

# Mixed Models in R

M. Hank H. Stevens

Department of Botany  
Miami University  
Oxford, Ohio, USA

Introduction

A Fixed Effect Approach

A Mixed Model Approach

# Acknowledgements

- Stephen Cox.
- Pinheiro and Bates (2000).
- Ben Bolker, Mollie Brooks, Connie Clark, Shane Geange, John Poulsen, Jada White: *Generalized mixed models: a practical guide for ecology and evolution*.

# What Is A Mixed Model?

- A statistical model with both *fixed* and *random* effects or factors.
- Fixed effect – under an investigator's control; repeatable; effect size of interest.
- Random effect – not under the investigator's control, therefore not repeatable; effects drawn from a distribution; effect size not of interest; variance of effect sizes of interest (or not).

## What Is A Fixed Effect?

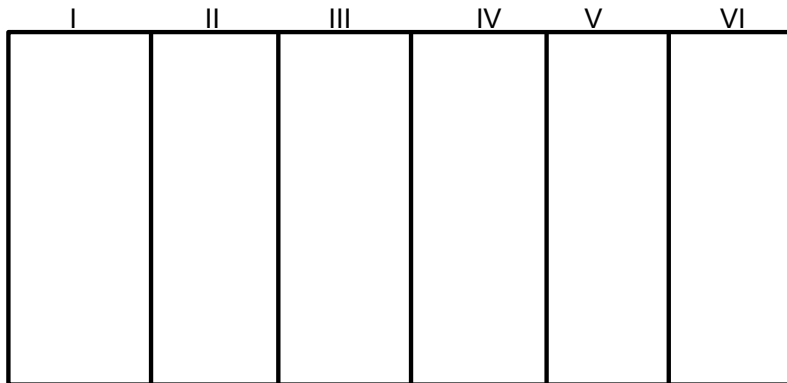
- An experimental treatment, or an environmental variable that varies deterministically (e.g., elevation, soil moisture).
- A variable for which we are interested in the *magnitude* of the effect.
- Under an investigator's control.
- Repeatable.

## What Is A Random Effect?

- E.g., experimental blocks in space or time, genotype, population.
- Levels are drawn from a larger distribution.
- Not under the investigator's control.
- Not repeatable.
- A variable for which we are *not* interested in the magnitude of a particular unit. Effect size not of interest;
- We may be interested in the *variance* of effect sizes of interest (or not).

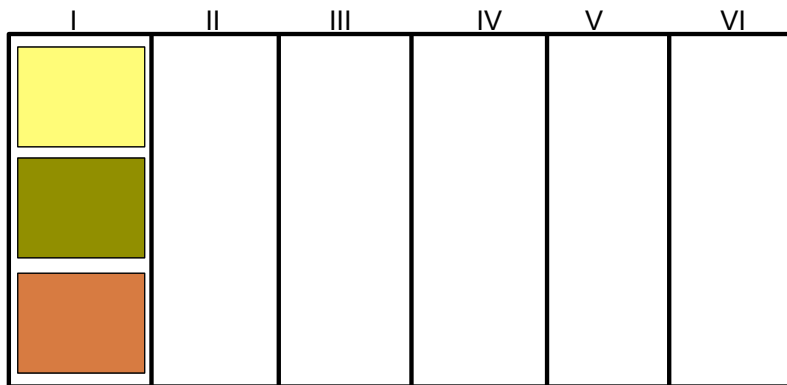
# Oats Experimental Design

Blocks



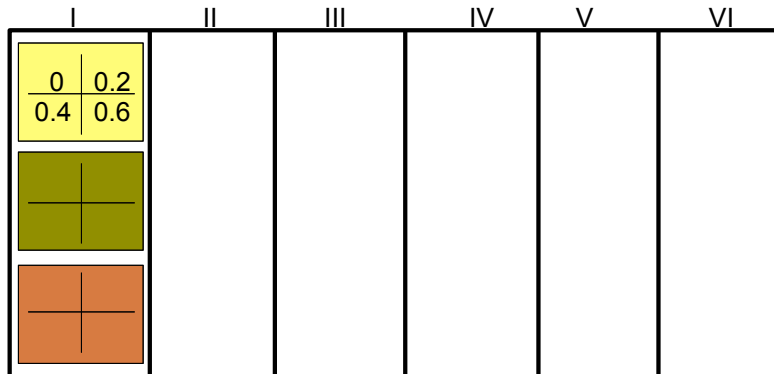
# Oats Experimental Design

Varieties (within Blocks)



