ANCOVA (Analysis of Covariance): A Visual, Intuitive Understanding

Michael Senderak, Merck & Co., Inc., Upper Gwynedd, PA

ABSTRACT

As programmers, we may be asked not only to prepare data for a research study, but also apply a sophisticated statistical analysis procedure selected by the researcher but unfamiliar to us. Understanding these analytical methods can help us both appreciate the research in which we participate and gain a sense of confidence and closure in our work.

One extremely valuable and widely applicable statistical method is ANCOVA, which allows the researcher to mathematically subtract out differences between non-equivalent groups and thereby more precisely estimate the effect of a treatment given to one group but not the other(s).

The SAS® code is deceptively simple. Three or four brief PROC GLM statements will yield a complete ANCOVA analysis. But how does ANCOVA work? Although the name itself and the mathematics behind it can be quite intimidating, the underlying logic can be readily grasped at an intuitive and visual level, as we will see in this poster.

INTRODUCTION

Often a researcher will test the effects of a treatment or intervention by applying it to one group and comparing the results to an untreated group. In the real world, however, it can be difficult, impractical and sometimes impossible to study equivalent groups. This usually occurs when only pre-existing groups are available for the study, or when groups become nonequivalent while the study is underway. The consequence is that any differences observed between groups cannot be easily attributed to the treatment.

ANCOVA allows us to 'subtract out' these pre-existing differences and thereby estimate the actual treatment effect.

AN EXAMPLE

At the start of the school year, a researcher is assigned two student classes to test the effectiveness of a new teaching method over the current method. The two classes are in the same school, at the same age level, taught at the same time during the day, etc.

ANCOVA is a valuable tool here because students were pre-assigned to classes by the school system before the start of the school year. Consequently, although the two classes are similar in age, etc., the researcher had no opportunity to assemble equivalent groups through random assignment of students, or through matching by academic record, etc.

Study Objective: Do students perform better with the new vs. current instruction method?

Study Design:

	Pre_Treatment (1st 3 Months)	**Treatment (Last 6 months)**
Class C (20 students)	Current Instruction	Current Instruction
Class N (20 students)	Current Instruction	New Instruction

Performance Measure: At the end of the year each student's improvement in test scores is calculated as his/her: (average Treatment Period exam score minus average Pre-Treatment exam score)

Results: Improvement scores are summarized in **Figure 1**.

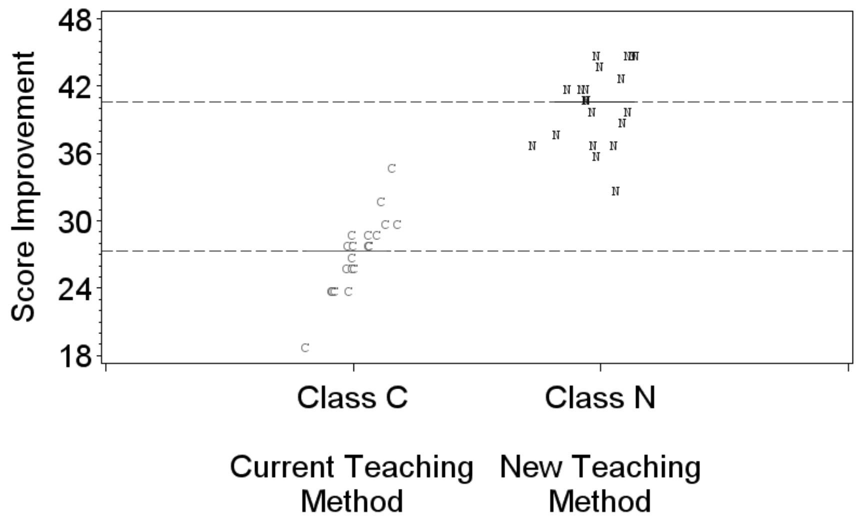

Figure 1. Plot of Student Improvement Scores Around Class Average. (Scores are spread out horizontally to reduce overlap.)

