

Notes on Maxwell & Delaney

PSYCH 710

4 Chapter 4 - Individual Comparisons of Means

4.1 Omnibus vs. Focused Tests

For the one-way ANOVAs that we have been discussing so far, a significant F test means that we reject the null hypothesis $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$ in favor of the alternative hypothesis $H_1: \alpha_j \neq 0$ (for at least one group, j). So, when $a \geq 3$, a significant omnibus F test does not tell us precisely how the group means differ from each other. To address this question, and make our inferences more precise, we need to use more focused comparisons of the group means.

There is another reason for using focused comparisons. In a sense, the omnibus F test examines all possible patterns of differences among the group means. The advantage of this test is obvious: we don't need to specify the pattern of effects in advance. However, there is a penalty to this general approach, namely reduced power. As we shall see, asking a focused question can result in a statistical test with much greater power than the omnibus F test.

The following example shows that a focused comparison among means can be significant even when the omnibus F test is not significant.

```
> set.seed(seed = 934)
> y <- c(rnorm(n = 60, mean = 100, sd = 20), rnorm(n = 10, 110, sd = 20))
> g <- factor(rep(seq(1, 7, 1), each = 10), labels = "g", ordered = FALSE)
```

If you examine the line `y <- c(rnorm(n=60, ..., sd=20))`; carefully, you will see that it consists of 60 random numbers drawn from a gaussian distribution with a mean of 100, and 10 random numbers drawn from a gaussian distribution with a mean of 110. Our grouping variable, `g`, will be used to group the `y`'s into 7 sets of 10 numbers each. Boxplots of the data are shown in Figure 1.

```
> boxplot(y~g)
```

Now we conduct a standard ANOVA.

```
> lm01 <- lm(y~g);
> anova(lm01);
```

Analysis of Variance Table

```
Response: y
      Df Sum Sq Mean Sq F value Pr(>F)
g       6   3431   571.8    1.726   0.13
Residuals 63  20873   331.3
```

Notice that the omnibus F test fails to find a significant difference among the groups. Next, we compare group 7 to the other groups. Specifically, we ask whether the mean of group 7 differs from the mean of the 6 other groups. To do this analysis, I first need to download some R commands:

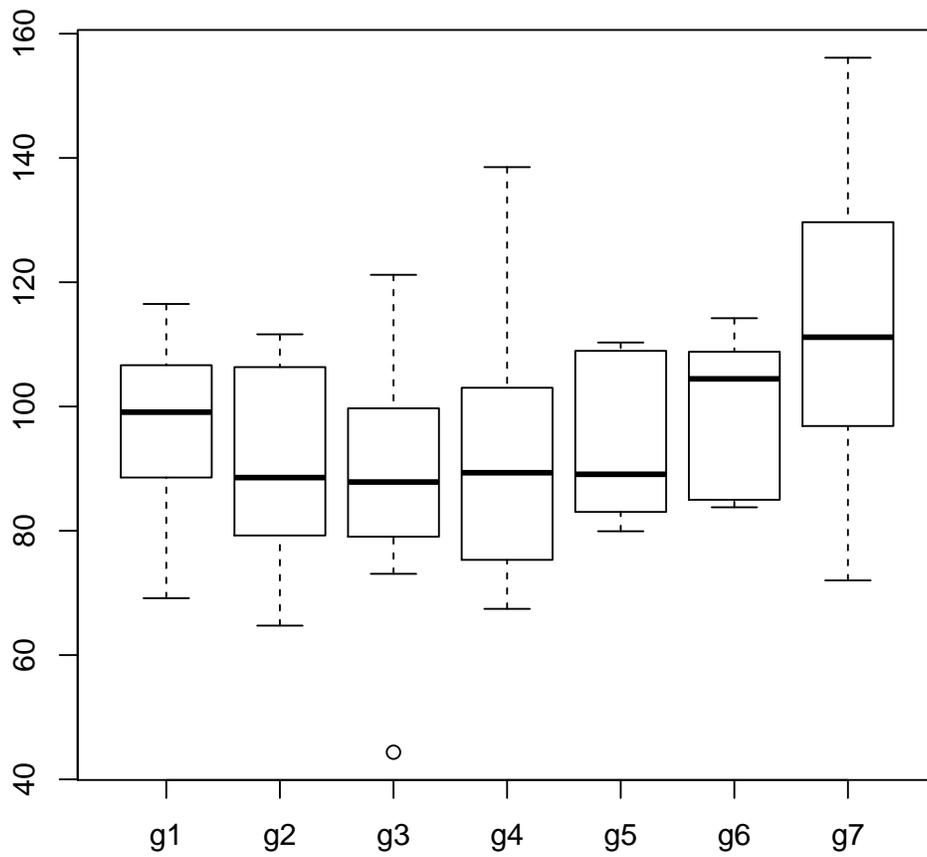


Figure 1: Boxplots of y data for different groups, g.

```
> source(url("http://psycserv.mcmaster.ca/bennett/psy710/Rscripts/linear_contrast_v2.R"))
```

```
[1] "loading function linear.comparison"
```

