Reliability Theory for Total Test Scores

Measurement Methods
Lecture 7
2/21/2006

Today's Class

- Reliability theory
- True score model
- Applications of the model

- Before we begin…do you have any questions about homework?

- Also, before we begin, allow me to hand out a document about the course project.

Reliability Theory

Basic Motivation

- Basic motivation for classical true-score theory is to provide a workable method for estimating the precision of measurement of a test score.
- Test score is the focus of this method.
- One of the first to develop in history.

Measurement By Analogy

- To begin our class, consider the process of measurement of a physical trait: length.
- We take our ruler/tape measure/whatever and use it to come up with the length of an object.
- If we wish to estimate the amount of error in our measurement, how would we proceed?

Multiple Measurements

- If we wish to estimate the error of measurement of the length of the object, we must take multiple measurements.
- The Mean of our measurements is the best estimate of the object's length.
- The Standard Deviation is the best estimate of the error in the measuring process.

Assumptions of Our Procedure

- To have the mean as our estimate of length and the SD as our estimate of measurement error there are several assumptions we must make…
- What do you think we are assuming?

Assumptions

- Replications are independent trials.
  - Making the errors for each trial independent of each other.
- The measurement instrument contains no source of constant error.
- The length of the object does not change over the time we take the measurements.

Transitioning to Psychological Measurement

- As soon as we move from our example to what we do in administering psychological tests, we see that our task is much more difficult.
- Replication, as a first example, becomes much more difficult.
  - Can you envision having to take our midterm 10 times just to get an estimate of the error in our measurement?

Problems with Replication

- There are further problems with repeated administrations of psychological tests to the same examinee(s).
  - Do the replications constitute independent trials?
    - Hence, do the results yield uncorrelated errors?
    - More exposures to the test will lead to "stereotyped" responses

Problems with Replication

- How much time should be allowed to lag between measurements?
- Are the psychological attributes constant over time?
  - Few psychological attributes are constant enough to be considered “traits.”
  - Maybe more appropriate to call many of these “psychological states.”
- As an example, consider mood.

Additional Concerns

- When it comes to length, it is fairly well understood that the number we arrive at based on a tape measure or ruler will represent the length of an object.

- When it comes to psychological attributes, the number we arrive at is not guaranteed to represent the attribute we intend to measure.

Psychological Attribute Measures

- To resolve the issue of what a score may represent psychologically, we consider three distinct (yet interrelated) concepts:

1. **Reliability** – the precision with which the test score measures the attribute.

2. **Validity** – the extent to which the test measures the attribute it was designed to measure.

3. **Generalizability** – The extent to which the composite test score generalizes beyond the specific items chosen to form the composite, to the domain of further indicators that might have been used.

Reliability – Where We Are Going

- You may feel that there is not a solution to the general problem: estimating the precision of measurement of a test score.

- We will postpone the conceptual issues of reliability by treating the classical true-score model as a piece of pure mathematics.

- In doing this, we will be able to illustrate the model.
  - Its assumptions.
  - When it can be applied.

The True-Score Model for Test Scores

Preliminaries

- Prior to introducing the true-score model, we introduce the following.

- Imagine we sample a single examinee:
  - At random.
  - From a population of interest.

- We administer a test of m items.
  - We form the total number right or number keyed score, Y.

Classical True-Score Model

- The classical true-score model hypothesizes that the total score consists of two components:
  - A portion representing the true score.
  - A portion representing the error of measurement.

- The model can be expressed mathematically:

$$Y = T + E$$

- Here T is the true-score.
- E is the error score.

Demonstration of True Score Model

- To show the true-score model, Table 5.1 (p. 64) lists the results of a simulation to demonstrate the sampling process for 10 examinees.
- These numbers were drawn from a distribution with a certain mean for T and E, and a certain variance for T and E.
- We will now simulate our own numbers for examinees to show how the process works.
  - We will be using R.
- For the moment we will work with Y (not Y').

