





One-Way Independent ANOVA

There goes my hero Watch him as he goes (to hospital)

Children wearing superhero costumes are more likely to harm themselves because of the unrealistic impression of invincibility that these costumes could create: For example, children have reported to hospital with severe injuries because of trying ‘to initiate flight without having planned for landing strategies’ (Davies, Surrudge, Hole, & Munro-Davies, 2007). I can relate to the imagined power that a costume bestows upon you; even now, I have been known to dress up as Fisher by donning a beard and glasses and trailing a goat around on a lead in the hope that it might make me more knowledgeable about statistics. Imagine we had data about the severity of **injury** (on a scale from 0, no injury, to 100, death) for children reporting to the emergency centre at hospitals and information on which superhero costume they were wearing (**hero**): Spiderman, superman, the hulk or a teenage mutant ninja turtle. The Data are in Table 1 and there are descriptive statistics in Output 1. The researcher hypothesized:

- Costumes of ‘flying’ superheroes (i.e. That is, the ones that travel through the air: Superman and Spiderman) will lead to more severe injuries than non-flying ones (the Hulk and Ninja Turtles).
- There will be a diminishing trend in injuries depending on the costume: Superman (most injuries because he flies), Spiderman (next highest injuries because although technically he doesn’t fly, he does climb buildings and throws himself about high up in the air), Hulk (doesn’t tend to fly about in the air much but does smash buildings and punch hard objects that would damage a child if they hit them)¹, and Ninja Turtles (let’s face it, they engage in fairly twee Ninja routines).

Table 1: Data showing the severity of injury sustained by 30 children wearing superhero costumes

Costume	Injury
 Superman	69
	32
	85
	66
	58
 Spiderman	52
	51
	31
	58
	20
	47
	37
49	
40	

¹ Some of you might take issue with this because you probably think of the hulk as a fancy bit of CGI that leaps skyscrapers. However, the ‘proper’ hulk, that is, the one that was on TV during my childhood in the late 1970s (see YouTube) was in fact a real man with big muscles painted green. Make no mistake, he was way scarier than any CGI, but he did not jump over skyscrapers.



Hulk	26
	43
	10
	45
	30
	35
	53
Ninja Turtle	41
	18
	18
	30
	30
	30
	41
	18
	25

Generating Contrasts

Based on what you learnt in the lecture, remember that we need to follow some rules to generate appropriate contrasts:

- **Rule 1:** Choose **sensible comparisons**. Remember that you want to compare only two chunks of variation and that if a group is singled out in one comparison, that group should be excluded from any subsequent contrasts.
- **Rule 2:** Groups coded with positive weights will be compared against groups coded with negative weights. So, assign one chunk of variation positive weights and the opposite chunk negative weights.
- **Rule 3:** The **sum of weights for a comparison should be zero**. If you add up the weights for a given contrast the result should be zero.
- **Rule 4:** If a group is not involved in a comparison, automatically assign it a weight of 0. If we give a group a weight of 0 then this eliminates that group from all calculations
- **Rule 5:** For a given contrast, the weights assigned to the group(s) in one chunk of variation should be equal to the number of groups in the opposite chunk of variation.

Figure 1 shows how we would apply Rule 1 to the Superhero example. We're told that we want to compare flying superheroes (i.e. Superman and Spiderman) against non-flying ones (the Hulk and Ninja Turtles) in the first instance. That will be contrast 1. However, because each of these chunks is made up of two groups (e.g., the flying superheroes chunk comprises both children wearing Spiderman and those wearing Superman costumes), we need a second and third contrast that breaks each of these chunks down into their constituent parts.

To get the weights (Table 2), we apply rules 2 to 5. Contrast 1 compares flying (Superman, Spiderman) to non-flying (Hulk, Turtle) superheroes. Each chunk contains two groups, so the weights for the *opposite* chunks are both 2. We assign one chunk positive weights and the other negative weights (in Table 2 I've chosen the flying superheroes to have positive weights, but you could do it the other way around). Contrast two then compares the two flying superheroes to each other. First we assign both non-flying superheroes a 0 weight to remove them from the contrast. We're left with two chunks: one containing the Superman group and the other containing the Spiderman group. Each chunk contains one group, so the weights for the *opposite* chunks are both 1. We assign one chunk positive weights and the other negative weights (in Table 2 I've chosen to give Superman the positive weight, but you could do it the other way around).

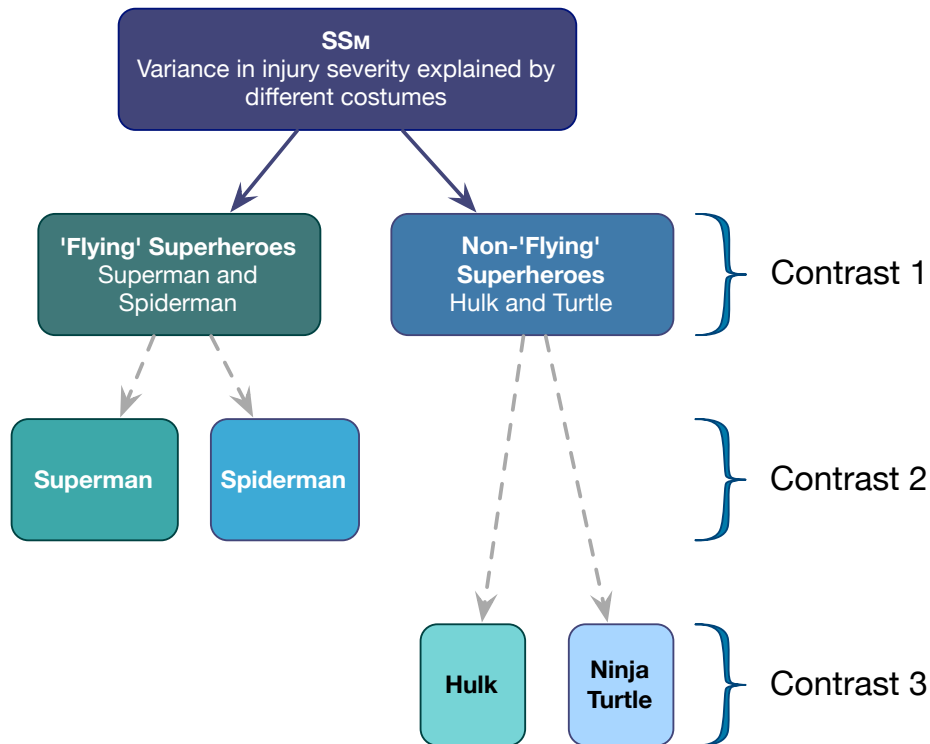


Figure 1: Contrasts for the Superman data

Table 2: Weights for the contrasts in Figure 1

Contrast	Superman	Spiderman	Hulk	Ninja Turtle
Contrast 1	2	2	-2	-2
Contrast 2	1	-1	0	0
Contrast 3	0	0	1	-1

Finally, Contrast three compares the two non-flying superheroes to each other. First we assign both flying superheroes a 0 weight to remove them from the contrast. We're left with two chunks: one containing the Hulk group and the other containing the Turtle group. Each chunk contains one group, so the weights for the *opposite* chunks are both 1. We assign one chunk positive weights and the other negative weights (in Table 2 I've chosen to give the Hulk the positive weight, but you could do it the other way around).

Note that if we add the weights we get 0 in each case (rule 3): $2 + 2 + (-2) + (-2) = 0$ (Contrast 1); $1 + (-1) + 0 + 0 = 0$ (Contrast 2); and $0 + 0 + 1 + (-1) = 0$ (Contrast 3).

