Looking at the questions for this week’s discussion board, I wanted to take an approach towards the question, how do you assess regression models? Regression models are a common use within the field and can be applied in several situations. However, they do not have a standard procedure for reviewing the accuracy of them once they have been fitted for a model. When focusing on a regression model and determining if it reaches its mark, it should be better than the fit of the mean model, which use the mean for every predicted value. (Grace-Martin, 2020) Since we want to ensure that our regression model stands over the mean model, how do we check that we are on the right path?

Author Karen Grace-Martins suggests three statistics to evaluate how well a model is fitted. The first method available to us is using R-squared and Adjusted R-Square. To find this value, you will obtain the difference of the sum of squares total and the sum of squares error follow by dividing that difference by the original sum of squares total. With this answer, we can see the proportional improvement in prediction from the regression model compared to the mean model which will indicate the goodness of fit for the model. (Grace-Martin, 2020) Now, with the Adjusted R-Square aspect, it will use the degrees of freedom to increase or decrease the fit of the model because with R-squared it can only reach an optimal value if it applies more or less predictors.

Looking at the next method at our disposal, we can perform an F-test for our regression model. For this test, it will evaluate the null hypothesis that all regression coefficients are equal to zero versus the alternative that at least one is not. (Grace-Martin, 2020) The F-test typically will play together with the R-squared that we mentioned within the first method. The test will determine if the R-squared that we calculated is strong and not a random value because of the data set that we have at the time. In general, the F-test will determine whether the relationships between response variables and the predictors are statistically reliable and can useful when you are determining prediction or explanation. (Grace-Martin, 2020)

Lastly, the third method for assessing is RMSE which stands for Root Mean Square Error. Within this method, it is searching for the most absolute fit of the model to the data. Moving along in the process, we want to investigate if the standard or observed data points are similar to the values that we are predicting using the model. In short, we want to make sure that we are on the right path and not predicting values that are not close at all to the actual values. When calculating RMSE, it is the square root of the variance of the residuals. After obtaining this value, we want to make sure that it is a lower value since that will indicate that it is a better fit of the model. RMSE is a good measure of how accurately the model predicts the response. (Grace-Martin, 2020)

At the end of the day, each one of these methods may or not may be the best approach for determining whether the model is accurate in predicting certain values. Because of this belief, it is up to those that are fitting this model to determine what will be the most useful statistic to use to measure the accuracy. Validation and evaluation is key to data science models. (Karbhari, 2018)

References:

Grace-Martin, K. (2020, January 16). Assessing the Fit of Regression Models. Retrieved July 08, 2020, from <https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/>

Karbhari, V. (2020, February 26). How to evaluate regression models? Retrieved July 08, 2020, from <https://medium.com/acing-ai/how-to-evaluate-regression-models-d183b4f5853d>

Malato, G. (2019, April 17). How to measure the goodness of a regression model. Retrieved July 08, 2020, from <https://towardsdatascience.com/how-to-measure-the-goodness-of-a-regression-model-60c7f87614ce>