13.3: ANOVA Table

Our ANOVA table in regression follows the exact same format as it did for ANOVA (hence the name). Our top row is our observed effect, our middle row is our error, and our bottom row is our total. The columns take on the same interpretations as well: from left to right, we have our sums of squares, our degrees of freedom, our mean squares, and our F statistic.

Table 13.3.1: ANOVA table in regression

Source	SS	df	MS	F
Model	$\sum(\widehat{Y}-\overline{Y})^2$	1	SS_M/df_M	MS_M/MS_E
Error	$\sum(Y-\widehat{Y})^2$	$N-2$	SS_E/df_E	
Total	$\sum(Y-\overline{Y})^2$	$N-1$		

As with ANOVA, getting the values for the SS column is a straightforward but somewhat arduous process. First, you take the raw scores of X and Y and calculate the means, variances, and covariance using the sum of products table introduced in our chapter on correlations. Next, you use the variance of X and the covariance of X and Y to calculate the slope of the line, b, the formula for which is given above. After that, you use the means and the slope to find the intercept, a, which is given alongside b. After that, you use the full prediction equation for the line of best fit to get predicted Y scores ($\widehat{Y}$) for each person. Finally, you use the observed Y scores, predicted Y scores, and mean of Y to find the appropriate deviation scores for each person for each sum of squares source in the table and sum them to get the Sum of Squares Model, Sum of Squares Error, and Sum of Squares Total. As with ANOVA, you won't be required to compute the SS values by hand, but you will need to know what they represent and how they fit together.

The other columns in the ANOVA table are all familiar. The degrees of freedom column still has $N-1$ for our total, but now we have $N-2$ for our error degrees of freedom and 1 for our model degrees of freedom; this is because simple linear regression only has one predictor, so our degrees of freedom for the model is always 1 and does not change. The total degrees of freedom must still be the sum of the other two, so our degrees of freedom error will always be $N-2$ for simple linear regression. The mean square columns are still the SS column divided by the df column, and the test statistic F is still the ratio of the mean squares. Based on this, it is now explicitly clear that not only do regression and ANOVA have the same goal but they are, in fact, the same analysis entirely. The only difference is the type of data we feed into the predictor side of the equations: continuous for regression and categorical for ANOVA.