Chapter 13: Factorial ANOVA

Smart Alex's Solutions

Task 1

*People's musical tastes tend to 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 possibility worries me immensely because the future seems incredibly 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 twenties'. So, I thought I'd do some research. I took two groups (**age**): young people (I arbitrarily decided that 'young' meant under 40 years of age) and older people (above 40 years of age). There were 45 people in each group, and I split each group into three smaller groups of 15 and assigned them to listen to Fugazi, ABBA or Barf Grooks (**music**). I got each person to rate it (**liking**) on a scale ranging from −100 (I hate this foul music) through 0 (I am completely indifferent) to +100 (I love this music so much I'm going to explode). The data are in the file **Fugazi.sav**, conduct a two-way independent ANOVA on them.*

SPSS output

The error bar chart of the music data (Figure 1) shows the mean rating of the music played to each group. It's clear from this chart that when people listened to Fugazi the two age groups were divided: the older ages rated it very low, but the younger people rated it very highly. A reverse trend is found if you look at the ratings for Barf Grooks: the youngsters give it low ratings, while the wrinkly ones love it. For ABBA the groups agreed: both old and young rated them highly.

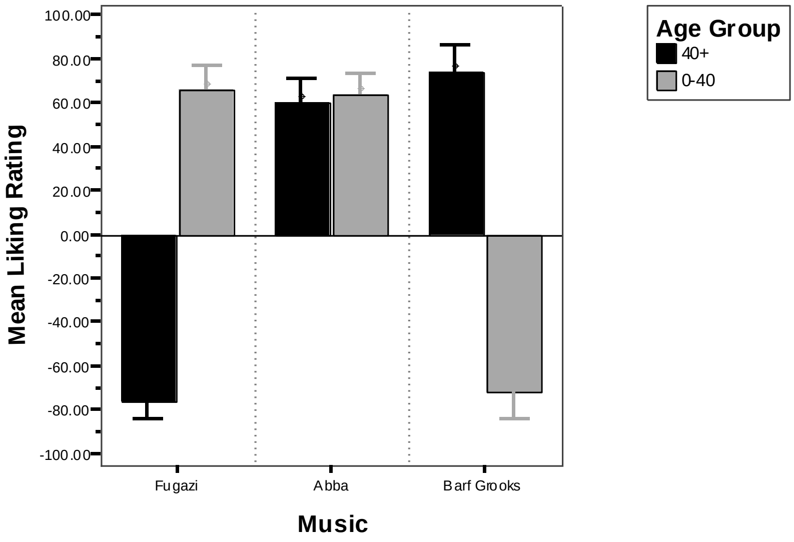

Figure 1

Output 1 shows Levene's test. For these data the significance value is .322, which is greater than the criterion of .05. This means that the variances in the different experimental groups are roughly equal (i.e., not significantly different), and that the assumption has been met.

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Liking Rating

F	df1	df2	Sig.
1.189	5	84	.322

Tests the null hypothesis that the error variance of t dependent variable is equal across groups.

a. Design: Intercept+MUSIC+AGE+MUSIC * AGE

Output 1

Output 2 shows the main ANOVA summary table.

Tests of Between-Subjects Effects

Dependent Variable: Liking Rating

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	392654.933[a]	5	78530.987	202.639	.000
Intercept	34339.600	1	34339.600	88.609	.000
MUSIC	**81864.067**	**2**	**40932.033**	105.620	**.000**
AGE	**.711**	**1**	**.711**	.002	**.966**
MUSIC * AGE	**310790.156**	**2**	**155395.078**	400.977	**.000**
Error	32553.467	84	387.541		
Total	459548.000	90			
Corrected Total	425208.400	89			

a. R Squared = .923 (Adjusted R Squared = .919)

Output 2

The main effect of music is shown by the *F*-ratio in the row labelled **music**; in this case the significance is .000, which is lower than the usual cut-off point of .05. Hence, we can say that there was a significant effect of the type of music on the ratings. To understand what this actually means, we need to look at the mean ratings for each type of music when we ignore whether the person giving the rating was old or young (Figure 2).

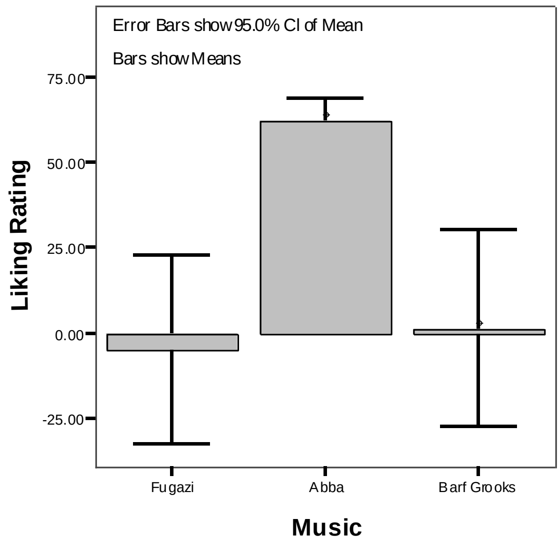

Figure 2

What this graph shows is that the significant main effect of music is likely to reflect the fact that ABBA were rated (overall) much more positively than the other two artists.

The main effect of age is shown in Output 2 by the *F*-ratio in the row labelled **age**; the probability associated with this *F*-ratio is .966, which is so close to 1 that it means that it is a virtual certainty that this *F* could occur by chance alone. Again, to interpret the effect we need to look at the mean ratings for the two age groups, ignoring the type of music to which they listened (Figure 3).

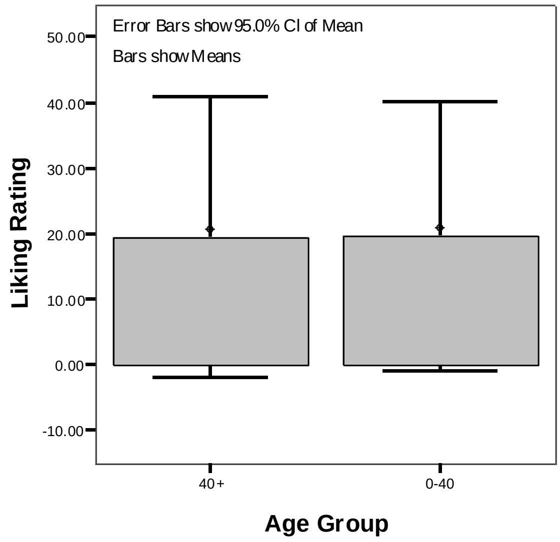

Figure 3

This graph shows that when you ignore the type of music that was being rated, older people and younger people, on average, gave almost identical ratings (i.e., the mean ratings in the two groups are virtually the same).

The interaction effect is shown in Output 2 by the *F*-ratio in the row labeled **music * age**; the associated significance value is small (.000) and is less than the criterion of .05. Therefore, we can say that there is a significant interaction between age and the type of music rated. To interpret this effect we need to look at the mean ratings in all conditions (Figure 1). The fact there is a significant interaction tells us that for certain types of music the different age groups gave different ratings. In this case, although they agree on ABBA, there are large disagreements in ratings of Fugazi and Barf Grooks.

