NRES_798_20_201501

Mixed-effects models

Mixed effects

- Categorical explanatory variables
  - Fixed effect
  - Random effect

- Why random vs fixed makes a difference
  - Influence how mean square values and F-ratio's are calculated (validity of p-value)
  - Fixed

$$F - ratio_A = \frac{explained\ variance}{unexplained\ variance} = \frac{MS_A}{\boxed{MS_{within\ groups}}}$$

$$F - ratio_{AB} = \frac{explained\ variance}{unexplained\ variance} = \frac{MS_{AB}}{MS_{within\ groups}}$$

  - Random

$$F - ratio_A = \frac{explained\ variance}{unexplained\ variance} = \frac{MS_A}{\boxed{MS_{AB}}}$$

$$F - ratio_{AB} = \frac{explained\ variance}{unexplained\ variance} = \frac{MS_{AB}}{MS_{within\ groups}}$$

Mixed effects

- Categorical explanatory variables
  - Fixed effect
  - Random effect

- Why random vs fixed makes a difference
  - Influence how mean square values and F-ratio's are calculated (validity of p-value)

  - Fixed effects focus on mean (i.e. factors directly related to state), random effects focus on variance (objects differ in many ways so we focus on how variance differs between objects).
  - Normal fixed effect multiway ANOVA's assume independence of errors
    - Random effects from a common group contravene this assumption
      - Nested designs, hierarchical structure

Fixed vs. Random effects

- Effects
  - Fixed:
    - If they are identical for all groups in a population
    - If the factor levels are informative (height class of measurement)
  - Random
    - If they are allowed to differ from group to group
    - If the factor levels are uninformative (individual, location of measurement)

    - Gelman (2005)

Mixed models

- Varieties of mixed models:
  - Linear mixed models (LMM)
  - Nonlinear mixed models (NLM)
  - Generalized linear mixed models (GLMM)

Mixed models

- R packages

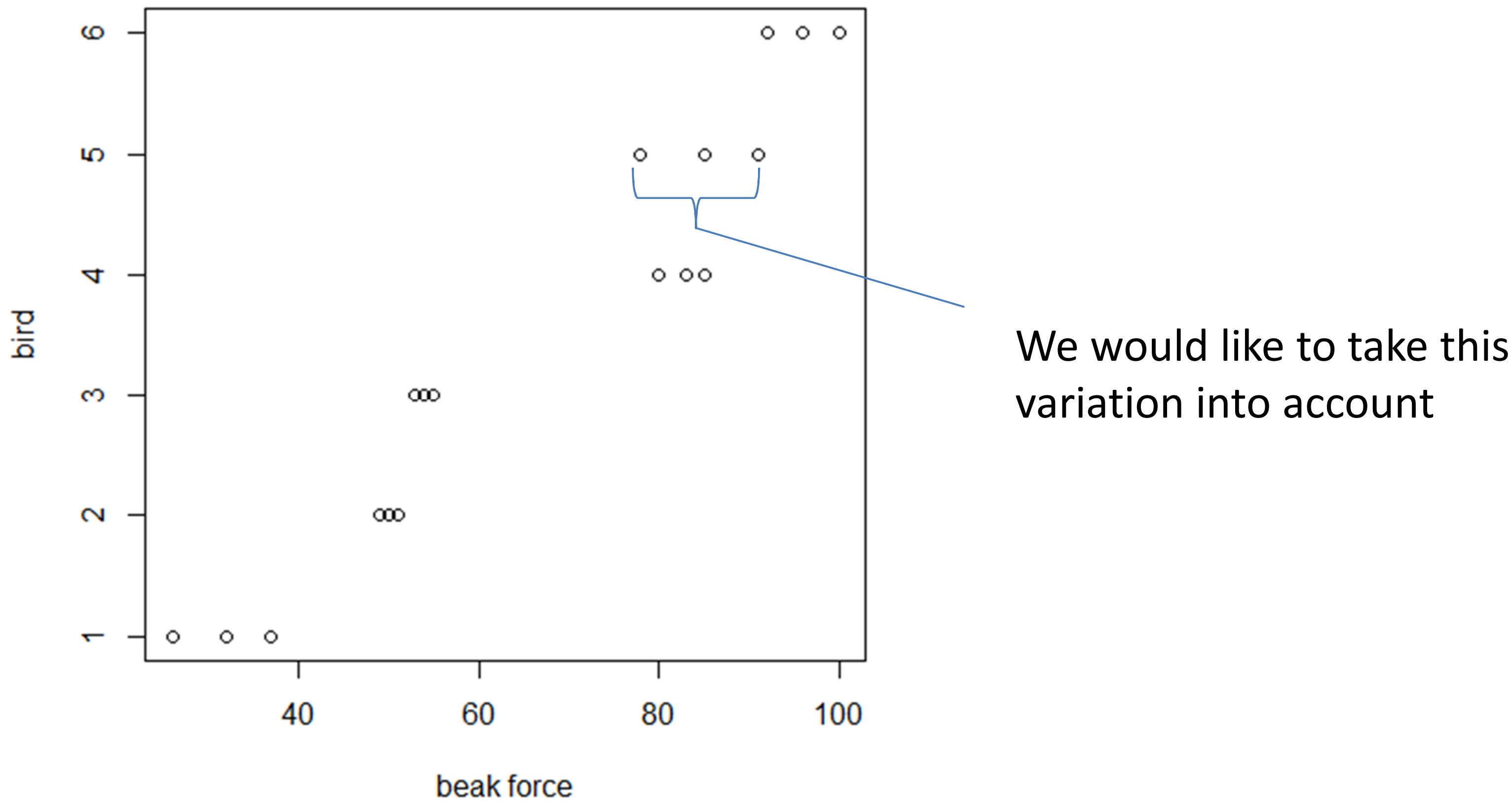
- “nlme”: Linear, non-linear mixed effects models
  – Slower, output easier to interpret
- “lme4”: Linear mixed effects models
  – Computationally “better”, more variable model design