We might be tempted to simply compare the average improvement in Class N over Class C to determine the effectiveness of the new teaching method.

However, students who score better at the beginning of the year (higher Pre-Treatment scores) may improve more throughout the year. This means that the improvement in Class N may be due in part to higher Pre-Treatment scores, making it difficult to estimate the improvement due specifically to the new teaching method.

(Conceivably in a study, the effect of a treatment also may be understated if a covariate is not taken into account.)

What We Would Like to Know:

Queston 1: Do the students in Class N still show improvement over Class C after we 'subtract out' class differences in Pre-Treatment scores?

Queston 2: If so, by how much?

Queston 3: Even after subtracting out Pre-Treatment scores, how confident are we that any improvement in Class N is due to the new teaching method and not to chance?

Even after subtracting out Pre-Treatment student performance, the 2 groups can still differ by chance alone. For example, if neither class received the new teaching method, it's unlikely that they would perform identically.

VISUALLY UNDERSTANDING HOW ANCOVA WORKS

Which of the 2 following graphs gives you the most confidence that the group differences are due to the new teaching method? **Figure 2a** shows the results for our study. **Figure 2b** shows another set of hypothetical scores for the same study.

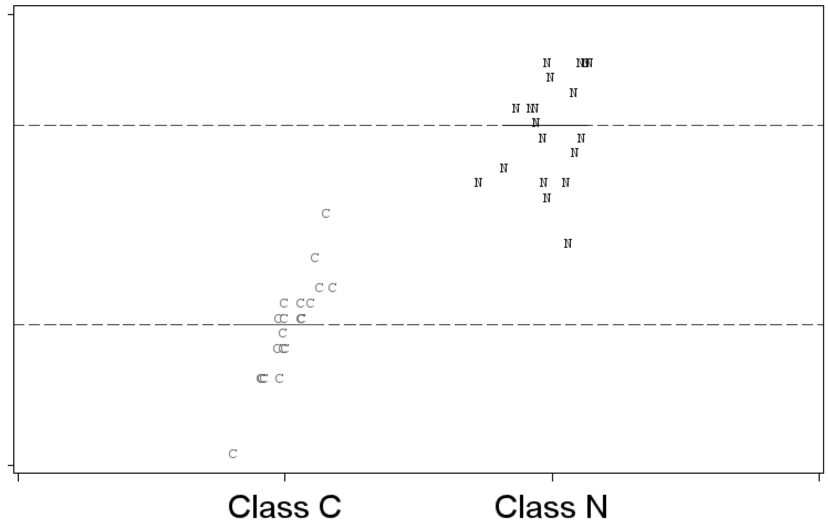

Figure 2a. Plot of Student Improvement Scores Around Their Class Average. (Scores are spread out horizontally to reduce overlap.)

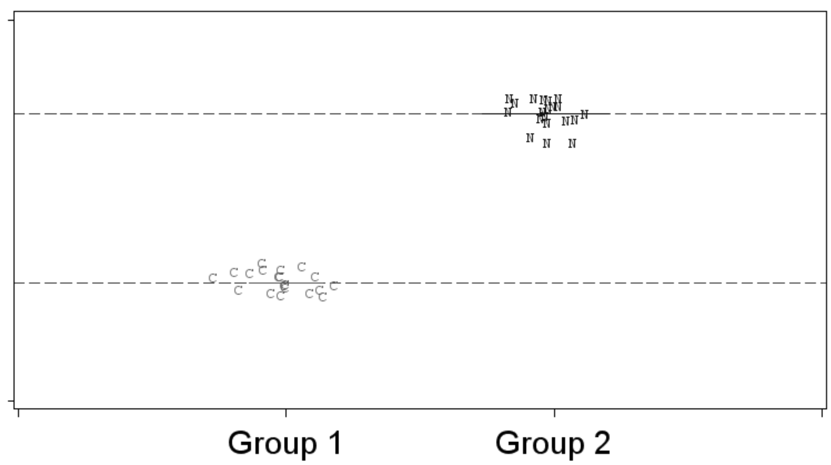

Figure 2b. Hypothetical Study: Plot of Student Improvement Scores Around Their Class Average. (Scores are spread out horizontally to reduce overlap.)