The `source` command loaded several commands that can be used to perform a linear comparison among means in a one-way design. Next, I have to specify my *contrast weights*:

```
> my.weights <- c(-1, -1, -1, -1, -1, -1, 6)
```

We will discuss the meaning of the contrast weights in the following section. Finally, I use the `linear.comparison` command, which was loaded into R's workspace by the previous `source` command, to perform the linear comparison:

```
> my.contrast <- linear.comparison(y, g, c.weights = my.weights)
```

```
[1] "computing linear comparisons assuming equal variances among groups"
```

```
[1] "C 1: F=7.218, t=2.687, p=0.009, psi=100.218, CI=(-2.897,203.332), adj.CI= (25.673,174.762)"
```

```
> my.contrast[[1]]$F
```

```
[1] 7.21771
```

```
> my.contrast[[1]]$p.2tailed
```

```
[1] 0.00921941
```

The F test for this *linear contrast*, or *linear comparison*, is significant, even though the omnibus F test was not.

4.1.1 using standard R commands

In this section I will show you how to perform the linear comparison with R's built-in commands (i.e., without using the `linear.comparison` command). The first step is to inform R that I want to perform a particular comparison among the various groups represented by the factor `g`:

```
> contrasts(g) <- my.weights
```

Next, I perform an ANOVA. Note that I'm using the `aov` command, rather than the `lm` command:

```
> my.aov <- aov(y~g)
```

Finally, I display the ANOVA table in a way that splits the group effect into various pieces:

```
> summary(my.aov, split=list(g=list(myContrast=1,others=2:6)))
```

The `split` syntax is complicated because it is a list of lists: It is used to inform R to split a `list` of terms in the linear model (e.g., `g`) into lists of pieces. In this case, we're telling R to split the effects related to factor `g` into two pieces: the first piece corresponds to our linear contrast, and the second piece is a combination of five `other` things that are not of interest. We will discuss what these pieces mean in the following sections. For now, let's examine the output of the `summary` command:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
g	6	3431	572	1.73	0.1296
g: myContrast	1	2391	2391	7.22	0.0092 **
g: others	5	1039	208	0.63	0.6794
Residuals	63	20873	331		

```
---
```

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

Notice that the F and p values listed in the ANOVA table are the same (to within rounding error) as those provided by `linear.comparison`. However, it is important to note that the two methods give the same results *only when there are equal n per group*. If the design is unbalanced, then `aov` yields different values for SS_{contrast} , F , and p . So you should only use `aov` to evaluate linear contrasts for balanced designs.

Finally, I want to show you how to perform a linear contrast with the `lm()` command:

```

> # list 1st column of weights (printed as 1 row):
> contrasts(g)[,1] # they are the same as before

g1 g2 g3 g4 g5 g6 g7
-1 -1 -1 -1 -1 -1 6

> my.lm <- lm(y~g) # construct the linear
> summary(my.lm) # list the coefficients and t tests

Call:
lm(formula = y ~ g)

Residuals:
    Min       1Q   Median       3Q      Max
-43.83 -11.38  -0.39   11.14  47.32

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  96.044      2.176   44.15 <2e-16 ***
g1           2.386      0.888    2.69  0.0092 **
g2          -6.153      5.756   -1.07  0.2891
g3          -3.150      5.756   -0.55  0.5861
g4           0.163      5.756    0.03  0.9775
g5           5.502      5.756    0.96  0.3428
g6          -5.084      5.756   -0.88  0.3804
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.2 on 63 degrees of freedom
Multiple R-squared:  0.141,    Adjusted R-squared:  0.0594
F-statistic: 1.73 on 6 and 63 DF,  p-value: 0.13

```

The first coefficient, g_1 , is the one that corresponds to our contrast. Notice that the t test for that coefficient yields the same p -value as our previous F test. Furthermore, the values of t and F are related: $t^2 = 2.69^2 = 7.23$ which is equal to our value of F (to within rounding error). In other words, the two-tailed t test in this summary table is equivalent to the F test in the ANOVA table listed above. Fortunately, `lm` works even with unbalanced designs (i.e., the values of t and p will be very close to those obtained with `linear.comparison`).

