DSC 423: Data Analysis and Regression
Assignment 06: Midterm Review

Your submission must include your name and student ID. Your submission must include the honor statement: "I have completed this work independently. The solutions given are entirely my own work." Your submission must be submitted as a PDF.

1. (5 points) What is the null and alternative Hypothesis of the F-test? …of a t-test? Explain how each one can be used in the analysis of your regression model.
2. (5 points) What are the four assumptions about residuals in the regression model? Why are these assumptions made? How can you verify your assumptions? How can you correct your model if the assumptions are not verified?
3. (5 points) How can you judge the quality of a model? What metrics can you use to compare models?
4. (5 points) Given a model that predicts y given x_1 and x_2 write the a) first order model, b) interaction model and c) complete second order model. Which is better, under which circumstances?
5. (5 points) In the model below, what is Beta-0, Beta-1, Beta-2? What is the regression line? Why was this line chosen? What is the SSE? Can you be certain that x_1 and x_2 should be in the model? What is R^2? What does that mean? What is MSE? What does that mean? RMSE? What does that mean?

The REG Procedure
Model: MODEL1
Dependent Variable: PRICE

Number of Observations Read	32
Number of Observations Used	32

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	4283063	2141531	120.19	<.0001
Error	29	516727	17818		
Corrected Total	31	4799790			

Root MSE	133.48467	R-Square	0.8923
Dependent Mean	1326.87500	Adj R-Sq	0.8849
Coeff Var	10.06008		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1338.95134	173.80947	-7.70	<.0001
AGE	1	12.74057	0.90474	14.08	<.0001
NUMBIDS	1	85.95298	8.72852	9.85	<.0001

6. (5 points) How can you validate your model? Give two distinctly different methods?
7. (5 points) Explain as if to a nonprofessional why adjusted-R^2 might be better than R^2.
8. (5 points) Define "parsimonious." Explain its relevance to building regression models.
9. (5 points) Explain how to incorporate categorical features into your model? Be specific.
10. (5 points) Compare and contrast the benefits and drawbacks of forward stepwise regression, backward stepwise regression, and all-possible regression.