

# Multiple Regression

Ch 2 in  
Sen & Srivastava

2.1

## Scalar form

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \quad \text{for } i=1, \dots, n, \text{ where}$$

$\beta_j$  are unknown constants

$x_{ij}$  are observable known constants

$\varepsilon_1, \dots, \varepsilon_n$  are unobservable random variables with

$$E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) = \sigma^2, \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for } i \neq j$$

$\uparrow$   
unknown

$y_i$  are observable random variables

## Matrix form

$$y = X\beta + \varepsilon, \text{ where}$$

WE WILL ALWAYS ASSUME  
 $n > k+1$

$\beta$  is a  $(k+1) \times 1$  vector of unknown constants

$X$  is an  $n \times (k+1)$  matrix of observable, known constants

$\varepsilon$  is an  $n \times 1$  unobservable random vector with

$$E(\varepsilon) = 0, \quad \text{Cov}(\varepsilon) = \sigma^2 I, \quad \sigma^2_{1 \times 1} \text{ unknown}$$

$y$  is an  $n \times 1$  observable random vector

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$y = X\beta + \varepsilon$

# Least Squares

2.2

$E(y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$  Estimate  $\beta_5$  by minimizing

$$S = \sum_{i=1}^n (y_i - E(y_i))^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})^2$$

$$\begin{aligned} \frac{\partial S}{\partial \beta_0} &= \frac{\partial}{\partial \beta_0} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})^2 \\ &= \sum_{i=1}^n \frac{\partial}{\partial \beta_0} (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})^2 \\ &= \sum_{i=1}^n 2(y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})(-1) \\ &= -2 \left( \sum_{i=1}^n y_i - n\beta_0 - \beta_1 \sum_{i=1}^n x_{i1} - \dots - \beta_k \sum_{i=1}^n x_{ik} \right) \stackrel{\text{set}}{=} 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial S}{\partial \beta_1} &= \sum_{i=1}^n \frac{\partial}{\partial \beta_1} (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})^2 \\ &= \sum_{i=1}^n 2(y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})(-x_{i1}) \\ &= -2 \sum_{i=1}^n (x_{i1} y_i - \beta_0 x_{i1} - \beta_1 x_{i1}^2 - \beta_2 x_{i1} x_{i2} - \dots - \beta_k x_{i1} x_{ik}) \\ &= -2 \left( \sum_{i=1}^n x_{i1} y_i - \beta_0 \sum_{i=1}^n x_{i1} - \beta_1 \sum_{i=1}^n x_{i1}^2 - \dots - \beta_k \sum_{i=1}^n x_{i1} x_{ik} \right) \\ &\quad \vdots \\ &\quad \stackrel{\text{set}}{=} 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial S}{\partial \beta_j} &= \sum_{i=1}^n \frac{\partial}{\partial \beta_j} (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_j x_{ij} - \dots - \beta_k x_{ik})^2 \\ &= \sum_{i=1}^n 2(y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_j x_{ij} - \dots - \beta_k x_{ik})(-x_{ij}) \\ &= -2 \left( \sum_{i=1}^n x_{ij} y_i - \beta_0 \sum_{i=1}^n x_{ij} - \dots - \beta_j x_{ij}^2 - \dots - \beta_k \sum_{i=1}^n x_{ij} x_{ik} \right) \\ &\quad \stackrel{\text{set}}{=} 0 \end{aligned}$$

$$\frac{\partial S}{\partial \beta_k} = -2 \left( \sum_{i=1}^n x_{ik} y_i - \beta_0 \sum_{i=1}^n x_{i1} - \dots - \beta_k \sum_{i=1}^n x_{i,k}^2 \right) \stackrel{\text{set}}{=} 0$$

