Fixed and Random Effects Models: Basic Concepts and Application to Neuroimaging Research

Sanford Braver & Todd Braver

Introduction

- Our first chance at collaboration and how it came about
- A father-son tag team talk
  - Sanford (Braver) & Son??
- Fixed vs. Random Effects Models
  - Why it's not just arcane ANOVA terminology

Outline

- **Part 1: Brief Refresher course (Sandy)**
  - Basic definitions, distinctions, and computations

- **Part 2: Contact with Neuroimaging & the Broader Issues (Todd)**
  - The state of affairs in neuroimaging data analysis

- **Part 3: The project we are conducting (Sandy)**
  - Monte Carlo simulations

- **Part 4: Insights we’ve gleaned so far (Todd)**
  - General Implications for Design & Analysis

Part 1:
Refresher on
Fixed vs. Random Effects Models

Definitions

- Fixed factor:
  - The levels (or exemplars) of the factor included in the investigation exhaust those in which the investigator has interest
- Random factor:
  - The levels of the factor included in the investigation are a random subset from a potentially infinite population of levels in which the investigator has interest
- Other possibilities
  - Finite

An Example Study

- Two conditions, Experimental (E) and Control (C), are being compared on different groups of subjects
- The Investigator had many Undergrad RAs (UG RAs) assisting in the project (e.g., 10)
  - each collects data on an equal number of E and C subjects
- In addition to exploring the effects of Condition (E vs C), the investigator is curious about differences, if any, in the data collected by the various UG RAs.
- Design:
  - UG RA included as second factor in a 2-factor, between subjects ANOVA

The Data Matrix

UG RA	Experimental	Control
Eve	s1	s1
	s2	s2
	s3	s3
	s4	s4
	s5	s5
	s6	s6
Sam	s1	s1
	s2	s2
	s3	s3
	s4	s4
	s5	s5
	s6	s6
Bob	s1	s1
	s2	s2
	s3	s3
	s4	s4
	s5	s5
	s6	s6
…		
RA10		

Fixed Effects Analysis

- In this study, Condition is always considered a Fixed Factor
- Treating UG RA as either a Fixed Factor or a Random Factor might make some sense
- Fixed Factor:
  - Inference: Do *these particular* RAs obtain different kind of data?
    - a kind of Lab Quality Control assessment
- Design:
  - If UG RA is Fixed as well as Treatment, both factors are Fixed
  - This is called a “Fixed Model” or “Fixed Effects” analysis

Random Effects Analysis

- Random Factor: The interest is in regarding these particular UG RAs as a random subset from a potentially infinite population of such UG RAs
  - Inference: How robust (or generalizable) over UG RAs is the effect of Condition?
- Design:
  - When one factor is Fixed and the Other Random, ANOVA parlance terms it a "Mixed Model".
  - While there is some overlap with this language as applied to "Mixed Model Analysis" (Greg Miller's talk), there is also a number of distinctions
  - For purposes of our talk, we will term it the "Random Effects Model"
    - traditional ANOVA nomenclature reserves the latter term for when BOTH factors are Random, a real rarity

What’s the statistical implication of this decision?

It changes the Error Term for the Fixed Factor

Note all Error Terms (denominator of F-ratios) are the same except that for the Fixed factor

Source	BOTH FACTORS FIXED		FACTOR *A* FIXED; FACTOR *B* RANDOM	
	E(*MS*)	Error Term	*E*(*MS*)	Error Term
A	$\sigma_\epsilon^2 + b(s)(\theta_A^2)$	*S/AB*	$\sigma_\epsilon^2 + s(\sigma_{A \times B}^2) + b(s)(\theta_A^2)$	*A* × *B*
B	$\sigma_\epsilon^2 + a(s)(\theta_B^2)$	*S/AB*	$\sigma_\epsilon^2 + a(s)(\sigma_B^2)$	*S/AB*
A × *B*	$\sigma_\epsilon^2 + s(\theta_{A \times B}^2)$	*S/AB*	$\sigma_\epsilon^2 + s(\sigma_{A \times B}^2)$	*S/AB*
S/AB	σ_ϵ^2		σ_ϵ^2	

From Keppel (1984)

