LECTURE NOTES #13: Reliability, Structural Equation Modeling, and Hierarchical Modeling

1. Introduction

These lecture notes present some basic intuitions underlying structural equations modeling (SEM). If you find this technique useful in your research, I suggest that you take a semester-long course on SEM.

2. Elementary Covariance Algebra

It will be useful to learn how to work with expectations and variances. I will now introduce some rules for how to manipulate expectation, variance and covariance. These rules will be useful as we develop intuition for structural equations modeling.

In the following, E refers to expectation, var refers to variance, and cov refers to covariance.

(a) If $a$ is some constant number, then

$$E(a) = a$$

“The average of a constant is a constant.” Think about a column in a data matrix. If all the numbers in that column are identical, then the mean will be the value of that number.

(b) If $a$ is some constant value real number and $X$ is a random variable with expectation $E(X)$, then

$$E(aX) = aE(X)$$

Multiplying by a constant then averaging is the same as averaging then multiplying by the constant.

(c) If $a$ is a constant real number and $X$ is a random variable, then

$$E(X + a) = E(X) + a$$
Adding a constant then averaging is the same as averaging then adding a constant.

(d) If $X$ is a random variable with expectation $E(X)$, and $Y$ is a random variable with expectation $E(Y)$, then

$$E(X + Y) = E(X) + E(Y)$$

The sum of two averages equals the average of two sums.

(e) Given some finite number of random variables, the expectation of the sum of those variables is the sum of their individual expectations. Thus,

$$E(X + Y + Z) = E(X) + E(Y) + E(Z)$$

(f) If $a$ is some constant real number, and if $X$ is a random variable with expectation $E(X)$ and variance $\sigma^2$, then the random variable $(X + a)$ has variance $\sigma^2$. In symbols,

$$\text{var}(X) = \text{var}(X + a)$$

(g) If $a$ is some constant real number, and if $X$ is a random variable with variance $\sigma^2$, the variance of the random variable $aX$ is

$$\text{var}(aX) = a^2 \sigma^2$$

(h) If $X$ and $Y$ are independent random variables, with variances $\sigma_X^2$ and $\sigma_Y^2$ respectively, then the variance of the sum $X + Y$ is

$$\sigma_{(X+Y)}^2 = \sigma_X^2 + \sigma_Y^2$$

Similarly, the variance of $X - Y$ when both are independent is

$$\sigma_{(X-Y)}^2 = \sigma_X^2 + \sigma_Y^2$$

(i) Given random variable $X$ with expectation $E(X)$ and the random variable $Y$ with expectation $E(Y)$, then if $X$ and $Y$ are independent,

$$E(XY) = E(X)E(Y)$$

The implication does not go in the other direction. If $E(XY) = E(X)E(Y)$, that does not imply that the two variables are independent.
(j) If $E(XY) \neq E(X)E(Y)$, the variables $X$ and $Y$ are not independent.

This statement is simply the contrapositive of the previous rule.

(k) Given the random variable $X$ with expectation $E(X)$ and the random variable $Y$ with expectation $E(Y)$, then the covariance of $X$ and $Y$ is
\[ \text{cov}(X,Y) = E(XY) - E(X)E(Y) \]

(l) Another definition of covariance is
\[ \text{cov}(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] \]

(m) If $a$ is a constant and $X$ is a random variable, then the covariance is
\[ \text{cov}(X,a) = 0 \]

“The covariance of a random variable with a constant is zero.”

(n) If $a$ is a constant and the variables $X$ and $Y$ are random, then
\[ \text{cov}(aX,Y) = a\text{cov}(X,Y) \]

(o) The variance of a random variable $X$ is equal to the covariance of $X$ with itself, i.e.,
\[ \text{cov}(X,X) = \text{var}(X) \]

(p) The sum operation between random variables is distributive for covariances, i.e.,
\[ \text{cov}(X,Y + Z) = \text{cov}(X,Y) + \text{cov}(X,Z) \]

(q) Now that we know about covariances, we can return to the definition of the sum of two random variables. I will now relax the restriction of independence. If $X$ and $Y$ are random variables, with variances $\sigma^2_X$ and $\sigma^2_Y$ respectively, then the variance of the sum $X + Y$ is
\[ \sigma^2_{(X+Y)} = \sigma^2_X + \sigma^2_Y + 2\text{Cov}(X,Y) \]

Similarly, the variance of $X - Y$ is
\[ \sigma^2_{(X-Y)} = \sigma^2_X + \sigma^2_Y - 2\text{Cov}(X,Y) \]
(r) Definition of the correlation: Given two random variables $X$ and $Y$, then the covariance divided by the standard deviation of each variable is the correlation $\rho$,

$$
\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}
$$

Note that the correlation is simply a normalized covariance.

The following results are useful for interactions:

(s) The expectation of the product of two variables is

$$
E(XY) = C(X, Y) + E(X)E(Y)
$$

(t) The variance of the product of two variables that are normally distributed is (using C for cov and V for var)

$$
V(XY) = E(X)^2V(Y) + E(Y)^2V(X) + 2E(X)E(Y)C(X, Y) + V(X)V(Y) + C(X, Y)^2
$$

(u) The covariance of two products with all variables distributed multivariate normal is

$$
$$

3. Covariance Matrices

Suppose you have three variables. There will be three variances (one for each variable) and three covariances (one for each possible pairwise combination). Making sense of these six numbers is facilitated if they are arranged in a matrix such as
Because $\text{cov}(x,y) = \text{cov}(y,x)$ (i.e., covariances are symmetric) the lower triangle of the above matrix mirrors the upper triangle\(^1\). This property is called symmetry and a matrix that has this property is called symmetric.

### 4. Correlation Matrices

Because a correlation is defined as

\[
\text{cor}(x,y) = \frac{\text{cov}(x,y)}{\sqrt{\text{var}(x)\text{var}(y)}}
\]

(13-1)

all information necessary to compute a correlation matrix is present in the covariance matrix.

The elements along the diagonal of a correlation matrix will equal one. The correlation matrix is also symmetric. Note that while it is easy to convert a covariance matrix into a correlation matrix, the conversion will not be possible going the other way (correlation matrix to covariance matrix) unless the variances of each variable are known.

A convenient way to represent both the covariance matrix and the correlation matrix is to display one matrix that has variances in the diagonal, covariances in the upper triangle, and correlations in the lower triangle. That is,

Because covariances are symmetric, the lower triangle mirrors the upper triangle. This property is called symmetry and a matrix that has this property is called symmetric.


\(^1\)Triangles are formed on either side of the main diagonal (which starts at the upper left and moves to the lower right).
problems are easier to deal with in the form of a covariance matrix and other problems are easier to deal with in the form of a correlation matrix.

5. Making a connection between a correlation matrix and regression

It will be instructive to see how a correlation matrix enters in techniques you already know. Here we will show the connection between a correlation matrix and regression. The “beta coefficients” for each variable in a regression can be computed from the correlation matrix of the variables involved. The “beta coefficients” are the regression weights when all variables (predictors and criterion) are standardized. In such a model there is no intercept. The following discussion is adapted from Edwards (1985).

Let $R$ be the correlation matrix of only the predictors, $b$ be the vector of “beta coefficients”, and $r$ is the vector of correlations between each predictor and the criterion. Multiple regression satisfies the following equation:

$$Rb = r \quad (13-2)$$

How one multiplies matrices is not important for our purposes. All you need to know now is that by “multiplying” $R$ and $b$ we can compute $r$, the correlation of the criterion with each predictor. But, usually we know $r$ and $R$ and not $b$; the task is to estimate $b$. So, think back to junior high school. The unknown is $b$ so “divide” both sides of the equation by $R$ to solve for $b$. Division is computationally a little more difficult for matrices, but the intuition is the same.

Back to the regression problem. Look at Equation 13-2—you’ll see that all we need to do is multiply both sides by the inverse of $R$.

$$Rb = r \quad (13-3)$$
$$R^{-1}Rb = R^{-1}r \quad (13-4)$$
$$b = R^{-1}r \quad (13-5)$$

This shows at all we need to compute the standardized beta coefficients from the regression are the correlations between predictors (matrix $R$) and the correlations between each predictor and the dependent variable (vector $r$)

Further, with knowledge of the variances we can transform the “standardized beta coefficients” computed through this method to “raw score regression coefficients”, i.e., the regression coefficients that result when the variables are not standardized. All one needs to do is multiply the “beta coefficients” by the ratio of the standard deviation of the criterion and the standard deviation of the predictor associated with the slope. That is,

$$\text{raw score reg coef for } x_i = \text{“beta” for } x_i \cdot \frac{sd(y)}{sd(x_i)} \quad (13-6)$$

---

$^2$For our purposes a vector can be treated as a matrix with one column—these vectors are, not surprisingly, called column vectors.
This section showed that the variance and correlation matrices contain all the information necessary to compute slopes in a regression. This demonstrates that these constructs are fundamental to statistics.

We now make use of these concepts to develop some intuition to a set of analyses including reliability, mediation and systems of linear equations (aka structural equation modeling).

