STAC67: Regression Analysis

Lecture 9

Sohee Kang

Feb. 10, 2021

$\hat{\beta} = (X'X)^{-1}X'Y$

$\hat{Y} = X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_{H}Y$

$\hat{e} = Y - HY = (I-H)Y$

Chapter 6: Mutiple Regression I

Models with Multiple Predictors

- Most Practical Problems have **more than one** potential predictor variable
- Goal is to determine effects (if any) of each predictor, controlling for others
- Can include polynomial terms to allow for nonlinear relations
- Can include product terms to allow for interactions when effect of one variable depends on level of another variable
- Can include "dummy" variables for categorical predictors
- First-Order Model with 2 Numeric Predictors

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \Rightarrow E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Example - Multiple Regression with Two predictor Variable

- This dataset, which can be found at UCI Machine Learning Repository contains a response variable **mpg** which stores the city fuel efficiency of cars, as well as several predictor variables for the attributes of the vehicles. we would like to model the fuel efficiency (mpg) of a car as a function of its weight (**wt**) and model year (**year**).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad i = 1, 2, \ldots, n, \quad where\ \epsilon_i \sim N(0, \sigma^2)$$

Bias Weight Year

$$(X'X)^{-1}X'Y$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix}$$

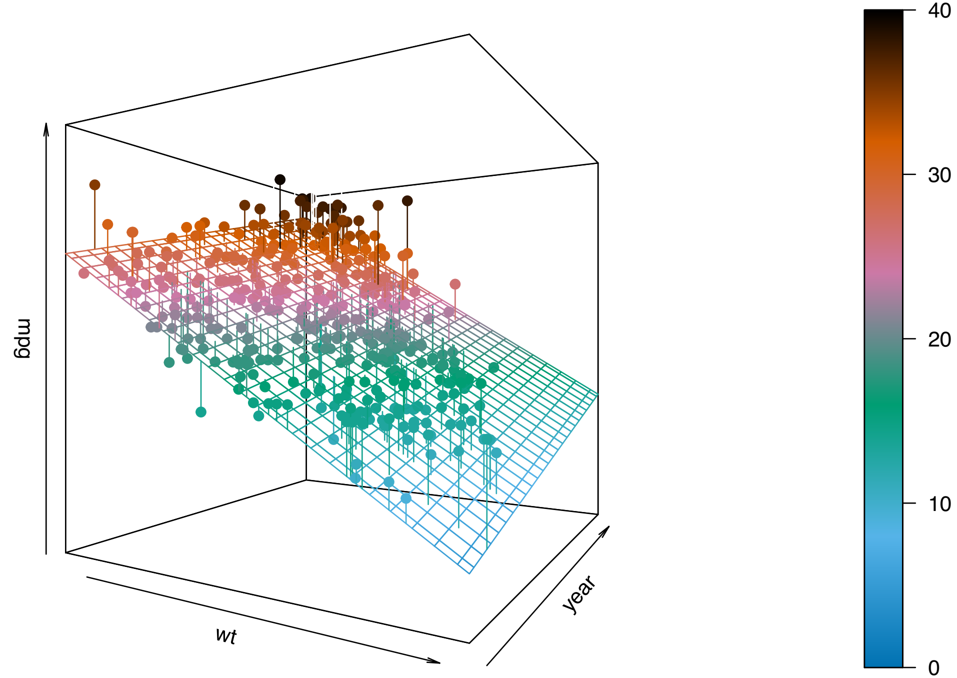
```
##
## Call:
## lm(formula = mpg ~ wt + year, data = autompg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.852 -2.292 -0.100  2.039 14.325
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.464e+01  4.023e+00  -3.638 0.000312 ***
## wt          -6.635e-03  2.149e-04 -30.881  < 2e-16 ***
## year         7.614e-01  4.973e-02  15.312  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.431 on 387 degrees of freedom
```

3-dimensional image

- what happens if we fit two simple regressions?

```
##  (Intercept)           wt
## 46.198395049 -0.007643021
```

```
## (Intercept)        year
##  -70.155395    1.231997
```



Interpretation of regression parameters

1. Change of $E(Y)$ per unit corresponding to a unit change in X_j, $\quad j = 1, \ldots, p$ when other predictor variables are held constant.
2. β_2: effect of X_2 after **adjusting** for X_1
3. β_1: effect of X_1 after **adjusting** for the effect of X_2

partial effect.

