Topic 13. Analysis of Covariance (ANCOVA, ST&D Chapter 17)

13. 1. Introduction

The analysis of covariance (ANCOVA) is a technique that is occasionally useful for improving the precision of an experiment. Suppose that in an experiment with a response variable *Y*, there is another variable *X*, such that *Y* is linearly related to *X*. Furthermore, suppose that the researcher cannot control *X* but can observe it along with Y. Such a variable *X* is called a **covariate** or a **concomitant variable**. The basic idea underlying ANCOVA is that precision in detecting the effects of treatments on *Y* can be increased by adjusting the observed values of *Y* for the effect of the concomitant variable. If such adjustments are not performed, the concomitant variable *X* could inflate the error mean square and make true differences in the response due to treatments harder to detect.

The concept is very similar to the use of blocks to reduce the experimental error. However, when the blocking variable is a continuous variable, the delimitation of the blocks can be very subjective.

The ANCOVA uses information about in *X* in two ways:

1. Variation in *Y* that is associated with variation in *X* is removed from the error variance (MSE), resulting in more precise estimates and more powerful tests
2. Individual observations of *Y* are adjusted to correspond to a common value of *X*, thereby producing group means that are not biased by *X*, as well as equitable group comparisons.

A sort of hybrid of ANOVA and linear regression analysis, ANCOVA is a method of adjusting for the effects of an uncontrollable nuisance variable. We will review briefly some concepts of regression analysis to facilitate this discussion

13. 2. Review of regression concepts.

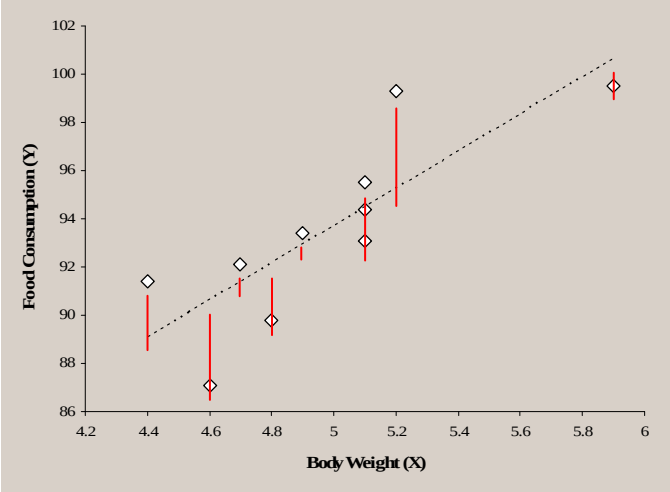
The equation of a straight line is $Y = a + bX$, where *Y* is the **dependent** variable and *X* is the **independent** variable. This straight line intercepts the *Y* axis at the value *a* so *a* is called the **intercept**. The coefficient *b* is the **slope** of the straight line and represents the change in *Y* for each unit change in *X* (i.e. rise/run). Any point (*X*,*Y*) on this line has an *X* coordinate, or **abscissa**, and a *Y* coordinate, or **ordinate**, whose values satisfy the equation $Y = a + bX$.

13.2.1 The principle of least squares

To find the equation of the straight line that best fits a dataset consisting of (*X*,*Y*) pairs, we use a strategy which relies on the concept of **least squares**. For each point in the dataset, we find its vertical distance from the putative best-fit straight line, square this distance, and then add together all the squared distances (i.e. vertical deviations). Of all the lines that could possibly be drawn through the scatter of data, the **line of best fit** is the one that minimizes this sum.

Example: Below is a scatterplot relating the body weight (X) of 10 animals to their individual food consumption (Y). The data are shown to the left.

Body weight (X)	Food consumption (Y)
4.6	87.1
5.1	93.1
4.8	89.8
4.4	91.4
5.9	99.5
4.7	92.1
5.1	95.5
5.2	99.3
4.9	93.4
5.1	94.4



13. 2. 2. Residuals

The vertical distance from an individual observation to the best-fit line is called the **residual** for that particular observation. These residuals, indicated by the solid red lines in the plot above, are the differences between the *actual* (observed) *Y* values and the *Y* values that the regression equation predicts. These residuals represent variation in *Y* that the independent variable (X) does not account for (i.e. they represent the *error* in the model).

13. 2. 3. Formulas to calculate *a* and *b*

Fortunately, finding the equation of the line of best fit does not require summing the residuals of the infinite number of possible lines and selecting the line with smallest sum of squared residuals. Calculus provides simple equations for the intercept *a* and the slope *b* that minimize the SS of the residuals (i.e. the SSE):

$$b = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} \equiv \frac{S(XY)}{SS(X)} \quad \text{and} \quad a = \overline{Y} - b\overline{X}$$

For the sample dataset given above, we find:

$$b = \frac{\left[(4.6 - 4.98)(87.1 - 93.56) + \ldots + (5.1 - 4.98)(94.4 - 93.56)\right]}{(4.6 - 4.98)^2 + \ldots + (5.1 - 4.98)^2} = 7.69$$

$$a = 93.56 - 7.69(4.98) = 55.26$$

Therefore, the equation of the line of best fit is $Y = 55.26 + 7.69X$.

13. 2. 4. Covariance

In the formula for the slope given above, the quantity S(XY) is called the **corrected sum of cross products**. Dividing S(XY) by (n – 1) produces a statistic called the **sample covariance between *X* and *Y***, which is a quantity that indicates the degree to which the values of the two variables vary together. If high values of *Y* (relative to $\overline{Y}$) are associated with high values of *X* (relative to $\overline{X}$), the sample covariance will be positive. If high values of *Y* are associated with low values of *X*, or vice-versa, the sample covariance will be negative. If there is no association between the two variables, the sample covariance will be close to zero.

