Regression Approach

- We can use multiple regression to produce results based on the factor effects model
  \[ Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \]

- Consider the restriction \( \sum \tau_i = 0 \)
- Because of this restriction, there are \( r - 1 \) regression coefficients /parameters
  \[ \sum \tau_i = 0 \rightarrow \tau_r = -\tau_1 - \tau_2 - \ldots - \tau_{r-1} \]
- Will use the following indicator variables
  \[
  X_{ijk} = \begin{cases} 
  1 & \text{if } i = k \\
  -1 & \text{if } i = r \\
  0 & \text{otherwise}
  \end{cases}
  \]

Example Page 685

- Kenton Food Company wants to test four different package designs for a new breakfast cereal
- Twenty “similar” stores were selected to be part of the experiment
- Package designs randomly and equally assigned to stores. Fire hit one store so it was dropped
- Since \( n_i \) not constant, the grand mean is not equal to the mean of the group means. Estimate of \( \mu \) based on
  \[ \mu = \frac{\sum n_i \mu_i}{n_T} \]
SAS Commands

proc means data=a1 noprint;
   class design;
   var cases;
   output out=a2 mean=mclass;
proc print data=a2;
proc means data=a2 mean;
   where _TYPE_ eq 1;
   var mclass;
run;

proc print data=a2;proc reg data=a1;
   model cases=x1 x2 x3;
run;
proc glm data=a1;
   class design;
   model cases=design / xpx inverse solution;
run;
### Output

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>588.22105</td>
<td>196.07368</td>
<td>18.59</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>158.20000</td>
<td>10.54667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>18</td>
<td>746.42105</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Parameter Estimates

| Parameter | DF | Estimate | Standard Error | t Value | Pr > |t| |
|-----------|----|----------|----------------|---------|------|---|
| Intercept | 1  | 18.67500 | 0.74853        | 24.95   | <.0001 | |
| x1        | 1  | -4.07500 | 1.27081        | -3.21   | 0.0059 | |
| x2        | 1  | -5.27500 | 1.27081        | -4.15   | 0.0009 | |
| x3        | 1  | 0.82500  | 1.37063        | 0.60    | 0.5562 | |

Notice that 18.675 is the mean of the means and 18.675-4.075=14.6, 18.675-5.275=13.4, 18.675+0.825=19.5, and 18.675+4.075+5.275-0.825=27.2, the treatment means. The same output we get from proc glm shown on the next page.

### Output

#### Sum of

<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>588.22105</td>
<td>196.073682</td>
<td>18.59</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>158.20000</td>
<td>10.546667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>18</td>
<td>746.42105</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Type III SS

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>design</td>
<td>588.22105</td>
<td>196.073682</td>
<td>18.59</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

#### Coeff Var | Root MSE | cases Mean

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Coeff Var</th>
<th>Root MSE</th>
<th>cases Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.788055</td>
<td>17.43042</td>
<td>3.247563</td>
<td>18.63158</td>
</tr>
</tbody>
</table>

#### Source | Type I SS | Mean Square | F Value | Pr > F |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>design</td>
<td>588.22105</td>
<td>196.073682</td>
<td>18.59</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>design</td>
<td>588.22105</td>
<td>196.073682</td>
<td>18.59</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

### SAS Regression Approach

- Constructs the following $r$ indicator variables
  \[ X_{ijk} = \begin{cases} 
  1 & \text{if } i = k \\
  0 & \text{otherwise}
  \end{cases} \]

- Because of the intercept (column of 1’s) there is complete dependence ($X’X$ doesn’t have an inverse)

1 = $c_1 X_1 + c_2 X_2 + ... + c_r X_r$

- SAS computes *generalized inverse* in its place

- Many generalized inverses each corresponding to a different constraint (constraint here is $\tau_r = 0$)
Output

<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>598.2210526</td>
<td>196.0736842</td>
<td>18.59</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>158.2000000</td>
<td>10.5466667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>18</td>
<td>746.4210526</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard Parameter Estimate Error t Value Pr > |t|
Intercept 27.20000000 B 1.45235441 18.73 <.0001
design 1 -12.60000000 B ... -7.70000000 B 2.17853162 -3.53 0.0030
Design 4 0.00000000 B . .

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Interpretation

- Generalized Inverse Matrix of the form

\[
\begin{bmatrix}
(X'X)^{-1} & (X'X)^{-1}X'Y \\
Y'X(X'X)^{-1} & Y'Y - Y'X(X'X)^{-1}X'Y
\end{bmatrix}
\]

- Parameter estimates in upper right corner and SSE in lower right corner
- The intercept is estimated by the mean in group 4 and the other \(b_i\)'s are the differences between the means of group \(i\) and group 4

Background Reading

- KNNL Section 16.3
- knnl686.sas
- KNNL Sections 17.1 - 17.8