ANCOVA with Regression Homogeneity

The purpose of the study was to compare the effectiveness of two different treatments in two populations. Both treatments have been repeatedly shown to work better than a no-treatment control. In practice, the different populations were most commonly given different treatments, and ours was the first study to compare the two treatments within both of the populations. Frankly we were "rooting" for Tx1, because it was less expensive and the treatment protocol had better initial acceptance ratings and higher completion rates.

Factorial ANOVA of DV = performance

SPSS code:

```
UNIANOVA performance BY population treatment
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(population)
  /EMMEANS=TABLES(treatment)
  /EMMEANS=TABLES(population*treatment) compare (treatment)
  /EMMEANS=TABLES(treatment*population) compare (population)
  /PRINT=DESCRIPTIVE.
```

The emeans tables aren't really necessary, but I wanted you to see how the code for these changes across the different analyses.

I did not "compare" for the main effects, because these each have 2 conditions.

Selected Output (I left out the emeans tables for main effects):

Descriptive Statistics

Dependent Variable:performance

population	treatment	Mean	Std. Deviation	N
pop1	tx1	100.4675	18.82079	16
	tx2	110.0657	24.96689	15
	Total	105.1118	22.17612	31
pop2	tx1	90.6118	27.21537	14
	tx2	79.5115	32.87875	15
	Total	84.8702	30.26975	29
Total	tx1	95.8681	23.24339	30
	tx2	94.7886	32.62256	30
	Total	95.3284	28.08812	60

Tests of Between-Subjects Effects

Dependent Variable:performance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7744.463[a]	3	2581.488	3.726	.016
Intercept	542162.345	1	542162.345	782.439	.000
population	6109.974	1	6109.974	8.818	.004
treatment	8.441	1	8.441	.012	.913
population * treatment	1603.030	1	1603.030	2.313	.134
Error	38803.136	56	692.913		
Total	591797.566	60			
Corrected Total	46547.599	59			

a. R Squared = .166 (Adjusted R Squared = .122)

These were very nice results, at first blush (we call that "foreshadowing"). The significant population difference made sense given known differences between the populations. The lack of a main effect for treatment "gave license" to picking between the treatments on bases other than differential effectiveness (e.g., treatment acceptability and completion). Finally, the lack of an interaction means that parallel decisions can be made for the two populations.

Pairwise Comparisons

Dependent Variable:performance

population	(I) treatment	(J) treatment	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
						Lower Bound	Upper Bound
pop1	tx1	tx2	-9.598	9.461	.315	-28.550	9.353
	tx2	tx1	9.598	9.461	.315	-9.353	28.550
pop2	tx1	tx2	11.100	9.782	.261	-8.495	30.696
	tx2	tx1	-11.100	9.782	.261	-30.696	8.495

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Pairwise Comparisons

Dependent Variable:performance

treatment	(I) population	(J) population	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
						Lower Bound	Upper Bound
tx1	pop1	pop2	9.856	9.633	.311	-9.442	29.154
	pop2	pop1	-9.856	9.633	.311	-29.154	9.442
tx2	pop1	pop2	30.554*	9.612	.002	11.299	49.809
	pop2	pop1	-30.554*	9.612	.002	-49.809	-11.299

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).
*. The mean difference is significant at the .050 level.

Factorial ANOVA of possible covariate = Motivation

No matter how pretty or desirable the data pattern and significance tests, there always are (or should be) nagging doubts about the validity of results from an ANOVA run on non-experimental data. One of the reasons we were rooting for Tx1 was that less motivated people tended to reject and/or drop out of Tx2, but would accept and complete Tx1. Since self-assignment to treatment was used in this study, it occurred to us to check motivation was playing a part in these results. Here are the results from that analysis → same factorial design, but asking for group differences on "treatment motivation".