This is so in order that the Expected Mean Square, [E(MS)] of the appropriate Error Term contains all the elements of the term to be tested except that due to the tested effect

The Data Matrix

UG RA	Experimental	Control
Eve	s1	s1
	s2	s2
	s3	s3
	s4	s4
	s5	s5
	s6	s6
Sam	s1	s1
	s2	s2
	s3	s3
	s4	s4
	s5	s5
	s6	s6
Bob	s1	s1
	s2	s2
	s3	s3
	s4	s4
	s5	s5
	s6	s6
...		
RA10		

# of conditions

# of RAs

Source	df	df(number)	MS
Condition	(a-1)	1	200
RA	(b-1)	9	167
Condition X RA	(a-1)(b-1)	9	100
Within (subjects within RA-Condition Groupings)	ab(n-1)	100	40

# of subjects or observations per cell

RA fixed

F(1,100)=5.0, p=.027

RA random

F(1,9)= 2.0, p=.19

Thus, typically Fixed Effect models are more powerful (more frequently attain significance) than Random Effect models, for two reasons:

- Bigger F-values, since MS Within is commonly smaller than MS Interaction
- Larger denominator df, which means Tabled F-values are smaller, easier to exceed
- Possibly because of this, most IVs in psychology are considered Fixed, and Random Effect models are both infrequent and somewhat esoteric
- This is so despite their correspondingly greater capability to provide statistically justified generalizability
  - cf. “Language as Fixed Effect” Fallacy (Herb Clark, 1973)
  - Statistically difficult in ANOVA to include more than one Random Factor simultaneously
  - (But not in Greg Miller’s Mixed Model Analysis)

The Key Exception

- Another experimental design:
  - Repeated Measures: Each subject tested in more than one experimental condition
    - Most common in cognitive behavioral and neuroscience studies
- This design is the key exception to the general preference for Fixed Effect models
  - Subject is treated as a Random Factor
  - Inference: Generalize from sample of subjects to wider population of interest

UG RA	Experimental	Control
Eve	s1	s1
	s2	s2
	s3	s3
	s4	s4
	s5	s5
	s6	s6
Sam	s1	s1
	s2	s2
	s3	s3
	s4	s4
	s5	s5
	s6	s6
Bob	s1	s1
	s2	s2
	s3	s3
	s4	s4
	s5	s5
	s6	s6
…		
RA10		

trial	E(experimental)						C(control)							
Subject	1	2	3	4	5	6	1	2	3	4	5	6	$\bar{E}$	$\bar{C}$
1	7	4					4	0					4.5	3.5
2	1	3					2	3					2.1	2.4
3														
4														
5														
6														
7														
8														
9														
10														

Data matrix could be rearranged as follows

If we substitute “Subject” for “UG RA” and “Trial” for “Subject”, we have:

Add some representative data

One simple, obvious, and quite appropriate device to simplify the analysis is to calculate the average E and the average C response for each subject

- What's left is a 2-dimensional array or 2 Factor design, with Subject as one factor, and average score for each type of trial (or Condition) as another factor: the typical Repeated Measures layout

Subject	$\bar{E}$	$\bar{C}$
1	4.5	3.5
2	2.1	2.4
3		
4		
5		
6		
7		
8		
9		
10		

•Of course, we'd treat Subject as a Random factor, since our interest is in generalizing to a population of subjects, of which ours are regarded as a random subset or sample

•Thus, the Error Term for Condition (a Fixed factor), as in all 2-factor designs with one Fixed factor and one Random factor, is the Subject X Condition interaction

•This makes it not a problem that the "Within" Error Term doesn't exist

•It'also intuitive, since we evaluate the average effect size against the variability in effect size, the interaction

•Exactly the same result, in terms both of the value of the test statistics and the dfs, would be obtained with the Matched t-test, or the One Sample t-test testing the single variable of the mean difference ($\bar{E}-\bar{C}$)

# of rows

Source	df
Condition (E vs C)	(a-1)=1
Subject	(n-1)
Condition XSubject	(a-1)(n-1) =(n-1)
~~Within~~	~~ab(n-1)~~=0

# of observations per cell=1

However, the Preceding Analysis Discards the Trial-by-Trial Data, Which One Might Want To Use Since:

- Subjects Are Expensive, But Trials Are “Cheap”
- The Within-Subject/Over-Trials Variability Might Well Be Small, While the Between-Subject Variability Might be Large
- Putting “Trials” Back In Would Leave the Preceding Analysis With a Non-Zero, Thus Potentially Usable, “Within” Term

Source	df
Condition (E vs C)	(a-1)=1
Subject	(n-1)
Condition X Subject	(a-1)(n-1)
Within	ab(t-1)

# of observations per cell=
Number of trials per subject

Subject	Experimental	Control
1	T1	T1
	T2	T2
	T3	T3
	T4	T4
	T5	T5
	T6	T6
2	T1	T1
	T2	T2
	T3	T3
	T4	T4
	T5	T5
	T6	T6
3	T1	T1
	T2	T2
	T3	T3
	T4	T4
	T5	T5
	T6	T6
…		
10		

- **Just as such data could be viewed in the repeated measures layout, so could it be viewed in the traditional way.**
- **In fact, it could be entered into software using the "one-line-per-level" setup as an alternative to the traditional "one-line-per-subject" setup**
- **See Page, Braver & MacKinnon "*A Guide to SPSS For Analysis of Variance" (2nd Ed)* (forthcoming from Erlbaum) for details**
- **As long as one specifies that the Error Term for Cond is the Cond X Subject interaction (in other words, the Random model) the results replicate the foregoing**

Subj	Cond	Trial	DV score
1	1	1	7
1	1	2	4
1	1	3	2
1	1	4	3
1	1	5	6
1	1	6	2
1	2	1	4
1	2	2	0
1	2	3	3
1	2	4	1
1	2	5	5
1	2	6	3
2	1	1	1
2	1	2	2
2	1	3	3
2	1	4	6
2	1	5	2
2	1	6	5

The Benefits of Using *Within* As the Error Term Are Countered By Its Disadvantages

- **If We Use This "Within" Term as the Error Term for Condition, the Implication Is That We're Treating Subjects As a Fixed Factor**
- **Thus, We Can Not Generalize Beyond the Current Set of Particular Subjects**
- **Instead the Fixed Effects generalization is over trials. The inference: Would the Condition effect replicate if the *same subjects* were run on a different set of *trials*?**
- **Seems like not the appropriate type of inference**
  - **Nonetheless this was the standard analysis in almost all neuroimaging analyses until 1998 (and still continues today)!!**

Subject	Experimental	Control
1	T1	T1
	T2	T2
	T3	T3
	T4	T4
	T5	T5
	T6	T6
2	T1	T1
	T2	T2
	T3	T3
	T4	T4
	T5	T5
	T6	T6
3	T1	T1
	T2	T2
	T3	T3
	T4	T4
	T5	T5
	T6	T6
…		
10		

An apparent alternative

- Could conduct individual analysis of this sort (a test over trials) for each subject
- Could be a t-test (Matched if can correspond each E trial with a certain C trial, otherwise Independent)
- Then could combine the result of each t-test over subjects
  - Meta-analysis
    - Of Effect Sizes
    - Of p-levels with Stouffer’s Z

    OR
  - One sample t-tests
    - Of t-test statistics
    - Of p-levels with Stouffer’s Z
- Howver, unless these use Random Effect meta-analyses, still has the constraint on generalizability over subjects

Part 2:
Contact with Neuroimaging and Broader Issues

Neuroimaging Analyses: Brief History and Sociology

- 1980's and early 1990's
  - PET imaging
  - Constraints:
    - Very Expensive -- only few subjects scanned per study
    - Radiation Concerns -- only few scans (data points) per subject
    - Low Temporal Resolution -- blocked designs
  - Upshot:
    - Limitation on types of experimental designs employed, statistics used

- 1990-1991
  - SPM paper published, SPM software made available (Karl Friston)
  - Very general approach
    - Calculate simple statistics (t,F,z) across each element (voxel) in the image
  - Widely accessible and Influential
    - Software made freely available from prominent center (Hammersmith), maintained and updated regularly
    - Accompanied by regular authoritative methodological reports Downside:
    - Methodology and Software are pretty opaque to users

