DESIGN CONCEPTS:

BLOCKS
REPLICATION
EXPERIMENTAL UNITS
RANDOM VERSUS FIXED EFFECTS

TREATMENT DESIGNS (PLANS) VS. EXPERIMENTAL DESIGNS

Outline:
Blocked designs
  Random Block Effects
  REML analysis
  Incomplete Blocks
Screening Designs
  Fractional Factorials
  Aliasing
  Confounding
Response Surfaces
  Quadratic Functions
  Central Composite Designs
  Searching for Optima
Extra topics – if time allows
Nested Designs
Split Plots & Repeated Measures

Blocks- collections of experimental units that are homogeneous. Examples: regions, customer clusters, individual customers.

Design (Randomized Complete Block or RCB)
Blocks have as many units in them as we have treatments and all treatments are randomly assigned to the units within each block. Blocks are usually random effects (definition below)

Example.1: A grocery chain wants to look at sales as a function of promotion type. They want to test three particular promotion types. They think there may be regional effects but can’t afford to test these out in all of their regions so they pick 5 regions at random, 3 stores per region, and assign each of the three promotion types randomly to the 3 stores. The response is sales in the promotion period.

Model: $Y_{ij} = \mu + \alpha_i + B_j + e_{i,j}$

$\mu$: a baseline value applying to all stores.
$\alpha_i$: the effect of promotion $i$, $i=1, 2, 3$. This is a fixed effect.

$B_j$: the effect of region $j$, $j = 1,2,3,4,5$. This is a random effect.

$e_{i,j}$: the error term for the store in region $j$ that got assigned promotion $i$

**Random versus fixed effects.**

**Fixed**: Inference is only for the levels we used (3 promotion types)

**Random**: Inference is for all levels, whether chosen for the experiment or not (regions in our case)

**Fixed**: $\alpha_i$ are fixed but unknown constants.

**Random**: $B_j \sim N(0,\sigma^2_B)$ for all $j$ (regions), not just the ones we selected.

**Fixed**: Would be the same if we ran the experiment again.

**Random**: Might be different in another run (randomly picking regions, we’d likely not get these same ones again).
Example 2: We give customers different offers (5 possibilities) from time to time with each customer getting exposed to all 5 offers. Response is spending as a result of the offer.

Are offers random? No. We do NOT claim that our results tell us about other types of offers besides these 5. These were not randomly selected from an infinite (or very large) population of offers but rather, were purposefully selected.

Are customers random? Yes. We would not want to say that our inference is restricted to only those customers that were selected from our customer database for the experiment, and we should have selected them at random!

Does it matter? It all depends … No, for the balanced case (each customer got all 5 promotions) here. Yes, if somehow some customers did not get all 5 offers during the test period.

Example 3: In clinical trials the blocks might be cooperating doctors and each doctor assigns the
three drugs under investigation randomly to three patients. Doctors are the blocks, drugs the treatments, and patients are the experimental units.

Example 4: We have 3 different machines for performing an outdoor task. We block on days because weather, traffic, pollution, etc. can vary from day to day. On each trial day (we might want to have days in different seasons), we assign 3 workers to the three machines and measure some response like time needed to complete the task or quality of the result (or both). For simplicity we’ll assume different sets of 3 workers for each test day.

Example 5: We select 12 delivery people from our 2000 delivery people nationwide, and 3 types of trucks on loan. We give 4 of them type A trucks, 4 type B, and 4 type C. We let them drive these for 3 months and measure miles per gallon. We hope to pick one of the types and get a deal on a volume order of trucks.
This design is NOT blocked. It is called a completely randomized design.

Data for example 5:

\( (\overline{Y}_* = \text{overall mean} = 15 ) \)

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>D</td>
<td>Y(_{ij})</td>
<td>(\overline{Y}_i)</td>
<td>((\overline{Y}<em>i - \overline{Y}</em>* )^2)</td>
<td>((Y_{ij} - \overline{Y}_*)^2)</td>
<td>((Y_{ij} - \overline{Y}_* )^2)</td>
<td></td>
</tr>
<tr>
<td>MPG</td>
<td>truck</td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>12.6</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>-3.4</td>
<td>11.56</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>14.2</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>-1.8</td>
<td>3.24</td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>17.4</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1.4</td>
<td>1.96</td>
</tr>
<tr>
<td>A</td>
<td>4</td>
<td>19.8</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>3.8</td>
<td>14.44</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>11.0</td>
<td>14</td>
<td>-1</td>
<td>1</td>
<td>-3.0</td>
<td>9.00</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>13.8</td>
<td>14</td>
<td>-1</td>
<td>1</td>
<td>-0.2</td>
<td>0.04</td>
</tr>
<tr>
<td>B</td>
<td>7</td>
<td>14.6</td>
<td>14</td>
<td>-1</td>
<td>1</td>
<td>0.6</td>
<td>0.36</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>16.6</td>
<td>14</td>
<td>-1</td>
<td>1</td>
<td>2.6</td>
<td>6.76</td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>12.4</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>-2.6</td>
<td>6.76</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>14.0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>-1.0</td>
<td>1.00</td>
</tr>
<tr>
<td>C</td>
<td>11</td>
<td>16.0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>1.00</td>
</tr>
<tr>
<td>C</td>
<td>12</td>
<td>17.6</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>2.6</td>
<td>6.76</td>
</tr>
</tbody>
</table>

\[
\text{Total Sum of Squares} = \sum_{i=1}^{3} \sum_{j=1}^{4} (Y_{ij} - \overline{Y}_*)^2 = 70.88, \ 12-1 = 11 df
\]

\[
\text{Treatment Sum of Squares} = \sum_{i=1}^{3} \sum_{j=1}^{4} (\overline{Y}_i - \overline{Y}_*)^2 = 4 \sum_{i=1}^{3} (\overline{Y}_i - \overline{Y}_*)^2 = 8, \ 3-1 = 2 df
\]
Error Sum of Squares = Total SS – Treatment SS = \[\sum_{i=1}^{3} \sum_{j=1}^{4} (Y_{ij} - \bar{Y}_i)^2 = 62.88,\ 11 - 2 = 9 \text{ df}\ (3 \text{ from each group})\]

