Chapter 15: Mixed design ANOVA

Smart Alex's Solutions

Task 1

In the previous chapter we looked at an example in which participants viewed a total of nine mock adverts over three sessions. In these adverts there were three products (a brand of beer, Brain Death; a brand of wine, Dangleberry; and a brand of water, Puritan). These could be presented alongside positive, negative or neutral imagery. Over the three sessions and nine adverts, each type of product was paired with each type of imagery (read the previous chapter if you need more detail). After each advert participants rated the drinks on a scale ranging from −100 (dislike very much) through 0 (neutral) to 100 (like very much). The design had two repeated-measures independent variables: the type of drink (beer, wine or water) and the type of imagery used (positive, negative or neutral). Imagine that we also knew each participant's gender. Men and women might respond differently to the products (because, in keeping with stereotypes, men might mostly drink lager whereas women might drink wine). Reanalyze the data, taking gender (a between-group variable) into account. The data are in the file ***MixedAttitude.sav****. Run a three-way mixed ANOVA on these data.*

Running the analysis

To carry out the analysis in SPSS, follow the same instructions that we did before. First of all, access the *define factors* dialog box by using the file path Analyze General Linear Model Repeated Measures... . We are using the same repeated-measures variables as in Chapter 13 of the book, so complete this dialog box exactly as shown there, and then click on Define to access the main dialog box. This box should be completed exactly as before, except that we must specify **gender** as a between-group variable by selecting it in the variables list and clicking ➡ to transfer it to the box labelled *Between-Subjects Factors* (Figure 1).

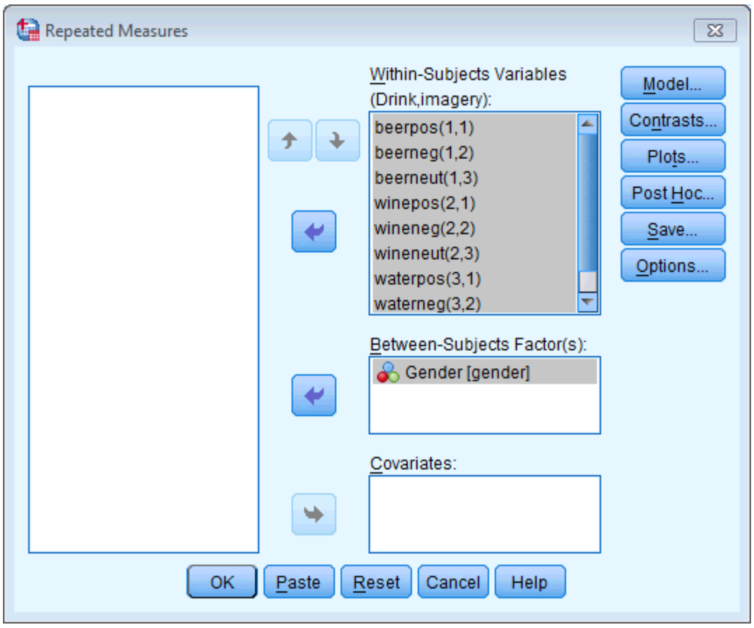

Figure 1

Gender has only two levels (male or female), so there is no need to specify contrasts for this variable; however, you should select simple contrasts for both **drink** and **imagery**. The addition of a between-group factor means that we can select *post hoc* tests for this variable by clicking on Post Hoc... . This action brings up the *post hoc* test dialog box, which can be used as previously explained. However, we need not specify any *post hoc* tests here because the between-group factor has only two levels. The addition of an extra variable makes it necessary to choose a different graph than the one in the previous example. Click on Plots... to access the dialog box and place **drink** and **imagery** in the same slots as for the previous example, but also place **gender** in the slot labelled *Separate Plots*. When all three variables have been specified, don't forget to click on Add to add this combination to the list of plots. By asking SPSS to plot the drink × imagery × gender interaction, we should get the same interaction graph as before, except that a separate version of this graph will be produced for male and female subjects.

As far as other options are concerned, you should select the same ones that were chosen in Chapter 13. It is worth selecting estimated marginal means for all effects (because these values will help you to understand any significant effects), but to save space I did not ask for confidence intervals for these effects because we have considered this part of the output in some detail already. When all of the appropriate options have been selected, run the analysis.

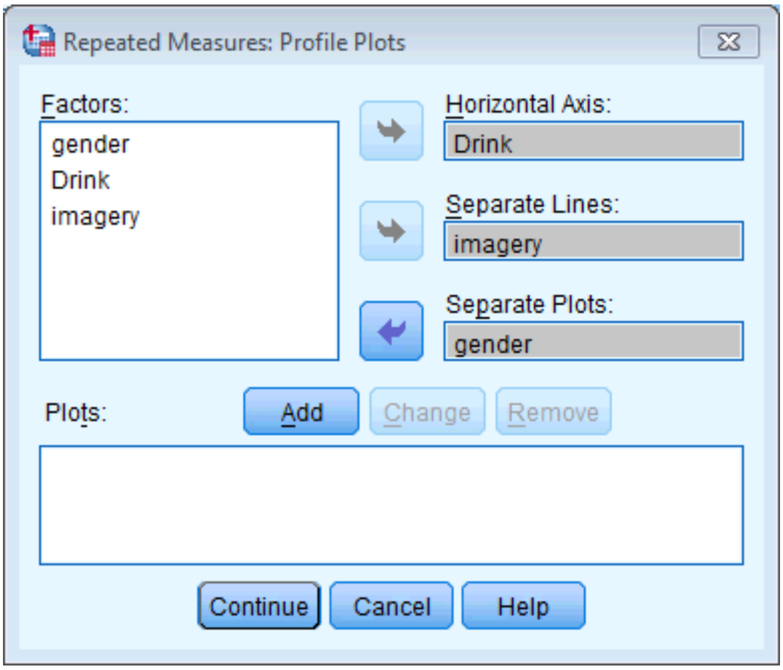

Figure 2

Overall analysis

The initial output is the same as in the two-way ANOVA example: there is a table listing the repeated-measures variables from the data editor and the level of each independent variable that they represent. The second table, shown in Output 1, contains descriptive statistics (mean and standard deviation) for each of the nine conditions split according to whether participants were male or female. The names in this table are the names I gave the variables in the data editor (therefore, your output may differ slightly). These descriptive statistics are interesting because they show us the pattern of means across all experimental conditions (so, we use these means to produce the graphs of the three-way interaction). We can see that the variability among scores was greatest when beer was used as a product, and that when a corpse image was used the ratings given to the products were negative (as expected) for all conditions except the men in the beer condition. Likewise, ratings of products were very positive when a sexy person was used as the imagery, irrespective of the gender of the participant or the product being advertised.

Descriptive Statistics

	Gender	Mean	Std. Deviation	N
Beer + Sexy	Male	24.8000	14.0063	10
	Female	17.3000	11.3925	10
	Total	21.0500	13.0080	20
Beer + Corpse	Male	20.1000	7.8379	10
	Female	-11.2000	5.1381	10
	Total	4.4500	17.3037	20
Beer + Person in Armchair	Male	16.9000	8.5434	10
	Female	3.1000	6.7074	10
	Total	10.0000	10.2956	20
Wine + Sexy	Male	22.3000	7.6311	10
	Female	28.4000	4.1150	10
	Total	25.3500	6.7378	20
Wine + Corpse	Male	-7.8000	4.9396	10
	Female	-16.2000	4.1312	10
	Total	-12.0000	6.1815	20
Wine + Person in Armchair	Male	7.5000	4.9721	10
	Female	15.8000	4.3919	10
	Total	11.6500	6.2431	20
Water + Sexy	Male	14.5000	6.7864	10
	Female	20.3000	6.3953	10
	Total	17.4000	7.0740	20
Water + Corpse	Male	-9.8000	6.7791	10
	Female	-8.6000	7.1368	10
	Total	-9.2000	6.8025	20
Water + Person in Armchair	Male	-2.1000	6.2973	10
	Female	6.8000	3.8816	10
	Total	2.3500	6.8386	20

Output 1

The results of Mauchly's sphericity test (Output 2) are different from the example in Chapter 13, because the between-group factor is now being accounted for by the test. The main effect of drink still significantly violates the sphericity assumption (W = 0.572, p = .009) but the main effect of imagery no longer does. Therefore, the F-value for the main effect of drink (and its interaction with the between-group variable **gender**) needs to be corrected for this violation.

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
DRINK	.572	9.486	2	.009	.700	.784	.500
IMAGERY	.965	.612	2	.736	.966	1.000	.500
DRINK * IMAGERY	.609	8.153	9	.521	.813	1.000	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the layers (by default) of the Tests of Within Subjects Effects table.

b. Design: Intercept+GENDER - Within Subjects Design: DRINK+IMAGERY+DRINK*IMAGERY

Output 2

The summary table of the repeated-measures effects in the ANOVA is split into sections for each of the effects in the model and their associated error terms (Output 3). The table format is the same as for the previous example, except that the interactions between gender and the repeated-measures effects are included also. We would expect to still find the effects that were previously present (in a balanced design, the inclusion of an extra variable should not affect these effects). By looking at the significance values it is clear that this prediction is true: there are still significant effects of the type of drink used, the type of imagery used, and the interaction of these two variables.

In addition to the effects already described, we find that gender interacts significantly with the type of drink used (so, men and women respond differently to beer, wine and water regardless of the context of the advert). There is also a significant interaction of gender and imagery (so, men and women respond differently to positive, negative and neutral imagery regardless of the drink being advertised). Finally, the three-way interaction between gender, imagery and drink is significant, indicating that the way in which imagery affects responses to different types of drinks depends on whether the subject is male or female. The effects of the repeated-measures variables have been outlined in Chapter 13 and the pattern of these responses will not have changed, so, rather than repeat myself, I will concentrate on the new effects and the forgetful reader should look back at Chapter 13!

