USING SPSS: ONE-WAY ANALYSIS OF COVARIANCE (ANCOVA)

1. ENTERING THE DATA:

For this example, we are using the HSB500 (High School and Beyond) data file. Variable names are shown in **bold**, and as such, should be substituted accordingly to model other examples.

To do a one-way analysis of covariance (ANCOVA) you have to enter the data in a specific manner. The first column "**subno**" is the subject number assigned to the student. The second column, "**faed**" is the four levels of the students' father's education, which is our independent variable (1 = Less than High School, 2 = High School Graduate, 3 = Some Vocational or College, and 4 = College Graduate or More). The third column contains the students' standardized mathematics achievement scores (**mathach**, range -8.333 to 25.000) – the dependent variable. The fourth column contains the students' **mosaic** score (range -28.000 to 56.000), which we will use in this example as the covariate.

The summary data (which is generated through the Case Summaries command in SPSS) for the four groups on both variables, **MATHACH** and **MOSAIC** are shown below:

Case Summaries

Father's Education Level		Mathematics Test Score	Mosaic Test Score
Less than High School	N	132	132
	Mean	10.86873	25.29545
	Std. Deviation	6.186632	9.946988
High School Graduate	N	129	129
	Mean	12.44186	27.18605
	Std. Deviation	6.448802	9.068332
Some Post-Secondary Education	N	105	105
	Mean	13.21907	27.00952
	Std. Deviation	6.283149	10.423019
College Graduate and Beyond	N	134	134
	Mean	15.83089	28.07836
	Std. Deviation	6.494654	7.621412
Total	N	500	500
	Mean	13.09803	26.88900
	Std. Deviation	6.604590	9.288203

Since we have a large sample size ($N = 500$), it is suggested that our *a priori* alpha level be set at .01.

Note: SPSS will produce a fair amount of output as part of the ANCOVA statistical process (request), of which only the applicable tables will be shown and discussed as part of this handout.

Below are the unadjusted means (which are part of the SPSS output generated from the ANCOVA request) for **MATHACH** for the four groups:

Descriptive Statistics

Dependent Variable: Mathematics Test Score

Father's Education Level	Mean	Std. Deviation	N
Less than High School	10.86873	6.186632	132
High School Graduate	12.44186	6.448802	129
Some Post-Secondary Education	13.21907	6.283149	105
College Graduate and Beyond	15.83089	6.494654	134
Total	13.09803	6.604590	500

2. CONDUCTING THE ONE-WAY ANCOVA

Before we get started – we must first test the assumptions underlying the ANCOVA. In addition to testing the assumption of independence and the assumption of normality we need to conduct a test of the homogeneity-of-regression (slopes) assumption. To conduct this test, follow these steps:

1. Click **Analyze**, click **General Linear Model**, then click **Univariate.**

2. Click the dependent variable (**mathach**), then click ► to move it to the Dependent Variable box.

3. Click the independent variable (**faed**), then click ► to move it to the Fixed Factor(s) box.

4. Click the covariate (**mosaic**), then click ► to move it to the Covariate(s) box.

5. Click **Model.**

6. Click **Custom** under Specify Model.

7. Click **faed(F)** – the independent variable under Factors & Covariates and click ► to make it appear in the Model box.

8. Click **mosaic(C)** – the covariate under Factors & Covariates and click ► to make it appear in the Model box.

9. Holding down the Ctrl key, click **faed(F)** and **mosaic(C)** in the Factors & Covariates box. Check to see that the default option Interaction is specified in the drop-down menu in the Build Term(s) box. If it is not, select it.

10. Click ► and **faed*mosaic** should now appear in the Model box.

11. Click **Continue.** This will bring you back to the **Univariate** screen.

12. Click **OK**.

The following type of table will appear in the output screen…

Tests of Between-Subjects Effects

Dependent Variable: mathach Mathematics Test Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3310.890[a]	7	472.984	12.609	.000
Intercept	3371.024	1	3371.024	89.866	.000
faed	290.025	3	96.675	2.577	.053
mosaic	1374.255	1	1374.255	36.635	.000
faed * mosaic	**72.562**	**3**	**24.187**	**.645**	**.587**
Error	18455.792	492	37.512		
Total	107545.825	500			
Corrected Total	21766.682	499			

a. R Squared = .152 (Adjusted R Squared = .140)

Before conducting an ANCOVA – the homogeneity-of-regression (slope) assumption should first be tested. The test evaluates the interaction between the covariate and the factor (independent variable) in the prediction of the dependent variable. A significant interaction between the covariate and the factor suggests that the differences on the dependent variable among groups vary as a function of the covariate.

