Notes 9d: ANCOVA

1. What is ANCOVA?

ANCOVA is a statistical procedure that enables one to compare groups on some quantitative dependent variable while simultaneously controlling for quantitative independent variables. Thus, ANCOVA combines both qualitative and quantitative independent variables.

ANCOVA is used because inclusion of the covariate in the model can (a) increase power to detect group differences and (b) precision of estimates. Both (a) and (b) are possible if the covariate is associated (correlated) with the DV and not correlated with the factor (the qualitative IV) of interest. A result of this inclusion of the covariate is the reduction of the model mean squared error (MSE) and thus the increase of corresponding F-ratios and reduction of standard errors (such as with pairwise comparisons among group means). Additionally, inclusion of the covariate (c) provides estimates of group means on the DV that statistically control or adjust, for differences on the covariate. For example, suppose one conducts a true experimental study in which students are randomly assigned to treatment and control condition, but the groups exhibit, through random chance, mean differences in IQ. This is illustrated in Table 1 below.

Table 1
Hypothetical ANCOVA Adjustment

Groups	Mean of IQ (Covariate)	Observed Posttest Achievement (DV Mean)	Adjusted Posttest Achievement (Estimated DV Mean)
Experimental Condition	106.00	85.00	82.00
Control Condition	98.00	75.00	78.00

Since the experimental group began the study with a higher mean IQ than the control group, ANCOVA can be used to provide a statistical estimate of group differences that account for this initial IQ difference. This estimate is known as an adjusted mean, predicted mean, or estimated marginal mean. The amount of adjustment varies according to the size of the difference between groups on the covariate, and the strength of the correlation between the covariate and the DV, and also the strength of the association between the covariate and the factor. In the above example, the 10 point mean difference in achievement was adjusted to a 4 point mean difference in achievement after taking into account the 8 point difference in IQ between the groups. In effect, researchers use ANCOVA to equalize group differences on covariates in an attempt to "level the playing field" so they can better estimate differences on the DV that removes differences due to the covariate.

It is important to understand limitations on ANCOVA because some estimates provided by ANCOVA may be unrealistic and therefore lead to inappropriate conclusions and inferences. Two sources to help explain issues with ANCOVA are linked below:

> Miller, G.A., & Chapman, J.P. (2001).Misunderstanding Analysis of Covariance. Journal of Abnormal Psychology, 110, 40-48.
>
> http://www.bwgriffin.com/gsu/courses/edur8132/notes/Miller_Chapman_ANCOVA_Abuses.pdf
>
> Engqvist, L. (2005) The mistreatment of covariate interaction terms in linear mode analyses of behavioral and evolutionary ecology studies. Animal Behavior, 70, 967-971.
>
> http://www.bwgriffin.com/gsu/courses/edur8132/notes/Engqvist_ANCOVA_Interaction_Term.pdf

While issues raised about interpretation of statistical estimates in ANCOVA are important, often researchers overlook these issues with other regression analysis models. This is unfortunately because all the issues discussed in the readings above also apply to other statistical analyses as well, but some reason these issues of interpretation only are raised when ANCOVA is used.

2. The Regression Equation

Suppose one is interested in modeling the following fictional achievement data. The covariate, the quantitative predictor, is motivation, and the factor, the qualitative predictor, is type of instruction with three categories. The data appear in Table 1.

Table 2
Fictional Achievement, Instruction, and Motivation Data

Achievement	Instruction Type	Motivation
70	coop	3
74	coop	4
77	coop	5
80	coop	6
83	coop	8
85	coop	8
75	self	2
78	self	2
81	self	4
86	self	6
88	self	6
89	self	7
85	lecture	3
87	lecture	3
90	lecture	5
96	lecture	7
95	lecture	8
98	lecture	8

Note: Motivation is scaled such that 1 = low motivation and 10 = high.

Of interest is whether achievement differs by instructional strategy after level of student motivation is controlled. It is possible that motivation could be very different for each student, so instruction group means will be adjusted to take into account differing levels of motivation within each group. In essence, those groups with lower motivation scores will have achievement means adjusted upward, and those with higher motivation scores will have achievement means adjusted downward. This is how ANCOVA adjusts for group differences on covariates.

The regression model is

$$Y \text{ (Achievement)} = b_0 + b_1\text{coop} + b_2 \text{ lecture} + b_3 \text{ motivation} + e$$

where “coop,” and “lecture” are dummy variables and the omitted category (reference group) is the self treatment.

3. SPSS Results

SPSS results of the ANCOVA, using the GENERAL LINEAR MODEL->UNIVARIATE command, are provided below:

Descriptive Statistics

Dependent Variable: achievement

instruction	Mean	Std. Deviation	N
coop	78.1667	5.63619	6
lecture	91.8333	5.26941	6
self	82.8333	5.70672	6
Total	84.2778	7.82008	18

Comment: The table above presents dependent variable (Achievement) means for each of the three groups involved. These are the raw, unadjusted means on the DV.

Tests of Between-Subjects Effects

Dependent Variable: achievement

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1015.687(a)	3	338.562	198.124	.000
Intercept	11512.556	1	11512.556	6737.051	.000
motivation	436.576	1	436.576	255.481	.000
instruction	562.171	2	281.085	164.489	.000
Error	23.924	14	1.709		
Total	128889.000	18			
Corrected Total	1039.611	17			

a R Squared = .977 (Adjusted R Squared = .972)

Comment: The table above presents ANOVA summary information. The covariate, Motivation, has an F-ratio = 255.48 and for the factor, Instruction, F = 164.49. Both are statistically significant at the .05 and also .01 levels since the corresponding p-values (column labeled "Sig.") are less than .01.

Parameter Estimates

Dependent Variable: achievement

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	71.765	.874	82.088	.000	69.890	73.640
motivation	2.460	.154	15.984	.000	2.130	2.790
[instruction=coop]	-7.536	.776	-9.714	.000	-9.200	-5.872
[instruction=lecture]	6.130	.776	7.902	.000	4.467	7.794
[instruction=self]	0(a)	.	.	.	.	.

a This parameter is set to zero because it is redundant.

Comment: The table above presents regression results for this ANCOVA model. Controlling for motivation, the adjusted mean difference between Cooperative Learning and Self-paced is -7.536, and the adjusted mean difference between Lecture and Self-paced is 6.13.

To obtain predicted values for each group, the following regression model can be used:

Y(Achievement)' = 71.765 + -7.536(coop) + 6.130(lecture) + 2.460(motivation = 5.2778)

To obtain predicted values, it is customary to use the mean of the covariate in the regression equation. The mean of the covariate can be found in SPSS using the DESCRIPTIVES command. Below is the output.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
motivation	18	2.00	8.00	5.2778	2.13667
Valid N (listwise)	18				

Predicted—or ***adjusted means***—for each of the three groups follows using the regression equation:

Adjusted Mean for Self-paced:

Y(Achievement)' =	71.765	-7.536(coop)	+ 6.130(lecture)	+ 2.460(motivation = 5.2778)
Y(Achievement)' =	71.765	-7.536(coop=0)	+ 6.130(lecture=0)	+ 2.460(motivation = 5.2778)
Y(Achievement)' =	71.765	-7.536*(0)	+ 6.130*(0)	+ 2.460*(5.2778)
Y(Achievement)' =	71.765			+ 12.9834
Y(Achievement)' =	**84.748**			

Adjusted Mean for Cooperative Learning:

Y(Achievement)' =	71.765	-7.536(coop)	+ 6.130(lecture)	+ 2.460(motivation = 5.2778)
Y(Achievement)' =	71.765	-7.536(coop=1)	+ 6.130(lecture=0)	+ 2.460(motivation = 5.2778)
Y(Achievement)' =	71.765	-7.536*(1)	+ 6.130*(0)	+ 2.460*(5.2778)
Y(Achievement)' =	71.765	-7.536		+ 12.9834
Y(Achievement)' =	**77.212**			

Adjusted Mean for Lecture:

Y(Achievement)' =	71.765	-7.536(coop)	+ 6.130(lecture)	+ 2.460(motivation = 5.2778)
Y(Achievement)' =	71.765	-7.536(coop=0)	+ 6.130(lecture=0)	+ 2.460(motivation = 5.2778)
Y(Achievement)' =	71.765	-7.536*(0)	+ 6.130*(1)	+ 2.460*(5.2778)
Y(Achievement)' =	71.765		+ 6.130	+ 12.9834
Y(Achievement)' =	**90.878**			

Comment: SPSS provides the adjusted means, or predicted means, in ANCOVA in the table below. SPSS refers to adjusted means as "Estimated Marginal Means." The values reported in the table below should correspond to the predicted means calculated above within round error. Note that SPSS reports the mean value of the covariate used to obtain the adjusted means in the table above; that value is reported as a footnote.

Comment: Note that adjusted means are derived by holding constant the covariate; the adjusted means are means predicted by the regression equation if all participants (or units) had the same covariate value. In this example, the ANCOVA model provides estimates of achievement for the three groups assuming the Motivation level in each group is 5.2778. This is what is meant by holding constant a variable.

Estimated Marginal Means

Estimates

Dependent Variable: achievement

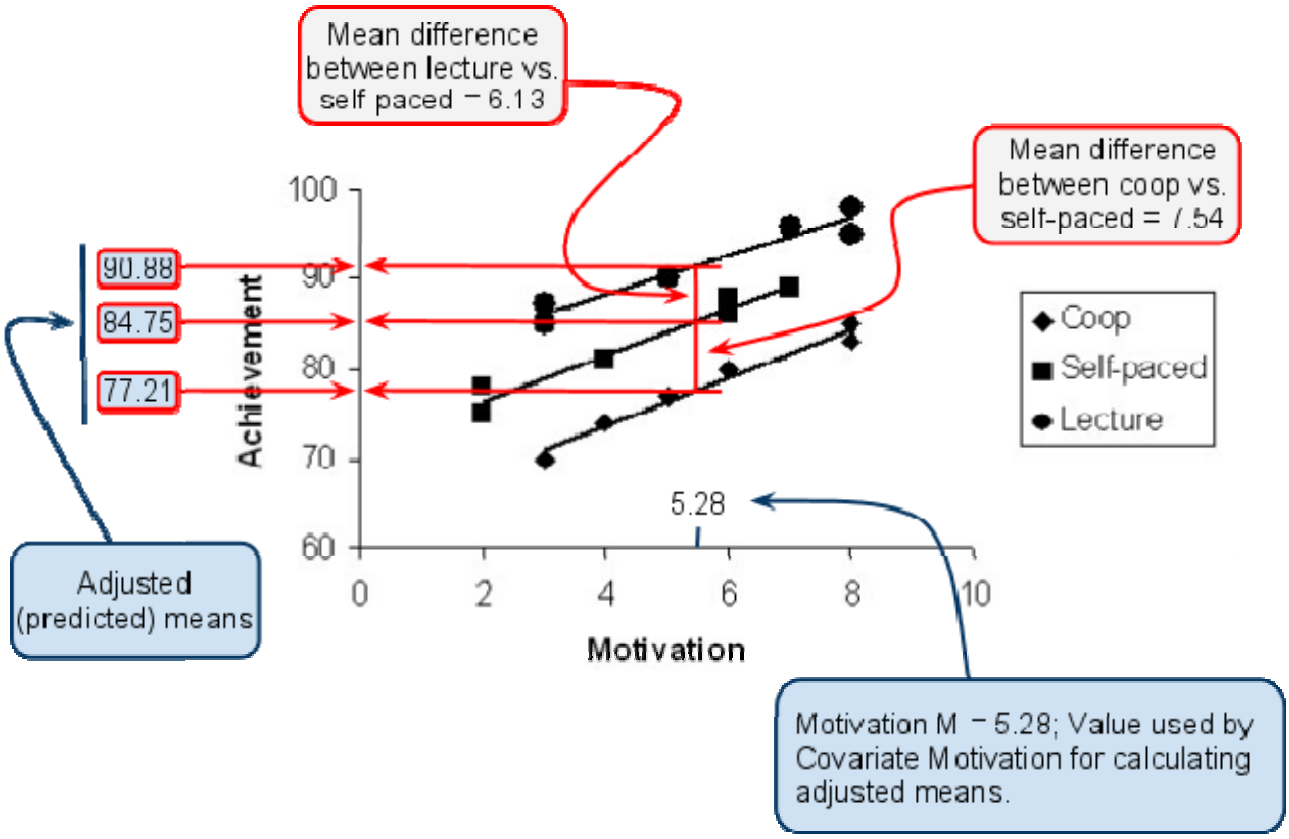
instruction	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
coop	77.210(a)	.537	76.058	78.362
lecture	90.877(a)	.537	89.725	92.029
self	84.746(a)	.547	83.573	85.919

a Covariates appearing in the model are evaluated at the following values: motivation = 5.2778.

Figure 1 below shows graphically how predicted, or adjusted, means are determined for each of the three groups. Note that the predicted means are identical to those calculated above.

Figure 1

Scatter plot of Motivation and Achievement with Regression Slopes and Predicted Means



4. Comparison of ANOVA to ANCOVA

As noted above some benefits of ANCOVA include increased power (e.g., larger F-ratios, smaller CI), increased precision (e.g., smaller standard errors, smaller CI), and statistical control/adjustments. To help illustrate these benefits of adding a covariate, consider results from an ANOVA and ANCOVA for the sample data in Table 2 above.

