12 Factorial ANOVA (GLM 3)

FIGURE 12.1
Andromeda coming to a living room near you in 1988 (from left to right: Malcolm, me and the two Marks)

12.1. What will this chapter tell me? ②

After persuading my two friends (Mark and Mark) to learn the bass and drums, I took the rather odd decision to *stop* playing the guitar. I didn't stop, as such, but I focused on singing instead. In retrospect, I'm not sure why because I am *not* a good singer. Mind you, I'm not a good guitarist either. The upshot was that a classmate, Malcolm, ended up as our guitarist. I really can't remember how or why we ended up in this configuration, but we called ourselves Andromeda, we learnt several Queen and Iron Maiden songs and we were truly awful. I have some tapes somewhere to prove just what a cacophony of tuneless drivel we produced, but the chances of these recordings appearing on the companion website are slim at best. Suffice it to say, you'd be hard pushed to recognize *which* Iron Maiden and Queen songs we were trying to play. I try to comfort myself with the fact that we were only 14 or 15 at the time, but even youth does not excuse the depths of ineptitude to which we sank. Still, we garnered a reputation for being too loud in school assembly and we did a successful tour of our friends' houses (much to their parents' amusement I'm sure). We even started to write a few songs (I wrote one called 'Escape from Inside', about the film *The Fly*, that contained the wonderful rhyming couplet of 'I am a fly, I want to die': genius!). The only thing that we did that resembled the activities of a 'proper' band was to split up due to 'musical differences'; these differences being that Malcolm wanted to write 15-part symphonies about a boy's journey to worship electricity pylons and discover a mythical beast called the cuteasauros, whereas I wanted to write songs about flies and dying (preferably both). When we could not agree on a musical direction the split became inevitable. We could have tested empirically the best musical direction for the band by writing and performing two songs: Malcolm his 15-part symphony and me my 3-minute song about a fly. If we played these songs to various people and measured their screams of agony then we could ascertain the best musical direction to gain popularity. We have two variables that predict screams: whether Malcolm or I wrote the song (songwriter), and whether the song was a 15-part symphony or a song about a fly (song type). The one-way ANOVA that we encountered in Chapter 10 cannot deal with two predictor variables – this is a job for factorial ANOVA.

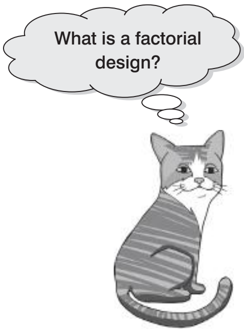
12.2. Theory of factorial ANOVA (independent design) ②

In the previous two chapters we have looked at situations in which we've tried to test for differences between groups when there has been a single independent variable (i.e., one variable has been manipulated). However, at the beginning of Chapter 10 I said that one of the advantages of ANOVA was that we could look at the effects of more than one independent variable (and how these variables interact). This chapter extends what we already know about ANOVA to look at situations where there are two (or more) independent variables. We've already seen in the previous chapter that it's very easy to incorporate a second variable into the ANOVA framework when that variable is a continuous variable (i.e., not split into groups), but now we'll move on to situations where there is a second independent variable that has been systematically manipulated by assigning people to different conditions.

12.2.1. Factorial designs ②

In Chapters 10 and 11 we have looked at the effects of a single independent variable on some outcome. However, independent variables often get lonely and want to have friends. Scientists are obliging individuals and often put a second (or third) independent variable into their designs to keep the others company. When an experiment has two or more independent variables it is known as a *factorial design* (this is because, as we have seen, variables are sometimes referred to as *factors*). There are several types of factorial design:

- **Independent factorial design**: In this type of experiment there are several independent variables or predictors and each has been measured using different entities (between groups). We discuss this design in this chapter.



- **Repeated-measures (related) factorial design**: This is an experiment in which several independent variables or predictors have been measured, but the same entities have been used in all conditions. This design is discussed in Chapter 13.
- **Mixed design**: This is a design in which several independent variables or predictors have been measured; some have been measured with different entities, whereas others used the same entities. This design is discussed in Chapter 14.

As you might imagine, analysing these types of experiments can get quite complicated. Fortunately, we can extend the ANOVA model that we encountered in the previous two chapters to deal with these more complicated situations. When we use ANOVA to analyse a situation in which there are two or more independent variables it is sometimes called **factorial ANOVA**; however, the specific names attached to different ANOVAs reflect the experimental design that they are being used to analyse (see Jane Superbrain Box 12.1). This section extends the one-way ANOVA model to the factorial case (specifically when there are two independent variables). In subsequent chapters we will look at repeated-measures designs, factorial repeated-measures designs and finally mixed designs.

JANE SUPERBRAIN 12.1

Naming ANOVAs ②

ANOVAs can be quite confusing because there appear to be lots of them. When you read research articles you'll quite often come across phrases like 'a two-way independent ANOVA was conducted', or 'a three-way repeated-measures ANOVA was conducted'. These names may look confusing but they are quite easy if you break them down. All ANOVAs have two things in common: they involve some quantity of independent variables, and these variables can be measured using either the same or different participants. If the same participants are used we typically use the term *repeated measures*, and if different participants are used we use the term *independent*. When there are two or more independent variables, it's possible that some variables use the same participants whereas others use different participants. In this case we use the term *mixed*. When we name an ANOVA, we are simply telling the reader how many independent variables we used and how they were measured. In general terms we could write the name of an ANOVA as:

- (number of independent variables)-way (how these variables were measured) ANOVA.

By remembering this you can understand the name of any ANOVA you come across. Look at these examples and try to work out how many variables were used and how they were measured:

- one-way independent ANOVA;
- two-way repeated-measures ANOVA;
- two-way mixed ANOVA;
- three-way independent ANOVA.

The answers you should get are:

- one independent variable measured using different participants;
- two independent variables both measured using the same participants;
- two independent variables: one measured using different participants and the other measured using the same participants;
- three independent variables all of which are measured using different participants.

12.3. Factorial ANOVA as regression ③

12.3.1. An example with two independent variables ②

Throughout this chapter we'll use an example that has two independent variables. This is known as a two-way ANOVA (see Jane Superbrain Box 12.1). I'll look at an example with two independent variables because this is the simplest extension of the ANOVAs that we have already encountered.

An anthropologist was interested in the effects of alcohol on mate selection at nightclubs. Her rationale was that after alcohol had been consumed, subjective perceptions of physical attractiveness would become more inaccurate (the well-known **beer-goggles effect**). She was also interested in whether this effect was different for men and women. She picked 48 students: 24 male and 24 female. She then took groups of eight participants to a nightclub and gave them no alcohol (participants received placebo drinks of alcohol-free lager), 2 pints of strong lager, or 4 pints of strong lager. At the end of the evening she took a photograph of the person that the participant was chatting up. She then got a pool of independent judges to assess the attractiveness of the person in each photograph (out of 100). The data are in Table 12.1 and **goggles.csv**.

Table 12.1 Data for the beer-goggles effect

Alcohol	*None*		*2 Pints*		*4 Pints*	
Gender	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*
	65	50	70	45	55	30
	70	55	65	60	65	30
	60	80	60	85	70	30
	60	65	70	65	55	55
	60	70	65	70	55	35
	55	75	60	70	60	20
	60	75	60	80	50	45
	55	65	50	60	50	40
Total	485	535	500	535	460	285
Mean	60.625	66.875	62.50	66.875	57.50	35.625
Variance	24.55	106.70	42.86	156.70	50.00	117.41

12.3.2. Extending the regression model ③

We saw in section 10.2.3 that one-way ANOVA could be conceptualized as a regression equation (a general linear model). In this section we'll consider how we extend this linear model to incorporate two independent variables. To keep things as simple as possible I want you to imagine that we have only two levels of the alcohol variable in our example (none and 4 pints). As such, we have two predictor variables, each with two levels. All of the general linear models we've considered in this book take the general form of:

$$\text{outcome}_i = (\text{model}) + \text{error}_i$$

For example, when we encountered multiple regression in Chapter 7 we saw that this model was written as (see equation (7.9)):

$$Y_i = (b_0 + b_1 X_{1i} + b_2 X_{2i} + \quad + b_n X_{ni}) + \varepsilon_i$$

Also, when we came across one-way ANOVA, we adapted this regression model to conceptualize our Viagra example, as (see equation (10.2)):

$$\text{libido}_i = (b_0 + b_2 \text{high}_i + b_1 \text{low}_i) + \varepsilon_i$$

In this model, the high and low variables were dummy variables (i.e., variables that can take only values of 0 or 1). In our current example, we have two variables: **gender** (male or female) and **alcohol** (none and 4 pints). We can code each of these with zeros and ones; for example, we could code gender as male = 0, female = 1, and we could code the alcohol variable as 0 = none, 1 = 4 pints. We could then directly copy the model we had in one-way ANOVA:

$$\text{attractiveness}_i = (b_0 + b_1 \text{gender}_i + b_2 \text{alcohol}_i) + \varepsilon_i$$

However, this model does not consider the interaction between gender and alcohol. If we want to include this term too, then the model simply extends to become (first expressed generally and then in terms of this specific example):

$$\begin{aligned} \text{attractiveness}_i &= (b_0 + b_1 A_i + b_2 B_i + b_3 AB_i) + \varepsilon_i \\ \text{attractiveness}_i &= (b_0 + b_1 \text{gender}_i + b_2 \text{alcohol}_i + b_3 \text{ interaction}_i) + \varepsilon_i \end{aligned} \tag{12.1}$$

The question is: how do we code the interaction term? The interaction term represents the combined effect of **alcohol** and **gender**; to get any interaction term in regression you simply multiply the variables involved. This is why you see interaction terms written as **gender** × **alcohol**, because in regression terms the interaction variable literally is the two variables multiplied by each other. Table 12.2 shows the resulting variables for the regression (note that the interaction variable is simply the value of the gender dummy variable multiplied by the value of the alcohol dummy variable). So, for example, a male receiving 4 pints of alcohol would have a value of 0 for the gender variable, 1 for the alcohol variable and 0 for the interaction variable. The group means for the various combinations of gender and alcohol are also included because they'll come in useful in due course.

Table 12.2 Coding scheme for factorial ANOVA

Gender	Alcohol	Dummy (Gender)	Dummy (Alcohol)	Interaction	Mean
Male	None	0	0	0	66.875
Male	4 Pints	0	1	0	35.625
Female	None	1	0	0	60.625
Female	4 Pints	1	1	1	57.500

To work out what the *b*-values represent in this model we can do the same as we did for the *t*-test and one-way ANOVA; that is, look at what happens when we insert values of our predictors (**gender** and **alcohol**). To begin with, let's see what happens when we look at men who had no alcohol. In this case, the value of **gender** is 0, the value of **alcohol** is 0 and the value of **interaction** is also 0. The outcome we predict (as with one-way ANOVA) is the mean of this group (66.875), so our model becomes:

$$\begin{aligned}
\text{attractiveness}_i &= (b_0 + b_1\text{gender}_i + b_2\text{alcohol}_i + b_3\,\text{interaction}_i) + \varepsilon_i \\
\bar{X}_{\text{Men, None}} &= b_0 + (b_1 \quad 0) + (b_2 \quad 0) + (b_3 \quad 0) \\
b_0 &= \bar{X}_{\text{Men, None}} \\
b_0 &= 66.875
\end{aligned}$$

So, the constant b_0 in the model represents the mean of the group for which all variables are coded as 0. As such it's the mean value of the base category (in this case men who had no alcohol).

Now, let's see what happens when we look at females who had no alcohol. In this case, the **gender** variable is 1 and the **alcohol** and **interaction** variables are still 0. Also remember that b_0 is the mean of the men who had no alcohol. The outcome is the mean for women who had no alcohol. Therefore, the equation becomes:

$$\begin{aligned}
\bar{X}_{\text{Women, None}} &= b_0 + (b_1 \times 1) + (b_2 \times 0) + (b_3 \times 0) \\
\bar{X}_{\text{Women, None}} &= b_0 + b_1 \\
\bar{X}_{\text{Women, None}} &= \bar{X}_{\text{Men, None}} + b_1 \\
b_1 &= \bar{X}_{\text{Women, None}} - \bar{X}_{\text{Men, None}} \\
b_1 &= 60.625 - 66.875 \\
b_1 &= -6.25
\end{aligned}$$

So, b_1 in the model represents the difference between men and women who had no alcohol. More generally we can say it's the effect of **gender** for the base category of **alcohol** (the base category being the one coded with 0, in this case no alcohol).

Now let's look at males who had 4 pints of alcohol. In this case, the **gender** variable is 0, the **alcohol** variable is 1 and the **interaction** variable is still 0. We can also replace b_0 with the mean of the men who had no alcohol. The outcome is the mean for men who had 4 pints. Therefore, the equation becomes:

$$\begin{aligned}
\bar{X}_{\text{Men, 4 Pints}} &= b_0 + (b_1 \times 0) + (b_2 \times 1) + (b_3 \times 0) \\
\bar{X}_{\text{Men, 4 Pints}} &= b_0 + b_2 \\
\bar{X}_{\text{Men, 4 Pints}} &= \bar{X}_{\text{Men, None}} + b_2 \\
b_2 &= \bar{X}_{\text{Men, 4 Pints}} - \bar{X}_{\text{Men, None}} \\
b_2 &= 35.625 - 66.875 \\
b_2 &= -31.25
\end{aligned}$$

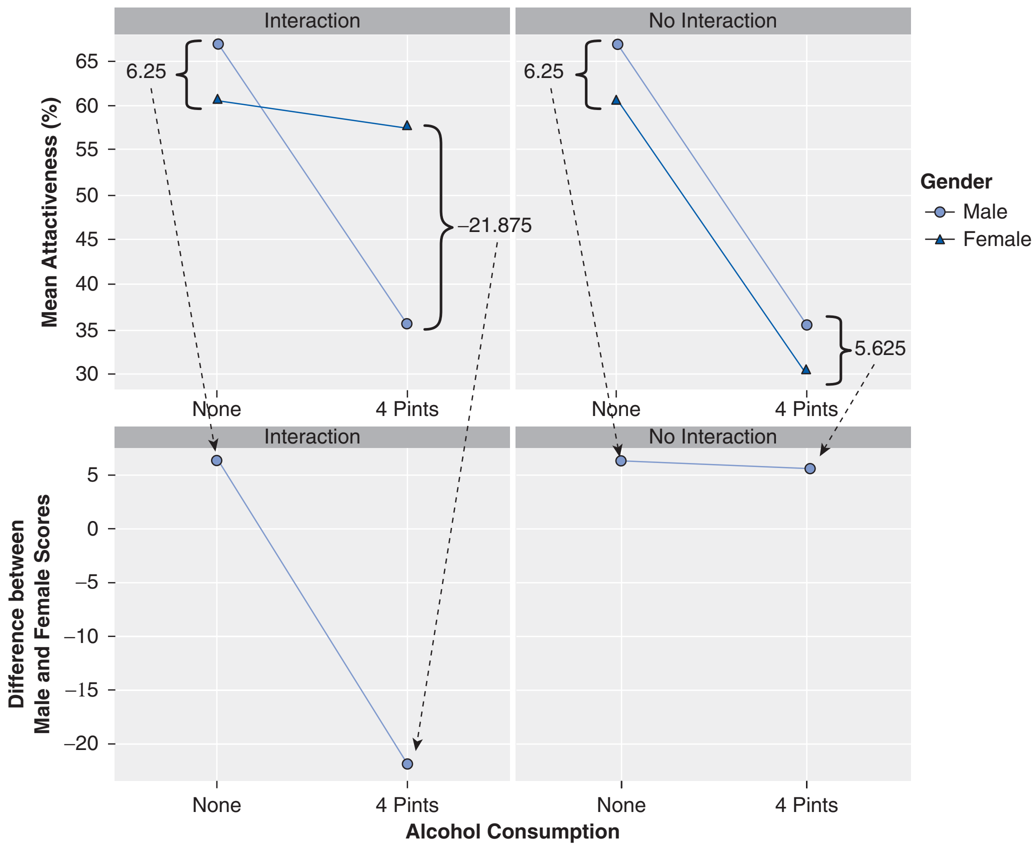
So, b_2 in the model represents the difference between having no alcohol and 4 pints in men. Put more generally, it's the effect of **alcohol** in the base category of **gender** (i.e., the category of **gender** that was coded with a 0, in this case men).

