Structural Equation Modeling

HGEN502, 2008

Danielle M. Dick

Hermine H. Maes

Role of model mediating between theory and data

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Theory

- Phenotype = Genotype+Environment (VP=VG+VE)
- Genetic factors (VG):
  - Additive (VA)
  - Dominance (VD)
- Environmental factors (VE):
  - Common / Shared (VC)
  - Specific / Unique (VE)
- Assumptions: additivity & independence of genetic and environmental effects

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Model Building

- **Write equations**

  or

- **Draw path diagrams**
  - ☐ Trace backwards, change direction at a 2-headed arrow, then trace forwards (implies that we can never trace through two-headed arrows in the same chain).
  - ☐ The expected covariance between two variables, or the expected variance of a variable, is computed by multiplying together all the coefficients in a chain, and then summing over all possible chains.

Model

- VP = VA + VC + VE
- ACE Model

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Study Design

- Uniquely quantify sources of variation of interest
- Subjects available in large enough samples to obtain stable estimates of sources of variation, and representative of general population
- Classical Twin Study (MZ & DZ twins reared together): Separation of genetic and shared environmental factors
- Equal environment assumption; Random mating; NO: GE correlation, G x E interaction, sex limitation, G x age interaction

QuickTime™ and a
TIFF (LZW) decompressor
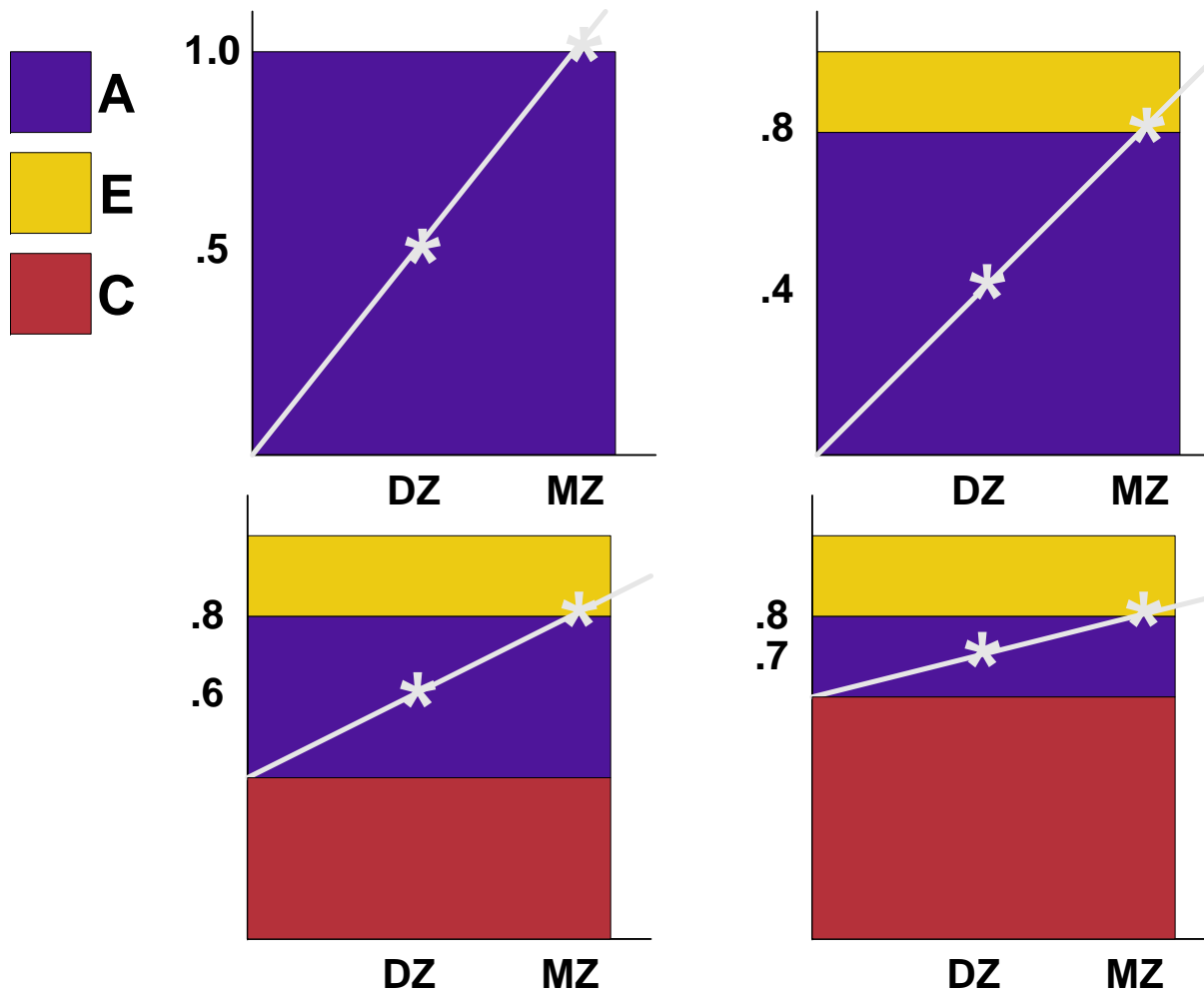are needed to see this picture.

Twin Correlations

Observed Statistics:
1. MZ correlation
2. DZ correlation
3. Variance

Unknown Parameters:
1. A: additive genetic
2. C: shared environment
3. E: unique environment

Data

- Collect phenotypic data from large sample of relatives/ twins (N > 100 - 10000)
  - Power session on Friday
- Calculate / Estimate means & variances
- Estimate covariances/ correlations between different types of relatives
- Fit 'the' model

Model Fitting

- **Formula's**
  - ☐ Falconer: $h^2_b = 2(rMZ - rDZ)$: broad sense heritability
  - ☐ Holzinger: $H = (rMZ - rDZ)(1 - rDZ)$
- **Analysis of Variance (Jinks & Fulker, 1970)**
- **Structural Equation Modeling (SEM)**
  - ☐ Includes Regression Analysis & Factor Analysis & most Linear Models
  - ☐ Evaluate series of equations
  - ☐ Solve for unknowns iteratively
  - ☐ Path Analysis: tool to derive expectations

SEM Advantages

- Systematic approach to hypothesis testing and parameter estimation
  - Evaluate goodness-of-fit of model - closeness of observed & expected values & Compare fit under alternative models
  - Obtain maximum likelihood estimates & Evaluate significance of parameters- effect size & sample size
- Both continuous and categorical variables
- Extendable: multiple variables; covariates; extended pedigrees; selected samples
- Make use of all available data

Model (Estimates)

- Goodness-of-fit: likelihood
- ML estimates and confidence intervals

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Revision

- **Test alternative models**
  - ACE
  - ADE
- **Test significance of parameters**
  - AE: test significance of C
  - CE: test significance of A
  - E: test significance of A&C, familial resemblance

Theory

- Conclusions regarding the best fitting model taking into account model fit and parsimony (which sources of variance are significant?)

- Conclusions regarding the contributions: magnitude of effects of genetic versus environmental factors (how big is their effect?)