Linear Models with Random Effects

Alexandra Chronopoulou

ILLINOIS

COLLEGE OF LIBERAL ARTS & SCIENCES

Department of Statistics
101 Illini Hall, MC-374
725 S. Wright St.
Champaign, IL 61820-5710

One-way ANOVA model

$$y_{ij} \;=\; \mu \;+\; \alpha_i \;+\; \varepsilon_{ij}$$

Previously, we assumed the α_is were parameters: fixed, unknown values. (We also gave a restriction, to make them identifiable.) These are **fixed effects**, corresponding to a **fixed factor**.

Now suppose that the α_is are <u>unobserved random variables</u>:

$$\alpha_i \;\sim\; N(0, \sigma_\alpha^2)$$

and assume they are <u>independent</u> of each other and of the ε_{ij}s.

These α_is are called **random effects**, and the corresponding factor variable is a **random factor**.

Fixed Versus Random Effects

– **Fixed effects** are appropriate when the levels of the factor are individually important or meaningful (e.g. treatments in a designed experiment, levels of education).

– **Random effects** are appropriate when the levels of the factor are meaningful only as representatives of a more general collection (e.g. as if sampled from a population, or representative of some hypothetical population).

Example:

Each level represents a factory worker.

– If we are interested in comparing these <u>particular</u> workers, we use a fixed factor.

– If these workers are just a *sample from the larger population of all workers,* and we want to know about that population (not just these particular workers), we use a random factor.

Example: Milk Data

- Bacterial counts are taken from randomly selected bottles of milk from different shipments.

- There are 5 different shipments, and 6 bottles are sampled from each shipment.

- These 5 shipments (from which the samples were taken) are of no particular importance. They are supposed to be representative of all shipments.

- Is there shipment-to-shipment variation in bacteria levels?

Data are in a file named "milk.txt"

```
milk = read.table("milk.txt", header=TRUE)
head(milk, 8)

##   bac shipment
## 1  24        a
## 2  15        a
## 3  21        a
## 4  27        a
## 5  33        a
## 6  23        a
## 7  14        b
## 8   7        b

dim(milk)

## [1] 30  2
```
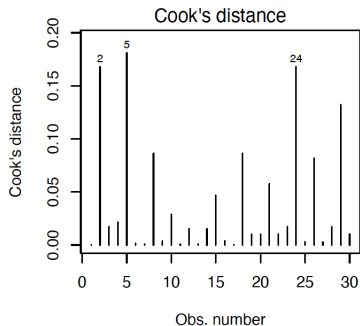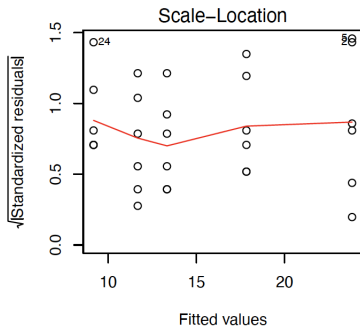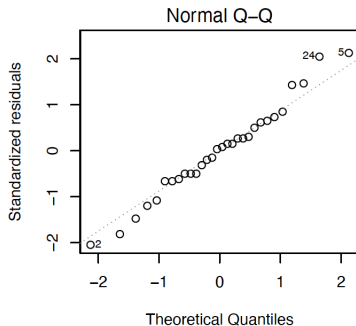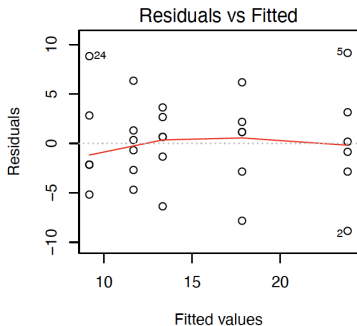
```
# We can use OLS for diagnostics
par(mfrow=c(2,2))
plot(milk.lm, which=c(3,4))
```

```
# We can use OLS for model checking diagnostics
milk.lm = lm(bac ~ shipment, data=milk)
par(mfrow=c(2,2))
plot(milk.lm, which=c(1,2))
```

ILLINOIS

– No apparent problems with the fit or the errors.

– These diagnostics do <u>not</u> check the normality assumption on the random effects. That is difficult in this example, because there are only 5 random effects.