Given that we found a main effect of music, and of the interaction between music and age, we can look at some of the *post hoc* tests to establish where the difference lies. Output 3 shows the result of Games–Howell *post hoc* tests. First, ratings of Fugazi are compared to ABBA, which reveals a significant difference (the value in the column labelled *Sig.* is less than .05), and then Barf Grooks, which reveals no difference (the significance value is greater than .05). In the next part of the table, ratings of ABBA are compared first to Fugazi (which just repeats the finding in the previous part of the table) and then to Barf Grooks, which reveals a significant difference (the significance value is below .05). The final part of the table compares Barf Grooks to Fugazi and ABBA, but these results repeat findings from the previous sections of the table.

Multiple Comparisons

Dependent Variable: Liking Rating

	(I) Music	(J) Music	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Games-Howell	Fugazi	Abba	-66.8667*	5.08292	.000	-101.1477	-32.5857
		Barf Grooks	-6.2333	5.08292	.946	-53.3343	40.8677
	Abba	Fugazi	66.8667*	5.08292	.000	32.5857	101.1477
		Barf Grooks	60.6333*	5.08292	.001	24.9547	96.3119
	Barf Grooks	Fugazi	6.2333	5.08292	.946	-40.8677	53.3343
		Abba	-60.6333*	5.08292	.001	-96.3119	-24.9547

Based on observed means.

*. The mean difference is significant at the .05 level.

Output 3

Task 2

Compute omega squared for the effects in Task 1 and report the results of the analysis.

Compute omega-squared

$$\hat{\sigma}_{\alpha}^{2} = \frac{(3-1)(40932.033 - 387.541)}{15 \times 3 \times 2} = 900.99$$

$$\hat{\sigma}_{\beta}^{2} = \frac{(2-1)(0.711 - 387.541)}{15 \times 3 \times 2} = -4.30$$

$$\hat{\sigma}_{\alpha\beta}^{2} = \frac{(3-1)(2-1)(155395.078 - 387.541)}{15 \times 3 \times 2} = 3444.61$$

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

$$\begin{aligned}\hat{\sigma}_{\text{total}}^{2} &= \hat{\sigma}_{\alpha}^{2} + \hat{\sigma}_{\beta}^{2} + \hat{\sigma}_{\alpha\beta}^{2} + \text{MS}_{\text{R}} \\ &= 900.99 - 4.30 + 3444.61 + 387.54 \\ &= 4728.84\end{aligned}$$

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

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

As such, for the main effect of music we get:

$$\omega_{\text{music}}^{2} = \frac{\hat{\sigma}_{\text{music}}^{2}}{\hat{\sigma}_{\text{total}}^{2}} = \frac{900.99}{4728.84} = .19$$

For the main effect of age we get:

$$\omega^2_{\text{age}} = \frac{\hat{\sigma}^2_{\text{age}}}{\hat{\sigma}^2_{\text{total}}} = \frac{-4.30}{4728.84} = -.001$$

For the interaction of music and age we get:

$$\omega^2_{\text{music} \times \text{age}} = \frac{\hat{\sigma}^2_{\text{music} \times \text{age}}}{\hat{\sigma}^2_{\text{total}}} = \frac{3444.61}{4728.84} = .73$$

Report the results

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*-ratios will be based on different degrees of freedom: they are derived from dividing the mean squares for the effect by the mean squares for the residual. For the effects of music and the music × age interaction, the model degrees of freedom were 2 ($df_M = 2$), but for the effect of age the degrees of freedom were only 1 ($df_M = 1$). For all effects, the degrees of freedom for the residuals were 84 ($df_R = 84$). We can, therefore, report the three effects from this analysis as follows:

- ✓ The results show that the main effect of the type of music listened to significantly affected the ratings of that music, $F(2, 84) = 105.62$, $p < .001$, $\omega^2 = .19$. The Games–Howell *post hoc* test revealed that ABBA were rated significantly higher than both Fugazi and Barf Grooks ($p < .01$ in both cases).
- ✓ The main effect of age on the ratings of the music was non-significant, $F(1, 84) = 0.002$, $p = .966$, $\omega^2 = -.001$.
- ✓ The music × age interaction was significant, $F(2, 84) = 400.98$, $p < .001$, $\omega^2 = .73$, indicating that different types of music were rated differently by the two age groups. Specifically, Fugazi were rated more positively by the young group ($M = 66.20$, $SD = 19.90$) than the old ($M = -75.87$, $SD = 14.37$); ABBA were rated fairly equally in the young ($M = 64.13$, $SD = 16.99$) and old groups ($M = 59.93$, $SD = 19.98$); Barf Grooks was rated less positively by the young group ($M = -71.47$, $SD = 23.17$) compared to the old ($M = 74.27$, $SD = 22.29$). These findings indicate that there is no hope for me — the minute I hit 40 I will suddenly start to love country and western music and will burn all of my Fugazi CDs (it will never happen … arghhhh!!!).

Task 3

In Chapter 3 we used some data that related to men and women's arousal levels when watching either Bridget Jones's Diary *or* Memento *(**ChickFlick.sav**). Analyse these data to see whether men and women differ in their reactions to different types of films.*

Output 4 shows Levene's test. For these data the significance value is .456, which is greater than the criterion of .05. This means that the variances in the different experimental groups are roughly equal (i.e. not significantly different), and that the assumption has been met.

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Arousal

F	df1	df2	Sig.
.889	3	36	.456

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + gender + film + gender * film

Output 4

Output 5 shows the main ANOVA summary table.

Tests of Between-Subjects Effects

Dependent Variable:Arousal

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1213.275[a]	3	404.425	9.920	.000
Intercept	16040.025	1	16040.025	393.433	.000
gender	87.025	1	87.025	2.135	.153
film	1092.025	1	1092.025	26.785	.000
gender * film	34.225	1	34.225	.839	.366
Error	1467.700	36	40.769		
Total	18721.000	40			
Corrected Total	2680.975	39			

a. R Squared = .453 (Adjusted R Squared = .407)

Output 5

The main effect of gender is shown by the *F*-ratio in the row labelled **gender**; in this case the significance is .153, which is greater than the usual cut-off point of .05. Hence, we can say that there was not a significant effect of gender on arousal during the films. To understand what this actually means, we need to look at the mean arousal levels for men and women, ignoring which film they watched (Figure 4).

