Chapter 23 - Two Factor Factorial
Designs with Unequal Sample Sizes

Every combination of a level of factor A with a level of factor B is observed. But the number of observations at these combinations differ. There are:

\[ n_{ij} = \text{(# obs. at level } i \text{ of factor } A \text{ and level } j \text{ of factor } B), \]

\[ n_{i.} = \sum_{j=1}^{b} n_{ij}, \quad n_{.j} = \sum_{i=1}^{a} n_{ij}, \quad \text{and } n_{.} = \frac{a \times b}{n_{ij}} \]

What is the problem?
and the sums of the cross-product terms are equal to zero. When the sample sizes are not the same, the cross-product terms do not sum to zero. Thus the total sum of squares does not separate into easily interpretable pieces. Therefore, the ANOVA process described earlier is no longer valid.

What can we do?
Main Effects of Factor B

\[ \begin{align*}
1 & \quad I_{B1} = \begin{cases} 1 & \text{if } B \text{ at level } 1 \\ 0 & \text{otherwise} \end{cases} \\
\vdots & \quad \vdots \\
b-1 & \quad I_{Bb-1} = \begin{cases} 1 & \text{if } B \text{ at level } b-1 \\ 0 & \text{otherwise} \end{cases} \\
b & \quad \text{all the above are zero}
\end{align*} \]

Main Effects Model

Likewise to test for a Main Effect of fact B, test

\[ H_0: \beta_a = \beta_{a+1} = \ldots = \beta_{a+b-2} = 0 \]

vs \( H_a: \text{not } H_0 \)
Interaction Model

If this is significant we stop and compare all ab cell means. If it is not significant, then test both main effects.

**Main Effects of A**

\[ H_0: \beta_1 = \beta_2 = \cdots = \beta_{a-1} = 0 \]  \hspace{1cm} \text{(a-1) terms}

\[ H_a: \text{not } H_0 \]

and

**Main Effect of B**

\[ H_0: \beta_a = \beta_{a+1} = \cdots = \beta_{a+b-2} = 0 \]  \hspace{1cm} \text{(b-1) terms}

\[ H_a: \text{not } H_0 \]
The follow-up procedures are similar to what was described in chapter 19, except generally the sample sizes differ, so Tukey's method of multiple comparison is not an option.

**Example:** (SAS Notes)

In this example, the overall test

Ho: no main effects and no interaction

Ha: not Ho

is significant

\[ F = 11.97 \quad \text{p-value} < .0001 \]

The interaction test
So by way of follow-up we examine the 9 cell means below, with a Bonferroni multiple comparison. Why? The simultaneous 95% C.I.'s show us which pairs differ.
What to do when a cell has no observations.

Sometimes patients leave a study, etc and a (a few) combinations of levels are not observed. How can we analyze the data and draw inferences about the factors in these cases.
Two possible ways to approach the analysis.

(A) Analyze a partial design

\[ \text{Diagram of a partial design} \]

Omit parts with partial information

(B) Assume structure among the means e.g. a Main Effects Model.
only want to take this approach if previous experience or a partial analysis led us to believe that the Main Effects structure is valid.