

9.2: ANCOVA in the GLM Setting - The Covariate as a Regression Variable

In this section, we will develop the statistical ANCOVA, which by definition is a general linear model that includes both ANOVA (categorical) predictors and regression (continuous) predictors. The simple linear regression model is:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (9.2.1)$$

where β_0 and β_1 are the intercept and the slope of the line, respectively. The significance of a regression is equivalent to testing $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ using the F statistic: $\frac{MS(Regr)}{MSE}$ where $MS(Regr)$ is the mean sum of squares for regression and MSE is the mean squared error. In this case of a simple linear regression, this test is equivalent to a t-test.

Now, in adding the regression variable to our one-way ANOVA model, we can envision a notational problem. In the balanced one-way ANOVA, we have the grand mean (μ), but now we also have the intercept β_0 .

To get around this, we can use

$$X^* = X_{ij} - \bar{X} \quad (9.2.2)$$

and get the following as an expression of our covariance model:

$$Y_{ij} = \mu + \tau_i + \gamma X^* + \epsilon_{ij} \quad (9.2.3)$$

Note that the above model fits into the general linear model (GLM) and the Type III (model fit) sums of squares for the treatment levels in this model are being corrected (or adjusted) for the regression relationship. This has the effect of evaluating the treatment levels "on the same playing field", that is, comparing the means of the treatment levels at the mean value of the covariate. This process effectively removes the variation due to the covariate that may otherwise be attributed to treatment level differences.

This page titled [9.2: ANCOVA in the GLM Setting - The Covariate as a Regression Variable](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Penn State's Department of Statistics](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.