If the interaction is significant – the results from an ANCOVA are not meaningful – and ANCOVA should not be conducted. In our example: The interaction source is labeled **faed*mosaic**. Our results suggest the interaction is not significant, $F(3, 492) = .645$, $p = .587$. That is, p (.587) > α (.01). Based on this finding, we can proceed with our ANCOVA analysis…

To conduct a one-way ANCOVA, follow these steps:

1. Click **Analyze**, click **General Linear Model**, then click **Univariate.**

2. Click **Reset**.

   If you have not exited SPSS – the prior commands will still be shown. As a precaution for avoiding possible errors – click the reset key and begin the procedure from the initial starting point.

3. Click the dependent variable (**mathach**), then click ► to move it to the Dependent Variable box.

4. Click the independent variable (**faed**), then click ► to move it to the Fixed Factor(s) box.

5. Click the covariate (**mosaic**), then click ► to move it to the Covariate(s) box.

6. Click on **Options.**

7. In the Factor(s) and Factor Interactions box, click the independent variable (in our example, **faed**).

    This will provide the adjusted (Estimated Marginal) means that will be used later (if needed) in post hoc procedures.

8. Click ► to move it to the Display Means for box.

9. Select **Descriptive statistics** in the Display box.

10. Select **Homogeneity tests** in the Display box.

11. Click **Continue.**

12. This will bring you back to the **Univariate** screen…

13. Click **Plots**.

14. Click the independent variable (**faed**), then click ► to move it to the Horizontal Axis: box.

15. Click **Add**.

16. Click **Continue.**

17. This will bring you back to the **Univariate** screen – click **OK.**

The following SPSS output will be produced…

3. THE SPSS OUTPUT:

Univariate Analysis of Variance

Descriptive Statistics

Dependent Variable: Mathematics Test Score

Father's Education Level	Mean	Std. Deviation	N
Less than High School	10.86873	6.186632	132
High School Graduate	12.44186	6.448802	129
Some Post-Secondary Education	13.21907	6.283149	105
College Graduate and Beyond	15.83089	6.494654	134
Total	13.09803	6.604590	500

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

Dependent Variable: mathach Mathematics Test Score

F	df1	df2	Sig.
1.102	3	496	**.348**

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

a. Design: Intercept+mosaic+faed

From the above output – we see that the underlying assumption of homogeneity of variance for the one-way ANCOVA has been met – as evidenced by $F(3, 496)$ = 1.102, p = .348. That is, p (.348) > α (.01).

The covariate is included in the analysis to control for the differences on the independent variable. While it is not the central focus of the analysis, the results are available as part of the output. The primary purpose of the test of the covariate is that it evaluates the relationship between the covariate and the dependent variable, controlling for the factor (i.e., for any particular group). In the current example, this relationship is significant, $F(1, 495)$ = 40.727, $p < .001$. Additionally, what this shows us is that there is a relationship (effect) between the covariate and the dependent variable. Had this not been significant, the question then would be on the appropriateness of the selection of the covariate. Recall that the covariate must be linearly related to the dependent variable.

Tests of Between-Subjects Effects

Dependent Variable: mathach Mathematics Test Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3238.328[a]	4	809.582	21.629	.000
Intercept	3360.839	1	3360.839	89.788	.000
mosaic	1524.455	1	1524.455	40.727	.000
faed	**1385.442**	**3**	**461.814**	**12.338**	**.000**
Error	18528.354	495	**37.431**		
Total	107545.825	500			
Corrected Total	21766.682	499			

a. R Squared = .149 (Adjusted R Squared = .142)

The results shown above are as follows: The group source (labeled **faed** on the SPSS output) evaluates the null hypothesis that the population adjusted means are equal. The results of the analysis indicate that this hypothesis should be rejected, $F(3, 495)$ = 12.338, $p < .001$. The test assesses the differences among the adjusted means for the four groups, which are reported in the Estimated Marginal Means box as 11.170 (Less than High School), 12.386 (High School Graduate), 13.196 (Some Post-Secondary Education), and 15.606 (College Graduate and Beyond). Typically, the results for the covariate are not reported in a Results section.

