Richard Gonzalez
Psych 613
Version 3.2 (Sep 2023)

LECTURE NOTES #4: Randomized Block, Latin Square, and Factorial Designs

Reading Assignment

- Read MD chs 7 and 8
- Read G chs 9, 10, 11

Goals for Lecture Notes #4

- Introduce multiple factors to ANOVA (aka factorial designs)
- Use randomized block and latin square designs as a stepping stone to factorial designs
- Understanding the concept of interaction

1. Factorial ANOVA

The next task is to generalize the one-way ANOVA to test several factors simultaneously. We will develop the logic of *k*-way ANOVA by using two intermediate designs: randomized block and the latin square designs. These designs are interesting in their own right as they highlight the important tool of controlling for blocking variables (also known as covariates), but our purpose will be to use them as "stepping stones" to factorial ANOVA.

The generalization is conducted by 1) adding more parameters to the structural model and 2) expanding the decomposition of the sums of squares. This is the sense in which we are extending the one way ANOVA. We will see that the material in LN3 on contrasts and post hoc tests will also apply to the factorial design.

For example, the two-way ANOVA has two treatments, say, A and B. The structural model is

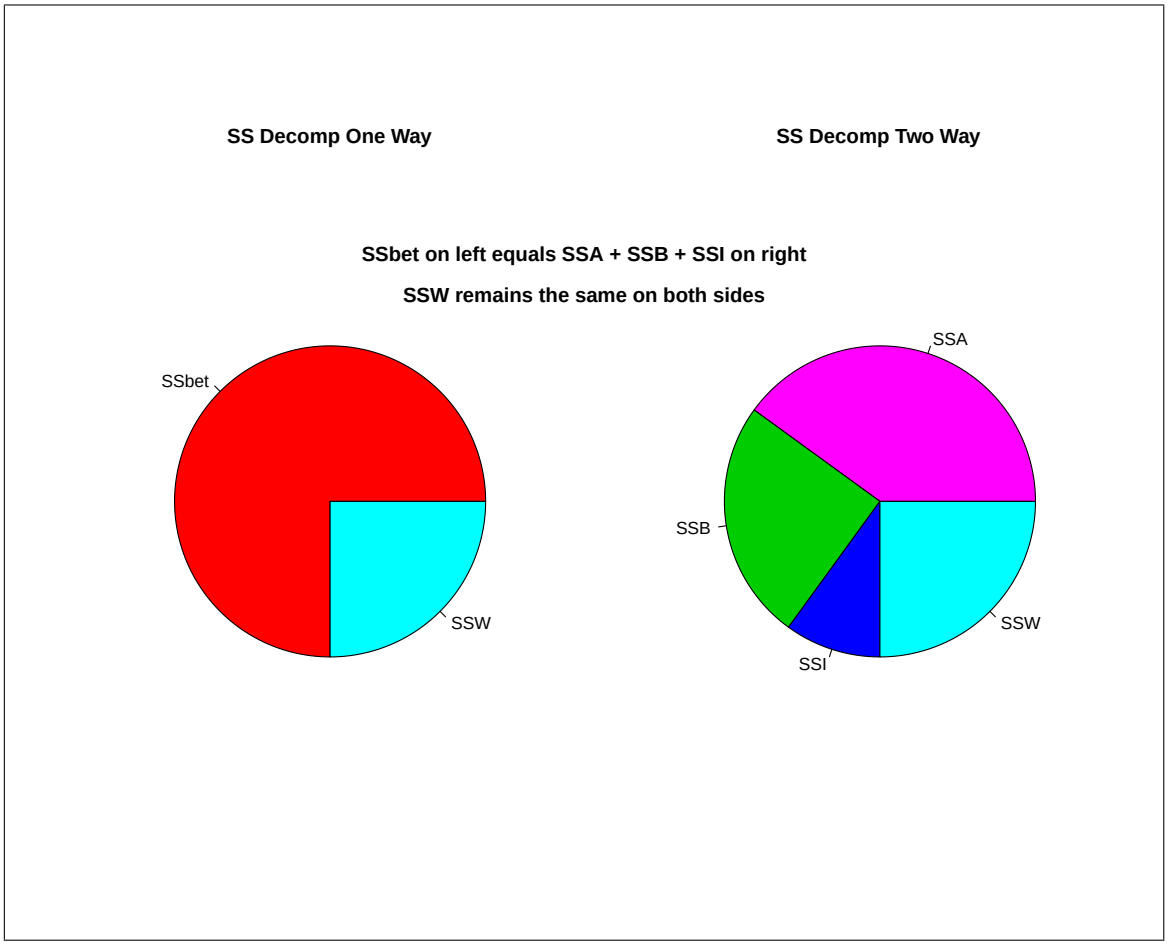
$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \quad (4\text{-}1)$$

where i refers to the levels of treatment A, j refers to the levels of treatment B, and k refers to subjects in the i,j treatment combinations. The main effect for treatment A tests whether

the αs are all equal to zero, the main effect for treatment B tests whether the βs are all equal to zero, and the interaction tests whether the $\alpha\beta$s are all equal to zero. The sum of squares decomposition for the two-way ANOVA involves four components (one for each term in the structural model)

$$\text{SST} \quad = \quad \text{SSA} + \text{SSB} + \text{SSI} + \text{SSW} \tag{4-2}$$

If we had ignored the factorial design structure and tested this design with a one-way ANOVA having I×J cells (the number of levels of treatment A times the number of levels of treatment B), SSW would be identical to the factorial ANOVA, and SSBetween = SSA + SSB + SSI. Thus, a factorial design is one particular way of decomposing the entire sum of squares between-subjects (Equation 4-2 does not contain a single SS for groups but has three different SS terms: SSA for the A factor, SSB for the B factor and SSI for the interaction). The pie chart below represents this relation between the one way ANOVA and the two way ANOVA. Also, we will see that a factorial ANOVA is identical to a ONEWAY ANOVA in terms of decomposing the SSB using a special set of orthogonal contrasts.



2. Randomized Block Design

The signature of this design is that it looks as though it has two factors. However, because there is only one subject per cell, the interaction term cannot be extracted[1]. In a two-way layout when there is one subject per cell, the design is called a randomized block design. When there are two or more subjects per cell (cell sizes need not be equal), then the design is called a two-way ANOVA.

Consider this example (Ott, p. 664). An experimenter tests the effects of three different insecticides on a particular variety of string beans. Each insecticide is applied to four plots (i.e., four observations). The dependent variable is the number of seedlings.

insecticide					row mean	$\hat{\alpha}$'s
1	56	49	65	60	57.50	$\hat{\alpha}_1 = -17.42$
2	84	78	94	93	87.25	$\hat{\alpha}_2 = 12.33$
3	80	72	83	85	80.00	$\hat{\alpha}_3 = 5.08$
					$\hat{\mu} = 74.92$	

Suppose we are unaware of the extra factor and "incorrectly" treat this design as a one-way ANOVA to test the effects of insecticide. In this case, we treat plots as "subjects" and get four observations per cell. The structural model is

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \tag{4-3}$$

The corresponding ANOVA source table appears below. We will use the ANOVA and MANOVA SPSS commands over the next few lectures (e.g., see Appendix 4) and the aov() command in R. Appendix 1 shows the SPSS commands that generated the following source table.

```
 Tests of Significance for SEED using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F

WITHIN CELLS              409.75       9     45.53
INSECT                   1925.17       2    962.58     21.14      .000
```

The main effect for insecticide is significant. Thus, the null hypothesis that population $\alpha_1 = \alpha_2 = \alpha_3 = 0$ is rejected.

Now, suppose each plot was divided into thirds allowing the three insecticides to be used on the same plot of land. The data are the same but the design now includes additional information.

[1]Tukey developed a special test of an interaction for the case when there is one observation per cell. But this concept of an interaction is restricted, and is not the same idea that is usually associated with the concept of an interaction. Tukey suggested this test could be used as a way to detect whether a transformation was needed, a transformation that would instill additivity. I will not discuss this test in this class. For details, however, see NWK p882-885. Note, in particular, their Equation 21.6 that specifies the special, restricted meaning of an interaction in the Tukey test for additivity.

insecticide	plot 1	plot 2	plot 3	plot 4	row mean	$\hat{\alpha}$'s
1	56	49	65	60	57.50	$\hat{\alpha}_1 = -17.42$
2	84	78	94	93	87.25	$\hat{\alpha}_2 = 12.33$
3	80	72	83	85	80.00	$\hat{\alpha}_3 = 5.08$
col mean	73.33	66.33	80.67	79.33	$\hat{\mu} = 74.92$	
$\hat{\beta}$'s	$\hat{\beta}_1 = -1.59$	$\hat{\beta}_2 = -8.59$	$\hat{\beta}_3 = 5.75$	$\hat{\beta}_4 = 4.41$		

structural model: randomized block design

In addition to the row means (treatment effects for insecticides) we now have column means (treatment effects for plots). The new structural model for this design is

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad (4\text{-}4)$$

Note that j is not an index for "subjects" but refers to plot. There is only one observation per each insecticide $\times$ plot cell. Plot is considered a "blocking variable". We include it not because we are necessarily interested in the effects of plot, but because we hope it will "soak up" variance from the error term (i.e., decreasing SS_{error}). Some of the variability may be systematically due to plot and we could gain power by reducing our error term. We acknowledge that some of the variability is due to plot of land and we want to use the statistical analysis to account for this positive feature of the study design.

The αs and βs are estimated by subtracting the grand mean from each row or column mean. See the previous table of data and means were I include the estimated αs and βs.

sum of squares decomposition: randomized block design

The sum of squares decomposition for this randomized block design is

$$SST = SS_{insect} + SS_{plot} + SS_{error} \quad (4\text{-}5)$$

where $SST = \sum(Y_{ij} - \overline{Y})^2$, $SS_{insect} = b\sum(\overline{Y}_i - \overline{Y})^2$, $SS_{plot} = t\sum(\overline{Y}_j - \overline{Y})^2$, b is the number of blocks and t is the number of treatment conditions. Recall that $\overline{Y}$ is the estimated grand mean. We will interpret SS_{error} later, but for now we can compute it relatively easily by rearranging Eq 4-5,

$$SS_{error} = SST - SS_{insect} - SS_{plot} \quad (4\text{-}6)$$

The long way to compute SS error directly from raw data is by the definitional formula

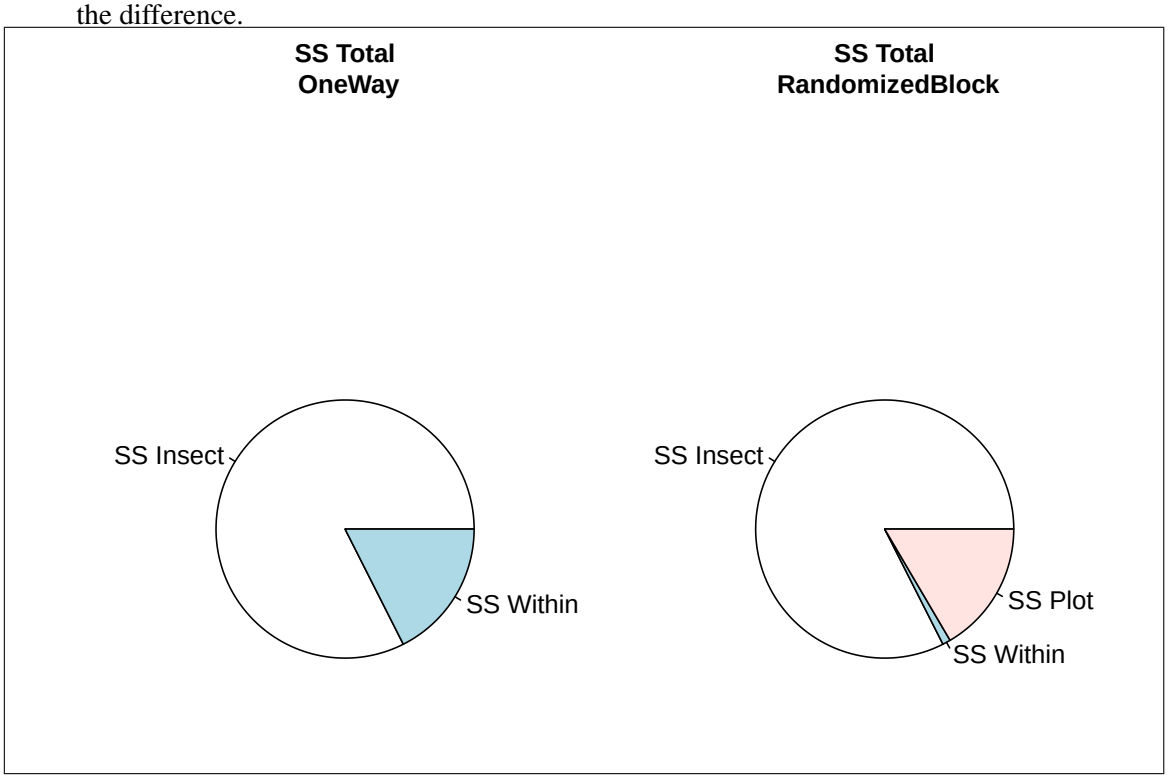
$$\Sigma(Y_{ij} - \overline{Y}_i - \overline{Y}_j + \overline{Y})^2 \quad (4\text{-}7)$$

where Y_{ij} is the score of in row i column j and the other terms are the row mean, the column mean and the grand mean, respectively.

The source table for the new model appears below. The appendices give the SPSS and R commands that produced this output.

```
Tests of Significance for SEED using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

RESIDUAL                  23.50       6      3.92
INSECT                  1925.17       2    962.58    245.77      .000
PLOT                     386.25       3    128.75     32.87      .000
```

Compare this partition of SST to the partition we saw for the one-way ANOVA. In both cases the SS_{insect} are identical but the SS_{error} (or $SS_{residual}$) term differs. What happened? The sum of squares in the one way ANOVA was 409.75. The randomized block design picks up that some of the error term is actually due to the variability of plot, thus the randomized block design performs an additional partition of sum of squares—the 409.75 is decomposed into 386.25 for plot and 23.50 for error. Thus, a major chunk of the original error term can be attributed to the plot factor. The randomized block design is based on a smaller error term, and consequently provides a more powerful test for the insecticide variable (assuming there wasn't also a massive drop in degrees of freedom). The side by side pie charts helps illustrate the difference.



An important distinction: while contrasts decompose SSB as we saw in Lecture Notes #3, blocking factors decompose the error SSW. The SSW on the left is relatively large (one way ANOVA) but the SSW on the right is relatively small because the blocking factor Plot soaks up a relatively large chunk of variability. The SSW on the left equals the sum of SSPlot and SSW on the right. The SS Insect for the key independent variable remains the same.

Notice how in this pair of pie charts, the SSW on the left (the one from the usual one-way ANOVA) is equal to SSW + SSPlot on the right. That is, adding the plot blocking factor to the design served to partition some of the sums of squares that would have been called error in the one-way ANOVA into sum of squares that can be attributable to the plot factor.

With this more complicated design a change in terminology is needed. The error term doesn't necessarily mean the variability of each cell observation from its corresponding cell mean. So rather than say "sums of squares within cells" we typically say the more general "sums of squares error" or "sums of squares residual." Still, some people get sloppy and say "sum of squares within."

The degrees of freedom for error has dropped too relative to the previous analysis without the blocking factor. The formula for the error term in a randomized block design is N - (number of rows minus one) - (number of columns minus 1) - 1.

The randomized block design allows you to use a variable as a "tool" to account for some variability in your data that would otherwise be called "noise" (i.e., be put in the SS_{error}). I like to call them blocking variables; others call them nuisance variables. But as in many things in academics, one researcher's blocking variable may be another researcher's critical variable of study.

I'll mention some examples of blocking variables in class. For a blocking variable to have the desired effect of reducing noise it must be related to the dependent variable. There is also the tradeoff between reduction of sums of squares error and reduction in degrees of freedom. For every new blocking factor added to the design, you are "penalized" on degrees of freedom error. Thus, one does not want to include too many blocking variables.

the paired t test is a randomized block design

A well-known example of a randomized block design is the before-after design where one group of subjects is measured at two points in time (also known as a paired t-test). Each subject is measured twice so if we conceptualized subject as a factor with N levels and time as a second factor with 2 levels, we have a randomized block design structure. The structural model has the grand mean, a time effect (which is the effect of interest), a subject effect and the error term. The subject factor plays the role of the blocking factor because the variability attributed to individual differences across subjects is removed from the error term, typically leading to a more powerful test.

3. Latin Square Design

The latin square design generalizes the randomized block design by having two blocking variables in addition to the treatment variable of interest. The latin square design has a three way layout, making it similar to a three way ANOVA. The latin square design permits only

three main effects to be estimated because the design is incomplete. It is an "incomplete" factorial design because not all cells are represented. For example, in a $4 \times 4 \times 4$ factorial design there are 64 possible cells. But, a latin square design where the treatment variable has 4 levels and each of the two blocking variables have 4 levels, there are only 16 cells. This type of design is useful when it is not feasible to test all the cells required in a three-way ANOVA, say, because of financial or time constraints in conducting a study with many cells.

The signature of a latin square design is that a given treatment of interest appears only once in a given row and a given column. There are many ways of creating "latin squares." Sudoku puzzles are a type of general latin squares with additional constraints such as no repetitions within particular 3 x 3 sub-blocks.

Here is a standard latin square for the case where there are one treatment with four levels, a blocking factor with four levels and a second blocking factor with four levels.

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

One simple way to create a latin square is to make new rows by shifting the columns of the previous row by one (see above).

structural model: latin square design

The structural model for the latin square design is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk} \quad (4\text{-}8)$$

where indices i, j, and k refer to the levels of treatment, blocking variable 1 and blocking variable 2, respectively. The symbol α_i refers to the treatment effect, and the two symbols β_j and γ_k refer to the two blocking variables. Thus, the decomposition yields three main effects but no interactions.

As with the randomized block design, each of the main effect terms α, β and γ are estimated by subtracting the grand mean from the respective row/column/effect mean.

sum of squares decomposition: latin square design

The sums of squares decomposition for the latin square design is

$$SST = SS_{treatment} + SS_{block1} + SS_{block2} + SS_{error} \quad (4\text{-}9)$$

Appendix 3 shows how to code data from a latin square design along with relevant SPSS code. The latin square example in R appears in the R appendix.

As with the randomized block design, the latin square design uses blocking variables to reduce the level of noise. For the latin square design to have the desired effect, the blocking variables must be related to the dependent variable, otherwise you lose power because of the loss of degrees of freedom for the error term.

4. I'll discuss the problem of performing contrasts and post hoc tests on randomized and latin square designs later in the context of factorial designs. Remember that these two designs are rarely used in psychology and I present them now as stepping stones to factorial ANOVA given that even when we have blocking factors we typically have more than one observation per cell (a notable exception occurs in repeated measures designs as we will see in LN5).

5. Checking assumptions in randomized block and latin square designs

   The assumptions are identical to the one-way ANOVA (equal variance, normality, and independence). In addition, we assume that the treatment variable of interest and the blocking variable(s) combine additively (i.e., the structural model holds).

   Because there is only one observation per cell, it is impossible to check the assumption on a cell-by-cell basis. Therefore, the assumptions are checked by collapsing all other variables. For example, in the string bean example mentioned earlier, one can check the boxplots of the three insecticides collapsing across the *plot* blocking variable, thus the assumptions are examined ignoring the blocking factor. Normal probability plots could also be examined in this way. I suggest you also examine the boxplot and normal probability plot on the blocking factor(s), collapsing over the treatment variable.

6. Factorial ANOVA

   We begin with a simple two-way ANOVA where each factor has two levels (i.e., 2×2 design). The structural model for this design is

   structural model: factorial anova

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \quad (4\text{-}10)$$

   where i refers to the levels of treatment A, j refers to the levels of treatment B, and k refers to the subjects in the i,j treatment combinations. The sum of squares decomposition is

   sum of squares decomposition: factorial design

$$SST = SS_A + SS_B + SS_{INT} + SS_{error} \quad (4\text{-}11)$$

   The assumptions in a factorial ANOVA are the same as in a one-way ANOVA: equality of cell variances, normality of cell distributions, and independence.

   The only difference between the two-way factorial and the randomized block design is that in the former more than one subject is observed per cell. This subtle difference allows the

estimation of the interaction effect as distinct from the error term. When only one subject is observed per cell, then the interaction and the error are confounded (i.e., they cannot be separated), unless one redefines the concept of interaction as Tukey did in this "test for additivity," as discussed in Maxwell & Delaney.

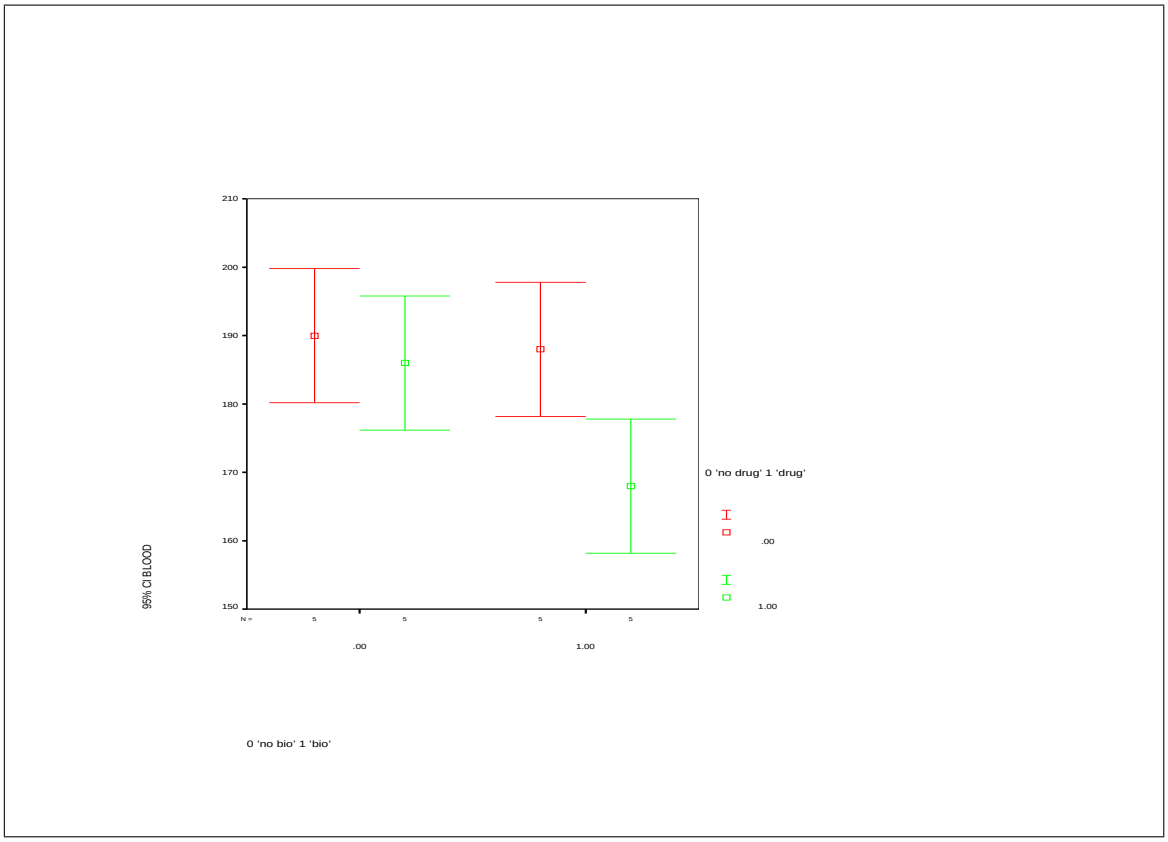
We now illustrate a factorial ANOVA with an example. The experiment involved the presence or absence of biofeedback (one factor) and the presence or absence of a new drug (second factor). The dependent variable was blood pressure (from Maxwell & Delaney, 1990). The structural model tested for these data was Equation 4-10.

The data are

bio & drug	bio only	drug only	neither
158	188	186	185
163	183	191	190
173	198	196	195
178	178	181	200
168	193	176	180

The cell means, marginal means (row and column means), and parameter estimates are (figure of means with 95% CI follows):

	biofeedback			
drug	present	absent	row means	α's
present	168	186	177	$\hat{\alpha} = -6$
absent	188	190	189	$\hat{\alpha} = 6$
col means	178	188		
$\hat{\beta}$'s	$\hat{\beta} = -5$	$\hat{\beta} = 5$	$\hat{\mu} = 183$	



The α's are computed by subtracting the grand mean from each row mean; the β's are computed by subtracting the grand mean from each column mean. The "interaction" estimates $\hat{\alpha\beta}_{ij}$ are computed as

$$\hat{\alpha\beta}_{ij} = \text{cell mean}_{ij} - \text{row mean}_i - \text{col mean}_j + \text{grand mean} \quad (4\text{-}12)$$

The four $\alpha\beta_{ij}$ estimates are

	biofeedback	
drug	present	absent
present	-4	4
absent	4	-4

The ϵs, or residuals, are computed the same way as before: observed score minus cell mean. Recall that residuals are what's left over from the observed score after the "model part" (grand mean, and all main effects and interactions) have been subtracted out. In symbols,

$$\epsilon_{ijk} = \text{observed - expected}$$

$$= \quad Y_{ijk} - \overline{Y}_{ij}$$

Each subject has his or her own residual term.