Finally, we can look at females who had 4 pints of alcohol. In this case, **gender** is 1, **alcohol** is 1 and **interaction** is also 1. We can also replace b_0, b_1 and b_2 with what we now know they represent. The outcome is the mean for women who had 4 pints. Therefore, the equation becomes:

$$\bar{X}_{\text{Women, 4 Pints}} = b_0 + (b_1 \times 1) + (b_2 \times 1) + (b_3 \times 1)$$
$$\bar{X}_{\text{Women, 4 Pints}} = b_0 + b_1 + b_2 + b_3$$
$$\bar{X}_{\text{Women, 4 Pints}} = \bar{X}_{\text{Men, None}} + (\bar{X}_{\text{Women, None}} - \bar{X}_{\text{Men, None}}) + (\bar{X}_{\text{Men, 4 Pints}} - \bar{X}_{\text{Men, None}}) + b_3$$
$$\bar{X}_{\text{Women, 4 Pints}} = \bar{X}_{\text{Women, None}} + \bar{X}_{\text{Men, 4 Pints}} - \bar{X}_{\text{Men, None}} + b_3$$
$$b_3 = \bar{X}_{\text{Men, None}} - \bar{X}_{\text{Women, None}} + \bar{X}_{\text{Women, 4 Pints}} - \bar{X}_{\text{Men, 4 Pints}}$$
$$b_3 = 66.875 - 60.625 + 57.500 - 35.625$$
$$b_3 = 28.125$$

So, b_3 in the model really compares the difference between men and women in the no-alcohol condition to the difference between men and women in the 4 pints condition. Put another way, it compares the effect of gender after no-alcohol to the effect of gender after 4 pints.[1] If you think about it in terms of an interaction graph, this makes perfect sense. For example, the top left-hand side of Figure 12.2 shows the interaction graph for these data. Now imagine we calculated the difference between men and women for the no-alcohol groups. This would

FIGURE 12.2 Breaking down what an interaction represents



[1] In fact, if you rearrange the terms in the equation you'll see that you can also phrase the interaction the opposite way around: it represents the effect of alcohol in men compared to women.

be the difference between the lines on the graph for the no-alcohol group (the difference between group means, which is 6.25). If we then do the same for the 4 pints group, we find that the difference between men and women is −21.875. If we plotted these two values as a new graph we'd get a line connecting 6.25 to −21.875 (see the bottom left-hand side of Figure 12.2). This reflects the difference between the effect of gender after no alcohol compared to after 4 pints. We know that beta values represent gradients of lines, and in fact b_3 in our model is the gradient of this line (this is $6.25 - (-21.875) = 28.125$).

Let's also see what happens if there isn't an interaction effect: the right-hand side of Figure 12.2 shows the same data except that the mean for the females who had 4 pints has been changed to 30. If we calculate the difference between men and women after no alcohol we get the same as before: 6.25. If we calculate the difference between men and women after 4 pints we now get 5.625. If we again plot these differences on a new graph, we find a virtually horizontal line. So, when there's no interaction, the line connecting the effect of gender after no alcohol and after 4 pints is flat and the resulting b_3 in our model would be close to 0 (remember that a zero gradient means a flat line). In fact its actual value would be $6.25 - 5.625 = 0.625$.

SELF-TEST

✓ The file **GogglesRegression.dat** contains the dummy variables used in this example. Just to prove that all of this works, use this file and run a multiple regression on the data.

The resulting table of coefficients is in Output 12.1. The important thing to note is that the beta value for the interaction (28.125) is the same as we've just calculated, which should hopefully convince you that factorial ANOVA is – as is everything, it would seem – just regression dressed up in a different costume.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   66.875      3.055  21.890  < 2e-16 ***
gender        -6.250      4.320  -1.447    0.159
alcohol      -31.250      4.320  -7.233 7.13e-08 ***
interaction   28.125      6.110   4.603 8.20e-05 ***
---
```

Output 12.1

What I hope to have shown you in this example is how even complex ANOVAs are just forms of regression (a general linear model). You'll be pleased to know (I'll be pleased to know for that matter) that this is the last I'm going to say about ANOVA as a general linear model. I hope I've given you enough background so that you get a sense of the fact that we can just keep adding independent variables into our model. All that happens is these new variables just get added into a multiple regression equation with an associated beta value (just like the regression chapter). Interaction terms can also be added simply by multiplying the variables that interact. These interaction terms will also have an associated beta value. So, any ANOVA (no matter how complex) is just a form of multiple regression.

12.4. Two-way ANOVA: behind the scenes ②

Now that we have a good conceptual understanding of factorial ANOVA as an extension of the basic idea of a linear model, we will turn our attention to some of the specific calculations that go on behind the scenes. The reason for doing this is that it should help you to understand what the output of the analysis means.

Two-way ANOVA is conceptually very similar to one-way ANOVA. Basically, we still find the total sum of squared errors (SS_T) and break this variance down into variance that can be explained by the experiment (SS_M) and variance that cannot be explained (SS_R). However, in two-way ANOVA, the variance explained by the experiment is made up of not one experimental manipulation but two. Therefore, we break the model sum of squares down into variance explained by the first independent variable (SS_A), variance explained by the second independent variable (SS_B) and variance explained by the interaction of these two variables ($SS_{A\times B}$) – see Figure 12.3.

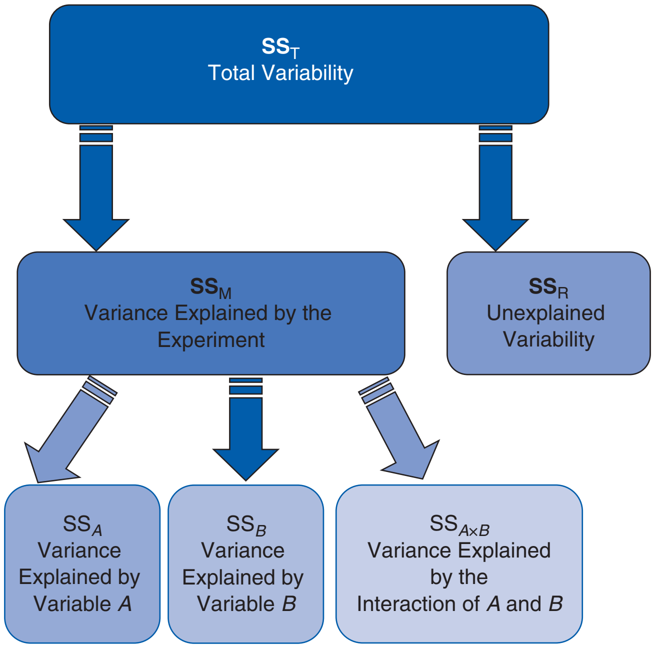


FIGURE 12.3 Breaking down the variance in two-way ANOVA

12.4.1. Total sums of squares (SS_T) ②

We start off in the same way as we did for a one-way ANOVA. That is, we calculate how much variability there is between scores when we ignore the experimental condition from which they came. Remember from one-way ANOVA (equation (10.4)) that SS_T is calculated using the following equation:

$$\begin{aligned} SS_T &= \sum_{i=1}^{N}(x_i - \bar{x}_{\text{grand}})^2 \\ &= s^2_{\text{grand}}(N-1) \end{aligned}$$

The grand variance is simply the variance of all scores when we ignore the group to which they belong. So if we treated the data as one big group it would look as follows:

65	50	70	45	55	30
70	55	65	60	65	30
60	80	60	85	70	30
60	65	70	65	55	55
60	70	65	70	55	35
55	75	60	70	60	20
60	75	60	80	50	45
55	65	50	60	50	40
Grand mean = 58.33					

If we calculate the variance of all of these scores, we get 190.78 (try this on your calculator if you don't trust me). We used 48 scores to generate this value, and so N is 48. As such the equation becomes:

$$
\begin{aligned}
\mathrm{SS_T} &= s^2_{\text{grand}}(N-1) \\
&= 190.78(48-1) \\
&= 8966.66
\end{aligned}
$$

The degrees of freedom for this SS will be $N - 1$, or 47.

12.4.2. The model sum of squares ($\mathrm{SS_M}$) ②

The next step is to work out the model sum of squares. As I suggested earlier, this sum of squares is then further broken into three components: variance explained by the first independent variable (SS_A), variance explained by the second independent variable (SS_B) and variance explained by the interaction of these two variables ($\mathrm{SS}_{A\times B}$).

Before we break down the model sum of squares into its component parts, we must first calculate its value. We know we have 8966.66 units of variance to be explained, and our first step is to calculate how much of that variance is explained by our experimental manipulations overall (ignoring which of the two independent variables is responsible). When we did one-way ANOVA we worked out the model sum of squares by looking at the difference between each group mean and the overall mean (see section 10.2.6). We can do the same here. We effectively have six experimental groups if we combine all levels of the two independent variables (three doses for the male participants and three doses for the females). So, given that we have six groups of different people we can then apply the equation for the model sum of squares that we used for one-way ANOVA (equation (10.5)):

$$
\mathrm{SS_M} = \sum_{n=1}^{k} n_k(\bar{x}_k - \bar{x}_{\text{grand}})^2
$$

The grand mean is the mean of all scores (we calculated this above as 58.33) and n is the number of scores in each group (i.e., the number of participants in each of the six experimental groups; eight in this case). Therefore, the equation becomes:

$$\begin{aligned}\mathrm{SS_M} &= 8(60.625-58.33)^2+8(66.875-58.33)^2+8(62.5-58.33)^2+\ldots\\ &\quad +8(66.875-58.33)^2+8(57.5-58.33)^2+8(35.625-58.33)^2\\ &= 8(2.295)^2+8(8.545)^2+8(4.17)^2+8(8.545)^2+8(-0.83)^2+8(-22.705)^2\\ &= 42.1362+584.1362+139.1112+584.1362+5.5112+4124.1362\\ &= 5479.167\end{aligned}$$

The degrees of freedom for this SS will be the number of groups used, k, minus 1. We used six groups and so $df = 5$.

At this stage we know that the model (our experimental manipulations) can explain 5479.167 units of variance out of the total of 8966.66 units. The next stage is to further break down this model sum of squares to see how much variance is explained by our independent variables separately.

12.4.2.1. The main effect of gender ($\mathrm{SS_A}$) ②

To work out the variance accounted for by the first independent variable (in this case, **gender**) we need to group the scores in the data set according to the gender to which they belong. So, basically we ignore the amount of drink that has been drunk, and we just place all of the male scores into one group and all of the female scores into another. So, the data will look like this (note that the first box contains the three female columns from our original table and the second box contains the male columns):

A_1: Female		
65	70	55
70	65	65
60	60	70
60	70	55
60	65	55
55	60	60
60	60	50
55	50	50

Mean Female = 60.21

A_2: Male		
50	45	30
55	60	30
80	85	30
65	65	55
70	70	35
75	70	20
75	80	45
65	60	40

Mean Male = 56.46

We can then apply the equation for the model sum of squares that we used to calculate the overall model sum of squares:

$$\mathrm{SS}_A = \sum_{n=1}^{k} n_k(\bar{x}_k - \bar{x}_{\text{grand}})^2$$

The grand mean is the mean of all scores (above) and n is the number of scores in each group (i.e., the number of males and females; 24 in this case). Therefore, the equation becomes:

$$\begin{aligned}
SS_{\text{gender}} &= 24(60.21-58.33)^2 + 24(56.46-58.33)^2 \\
&= 24(1.88)^2 + 24(-1.87)^2 \\
&= 84.8256 + 83.9256 \\
&= 168.75
\end{aligned}$$

The degrees of freedom for this SS will be the number of groups used, k, minus 1. We used two groups (males and females) and so $df = 1$. To sum up, the main effect of gender compares the mean of all males against the mean of all females (regardless of which alcohol group they were in).

12.4.2.2. The main effect of alcohol (SS_B) ②

To work out the variance accounted for by the second independent variable (in this case, **alcohol**) we need to group the scores in the data set according to how much alcohol was consumed. So, basically we ignore the gender of the participant, and we just place all of the scores after no drinks in one group, the scores after 2 pints in another group and the scores after 4 pints in a third group. So, the data will look like this:

B_1: None		B_2: 2 Pints		B_3: 4 Pints	
65	50	70	45	55	30
70	55	65	60	65	30
60	80	60	85	70	30
60	65	70	65	55	55
60	70	65	70	55	35
55	75	60	70	60	20
60	75	60	80	50	45
55	65	50	60	50	40
Mean None = 63.75		Mean 2 pints = 64.6875		Mean 4 pints = 46.5625	

We can then apply the same equation for the model sum of squares that we used for the overall model sum of squares and for the main effect of gender:

$$SS_B = \sum_{n=1}^{k} n_k(\bar{x}_k - \bar{x}_{\text{grand}})^2$$

The grand mean is the mean of all scores (58.33 as before) and n is the number of scores in each group (i.e., the number of scores in each of the boxes above, in this case 16). Therefore, the equation becomes:

$$\begin{aligned}
SS_{\text{alcohol}} &= 16(63.75-58.33)^2 + 16(64.6875-58.33)^2 + 16(46.5625-58.33)^2 \\
&= 16(5.42)^2 + 16(6.3575)^2 + 16(-11.7675)^2 \\
&= 470.0224 + 646.6849 + 2215.5849 \\
&= 3332.292
\end{aligned}$$

The degrees of freedom for this SS will be the number of groups used, k, minus 1 (see section 10.2.6). We used three groups and so $df = 2$. To sum up, the main effect of alcohol compares the means of the no-alcohol, 2-pints and 4-pints groups (regardless of whether the scores come from men or women).

12.4.2.3. The interaction effect ($SS_{A \times B}$) ②

The final stage is to calculate how much variance is explained by the interaction of the two variables. The simplest way to do this is to remember that the SS_M is made up of three components (SS_A, SS_B and $SS_{A \times B}$). Therefore, given that we know SS_A and SS_B we can calculate the interaction term using subtraction:

$$SS_{A\times B} = SS_M - SS_A - SS_B$$

Therefore, for these data, the value is:

$$\begin{aligned} SS_{A\times B} &= SS_M - SS_A - SS_B \\ &= 5479.167 - 168.75 - 3332.292 \\ &= 1978.125 \end{aligned}$$

The degrees of freedom can be calculated in the same way, but are also the product of the degrees of freedom for the main effects (either method works):

$$\begin{aligned} df_{A\times B} &= df_M - df_A - df_B \\ &= 5 - 1 - 2 \\ &= 2 \end{aligned} \qquad \begin{aligned} df_{A\times B} &= df_A \times df_B \\ &= 1 \times 2 \\ &= 2 \end{aligned}$$

12.4.3. The residual sum of squares (SS_R) ②

The residual sum of squares is calculated in the same way as for one-way ANOVA (see section 10.2.7) and again represents individual differences in performance or the variance that can't be explained by factors that were systematically manipulated. We saw in one-way ANOVA that the value is calculated by taking the squared error between each data point and its corresponding group mean. An alternative way to express this was as (see equation (10.7)):

$$\begin{aligned} SS_R &= \sum s_k^2 (n_k - 1) \\ &= s^2_{\text{group 1}}(n_1 - 1) + s^2_{\text{group 2}}(n_2 - 1) + s^2_{\text{group 3}}(n_3 - 1) + \ldots + s^2_{\text{group } n}(n_n - 1) \end{aligned}$$

So, we use the individual variances of each group and multiply them by one less than the number of people within the group (n). We have the individual group variances in our original table of data (Table 12.1) and there were eight people in each group (therefore, $n = 8$) and so the equation becomes:

$$\begin{aligned} SS_R &= s^2_{\text{group1}}(n_1 - 1) + s^2_{\text{group2}}(n_2 - 1) + s^2_{\text{group3}}(n_3 - 1) + s^2_{\text{group4}}(n_4 - 1) + \ldots \\ &\quad + s^2_{\text{group5}}(n_5 - 1) + s^2_{\text{group6}}(n_6 - 1) \\ &= 24.55(8-1) + 106.7(8-1) + 42.86(8-1) + 156.7(8-1) + 50(8-1) + 117.41(8-1) \\ &= (24.55 \times 7) + (106.7 \times 7) + (42.86 \times 7) + (156.7 \times 7) + (50 \times 7) + (117.41 \times 7) \\ &= 171.85 + 746.9 + 300 + 1096.9 + 350 + 821.87 \\ &= 3487.52 \end{aligned}$$

The degrees of freedom for each group will be one less than the number of scores per group (i.e., 7). Therefore, if we add the degrees of freedom for each group, we get a total of $6 \times 7 = 42$.

12.4.4. The *F*-ratios ②

Each effect in a two-way ANOVA (the two main effects and the interaction) has its own *F*-ratio. To calculate these we have to first calculate the mean squares for each effect by taking the sum of squares and dividing by the respective degrees of freedom (think back to section 10.2.8). We also need the mean squares for the residual term. So, for this example we'd have four mean squares calculated as follows:

$$
\begin{aligned}
\mathrm{MS}_A &= \frac{\mathrm{SS}_A}{df_A} = \frac{168.75}{1} = 168.75 \\
\mathrm{MS}_B &= \frac{\mathrm{SS}_B}{df_B} = \frac{3332.292}{2} = 1666.146 \\
\mathrm{MS}_{A\ B} &= \frac{\mathrm{SS}_{A\ B}}{df_{A\ B}} = \frac{1978.125}{2} = 989.062 \\
\mathrm{MS}_\mathrm{R} &= \frac{\mathrm{SS}_\mathrm{R}}{df_\mathrm{R}} = \frac{3487.52}{42} = 83.036
\end{aligned}
$$

The *F*-ratios for the two independent variables and their interactions are then calculated by dividing their mean squares by the residual mean squares. Again, if you think back to one-way ANOVA this is exactly the same process.

$$
\begin{aligned}
F_A &= \frac{\mathrm{MS}_A}{\mathrm{MS}_\mathrm{R}} = \frac{168.75}{83.036} = 2.032 \\
F_B &= \frac{\mathrm{MS}_B}{\mathrm{MS}_\mathrm{R}} = \frac{1666.146}{83.036} = 20.065 \\
F_{A\ B} &= \frac{\mathrm{MS}_{A\ B}}{\mathrm{MS}_\mathrm{R}} = \frac{989.062}{83.036} = 11.911
\end{aligned}
$$

Each of these *F*-ratios can be compared against critical values (based on their degrees of freedom, which can be different for each effect) to tell us whether these effects are likely to reflect data that have arisen by chance, or reflect an effect of our experimental manipulations (these critical values can be found in the Appendix). If an observed *F* exceeds the corresponding critical values then it is significant. **R** will calculate each of these *F*-ratios and their exact significance, but what I hope to have shown you in this section is that two-way ANOVA is basically the same as one-way ANOVA except that the model sum of squares is partitioned into three parts: the effect of each of the independent variables and the effect of how these variables interact.

12.5. Factorial ANOVA using R ②

12.5.1. Packages for factorial ANOVA in R ①

If you're using commands (which we recommend), then you will need the packages *car* (for Levene's test), *compute.es* (for effect sizes), *ggplot2* (for graphs), *multcomp* (for *post* *hoc* tests), *pastecs* (for descriptive statistics), *reshape* (for reshaping the data) and *WRS* (for robust tests). If you do not have these packages installed (some should be installed from previous chapters), you can install them by executing the following commands:

```
install.packages("car"); install.packages("compute.es"); install.packages
("ggplot2");install.packages("multcomp");install.packages("pastecs");install.
packages("reshape"); install.packages("WRS", repos="http://R-Forge.R-project.
org")
```

You then need to load these packages by executing these commands:

```
library(car); library(compute.es); library(ggplot2); library(multcomp);
library(pastecs); library(reshape); library(WRS)
```

12.5.2. General procedure for factorial ANOVA ①

To conduct factorial ANOVA you should follow this general procedure:

1. *Enter data*: you've probably gathered this much by now.
2. *Explore your data*: as always, we'll begin by graphing the data and computing descriptive statistics. You should check distributional assumptions and use Levene's test to check for homogeneity of variance (see Chapter 5).
3. *Construct or choose contrasts*: you need to decide what contrasts to do and to specify them appropriately for all of the independent variables in your analysis. If you want to use Type III sums of squares, *these contrasts must be orthogonal*.
4. *Compute the ANOVA*: you can then run the main analysis of variance. Depending on what you found in the previous step, you might need to run a robust version of the test.
5. *Compute contrasts or post hoc tests*: having conducted the main ANOVA, you can follow it up with *post hoc* tests or look at the results of your contrasts. Again, the exact methods you choose will depend upon what you unearth in step 2.

We will work through these steps in turn.

12.5.3. Factorial ANOVA using R Commander ②

Running factorial ANOVA using commands gives you much more versatility than R Commander. However, you can do a basic factorial ANOVA using R Commander. First load the data from the file **goggles.csv** by using the **Data⇒Import data⇒from text file, clipboard, or URL…** menu (see section 3.7.3). Note that this file is a comma-separated (not a tab-delimited) file. This data set has three variables: **gender**, which is entered as text ('Male' and 'Female'), **alcohol**, which is also entered as text ('None', '2 Pints' and '4 Pints'), and **attractiveness**, which is the outcome variable. I have called the dataframe *gogglesData*. Note that because **gender** and **alcohol** contained text strings, rather than numbers, **R** has assumed that these variables are factors.

We can explore the data by getting some descriptive statistics and testing the assumptions. This is explained in Chapter 5. Levene's test looks at whether variances across conditions are equal. Use the **Statistics⇒Variances⇒Levene's test…** menu to run the analysis as in Chapters 5 and 10. You will need to run separate tests for **alcohol** and **gender** (as you will see, by using commands we can also run the test for the interaction of these variables).

To do the ANOVA, use the **Statistics⇒Means⇒Multi-way ANOVA...** menu. The resulting dialog box is fairly self-explanatory (Figure 12.4). You need to enter a name for the model that you're going to create (I have chosen *gogglesModel*) in the box labelled *Enter name for model*, select any factors from the list labelled *Factors* (in this case we have two factors, **alcohol** and **gender**) and select the outcome variable (in this case **attractiveness**) from the list labelled *Response Variable*. You cannot do planned comparisons or *post hoc* tests using this menu. Click on OK to run the analysis. The resulting output is described in sections 12.5.8. Note that R Commander uses Type II sums of squares when computing a factorial ANOVA, which may or may not be what you want (see Jane Superbrain Box 11.1 in the previous chapter).

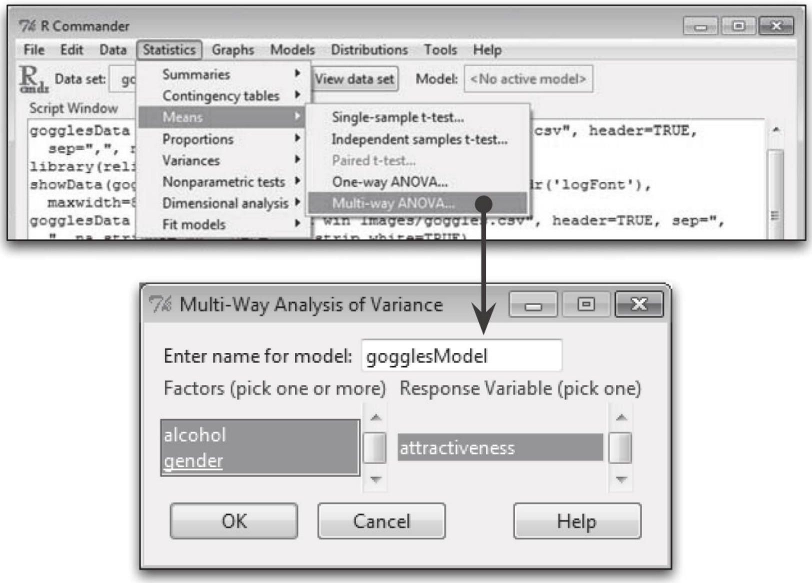


FIGURE 12.4
Factorial-way ANOVA using R Commander

12.5.4. Entering the data ②

The data for the example can be found in the file **goggles.csv**. You can load this data file by setting your working directory and executing:

```
gogglesData<-read.csv("goggles.csv", header = TRUE)
```

Note that we have used the *read.csv()* function because the data are stored in a comma-separated values file (.csv). If we look at the data in **R** we will see that levels of the between-group variables have been entered in single columns.

```
   gender alcohol attractiveness
1  Female    None             65
2  Female    None             70
3  Female    None             60
4  Female    None             60
5  Female    None             60
6  Female    None             55
7  Female    None             60
8  Female    None             55
```

```
9  Female 2 Pints          70
10 Female 2 Pints          65
11 Female 2 Pints          60
12 Female 2 Pints          70
13 Female 2 Pints          65
14 Female 2 Pints          60
15 Female 2 Pints          60
16 Female 2 Pints          50
17 Female 4 Pints          55
18 Female 4 Pints          65
19 Female 4 Pints          70
20 Female 4 Pints          55
21 Female 4 Pints          55
22 Female 4 Pints          60
23 Female 4 Pints          50
24 Female 4 Pints          50
25   Male    None          50
26   Male    None          55
27   Male    None          80
28   Male    None          65
29   Male    None          70
30   Male    None          75
31   Male    None          75
32   Male    None          65
33   Male 2 Pints          45
34   Male 2 Pints          60
35   Male 2 Pints          85
36   Male 2 Pints          65
37   Male 2 Pints          70
38   Male 2 Pints          70
39   Male 2 Pints          80
40   Male 2 Pints          60
41   Male 4 Pints          30
42   Male 4 Pints          30
43   Male 4 Pints          30
44   Male 4 Pints          55
45   Male 4 Pints          35
46   Male 4 Pints          20
47   Male 4 Pints          45
48   Male 4 Pints          40
```

These data were originally entered in Excel, and as you can see we need two different coding variables to represent gender and alcohol consumption. Therefore, in Excel, I created a variable called **gender** into which I typed 'Female' or 'Male'; because I have used words rather than numbers, when **R** imports the data it guesses that this variable is a factor (i.e., we don't need to explicitly convert it to a factor like we would had I used numbers to represent males and females). **R** will code this factor with the levels in alphabetical order (so, females will be level 1 and males level 2 of gender, which coincidentally is the same order as in the data file).

Next, I created a variable called **alcohol** and entered 'None', '2 Pints' or '4 Pints'. Again, **R** guesses that this variable is a factor when it imports the data, and organizes the levels of this variable alphabetically. The alphabetic ordering means that **R** has imported this factor with the groups ordered as '2 Pints', '4 Pints' and 'None'. This is because numbers (e.g., 2 and 4) are deemed to come before letters in the alphabet. Ideally, we might like the groups to be ordreed as they are in the data (i.e., 'None', '2 Pints' and '4 Pints'). To reorder the groups, we can use the *levels* option of the *factor()* function. All we need to do is type the levels in the order that we want them. So, by executing:

```
gogglesData$alcohol<-factor(gogglesData$alcohol, levels = c("None", "2
Pints", "4 Pints"))
```

we take the variable **alcohol** from the *gogglesData* dataframe, and we reorder the levels of the factor as 'None', '2 Pints' and '4 Pints' (*levels = c("None", "2 Pints", "4 Pints")*).

You can see from the data that there are 24 females followed by 24 males, and within these groups there are 8 people who had no alcohol, 8 who had two pints and 8 who consumed four pints. Finally, I created a variable called **attractiveness** into which I put the scores (out of 100) representing the attractiveness of the each participant's date.

If we wanted to enter the data directly into **R**, we would need to assign group codes for the **gender** and **alcohol** variables. We might code **gender** as 1 for females and 2 for males, and we might code **alcohol** as no alcohol = 1, 2 pints = 2 and 4 pints = 3. The way this coding works is as follows:

Gender	Alcohol	*Participant was*
1	1	Male who consumed no alcohol
1	2	Male who consumed 2 pints
1	3	Male who consumed 4 pints
2	1	Female who consumed no alcohol
2	2	Female who consumed 2 pints
2	3	Female who consumed 4 pints

We can create these two coding variables very quickly by using the *gl()* function (Chapter 3). Remember that this function takes the general form:

```
factor<-gl(number of levels, cases in each level, total cases, labels =
c("label1", "label2"…))
```

This function creates a factor variable called *factor*; you specify the number of levels or groups of the factor, how many cases are in each level/group, optionally the total number of cases (the default is to multiply the number of groups by the number of cases per group), and you can also use the *labels* option to list names for each level/group. For gender, we want 24 females followed by 24 males, so we can specify it as:

```
gender<-gl(2, 24, labels = c("Female", "Male"))
```

The numbers in the function tell **R** that we want 2 groups of 24 cases, the labels option then specifies the names to attach to these two groups. To create the alcohol variable we want 3 groups that each contain 8 cases. This will create 24 cases (3 × 8 = 24), or, put another way, it will create the codes for the first gender group (i.e., females). However, we want this pattern to be repeated for the second gender group also; we can do this by adding a third value to the function that is the total number of cases (i.e., 48). By specifying the total number of cases, the *gl()* function will repeat the pattern of 24 codes until it reaches this total number of cases – in other words if we specify 48 as the limit, it will repeat the pattern twice.

```
alcohol<-gl(3, 8, 48, labels = c("None", "2 Pints", "4 Pints"))
```

We can add the attractiveness values by creating a numeric variable in the usual way:

```
attractiveness<-c(65,70,60,60,60,55,60,55,70,65,60,70,65,60,
60,50,55,65,70,55,55,60,50,50,50,55,80,65,70,75,75,65,45,60,85,65,70,
70,80,60,30,30,30,55,35,20,45,40)
```

Finally, we can merge these variables into a dataframe called *gogglesData* by executing:

```
gogglesData<-data.frame(gender, alcohol, attractiveness)
```

12.5.5. Exploring the data ②

As ever, we'll look at some graphs first. Let's start with the means across the different conditions.

SELF-TEST

✓ Use *ggplot2* to plot a line graph (with error bars) of the attractiveness of the date with alcohol consumption on the *x*-axis and different-coloured lines to represent males and females.

The resulting plot (shown later in the chapter in Figure 12.8) is what is known as an **interaction graph**. These graphs are useful for interpreting significant interaction effects (should the analysis throw one up).

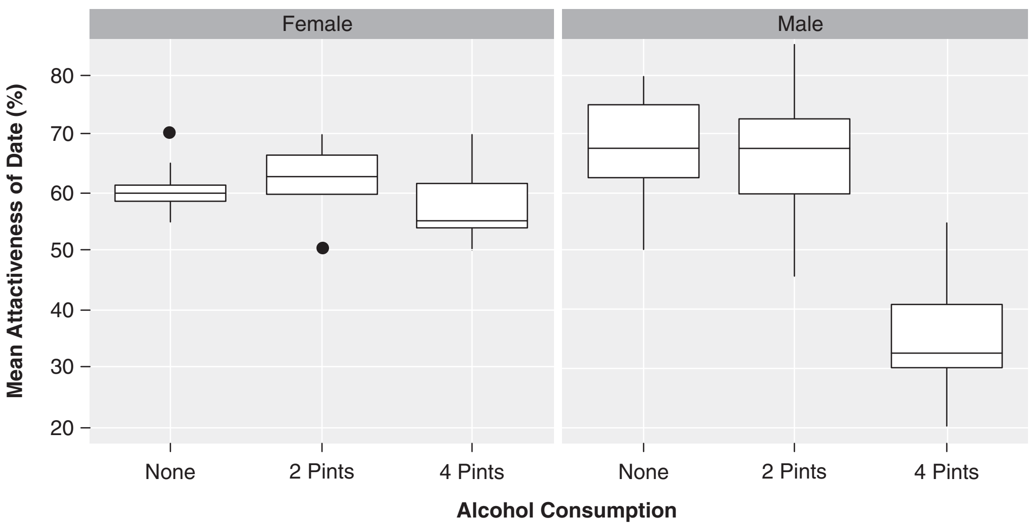
We can also look at boxplots for attractiveness scores for men and women at each level of alcohol consumption.

SELF-TEST

✓ Use *ggplot2* to plot boxplots of the attractiveness of the date at each level of alcohol consumption on the *x*-axis and different panels to represent males and females.

Figure 12.5 shows boxplots for these data. For females, the median score (the horizontal line in the middle of each box) does not change much across the doses of alcohol, and also the spread of their scores is relatively narrow; however, for males, the spread of scores is wider than for females, and the median attractiveness seems to fall dramatically after 4 pints.

FIGURE 12.5
Boxplots of the beer-goggles data



We have used the *by()* and *stat.desc()* functions before to get descriptive statistics for separate groups (see section 10.6.5 for more detail). Therefore, if we wanted to explore the effects of alcohol and gender on the attractiveness of the dates selected, we could do so by executing separate commands:

```
by(gogglesData$attractiveness, gogglesData$gender, stat.desc)
by(gogglesData$attractiveness, gogglesData$alcohol, stat.desc)
```

The resulting output is useful for interpreting the main effects of alcohol and gender on the attractiveness of mates. However, we are also interested in how these variables interact. This requires obtaining statistics for all combinations of **alcohol** and **gender**. To do this we need to use the *list()* function to create a list of variables that we can then feed into the *by()* function. If, for example, we execute *list(gogglesData$alcohol, gogglesData$gender)* we create a list (just like a shopping list) that contains the variables **alcohol** and **gender**. If we place this list within the *by()* function, then we will get descriptive statistics for all combinations of levels of the variables within the list. To see what I mean, execute:

```
by(gogglesData$attractiveness, list(gogglesData$alcohol,
gogglesData$gender), stat.desc)
```

The resulting (edited) output is in Output 12.2. Notice that the descriptive statistics are split by every combination of **gender** and **alcohol**, resulting in six different groups of information. So, for example, we can see that in the no-alcohol condition, males typically chatted up a female who was rated at about 67% on the attractiveness scale, whereas females selected a male who was rated as 61% on that scale. These means will be useful in interpreting the direction of any effects that emerge in the analysis.

```
: None
: Female
      median  mean    SE.mean CI.mean.0.95  var   std.dev   coef.var
      60.000  60.625  1.752     4.143      24.554 4.9551    0.0817
------------------------------------------------------------------
: 2 Pints
: Female
      median  mean    SE.mean CI.mean.0.95  var   std.dev   coef.var
      62.500  62.500 2.315    5.473         42.857 6.547    0.105
-------------------------------------------------------------------
: 4 Pints
: Female
      median  mean    SE.mean CI.mean.0.95  var   std.dev   coef.var
      55.000  57.500 2.500    5.912         50.000 7.071    0.123
-------------------------------------------------------------------
: None
: Male
      median  mean    SE.mean CI.mean.0.95  var   std.dev   coef.var
      67.500  66.875 3.652    8.636         106.696 10.329 0.154
-------------------------------------------------------------------
: 2 Pints
: Male
median  mean   SE.mean CI.mean.0.95  var  std.dev   coef.var        67.500
66.875    4.426   10.465         156.696 12.518    0.187
-------------------------------------------------------------------
: 4 Pints
: Male
      median  mean    SE.mean CI.mean.0.95  var   std.dev   coef.var
      32.500 35.625  3.831    9.059         117.411 10.836 0.304
```

Output 12.2

The final thing to do at this stage is to compute Levene's test (see Chapter 5 and section 10.3.1). We can again use the *leveneTest()* function from the *car* package here. If we want to do a Levene's test to see whether the variance in **attractiveness** differs across different gender and alcohol groups seperately, we can simply execute:

```
leveneTest(gogglesData$attractiveness, gogglesData$gender, center = median)
leveneTest(gogglesData$attractiveness, gogglesData$alcohol, center =
median)
```

However, as with the descriptive statistics, we're primarily interested in the interaction of these variables, so we would ideally like to know whether the variances differ across all six groups (not just the two gender groups and three alcohol groups). To do this, we can add the *interaction()* option to the *leveneTest()* function, which will compute Levene's test across any combination of groups for the variables specified within *interaction()*. In this case, we want to know whether the variances differ across all six groups that result from the combination of **gender** and **alcohol** (i.e., female_none, female_2 pints, female_4 pints, male_none, male_2 pints, male_4 pints). Therefore, we specify both variables within *interaction()*, that is, *interaction(gogglesData$alcohol, gogglesData$gender)*. The resulting command that we need to execute is therefore:

```
leveneTest(gogglesData$attractiveness,   interaction(gogglesData$alcohol,
gogglesData$gender), center = median)
```

Output 12.3 shows the results of Levene's test. We have encountered Levene's test numerous times before, so you should know that it tests whether there are any significant differences between group variances and so a non-significant result like the one we have here, $F(5, 42) = 1.425$, $p = .235$, is indicative of the assumption being met.