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
DRINK	Sphericity Assumed	2092.344	2	1046.172	11.708	.000
	Greenhouse-Geisser	2092.344	1.401	1493.568	11.708	.001
	Huynh-Feldt	2092.344	1.567	1334.881	11.708	.000
	Lower-bound	2092.344	1.000	2092.344	11.708	.003
DRINK * GENDER	Sphericity Assumed	4569.011	2	2284.506	25.566	.000
	Greenhouse-Geisser	4569.011	1.401	3261.475	25.566	.000
	Huynh-Feldt	4569.011	1.567	2914.954	25.566	.000
	Lower-bound	4569.011	1.000	4569.011	25.566	.000
Error(DRINK)	Sphericity Assumed	3216.867	36	89.357		
	Greenhouse-Geisser	3216.867	25.216	127.571		
	Huynh-Feldt	3216.867	28.214	114.017		
	Lower-bound	3216.867	18.000	178.715		
IMAGERY	Sphericity Assumed	21628.678	2	10814.339	287.417	.000
	Greenhouse-Geisser	21628.678	1.932	11196.937	287.417	.000
	Huynh-Feldt	21628.678	2.000	10814.339	287.417	.000
	Lower-bound	21628.678	1.000	21628.678	287.417	.000
IMAGERY * GENDER	Sphericity Assumed	1998.344	2	999.172	26.555	.000
	Greenhouse-Geisser	1998.344	1.932	1034.522	26.555	.000
	Huynh-Feldt	1998.344	2.000	999.172	26.555	.000
	Lower-bound	1998.344	1.000	1998.344	26.555	.000
Error(IMAGERY)	Sphericity Assumed	1354.533	36	37.626		
	Greenhouse-Geisser	1354.533	34.770	38.957		
	Huynh-Feldt	1354.533	36.000	37.626		
	Lower-bound	1354.533	18.000	75.252		
DRINK * IMAGERY	Sphericity Assumed	2624.422	4	656.106	19.593	.000
	Greenhouse-Geisser	2624.422	3.251	807.186	19.593	.000
	Huynh-Feldt	2624.422	4.000	656.106	19.593	.000
	Lower-bound	2624.422	1.000	2624.422	19.593	.000
DRINK * IMAGERY * GENDER	Sphericity Assumed	495.689	4	123.922	3.701	.009
	Greenhouse-Geisser	495.689	3.251	152.458	3.701	.014
	Huynh-Feldt	495.689	4.000	123.922	3.701	.009
	Lower-bound	495.689	1.000	495.689	3.701	.070
Error(DRINK*IMAGERY)	Sphericity Assumed	2411.000	72	33.486		
	Greenhouse-Geisser	2411.000	58.524	41.197		
	Huynh-Feldt	2411.000	72.000	33.486		
	Lower-bound	2411.000	18.000	133.944		

Output 3

The effect of gender

The main effect of gender is listed separately from the repeated-measures effects in a table labelled *Tests of Between-Subjects Effects*. Before looking at this table it is important to check the assumption of homogeneity of variance using Levene's test (Output 4). SPSS produces a table listing Levene's test for each of the repeated-measures variables in the data editor, and we need to look for any variable that has a significant value. The table showing Levene's test indicates that variances are homogeneous for all levels of the repeated-measures variables (because all significance values are greater than .05). If any values were significant, then this would compromise the accuracy of the *F*-test for gender, and we would have to consider transforming all of our data to stabilize the variances between groups (one popular transformation is to take the square root of all values). Fortunately, in this example a transformation is unnecessary.

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

	F	df1	df2	Sig.
Beer + Sexy	1.009	1	18	.328
Beer + Corpse	1.305	1	18	.268
Beer + Person in Armchair	1.813	1	18	.195
Wine + Sexy	2.017	1	18	.173
Wine + Corpse	1.048	1	18	.320
Wine + Person in Armchair	.071	1	18	.793
Water + Sexy	.317	1	18	.580
Water + Corpse	.804	1	18	.382
Water + Person in Armchair	1.813	1	18	.195

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

a. Design: Intercept+GENDER - Within Subjects Design: DRINK+IMAGERY+DRINK*IMAGERY

Output 4

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	1246.445	1	1246.445	144.593	.000
GENDER	58.178	1	58.178	6.749	.018
Error	155.167	18	8.620		

Output 5

Output 5 shows the ANOVA summary table for the main effect of gender, and (because the significance of .018 is less than the standard cut-off point of .05) we can report that there was a significant main effect of gender, $F(1, 18) = 6.75$, $p < .05$. This effect tells us that if we ignore all other variables, male subjects' ratings were significantly different than females. If you requested that SPSS display means for the gender effect you should scan through your output and find the table in a section headed *Estimated Marginal Means*. The table of means for the main effect of gender with the associated standard errors is plotted alongside. It is clear from this graph that men's ratings were generally significantly more positive than females.

Therefore, men gave more positive ratings than women regardless of the drink being advertised and the type of imagery used in the advert.

Estimates

Measure: MEASURE_1

Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Male	9.600	.928	7.649	11.551
Female	6.189	.928	4.238	8.140

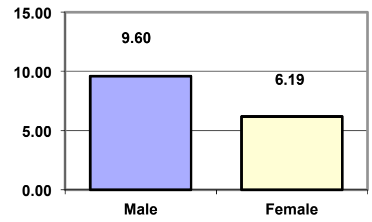

Figure 3

The interaction between gender and drink

Gender interacted in some way with the type of drink used as a stimulus. Remembering that the effect of drink violated sphericity, we must report Greenhouse–Geisser-corrected values for this interaction with the between-group factor. From the summary table (Output 3) we should report that there was a significant interaction between the type of drink used and the gender of the subject, $F(1.40, 25.22) = 25.57$, $p < .001$. This effect tells us that the type of drink being advertised had a different effect on men and women. We can use the estimated marginal means to determine the nature of this interaction (or we could have asked SPSS for a plot of gender × drink). The means and interaction graph (Figure 4) show the meaning of this result. The graph shows the average male ratings of each drink ignoring the type of imagery with which it was presented (circles). The women's scores are shown as squares. The graph clearly shows that male and female ratings are very similar for wine and water, but men seem to rate beer more highly than women — regardless of the type of imagery used. We could interpret this interaction as meaning that the type of drink being advertised influenced ratings differently in men and women. Specifically, ratings were similar for wine and water but males rated beer higher than women. This interaction can be clarified using the contrasts specified before the analysis.

2. Gender * DRINK

Measure: MEASURE_1

Gender	DRINK	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Male	1	20.600	2.441	15.471	25.729
	2	7.333	.765	5.726	8.940
	3	.867	1.414	-2.103	3.836
Female	1	3.067	2.441	-2.062	8.196
	2	9.333	.765	7.726	10.940
	3	6.167	1.414	3.197	9.136

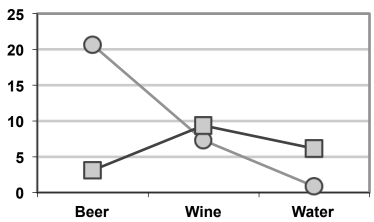

Figure 4

The interaction between gender and imagery

Gender interacted in some way with the type of imagery used as a stimulus. The effect of imagery did not violate sphericity, so we can report the uncorrected *F*-value. From the summary table (Output 3) we should report that there was a significant interaction between the type of imagery used and the gender of the subject ($F(2, 36) = 26.55$, $p < .001$). This effect tells us that the type of imagery used in the advert had a different effect on men and women. We can use the estimated marginal means to determine the nature of this interaction. The means and interaction graph (Figure 5) shows the meaning of this result. The graph shows the average male in each imagery condition ignoring the type of drink that was rated (circles). The women's scores are shown as squares. The graph clearly shows that male and female ratings are very similar for positive and neutral imagery, but men seem to be less affected by negative imagery than women — regardless of the drink in the advert. To interpret this finding more fully, we should consult the contrasts for this interaction.

3. Gender * IMAGERY

Measure: MEASURE_1

Gender	IMAGERY	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Male	1	20.533	1.399	17.595	23.471
	2	.833	1.092	-1.460	3.127
	3	7.433	1.395	4.502	10.365
Female	1	22.000	1.399	19.062	24.938
	2	-12.000	1.092	-14.293	-9.707
	3	8.567	1.395	5.635	11.498

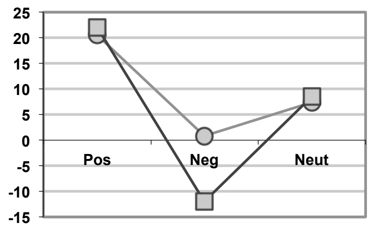

Figure 5

The interaction between drink and imagery

The interpretation of this interaction is the same as for the two-way ANOVA (see Chapter 13). You may remember that the interaction reflected the fact that negative imagery has a different effect than both positive and neutral imagery (because it decreased ratings rather than increasing them).

The interaction between gender, drink and imagery

The three-way interaction tells us whether the drink by imagery interaction is the same for men and women (i.e., whether the combined effect of the type of drink and the imagery used is the same for male subjects as for female subjects). We can conclude (Output 3) that there is a significant three-way drink × imagery × gender interaction, $F(4, 72) = 3.70$, $p = .009$. The nature of this interaction is shown up in the graph in Figure 6, which shows the imagery by drink interaction for men and women separately. The male graph shows that when positive imagery is used (circles), men generally rated all three drinks positively (the line with circles is higher than the other lines for all drinks). This pattern is true of women also (the line representing positive imagery is above the other two lines). When neutral imagery is used (triangles), men rate beer very highly, but rate wine and water fairly neutrally. Women, on the other hand rate beer and water neutrally, but rate wine more positively (in fact, the pattern of the positive and neutral imagery lines show that women generally rate wine slightly more positively than water and beer). So, for neutral imagery men still rate beer positively, and women still rate wine positively. For the negative imagery (squares), the men still rate beer very highly, but give low ratings to the other two types of drink. So, regardless of the type of imagery used, men rate beer very positively (if you look at the graph you'll note that ratings for beer are virtually identical for the three types of imagery). Women, however, rate all three drinks very negatively when negative imagery is used. The three-way interaction is, therefore, likely to reflect these sex differences in the interaction between drink and imagery. Specifically, men seem fairly immune to the effects of imagery when beer is being used as a stimulus, whereas women are not. The contrasts will show up exactly what this interaction represents.