- If X_1 and X_2 are **uncorrelated** then the estimates will be the same as the estimates in the simple models but in general, this is not true.
- Consider the following sequence of model:

1. Fit a simple linear model between Y and X_1, and find the residual, e_1.
2. Fit a regression with X_2 as a response and X_1 as a covariate and find the residual, e_2.
3. Fit a regression with e_1 as a response and e_2 as a covariate.

```
# Fit a simple regression between Y and X1 and find e1
```

The Multiple Regression Model in matrix notation

$$\underset{\sim}{Y} = \boldsymbol{X}\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$$

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}}_{} = \underbrace{\begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix}}_{n \times p+1} \underbrace{\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}}_{p+1 \times 1} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{n \times 1}$$

- Each column of $\boldsymbol{X}$ contains the values of a particular predictor variable.

  - X is the design matrix of constants with full column rank
  - ε is the vector of random errors (i.i.d $N(0,\sigma^2)$)
  - β is the vector of population params

Normality assumption and Random vector

- Common assumption ϵ_i iid $N(0, \sigma^2)$
- Under this assumption, the joint probability density function of $\underset{\sim}{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ is

$$f(\underset{\sim}{\epsilon}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma^2}\right\} = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\sum_{i=1}^{n} \frac{\epsilon_i^2}{2\sigma^2}\right)$$

Random vector $\underset{\sim}{Y}$

- $\underset{\sim}{Y}$ is a sum of two components:
  - a constant component: $\underset{\sim}{X}\vec{\beta}$
  - a random component : $\underset{\sim}{\epsilon}$
- Y_i are pairwise independent because: ϵ_i iid
- $Var(Y_i) = \sigma^2$
- The variance-covariance matrix of $\underset{\sim}{Y}$ is: $\sigma^2 \underline{I}_{(n \times n)}$

Estimation in Matrix Notation

- **Least Squres**: best in the sense that the sum of the squares of the errors is minimized, i.e. the minimizing criterion is:

$$
\begin{aligned}
Q(\beta_0,\beta_1,\dots,\beta_p) &= \sum_{i=1}^{n} \epsilon_i^2 \\
&= \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip})]^2 \\
&= \sum_{i=1}^{n} \left(Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 .
\end{aligned}
$$

Least square estimator of $\underset{\sim}{\beta}$: $\underset{\sim}{\widehat{\beta}}$ that minimizes criterion Q.

Least squares in matrix form

Exercise 1

Prove that the least squares criterion can be rewritten:

$$Q(\beta_0, \beta_1, \ldots, \beta_p) = (\underset{\sim}{Y} - \boldsymbol{X}\underset{\sim}{\beta})'(\underset{\sim}{Y} - \boldsymbol{X}\underset{\sim}{\beta})$$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - 2\underset{\sim}{\beta}'\boldsymbol{X}'\underset{\sim}{Y} + \underset{\sim}{\beta}'\boldsymbol{X}'\boldsymbol{X}\underset{\sim}{\beta}$$

$$= Y'Y - Y'XB - B'X'Y + B'X'XB$$

$$= Y'Y - 2B'X'Y + B'X'XB$$

$$\frac{dQ}{dB} = -2X'Y + 2X'XB = 0$$

$$X'XB = X'Y$$

$$\hat{B} = (X'X)^{-1}X'Y$$

Least squares estimator

Exercise 2

Show that the least squares estimator of $\underset{\sim}{\beta}$ is

$$\widehat{\underset{\sim}{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\underset{\sim}{Y}$$

Plan of the proof:

1. Write the normal equations (derivatives of Q set to 0).
   Hint: use (without proof) that $\frac{\partial \underset{\sim}{\beta}'\boldsymbol{X}'\underset{\sim}{Y}}{\partial \underset{\sim}{\beta}} = \boldsymbol{X}'\underset{\sim}{Y}$ and $\frac{\partial \underset{\sim}{\beta}'\boldsymbol{X}'\boldsymbol{X}\underset{\sim}{\beta}}{\partial \underset{\sim}{\beta}} = 2\boldsymbol{X}'\boldsymbol{X}\underset{\sim}{\beta}$.
2. Find the critical points (solution to the normal equations).
3. Show that the critical point is a minimum (we will skip this step).

Comment 1 Matrix $\boldsymbol{X}'\boldsymbol{X}$ is invertible because $\boldsymbol{X}$ is of full column rank.

