DSA 8020 R Lab 4: Model Selection and Model Checking

your name here

Contents

Savings rates in 50 countries

The savings data frame has 50 rows (countries) and 5 columns (variables):

1. `sr`: savings rate - personal saving divided by disposable income *This variable will be used as the response*
2. `pop15`: percent population under age of 15
3. `pop75`: percent population over age of 75
4. `dpi`: per-capita disposable income in dollars
5. `ddpi`: percent growth rate of dpi

The data is averaged over the period 1960-1970.

Data Source: Belsley, D., Kuh. E. and Welsch, R. (1980) *Regression Diagnostics* Wiley.

Load the dataset

Code:

```
data(savings, package = "faraway")
```

1. Perform the best subset selection and select the "best" model using R^2_{adj}

Code:

Answer:

2. Perform a stepwise selection using AIC

Code:

Answer:

3. Perform a general linear F-test (with $\alpha = 0.1$) to choose between the full model (i.e., using the all 4 predictors) and the reduce model that include `pop15`, `pop75`, and `ddpi` as the predictors

Code:

Answer:

4. Make a residual plot of the model selected by *AIC*. Then, comment on the model assumptions

Code:

Answer:

5. Use both histogram and qqplot to examine the normality assumption on error

Code:

Answer:

6. Calculate the leverage values to check if there is any high leverage points (i.e., $h > \frac{2p}{n}$)

Code:

Answer:

7. Compute jackknife residuals to identify outlier(s)

Code:

Answer:

8. Identifying influential observations by computing DFFITS

Code:

Answer: