Class Introduction and Overview; Review of Regression and Measurement; Introduction to R and the lavaan Package

PSQF 6249: Structural Equation Modeling
Lecture #1 – August 27, 2025

Homework using Your Own Data

- HW 1, 3, and 5 will require individual-specific item-level data
  - At least 6 items thought to measure one latent trait or 8 items thought to measure 2 latent traits
  - Any response format (ordinal, slider, RT, etc)
  - No sample size requirement, but ideally >100 respondents
  - Preferable source: data from your own research area that you care about and want to do something with anyway
  - Otherwise: any publicly-available data you can find, such as through the International Personality Item Pool; archives at ICPSR, Berkley, or Harvard; Healthy Minds Data; or Early Childhood Data

Our Other Responsibilities

- My job (besides providing materials and assignments):
  - **Answer questions** via email, in individual meetings, or in group-based zoom office hours—you can each work on homework during office hours and get (near) immediate assistance (and then keep working)
    - Email me first (but you can follow up with the TAs if they help you)

- Your job (in descending order of timely importance):
  - **Ask questions**—preferably in class, but any time is better than none
  - **Frequently review** the class material, focusing on mastering the vocabulary (words and symbols), logic, and procedural skills
  - Don't wait until the last minute to start homework, and don't be afraid to **ask for help if you get stuck** on one thing for more than 15 minutes
    - Please email me a screenshot of your code+error so I can respond easily
  - **Do the readings** for a broader perspective and additional examples (best after lecture; readings are for the whole unit, not just that day)
  - **Practice** using the software to implement the techniques you are learning **on data you care about**—this will help you so much more!

Class-Sponsored Statistical Software

- I will show examples using ***Mplus*** (currently v. 8.10)
  - Mplus is expensive to purchase, but it is available for free to course participants through the UIowa Virtual Desktop
  - Mplus is expensive to purchase, but it is available for free to course participants through the UIowa Virtual Desktop
  - Also, Mplus syntax is (relatively) easy to follow and replicate
  - But…I will mostly us R (and the lavaan package)

- That being said, M*plus* is not the only option:
  - R program lavaan can estimate some of the models covered, and can be used for some homework (canned or your own data)
  - STATA SEM or GSEM can be used to analyze your own data, but cannot be used for all homework (different missing data routines)
  - SAS CALIS and SPSS AMOS can only do models for continuous responses (as far as I know), so these won't work for our purposes

Today's Class

- Introduction and overview of the course
  - Syllabus information

- Review of prerequisites needed in this course
  - Linear Models: Analysis of Variance (ANOVA)/Regression
  - Concepts from Construct Measurement
  - Why use SEM?

- Introduction to R
  - How R and R Studio work
  - The lavaan package
  - How to use syntax starter files for class
  - Where to put your syntax for homework/project/your own analyses

Key Questions for Today's Lecture

1. What is Structural Equation Modeling (SEM)?
2. How does SEMs differ from linear models?
3. What other models/methods are SEMs related to?
4. Why use SEM?

Today's Data Set

- To introduce and motivate SEM, and to review some prerequisites, we will make use of an example data set

- Data come from a (simulated) sample of 150 participants who provided self-reports of a happiness scale and their marital status

- Participant responded one survey:
  - 5-item happiness survey (each item used roughly a 5-point Likert scale)
  - 1-item marital status question (are you married? Yes/No)

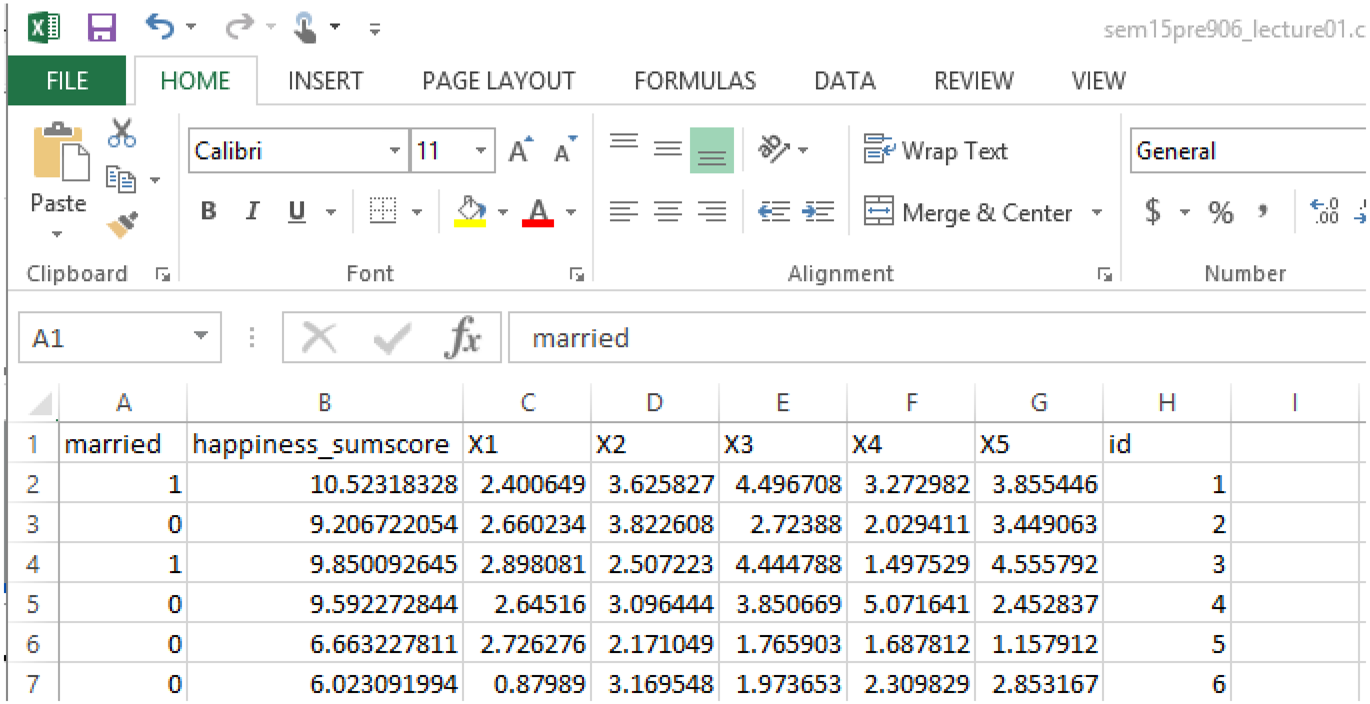
- The researchers were interested in the effects of marital status on happiness