13.2.5 Using SAS for regression analysis

PROC GLM can be used for regression analysis, as seen before when we discussed trend analysis. Representative code for the sample dataset above:

```
Data Example;
   Input X Y @@;
Cards;
4.6 87.1   5.1 93.1   4.8 89.8   4.4 91.4   5.9 99.5
4.7 92.1   5.1 95.5   5.2 99.3   4.9 93.4   5.1 94.4
;
Proc GLM;
   Model Y = X / solution;
Run;
Quit;
* Note that the independent/explanatory variable X is not declared in a
CLASS statement;
```

Output:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	90.835510	90.835510	16.23	0.0038
Error	8	44.768490	5.596061		
Corrected Total	9	135.604000			

F: tests if the model as a whole accounts for a significant proportion of *Y*.

R-Square	C.V.	Root MSE	Y Mean
0.669859	2.528430	2.3656	93.560

R-Square: measures how much variation in *Y* the model can account for.

This analysis tells us that the model accounts for a significant (p = 0.0038) amount of the variation in the experiment, nearly 67% of it (R-square = 0.67). This indicates that a great deal of the variation in food consumption among individuals is explained, through a simple linear relationship, by differences in body weight.

The "solution" option in the Model statement requests parameter estimates for this linear relationship. The output from this option:

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
INTERCEPT	55.26328125	5.80	0.0004	9.53489040
X	7.69010417	4.03	0.0038	1.90873492

Estimates: Calculates the INTERCEPT (a= 55.26) and the slope (b= 7.69). Therefore, the equation of the best-fit line for this dataset is $Y = 55.26 + 7.69X$, just as we found before. The p-values associated with these estimates (0.0004 for a and 0.0038 for b) are the probabilities that the true values of these parameters are different from zero.

13. 2. 6. Analysis of the adjusted *Y*'s

The experimental error in the previous analysis (MS_{error} = 5.596) from the previous analysis represents the variation in food consumption (Y) that would have been observed if all the animals used in the experiment had had the same initial body weight (X).

In the following table each Y value is adjusted using the regression equation to a common X. Any value of X can be used to adjust the Y's but the mean of the X (4.98) values is used as a representative value:

	X	***Y***	***Adjusted Y*** $= Y - b(X - \overline{X}_{..})$
	4.6	87.1	90.02224
	5.1	93.1	92.17719
	4.8	89.8	91.18422
	4.4	91.4	95.86026
	5.9	99.5	92.4251
	4.7	92.1	94.25323
	5.1	95.5	94.57719
	5.2	99.3	97.60818
	4.9	93.4	94.01521
	5.1	94.4	93.47719
X_{mean}=	4.98		
SSY		135.604	44.76849

The first adjusted value. 90.02224, is the food consumption expected for this animal *if its initial body weight had been* 4.98 ($\overline{X}$). Because X and Y are positively correlated, the adjusted food consumption for underweight animals is always higher than the observed values and the adjusted food consumption for overweight animals is always lower.

Note that the SS of the adjusted Y's is similar to the Total SS of the previous ANOVA and that the SS of the adjusted Y's is similar to the SS_{error}. The SS_{error} is the variation in

food consumption that we would have found if all the animals used in the experiment had had the same weight (assuming that "b" was estimated without error).

Note also the large reduction in the variation of the Y's that is obtained when the variation due to the regression is eliminated.

13. 3. ANCOVA example

The analysis of covariance is illustrated below with data from a pilot experiment designed to study oyster growth. Specifically, the goals of this experiment were:

1. To determine if exposure to artificially-heated water affects growth
2. To determine if position in the water column (surface vs. bottom) affects growth

In this experiment, twenty bags of ten oysters each were placed across 5 locations within the cooling water runoff of a power-generation plant (i.e. 4 bags / location). Each location is considered a treatment: TRT1: cool-bottom, TRT2: cool-surface, TRT3: hot-bottom, TRT4: hot-surface, TRT5: control (i.e. mid-depth and mid-temperature).

Each bag of ten oysters is considered to be one experimental unit. The oysters were cleaned and weighed at the beginning of the experiment and then again about one month later. The dataset consists of the initial weight and final weight for each of the twenty bags.

The code (from SAS System for linear models):

```
Data Oyster;
   Input Trt Rep Initial Final;
Cards;
1     1     27.2  32.6
1     2     32.0  36.6
1     3     33.0  37.7
1     4     26.8  31.0
2     1     28.6  33.8
2     2     26.8  31.7
2     3     26.5  30.7
2     4     26.8  30.4
3     1     28.6  35.2
3     2     22.4  29.1
3     3     23.2  28.9
3     4     24.4  30.2
4     1     29.3  35.0
4     2     21.8  27.0
4     3     30.3  36.4
4     4     24.3  30.5
5     1     20.4  24.6
5     2     19.6  23.4
5     3     25.1  30.3
5     4     18.1  21.8
;
Proc GLM;   * Simple overall regression;
   Model Final = Initial;
Proc Sort;
   By Trt;
```

```
Proc GLM;   * Simple regression, within each treatment level;
   Model Final = Initial;
   By Trt;
Proc GLM;   * The one-way ANOVA;
   Class Trt;
   Model Final = Trt;
Proc GLM;   * The ANCOVA;
   Class Trt;
   Model Final = Trt Initial;
Run;
Quit;
```

The first **Proc GLM** performs a simple linear regression of Final Weight on Initial Weight and shows that, for the experiment as a whole, there is a significant linear relationship between these two variables ($p < 0.0001$; $R^2 = 0.95$), as shown in the output below.