# Oats Experimental Design

Nitrogen Addition  
(within Varieties within Blocks)



## The 'setwd' and other functions

```
> setwd("~/Documents/Presentations/2008/ESAWorkshop")  
> options(width = 60)  
> library(nlme)  
> library(lattice)
```

## Variables in the Oats data frame

```
> data(Oats)
```

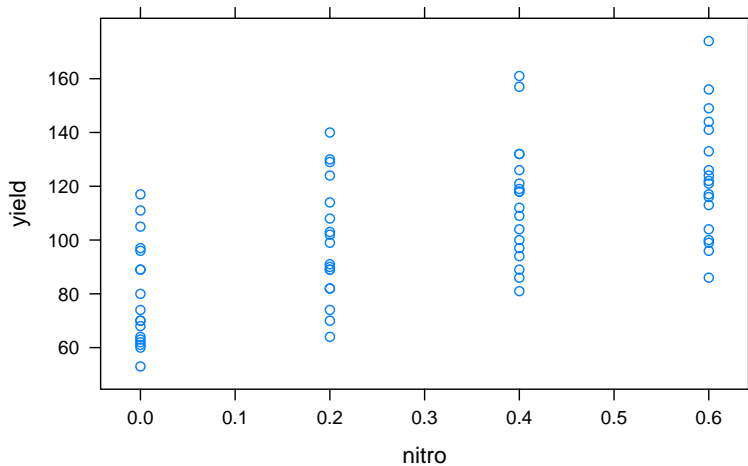
```
> summary(Oats)
```

Block	Variety	nitro	yield
VI :12	Golden Rain:24	Min. :0.00	Min. : 53
V :12	Marvellous :24	1st Qu.:0.15	1st Qu.: 86
III:12	Victory :24	Median :0.30	Median :102
IV :12		Mean :0.30	Mean :104
II :12		3rd Qu.:0.45	3rd Qu.:121
I :12		Max. :0.60	Max. :174

# Looking At The Data

```
> xyplot(yield ~ nitro, data = Oats)
```

```
xyplot(yield ~ nitro, data=Oats)
```

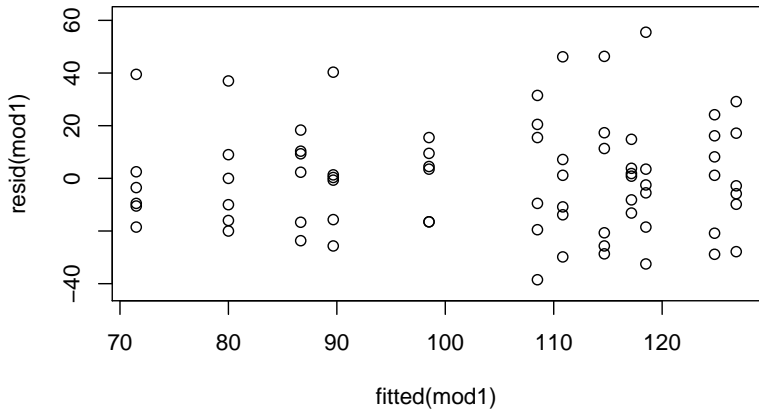


## A Simple Model of Yield

```
> nitroF <- factor(Oats$nitro)
> mod1 <- lm(yield ~ nitroF * Variety, data = Oats)
> plot(resid(mod1) ~ fitted(mod1), ylim = 1.1 *
+       range(resid(mod1)), main = "Fixed Effect Model")
```


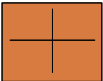
```
plot(resid(mod1) ~ fitted(mod1))
```

