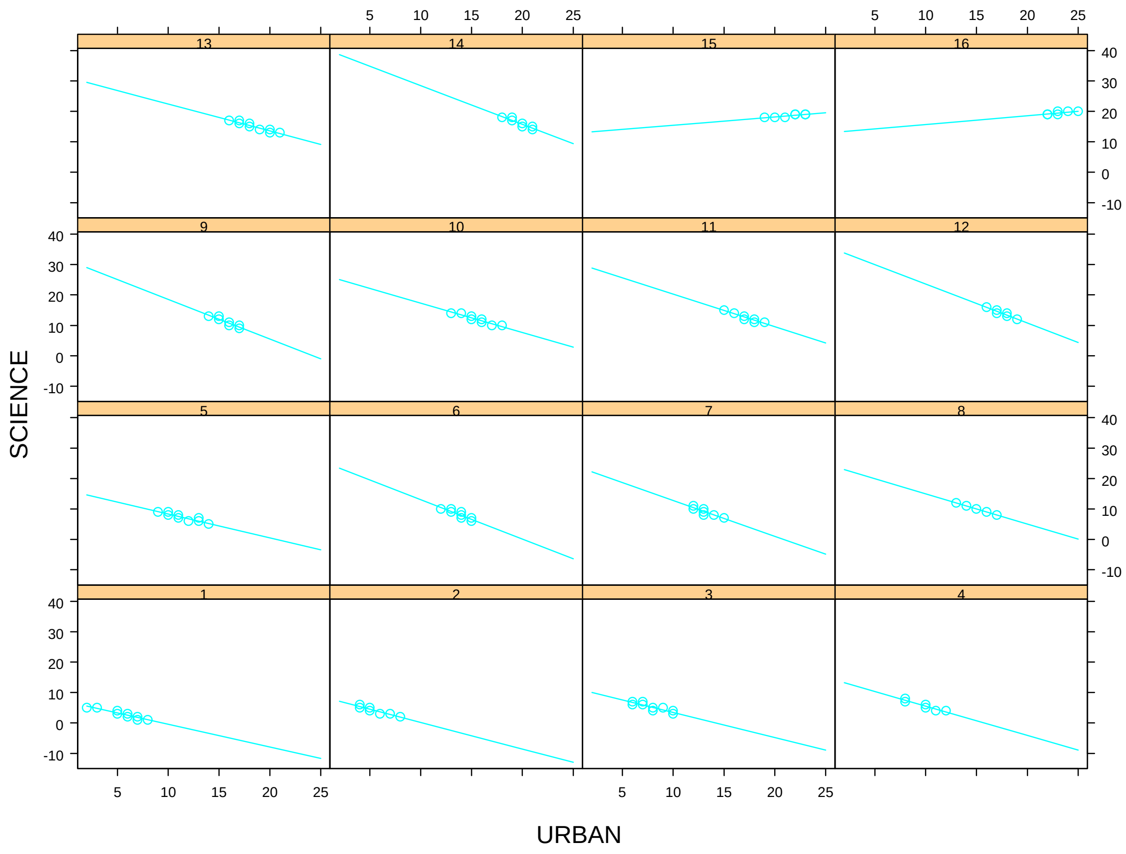
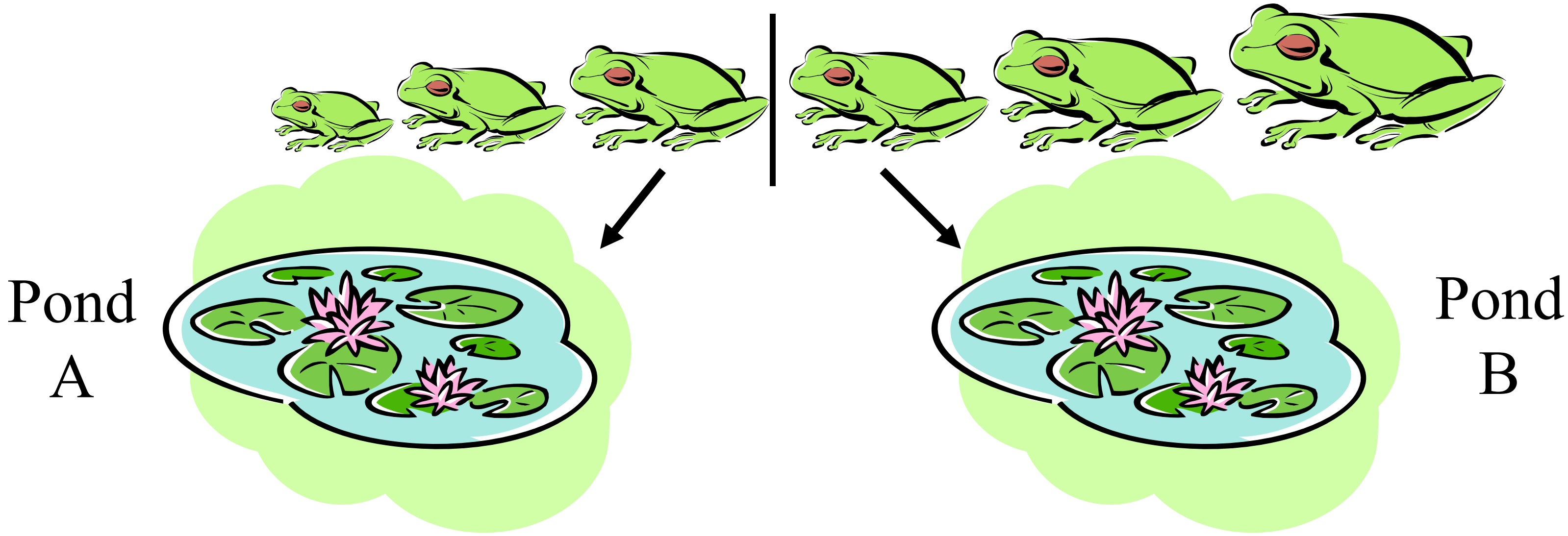
Introduction to Hierarchical Linear Modeling with R



First Things First

- Robinson (1950) and the problem of contextual effects
- The "Frog-Pond" Theory



A Brief History of Multilevel Models

- Nested ANOVA designs
- Problems with the ANCOVA design
  - "Do schools differ" vs. "Why schools differ?"
  - ANCOVA does not correct for intra-class correlation (ICC)

Strengths of Multilevel Models

- Statistical models that are not hierarchical sometimes ignore structure and report underestimated standard errors
- Multilevel techniques are more efficient than other techniques
- Multilevel techniques assume a general linear model and can perform all types of analyses

Multilevel Examples

- Students nested within classrooms
- Students nested within schools
- Students nested within classrooms within schools
- Measurement occasions nested within subjects (repeated measures)
- Students cross-classified by school and neighborhood
- Students having multiple membership in schools (longitudinal data)
- Patients within a medical center
- People within households

