ANOVA in SPSS (Practical)

## Analysis of Variance practical

In this practical we will investigate how we model the influence of a categorical predictor on a continuous response.

In this example, we will test formally for an association between a student's test score on the PISA science test, SCISCORE, and their parents' educational attainment, an indicator often used to capture a family's socio-economic resources. The variable PAREDU classifies highest parental educational qualification into three categories and, as such, this an appropriate predictor or independent variable to use in ANOVA (known as a factor with three levels in ANOVA terminology). The statistical test explores whether the mean science test score is different across the three educational groups. Evidence of achievement gaps of this kind is often used in the discussion and quantification of inter-generational social mobility.

In this case we are interested in the effect of **PAREDU** on our response variable **SCISCORE**. **PAREDU** has 3 categories: Low: GCSE or equiv, Medium: A-level or equiv, High: University degree. We will begin to look at this relationship graphically and look at the distribution of **SCISCORE** separately for each category and a good way to do this is via a box plot. To do this in SPSS:

- Select **Boxplot** from the **Legacy Dialogs** submenu of the **Graphs** menu.
- Select **Simple** and **Summaries for groups of cases** and click on the **Define** button.
- Transfer the **Science test score[SCISCORE]** variable to the **Variable** box.
- Transfer the **Highest qualification of parent[PAREDU]** variable to the **Category Axis** box.
- Click on the **OK** button.

This will produce a table detailing the numbers of valid observations which we do not show here and then a plot with a box for each category as shown below:
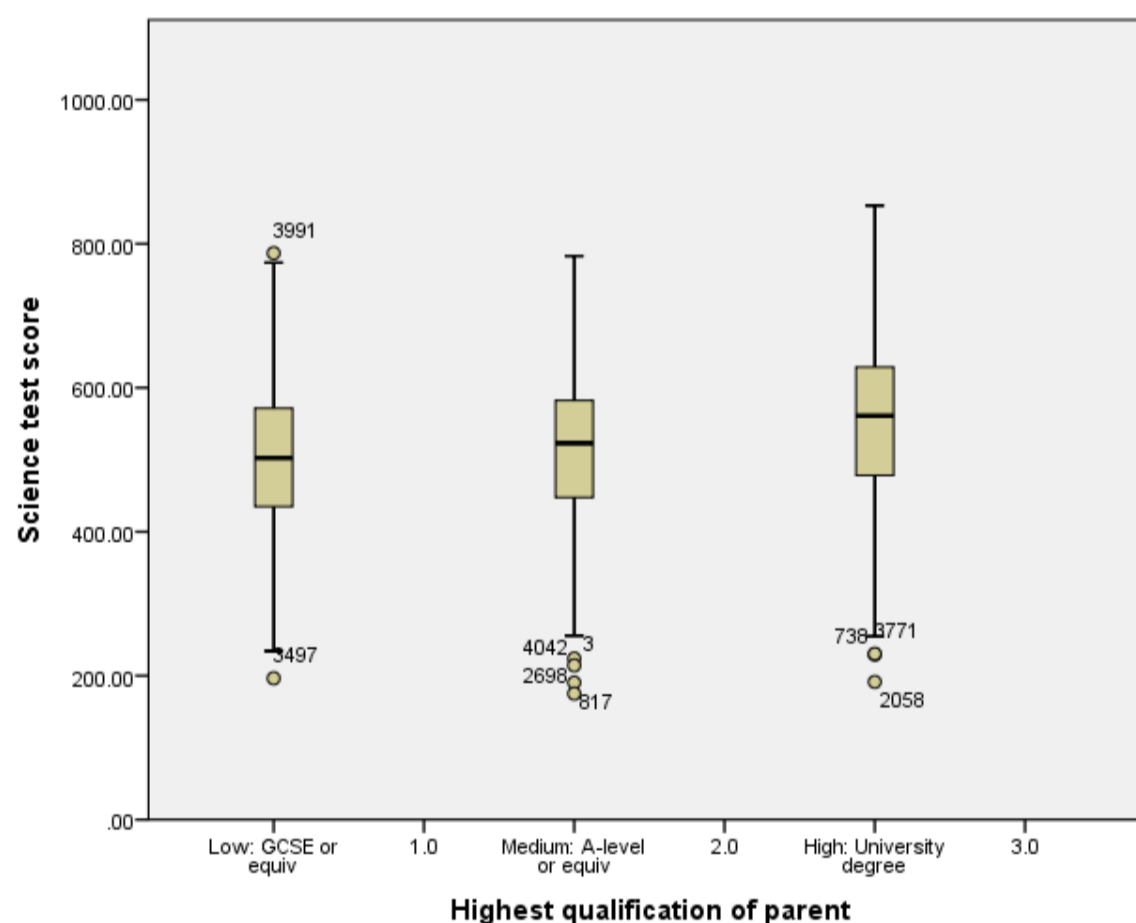

