Overview

- Now have two factors (A and B)
- Suppose each factor has two levels
- Could analyze as one factor with 4 levels
  - Trt 1: A high, B high
  - Trt 2: A high, B low
  - Trt 3: A low, B high
  - Trt 4: A low, B low
- Use contrasts to test for main effects an interaction

\[
A \text{ main effect} = \frac{\text{Trt1 + Trt2}}{2} - \frac{\text{Trt3 + Trt4}}{2}
\]

Example

An experiment is conducted to study the effect of hormones injected into test rats. There are two distinct hormones (A, B) each with two distinct levels. For purposes here, we will consider this to be four different treatments labeled \{A,a,B,b\}. Each treatment is applied to six rats with the response being the amount of glycogen (in mg) in the liver.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 106 101</td>
<td>120 86 132 97</td>
</tr>
<tr>
<td>a 51 98</td>
<td>85 50 111 72</td>
</tr>
<tr>
<td>B 103 84</td>
<td>100 83 110 91</td>
</tr>
<tr>
<td>b 50 66</td>
<td>61 72 85 60</td>
</tr>
</tbody>
</table>
Example

Three contrasts are of interest. They are:

**Comparison**  |  A  | a  | B  | b  
---|---|---|---|---
Hormone A vs Hormone B  | 1  | 1  | -1 | -1  
Low level vs High level  | 1  | -1 | 1  | -1  
Equivalence of level effect  | 1  | -1 | -1 | 1  

Can we reanalyze the experiment in such a way that these sum of squares are already separated?

Two-way Factorial

- Break up the treatments into the two factors
- Example also known as a $2^2$ factorial
- Investigates all combos of factor levels
- Single “replicate” involves $ab$ trials

Two-way Factorial Layout

- Often presented as table or plot

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Level</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data for Two Factor ANOVA

- $Y$ is the response variable
- Factor $A$ has levels $i = 1, 2, ..., a$
- Factor $B$ has levels $j = 1, 2, ..., b$
- $Y_{ijk}$ is the $k^{th}$ observation from cell $(i, j)$
- Chapter 19 assumes $n_{ij} = n$
- Chapter 20 assumes $n_{ij} = 1$
- Chapter 23 allows $n_{ij}$ to vary
Example Page 833

- Castle Bakery supplies wrapped Italian bread to a large number of supermarkets
- Bakery interested in the set up of their store display
  - Height of display shelf (top, middle, bottom)
  - Width of display shelf (regular, wide)
- Twelve stores equal in sales volume were selected
- Randomly assigned equally to each of 6 combinations
- $Y$ is the sales of the bread
  - $i = 1, 2, 3$ and $j = 1, 2$
  - $n_{ij} = n = 2$

SAS Commands

data a1; infile 'u:\www\datasets525\CH19TA07.txt';
  input sales height width;
proc print;
run;
data a1; set a1;
  if height eq 1 and width eq 1 then hw='1_BR';
  if height eq 1 and width eq 2 then hw='2_BW';
  if height eq 2 and width eq 1 then hw='3_MR';
  if height eq 2 and width eq 2 then hw='4_MW';
  if height eq 3 and width eq 1 then hw='5_TR';
  if height eq 3 and width eq 2 then hw='6_TW';
symbol1 v=circle i=none;
proc gplot data=a1;
  plot sales*hw/frame;

Sales Data

<table>
<thead>
<tr>
<th>Obs</th>
<th>sales</th>
<th>height</th>
<th>width</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>43</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>46</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>62</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>68</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>67</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>71</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>41</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>39</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>42</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>46</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
The Model

• All observations assumed independent
• All observations normally distributed with
  – a mean that may depend on levels of factors A and B
  – constant variance
• Often presented in terms of cell means or factor effects

The Cell Means Model

• Expressed mathematically
  \[ Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \]
  where \( \mu_{ij} \) is the theoretical mean or expected value of all observations in cell \((i, j)\)
• The \( \varepsilon_{ijk} \) are iid \( N(0, \sigma^2) \) which implies the \( Y_{ijk} \) are independent \( N(\mu_{ij}, \sigma^2) \)
• Parameters
  – \( \{\mu_{ij}\}, \; i = 1, 2, \ldots, a, \; j = 1, 2, \ldots, b \)
  – \( \sigma^2 \)

