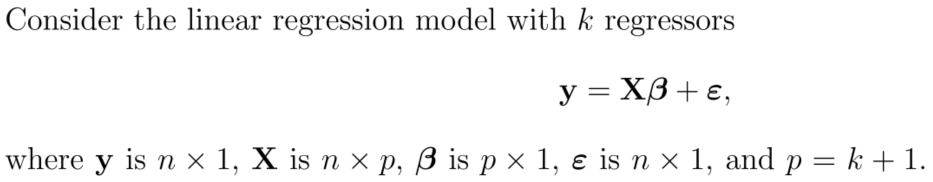
STAT2013 Regression Analysis

Assignment 4

1. From prediction point of view, please list criteria we consider.

2. Write down the formula of R^2 and R^2_{adj}. What is the disadvantage of R^2 ? Why do we propose R^2_{adj} compared with R^2 ?

3.

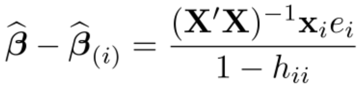


Consider the linear regression model with k regressors

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is $n \times 1$, $\mathbf{X}$ is $n \times p$, $\boldsymbol{\beta}$ is $p \times 1$, $\boldsymbol{\varepsilon}$ is $n \times 1$, and $p = k + 1$.

Suppose$\hat{\beta}$is the OLS estimator ofβobtained by using all *n* observations,$\widehat{\beta}_{\imath}$is the estimatorobtained with the *i*th observation deleted. Show that



$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)} = \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i e_i}{1 - h_{ii}}$$

Where h_{ii} is the *i*th diagonal element of the hat matrix= $X(X'X)^{-1}X'$, e_i is the ordinary residuals of *i*th observation when $\hat{y}$ is estimated by $\hat{\beta}$.

[Hint: If $X'X$ is a $k \times k$ matrix and x be it ith row vector then *(*$X'X - x'x$*)* denotes the $X'X$ matrix with the ith row withheld.

$$[X'X - x'x]^{-1} = （X'X）^{-1} + \frac{(X'X)^{-1}x'x(X'X)^{-1}}{1 - x(X'X)^{-1}x'}$$

4. Describe the following three variable selection procedures
   1) Forward selection 2) Backward elimination 3) stepwise

5. If you find few observations to be an outliers. What do you suggest to deal with outliers in building regression model?

6. Describe several ways to identify the influential point in regression analysis.

7. Multiple choice questions:
7.1. An acceptable residual plot exhibits:
A) Increasing error variance
B) Decreasing error variance
C) Constant error variance
D) A curved pattern
E) A mixture of increasing and decreasing error variance

7.2. Which one of the following is not an assumption about the residuals in a regression model?
A) Constant variance
B) Independence
C) Normality
D) Variance of zero
E) Mean of zero

7.3. All of the following are desirable outcomes for a multiple regression model except

A) High R^2

B) Small s
C) Small Cp statistic
D) Large Cook's distance measure
E) Large F statistic

7.4. In multiple regression analysis, a desirable residual plot has what type of appearance?
A) Curved
B) Cyclical
C) Fanning out
D) Funneling in
E) Horizontal band

7.5. Cook's distance measure is used in:
A) Determining if there is significant first order auto-correlation.
B) Identifying influential observations in multiple regression analysis.
C) Determining the significance of an independent variable.
D) Determining if there is significant multicollinearity.
E) Determining if the overall regression model is significant.

7.6. Which one of the following tools is not used to check the normality of residuals assumption for a multiple regression model?
A) Histogram
B) Stem-and-leaf display

C) Scatter diagram
D) Normal plot

7.7. In residual analysis, as the value of the leverage _____________, the value of the studentized residual_____________.
A) Decreases, increases
B) Increases, decreases
C) Decreases, decreases
D) None of the above

7.8. The logarithm transformation can be used
A) to overcome violations to the assumption that residuals are independent.
B) to change a linear independent variable into a nonlinear independent variable.
C) to overcome violations to the homoscedasticity assumption.
D) to test for possible violations to the homoscedasticity assumption.