My Favorite (Mixed) Models: An Overview

Rod Sturdivant

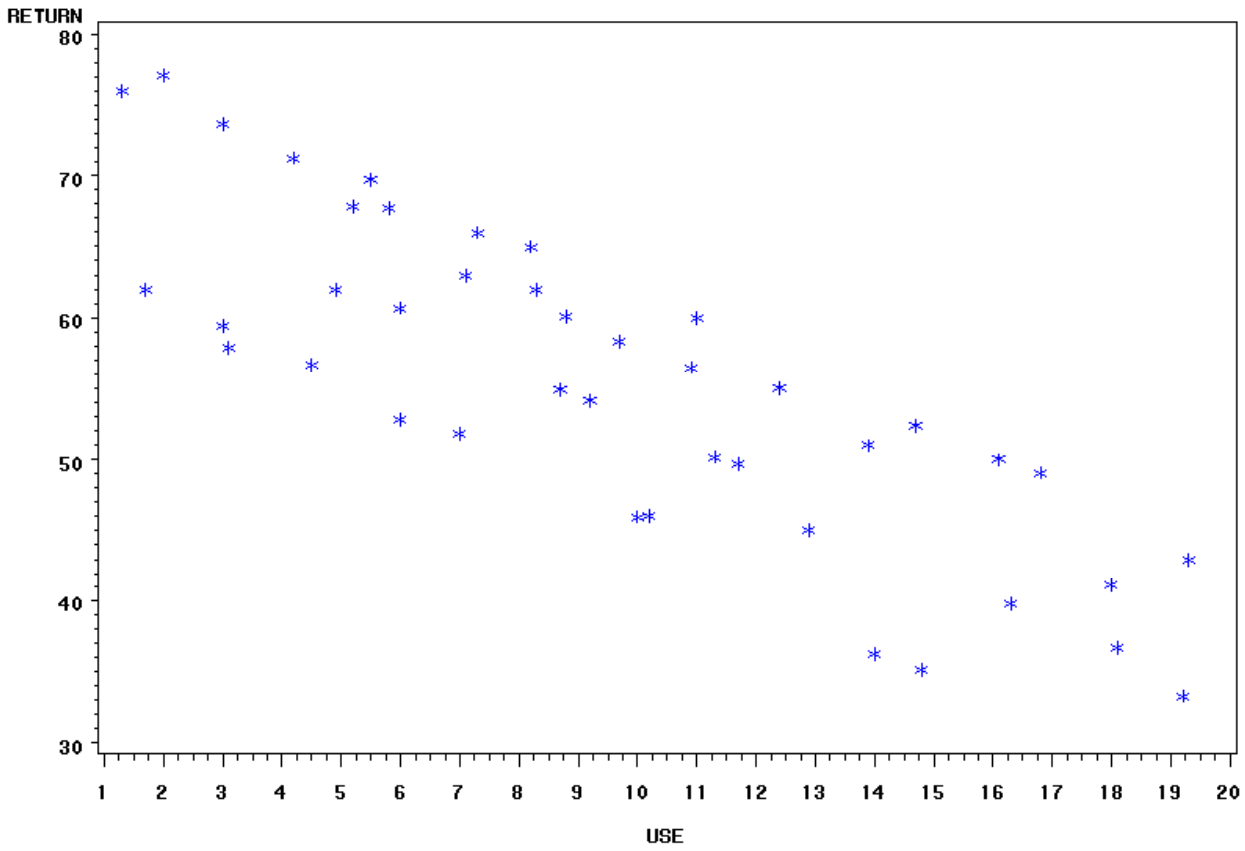
Center for Data Analysis and Statistics (CDAS)

Feb 2007

AGENDA

- The Problem
- The Traditional Approach
- The Mixed Model
- An Example
- References

The Problem…


RETURN
80
70
60
50
40
30
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
USE

Initial Model Specification

Reasonable? A straight line:

$$return_i = \beta_0 + \beta_1 \cdot use_i + \varepsilon_i$$

Where:

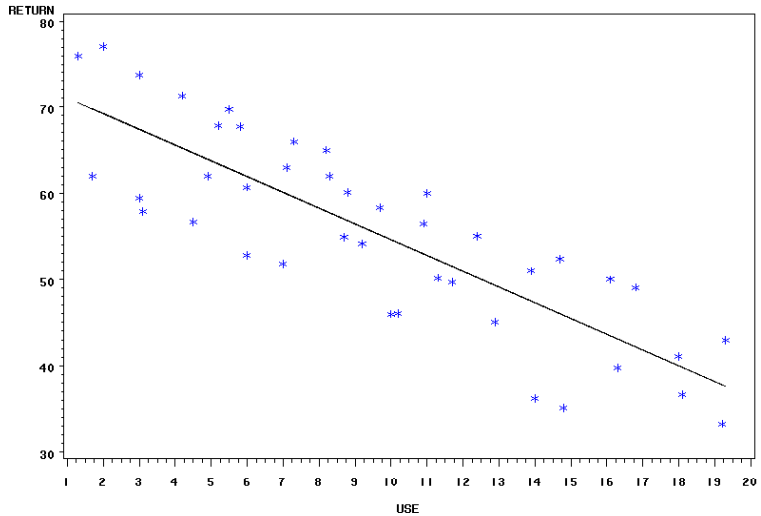
i is the subject index (I = 1,…,40)

ε_{ij} is an error term (assumed independent with mean 0 and constant variance – usually assume a normal distribution)

Initial Model Specification

Parameters estimated using least squares (or Maximum Likelihood) results in:

$$return_i = 72.9 - 1.8 \cdot use_i$$



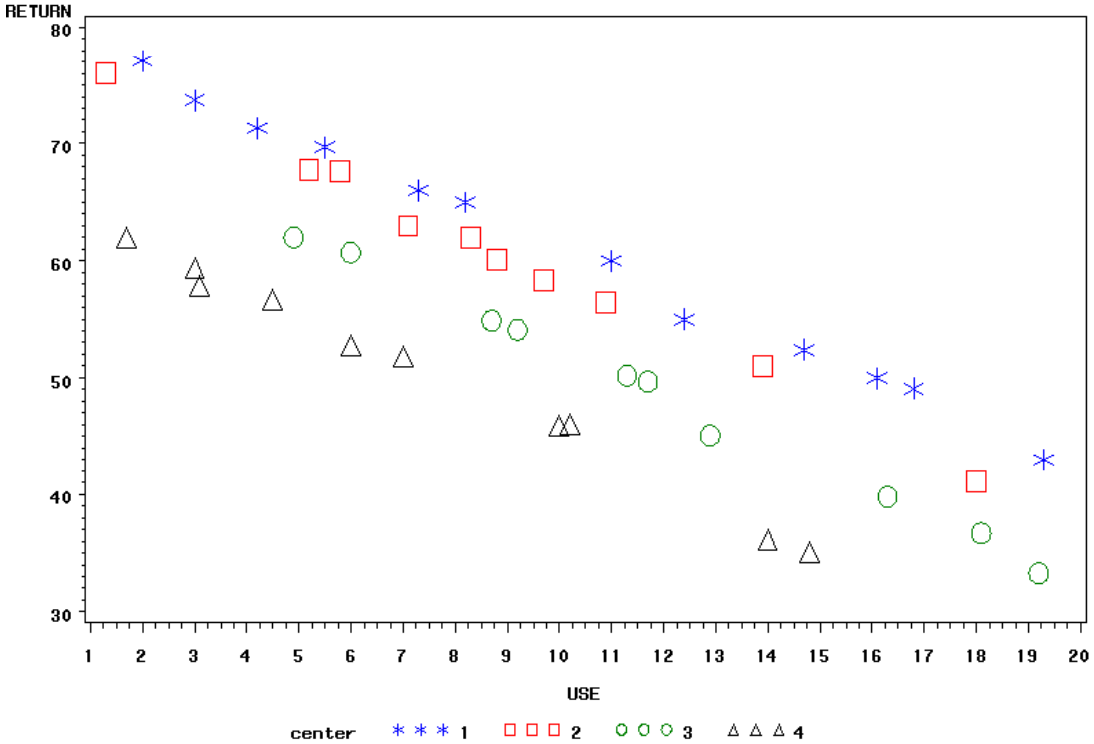
A “Good” Model?

- Regression is statistically significant ($p<0.0001$)
- Standard regression diagnostics reveal nothing suspicious.

BUT, WHAT IF…

A NEW PICTURE

The data came from 4 treatment centers:



Traditional Solution

- Use 0-1 design variables to represent the group differences
- Results in a “different line” for each group

4 Groups

3 Design variables

Center	D1	D2	D3
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

Revised Model Specification

$$return_i = \beta_0 + \beta_1 \cdot use_i$$
$$+ \beta_2 \cdot D1_i + \beta_3 \cdot D2_i + \beta_4 \cdot D3_i + \varepsilon_i$$

USING LEAST SQUARES THE ESTIMATED MODEL IS:

$$return_i = 80.8 - 2 \cdot use_i - 2.9 \cdot D1_i - 8.8 \cdot D2_i - 15.8 \cdot D3_i$$

Revised Model Specification

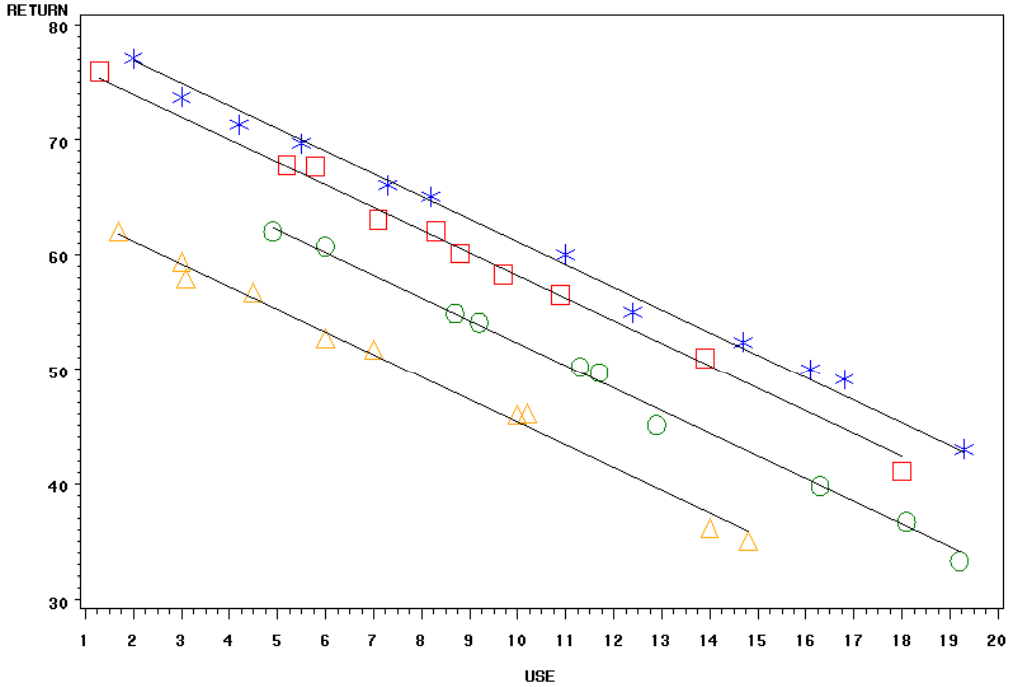
Note each center has a different estimated line:

$$return_i = 80.8 - 2 \cdot use_i - 2.9 \cdot D1_i - 8.8 \cdot D2_i - 15.8 \cdot D3_i$$

Center	Design Variables	Estimated Line
1	0, 0, 0	$return_i = 80.8 - 2 \cdot use_i$
2	1, 0, 0	$return_i = 77.9 - 2 \cdot use_i$
3	0, 1, 0	$return_i = 72.0 - 2 \cdot use_i$
4	0, 0, 1	$return_i = 65.0 - 2 \cdot use_i$

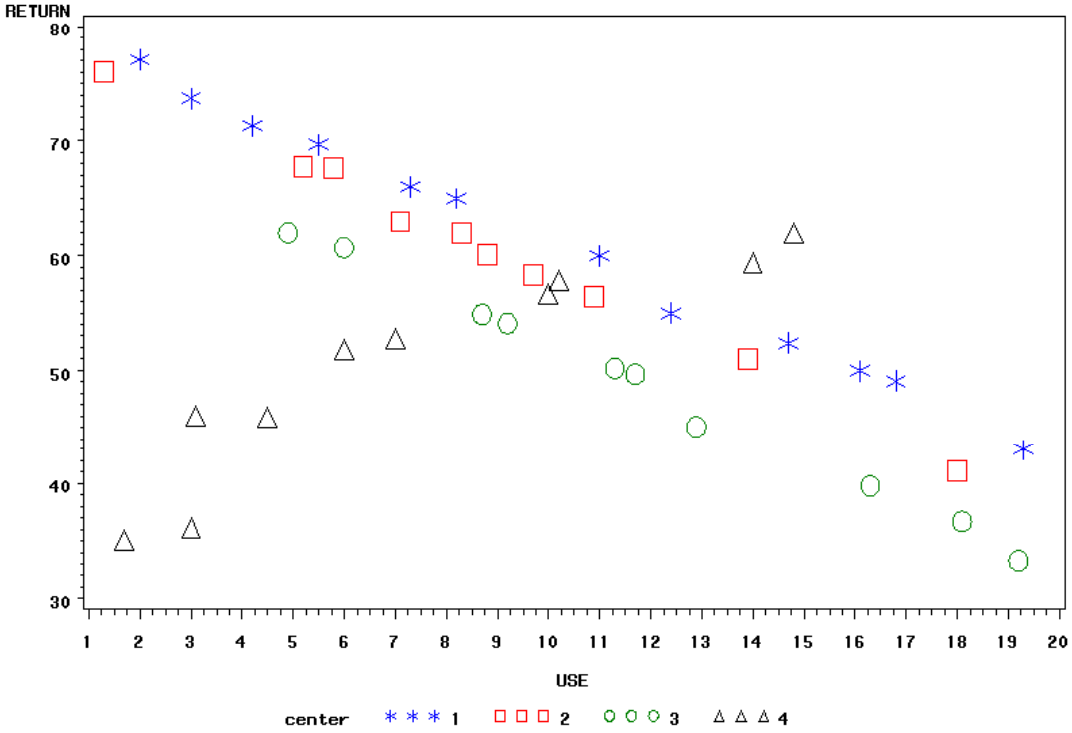
Model Results

- Excellent fit (R^2 increase from 0.69 to 0.99)
- All variables highly significant ($p < 0.0001$)



ONE MORE TWIST

Slopes are not all the same:



Revised Model Specification

ADD "use" and "D3" interaction:

$$return_i = \beta_0 + \beta_1 \cdot use_i + \beta_2 \cdot D1_i + \beta_3 \cdot D2_i + \beta_4 \cdot D3_i$$
$$+ \beta_4 \cdot D3_i \cdot use_i + \varepsilon_i$$

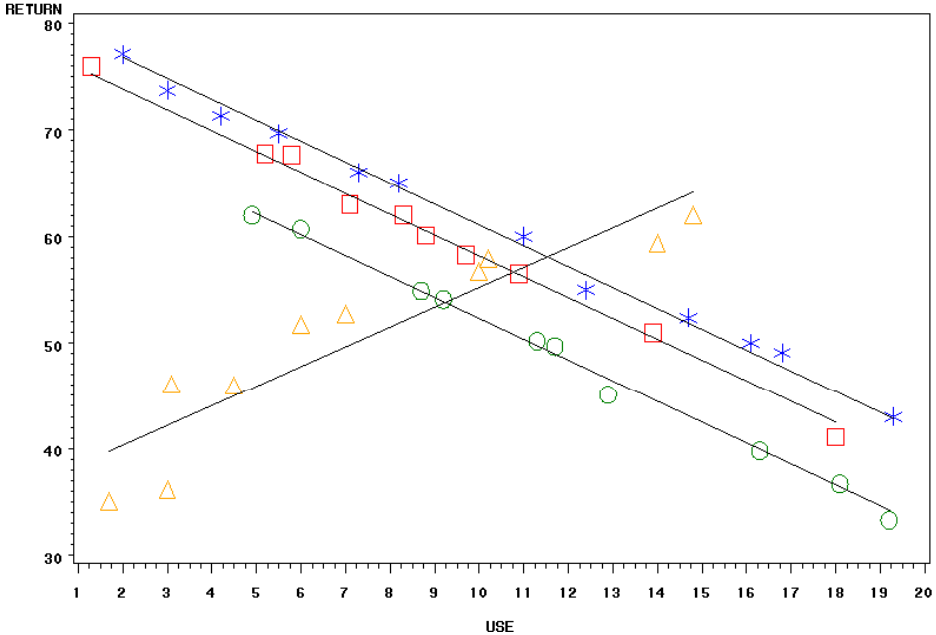
THE ESTIMATED MODEL IS:

$$return_i = 80.7 - 2 \cdot use_i - 2.9 \cdot D1_i - 8.9 \cdot D2_i - 44.1 \cdot D3_i + 3.8 \cdot D3_i \cdot use_i$$

The Center 4 Line Changes $return_i = 36.6 + 1.8 \cdot use_i$

Model Results

- Excellent fit (R^2 of 0.97)
- All variables highly significant ($p < 0.0001$ and d1 0.0015)



The Mixed Model

Q: What is wrong with the traditional model - it looked pretty good...

- Increase in the number of groups leads to large number of parameters
  - Data management and parameter estimation issues
  - Decrease in estimate precision (note: in our example, standard errors more than doubled when interaction added)
- May have a large number of predictors unique at the higher level only
- Observed “groups” are a random sample
- Independence assumption – data CORRELATED

Introduction

- Issue of clustered data
  - Common in many studies and fields
  - Standard models fail to adequately address
  - Inference and estimates affected
- Mixed models as a solution
  - Increasing use
  - Estimation algorithms and software available

Mixed Model Specification

- Random intercept only

Emphasis on hierarchy

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij}$$

where $\beta_{0j} = \beta_0 + \mu_{0j}$

Emphasis on fixed/random

$$y_{ij} = (\beta_{00} + \beta_1 x_{ij}) + (\mu_{0j} + \varepsilon_{ij})$$

$j = 1, \ldots, J$ Level 2 (group) index

$i = 1, \ldots, n_j$ Level 1 (subject) index

We assume: $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ Independent of $\mu_{0j} \sim N(0, \sigma_0^2)$

Random Slope Model

Emphasis on hierarchy

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}$$

where $\beta_{0j} = \beta_0 + \mu_{0j}$

$$\beta_{1j} = \beta_1 + \mu_{1j}$$

Emphasis on fixed/random

$$y_{ij} = (\beta_0 + \beta_1 x_{ij}) + (\mu_{0j} + \mu_{1j} x_{ij} + \varepsilon_{ij})$$

Additional assumptions: $\mu_{1j} \sim N(0, \sigma_1^2)$

$$\mathrm{cov}(\mu_{0j}, \mu_{1j}) = \sigma_{01}$$

Matrix Form of the Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

As an example, consider the matrices for the 2 level random slope model:

$$y_{ij} = (\beta_0 + \beta_1 x_{ij}) + (\mu_{0j} + \mu_{1j} x_{ij} + \varepsilon_{ij})$$

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{n_1 1} \\ y_{12} \\ \vdots \\ y_{n_2 2} \\ \vdots \\ y_{1J} \\ \vdots \\ y_{n_J J} \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} \\ \vdots & \vdots \\ 1 & x_{n_1 1} \\ 1 & x_{12} \\ \vdots & \vdots \\ 1 & x_{n_2 2} \\ \vdots & \vdots \\ 1 & x_{1J} \\ \vdots & \vdots \\ 1 & x_{n_J J} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad \mathbf{Z} = \begin{pmatrix} 1 & x_{11} & & & & & & \\ \vdots & \vdots & & & & & & \\ 1 & x_{n_1 1} & & & & & & \\ & & 1 & x_{12} & & & & \\ & & \vdots & \vdots & & & & \\ & & 1 & x_{n_2 2} & & & & \\ & & & & \ddots & & & \\ & & & & & & 1 & x_{1J} \\ & & & & & & \vdots & \vdots \\ & & & & & & 1 & x_{n_J J} \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_{01} \\ \mu_{11} \\ \vdots \\ \mu_{0J} \\ \mu_{1J} \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{n_1 1} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{n_2 2} \\ \vdots \\ \varepsilon_{1J} \\ \vdots \\ \varepsilon_{n_J J} \end{pmatrix}$$

Covariance Structure of the Linear Hierarchical Model

$$\mathrm{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z\Omega Z}' + \mathbf{W}$$

Where: $\mathbf{\Omega} = \mathrm{var}(\boldsymbol{\mu})$ and $\mathbf{W} = Var(\boldsymbol{\varepsilon})$

In the 2 level random slope example:

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} & & & \\ \sigma_{01} & \sigma_0^2 & & & \\ & & \ddots & & \\ & & & \sigma_0^2 & \sigma_{01} \\ & & & \sigma_{01} & \sigma_0^2 \end{bmatrix} \qquad \mathbf{W} = \mathrm{diag}(\sigma_\varepsilon^2)$$

Model Estimation

- Recent (10 years) computing advances allow estimation
- Maximum Likelihood (ML) Estimates
  - Iterative algorithms
  - Bayesian MCMC methods
- Software including MLwiN, HLM, Winbugs, SAS, VARCL

For Our Example

- Parameter Estimates

	Parameter	Estimate	Standard Error
Fixed	Intercept	66.9	10.2
	Slope	-1.0	0.96
Random	Level 1	4.0	0.97
	Intercept	412.8	293.6
	Slope	3.6	2.6
	Covariance	-38.1	27.3

NOTE: only estimate 6 parameters...regardless of number of groups!

Predictions

- Residual estimates (posterior means) for level *h* of model substitute parameter estimates in:

$$\mathbf{r}_h = \mathbf{R}_h' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad \text{where} \quad \mathbf{R}_h{}' = \mathbf{Z}_h \boldsymbol{\Omega}_h$$

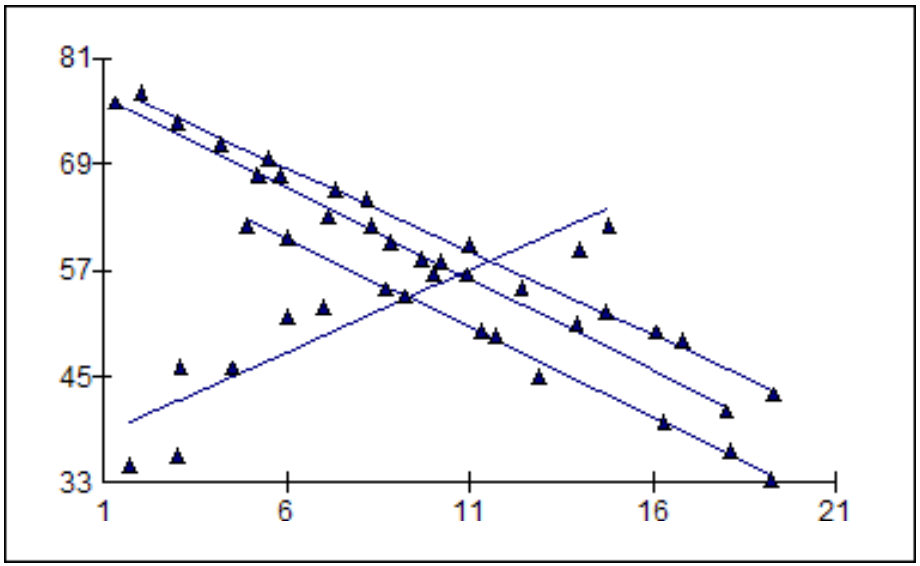
$$\operatorname{cov}(\mathbf{r}_h) = \mathbf{R}_h' \mathbf{V}^{-1} \mathbf{R}_h$$

(Design and covariance matrices for level h)

- Referred to as "shrunken" residuals. Example for level two intercept only case:

$$\hat{\boldsymbol{\mu}}_{0j} = \frac{n_j \hat{\sigma}_0^2}{n_j \hat{\sigma}_0^2 + \hat{\sigma}_\varepsilon^2} (\bar{\mathbf{y}}_{.j} - \bar{\mathbf{y}}_{..})$$

Model Results



Center (*j*)	Intercept	Slope	Std. Model
1	66.9+12.9 = 79.8	-1 - 0.86 = -1.86	80.7 -2
2	66.9+11.7 = 78.6	-1 - 1.03 = -2.03	77.8 -2
3	66.9+ 5.7 = 72.6	-1 - 0.98 = -1.98	71.8 -2
4	66.9-30.2 = 36.7	-1 + 2.87 = 1.87	36.6 1.8

Estimate intercept and slope from:

$$\hat{\beta}_{0j} = \hat{\alpha}_{00} + \hat{\tau}_{0j}$$

$$\hat{\beta}_{1j} = \hat{\alpha}_{10} + \hat{\tau}_{1j}$$

Model Extensions

There are numerous possible extensions at this point. In addition to additional predictors (with either fixed or random coefficients) at level 1...