Estimates / Inference

• Estimate \( \mu_{ij} \) by the sample mean of the observations in cell \((i, j)\)
  \[ \hat{\mu}_{ij} = \bar{Y}_{ij}. \]
• Can also estimate variance using observations in cell \((i, j)\)
  \[ s_{ij}^2 = \frac{\sum (Y_{ijk} - \bar{Y}_{ij.})^2}{n - 1} \]
• These \( s_{ij}^2 \) are combined for single estimate of \( \sigma^2 \)

ANOVA Table : \( n_{ij} = n \)

• Similar ANOVA table construction (\( \bar{Y}_{ij.} \) is fitted value)
  Source of Variation | df | SS
  Model \hspace{1cm} \hspace{1cm} ab - 1 | \hspace{1cm} n \sum \sum (\bar{Y}_{ij.} - \bar{Y}_{...})^2
  Error \hspace{1cm} ab(n - 1) | \hspace{1cm} \sum \sum \sum (Y_{ijk} - \bar{Y}_{ij.})^2
  Total \hspace{1cm} abn - 1 | \hspace{1cm} \sum \sum \sum (Y_{ijk} - \bar{Y}_{...})^2

  \[ \bar{Y}_{...} = \sum \sum \sum Y_{ijk}/abn \]
  \[ \bar{Y}_{ij.} = \sum Y_{ij.}/n \]
• Can further break down into Factor A, Factor B and interaction effects using contrasts
Factor Effects Model

- Breaks down cell means
  \[ \mu = \sum_i \sum_j \mu_{ij} / (ab) \]
  \[ \mu_i = \sum_j \mu_{ij} / b \text{ and } \mu_j = \sum_i \mu_{ij} / a \]
  \[ \alpha_i = \mu_i - \mu \text{ and } \beta_j = \mu_j - \mu \]
  \[ (\alpha\beta)_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j) \]

- Interaction effect is the difference between the cell mean and the additive or main effects model. Explains variation not explained by main effects.

Factor Effects Estimates

- Constraints on previous page result in
  \[ \hat{\mu} = \overline{Y} \]
  \[ \hat{\alpha}_i = \overline{Y}_{i..} - \overline{Y} \]
  \[ \hat{\beta}_j = \overline{Y}_{.j} - \overline{Y} \]
  \[ (\hat{\alpha\beta})_{ij} = \overline{Y}_{ij} - \overline{Y}_{i..} - \overline{Y}_{.j} + \overline{Y} \]

- The predicted value and residual are
  \[ \hat{Y}_{ijk} = \overline{Y}_{ij} \]
  \[ e_{ijk} = Y_{ijk} - \overline{Y}_{ij} \]

Questions about our Example

- Does the height of the display affect sales?
  - If yes, will need to do pairwise comparisons

- Does the width of the display affect sales?
  - If yes, will need to do pairwise comparisons

- Does the effect of height depend on the width?
- Does the effect of width depend on the height?
  - If yes to either of these last two, we have an interaction
Partitioning the Sum of Squares

\[ Y_{ijk} - \bar{Y} = (\bar{Y}_{i.} - \bar{Y}) + (\bar{Y}_{.j.} - \bar{Y}) + (Y_{ij.} - \bar{Y}_{i.} - \bar{Y}_{.j.} + \bar{Y}) + (Y_{ijk} - \bar{Y}_{ij.}) \]

- Consider \( \sum \sum \sum (Y_{ijk} - \bar{Y}) \)²
- Right hand side simplifies to
  \[ \sum_i (\bar{Y}_{i.} - \bar{Y})^2 + \sum_j (\bar{Y}_{.j.} - \bar{Y})^2 + \sum_{ij} (Y_{ij.} - \bar{Y}_{i.} - \bar{Y}_{.j.} + \bar{Y})^2 + \sum_{ijk} (Y_{ijk} - \bar{Y}_{ij.})^2 \]

