Unbalanced Factorial Designs

Or when does a mouse weigh more than an elephant?
Bad stuff can happen when things are unbalanced, if you’re not careful!
Unbalanced Designs

- three (common) ways to analyze them
- all of them involve differences of opinion over how the sums of squares should be calculated (the rest of the ANOVA is the same)

- Type I - sequential
- Type II
- Type III - simultaneous

- this is a very contentious issue
Imagine that the effects are hogs feeding at a trough. If there are two factors, A and B, then there are three hogs to be fed: the A main effect hog, the B main effect hog, and the AxB interaction hog.

Type I - the hogs feed sequentially

Type III - the hogs feed simultaneously

Type II - the main effects hogs feed simultaneously while the interaction hog waits its turn
What does it mean?

When the design is balanced, the independent variables are truly independent of each other. There is no confounding. This is why true experiments with random assignment are usually done using balanced designs. (Anything else would be silly.) In quasi-experiments using in tact groups, keeping the design balanced may not be possible and may even be unwise (because it would be a violation of random sampling). When the design is unbalanced, it is no longer the case that the IVs are independent of each other. In unbalanced designs, the IVs are “confounded” with each other. (I have put “confounded” in quotes because we will see later it ain’t necessarily so.)

In a Type I analysis, the first effect entered (A) grabs all of the total variability it can get. No confounds are removed, and therefore the A main effect is not controlled for confounding with the B main effect or with the interaction. (Another way to say this is A ignoring B, i.e., we are looking at the A effect while pretending that factor B does not exist. It’s almost like we’re doing a single-factor ANOVA on A.) The second effect entered (B) then gets to grab variability from what’s left. Any confounds between A and B are removed from the B main effect. Therefore, B is controlled for A. (Another way to say this is B with A removed; that is, any “confound” with A has been statistically removed before factor B is analyzed.)

In a Type III analysis, everything is controlled for everything else. All confounds are removed. (Another way to say this is every factor is treated as if entered last.)

You will notice, when comparing Type I and Type III analyses, that the Residuals line (error) is always the same, and the highest order interaction (last effect in the summary table) is also the same.
What Does This Accomplish?

(Think about these possibilities.)

A and B have only direct effects on the DV.  
A (and B) may have indirect effects on the DV.

A acts through B.
Both A and B may act through the interaction, although this is hard to conceptualize, and for the most part we’ll ignore it.

Example: I hit you with a stick (A). You get angry (B) and hit me back (DV).
Why Are Variables Related?

Question: When is a “confound” not a confound?
Answer: When it’s an effect.

total effect = direct effect + indirect (mediated) effects

What are the effects?
Hold on a second, Jack! I was taught that correlation does not imply causation.

The correct statement is: “Correlation” alone does not necessarily imply causation.
So...?

- True (randomized) experiments are arranged in such a way that there are only direct effects.
- Any relationship between A and B is a confound and should be removed.
- Type III sums of squares accomplish this. (Simultaneous analysis shows only direct effects. Indirect effects due to confounds are removed.)
- With intact groups ("quasi-experimental" designs), legitimate indirect effects may be present, and they may be important.
- We may want to see total effects.
- Type I sums of squares accomplish this. (Sequential analysis shows total effects, provided we have our causal model correct.)
The Moral (but Hardly the End) of the Story

- When a true experiment is mildly unbalanced by accident and at random, that’s what the Type III analysis was created for.
- If the unbalanced condition is more than mild, Type III can cost you a lot of power! (Think about Type II in that case.)
- With quasi-experimental designs, you MAY want to consider Type I (if you want to see total rather than just direct effects).
- What about Type II? I would use them for quasi-experiments when you’re not sure which factor should be A (entered first), or when there is no good reason to enter one of the factors first. I would also use them for true experiments which are more than mildly unbalanced.
Summary

<table>
<thead>
<tr>
<th>SAS</th>
<th>Overall &amp; Spiegel</th>
<th>Herr &amp; Gaebelein</th>
<th>Others (e.g., Macnaughton)</th>
<th>Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>Method III</td>
<td>HRC, HCR</td>
<td>hierarchical or sequential</td>
<td>A, B</td>
</tr>
<tr>
<td>Type II</td>
<td>Method II</td>
<td>EAD</td>
<td>higher terms omitted (HTO)</td>
<td>A</td>
</tr>
<tr>
<td>Type III</td>
<td>Method I</td>
<td>STP</td>
<td>higher terms included (HTI), A</td>
<td>B &amp; AxB, B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>partial, marginal</td>
<td></td>
</tr>
</tbody>
</table>

Note: B|A means B given, or controlled for, or removing (the “confound” with) A

Go here for the details:
http://ww2.coastal.edu/kingw/statistics/R-tutorials/unbalanced.html
Example

• What is the relationship between:
  • hostility (categorical: high and low)
  • anger (categorical: high and low)
  • aggression (numeric)
I propose the following...

