

The ANOVA Table for Regression

Supplement to Section 11.4 - B. Habing, 11/14/16

Just like the ANOVA table can be used to test the null hypothesis that $\mu_1 = \mu_2 = \dots = \mu_k$ when looking at the means of several groups, it can also be used in regression. In the case of means, the null hypothesis means that knowing the group membership provides no extra information about y . In the case of regression, the corresponding null hypothesis would be that knowing x provides no extra information about y . If we were to guess the same y value for every x , that would mean that the regression line was flat, that it had no slope. Therefore, the null hypothesis for the ANOVA table in regression is $H_0: \beta_1 = 0$ and the alternate hypothesis is $H_A: \beta_1 \neq 0$.

The main idea in setting up the ANOVA table for regression is that instead of comparing the individual observations to the group averages \bar{y}_i (our best guess for the observations in each group), we will compare them to $\hat{\beta}_0 + \hat{\beta}_1 x_i$ (the best guess for the observation at each x). In other words, we'll basically use the same table we used before and just replace the \bar{y}_i by the $\hat{\beta}_0 + \hat{\beta}_1 x_i$ and make sure the degrees of freedom work out correctly. When calculating the degrees of freedom now, instead of using k (one for each group average) we use 2 (one for each β). It is almost easier to do because we only have one subscript to worry about instead of two.

An example of a completed ANOVA table for regression can be seen in Figure 11.8 on page 605 and in Figure 11.14 on page 5610. Note that the "Root MSE" in the SAS output is the same square root of the MSE in the ANOVA table (this is called "S" in the MINITAB output). Also note that the p-value from the ANOVA table is the same as the p-value for the t-test for the slope variable (0.0354) and that the $t=3.66$ is the square root of the $F=13.36$. The t-test for $\beta_1 \neq 0$ will always agree with the F-test for $\beta_1 \neq 0$. The t-test has the advantage that it can be made to test $>$ or $<$. The ANOVA table has the advantage that it will be useful in many other situations in STAT 516.

The ANOVA Table for Comparing Means

Source	SS (<i>Sum of Squares, the numerator of the variance</i>)	DF (<i>the denominator</i>)	MS (<i>Mean Square, the variance</i>)	F
Treatment (or Between or Model)	$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2$	$k-1$	$MST = \frac{SST}{k-1}$	$F = \frac{MST}{MSE}$
Error (or Within)	$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$n-k$	$MSE = \frac{SSE}{n-k}$	
Total	$SS(Total) = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$n-1$		

The ANOVA Table for Regression

Source	SS (<i>Sum of Squares, the numerator of the variance</i>)	DF (<i>the denominator</i>)	MS (<i>Mean Square, the variance</i>)	F
Regression (or Model)	$SSR = \sum_{i=1}^n ((\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{y})^2$	$2-1=1$	$MSR = \frac{SSR}{1}$	$F = \frac{MSR}{MSE}$
Error	$SSE = \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$	$n-2$	$MSE = \frac{SSE}{n-2}$	
Total	$SS(Total) = \sum_{i=1}^n (y_i - \bar{y})^2$	$n-1$		

Because they are made in the same way, the five keys we saw for the ANOVA table for means also apply to the ANOVA table for regression.