Two-way ANOVA

- **Two categorical factors** – combinations of levels are treatment levels
- Breaks up one-way ANOVA to see marginal and interaction effects
- Cell means model use treatment level means (and constant error variance) to explain variability in observations
- Interaction treated (notationally) as product
– In regression expansion, use products of indicators
– Explains variability of cell means not explained by individual components.
– Is variability of one factor dependent on level of other factor?
– Can use interaction plots to diagnose if/where interaction occurs

• Factor effects explicitly includes interaction effect.
• Simple constraints (more than one) mean factor effects are simple functions of treatment means.
• Estimates of factor effects are simple functions of simple treatment sample means.
Inference

- Break down model $SS$ into factor effects and interaction $SS$

- Rules for degrees of freedom
  - *marginal effects* – number of levels $- 1$
  - *interaction effects* – products of marginal $df$’s
  - *total* – number of observations $- 1$
  - *model* – sum of marginal and interaction $df$’s
  - *error* – difference between total and model
• Get several $F$-tests

• Since effects are fixed use $MSE$ in denominator

• $p$-values calculated based on $F$ distribution and significant if small
**KNNL Example: ANOVA with GLM**

```proc glm data=bread;
   class height width;
   model sales=height width height*width;
```

The GLM Procedure

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>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>

Latest update April 5, 2016
<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>1544.0000000</td>
<td>772.000000</td>
<td>74.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>12.0000000</td>
<td>12.000000</td>
<td>1.16</td>
<td>0.3226</td>
</tr>
<tr>
<td>height * width</td>
<td>2</td>
<td>24.0000000</td>
<td>12.000000</td>
<td>1.16</td>
<td>0.3747</td>
</tr>
</tbody>
</table>

<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>1544.0000000</td>
<td>772.000000</td>
<td>74.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>12.0000000</td>
<td>12.000000</td>
<td>1.16</td>
<td>0.3226</td>
</tr>
<tr>
<td>height * width</td>
<td>2</td>
<td>24.0000000</td>
<td>12.000000</td>
<td>1.16</td>
<td>0.3747</td>
</tr>
</tbody>
</table>
Sums of Squares

- Type I $SS$ are again the sequential sums of squares (variables added in order). Thus height explains 1544, width explains 12 of what is left, and the interaction explains 24 of what is left after that.
• Type III SS is like Type II SS (variable added last) but it also adjusts for differing $n_{i,j}$. So if all cells have the same number of observations (balanced designs are nice - the variables height and width in our example are independent - no multicollinearity!) $SS_1$, $SS_2$, and $SS_3$ will all be the same.

• More details on SS later.

<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>
Results

• The interaction between height and width is not statistically significant ($F = 1.16; df = (2, 6); p = 0.37$). **NOTE: Check Interaction FIRST! If it is significant then main effects are left in the model, even if not significant themselves!** We may now go on to examine main effects since our interaction is not significant.

• The main effect of height is statistically significant ($F = 74.71; df = (2, 6); p = 4.75 \times 10^{-5}$).
• The main effect of width is not statistically significant

\[(F = 1.16; df = (1, 6); p = 0.32)\]
**Interpretation**

- The height of the display affects sales of bread.
- The width of the display has no apparent effect.
- The effect of the height of the display is similar for both the regular and the wide widths.

**Additional Analyses**

- We will need to do additional analyses to understand the height effect (factor $A$).
• There were three levels: bottom, middle and top. Based on the interaction picture, it appears the middle shelf increases sales.

• We could rerun the data with a one-way ANOVA and use the methods we learned in the previous chapters to show this (e.g. Tukey).
Cell Means Model

\[ Y_{i,j,k} = \mu_{i,j} + \epsilon_{i,j,k}, \text{ where} \]

• \( \mu_{i,j} \) is the theoretical mean or expected value of all observations in cell \((i, j)\).

• \( \epsilon_{i,j,k} \sim iid \ N(0, \sigma^2) \)

• \( Y_{i,j,k} \sim N(\mu_{i,j,k}, \sigma^2) \) are independent
• There are $ab + 1$ parameters of the model: $\mu_{i,j}$

$i = 1, \ldots, a, j = 1, \ldots, b$ and $\sigma^2$.