Which of these arrows would be shown by Type I sums of squares? Which of these arrows would be shown by Type III sums of squares?
The Salary Problem
(from Maxwell & Delaney)

We are interested in finding out if there is salary discrimination by gender at a certain company. We take a random sample of new employees and examine their starting salaries. We also record whether or not the person has a college degree. Here are the data (salaries are in thousands of dollars—it's old data!).

```
salary  gender  education
S1      24     female  degree   |
S2      26     female  degree   |
S3      25     female  degree   |
S4      24     female  degree   |
S5      27     female  degree   |
S6      24     female  degree   |
S7      27     female  degree   |
S8      23     female  degree   |
S9      15     female  no.degree|
S10     17     female  no.degree|
S11     20     female  no.degree|
S12     16     female  no.degree|
S13     25     male    degree   |
S14     29     male    degree   |
S15     27     male    degree   |
S16     19     male    no.degree|
S17     18     male    no.degree|
S18     21     male    no.degree|
S19     20     male    no.degree|
S20     21     male    no.degree|
S21     22     male    no.degree|
S22     19     male    no.degree|
```

sum = 268  mean = 22.333  var = 18.242  n = 12

sum = 221  mean = 22.100  var = 13.656  n = 10

The means are almost the same. Issue settled?
The Salary Problem
(continued)

Weighted vs. unweighted means
- as an aside, let’s look at Tom Prin’s firemen data again.

\[
\begin{array}{ccc}
A & B & C \\
8.961538 & 10.212121 & 11.062500 \\
\end{array}
\]

\[
\begin{array}{ccc}
A & B & C \\
26 & 33 & 16 \\
\end{array}
\]

Question: what is the grand mean?

\[
\frac{(8.961538 \times 26 + 10.212121 \times 33 + 11.0625 \times 16)}{(26 + 33 + 16)} \approx 9.96
\]

Which one is correct?
The Salary Problem
(continued)

We have a factorial design, so let’s look at the full factorial structure of the data.

<table>
<thead>
<tr>
<th></th>
<th>female</th>
<th>male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24, 26, 25, 24,</td>
<td>25, 29, 27</td>
</tr>
<tr>
<td>27, 24, 27, 23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>degree</td>
<td>n = 8</td>
<td>n = 3</td>
</tr>
<tr>
<td>sum</td>
<td>200 (25.0)</td>
<td>81 (27.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>15, 17, 20, 16</td>
<td>19, 18, 21, 20,</td>
</tr>
<tr>
<td></td>
<td>21, 22, 19</td>
<td></td>
</tr>
<tr>
<td>degree</td>
<td>n = 4</td>
<td>n = 7</td>
</tr>
<tr>
<td>sum</td>
<td>68 (17.0)</td>
<td>140 (20.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 12</td>
<td>n = 10</td>
</tr>
<tr>
<td>sum</td>
<td>268</td>
<td>221</td>
</tr>
<tr>
<td>mean</td>
<td>22.333</td>
<td>22.100</td>
</tr>
<tr>
<td>mean</td>
<td>21.000</td>
<td>23.500</td>
</tr>
</tbody>
</table>

--- weighted marginal means
--- unweighted marginal means

If the main effects are in the marginal means, what main effects are we seeing?
Next question: Is there an interaction? Draw a profile plot. What effects do we see in the profile plot? Are those effects seen in the weighted or unweighted marginal means?

Which kind of means control for educational level?
You control A for B by looking at A within levels of B.
You control gender for educational level by looking at gender within levels of educational level, or by looking at gender with education held constant.
The Salary Problem  
(continued)

salary.csv

The two ANOVAs in R:
> summary(aov(salary ~ gender * education))

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>1</td>
<td>0.30</td>
<td>0.30</td>
<td>0.107</td>
</tr>
<tr>
<td>education</td>
<td>1</td>
<td>272.39</td>
<td>272.39</td>
<td>98.061</td>
</tr>
<tr>
<td>gender:education</td>
<td>1</td>
<td>1.17</td>
<td>1.17</td>
<td>0.423</td>
</tr>
<tr>
<td>Residuals</td>
<td>18</td>
<td>50.00</td>
<td>2.78</td>
<td></td>
</tr>
</tbody>
</table>

> summary(aov(salary ~ education * gender))

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>education</td>
<td>1</td>
<td>242.23</td>
<td>242.23</td>
<td>87.202</td>
</tr>
<tr>
<td>gender</td>
<td>1</td>
<td>30.46</td>
<td>30.46</td>
<td>10.966</td>
</tr>
<tr>
<td>education:gender</td>
<td>1</td>
<td>1.17</td>
<td>1.17</td>
<td>0.423</td>
</tr>
<tr>
<td>Residuals</td>
<td>18</td>
<td>50.00</td>
<td>2.78</td>
<td></td>
</tr>
</tbody>
</table>

R does Type I ANOVA by default. In Type I ANOVA, the order in which the factors are entered makes a difference.

What are these effects?
The Salary Problem
(continued)

> source("http://ww2.coastal.edu/kingw/psyc480/functions/aovIII.R")
> aovIII(salary ~ gender * education, data=sal)  # Type III: order does not matter
Single term deletions

Model:
salary ~ gender * education

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>RSS</th>
<th>AIC</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;none&gt;</td>
<td>50.000</td>
<td>26.062</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>1</td>
<td>29.371</td>
<td>79.371</td>
<td>34.228</td>
<td>10.5734</td>
</tr>
<tr>
<td>education</td>
<td>1</td>
<td>264.336</td>
<td>314.336</td>
<td>64.507</td>
<td>95.1608</td>
</tr>
<tr>
<td>gender:education</td>
<td>1</td>
<td>1.175</td>
<td>51.175</td>
<td>24.573</td>
<td>0.4229</td>
</tr>
</tbody>
</table>

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Notice that the interaction is evaluated the same. Further analysis and explanation of this problem will be forthcoming.)