Latent Variable Analysis

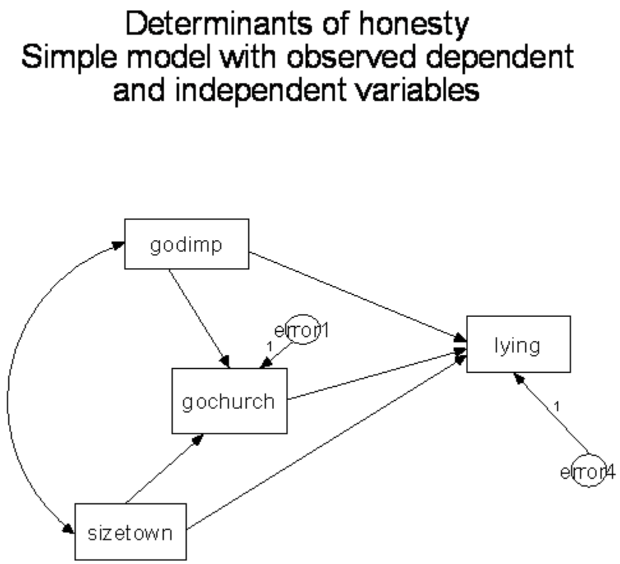
Path Analysis Recap

- **I. Path Diagram**
  - ***a. Exogeneous vs. Endogeneous Variables***
  - ***b. Dependent vs, Independent Variables***
  - ***c. Recursive vs. Non-Recursive Models***
- **II. Structural (Regression) Equations**
  - ***Normal Equations***
- **III. Estimating Path Coefficients**
- **IV. Identification**
  - ***a. Degrees of freedom***
  - ***b. Just Identified Models***
  - ***c. Overidentified Models***
  - ***d. Underidentified Models***

Path Analysis Recap

- **IV. Rules of decomposing the relationship between two variables**
- ***1. The components***
  - ***a. Direct effect***
    - path coefficient
  - ***Compound effects***
  - ***b. Indirect effect***
    - Start from the variable (Y) later in the causal chain to your right. Trace backwards (right to left) on arrows until you get to the other variable (X). You must always go against straight arrows (from arrow head to arrow tail).
  - ***c. Spurious effect (due to common causes)***
    - Start from variable Y. Trace backwards to a variable (Z) that has a direct or indirect effect on X. Move from Z to X.
  - ***d. Correlated (unanalyzed) effect***
    - It is like an indirect effect or a spurious effect due to common causes, except it includes one ,*and only a single one,* double headed arrow.
- ***2. Calculate compound paths by multiplying (path and/or correlation) coefficients encountered on the way***
  - ***Sewall Wright's rules***
  - **No loops**
    - Within one path you cannot go through the same variable twice.
  - **No going forward then backward**
    - Only common causes matter, common consequences (effects) don't.
  - **Maximum of one curved arrow per path**
- ***3. Add up all direct and compound effects***
  - **The sum is the total association**
    - In a just identified model the total association equals Pearson's correlation coefficient