These were obtained using the SPSS code from the last page, but with Motivation as the DV

Descriptive Statistics

Dependent Variable:motivation

population	treatment	Mean	Std. Deviation	N
pop1	tx1	29.8008	4.76610	16
	tx2	29.3136	3.99955	15
	Total	29.5650	4.34559	31
pop2	tx1	8.5099	5.41027	14
	tx2	49.9258	4.44038	15
	Total	29.9319	21.61144	29
Total	tx1	19.8650	11.89886	30
	tx2	39.6197	11.27472	30
	Total	29.7423	15.20819	60

Tests of Between-Subjects Effects

Dependent Variable:motivation

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12424.814[a]	3	4141.605	189.912	.000
Intercept	51702.101	1	51702.101	2370.787	.000
population	1.723	1	1.723	.079	.780
treatment	6267.853	1	6267.853	287.411	.000
population * treatment	6569.837	1	6569.837	301.258	.000
Error	1221.247	56	21.808		
Total	66722.494	60			
Corrected Total	13646.062	59			

a. R Squared = .911 (Adjusted R Squared = .906)

Here we find some group differences on Motivation, suggesting that the data pattern from the ANOVA may be spurious.

Pairwise Comparisons

Dependent Variable:motivation

population	(I) treatment	(J) treatment	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
						Lower Bound	Upper Bound
pop1	tx1	tx2	.487	1.678	.773	-2.875	3.849
	tx2	tx1	-.487	1.678	.773	-3.849	2.875
pop2	tx1	tx2	-41.416*	1.735	.000	-44.892	-37.939
	tx2	tx1	41.416*	1.735	.000	37.939	44.892

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).
*. The mean difference is significant at the .050 level.

Pairwise Comparisons

Dependent Variable:motivation

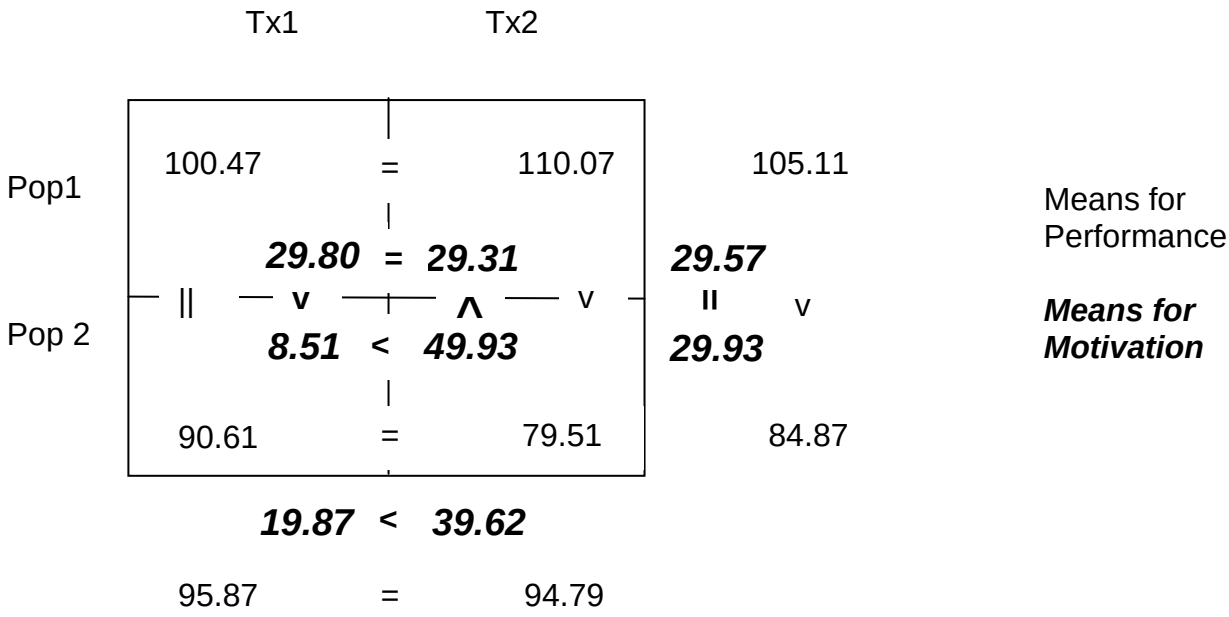
treatment	(I) population	(J) population	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
						Lower Bound	Upper Bound
tx1	pop1	pop2	21.291*	1.709	.000	17.867	24.714
	pop2	pop1	-21.291*	1.709	.000	-24.714	-17.867
tx2	pop1	pop2	-20.612*	1.705	.000	-24.028	-17.196
	pop2	pop1	20.612*	1.705	.000	17.196	24.028

