Two-factor studies

STAT 525
Chapter 19 and 20

Professor Olga Vitek

December 2, 2010
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 A or B effect

\[
A \text{ effect } = \frac{\text{Trt1 + Trt2}}{2} - \frac{\text{Trt3 + Trt4}}{2}
\]
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</td>
<td>106 101 120 86 132 97</td>
</tr>
<tr>
<td>a</td>
<td>51 98 85 50 111 72</td>
</tr>
<tr>
<td>B</td>
<td>103 84 100 83 110 91</td>
</tr>
<tr>
<td>b</td>
<td>50 66 61 72 85 60</td>
</tr>
</tbody>
</table>

Three contrasts are of interest. They are:

<table>
<thead>
<tr>
<th>Comparison</th>
<th>A</th>
<th>a</th>
<th>B</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hormone A vs Hormone B</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Low level vs High level</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Equivalence of level effect</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

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

• Break up trts into the two factors

• Also known as a $2^2$ factorial

• Investigates all combos of factor levels

• Single “replicate” involves $ab$ trials

• Often presented as table or plot

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

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;

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;
Scatterplot
The Model

• Same basic assumptions as regression

• All observations assumed independent

• All observations normally distributed with
  – a mean that may depend on the levels of factors $A$ and $B$
  – constant variance

• Often presented in terms of cell means or factor effects
The Cell Means Model

• Expressed numerically

\[ Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \]

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

- \( \{\mu_{ij}\}, \ i = 1, 2, .., a, \ j = 1, 2, ..., b \)

- \( \sigma^2 \)

19-9
Estimates / Inference

• Estimate $\mu_{ij}$ by the sample mean of the observations in cell $(i, j)$

$$\hat{\mu}_{ij} = \bar{Y}_{ij}. \quad \text{(19-10)}$$

• For each cell $(i, j)$, also estimate of the variance

$$s^2_{ij} = \sum (Y_{ijk} - \bar{Y}_{ij})^2 / (n - 1) \quad \text{(20-1)}$$

• These $s^2_{ij}$ are combined to estimate $\sigma^2$
**ANOVA Table**: \( n_{ij} = n \)

- Similar ANOVA table construction

- Plug in \( \bar{Y}_{ij} \) as fitted value

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>df</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>( ab - 1 )</td>
<td>( n \sum \sum \sum (\bar{Y}<em>{ij} - \bar{Y}</em>{...})^2 )</td>
</tr>
<tr>
<td>Error</td>
<td>( ab(n - 1) )</td>
<td>( \sum \sum \sum (Y_{ijk} - \bar{Y}_{ij})^2 )</td>
</tr>
<tr>
<td>Total</td>
<td>( abn - 1 )</td>
<td>( \sum \sum \sum (Y_{ijk} - \bar{Y}_{...})^2 )</td>
</tr>
</tbody>
</table>

\[
\bar{Y}_{...} = \frac{\sum \sum \sum Y_{ijk}}{abn}
\]

\[
\bar{Y}_{ij} = \frac{\sum Y_{ij}}{n}
\]

- Can further break down into Factor A, Factor B and interaction effects using contrasts
Factor Effects Model

• Statistical model is

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \]

\[ \begin{align*}
\mu & \text{ - grand mean} \\
\alpha_i & \text{ - } i^{th} \text{ level effect of factor A (ignores B)} \\
\beta_j & \text{ - } j^{th} \text{ level effect of factor B (ignores A)} \\
(\alpha\beta)_{ij} & \text{ - interaction effect of combination } ij
\end{align*} \]

• Over-parameterized model.