The data in the format needed for SPSS, the command syntax, and the ANOVA source table are:

```
Column 1:  bio-- 0 is not present, 1 is present
Column 2:  drug--0 is not present, 1 is present
Column 3:  blood pressure--dependent variable
Column 4:  recoding into four distinct groups (for later use)

Data are
1 1 158    1
1 1 163    1
1 1 173    1
1 1 178    1
1 1 168    1
1 0 188    2
1 0 183    2
1 0 198    2
1 0 178    2
1 0 193    2
0 1 186    3
0 1 191    3
0 1 196    3
0 1 181    3
0 1 176    3
0 0 185    4
0 0 190    4
0 0 195    4
0 0 200    4
0 0 180    4

  manova blood by biofeed(0,1) drug(0,1)
   /design biofeed,  drug, biofeed by drug.

  Tests of Significance for BLOOD using UNIQUE sums of squares
  Source of Variation          SS      DF        MS         F  Sig of F

  WITHIN CELLS            1000.00      16     62.50
  BIOFEED                  500.00       1    500.00      8.00      .012
  DRUG                     720.00       1    720.00     11.52      .004
  BIOFEED BY DRUG          320.00       1    320.00      5.12      .038
```

In this example all three effects (two main effects and interaction) are statistically significant at $\alpha = 0.05$. To interpret the three tests, we need to refer to the table of cell and marginal means (i.e., to find out which means are different from which and the direction of the effects). The same source table emerges in R:

```
summary(aov(blpr ~ bio * drug, data = blood.dat))


##             Df Sum Sq Mean Sq F value  Pr(>F)
```

```
## bio          1    500   500.0    8.00 0.01211 *
## drug         1    720   720.0   11.52 0.00371 **
## bio:drug     1    320   320.0    5.12 0.03792 *
## Residuals   16   1000    62.5
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Each effect in this source table has one degree of freedom in the numerator. Recall that contrasts have one degree of freedom in the numerator. This suggests that each of these three effects can be expressed equivalently through contrasts. In this case we have the main effect for biofeedback equivalent to the (1, 1, -1, -1) contrast, the main effect for drug therapy equivalent to the (1, -1, 1, -1) contrast, and the interaction equivalent to the (1, -1, -1, 1) contrast. Make sure you understand how the contrast weights correspond directly to the main effects and interaction tests above.

I'll now verify that these three contrasts lead to the same three tests as the factorial ANOVA. I tested these contrasts using ONEWAY. I needed to recode the design to make it into a one-way factorial. One method to "recode" is to enter a new column of numbers that uniquely specify each cell (i.e., create a new column in the data set that is the grouping variable with codes 1-4). The resulting output from the ONEWAY command is

```
oneway blood by treat(1,4)
  /contrast 1 1 -1 -1
  /contrast 1 -1 1 -1
  /contrast 1 -1 -1 1.
```

```
                                ANALYSIS OF VARIANCE
                                SUM OF          MEAN             F      F
        SOURCE           D.F.   SQUARES        SQUARES          RATIO  PROB.

 BETWEEN GROUPS             3    1540.0000      513.3333       8.2133  .0016
 WITHIN GROUPS             16    1000.0000       62.5000
 TOTAL                     19    2540.0000

            Grp 1       Grp 3
                 Grp 2       Grp 4

 CONTRAST  1   1.0   1.0  -1.0  -1.0
 CONTRAST  2   1.0  -1.0   1.0  -1.0
 CONTRAST  3   1.0  -1.0  -1.0   1.0

                                        POOLED VARIANCE ESTIMATE
                     VALUE      S. ERROR       T VALUE    D.F.       T PROB.
 CONTRAST  1       -20.0000        7.0711        -2.828    16.0         0.012
 CONTRAST  2       -24.0000        7.0711        -3.394    16.0         0.004
 CONTRAST  3       -16.0000        7.0711        -2.263    16.0         0.038

                                      SEPARATE VARIANCE ESTIMATE
                     VALUE      S. ERROR       T VALUE    D.F.       T PROB.
```

```
CONTRAST  1       -20.0000        7.0711        -2.828    16.0         0.012
CONTRAST  2       -24.0000        7.0711        -3.394    16.0         0.004
CONTRAST  3       -16.0000        7.0711        -2.263    16.0         0.038
```

By squaring the observed t values one gets exactly the same results as the three F's from the previous MANOVA output. For instance, take the t = -2.828 for contrast 1, square it, and you get 8.00, which is the same as the F for the main effect of biofeedback. This verifies that for a 2 $\times$ 2 design these three contrasts are identical to the tests in the factorial design. Also, what ONEWAY labels "between groups sums of squares" (1540) is the same as the sum of the three treatment sums of squares in the MANOVA (500 + 720 + 320). **Both analyses yield identical MSE and df for error**. In some cases, the separate variance t test will differ from the pooled variance t test, but in this contrived example all four cells have the identical standard deviation of 7.9 so the Welch correction is identical to the pooled version.

Important point about factorial designs and contrasts

The 2 $\times$ 2 factorial design breaks down into contrasts, as we have just seen. What does the above exercise demonstrate? The main point is that we can express a factorial, between-subjects design in terms of specific contrasts. How many contrasts does it take to complete a factorial design? The number of cells minus one (in other words, the sum of all the degrees of freedom associated with treatment effects and interactions). More complicated factorial designs such as 3 $\times$ 3 designs can also be partitioned into contrasts but the process is a bit trickier (i.e., getting the omnibus result from a set of orthogonal contrasts separately for each main effect and the interaction). I defer that discussion to a later time.

I presented one set of contrasts to analyze this design. But, there are other research questions one could ask and those questions would possibly imply different contrasts. One could imagine a researcher who just wanted to test whether getting a treatment differed from not getting a treatment. The planned contrast for that hypothesis is (1 1 1 -3). The researcher could also follow up the contrast with a Tukey test on the cell means (i.e., having found the above contrast to be significant it would make sense to "fish around" to see whether the three treatments differ from each other and whether the three treatments differ from the control). The output from the ONEWAY command for the (1 1 1 -3) contrast is listed below; the TUKEY pairwise comparisons on the **cell means** appears in Appendix 6. In order to get the Tukey test in SPSS on the cell means, one must convert the design into a one-way factorial as was done here (though the UNIANOVA procedure in SPSS can perform pairwise comparisons directly in a factorial design).

```
CONTRAST  4     -1.0  -1.0  -1.0   3.0

                                        POOLED VARIANCE ESTIMATE
                    VALUE      S. ERROR       T VALUE    D.F.       T PROB.
CONTRAST  4       28.0000       12.2474         2.286    16.0         0.036

                                      SEPARATE VARIANCE ESTIMATE
                    VALUE      S. ERROR       T VALUE    D.F.       T PROB.
CONTRAST  4       28.0000       12.2474         2.286     6.9         0.057
```

In Appendix 5 I show how to run contrasts directly in the MANOVA command. This is a complicated command so I'll explain the syntax in more detail latter. Corresponding R commands also appear in the appendix.

We will return to this example later because it raises some deep issues about contrasts and interactions.

7. Multiplicity

The factorial design produces more than one omnibus test per analysis. So new terminology is needed to conceptualize different error rates. Experiment-wise error rate is the overall error rate observed by the entire ANOVA. For example, the 2×2 factorial ANOVA above yielded three F tests. Obviously, if each is tested at $\alpha = 0.05$ the multiple comparison problem is present. Few researchers actually do a Bonferroni correction on main effects and interactions, though some recent textbooks argue that we should. A second concept is family-wise error, which refers to the error rate within a particular omnibus test (more correctly, the set of contrasts that one performs within a particular main effect or interaction). In the one-way ANOVA the distinction between family-wise and experiment-wise doesn't exist (both are equivalent) because there is only one omnibus test. Finally, we have the per-comparison error rate, which is the α level at which each single-df contrast is tested.

8. Contrasts and Post Hoc Tests

Performing contrasts and post hoc tests in factorial designs are (almost) identical to the way we did them for the one-way ANOVA. The subtle part is whether we are making comparisons between cell means or between marginal means. The difference creates a tiny change in the notation for the formulas. We need to take into account the appropriate number of observations that went into the particular means that are compared. The issue is merely one of proper book-keeping.

(a) Contrasts on Factorial Designs

Suppose you had a 3×3 factorial design and wanted to test the (1 -1 0) contrast for the row factor. One approach to testing this contrast treats the design as a one-way with 9 levels (the 9 cells). The contrast would be written out as (1, -1, 0, 1, -1, 0, 1, -1, 0). This "lumps" together cells 1, 4, 7 and compares them to cells 2, 5, 8 (ignoring the remaining three cells). The contrast has the correct number of subjects and the correct error term.

Another way to test this contrast would be to take the relevant marginal means (in this

example, the three row means) and apply the following formula to compute SSC

$$\frac{(\sum a_i \overline{Y}_i)^2}{\sum \frac{a_i^2}{s_i}} \tag{4-13}$$

where $\overline{Y}_i$ refers to the marginal means, i is the subscript over the marginal means, and s_i is the number of subjects that went into the computation of the i*th* marginal mean. Once you have SSC, then you divide it by the MSE from the factorial ANOVA table to get the F value (the degrees of freedom for the numerator is 1, the degrees of freedom in the denominator is the same as the error term).

(b) Post Hoc Tests on Factorial Designs

If you are performing pairwise comparisons of cell means, then the easiest thing to do is recode the independent variables into total number of cells and use the same code you would use for a one-way ANOVA. This means that the Tukey and Scheffe will be tested using the number of cells minus one. Few people would question this practice for the Tukey test but some people may argue that it is too conservative for the Scheffe. The logic of this critique would be correct if you limited yourself to interaction contrasts, then you could instead use the degrees of freedom for the interaction term, which is typically less than the number of cells minus 1 and corresponds to contrasts within the "interaction family". That is, the degrees of freedom for the interaction term is $(a-1)*(b-1)$ and this will be less than the number of cells $ab-1$ where a is the number of levels of Factor A and b is the number of levels of Factor B. Researchers usually want to test many contrasts so the practice I'm presenting of using "number of cells - 1" is not too far fetched. But it is good to be consistent so if you adopt the family-wise approach elsewhere in your analyses, then it may be good to limit yourself to interaction contrasts and base Tukey or Scheffe on the interaction degrees of freedom. Lately, though, I've been taking the perspective that if one wants to perform many post-hoc contrasts, then one might as well use the number of cells minus 1 ($ab-1$) instead of the interaction term ($(a-1)(b-1)$) and not worry about whether each contrast belongs to the main effect A family, main effect B family or interaction of A and B family.

However, if you want to perform post hoc tests on the marginal means (i.e. row or column means), you are stuck because most programs do not compute such tests. Just take out your calculator. The formulas for Tukey, SNK, and Scheffe in the factorial world are similar to the formulas in the one-way world. The subtle difference is, again, the number of subjects that went into the computation of the particular marginal means that are being compared and the number of marginal means. For both Tukey and Scheffe, common practice is to focus on the number of levels of the factor for that marginal as the basis for the relevant degrees of freedom.

Recall that for the one-way ANOVA the (equal n) formula for Tukey's W is

$$W = q_{\alpha}(T,v)\sqrt{\frac{MSE}{n}}. \tag{4-14}$$

The Tukey test for *marginal means* in a factorial design is

$$W = q_{\alpha}(k,v)\sqrt{\frac{MSE}{s}} \tag{4-15}$$

where k is the number of means in the "family" (i.e., the number of marginal means for row, or number of marginal means for column) and s is the total number of subjects that went into the computation of the marginal mean. The MSE and the associated df's v is straight from the factorial source table. A numerical example is given on page 4-27. Comparable changes are made to other posthoc pairwise tests such as the SNK.

In a similar way, the Scheffe test becomes

$$S = \sqrt{V(\hat{I})}\sqrt{df1\ F_{\alpha;\ df1,\ df2}} \tag{4-16}$$

Recall that

$$V(\hat{I}) = MSE\sum\frac{a_i^2}{s_i} \tag{4-17}$$

where df1 depends on the kind of test (either df for treatment A, df for treatment B, or df for interaction—all these df appear in the ANOVA source table) and s is the total number of subject that went into the computation of the relevant means.

Numerical examples of both Tukey and Scheffe are given starting on page 4-27. The key point is that the formulas for Tukey and Scheffe have the same structure but we need to keep track of the number of means that are being compared (which may not be the same as the number of groups) and the corresponding sample sizes that fed into each of those means.

(c) Planned Comparisons on Randomized Block and Latin Square Designs

Not easily done by most programs, so we'll do them by hand. Exactly the same way as the factorial design for marginal means. Use the MSE from the source table you constructed to test the omnibus effects (i.e., the source table corresponding to the structural model having no interactions).

(d) Effect size for contrasts in factorial designs

effect size

The effect size of a contrast (i.e., the analog to R^2 (see LN2), the percentage of variance accounted for by that contrast) is given by

$$\sqrt{\frac{\text{SSC}}{\text{SSC + SSE}}} \tag{4-18}$$

where SSC is the sum of squares for that contrast given by $\hat{\text{I}}^2/\sum \frac{a_i^2}{n_i}$ and SSE is the sum of squares error from the full source table. Equation 4-18 is the same quantity that I called p_c/(1-p_b) in Lecture Notes #3. Some people refer to Expression 4-18 as a partial R^2 because the sum of squares from the other contrasts are not included (or "partialled out").

A simpler and equivalent way to compute the contrast effect size is to do it directly from the t-test of the contrast. For any contrast, plug the observed t into this formula and you'll get effect size:

$$\sqrt{\frac{t^2}{t^2 + \text{df error}}} \tag{4-19}$$

where df error corresponds to the degrees of freedom associated with the MSE term of the source table. Equation 4-19 is useful when all you have available is the observed t, as in a published study. For example, turning to the (-1 -1 -1 3) contrast on page 4-13 we see the observed t was 2.286 with df = 16. Plugging these two values into the formula for effect size of the contrast we have

$$\begin{aligned} r &= \sqrt{\frac{2.286^2}{2.286^2 + 16}} \\ &= .496 \end{aligned}$$

9. Caveat

You need to be careful when making comparisons between cell means in a factorial design. The problems you may run into are not statistical, but involve possible confounds between the comparisons. Suppose I have a 2 $\times$ 2 design with gender and self-esteem (low/high) as factors. It is not clear how to interpret a difference between low SE/males and high SE/females because the comparison confounds gender and self-esteem. Obviously, these confounds would occur when making comparisons that crisscross rows and columns. If one stays within the same row or within the same column, then these potential criticisms do not apply as readily. [2] An example where crisscrossing rows and columns makes sense is the 2 $\times$ 2 design presented at the beginning of today's notes (the biofeedback and drug study). It makes perfect sense to compare the drug alone condition to the biofeedback alone condition.

[2] When you limit yourself to comparisons within a row or column you limit the number of possible comparisons in your "fishing expedition". There are some procedures that make adjustments, in the same spirit as SNK, to the Tukey and Scheffe tests to make use of the limited fishing expedition (see Cicchetti, 1976, *Psychological Bulletin, 77*, 405-408). But I'm not a fan of those tests and prefer original Tukey or Scheffe.

So, sometimes it is okay to make comparisons between cell means, just approach the experimental inferences with caution because there may be confounds that make interpretation difficult.

In a similar spirit, some people get fussy about interpreting main effects if there is also a significant interaction. Their logic is one should avoid intepreting main effects if the interaction is significant because the interaction means the pattern for one factor is qualified by the other factor(s) involved in the interaction. I think that is too strict a position because there are cases where a main effect can be interpreted even if the interaction is significant. We will see an example of this in a few pages with hte 2 x 3 study on lecture format. We will be able to interpret the main effect for the method of instruction factor clearly because across the board computer-aided instruction yielded higher scores than standard lecture, even though there is an interaction. The other main effect (for the content of the course factor) is difficult to interpret because we can't say that social science had lower scores than physical science and history because of the interaction—that pattern, while it shows up in the main effect—is mostly driven by the standard instruction condition not the computer-aided instruction condition. So, sometimes "across the board" statements are justified (we sometimes can interpret a main effect in the presence of an interaction) and other times they are not (we sometimes can't interpret a main effect in the presence of an interaction).

10. Factorial ANOVA continued...

Earlier we saw that for the special case of the 2×2 factorial design the three omnibus tests (two main effects and interaction) can be expressed in terms of three contrasts. This complete equivalence between omnibus tests and contrasts only holds if every factor has two levels (i.e., a 2×2 design, a $2 \times 2 \times 2$ design, etc.). More generally, whenever the omnibus test has one degree of freedom in the numerator, then it is equivalent to a contrast. However, if factors have more than two levels, such as in a 3×3 design, then the omnibus tests have more than one degree of freedom in the numerator and cannot be expressed as single contrasts. As we saw in the one-way ANOVA, the sums of squares between groups can be partitioned into separate "pieces", SSC_i, with orthogonal contrasts. If the omnibus test has two or more degrees of freedom in the numerator (say, T - 1 df's), then more than one contrast is needed (precisely, T - 1 orthogonal contrasts).

One conclusion from the connection between main effects/interactions and contrasts is that main effects and interactions are "predetermined" contrasts. That is, main effects and interactions are contrasts that make sense given a particular factorial design. For example, in a 2×2 we want to compare row means, column means, and cell means. Most statistics packages are programmed to create the relevant omnibus tests in factorial designs from sets of contrasts applied separately to each factor (and by implication, the interaction terms), though not all programs use orthogonal sets making interpretation sometimes difficult especially across programs. Indeed, one way to view a factorial designs is that they commit the data analyst to a predefined set of contrasts applied separately to each factor.

One is not limited to testing main effects and interactions. Sometimes other partitions of the sums of squares, or contrasts, may make more sense for a particular research question or study design. Consider the biofeedback/drug experiment discussed earlier. The three degrees of freedom could be partitioned as a) -1 -1 -1 3 to test if there is an effect of receiving a treatment, b) 2 -1 -1 0 to test if receiving both treatments together differs from the sum of receiving both treatments separately, and c) 0 -1 1 0 to test if there is a difference between drug alone and biofeedback alone. These three contrasts are orthogonal to each other. The oneway ANOVA output for these three contrasts is:

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	3	1540.0000	513.3333	8.2133	.0016
WITHIN GROUPS	16	1000.0000	62.5000		
TOTAL	19	2540.0000			

CONTRAST COEFFICIENT MATRIX

	Grp 1	Grp 2	Grp 3	Grp 4
CONTRAST 4	-1.0	-1.0	-1.0	3.0
CONTRAST 5	2.0	-1.0	-1.0	0.0
CONTRAST 6	0.0	-1.0	1.0	0.0

POOLED VARIANCE ESTIMATE

	VALUE	S. ERROR	T VALUE	D.F.	T PROB.
CONTRAST 4	28.0000	12.2474	2.286	16.0	0.036
CONTRAST 5	-38.0000	8.6603	-4.388	16.0	0.000
CONTRAST 6	-2.0000	5.0000	-0.400	16.0	0.694

SEPARATE VARIANCE ESTIMATE

	VALUE	S. ERROR	T VALUE	D.F.	T PROB.
CONTRAST 4	28.0000	12.2474	2.286	6.9	0.057
CONTRAST 5	-38.0000	8.6603	-4.388	8.0	0.002
CONTRAST 6	-2.0000	5.0000	-0.400	8.0	0.700

It is instructive to compare the results of these three contrasts to the results previously obtained for the main effects and interaction. Pay close attention to the difference between the two models and the null hypotheses they are testing.

11. Interactions

The null hypothesis for the standard interaction (i.e., the 1 -1 -1 1 contrast) in the biofeedback example is

$$\text{Ho:} \quad \mu_{\text{drug\&bio}} + \mu_{\text{neither}} - \mu_{\text{drug}} - \mu_{\text{bio}} = 0 \tag{4-20}$$

This can be written as

$$\text{Ho:} \quad (\mu_{\text{drug\&bio}} - \mu_{\text{neither}}) = (\mu_{\text{drug}} - \mu_{\text{neither}}) + (\mu_{\text{bio}} - \mu_{\text{neither}}) \tag{4-21}$$

In other words, is the effect of receiving both drug and biofeedback treatments simultaneously relative to receiving nothing equal to the additive combination of 1) receiving drug treatment alone relative to nothing and 2) receiving biofeedback treatment alone relative to nothing? Said still another way, is the effect of receiving both treatments simultaneously equal to the sum of receiving each treatment separately (each of those three terms are relative to receiving no treatment)? The presence of a significant interaction implies a nonadditive effect.

Equation 4-20 is different than the (0 -1 -1 2) contrast comparing both treatments alone to the two administered together. This contrast is a special case of the interaction, i.e., when $\mu_{\text{drug\&bio}} = \mu_{\text{neither}}$ then the two tests are the same. You need to be careful not to use the language of "interactions" if you deviate from the standard main effect/interaction contrasts. Sloppy language in calling contrasts such as (1 1 1 -3) "interaction contrasts" has led to much debate (e.g., see the exchange between Rosenthal, Abelson, Petty and others in *Psychological Science*, July, 1996)[3].

We will return to the issue of additivity later when discussing the role of transformations in interpreting interactions.

12. Using plots to interpret main effects and interactions as well as checking assumptions

With plots it is relative easy to see the direction and magnitude of the effects. If you place confidence intervals around the points (using the pooled error term) you can even get some sense of the variability under the hypothetical "repeated sampling" interpretation of confidence intervals developed in LN1, and visualize the patterns of statistical significance.

Some people view interactions as the most important test one can do. One function that interactions serve is to test whether an effect can be turned on or off, or if the direction of an effect can be reversed. Interactions allow one to test whether one factor moderates the effects of a second factor.

SPSS doesn't offer great graphics. You might want to invest time in learning a different program that offers better graphics options such as the statistical program R, or at least creating graphs in a spreadsheet like Excel. Here are some SPSS commands you may find useful:

[3] A related point has been made in the intersectionality literature (John Jackson et al, 2016) where their argument is that the intersectionality of, say, race and socioeconomic status (SES) should not be tested as an interaction between race and SES but as a series of contrasts that compare, for example, having two risk factors to no risk factors ("joint disparity") or contrasts over groups with one risk factor. Jackson et al refer to the interaction contrast in EQ 4-21 as "excess intersectional disparity" because each of the groups with one or two components is compared separately to the group without a disparity component. They provide measures based on simple contrasts that teases apart the effect of both risk factors alone and the effect of both risk factors being present. While their focus was on intersectionality and health disparities, their approach can be applied more generally and shows again the important relation between research questions and contrasts.

For a plot of only means in a two-way ANOVA:

```
graph
 /line(multiple)=mean(dv) by iv1 by iv2.
```

where dv represents the dependent variable, and iv1, iv2 represent the two independent variables.

plot errorbars

To have errorbars (i.e., 95% CIs around the means) do this instead:

```
graph
 /errorbar(ci 95) = dv by iv1 by iv2.
```

The first command uses mean(dv) but the second command uses dv. Ideally, it would be nice to have the first plot with confidence intervals too, but I haven't figured out how to do that in SPSS. Also, this shows how limited the menu system—on my version of SPSS the menu system only permits one independent variable per plot, but it is possible to get two independent variables in the same plot if you use syntax.

boxplot; spread and level plots

The boxplot and the spread and level plot to check for assumption violations and possible transformations can be done in a factorial design as well. The syntax is

```
examine dv by iv1 by iv2
 /plots boxplot spreadlevel.
```

Examples of various plots will be given in lecture.

Relevant R code for graphs in factorial designs is more involved and is developed in Appendix 8.

13. An example of a 2×3 factorial design

An example putting all of these concepts to use will be instructive.

Consider the following experiment on learning vocabulary words (Keppel & Zedeck, 1989). A researcher wants to test whether computer-aided instruction is better than the standard

lecture format. He believes that the subject matter may make a difference as to which teaching method is better, so he includes a second factor, type of lecture. Lecture has three levels: physical science, social science, and history. Fifth graders are randomly assigned to one of the six conditions. The same sixty "target" vocabulary words are introduced in each lecture. All subjects are given a vocabulary test. The possible scores range from 0-60.

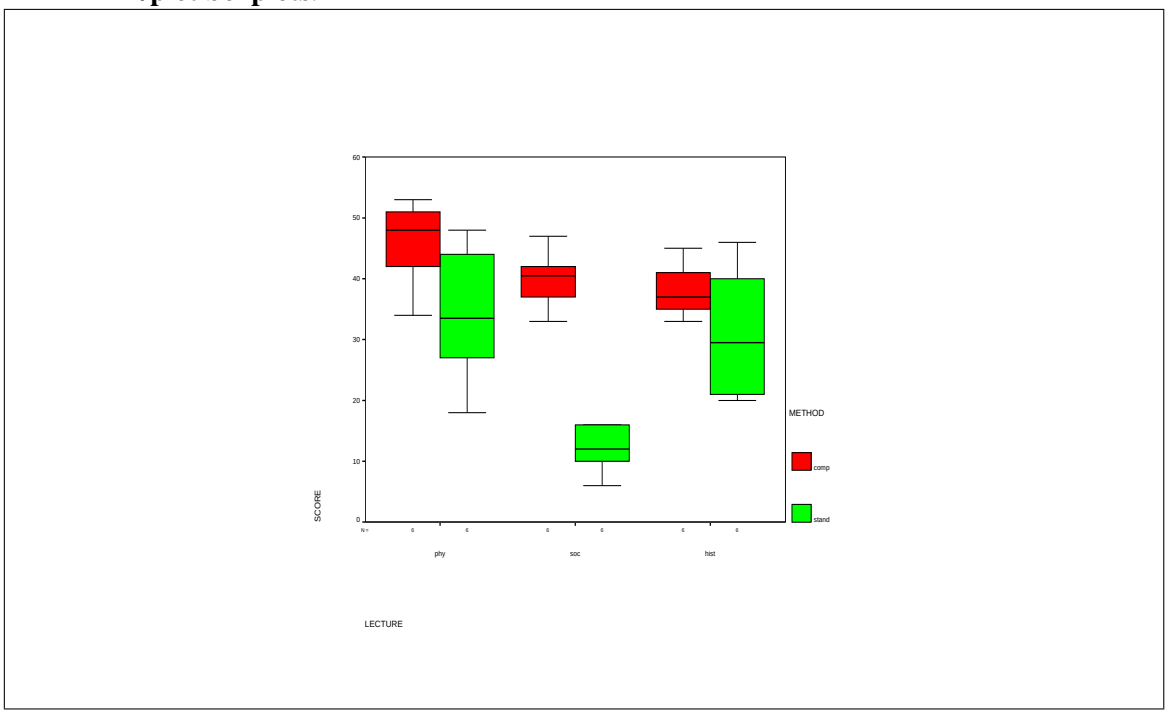
The data in the format for SPSS are listed below. The first column codes type of lecture, the second column codes method, the third column is the dependent measure, and the fourth column is a "grouping variable" to facilitate contrasts in ONEWAY. In Appendix 8 I show smarter ways of doing this coding that doesn't require re-entering codes.

