Reliability Theory for Total Test Scores

Measurement Methods

Lecture 8

Today's Class

- More reliability theory with the true score model
- Some applications of the model (to be continued next week).

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

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.

$$\rho_{TE} = 0$$

   – 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*.

$$\boxed{\sigma_Y^2 = \sigma_T^2 + \sigma_E^2}$$

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*.

$$\sigma_T^2 \leq \sigma_Y^2 \quad and \quad \sigma_E^2 \leq \sigma_Y^2$$

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*.

Notes About ρ_r

- The subscript *r* is preferred because:
  - ρ_r is the *r*eliability coefficient.
  - ρ_r is a *r*atio of variances.

- You will also find ρ_r denoted as $\rho_{YY'}$

Back to Our Simulation

- Although we sampled values for T, E, and E', in reality we would not be able to observe these values.
  - We can only observe Y and Y'.
  - That's three "unknowns" and two "knowns."

- We can estimate ρ_r from the correlation between Y and Y'

Our Simulation Values

- In generating the data we set the following values:
  - $\sigma_E^2 = 33$
  - $\sigma_{E'}^2 = 33$
  - $\sigma_T^2 = 825$

- So… $\sigma_Y^2 = 825 + 33 = 858$

- And… $\rho_r = 825/858 = 0.9615$

Working Toward a Measure of Error

- ρ_r is really more of a means to an end rather than a final destination.

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

$$\sigma_E^2 = \sigma_Y^2\left(1 - \rho_{YY'}\right)$$

- Really, we would like to estimate the variance of E.
  - Recall that E happens to be on the same scale as Y
  - We can re-express something we have seen before to get at σ_E^2:

Test Error Variance

- The variance of the test error can be estimated from samples using Y and Y'.

- To show how this would work from our simulation example:
  - $\sigma_E^2 = 858(1-0.9615) = 33.0$

Standard Error of Measurement

- We were working with test error variance previously, which has a non-standard unit metric (remember variance is the units squared).

- Instead, we consider a measure of error that is in the same unit – the standard error of measurement.
  - This is simply the square root of the test error variance.

Standard Error of Measurement

- For the population values (which we know in our example):

$$SEM(Y) = \sqrt{\sigma_E^2} = \sqrt{33} = 5.74$$

- For the sample, we just plug in our sample estimate of the error variance:

$$SEM(Y) = \sqrt{s_E^2}$$

Interpretations

- In interpreting the classical true-score model:
  - We think of T as being characteristic of the examinee.
  - We think of E as being characteristic of the test.

- In practice, we can expect that the reliability coefficient will vary according to the population sampled.
  - Under reasonable assumptions the variance of E will remain approximately invariant, however.

Applications of the True-Score Model

Applications of the Model

- To apply the true-score model, we wish to interpret T as the true of the examinee and E as the error of measurement.
  - We assume that a person's true score does not change between measurements.

- To the extent that factors outside of the test enter into the equation, the test-retest methods yield a valid coefficient of precision.

Methods Used to Estimate Reliability

- There are three main methods used to estimate the reliability coefficient:

    1. Test-retest methods.
    2. Parallel or alternate-form methods.
    3. Internal analysis.

- We will revisit #3 in the next Chapter.

Test-Retest Methods

- A test of *m* items is administered to a large sample of examinees at two points in time
  - Yielding Y and Y'.

- We estimate $\rho_{YY'}$ from the test scores.
  - We then call this entity ρ_r
  - We can then estimate σ_E^2 and SEM(Y).

Assumptions

- In doing this whole process, we are making two very strong assumptions:
  1. The true scores of the examinees do not change between administrations of the test.
     - If this is the case, the errors are independent.
     - $\rho_{YY'}$ can be taken to be an approximation of the coefficient of precision.
- One big problem with this assumption is that there is no way to tell if it is violated.

Assumptions, Continued

- Our second assumption under the retest methods:
  2. We must say that the retest true score is defined as the component of the observed score that does not change between administrations.

- This is the reason that the reliability coefficient is sometimes called the coefficient of stability.

Wrapping Up

- Classical true score theory sets some strict assumptions about tests.
- Some assumptions are not possible to test.
- There are various properties associated with the true-score model.
- We will build upon these concepts next time

Next Time

- More applications of the true-score model.
  - Last section of Chapter 5.