For the bread example, estimate the $\mu_{i,j}$ with $\bar{Y}_{i,j}$ which we can get from the means.
**Height * Width** statement:

\[ \hat{\mu}_{1,1} = \bar{Y}_{1,1.} = 45 \]
\[ \hat{\mu}_{1,2} = \bar{Y}_{1,2.} = 43 \]
\[ \hat{\mu}_{2,1} = \bar{Y}_{2,1.} = 65 \]
\[ \hat{\mu}_{2,2} = \bar{Y}_{2,2.} = 69 \]
\[ \hat{\mu}_{3,1} = \bar{Y}_{3,1.} = 40 \]
\[ \hat{\mu}_{3,2} = \bar{Y}_{3,2.} = 44 \]

As usual, \( \sigma^2 \) is estimated by \( MSE \).
Factor Effects Model

\[ \mu_{i,j} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j}, \]  

where

- \( \mu \) is the overall (grand) mean - it is \( \mu_{..} \) in KNNL
- \( \alpha_i \) is the main effect of Factor \( A \)
- \( \beta_j \) is the main effect of Factor \( B \)
- \( (\alpha\beta)_{i,j} \) is the interaction effect between \( A \) and \( B \).

Note that \( (\alpha\beta)_{i,j} \) is the name of a parameter all on its own and does not refer to the product of \( \alpha \) and \( \beta \).
Overall Mean

The overall mean is estimated as $\hat{\mu} = \bar{Y} = 51$ under the zero-sum constraining. (This is sales mean in the \texttt{glm} output). You can get a whole dataset of this value by using a \texttt{model} statement with no right hand side, e.g. \texttt{model sales=} ; and storing the predicted values.
Main Effects

The main effect of $A$ is estimated from the means height output. You can get a whole dataset of the $\hat{\mu}_i$ by running a model with just $A$, e.g. `model sales = height`; and storing the predicted values.
To estimate the $\alpha$’s, you then subtract $\hat{\mu}$ from each height mean.

$\hat{\mu}_1. = \bar{Y}_1.. = 44 \implies \hat{\alpha}_1 = 44 - 51 = -7$

$\hat{\mu}_2. = \bar{Y}_2.. = 67 \implies \hat{\alpha}_2 = 67 - 51 = +16$

$\hat{\mu}_3. = \bar{Y}_3.. = 42 \implies \hat{\alpha}_3 = 42 - 51 = -9$

This says that “middle” shelf height has the effect of a relative increase in sales by 16, while bottom and top decrease the sales by 7 and 9 respectively. Notice that these sum to zero so that there is no “net” effect (there are only 2 free parameters, $df_A = 2$).
The main effect of $B$ is similarly estimated from the means width output, or by storing the predicted values of model sales = width; then subtract $\hat{\mu}$ from each height mean.

\[
\begin{align*}
\hat{\mu}_1 &= \bar{Y}_1. = 50 \Rightarrow \hat{\beta}_1 = 50 - 51 = -1 \\
\hat{\mu}_2 &= \bar{Y}_2. = 52 \Rightarrow \hat{\beta}_2 = 52 - 51 = +1
\end{align*}
\]

Wide display increases sales by an average of 1, while regular display decreases sales by 1 (they sum to zero so there's only 1 free parameter, $df_B = 1$).
Interaction Effects

Recall that $\hat{\alpha}_i, j = \hat{\mu}_{i, j} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)$. This is the difference between the treatment mean and the value predicted by the overall mean and main effects only (i.e. by the additive model).

You can get the treatment means from the `means height*width` statement, or by the predicted values of `model sales=height*width;` then subtract the appropriate combination of the previously
estimated parameters.

$$(\hat{\alpha}_1 \hat{\beta}_1)_{11} = \bar{Y}_{11} - (\hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1) = 45 - (51 - 7 - 1) = +2$$

$$(\hat{\alpha}_1 \hat{\beta}_2)_{12} = \bar{Y}_{12} - (\hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_2) = 43 - (51 - 7 + 1) = -2$$

$$(\hat{\alpha}_2 \hat{\beta}_1)_{21} = \bar{Y}_{21} - (\hat{\mu} + \hat{\alpha}_2 + \hat{\beta}_1) = 65 - (51 + 16 - 1) = -1$$

$$(\hat{\alpha}_2 \hat{\beta}_2)_{22} = \bar{Y}_{22} - (\hat{\mu} + \hat{\alpha}_2 + \hat{\beta}_2) = 69 - (51 + 16 + 1) = +1$$

$$(\hat{\alpha}_3 \hat{\beta}_1)_{31} = \bar{Y}_{31} - (\hat{\mu} + \hat{\alpha}_3 + \hat{\beta}_1) = 40 - (51 - 9 - 1) = -1$$

$$(\hat{\alpha}_3 \hat{\beta}_2)_{32} = \bar{Y}_{32} - (\hat{\mu} + \hat{\alpha}_3 + \hat{\beta}_2) = 44 - (51 - 9 + 1) = +1$$
Notice that they sum in pairs (over $j$) to zero and also the sum over $i$ is zero for each $j$. Thus there are in reality only two free parameters here ($df_{AB} = 2$).
Doing this in SAS

