Book Review: "SAS Structural Equation Modeling 1.3 for JMP"

Brandy R. Sinco
Research Associate
University of Michigan
School of Social Work

Surprise – SEM is an add-in.

- The book begins by telling the reader to click on SAS / Structural Equation Modeling. But, this menu item requires an add-in.

- Download SEM13.jmpaddin from www.jmp.com.
- Installation is a bit confusing. After the add-in was installed, the menu looked the way the book said that it should.
- JMP tech support was very helpful, just like SAS.

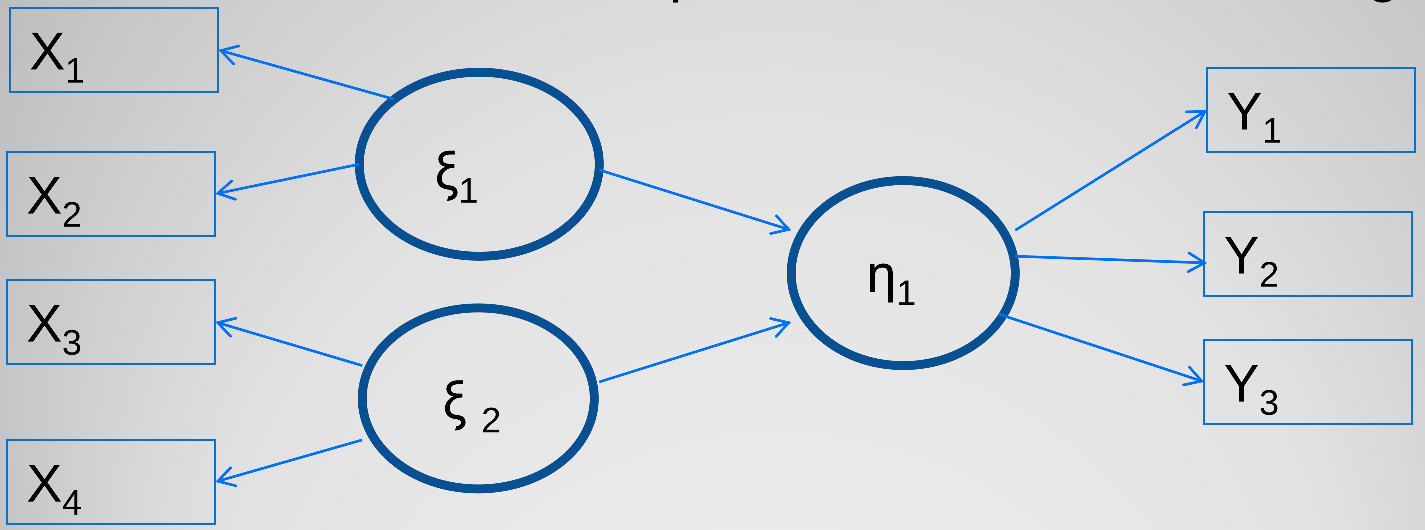
Contents of Book

- Linear Regression
- Path Analysis
- Confirmatory Factor Analysis
- Structural Equation Model
- Latent Growth Curve
- Single Group Analysis Window
- Model Library Window
- User Profile Window
- Properties Window
- Appendix: Frequently Asked Questions

- Book focuses on how to use JMP for SEM. It is not a textbook on SEM.