You probably chose **<u>Figure 2b</u>**. Here the students score more consistently around their group averages, suggesting that influences other than the teaching method they differ by are not affecting their scores.

And so we can see that the more we can reduce the distance, or variation, in student scores around their group average, the more confidently we can conclude the groups differ due to our treatment.

Now let's approach our research example step by step, to see the process of 'subtracting out' variation in ANCOVA.

<u>Figures 3a-3c</u> are plots of each student's improvement score against his/her Pre-Treatment score (20 Class C, 20 Class N data points). The 40 '<u>d</u>' values represent the variation we'd like to subtract out.

Visually, if we add all 40 '<u>d</u>' variation values in **<u>Figures 3a</u>** <u>and then in</u> **<u>3b</u>** separately, we can see the total variation would be less in **<u>Figure 3B</u>**. That is, we are able to 'subtract out' variation by using the group averages (**<u>Figure 3b</u>**)

rather than the average across all 40 students (**Figure 3a**), giving us confidence that the groups are performing differently (class N shows greater improvement (y-axis)).

In **Figure 3c** we introduce our 'covariate', Pre-Treatment scores, to help us 'subtract out' even more variation. Visually, we can see that the sum of d scores in **Figure 3c** is less than for **Figure 3b**. Here we fit the group lines to account not for score improvement alone (y-axis), but for the relationship between score improvement (y-axis) and Pre-Treatment scores (x-axis). This relationship represents how Pre-Treatment and improvement scores co-vary, hence the lines are named the ‘covariate lines’. These covariate best-fitting lines explain student improvement scores (vertical axis) using their Pre-Treatment scores (horizontal axis). In other words, they show that for both Class N and Class C those students with higher Pre-Treatment scores tend to show greater improvement during the Treatment period Each line is 'best fitting' in that the students' data points cluster more closely (i.e., with less variation) around them than any other line.

The only remaining variation we can't account for are the d values around the best fitting lines in **Figure 3c**. In statistical parlance, this is our 'error variation'.

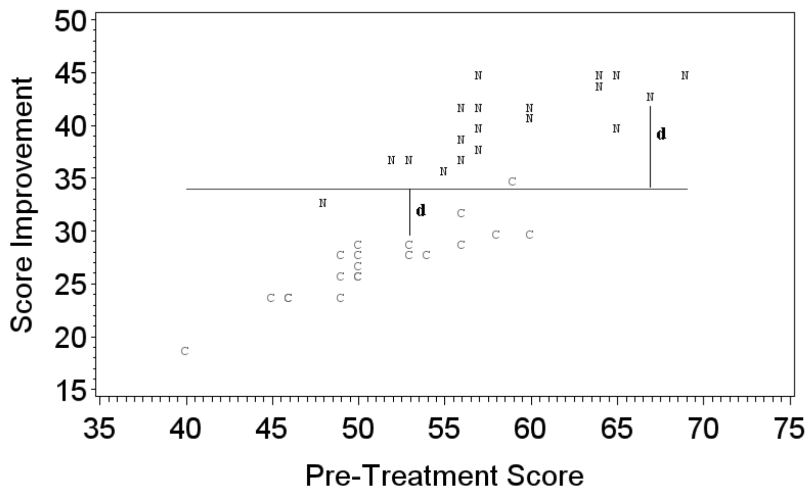

Figure 3a. Plot of Student Improvement Scores Around the Average of All 40 Students (Class N and Class C combined).

