Tim Xeriland
CEP 933
Unit 6

ANCOVA

The hypothesis is that there will be a significant difference in geometry scores between socioeconomic groups even when controlling for motivation. The null hypothesis is that there will be no significant difference. I think the logical first step would be to run a simple ANOVA to see if there is difference between the means in the three socioeconomic status groups for geometry scores. I will not take into consideration motivation. This will be simply to see if the means have a significant difference period.

One complication is that we have uneven sample sizes. For now let's ignore that and look at the variance.

Between-Subjects Factors

		Value Label	N
SES	1.00	low	87
	2.00	medium	87
	3.00	high	47

We can see there are some differences in the standard deviations. The differences do not look too large, but it is required to look at this more closely.

Descriptive Statistics

Dependent Variable: Geom

SES	Mean	Std. Deviation	N
low	41.8666	14.56218	87
medium	52.3133	18.01723	87
high	65.6715	14.12591	47
Total	51.0417	18.19806	221

Specifically, the Levene's test tells us there is not a significant difference with our variances so our assumption of homogeneity of variances has not been violated. This is needed to run our ANCOVA.

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

Dependent Variable: Geom

F	df1	df2	Sig.
2.297	2	218	.103

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

a. Design: Intercept + SES

Tests of Between-Subjects Effects

Dependent Variable: Geom

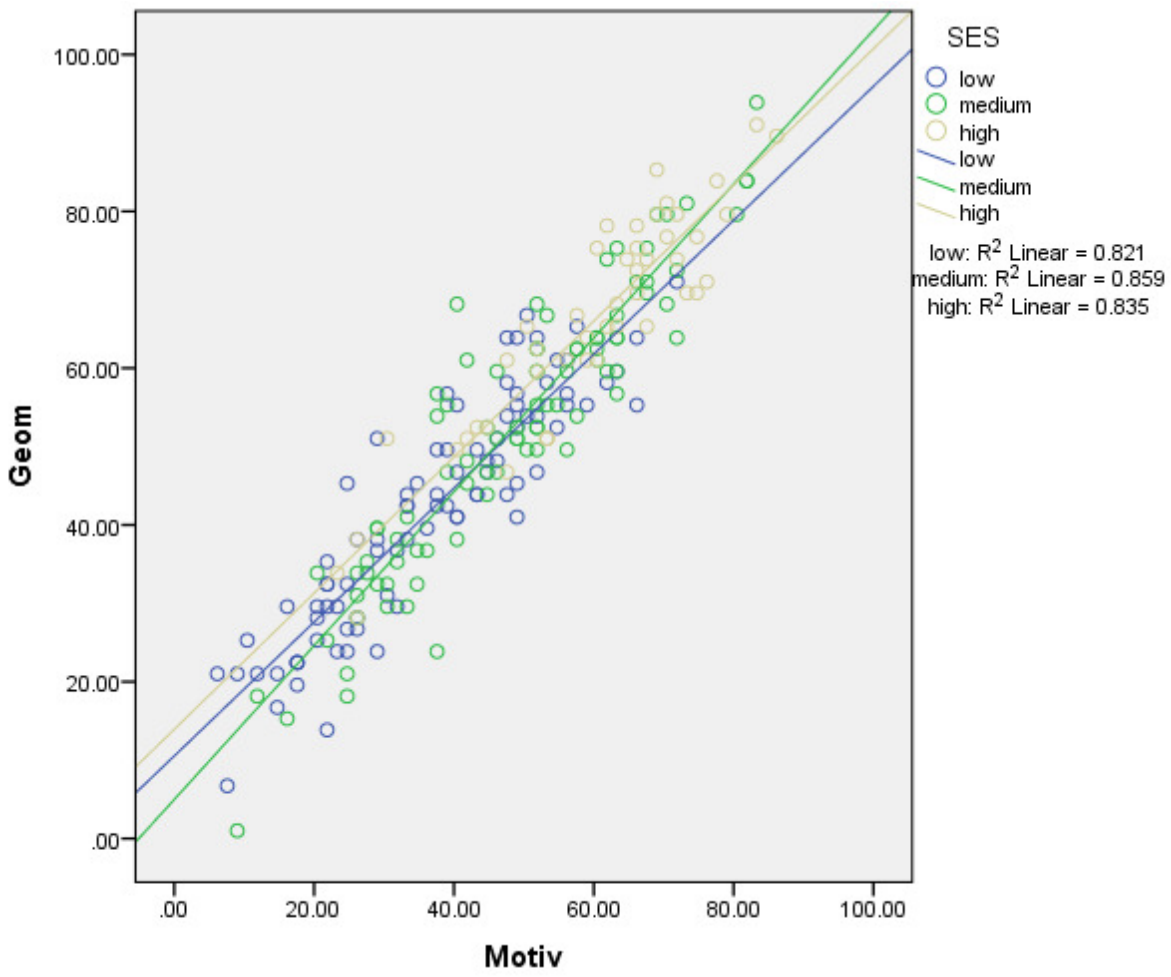
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	17524.079[a]	2	8762.040	34.520	.000	.241
Intercept	577259.761	1	577259.761	2274.270	.000	.913
SES	17524.079	2	8762.040	34.520	.000	.241
Error	55333.185	218	253.822			
Total	648618.068	221				
Corrected Total	72857.264	220				