Unlike `proc reg`, you are only allowed one `model` statement per call to `glm`. But this at least saves you doing the arithmetic by hand.

```sas
proc glm data=bread;
  class height width;
  model sales=;
  output out=pmu p=muhat;
proc glm data=bread;
  class height width;
  model sales=height;
  output out=pA p=Amean;
```
proc glm data=bread;
   class height width;
   model sales=width;
   output out=pB p=Bmean;
proc glm data=bread;
   class height width;
   model sales=height*width;
   output out=pAB p=ABmean;
data parmest;
   merge bread pmu pA pB pAB;
   alpha = Amean - muhat;
   beta = Bmean - muhat;
   alphabeta = ABmean - (muhat+alpha+beta);
proc print data=parmest;
<table>
<thead>
<tr>
<th>Obs</th>
<th>sales</th>
<th>height</th>
<th>width</th>
<th>muhat</th>
<th>Amean</th>
<th>Bmean</th>
<th>ABmean</th>
<th>alpha</th>
<th>beta</th>
<th>alphabeta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47</td>
<td>1</td>
<td>1</td>
<td>51</td>
<td>44</td>
<td>50</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>43</td>
<td>1</td>
<td>1</td>
<td>51</td>
<td>44</td>
<td>50</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>46</td>
<td>1</td>
<td>2</td>
<td>51</td>
<td>44</td>
<td>52</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>1</td>
<td>2</td>
<td>51</td>
<td>44</td>
<td>52</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>62</td>
<td>2</td>
<td>1</td>
<td>51</td>
<td>67</td>
<td>50</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>68</td>
<td>2</td>
<td>1</td>
<td>51</td>
<td>67</td>
<td>50</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>67</td>
<td>2</td>
<td>2</td>
<td>51</td>
<td>67</td>
<td>52</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>71</td>
<td>2</td>
<td>2</td>
<td>51</td>
<td>67</td>
<td>52</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>41</td>
<td>3</td>
<td>1</td>
<td>51</td>
<td>42</td>
<td>50</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>39</td>
<td>3</td>
<td>1</td>
<td>51</td>
<td>42</td>
<td>50</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>42</td>
<td>3</td>
<td>2</td>
<td>51</td>
<td>42</td>
<td>52</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Zero-sum Constraints

\[ \alpha. = \sum_{i} \alpha_i = 0 \]

\[ \beta. = \sum_{j} \beta_{j} = 0 \]

\[ (\alpha\beta)_{.j} = \sum_{i} (\alpha\beta)_{i,j} = 0 \quad \forall j \quad (\text{for all } j) \]

\[ (\alpha\beta)_{.i} = \sum_{i} (\alpha\beta)_{i,j} = 0 \quad \forall i \quad (\text{for all } i) \]

All of these constraints are satisfied by the above estimates.
Notice how these main and interaction effects fit together to give back the treatment means:

\[
\begin{align*}
45 & = 51 - 7 - 1 + 2 \\
43 & = 51 - 7 + 1 - 2 \\
65 & = 51 + 16 - 1 - 1 \\
69 & = 51 + 16 + 1 + 1 \\
40 & = 51 - 9 - 1 - 1 \\
44 & = 51 - 9 + 1 + 1 \\
\end{align*}
\]
SAS GLM Constraints

As usual, SAS has to do its constraints differently. As in one-way ANOVA, it sets the parameter for the last category equal to zero.

\[ \alpha_a = 0 \quad (1 \text{ constraint}) \]
\[ \beta_b = 0 \quad (1 \text{ constraint}) \]
\[ (\alpha\beta)_{a,j} = 0 \quad \text{for all } j \quad (b \text{ constraints}) \]
\[ (\alpha\beta)_{i,b} = 0 \quad \text{for all } i \quad (a \text{ constraints}) \]
The total is $1 + 1 + a + b - 1 = a + b + 1$ constraints (the constraint $(\alpha/\beta)_{a,b}$ is counted twice above).
Parameters and constraints

The cell means model has $ab$ parameters for the means.
The factor effects model has $(1 + a + b + ab)$ parameters.

- An intercept $(1)$
- Main effect of $A$ $(a)$
- Main effect of $B$ $(b)$
- Interaction of $A$ and $B$ $(ab)$
There are $1 + a + b + ab$ parameters and $1 + a + b$ constraints, so there are $ab$ remaining unconstrained parameters (or sets of parameters), the same number of parameters for the means in the cell means model. This is the number of parameters we can actually estimate.
KNNL page 833 (knnl833b.sas)

\(Y\) is the number of cases of bread sold

\(A\) is the height of the shelf display, \(a = 3\) levels: bottom, middle, top

\(B\) is the width of the shelf display, \(b = 2\): regular, wide

\(n = 2\) stores for each of the \(3 \times 2\) treatment combinations
proc glm with solution

We will get *different estimates* for the parameters here because a different constraint system is used.
proc glm data=bread;
  class height width;
  model sales=height width height*width/solution;
  means height*width;
## Solution output

