Unbalanced two-factor ANOVA

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

In the unbalanced case, there are 2 ways to define sums of squares for factors A and B.

1. SAS used notation that has gone beyond SAS. “Type III sums of squares” same as “partial SS” or “adjusted SS.” It’s the default in many programs, but not in R. Here are the full and reduced models being tested for Factor A, Factor B, and AB interaction:

(I’ve left the subscripts off in all of the following model statements.)

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 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 for the adjusted SS method.

NOTES:
- 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.
- Type III usually makes the most sense! It tests whether the population means are equal for the levels of a factor, where the means use equal weights for across levels of the other factor. But you have to work hard to get these in R.
- Type I tests whether weighted means are equal, with weights defined proportional to the sample size in the other factor. BUT that’s only for the factor listed first. For the one listed 2nd, the test is a strange combination of weights!
Seat Location x Alcoholic drinks example:

Default in R, Sequential, Type 1:  
Type 3; need “car” package first:

```
> anova(lm(GPA ~ Seat*Alc, data=SeatsData))
Analysis of Variance Table
Response: GPA
  Df Sum Sq Mean Sq F value  Pr(>F)
Seat  2 4.359  2.17932  7.0811 0.000969
Alc  2 1.122  0.56091  1.8225 0.163175
Seat:Alc 4 1.395  0.34878  1.1333 0.340593
Residuals 344 105.872  0.30777
```

```
> Anova(lm(GPA ~ Seat*Alc, data=SeatsData, contrasts=list(Seat=contr.sum, Alc=contr.sum)), type=3)
Anova Table (Type III tests)
Response: GPA
  Sum Sq Df F value  Pr(>F)
(Intercept) 1820.92  1 5916.5562 < 2e-16 ***
Seat  2.81  2  4.5584 0.01112 *
Alc  1.30  2  2.1163 0.12204
Seat:Alc  1.40  4  1.1333 0.34059
Residuals 105.87 344
```

```
> anova(lm(GPA ~ Alc*Seat, data=SeatsData))
Analysis of Variance Table
Response: GPA
  Df Sum Sq Mean Sq F value  Pr(>F)
Alc  2  2.312  1.15616  3.7566 0.024327
Seat  2  3.168  1.58407  5.1470 0.006273
Alc:Seat 4  1.395  0.34878  1.1333 0.340593
Residuals 344 105.872  0.30777
```

```
> Anova(lm(GPA ~ Alc*Seat, data=SeatsData, contrasts=list(Seat=contr.sum, Alc=contr.sum)), type=3)
Anova Table (Type III tests)
Response: GPA
  Sum Sq Df F value  Pr(>F)
(Intercept) 1820.92  1 5916.5562 < 2e-16 ***
Alc  1.30  2  2.1163 0.12204
Seat  2.81  2  4.5584 0.01112 *
Alc:Seat  1.40  4  1.1333 0.34059
Residuals 105.87 344
```

```
> anova(lm(GPA ~ Alc, data=SeatsData))
Analysis of Variance Table
Response: GPA
  Df Sum Sq Mean Sq F value  Pr(>F)
Alc  2  2.442  1.22111  3.8818 0.0215
Residuals 353 111.044  0.31457
```

Note that MSE is not the same as in two-factor model, so tests are not the same even when Alc is brought in first in two-factor model.
Random Effects Models

Factors can be either fixed or random.

A factor is called a fixed effects factor if the levels of the factor in the study are the only levels (categories) of interest.

A factor is called a random effects factor if the levels of the factor in the study represent a larger set of possible levels, and interest is in how much the response varies in the population of possible levels of the factor.

One-Factor Anova Example:
How well do California students learn to read by the end of first grade? Choose $K = 6$ schools in California. Randomly choose $n_k$ students in school $k$ to take a reading test.

- Factor = School ($k = 1$ to 6)
- Unit = student ($i = 1$ to $n_k$)
- $Y_{ik}$ = response variable = reading score for student $i$ in school $k$.

- 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, such as all elementary schools in California.
- For random effect situation, we don’t care about means for those particular schools. Instead, we want to know how much variability there is in reading scores across schools, compared to within a school.
RANDOM EFFECTS MODEL (One factor only):

\[ Y_{ik} = \mu_k + \varepsilon_{ik} = \mu + \alpha_k + \varepsilon_{ik} \quad \text{for } k = 1, \ldots, K; \ i = 1, \ldots, n_k \]

Where:
1. \( \varepsilon_{ik} \) is \( N(0, \sigma) \) as before
2. \( \mu_k \) is \( N(\mu, \sigma_\mu) \), so the \( \mu_k \) are random and not considered fixed as they were before.
3. \( \mu \) is a fixed constant = mean of all possible \( \mu_k \).
4. All \( \mu_k \) and \( \varepsilon_{ik} \) are independent.