Simple mixed effects

```
beak

## Grouped Data: travel ~ 1 | Rail
##    bird force
## 1     1    55
## 2     1    53
## 3     1    54
## 4     2    26
## 5     2    37
## 6     2    32
## 7     3    78
## 8     3    91
## 9     3    85
## 10    4    92
## 11    4   100
## 12    4    96
## 13    5    49
## 14    5    51
## 15    5    50
## 16    6    80
## 17    6    85
## 18    6    83
```

- What amount of force do birds create with their beaks?
- Mean beak force?



- Between-bird variability is grater than within-bird variability
- Within-bird variability is not constant
- Beak strength appears different for some birds

- Need to account for the classification factor (bird) in the analysis

Bird as a fixed effect

```r
# fixed effect model
b.lm1 <- lm(force ~ bird -1, data=beak)

b.lm1
```

```
##
## Call:
## lm(formula = force ~ bird - 1, data = beak)
##
## Coefficients:
## bird2  bird5  bird1  bird6  bird3  bird4
## 31.67  50.00  54.00  82.67  84.67  96.00
```

```
summary(b.lm1)

##
## Call:
## lm(formula = force ~ bird - 1, data = beak)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6667 -1.0000  0.1667  1.0000  6.3333
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## bird2   31.667      2.321   13.64 1.15e-08 ***
## bird5   50.000      2.321   21.54 5.86e-11 ***
## bird1   54.000      2.321   23.26 2.37e-11 ***
## bird6   82.667      2.321   35.61 1.54e-13 ***
## bird3   84.667      2.321   36.47 1.16e-13 ***
## bird4   96.000      2.321   41.35 2.59e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.021 on 12 degrees of freedom
## Multiple R-squared:  0.9978, Adjusted R-squared:  0.9967
## F-statistic: 916.6 on 6 and 12 DF,  p-value: 2.971e-15
```

We are interested in population value

12 DF?

Test for any difference between force

Model does not account for the fact that the 3 observations per bird are not independent

Random factor which is give by the grouping variable “bird”

```r
# random effect model
b.lme <- lmer(force ~ 1 + (1 | bird), REML=FALSE,data=beak)
summary(b.lme)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: force ~ 1 + (1 | bird)
##    Data: beak
##
##      AIC      BIC   logLik deviance df.resid
##    134.6    137.2    -64.3    128.6       15
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.61098 -0.28887  0.03454  0.21373  1.62222
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  bird     (Intercept) 511.86   22.624
##  Residual              16.17    4.021
## Number of obs: 18, groups:  bird, 6
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   66.500      9.285   7.162
```

