

## Comparing several means

### Self-test answers



- To illustrate exactly what is going on I have created a file called **dummy.dat**. This file contains the Viagra data but with two additional variables (**dummy1** and **dummy2**) that specify to which group a data point belongs (as in Table 10.2). Access this file and run multiple regression analysis using **libido** as the outcome and **dummy1** and **dummy2** as the predictors.

To read in the data, set your working directory to the location of the file and execute:

```
dummyData<-read.delim("Dummy.dat", header = TRUE)
```

We need to turn the **dose** variable into a factor because it is categorical. It is coded as Placebo = 1, Low Dose = 2, and High Dose = 3. Therefore, we can change **dose** to a factor as follows:

```
dummyData$dose<-factor(dummyData$dose, levels = c(1:3), labels = c("Placebo", "Low Dose", "High Dose"))
```

To create the regression model (*dummy.model*) and to print the summary of it, we execute:

```
dummy.model<-lm(libido~dummy1 + dummy2, data = dummyData)
summary(dummy.model)
```



- To illustrate these principles, I have created a file called **Contrast.dat** in which the Viagra data are coded using the contrast coding scheme used in this section. Run multiple regression analyses on these data using **libido** as the outcome and using **dummy1** and **dummy2** as the predictor variables (leave all default options).

To read in the data, set your working directory to the location of the file and execute:

```
contrastData<-read.delim("Contrast.dat", header = TRUE)
```

We need to turn the **dose** variable into a factor because it is categorical. It is coded as Placebo = 1, Low Dose = 2, and High Dose = 3. Therefore, we can change **dose** to a factor as follows:

```
contrastData$dose<-factor(contrastData$dose, levels = c(1:3), labels = c("Placebo", "Low Dose", "High Dose"))
```

To create the regression model (*contrast.model*) and to print the summary of it, we execute:

```
contrast.model<-lm(libido~dummy1 + dummy2, data = contrastData)
summary(contrast.model)
```



- Use *ggplot2* to produce a line chart with error bars showing bootstrapped confidence intervals for the Viagra data.

The image in the book was generated from this code:

```
line <- ggplot(viagraData, aes(dose, libido))

line + stat_summary(fun.y = mean, geom = "line", size = 1, aes(group=1), colour =
"#FF6633") + stat_summary(fun.data = mean_cl_boot, geom = "errorbar", width = 0.2,
size = 0.75, colour = "#990000") + stat_summary(fun.y = mean, geom = "point", size =
4, colour = "#990000") + stat_summary(fun.y = mean, geom = "point", size = 3, colour =
"#FF6633") + labs(x = "Dose of Viagra", y = "Mean Libido")
```



- Repeat the analysis with 10% trimmed means. How do your conclusions differ?

To use 10% trimmed means we would change the command to be (note *tr* = .1, which gives us 10% trim rather than the default of *tr* = .2, which trims 20%):

```
lincon(viagraWide, tr = .1)
```

The output is:

```
[1] "Note: confidence intervals are adjusted to control FWE"
[1] "But p-values are not adjusted to control FWE"
$test
  Group Group      test      crit      se      df
[1,]    1    2 1.212678 2.960000 0.8246211 8.000000
[2,]    1    3 3.055050 2.985397 0.9165151 7.719912
[3,]    2    3 1.963961 2.985397 0.9165151 7.719912

$psihat
  Group Group psihat ci.lower ci.upper p.value
[1,]    1    2   -1.0 -3.440879 1.44087853 0.25984505
[2,]    1    3   -2.8 -5.536161 -0.06383861 0.01638290
[3,]    2    3   -1.8 -4.536161 0.93616139 0.08644456
```

The confidence interval for the difference between groups 1 and 3 (i.e. placebo and high dose) no longer crosses zero, which implies a significant difference in libido between these two groups. This mirrors the results from the non-robust *post hoc* tests in the chapter but is different from when we used a 20% trim on the data.

## Oliver Twisted

### Please Sir, can I have some more ... Levene's test?



Levene's test is basically an ANOVA conducted on the absolute differences between the observed data and the median or mean from which the data came. To see what I mean, let's do a sort of manual Levene's test on the Viagra data. First, create the **dose** and **libido** variables and put them in a dataframe:

```
libido<-c(3,2,1,1,4,5,2,4,2,3,7,4,5,3,6)

dose<-c(rep(1,5),rep(2,5), rep(3,5))

dose<-factor(dose, levels = c(1:3), labels = c("Placebo", "Low Dose", "High Dose"))

viagraData<-data.frame(libido, dose)
```

The data look like this:

	dose	libido
1	Placebo	3
2	Placebo	2
3	Placebo	1
4	Placebo	1
5	Placebo	4

```

6   Low Dose      5
7   Low Dose      2
8   Low Dose      4
9   Low Dose      2
10  Low Dose      3
11  High Dose     7
12  High Dose     4
13  High Dose     5
14  High Dose     3
15  High Dose     6

```

We now want to add a variable containing the medians for each group. We can get the medians from:

```
by(viagraData$libido, viagraData$dose, summary)
```

```

viagraData$dose: Placebo
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.0    1.0    2.0    2.2    3.0    4.0
-----
viagraData$dose: Low Dose
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.0    2.0    3.0    3.2    4.0    5.0
-----
viagraData$dose: High Dose
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3      4      5      5      6      7

```

The medians for the placebo, low-dose and high-dose groups were 2, 3 and 5, respectively; we can add this variable to the dataframe by executing:

```
viagraData$median<-c(rep(2,5),rep(3,5), rep(5,5))
```

The data now look like this:

```

      dose libido median
1   Placebo     3      2
2   Placebo     2      2
3   Placebo     1      2
4   Placebo     1      2
5   Placebo     4      2
6   Low Dose     5      3
7   Low Dose     2      3
8   Low Dose     4      3
9   Low Dose     2      3
10  Low Dose     3      3
11  High Dose     7      5
12  High Dose     4      5
13  High Dose     5      5
14  High Dose     3      5
15  High Dose     6      5

```

We need to create a new variable called **difference** (short for ‘difference from group median’), which is each score subtracted from the median of the group to which that score belongs. In other words, we simply subtract the variable **libido** from the variable **median**.

```
viagraData$difference = viagraData$libido-viagraData$median
```

The data now look like this:

```

      dose libido median difference
1   Placebo     3      2          1
2   Placebo     2      2          0
3   Placebo     1      2         -1
4   Placebo     1      2         -1
5   Placebo     4      2          2
6   Low Dose     5      3          2
7   Low Dose     2      3         -1
8   Low Dose     4      3          1
9   Low Dose     2      3         -1
10  Low Dose     3      3          0
11  High Dose     7      5          2
12  High Dose     4      5         -1
13  High Dose     5      5          0
14  High Dose     3      5         -2
15  High Dose     6      5          1

```

Note that for person 1, the difference score is  $3 - 2 = 1$ , for person 2 it is  $2 - 2 = 0$ . As we move into the low-dose group we subtract the median of that group, so person 6’s difference is  $5 - 3 = 2$ , person

7 is  $2 - 3 = -1$ . In the high-dose group, the group median is 5, so for person 11 we get a difference of  $7 - 5 = 2$  and so on.

Think about what these differences are: they are deviations from the median. They represent variation from the median. When we compute the variance we square the values to get rid of the plus and minus signs (otherwise the positive and negative deviations will cancel out). Levene's test doesn't do this (because we don't want to change the units of measurement by squaring the values), but instead simply takes the absolute values; that is, it pretends that all of the deviations are positive.

We can get the absolute values of these differences (i.e. we need to make them all positive values), by creating a new variable (called *abs.diff*) and applying the *abs()* function to the **difference** variable:

```
viagraData$abs.diff = abs(viagraData$difference)
```

The data now look like this:

	dose	libido	median	difference	abs.diff
1	Placebo	3	2	1	1
2	Placebo	2	2	0	0
3	Placebo	1	2	-1	1
4	Placebo	1	2	-1	1
5	Placebo	4	2	2	2
6	Low Dose	5	3	2	2
7	Low Dose	2	3	-1	1
8	Low Dose	4	3	1	1
9	Low Dose	2	3	-1	1
10	Low Dose	3	3	0	0
11	High Dose	7	5	2	2
12	High Dose	4	5	-1	1
13	High Dose	5	5	0	0
14	High Dose	3	5	-2	2
15	High Dose	6	5	1	1

Note that the **abs.diff** scores are the same magnitude as the scores in **difference**, it's just that the minus signs have gone. These values still represent deviations from the mean, or variance, we just now don't have the problem of positive and negative deviations cancelling each other out.

Now, using what you learnt in the book, conduct a one-way ANOVA on these difference scores: **dose** is the independent variable and **diff** is the dependent variable. Call the resulting model *levenes*.