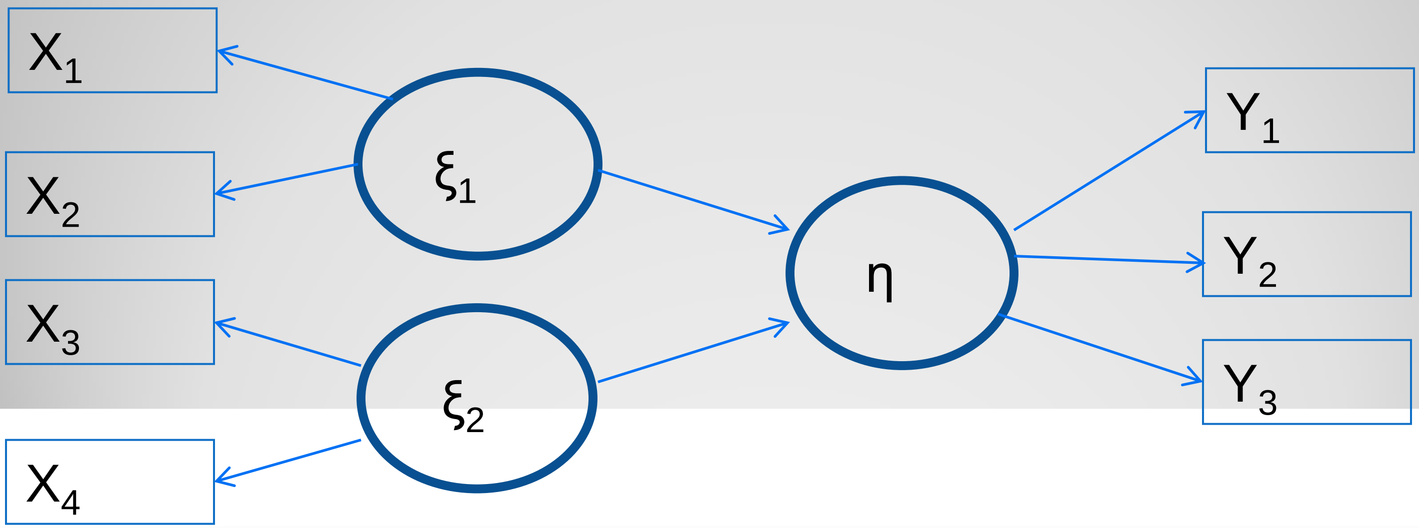
Brief Introduction to Structural Equation Modeling 1/2

- A system of linear equations based on a diagram that specifies the relationships between the variables.
- Variables can be manifest (observable X, Y) or latent (concept ξ, η).
- Exogenous ⟷ independent variables.
- Endogenous ⟷ outcome variables.
- Latent variables inside ellipse; manifest inside rectangle.



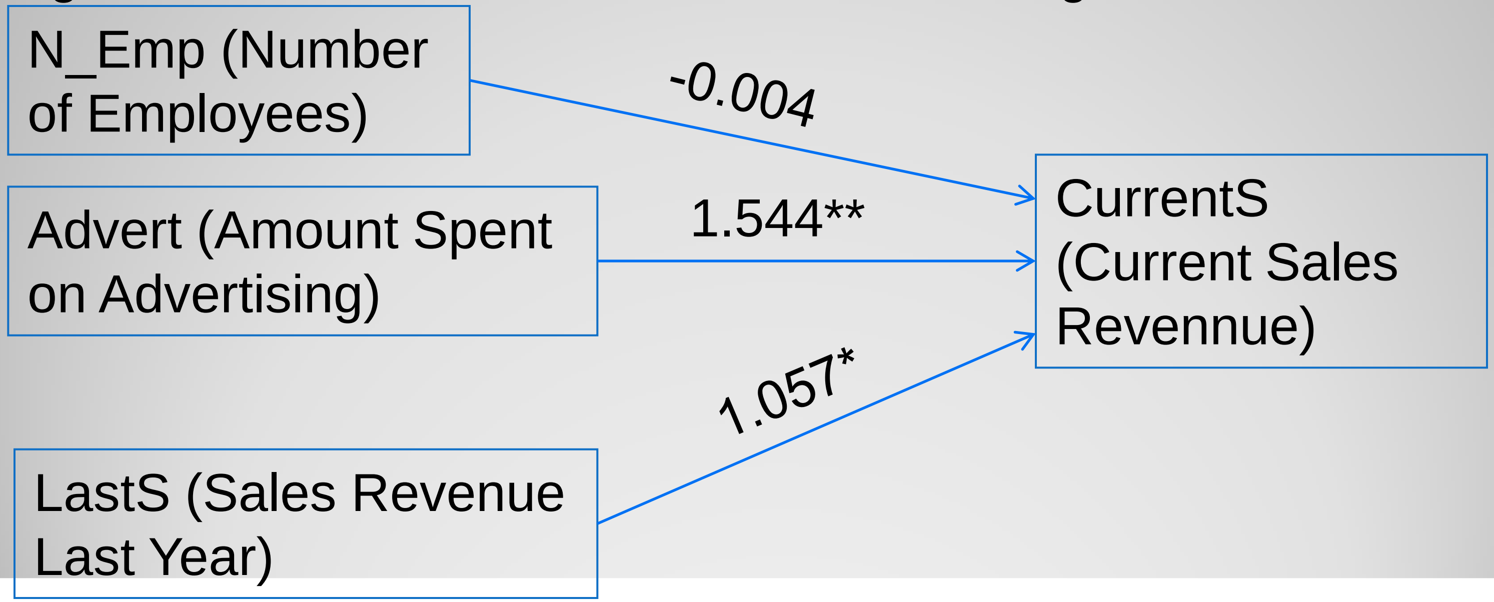
Brief Introduction to Structural Equation Modeling 2/2

