1. The director of admissions at a state university wished to determine how accurately students' grade-point averages at the end of their freshman year (Y); X1= the high school class rank as percentile where 99 indicates student is at or near the top of his or her class and 1 indicates student is at or near the bottom of the class; X2= ACT score; X3= Academic Year.
2. Create a development sample (70% of the data) and hold-out sample (%30) of the data, build your regression model on the development data, check the regression for the followings: normality, outliers and influential points, and multi collinearity document all your work and attach relevant r graphs.
3. Test the performance of the model on the hold out sample, and investigate the model stability on the hold out sample
4. Divide the cases into two groups, placing 247 cases with the smallest fitted values into group 1 and the remaining cases into group 2. Conduct the Brown-Forsythe test for constancy of the error variance, using α = .01. State the decision rule and conclusion.



1. We are interested in predicting the number of customers who complained about the service, use the attached data sets to answer the questions below:
2. Build a model to predict the number of complaints, perform the statistical tests that shows that model is significant
3. Find the predicted number complaints given the independent variables below

X1 X2 X3 X4 X5

606 41393 3 3.04 6.32



1. For the attached data sets, build a model to predict Y based on the independent variables and test if there is an autocorrelation persists in the data. If autocorrelation persists, remediate the autocorrelation.



1. For the following hospital data,

Y= Total cost

X1=Interventions

X2=Drugs

X3=Emergency room visits

X4=Complications

X5=Comorbidities

a-) use the best subset method to find optimal linear regression model

b) Compare your model in part a against the regression tree and Neural Network Model, and calculate the SSE for each model, which method has the lowest SSE?



5-)In a flu shot study, 159 clients were randomly selected and asked whether they actually received a flu shot. A client who received a flu shot was coded Y=1 and a client who did not receive a flu shot was coded Y =0. In addition. data were collected on their age (X1) and their health awareness. The latter data were combined into a health awareness index (X2), for which higher values indicate greater awareness. Also included in the data was client gender, where males were coded X3=1 and females were coded X3=0.

1. Fit a model to predict the probability of getting a flu shot and state the fitted response function.
2. Use the likelihood ratio test to determine whether X3 can be dropped from the regression model
3. What is the estimated probability that male clients aged 55 with a health awareness index of 60 will receive a flu shot? Obtain a 90% confidence interval for your prediction
4. Conduct Hosmer-Lemshow goodness of fit test for the appropriateness of the logistic regression function



Top of Form