Effect Sizes: Cohen's *d*

We discussed earlier in the module that it can be useful not just to rely on significance testing but also to quantify the effects in which we're interested. When looking at differences between means, a useful measure of effect size is Cohen's *d*. This statistic is very easy to understand because it is the difference between two means divided by some estimate of the standard deviation of those means:

$$\hat{d} = \frac{\bar{X}_1 - \bar{X}_2}{s}$$

I have put a hat on the *d* to remind us that we're really interested in the effect size in the population, but because we can't measure that directly, we estimate it from the samples (The hat means 'estimate of'). By dividing by the standard deviation we are expressing the difference in means in standard deviation units (a bit like a *z*-score). The standard deviation is a measure of 'error' or 'noise' in the data, so *d* is effectively a signal-to-noise ratio. The hat means 'estimate of'.



However, if we're using two means, then there will be a standard deviation associated with each of them so which one should we use? There are three choices:

1. If one of the group is a control group it makes sense to use that groups standard deviation to compute d (the argument being that the experimental manipulation might affect the standard deviation of the experimental group, so the control group SD is a 'purer' measure of natural variation in scores)
2. Sometimes we assume that group variances (and therefore standard deviations) are equal (homogeneity of variance) and if they are we can pick a standard deviation from either of the groups because it won't matter.
3. We use what's known as a 'pooled estimate', which is the weighted average of the two group variances. This is given by the following equation:

$$s_p = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}}$$

Let's look at an example. Say we wanted to estimate d for the effect of Superman costumes compared to Ninja Turtle costumes. Output 1 shows us the means, sample size and standard deviation for these two groups:

- Superman: $M = 60.33$, $N = 6$, $s = 17.85$, $s^2 = 318.62$
- Ninja Turtle: $M = 26.25$, $N = 8$, $s = 8.16$, $s^2 = 66.50$

Neither group is a natural control (you would need a 'no costume' condition really), but if we decided that Ninja Turtle (for some reason) was a control (perhaps because Turtles don't fly but supermen do) then d is simply:

$$\hat{d} = \frac{\bar{X}_{Experimental} - \bar{X}_{Control}}{s_{Control}} = \frac{60.33 - 26.25}{8.16} = 4.18$$

In other words, the mean injury severity for people wearing superman costumes is 4 standard deviations greater than for those wearing Ninja Turtle costumes. This is a pretty huge effect.

Cohen (1988, 1992) has made some widely used suggestions about what constitutes a large or small effect: $d = 0.2$ (small), 0.5 (medium) and 0.8 (large). Be careful with these benchmarks because they encourage the kind of lazy thinking that we were trying to avoid and ignore the context of the effect such as the measurement instruments and general norms in a particular research area.

Let's have a look at using the pooled estimate.

$$s_p = \sqrt{\frac{(6 - 1)17.85^2 + (8 - 1)8.16^2}{6 + 8 - 2}} = \sqrt{\frac{1593.11 + 466.10}{12}} = \sqrt{171.60} = 13.10$$

When the group standard deviations are different, this pooled estimate can be useful; however, it changes the meaning of d because we're now comparing the difference between means against all of the background 'noise' in the measure, not just the noise that you would expect to find in normal circumstances. Using this estimate of the standard deviation we get:

$$\hat{d} = \frac{\bar{X}_{Experimental} - \bar{X}_{Control}}{s_{Control}} = \frac{60.33 - 26.25}{13.10} = 2.60$$

Notice that d is smaller now; the injury severity for Superman costumes is about 2 standard deviations greater than for Ninja Turtle Costumes (which is still pretty big)



SELF-TEST: Compute Cohen's d for the effect of Superman costumes on injury severity compared to Hulk and Spiderman costumes. Try using both the standard deviation of the control (the non-Superman costume) and also the pooled estimate. (Answers at the end of the handout)

Running One-Way Independent ANOVA on SPSS

Let's conduct an ANOVA on the injury data. We need to enter the data into the data editor using a coding variable specifying to which of the four groups each score belongs. We need to do this because we have used a between-group design (i.e. different people in each costume). So, the data must be entered in two columns (one called **hero** which

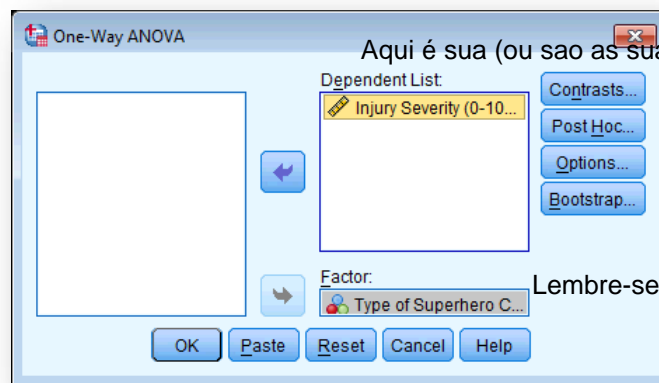
specifies the costume worn and one called **injury** which indicates the severity of injury each child sustained). You can code the variable **hero** any way you wish but I recommend something simple such as 1 = Superman, 2 = Spiderman, 3 = The Hulk, and 4 = Ninja Turtle.

→ Save these data in a file called **Superhero.sav**.



→ Independent Variables are sometimes referred to as **Factors**.

To conduct one-way ANOVA we have to first access the main dialogue box by selecting **Analyze** **Compare Means** **One-Way ANOVA...** (Figure 1). This dialogue box has a space where you can list one or more dependent variables and a second space to specify a grouping variable, or *factor*. Factor is another term for independent variable. For the injury data we need select only **injury** from the variable list and transfer it to the box labelled *Dependent List* by clicking on (or dragging it there). Then select the grouping variable **hero** and transfer it to the box labelled *Factor* by clicking on (or dragging it). If you click on **Contrasts...** you access the dialog box that allows you to conduct planned comparisons, and by clicking on **Post Hoc...** you access the post hoc tests dialog box. These two options will be explained during the next practical class.



Aqui é sua (ou são as suas) variáveis dependentes)

Lembre-se isto é sua variável independente (VI)

Figure 2: Main dialogue box for one-way ANOVA

Planned Comparisons Using SPSS

Click on **Contrasts...** to access the dialogue box that allows you to conduct planned comparisons. The dialogue box is shown in Figure 2 and has two sections. The first section is for specifying trend analyses. If you want to test for trends in the data then tick the box labelled *Polynomial* and select the degree of polynomial you would like. The Superhero data has four groups and so the highest degree of trend there can be is a cubic trend (see Field, 2013 Chapter 11). We predicted that the injuries will decrease in this order: Superman > Spiderman > Hulk > Ninja Turtle. This could be a linear trend, or possibly quadratic (a curved descending trend) but not cubic (because we're not predicting that injuries go down and then up).

It is important from the point of view of trend analysis that we have coded the grouping variable in a meaningful order. To detect a meaningful trend, we need to have coded the groups in the order in which we expect the mean injuries to descend; that is, Superman, Spiderman, Hulk, Ninja Turtle. We have done this by coding the Superman group with the lowest value 1, Spiderman with the next largest value (2), the Hulk with the next largest value (3), and the Ninja Turtle group with the largest coding value of 4. If we coded the groups differently, this would influence both whether a trend is detected, and if by chance a trend is detected whether it is meaningful. For the superhero we predict at most a quadratic trend (see above), so select the polynomial option (*Polynomial*), and then select a quadratic degree by clicking on **Linear** and then selecting *Quadratic* (the drop-down list should now say **Quadratic**) — see Figure 3. If a quadratic trend is selected SPSS will test for both linear and quadratic trends.