– For the one-way ANOVA model with a random factor, the random effects satisfy

$$E(\alpha_i) \;=\; 0 \qquad\qquad Var(\alpha_i) \;=\; \sigma_\alpha^2$$

so random effects contribute only to the variance structure of the model, not to the mean structure.

– The parameter σ_α^2 is generally unknown, and we usually seek to estimate it (or σ_α) and test the null hypothesis $\sigma_\alpha^2 = 0$.

Parameters like σ_α^2 (and σ^2) are called **variance components**.

– Under this model, the responses can be correlated:

$$\text{Cov}(y_{ij}, y_{ij'}) \;=\; \sigma_\alpha^2 \qquad \text{for } j \neq j'$$

So different observations from the same "class" (same level i of the random factor) may have a nonzero correlation.

– The **intraclass correlation coefficient (ICC)** is the correlation between y_{ij} and $y_{ij'}$ (for any i and $j \neq j'$):

$$\rho \;=\; \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma^2}$$

- A **fixed effects model** has only fixed factors.

- A **random effects model** has only random factors.

- A **mixed (effects) model** has both fixed and random factors.

The general form (matrix-vector):

$$\mathbf{y} \;=\; \mathbf{X}\beta \;+\; \mathbf{Z}\gamma \;+\; \varepsilon$$

where $\mathbf{X}$ and $\mathbf{Z}$ are known design matrices, **beta** contains the fixed effect (mean-related) parameters, and

$$\gamma \sim N(\mathbf{0}, \sigma^2 \mathbf{D}) \qquad \text{independent of} \qquad \varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

are the random effects and the errors, with $\mathbf{D}$ containing the (unknown) random effect parameters.

The variance components (in **D**) are typically estimated via one of three different methods:

- **ANOVA estimation**, based on quantities in an ANOVA table; complicated for general models

- maximum likelihood

- **restricted maximum likelihood (REML)**, generally less biased than maximum likelihood

For balanced data, REML and ANOVA estimation tend to coincide.

R package `lme4` with mixed model function `lmer`

```
library(lme4)

## Loading required package:  Matrix
```

REML estimation

```
milk.reml = lmer(bac ~ (1|shipment), data=milk)
```

```
summary(milk.reml)

## Linear mixed model fit by REML ['lmerMod']
## Formula: bac ~ (1 | shipment)
##    Data: milk
##
## REML criterion at convergence: 184.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.66731 -0.60005  0.06269  0.50602  2.14554
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  shipment (Intercept) 29.74    5.454
##  Residual             22.29    4.721
## Number of obs: 30, groups:  shipment, 5
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   15.167      2.587   5.863
```

– The formula `bac ~ (1|shipment)` specifies the one-way random effects ANOVA model. As usual, there is an automatically-added intercept (representing μ), and the term `(1|shipment)` represents the random effect term α_i.

– Function `lmer` uses the REML method, by default. We see that the REML estimates of the variance components are

$$\widehat{\sigma}_\alpha^2 \;\approx\; 29.74 \qquad\qquad \widehat{\sigma}^2 \;\approx\; 22.29$$

(The `Std.Dev.` column simply gives the square roots of these: $\widehat{\sigma}_\alpha$ and $\widehat{\sigma}$.)

– The only fixed effect is the intercept, μ.

```r
# ML estimation

milk.ml = lmer(bac ~ (1|shipment), data=milk, REML=FALSE)
```

```
summary(milk.ml)

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: bac ~ (1 | shipment)
##    Data: milk
##
##      AIC      BIC   logLik deviance df.resid
##    194.1    198.3    -94.1    188.1       27
##
## Scaled residuals:
##     Min       1Q   Median       3Q      Max
## -1.61636 -0.63533  0.08278  0.48918  2.19649
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  shipment (Intercept) 23.05    4.801
##  Residual             22.29    4.721
## Number of obs: 30, groups:  shipment, 5
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   15.167      2.314   6.555
```

– So the MLE for σ_α^2 is

$$\widehat{\sigma}_\alpha^2 \;\approx\; 23.05$$

smaller than the REML estimate 29.74. MLEs for variance components are often biased low.

– The estimate listed for σ^2 is apparently <u>not</u> the MLE, but is still the REML estimate.

– In this case, the estimate of μ has remained unchanged, but its standard error has changed.