(a) ANOVA Summary Table

SPSS ANOVA Results:

Tests of Between-Subjects Effects

Dependent Variable: achievement

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	579.111(a)	2	289.556	9.432	.002
Intercept	127849.389	1	127849.389	4164.475	.000
instruction	579.111	2	289.556	9.432	.002
Error	460.500	15	30.700		
Total	128889.000	18			
Corrected Total	1039.611	17			

a R Squared = .557 (Adjusted R Squared = .498)

SPSS ANCOVA Results:

Tests of Between-Subjects Effects

Dependent Variable: achievement

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1015.687(a)	3	338.562	198.124	.000
Intercept	11512.556	1	11512.556	6737.051	.000
motivation	436.576	1	436.576	255.481	.000
instruction	562.171	2	281.085	164.489	.000
Error	23.924	14	1.709		
Total	128889.000	18			
Corrected Total	1039.611	17			

a R Squared = .977 (Adjusted R Squared = .972)

Highlights:

(1) F-ratios

ANOVA Group F = 9.43
ANCOVA Group F = 164.49

Larger F-ratio means more power to detect group differences with ANCOVA.

(2) MSE (error variance)

ANOVA MSE = 30.70
ANCOVA MSE = 1.71

Note the smaller error variance with ANCOVA, so this will lead to larger F-ratios and smaller standard errors.

(b) Pairwise Comparisons

SPSS ANOVA Results:

Pairwise Comparisons

Dependent Variable: achievement

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
coop	lecture	-13.667(*)	3.199	.002	-22.284	-5.050
	self	-4.667	3.199	.496	-13.284	3.950
lecture	coop	13.667(*)	3.199	.002	5.050	22.284
	self	9.000(*)	3.199	.039	.383	17.617
self	coop	4.667	3.199	.496	-3.950	13.284
	lecture	-9.000(*)	3.199	.039	-17.617	-.383

Based on estimated marginal means

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

a Adjustment for multiple comparisons: Bonferroni.

SPSS ANCOVA Results:

Pairwise Comparisons

Dependent Variable: achievement

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
coop	lecture	-13.667(*)	.755	.000	-15.718	-11.616
	self	-7.536(*)	.776	.000	-9.645	-5.428
lecture	coop	13.667(*)	.755	.000	11.616	15.718
	self	6.130(*)	.776	.000	4.022	8.239
self	coop	7.536(*)	.776	.000	5.428	9.645
	lecture	-6.130(*)	.776	.000	-8.239	-4.022

Based on estimated marginal means

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

a Adjustment for multiple comparisons: Bonferroni.

Highlights:

(1) Mean Difference Standard Errors

ANOVA standard error = 3.199

ANCOVA standard error = 0.755 or 0.776

Smaller standard errors with ANCOVA, so smaller CI will result and this leads to greater power to detect differences if differences exist.

(2) Confidence Intervals

ANOVA 95%CI (Bonferroni adjusted) = Lecture vs. Coop 5.05 to 22.28 (width = 17.23)

ANCOVA 95%CI (Bonferroni adjusted) = Lecture vs. Coop 11.62 to 15.72 (width = 4.10)

Note the smaller—less wide—CI for ANCOVA. Smaller CI mean better precision of estimates and grater power to detect differences.

(3) Adjusted Mean Difference Estimates

ANOVA Self-paced vs. Coop Mean Difference = 4.667
ANCOVA Self-paced vs. Coop Mean Difference = 7.536

The difference values for the mean difference reflect the adjusting of the estimated mean difference among groups due to differences in Motivation, the covariate, between Self-paced and Coop Learning groups.

5. Reporting in APA Style

Using the SPSS tables posted above it is possible now to form results presentation that is suitable for publication. One exception, however, is the test for homogeneity of regression slopes. That will be explained below.

Table 3
ANCOVA Results and Descriptive Statistics for Achievement by Instruction Type and Motivation

Type of Instruction	Achievement			
	Observed Mean	Adjusted Mean	SD	n
Cooperative Learning	78.17	77.21	5.64	6
Lecture	91.83	90.88	5.27	6
Self-paced	82.83	84.75	5.71	6

Source	SS	df	MS	F
Motivation	436.58	1	436.58	255.48*
Instruction	562.17	2	281.09	164.49*
Error	23.92	14	1.71	

Note. R^2 = .98, Adj. R^2 = .97, adjustments based on Motivation mean = 5.28. Homogeneity of regression tested and not significant: F = 1.03, p>.05. Motivation regression coefficient = 2.46*.
* p < .05

Table 4
Multiple Comparisons and Mean Differences in Achievement by Instruction Type Controlling for Motivation

Comparison	Mean Difference	s.e.	Bonferroni Adjusted 95% CI
CL vs. Lecture	-13.67*	0.78	-15.72, -11.62
CL vs. Self-paced	-7.54*	0.78	-9.65, -5.43
Lecture vs. Self-paced	6.13*	0.78	4.02, 8.24

Note. Comparisons based upon ANCOVA adjusted means controlling for motivation mean of 5.28. CL = cooperative learning.
* p < .05, where p-values are adjusted using the Bonferroni method.

ANCOVA results show that there are statistically significant instructional differences in achievement and that student motivation is positively associated with achievement. Multiple comparisons show that all groups differ, statistically, in achievement. Both the observed and adjusted means show that students in the lecture group performed best, followed by student in the self-paced group, with students in cooperative learning performing worst. Results also show that the greater one's motivation, the greater will be one's achievement.

Comment: It is important to test for homogeneity of regression (interaction between covariates and factors), and if the interaction is not statistically significant, the interaction term should be removed from the model and the model re-estimated. In the above example the interaction term between Motivation and Instruction was tested and removed since it was not statistically significant. More on interactions is presented below.

6. Statistical Adjustments in ANCOVA

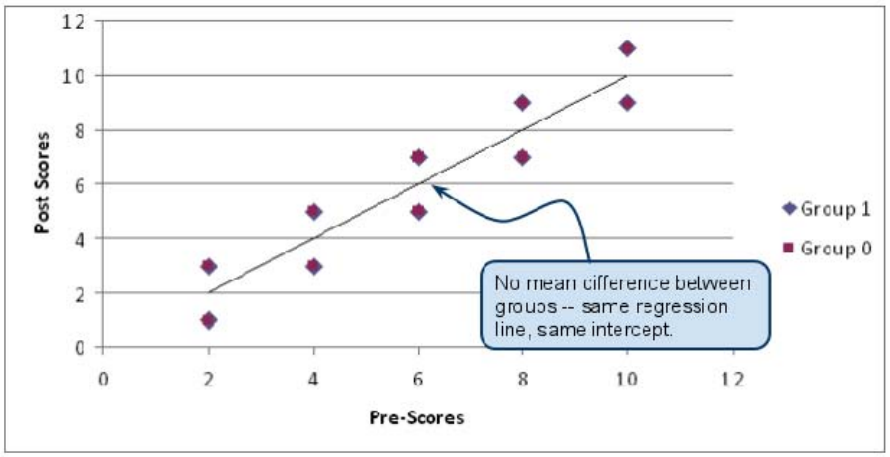
Data for these illustrations posted below in each table. Spreadsheet for data and scatterplots linked below.

http://www.bwgriffin.com/gsu/courses/edur8132/notes/Notes_9d_ANCOVA_Example_Separate_Lines.xls

Table 5
Pretest and Posttest Scores Same for Both Groups

Data for Illustration

pre	post	group
2	1	1
2	3	1
4	3	1
4	5	1
6	5	1
6	7	1
8	7	1
8	9	1
10	9	1
10	11	1
2	1	0
2	3	0
4	3	0
4	5	0
6	5	0
6	7	0
8	7	0
8	9	0
10	9	0
10	11	0



	Pretest	Posttest	Adjusted Posttest Mean
Group 1	6.00	6.00	6.00
Group 0	6.00	6.00	6.00
Grand Mean	6.00	6.00	

Covariates appearing in the model are evaluated at the following values: pre = 6.000.

Coefficients(a)

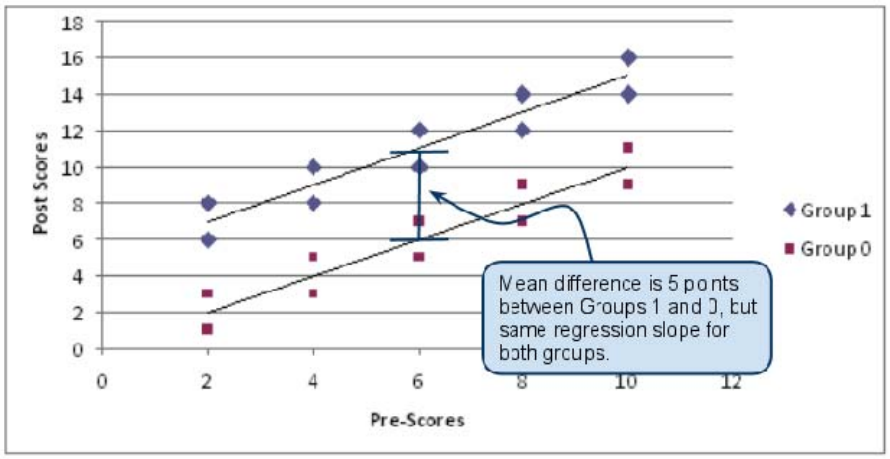
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	0.00	.618		.000	1.000
	group	.000	.485	.000	.000	1.000
	pre	1.000	.086	.943	11.662	.000

a Dependent Variable: post

Table 6
Posttest Scores 5 Points Higher for Group 1, All Else Same

Data for Illustration

pre	post	group
2	6	1
2	8	1
4	8	1
4	10	1
6	10	1
6	12	1
8	12	1
8	14	1
10	14	1
10	16	1
2	1	0
2	3	0
4	3	0
4	5	0
6	5	0
6	7	0
8	7	0
8	9	0
10	9	0
10	11	0



	Pretest	Posttest	Adjusted Posttest Mean
Group 1	6.00	11.00	11.00
Group 0	6.00	6.00	6.00
Grand Mean	6.00	8.50	

Covariates appearing in the model are evaluated at the following values: pre = 6.000.

Coefficients(a)

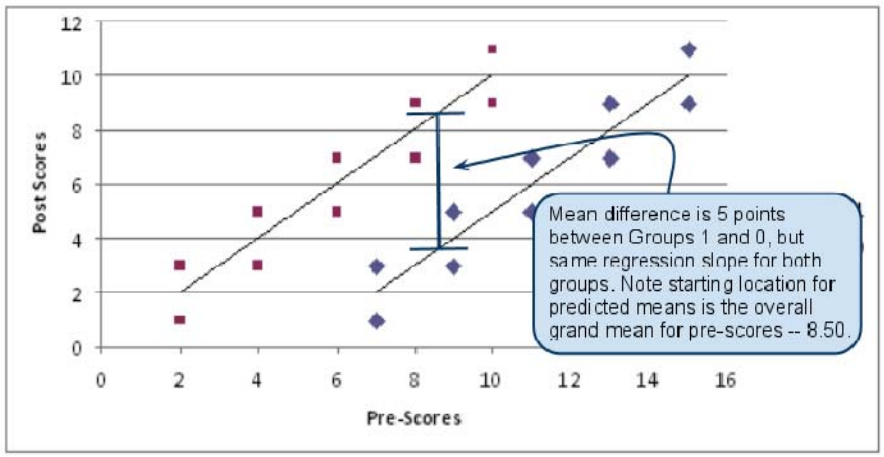
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.00	.618		.000	1.000
	group	5.000	.485	.640	10.308	.000
	pre	1.000	.086	.724	11.662	.000

a Dependent Variable: post

Table 7
Posttest Scores Same, but Pretest Scores 5 Points Higher for Group 1

Data for Illustration

pre	post	group
7	1	1
7	3	1
9	3	1
9	5	1
11	5	1
11	7	1
13	7	1
13	9	1
15	9	1
15	11	1
2	1	0
2	3	0
4	3	0
4	5	0
6	5	0
6	7	0
8	7	0
8	9	0
10	9	0
10	11	0



	Pretest	Posttest	Adjusted Posttest Mean
Group 1	11.00	6.00	3.50
Group 0	6.00	6.00	8.50
Grand Mean	8.50	6.00	

Covariates appearing in the model are evaluated at the following values: pre = 8.50.

Coefficients(a)

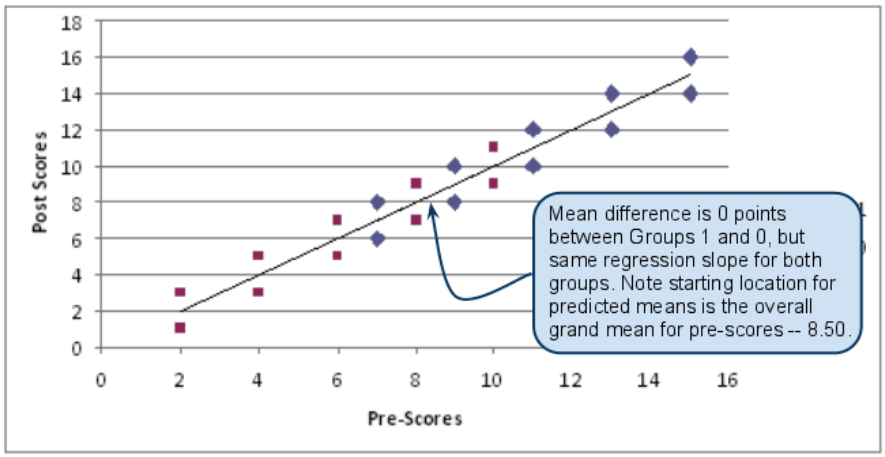
Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	0.00	.618		.000	1.000
	group	-5.000	.647	-.833	-7.723	.000
	pre	1.000	.086	1.258	11.662	.000

a Dependent Variable: post

Table 8
Both Posttest Scores and Pretest Scores 5 Points Higher for Group 1

Data for Illustration

pre	post	group
7	1	1
7	3	1
9	3	1
9	5	1
11	5	1
11	7	1
13	7	1
13	9	1
15	9	1
15	11	1
2	1	0
2	3	0
4	3	0
4	5	0
6	5	0
6	7	0
8	7	0
8	9	0
10	9	0
10	11	0



	Pretest	Posttest	Adjusted Posttest Mean
Group 1	11.00	11.00	8.50
Group 0	6.00	6.00	8.50
Grand Mean	8.50	8.50	

Covariates appearing in the model are evaluated at the following values: pre = 8.50.

