Unbalanced Designs – Mechanics

- Estimate of $\sigma^2$ becomes weighted average of treatment combination sample variances.

Types of $SS$

Difference depends on what hypotheses are tested and how differing sample sizes are dealt with.

- Same for last interaction
– Interaction effects appear in factor effects formulation for testing other factors.
– Therefore, if interaction effects large, can unduly affect results of marginal tests.

• Type I: Keep each observation weighted equally – upweights means with larger group sizes
– Gives undue influence to larger groups of other factor
– Other factors appear in factor effects contrast statement. Tests of a factor not
independent of result for other factor.

- If variable is not first in model, general linear test for contrast conditions on first variable (contrasts) being in the (sub)model. (Can’t do with sample contrast statement. Need test option.)

- **Type III:** Keep each treatment weighted equally – downweights observations from groups with larger group sizes
  - Keeps influence of factors the same
  - Outliers (and non-constant variance) have larger
effect in smaller groups

– Other factors do not appear in contrast statement.
Unbalanced Designs – Further Analysis Issues

• Further analysis depends on result of interaction analysis

• Analysis of marginal effects same as in balanced design

• Standard errors tend to be more complicated.
Three-way ANOVA

- Not much different in data or cell means models.
- More interactions (and constraints) in factor effects model
- Post-interaction analysis more complicated, requiring combining factors or looking at one factor level at a time.
Analysis of Three-way ANOVA

- Combine factors for interaction plots.
- Can average over factors with no involvement in significant interactions.
- Combine factors over significant interactions.
KNNL Example

- KNNL page 1018 (knnl1018.sas)
- $Y$ is exercise tolerance, minutes until fatigue on a bicycle test
- $A$ is gender, $a = 2$ levels: male = 1, female = 2
- $B$ is percent body fat, $b = 2$ levels: low = 1, high = 2
- $C$ is smoking history, $c = 2$ levels: light = 1, heavy = 2
- $n = 3$ persons aged 25-35 per $(i, j, k)$ cell
Read and check the data

data exercise;
  infile 'h:\System\Desktop\CH23TA04.DAT';
  input extol gender fat smoke;

Define variable for a plot

This is just to set a unique identifier for each treatment. There are other ways to do this.

data exercise;
  set exercise;
  gfs = 100*gender + 10*fat + smoke;
  proc print data=exercise;
<table>
<thead>
<tr>
<th>Obs</th>
<th>extol</th>
<th>gender</th>
<th>fat</th>
<th>smoke</th>
<th>gfs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>111</td>
</tr>
<tr>
<td>2</td>
<td>29.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>111</td>
</tr>
<tr>
<td>3</td>
<td>24.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>111</td>
</tr>
<tr>
<td>4</td>
<td>20.0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>211</td>
</tr>
<tr>
<td>5</td>
<td>21.9</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>211</td>
</tr>
<tr>
<td>6</td>
<td>17.6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>211</td>
</tr>
<tr>
<td>7</td>
<td>14.6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>121</td>
</tr>
<tr>
<td>8</td>
<td>15.3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>121</td>
</tr>
<tr>
<td>9</td>
<td>12.3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>121</td>
</tr>
<tr>
<td>10</td>
<td>16.1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>221</td>
</tr>
<tr>
<td>11</td>
<td>9.3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>221</td>
</tr>
<tr>
<td>12</td>
<td>10.8</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>221</td>
</tr>
<tr>
<td>13</td>
<td>17.6</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>112</td>
</tr>
<tr>
<td>14</td>
<td>18.8</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>112</td>
</tr>
<tr>
<td>15</td>
<td>23.2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>112</td>
</tr>
<tr>
<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>
</tr>
<tr>
<td>16</td>
<td>14.8</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>10.3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>11.3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>14.9</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>20.4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>12.8</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>10.1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>14.4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>6.1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Plot the data

proc sort data=exercise;
  by gender fat smoke;
title1 'Plot of the data';
symbol1 v=circle i=none c=black;
proc gplot data=exercise;
  plot extol*gfs/ haxis = 111 112 121 122 211
Find the means

```sas
proc means data=exercise;
  output out=exer2 mean=avextol;
  by gender fat smoke;
```

Make a two-variable combination of *fat* and *smoke*

This is helpful for plotting.

```sas
data exer2;
  set exer2;
  fs = fat*10 + smoke;
proc print data=exer2;
```
<table>
<thead>
<tr>
<th>Obs</th>
<th>gender</th>
<th>fat</th>
<th>smoke</th>
<th>avextol</th>
<th>fs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>25.9667</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>19.8667</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>14.0667</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>16.0333</td>
<td>22</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>19.8333</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>12.1333</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>12.0667</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10.2000</td>
<td>22</td>
</tr>
</tbody>
</table>
Plot the means

proc sort data=exer2; by fs;
title1 'Plot of the means';
symbol1 v='M' i=join c=black;
symbol2 v='F' i=join c=black;
proc gplot data=exer2;
   plot avextol*fs=gender / haxis = 11 12 21 22;
From this plot it appears that \textit{gender} probably doesn’t interact too much with the other variables.
Note: Interaction plots in the 3-variable model take the form of putting 2-factor combinations on the $X$-axis with separate lines for the third factor.

```
proc glm data=exercise;
   class gender fat smoke;
   model extol=gender|fat|smoke / solution;
   means gender*fat*smoke;
```

Recall that $gender\mid fat\mid smoke$ is short for

- gender fat smoke
- gender*fat
- gender*smoke
- fat*smoke
- gender*fat*smoke.
The GLM Procedure

Class Level Information

Class Levels Values
gender 2 1 2
fat 2 1 2
smoke 2 1 2

Number of observations 24

Dependent Variable: extol

Sum of

Source   DF    Squares    Mean Square    F Value    Pr > F
Model     7   588.5829167   84.0832738    9.01    0.0002
Error    16   149.3666667   9.3354167
Corrected Total 23   737.9495833

Fall 2015
<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>gender</td>
<td>1</td>
<td>176.5837500</td>
<td>176.5837500</td>
<td>18.92</td>
<td>0.0005</td>
</tr>
<tr>
<td>fat</td>
<td>1</td>
<td>242.5704167</td>
<td>242.5704167</td>
<td>25.98</td>
<td>0.0001</td>
</tr>
<tr>
<td>gender*fat</td>
<td>1</td>
<td>13.6504167</td>
<td>13.6504167</td>
<td>1.46</td>
<td>0.2441</td>
</tr>
<tr>
<td>smoke</td>
<td>1</td>
<td>70.3837500</td>
<td>70.3837500</td>
<td>7.54</td>
<td>0.0144</td>
</tr>
<tr>
<td>gender*smoke</td>
<td>1</td>
<td>11.0704167</td>
<td>11.0704167</td>
<td>1.19</td>
<td>0.2923</td>
</tr>
<tr>
<td>fat*smoke</td>
<td>1</td>
<td>72.4537500</td>
<td>72.4537500</td>
<td>7.76</td>
<td>0.0132</td>
</tr>
<tr>
<td>gender<em>fat</em>smoke</td>
<td>1</td>
<td>1.8704167</td>
<td>1.8704167</td>
<td>0.20</td>
<td>0.6604</td>
</tr>
</tbody>
</table>

All main effects are significant. *Gender* and *fat* appear to have bigger effects than *smoke*. The two-way interaction between *fat* and *smoke* is also significant.
SAS Parameter Estimates

Solution option on the model statement gives parameter estimates for the glm parameterization. These are as we have seen before; any main effect or interaction with a subscript of $a$, $b$, or $c$ is zero. These can be used to reproduce the cell means in the usual way.
| Parameter       | Estimate | Error       | t Value | Pr > |t| |
|-----------------|----------|-------------|---------|------|---|
| Intercept       | 10.2     | 1.76403105  | 5.78    | <.0001 |
| gender 1        | 5.83333333 | 2.49471664 | 2.34   | 0.0327 |
| gender 2        | 0.0      |             |        |      |
| fat 1           | 1.93333333 | 2.49471664 | 0.77   | 0.4497 |
| fat 2           | 0.0      |             |        |      |
| gender * fat 1 1| 1.9      | 3.52806211  | 0.54   | 0.5976 |
| gender * fat 1 2| 0.0      |             |        |      |
| gender * fat 2 1| 0.0      |             |        |      |
| gender * fat 2 2| 0.0      |             |        |      |
| smoke 1         | 1.86666667 | 2.49471664 | 0.75   | 0.4652 |
| smoke 2         | 0.0      |             |        |      |
| gender * smoke 1 1| -3.83333333 | 3.52806211 | -1.0   |       |
| gender * smoke 1 2| 0.0      |             |        |      |
| gender * smoke 2 1| 0.0      |             |        |      |
We can get the zero-sum constraints in the usual way (see the file `knnl1018.sas` for the code).
<table>
<thead>
<tr>
<th>Obs</th>
<th>gender</th>
<th>fat</th>
<th>smoke</th>
<th>mu</th>
<th>alpha</th>
<th>beta</th>
<th>gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>16.2708</td>
<td>2.7125</td>
<td>3.17917</td>
<td>1.7125</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>16.2708</td>
<td>2.7125</td>
<td>3.17917</td>
<td>-1.7125</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>16.2708</td>
<td>2.7125</td>
<td>-3.17917</td>
<td>1.7125</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>16.2708</td>
<td>2.7125</td>
<td>-3.17917</td>
<td>-1.7125</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>16.2708</td>
<td>-2.7125</td>
<td>3.17917</td>
<td>1.7125</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>16.2708</td>
<td>-2.7125</td>
<td>3.17917</td>
<td>-1.7125</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>16.2708</td>
<td>-2.7125</td>
<td>-3.17917</td>
<td>1.7125</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16.2708</td>
<td>-2.7125</td>
<td>-3.17917</td>
<td>-1.7125</td>
</tr>
</tbody>
</table>
Notice from the parameter estimates that $\beta\gamma$ is about the same size as $\gamma$. This makes it pretty hard to interpret the main effect of *smoke*.
Looking at this plot, it appears that smoking decreases tolerance for those of low body fat, but makes almost no difference for those at the high body fat.
Example Approach