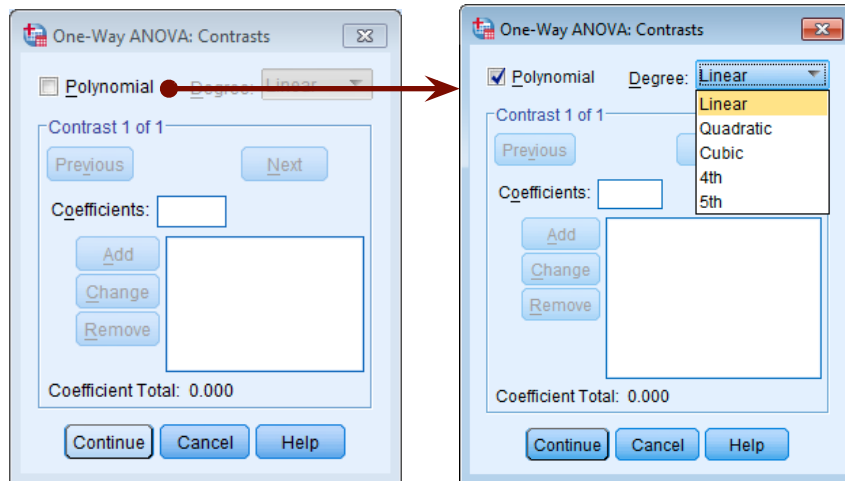


Figure 3: Dialog box for conducting planned comparisons

To conduct planned comparisons we need to tell SPSS what weightings to assign to each group. The first step is to decide which comparisons you want to do and then what weights must be assigned to each group for each of the contrasts (see Field, 2013, Chapter 11). We saw earlier in this handout what sensible contrasts would be, and what weights to give them (see Figure 1 and Table 2). To enter the weights in Table 2 we use the lower part of the dialogue box in Figure 3.

Entering Contrast 1

We will specify contrast 1 first. It is important to make sure that you enter the correct weighting for each group, so you should remember that the first weight that you enter should be the weight for the *first* group (that is, the group coded with the lowest value in the data editor). For the superhero data, the group coded with the lowest value was the Superman group (which had a code of 1) and so we should enter the weighting for this group first. Click in the box labelled *Coefficients* with the mouse and then type '2' in this box and click on **Add**. Next, we input the weight for the second group, which was the Spiderman group (because this group was coded in the data editor with the second highest value). Click in the box labelled *Coefficients* with the mouse and then type '2' in this box and click on **Add**. Next, we input the weight for Hulk group (because it had the next largest code in the data editor), so click in the box labelled *Coefficients* with the mouse and type '-2' and click on **Add**. Finally, we input the code for the last group (the one with the largest code in the data editor), which was the Ninja Turtle group — click in the box labelled *Coefficients* with the mouse and type '-2' and click on **Add**. The box should now look like Figure 4 (left).

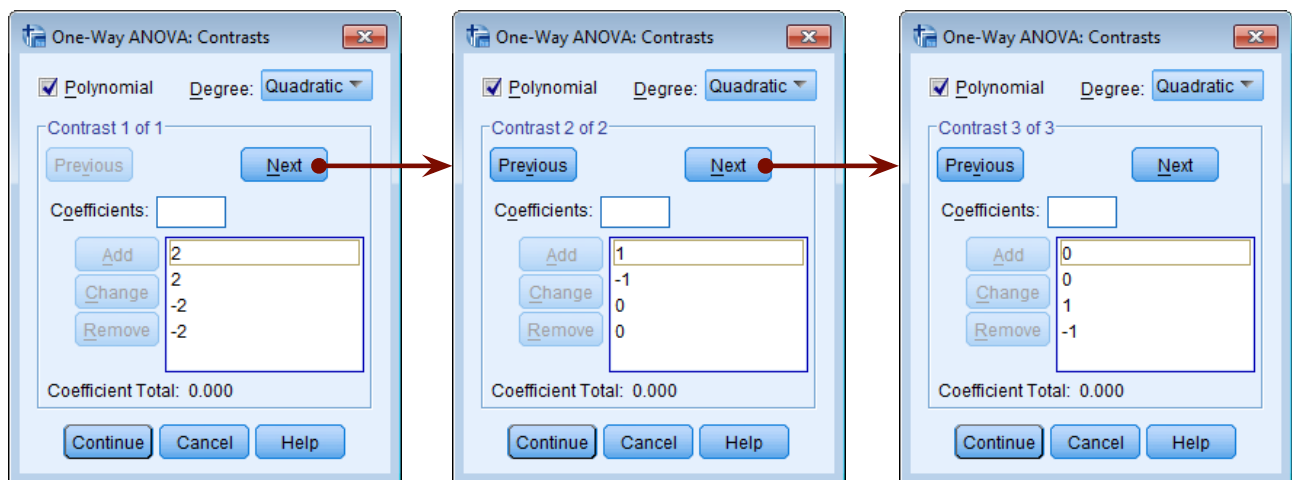


Figure 4: *Contrasts* dialog box completed for the three contrasts of the Superhero data



Once you have input the weightings you can change or remove any one of them by using the mouse to select the weight that you want to change. The weight will then appear in the box labelled *Coefficients* where you can type in a new weight and then click on **Change**. Alternatively, you can click on any of the weights and remove it completely by clicking **Remove**. Underneath the weights SPSS calculates the coefficient total, should equal zero (If you've used the correct weights). If the coefficient number is anything other than zero you should go back and check that the contrasts you have planned make sense and that you have followed the appropriate rules for assigning weights.

Entering Contrast 2

Once you have specified the first contrast, click on **Next**. The weightings that you have just entered will disappear and the dialogue box will now read *contrast 2 of 2*. The weights for contrast 2 should be: 1 (Superman group), -1 (Spiderman group), 0 (Hulk group) and 0 (Ninja Turtle group). We can specify this contrast as before. Remembering that the first weight we enter will be for the Superman group, we must enter the value 1 as the first weight. Click in the box labelled *Coefficients* with the mouse and then type '1' and click on **Add**. Next, we need to input the weight for the Spiderman group by clicking in the box labelled *Coefficients* and then typing '-1' and clicking on **Add**. Then the Hulk group: click in the box labelled *Coefficients*, type '0' and click on **Add**. Finally, we need to input the weight for the Ninja Turtle group by clicking in the box labelled *Coefficients* and then typing '0' and clicking on **Add** (see Figure 4, middle).

Entering Contrast 3

Click on **Next**, and you can enter the weights for the final contrast. The dialogue box will now read *contrast 3 of 3*. The weights for contrast 3 should be: 0 (Superman group), 0 (Spiderman group), 1 (Hulk group) and -1 (Ninja Turtle group). We can specify this contrast as before. Remembering that the first weight we enter will be for the Superman group, we must enter the value 0 as the first weight. Click in the box labelled *Coefficients*, type '0' and click on **Add**. Next, we input the weight for the Spiderman group by clicking in the box labelled *Coefficients* and then typing '0' and clicking on **Add**. Then the Hulk group: click in the box labelled *Coefficients*, type '1' and click on **Add**. Finally, we input the weight for the Ninja Turtle group by clicking in the box labelled *Coefficients*, typing '-1' and clicking on **Add** (see Figure 4, right).