```
1 1 53 1
1 1 49 1
1 1 47 1
1 1 42 1
1 1 51 1
1 1 34 1
1 2 44 2
1 2 48 2
1 2 35 2
1 2 18 2
1 2 32 2
1 2 27 2
2 1 47 3
2 1 42 3
2 1 39 3
2 1 37 3
2 1 42 3
2 1 33 3
2 2 13 4
2 2 16 4
2 2 16 4
2 2 10 4
2 2 11 4
2 2 6 4
3 1 45 5
3 1 41 5
3 1 38 5
3 1 36 5
3 1 35 5
3 1 33 5
3 2 46 6
3 2 40 6
3 2 29 6
3 2 21 6
3 2 30 6
3 2 20 6
```

One advantage of using the MANOVA command (in addition to those already mentioned) is that MANOVA has a built in subcommand to print boxplots[4] :

[4]It is possible to get cell by cell boxplots and normal plots using the examine command: examine dv by iv1 by iv2 /plot npplot, boxplot. Note the double use of the BY keyword, omitting the second BY produces plots for marginal means instead of cell means.

/plot boxplots.

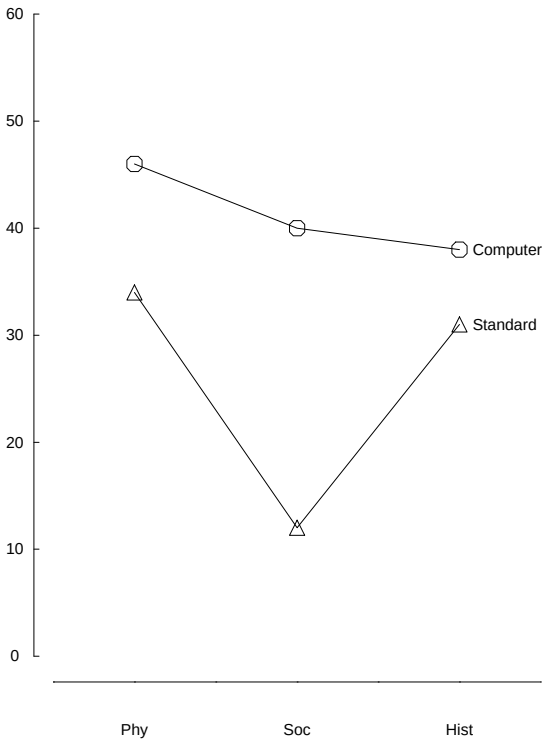


The equality of variance assumption appears okay—not great, but we can live with it because of the small sample size. No evidence of outliers.

The cell and marginal means are

	Lecture			
	phy	soc	hist	row marginal
comp	46	40	38	41.33
standard	34	12	31	25.67
column marginal	40	26	34.5	

The means plotted:



The structural model is

$$Y \quad = \quad \mu + \alpha + \beta + \alpha\beta + \epsilon \tag{4-22}$$

The treatment effects are ($\hat{\mu} = 33.5$):

		Lecture		
		phy	soc	hist
comp		$\hat{\alpha}_1 = 7.83$	$\hat{\alpha}_1 = 7.83$	$\hat{\alpha}_1 = 7.83$
		$\hat{\beta}_1 = 6.50$	$\hat{\beta}_2 = -7.50$	$\hat{\beta}_3 = 1.00$
		$\hat{\alpha\beta}_{11} = -1.83$	$\hat{\alpha\beta}_{12} = 6.17$	$\hat{\alpha\beta}_{13} = -4.33$
standard		$\hat{\alpha}_2 = -7.83$	$\hat{\alpha}_2 = -7.83$	$\hat{\alpha}_2 = -7.83$
		$\hat{\beta}_1 = 6.50$	$\hat{\beta}_2 = -7.50$	$\hat{\beta}_3 = 1.00$
		$\hat{\alpha\beta}_{21} = 1.83$	$\hat{\alpha\beta}_{22} = -6.17$	$\hat{\alpha\beta}_{23} = 4.33$

The ANOVA source table:

```
manova score by lecture(1,3) method(1,2)
 /design lecture method lecture by method.

Tests of Significance for SCORE using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN CELLS           1668.00      30     55.60
LECTURE                1194.00       2    597.00     10.74      .000
METHOD                 2209.00       1   2209.00     39.73      .000
LECTURE BY METHOD       722.00       2    361.00      6.49      .005
```

All three effects (two main effects and interaction) are statistically significant at $\alpha = 0.05$. Two of the three F tests are omnibus tests that are not easy to interpret without further analysis. Which ones are those? (Hint: just look at the df column in the source table.)

A variant of the previous command prints out more information such as boxplot as well as marginal and cell means (though fine to use the EXAMINE command for all the additional assumption-checking plots).

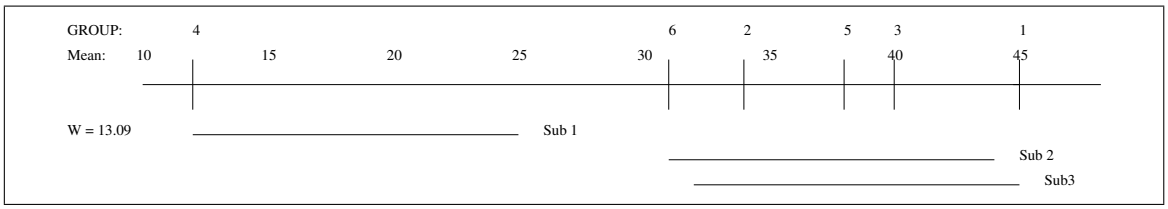
```
manova score by lecture(1,3) method(1,2)
/print error(stddev)
/omeans table(method) table(lecture) table (lecture by method)
/plot boxplot
/design lecture method lecture by method.
```

Now, suppose you had planned to test the contrast comparing the history conditions to the average of the physical sciences and social sciences. An easy way to do this would be to create a new grouping variable, ranging from 1 to 6, that codes the six cells. We conveniently have that variable entered in the raw data file. The contrast can be tested through the ONEWAY command. Suppose you also want to test the difference between the physical sciences and social sciences conditions, but only within the standard lecture format. That could also be tested within the ONEWAY command. You may also want to test all possible pairwise differences between cells (using Tukey). The following command illustrates these tests.

```
oneway score by groups(1,6)
  /contrasts = 1 1 1 1 -2 -2
  /contrasts = 1 1 -1 -1 0 0
  /contrasts = 1 -1 1 -1 1 -1
  /contrasts = 0 1 0 -1 0 0
  /ranges= tukey.

                              ANALYSIS OF VARIANCE
                              SUM OF          MEAN             F      F
       SOURCE           D.F.  SQUARES        SQUARES          RATIO  PROB.
BETWEEN GROUPS             5  4125.0000       825.0000       14.8381  .0000
WITHIN GROUPS             30  1668.0000        55.6000
TOTAL                     35  5793.0000

 CONTRAST COEFFICIENT MATRIX
```

```
              Grp 1       Grp 3       Grp 5
                    Grp 2       Grp 4       Grp 6
 CONTRAST  1   1.0   1.0   1.0   1.0  -2.0  -2.0
 CONTRAST  2   1.0   1.0  -1.0  -1.0   0.0   0.0
 CONTRAST  3   1.0  -1.0   1.0  -1.0   1.0  -1.0
 CONTRAST  4   0.0   1.0   0.0  -1.0   0.0   0.0

                                        POOLED VARIANCE ESTIMATE
                     VALUE      S. ERROR       T VALUE    D.F.       T PROB.
 CONTRAST  1        -6.0000       10.5451        -0.569    30.0         0.574
 CONTRAST  2        28.0000        6.0882         4.599    30.0         0.000
 CONTRAST  3        47.0000        7.4565         6.303    30.0         0.000
 CONTRAST  4        22.0000        4.3050         5.110    30.0         0.000

                                      SEPARATE VARIANCE ESTIMATE
                     VALUE      S. ERROR       T VALUE    D.F.       T PROB.
 CONTRAST  1        -6.0000       10.8812        -0.551    12.3         0.591
 CONTRAST  2        28.0000        5.8878         4.756    12.1         0.000
 CONTRAST  3        47.0000        7.4565         6.303    18.9         0.000
 CONTRAST  4        22.0000        4.7610         4.621     6.2         0.003

THE RESULTS OF ALL PAIRWISE COMPARISONS BETWEEN CELLS.  NOTE THAT THE q
VALUE USES 6 DF'S IN THE NUMERATOR (NUMBER OF CELL MEANS) AND n IS THE NUMBER
OF SUBJECTS IN THE CELL.

 TUKEY-HSD PROCEDURE
 RANGES FOR THE 0.050 LEVEL -

          4.30   4.30   4.30   4.30   4.30


   HOMOGENEOUS SUBSETS   (SUBSETS OF GROUPS, WHOSE HIGHEST AND LOWEST MEANS
                          DO NOT DIFFER BY MORE THAN THE SHORTEST
                          SIGNIFICANT RANGE FOR A SUBSET OF THAT SIZE)

 SUBSET  1
 GROUP       Grp 4
 MEAN       12.0000

 SUBSET  2
 GROUP       Grp 6        Grp 2        Grp 5        Grp 3
 MEAN       31.0000      34.0000      38.0000      40.0000

 SUBSET  3
 GROUP       Grp 2        Grp 5        Grp 3        Grp 1
 MEAN       34.0000      38.0000      40.0000      46.0000
```

One way to understand the Subsets output is to use the picture below. Order the means along the number line and draw a line that is as long as W (which here equals 13.09) and has one end on the smallest mean. Group means not included within that line are statistically different from the smallest mean. In this case, all groups are different from Group 4, creating the first subset. Then move the "line marker" over to the next mean. Any mean not included in the span of the line marker is different from the second mean—Group 1 is different from Group 6 (we've already determined that Group 4 differs from Group 6 in the previous subset). Notice that all groups within the line marker are not different from Group 6 by Tukey. Continue moving the "line marker" to each new Group until the upper end of the line marker matches the greatest mean; there we see that Group 6 differs from Group 1 (and also Group 4 differs from Group 1, but we already knew that from Subset 1).



Suppose you want to test all possible pairwise differences between the three column means (lecture). Perform a Tukey test ($\alpha = 0.05$) using the MSE from the ANOVA source table, look up the q value using degrees of freedom in the numerator equal to 3 (there are three marginal means), and be sure to take into account the correct number of subjects that went into that computation of the marginal mean. Note that for this case on marginal means we will have a different W than in the previous case where we looked at cell means (different number of means, different number of subjects per mean)—recall the discussion on page 4-16 that explained k and s (the number of means and the number of observations per mean, respectively).

$$\begin{aligned} W &= q(k,v)\sqrt{\frac{MSE}{s}} \\ &= 3.49\sqrt{\frac{55.6}{12}} \\ &= 7.51 \end{aligned} \tag{4-23}$$

Thus, any (absolute value of the) pairwise difference must exceed 7.51 to be significant with an overall $\alpha = 0.05$. Looking at the column marginals, we see that physical v. social and social v. history are significant comparisons, but physical v. history is not.

An example of a Scheffe-corrected contrast test: suppose we had not planned the (1, 1, -2) contrast on the lecture marginal means but got the idea to test it after seeing the results of the study. We could perform a Scheffe test to adjust for the post hoc testing. One indirect way of computing the Scheffe test in SPSS or R is to convert the problem into a one-way ANOVA. However, even though we are implementing the contrast on six cells, we intend to interpret the contrast as a main effect over three marginal means, hence k = 3. Plugging numbers from the SPSS output into the Scheffe formula we have

$$\begin{aligned} S &= \sqrt{V(\hat{I})(k\text{-}1)F_{\alpha,k\text{-}1,\,v}} \\ &= 10.55\sqrt{2 \times 3.32} \\ &= 27.19 \end{aligned} \tag{4-24}$$

The estimate of the contrast value is -6, so we fail to reject the null hypothesis under the Scheffe test since the absolute value is less than S. The estimate of the contrast ($\hat{I}$) as well as its standard error (10.55) are available in the ONEWAY output. Note that the $\hat{I}$ and se($\hat{I}$) correspond to the (1 1 1 1 -2 -2) contrast on the cell means, but I still plug in k = 3 into the Scheffe formula because the correction for multiplicity applies to the three marginal means

(not the six cells that I used merely as a computational shortcut). Equivalently, I could have computed the (1 1 -2) on the marginal means directly by hand. This yields an $\hat{I}$ = -3 and a se($\hat{I}$) = 5.27 (this se is based on the three marginal means). The Scheffe test computed on the marginals in this manner is identical because the two sets of numbers (the pair from the SPSS output of $\hat{I}$ -6 and se($\hat{I}$) of 10.55 and the respective pair I computed by hand on the marginals of -3 and 5.27) differ by a factor of 2, so tests regarding whether $\hat{I}$ is greater than S are the same.

There is a philosophical "bias" that I have about how we should correct for multiple comparisons in a factorial design. I used to follow a correct-within-each-family approach, also known as "family-wise", so we correct within each main effect, within each set of interactions, etc. This compares with another approach that doesn't consider family and just corrects for all possible tests in a design. I know it can be confusing if we sometimes use as a shortcut method the approach of treating a factorial design as a one-way ANOVA with the same number of cells. But if we conduct the one-way ANOVA in terms of the set of contrasts that correspond to main effects and interactions, we are indeed creating a set of contrasts based on particular families so one could argue that it makes sense to correct for multiple contrasts accordingly by family. It requires careful attention for each contrast and accounting whether it is a main effect contrast or an interaction contrast. Or we can just through up our hands and say we have cell means, want to apply contrasts on those cell means and not worry about whether the contrasts belong to main effect families or interaction families. This problem continues to be hotly debated in the literature.

14. Working out contrasts in factorial designs

The 2x3 lecture example has six cells so we know there are 5 orthogonal contrasts. The source table tells us there is 1 df for the main effect of lecture, 2 dfs for the main effect of content and 2 dfs for the interaction, for a total of 5. Let's work through those contrasts.

Let's say I use (1, -1) for the two levels of lecture. I can think of that as a contrast on the marginal means as in

	Lecture			
	phy	soc	hist	row marginal
comp				1
standard				-1
column marginal				

I can equivalently represent it as a contrast over six cells that has identical values across a row because we are after all comparing row 1 to row 2.

	Lecture			
	phy	soc	hist	row marginal
comp	1	1	1	
standard	-1	-1	-1	
column marginal				

For the content factor there are 3 groups so we need two contrasts. Let's say we select (1, 0, -1) and (1, -2, 1). Each of these can be represented equivalently as either contrasts applied to the three column means or contrasts applied to the six cells.

In column format we have

	Lecture			
	phy	soc	hist	row marginal
comp				
standard				
column marginal	1	0	-1	

and

	Lecture			
	phy	soc	hist	row marginal
comp				
standard				
column marginal	1	-2	1	

The same two contrasts in cell format (i.e., across the six cells) can be expressed this way:

	Lecture			
	phy	soc	hist	row marginal
comp	1	0	-1	
standard	1	0	-1	
column marginal				

and

	Lecture			
	phy	soc	hist	row marginal
comp	1	-2	1	
standard	1	-2	1	
column marginal				

So far, I've listed out three of the five contrasts. We still have two more corresponding to the interaction contrasts. The easiest way to see this to take the perspective of the cell format (i.e., the contrasts expressed as weights over the six cells). We compute all possible products of a main effect contrast from one factor with the main effect contrast from the other factor. In this example there is one main effect contrast for the lecture factor and two main effect contrasts for the content factor, this leads to two products. The first one is the product of the lecture contrast with the first content contrast:

	Lecture			
	phy	soc	hist	row marginal
comp	1*1=1	0*1=0	-1*1=-1	
standard	1*-1=-1	0*-1=0	-1*-1=1	
column marginal				

and the second interaction contrast is

	Lecture			
	phy	soc	hist	row marginal
comp	1*1=1	-2*1=-2	1*1=1	
standard	1*-1=-1	-2*-1=2	1*-1=-1	
column marginal				

That completes an example set of five contrasts possible in the 6 cells formed by the 2 x 3 design.

For completeness, the technical matrix operation used when creating interaction contrasts by multiplying contrasts defined on the marginals is the Kronecker product. R has it as a command and here are the two interaction contrasts I defined above written as Kronecker products (listed out as the three row 1 values then the three row 2 values:

```
kronecker(c(1, -1), c(1, 0, -1))
```

```
## [1]  1  0 -1 -1  0  1
```

```
kronecker(c(1, -1), c(1, -2, 1))
```

```
## [1]  1 -2  1 -1  2 -1
```

15. Expected Mean Square Terms in a Factorial Design

Recall that the F test is a ratio of two variances. In a two-way ANOVA there are three F tests (two main effects and one interaction). The estimates of the variance are as follows (for cells having equal sample sizes)

$$\text{MSE} \quad \text{estimates} \quad \sigma_\epsilon^2 \tag{4-25}$$

$$\text{MSA} \quad \text{estimates} \quad \sigma_\epsilon^2 + \frac{\text{nb}\sum \alpha^2}{\text{a}-1} \tag{4-26}$$

$$\text{MSB} \quad \text{estimates} \quad \sigma_\epsilon^2 + \frac{\text{na}\sum \beta^2}{\text{b}-1} \tag{4-27}$$

$$\text{MSAB} \quad \text{estimates} \quad \sigma_\epsilon^2 + \frac{\text{n}\sum \alpha\beta^2}{(\text{a}-1)(\text{b}-1)} \tag{4-28}$$

where n is the cell sample size, a is the number of levels for factor A, and b is the number of levels for factor B.

Recall the "don't care + care divided by don't care" logic introduced for the one way ANOVA. Each of the three omnibus tests in the two way factorial ANOVA uses the MSE term as the error term. In Lecture Notes 5 we will encounter cases where the MSE term will not be the appropriate error term.

Here are a couple of references that discuss details of how to compute the expected mean square terms. This could be useful as we introduce more complicated designs in ANOVA. In class and in these lecture notes I'll merely assert the expected mean square terms (and you can memorize them), but those students interested in how to derive the expected mean square terms can consult these papers. Millman and Glass (1967), Rules of thumb for writing the ANOVA table, *Journal of Educational Measurement, 4,* 41-51; Schultz (1955). Rules of thumb for determining expectations of mean squares in analysis of variance. *Biometrics, 11,* 123-135. There are more if you do a Google search, but these two are classics.

16. Extensions to more than two factors

There can be any number of factors. For instance, a $2 \times 2 \times 2$ ANOVA has three factors each with two levels. It contains three main effects, three two-way interactions, and one three-way interaction. The main effects and two-way interactions are interpreted exactly the same way

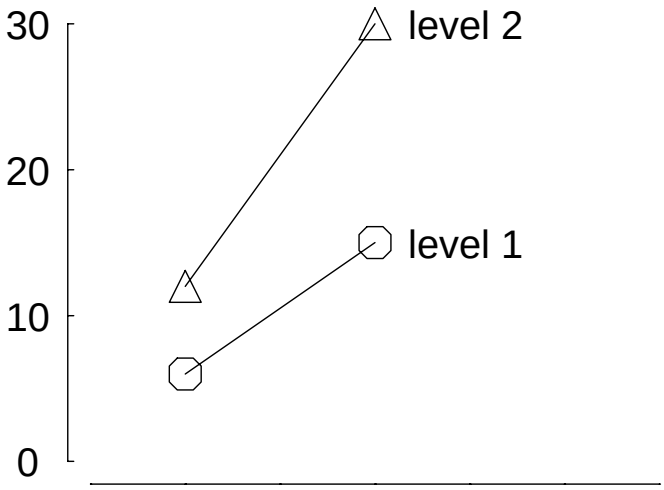
as with a 2×2 design. For instance, the main effect of the first factor compares just two means, the two marginal means of that factor collapsing over all other cells in the design. The new idea is the three-way interaction. One way to interpret a three-way interaction is to decompose it to smaller parts. For instance, take just the four cells where the third factor is at the first level, which creates a 2×2 interaction between the four remaining means. Interpret that interaction. Now compute the 2×2 interaction for the cells when the third factor is at the second level. Interpret that two-way interaction. A statistically significant three-way interaction means that the two-way interaction tested at one level of the third factor differs from the two-way interaction tested at the second level of the third factor.

This logic extends to any size design, such as a $3 \times 4 \times 2$ means there are three factors, where the first factor has three levels, the second factor has four levels and the third factor has two levels.

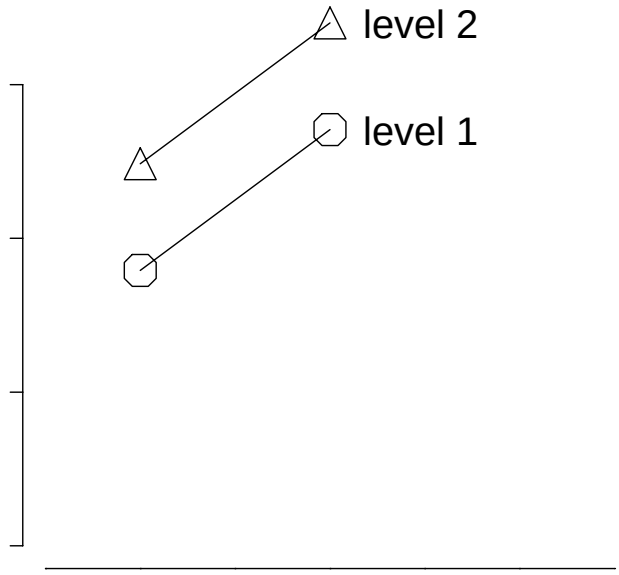
17. Interactions and Transformations

In a previous lecture we discussed transformation as a remedial technique for violations of the equality of variance assumption and correcting skewed distributions. Another application of transformations is to produce additivity. That is, sometimes it is possible to "eliminate", or reduce, an interaction by applying an appropriate transformation. In some applications an additive model is easier to interpret than a model with an interaction. Sometimes interactions are statistically significant because we are applying the wrong structural model to the data set.

For example, consider the 2×2 plot below. The lines are not parallel so we suspect an interaction (whether or not the interaction is is statistically significant is a question of power).



But, watch what happens when we apply a log transformation to the dependent variable (i.e., we transform the y-axis to log units). The non-parallelism goes away.



Why did this happen? The first plot shows the lines are not parallel. We hastily conclude that an interaction is present. However, using the fact that log(xy) = log(x) + log(y), if we had applied the log transform we would have converted the multiplicative effect into an additive

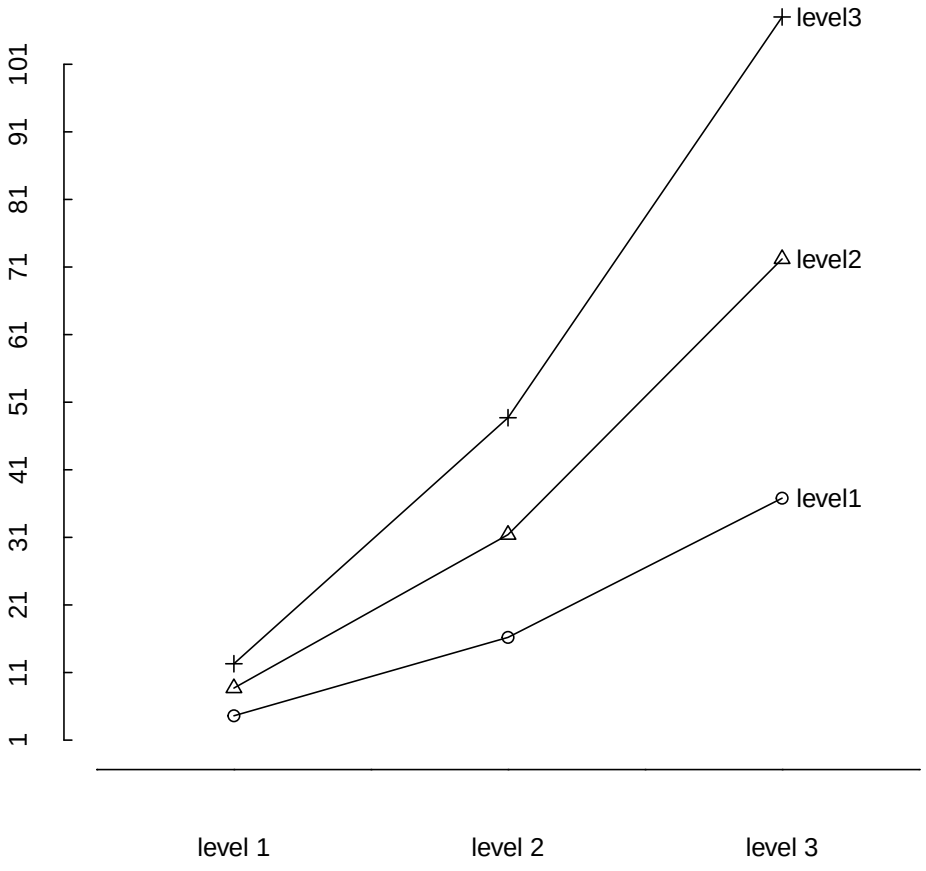
effect. For example, the lower left cell is log(6) = log(2) + log(3). Perhaps the investigator decided to use a log transformation in order to satisfy the equality of variance assumption. You can see that the decision to use a transformation to improve one assumption on, say, equality of variance, can influence the presence or absence of the interaction term.

Interactions are scale dependent

Which scale is correct? There is a sense in which they both are. The point is that what appears as an interaction may really be a question of having the right scale (i.e., having the right model). You can see how two researchers, one who works from the first, untransformed plot and one who works from the second, transformed plot, can reach different conclusions about the interaction. Unfortunately, there is no general statement that could be made to guide researchers about which transformations are permissible and under which situations. This issue has been covered somewhat in the literature on binary dependent variables such as logistic regression, but the issue of scale dependency of interactions for other types of dependent variables like normally distributed variables has been mostly ignored.

Let's do another example. Imagine Galileo Galilei studying the relation between distance, time, and acceleration using ANOVA[5]. What would that look like? How about a 3×3 design with three levels of time (shortest, shorter, short) and three levels of acceleration (little gravity, some gravity, more gravity; say by varying the angle of the incline). He rolls balls down inclines and measures the distance traveled. Here are some simulated results.

[5]There are many legends about Galileo and how he accomplished his science given the instruments he had available. For instance, when he collected data for the incline experiments, legend has it that he sang opera to mark time, and when he tested his intuitions about pendulums—that the period of the pendulum is independent of the mass of the object at the end—he used his pulse to mark time. I don't know whether these are true, but they are interesting and plausible stories given that he couldn't go to the local Meijer's store to buy a stop watch.



Actually, these were created by defining time as 2, 4, and 6 units, defining acceleration as 2, 4, and 6 units, applying the formula

$$\text{distance} \quad = \quad 0.5\text{at}^2 \tag{4-29}$$

and adding some random noise (from a normal distribution with mean 0 and variance 9). Let's see what happens when we apply an additive structural model to something that was created with a multiplicative model. Run MANOVA, examine boxplots, cell and marginal means, and parameter estimates.