Simple regression

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	342.358	342.358	377.79	<.0001
Error	18	16.312	0.906		
Corrected Total	19	358.669			

R-Square	Coeff Var	Root MSE	Final Mean
0.954522	3.086230	0.951948	30.84500

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Initial	1	342.3578	342.3578	377.79	<.0001

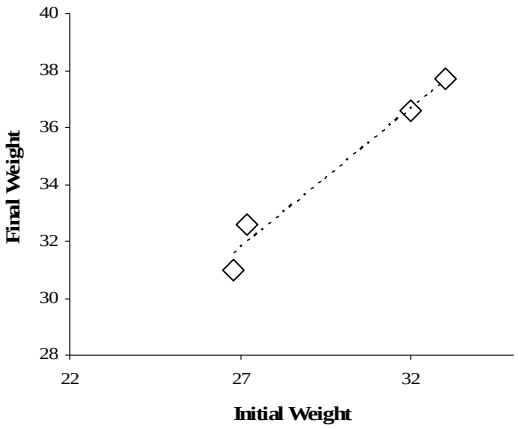
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	3.764686495	1.40940934	2.67	0.0156
Initial	1.051254406	0.05408550	19.44	<.0001

This strong dependence of Final Weight on Initial Weight suggests that Initial Weight may be a useful **covariable** for this analysis. The second **Proc GLM** carries out a similar analysis *within each treatment group separately*. This analysis reveals the fact that the slope of this regression is fairly uniform across all treatment levels. This is important because in ANCOVA, all treatment groups are adjusted by the *same* slope. The estimates of the slopes within each treatment group:

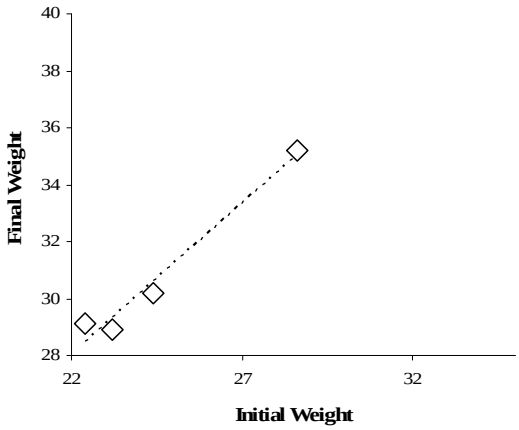
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Slope(Trt1)	**0.982646773**	0.10941830	8.98	0.0122
Slope(Trt2)	**1.501355014**	0.39230860	3.83	0.0620
Slope(Trt3)	**1.056066579**	0.12801209	8.25	0.0144
Slope(Trt4)	**1.056925032**	0.06842649	15.45	0.0042
Slope(Trt5)	**1.223886048**	0.02530981	48.36	0.0004

This similarity can be seen in the following scatterplots of Final vs. Initial Weight for treatment levels 1 and 3 below:

TRT1: $Y = 5.24 + 0.98X$



TRT3: $Y = 4.82 + 1.06X$



In the third **Proc GLM**, the **CLASS** statement specifies that TRT is the only classification variable and the analysis is a simple ANOVA of a CRD with four replications. The output:

The ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	198.407	49.602	4.64	0.0122
Error	15	160.262	10.684		
Corrected Total	19	358.669			

R-Square	Coeff Var	Root MSE	Final Mean
0.553	10.597	3.269	30.84500

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Trt	4	198.407	49.602	4.64	0.0122

From these results, we would conclude that location does affect oyster growth (P = 0.0122). This particular model explains roughly **55%** of the observed variation.

Finally, in the last **Proc GLM** (the ANCOVA), we ask the question: What is the effect of location on Final Weight, adjusting for differences in Initial Weight? That is, what would the effect of Location be if all twenty bags of oysters had started with the same initial weight? Notice that the continuous variable "**Initial**" does not appear in the **Class** statement, designating it as a regression variable, or a **covariate**. This is similar to the regression example in 13.2.4. The ANCOVA output:

Source	DF	S. of Squares	Mean Square	F Value	Pr > F
Model	5	354.447	70.889	235.05	<.0001
Error	14	4.222	0.302		
Corrected Total	19	358.669			

R-Square	Coeff Var	Root MSE	Final Mean
0.988	1.780	0.549	30.845

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Trt	4	**198.407**	49.602	164.47	<.0001
Initial	1	156.040	156.040	517.38	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Trt	4	12.089	3.022	10.02	0.0005
Initial	1	156.040	156.040	517.38	<.0001

There are several things to notice here. First, the Type I SS for TRT (**198.407**) is the **unadjusted treatment SS** and is the same as the one found in the one-way ANOVA. If we subtract this SS from the Total SS, we obtain the error SS for the simple one-way ANOVA (358.6695 – 198.407 = 160.2625).

The Type III SS for TRT (12.1) is the **adjusted treatment SS** and allows us to test the treatment effects, adjusting for all other factors included in the model. The reason adjustments are needed is because the two factors in the model, class variable TRT and regression variable INITIAL, are not orthogonal to one another. Because INITIAL is a continuous variable, the design is not balanced, even though there are no missing data (i.e. not all levels of TRT are present in all levels of Initial). This lack of orthogonality necessitates the use of partial sums of squares.

