# lab-06

Tim Lanthier

2/24/2022

## Lab 06: Model Selection + Diagnostics

```
library(tidyverse)
library(knitr)
library(broom)
library(leaps)
library(rms)
library(Sleuth3) #case1201 data
```

In this lab we will be working with SAT data from the 1982 exam. The dataset can be found in the Sleuth3 package (case 1201).

### Model Selection

We will start with a full linear model including all possible predictor variables and no interaction terms.

```
sat_scores <- Sleuth3::case1201
full_model <- lm(SAT ~ Takers + Income + Years + Public + Expend + Rank , data = sat_scores)
tidy(full_model) %>% kable(digits = 3)
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -94.659 | 211.510 | -0.448 | 0.657 |
| Takers | -0.480 | 0.694 | -0.692 | 0.493 |
| Income | -0.008 | 0.152 | -0.054 | 0.957 |
| Years | 22.610 | 6.315 | 3.581 | 0.001 |
| Public | -0.464 | 0.579 | -0.802 | 0.427 |
| Expend | 2.212 | 0.846 | 2.615 | 0.012 |
| Rank | 8.476 | 2.108 | 4.021 | 0.000 |

Now we will conduct backwards selection on our full model. We will start by using adjusted $R^2$ as our selection criterion.

```
model_select <- regsubsets(SAT ~ Takers + Income + Years + Public + Expend +
                             Rank , data = sat_scores, method = "backward")
select_summary <- summary(model_select)
coef(model_select,1:6)
```

```
## [[1]]
## (Intercept)        Rank
##  183.418763    9.557949
##
## [[2]]
```

```
## (Intercept)       Years        Rank
## -243.930900   27.382901    9.351603
##
## [[3]]
## (Intercept)       Years      Expend        Rank
## -303.724295   26.095227    1.860866    9.825794
##
## [[4]]
## (Intercept)       Years      Public      Expend        Rank
## -204.598232   21.890482   -0.663798    2.241640   10.003169
##
## [[5]]
##  (Intercept)       Takers        Years       Public       Expend         Rank
## -100.4736967   -0.4620796   22.6688085   -0.4522606    2.1859091    8.4964099
##
## [[6]]
##   (Intercept)        Takers        Income         Years        Public
## -94.659108883  -0.480080120  -0.008195013  22.610081908  -0.464152292
##        Expend          Rank
##   2.212004850   8.476216985
```

So using backward selection with adjusted $R^2$ as our criterion, our best model includes `Years`, `Public`, and `Expend` as our predictor variables. Now we will run backward selection with BIC as the selection criterion.

```
model_select <- regsubsets(SAT ~ Takers + Income + Years + Public + Expend +
                             Rank , data = sat_scores, method = "backward")
select_summary <- summary(model_select)
coef(model_select,1:6)
```

```
## [[1]]
## (Intercept)        Rank
##  183.418763    9.557949
##
## [[2]]
## (Intercept)       Years        Rank
## -243.930900   27.382901    9.351603
##
## [[3]]
## (Intercept)       Years      Expend        Rank
## -303.724295   26.095227    1.860866    9.825794
##
## [[4]]
## (Intercept)       Years      Public      Expend        Rank
## -204.598232   21.890482   -0.663798    2.241640   10.003169
##
## [[5]]
##  (Intercept)       Takers        Years       Public       Expend         Rank
## -100.4736967   -0.4620796   22.6688085   -0.4522606    2.1859091    8.4964099
##
## [[6]]
##   (Intercept)        Takers        Income         Years        Public
## -94.659108883  -0.480080120  -0.008195013  22.610081908  -0.464152292
##        Expend          Rank
##   2.212004850   8.476216985
```