Pseudo replication accounted for in grouping

Population estimate of bird beak force

```r
# Extract fixed-effects estimates
fixef(b.lme)

## (Intercept)
##        66.5

# Extract random-effects estimates
ranef(b.lme)

## $bird
##   (Intercept)
## 2   -34.47043
## 5   -16.32810
## 1   -12.36977
## 6    15.99824
## 3    17.97740
## 4    29.19266

# Extract full beak strenth estimates per bird
fixef(b.lme) + ranef(b.lme)$bird

##   (Intercept)
## 2    32.02957
## 5    50.17190
## 1    54.13023
## 6    82.49824
## 3    84.47740
## 4    95.69266
```

Fixed model

```
## Coefficients:
##         Estimate
## bird2     31.667
## bird5     50.000
## bird1     54.000
## bird6     82.667
## bird3     84.667
## bird4     96.000
```

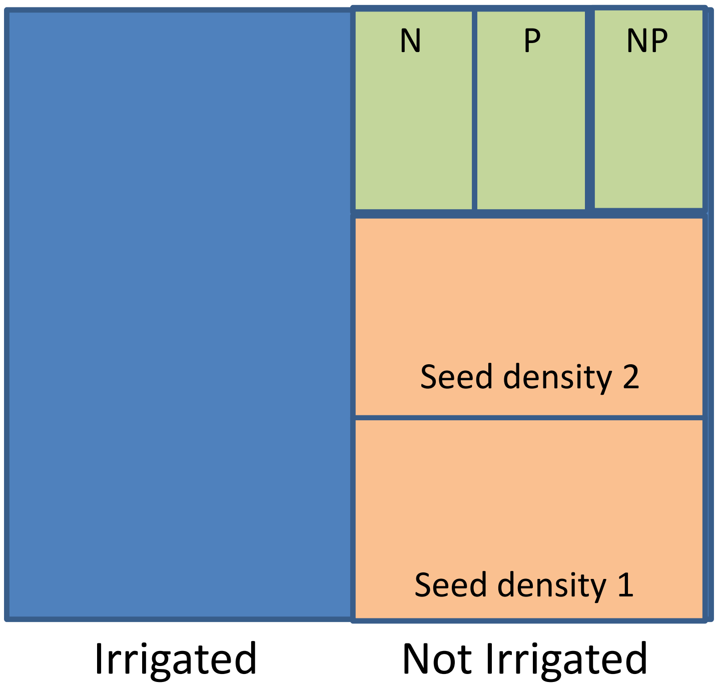
Hierarchical model design (nested, split plot)

- Country, Province, City, Neighborhood
- Region, Watershed, River, Section, Reach
- Region, Area, Site, Plot, Aspect
- Block, Irrigation, Density, Fertilization

Fixed vs. random? Independent vs. nested?

- Yield ~ irrigation + density + fertilization

- Blocks : 4 fields
- Irrigation : 2 levels
- Seeding density: 3 levels
- Fertilization : 3 levels



Hierarchical, split plot

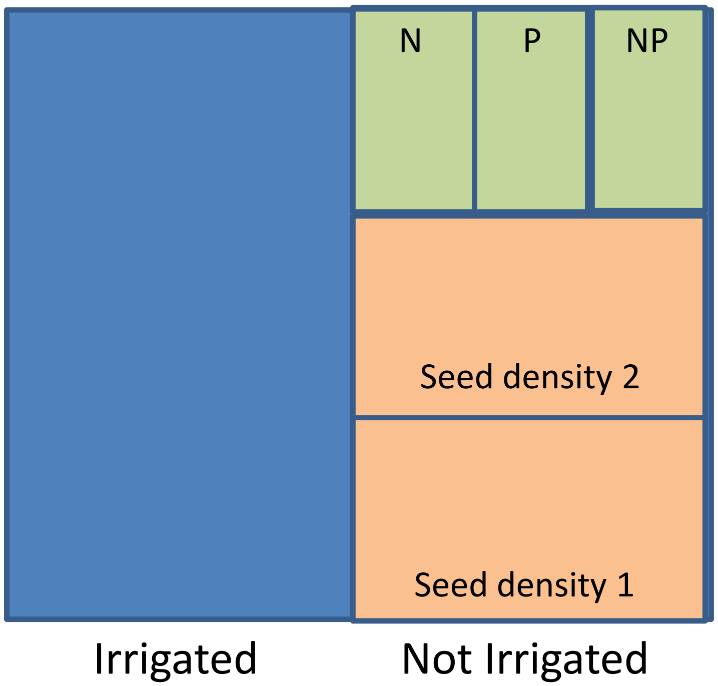
```
attach(yields)
names(yields)

## [1] "yield"      "block"      "irrigation" "density"    "fertilizer"

head(yields)

##   yield block irrigation density fertilizer
## 1    90     A    control     low          N
## 2    95     A    control     low          P
## 3   107     A    control     low         NP
## 4    92     A    control  medium          N
## 5    89     A    control  medium          P
## 6    92     A    control  medium         NP
```