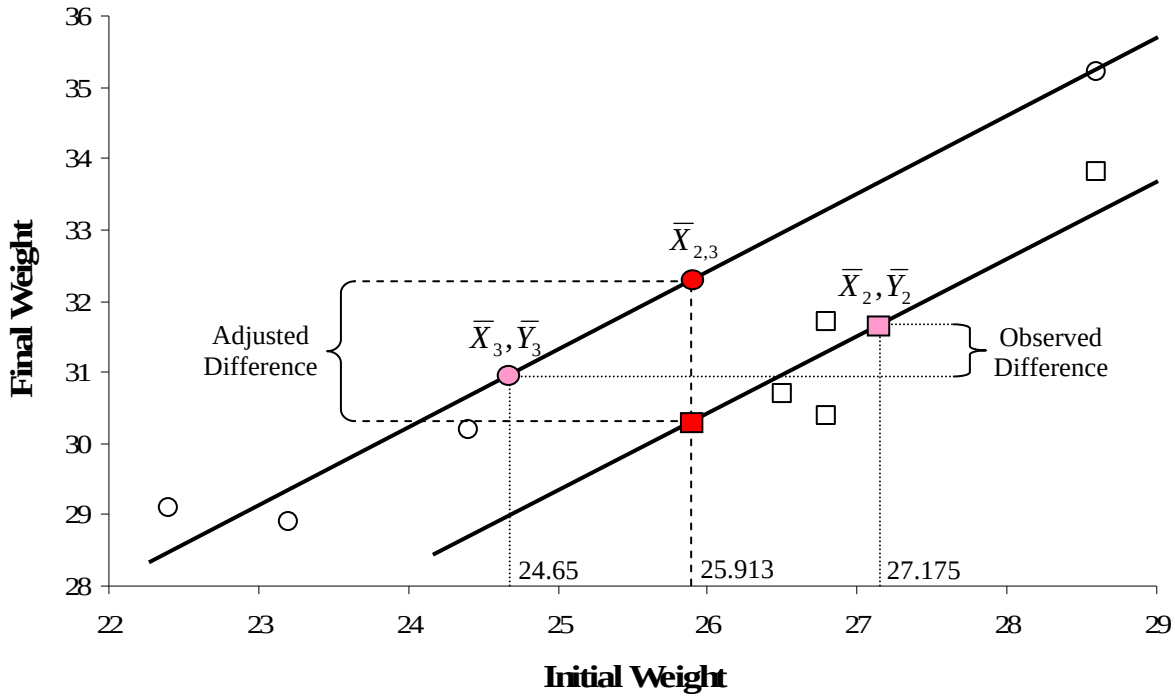
Type III SS produces the appropriate results in ANCOVA.

Even though the adjusted Trt MS (3.02) is much smaller than the unadjusted TRT MS (49.60), the reduction in the MSE is also quite large (from 10.68 in the ANOVA to 0.30 in the ANCOVA). It is this reduction in the experimental error that drives the increase in the F statistic for TRT from 4.64 in the simple one-way ANOVA to 10.02 in the ANCOVA. The power of the test for treatment differences increases when the covariate is included because most of the error in the simple ANOVA is due to variation among INITIAL values.

Similarly, the true test of the significance of the linear components of the relationship between INITIAL (X) and FINAL (Y) uses an INITIAL SS that is *adjusted for the effects of treatment*. In this case, notice that the INITIAL SS decreased (from 342.36 in the simple regression to 156.04 in the ANCOVA) because some of the observed variation can be attributed to the treatments. But again, the MSE also decreased significantly (from 0.91 to 0.30), ultimately leading to a more sensitive test for INITIAL.

13. 3. 1. Graphic interpretation of the ANCOVA example

The following scatterplot shows the data for treatments 2 (white squares) and 3 (white circles) from the oyster example. The mean final weight of treatment 3 (pink circle, 30.85) is seen to be slightly lower than the mean final weight of treatment 2 (pink square, 31.65).



For each treatment, variation in *X* is seen to contribute to variation in *Y*, as indicated by the common regression lines (solid lines). Because of this relationship, differences in the initial average weights of oysters assigned to each treatment can contribute greatly to the observed differences between the final average weights. For example, Treatment 3 started with an initial average weight of 24.65, while Treatment 2 started with an initial average weight of 27.175. It is therefore likely that the final difference in weights (24.65 < 27.175) is not a good indicator of the treatment effects because the difference is due to *both* treatment effects *and* the differences in initial weights.

Thus the need to adjust the observed treatment means to some common initial weight. In the schematic above, this common initial weight is the overall mean (25.913). Adjustment consists of sliding the values of Treatment 3 up its regression line and the values of Treatment 2 down its regression line such that the initial weights of the two treatments are equal to the overall mean. By adjusting in this way, we see that the real effect of Treatment 3 is to increase the final weights of the oysters relative to Treatment 2. This effect was completely hidden in the original data.

13. 3. 2. Least squares means or adjusted means

If a MEANS statement is included in the previous example, it will produce the unadjusted treatment means of all continuous (i.e. non-CLASS) variables in the model. As discussed in the graphic example above, these means and the comparisons among them are not strictly appropriate.

To compare the true effects of the treatments, unbiased by differences in initial weights, the treatment means should be adjusted to what their values *would have been if all oysters had had the same initial weight*. As we saw before with unbalanced designs, these adjusted means can be calculated in SAS via the **LSMEANS** (least-squares means) statement. For example, the statement:

```
LSMeans Trt / StdErr Pdiff Adjust = Tukey;
```

will generated the estimated least-squares means followed by their standard errors (StdErr option) and the p-values for all pairwise tests of treatment differences (Pdiff option), according to the Tukey method of means separations (Adjust = Tukey option).

Note the large differences between the unadjusted to adjusted treatments means for the variable FINAL weight in the table below:

TRT	Unadjusted Means	Adjusted LS Means	Calculation $\bar{Y}_{adj_i} = \bar{Y}_i - \beta(\bar{X}_i - \bar{X})$
1	34.475	30.153	34.475 - 1.08318 (29.75 - 25.76)
2	31.650	30.117	31.650 - 1.08318 (27.18 - 25.76)
3	30.850	32.052	30.850 - 1.08318 (24.65 - 25.76)
4	32.225	31.504	32.225 - 1.08318 (26.43 - 25.76)
5	25.025	30.398	25.025 - 1.08318 (20.80 - 25.76)

These differences are due to the large differences in initial weights among the treatment groups (TRT 5, for example, was assigned much smaller oysters than other treatments). In calculating these adjusted means, the coefficient β = 1.08318 is a weighted average of the slopes of the linear regressions for each of the five treatment groups. To obtain this coefficient, simply add the solution option to the Model statement:

```
Model Final = Trt Initial / solution;
```

The estimated slope obtained with this solution statement is identical to the slope obtained by performing individual ANOVAs on both X and Y, calculating their respective residuals, and then running a regression of the **Y residuals on the X residuals**. As suggested in the above table, this slope can be used to create a new adjusted response variable:

$$Z = Y - \beta(X - \bar{X})$$

This adjusted response variable (Z) is very handy because it can be used to perform a Levene's test for homogeneity of variances as well as a Tukey's test for non-additivity, if the design is an RCBD with only one observation per block-treatment combination.