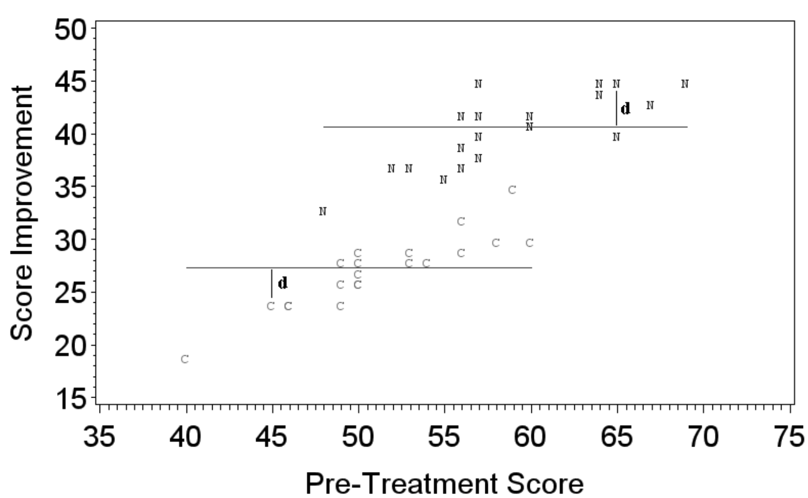

Figure 3b. Plot of Student Improvement Scores Around their Class Average (Class N and Class C separately).

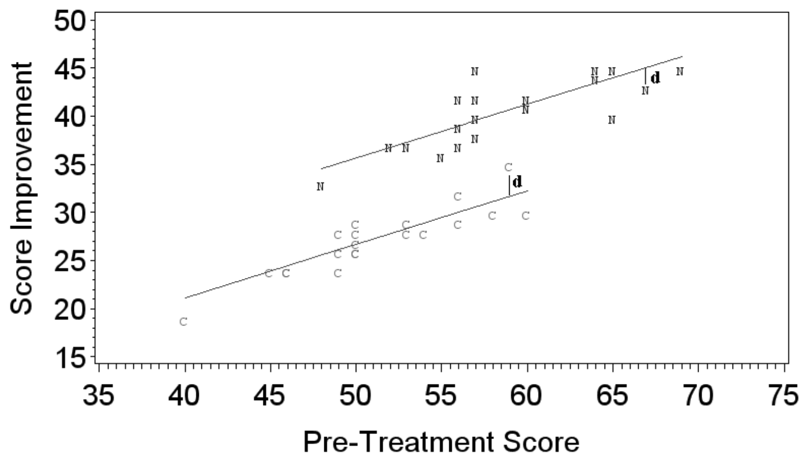

Figure 3c. Introducing the Covariate: Plot of Student Score Variation around their Class Covariate Line.

The best-fitting 'covariate' lines in **Figure 3c** are crucial to understanding ANCOVA. Because they allow us to compare groups using improvement scores and Pre-Treatment scores simultaneously, we can use the lines to estimate group differences in improvement while accounting for Pre-Treatment differences.

Let's see how using **Figure 4** below.