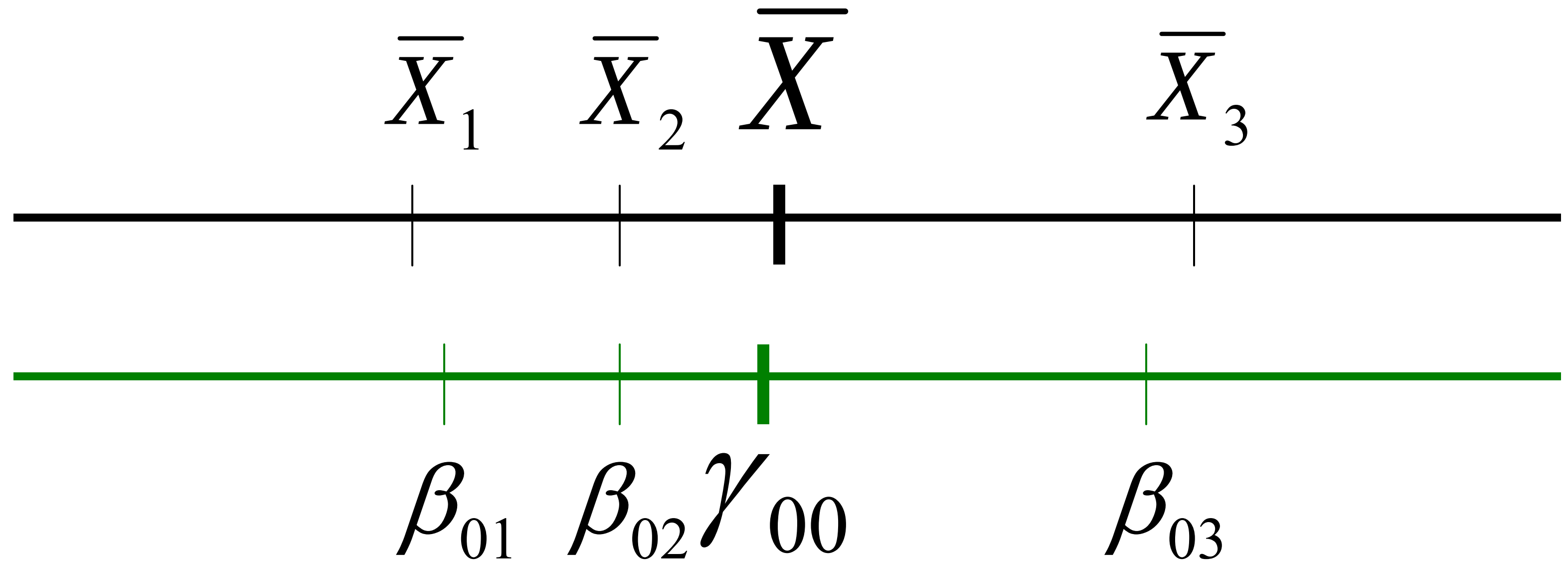
Children Nested In Families!!

Do we really need HLM/MLM?

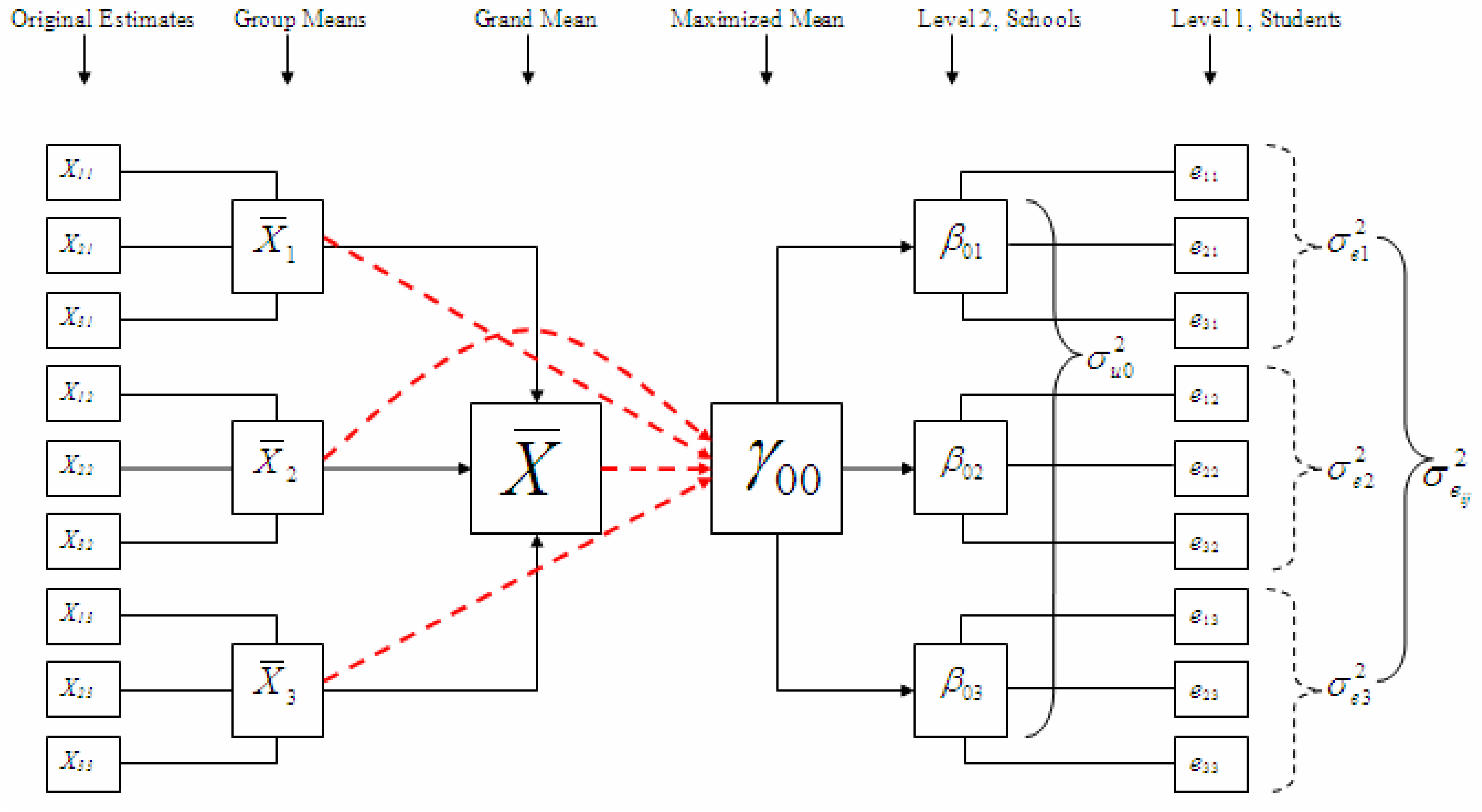
- “All data are multilevel!”
- The problem of independence of observations
- The “inefficiency” of OLS techniques

Differences in HLM and Other Methods

- HLM is based on Maximum Likelihood and Empirical Bayesian estimation techniques
- 1 + 1 = 1.5



Graphical Example of Multilevel ANOVA



Notating the HLM ANOVA

- The full model would be:

$$y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

- Level-1 model is:

$$y_{ij} = \beta_{0j} + e_{ij}$$

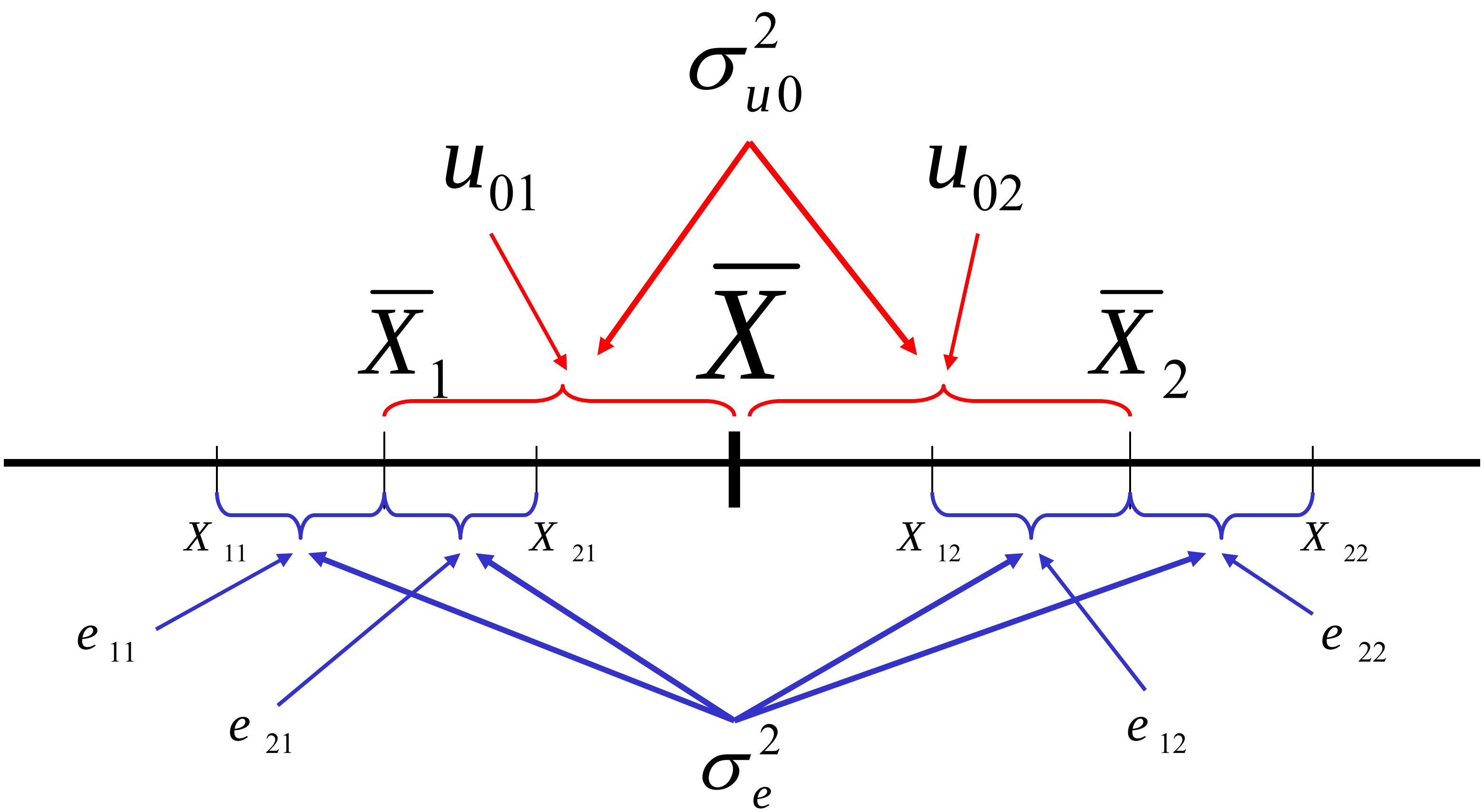
$$\begin{cases} y_{11} = \beta_1 + e_{11} \\ y_{21} = \beta_1 + e_{21} \\ \cdots \\ y_{ij} = \beta_j + e_{ij} \end{cases}$$

