Test Homogeneity – The Single-Factor Model

Measurement Methods

Lecture 10

Today's Class

- Test Homogeneity.

- The Single Factor Model.
  - AKA the Spearman model.

Initial Thoughts

- Today we will be discussing how we come to understand if tests are homogeneous.
- Primary to our purpose will be the covariance (or correlation) matrix of associations between items.
- We will use the covariance/correlation matrix to test the homogeneity of a test with the Spearman single-factor model.

If Two Variables are Correlated…

- If two variables are correlated, there are at least three ways we can "explain" the presence of the relationship between them:

1. One variable (partly) causes the other.

2. Both variables are related effects of a common cause.

3. Both variables are correlated because they measure, or indicate, something in common.

Example of #1

1. One variable (partly) causes the other.

Observing a rat in an activity cage, we say that hunger causes activity.

After a good workout, we say that activity causes hunger.

We do not say that hunger and activity are merely contingently associated.

Both the meaning and the verification of causal claims are deep and controversial matters.

Example of #2

2. Both variables are related effects of a common cause.

   Distinct stock prices vary together from the impact of political events on the psyches of market players

Example of #3

3. Both variables are correlated because they measure, or indicate, something in common.

   Some tests contain items that can be scored for more than one trait - this correlation would generally be regarded as spurious.

Measurement of Something In Common

- By measuring something common, nothing quite so literal is meant.
- The notion is that variables are indicators of the attribute.
  - You can think of indictors as being symptoms or manifestations.
- For example – take extraversion.
  - An abstract concept whose instances are the recognized extravert behaviors.
    - It is circular to say extraversion “causes” its manifestations.

Where We Are Headed

- Really, the goal of this chapter is to look at #3.
  - Correlated items measuring something in common.
- We are going to introduce a statistical model to refine our conception of a homogeneous test.
  - A test where all the items measure the same thing.
- We will see that the true-score model from the past chapter is a special case of the new model.

Test Homogeneity

Preliminary Distinctions

- In this chapter we do not need to distinguish between a test composed of *m* items and a test composed of *m* subtests.
  - Subtests can include item bundles, testlets, test batteries, etc…

- All results in this chapter apply to whatever measurements we regard as the basic sets of scores to be combined into a global test score.

Preliminary Information

- Suppose we have a test with *m* items.
  - $j=1,\ldots,m$
- The items have scores $X_1, X_2, \ldots, X_m$.
- In our population of interest, the items have pairwise covariances σ_{jk}.
  - We wish to give a statistical model to the idea that the pairs of covariances are non-zero because the items measure ***<u>one</u>*** attribute in common.

Extending the True Score Model

- The phrase "measure something in common" requires an extension of the true score model.

- Recall the true score model with:

  $Y = T + E$ $\qquad\qquad$ $Y' = T + E'$

- Here T represents the attribute common to the test forms.
  - T is assumed to be measured equally well by either form.
  - E and E' are due to unique or idiosyncratic properties of the particular items in the separate forms.

Item Level Model

- Really, what we need to create is a model that will work at the level of the items.
  - The true-score model only works at the level of the test score.
- If we applied the true-score model at the level of the items, we would get:

$$X_j = T + E_j$$

- We will see that the above model is a special case of the to-be-introduced single factor model.

What is Wrong with This Model?

- We knew that the classical true-score model had problems, but what's so bad about

$$X_j = T + E_j?$$

- Really, three things jump out at me with a model like this:

1. Nothing to differentiate difficulties of items.
2. Every item measures T equally well.
3. Nothing to differentiate unique error variances of items.

Can This Model Be Improved?

- We quickly learned that our "new" item-level model doesn't do a whole lot for us in general.

- We could improve this model by thinking about what we are measuring in common.
  - Lets let *F* represent the attribute we are trying to measure.
  - Each examinee will have a certain score on *F*.

- Taking it a step further, we say that *F* is the ***common factor*** that ties items together – the common thing measured by a homogeneous test.
  - Can we create a model that would relate an examinee's score on *F* to their performance on an item?

The Spearman Single Factor Model

- In relating an examinee's level of the common factor to their performance on an item, we introduce the Spearman single factor model:

$$X_j = \mu_j + \lambda_j F + E_j$$

- X_j is the score for an examinee on the j^{th} item.