Storing Data: CSV Files

- Although the data for today are simulated from within R and this step is unnecessary, I saved the data to a file (available online) to show you how to export data from R and how to import data into R
  - Note the variables are not integers (more on that later in the semester)
  - More on data import/export in R later in this lecture
- The file "sem15psqf6249_lecture01.csv" is a comma-delimited file which has the following characteristics:
  - It is a plain ASCII text file (can be opened in any text editing program)
  - Each row represents one observation
  - Each variable is separated by commas
- I prefer comma-delimited files as Microsoft Excel opens them by default when double clicked

Comma-Delimited File Example: sem15pre906_lecture01.csv

- In Excel:



	A	B	C	D	E	F	G	H	I
1	married	happiness_sumscore	X1	X2	X3	X4	X5	id	
2	1	10.52318328	2.400649	3.625827	4.496708	3.272982	3.855446	1	
3	0	9.206722054	2.660234	3.822608	2.72388	2.029411	3.449063	2	
4	1	9.850092645	2.898081	2.507223	4.444788	1.497529	4.555792	3	
5	0	9.592272844	2.64516	3.096444	3.850669	5.071641	2.452837	4	
6	0	6.663227811	2.726276	2.171049	1.765903	1.687812	1.157912	5	
7	0	6.023091994	0.87989	3.169548	1.973653	2.309829	2.853167	6	

- In textpad:

```
"married","happiness_sumscore","X1","X2","X3","X4","X5","id"
1,10.5231832834812,2.40064882043704,3.62582658284954,4.49670788019466,3.27298246417364,3.85544574026294,"1"
0,9.20672205429465,2.6602340824522,3.82260804121246,2.72387993062999,2.02941130257217,3.44906263758752,"2"
1,9.85009264509685,2.89808122777778,2.50722330594268,4.4447881113764,1.49752920034449,4.55579165083162,"3"
0,9.5922728439898,2.64516042325382,3.09644389873738,3.8506685219986,5.07164072684963,2.45283711008061,"4"
0,6.6632278106384,2.72627566251562,2.17104926157036,1.76590288655242,1.68781204967547,1.15791209450522,"5"
0,6.02309199433131,0.8798902461844,3.16954830961018,1.97365343853673,2.30982884500968,2.85316716317116,"6"
1,7.20640323945185,4.17363897151656,1.15698176288563,1.87578250504966,1.83130191328153,2.23226510144894,"7"
1,10.3215420501241,2.61862547732736,3.69403431314022,4.00888225965652,2.88186735297125,2.01380963101838,"8"
0,7.8586950140758,1.42293485988317,2.68884865283425,3.74691150135837,4.20570686297511,4.34466595635643,"9"
1,11.9138068119104,3.7389620438709,2.77281993294049,5.40202483509902,3.47578933304646,2.73946571794411,"10"
1,5.533332203772,1.68690347148075,2.70433916695298,1.14208956533828,4.03228411696483,4.40844671299883,"11"
```

THE GENERAL LINEAR MODEL

The General Linear Model

- The general linear model incorporates many different labels of analyses under one unifying umbrella:

	Categorical Xs	Continuous Xs	Both Types of Xs
Univariate Y	ANOVA	Regression	ANCOVA
Multivariate Ys	MANOVA	Multivariate Regression	MANCOVA

- The typical assumption is that error is normally distributed – meaning that the data are **conditionally** normally distributed

- Models for non-normal outcomes (e.g., dichotomous, categorical, count) fall under the *Generalized* Linear Model, of which the GLM is a special case (i.e., for when model residuals can be assumed to be normally distributed)

General Linear Models: Conditional Normality

$$Y_p = \boxed{\beta_0 + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p} + \boxed{e_p}$$

- **Model for the Means (Predicted Values):**
  - Each person's expected (predicted) outcome is a function of his/her values on x and z (and their interaction)
  - y, x, and z are each measured only once per person (*p* subscript)

- **Model for the Variance:**
  - $e_p \sim N(0, \sigma_e^2) \rightarrow$ **ONE** residual (unexplained) deviation
  - e_p has a mean of 0 with some estimated constant variance σ_e^2, is normally distributed, is unrelated to x and z, and is unrelated across people (across all observations, just people here)

Building a Linear Model for Predicting a Person's Happiness

- We will now build a linear model for predicting a person's happiness, using marital status as the predictor
- We will use the total score of the 5-item happiness scale as our dependent variable (happiness_sumscore)
- We will code marital status with a 1 if a person is married and a 0 if they are not married
  - Called reference or dummy coding
- Our beginning model is that of an **empty model** – no predictors for happiness (an **unconditional model**)
- Our ending model adds the marital status predictor

Model 0: The Empty Model

- Linear model: $Y_p = \beta_0 + e_p$

  where $e_p \sim N(0, \sigma_e^2)$

- Estimated Parameters (from lavaan output on html page):
  - $\beta_0 = 8.813\ (0.154)$
  - $\sigma_e^2 = 3.535\ (0.408)$

```
##
##                          Estimate  Std.err  Z-value  P(>|z|)
## Intercepts:
##     happnss_smscr           8.813    0.154   57.408    0.000
##
## Variances:
##     happnss_smscr           3.535    0.408    8.660    0.000
```

Notes on Empty Models

- As there are no predictors in the model, each person's predicted value is given by the intercept

- The intercept happens to be equal to the overall mean of the dependent variable:

```
#compare output to sample statistics:
```

```
mean(data02$happiness_sumscore)
```

```
## [1] 8.813357
```

- At a glance, the estimated residual variance is not quite equal to the sample variance:

```
#variance is not quite (lavaan is ML (divides by N)/var() function is unbiased (divides by N-1))
var(data02$happiness_sumscore)
```

```
## [1] 3.559059
```

Re-examining the Concept of Variance

- Variability is a central concept in advanced statistics
  - In multivariate statistics and SEM, covariance is also central

- Two formulas for the variance (about equal when N is big):

$$S_Y^2 = \frac{1}{N-1}\sum_{p=1}^{N}\left(Y_p - \bar{Y}\right)^2$$

Unbiased or "sample"

$$S_Y^2 = \frac{1}{N}\sum_{p=1}^{N}\left(Y_p - \bar{Y}\right)^2$$

Biased/ML or "population"

Here: p = person; 1 = variable number one

- Lavaan uses maximum likelihood to estimate parameters, so it's variance is the ML version

Variance/SD as a More General Statistical Concept

- Variance (and the standard deviation) is a concept that is applied across statistics – not just for data
  - Statistical parameters have variance
  - e.g. The sample mean $\bar{Y}$ has a "standard error" (SE) of $S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}}$
- The standard error is another name for standard deviation
  - So "standard error of the mean" is equivalent to "standard deviation of the mean"
  - Usually "error" refers to parameters; "deviation" refers to data
  - Variance of the mean would be $S_{\bar{Y}}^2 = \frac{S_Y}{N}^2$
- More generally, variance = error
  - You can think about the SE of the mean as telling you how far off the mean is for describing the data