- Level-2 model is:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

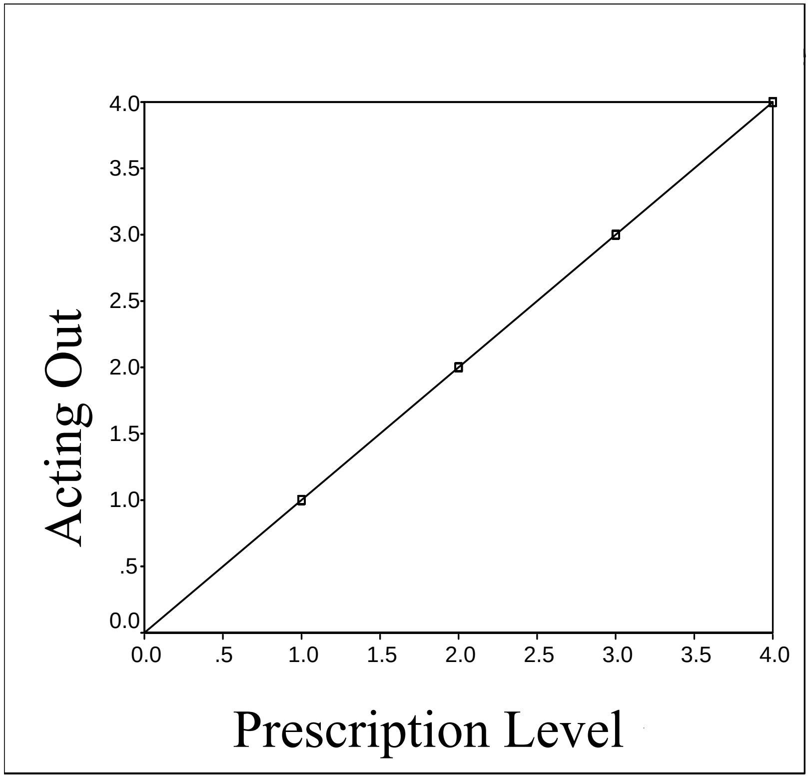
$$\begin{cases} \beta_1 = \gamma_{00} + u_1 \\ \beta_2 = \gamma_{00} + u_2 \\ \cdots \\ \beta_j = \gamma_{00} + u_j \end{cases}$$