Since there appears to be a $fat$ by $smoke$ interaction, let’s run a two-way ANOVA (no additional interaction) using the $fat \times smoke$ variable and $gender$. This will consider the four $fs$ categories separately. We will also use the interaction plot to describe the interaction.

```r
proc glm data=exercise;
  class gender fs;
  model extol=gender fs;
  means gender fs/tukey;
```
<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>561.9916667</td>
<td>140.4979167</td>
<td>15.17</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>19</td>
<td>175.9579167</td>
<td>9.2609430</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>23</td>
<td>737.9495833</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>gender</td>
<td>1</td>
<td>176.5837500</td>
<td>176.5837500</td>
<td>19.07</td>
<td>0.0003</td>
</tr>
<tr>
<td>fs</td>
<td>3</td>
<td>385.4079167</td>
<td>128.4693056</td>
<td>13.87</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

Notice that the $SS$ for $gender$ is the same as before. Also, the $SS$ now shown for $fs$ is the sum of the $SS$ for $fat$, $smoke$, and $fat \times smoke$ in the original model. The $SS$ for the remaining interaction terms has now
been incorporated into the error term. $SSE$ has gone up, but $MSE$ has actually gone down a little.
Different means for \textit{gender}

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>\textit{gender}</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>18.983</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>13.558</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>

(Well, we knew that since \textit{gender} was significant)
Tukey comparisons for $f_s$

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean</th>
<th>N</th>
<th>fs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>22.900</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>B</td>
<td>16.000</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>B</td>
<td>13.117</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>B</td>
<td>13.067</td>
<td>6</td>
<td>21</td>
</tr>
</tbody>
</table>

Category $f_s = 1$ is the low body fat and light smoking history group. The other three groups were not significantly different from each other.
Chapter 25: One-way Random Effects Design

Fixed Effects vs Random Effects

- Up to this point we have been considering "fixed effects models", in which the levels of each factor were fixed in advance of the experiment and we were interested in differences in response among those specific levels.
Now we will consider “random effects models”, in which the factor levels are meant to be representative of a general population of possible levels. We are interested in whether that factor has a significant effect in explaining the response, but only in a general way. For example, we’re not interested in a detailed comparison of level 2 vs. level 3, say.

When we have both fixed and random effects, we call it a “mixed effects model”. The main SAS procedure we will use is called “proc mixed” which allows for fixed and random effects, but we can also use
glm with a random statement. We’ll start first with a single random effect.

- In some situations it is clear from the experiment whether an effect is fixed or random. However there are also situations in which calling an effect fixed or random depends on your point of view, and on your interpretation and understanding. So sometimes it is a personal choice. This should become more clear with some examples.
Data for one-way design

- $Y$, the response variable
- Factor with levels $i = 1$ to $r$
- $Y_{i,j}$ is the $j$th observation in cell $i$, $j = 1$ to $n_i$
- A balanced design has $n = n_i$
KNNL Example

- KNNL page 1036 (knnl1036.sas)
- $Y$ is the rating of a job applicant
- Factor $A$ represents five different personnel interviewers (officers), $r = 5$ levels
- $n = 4$ different applicants were randomly chosen and interviewed by each interviewer (i.e. 20 applicants) ($applicant$ is not a factor since no applicant was interviewed more than once)
• The interviewers were selected at random from the pool of interviewers and the applicants were randomly assigned to interviewers.

• Here we are not so interested in the differences between the five interviewers that happened to be picked (i.e. does Joe give higher ratings than Fred, is there a difference between Ethel and Bob). Rather we are interested in quantifying and accounting for the effect of “interviewer” in general. There are other interviewers in the “population” (at the company) and we want to make inference about them too.
Another way to say this is that with fixed effects we were primarily interested in the means of the factor levels (and the differences between them). With random effects, we are primarily interested in their variances.
Read and check the data

data interview;
  infile 'h:\System\Desktop\CH24TA01.DAT';
  input rating officer;
proc print data=interview;

Obs    rating  officer
1      76      1
2      65      1
3      85      1
4      74      1
5      59      2
6      75      2
7      81      2
8      67      2
9      49      3
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>63</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>61</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>46</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>74</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>71</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>85</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>89</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>66</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>84</td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>80</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>79</td>
<td>5</td>
</tr>
</tbody>
</table>
Plot the data

title1 'Plot of the data';
symbol1 v=circle i=none c=black;
proc gplot data=interview;
    plot rating*officer;
run;
Random effects model (cell means)

This model is also called

- ANOVA Model II
- A variance components model

\[ Y_{i,j} = \mu_i + \epsilon_{i,j} \]
\[ Y_{i,j} = \mu_i + \epsilon_{i,j} \]