Lavaan's Variance Estimate, Revisited

- Recall the variance estimate

```
## Variances:
##      happnss_smscr     3.535     0.408     8.660     0.000
```

- If we take the sample (N-1) variance, multiply by N-1, then divide by N, we will get this estimated ML value

```
#variance is not quite (lavaan is ML (divides by N)/var() function is unbiased (divides by N-1))
var(data02$happiness_sumscore)
```

```
## [1] 3.559059
```

```
#convert unbiased (N-1) variance to N version and get the same result:
(var(data02$happiness_sumscore)*(dim(data02)[1]-1))/(dim(data02)[1])
```

```
## [1] 3.535332
```

Model 1: Predicting Happiness from Marital Status

- Linear model: $Y_p = \beta_0 + \beta_1 Married_p + e_p$

    where $e_p \sim N(0, \sigma_e^2)$

- Estimated Parameters (from lavaan output on html page):

    - $\beta_0 = 8.471\ (0.205)$
    - $\beta_1 = 0.744\ (0.302)$
    - $\sigma_e^2 = 3.398\ (0.392)$

```
##                        Estimate  Std.err  Z-value  P(>|z|)
## Regressions:
##   happiness_sumscore ~
##     married               0.744    0.302    2.462    0.014
##
## Intercepts:
##     happnss_smscr         8.471    0.205   41.360    0.000
##
## Variances:
##     happnss_smscr         3.398    0.392    8.660    0.000
```

Model 1: Parameter Interpretation

- Linear model: $Y_p = \beta_0 + \beta_1 Married_p + e_p$

  where $e_p \sim N(0, \sigma_e^2)$

- Marital status was coded 0/1 using what is called reference or dummy coding:
  - Intercept becomes mean of the “reference” group (the 0 group)
  - Slopes become the difference in the means between reference and non-reference groups
  - For C categories, C-1 predictors are created

- $\beta_0 = 8.471\ (0.205)$
  - Predicted value of happiness when all predictors are equal to zero
  - Mean happiness value for people who are not married ($Married_p = 0$)

Model 1: Parameter Interpretation

- Linear model: $Y_p = \beta_0 + \beta_1 Married_p + e_p$

  where $e_p \sim N(0, \sigma_e^2)$

- Marital status was coded 0/1 using what is called reference or dummy coding:
  - Intercept becomes mean of the "reference" group (the 0 group)
  - Slopes become the difference in the means between reference and non-reference groups
  - For C categories, C-1 predictors are created

- $\beta_1 = 0.744\ (0.302)$
  - Change in predicted value of happiness for one-unit change in $Married$
  - Because $Married$ is a coded variable, this slope is the **difference** in happiness between those who are married ($Married_p = 1$) and those who are not ($Married_p = 0$)
  - Married people report higher happiness (.744 units higher)

More on Categorical Predictors

- Marital status was coded as 1=married and 0=non-married

- What about the opposite?

- **<u>All coding choices can be recovered from the model:</u>**
  - Predicted happiness for married persons (mean happiness for married=1):

$$Y_p = \beta_0 + \beta_1 = 8.471 + .744 = 9.215$$

  - Predicted happiness for non-married persons:

$$Y_p = \beta_0 = 8.471$$

- What would β_0 and β_1 be if we coded NonMarried= 1?

Hypothesis Tests for Parameters

- To determine if the regression slope is significantly different from zero, we must use a hypothesis test:

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

- From your previous statistics courses you may have found this an ANOVA table: sums of squares – F-test

- Because we are using ML, however, ANOVA tables aren't directly given

- Instead, we can use what is called a "Wald" test:

$$Z = \frac{\beta_1}{se(\beta_1)} = \frac{.744}{.302} = 2.462$$

```
##                      Estimate  Std.err  Z-value  P(>|z|)
## Regressions:
##   happiness_sumscore ~
##     married             0.744    0.302    2.462    0.014
##
```

Finally, Interpreting the Residual Variance

- There appears to be a significant difference in happiness between married and non-married people
  - We can now describe how much variance marital status explained by comparing our empty model and predictor model variance estimates

- From the empty model (Model 0): $\sigma_e^2 = 3.535\ (0.408)$

- From the predictor model (Model 1): $\sigma_e^2 = 3.398\ (0.392)$

- We can calculate the proportion reduction in variance (called R^2):

$$R^2 = \frac{3.535 - 3.398}{3.535} = .039$$

Standardized Coefficients

- The scale of the (unstandardized) slope coefficients is given in terms of UNITS of Y (SD Y) per UNITS of X (SD X)
  - Y goes up β_1 UNITS of Y for every 1 UNIT of X
    - Happiness (Y) goes up .744 units (where 1 unit is SD = $\sqrt{3.535} = 1.88$ score points) for every 1 unit of marital status (where 1 unit is SD = .497)*
      - *Not typically computed like this for coded variables
    - If the UNITS of X differ for the various IVs in a model, it can be hard to compare relative strengths of coefficients

- Standardized coefficients are the coefficients that would be obtained if Y and X were standardized:
  - Standardized = variance of 1 (i.e. z-scores used for analysis)

- Standardized coefficients are useful for comparing the relative effects of each IV in the model

Computing our Standardized Coefficients

- Our standardized estimates are:

```
#display standardized parameter output
standardizedSolution(model01.fit)
```

```
##                 lhs op               rhs est.std    se     z pvalue
## 1 happiness_sumscore  ~           married   0.197 0.080 2.462  0.014
## 2 happiness_sumscore ~~ happiness_sumscore   0.961 0.111 8.660  0.000
```

- Here, the interpretation is that married people are .197 **standard deviations** higher in happiness than are non-married people (note variance is $1\text{-}R^2$)
  - We use standardized estimates frequently in SEM

- We get this by (SDs are from previous slide):

$$b_1 = \beta_1 \frac{SD(X)}{SD(Y)} = 0.744 \frac{0.498}{1.88} = .197$$

REVIEW OF MEASUREMENT (RELIABILITY)

Measured Constructs, Reliability, and the GLM

- The general linear model (regression/ANOVA combo) assumed that **<u>all</u>** variables were measured without error
- Surveys, questionnaires, and tests are all psychological instruments that are imprecise
  - Adding up scores is a very naïve statistical model – so the scores all have measurement error
- Related to the measurement error of a test score is the reliability of a test
  - The GLM assumes Reliability = 1 for all variables
- In our example so far, our happiness variable was the sum score for a 5-item happiness scale

Latent Traits Need Test Theory

- "**Test theory**" is an abbreviated expression for:
  - "Theory of Psychological Tests and Measurements"
  - Or "Psychometric Theory" (even when not used in Psychology)

- Test theory is a general collection of **statistical models** for evaluating the development and use of instruments
  - **Operationalize** practical problems in measurement
  - **Provide answers to** practical problems in measurement
  - So yes, measurement models are indeed statistical models!