Understanding Errors



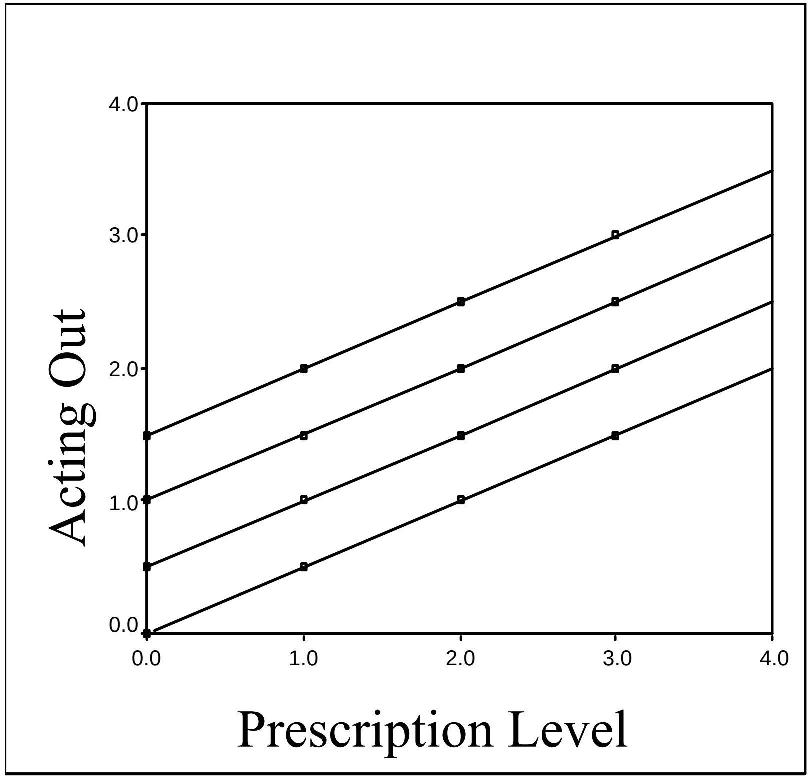
Fixed vs. Random Coefficients

- Fixed Slopes and Intercepts



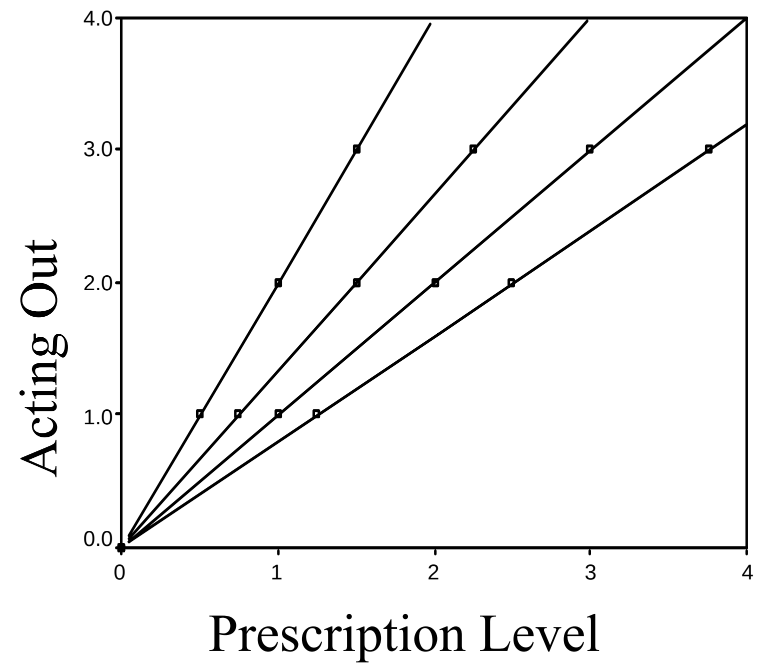
Fixed vs. Random Coefficients

- Fixed Slopes and Intercepts
- Random Intercepts and Fixed Slopes



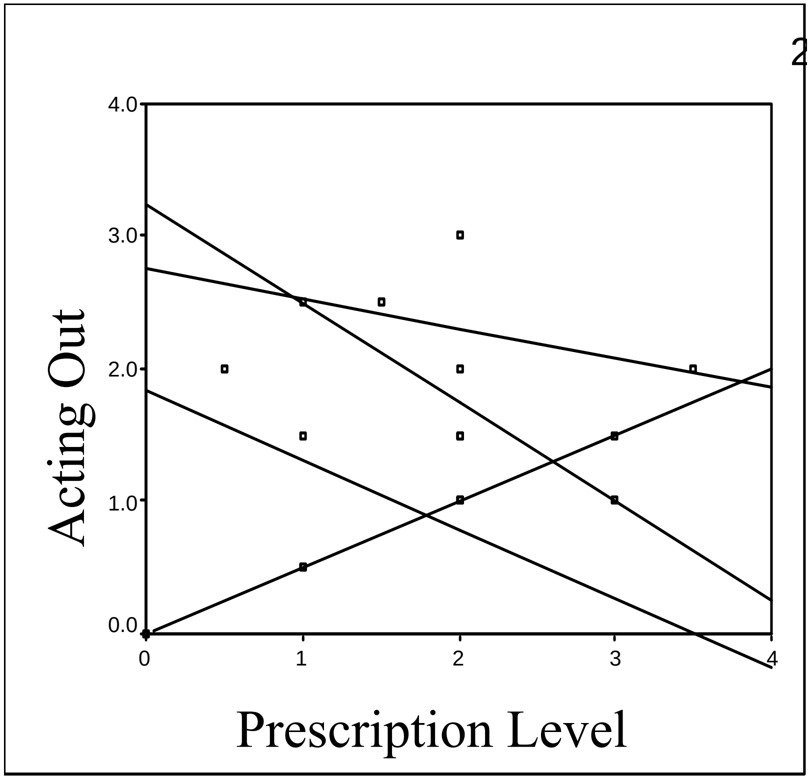
Fixed vs. Random Coefficients

- Fixed Slopes and Intercepts
- Random Intercepts and Fixed Slopes
- Fixed Intercepts and Random Slopes



Fixed vs. Random Coefficients

- Fixed Slopes and Intercepts
- Random Intercepts and Fixed Slopes
- Fixed Intercepts and Random Slopes
- Random Slopes and Intercepts



Let's Give This A Shot!!!

- An example where we use a child's level of "urbanicity" (a SES composite) to predict their science achievement
- Start with Multilevel ANOVA (also called the "null model")

$$science_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

Grand mean

Group deviation

Individual diff.

Intraclass Correlation

- The proportion of total variance that is *between* the groups of the regression equation
- "The degree to which individuals share common experiences due to closeness in space and/or time" Kreft & de Leeuw, 1998.
- a.k.a – ICC is the proportion of group-level variance to the total variance
- LARGE ICC DOES NOT EQUAL LARGE DIFFERENCES BETWEEN MLM AND OLS (Roberts, 2002)
- Formula for ICC: $$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{e0}^2}$$

Statistical Significance???

- Chi-square vs. degrees of freedom in determining model fit
- The problem with the df
- Can also compute statistical significance of variance components (only available in some packages)

The Multilevel Model – Adding a Level-1 Predictor

- Consider the following 1-level regression equation:
  - $y = a + bx + e$
    - y = response variable
    - a = intercept
    - b = coefficient of the response variable (slope)
    - x = response variable
    - e = residual or error due to measurement

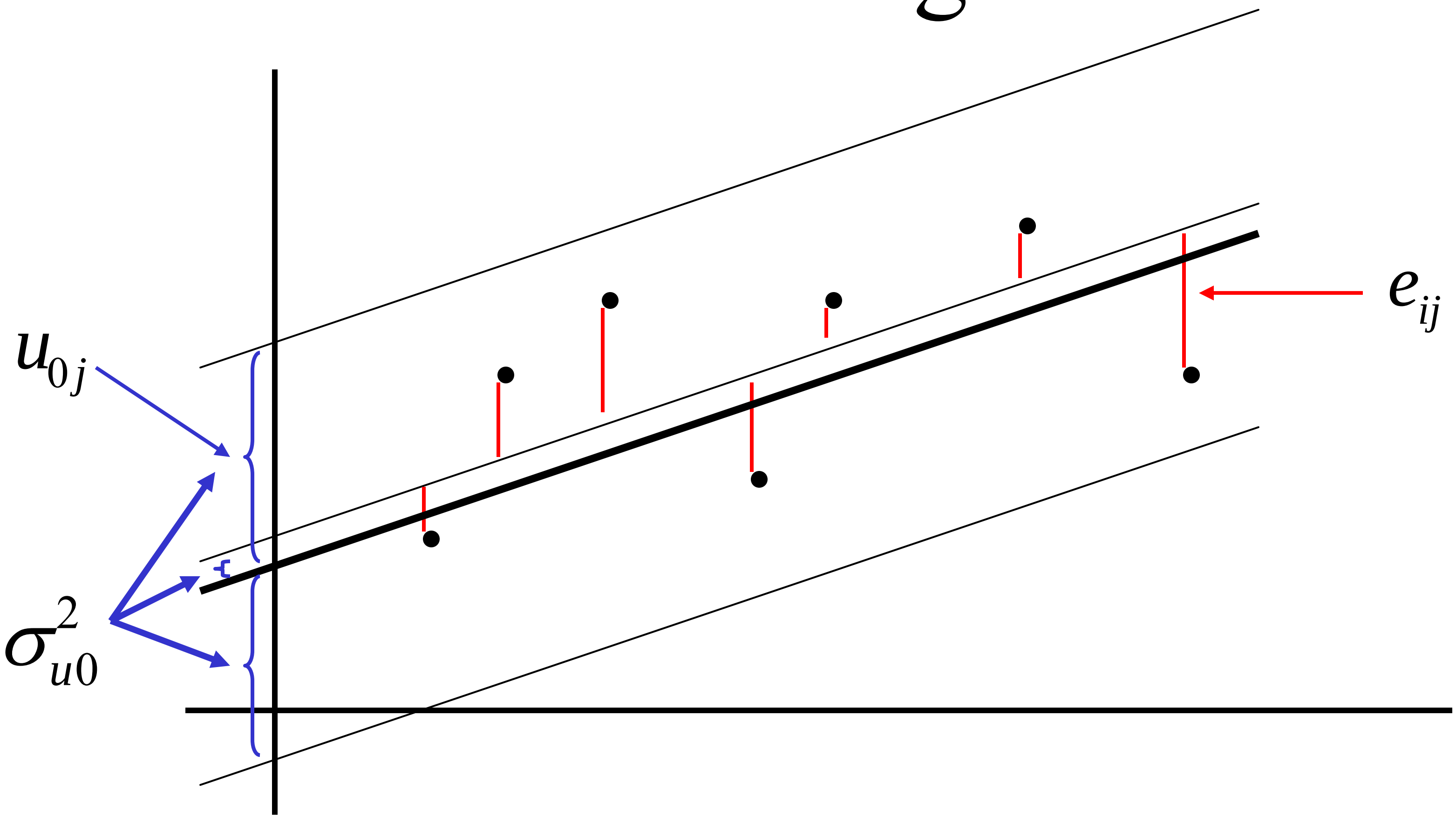
The Multilevel Model (2)

- The fixed coefficients multilevel model is a slight variation on the OLS regression equation:
  - $y_{ij} = a + bx_{ij} + u_j + e_{ij}$
    - Where "i" defines level-1, "j" defines level-2, u_j is the level-2 residual and e_{ij} is the level-1 residual
- Using slightly different annotation we can transform the above equation to:
  - $y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + e_{ij}$
    - Where γ_{00} now defines the constant/intercept "a" and γ_{10} defines the slope

The Multilevel Model (3)

