Random Effects Models

A factor is called a random effects factor if the levels of the factor represent a larger set of interest.

Examples:

1. Medicine: How accurate are labs for testing for a certain disease? Do labs differ in their accuracy? Suppose we have (different) people tested at 3 different labs.
Factor = Lab (i = 1, 2, 3)
Unit = a person having a medical test
Y_{ij} = accuracy rating of the test for person j and lab i

n_i = number of people tested at Lab *i*

Lab is a fixed effect if we care only about those labs.
Lab is a random effect if the 3 labs are a random sample of all such labs.

2. Education: How well do California students learn to read by the end of first grade? Choose 6 schools in California. Then randomly choose n_i students in school *i* to take a reading test.
Factor = School (*i* = 1 to 6)
Unit = student (*j* = 1 to n_i)
Y_{ij} = reading score for student *j* in school *i*.

School is a fixed effect if we care only about those 6 schools.
School is a random effect if those schools are randomly sampled from a larger set of interest.

3. Psychology: Compare therapists for effectiveness.
Factor = therapist
Unit = patient
Y_{ij} = change in score on depression test after one year of therapy for patient *j*, therapist *i*.

Therapist is a fixed effect if we are interested in those specific therapists
Therapist is a random effect if the therapists are randomly selected from all therapists of interest.

RANDOM EFFECTS MODEL (One factor only):

$Y_{ij} = \mu_i + \varepsilon_{ij} = \mu_. + \tau_i + \varepsilon_{ij}$ for $i = 1, \ldots, k; j = 1, \ldots, n_i$

Where:
1. ε_{ij} is $N(0, \sigma^2)$ as before
2. μ_i is $N(\mu_., \sigma_\mu^2)$, so the μ_i are *not* considered fixed as they were before.
3. $\mu_.$ is a fixed constant.
4. All μ_i and ε_{ij} are independent.

<u>See picture drawn on white board</u>: Normal curve with $\mu_.$ as the mean and σ_μ^2 as the variance, each μ_i drawn from it. Then normal curves with each μ_i as the mean and σ^2 as the variance, each Y_{ij} drawn from that.

EXAMPLE: Therapists and depression scores

$\mu_.$ is the overall mean change in depression scores for the *population* of *all possible patients* (not just those treated) for *all possible therapists* of the type the sample was drawn from.

μ_i is the overall mean change in depression scores for the *population* of *all possible patients* if they were to have therapist i.

σ^2 is the variance of the changes in depression scores for the population for *any* particular therapist.

σ_μ^2 is the variance of the *means* μ_i for all possible therapists (not just the ones in the study), so it's the variability *across* therapists in the population.

The variance of Y_{ij} is made of *two* components. It's variance of $(\mu_i + \varepsilon_{ij}) = \sigma^2 + \sigma_\mu^2$ = variance across all possible patients and all possible therapists.

<u>To summarize</u> fixed versus random effects model for one factor:

Fixed effects model, Y_{ij} is $N(\mu_i, \sigma^2)$ and all independent.

Random effects model, Y_{ij} is $N(\mu_., \sigma^2 + \sigma_\mu^2)$ and *not* all independent.

Mixed Models - Fixed And Random Effects; Simplest - Randomized Block Design

Blocks are similar to pairing when there are only two factor levels (groups). They are used to reduce extraneous variability in responses. Often, block = a person or "subject."

Example: Does skin response change for different *imagined* responses?
Y_{ij} = Skin potential for person (block) i and emotion j. Note that i = 1 to 8 and j = 1 to 4 *for each person*.

Model: $Y_{ij} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij}$ where ε_{ij} are $N(0, \sigma^2)$

Blocks = Factor A = person, fixed or random. If fixed, $\sum_{i=1}^{a} \rho_i = 0$; if random, ρ_i are $N(0, \sigma^2_\rho)$

Treatment = Factor B = emotion. This is a fixed effect, and thus $\sum_{j=1}^{b} \tau_j = 0$.