13. 3. 3. Contrasts

In this particular oyster example, the adjusted treatment means from the ANCOVA can be analyzed further with four orthogonal contrasts, as shown:

```
Proc GLM Order = Data;
   Class Trt;
   Model Final = Trt Initial;
   LSMeans Trt;
      Contrast 'Control vs. Treatment'    TRT -1 -1 -1 -1 4;
      Contrast 'Bottom vs. Top'           TRT -1  1 -1  1 0;
      Contrast 'Cool vs. Hot'             TRT -1 -1  1  1 0;
      Contrast 'Interaction Depth*Temp'   TRT  1 -1 -1  1 0;
```

The output:

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Trt	4	**12.089**	3.022	10.02	0.0005
Initial	1	156.040	156.040	517.38	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Control vs. Treatment	1	0.5200	0.5200	1.72	0.2103
Bottom vs. Top	1	0.3388	0.3388	1.12	0.3071
Cool vs. Hot	1	8.5911	8.5910	28.49	**0.0001**
Interactions Depth*Temp	1	0.2293	0.2293	0.76	0.3979
		9.6792			

The output indicates that oyster growth is only significantly affected by differences in temperature (cool vs. hot). Although constructed to be orthogonal, these contrasts are not orthogonal to the covariable; **therefore, their sums of squares do not add to the adjusted treatment SS.**

Now consider this: If the covariable is **not** included in the model, these exact same contrasts produce completely different results:

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Control vs. Treatment	1	169.362	169.362	15.85	**0.0012**
Bottom vs. Top	1	2.102	2.102	0.20	0.6637
Cool vs. Hot	1	9.302	9.302	0.87	0.3655
Interactions Depth*Temp	1	17.640	17.640	1.65	0.2183

Here, the effect of temperature is completely obscured by the differences in the initial weights among the treatments. The significant difference between the control (TRT 5) and the treatments (TRT 1 – 4) is due entirely to the lower initial weights of the oysters placed in the control bags. Note in this case that the contrast SS sum perfectly to the TRT SS (198.407).

13. 4 ANCOVA model

Recall the ANOVA model for a CRD:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

The regression model is:

$$Y_i = \mu + \beta(X_i - \overline{X}_.) + \varepsilon_i$$

ANCOVA is a combination of ANOVA and regression, as reflected in its linear model:

$$Y_{ij} = \mu + \tau_i + \beta(X_{ij} - \overline{X}_{..}) + \varepsilon_{ij}$$

Extending this concept, the linear model for ANOCOVA within any given design (e.g. CRD, RCBD, LS, etc.) is simply the linear model for the ANOVA *plus* an additional term for the concomitant variable. For the CRD, the formula can be slightly rearranged:

$$Y_{ij} - \beta(X_{ij} - \overline{X}_{..}) = \mu + \tau_i + \varepsilon_{ij}$$

And here you have it: **An ANCOVA on the original response variable (Y) is equivalent to a regular ANOVA on values of Y that have been adjusted according to their linear dependence on X**. In the discussion which follows, we denote these regression-adjusted values with the letter Z.

13. 5 Assumptions of the ANCOVA

1. **The X's are fixed, measured without error, and independent of treatments.**

   This means that inferences are appropriate for interpolated (i.e. within range) rather extrapolated (outside range) values, that the measurement error for X is trivial relative to the observed variation in X, and that the treatments themselves will not affect the X values.

2. **The regression of Y on X, after accounting for treatment differences, is linear and independent of treatments.**

   This means that the regression is assumed to be approximately linear within the given range of X values and that the slopes of the regressions within each of the treatment levels are not significantly different from one another. Note that a linear relation is often a reasonably good approximation even in nonlinear cases, provided the values of the independent variable do not span too wide a range.

3. **The residuals are normally and independently distributed with zero mean and a common variance.**

   These are the normal assumptions for the validity of *F* tests.

13. 5. 1. Independence of X values from the treatments

If the covariable is measured **before** the treatments are applied, like in the previous oyster example, the independence of the treatments and the concomitant variable is satisfied by definition. However, if the concomitant variable is measured **after** the experiment, the independence of the covariable and the treatments should be tested.

An **ANOVA** which uses the covariable **(X) as the response variable** is appropriate to test this hypothesis. The null hypothesis is that there are no significant differences in X among treatment levels, and we expect to find no significant differences in order to be able to perform a standard covariance analysis.