```
COMMANDS

  manova distpe by accel(1,3) time(1,3)
   /print parameters(est)
```

```
   /omeans=tables(accel time accel by time)
   /plot boxplots
   /design constant  accel time accel by time.

[BOXPLOTS OMITTED]

MEANS FOR ACCEL MARGINAL
  Combined Observed Means for ACCEL
  Variable .. DISTPE
         ACCEL
             1        WGT.    17.55063
                    UNWGT.    17.55063
             2        WGT.    35.74165
                    UNWGT.    35.74165
             3        WGT.    55.65252
                    UNWGT.    55.65252

MEANS FOR TIME MARGINAL
  Combined Observed Means for TIME
  Variable .. DISTPE
          TIME
             1        WGT.     6.20152
                    UNWGT.     6.20152
             2        WGT.    31.09214
                    UNWGT.    31.09214
             3        WGT.    71.65113
                    UNWGT.    71.65113

CELL MEANS
  Combined Observed Means for ACCEL BY TIME
  Variable .. DISTPE
                      ACCEL           1           2           3
          TIME
             1        WGT.     2.61239     5.68376    10.30842
                    UNWGT.     2.61239     5.68376    10.30842
             2        WGT.    13.24837    31.35449    48.67356
                    UNWGT.    13.24837    31.35449    48.67356
             3        WGT.    36.79112    70.18670   107.97558
                    UNWGT.    36.79112    70.18670   107.97558

ANOVA SOURCE TABLE
  Tests of Significance for DISTPE using UNIQUE sums of squares
  Source of Variation          SS      DF        MS         F  Sig of F

  WITHIN CELLS             408.43      45      9.08
  CONSTANT               71213.81       1  71213.81   7846.14      .000
  ACCEL                  13074.66       2   6537.33    720.27      .000
  TIME                   39289.35       2  19644.68   2164.40      .000
  ACCEL BY TIME           6091.87       4   1522.97    167.80      .000

PARAMETER ESTIMATES
  --- Individual univariate .9500 confidence intervals
  CONSTANT

   Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

         1   36.3149328     .40997   88.57844     .00000   35.48920   37.14066
  ACCEL

   Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

         2   -18.764306     .57979  -32.36386     .00000  -19.93207  -17.59654
         3   -.57328354     .57979    -.98877     .32806   -1.74104     .59448
  TIME
```

Parameter	Coeff.	Std. Err.	t-Value	Sig. t	Lower -95%	CL- Upper
4	-30.113408	.57979	-51.93830	.00000	-31.28117	-28.94565
5	-5.2227910	.57979	-9.00804	.00000	-6.39055	-4.05503

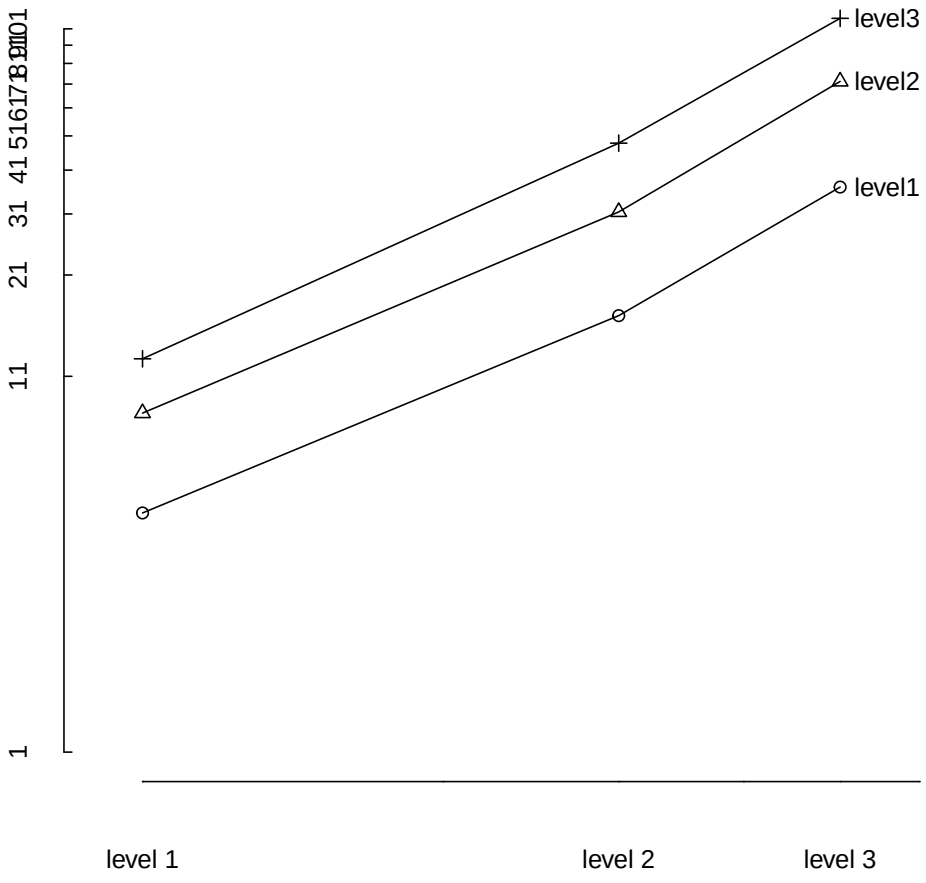
ACCEL BY TIME

Parameter	Coeff.	Std. Err.	t-Value	Sig. t	Lower -95%	CL- Upper
6	15.1751725	.81995	18.50744	.00000	13.52371	16.82664
7	.920536093	.81995	1.12267	.26753	-.73093	2.57200
8	.055517204	.81995	.06771	.94632	-1.59595	1.70698
9	.835632759	.81995	1.01913	.31359	-.81583	2.48710

The source table shows all three omnibus effects (two main effects and interaction) are statistically significant. We conclude that time and acceleration have an effect on distance independent of each of the two main effects. I doubt Galileo would have become famous reporting those kinds of results.

Now compare what happens when we transform the data to make the multiplicative model into an additive model. We will use logs to get a structural model that does not include an interaction.[6]

[6]Test question: Why might a boxplot look worse (additional skewness, different variances, outliers) after the transformation?



[BOXPLOTS OMITTED]

```
MEANS FOR ACCEL MARGINAL
  Combined Observed Means for ACCEL
  Variable .. LNDISTPE
          ACCEL
              1        WGT.     2.56839
                     UNWGT.     2.56839
              2        WGT.     3.23093
                     UNWGT.     3.23093
              3        WGT.     3.70756
                     UNWGT.     3.70756

MEANS FOR TIME MARGINAL
  Combined Observed Means for TIME
  Variable .. LNDISTPE
           TIME
```

```
              1        WGT.     1.91560
                     UNWGT.     1.91560
              2        WGT.     3.38019
                     UNWGT.     3.38019
              3        WGT.     4.21110
                     UNWGT.     4.21110

CELL MEANS
  Combined Observed Means for ACCEL BY TIME
  Variable .. LNDISTPE
                      ACCEL           1           2           3
          TIME
             1        WGT.     1.33736     1.91127     2.49816
                    UNWGT.     1.33736     1.91127     2.49816
             2        WGT.     2.71278     3.50314     3.92465
                    UNWGT.     2.71278     3.50314     3.92465
             3        WGT.     3.65503     4.27838     4.69988
                    UNWGT.     3.65503     4.27838     4.69988

ANOVA SOURCE TABLE
  Tests of Significance for LNDISTPE using UNIQUE sums of squares
  Source of Variation          SS      DF        MS         F  Sig of F

  WITHIN CELLS               4.93      45       .11
  CONSTANT                 542.28       1    542.28   4954.43      .000
  ACCEL                     11.78       2      5.89     53.83      .000
  TIME                      48.63       2     24.31    222.14      .000
  ACCEL BY TIME               .12       4       .03       .27      .897

PARAMETER ESTIMATES
  --- Individual univariate .9500 confidence intervals
  CONSTANT

   Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

         1   3.16896113     .04502   70.38773     .00000    3.07828    3.25964
  ACCEL

   Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

         2   -.60057346     .06367   -9.43259     .00000    -.72881    -.47234
         3   .061970386     .06367     .97331     .33560    -.06627     .19021
  TIME

   Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

         4   -1.2533624     .06367  -19.68528     .00000   -1.38160   -1.12512
         5   .211227082     .06367    3.31753     .00180     .08299     .33947
  ACCEL BY TIME

   Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

         6   .022334344     .09004     .24804     .80523    -.15902     .20369
         7   -.06683645     .09004    -.74227     .46178    -.24819     .11452
         8   -.06629596     .09004    -.73627     .46539    -.24765     .11506
         9   .060979685     .09004     .67723     .50173    -.12038     .24234
```

The interaction is no longer significant. This factorial design is telling us that, on the log scale, there are only two main effects. There is no interaction effect because the distance traveled in the log scale is simply the sum of appropriate row and column effects. Actually, the lack of interaction in the log scale suggests multiplication in the original scale so that gets

at the idea that acceleration and time multiply, which gets us part of the way there to the law relating distance, acceleration and time.

These examples highlight two main points: 1) the need to be careful about interpreting particular kinds of interactions and 2) the importance of thinking about the structural model you are using.

Interactions are sensitive to the particular scale one is using. There are no decent nonparametric procedures to handle the general case of factorial designs (see Busemeyer, 1980, *Psych Bull, 88*, 237-244 for a nice explanation of the problem of scale types and interactions). Developing nonparametric tests for interactions is an active area of research in statistics.

In psychology experiments we typically do not have strong justifications for using a particular scale or for fitting a particular model. In other words, we don't know what the true model is, nor do we have a clue as to what the real model would look like (unlike our counterparts in physics). What we need are methods to distinguish "real" interactions from spurious ones. Such trouble-shooting techniques exist. They involve residuals and are called "residual analysis." We will discuss them in the second half of the course in the context of regression. There's a sense in which residual analysis is the complement of what we have been focusing on in this class so far. The first half we talked about fitting structural models and interpreting the parameter estimates (what the model is telling us). Residual analysis, on the other hand, looks at what the model is not picking up, at what the model is missing. By looking at how the model is "not working" we can get a better understanding of what is happening in the statistical analysis and how to improve matters.

Before we can perform residual analysis in a sophisticated way, we need go over regression, develop the general linear model, and recast the ANOVA framework into this general linear model. The end result will be a general technique that can inform us both about what the model is doing (the "positive side") and inform us about what the model is not doing (the "negative side").

Let's wrap up ANOVA with Lecture Notes #5 and start regression....

Appendix 1: One-way ANOVA in SPSS MANOVA

```
Data in this form
1 56
1 49
1 65
1 60
2 84
2 78
2 94
2 93
3 80
3 72
3 83
3 85

Command syntax:

    manova seed by insect(1,3)
      /design = insect.
```

Appendix 2: Randomized Block Design in SPSS MANOVA

```
Data in this form

1 1 56
1 2 49
1 3 65
1 4 60
2 1 84
2 2 78
2 3 94
2 4 93
3 1 80
3 2 72
3 3 83
3 4 85

Command Syntax:

    manova seed by insect(1,3) plot(1,4)
      /design = insect, plot.
```

Note how the *design* subcommand parallels the terms in the structural model.

```
Suppose we forgot that this is a randomized block design (i.e., one
subject per cell) and treated the design as a two-way anova.
The command syntax for a two-way ANOVA is

     manova seed by insect(1,3) plot(1,4)
      /design = insect, plot, insect by plot.

The corresponding output is

 * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * *
  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
  *                    *                                           *
  *   W A R N I N G    *  Too few degrees of freedom in RESIDUAL   *
  *                    *  error term (DF = 0).                     *
  *                    *                                           *
  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 * * * * * * A N A L Y S I S   O F   V A R I A N C E -- DESIGN   1 * * * * * *

  Tests of Significance for SEED using UNIQUE sums of squares
  Source of Variation          SS      DF        MS         F  Sig of F

  RESIDUAL                    .00       0       .
  INSECT                   1925.17      2    962.58       .        .
  PLOT                      386.25      3    128.75       .        .
  INSECT BY PLOT             23.50      6      3.92       .        .
```

Note what happened. The partitioning of sums of squares is identical to the previous source table. However, because the program attempts to fit four terms when only three are available, the residual sum of squares is left with zero. The interaction term picks up what before we called residual. Had there been more than one subject per cell, the variability within cells (i.e., mean square residual) could have been computed.

The main point. We showed that the SSerror in the randomized block design can be interpreted as the variability of the cell observations. That is, in this misspecified design, the SSerror is equivalent to the "interaction term" between the two variables. Because there is only one observation per cell, the "interaction" is equivalent to the error term. When more than one observation per cell is available, then the interaction term is separate from the residual term.

Contrasts for randomized block designs can be conducted directly in the MANOVA command through this syntax. I'll explain this in more detail in the factorial design section of these lecture notes. I specify an orthogonal set for the insect factor and don't specify anything for the plot factor, but SPSS fills in a default set of contrasts for the blocking factor. The error term is correctly adjusted for the blocking factor, and even though the insect contrasts ignore the plot factor, the standard error for the contrast ihat uses the correct error term (SS and df) that has properly adjusted for the blocking factor.

```
manova seed by insect(1,3) plot(1,4)
 /contrast(insect) = special(1 1 1
                             1 -1 0
                             1 1 -2)
 /design = insect(1) insect(2) plot.
```

Tests of Significance for seed using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
RESIDUAL	23.50	6	3.92		
INSECT(1)	1770.13	1	1770.13	451.95	.000
INSECT(2)	155.04	1	155.04	39.59	.001
PLOT	386.25	3	128.75	32.87	.000
(Model)	2311.42	5	462.28	118.03	.000
(Total)	2334.92	11	212.27		

R-Squared = .990
Adjusted R-Squared = .982

Estimates for seed --- Individual univariate .9500 confidence intervals

INSECT(1)

Parameter	Coeff.	Std. Err.	t-Value	Sig. t	Lower -95%	CL- Upper
2	-29.7500000000	1.39940	-21.25904	.00000	-33.17422	-26.32578

INSECT(2)

Parameter	Coeff.	Std. Err.	t-Value	Sig. t	Lower -95%	CL- Upper
3	-15.2500000000	2.42384	-6.29167	.00075	-21.18092	-9.31908

PLOT

Parameter	Coeff.	Std. Err.	t-Value	Sig. t	Lower -95%	CL- Upper
4	-1.5833333333	.98953	-1.60009	.16070	-4.00462	.83796
5	-8.5833333333	.98953	-8.67416	.00013	-11.00462	-6.16204
6	5.7500000000	.98953	5.81085	.00114	3.32871	8.17129

Appendix 3: Latin Square Design in SPSS MANOVA

Example from Ott (p. 675). A petroleum company was interested in comparing the miles per gallon achieved by four different gasoline blends (I, II, III, IV). Because there can be considerable variability due to drivers and due to car models, these two extraneous sources of variability were included as blocking variables in the following Latin square design.

Each operator drove each car model over a standard course with the assigned gasoline blend by the Latin square design below. The miles per gallon are the dependent variable.

	Car Model			
Driver	1	2	3	4
1	IV	II	III	I
2	II	III	I	IV
3	III	I	IV	II
4	I	IV	II	III

Data from Table 14.14 (p. 676) is coded as follows:

```
Column 1:  Blocking variable--Driver
Column 2:  Blocking variable--Car model
Column 3:  Manipulation--Gasoline Blend
Column 4:  Dependent measure--miles per gallon

1 1 4 15.5
1 2 2 33.9
1 3 3 13.2
1 4 1 29.1
2 1 2 16.3
2 2 3 26.6
2 3 1 19.4
2 4 4 22.8
3 1 3 10.8
3 2 1 31.1
3 3 4 17.1
3 4 2 30.3
4 1 1 14.7
4 2 4 34.0
4 3 2 19.7
4 4 3 21.6
```

The command syntax is

```
   manova mpg by gas(1,4) driver(1,4) model(1,4)
    /design = gas, driver, model.
```

* * * * * * A N A L Y S I S O F V A R I A N C E * * * * * *

Tests of Significance for MPG using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
RESIDUAL	23.81	6	3.97		
GAS	108.98	3	36.33	9.15	.012
DRIVER	5.90	3	1.97	.50	.699
MODEL	736.91	3	245.64	61.90	.000

To understand these tests of significance we need to examine the means and parameter estimates.

grand mean		Parameter estimates
mu	= 22.26	obs mean - grand mean
treatment means		
I	= 23.57	1.31
II	= 25.05	2.79
III	= 18.05	-4.21
IV	= 22.35	0.09
Driver means		
1	= 22.92	0.66
2	= 21.27	-0.99
3	= 22.32	0.06
4	= 22.50	0.24
Car model means		
1	= 14.32	-7.94
2	= 31.40	9.14
3	= 17.35	-4.91
4	= 25.95	3.69

There appears to be a fairly large effect of car model, but the driver doesn't seem to matter much on mpg. The critical comparisons are the treatment means---miles per gallon of the four different gasoline types.

What if you had forgotten this is a latin square design
and performed a three-way ANOVA? Command syntax and source table below.

```
  manova mpg by gas(1,4) driver(1,4) model(1,4)
   /design = gas, driver, model, gas by driver, gas by model,
             driver by model, gas by driver by model.
```

```
* * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * *

 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 *                    *                                            *
 *   W A R N I N G    * UNIQUE sums-of-squares are obtained assuming  *
 *                    * the redundant effects (possibly caused by     *
 *                    * missing cells) are actually null.             *
 *                    * The hypotheses tested may not be the          *
 *                    * hypotheses of interest.  Different reorderings *
 *                    * of the model or data, or different contrasts  *
 *                    * may result in different UNIQUE sums-of-squares. *
 *                    *                                            *
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 *                    *                                            *
 *   W A R N I N G    *  Too few degrees of freedom in RESIDUAL      *
 *                    *  error term (DF = 0).                        *
 *                    *                                            *
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

  Tests of Significance for mpg using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F

 RESIDUAL                    .00       0       .
 GAS                       108.98      3     36.33       .        .
 DRIVER                      5.90      3      1.97       .        .
 MODEL                     247.73      3     82.58       .        .
 GAS BY DRIVER              23.81      6      3.97       .        .
 GAS BY MODEL                .00      0       .         .        .
 DRIVER BY MODEL             .00      0       .         .        .
 GAS BY DRIVER BY MOD        .00      0       .         .        .
 EL

 (Model)                   875.60     15     58.37       .        .
 (Total)                   875.60     15     58.37
```

This partition does not equal the correct latin square
partition we observed previously. This is because there are missing
cells---this ANOVA is assuming 64 cells but we only supplied 16. I
don't understand where the missing sum of squares went. The MODEL term
should be 736.91 and if you use that value the total sum of squares
is 875.60. The case of the missing sum of squares....

Finally, compare the results of the latin square design to case where only gasoline type is a factor (i.e., one-way ANOVA ignoring the two other variables). Command syntax and source table follow.

```
   manova mpg by gas(1,4)
    /design = gas.
```

* * * * * * A N A L Y S I S O F V A R I A N C E * * * * * *

Tests of Significance for MPG using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	766.62	12	63.88		
GAS	108.98	3	36.33	.57	.646

Compare this SS and F with the ones from the latin square. Driver and car model add considerable noise to the dependent variable. In this example, the noise is large enough to mask the effects of treatment.

Appendix 4: Analysis of Variance Commands in SPSS

SPSS provides several ways of calculating analysis of variance. You already know how to use the **ONEWAY** command. We will discuss two additional commands: **ANOVA** and **MANOVA**.

Unfortunately, there are minor differences between different versions of SPSS (both platform and version number) in the subcommands of ANOVA and MANOVA. I list commands that should work for most recent versions. If you are having trouble with your platform/version, look at the online syntax chart to find the way to perform these commands using your setup.

1. The **ANOVA** command

   We'll be interested in this command as a stepping stone to the more complicated, but more useful, **MANOVA** command.

   Nice features of the **ANOVA** command

   (a) Simple to use

   (b) Prints cell and marginal means

   (c) Prints the estimates of the terms in the structural model

   (d) Allows one to do randomized block and latin square designs by not including interaction terms in the structural model

   (e) Can perform analysis of covariance

   Ugly features of the **ANOVA** command

   (a) Doesn't do contrasts or *post hoc* tests

   (b) Be careful: different versions of SPSS have different defaults for handling unequal sample sizes

   (c) When using the correct approach to the unequal sample size problem (regression approach) the terms of the structural model cannot be printed (at least on my version of SPSS)

(d) Does not allow one to examine assumptions (**MANOVA** does)

(e) Does not allow diagnostics for the analysis of covariance

2. Using the **ANOVA** command

I think that once you see all the bells and whistles on the **MANOVA** command you probably won't use **ANOVA**. However, it is still worth spending a little time showing what the **ANOVA** command is all about.

The syntax is very similar to the **ONEWAY** command. The command line reads

ANOVA `dv` by `grouping1(a,b) grouping2(c,d) grouping3(e,f).`

where `dv` stands for name of the dependent variable, and the grouping# refers to the grouping variables (up to 10).

There are three subcommands of interest. One is the**/maxorders** subcommand. Including the subcommand

/maxorders = 1

will suppress all interaction terms. This is useful for fitting randomized block designs and latin square designs. The default for /maxorders is all; you can specify maxorder to be any integer relevant to your design. The **MANOVA** command also has this feature within its **/design** subcommand.

You should also specify how the ANOVA command should deal with unequal sample sizes. The options are unique, experimental, and hierarchical.