- From the previous model:
  - $y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + e_{ij}$
- We can then transform this model to:
  - $y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + e_{ij}$ ← Level-1 Model
  - $\beta_{0j} = \gamma_{00} + u_{0j}$ ← Level-2 Model
  - $\beta_{1j} = \gamma_{10}$
  - With variances $u_{0j} = \sigma^2_{u0}$ $e_{ij} = \sigma^2_{e_{ij}}$

Understanding Errors



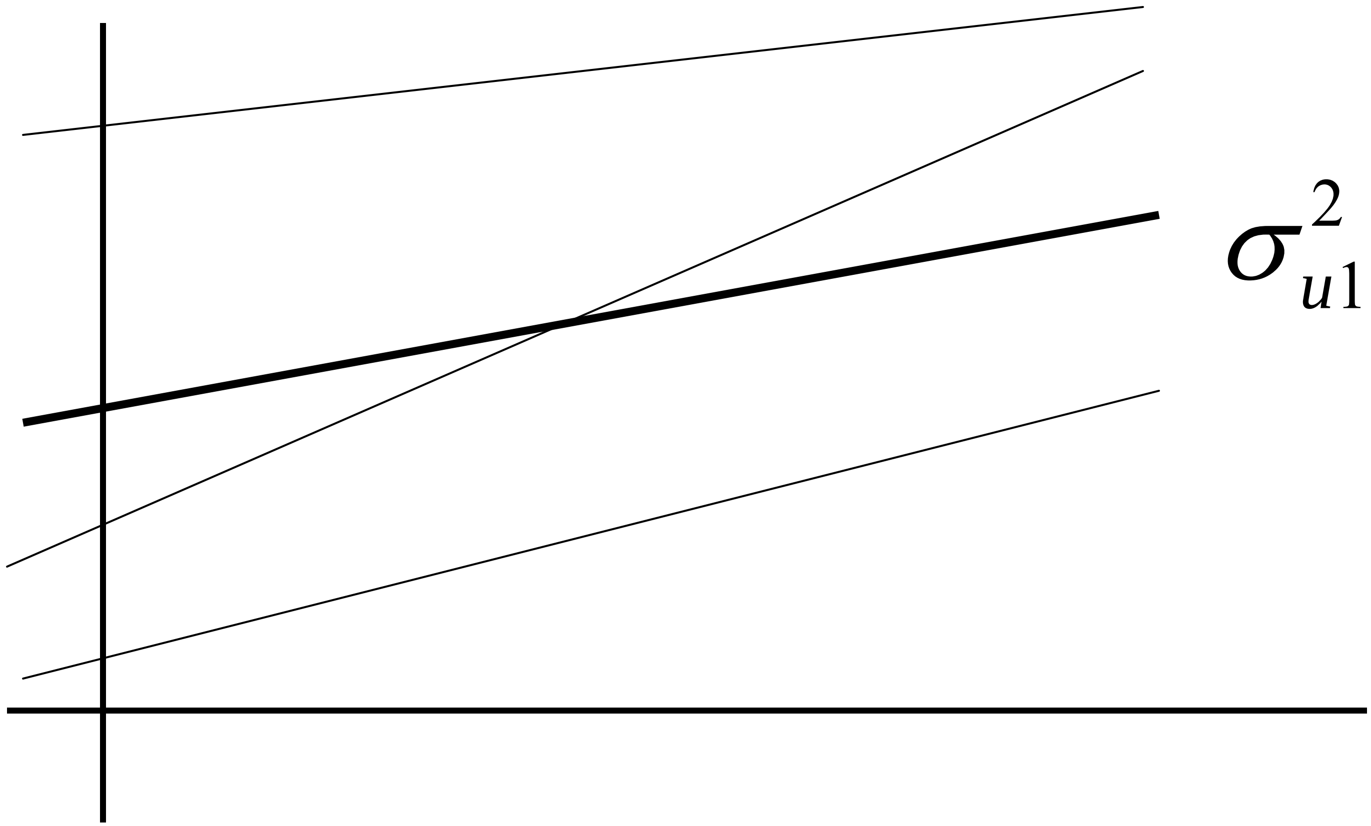
Adding a Random Slope Component

- Suppose that we have good reason to assume that it is inappropriate to "force" the same slope for "urbanicity" on each school
  - Level-1 Model → $y_{ij} = \beta_{0j}x_0 + \beta_{1j}x_{1ij} + r_{ij}$
  - Level-2 Model → $\beta_{0j} = \gamma_{00} + u_{0j}$

    $\beta_{1j} = \gamma_{10} + u_{1j}$

Complete Model → $science_{ij} = \gamma_{00} + u_{0j} + (\gamma_{10} + u_{1j})urban + r_{ij}$

Understanding Errors Again



σ_{u1}^2 ????

Model Fit Indices

- Chi-square $-2 * \ell$
- Akaike Information Criteria

$$AIC = -2 * \ell + 2K$$

- Bayesian Informaiton Criteria

$$BIC = -2 * \ell + K * Ln(N)$$

To Center or Not to Center

- In regression, the intercept is interpreted as the expected value of the outcome variable, when all explanatory variables have the value zero
- However, zero may not even be an option in our data (e.g., Gender)
- This will be especially important when looking at cross-level interactions
- General rule of thumb: If you are estimating cross-level interactions, you should grand mean center the explanatory variables.

An Introduction to R

R as a Big Calculator

- Language was originally developed by AT&T Bell Labs in the 1980's
- Eventually acquired by MathSoft who incorporated more of the functionality of large math processors
- The commands window is like a big calculator

```
> 2+2
[1] 4
```

```
> 3*5+4
[1] 19
```

Object Oriented Language

- A Big Plus for R is that it utilizes object oriented language.

```
> x<-2+4
> x
[1] 6

> y<-3+5
> y
[1] 8

> x+y
[1] 14
```

```
> x<-1:10
> x
 [1]  1  2  3  4  5  6
7  8  9 10
> mean(x)
[1] 5.5
> 2*x
 [1]  2  4  6  8 10 12 14 16
18 20
> x^2
 [1]   1   4   9  16  25  36
49  64  81 100
```

Utilizing Functions in R

- R has many "built in" functions (c.f., "Language Reference" in the "Help" menu)
- Functions are commands that contain "arguments"
- `seq` function has 4 arguments
  - `seq(from, to, by, length.out, along.with)`

```
> ?seq
> seq(from=1, to=100, by=10)
 [1]  1 11 21 31 41 51 61 71
81 91
```

```
> seq(1, 100, 10)
> seq(1, by=10,
length=4)
[1]  1 11 21 31
```

Making Functions in R

```
> squared<-function(x){x^2}
> squared(5)
[1] 25

> inverse<-function(x){1/x}
> num<-c(1,2,3,4,5)
> inverse(num)
[1] 1.0000000 0.5000000 0.3333333
0.2500000 0.2000000
```