- F is the examinee's measure of the common attribute.

The Spearman Single Factor Model, Continued

$$X_j = \mu_j + \lambda_j F + E_j$$

- E_j is the examinee's measure of the unique or idiosyncratic property of item *j*.
  - The amount by which item *j* is shifted.
- μ_j is the overall mean for an item.
  - Allowing for differing item difficulties.
- λ_j is called the factor loading of item *j*.
  - We will come to discuss this factor loading quite a bit.
  - This is where the "magic" happens.

Factor Loadings

- The term factor loading has a long history in Psychology.
- It is the extent to which the item is “loaded” onto the Factor.
  - Some items load more highly on to the factor than other.
- The factor loadings of items reveal much about a test’s structure.

More on Factor Loadings

- The factor loading is similar to a regression weight:
  - It represents the amount of change in the item per one-unit increase in the factor score.

- It measures how sensitively each item functions as an indicator of the common factor F.
  - Items with relatively large λ_j are better indicators of F than items with relatively small λ_j.

- The factor loading is a measure of the discriminating power of the item.
  - How well the item discriminates between examinees with low and high values of F.

The Factor Model and Test Homogeneity

- We can use the single factor model to test whether our items measure a single trait homogeneously.

- If the model appears to "fit" the data adequately, then we can conclude we have a sufficiently homogeneous test.

- If the model does not "fit" the data, then our test is not homogeneous in measuring F.

Single Factor Model Specifics

- We need to define a few more things about our factor model:
  - The unique component, E_j, is independent of the common factor, F.
    - Remember independence means that $\text{Cov}(E_j,F) = \text{Corr}(E_j,F) = 0$
  - The unique components of any two items j and k are independent:
    - $\text{Cov}(E_j,E_k) = \text{Corr}(E_j,E_k) = 0$
  - The mean for the unique component is zero.

- We also have to set the scale for F.
  - We must pick it’s mean and variance.
  - For most of our purposes, it serves us well to think of F as being a standardized measure.
    - Mean of zero.
    - Standard Deviation/Variance of one.

What Does The Common Factor Model Say About Our Items?

- So, what can we say the model predicts about our items, marginally?

- What is the model-predicted item mean?

- What is the model-predicted item variance?

Model Predicted Item Mean

- The mean for an item under the single factor model can be found by the algebra of expectations:

$$
\begin{aligned}
\mathrm{E}(\mathrm{X}_\mathrm{j}) &= \mathrm{E}(\mu_j + \lambda_j F + E_j) \\
&= \mathrm{E}(\mu_j) + \mathrm{E}(\lambda_j F) + \mathrm{E}(E_j) \\
&= \mu_j + \lambda_j \mathrm{E}(F) + \mathrm{E}(E_j) \\
&= \mu_j + \lambda_j * 0 + 0 \\
&= \mu_j
\end{aligned}
$$

Item Mean = Nada

- We note that our model says that our item mean should be our item mean parameter.
- Generally, we are not concerned with such a quantity because it tells us information only marginally.
  - No information about how the item measures the common factor.

Model Predicted Item Variance

- The variance for an item under the single factor model can be found by the algebra of expectations:

$$
\begin{aligned}
\mathrm{Var}(X_j) &= \mathrm{Var}(\mu_j + \lambda_j F + E_j) \\
&= \mathrm{Var}(\lambda_j F + E_j) \\
&= \mathrm{Var}(\lambda_j F) + \mathrm{Var}(E_j) + 2\,\mathrm{Cov}(F, E_j) \\
&= \lambda_j^2\,\mathrm{Var}(F) + \mathrm{Var}(E_j) \\
&= \lambda_j^2 + \Psi_j^2
\end{aligned}
$$

We Typically Set this to One

Is zero by independence

We define the variance of E to be the unique variance of the item.