Coefficients(a)

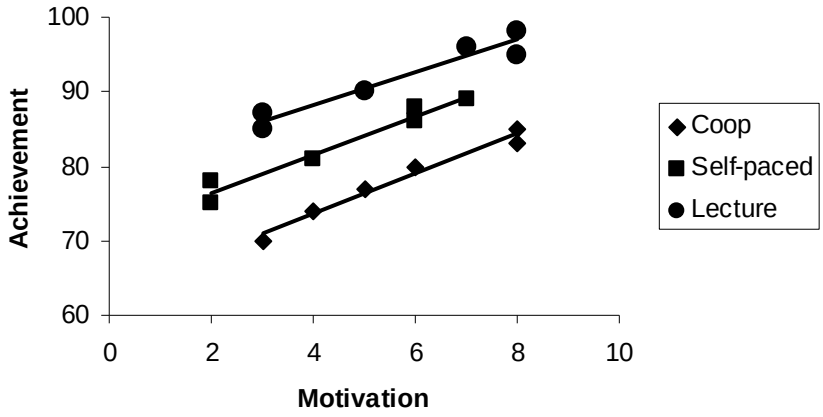
Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	0.00	.618		.000	1.000
	group	0.00	.647	.000	.000	1.000
	pre	1.000	.086	.967	11.662	.000

a Dependent Variable: post

7. Homogeneity of Regression

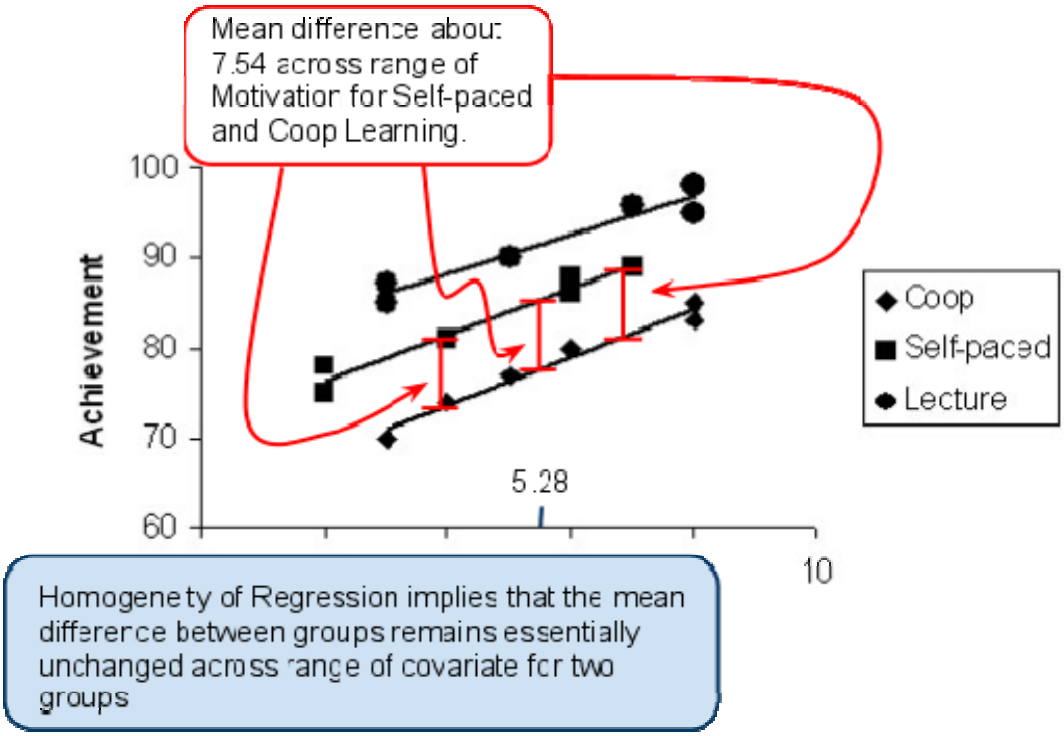
An assumption of ANCOVA is that the covariate regression slopes are parallel for each group. Stated differently, each group has covariate regression slope that is similar. This is known as ***homogeneity of regression slopes***. The regression slopes for the fictional achievement are presented below in Figure 2.

Figure 2
Scatterplot of Motivation and Achievement Showing Homogeneity of Regression Slopes



As Figure 2 shows, the regression slopes for each group are approximately the same—they are approximately parallel. This is important because it means one may calculate and test main effects for the ANCOVA factor (instruction in this case) since mean differences among groups is approximately equal throughout the range of the covariate. Figure 3 illustrates this.

Figure 3
Mean Difference Similar Across Range of Covariate If Homogeneity Assumption Met



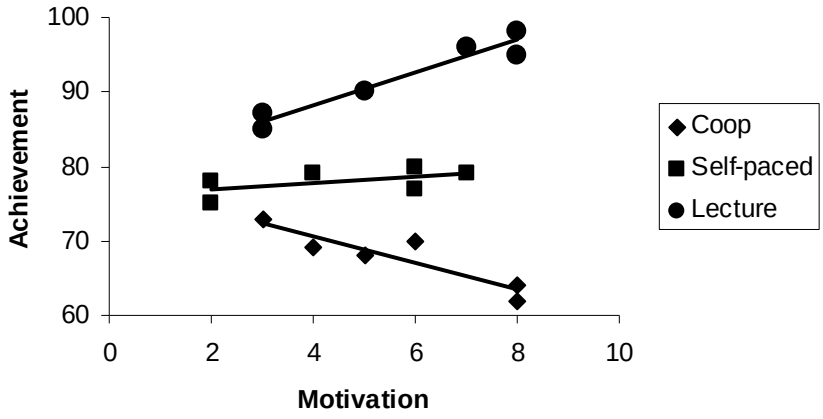
If the assumption of homogeneity of regression slopes is violated, this means an interaction between the factor and covariate exists. Below in Table 5 are data that violate homogeneity of regression.

Table 5
Fictional Achievement, Instruction, and Motivation Data with Interaction

Achievement	Instruction Type	Motivation
73	coop	3
69	coop	4
68	coop	5
70	coop	6
62	coop	8
64	coop	8
75	self	2
78	self	2
79	self	4
77	self	6
80	self	6
79	self	7
85	lecture	3
87	lecture	3
90	lecture	5
96	lecture	7
95	lecture	8
98	lecture	8

A graphical display of the interaction is provided in Figure 4.

Figure 4
Violation of Homogeneity of Regression; Interaction Between Motivation and Instruction



Why does heterogeneity of regression slopes present a problem for ANCOVA? Actually an interaction between factor and covariate is not a problem statistically—interactions are handled well in ANOVA and regression models. The problem occurs when one fails to recognize and appropriately model the interaction. Suppose, for instance, that the interaction was not test for the data in Table 5. The marginal mean differences for ANCOVA ignoring the interaction appear below in the SPSS output.

Tests of Between-Subjects Effects

Dependent Variable: ach

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1779.423(a)	3	593.141	37.562	.000
Intercept	13342.200	1	13342.200	844.914	.000
Motivation	15.090	1	15.090	.956	.345
Instruction	1757.588	2	878.794	55.651	.000
Error	221.077	14	15.791		
Total	114813.000	18			
Corrected Total	2000.500	17			

a R Squared = .889 (Adjusted R Squared = .866)

Estimated Marginal Means

Dependent Variable: ach

group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
coop	67.489(a)	1.632	63.988	70.990
lecture	91.656(a)	1.632	88.154	95.157
self	78.356(a)	1.663	74.790	81.922

a Covariates appearing in the model are evaluated at the following values: motv = 5.2778.

Pairwise Comparisons

Dependent Variable: ach

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
coop	lecture	-24.167(*)	2.294	.000	-30.402	-17.931
	self	-10.867(*)	2.358	.001	-17.276	-4.458
lecture	coop	24.167(*)	2.294	.000	17.931	30.402
	self	13.300(*)	2.358	.000	6.891	19.709
self	coop	10.867(*)	2.358	.001	4.458	17.276
	lecture	-13.300(*)	2.358	.000	-19.709	-6.891

Based on estimated marginal means

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

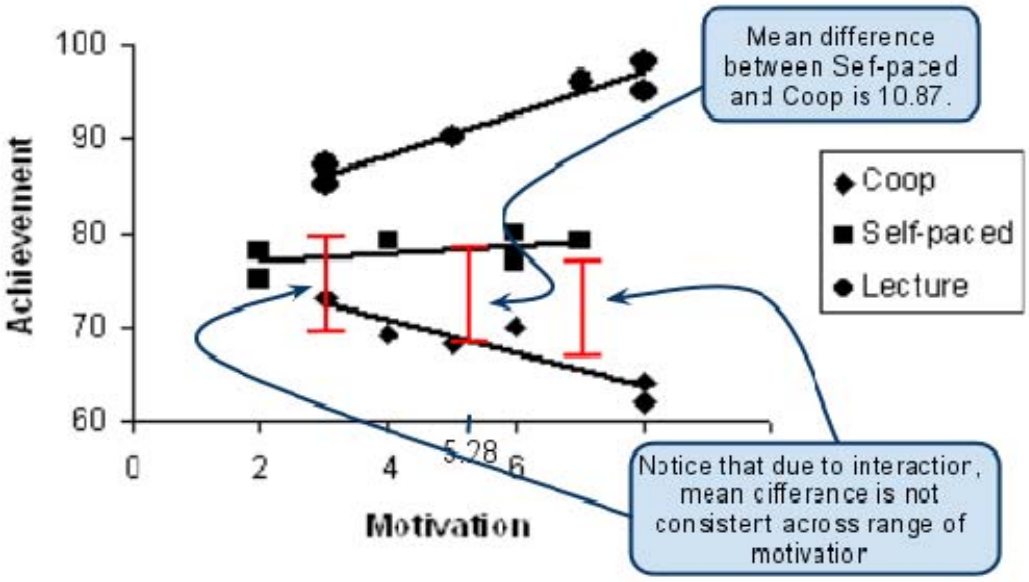
a Adjustment for multiple comparisons: Bonferroni.

According to the table of multiple comparisons above (which compares marginal means for each group), the average difference between groups are as follows:

Comparison	Mean Difference
Lecture vs. Coop	24.17
Lecture vs. Self-paced	13.30
Self-paced vs. Coop	10.87

These differences are plotted on the scatter plot below in Figure 5 for the Self-paced vs. Cooperative Learning comparison.

Figure 5
Lack of Consistent Mean Differences Between Groups Due to Interaction



As Figure 5 shows, ignoring the interaction means estimating mean differences among groups that will be incorrect for some regions of the covariate. In the above example, while the mean difference between Self-paced and Cooperative Learning is 10.87 points when Motivation = 5.28, the mean difference between Self-paced and Cooperative Learning is less when Motivation = 3 and greater when Motivation = 7. This difference, by definition, is an interaction—the statistical effects of one IV upon the DV differ across regions or categories of a second IV.

8. Testing for Homogeneity of Regression

The test of whether regression slopes are parallel for all groups is simple and requires only that an interaction be modeled. SPSS ANCOVA results for a model with an interaction for data in Table 5 are presented below.

Tests of Between-Subjects Effects

Dependent Variable: achievement

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1966.235(a)	5	393.247	137.721	.000
Intercept	13435.528	1	13435.528	4705.339	.000
instruction	5.401	2	2.700	.946	.416
motivation	6.221	1	6.221	2.179	.166
instruction * motivation	186.812	2	93.406	32.712	.000
Error	34.265	12	2.855		
Total	114813.000	18			
Corrected Total	2000.500	17			

a R Squared = .983 (Adjusted R Squared = .976)

The F-ratio for the interaction, F = 32.71 (p = .000), shows that the interaction is statistically significant at the .05 and .01 levels. This indicates that the assumption of homogeneity of regression is violated.

9. Addressing Heterogeneity of Regression

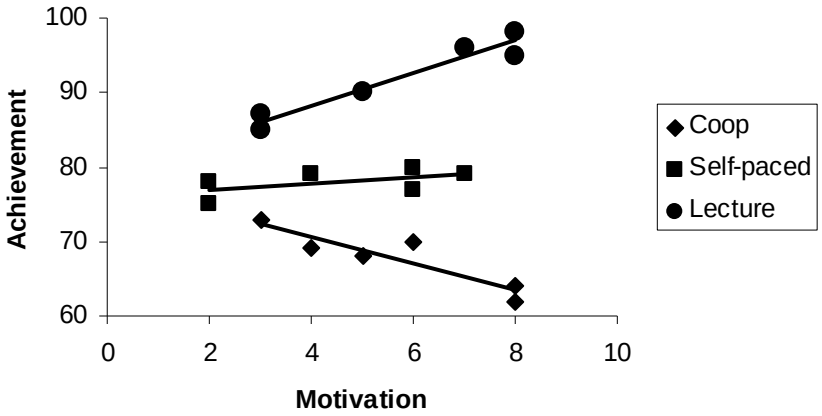
There are several methods statisticians use to address and interpret interactions in ANCOVA. Some are discussed below.

