ANALYSIS OF COVARIANCE PRIMER

In the normal experimental design, Ss are randomly assigned to different treatment levels ... and then we might perform a simple ONE factor ANOVA to test the null hypothesis of no treatment effects in the overall populations. Consider the simplest of these “experiments”, a two group study where 20S have been randomly assigned (n=10 to each group) to two conditions: Experimental and Control. In the data table below, YE means the dependent variable data (% correct scores on a science test) for the Experimental group and YC is the comparable data for the Control group. With the DESC output, you will note that the mean E is 68.1 and the mean C is 61. Then question here of course is whether we would, after our statistical test, reject the null hypothesis.

The first thing we would do, if we were going to apply ANOVA to these data, would be to partition the sums of squares: SS(T) = SS(BG) + SS(WG). To do this, we would pile all the data from YE and YC into a combo column .. TOTY ... and find the SS around the mean of that column (we would subtract 64.55, the grand mean, from each score, square the deviation, then add up the squared deviations0. This would be our SS(T). Then we would do the same thing WITHIN each group ... square the deviations around EACH group mean, add them up, then sum across both groups. This would be our SS(WG). Finally, we would find the SS(BG) by subtracting the grand mean of 64.55 from each group mean (68.1 - 64.55 ... and 61 - 64.55), square these deviations, weigh each squared deviation by n, then sum across the two values. These values could easily be found by doing the AOVO ... in Minitab.

Row	YE	YC	TOTY
1	78	64	78
2	74	44	74
3	67	50	67
4	78	84	78
5	58	45	58
6	58	71	58
7	83	71	83
8	74	45	74
9	46	68	46
10	65	68	65
11			64
12			44
13			50
14			84
15			45
16			71
17			71
18			45
19			68
20			68

MTB > desc c2 c4 c6

Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev	SE Mean
YE	10	68.10	70.50	69.00	11.50	3.64
YC	10	61.00	66.00	60.25	13.98	4.42
TOTY	20	64.55	67.50	64.61	12.98	2.90

```
MTB > aovo c2 c4
```

One-way Analysis of Variance

```
Analysis of Variance
Source     DF        SS        MS        F        P
Factor      1       252       252     1.54    0.231
Error      18      2949       164
Total      19      3201
                                  Individual 95% CIs For Mean
                                  Based on Pooled StDev
Level       N      Mean     StDev  ------+---------+---------+---------+
YE         10     68.10     11.50            (-----------*-----------)
YC         10     61.00     13.98   (-----------*-----------)
                                  ------+---------+---------+---------+
Pooled StDev =    12.80               56.0      63.0      70.0      77.0
MTB >
```

By doing this, we find our overall F test statistic to be 1.54 ... and given the p value of more than .05, we would RETAIN the null hypothesis. Our conclusion here is that there is not enough evidence to suggest that the experimental treatment is any better than the control condition. This makes us MAD of course but, such is life.

BUT WAIT A MINUTE ... !!

In the typical experimental design, the WITHIN GROUP VARIANCE is considered to be error. We would like to think that the treatment E and C would impact on every S within the E and C groups, the same. But of course, some of the Ss in the E and some of the Ss in the C, do better or poorer on the science test than other Ss. Why?

Well, maybe it is possible that the randomization into the two groups put within each group, Ss who VARY WITH RESPECT TO SOME OTHER FACTOR ... like ability or intelligence (IQ). If that is the case, then some of the within group variability in the E group ... and in the C group ... could be due to varying IQ of Ss. Thus, perhaps it is possible that SOME of the variability is due to IQ and therefore is understandable and predictable ... and hence, should NOT be considered error in the usual sense. If we can EXPLAIN the variability, then we account for it AND, it is not error any more.

In the experiment above, what if we had gathered IQ data on all 20 Ss PRIOR TO THE STUDY. We would do it this way since, we would not want to have someone say that the treatments themselves impacted on the IQ scores that we have ... so our assessment of IQ is made independent and collected before randomly assigning Ss to E and C, and before doing the experiment and gathering the science test data. Look at the expanded data table below. (The data are on the next page.)

Notice that the mean IQE and for IQC ... are in the same ballpark ... a few points different, but .. with a small sample like we have, this would be expected and would not be considered unusual.