Model Predicted Item Pair Covariances

- The covariance for a pair of items under the single factor model can be found by the algebra of expectations:

$$
\begin{aligned}
\mathrm{Cov}(\mathrm{X_j}, \mathrm{X_k}) &= \mathrm{Cov}(\mu_j + \lambda_j F + E_j, \mu_k + \lambda_k F + E_k) \\
&= \mathrm{Cov}(\lambda_j F + E_j, \lambda_k F + E_k) \\
&= \mathrm{Cov}(\lambda_j F, \lambda_k F) + \cancel{\mathrm{Cov}(\lambda_j F, E_k)} + \cancel{\mathrm{Cov}(\lambda_k F, E_j)} \\
&\qquad + \cancel{\mathrm{Cov}(E_j, E_k)} \\
&= \lambda_j \lambda_k \mathrm{Cov}(F, F) \\
&= \lambda_j \lambda_k
\end{aligned}
$$

The covariance of a variable with itself is its variance.

The variance of *F* is set to one.

Extrapolating to the Covariance Matrix

- Ok…we have seen what the model predicts our variance to be for each item.
- We have seen what the model predicts our covariance to be for each pair of items.
- So how does our model-predicted covariance matrix look?

$$\mathbf{\Sigma} = \begin{bmatrix} \lambda_1^2 + \psi_1^2 & \lambda_1 \lambda_2 & \lambda_1 \lambda_3 & \lambda_1 \lambda_4 & \lambda_1 \lambda_5 \\ \lambda_1 \lambda_2 & \lambda_2^2 + \psi_2^2 & \lambda_2 \lambda_3 & \lambda_2 \lambda_4 & \lambda_2 \lambda_5 \\ \lambda_1 \lambda_3 & \lambda_2 \lambda_3 & \lambda_3^2 + \psi_3^2 & \lambda_3 \lambda_4 & \lambda_3 \lambda_5 \\ \lambda_1 \lambda_4 & \lambda_2 \lambda_4 & \lambda_3 \lambda_4 & \lambda_4^2 + \psi_4^2 & \lambda_4 \lambda_5 \\ \lambda_1 \lambda_5 & \lambda_2 \lambda_5 & \lambda_3 \lambda_5 & \lambda_4 \lambda_5 & \lambda_5^2 + \psi_5^2 \end{bmatrix}$$

Model Parameters

- So, for our single factor model, we have a set of parameters we need to estimate:
  - m factor loadings – $\lambda_1, \lambda_2, \ldots, \lambda_m$
  - m unique variances – $\Psi_1^2, \Psi_2^2, \ldots, \Psi_m^2$

- Note that item means are not necessary to be estimated.
  - We have these already.

- How do we go about getting estimates of these parameters?

- How do we estimate things in general?

Model Estimation

- The key in model estimation is to find a set of parameters that minimize the discrepancy between the observed covariance matrix and the model-predicted covariance matrix.

$$\mathbf{\Sigma} = \begin{bmatrix} \lambda_1^2+\psi_1^2 & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4 & \lambda_1\lambda_5 \\ \lambda_1\lambda_2 & \lambda_2^2+\psi_2^2 & \lambda_2\lambda_3 & \lambda_2\lambda_4 & \lambda_2\lambda_5 \\ \lambda_1\lambda_3 & \lambda_2\lambda_3 & \lambda_3^2+\psi_3^2 & \lambda_3\lambda_4 & \lambda_3\lambda_5 \\ \lambda_1\lambda_4 & \lambda_2\lambda_4 & \lambda_3\lambda_4 & \lambda_4^2+\psi_4^2 & \lambda_4\lambda_5 \\ \lambda_1\lambda_5 & \lambda_2\lambda_5 & \lambda_3\lambda_5 & \lambda_4\lambda_5 & \lambda_5^2+\psi_5^2 \end{bmatrix}$$

- To demonstrate, lets look at an example from the SWLS.

