Repeated Measures ANOVA

Issues with Repeated Measures Designs

Repeated measures is a term used when the same entities take part in all conditions of an experiment. So, for example, you might want to test the effects of alcohol on enjoyment of a party. In this type of experiment it is important to control for individual differences in tolerance to alcohol: some people can drink a lot of alcohol without really feeling the consequences, whereas others, like me, only have to sniff a pint of lager and they fall to the floor and pretend to be a fish. To control for these individual differences we can test the same people in all conditions of the experiment: so we would test each subject after they had consumed one pint, two pints, three pints and four pints of lager. After each drink the participant might be given a questionnaire assessing their enjoyment of the party. Therefore, every participant provides a score representing their enjoyment before the study (no alcohol consumed), after one pint, after two pints, and so on. This design is said to use repeated measures.

What is Sphericity?

We have seen that parametric tests based on the normal distribution assume that data points are independent. This is not the case in a repeated measures design because data for different conditions have come from the same entities. This means that data from different experimental conditions will be related; because of this we have to make an additional assumption to those of the independent ANOVAs you have so far studied. Put simply (and not entirely accurately), we assume that the relationship between pairs of experimental conditions is similar (i.e. the level of dependence between pairs of groups is roughly equal). This assumption is known as the assumption of **sphericity**. The more accurate but complex explanation is as follows. Table 1 shows data from an experiment with three conditions. Imagine we calculated the differences between pairs of scores in all combinations of the treatment levels. Having done this, we calculated the variance of these differences. Sphericity is met when these variances are roughly equal. In these data there is some deviation from sphericity because the variance of the differences between conditions A and B (15.7) is greater than the variance of the differences between A and C (10.3) and between B and C (10.7). However, these data have *local circularity* (or local sphericity) because two of the variances of differences are very similar.

Table 1: Hypothetical data to illustrate the calculation of the variance of the differences between conditions

Condition A	Condition B	Condition C	A−B	A−C	B−C
10	12	8	−2	2	4
15	15	12	0	3	3
25	30	20	−5	5	10
35	30	28	5	7	2
30	27	20	3	10	7
		Variance:	15.7	10.3	10.7

What is the Effect of Violating the Assumption of Sphericity?

The effect of violating sphericity is a loss of power (i.e. an increased probability of a Type II error) and a test statistic (*F*-ratio) that simply cannot be compared to tabulated values of the *F*-distribution (for more details see Field, 2009; 2013).

Assessing the Severity of Departures from Sphericity

Departures from sphericity can be measured in three ways:

1. Greenhouse and Geisser (1959)
2. Huynh and Feldt (1976)
3. The Lower Bound estimate (the lowest possible theoretical value for the data)

The Greenhouse-Geisser and Huynh-Feldt estimates can both range from the lower bound (the most severe departure from sphericity possible given the data) and 1 (no departure from sphjercitiy at all). For more detail on these estimates see Field (2013) or Girden (1992).

SPSS also produces a test known as Mauchly's test, which tests the hypothesis that the variances of the differences between conditions are equal.

TIP

- → If Mauchly's test statistic is significant (i.e. has a probability value less than .05) we conclude that there are significant differences between the variance of differences: the condition of sphericity has not been met.
- → If, Mauchly's test statistic is nonsignificant (i.e. $p > .05$) then it is reasonable to conclude that the variances of differences are not significantly different (i.e. they are roughly equal).
- → If Mauchly's test is significant then we cannot trust the *F*-ratios produced by SPSS.
- → Remember that, as with any significance test, the power of Mauchley's test depends on the sample size. Therefore, it must be interpreted within the context of the sample size because:
    - In small samples large deviations from sphericity might be deemed non-significant.
    - In large samples, small deviations from sphericity might be deemed significant.

Correcting for Violations of Sphericity

Fortunately, if data violate the sphericity assumption we simply adjust the defrees of freedom for the effect by multiplying it by one of the aforementioned sphericity estimates. This will make the degrees of freedom smaller; by reducing the degrees of freedom we make the *F*-ratio more conservative (i.e. it has to be bigger to be deemed significant). SPSS applies these adjustments automatically.

TIP

Which correction should I use?

- → Look at the Greenhouse-Geisser estimate of sphericity (ε) in the SPSS handout.
- → When ε > .75 then use the Huynh-Feldt correction.
- → When ε < .75 then use the Greenhouse-Geisser correction.

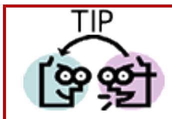
One-Way Repeated Measures ANOVA using SPSS

"I'm a celebrity, get me out of here" is a TV show in which celebrities (well, I mean, they're not really are they … I'm struggling to know who anyone is in the series these days) in a pitiful attempt to salvage their careers (or just have careers in the first place) go and live in the jungle and subject themselves to ritual humiliation and/or creepy crawlies in places where creepy crawlies shouldn't go. It's cruel, voyeuristic, gratuitous, car crash TV, and I *love* it. A particular favourite bit is the Bushtucker trials in which the celebrities willingly eat things like stick insects, Witchetty grubs, fish eyes, and kangaroo testicles/penises, nom nom noms….

I've often wondered (perhaps a little too much) which of the bushtucker foods is most revolting. So I got 8 celebrities, and made them eat four different animals (the aforementioned stick insect, kangaroo testicle, fish eye and Witchetty grub) in counterbalanced order. On each occasion I measured the time it took the celebrity to retch, in seconds. The data are in Table 2.

Entering the Data

The independent variable was the animal that was being eaten (stick, insect, kangaroo testicle, fish eye and witchetty grub) and the dependent variable was the time it took to retch, in seconds.



→ Levels of repeated measures variables go in different columns of the SPSS data editor.

Therefore, separate columns should represent each level of a repeated measures variable. As such, there is no need for a coding variable (as with between-group designs). The data can, therefore, be entered as they are in Table 2.

- Save these data in a file called **bushtucker.sav**

Table 2: Data for the Bushtucker example

Celebrity	Stick Insect	Kangaroo Testicle	Fish Eye	Witchetty Grub
1	8	7	1	6
2	9	5	2	5
3	6	2	3	8
4	5	3	1	9
5	8	4	5	8
6	7	5	6	7
7	10	2	7	2
8	12	6	8	1

Draw an error bar chart of these data. The resulting graph is in Figure 1.