Your command should look like this:

```
levenes<-aov(abs.diff~dose, data = viagraData)
```

Get the summary data by executing:

```
summary(levenes)
```

You'll find that the *F*-ratio for this analysis is 0.118, which is very non-significant at  $p = 0.89$ ; that is, the same values as Levene's test in the book!

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dose	2	0.1333	0.06667	0.1176	0.89
Residuals	12	6.8000	0.56667		

Levene's test is, therefore, testing whether the 'average' absolute deviation from the median is the same in the three groups. Clever, eh?

Programs like SPSS use the deviation from the mean rather than the median (which is a less good measure). The group means were 2.2, 3.2 and 5. If you want to see the same process that we have just examined but using the means you can adapt the code as follows:

```
viagraData$mean<-c(rep(2.2,5),rep(3.2,5), rep(5,5))
viagraData$difference = viagraData$libido-viagraData$mean
viagraData$abs.diff = abs(viagraData$difference)
levenes<-aov(abs.diff~dose, data = viagraData)
summary(levenes)
```

You'll find that the *F*-ratio for this analysis is 0.092, which is again non-significant at  $p = .913$ .

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dose	2	0.0853	0.04267	0.0917	0.913
Residuals	12	5.5840	0.46533		

## Please Sir, can I have some more ... Welch's $F$ ?



The Welch (1951)  $F$ -ratio is somewhat complicated (which is why it's stuck on the website). First we have to work out a weight that is based on the sample size,  $n_k$ , and variance,  $s_k^2$ , for a particular group:

$$w_k = \frac{n_k}{s_k^2}$$

We also need to use a grand mean based on a weighted mean for each group. So we take the mean of each group,  $\bar{x}_k$ , and multiply it by its weight,  $w_k$ , do this for each group and add them up, then divide this total by the sum of weights:

$$\bar{x}_{\text{grand}}^{\text{Welch}} = \frac{\sum w_k \bar{x}_k}{\sum w_k}$$

The easiest way to do this is in table form:

Group	Variance $s^2$	Sample size, $n_k$	Weight $w_k$	Mean $\bar{x}_k$	$w_k \times \bar{x}_k$
Placebo	1.70	5	2.941	2.2	6.4702
Low Dose	1.70	5	2.941	3.2	9.4112
High Dose	2.50	5	2.000	5.0	10.000
			$\Sigma = 7.882$		$\Sigma = 25.8814$

So, we get:

$$\bar{x}_{\text{grand}}^{\text{Welch}} = \frac{25.8814}{7.882} = 3.284$$

Think back, to equation (8.5), the model sum of squares was:

$$SS_M = \sum n_k (\bar{x}_k - \bar{x}_{\text{grand}})^2$$

In Welch's  $F$  this is adjusted to incorporate the weighting and the adjusted grand mean:

$$SS_M^{\text{Welch}} = \sum w_k (\bar{x}_k - \bar{x}_{\text{grand}}^{\text{Welch}})^2$$

And to create a mean squares we divide by the degrees of freedom,  $k - 1$ :

$$\begin{aligned} MS_M^{\text{Welch}} &= \frac{\sum w_k (\bar{x}_k - \bar{x}_{\text{grand}}^{\text{Welch}})^2}{k - 1} \\ &= \frac{2.941(2.2 - 3.284)^2 + 2.941(3.2 - 3.284)^2 + 2(5 - 3.284)^2}{2} \\ &= 4.683 \end{aligned}$$

We now have to work out a term called lambda, which is based again on the weights:

$$\Lambda = \frac{3 \sum \left( 1 - \frac{w_k}{\sum w_k} \right)^2}{\frac{n_k - 1}{K^2 - 1}}$$

This looks horrendous, but is just based on the sample size in each group, the weight for each group (and the sum of all weights), and the total number of groups,  $K$ . For the Viagra data this gives us:

$$\Lambda = \frac{3 \left( \frac{\left(1 - \frac{2.941}{7.882}\right)^2}{5-1} + \frac{\left(1 - \frac{2.941}{7.882}\right)^2}{5-1} + \frac{\left(1 - \frac{2}{7.882}\right)^2}{5-1} \right)}{3^2 - 1}$$

$$= \frac{3(0.098 + 0.098 + 0.139)}{8}$$

$$= 0.126$$

The  $F$  ratio is then given by:

$$F_W = \frac{MS_M^{\text{Welch}}}{1 + \frac{2\Lambda(K-2)}{3}}$$

So, for the Viagra data we get:

$$F_W = \frac{4.683}{1 + \frac{(2 \times 0.126)(3-2)}{3}} = \frac{4.683}{1.084} = 4.32$$

The model degrees of freedom stay the same at  $K-1$  (in this case 2), but the residual degrees of freedom,  $df_R$ , are  $1/\Lambda$  (in this case,  $1/0.126 = 7.94$ ).

## Oliver Twisted

Please Sir, can I have some more ... omega?



If you don't want to compute omega squared by hand, you could easily write yourself a function that describes the equation in the book. The reason why writing a function is a good idea is because once you have executed the function once, you can use it over and over again within your current **R** session. This saves you having to write out the equation every time you want to compute omega squared.

If you access the **GLM1.R** file from the companion website you will see that I have written the function for you. The function looks like this:

```
omega<-function(SSm, SSr, dfm, MSr)
{
  SSt = SSm + SSr
  omega = (SSm-(dfm*MSr))/(SSt+MSr)
  print(paste("Omega-Squared: ", omega))
}
```

We'll try to break this down into its constituent parts:

- *omega*: This is the name I have given to the function. If we want to use it in the future, we therefore would type *omega()* just like we do with built-in functions.
- *function(SSm, SSr, dfm, MSr)*: This defines what values are input to the function. To calculate omega squared we need to know four things: the model sum of squares ( $SS_M$ ), the residual sum of squares ( $SS_R$ ) so that we can calculate the total sum of squares, the model degrees of freedom, and the residual mean squares model sum of squares ( $MS_R$ ). When we use the function we will have to input these four values. For example, to calculate omega squared for the example in the book we would simply execute *omega(20.133, 23.600, 2, 1.9667)*, which will feed the four values we have listed into the function *omega()*, which we're about to define.

- `{}`: Anything between these brackets is the function itself. It contains a list of instructions that tell **R** what to do.
- `SSt = SSm + SSr`: The first instruction is to create an object called *SSt*, which is simply the total sum of squares, which is calculated by adding together the model and residual sums of squares (which we entered into the function).
- `omega = (SSm-(dfm*MSr))/(SSt+MSr)`: This creates an object, *omega*, which is the value of omega squared. Notice that this is simply the equation for omega squared from the book. It will take the values entered into the function and place them into this equation.
- `print(paste("Omega-Squared: ", omega))`: This command prints *omega* (the value of omega squared). To make the output nicer we use the *paste()* command to 'paste' together a text string explaining what the statistic is, and the value of the statistic itself.

Once you have executed the above function you can compute omega squared by referring to *omega()* (remembering to include the necessary information for *SSm*, *SSr*, *dfm*, and *MSr*). For example, to compute omega-squared for the Viagra example (see the book chapter), we would execute:

```
omega(20.133, 23.600, 2, 1.9667)
```

The output will look like this:

```
[1] "Omega-Squared: 0.35447935106795"
```

This value matches that computed in the chapter.

## Labcoat Leni's real research

### Scraping the barrel?

#### Problem

Gallup, G. G. J., et al. (2003). *Evolution and Human Behavior*, 24, 277–289.



Evolution has endowed us with many beautiful things (cats, dolphins, the Great Barrier Reef, etc.), all selected to fit their ecological niche. Given evolution's seemingly limitless capacity to produce beauty, it's something of a wonder how it managed to produce such a monstrosity as the human penis. One theory is that the penis evolved into the shape that it is because of sperm competition. Specifically, the human penis has an unusually large glans (the 'bell end' as it's affectionately known) compared to other primates, and this may have evolved so that the penis can displace seminal fluid from other males by 'scooping it out'

during intercourse. To put this idea to the test, Gordon Gallup and his colleagues came up with an ingenious study (Gallup et al., 2003). Armed with various female masturbatory devices from Hollywood Exotic Novelties, an artificial vagina from California Exotic Novelties, and some water and cornstarch to make fake sperm, they loaded the artificial vagina with 2.6 ml of fake sperm and inserted one of three female sex toys into it before withdrawing it. Over several trials, three different female sex toys were used: a control phallus that had no coronal ridge (i.e. no bell end), a phallus with a minimal coronal ridge (small bell end) and a phallus with a coronal ridge.

They measured sperm displacement as a percentage using the following equation (included here because it is more interesting than all of the other equations in this book):

$$\frac{\text{weight of vagina with semen} - \text{weight of vagina following insertion and removal of phallus}}{\text{weight of vagina with semen} - \text{weight of empty vagina}} \times 100$$

As such, 100% means that all of the sperm was displaced by the phallus, and 0% means that none of the sperm was displaced. If the human penis evolved as a sperm displacement device then Gallup et al. predicted: (1) that having a bell end would displace more sperm than not; and (2) the phallus with

the larger coronal ridge would displace more sperm than the phallus with the minimal coronal ridge. The conditions are ordered (no ridge, minimal ridge, normal ridge) so we might also predict a linear trend. The data can be found in the file **Gallup et al.csv**. Draw an error bar graph of the means of the three conditions. Conduct a one-way ANOVA with planned comparisons to test the two hypotheses above. What did Gallup et al. find?

## Solution

First, load the file **Gallup et al.csv** from the companion website into a dataframe called *GallupData* by executing (assuming you have set your working directory to be where the file is located):

```
GallupData = read.csv("Gallup et al.csv", header = TRUE)
```

Let's do the graph first. There are two variables in the data editor: **Phallus** (the independent variable, with three levels: no ridge, minimal ridge and normal ridge) and **Displace** (the dependent variable, the percentage of sperm displaced). The graph should therefore plot **Phallus** on the x-axis and **Displace** on the y-axis.

If you have not already done so, you need to install the *ggplot2* package:

```
install.packages("ggplot2")
```

Even if you have previously installed the *ggplot2* package, you will still need to activate it by executing:

```
library(ggplot2)
```

To plot a categorical variable in *ggplot2* it needs to be recognized as a factor, so we need to create a new variable in the *GallupData* dataframe called **GallupData\$Phallus**, which is just the **Phallus** variable converted into a factor:

```
GallupData$Phallus<-factor(GallupData$Phallus, levels = c(1:3), labels = c("No Coronal Ridge", "Minimal Coronal Ridge", "Coronal Ridge"))
```

We are now ready to plot the graph. As always we first create the plot object and define the variables that we want to plot as aesthetics:

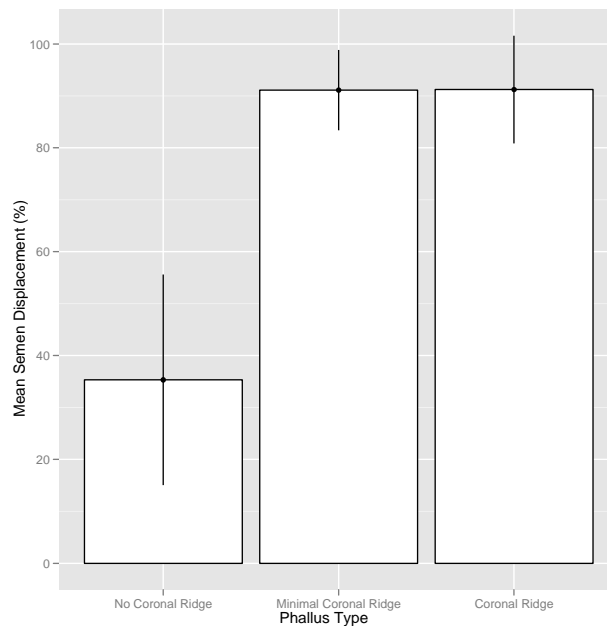
```
bar <- ggplot(GallupData, aes(Phallus, Displace))
```

We can then create an error bar graph by executing the following command:

```
bar + stat_summary(fun.y = mean, geom = "bar", fill = "White", colour = "Black") +  
stat_summary(fun.data = mean_cl_normal, geom = "pointrange") + labs(x = "Phallus  
Type", y = "Mean Semen Displacement (%)")
```

The final graph looks like this:





This graph shows that having a coronal ridge results in more sperm displacement than not having one. The size of ridge made very little difference.

Before we get the ANOVA itself we need to compute the Levene's test to see whether the variance in percentage semen displacement (the outcome) varies across different phallus types:

```
leveneTest(GallupData$Displace, GallupData$Phallus, center = "median")
```

Levene's Test for Homogeneity of Variance (center = "median")

group	Df	F value	Pr(>F)
2	2	1.1516	0.3487
12	12		

The output tells us that Levene's test is not significant,  $F(2, 12) = 1.12$ ,  $p > .05$ , so we can assume that variances are equal.

To run the one-way ANOVA we can execute the following command:

```
GallupModel<-aov(Displace~Phallus, data = GallupData)
```

```
Summary(GallupModel)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Phallus	2	10397.7	5198.8	41.559	4.032e-06 ***
Residuals	12	1501.1	125.1		

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

We can conduct a trend analysis by executing:

```
contrasts(GallupData$Phallus)<-contr.poly(3)
```

The '3' just tells *contr.poly()* how many groups there are in the predictor variable. Having set the contrast, we can create a new model using *aov()*, by executing:

```
GallupTrend<-aov(Displace~Phallus, data = GallupData)
```

To access the contrasts we need the model parameters, which are obtained by executing:

```
summary.lm(GallupTrend)
```

Call:  
aov(formula = Displace ~ Phallus, data = GallupData)

Residuals:

Min	1Q	Median	3Q	Max

```

-13.8276 -7.3288 0.1826 4.1917 26.4298

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  72.549      2.888   25.122 9.60e-12 ***
Phallus.L    39.532      5.002    7.904 4.26e-06 ***
Phallus.Q   -22.731      5.002   -4.545 0.000672 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.18 on 12 degrees of freedom
Multiple R-squared: 0.8738,    Adjusted R-squared: 0.8528
F-statistic: 41.56 on 2 and 12 DF, p-value: 4.032e-06

```

The main ANOVA tells us that there was a significant effect of the type of phallus,  $F(2, 12) = 41.56$ ,  $p < .001$ . (This is exactly the same result as reported in the paper on page 280.) There is a significant linear trend,  $t(12) = 7.90$ ,  $p < .001$ , indicating that more sperm was displaced as the ridge increased (however, note from the graph that this effect reflects the increase in displacement as we go from no ridge to having a ridge; there is no extra increase from 'minimal ridge' to 'coronal ridge'). There is also a significant quadratic trend, indicating that the pattern of means is curvilinear (i.e., is represented by a curve that has one bend),  $t(12) = -4.55$ ,  $p < .001$ . This suggests that having a coronal ridge compared to having no coronal ridge enhances semen displacement but then as the coronal ridge increases beyond a certain point the percentage semen displacement decreases again. This makes sense when looking at the error bar graph, there was very little difference between 'minimal coronal ridge' and 'coronal ridge'.

To test our hypotheses we need to conduct some planned comparisons. To do this we need to tell R what weights to assign to each group:

	Group		
	No Ridge (Control)	Minimal Ridge	Coronal Ridge
<b>Contrast 1</b>	-2	1	1
<b>Contrast 2</b>	0	-1	1

Contrast 1 tests hypothesis 1: (1) that having a bell end will displace more sperm than not. To test this we compare the two conditions with a ridge against the control condition (no ridge). So we compare chunk 1 (no ridge) to chunk 2 (minimal ridge, coronal ridge). The numbers assigned to the groups are the numbers of groups in the opposite chunk, and then we randomly assign one chunk to be a negative value (the codes 2 -1 -1 would work fine as well).

Contrast 2 tests hypothesis 2: (2) the phallus with the larger coronal ridge will displace more sperm than the phallus with the minimal coronal ridge. First we get rid of the control phallus by assigning a code of 0; next we compare chunk 1 (minimal ridge) to chunk 2 (coronal ridge). The numbers assigned to the groups are the numbers of groups in the opposite chunk, and then we randomly assigned one chunk to be a negative value (the codes 0 1 -1 would work fine as well).

To conduct these planned comparisons, we can execute the following commands:

```

contrast1<-c(-2,1,1)
contrast2<-c(0,-1,1)

```

Having created these contrast variables we now need to bind them together using `cbind()`:

```

contrasts(GallupData$Phallus)<-cbind(contrast1,
contrast2)

```

This command sets the contrast property of **Phallus** to contain the weights for the two contrasts that we want to conduct. If you have a look at the **Phallus** variable by executing:

```

GallupData$Phallus

               contrast1 contrast2
No Coronal Ridge        -2         0

```

```
Minimal Coronal Ridge      1      -1
Coronal Ridge              1       1
Levels: No Coronal Ridge Minimal Coronal Ridge Coronal Ridge
```

This output tells us that we entered our weights correctly. 😊

Once we have set the contrast attribute we create a new model using `aov()`, by executing:

```
gallupPlanned<-aov(Displace~Phallus, data = GallupData)

summary.lm(gallupPlanned)

Call:
aov(formula = Displace ~ Phallus, data = GallupData)

Residuals:
    Min       1Q   Median       3Q      Max
-13.8276  -7.3288   0.1826   4.1917  26.4298