When all of the planned contrasts have been specified click on **Continue** to return to the main dialogue box.

Post Hoc Tests in SPSS

Normally if we have done planned comparisons we should not do post hoc tests (because we have already tested the hypotheses of interest). Likewise, if we choose to conduct post hoc tests then planned contrasts are unnecessary (because we have no hypotheses to test). However, for the sake of space we will conduct some post hoc tests on the superhero data. Click on **Post Hoc...** in the main dialogue box to access the post hoc tests dialogue box (Figure 5). The choice of comparison procedure will depend on the exact situation you have and whether it is more important for you to keep strict control over the familywise error rate or to have greater statistical power. I have drawn some general guidelines:



Field (2013) recommends:

- When you have equal sample sizes and you are confident that your population variances are similar then use R-E-G-W-Q or Tukey because both have good power and tight control over the Type I error rate.
- If sample sizes are slightly different then use Gabriel's procedure because it has greater power, but if sample sizes are very different use Hochberg's GT2.
- If there is any doubt that the population variances are equal then use the Games-Howell procedure because this seems to generally offer the best performance.

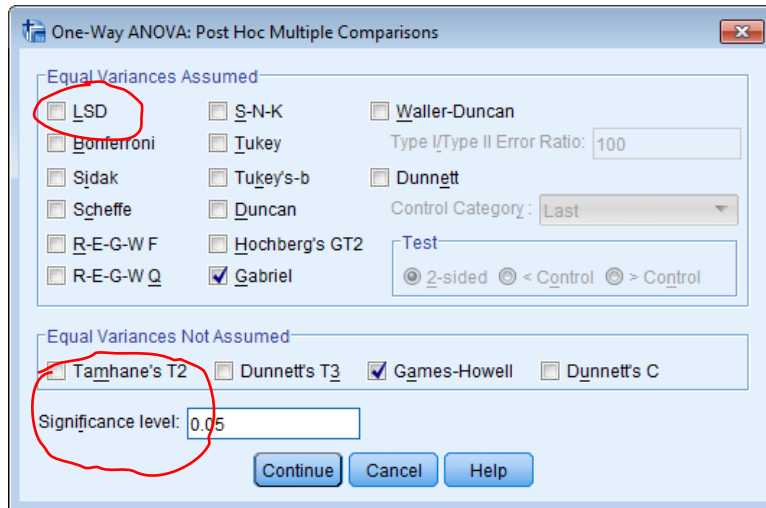


Figure 5: Dialogue box for specifying post hoc tests

I recommend running the Games-Howell procedure in addition to any other tests you might select because of the uncertainty of knowing whether the population variances are equivalent. For the superhero data there are slightly unequal sample sizes and so we will use Gabriel's test (see Tip above). When the completed dialogue box looks like Figure 5 click on [Continue](#) to return to the main dialogue box.

Options

The additional options for one-way ANOVA are fairly straightforward. The dialog box to access these options can be obtained by clicking on [Options...](#). First you can ask for some descriptive statistics, which will display a table of the means, standard deviations, standard errors, ranges and confidence intervals for the means of each group. This is a useful option to select because it assists in interpreting the final results. You can also select *Homogeneity-of-variance* tests. Earlier in the module we saw that there is an assumption that the variances of the groups are equal and selecting this option tests this assumption using Levene's test (see your handout on bias). SPSS offers us two alternative versions of the *F*-ratio: the Brown-Forsythe *F* (1974), and the Welch *F* (1951). These alternative *F*s can be used if the homogeneity of variance assumption is broken. If you're interested in the details of these corrections then see Field (2013), but if you've got better things to do with your life then take my word for it that they're worth selecting just in case the assumption is broken. You can also select a *Means plot* which will produce a line graph of the means. Again, this option can be useful for finding general trends in the data. When you have selected the appropriate options, click on [Continue](#) to return to the main dialog box. Click on [OK](#) in the main dialog box to run the analysis.

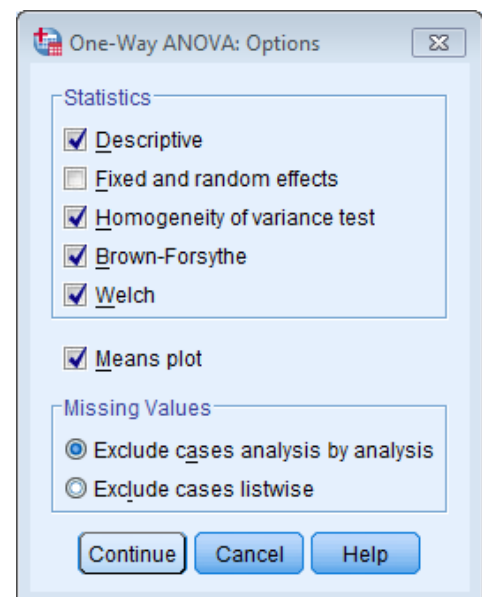


Figure 6: Options for One-Way ANOVA

Bootstrapping

Also in the main dialog box is the alluring [Bootstrap...](#) button. We have seen in the module that bootstrapping is a good way to overcome bias, and this button glistens and tempts us with the promise of untold riches, like a diamond in a bull's rectum. However, if you use bootstrapping it'll be as disappointing as if you reached for that diamond only to discover that it's a piece of glass. You might, not unreasonably, think that if you select bootstrapping it'd do a nice bootstrap of the *F*-statistic for you. It won't. It will bootstrap confidence intervals around the means (if you ask for descriptive statistics), contrasts and differences between means (i.e., the post hoc tests). This, of course, can be useful, but the main test won't be bootstrapped.



Output from One-Way ANOVA

Descriptive Statistics

Figure 7 shows the ‘means plot’ that we asked SPSS for. A few important things to note are:

- ✘ It looks horrible.
- ✘ SPSS has scaled the y-axis to make the means look as different as humanly possible. Think back to week 1 when we learnt that it was very bad to scale your graph to maximise group differences – SPSS has not read my handout☹
- ✘ There are no error bars: the graph just isn’t very informative because we aren’t given confidence intervals for the mean.

Paste something like this into one of your lab reports and watch your tutor recoil in horror and your mark plummet! The moral is, never let SPSS do your graphs for you☹



→ Using what you learnt in week one draw an error bar chart of the data. (Your chart should ideally look like Figure 8).