(a) Visual display and General Interpretation

With this approach one presents a graphical display of the interaction and provides a general discussion of results. This approach is less technical and does not present inference tests of comparisons; this omission may be viewed by some as a serious limitation. Below is an example of how one may present and discuss the interaction between Motivation and Instruction for data found in the Table 5 above.

> Results of the ANCOVA revealed an interaction between motivation and type of instruction. The interaction is plotted in Figure 4 (reposted below). In general differences in achievement among cooperative learning, self-paced, and lecture were less when motivation is low, but greater when motivation is high. Students in Lecture condition tended to perform better as motivation increases, while students in cooperative-learning tended to perform worse as motivation increases. Students in self-paced instruction tended to perform similar across the range of motivation although there seems to be a slight increase in achievement as motivation increases. On average those in lecture tended to perform at a higher level of achievement, those in self-paced performed, on average, below those in lecture, and those in cooperative learning tended to perform worse.

Figure 4 (Reposted)
Violation of Homogeneity of Regression; Interaction Between Motivation and Instruction



(b) Select Point Values Along Covariate and Perform Comparisons

An alternative method for presenting the interaction is to select various point values along the Motivation scale and perform pairwise comparisons among the groups at those defined points. SPSS automatically reports pairwise comparisons for the mean value of Motivation (5.28). One may additionally choose to use two other points on the Motivation scale, such as M ± SD. For Motivation SD = 2.14, so this would lead to 5.28 ± 2.14, thus the following three values of Motivation would be used to perform pairwise comparisons:

Low Motivation: 5.28 – 2.14 = 3.14
Mid Motivation: Mean = 5.28 (comparisons given by default in SPSS)
High Motivation: 5.28 + 2.14 = 7.42

These three points enables one to test group comparisons for low, middle, and high values of Motivation. Repeat the pairwise comparisons for each of three values of Motivation and report three tables of multiple comparisons. This process is illustrated below.

(1) Mid Motivation = 5.28

If one requests SPSS to present pairwise comparisons, then the mean of the covariate, or covariates, will be used as the point on the covariate scale. SPSS results are presented below.

Estimated Marginal Means

Dependent Variable: Achievement

Instruction	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
coop	68.353(a)	.704	66.819	69.888
lecture	90.984(a)	.701	89.457	92.512
self	78.364(a)	.741	76.749	79.979

a Covariates appearing in the model are evaluated at the following values: Motivation = 5.2778.

Pairwise Comparisons

Dependent Variable: Achievement

(I) Instruction	(J) Instruction	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
coop	lecture	-22.631(*)	.994	.000	-25.394	-19.869
	self	-10.011(*)	1.023	.000	-12.853	-7.169
lecture	coop	22.631(*)	.994	.000	19.869	25.394
	self	12.620(*)	1.020	.000	9.784	15.456
self	coop	10.011(*)	1.023	.000	7.169	12.853
	lecture	-12.620(*)	1.020	.000	-15.456	-9.784

Based on estimated marginal means

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

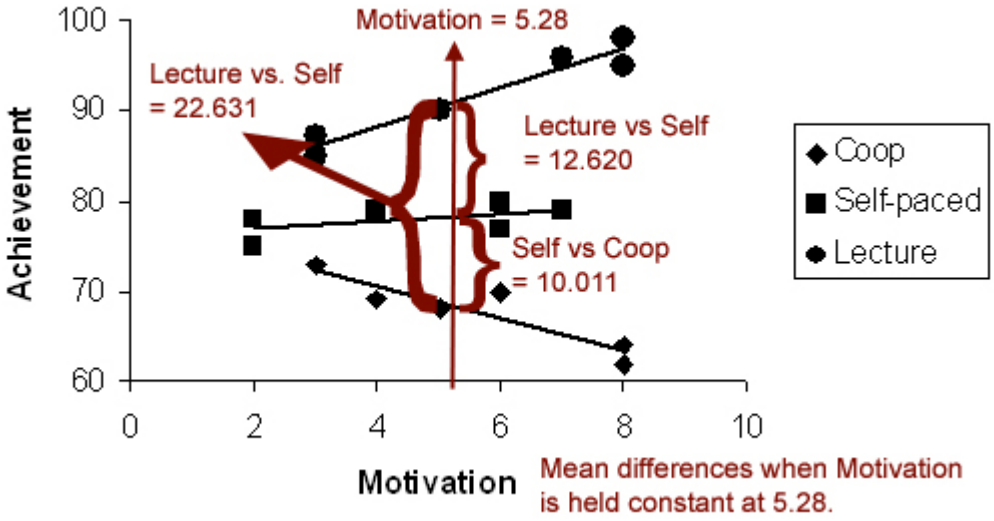
a Adjustment for multiple comparisons: Bonferroni.

The estimated marginal means—the adjusted or predicted means—are reported in the first table above to be coop = 68.353, lecture = 90.984, and self = 78.364 when Motivation = 5.2778 (see footnote for first table above). The following pairwise comparisons hold (mean differences are calculated below, but are also displayed in the SPSS Pairwise Comparison table above):

Comparison	Mean Difference
Lecture vs. Coop	90.984 – 68.353 = 22.631
Lecture vs. Self	90.984 – 78.364 = 12.620
Self vs. Coop	78.364 – 68.353 = 10.011

Graphically these differences are displayed in Figure 6 below.

Figure 6
Estimated Instructional Mean Differences in Achievement when Motivation = 5.28



(2) Low Motivation = 3.14

To obtain pairwise comparisons among Instructional groups when Motivation = 3.14, one must use the Paste command in SPSS to obtain ANOVA command syntax. Below is the SPSS syntax.

```
UNIANOVA
  Achievement  BY Instruction  WITH Motivation
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /EMMEANS = TABLES(Instruction) WITH(Motivation=MEAN) COMPARE ADJ(BONFERRONI)
  /PRINT = DESCRIPTIVE
  /CRITERIA = ALPHA(.05)
  /DESIGN = Instruction Motivation Instruction*Motivation .
```

Make the following change to request comparisons for Motivation = 3.14:

```
UNIANOVA
  Achievement  BY Instruction  WITH Motivation
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /EMMEANS = TABLES(Instruction) WITH(Motivation=3.14) COMPARE ADJ(BONFERRONI)
  /PRINT = DESCRIPTIVE
  /CRITERIA = ALPHA(.05)
  /DESIGN = Instruction Motivation Instruction*Motivation .
```

Output from the above command appears below. Note the value of Motivation in the footnote of the table below.

Estimates

Dependent Variable: Achievement

Instruction	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
coop	72.128(a)	1.153	69.615	74.641
lecture	86.318(a)	1.069	83.989	88.647
self	77.363(a)	.837	75.540	79.187

a Covariates appearing in the model are evaluated at the following values: Motivation = 3.14.

Pairwise Comparisons

Dependent Variable: Achievement

(I) Instruction	(J) Instruction	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
coop	lecture	-14.190(*)	1.573	.000	-18.561	-9.819
	self	-5.236(*)	1.425	.010	-9.197	-1.274
lecture	coop	14.190(*)	1.573	.000	9.819	18.561
	self	8.954(*)	1.358	.000	5.181	12.728
self	coop	5.236(*)	1.425	.010	1.274	9.197
	lecture	-8.954(*)	1.358	.000	-12.728	-5.181

Based on estimated marginal means

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

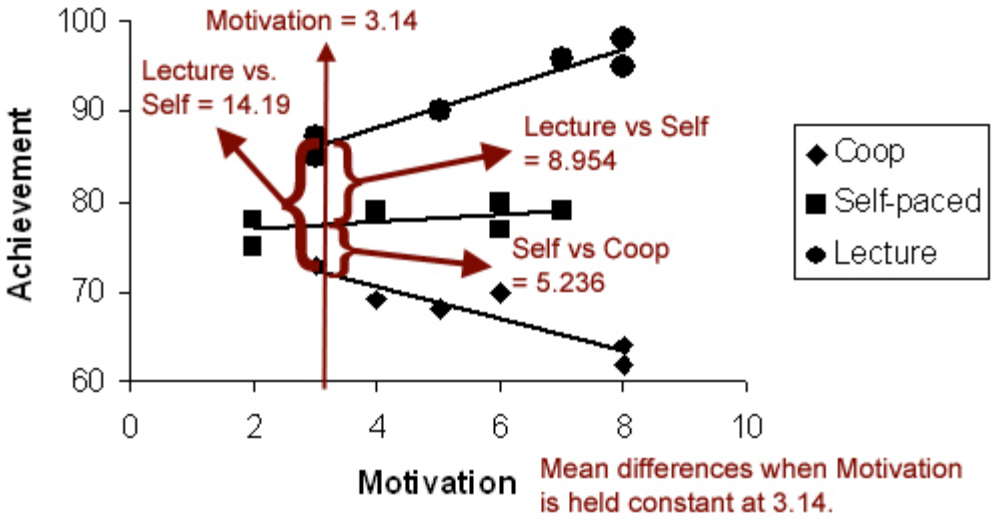
a Adjustment for multiple comparisons: Bonferroni.

When Motivation = 3.14, the following pairwise comparisons hold (mean differences are calculated below, but are also displayed in the SPSS Pairwise Comparison table above):

Comparison	Mean Difference
Lecture vs. Coop	86.318 – 72.128 = 14.190
Lecture vs. Self	86.318 – 77.363 = 8.954
Self vs. Coop	77.363 – 72.128 = 5.236

The results of these comparisons are show in Figure 7.

Figure 7
Estimated Instructional Mean Differences in Achievement when Motivation = 3.14



(3) High Motivation = 7.42

Pairwise comparisons among Instructional groups when Motivation = 7.42 may be obtained with the following change to the SPSS commands:

```
UNIANOVA
  Achievement  BY Instruction  WITH Motivation
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /EMMEANS = TABLES(Instruction) WITH(Motivation=7.42) COMPARE ADJ(BONFERRONI)
  /PRINT = DESCRIPTIVE
  /CRITERIA = ALPHA(.05)
  /DESIGN = Instruction Motivation Instruction*Motivation .
```

Estimates

Dependent Variable: Achievement

Instruction	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
coop	64.571(a)	.942	62.518	66.623
lecture	95.661(a)	.893	93.716	97.606
self	79.367(a)	1.230	76.688	82.046

a Covariates appearing in the model are evaluated at the following values: Motivation = 7.42.

Pairwise Comparisons

Dependent Variable: Achievement

(I) Instruction	(J) Instruction	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
coop	lecture	-31.090(*)	1.298	.000	-34.697	-27.482
	self	-14.796(*)	1.549	.000	-19.101	-10.491
lecture	coop	31.090(*)	1.298	.000	27.482	34.697
	self	16.294(*)	1.520	.000	12.070	20.517
self	coop	14.796(*)	1.549	.000	10.491	19.101
	lecture	-16.294(*)	1.520	.000	-20.517	-12.070

Based on estimated marginal means

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

a Adjustment for multiple comparisons: Bonferroni.

Figure 8 plots the comparisons at Motivation = 7.42. Results of these three analyses could be presented in a table like that following below. In addition, some simple graphical display showing the interactions and points along the covariate indicating where comparisons were made would be helpful. See, for example, Figure 9.

Table X

Multiple Comparisons of Achievement by Instructional Groups for Three Levels of Motivation

Comparison	Adjusted Mean Difference	Standard Error of Difference	Bonferroni Adjusted 95% CI
Motivation = M + SD = 7.42			
Lecture vs. Coop	31.09	X	x, x
Lecture vs. Self	16.29	X	x, x
Self vs. Coop	14.80	X	x, x
Motivation = M = 5.28			
Lecture vs. Coop	22.63	X	x, x
Lecture vs. Self	12.62	X	x, x
Self vs. Coop	10.01	X	x, x
Motivation = M – SD = 3.14			
Lecture vs. Coop	14.19	X	x, x
Lecture vs. Self	8.95	X	x, x
Self vs. Coop	5.24	x	x, x

Replace X with appropriate values.

