

Some Advanced Approaches

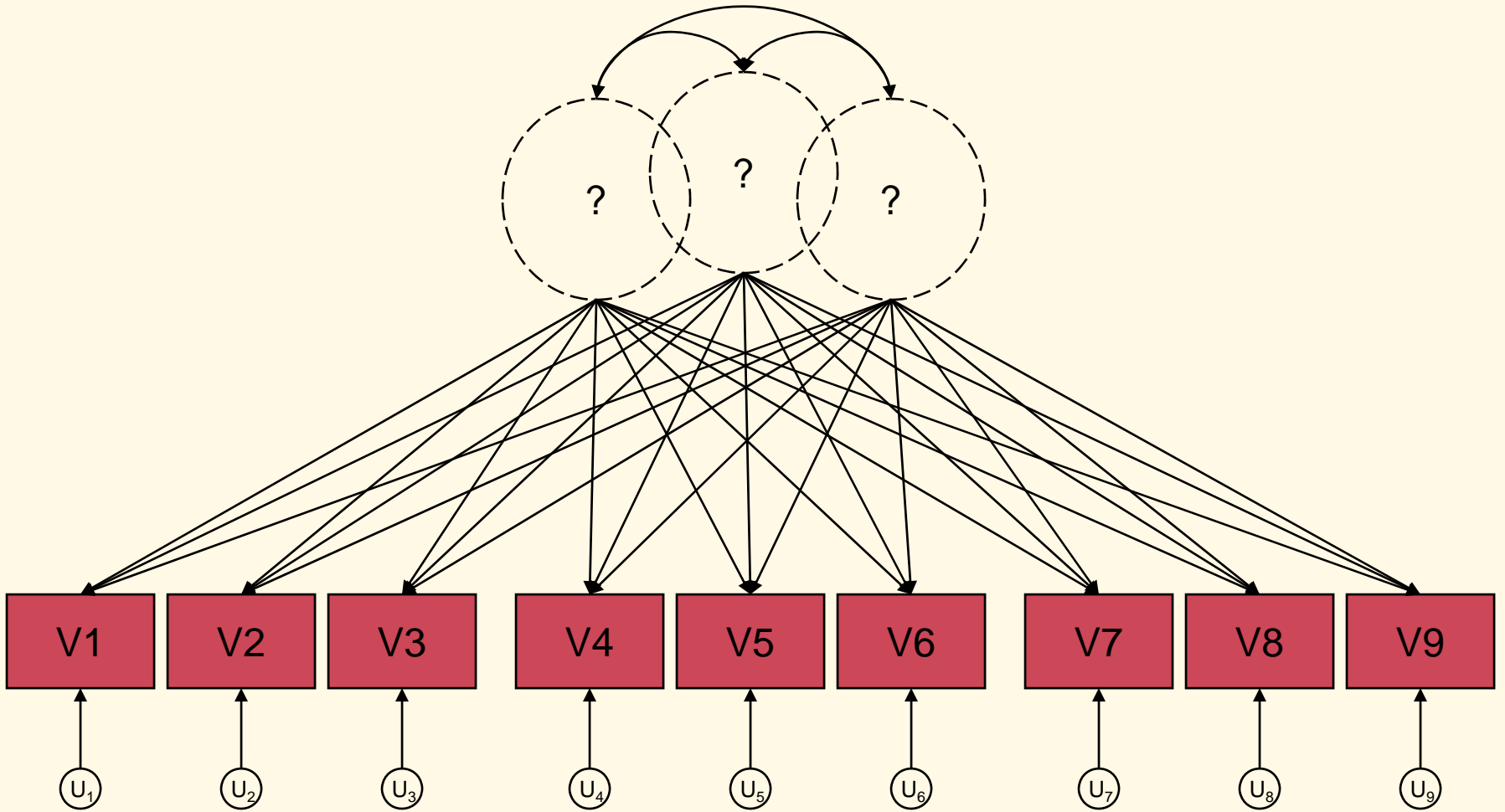


Confirmatory Factor Analysis



- First recall factor analysis as we have presented
- Given a set of variables, the goal is to find the underlying structure
- The basic model is $X = \lambda\xi + \delta$
 - Each observed variable is a function of a common factor and unique factor
- It is exploratory in that we do not specify a structure beforehand
 - Each observed variable is a function of all factors delineated
- However, with EFA we can:
 - Specify how many factors we want the solution to have
 - Specify whether or not factors are to be correlated
 - Test for hierarchical factor structure
 - Compare relative fits of competing factor structure models
- In this sense even EFA can be seen or used as partly confirmatory

