Last Time: Statistical Model
- Another way to summarize our results is in terms of treatment effects where instead of comparing groups to each other, we compare them to an overall mean (deviations)
  - $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ where our typical model (e.g., ANOVA, t-tests) assume the random errors follow a normal distribution with mean 0 and variance $\sigma^2$.
    - These assumptions about the random errors is a slightly different model from when we assume the randomness all comes from random assignment
    - To estimate $\mu$, we use the overall mean response $\bar{y}$ (average of group means)
    - To estimate $\hat{\alpha}_i = \bar{y}_i - \bar{y}$ (Fit Model > Estimates > Expanded Estimates)
    - So we predict a value in group $i$ by $\bar{y}_i$, so these are the “fitted values”
- Partitioning the variability in the response variable
  - SSTotal = SSGroups + SSErrror = $\sum \alpha_i^2 + \sum \epsilon_i^2$
  - R$^2$ = SSG/SST x 100% = percentage of variability in $y$ explained by the EV
- In general, we can use SSE as a way to compare the performance of a model

Example: Researchers wanted to compare the effects of caffeine with theobromine, an alkaloid found in chocolate with similar molecular structure as caffeine. To measure the effects of these two different chemicals, researchers trained subjects to tap their fingers in such a way that the rate could be measured. After learning/practicing, researchers had participants either take a caffeine pill (200mg), a theobromine pill (200mg), or a placebo, wait two hours, and then measured the rate of finger tapping (taps per minute).

(a) Do you think the ANOVA F-test will be statistically significant? No? Small sample sizes, differences aren't large compared to means & compared to SDs.

(b) Complete the ANOVA table below

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>8</td>
<td>872</td>
<td>109</td>
<td>22.36</td>
<td>.0005</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>581</td>
<td>65</td>
<td>45.67</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>623</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(c) How much of the variability in finger tapping rates is explained by the type of chemical?

\[
R^2 = \frac{872}{468.3} \times 100\% = 18.36\%
\]

(d) What are some reasons we are failing to find a statistically significant difference here? What are some strategies for improving our study?

- Larger sample sizes
- Want to reduce the variability in the data
- Account for that person to person variability

(e) How could we design a study that would take into account the fact that some people are faster finger tapers than others? Where is the randomness in your design?

- Have each person use all methods
- Liked a matched pairs design but now
- We 3 treatments

**Definition:** A randomized block study design creates blocks of experimental units that are similar to each other, assigns the treatments within each block, and then analyzes the data accounting for block to block variations. When there are only two groups being compared, a block study design is called a matched pairs design. The term block comes from the first ‘block designs’ which were agricultural experiments in large fields where separate parts of the field were called ‘blocks.’

Below are the same results, but now grouped for each of the four subjects

(f) Was our suspicion of “significant” subject to subject variation accurate? How much of the variability in finger tapping is explained by which subject is doing the tapping?

\[
P^2 = \frac{5478}{6682} = 0.827
\]

Can run an ANOVA on blocks

\[F = 12.13 \quad p = 0.0024\]

Strong evidence of genuine differences in tapping rates across people

(g) How do you suggest now analyzing the data?

Focus on variability among the 3 treatments within each person
The table below shows the results by subject:

<table>
<thead>
<tr>
<th>Subject</th>
<th>Placebo</th>
<th>Caffeine</th>
<th>Theobromine</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>446</td>
<td>453</td>
<td>472</td>
<td>457</td>
</tr>
<tr>
<td>2</td>
<td>451</td>
<td>466</td>
<td>460</td>
<td>459</td>
</tr>
<tr>
<td>3</td>
<td>455</td>
<td>474</td>
<td>481</td>
<td>470</td>
</tr>
<tr>
<td>4</td>
<td>496</td>
<td>523</td>
<td>511</td>
<td>510</td>
</tr>
<tr>
<td>Means</td>
<td>462</td>
<td>479</td>
<td>481</td>
<td>474</td>
</tr>
</tbody>
</table>

(h) Within each subject, does one of the chemicals seem to correspond to faster or slower tapping? What is the average placebo treatment effect?

(i) If we assume “additive effects,” what can we use for a model equation to predict a particular tapping rate outcome?

\[ y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \]

(j) Use the above data to estimate each of the block effects.

(k) So one way to account for the block effects is to subtract them off! In other words, our “block adjusted” data values will be the original data values minus the block effect.

(l) Carry out a one-way ANOVA on the block adjusted data. Compare the ANOVA table to what we found in (b).
(m) What has changed and not changed in our ANOVA table and why??

- SSG & MSG stay the same because this doesn't change $g_1, g_2, g_3, g_4, g_5$
- SSE & MSE because the within group variability is a lot smaller when the person-to-person variability

(n) Where did all that extra unexplained variation go?

$\frac{(0.17)^2 + (-1.59 + 6.49)^2 + 3.6^2}{6} \times 3 = 6478$

$6482 - 1826 = 6478$

To better keep track of this, we will set up a two-way ANOVA table

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>872</td>
<td>436</td>
<td>7.18</td>
<td>.0210</td>
</tr>
<tr>
<td>Blocks</td>
<td>10</td>
<td>5478</td>
<td>5478</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>532</td>
<td>88.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>6682</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(o) Explain how the F statistic for treatments is calculated, how do you interpret this value, and how the p-value is determined. What conclusion can we draw?

$F \frac{MS_{Treatment}}{MS_{Error}}$

$0.0210$ small at least one treatment differs after adjusting for subject

Technology Instructions:
- JMP: Use Analyze > Fit Model and put both the block variable and the treatment variable in the Construct Model Effects box
- R: summary(aov(taps~drug+person))

Validity Conditions

Validity conditions for ANOVA with a blocking variable follow in the same spirit as the validity conditions for ANOVA without a blocking variable, essentially they are the same conditions as we saw before, applied to the adjusted data. Below are some common ways to assess these conditions:

- You can plot residuals vs. treatment or residuals vs. predicted values (aka fits, aka $\hat{y}$).
  Hope to see constant spread of across all the fitted values, vs. some kind of fanning/changes in variability with the predicted values.
- Normal probability plot of residuals
  Hope for straight line / large p-value

Multiple Comparisons can also be run for the treatment variable and/or the blocking variable.
You can also get confidence intervals for the block effects and/or for the treatment effects.