A boxplot uses the 1st and 3rd quantiles of the data to form the box with the median presented as a line in the middle of the box. In this case the medians are for each category as follows: Low: GCSE or equiv = 502.5175, Medium: A-level or equiv = 523.0240 and for High: University degree = 561.2170 and we see that category High: University degree has the highest median whilst category Low: GCSE or equiv has the lowest median.

In fact we will be fitting an analysis of variance (ANOVA) model and this is based on group means rather than medians and so an alternative error bar plot is useful. To get this in SPSS do the following:

- Select **Error Bar** from the **Legacy Dialogs** submenu of the **Graphs** menu.
- Select **Simple** and **Summaries for groups of cases** as for the boxplot and click on the **Define** button.
- Transfer the **Science test score[SCISCORE]** variable to the **Variable** box.
- Transfer the **Highest qualification of parent[PAREDU]** variable to the **Category Axis** box.
- Click on the **OK** button.

The graph will appear as shown below:
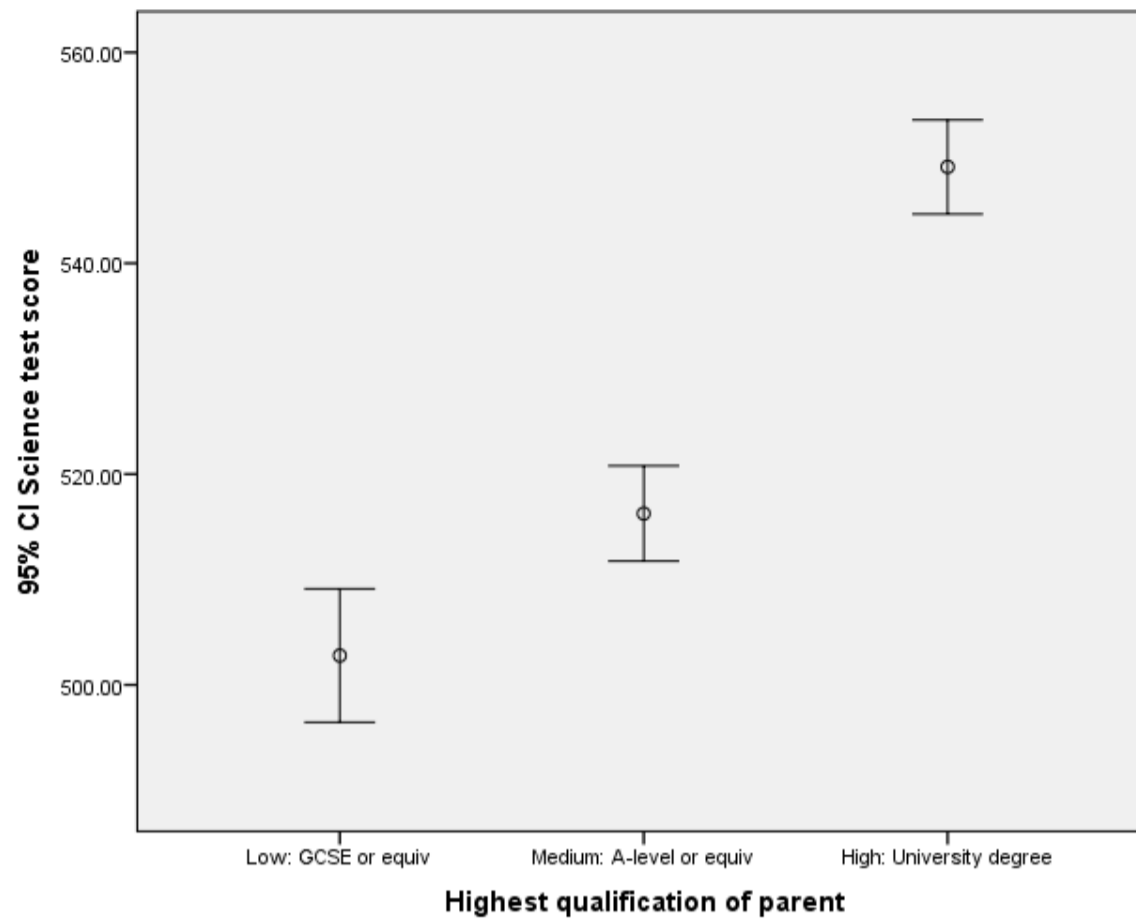
Here we are now plotting the means rather than the medians and the error bars represent 95% confidence intervals for the different groups. The highest mean is 549.1397 for category High: University degree whilst the lowest mean is 502.7856 for category Low: GCSE or equiv. Here we see that all pairs of categories overlap and so we might not expect an effect of **PAREDU** on **SCISCORE**.

