Lab 8 – Nested ANOVA

October 8 & 9, 2017
FANR 6750

Richard Chandler and Bob Cooper

Outline

Scenario

We subsample each experimental unit

For example

- We count larvae at multiple subplots within a plot
- We weigh multiple chicks in a brood

We're interested in treatment effects at the experimental (whole) unit level, not the subunit level

The additive model

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$$

Because we want our inferences to apply to all experimental units, not just the ones in our sample, β_{ij} is random.

Specifically:

$$\beta_{ij} \sim \text{Normal}(0, \sigma_B^2)$$

And as always,

$$\varepsilon_{ijk} \sim \text{Normal}(0, \sigma^2)$$

Hypotheses

Treatment effects

$H_0 : \alpha_1 = \cdots = \alpha_a = 0$

H_a : at least one inequality

Random variation among experimental units

$H_0 : \sigma_B^2 = 0$

$H_a : \sigma_B^2 > 0$

Example data

Import data

```
gypsyData <- read.csv("gypsyData.csv")
str(gypsyData)

## 'data.frame': 36 obs. of  3 variables:
##  $ larvae   : num  16 16 15.8 14.2 13.9 14.2 13.5 13.4 14 13.1
##  $ Treatment: Factor w/ 3 levels "Bt","Control",..: 1 1 1 1 1
##  $ Plot     : int  1 1 1 1 2 2 2 2 3 3 ...
```

Convert `Plot` to a factor and then cross-tabulate

```
gypsyData$Plot <- factor(gypsyData$Plot)
table(gypsyData$Treatment, gypsyData$Plot)

##
##            1 2 3 4 5 6 7 8 9
##   Bt       4 4 4 0 0 0 0 0 0
##   Control  0 0 0 4 4 4 0 0 0
##   Dimilin  0 0 0 0 0 0 4 4 4
```


Incorrect analysis

```
aov.wrong <- aov(larvae ~ Treatment + Plot,
                 data=gypsyData)
```

```
summary(aov.wrong)

##             Df Sum Sq Mean Sq F value Pr(>F)
## Treatment    2 215.39  107.69  208.89 <2e-16 ***
## Plot         6  11.17    1.86    3.61 0.0093 **
## Residuals   27  13.92    0.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The denominator degrees-of-freedom are wrong

Correct analysis

```
aov.correct <- aov(larvae ~ Treatment + Error(Plot),
                   data=gypsyData)
```

```
summary(aov.correct)

##
## Error: Plot
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Treatment  2 215.39  107.69   57.87 0.00012 ***
## Residuals  6  11.17    1.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 27  13.92  0.5156
```

What happens if we analyze plot-level means?

The `aggregate` function is similar to `tapply` but it works on entire `data.frames`. Here we get averages for each whole plot.

```
plotData <- aggregate(formula=larvae ~ Treatment + Plot,
                      data=gypsyData, FUN=mean)
```

```
plotData

##   Treatment Plot larvae
## 1        Bt    1  15.50
## 2        Bt    2  13.75
## 3        Bt    3  14.00
## 4   Control    4  18.25
## 5   Control    5  18.75
## 6   Control    6  19.25
## 7   Dimilin    7  12.50
## 8   Dimilin    8  13.50
## 9   Dimilin    9  13.00
```


F and p values are the same as before

```
aov.plot <- aov(larvae ~ Treatment, data=plotData)
summary(aov.plot)

##             Df Sum Sq Mean Sq F value  Pr(>F)
## Treatment    2  53.85  26.924   57.87 0.00012 ***
## Residuals    6   2.79   0.465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov.correct)

##
## Error: Plot
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Treatment  2 215.39  107.69   57.87 0.00012 ***
## Residuals  6  11.17    1.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 27  13.92  0.5156
```


Issues

When using using `aov` with `Error` term:

- You can't use `TukeyHSD`
- You don't get a direct estimate of σ_B^2
- Doesn't handle unbalanced designs well
- But, you can use `model.tables` and `se.contrast`

An alternative is to use `lme` function in `nlme` package

- Possible to get direct estimates of σ_B^2 and other variance parameters
- Handles very complex models and unbalanced designs
- Possible to do multiple comparisons and contrasts using the the `glht` function in the `multcomp` package.
- But...
- Only works if there random effects
- ANOVA tables aren't as complete as `aov`

Using the `lme` function

```
library(nlme)
library(multcomp)
lme1 <- lme(larvae ~ Treatment, random=~1|Plot,
            data=gypsyData)
```

```
anova(lme1, Terms="Treatment")

## F-test for: Treatment
##   numDF denDF  F-value p-value
## 1     2     6 57.86567   1e-04
```

Variance parameter estimates

The first row shows the estimates of σ_B^2 and σ_B. The second row shows the estimates of σ^2 and σ

```
VarCorr(lme1)

## Plot = pdLogChol(1)
##             Variance  StdDev
## (Intercept) 0.3363889 0.5799904
## Residual    0.5155556 0.7180220
```

There is more random variation within whole units than among whole units (after accounting for treatment effects)

Extract the plot-level random effects

These are the β_{ij}'s

```
round(ranef(lme1), 2)

##   (Intercept)
## 1        0.78
## 2       -0.48
## 3       -0.30
## 4       -0.36
## 5        0.00
## 6        0.36
## 7       -0.36
## 8        0.36
## 9        0.00
```


Multiple comparisons

```
tuk <- glht(lme1, linfct=mcp(Treatment="Tukey"))
```

```
summary(tuk)

##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme.formula(fixed = larvae ~ Treatment, data = gypsyData, random = ~1 |
##     Plot)
##
## Linear Hypotheses:
##                       Estimate Std. Error z value Pr(>|z|)
## Control - Bt == 0        4.3333     0.5569   7.781   <0.001 ***
## Dimilin - Bt == 0       -1.4167     0.5569  -2.544   0.0295 *
## Dimilin - Control == 0  -5.7500     0.5569 -10.324   <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```


Assignment

To determine if salinity affects adult fish reproductive performance, a researcher places one pregnant female in a tank with one of three salinity levels: low, medium, and high, or a control tank. A week after birth, two offspring (fry) are measured.

Run a nested ANOVA using `aov` and `lme` on the `fishData.csv` dataset. Answer the following questions:

(1) What are the null and alternative hypotheses?

(2) Does salinity affect fry growth?

(3) If so, which salinity levels differ?

(4) Is there more random variation among or within experimental units?

Upload your self-contained **R** script to ELC at least one day before your next lab