Hypothesis Testing

- Can show: Fixed Case
  \[ \text{E(MSE)} = \sigma^2 \]
  \[ \text{E(MSA)} = \sigma^2 + bn \sum \alpha_i^2 / (a-1) \]
  \[ \text{E(MSB)} = \sigma^2 + an \sum \beta_j^2 / (b-1) \]
  \[ \text{E(MSAB)} = \sigma^2 + n \sum (\alpha \beta)_{ij}^2 / (a-1)(b-1) \]
- Use F-test to test for A, B, and AB effects
  \[ F^* = \frac{\text{SSA}/(a-1)}{\text{SSE}/(ab(n-1))} \]
  \[ F^* = \frac{\text{SSB}/(b-1)}{\text{SSE}/(ab(n-1))} \]
  \[ F^* = \frac{\text{SSAB}/(a-1)(b-1)}{\text{SSE}/(ab(n-1))} \]

SAS Commands

\begin{verbatim}
proc glm data=a1;
  class height width;
  model sales=height width height*width;
  means height width height*width;
proc means data=a1;
  var sales; by height width;
  output out=a2 mean=avsales;
symbol1 v=square i=join c=black;
symbol2 v=diamond i=join c=black;
proc gplot data=a2;
  plot avsales*height=width/frame;
run;
\end{verbatim}
The GLM Procedure

Class Level Information
Class Levels Values
height 3 1 2 3
width 2 1 2

Number of observations 12

Sum of
Source DF Squares Mean Square F Value Pr > F
Model 5 1580.000000 316.000000 30.58 0.0003
Error 6 62.000000 10.333333
Corrected Total 11 1642.000000

R-Square Coeff Var Root MSE sales Mean 0.962241 6.303040 3.214550 51.000000
Source DF Type I SS Mean Square F Value Pr > F
height 2 1544.000000 772.000000 74.71 <.0001
width 1 12.000000 12.000000 1.16 0.3226
height*width 2 24.000000 12.000000 1.16 0.3747

Level of Level of
height width N Mean Std Dev
1 1 2 45.0000000 2.82842712
1 2 2 43.0000000 4.24264069
2 1 2 65.0000000 4.24264069
2 2 2 69.0000000 2.82842712
3 1 2 40.0000000 1.41421356
3 2 2 44.0000000 2.82842712

proc glm data=a1;
class height width;
model sales=height|width;
means height / tukey lines;
run;
```
Results

• There appears to be no interaction between height and width ($P=0.37$) → The effect of width (or height) is the same regardless of height (or width). Because of this, we can focus on the main effects (averages out the other effect).

• The main effect for width is not statistically significant ($P=0.32$) → Width does not affect sales of bread

• The main effect for height is statistically significant ($P < 0.0001$). From the scatterplot and interaction plot, it appears the middle location is better than the top and bottom. Pairwise testing (adjusting for multiple comparisons) can confirm this.

Pooling Insignificant Terms

• Some argue that an insignificant interaction should be dropped from the model (i.e., pooled with error)

• See last GLM call

\[ SSE^* = SSE + SSAB \]
\[ df^*_E = ab(n - 1) + (a - 1)(b - 1) \]

• This increases DF but could inflate $\hat{\sigma}^2$

• Possibly result in a Type II error

• Rule of thumb: Only pool when dfe small (e.g., $< 5$) and P-value of the interaction is large (e.g., $> 0.25$)

Output

Tukey’s Studentized Range (HSD) Test for sales

Error Degrees of Freedom 6
Error Mean Square 10.33333
Critical Value of Studentized Range 4.33902
Minimum Significant Difference 6.974

Mean  N  height
A  67.000  4 2
B  44.000  4 1
B  42.000  4 3

**** POOLING ****

Error Degrees of Freedom 8
Error Mean Square 10.75
Critical Value of Studentized Range 4.04101
Minimum Significant Difference 6.6247

Mean  N  height
A  67.000  4 2
B  44.000  4 1
B  42.000  4 3

Background Reading

• KNNL Chapter 19

• knnl833.sas