- Variables can be latent (depression, anxiety) or observable (temperature, voltage, height, weight).
- X = exogenous, manifest; Y = endogenous, manifest.
- ξ (Xi) = exogenous, latent; η (Eta) = endogenous, latent.
- $X_1 = \tau_{X1} + \xi_1\Lambda_{X1} + \delta_1$; $X_2 = \tau_{X2} + \xi_1\Lambda_{X2} + \delta_2$.
- $X_3 = \tau_{X3} + \xi_2\Lambda_{X3} + \delta_3$; $X_4 = \tau_{X4} + \xi_2\Lambda_{X4} + \delta_4$.
- $\eta = \alpha + \xi_1\Gamma_1 + \xi_2\Gamma_2 + \zeta$.
- $Y_1 = \tau_{Y1} + \eta\Lambda_{Y1} + \varepsilon_1$; $Y_2 = \tau_{Y2} + \eta\Lambda_{Y2} + \varepsilon_2$;
- $Y_3 = \tau_{Y3} + \eta\Lambda_{Y3} + \varepsilon_3$.



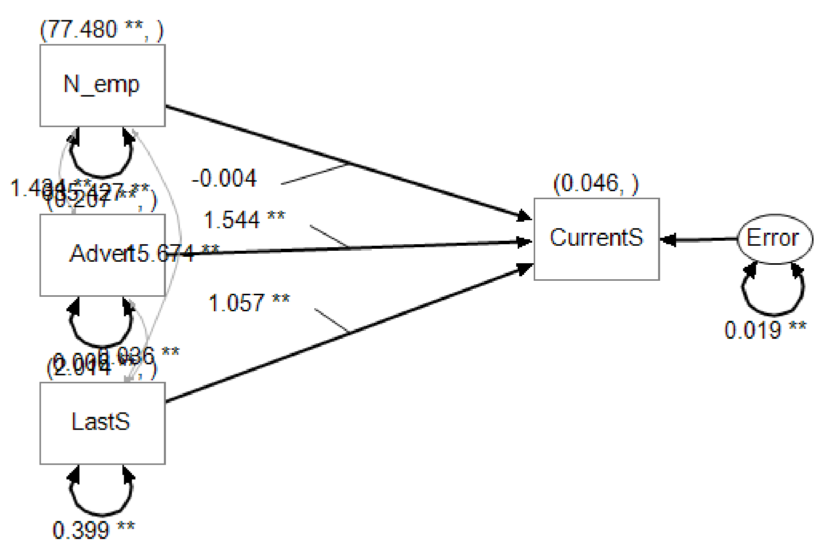
Ch 3, 4: Linear Regression, Drawing Diagram

- Linear regression is path analysis (SEM with manifest variables) when there is one outcome (endogenous variable).
- In real life, use Prog Reg for linear regression because least squares produces unbiased estimates and variances.
- SEM estimation by maximum likelihood. Coefficient values same as regression, but variances are biased. For large N, bias minimal.