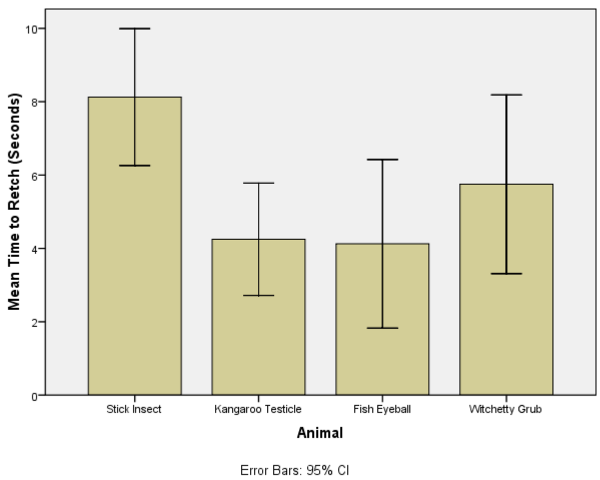


Figure 1: Graph of the mean time to retch after eating each of the four animals (error bars show the 95% confidence interval)

To conduct an ANOVA using a repeated measures design, activate the *define factors* dialog box by selecting Analyze General Linear Model ▸ Repeated Measures... . In the *Define Factors* dialog box (Figure 2), you are asked to supply a name for the within-subject (repeated-measures) variable. In this case the repeated measures variable was the type of animal eaten in the bushtucker trial, so replace the word *factor1* with the word *Animal*. The name you give to the repeated measures variable cannot have spaces in it. When you have given the repeated measures factor a name, you have to tell the computer how many levels there were to that variable (i.e. how many experimental conditions there were). In this case, there were 4 different animals eaten by each person, so we have to enter the number 4 into the box labelled *Number of Levels*. Click on Add to add this variable to the list of repeated measures variables. This variable will now appear in the white box at the bottom of the dialog box and appears as *Animal(4)*. If your design has several repeated measures variables then you can add more factors to the list (see Two Way ANOVA example below). When you have entered all of the repeated measures factors that were measured click on Define to go to the *Main Dialog Box*.

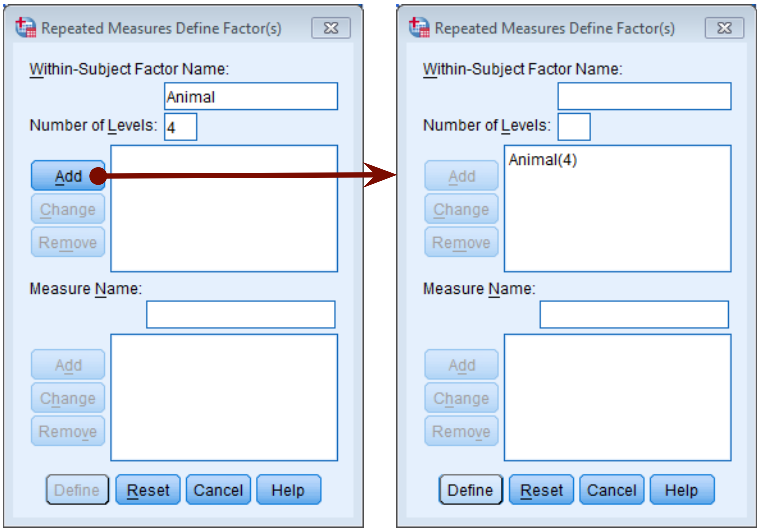


Figure 2: *Define Factors* dialog box for repeated measures ANOVA

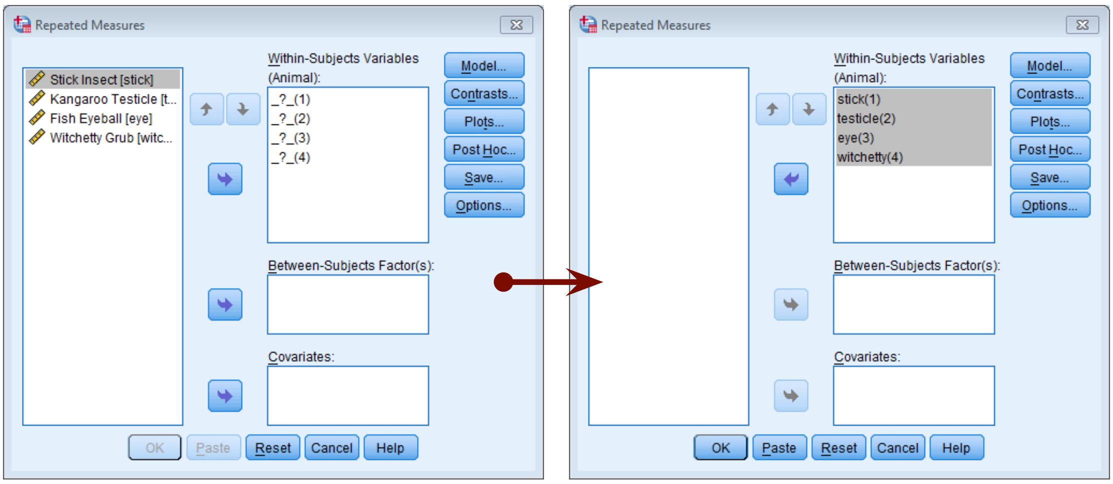


Figure 3: Main dialog box for repeated measures ANOVA

The main dialog box (Figure 3) has a space labelled *within subjects variable list* that contains a list of 4 question marks proceeded by a number. These question marks are for the variables representing the 4 levels of the independent variable. The variables corresponding to these levels should be selected and placed in the appropriate space. We have only 4 variables in the data editor, so it is possible to select all four variables at once (by clicking on the variable at the top, holding the mouse button down and dragging down over the other variables). The selected variables can then be transferred by dragging them or clicking on .

When all four variables have been transferred, you can select various options for the analysis. There are several options that can be accessed with the buttons at the bottom of the main dialog box. These options are similar to the ones we have already encountered.

Post Hoc Tests

There is no proper facility for producing post hoc tests for repeated measures variables in SPSS (you will find that if you access the *post hoc* test dialog box it will not list any repeated-measured factors). However, you can get a basic set of post hoc tests clicking Options... in the main dialog box. To specify *post hoc* tests, select the repeated measures variable (in this case **Animal**) from the box labelled *Estimated Marginal Means: Factor(s) and Factor Interactions* and transfer it to the box labelled *Display Means for* by clicking on ➡ (Figure 4). Once a variable has been transferred, the box labelled *Compare main effects* (☑ Compare main effects) becomes active and you should select this option. If this option is selected, the box labelled *Confidence interval adjustment* becomes active and you can click on LSD(none) to see a choice of three adjustment levels. The default is to have no adjustment and simply perform a Tukey LSD *post hoc* test (this is not recommended). The second option is a Bonferroni correction (which we've encountered before), and the final option is a Sidak correction, which should be selected if you are concerned about the loss of power associated with Bonferroni corrected values. When you have selected the options of interest, click on Continue to return to the main dialog box, and then click on OK to run the analysis.