Example: A just identified model



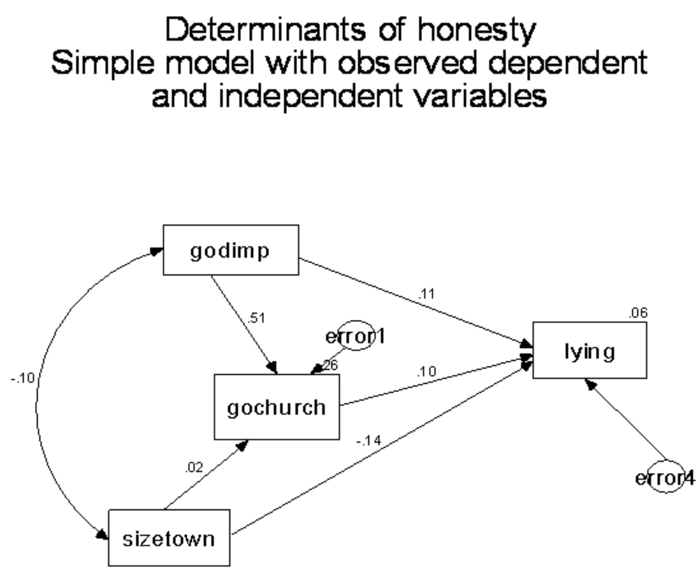
Correlations

		lying	sizetown	gochurch	godimp
lying	Pearson Correlation	1	-.158**	.160**	.175**
	Sig. (2-tailed)		.000	.000	.000
	N	1732	1732	1732	1732
sizetown	Pearson Correlation	-.158**	1	-.034	-.102**
	Sig. (2-tailed)	.000		.163	.000
	N	1732	1732	1732	1732
gochurch	Pearson Correlation	.160**	-.034	1	.508**
	Sig. (2-tailed)	.000	.163		.000
	N	1732	1732	1732	1732
godimp	Pearson Correlation	.175**	-.102**	.508**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	1732	1732	1732	1732

**. Correlation is significant at the 0.01 level (2-tailed).

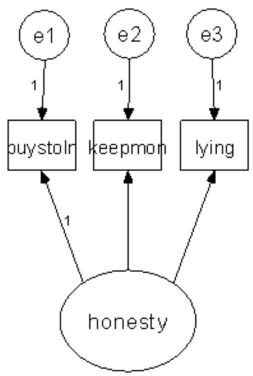
6 equations (correlations)
6 unknowns (5 paths and 1 correlation)

Standardized Estimates



Latent Variable and Its Indicators

Estimating the latent variable separately



Correlations

		buystoln	keepmon	lying
buystoln	Pearson Correlation	1	.276**	.371**
	Sig. (2-tailed)		.000	.000
	N	1732	1732	1732
keepmon	Pearson Correlation	.276**	1	.457**
	Sig. (2-tailed)	.000		.000
	N	1732	1732	1732
lying	Pearson Correlation	.371**	.457**	1
	Sig. (2-tailed)	.000	.000	
	N	1732	1732	1732

**. Correlation is significant at the 0.01 level (2-tailed).

The three observed variables are indicators of the latent variable Honesty which is a concept. They are effect indicators because they are the effects of the latent variable.

3 equations (correlations)
3 unknowns (paths)

Structural Equations:

(1) $B=p_{bh}*H+e1$
(2) $K=p_{kh}*H+e2$
(3) $L=p_{lh}*H+e3$

Normal Equations:

If we just multiply each equation by its independent variable we will not get anywhere. Take the 1st equation:

$r_{bh}= p_{bh} *r_{hh}+r_{he1}$ $r_{hh}=1$ and $r_{he1}=0$ so $r_{bh}= p_{bh}$ but what is r_{bh}?

So we must multiply each equation by the other two

(1) $B=p_{bh}*H+e1$ multiplied by (2) $K=p_{kh}*H+e2$

$B*K=(p_{bh}*H+e1)*(p_{kh}*H+e2)= p_{bh}*H*p_{kh+} *H+p_{bh}*H*e2 + p_{kh}*H*e1+ e1*e2$

Turn it into a normal equation

$r_{bk} = p_{bh}*p_{kh}* r_{hh} +p_{bh}*r_{he2}*+ p_{kh}*r_{he1} +r_{e1e2}$

because $r_{hh} =1$ and $r_{he2} =0$ and $r_{he1} =0$ and $r_{e1e2} =0$

$r_{bk} = p_{bh}*p_{kh}$ this also follows from the rules of decomposing relationship between two variables K and B are related only through their common cause of H

the same way we can calculate two other normal equations:

$r_{bl} = p_{bh}*p_{lh}$
$r_{lk} = p_{lh}*p_{kh}$

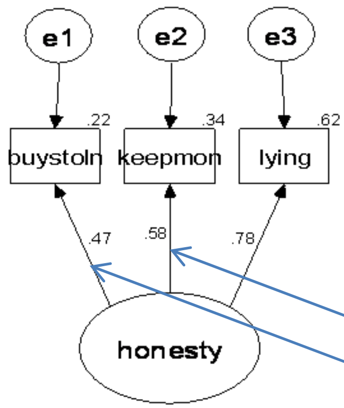
Finding the Path Coefficients

- **Normal Equations:**

- **(1) $r_{bk} = p_{bh} * p_{kh}$**
- **(2) $r_{bl} = p_{bh} * p_{lh}$**
- **(3) $r_{lk} = p_{lh} * p_{kh}$**
- We express p_{bh} from (1)
- $r_{bk} / p_{kh} = p_{bh}$
- We substitute p_{bh} in (2)
- $r_{bl} = (r_{bk} / p_{kh}) * p_{lh}$
- **We express p_{lh}**
- $r_{bl} / (r_{bk} / p_{kh}) = p_{lh}$
- **We substitute p_{lh} in (3)**
- $r_{lk} = (r_{bl} / (r_{bk} / p_{kh})) * p_{kh} = p_{kh} * p_{kh} * r_{bl} / r_{bk}$ → $p_{kh}^2 = r_{lk} * r_{bk} / r_{bl}$
- $p_{kh}^2 = .457 * .276 / .371 = .34$ → $p_{kh} = \sqrt{.34}$ **=+/-.583** *Notice that this number can be +.583 or -.583 because the latent*
- *variable can be scaled in either direction (it can measure honesty or dishonesty).*
- *We choose +.583 and the latent variable will be scaled in the same direction as K.*
- We can get p_{bh} by substituting in (1)
- $.274 = p_{bh} * .583$ → **$p_{bh} = .470$**
- And we can get p_{lh} by substituting in (3)
- $.457 = p_{lh} * .583$ → **$p_{lh} = .784$**

The Measurement Model Calculated by AMOS

Estimating the latent variable separately



Correlations

		buystoln	keepmon	lying
buystoln	Pearson Correlation	1	.276**	.371**
	Sig. (2-tailed)		.000	.000
	N	1732	1732	1732
keepmon	Pearson Correlation	.276**	1	.457**
	Sig. (2-tailed)	.000		.000
	N	1732	1732	1732
lying	Pearson Correlation	.371**	.457**	1
	Sig. (2-tailed)	.000	.000	
	N	1732	1732	1732

**. Correlation is significant at the 0.01 level (2-tailed).

$r_{bk}=p_{hb}*p_{hk}=.47*.58\approx.276$

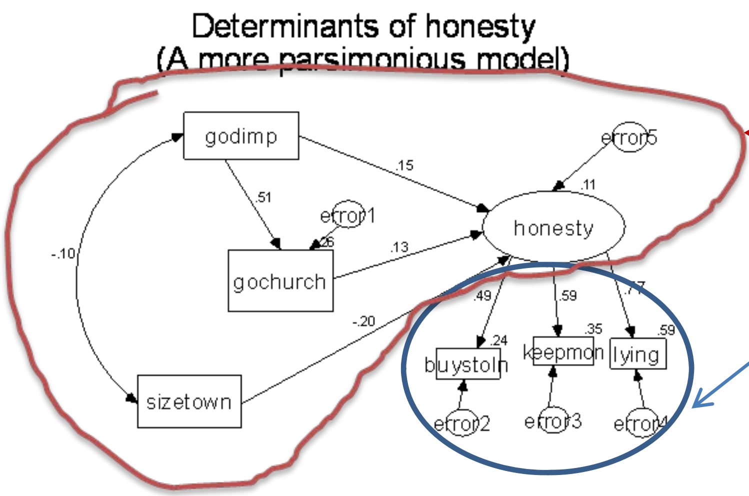
$r_{bl}=p_{hb}*p_{hl}=.47*.78\approx.371$

$r_{lk}=p_{hl}*p_{hk}=.78*.58\approx.457$

The paths and R-squares tell us how good each indicator is measuring the latent variable.