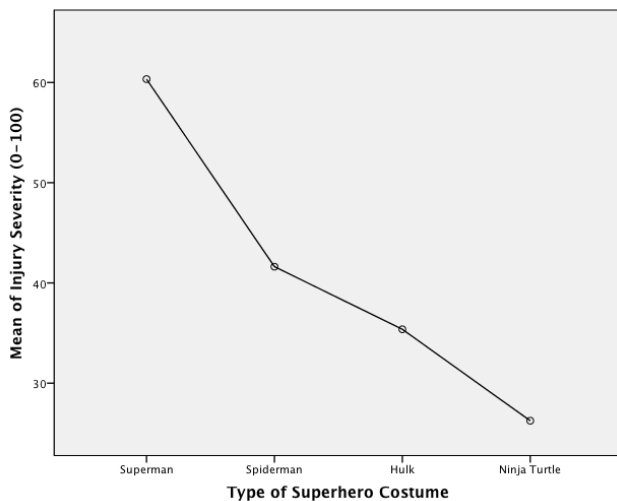


Figure 7: Crap graph automatically produced by SPSS

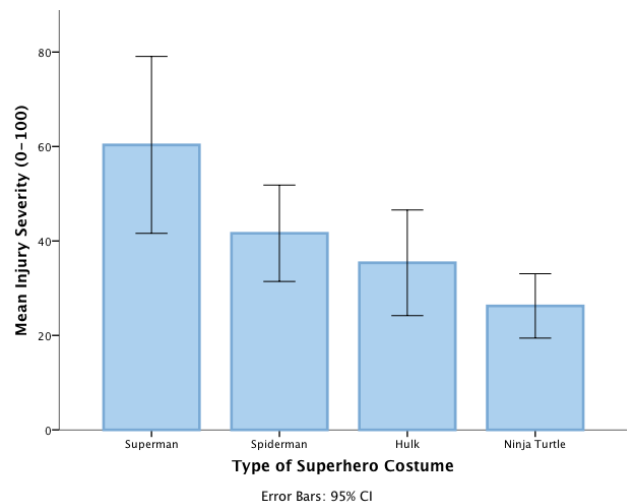


Figure 8: Nicely edited error bar chart of the injury data

Figure 8 shows an error bar chart of the injury data. The means indicate that some superhero costumes do result in more severe injuries than others. Notably, the Ninja Turtle costume seems to result in less severe injuries and the Superman costume results in most severe injuries. The error bars (the I shapes) show the 95% confidence interval around the mean.

Lembre-se do SEM



→ Think back to the start of the module, what does a confidence interval represent?

If we were to take 100 samples from the same population, the true mean (the mean of the population) would lie somewhere between the top and bottom of that bar in 95 of those samples. In other words, these are the limits between which the population value for the mean injury severity in each group will (probably) lie. If these bars do not overlap then we expect to get a significant difference between means because it shows that the population means of those two samples are likely to be different (they don’t fall within the same limits). So, for example, we can tell that Ninja Turtle related injuries are likely to be less severe than those of children wearing superman costumes (the error bars don’t overlap) and Spiderman costumes (only a small amount of overlap)



Descriptives

Injury Severity (0-100)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Superman	6	60.33	17.851	7.288	41.60	79.07	32	85
Spiderman	8	41.63	12.212	4.317	31.42	51.83	20	58
Hulk	8	35.38	13.384	4.732	24.19	46.56	10	53
Ninja Turtle	8	26.25	8.155	2.883	19.43	33.07	18	41
Total	30	39.60	17.158	3.133	33.19	46.01	10	85

Output 1

The table of descriptive statistics verifies what the graph shows: that the injuries for Superman costumes were most severe, and for ninja turtle costumes were least severe. This table also provides the confidence intervals upon which the error bars were based.

Levene's test

Levene's test (think back to your lecture and handout on bias) is designed to test the null hypothesis that the variances of the groups are the same. In this case Levene's test is testing whether the variances of the four groups are significantly different.



→ If Levene's test is significant (i.e. the value of sig. is *less than .05*) then we can conclude that the variances are significantly different. This would mean that we had violated one of the assumptions of ANOVA and we would have to take steps to rectify this matters by (1) transforming all of the data (see your handout on bias), (2) bootstrapping (not implemented in SPSS for ANOVA, or (3) using a corrected test (see below). Remember that how we interpret Levene's test depends on the size of sample we have (see the handout on bias).

Test of Homogeneity of Variances

Injury Severity (0-100)

Levene Statistic	df1	df2	Sig.
.891	3	26	.459

Output 2

For these data the variances are relatively similar (hence the high probability value). Typically people would interpret this result as meaning that we can assume homogeneity of variance (because the observed p -value of .459 is greater than .05). However, our sample size is fairly small (some groups had only 6 participants). The small sample (per group) will limit the power of Levene's test to detect differences between the variances. We could also look at the variance ratio. The smallest variance was for the Ninja Turtle costume ($8.16^2 = 66.59$) and the largest was for Superman costumes ($17.85^2 = 318.62$). The ratio of these values is $318.62/66.59 = 4.78$. In other words, the largest variance is almost five time larger than the smallest variance. This difference is quite substantial. Therefore, we might reasonably assume that variances are not homogenous.

For the main ANOVA, we selected two procedures (Brown-Forsythe and Welch) that should be accurate when homogeneity of variance is not true. So, we should perhaps inspect these F -values in the main analysis. We might also choose a method of post hoc test that does not rely on the assumption of equal variances (e.g., the Games-Howell procedure).



The Main ANOVA

Output 3 shows the main ANOVA summary table. The output you will see is the table at the bottom of Output 3, this is a more complicated table than a simple ANOVA table because we asked for a trend analysis of the means (by selecting the select the polynomial option Polynomial in Figure 3). The top of Output 3 shows what you would see if you hadn't done the trend analysis just note that the *Between Groups*, *Within Groups* and *Total* rows in both tables are the same — it's just that the bottom table decomposes the *Between Groups* effect into linear and quadratic trends.

The table you'll see is divided into between group effects (effects due to the experiment) and within group effects (this is the unsystematic variation in the data). The between-group effect is the overall experimental effect (the effect of wearing different costumes on the severity of injuries). In this row we are told the sums of squares for the model ($SS_M = 4180.62$). The sum of squares for the model represents the total experimental effect whereas the mean squares for the model represents the average experimental effect. The row labelled *within group* gives details of the unsystematic variation within the data (the variation due to natural individual differences in physique and tolerance to injury). The table tells us how much unsystematic variation exists (the residual sum of squares, SS_R). It then gives the average amount of unsystematic variation, the residual mean squares (MS_R). The test of whether the group means are the same is represented by the *F*-ratio for the combined between-group effect. The value of this ratio is 8.32. The final column labelled *sig.* indicates how likely it is that an *F*-ratio of at least that size would have occurred if there were no differences between means. In this case, there is a probability of 0.000 (that's less than a .1% chance!). We have seen that, scientists tend to use a cut of point of .05 as their criterion for statistical significance. Hence, because the observed significance value is less than .05 we can say that there was a significant effect of the costume worn on the severity of injuries sustained. However, at this stage we still do not know exactly what the effect of each costume was (we don't know which groups differed). Also, we know from previous lectures that thinking about significance in this black and white way is not always helpful and we should consider other evidence (such as the effect sizes we computed at the beginning of this handout).

ANOVA

Injury Severity (0-100)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4180.617	3	1393.539	8.317	.000
Within Groups	4356.583	26	167.561		
Total	8537.200	29			

ANOVA

Injury Severity (0-100)

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		4180.617	3	1393.539	8.317	.000
	Linear Term	Unweighted	4094.696	1	4094.696	24.437	.000
		Weighted	3927.138	1	3927.138	23.437	.000
		Deviation	253.479	2	126.740	.756	.479
	Quadratic Term	Unweighted	169.551	1	169.551	1.012	.324
		Weighted	160.976	1	160.976	.961	.336
		Deviation	92.503	1	92.503	.552	.464
Within Groups			4356.583	26	167.561		
Total			8537.200	29			

Output 3

Trend Analysis

The trend analysis breaks down the experimental effect into that which can be explained by both a linear and quadratic relationship. It's confusing that for both trends you get three rows (labelled *Unweighted*, *Weighted* and



Deviation)² but focus on the row labelled *Weighted*³. First, let's look at the linear component. This comparison tests whether the means change across groups in a linear way. The sum of squares and mean squares are given, but the most important things to note are the value of the *F*-ratio and the corresponding significance value. For the linear trend the *F*-ratio is 23.44 and this value is significant at a $p < .001$ level of significance. Looking at Figure 8 we can interpret this trend as the mean severity of injuries decreased proportionately across the four superhero costumes. Moving onto the quadratic trend, this comparison is testing whether the pattern of means is curvilinear (i.e. is represented by a curve with one bend in). Figure 8 does not particularly suggest that the means can be represented by a curve and the results for the quadratic trend bear this out. The *F*-ratio for the quadratic trend is non-significant, $F(1, 26) = 0.91, p = .336$.