4. Gender * DRINK * IMAGERY

Measure: MEASURE_1

Gender	DRINK	IMAGERY	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Male	1	1	24.800	4.037	16.318	33.282
		2	20.100	2.096	15.697	24.503
		3	16.900	2.429	11.797	22.003
	2	1	22.300	1.939	18.227	26.373
		2	-7.800	1.440	-10.825	-4.775
		3	7.500	1.483	4.383	10.617
	3	1	14.500	2.085	10.119	18.881
		2	-9.800	2.201	-14.424	-5.176
		3	-2.100	1.654	-5.575	1.375
Female	1	1	17.300	4.037	8.818	25.782
		2	-11.200	2.096	-15.603	-6.797
		3	3.100	2.429	-2.003	8.203
	2	1	28.400	1.939	24.327	32.473
		2	-16.200	1.440	-19.225	-13.175
		3	15.800	1.483	12.683	18.917
	3	1	20.300	2.085	15.919	24.681
		2	-8.600	2.201	-13.224	-3.976
		3	6.800	1.654	3.325	10.275

Male Female

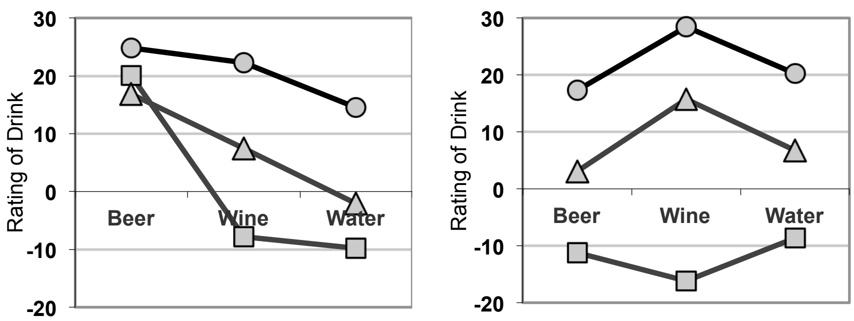

Figure 6 Graphs showing the drink by imagery interaction for men and women. Lines represent positive imagery (circles), negative imagery (squares) and neutral imagery (triangles)

Contrasts for repeated-measures variables

We requested simple contrasts for the **drink** variable (for which water was used as the control category) and for the **imagery** category (for which neutral imagery was used as the control category). The table (Output 5) is the same as for the previous example except that the added effects of **gender** and its interaction with other variables are now included. So, for the main effect of drink, the first contrast compares level 1 (beer) against the base category (in this case, the last category, water); this result is significant, $F(1, 18) = 15.37$, $p < .01$. The next contrast compares level 2 (wine) with the base category (water) and confirms the significant difference found when gender was not included as a variable in the analysis, $F(1, 18) = 19.92$, $p < .001$. For the imagery main effect, the first contrast compares level 1 (positive) to the base category (neutral) and verifies the significant effect found by the *post hoc* tests, $F(1, 18) = 134.87$, $p < .001$. The second contrast confirms the significant difference found for the negative imagery condition compared to the neutral, $F(1, 18) = 129.18$, $p < .001$. No contrast was specified for gender.

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	DRINK	IMAGERY	Type III Sum of Squares	df	Mean Square	F	Sig.
DRINK	Level 1 vs. Level 3		1383.339	1	1383.339	15.371	.001
	Level 2 vs. Level 3		464.006	1	464.006	19.923	.000
DRINK * GENDER	Level 1 vs. Level 3		2606.806	1	2606.806	28.965	.000
	Level 2 vs. Level 3		54.450	1	54.450	2.338	.144
Error(DRINK)	Level 1 vs. Level 3		1619.967	18	89.998		
	Level 2 vs. Level 3		419.211	18	23.290		
IMAGERY		Level 1 vs. Level 3	3520.089	1	3520.089	134.869	.000
		Level 2 vs. Level 3	3690.139	1	3690.139	129.179	.000
IMAGERY * GENDER		Level 1 vs. Level 3	.556	1	.556	.021	.886
		Level 2 vs. Level 3	975.339	1	975.339	34.143	.000
Error(IMAGERY)		Level 1 vs. Level 3	469.800	18	26.100		
		Level 2 vs. Level 3	514.189	18	28.566		
DRINK * IMAGERY	Level 1 vs. Level 3	Level 1 vs. Level 3	320.000	1	320.000	1.686	.211
		Level 2 vs. Level 3	720.000	1	720.000	8.384	.010
	Level 2 vs. Level 3	Level 1 vs. Level 3	36.450	1	36.450	.223	.642
		Level 2 vs. Level 3	2928.200	1	2928.200	31.698	.000
DRINK * IMAGERY * GENDER	Level 1 vs. Level 3	Level 1 vs. Level 3	441.800	1	441.800	2.328	.144
		Level 2 vs. Level 3	480.200	1	480.200	5.592	.029
	Level 2 vs. Level 3	Level 1 vs. Level 3	4.050	1	4.050	.025	.877
		Level 2 vs. Level 3	405.000	1	405.000	4.384	.051
Error(DRINK*IMAGERY)	Level 1 vs. Level 3	Level 1 vs. Level 3	3416.200	18	189.789		
		Level 2 vs. Level 3	3416.200	18	189.789		
	Level 2 vs. Level 3	Level 1 vs. Level 3	1545.800	18	85.878		
		Level 2 vs. Level 3	1662.800	18	92.378		

Output 5

Drink × gender interaction 1: beer vs. water, male vs. female

The first interaction term looks at level 1 of drink (beer) compared to level 3 (water), comparing male and female scores. This contrast is highly significant, $F(1, 18) = 28.97$, $p < .001$. This result tells us that the increased ratings of beer compared to water found for men are not found for women. So, in the graph the squares representing female ratings of beer and water are roughly level; however, the circle representing male ratings of beer is much higher than the circle representing water. The positive contrast represents this difference, and so we can conclude that male ratings of beer (compared to water) were significantly greater than women's ratings of beer (compared to water).

Drink × gender interaction 2: wine vs. water, male vs. female

The second interaction term compares level 2 of drink (wine) to level 3 (water), contrasting male and female scores. There is no significant difference for this contrast, $F(1, 18) = 2.34$, $p = 0.14$, which tells us that the difference between ratings of wine compared to water in males is roughly the same as in females.

Therefore, overall, the drink × gender interaction has shown up a difference between males and females in how they rate beer (regardless of the type of imagery used).

Imagery × gender interaction 1: positive vs. neutral, male vs. female

The first interaction term looks at level 1 of imagery (positive) compared to level 3 (neutral), comparing male and female scores. This contrast is not significant ($F < 1$). This result tells us that ratings of drinks presented with positive imagery (relative to those presented with neutral imagery) were equivalent for males and females. This finding represents the fact that in the earlier graph of this interaction the squares and circles for both the positive and neutral conditions overlap (therefore male and female responses were the same).

Imagery × gender interaction 2: negative vs. neutral, male vs. female

The second interaction term looks at level 2 of imagery (negative) compared to level 3 (neutral), comparing male and female scores. This contrast is highly significant, $F(1, 18) = 34.13$, $p < .001$. This result tells us that the difference between ratings of drinks paired with negative imagery compared to neutral was different for men and women. Looking at the earlier graph of this interaction, this finding represents the fact that for men, ratings of drinks paired with negative imagery were relatively similar to ratings of drinks paired with neutral imagery (the circles have a fairly similar vertical position). However, if you look at the female ratings, then drinks were rated much less favourably when presented with negative imagery than when presented with neutral imagery (the square in the negative condition is much lower than the neutral condition).

Therefore, overall, the imagery × gender interaction has shown up a difference between males and females in terms of their ratings of drinks presented with negative imagery compared to neutral; specifically, men seem less affected by negative imagery.

Drink × imagery × gender interaction 1: beer vs. water, positive vs. neutral imagery, male vs. female

The first interaction term compares level 1 of drink (beer) to level 3 (water), when positive imagery (level 1) is used compared to neutral (level 3) in males compared to females, $F(1, 18) = 2.33$, $p = .144$. The non-significance of this contrast tells us that the difference in ratings when positive imagery is used compared to neutral imagery is roughly equal when beer is used as a stimulus and when water is used, and these differences are equivalent in male and female subjects. In terms of the interaction graph it means that the distance between the circle and the triangle in the beer condition is the same as the distance between the circle and the triangle in the water condition and that these distances are equivalent in men and women.

Drink × imagery × gender interaction 2: beer vs. water, negative vs. neutral imagery, male vs. female

The second interaction term looks at level 1 of drink (beer) compared to level 3 (water), when negative imagery (level 2) is used compared to neutral (level 3). This contrast is significant, $F(1, 18) = 5.59$, $p < .05$. This result tells us that the difference in ratings between beer and water when negative imagery is used (compared to neutral imagery) is different between men and women. If we plot ratings of beer and water across the negative and neutral conditions, for males (circles) and females (squares) separately, we see that ratings after negative imagery are always lower than ratings for neutral imagery except for men's ratings of beer, which are actually higher after negative imagery. As such, this contrast tells us that the interaction effect reflects a difference in the way in which males rate beer compared to females when negative imagery is used compared to neutral. Males and females are similar in their pattern of ratings for water but different in the way in which they rate beer.

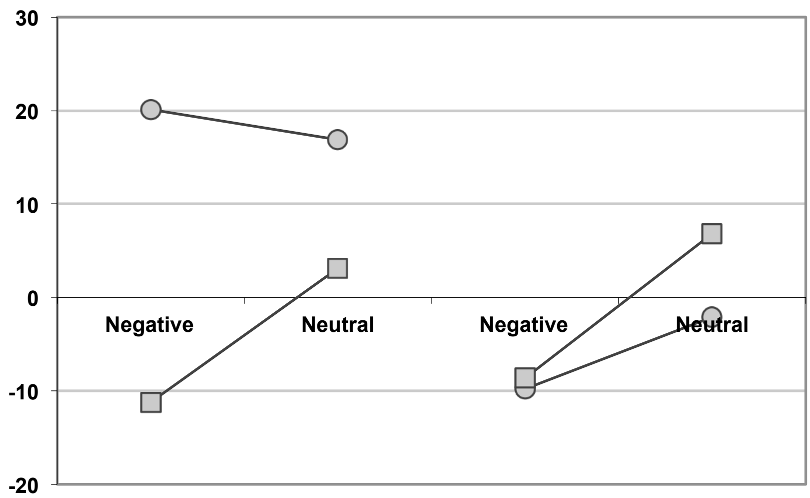

Figure 7

Drink × imagery × gender interaction 3: wine vs. water, positive vs. neutral imagery, male vs. female.

The third interaction term looks at level 2 of drink (wine) compared to level 3 (water), when positive imagery (level 1) is used compared to neutral (level 3) in males compared to females. This contrast is non-significant, $F(1, 18) < 1$. This result tells us that the difference in ratings when positive imagery is used compared to neutral imagery is roughly equal when wine is used as a stimulus and when water is used, and these differences are equivalent in male and female subjects. In terms of the interaction graph it means that the distance between the circle and the triangle in the wine condition is the same as the distance between the circle and the triangle in the water condition and that these distances are equivalent in men and women.