Mean Square = (Sum of Squares)/df

\[F = \frac{(\text{Treatment Mean Square})}{(\text{Error Mean Square})}\]

Analysis of Variance (ANOVA)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SSq</th>
<th>MnSq</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truck</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>4/6.987 =0.57</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>62.88</td>
<td>6.987</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>70.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model: MPG = mean + truck effect + error

\[Y_{ij} = \mu + \tau_i + e_{ij}\]

(1) Matrix Formulation – the modern approach
\[
Y = \begin{pmatrix} 12.6 \\ 11.0 \\ 12.4 \\ 14.2 \\ 13.8 \\ 14.0 \\ 17.4 \\ 14.6 \\ 16.0 \\ 19.8 \\ 16.6 \\ 17.6 \end{pmatrix}, \quad X = \begin{pmatrix} 1100 \\ 1010 \\ 1001 \\ 1100 \\ 1010 \\ 1001 \\ 1100 \\ 1010 \\ 1001 \\ 1100 \\ 1010 \\ 1001 \end{pmatrix}, \quad Y = \begin{pmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{24} \\ e_{31} \\ e_{32} \\ e_{33} \end{pmatrix}
\]

\[
Y = X\beta + e
\]

Assuming \(e_{ij} \sim N(0, \sigma^2)\) this is just regression!

\[
X'X = \begin{pmatrix} 12 & 4 & 4 & 4 \\ 4 & 4 & 0 & 0 \\ 4 & 0 & 4 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix}
\]

\[
(X'X)^{-1} = ?
\]

How to solve \((X'X)\hat{\beta} = (X'Y)\) ? Many solutions!

Problem – can’t compute inverse. Back to square 1! Reformulate the equation to give
Omit last column of $X$:

$$
\begin{pmatrix}
  12.6 \\
  11.0 \\
  12.4 \\
  14.2 \\
  13.8 \\
  14.0 \\
  17.4 \\
  14.6 \\
  16.0 \\
  19.8 \\
  16.6 \\
  17.6
\end{pmatrix}
= 
\begin{pmatrix}
  110 \\
  101 \\
  100 \\
  110 \\
  101 \\
  100 \\
  110 \\
  101 \\
  100 \\
  110 \\
  101 \\
  100
\end{pmatrix}
+ 
\begin{pmatrix}
e_{11} \\
e_{12} \\
e_{13} \\
e_{14} \\
e_{21} \\
e_{22} \\
e_{23} \\
e_{24} \\
e_{31} \\
e_{32} \\
e_{33} \\
e_{34}
\end{pmatrix}

Y = X\beta + e

Checks:

- lines 3, 6, 9, 12: $Y = \mu + \tau_3 + 0 + 0 + e$ OK
- lines 1, 4, 7, 10: $Y = \mu + \tau_3 + (\tau_1 - \tau_3) + 0 + e = \mu + \tau_1 + 0 + e$ OK
- lines 2, 5, 8, 11: $Y = \mu + \tau_3 + (\tau_2 - \tau_3) + 0 + e = \mu + \tau_2 + 0 + e$ OK

**Can** compute this one!

$$
XX = \begin{pmatrix}
  12 & 4 & 4 \\
  4 & 4 & 0 \\
  4 & 0 & 4
\end{pmatrix}
(XX)^{-1} = \frac{1}{4}
\begin{pmatrix}
  1 & -1 & -1 \\
  -1 & 2 & 1 \\
  -1 & 1 & 2
\end{pmatrix}
\text{(can you check this?)}

$$
XY = \begin{pmatrix}
  1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
  1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
  0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{pmatrix}
= \begin{pmatrix}
  \sum_{i=1}^{3} \sum_{j=1}^{4} Y_{ij} \\
  \sum_{j=1}^{4} Y_{1j} \\
  \sum_{j=1}^{4} Y_{2j}
\end{pmatrix}
= \begin{pmatrix}
  12.6 \\
  11.0 \\
  12.4 \\
  14.2 \\
  13.8 \\
  14.0 \\
  17.4 \\
  14.6 \\
  16.0 \\
  19.8 \\
  16.6 \\
  17.6
\end{pmatrix}
= \begin{pmatrix}
  180 \\
  64 \\
  56
\end{pmatrix}
Sample means – truck type A 16, truck type B 14, truck type C 15.

\[
\hat{\beta} = \begin{pmatrix}
\mu + \tau_3 \\
\tau_1 - \tau_3 \\
\tau_2 - \tau_3
\end{pmatrix} = (X'X)^{-1}(X'Y) = \frac{1}{4} \begin{pmatrix}
1 & -1 & -1 \\
-1 & 2 & 1 \\
-1 & 1 & 2
\end{pmatrix} \begin{pmatrix}
180 \\
64 \\
56
\end{pmatrix} = \frac{1}{4} \begin{pmatrix}
60 \\
4 \\
-4
\end{pmatrix} = \begin{pmatrix}
15 \\
1 \\
-1
\end{pmatrix} = \begin{pmatrix}
C_{\text{mean}} \\
A_{\text{mean}} - C_{\text{mean}} \\
B_{\text{mean}} - C_{\text{mean}}
\end{pmatrix}
\]

*** Run program DOE1.sas ***

(2) Discussion:
Suppose \( \mu + \tau_1 = 17, \mu + \tau_2 = 16, \mu + \tau_3 = 12 \)

Even \textit{knowing} these, we cannot uniquely find \( \mu, \tau_1, \tau_2, \tau_3 \), let alone when these are estimated by the sample means.

For example \( \mu = 3, \tau_1 = 14, \tau_2 = 13, \tau_3 = 9 \) would work as would \( \mu = 15, \tau_1 = 2, \tau_2 = 1, \tau_3 = -3 \) (\textit{“effects”} or \textit{“deviations”} solution) or the \textit{“reference cell”} or \textit{“GLM”} solution \( \mu = 12, \tau_1 = 5, \tau_2 = 4, \tau_3 = 0 \)

Using observed sample means (all we really have) we estimate \( \mu + \tau_1 \) as 16, \( \mu + \tau_2 \) as 14, and \( \mu + \tau_3 \) as 15 giving GLM (reference cell) solutions 15, 1, -1. We can estimate \( \mu + \tau_1, \mu + \tau_2, \) and \( \mu + \tau_3 \) and any linear combination of them, like \( \tau_1 - (\tau_2 + \tau_3)/2 \)

\begin{equation}
(1.5 \text{ mpg})
\end{equation}
Demo shows no truck effect, lot of driver effects (captured in error term).