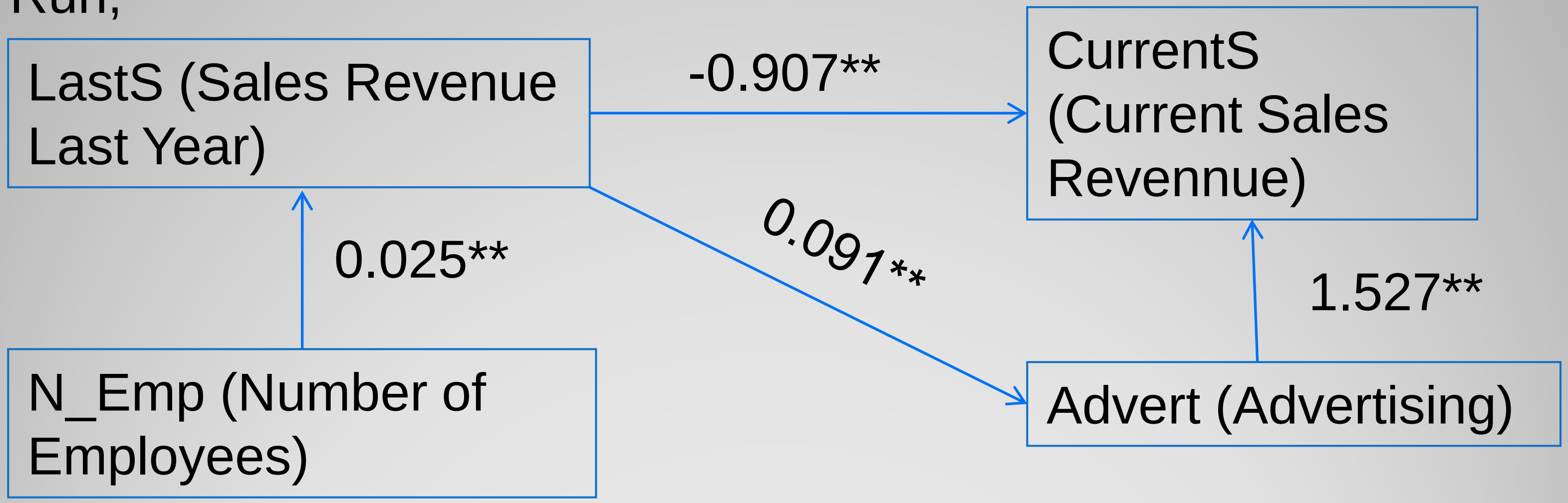
SEM diagram editor in JMP

- Model drawing in JMP is user-friendly; book is helpful guide.
- Covariances between exogenous variables in model by default, don't have to draw them like in AMOS.
- Option to display or hide covariances, error terms.
- For displaying model in presentation, I prefer to draw it in Word or Power Point with shapes, text boxes, and arrows.



Ch 5: Path Analysis (Manifest Variables Only)

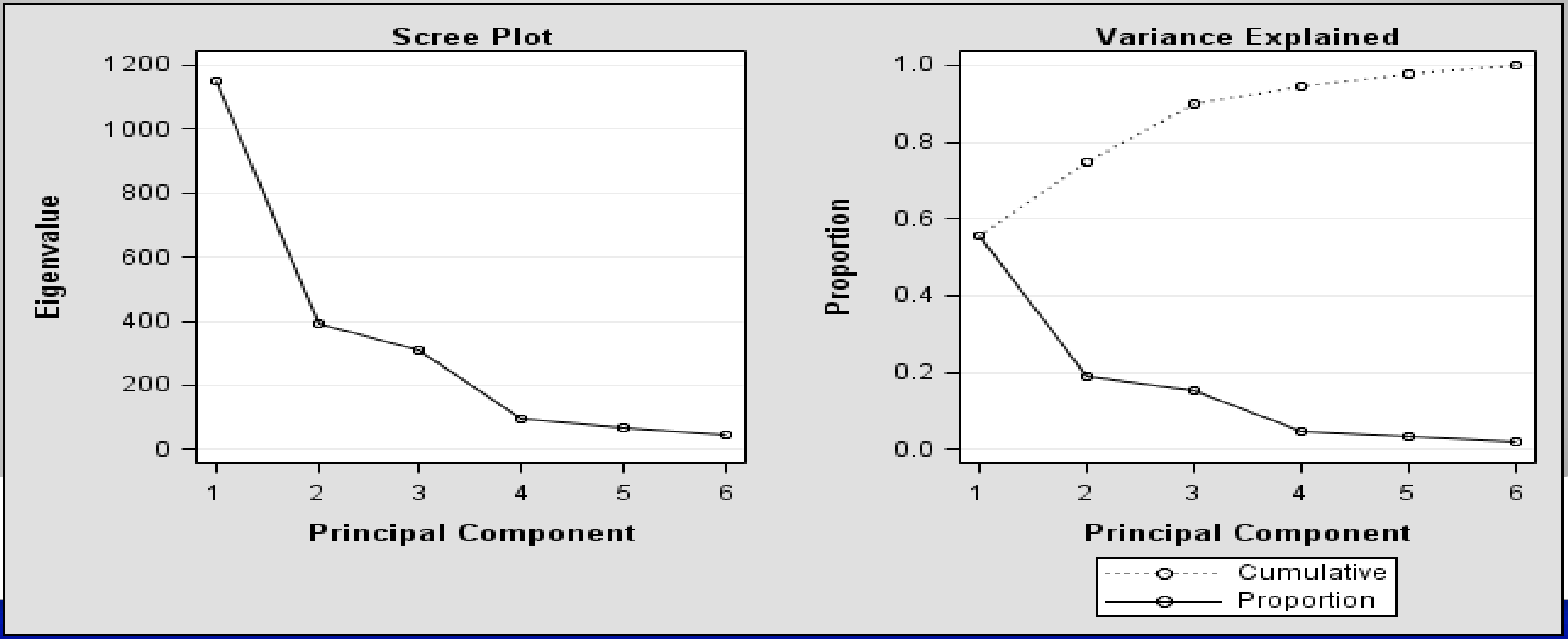
- Proc Calis data=Sales method=ml outest=semEstimates;_
- fitindex on(only)=[AGFI BentlerCFI ChiSq Df ProbChi nObs ProbClFit PGFI RMSEA LL_RMSEA UL_RMSEA SRMSR];
- Path Advert   <- LastS, CurrentS <- Advert, CurrentS <- LastS, LastS   <- N_emp;
- Run;



