STA302 weeks 7–8

Mark Ebden 2018. Section 5.2

With grateful acknowledgment to Alison Gibbs

Using Matrix SLR

Recall our question from last week: How do we solve the least-squares estimates of the regression coefficients, in matrix form?

In other words: in *matrix form* we seek b_0 and b_1 that minimize the sum of squares of residuals,

$$\text{RSS} = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Least-squares estimation of regression coefficients

Let's start by speaking of RSS in terms of $\hat{\beta}$, and getting rid of the summation:

$$\text{RSS}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

Why this works: Easiest is to begin with the matrix RHS, multiply it out, and arrive at the $\sum$ LHS. Building on last week:

$n \times 2$

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$

Textbook tip: When Simon Sheather writes RSS(β), consider RSS($\hat{\beta}$).

Least-squares estimation of regression coefficients

Let's continue:

$$\begin{aligned}
\mathrm{RSS}(\hat{\beta}) &= (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\
&= (\mathbf{Y}' - \hat{\beta}'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\hat{\beta}) \quad \text{from pp 22 \& 24 of RMA} \\
&= \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\
&= \mathbf{Y}'\mathbf{Y} - 2\hat{\beta}'\mathbf{X}'\mathbf{Y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}
\end{aligned}$$

scalar

$$\frac{\partial \mathrm{RSS}(\hat{\beta})}{\partial \hat{\beta}} = 0 - \underbrace{2\mathbf{X}'\mathbf{Y}}_{\text{by Lemma 1}} + \underbrace{2\mathbf{X}'\mathbf{X}\hat{\beta}}_{\text{By Lemma 2}}$$

Setting the derivative to zero as before in Ch 2,

$$2\mathbf{X}'\mathbf{X}\hat{\beta} = 2\mathbf{X}'\mathbf{Y}$$

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

A reflection on the inverse

$$\left(\sum x_i y_i\right)\frac{1}{n}$$

$$\overline{xy}\cdot n$$

Our $\hat{\beta}$ expression is a concise way to write the estimators for linear regression, compared to $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

Should we always set $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$?

See Answer Slides 1-2.

A closer look at $\mathbf{X}'\mathbf{X}$

What does $\mathbf{X}'\mathbf{X}$ simplify to? Is it symmetric?

yes

Recall: $\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$

$$\begin{pmatrix} n & n\bar{x} \\ n\bar{x} & n\overline{x^2} \end{pmatrix}$$

See Answer Slide 3.

A closer look at $(\mathbf{X}'\mathbf{X})^{-1}$

$(nA)^{-1} = \frac{1}{n} A^{-1}$

What does $(\mathbf{X}'\mathbf{X})^{-1}$ simplify to? Recall:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$\left[n \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix} \right]^{-1} = \frac{1}{n} \begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}^{-1}$$

$$= \frac{1}{n} \cdot \frac{1}{\overline{x^2} - \bar{x}^2} \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

$$= \frac{1}{S_{xx}} \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$ by the aid sheet

See Answer Slide 4.

Bringing this together

$\overline{xy} \neq \bar{x}\,\bar{y}$ 6:11

Given our expression for $(\mathbf{X}'\mathbf{X})^{-1}$, what is $\hat{\beta}$ and hence our $\hat{\beta}_0$ and $\hat{\beta}_1$?

Recall that

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

$$\hat{\beta} = \frac{1}{S_{xx}} \begin{pmatrix} \overline{x^2}\,\bar{y}\,n - \bar{x}\,\overline{xy}\,n \\ -\bar{x}\,\bar{y}\,n + \overline{xy}\,n \end{pmatrix}$$

2×1

$$= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$

$\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}$

slide 5

Recall our R code for β_0 and β_1

e.g. handling question 1 of Chapter 2:

```
X <- read.csv("playbill.csv")
y <- X$CurrentWeek; x <- X$LastWeek
my <- mean(y); mx <- mean(x); n <- length(x)
Sxy <- sum((x-mx)*(y-my)); Sxx <- sum((x-mx)^2)
b1 <- Sxy/Sxx # (2.4), beta-hat-1
b0 <- my - b1*mx # (2.3), beta-hat-0
yHat <- b1*x + b0 # (2.1)
```