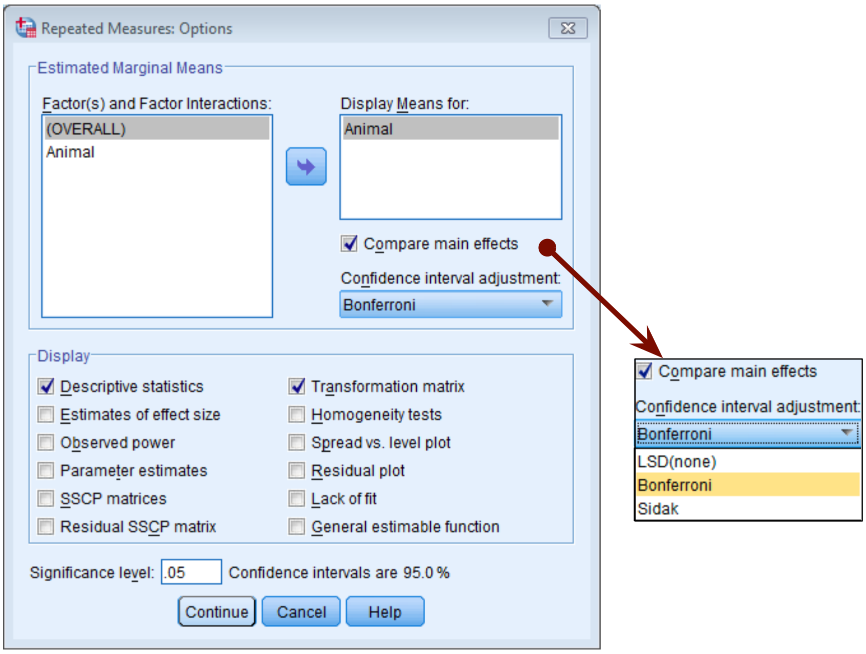


Figure 4: Options dialog box

Output for Repeated Measures ANOVA

Descriptive statistics and other Diagnostics

Within-Subjects Factors

Measure: MEASURE_1

Animal	Dependent Variable
1	stick
2	testicle
3	eye
4	witchetty

Descriptive Statistics

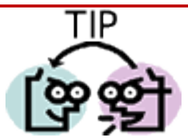
	Mean	Std. Deviation	N
Stick Insect	8.13	2.232	8
Kangaroo Testicle	4.25	1.832	8
Fish Eyeball	4.13	2.748	8
Witchetty Grub	5.75	2.915	8

Output 1

Output 1 shows the initial diagnostics statistics. First, we are told the variables that represent each level of the independent variable. This box is useful mainly to check that the variables were entered in the correct order. The following table provides basic descriptive statistics for the four levels of the independent variable. From this table we can see that, on average, the quickest retching was after the kangaroo testicle and fish eyeball (implying they are more disgusting).

Assessing Sphericity

Earlier you were told that SPSS produces a test that looks at whether the data have violated the assumption of sphericity. The next part of the output contains information about this test.



→ Mauchly's test should be nonsignificant if we are to assume that the condition of sphericity has been met.

→ Sometimes when you look at the significance, all you see is a dot. There is no significance value. The reason that this happens is that you need at least three conditions for sphericity to be an issue. Therefore, if you have a repeated-measures variable that has only two levels then sphericity is met, the estimates computed by SPSS are 1 (perfect sphericity) and the resulting significance test cannot be computed (hence why the table has a value of 0 for the chi-square test and degrees of freedom and a blank space for the significance). It would be a lot easier if SPSS just didn't produce the table, but then I guess we'd all be confused about why the table hadn't appeared; maybe it should just print in big letters 'Hooray! Hooray! Sphericity has gone away!' We can dream.

Output 2 shows Mauchly's test for these data, and the important column is the one containing the significance vale. The significance value is .047, which is less than .05, so we must accept the hypothesis that the variances of the differences between levels were significantly different. In other words the assumption of sphericity has been violated. We could report Mauchly's test for these data as:

→ Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 11.41$, $p = .047$.

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
Animal	.136	11.406	5	.047	.533	.666	.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
Within Subjects Design: Animal

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 2

The Main ANOVA

Output 3 shows the results of the ANOVA for the within-subjects variable. The table you see will look slightly different (it will look like Output 4 in fact), but for the time being I've simplified it a bit. Bear with me for now. This table can be read much the same as for One-way independent ANOVA (see your handout). The significance of *F* is .026, which is significant because it is less than the criterion value of .05. We can, therefore, conclude that there was a significant difference in the time taken to retch after eating different animals. However, this main test does not tell us which animals resulted in the quickest retching times.

Although this result seems very plausible, we saw earlier that the assumption of sphericity had been violated. I also mentioned that a violation of the sphericity assumption makes the *F*-test inaccurate. So, what do we do? Well, I mentioned earlier on that we can correct the degrees of freedom in such a way that it is accurate when sphericity is violated. This is what SPSS does. Output 4 (which is the output you will see in your own SPSS analysis) shows the main ANOVA. As you can see in this output, the value of *F* does not change, only the degrees of freedom. But the effect of changing the degrees of freedom is that the *significance* of the value of *F* changes: the effect of the type of animal is less significant after correcting for sphericity.

Tests of Within-Subjects Effects

Measure: MEASURE_1
Sphericity Assumed

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Animal	83.125	3	27.708	3.794	.026
Error(Animal)	153.375	21	7.304		

Output 3

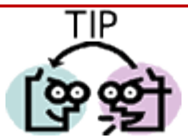
Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Animal	Sphericity Assumed	83.125	3	27.708	3.794	.026
	Greenhouse-Geisser	83.125	1.599	52.001	3.794	.063
	Huynh-Feldt	83.125	1.997	41.619	3.794	.048
	Lower-bound	83.125	1.000	83.125	3.794	.092
Error(Animal)	Sphericity Assumed	153.375	21	7.304		
	Greenhouse-Geisser	153.375	11.190	13.707		
	Huynh-Feldt	153.375	13.981	10.970		
	Lower-bound	153.375	7.000	21.911		

Output 4

The next issue is which of the three corrections to use. Earlier I gave you some tips and they were that when $\varepsilon > .75$ then use the Huynh-Feldt correction, and when $\varepsilon < .75$, or nothing is known about sphericity at all, then use the Greenhouse-Geisser correction; ε is the estimate of sphericity from Output 2 and these values are .533 and .666 (the correction of the beast ….); because these values are less than .75 we should use the Greenhouse-Geisser corrected values. Using this correction, *F* is not significant because its *p* value is .063, which is more than the normal criterion of .05.



- In this example the results are quite weird because uncorrected they are significant, and applying the Huynh-Feldt correction they are also significant. However, with the Greenhouse-Geisser correction applied they are not.
- This highlights how arbitrary the whole .05 criterion for significance is. Clearly, these *F*s represent the same sized effect, but using one criterion they are 'significant' and using another they are not.

Post Hoc Tests

Given the main effect was not significant, we should not follow this effect up with post hoc tests, but instead conclude that the type of animal did not have a significant effect on how quickly contestants retched (perhaps we should have used beans on toast as a baseline against which to compare …).

However, just to illustrate how you would interpret the SPSS output I have reproduced it in Output 5: the difference between group means is displayed, the standard error, the significance value and a confidence interval for the difference between means. By looking at the significance values we can see that the only significant differences between group means is between the stick insect and the kangaroo testicle, and the stick insect and the fish eye. No other differences are significant.

Pairwise Comparisons

Measure: MEASURE_1

(I) Animal	(J) Animal	Mean Difference (I–J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	3.875*	.811	.012	.925	6.825
	3	4.000*	.732	.006	1.339	6.661
	4	2.375	1.792	1.000	-4.141	8.891
2	1	-3.875*	.811	.012	-6.825	-.925
	3	.125	1.202	1.000	-4.244	4.494
	4	-1.500	1.336	1.000	-6.359	3.359
3	1	-4.000*	.732	.006	-6.661	-1.339
	2	-.125	1.202	1.000	-4.494	4.244
	4	-1.625	1.822	1.000	-8.249	4.999
4	1	-2.375	1.792	1.000	-8.891	4.141
	2	1.500	1.336	1.000	-3.359	6.359
	3	1.625	1.822	1.000	-4.999	8.249

Based on estimated marginal means

*. The mean difference is significant at the

b. Adjustment for multiple comparisons: Bonferroni.