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>44.00000000</td>
<td>$B^* = \hat{\mu}$</td>
</tr>
<tr>
<td>height 1</td>
<td>-1.00000000</td>
<td>$B = \hat{\alpha}_1$</td>
</tr>
<tr>
<td>height 2</td>
<td>25.00000000</td>
<td>$B^* = \hat{\alpha}_2$</td>
</tr>
<tr>
<td>height 3</td>
<td>0.00000000</td>
<td>$B = \hat{\alpha}_3$</td>
</tr>
<tr>
<td>width 1</td>
<td>-4.00000000</td>
<td>$B = \hat{\beta}_1$</td>
</tr>
<tr>
<td>width 2</td>
<td>0.00000000</td>
<td>$B = \hat{\beta}_2$</td>
</tr>
<tr>
<td>height * width</td>
<td>1 1</td>
<td>6.00000000 B</td>
</tr>
<tr>
<td>height * width</td>
<td>1 2</td>
<td>0.00000000 B</td>
</tr>
<tr>
<td>height * width</td>
<td>2 1</td>
<td>0.00000000 B</td>
</tr>
<tr>
<td>height * width</td>
<td>2 2</td>
<td>0.00000000 B</td>
</tr>
<tr>
<td>height * width</td>
<td>3 1</td>
<td>0.00000000 B</td>
</tr>
<tr>
<td>height * width</td>
<td>3 2</td>
<td>0.00000000 B</td>
</tr>
</tbody>
</table>
It also prints out standard errors, $t$-tests and $p$-values for testing whether each parameter is equal to zero. That output has been omitted here but the significant ones have been starred.

Notice that the last $\alpha$ and $\beta$ are set to zero, as well as the last $\hat{\alpha\beta}$ in each category. *They no longer sum to zero.*
Means

The estimated treatment means are

$$\hat{\mu}_{i,j} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + (\hat{\alpha}\hat{\beta})_{i,j}.$$
## ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$a - 1$</td>
<td>$SSA$</td>
<td>$MSA$</td>
<td>$MSA/MSE$</td>
</tr>
<tr>
<td>$B$</td>
<td>$b - 1$</td>
<td>$SSB$</td>
<td>$MSB$</td>
<td>$MSB/MSE$</td>
</tr>
<tr>
<td>$AB$</td>
<td>$(a - 1)(b - 1)$</td>
<td>$SSAB$</td>
<td>$MSAB$</td>
<td>$MSAB/MSE$</td>
</tr>
<tr>
<td>Error</td>
<td>$ab(n - 1)$</td>
<td>$SSE$</td>
<td>$MSE$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$abn - 1$</td>
<td>$SSTO$</td>
<td>$MST$</td>
<td></td>
</tr>
</tbody>
</table>
Expected Mean Squares

\[ E(MSE) = \sigma^2 \]
\[ E(MSA) = \sigma^2 + \frac{nb}{a-1} \sum_i \alpha_i^2 \]
\[ E(MSB) = \sigma^2 + \frac{na}{b-1} \sum_j \beta_j^2 \]
\[ E(MSAB) = \sigma^2 + \frac{n}{(a-1)(b-1)} \sum_{i,j} (\alpha \beta)_{i,j}^2 \]

Here, \( \alpha_i, \beta_j \), and \( (\alpha \beta)_{i,j} \) are defined with the usual zero-sum constraints.
Analytical strategies

- Run the model with main effects and the two-way interaction.
- Plot the data, the means and look at the residuals.
- Check the significance test for the interaction.
What if $AB$ interaction is not significant?

If the $AB$ interaction is not statistically significant, you could rerun the analysis without the interaction (see discussion of pooling KNNL Section 19.10). This will put the $SS$ and $df$ for $AB$ into Error. Results of main effect hypothesis tests could change because $MSE$ and denominator $df$ have changed (more impact with small sample size). **If one main effect is not significant...**

- There is no evidence to conclude that the levels of this explanatory variable are associated with different
means of the response variable.

- Model could be rerun without this factor giving a one-way ANOVA.

If neither main effect is significant...

- Model could be run as $Y = \_;$ (i.e. no factors at all)

- A one population model

- This seems silly, but this syntax can be useful for getting parameter estimates in the preferred constraint system (see below).
For a main effect with more than two levels that is significant, use the \texttt{means} statement with the Tukey multiple comparison procedure. Contrasts and linear combinations can also be examined using the \texttt{contrast} and \texttt{estimate} statements.
If $AB$ interaction is significant but not important

- Plots and a careful examination of the cell means may indicate that the interaction is not very important even though it is statistically significant.
- For example, the interaction effect may be much smaller in magnitude than the main effects; or may only be apparent in a small number of treatments.
- Use the marginal means for each significant main effect to describe the important results for the main effects.
• You may need to qualify these results using the interaction.