- 3 branches of measurement models for latent traits that are inter-related… you likely know one of these already

Classical Test Theory (CTT)

- What you first learned about measurement probably falls under the category of Classical Test Theory (CTT):
  - Writing items and building scales (or “tests”)
  - Item analysis for differentiating “good” from “bad” items
  - Evaluating dimensionality underlying the items
  - Interpretating scale or test “scores”
  - Evaluating reliability and construct validity

- Big picture: We will view CTT as a model with a restrictive set of assumptions within a more general family of latent trait measurement models

What is a ‘latent trait’?

- **Latent trait** = Unobservable construct (“factor”)
  - Many types of variables: ability, attitude, tendency, etc.
  - e.g., “Intelligence”, “Extroversion”, “Depression”

- But how can we measure something unobservable?
  - Build **measurement models** by which to represent them!

- Big picture: Latent traits can be measured using observed responses → “**items**” or “**indicators**”
  - A new latent variable is created from the common variance across indicators thought to measure the same construct
  - But not all constructs should use latent trait measurement models! (e.g., formative vs. **reflective indicators**)

Differences Among Latent Trait Measurement Models (LTMMs)

- What do we call the latent trait measured by the indicators?
  - Classical Test Theory (CTT) → "True Score" (T)
  - Confirmatory Factor Analysis (CFA) → "Factor Score" (F)
  - Item Factor Analysis (IFA) → "Factor Score" (F)
  - Item Response Theory (IRT) → "Theta" (θ)

- Fundamental difference in approach:
  - CTT → unit of analysis is the WHOLE TEST (item sum or mean)
    - Sum = latent trait, so items and persons are inherently tied together → bad
    - Only using the sum requires restrictive assumptions about the items
  - CFA, IFA, IRT, and other LTMMs → unit of analysis is the ITEM
    - Model of how item response relates to a separately estimated latent trait
    - Provides way of separating item and person properties → good for flexibility
    - Different names of models are used for differing item response formats
    - Provides a framework for testing adequacy of measurement models

Latent Trait Measurement Models (LTMMs)

- Families of latent trait measurement models are labeled differently based on their indicators' response format:
  - Continuous responses? → Confirmatory Factor Models
  - Categorical responses? → Item Response Theory or Item Factor Models
  - Measurement models for other response types exist too (like counts), but they don't necessarily have special names (I say "generalized")

- Other relevant, related terms:
  - "Structural Equation Modeling" (SEM) is correlation or regression among the latent traits defined by the measurement models
    - Things that can go wrong in SEM most often reflect problems with the measurement models—that is why we spend most of the semester on this!
  - "Path Analysis" is just regression among observed variables only
  - "Mediation" is just regression with a better marketing campaign
  - "Moderation" is an interaction term with a better marking campaign

A Brief History of Test Theory…

- Motivated by problems in education and psychology
  - Education → Assessment of academic abilities
  - Psychology → Understand structure of intelligence or personality
  - Piecemeal approach; also barriers from technical presentation
  - Theories developed before availability of computing power, so approximations were developed that could actually be used (with remnants that unfortunately still get used, like alpha and EFA)

- 1904: Charles Spearman published two seminal papers
  - One showed how to estimate amount of error in test scores
    - Led to classical true score theory (aka, classical test theory)
  - Other showed how to recognize from test data that the tests measure just one psychological attribute in common ("G")
    - Led to common factor theory (aka, confirmatory factor analysis)

Sum Score = Classical Test Theory: Basics of CTT

- In CTT, the ***test*** is the unit of analysis:

$$Y_{total} = T + e$$

  - **True score T:** best estimate of "construct"
  - **Error e:** mean of zero, uncorrelated with T
- Variance of test scores: $\sigma_Y^2 = \sigma_T^2 + \sigma_e^2$
- Goal is to quantify ***reliability*** :: proportion of test variance accounted for by true score variance:

$$\rho = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2}$$

- Items assumed exchangeable (they all count the same)
  - More items means higher reliability, regardless of type

Classical Test Theory, continued

- CTT unit of analysis is the TEST or SCALE (sum/mean of items)
  - Want to quantify how much of observed test score variance is due to "true score" variance versus "error" variance
  - "Error" is a unitary construct in CTT (and error is always bad)
  - Goal is then to reduce "error" variance as much as possible
    - Standardization of testing conditions (make confounds constants)
    - Aggregation → more items are better (errors should cancel out)
  - Items are exchangeable; item properties are NOT taken into account in indicating the latent trait of a given person (which is just the sum)

- Followed by *generalizability theory* to distinguish kinds of error
  - e.g., item variance, person variance, rater variance, occasion variance
  - Modern analog: mixed-effects (multilevel) models with crossed random effects for each (random) sampling dimension and their interactions

Classical Test Theory, continued

- Brief history of solutions for quantifying reliability:
  - 1904: Spearman: from alternate forms or test-retest
  - 1945: Guttman: from the relations between the items within a test (i.e., coefficient alpha)
  - 1951: Cronbach further developed Guttman's work → "Cronbach's alpha"
    - Called "Guttman-Cronbach alpha" by McDonald (and no one else)
    - Cronbach's work further elaborated into generalizability theory
    - And no, a good alpha doesn't mean anything—stay tuned for why!
  - 1950: Gulliksen classic text for CTT
    - See also Nunnally's texts from the 1970's–1990's

- More CTT specifics in upcoming classes…
- *Next, tracing the other contribution of Spearman…*

So…our 5 Happiness Items are Now the Focal Point

- The happiness sum score was just the sum of the 5 happiness items in our data set for each person:

$$Y_p = X_1 + X_2 + X_3 + X_4 + X_5 = \sum_{i=1}^{I} X_i$$

- This brings up discussions of multi-variable statistics:
  - Correlation and covariance

- The sample correlation matrix of these data are:

	x1	x2	x3	x4	x5
x1	1.00000000	-0.06354391	0.06019113	0.02107528	0.15239670
x2	-0.06354391	1.00000000	0.07555098	-0.02271474	0.03080322
x3	0.06019113	0.07555098	1.00000000	0.05662233	0.15168240
x4	0.02107528	-0.02271474	0.05662233	1.00000000	0.16500076
x5	0.15239670	0.03080322	0.15168240	0.16500076	1.00000000

Correlation of Variables

- The Pearson correlation is often used to describe the association between a pair of variables:

$$r_{Y_1,Y_2} = \frac{\frac{1}{N-1}\sum_{p=1}^{N}\left(Y_{1p} - \bar{Y}_1\right)\left(Y_{2p} - \bar{Y}_2\right)}{S_{Y_1} S_{Y_2}}$$

- The correlation is **unitless** as it ranges from -1 to 1 for continuous variables, regardless of their variances
    - Pearson correlation of binary/categorical variables with continuous variables is called a point-biserial (same formula)
    - Pearson correlation of binary/categorical variables with other binary/categorical variables has bounds within -1 and 1

Reliability Measured by Alpha

- For quantitative items (items with a scale – although used on categorical items), this is often indexed by Cronbach's Alpha…
  - Or 'Guttman-Cronbach alpha' (Guttman 1945 > Cronbach 1951)
  - Another reduced form of alpha for binary items: KR 20

- Alpha is described in multiple ways:
  - Is the mean of all possible split-half correlations
  - Is expected correlation with hypothetical alternative form of the same length
  - Is lower-bound estimate of reliability under assumption that all items are tau-equivalent (more about that later)
  - As an index of "internal consistency"
    - Although nothing about the index indicates consistency!

