STAC67: Regression Analysis

Lecture 10

Sohee Kang

Feb. 11, 2021

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 $\boldsymbol{X}'\boldsymbol{X}$

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

$$= \begin{pmatrix} n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{ip} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} & \cdots & \sum X_{i1}X_{ip} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{i2}X_{ip} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{ip} & \sum X_{i1}X_{ip} & \sum X_{i2}X_{ip} & \cdots & \sum X_{ip}^2 \end{pmatrix} \in \mathbb{R}^{p+1 \times p+1}$$

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

```
##                                         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 $\boldsymbol{X}'\underset{\sim}{Y}$.

① ②

$$① = \begin{bmatrix} 390 & 1162338 & 29622 \\ 1162338 & 3745\ldots & 8791\ldots \\ 29622 & 8791\ldots & 2551\ldots \end{bmatrix}$$

$$② = \begin{bmatrix} 9133\ldots \\ 2506\ldots \\ 7062\ldots \end{bmatrix}$$

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

```
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)
```

while weight is constant, every 1 increase of year will result in 0.761401955 increase of MPG. on average

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

The estimated regression surface is: $\hat{Y} = X\hat{B}$

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

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

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.

$H' = (X(X'X)^{-1}X')' = X(X'X)^{-1}X' = H$ $\qquad$ $HH = X\cancel{(X'X)^{-1}X'X}(X'X)^{-1}X' = H$

Residuals

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

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

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

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

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

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

$(I-H)' = I' - H' = I - H$

$(I-H)(I-H) = I - H \cancel{-H-H+HH}$
$= I - H$

mpg Example

```
Yhat = X %*% beta.hat
class(Yhat)
```

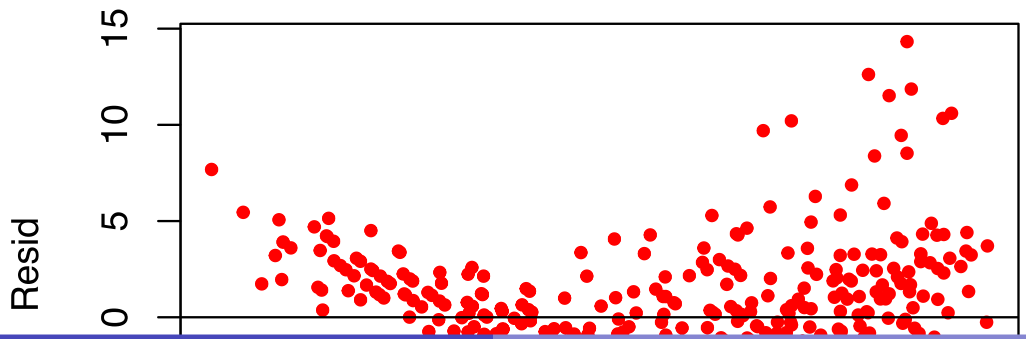
```
## [1] "matrix"
```

```
Yhat = as.vector(Yhat)