EFA

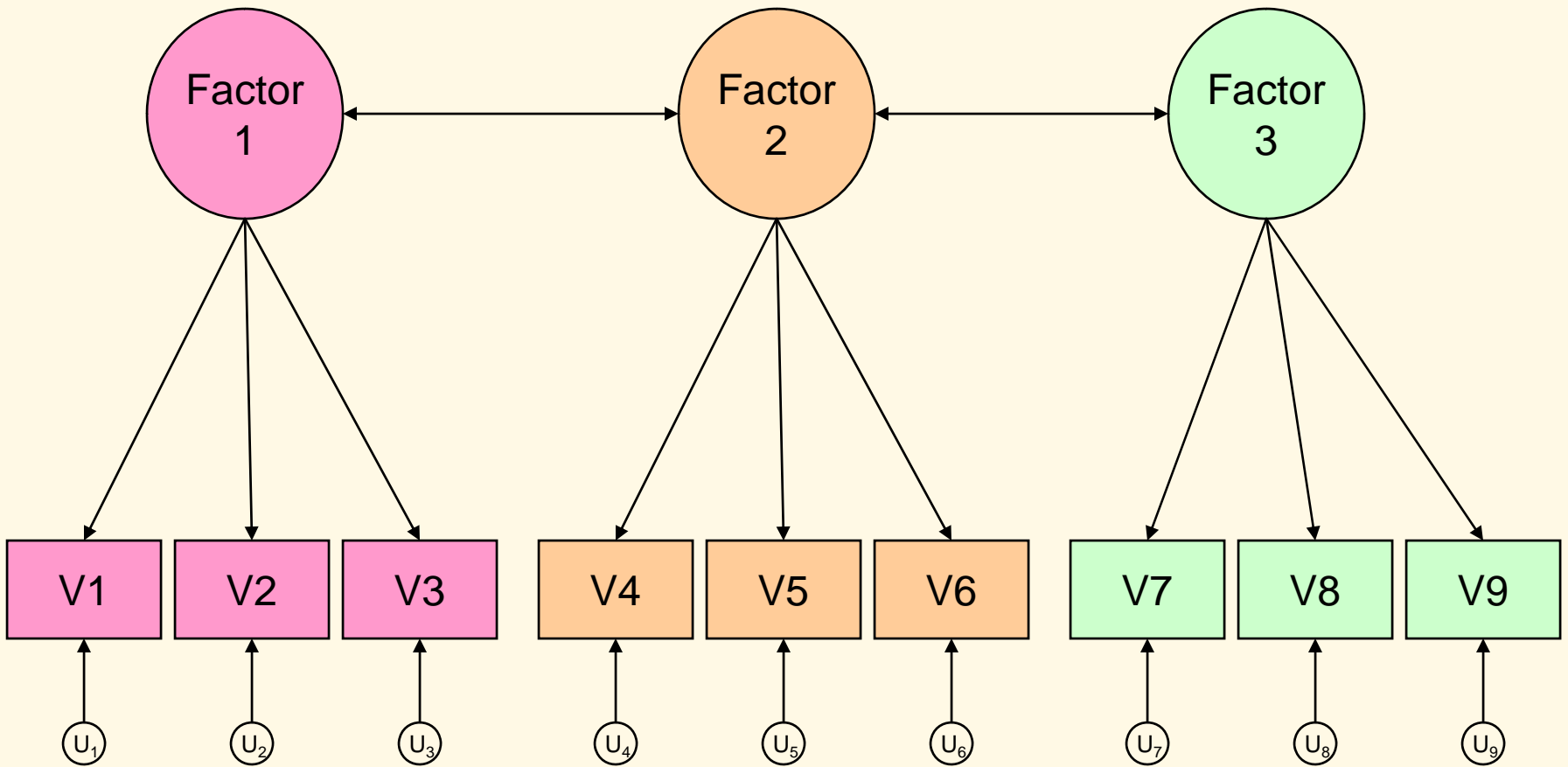


Confirmatory Factor Analysis



- With CFA the underlying model is the same, however all elements are specified before analysis is conducted
 - A factor structure is specified with k factors and the variables are specifically noted to load only on certain factors (i.e. constrained to have loadings of zeros on other factors)
 - ✦ $X_1 = \lambda_1 \xi_1 + 0 \xi_2 + \delta_1$
- As such, CFA has the same specification and identification issues that we saw with path analysis
- However, in common application modification indices are examined and the CFA model changed, thus turning it into an exploratory enterprise (and it is exploratory if no subsequent validation is carried out after changes)
- As noted before, even when starting with a confirmatory approach, some analyses become exploratory