Here is the important question at this stage? Does IQ predict science test scores? One way to find this out would be to stack all the science data in a column and the associated IQ scores in a column, and then do the correlation ... it is ...

```
MTB > corr c5 c6
```

Correlations (Pearson)

```
Correlation of TOTIQ and TOTY = 0.694, P-Value = 0.001
```

Row	IQE	YE	IQC	YC	TOTIQ	TOTY
1	106	78	101	64	106	78
2	100	74	98	44	100	74
3	94	67	96	50	94	67
4	109	78	110	84	109	78
5	93	58	94	45	93	58
6	97	58	105	71	97	58
7	107	83	102	71	107	83
8	100	74	101	45	100	74
9	87	46	104	68	87	46
10	96	65	113	68	96	65
11					101	64
12					98	44
13					96	50
14					110	84
15					94	45
16					105	71
17					102	71
18					101	45
19					104	68
20					113	68

```
MTB > desc c1-c6
```

Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev	SE Mean
IQE	10	98.90	98.50	99.12	6.94	2.19
YE	10	68.10	70.50	69.00	11.50	3.64
IQC	10	102.40	101.50	102.13	5.91	1.87
YC	10	61.00	66.00	60.25	13.98	4.42
TOTIQ	20	100.65	100.50	100.72	6.52	1.46
TOTY	20	64.55	67.50	64.61	12.98	2.90

If the r value is .694, then what percent of the science test score variance is due to the predictor of IQ? Why, that would be r square. And, that is in this case:

```
MTB > let k1=.694**2
MTB > prin k1
```

Data Display

```
K1    0.481
```

Thus, about 48% of the variance on science test scores can be accounted for by the IQ variable ... or, COVARIATE.

Well, when we did the partitioning of the sums of squares based on the science test scores, we found that the SS(T) was: (see anova table above)

SS(T) = 3201

But, we also saw that a little bit more than 48% of the science test scores was due to IQ ... and therefore explainable BASED on variations IN IQ. What would be leftover from the SS(T) of 3201 ...

3201 * .481 = 1539.7

So, 1539.7 of the SS(T) is due to IQ, and therefore ...

3201 - 1539.7 = SS(T) - SS(due to IQ) = 1661.3

Thus, of the overall SS(T) which was 3201 ... we find that AFTER YOU REMOVE THE PART OF IT THAT IS PREDICTABLE based on IQ, you still have 1661.3 of the SS(T) that is not accounted for.

With our ADJUSTED SS(T) ... we have to start all over again partitioning it into a SS(BG) part and a SS(WG) part.

NOTE: Another way to find the adjusted SS(T) is to use IQ to predict science scores, get the RESIDUAL VALUES (these are the science values made independent of IQ) ... and find the SS for these residuals.

The regression equation is
TOTY or science scores = - 74.4 + 1.38 TOTIQ

Predictor	Coef	StDev	T	P
Constant	-74.35	34.07	-2.18	0.043
TOTIQ	1.3801	0.3378	4.09	0.001

S = 9.606 R-Sq = 48.1% R-Sq(adj) = 45.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1539.9	1539.9	16.69	0.001
Residual Error	18	1661.0	92.3		
Total	19	3201.0			

Notice that the SS for regression, which is DUE to IQ is 1539.9 and the SS for residual error ... = 1661. The 1661 is the part of the SS(T) that is NOT explained by IQ ... and again is called our ADJUSTED SS(T).

==========

ADJUSTING THE SS(WG)

To get our adjusted SS(T), we used the combo data and the correlation between IQ and science scores, found r square, then reduced the SS(T) by the r squared value.
But, when we are ready to adjust the SS(WG) ... we have to go within each of the groups and do some work. Remember, in the regular ANOVA, we would find the squared deviations around each group mean ... then add them up for both groups. For our example here, out original SS(WG) turned out to be 2949. How will we adjust this? Since the SS(T) was adjusted by the r square between IQ and science scores for the OVERALL group, we must do a similar thing and make an adjustment based on the r square between IQ and science scores, BUT WITHIN EACH GROUP SEPARATELY. Here is how we can do that ... though the method shown will not be totally self evident.

First, let's see what the r values are between IQ and science scores within each group.

Correlations (Pearson)

Correlation of IQE and YE = 0.919, P-Value = 0.000

MTB > corr c3 c4

Correlations (Pearson)

Correlation of IQC and YC = 0.780, P-Value = 0.008

Thus, within the E group, that r is .919 and, within the C group that r = .78. The r squares respectively would be ... .919 squared = 0.845 and .78 squared = .608. But, when we found the original SS(WG), we simply added the two SS values together, one for each group. But, now we are using correlation types of values, if we add them together, we could get a value of larger than 1 and, that does not make any sense in the context of correlations. Another way would be to think about averaging the two r values .... .919 and .78 ... and then using that average as sort of a ballpark figure for the within group correlation. If we did that, we would get about (.919 + .78) / 2 = .85. But, there is a computational formula (see in Reference at end of handout) that gives us a POOLED VALUE for the within group correlation between IQ and science scores and that value works out to be about 0.835, which is slightly less than the average of .919 and .78 (= .85).

