Growth Curve Modeling to Measure Impact of Temperature and Usage Amount on Detergent Performance

Zhiwu Liang, Procter & Gamble

A. Narayanan, University of Cincinnati

University of Cincinnati | Carl H. Lindner College of Business

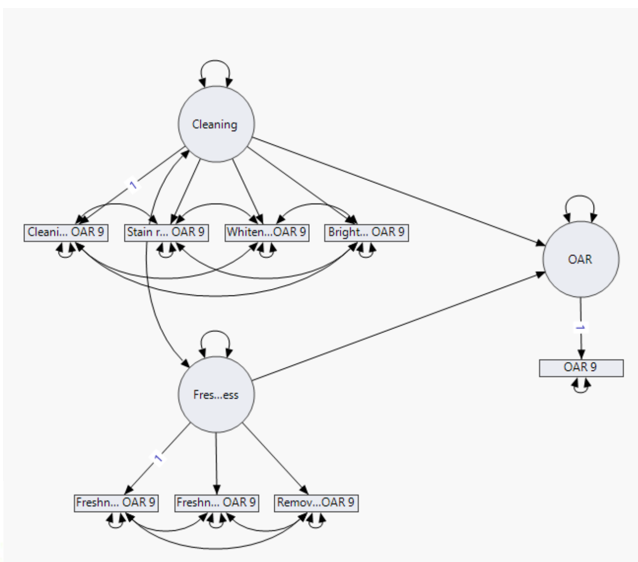
Outline

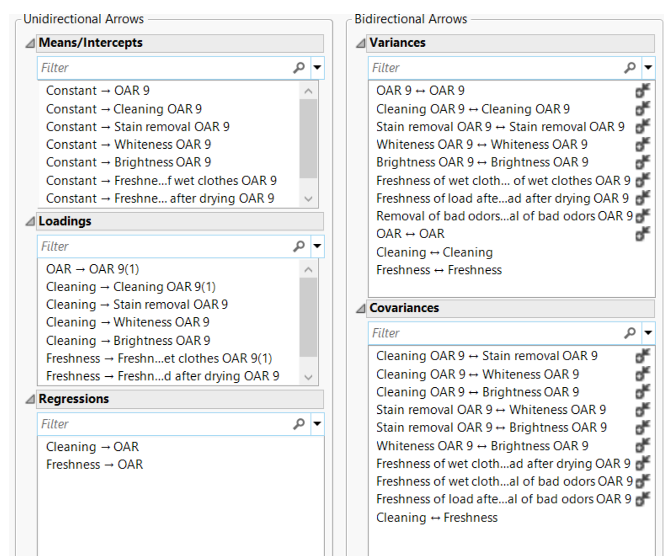
- Introduction of SEM (Structural Equations Modeling)
- A bit of data for modeling
- Introduction to Growth Curve Modeling
- Model building process
- JMP Demo and Results
- Conclusions and next steps

University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

SEM (Structural Equations Modelling)

- SEM is a multivariate technique that is used to test a set of relationships between observed and latent variables by comparing the model predicted covariance matrix and observed covariance matrix. In SEM, observed variables are manifest variables as indicators for latent variables and latent variables form the regression model to build a network.
- Example for an Overall rating model with 3 latent variables and 8 observed variables





University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

Survey data from detergent test diary results

- There are 119 consumers in 2 groups, 60 consumers use control product (Ariel SUD, coded as 0), 59 consumers use test product (Ecolabel, coded as 1)
- Each consumer use their own products in the first 4 weeks, then go to use the products assigned to them for next 8 weeks (from week 5 to week 12).
- Consumer will fill out questionnaire after each wash by providing information about washing temperature, number of pods used, soiled level of fabric and overall rating for load results
- The modeling objective is to test if there is product effect to overall rating, washing temperature to overall rating and number of pods used for overall rating

University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

Data Activation for Model Building

1. Aggregated diary data to weekly data for each panelist based on average washing temperature, number of pods used and OAR in each week