```
/method = unique
```

We will deal with the topic of unequal n's later. For now, take my word that the "unique approach" (aka "regression approach") is preferred; I'll explain in LN5. If all cells have the same number of subjects, then all three techniques (unique, hierarchical, experimental) will yield the identical result. However, when the sample sizes differ, then the three approaches will differ in the context of factorial ANOVAs.

3. The **MANOVA** command

   A very general command. Can do any factorial design. For now we will focus on the special case of univariate statistics (i.e., one dependent variable). The same command can be used to analyze repeated-measures and multiple dependent measures

   Nice features of the **MANOVA** command

   (a) The user specifies the structural model

   (b) Perform fixed- and random-effects models

   (c) Test contrasts

   (d) The default for handling unequal sample sizes (regression approach) is the best of the three approaches

   (e) Informative print outs on the parameter estimates

   (f) Generates boxplots and normal plots (a nice feature in complicated designs)

   (g) Some versions of SPSS also calculate the power of a nonsignificant test (one of my favorite features). This power calculation can be used to estimate the number of subjects needed to achieve a significant result given specified Type I and Type II error rates.

   Ugly features of the **MANOVA** command

   (a) Quite complicated; many subcommands; easy to get the incorrect analysis without being aware that it is incorrect

   (b) Cannot perform *post hoc* tests

   (c) The analysis of covariance capabilities are not very useful. We will perform analysis of covariance with the **REGRESSION** command later.

4. **MANOVA** syntax

The two necessary requirements are that the command line specify the dependent variable and list the independent variables. This is similar to the **ANOVA** command.

For example, a $2 \times 3 \times 4$ factorial design using GPA as a dependent variable would have the command line:

manova gpa by treat1(1,2) treat2(1,3) treat3(1,4)

5. Important **MANOVA** subcommands used in the analysis of a between-subjects analysis of variance

I list the ones that I will talk about today. Other subcommands will be discussed later.

DESIGN The **DESIGN** command is how the user specifies the structural model. If the user wants all main effects and interactions in the model (i.e., a complete factorial design), then the **DESIGN** subcommand can be omitted. That is, if the **DESIGN** subcommand is omitted, the program assumes the user wants all possible terms in the structural model. I'm compulsive about these things and always write out the complete design (even when the default would automatically do it for me).

While most of the time you will use the default setting on the **DESIGN** subcommand it is nice feature to have around (e.g., to calculate Latin Square designs, to test specific models that don't include all possible main effects and interactions).

PRINT This subcommand has several options. The most useful include:

(a) *CELLINFO*: prints means, standard deviations, and confidence intervals (the latter, according to the manual, if one has SET WIDTH WIDE).

(b) *PARAMETERS*: prints the estimates of the parameters of the structural model (assuming one has not specified *SPECIAL* contrasts—see below). See the manual for more details.

OMEANS Prints observed means. Nice for getting the marginal means. The *TABLE* keyword allows one to specify the factors.

CONTRAST A nice subcommand with many features. The most useful one is *SPECIAL*. It allows a user to define the set of orthogonal contrast to test. For example, consider the case of a 3×3 factorial design with treatA and treatB as the independent variables. The researcher is interested in testing the contrasts (1, -1, 0) and (1, 1, -2) on treatA. The command is:

```
manova dv by treatA(1,3) treatB(1,3)
  /contrast(treatA) = special( 1  1  1
                               1 -1  0
```

```
                                    1  1 -2)
  /design treatA(1), treatA(2), treatB, treatA(1) by treatB,
          treatA(2) by treatB.
```

The first line of that command contains the variable list. The **CONTRAST** command defines an orthogonal set of contrasts for treatA. Note that the unit vector must be included in the specification of *SPECIAL*. The **PARTITION** subcommand is not required but I include it for completeness. **PARTITION** lets SPSS know that you want to split the factor treatA into two pieces each with one degree of freedom. Finally, the **DESIGN** subcommand tells SPSS to calculate each contrast for treatA (as defined by the **CONTRAST** and **PARTITION** subcommands), the main effect for treatB, and separately the interaction of each contrast on treatA with the main effect on treatB.

There are other features of the **CONTRAST** subcommand, but they are not as useful as user-defined *SPECIAL*. If you don't specify the **CONTRAST** subcommand, SPSS will by default calculate the *DEVIATION* contrasts—which are equivalent to the parameters of the structural model (i.e., the α's, β's, $\alpha\beta$'s, etc.).

PLOT Can produce boxplots and normal plots. Unfortunately, on some versions of SPSS normal plots are on the marginals even though the boxplots are cell by cell. However, on some versions of SPSS both the boxplots and the normal plots are cell by cell. See Footnote 4 on page 4-22 for an alternative using the EXAMINE command.

You can use this line, for example, to get boxplots and normal probability plots.

```
/PLOT boxplot normal
```

ERROR Useful for random-effects models, nested designs, and dealing with unequal sample sizes. More details on this later, but now I'll give an example. Suppose you have a 2 factor, factorial between subjects ANOVA with the usual structural model Y = $\mu + \alpha + \beta + \alpha\beta + \epsilon$. In this full model, the MSE is the usual error term and corresponds to the keyword WITHIN in SPSS (e.g., /ERROR WITHIN). However, in some situations that we will discuss later, you may want to have different error terms. For instance, you may want to run the structural model without the interaction $\mu + \alpha + \beta + \epsilon$. In this case, there are three things you could do with the error term. (1) You could keep the same MSE (that is, ignore the portion of sum of squares that the interaction picks up). In this case, /ERROR WITHIN tells SPSS to ignore the interaction completely. (2) You may want to use the interaction term as the error term (and NOT use MSE). In this case /ERROR RESIDUAL tells SPSS to ignore MSE and only use what is left over in the structural model (i.e., the residual; the terms not explicitly stated in the /DESIGN line), which here is the interaction term. (3) You may want to combine the SSE and SSint into a single pooled error term. In this case, /ERROR WITHIN+RESIDUAL tells SPSS to combine into a single error term the MSE along with any term not explicitly stated in the design line. The default in SPSS is /ERROR WITHIN+RESIDUAL. So, if you omit a term in the design line it automatically gets incorporated into the error term.

POWER SPSS computes power within the MANOVA command by adding these two lines to your MANOVA syntax.

```
/power exact
/print signif(all) parameters(all)
```

By default, MANOVA prints source tables to two values after the decimal places. Sometimes you want more digits after the decimal place, such as when you need the MSE to compute something by hand. Currently, there isn't a straightforward way around the rounding problem. Here are two temporary kludges:

(a) multiply your raw data by a large number such as 100 or 1000, analyze the new data, and then divide the MSE in the source table by that number squared

(b) add this line in your MANOVA command

```
/print error(stddev)
```

which causes an extra line of output—the square root of the MSE—to be printed to five values after the decimal (at least on my version of SPSS does this). Simply square that value to get MSE.

6. **UNIANOVA** command

The UNIANOVA command is similar to to the MANOVA command with a few exceptions. It does not require one to specify the levels of the independent variables. So, one writes

```
UNIANOVA gpa by treat1, treat2, treat3
```

rather than

```
UNIANOVA gpa by treat1(1,2), treat2(1,2), treat3(1,2)
```

For the biofeedback example we could use the UNIANOVA command as in

```
UNIANOVA
  blpr  BY biofeed drug
```

```
  /METHOD = SSTYPE(3)
  /POSTHOC = biofeed drug ( TUKEY )
  /PRINT = OPOWER TEST(LMATRIX)
  /DESIGN = biofeed drug biofeed*drug .
```

We will talk about SSTYPE in Lecture Notes #5. Note the use of posthoc to define TUKEY tests on the marginals. In this simple example, because there are two rows there is no TUKEY test to compute on rows (and similarly, since there are two columns TUKEY cannot be computed on the column factor either). The PRINT subcommand also lists two things I want to mention: OPOWER prints out the observed power and TEST(LMATRIX) presents a different way of printing out the contrasts that SPSS actually computes. Recall we also found it helpful to print out the contrasts within the MANOVA command.

The UNIANOVA command does have the ability to perform Tukey tests on the marginal means, but not the individual cell means. It also computes Scheffe tests, but as pairwise means rather than for arbitrary contrasts on marginal means (not on cell means). Contrasts can be called the same way using the same "contrast()=special" syntax.

There is also an LMATRIX subcommand that let's you specify specific contrasts by name without needing an entire set of orthogonal contrasts as required by the "contrast()=special".

```
UNIANOVA
  blpr  BY biofeed drug
  /LMATRIX = "interaction contrast" biofeed*drug 1 -1 -1 1
  /DESIGN = biofeed drug biofeed*drug .
```

If there is more than one contrast for an effect that you want to test you separate the lines by a semicolon just repeating the part that goes after the label definition (i.e,. after the quotes). I'll explain the meaning of LMATRIX in Lecture Notes

I'm not as familiar with the UNIANOVA command, so best to look at the online help (e.g., the syntax guide within SPSS for UNIANOVA). From what I can tell it is a special case of the GLM command discussed next, so might as well focus on GLM if you want the latest and greatest SPSS has to offer.

7. **GLM** command

The GLM command is a relatively new command in SPSS, following similar syntax to the UNIANOVA command.

For example, the randomized block design with contrasts can be specified like this:

```
GLM seed BY insect plot
 /LMATRIX "1 v 2" insect 1 -1 0
 /LMATRIX  "1&2 v 3" insect 1 1 -2
 /DESIGN insect plot.
```

Contrasts can be listed as separate LMATRIX subcommands. Each LMATRIX contrast is assigned a label so that during the printing you get more readable output.

The GLM command also has the POSTHOC marginal means capability as mentioned above in the UNIANOVA command. There needs to be more than two levels in factor for post hoc tests. If there are only two levels, then there is only one pairwise test on the marginal means for that factor so multiple comparisons are not an issue.

```
GLM dv BY factor1 factor2
 /posthoc factor1 factor2 (tukey).
```

I don't know how to make GLM do posthoc tests like Tukey on individual cells in a factorial design (the syntax right above performs Tukey on the marginal means). You can get SPSS to do a Bonferroni correction for all possible cell means using the subcommand below, but it doesn't support the Tukey test.

```
/emmeans = tables(factor1*factor2)
                  compare(factor1) adj(bonferroni)
```

Appendix 5: Contrasts in SPSS MANOVA

```
Command syntax
  manova blood by biofeed(0,1) drug(0,1)
   /contrast(biofeed) = special(1 1
                                1 -1)
   /contrast(drug) = special(1 1
                             1 -1)
   /design biofeed(1) drug(1) biofeed(1) by drug(1).

Source table

 * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * *

  Source of Variation          SS      DF        MS         F  Sig of F

  WITHIN CELLS            1000.00      16     62.50
  BIOFEED(1)               500.00       1    500.00      8.00      .012
  DRUG(1)                  720.00       1    720.00     11.52      .004
  BIOFEED(1) BY DRUG(1)    320.00       1    320.00      5.12      .038
```

The tests printed in this source table are identical to the tests of the three contrasts. Recall that all Fs with one degree of freedom in the numerator are identical to t^2. Contrasts always have one degree of freedom in the numerator.

Appendix 6: Tukey HSD on 2 × 2

```
Output from ONEWAY on TUKEY.  These pairwise comparisons
correspond to cell means, not marginal means.  For cell means, the
TUKEY test (as well as all other post hoc tests like Scheffe, SNK, etc.)
can be done with ONEWAY. However, if you are doing post hoc tests on the
marginal means, then you must use the UNIANOVA or the GLM command,
or you can perform the calculations by hand.

 TUKEY-HSD PROCEDURE
 RANGES FOR THE 0.050 LEVEL -

           4.04   4.04   4.04

 THE RANGES ABOVE ARE TABLE RANGES.
 THE VALUE ACTUALLY COMPARED WITH MEAN(J)-MEAN(I) IS..
         5.5902 * RANGE * DSQRT(1/N(I) + 1/N(J))

    (*) DENOTES PAIRS OF GROUPS SIGNIFICANTLY DIFFERENT AT THE 0.050 LEVEL

                           G G G G
                           r r r r
                           p p p p

      Mean       Group     1 3 2 4

    168.0000    Grp 1
    186.0000    Grp 3      *
    188.0000    Grp 2      *
    190.0000    Grp 4      *

   HOMOGENEOUS SUBSETS   (SUBSETS OF GROUPS, WHOSE HIGHEST AND LOWEST MEANS
                          DO NOT DIFFER BY MORE THAN THE SHORTEST
                          SIGNIFICANT RANGE FOR A SUBSET OF THAT SIZE)

 SUBSET  1

 GROUP       Grp 1
 MEAN      168.0000

 SUBSET  2

 GROUP       Grp 3         Grp 2         Grp 4
 MEAN      186.0000      188.0000      190.0000
```

Appendix 7: Another example of a factorial design (SPSS)

Here is an experiment testing the effects of wall color and room size on anxiety level during an interview (Hays, 1988). Assume that the anxiety measure is reliable. Lower numbers reflect lower anxiety scores. The four room colors were red, yellow, green, and blue (coded as 1, 2, 3, and 4, respectively). The three room sizes were small, medium, and large (coded 1, 2, and 3).

It will be useful to try the various SPSS features to get a feel for what happens. I will now perform for your viewing pleasure several different runs of MANOVA on the same data to illustrate many features. Of course, in most data analysis situations only one run of MANOVA is required. All assumptions have been checked and appear okay; for simplicity here I focus on the statistical analysis.

Data

```
1 1 160
1 1 155
1 1 170
1 2 134
1 2 139
1 2 144
1 3 104
1 3 175
1 3 96
1 4 86
1 4 71
1 4 112
2 1 175
2 1 152
2 1 167
2 2 159
2 2 156
2 2 159
2 3 83
2 3 89
2 3 79
2 4 110
2 4 87
2 4 100
3 1 180
3 1 154
3 1 141
3 2 170
3 2 133
3 2 128
3 3 84
3 3 86
3 3 83
3 4 105
3 4 93
3 4 85
```

```
FIRST LET'S SEE WHAT HAPPENS IF ALL WE DO IS SPECIFY THE GROUPING
FACTORS IN THE FIRST LINE. WE ONLY GET THE SOURCE TABLE.

 manova anxiety by size(1,3) color(1,4) .

 Tests of Significance for ANXIETY using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F

 WITHIN CELLS            7453.33      24    310.56
 SIZE                     477.56       2    238.78       .77      .475
 COLOR                  31526.44       3  10508.81     33.84      .000
 SIZE BY COLOR           3664.22       6    610.70      1.97      .111


 NOW ADD A PRINT SUBCOMMAND TO GIVE INFO ON THE CELL MEANS AND PARAMETER
 ESTIMATES OF THE STRUCTURAL MODEL.  THE OMEANS COMMANDS GIVES MARGINAL
 MEANS AS WELL. NOTE THAT THE DESIGN COMMANDS DOESN'T SAY ANYTHING SO IT
 DEFAULTS TO THE PUTTING ALL POSSIBLE TERMS IN THE MODEL.

 manova anxiety by size(1,3) color(1,4)
   /print  cellinfo(means) parameters(estim negsum)
   /omean tables(size, color, size by color)
   /design .
```

```
Cell Means and Standard Deviations
Variable .. ANXIETY
     FACTOR           CODE        Mean  Std. Dev.     N   95 percent Conf. Interval

 SIZE            small
  COLOR           red          161.667      7.638     3    142.694    180.640
  COLOR           yellow       139.000      5.000     3    126.579    151.421
  COLOR           green        125.000     43.486     3     16.975    233.025
  COLOR           blue          89.667     20.744     3     38.134    141.199
 SIZE            medium
  COLOR           red          164.667     11.676     3    135.661    193.672
  COLOR           yellow       158.000      1.732     3    153.697    162.303
  COLOR           green         83.667      5.033     3     71.163     96.170
  COLOR           blue          99.000     11.533     3     70.351    127.649
 SIZE            large
  COLOR           red          158.333     19.858     3    109.003    207.663
  COLOR           yellow       143.667     22.942     3     86.675    200.658
  COLOR           green         84.333      1.528     3     80.539     88.128
  COLOR           blue          94.333     10.066     3     69.327    119.340
For the entire sample          125.111     35.100    36    113.235    136.987


Combined Observed Means for SIZE
Variable .. ANXIETY
         SIZE
        small        WGT.   128.83333
                   UNWGT.   128.83333
       medium        WGT.   126.33333
                   UNWGT.   126.33333
        large        WGT.   120.16667
                   UNWGT.   120.16667

Combined Observed Means for COLOR
Variable .. ANXIETY
        COLOR
          red        WGT.   161.55556
                   UNWGT.   161.55556
       yellow        WGT.   146.88889
                   UNWGT.   146.88889
        green        WGT.    97.66667
                   UNWGT.    97.66667
         blue        WGT.    94.33333
                   UNWGT.    94.33333

Combined Observed Means for SIZE BY COLOR
Variable .. ANXIETY
                     SIZE       small      medium       large
        COLOR
          red        WGT.   161.66667   164.66667   158.33333
                   UNWGT.   161.66667   164.66667   158.33333
       yellow        WGT.   139.00000   158.00000   143.66667
                   UNWGT.   139.00000   158.00000   143.66667
        green        WGT.   125.00000    83.66667    84.33333
                   UNWGT.   125.00000    83.66667    84.33333
         blue        WGT.    89.66667    99.00000    94.33333
                   UNWGT.    89.66667    99.00000    94.33333

Tests of Significance for ANXIETY using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN CELLS           7453.33      24    310.56
SIZE                    477.56       2    238.78       .77      .475
```

```
COLOR                  31526.44       3  10508.81     33.84      .000
SIZE BY COLOR           3664.22       6    610.70      1.97      .111
Estimates for ANXIETY
--- Individual univariate .9500 confidence intervals
SIZE

 Parameter           Coeff.     Std. Err.      t-Value   Sig. t   Lower -95%   CL- Upper

       2       3.7222222222       4.15368       .89613   .37909     -4.85056   12.29500
       3       1.2222222222       4.15368       .29425   .77110     -7.35056    9.79500
       4      -4.9444444444         .             .        .           .          .
COLOR

 Parameter           Coeff.     Std. Err.      t-Value   Sig. t   Lower -95%   CL- Upper

       4      36.4444444444       5.08720      7.16395   .00000     25.94497    46.94391
       5      21.7777777778       5.08720      4.28089   .00026     11.27831    32.27725
       6     -27.4444444444       5.08720     -5.39480   .00002    -37.94391   -16.94497
       7     -30.7777777778         .              .          .       .           .
SIZE BY COLOR

 Parameter           Coeff.     Std. Err.      t-Value   Sig. t   Lower -95%   CL- Upper

       7      -3.6111111111       7.19439      -.50193   .62029    -18.45960    11.23738
       8     -11.6111111111       7.19439     -1.61391   .11962    -26.45960     3.23738
       9      23.6111111111       7.19439      3.28188   .00315      8.76262    38.45960
      10       1.8888888889       7.19439       .26255   .79514    -12.95960    16.73738
      11       9.8888888889       7.19439      1.37453   .18198     -4.95960    24.73738
      12     -15.2222222222       7.19439     -2.11585   .04492    -30.07072     -.37373
      13      -4.9444444444         .             .        .           .          .

 Note:  Overlap in parameter numbering is due to the NEGSUM option.

TURNS OUT THE PARAMETER VALUES ARE IDENTICAL TO THE ESTIMATES OF THE
STRUCTURAL MODEL (THE ALPHAS, BETAS, ALPHABETAS), BUT WE DON'T KNOW THE
CONTRASTS CORRESPONDING TO THOSE "I HATS".  WE COULD USE DESIGN(SOLUTION) TO
FORCE SPSS TO PRINT OUT THE CONTRASTS IT USED.

 NOW ASK FOR A PRINT OUT OF THE DESIGN MATRIX AS WELL AS ORTHOGONAL
 PARAMETER ESTIMATES.  I ALSO SPECIFY SPECIFIC CONTRASTS I WANT TO TEST.
 NOTE HOW I SPECIFY SEPARATELY FOR EACH FACTOR.

manova anxiety by size(1,3) color(1,4)
  /print parameters(estim) design(solution)
  /contrast(size) = special(1 1 1
                1 -1 0
                1 1 -2)
  /contrast(color) = special(1 1 1 1
                1 1 -1 -1
                1 -1 -1 1
                1 -1 1 -1)
  /design size color size by color.

Solution Matrix for Between-Subjects Design
 1-SIZE  2-COLOR
FACTOR                                    PARAMETER

1   2          1         2         3         4         5         6         7         8         9      10

1   1     -.50000    .61237    .35355    .50000   -.50000   -.50000    .61237    .61237    .61237 .35355
1   2     -.50000    .61237    .35355    .50000    .50000    .50000    .61237   -.61237   -.6123  .35355
1   3     -.50000    .61237    .35355   -.50000    .50000   -.50000   -.61237   -.61237    .61237 -.35355
1   4     -.50000    .61237    .35355   -.50000   -.50000    .50000   -.61237    .61237   -.61237 -.35355
2   1     -.50000   -.61237    .35355    .50000   -.50000   -.50000   -.61237   -.61237   -.61237  .35355
```

```
2   2     -.50000   -.61237    .35355    .50000    .50000    .50000   -.61237    .61237    .61237  .35355
2   3     -.50000   -.61237    .35355   -.50000    .50000   -.50000    .61237    .61237   -.61237 -.35355
2   4     -.50000   -.61237    .35355   -.50000   -.50000    .50000    .61237   -.61237    .61237 -.35355
3   1     -.50000    .00000   -.70711    .50000   -.50000   -.50000    .00000    .00000    .00000 -.70711
3   2     -.50000    .00000   -.70711    .50000    .50000    .50000    .00000    .00000    .00000 -.70711
3   3     -.50000    .00000   -.70711   -.50000    .50000   -.50000    .00000    .00000    .00000  .70711
3   4     -.50000    .00000   -.70711   -.50000   -.50000    .50000    .00000    .00000    .00000  .70711

1   2          11        12

1   1      .35355    .35355
1   2     -.35355   -.35355
1   3     -.35355    .35355
1   4      .35355   -.35355
2   1      .35355    .35355
2   2     -.35355   -.35355
2   3     -.35355    .35355
2   4      .35355   -.35355
3   1     -.70711   -.70711
3   2      .70711    .70711
3   3      .70711   -.70711
3   4     -.70711    .70711

 Tests of Significance for ANXIETY using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F

 WITHIN CELLS           7453.33      24    310.56
 SIZE                    477.56       2    238.78       .77      .475
 COLOR                 31526.44       3  10508.81     33.84      .000
 SIZE BY COLOR          3664.22       6    610.70      1.97      .111
 Estimates for ANXIETY
 --- Individual univariate .9500 confidence intervals
 SIZE

  Parameter           Coeff.    Std. Err.     t-Value     Sig. t     Lower -95%    CL- Upper

        2       2.5000000000      7.19439      .34749     .73125      -12.34849     17.34849
        3      14.8333333333     12.46105     1.19038     .24554      -10.88501     40.55168
 COLOR

  Parameter           Coeff.    Std. Err.     t-Value     Sig. t     Lower -95%    CL- Upper

        4     116.4444444444     11.74839     9.91152     .00000       92.19696    140.69193
        5      11.3333333333     11.74839      .96467     .34433      -12.91415     35.58082
        6      18.0000000000     11.74839     1.53212     .13857       -6.24749     42.24749
 SIZE BY COLOR

  Parameter           Coeff.    Std. Err.     t-Value     Sig. t     Lower -95%    CL- Upper

        7     -54.0000000000     28.77756    -1.87646     .07280     -113.39397      5.39397
        8     -34.6666666667     28.77756    -1.20464     .24009      -94.06064     24.72730
        9      66.6666666667     28.77756     2.31662     .02938        7.27270    126.06064
       10     -20.6666666667     49.84420     -.41463     .68210     -123.54004     82.20671
       11     -40.0000000000     49.84420     -.80250     .43014     -142.87338     62.87338
       12      40.0000000000     49.84420      .80250     .43014      -62.87338    142.87338

THE ABOVE PARAMETER ESTIMATES CORRESPOND TO CONTRAST VALUES AND THE
T-TESTS ARE BASED ON THE POOLED ESTIMATES.

NOW LET'S SPECIFY EACH CONTRAST SPECIFICALLY IN THE DESIGN
STATEMENT. FOR COMPLETENESS I ALSO ASK FOR DESIGN(SOLUTION) SO YOU CAN
COMPARE THE OUTPUT OF THE SOURCE TABLE WITH THE OUTPUT OF THE INDIVIDUAL
CONTRASTS. NOTE THE P-VALUES ARE IDENTICAL. THE SOURCE TABLE PRESENTS
EVERYTHING IN A COMPACT MANNER, WHERE AS TO UNDERSTAND THE PARAMETERS
```

ONE NEEDS TO LOOK AT THE ACTUAL CONTRASTS USED BY SPSS IN THE SOLUTION
MATRIX TO BE SURE WHICH CONTRASTS CORRESPOND TO WHICH PARAMETER (AND TO
KNOW WHAT WEIGHTS WERE USED TO MAKE SENSE OF I HAT.