• Keep the interaction in the model.

• Carefully interpret the marginal means as averages over the levels of the other factor.

• KNNL also discuss ways that transformations can sometimes eliminate interactions.
If $AB$ interaction is significant and important

The interaction effect is so large and/or pervasive that main effects cannot be interpreted on their own.

Options include the following:

- Treat as a one-way ANOVA with $ab$ levels; use Tukey to compare means; contrast and estimate can also be useful.
- Report that the interaction is significant; plot the means and describe the pattern.
- Analyze the levels of $A$ for each level of $B$ (use a
by statement) or vice versa
### Strategy for the bread example

### Previous results with interaction

The GLM Procedure

Dependent Variable: sales

<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>1580.0000000</td>
<td>316.0000000</td>
<td>30.58</td>
<td>0.0003</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>62.0000000</td>
<td>10.3333333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11</td>
<td>1642.0000000</td>
<td></td>
<td></td>
<td></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>height</td>
<td>2</td>
<td>1544.0000000</td>
<td>772.0000000</td>
<td>74.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>12.0000000</td>
<td>12.0000000</td>
<td>1.16</td>
<td>0.326</td>
</tr>
<tr>
<td>height*width</td>
<td>2</td>
<td>24.0000000</td>
<td>12.0000000</td>
<td>1.16</td>
<td>0.374</td>
</tr>
</tbody>
</table>
**Rerun without interaction**

```plaintext
proc glm data=bread;
   class height width;
   model sales=height width;
   means height / tukey lines;
```

<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>3</td>
<td>1556.000000</td>
<td>518.666667</td>
<td>48.25</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>86.000000</td>
<td>10.750000</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>

<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>71.81</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>12.000000</td>
<td>12.000000</td>
<td>1.12</td>
<td>0</td>
</tr>
</tbody>
</table>
**Pooling** $SS$

\[ \text{Data} = \text{Model} + \text{Residual} \]

When we remove a term from the ‘model’, we put this variation and the associated $df$ into ‘residual’. This is called *pooling*. A benefit is that we have more $df$ for error and a simpler model. A drawback is that if the $SS$ for that term is large it will increase the $SSE$ too much. Therefore we would only want to do this for insignificant terms, i.e. those with small $SS$, most often the interaction term.
This strategy can be beneficial in small experiments where the $df_E$ is very small.

Do not remove the main effect and leave the interaction term. Typically we are pooling $SSE$ and $SSAB$.

In our example, $MSh$ and $MSw$ have not changed, but $MSE$, $F$’s, and $p$-values have changed. In this case $MSE$ went up, but in other cases it might go down.

Note $SSE: 62 + 24 = 86$; and $df_E: 6 + 2 = 8$. 
Tukey Output

<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>

As we noticed from the plot, the middle shelf is significantly different (better in terms of sales if we look at the plot) from the other two.
Specification of contrast and estimate statements

- When using the `contrast` statement, you can double check your results with the `estimate` statement.

- The order of factors is determined by the order in the `class` statement, not the order in the `model` statement.

- **Contrasts to be examined should come *a priori* from research questions, not from questions that arise after looking at the plots and means.**
**Contrast-Estimate Example**

For the bread example, suppose we want to compare the average of the two height = middle cells with the average of the other four cells; i.e. look at “eye-level” vs. “not eye-level” (for the average person). With this approach, the contrast should correspond to a research question formulated before examining the data. First, formulate the question as a contrast in terms of the cell means model:

\[
H_0 : \frac{\left( \mu_{2,1} + \mu_{2,2} \right)}{2} = \frac{\left( \mu_{1,1} + \mu_{1,2} + \mu_{3,1} + \mu_{3,2} \right)}{4}
\]
or $L = 0$, where

$$L = -0.25\mu_{1,1} - 0.25\mu_{1,2} + 0.5\mu_{2,1} + 0.5\mu_{2,2} - 0.25\mu_{3,1} - 0.25\mu_{3,2}$$
Then translate the contrast into the factor effects model using
\[ \mu_{i,j} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} \]

\[-0.25\mu_{1,1} = -0.25(\mu + \alpha_1 + \beta_1 + \alpha\beta_{1,1})\]
\[-0.25\mu_{1,2} = -0.25(\mu + \alpha_1 + \beta_2 + \alpha\beta_{1,2})\]
\[0.50\mu_{2,1} = +0.50(\mu + \alpha_2 + \beta_1 + \alpha\beta_{2,1})\]
\[0.50\mu_{2,2} = +0.50(\mu + \alpha_2 + \beta_2 + \alpha\beta_{2,2})\]
\[-0.25\mu_{3,1} = -0.25(\mu + \alpha_3 + \beta_1 + \alpha\beta_{3,1})\]
\[-0.25\mu_{3,2} = -0.25(\mu + \alpha_3 + \beta_2 + \alpha\beta_{3,2})\]
\[ L = (-0.5\alpha_1 + \alpha_2 - 0.5\alpha_3) \]

\[ + (-0.25\alpha\beta_{1,1} - 0.25\alpha\beta_{1,2} + 0.5\alpha\beta_{2,1} + 0.5\alpha\beta_{2,2} \]

\[ - 0.25\alpha\beta_{3,1} - 0.25\alpha\beta_{3,2}) \]