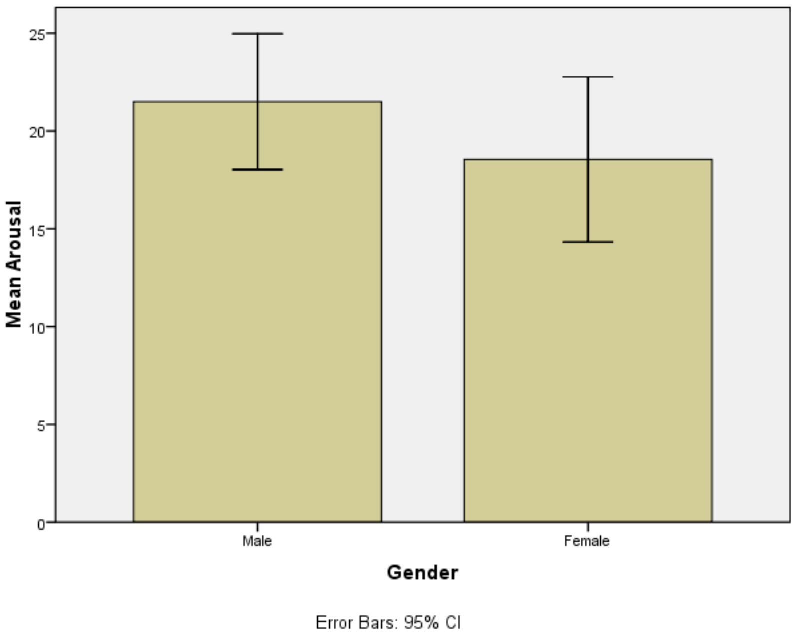

Figure 4

What this graph shows is that arousal levels were quite similar for men and women in general; this is why the main effect of gender was non-significant.

The main effect of film is shown in Output 5 by the *F*-ratio in the row labelled **film**; the probability associated with this *F*-ratio is .000, which is less than the critical value of .05, hence we can say that arousal levels were significantly different in the two films. Again, to interpret the effect we need to look at the mean arousal levels, but this time comparing the two films (and ignoring whether the person was male or female). Figure 5 shows that when you ignore the gender of the person, arousal levels were significantly higher for *Memento* than *Bridget Jones's Diary*.

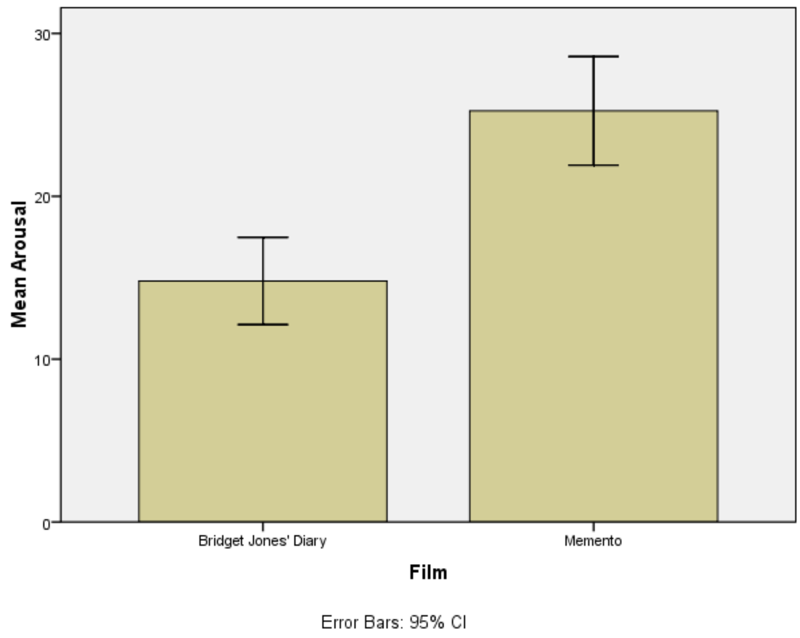

Figure 5

The interaction effect is shown in Output 5 by the *F*-ratio in the row labelled **gender * film**; the associated significance value is .366, which is greater than the criterion of .05. Therefore, we can say that there is not a significant interaction between gender and the type of film watched. To interpret this effect we need to look at the mean arousal in all conditions (Figure 6).

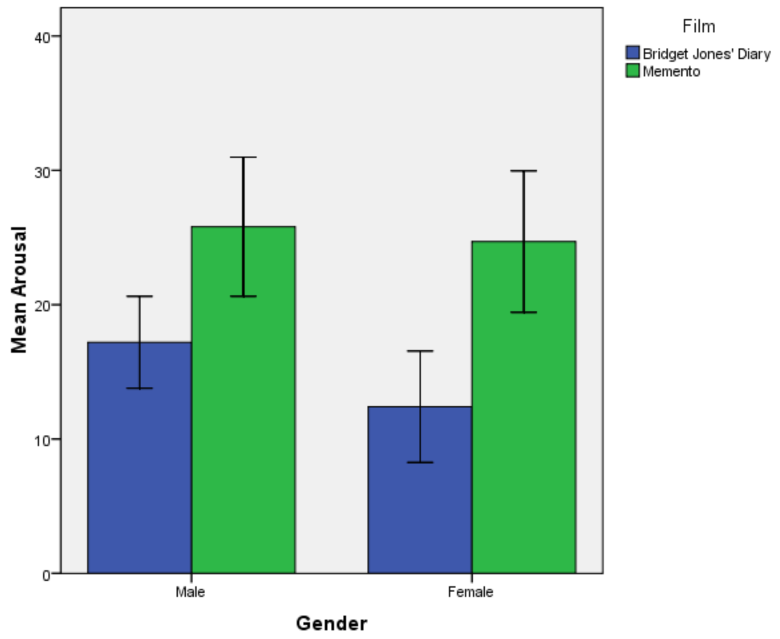

Figure 6

This graph shows the non-significant interaction: arousal levels are higher for *Memento* compared to *Bridget Jones's Diary* in both men and women (i.e., the difference between the green and blue bars is more or less the same for men and women).

Task 4

Compute omega squared for the effects in Task 3 and report the results of the analysis

Compute omega-squared

$$\hat{\sigma}_{\alpha}^2 = \frac{(2-1)(87.03-40.77)}{10 \times 2 \times 2} = 1.16$$

$$\hat{\sigma}_{\beta}^2 = \frac{(2-1)(1092.03-40.77)}{10 \times 2 \times 2} = 1091.01$$

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{(2-1)(2-1)(34.23-40.77)}{10 \times 3 \times 2} = -0.16$$

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

$$\begin{aligned}\hat{\sigma}_{\text{total}}^2 &= \hat{\sigma}_{\alpha}^2 + \hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + MS_R \\ &= 1.16 + 1091.01 - 0.16 + 40.77 \\ &= 1132.78\end{aligned}$$

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

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

As such, for the main effect of gender we get:

$$\omega_{\text{gender}}^2 = \frac{\hat{\sigma}_{\text{gender}}^2}{\hat{\sigma}_{\text{total}}^2} = \frac{1.16}{1132.78} = .01$$

For the main effect of film we get:

$$\omega_{\text{film}}^2 = \frac{\hat{\sigma}_{\text{film}}^2}{\hat{\sigma}_{\text{total}}^2} = \frac{1091.01}{1132.78} = .96$$

For the interaction we get:

$$\omega^2_{\text{gender} \times \text{film}} = \frac{\hat{\sigma}^2_{\text{gender} \times \text{film}}}{\hat{\sigma}^2_{\text{total}}} = \frac{-0.16}{1132.78} = -.0001$$

Interpreting and writing the result

We can report the three effects from this analysis as follows:

- ✓ The results show that the main effect of the type of film significantly affected arousal during that film, $F(1, 36) = 26.79$, $p < .001$, $\omega^2 = .96$; Arousal levels were significantly higher during *Memento* compared to *Bridget Jones's Diary*.
- ✓ The main effect of gender on arousal levels during the films was non-significant, $F(1, 84) = 2.14$, $p = .153$, $\omega^2 = .01$.
- ✓ The gender × film interaction was non-significant, $F(1, 36) = 0.84$, $p = .366$, $\omega^2 = -.0001$. This showed that arousal levels were higher for *Memento* compared to *Bridget Jones's Diary* in both men and women.