### Fixed Effect Model



# What Are The Data?

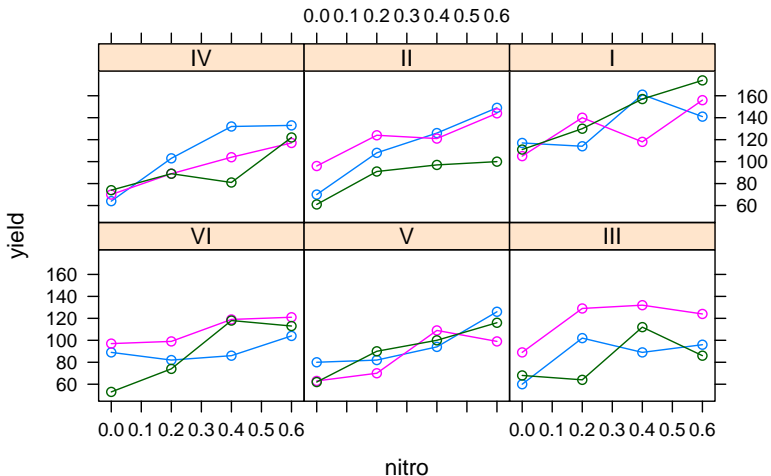
Nitrogen Addition  
(within Varieties within Blocks)

I	II	III	IV	V	VI				
<table border="1"><tr><td>0</td><td>0.2</td></tr><tr><td>0.4</td><td>0.6</td></tr></table>	0	0.2	0.4	0.6					
0	0.2								
0.4	0.6								
									
									

## An Inside Look At Our Data

```
> xyplot(yield ~ nitro | Block, data = Oats,  
+        groups = Variety,  
+        type = "b", layout = c(3, 2, 1))
```

```
xyplot(yield ~ nitro|Block, groups=Variety)
```



## A Mixed Model of Yield

$$\text{Yield} = \text{Intercept} + \text{N} + \text{Variety} + \text{N} \times \text{Variety} + \\ \text{Block} + \text{Whole Plot} + \text{Subplot (residual)}$$

## A Mixed Model of Yield

```
> mod3 <- lme(yield ~ nitroF * Variety, data = Oats,  
+           random = ~1 | Block/Variety)
```

## What is a "model object"?

```
> names(mod3)
```

```
[1] "modelStruct" "dims"          "contrasts"  
[4] "coefficients" "varFix"        "sigma"  
[7] "apVar"        "logLik"        "numIter"  
[10] "groups"       "call"          "terms"  
[13] "method"       "fitted"        "residuals"  
[16] "fixDF"        "na.action"     "data"
```

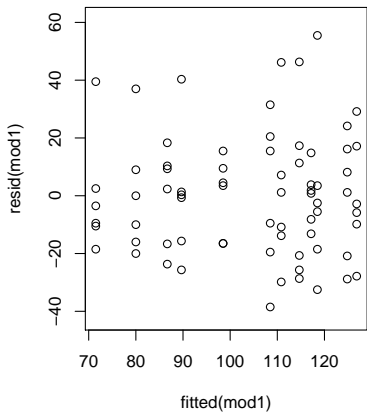
## Residuals vs. Fitted

```
> plot(resid(mod3) ~ fitted(mod3), ylim = 1.1 *  
+      range(resid(mod1)), main = "With Block/Variety")
```

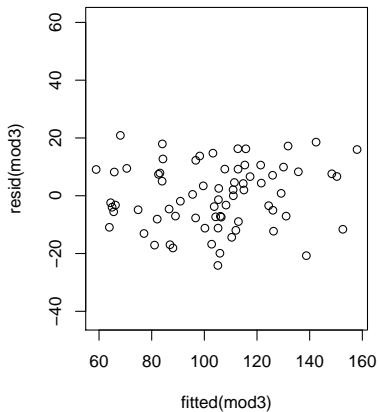
Note: `resid()` and `fitted` are *extractor functions*.

# Residuals vs. Fitted

**Fixed Effect Model**



**With Block/Variety**



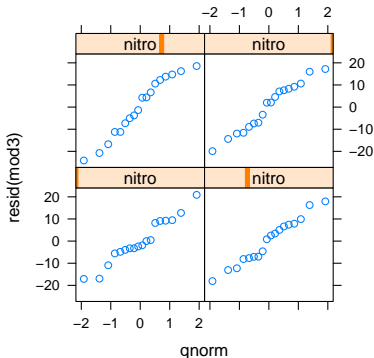
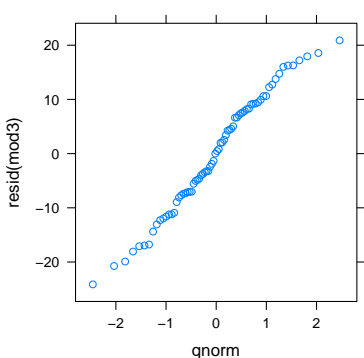
# Evaluating A Mixed Model

