# Structural Equation Models[1]

STA431 Winter/Spring 2017

[1] See last slide for copyright information.

# Structural Equation Models

- An extension of multiple regression.
- Can incorporate measurement error.
- More than one regression-like equation.
- All the variables are random.
- An explanatory variable in one equation can be the response variable in another equation.

# Measurement Error

- What you see is not what you really want.
- *Latent variable*: A random variable whose values cannot be directly observed. For example, family income last year.
- Contrast with *Observable variable*. For example, reported family income last year.
- Usually, interest is in relationships between latent variables.
- But all you have in your data set are the observable variables.

# Doubly Labeled Water

Participants drink water that is enriched with respect to two isotopes, and urine samples allow the measurement of energy expenditure (Graphics used without permission).

*Measurement Error in Nonlinear Models*: Carroll et al., 2006, p. 8

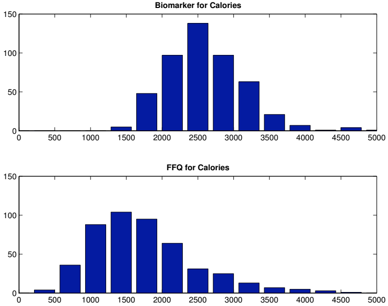
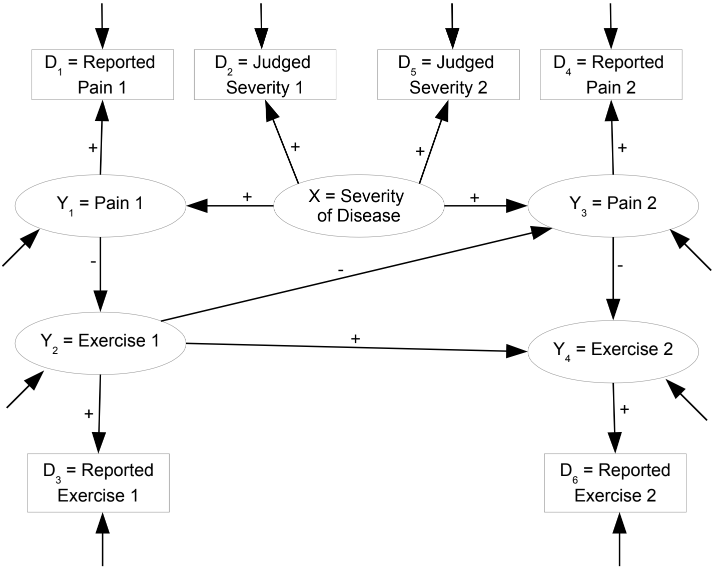



Figure 1.5 *OPEN Study data, histograms of energy (calories) using a biomarker (top panel) and a food frequency questionnaire (bottom panel). Note how individuals report far fewer calories than they actually consume.*

# Path diagrams

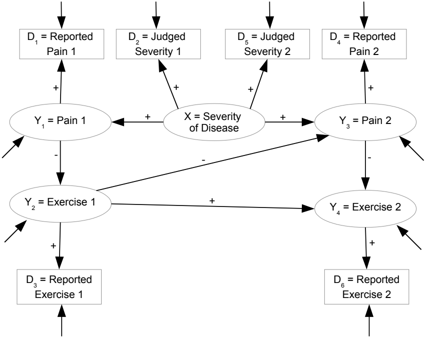
## Example: Exercise and arthritis pain



# Comments

- Latent variables are in ovals, observable variables are in boxes.
- Error terms seem to come from nowhere – often not shown.
- There is real modeling here. Lots of theoretical input is required.
- These are usually interpreted as *causal* models: Models of influence.
- $A \to B$ means $A$ has an influence on $B$.
- But the data are usually observational.

# Path diagrams correspond to systems of equations

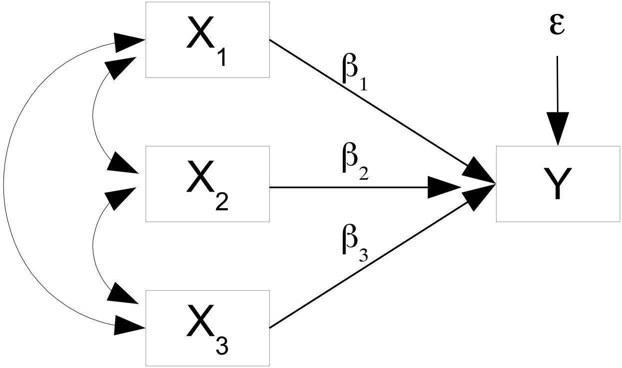



$$
\begin{aligned}
Y_{i,1} &= \beta_{0,1} + \beta_1 X_i + \epsilon_{i,1} \\
Y_{i,2} &= \beta_{0,2} + \beta_2 Y_{i,1} + \epsilon_{i,2} \\
Y_{i,3} &= \beta_{0,3} + \beta_3 X_i + \beta_4 Y_{i,2} + \epsilon_{i,3} \\
Y_{i,4} &= \beta_{0,4} + \beta_5 Y_{i,2} + \beta_6 Y_{i,3} + \epsilon_{i,4} \\
D_{i,1} &= \lambda_{0,1} + \lambda_1 Y_{i,1} + e_{i,1} \\
D_{i,2} &= \lambda_{0,2} + \lambda_2 X_i + e_{i,2} \\
D_{i,3} &= \lambda_{0,3} + \lambda_3 Y_{i,2} + e_{i,3} \\
D_{i,4} &= \lambda_{0,4} + \lambda_4 Y_{i,3} + e_{i,4} \\
D_{i,5} &= \lambda_{0,5} + \lambda_2 X_i + e_{i,5} \\
D_{i,6} &= \lambda_{0,6} + \lambda_5 Y_{i,4} + e_{i,6}
\end{aligned}
$$

Multivariate normal model is standard.

# Regression with observable variables

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \epsilon_i$$



# Tools

- Scalar variance-covariance calculations
- Matrices
- Random vectors
- Multivariate normal
- Maximum likelihood
- A little large-sample theory
- SAS

# Copyright Information