Random effects and mixed modeling:

General linear models +

fixed and random effects =

General linear mixed models

Outline

- Fixed versus random effects
  - Example with an ANOVA
  - Reasons to use one or the other
- Mixed models
- Lab
  - Going from fixed to random effects

General Linear Models

Linear Regression:

Continuous response, one continuous explanatory variable

T-test:

Continuous response, one discrete explanatory variable with only two categories

One-way ANOVA (Analysis of Variance):

Continuous response, one discrete explanatory variable with more than two categories

Two-way ANOVA:

Continuous response, two discrete explanatory variable

ANCOVA (Analysis of Covariance):

Continuous response, one discrete explanatory variable and one continuous explanatory variable

General Linear Models

So far, we assumed that each parameter in our model is estimated separately and independently:

$$y_i = \alpha + \beta x1_i + \delta x2_i + \cdots + \varepsilon_i$$

$$\varepsilon_i \sim Norm(0, \sigma^2)$$

We refer to each of the parameters as “fixed effects”.

Fixed and random effects

- ***Fixed effects:*** factors whose levels are experimentally determined or whose interest lies in the specific effects of each level, such as effects of covariates, differences among treatments and interactions.
- ***Random effects:*** factors whose levels are sampled from a larger population, or whose interest lies in the variation among them rather than the specific effects of each level.
- The precise definitions of ‘fixed’ and ‘random’ are controversial; the status of particular variables depends on experimental design and context

Fixed and random effects

One-way (factor) ANOVA – fixed effects

- Assume that we measured wing length in five different populations of little owls (*Athene noctua*)
- Is there a difference in wing length among the populations?

Fixed and random effects

One-way (factor) ANOVA – fixed effects

With a means parameterization, we can write the model:

$$y_i = \alpha_{j(i)} + \varepsilon_i$$

$$\varepsilon_i \sim Norm(0, \sigma^2)$$

y_i = observed wing length of owl i in population j

$\alpha_{j(i)}$ = expected wing length of an owl in population j

ε_i = the random wing deviation of owl i from its population mean

Fixed and random effects

- In this model, the levels (groups) of the factor were assumed to be fixed by design; we have a special interest in the particular five little owl populations that we want to compare and have no interest in generalizing to other populations

- But what if aren't actually interested in the differences among the sites? What if instead we considered these five groups a random sample of some larger population of little owl and wanted to generalize our results?

Fixed and random effects

One-way (factor) ANOVA – random effects

In such a case, we can re-write the model:

$$y_i = \alpha_{j(i)} + \varepsilon_i$$

$$\varepsilon_i \sim Norm(0, \sigma^2)$$

$$\alpha_{j(i)} \sim \text{Normal } (\mu, \tau^2)$$

- y_i , $\alpha_{j(i)}$, and ε_i have the same interpretation but now each of the $\alpha_{j(i)}$ parameters are no longer assumed to be independent
- Instead, they come from a second normal distribution with a mean of μ and a variance of τ^2

Fixed and random effects

One-way (factor) ANOVA – random effects

In such a case, we can re-write the model:

$$y_i = \alpha_{j(i)} + \varepsilon_i$$

$$\varepsilon_i \sim Norm(0, \sigma^2)$$

$$\alpha_{j(i)} \sim \text{Normal}\ (\mu, \tau^2)$$

What is the interpretation of μ and τ^2 ?

$\mu =$ mean winglength across all the five little owl populations

$\tau^2 =$ variance in winglength across populations

Fixed and random effects

Which to choose?

Fixed Effects

- You have a particular interest in the studied factor levels
- You have included all conceivable levels in a study
- No interest in the variance among levels
- No interest in generalizing to factor levels that you did not study

Fixed and random effects

Which to choose?

Random Effects

- You don’t have a particular interest in the studied factor levels and/or you could not have sampled all levels
- Interested in the variation among levels (but may still want to understand the effects for the observed levels)
- You want to generalize to a larger population (e.g., levels are more like samples)

Fixed and random effects

Three reasons to go from fixed to random effects:

- Extrapolation to a wider population of inference
- Improved accounting for system uncertainty

  Randomness in ε and both τ. Acknowledges that repeating our study would result in different parameter estimates

- Efficiency of estimation -> shrinkage

  Parameters are no longer independent and will be pulled to the mean, "borrowing strength"

Mixed models

- In many situations, it is useful to make use of both fixed and random effects
- **Mixed models** contain both fixed and random effects
- Introduced by Fisher when studying correlations in traits among relatives
- Particularly useful for:
    - Repeated measures
    - Missing data

Mixed models in R

- Can be hard to fit mixed and random effects models using canned functions in most software packages

    - algorithms do not always converge, especially when the number of groups is small

Mixed models in R

- Can be hard to fit mixed and random effects models using canned functions in most software packages
- In R, can use:

  - ***aov*** function with "Error" term -- if predictors are all categorical and design is balanced

  - ***nlme*** package

  - ***lme4*** package – allows for unbalanced data sets, random effects on parameters (e.g., slope in an ANCOVA), and nonlinear models (not discussed here

Mixed models in R

- What to do when those don’t work? Suggestions from Bolker:

1. Fit fixed-effects models instead. May loose power -> conservative results
2. For nested design (e.g., subsamples in blocks) -> collapse groups data by computing means and do a single level analysis
3. Fit model ignoring blocks and examine the variation in residuals between blocks -> must show between-group variation in residuals is both statistically and biologically irrelevant
4. Switch to a Bayesian analysis -> computing/ software, while more difficult to learn, is much more flexible

Fixed and random effects

Lab: ANOVA

- Review the ANOVA lab from last week
- Modify the data simulation with an assumption that means of populations are random drawn from a common population-level grand mean.