Note the \( \beta \)'s do not appear in this contrast because we are looking at height only and averaging over width (this would not necessarily be true in an unbalanced design).
proc glm with contrast and estimate

(knnl860.sas)

proc glm data=bread;
    class height width;
    model sales=height width height*width;
    contrast 'middle vs others'
        height -.5 1 -.5
        height*width -.25 -.25 .5 .5 -.25 -.25;
    estimate 'middle vs others'
        height -.5 1 -.5
        height*width -.25 -.25 .5 .5 -.25 -.25;
    means height*width;

Latest update April 5, 2016
## 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>
</tbody>
</table>

### Standard

| Parameter       | Estimate | Error         | t Value | Pr > |t| |
|-----------------|----------|---------------|---------|------| -|
| middle vs others | 24.000000 | 1.96850197 | 12.19   | <.0001 |
Check with \textit{means}

\[
\hat{L} = \frac{(65 + 69)}{2} - \frac{(45 + 43 + 40 + 44)}{4} = 24
\]
Combining with Quantitative Factors

Sometimes a factor can be interpreted as either categorical or quantitative. For example, “low, medium, high” or actual height above floor. If there are replicates for a quantitative factor we could use either regression or ANOVA. Recall that GLM will treat a factor as quantitative unless it is listed in the class statement. Notice that you can use ANOVA even if the relationship with the quantitative variable is non-linear, whereas with regression you would have to find that relationship.
One Quantitative factor and one categorical

- Plot the means vs the quantitative factor for each level of the categorical factor
- Consider linear and quadratic terms for the quantitative factor
- Consider different slopes for the different levels of the categorical factor; i.e., interaction terms.
- Lack of fit analysis can be useful (recall trainhrs example).
Two Quantitative factors

- Plot the means vs $A$ for each level of $B$
- Plot the means vs $B$ for each level of $A$
- Consider linear and quadratic terms.
- Consider products to allow for interaction.
- Lack of fit analysis can be useful.
Chapter 20: One Observation per Cell

For $Y_{i,j,k}$, as usual

- $i$ denotes the level of the factor $A$
- $j$ denotes the level of the factor $B$
- $k$ denotes the $k$th observation in cell $(i, j)$
- $i = 1, \ldots , a$ levels of factor $A$
- $j = 1, \ldots , b$ levels of factor $B$
Now suppose we have $n = 1$ observation in each cell $(i, j)$. We can no longer estimate variances separately for each treatment. The impact is that we will not be able to estimate the interaction terms; we will have to assume no interaction.
Factor Effects Model

\[ \mu_{i,j} = \mu + \alpha_i + \beta_j \]

- \( \mu \) is the overall mean
- \( \alpha_i \) is the main effect of \( A \)
- \( \beta_j \) is the main effect of \( B \)

Because we have only one observation per cell, we do not have enough information to estimate the interaction in the usual way. We assume no interaction.
Constraints

- Text: $\sum \alpha_i = 0$ and $\sum \beta_j = 0$
- SAS `glm`: $\alpha_a = \beta_b = 0$

ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$a - 1$</td>
<td>$SSA$</td>
<td>$MSA$</td>
<td>$MSA/MSE$</td>
</tr>
<tr>
<td>$B$</td>
<td>$b - 1$</td>
<td>$SSB$</td>
<td>$MSB$</td>
<td>$MSB/MSE$</td>
</tr>
<tr>
<td>Error</td>
<td>$(a - 1)(b - 1)$</td>
<td>$SSE$</td>
<td>$MSE$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$ab - 1$</td>
<td>$SSTO$</td>
<td>$MST$</td>
<td></td>
</tr>
</tbody>
</table>
Expected Mean Squares

\[
E(MSE) = \sigma^2
\]

\[
E(MSA) = \sigma^2 + \frac{b}{a-1} \sum_{i} \alpha_i^2
\]

\[
E(MSB) = \sigma^2 + \frac{a}{b-1} \sum_{j} \beta_j^2
\]

Here, \(\alpha_i\) and \(\beta_j\) are defined with the zero-sum factor effects constraints.
KNNL Example