- Alpha, however, is calculated using covariance matrices instead of correlation matrices

Covariance of Variables: Association with Units

- The numerator of the correlation coefficient is the covariance of a pair of variables:

$$S_{Y_1,Y_2} = \frac{1}{N-1}\sum_{p=1}^{N}(Y_{1p} - \bar{Y}_1)(Y_{2p} - \bar{Y}_2)$$

Unbiased or "sample"

$$S_{Y_1,Y_2} = \frac{1}{N}\sum_{p=1}^{N}(Y_{1p} - \bar{Y}_1)(Y_{2p} - \bar{Y}_2)$$

Biased/ML or "population"

- The covariance uses the units of the original variables (but now they are multiples):

- The covariance of a variable with itself is the variance

- The covariance is often used in SEM and multivariate analyses because it ties directly into multivariate distributions
  - But…covariance and correlation are easy to switch between

Going from Covariance to Correlation

- If you have the covariance matrix (variances and covariances):

$$r_{Y_1,Y_2} = \frac{S_{Y_1,Y_2}}{S_{Y_1} S_{Y_2}}$$

- If you have the correlation matrix and the standard deviations:

$$S_{Y_1,Y_2} = r_{Y_1,Y_2} S_{Y_1} S_{Y_2}$$

Where Alpha Comes From

- First, we need the covariance matrix of all the items

	x1	x2	x3	x4	x5
x1	1.19557420	-0.07495136	0.06737788	0.02346632	0.16606757
x2	-0.07495136	1.16368298	0.08343611	-0.02495217	0.03311574
x3	0.06737788	0.08343611	1.04807695	0.05902929	0.15475788
x4	0.02346632	-0.02495217	0.05902929	1.03697078	0.16745194
x5	0.16606757	0.03311574	0.15475788	0.16745194	0.99321203

- The **sum of the item variances** is given by:

$$ItemVar(Y) = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 + \sigma_5^2 = 5.437$$

- The **variance of the sum of the items** is given by the sum of ALL the item variances and covariances:

$$TotalVar(Y) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sigma_{ij} = 6.747$$

Guttman-Cronbach's Alpha

$$\alpha_{GC} = \frac{I}{I-1} \cdot \frac{TotalVar(Y) - ItemVar(Y)}{TotalVar(Y)}$$

$$\alpha_{GC} = \frac{5}{5-1} \cdot \frac{6.747 - 5.437}{6.747} = 0.243$$

- Numerator reduces to just the covariance among items
  - Sum of the item variances…
    - Var(X) + Var(Y) = Var(X) + Var(Y) → just the item variances
  - Variance of total Y (the sum of the items)…
    - Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) → PLUS covariances
  - So, if the items are related to each other, the variance of the total Y item sum should be bigger than the sum of the item variances
    - How much bigger depends on how much covariance among the items – the primary index of relationship
- **Therefore – the happiness scale is not perfectly reliable**
  - Violates the assumption of the GLM/Regression

Confirmatory Factor Analysis (CFA) Models

- Main idea: Build a measurement model of which response indicators should "”go together” to measure the same thing
  - **CFA = Linear regression model** predicting each continuous observed outcomes (“indicators”) from **latent** trait (unobserved) predictor(s)

- Differs from exploratory factor analysis (that is NOT a model):
  - In CFA *you* impose the number and content of factors
  - In CFA alternative models are COMPARABLE and TESTABLE

- Uses of confirmatory factor analysis models:
  - Analyze relationships among indicators that have normal, continuous distributions (or “incorrectly” to analyze ordinal response indicators)
  - Provide separation of persons, items, and occasions (as in any LTMM)

Confirmatory Factor Analysis (CFA)

- **The CFA unit of analysis is the ITEM (as in any LTMM):**

  $\boldsymbol{y_{is} = \mu_i + \lambda_i F_s + e_{is}}$ → **both items AND subjects matter**

  Should look familiar…
  $y_{is} = \beta_{0i} + \beta_{1i} x_s + e_{is}$

  - Observed response for item *i* and subject *s*
    = intercept of item *i* (μ)
    + subject *s*'s latent trait/factor (F), item-weighted by λ
    + error (e) of item *i* and subject *s*

- **Dimensionality** → part of the model (usually 1 latent trait per item)
  - Local Independence → e residuals are independent after controlling for factor(s)
  - The factor is the reason why item responses were correlated in the first place!
  - If not, you can **augment the model** to address unintended multidimensionality

- **Linear model** → a one-unit change in latent trait/factor F_s creates same increase in the expected response y_{is} along all points of y_{is}
  - Won't work well for binary or ordinal data… thus, we need another LTMM

- Items can now differ from each other in how much they relate to the latent trait, *but a "good item" is assumed equally good for everybody!*

A Brief History of Common Factor Theory

- 1900’s: Spearman’s “G” single-factor models
  - Development of techniques designed to find a common factor
  - Led to development of other IQ tests (Stanford-Binet, Wechsler)

- 1930’s and 1940’s: Thurstone elaborated Spearman’s “G” unidimensional model into a “multiple factor” model
  - Beginnings of exploratory factor analysis to do so
  - Later applied in other personality tests (e.g., MMPI)

- 1940’s and 1950’s: Guttman’s work
  - Factor analysis and test development is about generalizing from measures we have created to more measures of the same kind
  - Thus, need to think about measurement structure before-hand

A Brief History of Common Factor Theory

- 1940s: Lawley → rigorous foundation for statistical treatment of common factor analysis
  - But had to wait for better computers to be able to do it!
- 1952: Lawley → beginnings of confirmatory factor model
  - Later extended by Howe and Bargmann (1950's)
  - Further extended by Jöreskog (the King of LISREL in1970's)
- But this linear model pry should not be applied to binary, ordinal, or other not-continuous responses…
  - Predicted response will go past possible response options
  - Errors can't be normally distributed with constant variance
- So then what? Item Response Theory to the rescue…
  - *aka*, LTMM for generalized response formats