Task 5

*In Chapter 3 we used some data that related to learning in men and women when either reinforcement or punishment was used in teaching (**Method Of Teaching.sav**). Analyse these data to see whether men and women's learning differs according to the teaching method used.*

To answer this question we need to conduct a 2 (**Method**: electric shock vs. being nice) × 2 (**Gender**: male vs. female) two-way independent ANOVA on scores on an SPSS exam.

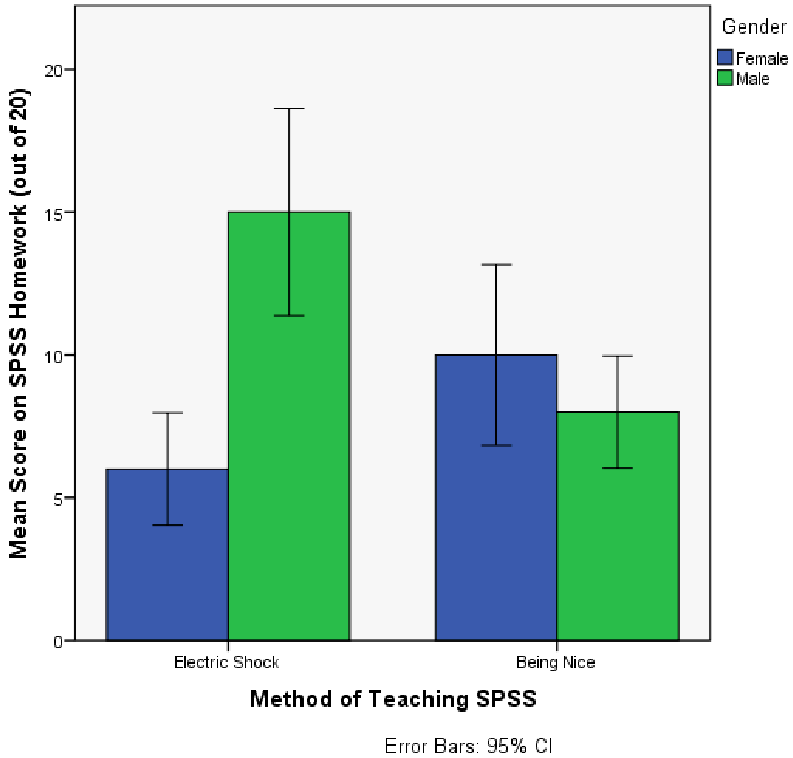

Figure 7

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Score on SPSS Homework (out of 20)

F	df1	df2	Sig.
.667	3	16	.585

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Method + Gender + Method * Gender

Output 6

Output shows Levene's test. For these data the significance value is .585, which is greater than the criterion of .05. This means that the variances in the different experimental groups are roughly equal (i.e. not significantly different), and that the assumption has been met.

1. Method of Teaching SPSS

Dependent Variable: Score on SPSS Homework (out of 20)

Method of Teaching SPSS	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Electric Shock	10.500	.707	9.001	11.999
Being Nice	9.000	.707	7.501	10.499

Output 7

2. Gender

Dependent Variable: Score on SPSS Homework (out of 20)

Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Female	8.000	.707	6.501	9.499
Male	11.500	.707	10.001	12.999

Output 8

3. Method of Teaching SPSS * Gender

Dependent Variable: Score on SPSS Homework (out of 20)

Method of Teaching SPSS	Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Electric Shock	Female	6.000	1.000	3.880	8.120
	Male	15.000	1.000	12.880	17.120
Being Nice	Female	10.000	1.000	7.880	12.120
	Male	8.000	1.000	5.880	10.120

Output 9

Tests of Between-Subjects Effects

Dependent Variable: Score on SPSS Homework (out of 20)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	223.750[a]	3	74.583	14.917	.000
Intercept	1901.250	1	1901.250	380.250	.000
Method	11.250	1	11.250	2.250	.153
Gender	61.250	1	61.250	12.250	.003
Method * Gender	151.250	1	151.250	30.250	.000
Error	80.000	16	5.000		
Total	2205.000	20			
Corrected Total	303.750	19			

a. R Squared = .737 (Adjusted R Squared = .687)

Output 10

Output shows the results from the main ANOVA. We can see that there was no significant main effect of **Method**, indicating that when we ignore the gender of the participant both methods of teaching had similar effects on the results of the SPSS exam, $F(1, 16) = 2.25$, $p = .153$. This result is not surprising when we look at the means in Output , as here we can see that being nice ($M = 9.0$) and electric shock ($M = 10.5$) had similar means. There was a significant main effect of **Gender**, indicating that if we ignore the method of teaching, men and women scored differently on the SPSS exam, $F(1, 16) = 12.50$, $p = .003$. If we look at the means, we can see that on average men ($M = 11.5$) scored higher than women ($M = 8.0$). There was also a significant **Gender** × **Method** interaction effect, $F(1, 16) = 30.25$, $p < .001$, indicating that the two teaching methods had different effects in men and women. If we look at the graph in Figure we can see that for men, using an electric shock resulted in significantly higher exam scores than being nice, whereas for women, the being nice teaching method resulted in significantly higher exam scores than when an electric shock was used.

Task 6

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.sav****. Draw an error bar graph (lines) and analyse these data.*

To do a multiple line chart for means that are independent (i.e., have come from different groups) we need to double-click on the multiple line chart icon in the chart builder (see the book chapter). All we need to do is to drag our variables into the appropriate drop zones. Select **Screams** from the variable list and drag it into Y-Axis?; select **Song_Type** from the variable list and drag it into X-Axis?; finally, select the **Songwriter** variable and drag it into Cluster on X: set color. This will mean that lines representing Andy's and Malcolm's songs will be displayed in different colours. Select error bars in the *properties* dialog box and click on Apply to apply them to the chart builder. Click on OK to produce the graph.