Divide by -2 & re-arrange, obtaining

$$\beta_0 n + \beta_1 \sum_{i=1}^n x_{i1} + \beta_2 \sum_{i=1}^n x_{i2} + \dots + \beta_k \sum_{i=1}^n x_{i,k} = \sum_{i=1}^n y_i$$

$$\beta_0 \sum_{i=1}^n x_{i1} + \beta_1 \sum_{i=1}^n x_{i1}^2 + \beta_2 \sum_{i=1}^n x_{i1} x_{i2} + \dots + \beta_k \sum_{i=1}^n x_{i1} x_{i,k} = \sum_{i=1}^n x_{i1} y_i$$

⋮

$$\beta_0 \sum_{i=1}^n x_{i,n} + \beta_1 \sum_{i=1}^n x_{i1} x_{i,n} + \beta_2 \sum_{i=1}^n x_{i2} x_{i,n} + \dots + \beta_k \sum_{i=1}^n x_{i,k}^2 = \sum_{i=1}^n x_{i,n} y_i$$

In matrix form

Leave up

$$\begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \dots & \sum_{i=1}^n x_{i,k} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1} x_{i2} & \dots & \sum_{i=1}^n x_{i1} x_{i,k} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i1} x_{i2} & \sum_{i=1}^n x_{i2}^2 & \dots & \sum_{i=1}^n x_{i2} x_{i,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{i,n} & \sum_{i=1}^n x_{i1} x_{i,n} & \sum_{i=1}^n x_{i2} x_{i,n} & \dots & \sum_{i=1}^n x_{i,n}^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1} y_i \\ \vdots \\ \sum_{i=1}^n x_{i,n} y_i \end{pmatrix}$$

$$X'X$$

$$\beta = X'y$$

I don't think it's so easy to recognize. See next page

$$X'X =$$

write this first

2.4

$$\begin{pmatrix} | & | & | & \dots \\ x_{11} & x_{21} & x_{31} & \dots \\ x_{12} & x_{22} & x_{32} & \dots \\ \vdots & \vdots & \vdots & \\ x_{1k} & x_{2k} & x_{3k} & \dots \end{pmatrix} \begin{pmatrix} | \\ x_{n1} \\ x_{n2} \\ \vdots \\ x_{nk} \end{pmatrix} = \begin{pmatrix} | & x_{11} & x_{12} & \dots & x_{1k} \\ | & x_{21} & x_{22} & \dots & x_{2k} \\ | & x_{31} & x_{32} & \dots & x_{3k} \\ \vdots & \vdots & \vdots & & \vdots \\ | & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

$$= \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \dots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \dots & \sum_{i=1}^n x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{i1}x_{ik} & \sum_{i=1}^n x_{i2}x_{ik} & \dots & \sum_{i=1}^n x_{ik}^2 \end{pmatrix}$$

$$X'Y =$$

2.5

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \dots & x_{nk} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \dots & x_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1} y_i \\ \vdots \\ \sum_{i=1}^n x_{in} y_i \end{pmatrix}$$

$(k+1) \times n$        $n \times 1$        $n \times 1$

So we have the normal equations

$$(X'X)\beta = X'Y$$

Our text says that the least squares estimates  $b$  of  $\beta$  must satisfy

$$(X'X)b = X'Y \quad (2.8)$$

And if  $X'X$  has an inverse,

$$b = (X'X)^{-1} X'Y \quad (2.9)$$

If no inverse, uses generalized inverses, but we won't go there. ( $B = A^-$  means  $ABA = A$ )

When does  $(X'X)^{-1}$  exist?

Theorem  $(X'X)^{-1}$  exists if

2.6

and only if the cols of  $X$  are linearly independent.

Proof

First note  $X'X$  is symmetric, since  $(X'X)' = X'X$ , and also non-negative definite, since

$$a'X'Xa = (Xa)'Xa = \sum_{i=1}^n \beta_i^2 \geq 0$$

(a) Cols of  $X$  are linearly independent.