6. Reliability

What is reliability? The standard definition uses the simple additive model

\[ X = T + \epsilon \] (13-7)

where \( X \) is the observed score, \( T \) is the true score, and \( \epsilon \) is measurement error. Think of a bathroom scale—you have a true weight \( T \) but you observe a reading \( X \) that includes measurement error from the scale.

The concept of reliability is based on variances. The reliability of variable \( X \) is defined as

\[
\text{reliability of } X = \frac{\text{var}(T)}{\text{var}(T) + \text{var}(\epsilon)}
\] (13-8)

\[
= \frac{\text{var}(T)}{\text{var}(X)}
\] (13-9)

In words, reliability is defined as a proportion of observed variance that is true variance. Reliability is interpreted as a proportion—reliability cannot be negative.

But in practice we don’t know the true score \( T \) (nor the error \( \epsilon \)) so what can we do? How do we separate the true score from the observed score? One approach is a test-retest paradigm\(^3\).

Test the person on the same variable twice. Assuming no changes over time in the true score\(^4\), time 1 is influenced by the same “latent variable” as time 2. The model assumes that the only difference between the two time periods is the error, which I’ll denote \( \epsilon_1 \) and \( \epsilon_2 \).

So, we have \( X_1 = T + \epsilon_1 \) and \( X_2 = T + \epsilon_2 \). We could get a handle on the true score \( T \) by examining the covariance between the two observed variables \( X_1 \) and \( X_2 \). The true component \( T \) is what both \( X_1 \) and \( X_2 \) have in common. We need to assume that the two errors are independent from \( T \) and are not correlated with each other (you will see how that assumption simplifies a complicated expression).

\(^3\)There are many other approaches that I will not cover here.

\(^4\)This is an important assumption. There cannot be change in the intervening time period, so if you expect “growth” or change to occur between the first and second administration, then this concept of reliability is not useful.
cov(X₁, X₂) = cov(T + ɛ₁, T + ɛ₂)
= cov(T, T) + cov(T, ɛ₁) + cov(T, ɛ₂) + cov(ɛ₁, ɛ₂)
= cov(T, T) under independence
= var(T)

Recall that the correlation between two variables is defined as the covariance divided by the product of the standard deviations.

\[
\text{cor} = \frac{\text{cov}(X, Y)}{\text{sd}(X)\text{sd}(Y)} \quad (13-10)
\]

To compute the correlation of X₁ and X₂ we just need the covariance (computed above) and the standard deviations.

\[
\text{cor}(X₁, X₂) = \frac{\text{cov}(X₁, X₂)}{\sqrt{\text{var}(X₁)\text{var}(X₂)}}
\]
\[
= \frac{\text{var}(T)}{\sqrt{\text{var}(T + ɛ₁)\text{var}(T + ɛ₂)}}
\]
\[
= \frac{\text{var}(T)}{\text{var}(T + ɛ)} \quad \text{assume var of ɛs are equal & ind}
\]
\[
= \text{reliability}
\]

Thus, a correlation of a test-retest variable is equivalent to reliability. For this reason, test-retest reliability is usually denoted rₓₓ. This is a case where a raw correlation can be interpreted as a proportion of true variance to observed variance—there is no need to square a correlation in a test-retest situation when it is interpreted as a reliability. In fact, taking the square root of the test-retest reliability leads to an estimate of the correlation between true score and observed score.

This same logic extends to the true correlation between two variables, each measured with their own error. For example, suppose variables X and Y are each measured with error. The correlation between the “true” component of X and the “true” component of Y is equal to:

\[
r_{XₜYₜ} = \frac{r_{XY}}{\sqrt{r_{XX}r_{YY}}} \quad (13-11)
\]

This equation is very important in test theory. It shows that a true correlation is equal to the observed correlation (the numerator in the right hand side) divided by the product of the
A few words on latent variables. Latent variables seem to rub people the wrong way at first. It seems we should only base empirical science on what we can directly observe. Actually, there are many examples in science of latent variables (things we don’t directly observe). One famous example is the black hole in astronomy. A black hole cannot be observed directly, but the presence of a black hole can be observed indirectly, such as the pattern light makes when it travels near a black hole. This serves as an analogy for how latent variables are used in structural equation modeling. One has a set of observations, or indicators, and they jointly determine the unobserved, latent variable. The general point is that “latent variables” are sometimes used in science; they are not something limited to social science or statistics or, more specifically, to structural equations modeling.

7. Spearman-Brown (SB) Formula

For a complete derivation of the SB formula see Nunnally’s *Psychometric Theory*. 

The logic of the SB formula is to estimate the reliability of a $k$ item test from knowledge of the reliability of individual items. For example, suppose I have a 5 item questionnaire and I know the intercorrelations of each item with the other items. However, I do not know (without
further computation) how a person’s score on this 5 item test (i.e., the sum over the 5 items) would correlate with their summed score on another 5 items (where the items are sampled from the same domain), or how the sum of these five items today would correlate with the same five items tomorrow. In other words, we would like to know the reliability of the five item test as a total score, but all we have are the individual reliabilities of each item.

We denote the correlation of an item with another item as $r_{11}$ (some books use $r_{ij}$ to denote item $i$ with item $j$). The correlation of a sum of $k$ items with another sum of $k$ items is denoted $r_{kk}$.

Now comes the famous SB formula. To estimate $r_{kk}$ we take

$$r_{kk} = \frac{kr_{11}}{1 + (k - 1)r_{11}}$$

Thus, from the interitem correlations between $k$ items we estimate the reliability of the sum of $k$ items. This form is known as the standardized version because it is based on correlations. There is also an unstandardized version that I present below. SPSS computes both.

The limit of SB as $k$ goes to infinity (an infinitely large test) is 1, which represents perfect reliability. This means that as the number of items in a scale grows, the reliability will increase as well.

8. Cronbach’s $\alpha$

What do you do if you have many different inter-item correlations, so there are several estimates of $r_{11}$? You want to estimate a single $r_{kk}$ using Spearman-Brown, but how do you choose among the different observed $r_{11}$s?

Nothing deep here. Simply take the average correlation, apply S-B to the average, and you get what’s known as the standardized Cronbach’s $\alpha$. If you had thought of that, then everyone would be saying YOURNAMEGOESHERE’s $\alpha$.

So, to estimate $r_{11}$ from $k$ items compute all possible correlations and average (yes, average) the correlations to estimate a single $r_{11}$. In symbols,

$$r_{11} = \frac{\sum_{ij} r_{ij}}{k \choose 2}$$

where $k \choose 2$ is “$k$ choose 2”.
Let’s do this example from Nunnally & Bernstein. Imagine we have 3 judges who rate 8 subjects. We want to know the reliability of these three judges. One way to frame this question is how much do these three judges intercorrelate (here just take the average interitem correlation). Another way to frame this question is to ask what is the reliability of the sum of the three judges (i.e., if these judges were to rate another nine subjects and we summed the three judge’s scores, what would be the correlation of the original nine sums with the new nine sums?). This latter question is answered by applying the SB formula on the average interitem correlation (which is called Cronbach’s $\alpha$).

Example from data in Nunnally & Bernstein.

<table>
<thead>
<tr>
<th>Items</th>
<th>Judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The SPSS syntax for this problem is

```plaintext
data list free / j1 j2 j3.

begin data
1 1 1
1 1 1
1 1 1
0 0 0
0 1 0
0 0 0
0 1 1
0 0 0
0 1 1

end data.

reliabilities variables = j1 j2 j3
/statistics all.
```

The intercorrelations of these judges are
If you take the average correlation and apply SB, then you get the standardized alpha. In this example the average correlation \( r \) is \( .5997 \), apply SB to that and you get \( .8180 \). SPSS reports a “standardized item alpha” = \( .8180 \), which is the same thing!

Some textbooks report different ways of computing alpha. For example, using the covariance matrix (so this is the nonstandardized form of alpha)

\[
\begin{pmatrix}
J1 & J2 & J3 \\
J1 & 1 & \\
J2 & .4472 & 1 \\
J3 & .5774 & .7746 & 1
\end{pmatrix}
\]

Sum of all elements in the covariance matrix (9 terms) is 1.6965. The sum of the diagonal elements is 1.7679. Cronbach’s alpha is given by (Nunnally & Bernstein 6-26)

\[
\alpha = \frac{k}{k - 1} \left( 1 - \frac{\sum \sigma_a^2}{\sigma_y^2} \right)
\]

(13-12)

\[
= \frac{3}{2} \left( \frac{1.6965 - .7679}{1.6965} \right)
\]

(13-13)

\[
= .821
\]

(13-14)

This is the same as what SPSS reports in the unstandardized form.