Drink × imagery × gender interaction 4: wine vs. water, negative vs. neutral imagery, male vs. female

The final interaction term looks at level 2 of drink (wine) compared to level 3 (water), when negative imagery (level 2) is used compared to neutral (level 3). This contrast is very close to significance, $F(1, 18) = 4.38$, $p = .051$. This result tells us that the difference in ratings between wine and water when negative imagery is used (compared to neutral imagery) is different between men and women (although this difference has not quite reached significance). If we plot ratings of wine and water across the negative and neutral conditions, for males (circles) and females (squares), we see that ratings after negative imagery are always lower than ratings for neutral imagery, but for women rating wine the change is much more dramatic (the line is steeper). As such, this contrast tells us that the interaction effect reflects a difference in the way in which females rate wine differently to males when neutral imagery is used compared to when negative imagery is used. Males and females are similar in their pattern of ratings for water but different in the way in which they rate wine. It is noteworthy that this contrast was not significant using the usual .05 level; however, it is worth remembering that this cut-off point was set in a fairly arbitrary way, and so it is worth reporting these close effects and letting your reader decide whether they are meaningful or not. There is also a growing trend towards reporting effect sizes in preference to using significance levels.

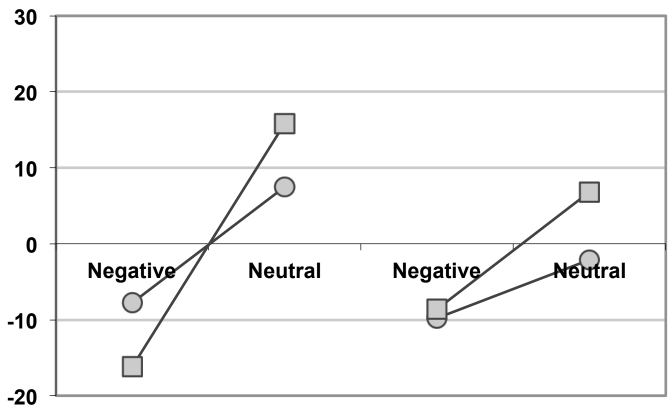

Figure 2

Summary

These contrasts again tell us nothing about the differences between the beer and wine conditions (or the positive and negative conditions) and different contrasts would have to be run to find out more. However, what is clear so far is that differences exist between men and women in terms of their ratings of beer and wine. It seems as though men are relatively unaffected by negative imagery when it comes to beer. Likewise, women seem more willing to rate wine positively when neutral imagery is used than men are. What should be clear from this is that complex ANOVA in which several independent variables are used results in complex interaction effects that require a great deal of concentration to interpret (imagine interpreting a four-way interaction!). Therefore, it is essential to take a systematic approach to interpretation, and plotting graphs is a particularly useful way to proceed. It is also advisable to think carefully about the appropriate contrasts to use to answer the questions you have about your data. It is these contrasts that will help you to interpret interactions, so make sure you select sensible ones.

Task 2

> *Text messaging and Twitter encourage communication using abbreviated forms of words (if u no wat I mean). A researcher wanted to see the effect this had on children's understanding of grammar. One group of 25 children was encouraged to send text messages on their mobile phones over a six-month period. A second group of 25 was forbidden from sending text messages for the same period (to ensure adherence, this group were given armbands that administered painful shocks in the presence of a phone signal). The outcome was a score on a grammatical test (as a percentage) that was measured both before and after the experiment. The data are in the file* ***TextMessages.sav****. Does using text messages affect grammar?*

The line chart (with error bars) in Figure 9 shows the grammar data. The circles show the mean grammar score before and after the experiment for the text message group and the controls. The means before and after are connected by a line for the two groups separately. It's clear from this chart that in the text message group grammar scores went down dramatically over the six-month period in which they used their mobile phone. For the controls, their grammar scores also fell but much less dramatically.

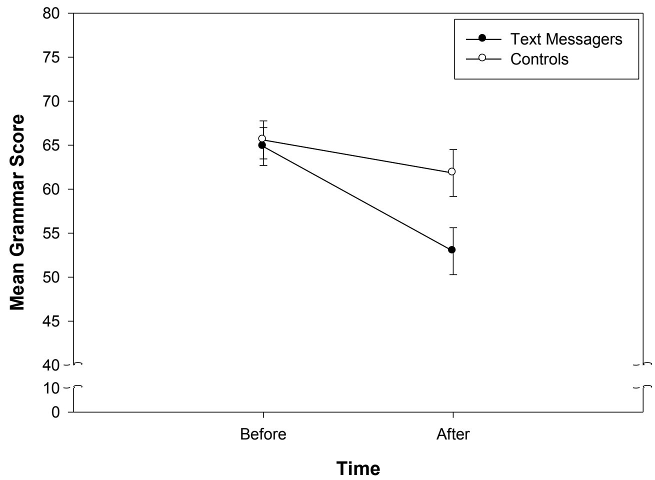

Figure 9: Line chart (with error bars showing the standard error of the mean) of the mean grammar scores before and after the experiment for text messagers and controls

Descriptive Statistics

	Group	Mean	Std. Deviation	N
Grammer at Time 1	Text Messagers	64.8400	10.67973	25
	Controls	65.6000	10.83590	25
	Total	65.2200	10.65467	50
Grammar at Time 2	Text Messagers	52.9600	16.33116	25
	Controls	61.8400	9.41046	25
	Total	57.4000	13.93278	50

Output 6

Output 6 shows the table of descriptive statistics from the two-way mixed ANOVA; the table has means at time 1 split according to whether the people were in the text messaging group or the control group, and then the means for the two groups at time 2. These means correspond to those plotted in Figure 9.

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

					Epsilon[a]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	1.000	.000	0	.	1.000	1.000	1.000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept+GROUP
Within Subjects Design: TIME

Output 7

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

	F	df1	df2	Sig.
Grammer at Time 1	.089	1	48	.767
Grammar at Time 2	3.458	1	48	.069

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

a. Design: Intercept+GROUP
Within Subjects Design: TIME

Output 8

We know that when we use repeated-measures we have to check the assumption of sphericity (Output 7). We also know that for independent designs we need to check the homogeneity of variance assumption. If the design is a mixed design then we have both repeated and independent measures, so we have to check both assumptions. In this case, we have only two levels of the repeated measure so the assumption of sphericity does not apply. Levene's test produces a different test for each level of the repeated-measures variable (Output 8). In mixed designs, the homogeneity assumption has to hold for every level of the repeated-measures variable. At both levels of time, Levene's test is non-significant (p = 0.77 before the experiment and p = .069 after the experiment). This means the assumption has not been broken at all (but it was quite close to being a problem after the experiment).

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	1528.810	1	1528.810	15.457	.000
	Greenhouse-Geisser	1528.810	1.000	1528.810	15.457	.000
	Huynh-Feldt	1528.810	1.000	1528.810	15.457	.000
	Lower-bound	1528.810	1.000	1528.810	15.457	.000
TIME * GROUP	Sphericity Assumed	412.090	1	412.090	4.166	.047
	Greenhouse-Geisser	412.090	1.000	412.090	4.166	.047
	Huynh-Feldt	412.090	1.000	412.090	4.166	.047
	Lower-bound	412.090	1.000	412.090	4.166	.047
Error(TIME)	Sphericity Assumed	4747.600	48	98.908		
	Greenhouse-Geisser	4747.600	48.000	98.908		
	Huynh-Feldt	4747.600	48.000	98.908		
	Lower-bound	4747.600	48.000	98.908		

Output 9

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	375891.610	1	375891.610	1933.002	.000
GROUP	580.810	1	580.810	2.987	.090
Error	9334.080	48	194.460		

Output 10

Outputs 9 and 10 shows the main ANOVA summary tables. Like any two-way ANOVA, we still have three effects to find: two main effects (one for each independent variable) and one interaction term. The main effect of time is significant, so we can conclude that grammar scores were significantly affected by the time at which they were measured. The exact nature of this effect is easily determined because there were only two points in time (and so this main effect is comparing only two means). Figure 10 shows that grammar scores were higher before the experiment than after. So, before the experimental manipulation scores were higher than after, meaning that the manipulation had the net effect of significantly reducing grammar scores. This main effect seems rather interesting until you consider that these means include both text messagers and controls. There are three possible reasons for the drop in grammar scores: (1) the text messagers got worse and are dragging down the mean after the experiment; (2) the controls somehow got worse; or (3) the whole group just got worse and it had nothing to do with whether the children text-messaged or not. Until we examine the interaction, we won't see which of these is true.

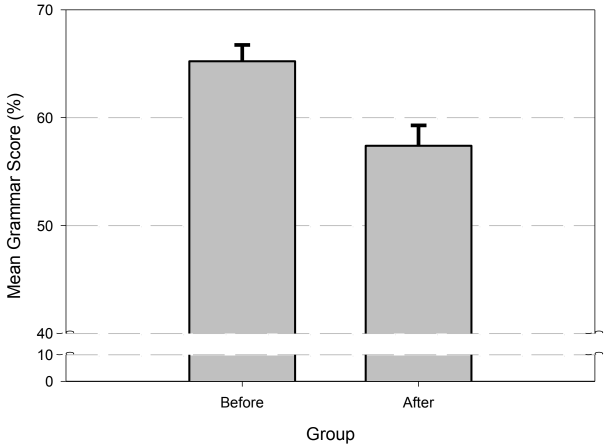

Figure 10

The main effect of group is shown by the *F*-ratio in Output 10. The probability associated with this *F*-ratio is .09, which is just above the critical value of .05. Therefore, we must conclude that there was no significant main effect on grammar scores of whether children text-messaged or not. Again, this effect seems interesting enough, and mobile phone companies might certainly choose to cite it as evidence that text messaging does not affect your grammatical ability. However, remember that this main effect ignores the time at which grammatical ability is measured. It just means that if we took the average grammar score for text messagers (that's including their score both before and after they started using their phone), and compared this to the mean of the controls (again including scores before and after) then these means would not be significantly different. The graph shows that when you ignore the time at which grammar was measured, the controls have slightly better grammar than the text messagers, but not significantly so.

Main effects are not always that interesting and should certainly be viewed in the context of any interaction effects. The interaction effect in this example is shown by the *F*-ratio in the row labelled **Time*Group**, and because the probability of obtaining a value this big by chance is .047, which is just less than the criterion of .05, we can say that there is a significant interaction between the time at which grammar was measured and whether or not children were allowed to text-message within that time. The mean ratings in all conditions help us to interpret this effect. The significant interaction tells us that the change in grammar scores was significantly different in text messagers compared to controls. Looking at the interaction graph, we can see that although grammar scores fell in controls, the drop was much more marked in the text messagers; so, text messaging does seem to ruin your ability at grammar compared to controls.[1]

[1] It's interesting that the control group means dropped too. This could be because the control group were undisciplined and still used their mobile phones, or it could just be that the education system in this country is so underfunded that there is no one to teach English anymore!

