

# Random Effects

Gary W. Oehlert

School of Statistics  
University of Minnesota

October 4, 2014

What a strange world we live in, said Alice to the Queen of Hearts.  
Lewis Carroll

# Fixed Effects

Everything we have done so far has dealt with fixed effects:

- The treatment effects  $\alpha_i$  may be unknown, but they are constants.
- If we ran the experiment over again, we would expect the same treatment effects.
- We can always increase the power of all of our tests to 1 simply by increasing  $n$ .
- We do inference on the  $\alpha_i$ s, and we do it via t-tests and F-tests.

It is time for a new world, one where each of those statements is false.

# Random Effects

Random effects models look very similar to fixed effects models. E.g., we could have

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

or

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

The difference is in the assumptions we make for the treatment effects.

## Fixed effects:

- $\alpha_i$ s and  $\beta_j$ s are constants that add to zero (or some other constraint).

## Random effects:

- $\alpha_i$ s are random variables that follow a  $N(0, \sigma_\alpha^2)$  distribution. ( $\beta_j$ s follow  $N(0, \sigma_\beta^2)$  and so on.)
- $\alpha_i$ s are independent of  $\epsilon_{ijk}$ s and  $\beta_j$ s and so on.
- We usually assume that all random effects within a given term are independent of each other (but there are cases where some factors are fixed and some are random when there can be some dependence in interaction effects).

Fixed effects:

- $\text{Var}(y_{ij}) = \sigma^2$ .
- All  $y_{ij}$ s are independent of each other.

Random effects:

- $\text{Var}(y_{ij}) = \sigma_\alpha^2 + \sigma^2$ .
- $\text{cor}(y_{ij}, y_{kl}) = \begin{cases} 0 & i \neq k \\ \sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma^2) & i = k; j \neq l \\ 1 & i = k; j = l \end{cases}$

Note:  $\sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma^2)$  is called the “intra-class correlation”; geneticists might call it the heritability.

Fixed effects:

- Interest and inference is about the  $\alpha_j$ s.

Random effects:

- Interest and inference (estimates, confidence intervals, tests) is mostly about  $\sigma_\alpha^2$ .
- In some cases there is interest in the  $\alpha_j$ s.

The variances  $\sigma_\alpha^2$  etc are called components of variance or variance components.

How and why are things so very different?

Fixed effects:

- The treatments are the treatments and they are unchanging.
- If we rerun the experiment, we are still studying the same treatments.

Random effects:

- The treatments are a random sample from a population of potential treatments.
- If we rerun the experiment, we are looking at an entirely new sample of treatments.
- Inference is on the population of potential treatments.



We use random effects models in cases like

- ① We have actually sampled treatments from a population.
- ② Treatments are supposed to represent a population of treatments about which we wish to learn.
- ③ Sampling patterns in the collection of units have induced similarities (correlations) among our units that can be modeled via random effects.

Point 3 is important: random effects are a model for the variability (variance and covariances) in the data.

Adding a random  $\alpha_i$  might look like we are affecting the mean, but we are not; we are affecting the random variation and the covariance.

Other sources of variation and covariation in the data can be modeled with random effects.

Here is one more way that random effects are different.

Fixed effects:

- Hierarchical models should be our default models.

Random effects:

- Non-hierarchical models for random terms make sense in most cases (fixed terms should still generally be hierarchical).

In fixed effects, the factorial breakdown is kind of arbitrary, and non-hierarchical models make sense only in specific circumstances.

In random effects, the variance/covariance structures for non-hierarchical models still make sense.

# Fork in the road



We now must choose between “old school” and “new school” .

The text describes old school. In the old school, you compute an ANOVA table for an RE model as if it were FE. Then you fix what needs to be fixed to make sense in the RE setting.

For simple cases, old school is dead easy and works well. For somewhat more complicated cases, old school can be rather difficult. For really complicated cases, old school is virtually impossible.