Thus, we will use the pooled within groups r between IQ and science scores as .835 .... and therefore, that proportion of variance will be r square or
.697

Therefore, our SS(WG) was originally 2949 but, .697 or almost 70% of this is due to the covariation of IQ with science scores. So,

So far, our ADJ SS(T) = 1661.3
our ADJ SS(WG) = 893.55

and therefore by subtraction (yes, I know this is the chicken way) ...

ADJ SS(BG) = 1661.3 - 893.55 = 767.75

Now this gives us a new set of SS values to work with and, therefore, that will impact on our ANOVA summary table ... as follows.

ANOVA SUMMARY TABLE AFTER REMOVING COVARIATE, IQ

Source	df	SS	MS	F	P
BG	1	767.75	767.75	767.75/52.56 = 14.61	< .001
WG	17*	893.55	52.56		

TOT ... * for adjusted SS(WG), we lose 1 df since we are using a covariate

COMPARISON OF ORIGINAL ANOVA AND, THE ANCOVA ... WITH IQ AS THE COVARIATE

The original summary table was:

```
Analysis of Variance
Source     DF        SS        MS        F        P
Factor      1       252       252     1.54    0.231
Error      18      2949       164
Total      19      3201
```

And the adjusted table using the covariate is given above. Notice the difference. In the original table, the F ratio did NOT allow us to reject the null. But, after removing the covariate IQ from the SS(T) ... and adjusting the SS(T), SS(BG), and SS(WG), the summary table DOES allow the rejection of the null hypothesis. Relatively speaking, the correlation between IQ and science scores has lowered the error SS and, has raised the treatment SS. In a sense, our ANCOVA has controlled for IQ within the data ... in a sense, made it as though each S had the same identical IQ.

MINITAB ANCOVA ROUTINE

To do ANCOVA in Minitab, you have to use the General Linear Model ... pull down STAT, click on ANOVA, and then slide down to GLM. To use this, you will have to stack the data for IQ and science scores into columns like c5 and c6 ... but also have a group codes column (like 1s and 2s) in c7. When you get to the dialog box in Minitab for GLM, you will click on c6 for the RESPONSES, and highlight c7 for the MODEL, and then click on COVARIATES and click on c5 since IQ is the covariate. Then do your Oks ... to get the analysis done. What you will see is:

General Linear Model

```
Factor    Type Levels Values
Group     fixed     2 1 2

Analysis of Variance for TOTY, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
TOTIQ       1     1539.9     2057.9     2057.9   39.26  0.000
Group       1      770.0      770.0      770.0   14.69  0.001
Error      17      891.0      891.0       52.4
Total      19     3200.9

Term          Coef     StDev        T      P
Constant   -102.47     26.70    -3.84  0.001
TOTIQ       1.6594    0.2648     6.27  0.000
```

Now, the Adj SS values we got more or less by hand were a wee different than in the table but, close enough for government work! The BG factor in this table is called Group ... and the F ratio is 14.69 ... and with a p value of .001, we reject the null. Don't worry about the other stuff. The main thing is that the F ratio originally did not allow us to reject the null but, partialling out the IQ covariate now does allow us to reject the null.

ADJUSTING THE DEPENDENT VARIABLE MEANS

If you think about it, what we have done is to adjust our ANOVA to take into account the predictability of the criterion variable (science scores) using our covariate IQ. We found that since IQ correlated with science scores, we were able to REDUCE the within groups error and thus, make our F test more sensitive to REAL treatment effects. Since our F ratio was larger, it is as though we had started off with larger differences in means, than we first observed ... which is like having a larger numerator term in the F ratio. Look at the means again as shown below.