a. R Squared = .241 (Adjusted R Squared = .234)

Looking at our variable of socioeconomic status (SES) we can see that it has an F-value of 34.5 and is significant for $p < .001$. Also, our eta squared tells us that SES explains approximately 24% of the variability in geometry scores.

After doing an ANOVA we see geometry score does appear to differ in a statistically significant way. The question becomes what does this really mean? For example, maybe what we are really seeing is that students that come from a higher SES are just more motivated. Those that come from higher SES have greater opportunity for schooling after high school and this may motivate them to learn the material and therefore they do better.

By doing a ANCOVA we can control for motivation and see if we still have an effect. The first step we want to take is to test the homogeneity of variance. The covariate has to have the same correlation with geometry scores for all three of our SES groups.



We can start by seeing if the covariate has the same slope as the factors. Above is a scatterplot with the lines representing the slopes. While they have the same general slope they are not parallel and the lines cross each other. Either way this is not a definitive test, but more to give us a general idea. To give us a definitive test we need to run the model. Once the model is run you first check to see that it meets the assumption of homogeneity of variance. Below is Levene's test which was not found to be significant and therefore it is safe to assume homogeneity of variance.

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

Dependent Variable: Geom

F	df1	df2	Sig.
.302	2	218	.739

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

a. Design: Intercept + Motiv + SES + SES * Motiv

Now we move on to the full source table which gives the sums of squares. This is checked to see if we have a statistically significant interaction. For the interaction of SES and Motivation we find a P value of .08 which is not statistically significant. This does tells us definitively that the slope for motivation is not different across the classroom groups. In other words, motivation is a good covariate.

Tests of Between-Subjects Effects

Dependent Variable: Geom

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	64132.819[a]	5	12826.564	316.090	.000	.880
Intercept	1572.522	1	1572.522	38.752	.000	.153
Motiv	39039.067	1	39039.067	962.055	.000	.817
SES	248.899	2	124.449	3.067	.049	.028
SES * Motiv	207.829	2	103.915	2.561	.080	.023
Error	8724.445	215	40.579			
Total	648618.068	221				
Corrected Total	72857.264	220				

a. R Squared = .880 (Adjusted R Squared = .877)

Because the interaction is irrelevant we can remove it from the model and run the ANCOVA model.

Between-Subjects Factors

		Value Label	N
SES	1.00	low	87
	2.00	medium	87
	3.00	high	47

Descriptive Statistics

Dependent Variable: Geom

SES	Mean	Std. Deviation	N
low	41.8666	14.56218	87
medium	52.3133	18.01723	87
high	65.6715	14.12591	47
Total	51.0417	18.19806	221

Looking at the new Levene's Test we see it is not statistically significant and therefore we can assume homogeneity of variance across the groups.

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

Dependent Variable: Geom

F	df1	df2	Sig.
.228	2	218	.796

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

a. Design: Intercept + Motiv + SES

Looking at the source table that was produced by SPSS we can see that both the factor (SES) and the covariat (Motiv) are statistically significant.

Tests of Between-Subjects Effects

Dependent Variable: Geom

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	63924.990[a]	3	21308.330	517.663	.000	.877
Intercept	1845.803	1	1845.803	44.842	.000	.171
Motiv	46400.911	1	46400.911	1127.260	.000	.839
SES	271.238	2	135.619	3.295	.039	.029
Error	8932.274	217	41.163			
Total	648618.068	221				
Corrected Total	72857.264	220				

a. R Squared = .877 (Adjusted R Squared = .876)

Using the estimated marginal means will get us the mean value for motivation, which is 46.0992.

1. Grand Mean

Dependent Variable: Geom

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
51.394[a]	.453	50.500	52.288

a. Covariates appearing in the model are evaluated at the following values: Motiv = 46.0992.

2. SES

Dependent Variable: Geom

SES	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
low	50.472[a]	.734	49.025	51.918
medium	50.369[a]	.690	49.009	51.730
high	53.341[a]	1.005	51.360	55.323

a. Covariates appearing in the model are evaluated at the following values: Motiv = 46.0992.