Output 5

Reporting One-Way Repeated Measures ANOVA

We can report repeated measures ANOVA in the same way as an independent ANOVA (see your handout). The only additional thing we should concern ourselves with is reporting the corrected degrees of freedom if sphericity was violated. Personally, I'm also keen on reporting the results of sphericity tests as well. Therefore, we could report the main finding as:

→ Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 11.41$, $p = .047$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .53$). The results show that there was no significant effect of which animal was eaten on the time taken to retch, $F(1.60, 11.19) = 3.79$, $p = .06$. These results suggested that no animal was significantly more disgusting to eat than the others.

Two-Way Repeated Measures ANOVA Using SPSS

As we have seen before, the name of any ANOVA can be broken down to tell us the type of design that was used. The 'two-way' part of the name simply means that two independent variables have been manipulated in the experiment. The 'repeated measures' part of the name tells us that the same participants have been used in all conditions. Therefore, this analysis is appropriate when you have two repeated-measures independent variables: each participant does all of the conditions in the experiment, and provides a score for each permutation of the two variables.

A Speed-Dating Example

It seems that lots of magazines go on all the time about how men and women want different things from relationships (or perhaps it's just my wife's copies of Marie Clare's, which obviously I don't read, honestly). The big question to which we all want to know the answer is are looks or personality more important. Imagine you wanted to put this to the test. You devised a cunning plan whereby you'd set up a speed-dating night. Little did the people who came along know that you'd got some of your friends to act as the dates. Specifically you found 9 men to act as the date. In each of these groups three people were extremely attractive people but differed in their personality: one had tonnes of charisma, one had some charisma, and the third person was as dull as this handout. Another three people were of average attractiveness, and again differed in their personality: one was highly charismatic, one had some charisma and the third was a dullard. The final three were, not wishing to be unkind in any way, butt-ugly and again one was charismatic, one had some charisma and the final poor soul was mind-numbingly tedious. The participants were heterosexual women who came to the speed dating night, and over the course of the evening they speed-dated all 9 men. After their 5 minute date, they rated how much they'd like to have a proper date with the person as a percentage (100% = 'I'd pay large sums of money for your phone number', 0% = 'I'd pay a large sum of money for a plane ticket to get me as far away as possible from you'). As such, each woman rated 9 different people who varied in their attractiveness and personality. So, there are two repeated measures variables: **looks** (with three levels because the person could be attractive, average or ugly) and **personality** (again with three levels because the person could have lots of charisma, have some charisma, or be a dullard).

Data Entry

To enter these data into SPSS we use the same procedure as the one-way repeated measures ANOVA that we came across in the previous example.

TIP

→ Levels of repeated measures variables go in different columns of the SPSS data editor.

If a person participates in all experimental conditions (in this case she dates all of the men who differ in attractiveness and all of the men who differ in their charisma) then each experimental condition must be represented by a column in the data editor. In this experiment there are nine experimental conditions and so the data need to be entered in nine columns. Therefore, create the following nine variables in the data editor with the names as given. For each one, you should also enter a full variable name for clarity in the output.

att_high	Attractive	+	High Charisma
av_high	Average Looks	+	High Charisma
ug_high	Ugly	+	High Charisma
att_some	Attractive	+	Some Charisma
av_some	Average Looks	+	Some Charisma
ug_some	Ugly	+	Some Charisma
att_none	Attractive	+	Dullard
av_none	Average Looks	+	Dullard
ug_none	Ugly	+	Dullard

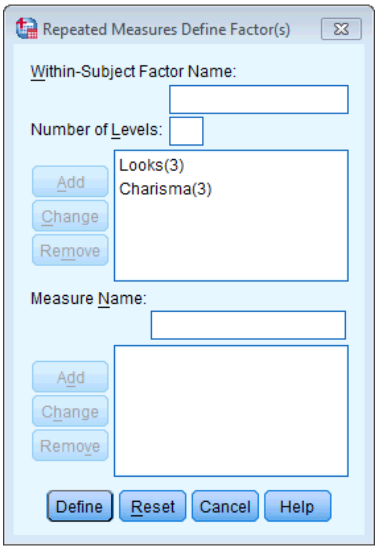


Figure 5: *Define factors* dialog box for factorial repeated measures ANOVA

The data are in the file **FemaleLooksOrPersonality.sav** from the course website. First we have to define our repeated measures variables, so access the *define factors* dialog box select Analyze ▸ General Linear Model ▸ Repeated Measures... . As with one-way repeated measures ANOVA (see the previous example) we need to give names to our repeated measures variables and specify how many levels they have. In this case there are two within-subject factors: **looks** (attractive, average or ugly) and **charisma** (high charisma, some charisma and dullard). In the *define factors* dialog box replace the word *factor1* with the word *looks*. When you have given this repeated measures factor a name, tell the computer that this variable has 3 levels by typing the number 3 into the box labelled *Number of Levels* (Figure 5). Click on Add to add this variable to the list of repeated measures variables. This variable will now appear in the white box at the bottom of the dialog box and appears as *looks(3)*.

Now repeat this process for the second independent variable. Enter the word *charisma* into the space labelled *Within-Subject Factor Name* and then, because there were three levels of this variable, enter the number 3 into the space labelled *Number of Levels*. Click on Add to include this variable in the list of factors; it will appear as *charisma(3)*. The finished dialog box is shown in Figure 5. When you have entered both of the within-subject factors click on Define to go to the main dialog box.

The main dialog box is shown in Figure 6. At the top of the *Within-Subjects Variables* box, SPSS states that there are two factors: **looks** and **charisma**. In the box below there is a series of question marks followed by bracketed numbers. The numbers in brackets represent the levels of the factors (independent variables). In this example, there are two independent variables and so there are two numbers in the brackets. The first number refers to levels of the first factor listed above the box (in this case **looks**). The second number in the bracket refers to levels of the second factor listed above the box (in this case **charisma**). We have to replace the question marks with variables from the list on the left-hand side of the dialog box. With between-group designs, in which coding variables are used, the levels of a particular factor are specified by the codes assigned to them in the data editor. However, in repeated measures designs, no such coding scheme is used and so we determine which condition to assign to a level at this stage. The variables can be entered as follows:

att_high	→	_?_(1,1)
att_some	→	_?_(1,2)
att_none	→	_?_(1,3)
av_high	→	_?_(2,1)
av_some	→	_?_(2,2)
av_none	→	_?_(2,3)
ug_high	→	_?_(3,1)
ug_some	→	_?_(3,2)
ug_none	→	_?_(3,3)