```
manova anxiety by size(1,3) color(1,4)
  /print design(solution)
  /contrast(size) = special( 1 1 1
                             1 -1 0
                             1 1 -2)
  /contrast(color)=special(1 1 1 1
                           1 1 -1 -1
                           1 -1 -1 1
                           1 -1 1 -1)
  /design size(1) size(2) color(1) color(2) color(3)
       size(1) by color(1), size(1) by color(2),
       size(1) by color(3), size(2) by color(1),
       size(2) by color(2), size(2) by color(3).
```

Solution Matrix for Between-Subjects Design
1-SIZE 2-COLOR

FACTOR		PARAMETER									
1	2	1	2	3	4	5	6	7	8	9	10
1	1	-.50000	.61237	-.35355	.50000	-.50000	.50000	-.61237	.61237	-.61237	.35355
1	2	-.50000	.61237	-.35355	.50000	.50000	-.50000	-.61237	-.61237	.61237	.35355
1	3	-.50000	.61237	-.35355	-.50000	.50000	.50000	.61237	-.61237	-.61237	-.35355
1	4	-.50000	.61237	-.35355	-.50000	-.50000	-.50000	.61237	.61237	.61237	-.35355
2	1	-.50000	-.61237	-.35355	.50000	-.50000	.50000	.61237	-.61237	.61237	.35355
2	2	-.50000	-.61237	-.35355	.50000	.50000	-.50000	.61237	.61237	-.61237	.35355
2	3	-.50000	-.61237	-.35355	-.50000	.50000	.50000	-.61237	.61237	.61237	-.35355
2	4	-.50000	-.61237	-.35355	-.50000	-.50000	-.50000	-.61237	-.61237	-.61237	-.35355
3	1	-.50000	.00000	.70711	.50000	-.50000	.50000	.00000	.00000	.00000	-.70711
3	2	-.50000	.00000	.70711	.50000	.50000	-.50000	.00000	.00000	.00000	-.70711
3	3	-.50000	.00000	.70711	-.50000	.50000	.50000	.00000	.00000	.00000	.70711
3	4	-.50000	.00000	.70711	-.50000	-.50000	-.50000	.00000	.00000	.00000	.70711

1	2	11	12
1	1	-.35355	.35355
1	2	.35355	-.35355
1	3	.35355	.35355
1	4	-.35355	-.35355
2	1	-.35355	.35355
2	2	.35355	-.35355
2	3	.35355	.35355
2	4	-.35355	-.35355
3	1	.70711	-.70711
3	2	-.70711	.70711
3	3	-.70711	-.70711
3	4	.70711	.70711

Tests of Significance for ANXIETY using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	7453.33	24	310.56		
SIZE(1)	37.50	1	37.50	.12	.731
SIZE(2)	440.06	1	440.06	1.42	.246
COLOR(1)	30508.44	1	30508.44	98.24	.000
COLOR(2)	289.00	1	289.00	.93	.344
COLOR(3)	729.00	1	729.00	2.35	.139
SIZE(1) BY COLOR(1)	1093.50	1	1093.50	3.52	.073

```
SIZE(1) BY COLOR(2)      450.67       1    450.67      1.45      .240
SIZE(1) BY COLOR(3)     1666.67       1   1666.67      5.37      .029
SIZE(2) BY COLOR(1)       53.39       1     53.39       .17      .682
SIZE(2) BY COLOR(2)      200.00       1    200.00       .64      .430
SIZE(2) BY COLOR(3)      200.00       1    200.00       .64      .430

(Model)                35668.22      11   3242.57     10.44      .000
(Total)                43121.56      35   1232.04

R-Squared =           .827
Adjusted R-Squared =  .748

SIZE(1)
 Parameter           Coeff.        Std. Err.     t-Value   Sig. t       Lower -95% CL- Upper
       2       2.5000000000          7.19439      .34749   .73125        -12.34849  17.34849

SIZE(2)
 Parameter           Coeff.        Std. Err.     t-Value   Sig. t       Lower -95%  CL- Upper
       3      14.8333333333         12.46105     1.19038   .24554        -10.88501   40.55168

COLOR(1)
 Parameter           Coeff.        Std. Err.    t-Value   Sig. t       Lower -95%   CL- Upper
       4     116.4444444444         11.74839    9.91152   .00000         92.19696   140.69193

COLOR(2)
 Parameter           Coeff.        Std. Err.    t-Value   Sig. t       Lower -95%   CL- Upper
       5      11.3333333333         11.74839     .96467   .34433        -12.91415    35.58082


COLOR(3)
 Parameter           Coeff.        Std. Err.    t-Value   Sig. t       Lower -95%   CL- Upper
       6      18.0000000000         11.74839    1.53212   .13857         -6.24749    42.24749

SIZE(1) BY COLOR(1)
 Parameter           Coeff.        Std. Err.    t-Value   Sig. t       Lower -95%   CL- Upper
       7     -54.0000000000         28.77756   -1.87646   .07280       -113.39397     5.39397

SIZE(1) BY COLOR(2)
 Parameter           Coeff.        Std. Err.   t-Value   Sig. t       Lower -95%   CL- Upper
       8     -34.6666666667         28.77756  -1.20464   .24009        -94.06064    24.72730

SIZE(1) BY COLOR(3)
 Parameter           Coeff.        Std. Err.    t-Value   Sig. t       Lower -95%   CL- Upper
       9      66.6666666667         28.77756    2.31662   .02938          7.27270   126.06064

SIZE(2) BY COLOR(1)
 Parameter           Coeff.        Std. Err.    t-Value   Sig. t       Lower -95%   CL- Upper
      10     -20.6666666667         49.84420    -.41463   .68210       -123.54004    82.20671

SIZE(2) BY COLOR(2)
 Parameter           Coeff.        Std. Err.    t-Value   Sig. t       Lower -95%   CL- Upper
      11     -40.0000000000         49.84420    -.80250   .43014       -142.87338    62.87338

SIZE(2) BY COLOR(3)
 Parameter           Coeff.        Std. Err.    t-Value   Sig. t       Lower -95%   CL- Upper
      12      40.0000000000         49.84420     .80250   .43014        -62.87338   142.87338

FINALLY, THE BRAIN TEASER!  LOOK AT THESE CONTRASTS.  ARE THEY
LEGITIMATE?  TAKE A LOOK AT THE ESTIMATES OF THE CONTRAST VALUES.  SEE
ANYTHING THAT MAKES SENSE?
```

manova anxiety by size(1,3) color(1,4)

```
/print parameters(estim)
/contrast(size) = special(1 0 0
                          0 1 0
                          0 0 1)
/contrast(color) = special(1 0 0 0
                           0 1 0 0
                           0 0 1 0
                           0 0 0 1)
/design size color size by color.


*** WARNING *** Possible error in special contrasts for SIZE

                Row 2 sums to       1.0000000
                Row 3 sums to       1.0000000


*** WARNING *** Possible error in special contrasts for COLOR

                Row 2 sums to       1.0000000
                Row 3 sums to       1.0000000
                Row 4 sums to       1.0000000


Tests of Significance for ANXIETY using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN CELLS            7453.33      24    310.56
SIZE                  156553.67       2  78276.83    252.05      .000
COLOR                 128958.33       3  42986.11    138.42      .000
SIZE BY COLOR         235248.33       6  39208.06    126.25      .000
Estimates for ANXIETY
--- Individual univariate .9500 confidence intervals
SIZE

 Parameter           Coeff.        Std. Err.     t-Value     Sig. t       Lower -95%  CL- Upper
       2     164.6666666667         10.17441    16.18440     .00000        143.66773  185.66561
       3     158.3333333333         10.17441    15.56193     .00000        137.33439  179.33227
COLOR

 Parameter           Coeff.        Std. Err.     t-Value     Sig. t       Lower -95%  CL- Upper
       4     139.0000000000         10.17441    13.66173     .00000        118.00106  159.99894
       5     125.0000000000         10.17441    12.28573     .00000        104.00106  145.99894
       6      89.6666666667         10.17441     8.81296     .00000         68.66773  110.66561
SIZE BY COLOR

 Parameter           Coeff.        Std. Err.     t-Value     Sig. t       Lower -95%   CL- Upper
       7     158.0000000000         10.17441    15.52916     .00000        137.00106   178.99894
       8      83.6666666667         10.17441     8.22325     .00000         62.66773   104.66561
       9      99.0000000000         10.17441     9.73030     .00000         78.00106   119.99894
      10     143.6666666667         10.17441    14.12040     .00000        122.66773   164.66561
      11      84.3333333333         10.17441     8.28877     .00000         63.33439   105.33227
      12      94.3333333333         10.17441     9.27163     .00000         73.33439   115.33227
```

Appendix 8: R code for factorial ANOVA

The R syntax for factorial ANOVAs is a simple extension of the one-way ANOVA version described in LN2 and LN3. The aov() command takes the structural model as the primary argument, so the only difference is how you specify the structural model. Everything else, such as defining each predictor variable as a factor, attaching contrasts to factors, etc., is the same.

The code below assumes equal number of subjects per cell and an orthogonal set of contrasts. More general code will be given in Lecture Notes 5.

Random Block Design

For the seedlings randomized block design presented in the first few pages of these lecture notes we have the following (note the use of the plus sign between the two factors):

```
seed.data <- read.table("data.rbd", header = F)
colnames(seed.data) <- c("insect", "plot", "dv")
seed.data$insect <- factor(seed.data$insect)
seed.data$plot <- factor(seed.data$plot)
out <- aov(dv ~ insect + plot, data = seed.data)
summary(out)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## insect       2 1925.2   962.6  245.77 1.75e-06
## plot         3  386.2   128.7   32.87 0.000404
## Residuals    6   23.5     3.9
##
## insect      ***
## plot        ***
## Residuals
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Some people like to call out the blocking factor's sum of squares outside of the source table and have one table for blocking factors and one table for the "key" factor. One way to accomplish this is the following command. The SS, df, MS and F terms are all identical to before, but the output is partitioned into a piece that is for the blocking factor and another piece this the the factor of interest. This uses the Error() command in the formula to the aov command, and there are two "errors"—one that is due to plot and another that is within groups.

```
summary(aov(dv ~ insect + Error(plot), data = seed.data))


##
## Error: plot
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  3  386.2   128.8
##
## Error: Within
##           Df Sum Sq Mean Sq F value   Pr(>F)
## insect     2 1925.2   962.6   245.8 1.75e-06 ***
## Residuals  6   23.5     3.9
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If you try to run a model that can't be estimated, R doesn't produce error messages but doesn't print an F or a p-value.

```
out <- aov(dv ~ insect * plot, data = seed.data)
summary(out)


##             Df Sum Sq Mean Sq
## insect       2 1925.2   962.6
## plot         3  386.2   128.7
## insect:plot  6   23.5     3.9
```

If you run this impossible-to-estimate model in a Bayesian framework, you will see what looks like complete output making it appear that all the terms, including interactions, can be estimated. However, the posterior results for the nonexistent parameters will be essentially identical to the prior distribution because the data cannot move away from the prior distribution. If you aren't careful, you will interpret the posterior when it is merely reflecting the prior. This is because there are no data to "update the prior." You can see this by running different priors and re-estimating the model. The resulting posteriors for the non-existent parameters will simply follow the prior distribution that was specified; then you will know something has gone awry.

Contrasts in R for randomized blocked designs can be accomplished in this manner. More explanation in the factorial design section later in these lecture notes. It is necessary to define manually the contrast weights to each factor to avoid the issues with fit.contrast described in LN3 surrounding nonorthogonal contrasts. R uses default dummy code contrasts with can create nonorthogonality issues with interactions; even though we aren't using interactions in randomized block designs, it is a good idea to get in the habit of manually defining orthogonal sets of contrasts when using R. The only difference in syntax between a randomized block design and a factorial design is that one

uses the plus sign in the aov() call rather than the asterisk, as the latter includes the interaction. Everything else remains the same.

```
library(gmodels)
seed.data$insect <- factor(seed.data$insect)
seed.data$plot <- factor(seed.data$plot)
# can define key contrasts of interest for primary factor
contrasts(seed.data$insect) <- cbind(c(1, -1, 0), c(1, 1, -2))
# arbitrary set of orthogonal contrasts for the blocking factor
contrasts(seed.data$plot) <- cbind(c(1, -1, 0, 0), c(1, 1, -2, 0), c(1, 1, 1, -3

out.aov <- aov(dv ~ insect + plot, data = seed.data)
fit.contrast(out.aov, "insect", c(1, -1, 0), df = T, conf.int = 0.95)


##                      Estimate Std. Error   t value
## insect c=( 1 -1 0 )   -29.75   1.399405 -21.25904
##                          Pr(>|t|) DF  lower CI
## insect c=( 1 -1 0 ) 7.063242e-07  6 -33.17422
##                       upper CI
## insect c=( 1 -1 0 ) -26.32578
```

We can verify these computations by hand. The ihat is 57.5 - 87.25 = -29.75. Each of the insecticide means is based on 4 observations (one observation per plot) so the n in the se(ihat) calculation is 4. The MSE is the 3.92 from the ANOVA source table that has removed the blocking factor from both sum of squares residual and the degrees of freedom residual. Plug those numbers into the se(ihat) formula from LN3, and you'll get 1.4. The t observed is $-21.25 = \frac{-29.75}{1.4}$. If you want to minimize round off error, you can print the MSE term with more digits using the following print command, and then you'll get exactly the same se(ihat) as in the fit.contrast() command (and also the SPSS example) above.

```
print(summary(out.aov), digits = 10)


##             Df      Sum Sq     Mean Sq   F value
## insect       2 1925.166667 962.5833333 245.76596
## plot         3  386.250000 128.7500000  32.87234
## Residuals    6   23.500000   3.9166667
##                  Pr(>F)
## insect      1.7538e-06 ***
## plot        0.00040368 ***
## Residuals
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Latin Square Design

Here is the latin square design example. I read in the data, define factors (otherwise aov won't treat the design correctly, recall LN2), and call aov with three main effects.

```
data.lsd <- read.table("data.lsd", header = F)
names(data.lsd) <- c("driver", "car", "gas", "dv")
data.lsd$driver <- factor(data.lsd$driver)
data.lsd$car <- factor(data.lsd$car)
data.lsd$gas <- factor(data.lsd$gas)

data.lsd
```

```
##    driver car gas   dv
## 1       1   1   4 15.5
## 2       1   2   2 33.9
## 3       1   3   3 13.2
## 4       1   4   1 29.1
## 5       2   1   2 16.3
## 6       2   2   3 26.6
## 7       2   3   1 19.4
## 8       2   4   4 22.8
## 9       3   1   3 10.8
## 10      3   2   1 31.1
## 11      3   3   4 17.1
## 12      3   4   2 30.3
## 13      4   1   1 14.7
## 14      4   2   4 34.0
## 15      4   3   2 19.7
## 16      4   4   3 21.6
```

```
summary(aov(dv ~ driver + car + gas, data = data.lsd))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## driver       3    5.9    1.97   0.495   0.6987
## car          3  736.9  245.64  61.903 6.63e-05
## gas          3  109.0   36.33   9.155   0.0117
## Residuals    6   23.8    3.97
##
## driver
## car         ***
## gas         *
## Residuals
```

```
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With this design I cannot estimate the two-way interactions nor the three way interaction because not only is there one observation per cell but there are only 16 cells rather than the required 64 as described in the text. I show what happens if you forget that the interactions cannot be estimated and ask R to estimate the interactions terms (i.e., use $*$ instead of $+$, which could be a natural mistake as one types commands quickly). As with the random block design, no error message but columns of the source table are missing, hinting there is a problem.

```
summary(aov(dv ~ driver * car * gas, data = data.lsd))


##             Df Sum Sq Mean Sq
## driver       3    5.9    1.97
## car          3  736.9  245.64
## gas          3  109.0   36.33
## driver:car   6   23.8    3.97
```

As in the randomized block case, in the Bayesian context if you ask for all possible interactions when you have a Latin Square design you will get output that looks like it ran smoothly, but all the interactions are mere reflections of the prior and are not 'updated' with data.

Factorial Design

First, we'll read in the teaching instruction data (computer vs standard across three subjects). I show how to label each factor and its levels.

```
teaching.data <- read.table("data.nl9",header=F)
colnames(teaching.data) <- c("subject","instruction",
                             "dv","group")
teaching.data$subject <-
  factor(teaching.data$subject,
         levels=1:3,labels=c("phy","soc","hist"))
teaching.data$instruction <-
  factor(teaching.data$instruction,
         levels=1:2,labels=c("computer","standard"))
```

There are several ways of computing cell and row/column marginal means in R. One is through the aggregate command, another is through the effects package, which requires a model,such as aov(), to have already been conducted. If there is an equal number of subjects in each group (balanced

design), and all terms are included in the model, then these different approaches will converge. But, if the sample size is unbalanced or terms are not included in the structural model, then the methods may lead to different results for the means. In LN4 I include all allowable model terms and have balanced samples, so the different methods converge, but in LN5 we'll examine the case of unbalanced designs and not including all the possible terms in the structural model.

The aggregate method to compute cell, row and column means produces this output:

```
aggregate(dv ~ instruction, teaching.data, mean)
```

```
##   instruction       dv
## 1    computer 41.33333
## 2    standard 25.66667
```

```
aggregate(dv ~ subject, teaching.data, mean)
```

```
##   subject   dv
## 1     phy 40.0
## 2     soc 26.0
## 3    hist 34.5
```

```
aggregate(dv ~ instruction * subject, teaching.data, mean)
```

```
##   instruction subject dv
## 1    computer     phy 46
## 2    standard     phy 34
## 3    computer     soc 40
## 4    standard     soc 12
## 5    computer    hist 38
## 6    standard    hist 31
```

I'll show the output of the means from the effects command after running the aov.

A factorial design is estimated by using an asterisk $*$ between the factors. The $*$ notation means to include the interaction as well as all lower terms such as the main effects of each factor listed on either side of the asterisk.

```
out <- aov(dv ~ subject * instruction, data = teaching.data)
summary(out)
```

```
##                    Df Sum Sq Mean Sq F value
```

```
## subject              2   1194   597.0  10.737
## instruction          1   2209  2209.0  39.730
## subject:instruction  2    722   361.0   6.493
## Residuals           30   1668    55.6
##                        Pr(>F)
## subject              0.000304 ***
## instruction          5.99e-07 ***
## subject:instruction 0.004539 **
## Residuals
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

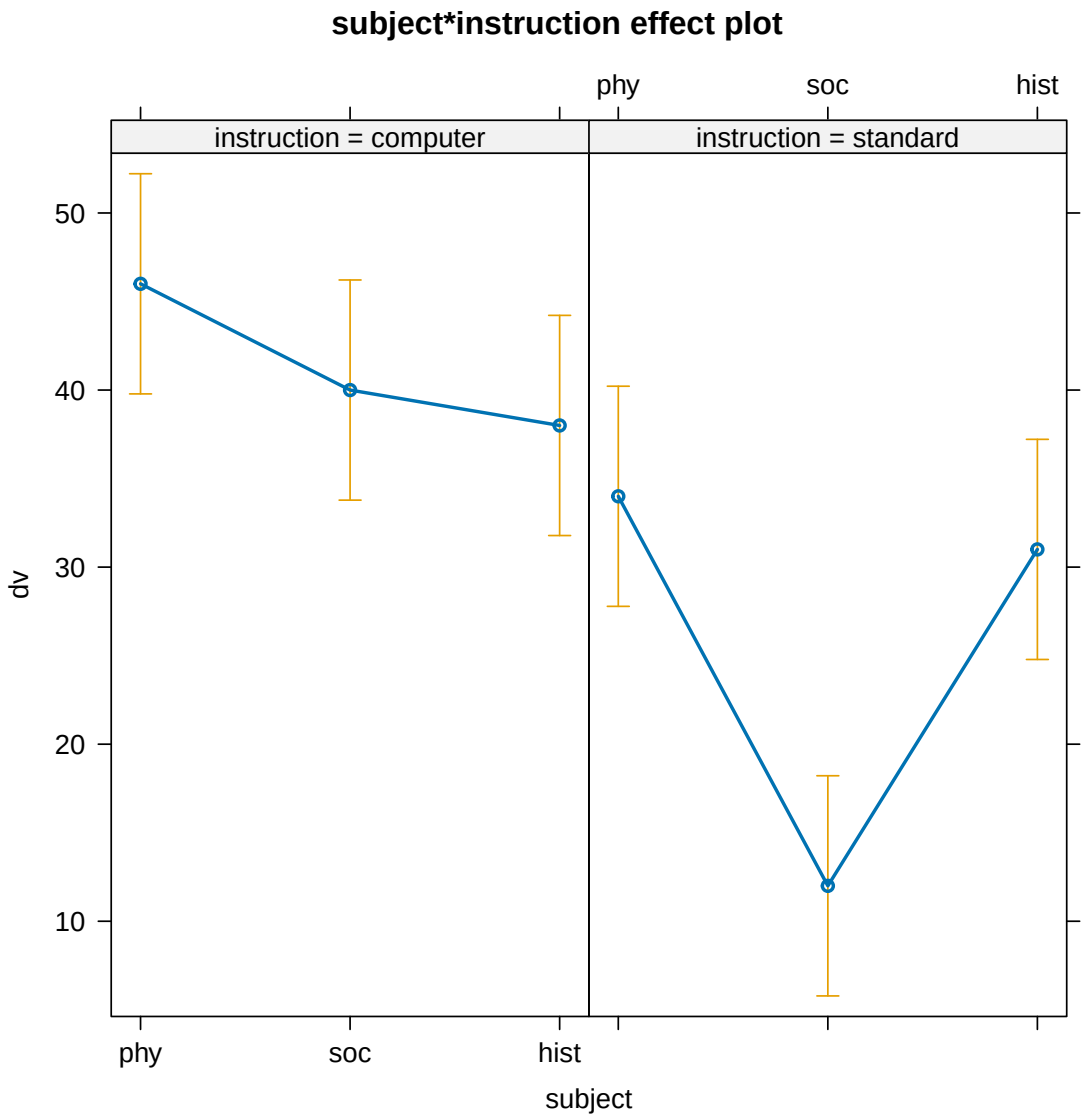
Or, equivalently, if you want to write out manually the complete structural model you do the following where the colon ":" denotes the specific interaction term (analogous to the word BY in SPSS)

```
out <- aov(dv ~ subject + instruction + subject:instruction,
           data=teaching.data)
summary(out)


##                     Df Sum Sq Mean Sq F value
## subject              2   1194   597.0  10.737
## instruction          1   2209  2209.0  39.730
## subject:instruction  2    722   361.0   6.493
## Residuals           30   1668    55.6
##                        Pr(>F)
## subject              0.000304 ***
## instruction          5.99e-07 ***
## subject:instruction 0.004539 **
## Residuals
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Again, the asterisk $*$ saves typing because you don't have to write all the lower effect terms. We will soon see cases where we do not necessarily want all the lower order terms, so knowing how to write specific terms manually will be useful.

You can use the output of aov and the effects command in the effects library to produce the cell and row/marginal means

```
library(effects)
effect(term = "instruction", mod = out)


##
##  instruction effect
## instruction
## computer standard
## 41.33333 25.66667


effect(term = "subject", mod = out)


##
##  subject effect
## subject
##  phy  soc hist
## 40.0 26.0 34.5


effect(term = "subject*instruction", mod = out)


##
##  subject*instruction effect
##        instruction
## subject computer standard
##    phy        46       34
##    soc        40       12
##    hist       38       31


# Some benefits of the effects command include plots
plot(effect(term = "subject*instruction", mod = out))
```



The effects command is nice but it requires a model to be run first. If the data are balanced and all the possible terms are in the model (i.e., a saturated model), then the means from the aggregate command will equal the means from the effects command. This won't always happen though as we will see in LN5. Also, the confidence interval around the means in the default effects plot is based on the pooled error term (i.e., the MSE from the ANOVA), so if the equal variance assumption doesn't hold, then the confidence interval is not a good representation of the variability across the plotted means.

The above models can be estimated using the lm() command instead of the aov() command. I recommend for now you limit the use of lm() command for when you have orthogonal contrasts and equal sample sizes in each group; one can run lm() with nonorthogonal contrasts and/or unequal sample sizes but one needs to be careful with interpreting the betas in the regression. More about

this in LN 5 and also when we get to regression in LN6-LN9; the issue was touched on in LN2 and LN3 of how lm() can go awry due to the default dummy code contrasts that create issues when the model include interactions. One still needs to run the anova() command on the output of the lm() command in order to print the source table (and there are a few options for this command and related commands, depending how nonorthogonality is handled, that we will go over in LN5).

In R, you compute contrasts on factorial designs the same way as described in LN3 for the one-way ANOVA. Each factor will have contrasts defined for it using the contrast() command, and you print the source table and t-tests as described in LN3.

I prefer NOT to use the default contrasts in R as they can lead to problems in the case of factorial designs. I typically define my own contrasts for each factor before doing any analyses. The default R contrasts are dummy codes and they can produce strange results For example, it can produce incorrect results that affect downstream analyses, such as affecting the aov() results and then when using fit.contrasts, which depends on aov() output. I'll illustrate the issues with the teaching instruction example where the subject factor has 3 levels and the instruction factor has 2 levels. First, let me produce the issue.