2. Data exploration revealed that OAR from week 9 to week 12 are stable and we decided to use data from these 4 weeks for our model

3. Build intercept only model for OAR from week 9 to week 12

4. Add manifest variables including wash temperature week 5 to 12, number of pods used week 5 to 12 and product (0/1) into the OAR model

5. Build growth curve (linear) for wash temperature, number of pods used by setting week 9 as anchor (means 5 to 8 weeks are lag manifest variables)

6. Build regression model from product, intercept of temperature, slope of temperature, intercept of pods#, slope of pods# to intercept of OAR

What is Latent Growth Curve Modeling (LGCM)?

- LGCM is a set of statistical models to understanding change in processes over time
- LGCM can be considered a special case of structural equations modeling (SEM)
- LGCM has all the advantages of SEM of specifying and testing relationships among observed and latent variables

- LGCM is an application of confirmatory factor analysis (CFA) with an embedded mean structure

University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

Evaluating Model Fit in LGCM

- Every model implies a covariance matrix and mean structure
- The model is tested using observed and predicted structures
- Exact fit between population and predicted results is tested using a chi-square test $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$

There are some watchouts with the chi-square test:

- ➢ The test statistic is a function of the sample size
- ➢ Test is global and does not reflect predictive ability or local fit
- ➢ Exact fit is too rigid (we know all models are …..)

University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

Evaluating Model Fit – Alternate Measures

- Root Mean Square Error of Approximation (RMSEA)
  - It measures model misfit adjusting for sample size
  - It is badness-of-fit measure (lower number is better)
  - Added benefit of confidence interval
  - Suggested model fit criteria (upper bound of RMSEA < .10)
- Comparative Fit Index (CFI) and Non-Normed Fit Index (NNFI)
  - Measure of relative fit
  - How good is the proposed model compared to the *worst model (model of no relationship)*?
  - It is goodness-of-fit measure (higher number is better)
  - Suggested model fit criteria (CFI & NNFI => 0.95)
- Standardized Root Mean Squared Residual (SRMR)
  - It is an average squared residual (like RMSE in regression)
  - It is a badness-of-fit measure (lower number is better)
  - Suggested model fit criteria (SRMR < .08)

Finally, DO NOT FORGET TO CHECK RESIDUALS!

Look at totality of fit (not just one fit statistic)

University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

Longitudinal Processes to Study

- Success Criteria (as measured by Overall Satisfaction Rating)
- Temperature Setting (in Celsius)
- Amount of Product Used (#pods)

Time Varying Covariates

- Type of Product (Ariel or Ecolabel)

Time Invariant Covariates

University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

Modeling Strategy

- Graph process trajectories (via Graph Builder)
- Select Latent Growth Curve Model (LGCM) for each process
- Build from univariate LGCM to multivariate LGCM (also called Parallel Process Model) to associate different processes
- Choose the right model based on fit statistics ***and substantive knowledge***
- Use simplification strategies and interpret the simplest model

JMP Demo

Conclusions

- We used Graph Builder platform to explore trajectories and SEM platform to build LGCModels
- We extended univariate models to multivariate models which were all within acceptable range of fit

- Product has significant negative impact on OAR latent variable which means Ariel has significantly better acceptance by consumer than Ecolabel product;
- Product has significant positive impact on number of pods latent variable for both intercept and slope which means Ecolabel product is used more than Ariel at the initial time period as well from week to week;
- Intercept of washing temperature has negative impact on OAR which means that higher temp wash is not preferred by consumer, cold wash conditions will provide higher consumer acceptance

P&G

University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

Next Steps

- Modeling results provide us confidence to claim cold wash and using less product for Ariel compared to Ecolabel product;
- Confirm consumer habit change with longitudinal data;
- Will bring more background variables (like using additive, washing cycle, load size) in next big sample size consumer test for a complex growth curve model

University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

Questions?

University of CINCINNATI
CARL H. LINDNER COLLEGE OF BUSINESS

P&G