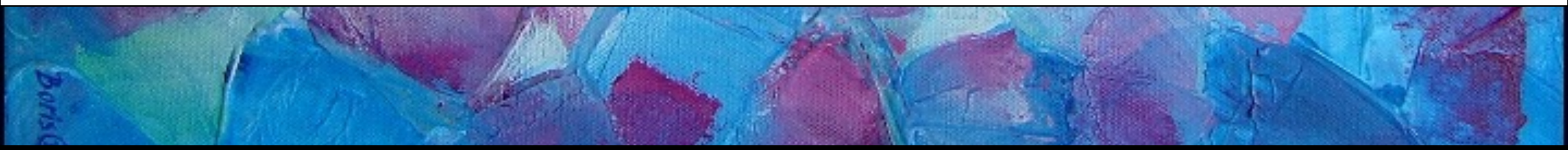
VI. RELIABILITY



★ IN THIS CHAPTER:

- Types of Reliability
- How to Assess Reliability
- Coder Discrepancy: Procedures and Solutions

RELIABILITY IS CONSISTENCY OR STABILITY IN MEASUREMENT.

Researchers cannot expect 100% reliability, but they should strive for the highest level possible. There are multiple types to consider.

★ TEST-RETEST RELIABILITY

When measuring a supposedly unchanging characteristic (e.g. an individual's personality), the test produces the **same results** each time it is performed.

★ INTER-OBSERVER RELIABILITY

Two researchers *observing* the same situation transform the sensed information into data in similar ways.

★ INTER-RATER RELIABILITY

Two coders *examining* the same material (e.g. ethnography) code variables from it similarly.

This type is the most relevant to our methods. In this section, we will discuss how to test for reliability between coders and how to deal with discrepancies in their coding.

HOW CAN WE TEST RELIABILITY BETWEEN CODERS?

THERE ARE TWO METHODS FOR ASSESSING INTER-RATER RELIABILITY IN CROSS-CULTURAL RESEARCH.

The first is to find the **percentage of agreement.** The researcher simply calculates the percentage of cases in which coders agreed for a particular variable. The closer the value is to 100%, the higher the inter-rater reliability for that variable. The major drawback of this method is that the *degree* of disagreement is not measured. Additionally, this method can only be used with a nominal or ordinal variable with a discrete number of scale ratings or scores.

WE CAN ALSO ASSESS INTER-RATER RELIABILITY BY CALCULATING A COEFFICIENT OF ASSOCIATION.

This method tests how closely coders' scores for the same variable predict one another. Coefficient of association scores can vary from -1 to +1. We want our score to be as close to +1 as possible, representing a near-perfect correlation. However, note that a systematic difference in coding (e.g. Coder A usually scores a case one point higher than Coder B) can still produce a high coefficient of association because the coders' scores still predict each other to a high degree.

To account for the degree of disagreement and systematic error, you should calculate both the percentage of agreement and coefficient of association for each coded variable.

HOW CAN CODER DISCREPANCIES BE RESOLVED?

4 COMMON METHODS FOR DISCREPANCY RESOLUTION

1. **Using only one coder's ratings.** This is most commonly employed when one coder rated all cases and a second coder rated a small, random subsample with a high degree of reliability.

2. **Summing or averaging codes**. The new score will represent the coders equally, and it should be closer to the "true" score. However, this method is not ideal if one coder seems to be more accurate than another. Discussing discrepancies can be a better option.

3. **Resolution method**. In this method, coders compare scores and jointly decide on a final resolution code. Discussing discrepancies provides an opportunity to ensure that coders are working with the same material and under the same assumptions. Resolution is especially important in pretesting because coding discrepancies can point to ambiguity in coding rules.

4. **Dropping serious disagreements**. Serious disagreements in coding may reflect ambiguous data. It often makes the most sense to simply drop such cases from your sample. However, if you use one of the other methods to decide upon values instead, you should include a reliability rating for each case. This will allow you to test the sample with and without the most problematic cases.

SUMMARY

- **Reliability** is consistency or stability in measurement.
- The main types of reliability are:
  - ★ Test-retest Reliability
  - ★ Inter-Observer Reliability
  - ★ Inter-Rater Reliability
- Two common methods for assessing reliability are:
  - ★ **Percentage of agreement** (only for nominal variables or an ordinal scale with a few points)
  - ★ A **coefficient of association** (more appropriate when there are many ordered scale positions)
- Because you can have a perfect coefficient if one coder is systematically higher than the other coder by a fixed amount, using both methods is often helpful.
- There are various strategies for resolving discrepancies between coders, each with advantages and disadvantages.

CITATION AND TERMS OF USE