Extra Sum of Squares and Lack of Fit Tests for ANCOVA

Brooke Coneeny

The data set we are going to use for this lecture revolves around the game of soccer. It includes a male soccer players weight, height and position

# Saturated Model

The linear models we have focused on have taken on the form

for

Using our soccer data, our linear model for regressing weight on height would take the form:

We are going to direct our focus to multiple regressions that model data sets involving a categorical explanatory variable. These variables can be described as distinct values. In the case of our soccer example, the categorical variable is which can be broken into distinct categories (midfield, defense, forward, and goalkeepers). Breaking up positions in this way gives us our , which can represented by the following:

# Lack of Fit Output

A lack of fit test is used to asses the quality of saturated models, meaning we want to assess if the fitted model adequately models the response variable. In this case, we want to know how well our above model predicts a player’s weight. When performing this test, the sum of squares error is separated into two components: and .

### Pure Error

The pure error represents, assuming a normal model, the error within each group. This error is assumed to be constant throughout all groups. In the case of our soccer example, the error within defense, forward, midfield, and goal keepers is assumed to be the same.

This error can be found by

Where is the number of distinct combinations of ’s and is the number of observations within the combination of variables. This means each represents a observation in a group . We can also say, under the null hypothesis (the linear model accurately predicts the expected value of and follows the assumptions of a normal model), that

With being the number of distinct .

### Lack-of-Fit Error

The lack-of-fit error represents the error of the saturated model. This error is calculated by finding the distance of a group mean from the corresponding line estimate.

This error can be found by the following:

Where is the number of distinct combinations of ’s. We can also say, under the same null hypothesis as above, that

With being the number of distinct and being the number of parameters.

### Max R-Squared

The max can be used to asses how well the saturated model fits the data. The formula for this value is:

In our soccer example, using this equation we get

### F-test

Using our knowledge of the pure error and lack-of-fit error from above, we can now construct an test for our saturated mode under the following hypotheses

The relationship assumed in the saturated model is reasonable, there is no lack of fit The relationship assumed in the model is not reasonable, there is lack of fit

The distribution for statistic can be found by:

Where is once again the number of distinct and is the number of parameters.

Normally, we are looking for a large F-statistic, along with a small p value, in order to reject the null hypothesis. However, a large value of this statistic would provide evidence that the linearity assumption fails and there is a lack of fit in our mode. We therefore do not want a large value of F (we do not want to reject the null hypothesis). Also note that failing to reject the null hypothesis is proof that the normal linear model is correct! It simply means our model predicts better than the intercept-only model.

Redirecting our attention to the soccer example once again, let us calculate the F-Statistic.

We can see from the output table is not significant. This means we do not reject the null hypothesis and we can assume the saturated model is reasonable.

# Effect F Test Output

The effect test is useful when dealing with models that include a categorical explanatory variable. It allows us to determine the significance of the variable as a whole, opposed to individual t-tests for the different variables within the category.

## Output

To make this more clear, let us direct our attention to the sample output gathered from our soccer data model.

### Connections Between the T-test and Effect Test: Continuous Variable

We are going to specifically examine the differences between the tests, focusing on our continuous variable: height

T-test 🡪

The result of this test can be found in the parameter estimate table. The p value is less than , which means it is significant. Therefore, given that we have the other predictors, the addition of height in our model would be valuable.

Effect F-test 🡪 The null hypothesis for the effect f-test of height is very similar to the above t-test. It is testing, given we have a model containing position indicator variables, does adding height help to predict weight more accurately?

The result of this test can be found in the effect tests table. The p value is less than , and is therefore significant. This tells us that, given we have a model with position indicator variables, the addition of height in our model would be valuable.

It is important to note that the t-test and partial f-test for height are testing the same hypothesis and that the f-statistic (358.8653) is the squared value of the t-statistic (18.94).

### Connections Between the T-test and Effect Test: Categorical Variable

Now let us direction our attention to the difference between the individual t-tests for each of the indicator variables and the effect test of the position variable as a whole.

In our parameter estimate table we can see there is a t-test performed for each position indicator. Looking at the goalkeeper variable specifically, our t-test is testing

and

As we can see from our table, the result was significant (p-value = 0.0105), telling us that including the goalkeeper indicator in our model is beneficial.

In our effect tests output, we can see that there is a f-test performed for the position variable as a whole. This test is different than performing a t-test for each position indicator.

This f-test is testing, given a model with height as an explanatory variable, is the addition of all position indicators (as a whole) useful for predicting the weight of male soccer players? We can see from the results that the p-value (0.0021) is less than 0.05, and therefore significant. The f-test informs us that although not all of the position indicators were significant for their individual t-tests, the inclusion of these predictors in the model as a whole is significant.

### Extra Sum of Squares F Statistic

To find the extra sum of squares f-statistic for position, we need to find the sum of squares for position.

# Interaction Model

We have been using a saturated model throughout this lecture which includes a height predictor alongside indicator variables for each of the positions in soccer. Let us now examine an interaction model, using the same data set.

### Why Add Interactions

Interactions are used in models when the extent to which is associated with a change in an depends on the value of a different . In the case of our soccer example, we are creating an interaction between height and position because we believe some of the relationship between weight and height is dependent on position. To create an interaction term, we simply multiply the two explanatory variables together.

Adding interactions to our above model would give us the following ’s:

The mean for individuals in each position are:

### Extra Sum of Squares F Statistic

Similar to the work we have done above, we are going to perform an effect f-test on our model. An effect f-test looks at a subset of explanatory variables and assess whether they are useful for predicting the response variable. In this case, the effect f-test will determine whether the subset of interaction variables are beneficial when predicting the weight of male soccer players.

The f-statistic for this test is equal to

Using this equation, we get the following calculations:

Giving us an f-stat =

This directly matches the f-statistic in the effect test output, confirming our calculations. As we can see in the output, this f-statistic is not significant. This means we cannot reject the null hypothesis that the for the interaction terms are all 0 (). Using an f-test in this example, opposed to a t-test, was helpful because it allowed us to determine if the interactions as a whole were significant, whereas a t-test could only look at the terms individually.