Analysis of Co-variance

* Analysis of co-variance (ANCOVA) is a statistical test related to ANOVA
* It tests whether there is significant difference between the groups after controlling for variance explained by a co-variate
* A covariate is a continuous variable that correlates with the dependent variable.

The ANCOVA model

ANCOVA evaluates whether the means of a [dependent variable](https://en.wikipedia.org/wiki/Dependent_variable) (DV) are equal across levels of a categorical [independent variable](https://en.wikipedia.org/wiki/Independent_variable) (IV) often called a treatment, while statistically controlling for the effects of other continuous variables that are not of primary interest, known as [covariates](https://en.wikipedia.org/wiki/Covariate) (CV) or nuisance variables. Mathematically, ANCOVA decomposes the variance in the DV into variance explained by the CV(s), variance explained by the categorical IV, and residual variance. Intuitively, ANCOVA can be thought of as 'adjusting' the DV by the group means of the CV(s)

The ANCOVA model assumes a linear relationship between the response (DV) and covariate (CV):

*Yij = + Ti + B(xij - x̄ ) + €ij*

where *x̄* represents the overall mean of 𝑋. If 𝑋 and 𝑌 are closely related, you would expect that the errors, 𝑒2𝑖j, would be much smaller than the errors, 𝑒1𝑖j, giving you more precise results.

The classical analysis of covariance is useful for many reasons, but it does have the (highly) restrictive assumption that the slope is constant over all the groups. This assumption is often violated, which limits the technique’s usefulness.

Assumptions

The following assumptions are made when using the F-test.

1. The response variable is continuous.

2. The treatments do not affect the value of the covariate, 𝑋𝑖j.

3. The 𝑒2𝑖j follow the normal probability distribution with mean equal to zero.

4. The variances of the 𝑒2𝑖j are equal for all values of i and j.

5. The individuals are independent.

Limitations

There are few limitations when using these tests. Sample sizes may range from a few to several hundred. If your data are discrete with at least five unique values, you can assume that you have met the continuous variable assumption. Perhaps the greatest restriction is that your data comes from a random sample of the population. If you do not have a random sample, the F-test will not work.