- Option to compare fit indices from different models.
- Convergence criterion in SAS log (ABSGCONV=.00001) satisfied.
- Classic Paper: Wright, S. (1934). "The Method of Path Coefficients", Annals of Mathematical Statistics, 5(3), 161-215.

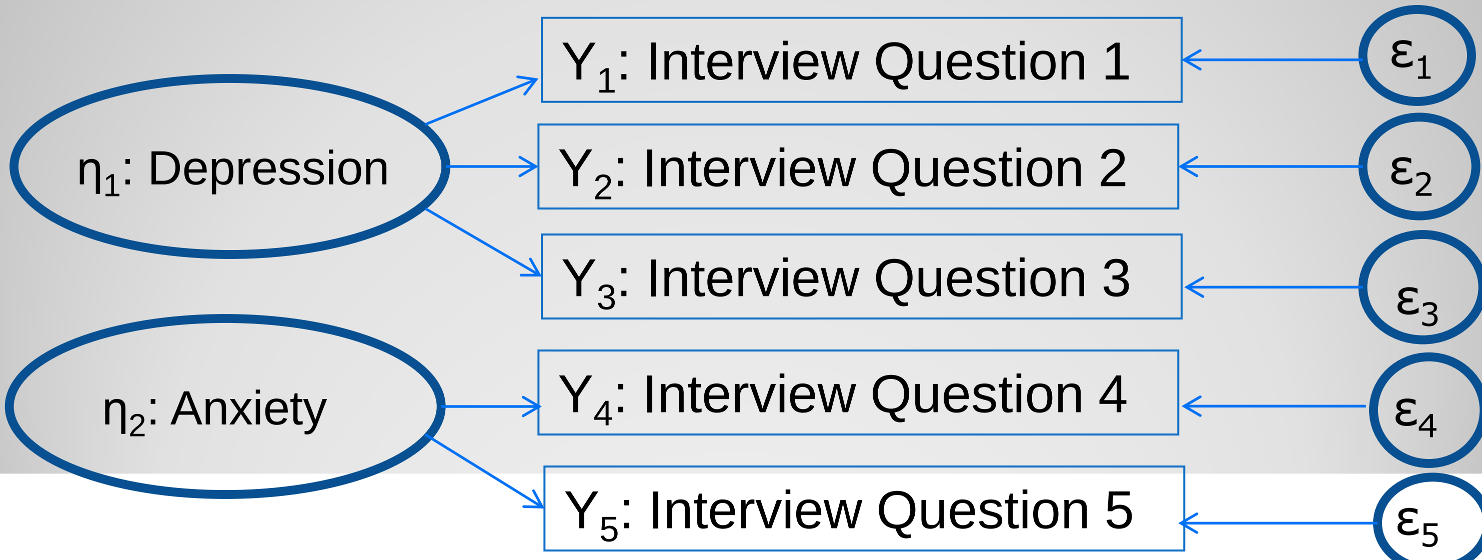
Ch 6: Confirmatory Factor Analysis, 1/3

- General Factor Equation. $Y = \tau_Y + \eta\Lambda_Y + \varepsilon$.
- Y = Endogenous manifest variables, n observations, q variables.
- η = Exogenous latent variables with mean 0, variance 1
- τ_Y = Y-intercepts, Λ_Y = factor loadings, ε = errors.
- EFA (Exploratory Factor Analysis). Proc Factor. Number of factors unknown. Hypothesis of number of factors based on scree plot, want to explain at least 70% of variance in Y by common factors.



Ch 6: Confirmatory Factor Analysis, 2/3

- Confirmatory Factor Analysis. Proc CALIS.
- System of equations with specific number of factors. Assess goodness of fit with indices for absolute fit (analogous to R^2), incremental fit, predictive fit, parsimony. Much wider array of indices to evaluate goodness of fit beyond % variance explained by factors.
- $Y_1 = \tau_{Y1} + \eta_1\Lambda_{Y1} + \varepsilon_1$; $Y_2 = \tau_{Y2} + \eta_1\Lambda_{Y2} + \varepsilon_2$;
- $Y_4 = \tau_{Y4} + \eta_2\Lambda_{Y4} + \varepsilon_4$; $Y_5 = \tau_{Y5} + \eta_2\Lambda_{Y5} + \varepsilon_5$.



