The topic for this week is two-factorial ANOVA each with two levels. One example of this type of research design is the analysis of WEIGHT OF COTTON SEED (outcome variable) by the SPACING OF PLANT (close together, far apart) and VARIETY OF PLANT (large type, small type) (Littell et al., 1991).

Previously, ANOVA with many independent samples had this research design:

```
1 2 3 4 5 6 7 8 9
```

The two-factorial ANOVA can be conceptualized in as follows:

<table>
<thead>
<tr>
<th></th>
<th>FACTOR 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>FACTOR 2</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
</tbody>
</table>

Similar to the one-way ANOVA, two-factorial ANOVA still examines multiple independent samples, but in the latter case there may be significant differences not only in the different levels of Factor 1 as well as in the different levels of Factor 2 but also in the interaction between Factors 1 and 2.

A significant interaction means that combinations of the predictor variables lead to meaningful differences in the outcome. As an example, marathon-running time (outcome variable) might differ significantly not only in gender (male, female) and in number of training hours (high, low) but also in the combination of gender and number of training hours (males who train frequently, males who do not train frequently, females who train frequently, females who do not train frequently).

```
Males - High Training | Females - High Training | Males - Low Training | Females - Low Training
```

If there were significant differences among these combined four groups, there would be a significant interaction effect.

Below is a portion of SAS code that includes the data step, proc anova and proc glm of a two-factorial anova:
data cotton;
  infile " a:\bolls.dat"; * use on PC;
* infile "bolls.dat"; * use on mainframe;
input variety spacing plant seed lint bract;
label variety="Plant Variety"
  spacing="Plant Spacing"
  plant="Plant Number"
  seed="Weight of Seed"
  lint="Weight of Lint"
  bract="Weight of Bract"
;
proc sort;     * the sort procedure is necessary for;
  by variety spacing;    * analyses with classification variables;
* such as variety and spacing here;
  title "Mature Cotton Boll Study";

proc anova;     * when the data cells are balanced;
  class variety spacing;   * procs anova and glm are analogous;
  model seed=variety spacing variety*spacing;

proc glm;     * proc glm needed for unbalanced data;
  class variety spacing;   * anova not intended for UNBAL data;
  model seed=variety spacing variety*spacing / ss1 ss2 ss3;

This code produces an ANOVA table similar to the ones we have already seen. We
will talk more about the statistics in future classes. One should be familiar with
the following concepts in factorial analysis of variance:

1. SS1 – Type I Sums of Squares, also called sequential sums of squares, are
   the incremental improvement in error SS as each effect is added to the
   model. They depend on the order in which effects are specified in the
   model.
2. SS2 – Type II Sums of Squares are the reduction in error SS due to adding
   the term after all other terms have been added to the model except terms
   that contain the effect being tested. SS computations will differ if the cells
   are balanced or unbalanced but there are not due to the ordering of effects
   in the model.
3. SS3 – Type III Sums of Squares, referred to as partial sums of squares, can
   be thought of as the improvement in error SS associated with each effect
   independent of improvement in error SS associated with the other effects.
   Each effect is evaluated for the amount of variability that it alone explains.