– For the <u>fixed effects</u> (in β), likelihood ratio tests are available. (For this to work, the variance components should be estimated with MLE, not REML.)

(These LRTs are sometimes unreliable, so a <u>parametric bootstrap</u> approach can be used — later.)

– The methods of generalized least squares (F-tests and t-tests) could alternatively be used (though this would ignore the additional uncertainty of replacing $\mathbf{D}$ with $\hat{\mathbf{D}}$).

– There are also confidence intervals for fixed effect parameters based on the Wald approach or (perhaps more reliably) on profile likelihood.

Testing the Random Effect Variance

- For the random effects, the null hypothesis is usually that a variance component equals zero.

- For technical reasons, the usual chi-square approximation in the LRT often fails to be adequate (most often leading to a test that is too conservative).

- An improvement is to use the parametric bootstrap to perform the LRT (see example later).

- Methods based on ANOVA are also available, and can be useful in single-factor or balanced cases.

- Profile likelihood confidence intervals for variance components can be computed (but may have problems, as the LRT does).

In the milk example, let's try an ordinary (fixed-effects) ANOVA first

```
anova(milk.lm)

## Analysis of Variance Table
##
## Response: bac
##            Df Sum Sq Mean Sq F value    Pr(>F)
## shipment    4 803.00 200.750  9.0076 0.0001197 ***
## Residuals  25 557.17  22.287
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

So shipment would be declared highly significant as a fixed factor.

For a <u>one-way</u> ANOVA model like this one, this F-test is also valid for testing the null hypothesis that $\sigma_\alpha^2 = 0$ versus the alternative $\sigma_\alpha^2 > 0$. Therefore, `shipment` would be declared highly significant as a random factor.

An ordinary LRT of $H_0 : \sigma_\alpha^2 = 0$ can be conducted as follows:

```r
nullmod = lm(bac ~ 1, data=milk)
llrt = as.numeric(2 * (logLik(milk.ml) - logLik(nullmod)))
llrt

## [1] 11.43066

pchisq(llrt,1,lower=FALSE)  # ordinary LRT p-value

## [1] 0.000722422
```

Paper production experiment with $n = 20$: do different operators affect the brightness of paper produced?

Variables:

- `bright` - Brightness of paper pulp as measured by reflectance meter
- `operator` - Shift operator a-d

```
##    bright operator
## 1   59.8        a
## 2   60.0        a
## 3   60.8        a
## 4   60.8        a
## 5   59.8        a
## 6   59.8        b
## 7   60.2        b
## 8   60.4        b
```

ILLINOIS

```
# Operator means

aov(bright ~ -1 + operator, data=pulp)$coef

## operatora operatorb operatorc operatord
##    60.24     60.06     60.62     60.68
```

```
# Comparison of operators
par(mfrow=c(2,1))
boxplot(bright ~ operator, data=pulp)
```

ILLINOIS
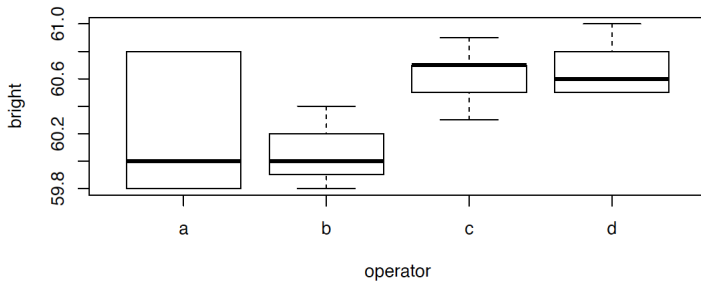
Mixed model:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: bright ~ (1 | operator)
##    Data: pulp
##
## REML criterion at convergence: 18.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4666 -0.7595 -0.1244  0.6281  1.6012
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  operator (Intercept) 0.06808  0.2609
##  Residual             0.10625  0.3260
## Number of obs: 20, groups:  operator, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  60.4000     0.1494   404.2
```