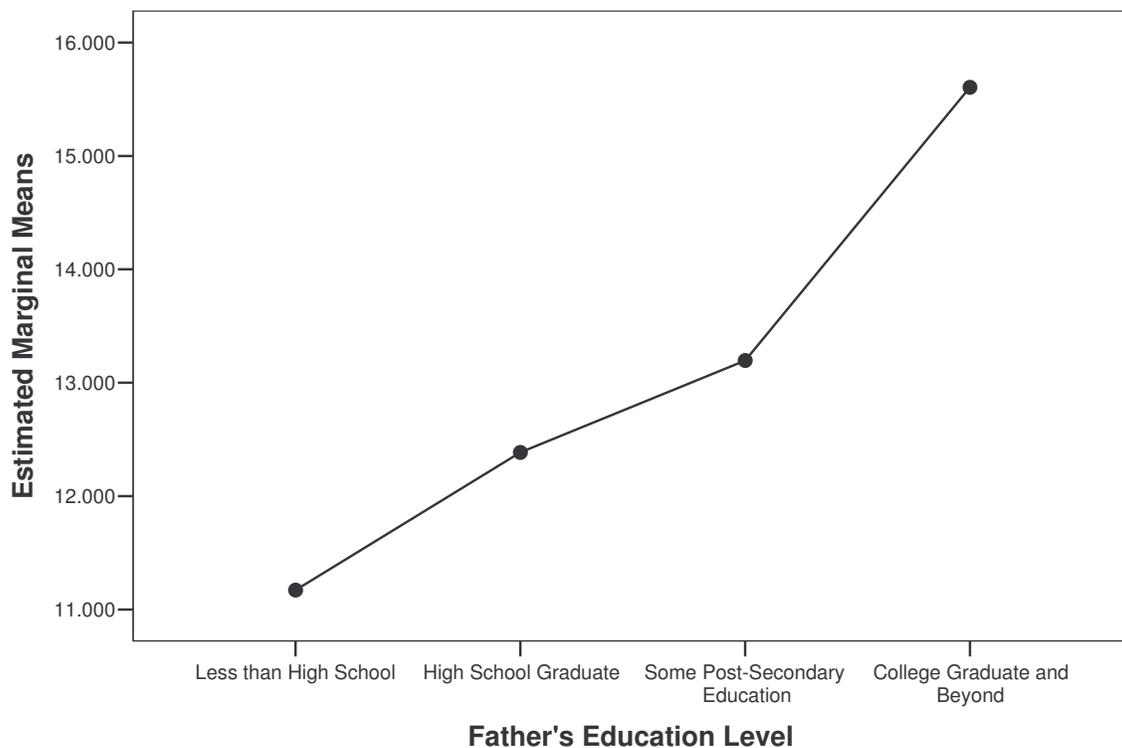# Estimated Marginal Means

Father's Education Level

Dependent Variable: Mathematics Test Score

Father's Education Level	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Less than High School	11.170[a]	.535	10.120	12.221
High School Graduate	12.386[a]	.539	11.327	13.444
Some Post-Secondary Education	13.196[a]	.597	12.023	14.369
College Graduate and Beyond	15.606[a]	.530	14.565	16.646

a. Covariates appearing in the model are evaluated at the following values: Mosaic Test Score = 26.88900.

Profile Plots

Estimated Marginal Means of Mathematics Test Score

The Profile Plot will give us a visual picture of what is going on with our study. As we can see the line represents the estimated marginal means for the mathematics test score at each of the levels of the father's education. These values correspond to those found in the estimated marginal means table shown earlier.

4.   CALCULATING THE MEASURE OF ASSOCIATION (ω^2)

Calculating the measure of association (omega squared) for the ANCOVA is very similar to that for the One-Way ANOVA. We only need to make a few minor adjustments to the formula – to account for the adjusted values of interest…

$$\omega^2 = \frac{SS_B^{'} - (K-1)MS_W^{'}}{SS_T^{'} + MS_W^{'}}$$

For our example – we substitute into the formula and get:

$$\omega^2 = \frac{1385.442 - (4-1)37.431}{21766.682 + 37.431} = \frac{1385.442 - 3(37.431)}{21804.113} = \frac{1385.44 - 112.293}{21804.113} = \frac{1273.149}{21804.113}$$

$$= 0.058390314 = \textbf{.06 } (\approx \textbf{6\%})$$

This means that the four levels of fathers education (the independent variable) account for approximate 6% of the total variance in the students mathematics achievement scores (the dependent variable), controlling for the effect of their mosaic test score (the covariate).