Item Response Theory (IRT) Models

- IRT resulted from combination of ideas from factor analysis and phi-gamma law of psychophysics
  - When detecting stimuli of varying intensity (e.g., light), the response follows a smooth, S-shaped curve that can be represented by the cumulative normal distribution
  - That response function also works to model probability of a correct response given (1 to 4) model parameters

- 1950: Lazarsfeld: Introduced "latent structure analysis" → factor analysis for binary item responses
  - Beginnings of item response theory (which is not a theory per se, but another set of latent trait measurement models)

Item Response Theory (IRT) Models

- Linear regression is to confirmatory factor models as to:
  - Logistic regression is to binary IRT models
  - Ordinal/nominal regression is to “polytomous” IRT models
  - IRT = generalized linear model predicting each categorical observed outcome indicator from latent predictors using link functions
  - Term “IRT” usually goes with full-information estimation (use all data)

- A “Rasch model” is a restricted version of an IRT model (but don’t let any Rasch people hear you saying that)

- Uses of IRT models:
  - *Correctly* analyze categorical indicators (binary, ordinal, or nominal)
  - Examine sensitivity of measurement across range of latent trait
  - Provide separation of persons, items, and occasions (as in any LTMM)

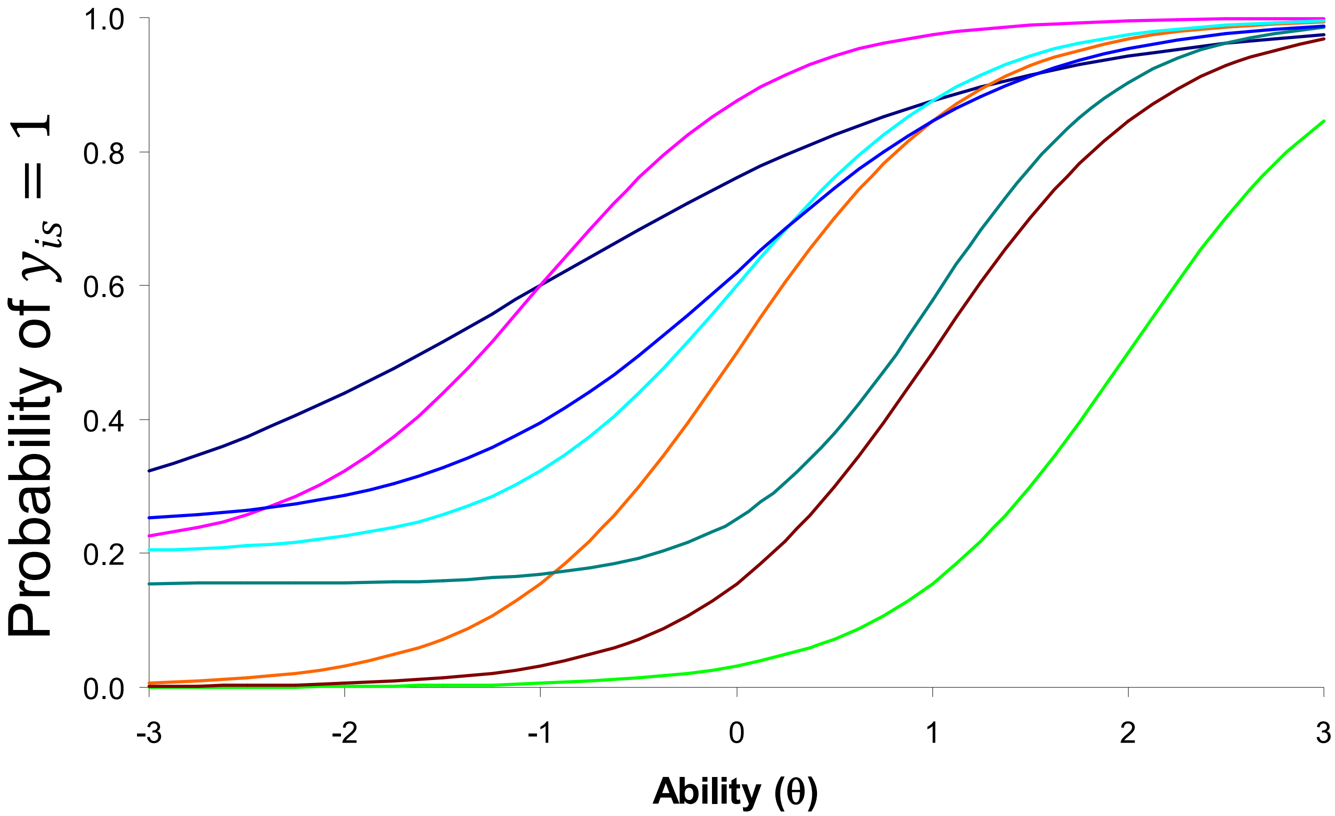
Item “Characteristic” Curves

a = Discrimination = slope of ‘line’

b = Difficulty = location of ‘line’

c = Lower Asymptote of ‘line’

d = Upper Asymptote of ‘line’



Note: Theta ability has a nonlinear relationship to probability of item response, but a linear relationship to its link-transformed mean…

Item Response Theory, continued

- The **IRT** unit of analysis is the individual **ITEM** (as in any LLTM)
  $\mathbf{Link}(\boldsymbol{y_{is}}) = \boldsymbol{a_i}(\boldsymbol{\theta_s} - \boldsymbol{b_i})$ → both items AND subjects matter
  - Response probability is predicted via a link (transformation) function (usually logit or probit, in which probit is called “ogive” in IRT)
  - Items and persons are located on the same latent metric
  - Probability of getting an item right depends (at least) on the subject’s ability ($\boldsymbol{\theta_s}$ **= “Theta”**) and the item’s difficulty ($\boldsymbol{b_i}$), weighted by its discrimination ($\boldsymbol{a_i}$, how related the item is to the latent trait)
  - **“Item factor analysis” (IFA)** re-arranges IRT model into something that looks more like CFA (and usually uses limited information estimation)

- All items are NOT created equal (not exchangeable)
  - Having items that differ in their properties is a GOOD THING, because you can customize tests for different groups or purposes
  - Reliability (“information”) varies across trait level, and depends specifically on how well the items’ difficulty matches subjects’ traits

Item Response Theory, continued

- 1952: Lord's seminal paper: Spearman's single-factor model can be applied to dichotomous items
  - Binary responses modeled by normal ogive function ("probit")
  - Later work used easier logit link instead (logit ≈ probit*1.7)
  - Elaborated in 1960s by Birnbaum (and others)

- 1968: Lord & Novick → first CTT text to also include IRT
  - Well-connected to emerging scholars in both educational testing and psychometric methods… and BOOM…

- 1960: Separate work by Rasch (common 'a' parameter)
  - Restricted IRT model, but with desirable properties if it fits…
  - … and a very different philosophical viewpoint (as "the" model)

A Unified View of Test Theory

- Classical test theory can be viewed as a restricted form of the common factor model, but the focus is the TEST…
  - Originated by Spearman, elaborated by Thurstone, formalized by Lawley, and made practical in software by Jöreskog
- Item response theory (and Rasch) models are common factor models used for binary or ordinal responses…
  - Developed by Lord, Birnbaum, Rasch, and their students
- Confirmatory factor analysis are common factor models for continuous responses…
  - Approximation for ordinal data with varying degrees of success
- Latent traits can also be indicated by other kinds of non-normal responses (count, zero-inflated, two-part/hurdle)….
  - But they don’t have special names (I’d call it “generalized SEM”)
  - Other response data (e.g., eye fixation, RT) can be used, too!