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    72.54891     2.88784   25.122 9.60e-12 ***
Phalluscontrast1 18.61686     2.04201    9.117 9.63e-07 ***
Phalluscontrast2  0.05679     3.53686    0.016  0.987
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.18 on 12 degrees of freedom
Multiple R-squared:  0.8738, Adjusted R-squared:  0.8528
F-statistic: 41.56 on 2 and 12 DF, p-value: 4.032e-06
```

Contrast 1 tells us that hypothesis 1 is supported: having some kind of ridge led to greater sperm displacement than not having a ridge,  $t(12) = 9.12$ ,  $p < .001$ . Contrast 2 shows that hypothesis 2 is not supported: the amount of sperm displaced by the normal coronal ridge was not significantly different from the amount displaced by a minimal coronal ridge,  $t(12) = 0.02$ ,  $p = .99$ .

## Smart Alex's solutions

### Task 1

- Imagine that I was interested in how different teaching methods affected students' knowledge. I noticed that some lecturers were aloof and arrogant in their teaching style and humiliated anyone who asked them a question, while others were encouraging and supportive of questions and comments. I took three statistics courses where I taught the same material. For one group of students I wandered around with a large cane and beat anyone who asked daft questions or got questions wrong (*punish*). In the second group I used my normal teaching style, which is to encourage students to discuss things that they find difficult and to give anyone working hard a nice sweet (*reward*). The final group I remained indifferent to and neither punished nor rewarded their efforts (*indifferent*). As the dependent measure I took the students' exam marks (*percentage*). Based on theories of operant conditioning, we expect punishment to be a very unsuccessful way of reinforcing learning, but we expect reward to be very successful. Therefore, one prediction is that reward will produce the best learning. A second hypothesis is that punishment should actually retard learning such that it is worse than an indifferent approach to learning. The data are in the file **Teach.dat**. Carry out a one-way ANOVA and use planned comparisons to test the hypotheses that: (1) reward results in better exam results than either punishment or indifference; and (2) indifference will lead to significantly better exam results than punishment.

First, load the Teach.dat data:

```
teachData = read.delim("Teach.dat", header = TRUE)
```

To get some descriptive statistics for each group we can use the `by()` function:

```
by(teachData$exam, teachData$group, stat.desc)
```

```
teachData$group: 1 (Punish)
  nbr.val  nbr.null  nbr.na    min    max    range    sum
10.0000000 0.0000000 0.0000000 45.0000000 57.0000000 12.0000000 500.0000000
  median    mean  SE.mean CI.mean.0.95    var    std.dev    coef.var
49.5000000 50.0000000 1.30809446 2.95911525 17.1111111 4.13655788 0.08273116
-----
teachData$group: 2 (Indifferent)
  nbr.val  nbr.null  nbr.na    min    max    range    sum
10.0000000 0.0000000 0.0000000 46.0000000 67.0000000 21.0000000 560.0000000
  median    mean  SE.mean CI.mean.0.95    var    std.dev    coef.var
54.5000000 56.0000000 2.2459841 5.0807690 50.4444444 7.1024253 0.1268290
-----
teachData$group: 3 (Reward)
  nbr.val  nbr.null  nbr.na    min    max    range    sum
10.0000000 0.0000000 0.0000000 58.0000000 71.0000000 13.0000000 654.0000000
  median    mean  SE.mean CI.mean.0.95    var    std.dev    coef.var
66.0000000 65.4000000 1.35973854 3.07594227 18.48888889 4.29987080 0.06574726
```

This output shows a table of descriptive statistics; we're told the means, standard deviations and standard errors of the means for each experimental condition. These diagnostics are important for interpretation later on. It looks as though marks are highest after reward and lowest after punishment.

We need to set the categorical variable **group** to be a factor:

```
teachData$group<-factor(teachData$group, levels = c(1:3), labels =
c("Punish","Indifferent", "Reward"))
```

We can then conduct Levene's test:

```
leveneTest(teachData$exam, teachData$group, center = "median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  2  1.7343  0.1956
      27
```

For these data, the assumption of homogeneity of variance has been met, because our significance is .20, which is bigger than the criterion of .05.

Next, we can run the one-way ANOVA by executing:

```
teachModel<-aov(exam~group, data = teachData)
```

```
summary(teachModel)
```

```
      Df Sum Sq Mean Sq F value    Pr(>F)
group   2 1205.1   602.53   21.008 3.145e-06 ***
Residuals 27   774.4    28.68
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the ANOVA output above, the observed significance value is less than .05 and so we can say that there was a significant effect of teaching style on exam marks. However, at this stage we still do not know exactly what the effect of the teaching style was (we don't know which groups differed).

Because there were specific hypotheses, we can specify some contrasts. In the data set, the codes for the categorical variable **group** were set as: Punish = 1, Indifferent = 2 and Reward = 3.

```
contrast1<-c(1,1,-2)
contrast2<-c(1,-1,0)
```

```
contrasts(teachData$group)<-cbind(contrast1, contrast2)
```

```
teachData$group
```

```
      contrast1 contrast2
Punish         1         1
Indifferent    1        -1
Reward        -2         0
Levels: Punish Indifferent Reward
```

The above output shows the codes I used. The first contrast compares reward (coded with -2) against punishment and indifference (both coded with 1). The second contrast compares punishment (coded with 1) against indifference (coded with -1). Note that the codes for each contrast sum to zero, and that in contrast 2, reward has been coded with a 0 because it is excluded from that contrast.

```
teachPlanned<-aov(exam~group, data = teachData)

summary.lm(teachPlanned)

Call:
aov(formula = exam ~ group, data = teachData)

Residuals:
    Min       1Q   Median       3Q      Max
 -10.0    -3.3    -0.7     3.9    11.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   57.1333     0.9778  58.432 < 2e-16 ***
groupcontrast1 -4.1333     0.6914  -5.978 2.24e-06 ***
groupcontrast2 -3.0000     1.1975  -2.505 0.0186 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.356 on 27 degrees of freedom
Multiple R-squared: 0.6088, Adjusted R-squared: 0.5798
F-statistic: 21.01 on 2 and 27 DF, p-value: 3.145e-06
```

The output above shows the significance of the two contrasts specified. The *t*-test for the first contrast tells us that reward was significantly different from punishment and indifference (it's significantly different because the value in the column labelled *Pr(>|t|)* is less than .05). Looking at the means, this tells us that the average mark after reward was significantly higher than the average mark for punishment and indifference combined. The second contrast (together with the descriptive statistics) tells us that the marks after punishment were significantly lower than after indifference (again, it's significantly different because the value in the column labelled *Pr(>|t|)* is less than .05). As such we could conclude that reward produces significantly better exam grades than punishment and indifference, and that punishment produces significantly worse exam marks than indifference. So lecturers should reward their students, not punish them!

### Calculating the effect size

The output from the ANOVA above provides us with three measures of variance: the between-group effect ( $SS_M$ ), the within-subject effect ( $MS_R$ ) and the total amount of variance in the data ( $SS_T$ ; remember that  $SS_T = SS_M + SS_R$ ). We can use these to calculate omega squared ( $\omega^2$ ):

$$\omega^2 = \frac{SS_M - df_M \times MS_R}{SS_T + MS_R}$$

$$\begin{aligned}\omega^2 &= \frac{1205.1 - 2 \times 28.68}{1979.5 + 28.68} \\ &= \frac{1147.74}{2008.18} \\ &= .57 \\ \omega &= .76\end{aligned}$$

For the contrasts (look at the contrasts output to find these numbers) the effect sizes will be:

$$\begin{aligned}r_{\text{contrast}} &= \sqrt{\frac{t^2}{t^2 + df}} \\ r_{\text{contrast1}} &= \sqrt{\frac{(-5.978)^2}{(-5.978)^2 + 27}} \\ &= 0.75\end{aligned}$$

If you think back to our benchmarks for effect sizes this represents a huge effect (it is well above .5, the threshold for a large effect). Therefore, as well as being statistically significant, this effect is large and so represents a substantive finding. For contrast 2 we get:

$$r_{\text{contrast 2}} = \sqrt{\frac{(-2.505)^2}{(-2.505)^2 + 27}} = 0.43$$

This too is a substantive finding and represents a medium to large effect size.

### Interpreting and writing the result

The correct way to report the main finding would be:

- ✓ All significant values are reported at  $p < .05$ . There was a significant effect of teaching style on exam marks,  $F(2, 27) = 21.01$ ,  $\omega^2 = .57$ . Planned contrasts revealed that reward produced significantly better exam grades than punishment and indifference,  $t(27) = -5.98$ ,  $r = .75$ , and that punishment produced significantly worse exam marks than indifference,  $t(27) = -2.51$ ,  $r = .43$ .

## Task 2

- Earlier in this chapter we encountered some data relating to children's injuries while wearing superhero costumes. Children reporting to the emergency centre at hospitals had the severity of their injury (**injury**) assessed (on a scale from 0, no injury, to 100, death). In addition, a note was taken of which superhero costume they were wearing (**hero**): Spiderman, Superman, the Hulk or a Teenage Mutant Ninja Turtle. Use one-way ANOVA and multiple comparisons to test the hypotheses that different costumes are associated with more severe injuries.

