Some Designs

**Nested Effects:** When we have

- arbitrarily numbered units within each treatment group
- which are subsampled then
  - the arbitrarily numbered units are a “nested” effect.

Tanks of fish: 3 treatments, 4 tanks per treatment, 3 fish per tank

<table>
<thead>
<tr>
<th>Tank 1</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tank 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tank 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tank 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The tanks are arbitrarily numbered.
- There is no particular connection between tank 1 of treat 1 and tank 1 of treat 2.
- We could just as well renumber the tanks 3, 2, 1, 4 in treatment 2.

In this case we says that “tanks are nested within treatments”

- With notation Tank(treatment)
- Within each treatment we have
  - arbitrarily numbered tanks
  - multiple measurements from each tank.
- If we only had one fish per tank,
  - the fish to fish variability would be the Error in the resulting ANOVA table.
  - We would not need to specify any nested effects.

We do **not** have 12 independent fish for each treatment.
- We have 4 independent tanks that were randomized to each treatment.
- Analyzing these data as a single factor ANOVA with 12 replicates per treatment would violate the rules,
  - a "pseudo-replication" mistake.
In contrast to a nested factor, when we have a factorial design, the factors are not arbitrarily numbered.

<table>
<thead>
<tr>
<th>Fertilizer</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 gal)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2 gal)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3 gal)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4 gal)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Water level 1 for fertilizer 1 is the same as water level 1 for fertilizer 2.
- We cannot arbitrarily change the water level designators for an observation.

**Multivariate Measurements:**
- Multiple values for the same person, animal, company, or other unit.
  - Vitamin D, parathyroid hormone and bone density for same person

**Repeated Measures Designs:** (Special case of multivariate)
- Repeatedly taking the same measurement on a single person, animal, or unit.
  - Lung resistance at increasing levels of methacholine.

**Longitudinal:** (Special case of repeated measures)
- Repeated measures over time.
  - Returns for a given stock fund monthly for 5 years.

There are usually better ways to handle repeated measures designs than in the text.
- If you have a repeated measures design, seek out more detailed help from books, websites, or people.
  - *SAS for Mixed Models*
  - *SAS for Linear Models*
• Random Effects
  o The levels of the factor are chosen (ideally randomly) to represent a population of possible units.
• Fixed Effects
  o Our scope of inference is just to these particular levels of the factor

Whether a factor is fixed or random primarily affects tests for other effects.

Tanks of fish: A random effect
• The tanks are chosen to represent investigate how the treatments affect fish in tanks in general, not just the particular tanks used in the experiment.
• Tank is also a nested effect, tank within treat, tank(treat)
• Almost always nested effects are random.
  o In any nested example I give you, the nested effect will be random.

Variances:
• Each ANOVA model in a SAS model statement automatically has variance corresponding to random error as part of the model.
• In specifying the model, we only need to include random effects when there is more than one random variance component to consider.
  o Random tank(treat);
  o Variance from fish to fish in same tank is the Error variance.

Blood pressures:
• 2 people.
• Each person has 10 BP values, 5 with and 5 without the medication.
• Person is random if
  o We are interested in how the med affects BP in the larger population of persons
  o On average does medication change BP in the population?
• Person is fixed if
  o We are interested in how the med affects BP in just these 2 persons
  o On average does medication change BP in these 2 persons?
• Likely person is a random effect here.
Effects of a medication in Men and Women.
- Sex: Female, Male. Fixed effect.
  - These are not sexes chosen to represent a larger population of possible sexes.

Effects of temperature on line strength.
- Fishing Line: Stren, XL, XT
- This is a fixed effect if we the scope of inferences is just about whether temperature affects these particular types of line.

Paints P₁, P₂, P₃, P₄ [Example 15.1 p. 858]
- If we are interested in these particular paints, the factor paint is a fixed effect.
- If these paints were chosen from a population of possible paints to see how much variability we have from paint to paint in this wider population of paints, then paint would be a random effect.
- Likely in Example 15.5 paint is a fixed effect.

Weed control [Example 9.3 p. 434]
- We have 2 biological and 2 chemical weed controlling agents.
- If we are drawing conclusions about these particular agents, we have a fixed effects situation.
- If these agents are chosen to represent biological agents in general and chemical agents in general, then we would be in a random effects situation.
- Comparing biological and chemical agents:
  - Fixed Effects: Are these 2 particular biological agents on average different from these particular 2 chemical agents?
  - Random Effect: Are biological agents in general on average different from chemical agents?
**Split Plot Designs:** Different factors are randomized to different subunits.

**Watering and plant variety:**
- We take a field and divide into 6 main plots
- We randomly assign 2 main plots to each watering level: Low, medium, high
- In each main plot we divide the main plot into 5 subplots.
  - In each subplot we randomly assign subplots to each of 5 tomato varieties.
- Watering level is the main plot effect.
- Variety is the subplot effect. Varieties are likely fixed effects.
- Plots and subplots random effects.
  - Our interest is in how the treatment affect plants in locations in general
    - Not just in these particular plots

**Beating and cooking temperatures:** We make 12 batches of cookies.
- We randomly assign 4 batches to each level of beating: Low, medium, high.
- We subdivide the batches into 3 trays.
  - Trays are randomly assigned to a temperature: 300°, 310°, 320°.
- The batches are “main plots”. Beating is a main plot effect.
- The trays are “sub-plots”. Temperature is a subplot effect.
- Beating and temperature are fixed effects.
- Batches and trays are random effects.
  - We want to draw conclusions about beating and temperature in general
    - Not just for these particular batches or trays.

**Randomized block design:** Randomize units within each block.
- For a "randomized block design" there is one replicate of each treatment in each block.
- A "generalized random block design" has more than one replicate per treatment in a given block.
- An "incomplete randomized block design" has only some of the treatments replicated in each block, for example if days are blocks and we can't do all treatments in a single day.
- Potentially better insurance against bias
- Make each block as uniform as possible, small variance.
- Variability from block to block is removed from error variance.
  - Estimates more precise, confidence intervals narrower
  - Tests have more power to detect alternatives.