The table below gives us the parameter estimates. We can see the slope for motivation is .914 and we know that is statistically significant. Also, we have the slopes for SES 1 and 2. This is the same result we would find if we ran a regression.

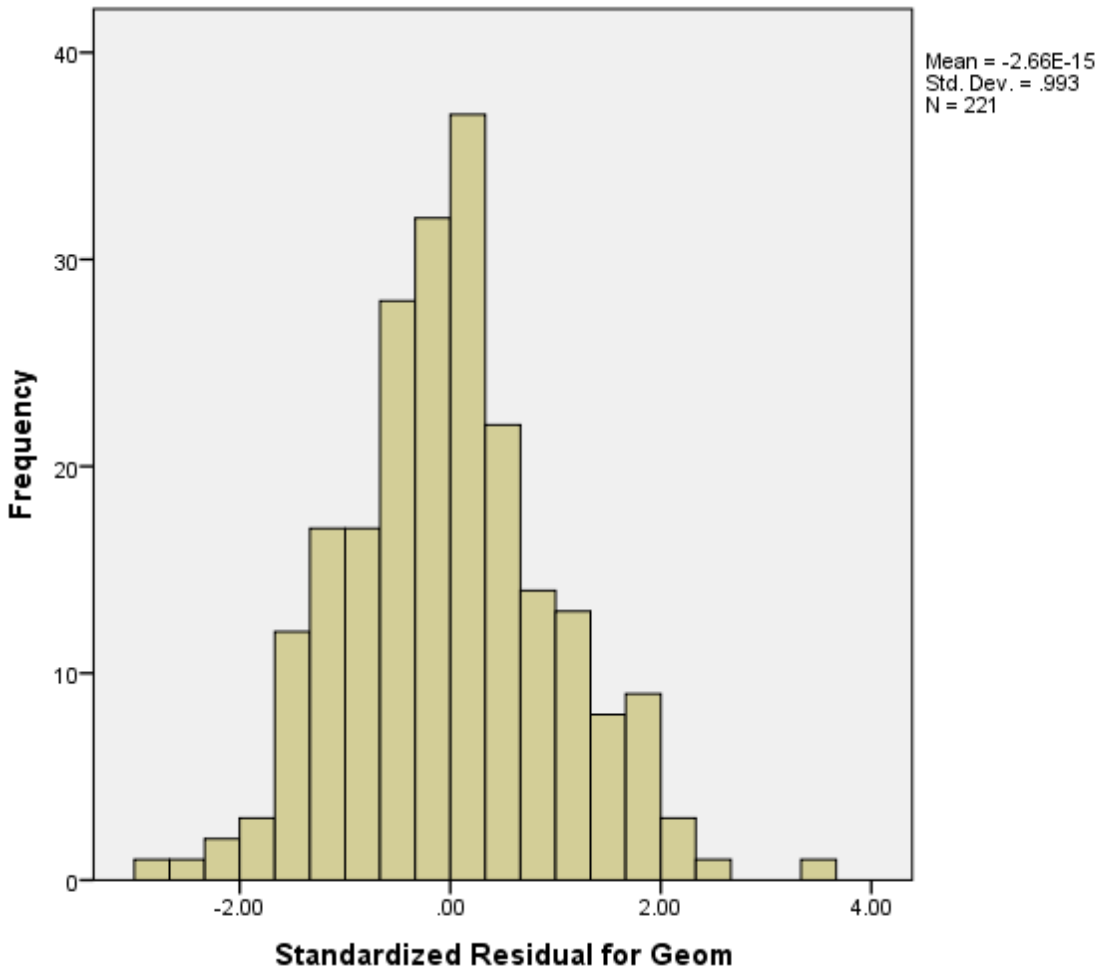
Parameter Estimates

Dependent Variable: Geom

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	11.227	1.872	5.996	.000	7.536	14.917
Motiv	.914	.027	33.575	.000	.860	.967
[SES=1.00]	-2.870	1.318	-2.177	.031	-5.468	-.272
[SES=2.00]	-2.972	1.202	-2.473	.014	-5.341	-.603
[SES=3.00]	0[a]	.	.	.	.	.

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

The last step when doing a ANCOVA analysis is to check the residuals for normality. We can do this by running a histogram of the residuals. Viewing the histogram below suggests that the assumption of normality of residuals is reasonable.


Mean = -2.66E-15
Std. Dev. = .993
N = 221
Frequency
40
30
20
10
0
-2.00
.00
2.00
4.00
Standardized Residual for Geom