Planned Contrasts

Following a significant one-way analysis of variance (ANOVA), the researcher may be interested in following up the analysis with some specific comparisons. In the case of the *planned contrast* or *planned comparison*, only a few predicted or *a priori* hypotheses are of interest, and familywise error is not likely to be a serious concern. *Post hoc* tests that adjust for familywise error typically follow a significant one-way ANOVA when many or all possible comparisons are of interest. Philosophically, the distinction between an *a priori* and *post hoc* test has to do with whether or not the group means compared were predicted to be different in advance or are decided after looking at the results. Statistically, the distinction also concerns whether there are a few or many contrasts conducted. Statisticians will cite either the philosophical or the statistical reason for deciding between the two approaches. We know that with many contrasts, familywise error becomes a problem, but, if not very many contrasts are performed (e.g., 2, 3, 4?) and there are *a priori* hypotheses about what groups may differ, then it is likely to be safe to use a planned contrast approach.

Planned contrasts typically involve the comparison of just two means. More complicated tests can be conducted (e.g., in a three-group design, the average of two groups might be compared to the third group), but I will not get into demonstrating more complicated comparisons in this handout (see Keppel & Wickens, 2004, for more detail). The approach is to develop a set of weights that eliminate any group means that are not involved in the comparison by giving them a zero weight and to specify the group means to be compared by giving them opposite values, usually -1 and +1. Thus, the first step is to obtain the weighted sum, $\hat{\psi}$, that gives the appropriate difference between the two means that one wishes to compare.

$$\hat{\psi} = \sum w_j \bar{Y}_j$$

In the formula, $\bar{Y}_j$ represents the group mean for each cell and w_i represents the contrast weights or "coefficients." In a three-group design, a comparison between the first and second group means uses the weights of -1, +1, and 0. The third mean drops out because it is multiplied by 0. The next step is to compute the standard error: [1]

$$s_{\hat{\psi}} = \sqrt{MS_{s/A} \sum \frac{w_j^2}{n_j}}$$

The mean-square error, $MS_{s/A}$, is then obtained from the full one-way (omnibus) ANOVA. Here, we use an estimate of error derived from using within-group variability of <u>all cases</u> in the study. This approach gives a more stable estimate of error and a more powerful statistical test than if we simply conducted a standard t test. The next step is simply to compute the t value using the familiar ratio of the difference to the standard error.

$$t_{contrast} = \frac{\hat{\psi}}{s_{\hat{\psi}}}$$

Although still fairly common practice, it is not advisable to conduct a standard t test after a significant ANOVA. Statistical power is lower with the standard t test compared than it is with the planned contrast version for two reasons: a) the sample size is smaller with the t test, because only the cases in the two groups are selected; and b) in the planned contrast the error term is smaller than it is with the standard t test because it is based on all the cases from the ANOVA. As with a standard t test, we can use a

[1] I've followed the textbook (Myers, Well, & Lorch, 2010) by computing a planned t-test, but many books and some software packages us an F-test. In the F-test version, the contrast weights are used to compute the numerator sum of squares. The mean square for the contrast is then divided by the means square error in a familiar F ratio. The two tests are statistically equivalent, however, because $t_\psi^2 = F_\psi$.

Welch's robust approach if there are concerns about equal variances (see Myers, Well, & Lorch, p. 245, for details).

I will use the study strategy example from the one-way ANOVA handout to illustrate the computation of the planned contrast: reading only (M = 6), retrieval practice (M = 9), and concept mapping (M = 6). The omnibus ANOVA was significant (see ANOVA Example handout), and it might be desirable to follow the test with a comparison of the means for reading only and retrieval practice groups for theoretical or policy reasons. Thus, the comparison involves the first two groups, and the contrast weights should be -1, +1, and 0.

$$\begin{aligned}\hat{\psi} &= \sum w_j \bar{Y}_j \\ &= -1(6) + 1(9) + 0(6) \\ &= -6 + 9 \\ &= 3\end{aligned}$$

$$\begin{aligned}s_{\hat{\psi}} &= \sqrt{MS_{s/A} \sum \frac{w_j^2}{n_j}} \\ &= \sqrt{1.833\left[\frac{(-1)^2}{5} + \frac{(1)^2}{5} + \frac{(0)^2}{5}\right]} \\ &= \sqrt{1.83\left[\frac{2}{5}\right]} \\ &= \sqrt{.732} \\ &= .856\end{aligned}$$

$$\begin{aligned}t_{contrast} &= \hat{\psi} / s_{\hat{\psi}} \\ &= \frac{3}{.856} \\ &= 3.50\end{aligned}$$

The critical value for a t test with $df = N - a$ = 15 -3 = 12 is 2.179. Because our calculated value of 3.50 is greater the critical value, the difference is significant.