Based on estimated marginal means

*. The mean difference is significant at the .050 level.
a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

These results show that there are group differences on motivation, which could be influencing the observed group differences on the DV! Let's gather this information together and take a look at the multivariate data patterns and consider the utility of performing a factorial ANCOVA on these data.

Summarizing the Covariate Pattern & Anticipating the ANCOVA Results!



Given the patterns of group differences for Performance and for the confound/covariate Motivation, what patters do we expect when use a factorial ANCOVA to look at the relationship between Treatment, Population and Performance, when controlling for Motivation?

Remember that we expect a smaller error term, and, so, more powerful comparisons from the ANCOVA than from the ANOVA.

- SE of Tx for Pop1 – no confounding, so expect correct and uncorrected comparisons to be the same
- SE of Tx for Pop 2 – Tx2 has higher mean motivation, so expect Tx1 will have higher mean than Tx2 after correction

- SE of Pop for Tx1 – large population difference on motivation, should “offset” or maybe even “flip” the direction of the corrected mean difference, relative to the uncorrected mean difference
- SE for Pop forf Tx2 -- large population difference on motivation, should “offset” or maybe even “flip” the direction of the corrected mean difference, relative to the uncorrected mean difference

- Interaction -- Taking the expected patterns of the corrected simple effects, we expect to find a significant interaction from the factorial ANCOVA

- ME of Pop – no confounding, so expect corrected and uncorrected comparisons to be the same

- ME of Tx -- – Tx2 has higher mean motivation, so expect Tx1 will have higher mean than Tx2 after correction

Factorial ANCOVA – testing homogeneity of regression slope assumption

ANCOVA allows us to make comparisons between groups while holding participants "constant at the covariate mean" and, by inference, to learn about group differences for participants with other values of the covariate. This assumes that the slope of the relationship between the covariate and the dependent variable is the same for all groups. Said differently, the inference that group difference comparisons controlling the covariate at its mean tell us about group differences for other values of the covariate assumes that there are no interactions between the covariate and the groping variables.

It is simple enough to test this assumption! Simply include all the interactions between the grouping variables (and their interaction) and the covariate.

```
UNIANOVA performance BY population treatment WITH motivation
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /DESIGN=population treatment population*treatment
          motivation
          motivation*population motivation*treatment
          motivation*population*treatment.
```

Support for the homogeneity of regression slope assumption is gained with **none** of these interactions are significant (with all the usual concerns about significance tests, statistical power, etc.).

Tests of Between-Subjects Effects

Dependent Variable:performance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11596.961[a]	7	1656.709	2.465	.029
Intercept	2293.619	1	2293.619	3.412	.070
population	733.050	1	733.050	1.091	.301
treatment	511.223	1	511.223	.761	.387
population * treatment	796.215	1	796.215	1.185	.281
motivation	3095.490	1	3095.490	4.606	.037
population * motivation	414.446	1	414.446	.617	.436
treatment * motivation	.614	1	.614	.001	.976
population * treatment * motivation	.023	1	.023	.000	.995
Error	34950.638	52	672.128		
Total	591797.566	60			
Corrected Total	46547.599	59			

a. R Squared = .249 (Adjusted R Squared = .148)

As you can see, none of the motivation interactions were significant, tell us that the slope of the regression line for motivation and performance are the same for Pop1 vs. Pop2, Tx2 vs. Tx2, and for the four conditions of their interaction.

In general, when the homogeneity of regression slope assumption is supported, the factorial ANCOVA is (re)run excluding the nonsignifiicant covariate interactions.

If the homogeneity of regression slope assumption is not supported – when one or more of the interactions are significant -- then further analysis should include those interactions. We will not pursue these analyses in this class, but will refer you to models that include interactions between categorical and quantiatitative variables in Psyc942 and the psyc930 GLM module.