Ch 6: Confirmatory Factor Analysis, 3/3

- Example in book showed how to create latent variables and compare two models side-by-side with the comparisons tab,
- How to set the correlation between factors to zero (helpful because Proc CALIS calculates covariances between factors by default, good to know how to override defaults).
- How to set factor loadings to 1, but not why an analyst would want to set a factor loading to 1?
- According to "Structural Equation Modeling" by David Garson (2011), pp. 63-64, each latent variable has to be assigned a metric. This can be done by constraining the latent variables to have variances of 1 or by constraining one of the factor loadings to be 1. If the factor loading from Y_i is constrained to be 1, then Y_i is the reference variable. Similar concept to having a reference group in ANOVA.

Ch 7: Structural Equation Model 1 of 3

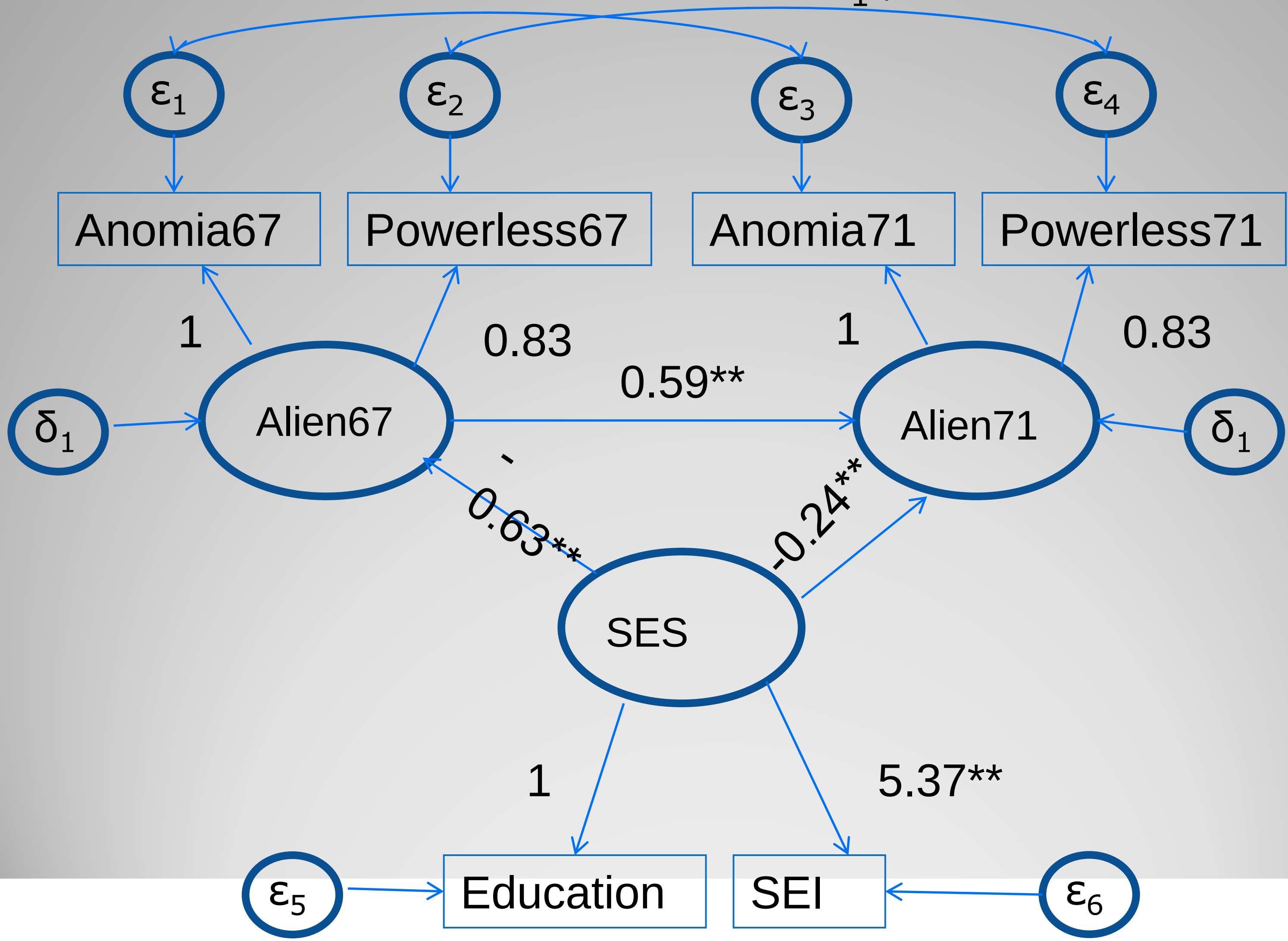
- Proc CALIS (Covariance Analysis and Linear Structural Equations).
- Dataset: Wheaton (1977). Study of political alienation over time.