We have mentioned summary statistics here (medians and means) and we can access these via the **Explore** option as follows:

- Choose **Explore** from the **Descriptives** submeny within the **Analyse** menu.
- Add **Science test score[SCISCORE]** to the **Dependent list**.
- Add **Highest qualification of parent[PAREDU]** to the **Factor list.**
- Click on the **OK** button.

The summary statistics will appear as below.

Descriptives

Highest qualification of parent				Statistic	Std. Error
Science test score	Low: GCSE or equiv	Mean		502.7856	3.22106
		95% Confidence Interval for Mean	Lower Bound	496.4639	
			Upper Bound	509.1072	
		5% Trimmed Mean		502.0548	
		Median		502.5175	
		Variance		9379.215	
		Std. Deviation		96.84635	
		Minimum		196.38	
		Maximum		787.00	
		Range		590.62	
		Interquartile Range		136.69	
		Skewness		.055	.081
		Kurtosis		-.165	.162
	Medium: A-level or equiv	Mean		516.2518	2.30579
		95% Confidence Interval for Mean	Lower Bound	511.7294	
			Upper Bound	520.7743	
		5% Trimmed Mean		516.7643	
		Median		523.0240	
		Variance		9043.655	
		Std. Deviation		95.09813	
		Minimum		175.22	
		Maximum		782.99	
		Range		607.77	
		Interquartile Range		134.72	
		Skewness		-.133	.059
		Kurtosis		-.289	.119
	High: University degree	Mean		549.1397	2.27646
		95% Confidence Interval for Mean	Lower Bound	544.6754	
			Upper Bound	553.6040	
		5% Trimmed Mean		551.2405	
		Median		561.2170	
		Variance		11162.656	
		Std. Deviation		105.65347	
		Minimum		191.46	
		Maximum		853.13	
		Range		661.67	
		Interquartile Range		150.39	
		Skewness		-.342	.053
		Kurtosis		-.306	.105

If we look at category Low: GCSE or equiv we can see the median = 502.5175 which is the middle line in the earlier boxplot and the IQR = 136.69 which is the width of the box in the plot. For the error bars plot for category Low: GCSE or equiv we see that the mean = 502.7856 which is the x in the middle of the bars whilst the values (496.4639, 509.1072) quoted under lower and upper bound for the 95% Confidence interval for the mean represent the ends of the bars in the plot. We can find similar numbers here for the other categories.

We will next go on to perform an ANOVA to test for significant effects of the predictor. This is done as follows:

- Choose **Univariate** from the **General Linear Model** -> **Analyse** menu.
- Choose **Science test score[SCISCORE]** as the **Dependent Variable**.
- Choose **Highest qualification of parent[PAREDU]** as a **Fixed Factor**.
- Click **Options** and in the window tick **Descriptive statistics** and **Homogeneity tests**.
- Click on **Continue** to return to the main window.
- Click on **Plots...** and transfer **Highest qualification of parent[PAREDU]** to the **Horizontal Axis** box.
- Click **Add** and then **Continue**.
- Click on **Post hoc...**.
- Transfer **Highest qualification of parent[PAREDU]** to the **post hoc test for** box and select **LSD**, **Bonferroni** and **Tukey**.