Neuroimaging Analyses: Brief History and Sociology

- **1992 - 1997**
  - Development of fMRI
    - Less constraints: Scan more subjects, Collect more samples per subject
    - Event-related fMRI: Trial-based analyses
      - Very similar to standard cognitive behavioral experimental designs
  - Development of General Linear Model in SPM software
    - Flexible approach that can handle more complex experimental designs
    - Can be used for event-related fMRI as well
    - Downside: Increased complexity, increased opaqueness
  - SPM becomes dominant software tool for neuroimaging
    - Probably used by about 80% neuroimaging researchers
- **1998**
  - Conference presentation (Holmes & Friston) introduces distinction between Fixed and Random effects analysis to field
  - Adds revelation that SPM until then was only Fixed effect, adds capability for Random effects

“Current methods of assessing functional neuroimaging data utilise only inter-scan variance … [they] assume the [subject] parameters are *fixed effects* and test for significant mean activation for these subjects….When subjects respond differentially, as they almost certainly do, this intra-subject variance is inappropriate for inter-subject inference…. Appropriate *random effects* analyses for multi-subject functional neuroimaging data are essential for valid population inference. They are standard in other fields, and are beginning to be demanded by discerning journals” *(Holmes & Friston, 1998)*

Neuroimaging Analyses: Brief History and Sociology

- 1999
  - Publication by Friston & colleagues of papers discussing distinction between Fixed & Random effects models
  - Semi-defense of Fixed effects with new analysis to show Fixed effects analyses can be used for population inference
  - Downside: Issue becomes even more confused
- The current state of affairs
  - Many groups now use Random Effects regularly, but many still use Fixed Effects
  - Also, other alternatives being considered (e.g., meta-analysis approaches)
  - Lack of clarity regarding the implications and consequences of this choice

Part 3: Our Research Project

Project Overview

- Our goal:
  - Systematically investigate the similarities and differences between different analysis approaches
    - For example: Under what conditions of the data do Fixed and Random effects analyses diverge?
- Secondary goal:
  - Are there other viable alternatives?
  - Consideration of data structure suggests multiple possible routes to statistical inference
    - Some approaches are already being utilized in neuroimaging: meta-analysis (group-average Z)

Our Approach

- Conduct “Monte Carlo” Simulations or Experiments
  - Empirical Comparisons of different methods
  - Simulated on the computer -- SAS statistical software
- The technique
  - Set up certain desired “parameters”
    - e.g., number of trials, number of subjects, between subject variability, within subject variability
  - Have computer generate a random sample from a population with those parameters
  - Calculate significance using each of the alternative methods
  - Do this for many such random samples (replications)--we’re doing 1000
  - Tabulate percent significance by each statistical method
  - Repeat entire process aAt least twice
    - Once when the null hypothesis is true: Type I Error or False Positives
      - Acceptable techniques contain the rate of significance at alpha level (e.g., .05)
    - Again when null is false: Power
      - Preferred techniques. among those acceptable on False Positive containment basis, have highest power
  - Repeat with different parameter settings
    - e.g., low within-subject variability, high between-subject variability

Techniques to be Empirically Compared

1. Fixed Effect ANOVA (i.e., Friston's)
2. Random Effect ANOVA (same as Matched t-test on average E vs average C; same as Mulitvariate Approach Repeated Measures Contrast)
3. Univariate Approach Repeated Measures Contrast
4. t-tests on each subject
   a. Matched vs Independent
   b. Combined by Meta-Analysis
      A. Stouffer's Z-tests on p-level
      B. Effect Size method
   c. Each with tests both of overall significance and heterogeneity of Effect Size over Subjects
   d. Or tested with One Sample t-tests

ALSO intend to add

5. Mixed Model Analysis
6. Cross-Sectional Interrupted Time Series Analysis (Simonton)
7. Techniques that correct for the Non-Independence over Trials (Kenny)
8. Quasi-F's (Clark) to generalize over trials and subjects (two Random Factors) simultaneously