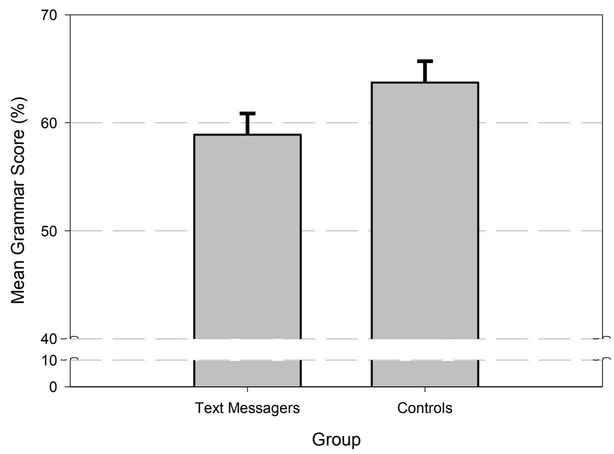

Figure 11

Writing the result

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

- ✓ The results show that the grammar ratings at the end of the experiment were significantly lower than those at the beginning of the experiment, $F(1, 48) = 15.46$, $p < .001$, $r = .61$.
- ✓ The main effect of group on the grammar scores was non-significant, $F(1, 48) = 2.99$, $p = .09$, $r = .27$. This indicated that when the time at which grammar was measured is ignored, the grammar ability in the text message group was not significantly different from the controls.
- ✓ The time × group interaction was significant, $F(1, 48) = 4.17$, $p = .047$, $r = .34$, indicating that the change in grammar ability in the text message group was significantly different from the change in the control groups. These findings indicate that although there was a natural decay of grammatical ability over time (as shown by the controls) there was a much stronger effect when participants were encouraged to use text messages. This shows that using text messages accelerates the inevitable decline in grammatical ability.

Task 3

A researcher hypothesized that Big Brother *(see Chapter 1) contestants start off with personality disorders that are exacerbated by being forced to live with people as attention-seeking as themselves. To test this hypothesis, she gave eight contestants a questionnaire measuring personality disorders before and after they entered the house. A second group of eight people were given the questionnaires at the same time, these people were short listed to go into the house, but never actually went in. The data are in* ***BigBrother.sav****. Does the* Big Brother *house give you a personality disorder?*

Running the analysis

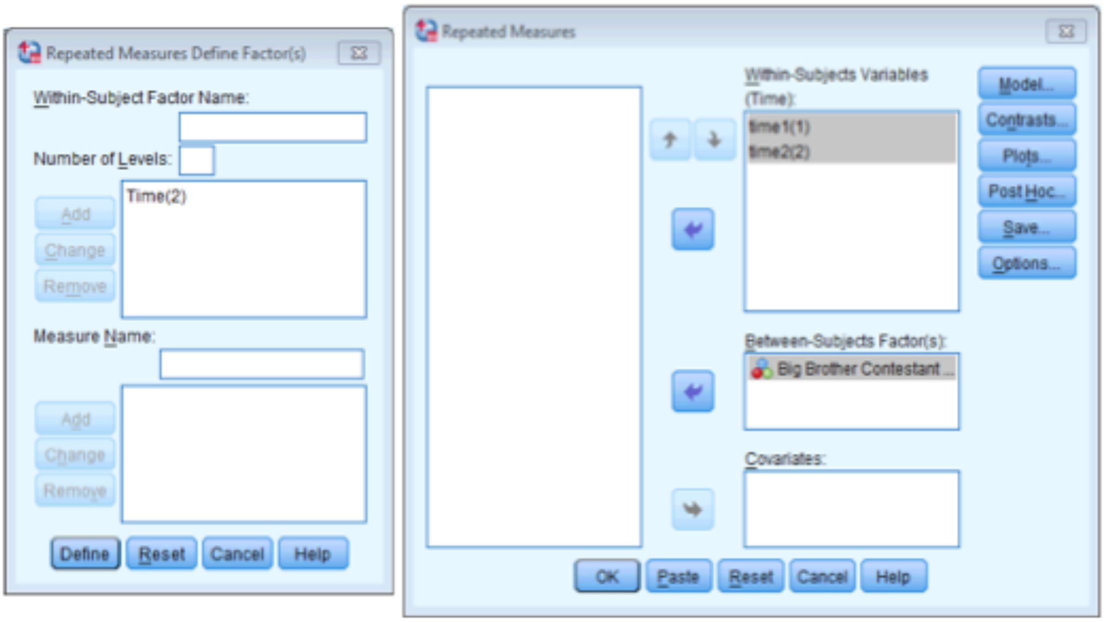

Figure 12

SPSS output

The bar graph (with error bars) in Figure 13 shows the *Big Brother* data. It's clear from this graph that in the *Big Brother* contestant group the mean personality disorder score increased from time 1 (before entering the house) to time 2 (after leaving the house). However, in the no treatment control group the mean personality disorder score decreased from time 1 to time 2.

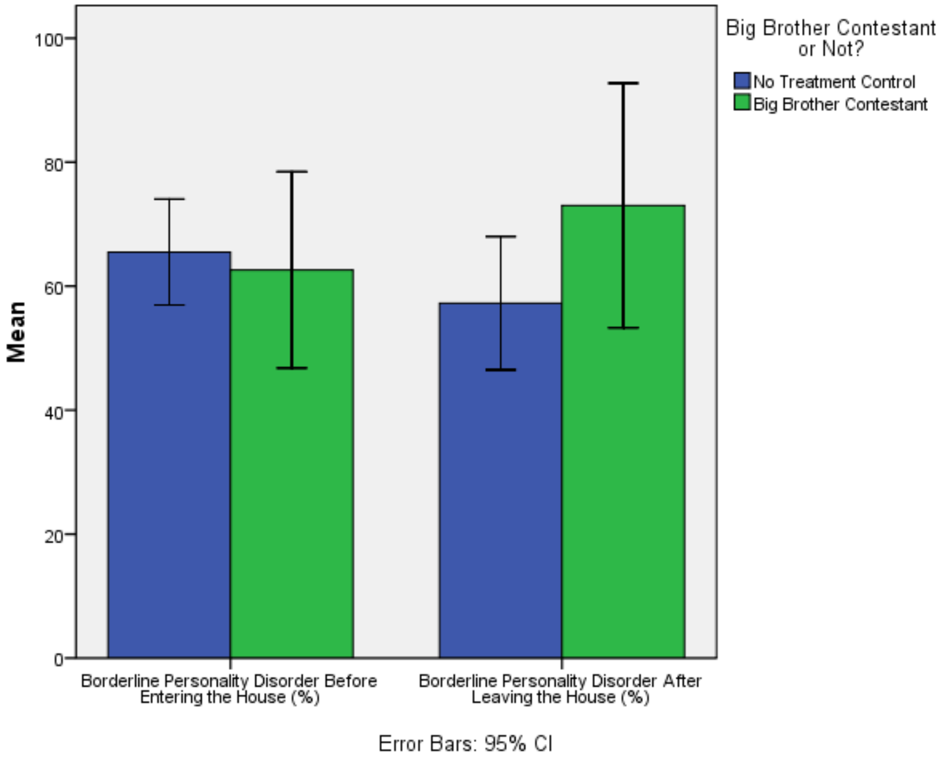

Figure 13: Error bar chart of mean personality disorder score before entering and after leaving the *Big Brother* house

Descriptive Statistics

	Big Brother Contest...	Mean	Std. Deviation	N
Borderline Personality Disorder Before Entering the House (%)	No Treatment Control	65.50	10.212	8
	Big Brother Contestant	62.62	18.951	8
	Total	64.06	14.780	16
Borderline Personality Disorder After Leaving the House (%)	No Treatment Control	57.25	12.870	8
	Big Brother Contestant	73.00	23.598	8
	Total	65.12	20.083	16

Output 11

Output 11 shows the table of descriptive statistics from the two-way mixed ANOVA; the table has mean borderline personality disorder (BPD) scores before entering the *Big Brother* house split according to whether the people were a contestant or not, and then the means for the two groups after leaving the house. These means correspond to those plotted in Figure 13.

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subject...	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	1.000	.000	0	.	1.000	1.000	1.000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept + bb
Within Subjects Design: Time

Output 12

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

	F	df1	df2	Sig.
Borderline Personality Disorder Before Entering the House (%)	4.151	1	14	.061
Borderline Personality Disorder After Leaving the House (%)	3.356	1	14	.088

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

a. Design: Intercept + bb
Within Subjects Design: Time

Output 13

We know that when we use repeated-measures we have to check the assumption of sphericity (Output 12). However, we also know that for sphericity to be an issue we need at least three conditions. We have only two conditions here so sphericity does not need to be tested (and, therefore, SPSS produces a blank in the column labeled *Sig.*). We also need to check the homogeneity of variance assumption (Output 13). Levene's test produces a different test for each level of the repeated-measures variable. In mixed designs, the homogeneity assumption has to hold for every level of the repeated-measures variable. At both levels of time, Levene's test is non-significant (p = 0.061 before entering the *Big Brother* house and p = .088 after leaving). This means the assumption has not been significantly broken (but it was quite close to being a problem).

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Sphericity Assumed	9.031	1	9.031	.093	.765
	Greenhouse-Geisser	9.031	1.000	9.031	.093	.765
	Huynh-Feldt	9.031	1.000	9.031	.093	.765
	Lower-bound	9.031	1.000	9.031	.093	.765
Time * bb	Sphericity Assumed	693.781	1	693.781	7.149	.018
	Greenhouse-Geisser	693.781	1.000	693.781	7.149	.018
	Huynh-Feldt	693.781	1.000	693.781	7.149	.018
	Lower-bound	693.781	1.000	693.781	7.149	.018
Error(Time)	Sphericity Assumed	1358.688	14	97.049		
	Greenhouse-Geisser	1358.688	14.000	97.049		
	Huynh-Feldt	1358.688	14.000	97.049		
	Lower-bound	1358.688	14.000	97.049		

Output 14

Tests of Between-Subjects Effects

Measure:MEASURE_1
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	133515.281	1	133515.281	269.235	.000
bb	331.531	1	331.531	.669	.427
Error	6942.688	14	495.906		

Output 15

Output 14 and 15 show the main ANOVA summary tables. Like any two-way ANOVA, we still have three effects to find: two main effects (one for each independent variable) and one interaction term. The main effect of time is not significant, so we can conclude that BPD scores were significantly affected by the time at which they were measured. The exact nature of this effect is easily determined because there were only two points in time (and so this main effect is comparing only two means). Figure 14 shows that BPD scores were not significantly different after leaving the *Big Brother* house compared to before entering it.