- Are the residuals normally distributed?
- Mean zero, with constant variance?

```
> qqmath(~resid(mod3), data = Oats)
```

```
> qqmath(~resid(mod3) | nitroF, data = Oats)
```

```
qqmath(~resid(mod3)|nitroF, data=Oats)
```



*Are the points along a straight line?*

# Evaluating A Mixed Model

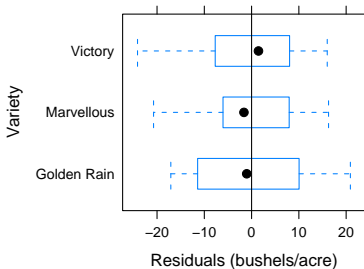
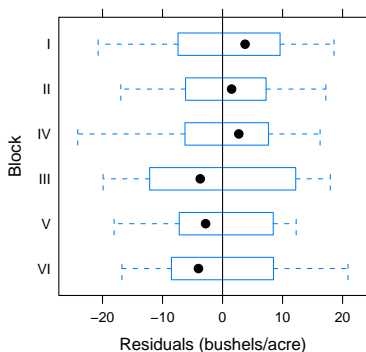
- Are the residuals normally distributed?
- Mean zero, with constant variance?

## Evaluating A Mixed Model

- Are the residuals normally distributed?
- Mean zero, with constant variance?

```
> plot(mod3, Block    resid(.), abline = 0)
> plot(mod3, Variety  resid(.), abline = 0)
```

```
plot(mod3, Variety ~ resid(.), abline=0)
```



*Is the variation similar among groups?*

# Modeling Your Data

Great, but is it the right model?

# Modeling Your Data

Great, but is it the right model?

## **Model Comparison**

# Modeling Your Data

Great, but is it the right model?

## **Model Comparison**

*... because not all random effects are real.*

# Model Comparison

## Principle of Parsimony Applied

- Define two models, e.g.,
  - mean vs. mean + effect.
  - most complex model vs. most complex - least likely term.
- Compare two models.
  - For random effects, compare models with likelihood ratio test, (LRT), and information criteria (AIC, BIC).
  - For fixed effects, compare models with ANOVA.

# Modeling Your Data

Is a simpler model more parsimonious?

```
> mod3 <- lme(yield ~ nitroF * Variety, data = Oats,  
+           random = ~1 | Block/Variety)  
> mod2 <- lme(yield ~ nitroF * Variety, data = Oats,  
+           random = ~1 | Block)  
> anova(mod2, mod3)
```

## Details

```
> anova(mod2, mod3)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
mod2	1	14	564.69	594.01	-268.35			
mod3	2	15	559.03	590.44	-264.51	1 vs 2	7.6615	0.0056

Note,

- AIC is lower for mod3. Suggests that mod3 is more parsimonious.
- LRT test (and the variation among subplots) is statistically significant.
- A parsimonious description of the data and appropriate test of nitrogen should account for variation among subplots.

# At Last Annoying Detail

Random Effects are Assumed to be  
**Normally Distributed**

## Check Normality of Random Effects