Advantages of LTMM Framework (CFA, IRT, IFA, and beyond)

- Explicit, testable models of dimensionality
- Concrete guidelines for selecting items to build scales
- Assess measurement sensitivity across range of latent trait (i.e., know where the “holes” of imprecision are)
- Provide comparability across persons, items (different forms scales or different scales), and occasions
- Examine comparability across groups or repeated measures
  - Confirmatory factor analysis → “Measurement invariance”
  - Item response theory → “Differential item functioning”
- Internal and external evidence for construct validity
- Generalized measurement models can even accommodate different response formats within the same instrument

Disadvantages of LTMM Framework

- Primary: Required sample size
  - Casts of 100s for sure, and preferably 1000s
    - Bayesian methods are much more flexible!
  - Uses maximum likelihood (limited-info WLSMV estimator in Mplus can also be used for multidimensional IRT models)
    - REML is not available for smaller samples (as it is in MLM software)

- Technical difficulties
  - Estimation is harder, especially in multidimensional IRT
  - References written in Greek (literally)
    - Except your textbook and selected readings, so please read them!

- Misnomers about what LTMM (within SEM) can do…
  - Bad items are still bad items, no matter what model is used
  - No, SEM is still not “causal” modeling

MOTIVATING STRUCTURAL EQUATION MODELING

Not 100% Reliable: What's the Big Deal?

- When the GLM (regression/ANOVA/all other forms) are used with variables that are not measured perfectly, several deleterious things can happen:
  - The effects may be biased
    - Depends on variable type (i.e., how continuous are your measures)
  - The standard errors may be biased
    - Happens any time
  - Therefore…hypothesis tests may not be accurate

- Basically, any conclusions you make can be drawn into question when you have not-100%-reliable variables used
  - The issue is with just how reliable is reliable enough

The Answer…Don't Use Aggregates – USE SEM

- Structural Equation Modeling seeks to determine the relationship between
  - Latent constructs only
    - Latent our example: Happiness
  - Latent and observed constructs
    - From our example: how does marital status factor into the model?
  - Complex relationships between latent constructs and observed variables
    - More variables needed for our example
    - Includes mediation models

- SEM is a generalization of linear modeling using observed and latent (sometimes called random) variables
  - I tend to think of SEM as a part of a bigger picture…you will see that SEM people think everything is part of SEM

Path Diagram of Our Regression Example

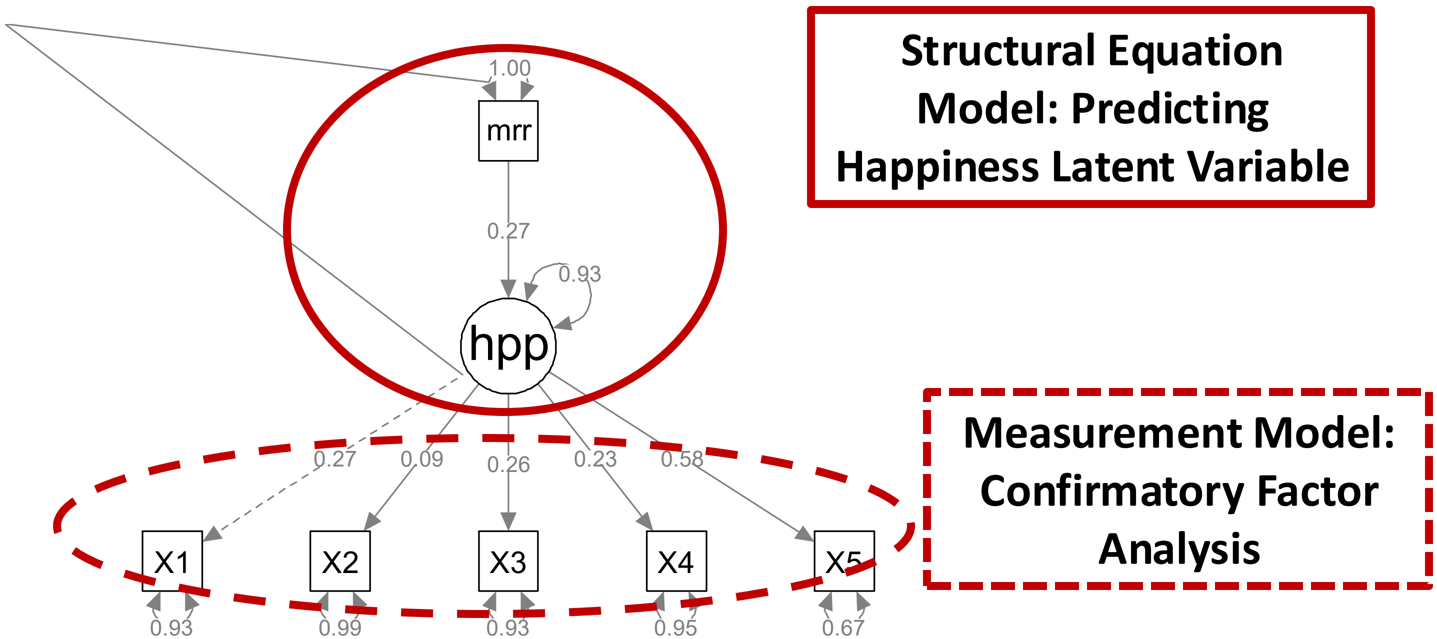
- A common way of depicting SEMs is with a path diagram → a pictorial representation of the statistical model
  - Observed variables: Squares
  - Latent variables: Circles
  - Direct effects: Arrows with one head
  - Indirect effects: Arrows with two heads