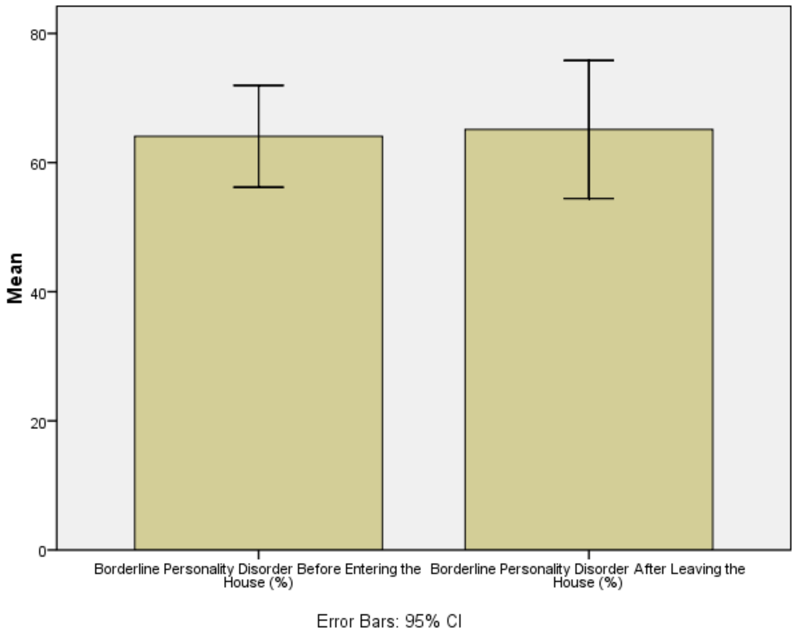

Figure 14

The main effect of group (**bb**) is shown by the *F*-ratio in Output 15. The probability associated with this *F*-ratio is .43, which is above the critical value of .05. Therefore, we must conclude that there was no significant main effect on BPD scores of whether the person was a *Big Brother* contestant or not. The graph shows that when you ignore the time at which BPD was measured, the contestants and controls are not significantly different.

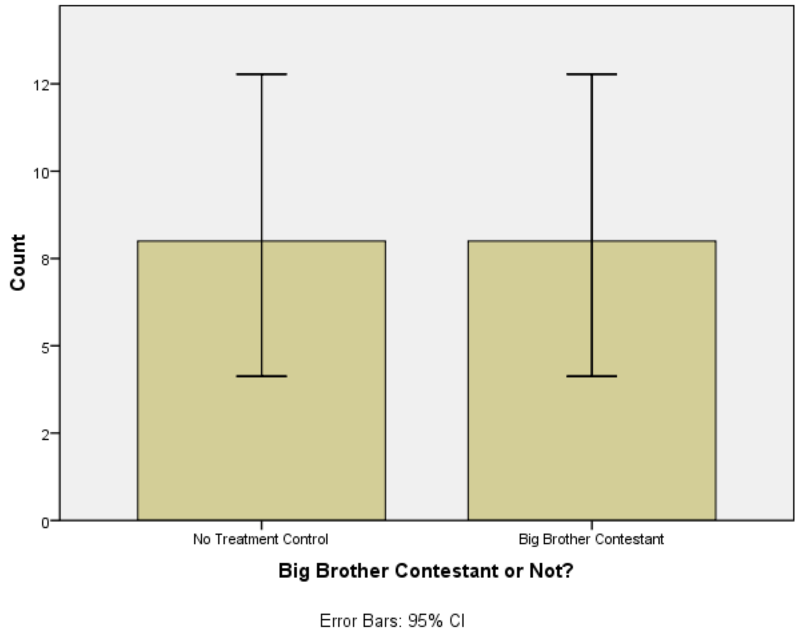

Figure 15

The interaction effect in this example is shown by the *F*-ratio in the row labelled **Time*bb** in Output 14, and because the probability of obtaining a value this big is .018, which is less than the criterion of .05, we can say that there is a significant interaction between the time at which BPD was measured and whether or not the person was a contestant. The mean ratings in all conditions (and on the interaction graph) help us to interpret this effect. The significant interaction seems to indicate that for controls BPD scores went down (slightly) from before entering the house to after leaving it, but for contestants these opposite is true: BPD scores increased over time.

Writing the results

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

- ✓ The main effect of group was not significant, $F(1, 14) = 0.67$, $p = .43$, indicating that across both time points borderline personality disorder scores were similar in *Big Brother* contestants and controls.
- ✓ The main effect of time was not significant, $F(1, 14) = 0.09$, $p = .77$, indicating that across all participants borderline personality disorder scores were similar before entering the house and after leaving it.
- ✓ The time × group interaction was significant, $F(1, 14) = 7.15$, $p = .018$, indicating that although borderline personality disorder scores decreased for controls from before entering the house to after leaving it, scores *increased* for the contestants.

Task 4

Angry Birds is a video game in which you fire birds at pigs. Some daft people think this sort of thing makes people more violent. A (fabricated) study was set up in which people played Angry Birds and a control game (Tetris) over a two-year period (one year per game). They were put in a pen of pigs for a day before the study, and after 1 month, 6 months and 12 months. Their violent acts towards the pigs were counted. The data are in the file ***Angry Pigs.sav****. Does playing Angry Birds make people more violent to pigs compared to a control game?*

To answer this question we need to conduct a 2 (**BaselineGame**: Angry Birds vs. Tetris) × 4 (**Time**: Baseline, 1 month, 6 months and 12 months) two-way mixed ANOVA with repeated measures on the time variable. Your completed dialog boxes should look like Figure and 17.

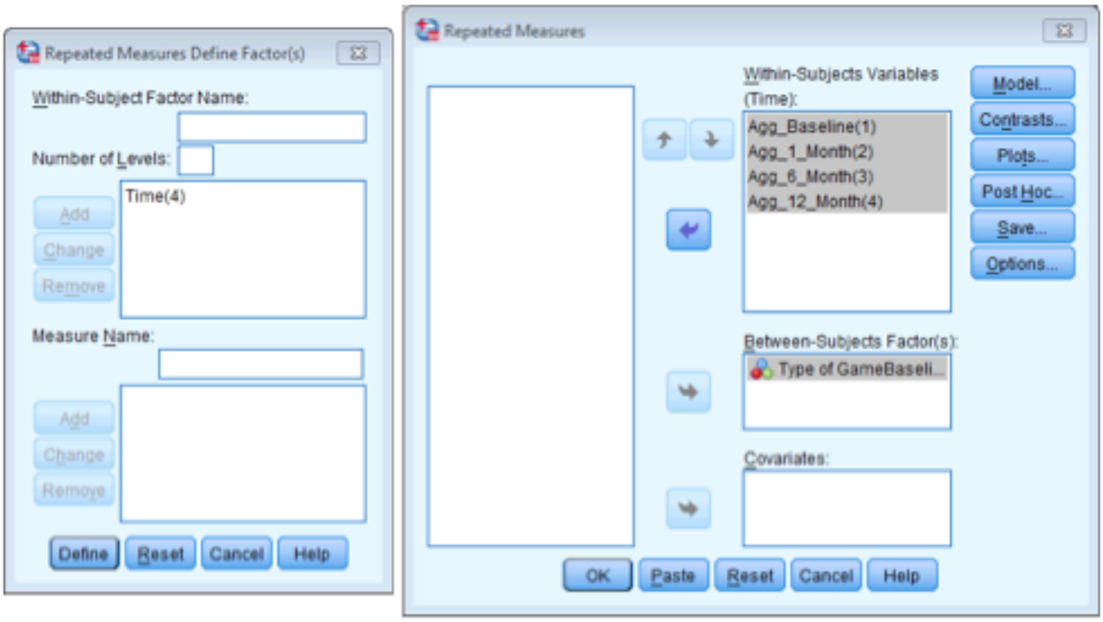

Figure 16

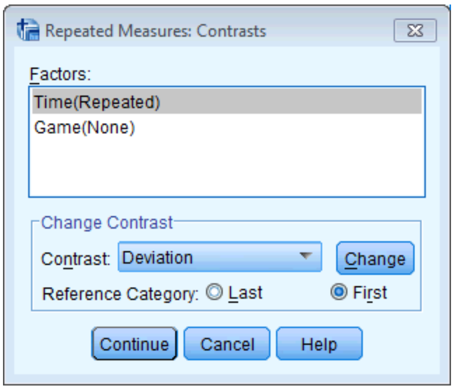

Figure 17

SPSS output

Figure is a line graph of the angry pigs data. We can see that when participants played Tetris (blue line) in general their aggressive behaviour towards pigs decreased over time (except for between 6 months and 12 months when it actually increased slightly). However, when participants played Angry Birds, their aggressive behaviour towards pigs increased over time.

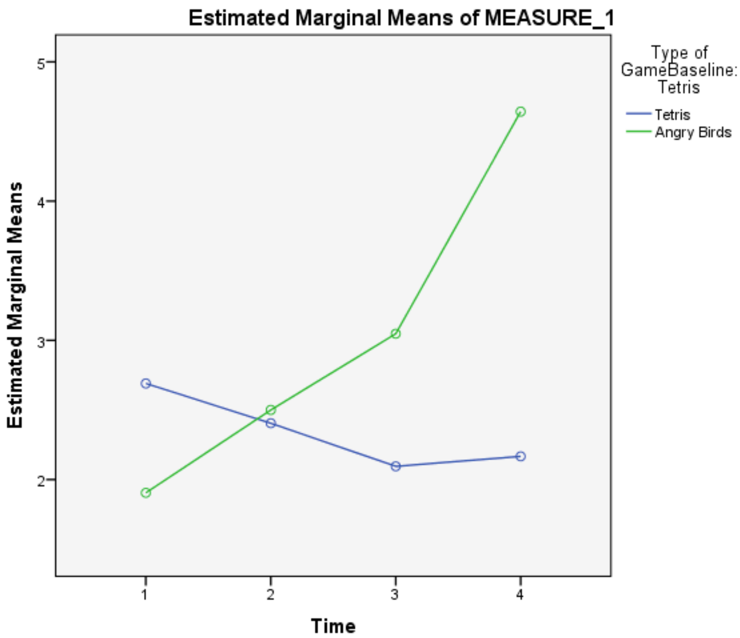

Figure 18

Type of GameBaseline: Tetris * Time

Measure: MEASURE_1

Type of GameBaseline: Tetris	Time	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tetris	1	2.690	.181	2.331	3.050
	2	2.405	.226	1.954	2.855
	3	2.095	.257	1.584	2.607
	4	2.167	.296	1.578	2.755
Angry Birds	1	1.905	.181	1.545	2.264
	2	2.500	.226	2.050	2.950
	3	3.048	.257	2.536	3.559
	4	4.643	.296	4.054	5.231

Output 16

Output shows the estimated marginal means for the interaction between Tetris and time. These values correspond with those plotted in Figure .