- KNNL page 882 (knnl882.sas)
- $Y$ is the premium for auto insurance
- $A$ is the size of the city, $a = 3$ levels: small, medium and large
- $B$ is the region, $b = 2$: East, West
- $n = 1$
- the response is the premium charged by a particular company
### The data

```sas
data carins;
infile 'H:\System\Desktop\CH21TA02.DAT';
  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 print data=carins;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>premium</th>
<th>size</th>
<th>region</th>
<th>sizea</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>140</td>
<td>1</td>
<td>1</td>
<td>1_small</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>1</td>
<td>2</td>
<td>1_small</td>
</tr>
<tr>
<td>3</td>
<td>210</td>
<td>2</td>
<td>1</td>
<td>2_medium</td>
</tr>
<tr>
<td>4</td>
<td>180</td>
<td>2</td>
<td>2</td>
<td>2_medium</td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>3</td>
<td>1</td>
<td>3_large</td>
</tr>
</tbody>
</table>
```
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>200</td>
<td>3</td>
<td>2</td>
<td>3_large</td>
<td></td>
</tr>
</tbody>
</table>
proc glm data=carins;
  class sizea region;
  model premium=sizea region/solution;
  means sizea region / tukey;
  output out=preds p=muhat;

The GLM Procedure

Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>sizea</td>
<td>3</td>
<td>1_small 2_medium 3_large</td>
</tr>
<tr>
<td>region</td>
<td>2</td>
<td>1 2</td>
</tr>
</tbody>
</table>

Number of observations 6
<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>10650.00000</td>
<td>3550.00000</td>
<td>71.00</td>
<td>0.0139</td>
</tr>
<tr>
<td>Error</td>
<td>2</td>
<td>100.00000</td>
<td>50.00000</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>

Notice that we only have 5 total df. If we had interaction in the model it would use up another 2 df and there would be 0 left to estimate error.

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Coeff Var</th>
<th>Root MSE</th>
<th>premium</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.990698</td>
<td>4.040610</td>
<td>7.071068</td>
<td>175.0000</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>DF</td>
<td>Type I SS</td>
<td>Mean Square</td>
<td>F Value</td>
</tr>
<tr>
<td>--------</td>
<td>----</td>
<td>-------------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>sizea</td>
<td>2</td>
<td>9300.000000</td>
<td>4650.000000</td>
<td>93.00</td>
</tr>
<tr>
<td>region</td>
<td>1</td>
<td>1350.000000</td>
<td>1350.000000</td>
<td>27.00</td>
</tr>
</tbody>
</table>

Both main effects are significant.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>195.00000000 B</td>
<td>5.77350269</td>
<td>33.77</td>
<td>0.0009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sizea 1_small</td>
<td>-90.00000000 B</td>
<td>7.07106781</td>
<td>-12.73</td>
<td>0.0061</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sizea 2_medium</td>
<td>-15.00000000 B</td>
<td>7.07106781</td>
<td>-2.12</td>
<td>0.1679</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sizea 3_large</td>
<td>0.00000000 B</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>region 1</td>
<td>30.00000000 B</td>
<td>5.77350269</td>
<td>5.20</td>
<td>0.0351</td>
<td></td>
<td></td>
</tr>
<tr>
<td>region 2</td>
<td>0.00000000 B</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Check vs predicted values (\( \hat{\mu} \))

<table>
<thead>
<tr>
<th>region</th>
<th>sizea</th>
<th>muhat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1_small</td>
<td>135 = 195 − 90 + 30</td>
</tr>
<tr>
<td>2</td>
<td>1_small</td>
<td>105 = 195 − 90</td>
</tr>
<tr>
<td>1</td>
<td>2_medium</td>
<td>210 = 195 − 15 + 30</td>
</tr>
<tr>
<td>2</td>
<td>2_medium</td>
<td>180 = 195 − 15</td>
</tr>
<tr>
<td>1</td>
<td>3_large</td>
<td>225 = 195 + 30</td>
</tr>
<tr>
<td>2</td>
<td>3_large</td>
<td>195 = 195</td>
</tr>
</tbody>
</table>
Multiple Comparisons $size$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>sizea</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>210.000</td>
<td>2</td>
<td>3_large</td>
</tr>
<tr>
<td>A</td>
<td>195.000</td>
<td>2</td>
<td>2_medium</td>
</tr>
<tr>
<td>B</td>
<td>120.000</td>
<td>2</td>
<td>1_small</td>
</tr>
</tbody>
</table>

The ANOVA results told us that $size$ was significant; now we additionally know that small is different from medium and large, but that medium and large do not differ significantly.
Multiple Comparisons \textit{region}

<table>
<thead>
<tr>
<th>A</th>
<th>190.000</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>160.000</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

The ANOVA results told us that these were different since region was significant (only two levels) …

So this gives us no new information.
Plot the data

symbol1 v='E' i=join c=black;
symbol2 v='W' i=join c=black;
title1 'Plot of the data';
proc gplot data=preds;
    plot premium*sizea=region;
The lines are not quite parallel, but the interaction, if any, does not appear to be substantial. If it was, our analysis would not be valid and we would need to collect more
data.
Plot the estimated model

title1 'Plot of the model estimates';
proc gplot data=preds;
   plot muhat*sizea=region;

Notice that the model estimates produce completely
parallel lines.
Tukey test for additivity

If we believe interaction is a problem, this is a possible way to test it without using up all our df.