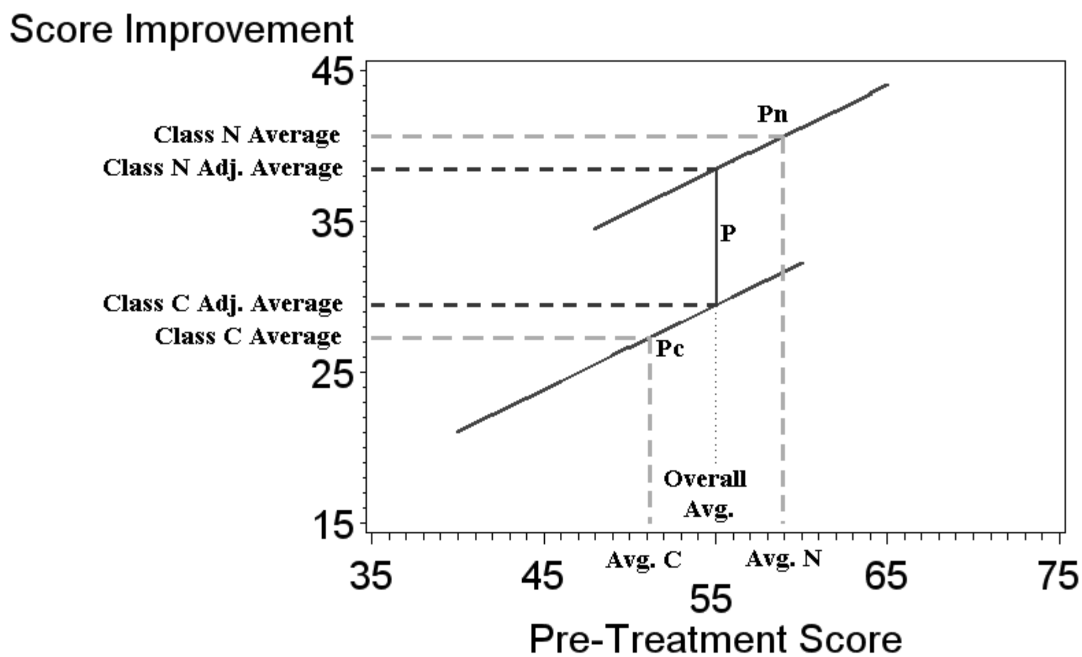

Figure 4. Graphic illustration of ANCOVA. Covariate Lines and Adjusted Means.

With **Figure 4**, we can now visually grasp the answers to our first two questions:

Answer to Question 1:

> The answer is yes, students in Class N still show improvement over Class C after we 'subtract out' class differences in Pre-Treatment scores. Relying on our best-fitting lines to represent our two classrooms, we see that Class N lies above Class C on the vertical axis, performance improvement.

Answer to Question 2:

The answer as to how much Class N improved:

After adjusting for student differences in Pre-Treatment scores, Class N differs in performance improvement by the distance between the two best-fitting lines.

To see this visually in **Figure 4**:

Choose a Pre-Treatment score (horizontal axis), say 55, and visually draw a vertical line between the two best-fitting lines. If a Class C and Class N student both scored 55 during Pre-Treatment (x-axis), they would be expected to differ in score improvement (y-axis) by the distance P between the two best-fitting lines.

Essentially, the best-fitting lines allow us a way to conceptually match students on Pre-Treatment scores and then estimate their difference in improvement.

Because the two best-fitting lines are parallel, this same level of improvement can be expected regardless of Pre-Treatment score.

Least Squares Means

Related to Question 2 is the concept of least squares (adjusted) means. Often researchers are interested not only in the amount of Post-treatment difference between groups after adjusting for the covariate (our Pre-Treatment scores), but also the group averages after similarly adjusting. Least Squares means are the average class scores after ‘subtracting out’ Pre-Treatment differences. Let's visually derive these on **Figure 4**, again using our best-fitting lines.

The dashed lines meeting at point PN and PC relate the unadjusted improvement group averages to the Pre-Treatment group averages.

Note that these improvement averages are related to different average Pre-Treatment scores. We want to equate Classes N and C on Pre-Treatment differences, and then find the related improvement averages.

To do so in ANCOVA, we shift point PC upward along its best-fitting line and point PN downward along its line, until they meet at point P, midway between the two Pre-Treatment averages. They are now matched on Pre-Treatment (x-axis), and we simply read the improvement scores (y-axis) corresponding to the best-fitting lines.

Note that the adjusted class averages are closer in value than are the non-adjusted means, indicating that the greater improvement in Class N is in part due to higher Pre-Treatment scores in Class N.

QUANTIFYING CONFIDENCE: STATISTICAL SIGNIFICANCE

To address Question 3, i.e., our level of confidence that the observed greater improvement in Class N is not due to chance, let's first understand statistical significance.

We’ve seen (**Figures 2a vs. 2b**) that intuitively, the less variation around the class best-fitting lines, the greater is our confidence of a true class difference that is not due to chance.

It also makes intuitive sense that greater the distance between class best fitting lines, the greater is our confidence that there is a true difference. This intuitive reasoning is the very logic around statistical significance, as embodied in the F statistic.