Robust tests

We saw earlier on that the assumption of homogeneity of variance was questionable (at least in terms of the variance ratio). Therefore, we should inspect Output 4, which has the Brown-Forsythe and Welch versions of the *F*-ratio. If you look at this table you should notice that both test statistics are still highly significant (the value of *Sig.* in the table is less than .05). Therefore, we can say that there was a significant effect of the costume worn on the severity of injuries.

Robust Tests of Equality of Means

Injury Severity (0–100)

	Statistic ^a	df1	df2	Sig.
Welch	7.100	3	13.018	.005
Brown-Forsythe	7.679	3	16.926	.002

a. Asymptotically F distributed.

Output 4

To find out where the differences between groups lie, you need to carry out further comparisons. There are two choices of comparison: the first is a planned comparison, in which predictions about which groups will differ were made prior to the experiment, and post hoc tests for which all groups are compared because no prior hypotheses about group differences were made. Let's look at these in turn.

Output for Planned Comparisons

We told SPSS to conduct three planned comparisons: one to test whether 'flying' superhero costumes led to worse injuries than 'non-flying' superhero costumes; the second to compare injury severity for the two flying superhero costumes (Superman vs. Spiderman costumes); and the third to compare injury severity for the two non-flying superhero costumes (Hulk vs. Ninja Turtle costumes). Output 5 shows the results of the planned comparisons that we requested. The first table displays the contrast coefficients and it is well worth looking at this table to double check that the contrasts are comparing what they are supposed to: they should correspond to Table 2, which they do. If they don't then you've entered the weights incorrectly (see Figure 4).

The second table gives the statistics for each contrast. The first thing to notice is that statistics are produced for situations in which the group variances are equal, and when they are unequal. Typically, if Levene's test was significant then you should read the part of the table labelled *Does not assume equal variances*; if Levene's test was not significant you use the part of the table labelled *Assume equal variances*. For these data Levene's test was not significant implying that we can assume equal variances; however, the variance ratio suggested that actually this assumption of homogeneity might be unreasonable (and that Levene's test might have been non-significant because of the small sample size). Therefore, based on the variance ratio we probably should not assume equal variances and instead use the part of the table labelled *Does not assume equal variances*.

² If you have equal sample sizes you just get two versions: one labeled *Contrast* and the other labeled *Deviation*. With unequal sample sizes, SPSS produces an unweighted and weighted version of the contrast. The weighted version factors in the different samples sizes.

³ If you have equal sample sizes then focus on the row labelled *Contrast*.

The table tells us the value of the contrast itself, the associated t -test and the two-tailed significance value. Hence, for contrast 1, we can say that injury severity was significantly different in kids wearing costumes of flying superheroes compared to those wearing non-flying superhero costumes, $t(15.10) = 3.99, p = .001$. Contrast 2 tells us that injury severity was not significantly different in those wearing Superman costumes compared to those wearing Spiderman costumes, $t(8.39) = 2.21, p = .057$. Finally, contrast 2 tells us that injury severity was not significantly different in those wearing Hulk costumes compared to those wearing Ninja Turtle costumes, $t(11.57) = 1.65, p = .126$.



→ **Contrast 2:** Bear in mind what we've discussed before on this module about sample size and significance. This effect is quite close to significance ($p = .057$) and is based on a small sample. Note also, that if we had assumed equal variances the p -value is below the .05 threshold. It would be particularly useful here to look at the effect size – in fact if you did the earlier self test you'll have found that $d = 1.53$ and 1.26 (using the pooled variance estimate), which is a very large effect. There is clearly something going on between the Superman and Spiderman conditions that is not reflected in the significance value because of the very small sample (remember that p is affected by the sample size).

Contrast Coefficients

Contrast	Type of Superhero Costume			
	Superman	Spiderman	Hulk	Ninja Turtle
1	2	2	-2	-2
2	1	-1	0	0
3	0	0	1	-1

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Injury Severity (0-100)	Assume equal variances	1	80.67	19.054	4.234	26	.000
		2	18.71	6.991	2.676	26	.013
		3	9.13	6.472	1.410	26	.170
	Does not assume equal variances	1	80.67	20.244	3.985	15.097	.001
		2	18.71	8.471	2.209	8.388	.057
		3	9.13	5.541	1.647	11.568	.126

Output 5

Output for Post Hoc Tests

If we had no specific hypotheses about the effect that different superhero costumes would have on the severity of injuries, then we could carry out post hoc tests to compare all groups of participants with each other. In fact, we asked SPSS to do this (see earlier) and the results of this analysis are shown in Output 6. This table shows the results of Gabriel's test and the Games-Howell procedure, which were specified earlier on. If we look at Gabriel's test first, each group of children is compared with all of the remaining groups. For each pair of groups the difference between group means is displayed, the standard error of that difference, the significance level of that difference and a 95% confidence interval. First of all, the Superman group is compared to the Spiderman group and reveals a nonsignificant difference ($Sig.$ is greater than .05), but when compared to the Hulk group ($p = .008$) and the Turtle group ($p < .001$) there is a significant difference ($Sig.$ is less than .05).

Next, the Spiderman group is compared to all other groups. The first comparison (to Superman) is identical to the one that we have already looked at. The only new information is the comparison of the Spiderman group to the Hulk ($p = .907$, not significant) and Turtle (not significant, $p = .136$) groups.

Finally, the Hulk group is compared to all other groups. Again, the first two comparisons replicate effects that we have already seen in the table, the only new information is the comparison of the Hulk group with the Ninja Turtle group ($p = .650$, not significant).

The rest of the table describes the Games-Howell tests and a quick inspection reveals two differences to the conclusions found with the Gabriel test: (1) the Superman and Hulk groups no longer differ significantly ($p = .073$ instead of .008); (2) the Spiderman and Turtle groups just about differ significantly ($p = .050$ instead of .136). In this



situation, what you conclude depends upon whether you think it's reasonable to assume that population variances differ. We can use the samples as a guide. Table 3 shows the variances in each group, and also the variance ratios for all pairs of groups (i.e. the larger of the two variances divided by the smaller). Note that all but one variance ratios are close to or above 2 (indicating heterogeneity). Most important let's look at the two comparisons where the Games-Howell test differs to the Gabriel test:

1. Superman vs. Hulk: the variance ratio is below 2 (although close) so we might choose to report Gabriel's test and accept a significant difference.
2. Spiderman vs. Turtle: the variance ratio is above 2 so we might choose to report the Games-Howell test and accept a significant difference.

The main point to take home is that data analysis can lead you into complex situations in which you have to make informed decisions about how to interpret the data.