4.2 Complex Comparisons

The previous examples used a linear contrast to test the null hypothesis:

$$-1(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) + 6\mu_7 = 0$$

which is equivalent to the null hypotheses

$$6\mu_7 - 1(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) = 0$$

and, by multiplying both sides of the equation by $\frac{1}{6}$,

$$\mu_7 - \frac{1}{6}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) = 0$$

which is equivalent (finally), to

$$\mu_7 = \frac{1}{6}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) \tag{1}$$

This last form of the hypothesis is perhaps the easiest to interpret: the null hypothesis is that the mean of group 7 does not differ from the mean of the other group means. We rejected that null hypothesis to accept the alternative

$$\mu_7 \neq \frac{1}{6}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6)$$

The general form of a linear contrast is

$$H_0 : c_1\mu_1 + c_2\mu_2 + \dots + c_a\mu_a = \Psi = 0 \quad (2)$$

where the weights of the linear contrast are constrained to sum to zero

$$\sum_{j=1}^a c_j = 0 \quad (3)$$

and the estimation of the comparison, $\hat{\Psi}$, is

$$\hat{\Psi} = \sum_{j=1}^a (c_j \bar{Y}_j) \quad (4)$$

In the example from the preceding section, $\hat{\Psi} = 100.2$.

```
> my.contrast[[1]]$psi
```

```
[1] 100.218
```

If there were four groups, and we wanted to compare group 3 to the mean of groups 1 and 2, then we would use the following contrast

$$\Psi = (1)\mu_1 + (1)\mu_2 - 2\mu_3 + (0)\mu_4$$

and our coefficients would be $c=(1, 1, -2, 0)$. If we wanted to compare group 1 to group 4, the coefficients would be $c=(1, 0, 0, -1)$. If we wanted to compare groups 1 and 2 to groups 3 and 4, the coefficients would be $c=(1, 1, -1, -1)$.

A linear contrast is evaluated using our standard F test, which in this case is equal to

$$F = \frac{(\Psi^2) / \sum_{j=1}^a (c_j^2 / n_j)}{MS_W} \quad (5)$$

when n is equal across groups, Equation 5 simplifies to

$$F = \frac{(n\Psi^2) / \sum_{j=1}^a (c_j^2)}{MS_W} \quad (6)$$

The numerators in Equations 5 and 6 are $SS_{contrast}$. **There is only 1 degree of freedom for each contrast**, so $SS_{contrast} = MS_{contrast}$. The degrees of freedom for the denominator is $N - a$, where a is the number of groups or, equivalently, the number of parameters in the full model (i.e., an intercept and $a - 1$ α 's). Most of these values are contained in the variable returned by `linear.comparison`:

```
> my.contrast[[1]]$F
```

```
[1] 7.21771
```

```
> my.contrast[[1]]$SS.contrast
[1] 2391.33
> my.contrast[[1]]$df1
[1] 1
> my.contrast[[1]]$df2
[1] 63
```

We can use these quantities to calculate $MS_W = SS_{contrast}/F = 331.3$. The p-value was 0.00922, which is less than $\alpha = .05$, so we reject the *two-tailed* null hypothesis described by Equation 1.

Examination of Equation 5 shows that Ψ is divided by a weighted sum of contrast weights. This division means that the values of $SS_{contrast}$, F , and p are not altered by multiplying the contrast weights by a constant. In other words, contrast weights of (for example) $c=(1, 1, 1, -3)$, $c=(2, 2, 2, -6)$, and $c=(1/3, 1/3, 1/3, -1)$ will all give the same values of $SS_{contrast}$, F , and p (although the value of Ψ will differ; see Equation 4).

4.2.1 one-tailed tests

How can we evaluate the following hypotheses?

$$H_0 : \mu_7 \leq \frac{1}{6}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) \quad (7)$$

$$H_1 : \mu_7 > \frac{1}{6}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) \quad (8)$$

A one-tailed test is appropriate here. One-tailed tests can be done easily with t tests, and therefore it would be useful to convert our F statistic (Equation 5) into a t statistic. In fact, it can be shown that

$$t = \frac{\Psi / \sqrt{\sum_{j=1}^a (c_j^2 / n_j)}}{\sqrt{MS_W}} \quad (9)$$

follows a t distribution with $N - a$ degrees of freedom. A comparison of Equations 5 and 9 will show that $F = t^2$, and therefore $t = \pm\sqrt{F}$. In other words, it is possible to recover the magnitude of t simply by taking the square-root of F . The sign of t can be recovered by noting the sign of $\hat{\Psi}$: if $\hat{\Psi} < 0$, then $t = -\sqrt{F}$, but if $\hat{\Psi} > 0$ then $t = \sqrt{F}$. In our example, $\hat{\Psi} = 100.2$, and therefore $t = \sqrt{7.2177} = 2.68$.

Finally, we can evaluate the null hypothesis. Our null hypothesis predicts that $\Psi \leq 0$. Therefore, when the null hypothesis is true, the largest possible value of Ψ in the population is zero, and observed values of Ψ (i.e., $\hat{\Psi}$) that are greater than zero are due only to sampling variation. This is a long-winded way of saying that large, positive values of $\hat{\Psi}$ are unlikely if the null hypothesis is true. These statements about $\hat{\Psi}$ can be extended to t (Eq. 9): large, positive values of t are unlikely if the null hypothesis is true. We can calculate the probability of obtaining a value of t that is equal to or greater than our observed value of $t = 2.68$ given the assumption that $\Psi = 0$:

```
> pt(2.68, df=63, lower.tail=FALSE)
[1] 0.00469164
```

Note that the degrees of freedom (df) correspond to the value in the denominator in Equation 5 and so were set to $N - a = 63$. Setting `lower.tail` to `FALSE` tells R that we want to know the probability of getting a t value that is *larger* than 2.68; setting `lower.tail` to `TRUE` (the default value) would return the probability of getting a value that is less than 2.68. The probability $p = 0.0046$ represents the probability of getting our result, or something more extreme, if the hypothesis that $\Psi = 0$ was true. Suppose, instead, we had assumed that Ψ was some value less than zero? In that case, the probability of getting a value of t that was equal to or greater than our observed value would be even lower. Therefore, the value $p = 0.0046$ represents the *highest* probability of obtaining our result (or something more extreme) given that the null hypothesis $\Psi \leq 0$ is true. Adopting a standard Type I error rate (i.e., $\alpha = .01$ or $.05$) would lead us to reject the null hypothesis in favour of the alternative, $\Psi > 0$.

4.3 Unequal Variances

The denominator in Equation 5 is MS_W , which is an estimate of σ_e^2 that is derived from *all* of the groups, not just the ones being compared. When the variance is homogeneous across groups, MS_W is a better estimator of σ_e^2 than an estimate obtained from a subset of the groups.

What if the variance is not homogeneous? It turns out that F tests of linear contrasts are *not* robust to violations of the homogeneity of variance assumptions. Moreover, it is not the case that violations of the assumption have the effect of simply lowering α : sometimes α will be increased and other times it will be decreased. When the groups have unequal variances, therefore, we use a different method for estimating σ_e^2

$$\hat{\sigma}_e^2 = \frac{\sum_{j=1}^a w_j s_j^2}{\sum_{j=1}^a w_j} \quad (10)$$

where

$$w_j = c_j^2/n_j \quad (11)$$

Using this new estimate of σ_e^2 , the formula for F becomes

$$F = \frac{(\Psi^2)/\sum_{j=1}^a (c_j/n_j)}{\sum_{j=1}^a [(c_j^2/n_j)s_j^2] / \sum_{j=1}^a (c_j^2/n_j)} \quad (12)$$

The numerator degrees of freedom remains 1, but the denominator degrees of freedom no longer is $N - a$. Instead, it is given by the formula

$$df = \frac{\left[\sum_{j=1}^a (c_j^2 s_j^2 / n_j) \right]^2}{\sum_{j=1}^a [(c_j^2 s_j^2 / n_j)^2 / (n_j - 1)]} \quad (13)$$

Fortunately, we rarely have to use Equations 12 and 13 to calculate F and df . Instead, we can use `linear.comparison()` to do the heavy lifting:

```
> source(url("http://psycserv.mcmaster.ca/bennett/psy710/Rscripts/linear_contrast_v2.R"))
[1] "loading function linear.comparison"
> contrast.unequal.var <- linear.comparison(y, g, c.weights = list(c(-1, -1, -1, -1, -1, -1, 6)), var.equal = FALSE)
[1] "computing linear comparisons assuming unequal variances among groups"
[1] "C 1: F=3.772, t=1.942, p=0.080, psi=100.218, CI=(-14.451,214.886), adj.CI= (-14.451,214.886)"
> contrast.unequal.var[[1]]$psi
```

```

[1] 100.218
> contrast.unequal.var[[1]]$F
[1] 3.77214
> contrast.unequal.var[[1]]$df2
[1] 10.1991
> contrast.unequal.var[[1]]$p.2tailed
[1] 0.0802195
> my.contrast[[1]]$psi
[1] 100.218
> my.contrast[[1]]$F
[1] 7.21771
> my.contrast[[1]]$p.2tailed
[1] 0.00921941

```

Notice that the variable `var.equal` has been set to `FALSE`. Also notice that $\hat{\Psi} = 100.2$, which is the same value we obtained when we assumed equal variances. However, the estimate of σ_e^2 is different from the previous value, which reduces F from 7.218 to 3.772, and the degrees of freedom in the denominator have been reduced from 63 to 10.2, and therefore the contrast is no longer significant ($p = 0.08$). Hence, when we assume unequal variances across groups, we do not reject the null hypothesis that $\mu_7 - \frac{1}{6}(\mu_1 + \mu_2 + \dots + \mu_6) = 0$.

Of course, we can test the assumption of constant variance:

```

> bartlett.test(y~g)

    Bartlett test of homogeneity of variances

data:  y by g
Bartlett's K-squared = 9.6411, df = 6, p-value = 0.1406

```

The Bartlett test is not significant, even if we set $\alpha = 0.1$, and therefore the assumption of homogeneity of variance appears reasonable for these data. Therefore, we can justify using the more powerful linear comparison that assumes constant variance.

4.4 Example #1

This example is question #3 in the Exercises for Chapter 4.

A psychologist collected data from three groups. The sample means are: $\bar{Y}_1 = 12$, $\bar{Y}_2 = 10$, $\bar{Y}_3 = 6$. $MS_W = 25$, and there are 10 subjects per group. The psychologist wants to compare the average of the means of Groups 1 and 2 to the mean of Group 3. We will use contrast coefficients $c=(1, 1, -2)$ to evaluate this comparison. The null and alternative hypotheses are

$$H_0 : (1)\mu_1 + (1)\mu_2 - (2)\mu_3 = 0 \quad (14)$$

$$H_1 : (1)\mu_1 + (1)\mu_2 - (2)\mu_3 \neq 0 \quad (15)$$

$$(16)$$

Because we only have the group means, and not the raw scores, we need to use Equation 4 to calculate Ψ :

$$\hat{\Psi} = \sum_{j=1}^a c_j \bar{Y}_j = (1)12 + (1)10 - (2)6 = 12 + 10 - 12 = 10$$

Next, we use Equation 5 to calculate F :

$$F = \frac{10^2 / (1/10 + 1/10 + 4/10)}{25} = \frac{100 / (6/10)}{25} = \frac{166.67}{25} = 6.67$$

with numerator and denominator degrees of freedom of 1 and $30 - 3 = 27$, respectively. The probability of obtaining this F , or one more extreme, if the null hypothesis is true, is

```
> pf(6.67, 1, 27, lower.tail=FALSE)
```

```
[1] 0.0155463
```

```
> 1 - pf(6.67, 1, 27)
```

```
[1] 0.0155463
```

which is less than $\alpha = .05$, so we reject the null hypothesis in favor of the alternative. Note, by the way, that the two forms of the `pf` command give the same result.

4.5 Example #2

We will now compute contrasts on the set of (fictitious) blood pressure data listed in Table 4.1 in Maxwell and Delaney. The data, shown in Table 1, are from four groups of subjects corresponding to four treatments for hypertension: drug therapy (`aDrug`), biofeedback (`bFeedback`), dietary modification (`cDiet`), and a treatment that combines all aspects of the other treatments (`dCombo`). The dependent variable is blood pressure after treatment and is stored in `bp$blood`.

```
> bp <- read.csv(url("http://psycserv.mcmaster.ca/bennett/psy710/datasets/maxwell_tab41.csv"))
> names(bp)
```

```
[1] "blood" "group"
```

```
> levels(bp$group)
```

```
[1] "aDrug"      "bFeedback"  "cDiet"      "dCombo"
```

We are going to do three contrasts. First, we want to compare `dCombo` to the average of the other groups, so we will use a contrast `c=(1/3, 1/3, 1/3, -1)`. Next, we will compare `cDiet` to the average of `aDrug` and `bFeedback` with a contrast `c=(1/2, 1/2, -1, 0)`. Finally, we will compare `aDrug` to `bFeedback` with the contrast `c=(1, -1, 0, 0)`. (Note that we must set the contrast for *all* groups, even ones that are not included in the contrast.) We first construct a `list` that has all three contrasts.

```
> blood.contrasts <- list(c(1/3, 1/3, 1/3, -1), c(1/2, 1/2, -1, 0), c(1, -1, 0, 0))
```

And then we pass the `bp$blood`, `bp$group`, and `blood.contrasts` to the function `linear.comparison`:

```
> bp.results <- linear.comparison(bp$blood, bp$group, blood.contrasts, var.equal = TRUE)
```

	blood	group
1	84.00	aDrug
2	95.00	aDrug
3	93.00	aDrug
4	104.00	aDrug
5	81.00	bFeedback
6	84.00	bFeedback
7	92.00	bFeedback
8	101.00	bFeedback
9	80.00	bFeedback
10	108.00	bFeedback
11	98.00	cDiet
12	95.00	cDiet
13	86.00	cDiet
14	87.00	cDiet
15	94.00	cDiet
16	91.00	dCombo
17	78.00	dCombo
18	85.00	dCombo
19	80.00	dCombo
20	81.00	dCombo

Table 1: Blood Pressure Data (Table 4.1)

```
[1] "computing linear comparisons assuming equal variances among groups"
[1] "C 1: F=4.815, t=2.194, p=0.043, psi=9.333, CI=(2.554,16.112), adj.CI= (-2.036,20.703)"
[1] "C 2: F=0.012, t=0.110, p=0.913, psi=0.500, CI=(-7.767,8.767), adj.CI= (-11.601,12.601)"
[1] "C 3: F=0.321, t=0.566, p=0.579, psi=3.000, CI=(-10.210,16.210), adj.CI= (-11.163,17.163)"

> bp.results.unequal <- linear.comparison(bp$blood,bp$group, blood.contrasts, var.equal = FALSE)

[1] "computing linear comparisons assuming unequal variances among groups"
[1] "C 1: F=8.519, t=2.919, p=0.014, psi=9.333, CI=(2.298,16.369), adj.CI= (0.320,18.346)"
[1] "C 2: F=0.016, t=0.128, p=0.900, psi=0.500, CI=(-8.009,9.009), adj.CI= (-10.361,11.361)"
[1] "C 3: F=0.232, t=0.481, p=0.643, psi=3.000, CI=(-11.407,17.407), adj.CI= (-15.862,21.862)"
```

We can list the results for comparison n using the syntax `bp.results[[n]]` (note the double brackets!). The results for each comparison are quite long, so I'll show only one complete record:

```
> bp.results[[1]]

$contrast
[1] 0.333333 0.333333 0.333333 -1.000000

$F
[1] 4.81505

$t
[1] 2.19432

$df1
[1] 1
```

```

$df2
[1] 16

$p.2tailed
[1] 0.0433206

$psi
[1] 9.33333

$confinterval
[1] 2.5543 16.1124

$adj.confint
[1] -2.03615 20.70282

$alpha
[1] 0.05

$SS.contrast
[1] 324.414

$d.effect.size
[1] 1.13707

$R2.alerting
[1] 0.969702

$R2.effect.size
[1] 0.229665

$R2.contrast
[1] 0.231325

```

The contrast coefficients are listed, as are the values of F , $t = \sqrt{F}$, p and $SS_{contrast}$. Also listed are the value of $\hat{\Psi}$ and the $100(1 - \alpha)\%$ confidence interval for Ψ . The last four values are measures of effect size and association strength, which we'll discuss later.

The F and p values for all three contrasts are

```

> for (kk in 1:3) {print(sprintf("Contrast %2i, F = %4.3f, p = %4.3f",kk,
+ bp.results[[kk]]$F, bp.results[[kk]]$p.2tailed))}

[1] "Contrast  1, F = 4.815, p = 0.043"
[1] "Contrast  2, F = 0.012, p = 0.913"
[1] "Contrast  3, F = 0.321, p = 0.579"

```

The results from the contrasts assuming unequal group variances are

```

> for (kk in 1:3) {print(sprintf("Contrast %2i, F = %4.3f, p = %4.3f",kk,
+ bp.results.unequal[[kk]]$F, bp.results.unequal[[kk]]$p.2tailed))}

[1] "Contrast  1, F = 8.519, p = 0.014"
[1] "Contrast  2, F = 0.016, p = 0.900"
[1] "Contrast  3, F = 0.232, p = 0.643"

```

The contrast between dCombo and the other groups is significant, but the other contrasts are not. The value of the contrast is

```

> bp.results[[1]]$psi

```

```
[1] 9.33333
```

```
> bp.results.unequal[[1]]$psi
```

```
[1] 9.33333
```

In other words, using the contrast that was calculated assuming equal group variances,

$$\hat{\Psi} = \frac{1}{3}(\mu_1 + \mu_2 + \mu_3) - \mu_4 = 9.333$$

The 95% confidence interval for the comparison (i.e., the difference between the mean of group 4 and the mean of the first three groups) is

```
> bp.results[[1]]$confinterval
```

```
[1] 2.5543 16.1124
```

```
> bp.results.unequal[[1]]$confinterval
```

```
[1] 2.29767 16.36900
```

4.5.1 further calculations

You should repeat the analyses after multiplying the contrast weights by a scalar. For example, try these coefficients:

```
> blood.contrasts <- list(c(1/3, 1/3, 1/3, -1)* 3, c(1/2, 1/2, -1, 0) * 2, c(1, -1, 0, 0))
```

How do the results change? How do they stay the same?

4.5.2 using lm

In this section we will perform the linear comparisons using the `lm()` function. First, we need to create our contrast weights and then link them to your grouping variable:

```
> c1 <- c(1/3,1/3,1/3,-1); # 1st contrast
> c2 <- c(1/2,1/2,-1,0); # 2nd contrast
> c3 <- c(1,-1,0,0); # 3rd contrast
> cbind(c1,c2,c3) # cbind joins the 3 contrasts as 3 columns

      c1  c2 c3
[1,] 0.333333 0.5 1
[2,] 0.333333 0.5 -1
[3,] 0.333333 -1.0 0
[4,] -1.000000 0.0 0

> contrasts(bp$group) <- cbind(c1,c2,c3) # link contrasts to group
> contrasts(bp$group)
```

```
      c1  c2 c3
aDrug  0.333333 0.5 1
bFeedback 0.333333 0.5 -1
cDiet  0.333333 -1.0 0
dCombo -1.000000 0.0 0
```

Next we construct the linear models and list the coefficients and t tests:

```
> bp.lm.01 <- lm(blood~group,data=bp)
> summary(bp.lm.01)
```

Call:

```
lm(formula = blood ~ group, data = bp)
```

Residuals:

```
    Min       1Q   Median       3Q      Max
-11.00  -5.25   0.00   3.75  17.00
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   90.000      1.854   48.53  <2e-16 ***
groupc1         7.000      3.190    2.19   0.043 *
groupc2         0.333      3.018    0.11   0.913
groupc3         1.500      2.649    0.57   0.579
---
```

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

Residual standard error: 8.21 on 16 degrees of freedom

Multiple R-squared: 0.237, Adjusted R-squared: 0.0937

F-statistic: 1.66 on 3 and 16 DF, p-value: 0.216

Notice that the p -values for our t tests are the same as the p -values for our F tests (assuming equal group variances), and that $F = t^2$ for each contrast. The first comparison is significant, but the other two are not.