Attitude about lying (LYING) is the best indicator of honesty (.78). 62 percent of what people say about their attitude about lying reflects their attitude about honesty. The rest is error (e3).

Causal Model with Latent Variable

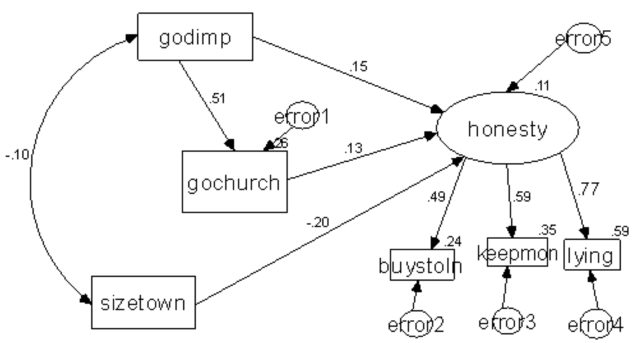


- Notice that we have 7 paths and 1 correlation or 8 coefficients to estimate.
- We have 6*(6-1)/2=15 normal equations (correlations)
- We have 15-8=7 degrees of freedom
  - We can test the entire model
- The model has a
- substantive part (relationships among concepts) and a
- measurement part (relationships among concepts and indicators).

- *IMPORTANT:*
- *Measurement CANNOT be separated from substantive theory. In fact, AMOS estimates the two simultaneously. If you change the substantive model, the measurement model may change as well.*

Evaluating Your Output

Determinants of honesty
(A more parsimonious model)



- Things to look for:
- 1. ***Could AMOS do the job?***
  - Did the model converge?
    - It should say: Minimum was achieved
- 2. ***Is your measurement model good?***
  - Are the indicators strong enough?
    - Direct effects of latent variables on indicators
      - R-squared of predicting indicators
  - Are their relative weights reasonable?
- 3. ***What does your substantive model say?***
  - Direct effects path coefficients
  - Indirect effects
- 3. ***How well are you predicting endogenous variables?***
  - Fitting each endogenous variable
    - R-square
- 4. ***Did you draw the right model/picture?***
  - Fitting the ***entire*** model
  - Chi-square test – statistical significance
    - Does the model significantly diverge from the data?
  - Various fit measures
    - How much does the model diverge on some standardized scale

How AMOS Fits Your Model

Sample Correlations (Group number 1)

	sizetown	godimp	gochurch	lying	buystoln	keepmon
sizetown	1.000					
godimp	-.102	1.000				
gochurch	-.034	.508	1.000			
lying	-.158	.175	.160	1.000		
buystoln	-.129	.158	.108	.371	1.000	
keepmon	-.130	.128	.125	.457	.276	1.000

Implied Correlations (Group number 1 - Default model)

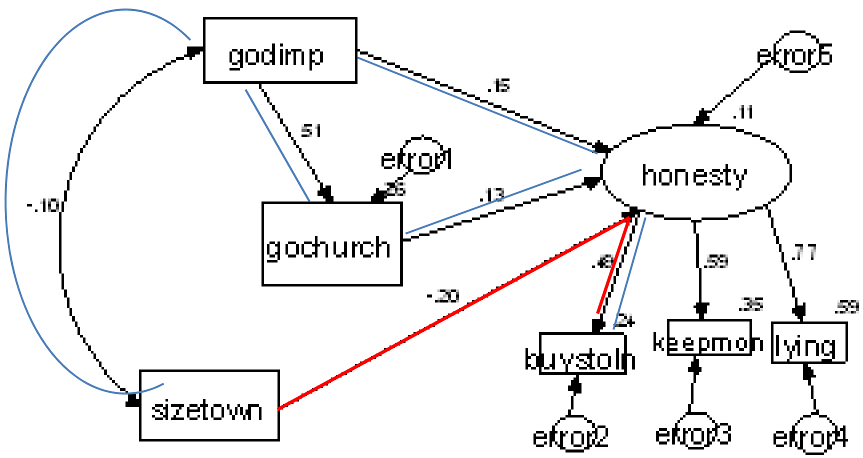
	sizetown	godimp	gochurch	lying	buystoln	keepmon
sizetown	1.000					
godimp	-.102	1.000				
gochurch	-.052	.508	1.000			
lying	-.168	.183	.164	1.000		
buystoln	-.107	.116	.104	.373	1.000	
keepmon	-.130	.141	.127	.452	.288	1.000