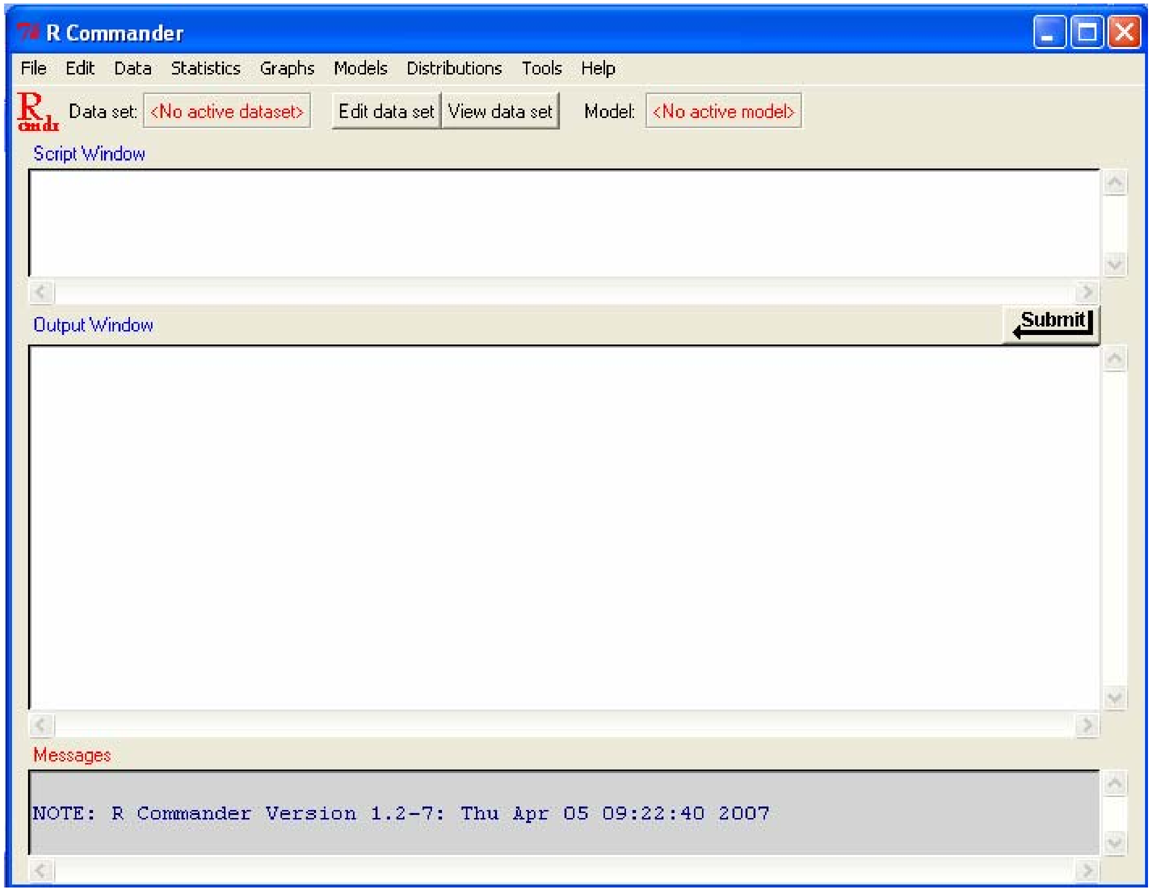
Sampling from a Single Variable

- The `sample` function is used to draw a random sample from a single vector of scores
- `sample(x, size, replace, prob)`
  - `x =` dataset
  - `size =` size of the sample to be drawn
  - `replace =` toggles on and off sampling with replacement (default = F)
  - `prob =` vector of probabilities of same length as x

Sampling a Vector (cont.)

```
> x<-1:30
> sample(x, 10, replace=T)
 [1]  8 14 27  2 30 16  4  9  9  2

> x<-1:5
> sample(x, 10, replace=T)
 [1] 3 2 2 3 3 4 1 1 2 3
```

Rcmdr – library(Rcmdr)



Reading a Dataset

```
> example<-read.table(file, header=T)
> example<-read.table(“c:/aera/example.txt”, header=T)
> head(example)
```

Science Achievement/Urbanicity

- Back to the Multilevel ANOVA example, let's perform an OLS using "urbanicity" to predict science achievement

```
> summary(lm(SCIENCE~URBAN, example))

Call:
lm(formula = SCIENCE ~ URBAN, data = example)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3358 -2.1292  0.4919  2.0432  5.0090

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.25108    0.59371  -2.107   0.0367 *
URBAN        0.82763    0.03863  21.425   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.592 on 158 degrees of freedom
Multiple R-Squared: 0.7439,     Adjusted R-squared: 0.7423
F-statistic:   459 on 1 and 158 DF,  p-value: < 2.2e-16
```

$$\hat{y} = -1.25 + 0.83(urban) + r$$

R's Functionality

Try these:

```
> linear<-lm(SCIENCE~URBAN, example)
> summary(linear)
> plot(SCIENCE~URBAN, example)
> abline(linear)
> plot(linear)
> names(linear)
> linear$coefficients
```

lmer – `library(lme4)`

lmer(lme4)　　　　R Documentation

Fit (Generalized) Linear Mixed-Effects Models

Description

Fit a linear or generalized linear mixed-effects model with nested or crossed grouping factors for the random effects.

Usage

```
lmer(formula, data, family, method, control, start,
     subset, weights, na.action, offset, contrasts,
     model, ...)
lmer2(formula, data, family, method, control, start,
      subset, weights, na.action, offset, contrasts,
      model, ...)
```

Arguments

`formula`	a two-sided linear formula object describing the fixed-effects part of the model, with the response on the left of a ~ operator and the terms, separated by + operators, on the right. The vertical bar character "`|`" separates an expression for a model matrix and a grouping factor.
`data`	an optional data frame containing the variables named in `formula`. By default the variables are taken from the environment from which `lmer` is called.
`family`	a GLM family, see `glm`. If `family` is missing then a linear mixed model is fit; otherwise a generalized linear mixed model is fit.

Running lmer for "example"

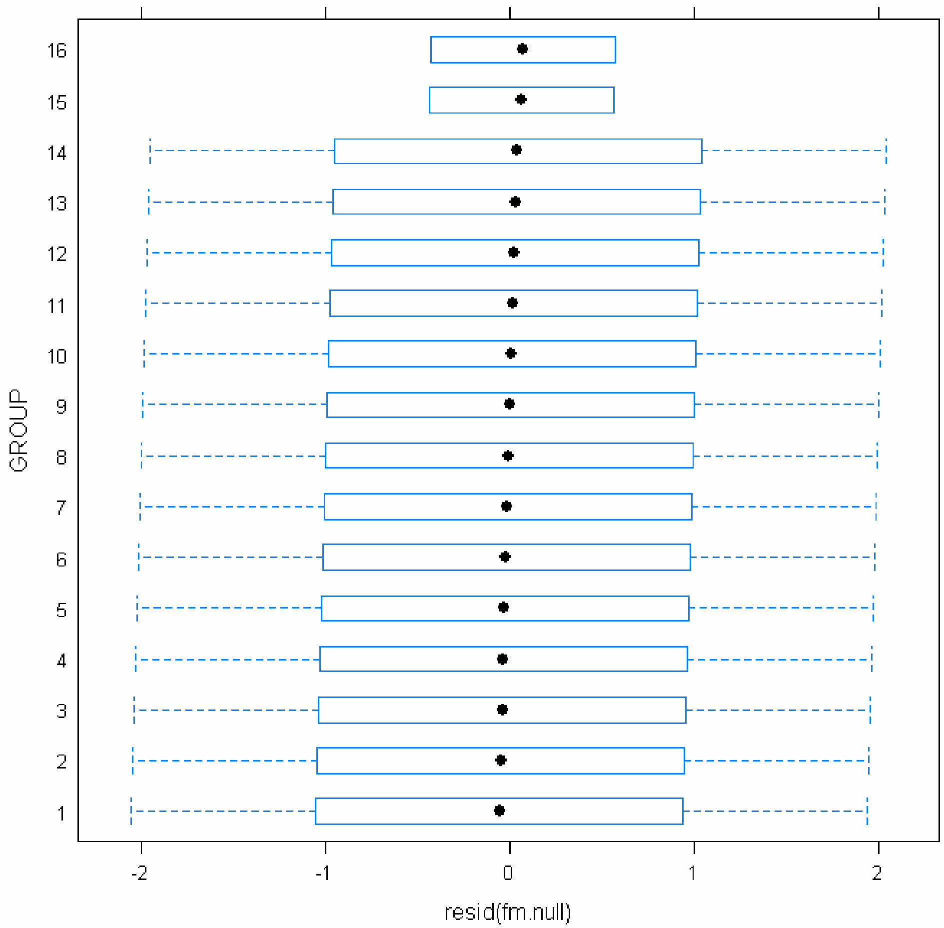
```
> fm.null<-lmer(SCIENCE~1 + (1|GROUP), example)
> summary(fm.null)
Linear mixed-effects model fit by REML
Formula: SCIENCE ~ 1 + (1 | GROUP)
   Data: example
   AIC BIC logLik MLdeviance REMLdeviance
 641.9 648 -318.9      640.2        637.9
Random effects:
 Groups   Name        Variance Std.Dev.
 GROUP    (Intercept) 25.5312  5.0528
 Residual              1.9792  1.4068
number of obs: 160, groups: GROUP, 16

Fixed effects:
            Estimate Std. Error t value
(Intercept)   10.687      1.268   8.428
```

