STAT 51200--FALL 2022
Homework 8

1. Read Chapters 7-8 in the text.
2. Do Problems: 7.7, 7.8, 7.10, 7.27 8.3, 8.10, 8.12, 8.14, 8.16, 8.20 in the textbook.
3. Additionally, refer the Homes.dat on the Canvas course site, which consists of records of:

   Y = sale price (x $1,000)

   x_1 = square footage (x 100)

   x_2 = number of bedrooms

   x_3 = number of bathrooms

   x_4 = total number of rooms

   x_5 = age of the home

   x_6 = car garage (yes=1, no=1)

   x_7 = good view (yes=1, no=0),

   for some *n=50* single-family homes. The seven variables *x_1,…, x_7* are possible predictors variables for the home's selling-price *y*.

1) Make a regression of *y* on these seven variables, and identify the variables which are **less likely** to be important predictors. That is, make a global test for the utility of the regression model including all seven independent variables and then perform separate individual tests for each of the parameters.
2) Fit a regression model to the data which includes those independent variables **you think (**based on your answer to 1) above), are likely to be important predictors. Discuss the resulting model and compare it to the model you fitted in 1).
3) One method of selection is to fit the data to **all possible regression models** (with one, with two, with three, etc. independent predictors). For each such model, to calculate the values of *R^2*, *MSE* and that of *C_p*. Then to select as the *best* sub-model (with *p* independent variables), the one which has the large *R^2*, small *MSE*, and a smaller bias of *C_p*, (with *C_p* approximately equal

to *p+1*). To do just that, run the following SAS procedure:

```
PROC RSQUARE CP MSE;
MODEL Y=X1-X7;
```

a) The output is self-explanatory. Based on this output and the above criterion, identify the *best* subset model.

b) To allow SAS to select for you the *best* subset model, run:

```
PROC RSQUARE OUTEST=BEST CP MSE SELECT=1;
MODEL Y=X1-X7;
```

To get three plots similar to those on pages 362 in the text, use:

```
PROC PLOT;
PLOT _CP_ * _P_='*'  _P_*_P_='+'/overlay;
PLOT _MSE_ * _P_='*';
PLOT _RSQ_*_P_='*';
```

Use these plots to determine the best sub-model.

4) Another method of selection is to use stepwise regression procedure.

```
PROC STEPWISE;
```

There are three techniques to do stepwise selection

- MODEL Y=X1-X7/FORWARD SLENTRY=0.10;

which starts with one independent variable (the most highly correlated with y) and then adds more independent variables according to the SLENTRY=0.10 criterion. Each time, the individual (F) tests are taking into account the already included variables.

- MODEL Y= X1-X7/BACKWARD SLSTAY=0.10;

which starts with the model including all (seven) variables and then eliminates variables (one by one) from the regression according to the SLSTAY=0.10 criterion.

- MODEL Y=X1-X7/STEPWISE SLENRTY=0.10 SLSTAY=0.10;

which is a sequential combination of *Forward* and *Backward* procedures above.