- The \( \mu_i \) are iid \( N(\mu, \sigma^2_A) \). NOTE!!!!!! THIS IS DIFFERENT!!!!!
- The \( \epsilon_{i,j} \) are iid \( N(0, \sigma^2) \)
- \( \mu_i \) and \( \epsilon_{i,j} \) are independent
- \( Y \sim N(\mu, \sigma^2_A + \sigma^2) \)
Now the $\mu_i$ are random variables with a common mean. The question of “are they all the same” can now be addressed by considering whether the \textit{variance} of their distribution, $\sigma^2_A$, is zero. Of course, the estimated means will likely be different from each other; the question is whether the difference can be explained by error ($\sigma^2$) alone.
The text uses the symbol $\sigma^2_{\mu}$ instead of $\sigma^2_A$; they are the same thing. I prefer the latter notation because it generalizes more easily to more than one factor, and also to the factor effects model.
Two Sources of Variation

Observations with the same $i$ (e.g. the same interviewer) are dependent, and their covariance is $\sigma_A^2$. The components of variance are $\sigma_A^2$ and $\sigma^2$. We want to get an idea of the relative magnitudes of these variance components.
Random factor effects model

Same basic idea as before... $\mu_i = \mu + \alpha_i$. The model is $Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}$.

$$\alpha_i \sim N(0, \sigma^2_A)$$

$$\epsilon_{i,j} \sim N(0, \sigma^2)$$

$$Y_{i,j} \sim N(\mu, \sigma^2_A + \sigma^2)$$

The book uses $\sigma^2_\alpha$ instead of $\sigma^2_A$ here. Despite the different notations, $\sigma^2_\alpha$ and $\sigma^2_\mu$ are really the same thing, because $\mu_i$ and $\alpha_i$ differ only by an additive constant.
(\mu), so they have the same variance. That is why in these notes I’m using the same symbol \sigma^2_A to refer to both. (With two factors we will have to distinguish between these.)
Parameters

There are two important parameters in these models: $\sigma_A^2$ and $\sigma^2$ (also $\mu$ in the F.E.M.).

The cell means $\mu_i$ are random variables, not parameters.

We are sometimes interested in estimating $\frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} = \frac{\sigma_A^2}{\sigma_Y^2}$.

In some applications it is called the *intraclass correlation coefficient*. It is the correlation between two observations with the same $i$. 
**ANOVA Table**

- The terms and layout of the ANOVA table are the same as what we used for the fixed effects model.
- The expected mean squares ($EMS$) are different because of the additional random effects, so we will estimate parameters in a new way.
- Hypotheses being tested are also different.
EMS and parameter estimates

\[ E(MSE) = \sigma^2 \] as usual. We use \( MSE \) to estimate \( \sigma^2 \).

\[ E(MSA) = \sigma^2 + n\sigma^2_A \]. Note that this is different from before. From this you can see that we should use \( \frac{(MSA - MSE)}{n} \) to estimate \( \sigma^2_A \).
Hypotheses

\[ H_0 : \sigma^2_A = 0 \]
\[ H_1 : \sigma^2_A \neq 0 \]

The test statistic is \( F = \frac{MSA}{MSE} \) with \( r - 1 \) and \( r(n - 1) \) degrees of freedom (since this ratio is 1 when the null hypothesis is true); reject when \( F \) is large, and report the \( p \)-value. Note that in the one factor analysis, the test is the same it was before. This WILL NOT be the case as we add more factors.
SAS Coding and Output

Run **proc glm** with a **random** statement

```
proc glm data=interview;
    class officer;
    model rating=officer;
    random officer;
```

<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>4</td>
<td>1579.700000</td>
<td>394.925000</td>
<td>5.39</td>
<td>0.0068</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>1099.250000</td>
<td>73.283333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>19</td>
<td>2678.950000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Random statement output

Source officer  Type III Expected Mean Square Var(Error) + 4 Var(officer)

This is SAS’s way of saying $E(MSA) = \sigma^2 + 4\sigma^2_A$

(note $n = 4$ replicates).
This procedure gets the “variance components”.

**proc varcomp**

```plaintext
proc varcomp data=interview;
   class officer;
   model rating=officer;
```

<table>
<thead>
<tr>
<th>MIVQUE(0) Estimates</th>
<th>rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var(officer)</td>
<td>80.41042</td>
</tr>
<tr>
<td>Var(Error)</td>
<td>73.28333</td>
</tr>
</tbody>
</table>

(Other methods are available for estimation; `mivque` is the default.)
SAS is now saying

\[
\text{Var}(\text{Error}) = \hat{\sigma}^2 = 73.28333 \text{ (notice this is just } MSE) \\
\text{Var}(\text{officer}) = \hat{\sigma}_\mu^2 = 80.41042 = \frac{(394.925 - 73.283)}{4} \\
= \frac{(MSA - MSE)}{n}.
\]

As an alternative to using \texttt{proc glm} with a \texttt{random} statement, and \texttt{proc varcomp}, you could instead use \texttt{proc mixed}, which has some options specifically for mixed models.
**proc mixed**

```plaintext
proc mixed data=interview cl;
    class officer;
    model rating=;
    random officer/vcorr;
```

- The `cl` option after `data=interview` asks for the confidence limits.
- The `class` statement lists all the categorical variables just as in `glm`.
The model rating=; line looks strange. In proc mixed, the model statement lists only the fixed effects. Then the random effects are listed separately in the random statement. In our example, there were no fixed effects, so we had no predictors on the model line. We had one random effect, so it went on the random line.

This is different from glm, where all the factors (fixed and random) are listed on the model line, and then the random ones are repeated in the random statement.
• Just in case you’re not confused enough, `proc varcomp` assumes all factors are random effects unless they are specified as fixed...
Proc mixed gives a huge amount of output. Here are some pieces of it.

. Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Estimate</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>officer</td>
<td>80.4104</td>
<td>0.05</td>
<td>24.4572</td>
<td>1498.97</td>
</tr>
<tr>
<td>Residual</td>
<td>73.2833</td>
<td>0.05</td>
<td>39.9896</td>
<td>175.54</td>
</tr>
</tbody>
</table>

The estimated intraclass correlation coefficient is

\[
\frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + \hat{\sigma}_Y^2} = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_Y^2} = \frac{80.4104}{80.4104 + 73.2833} = 0.5232.
\]

About half the variance in rating is explained by interviewer.
Output from `vcorr` option

This gives the intraclass correlation coefficient.

<table>
<thead>
<tr>
<th>Row</th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
<th>Col4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000</td>
<td>0.5232</td>
<td>0.5232</td>
<td>0.5232</td>
</tr>
<tr>
<td>2</td>
<td>0.5232</td>
<td>1.0000</td>
<td>0.5232</td>
<td>0.5232</td>
</tr>
<tr>
<td>3</td>
<td>0.5232</td>
<td>0.5232</td>
<td>1.0000</td>
<td>0.5232</td>
</tr>
<tr>
<td>4</td>
<td>0.5232</td>
<td>0.5232</td>
<td>0.5232</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Confidence Intervals

- For $\mu$ the estimate is $\bar{Y}_{..}$, and the variance of this estimate under the random effects model becomes
  $$\sigma^2 \{ \bar{Y}_{..} \} = \frac{(n \sigma_A^2 + \sigma^2)}{r n}$$
  which may be estimated by
  $$s^2 \{ \bar{Y}_{..} \} = \frac{(MSA)}{r n}.$$  See page 1038 for derivation if you like. To get a CI we use a $t$ critical value with $r - 1$ degrees of freedom.

- Notice that the variance here involves a combination of the two errors and we end up using $MSA$ instead of $MSE$ in the estimate (we used $MSE$ in the fixed
effects case).

- We may also get point estimates and CI’s for $\sigma^2$, $\sigma_A^2$, and the intraclass correlation $\frac{\sigma_A^2}{\sigma^2_A + \sigma^2}$. See pages 1040-1047 for details. All of these are available in `proc mixed`. 
Applications

• In the KNNL example we would like $\frac{\sigma^2_\mu}{(\sigma^2_\mu + \sigma^2)}$ to be small, indicating that the variance due to interviewer is small relative to the variance due to applicants.

• In many other examples we would like this quantity to be large. One example would be measurement error - if we measure $r$ items $n$ times each, $\sigma^2$ would represent the error inherent to the instrument of measurement.
Two-way Random Effects Model

Data for two-way design

- $Y$, the response variable
- Factor $A$ with levels $i = 1$ to $a$
- Factor $B$ with levels $j = 1$ to $b$
- $Y_{i,j,k}$ is the $k$th observation in cell $(i, j) \ k = 1$ to $n_{i,j}$
- For balanced designs, $n = n_{i,j}$
KNNL Example

- KNNL Problem 25.15, page 1080
  (knnl1080.sas)

- $Y$ is fuel efficiency in miles per gallon

- Factor $A$ represents four different drivers, $a = 4$ levels

- Factor $B$ represents five different cars of the same model, $b = 5$

- Each driver drove each car twice over the same
40-mile test course \((n = 2)\)
Read and check the data

data efficiency;
    infile 'h:\System\Desktop\CH24PR15.DAT';
    input mpg driver car;
proc print data=efficiency;

    Obs  mpg  driver  car
        1   25.3   1    1
        2   25.2   1    1
        3   28.9   1    2
        4   30.0   1    2
        5   24.8   1    3
        6   25.1   1    3
        7   28.4   1    4
        8   27.9   1    4
...