Descriptive Statistics

	Type of GameBaseline: Tetris	Mean	Std. Deviation	N
Baseline Aggression	Tetris	2.69	1.370	42
	Angry Birds	1.90	.932	42
	Total	2.30	1.230	84
1 Month Aggression	Tetris	2.40	1.191	42
	Angry Birds	2.50	1.700	42
	Total	2.45	1.460	84
6 Months Aggression	Tetris	2.10	1.165	42
	Angry Birds	3.05	2.048	42
	Total	2.57	1.724	84
12 Months Aggression	Tetris	2.17	1.286	42
	Angry Birds	4.64	2.387	42
	Total	3.40	2.277	84

Output 17

Output shows the table of descriptive statistics from the two-way mixed ANOVA.

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Epsilon[b] Huynh-Feldt	Epsilon[b] Lower-bound
Time	.908	7.768	5	.170	.938	.987	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + Game
Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Output 18

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

	F	df1	df2	Sig.
Baseline Aggression	10.423	1	82	.002
1 Month Aggression	6.554	1	82	.012
6 Months Aggression	5.538	1	82	.021
12 Months Aggression	14.106	1	82	.000

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

a. Design: Intercept + Game
Within Subjects Design: Time

Output 19

We know that when we use repeated-measures we have to check the assumption of sphericity. We also know that for independent designs we need to check the homogeneity of variance assumption. If the design is a mixed design then we have both repeated and independent measures, so we have to check both assumptions. Output shows the results of Mauchly's test for our repeated-measures variable **Time**. The value in the column labelled *Sig* is .170, which is larger than the cut off of .05, therefore it is non-significant and the assumption of sphericity has been met. Levene's test produces a different test for each level of the repeated-measures variable. In mixed designs, the homogeneity assumption has to hold for every level of the repeated-measures variable. Output reveals that at each level of the variable **Time**, Levene's test is significant ($p < 0.05$ in every case). This means the assumption has been broken.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Sphericity Assumed	61.747	3	20.582	8.924	.000
	Greenhouse-Geisser	61.747	2.815	21.933	8.924	.000
	Huynh-Feldt	61.747	2.961	20.852	8.924	.000
	Lower-bound	61.747	1.000	61.747	8.924	.004
Time * Game	Sphericity Assumed	121.604	3	40.535	17.574	.000
	Greenhouse-Geisser	121.604	2.815	43.194	17.574	.000
	Huynh-Feldt	121.604	2.961	41.065	17.574	.000
	Lower-bound	121.604	1.000	121.604	17.574	.000
Error(Time)	Sphericity Assumed	567.399	246	2.306		
	Greenhouse-Geisser	567.399	230.855	2.458		
	Huynh-Feldt	567.399	242.822	2.337		
	Lower-bound	567.399	82.000	6.919		

Output 20

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	604.019	1	604.019	789.896	.000
Game	9.840	1	9.840	12.868	.001
Error	62.704	82	.765		

Output 21

Output and 21 show the main ANOVA summary tables. Like any two-way ANOVA, we still have three effects to find: two main effects (one for each independent variable) and one interaction term. The main effect of **Game** was significant, indicating that (ignoring the time at which the aggression scores were measured), the type of game being played significantly affected participant's aggression towards pigs. The main effect of **Time** was also significant, so we can conclude that (ignoring the type of game being played), aggression was significantly different at different points in time. However, the effect that we are most interested in is the **Time × Game** interaction, which was also significant. This effect tells us that changes in aggression scores over time were different when participants played Tetris compared to when they played Angry Birds. Looking at the graph in Figure , we can see that for Angry Birds, aggression scores increase over time, whereas for Tetris, aggression scores decreased over time. To investigate the exact nature of this interaction effect we can look at some contrasts.

I chose to use the repeated contrast (Figure), which compares aggression scores for the two games at each time point against the previous time point. The results of these contrasts are in Output .

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Level 1 vs. Level 2	2.012	1	2.012	.575	.450
	Level 2 vs. Level 3	1.190	1	1.190	.255	.615
	Level 3 vs. Level 4	58.333	1	58.333	12.624	.001
Time * Game	Level 1 vs. Level 2	16.298	1	16.298	4.661	.034
	Level 2 vs. Level 3	15.429	1	15.429	3.300	.073
	Level 3 vs. Level 4	48.762	1	48.762	10.553	.002
Error(Time)	Level 1 vs. Level 2	286.690	82	3.496		
	Level 2 vs. Level 3	383.381	82	4.675		
	Level 3 vs. Level 4	378.905	82	4.621		

Output 22

We are most interested in the **Time × Game** interaction. We can see that the first contrast (Level 1 vs. Level 2) was significant, *p* = .034, indicating that the change in aggression scores from the baseline to 1 month was significantly different for Tetris and Angry birds. If we look at the graph in Figure , we can see that on average, aggression scores decreased from baseline to 1 month when participants played Tetris. However, aggression scores increased from baseline to 1 month when participants played Angry Birds. The second contrast (Level 2 vs. Level 3) was non-significant (p = .073), indicating that the change in aggression scores from 1 month to 6 months was similar when participants played Tetris compared to when they played Angry Birds. Looking at the graph, we can see that aggression scores increased for Angry Birds but decreased for Tetris – according to the contrast, not significantly so. The final contrast (Level 3 vs. Level 4) was significant, p = .002. Again looking at the graph, we can see that for Angry Birds aggression scores increased dramatically from 6 to 12 months, whereas for Tetris they stayed fairly stable.

Writing the result

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

- ✓ The results show that the aggression scores were significantly higher when participants played Angry Birds compared to when they played Tetris, $F(1, 82) = 12.87$, $p = .001$.
- ✓ The main effect of Time on the aggression scores was significant, $F(3, 246) = 8.92$, $p < .001$. This indicated that when the game which participants played is ignored, aggressive behaviour was significantly different across the four time points.
- ✓ The time × game interaction was significant, $F(3, 246) = 17.57$, $p < .001$, indicating that the change in aggression scores when participants played Tetris was significantly different from the change in aggression scores when they played Angry Birds. Looking at the line graph, we can see that these findings indicate that when participants played Tetris, their aggressive behaviour towards pigs significantly decreased over time, whereas when they played Angry birds their aggressive behaviour towards pigs significantly increased over time.

Task 5

*A different study was conducted with the same design as in Task 4. The only difference was that the participants' violent acts in real life were monitored before the study, and after 1 month, 6 months and 12 months. Does playing Angry Birds make people more violent in general compared to a control game? (**Angry Real.sav**)*

We need to run the same analysis on the **Angry Real.sav** data as we did in Task 4.

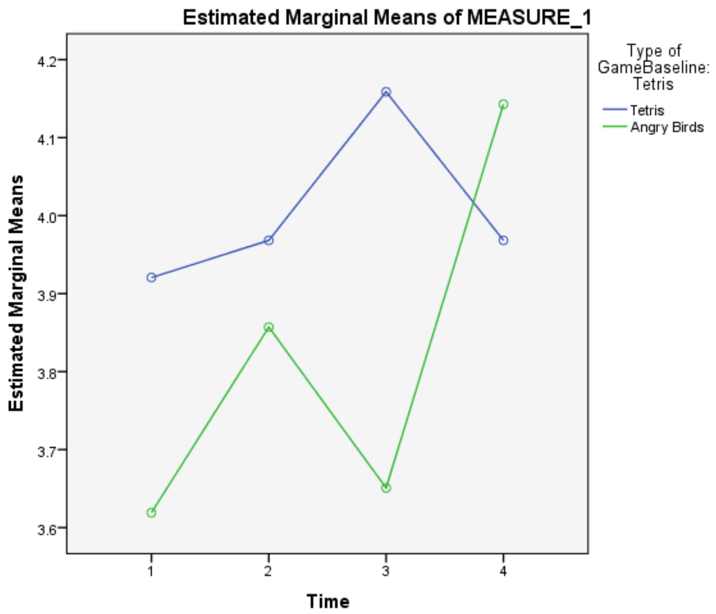

Figure 19

Looking at Figure in comparison to Figure , we can see that aggressive behaviour in the real world was more erratic for the two video games than aggressive behaviour towards pigs. Figure shows that for Tetris, aggressive behaviour in the real world increased from time 1 (baseline) to time 3 (6 months) and then decreased from time 3 (6 months) to time 4 (12 months). For Angry Birds, on the other hand, aggressive behaviour in the real world initially increased from baseline to 1 month, it then decreased from 1 month to 6 months and then dramatically increased from 6 months to 12 months. The graph also shows that more aggressive behaviour was displayed when participants played Tetris compared to when they played Angry Birds at baseline, 1 month and 6 months; however, at 12 months participants engaged in more aggressive behaviour when playing Angry Birds compared to Tetris.

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	.982	2.288	5	.808	.988	1.000	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + Game
Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Output 23

Output shows the results of Mauchly's test for our repeated measures variable **Time**. The value in the column labelled *Sig* is .808 which is larger than the cut off of .05, therefore it is non-significant and the assumption of sphericity has been met.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Sphericity Assumed	5.149	3	1.716	.867	.458
	Greenhouse-Geisser	5.149	2.963	1.738	.867	.457
	Huynh-Feldt	5.149	3.000	1.716	.867	.458
	Lower-bound	5.149	1.000	5.149	.867	.354
Time * Game	Sphericity Assumed	7.958	3	2.653	1.340	.261
	Greenhouse-Geisser	7.958	2.963	2.686	1.340	.261
	Huynh-Feldt	7.958	3.000	2.653	1.340	.261
	Lower-bound	7.958	1.000	7.958	1.340	.249
Error(Time)	Sphericity Assumed	736.643	372	1.980		
	Greenhouse-Geisser	736.643	367.360	2.005		
	Huynh-Feldt	736.643	372.000	1.980		
	Lower-bound	736.643	124.000	5.941		

Output 24

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	1927.004	1	1927.004	3332.044	.000
Game	1.096	1	1.096	1.895	.171
Error	71.712	124	.578		

Output 25

Output and Output show the main ANOVA summary tables. The main effect of **Game** was non-significant, indicating that (ignoring the time at which the aggression scores were measured), the type of game being played did not significantly affect participants' aggression in the real world. The main effect of **Time** was also non-significant, so we can conclude that (ignoring the type of game being played), aggression was not significantly different at different points in time. The effect that we are most interested in is the **Time × Game** interaction, which was again non-significant. This effect tells us that change in aggression scores over time were not significantly different when participants played Tetris compared to when they played Angry Birds. Because none of the effects were significant it doesn't make sense to conduct any contrasts. Therefore, we can conclude that playing Angry Birds does not make people more violent in general, just towards pigs.