New problem, same data (for illustration).
Now suppose each of 4 drivers drove all 3 types of truck. Truck comparison like $\tau_1 - (\tau_2 + \tau_3)/2$ will be free of driver effects. If driver 1 has effect -3 for all trucks, we see that $(\tau_1 -3) - (\tau_2 -3+\tau_3-3)/2 = \tau_1 - (\tau_2 +\tau_3)/2 :$ driver effect cancels out.

New Model:
\[
\text{MPG} = \text{mean} + \text{truck effects} + \text{driver effects} + \text{error}
\]
\[
Y_{ij} = \mu + \tau_i + D_j + e_{ij}
\]

Assumptions: $D_j \sim \text{N}(0, \sigma^2_D)$, $e_{ij} \sim \text{N}(0, \sigma^2)$

Implications:
\[
\overline{Y}_i - \overline{Y}_s = \tau_i - \tau_s + \overline{D} - \overline{D} + \overline{e}_i - \overline{e}_s
\]
\[
\overline{Y}_i - \overline{Y}_s = \tau_i - \tau_s + \overline{e}_i - \overline{e}_s
\]

Expected value: $E\{\overline{Y}_i - \overline{Y}_s = \tau_i - \tau_s\}$

Variance (variance of $\overline{e}_i - \overline{e}_s$) $\sigma^2 / n_i + \sigma^2 / n_s$

**********Using the “three rules”**********
********** The “three rules” **********

(1) Variance of a mean is $\sigma^2 / n$

(2) Variance of a sum or difference of independent variables is sum of variances.

(3) Variance of $C$ (constant) times a variable is $C^2$ times variance of original variable.

Fixed block (driver) effects (unusual):

$$
\begin{pmatrix}
12.6 \\
11.0 \\
12.4 \\
14.2 \\
13.8 \\
14.0 \\
17.4 \\
14.6 \\
16.0 \\
19.8 \\
16.6 \\
17.6
\end{pmatrix}
= 
\begin{pmatrix}
110 & 100 \\
101 & 100 \\
100 & 100 \\
110 & 010 \\
101 & 010 \\
100 & 010 \\
110 & 001 \\
101 & 001 \\
100 & 001 \\
110 & 000 \\
101 & 000 \\
100 & 000
\end{pmatrix}
\begin{pmatrix}
\mu + \tau_3 + D_4 \\
\tau_1 - \tau_3 \\
\tau_2 - \tau_3 \\
D_1 - D_4 \\
D_2 - D_4 \\
D_3 - D_4
\end{pmatrix}
+ 
\begin{pmatrix}
e_{11} \\
e_{12} \\
e_{13} \\
e_{14} \\
e_{21} \\
e_{22} \\
e_{23} \\
e_{24} \\
e_{31} \\
e_{32} \\
e_{33} \\
e_{34}
\end{pmatrix}

Y = X\beta + e

Sample means: $\bar{Y}_i = \mu + \tau_i + \bar{D} + \bar{e}_i$.  
Expected value $E\{\bar{Y}_i\} = \mu + \tau_i + \bar{D}$, variance $\sigma^2/4$
Random block (driver) effects (usual):

\[
\begin{bmatrix}
12.6 & 11.0 & 12.4 & 14.2 & 13.8 & 14.0 & 17.4 & 14.6 & 16.0 & 19.8 & 16.6 & 17.6 \\
110 & 101 & 100 & 110 & 101 & 100 & 110 & 101 & 100 & 110 & 101 & 100
\end{bmatrix}
\begin{bmatrix}
\mu + \tau_3 \\
\tau_1 - \tau_3 \\
\tau_2 - \tau_3
\end{bmatrix}
+ 
\begin{bmatrix}
D_1 & D_2 & D_3 & D_4
\end{bmatrix}
+ 
\begin{bmatrix}
e_{11} \\
e_{12} \\
e_{13} \\
e_{14} \\
e_{21} \\
e_{22} \\
e_{23} \\
e_{24} \\
e_{31} \\
e_{32} \\
e_{33} \\
e_{34}
\end{bmatrix}
= 
Y = X\beta + Z\gamma + e
\]

Assumptions: $\gamma \sim N(0, G)$  $e \sim N(0, R)$

Our example:

\[
G = \begin{bmatrix}
\sigma_D^2 & 0 & 0 & 0 \\
0 & \sigma_D^2 & 0 & 0 \\
0 & 0 & \sigma_D^2 & 0 \\
0 & 0 & 0 & \sigma_D^2
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\sigma_D^2 = I\sigma_D^2 \\
R = I\sigma^2
\]

Sample means: $\bar{Y}_{i*} = \mu + \tau_i + \bar{D} + \bar{e}_{i*}$

Expected value: $E\{\bar{Y}_{i*}\} = \mu + \tau_i$, \hspace{1cm} (not $\mu + \tau_i + \bar{D}$)

Variance (of $\bar{D} + \bar{e}_{i*}$) is $\sigma_D^2 / 4 + \sigma^2 / 4$, \hspace{1cm} (not $\sigma^2 / 4$)

*** Run program DOE2.sas ***

*** What is standard error for A mean (LSMEANS) ?? ***

*** How does this relate to formulas above??***
REML estimation:

Data (12, 18, 13, 21) mean 16  SSq=16+4+9+25=54
Simple model Y~N(μ,σ²)

Log likelihood:

\[-\frac{n}{2} \ln(2πσ²) - \sum_{i=1}^{n} \frac{(Y_i - μ)^2}{2σ²}\]

Set derivative w.r.t. μ to 0

\[-2\sum_{i=1}^{n} \frac{(Y_i - μ)}{2σ²} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} (Y_i - μ) = 0 \quad \Rightarrow \quad μ_{MLE} = \bar{Y}\]

“maximum likelihood estimate or MLE.”