A completely different way to compute Cronbach’s \( \alpha \) is through the intraclass correlation and then apply the SB formula. To compute the intraclass, first compute the ANOVA source table using items and judge as two factors (no interaction term is fitted):

Source table (k=3 judges)

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>items</td>
<td>.5655</td>
<td>3.958</td>
</tr>
<tr>
<td>judge</td>
<td>.1250</td>
<td>.58</td>
</tr>
<tr>
<td>residual</td>
<td>.1012</td>
<td>1.417</td>
</tr>
<tr>
<td>total</td>
<td>.1012</td>
<td>1.417</td>
</tr>
</tbody>
</table>
The ANOVA intraclass is .6046. This is an index of similarity of one judge to the next; it is not the average interjudge correlation. The equation for the ANOVA intraclass is

\[
\text{intraclass} = \frac{\text{MSB-MSW}}{\text{MSB} + (k-1)\text{MSW}} \quad (13-15)
\]

\[
= \frac{.5655 - .1012}{.5655 + (2)\cdot.1012} \quad (13-16)
\]

\[
= .6046 \quad (13-17)
\]

If you apply Spearman-Brown to this intraclass

\[
r_{kk} = \frac{k r_{11}}{1 + (k - 1)r_{11}} \quad (13-18)
\]

\[
= \frac{3(.6046)}{1 + 2(.6046)} \quad (13-19)
\]

\[
= .821 \quad (13-20)
\]

This last number is interpreted in terms of how this set of 3 judges will correlate to another set of 3. Again, the same number as before, the unstandardized Cronbach’s \( \alpha \).

With a little algebra both steps (the intraclass and the SB formula) can be condensed into a single expression:

\[
\frac{\text{MSB-MSW}}{\text{MSB}} \quad (13-21)
\]

This is what appears on page 277 Eq 7-20 in Nunnally & Bernstein. Cronbach’s \( \alpha \) is identical to applying the SB formula to the intraclass correlation, and the simple equation in these notes Eq 13-21 shows that the relevant info is embedded in ANOVA. So ANOVA also shows up in reliability theory.

It may seem strange to use the sum of squares (SS) for items rather than SS for judges (aka raters). Judges are conceptualized as a blocking factor so we want to remove from the error term any effect due to mean differences between the judges. So, let’s suppose the judges perfectly agree with each other. Then all three scores for each item will be the same, there won’t be any within item variance (so MSW goes to 0) and all the remaining variance is between items (MSB). In this extreme case when MSW = 0 the intraclass will be equal to 1. The other extreme is that all variance is within items (i.e., all three judges always disagree with each other), so MSB=0 and the intraclass in Equation 13-15 goes to -1/(k-1). Note that the intraclass correlation is asymmetric as it only approaches 0 when k is very large.

9. Reliability & Regression with One Predictor

How does measurement error influence a linear regression with one predictor? When the dependent variable has measurement error there is no serious problem because that is exactly
the kind of error regression is designed to deal with (but there will be loss of statistical power). For instance, take the usual

\[ Y = \beta_0 + \beta_1 X + \epsilon \]  

(13-22)

The \( \epsilon \) term captures both measurement error on \( Y \) and the prediction error of the equation. A simple regression cannot distinguish measurement noise in the criterion from noise in the prediction. The “true component” of \( Y \) is taken to be a linear transformation of the variable \( X \). Of course, measurement error on \( Y \) reduces the power of the tests (e.g., increases the size of the confidence interval around \( \beta_1 \)) but does not introduce bias. The requirement is that the error be independent from all the effects in the model. Violations of these assumptions can be dealt with easily (e.g., if errors are correlated across observations, as in a time series, then there exist techniques to take into account the “autocorrelation”).

There is no provision in the standard regression model for measurement error on \( X \). More generally, there is no provision in the standard regression model for measurement error on any predictor variable. But frequently regressions involve predictors that contain measurement error (e.g., the total score the self-esteem scale might be used as a predictor, the behavioral coding of a rat’s licking behavior may have error in measurement, etc). It turns out that measurement error on \( X \) not only reduces power, but it also introduces bias. The problem associated with bias are, in general, more serious than reduction in power. Thus, the inability for regression to handle measurement error in the predictor(s) is a very serious limitation.

We now turn to an examination of the effects of having a “noisy” predictor in a simple regression equation with just one predictor; this discussion is adapted from Neter, Wasserman, and Kutner, (1985). The only machinery we need to reach a better understanding is covariance algebra. Consider a simple linear regression as displayed in Equation 13-22. The kind of regression model we want to study is

\[ Y = \beta_0 + \beta_1 X_T + \epsilon \]  

(13-23)

where \( X_T \) denotes the true \( X \) score. However, when there is measurement error on \( X \) (I’m using \( X \) to denote the observed score), then the desired model can be expressed as

\[ Y = \beta_0 + \beta_1 (X - \epsilon_x) + \epsilon \]  

(13-24)

where \( \epsilon_x \) is the measurement error on \( X \). Note that the previous two equations are identical because \( X_T = X - \epsilon_x \).

This last equation shows that a crucial assumption in regression is being violated: independence of error terms. By re-arranging terms we get

\[ Y = \beta_0 + \beta_1 X + (\epsilon - \beta_1 \epsilon_x) \]  

(13-25)
and can see that the total error is correlated with $X^5$.

I have just shown that the independence assumption is violated when there is measurement error on $X$. Now let’s see exactly what happens to the $\beta_1$ term in the regression equation when there is measurement error on $X$. Below I show that the $\beta_1$ under measurement error for $X$ will always be less than the true $\beta_1$ (i.e., the true beta without measurement error).

We need to define some terms. $Y$ is the dependent variable, $X$ is the observed predictor, $X_T$ is the true predictor, $\epsilon_X$ is the measurement error on $X$, and $\epsilon$ is the measurement error on $Y$.

The $\beta_1$ for the regression of $Y$ on $X$ is equal to

$$\frac{\text{cov}(X,Y)}{\text{var}(X)}$$

(this is just the standard definition of the slope). Similarly, the true $\beta_1$, denoted $\beta_1^{\text{true}}$, for the regression of $Y$ on $X_T$ is

$$\frac{\text{cov}(X_T,Y)}{\text{var}(X_T)}$$

The covariance of $X$ and $Y$ is equal to (under independence of $\epsilon_X$ from $Y$)

$$\text{cov}(X_T + \epsilon_x, Y) = \text{cov}(X_T, Y) = \beta_1^{\text{true}} \cdot \text{var}(X_T)$$

One more definition. The reliability, as discussed above, will be denoted $r_{xx}$. The $xx$ subscript suggests the interpretation of the correlation of $X$ with itself (as in a test-retest situation). Following the definition of reliability (Equation 13-9)

$$r_{xx} = \frac{\text{var}(X_T)}{\text{var}(X)}$$

Now the last step, recall that the slope using the observed predictor is

$$\frac{\text{cov}(X,Y)}{\text{var}(X)}$$

But I just showed that the $\text{cov}(X,Y) = \beta_1^{\text{true}} \cdot \text{var}(X_T)$, so

$$\beta_1 = \frac{\text{cov}(X,Y)}{\text{var}(X)}$$

---

5 As a do-it-yourself exercise, you might want to show the covariance between $X$ and $(\epsilon + \beta_1\epsilon_x)$ equals $\beta_1 \text{var}(\epsilon_x)$. This value will not usually be zero and thus, there is a violation of the independence assumption.
Lecture Notes #13: Reliability and Structural Equation Modeling

\[
\beta_1 = \frac{\text{var}(X_T)}{\text{var}(X)} = \beta_{1\text{true}} r_{xx}
\]  
(13-33)  
(13-34)

Now because \(r_{xx}\), the reliability of \(X\), is always a number between 0 and 1 \(\beta_1\) will be less than or equal to \(\beta_{1\text{true}}\). So measurement error introduces systematic bias on the slope parameter. Things get more complicated, of course, with multiple predictors.

What do you when your predictor has measurement error? There are several things to do to deal with measurement error. First, if you found significance with a “noisy” predictor, then you are probably okay because noise probably made the test less powerful\(^6\). The problem comes when you fail to find significance—it might be that the true \(X\) variable is related, but the measurement error masked that relationship. Second, a technique that some people have used involves the use of “instrumental variables.” An instrumental variable is correlated with the true \(X\) but not with the measurement error on \(X\). Instrumental variables can then be included in the analyses in a way that helps adjust for the error issue. A practical problem with this technique is that it is difficult to know when a variable is correlated with true \(X\) but not with \(\epsilon_X\). A third technique involves the use of structural equations models (SEM). These models include measurement error directly in the regression equation. Constructing a model and developing an algorithm for estimating/testing the parameters of the model is, obviously, the best thing to do.

10. Reliability and Multiple Regression

All bets are off when trying to get a handle on the effects of measurement error on predictors in the context of multiple regression. Measurement error can increase, decrease, or even change the sign of correlations and betas at the observed level. As you might suspect, getting a deep understanding of the effects of measurement error in the context of multiple regression is not easy. Following Cohen and Cohen (1983, pp. 406-413) we will build intuition by using the notion of partial correlations.