See picture drawn on white board: Normal curve with \( \mu \) as the mean and \( \sigma_\mu \) as the standard deviation, each \( \mu_k \) drawn from it. Then normal curves with each \( \mu_k \) as the mean and \( \sigma \) as the standard deviation, each \( Y_{ik} \) drawn from that.

EXAMPLE: California schools and Reading scores
Randomly sample \( K = 6 \) schools.
\( \mu \) = the overall mean reading score for all first-grade students in California.
\( \mu_k \) = the mean reading score for all first-grade students at School \( k \); it is \( N(\mu, \sigma_\mu) \)
\( \sigma_\mu \) = Standard deviation among all possible \( \mu_k \) for all elementary schools in California
\( \sigma \) = Standard deviation of reading scores for all first-grade students within each school.
EXAMPLE 2: Therapists and depression scores
Compare therapists for effectiveness.
Factor = therapist
Unit = patient
\( Y_{ik} \) = change in depression test score after one year of therapy for patient \( i \), therapist \( k \).

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.

\( \mu \) = 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.

\( \mu_k \) = the overall mean change in depression scores for the population of all possible patients if they were to have therapist \( k \).

\( \sigma \) is the standard deviation of the changes in depression scores for the population for any particular therapist.

\( \sigma_\mu \) is the standard deviation of the means \( \mu_k \) for all possible therapists (not just the ones in the study), so it’s the variability across therapists in the population.
EXAMPLE 3: 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 \( (k = 1, 2, 3) \)

Unit = a person having a medical test

\( Y_{ik} \) = accuracy rating of the test for person \( i \) and lab \( k \)

\( n_k \) = number of people tested at Lab \( k \)

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.

MORE DETAILS ABOUT THE MODEL (Using therapist example)

The test of interest for Random effects model is \( H_0: \sigma^2 = 0 \) vs \( H_a: \sigma^2 > 0 \)

If \( H_0: \sigma^2 = 0 \) is true, it means that *all* \( \mu_k \) are equal, not just those in the study.

The variance of \( Y_{ik} \) is made of *two* components. It is the variance of \( (\mu_k + \varepsilon_{ik}) \)

\( = \sigma^2 + \sigma_\mu^2 \) = variance across all possible patients and all possible therapists.

To summarize fixed versus random effects model for one factor:

**Fixed effects:** \( Y_{ik} \) are \( N(\mu_k, \sigma) \) and all independent; test \( H_0: \mu_1 \) to \( \mu_K \) all equal.

**Random effects:** \( Y_{ik} \) are \( N(\mu, \sqrt{\sigma^2 + \sigma_\mu^2}) \) and *not* all independent; test \( H_0: \sigma_\mu^2 = 0 \).
ANALYSIS:

For one-factor ANOVA, the F-test is the same whether it’s a fixed or random effect.

- If factor is a fixed effect, test:
  \[ H_0: \mu_1 \text{ to } \mu_K \text{ all equal} \]
  \[ H_a: \text{ at least one is different} \]
- If factor is a random effect, test:
  \[ H_0: \sigma_\mu^2 = 0 \]
  \[ H_a: \sigma_\mu^2 > 0 \]

In both cases, use \( F = \frac{MS_{Groups}}{MSE} \), and compare to \( F \) with \( df = K-1, n-K \)

For two-factor ANOVA (and higher): Denominator of tests change. Use R package \texttt{lme4}. (We will not have time to do this.) \( L\text{me} = \) “linear mixed effects”
Mixed Models - Fixed and Random Effects; Simplest - Randomized Block Design

Recall from last lecture, **blocks** are to reduce known, extraneous variability in responses. Blocks are similar to pairing (used when there are only two factor levels or groups). Often, a block = a person or “subject.”

Blocks are *almost always* considered to be random effects, because interest is not in those specific individuals or blocks.

Examples:

1. Last time, Factor A = Exam type, Factor B = block = Student
   Each student took an exam of each type. There was only $n = 1$ observation per AB (exam type x student) combination.
   - Response = exam score

There are 2 Factors:
   - Exam type, fixed, 4 levels
   - Student, random, 5 levels

There is $n = 1$ observation in each cell
2. Expanded California reading example. Suppose the Department of Education wants to compare 3 methods for teaching reading to 1st graders. Schools differ, so randomly choose 6 schools to do the experiment, then randomly assign $n = 30$ students within each school to learn under each method.

- **Response** = reading score

There are 2 Factors:
- Factor A = Learning method, with 3 levels, *fixed effect*
- Factor B = School, with 6 levels, *random effect*

There are $n = 30$ students in each cell, where a cell is an AB (Learning method x School) combination.

Note: “Student” is not a factor because there is only one observation *per* student.
NESTED FACTORS

Whether they are fixed or random, Factors A and B are **crossed** if all of the specific levels of Factor B are measured at all of the specific 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.