Note: We cannot include *interaction* in the model because there is no way to estimate both interaction and error. This is because there is only one observation in each *ij* combination.

ANOVA Table for Randomized Block Design (blocks as fixed effect*):

Source	SS	df	MS	E{MS}	F*
Blocks (Subjects)	SSBL	a – 1	MSB	$\sigma^2 + \frac{b\sum \rho_i^2}{a-1}$	Not tested
Treatments	SSTR	b – 1	MSTR	$\sigma^2 + \frac{a\sum \tau_i^2}{b-1}$	$\frac{MSTR}{MSE}$
Error = Blks x Trts	SSE	(a – 1)(b – 1)	MSE	σ^2	
TOTAL	SSTO	ab – 1			

*When blocks considered to be random, the only difference is E{MSBL} = $\sigma^2 + b\,\sigma^2_\rho$

Example: H_0: all $\tau_j = 0$ (*population* mean skin response same for all 4 imagined emotions)

F* = 3.47, *p*-value = .034, reject the null hypothesis, conclude population means differ.

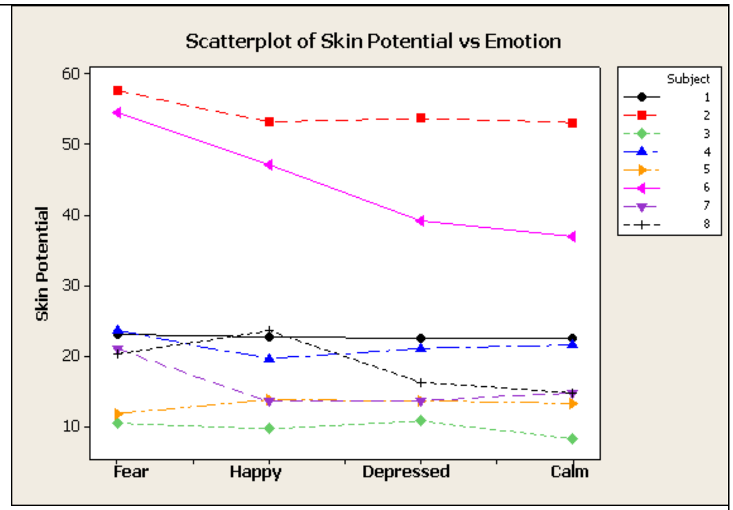
Tukey multiple comparisons: Calm differs from Fearful, all other C.I.s cover 0.

See computer output for what happens if blocks are ignored. F* = 0.13, *p*-value = 0.941!

Clearly a good idea to use blocks. Natural variability *across* people is huge. Blocks control for that. See plot. Two people have very high skin potential measurements in general, others are much lower.

Randomized Block Design (Subjects = Blocks, Emotions = Treatments)
Eight subjects were hypnotized and asked to imagine four emotions. Session order was randomized for each subject. Skin potential was measured as the response variable.
Ref: "Physiological effects during hypnotically requested emotions" ***Psychosomatic Medicine***, 1963, pp. 334-43. (Also Cobb, ***Des. & Analysis of Exp.***)

Subj.	Calm	Depressed	Fearful	Happy	Mean
1	22.600	22.500	23.100	22.700	22.7
2	53.100	53.700	57.600	53.200	54.4
3	8.300	10.800	10.500	9.700	9.8
4	21.600	21.100	23.600	19.600	21.5
5	13.300	13.700	11.900	13.800	13.2
6	37.000	39.200	54.600	47.100	44.5
7	14.800	13.700	21.000	13.600	15.8
8	14.800	16.300	20.300	23.600	18.7
Mean:	23.2	23.9	27.8	25.4	



Analysis of Variance

Source	DF	SS	MS	F	P
Subject	7	7021.87	1003.12	102.86	0.000
Emotions	3	101.43	33.81	3.47	0.034
Subj.xEmotion	21	204.80	9.75		
Total	31	7328.10			