Multiple Comparisons

Dependent Variable: Injury Severity (0-100)

	(I) Type of Superhero Costume	(J) Type of Superhero Costume	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Gabriel	Superman	Spiderman	18.708	6.991	.070	-1.05	38.46
		Hulk	24.958*	6.991	.008	5.20	44.71
		Ninja Turtle	34.083*	6.991	.000	14.33	53.84
	Spiderman	Superman	-18.708	6.991	.070	-38.46	1.05
		Hulk	6.250	6.472	.907	-12.09	24.59
		Ninja Turtle	15.375	6.472	.136	-2.96	33.71
	Hulk	Superman	-24.958*	6.991	.008	-44.71	-5.20
		Spiderman	-6.250	6.472	.907	-24.59	12.09
		Ninja Turtle	9.125	6.472	.650	-9.21	27.46
	Ninja Turtle	Superman	-34.083*	6.991	.000	-53.84	-14.33
		Spiderman	-15.375	6.472	.136	-33.71	2.96
		Hulk	-9.125	6.472	.650	-27.46	9.21
Games-Howell	Superman	Spiderman	18.708	8.471	.197	-8.13	45.55
		Hulk	24.958	8.689	.073	-2.19	52.10
		Ninja Turtle	34.083*	7.837	.016	7.69	60.48
	Spiderman	Superman	-18.708	8.471	.197	-45.55	8.13
		Hulk	6.250	6.406	.765	-12.39	24.89
		Ninja Turtle	15.375*	5.192	.050	.00	30.75
	Hulk	Superman	-24.958	8.689	.073	-52.10	2.19
		Spiderman	-6.250	6.406	.765	-24.89	12.39
		Ninja Turtle	9.125	5.541	.392	-7.42	25.67
	Ninja Turtle	Superman	-34.083*	7.837	.016	-60.48	-7.69
		Spiderman	-15.375*	5.192	.050	-30.75	.00
		Hulk	-9.125	5.541	.392	-25.67	7.42

*. The mean difference is significant at the 0.05 level.

Injury Severity (0-100)

Type of Superhero Costume	N	Subset for alpha = 0.05	
		1	2
Gabriel ^{a,b} Ninja Turtle	8	26.25	
Hulk	8	35.38	
Spiderman	8	41.63	41.63
Superman	6		60.33
Sig.		.164	.057

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 7.385.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Output 6



Table 3: Variances and variance ratios for all groups in the Superhero data

	Variance	Variance Ratio to Spiderman	Variance Ratio to Hulk	Variance Ratio to Ninja Turtle
Superman	318.66	2.14	1.78	4.79
Spiderman	149.13		0.83	2.24
Hulk	179.13			2.69
Ninja Turtle	66.50			

Reporting Results from One-Way Independent ANOVA

When we report an ANOVA, we give details of the F -ratio and the degrees of freedom from which it was calculated. For the experimental effect in these data the F -ratio was derived from dividing the mean squares for the effect by the mean squares for the residual. Therefore, the degrees of freedom used to assess the F -ratio are the degrees of freedom for the effect of the model ($df_M = 3$) and the degrees of freedom for the residuals of the model ($df_R = 26$):

- ✓ There was a significant effect of the costume worn on the severity of injuries sustained, $F(3, 26) = 8.32, p < .001$.

Notice that the value of the F -ratio is preceded by the values of the degrees of freedom for that effect. However, in this example, we had some evidence that homogeneity of variance was violated, so we might report the alternative statistics (which can be done in the same way). Notice that the degrees of freedom have changed to reflect how the F -ratio was calculated, and that the value of F itself is different. Note also that unless $p < .001$ it is good practice to report the exact p -value; this is because it is more informative to know the exact value of p than to know only that it was bigger or smaller than .05. The APA recommends reporting exact p -values.

- ✓ The assumption of homogeneity of variance was violated; therefore, the Brown-Forsythe F -ratio is reported. There was a significant effect of the costume worn on the severity of injuries sustained, $F(3, 16.93) = 7.68, p = .005$.
- ✓ The assumption of homogeneity of variance was violated; therefore, the Welch F -ratio is reported. There was a significant effect of the costume worn on the severity of injuries sustained, $F(3, 13.02) = 7.10, p = .002$.

We can report contrasts and trends in much the same way:

- ✓ The mean severity of injuries decreased proportionately across the four superhero costumes, $F(1, 26) = 23.44, p < .001$.
- ✓ Planned contrasts revealed that injury severity was significantly different in children wearing costumes of flying superheroes compared to those wearing non-flying superhero costumes, $t(15.10) = 3.99, p = .001$. Injury severity was not significantly different in those wearing Superman costumes compared to those wearing Spiderman costumes, $t(8.39) = 2.21, p = .057$, nor between those wearing Hulk costumes compared to those wearing Ninja Turtle costumes, $t(11.57) = 1.65, p = .126$.



SELF-TEST: Compute Cohen's d for the effect of Spiderman costumes on injury severity compared to Hulk and Ninja Turtle costumes, and between the Hulk and Ninja Turtle conditions. Try using both the standard deviation of the control (the non-Superman costume) and also the pooled estimate. (Answers at the end of the handout)

Post hoc tests are usually reported just with p -values and effect sizes:

- ✓ In general homogeneity could not be assumed between pairs of groups except for the Hulk group with both Superman and Spiderman. Where homogeneity could not be assumed Games-Howell post hoc tests were used, where local homogeneity could be assumed Gabriel's test was used. These tests revealed significant differences between the Superman group and both the Hulk, $p = .008, d = 1.62$, and Ninja Turtle, $p = .016, d =$



2.60, groups and the Spiderman and Turtle groups, $p = .050$, $d = 1.48$. There were no significant differences between the Spiderman and both Superman, $p = .197$, $d = 1.26$, and Hulk, $p = .907$, $d = 0.49$, groups, or between the Hulk and Ninja Turtle group, $p = .392$, $d = 0.82$.

Guided Task

The University was interested on the effects of different statistics classes on aggression in undergraduates. Following one of three types of statistic class (workshops, lectures and an exam) 6 students were placed in a relaxation room in which there was a dartboard with the face of their lecturer pinned to it. The number of darts that each student threw at the dartboard was measured.

Class	Darts
Workshop	10
	8
	15
	26
	28
	12
Lecture	12
	6
	30
	24
	18
	13
Exam	18
	40
	35
	29
	30
	25



- Enter the data into SPSS.
- Save the data onto a disk in a file called **StatsClass.sav**
- Draw an error-bar chart of the data.
- Carry out one-way ANOVA to find out whether the type of statistics class affects aggression.
- Extra activity if you have time: Use what you learnt in week 1 to screen the data.

Are the data normally distributed? (Report some relevant statistics in your answer in APA format)



Your Answer:

Based on the error bar chart, which groups (if any) do you think will be significantly different and why?

Your Answer:

What is the assumption of homogeneity of variance? Has this assumption been met (quote relevant statistics in APA format)? If the assumption has not been met what action should be taken?

Your Answer:

Does the type of statistics class significantly affect aggression (quote relevant statistics in APA format)? What does the *F*-value represent?



Your Answer:

What does the p value reported above mean?

Your Answer:

Unguided Task 1

In the lectures (and my book) we used an example about the drug Viagra. Suppose we tested the belief that Viagra increases libido by taking three groups of people and administering one group with a placebo (such as a sugar pill), one group with a low dose of Viagra and one with a high dose. The dependent variable was an objective measure of libido. In the lecture we established two useful planned comparisons that we could do to test these hypotheses:

1. Having a dose of Viagra will increase libido compared to not having any.
2. Having a high dose of Viagra will increase libido more than a low dose.



