

4.4: Analysis of Covariance (ANCOVA)

As mentioned earlier this chapter, there are two ways to add control variables into a research study. One is through design, such as randomized block design. The other is through statistical control, known as analysis of covariance. The control variables are called covariates. Covariates usually have an impact on the dependent variable and thus can be included into an ANOVA analysis. We can use this technique analysis of covariance (ANCOVA).

Analysis of Covariance (ANCOVA)

Recall in randomized block design/ANOVA, we would utilize the control variable in the design stage by sorting participants and grouping them into different blocks based on the control variable. By doing that, the participants will be similar on the control variable before they are assigned into different treatment or intervention groups. However, in ANCOVA, we don't do anything with the control variable, or the covariate, in the design stage. Instead, we simply collect the data on the covariate along with other data we would collect from participants, and analyze the covariate during the data analysis stage. That's why this is considered a statistical control technique.

Still using the same example as used in randomized block design, we are conducting an experiment on the effect of cell phone use (yes vs. no) on driving ability and include driving experience (as measured by months of driving) as a control variable. Instead of grouping participants into different blocks based on their driving experience, in ANCOVA, we would treat driving experience as a covariate and simply collect data on it and analyze it using ANCOVA technique. Or another example is, say we want to test the impact of different teaching methods on students' performance in an introductory calculus course. Previous research has established that math scores on college entrance test impacts students' performance in calculus courses. Therefore, we would include ACT math scores as the covariate. Notice that, in both cases, driving experience (as measured by months of driving) and ACT math scores, are intervally scaled variables. This is typically the expectations of ANCOVA, to have the covariate on an interval scale. As a matter of fact, we would expect there to be a linear relationship between the covariate and the dependent variable. More on this in a little bit.

ANCOVA is an extension of ANOVA. The main advantage of using ANCOVA over using ANOVA is that by adding covariates into the study/model, we are minimizing the effect of the covariates on the dependent variable. Recall, the covariates are known to have an influence on the dependent variable, which is why they are included in the study in the first place. By controlling for the effect of covariate, we are reducing its threat to confound the results, and this gives us more confidence to establish that the intervention, or the independent variable, causes the change in the dependent variable. In the example above, by controlling for the effect of driving experience on driving ability, we are more certain it is the cell phone use that causes the change in driving ability.

How to Use ANCOVA

The analyses of ANCOVA is fairly complex. Without getting into the details of computations, this section provides a brief overview of how to use ANCOVA.

Just like ANOVA, ANCOVA uses Fisher's F test. Therefore, the key to understand ANCOVA is still partitioning of the variance. When adding a covariate into the study, it essentially becomes another predictor in the model, even though it is not the researchers' main focus or interest. So we will partition out the variance that can be explained by this variable. As a result, there will be three parts of the variance in ANCOVA, SS intervention, SS covariate, and SS error, and together they make up SS total. If we were to compare this to ANOVA, the difference is SS covariate. Without the covariate, ANOVA has two parts of variance, SS intervention and SS error. By adding the covariate, we partition out some of the error variance and attribute it to the covariate. In doing so, the error variance will be reduced. As we have seen many times, in F tests, we look at the ratio of effect and error. When the numerator (i.e., effect) decreases, the calculated F is going to be larger. We will achieve a smaller P obtained value, and are more likely to reject the null hypothesis. In other words, good covariates decrease error, which increases statistical power. This is another main advantage of ANCOVA (besides control mentioned above), assuming the covariate we selected is a decent one based on theoretical/empirical evidences.

As mentioned above, covariates should be either interval or ratio. That's because ANCOVA essentially uses a linear regression model. With the linear model, the computation can be rather complicated. But conceptually, by including the covariate into the model, ANCOVA adjusts each group mean on the outcome variable. Using the same example of studying the effect of cell phone use (yes vs. no) on driving ability with driving experience as a covariate, it is possible that one of the treatment group (the no cell phone use) happens to be higher on the covariate, that is, have more driving experience than the other treatment group (cell phone use).

use). Accounting for that, ANCOVA will lower the no-cell-phone-use group's average score on the dependent variable, driving ability. You probably have guessed, since the other group (cell phone use) is lower on the covariate, that is, less driving experience, ANCOVA will increase its group average score on the dependent variable, driving ability. Mathematically, this allows us to compare the means of the treatment groups at the mean/average value of the covariate. In other words, the treatment groups in the study will be "adjusted" for the linear model, so that the "playing field is leveled". By doing so, ANCOVA allows us to find the best estimates of how different treatment groups would have scored on the dependent variable if they all had statistically equivalent means on the covariate.

Assumptions of ANCOVA

ANCOVA shares the assumptions of ANOVA. In addition, there are three assumptions that are unique to ANCOVA.

First, in ANCOVA, the independent variable and the covariate must be independent from each other. In other words, the levels or groups of the intervention/treatment should have no influence on the covariate. In the example above, this means cell phone use treatment groups (yes vs. no) should be independent of, or have no influence on, driving experience. When this assumption is violated, the effect of the independent variable (treatment) and the effect of the covariate overlaps. Translate that into statistical calculations, the treatment and the covariate would share some of the variance. This will skew the analysis and make the results biased. When covariate driving experience is affected by the independent variable cell phone use treatments (yes vs. no), adding driving experience into the model as a covariate does not control for the differences between treatment groups on the dependent variable driving ability. The ANCOVA results will be inaccurate.

Second, the relationship between the covariate and the dependent variable must be linear. In the example above, this means driving experience and driving ability is expected to have a linear relationship. It is critical to first examine the nature of the relationship between the covariate and the dependent variable, for example through scatter plots, before performing ANCOVA. If the relationship is not linear, the adjustments ANCOVA makes will be biased and the results will be inaccurate.

In addition, these regression lines (on the covariate and the dependent variable) from different treatment groups must be parallel to each other. In other words, different treatment groups should have similar slopes. In the above example, this mean for both groups (cell phone use and no cell phone use), the slope for the relationship between driving experience and driving ability should be similar. This assumption is called homogeneity of regression slopes. This is one of the most important assumptions of ANCOVA as it allows us to "adjust" for the group means. If this assumption is violated, it means there is an interaction between the independent variable and the covariate. In this case, ANCOVA will be biased and the results will be inaccurate.

References

Field, A. (2013). *Discovering Statistics Using IBM Statistics*. London: Sage Publications.

This page titled [4.4: Analysis of Covariance \(ANCOVA\)](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Yang Lydia Yang](#).