Likelihood becomes

\[-\frac{n}{2} \ln(2π) - \frac{n}{2} \ln(σ²) - \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{2σ²} = -\frac{n}{2} \ln(2π) - \frac{n}{2} \ln(σ²) - \frac{54}{2σ²}\]

Set derivative w.r.t σ² to 0:

\[0 - \frac{n}{2σ²} + \frac{54}{2(σ²)^2} = 0 \quad \Rightarrow \quad -nσ² + 54 = 0 \quad \Rightarrow \quad σ_{MLE}² = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})²}{n} = 54 / n\]
This \( \sigma_{MLE}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \) is biased downward because
\[ \bar{Y} \neq \mu_{TRUE} \], expected value is \( \frac{n-1}{n} \sigma^2 \)

\[
\begin{pmatrix}
.25 & .25 & .25 & .25 \\
.5 & .5 & -.5 & -.5 \\
.5 & -.5 & .5 & -.5 \\
.5 & -.5 & -.5 & .5
\end{pmatrix}
\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} \sim N \left( \begin{pmatrix} \mu \\ 4(1/16) \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right) \sigma^2 \text{ where we used}
\]

\[
\begin{pmatrix}
.25 & .25 & .25 & .25 \\
.5 & .5 & -.5 & -.5 \\
.5 & -.5 & .5 & -.5 \\
.5 & -.5 & -.5 & .5
\end{pmatrix}
\begin{pmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \sim N \left( \begin{pmatrix} 4(1/16) \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right) \sigma^2
\]

This shows that these “orthogonal contrasts” satisfy

\[
\begin{pmatrix}
1 & 1 & -1 & -1 \\
0.5 & 1 & -1 & 1 \\
1 & -1 & 1 & -1
\end{pmatrix}
\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \end{pmatrix} \sigma^2 \right)
\]

Three independent random variables with known mean (0). Maximize their likelihood.
\[
\begin{bmatrix}
0.5 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 \\
\end{bmatrix} \begin{bmatrix}
12 \\
18 \\
13 \\
21 \\
\end{bmatrix} = \begin{bmatrix} -2 \\ -7 \\ 1 \end{bmatrix}
\]

sum of squares \(4+49+1=54\) (same as before! – always). Now compute Maximum Likelihood Estimator (means known to be 0, 3 observations) \(\frac{54}{3} = \frac{\sum (y_i - \bar{y})^2}{n}\) which is unbiased.

General: \(Y = X\beta + Z\gamma + e\)

Multiply through by \(I - X(X'X)^{-1}X'\)

\[
( I - X(X'X)^{-1}X' )Y = ( I - X(X'X)^{-1}X' )X\beta + ( I - X(X'X)^{-1}X' )Z\gamma + ( I - X(X'X)^{-1}X' )e
\]

From this pick n-(rank of X) linearly independent linear combinations with mean 0 and maximize likelihood. This is called REML estimation. This is how PROC MIXED works. REML gives less biased (not always unbiased) variance component estimates than does unrestricted maximum likelihood.

****** end of optional mathematical material ******

Executive summary: Use PROC MIXED when you have random effects other than the error term!
General mixed models without balance – driver 3 is sick on last day when he was supposed to drive truck type B and driver 1 skips out on her truck type A run.

*** run DOE3.sas ***

GLM results – fixed effect estimates

The GLM Procedure
Least Squares Means

<table>
<thead>
<tr>
<th>truck</th>
<th>MPG LSMEAN</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>16.1714286</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>14.1714286</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>15.0000000</td>
<td>3</td>
</tr>
</tbody>
</table>

Least Squares Means for effect truck
Pr > |t| for H0: LSMean(i)=LSMean(j)

<table>
<thead>
<tr>
<th>i/j</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0429</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.1250</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.1250</td>
<td>0.2429</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.

PROC MIXED results

Least Squares Means

| Effect | truck | Estimate | Error | DF | t Value | Pr > |t| |
|--------|-------|----------|-------|----|---------|------|---|
| truck  | A     | 16.2191  | 1.3216| 4  | 12.27   | 0.0003|
| truck  | B     | 14.1469  | 1.3216| 4  | 10.70   | 0.0004|
| truck  | C     | 15.0000  | 1.2934| 4  | 11.60   | 0.0003|

Differences of Least Squares Means

| Effect | truck | _truck | Estimate | Error | DF | t Value | Pr > |t| |
|--------|-------|--------|----------|-------|----|---------|------|---|
| truck  | A     | B      | 2.0722   | 0.6799| 4  | 3.05    | 0.0381|
| truck  | A     | C      | 1.2191   | 0.6043| 4  | 2.02    | 0.1138|
| truck  | B     | C      | -0.8531  | 0.6043| 4  | -1.41   | 0.2309| (0.2429)
What are “LSMEANS?”

Treating drivers as fixed, LSMEANS are estimates of $\mu + \bar{D} + \tau_i$

$$\text{GLM: } 17.886 + (-5.771 - 1.771 + 0)/4 + 1.171 = 16.171$$

| Effect     | Estimate | Error | DF | t Value | Pr > |t| |
|------------|----------|-------|----|---------|-------|
| Intercept  | 17.8857  | 0.5525 | 32.37 | <.0001  |
| truck A    | 1.1714   | 0.6052 | 1.94  | 0.1250  |
| truck B    | -0.8286  | 0.6052 | -1.37 | 0.2429  |
| truck C    | 0.0000   | .      | .     | .       |
| driver 1   | -5.7714  | 0.7223 | -7.99 | 0.0013  |
| driver 2   | -4.0000  | 0.6234 | -6.42 | 0.0030  |
| driver 3   | -1.7714  | 0.7223 | -2.45 | 0.0703  |
| driver 4   | 0.0000   | .      | .     | .       |

Treating drivers as random, LSMEANS are estimates of $\mu + \tau_i$

Why? Random effect model assumes $E\{D\}=0$, no need to adjust by sample average of these particular 4 drivers.

$$\text{MIXED: } 15.0000 + 1.2191 = 16.219$$

| Effect   | Estimate | Error | DF | t Value | Pr > |t| |
|----------|----------|-------|----|---------|-------|
| Intercept| 15.0000  | 1.2934 | 3  | 11.60   | 0.0014 |
| truck A  | 1.2191   | 0.6043 | 4  | 2.02    | 0.1138 |
| truck B  | -0.8531  | 0.6043 | 4  | -1.41   | 0.2309 |
| truck C  | 0.0000   | .      | .  | .       | .     |

Why LSMEANS are better than MEANS
(1) For balanced data they are the same
(2) Salary data:

<table>
<thead>
<tr>
<th></th>
<th>Executives</th>
<th>Workers</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>200</td>
<td>50, 48, 49, 53</td>
<td>80</td>
</tr>
<tr>
<td>Females</td>
<td>170, 190, 180</td>
<td>28, 32</td>
<td>120</td>
</tr>
</tbody>
</table>

Conclusion: Females make $40\,K$ more than males.
Table of means below: means of means are LSMEANS for data like these. LSMEANS use the estimates from any data, balanced or not, to estimate what means would have been if the data had been balanced and all continuous variables (like baseline blood pressure, age) were set at the average observed value.