```
Levene's Test for Homogeneity of Variance
      Df F value Pr(>F)
group  5  1.4252 0.2351
      42
```

Output 12.3

12.5.6. Choosing contrasts ②

We saw in Chapter 10 that it's useful to follow up ANOVA with contrasts that break down the main effects and tell us where the differences between groups lie. For one-way ANOVA, we entered codes that define the contrasts we want to do. We can follow the same procedure for factorial ANOVA except that we have to define contrasts for *all* of the independent variables. One very important consideration here is that if we want to look at Type III sums of squares (see Jane Superbrain Box 11.1) then *we must use an orthogonal contrast for these sums of squares to be computed correctly*.

We encountered an orthogonal contrast in Table 10.6: the Helmert contrast. This contrast will give you what you want in many different situations; however, if it doesn't and you want to define your own contrasts then this can be done in the same way as we discussed in Chapter 10 (see Oliver Twisted).

The effect of **gender** has only two levels, so we could code an orthogonal contrast as simply −1 (females) and 1 (males). Remember that when we code contrasts anything with a positive sign is compared to anything with a negative sign, so this contrast will compare males to females.

OLIVER TWISTED

Please Sir, can I have some more … contrasts?

'This example is too similar to the one in Chapter 10', sulks Oliver as he stamps his feet on the floor. 'It smells of rotting cabbage.' I think actually, Oliver, the stench of rotting cabbage is probably because you stood your Dickensian self under a window when someone emptied his or her toilet bucket into the street. On the website I've prepared a different (slightly more complicated) example of how to specify your own contrasts to give you a bit more practice.

The effect of **alcohol** has three levels: none, 2 pints and 4 pints. The no-alcohol group is a control, so, following the advice from Chapter 10, our first contrast might compare the no-alcohol group to the remaining categories (that is, all of the groups that had some alcohol). We need a second contrast then to separate the two alcohol groups. The resulting codes are in Table 12.3; this scenario is basically the same as the Viagra data in Chapter 10 so reread that chapter if you don't understand the values in the table.

Table 12.3 Orthogonal contrasts for the **alcohol** variable

Group	$Contrast_1$	$Contrast_2$
No Alcohol	–2	0
2 Pints	1	–1
4 Pints	1	1

Setting contrasts for the two variables will also produce parameter estimates for the interaction term. So, in this case, we'll get not only a contrast comparing no alcohol to the combined effect of 2 and 4 pints, but also one that tests whether this effect is different in men and women. Similarly, contrast 2 tests whether the 2- and 4-pints groups differ, but we will also get a parameter estimate that tests whether the difference between the 2- and 4-pints groups is affected by the gender of the participant. To set the orthogonal contrasts we execute:

```
contrasts(gogglesData$alcohol)<-cbind(c(-2, 1, 1), c(0, -1, 1))
contrasts(gogglesData$gender)<-c(-1, 1)
```

The first command sets the two contrasts for **alcohol**, just as we did in Chapter 10; the second sets a single contrast for **gender**. We can check that we have set the contrast correctly by executing the name of the variable and looking at the contrast attribute:

```
> gogglesData$alcohol

attr(,"contrasts")
         [,1] [,2]
None       -2    0
2 Pints     1   -1
4 Pints     1    1
Levels: None 2 Pints 4 Pints

> gogglesData$gender
```

```
attr(,"contrasts")
       [,1]
Female   -1
Male      1
Levels: Female Male
```

Remembering that positive numbers are compared with negative and a zero means that the group is not involved at all, we can see that for alcohol we have set the first contrast to compare ‘none’ with the 2- and 4-pints groups (combined) and a second contrast that ignores the no-alcohol group and compares only the 2-pints against the 4-pints group.

12.5.7. Fitting a factorial ANOVA model ②

To create a factorial ANOVA model we can use the *aov()* function that we have used in the previous two chapters (see section 10.6.6.1). Remember that the *aov()* function is just the *lm()* function in disguise, so we can use what we learnt in Chapter 7 to add new variables into our ANOVA model. Remember, that to add a predictor, we simply write ‘+ variableName’ into the model. In the current model we wish to predict **attractiveness** scores from both **gender** and **alcohol** so our model is simply ‘attractiveness ~ gender + alcohol’, isn’t it? Actually, it’s not, because we also need to include the interaction term. To specify an interaction term we link variable names with a colon. For example, the interaction of **gender** and **alcohol** would be written in **R** as *gender:alcohol* (or indeed *alcohol:gender*, it doesn’t matter). Therefore, to specify the model including the interaction term, we could execute:

```
gogglesModel<-aov(attractiveness ~ gender + alcohol + gender:alcohol, data = gogglesData)
```

This command creates a model called *gogglesModel*, which includes the two independent variables and their interaction.

The above method is good because it makes very explicit the predictors in the model (and is a useful reminder that we’re simply using a linear model, as we have throughout the book so far). However, there is a quicker method. You can include two variables and their interactions in a model by specifying *variable1*variable2* as the predictor. Doing so will enter not just the interaction but also the effects of the individual variables as well. So, for example, this command:

```
gogglesModel<-aov(attractiveness ~ alcohol*gender, data = gogglesData)
```

does exactly the same thing as the previous command (see R’s Souls’ Tip 12.1).

We had a fairly lengthy discussion about sums of squares in the previous chapter (see Jane Superbrain Box 11.1) and I refer you back there if what I’m about to say doesn’t make any sense. If we want to look at the Type III sums of squares for the model, we need to also execute this command after we have created the model:

```
Anova(gogglesModel, type="III")
```

This takes the model that we have just created (*gogglesModel*) but, rather than displaying the Type I sums of squares (the default), it will show us the Type III sums of squares.

12.5.8. Interpreting factorial ANOVA ②

Output 12.4 tells us whether any of the independent variables have had an effect on the dependent variable. The important things to look at in the table are the significance values

R's Souls' Tip 12.1 Specifying more complex designs ②

It follows that if you have three independent variables then you can simply add the third variable into the model in the same way. For example, if we had also measured whether the **lighting** at the club was dim or bright (which would affect how well you could see your date), then we could specify the model as:

```
gogglesModel<-aov(attractiveness ~ gender*alcohol*lighting, data = gogglesData)
```

Note that we have used 'gender*alcohol*lighting' as the predictors, which will add in the three main effects but also all of the interactions between these variables.

of the independent variables. The first thing to notice is that there is a significant main effect of **alcohol** (because the significance value is less than .05). The *F*-ratio is highly significant, indicating that the amount of alcohol consumed significantly affected whom the participant would try to chat up. This means that overall, when we ignore whether the participant was male or female, the amount of alcohol influenced their mate selection. The best way to see what this means is to look at a bar chart of the average attractiveness at each level of alcohol (ignore gender completely). This graph displays the means in Output 12.2 that we calculated in section 12.4.2.2.

```
Anova Table (Type III tests)

Response: attractiveness
               Sum Sq Df   F value    Pr(>F)
(Intercept)    163333  1 1967.0251 < 2.2e-16 ***
gender            169  1    2.0323    0.1614
alcohol          3332  2   20.0654 7.649e-07 ***
gender:alcohol   1978  2   11.9113 7.987e-05 ***
Residuals        3488 42
```

Output 12.4

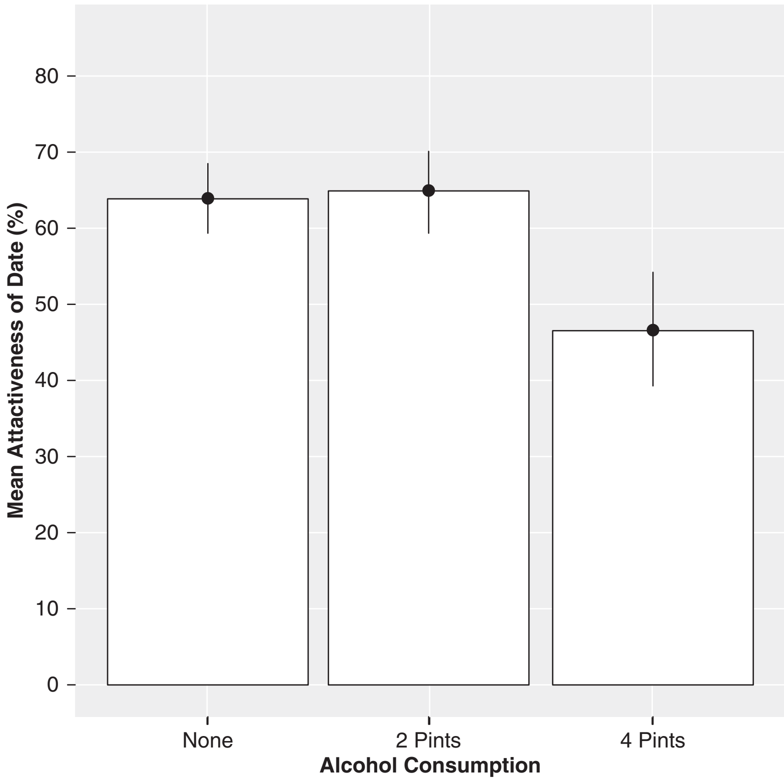
SELF-TEST

✓ Plot error bar graphs of the main effects of **alcohol** and **gender**.

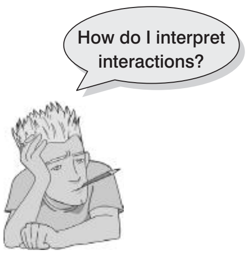
Figure 12.6 clearly shows that when you ignore gender the overall attractiveness of the selected mate is very similar when no alcohol has been drunk and when 2 pints have been drunk (the means of these groups are approximately equal). Hence, this significant main effect is *likely* to reflect the drop in the attractiveness of the selected mates when 4 pints have been drunk. This finding seems to indicate that a person is willing to accept a less attractive mate after 4 pints.

Output 12.4 also tells us about the main effect of gender. This time the *F*-ratio is not significant (p = .161, which is larger than .05). This effect means that overall, when we ignore how much alcohol had been drunk, the gender of the participant did not influence the attractiveness of the partner that the participant selected. In other words, other things being equal, males and females selected equally attractive mates. The bar chart (that you have hopefully produced from the self-test) of the average attractiveness of mates for men and

FIGURE 12.6 Graph showing the main effect of alcohol



women (ignoring how much alcohol had been consumed) reveals the meaning of this main effect. Figure 12.7 plots the means in Output 12.2 that we calculated in section 12.4.2.1. This graph shows that the average attractiveness of the partners of male and female participants was fairly similar (the means are different by only 4%). Therefore, this non-significant effect reflects the fact that the mean attractiveness was similar. We can conclude from this that, other things being equal, men and women chose equally attractive partners.



Finally, Output 12.4 tells us about the interaction between the effect of **gender** and the effect of **alcohol**. The *F*-value is highly significant (because the *p*-value is less than .05). What this actually means is that the effect of alcohol on mate selection was different for male participants than it was for females. *In the presence of this significant interaction it makes no sense to interpret the main effects*. Figure 12.8 shows the plot that we produced earlier as a self-test task; this graph tells us something about the nature of this interaction effect.

Figure 12.8 shows that for women, alcohol has very little effect: the attractiveness of their selected partners is quite stable across the three conditions (as shown by the near-horizontal line). However, for the men, the attractiveness of their partners is stable when only a small amount has been drunk, but rapidly declines when 4 pints have been drunk. Non-parallel lines usually indicate a significant interaction effect. In this particular graph the lines actually cross, which indicates a fairly large interaction between independent variables. The lines tell us that alcohol has little effect on mate selection until 4 pints have been drunk and that the effect of alcohol is prevalent only in male participants. In short, the results show that women maintain high standards in their mate selection regardless of alcohol, whereas men have a few beers and then try to get off with anything on legs. One interesting point that these data demonstrate is that we earlier concluded that alcohol significantly affected how attractive a mate was selected (the **alcohol** main effect);

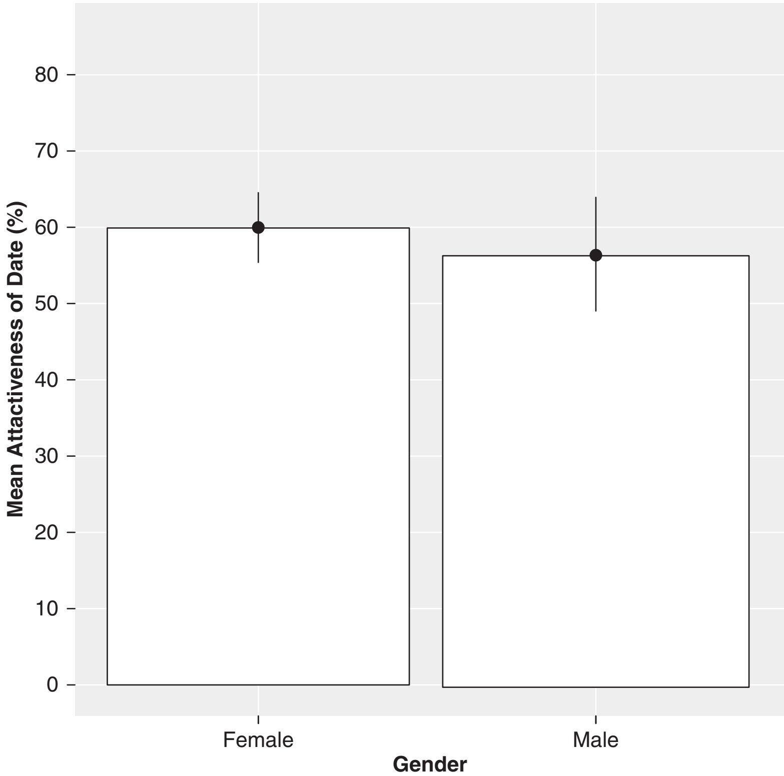


FIGURE 12.7 Graph to show the main effect of gender on mate selection

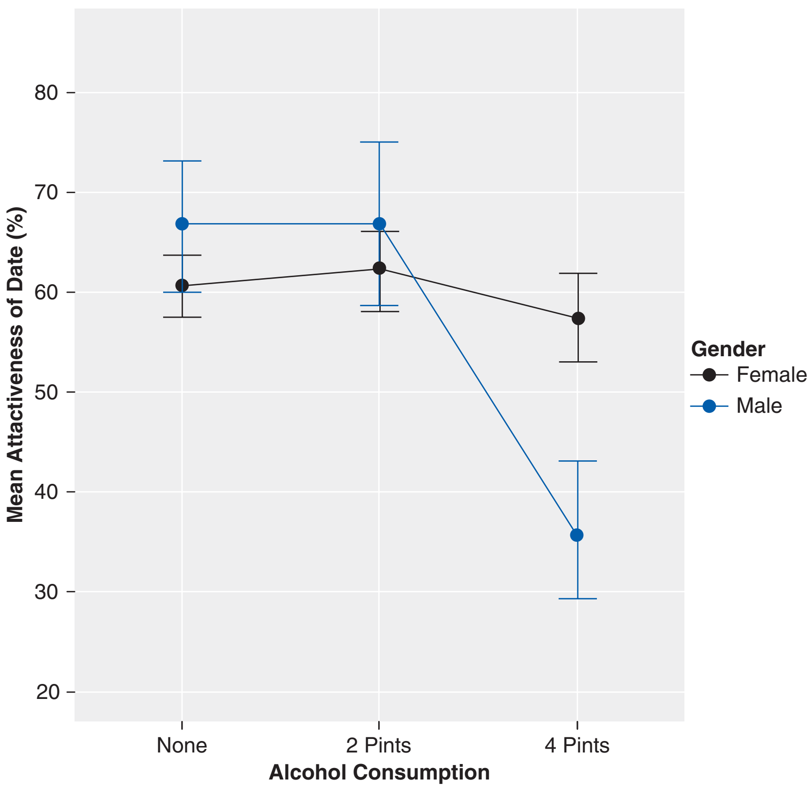


FIGURE 12.8 Graph of the interaction of gender and alcohol consumption in mate selection

however, the interaction effect tells us that this is true only in males (females appear unaffected). This shows why main effects should not be interpretted when a significant interaction involving those main effects exists.

12.5.9. Interpreting contrasts ②

To see the output for the contrasts that we specified, execute:

```
summary.lm(gogglesModel)
```

Doing so will display the parameter estimates for the model (Output 12.5). Let's look at each effect in the analysis in turn:

- **gender1**: This is the contrast for the main effect of gender; because gender has only two groups this is the same as the effect of **gender** from Output 12.4. (Quite literally, in fact: the *t*- and *F*-statistics are directly related by $F = t^2$. Our *t*-value for this contrast is –1.426, and the value of *F* for the effect of gender is $-1.426^2 = 2.03$).
- **alcohol1**: This contrast compares the no-alcohol group to the two alcohol groups. This tests whether the mean of the no-alcohol group (63.75) is different than the mean of the 2-pints and 4-pints groups combined ((64.69 + 46.56)/2 = 55.625). This is a difference of –8.125 (55.63 − 63.75). As explained in Chapter 10, the estimate for this difference is this difference divided by the number of groups involved in the contrast (–8.125/3 = –2.708). The *p*-value is .006, which is smaller than .05, indicating a significant difference. So we could conclude that the effect of alcohol is that any amount of alcohol reduces the attractiveness of the dates selected compared to when no alcohol is drunk. Of course this is misleading because, in fact, the means for the no-alcohol and 2-pints groups are fairly similar (63.75 and 64.69), so 2 pints of alcohol don't reduce the attractiveness of selected dates. The comparison is significant because it's testing the combined effect of 2 and 4 pints; 4 pints has such a drastic effect that it drags down the overall mean. This example shows why you need to be careful about how you interpret contrasts: you need to have a look at the next contrast as well.
- **alcohol2**: This contrast tests whether the mean of the 2-pints group (64.69) is different than the mean of the 4-pints group (46.56). This is a difference of –18.13 (46.56 – 64.69); as explained in Chapter 10, the estimate is this value divided by the number of groups involved in the contrast (–18.13/2 = 9.06). The *p*-value is .000, which is smaller than .05, and therefore indicates a significant difference between the groups. We can conclude that having 4 pints significantly reduced the attractiveness of selected dates compared to having only 2 pints.
- **gender1:alcohol1**: This contrast tests whether the effect of **alcohol1** described above is different in men and women. It answers the question: is the effect of alcohol compared to no alcohol on the attractiveness of dates comparable in men and women? The *p*-value is .010, which is significant, so the answer is no, the extent to which alcohol vs. no alcohol has an effect on date attractiveness is different in men and women. Figure 12.9 (left) shows what this contrast is testing. The 'Alcohol' group is the combined 2- and 4-pints groups. For the women, the difference in means between the no-alcohol group and the other groups combined is 60 – 60.625 = –0.625 (the line is flat, reflecting this small difference). For the men, the difference between the two means is 51.25 – 66.875 = –15.625 (the line for males on the graph slopes down, reflecting this decrease). This contrast tests whether –0.625 (the difference for females) is significantly different from 15.625 (the difference for males). In terms of the graph, it tests whether the lines for males and females have different slopes.
- **gender1:alcohol2**: This contrast tests whether the effect of **alcohol2** described above is different in men and women. It answers the question: is the effect of 2 pints compared to 4 pints on the attractiveness of dates comparable in men and women? The *p*-value is .000, which is significant, so the answer is no, the extent to which 2 vs. 4

pints has an effect on date attractiveness is different in men and women). Figure 12.9 (right) shows what this contrast is testing. For the women, the difference in means between the 2- and 4-pints groups is 57.50 – 62.50 = –5 (the line slopes down slightly). For the men, the difference between the two means is 35.625 – 66.875 = –31.25 (the line for males on the graph slopes down much more than for females). This contrast tests whether –5 (the difference for females) is significantly different from –31.25 (the difference for males). In terms of the graph, it tests whether the lines for males and females have different slopes.

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          58.333      1.315  44.351  < 2e-16 ***
gender1              -1.875      1.315  -1.426 0.161382
alcohol1             -2.708      0.930  -2.912 0.005727 **
alcohol2             -9.062      1.611  -5.626 1.37e-06 ***
gender1:alcohol1     -2.500      0.930  -2.688 0.010258 *
gender1:alcohol2     -6.562      1.611  -4.074 0.000201 ***
```

Output 12.5

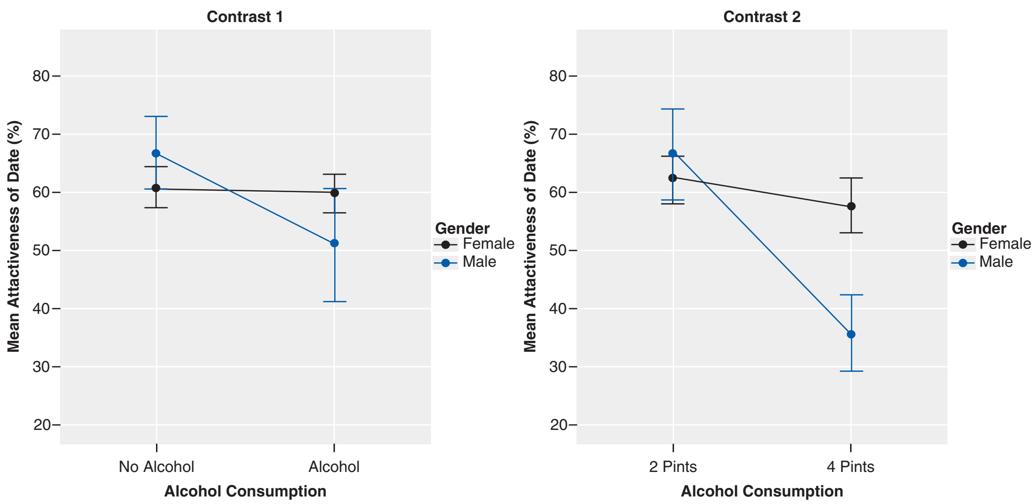


FIGURE 12.9 Graphical displays of the contrasts for the beer-goggles data

12.5.10. Simple effects analysis ③

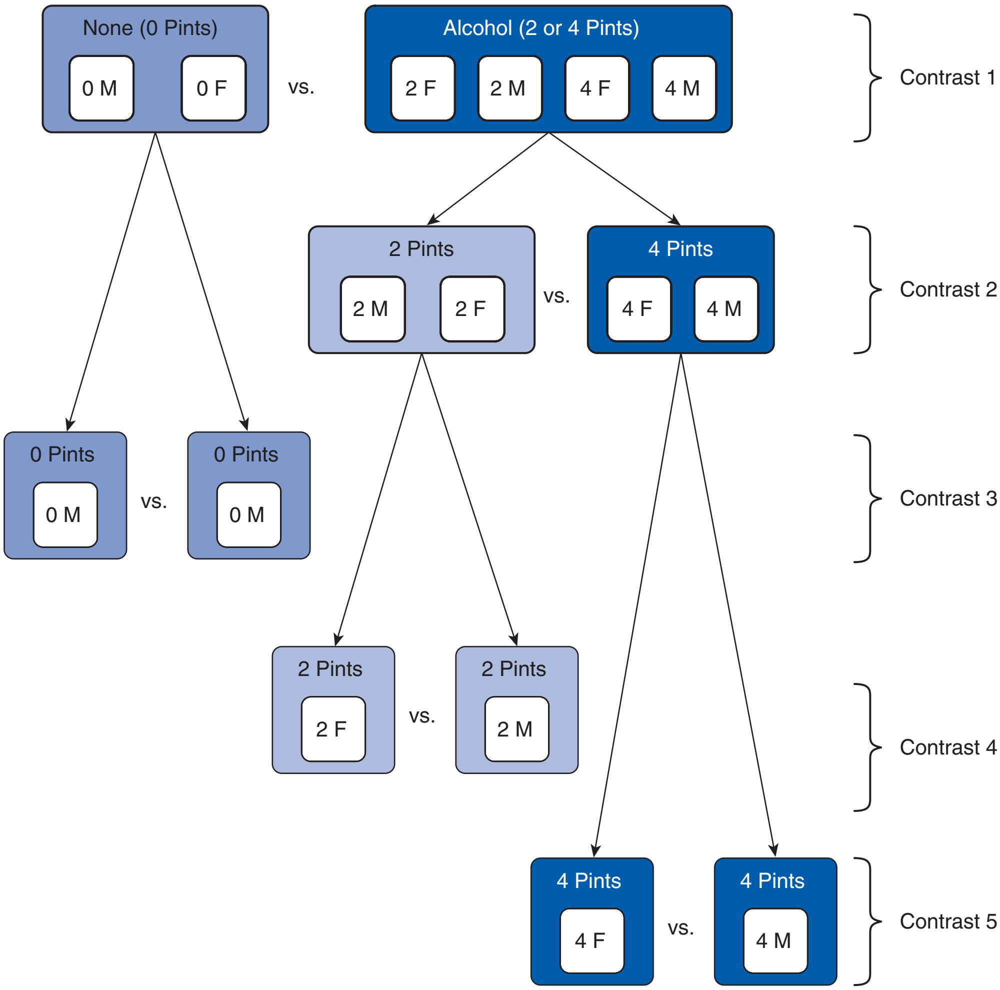
A popular way to break down an interaction term is to use a technique called **simple effects analysis**. This analysis looks at the effect of one independent variable at individual levels of the other independent variable. So, for example, in our beer-goggles data we could do simple effects analysis looking at the effect of gender at each level of alcohol. This would mean taking the average attractiveness of the date selected by men and comparing it to that for women after no drinks, then making the same comparison for 2 pints and then finally for 4 pints. Another way of looking at this is to say we would compare each black dot to the corresponding blue dot in Figure 12.8: based on the graph, we might expect to find no difference after no alcohol and after 2 pints (in both cases the black and blue dots are located in about the same position) but we would expect a difference after 4 pints (because the black and blue dots are quite far apart). The alternative way to do it would be to compare the mean attractiveness after no alcohol, 2 pints and 4 pints for men and then in a separate analysis do the same but for women. (This would be a bit like doing a one-way ANOVA on the effect of alcohol in men, and then doing a different one-way ANOVA for the effect of alcohol in women.)

OLIVER TWISTED

Please Sir, can I have some more … simple effects?

'I want to impress my friends by doing a simple effects analysis by hand', boasts Oliver. You don't really need to know how simple effects analyses are calculated to run them, Oliver, but seeing as you asked, it is explained in the additional material available from the companion website.

FIGURE 12.10 Schematic representation of the contrasts and codes for simple effects analysis on the goggles data



		0 M	0 F	2 F	2 M	4 F	4 M
	Contrast 1	–2	–2	1	1	1	1
	Contrast 2	0	0	–1	–1	1	1
Simple effect of Gender	Contrast 3	–1	1	0	0	0	0
	Contrast 4	0	0	1	–1	0	0
	Contrast 5	0	0	0	0	–1	1

R's Souls' Tip 12.2 Simple effects analysis in R ③

Unfortunately, simple effects are not that easy to do in **R**. The first thing we need to do is create a variable in the dataframe that merges the variables of interest into a single factor. In other words, rather than have **alcohol** and **gender** as separate variables, we want a new variable that simply codes the six groups that result from combining all levels of **alcohol** and **gender**. We can do this using the *gl()* function to add a variable (*simple*) to the dataframe that is six groups each containing eight observations:

```
gogglesData$simple<-gl(6,8)
```

We can then use the *factor()* function to specify labels for these six groups:

```
gogglesData$simple<-factor(gogglesData$simple, levels = c(1:6), labels = c("F_None","F_2pints", "F_4pints","M_None","M_2pints", "M_4pints"))
```

The data now look like this (I've edited out cases to save space):

```
   gender alcohol   alcohol2 attractiveness   simple
1  Female    None No Alcohol             65   F_None
2  Female    None No Alcohol             70   F_None
…     …       …        …                  …       …
9  Female 2 Pints    Alcohol             70 F_2pints
10 Female 2 Pints    Alcohol             65 F_2pints
…     …       …        …                  …       …
17 Female 4 Pints    Alcohol             55 F_4pints
18 Female 4 Pints    Alcohol             65 F_4pints
…     …       …        …                  …       …
25   Male    None No Alcohol             50   M_None
26   Male    None No Alcohol             55   M_None
…     …       …        …                  …       …
33   Male 2 Pints    Alcohol             45 M_2pints
34   Male 2 Pints    Alcohol             60 M_2pints
…     …       …        …                  …       …
47   Male 4 Pints    Alcohol             45 M_4pints
48   Male 4 Pints    Alcohol             40 M_4pints
```

Note that we have added the variable *simple*, which codes whether a person was male or female and how much alcohol they had in a single variable.

Next, we create contrasts that break these six groups up using the standard rules for planned contrasts. Figure 12.10 shows how we would break the groups up into five contrasts to do a simple effects analysis of **gender**. The first contrast compares no alcohol to alcohol (2 or 4 pints combined). Remember that these two 'chunks' of variation are made up of the different gender groups and so need to be broken down further. For example, the no-alcohol group is made up of the males that had no alcohol ('0 M') and the females that had no alcohol ('0 F'), and the alcohol chunk contains the males and females that had 2 pints ('2 M' and '2 F') and the males and females that had 4 pints ('4 M' and '4 F'). The second contrast breaks down the 'alcohol' chunk to compare 2 pints against 4 pints. Again, remember that both chunks at this stage are made up of the two corresponding gender groups. The third contrast takes the no-alcohol 'chunk' and compares the two gender groups contained within it. This contrast is the simple effect of gender when no alcohol was consumed. The fourth contrast takes the 2-pint 'chunk' and breaks the variance down to compare the two gender groups contained within it. This contrast is the simple effect of gender when 2 pints were consumed. Finally, the fifth contrast takes the 4-pint 'chunk' and compares the two gender groups contained within it. This contrast is the simple effect of gender

(Continued)

(Continued)

when 4 pints were consumed. If you look back to Chapter 10 you'll see that these contrasts conform to the rules of orthogonal contrasts, and that the codes in Figure 12.10 specify the contrasts.

To create these contrasts in **R** we can create five variables (one for each contrast) that contain the codes for the respective groups. (Bear in mind that in the dataframe the groups are ordered as: female none, female 2 pints, female 4 pints, male none, male 2 pints, male 4 pints, and we have to order the codes accordingly.) I have also labelled the contrasts in a way that tells us something about what they represent:

```
alcEffect1<-c(-2, 1, 1, -2, 1, 1)
alcEffect2<-c(0, -1, 1, 0, -1, 1)
gender_none<-c(-1, 0, 0, 1, 0, 0)
gender_twoPint<-c(0, -1, 0, 0, 1, 0)
gender_fourPint<-c(0, 0, -1, 0, 0, 1)
```

To tidy things up lets merge these variables into an object called *simpleEff*:

```
simpleEff<-cbind(alcEffect1, alcEffect2, gender_none, gender_twoPint, gender_fourPint)
```

We can now set the contrasts for the variable **simple** to be this object:

```
contrasts(gogglesData$simple)<-simpleEff
```

We then create a new model in which **attractiveness** is predicted from **simple** (which, remember, contains both the effects of alcohol and gender but coded so that the contrasts give us simple effects):

```
simpleEffectModel<-aov(attractiveness ~ simple, data = gogglesData)
```

To see the contrasts we use *summary.lm()* on the newly created model:

```
summary.lm(simpleEffectModel)
```

The resulting output contains the parameter estimates for the five contrasts. Looking at the significance values for each simple effect, it appears that there was no significant difference between men and women when they drank no alcohol, p = .177, or when they drank 2 pints, p = .34, but there was a very significant difference, p < .001, when 4 pints were consumed (which, judging from the interaction graph, reflects the fact that the mean for men is considerably lower than for women).

```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)              58.333      1.315  44.351  < 2e-16 ***
simplealcEffect1         -2.708      0.930  -2.912  0.00573 **
simplealcEffect2         -9.062      1.611  -5.626 1.37e-06 ***
simplegender_none         3.125      2.278   1.372  0.17742
simplegender_twoPint      2.188      2.278   0.960  0.34243
simplegender_fourPint   -10.938      2.278  -4.801 2.02e-05 ***
```

12.5.11. *Post hoc* analysis ②

The variable **alcohol** has three levels and so you might want to perform *post hoc* tests to see where the differences between groups lie. I want to stress again that the significant main effect of alcohol that we observed should not be interpreted given the significant interaction with gender. Therefore, I'm covering *post hoc* tests here for illustrative purposes: if this was a real piece of research I would focus on the interaction effect and not perform *post hoc* tests on **alcohol**.

We saw in Chapter 10 that we can specify Bonferroni *post hoc* tests using the *pairwise.t.test()* function and Tukey tests using *glht()*. Refer back to that chapter for details of these functions, but for the present example we could obtain *post hoc* tests for alcohol by executing either of these commands:

```
pairwise.t.test(gogglesData$attractiveness, gogglesData$alcohol, p.adjust.
method = "bonferroni")
postHocs<-glht(gogglesModel, linfct = mcp(alcohol = "Tukey"))
summary(postHocs)
confint(postHocs)
```

The resulting *post hoc* tests are shown in Outputs 12.6 (Bonferroni) and 12.7 (Tukey); they both break down the main effect of **alcohol** and can be interpreted as if a one-way ANOVA had been conducted on the **alcohol** variable (i.e., the reported effects for alcohol are collapsed with regard to gender). The Bonferroni and Tukey tests show the same pattern of results: when participants had drunk no alcohol or 2 pints of alcohol, they selected equally attractive mates. However, after 4 pints had been consumed, participants selected significantly less attractive mates than after both 2 pints ($p < .001$) and no alcohol ($p <$ .001). It is interesting to note that the mean attractiveness of partners after no alcohol and 2 pints was so similar that the probability of the obtained difference between those means is 1 (i.e., completely probable).

```
       Pairwise comparisons using t tests with pooled SD
data:  gogglesData$attractiveness and gogglesData$alcohol

        None    2 Pints
2 Pints 1.00000 -
4 Pints 0.00024 0.00011

P value adjustment method: bonferroni
```

Output 12.6

```
        Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = attractiveness ~ gender + alcohol + gender:alcohol,
    data = gogglesData)

Linear Hypotheses:
                       Estimate Std. Error t value Pr(>|t|)
2 Pints - None == 0      0.9375     3.2217   0.291    0.954
4 Pints - None == 0    -17.1875     3.2217  -5.335 1.01e-05 ***
4 Pints - 2 Pints == 0 -18.1250     3.2217  -5.626  < 1e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

> confint(postHocs)

        Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = attractiveness ~ gender + alcohol + gender:alcohol,
    data = gogglesData)
```

```
Quantile = 2.4303
95% family-wise confidence level


Linear Hypotheses:
                       Estimate lwr      upr
2 Pints - None == 0      0.9375  -6.8921   8.7671
4 Pints - None == 0    -17.1875 -25.0171  -9.3579
4 Pints - 2 Pints == 0 -18.1250 -25.9546 -10.2954
```

Output 12.7

12.5.12. Overall conclusions ②

In summary, we should conclude that alcohol has an effect on the attractiveness of selected mates. Overall, after a relatively small dose of alcohol (2 pints) humans are still in control of their judgements and the attractiveness levels of chosen partners are consistent with a control group (no alcohol consumed). However, after a greater dose of alcohol, the attractiveness of chosen mates decreases significantly. This effect is what is referred to as the 'beer-goggles effect'. More interesting, the interaction shows a gender difference in the beer-goggles effect. Specifically, it looks as though men are significantly more likely to pick less attractive mates when drunk. Women, in comparison, manage to maintain their standards despite being drunk. What we still don't know is whether women will become susceptible to the beer-goggles effect at higher doses of alcohol.