- We essentially have to search for values of our model parameters to minimize the distance from Σ to S.

$$\boldsymbol{S} = \begin{bmatrix} 2.566 & 1.560 & 1.487 & 1.195 & 1.425 \\ 1.560 & 2.493 & 1.283 & 0.845 & 1.313 \\ 1.487 & 1.283 & 2.462 & 1.127 & 1.313 \\ 1.195 & 0.845 & 1.127 & 2.769 & 1.323 \\ 1.425 & 1.313 & 1.313 & 1.323 & 3.356 \end{bmatrix}$$

Model Estimation

- We could try out numerous values for our model parameters.
- For instance consider the following:

λ	Ψ
1.0	0.5
1.0	0.5
1.0	0.5
1.0	0.5
1.0	0.5

- These parameters give us a model predicted covariance matrix of:

$$\hat{\Sigma} = \begin{bmatrix} 1.5 & 1.0 & 1.0 & 1.0 & 1.0 \\ 1.0 & 1.5 & 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.5 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.5 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.5 \end{bmatrix}$$

Realistically, Though…

- In reality, a computer will search through the parameter space and find a set that will minimize the discrepancy between Σ and S.

- There are many ways of defining a discrepancy, though.

- Our book introduces a very simple one: the unweighted least squares (ULS) function.

$$q_u = \frac{1}{m^2} \sum_j \sum_k (s_{jk} - \sigma_{jk})^2$$

Example Results

Model Estimates

Model Predicted Covariance Matrix

Discrepancies: From S - Σ

TABLE 6.4
Satisfaction With Life Scale—Spearman Analysis

	(a)		(b)				
	λ	ψ^2	1	2	3	4	5
1	1.290	0.901	2.565	1.424	1.481	1.328	1.529
2	1.104	1.274	1.424	2.493	1.267	1.051	1.308
3	1.148	1.144	1.481	1.267	2.462	1.093	1.360
4	0.952	1.863	1.328	1.051	1.093	2.769	1.128
5	1.185	1.951	1.529	1.308	1.360	1.128	3.355

(c)

	1	2	3	4	5
1	.0	.135	.006	−.033	−.104
2	.135	.0	.015	−.206	.004
3	.006	.015	.0	.035	−.048
4	−.033	−.206	.035	.0	.195
5	−.104	.004	−.048	.195	.0

- To provide an example, consider the following estimates provided in our textbook (we will learn how to get estimates next week).

Great, But Does The Model "Fit?"

- Once you get model parameter estimates, you must check to see if the model "fits."
  - Measures of model fit are plentiful.
  - For now, we will consider one measure.

- Basically, we want to say a model "fits" when the discrepancy between the observed and model-predicted covariance matrix is small.

- So, we need a function that will quantify the size of the model discrepancy.

The Goodness-of-Fit Index

- Introducing the GFI – an index based on our ULS discrepancy function:

$$q_u = \frac{1}{m^2} \sum_j \sum_k \left(s_{jk} - \sigma_{jk}\right)^2$$

- The GFI is then:

$$GFI = 1 - \frac{q_u}{c}$$

- Where:

$$c = \frac{1}{m^2} \sum_j \sum_k s_{jk}^2$$

Properties of the GFI

- The GFI weights the size of the discrepancies by the size of the covariances in the matrix.
  - That is why we divide by c.
  - Actually, q_u is the average discrepancy.

- If the model fits perfectly, the GFI will be 1.0.

- The GFI will approach 0.0 as the fit gets worse.

- The computer package you use to estimate your model will tell you the goodness of fit.

What is the GFI of the SWLS?

- The GFI of the SWLS was 0.9968.

- From practical experience, we come to learn this is very good fit.
  - Rod says that 0.90 is acceptable and the fit is “good” when the GFI is above 0.95.

- It is also good practice to look at the discrepancy matrix to see the fit.
  - This will tell you a better story of why a model does not fit.
  - It will also tell you if some pairs of items are not well predicted by the model.

Back To Test Homogeneity

- By applying the single factor model to the SWLS data, we were in essence trying to determine if the test was homogeneous.
  - If the model fit, we say the test measured a single factor – and was homogeneous.
  - If the model didn't fit, we say the test measured more than a single factor – and was not homogeneous.

- By our standards, we would say our model fit, and that the SWLS was a homogeneous test.

Wrapping Up

- Today we introduced the single factor model in all its glory.
  - It provides a more detailed way of looking at test items.
  - We will show how we can test the assumptions of the true-score model with the single-factor model.
  - We will also show how we can estimate reliability from the single-factor model.

Next Time

- Testing the true-score model with the single-factor model.

- Test reliability.
  - Chapter 6 – Part 2.

- How to fit the factor model in SAS.