General form of $X'X$

$$
\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & X_{31} & \cdots & X_{n1} \\ X_{12} & X_{22} & X_{32} & \cdots & X_{n2} \\ \vdots & \vdots & \vdots & & \vdots \\ X_{1p} & X_{2p} & X_{3p} & \cdots & X_{np} \end{pmatrix} \begin{pmatrix} 1 & X_{11} & \cdots & X_{1j} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2j} & \cdots & X_{2p} \\ 1 & X_{31} & \cdots & X_{3j} & \cdots & X_{3p} \\ \vdots & & \vdots & & \vdots & \\ 1 & X_{n1} & \cdots & X_{nj} & \cdots & X_{np} \end{pmatrix}
$$

$$
= \begin{pmatrix} n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{ip} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} & \cdots & \sum X_{i1}X_{ip} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{i2}X_{ip} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{ip} & \sum X_{i1}X_{ip} & \sum X_{i2}X_{ip} & \cdots & \sum X_{ip}^2 \end{pmatrix}
$$

General form of $\boldsymbol{X}'\underset{\sim}{Y}$

$$\boldsymbol{X}'\underset{\sim}{Y} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & X_{31} & \cdots & X_{n1} \\ X_{12} & X_{22} & X_{32} & \cdots & X_{n2} \\ \vdots & \vdots & \vdots & & \vdots \\ X_{1p} & X_{2p} & X_{3p} & \cdots & X_{np} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum Y_i \\ \sum X_{i1} Y_i \\ \sum X_{i2} Y_i \\ \vdots \\ \sum X_{ip} Y_i \end{pmatrix}$$

Example: mpg data

```
library(dplyr)
mpg.data = select(autompg, X, mpg, wt, year)
head(mpg.data)
```

```
##                                        X mpg   wt year
## 1 8 cylinder 70 chevrolet chevelle malibu  18 3504   70
## 2         8 cylinder 70 buick skylark 320  15 3693   70
## 3        8 cylinder 70 plymouth satellite  18 3436   70
## 4             8 cylinder 70 amc rebel sst  16 3433   70
## 5               8 cylinder 70 ford torino  17 3449   70
## 6          8 cylinder 70 ford galaxie 500  15 4341   70
```

$$\sum_{i=1}^{n} X_{i1} = 1162338 \qquad \sum_{i=1}^{n} X_{i2} = 29622 \qquad \sum_{i=1}^{n} X_{i1}X_{i2} = 87911306$$

$$\sum_{i=1}^{n} X_{i1}^2 = 3745687164 \qquad \sum_{i=1}^{n} X_{i2}^2 = 2255160$$

$$\sum_{i=1}^{n} Y_i = 9133.6 \qquad \sum_{i=1}^{n} X_{i1}Y_i = 25069783.4 \qquad \sum_{i=1}^{n} X_{i2}Y_i = 700206.4$$

Exercise

Provide $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{X}'\underset{\sim}{Y}$.

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} =$$

```
n = dim(mpg.data)[1]
X = cbind(rep(1, n), mpg.data$wt, mpg.data$year)
solve(t(X)%*%X)
```

```
##               [,1]          [,2]          [,3]
## [1,]  1.374834e+00 -3.280479e-05 -1.677993e-02
## [2,] -3.280479e-05  3.920470e-09  2.780689e-07
## [3,] -1.677993e-02  2.780689e-07  2.100115e-04
```

Exercise

Give $\underset{\sim}{\widehat{\beta}}$, the estimated regression surface, and interpret the parameters.

$$\underset{\sim}{\widehat{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\underset{\sim}{Y})$$

```
beta.hat = solve(t(X)%*%X)%*%(t(X)%*%mpg.data$mpg)
beta.hat
```

```
##                [,1]
## [1,] -14.637641945
## [2,]  -0.006634876
## [3,]   0.761401955
```

```
coefficients(fit)
```

```
##   (Intercept)            wt          year
## -14.637641945  -0.006634876   0.761401955
```

The estimated regression surface is:

Fitted Values

- The estimated values of the mean of Y for the values of the predictor variables in the sample are

$$\underset{\sim}{\widehat{Y}} = \boldsymbol{X}\underset{\sim}{\widehat{\beta}}.$$

- This vector is called the vector of fitted values.
- It can be rewritten as a linear function of Y as

$$\underset{\sim}{\widehat{Y}} =$$

where,

Exercise: show that $\boldsymbol{H}$ s a projection matrix. That is, show that $\boldsymbol{H}$ is a symmetric ($\boldsymbol{H}' = \boldsymbol{H}$) and idempotent ($\boldsymbol{H}\boldsymbol{H} = \boldsymbol{H}$) matrix.

Residuals

- A residual is the deviation of the observed value of Y to the corresponding fitted value.
- The vector of residuals is

$$\underset{\sim}{e} = \underset{\sim}{Y} - \widehat{\underset{\sim}{Y}}.$$

- It can be expressed as a linear function of Y as

$$\underset{\sim}{e} =$$

- Reminder: $\widehat{\underset{\sim}{\beta}}$ was chosen so that $\underset{\sim}{e}'\underset{\sim}{e}$ is minimum.

Exercise Show that $\boldsymbol{I} - \boldsymbol{H}$ is a projection matrix.