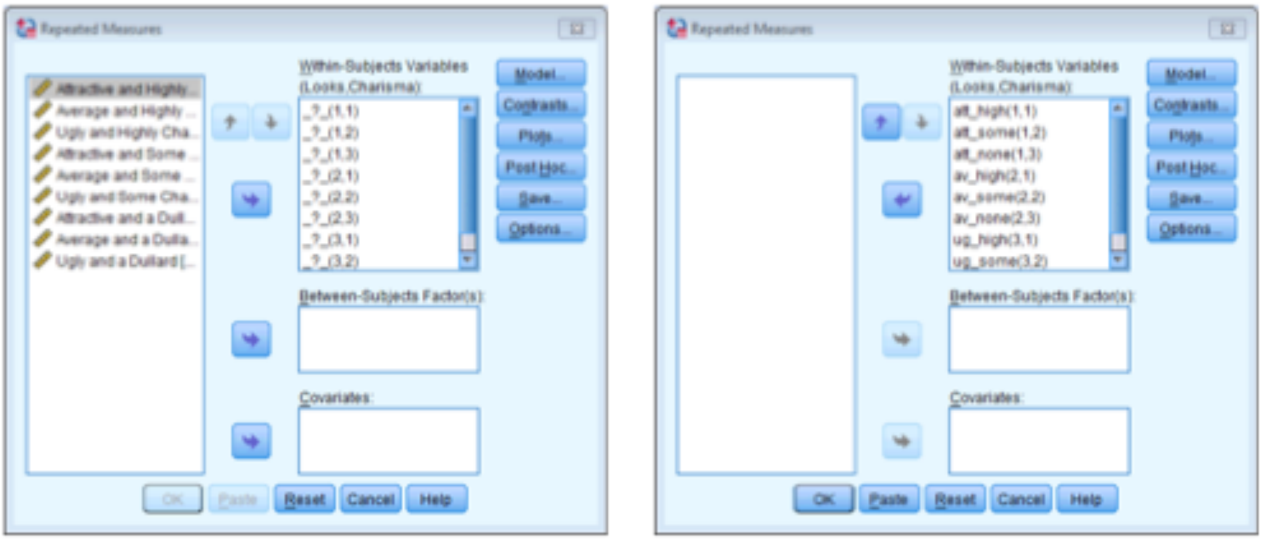


Figure 6: Main repeated measures dialog box

The completed dialog box should look exactly like Figure 6. I've already discussed the options for the buttons at the bottom of this dialog box, so I'll talk only about the ones of particular interest for this example.

Other Options

The addition of an extra variable makes it necessary to choose a different graph to the one in the previous handout. Click on Plots... to access the dialog box in Figure 7. Place **looks** in the slot labelled *Horizontal Axis:* and **charisma** in slot labelled *Separate Line*. When both 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 looks × charisma interaction, we should get the interaction graph for looks and charisma. You could also think about plotting graphs for the two main effects (e.g. looks and charisma). As far as other options are concerned, you should select the same ones that were chosen for the previous example. It is worth selecting estimated marginal means for all effects (because these values will help you to understand any significant effects).

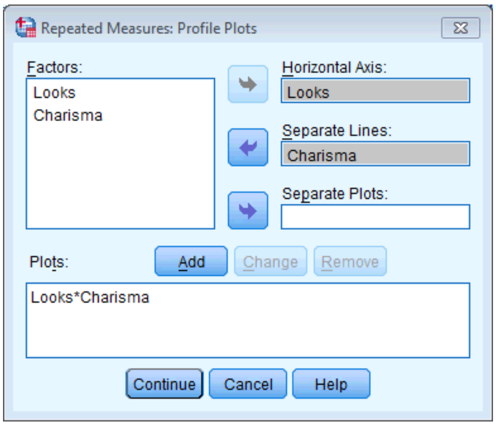


Figure 7: *Plots* dialog box for a two-way repeated measures ANOVA

Descriptives and Main Analysis

Output 6 shows the initial output from this ANOVA. The first table merely lists the variables that have been included from the data editor and the level of each independent variable that they represent. This table is more important than it might seem, because it enables you to verify that the variables in the SPSS data editor represent the correct levels of the independent variables. The second table is a table of descriptives and provides the mean and standard deviation for each of the nine conditions. The names in this table are the names I gave the variables in the data editor (therefore, if you didn't give these variables full names, this table will look slightly different). The values in this table will help us later to interpret the main effects of the analysis.

Within-Subjects Factors

Measure: MEASURE_1

Looks	Charisma	Dependent Variable
1	1	att_high
	2	att_some
	3	att_none
2	1	av_high
	2	av_some
	3	av_none
3	1	ug_high
	2	ug_some
	3	ug_none

Descriptive Statistics

	Mean	Std. Deviation	N
Attractive and Highly Charismatic	89.60	6.637	10
Attractive and Some Charisma	87.10	6.806	10
Attractive and a Dullard	51.80	3.458	10
Average and Highly Charismatic	88.40	8.329	10
Average and Some Charisma	68.90	5.953	10
Average and a Dullard	47.00	3.742	10
Ugly and Highly Charismatic	86.70	5.438	10
Ugly and Some Charisma	51.20	5.453	10
Ugly and a Dullard	46.10	3.071	10

Output 6

Output 7 shows the results of Mauchly's sphericity test for each of the three effects in the model (two main effects and one interaction). The significance values of these tests indicate that for the main effects of **Looks** and **Charisma** the assumption of sphericity is met (because $p > .05$) so we need not correct the F-ratios for these effects. However, the Looks × Charisma interaction has violated this assumption and so the F-value for this effect should be corrected.

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
Looks	.904	.810	2	.667	.912	1.000	.500
Charisma	.851	1.292	2	.524	.870	1.000	.500
Looks * Charisma	.046	22.761	9	.008	.579	.791	.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 Tests of Within-Subjects Effects table.

b.

Design: Intercept
Within Subjects Design: Looks+Charisma+Looks*Charisma

Output 7

Output 8 shows the results of the ANOVA (with corrected *F* values). The output is split into sections that refer to each of the effects in the model and the error terms associated with these effects. The interesting part is the significance values of the *F*-ratios. If these values are less than .05 then we can say that an effect is significant. Looking at the significance values in the table it is clear that there is a significant main effect of how attractive the date was (Looks), a significant main effect of how charismatic the date was (Charisma), and a significant interaction between these two variables. I will examine each of these effects in turn.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Looks	Sphericity Assumed	3308.867	2	1654.433	66.437	.000
	Greenhouse-Geisser	3308.867	1.824	1813.723	66.437	.000
	Huynh-Feldt	3308.867	2.000	1654.433	66.437	.000
	Lower-bound	3308.867	1.000	3308.867	66.437	.000
Error(Looks)	Sphericity Assumed	448.244	18	24.902		
	Greenhouse-Geisser	448.244	16.419	27.300		
	Huynh-Feldt	448.244	18.000	24.902		
	Lower-bound	448.244	9.000	49.805		
Charisma	Sphericity Assumed	23932.867	2	11966.433	274.888	.000
	Greenhouse-Geisser	23932.867	1.740	13751.549	274.888	.000
	Huynh-Feldt	23932.867	2.000	11966.433	274.888	.000
	Lower-bound	23932.867	1.000	23932.867	274.888	.000
Error(Charisma)	Sphericity Assumed	783.578	18	43.532		
	Greenhouse-Geisser	783.578	15.663	50.026		
	Huynh-Feldt	783.578	18.000	43.532		
	Lower-bound	783.578	9.000	87.064		
Looks * Charisma	Sphericity Assumed	3365.867	4	841.467	34.912	.000
	Greenhouse-Geisser	3365.867	2.315	1453.670	34.912	.000
	Huynh-Feldt	3365.867	3.165	1063.585	34.912	.000
	Lower-bound	3365.867	1.000	3365.867	34.912	.000
Error(Looks*Charisma)	Sphericity Assumed	867.689	36	24.102		
	Greenhouse-Geisser	867.689	20.839	41.638		
	Huynh-Feldt	867.689	28.482	30.465		
	Lower-bound	867.689	9.000	96.410		

Output 8

The Main Effect of Looks

We came across the main effect of looks in Output 8.