Variable	N	Mean
IQE	10	98.90
YE	10	68.10
IQC	10	102.40
YC	10	61.00
TOTIQ	20	100.65
TOTY	20	64.55

Now, you see that the E mean on science was 68.1 while the mean of C was 61. But also notice that the means on the IQ measure, favored the control group. Thus, JUST DUE TO RANDOM ASSIGNMENT, the control group appears to be a bit smarter than the experimental group. Thus, whatever effect the treatment had, the difference in ability probably did not let the treatment effect work quite as well in the E group compared to the C group. Stated another way, the advantage of ability in the control group probably made the control group's MEAN on science a bit CLOSER to the mean in the experimental group than it should have been IF BOTH GROUPS HAD BEEN EQUAL ON IQ AT THE BEGINNING. Does this make sense?

So, what if both groups HAD been of equal IQ ability at the start, what might the means on the science scores have looked liked? Note above that I have also listed out the mean IQ score for both groups together .... this is the grand mean of course. Thus, consider this: what might the means on science have been in E and C if both groups had a mean IQ of 100.65? Since the mean in E on science was lower than the mean on science in C, it seems like we would boost up the mean on science a bit (if their ability had been a bit higher (from 98.9 to 100.65) while the mean on science would need to be reduced a bit if their IQ ability had been a bit lower (from 102.4 to 100.65).

There is a way to make these estimates for adjusting the means ... it is based on the regression line using IQ to estimate science scores ... and the general form of this looks like:

Adjusted Mean = grp mean science - slope value of reg line (grp mean IQ - grand mean IQ)

Without going into how the slope value is calculated, the formulas for the E and C groups would be:

Adj Mean E = 68.1 - 1.66 (98.9 - 100.5) = 70.76

Adj Mean C = 61 - 1.66(102.4 - 100.5) = 57.85

Thus, given that the IQ means were in the opposite direction of the actual science means for E and C, the effect of this adjustment was to boost the mean for E and lower the mean for C. Thus, the adjusted means show a LARGER difference than what was seen originally, which is consistent with the fact that the adjusted ANOVA produces a larger F ratio than we found originally.

ASSUMPTIONS OF ANCOVA

There are numerous assumptions for properly applying and interpreting the results from ANCOVA, but two of particular importance are the following.

1. All the regular assumptions for ANOVA apply, particularly the one about RANDOM ASSIGNMENT OF Ss TO TREATMENT LEVELS/CONDITIONS. This is really crucial since, you will find many inappropriate uses of ANCOVA in situations where there was NOT random assignment of Ss, and people use ANCOVA as a post hoc way making the groups appear as though random assignment had been used.

2. It is also assumed that the regression equations (one for using IQ to predict science scores in E group, and another equation for the comparable prediction in the C group) have the same slope

values. Or, another way to say that is that the two regression lines in this case would be parallel. (Note: they don't have to be one and the same lines since if there is a treatment effect, one line might be higher ... ie, have a different intercept ... than the other.) Now, violations of this assumption might not prevent the ANCOVA from still reducing within group error and thus increasing the F ratio, BUT, the interpretation of the adjusted MEANs tht we made above would be suspect. Remember, I used the same slope value (1.66 in this case) to make BOTH mean adjustments.

I cannot too strongly emphasize how often ANCOVA is applied inappropirately. This is due to two things in my opinion: 1) the fact that many designs for research either can't or don't utilize random assignment of Ss to conditions, and/or 2) have done 1 AND think that ANCOVA can salvage their design flaw. Well, it cannot. As you can see, ... if you believe me of course, ... ANCOVA has TWO sets of assumptions ... those for ANOVA and those related to regression. This makes the set of assumptions that are relevant to ANCOVA stricter than simple ANOVA ... not more relaxed. It puts more of an onus on the investigator ... not less. And finally, it will not make up for a procedural flaw in the way that you have assigned Ss to the different treatment conditions.

The example I have presented is at the most elemental level of the way in which you would use ANCOVA and do the necessary calculations. I have opted for a regression and r squared approach to do this ... rather than other ways for doing the necessary partitioning and SS adjustments. If you pick up 5 books that show ANCOVA, you will find 5 ways to doing the work. The main things I have tried to show you are: A) the logic behind ANCOVA, and ... B) the assumptions (and no nos) made when using this method of analysis. The bottom line is: in cases where there might be some treatment effect but the within group (or error) variation is large, it might be possible to reduce the error variance by the use of a covariate. If this is the case, you might find that your F ratio gives you a better chance of rejecting the null hypothesis. However, ANCOVA is not a panacea for a sloppy research design.

Take my word for it and ... put this one in the bank!

REFERENCE

The numerical example and general approach for this handout were taken from:

Glass and Hopkins, Statistical methods in psychology and education, Prentice-Hall, 1984 ... page 496.