MTH 325, Quiz #7, Spring 2024 Name KEY

Instructions: Answer each question as thoroughly as possible. Round answers to 4 decimal places as needed. Exact answers are best when possible. Be sure to answer all parts of each question.

1. Using the same data from Quiz #5 (**325quiz5data.xlsx**), perform three types of model selection procedures to predict Credit Card Debt from the remaining numerical variables:
   a. Best subset selection
   b. Backward selection
   c. LASSO (penalized) regression (see Lab #6 for code examples)

   Report the results of the coefficients and variables in the model in each case. Compare the results using the following criteria:
   i. The R^2 value if available
   ii. The values of the coefficients
   iii. The residual standard error
   iv. The AIC and BIC
   v. Which model is the simplest (has the fewest variables)? Did any come out the same?
   vi. Using plot(modelname) in R, create diagnostic plots for each model.

   Based on this information, write a paragraph explaining how you would choose from among these models. You are free to bring in additional criteria as needed.

Best Subsets - 5 var model $R^2 = .2477$ Adj $R^2 = .2433$
4 var model $R^2 = 0.2474$ Adj $R^2 = 0.2439$
3 var model $R^2 = 0.2467$ Adj $R^2 = 0.2441$
2 var model $R^2 = 0.2461$ Adj $R^2 = 0.2447$ Age + Mall Trips
1 var model $R^2 = 0.2457$ Adj $R^2 = 0.2448$ Mall Trips

Backward Selection settled on 1 var model Mall Trips
Only one with all coeff. significant
CCDebt = 239.90 Mall Trips + 295.34

Lasso kept Age and Mall Trips w/ $R^2 = 0.2449$

CC Debt = 234.81 * Mall Trips − 0.41327 * Age + 382.7795

The coeffs of mall trips are not that dissimilar, but the intercept is off by nearly $100.

Given that the R^2 values are all pretty low and very similar, I'd go w/ parsimony and choose the simplest one-var model