This means  $Xa = 0$  implies  $a = 0$   
 $\uparrow$   
 $(k+1)x_1$

To show  $X'X$  p.d., look at  $a'X'Xa \geq 0$  because we've seen it's non-negative definite. The only vector  $a$  that makes it zero is  $a=0$ , because

$$a'X'Xa = \sum \beta_i^2 = 0 \text{ implies } \beta = Xa = 0 \Rightarrow a = 0$$

by linear independence. Hence

(b)  $X'X$  is positive definite

Therefore the eigenvalues are all positive. Since  $X'X$  is symmetric, we have the Spectral Decomposition  $X'X = CDC'$ , and  $(X'X)^{-1} = CD^{-1}C'$ , so that

(c)  $(X'X)^{-1}$  exists

If  $(X'X)^{-1}$  exists. To show cols of  $X$  l.i., let  $Xa = 0$   
† show  $a = 0$ .

$$Xa = 0 \Rightarrow X'Xa = X'0 = 0$$

$$\Rightarrow (X'X)^{-1}X'Xa = X'0 = 0 \Rightarrow a = 0 \text{ and } (c) \Rightarrow (a)$$

We will always assume that the columns of  $X$  are linearly independent, which just means that the independent variables are not redundant.

2.7

We have solved the normal equations to obtain the least-squares estimator

$$b = (X'X)^{-1} X'y \quad b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}$$

At least that's the point where all the derivatives are zero. We will prove  $S$  is MINIMIZED there later.

First more notation & concepts

Predicted  $y$

$$\hat{y} = Xb$$

$$\hat{y}_i = b_0 + b_1 x_{i1} + \dots + b_k x_{ik}$$

A point on the least-squares plane

Residual

$$e = y - \hat{y}$$

$$e_i = y_i - \hat{y}_i$$

vertical distance of  $y_i$  from the least-squares plane

The "hat" matrix  $H$

$$\hat{y} = Xb = \underbrace{X(X'X)^{-1}X'}_{H_{n \times n}} y = Hy$$

The hat matrix puts a hat on  $y$

- Symmetric (Show  $H' = H$ )
- Idempotent meaning  $HH = H$

$$\begin{aligned} HH &= (X(X'X)^{-1}X')(X(X'X)^{-1}X') \\ &= X(X'X)^{-1} \underbrace{X'X(X'X)^{-1}}_I X' = X(X'X)^{-1}X' = H \end{aligned}$$

Also note

$$\begin{aligned} e &= y - \hat{y} = y - Hy = \underbrace{I}_{n \times n} y - Hy \\ &= (I - H)y = My \quad \text{book's notation} \end{aligned}$$

Also symmetric & idempotent

Show



Thm 2.1 (two best part)

from the text  $X'e = 0$

2.9

$$\begin{aligned}
X'e &= X'(y - \hat{y}) = X'y - X'Xb \\
&= X'y - X'X(X'X)^{-1}X'y \\
&= X'y - X'y = 0
\end{aligned}$$

Now we will see that  $b$  actually minimize

$$S = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})^2 = (y - X\beta)'(y - X\beta)$$

$$S = (y - \hat{y} + \hat{y} - X\beta)'(y - \hat{y} + \hat{y} - X\beta)$$

$$= (e + Xb - X\beta)'(e + Xb - X\beta)$$

$$= (e + X(b - \beta))'(e + X(b - \beta))$$

$$= e'e + \underbrace{e'X(b - \beta)}_{0 \text{ (by } X'e)} + (X(b - \beta))'e + (X(b - \beta))'X(b - \beta)$$

$$= e'e + (b - \beta)' \underbrace{X'e}_{0 \text{ (by } X'e)} + (b - \beta)'X'X(b - \beta)$$

$$= e'e + (b - \beta)'X'X(b - \beta)$$

$$\uparrow$$

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$\uparrow$   
 POSITIVE definite  $\textcircled{A} \leftarrow \textcircled{B}$

p. 37 but not even an optimization number. it's an important stopping sign

So the second term is POSITIVE except when  $\beta = b$ , at then it's zero - unique minimum

# Projections

Not in  
the text

2.10

$$\mathcal{V} = \left\{ v \in \mathbb{R}^n : v = Xa, a \in \mathbb{R}^{k+1} \right\}$$

