

15.3: Hypothesis Testing- Slope to ANOVAs

In regression, we are interested in predicting Y scores and explaining variance using a line, the slope of which is what allows us to get closer to our observed scores than the mean of Y can. Thus, our hypotheses can concern the slope of the line, which is estimated in the prediction equation by b .

Research Hypothesis

Specifically, we want to test that the slope is not zero. The research hypothesis will be that there is an explanatory relation between the variables.

- RH: $\beta > 0$
- RH: $\beta < 0$
- RH: $\beta \neq 0$

A non-zero slope indicates that we can explain values in Y based on X and therefore predict future values of Y based on X .

Null Hypothesis

Thus, the null hypothesis is that the slope is zero, that there is *no* explanatory relation between our variables

$$\text{Null Hypothesis : } \beta = 0$$

Regression Uses a ANOVA Summary Table

Did you notice that we don't have a test statistic yet (like t , F of ANOVA, or Pearson's r yet? To test the null hypothesis, we use the F statistic of ANOVA from an ANOVA Summary Table compared to a critical value from the F distribution table.

Our ANOVA table in regression follows the exact same format as it did for ANOVA (Table 15.3.1). 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 15.3.1: ANOVA Table for Regression

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

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 (\hat{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 have for the IV (predictor): a quantitative variable for for regression and groups (qualitative) for ANOVA. The DV is quantitative for both ANOVAs and regressions/correlations.

With a completed ANOVA Table, we follow the same process of null hypothesis significance testing by comparing our calculated F-score to a critical F-score to determine if we retain or reject the null hypothesis. In ANOVAs, the null hypothesis was that all of the means would be similar, but with correlations (which are what regression is based on), the null hypothesis says that there is no linear relationship. However, what we are really testing is how much variability in the criterion variable (y) can be explained by variation in the predictor variable (x). So, for regression using ANOVA, the null hypothesis is saying that the predictor variable does not explain variation in the criterion variable.

This is a little confusing, so let's take a look at an example of regression in action.

This page titled [15.3: Hypothesis Testing- Slope to ANOVAs](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michelle Oja](#).

- [13.4: Hypothesis Testing in Regression](#) by [Foster et al.](#) is licensed [CC BY-NC-SA 4.0](#). Original source: <https://irl.umsl.edu/oer/4>.
- [13.3: ANOVA Table](#) by [Foster et al.](#) is licensed [CC BY-NC-SA 4.0](#). Original source: <https://irl.umsl.edu/oer/4>.