First load in the Superhero.dat data:

```
superData = read.delim("Superhero.dat", header = TRUE)
```

Set the categorical variable **hero** to be a factor:

```
superData$hero <- factor(superData$hero, levels = c(1:4), labels =  
c("Spiderman", "Superman", "Hulk", "Ninja Turtle"))
```

We can get some descriptive statistics:

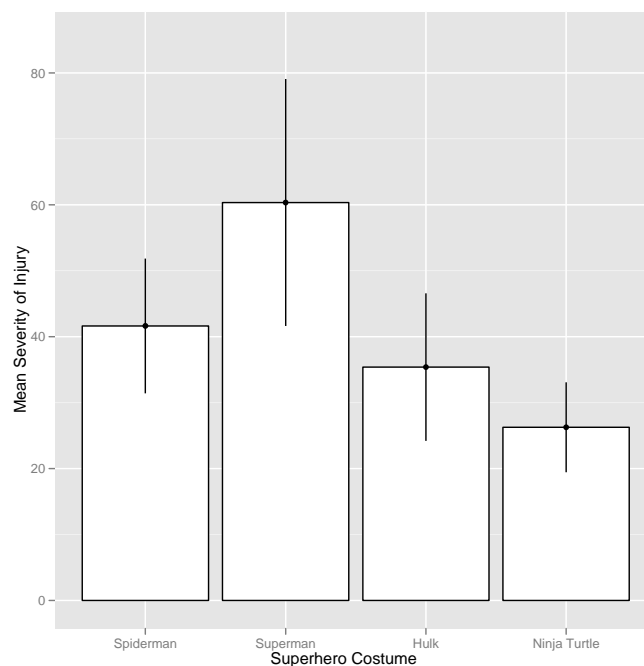
```
by(superData$injury, superData$hero, stat.desc)
```

```
superData$hero: Spiderman
  nbr.val  nbr.null  nbr.na      min      max
 8.0000000  0.0000000  0.0000000 20.0000000 58.0000000
  range      sum      median    mean    SE.mean
38.0000000 333.0000000 43.5000000 41.6250000 4.3174790
CI.mean.0.95  var      std.dev    coef.var
10.2092156 149.1250000 12.2116747 0.2933736
-----
superData$hero: Superman
  nbr.val  nbr.null  nbr.na      min      max
 6.0000000  0.0000000  0.0000000 32.0000000 85.0000000
  range      sum      median    mean    SE.mean
53.0000000 362.0000000 62.0000000 60.3333333 7.2877370
CI.mean.0.95  var      std.dev    coef.var
18.7337244 318.6666667 17.8512371 0.2958769
-----
superData$hero: Hulk
  nbr.val  nbr.null  nbr.na      min      max
 8.0000000  0.0000000  0.0000000 10.0000000 53.0000000
  range      sum      median    mean    SE.mean
43.0000000 283.0000000 38.0000000 35.3750000 4.7318733
CI.mean.0.95  var      std.dev    coef.var
11.1891024 179.1250000 13.3837588 0.3783395
-----
superData$hero: Ninja Turtle
```

	nbr.val	nbr.null	nbr.na	min	max
	8.0000000	0.0000000	0.0000000	18.0000000	41.0000000
	range	sum	median	mean	SE.mean
	23.0000000	210.0000000	27.5000000	26.2500000	2.8831406
	CI.mean.0.95	var	std.dev	coef.var	
	6.8175443	66.5000000	8.1547532	0.3106573	

Looking at the means in the output above, it seems that children wearing a Ninja Turtle costume had the least severe injuries, whereas children wearing a Superman costume had the most severe injuries. The error bar graph below also shows this pattern.

```
bar <- ggplot(superData, aes(hero, injury))
bar + stat_summary(fun.y = mean, geom = "bar", fill = "White", colour = "Black") +
  stat_summary(fun.data = mean_cl_normal, geom = "pointrange") + labs(x = "Superhero
  Costume", y = "Mean Severity of Injury")
```



Before we conduct the ANOVA, we need to conduct Levene's test:

```
leveneTest(superData$injury, superData$hero, center = "median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
      Df    F value    Pr(>F)
group 3    0.827    0.491
     26
```

For these data, the assumption of homogeneity of variance has been met, because our significance is .49, which is much bigger than the criterion of .05.

We can now run the ANOVA:

```
superModel<-aov(injury~hero, data = superData)
```

```
summary(superModel)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Hero	3	4180.6	1393.54	8.3166	0.0004828 ***
Residuals	26	4356.6	167.56		

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

In the ANOVA output above, the observed significance value is less than .05 and so we can say that there was a significant effect of superhero costume on injury severity. However, at this stage we still do not know exactly what the effect of superhero costume was (we don't know which groups differed).

Because there were no specific hypotheses, only that the groups would differ, we can't look at planned contrasts but we can conduct some *post hoc* tests. I am going to conduct a Benjamini–Hochberg (BH) test by executing:

```
pairwise.t.test(superData$injury, superData$hero, p.adjust.method = "BH")

Pairwise comparisons using t tests with pooled SD

data:  superData$injury and superData$hero

      Spiderman Superman Hulk
Superman  0.02544      -      -
Hulk      0.34311  0.00426      -
Ninja Turtle 0.03777  0.00028  0.20452

P value adjustment method: BH
```

Looking at the BH corrected tests we find that wearing a Spiderman costume is significantly different from wearing a Superman or Ninja Turtle costume but not significantly different from wearing a Hulk costume. Wearing a Superman costume is significantly different from wearing any of the other costumes. Finally, wearing a Hulk costume is not significantly different from wearing a Ninja Turtle costume.

The *post hoc* test has shown us which differences between means are significant. However, if we want to see the direction of the effects we can look back to the means and error bar graph above. Looking at the graph and means, we can conclude that wearing a Superman costume resulted in significantly more severe injuries than any of the other costumes (probably because these children were jumping of things in an attempt to fly the most ☺), and wearing a Spiderman costume resulted in significantly more severe injuries than wearing a Ninja Turtle costume but not significantly more severe injuries than wearing a Hulk costume. Wearing a Hulk costume did not result in significantly more severe injuries than wearing a Ninja costume.

### Calculating the effect size

The output from the ANOVA above provides us with three measures of variance: the between-group effect ( $SS_M$ ), the within-subject effect ( $MS_R$ ) and the total amount of variance in the data ( $SS_T$ ). We can use these to calculate omega squared ( $\omega^2$ ):

$$\omega^2 = \frac{SS_M - df_M \times MS_R}{SS_T + MS_R}$$

$$\omega^2 = \frac{4180.6 - 3 \times 167.56}{8537.2 + 167.56}$$

$$= \frac{3677.92}{8704.76}$$

$$= .42$$

$$\omega = .65$$

### Interpreting and writing the result

The correct way to report the main finding would be:

- ✓ All significant values are reported at  $p < .05$ . There was a significant effect of superhero costume on severity of injury,  $F(3, 26) = 8.32$ ,  $\omega^2 = .42$ . Benjamini–Hochberg tests revealed wearing a Superman costume resulted in significantly more severe injuries than any of the other costumes, all  $ps < .05$ . Wearing a Spiderman costume resulted in significantly more severe injuries than wearing a Ninja Turtle costume  $p < .05$  but not significantly more severe injuries than wearing a Hulk costume  $p = .34$ . Wearing a Hulk costume did not result in significantly more severe injuries than wearing a Ninja Turtle costume,  $p = .20$ .



## Task 3

- ✓ In Chapter 15 there are some data looking at whether eating soya meals reduces your sperm count. Have a look at this section, access the data for that example, but analyse them with ANOVA. What's the difference between what you find and what is found in section 15.6.4? Why do you think this difference has arisen?

First load in the **Soya.dat** data:

```
soyaData = read.delim("Soya.dat", header = TRUE)
```

Next, set the categorical variable **soya** to be a factor:

```
soyaData$Soya<-factor(soyaData$Soya, levels = c(1:4), labels = c("No Soya Meals", "1 Soya Meal Per Week", "4 Soya Meals Per Week", "7 Soya Meals Per Week"))
```

Let's have a look at some descriptive statistics by executing:

```
by(soyaData$Sperm, soyaData$Soya, stat.desc)
```

```
soyaData$Soya: No Soya Meals
  nbr.val  nbr.null  nbr.na      min      max      range      sum
20.0000000 0.0000000 0.0000000 0.3510965 21.0800000 20.7289035 99.7352551
  median      mean  SE.mean CI.mean.0.95      var      std.dev      coef.var
 3.0954201  4.9867628  1.1368993  2.3795576  25.8508002  5.0843682  1.0195729
-----
soyaData$Soya: 1 Soya Meal Per Week
  nbr.val  nbr.null  nbr.na      min      max      range      sum
20.0000000 0.0000000 0.0000000 0.3255391 18.4700000 18.1444609 92.1049030
  median      mean  SE.mean CI.mean.0.95      var      std.dev      coef.var
 2.5950621  4.6052451  1.0448316  2.1868577  21.8334621  4.6726290  1.0146320
-----
soyaData$Soya: 4 Soya Meals Per Week
  nbr.val  nbr.null  nbr.na      min      max      range      sum
20.0000000 0.0000000 0.0000000 0.4025473 18.2100000 17.8074527 82.2014229
  median      mean  SE.mean CI.mean.0.95      var      std.dev      coef.var
 2.9428655  4.1100711  0.9860858  2.0639013  19.4473048  4.4099098  1.0729522
-----
soyaData$Soya: 7 Soya Meals Per Week
  nbr.val  nbr.null  nbr.na      min      max      range      sum
20.0000000 0.0000000 0.0000000 0.3100000  4.1100000  3.8000000 33.0602432
  median      mean  SE.mean CI.mean.0.95      var      std.dev      coef.var
 1.3325585  1.6530122  0.2479015  0.5188639  1.2291033  1.1086493  0.6706843
```

From this output, it looks as though as soya intake increases, sperm counts do indeed decrease.

We should next conduct Levene's test:

```
leveneTest(soyaData$Sperm, soyaData$Soya, center = "median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
  Df F value Pr(>F)
group 3  2.8601 0.04239 *
    76
```

For these data, it appears that the assumption of homogeneity of variance has been broken, because our significance is .04, which is smaller than the criterion of .05. In fact, these data also violate the assumption of normality (see Chapter 15 on non-parametric statistics).

#Run the one-way ANOVA:

```
soyaModel<-aov(Sperm~Soya, data = soyaData)
```

```
summary(soyaModel)
```

```
          Df Sum Sq Mean Sq F value    Pr(>F)
Soya      3   135.13    45.043    2.6356  0.05577 .
Residuals 76   1298.85    17.090
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The main ANOVA summary table shows us that because the observed significance value is greater than .05 we can say that there was no significant effect of soya intake on men's sperm count. This is strange because if you read Chapter 15 from where this example came, the Kruskal-Wallis test

produced a significant result! The reason for this difference is that the data violate the assumptions of normality and homogeneity of variance. As I mention in Chapter 15, although parametric tests have more power to detect effects when their assumptions are met, when their assumptions are violated non-parametric tests have more power! This example was arranged to prove this point: because the parametric assumptions are violated, the non-parametric tests produced a significant result and the parametric test did not because, in these circumstances, the non-parametric test has the greater power!

```
oneway.test(Sperm~Soya, data = soyaData)
```

One-way analysis of means (not assuming equal variances)

```
data: Sperm and Soya
F = 6.2844, num df = 3.000, denom df = 34.657, p-value = 0.001607
```

This output shows Welch's  $F$ ; note that the Welch test agrees with the non-parametric test in that the significance of  $F$  is below the .05 threshold. However, in these circumstances, because normality *and* homogeneity of variance have been violated we'd use a non-parametric test anyway!

## Task 4

- ✓ Students (and lecturers for that matter) love their mobile phones, which is rather worrying given some recent controversy about links between mobile phone use and brain tumours. The basic idea is that mobile phones emit microwaves, and so holding one next to your brain for large parts of the day is a bit like sticking your brain in a microwave oven and hitting the 'cook until well done' button. If we wanted to test this experimentally, we could get six groups of people and strap a mobile phone to their heads (so they can't remove it). Then, by remote control, we turn the phones on for a certain amount of time each day. After 6 months, we measure the size of any tumour (in  $\text{mm}^3$ ) close to the site of the phone antenna (just behind the ear). The six groups experienced 0, 1, 2, 3, 4 or 5 hours per day of phone microwaves for six months. The data are in **Tumour.dat** (from Field & Hole, 2003).

As always, we first need to load in the data:

```
tumourData = read.delim("Tumour.dat", header = TRUE)
```

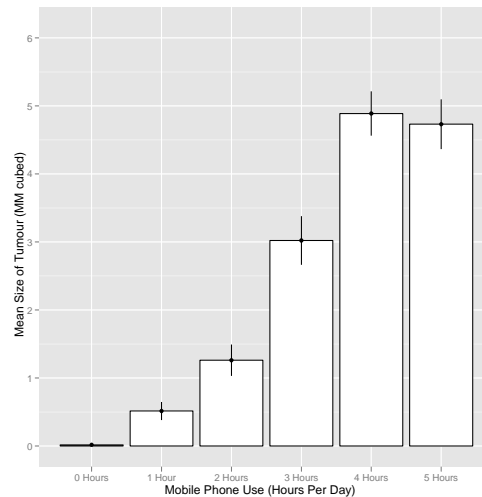
Next we need to set the categorical variable **usage** as a factor:

```
tumourData$usage<-factor(tumourData$usage, levels = c(0:5), labels = c("0 Hours", "1 Hour", "2 Hours", "3 Hours", "4 Hours", "5 Hours"))
```

Now let's plot an error bar graph of these data:

```
bar <- ggplot(tumourData, aes(usage, tumour))

bar + stat_summary(fun.y = mean, geom = "bar", fill = "White", colour = "Black") +
stat_summary(fun.data = mean_cl_normal, geom = "pointrange") + labs(x = "Mobile Phone Use (Hours Per Day)", y = "Mean Size of Tumour (MM cubed)")
```



The error bar chart of the mobile phone data shows the mean size of brain tumour in each condition, and the funny 'I' shapes show the confidence intervals of these means. Note that in the control group (0 hours), the mean size of the tumour is virtually zero (we wouldn't actually expect them to have a tumour) and the error bar shows that there was very little variance across samples. We'll see later that this is problematic for the analysis.

We can request some descriptive statistics:

```
by(tumourData$tumour, tumourData$usage, stat.desc)
```

```
tumourData$usage: 0 Hours
  nbr.val  nbr.null  nbr.na    min    max    range    sum
2.000000e+01 0.000000e+00 0.000000e+00 7.442466e-04 3.711899e-02 3.637475e-02 3.509991e-01
  median    mean  SE.mean CI.mean.0.95    var    std.dev    coef.var
1.607072e-02 1.754996e-02 2.711699e-03 5.675651e-03 1.470662e-04 1.212709e-02 6.910037e-01
-----
tumourData$usage: 1 Hour
  nbr.val  nbr.null  nbr.na    min    max    range    sum
20.00000000 0.00000000 0.00000000 0.00287482 0.94027839 0.93740357 10.29772157
  median    mean  SE.mean CI.mean.0.95    var    std.dev    coef.var
0.54767640 0.51488608 0.06354728 0.13300598 0.08076513 0.28419207 0.55195136
-----
tumourData$usage: 2 Hours
  nbr.val  nbr.null  nbr.na    min    max    range    sum
20.00000000 0.00000000 0.00000000 0.4775269 2.3420716 1.8645447 25.2272512
  median    mean  SE.mean CI.mean.0.95    var    std.dev    coef.var
1.2400708 1.2613626 0.1100540 0.2303457 0.2422376 0.4921764 0.3901943
-----
tumourData$usage: 3 Hours
  nbr.val  nbr.null  nbr.na    min    max    range    sum
20.00000000 0.00000000 0.00000000 1.7742355 4.3064064 2.5321709 60.4318812
  median    mean  SE.mean CI.mean.0.95    var    std.dev    coef.var
2.9401358 3.0215941 0.1711835 0.3582912 0.5860759 0.7655559 0.2533616
-----
tumourData$usage: 4 Hours
  nbr.val  nbr.null  nbr.na    min    max    range    sum
20.00000000 0.00000000 0.00000000 3.0383907 6.0488796 3.0104889 97.7559078
  median    mean  SE.mean CI.mean.0.95    var    std.dev    coef.var
5.1272139 4.8877954 0.1556873 0.3258572 0.4847705 0.6962546 0.1424476
-----
tumourData$usage: 5 Hours
  nbr.val  nbr.null  nbr.na    min    max    range    sum
20.00000000 0.00000000 0.00000000 2.7027423 6.1398957 3.4371534 94.6118789
  median    mean  SE.mean CI.mean.0.95    var    std.dev    coef.var
4.8654203 4.7305939 0.1747768 0.3658121 0.6109388 0.7816257 0.1652278
```

This output shows the means, standard deviations and standard errors of the means for each experimental condition. The means should correspond to those plotted in the graph. These diagnostics are important for interpretation later on.

Next we can conduct Levene's test:

```
leveneTest(tumourData$tumour, tumourData$usage, center = "median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  5  8.3953 8.68e-07 ***
114
```

For these data, the assumption of homogeneity of variance has been violated, because our significance is less than .0001, which is considerably smaller than the criterion of .05. In these situations, we have to try to correct the problem and we can either transform the data or choose Welch's *F*.