**Example:** Compare on-time performance of airlines coming into John Wayne Airport.

\[ Y_{ijk} = \text{Minutes late for Flight } i, \text{ from City } j \text{ on Airline } k. \]

Factor A = Airline, \( K = 3 \) levels, which are American, Delta and United.

Factor B = Originating city, \( J = 2 \) levels for each airline, but *different* cities for each airline. They don’t all fly from the same places.

Randomly sample \( n = 10 \) flights out of the past year for each city/airline combination.
### AIRLINE

<table>
<thead>
<tr>
<th>CITY</th>
<th>American ($k = 1$)</th>
<th>Delta ($k = 2$)</th>
<th>United ($k = 3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dallas ($j = 1$)</td>
<td>Group 1 = 10 flights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago ($j = 2$)</td>
<td>Group 2 = 10 flights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlanta ($j = 1$)</td>
<td></td>
<td>Group 3 = 10 flights</td>
<td></td>
</tr>
<tr>
<td>Minneap. ($j = 2$)</td>
<td></td>
<td>Group 4 = 10 flights</td>
<td></td>
</tr>
<tr>
<td>Denver ($j = 1$)</td>
<td></td>
<td></td>
<td>Group 5 = 10 flights</td>
</tr>
<tr>
<td>San Fran. ($j = 2$)</td>
<td></td>
<td></td>
<td>Group 6 = 10 flights</td>
</tr>
</tbody>
</table>

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

Model includes Airline + City (nested under Airline):

$Y_{ijk} = \mu.. + \alpha_k + \beta_{j(k)} + \varepsilon$

where $\mu.. =$ overall mean

$\alpha_k =$ how airline $k$ differs from the overall average $= \mu_k - \mu..$

$\beta_{j(k)} =$ how city $j$ (within airline) differs from Airline $k$ overall average $= \mu_{jk} - \mu_k$

Examples:

$\alpha_1 =$ How much American’s average differs from the overall average

$\beta_{1(2)} =$ How much the average for Atlanta (city 1 within airline 2) differs from Delta’s combined average
Test for Factor A: Do all airlines have same average on time performance? (All $\alpha_k = 0$)

Test for Factor B: Does the on time performance within airlines change based on what city the flight is from? (All $\beta_{j(k)} = 0$)

NOTE: Can’t have AB 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. Subjects can be treated as blocks in this case!

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

<table>
<thead>
<tr>
<th>Type is determined by what groups are used for each level of the Factor. Same or different?</th>
<th>Factor A Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td><em>Between subjects</em></td>
<td>Group 1</td>
</tr>
<tr>
<td>No repeated measures; different groups used</td>
<td></td>
</tr>
<tr>
<td><em>Within subjects</em></td>
<td>Group 1</td>
</tr>
<tr>
<td>Repeated measures; same group used</td>
<td></td>
</tr>
</tbody>
</table>
Recall Randomized Block Design (RBD) from last lecture

It was a special case of a Repeated Measures design:

- “Blocks” are individuals
  - Example: the 5 students, Student= Block

- All factor levels measured for each block
  - Example: Each level of the factor (exam type) measured for each student

- Therefore, using the distinction of “Between subjects” and “Within subjects”
  - The *same* group of students is used for each level of the “Exam” Factor
  - So Exam is a “Within subjects” or repeated measures Factor

- So RBD is a special case of repeated measures design

- Blocks are *crossed* with the Factor because “Person #1” “Person #2” etc. means the *same person* at all levels of the Factor. Therefore, we can treat “Person” as a Factor (i.e. a Block)
Another Randomized Block Design Example

Does listening to Mozart increase IQ, at least temporarily?

- 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.
  - Factor A is crossed with Student.
  - Factor A is a “Within subjects” factor (same group for all levels)
  - Model=Treatment + Blocks (can’t test interaction; only one response per cell)

Y = results of an IQ test

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

<table>
<thead>
<tr>
<th>Level:</th>
<th>Factor A = Treatment = Listening Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Units tested</td>
<td>Mozart</td>
</tr>
<tr>
<td>Group 1</td>
<td>119</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Units tested</th>
<th>Mozart</th>
<th>Relaxation tape</th>
<th>Silence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 3</td>
<td></td>
</tr>
</tbody>
</table>

In that case, Model has one factor only = Listening condition (no blocks)
ANOVA Table for the actual experiment (randomized block):

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>(Not used)</td>
<td>35</td>
<td>(Not used)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>1752</td>
<td>2</td>
<td>876</td>
<td>7.1</td>
<td>.002</td>
</tr>
<tr>
<td>Error</td>
<td>8610</td>
<td>70</td>
<td>123</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reject null hypothesis of equal average IQ after each condition (in population, if they were to listen). Multiple comparisons 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. Remember that an AB (for example) interaction cannot be included if:

  o There is only one observation in each combination of factors A and B and/or

  o The levels of A are nested under levels of B, or vice versa