Notice: No p-values!!

$$SCIENCE = \gamma_{00} + u_{0j} + e_{ij}$$

$$SCIENCE = 10.687 + u_{0j} + e_{ij}$$

```
bwplot(GROUP~resid(fm.null), example)
```



Examining Empirical Bayesian Estimated Group Means

```
> with(example, tapply(SCIENCE, GROUP, mean))
> coef(fm.null)
```

Group	OLS Est.	EB Est.	Group	OLS Est.	EB Est.
1	3.00	3.06	9	11.00	11.00
2	4.00	4.05	10	12.00	11.99
3	5.00	5.04	11	13.00	12.98
4	6.00	6.04	12	14.00	13.97
5	7.00	7.03	13	15.00	14.97
6	8.00	8.02	14	16.00	15.96
7	9.00	9.01	15	18.50	18.44
8	10.00	10.01	16	19.50	19.43

Adding our "Urban" Predictor

```
> fm1<-lmer(SCIENCE~URBAN + (1|GROUP), example)
> summary(fm1)
Linear mixed-effects model fit by REML
Formula: SCIENCE ~ URBAN + (1 | GROUP)
   Data: example
   AIC   BIC logLik MLdeviance REMLdeviance
 506.1 515.3 -250.0      499.4        500.1
Random effects:
 Groups   Name        Variance Std.Dev.
 GROUP    (Intercept) 86.45595 9.29817
 Residual              0.65521 0.80945
number of obs: 160, groups: GROUP, 16

Fixed effects:
            Estimate Std. Error t value
(Intercept)  22.3029     2.4263   9.192
URBAN        -0.8052     0.0480 -16.776

Correlation of Fixed Effects:
      (Intr)
URBAN -0.285
```

coef(fm1)

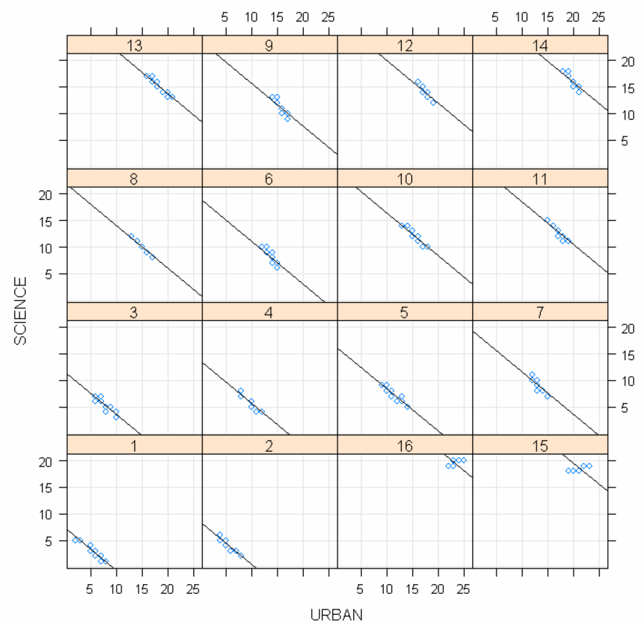
```
An object of class "coef.lmer"
[[1]]
   (Intercept)      URBAN
1     7.359555 -0.8052278
2     8.519721 -0.8052278
3    11.530509 -0.8052278
4    13.656217 -0.8052278
5    16.425619 -0.8052278
6    19.195021 -0.8052278
7    19.631031 -0.8052278
8    22.078586 -0.8052278
9    23.721524 -0.8052278
10   24.479381 -0.8052278
11   26.524627 -0.8052278
12   28.087102 -0.8052278
13   29.810501 -0.8052278
14   31.936209 -0.8052278
15   35.641243 -0.8052278
16   38.249722 -0.8052278
```

Comparing Models

```
> anova(fm.null, fm1)
Data: example
Models:
fm.null: SCIENCE ~ 1 + (1 | GROUP)
fm1: SCIENCE ~ URBAN + (1 | GROUP)
        Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
fm.null  2  644.17  650.32 -320.08
fm1      3  505.37  514.60 -249.69 140.79      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Graphing

```
with(fm1, {
    cc <- coef(.)$GROUP
    xyplot(SCIENCE ~ URBAN | GROUP,
                 index.cond = function(x, y) coef(lm(y ~ x))[1],
                 panel = function(x, y, groups, subscripts, ...) {
                     panel.grid(h = -1, v = -1)
                     panel.points(x, y, ...)
                     subj <- as.character(GROUP[subscripts][1])
                     panel.abline(cc[subj,1], cc[subj, 2])
                 })
})
```



Adding a Random Coefficient

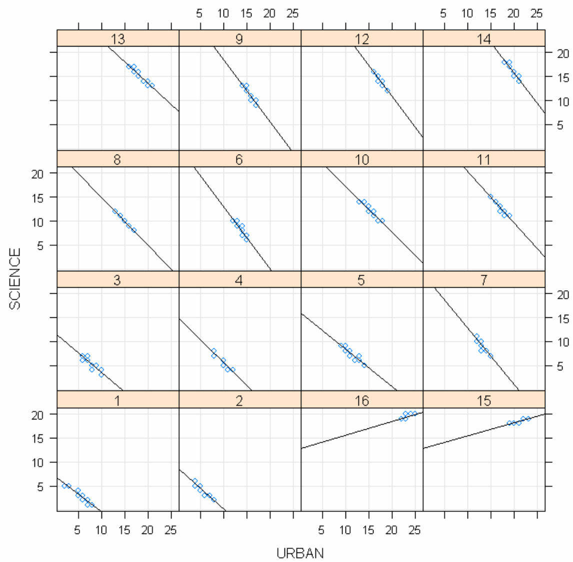
```
> fm2<-lmer(SCIENCE~URBAN + (URBAN|GROUP), example)
> summary(fm2)
Linear mixed-effects model fit by REML
Formula: SCIENCE ~ URBAN + (URBAN | GROUP)
   Data: example
   AIC   BIC logLik MLdeviance REMLdeviance
 422.2 437.5 -206.1      413.2        412.2
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 GROUP    (Intercept) 113.65372 10.66085
          URBAN         0.25204  0.50204 -0.626
 Residual               0.27066  0.52025
number of obs: 160, groups: GROUP, 16

Fixed effects:
            Estimate Std. Error t value
(Intercept)  22.3913     2.7176   8.239
URBAN        -0.8670     0.1298  -6.679

Correlation of Fixed Effects:
      (Intr)
URBAN -0.641
```

coef(fm2)

```
An object of class "coef.lmer"
[[1]]
   (Intercept)      URBAN
1     7.038437 -0.7468619
2     8.901233 -0.8742708
3    11.668557 -0.8227891
4    15.130206 -0.9607240
5    16.185638 -0.7849262
6    26.029030 -1.2969941
7    24.550979 -1.1780460
8    24.894060 -0.9929589
9    31.570587 -1.3020201
10   26.967287 -0.9657138
11   30.982799 -1.0705419
12   36.360597 -1.2779387
13   31.267775 -0.8843021
14   41.202393 -1.2731132
15   12.723743  0.2710856
16   12.787665  0.2879866
```

Graphing

```
with(fm2, {
    cc <- coef(.)$GROUP
    xyplot(SCIENCE ~ URBAN | GROUP,
                 index.cond = function(x, y) coef(lm(y ~ x))[1],
                 panel = function(x, y, groups, subscripts, ...) {
                     panel.grid(h = -1, v = -1)
                     panel.points(x, y, ...)
                     subj <- as.character(GROUP[subscripts][1])
                     panel.abline(cc[subj,1], cc[subj, 2])
                 })
})
```



Comparing Models Again

```
> anova(fm.null, fm1, fm2)
Data: example
Models:
fm.null: SCIENCE ~ 1 + (1 | GROUP)
fm1: SCIENCE ~ URBAN + (1 | GROUP)
fm2: SCIENCE ~ URBAN + (URBAN | GROUP)
        Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
fm.null  2  644.17  650.32 -320.08
fm1      3  505.37  514.60 -249.69 140.79      1  < 2.2e-16 ***
fm2      5  423.22  438.60 -206.61  86.15      2  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing Three Models
Which Model is Best?