CFA



Distinction



- **EFA**

- Specification of only observed variables (and perhaps number of factors)
- All observed variables are caused by *all* common factors
- Unique factors of observed variables are uncorrelated
- Common factors are not correlated with unique factors

- **CFA**

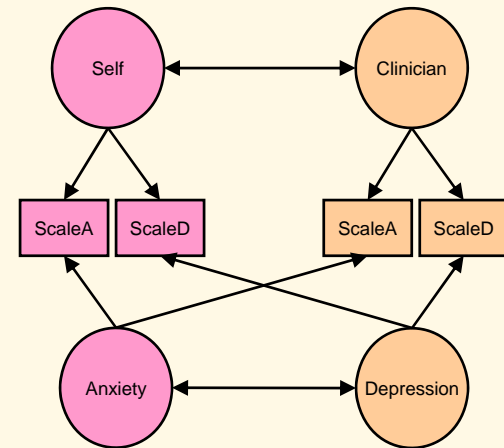
- Researcher imposes substantively motivated constraints, determining which:
 - ✦ Common factors are correlated
 - ✦ Observed variables are caused by which common factors
 - Unidimensional vs. Multidimensional
 - ✦ Unique factors are correlated

Special Types of CFA



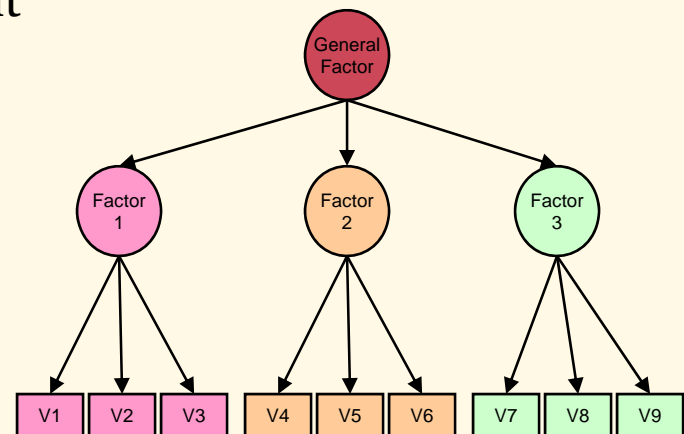
- **Multitrait Multimethod**

- Evaluate convergent and discriminant validity of a set of tests that vary in their measurement method
- Derive separate estimates of the effect of traits versus measurement methods on the observed scores
 - ✦ Factors for different formats
 - ✦ Factors for different sources (e.g. parent vs. clinician)



- **Hierarchical structure**

- Higher order factors



Other Issues

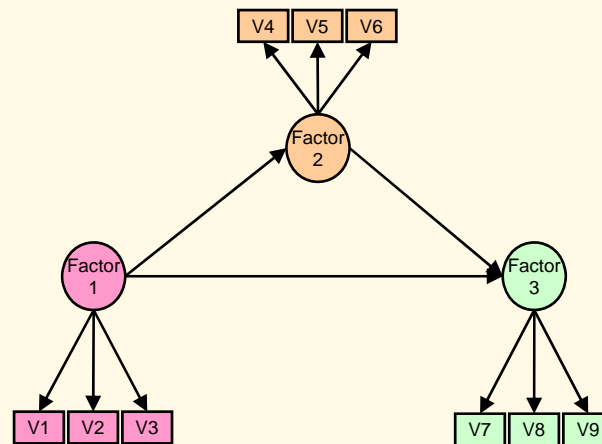


- Estimation of Factor Scores
 - There are different methods available, but analyses with factor scores may not hold well with new data
- Items vs. Scales
 - Items are less reliable and tend to be categorical or ordinal though would be used in the process of scale development in particular
 - ✦ An approach such IRT may be more appropriate however
- Model Search
 - E.g. Tetrad