- Yield ~ irrigation + density + fertilization



- Blocks : 4 fields
- Irrigation : 2 levels
- Seeding density: 3 levels
- Fertilization : 3 levels

Hierarchical, split plot

Nesting hierarchy

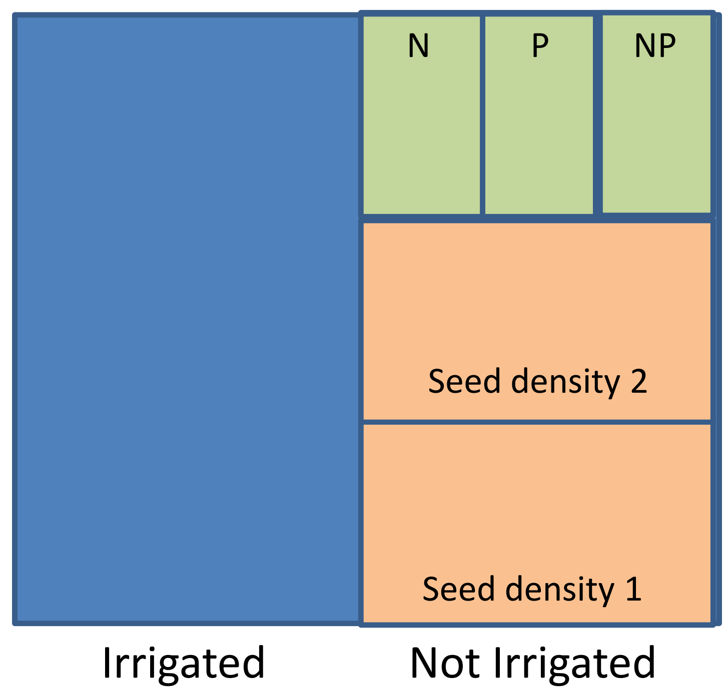
Fixed

Random

yd3 <- **lme**(yield ~ irrigation * density * fertilizer, random = ~ 1 | block/irrigation/density, data=yields)

- Yield ~ irrigation + density + fertilization

- Blocks : 4 fields
- Irrigation : 2 levels
- Seeding density: 3 levels
- Fertilization : 3 levels



Hierarchical, split plot

yd3 <- **lme**(yield ~ irrigation * density * fertilizer, random = ~ 1 | block/irrigation/density, data=yields)

```
anova(yd3)
```

```
##                                 numDF denDF   F-value p-value
## (Intercept)                         1    36 2674.6296  <.0001
## irrigation                          1     3   30.9207  0.0115
## density                             2    12    3.7842  0.0532
## fertilizer                          2    36   11.4493  0.0001
## irrigation:density                  2    12    5.9119  0.0163
## irrigation:fertilizer               2    36    5.5204  0.0081
## density:fertilizer                  4    36    0.8826  0.4841
## irrigation:density:fertilizer       4    36    0.6795  0.6107
```

yd3 <- **lme**(yield ~ irrigation * density * fertilizer, random = ~ 1 | block/irrigation/density, data=yields)

```
anova(yd3)
```

```
##                               numDF denDF   F-value p-value
## (Intercept)                       1    36 2674.6296  <.0001
## irrigation                        1     3   30.9207  0.0115
## density                           2    12    3.7842  0.0532
## fertilizer                        2    36   11.4493  0.0001
## irrigation:density                2    12    5.9119  0.0163
## irrigation:fertilizer             2    36    5.5204  0.0081
## density:fertilizer                4    36    0.8826  0.4841
## irrigation:density:fertilizer     4    36    0.6795  0.6107
```