One-way Analysis of Variance (*Ignoring Blocks*)

Source	DF	SS	MS	F	P
Emotions	3	101	34	0.13	0.941
Error	28	7227	258		
Total	31	7328			

Tukey 95.0% Simultaneous Confidence Intervals
Response Variable Skin Potential
All Pairwise Comparisons among Levels of Emotions

```
Emotions = Calm subtracted from:
Emotions     Lower    Center     Upper  ----+---------+---------+---------+--
Depresse    -3.663    0.6875     5.038         (--------*---------)
Fear         0.287    4.6375     8.988                 (--------*---------)
Happy       -2.125    2.2250     6.575            (--------*---------)
                                        ----+---------+---------+---------+--
                                          -5.0       0.0       5.0      10.0

Emotions = Depresse subtracted from:
Emotions     Lower    Center     Upper  ----+---------+---------+---------+--
Fear        -0.400     3.950     8.300               (---------*---------)
Happy       -2.813     1.538     5.888          (---------*---------)
                                        ----+---------+---------+---------+--
                                          -5.0       0.0       5.0      10.0

Emotions = Fear subtracted from:
Emotions     Lower    Center     Upper  ----+---------+---------+---------+--
Happy       -6.763    -2.412     1.938  (--------*---------)
                                        ----+---------+---------+---------+--
```

NESTED FACTORS

Factors A and B are **crossed** if all levels of Factor B are measured at all levels of Factor A.

Factor B is **nested** under Factor A if the actual *levels* of Factor B *differ* for each level of Factor A.

<u>Example</u>: Compare on-time performance of airlines coming into John Wayne Airport.

Factor A = Airline, 3 "levels" are American, Delta and United

Factor B = City, 2 levels for each airline, so *j* = 2, but *different* cities for each airline.

Randomly sample 10 flights out of the past year for each city/airline combination.

	AIRLINE		
CITY	**American** (i = 1)	**Delta** (i = 2)	**United** (i = 3)
Dallas (j = 1)	Group 1 = 10 flights		
Chicago (j = 2)	Group 2 = 10 flights		
Atlanta (j = 1)		Group 3 = 10 flights	
Minneap. (j = 2)		Group 4 = 10 flights	
Denver (j = 1)			Group 5 = 10 flights
San Fran. (j = 2)			Group 6 = 10 flights

Y_{ijk} = Minutes late for Flight *k*, from City *j* on Airline *i*.

Model has Airline + City (nested under Airline):
$Y_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \varepsilon$

where $\mu_{..}$ = overall mean
α_i = how airline *i* differs from the overall average = $\mu_{i.} - \mu_{..}$
$\beta_{j(i)}$ = how city *j* differs from Airline *i* overall average = $\mu_{ij} - \mu_i$

Examples:
α_1 = How much American's average differs from the overall average
$\beta_{1(2)}$ = How much the average for Atlanta differs from Delta's combined average

<u>Test for Factor A</u>: Do all airlines have same average "on time" performance? (All $\alpha_i = 0$)
<u>Test for Factor B</u>: Does the "on time" performance within airlines change based on what city the flight is from? (All $\beta_{j(i)} = 0$)

NOTE: Can't have A*B interaction because levels of B are *different* for each level of A.

REPEATED MEASURES DESIGNS

- Same units (subjects, participants) at *all* levels of one or more factors.
- Randomized block design is a special case of a repeated measures design.

Definition of two types of factors:
- *Between subjects factor* or *Grouping factor*: The subjects are *nested* under the factor.
- *Within subjects factor* or *repeated measures factor*: The subjects are *crossed* with the factor, so all subjects are measured at ("within") all levels of the factor. Measurements are *repeatedly taken* on the same subjects.

NOTE: Can't have interactions between Subjects and Grouping factors because each subject is only measured at one level of a grouping factor.