Structural Equation Modeling



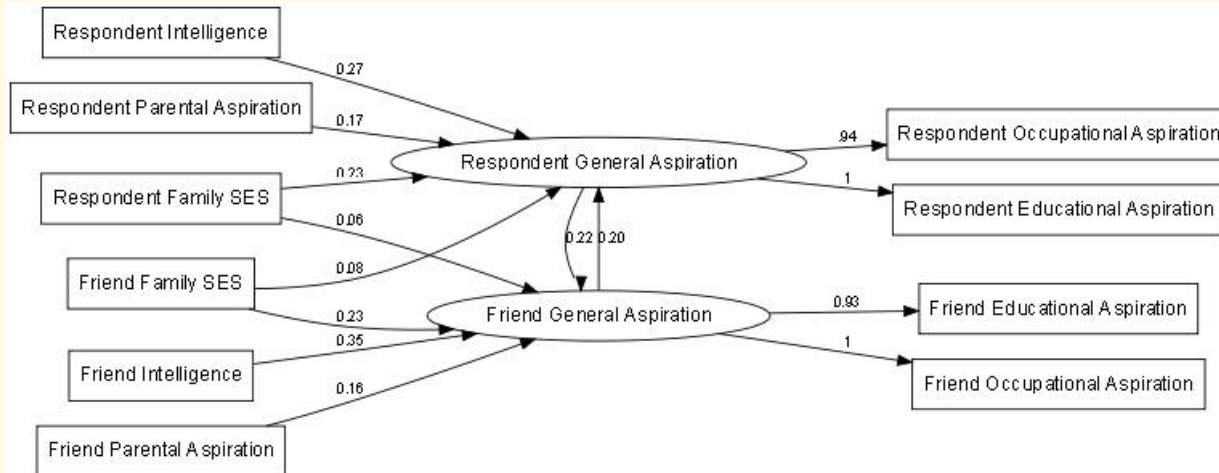
- CFA with structural relations applied
 - Aka Hybrid Model with both measurement and structural models
- In essence the measurement model is first estimated and the correlations between factors then serve as input to estimate the coefficients between the latent variables
 - Programs do this simultaneously however



SEM Example



- Peer influences on occupational choice
 - Duncan Haller Portes
- Might look fancy but in terms of conceptual theory you really just have two four predictor models (and fairly obvious ones at that) whose outcomes predict one another



SEM Considerations

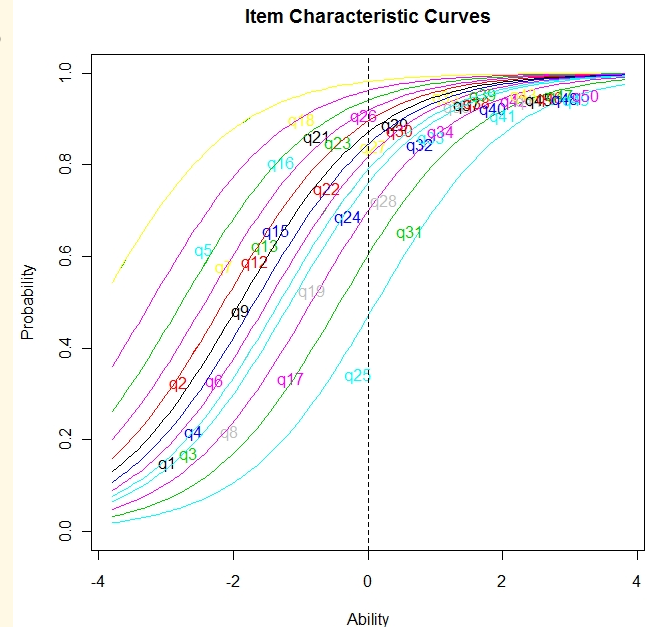


- Sample size requirements are huge for real data
 - The heuristic of $N = 300$ is rarely adequate except for well-behaved data and extremely simple models
- May be fully or partially latent models
 - Latter includes structural relations with one or more observed variables
 - The prior example is partially latent
- May include single indicator variables though that comes with additional issues and usually one would want at least three
- May have causal indicators for factors
 - If theoretically justified, one may have the observed variables cause the factor or have a mix (*Multiple Indicators Multiple Causes* ‘MIMIC’)
 - Such factors are always endogenous
- Latent Class Analysis involves presumed categorical latent variables
- Latent Growth Models involve the application of SEM to longitudinal data

Item Response Theory



- Factor analysis as we've discussed it may be specifically used for the purposes of scale development
- IRT is a form of modeling that can also be used for this and is in many ways superior for ability measures of various kinds of tests
- It assumes reliability is not constant across individuals, and frames the problem in terms of probability of 'correct' response as a function of a person's ability
- As it can distinguish item bias from ability, its use is extremely important in construction of tests that are equivalent in different contexts
- There are three main models
 - One parameter (Rasch): $p(\text{correct})$ is a function of ability
 - ✦ Assumes items discriminate equally and guessing is irrelevant
 - Two parameter model: adds in an item discrimination factor
 - Three parameter model: allows for guessing
- Extends to polytomous items (e.g. Likert) as well as multidimensional models that extend beyond just ability