```
> str(ranef(mod3))
```

List of 2

```
$ Block :'data.frame':      6 obs. of  1 variable:
```

```
..$ (Intercept): num [1:6] -6.26 -10.58 -6.53 -4.71
```

```
..- attr(*, "effectNames")= chr "(Intercept)"
```

```
$ Variety:'data.frame':     18 obs. of  1 variable:
```

```
..$ (Intercept): num [1:18] -5.64  3.86 -1.32  1.12 -9.85
```

```
..- attr(*, "effectNames")= chr "(Intercept)"
```

```
- attr(*, "label")= chr "Random effects"
```

```
- attr(*, "level")= int 2
```

```
- attr(*, "standardized")= logi FALSE
```

```
- attr(*, "grpNames")= chr [1:2] "Block" "Variety %in% Blo
```

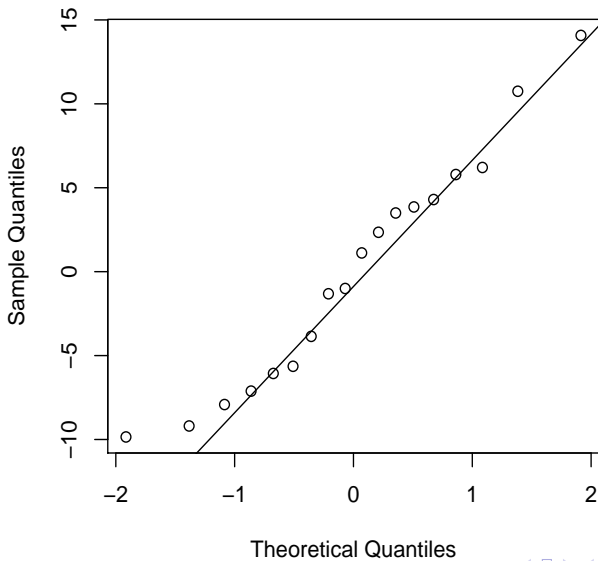
```
- attr(*, "class")= chr [1:2] "ranef.lme" "list"
```

## Check Normality of Random Effects

```
> subplotRE <- ranef(mod3)$Variety[, 1]
> qqnorm(subplotRE)
> qqline(subplotRE)
```

# Q-Q Plot of Subplot Effects

## Normal Q-Q Plot



# The Results We Were Looking For

## Confidence Intervals

```
> intervals(mod3)$fixed[, 1:6, ]  
> plot(intervals(mod3))
```

## F-tests

```
> anova(mod3)
```

## Confidence Intervals

```
> intervals(mod3)$reStruct
```

```
$Block
```

```
          lower  est.  upper  
sd((Intercept)) 6.7708 14.645 31.676
```

```
$Variety
```

```
          lower  est.  upper  
sd((Intercept)) 5.5005 10.299 19.282
```

```
attr("label")
```

```
[1] "Random Effects:"
```

```
> intervals(mod3)$sigma
```

```
  lower  est.  upper  
10.823 13.307 16.361
```

```
attr("label")
```

```
[1] "Within-group standard error:"
```

## Confidence Intervals

```
> intervals(mod3)$fixed
```

	lower	est.	upper
(Intercept)	61.6576	80.00000	98.342
nitroF0.2	3.0257	18.50000	33.974
nitroF0.4	19.1924	34.66667	50.141
nitroF0.6	29.3591	44.83333	60.308
VarietyMarvellous	-14.9798	6.66667	28.313
VarietyVictory	-30.1464	-8.50000	13.146
nitroF0.2:VarietyMarvellous	-18.5506	3.33333	25.217
nitroF0.4:VarietyMarvellous	-26.0506	-4.16667	17.717
nitroF0.6:VarietyMarvellous	-26.5506	-4.66667	17.217
nitroF0.2:VarietyVictory	-22.2173	-0.33333	21.551
nitroF0.4:VarietyVictory	-17.2173	4.66667	26.551
nitroF0.6:VarietyVictory	-19.7173	2.16667	24.051

```
attr(,"label")  
[1] "Fixed effects:"
```

# F-tests

```
> anova(mod3)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	45	245.143	<.0001
nitroF	3	45	37.686	<.0001
Variety	2	10	1.485	0.2724
nitroF:Variety	6	45	0.303	0.9322

# Comparing Levels of Nitrogen

*Think hard* before analyzing your data.

Approaches:

- *A priori* contrasts.
  - dummy coding (“treatment” contrasts).
  - polynomial contrasts.
  - Helmert and sum contrasts (both include the overall mean).
- `multcomp` package.

# Polynomial Contrasts

## Quantitative Relations:

- Assumes levels are equally spaced.
- Provides  $i - 1$  order polynomial.

```
> mod3.o <- lme(yield ~ as.ordered(nitroF) + Variety,  
+ data = Oats, random = ~1 | Block/Variety)
```

# Polynomial Contrasts

```
> round(summary(mod3.o)$tTable, 3)
```

	Value	Std.Error	DF	t-value	p-value
(Intercept)	104.500	7.797	51	13.402	0.000
as.ordered(nitroF).L	32.945	3.005	51	10.963	0.000
as.ordered(nitroF).Q	-5.167	3.005	51	-1.719	0.092
as.ordered(nitroF).C	-0.447	3.005	51	-0.149	0.882
VarietyMarvellous	5.292	7.079	10	0.748	0.472
VarietyVictory	-6.875	7.079	10	-0.971	0.354