Chrissa Muller

Assignment (regression extensions)

1. The attached wages.xlsx dataset contains data on the hourly wage, age, and education of 80 workers at a single manufacturing firm. You are interested in whether age has a quadratic effect on predicted wages.

> a. Plot Wage against Age and evaluate whether a linear or quadratic model would better capture the relationship.

I believe that a quadratic model would better capture the relationship between wage and age because of the upside down U shape. Wages tend to increase with age in the beginning of someone's career and then peaks high around the middle of their career and slowly declines towards the end of their career and older ages.

> b. Estimate a multiple regression model of Wage using Age and Education as independent (X) variables; assume a standard linear relationship between Wage and Age.

This multiple regression model showed us that even though age and education are both significant predictors of wage, it did not fully show the curve that was displayed in the previous scatterplot. The residuals did show a noticeable pattern.

> c. Estimate another multiple regression model of Wage using Age and Education as independent (X) variables; this time fit Age using a quadratic relationship. Verify your choice from part a. by comparing the distribution of residuals and the goodness of fit between the models in parts b and c.

After doing this second regression model, it does confirm that a quadratic model has a substantially better fit than a linear model in this case for age. The higher R squared value helped me confirm this and the distribution is more symmetric and I don't notice any patterns in the residuals. The R squared value for the linear model is 0.6186944, while for the quadratic model it is 0.8323164.

> d. Use the appropriate model to predict hourly wages for someone with 16 years of education and age equal 30, 50, or 70.

I used the quadratic model and the hourly wages for someone who has 16 years of education is about $25.85 at age 30, about $31.54 at age 50, and about $26.56 at age 70.

> e. According to the model, at what age will someone with 16 years of education attain the highest wages?

Using the quadratic model again, someone with 16 years of education will attain the highest wages at about age 50.67. I used the coefficients and turning point formula to find this age value.

2. The AnnArbor.xlsx file contains data on a portion of the rental market in Ann Arbor, MI. The data includes the monthly rent, number of bedrooms, number of bathrooms, and the square footage of 40 rental properties.

     a. Plot Rent against each of the three predictor variables and evaluate whether the relationship is best captured by a line or a curve. Identify variables that may benefit from a log-transformation.

After plotting rent against the three predictor variables, I believe that the relationship is better captured by a slight curve. Not all of the relationships show linear relationships meaning that a straight line wouldn't be as ideal. I think square footage and even rent itself would benefit from a log transformation and help improve the model overall.

     b. Estimate a multiple regression model (with any appropriate log-transformations) to predict rent for a 1,600-square-foot rental with 3 bedrooms and 2 bathrooms.

After using log-transformations for square footage and rent and running a multiple regression model with all the variables, the predicted rent for a 1,600-square-foot rental with 3 bedrooms and 2 bathrooms is $1468.24.