yd2 <- **lm**(yield ~ irrigation * density * fertilizer)

```
anova(yd2)
```

```
## Analysis of Variance Table
##
## Response: yield
##                               Df Sum Sq Mean Sq F value    Pr(>F)
## irrigation                     1 8277.6  8277.6 59.5746 2.813e-10
***
## density                        2 1758.4   879.2  6.3276 0.0033972
**
## fertilizer                     2 1977.4   988.7  7.1160 0.0018070
**
## irrigation:density             2 2747.0  1373.5  9.8853 0.0002197
***
## irrigation:fertilizer          2  953.4   476.7  3.4310 0.0395615 *
## density:fertilizer             4  304.9    76.2  0.5486 0.7008151
## irrigation:density:fertilizer  4  234.7    58.7  0.4223 0.7918283
## Residuals                     54 7503.0   138.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hierarchical

yd3 <- **lme**(yield ~ irrigation * density * fertilizer, random = ~ 1 | block/irrigation/density, data=yields)

yd2 <- **lm**(yield ~ irrigation * density * fertilizer)

```
anova(yd3,yd2)
```

```
##      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## yd3      1 22 481.6212 525.3789 -218.8106
## yd2      2 19 482.6387 520.4294 -222.3193 1 vs 2 7.017463  0.0713
```

- Model selection often better than reliance on p-value
- Model selection should not be done non-critically
- Data and experiment structure will define which models are appropriate

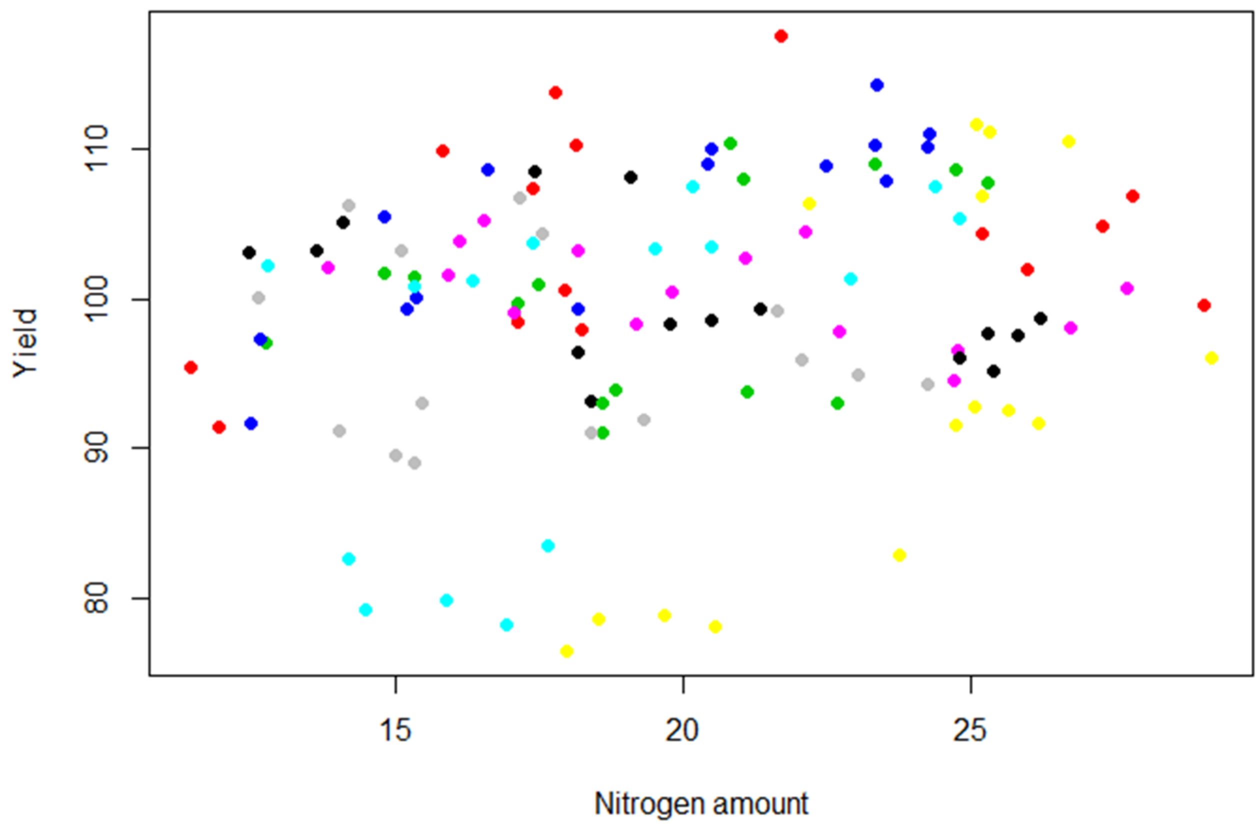
Mixed model regression

- Incorporate random factors into a regression framework

- 24 farms (4 points/farm)
- Measure soil N
- Record yield (plant size)