- Enter the data into SPSS creating two variables: one called **dose** which specifies how much Viagra the person was given and one called **libido** which indicates the person's libido over the following week). You can code the variable **dose** any way you wish but I recommend something simple such as 1 = placebo, 2 = low dose and 3 = high dose.
- Save the data onto a disk in a file called **Viagra.sav**.
- Using what we learnt in the lectures, do a one-way ANOVA to test whether Viagra had a significant effect on libido and also define contrasts to test the two hypotheses.
- A complete answer can be found in Chapter 11 of my textbook (Field, 2013)

Table 4: Data showing how libido differs after different doses of Viagra

Dose	Libido
Placebo	3
	2
	1
	1
	4
Low Dose Viagra	5
	2



	4
	2
	3
High Dose Viagra	7
	4
	5
	3
	6

Unguided Task 2

The organisers of the Rugby world cup were interested whether certain teams were more aggressive than others. Over the course of the competition, they noted the injury patterns of players in the England squad after certain games. The dependent variable was the number of injuries sustained by each player in a match, and the independent variable was the team they had played. Different players were used in the different matches (to avoid injuries from previous matches carrying over into new matches).



- Enter the data into SPSS and save the data onto a disk in a file called **rugby.sav**.
- Conduct one-way ANOVA to see whether the number of injuries inflicted differed across the rugby teams.
- Conduct planned comparisons testing the hypotheses that: (1) Tonga cause more injuries than all of the other teams; (2) Japan cause less injuries than Wales and New Zealand; (3) Wales and New Zealand are no different in terms of the injuries inflicted.

Team	Injuries
Japan	0
	35
	31
	29
	20
	7
	43
	16
Wales	30
	40
	27
	25
	40
	15
	30
	46
New Zealand	16
	33
	25
	32
	20
	54
	57



	19
	55
	57
	55
Tonga	57
	56
	53
	59
	55

Unguided Task 3

I read a story in a newspaper recently claiming that scientists had discovered that the chemical genistein, which is naturally occurring in Soya, was linked to lowered sperm counts in western males. In fact, when you read the actual study, it had been conducted on rats, it found no link to lowered sperm counts but there was evidence of abnormal sexual development in male rats (probably because this chemical acts like oestrogen). The journalist naturally interpreted this as a clear link to apparently declining sperm counts in western males (bloody journalists!). Anyway, as a Vegetarian who eats lots of Soya products and probably would like to have kids one day, imagine I wanted to test this idea in humans rather than rats. I took 80 males and split them into four groups that varied in the number of Soya meals they ate per week over a year long period. The first group was a control group and they had no Soya meals at all per week (i.e. none in the whole year); the second group had 1 Soya meal per week (that's 52 over the year); the third group had 4 Soya meals per week (that's 208 over the year) and the final group had 7 Soya meals a week (that's 364 over the year). At the end of the year, all of the participants were sent away to produce some sperm that I could count (when I say 'I', I mean someone in a laboratory as far away from me as humanly possible). Data are below (although in a different format to how it should be entered into SPSS).



- Enter the data into SPSS (bear in mind that they are not entered in the same way as the table below).
- Save the data onto a disk in a file called **Sperm.sav**
- Are the data normally distributed?
- Is the assumption of homogeneity of variance met?
- Carry out one-way ANOVA to find out whether Genistein affects sperm counts.
- Test for a linear trend and do post hoc tests. Which groups differ from which?
- Answers can be found on the companion website for my textbook (Smart Alex answers)

No Soya Sperm (Millions)	1 Soya Meal Sperm (Millions)	4 Soya Meals Sperm (Millions)	7 Soya Meals Sperm (Millions)
0.35	0.33	0.40	0.31
0.58	0.36	0.60	0.32
0.88	0.63	0.96	0.56
0.92	0.64	1.20	0.57
1.22	0.77	1.31	0.71
1.51	1.53	1.35	0.81
1.52	1.62	1.68	0.87
1.57	1.71	1.83	1.18
2.43	1.94	2.10	1.25



2.79	2.48	2.93	1.33
3.40	2.71	2.96	1.34
4.52	4.12	3.00	1.49
4.72	5.65	3.09	1.50
6.90	6.76	3.36	2.09
7.58	7.08	4.34	2.70
7.78	7.26	5.81	2.75
9.62	7.92	5.94	2.83
10.05	8.04	10.16	3.07
10.32	12.10	10.98	3.28
21.08	18.47	18.21	4.11

Unguided Task 4

People love their mobile phones, which is rather worrying given some recent controversy about links between mobile phone use and brain tumours. The basic idea is that mobile phones emit microwaves, and so holding one next to your brain for large parts of the day is a bit like sticking your brain in a microwave oven and selecting the 'cook until well done' button. If we wanted to test this experimentally, we could get 6 groups of people and strap a mobile phone on their heads (that they can't remove). Then, by remote control, we turn the phones on for a certain amount of time each day. After 6 months, we measure the size of any tumour (in mm³) close to the site of the phone antennae (just behind the ear). The six groups experienced 0, 1, 2, 3, 4 or 5 hours per day of phone microwaves for 6 months. Data are below (although in a different format to how it should be entered into SPSS). (This example is from Field & Hole, 2003, so there is a very detailed answer in there if you're interested).



- Enter the data into SPSS (bear in mind that they are not entered in the same way as the table below).
- Save the data onto a disk in a file called **Tumour.sav**
- Are the data normally distributed?
- Is the assumption of homogeneity of variance met?
- Carry out one-way ANOVA to find out whether mobile phone use causes brain tumours. Test for a linear trend.
- Answers can be found on the companion website for my textbook (Smart Alex answers)

0 Hours per day	1 Hours per day	2 Hours per day	3 Hours per day	4 Hours per day	5 Hours per day
0.02	0.77	1.29	4.31	4.65	5.17
0.00	0.74	1.08	2.47	5.16	5.03
0.01	0.22	1.07	2.04	4.06	6.14
0.01	0.94	0.48	3.32	4.61	4.90
0.04	0.62	1.26	3.18	5.32	4.65
0.04	0.33	0.52	2.24	4.84	3.88
0.01	0.47	0.64	3.80	5.20	5.25
0.03	0.78	1.69	2.86	5.40	2.70
0.00	0.76	1.85	2.58	3.04	5.31
0.01	0.03	1.22	4.09	4.73	5.36
0.01	0.93	1.69	3.51	5.18	5.43

0.02	0.39	2.34	3.02	5.09	4.43
0.03	0.62	1.54	3.63	5.65	4.83
0.02	0.48	1.87	4.14	3.83	4.13
0.01	0.15	0.83	2.82	4.88	4.05
0.04	0.72	0.98	1.77	5.21	3.67
0.03	0.38	0.98	2.56	5.30	4.43
0.02	0.27	1.39	3.59	6.05	4.62
0.01	0.00	1.63	2.64	4.13	5.11
0.01	0.69	0.87	1.84	5.42	5.55

Multiple Choice Test



Complete the multiple choice questions for **Chapter 11** 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 13) and do them again until you get them all correct.

This handout is based on 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).

Effect Size Answers

Table 5: Cohen's d for the superhero data groups

	Mean 1	Mean 2	d (control)	d (pooled)
Superman v Spiderman	60.33	41.63	1.53	1.26
Superman v Hulk	60.33	35.38	1.86	1.62
Superman v Ninja	60.33	26.25	4.18	2.60
Spiderman v Hulk	41.63	35.38	0.47	0.49
Spiderman v Ninja	41.63	26.25	1.88	1.48
Hulk v Ninja	35.38	26.25	1.12	0.82

References

- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences (2nd edition)*. New York: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Davies, P., Surridge, J., Hole, L., & Munro-Davies, L. (2007). Superhero-related injuries in paediatrics: a case series. *Archives of Disease in Childhood*, 92(3), 242-243. doi: 10.1136/adc.2006.109793
- Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (4th ed.). London: Sage.