12.5.13. Plots in factorial ANOVA ②

We saw in the previous two chapters that the *aov()* function automatically generates some plots that we can use to test the assumptions. We can see these graphs by executing:

```
plot(gogglesModel)
```

The results are in Figure 12.11. The first graph (on the left) can be used for testing homogeneity of variance: if it has a funnel shape then we're in trouble. The plot we have does show funnelling (the spread of scores is wider at some points than at others), which implies that the residuals might be heteroscedastic (a bad thing). The second plot (on the right) is a Q-Q plot (see Chapter 5), which tells us about the normality of residuals in the model. We want our residuals to be normally distributed, which means that the dots on the graph should hover around the diagonal line. On our plot this is the case, suggesting that we can assume normality of our residuals/errors.

12.6. Interpreting interaction graphs ②

Interactions are very important, and the key to understanding them is being able to interpret interaction graphs. We've already had a look at one interaction graph when we interpreted the analysis in this chapter. We used Figure 12.8 to conclude that the interaction probably reflected the fact that men and women chose equally attractive dates after no alcohol and 2 pints, but that at 4 pints men's standards dropped significantly more than

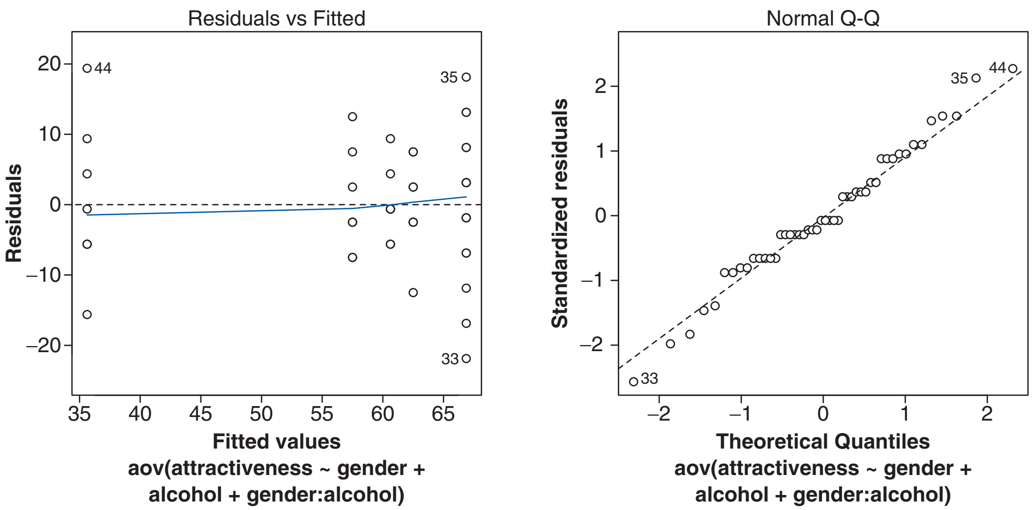


FIGURE 12.11 Plots of the beer-goggles model

CRAMMING SAM'S TIPS — Two-way independent ANOVA

- Two-way independent ANOVA compares several means when there are two independent variables and different participants have been used in all experimental conditions. For example, if you wanted to know whether different teaching methods worked better for different subjects, you could take students from four courses (Psychology, Geography, Management and Statistics) and assign them to either lecture-based or book-based teaching. The two variables are course and method of teaching. The outcome might be the end of year mark (as a percentage).
- Test for homogeneity of variance using *Levene's test*. If the *p*-value is less than .05 then the assumption is violated.
- A 'main effect' is the effect of a variable in isolation, whereas an 'interaction' represents the combined effect of two or more variables.
- In the main analysis you'll get a summary table containing a main effect of each predictor variable and an effect of the interaction between the two variables; if the *p*-value is less than .05 then the effect is significant. For main effects consult *post hoc* tests to see which groups differ, and for the interaction look at contrasts, an interaction graph or conduct simple effects analysis. If the interaction effect is significant it makes little sense to interpret or do further analysis on the main effects.
- For *post hoc* tests, look at the *p*-value of each test to discover if your comparisons are significant (they will be if the significance value is less than .05).
- Test the same assumptions as for one-way independent ANOVA (see Chapter 10).

women's. Imagine we'd got the profile of results shown in Figure 12.12; do you think we would've still got a significant interaction effect?

This profile of data probably would also give rise to a significant interaction term because, although the attractiveness of men and women's dates is similar after no alcohol and 4 pints of alcohol, there is a big difference after 2 pints. This reflects a scenario in which the beer-goggles effect is equally big in men and women after 4 pints (and doesn't exist after no alcohol) but kicks in quicker for men: the attractiveness of their dates plummets after 2 pints, whereas women maintain their standards until 4 pints (at which point they'd happily date an unwashed skunk). Let's try another example. Is there a significant interaction in Figure 12.13?

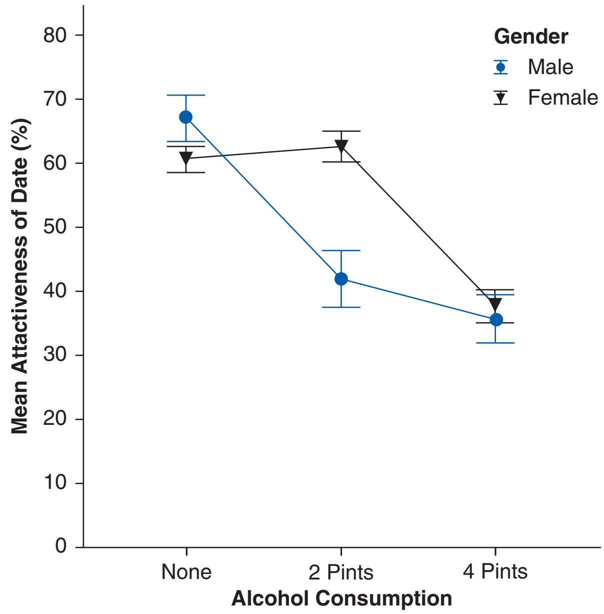


FIGURE 12.12
Another interaction graph

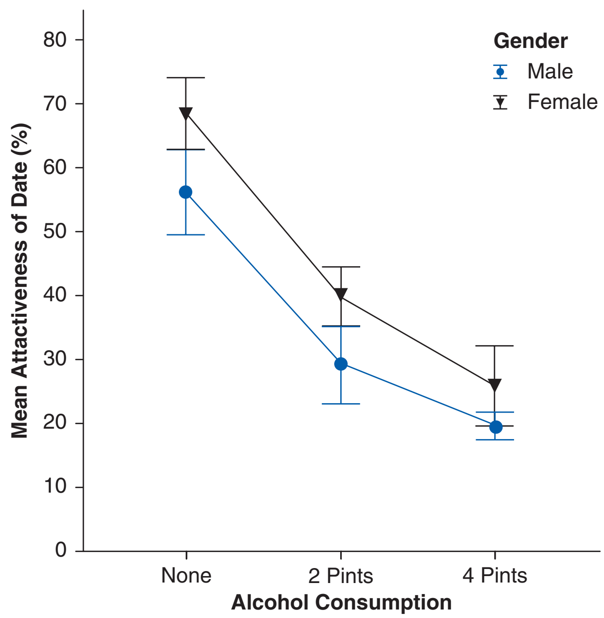


FIGURE 12.13
A 'lack of' interaction graph

For the data in Figure 12.13 there is unlikely to be a significant interaction because the effect of alcohol is the same for men and women. So, for both men and women, the attractiveness of their dates after no alcohol is quite high, but after 2 pints all types drop by a similar amount (the slope of the male and female lines is about the same). After 4 pints there is a further drop and, again, this drop is about the same in men and women (the lines again slope at about the same angle). The fact that the line for males is lower than for females just reflects the fact that across all conditions, men have lower standards than their female counterparts: this reflects a main effect of gender (i.e., males generally chose less attractive dates than females at all levels of alcohol). There are two general points that we can make from these examples:

- Non-parallel lines on an interaction graph imply significant interactions. However, it's important to remember that this doesn't mean that non-parallel lines automatically mean that the interaction is significant: whether the interaction is significant will depend on the degree to which the lines are not parallel.
- If the lines on an interaction graph cross then obviously they are not parallel and this can give away a possible significant interaction. However, contrary to popular belief, it isn't *always* the case that if the lines of the interaction graph cross then the interaction is significant.

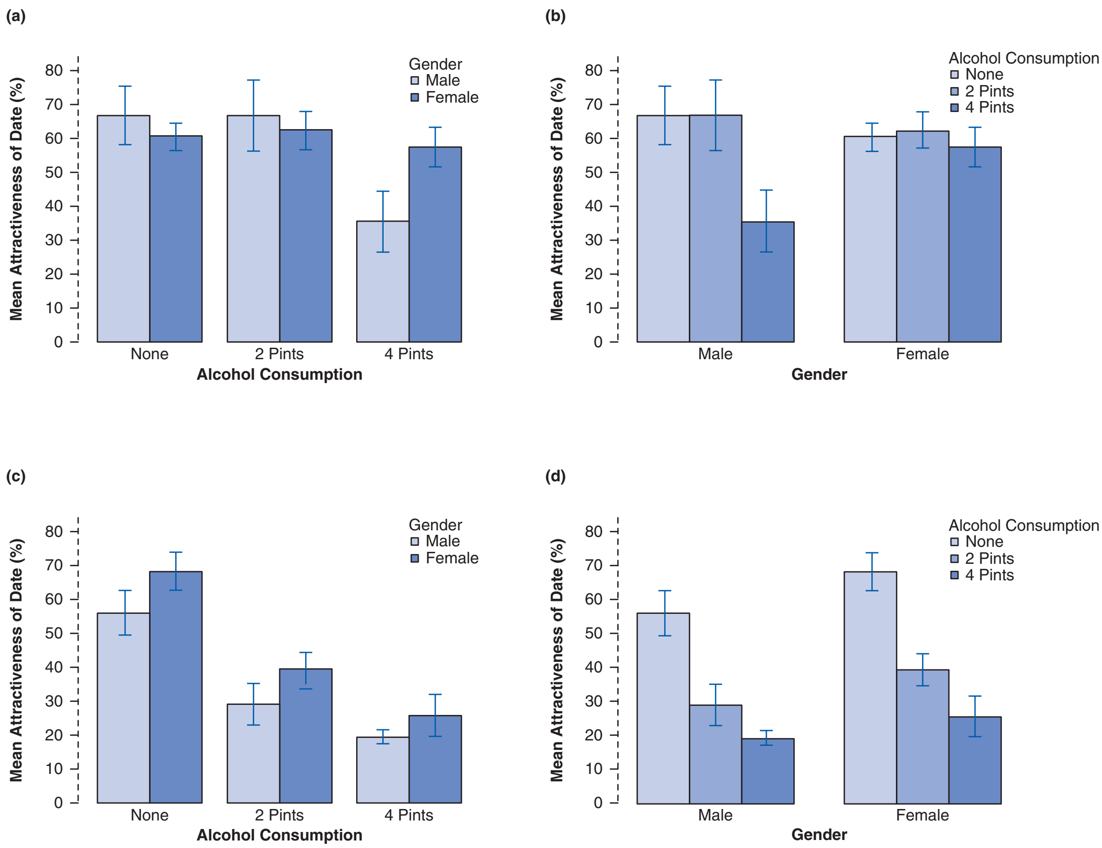


FIGURE 12.14 Bar charts showing interactions between two variables

A further complication is that sometimes people draw bar charts rather than line charts. Figure 12.14 shows some bar charts of interactions between two independent variables. Panels (a) and (b) actually display the data from the example used in this chapter (in fact, why not have a go at plotting them). As you can see, there are two ways to present the same data: panel (a) shows the data when levels of alcohol are placed along the x-axis and different-coloured bars are used to show means for males and females, and panel (b) shows the opposite scenario in which gender is plotted on the x-axis and different colours distinguish the dose of alcohol. Both of these graphs show an interaction effect.

What you're looking for is for the differences between coloured bars to be different at different points along the *x*-axis. So, for panel (a) you'd look at the difference between the light and dark blue bars for no alcohol, and then look to 2 pints and ask, 'Is the difference between the bars different than when I looked at no alcohol?' In this case the dark and light blue bars look the same at no alcohol as they do at 2 pints – hence, no interaction. However, we'd then move on to look at 4 pints, and we'd again ask, 'Is the difference between the light and dark blue bars different than what it has been in any of the other conditions?' In this case the answer is yes: for no alcohol and 2 pints, the light and dark blue bars were about the same height, but at 4 pints the dark blue bar is much higher than the light one. This shows an interaction: the pattern of responses changes at 4 pints. Panel (b) shows the same thing but plotted the other way around. Again we look at the pattern of responses. So, first we look at the men and see that the pattern is that the first two bars are the same height, but the last bar is much shorter. The interaction effect is shown up by the fact that for the women there is a different pattern: all three bars are about the same height.

SELF-TEST

✓ What about panels (c) and (d): do you think there is an interaction?

Again, they display the same data in two different ways, but it's different data than what we've used in this chapter. First let's look at panel (c): for the no-alcohol data, the dark bar is a little bit bigger than the light one; moving on to the 2-pints data, the dark bar is also a little bit taller than the light bar; and finally for the 4-pints data the dark bar is again higher than the light one. In all conditions the same pattern is shown – the dark blue bar is a bit higher than the light blue one (i.e., females pick more attractive dates than men regardless of alcohol consumption) – therefore, there is no interaction. Looking at panel (d), we see a similar result. For men, the pattern is that attractiveness ratings fall as more alcohol is drunk (the bars decrease in height) and then for the women we see the same pattern: ratings fall as more is drunk. This again is indicative of no interaction: the change in attractiveness due to alcohol is similar in men and women.

12.7. Robust factorial ANOVA ③

As with one-way ANOVA, Wilcox (2005) describes robust procedures for conducting factorial ANOVA. To access these we need to load the *WRS* package (see section 5.8.4.). There are four functions that we will look at:

- **t2way()**: This performs a two-way independent ANOVA on trimmed means.
- **mcp2atm()**: This performs *post hoc* tests for a two-way independent design based on trimmed means.
- **pbad2way()**: This performs a two-way independent ANOVA using M-measures of location (e.g., the median) and a bootstrap.
- **mcp2a()**: This performs *post hoc* tests for the above function.

The first problem we have is that these functions need the data to be in wide format rather than long (see Chapter 3). Figure 12.15 shows the existing data format (long) and how we need it to look (wide). Essentially we want levels of our two factors to be represented in different columns. Therefore, rather than a dataframe with three columns and 48 rows, we want one with six columns and eight rows.

We could re-enter the data in the wide format (which is very tempting when you've spent half an hour trying to work out how to get **R** to restructure it for you), but we're going to look at how to use *melt()* and *cast()* to do the restructuring for us. To get the restructuring to work, we need to add a variable to our dataframe that identifies the rows in the wide format. Notice in Figure 12.15 that the data are made up of six chunks that represent the combinations of **gender** and **alcohol**, and each chunk contains eight rows. We want to move these chunks from being stacked on top of each other to being beside each other. To do this, **R** needs to know what row a particular score will end up in when we move each block of scores from the stack into the columns. The easiest approach is simply to create a variable (called **row**) that identifies within each chunk the row number of a given score. In other words, it will be a value from 1 to 8 telling us whether the score is the first, second, third, etc. score within the chunk. At the moment, the chunks are stacked on top of each other so we want a variable that is the sequence of numbers 1 to 8 repeated for each of the six chunks. We can add this variable to the dataframe by executing:

```
gogglesData$row<-rep(1:8, 6)
```

This command uses the *rep()* function to create a variable **row** in the dataframe *gogglesData*, that is, the numbers 1 to 8 repeated six times (*rep(1:8, 6)*). The dataframe now looks like this (edited):

```
   gender alcohol attractiveness row
1  Female    None             65   1
2  Female    None             70   2
3  Female    None             60   3
4  Female    None             60   4
5  Female    None             60   5
6  Female    None             55   6
7  Female    None             60   7
8  Female    None             55   8
9  Female  2 Pints            70   1
10 Female  2 Pints            65   2
11 Female  2 Pints            60   3
12 Female  2 Pints            70   4
13 Female  2 Pints            65   5
14 Female  2 Pints            60   6
15 Female  2 Pints            60   7
16 Female  2 Pints            50   8
```

Note that the structure is the same as before – it's just that we have a new variable called **row** that identifies the scores within each combination of **gender** and **alcohol** as a value from 1 to 8.