Variables:

Variable	Description	Type
Anomia67 Anomia71	Scores on Anomia Scale in 1967 and 1971	Manifest
Powerlessness67 Powerlessness71	Scores on Powerless Scale in 1967 and 1971	Manifest
Alien67 Alien71	Political Alienation in 1967 and 1971	Latent
Education	Years of School as of 1966	Manifest
SEI	Duncan's Socioeconomic Scale, assessed in 1966	Manifest
SES	Socioeconomic Factor	Latent

Ch 7: Structural Equation Model 2 of 3

- Alien71 = 0.59*Alien67 – 0.24*SES + δ_1 , $R^2 = 0.50$.



Ch 7: Structural Equation Model 3 of 3

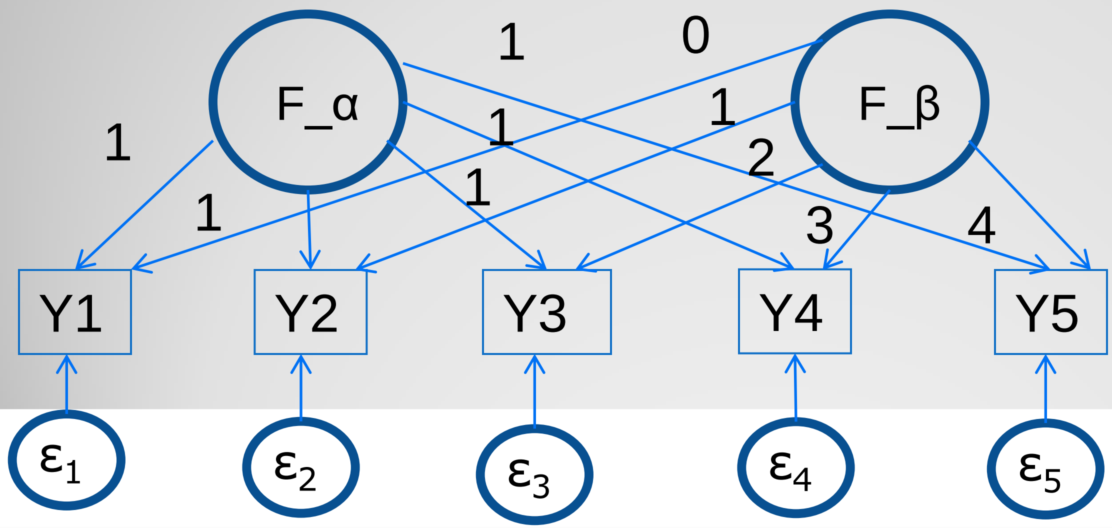
- Author of chapter said that model had excellent fit and provided insight into the relationships among the variables.
- RMSEA (Root Mean Square Error of Approximation) = 0.0231.
- CFI (Bentler's Comparative Fit Index) = 0.9979, close to 1.
- High R^2's for equations implied by model.
- Significant p-values for coefficients in model
- Chi-Square p-value = 0.1419.
- Want non-significant p-value because null hypothesis, H0, is that model's covariance matrix equals the sample covariance matrix, $\Sigma = S$.
- Alternative hypothesis is that $\Sigma \neq S$.

Ch 8: Latent Growth Curve 1/4

- Population average slope and intercept .
- Y=outcome, T=Time, i=subect index, ε = error..
- $Y_i = \alpha + \beta t_i + \varepsilon_i$. Growth pattern over time is assumed to be linear.

- Random effects allow estimation of individual growth curve for each subject.
- $Y_i = \alpha_i + \beta_i t_i + \varepsilon_i$.

- If 5 timepoints, estimate for subject i at time 1 is $Y1_i = \alpha_i$.
- At time 2, $Y2_i = \alpha_i + \beta_i$; at time 5, $Y5_i = \alpha_i + 4\beta_i$.
- In SEM, factors A and B function as the random intercept and slope.