Old school also has the embarrassment of negative estimates of variances. (This can be fixed, but it lacks elegance.)

New school is the REML approach as described in the handout.  
New school models the random effects directly from the start.

New school is never dead easy; even simple cases are a little difficult. However, with new school even really complicated cases are not much more difficult than the simple cases.

New school never gives negative estimates of variance.

We will mostly do new school, with a few allusions to old school.  
We will also find that new and old agree in some simple cases.

# Normal Likelihood

The probability density for a single normal is

$$f(y_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(y_i - \mu)^2}{2\sigma^2} \right]$$

(the canonical “bell shaped curve.”) The probability density for the  $n$  independent results is the product of the individual densities:

$$f(y_1, \dots, y_n; \mu, \sigma^2) = \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \right]^n \exp \left[ -\sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2} \right]$$

We can adapt this for more complicated mean structures (e.g., regression or factorial) and more complicated variance structures (e.g., random effects).

Turn this around and think of it as a function of the parameters with the data as given, then we get the likelihood:

$$L(\mu, \sigma^2; y_1, \dots, y_n) = \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \right]^n \exp \left[ - \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2} \right]$$

We usually work with the log likelihood

$$\begin{aligned} \ell(\mu, \sigma^2; y_1, \dots, y_n) &= \ln[L(\mu, \sigma^2; y_1, \dots, y_n)] \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2} \end{aligned}$$

If you ever wondered why we minimize sums of squares, look no further.



# Maximum Likelihood

Maximum likelihood estimates of parameters are those parameters that make the observed data most likely, that is, have the highest likelihood.

MLEs have lots of desirable properties (mostly for large samples).

From calculus, to maximize take the derivatives of  $\ell(\mu, \sigma^2)$  with respect to  $\mu$  and  $\sigma^2$ , set the derivatives to zero, and solve for the parameter values.

In the example,

$$\hat{\mu} = \bar{y} = \sum_{i=1}^n y_i/n$$
$$\hat{\sigma}^2 = \sum_{i=1}^n (y_i - \hat{\mu})^2/n$$

Note that the divisor in the estimated variance is  $n$ , not the more common  $n - 1$ . This  $n$  shows up for regression, factorials, and so on.

# Mixed Effects

observation = fixed effects part + random effects part + error

This is mixed effects, because it includes both fixed and random.

The fixed effects part can be anything we've talked about to date, e.g., qualitative factors, polynomial terms, etc.

The random effects part can be very general, but we will only talk about terms that are additive random terms for factor levels like  $\alpha_i$ s or  $\alpha\beta_{ij}$ s.

The parameters of the mixed model are the usual fixed effects we are used to, the variances of the random effects, and the variance of error.

One of the problems with MLE for  $\sigma^2$  is a tendency to systematically underestimate (i.e., it uses  $n$  in the denominator when it should use something smaller).

This brings us to . . .

# Restricted Maximum Likelihood

This is REML, also called Residual Maximum Likelihood. REML is a modification to ML that cures the problem of bias in variance estimates.

Short version:

- 1 Get residuals of data from the fixed effects model.
- 2 Estimate variances using ML based on residuals (which have a complicated covariance structure).
- 3 Using variance estimates as “true,” get best estimates of fixed effects parameters.

(Fixed effects estimates in steps 1 and 3 could differ.)

We can use ML to compare nested models that differ in fixed or random terms.

We can only use REML to compare nested models that differ in random terms. (Remember, the first thing REML does is get rid of the fixed stuff.)

To compare models using likelihood, the standard approach is

- 1 Take minus twice the difference of the log likelihoods of two nested models;
- 2 Treat that as chi square with df equal to difference in number of parameters;
- 3 Reject included model in favor of larger model if p-value is small enough.

Note: minus twice the log likelihood is called the Deviance. We have an ML deviance and a REML deviance; they are not comparable.

We can also compare models using AIC or BIC:

$$AIC = -2\ell + 2k = \text{Deviance} + 2k$$

$$BIC = -2\ell + \log(n)k = \text{Deviance} + \log(n)k$$

where  $k$  is the number of parameters in the model.