• Must include \( a + b + 1 \) model constraints.

\[ \sum_i \alpha_i = 0 \quad \sum_j \beta_j = 0 \quad \sum_i (\alpha\beta)_{ij} = 0 \quad \sum_j (\alpha\beta)_{ij} = 0 \]
Factor Effects Model

- Breaks down cell means

\[ \mu = \frac{\sum_i \sum_j \mu_{ij}}{(ab)} \]

\[ \mu_i. = \frac{\sum_j \mu_{ij}}{b} \text{ and } \mu_j = \frac{\sum_i \mu_{ij}}{a} \]

\[ \alpha_i = \mu_i. - \mu \text{ and } \beta_j = \mu_j - \mu \]

\[ (\alpha\beta)_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j) \]

- 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

• Previous constraints result in

\[
\hat{\mu} = \overline{Y}_\ldots
\]

\[
\hat{\alpha}_i = \overline{Y}_{i.} - \overline{Y}_\ldots
\]

\[
\hat{\beta}_j = \overline{Y}_{.j} - \overline{Y}_\ldots
\]

\[
(\hat{\alpha}\hat{\beta})_{ij} = \overline{Y}_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_\ldots
\]

• The predicted value and residual are

\[
\hat{Y}_{ijk} = \overline{Y}_{ij}.
\]

\[
e_{ijk} = Y_{ijk} - \overline{Y}_{ij}.
\]
Questions

• 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, we have an interaction
Partitioning the Sum of Squares

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

- Consider \( \sum \sum \sum (Y_{ijk} - \bar{Y}_{...})^2 \)
- Right hand side simplifies to

\[
bn \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2 + \\
an \sum_j (\bar{Y}_{.j} - \bar{Y}_{...})^2 + \\
n \sum_i \sum_j (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2 + \\
\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij})^2
\]
Partitioning the Sum of Squares

- Can be written as

\[ \text{SSTO} = \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSE} \]

- Degrees of freedom also broken down

- Under normality, all \( \text{SS}/\sigma^2 \) independent

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>SSA</td>
<td>( a - 1 )</td>
<td>MSA</td>
</tr>
<tr>
<td>Factor B</td>
<td>SSB</td>
<td>( b - 1 )</td>
<td>MSB</td>
</tr>
<tr>
<td>Interaction</td>
<td>SSAB</td>
<td>((a - 1)(b - 1))</td>
<td>MSAB</td>
</tr>
<tr>
<td>Error</td>
<td>SSE</td>
<td>( ab(n - 1) )</td>
<td>MSE</td>
</tr>
<tr>
<td>Total</td>
<td>SSTO</td>
<td>( abn - 1 )</td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis Testing

• Can show: Fixed Case

\[ E(\text{MSE}) = \sigma^2 \]

\[ E(\text{MSA}) = \sigma^2 + bn \sum \alpha_i^2 / (a - 1) \]

\[ E(\text{MSB}) = \sigma^2 + an \sum \beta_j^2 / (b - 1) \]

\[ E(\text{MSAB}) = \sigma^2 + n \sum (\alpha \beta)_{ij}^2 / (a - 1)(b - 1) \]

• Use F-test to test for A, B, and AB effects

\[ F^* = \frac{\text{SSA} / (a - 1)}{\text{SSE} / (ab(n - 1))} \]

\[ F^* = \frac{\text{SSB} / (b - 1)}{\text{SSE} / (ab(n - 1))} \]

\[ F^* = \frac{\text{SSAB} / (a - 1)(b - 1)}{\text{SSE} / (ab(n - 1))} \]
data a1; infile 'u:\.www\datasets525\CH19TA07.txt';
  input sales height width;

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;

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

proc glm data=a1;
  class height width;
  model sales=height width;
  means height / tukey lines;
run;
## Output

The GLM Procedure

### Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>height</td>
<td>3</td>
<td>1 2 3</td>
</tr>
<tr>
<td>width</td>
<td>2</td>
<td>1 2</td>
</tr>
</tbody>
</table>

Number of observations 12

### Sum of Squares

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>1580.000000</td>
<td>316.000000</td>
<td>30.58</td>
<td>0.0003</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>62.000000</td>
<td>10.333333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11</td>
<td>1642.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### R-Square, Coeff Var, Root MSE, sales Mean

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Coeff Var</th>
<th>Root MSE</th>
<th>sales Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.962241</td>
<td>6.303040</td>
<td>3.214550</td>
<td>51.00000</td>
</tr>
</tbody>
</table>

### Type I SS

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>height</td>
<td>2</td>
<td>1544.000000</td>
<td>772.000000</td>
<td>74.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>12.000000</td>
<td>12.000000</td>
<td>1.16</td>
<td>0.3226</td>
</tr>
<tr>
<td>height*width</td>
<td>2</td>
<td>24.000000</td>
<td>12.000000</td>
<td>1.16</td>
<td>0.3747</td>
</tr>
</tbody>
</table>

