George Briggs

Test 2v1 Corrections

Problem 4
Answer: A, B, D
Explanation: These points are all above or below the green horizontal lines.

Problem 5
d)
There is a statistically significant relationship between smoking status and life expectancy. This is because the term ‘SmokerYes’ has a p value of 2e-16 which is smaller than even an alpha value of .01. This suggests that there is a statistically significant relationship or that the coefficient is not likely equal to 0.

Problem 6
b)
Null Hypothesis: There is no relationship between all our coefficients (weight, transmission_type, weight*transmission_type) and our response variable, mpg. In other words, all our coefficients are equal to 0.
Alternate hypothesis: Not all of our coefficients are simultaneously equal to zero.

Problem 7
a)
Null Hypothesis: Our categorical variable, Transimission_type, has no correlation with our dependent variable (mpg) and therefore, the coefficient is equal to 0.
Alternative Hypothesis: Our categorical variable, transmission_type, is correlated with our dependent variable and therefore changes in its value reflect changes in mpg, so its value is not equal to 0.

With a p-value of 0.0437, we can reject the null hypothesis in favor of the alternative with an alpha value of .05.

b)
The test in 7a is more reliable than the test in 6b. We have to look at our MLR assumptions when deciding which model is more reliable. In model 6b, we have multiple predictor variables that could be at risk of multicollinearity. This would not be a concern in our model in 7a since there is only one predictor term. We would also need to analyze our residual plots to make sure that they are normally distributed with constant variance and model a linear relationship. We could compare the plots of the two models when deciding which is more reliable.

Problem 8
I would add the new predictor term to our regression model then compare the adjusted R^2 term to see if it increased when we added the term.

I would also construct an added variable plot comparing the 2 sets of residuals to determine how helpful our new term is by analyzing R^2.
Finally, I would construct a scatter plot matrix to see how displacement is related to our quantitative predictors and box plots to see how displacement is related to our categorical predictors.