TIP

→ We can report that 'there was a significant main effect of looks, $F(2, 18) = 66.44$, $p < .001$.'

→ This effect tells us that if we ignore all other variables, ratings were different for attractive, average and unattractive dates.

If you requested that SPSS display means for the looks effect (I'll assume you did from now on) you will find the table in a section headed *Estimated Marginal Means*. Output 9 is a table of means for the main effect of looks with the associated standard errors. The levels of looks are labelled simply 1, 2 and 3, and it's down to you to remember how you entered the variables (or you can look at the summary table that SPSS produces at the beginning of the output—see Output 6). If you followed what I did then level 1 is attractive, level 2 is average and level 3 is ugly. To make things easier, this information is plotted in Figure 8: as attractiveness falls, the mean rating falls too. This main effect seems to reflect that the women were more likely to express a greater interest in going out with attractive men than average or ugly men. However, we really need to look at some contrasts to find out exactly what's going on (see Field, 2013 if you're interested).

Estimates

Measure: MEASURE_1

Looks	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	76.167	1.013	73.876	78.457
2	68.100	1.218	65.344	70.856
3	61.333	1.018	59.030	63.637

Output 9

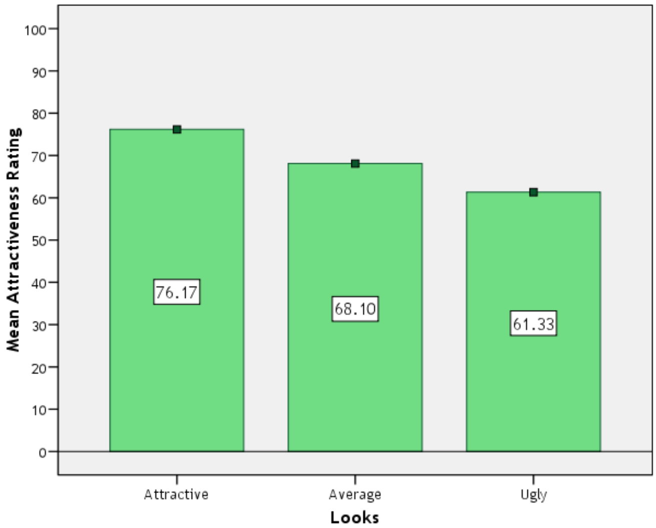


Figure 8: The main effect of looks

The Effect of Charisma

The main effect of charisma was in Output 8.

TIP

→ We can report that there was a significant main effect of charisma, $F(2, 18) = 274.89$, $p < .001$.

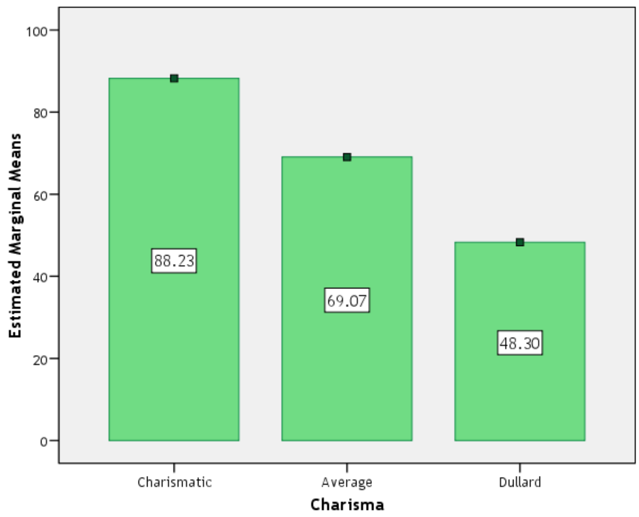
→ This effect tells us that if we ignore all other variables, ratings were different for highly charismatic, a bit charismatic and dullard people.

The table labelled *CHARISMA* in the section headed *Estimated Marginal Means* tells us what this effect means (Output 10.). Again, the levels of charisma are labelled simply 1, 2 and 3. If you followed what I did then level 1 is high charisma, level 2 is some charisma and level 3 is no charisma. This information is plotted in Figure 9: As charisma declines, the mean rating falls too. So this main effect seems to reflect that the women were more likely to express a greater interest in going out with charismatic men than average men or dullards. Again, we would have to look at contrasts or post hoc tests to break this effect down further.

Estimates

Measure: MEASURE_1

Charisma	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	88.233	1.598	84.619	91.848
2	69.067	1.293	66.142	71.991
3	48.300	.751	46.601	49.999



Output 10

Figure 9: The main effect of charisma

The Interaction between Looks and Charisma

Output 8 indicated that the attractiveness of the date interacted in some way with how charismatic the date was.

TIP

→ We can report that 'there was a significant interaction between the attractiveness of the date and the charisma of the date, $F(2.32, 20.84) = 34.91$, $p < .001$'.

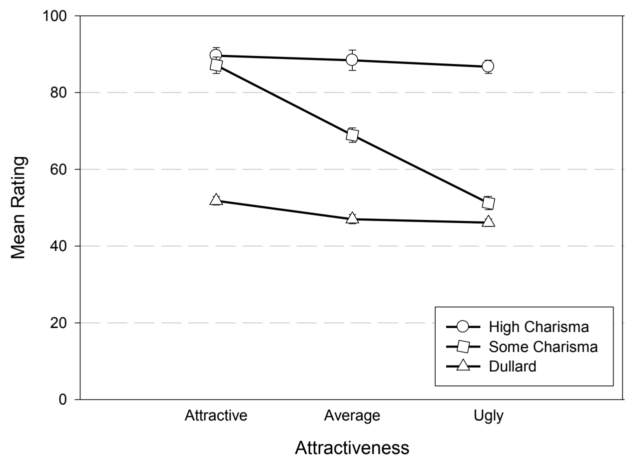
→ This effect tells us that the profile of ratings across dates of different levels of charisma was different for attractive, average and ugly dates.

The estimated marginal means (or a plot of looks × charisma using the dialog box in Figure 4) tell us the meaning of this interaction (see Figure 10 and Output 11).

3. Looks * Charisma

Measure: MEASURE_1

Looks	Charisma	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	89.600	2.099	84.852	94.348
	2	87.100	2.152	82.231	91.969
	3	51.800	1.093	49.327	54.273
2	1	88.400	2.634	82.442	94.358
	2	68.900	1.882	64.642	73.158
	3	47.000	1.183	44.323	49.677
3	1	86.700	1.719	82.810	90.590
	2	51.200	1.724	47.299	55.101
	3	46.100	.971	43.903	48.297



Output 11

Figure 10: The looks × charisma interaction

The graph shows the average ratings of dates of different levels of attractiveness when the date also had high levels of charisma (circles), some charisma (squares) and no charisma (triangles). Look first at the highlight charismatic dates. Essentially, the ratings for these dates do not change over levels of attractiveness. In other words, women's ratings of dates for highly charismatic men was unaffected by how good looking they were – ratings were high regardless of looks. Now look at the men who were dullards. Women rated these dates as low regardless of how attractive the man was. In other words, ratings for dullards were unaffected by looks: even a good looking man gets low ratings if he is a dullard. So, basically, the attractiveness of men makes no difference for high charisma (all ratings are high) and low charisma (all ratings are low). Finally, let's look at the men who were averagely charismatic. For these men attractiveness had a big impact – attractive men got high ratings, and unattractive men got low ratings. If a man has average charisma then good looks would pull his rating up, and being ugly would pull his ratings down. A succinct way to describe what is going on would be to say that the Looks variable only has an effect for averagely charismatic men.