Figure 8
Estimated Instructional Mean Differences in Achievement when Motivation = 7.42

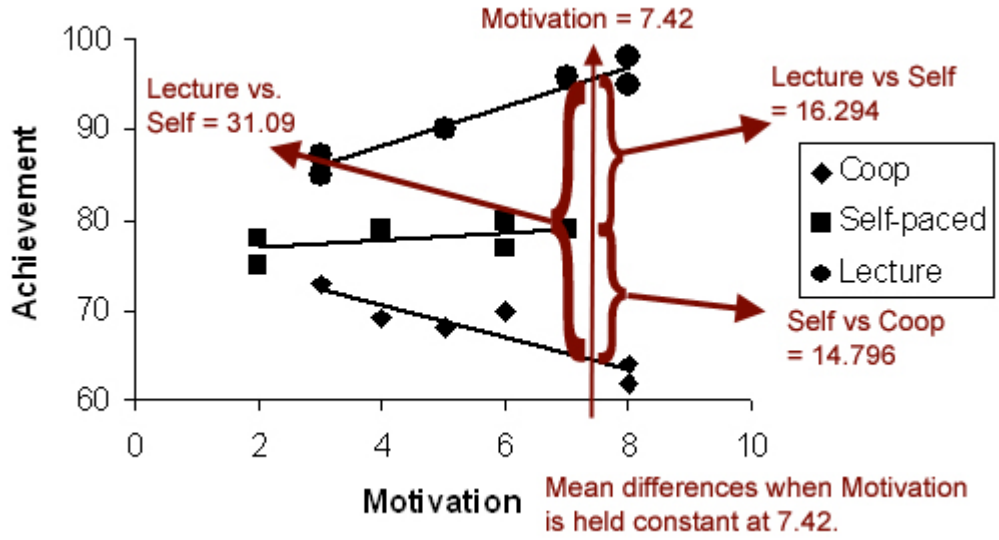
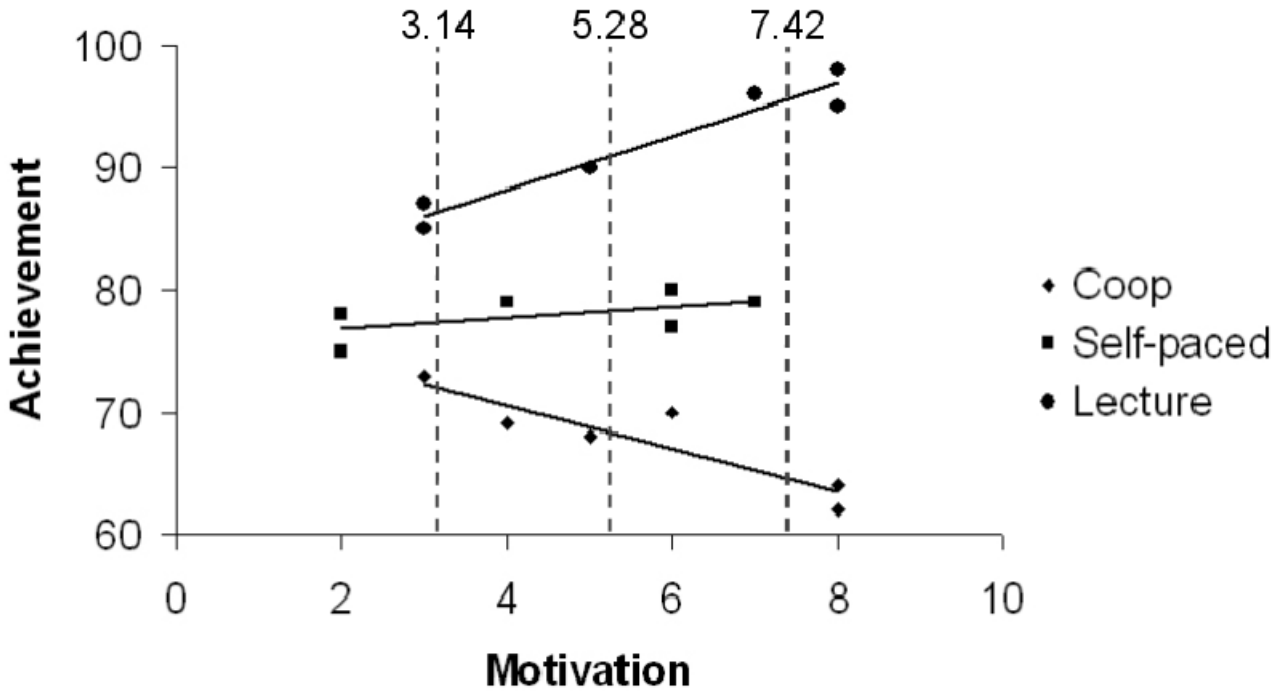


Figure 9
Points Along Motivation Scale at which Multiple Comparisons of Instruction Performed



(c) Johnson-Neyman Approach

This approach, developed by Johnson and Neyman (1936), is technical but useful. Illustration and discussion will be added. An example can be found in the citation below:

Fraas, J.W, & Newman, I. (date?). The Use of the Johnson-Neyman Confidence Bands and Multiple Regression Models to Investigate Interaction Effects: Important Tools for Educational Researchers and Program Evaluators. Multiple Linear Regression Viewpoints, 24, 14-24.

http://www.bwgriffin.com/gsu/courses/edur8132/notes/Fraas_Johnson_Neyman_interaction_procedure.pdf

A scarily technical illustration can be found here:

Charles J. Kowalskia, Emet D. Schneiderman, Stephen M. Willis (1994). ANCOVA for nonparallel slopes: the Johnson-Neyman technique. International Journal of Bio-Medical Computing 37 (1994) 273-286

http://deepblue.lib.umich.edu/bitstream/2027.42/31210/1/0000112.pdf

(d) Two-way ANOVA Approach

Another option, which is similar to option (b) above, is to convert Motivation from a quantitative variable to a qualitative variable by artificially creating categories, such as Low vs. High, or Low, Middle, and High, then perform two-way ANOVA. This allows for easier presentation of interaction and tests of group differences. A drawback of this approach is deciding how to create categories and the loss of power and precision by reducing a quantitative variable to a qualitative variable.

Illustration of this approach may be added later.

9. ANCOVA Extensions

ANCOVA can be extended to include multiple factors and multiple covariates. If more than one factor is involved, the analysis and reporting follows the format used in Two-way ANOVA and multi-way ANOVA with the exception of adding covariate information (e.g., covariates' SS, df, MS, and F-ratios in the summary table; adjusted means). For each covariate report mean values and regression coefficients so readers know what values were used to obtain predicted—adjusted—means on the DV and how the covariates are related to the DV (e.g., whether coefficients are positive or negative). One should also test for possible interactions between each factor and each covariate.

10. Exercises

10a Avoidance

Taken from Stevens (2007). Examine the effect of behavioral rehearsal (BH), and behavioral rehearsal plus cognitive restructuring (BH+CR; a combination treatment) on reducing anxiety and facilitating social skills for female college freshmen. The 33 participants were randomly assigned (11 to each group) to Training Conditions of either the BH condition, a control group with no treatment, or the BH+CR condition. A pretest-posttest control group design was employed and participants were pretested before treatment and posttested after treatment on several variables. One variable measured was avoidance. Avoidance is a conscious or unconscious defense mechanism by which a person tries to escape from unpleasant situations or feelings, such as anxiety and pain. The scores for avoidance are given in the table below. Is there any evidence that treatments helped alter avoidance tendencies? It is not clear whether higher scores reflect more or less avoidance. Data are presented below.

Avoidance Data

BH		Control		BH+CR	
Avoid	Pre-Avoidance	Avoid	Pre-Avoidance	Avoid	Pre-Avoidance
91	70	107	115	121	96
107	121	76	77	140	120
121	89	116	111	148	130
86	80	126	121	147	145
137	123	104	105	139	122
138	112	96	97	121	119
133	126	127	132	141	104
127	121	99	98	143	121
114	80	94	85	120	80
118	101	92	82	140	121
114	112	128	112	95	92

10b Student Ratings

Student ratings are a common assessment tool in universities across the USA. Some argue that an instructor's reputation may taint student ratings by creating a presupposition prior to any exposure to an instructor. Additionally, some faculty members believe the instructor's sex offer some advantage in ratings. Lastly, it many believe that the more intrinsic interest students have in a given subject, the higher will be their rating for that course and instructor. To address these questions, data were collected from a number of courses and ratings from over 900 students were recorded. These data are located here:

http://www.bwgriffin.com/gsu/courses/edur8132/data/Student_Ratings.sav

Ratings on several dimensions (e.g., evaluation fairness, clarity, organization, openness) of teaching were recorded and the mean was calculated across all dimensions. That overall mean rating is identified in the data file above as follows:

mean_rating_4to14 = mean rating for items 4 to 14 on questionnaire; represents mean rating of instruction

The three predictors are inst_sex (0 = female, 1 = male), pos_neg (Instruction Reputation, coded as 1 = negative, 2 = no information, 3 = positive), and intrin (Intrinsic Motivation where higher scores indicate greater natural interest in the subject matter for the course evaluated).

10c Mathematics Achievement

Reciprocal peer tutoring (RPT) is a method by which students develop test items, administer test items to a partner, then tutor their partner on those items answered incorrectly by their partner. Does RPT enhance mathematics achievement? Also, if RPT is effective in enhancing mathematics learning, what is it about RPT that makes it effective? Does the tutoring part play the key role, or does developing the test items make students learn the material better? To address this question, a randomized study was developed with four conditions:

1. RPT = Dyads of students are created and each student develops three sets of 10-item tests on relevant mathematical topics, administers those tests to their partner, and then tutors their partner on those items answered incorrectly. Partners then reverse roles.
2. Tutor Only = As above, dyads are created and then instructed to tutor each other on assigned mathematical concepts.
3. Test Only = Students are instructed construct three 10-item tests on relevant mathematical topics, and then submit those tests to the instructor for review.
4. Control = Students in this condition are asked to study individually assigned mathematical material.

Before the experiment began, students were pretested on mathematical concepts that would be taught over the next 4 weeks. Students were randomly assigned to one of the four conditions described above. Over the course of the 4 weeks students participated in their assigned conditions. At the end of the 4 weeks a posttest was administered. Both pretest and posttest reflect percentage correct scores. The data for this study appears below.

Mathematical Performance Data

RPT		Tutor Only		Test Only		Control	
Math Posttest	Math Pretest	Math Posttest	Math Pretest	Math Posttest	Math Pretest	Math Posttest	Math Pretest
88	51	84	43	86	51	66	41
88	45	86	45	72	32	73	38
89	55	86	50	68	34	76	35
93	46	90	46	78	43	73	46
94	34	75	34	81	45	78	45
96	58	95	58	82	46	81	52
97	43	96	43	89	55	83	48
98	55			92	55	85	59
				92	58	87	49
						90	62

11. Answers to Exercises

10a Avoidance: SPSS Results

Check for interaction between factor and covariate (Training × Pre-Avoidance Scores). If that is not significant, remove the interaction and rerun the ANCOVA with just main effects for Training Condition and Pre-Avoidance scores. The ANCOVA summary table with the interaction included is presented below. Note the interaction is not statistically significant (F = 0.668, p = 0.521).

Tests of Between-Subjects Effects

Dependent Variable: Avoidance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9182.548(a)	5	1836.510	16.115	.000
Intercept	1590.068	1	1590.068	13.952	.001
Training	345.196	2	172.598	1.514	.238
Preavoidance	5298.802	1	5298.802	46.494	.000
Training * Preavoidance	152.306	2	76.153	.668	.521
Error	3077.088	27	113.966		
Total	474588.000	33			
Corrected Total	12259.636	32			

a R Squared = .749 (Adjusted R Squared = .703)

Since the interaction was not statistically significant it was removed. Below are SPSS results for the ANCOVA without the interaction effect.

Descriptive Statistics

Dependent Variable: Avoidance

Training	Mean	Std. Deviation	N
BH	116.9091	17.23052	11
BH+CR	132.2727	16.16843	11
Control	105.9091	16.78961	11
Total	118.3636	19.57329	33

Tests of Between-Subjects Effects

Dependent Variable: Avoidance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9030.243(a)	3	3010.081	27.031	.000
Intercept	1719.396	1	1719.396	15.440	.000
Preavoidance	5172.606	1	5172.606	46.450	.000
Training	1915.446	2	957.723	8.600	.001
Error	3229.394	29	111.358		
Total	474588.000	33			
Corrected Total	12259.636	32			

a R Squared = .737 (Adjusted R Squared = .709)

Parameter Estimates

Dependent Variable: Avoidance

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	33.653	11.069	3.040	.005	11.014	56.291
Preavoidance	.700	.103	6.815	.000	.490	.910
[Training=BH]	11.000	4.500	2.445	.021	1.797	20.203
[Training=BH+CR]	19.042	4.626	4.116	.000	9.581	28.504
[Training=Control]	0(a)	.	.	.	.	.

a This parameter is set to zero because it is redundant.

Comment: The table below provides the adjusted posttest avoidance means for the three groups taking into account their initial pre-avoidance scores.

Estimates

Dependent Variable: Avoidance

Training	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
BH	119.349(a)	3.202	112.801	125.898
BH+CR	127.392(a)	3.261	120.722	134.062
Control	108.349(a)	3.202	101.801	114.898

a Covariates appearing in the model are evaluated at the following values: Preavoidance = 106.6667.

Pairwise Comparisons

Dependent Variable: Avoidance

(I) Training	(J) Training	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
BH	BH+CR	-8.042	4.626	.278	-19.797	3.712
	Control	11.000	4.500	.062	-.433	22.433
BH+CR	BH	8.042	4.626	.278	-3.712	19.797
	Control	19.042(*)	4.626	.001	7.288	30.797
Control	BH	-11.000	4.500	.062	-22.433	.433
	BH+CR	-19.042(*)	4.626	.001	-30.797	-7.288

Based on estimated marginal means

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

a Adjustment for multiple comparisons: Bonferroni.