Now we have changed the data set we need to make it molten so that we can cast the data into the wide format. To do this we use the *melt()* function (see section 3.9.4). Remember that in this function we differentiate variables that identify attributes of the scores (in this case, **gender**, **alcohol**, and **row** all tell us about a given attractiveness score, for example, that it was the first score in the male group who drank 2 pints) from the scores or measured variables themselves. Attributes are specified with the *id* option, and scores with the *measured* option. Therefore, we can create a molten dataframe called *gogglesMelt* by executing:

```
gogglesMelt<-melt(gogglesData, id = c("row", "gender", "alcohol"), measured = 
c("attractiveness"))
```

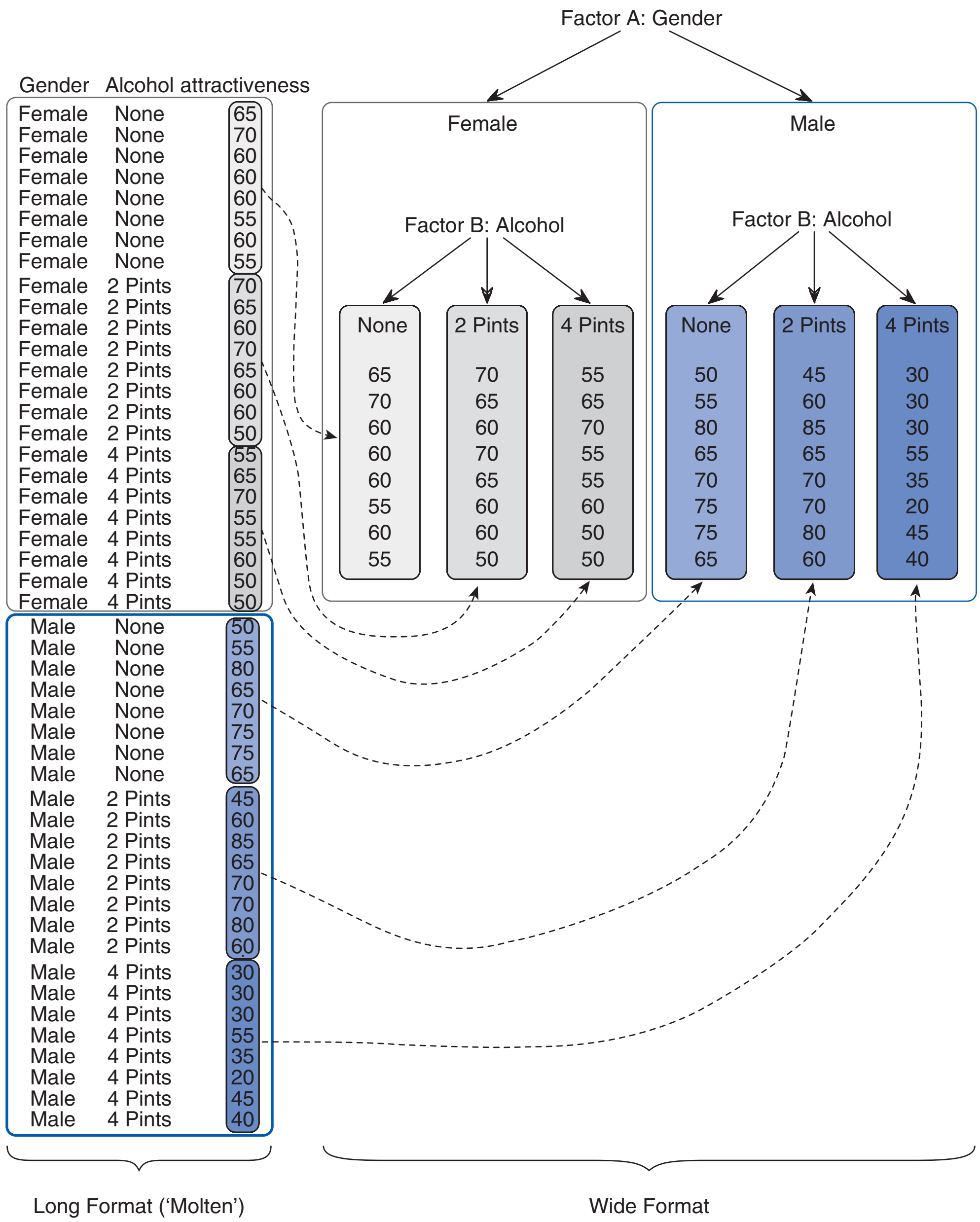


FIGURE 12.15 Restructuring the data for robust factorial ANOVA

Having melted the data, we want to cast it in the wide format using *cast()*. To do this we use a formula in the form: *variables specifying the rows ~ variables specifying the columns*. In this case, **row** tells us in which row to place a score, and we want the alcohol and gender variables split across different columns, so we'd use the formula: *row ~ gender + alcohol*. Therefore, we can make a wide dataframe called *gogglesWide* by executing:

```
gogglesWide<-cast(gogglesMelt,  row ~ gender + alcohol)
```

Note that we have applied this command to the molten data set (*gogglesMelt*). The result is that the data have been transformed from the long format to the wide format. However, because we added the variable **row** to the dataframe, our new dataframe also contains this variable, and for the analysis we want only the **alcohol** and **gender** variables, therefore, we want to remove **row**. We can do this by executing:

```
gogglesWide$row<-NULL
```

which basically zaps the variable **row** into oblivion. If you look at the dataframe you'll see a lovely wide format set of data:

```
gogglesWide
```

```
F_None   F_2 Pints   F_4 Pints   M_None   M_2 Pints   M_4 Pints
65         70          55          50         45          30
70         65          65          55         60          30
60         60          70          80         85          30
60         70          55          65         65          55
60         65          55          70         70          35
55         60          60          75         70          20
60         60          50          75         80          45
55         50          50          65         60          40
```

It's important to note the order of the columns because this affects how we specify the robust analysis. In this case, the hierarchy of the independent variables is **gender** followed by **alcohol**. In other words, we have taken the six groups and first divided them into male and female, then within the male and female groups we have subdivided according to the amount of alcohol they drank. We would say that **gender** is factor A and **alcohol** factor B. If this idea is not clear then Figure 12.15 might help you to visualize it. As such, the order of the columns reflects a 2 × 3 design (2 levels of gender divided up into 3 levels of alcohol). If the columns were ordered as F_None, M_None, F 2 Pints, M 2 Pints, F 4 Pints, M 4 Pints, then we would have a 3 × 2 design (3 levels of alcohol each divided up into 2 levels of gender). In this case factor A would be **alcohol** and factor B **gender**.

The function *t2way()* takes the general form:

```
t2way(levels of factor A, levels of factor B, data, tr = .2, alpha = .05)
```

As with other functions we've encountered, the level of trimming is by default 20% (*tr* = .2) but can be changed by including the *tr* = option. Also the default alpha level is .05 but can be changed by including the *alpha* = option. Assuming we are happy with the default level of trimming, we need only specify the dataframe (*gogglesWide*) and the levels of factor A (2 in this case as explained above) and factor B (3 in this case). Therefore, we can do a robust two-way factorial ANOVA based on trimmed means by executing:

```
t2way(2,3, gogglesWide)
```

The function *pbad2way()* has a similar format:

```
pbad2way(levels of factor A, levels of factor B, data, est = mom, nboot = 2000)
```

The main differences are an option to control the number of bootstrap samples (*nboot*), although the default of 2000 is fine, and an option *est* to control the M-estimator that you want to use. You can use *est* = *median* (to use the median) or *est* = *mom* (to use a method based on identifying and removing outliers). In smaller samples you might find that *est* = *mom* throws up an error message, in which case switch to *est* = *median*. If we're happy with 2000 bootstrap samples and using *mom* rather than *median* then we can run the analysis for the current data by executing:[2]

```
pbad2way(2,3, gogglesWide)
```

The output of both of these commands is shown in Output 12.8. For *t2way()* (left-hand side of Output 12.8) we are given a test statistic for factor A (*$Qa*), factor B (*$Qb*) and their interaction (*$Qab*) as well as the corresponding *p*-value (*$A.p.value*, *$B.p.value*, and

[2] If you want to compare medians then execute:

```
pbad2way(2,3, gogglesWide, est = median)
```

$AB.p.value respectively). Remember that factor A was gender and factor B alcohol; therefore, we could conclude that there was no significant main effect of gender, $Q = 1.67$, $p = .209$, but there was a significant main effect of alcohol, $Q = 48.28$, $p = .001$, and a significant gender × alcohol interaction, $Q = 26.26$, $p = .001$. The bottom of the output shows the trimmed means on which these results are based: factor A (gender) is represented by rows, and factor B (alcohol) by columns. So, for example, the trimmed mean of the attractiveness score for females who drank 2 pints was 63.3.

The output of *pbad2way()* (right-hand side of Output 12.8) tells us much the same things but we get only *p*-values and no test statistics: there was no significant main effect of gender, $p = .171$, but there was a significant main effect of alcohol, $p < .001$, and a significant gender × alcohol interaction, $p < .001$.

t2way()	*pbad2way()*

`$Qa`
`[1] 1.666667`
 	
`$A.p.value`
`[1] 0.209`
 	
`$Qb`
`[1] 48.2845`
 	
`$B.p.value`
`[1] 0.001`
 	
`$Qab`
`[1] 26.25718`
 	
`$AB.p.value`
`[1] 0.001`
 	
`$means`
`     [,1]     [,2]     [,3]`
`[1,] 60.0 63.33333 56.66667`
`[2,] 67.5 67.50000 35.00000`
	`$sig.levelA`
	`[1] 0.171`
	
	`$sig.levelB`
	`[1] 0`
	
	`$sig.levelAB`
	`[1] 5e-04`

Output 12.8

The *post hoc* tests for each analysis are conducted using the same command structure. That is, we define the number of levels of factor A, then factor B, then indicate the dataframe. Therefore, to run *post hoc* tests based on a 20% trimmed mean, we execute:[3]

```
mcp2atm(2,3, gogglesWide)
```

To conduct *post hoc* tests based on an M-estimator we execute:[4]

```
mcp2a(2,3, gogglesWide)
```

Output 12.9 shows the *post hoc* tests based on trimmed means (*mcp2atm*). The main effect of gender is tested by *$Factor.A$test* and *$Factor.A$psihat*. We have two choices.

[3] Obviously if you changed the level of trim for the main analysis you would need to do the same here. For example, for 10% trimmed means:

```
t2way(2,3, gogglesWide, tr = .1)
mcp2atm(2,3, gogglesWide, tr = .1)
```

[4] Remember that if you chose the median as your M-estimator then you would need to execute:

```
mcp2a(2,3, gogglesWide, est = median)
```

The first is to interpret the value in the column labelled *test* against the critical value (*crit*): if the test value is larger than the critical value then the test is significant (at $p < .05$). In this case, 1.29 is smaller than 2.06 so the result is non-significant. The second choice is to interpret *psihat* and its confidence interval and *p*-value. We should focus on interpreting the confidence interval because (unlike the *p*-value) it is corrected for the number of tests. In this case the confidence interval crosses zero, which indicates a non-significant result. These tests of gender, because it contains only two levels, basically just confirm what we already know from the main analysis.

The effect of alcohol (*$Factor.B$test* and *$Factor.B$psihat*) is more interesting because it breaks down the main effect of alcohol. There are three contrasts to interpret, but how do we know what they mean? To interpret these contrasts we need to look at the contrast codes for factor A, B and the interaction at the bottom of the output. The rows labelled [1,] … [6,] relate to the six columns of data. In other words they are: F_None, F_2 Pints, F_4 Pints, M_None, M_2 Pints, M_4 Pints. Remembering that groups with positive codes are compared against groups with negative codes, *$conA* tells us that A was the effect of gender (you have the three female groups coded with 1 and the three male groups coded with −1). Similarly, *$conB* tells us that B was the effect of alcohol split into three contrasts. Each contrast is in a separate column. We could rewrite this matrix as:

```
             Con1 Con2  Con3
F_None         1    1    0
F_2 Pints     -1    0    1
F_4 Pints      0   -1   -1
M_None         1    1    0
M_2 Pints     -1    0    1
M_4 Pints      0   -1   -1
```

Remembering that 0 means that the group is not involved, and that positives are compared to negatives, the first contrast (column 1) compares 2 pints to none, the second contrast (column 2) is 4 pints compared to none, and the third (column 3) is 2 pints compared to 4 pints).

Finally, the codes for the interaction (*$conAB*) are the same as for the main effect of alcohol except that the plus and minus signs are reversed for males and females, which tests whether the effect of alcohol differs across gender. In other words, contrast 1 compares whether the difference between 2 pints and no alcohol is different in men and women.

For the main effect of alcohol, contrast 1 is not significant (−0.52 is smaller than 2.68 and the confidence interval for *psihat* crosses zero), but contrasts 2 (5.75 is greater than 2.65 and the confidence interval for *psihat* does not contain zero) and 3 (6.18 is greater than 2.64 and the confidence interval for *psihat* does not contain zero) are. This indicates a significant difference in attractiveness scores for 4 pints compared to both no alcohol and 2 pints, but not between 2 pints and no alcohol.

For the interaction term, we get the same profile of results: contrast 1 is not significant (−0.52 is smaller than 2.68 and the confidence interval for *psihat* crosses zero), but contrasts 2 (−4.68 is greater than 2.65 – you can ignore the minus sign – and the confidence interval for *psihat* does not contain zero) and 3 (−4.08 is greater than 2.64 and the confidence interval for *psihat* does not contain zero) are. These findings indicate that the difference in attractiveness scores for 4 pints compared to both no alcohol and 2 pints differed in men and women, but that the lack of difference between 2 pints and no alcohol was similar for males and females. This profile of results tells the same story as the factorial ANOVA that we interpreted in the main part of the chapter.

```
$Factor.A
$Factor.A$test
     con.num     test     crit       se       df
[1,]       1 1.290994 2.065879 7.745967 23.57301

$Factor.A$psihat
     con.num psihat  ci.lower ci.upper   p.value
[1,]       1     10 -6.002228 26.00223 0.2092233

$Factor.B
$Factor.B$test
     con.num       test     crit       se       df
[1,]       1 -0.5203149 2.678921 6.406377 14.50207
[2,]       2  5.7486837 2.647995 6.233311 16.11968
[3,]       3  6.1847459 2.636865 6.332785 16.81814

$Factor.B$psihat
     con.num    psihat  ci.lower ci.upper      p.value
[1,]       1 -3.333333 -20.49551 13.82885 6.106962e-01
[2,]       2 35.833333  19.32755 52.33911 2.905447e-05
[3,]       3 39.166667  22.46796 55.86537 1.047835e-05

$Factor.AB
$Factor.AB$test
     con.num       test     crit       se       df
[1,]       1 -0.5203149 2.678921 6.406377 14.50207
[2,]       2 -4.6791611 2.647995 6.233311 16.11968
[3,]       3 -4.0793005 2.636865 6.332785 16.81814

$Factor.AB$psihat
     con.num     psihat  ci.lower  ci.upper      p.value
[1,]       1  -3.333333 -20.49551  13.82885 0.6106961628
[2,]       2 -29.166667 -45.67245 -12.66089 0.0002466289
[3,]       3 -25.833333 -42.53204  -9.13463 0.0007964981


$All.Tests
[1] NA

$conA
     [,1]
[1,]    1
[2,]    1
[3,]    1
[4,]   -1
[5,]   -1
[6,]   -1

$conB
     [,1] [,2] [,3]
[1,]    1    1    0
[2,]   -1    0    1
[3,]    0   -1   -1
[4,]    1    1    0
[5,]   -1    0    1
[6,]    0   -1   -1
```

```
$conAB
     [,1] [,2] [,3]
[1,]    1    1    0
[2,]   -1    0    1
[3,]    0   -1   -1
[4,]   -1   -1    0
[5,]    1    0   -1
[6,]    0    1    1
```

Output 12.9

Output 12.10 shows the *post hoc* tests based on an M-estimator (*mcp2a*). The interpretation of these results is exactly the same as for the trimmed means. If the value of *sig.test* is less than the critical value (*sig.crit*) and the confidence interval does not cross zero then the contrast is significant. For the main effect of alcohol we find a significant difference in attractiveness scores for 4 pints compared to both no alcohol, $\hat{\psi} = 35.80$, $p < .001$, and 2 pints, $\hat{\psi} = 40.80$, $p < .001$, but not between 2 pints and no alcohol, $\hat{\psi} = -5$, $p = .383$. Similarly, for the interaction term, males and females were comparable in terms of the difference in attractiveness ratings between 4 pints compared to both no alcohol, $\hat{\psi} = -32.23$, $p < .001$, and 2 pints, $\hat{\psi} = -27.23$, $p < .01$, but not between 2 pints and no alcohol, $\hat{\psi} = -5$, $p = .318$.

```
$FactorA
     con.num   psihat sig.test sig.crit  ci.lower ci.upper
[1,]       1 14.46429   0.1515    0.025 -10.08929 28.23214

$FactorB
     con.num   psihat sig.test sig.crit  ci.lower ci.upper
[1,]       1 -5.00000   0.3825    0.025 -18.83929 13.24405
[2,]       2 35.80357   0.0000    0.025  20.62500 51.84524
[3,]       3 40.80357   0.0000    0.025  21.25000 55.20833

$Interactions
     con.num    psihat sig.test sig.crit  ci.lower   ci.upper
[1,]       1  -5.00000   0.3180    0.025 -19.37500  12.500000
[2,]       2 -32.23214   0.0005    0.025 -45.20833 -13.750000
[3,]       3 -27.23214   0.0015    0.025 -41.96429  -9.583333
```

Output 12.10

OLIVER TWISTED

Please Sir, can I have some more … robust methods?