I ILLINOIS

We see that the REML variance component estimates are

$$\tilde{\sigma}_\alpha^2 = 0.0681 \qquad \tilde{\sigma}^2 = 0.10625$$

so the estimated intraclass correlation is

$$\tilde{\rho} = \frac{0.0681}{0.0681 + 0.10625} = 0.39$$

Is the correlation significant? We wish to test $H_0 : \sigma_\alpha^2 = 0$

ANOVA test:

```
# Treats operator as a fixed effect
summary(aov(bright~operator, data=pulp))

##             Df Sum Sq Mean Sq F value Pr(>F)
## operator     3   1.34  0.4467   4.204 0.0226 *
## Residuals   16   1.70  0.1062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Suggests operator variation is present.

Next: compare random effect likelihood ratio chi-square test and parametric bootstrap

Compare the models:

$$\text{Full Model: } y_{ij} = \mu + \alpha_i + e_{ij}, \quad \alpha_i \text{ indep. } N(0, \sigma_\alpha^2), \quad \sigma_\alpha^2 > 0$$

$$\text{Null Model: } y_{ij} = \mu + e_{ij}$$

```r
# Calculate LLRT for m0 versus m1 and p-value
fullmod = lmer(bright~ 1+ (1|operator), data=pulp, REML=FALSE)
nullmod = lm(bright ~ 1, pulp)
lrtstat = as.numeric(2*(logLik(fullmod)-logLik(nullmod)))
lrtstat

## [1] 2.568371

1 - pchisq(lrtstat, 1) # p-value

## [1] 0.1090199
```

Parametric Bootstrap

In our example, n is small and the number of random effects is very small, which means the power may be low.

It also means the chi-square approximation might not be very good.

The **parametric bootstrap** may be more accurate in small samples. Here are the steps:

1. Compute the LR statistics for the null and alternative models
2. Generate data under the null hypothesis model
3. Fit the null and alternative model for the generated data
4. Compute the LR statistic
5. Repeat steps 2 to 4 many times
6. Find the Bootstrap probability of exceeding the observed LR value

Apply to the pulp data:

```
# Generate bootstrap LLR values under null model
B = 1000
lrbs = rep(0, B)
set.seed(12357)
for(i in 1:B){
  y = unlist(simulate(nullmod)) #null bootstrap sample
  m0 = lm(y~1)
  m1 = suppressMessages(lmer(y~ 1+ (1|operator),pulp,REML=FALSE))
  lrbs[i]=as.numeric(2*(logLik(m1)-logLik(m0)))
}
summary(lrbs)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.23465 0.00971 8.03386
```

```r
par(mfrow=c(2,2))
d = density(lrbs[lrbs>0])
plot(d$x, d$y, type="l", xlab="T", ylab="", main="density(T | T > 0)" );
abline(v=lrtstat, lty=2); text(lrtstat+0.1, 2, "<= Observed Value", pos=4)
hist(lrbs); abline(v=lrtstat, lty=2)
```
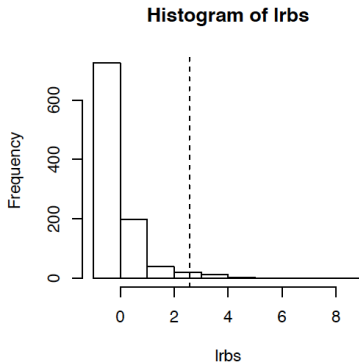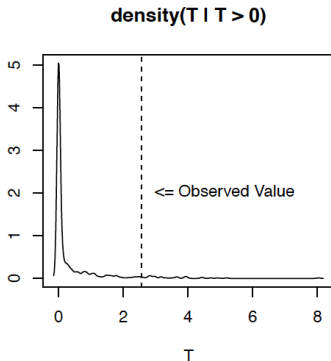
p-value and standard error for the p-value

```
pvalue = mean(lrbs>lrtstat) # Estimated p-value
pvalue

## [1] 0.03

#p-value standard error
sqrt(pvalue*(1-pvalue)/B)

## [1] 0.005394442
```

Alternatively, we can use the bootstrap to get confidence intervals for the random effects based on the full model:

```
suppressMessages(confint(fullmod, method="boot"))

##                    2.5 %      97.5 %
## .sig01        0.0000000   0.3667739
## .sigma        0.2131938   0.4271960
## (Intercept) 60.1380659  60.6525116
```

Since the confidence interval for .sig01 includes zero, a bootstrap test of this coefficient fails to reject the null hypothesis. (We suppressed messages to avoid repeated warnings about estimates at the boundary value zero, which we expect.)