The following code is included only as an example, because the test is not required in the oyster example.

```
Proc GLM;
   Title 'Test for independence of treatments and covariable';
   Class Trt;
   Model Initial = Trt;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	176.793	44.198	4.98	0.0093
Error	15	132.995	8.866		
Corrected Total	19	309.788			

R-Square	Coeff Var	Root MSE	Initial Mean
0.570690	11.55916	2.977639	**25.76000**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Trt	4	176.793	44.198	4.98	0.0093

In this case, we see that the differences in initial weight are highly significant. When the treatments are applied *after* measuring the covariable (i.e. when the covariable is not affected by the treatments) and the ANOVA of the covariable is found to be significant, the selected covariable will most likely have an effect on the final results. This is due to the fact that the experimental units across the treatments are not very uniform and adjusting for this lack of uniformity will most likely increase the precision of tests.

ANCOVA can be used where the *X* values are affected by the treatments, but results should be interpreted with caution.

13. 5. 2. Test for heterogeneity of slopes.

ANCOVA assumes homogeneity of covariate regression coefficients. This is ANCOVA's "equality of regressions" or "homogeneity of regressions" assumption. Because a single slope is used to adjust all observations in the experiment, the covariate coefficients (i.e. the slopes of the regression lines) must be the same for each level of the

categorical variable being analyzed (i.e. all slopes must estimate the same common slope β).

Said another way: The adjustment of the *Y* values using a single β for all treatments is based on the assumption that there is a constant regression relationship among groups. The test for heterogeneity of slopes tests the validity of this assumption; that is, it tests whether or not the regression coefficients are constant over groups.

The null hypothesis of this test is H_o: $\beta_1 = \beta_2 = ... = \beta_i$

A regression relationship that differs among treatment groups reflects an **interaction between the treatment groups and the independent variable or covariate**. In fact, to test the assumption of homogeneity of slopes, one simply specifies and analyzes this phenomenon as an interaction within **Proc GLM**. For example:

```
Proc GLM;
   Title 'Test for homogeneity of slopes';
   Class Trt;
   Model Final = Trt Initial Trt*Initial;
```

In this code, the expression "**Trt*Initial**" probes the treatment*regression interaction within the overall ANCOVA model. This factor in the model produces the appropriate statistics for estimating the differences among regressions across the different levels of Trt.

The output:

Dependent Variable: Final

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	355.835	39.537	139.51	<.0001
Error	10	2.834	0.283		
Corrected Total	19	358.669			

R-Square	Coeff Var	Root MSE	Final Mean
0.992099	1.725901	0.532354	30.84500

Source	DF	Type III SS	Mean Square	F Value	Pr > F
~~Trt~~	~~4~~	~~1.6962~~	~~0.4240~~	~~1.50~~	~~0.2752~~
~~Initial~~	~~1~~	~~68.5289~~	~~68.5289~~	~~241.81~~	~~<.0001~~
Trt*Initial	4	1.3883	0.3471	1.22	**0.3602**

The last row **Trt*Initial** is an additional SS due to differences in regression coefficients among the groups specified by the classification variable TRT. If $P > 0.05$ for this interaction, we do not reject the hypothesis of homogeneity of slopes and are justified in proceeding with a standard analysis of covariance.

In more complex models, a similar strategy is followed. That is, to test for homogeneity of slopes across all levels of the factor(s) being tested, you need to run a **Proc GLM** with the standard ANCOVA model, to which the **Trt*X** interaction is added. For example, in an RCBD with one treatment factor (TRT) and a covariable (X):

```
Proc GLM;
   Title 'Test for homogeneity of slopes';
   Class Block Trt;
   Model Y = Block Trt X Trt*X;
```

13. 5. 3. Normality of residuals.

To test for normality of residuals, simply request the residuals from the **ANCOVA** model and perform a **Shapiro-Wilk** test on those residuals, as seen previously:

```
Proc GLM;
   Class Trt;
   Model Final = Trt Initial;
   Output out = OysterPR p = Pred r = Res;

Proc Univariate Data = OysterPR normal;
   Var Res;

Proc Plot Data = OysterPR;
   Plot Res*Pred = Trt;
```

13.5.4 Homogeneity of variances (and example of a complete ANCOVA)

Recall that Levene's Test is only defined for one-way ANOVAs. To test for homogeneity of variances in an ANCOVA, therefore, it is customary to adjust the response variable (Y) according to its regression on X and perform a Levene's Test on the adjusted variable (Z). Below is the code for the complete analysis of the oyster experiment, including the Levene's test:

```
Data Oyster;
   Input Trt Rep Initial Final;
   Z = Final - 1.083179819*(Initial-25.76);
Cards;
1     1     27.2  32.6
1     2     32.0  36.6
1     3     33.0  37.7
1     4     26.8  31.0
2     1     28.6  33.8
2     2     26.8  31.7
2     3     26.5  30.7
2     4     26.8  30.4
3     1     28.6  35.2
3     2     22.4  29.1
3     3     23.2  28.9
3     4     24.4  30.2
4     1     29.3  35.0
4     2     21.8  27.0
4     3     30.3  36.4
4     4     24.3  30.5
5     1     20.4  24.6
5     2     19.6  23.4
5     3     25.1  30.3
5     4     18.1  21.8
;
Proc GLM;
```

```
   Title 'One-way ANOVAs for INITIAL and FINAL';
   Class Trt;
   Model Initial Final = Trt;

Proc GLM;
   Title 'General regression';
   Model Final = Initial;

Proc GLM Order = Data;
   Title 'The ANCOVA';
   Class Trt;
   Model Final = Trt Initial / solution;
     LSMeans Trt / StdErr Pdiff Adjust = Tukey;
     Contrast 'Control vs. Treatment'    TRT -1 -1 -1 -1 4;
     Contrast 'Bottom vs. Top'           TRT -1  1 -1  1 0;
     Contrast 'Cool vs. Hot'             TRT -1 -1  1  1 0;
     Contrast 'Interactions Depth*Temp'  TRT  1 -1 -1  1 0;

Proc GLM Order = Data;
   Title 'ANOVA on Z';
   Class Trt;
   Model Z = Trt;
   Output out = OysterPRz p = Pred r = Res;

Proc Univariate Data = OysterPRz normal;
   Title 'Normality of residuals';
   Var Res;

Proc Plot Data = OysterPRz;
   Plot Res*Pred = Trt;
Proc GLM;
   Title 'Homogeneity of variances';
   Class Trt;
   Model Z = Trt;
   Means Trt / hovtest = Levene;