Guided Example

A clinical psychologist was interested in the effects of antidepressants and cognitive behaviour therapy on suicidal thoughts. Four depressives took part in four conditions: placebo tablet with no therapy for one month, placebo tablet with cognitive behaviour therapy (CBT) for one month, antidepressant with no therapy for one month, and antidepressant with cognitive behaviour therapy (CBT) for one month. The order of conditions was fully counterbalanced across the 4 participants. Participants recorded the number of suicidal thoughts they had during the final week of each month. The data are below:

Table 3: Data for the effect of antidepressants and CBT on suicidal thoughts

Drug:	Placebo		Antidepressant	
Therapy:	None	CBT	None	CBT
Andy	70	60	81	52
Laura	66	52	70	40
Fidelma	56	41	60	31
Becky	68	59	77	49
Mean	**65**	**53**	**72**	**43**

The SPSS output you get for these data should look like the following:

Within-Subjects Factors

Measure: MEASURE_1

DRUG	THERAPY	Dependent Variable
1	1	PLNONE
	2	PLCBT
2	1	ANTNONE
	2	ANTCBT

Descriptive Statistics

	Mean	Std. Deviation	N
Placebo - No Therapy	65.0000	6.2183	4
Placebo - CBT	53.0000	8.7560	4
Antidepressant - No Therapy	72.0000	9.2014	4
Antidepressant - CBT	43.0000	9.4868	4

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
DRUG	1.000	.000	0	.	1.000	1.000	1.000
THERAPY	1.000	.000	0	.	1.000	1.000	1.000
DRUG * THERAPY	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
Within Subjects Design: DRUG+THERAPY+DRUG*THERAPY

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
DRUG	Sphericity Assumed	9.000	1	9.000	1.459	.314
	Greenhouse-Geisser	9.000	1.000	9.000	1.459	.314
	Huynh-Feldt	9.000	1.000	9.000	1.459	.314
	Lower-bound	9.000	1.000	9.000	1.459	.314
Error(DRUG)	Sphericity Assumed	18.500	3	6.167		
	Greenhouse-Geisser	18.500	3.000	6.167		
	Huynh-Feldt	18.500	3.000	6.167		
	Lower-bound	18.500	3.000	6.167		
THERAPY	Sphericity Assumed	1681.000	1	1681.000	530.842	.000
	Greenhouse-Geisser	1681.000	1.000	1681.000	530.842	.000
	Huynh-Feldt	1681.000	1.000	1681.000	530.842	.000
	Lower-bound	1681.000	1.000	1681.000	530.842	.000
Error(THERAPY)	Sphericity Assumed	9.500	3	3.167		
	Greenhouse-Geisser	9.500	3.000	3.167		
	Huynh-Feldt	9.500	3.000	3.167		
	Lower-bound	9.500	3.000	3.167		
DRUG * THERAPY	Sphericity Assumed	289.000	1	289.000	192.667	.001
	Greenhouse-Geisser	289.000	1.000	289.000	192.667	.001
	Huynh-Feldt	289.000	1.000	289.000	192.667	.001
	Lower-bound	289.000	1.000	289.000	192.667	.001
Error(DRUG*THERAPY)	Sphericity Assumed	4.500	3	1.500		
	Greenhouse-Geisser	4.500	3.000	1.500		
	Huynh-Feldt	4.500	3.000	1.500		
	Lower-bound	4.500	3.000	1.500		

1. DRUG

Measure: MEASURE_1

DRUG	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	59.000	3.725	47.146	70.854
2	57.500	4.668	42.644	72.356

2. THERAPY

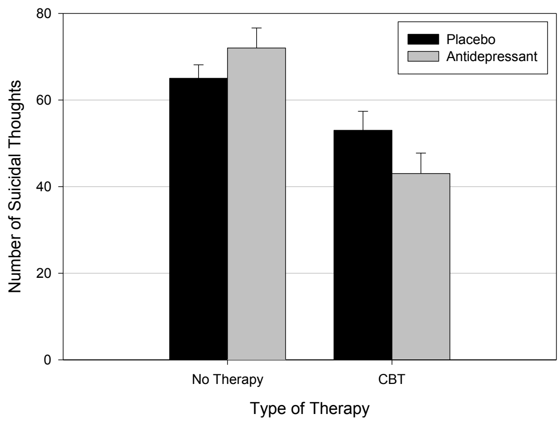
Measure: MEASURE_1

THERAPY	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	68.500	3.824	56.329	80.671
2	48.000	4.546	33.532	62.468

3. DRUG * THERAPY

Measure: MEASURE_1

DRUG	THERAPY	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	65.000	3.109	55.105	74.895
	2	53.000	4.378	39.067	66.933
2	1	72.000	4.601	57.358	86.642
	2	43.000	4.743	27.904	58.096



→ Enter the data into SPSS.

→ Save the data onto a disk in a file called **suicidaltutors.sav**.

→ Conduct the appropriate analysis to see whether the number of suicidal thoughts patients had was significantly affected by the type of drug they had, the therapy they received or the interaction of the two..

What are the independent variables and how many levels do they have?

Your Answer:

What is the dependent variable?

Your Answer:

What analysis have you performed?

Your Answer:

Describe the assumption of sphericity. Has this assumption been met? (Quote relevant statistics in APA format).

Your Answer:

Report the main effect of therapy in APA format. Is this effect significant and how would you interpret it?

Your Answer:

Report the main effect of 'drug' in APA format. Is this effect significant and how would you interpret it?

Your Answer:

Report the interaction effect between drug and therapy in APA format. Is this effect significant and how would you interpret it?

Your Answer:

In your own time …

Task 1

There is a lot of concern among students as to the consistency of marking between lecturers. It is pretty common that lecturers obtain reputations for being 'hard markers' or 'light markers' but there is often little to substantiate these reputations. So, a group of students investigated the consistency of marking by submitting the same essay to four different lecturers. The mark given by each lecturer was recorded for each of the 8 essays. It was important that the same essays were used for all lecturers because this eliminated any individual differences in the standard of work that each lecturer was marking. The data are below.

- → Enter the data into SPSS.
- → Save the data onto a disk in a file called **tutor.sav**.
- → Conduct the appropriate analysis to see whether the tutor who marked the essay had a significant effect on the mark given.
- → What analysis have you performed?
- → Report the results in APA format?
- → Do the findings support the idea that some tutors give more generous marks than others?

The answers to this task are on the companion website for my SPSS book.