To conduct a one-way ANOVA we can execute the following commands:

```
tumourModel<-aov(tumour~usage, data = tumourData)

summary(tumourModel)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
usage	5	450.66	90.133	269.73	< 2.2e-16 ***
Residuals	114	38.09	0.334		

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

The main ANOVA summary table shows us that because the observed significance value is less than .05 we can say that there was a significant effect of mobile phones on the size of tumour. However, at this stage we still do not know exactly what the effect of the phones was (we don't know which groups differed).

We can get the Welch *F* by executing:

```
oneway.test(tumour~usage, data = tumourData)

One-way analysis of means (not assuming equal variances)

data: tumour and usage
F = 414.926, num df = 5.00, denom df = 44.39, p-value < 2.2e-16
```

This output shows the Welch *F*, which is useful because homogeneity of variance was violated. Luckily our conclusions remain the same: both *F*s have significance values less than .05.

Because there were no specific hypotheses, we can just carry out *post hoc* tests:

```
pairwise.t.test(tumourData$tumour, tumourData$usage, p.adjust.method = "BH")
```

Pairwise comparisons using t tests with pooled SD

```
data: tumourData$tumour and tumourData$usage
```

	0 Hours	1 Hour	2 Hours	3 Hours	4 Hours
1 Hour	0.008	-	-	-	-
2 Hours	6.2e-10	9.5e-05	-	-	-
3 Hours	< 2e-16	< 2e-16	3.1e-16	-	-
4 Hours	< 2e-16	< 2e-16	< 2e-16	< 2e-16	-
5 Hours	< 2e-16	< 2e-16	< 2e-16	1.2e-15	0.392

P value adjustment method: BH

It is clear from the table that each group of participants is compared to all of the remaining groups. At the bottom of the table, we can see that group 5 is compared to the groups that used their phones 0, 1, 2, 3 and 4 hours a day and reveals a significant difference in all cases (all the values in the columns are less than .05) except for the comparison between the 4- and 5-hour groups. In the next part of the table, the 4-hour group is compared to all other groups. Again all comparisons are significant (all the values in the columns are less than .05). In fact, all of the comparisons appear to be highly significant except the comparison between the 4- and 5-hour groups, which is non-significant because the value is bigger than .05.

### Calculating the effect size

The main ANOVA output that we obtained earlier provided us with three measures of variance: the between group effect ( $SS_M$ ), the within subject effect ( $MS_R$ ) and the total amount of variance in the data ( $SS_T$ ). We can use these to calculate omega squared ( $\omega^2$ ):

$$\omega^2 = \frac{SS_M - df_M \times MS_R}{SS_T + MS_R}$$

$$\omega^2 = \frac{450.66 - 5 \times 0.334}{488.75 + 0.334}$$

$$\begin{aligned}
 &= \frac{448.99}{488.416} \\
 &= .92 \\
 \omega &= .96
 \end{aligned}$$

**Interpreting and writing the result**

We could report the main finding as:

- ✓ Levene's test indicated that the assumption of homogeneity of variance had been violated,  $F(5, 114) = 8.40$ ,  $p < .001$ . Transforming the data did not rectify this problem and so  $F$ -tests are reported nevertheless. The results show that using a mobile phone significantly affected the size of brain tumour found in participants,  $F(5, 114) = 269.73$ ,  $p < .001$ ,  $\omega^2 = .92$ . The effect size indicated that the effect of phone use on tumour size was substantial.

The next thing that needs to be reported are the *post hoc* comparisons. It is customary just to summarize these tests in very general terms like this:

- ✓ Benjamini–Hochberg *post hoc* tests revealed significant differences between all groups ( $p < .001$  for all tests) except between 4 and 5 hours (*ns*).

## Task 5

- ✓ Using the Glastonbury data from Chapter 7 (**GlastonburyFestivalRegression.dat**), carry out a one-way ANOVA on the data to see if the change in hygiene (**change**) is significantly different across people with different musical tastes (**music**). Do a simple contrast to compare each group against 'No Affiliation'. Compare the results to those described in section 7.11.

Remember to load in the data:

```
festivalData<-read.delim("GlastonburyFestivalRegression.dat", header = TRUE)
```

and then make the categorical variable **music** a factor:

```
festivalData$music<-factor(festivalData$music, levels = c(1:4), labels = c("Indie kid", "Metaller", "Crusty", "No Musical Affiliation"))
```

Conduct Levene's test:

```
leveneTest(festivalData$change, festivalData$music, center = "median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  3  0.7673 0.5145
      119
```

Levene's test is non-significant, showing that variances were roughly equal,  $F(3, 119) = 0.77$ ,  $p > .05$ , across crusties, metallers, indie kids and people with no affiliation.

Next, we can run the one-way ANOVA by executing:

```
festivalModel<-aov(change~music, data = festivalData)
```

```
summary(festivalModel)
```

```

              Df    Sum Sq   Mean Sq  F value    Pr(>F)
music           3     4.646    1.54882    3.2704  0.02369 *
Residuals      119    56.358    0.47359
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
687 observations deleted due to missingness
```

Looking at the ANOVA output above, we could say that the change in hygiene scores was significantly different across the different musical groups,  $F(3, 119) = 3.27$ ,  $p < .05$ . Compare the values in this output to the one in Chapter 7, section 7.12.2, in which we analysed these data as a regression:

```
Call:
lm(formula = change ~ crusty + indie.kid + metallar, data = gfr)

Residuals:
    Min       1Q   Median       3Q      Max
-1.82569 -0.50489  0.05593  0.42430  1.59431

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.55431    0.09036  -6.134 1.15e-08 ***
crustyTRUE   -0.41152    0.16703  -2.464  0.0152  *
indie.kidTRUE -0.40998    0.20492  -2.001  0.0477  *
metallarTRUE  0.02838    0.16033   0.177  0.8598

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

Residual standard error: 0.6882 on 119 degrees of freedom
(687 observations deleted due to missingness)
Multiple R-squared:  0.07617,    Adjusted R-squared:  0.05288
F-statistic: 3.27 on 3 and 119 DF,  p-value: 0.02369
```

If you look at the values that I have highlighted, you will see that they are exactly the same! This should, I hope, re-emphasize to you that regression and ANOVA are the same analytic system!

## Task 6

- ✓ Labcoat Leni's Real Research 15.2 describes an experiment (Çetinkaya & Domjan, 2006) on quails with fetishes for terrycloth objects (really, it does). In this example, you are asked to analyse two of the variables that they measured with a Kruskal–Wallis test. However, there were two other outcome variables (time spent near the terrycloth object and copulatory efficiency). These data can be analysed with one-way ANOVA. Read Labcoat Leni's Real Research 15.2 to get the full story, then carry out two one-way ANOVAs and Bonferroni *post hoc* tests on the aforementioned outcome variables.

Load the data into R:

```
quailData<-read.delim("Cetinkaya & Domjan.dat", header = TRUE)
```

Set the categorical variable **Groups** to be a factor:

```
quailData$Groups<-factor(quailData$Groups, levels = c(1:3), labels =
c("Fetishistics", "NonFetishistics", "Control"))
```

Let's begin by drawing some error bar charts. To create the first error bar chart you could execute:

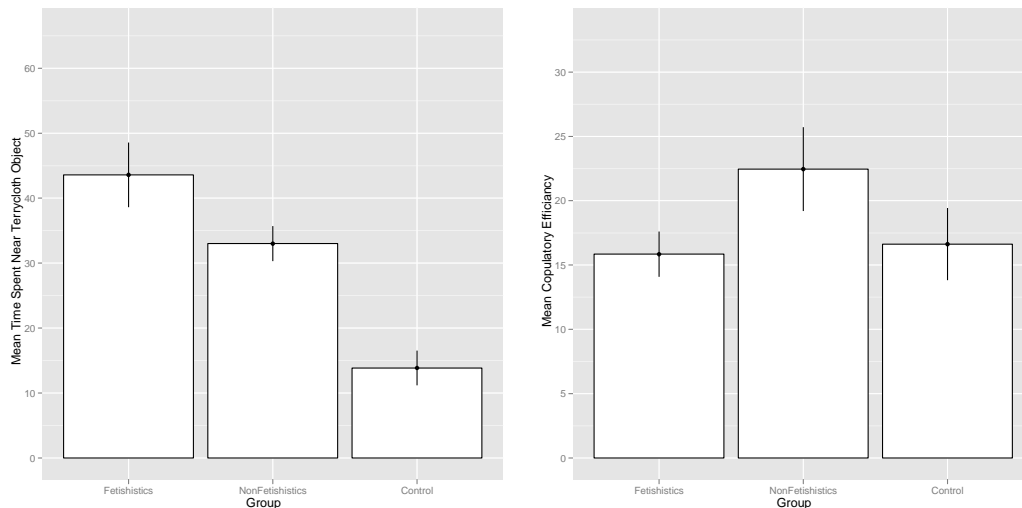
```
bar <- ggplot(quailData, aes(Groups, Duration))

bar + stat_summary(fun.y = mean, geom = "bar", fill = "White", colour = "Black") +
stat_summary(fun.data = mean_cl_normal, geom = "pointrange") + labs(x = "Group", y
= "Mean Time Spent Near Terrycloth Object")
```

To create the second error bar chart you could execute:

```
bar <- ggplot(quailData, aes(Groups, Efficiency))

bar + stat_summary(fun.y = mean, geom = "bar", fill = "White", colour = "Black") +
stat_summary(fun.data = mean_cl_normal, geom = "pointrange") + labs(x = "Group", y
= "Mean Copulatory Efficiency")
```



For the first ANOVA we are looking at whether the group (fetishistic, non-fetishistic or control group) had a significant effect on the time spent near the terrycloth object.

First we need to conduct Levene's tests by executing:

```
leveneTest(quailData$Duration, quailData$Groups, center = "median")

Levene's Test for Homogeneity of Variance (center = "median")
  Df F value Pr(>F)
group 2  1.6938 0.1931
  56
```

This tells us that the homogeneity of variance assumption is met. Therefore, we can use the normal *F*s and Bonferroni *post hoc* tests (which is what the authors of this paper reported).

To conduct the one-way ANOVA we can execute:

```
quailModel.1<-aov(Duration~Groups, data = quailData)

summary(quailModel.1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Groups	2	9880.6	4940.3	91.38	< 2.2e-16 ***
Residuals	56	3027.5	54.1		

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

The outputs tell us that the group (fetishistic, non-fetishistic or control group) had a significant effect on the time spent near the terrycloth object. To find out exactly what's going on we can look at some *post hoc* tests. You were asked to do a Bonferroni *post hoc* test. To do so, execute the following command:

```
pairwise.t.test(quailData$Duration, quailData$Groups, p.adjust.method = "bonferroni")

Pairwise comparisons using t tests with pooled SD
data:  quailData$Duration and quailData$Groups
```

	Fetishistics	NonFetishistics
NonFetishistics	0.00045	-
Control	< 2e-16	1.7e-10

```
P value adjustment method: bonferroni
```

If we want to get confidence intervals of the difference between group means, we can do a Tukey test by executing:

```
postHocs.1<-glht(quailModel.1, linfct = mcp(Groups = "Tukey"))

summary(postHocs.1)
```

## Simultaneous Tests for General Linear Hypotheses

## Multiple Comparisons of Means: Tukey Contrasts

```
Fit: aov(formula = Duration ~ Groups, data = quailData)

Linear Hypotheses:
              Estimate Std. Error t value
NonFetishistics - Fetishistics == 0  -10.588      2.605  -4.065
Control - Fetishistics == 0          -29.736      2.277 -13.062
Control - NonFetishistics == 0       -19.148      2.368  -8.087
              Pr(>|t|)
NonFetishistics - Fetishistics == 0  0.000458 ***
Control - Fetishistics == 0          < 1e-04 ***
Control - NonFetishistics == 0       < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

confint(postHocs.1)
```

## Simultaneous Confidence Intervals

## Multiple Comparisons of Means: Tukey Contrasts

```
Fit: aov(formula = Duration ~ Groups, data = quailData)

Quantile = 2.4053
95% family-wise confidence level

Linear Hypotheses:
              Estimate    lwr      upr
NonFetishistics - Fetishistics == 0 -10.5882 -16.8532  -4.3233
Control - Fetishistics == 0          -29.7364 -35.2120 -24.2608
Control - NonFetishistics == 0       -19.1481 -24.8434 -13.4529
```

We then need to conduct another one-way ANOVA to investigate whether the group (fetishistic, non-fetishistic or control group) had a significant effect on copulatory efficiency. To do this, we follow the same procedure as before.

First, conduct a Levene's test:

```
leveneTest(quailData$Efficiency, quailData$Groups, center = "median")

Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  2  1.4451 0.2444
      56
```

This tells us that the homogeneity of variance assumption is met. Therefore, we can use the normal *F*s and Bonferroni *post hoc* tests (which is what the authors of this paper reported).

Now conduct the one-way ANOVA:

```
quailModel.2<-aov(Efficiency~Groups, data = quailData)

summary(quailModel.2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Groups	2	427.56	213.78	6.0373	0.004224 **
Residuals	56	1982.97	35.41		

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

This output tells us that the group (fetishistic, non-fetishistic or control group) had a significant effect on copulatory efficiency. To find out exactly what's going on we can look at some *post hoc* tests. You were asked to do a Bonferroni *post hoc* test for each ANOVA. To do so, execute the following command:

```
pairwise.t.test(quailData$Efficiency, quailData$Groups, p.adjust.method =
"bonferroni")

Pairwise comparisons using t tests with pooled SD
```



```
data: quailData$Efficiency and quailData$Groups
```

```

      Fetishistics NonFetishistics
NonFetishistics 0.0081 -
Control         1.0000      0.0106

```

```
P value adjustment method: bonferroni
```

Again, if we want to get confidence intervals of the difference between group means we can do a Tukey test by executing:

```
postHocs.2<-glht(quailModel.2, linfct = mcp(Groups = "Tukey"))
```

```
summary(postHocs.2)
```

```
Simultaneous Tests for General Linear Hypotheses
```

```
Multiple Comparisons of Means: Tukey Contrasts
```

```
Fit: aov(formula = Efficiency ~ Groups, data = quailData)
```

```
Linear Hypotheses:
```

	Estimate	Std. Error	t value
NonFetishistics - Fetishistics == 0	6.6146	2.1080	3.138
Control - Fetishistics == 0	0.7794	1.8424	0.423
Control - NonFetishistics == 0	-5.8353	1.9163	-3.045

	Pr(> t )
NonFetishistics - Fetishistics == 0	0.00748 **
Control - Fetishistics == 0	0.90592
Control - NonFetishistics == 0	0.00969 **

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

```
confint(postHocs.2)
```

```
Simultaneous Confidence Intervals
```

```
Multiple Comparisons of Means: Tukey Contrasts
```

```
Fit: aov(formula = Efficiency ~ Groups, data = quailData)
```

```
Quantile = 2.4051
```

```
95% family-wise confidence level
```

```
Linear Hypotheses:
```

	Estimate	lwr	upr
NonFetishistics - Fetishistics == 0	6.6146	1.5446	11.6846
Control - Fetishistics == 0	0.7794	-3.6518	5.2106
Control - NonFetishistics == 0	-5.8353	-10.4442	-1.2263

The authors reported as follows (pp. 429–430): ‘A one-way ANOVA indicated significant group differences,  $F(2, 56) = 91.38$ ,  $p < .05$ ,  $\eta^2 = 0.76$ . Subsequent pairwise comparisons (with the Bonferroni correction) revealed that fetishistic male quail stayed near the CS [the terrycloth object] longer than both the nonfetishistic male quail (mean difference = 10.59 s; 95% CI = 4.16, 17.02;  $p < .05$ ) and the control male quail (mean difference = 29.74 s; 95% CI = 24.12, 35.35;  $p < .05$ ). In addition, the nonfetishistic male quail spent more time near the CS than did the control male quail (mean difference = 19.15 s; 95% CI = 13.30, 24.99;  $p < .05$ ).’

Look at the graph, the ANOVA output and the *post hoc* tests to see from where the values that they report come. As you will see, the ANOVA values and *post hoc*  $p$ -values match those reported in the paper. However, the mean differences and their confidence intervals are slightly different. This is because we used a different method than that used in the paper. The author of the paper used Bonferroni corrected  $ps$ , whereas we have used Tukey. This is because the author of the paper used a different statistical package to analyse the data.

For the copulatory efficiency outcome the authors reported as follows (p. 430): ‘A one-way ANOVA yielded a significant main effect of groups,  $F(2, 56) = 6.04$ ,  $p < .05$ ,  $\eta^2 = 0.18$ . Paired comparisons (with the Bonferroni correction) indicated that the nonfetishistic male quail copulated with the live female quail ... more efficiently than both the fetishistic male quail (mean difference = 6.61; 95% CI = 1.41, 11.82;  $p < .05$ ) and the control male quail (mean difference = 5.83; 95% CI = 1.11, 10.56;  $p < .05$ ). The

difference between the efficiency scores of the fetishistic and the control male quail was not significant (mean difference = 0.78; 95% CI = -5.33, 3.77;  $p > .05$ ).'

These results show that male quails do show fetishistic behaviour (the time spent with the terrycloth) and that this affects their copulatory efficiency (they are less efficient than those that don't develop a fetish, but it's worth remembering that they are no worse than quails that had no sexual conditioning – the controls). If you look at Labcoat Leni's Box 15.2 then you'll also see that this fetishistic behaviour may have evolved because the quails with fetishistic behaviour manage to fertilize a greater percentage of eggs (so their genes are passed on!).