Properties of *T* and *E*

1. *T* and *E* are measured on the scale of *Y*
   - They are bounded within the range of Y.
   - They have the same floor and ceiling.

2. *T* and *E* are uncorrelated.

$$\boxed{\rho_{TE} = 0}$$

   – In our example, this means that *E* is chosen independently of *T*.

Properties of *T* and *E*

3. The variance of *Y* is the sum of the variances of *T* and *E*.

$$\sigma_Y^2 = \sigma_T^2 + \sigma_E^2$$

This can be shown by the algebra of expectations:

Var(Y) = Var(T+E) = Var(T) + Var(E) + 2Cov(T,E) = Var(T) + Var(E)

Properties of *T* and *E*

4. Variances of *T* and *E* are both less than and at most equal to the variance of *Y*.

$$\sigma_T^2 \leq \sigma_Y^2 \quad \textit{and} \quad \sigma_E^2 \leq \sigma_Y^2$$

Properties of *T* and *E*

5. The ratio of the variance of *T* to the variance of *Y*,

$$\rho_r = \frac{\sigma_T^2}{\sigma_Y^2} = \frac{\sigma_T^2}{\left(\sigma_T^2 + \sigma_E^2\right)}$$

- This term is bounded by zero and one.
- By definition this is called the **reliaility coefficient** of *Y*.

Further Information

- The properties of the classical true-score model by themselves are relatively uninformative.

- To further expand our example, consider if we get, from our same sample of examinees, a second total test score, called *Y'*.

The Second Score

- You can envision generating the second score for each examinee by having the same T for each person.
- The error score, E', however, would be independently drawn (but it would have the same error variance, σ^2_E.
- The classical true-score theory formulation for the new total test score would then be:

$$Y' = T + E'$$

Simulated Data

- Using our former parameters, we can simulate the data for Y', using a process similar to that used for Y.
  - We use the same T for each simulated examinee.
  - We draw a new E for each examinee.

More Properties

- For each examinee, *Y* and *Y'* have the same randomly drawn *T* value.

- Each has an independently drawn *E* and *E'* value.

- By construction, *E* and *E'* are uncorrelated with:
  - *T*.
  - Each other.

Independent Error Implications

- Because of the independence of error terms, we get the following result:

$$\rho_{TE} = \rho_{TE'} = \rho_{EE'} = 0$$

- The correlation between each of these elements is zero.

More Implications

- Another property we set was for the variances of *E* and *E'* to be equal:

$$\sigma^2_{E'} = \sigma^2_E$$

- What follows from this result is that the variance of *Y* is equal to that of the variance of *Y'*:

$$\sigma^2_{Y'} = \sigma^2_Y$$

Variance Formation Practice

- Show that $\sigma^2_{Y'} = \sigma^2_Y$

More Properties

- A further property of the two tests now follows:

$$\rho_{YY'} = \rho_r = \frac{\sigma_T^2}{\sigma_Y^2}$$

- Note that $\rho_{YY'}$ is the correlation between Y and Y'.

- This hold important consequences.
  - ρ_r can now be computed from observations.
  - Reliability can be estimated from finite samples.

How Does That Happen?

- It is not usually the case that a variance ratio will equal a correlation.

- In our case, it is easy to show why:

$$\mathrm{Cov}(Y, Y') = \mathrm{Cov}[(T+E),(T+E')]$$

$$= \sigma_{TT} + \sigma_{TE} + \sigma_{TE'} + \sigma_{EE'}$$

$$= \sigma_T^2$$

More Connections

- We note that:

$$\rho_r = \rho^2_{YT} = \rho^2_{Y'T}$$

- The reliability coefficient is the square of the correlation between *Y* and *T* or *Y'* and *T*.

Wrapping Up

- Today, we scratched the surface of concepts about reliability.
- To do so, we used classical true-score theory.
- We will build upon these concepts next time

Next Time

- More of Chapter 5 – Reliability Theory for Total Test Scores.