### Type III SS

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>height</td>
<td>2</td>
<td>1544.000000</td>
<td>772.000000</td>
<td>74.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>12.000000</td>
<td>12.000000</td>
<td>1.16</td>
<td>0.3226</td>
</tr>
<tr>
<td>height*width</td>
<td>2</td>
<td>24.000000</td>
<td>12.000000</td>
<td>1.16</td>
<td>0.3747</td>
</tr>
</tbody>
</table>
## Output

<table>
<thead>
<tr>
<th>Level of</th>
<th>-----------sales-----------</th>
</tr>
</thead>
<tbody>
<tr>
<td>height</td>
<td>N</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of</th>
<th>-----------sales-----------</th>
</tr>
</thead>
<tbody>
<tr>
<td>width</td>
<td>N</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of</th>
<th>Level of</th>
<th>-----------sales-----------</th>
</tr>
</thead>
<tbody>
<tr>
<td>height</td>
<td>width</td>
<td>N</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Interaction Plot
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 ($P < 0.00001$). 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 be performed.
Pooling

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

\[
\text{SSE}^* = \text{SSE} + \text{SSAB}
\]

\[
\text{df}^*_E = ab(n-1) + (a-1)(b-1)
\]

- Increases DF but could inflate \(\hat{\sigma}^2\)

- Could encounter Type II error

- Only pool when df small (e.g., < 5) and P-value large (e.g., > 0.25)
Output

Tukey’s Studentized Range (HSD) Test for sales

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>height</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>67.000</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>44.000</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>42.000</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

**** POOLING ****

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>height</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>67.000</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>44.000</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>42.000</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
General Plan

- Construct scatterplot / interaction plot
- Run full model
- Check assumptions
  - Residual plots
  - Histogram / QQplot
  - Ordered residuals plot
- Check significance of interaction
Interaction Not Significant

- Determine whether pooling is beneficial
- If yes, rerun analysis without interaction
- Check significance of main effects
- If factor insignificant, determine whether pooling is beneficial
- If yes, rerun analysis as one-way ANOVA
- If statistically significant factor has more than two levels, use multiple comparison procedure to assess differences
- Contrasts and linear combinations can also be used
Interactions

- **Opposite behavior** (no Factor 2 effect?)

- **Not similar increase** (still Factor 2 effect?)
If Interaction Significant

• Determine if interaction “important”
  – May not be of practical importance
  – May be like plot #2
  – Often due to one cell mean

• If no, use previously described methods making sure to leave interaction in the model (no pooling). Carefully interpret the marginal means as averages over the levels of the other factor and not a main effect

• If yes, take approach of one-way ANOVA with \( ab \) levels. Use linear combinations to compare various means (e.g., levels of factor \( A \) for each level of factor \( B \)). Use the interaction plots for discussion purposes.
Using Estimate Statement

- Must formulate in terms of factor effects model

- Order of factors determined by order in class statement not the model statement

- Example of a contrast in the Castle Bread Company problem:
  - Equivalent null hypotheses:
    
    $H_0 : \mu_2 = \mu_1 + \mu_3.$
    
    $H_0 : \mu_2 - \mu_1 - \mu_3 = 0$
    
    $H_0 : \frac{1}{2}(\mu_{21} + \mu_{22}) - \frac{1}{2}(\mu_{11} + \mu_{12}) - \frac{1}{2}(\mu_{31} + \mu_{32}) = 0$
    
    $H_0 : (\mu_{21} + \mu_{22}) - (\mu_{11} + \mu_{12}) - (\mu_{31} + \mu_{32}) = 0$
Using Estimate Statement

- Rewriting in factor effect terms

\[
[\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}] + [\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}]
- [\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}] + [\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}]
- [\mu + \alpha_3 + \beta_1 + (\alpha\beta)_{31}] + [\mu + \alpha_3 + \beta_2 + (\alpha\beta)_{32}]
= 0
\]