Proc GLM;
   Title 'Independence of Trt and X, unnecessary for this experiment';
    Class Trt;
   Model Initial = Trt;

Proc GLM;
   Title 'Homogeneity of slopes';
   Class Trt;
   Model Final = Trt Initial Trt*Initial;
Run;Quit;
```

Note that in an RCBD with one replication per block-treatment combination, the adjusted Z values can be also be used for the Tukey Test for Non-additivity. Example code:

```
Proc GLM;
   Title 'ANOVA on Z = adjusted Y';
   Class Block Trt;
   Model Z = Block Trt;
   Output out = OysterZPr p = PredZ r = ResZ;

Proc GLM Data = OysterZPR;
```

```
Title 'Tukey test for non-additivity';
Class Block Trt;
Model Z = Block Trt PredZ*PredZ;
```

13. 6. Increase in precision due to covariance (ST&D 17.6)

To test the effectiveness of the covariate as a means of error control, a comparison can be made between the experimental errors, with and without the covariance adjustment. In the oyster example:

ANOVA of Y:	MSE = 10.68417	df_{error} = **15**
ANCOVA of Y:	MSE = 0.30159	df_{error} = **14**

To make the comparison between these two errors, the 2nd value (ANCOVA MSE) must be adjusted upward to allow for sampling error in the regression coefficient. The appropriate adjustment involves the TRT SS (176.793) and the error SS (132.995) from the ANOVA of X.

The **effective ANCOVA MSE**, adjusting for sampling error in X, is therefore given by:

$$MS_{Error\ Adj.Y}\left[1+\frac{TRT\ \ SS_{X\ \ variable}}{(t-1)\ \ Error\ SS_{X\ \ variable}}\right]=0.30159\left[1+\frac{176.793}{4*132.995}\right]=0.402$$

An estimate of the relative precision is

$RE_{ANCOVA:ANOVA}$ = $MS_{Error\ ANOVA}$/ **Effective** $MS_{Error\ Adj.\ Y}$ = 10.68417 / 0.402= 26.6,

This indicates that each replication, adjusted for the effect of the covariable, is as effective as 26.6 replications without such adjustment. The ANCOVA in this example is 26.6 times more precise than the ANOVA in detecting the effects of the treatments.

13. 7. Comparison between ANCOVA and ANOVA of the differences (final –initial)

Can we obtain a similar result just by subtracting final weight – initial weight?

We will compare the previous **ANCOVA** result:

Dependent Variable: ***FINAL*** **(weight)**

Source	DF	Type III SS	MS	F Value	Pr > F
TRT	4	12.09	3.02	**10.02**	0.0005
INITIAL	1	156.04	156.04	517.38	0.0001

With an **ANOVA** of the differences (*DIF*= final – initial)

Dependent Variable: ***DIF*** **(final - initial)**

Source	DF	Type III SS	MS	F Value	Pr > F
TRT	4	11.98	3.00	**8.74**	0.0008
Error	15	5.14	0.34		
Corrected Total	19	17.13			

The *P* value of the ANOVA of the differences (final – initial, *P*= 0.0008) is better than the ANOVA of the uncorrected values (*P*=0.0122), but still not as good as the ANCOVA P value (*P*= 0.0005).

Why?

Because the subtraction of the initial weight eliminates its initial effect, but not the effect on the increase in weight during the treatment (large animals may gain weight faster or slower than average animals)

An ANCOVA of the differences (final-initial) produces exactly the same *P* value for treatment as the ANCOVA of the uncorrected data. This shows that the ANCOVA adjusts both for the initial difference and for the effect that that initial difference has on the effect of the treatments.

ANCOVA of the differences (final –initial)

Source	DF	Type III SS	MS	F Value	Pr > F
TRT	4	12.09	3.02	**10.02**	0.0005
INITIAL	1	0.92	0.92	3.05	0.1026

13. 8. Comparison between ANCOVA and ANOVA of ratios

An experiment was conducted to study the effect of stress on the presence of enzyme A in the liver (Bulletin de la Societé de Chimie Biologique, 1954). The researcher measured the total activity of enzyme A (Variable "A") from liver homogenates of 10 control and 10 shocked animals. He also measured the total amount of nitrogen in the homogenates (Variable "N") as an indicator of total enzyme activity (i.e. total protein content) in the liver. He took this measure as a way to correct the measurements for A by total enzyme activity within each animal. Since he knew that A is correlated with N (but did not know about ANCOVA), he decided to analyze the ratio (A/N), the activity of enzyme per unit of protein, as his response variable.

The data:

Control animals			Shocked animals		
N	**A**	**A/N**	**N**	**A**	**A/N**
84	76	90.4	122	108	88.5
28	38	133.9	98	158	161.2
166	72	43.4	115	58	50.0
98	64	65.3	86	65	75.5
105	53	50.0	69	40	58.0
84	28	32.8	86	65	75.5
72	31	43.0	102	82	80.3
80	28	34.3	112	94	84.1
84	28	32.7	98	65	66.3
105	49	46.1	74	76	102.7

ANOVA of the ratio A/N

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model = TRT	1	3650.4020	3650.4020	3.66	**0.0719**
Error	18	17970.7180	998.3732		
Corrected Total	19	21621.1200			

The ANOVA indicates that there are no significant differences between treatments (p = 0.0719). This result is not unexpected when you recognize the large variance within groups for variable A/N. Note, for example, the big difference between the extreme values of the control group (32.7 and 133.9).

ANCOVA of the variable A, using N as a covariable.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	8079.3923	4039.6961	6.31	0.0089
Error	17	10882.4077	640.1416		
Corrected Total	19	18961.8000			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	1	5108.8082	5108.8082	7.98	**0.0117**
N	1	2162.5923	2162.5923	3.38	0.0836

In contrast to the result from the ANOVA of (A/N), the ANCOVA of A, using N as a covariable, shows significant differences between treatments (p = 0.0117). The increased precision of the ANCOVA is evident, but *why* is ANCOVA the superior analysis?

The use of ANOVA to analyze ratios $Z = Y/X$ is not correct. Both X and Y, being random variables, exhibit random variation; consequently, the derived variable Z will exhibit a variance that is some combination of the variances of X and Y. Variation in the numerator (Y) affects Z in a linear way, but variation in the denominator affects Z in a hyperbolic way ($Z = 1/X$ is the equation of a hyperbole). This means that the magnitude of the error of Z depends not only on the error of X but also on the absolute value of X, with the error being higher for low values of X. Such a relationship clearly affects the homogeneity of variances.

The best way to analyze such ratios is an ANCOVA of the numerator using the denominator as the covariable.

13.9. Uses of ANCOVA (ST&D p 429)

The most important uses of the analysis of covariance are