Factorial ANCOVA

```
UNIANOVA performance BY population treatment WITH motivation
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(population)
                  WITH(motivation=MEAN)

  /EMMEANS=TABLES(treatment)
                  WITH(motivation=MEAN)

  /EMMEANS=TABLES(population*treatment)
                  WITH(motivation=MEAN) compare (treatment)

/EMMEANS=TABLES(treatment*population)
                  WITH(motivation=MEAN) compare (population)

  /PRINT=DESCRIPTIVE.
```

← identified DV, IVs & Covariate

← gets corrected Population marginal means (k=2, so no comparison needed)

← gets corrected Treatment marginal means (k=2, so no comparison needed)

← gets corrected cell means (k=4) (SE of Tx for each Population)

← gets corrected cell means (k=4) (SE of Pop for each Treatment)

Tests of Between-Subjects Effects

Dependent Variable:performance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11166.527[a]	4	2791.632	4.340	.004
Intercept	2919.484	1	2919.484	4.538	.038
motivation	3422.063	1	3422.063	5.320	.025
population	5763.139	1	5763.139	8.959	.004
treatment	2990.978	1	2990.978	4.649	.035
population * treatment	4839.976	1	4839.976	7.524	.008
Error	35381.072	55	643.292		
Total	591797.566	60			
Corrected Total	46547.599	59			

a. R Squared = .240 (Adjusted R Squared = .185)

Motivation is related to Performance, after controlling for the main and interaction effects.

There is a Population effect after controlling for motivation and the other design effects.

There is a Treatment effect after controlling for motivation and the other design effects.

There is an interaction of Population & Motivation after controlling for Motivation and the related main effects.

Estimates

Dependent Variable:performance

population	treatment	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
pop1	tx1	100.370[a]	6.341	87.662	113.077
	tx2	110.783[a]	6.556	97.645	123.922
pop2	tx1	126.154[a]	16.835	92.416	159.892
	tx2	45.725[a]	16.046	13.569	77.882

a. Covariates appearing in the model are evaluated at the following values: motivation = 29.7423.

1. population

Dependent Variable:performance

population	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
pop1	105.577[a]	4.560	96.439	114.714
pop2	85.940[a]	4.728	76.465	95.415

a. Covariates appearing in the model are evaluated at the following values: motivation = 29.7423.

2. treatment

Dependent Variable:performance

treatment	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
tx1	113.262[a]	8.977	95.272	131.251
tx2	78.254[a]	8.534	61.151	95.358

a. Covariates appearing in the model are evaluated at the following values: motivation = 29.7423.

Pairwise Comparisons

Dependent Variable:performance

population	(I) treatment	(J) treatment	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
						Lower Bound	Upper Bound
pop1	tx1	tx2	-10.414	9.122	.259	-28.695	7.868
	tx2	tx1	10.414	9.122	.259	-7.868	28.695
pop2	tx1	tx2	80.428*	31.502	.013	17.298	143.559
	tx2	tx1	-80.428*	31.502	.013	-143.559	-17.298

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).
*. The mean difference is significant at the .050 level.

As expected there is no Tx SE for Pop1.

For Pop2, Tx1 outperforms Tx2, also as expected.

Unlike the ANOVA, the ANCOVA shows a interaction – the treatments are not similarly comparable for Pop1 & Pop2!

Pairwise Comparisons

Dependent Variable:performance

treatment	(I) population	(J) population	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
						Lower Bound	Upper Bound
tx1	pop1	pop2	-25.784	18.026	.158	-61.909	10.341
	pop2	pop1	25.784	18.026	.158	-10.341	61.909
tx2	pop1	pop2	65.058*	17.595	.001	29.798	100.318
	pop2	pop1	-65.058*	17.595	.001	-100.318	-29.798

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).
*. The mean difference is significant at the .050 level.

As expected, when using Tx2, Pop1 out-performed Pop 2.

However, contrary to our expectation, there was no Population effect when Tx1 was used (though the effect size of r=.45 suggests this null may be a power problem).