Prepare the data for a plot, and plot the data

data efficiency;
  set efficiency;
  dc = driver*10 + car;
title1 'Plot of the data';
symbol1 v=circle i=none c=black;
proc gplot data=efficiency;
  plot mpg*dc;
Find and plot the means

```
proc means data=efficiency;
  output out=effout mean=avmpg;
  var mpg;
  by driver car;
title1 'Plot of the means';
symbol1 v='A' i=join c=black;
symbol2 v='B' i=join c=black;
symbol3 v='C' i=join c=black;
symbol4 v='D' i=join c=black;
symbol5 v='E' i=join c=black;
proc gplot data=effout;
  plot avmpg*driver=car;
```
Random Effects Model

Random cell means model

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

- \( \mu_{i,j} \sim N(\mu, \sigma^2_\mu) \). NOTE!!!!!! THIS IS DIFFERENT!!!
- \( \epsilon_{i,j,k} \sim iid\ N(0, \sigma^2) \) as usual
- \( \mu_{i,j}, \epsilon_{i,j,k} \) are independent
- The above imply that \( Y_{i,j,k} \sim N(\mu, \sigma^2_\mu + \sigma^2) \)
Dependence among the $Y_{i,j,k}$ can be most easily described by specifying the covariance matrix of the vector $(Y_{i,j,k})$
Random factor effects model

\[ Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{i,j} + \epsilon_{i,j,k}, \text{ where} \]

\[
\alpha_i \sim N(0, \sigma^2_A) \\
\beta_j \sim N(0, \sigma^2_B) \\
(\alpha \beta)_{i,j} \sim N(0, \sigma^2_{AB}) \\
\sigma^2_Y = \sigma^2_A + \sigma^2_B + \sigma^2_{AB} + \sigma^2
\]

Now the component \( \sigma^2_\mu \) from the cell means model can be divided up into three components - \( A \), \( B \), and \( AB \).

That is, \( \sigma^2_\mu = \sigma^2_A + \sigma^2_B + \sigma^2_{AB} \)
Parameters

- There are five parameters in this model: $\mu$, $\sigma_A^2$, $\sigma_B^2$, $\sigma_{AB}^2$, $\sigma^2$
- The cell means are random variables, not parameters!!!

ANOVA Table

- The terms and layout of the ANOVA table are the same as what we used for the fixed effects model
- However, the expected mean squares ($EMS$) are
different.
EMS and parameter estimates

\[
\begin{align*}
E(\text{MSA}) & = \sigma^2 + b n \sigma_A^2 + n \sigma_{AB}^2 \\
E(\text{MSB}) & = \sigma^2 + a n \sigma_B^2 + n \sigma_{AB}^2 \\
E(\text{MSAB}) & = \sigma^2 + n \sigma_{AB}^2 \\
E(\text{MSE}) & = \sigma^2
\end{align*}
\]

Estimates of the variance components can be obtained from these equations or other methods.
Note the patterns in the $EMS$: (these hold for balanced data).

They all contain $\sigma^2$. For $MSA$, it also contains all the $\sigma^2$'s that have an $A$ in the subscript ($\sigma^2_A$ and $\sigma^2_{AB}$); similarly for the other $MS$ terms.

The coefficient of each term (except the first) is the product of $n$ and all letters not represented in the subscript. It is also the total number of observations at each fixed level of the level corresponding to the subscript (e.g. there are $nb$ observations for each level of $A$).
Hypotheses

\[ H_{0A} : \sigma_A^2 = 0; \quad H_{1A} : \sigma_A^2 \neq 0 \]
\[ H_{0B} : \sigma_B^2 = 0; \quad H_{1B} : \sigma_B^2 \neq 0 \]
\[ H_{0AB} : \sigma_{AB}^2 = 0; \quad H_{1AB} : \sigma_{AB}^2 \neq 0 \]
Hypothesis $H_{0A}$

- $H_{0A} : \sigma_A^2 = 0; H_{1A} : \sigma_A^2 \neq 0$
- $E(MSA) = \sigma^2 + bn\sigma_A^2 + n\sigma_{AB}^2$
- $E(MSAB) = \sigma^2 + n\sigma_{AB}^2$
- $E(MSE) = \sigma^2$

- Need to look for the ratio that will be 1 when $H_0$ is true and bigger than 1 when it is false. So this hypothesis will be tested by $F = \frac{MSA}{MSAB}$ (not the usual fixed effects test statistic). The degrees of freedom for the
test will be the degrees of freedom associated to those mean squares: \( a - 1, (a - 1)(b - 1) \).

- Notice you can no longer assume that the denominator is \( MSE \)!!!! (Note that the test using \( MSE \) is done by SAS, but it is not particularly meaningful (it sort of tests both main and interaction at once).)
Hypothesis $H_{0B}$

- $H_{0B} : \sigma_B^2 = 0$; $H_{1B} : \sigma_B^2 \neq 0$
- $E(MSB) = \sigma^2 + an\sigma_B^2 + n\sigma_{AB}^2$
- $E(MSAB) = \sigma^2 + n\sigma_{AB}^2$
- $E(MSE) = \sigma^2$
- So $H_{0B}$ is tested by $F = \frac{MSB}{MSAB}$ with degrees of freedom $b - 1, (a - 1)(b - 1)$. 
Hypothesis $H_{0AB}$