The space spanned by the columns of  $X$

Some important vectors are in  $\mathcal{V}$

$$y = X\beta + \varepsilon \quad E(y) = X\beta \in \mathcal{V} \quad \beta \text{ is an } a$$

$$\hat{y} = Xb \in \mathcal{V} \quad b \text{ is an } a$$

Q Is  $y \in \mathcal{V}$ ? <sup>No!</sup> Suppose the answer is Yes

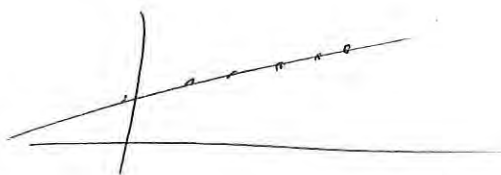
$$y = Xa \Rightarrow X'Xa = X'y$$

$$\Rightarrow (X'X)^{-1}X'Xa = (X'X)^{-1}X'y$$

$$\Rightarrow a = b. \text{ That is } y = Xb, \text{ exactly}$$

In scalar form,

$$y_i = b_0 + b_1 x_{i1} + \dots + b_k x_{ik}, \text{ exactly}$$



All points are  
exactly on the least  
squares plane

$$e_i = y_i - \hat{y}_i = 0 \quad \forall i$$

So assume  $y \notin V$ , because otherwise something funny is going on, with the model.

2.11

What's ~~the best~~ point in  $V$  that is closest to  $y$ ? Euclidean distance is

$$\sqrt{(y_1 - p_1)^2 + (y_2 - p_2)^2 + \dots + (y_n - p_n)^2}$$

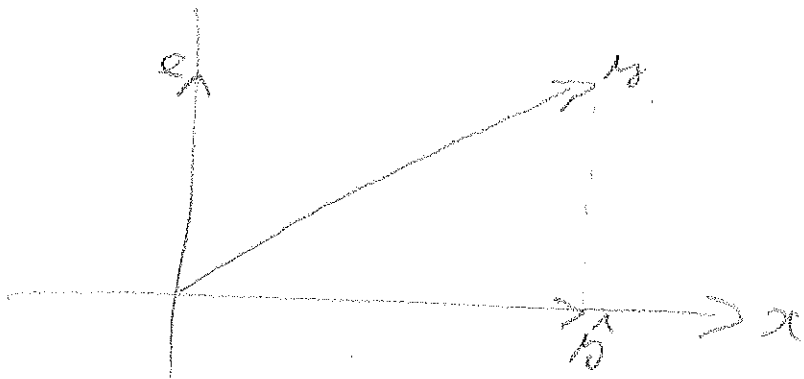
$$p \in V = Xa, \text{ some } a$$

minimize  $(y - p)'(y - p) = (y - Xa)'(y - Xa)$  over all  $a$

Hey, we've already done this, calling  $a$   $\beta$  by the name

$$\text{Answer is } p = Xb = \hat{y}$$

So  $\hat{y}$  is the point in  $V$  that is closest to  $y$



$$e = y - \hat{y}$$

$$\Leftrightarrow y = \hat{y} + e$$

$\hat{y}$  is the projection (shadow) of  $y$  on  $V$

$$\hat{y} = Hy, \text{ } H \text{ is the Projection Operator}$$

$y$  is special to us, but mathematically it could be any point in  $\mathbb{R}^n$ . Let  $z \in \mathbb{R}^n$ . Then  $Hzy$  is the point in  $V$  closest to  $Pz$

# Projections

2.12

- Picture suggests  $\vec{y} \perp e$ . True?

$$\vec{y} \cdot \vec{e} = (Xb)'e = b' \underbrace{X'e}_0 = 0$$

- In fact,  $e$  is perpendicular to any point in  $\mathcal{V}$ . Let  $v \in \mathcal{V}$  so  $v = Xa$   
 $v' e = (Xa)' e = a' X' e = 0$

- If you shine the light on any point in  $\mathcal{V}$ , shadow should be right on the point. Let  $v \in \mathcal{V}$ .