- Click on the **Continue** button.
- Click on the **OK** button.

This will produce lots of outputs so we will talk about these in turn. First Levene's test:

Levene's Test of Equality of Error Variances

Science test score

F	df1	df2	Sig.
12.142	2	4756	.000

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

The Levene's test is used to test one of the underlying assumptions of the ANOVA which is the homogeneity of variances i.e. that the residual variances are equal in each group. This test requires a test statistic that has value here 12.142 and under the hypothesis of equal variances this statistic follows an F distribution with 2 and 4756 degrees of freedom where 2 is the number of categories - 1 and 4756 is the number of observations - the number of categories. Here we see the p value is .000 which is less than 0.05 and therefore significant which means we can reject the null hypothesis and so the assumption of equal variances is not true and we should adjust the variable in some way (maybe via a transformation) if we wish to use an ANOVA. The ANOVA itself is described by the ANOVA table given below:

Tests of Between-Subjects Effects

Science test score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1770497.879[a]	2	885248.939	87.939	.000
Intercept	1139386327.066	1	1139386327.066	113184.602	.000
PAREDU	1770497.879	2	885248.939	87.939	.000
Error	47876842.524	4756	10066.620		
Total	1379293783.089	4759			
Corrected Total	49647340.403	4758			

a. R Squared = .036 (Adjusted R Squared = .035)

The above table gives all the information required for us to decide if **PAREDU** is a significant predictor of **SCISCORE**. We will here go through the table column by column to explain how the ANOVA works. SPSS gives rather a lot of rows in its ANOVA tables largely because it also allows one to test the intercept term which we are less interested in here. So we will begin with the Type III Sum of Squares (SS) column and for the row Corrected total we see the value 49647340.403. This value is calculated by for each observation taking the value for **SCISCORE** and subtracting the overall mean of **SCISCORE**. These values are then squared and their sum is the value 49647340.403 we see in the table. (Note the value 1379293783.089 in the Total row is calculated similarly but without subtracting the mean from each observation). Next the value in row **PAREDU** is calculated by working out the mean of **SCISCORE** for each category in **PAREDU**. We take these category means and subtract the overall mean from them and again square and sum them to give the value 1770497.879. The Corrected Model row you will see in the one way ANOVA simply repeats the **PAREDU** row. Finally the Error row sum of squares is calculated by subtracting the **PAREDU** SS value from the Corrected total SS value i.e. 47876842.524 = 49647340.403 - 1770497.879, which is effectively the sums of squares not explained by **PAREDU**. For **PAREDU** to be a significant predictor of **SCISCORE** we hope its SS value is large relative to the Error SS but these numbers are based on different sample sizes and so we first need to adjust them to reflect this so the next column is the degrees of freedom (df) column. Here we see we have 4759 total degrees of freedom which represents the number of observations but 4758 corrected total df as we lose one by estimating the mean. For **PAREDU** we have 2 df which is the number of categories - 1 again losing one as if we knew the mean and all bar one of the category means we could calculated the last one. Finally the Error df is 4756 which again is calculated by subtraction i.e. 4756 = 4758 - 2. We next use the df to adjust the SS into Mean Squares (MS) and so MS for **PAREDU** is SS for **PAREDU** divided by df for **PAREDU** which means 885248.939 = 1770497.879 / 2. Similarly for the Error column we have an MS of 10066.620. These two mean squares are now on the same scale and so we can look at their relative sizes by taking their ratio so F = 87.939 = 885248.939 / 10066.620. This test statistic follows an F distribution with 2 and 4756 degrees of freedom and equates to a p value of .000 given in the Sig. column. This p value is less than 0.05 and so we can reject the null hypothesis and we find that **PAREDU** is a significant predictor of **SCISCORE**.