Let's understand this logic, first without considering our covariate. Conceptually, for **Figure 3b**:

F	represents	variation between group averages / variation (the 'd' scores) around each group average
As the value of F increases (increase in numerator, and/or decrease in denominator), our confidence increases that our group differences are not due to chance.		

The numeric value of F is directly related to the mathematical probability of finding a given difference (variation) between group averages as a result of chance.

And if you know the group averages and the 'd' values, you can calculate F and the corresponding probability.

Normally, if the probability is small, < .05, we conclude our groups differed due to our treatment and not to chance.

Statistical Significance in ANCOVA

As we've seen in **Figure 3b vs. 3c**, in ANCOVA we replace the lines representing group averages (**Figure 3b**) with the covariate best-fitting lines (**Figure3c**). Correspondingly, we now use the following F statistic:

F_{cov}	represents	variation related to the *adjusted* group means*** / variation (the 'd' scores) around each group *covariate* line
***(Calculation of this variation is a bit more involved, because there is variation around the group means and around the covariate lines to consider.)		

Recall that in our study the difference between the adjusted means is smaller than the difference between the nonadjusted averages.

You may think that this reduces our chance of finding any actual improvement due to the new teaching method (since this makes the F numerator smaller).

However, this reduced difference is examined using a much smaller denominator (variation around the group covariate lines is much smaller than around the group averages), yielding a much more sensitive test (higher F value) of class differences due to teaching method. The greater the value of F, the lower the probability that the observed group differences are due to chance.

Let's address Question 3 in the next section using the F_{cov} statistic computed by PROC GLM.

PROC GLM ANALYSIS

```
data all;
   input post_treat pre_treat classroom $;
   cards;
24 46 C
24 49 C
26 50 C
       :
       :
       :
45 64 N
37 52 N
38 57 N
;
run;

proc glm data=all;
   class classroom;
   model post_treat=pre_treat classroom;
   lsmeans classroom /pdiff out=adjmeans;
run;
```

Letters on the following printouts correspond to the descriptions below.

Analysis of Covariance

The GLM Procedure

Dependent Variable: post_treat

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2090.979154	1045.489577	295.47	<.0001
Error	37	130.920846 **A**	3.538401		
Corrected Total	39	2221.900000			

R-Square	Coeff Var	Root MSE	y Mean
0.941077	5.540689	1.881064	33.95000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
pre_treat	1	1579.671382	1579.671382	446.44	<.0001
classroom	1	511.307773	511.307773	144.50	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
pre_treat	1	322.0791544	322.0791544	91.02	<.0001
classroom	1	**B** 511.3077729	511.3077729	**C** 144.50	**D** <.0001

Analysis of Covariance

The GLM Procedure

Least Squares Means

method	y LSMEAN **E**	H0:LSMean1= LSMean2 Pr > |t|
C	29.4581475	<.0001 **F**
N	38.4418525	

A: The 'd' values around the covariate lines in **<u>Figure 3c</u>** after first squaring each and then summing all of them together

- This is the basis for the denominator in our F_{cov}
- (The values are squared before summing to meet the mathematical requirements for calculating probability)

B: Variation involving the difference between adjusted group means

- This is the basis for our F_{cov} numerator

C: Our F_{cov}

D: The probability that the adjusted means differ due to chance

- If a study has 3 or more adjusted means (e.g., 1 old and 2 new teaching methods), this is the probability that *any* difference among groups is due to chance
- If the probability for F_{cov} is acceptably low (e.g., < .05), then use the information described in Item **F** below to identify where the difference among means is occurring

E: Least Squares (adjusted) Means

F: For each possible pair of least square means in the study, the probability that the difference between them is due to chance (We have only 2 means in our study)

<u>Answer to Question 3:</u>

Looking at **F** above, we can be highly confident that the observed difference between Class N and Class C adjusted means is not due to chance (chance probability is less than 0.0001).

Referring to sections **E** and **F** on the GLM output, we can conclude that after subtracting out the influence of Pre-Treatment scores, there is only a 0.0001 probability that the observed improvement (9-point difference in adjusted means) in Class N over Class C results from chance differences between groups.