Recall from Lecture Notes 8 that a partial correlation involves a correlation between two variables of interest where the effects of a third variable (typically a nuisance variable or a covariate, but sometimes a theoretically meaningful variable such as a mediator) have been removed. One way to think about this is to imagine a correlation between \(Y\) and \(X\) where you want to “partial out” the effects of a variable \(W\). The partial correlation between \(Y\) and \(X\) (partialling the effects of \(W\)) is equivalent to performing two regressions (\(Y\) on \(W\) and \(X\) on \(W\)) and correlating the residuals from the two regressions. In other words, the regression of \(Y\) on \(W\) creates residuals that have the linear effect of \(W\) “removed” from \(Y\) and the regression of

\(^6\)This may not generally be the case because the above result was obtained under the assumption that the measurement error on \(X\) is independent from \(Y\). How violations of such assumptions influence both \(\beta_1\) and the variance of \(\beta_1\) is very complicated and no simple statement can be made.
X on W creates residuals that have the linear effect of W “removed” from X. The correlation of these two sets of residuals is the partial correlation.

A relatively simple way to compute the partial correlation is by the formula

$$ r_{yx,w} = \frac{r_{yx} - r_{yw}r_{xw}}{\sqrt{(1 - r^2_{yw})(1 - r^2_{xw})}} $$

(13-35)

The partial correlation formula (Equation 13-35) provides a convenient way to understand the effects of measurement error on each of the variables. Cohen and Cohen (1983, p406-413) discuss the different effects of having measurement error on Y, X, and W. Here is a table showing the effects of measurement error on the observed partial correlation. We allow measurement error on variable W and assess the effects on the partial correlation $r_{yx,w}$. I only display a few cases (all for reliability .7), but the effects are more dramatic for lower reliability.

<table>
<thead>
<tr>
<th>example</th>
<th>$r_{yx}$</th>
<th>$r_{yw}$</th>
<th>$r_{xw}$</th>
<th>$r_{yw}$</th>
<th>$r_{y,w}$</th>
<th>true $r_{y,w}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.3</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>0</td>
<td>-.23</td>
</tr>
<tr>
<td>2</td>
<td>.5</td>
<td>.7</td>
<td>.5</td>
<td>.7</td>
<td>.24</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>.5</td>
<td>.7</td>
<td>.6</td>
<td>.7</td>
<td>.14</td>
<td>-.26</td>
</tr>
<tr>
<td>4</td>
<td>.5</td>
<td>.3</td>
<td>.8</td>
<td>.7</td>
<td>.45</td>
<td>.57</td>
</tr>
<tr>
<td>5</td>
<td>.5</td>
<td>.3</td>
<td>.6</td>
<td>.7</td>
<td>.42</td>
<td>.37</td>
</tr>
</tbody>
</table>

Look at how discrepant the observed partial is from the true partial. That is with a reliability of .7, which many people take to be acceptable.

The structural model representing this multiple regression with a predictor having error is depicted below.

![Structural Model Diagram](image)

11. Instrumental Variables
Instrumental variables is a popular technique in some of the social sciences that allows one to exploit a randomization-like structure to get better causal estimates even in the context of a correlational design. A simple example, taken from Gennetian et al., 2008, *Developmental Psychology*, 44, 381-, involves the relation between mother’s education attainment (ME) and child’s performance (CP). At best the relation between those two variables is a correlation and without additional information or assumptions one could not move toward “causal inferences” of the effect of ME on CP. Instrumental variables serve the role of that additional information allowing one to do a little more with correlational information; random assignment is another way to move toward causal inferences, but random assignment is not always possible.

Suppose that moms are randomly assigned to a treatment condition (T), such as an intervention designed to influence her educational attainment ME (but not influence in a direct way the child’s educational performance). Mom’s educational attainment (ME) is assessed after this intervention. One can run a regression with ME = β₀ + β₁T + ε₁, which if T is a dummy code then the slope represents the difference between treatment and control. So that can establish the “causal effect” of the treatment T on ME. But we still don’t know about the outcome of interest CP. Do changes in ME “cause” subsequent changes in CP? Is it possible to influence CP by influencing ME?

We can make use of some of the structure in this setup and make a few assumptions. We will use what is called a “two stage least squares estimation”, which is just a fancy way of saying run one regression, take relevant output from that regression and use it in a second regression (so two stages). The first regression is the one I mentioned in the previous paragraph: take the predicted values ŶE. This represents the predicted educational attainment of mom given her random assignment to treatment T; those predicted scores do not include the residuals (i.e., noise) and hence any correlation between the residuals ε₁ and other variables have been removed. In other words, ME and CP could be correlated because they share a common variable, but by using the predicted values ŶME that are associated with the treatment T we reduce the impact of such issues like common factor and correlated residuals.

Then use that predicted value of ME in a second equation to predict the child’s educational performance, i.e., the regression CP = β₀′ + β₁′ME + ε₂ where primes are used to distinguish the slope and intercept from the previous regression. The two sets of residuals are denoted by 1 and 2 for first and second regression. If we make some assumptions, such as the correlation between the treatment and the residuals in the second regression is 0, we can derive an estimate of the direct effect (some would say the causal link) between mother’s educational attainment and child’s performance. This example used a random assignment, but econometricians have shown that even when there isn’t random assignment, as long as some additional assumptions are met, one can use this logic to derive causal estimates even from a correlation design.

The two step regression I outlined above yields the right parameter estimate, but the standard error for the regression slope β₁’ will be off because the second regression does not properly
model the correlation between the epsilons. There are specialized programs, including one in SPSS, that run instrumental variable analyses and yield the proper error term. In SPSS it is known as a “two stage least squares.” Here is some example syntax using the variables in the example I’ve been discussing.

2SLS CP WITH ME  
/INSTRUMENTS T  
/CONSTANT.

You specify the dv and critical predictor in the first line and define the instrument in the second line.

We can use our new knowledge in covariance algebra to figure out what all the fuss is about with instrumental variables. We know that a slope is equal to the ratio of the covariance between the dv and the predictor and the variance of the predictor (LN#6). Thus, in the second regression just described, the slope $\beta'_1$ can be rewritten as

$$\text{cov}(\text{CP}, \hat{\text{ME}})/\text{var}(\hat{\text{ME}})$$  \hspace{1cm} (13-36)

Substituting in the regression equation for $\hat{\text{ME}}$ leads to the familiar equation for the instrumental slope after making some assumptions about correlated error

$$\beta'_1 = \frac{\text{cov}(\text{CP}, \text{T})}{\text{cov}(\text{ME}, \text{T})}$$  \hspace{1cm} (13-37)

This slope amounts to a ratio of two differences: the difference in treatment for the dependent variable CP over the difference in treatment for the predictor variable mother’s educational attainment.

This equation can be derived from taking Equation 13-36 and using covariance algebra to re-express the terms to simplify to Equation 13-37. That is,

$$\beta'_1 = \frac{\text{cov}(\text{CP}, \hat{\text{ME}})}{\text{var}(\hat{\text{ME}})}$$

$$= \frac{\text{cov}(\text{CP}, \beta_0 + \beta_1 \text{T})}{\text{var}(\beta_0 + \beta_1 \text{T})}$$

$$= \frac{\beta_1 \text{cov}(\text{CP}, \text{T})}{\beta_1^2 \text{var}(\text{T})}$$

$$= \frac{\text{cov}(\text{CP}, \text{T})}{\text{cov}(\text{ME}, \text{T})}$$
Note that \( \hat{\text{ME}} \) (predicted ME from the first regression in the two stage model) in this setup does not have a residual variance so the numerator and denominator in the second line don’t include an \( \epsilon_1 \). The last line follows because \( \beta_1 = \frac{\text{cov}(\text{ME}, T)}{\text{var}(T)} \).

I introduce this example here to illustrate how covariance algebra can illuminate some complicated statistical ideas. Instrumental variables are not easy to understand but by using covariance algebra the idea and the importance of its assumptions fall out naturally.

The instrumental variables logic can be applied in a structural equation modeling (SEM) context as well. One needs to be careful because SEMs are usually estimated in a maximum likelihood context whereas instrumental variables are traditionally estimated in a two stage least squares context, so the estimates and their standard errors may differ slightly if one compares the output from SEM to the output from a traditional instrumental variables program.

The SEM representation of instrumental variables is given in this Figure

![Diagram](image)

This is similar to the mediation model but the focus in instrumental variables is on the intermediate path between ME and CP (slope \( \beta_1 \) above) rather than whether there is a direct path between T and CP (as in the case of mediation). One of the key assumptions of an instrument is that it NOT have a direct path to the final variable, so the instrument should not directly influence CP except through the mediator. In the mediation literature this is called perfect mediation.

The IV literature has correctly addressed the issue of correlated residuals. Related issues of measurement error have been pretty much ignored in the mediation literature (though mentioned in the original Baron & Kenney article). As we saw earlier in these lecture notes, measurement error in the context of a multiple regression can create serious problems.