- Simplifies to

\[
-2\mu - 2\alpha_1 + 2\alpha_2 - 2\alpha_3 - \beta_1 - \beta_2
- (\alpha\beta)_{11} - (\alpha\beta)_{12} + (\alpha\beta)_{21} + (\alpha\beta)_{22} - (\alpha\beta)_{31} - (\alpha\beta)_{32}
= 0
\]

- Use these coefficients in the contrast statement

  - use solution to see the order of the coefficients in SAS
Using Slice Statement

• Slice option performs one-way ANOVA for fixed level of other factor

• Can also express that as contrast statement

• Following output presents results from two contrasts

  - $H_0 : 2\mu_2. = \mu_1. + \mu_3.$

  - $H_0 : \mu_{11} = \mu_{21} = \mu_{31}$

• See if you can come up with the same contrast statements
SAS Commands

options nocenter ls=75;

data a1; infile 'u:\.www\datasets525\CH19TA07.DAT';
   input sales height width;

proc glm data=a1;
   class height width;
   model sales=height width height*width;
   estimate 'middle is sum of other two heights'
      intercept -2
      height -2 2 -2
      width -1 -1
      height*width -1 -1 1 1 -1 -1;
   contrast 'middle two vs all others'
      height -.5 1 -.5
      height*width -.25 -.25 .5 .5 -.25 -.25;
   estimate 'middle two vs all others'
      height -.5 1 -.5
      height*width -.25 -.25 .5 .5 -.25 -.25;
   contrast 'height same for normal width'
      height 1 -1 0 height*width 1 0 -1 0 0 0,
      height 0 1 -1 height*width 0 0 1 0 -1 0;
   means height*width;

proc glm data=a1;
   class height width;
   model sales=height width height*width;
   lsmeans height*width / slice=width;
## Output

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>middle vs others</td>
<td>1</td>
<td>1536.000000</td>
<td>1536.000000</td>
<td>148.65</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>height for normal</td>
<td>2</td>
<td>700.000000</td>
<td>350.000000</td>
<td>33.87</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

| Parameter                  | Estimate | Error      | t Value | Pr > |t| |
|---------------------------|----------|------------|---------|-------|-----|
| middle is sum of others   | -38.00000000 | 5.56776436 | -6.83  | 0.0005 |
| middle vs others          | 24.00000000 | 1.96850197 | 12.19  | <.0001 |

<table>
<thead>
<tr>
<th>Level of height</th>
<th>Level of width</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>45.00000000</td>
<td>2.82842712</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>43.00000000</td>
<td>4.24264069</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>65.00000000</td>
<td>4.24264069</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>69.00000000</td>
<td>2.82842712</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>40.00000000</td>
<td>1.41421356</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>44.00000000</td>
<td>2.82842712</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum of height*width Effect Sliced by width for sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>width</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>
Multiple comparisons of means: no interaction

- Critical values when comparing means of factor A:

  Tukey (all pairs): $\frac{1}{\sqrt{2}} q (1 - \alpha, a, (n - 1)ab)$

  Bonferroni (g comparisons): $t \left(1 - \frac{\alpha}{(2g)}, (n - 1)ab\right)$

  Scheffé (all contrasts): $(a-1)F \left(1 - \alpha, a - 1, (n - 1)ab\right)$
Multiple comparisons of means: interaction present

- Critical values when comparing means of two combinations of factors:

  Tukey (all pairs): \( \frac{1}{\sqrt{2}} q (1 - \alpha, ab, (n - 1)ab) \)

  Bonferroni (g comparisons): \( t (1 - \alpha/(2g), (n - 1)ab) \)

  Scheffé (all contrasts): \( (ab-1)F (1 - \alpha, ab - 1, (n - 1)ab) \)
Ch. 20: One Observation Per Cell

• Do not have enough information to estimate *both* the interaction effect and error variance

• With interaction, error degrees of freedom is \(ab(n - 1) = 0\)

• Common to assume there is no interaction (i.e., pooling)
  
  - \(SSE^* = SSAB + 0\)
  
  - \(df_E^* = df_{AB} + 0\)

• Can also test for less general type of interaction that requires fewer degrees of freedom
Tukey’s Test for Additivity