- From our previous GLM example

- Here MRR is marital status and hp_ is the happiness sum score

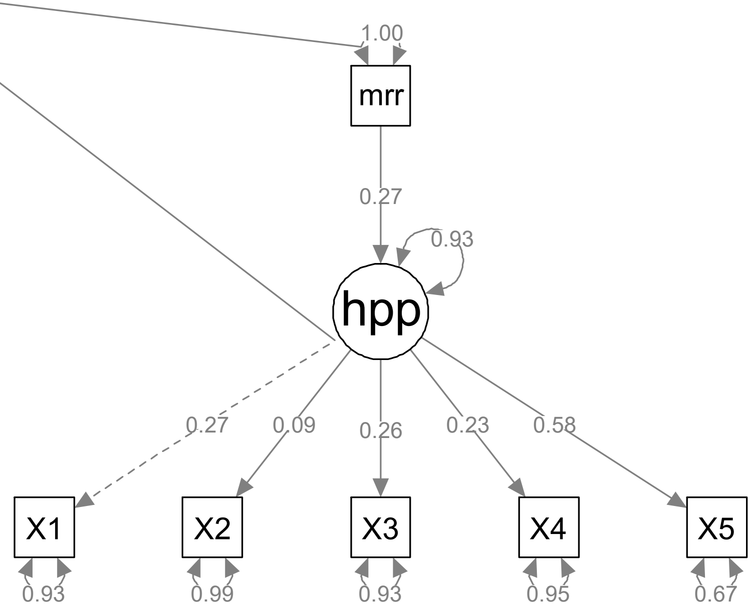
In SEM, We Don't Need a Sum Score

- Variables that are measured with error are represented as "latent" constructs in SEM
  - The latent variables are estimated directly by the model
  - Any equations involving latent variables are estimated simultaneously
- A more accurate depiction of our example:



Simultaneous Equations Implied by Path Diagram

SEM is often called Path Analysis with Latent Variables



$$X_{p1} = \mu_1 + \lambda_1 HPP_p + e_{p1}$$
$$X_{p2} = \mu_2 + \lambda_2 HPP_p + e_{p2}$$
$$X_{p3} = \mu_3 + \lambda_3 HPP_p + e_{p3}$$
$$X_{p4} = \mu_4 + \lambda_4 HPP_p + e_{p4}$$
$$X_{p5} = \mu_5 + \lambda_5 HPP_p + e_{p5}$$
$$HPP_p = \beta_0 + \beta_1 Married_p + e_p^{HPP}$$

Measurement Models

- Measurement models can be divided into families of models based on **response format alone:**
  - In most of this course: continuous variable responses measuring a latent construct:: **Confirmatory Factor Models**
  - Non-continuous variable responses→ item response theory (and other names)

- Both of these families fall under a larger framework: **Generalized Linear Latent and Mixed Models**
  - Provide measurement models for other types of responses

- Other relevant families we will be discussing:
  - **Structural Equation Models** :: provides estimates of correlations amongst latent variables in measurement models
  - **Path Analysis** :: simultaneous regression among observed variables

Comparing Results for GLM and SEM Analyses

- The results of these two analyses show how SEM can and should be used:

- Recall the standardized coefficient from our GLM analysis:

```
##                  lhs op              rhs est.std    se     z pvalue
## 1 happiness_sumscore  ~          married   0.197 0.080 2.462  0.014
```

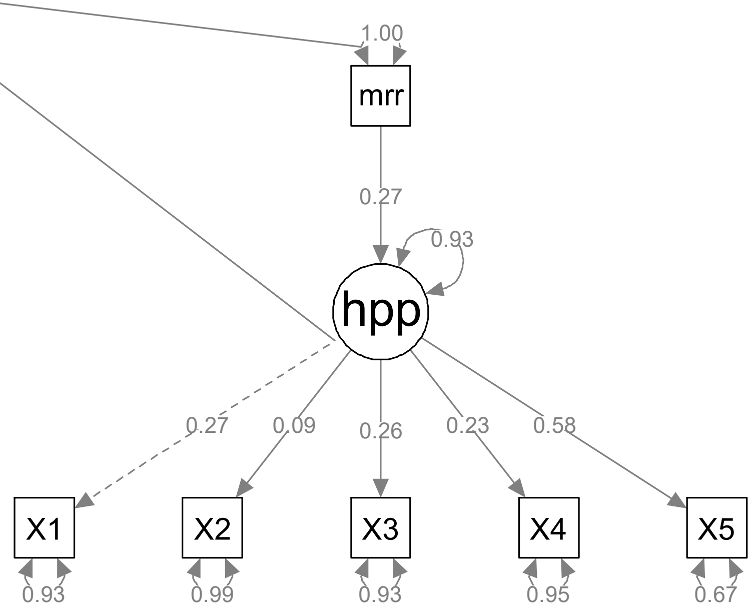
- Now here is the same coefficient from the SEM analysis:

```
##          lhs op       rhs est.std    se      z pvalue
## 6  happiness  ~   married   0.272 0.233  1.168  0.243
```

- The GLM β_1 is lower (biased due to unreliable Y)
- The GLM β_1 standard error is lower (result of assuming perfect reliability for Y: more power that you really have)
- The p-value is non-significant → the conclusion changes

Simultaneous Equations Implied by Path Diagram

SEM is often called Path Analysis with Latent Variables



$$X_{p1} = \mu_1 + \lambda_1 HPP_p + e_{p1}$$
$$X_{p2} = \mu_2 + \lambda_2 HPP_p + e_{p2}$$
$$X_{p3} = \mu_3 + \lambda_3 HPP_p + e_{p3}$$
$$X_{p4} = \mu_4 + \lambda_4 HPP_p + e_{p4}$$
$$X_{p5} = \mu_5 + \lambda_5 HPP_p + e_{p5}$$
$$HPP_p = \beta_0 + \beta_1 Married_p + e_p^{HPP}$$

The (Really) Big Picture

- Statistical distributions are what drive the process
  - Each distribution is described by a set of parameters
  - Think of the normal distribution (mean and variance)

- Each of the lines represents model parameters
  - The statistical distribution of the boxes and circles are described by the model parameters

- Model parameters provide constraints to the statistical distribution parameters
  - Reduce complexity of model
  - Provide for meaningful inference

- A model is bound by distributions assumed and, hence, the number of possible parameters
  - We will learn statistics and path models
    - Both are needed to be good at SEM

In-Class Video Demonstration

INTRODUCTION TO R AND R STUDIO

WRAPPING UP AND REFOCUSING

Key Questions for Today's Lecture

1. What is Structural Equation Modeling (SEM)?

2. How does SEMs differ from linear models?

3. What other models/methods are SEMs related to?

4. Why use SEM?

Wrapping Up

- Today we covered the structure of the course, a review of the prerequisites, and an introduction to R and R studio

- First Homework:
  - Assigned next week (Sept 3) → Due September 9th

- First Reading Assessment:
  - Next week (on Kaplan's Chapter 2)