```
#using the fit.contrast function in gmodels (see LN3)
library(gmodels)
out <- aov(dv~subject*instruction, data=teaching.data)
fit.contrast(out,"subject", show.all=T, c(-1,1,0),
             df = T, conf.int = 0.95)


##                       Estimate Std. Error
## subject c=( -1 1 0 )       -6   4.305036
##                          t value  Pr(>|t|) DF
## subject c=( -1 1 0 ) -1.393717 0.1736427 30
##                           lower CI upper CI
## subject c=( -1 1 0 ) -14.79206 2.792056
```

You'll get an incorrect test because the aov() was based on default contrasts, which are not orthogonal, so the decomposition of sum of squares was off, and it produces issues for the downstream analysis used by fit.contrast() command. You can tell something is wrong because the marginal mean for phy is 40 and the marginal mean for soc is 26, so the difference should be 14, not 6 as reported by the the fit.contrast command.

Instead, one needs to define all contrasts for all factors before calling the aov() command as in

```
#these two lines save the contrast matrices directly to the factor
contrasts(teaching.data$subject) <- cbind(c(-1,1,0), c(1,1,-2))
contrasts(teaching.data$instruction) <- cbind(c(-1,1))

#the factors now "carry" those contrasts in subsequent calls;
```

```
#same two lines as in the previous "chunk" that earlier produced
#incorrect results now work correctly
out <- aov(dv~subject*instruction, data=teaching.data)
fit.contrast(out,"subject", show.all=T, c(-1,1,0), df = T,
             conf.int = 0.95)


##                        Estimate Std. Error  t value
## subject c=( -1 1 0 )        -14    3.04412 -4.59903
##                            Pr(>|t|) DF  lower CI
## subject c=( -1 1 0 ) 7.210211e-05 30 -20.21692
##                           upper CI
## subject c=( -1 1 0 ) -7.783078
```

and this yields the correct result for the fit.contrast to make use of when it tests the -1, 1, 0 marginal contrast on the subject factor (a difference of 14). You can see how this tiny point can lead to errors that are not easy to catch unless you are sanity checking your results. This is the same t observed of 4.599 shown in the SPSS output (contrast 2, main body of lecture notes).

An example of testing the (1, -1) contrast on the two levels of the instruction factor is

```
fit.contrast(out,"instruction", show.all=T, c(1,-1),
             df = T, conf.int = 0.95)


##                          Estimate Std. Error
## instruction c=( 1 -1 ) 15.66667   2.485514
##                           t value     Pr(>|t|) DF
## instruction c=( 1 -1 ) 6.303191 5.993468e-07 30
##                          lower CI upper CI
## instruction c=( 1 -1 ) 10.59057 20.74276
```

This leads to the same t observed of 6.303 shown in the SPSS output (contrast 3, main body of lecture notes).

For a completely between-subjects factorial design (what we're doing throughout this set of lecture notes), you could instead convert the factorial structure into a single one-way ANOVA (e.g., a 2x3 is recoded as a single factor with six groups), and then you apply the same commands as described in LN3. This just makes use of the fact that a completely between-subjects factorial ANOVA is equivalent to a one-way ANOVA with as many group as cells in the factorial design (see last Appendix in this lecture notes). In that case, the code I wrote for the separate variance contrast test and the Scheffe test (LN3) will work for recoded factorial designs. You can also use the tukeyHSD() command to test all possible pairwise differences in the individual cells in the factorial design. This recoding procedure to convert a factorial into a one-way ANOVA unfortunately loses the ability to perform Tukey tests on the marginal means because the main effect information is

lost. But you can compute those tests by hand using the formulas given earlier or use combine other approaches to pairwise tests (just keep track of number of groups and sample size in each group). For example, to perform Tukey tests on the cell means you can use the recoding method described in this paragraph and for Tukey tests on the marginal means you can use the emmeans command in the emmeans package as in

First, I'll show eemeans in the context of marginal effects using the factorial design, then show how to use emmeans in the recoded version of that design using six groups.

```
library(emmeans)
out <- aov(dv ~ subject * instruction, data = teaching.data)
out.emmeans <- emmeans(out, "subject")
summary(pairs(out.emmeans), adjust = "tukey", infer = T)


##  contrast   estimate   SE df lower.CL upper.CL
##  phy - soc      14.0 3.04 30      6.5   21.505
##  phy - hist      5.5 3.04 30     -2.0   13.005
##  soc - hist     -8.5 3.04 30    -16.0   -0.995
##  t.ratio p.value
##    4.599  0.0002
##    1.807  0.1846
##   -2.792  0.0238
##
## Results are averaged over the levels of: instruction
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
```

The infer=T option prints out both the p value and the confidence interval using the Tukey criterion.

The emmeans command also can (1) print marginal means as in

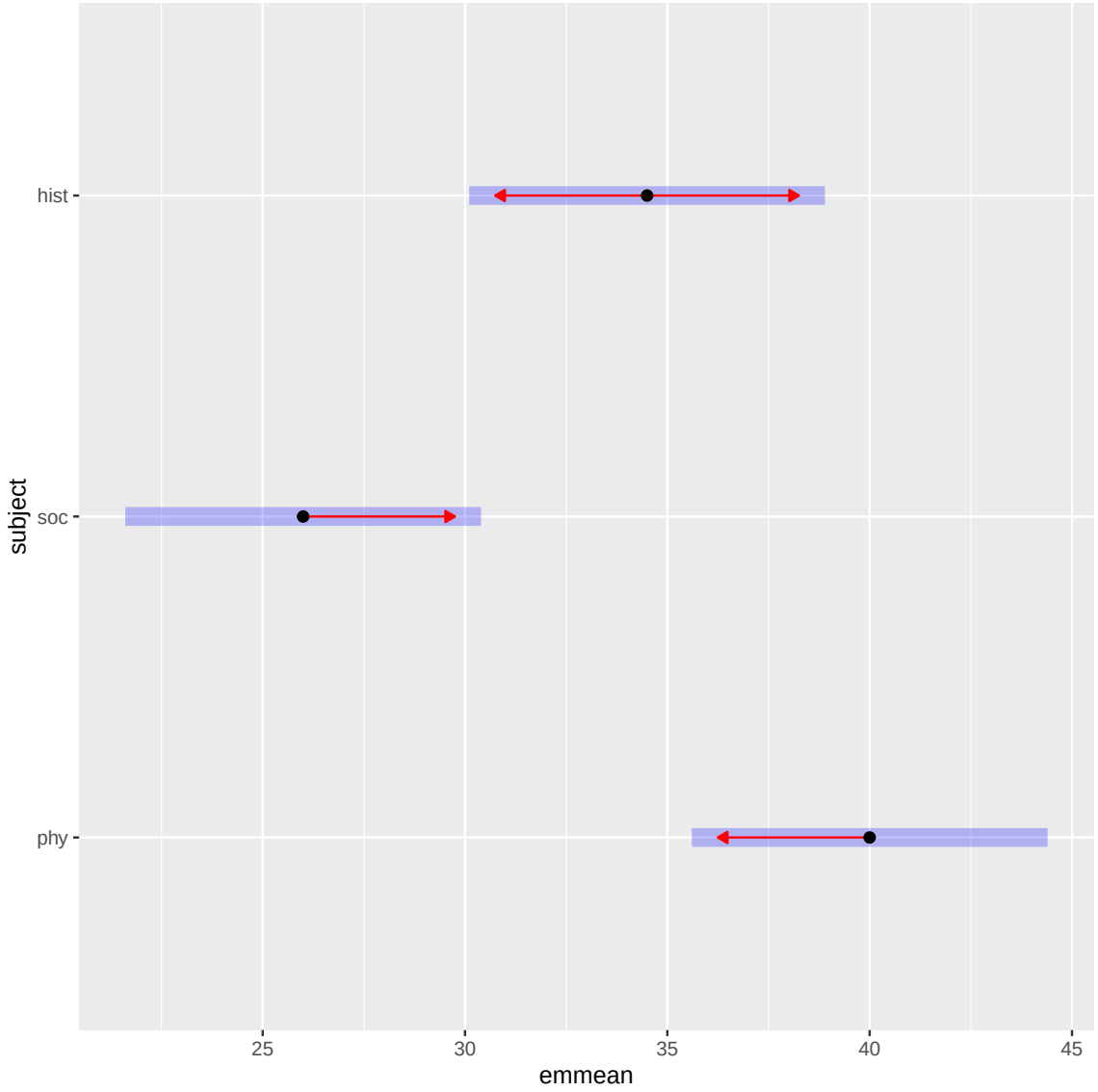
```
out.emmeans


##  subject emmean   SE df lower.CL upper.CL
##  phy       40.0 2.15 30     35.6     44.4
##  soc       26.0 2.15 30     21.6     30.4
##  hist      34.5 2.15 30     30.1     38.9
##
## Results are averaged over the levels of: instruction
## Confidence level used: 0.95
```

and it can (2) deal with doing a special type of Tukey tests across levels of another factor

```
emmeans(out, pairwise ~ subject | instruction)


## $emmeans
## instruction = computer:
##  subject emmean   SE df lower.CL upper.CL
##  phy         46 3.04 30    39.78     52.2
##  soc         40 3.04 30    33.78     46.2
##  hist        38 3.04 30    31.78     44.2
##
## instruction = standard:
##  subject emmean   SE df lower.CL upper.CL
##  phy         34 3.04 30    27.78     40.2
##  soc         12 3.04 30     5.78     18.2
##  hist        31 3.04 30    24.78     37.2
##
## Confidence level used: 0.95
##
## $contrasts
## instruction = computer:
##  contrast   estimate   SE df t.ratio p.value
##  phy - soc         6 4.31 30   1.394  0.3568
##  phy - hist        8 4.31 30   1.858  0.1684
##  soc - hist        2 4.31 30   0.465  0.8883
##
## instruction = standard:
##  contrast   estimate   SE df t.ratio p.value
##  phy - soc        22 4.31 30   5.110  <.0001
##  phy - hist        3 4.31 30   0.697  0.7671
##  soc - hist      -19 4.31 30  -4.413  0.0003
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

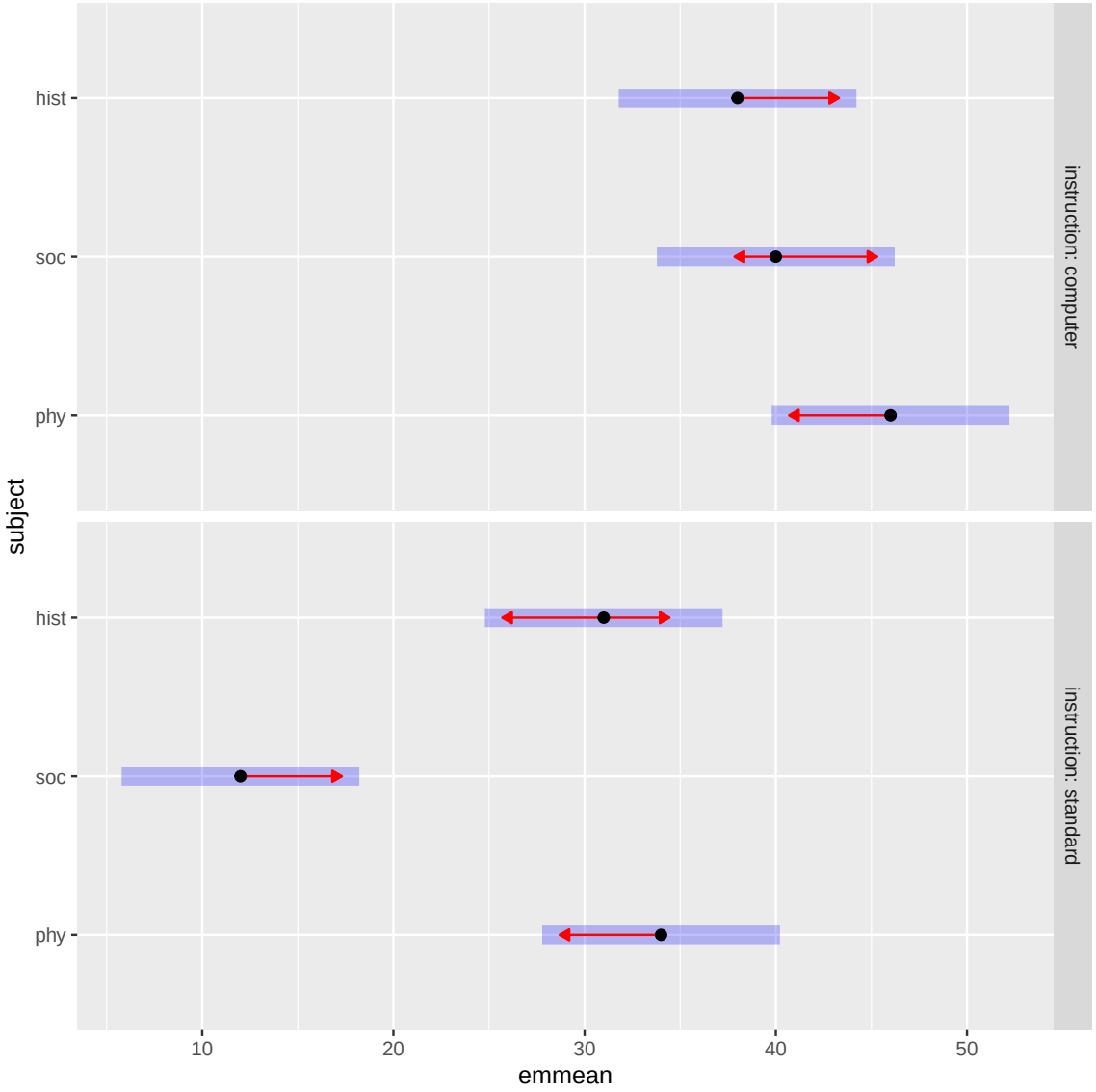
where we first see the 6 cell means with SE and CIs, and then two separate Tukey tests (one for the 3 levels of subject for instruction = computer and another for the 3 levels of subject for instruction = standard) and (3) produce nice plots illustrating the Tukey tests (blue bands are 95% CIs around the mean and the red arrows can be used for Tukey in that nonoverlapping red areas correspond to significant tests by the Tukey criterion) as well a plot of the 6 cell means. Here is a version of the 3 marginal means for the subject factor

```
plot(out.emmeans, comparisons = T, adjust = "tukey", int.adjust = "none")
```



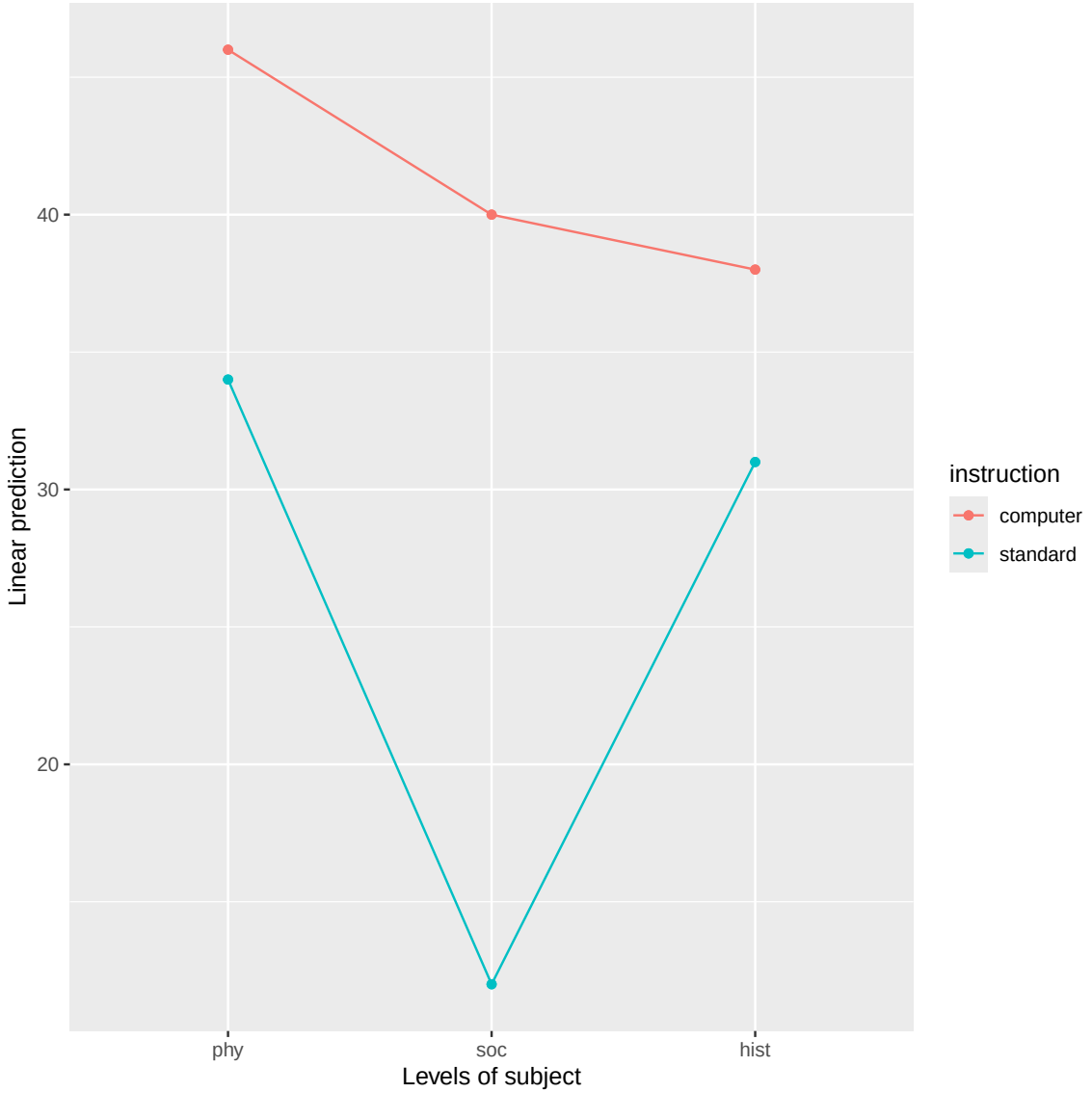
a version of the 3 means for the subject factor separately for each level of instruction

```
plot(emmeans(out, pairwise ~ subject | instruction),
     comparisons=T,
     adjust="tukey",int.adjust="none")
```



as well as a line graph of the 6 cell means

```
emmip(out, instruction ~ subject)
```



I recommend looking at the various options in the emmeans package as well as the related glht package that I introduced in a previous lecture notes for computing power; they have some nice features for comparisons and integration with the emmeans package.

The recoded version of this design using six cells testing the (0,1,0,-1,0,0) contrast to compare with the earlier SPSS output (contrast 4, body of lecture notes). This is the comparison of physical science to social science within the standard lecture condition.

```
teaching.data$group <- factor(teaching.data$group)
out <- aov(dv~group, data=teaching.data)
summary(out)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## group        5   4125   825.0   14.84 2.38e-07
## Residuals   30   1668    55.6
##
## group       ***
## Residuals
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit.contrast(out, "group", show.all = T, c(0,1,0,-1,0,0),
             df = T, conf.int = 0.95)
```

```
##                            Estimate Std. Error
## group c=( 0 1 0 -1 0 0 )         22   4.305036
##                             t value     Pr(>|t|) DF
## group c=( 0 1 0 -1 0 0 ) 5.110294 1.706157e-05 30
##                            lower CI upper CI
## group c=( 0 1 0 -1 0 0 ) 13.20794 30.79206
```

This is the identical output as in SPSS with a t observed of 5.11.

If we want to get fancy, we can test interaction contrasts using the six group recoded version. Say I want to test the interaction of (1,-1) on instruction and (1,-2,1) on content. As computed earlier, this is the (1,-1,-2,2,1,-1) where the order is phy:comp, phy:stand, soc:comp, soc:stand, hist:comp, hist:stand.

```
fit.contrast(out, "group", show.all = T, c(1,-1,-2,2,1,-1),
             df = T, conf.int = 0.95)
```

```
##                              Estimate Std. Error
## group c=( 1 -1 -2 2 1 -1 )        -37   10.54514
##                               t value    Pr(>|t|)
## group c=( 1 -1 -2 2 1 -1 ) -3.508725 0.001442978
##                              DF  lower CI  upper CI
## group c=( 1 -1 -2 2 1 -1 ) 30 -58.53605 -15.46395
```

And if you want to perform this test in the context of a factorial design, just define each of the two contrasts on the marginal and let the lm() command compute the interaction contrasts automatically to achieve the identical result.

```
contrasts(teaching.data$instruction) <- c(1, -1)
# subject has two ortho contrasts; put kry contrast first so easy to spot
contrasts(teaching.data$subject) <- cbind(c(1, -2, 1), c(1, 0, -1))
out.lm <- lm(dv ~ subject * instruction, data = teaching.data)
summary(out.lm)


##
## Call:
## lm(formula = dv ~ subject * instruction, data = teaching.data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -16.00  -3.25  -0.50   4.00  15.00
##
## Coefficients:
##                         Estimate Std. Error t value
## (Intercept)              33.5000     1.2428  26.956
## subject1                  3.7500     0.8788   4.267
## subject2                  2.7500     1.5221   1.807
## instruction1              7.8333     1.2428   6.303
## subject1:instruction1    -3.0833     0.8788  -3.509
## subject2:instruction1     1.2500     1.5221   0.821
##                         Pr(>|t|)
## (Intercept)              < 2e-16 ***
## subject1                0.000182 ***
## subject2                0.080840 .
## instruction1            5.99e-07 ***
## subject1:instruction1   0.001443 **
## subject2:instruction1   0.417980
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.457 on 30 degrees of freedom
## Multiple R-squared:  0.7121,Adjusted R-squared:  0.6641
## F-statistic: 14.84 on 5 and 30 DF,  p-value: 2.382e-07
```

You see a lot of output but focusing on the first interaction contrast (the first subject contrast with the first instruction contrast) we have the same output as the previous fit.contrast command, both yield a t observed of -3.509.

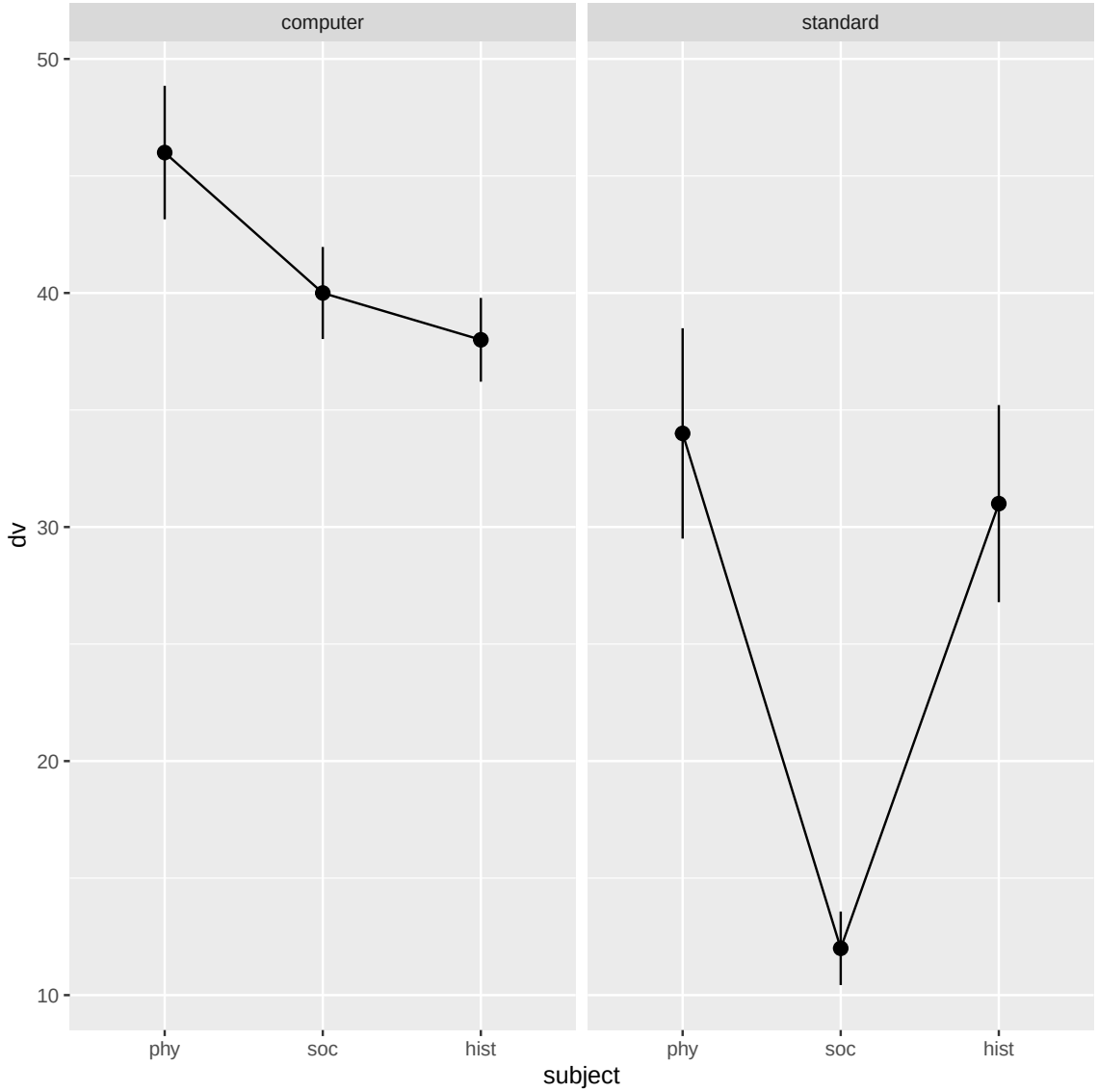
For completeness, here is the full set of 5 contrasts plus the grand mean contrast that the lm() automatically created by taking the three main effect contrasts supplied and multiplying them. The fancy syntax here is explained in the next example with the hays data set.

```
library(dplyr)
data.frame(model.matrix(out.lm)) %>%
    distinct()
```