<table>
<thead>
<tr>
<th></th>
<th>Executives</th>
<th>Workers</th>
<th>LSMEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>200</td>
<td>50</td>
<td>125</td>
</tr>
<tr>
<td>Females</td>
<td>180</td>
<td>30</td>
<td>105</td>
</tr>
</tbody>
</table>

LSMEANS show males make more than females

***** run DOE4.sas *******

Factorial Treatment Arrangements:

Room Price: 120, 140, 160, 180  
Location: City, Resort, Other

Problem: higher price $\rightarrow$ lower occupancy $\rightarrow$ lower profit  
Problem: lower price with no occupancy change $\rightarrow$ lower profit

Y= profit

Twelve treatment combinations, 3 hotels in each category  
Price and location are fixed effects.

Q1: Does optimal pricing depend on location?  
Q2: Averaged over price, how do location profits compare?
### Table of Profit Means (of 3 hotels each):

<table>
<thead>
<tr>
<th>Location</th>
<th>Price $120</th>
<th>Price $140</th>
<th>Price $160</th>
<th>Price $180</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>2557.33</td>
<td>3722.33</td>
<td>4721.67</td>
<td>4083.00</td>
<td>3771.08</td>
</tr>
<tr>
<td>Other</td>
<td>3981.33</td>
<td>3879.67</td>
<td>2553.67</td>
<td>-77.33</td>
<td>2584.33</td>
</tr>
<tr>
<td>Resort</td>
<td>4846.00</td>
<td>5327.67</td>
<td>5388.33</td>
<td>4763.33</td>
<td>5081.33</td>
</tr>
<tr>
<td><strong>Means</strong></td>
<td>4221.22</td>
<td>4309.89</td>
<td>3794.89</td>
<td>2923.00</td>
<td>3812.25</td>
</tr>
</tbody>
</table>

\[
\text{SS(table)} = 3 \left( (2557.33 - 3812.25)^2 + \ldots + (4763.33 - 3812.25)^2 \right) = 77,948,984.08 \quad (11 \text{ df})
\]

\[
\text{SS(price)} = 9 \left( (4221.22 - 3812.25)^2 + \ldots + (2923.00 - 3812.25)^2 \right) = 10,853,727.42 \quad (3 \text{ df})
\]

\[
\text{SS(location)} = 12 \left( (3771.08 - 3812.25)^2 + \ldots + (5081.33 - 3812.25)^2 \right) = 37,440,558.50 \quad (2 \text{ df})
\]

\[
\text{SS(PxL interaction)} = \text{SS(table)} - \text{SS(location)} - \text{SS(price)} = 29,654,698.17 \quad (11 - 3 - 2 = 6 \text{ df})
\]

### Splitting price effects into linear, quadratic, cubic effects

1. (optional – by hand)
   Locate table of “orthogonal polynomial coefficients” \( c_{ij} \) for 4 levels (e.g. linear: \( c_{11} = -3, c_{12} = -1, c_{13} = 1, c_{14} = 3 \))

<table>
<thead>
<tr>
<th>Location</th>
<th>Effect</th>
<th>Denominator</th>
<th>SSq=(Effect)^2/den</th>
</tr>
</thead>
<tbody>
<tr>
<td>means</td>
<td>4221.22</td>
<td>4309.89</td>
<td>3794.89</td>
</tr>
<tr>
<td>linear</td>
<td>-3</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>quadratic</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>cubic</td>
<td>-1</td>
<td>3</td>
<td>-3</td>
</tr>
</tbody>
</table>

Compute \( \text{effect}_i = \sum_{j=1}^{4} c_{ij} Y_j \) and denominator: \( \sum_{j=1}^{4} c_{ij}^2 / n \) (n obs per mean)

Sum of squares is \( \text{effect}^2/den \)

2. Optional: How did they get those orthogonal polynomial coefficients?

\[
\begin{pmatrix}
  1 & 1 & 1 & 1 \\
  1 & 2 & 4 & 8 \\
  1 & 3 & 9 & 27 \\
  1 & 4 & 16 & 64
\end{pmatrix}
\]

Regress each column on the ones to its left. Replace each column by its residuals (multiplied by a constant to make the entries integers). New columns contain orthogonal polynomial coefficients.
Details: Regress column 2 (call it Y) on column 1 (treat it as X).

\[ X = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad X'X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = 4 \quad b = (X'X)^{-1}X'Y = 4^{-1} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = 2.5 \]

\[ \text{residual } = Y - Xb = \begin{pmatrix} 1/3 \\ 2/3 \\ 4/3 \end{pmatrix} - 2.5 \begin{pmatrix} 1/2 \\ 1 \\ 1/2 \end{pmatrix} = \begin{pmatrix} -3/3 \\ -1/3 \\ 1/3 \end{pmatrix} \]

Gives linear coefficients \( \begin{pmatrix} -3/3 \\ -1/3 \\ 1/3 \end{pmatrix} \).