Part 4:   Insights we’ve gleaned so far

When Fixed and Random Models diverge

- **Preliminary simulations showed:**
  - Under null conditions the two models are pretty similar
  - In real data sets (behavioral - RT data), sometimes Fixed much more significant
- **Key Insight**
  - The magnitude of the Subject x Condition interaction term will directly indicate the difference between Fixed Effect and Random Effect F-ratios for Main Effect of Condition
  - This can be shown analytically:
    - Ratio of Fixed Effect to Random Effect F-ratios for Condition effect is directly equal to Subject X Condition F-ratio
    - When SxC = 1 (i.e., expected value under null), Fixed = Random (F-ratio, not p-value)
    - When SxC >> 1 (i.e., significant effect), Fixed >> Random

$$\text{Interaction} = \frac{S \times C}{\text{Within Error}}$$

$$\text{Fixed} = \frac{C}{\text{Within Error}}$$

$$\text{Random} = \frac{C}{S \times C}$$

$$\text{Interaction} = \frac{\text{Fixed}}{\text{Random}}$$

How common are Subject x Condition interactions?

- Maybe very common!
  - At least in behavioral data sets

- Examples:
  - AX-CPT: Effect of monetary incentive on BX RT
    - Random Effect: $F(1,24)=0.14$, $p=.711$
    - Fixed Effect: $F(1,395)=0.53$, $p=.468$
    - Subj X Cond: $F(24,395)=3.76$, $p<.000001$
  - Stroop: Effect of Cong/Inc probability on Neutral RT
    - Random Effect: $F(1,29)=1.05$, $p=.314$
    - Fixed Effect: $F(1,1799)= 2.93$, $p=.087$
    - Subj X Cond: $F(29,1799)=2.80$, $p=.000001$
  - Nback: Effect of WM Load (1B vs. 3B) on Target RT
    - Random Effect: $F(1,11)= 14.7$, $p=.003$
    - Fixed Effect: $F(1,1553)= 95.5$, $p<.000001$
    - Subj X Cond: $F(11,1553)= 6.49$, $p<.000001$

Implications of Subject x Condition Interactions

- SxC interactions are typically never directly examined!
  - In cognitive experimental studies
  - In neuroimaging studies
  - It is not yet established how prevalent they are -- Best Guess is Pretty Prevalent

- The standard ANOVA dictum:
  - Condition effect differs in magnitude across subjects
    - More than would be expected by chance if they came from same population
  - An interaction qualifies interpretation of main effect
  - Need to examine the pattern of interaction!
    - Least problematic
      - Virtually all subjects show effect, but to differing extents
    - But could also find -
      - Cross-over effect: Some subjects show significant "reverse" pattern
      - Null responders: Non-trivial proportion of subjects show no effect
        - What proportion is a cause for concern? (e.g., 50% or more)

- The upside
  - Variability in condition effect across subjects is informative!
  - Potential to find individual difference effects, covariates, other moderators
    - Without sufficient sample variability there can be no power to detect correlates

Take Home messages

Both behavioral and neuroimaging studies are not appropriately utilizing the data that is collected!!

- Imaging Studies
  - Results from Fixed Effects models can lead to misleading inference if due to large Subject x Condition interactions
  - Random Effects analyses ignore within-subject variability and are completely dependent upon sample size for power
- Behavioral studies (i.e., Repeated Measures designs such as RT studies)
  - Do not test for Subject x Condition interactions
  - Are blind to the degree and pattern of variability in effect size across studied subjects
- It is not yet clear which analysis model -- Random, Fixed, or Alternative -- is best suited for the types of data typically acquired in behavioral and imaging studies
  - Hopefully our Monte Carlo project will provide useful answers!

trial	E(experimental)						C(control)					
Subject	1	2	3	4	5	6	1	2	3	4	5	6
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												

- **An apparently different analysis retains the full trial-by-trial data**
- **Then, conducting a “Repeated Measures Contrast” involves giving contrast weights of +1 to all the E trials and -1 to all the C trials**
- **This would yield the same numerator as the previous analysis**
- **But the denominator would differ; it would be the full Trials X Subjects interaction, with df=(n-1)(t-1)**
- **This is only so in the “univariate” analysis (or under the “mixed” model), which in turn depends on the rarely met but influential “sphericity” assumption**
- **If not assumed or not met, the alternative “multivariate” approach (which makes no such assumption) would use the MS for EvsC Contrast X Subject as the Error**
- **This in turn is exactly equal, both in test value and in df, to the preceding**