```r
attach(yields2)
farm<-factor(farm)

plot(N,size,pch=16,col=farm,xlab="Nitrogen amount",ylab="Yield")
```



Farms (random factor)

- Different intercepts?
- Different intercepts and slopes?

Mixed model regression

Unique intercept for each farm

```
m1 <- lme(size ~ N + farm, random = ~1|farm,method="ML")
m2 <- lme(size ~ N * farm, random = ~1|farm,method="ML")
```

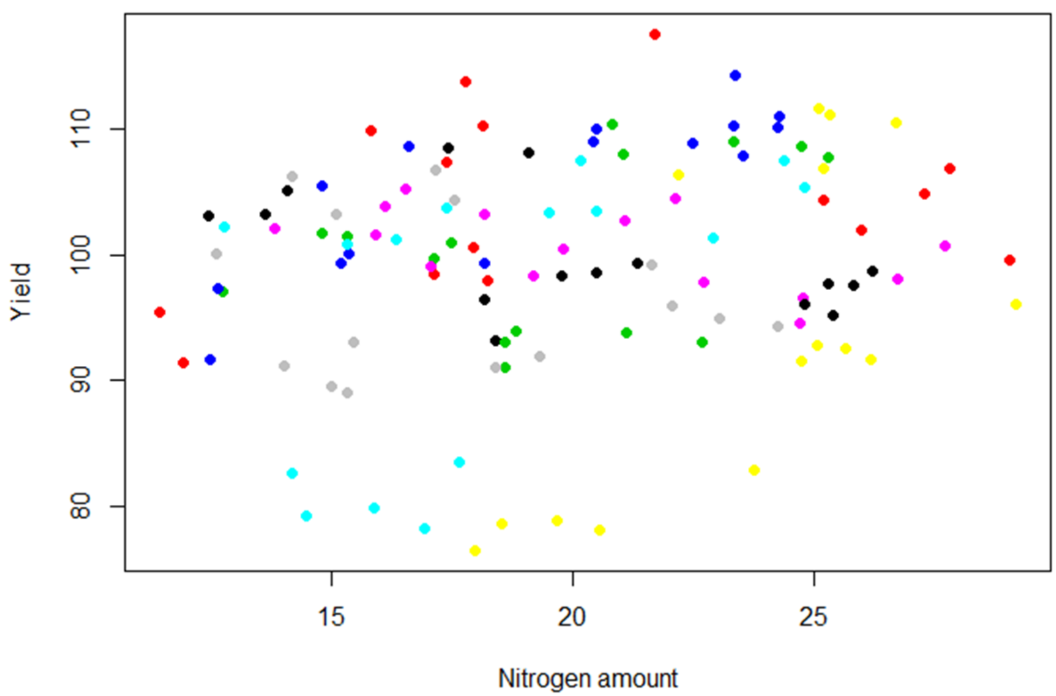
Unique intercept and slope for each farm

Mixed model regression

```
m1 <- lme(size ~ N + farm, random = ~1|farm,method="ML")
m2 <- lme(size ~ N * farm, random = ~1|farm,method="ML")

anova(m1,m2)
```

```
##     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## m1      1 27 524.2971 599.5594 -235.1486
## m2      2 50 542.9035 682.2781 -221.4518 1 vs 2 27.39359  0.2396
```



Statistical methods for ecologists

- Ideally know statistical framework (strengths, weaknesses, needs) prior to experimental design and data collection
- Always have clearly defined objective and hypothesis (the more detailed the better).
- Consider the distribution of your data (what type of random process underlies your data)
- Know your data (simple plots, distribution, correlations)

- Know why you are creating a statistical model (hypothesis testing, parameter estimation)
- Simple models are almost always better (easier to interpret)
- Distinguish between statistical effect size and ecological effect size
- Model comparison often provide more detail and insight compared to significance tests