Quadratic: Regress column 3 on column 1 and column 2.

\[ X = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -3 \\ -1 \\ 1 \\ 1 \\ 1 \\ 3 \end{pmatrix} \quad X'X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & -3 & -1 \\ 1 & -3 & -1 \\ 1 & -3 & -1 \\ 1 & -3 & -1 \\ 1 & 3 & 1 \end{pmatrix} = 4 \quad b = (X'X)^{-1}X'Y = \begin{pmatrix} 1/4 \\ 0 \\ 1/20 \end{pmatrix} \begin{pmatrix} 1/4 \\ 0 \\ 1/20 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 9 \\ 16 \end{pmatrix} = \begin{pmatrix} 1/4 \\ 0 \\ 1/20 \end{pmatrix} \begin{pmatrix} 30 \\ 50 \end{pmatrix} = \begin{pmatrix} 7.5 \\ 2.5 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 5 \end{pmatrix} \]

\[ \text{residual } = Y - Xb = \begin{pmatrix} 1 \\ 4 \\ 9 \\ 16 \end{pmatrix} - \begin{pmatrix} 1 \\ 4 \\ 9 \\ 16 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ -1 \\ -1 \end{pmatrix} \]

Cubic: Regress column 4 on others.

\[ b = (X'X)^{-1}X'Y = \begin{pmatrix} 1/4 \\ 0 \\ 1/20 \\ 0 \\ 1/4 \\ 0 \\ 0 \\ 1/4 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -3 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1/4 \\ 0 \\ 1/20 \\ 0 \\ 1/4 \\ 0 \\ 0 \\ 1/4 \end{pmatrix} \begin{pmatrix} 100 \\ 208 \end{pmatrix} = \begin{pmatrix} 25 \\ 10.4 \end{pmatrix} \]

\[ \text{residual } = Y - Xb = \begin{pmatrix} 1 \\ 8 \\ 27 \\ 64 \end{pmatrix} - \begin{pmatrix} 1 \\ 8 \\ 27 \\ 64 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1.3 \\ -0.3 \\ 1.3 \end{pmatrix} = \begin{pmatrix} 0 \\ -0.9 \\ 27.9 \\ -0.9 \end{pmatrix} = \begin{pmatrix} 3 \\ -3 \end{pmatrix} \]

**** Run DOE5.sas ****
The $2^N$ experiment:

Several (N) factors, each at 2 levels.

Simple example $2^2$:

Gender: Female, Male  code as $X_G = -1, 1$
Dose: 5, 75  code as $X_D = -1, 1$

Ten clinics, 4 patients each. Measure side effect: change in blood pressure (after drug – before drug).

Means of 10:

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>(Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dose 5</td>
<td>-2.00</td>
<td>4.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Dose 75</td>
<td>-24.00</td>
<td>26.00</td>
<td>1.00</td>
</tr>
<tr>
<td>(Mean)</td>
<td>-13.00</td>
<td>15.00</td>
<td>(1.00 overall)</td>
</tr>
</tbody>
</table>

Discussion:

If we look at the average response at the two doses, it is first of all very small (1.00) and second, it is the same at both doses (we say there is no *main effect* of dose) which is interesting in light of the fairly large and potentially harmful effects within the females and within the males. These we call the *simple effects* of dose within the males and females. Note that it would be dangerous and incorrect to say that there is a negligible (1 point) change in blood pressure when dose is increased from 5 to 75.
Solution: See if there is interaction. If so, report simple effects. We define and discuss interactions and simple effects but first -

** Run DOE6.sas **

Means of 10:

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>(Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dose 5</td>
<td>-2.00</td>
<td>4.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Dose 75</td>
<td>-24.00</td>
<td>26.00</td>
<td>1.00</td>
</tr>
<tr>
<td>(Mean)</td>
<td>-13.00</td>
<td>15.00</td>
<td>1.00 overall</td>
</tr>
</tbody>
</table>

Simple effect of Dose within Females

\((-1)(-2.00) + (1)(-24.00) + (0)(4.00) + (0)(26.00) = -22\)

Simple effect of Dose within Males:

\((0)(-2.00) + (0)(-24.00) + (-1)(4.00) + (1)(26.00) = 22\)

Simple effect of gender at low dose

\((-1)(-2.00) + (0)(-24.00) + (1)(4.00) + (0)(26.00) = 6\)

Simple effect of gender at high dose

\((0)(-2.00) + (-1)(-24.00) + (0)(4.00) + (1)(26.00) = 50 !!!\)

Definition:

Interaction is \(\frac{1}{2}\) (difference of simple effects):

\(\frac{1}{2}(22 - (-22)) = \frac{1}{2}(50-6)\)

Note that one way to compute the main effect of gender is to take \(\frac{1}{2}\) (sum of the simple effects of gender)

\(\frac{1}{2}(6+50) = 28 = 15 - (-13)\)

With n observations per mean and the 1s, -1s, and 0s called “coefficients” \(c_i\), there is a one degree of freedom sum of squares for each effect. Dividing that mean square by the error mean square gives an F test with 1 numerator degree of freedom. That F is also the square of the t test for the associated coefficient.
As before, the sum of squares is:

\[
(\text{effect})^2 / \sum_{i=1}^{k} (c_i^2 / n)
\]

where \( k \) is the number of means (4 in our case) being combined. Noting that main effects and interactions have the \( \frac{1}{2} \) multiplier which the others do not, we can organize the various effects and their sums of squares in a table. Note that this is being presented to illustrate what your statistical software is computing. Further, these formulas will no longer work when the data become unbalanced. The matrix approach that SAS uses is general!

<table>
<thead>
<tr>
<th>Effect</th>
<th>D low G low</th>
<th>D low G high</th>
<th>D hi G lo</th>
<th>D hi G hi</th>
<th>effect</th>
<th>Denom ( \sum_{i=1}^{4} (c_i^2 / 10) )</th>
<th>SSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>-( \frac{1}{2} )</td>
<td>-( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} (0) )</td>
<td>0.1</td>
<td>0(^2/.1=0)</td>
</tr>
<tr>
<td>G</td>
<td>-( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
<td>-( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} (56) )</td>
<td>0.1</td>
<td>7,840</td>
</tr>
<tr>
<td>D in G_lo</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-22</td>
<td>0.2</td>
<td>2,420</td>
</tr>
<tr>
<td>D in G_hi</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>22</td>
<td>0.2</td>
<td>2,420</td>
</tr>
<tr>
<td>G in D_lo</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.2</td>
<td>20</td>
</tr>
<tr>
<td>G in D_hi</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>50</td>
<td>0.2</td>
<td>12,500</td>
</tr>
<tr>
<td>D*G intrctn</td>
<td>( \frac{1}{2} )</td>
<td>-( \frac{1}{2} )</td>
<td>-( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} (44) )</td>
<td>0.1</td>
<td>4840</td>
</tr>
</tbody>
</table>

Example: More factors
Generalizations include more factors and factors at more than 2 levels. Let’s consider a marketing example. I may have displays at the end of an aisle or in the aisle itself for my product.
(Factor E, levels -1, 1) and there may be a sign or not. My product could be displayed on a low, middle, or high level shelf (factor H, levels -1, 0, 1) and I think there may be different purchasing habits in different towns so I will block my experiment on towns, assuming a random selection of 5 towns with each having at least 12 stores (why?). Alternatively I could block on (5) stores and randomly apply each treatment for a month. How long would that experiment take?