For the 150-point tests used in this study, this translates to a 6% improvement attributable to the new teaching method.

CONCLUSION OF THE ANCOVA ANALYSIS

After adjusting for initial performance differences, students subsequently receiving the new teaching method exhibited a 6% improvement in test results over students receiving the traditional method of instruction. This difference is statistically significant (p < .0001).

ASSUMPTIONS IN ANCOVA

ANCOVA relies on a number of assumptions. Many of these are common across a variety of statistical analysis procedures. Let's instead focus on those additional assumptions specific to ANCOVA:

- ANCOVA assumes that the covariate lines do not have a slope of zero, i.e., they are not horizontal. In our study, flat (slope=0) covariate lines would imply that Pre-Treatment scores are not related to improvement.
  - That is, performance (vertical axis) remains the same in each group across Pre-Treatment scores (horizontal axis).
  - Hence the covariate line for each group would perfectly overlap the group average lines shown in **Figure 3b**.
  - As such, introducing a covariate offers no useful adjustment to the group averages because it will not subtract out any additional variation in student improvement scores

- ANCOVA assumes that the covariate lines for the groups are parallel to each other. If the covariate lines are not parallel, then the amount of adjustment needed to subtract out Pre-Treatment performance depends on the student's Pre-Treatment score itself.
  - In **Figure 5** below, only students scoring 55 during Pre-Treatment will have Pre-Treatment adjustments equal to the adjusted averages calculated by ANCOVA.
  - Hence, the calculated adjusted means cannot adequately represent all students in the study.

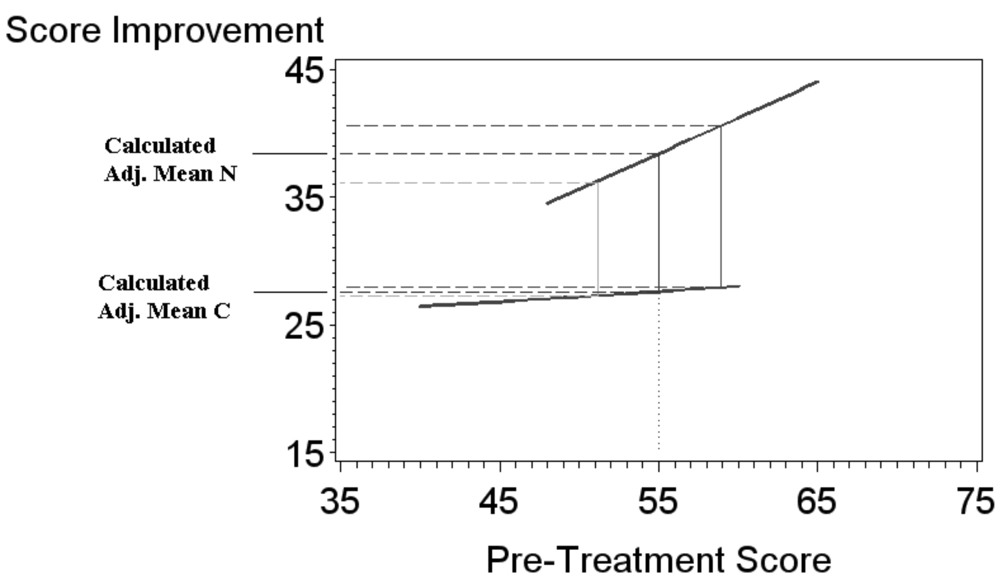

Figure 5. Graphic illustration of Non-Parallel Covariate Lines. Calculated Adjusted (Least Squares) Means No Longer Represent All Students.

- ANCOVA assumes that the treatment does not affect the covariate. If the treatment affects the covariate, the location of the covariate line for the treatment group is altered.
  - In our study, we do not violate this assumption.
    - Pre-Treatment scores (the covariate) are collected before students ever receive the new teaching method.
  - But for demonstration purposes, let's consider some hypothetical study where the Treatment affects the covariate.