Now suppose we were performing a directional test. Specifically, let's suppose that we wanted to know if average blood pressure in the first three groups was greater than blood pressure in the fourth group (dCombo). Given the way we set up our weights – i.e., positive weights were assigned to the first three groups and a negative weight was assigned to the fourth group – then our hypothesis predicts the value of our contrast should be greater than zero, which implies that the value of t for this contrast also should be greater than zero. The null hypothesis for this test is that $\psi \leq 0$, and the alternative hypothesis is that $\psi > 0$. So the first thing we should do is determine if our value of t is less than zero; if it is, then we would fail to reject the null hypothesis and our analysis would halt. However, inspection of the summary table listed above shows that $t = 2.19$ for the first contrast, which is greater than zero, and so our analysis must next determine if it is *significantly* greater than zero. The p -value listed in the table is for a two-tailed test. To get the p -value for a one-tailed test we need to divide it by two: $p_{one-tailed} = 0.043/2 = 0.0215$. Therefore, we would reject the null hypothesis $\psi \leq 0$ in favor of the alternative hypothesis that $\psi > 0$. In a journal paper you would report this result thusly: the contrast was significant ($t(16) = 2.19$, $p = 0.0215$ one-tailed). Notice that the degrees of freedom for the t test are listed in the line that reports the residual standard error, near the bottom of the summary.

Finally, you should consider how our analysis would be affected if the weights for the first linear contrast were multiplied by -1 to yield $[-1/3, -1/3, -1/3, 1]$. It should be obvious that reversing the sign of the weights will reverse the sign of ψ and, consequently, the sign of t in the summary table, yielding $t = -2.19$. Because the p -value in the table is for a two-tailed test, it will remain the same:

```
> c1 <- c(-1/3,-1/3,-1/3,1); # reverse the sign of the 1st contrast weights
> contrasts(bp$group) <- cbind(c1,c2,c3) # link contrasts to group
> bp.lm.02 <- lm(blood~group,data=bp) # new anova
> summary(bp.lm.02) # list coefficients and p-values
```

Call:

```
lm(formula = blood ~ group, data = bp)
```

Residuals:

```
   Min       1Q   Median       3Q      Max
-11.00  -5.25   0.00   3.75  17.00
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   90.000      1.854   48.53  <2e-16 ***
groupc1       -7.000      3.190   -2.19   0.043 *
groupc2        0.333      3.018    0.11   0.913
groupc3        1.500      2.649    0.57   0.579
```

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

Residual standard error: 8.21 on 16 degrees of freedom

Multiple R-squared: 0.237, Adjusted R-squared: 0.0937

F-statistic: 1.66 on 3 and 16 DF, p-value: 0.216

Does this mean that we now fail to reject the null hypothesis, and that our contrast is no longer significant? The answer is, no. To see why, we need to re-construct our null and alternative hypothesis. We predicted that the means of the first three groups will be greater than the mean of the fourth group. If our prediction is correct, then – *given the values of the weights used to construct our linear contrast* – we would expect $\psi < 0$. Therefore, the null hypothesis for this set of weights is $\psi \geq 0$, and the alternative is $\psi < 0$. Based on the results of our t test, we would reject the null hypothesis ($t(16) = -2.19$, $p = 0.0215$, one-tailed). The important point here is that the null and alternative hypotheses for directional tests of linear contrasts are affected by the sign of the contrast weights.

4.6 Measures of Effect Size & Association Strength

A common measure of the standardized difference between two means is Cohen's d

$$d = (\mu_1 - \mu_2) / \sigma_e$$

which can be estimated from the data by

$$\hat{d} = (\bar{Y}_1 - \bar{Y}_2) / \sqrt{MS_W} \quad (17)$$

In the case of two or more groups, we can define a standardized difference *for a contrast* among population means as

$$d = 2\Psi / \left(\sigma_e \left[\sum_{j=1}^a |c_j| \right] \right)$$

We can estimate d from data by

$$d = 2\hat{\Psi} / \left(\sqrt{\text{MS}_W} \left[\sum_{j=1}^a |c_j| \right] \right) \quad (18)$$

When the comparison is between two group means (e.g., $c=(-1,1,0,0,\dots)$), Equations 17 and 18 are the same.