Some would argue that a natural use of instrumental variables in behavioral work is when one has a manipulation check and wants to examine the effect of the manipulation check variable on the dependent variable. In this case the variable representing the manipulation is the instrument (e.g., group 1 vs group 2). The manipulation check becomes the predictor of interest such that the researcher wants to examine how the variability in the manipulation played out (i.e., some people responded more to the manipulation than others), so the key
question becomes how the manipulation check predicts the dependent variable instead of how the intended manipulation predicts the dependent variable. The former acts as though everyone responded the same way to the treatment, the latter uses the “actual” response to the manipulation as measured by the manipulation check. In other words, if one uses the actual manipulation variable as the only predictor, then one examines, in a sense, the “intent to treat” because the analysis focuses on the condition the subjects are in not how they responded to the treatment. However, if one uses the manipulation variable as an instrument, runs a regression with the manipulation check as the dependent variable, uses the predicted values from that regression as predictors in the second stage regression, makes several assumptions, then one examines the effect of the manipulation check (how the manipulation actually played out) on the dependent variable. This approach though still uses a type of “average treatment effect” and glosses over heterogeneity in the manipulation check and some of the assumptions may be difficult to justify such as whatever the reason some people responded differently to the manipulation as measured by the manipulation check is unrelated to the residual in the regression that contains the key dependent variable of interest.

The potential use of instrumental variables in correlation studies is interesting, though personally I don’t find it very compelling. One still needs to make a set of assumptions that cannot easily be empirically validated. The pessimistic take is that if one doesn’t use instrumental variables, then in order to interpret the correlation as “causal” one must make lots of handy wavy arguments (i.e., assumptions) about there not being a third variable, that causation goes in one direction, etc. In order to use instrumental variables, one’s hands are waving just as much, but in the end one has a formula that leads to an estimate of the “causal relation.” While I recognize that randomized studies cannot always be carried out, I can’t break away from my traditional upbringing that “cause” should be reserved for randomized experiments. I can’t distinguish whether the pair of waving hands from the person who uses instrumental variables is more or less convincing than the pair of waving hands from the person who doesn’t and tries to convince me why the reported correlation should be interpreted in a causal manner. I’m happy using correlations in my research as they provide important information about associations and also using randomized designs when possible to address cause. But when it comes to waving, I prefer to give the parade wave as I ride by in my statistics float to all those who claim they can squeeze cause out of correlation.

12. Mediation

The basic mediation analysis uses a similar structure to instrumental variables but the emphasis is on a different part of the path diagram. Whereas in instrumental variables the focus is on examining the path between M and Y, in mediation analysis the question is whether X influences Y directly (the curved path) or whether X influences Y indirectly through M.
A colleague gave me a nice way to conceptualize the meaning of an indirect path. Suppose you tell me a secret and I turn around and tell your advisor. One wouldn’t say that you told your advisor (i.e., there was no direct communication link between you and your advisor). I was the middleman who took input from you and spewed it to your advisor. I’m the “mediator” in that example—there are two communication links, one from you to me and another from me to your advisor.

A lot of behavioral science starts off being simple stimulus-response (questions such as does X influence Y). Then those questions become black box questions such as the mechanism or process through which X influences Y is that X influences some intermediate process and that process influences Y. If I pay my son to spend more time studying, and then he gets a high grade on a test, the stimulus-response operational language is that payment led to high grade. But the black box view says, no, payment for study led my son to spend more time studying, which meant he learned the material better, which led to better comprehension, which led to a higher performance on the exam compared to the counterfactual where I didn’t pay him to study more thereby breaking the chain. You see these kinds of black box theories in many behavioral science literatures. The treatment affected behavior, not because it directly changed behavior but because it changed a psychological state and in turn that new psychological state contributed to the change in behavior.

Testing mediation is tricky. The modern method is to run an SEM model and test the product of two $\beta$s (the beta in the path from X to M times the beta in the path from M to Y). Testing products of two betas is not easy, and has led to specialized methods using bootstrap and Bayesian approaches.

Even if one can conduct a randomized experiment by manipulating X, it isn’t clear that the randomized experimental approach really helps because the relation between M and Y isn’t manipulated if one only manipulates X. If one manipulates both X and M, then that isn’t directly testing mediation—one could test, for example, for main effects and interaction of X and M on Y. But that isn’t testing the black box proposition that X influences M (you tell me a secret) and in turn M influences Y (I then tell the secret to your advisor). So manipulation doesn’t necessarily help with these kind of black box models.

If there is time at the end of the term I’ll come back and talk more about mediation, but we should really offer an entire course on the variants of mediation analysis. I just want to introduce the concept of mediation so you have some familiarity if you encounter it in your
research literature.

13. Structural Equations Modeling

SEM is a general procedure that can do all ANOVA, regression, PCA, MANOVA, and canonical correlation. The framework also allows for measurement error. It is possible to mix and match these techniques, such as perform a PCA with measurement error within a regression model.

Words of Caution: Structural equations programs such as LISREL, EQS, CALIS, EZPATH, Amos, Mplus and the like are very tricky. It is easy for an inexperienced user (sometimes even an experienced user) to make a mistake and not be aware that he or she has made one. It is quite tempting to play with structural/measurement models (add a path here, remove a path over there, correlate an error, free a variance, etc). One little change can have a dramatic impact on the subtle features of the model (e.g., identification). You should develop a deep distrust of structural equations models that appear in the literature. I hope we can instill an appreciation for the elegance and necessity of this type of modeling as well as the motivation for the need to seek further training.

I will illustrate with one application of SEM showing how canonical correlation can be extended to include measurement error so that one can assess the reliability of each observed variable along with the factor structure.

14. Factor Models in SEM

Structural equations modeling presents an excellent way to estimate correlations between latent variables, or factors, because the technique automatically takes into account measurement error (i.e., one does not need to correct for attenuation). Also, the analysis automatically provides factor loadings and reliabilities along with the relation between the factors.

Recall the self-esteem example from Lecture Notes #12 where I illustrated canonical correlation. Figure 13-2 illustrates the generalization in graphical form of canonical correlation to allow for measurement error. The new addition is that there are now 10 circles labelled E1 to E10 pointing to their respective observed variables. For example, this shows that observed variable V01 has two arrows point to it: the unobserved factor F1 and the unobserved measurement error E1 thus representing the regression equation

\[ V01 = \beta F1 + E1 \]

There are nine such equations for each of the remaining nine observed variables V02 to V10. Each of these equations separately assesses the measurement error of each variable (the unobserved E) and the common component, the unobserved factor F. One would need to estimate
the 10 $\beta$s, the variances of the two factors, and the variances of the unobserved errors, the 10 $E$'s. There are a few details to have to worry about involving identification of parameters as we will soon see. Below I present the logic of goodness of fit measures and tests, which compare the observed covariance matrix to the model implied covariance matrix. In this case the

15. Models

Recall that in regression models one is assessing the fit between the observed score $y$ and the score predicted by the model, $\hat{y}$. A residual is the difference between the observed score and the predicted score:

$$\text{residual} = \text{observed score} - \text{predicted score} \quad (13-38)$$

In multiple regression the score predicted by the model is completely determined by the regression coefficients. One way to think about regression coefficients is that they are the weights that minimize the sum of squared residuals (where the sum is taken over cases); that is, for each subject

$$\text{residual} = \text{observed score} - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots) \quad (13-39)$$
and the coefficients are selected so $\sum$(residual$^2$) (sum of squared residuals over subjects) is minimized. The residual for a case is the estimated $\epsilon$ for that case.

This concept of difference between observed and predicted can be extended to covariance and correlation matrices.

$$\text{residual matrix} = \text{observed matrix} - \text{predicted matrix} \quad (13-40)$$

The difference is taken element by element. For example, assume the observed covariance matrix

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>w</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>0.25</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>w</td>
<td>0.13</td>
<td>3.24</td>
<td>0.87</td>
</tr>
<tr>
<td>y</td>
<td>0.05</td>
<td>0.87</td>
<td>0.95</td>
</tr>
</tbody>
</table>

and the expected covariance matrix from some model

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>w</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>0.25</td>
<td>0.30</td>
<td>0.15</td>
</tr>
<tr>
<td>w</td>
<td>0.30</td>
<td>3.56</td>
<td>1.46</td>
</tr>
<tr>
<td>y</td>
<td>0.15</td>
<td>1.46</td>
<td>1.37</td>
</tr>
</tbody>
</table>

The matrix of residuals is

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>w</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>0.00</td>
<td>-0.17</td>
<td>-0.10</td>
</tr>
<tr>
<td>w</td>
<td>-0.17</td>
<td>-0.327</td>
<td>-0.59</td>
</tr>
<tr>
<td>y</td>
<td>-0.10</td>
<td>-0.59</td>
<td>-0.75</td>
</tr>
</tbody>
</table>

A matrix of residuals is difficult to interpret because the metric is not well-specified (e.g., is -0.75 residual something to worry about?). So we need a metric that can be interpreted as goodness of fit of the model. Such an index is usually abbreviated GFI (goodness of fit index).