- **Figure 6** below shows an example of this visually. If each Treatment Group member's covariate score is increased slightly, the covariate line is shifted to the right from A_1 to A_2.
- Line A_2 is closer to Line B than is A_1. This implies that Line A_2's calculated adjusted mean no longer measures the 'true' original distance between the covariate lines A and B.

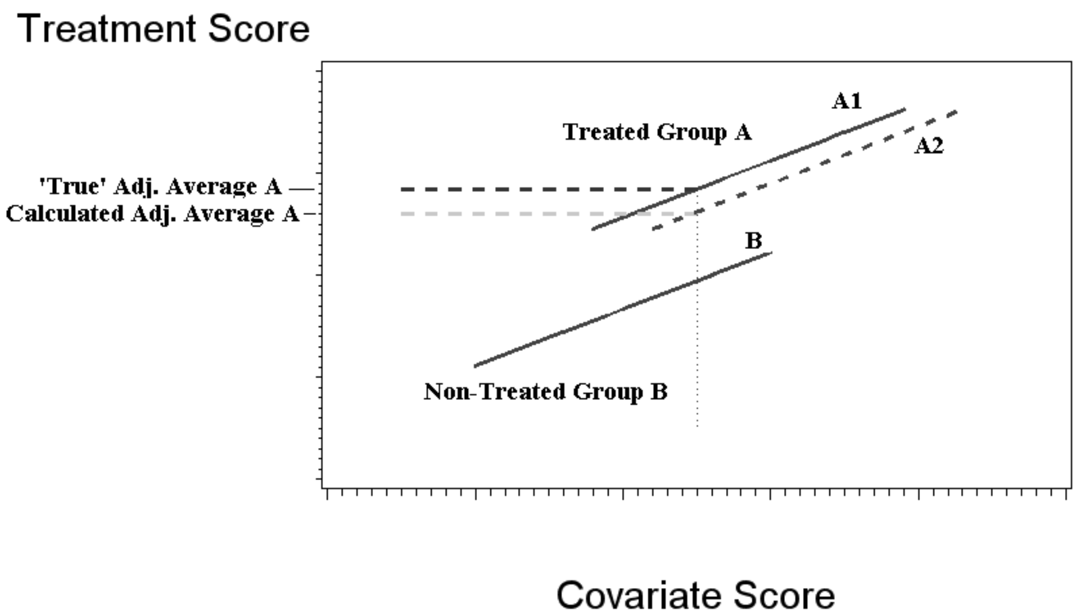

Figure 6. Hypothetical Covariate Line Shift Due to Treatment Influencing the Covariate (Pre-Treatment Score). Calculated Adjusted (Least Squares) Mean No Longer Represents the Treatment Group.

CONCLUSION

Analysis of Covariance offers researchers a valuable statistical procedure for subtracting out Pre-Treatment differences between non-equivalent groups, allowing a more accurate estimate of the impact of the treatment given to one group but not the other(s).

By positioning each group member on a Treatment by Pre-Treatment (covariate) plot and identifying the line that best fits these points for each group, we can visually grasp the strategy ANCOVA uses to estimate the Treatment effect and the corresponding level of statistical significance after accounting for Pre-Treatment group differences. As we've also seen in this poster, this visual approach can be extended to understand the assumptions ANCOVA must rely upon to yield interpretable results.

This visual and intuitive approach in turn offers a framework for interpreting ANCOVA results produced from PROC GLM, and allows us as programmers to better appreciate the research in which we participate.

REFERENCES

Lentner, M. and Bishop, T. 1986. *Experimental Design and Analysis*, Blacksburg, VA: Valley Book Co.

Winer, B.J. 1971. *Statistical Principles in Experimental Design*, 2nd ed., New York, NY: McGraw-Hill.

CONTACT INFORMATION

Name: Michael Senderak
Enterprise: Merck & Co., Inc.
Address: 351 North Sumneytown Pike
City, State ZIP: North Wales, PA 19454
E-mail: michael.senderak@merck.com