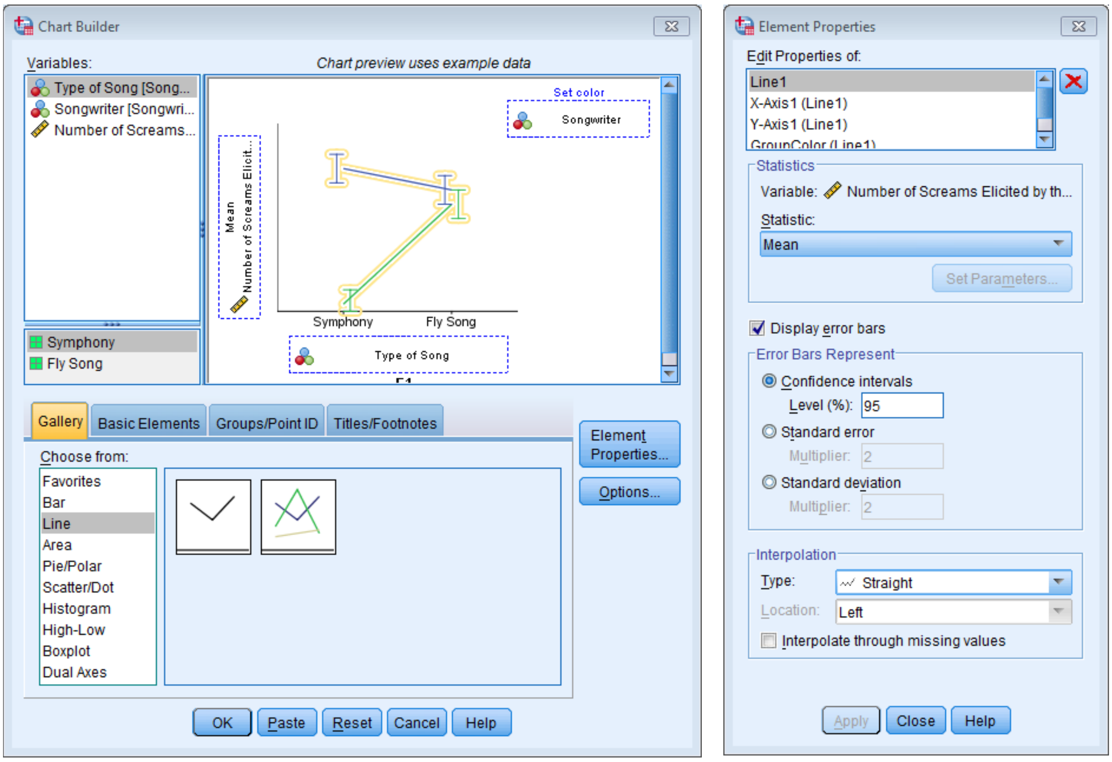

Figure 8

The resulting graph is shown in Figure 9.

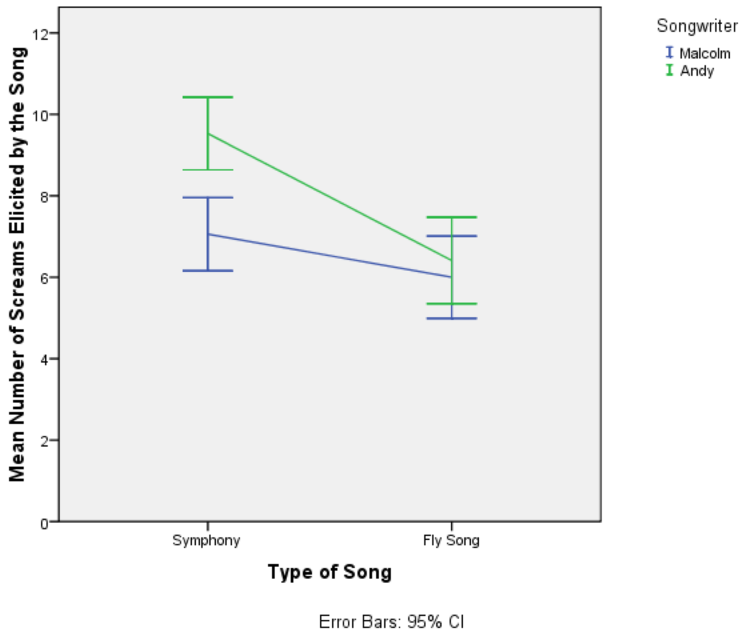

Figure 9

Output 11 shows Levene's test. For these data the significance value is .817, which is greater than the cut-off criterion of .05. This means that the variances in the different experimental groups are roughly equal (i.e. not significantly different), and that the assumption has been met.

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Number of Screams Elicited by the Song

F	df1	df2	Sig.
.311	3	64	.817

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Song_Type + Songwriter + Song_Type * Songwriter

Output 11

Tests of Between-Subjects Effects

Dependent Variable:Number of Screams Elicited by the Song

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	127.456[a]	3	42.485	11.963	.000
Intercept	3574.250	1	3574.250	1006.414	.000
Song_Type	74.132	1	74.132	20.874	.000
Songwriter	35.309	1	35.309	9.942	.002
Song_Type * Songwriter	18.015	1	18.015	5.072	.028
Error	227.294	64	3.551		
Total	3929.000	68			
Corrected Total	354.750	67			

a. R Squared = .359 (Adjusted R Squared = .329)

Output 1

Output 12 shows the main ANOVA summary table. The main effect of the type of song is shown by the *F*-ratio in the row labelled **Song_Type**; in this case the significance is .000, which is smaller than the usual cut-off point of .05. Hence, we can say that there was a significant effect of the type of song on the number of screams elicited while it was played. The graph shows that the two symphonies elicited significantly more screams of agony than the two songs about flies.

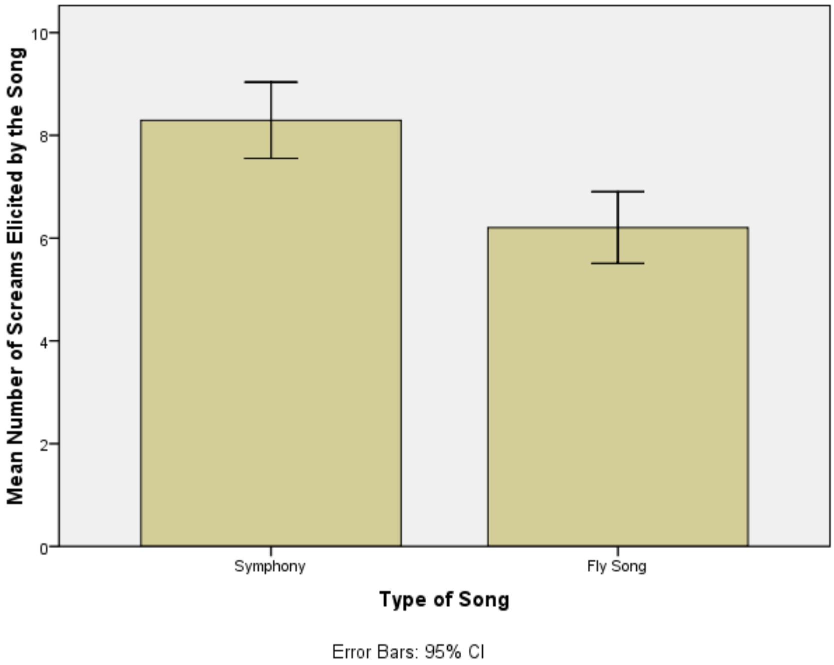

Figure 10

The main effect of the songwriter was significant because the significance of the *F*-ratio for this effect is .002, which is less than the critical value of .05, hence we can say that Andy and Malcolm differed in the reactions to their songs. The graph tells us that Andy's songs elicited significantly more screams of torment from the audience than Malcolm's songs.