Y = mpg.data$mpg

Resid = Y - Yhat

plot(Yhat, Resid, col="red", pch=20)
abline(c(0,0))
```



Linear function of $\underset{\sim}{Z}$

- A vector $\underset{\sim}{U}$ of size $k \times 1$ is a linear function of $\underset{\sim}{Z}$ if it can be written

$$\underset{\sim}{U} = \mathbf{A}\underset{\sim}{Z}$$

for a matrix of constants $\mathbf{A}$ (i.e. whose elements are not random).

Exercise

Show that $\underset{\sim}{\hat{\beta}}$, $\underset{\sim}{\hat{Y}}$, and $\underset{\sim}{e}$ are linear functions of $\underset{\sim}{Y}$.

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

$$\hat{Y} = HY \qquad A = H$$

$$e = (I-H)Y \qquad A = I-H$$

Properties of a linear function of a random vector

- Consider a linear function

$$\underset{\sim}{U} = \mathbf{A}\underset{\sim}{Z}$$

of a random vector $\underset{\sim}{Z}$. We have

$$E(\underset{\sim}{U}) = \mathbf{A}E(\underset{\sim}{Z}), \qquad Var(\underset{\sim}{U}) = \mathbf{A}Var(\underset{\sim}{Z})\mathbf{A}'$$

Exercise

Consider $Z_1, Z_2, \ldots, Z_n$ are iid with an μ and σ^2.

1. Show that the mean $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$ is a linear function of $\underset{\sim}{Z}$ and give $\mathbf{A}$ such that $\bar{Z} = \mathbf{A}\underset{\sim}{Z}$.

$$A = \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}$$

2. Apply the properties shown above to derive the mean and variance of $\bar{Z}$.

$$E(\bar{Z}) = A\,E(\underset{\sim}{Z}) = \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix} \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} = \mu$$

$$Var(\bar{Z}) = A\,Var(\underset{\sim}{Z})A' = \frac{\sigma^2}{n}$$

Properties of $\underset{\sim}{\widehat{\beta}}$

(

Show the following properties of $\underset{\sim}{\widehat{\beta}}$.

- The mean of varinace-covariance matrix of $\underset{\sim}{\hat{\beta}}$ are

$$E(\underset{\sim}{\hat{\beta}}) = \underset{\sim}{\beta} \qquad Var(\underset{\sim}{\hat{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

Moreover, when $\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \sigma^2\mathbf{I})$, we have: $\hat{B} \sim N\left(B, (X'X)^{-1}\sigma^2\right)$

Proof of Properties of $\underset{\sim}{\widehat{\beta}}$

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

$Var(\hat{B}) = (X'X)^{-1}X'\sigma^2 I\, X(X'X)^{-1} = \cancel{(X'X)^{-1}X'X}(X'X)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2$

mpg Example

Exercise

What is the variance of $\hat{\beta}_2$ in mpg example? What is the covariance between $\hat{\beta}_0$ and $\hat{\beta}_2$? Give your answer in function of σ^2 (unknown)

Properties of $\underset{\sim}{\widehat{Y}}$

Show the following properties of $\underset{\sim}{\widehat{Y}}$.

- The mean of varinace-covariance matrix of $\underset{\sim}{\widehat{Y}}$ are

$$E(\hat{Y}) = E(X\hat{B}) = XB = E(Y)$$
$$V(\hat{Y}) = V(X\hat{B}) = X(X'X)^{-1}\sigma^2 X' = \sigma^2 H$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Moreover, when $\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \sigma^2\mathbf{I})$, we have

$$\hat{Y} \sim N(XB, \sigma^2 H)$$

Properties of $\widehat{\underset{\sim}{Y}}$

- $\widehat{\underset{\sim}{Y}}$ is an unbiased estimator of $E(\underset{\sim}{Y})$ (if the model is correct!)
- The variances of any subset of the $\widehat{Y}_i$'s can be determined using

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

where subscript r indicates that we consider only the rows that correspond to the $\widehat{Y}_i$'s of interest. We get

$$Var(\widehat{\underset{\sim}{Y}}_r) = \boldsymbol{X}_r(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_r'\sigma^2$$

Proof of the properties of $\underset{\sim}{\widehat{Y}}$

$$Var(\hat{Y}_r) = X_r \, Var(\hat{B}) X_r'$$

$$= X_r (X'X)^{-1} X_r' \, \sigma^2$$

mpg Example

Exercise

What is the variance of the first two fitted values $\hat{Y}_1$ and $\hat{Y}_2$ in mpg example? What is their covariance? Give your answers in function of σ^2 (unknown)

$$Var\left(\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \end{bmatrix}\right) = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \end{bmatrix} (X'X)^{-1} \begin{bmatrix} 1 & 1 \\ x_{11} & x_{21} \\ x_{12} & x_{22} \end{bmatrix} \sigma^2$$

$$X_r = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 3504 & 70 \\ 1 & 3693 & 70 \end{bmatrix}$$

Properties of $\underset{\sim}{e}$

$Var(e) = (I-H) H \sigma^2 (I-H)'$

$=$

Show the following properties of $\underset{\sim}{e}$.

The mean and variance-covariance matrix of $\underset{\sim}{e}$ are:

$E(e) = 0 \qquad Var(e) = (I-H)\sigma^2$

when $\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \sigma^2 \boldsymbol{I})$ we have

$e \sim N(0, (I-H)\sigma^2)$

Comment The diagonal elements H_{ii} of $\boldsymbol{H}$ satisfy $0 \leq H_{ii} \leq 1$ because

$Var(\hat{Y}_i) = H_{ii}\sigma^2$

$Var(e_i) = (1-H_{ii})\sigma^2$

Since $Var \geq 0$, $H_{ii} \in [0,1]$

Proof of Properties of $\underset{\sim}{e}$

$$E(e) = E(Y) - E(\hat{Y}) = XB - XB = 0$$

$$Var(e) = Var((I-H)Y) = (I-H) I \sigma^2 (I-H)^2$$
$$= (I-H)\sigma^2$$

Comments on the Variance

We have

$$Var(\underset{\sim}{Y}) = Var(\underset{\sim}{\widehat{Y}}) + Var(\underset{\sim}{e})$$

$$\sigma^2 I = \sigma^2 H + (I - H)\sigma^2$$

$$= \sigma^2 I$$

So data points having lower variance of $\hat{y}_i$ have higher variance on e_i and vice versa.

Exercise

$$H = X'(X'X)^{-1}X$$
$$= \left[\frac{1}{n}\right]_{ij}$$

Consider the model with only one constant:

$$y_i = \beta_0 + \epsilon_i$$

1. What is the design matrix $\boldsymbol{X}$? $\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$
2. What is the value of H_{ii} for any i under this model? $\frac{1}{n}$
3. What is the variance of $\widehat{Y}_i$? $\frac{1}{n} \cdot \sigma^2$
4. This is a well known formula for the variance of an estimator. Which one? Variance of sample mean.
5. Is this result surprising? Explain.

$$\hat{\beta}_0 = \frac{1}{n} \sum Y_i$$

Properties of a prediction $\widehat{Y}_{pred_0}$

- The prediction of a new observation

$$Y_0 = \underset{\sim}{x}'_0 \underset{\sim}{\beta} + \epsilon_0$$

for a vector of values of independent variables $\underset{\sim}{x}_0$ is

$$\widehat{Y}_{pred_0} = \underset{\sim}{x}'_0 \underset{\sim}{\widehat{\beta}}$$

- The mean and variance of $\widehat{Y}_{pred_0}$ are

Moreover, when $\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \sigma^2 \boldsymbol{I})$ then

Properties of a prediction $\widehat{Y}_{pred_0}$

- The variance of a prediction for a new observation is larger than the variance of the estimator of the mean response even though the point estimate is the same. That is, for a vector of values of predictor variables $\underset{\sim}{x}_0$
  - Prediction:

$$\widehat{Y}_{pred_0} = \underset{\sim}{x}_0' \underset{\sim}{\widehat{\beta}} \quad \text{with variance} \quad [1 + \underset{\sim}{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\underset{\sim}{x}_0]\sigma^2$$

  - Mean response:

$$\widehat{Y}_0 = \underset{\sim}{x}_0' \underset{\sim}{\widehat{\beta}} \quad \text{with variance} \quad \underset{\sim}{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\underset{\sim}{x}_0\sigma^2$$

Estimates and precision: summary

Consider the model $\underset{\sim}{Y} = \boldsymbol{X}\underset{\sim}{\beta} + \epsilon$, with $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$, $Var(\underset{\sim}{\epsilon}) = \boldsymbol{I}\sigma^2$