*** Run DOE8.SAS ***

Here is the analysis of variance table using the code

```sas
proc glm data=next;
class E H sign town;
model sales = town E|H|sign;
```

<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>15</td>
<td>9026.41667</td>
<td>601.76111</td>
<td>16.75</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>44</td>
<td>1580.56667</td>
<td>35.92197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>10606.98333</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model sum of square contains the block (TOWN) sum of squares 604.23 plus the treatment sum of squares 8422.18. These sums of squares can be obtained from the town and treatment means. Each sum of squares is obtained by the familiar formula in which each mean, for example the 12 treatment means, is subtracted from the overall mean and that difference is squared and multiplied by the number of observations going into the treatment mean, 5 in our example, to give a contribution to the sum of squares. These are totaled up across the levels, for example the 12 levels of treatment
combinations, to give the sum of squares. Here are the computations for town and treatments:

_TYPE_=8 (these are town means)

<table>
<thead>
<tr>
<th>town</th>
<th>E</th>
<th>H</th>
<th>sign</th>
<th><em>FREQ</em></th>
<th>mn_sales</th>
<th>diff</th>
<th>contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>.</td>
<td>-</td>
<td>12</td>
<td>99.250</td>
<td>0.26667</td>
<td>0.85333</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>.</td>
<td>-</td>
<td>12</td>
<td>93.083</td>
<td>-5.90000</td>
<td>417.72</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>.</td>
<td>-</td>
<td>12</td>
<td>99.167</td>
<td>0.18333</td>
<td>0.40</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>.</td>
<td>-</td>
<td>12</td>
<td>101.167</td>
<td>2.18333</td>
<td>57.20</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>.</td>
<td>-</td>
<td>12</td>
<td>102.250</td>
<td>3.26667</td>
<td>128.05</td>
</tr>
<tr>
<td>------</td>
<td>---</td>
<td>---</td>
<td>------</td>
<td>--------</td>
<td>----------</td>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>604.23</td>
</tr>
</tbody>
</table>

_TYPE_=14 (these are the treatment means)

<table>
<thead>
<tr>
<th>E</th>
<th>H</th>
<th>sign</th>
<th><em>FREQ</em></th>
<th>mn_sales</th>
<th>diff</th>
<th>contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>-1</td>
<td>no</td>
<td>5</td>
<td>70.8</td>
<td>-28.1833</td>
<td>3971.50</td>
</tr>
<tr>
<td>no</td>
<td>-1</td>
<td>yes</td>
<td>5</td>
<td>89.0</td>
<td>-9.9833</td>
<td>498.33</td>
</tr>
<tr>
<td>no</td>
<td>0</td>
<td>no</td>
<td>5</td>
<td>86.0</td>
<td>-12.9833</td>
<td>842.83</td>
</tr>
<tr>
<td>no</td>
<td>0</td>
<td>yes</td>
<td>5</td>
<td>96.0</td>
<td>-2.9833</td>
<td>44.50</td>
</tr>
<tr>
<td>no</td>
<td>1</td>
<td>no</td>
<td>5</td>
<td>101.0</td>
<td>2.0167</td>
<td>20.33</td>
</tr>
<tr>
<td>no</td>
<td>1</td>
<td>yes</td>
<td>5</td>
<td>103.0</td>
<td>4.0167</td>
<td>80.67</td>
</tr>
<tr>
<td>yes</td>
<td>-1</td>
<td>no</td>
<td>5</td>
<td>104.8</td>
<td>5.8167</td>
<td>169.17</td>
</tr>
<tr>
<td>yes</td>
<td>-1</td>
<td>yes</td>
<td>5</td>
<td>102.8</td>
<td>3.8167</td>
<td>72.83</td>
</tr>
<tr>
<td>yes</td>
<td>0</td>
<td>no</td>
<td>5</td>
<td>104.8</td>
<td>5.8167</td>
<td>169.17</td>
</tr>
<tr>
<td>yes</td>
<td>0</td>
<td>yes</td>
<td>5</td>
<td>106.8</td>
<td>7.8167</td>
<td>305.50</td>
</tr>
<tr>
<td>yes</td>
<td>1</td>
<td>no</td>
<td>5</td>
<td>119.8</td>
<td>20.8167</td>
<td>2166.67</td>
</tr>
<tr>
<td>yes</td>
<td>1</td>
<td>yes</td>
<td>5</td>
<td>103.0</td>
<td>4.0167</td>
<td>80.67</td>
</tr>
<tr>
<td>------</td>
<td>---</td>
<td>------</td>
<td>--------</td>
<td>----------</td>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8422.18</td>
</tr>
</tbody>
</table>

We now want to break the treatment sum of squares down into pieces due to main effects and interactions. The three way (E by sign by H) “table” of means we used so far “contains” all the
main effects, two way interactions and the three way interaction. Using two-way tables we can get the sums of squares for any two main effects and their interaction. Once we have all of these, a simple subtraction from the treatment sum of squares will give the remaining sum of squares, the one for the three way interaction. For example, let’s examine the 2 by 3 table for E*H. First, the PROC MEANS printout:

\_TYPE\_ =6

<table>
<thead>
<tr>
<th>E</th>
<th>H</th>
<th>sign</th>
<th><em>FREQ</em></th>
<th>mn_sales</th>
<th>diff</th>
<th>contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>-1</td>
<td>-</td>
<td>10</td>
<td>79.9</td>
<td>-19.0833</td>
<td>3641.74</td>
</tr>
<tr>
<td>no</td>
<td>0</td>
<td>-</td>
<td>10</td>
<td>91.0</td>
<td>-7.9833</td>
<td>637.34</td>
</tr>
<tr>
<td>no</td>
<td>1</td>
<td>-</td>
<td>10</td>
<td>102.0</td>
<td>3.0167</td>
<td>91.00</td>
</tr>
<tr>
<td>yes</td>
<td>-1</td>
<td>-</td>
<td>10</td>
<td>103.8</td>
<td>4.8167</td>
<td>232.00</td>
</tr>
<tr>
<td>yes</td>
<td>0</td>
<td>-</td>
<td>10</td>
<td>105.8</td>
<td>6.8167</td>
<td>464.67</td>
</tr>
<tr>
<td>yes</td>
<td>1</td>
<td>-</td>
<td>10</td>
<td>111.4</td>
<td>12.4167</td>
<td>1541.74</td>
</tr>
</tbody>
</table>

\[72.9 \quad 91.0 \quad 102.0 \quad Mn=90.967\]
\[103.8 \quad 105.8 \quad 111.4 \quad Mn=107.000\]
\[Mn=91.85 \quad Mn=98.40 \quad Mn=106.70\]

\[6608.48\]

E H sign _FREQ_ mn_sales diff contribution
- -1 - 20 91.85 -7.13333 1017.69
- 0 - 20 98.40 -0.58333 6.81
- 1 - 20 106.70 7.71667 1190.94

\[2215.43\]

E H sign _FREQ_ mn_sales diff contribution
no . - 30 90.967 -8.01667 1928.01
yes . - 30 107.000 8.01667 1928.01

\[3856.02\]
The two way table sum of squares $6608.48$ contains the $E^*H$ interaction sum of square plus the $H$ sum of squares $2215.43$ and the $E$ sum of squares $3856.02$ so the interaction sum of squares is $6608.48 - 2215.43 - 3856.02 = 537$. With 6 entries the table has $6-1=5$ df while $H$ has $3-1=2$ and $E$ has 1 so the interaction has $5-2-1=2$ degrees of freedom.

Note that these formulas work only for balanced data whereas the general matrix approach of SAS works for balanced or unbalanced data. Because the data are balanced, the Type I (sequential) and Type III (partial) sums of squares are the same as each other. Here are the Type III sums of squares from the code above:

<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>town</td>
<td>4</td>
<td>604.233333</td>
<td>151.058333</td>
<td>4.21</td>
<td>0.0057</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>3856.016667</td>
<td>3856.016667</td>
<td>107.34</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>H</td>
<td>2</td>
<td>2215.433333</td>
<td>1107.716667</td>
<td>30.84</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>E*H</td>
<td>2</td>
<td>537.033333</td>
<td>268.516667</td>
<td>7.47</td>
<td>0.0016</td>
</tr>
<tr>
<td>sign</td>
<td>1</td>
<td>74.816667</td>
<td>74.816667</td>
<td>2.08</td>
<td>0.1561</td>
</tr>
<tr>
<td>E*sign</td>
<td>1</td>
<td>920.416667</td>
<td>920.416667</td>
<td>25.62</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>H*sign</td>
<td>2</td>
<td>707.033333</td>
<td>353.516667</td>
<td>9.84</td>
<td>0.0003</td>
</tr>
<tr>
<td>E<em>H</em>sign</td>
<td>2</td>
<td>111.433333</td>
<td>55.716667</td>
<td>1.55</td>
<td>0.2230</td>
</tr>
</tbody>
</table>

The $E^*H^*sign$ is whatever it takes to make all but the town sum of squares add up to the treatment sum of squares we computed earlier. Notice how the mean squares measure average variation due to the levels of the treatment in question and how the F ratio compares that variation to the variation due to noise. Hence large F's imply added variability due to nonzero treatment effects.
In our case, the three way interaction is insignificant at the 5% level. This will not always be the case so it is of interest to understand that a three way interaction is the failure of the two way interactions of two of the factors to have the same pattern across the levels of the third factor and in general a k-way interaction is the failure of the patterns of k-1 interactions to be the same across the levels of the remaining factor. Here for example are H*sign interaction plots for two levels of the E factor:

The end of aisle sales (right side) are higher than the mid aisle sales (left) but that is an end of aisle effect, *not* an interaction. It is the fact that the *pattern* on the left (two almost straight lines converging together) differs from that on the right (two crossing kinked lines that are almost horizontal at first) that constitutes the three way interaction. Apparently the error variation is enough to explain the differences and there is not evidence at the usual level of a real three way interaction. Note that we do not
see the amount of variation in the individual data points in these plots.

We could use contrasts to ask questions such as: “Is a high display in mid aisle as advantageous as a more expensive end of aisle display at a low level, and does this answer depend on whether there is a sign for my product?” (see the little X pattern between the two interaction plots)

<table>
<thead>
<tr>
<th>Coefficients for overall contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>End \ Ht</strong></td>
</tr>
<tr>
<td>no</td>
</tr>
<tr>
<td>yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients for contrast interaction with sign factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>(end,sign)</td>
</tr>
<tr>
<td>(no, yes)</td>
</tr>
<tr>
<td>(yes, yes)</td>
</tr>
<tr>
<td>(no, no)</td>
</tr>
<tr>
<td>(yes, no)</td>
</tr>
</tbody>
</table>

One way to do this, as illustrated in the demo, is just to create a treatment variable to delineate the 12 treatments then be sure you assign the above coefficients to the proper levels. Noticing that we are just talking about a linear combination (weighted sum) of 4 means, we can also just use the 3 rules to formulate a t-test whose square would be the F test for the contrast.

Now suppose you want to run an experiment with 12 factors each at 2 levels. Then for even a single observation at each
treatment combination you would need $2^{12} = 4096$ observations and you would have 12 main effects plus $12(11)/2 = 66$ two way interactions leaving 4015 degrees of freedom for 3 way and higher interactions. Perhaps we can reduce the number of observations needed by giving up on estimating high order interactions. To do so, we next look at “fractional factorials.”