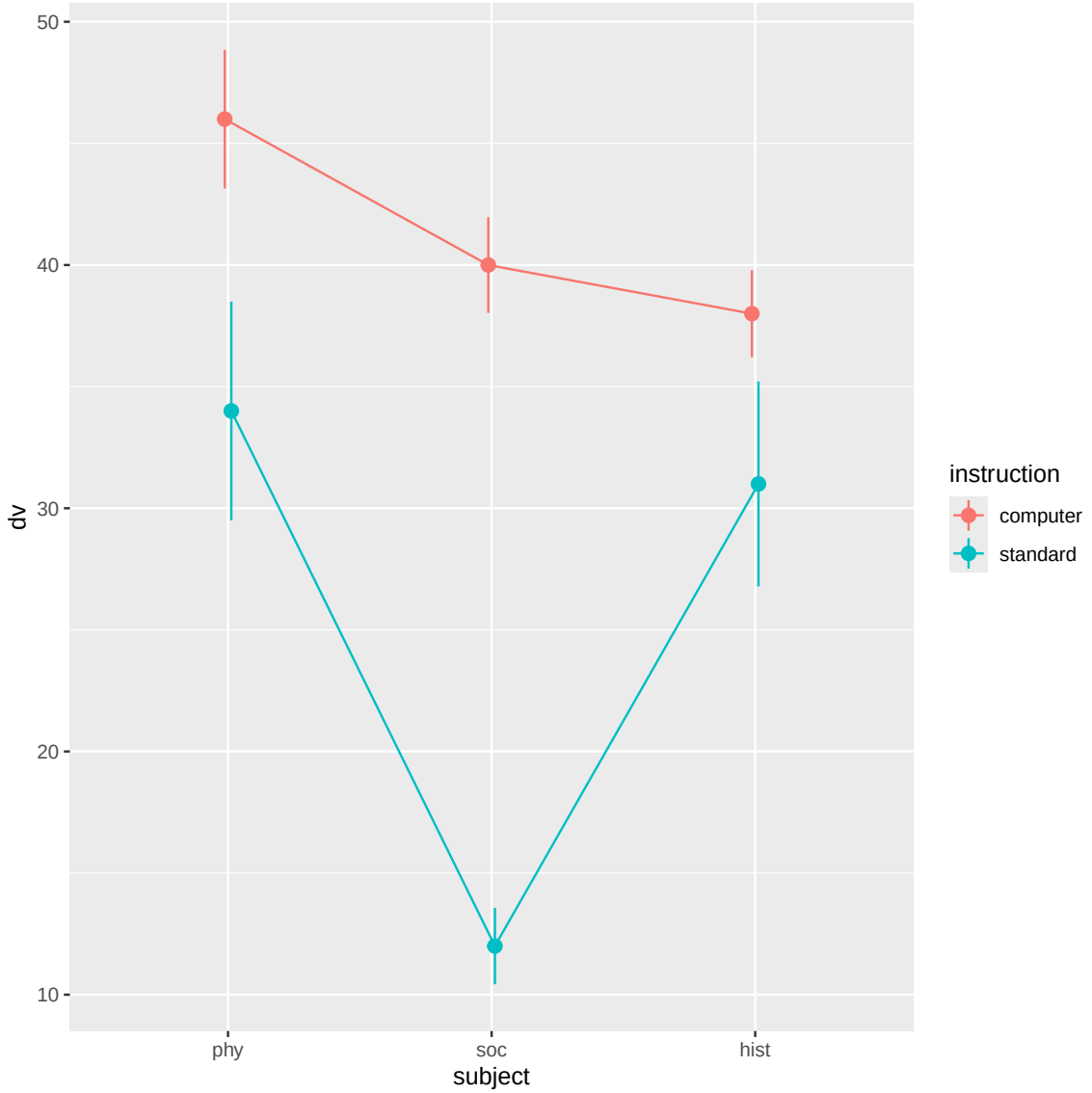
```
##    X.Intercept. subject1 subject2 instruction1
## 1             1        1        1            1
## 7             1        1        1           -1
## 13            1       -2        0            1
## 19            1       -2        0           -1
## 25            1        1       -1            1
## 31            1        1       -1           -1
##    subject1.instruction1 subject2.instruction1
## 1                      1                     1
## 7                     -1                    -1
## 13                    -2                     0
## 19                     2                     0
## 25                     1                    -1
## 31                    -1                     1
```

More example syntax for graphs: Here are a couple of different versions of ggplots for these data. The first version shows each of the therapists in different panels; the second uses one panel and color. In the second plot I used "dodge" so that two points would not be superimposed (not relevant here because the points in the two instruction groups don't overlap).

```
library(ggplot2)
ggplot(teaching.data, aes(subject, dv)) +
  stat_summary(fun.data="mean_se") +
  stat_summary(aes(group=instruction),fun.data="mean_se",
               geom="line") + facet_wrap(~instruction)
```




```
ggplot(teaching.data, aes(subject, dv, color = instruction)) + stat_summary(fun.
    position = position_dodge(width = 0.05)) + stat_summary(aes(group = instruct
    fun.data = "mean_se", geom = "line", position = position_dodge(width = 0.05)
```



Finally, here is the recoding of the 2x3 into a one way with 6 cells.

```
newcode <- 2 * (as.numeric(teaching.data$subject) - 1) +
  as.numeric(teaching.data$instruction)
newcode <- factor(newcode)

#double check it worked as expected
data.frame(subject=teaching.data$subject,
           instruction=teaching.data$instruction,
           newcode)

##    subject instruction newcode
```

```
## 1       phy    computer       1
## 2       phy    computer       1
## 3       phy    computer       1
## 4       phy    computer       1
## 5       phy    computer       1
## 6       phy    computer       1
## 7       phy    standard       2
## 8       phy    standard       2
## 9       phy    standard       2
## 10      phy    standard       2
## 11      phy    standard       2
## 12      phy    standard       2
## 13      soc    computer       3
## 14      soc    computer       3
## 15      soc    computer       3
## 16      soc    computer       3
## 17      soc    computer       3
## 18      soc    computer       3
## 19      soc    standard       4
## 20      soc    standard       4
## 21      soc    standard       4
## 22      soc    standard       4
## 23      soc    standard       4
## 24      soc    standard       4
## 25     hist    computer       5
## 26     hist    computer       5
## 27     hist    computer       5
## 28     hist    computer       5
## 29     hist    computer       5
## 30     hist    computer       5
## 31     hist    standard       6
## 32     hist    standard       6
## 33     hist    standard       6
## 34     hist    standard       6
## 35     hist    standard       6
## 36     hist    standard       6


#by now you should know what happens if you forget to execute
#factor(newcode)
out.oneway <- aov(dv ~ newcode, teaching.data)

#contrast to compare subject 1 to subject 2 ignoring subject 3
#and collapsing over the instruction factor
#which is equivalent to the 1,-1,0 marginal contrast on the
```

```
#subject factor (see above)
fit.contrast(out.oneway, "newcode",
             show.all=T, c(.5,.5, -.5,-.5, 0, 0),
             df=T, conf.int=.95)
```

```
##                                        Estimate
## newcode c=( 0.5 0.5 -0.5 -0.5 0 0 )       14
##                                        Std. Error
## newcode c=( 0.5 0.5 -0.5 -0.5 0 0 )    3.04412
##                                        t value
## newcode c=( 0.5 0.5 -0.5 -0.5 0 0 ) 4.59903
##                                            Pr(>|t|)
## newcode c=( 0.5 0.5 -0.5 -0.5 0 0 ) 7.210211e-05
##                                        DF lower CI
## newcode c=( 0.5 0.5 -0.5 -0.5 0 0 ) 30 7.783078
##                                        upper CI
## newcode c=( 0.5 0.5 -0.5 -0.5 0 0 ) 20.21692
```

```
#Tukey among the 6 cell means
TukeyHSD(out.oneway,"newcode")
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = dv ~ newcode, data = teaching.data)
##
## $newcode
##     diff        lwr        upr     p adj
## 2-1  -12 -25.094172   1.094172 0.0873690
## 3-1   -6 -19.094172   7.094172 0.7303314
## 4-1  -34 -47.094172 -20.905828 0.0000001
## 5-1   -8 -21.094172   5.094172 0.4459890
## 6-1  -15 -28.094172  -1.905828 0.0174701
## 3-2    6  -7.094172  19.094172 0.7303314
## 4-2  -22 -35.094172  -8.905828 0.0002286
## 5-2    4  -9.094172  17.094172 0.9357513
## 6-2   -3 -16.094172  10.094172 0.9808926
## 4-3  -28 -41.094172 -14.905828 0.0000048
## 5-3   -2 -15.094172  11.094172 0.9970248
## 6-3   -9 -22.094172   4.094172 0.3189838
## 5-4   26  12.905828  39.094172 0.0000174
## 6-4   19   5.905828  32.094172 0.0015496
## 6-5   -7 -20.094172   6.094172 0.5885943
```

But other bells and whistles like Tukey with Welch or Scheffe on arbitrary contrasts with/without Welch need to be done by hand or can make use of the functions described in LN3 after converting the factorial between subjects design into a one-way between subjects design as described here.

Biofeedback example in R

I'll simply provide the ANOVA source table and basic tests of the contrasts on the main effects, which are identical to the omnibus F tests because there have 1 degree of freedom in the numerator. The rest of the syntax, including plots, follows the previous 2 x 3 example. In a 2 x 2 there are no Tukey tests to compute.

```
data.bio <-
  read.table("/Users/gonzo/rich/Teach/Gradst~1/unixfiles/lectnotes/lect4/data.nl
names(data.bio) <- c("biofeed", "drug", "blood", "group")
data.bio[,1] <- factor(data.bio[,1])
data.bio[,2] <- factor(data.bio[,2])
contrasts(data.bio[,1])<-  cbind(c(1,-1))
contrasts(data.bio[,2])<-   cbind(c(1,-1))
data.bio <- data.frame(data.bio)

out.bio <- aov(blood ~ biofeed*drug, data.bio)
summary(out.bio)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## biofeed       1    500   500.0    8.00 0.01211 *
## drug          1    720   720.0   11.52 0.00371 **
## biofeed:drug  1    320   320.0    5.12 0.03792 *
## Residuals    16   1000    62.5
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(gmodels)
fit.contrast(out.bio,"biofeed", show.all=T, c(1,-1),
             df = T, conf.int = 0.95)
```

```
##                    Estimate Std. Error  t value
## biofeed c=( 1 -1 )       10   3.535534 2.828427
##                       Pr(>|t|) DF lower CI
## biofeed c=( 1 -1 ) 0.01210934 16 2.505003
##                    upper CI
## biofeed c=( 1 -1 )   17.495
```

```
fit.contrast(out.bio,"drug", show.all=T, c(1,-1),
             df = T, conf.int = 0.95)


##                  Estimate Std. Error  t value
## drug c=( 1 -1 )        12   3.535534 3.394113
##                      Pr(>|t|) DF lower CI upper CI
## drug c=( 1 -1 ) 0.003705919 16 4.505003   19.495
```

Additional factorial Example (material similar to Appendix 7 in SPSS)

This is a 3x4 factorial design testing the effects of wall color and room size on anxiety level during an interview (Hays, 1988). Assume that the anxiety measure is reliable. Lower numbers reflect lower anxiety scores. The four room colors were red, yellow, green, and blue (coded as 1, 2, 3, and 4, respectively). The three room sizes were small, medium, and large (coded 1, 2, and 3).

```
hays.data <- read.table("hays.dat",header=F)
names(hays.data) <- c("size","color","anxiety")
hays.data$size <- factor(hays.data$size, levels=1:3,
                         labels=c("small","medium","large"))
hays.data$color <- factor(hays.data$color, levels=1:4,
                          labels=c("red","yellow","green","blue"))
```

First, conduct an ANOVA. Unfortunately, this only produces the source table and not the output of the contrasts.

```
hays.aov <- aov(anxiety ~ size * color, hays.data)
summary(hays.aov)


##             Df Sum Sq Mean Sq F value   Pr(>F)
## size         2    478     239   0.769    0.475
## color        3  31526   10509  33.839 8.73e-09
## size:color   6   3664     611   1.966    0.111
## Residuals   24   7453     311
##
## size
## color       ***
## size:color
## Residuals
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To see what contrast values R selected by default you can do the contrast command to print out the contrasts. You can see these are dummy codes, and this will create issues with interactions as we saw before (such as when dealing with the fit.contrast command).

```
contrasts(hays.data$size)
```

```
##        medium large
## small       0     0
## medium      1     0
## large       0     1
```

```
contrasts(hays.data$color)
```

```
##        yellow green blue
## red         0     0    0
## yellow      1     0    0
## green       0     1    0
## blue        0     0    1
```

To see what complete set of contrasts, including the interaction contrasts, were assigned to each cell (actually each participant in each cell, but within a cell all participants are assigned equal values) you can use the model.matrix() command. To save space, I'm selecting only distinct rows in the model.matrix. The %>% syntax is the pipe command. This means "send the output from the command on the left side and make it the input to the command on the right side."

```
library(dplyr)
data.frame(model.matrix(hays.aov)) %>%
    distinct()
```

```
##    X.Intercept. sizemedium sizelarge coloryellow
## 1             1          0         0           0
## 4             1          0         0           1
## 7             1          0         0           0
## 10            1          0         0           0
## 13            1          1         0           0
## 16            1          1         0           1
## 19            1          1         0           0
## 22            1          1         0           0
## 25            1          0         1           0
## 28            1          0         1           1
## 31            1          0         1           0
## 34            1          0         1           0
```

```
##    colorgreen colorblue sizemedium.coloryellow
## 1           0         0                       0
## 4           0         0                       0
## 7           1         0                       0
## 10          0         1                       0
## 13          0         0                       0
## 16          0         0                       1
## 19          1         0                       0
## 22          0         1                       0
## 25          0         0                       0
## 28          0         0                       0
## 31          1         0                       0
## 34          0         1                       0
##    sizelarge.coloryellow sizemedium.colorgreen
## 1                      0                      0
## 4                      0                      0
## 7                      0                      0
## 10                     0                      0
## 13                     0                      0
## 16                     0                      0
## 19                     0                      1
## 22                     0                      0
## 25                     0                      0
## 28                     1                      0
## 31                     0                      0
## 34                     0                      0
##    sizelarge.colorgreen sizemedium.colorblue
## 1                     0                     0
## 4                     0                     0
## 7                     0                     0
## 10                    0                     0
## 13                    0                     0
## 16                    0                     0
## 19                    0                     0
## 22                    0                     1
## 25                    0                     0
## 28                    0                     0
## 31                    1                     0
## 34                    0                     0
##    sizelarge.colorblue
## 1                    0
## 4                    0
## 7                    0
## 10                   0
```

```
## 13                    0
## 16                    0
## 19                    0
## 22                    0
## 25                    0
## 28                    0
## 31                    0
## 34                    1
```

I'll redo that example with a particular set of orthogonal contrasts. I'll re-assign a specific set of contrasts to each factor. The resulting ANOVA source table is identical as with the default set of contrasts. I also print out each of the contrast tests using the "split" option in the summary command. The results are identical is in Appendix 7 with SPSS.

```
contrasts(hays.data$size) <- cbind(c(1, -1, 0),c(1,1,-2))
contrasts(hays.data$color) <- cbind(c(1, 1, -1, -1),
                                    c(1,-1,-1,1),c(1,-1,1,-1))

hays.aov2 <- aov(anxiety ~ size * color, hays.data)
summary(hays.aov2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## size         2    478     239   0.769    0.475
## color        3  31526   10509  33.839 8.73e-09
## size:color   6   3664     611   1.966    0.111
## Residuals   24   7453     311
##
## size
## color       ***
## size:color
## Residuals
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(hays.aov2, split=list(size=list("s.C1"=1, "s.C2" = 2),
                              color=list("c.C1"=1, "c.C2"=2,"c.C3"=3)))
```

```
##                          Df Sum Sq Mean Sq F value
## size                      2    478     239   0.769
##   size: s.C1              1     37      37   0.121
##   size: s.C2              1    440     440   1.417
```

```
## color                     3  31526   10509  33.839
##   color: c.C1             1  30508   30508  98.238
##   color: c.C2             1    289     289   0.931
##   color: c.C3             1    729     729   2.347
## size:color                6   3664     611   1.966
##   size:color: s.C1.c.C1   1   1094    1094   3.521
##   size:color: s.C2.c.C1   1     53      53   0.172
##   size:color: s.C1.c.C2   1    451     451   1.451
##   size:color: s.C2.c.C2   1    200     200   0.644
##   size:color: s.C1.c.C3   1   1667    1667   5.367
##   size:color: s.C2.c.C3   1    200     200   0.644
## Residuals                24   7453     311
##                            Pr(>F)
## size                       0.4746
##   size: s.C1               0.7313
##   size: s.C2               0.2455
## color                    8.73e-09 ***
##   color: c.C1            5.85e-10 ***
##   color: c.C2              0.3443
##   color: c.C3              0.1386
## size:color                 0.1107
##   size:color: s.C1.c.C1    0.0728 .
##   size:color: s.C2.c.C1    0.6821
##   size:color: s.C1.c.C2    0.2401
##   size:color: s.C2.c.C2    0.4301
##   size:color: s.C1.c.C3    0.0294 *
##   size:color: s.C2.c.C3    0.4301
## Residuals
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The resulting model.matrix() command for this set of contrasts is

```
data.frame(model.matrix(hays.aov2)) %>%
    distinct()


##    X.Intercept. size1 size2 color1 color2 color3
## 1             1     1     1      1      1      1
## 4             1     1     1      1     -1     -1
## 7             1     1     1     -1     -1      1
## 10            1     1     1     -1      1     -1
## 13            1    -1     1      1      1      1
## 16            1    -1     1      1     -1     -1
```

```
## 19             1    -1     1     -1     -1      1
## 22             1    -1     1     -1      1     -1
## 25             1     0    -2      1      1      1
## 28             1     0    -2      1     -1     -1
## 31             1     0    -2     -1     -1      1
## 34             1     0    -2     -1      1     -1
##    size1.color1 size2.color1 size1.color2
## 1             1             1             1
## 4             1             1            -1
## 7            -1            -1            -1
## 10           -1            -1             1
## 13           -1             1            -1
## 16           -1             1             1
## 19            1            -1             1
## 22            1            -1            -1
## 25            0            -2             0
## 28            0            -2             0
## 31            0             2             0
## 34            0             2             0
##    size2.color2 size1.color3 size2.color3
## 1             1             1             1
## 4            -1            -1            -1
## 7            -1             1             1
## 10            1            -1            -1
## 13            1            -1             1
## 16           -1             1            -1
## 19           -1            -1             1
## 22            1             1            -1
## 25           -2             0            -2
## 28            2             0             2
## 31            2             0            -2
## 34           -2             0             2
```

Of course, one could convert this to a oneway ANOVA design with 12 groups (3x4) and compute the 11 orthogonal contrasts to run in the fit.contrast() command. The results would be identical if you selected the same set of 11 orthogonal contrasts (i.e., the same 2 for size, same 3 for color, and same 6 for the interactions).

Appendix 9: Coding Factorial ANOVAs as One-way ANOVAs

(adapted from a handout prepared by Steve Holste, a former TA at the University of Washington) In the lecture notes I showed how a factorial ANOVA can be recoded into a One-way ANOVA. In the examples I did the coding manually in the data set by creating a new column of codes. For instance, a 3×4 ANOVA can be represented as follows

	Variable B			
	1	2	3	4
1				
2				
3				

Now I'll put in the codes for row and column denoting each cell.

	Variable B			
Variable A	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4

Suppose we observe two subjects in each of the 12 cells as follows:

	Variable B			
Variable A	1	2	3	4
1	8 3	5 9	2 4	6 3
2	3 1	2 5	4 8	6 7
3	2 2	3 5	1 9	8 1

SPSS

In SPSS I would enter the data as follows (first column codes the row variable, the second column codes the column variable, and the third column lists data).

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

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

To recode this factorial design as a oneway ANOVA I could do one of three things:

1. Create a new column of data manually that codes the 12 groups from 1 to 12. This is tedious and will work properly, but is probably not something you want to do when you have 100s or 1000s of subjects because of all the typing (and chances to make errors).

2. Use a series of if-then statements in SPSS. For instance, the following creates a new set of codes ranging from 1 to 12:

```
if (varA = 1 and varB = 1) newcode = 1.
if (varA = 1 and varB = 2) newcode = 2.
if (varA = 1 and varB = 3) newcode = 3.
if (varA = 1 and varB = 4) newcode = 4.
if (varA - 2 and varB = 1) newcode = 5.
ETC
if (varA = 3 and varB = 4) newcode = 12.
execute.
```

R The analogous version of this in R would be

```
#initialize variable; N is total number of subjects
newcode <- rep(0,N)

#use ifelse which has form BOOLEAN, return if TRUE, return if FALSE
```

```
newcode <- ifelse(varA == 1 & varB == 1, 1, newcode)
newcode <- ifelse(varA == 1 & varB == 2, 2, newcode)
newcode <- ifelse(varA == 1 & varB == 3, 3, newcode)
newcode <- ifelse(varA == 1 & varB == 4, 4, newcode)
newcode <- ifelse(varA == 2 & varB == 1, 5, newcode)
ETC
newcode <- ifelse(varA == 3 & varB == 4, 12, newcode)
```

The variable newcode will now be numbers 1-12 depending on the combination of varA and varB. Note that if you have levels of factor defined as strings you would have something like varA == "level 1 label" rather than the numerical varA == 1.

3. The "I can do it in one line" procedure.

SPSS

In SPSS, the following line creates a new grouping code ranging from 1 to 12:

```
compute newcode = 4 * (varA - 1) + varB.
```

On some versions of SPSS you may need to follow that compute line with the "execute." command.

Double check that this formula works as advertised. When varA = 1 and varB = 1, newcode will be 1; ETC.; when varA = 3 and varB = 4, newcode will be 12.

The general formula for this "one-liner" is:

```
compute newcode = [number of levels in varB] * (varA - 1) + varB
```

If your codes do not start at 1, you'll need to adjust the formula accordingly.

R

For R, the one liner is

```
newcode <- 4 * (varA - 1) + varB
```

where the 4 is the number of levels for varB and the varA and varB terms refer the actual variables names used for those two predictors This syntax for the example in this appendix will

create the values 1-12 depending on the combination of varA and varB and store the values in the new variable called newcode. newcode is a grouping code with 12 levels emerging from the 3 levels of varA and the 4 levels of varB.

I'll run this R code for the teaching example data set from the previous Appendix. This syntax though won't work on factors directly. It is necessary to use the numeric form of these predictors like levels 1, 2, 3 and 4. So, it is easier to apply before converting the variables to factors. One needs to double check the conversion worked as expected as there can be issues when converting factors back to their numeric values.

```
teaching.data$subject.num <- as.integer(teaching.data$subject)
teaching.data$instruction.num <- as.integer(teaching.data$instruction)

teaching.data$newcode <- 2 * (teaching.data$subject.num - 1) +
  teaching.data$instruction.num
teaching.data
```

```
##    subject instruction dv group subject.num
## 1      phy    computer 53     1           1
## 2      phy    computer 49     1           1
## 3      phy    computer 47     1           1
## 4      phy    computer 42     1           1
## 5      phy    computer 51     1           1
## 6      phy    computer 34     1           1
## 7      phy    standard 44     2           1
## 8      phy    standard 48     2           1
## 9      phy    standard 35     2           1
## 10     phy    standard 18     2           1
## 11     phy    standard 32     2           1
## 12     phy    standard 27     2           1
## 13     soc    computer 47     3           2
## 14     soc    computer 42     3           2
## 15     soc    computer 39     3           2
## 16     soc    computer 37     3           2
## 17     soc    computer 42     3           2
## 18     soc    computer 33     3           2
## 19     soc    standard 13     4           2
## 20     soc    standard 16     4           2
## 21     soc    standard 16     4           2
## 22     soc    standard 10     4           2
## 23     soc    standard 11     4           2
## 24     soc    standard  6     4           2
## 25    hist    computer 45     5           3
```

```
## 26    hist    computer 41     5            3
## 27    hist    computer 38     5            3
## 28    hist    computer 36     5            3
## 29    hist    computer 35     5            3
## 30    hist    computer 33     5            3
## 31    hist    standard 46     6            3
## 32    hist    standard 40     6            3
## 33    hist    standard 29     6            3
## 34    hist    standard 21     6            3
## 35    hist    standard 30     6            3
## 36    hist    standard 20     6            3
##    instruction.num newcode
## 1                1       1
## 2                1       1
## 3                1       1
## 4                1       1
## 5                1       1
## 6                1       1
## 7                2       2
## 8                2       2
## 9                2       2
## 10               2       2
## 11               2       2
## 12               2       2
## 13               1       3
## 14               1       3
## 15               1       3
## 16               1       3
## 17               1       3
## 18               1       3
## 19               2       4
## 20               2       4
## 21               2       4
## 22               2       4
## 23               2       4
## 24               2       4
## 25               1       5
## 26               1       5
## 27               1       5
## 28               1       5
## 29               1       5
## 30               1       5
## 31               2       6
## 32               2       6
```

```
## 33                2        6
## 34                2        6
## 35                2        6
## 36                2        6
```

We see this code reproduces the previous values I hand coded into the variable group.

Any of the above procedures can be used to achieve the identical result of re-coding a factorial, between-subjects ANOVA into a oneway ANOVA.