• Consider special type of interaction

• Assume following model

\[ Y_{ij} = \mu + \alpha_i + \beta_j + \theta \alpha_i \beta_j + \varepsilon_{ij} \]

• Uses up only one degree of freedom

• Other variations possible (e.g., \( \theta_i \beta_j \))

• Want to test \( H_0 : \theta = 0 \)

• Will use regression after estimating factor effects to test \( \theta \)
• $Y$ is the premium for auto insurance

• Factor $A$ is the size of the city
  – $a = 3$: small, medium, large

• Factor $B$ is the region
  – $b = 2$: east, west

• Only one city per cell was observed
SAS Commands

data a1; infile 'u:\.www\datasets525\CH20TA02.txt';
    input premium size region;

if size=1 then sizea='1_small ';
if size=2 then sizea='2_medium';
if size=3 then sizea='3_large ';

proc glm data=a1;
    class sizea region;
    model premium=sizea region / solution;
    means sizea region / tukey;

symbol1 v='E' i=join c=black; symbol2 v='W' i=join c=black;
title1 'Plot of the data';
proc gplot data=a2;
    plot premium*sizea=region/frame;
Scatterplot

Plot of the data
SAS Commands

proc glm data=a1;
   model premium=;
   output out=aall p=muhat;

proc glm data=a1;
   class size;
   model premium=size;
   output out=aA p=muhatA;

proc glm data=a1;
   class region;
   model premium=region;
   output out=aB p=muhatB;

data a2; merge aall aA aB;
   alpha=muhatA-muhat;
   beta=muhatB-muhat;
   atimesb=alpha*beta;

proc print data=a2;
   var size region atimesb;

proc glm data=a2;
   class size region;
   model premium=size region atimesb/solution;
run;
Output

<table>
<thead>
<tr>
<th>Obs</th>
<th>size</th>
<th>region</th>
<th>atimesb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-825</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>825</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>300</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>-300</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1</td>
<td>525</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>-525</td>
</tr>
</tbody>
</table>

Note: These estimates are based on the factor effects model where $\sum \alpha = 0$ and $\sum \beta = 0$. While not shown, the following were used to compute atimesb: $\hat{\mu} = 175$, $\hat{\mu}_1 = 120$, $\hat{\mu}_2 = 195$, $\hat{\mu}_3 = 210$, $\hat{\mu}_1 = 190$, and $\hat{\mu}_2 = 160$. 
## Output

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>10737.09677</td>
<td>2684.27419</td>
<td>208.03</td>
<td>0.0519</td>
</tr>
<tr>
<td>Error</td>
<td>1</td>
<td>12.90323</td>
<td>12.90323</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>5</td>
<td>10750.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Coeff Var</th>
<th>Root MSE</th>
<th>premium Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.998800</td>
<td>2.052632</td>
<td>3.592106</td>
<td>175.00000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>size</td>
<td>2</td>
<td>9300.000000</td>
<td>4650.000000</td>
<td>360.37</td>
<td>0.0372</td>
</tr>
<tr>
<td>region</td>
<td>1</td>
<td>1350.000000</td>
<td>1350.000000</td>
<td>104.62</td>
<td>0.0620</td>
</tr>
<tr>
<td>atimesb</td>
<td>1</td>
<td>87.096774</td>
<td>87.096774</td>
<td>6.75</td>
<td>0.2339</td>
</tr>
</tbody>
</table>

| Parameter | Estimate | Error | t Value | Pr > |t| |
|-----------|----------|-------|---------|-------|-------|
| Intercept | 195.0000000 B | 2.93294230 | 66.49 | 0.0096 |
| size 1    | -90.0000000 B | 3.59210604 | -25.05 | 0.0254 |
| size 2    | -15.0000000 B | 3.59210604 | -4.18  | 0.1496 |
| size 3    | 0.0000000 B   | .       | .      | .    |
| region 1  | 30.0000000 B  | 2.93294230 | 10.23  | 0.0620 |
| region 2  | 0.0000000 B   | .       | .      | .    |
| atimesb   | -0.0064516    | 0.00248323 | -2.60  | 0.2339 |

Note: These are the same parameter estimates as the original model without the interaction term.