At this point in the analysis, we know that the main effect for the independent variable (**faed**) is significant controlling for the effect of the covariate (**mosaic**). Using APA guidelines, we would report this statistical information as follows:

$F(3, 495) = 12.34$, $p < .001$, $\omega^2 = .06$

   Or $F(3, 495) = 12.34$, $p < .01$, $\omega^2 = .06$

5.   CONDUCTING PAIRWISE COMPARISONS FOLLOWING A SIGNIFICANT MAIN EFFECT

In order to conduct pariwise comparisons – <u>you can conduct them by hand</u> – using the Tukey HSD (equal ns) or Tukey/Kramer (unequal ns) to calculate the Q statistic. However, you can conduct pairwise comparisons by performing the /lmatrix command (maintaining the appropriate error term) to evaluate whether there are significant differences in the adjusted means between the groups.

<u>**IF** the Main Effect is significant</u> – you will need to conduct post hoc follow-up. To conduct the pairwise comparisons among the four groups, follow these steps:

1.   Click **Analyze**, click **General Linear Model**, then click **Univariate.**

2.   If you have not exited SPSS, the appropriate options should already be chosen, if not, conduct Steps 2 through 5 from the preceding illustration of conducting the one-way ANCOVA.

3.   Click **Paste** to exit the General Linear Model dialog box and go to the Syntax1: SPSS Syntax Editor window.

4. Remove all but the first three lines:

    UNIANOVA
      mathach BY faed WITH mosaic
      /METHOD = SSTYPE(3)

5. Type six separate /lmatrix commands to compare the adjusted means for Group 1 and Group 2, Group 1 and Group 3, Group 1 and Group 4, Group 2 and Group 3, Group 2 and Group 4, and Group 3 and Group 4. The completed syntax is shown below. Don't forget the period (.) at the end of the syntax.

    UNIANOVA
      mathach BY faed WITH mosaic
      /METHOD = SSTYPE(3)
    /LMATRIX 'faed 1 VS faed 2'
    faed 1 -1 0 0
    /LMATRIX 'faed 1 VS faed 3'
    faed 1 0 -1 0
    /LMATRIX 'faed 1 VS faed 4'
    faed 1 0 0 -1
    /LMATRIX 'faed 2 VS faed 3'
    faed 0 1 -1 0
    /LMATRIX 'faed 2 VS faed 4'
    faed 0 1 0 -1
    /LMATRIX 'faed 3 VS faed 4'
    faed 0 0 1 -1.

6. Highlight the entire syntax, click **Run**, and then click **Selection.**

6. EXPLAINING THE /LMATRIX COMMANDS FOR PAIRWISE COMPARISONS

The results of the first pariwise comparison are shown below: This comparison evaluates the difference in adjusted means between the first and second groups. For each pariwise comparison (labeled Custom Hypothesis Tests), the Contrast Estimate in the Contrast Results box reflects the difference between the two adjusted means for that contrast. The Test Results box reports the *F* test, degrees of freedom, and the *p* value (Sig.) for a comparison.

Custom Hypothesis Tests #1

Contrast Results (K Matrix)[a]

Contrast		Dependent Variable Mathematics Test Score
L1	Contrast Estimate	-1.215
	Hypothesized Value	0
	Difference (Estimate - Hypothesized)	-1.215
	Std. Error	.760
	Sig.	.110
	95% Confidence Interval for Difference Lower Bound	-2.707
	Upper Bound	.277

a. Based on the user-specified contrast coefficients (L') matrix: faed 1 VS faed 2

Test Results

Dependent Variable: MATHACH Mathematics Test Score

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	95.804	1	95.804	2.559	.110
Error	18528.354	495	37.431		

The difference in adjusted means (labeled Contrast Estimates) is -1.215 (which varies from 11.170 – 12.386, do to rounding). The Sig. (*p*) value is .110, however, it cannot be compared to the *a priori* alpha level (for this example, .01) without making an adjustment to protect against inflating the Type I error rate.

While there are different options available – one suggestion is using the Bonferroni adjustment. For this example, since we had six unique pairwise comparisons – the adjusted alpha (α') would be .0017 (.01/6). Comparing our obtained *p* value of .110 to the adjusted alpha (.0017) we would conclude that there was not a significant difference between Groups 1 and 2. The difference that does exist can be attributed to sample fluctuation.

Let's take a look at one more comparison – comparing Groups 1 and 4:

The difference in adjusted means is -4.435. The Sig. (*p*) value is .000, however, remember it cannot be compared to the *a priori* alpha level (for this example, .01) without making an adjustment to protect against inflating the Type I error rate.