10a Avoidance: APA Results

Table 10a1
ANCOVA Results and Descriptive Statistics for Avoidance by Training Condition and Pre-Avoidance Scores

Type of Training	Avoidance			
	Observed Mean	Adjusted Mean	SD	n
Behavioral Rehearsal (BH)	116.91	119.35	17.23	11
BH + Cognitive Restructuring	132.27	127.39	16.17	11
Control	105.91	108.35	16.79	11

Source	SS	df	MS	F
Pre-avoidance	5172.61	1	5172.61	46.45*
Training	1915.45	2	957.72	8.60*
Error	3229.39	29	111.36	

Note. $R^2 = .74$, Adj. $R^2 = .71$, adjustments based on Pre-avoidance mean = 106.67. Homogeneity of regression tested and not significant: F = 0.67, p>.05. Pre-avoidance regression coefficient = 0.70*.
* p < .05

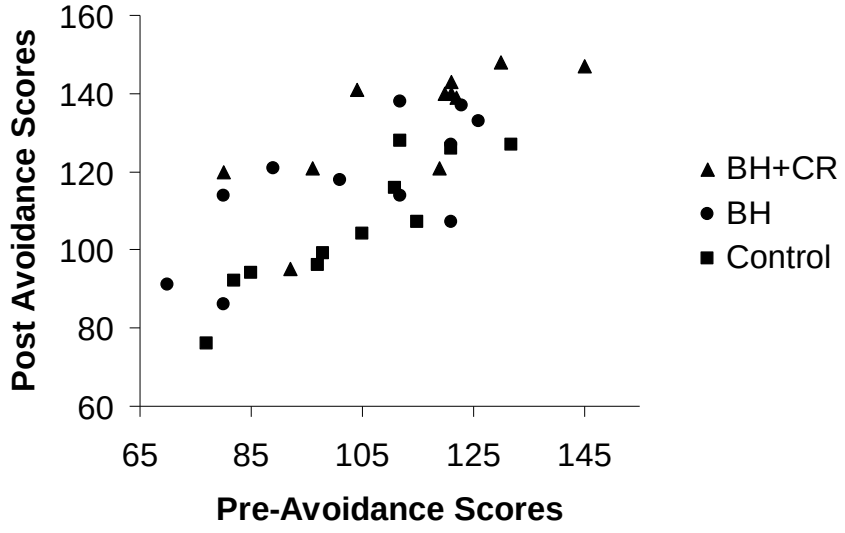
Table 10a2
Multiple Comparisons and Mean Differences in Avoidance by Training Condition

Comparison	Mean Difference	s.e.	Bonferroni Adjusted 95% CI
BH vs. BH+CR	-8.04	4.63	-19.80, 3.71
BH vs. Control	11.00	4.50	-0.43, 22.43
BH+CR vs. Control	19.04*	4.63	7.29, 30.80

Note. Comparisons based upon ANCOVA adjusted means controlling for pre-avoidance mean of 106.67. BH = Behavioral Rehearsal and CR = Cognitive Restructuring.
* p < .05, where p-values are adjusted using the Bonferroni method.

ANCOVA results indicate that mean avoidance scores differ by training conditions, and that there is a positive association between pre-avoidance and post-avoidance scores. Students in both BH and BH+CR conditions display adjusted avoidance means that are higher than the mean for the control students, but only those in the BH+CR condition are statistically higher according to the table of multiple comparisons. Figure 10a1 below shows the nature of the association for each of the conditions.

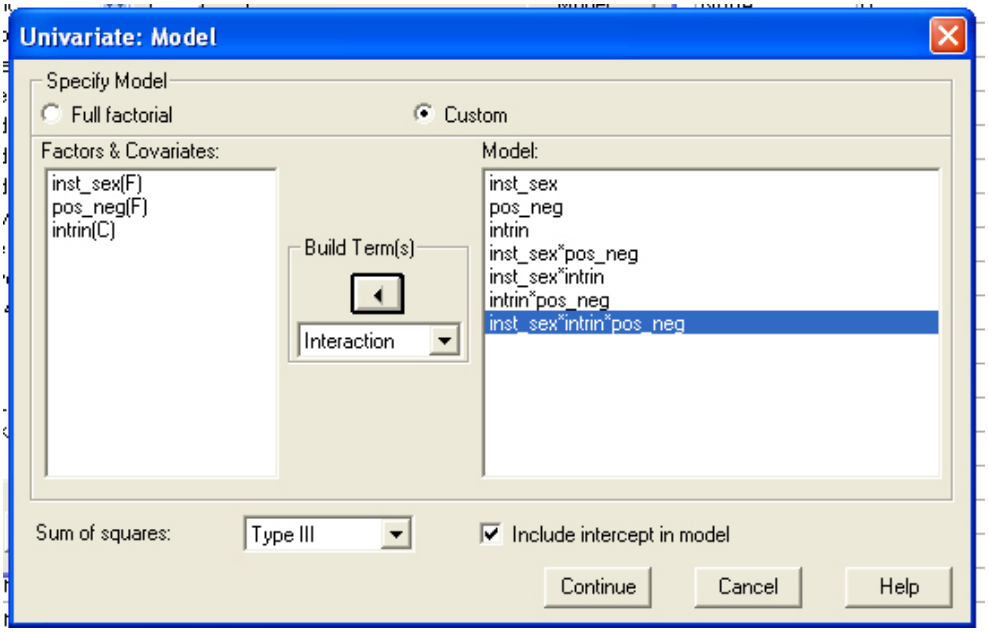
Figure 10a1
Scatter Plot of Pre and Post Avoidance Scores



10b Student Ratings: SPSS Results

Since there are two factors (teacher sex and reputation) and one covariate (intrinsic motivation), it will be necessary to form and test the following interactions (formation in SPSS shown below in the screen shot):

Sex × Reputation
Sex × Intrinsic Motivation
Reputation × Intrinsic Motivation
Sex × Reputation × Intrinsic Motivation (this is a three-way interaction)



The relevant tests for these interactions can be found in the ANOVA summary table from SPSS:

Tests of Between-Subjects Effects

Dependent Variable: mean_rating_4to14

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	193.994(a)	11	17.636	42.683	.000
Intercept	172.479	1	172.479	417.440	.000
inst_sex	.848	1	.848	2.051	.152
pos_neg	1.121	2	.560	1.356	.258
intrin	62.239	1	62.239	150.633	.000
inst_sex * pos_neg	.181	2	.090	.219	.803
inst_sex * intrin	.321	1	.321	.777	.378
pos_neg * intrin	.452	2	.226	.547	.579
inst_sex * pos_neg * intrin	.027	2	.013	.032	.968
Error	375.170	908	.413		
Total	16771.248	920			
Corrected Total	569.164	919			

a R Squared = .341 (Adjusted R Squared = .333)

Note that the 3-way interaction of Sex × Reputation × Intrinsic Motivation is not statistically significant ($F = 0.03$, $p = 0.97$). As a result, that interaction can be removed and the model re-run with only the two-way interactions present:

Sex × Reputation
Sex × Intrinsic Motivation
Reputation × Intrinsic Motivation

Tests of Between-Subjects Effects

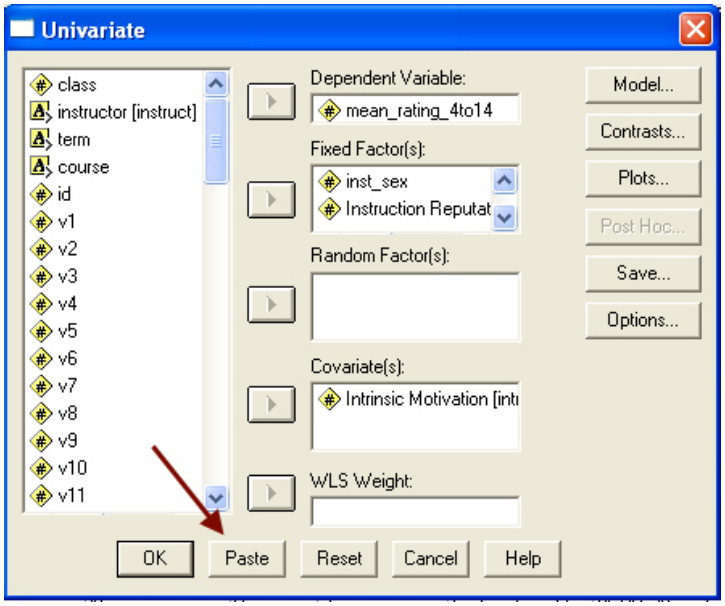
Dependent Variable: mean_rating_4to14

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	193.967(a)	9	21.552	52.272	.000
Intercept	199.369	1	199.369	483.547	.000
inst_sex	1.339	1	1.339	3.247	.072
pos_neg	1.148	2	.574	1.392	.249
intrin	72.364	1	72.364	175.512	.000
inst_sex * pos_neg	4.309	2	2.155	5.226	.006
inst_sex * intrin	.504	1	.504	1.222	.269
pos_neg * intrin	.715	2	.357	.867	.421
Error	375.197	910	.412		
Total	16771.248	920			
Corrected Total	569.164	919			

a R Squared = .341 (Adjusted R Squared = .334)

The results above show that two of the two-way interactions were not significant, Sex × Intrinsic Motivation ($F = 1.22$, $p = .27$) and Reputation × Intrinsic Motivation ($F = 0.87$, $p = 0.42$). Both of these should be removed and the model re-run to include the one significant interaction of Sex × Reputation.

Since there is an interaction present, it is necessary to alter the syntax commands to obtain simple main effects in ANCOVA just like done earlier with two-way ANOVA. The command alternations needed are presented below. To obtain the commands, click on Paste once the ANCOVA options are selected in SPSS.



Alter the following syntax from this:

```
UNIANOVA
  mean_rating_4to14  BY inst_sex pos_neg  WITH intrin
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /PLOT = PROFILE( inst_sex*pos_neg pos_neg*inst_sex )
  /EMMEANS = TABLES(inst_sex) WITH(intrin=MEAN) COMPARE ADJ(BONFERRONI)
  /EMMEANS = TABLES(pos_neg) WITH(intrin=MEAN) COMPARE ADJ(BONFERRONI)
  /EMMEANS = TABLES(inst_sex*pos_neg) WITH(intrin=MEAN)
  /PRINT = DESCRIPTIVE PARAMETER
  /CRITERIA = ALPHA(.05)
  /DESIGN = inst_sex pos_neg intrin inst_sex*pos_neg .
```

to this:

```
UNIANOVA
  mean_rating_4to14  BY inst_sex pos_neg  WITH intrin
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /PLOT = PROFILE( inst_sex*pos_neg pos_neg*inst_sex )
  /EMMEANS = TABLES(inst_sex) WITH(intrin=MEAN) COMPARE ADJ(BONFERRONI)
  /EMMEANS = TABLES(pos_neg) WITH(intrin=MEAN) COMPARE ADJ(BONFERRONI)
  /EMMEANS = TABLES(inst_sex*pos_neg) WITH(intrin=MEAN) COMPARE(inst_sex) ADJ(BONFERRONI)
  /EMMEANS = TABLES(pos_neg*inst_sex) WITH(intrin=MEAN) COMPARE(pos_neg) ADJ(BONFERRONI)
  /PRINT = DESCRIPTIVE PARAMETER
  /CRITERIA = ALPHA(.05)
  /DESIGN = inst_sex pos_neg intrin inst_sex*pos_neg .
```

Comment: Note the line added reversed the order in the TABLES(A*B) from inst_sex*pos_neg to pos_neg*inst_sex. This change will provide a helpful descriptive table of achievements. One table will show achievement means for each reputation level separately by sex, the other will show achievement means for each sex separately by reputation levels.

SPSS results for this final model are presented below.

Descriptive Statistics

Dependent Variable: mean_rating_4to14

inst_sex	Instruction Reputation	Mean	Std. Deviation	N
Female	Negative Info	3.7027	.87067	111
	No Info	4.4739	.62408	287
	Positive Info	4.4831	.62157	159
	Total	4.3228	.74566	557
Male	Negative Info	3.6025	.67818	59
	No Info	4.0112	.86345	235
	Positive Info	4.3162	.54403	69
	Total	4.0028	.81000	363
Total	Negative Info	3.6679	.80843	170
	No Info	4.2656	.77571	522
	Positive Info	4.4326	.60289	228
	Total	4.1965	.78698	920

Comment: The means tabled above can be used to plot a graph showing the interaction between sex and reputation.

Tests of Between-Subjects Effects

Dependent Variable: mean_rating_4to14

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	192.901(a)	6	32.150	78.012	.000
Intercept	296.635	1	296.635	719.782	.000
inst_sex	3.476	1	3.476	8.435	.004
pos_neg	27.744	2	13.872	33.661	.000
intrin	100.816	1	100.816	244.630	.000
inst_sex * pos_neg	4.172	2	2.086	5.061	.007
Error	376.263	913	.412		
Total	16771.248	920			
Corrected Total	569.164	919			

a R Squared = .339 (Adjusted R Squared = .335)

Comment: The interaction between sex and reputation is statistically significant (p = .007). As a result, a table of simple main effects can be used to show mean differences.

Comment: Some SPSS omitted since the table of main interest will be partial adjusted means and simple main effect pairwise comparisons.

Estimates

Dependent Variable: mean_rating_4to14

inst_sex	Instruction Reputation	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Female	Negative Info	3.793(a)	.061	3.673	3.913
	No Info	4.413(a)	.038	4.339	4.488
	Positive Info	4.377(a)	.051	4.276	4.477
Male	Negative Info	3.800(a)	.085	3.634	3.966
	No Info	4.081(a)	.042	3.998	4.163
	Positive Info	4.264(a)	.077	4.112	4.416

a Covariates appearing in the model are evaluated at the following values: Intrinsic Motivation = 3.6036.

Comment: The table above shows achievement means for each reputation level separately by sex. The table below shows achievement means for each sex separately by reputation level.

Estimates

Dependent Variable: mean_rating_4to14

Instruction Reputation	inst_sex	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Negative Info	Female	3.793(a)	.061	3.673	3.913
	Male	3.800(a)	.085	3.634	3.966
No Info	Female	4.413(a)	.038	4.339	4.488
	Male	4.081(a)	.042	3.998	4.163
Positive Info	Female	4.377(a)	.051	4.276	4.477
	Male	4.264(a)	.077	4.112	4.416

a Covariates appearing in the model are evaluated at the following values: Intrinsic Motivation = 3.6036.

Comment: The two tables below show simple main effects, the first by levels of reputation and the second by levels of sex.

Pairwise Comparisons

Dependent Variable: mean_rating_4to14

Instruction Reputation	(I) inst_sex	(J) inst_sex	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
						Lower Bound	Upper Bound
Negative Info	Female	Male	-.007	.104	.947	-.210	.197
	Male	Female	.007	.104	.947	-.197	.210
No Info	Female	Male	.333(*)	.057	.000	.221	.445
	Male	Female	-.333(*)	.057	.000	-.445	-.221
Positive Info	Female	Male	.113	.093	.224	-.069	.295
	Male	Female	-.113	.093	.224	-.295	.069

Based on estimated marginal means

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

a Adjustment for multiple comparisons: Bonferroni.