- $H_{0AB} : \sigma_{AB}^2 = 0$; $H_{1AB} : \sigma_{AB}^2 \neq 0$

- $E(MS_{AB}) = \sigma^2 + n\sigma_{AB}^2$

- $E(MSE) = \sigma^2$

- So $H_{0AB}$ is tested by $F = \frac{MS_{AB}}{MSE}$ with degrees of freedom $(a - 1)(b - 1), ab(n - 1)$. 
Run `proc glm`

```plaintext
proc glm data=efficiency;
  class driver car;
  model mpg=driver car driver*car;
  random driver car driver*car/test;
```
### Model and error 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>19</td>
<td>377.44475</td>
<td>19.8655132</td>
<td>113.03</td>
<td>&lt;</td>
</tr>
<tr>
<td>Error</td>
<td>20</td>
<td>3.51500</td>
<td>0.1757500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>39</td>
<td>380.95975</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Factor effects output

<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>driver</td>
<td>3</td>
<td>280.2847500</td>
<td>93.4282500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>car</td>
<td>4</td>
<td>94.7135000</td>
<td>23.6783750</td>
<td></td>
<td></td>
</tr>
<tr>
<td>driver*car</td>
<td>12</td>
<td>2.4465000</td>
<td>0.2038750</td>
<td>1.16</td>
<td>0.3715</td>
</tr>
</tbody>
</table>
Only the interaction test is valid here: the test for interaction is \( MSAB/MSE \), but the tests for main effects should be \( MSA/MSAB \) and \( MSB/MSAB \) which are done with the `test` statement, not \( /MSE \) as is done here. (However, if you do this the main effects are significant as shown below.)

*Lesson: just because SAS spits out a \( p \)-value, doesn’t mean it is for a meaningful test!*
Random statement output

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>driver</td>
<td>Var(Error) + 2 Var(driver*car) + 10 Var(driver)</td>
</tr>
<tr>
<td>car</td>
<td>Var(Error) + 2 Var(driver*car) + 8 Var(car)</td>
</tr>
<tr>
<td>driver*car</td>
<td>Var(Error) + 2 Var(driver*car)</td>
</tr>
</tbody>
</table>
**Random/test output**

```
. The GLM Procedure
Tests of Hypotheses for Random Model Analysis of Variance
Dependent Variable: mpg

<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>driver</td>
<td>3</td>
<td>280.284750</td>
<td>93.428250</td>
<td>458.26</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>car</td>
<td>4</td>
<td>94.713500</td>
<td>23.678375</td>
<td>116.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>12</td>
<td>2.446500</td>
<td>0.203875</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Error: MS(driver*car)
```

This last line says the denominator of the $F$-tests is the $MS_{AB}$.
<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>driver*car</td>
<td>12</td>
<td>2.446500</td>
<td>0.203875</td>
<td>1.16</td>
<td>0.3715</td>
</tr>
<tr>
<td>Error: MS(Error)</td>
<td>20</td>
<td>3.515000</td>
<td>0.175750</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the interaction term, this is the same test as was done above.


```
proc varcomp
proc varcomp data=efficiency;
   class driver car;
   model mpg=driver car driver*car;
   MIVQUE(0) Estimates
   Variance Component     mpg
   Var(driver)             9.32244
   Var(car)                2.93431
   Var(driver*car)         0.01406
   Var(Error)              0.17575
```
Mixed Models

Two-way mixed model

Two way mixed model has

- One fixed main effect
- One random main effect
- The interaction is considered a random effect
Tests

- Fixed main effect is tested by interaction in the denominator
- Random main effect is tested by error
- Interaction is tested by error
- Notice that these are *backwards* from what you might intuitively extrapolate from the two-way random effects and two-way fixed effects model

See Table 25.5 (page 1052) and below for the $EMS$ that justify these statements. Also see Table 25.6 for the
tests (page 1053).
Notation for two-way mixed model

\( Y \), the response variable
\( A \), the fixed effect (\( a \) levels)
\( B \), the random effect (\( b \) levels)

We’ll stick to balanced designs (\( n_i,j = n \))
Factor effects parameterization

\[ Y_{i,j,k} = \mu + \alpha_i + \beta_j + \alpha \beta_{i,j} + \epsilon_{i,j,k} \]

Where

- \( \mu \) is the overall mean,
- \( \alpha_i \) are fixed (but unknown) fixed main effects with \( \sum \alpha_i = 0 \),
- \( \beta_j \) are \( N(0, \sigma_B^2) \) independent random main effects,
- \( \alpha \beta_{i,j} \) are random interaction effects.
Randomness is “catching” so the interaction between a fixed and a random effect is considered random and has a distribution. However, the interactions are also subject to constraints kind of like fixed effects.

\[ \alpha \beta_{i,j} \sim N \left( 0, \frac{a-1}{a} \sigma_{AB}^2 \right) \]

subject to the constraint

\[ \sum_i (\alpha \beta)_{i,j} = 0 \text{ for each } j. \]

Because of the constraints, \( \alpha \beta_{i,j} \) having the same \( j \) (but different \( i \)) are negatively correlated, with covariance

\[ \text{Cov}(\alpha \beta_{i,j}, \alpha \beta_{i',j}) = -\frac{\sigma_{AB}^2}{a}. \]
Expected Mean Squares

\[ E(\text{MSA}) = \sigma^2 + \frac{nb}{a-1} \sum_i \alpha_i^2 + n\sigma_{\alpha\beta}^2 \]

\[ E(\text{MSB}) = \sigma^2 + na\sigma_{\beta}^2 \]

\[ E(\text{MSAB}) = \sigma^2 + n\sigma_{\alpha\beta}^2 \]