Our next step after establishing whether or not there are significant differences in **SCISCORE** for the different categories of **PAREDU** is to do a post-hoc multiple comparisons test. We selected three different tests and we will look at these in turn. The first we will look at is Tukey's HSD (honest significant difference) shown below:

Multiple Comparisons

Science test score

Tukey HSD

(I) Highest qualification of parent	(J) Highest qualification of parent	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Low: GCSE or equiv	Medium: A-level or equiv	-13.4663*	4.12961	.003	-23.1479	-3.7847
	High: University degree	-46.3541*	3.97607	.000	-55.6757	-37.0325
Medium: A-level or equiv	Low: GCSE or equiv	13.4663*	4.12961	.003	3.7847	23.1479
	High: University degree	-32.8878*	3.25446	.000	-40.5177	-25.2579
High: University degree	Low: GCSE or equiv	46.3541*	3.97607	.000	37.0325	55.6757
	Medium: A-level or equiv	32.8878*	3.25446	.000	25.2579	40.5177

Based on observed means.
The error term is Mean Square(Error) = 10066.620.

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

Here we see a table with one row for each pair of categories (in fact there are two for each pairing as the reverse ordering is also included) and the first two columns identify which pair of categories are being compared so for example the first pairing is Low: GCSE or equiv and Medium: A-level or equiv. The next column gives the difference in means between this pair of categories and so we for example that the mean of Low: GCSE or equiv minus the mean of Medium: A-level or equiv has value -13.4663 and in the following column the associated standard error 4.12961 which will be based on the number of observations in each group and the overall variability. These two columns will be the same regardless of which posthoc test you choose whilst the last three columns differ by test. The Sig. column gives a p value for a difference in mean between the two categories and the final 2 columns give an associated confidence interval. For example the difference between Low: GCSE or equiv and Medium: A-level or equiv has a p value .003. which is less than 0.05 and therefore a significant difference and the confidence interval is (-23.1479, -3.7847) which doesn't contains 0 which agrees with this. The Tukey HSD procedure balances good power with a reasonably tight control of the Type I error rate. We see that all of the pairs of categories are significantly different. We can compare these pairings with our initial error bars.

Next we will look at the LSD procedure with the results shown below:

Multiple Comparisons

Science test score

LSD

(I) Highest qualification of parent	(J) Highest qualification of parent	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Low: GCSE or equiv	Medium: A-level or equiv	-13.4663*	4.12961	.001	-21.5622	-5.3703
	High: University degree	-46.3541*	3.97607	.000	-54.1490	-38.5592
Medium: A-level or equiv	Low: GCSE or equiv	13.4663*	4.12961	.001	5.3703	21.5622
	High: University degree	-32.8878*	3.25446	.000	-39.2681	-26.5076
High: University degree	Low: GCSE or equiv	46.3541*	3.97607	.000	38.5592	54.1490
	Medium: A-level or equiv	32.8878*	3.25446	.000	26.5076	39.2681

Based on observed means.
The error term is Mean Square(Error) = 10066.620.

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

The LSD procedure is less conservative and we should therefore get at least as many pairs that are statistically significantly different. We see that all of the pairs of categories are significantly different.

Finally we look at the Bonferroni procedure with the results as shown below:

Multiple Comparisons

Science test score

Bonferroni

| (I) Highest qualification of parent	(J) Highest qualification of parent	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Low: GCSE or equiv	Medium: A-level or equiv	-13.4663*	4.12961	.003	-23.3560	-3.5766
	High: University degree	-46.3541*	3.97607	.000	-55.8761	-36.8321
Medium: A-level or equiv	Low: GCSE or equiv	13.4663*	4.12961	.003	3.5766	23.3560
	High: University degree	-32.8878*	3.25446	.000	-40.6817	-25.0940
High: University degree	Low: GCSE or equiv	46.3541*	3.97607	.000	36.8321	55.8761
	Medium: A-level or equiv	32.8878*	3.25446	.000	25.0940	40.6817

Based on observed means.
The error term is Mean Square(Error) = 10066.620.

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

The Bonferroni procedure is very conservative and of the three selected we will generally get the least pairs that are statistically significantly different. We see that all of the pairs of categories are significantly different.

The options we have chosen also give us a marginal means plot as shown below:
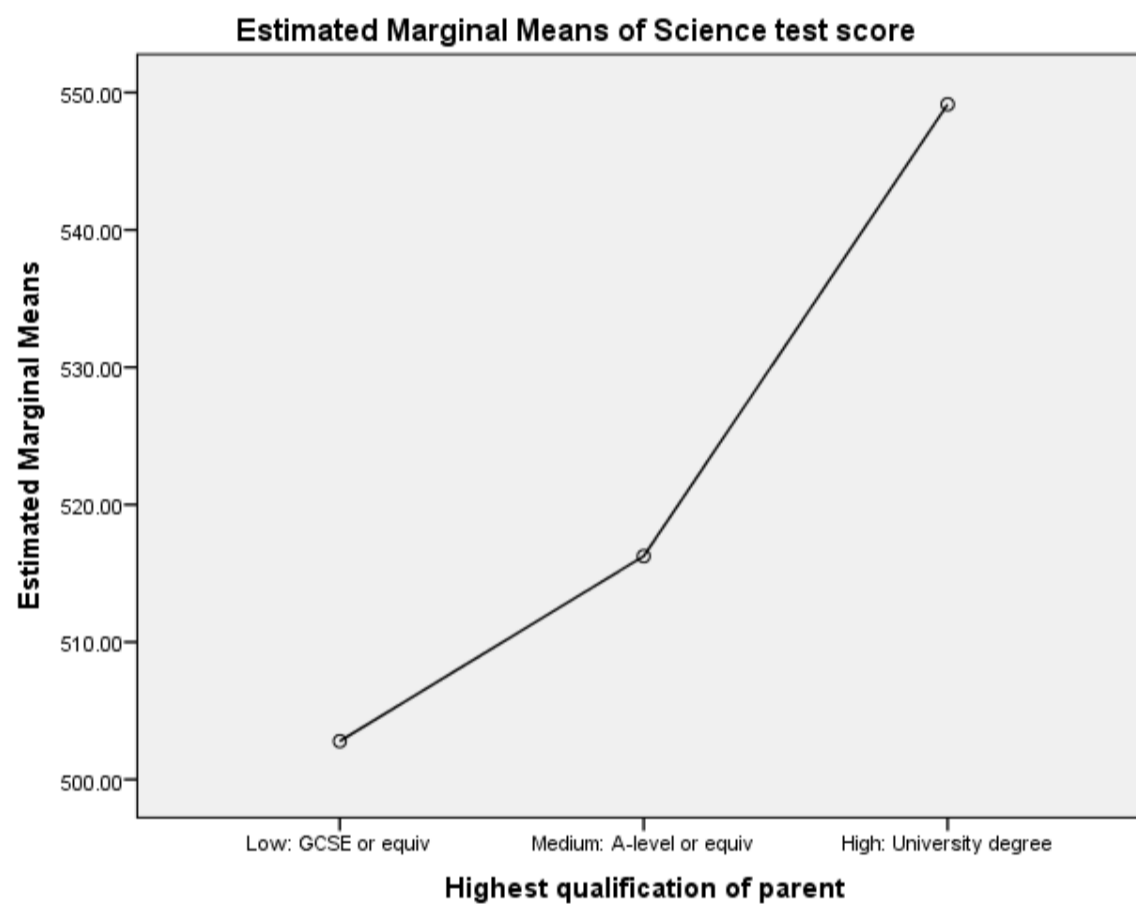
In general a marginal means plot will show the means for each category adjusting for any other variables in the model but here we are simply fitting a one way ANOVA and so we will simply see the means for each group plotted together so for example we see the mean for category Low: GCSE or equiv is 502.7856 and for category Medium: A-level or equiv is 516.2518.

Finally we have asked SPSS to find homogeneous subsets within our categories. These are groups of categories that cannot be separated (using in this case Tukey's HSD) as statistically different within the category.

Science test score

Tukey HSD[a,b,c]

Highest qualification of parent	N	Subset 1	Subset 2	Subset 3
Low: GCSE or equiv	904	502.7856		
Medium: A-level or equiv	1701		516.2518	
High: University degree	2154			549.1397
Sig.		1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

Based on observed means.

The error term is Mean Square(Error) = 10066.620.

a. Uses Harmonic Mean Sample Size = 1389.959.

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

c. Alpha = .05.

In this case we see there are 3 identified subsets with categories Low: GCSE or equiv in subset 1; Medium: A-level or equiv in subset 2; High: University degree in subset 3;

We have seen that there are large and statistically significant differences in mean science test scores between parental education groups. A sense of the magnitudes involved is given by recalling PISA's guidance that 30 test score points equates to the increment associated with one year of schooling. Post-hoc tests showed a statistically significant difference between the low and medium education groups, and an even larger one between the medium and high education groups.