The fit of the entire model is evaluated by comparing the observed and implied correlations (covariances). *(AMOS really works with unstandardized variables and uses covariances rather than correlations. But for the sake of simplicity we assume that the world is standardized.)*

AMOS compares these two tables as you did in 205 when you calculated Chi-Square for a table comparing cell by cell the predicted (or implied) and the observed values. There you compared frequencies, here AMOS compares correlations (covariances).
Notice that here your model is good if Chi-Square is NOT significant because it means that the discrepancy between your model's predictions and the data is insignificant.
Also notice that

Let's take the correlation between BUYSTOLN and SIZETOWN.
Observed: -.129,
Implied: -.107.
Our model does not predict this correlation very well.
How is the implied correlation computed?
It is computed using the rules of path analysis.

The Implied Correlation Between BUYSTOLN and SIZETOWN



No direct effect

Indirect effect through HONESTY

-.20*.49=-.098

No spurious effect due to common causes (SIZETOWN is exogenous)

Correlated/Unanalyzed effects

through GODIMP and HONESTY

-.10*.15*.49= -.007

through GODIMP and GOCHURCH and HONESTY

-.10*.51*.13*.49=-.003

Implied correlation is (-.098)+(-.007)+(-.003)=-.108 ≈-.107

Evaluating the Fit of the Entire Model

- **Result (Default model)**
- Minimum was achieved
- Chi-square = 8.726
- Degrees of freedom = 7
- Probability level = .273

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.012	.000	.033	1.000
Independence model	.228	.218	.238	.000

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.994	.986	.999	.997	.999
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Model	AIC	BCC	BIC	CAIC
Default model	36.726	36.840	113.125	127.125
Saturated model	42.000	42.171	156.598	177.598
Independence model	1378.717	1378.766	1411.459	1417.459

- Chi-Square:
  - Measure of *statistical significance* of the fit (it is like the F-statistics for R-square
  - A Chi-Square is big if
    - You have a poor fit and/or you have a large N
  - Here our Chi-Square is 8.726 with 7 degrees of freedom
  - The probability level tells you the likelihood of getting this discrepancy between implied and observed correlation/covariance by chance when in the population your model would have a perfect fit
    - Your Chi-Square is NOT significant at the .05 or .1 level. It means that your fit is GOOD. The discrepancy is insignificant.
- Measures of FIT
  - It measures how close the path coefficients reproduce the correlation/covariance matrix (it is like R-square)
    - Default model – your model
    - Saturated model – model with 0 degree of freedom (d.f.)
    - Independence model – model with k*(k-1)/2 d.f. where k is the number of observed variables (all variables are unconnected)
  - **RMSEA:** Root Mean Square Error of Approximation. Exact fit is 0. A value under .05 is considered a close fit, and a value of .1 or larger means the model fit is very bad.
  - **CFI:** Comparative Fit Index. It is between 0 that indicates no fit (the Independence Model) and 1 that indicates perfect fit (Saturated Model).
  - **BIC**: Bayesian Information Criterion -- it is based on Bayesian probabilities and it compares two models and asks how likely is the second model if the population is accurately described by the first one. To evaluate BIC of the (default) model the BIC of the saturated model must be subtracted from it. If the difference is
  - $0 < BIC_{default} - BIC_{saturated}$ bad
  - $-2 \geq BIC_{default} - BIC_{saturated} \geq -6$ good
  - $-10 \geq BIC_{default} - BIC_{saturated}$ excellent