Application of the matrix approach to a small but real dataset

```
Q <- read.csv("playbill.csv")
Y <- Q$CurrentWeek; n <- length(Y)
X <- matrix(c(rep(1,n),Q$LastWeek),ncol=2,byrow=FALSE)
BetaHat <- solve(t(X)%*%X)%*%t(X)%*%Y
Yhat <- X%*%BetaHat
print(BetaHat)
```

```
##              [,1]
## [1,] 6804.8860355
## [2,]    0.9820815
```

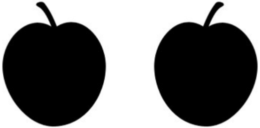
Optional material: In "Octave", `BetaHat = inv(X'*X)*X'*Y`

Properties of least squares estimates

Recall that $\mathrm{E}(\hat{\beta}_0) = \beta_0$ and $\mathrm{E}(\hat{\beta}_1) = \beta_1$, and that

$$\operatorname{var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{S_{xx}}, \qquad \operatorname{var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right] \quad \text{and} \quad \operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \bar{x}}{S_{xx}}$$

Let's now confirm that our new equations give this as well.



$$E(\underset{\sim}{\hat{\beta}}) = E\left[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}\right] \qquad \underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e}$$

tildes underneath

$$= \underbrace{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}}_{I}\,(\underset{\sim}{\beta})$$

$$= I \quad \beta$$

$$E(\hat{\beta}_0) = \beta_0$$

$$E(\hat{\beta}_1) = \beta_1$$

$$E(\underset{\sim}{\hat{\beta}}) = \underset{\sim}{\beta}$$

$$\cancel{\operatorname{var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}}$$

$$E(AY) = AE(Y)$$

$$\mathrm{var}(\underset{\sim}{\hat{\beta}}) = \mathrm{var}\left((X'X)^{-1}X'Y\right)$$

$$= (X'X)^{-1}X' \underbrace{\mathrm{var}(Y)}_{\downarrow} \left[(X'X)^{-1}X'\right]' \qquad (AB)' = B'A'$$

$$= \quad '' \quad \sigma^2 I \quad ''$$

$$\frac{n}{S_{xx}}\begin{pmatrix} \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

$$= \sigma^2 (X'X)^{-1}X' X \cancel{\left[(X'X)^{-1}\right]'}$$

ignore the transpose because symmetric

$$\mathrm{var}(\underset{\sim}{\hat{\beta}}) = \sigma^2 (X'X)^{-1}$$

$$\cancel{\mathrm{var}(\hat{\beta}_0) = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}$$

$$\mathrm{var}(\hat{\beta}) = \sigma^2 \begin{bmatrix} \frac{\overline{x^2}}{S_{xx}} & -\frac{\bar{x}}{S_{xx}} \\ -\frac{\bar{x}}{S_{xx}} & \frac{1}{S_{xx}} \end{bmatrix} = \begin{pmatrix} \mathrm{var}(\hat{\beta}_0) & \mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \mathrm{cov}(\cdot\,\cdot) & \mathrm{var}(\hat{\beta}_1) \end{pmatrix}$$

RHS

$$\mathrm{LHS} = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$$

Show LHS=RHS

$$= \sigma^2\left(\frac{\cancel{n}\overline{x^2} - \cancel{n\bar{x}^2} + \cancel{\bar{x}^2 n}}{\cancel{n}S_{xx}}\right)$$

from the aid sheet

Exercise: RHS → LHS

$$= \frac{\sigma^2\,\overline{x^2}}{S_{xx}} \checkmark$$

$$\mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2\bar{x}}{S_{xx}} \checkmark$$

Next steps

- After the midterm we'll continue in Chapter 5, covering all of it eventually
- At TA office hours, often only zero or one student shows up (e.g. 17 October most recently) so the plan is that no new ones will be added besides those already in place