$$Hv = HXa = X(X'X)^{-1} X'a = Xa = v$$

yes

## Expected values and variance-covariance

2.13

$$\bullet b = (X'X)^{-1} X'y$$

$$E(b) = (X'X)^{-1} X' E(y) = (X'X)^{-1} X' X \beta = \beta$$

unbiased

$$\begin{aligned} \text{cov}(b) &= \text{cov}\left((X'X)^{-1} X'y\right) = (X'X)^{-1} X' \text{cov}(y) (X'X)^{-1} \\ &= (X'X)^{-1} X' \sigma^2 I_n X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} X' X (X'X)^{-1} = \sigma^2 (X'X)^{-1} \end{aligned}$$

(Book shows  $b \xrightarrow{p} \beta$ , consistent. We will skip the large-sample theory.)

$$\bullet \hat{y}_0 = Xb$$

$$E(\hat{y}_0) = E(Xb) = X E(b) = X \beta = E(y_0)$$

So  $\hat{y}_0$  is described as "predicted  $y_0$ ," but statistically it's an unbiased estimate of  $E(y_0)$ , which is also a reasonable prediction.

$$\begin{aligned} \text{cov}(\hat{y}_0) &= \text{cov}(Xb) = X \text{cov}(b) X' \\ &= X \sigma^2 (X'X)^{-1} X' = \sigma^2 H \end{aligned}$$

2.14

$$e = y - \hat{y}$$

$$E(e) = E(y) - E(\hat{y}) = X\beta - X\beta = 0$$

$$\begin{aligned} \text{cov}(e) &= \text{cov}(y - Hy) = \text{cov}(Iy - Hy) \\ &= \text{cov}((I-H)y) = (I-H)\text{cov}(y)(I-H)' \\ &= (I-H)\sigma^2 I_n (I-H) = \sigma^2 (I-H) \end{aligned}$$

Book has  $M = (I-H)$

$$e = y - \hat{y} = y - Hy = (I-H)y$$

$$= (I-H)(X\beta + \varepsilon) = (I-H)X\beta + (I-H)\varepsilon$$

$$= X\beta - HX\beta + (I-H)\varepsilon$$

$$= X\beta - X(X'X)^{-1}X'X\beta + (I-H)\varepsilon = M\varepsilon$$

And could have done

$$e = M\varepsilon \Rightarrow \text{cov}(e) = M\sigma^2 I M' = \sigma^2 M$$

or  $\sigma^2 (I-H)$

$M$  is useful it seems

## Estimation of $\sigma^2$

2.15

Parameters of the model are  $(\beta, \sigma^2)$

Unbiased estimate of  $\beta$  is  $b$ .

How should we estimate  $\sigma^2$ ?

Seek an UNBIASED ESTIMATE

$$E(\hat{\sigma}^2) = \sigma^2$$

Use  $\text{tr}(AB) = \text{tr}(BA)$

For example,  $\text{tr}(H) = \text{tr}\left(\underset{n \times (k+1)}{X} \underset{(k+1) \times n}{(X'X)^{-1} X'}\right)$

$$= \text{tr}\left(\underset{(k+1) \times (k+1)}{(X'X)^{-1} X' X}\right) = \text{tr}(I_{k+1}) = k+1$$

So  $\text{tr}(H) = k+1$

Start with spread of points  
around the least-squares plane

2.16

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = e'e$$

Compare text

$$E(e'e) = E(\text{tr}(e'e)) = E(\text{tr}((ME)'ME))$$

$$= E(\text{tr}(\varepsilon'M'M\varepsilon)) = E(\text{tr}(\varepsilon'MM\varepsilon))$$

$$= E(\text{tr}(\varepsilon'M\varepsilon)) = E(\text{tr}(M\varepsilon\varepsilon'))$$

$$= \text{tr}(M E(\varepsilon\varepsilon'))$$

$$\text{Now } \text{cov}(\varepsilon) = E(\varepsilon - 0)(\varepsilon - 0)' = \sigma^2 I_n$$

$$= \text{tr}(M \sigma^2 I_n) = \sigma^2 \text{tr}(M)$$

$$= \sigma^2 \text{tr}(I - H) = \sigma^2 (\text{tr}(I_n) - \text{tr}(H))$$

$$= \sigma^2 (n - k - 1) \quad (!)$$

Theorem 2.3 says

$$s^2 = \frac{e'e}{n - k - 1} \text{ is an unbiased estimator of } \sigma^2$$

Consistency Skipped

- $s^2$  looks like the sample variance, and we'll see that the usual sample variance is a special case.
- $e'e$  is often called SSE  $\equiv$   
 $s^2$  is called MSE



# Decomposition of SS

2.17

(See Thm 2.4 & Cor 2.2)

Suppose the model has an intercept (not all do).  
Then because

Leave up  $\rightarrow$   $X'e = 0$   $\implies \sum_{i=1}^n e_i = 0$

$(k+1) \times n$   $n \times 1$   $(k+1) \times 1$

EXPLAIN

Theorem (Cor 2.2 in the text)

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

SSTO = SSE + SSR

Total variation to explain

Unexplained

Must be explained

$$\begin{aligned} \text{SSTO} &= \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (e_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n e_i^2 + 2 \sum_{i=1}^n e_i (\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ &= \text{SSE} + \text{SSR} + 2 \left( \sum_{i=1}^n e_i \hat{y}_i - \bar{y} \sum_{i=1}^n e_i \right) \\ &\quad \quad \quad e' \hat{y} = \hat{y}' e \quad \quad \quad \underbrace{\sum_{i=1}^n e_i}_0 \\ &= \text{SSE} + \text{SSR} + 2 \underbrace{(Xb)' e}_0 \end{aligned}$$

$$R^2 = \frac{\text{Explained}}{\text{Total}} = \frac{\text{SSR}}{\text{SSTO}} = \frac{\text{SSTO} - \text{SSE}}{\text{SSTO}} = 1 - \frac{\text{SSE}}{\text{SSTO}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Proportion Explained

Then

For simple regression,  $R^2 = r^2$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Have

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad \text{and}$$

2.18

$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$r =$

$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

read on formula sheet

So that  $b_1 = r \frac{S_y}{S_x}$

$$= r \frac{\sqrt{\sum (y_i - \bar{y})^2 / (n-1)}}{\sqrt{\sum (x_i - \bar{x})^2 / (n-1)}}$$

Now  $R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

$$\hat{y}_i = b_0 + b_1 x_i = \bar{y} - b_1 \bar{x} + b_1 x_i = \bar{y} + b_1 (x_i - \bar{x})$$

$$\text{So } R^2 = \frac{\sum_{i=1}^n (\bar{y} + b_1 (x_i - \bar{x}) - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$= \frac{b_1^2 \sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2} = \left( r \frac{S_y}{S_x} \right)^2 \frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}$$

$$\frac{\sum (y_i - \bar{y})^2 / (n-1)}{\sum (x_i - \bar{x})^2 / (n-1)} \frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}$$

$$= r^2 \frac{\sum (y_i - \bar{y})^2 / (n-1)}{\sum (x_i - \bar{x})^2 / (n-1)} \frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}$$

$$= r^2$$

# The Gauss-Markov Theorem

2.19

Idea is that  $b$  is the "best" estimate of  $\beta$

Seek to estimate  $l'\beta = l_1\beta_0 + l_2\beta_1 + \dots + l_{k+1}\beta_k$

For example, any  $\beta_i$

$$\begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \beta_2$$

$l'$   $\beta$

or difference, or average, or  $E(y)$  for a new set of  $x$  values

$$\begin{pmatrix} 1 & x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Natural choice is  $l' \hat{b}$

• unbiased  $E(l'\hat{b}) = l'E(\hat{b}) = l'\beta \quad \forall \beta$

• Linear combination of  $y$

$$l'\hat{b} = \underbrace{l'(X'X)^{-1}X'}_{1 \times n} y = c_0' y$$

G-M Theorem says

2.20

$l'b$  is better than any other unbiased linear combination  $c'y$  in the sense of having smaller variance  
Say it's B. L. U. E.

$$\text{Var}(c'y) = c' \sigma^2 I_n c = \sigma^2 c'c = \sigma^2 \sum_{i=1}^n c_i^2$$

Want to find  $c_1, \dots, c_n$  to make this as small as possible, but  $c_1 = \dots = c_n = 0$  does it, so not just any  $c$ .

Want it unbiased for  $l'\beta$ , so

$$E(c'y) = l'\beta \quad \forall \beta \in \mathbb{R}^{k+1}$$

$$c'E(y) = c'X\beta$$

The clever part: It must hold for all  $\beta \in \mathbb{R}^{k+1}$ , even strange, unnatural examples.

$$\underbrace{c'X}_{\substack{n' \\ (1 \times (k+1))}} \beta = l'\beta$$

$$(r_1 \ r_2 \ r_3 \ r_4) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = (l_1 \ l_2 \ l_3 \ l_4) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

Erase & put zeros & ones, so  $l' = c'X$

# Theorem 2.5

2.21

For the general linear regression model  $y = X\beta + \varepsilon$  with  $E(\varepsilon) = 0$  and  $\text{cov}(\varepsilon) = \sigma^2 I$ ,  $l'b$  has the smallest variance of any linear combination of  $y$  values that is unbiased for  $l'\beta$

Proof

- $l'b = l'(X'X)^{-1}X'y = c_0'y$  is a linear combination of  $y$  values with  $\boxed{c_0' = l'(X'X)^{-1}X' \Leftrightarrow c_0 = X(X'X)^{-1}l}$  \*
- $E(l'b) = l'E(b) = l'\beta$ , unbiased
- Let  $c'y$  be a general linear combination with  $E(c'y) = c'X\beta = l'\beta$  for all  $\beta$ , so that  $\boxed{c'X = l' \Leftrightarrow l = X'c}$  \*\*

$$\begin{aligned} \text{Now } \text{Var}(c'y) - \text{Var}(l'b) &= c'\sigma^2 I_n c - \sigma^2 l'(X'X)^{-1}l \\ &\stackrel{**}{=} \sigma^2 (c'c - c'X(X'X)^{-1}X'c) \\ &= \sigma^2 (c'c - c'Hc) = \sigma^2 (c'Ic - c'Hc) \\ &= \sigma^2 c'(I-H)c = \sigma^2 c'(I-H)(I-H)c \\ &= \sigma^2 \underbrace{c'(I-H)}_{z'} \underbrace{(I-H)c}_{z} \geq 0 \text{ and } \text{Var}(c'y) \geq \text{Var}(l'b) \end{aligned}$$

If the variances are equal,

$$(I-H)c = 0 \Leftrightarrow c = Hc = X(X'X)^{-1}X'c = X(X'X)^{-1}l \stackrel{(*)}{=} c_0$$

So  $c_0'y = l'b$  is the unique linear combination with the smallest variance.  $\square$

The geometry is remarkable

2.22

Because  $l = X\hat{c}$ , we've found that

$$c_0 = Hc$$

For any  $c \in \mathbb{R}^n$  with  $E(c; y_0) = l^T \beta$

∞ many

That is,  $c_0$  is the projection of  $c$

on  $\mathcal{V} = \left\{ v \in \mathbb{R}^n : v = Xa, a \in \mathbb{R}^{k+1} \right\}$

and indeed  $c_0 = X \underbrace{(X^T X)^{-1} X^T c}_{a_{(k+1) \times 1}} \in \mathcal{V}$

$c$  is like a  $\vec{c}$

$c_0$  is like a  $\hat{c}$

$c - c_0$  is like an  $e$

(so should have  $c - c_0 \perp v \quad \forall v \in \mathcal{V}$ )

HW?