Quantity	Estimator	Variance of the estimator
$\underset{\sim}{\beta}$		
$E(\underset{\sim}{Y})$		
$\underset{\sim}{\epsilon}$		
Y_0		

Analysis of Variance and Quadratic Forms

- Quadratic forms of $\underset{\sim}{Y}$: $\underset{\sim}{Y}'\boldsymbol{A}\underset{\sim}{Y}$, where $\boldsymbol{A}$ is a symmetric matrix of coefficients called defining matrix
- Next section: Study the properties of residual, regression and total sum of squares and sum of squares used in inference
- They are all quadratic forms of $\underset{\sim}{Y}$

Partitioning of total sum of squares

- We know that

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

- We will generalize the partitioning of the total sum of squares that we had for simple linear regression, i.e.

$$SST = SSR + SSE$$

to multiple linear regression.

Total sum of squares

- $SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y} \right)^2$ in matrix notation:

Exercise

Show that $SST = \underset{\sim}{Y}' \underset{\sim}{Y} - \frac{1}{n} \underset{\sim}{Y}' \boldsymbol{J} \underset{\sim}{Y}$, where $\boldsymbol{J}$ is the $n \times n$ square matrix with all elements equal to 1.

$SSR = SST - SSE$

$= Y'Y - Y'JY - (Y'-Y))$

- SST is a quadratic form of $\underset{\sim}{Y}$ because

- The defining matrix associated is

Residual sum of squares

- $SSE = \sum_{i=1}^{n} e_i^2$ in matrix notation:

Exercise

Show that $SSE = \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{\widehat{\beta}}'\boldsymbol{X}'\underset{\sim}{Y}$

- SSE is a quadratic forms of $\underset{\sim}{Y}$ because

$$SSE = \underset{\sim}{Y}'(\boldsymbol{I} - \boldsymbol{H})\underset{\sim}{Y}$$

- The defining matrix is $\boldsymbol{I} - \boldsymbol{H}$.

Regression sum of squares

- $SSR = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2$ in matrix notation:

Exercise

Show that $SSR = \widehat{\underset{\sim}{\beta}}'\boldsymbol{X}'\underset{\sim}{Y} - \frac{1}{n}\underset{\sim}{Y}'\boldsymbol{J}\underset{\sim}{Y}$

Regression sum of squares

- SSR is a quadratic forms of $\underset{\sim}{Y}$ because

- The defining matrix associated is:

Exercise

Check that

$$SST = SSR + SSE$$

Degrees of freedom

- For now: the number of values in the calculation of a statistic that can freely vary.
- SST has degrees of freedom
- SSE has degrees of freedom
- SSR has degrees of freedom.

Mean squares

- Mean squares: sum of squares divided by its associated degrees of freedom
- Regression mean squares:

$$MSR = \frac{SSR}{}$$

- Residual mean squares:

$$MSE = \frac{SSE}{}$$

Analysis of variance table

- Analysis of variance (ANOVA) table to display the sum of squares and degrees of freedom

Source of variation	Sum of squares	df	Mean squares
Regression	SSR		$MSR = \frac{SSR}{}$
Residual	SSE		$MSE = \frac{SSE}{}$
Total	SST		

- The results in the ANOVA table will be used to construct a global test for the regression coefficients.

Properties of a quadratic form of a random vector

- Consider a quadratic form

$$U = \underset{\sim}{Z}'\boldsymbol{A}\underset{\sim}{Z}$$

of a random vector $\underset{\sim}{Z}$ where $\boldsymbol{A}$ is a symmetric matrix (the defining matrix). We have

$$E(\underset{\sim}{Z}'\boldsymbol{A}\underset{\sim}{Z}) =$$

$$Var(\underset{\sim}{Z}'\boldsymbol{A}\underset{\sim}{Z}) =$$

Unbiased estimator of σ^2

Exercise: Show that $E(\underset{\sim}{e}'\underset{\sim}{e}) = (n - p')\sigma^2$.

Hint: use that $tr(\boldsymbol{P}) = p'$ (without proof) and the quadratic formulation of $\underset{\sim}{e}'\underset{\sim}{e}$.

Exercise:Show that the estimator

$$s^2 = \frac{\underset{\sim}{e}'\underset{\sim}{e}}{n - p'}$$

is an unbiased estimator of σ^2.