we call the internal consistency.

Computation: Internal Consistency



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ Like test-retest reliability, internal consistency can only be assessed by collecting and analyzing data.
- ▶ The most common measure of internal consistency is a statistic called **Cronbach's α** .
- ▶ A value of **+ .80 or greater** is generally taken to indicate good internal consistency.

Inter-Rater Reliability



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ Many measures involve significant judgement on the part of an observer or a rater.
- ▶ Inter-rater reliability is the extent to which different observers are consistent in their judgments.

Example

- ▶ To measure university students' social skills, we could make video recordings of them as they interacted with another student whom they are meeting for the first time.
- ▶ Then we could have two or more observers watch the videos and rate each student's level of social skills.

Computation: Inter-Rater Reliability



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ It is assessed using **Cronbach's** α when the judgements are quantitative.
- ▶ In case of categorical judgements an analogous statistic called **Cohen's** κ is used to assess the inter-rater reliability.

Validity



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ Validity is the extent to which the scores from a measure represent the variable they are intended to.
- ▶ A measure can be extremely reliable but have no validity.

Example

- ▶ Suppose people's index finger length reflects their self-esteem.
- ▶ One can measure self-esteem by holding a ruler up to people's index fingers.
- ▶ Although this measure would have extremely good test-retest reliability, it would have absolutely no validity.



- ▶ There are three types of validity:
 1. Face Validity
 2. Content Validity
 3. Criterion Validity

Face Validity



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ Face validity is the extent to which a measurement method appears **on its face** to measure the construct of interest.
- ▶ It is a very weak kind of evidence that a measurement method is measuring what it is supposed to.
- ▶ The reason is that it is based on people's intuitions, which are frequently wrong.



Example

- ▶ Self-Esteem Questionnaire
 - ▶ Whether respondents see themselves as a person of worth?
 - ▶ Whether respondents think they have good qualities?
 - ▶ Finger length of respondent.

Content Validity



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ Content validity is the extent to which a measurement **covers** the construct of interest.
- ▶ Like face validity, content validity is not usually assessed quantitatively.
- ▶ It is assessed by carefully checking the measurement method against the conceptual definition of the construct.



Example

- ▶ Suppose test anxiety is defined as involving:
 - ▶ Sympathetic nervous system activation (leading to nervous feelings)
 - ▶ Negative thoughts
- ▶ Therefore, measure of test anxiety should include items about both nervous feelings and negative thoughts.

Criterion Validity



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ Criterion validity is the extent to which scores on a measure are correlated with other variables known as criteria.



Example

- ▶ Scores on a measure of test anxiety should be negatively correlated with their performance on an exam.
 - ▶ If anxiety scores were negatively correlated with exam performance, then this suggests that these scores really represent test anxiety.
 - ▶ But if it were found that people scored equally well on the exam regardless of their test anxiety scores, then this would cast doubt on the validity of the measure.



- ▶ A criterion can be any variable that should be correlated with the construct being measured.

Example

- ▶ We can expect test anxiety scores to be negatively correlated with exam performance and course grades, and
- ▶ Anxiety scores positively correlated with general anxiety and with blood pressure during an exam.

Criterion Validity



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ When the criterion is measured at the same time as the construct, criterion validity is referred to as **concurrent validity**.
- ▶ When the criterion is measured at some point in the future (after the construct has been measured), it is referred to as **predictive validity**.

Computation: Criterion Validity



RMDA, 2019

Learning
Objectives

Introduction

Reliability

Validity

Key Takeaways

- ▶ Same as Test-Retest Reliability.



- ▶ Researchers do not simply assume that their measures work. Instead, they conduct research to show that they work.
- ▶ There are two distinct criteria by which researchers evaluate their measures: **reliability and validity**.
- ▶ Reliability is consistency **across time** (test-retest reliability), **across items** (internal consistency), and **across researchers** (inter-rater reliability).
- ▶ Validity is the extent to which the scores actually represent the variable they are intended to.



- ▶ The reliability and validity of a measure is not established by any single study but by the pattern of results across multiple studies.
- ▶ The assessment of reliability and validity is an ongoing process.

For Further Reading I



RMDA, 2019

Appendix

For Further Reading

 Paul C. Price, Rajiv S. Jhangiani, and I-Chant A. Chiang.

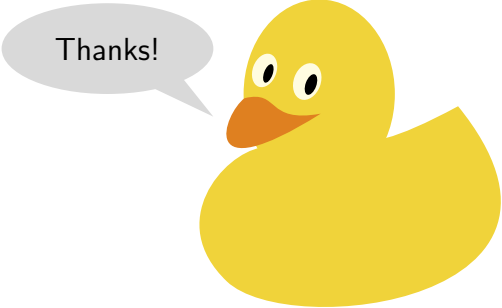
Research Methods in Psychology.

BCcampus OpenEd, 2015.

 J. T. Cacioppo and R.E. Petty.

The need for cognition.

Journal of Personality and Social Psychology, 42:
116–131, 1982.

A cartoon illustration of a yellow duck with a large orange beak and two simple black eyes. A grey speech bubble is positioned to the left of the duck's head, containing the text "Thanks!".

Thanks!