Task 6

My wife believes that she has received less friend requests from random men on Facebook since she changed her profile picture to a photo of us both. Imagine we took 40 women who had profiles on a social networking website; 17 of them had a relationship *status of 'single' and the remaining 23 had their status as 'in a relationship' (**relationship_status**). We asked these women to set their profile picture to a photo of them on their own (**alone**) and to count how many friend request they got from men over 3 weeks, then to switch it to a photo of them with a man (**couple**) and record their friend requests from random men over 3 weeks. The data are in the file **ProfilePicture.sav**. Run a mixed ANOVA to see if friend requests are affected by relationship status and type of profile picture.*

We need to run a 2 (**relationship_status**: single vs. in a relationship) × 2(**photo**: couple vs. alone) mixed ANOVA with repeated measures on the second variable. Your completed dialog box should look like Figure .

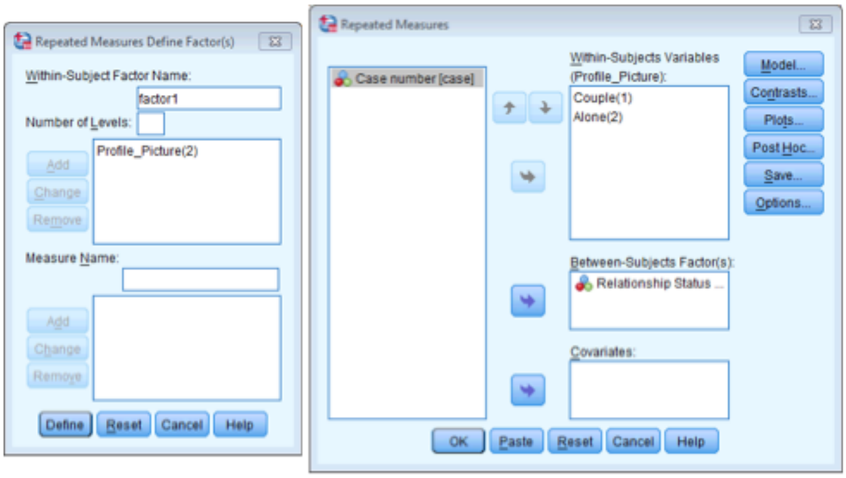

Figure 20

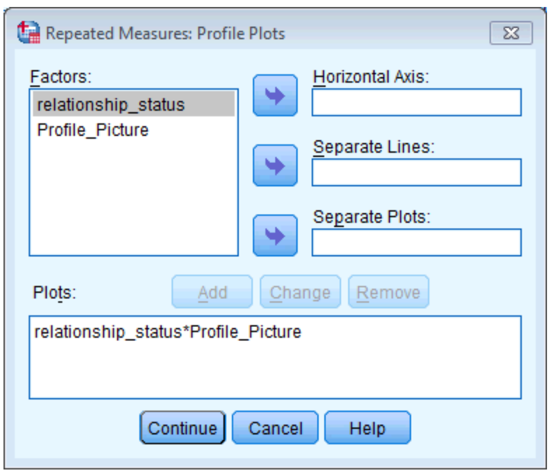

Figure 21

I also requested a line graph of the two-way interaction between relationship status and profile picture to help us interpret the results of the ANOVA (Figure). There is no need to request contrasts or *post ho*c tests because both of our independent variables have only two levels and therefore we will be able to interpret the direction of any effects by looking at the graph or the estimated marginal means.

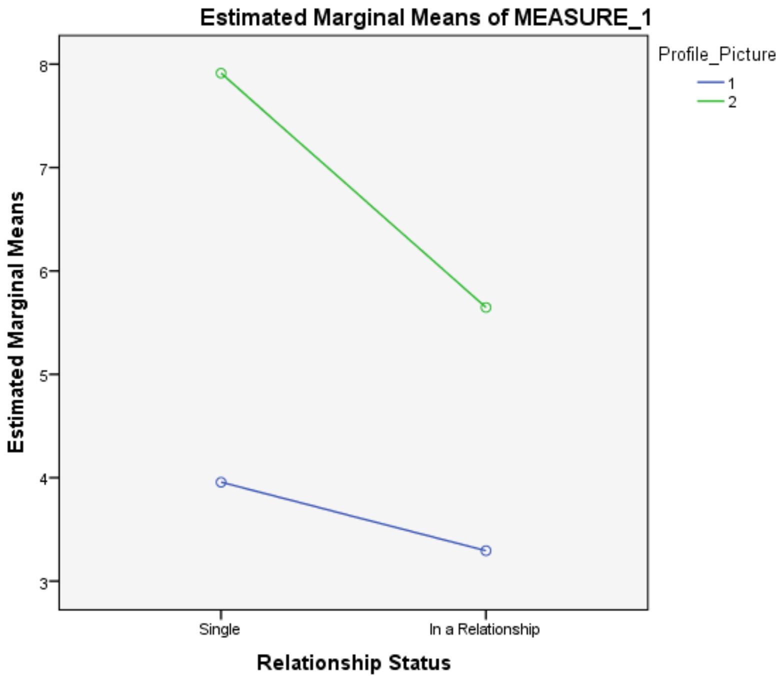

Figure 22: Line chart of the mean number of friends of single women and women in a relationship when displaying a photo of themselves alone (green line) and with a partner (blue line).

Figure indicates that in both photo conditions, single women received more friend requests than women who were in a relationship. The number of friend requests increased in both single women and those who were in a relationship when they displayed a profile picture of themselves alone (green line) compared to with a partner (blue line). However, for single women this increase was greater than for women who were in a relationship.

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Profile_Picture	1.000	.000	0	.	1.000	1.000	1.000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + relationship_status
Within Subjects Design: Profile_Picture

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Output 26

We know that when we use repeated measures we have to check the assumption of sphericity. However, we also know that for sphericity to be an issue we need at least three conditions. We have only two conditions here so sphericity does not need to be tested. Therefore, if we look at the results of Mauchly's test (Output), we can see that in the column labelled *Sig.* there is simply a full stop to indicate that the sphericity assumption does not apply to these data.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Profile_Picture	Sphericity Assumed	194.568	1	194.568	114.773	.000
	Greenhouse-Geisser	194.568	1.000	194.568	114.773	.000
	Huynh-Feldt	194.568	1.000	194.568	114.773	.000
	Lower-bound	194.568	1.000	194.568	114.773	.000
Profile_Picture * relationship_status	Sphericity Assumed	12.568	1	12.568	7.414	.010
	Greenhouse-Geisser	12.568	1.000	12.568	7.414	.010
	Huynh-Feldt	12.568	1.000	12.568	7.414	.010
	Lower-bound	12.568	1.000	12.568	7.414	.010
Error(Profile_Picture)	Sphericity Assumed	64.419	38	1.695		
	Greenhouse-Geisser	64.419	38.000	1.695		
	Huynh-Feldt	64.419	38.000	1.695		
	Lower-bound	64.419	38.000	1.695		

Output 27

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	2116.713	1	2116.713	822.655	.000
relationship_status	41.913	1	41.913	16.289	.000
Error	97.775	38	2.573		

Output 28

Output and Output show the main ANOVA summary tables. Like any two-way ANOVA, we still have three effects to find: two main effects (one for each independent variable) and one interaction term. The main effect of **relationship_status** is significant, so we can conclude that, ignoring the type of profile picture, the number of friend requests was significantly affected by the relationship status of the woman. The exact nature of this effect is easily determined because there were only two levels of relationship status (and so this main effect is comparing only two means).

1. Relationship Status

Measure: MEASURE_1

Relationship Status	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	5.935	.237	5.456	6.414
In a Relationship	4.471	.275	3.914	5.027

Output 29

If we look at the estimated marginal means for relationship status in Output , we can see that the number of friend requests was significantly higher for single women (M = 5.94) compared to women who were in a relationship (M = 4.47).

The main effect of **Profile_Picture** is shown by the F-ratio in Output . The probability associated with this F-ratio is shown as .000, which is smaller than the critical value of .05. Therefore, we can conclude that when ignoring relationship status, there was a significant main effect of whether the person was alone in their profile picture or with a partner on the number of friend requests.

2. Profile_Picture

Measure: MEASURE_1

Profile_Picture	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	3.625	.188	3.245	4.006
2	6.780	.272	6.230	7.330

Output 30

Looking at the estimated marginal means for the profile picture variable (Output), we can see that the number of friend requests was significantly higher when women were alone in their profile picture (*M* = 6.78) than when they were with a partner (*M* = 3.63). Note: we know that 1 = 'in a couple' and 2 = 'alone' in Output because this is how we coded the levels of the profile picture variable in the *define* dialog box in Figure .

The interaction effect is the effect that we are most interested in and it is shown by the significance of the *F*-ratio in Output . We can see that the significance of the F-ratio is .010, which is less than the criterion of .05; therefore, we can say that there is a significant interaction between the relationship status of women and whether they had a photo of themselves alone or with a partner. The estimated marginal means in Output and the interaction graph (Figure) help us to interpret this effect. The significant interaction seems to indicate that when displaying a photo of themselves alone rather than with a partner, the number of friend requests increases in both women in a relationship and single women. However, for single women this increase is greater than for women who are in a relationship.

3. Relationship Status * Profile_Picture

Measure: MEASURE_1

Relationship Status	Profile_Picture	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Single	1	3.957	.245	3.460	4.453
	2	7.913	.354	7.196	8.630
In a Relationship	1	3.294	.285	2.717	3.871
	2	5.647	.412	4.813	6.481

Output 31

Writing the results

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

- ✓ The main effect of relationship status was significant, $F(1, 38) = 16.29$, $p < .001$, indicating that single women received more friend requests than women who were in a relationship, regardless of their type of profile picture.
- ✓ The main effect of profile picture was significant, $F(1, 38) = 114.77$, $p < .001$, indicating that across all women, the number of friend requests was greater when displaying a photo alone rather than with a partner.

- ✓ The relationship status × profile picture interaction was significant, $F(1, 38) = 7.41$, $p = .010$, indicating that although number of friend requests increased in all women when they displayed a photo of themselves alone compared to when they displayed a photo of themselves with a partner, this increase was significantly greater for single women than for women who were in a relationship.