Your textbook describes several measures of association strength that are based on $\text{SS}_{contrast}$. The first is $R_{alerting}^2$ which is defined as

$$R_{alerting}^2 = \text{SS}_{contrast} / \text{SS}_B \quad (19)$$

Recall that SS_B is a measure of the variation among group means:

$$\text{SS}_B = \sum_{j=1}^a [n_j (\bar{Y}_j - \bar{Y}_u)^2]$$

So, Equation 19 expresses $\text{SS}_{contrast}$ as a proportion of the “total variation” among groups. $R_{alerting}^2$ can vary between 0 and 1. As your book notes, another interpretation of $R_{alerting}^2$ is that it is the *squared* correlation between the comparison coefficients and the group means when the group n 's are equal.

Another measure of association strength is $R_{effectsize}^2$:

$$R_{effectsize}^2 = \text{SS}_{contrast} / \text{SS}_{Total} \quad (20)$$

Comparing Equations 19 and 20 shows that they have the same numerators but different denominators. $R_{effectsize}^2$ describes the between-group variability associated with the contrast relative to the total variability among the dependent measures.

The third and final measure of association strength does not depend on SS_{Total} or SS_B :

$$R_{contrast}^2 = \text{SS}_{contrast} / (\text{SS}_{contrast} + \text{SS}_W) \quad (21)$$

Your textbook describes one advantage that $R_{contrast}^2$ has over the other two measures of association strength. Suppose you are conducting a linear contrast that compares groups B and C to each other, and ignores group A . It turns out that the values of $R_{alerting}^2$ and $R_{effectsize}^2$ depend on the mean of group A even though that group is ignored in our contrast. The value of $R_{contrast}^2$ is less dependent on the value of the ignored group A .

None of these measures is necessarily better than the other two; they simply convey different information. You should note that it is possible for $R_{alerting}^2$ to be close to its maximum value (i.e., ≈ 1) but $R_{effectsize}^2$ to be very small. (How could this happen?)

4.7 Orthogonal Contrasts

In this section we introduce the notion of *orthogonal* contrasts. When sample sizes are equal, two contrasts are said to be orthogonal if and only if

$$\sum_{j=1}^a (c_{1j}c_{2j}) = 0 \quad (22)$$

where c_{1j} and c_{2j} are the coefficients, or weights, of the two contrasts. When sample sizes are unequal, contrasts are orthogonal if and only if

$$\sum_{j=1}^a (c_{1j}c_{2j}/n_j) = 0 \quad (23)$$

So, when sample sizes are equal, the pair of contrasts $c_1 = (1, 1, 1, -3)$ and $c_2 = (1, 1, -2, 0)$ are orthogonal because

$$(1)(1) + (1)(1) + (1)(-2) + (-3)(0) = 0$$

However, $c_1 = (1, 1, 1, -3)$ and $c_3 = (1, 1, 0, -3)$ are not orthogonal because

$$(1)(1) + (1)(1) + (1)(0) + (-3)(-3) \neq 0$$

A *set* of contrasts is said to be mutually orthogonal if and only if every possible pair of contrasts is orthogonal. For instance the set of contrasts

$$\begin{aligned} \Psi_1 &= (1)\mu_1 + (1)\mu_2 + (-1)\mu_3 + (-1)\mu_4 \\ \Psi_2 &= (1)\mu_1 + (-1)\mu_2 + (0)\mu_3 + (0)\mu_4 \\ \Psi_3 &= (0)\mu_1 + (0)\mu_2 + (1)\mu_3 + (-1)\mu_4 \end{aligned}$$

is (mutually) orthogonal because each pair of contrasts is orthogonal.

The concept of orthogonality is important for the following reason: multiple orthogonal contrasts of a set of group means provide independent pieces of information about the way the means differ. In other words, orthogonal contrasts provide separate and unique information about group differences. Multiple contrasts that are *not* orthogonal, on the other hand, provide overlapping, correlated, and partially-redundant information about group differences. This special characteristic of orthogonal contrasts is represented in the Equation 24, which is true *only* for orthogonal sets of contrasts:

$$\sum_{j=1}^{a-1} SS_{contrast,j} = SS_B \quad (24)$$

Notice that the summation in Equation 24 is over $a - 1$ contrasts. This leads to the second important property of orthogonal contrasts: If there are a groups, then the largest possible set of orthogonal contrasts will have $a - 1$ elements. I will refer to a set of $a - 1$ orthogonal contrast as constituting a *complete* orthogonal set. Note that I am *not* suggesting that there are only $a - 1$ orthogonal contrasts. In fact, there are an infinite number of orthogonal contrasts, and there are an infinite number of orthogonal *sets* of contrasts. However, each complete orthogonal set will contain $a - 1$ elements.

These ideas about orthogonality, and the decomposition of SS_B into independent pieces, will be important when we discuss multiple comparisons.