Ch 8: Latent Growth Curve 2/4

- Equations for latent growth curve, in terms of factors
- $Y1_i = F\_\alpha_i + \varepsilon_i$, $Y2_i = F\_\alpha_i + F\_\beta_i t_i + \varepsilon_i$; $Y3_i = F\_\alpha_i + 2F\_\beta_i t_i + \varepsilon_{i,}$
- $Y4_i = F\_\alpha_i + 3F\_\beta_i t_i + \varepsilon_{i,}$ $Y5_i = F\_\alpha_i + 4F\_\beta_i t_i + \varepsilon_i$.

- Proc Mixed or Proc CALIS?
- Dataset, Growth contains ID, Y1-Y5.

```
ods html newfile=proc path="c:\tempfiles";   ods graphics on;
proc calis method=ml data=growth plots=residuals;
 lineqs
  y1 = 0. * Intercept + f_alpha  + e1,
  y2 = 0. * Intercept + f_alpha  +  1 * f_beta + e2,
  y3 = 0. * Intercept + f_alpha  +  2 * f_beta + e3,
  y4 = 0. * Intercept + f_alpha  +  3 * f_beta + e4,
  y5 = 0. * Intercept + f_alpha  +  4 * f_beta + e5;

 mean   f_alpha f_beta;
run; ods graphics off;  ods html close;
```

Ch 8: Latent Growth Curve 3/4

- ***** Proc Mixed Fixed Slope and Intercept ***;**
- ods html newfile=proc path="c:\tempfiles"; ods graphics on;
- Proc Mixed Data=LongGrow Method=REML NOCLPRINT;
- Class ID;
- Model Y=TimeB/ Solution Influence(effect=ID Est);
- Repeated / Type=UN Subject=ID R RCorr;
- Run;
- ods graphics off; ods html close;

- ***** Proc Mixed Random Slope and Intercept ***;**
- ods html newfile=proc path="c:\tempfiles"; ods graphics on;
- Proc Mixed Data=LongGrow Method=REML NOCLPRINT;
- Class ID;
- Model Y=TimeB/ Solution Influence(effect=ID Est);
- Random Int TimeB / type=un sub=id solution;
- Run;
- ods graphics off; ods html close;

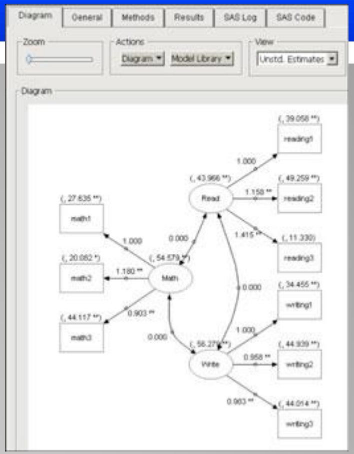
Ch 8: Latent Growth Curve 4/4

Method	F_α	F_β
Proc CALIS	14.03	3.97
Proc Mixed, Population Slope And Intercent	14.92	3.93
Proc Mixed, Random Slope and Intercept	14.16	4.05

- Estimates for population slope and intercept similar.
- Advantage of Proc Mixed with random slope and intercept provides subject level estimates **(see below)**. Other methods do not.
- However, Proc CALIS provides in-depth goodness of fit indices, absolute fit (GFI), incremental fit (CFI), parsimony (RMSEA), predictive fit (SRMR).

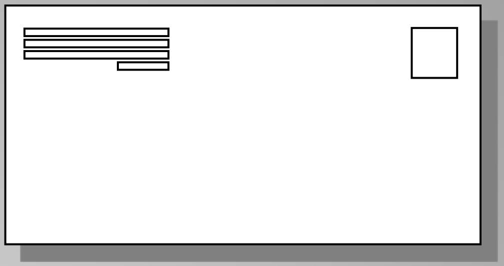
Effect	id	Estimate	Std Err
Intercept	1	2.81	1.49
timeb	1	0.84	0.51
Intercept	2	-2.13	1.49
timeb	2	-0.79	0.51
Intercept	3	-4.55	1.49
timeb	3	1.65	0.51

Conclusions – Book Worth Reading



- Useful reference book for how to set up structural equation models with a graphical user interface for single group analysis.
- Helpful tool in seeing how Proc CALIS code is generated from a structural equation model diagram.
- Easy-to-follow examples of Proc CALIS.
- Shows how to set covariances to zero, how to set paths to constants, and how to add intercepts.
- Not a self-teaching book on SEM; need to learn SEM theory through other books or courses.

Contact



- Brandy R. Sinco
- Statistician and Programmer Analyist
- University of Michigan School of Social Work
- 1080 S. University St.
- Ann Arbor, MI 48109-1106
- 734-763-7784
- E-Mail: brsinco@umich.edu