$$return_{ij} = \beta_{0j} + \beta_{1j} use_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \alpha_{00k} + \alpha_{01} z_{0j} + \tau_{0j}$$

$$\alpha_{00k} = \alpha_{000} + \tau_{00k}$$

$$\beta_{1j} = \alpha_{10} + \alpha_{11} z_{1j} + \tau_{1j}$$

Additional level

Center specific predictor variable

Example

Batting Average in DAY/NIGHT games

- MIXED MODEL (SAS PROC MIXED)

Solution for Fixed Effects

Effect	DAY	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		0.2850	0.005203	19	54.77	<.0001
DAY	N	-0.0045	0.001027	379	-4.38	<.0001

In data: Day average = 0.28499 Night average = 0.28048

No random effect for Player

Parameter		Estimate	Error	t Value	Pr > |t|
Intercept		0.2850	0.00174983	162.87	<.0001
DAY	N	-0.0045	0.00247463	-1.82	0.0696

Random effects

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	PLAYER	0.000531
Residual		0.000105

Player	Day	Night	Model Day	Model Night
1	0.24943	0.24545	0.25004	0.24554
2	0.26905	0.26712	0.27048	0.26598
3	0.29949	0.29169	0.29772	0.29321
4	0.26852	0.26626	0.26979	0.26529

Acute Myocardial Infarction (AMI)

- Article by Austin, Tu, Alter in the American Heart Journal (JAN 2001)
- Analyzed patients admitted to Ontario hospitals between 1994 and 1999 (>100,000)
- Compared traditional to hierarchical logistic models
- 3 levels: patients, physicians and hospitals
- Separate analysis for 9 outcomes – classified as fatal, non-fatal and processes of care

Study Findings

- Found patient level variables agreed
- Hierarchical method led to different conclusions 44% of the time for hospital level factors
  - Traditional methods overestimated the statistical significance of these factors
- Traditional models tended to underestimate the magnitude of physician level factors

References

BOOKS

- Bryk and Raudenbush (1992) – good diagnostic sections
- Longford (1993) – can be tough to read; Fisher scoring
- Goldstein (1995) – details sometimes omitted, but many of technical issues are there (download for free)
- Hox (1995) – appears very introductory (download for free)
- Kreft and DeLeeuw (1998) – Nice introduction with many examples using MLn
- Snijders and Bosker (1999) – Good introduction; some math left out. Best sections on model checking/diagnostics
- Goldstein and Leyland Eds (2001) – Great for intuitive understanding; recent and good source of further references

PAPERS

- Topic dependent; I have lists...
- Useful Websites:
  - Multilevel models project (UK) - Goldstein http://multilevel.ioe.ac.uk/index.html
  - Multilevel Analysis Page (Netherlands) - Snijders http://stat.gamma.rug.nl/multilevel.htm
  - Multilevel Modeling Page (Leipzig) – Mayerhofer http://www.lrz-muenchen.de/~wlm/wlmmule.htm#Author
  - LAMMP (Michigan) – Raudenbush http://www-personal.engin.umich.edu/~gibsong/

Marginal Models

- Example is Generalized Estimating Equations (GEE)
- Primary interest in fixed parameters
- Random structure specified – “nuisance parameters”
- Hierarchical models consider random parameters of interest themselves
  - Estimate fixed and random simultaneously
  - Focus of our proposed research

Linear Hierarchical Model Comparison

- Using the Deviance
  - IGLS (ML estimation only)
  - Comparing models with different random parts must have the same fixed part

Given the maximum of the likelihood function $D_i = -2\log(L_i)$

Difference in Deviance to compare two models $D_1 - D_2 \sim \chi^2(p_2 - p_1)$

Where the number of parameters in the models $p_1 < p_2$

- Wald Tests of Fixed Parameters
  - Concern about appropriateness of assymptotic standard normal distribution

Test of hypothesis $H_0 : \beta_h = 0$

$$T(\beta_h) = \frac{\hat{\beta}_h}{\mathrm{SE}(\hat{\beta}_h)}$$

Linear Hierarchical Model Comparison

- Wald test in small samples compare to a t-distribution
- Random parameters (Bryk & Raudenbush)
  - Obtain LS estimates of parameter within each group and use chi-square test of equality across groups
- Multivariate Wald test available for a group of parameters
- Other tests all appear to have problems and/or are not available in most software packages