Unfortunately, the chi-square approximation in step 2 is not terribly good for testing random effects.

To get better p-values, we can simulate the distribution of the chi-square test. There is a function in R that will do that.

Calves example in R.



We would like to have confidence intervals for our parameters, including better confidence intervals for mean parameters than simply estimate plus or minus two SE.

What to do, what to do?

# Bayesian Statistics in 60 seconds

Bayesian statistics assume that all parameters are random variables. We express our knowledge about them via a distribution.

Before seeing data, we have a prior distribution for the parameters.

As before, we have a likelihood for data given the parameters.

After seeing the data, we can update our distribution for the parameters based on the data (now it is a posterior distribution). This update uses Bayes' Rule, and thus we have Bayesian statistics.

Computing the posterior can be **very** difficult, but researchers have developed ways (called Markov Chain Monte Carlo or MCMC) to sample (approximately) from the posterior. Theory says that the approximation is good enough “in large enough samples.”

We will sample from the posterior and then make inferences based on the samples.

If somehow making a very difficult problem very easy seems too good to be true, well, you're right.

- 1 MCMC sampling can be very slow.
- 2 MCMC chains can have “transient” behavior before settling down to the posterior.
- 3 We can sometimes say that a chain is still transient, but we can rarely guarantee that a chain is not still transient.
- 4 How you set up your MCMC can matter, a lot.
- 5 Your only real fix is to lengthen the chain (see item #1), but how much (see item #3)?

Bayesian statistics have many optimality properties. In fact, there are many classes of problems for which you can show that you cannot do better than the Bayesian approach.

However, the prior distribution injects an element of subjectivity, and many people find that anathema.

And, as we've said, Bayesian results can be very difficult to compute.

Back to the calves example.

In principle, there is nothing really different to do when we use REML on more complicated models. This is in direct contrast to the situation in the “old school” approach, where random effects models with more terms are more complicated to analyze, even in simple, balanced cases.

Resistors example.

# Nesting and Crossing

So far our multi-factor models have all been crossed models (although we have not discussed it as such).

When two factors are crossed, each level of one factor appears with each level of the other factor. For example, in the resistors data, each operator appeared with each part, so part and operator were crossed.

In contrast, if factor B nests in factor A, then each level of factor A has a completely different set of factor B effects.

In notation, for an effect B nested in A we write it as B(A).

As an effect, we write B nested in A as  $\beta_{j(i)}$ .

If B nests in A, then A and B cannot interact.

You can have C nested in B nested in A. In a different experiment, you might find C nested in the AB combinations. Almost anything is possible.

Nested factors are almost always random effects, but it is possible to find examples where they could be fixed effects.



Suppose we have six randomly chosen male rats. Each male rat is bred to five different female rats (also assumed to be a random selection of female rats). We then measure the four-week growth of each pup that is born.

In this case, female is nested in male, because we have a different set of females for each male. Pups are nested within each female.

Is pup an effect or is it error? Does it matter? What if we had two measurements for each pup?

Even in our basic fixed effects models, we could have written error as nested in the treatments, e.g.,  $y_{ink} = \mu_{ij} + \epsilon_{k(ij)}$ , because errors are nested; we just didn't bother to point it out until we had other nested things.

We are interested in the effects of apple variety and treatment on growing. We use three varieties of apples (Macintosh, Honey Crisp, Harrelson). From each variety, we choose six apples at random. Each apple is then cut into four slices. The four slices from an apple are then randomly assigned to four treatments: control, orange juice, lemon juice, salt water.

In this design, apples are nested in variety, but both variety and individual apple cross with treatment (each treatment appears with each variety and each apple, but there are different apples for each variety).

Why do we use nesting?

- Genuinely subsampled treatments.
- Constrained by logistics, e.g., need to do things independently at different sites.
- Correlation structure among units may be modeled by nesting.