16. **GFI**

The material in the next two sections is adapted from Bollen (1989) and Everitt (1984).
First, let’s make a connection between GFI and the familiar concept of $R^2$. Recall that

$$R^2 = \frac{SSR}{SST}$$

(13-41)

$$= \frac{SST - SSE}{SST}$$

(13-42)

$$= 1 - \frac{SSE}{SST}$$

(13-43)

The last equation re-expresses $R^2$ as 1 minus the ratio of observed error over total error. The observed total error can be interpreted as a “baseline” to compare the observed error. Note that the baseline (the denominator) is based on observed scores not expected scores.

The sum of the squared elements in the residual matrix is a nice candidate for the analog to SSE. Square every element in the residual matrix, then sum all the squared elements. The analog to SST is just the sum of the squared elements in the observed covariance matrix (yes, squared variances). This leads to

$$GFI = 1 - \frac{\sum \text{squared terms from residual matrix}}{\sum \text{squared terms from observed cov matrix}}$$

(13-44)

How is this index interpreted? When the model is perfect, then the sum of squared residuals will be 0 and $GFI = 1$. When the model is “terrible” in the sense that the model predicts the same number for every subject, the sum of squared residuals will equal the sum of squared terms from the covariance matrix. Note that this “terrible” model could arise in cases where you know nothing and the best prediction for each subject on each variable is the mean of that variable. It is possible for this index to be negative.

NOTE: this particular GFI index is now outdated and has been replaced with better indices. The new indices are computationally intensive and are not as illuminating as the one presented here. See Bollen (1989) for newer treatments. All GFI measures are based on the same intuition. I present this particular form only because it illustrates the ideas in the simplest way.

17. Testing Fit

A simple Chi-square test for testing the fit of the model is

$$X^2 = \frac{(N-1)}{2} \text{sum of squared terms from residual matrix}$$

(13-45)

The computed $X^2$ can be compared to the tabled Chi-square value. The degrees of freedom are

$$\frac{k(k+1)}{2} - p$$

(13-46)

where $k$ is the total number of variables (i.e., number of rows or columns in the covariance matrix) and $p$ is the number of parameters that are being estimated (e.g., the total number of
regression coefficients that are used to compute the expected covariance matrix, not counting any parameter that is associated with error). The total number of parameters is simply the total number of slopes present in the model. Note that the intercepts don’t enter here because adding or subtracting a constant has no effect on the variance or the covariance.

There are lots of peculiar things about this Chi-square test. First, the number of subjects does not enter into the computation of degrees of freedom. The reason is that what is being taken as data is the covariance matrix not the individual data points. The parameter $k$, the number of variables, captures the “size” of the covariance matrix. Second, the null hypothesis here is the model you are trying to fit. So you want to accept the null hypothesis. A significant Chi-square means that the model you are testing does not fit the data, where by “fitting the data” we mean does the covariance matrix implied by the model match the covariance matrix of the data. Third, the sample size does enter into the computation of the actual Chi-square statistic, even though it did not enter into the GFI measure. So, with enough subjects (i.e., power) you will usually be able to reject the null hypothesis. There is an unusual tradeoff here: in order to be sure about accepting the null hypothesis you need lots of power but too much power might lead to erroneous rejection of a reasonable model.

NOTE: this particular Chi-square test is now outdated and has been replaced with better tests. The new tests are computationally intensive and are not as illuminating as the one presented here. The newer tests do try to get around some of the problems of this test. Also, some of the newer tests count parameters a little differently. We will discuss these newer tests later in the context of structural equations modeling.

18. Theoretical v. Empirical Models

If you know the true parameters, you can compute the expected covariance matrix and compare it to the observed covariance matrix using the GFI and the Chi-square test. However, typically one does not know the true parameters of the models. To what matrix is the observed covariance matrix compared when the true model is not known?

The answer is easy. Assume a model, and run the regressions to estimate the relevant path coefficients. Now you have a model with numerical estimates. You can use the slopes estimated from these regression equations to create the model-implied covariance matrix. For example, suppose I was testing the model that $w$ mediates the relation between $x$ and $y$. Following Baron & Kenny I would run two regressions (a third regression is an intermediate one and is not used here): regress $y$ on both $w$ and $x$ and regress $x$ on $w$. Suppose the two regressions resulted in the following slopes and intercepts

$$y = 3.84 + .46w + .27x$$  \hspace{1cm} (13-47)

$$w = -2.22 + 1.62x$$  \hspace{1cm} (13-48)

With this information one can compute the model-implied covariance matrix and compare it
to the observed covariance. That is, the regression output provides estimates that are used to compute the “expected” covariance matrix. This “expected” covariance matrix can be compared to the observed covariance matrix to see how well the model is doing in capturing the observed data. The GFI index and $X^2$ test can also be calculated. This is how the program works. Instead of known the regression betas, they are estimated in order to maximize GFI. That is, the program searches the optimal combination of betas that produce the expected covariance matrix that is closest to the observed covariance matrix.

In this way we can simultaneously assess the fit of several regression models. This is useful for more realistic testing where there may be several dependent variables and multiple hypotheses over these variables.

Thus, one use of SEM is to estimate several regression equations simultaneously (i.e., a system of regression equations).

19. SEM example

Let’s do a simpler example with just six observed variables and two factors. I will use the six equations to set up a model that predicts a 6x6 covariance matrix. This model implied covariance matrix will be compared to the observed covariance matrix to estimate parameters ($\beta$s, factor variances, error variances, factor covariances). The six equations are listed here:

\begin{align*}
  x_{1A} & = \beta_1 F_1 + E_1 \\
  x_{1B} & = \beta_2 F_1 + E_2 \\
  x_{1C} & = \beta_2 F_1 + E_3 \\
  x_{2A} & = \beta_3 F_2 + E_4 \\
  x_{2B} & = \beta_4 F_2 + E_5 \\
  x_{2C} & = \beta_4 F_2 + E_6
\end{align*}

Here we have six observed variables, $x_{1A}$ to $x_{2C}$. Each variable has its own error (denoted $E_1$ to $E_6$). There are two factors denoted $F_1$ and $F_2$. This can be interpreted as a set of six regression equations, where the goal is to estimate the six $\beta$s and the six error variances using two unknown factors. In this simple model we will set the variances of the unknown factors to 1 and force the two factors to be independent. We will relax the second assumption (uncorrelated factors) later. I will use the SEM program called AMOS.

Here is the AMOS syntax. AMOS has a nice feature that it is possible to draw the SEM diagram on the screen to build the model (rather than entering syntax) but you really need to know what you are doing since there are constraints that need to be set and such. Best to start
with syntax and then when you understand the program work your way over to the graphical
interface.

```vbnet
#Region "Header"
Imports System
Imports System.Diagnostics
Imports Microsoft.VisualBasic
Imports AmosEngineLib
Imports AmosGraphics
Imports AmosEngineLib.AmosEngine.TMatrixID
Imports PBayes
#End Region

Module MainModule
Public Sub Main()
    Dim Sem As AmosEngineLib.AmosEngine = New AmosEngineLib.AmosEngine
    Sem.TextOutput()
    Sem.Standardized()
    Sem.Smc()
    Sem.Dispose()
    End Sub
End Module

Sem.BeginGroup("f:\rich\teach\multstat\unixfiles\lectnotes\sem\rg1.sav")
    Sem.AStructure("x1A = (1) F1 + (1) err_x1A")
    Sem.AStructure("x1B = F1 + (1) err_x1B")
    Sem.AStructure("x1C = F1 + (1) err_x1C")
    Sem.AStructure("x2A = (1) F2 + (1) err_x2A")
    Sem.AStructure("x2B = F2 + (1) err_x2B")
    Sem.AStructure("x2C = F2 + (1) err_x2C")
    Sem.AStructure("err_x1A (ex1A)")
    Sem.AStructure("err_x1B (ex1B)")
    Sem.AStructure("err_x1C (ex1C)")
    Sem.AStructure("err_x2A (ex2A)")
    Sem.AStructure("err_x2B (ex2B)")
    Sem.AStructure("err_x2C (ex2C)")
    Sem.AStructure("F1 (VF1)")
    Sem.AStructure("F2 (VF2)")
Sem.Dispose()
End Sub
End Module
```
Let's walk through some of the syntax. The first few lines determine some of the output, such as setting text output and printing the standardized solution as well as the raw solution. The Smc piece adds the reliabilities to the output. The file containing the SPSS data is “rg1.sav”. The next block of output sets up the six regression equations. For each observed variable we define a factor as a predictor and set up measurement error. Then we define the six errors of measurement and the variances of the two factors are set to 1. The correlation between the two factors is set to 0.

The output is presented in Figure 13-3. The output includes the six regression equations equations, with betas and their individual tests, the standarized betas, the six error variances, tests and goodness of fit measures for the entire model (all six regressions simultaneously fit and tested to the covariance matrix), and finally the reliabilities of each variable.

One may be interested in testing the correlation between the two factors. You run the same syntax except change the one line that fixes the correlation to 0 so that the program estimates the correlation. The one line of syntax is changed to the following where “corr” is an arbitrary label I assign to that parameter:

```
Sem.AStructure ("F1 <> F2 (corr)")
```

After running this I get new output which I present only snippets in Figure 13-4. The covariance between the two factors is .206 and is statistically significant.