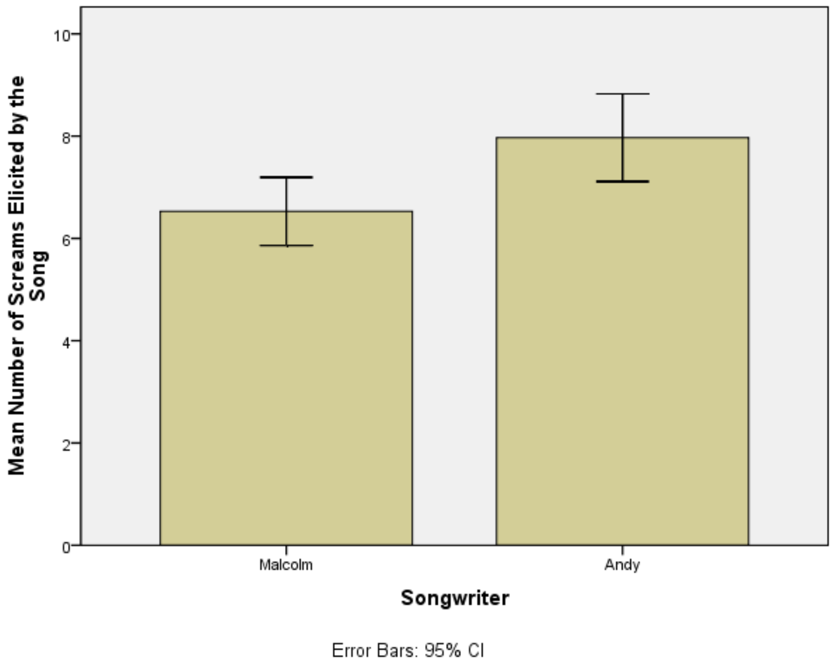

Figure 11

The interaction effect was significant too because the associated significance value is .028, which is less than the criterion of .05. Therefore, we can say that there is a significant interaction between the type of song and who wrote it on people's appreciation of the song. The line graph that you drew earlier on (Figure 9) tells us that although reactions to Malcolm's and Andy's were fairly similar for the flies song, they differed quite a bit for the symphony: Andy's symphony elicited more screams of torment than Malcolm's. We can conclude that in general Malcolm was a better songwriter than Andy, but the interaction tells us that this effect is true mainly for symphonies.

Task 7

Compute omega squared for the effects in Task 6 and report the results of the analysis.

Calculating effect sizes

$$\hat{\sigma}_{\alpha}^2 = \frac{(2-1)(74.13-3.55)}{17 \times 2 \times 2} = 1.04$$

$$\hat{\sigma}_{\beta}^2 = \frac{(2-1)(35.31-3.55)}{17 \times 2 \times 2} = 0.47$$

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{(2-1)(2-1)(18.02-3.77)}{17 \times 2 \times 2} = 0.21$$

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

$$\begin{aligned}\hat{\sigma}_{\text{total}}^2 &= \hat{\sigma}_{\alpha}^2 + \hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + MS_R \\ &= 1.04 + 0.47 + 0.21 + 3.77 \\ &= 5.49\end{aligned}$$

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

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

As such, for the main effect of song type we get:

$$\omega_{\text{Type of Song}}^2 = \frac{\hat{\sigma}_{\text{Type of Song}}^2}{\hat{\sigma}_{\text{total}}^2} = \frac{1.04}{5.49} = 0.19$$

For the main effect of songwriter we get:

$$\omega_{\text{Songwriter}}^2 = \frac{\hat{\sigma}_{\text{Songwriter}}^2}{\hat{\sigma}_{\text{total}}^2} = \frac{0.47}{5.49} = .09$$

For the interaction we get:

$$\omega^2_{\text{Type of song} \times \text{Songwriter}} = \frac{\hat{\sigma}^2_{\text{Type of song} \times \text{Songwriter}}}{\hat{\sigma}^2_{\text{total}}} = \frac{0.21}{5.49} = .04$$

Interpreting and writing the result

We can, report the three effects from this analysis as follows:

- ✓ The results show that the main effect of the type of song significantly affected screams elicited during that song, $F(1, 64) = 20.87$, $p < .001$, $\omega^2 = .19$; the two symphonies elicited significantly more screams of agony than the two songs about flies.
- ✓ The main effect of the songwriter significantly affected screams elicited during that song, $F(1, 64) = 9.94$, $p < .001$, $\omega^2 = .09$; Andy's songs elicited significantly more screams of torment from the audience than Malcolm's songs.
- ✓ The song type × songwriter interaction was significant, $F(1, 64) = 5.07$, $p < .05$, $\omega^2 = .04$. Although reactions to Malcolm's and Andy's were fairly similar for the flies song, they differed quite a bit for the symphony: Andy's symphony elicited more screams of torment than Malcolm's.

Task 8

*Using SPSS Tip 13.1, change the syntax in **GogglesSimpleEffects.sps** to look at the effect of alcohol at different levels of gender.*

The correct syntax to use is:

```
glm Attractiveness by gender alcohol

 /emmeans = tables(gender*alcohol)compare(alcohol).
```

Note that all we change is compare(gender) to compare(alcohol). The pertinent part of the output is in Output .

Univariate Tests

Dependent Variable: Attractiveness of Date

Gender		Sum of Squares	df	Mean Square	F	Sig.
Male	Contrast	5208.333	2	2604.167	31.362	.000
	Error	3487.500	42	83.036		
Female	Contrast	102.083	2	51.042	.615	.546
	Error	3487.500	42	83.036		

Each F tests the simple effects of Alcohol Consumption within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

Output 2

What this output shows is a significant effect of alcohol for males ($p < .001$) but not females ($p = .546$). Think back to the chapter; this reflects the fact that men choose very unattractive dates after 4 pints. However, there is no significant effect of alcohol at level 2 of gender. This tells us that women are not affected by the beer-goggles effect: the attractiveness of their dates does not change as they drink more.

Calculating the effect size

These effects have *df* = 2 in the model so we can't calculate an effect size (well, technically we can calculate ω^2 but I'm not entirely sure how useful that is).

Task 9

*There are reports of increases in injuries related to playing Nintendo Wii (http://ow.ly/ceWPj). These injuries were attributed mainly to muscle and tendon strains. A researcher 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.sav**. Conduct a three-way ANOVA to test whether athletes are less prone to injury, and whether the prevention programme worked.*

To answer this question we need to conduct a 2 (**Athlete**: athlete vs. non-athlete) × 2 (**Wii**: playing Wii vs. watching Wii) × 2(**Stretch**: stretching vs. no stretching) three-way independent ANOVA. Your completed dialog box should look like Figure .

