Liam Manikowski

Stat Task 2 Corrections: 4/21/22

MC:
2: A, D

Problem 5:
c. The average difference in lifespan between smokers and non-smokers is 23.4392 years.

Problem 6:
b. $H_0: B_1 = B_2 = B_3 = 0$ $H_A: B_i \neq 0$ (any coefficient $\neq 0$)
Our p value = $1.669e-12 < .05$
Given the significant p-value shown above, we can conclude that at least one of our coefficients are different from 0 ($\neq 0$). This information tells us that there is a linear association between the predictor variables and the response variable.

Problem 7:
a. $H_0: B_2 = B_3 = 0$
$H_A: B_2$ or $B_3 \neq 0$
b. Reliability has to do w/ the appropriateness of the model assumptions, which are that the errors are independent, there's constant variance, and that it follows a normal distribution. With respect to these tests, I would say that Model 3 is better b/c the fitted values vs. standardized residuals plot shows the strongest evidence of meeting our constant variance and normality assumptions.

Problem 8:

A first measure would be to individually test for the variables statistical significance by running a hypothesis test using its p-value.

I would also want to check the residuals plot to assess for the fit and look for if the data pts were evenly distributed above/below the zero line and make sure that there were only small differences between theoretical and observed qualities.

W/ a test on the effectiveness of the entire model, I'd pay attention to R^2 / R^2_{adj}.

If R^2 increases but R^2_{adj} decreases, we are overcomplicating the model by adding this predictor.

I'd want to look at a scatterplot of weight vs displacement to see if there's a linear relationship. If there is, we would have multicollinearity and be overcomplicating the model.