© Prof. Andy Field, 2012

Table 4: Marks of 8 essays by 4 different tutors

Essay	Tutor 1 (Dr. Field)	Tutor 2 (Dr. Smith)	Tutor 3 (Dr. Scrote)	Tutor 4 (Dr. Death)
1	62	58	63	64
2	63	60	68	65
3	65	61	72	65
4	68	64	58	61
5	69	65	54	59
6	71	67	65	50
7	78	66	67	50
8	75	73	75	45

Task 2

In a previous handout we came across the beer-goggles effect: a severe perceptual distortion after imbibing vast quantities of alcohol. Imagine we wanted to follow this finding up to look at what factors mediate the beer goggles effect. Specifically, we thought that the beer goggles effect might be made worse by the fact that it usually occurs in clubs, which have dim lighting. We took a sample of 26 men (because the effect is stronger in men) and gave them various doses of alcohol over four different weeks (0 pints, 2 pints, 4 pints and 6 pints of lager). This is our first independent variable, which we'll call *alcohol consumption*, and it has four levels. Each week (and, therefore, in each state of drunkenness) participants were asked to select a mate in a normal club (that had dim lighting) and then select a second mate in a specially designed club that had bright lighting. As such, the second independent variable was whether the club had dim or bright lighting. The outcome measure was the attractiveness of each mate as assessed by a panel of independent judges. To recap, all participants took part in all levels of the alcohol consumption variable, and selected mates in both brightly- and dimly-lit clubs. This is the example I presented in my handout and lecture in writing up laboratory reports.

→ Enter the data into SPSS.

→ Save the data onto a disk in a file called **BeerGogglesLighting.sav**.

→ Conduct the appropriate analysis to see whether the amount drunk and lighting in the club have a significant effect on mate selection.

→ What analysis have you performed?

→ Report the results in APA format?

→ Do the findings support the idea that mate selection gets worse as lighting dims and alcohol is consumed?

For answers look at the companion website for my SPSS book.

Table 5: Attractiveness of dates selected by people under different lighting and levels of alcohol intake

Dim Lighting				Bright Lighting			
0 Pints	2 Pints	4 Pints	6 Pints	0 Pints	2 Pints	4 Pints	6 Pints
58	65	44	5	65	65	50	33
67	64	46	33	53	64	34	33
64	74	40	21	74	72	35	63
63	57	26	17	61	47	56	31
48	67	31	17	57	61	52	30
49	78	59	5	78	66	61	30
64	53	29	21	70	67	46	46

83	64	31	6	63	77	36	45
65	59	46	8	71	51	54	38
64	64	45	29	78	69	58	65
64	56	24	32	61	65	46	57
55	78	53	20	47	63	57	47
81	81	40	29	57	78	45	42
58	55	29	42	71	62	48	31
63	67	35	26	58	58	42	32
49	71	47	33	48	48	67	48
52	67	46	12	58	66	74	43
77	71	14	15	65	32	47	27
74	68	53	15	50	67	47	45
73	64	31	23	58	68	47	46
67	75	40	28	67	69	44	44
58	68	35	13	61	55	66	50
82	68	22	43	66	61	44	44
64	70	44	18	68	51	46	33
67	55	31	13	37	50	49	22
81	43	27	30	59	45	69	35

Task 3

Imagine I wanted to look at the effect alcohol has on the 'roving eye' (apparently I am rather obsessed with experiments involving alcohol and dating for some bizarre reason). The 'roving eye' effect is the propensity of people in relationships to 'eye-up' members of the opposite sex. I took 20 men and fitted them with incredibly sophisticated glasses that could track their eye movements and record both the movement and the object being observed (this is the point at which it should be apparent that I'm making it up as I go along). Over 4 different nights I plied these poor souls with either 1, 2, 3 or 4 pints of strong lager in a pub. Each night I measured how many different women they eyed-up (a women was categorized as having been eyed up if the man's eye moved from her head to toe and back up again). To validate this measure we also collected the amount of dribble on the man's chin while looking at a woman.

Table 6: Number of women 'eyed-up' by men under different doses of alcohol

1 Pint	2 Pints	3 Pints	4 Pints
15	13	18	13
3	5	15	18
3	6	15	13
17	16	15	14
13	10	8	7
12	10	14	16
21	16	24	15

10	8	14	19
16	20	18	18
12	15	16	13
11	4	6	13
12	10	8	23
9	12	7	6
13	14	13	13
12	11	9	12
11	10	15	17
12	19	26	19
15	18	25	21
6	6	20	21
12	11	18	8

→ Enter the data into SPSS.

→ Save the data onto a disk in a file called **RovingEye.sav**.

→ Conduct the appropriate analysis to see whether the amount drunk has a significant effect on the roving eye.

→ What analysis have you performed?

→ Report the results in APA format?

→ Do the findings support the idea that males tend to eye up females more after they drink alcohol?

For answers look at the companion website for my SPSS book.

Task 4

Western people can become obsessed with body weight and diets, and because the media are insistent on ramming ridiculous images of stick-thin celebrities down into our eyes and brainwashing us into believing that these emaciated corpses are actually attractive, we all end up terribly depressed that we're not perfect (because we don't have a couple of red slugs stuck to our faces instead of lips). This gives evil corporate types the opportunity to jump on our vulnerability by making loads of money on diets that will apparently help us attain the body beautiful! Well, not wishing to miss out on this great opportunity to exploit people's insecurities I came up with my own diet called the 'Andikins diet'[1]. The basic principle is that you eat like me: you eat no meat, drink lots of Darjeeling tea, eat shed-loads of smelly European cheese with lots of fresh crusty bread, pasta, and eat chocolate at every available opportunity, and enjoy a few beers at the weekend. To test the efficacy of my wonderful new diet, I took 10 people who considered themselves to be in need of losing weight (this was for ethical reasons – you can't force people to diet!) and put them on this diet for two months. Their weight was measured in Kilograms at the start of the diet and then after 1 month and 2 months.

Table 7: Weight (Kg) at different times during the Andikins diet

Before Diet	After 1 Month	After 2 Months

[1] Not to be confused with the Atkins diet obviously☺

63.75	65.38	81.34
62.98	66.24	69.31
65.98	67.70	77.89
107.27	102.72	91.33
66.58	69.45	72.87
120.46	119.96	114.26
62.01	66.09	68.01
71.87	73.62	55.43
83.01	75.81	71.63
76.62	67.66	68.60

- → Enter the data into SPSS.
- → Save the data onto a disk in a file called **AndikinsDiet.sav**.
- → Conduct the appropriate analysis to see whether the diet is effective.
- → What analysis have you performed?
- → Report the results in APA format?
- → Does the diet work?

… And Finally The Multiple Choice Test!

Complete the multiple choice questions for **Chapter 14** on the companion website to Field (2013): http://www.uk.sagepub.com/field4e/study/mcqs.htm. If you get any wrong, re-read this handout (or Field, 2013, Chapter 14) and do them again until you get them all correct.

Copyright Information

This handout contains material from:

Field, A. P. (2013). *Discovering statistics using SPSS: and sex and drugs and rock 'n' roll (4th Edition)*. London: Sage.

This material is copyright Andy Field (2000, 2005, 2009, 2013).

References

Field, A. P. (2013). Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll (4th ed.). London: Sage.

Girden, E. R. (1992). ANOVA: Repeated measures. Sage university paper series on quantitative applications in the social sciences, 07-084. Newbury Park, CA: Sage.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95–112.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomised block and split-plot designs. Journal of Educational Statistics, 1(1), 69-82.