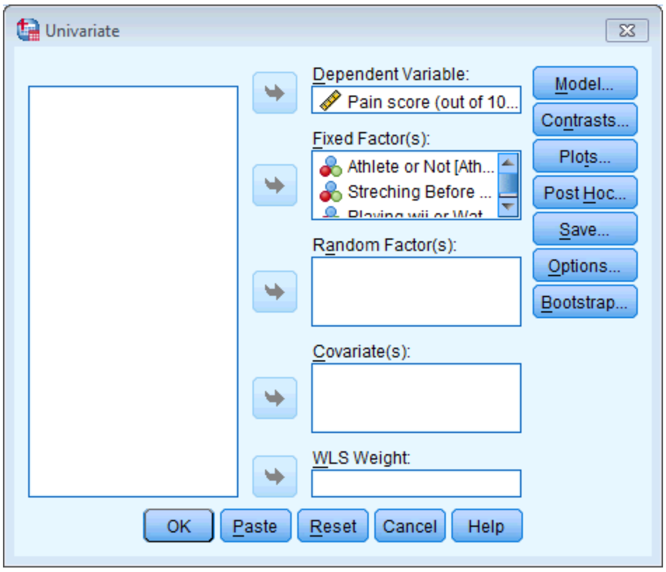

Figure 12

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Pain score (out of 10)

F	df1	df2	Sig.
2.732	7	112	.012

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Athlete + Stretch + Wii + Athlete * Stretch + Athlete * Wii + Stretch * Wii + Athlete * Stretch * Wii

Output 14

Output shows the results of Levene's test. This is significant, $F(7, 112) = 2.14$, $p < .05$, suggesting that the assumption of homogeneity of variance has been violated.

Tests of Between-Subjects Effects

Dependent Variable: Pain score (out of 10)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	302.525[a]	7	43.218	28.295	.000
Intercept	1003.408	1	1003.408	656.947	.000
Athlete	99.008	1	99.008	64.822	.000
Stretch	16.875	1	16.875	11.048	.001
Wii	85.008	1	85.008	55.656	.000
Athlete * Stretch	1.875	1	1.875	1.228	.270
Athlete * Wii	69.008	1	69.008	45.181	.000
Stretch * Wii	21.675	1	21.675	14.191	.000
Athlete * Stretch * Wii	9.075	1	9.075	5.942	.016
Error	171.067	112	1.527		
Total	1477.000	120			
Corrected Total	473.592	119			

a. R Squared = .639 (Adjusted R Squared = .616)

Output 15

Output is the main ANOVA table. The results show that there was a significant main effect of **Athlete**, $F(1, 112) = 64.82$, $p < .001$. To help us interpret this significant effect we could plot an error bar chart of the mean pain score in athletes and non-athletes (not taking into account whether or not they stretched and whether or not they played on the Wii or watched). The resulting graph is shown in Figure and reveals that, on average, athletes had significantly lower injury scores than non-athletes.

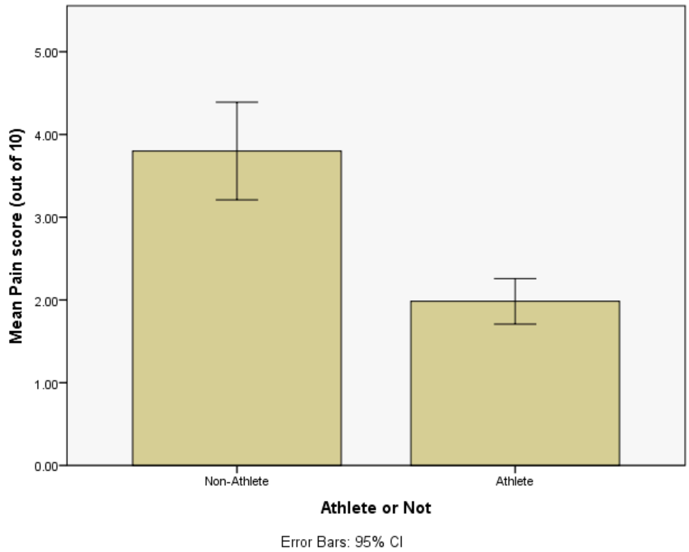

Figure 13

Output also reveals a significant main effect of **Stretch**, $F(1, 112) = 11.05$, $p < .01$. The graph in Figure shows that stretching significantly decreased injury score compared to not stretching. However, the two-way interaction graph (**Error! Reference source not found.**Figure 17) shows us that this is true only for athletes and non-athletes who played on the Wii, not for those in the control group (you can also see this pattern in the three-way interaction graph, Figure 19). This is an example of how main effects can sometimes be misleading.

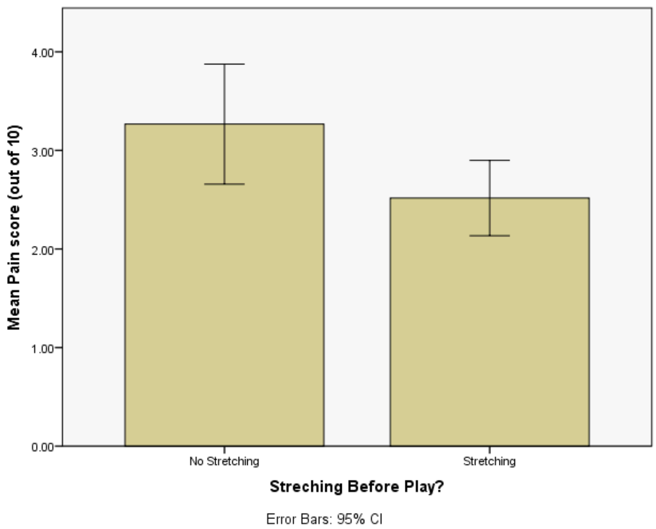

Figure 14

There was also a significant main effect of **Wii**, *F*(1, 112) = 55.66, *p* < .001. Figure tells us (not surprisingly) that playing on the Wii resulted in a significantly higher injury score compared to watching other people playing on the Wii (control).

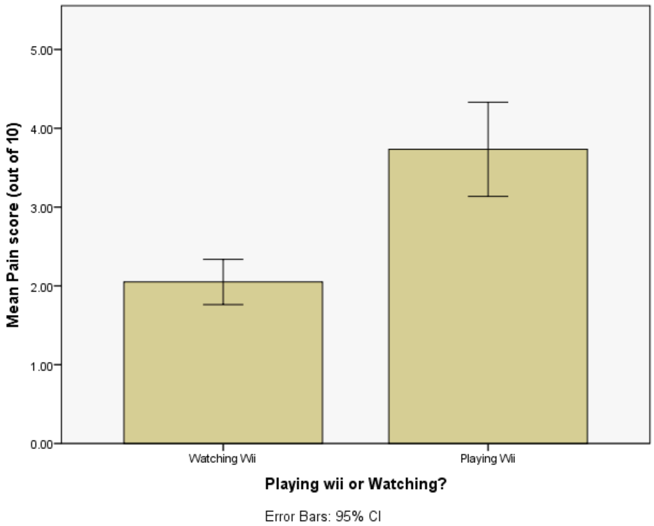

Figure 15

There was not a significant **Athlete** × **Stretch** interaction *F*(1, 112) = 1.23, *p* > .05. The interaction effect (Figure) shows that (not taking into account playing vs. watching the Wii) while non-athletes had higher injury scores than athletes overall, stretching decreased the number of injuries in both athletes and non-athletes by roughly the same amount. Parallel lines usually indicate a non-significant interaction effect, and so it is not surprising that the interaction between stretch and athlete was non-significant.