It turns out that the test presented in most SEM programs for individual variables is incorrectly defined (I can supply a paper to those who are interested). The best way to test parameters such as the covariance or correlation between factors is to compare the Chi square tests across the two runs, the full and reduced model where the full model is when the parameter is free and the reduced model is when the parameter is set to the value of the null hypothesis (which is usually 0). Looking at these two outputs we have a Chisq of 9.25 with 8 degrees of freedom for the full model and a Chisq of 13.724 with 9 degrees of freedom for the reduced model. Obviously, the reduced model fits worse (a greater Chisq) because it has one fewer parameter to fit the data. There is one degree of freedom difference because the only change from the reduced to the full model is that we estimate one more parameter (the value of the covariance between the two factors). The difference between those two Chisq values is itself a Chisq distribution, so we can take 13.724 - 9.25 = 4.474, with one degree of freedom (9-8=1). You then look up the p-value for a Chisq of 4.474 with 1 d.f., which is about 0.035.

20. What to Report from an SEM Run?

What do you do after you run all these models? You would like to decide which model “fits” the data better so you need to supply the necessary information. You can organize the output
Figure 13-3: Two factor model with six observed variables. Factor variances set to 1.

Regression Weights: (Group number 1 - Model 1)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>S.E.</th>
<th>C.R.</th>
<th>P</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1A &lt;--- F1</td>
<td>1.03924</td>
<td>.08134</td>
<td>12.7722</td>
<td>*</td>
<td>beta1</td>
</tr>
<tr>
<td>x1B &lt;--- F1</td>
<td>1.26835</td>
<td>.10563</td>
<td>12.0076</td>
<td>*</td>
<td>beta2</td>
</tr>
<tr>
<td>x1C &lt;--- F1</td>
<td>.87597</td>
<td>.08440</td>
<td>10.3790</td>
<td>*</td>
<td>beta3</td>
</tr>
<tr>
<td>x2A &lt;--- F2</td>
<td>.98133</td>
<td>.08333</td>
<td>11.7768</td>
<td>*</td>
<td>beta4</td>
</tr>
<tr>
<td>x2B &lt;--- F2</td>
<td>1.03091</td>
<td>.07877</td>
<td>13.0876</td>
<td>*</td>
<td>beta5</td>
</tr>
<tr>
<td>x2C &lt;--- F2</td>
<td>1.39521</td>
<td>.10253</td>
<td>13.6082</td>
<td>*</td>
<td>beta6</td>
</tr>
</tbody>
</table>

Standardized Regression Weights: (Group number 1 - Model 1)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1A &lt;--- F1</td>
<td>.928</td>
</tr>
<tr>
<td>x1B &lt;--- F1</td>
<td>.891</td>
</tr>
<tr>
<td>x1C &lt;--- F1</td>
<td>.806</td>
</tr>
<tr>
<td>x2A &lt;--- F2</td>
<td>.867</td>
</tr>
<tr>
<td>x2B &lt;--- F2</td>
<td>.926</td>
</tr>
<tr>
<td>x2C &lt;--- F2</td>
<td>.947</td>
</tr>
</tbody>
</table>

Variance: (Group number 1 - Model 1)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>S.E.</th>
<th>C.R.</th>
<th>P</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>err_x1A</td>
<td>.175</td>
<td>.056</td>
<td>3.121</td>
<td>.002</td>
<td>ex1A</td>
</tr>
<tr>
<td>err_x1B</td>
<td>.416</td>
<td>.093</td>
<td>4.449</td>
<td>***</td>
<td>ex1B</td>
</tr>
<tr>
<td>err_x1C</td>
<td>.413</td>
<td>.065</td>
<td>6.377</td>
<td>***</td>
<td>ex1C</td>
</tr>
<tr>
<td>err_x2A</td>
<td>.318</td>
<td>.051</td>
<td>6.269</td>
<td>***</td>
<td>ex2A</td>
</tr>
<tr>
<td>err_x2B</td>
<td>.177</td>
<td>.040</td>
<td>4.444</td>
<td>***</td>
<td>ex2B</td>
</tr>
<tr>
<td>err_x2C</td>
<td>.223</td>
<td>.066</td>
<td>3.363</td>
<td>***</td>
<td></td>
</tr>
</tbody>
</table>

Squared Multiple Correlations: (Group number 1 - Model 1)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2C</td>
<td>.89717</td>
</tr>
<tr>
<td>x2B</td>
<td>.85708</td>
</tr>
<tr>
<td>x2A</td>
<td>.75187</td>
</tr>
<tr>
<td>x1C</td>
<td>.65032</td>
</tr>
<tr>
<td>x1B</td>
<td>.79467</td>
</tr>
<tr>
<td>x1A</td>
<td>.86063</td>
</tr>
</tbody>
</table>

MODEL 1

<table>
<thead>
<tr>
<th></th>
<th>NPAR</th>
<th>CMIN</th>
<th>DF</th>
<th>P</th>
<th>CMIN/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default model</td>
<td>12</td>
<td>13.724</td>
<td>9</td>
<td>.132</td>
<td>1.525</td>
</tr>
</tbody>
</table>
Figure 13-4: Two factor model with six observed variables. Factor variances set to 1 and are allowed to be correlated.

<table>
<thead>
<tr>
<th>Covariances: (Group number 1 - Model 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>F1 &lt;--&gt; F2</td>
</tr>
</tbody>
</table>

| CMIN |
| Model | NPAR | CMIN | DF | P | CMIN/DF |
| Default model | 13 | 9.25471 | 8 | .32127 | 1.15684 |
| Saturated model | 21 | .00000 | 0 | | |
| Independence model | 6 | 576.53262 | 15 | .00000 | 38.43551 |

| RMR, GFI |
| Model | RMR | GFI | AGFI | PGFI |
| Default model | .05633 | .97575 | .93636 | .37172 |
| Saturated model | .00000 | 1.00000 | |
| Independence model | .66869 | .42338 | .19273 | .30241 |
in a way that facilitates comparison. For example, here is one suggestion (you might prefer a different way). You could also list the residual matrix of each model and eye-ball to see where each of them fails. Once you decide on a best model, then you present the relevant parameter estimates and interpret the parameters of the best fitting model.

<table>
<thead>
<tr>
<th>Model</th>
<th># parameters</th>
<th>GFI</th>
<th>CFI</th>
<th>RMSEA</th>
<th>Chi-square</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Keep in mind that there is a trade-off between number of parameters and how well a model fits. Are there some models that clearly stick out as not being good in comparison to the rest?

Recall from a previous lecture that it is possible to compare two models by taking the difference between the two Chi-square values and subtracting to two degrees of freedom (always taking the difference in the order “reduced minus full”). This works when one model is nested within the other (i.e., the two models are identical except that one or more paths in the full model are forced to zero in the reduced model).

Whenever you have a standard error for a parameter, you can create a confidence interval around that parameter by using the usual “plus or minus 1.96 times the standard error” rule. But as I hinted above, the tests as defined in most SEM programs give poor approximations to the standard error of parameters. It is best to always test parameters using the “reduced minus full” approach I showed above; that is, set the parameter to the value of the null hypothesis (reduced) and rerun the model with the parameter as a free parameter (full model). See Gonzalez & Griffin (2001) for an explanation for why this approach should be used over the other possible approaches.

21. SEM and Hierarchical Linear Models

There is a simple connection between SEM and another kind of model that is popular these days called multilevel models, or hierarchical linear models.

We can think of the observed “squares” as nested within the unobserved “circles.” Think back to ANOVA designs when we talked about random effects and nested effects. Recall that we discussed a repeated measures design as having a subjects factor that is treated as a random effect and repeated observations being connected to a given subject. For example, if there are three observations per subject we would write an ANOVA design as

\[ Y_{ij} = \mu + \alpha_i + \pi_j + \epsilon_{ij} \]  

(13-55)
with \( \pi \) a random effect representing subjects.

We saw this same design in regression where we treated each subject as a factor and created N-1 dummy codes to properly account for the subject variance.

We can represent this as a structural equation model with latent variable representing the intercept term and slope terms.

We can represent this as a hierarchical linear model. This is done in SPSS using the MIXED command. The general syntax of the MIXED command looks like this:

```
MIXED DV with X1 X2 etc XK
   /print = solution
   /method=ml
   /fixed = X1 X2 etc XK
   /random = intercept X1 | subject(ID) covtype(UN).
```

This syntax has a dependent variable DV where the intercept and the slope for X1 are random (and nested within the variable code ID with covtype(UN) meaning unstructured, or free, variance and covariance between intercept and random slope). The fixed part has a common slope for each subject. If the DV is a long column of data with all the time 1 data, then all the time 2 data, etc., and the Xs contain time contrasts (like the linear contrast -3 -1 1 3 for, say, four times), then we can extend the simple repeated measures ANOVA by allowing each subject to have a unique slope and intercept. In this case the simple \( Y_{ij} = \mu + \alpha_i + \pi_j + \epsilon_{ij} \) model adds one more term to allow for a random slope per subject as well as the random intercept.