Again using the Bonferroni adjustment – the adjusted alpha (α') would be .0017 (.01/6). Comparing our obtained *p* value of .000 to the adjusted alpha (.0017) we would conclude that there was a significant difference between Groups 1 and 4. The probability that the observed difference between the adjusted means would have occurred by chance, if the null hypothesis is true, is less than .0083. Note that in the printout, "Sig." shows .000, which implies that the probability of such an occurrence is actually less than .001.

Custom Hypothesis Tests #3

Contrast Results (K Matrix)[a]

		Dependent Variable
Contrast		Mathematics Test Score
L1	Contrast Estimate	-4.435
	Hypothesized Value	0
	Difference (Estimate - Hypothesized)	-4.435
	Std. Error	.755
	Sig.	.000
	95% Confidence Interval for Difference Lower Bound	-5.918
	Upper Bound	-2.952

a. Based on the user-specified contrast coefficients (L') matrix: faed 1 VS faed 4

Test Results

Dependent Variable: MATHACH Mathematics Test Score

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	1292.374	1	1292.374	34.527	.000
Error	18528.354	495	37.431		

Since we found a significant pairwise difference – we will need to follow this up with determining the Effect Size, measuring the magnitude of this effect.

Using the formula: $ES = \dfrac{\overline{X}'_i - \overline{X}'_k}{\sqrt{MS'_w}}$ *where adjusted values are used…*

Note: the MS'_W comes from the ANCOVA summary table (and SPSS provides it in the "Test Results" table) … and the means will come from the absolute values of the estimated marginal means table…

For this example: $ES = \dfrac{4.435}{\sqrt{37.431}} = \dfrac{4.435}{6.1180879} = .7248997 \quad = \quad \textbf{.72}$

7.   INTERPRETATION (RESULTS):

A one-way analysis of covariance (ANCOVA) was conducted for this study. The independent variable, father's educational level, included four levels: less than high school, high school graduate, some vocational or college, and college graduate or more. The dependent variable was the students' mathematics achievement scores and the covariate was the students' score on the mosaic test. A preliminary analysis evaluating the homogeneity-of-regression (slopes) assumption indicated that the relationship between the covariate and the dependent variable did not differ significantly as a function of the independent variable, $F(3, 492) = .65$, $p = .587$. The ANCOVA was significant, $F(3, 495) = 12.34$, $p < .001$ (See Table 1). However, only 6% ($\omega^2 = .06$) of the total variance in mathematics achievement scores was accounted for by the four levels of fathers' education controlling for the effect of the students' mosaic test scores.

Table 1

Analysis of Co-Variance for Mathematics Achievement by Father's Education Level

Source	SS	df	MS	F	p
Mosaic Test	1,524.46	1	1,524.46	40.73	.000
Father's Education	1,385.44	3	461.81	12.34	.000
Error	18,528.35	495	37.43		
Total	21,766.68	499			

Follow-up tests were conducted to evaluate pairwise differences among the adjusted means for father's educational level. The Bonferroni procedure was used to control for Type I error across the six pairwise comparisons ($\alpha' = .01/6 = .0017$). The results showed that students' whose fathers were college graduates or more ($M = 15.61$) had significantly higher mathematics achievement scores, controlling for the effect of their mosaic test, than students whose fathers had some vocational or college education ($M = 13.20$), were high school graduates ($M = 12.39$), and had less than a high school education ($M = 11.17$). The effects sizes for these significant adjusted mean differences were .72, .53, and .39, respectively. (See Table 2)

Table 2

Pairwise Comparisons and Effect Sizes of Mathematics Achievement by Fathers' Education

Group	Mean	Adjusted Mean	Adjusted Mean Differences ($\overline{X}'_i - \overline{X}'_k$) (Effect Sizes are indicated in parentheses)			
			1.	2.	3.	4.
1. Less than HS	10.87	11.17	——			
2. HS Graduate	12.44	12.39	1.22	——		
3. Some Vocational or College Education	13.22	13.20	2.03	.81	——	
4. College Graduate	15.83	15.61	4.44* (.72)	3.22* (.53)	2.41* (.39)	——

* $p < .0017$

Note: The Bonferroni adjustment is considered to be a rather conservative adjustment – and as such should be used with that in mind. An alternative (somewhat less conservative method) is to use the Holm's Sequential Bonferroni adjustment method to control for Type I error or to hand calculate the pairwise comparisons using the Tukey or Tukey-Kramer method.