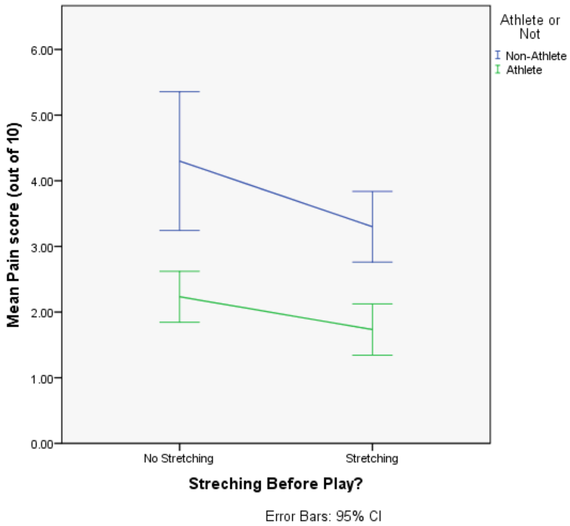

Figure 16

There was a significant **Athlete** × **Wii** interaction *F*(1, 112) = 45.18, *p* < .001. The interaction graph (Figure) shows that (not taking stretching into account) when playing on the Wii, non-athletes suffered significantly higher injury scores than athletes. However, when watching other people playing on the Wii, athletes and non-athletes had very similar injury scores.

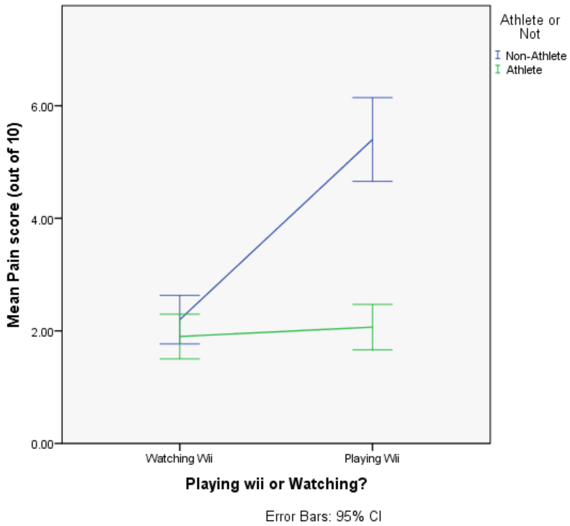

Figure 17

There was a significant **Stretch** × **Wii** interaction *F*(1, 112) = 14.19, *p* < .001. Figure shows that (not taking athlete into account) stretching before playing on the Wii significantly decreased injury scores, but stretching before watching other people playing on the Wii did not significantly reduce injury scores. This is not surprising as watching other people playing on the Wii is unlikely to result in sports injury!

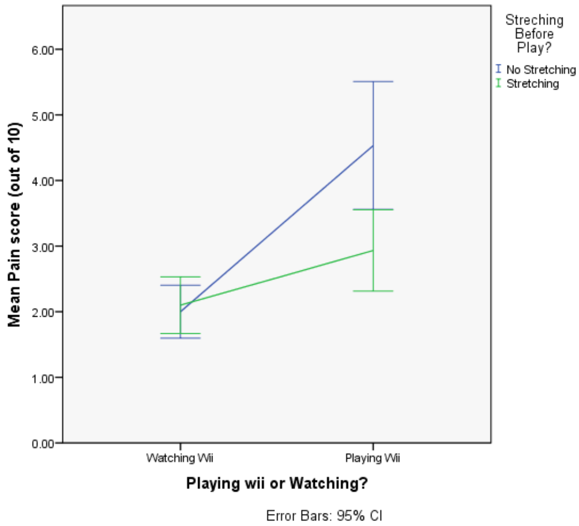

Figure 18

There was a significant **Athlete** × **Stretch** × **Wii** interaction $F(1, 112) = 5.94$, $p < .05$. What this actually means is that the effect of stretching and playing on the Wii on injury score was different for athletes than it was for non-athletes. In the presence of this significant interaction it makes no sense to interpret the main effects. Figure shows the interaction graph for this three-way effect. I produced this graph using the statistics package R (Field, Miles, & Field, 2012). However, you could produce two similar graphs using SPSS by first splitting the file by the variable **Athlete**. The graph shows that for athletes, stretching and playing on the Wii has very little effect: their mean injury score is quite stable across the two conditions (whether they played on the Wii or watched other people playing on the Wii, stretched or did no stretching). However, for the non-athletes, watching other people play on the Wii compared to not stretching and playing on the Wii rapidly declines their mean injury score. The interaction tells us that stretching and watching rather than playing on the Wii both result in a lower injury score and that this is true only for non-athletes. In short, the results show that athletes are able to minimize their injury level regardless of whether they stretch before exercise or not, whereas non-athletes only have to bend slightly and they get injured!….although I wonder if we would get the same results using the Arsenal football team. ☺

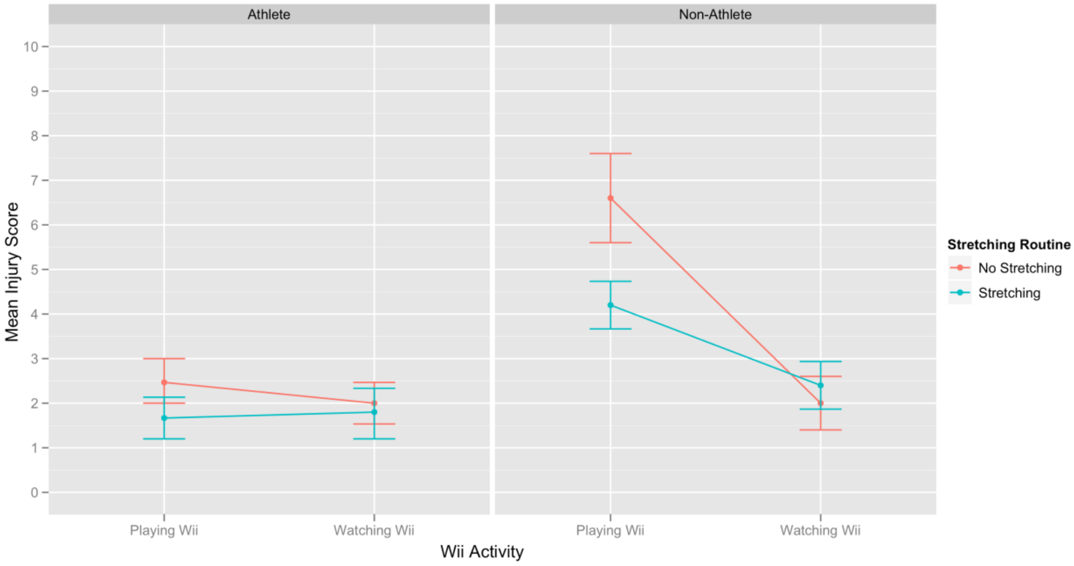
Athlete
Non-Athlete
Mean Injury Score
10
9
8
7
6
5
4
3
2
1
0
Playing Wii
Watching Wii
Playing Wii
Watching Wii
Wii Activity
Stretching Routine
No Stretching
Stretching

Figure 19