	M_0: Null model		M_1: + Urbanicity		M_2: + random est.	
Fixed Effects:	estimate	s.e.	estimate	s.e.	estimate	s.e.
Intercept	10.688	1.268	22.303	2.426	22.391	2.717
Urbanicity			-0.805	0.048	-0.867	0.130
Random Effects:						
σ_{ε}^2	1.979		0.655		0.271	
σ_{u0}^2	25.531		86.456		113.603	
σ_{u1}^2					0.253	
σ_{u01}					-3.344	
Fit:						
X^2	637.856		500.094		412.171	
AIC	643.856		508.094		424.171	
BIC	653.063		520.344		442.547	

OLS vs. EB Random Estimates

Notice that estimates that are further away from our grand slope estimate (-0.87) are "shrunk" further back to the mean.

Group	OLS Est.	EB Est.	Group	OLS Est.	EB Est.
1	-0.76	-0.75	9	-1.35	-1.30
2	-0.89	-0.87	10	-0.97	-0.97
3	-0.83	-0.82	11	-1.09	-1.07
4	-0.97	-0.96	12	-1.33	-1.28
5	-0.78	-0.78	13	-0.89	-0.88
6	-1.35	-1.30	14	-1.35	-1.27
7	-1.21	-1.18	15	***0.32***	***0.27***
8	-1.00	-0.99	16	***0.37***	***0.28***

R^2 in HLM

Level-1 Equation

$$R_1^2 = \frac{\sigma^2_{e|b} - \sigma^2_{e|m}}{\sigma^2_{e|b}}$$

$$R_1^2 = \frac{1.979 - 0.271}{1.979} = .863$$

Level-2 Equation

$$R_2^2 = \frac{\sigma^2_{u0|b} - \sigma^2_{u0|m}}{\sigma^2_{u0|b}}$$

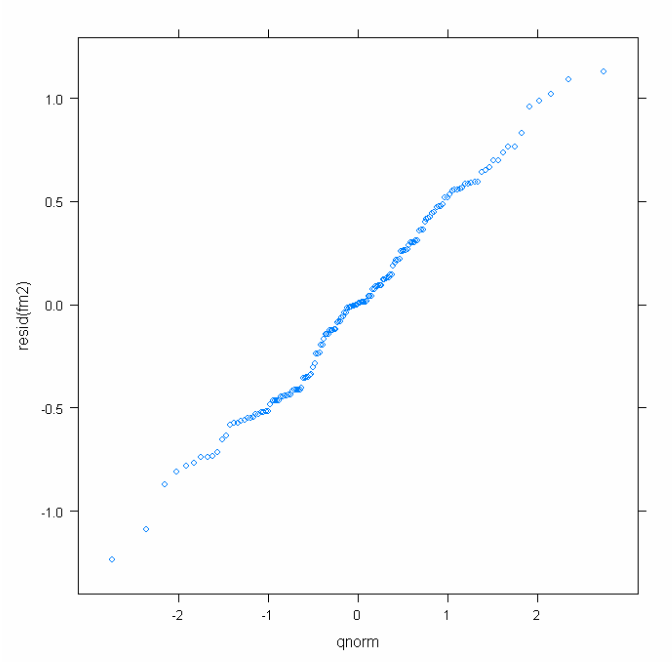
$$R_2^2 = \frac{25.531 - 86.456}{25.531} = -2.386$$

??? -238% Variance Explained???

Distributional Assumptions

- Level-1 errors are independent and identically normally distributed with a mean of 0 and a variance σ^2 in the population

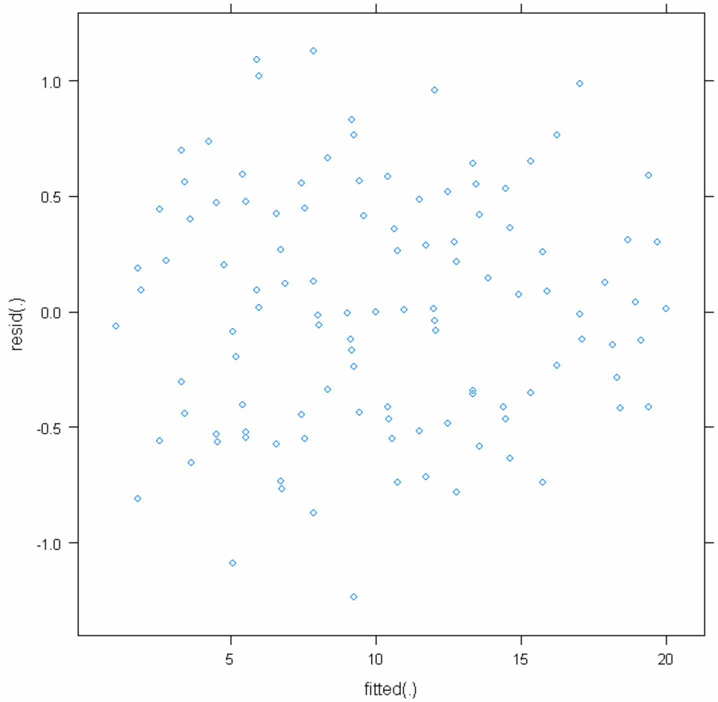
```
> qqmath(~resid(fm2))
```

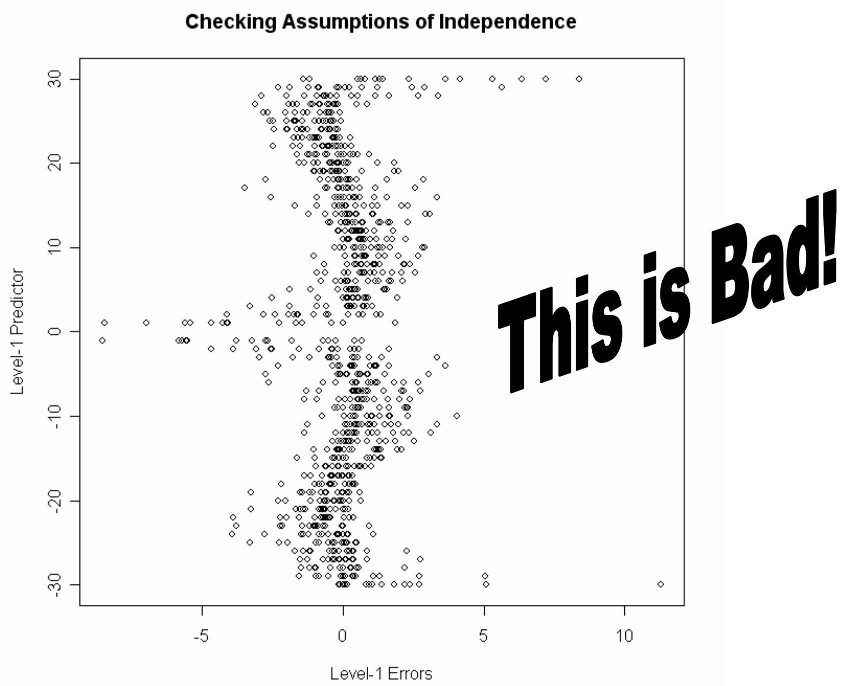


Distributional Assumptions

- Level-1 predictor is independent of Level-1 errors

```
> with(fm2, xyplot(resid(.) ~ fitted(.)))
```





An Example for Homework

HLM-Online.com

- http://www.hlm-online.com/datasets/education/
- Look at Dataset 2 (download dataset from online)
- Printout is in your packet

Other Software Packages for HLM Analysis

- Good Reviews at http://www.cmm.bristol.ac.uk/
  - MLwiN
  - SAS – PROC MIXED, PROC NLMIXED, GLIMMIX
  - S-PLUS – lme, nlme, glme
  - M*plus*
  - SPSS – Version 12 and later
  - STATA