To obtain this contrast in SPSS, a contrast subcommand can be added to the one-way ANOVA using the following syntax and this produces a *t* test version of the planned comparison.[2] For this example, which involves a very small sample size and very different variances, it would be wise to look at the Welch's test ("Does not assume equal variances" row).

```
ONEWAY recall BY groups
  /CONTRAST= -1 1 0.
```

ANOVA

recall number of new terms recalled

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	30.000	2	15.000	8.182	.006
Within Groups	22.000	12	1.833		
Total	52.000	14			

Contrast Coefficients

	groups learning strategy experimental groups		
Contrast	1.00 reading only	2.00 retrieval practice	3.00 concept mapping
1	-1	1	0

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)	95% Confidence Interval Lower	Upper
recall number of new terms recalled	Assumes equal variances	1	3.0000	.85635	3.503	12	.004	1.1342	4.8658
	Does not assume equal variances	1	3.0000	1.00000	3.000	5.882	.025	.5412	5.4588

R

The R function I use is an F-test version. The planned contrasts can also be stated in terms of an F test. In fact, the simple, old-fashioned way to conduct a planned comparison was to conduct an F test using only the two groups you want to compare, use the mean square from that analysis (or sum of squares

[2] The MANOVA command can also be used, but the subcommand requires that you specify a -1 comparisons, where a is the number of levels (see the simple effects analysis handout for more details).

because $MS_{A(1vs2)} = SS_{A(1vs2)}/1$), but divide it by the mean square error from the full omnibus ANOVA, with $F = MS_{A(1vs2)}/MS_{s/A}$. As it turns out, this method is equal to the t contrast method above, because $t^2 = F$.

I do not know of any R package that provides a Welch's adjustment for unequal variances for the planned comparison, so you would have to rely on assumed equal variances (which is generally unproblematic for studies with larger sample sizes and less dramatic differences between variances).

Note that if you have missing data, R may have problems running the analysis with the approach given below. And, if the conditions are not in ascending order in the data set, the analysis will not be correct. I had no missing data and my data were ordered, but for other data sets, you may need the first two statements I have commented out.

```
#contrasts do not work with missing data, so use listwise deletion
#d <- d[complete.cases(d), ]

#need to sort data first
#library(dplyr)
#d = (d[order(d$groups, decreasing = FALSE), ] )

> #planned contrasts -- there must be a-1 contrasts for results to be correct
> #I will ignore the second test, because it is not one that I wanted
> #compare the first and second means
> #(each command specifies number of cases in each group)
> d$c1 <- rep(c(-1, 1, 0), each = 5)
> #compare the first and third means
> d$c3 <- rep(c(-1, 0, 1), each = 5)
>
> #request the first contrast, c1, and second contrast, c3
> anova(lm(recall ~ c1 + c3, d))


Analysis of Variance Table

Response: recall
          Df Sum Sq Mean Sq F value   Pr(>F)
c1         1   22.5 22.5000 12.2727 0.004356 **
c3         1    7.5  7.5000  4.0909 0.065982 .
Residuals 12   22.0  1.8333
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Notice that the square of the *t*-value from SPSS (assumed equal variances) equals the c1 *F* from R, $(3.503)^2 = 12.27$).

Write-Up

(assumes omnibus ANOVA test already reported)

A planned contrast, using the Welch's adjustment for unequal variances, indicated that student recall in the retrieval practice group (M = 9.00, SD = 1.00) was significantly higher than in the reading only group (M = 6.00, SD = 2.00), $t(5.882) = 3.00$, $p = .025$.

Note: In most instances in practice, the sample size will be sufficient (say > 15 per group) and the variances will not be dramatically different (e.g., not larger than a 4:1 ratio), so reporting the equal variances assumed value is unlikely lead to incorrect decisions (see Myers et al., 2010, p. 137). In this instance, it does not make a difference in the statistical decision even with the more extreme circumstances.