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

SAS (\texttt{proc glm}) writes these out for you but it uses the notation \( Q(A) \) to denote the fixed quantity \( \frac{nb}{a-1} \sum_i \alpha_i^2 \). It uses the names \( \text{Var}(\text{Error}) = \sigma^2 \), \( \text{Var}(B) = \sigma_B^2 \), and \( \text{Var}(A \times B) = \sigma_{AB}^2 \). (It doesn’t
actually use the names $A$ and $B$; it uses the variable names.)
Looking at these EMS, we can see that different denominators will be needed to test for the various effects.

\[
\begin{align*}
H_{0A} & : \quad \text{all } \alpha_i = 0 \text{ is tested by } F = \frac{MSA}{MSAB} \\
H_{0B} & : \quad \sigma_B^2 = 0 \text{ is tested by } F = \frac{MSB}{MSE} \\
H_{0AB} & : \quad \sigma_{AB}^2 = 0 \text{ is tested by } F = \frac{MSAB}{MSE}.
\end{align*}
\]

So, though it seems counterintuitive at first, the fixed effect is tested by the interaction, and the random effect is tested by the error.
Example: KNNL Problem 25.16

(knnl1080a.sas)

$Y$ - service time for disk drives

$A$ - make of drive (fixed, with $a = 3$ levels)

$B$ - technician performing service (random, with $b = 3$ levels)

The three technicians for whom we have data are selected at random from a large number of technicians who work at the company.
data service;
  infile 'h:\stat512\datasets\ch19pr16.dat';
  input time tech make k;
  mt = make*10+tech;
proc print data=service;
proc glm data=service;
  class make tech;
  model time = make tech make*tech;
  random tech make*tech/test;
The GLM Procedure

Dependent Variable: time

<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>8</td>
<td>1268.177778</td>
<td>158.522222</td>
<td>3.05</td>
<td>0.0101</td>
</tr>
<tr>
<td>Error</td>
<td>36</td>
<td>1872.400000</td>
<td>52.011111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>44</td>
<td>3140.577778</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>make</td>
<td>2</td>
<td>28.311111</td>
<td>14.155556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tech</td>
<td>2</td>
<td>24.577778</td>
<td>12.288889</td>
<td>0.24</td>
<td>0.7908</td>
</tr>
<tr>
<td>make*tech</td>
<td>4</td>
<td>1215.288889</td>
<td>303.822222</td>
<td>5.84</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

We have $MS_A = 14.16$, $MS_B = 12.29$, 

\[ MSAB = 303.82, \text{ and } MSE = 52.01. \]
The GLM Procedure

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>make</td>
<td>( \text{Var(Err)} + 5 \times \text{Var(make*tech)} + Q(\text{make}) )</td>
</tr>
<tr>
<td>tech</td>
<td>( \text{Var(Err)} + 5 \times \text{Var(make*tech)} + 15 \times \text{Var(tech)} )</td>
</tr>
<tr>
<td>make*tech</td>
<td>( \text{Var(Err)} + 5 \times \text{Var(make*tech)} )</td>
</tr>
</tbody>
</table>

To test the fixed effect make we must use the interaction:

\[
F_A = \frac{MSA}{MSAB} = \frac{14.16}{303.82} = 0.05 \ldots
\]

with 2,4 df \( (p = 0.955) \)

To test the random effect tech and the interaction, we use error:

\[
F_B = \frac{MSB}{MSE} = \frac{12.29}{52.01} = 0.24 \ldots \text{with 2, 36 df (p = 0.7908)}
\]
\[ F_{AB} = \frac{MS_{AB}}{MSE} = \frac{303.82}{52.01} = 5.84 \ldots \]

with 4, 36 df (\( p = 0.001 \))
The GLM Procedure

Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: time

<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>make</td>
<td>2</td>
<td>28.311111</td>
<td>14.555556</td>
<td>0.05</td>
<td>0.9550</td>
</tr>
<tr>
<td>tech</td>
<td>2</td>
<td>24.577778</td>
<td>12.288889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error: MS(make*tech)</td>
<td>4</td>
<td>1215.288889</td>
<td>303.822222</td>
<td>5.84</td>
<td>0.0010</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>make*tech</td>
<td>4</td>
<td>1215.288889</td>
<td>303.822222</td>
<td>5.84</td>
<td>0.0010</td>
</tr>
<tr>
<td>Error: MS(Error)</td>
<td>36</td>
<td>1872.400000</td>
<td>52.011111</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Three-way models**

We can have zero, one, two, or three random effects (etc)

\[ EMS \] indicate how to do tests

In some cases the situation is complicated and we need approximations, e.g. when all are random, use

\[ MS(AB) + MS(AC) - MS(ABC) \] to test \( A \).