1. To control error and thereby to increase precision
2. To adjust treatment means of the dependent variable for differences in the values of corresponding independent variables.
3. To assist in the interpretation of data, especially regarding nature of treatment effects.
4. To estimate missing data.

13. 9. 1. Error control

Reduction of the experimental error (MSE) is accomplished through experimental design or through the use of one or more covariates; and both strategies may be used simultaneously. Covariance can be used as an effective means of error control when variation in the dependent variable Y is partly attributable to variation in some continuous, independent variable X.

ANCOVA increases the precision with which treatment effects can be measured by removing, via linear regression, certain recognized effects that cannot be or have not been controlled effectively by experimental design. For example, in a cattle-feeding experiment, to compare the effects of several rations on gain in weight, animals will always vary in initial weight. Now, if initial weight is correlated with gain in weight, a portion of the experimental error for gain can be the result of differences in initial weight. Through covariance analysis, this portion may be computed and eliminated from the experimental error, thereby increasing the researcher's ability to detect differences due to the rations.

13. 9. 2. Adjustment of treatment means

When observed variation in *Y* is partly attributable to variation in *X*, variation among treatment *Y* means will also be affected by variation among treatment *X* means. For valid comparisons, the treatment *Y* means should be adjusted to make them *the best estimates of what they would have been* had all treatment *X* means been the same. For illustration, consider canning peas. This crop increases rapidly in yield as the peas progress in maturity. In a trial to evaluate the yields of different varieties, however, it is very difficult (read: impossible) to harvest all the varieties at the same state of maturity. Given the strong dependence of yield on maturity, therefore, an analysis of yields unadjusted for maturity differences may have little value. Worse, such an analysis may lead to completely wrong conclusions. In contrast, a comparison of yields adjusted for maturity differences (i.e. using maturity as a covariable) could have great value.

In field experiments, yields are often adjusted for differences in plot productivity as determined by uniformity trials. A uniformity trial measures the yields from plots handled in a uniform manner prior to the execution of the main experiment. With annual crops, the increased precision resulting from the use of uniformity data rarely pays; however, with long-lived perennials such as tree crops, there is often much to be gained.

In animal feeding experiments, differences among treatment means unadjusted for the amount of food consumed may be due to differences in the nutritive value of the rations, to differences in the amounts consumed, or to both. If differences among mean gains in weight for the different rations are adjusted to a common food intake, the adjusted means will indicate whether or not the rations differ in nutritive value.

Here you can begin to see how covariables can help not only in increasing the ability to detect differences in treatments but also in understanding some of the reasons underlying *how* the treatments produce different effects.

13.7.3 Interpretation of data

As hinted above, covariance analysis often aids the experimenter in understanding the principles underlying the results of an investigation.

Recall that the adjusted treatment Y means estimate the values expected when the treatment X means are the same for all treatment groups. In cases where X is itself influenced by the treatments, adjustment by X removes part of the treatment effects on Y. In such situations, the results of the ANCOVA must be interpreted carefully.

As an example, consider a fertilizer trial on sugar beets, in which the researcher is interested in testing the effects of treatments on yield (Y). It is possible that the treatments may cause differences in stand (germination and establishment of the crop). In such cases, the analysis of yield (Y) adjusted for stand differences (X) will remove part of the treatment effect, thereby leading the experimenter to a wrong interpretation of the data. An ANCOVA can still supply useful information in this case, however. Total yield is a function of average weight per beet and of average stand per plot. Now, if stand is influenced by treatments, the ANCOVA of yield *adjusted for stand differences* would test specifically if the treatments affect *individual beet weights* rather than yield in general.

While an adjustment in observed yield that is proportional to the number of established plants is sometimes practiced in agricultural experiments, this procedure is not recommended. This is because yields are rarely proportional to the number of plants per plot, so such an approach usually results in an over-correction for the plots with smallest stands. *ANCOVA provides a better method for adjusting yield data for differences in plant number*.

In situations where real differences among treatments for the independent variable do occur but are not the direct effect of the treatments, adjustment can be made without affecting interpretation. For example, consider a variety trial for which seed of the various varieties or strains was produced in different areas. Such seed may differ widely in germination rates, not because of inherent variety differences but as a result of the different environments in which the seed was produced. Consequently, differences in stand may occur even if planting rate is controlled. In this situation, the use of covariance for both error control and yield adjustment is warranted.

13.7.4. Estimation of missing data

Finally, the formulas given previously to estimate the values of missing data are based on the principle of minimizing the residual sum of squares. Such an approach, however, biases the treatment sum of squares upward. Covariance can be used to estimate missing values in a way that minimizes the residual sum of squares without biasing the treatment sum of squares. Such a procedure is simple to carry out though more difficult to describe than the previous procedure which required little more than a single formula.