Pairwise Comparisons

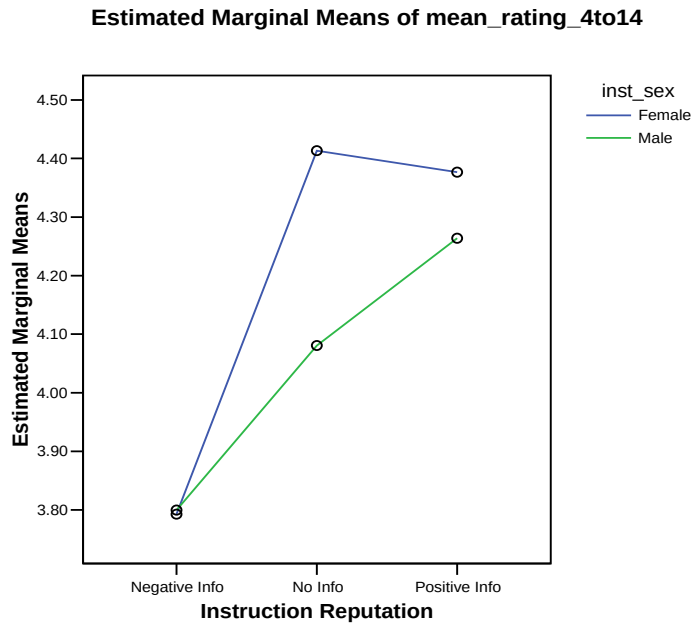
Dependent Variable: mean_rating_4to14

inst_sex	(I) Instruction Reputation	(J) Instruction Reputation	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
						Lower Bound	Upper Bound
Female	Negative Info	No Info	-.621(*)	.072	.000	-.794	-.447
		Positive Info	-.584(*)	.080	.000	-.777	-.391
	No Info	Negative Info	.621(*)	.072	.000	.447	.794
		Positive Info	.037	.064	1.000	-.116	.189
	Positive Info	Negative Info	.584(*)	.080	.000	.391	.777
		No Info	-.037	.064	1.000	-.189	.116
Male	Negative Info	No Info	-.281(*)	.094	.008	-.506	-.056
		Positive Info	-.464(*)	.115	.000	-.740	-.188
	No Info	Negative Info	.281(*)	.094	.008	.056	.506
		Positive Info	-.183	.088	.115	-.395	.028
	Positive Info	Negative Info	.464(*)	.115	.000	.188	.740
		No Info	.183	.088	.115	-.028	.395

Based on estimated marginal means

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

a Adjustment for multiple comparisons: Bonferroni.



Parameter Estimates

Dependent Variable: mean_rating_4to14

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	2.764	.126	21.970	.000	2.517	3.011
[inst_sex=.00]	.113	.093	1.218	.224	-.069	.295
[inst_sex=1.00]	0(a)	.	.	.	.	.
[pos_neg=1.00]	-.464	.115	-4.038	.000	-.690	-.239
[pos_neg=2.00]	-.183	.088	-2.076	.038	-.356	-.010
[pos_neg=3.00]	0(a)	.	.	.	.	.
intrin	.416	.027	15.641	.000	.364	.469
[inst_sex=.00] * [pos_neg=1.00]	-.120	.139	-.862	.389	-.392	.153
[inst_sex=.00] * [pos_neg=2.00]	.220	.109	2.026	.043	.007	.433
[inst_sex=.00] * [pos_neg=3.00]	0(a)	.	.	.	.	.
[inst_sex=1.00] * [pos_neg=1.00]	0(a)	.	.	.	.	.
[inst_sex=1.00] * [pos_neg=2.00]	0(a)	.	.	.	.	.
[inst_sex=1.00] * [pos_neg=3.00]	0(a)	.	.	.	.	.

a This parameter is set to zero because it is redundant.

Comment: Above the above table reports regression results. Not covariate b = 0.416.

10b Student Ratings: APA Results

Table 10b1

Descriptive Statistics of Student Ratings by Instructor Reputation and Instructor Sex

	Reputation											
	Negative				No Info				Positive			
Sex	M	Adj. M	SD	n	M	Adj. M	SD	n	M	Adj. M	SD	n
Female	3.70	3.79	0.87	111	4.47	4.41	0.62	287	4.48	4.38	0.62	159
Male	3.60	3.80	0.68	59	4.01	4.08	0.86	235	4.32	4.26	0.54	69

Note. Grand mean = 4.20 (SD = 0.79), n = 920. Adjusted mean based upon Intrinsic Motivation = 3.60.

Table 10b2

ANCOVA Summary for Student Ratings by Instructor Reputation, Instructor Sex, and Intrinsic Motivation

Source	SS	df	MS	F
Reputation (R)	27.74	2	13.87	33.66*
Sex (S)	3.48	1	3.48	8.44*
Intrinsic Mot.	100.82	1	100.82	244.63*
R × S	4.17	2	2.09	5.06*
Error	376.26	913	0.41	

Note: R^2 = .34, adj. R^2 = .34. Homogeneity of regression and other possible two-way and three-way interactions were not significant at the .05 level. Intrinsic Motivation regression coefficient = 0.42*.

* $p < .05$

Table 10b3

Comparisons of Mean Differences in Student Ratings by Reputation and Instructor Sex

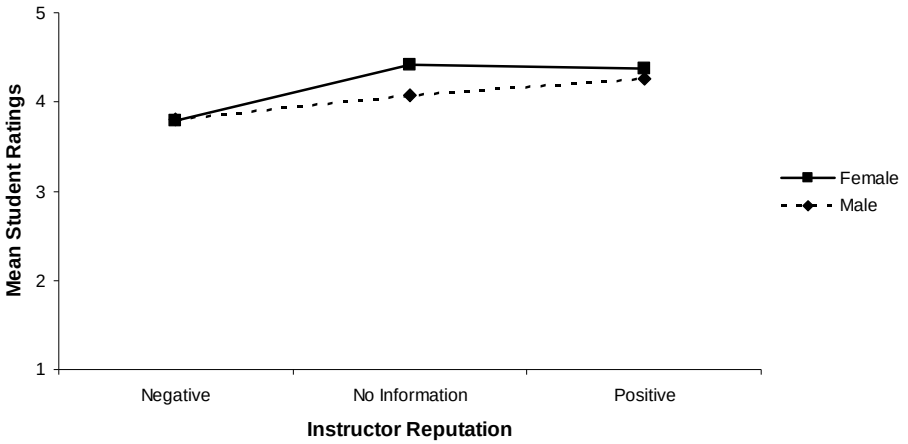
Reputation Comparison by Instructors' Sex	Estimated Mean Difference	Standard Error of Difference	Bonferroni Adjusted 95% CI
Females			
Positive vs. No Information	-0.04	0.06	-0.19, 0.12
Positive vs. Negative	0.58*	0.08	0.39, 0.78
No Information vs. Negative	0.62*	0.07	0.45, 0.79
Males			
Positive vs. No Information	0.18	0.09	-0.03, 0.40
Positive vs. Negative	0.46*	0.12	0.19, 0.74
No Information vs. Negative	0.28*	0.09	0.06, 0.51

Note. Comparisons based upon ANCOVA adjusted means controlling for Intrinsic Motivation mean of 3.60.

* $p < .05$, where p-values are adjusted using the Bonferroni method

Figure 10b1

Plot of Marginal Means of Student Ratings for Instructor Sex, Reputation, and Controlling for Student Intrinsic Motivation



Results show that student ratings are statistically related to instructor sex, instructor reputation, and student intrinsic motivation for the subject. In addition, there is a statistically significant interaction between instructor sex and reputation. As the tables and figure above shows, females tend to receive higher ratings than males, although this difference is statistically significant only when both males and females have no known reputation as instructors. In both the negative and positive reputation levels, the differences in adjusted ratings are not statistically significant. Instructors with positive reputations tend to receive higher ratings, although female instructors receive similar ratings whether their reputation is positive or not known. Lastly, intrinsic motivation is positively related to ratings – the more interest students have in the subject matter of the course, the greater will be those students' ratings.

Comment: One may also opt to report the following table of comparisons if a focus of the study is upon sex differences:

Table 10b4

Comparisons of Mean Differences in Student Ratings by Instructor Sex and Reputation

Sex Comparison by Instructors' Reputation	Estimated Mean Difference	Standard Error of Difference	Bonferroni Adjusted 95% CI
Positive Reputation			
Female vs. Male	0.11	0.09	-0.07, 0.30
No Known Reputation			
Female vs. Male	0.33*	0.06	0.22, 0.45
Negative Reputation			
Female vs. Male	-0.01	0.10	-0.21, 0.20

Note. Comparisons based upon ANCOVA adjusted means controlling for Intrinsic Motivation mean of 3.60.
* $p < .05$, where p-values are adjusted using the Bonferroni method

10c Mathematics Achievement: SPSS Results

Consistent with ANCOVA analyses, one should form first an interaction to test homogeneity of regression. Results show that there is a statistically significant interaction between Instructional Condition and Mathematics Pretest Scores. Reported below are results from SPSS.

Descriptive Statistics

Dependent Variable: math_posttest

instruction	Mean	Std. Deviation	N
Control	79.2000	7.39068	10
RPT	92.8750	4.08613	8
Test_Only	82.2222	8.49673	9
Tutor	87.4286	7.16140	7
Total	84.9118	8.59305	34

Tests of Between-Subjects Effects

Dependent Variable: math_posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1942.590(a)	7	277.513	14.602	.000
Intercept	2963.563	1	2963.563	155.931	.000
study_condition	341.008	3	113.669	5.981	.003
math_pretest	612.298	1	612.298	32.217	.000
study_condition * math_pretest	222.032	3	74.011	3.894	.020
Error	494.146	26	19.006		
Total	247577.000	34			
Corrected Total	2436.735	33			

a R Squared = .797 (Adjusted R Squared = .743)

Parameter Estimates

Dependent Variable: math_posttest

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	56.225	11.198	5.021	.000	33.206	79.243
[instruction=Control]	-8.715	13.857	-.629	.535	-37.200	19.769
[instruction=RPT]	36.279	15.129	2.398	.024	5.181	67.377
[instruction=Test_Only]	-16.073	13.720	-1.171	.252	-44.276	12.129
[instruction=Tutor]	0(a)	.	.	.	.	.
math_pretest	.685	.243	2.817	.009	.185	1.184
[instruction=Control] * math_pretest	-.018	.296	-.059	.953	-.627	.591
[instruction=RPT] * math_pretest	-.677	.320	-2.117	.044	-1.334	-.020
[instruction=Test_Only] * math_pretest	.219	.295	.742	.465	-.388	.826
[instruction=Tutor] * math_pretest	0(a)	.	.	.	.	.

a This parameter is set to zero because it is redundant.

Estimates

Dependent Variable: math_posttest

study_condition	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Control	78.906(a)	1.381	76.068	81.744
RPT	92.865(a)	1.565	89.647	96.083
Test Only	82.677(a)	1.456	79.685	85.669
Tutor Only	88.447(a)	1.687	84.979	91.915

a Covariates appearing in the model are evaluated at the following values: math_pretest = 47.0588.

Pairwise Comparisons

Dependent Variable: math_posttest

(I) study_condition	(J) study_condition	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
Control	RPT	-13.959(*)	2.087	.000	-19.919	-7.999
	Test Only	-3.771	2.006	.428	-9.500	1.957
	Tutor Only	-9.541(*)	2.180	.001	-15.766	-3.317
RPT	Control	13.959(*)	2.087	.000	7.999	19.919
	Test Only	10.188(*)	2.138	.000	4.084	16.292
	Tutor Only	4.418	2.301	.396	-2.154	10.990
Test Only	Control	3.771	2.006	.428	-1.957	9.500
	RPT	-10.188(*)	2.138	.000	-16.292	-4.084
	Tutor Only	-5.770	2.228	.093	-12.133	.592
Tutor Only	Control	9.541(*)	2.180	.001	3.317	15.766
	RPT	-4.418	2.301	.396	-10.990	2.154
	Test Only	5.770	2.228	.093	-.592	12.133

Based on estimated marginal means

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

a Adjustment for multiple comparisons: Bonferroni.

Comment: By default, if multiple comparisons are requested SPSS presents comparisons at the mean value of covariates. In the above example, the mean value of Mathematics Pretest scores is 47.058.