We can estimate the variability of, say, the slope. In more complex models we can predict the variability or use the variability to predict other data.

The latent growth model can be represented in a standard SEM diagram like this:
with the unit contrast (1, 1, 1, 1) defining the latent variable called intercept and the contrast (-3,-1,1,3) defining the latent variable called slope. This example has four observed times on the same variable, labeled T1 to T4. By setting the factor paths to specific contrast values, the latent variables are defined just like contrast values in ANOVA. Each subject is assigned a contrast score (we called them “I hat” last semester) separately for intercept and for slope. Thus, each subject has his/her own slope and intercept. This SEM representation will accomplish the same thing as the MIXED model above. The traditional latent growth curve model sets all error variances across time to be equal, but the SEM framework can estimate a more general model with separate error variances over time and more complicated error structures.

All of these formulations—repeated measures, regression, SEM, multilevel model—will yield the identical results if they are each run in a particular way and there are no missing data. But the reason for the more complicated models is that they provide more flexibility, such as each subject has his or her own slope or better ways of handling missing data such as full-information maximum likelihood (FIML).

We end the course with the most difficult topic of the year: SEM. I outlined how we can use SEM to run all the designs we’ve covered this year (except for tree structures). SEM can even be used to create new analyses such as PCA with error, or regression with latent variables, or ANOVA on PCA-like factors, etc. SEM also makes a connection to a relatively complicated technique, multilevel modeling, which all ties back to the most elementary of all statistical tests—the paired t test, which is just the basic ANOVA model with a grand mean, a fixed effect $\alpha$ for the two times, a random effect $\pi$ for subject variability and an error term $\epsilon$.

Hope you enjoyed the journey. May all your results be statistically significant!
Appendix 1: Amos

Amos is a program that runs SEM. It is common at the University of Michigan because there is a site license for it. SPSS bought AMOS a few years ago, so the interface between AMOS and SPSS is relatively seamless.

There are plenty of tutorials for AMOS on the web. Most rely on the graphical interface where the user enters circles, squares, arrows and double headed arrows. That is fine for the simplest of models, but for more realistic models the syntax is more efficient. The syntax is a little cumbersome (written in visual basic style), but it can do most things. Unfortunately, some of Amos’ bells and whistles are not available through syntax file.

Appendix 2: Running SEM in SAS

I used to use SAS for teaching SEM but I’m migrating over to AMOS. I’ll still leave the syntax description in case there are some SAS users out there who are interested.

SAS has an SEM procedure called CALIS and a built in canonical correlation program. CALIS allows you to use different interfaces such as EQS, LISREL, RAM, and COSAN, essentially these four programs are rolled up into one. SAS is available on some LSA PCs. I mention SAS here in case any of come across this program in your research. The SAS SEM procedure is not as general as other programs such as LISREL, EQS, and AMOS, which allow multiple group analysis—the SAS SEM procedure just performs single group analyses.

Running SAS

Much like SPSS syntax it is possible to put all SAS syntax into a text file and run the syntax file within SAS.

SAS commands

I don’t want to get into all the details of SAS. Here I will just supply the information you will need to run the homework. I am not a SAS user so I doubt I could help much with any deviation you might want to try. Of course, you are welcome to play and learn.

At the top of the data file enter the following commands:

```
DATA givethisname;
TITLE 'data file for homework 5';
INPUT sub x1A x1B x2A x2B;
CARDS;
[first line of data goes here]
[second line of data goes here]
etc
```
So, you need to put four lines of command at the top of datafile and two lines of commands at the bottom of data file. Later, I’ll show you how to keep the data file and the command file separate.

DATA givethisname;
TITLE ‘data file for homework 5’;
INPUT sub x1A x1B x2A x2B;
CARDS;
[first line of data goes here]
[second line of data goes here]
etc
[last line of data goes here]
;
RUN;

DATA just names the current set of data that is being used. In SAS you can read in several data sets in the same run and do analyses on several different data sets. This is convenient if you have your data entered in many different files.

TITLE is a string that appears on top of every page. This is helpful for keeping track of output.

INPUT defines the variable names (and implicitly the number of variables).

CARDS informs SAS that data follows.

Note that the semicolons are important.

Running CALIS

The general syntax for the structural equations procedure CALIS is

TITLE ‘PUT A DESCRIPTIVE TITLE HERE’;
PROC CALIS COV METHOD = ML DATA=givethisname RESIDUALS EDF;
LINEQS
[structural equations go here];

STD
[relevant variances to be estimated or constrained go here];

COV
[relevant covariances to be estimated or constrained go here];

RUN;
The above can go on a newline right after any “RUN;” command. So, you could string one PROC CALIS command right after another because CALIS ends with “RUN;”.

TITLE is helpful especially when you string together several CALIS runs in one batch.

PROC CALIS is always required.

COV tells the program to use a covariance matrix. When COV is omitted CALIS will work from the correlation matrix by default. Covariances lead to raw regression coefficients; correlation matrices lead to standardized regression coefficients. There are many reasons why one should get in the habit of using a covariance matrix (see Cudeck, 1989).

METHOD = ML instructs CALIS to use a maximum likelihood estimation procedure. Sometimes we’ll use METHOD = GLS (which stands for generalized least squares).

DATA=givethisaname tells CALIS which datafile to use. This must be a data that has already been read in with the DATA INPUT CARDS sequence mentioned above (or read in from a file or entered directly into sas manually–I will not discuss these alternatives here).

RESIDUAL prints the residual covariance matrix (observed matrix minus model-implied matrix).

EDF is an optional argument necessary when you enter the covariance or correlation matrix rather than the raw data. EDF stands for effective degrees of freedom and it is equal to N - 1, where N is the sample size. Usage is: EDF=156 (representing 157 subjects).

LINEQS is required. This tells CALIS which interface to use.

I will describe the structural model, the terms that go in STD and the terms that go in COV through the examples below. In some problems, as you will see below, STD or COV may not be required.

You might also find the residual matrix helpful to understand where the model is failing. To get the residual matrix in the output just add “residual” on the PROC CALIS line.

CALIS conventions and other misc things

Use F* for factors, E* for errors on observed variables, and D* for errors on latent variables (disturbances). The “*” denotes a number such as F1 for factor 1 or F2 for factor 2.

Every F*, E*, and D* parameter must be listed in the STD section with either a fixed value or a free value. There are shortcuts to save on typing but I won’t bother teaching them because the homework assignments will contain easy models.

I kept the output to a minimum in this handout. You might want to try entering the subcommand RESIDUAL (anywhere between PROC CALIS and the first semi-colon). This prints the residual variance-covariance matrix from which everything is based on. If you are daring you could include the subcommand ALL (again anywhere between PROC CALIS and the first semi-colon). The subcommand ALL will give you more output than you could every imagine....
CALIS will complain if it spots an identification problem. It will even give you some hints on how to correct it (i.e., it spots which parameters are linearly related).

Canonical Correlation

SAS has a built in procedure for performing canonical correlations. This will come in handy when you don’t want to tinker with SPSS’s macro or do those iterated regressions we talked about earlier this semester. Here is the SAS syntax:

After any “RUN;” command enter

```
TITLE 'CANCOR ANALYSIS';
PROC CANCORR DATA=rg1 ;
VAR x1A x1B;
WITH x2A x2B;
RUN;
```

You can have as many variables as you want after the VAR (i.e., set1 variables) and after the WITH (i.e., set2 variables).

The raw canonical coefficients are the weights that can be used to created the weighted sum. The correlation between these two weighted sums will be identical to the canonical correlation printed in the output. SAS also prints out an adjusted canonical correlation, which is analogous to the adjusted $R^2$ in regression.
Appendix 3: R

Cronbach’s α

There are two packages (coefficientalpha and psy) that offer versions of Cronbach’s α.

SEM

My preferred package for running SEM in R is lavaan. The syntax is very similar to Mplus so it is easy to go back and forth. There is an older package called sem that is pretty good too. For those who want to get very involved in SEM possibly even doing your own simulations and writing more advanced code to develop new methods, then the package openMX would be relevant.

Mediation tests can be worked into SEM programs. The lavaan package has a nice way of testing mediation including bootstrapping the indirect effect as a product term. Here is an example from the lavaan package documentation:

```r
set.seed(1234)
X <- rnorm(100)
M <- 0.5*X + rnorm(100)
Y <- 0.7*M + rnorm(100)
Data <- data.frame(X = X, Y = Y, M = M)
model <- ' # direct effect
    Y ~ c*X
    # mediator
    M ~ a*X
    Y ~ b*M
    # indirect effect (a*b)
    ab := a*b
    # total effect
    total := c + (a*b)
'
fit <- sem(model, data = Data)
summary(fit)
```

Instrumental variables

The sem package has a function to compute instrumental variables analysis.