Lots of examples in R.

I think that figuring out what is fixed and what is random, what is crossed and what is nested is lots of fun, but also very challenging. You will get lots of practice.

More about this later, but the general procedure in describing a model for an experiment is:

- 1 Determine the sources of variation.
- 2 Determine which cross and which nest.
- 3 Determine which are random and which are fixed.
- 4 Fill out the possible interactions.

# Hasse diagram

We will also use a graphical representation called a Hasse diagram to help us keep track of terms in a model.

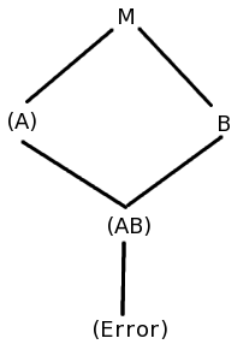
In old school non-REML analysis, the Hasse diagram was the key to doing all kinds of random and mixed effects models. In REML analysis, the diagram is not so central, but even a simplified form is still helpful for arranging your thoughts about the appropriate model for a situation.

A Hasse diagram is a graph of nodes that represent terms in a model.

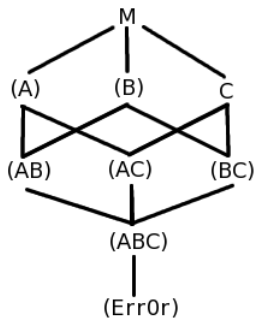
The nodes are connected by line segments that illustrate relationships between nodes (nesting, crossing, etc).

We will use an incomplete Hasse diagram, where each node just gives a label for the corresponding term and possibly parentheses as an indication that the term is a random term.

A complete Hasse diagram also includes subscripts that indicate numbers of levels and degrees of freedom.

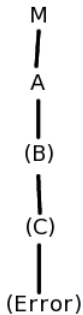


Two factors, A and B. A is random, B is fixed. Both cross.

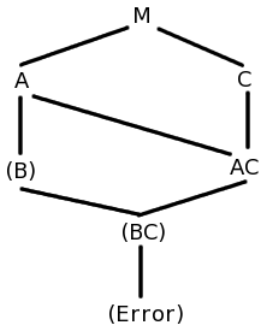


Three factors, A and B are random;  
C is fixed. All cross.





Three factors, A is fixed; B and C are random. C nests in B which nests in A.



Three factors, A and C are fixed; B is random and nests in A. C crosses with A and B.

## Exercise

We have four different kinds of cheddar cheese. We have selected 10 students at random with a rural background, and 10 students at random with an urban background. Each student will eat and rate for bitterness 8 bites of cheese (two of each of the four types in random order).

Many insects respond to odors (of food, for example), but their responses may depend on many factors. In this study, scientists wish to study the effects of genetic background, type of food, and concentration of odor on a physiological response in bees.

The experiment involves 312 bees, 52 from each of six colonies (hives).

The 52 bees from each colony are divided at random into two groups of 26. Group 1 gets control feed and group 2 gets experimental feed.

Each bee is then exposed to three different concentrations of the odor in random order, with a wash-out odor exposure between each experimental condition.

Each bee is measured for its response after exposure to each experimental condition.

We are conducting an experiment on plant growth. The experiment will be conducted at two sites, and our interest is in how the plants grow at these particular two sites. There are four populations of plants, and from within each population we have randomly chosen 10 sub-populations called lines. At each site there are 8 parcels of land called quadrats. At each site, each population is randomly assigned to two quadrats. At each quadrat, all 10 lines from the chosen population are grown. Finally, we take two biomass measurements from each line. This gives us a total of 320 biomass measurements.

A psychologist wishes to explore a theory about interactions between men and women. She has identified four male personality types and five female personality types. Five men of each type are selected at random (total 20 men), and 5 women of each type are selected at random (total 25 women). Each of the 20 men will meet each of the 25 women for a five minute conversation. The conversation is taped, and the psychologist rates each conversation for its degree of interaction.