Stat 21 Test 2 v2 Corrections

Jake Bohman

(1) A, D

(2) B, C

(5) c) The average difference in lifespan between smokers and non-smokers is 23.44 years, accounting for alcohol consumption, since SmokerYes decreases the expected life expectancy by this much, meaning that smokers on average live 23.44 less years than non-smokers, accounting for alcohol consumption.

(6) a) I would pick model 3 since it has the highest adjusted R^2 value, has the best constant variance in the fitted values vs Residuals plot, appears to be the most normal in the normal quantile plot, and has low p-values for each variable.

(7) b) The test in problem 6b is more reliable, ~~since [illegible]~~ since it relies on less assumptions.

(8) I would first see if adding it to the model increases the adjusted R^2 at all, if no then immediately reject. I might check the plot of residuals for engine alone vs residuals of the other model to see if engine contributes to explaining the variability of mpg while checking for multi-collinearity. Additionally, I may check the variance inflation factor to find if the variable is related to many others to see if the variable truly adds more valuable information or not.