ICC from a 5700 final

Generalized Linear Models



- GLMs allow for prediction of variables with different types of distributions via a ‘link’ function
 - Typical linear regression is a special case that uses the normal distribution and identity link
 - Logistic regression uses the logit link function
 - Poisson regression uses the log function, and is used to model count outcomes
- Others are available that might be applicable to specific research situations
- Gist is that the model is in some manner still of the basic linear form you know from regression

Multilevel Modeling



- Mixed effects models, Random effects models, Hierarchical Linear modeling etc.
- In psychology we'd typically be interested in individual, family or other grouping factors in which data of interest can be seen to be nested within
- With ML models, the parameters to be estimated are allowed to vary across different levels of some factor
 - Intercepts
 - Slopes
 - Both

MLM: Comparison to the familiar



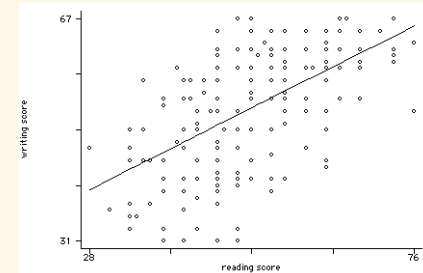
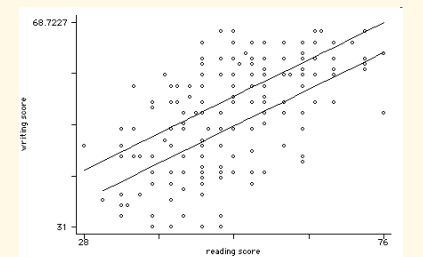
- As an example, repeated measures design can be formulated as a special case where the structural model is the same as between groups ANOVA but the intercept is allowed to vary over individuals
 - On the right we have the initial model where scores are a result of the effect of the a repeated measure (T for time)
 - However, the intercept is allowed to vary across subjects (Z)
 - The complete model is then provided

- With ANCOVA we see a covariate-DV relationship whose intercept is fixed but different for each group while the slopes are the same
 - The slopes *have to be* for ANCOVA estimates to be correct
- If we actually did take a random effects approach however, the intercepts are assumed to be drawn randomly from a certain distribution (i.e. normal distribution with mean γ_0 and some variance)
 - If there is no variance, it collapses to a regular regression model
 - Likewise if there is no Covariate-DV correlation, the ANCOVA also collapses to the simple regression model
- In the end there is only one model in the random effects approach, but the variance for the parameters are estimated differently

$$y = b_0 + b_1T_1 + e$$

$$b_0 = \gamma_0 + \gamma_1Z_1 + u$$

$$y = \gamma_0 + b_1T_1 + \gamma_1Z_1 + (e + u)$$



Multilevel Modeling



- MLM is not going to do much for if there are not many levels on the grouping factor
- Furthermore, it is easy to become overwhelmed with parameter estimates with a multilevel design, as models are being tested for Level 1, Level 2 etc. and there may be cross-level interactions posited as well
- Some questions to ask yourself include:
 - Do you want to estimate a random effect?
 - ✦ Almost everything you've been trained on assumes fixed effects in the population
 - Is there enough variability across levels to warrant such an approach?
 - Are you willing to trade model complexity in terms of predictors (i.e. the ability to capture reality with more predictors) for model complexity across levels?

Summary



- CFA is conceptually not too different from EFA
 - EFA explore the factor structure
 - CFA test the factor structure
- SEM is a combination of CFA and path analysis, but often is applied to overly simplistic models or overly complex ones in practice
 - Example: A four factor SEM with three factors causing the fourth is still only has four theoretical variables, and when is that typically adequate to explain human phenomena?
 - Example: 10 factors with multiple interrelationships, moderation etc. is probably not something one can easily apply practically
 - Path analysis with very reliable measures will tell you essentially the same thing and more efficiently
- Generalized linear models are no more complex by definition than your regular linear models
- IRT and MLM can be seen as alternate forms of classical techniques but with different assumptions about how the parameters of interest should be estimated
- While it may seem like these are more advanced techniques, or automatic ‘dissertation/publication worthy’ sorts of endeavors, they aren’t any more so just because they are more computationally complex, and certainly not if they are done poorly