One additional term is added to the model \((\theta)\), replacing the \((\alpha\beta)_{i,j}\) with the product:

\[
\mu_{i,j} = \mu + \alpha_i + \beta_j + \theta \alpha_i \beta_j
\]

We use one degree of freedom to estimate \(\theta\), leaving one left to estimate error. Of course, this only tests for interaction of the specified form, but it may be better than
nothing.

There are other variations on this idea, such as $\theta_i \beta_j$. 
Find \( \hat{\mu} \) (grand mean)

*(knnl888.sas)*

```
proc glm data=carins;
  model premium=;
  output out=overall p=muhat;
proc print data=overall;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>premium</th>
<th>size</th>
<th>region</th>
<th>muhat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>140</td>
<td>1</td>
<td>1</td>
<td>175</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>1</td>
<td>2</td>
<td>175</td>
</tr>
<tr>
<td>3</td>
<td>210</td>
<td>2</td>
<td>1</td>
<td>175</td>
</tr>
<tr>
<td>4</td>
<td>180</td>
<td>2</td>
<td>2</td>
<td>175</td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>3</td>
<td>1</td>
<td>175</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
<td>3</td>
<td>2</td>
<td>175</td>
</tr>
</tbody>
</table>
Find $\hat{\mu}_A$ (treatment means)

```
proc glm data=carins;
    class size;
    model premium=size;
    output out=meanA p=muhatA;
proc print data=meanA;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>premium</th>
<th>size</th>
<th>region</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>140</td>
<td>1</td>
<td>1</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>1</td>
<td>2</td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td>210</td>
<td>2</td>
<td>1</td>
<td>195</td>
</tr>
<tr>
<td>4</td>
<td>180</td>
<td>2</td>
<td>2</td>
<td>195</td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>3</td>
<td>1</td>
<td>210</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
<td>3</td>
<td>2</td>
<td>210</td>
</tr>
</tbody>
</table>
```
Find $\hat{\mu}_B$ (treatment means)

```
proc glm data=carins;
  class region;
  model premium=region;
  output out=meanB p=muhatB;
proc print data=meanB;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>premium</th>
<th>size</th>
<th>region</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>140</td>
<td>1</td>
<td>1</td>
<td>190</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>1</td>
<td>2</td>
<td>160</td>
</tr>
<tr>
<td>3</td>
<td>210</td>
<td>2</td>
<td>1</td>
<td>190</td>
</tr>
<tr>
<td>4</td>
<td>180</td>
<td>2</td>
<td>2</td>
<td>160</td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>3</td>
<td>1</td>
<td>190</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
<td>3</td>
<td>2</td>
<td>160</td>
</tr>
</tbody>
</table>
Combine and Compute

data estimates;
   merge overall meanA meanB;
   alpha = muhatA - muhat;
   beta = muhatB - muhat;
   atimesb = alpha*beta;
proc print data=estimates;
   var size region alpha beta atimesb;

Obs   size  region  alpha  beta  atimesb
1    1      1       -55     15    -825
2    1      2       -55    -15     825
3    2      1       20      15     300
4    2      2       20    -15    -300
5    3      1       35      15     525
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>35</td>
<td>−15</td>
<td>−525</td>
<td></td>
</tr>
</tbody>
</table>
```r
proc glm data=estimates;
  class size region;
  model premium=size region atimesb/solution;

  .

  Sum of
  Source     DF    Squares  Mean Square   F Value  Pr > F
  Model       4  10737.09677   2684.27419    208.03    0.0519
  Error       1   12.90323    12.90323
  Corrected Total  5  10750.00000

  R-Square  Coeff Var   Root MSE  premium Mean
  0.998800   2.052632   3.592106   175.0000

  Source     DF  Type I SS    Mean Square    F Value  Pr > F
  size       2  9300.000000   4650.000000    360.37    0.0372
  region     1  1350.000000   1350.000000    104.62    0.0620
  atimesb    1   87.096774    87.096774     6.75      0.2339

Latest update April 5, 2016
```
| 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 |

The test for $a \times b$ is testing $H_0 : \theta = 0$, which is not rejected. According to this, the interaction is not significant. Notice the increased $p$-values on the main effect tests, because we used up a df.