10c Mathematics Achievement: SPSS—Multiple Comparisons at Three Values of Mathematics Pretest Scores

To help test whether group differences exist along different ranges of the covariate, use SPSS to perform multiple comparisons at select values of mathematics pretest scores. In discussion of this approach earlier, the three points selected for the covariate were the mean, the mean + SD, and the mean – SD. Sometimes those values may not be useful given the nature of the interaction. The interaction should be plotted and examined to determine whether such values will likely help present the nature of the interaction. Below is a scatter plot with regression lines that demonstrate the interaction and the points represented by mathematics pretest scores at M, M+SD, and M–SD. Using FREQUENCIES command and requesting descriptive statistics, SPSS produces the M and SD for mathematics pretest scores:

Statistics

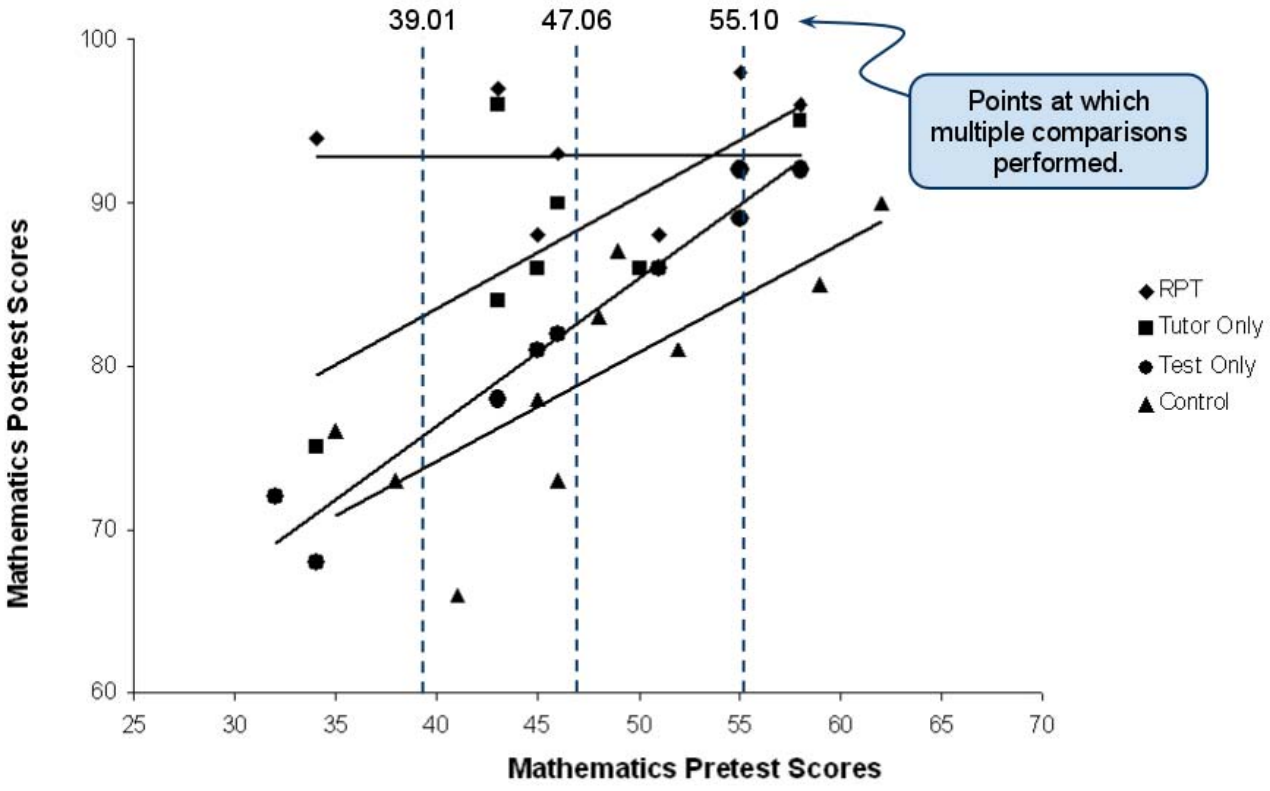
math_pretest

N	Valid	34
	Missing	0
Mean		47.0588
Std. Deviation		8.04510

The three values of Mathematics Pretest scores would be:

M = 47.0588
M+SD = 47.0588 + 8.0451 ≈ 55.10
M–SD = 47.0588 – 8.0451 ≈ 39.01

Figure 10c1
Scatter Plot of Pre- and Post-Mathematics Scores by Study Condition with Points for Multiple Comparisons Indicated



To obtain the SPSS pairwise comparisons for mathematics pretest scores of 39.01 and 55.10, use the following alterations to the command syntax:

```
UNIANOVA
  math_posttest  BY study_condition  WITH math_pretest
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /EMMEANS = TABLES(study_condition) WITH(math_pretest=MEAN) COMPARE ADJ(BONFERRONI)
  /PRINT = DESCRIPTIVE PARAMETER
  /CRITERIA = ALPHA(.05)
  /DESIGN = study_condition math_pretest math_pretest*study_condition .
```

Change the above to this:

```
UNIANOVA
  math_posttest  BY study_condition  WITH math_pretest
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /EMMEANS = TABLES(study_condition) WITH(math_pretest=MEAN) COMPARE ADJ(BONFERRONI)
  /EMMEANS = TABLES(study_condition) WITH(math_pretest=39.01) COMPARE ADJ(BONFERRONI)
  /EMMEANS = TABLES(study_condition) WITH(math_pretest=55.10) COMPARE ADJ(BONFERRONI)
  /PRINT = DESCRIPTIVE PARAMETER
  /CRITERIA = ALPHA(.05)
  /DESIGN = study_condition math_pretest math_pretest*study_condition .
```

Reported below are the additional pairwise comparison that result from the above command.

Mathematics Pretest = 39.01

Estimates

Dependent Variable: math_posttest

study_condition	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Control	73.536(a)	1.992	69.441	77.631
RPT	92.803(a)	2.483	87.699	97.907
Test Only	75.404(a)	1.925	71.446	79.361
Tutor Only	82.936(a)	2.293	78.222	87.649

a Covariates appearing in the model are evaluated at the following values: math_pretest = 39.01.

Pairwise Comparisons

Dependent Variable: math_posttest

(I) study_condition	(J) study_condition	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
Control	RPT	-19.267(*)	3.183	.000	-28.357	-10.177
	Test Only	-1.868	2.770	1.000	-9.779	6.043
	Tutor Only	-9.400(*)	3.038	.028	-18.074	-.726
RPT	Control	19.267(*)	3.183	.000	10.177	28.357
	Test Only	17.400(*)	3.142	.000	8.428	26.372
	Tutor Only	9.867(*)	3.380	.043	.216	19.519
Test Only	Control	1.868	2.770	1.000	-6.043	9.779
	RPT	-17.400(*)	3.142	.000	-26.372	-8.428
	Tutor Only	-7.532	2.994	.110	-16.082	1.018
Tutor Only	Control	9.400(*)	3.038	.028	.726	18.074
	RPT	-9.867(*)	3.380	.043	-19.519	-.216
	Test Only	7.532	2.994	.110	-1.018	16.082

Based on estimated marginal means

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

a Adjustment for multiple comparisons: Bonferroni.

Mathematics Pretest = 55.10

Estimates

Dependent Variable: math_posttest

study_condition	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Control	84.270(a)	1.886	80.393	88.148
RPT	92.927(a)	2.081	88.649	97.204
Test Only	89.944(a)	2.039	85.752	94.135
Tutor Only	93.953(a)	2.842	88.111	99.796

a Covariates appearing in the model are evaluated at the following values: math_pretest = 55.10.

Pairwise Comparisons

Dependent Variable: math_posttest

(I) study_condition	(J) study_condition	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
Control	RPT	-8.656(*)	2.808	.029	-16.676	-.637
	Test Only	-5.673	2.778	.308	-13.605	2.258
	Tutor Only	-9.683	3.411	.052	-19.423	.058
RPT	Control	8.656(*)	2.808	.029	.637	16.676
	Test Only	2.983	2.913	1.000	-5.336	11.302
	Tutor Only	-1.026	3.523	1.000	-11.085	9.032
Test Only	Control	5.673	2.778	.308	-2.258	13.605
	RPT	-2.983	2.913	1.000	-11.302	5.336
	Tutor Only	-4.009	3.498	1.000	-13.998	5.979
Tutor Only	Control	9.683	3.411	.052	-.058	19.423
	RPT	1.026	3.523	1.000	-9.032	11.085
	Test Only	4.009	3.498	1.000	-5.979	13.998

Based on estimated marginal means

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

a Adjustment for multiple comparisons: Bonferroni.

10c Mathematics Achievement: APA Presentation

Table 10c1

ANCOVA Results and Descriptive Statistics for Mathematics Achievement by Study Condition and Pre-Mathematics Scores

	Mathematics Scores			
	Observed Mean	Adjusted Mean	SD	n
RPT	92.88	varies	4.09	8
Test Only	82.22	varies	8.50	9
Tutor Only	87.43	varies	7.16	7
Control	79.20	varies	7.39	10

Source	SS	df	MS	F
Study Condition (SC)	341.01	3	113.67	5.98*
Math. Pretest (P)	612.30	1	612.30	32.22*
SC × P	222.03	3	74.01	3.89*
Error	494.15	26	19.01	

Note. $R^2 = .80$, Adj. $R^2 = .74$.

* $p < .05$

Table 10c2

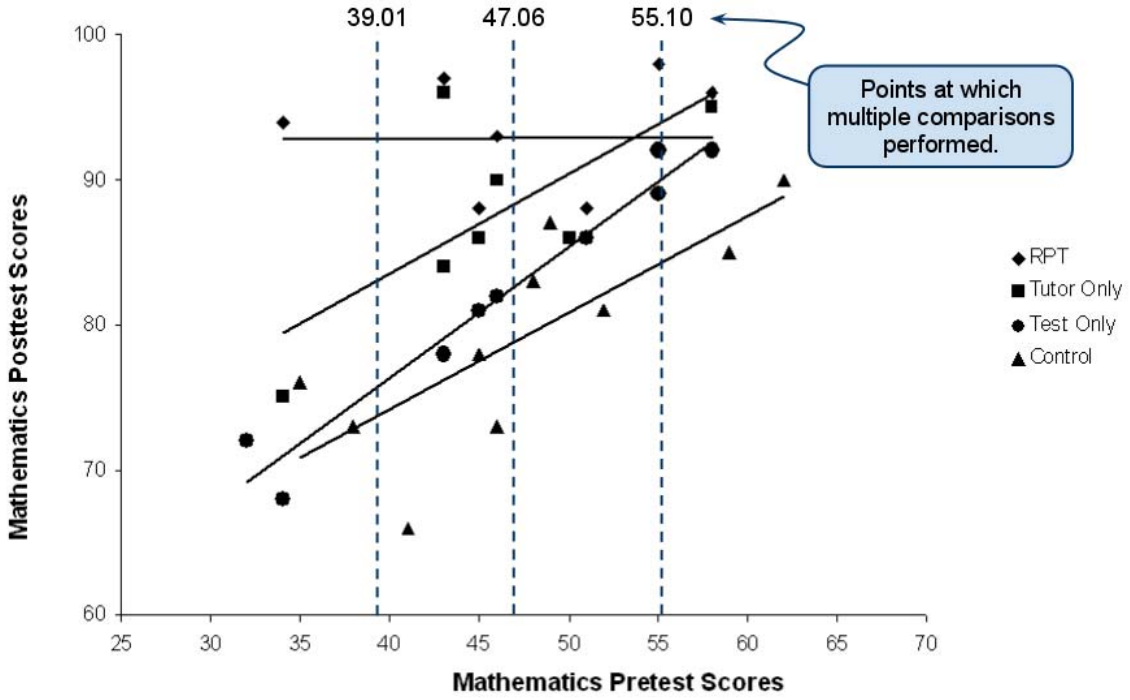
Comparisons of Mean Differences in Mathematics Achievement by Study Condition (RPT vs. others)

Achievement Comparison by Study Condition for Levels of Pretest Performance	Estimated Mean Difference	Standard Error of Difference	Bonferroni Adjusted 95% CI
Math. Pretest = 39.01			
RPT vs. Control	19.27*	3.18	10.18, 28.36
RPT vs. Test Only	17.40*	3.14	8.43, 26.37
RPT vs. Tutor Only	9.87*	3.38	0.22, 19.52
Math. Pretest = 47.06			
RPT vs. Control	13.96*	2.09	8.00, 19.92
RPT vs. Test Only	10.19*	2.14	4.08, 16.29
RPT vs. Tutor Only	4.42	2.30	-2.15, 10.99
Math. Pretest = 55.10			
RPT vs. Control	8.66*	2.81	0.64, 16.68
RPT vs. Test Only	2.98	2.91	-5.34, 11.30
RPT vs. Tutor Only	-1.03	3.52	-11.09, 9.03

Note. Comparisons based upon ANCOVA adjusted means controlling for Mathematics Pretest with the scores specified within the table.

* $p < .05$, where p-values are adjusted using the Bonferroni method

Figure 10c1 (Reposted)
Scatter Plot of Pre- and Post-Mathematics Scores by Study Condition with Points for Multiple Comparisons Indicated



ANCOVA results, in Table 10c1, show there is a statistically significant interaction between Mathematics pretest scores and Study Condition. Figure 10c1 displays this interaction. Note that the association between pretest and posttest scores is positive and similar for the Tutor Only, Test Only, and Control groups. This means that those who performed better on the pretest also tended to perform better on the posttest; similarly, those who performed at a lower level on the pretest also tended to perform at a lower level on the posttest. The RPT group, however, displayed a different pattern—the slope between pretest and posttest mathematics scores is flat which indicates that students who scored low on the pretest tended to perform on the posttest as well as students who scored high on the pretest. Several multiple comparisons were performed to help examine these scores. Since the RPT group clearly differs from the other three groups, only the RPT vs. other group comparisons are reported. For these multiple comparisons, three points along the mathematics pretest scores range were examined, the mean (47.06), the mean + SD (55.10), and the mean – SD (39.01). Table 10c2 contains results of these comparisons. When pretest scores are low (39.01), students in the RPT group achieved higher scores than their counterparts in the other three groups. For mid-level pretest scores (47.06), students in the RPT group achieved higher mathematics scores than did students in the Control or Test Only conditions, but students in the Tutor Only condition displayed similar mean performance to those in the RPT condition. Lastly, when pretest scores are high (55.10), students in the three treatment conditions appear to display similar levels of achievement and the only statistically significant difference observed was between those in RPT and Control conditions. In summary, with RPT it appears that all students were able to perform well on the posttest despite their initial level of understanding as measured by the mathematics pretest. Overall the RPT group appears to have better performance than those in the Tutor only, Test only, or Control conditions under most ranges of initial understanding of mathematics.