Simplest case: One factor, could be either a Between subjects or Within subjects factor

Which type determined by what groups are used	Factor A Levels		
	A1	A2	A3
Between subjects No repeated measures	Group 1	Group 2	Group 3
Within subjects Repeated measures	Group 1	Group 1	Group 1

For Randomized Block Design, the "Treatment Factor" is a Within subject factors.

Example: Imagined emotion – all subjects measured for ("within") all 4 emotions

Expanded version of the "Emotions" example: Have one group of subjects do it under hypnosis, and the other group not under hypnosis. Everyone imagines all 4 emotions:

	Happy	Fear	Calm	Depressed
No hypnosis	Group 1	Group 1	Group 1	Group 1
Hypnosis	Group 2	Group 2	Group 2	Group 2

Factor A: Hypnotized or not, separate groups under each condition; "Between subjects"
Factor B: Imagined emotion, same groups do all 4 emotions; "Within subjects"

Can't have Subject * hypnosis interaction, because no subjects were measured at both hypnosis and no hypnosis.

Can have Emotion * Hypnosis because all emotions measured for both hypnosis and no hypnosis.

So the model would have Emotion + Hypnosis + Subject + Emotion*Hypnosis

Another Randomized Block Design Example

Does listening to Mozart increase IQ, at least temporarily?

- Study done at UCI, published in *Nature*, Oct 14, 1993, page 611
- Randomized block design:
- Factor A: Three listening conditions (Mozart, relaxation tape, silence)
- Block: Student; there were 36 students and they *all* listened to *all* conditions.
- So Factor A is *crossed* with Student.

Y = results of an IQ test

Picture it this way, where Group = group of units tested:

	Listening Condition		
	Mozart	Relaxation tape	Silence
Units tested	Group 1	Group 1	Group 1

Model has Treatment + Blocks

If *different* people had been used, it would be called a *completely randomized design:*

	Listening Condition		
	Mozart	Relaxation tape	Silence
Units tested	Group 1	Group 2	Group 3

In that case, Model has Treatment only (no blocks)

ANOVA Table for the actual experiment (randomized block):

Source	SS	df	MS	F	p-value
Blocks	Not used	35	Not used		
Treatments	1752	2	876	7.1	.002
Error	8610	70	123		

Reject null hypothesis of equal average IQ after each condition (in population, if they were to listen). Multiple comparison showed that Mozart average was significantly higher than the other two conditions, which did not differ significantly from each other.

SUMMARY OF WHAT YOU NEED TO BE ABLE TO IDENTIFY

If you can identify each of these, you can usually tell your computer software what model to use. You don't need to know how to work out the details yourself; you do need to know how to interpret the results:

- Each factor, including Subjects if that should be a factor. It should only be a factor if each subject (unit) is measured more than once.
- Number of levels of each factor, and what the levels are.
- Whether each factor is fixed or random.
- Whether any factors are nested under other factors
- Which interactions can be included in the model

Some examples for practice (answers will be on website):

1. Three methods for treating migraine headaches are to be compared. Ninety persistent migraine sufferers are recruited for the study and 30 are randomly assigned to each of the three methods. Response is a measure of decrease in pain intensity.

2. A study is done to determine whether two issues affect test performance for students. The first issue is whether or not there is time pressure to finish. The second issue is whether the student takes the test in the same classroom as where the class has been held, or in an unfamiliar classroom. Students from a large class are randomly assigned to one of these four conditions, and performance on the test is the response variable.

3. A company has a national chain of hundreds of weight loss clinics, which offer a combination of diet and exercise programs. They have two diet plans and three exercise programs, and want to know what works best. They randomly select 10 of their clubs to participate in an experiment. Within each club they recruit 120 volunteers, and randomly assign 20 of them to each diet x exercise combination. The response variable is amount of weight lost over a 10 week period.