'These robust tests are not nearly complicated enough', salivates Oliver with a maniacal look in his eye and suspiciously empty bowl of additive-ridden sweets by his side. 'I want to add in a third independent variable, and then I want the magic number ferret to lick the brains from my skull.' Oh dear, he's lost it. You can lose it too by finding out how to do a robust three-way independent ANOVA on the companion website. If you're lucky you might get a brain licking too or, at the very least, a headache.

12.8. Calculating effect sizes ③

As we saw in previous chapters (e.g., section11.6), we can use omega squared (ω^2) as an effect size measure. The calculation of ω^2 becomes somewhat more cumbersome in factorial designs ('somewhat' being one of my characteristic understatements!). Howell (2006), as ever, does a wonderful job of explaining the complexities of it all (and has a nice table summarizing the various components for a variety of situations). Condensing all of this down, I'll just say that we need to first compute a variance component for each of the effects (the two main effects and the interaction term) and the error, and then use these to calculate effect sizes for each. If we call the first main effect A, the second main effect B and the interaction effect $A \times B$, then the variance components for each of these are based on the mean squares of each effect and the sample sizes on which they're based:

$$\hat{\sigma}_\alpha^2 = \frac{(a-1)(\text{MS}_A - \text{MS}_\text{R})}{nab}$$

$$\hat{\sigma}_\beta^2 = \frac{(b-1)(\text{MS}_B - \text{MS}_\text{R})}{nab}$$

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{(a-1)(b-1)(\text{MS}_{A\times B} - \text{MS}_\text{R})}{nab}$$

In these equations, a is the number of levels of the first independent variable, b is the number of levels of the second independent variable and n is the number of people per condition.

We also need to estimate the total variability, and this is just the sum of these other variables plus the residual mean squares:

$$\hat{\sigma}_\text{total}^2 = \hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \text{MS}_\text{R}$$

The effect size is then the variance estimate for the effect in which you're interested divided by the total variance estimate:

$$\omega_\text{effect}^2 = \frac{\hat{\sigma}_\text{effect}^2}{\hat{\sigma}_\text{total}^2}$$

We can write a function in **R** to compute the effect sizes for us (see R's Souls' Tip 6.2). This process might seem like a faff, but remember that once you have the function written, you can use it again and again. Output 12.4 gives us the sums of squares for each effect and the interaction, so it would be nice to be able to enter these values to get the resulting omega squared. We can write and execute this function:

```
omega_factorial<-function(n, a, b, SSa, SSb, SSab, SSr)
{
	MSa<-SSa/(a-1)
	MSb<-SSb/(b-1)
	MSab<-SSab/((a-1)*(b-1))
	MSr<-SSr/(a*b*(n-1))
	varA<-((a-1)*(MSa-MSr))/(n*a*b)
	varB<-((b-1)*(MSb-MSr))/(n*a*b)
	varAB<-((a-1)*(b-1)*(MSab-MSr))/(n*a*b)
	varTotal<-varA + varB + varAB + MSr
	print(paste("Omega-Squared A: ", varA/varTotal))
	print(paste("Omega-Squared B: ", varB/varTotal))
	print(paste("Omega-Squared AB: ", varAB/varTotal))
	}
```

This creates a function called *omega_factorial*.[5] First, we tell **R** that we want to be able to input n, a, b, SSa, SSb, SSab, and SSr into the function (these are specified in brackets). This means that to use the function we have to input these values in brackets in the correct order. The rest of the function uses these values to compute the various values of ω^2. The first four commands take the sums of squares and convert them to mean squares by dividing by the degrees of freedom (rather than have you input the degrees of freedom by hand, we calculate them from *a* and *b*, the number of levels of the two independent variables). The next four lines calculate the variance estimates in the equations above; for example, *varA* computes $\hat{\sigma}^2_\alpha$ by writing out the equation above in **R**-speak (because of how I have labelled everything in the function you should be able to compare directly the command in the function with the equation above). The final three lines print some text (in speech marks) that describes which ω^2 we're calculating followed by each variance estimate divided by the total variance estimate (i.e., $\hat{\sigma}^2_{\text{effect}}/\hat{\sigma}^2_{\text{total}}$).

Having executed this function we can use it to calculate ω^2 in the current data by using the values of *n* (8 people per group), *a* (levels of gender = 2), *b* (levels of alcohol = 3) and the four sums of squares from Output 12.4:

```
omega_factorial(8, 2, 3, 169, 3332, 1978, 3488)
```

Executing this command will print the following to the console:

```
[1] "Omega-Squared A:  0.00949745068429"
[1] "Omega-Squared B:  0.34982188991376"
[1] "Omega-Squared AB:  0.200209417472152"
```

For the main effect of gender we get $\omega^2_{\text{gender}} = 0.009$; for the main effect of alcohol we get $\omega^2_{\text{alcohol}} = 0.350$; and for the interaction $\omega^2_{\text{gender}\ \ \text{alcohol}} = 0.200$.

I have mentioned several times that it is perhaps more useful to quantify focused differences (i.e., between two things) than overall effects. In the case of a factorial ANOVA when there is a significant interaction, we might compute effect sizes for the simple effects (section 12.5.10). In other words, compute the differences between means for one independent variable at different levels of the other independent variable. In the current example, we might compute effect sizes for the effect of gender at different levels of alcohol. We could again use the *mes()* function from the *calculate.es* package:

```
mes(mean_males, mean_females, sd_males, sd_females, n_males, n_females)
```

We have all the information we need to use the *mes()* function in Output 12.2. For example, if we want to compare men and women who drank no alcohol we would execute:

```
mes(66.875, 60.625, 10.3293963, 4.95515604, 8, 8)
```

We have entered the mean of the men who drank no alcohol (66.875), the mean of women who drank no alcohol (60.625), the corresponding standard deviations (10.329 and 4.955), and the sample sizes (both 8).

Similarly we can get effect sizes for the difference between men and women who drank 2 pints by executing:

```
mes(66.875, 62.5, 12.5178444, 6.5465367, 8, 8)
```

Finally, the difference between men and women who drank 4 pints can be quantified by executing:

```
mes(35.625, 57.5, 10.8356225, 7.0710678, 8, 8)
```

The (edited) outputs of these commands are shown in Output 12.11. The difference in attractiveness scores between males and females who drank no alcohol is a medium effect

[5] If you install the package *DSUR*, which we produced for this book, you can use this function without executing these commands.

(the means are under a standard deviation different), $d = 0.77$, $r = .36$; the difference between males and females who drank 2 pints is a fairly small effect (there is less than half a standard deviation difference between the group means), $d = 0.44$, $r = .21$; finally, the difference between males and females who drank 4 pints is a very large effect (the means are more than 2 standard deviation apart), $d = -2.39$, $r = -.77$.

```
No Alcohol: Males vs. Females

$MeanDifference
        d     var.d         g     var.g
0.7715168 0.2686012 0.7294340 0.2400984

$Correlation
         r      var.r
0.35990788 0.04428981

2 Pints: Males vs. Females

$MeanDifference
        d     var.d         g     var.g
0.4379891 0.2559948 0.4140988 0.2288298

$Correlation
        r     var.r
0.2139249 0.0556082

4 Pints: Males vs. Females

$MeanDifference
         d      var.d          g      var.g
-2.3909552  0.4286458 -2.2605394  0.3831598

$Correlation
          r        var.r
-0.767030763  0.007475955
```

Output 12.11

12.9. Reporting the results of two-way ANOVA ②

As with the other ANOVAs we've encountered, we have to report the details of the F-ratio and the degrees of freedom from which it was calculated. For the various effects in these data the F-ratio will be based on different degrees of freedom: it was derived from dividing the mean squares for the effect by the mean squares for the residual. For the effects of alcohol and the alcohol × gender interaction, the model degrees of freedom were 2 ($df_M = 2$), but for the effect of gender the degrees of freedom were only 1 ($df_M = 1$). For all effects, the degrees of freedom for the residuals were 42 ($df_R = 42$). We can, therefore, report the three effects from this analysis as follows:

- ✓ There was a significant main effect of the amount of alcohol consumed at the nightclub, on the attractiveness of the mate they selected, $F(2, 42) = 20.07$, $p < .001$, $\omega^2 = .35$. The Bonferroni *post hoc* tests revealed that the attractiveness of selected dates was significantly lower after 4 pints than both after 2 pints and no alcohol (both *ps* < .001). The attractiveness of dates after 2 pints and no alcohol were not significantly different.

- ✓ There was a non-significant main effect of gender on the attractiveness of selected mates, $F(1, 42) = 2.03$, $p = .161$, $\omega^2 = .009$.
- ✓ There was a significant interaction effect between the amount of alcohol consumed and the gender of the person selecting a mate, on the attractiveness of the partner selected, $F(2, 42) = 11.91$, $p < .001$, $\omega^2 = .20$. This indicates that male and female genders were affected differently by alcohol. Specifically, the attractiveness of partners was similar in males ($M = 66.88$, $SD = 10.33$) and females ($M = 60.63$, $SD = 4.96$) after no alcohol, $d = 0.77$; the attractiveness of partners was also similar in males ($M = 66.88$, $SD = 12.52$) and females ($M = 62.50$, $SD = 6.55$) after 2 pints, $d = 0.44$; however, attractiveness of partners selected by males ($M = 35.63$, $SD = 10.84$) was significantly lower than those selected by females ($M = 57.50$, $SD = 7.07$) after 4 pints, $d = -2.39$.

Labcoat Leni's Real Research 12.1 Don't forget your toothbrush? ②

Davey, G. C. L., et al. (2003). *Journal of Behavior Therapy & Experimental Psychiatry, 34*, 141–160.

We have all experienced that feeling after we have left the house of wondering whether we locked the door, or closed the window, or whether we remembered to remove the bodies from the fridge in case the police turn up. This behaviour is normal; however, people with obsessive compulsive disorder (OCD) tend to check things excessively. They might, for example, check whether they have locked the door so often that it takes them an hour to leave their house. It is a very debilitating problem.

One theory of this checking behaviour in OCD suggests that it is caused by a combination of the mood you are in (positive or negative) interacting with the rules you use to decide when to stop a task (do you continue until you feel like stopping, or until you have done the task as best as you can?). Davey, Startup, Zara, MacDonald, and Field (2003) tested this hypothesis by inducing a negative, positive or no mood in different people and then asking them to imagine that they were going on holiday and to generate as many things as they could that they should check before they went away. Within each mood group, half of the participants were instructed to generate as many items as they could (known as an 'as many as can' stop rule), whereas the remainder were asked to generate items for as long as they felt like continuing the task (known as a 'feel like continuing' stop rule). The data are in the file **Davey2003.dat**.

Davey et al. hypothesized that people in negative moods, using an 'as many as can' stop rule, would generate more items than those using a 'feel like continuing' stop rule. Conversely, people in a positive mood would generate more items when using a 'feel like continuing' stop rule compared to an 'as many as can' stop rule. Finally, in neutral moods, the stop rule used shouldn't affect the number of items generated. Draw an error bar chart of the data and then conduct the appropriate analysis to test Davey et al.'s hypotheses.

Answers are in the additional material on the companion website (or look at pages 148–149 in the original article).

What have I discovered about statistics? ②

This chapter has been a whistle-stop tour of factorial ANOVA. In fact we'll come across more factorial ANOVAs in the next two chapters, but for the time being we've just looked at the situation where there are two independent variables, and different people have been used in all experimental conditions. We started off by discovering that even complex ANOVAs are simply regression analyses in disguise. We moved on to look at how to calculate the various sums of squares in this analysis, but, most important, we saw that we get three effects: two main effects (the effect of each of the independent variables) and an interaction effect. We moved on to see how this analysis is done using **R** and how the output is interpreted. Much of this was similar to the ANOVAs we've come across in previous chapters, but one big difference was the interaction term. We spent a bit of time exploring interactions (and especially interaction graphs) to see what an interaction looks like and how to spot it. The brave readers also found out how to follow up an interaction with simple effects analysis. Finally, we discovered that calculating effect sizes in factorial designs is a complete headache and should be attempted only by the criminally insane. So far we've steered clear of repeated-measures designs, but in the next chapter I have to resign myself to the fact that I can't avoid explaining them for the rest of my life.☹

We also discovered that no sooner had I started my first band than it disintegrated. I went with drummer Mark to sing in a band called the Outlanders, who were much better musically but were not, if the truth were told, metal enough for me. They also sacked me after a very short period of time for not being able to sing like Bono (an insult at the time, but in retrospect …).

R packages used in this chapter

car
compute.es
ggplot2
multcomp
pastecs
reshape
WRS

R functions used in this chapter

Anova()
aov()
by()
cast()
contrasts()
confint()
factor()
ggplot()
gl()
glht()
mcp2a()
mcp2atm()
leveneTest()
list()
lm()
melt()
mes()
pairwise.t.test()
pbad2way()
plot()
read.csv()
rep()
stat.desc()
summary()
summary.lm()
t2way()

Key terms that I've discovered

Beer-goggles effect
Factorial ANOVA
Independent factorial design
Interaction graph
Mixed design
Related factorial design
Simple effects analysis

Smart Alex's tasks

- **Task 1**: People's musical tastes change as they get older (my parents, for example, after years of listening to relatively cool music when I was a kid, subsequently hit their mid-forties and developed a worrying obsession with country and western music). This worries me immensely because the future seems bleak if it is spent listening to Garth Brooks and thinking 'oh boy, did I underestimate Garth's immense talent when I was in my 20s'. So, I did some imaginary research to find out whether my fate really was sealed, or whether it's possible to be old and like good music too. First, I got two groups of people (45 people in each group): one group contained young people (which I arbitrarily decided was under 40 years of age) and the other group contained more mature individuals (above 40 years of age). This is my first independent variable, **age**. I then split each of these groups of 45 into three smaller groups of 15 and assigned them to listen to Fugazi (who everyone knows are the coolest band on the planet),[6] ABBA or Barf Grooks (a less well-known country and western musician not to be confused with anyone real who produces music that makes me want to barf). This is my second independent variable, **music**. After listening to the music I got each person to rate it on a scale ranging from −100 (please poke a pencil through my eardrum so I don't have to listen any more) through 0 (I am completely indifferent) to +100 (I love this music so much, it gives me a tingle down my spine). This variable is called **liking**. The data are in the file **fugazi.dat**. Conduct a two-way independent ANOVA on them. ②
- **Task 2**: In Chapter 3 we used some data that related to men and women's psychological arousal levels when watching either *Bridget Jones's Diary* or *Memento* (**ChickFlick.dat**). Analyse these data to see whether men and women differ in their reactions to different types of films. ②
- **Task 3**: At the start of this chapter I described a way of empirically researching whether I wrote better songs than my old band mate Malcolm, and whether this depended on the type of song (a symphony or song about flies). The outcome variable would be the number of screams elicited by audience members during the songs. These data are in the file **Escape From Inside.dat**. Draw an error bar graph (lines) and analyse and interpret these data. ②
- **Task 4**: Using R's Souls' Tip 12.2, conduct a simple effects analysis of the effect of alcohol at different levels of gender (which is the opposite to the example in the chapter). ③
- **Task 5**: Back in 2008, hospitals were reporting an increase in injuries related to playing Nintendo Wii (http://www.telegraph.co.uk/news/uknews/1576244/Spate-of-injuries-blamed-on-Nintendo-Wii.html). These injuries were attributed mainly to

[6] See http://www.dischord.com

muscle and tendon strains. A researcher was interested to see whether these injuries could be prevented. She hypothesized that a stretching warm-up before playing Wii would help lower injuries, and that athletes would be less susceptible to injuries because their regular activity makes them more flexible. She took 60 athletes and 60 non-athletes (**athlete**), half of them played Wii and half watched others playing as a control (**wii**), and within these groups half did a 5-minute stretch routine before playing/watching whereas the other half did not (**stretch**). The outcome was a pain score out of 10 (where 0 is no pain, and 10 is severe pain) after playing for 4 hours (**injury**). The data are in the file **Wii.dat**. Conduct a three-way ANOVA to test whether athletes are less prone to injury, and whether the prevention programme worked. ③

The answers are on the companion website. Task 1 is an example from Field and Hole (2003) and so has a more detailed answer if you feel like you want it.

Further reading

Howell, D. C. (2006). *Statistical methods for psychology* (6th ed.). Belmont, CA: Duxbury. (Or you might prefer his *Fundamental Statistics for the Behavioral Sciences,* also in its 6th edition, 2007.)

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioural research: A correlational approach*. Cambridge: Cambridge University Press. (This is quite advanced but really cannot be bettered for contrasts and effect size estimation.)

Rosnow, R. L., & Rosenthal, R. (2005). *Beginning behavioral research: A conceptual primer* (5th ed.). Upper Saddle River, NJ: Pearson/Prentice Hall. (Has some wonderful chapters on ANOVA, with a particular focus on effect size estimation, and some very insightful comments on what interactions actually mean.)

Interesting real research

Davey, G. C. L., Startup, H. M., Zara, A., MacDonald, C. B., & Field, A. P. (2003). Perseveration of checking thoughts and mood-as-input hypothesis. *Journal of Behavior Therapy & Experimental Psychiatry*, *34*, 141–160.