4. An employer wants to compare three types of computer keyboards to see which type to buy. They randomly choose 10 employees whose main job is to type. They have each of the 10 employees try each keyboard for a day, with the order randomly assigned. The response is a measure of productivity for that day.

5. Same as #4, except now the company wants to know if there is a difference between males and females, given that males tend to have bigger hands than females do. Therefore, they randomly choose 10 males and 10 females, and have each of these 20 employees try each keyboard for a day.

<u>Continuing random effects model</u>: For two units from the *same* level (same *i*), covariance between Y_{ij} and $Y_{ij'}$ is σ_μ^2 and correlation is $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma^2} = \frac{\text{Variance of } \mu_i}{\text{Variance of } Y_{ij}}$ called the *intraclass correlation*. It represents the correlation between 2 population units in the *same* class or level of the factor. It's *large* if the variability is large *between* levels compared to *within* levels (large σ_μ^2 compared to σ^2). This makes sense, because then knowing the value for one unit *within* a level gives good info about other units in that same level. If there's lots of variability *within* a level, then knowing one value doesn't help so much with others. It also represents the percent of variability in the Y_{ij} that can be attributed to the variability in the μ_i's. (<u>See picture on white board</u>.)

QUESTIONS OF INTEREST FOR THE RANDOM EFFECTS MODEL:

1. Test H_0: $\sigma_\mu^2 = 0$ vs Ha: $\sigma_\mu^2 > 0$. If the null hypothesis is true, all *possible* μ_i are equal.
Example: Are *all* therapists equally effective, on average?
How to do this: EASY! Still use the same F test as in fixed effects one-way ANOVA,
$F^* = MSR/MSE$. In this case, $E\{MSE\} = \sigma^2$ and $E\{MSR\} = \sigma^2 + n' \sigma_\mu^2$ where n' is a weighted sample size.

2. Estimate $\mu_{.}$, Use $\bar{Y}_{..}$ = the overall mean. To get a confidence interval, use MSR instead of MSE as estimate of variance, because now the natural variability in the Y's includes both σ^2 and σ_μ^2.

3. Estimate each component of the variances.

4. Estimate the intraclass correlation.

Details are messy – use computer!

Unbalanced two-factor ANOVA

The term "unbalanced" means that the sample sizes n_{ij} are not all equal. A balanced design is one in which all $n_{ij} = n$.

In the unbalanced case, there are two ways to define sums of squares for factors A and B.

The method SAS calls Type III sums of squares goes by the name "partial SS" or "adjusted SS." It's the default in many programs, but not in R. For this method, here are the full and reduced models being tested for Factor A, Factor B, and the AB interaction:

(I've left the subscripts off in all of the following model statements. They should be obvious.)

Factor A:
Full model is $Y = \mu + \alpha + \beta + \alpha\beta + \varepsilon$
Reduced model is $Y = \mu + \beta + \alpha\beta + \varepsilon$

Factor B:
Full model is $Y = \mu + \alpha + \beta + \alpha\beta + \varepsilon$
Reduced model is $Y = \mu + \alpha + \alpha\beta + \varepsilon$

AB interaction:
Full model is $Y = \mu + \alpha + \beta + \alpha\beta + \varepsilon$
Reduced model is $Y = \mu + \alpha + \beta + \varepsilon$

The method SAS calls Type I sums of squares is usually called sequential sums of squares. It is the default in R. To get it in programs like Stata and Minitab, you need to ask for the "sequential" sums of squares.

Factor A:
Full model is $Y = \mu + \alpha + \varepsilon$
Reduced model is $Y = \mu + \varepsilon$

Factor B:
Full model is $Y = \mu + \alpha + \beta + \varepsilon$
Reduced model is $Y = \mu + \alpha + \varepsilon$

AB interaction: Same as above.

For Type I SS, notice that it matters what order you use to name the factors, whereas it doesn't matter for Type III SS.