Multiple Regression

2.1

Ch 2 in Sen & Srivastava

Scalar form

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

β_j are unknown constants

x_{ij} are observable known constants

$\varepsilon_1 \ldots \varepsilon_n$ are unobservable random variables with

$E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$ (unknown), $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$

y_i are observable random variables

Matrix form

WE WILL ALWAYS ASSUME $n > k+1$

$y = X\beta + \varepsilon$, where

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

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

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

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

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

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

$$y = X \quad \beta + \varepsilon$$

Least Squares

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

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

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

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_0} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2$$

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

$$= -2\left(\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_{i1} - \cdots - \beta_k \sum_{i=1}^{n} x_{ik} \right) \stackrel{set}{=} 0$$

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

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

$$= -2 \sum_{i=1}^{n} (x_{i1} y_i - \beta_0 x_{i1} - \beta_1 x_{i1}^2 - \beta_2 x_{i1} x_{i2} - \cdots - \beta_k x_{i1} x_{ik})$$

$$= -2\left(\sum_{i=1}^{n} x_{i1} y_i - \beta_0 \sum_{i=1}^{n} x_{i1} - \beta_1 \sum_{i=1}^{n} x_{i1}^2 - \cdots - \beta_k \sum_{i=1}^{n} x_{i1} x_{ik} \right) \stackrel{set}{=} 0$$

$$\vdots$$

$$\frac{\partial S}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_j x_{ij} - \cdots - \beta_k x_{ik})^2$$

$$= \sum_{i=1}^{n} 2 (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_j x_{ij} - \cdots - \beta_k x_{ik})(-x_{ij})$$

$$= -2\left(\sum_{i=1}^{n} x_{ij} y_i - \beta_0 \sum_{i=1}^{n} x_{i1} x_{ij} - \cdots - \beta_j x_{ij}^2 - \cdots - \beta_k \sum_{i=1}^{n} x_{ij} x_{ik} \right) \stackrel{set}{=} 0$$

continuing

2.3

$$\frac{\partial S}{\partial \beta_k} = -2\left(\sum_{i=1}^{n} x_{ik} y_i - \beta_0 \sum_{i=1}^{n} x_{ik} - \cdots - \beta_k \sum_{i=1}^{n} x_{ik}^2 \right) \overset{set}{=} 0$$

Divide by -2 & re-arrange, obtaining

$$\beta_0 n + \beta_1 \sum_{i=1}^{n} x_{i1} + \beta_2 \sum_{i=1}^{n} x_{i2} + \cdots + \beta_k \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i$$

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

$$\vdots$$

$$\beta_0 \sum_{i=1}^{n} x_{ik} + \beta_1 \sum_{i=1}^{n} x_{i1} x_{ik} + \beta_2 \sum_{i=1}^{n} x_{i1} x_{ik} + \cdots + \beta_k \sum_{i=1}^{n} x_{ik}^2 = \sum_{i=1}^{n} x_{ik} y_i$$

In matrix form

Leave up

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ik} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i1}x_{i2} & \sum_{i=1}^{n} x_{i2}^2 & \cdots & \sum_{i=1}^{n} x_{i2}x_{ik} \\ \vdots & \vdots & & & \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{i1}x_{ik} & \sum_{i=1}^{n} x_{i2}x_{ik} & \cdots & \sum_{i=1}^{n} x_{ik}^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik} y_i \end{pmatrix}$$

$$X'X \qquad \beta = X'y$$

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

$$X'X =$$

2.4

Write this first

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

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

2.5

$X'y =$

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & & & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{pmatrix}_{(k+1)\times n} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1} = \begin{pmatrix} \sum y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik} y_i \end{pmatrix}$$

So we have the normal equations

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

Our text says that the least squares estimate b of β must satisfy

$$(X'X)\,b = X'y \qquad (2.8)$$

And if $X'X$ has an inverse,

$$b = (X'X)^{-1}X'y \qquad (2.9)$$

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 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 $a'X'Xa = \underset{1\times n}{(Xa)'}\underset{n\times 1}{Xa}$

$$= z'z = \sum_{i=1}^{n} z_i^2 \geq 0$$

(a) Cols of X are linearly independent.

This means $\underset{(k+1)\times 1}{Xa} = 0$ implies $a = 0$

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

$a'X'Xa = z'z = 0$ implies $z = Xa = 0 \Rightarrow a = 0$ 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 $a=0$.

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

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

2.7

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

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

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

At least that's the point where all the derivatives are zero. We will prove S is <u>MINIMIZED</u> there later.
First more notation & concepts

Predicted y $\qquad \hat{y} = Xb \qquad \hat{y}_i = b_0 + b_1 x_{i1} + \cdots + b_k x_{ik}$

A point on the least-squares plane

Residual $\qquad e = y - \hat{y} \qquad e_i = y_i - \hat{y}_i$

vertical distance of y_i from the least-squares plane

2.8

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 $HH = H$

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

$$= X(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}X' = X(X'X)^{-1}X' = H$$

Also note

$$e = y - \hat{y} = y - Hy = \underset{n\times n}{I}y - Hy$$

$$= (I - H)y = My \quad \text{book's notation}$$

Also symmetric & idempotent

Show

Thm 2.1 (the last part) from the text $X'e = 0$

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

Now we will see that b actually minimize

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

$$
\begin{aligned}
S &= (y - \hat{y} + \hat{y} - X\beta)'(y - \hat{y} + \hat{y} - X\beta) \\
&= (e + Xb - X\beta)'(e + Xb - X\beta) \\
&= (e + X(b - \beta))'(e + X(b - \beta)) \\
&= e'e + \underbrace{e'X}_{0_{1\times k+1}}(b - \beta) + (X(b - \beta))'e + (X(b - \beta))'X(b - \beta) \\
&= e'e + (b - \beta)'\underbrace{X'e}_{0_{(k+1)\times 1}} + (b - \beta)'X'X(b - \beta) \\
&= e'e + (b - \beta)'\underbrace{X'X}(b - \beta)
\end{aligned}
$$

P. 37 but not even an equation number. It's an important stepping stone

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

POSITIVE definite

$$
\begin{tikzcd}
& a \arrow[dr] & \\
c \arrow[ur] & & b \arrow[ll, Rightarrow]
\end{tikzcd}
$$

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

Projections

Not in the text

2.10

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

The space spanned by the columns of X

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

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

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

Is $y \in \mathcal{V}$? No! Suppose the answer is Yes

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

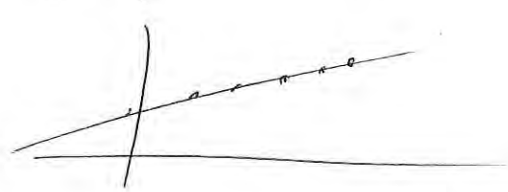
$$\Rightarrow (X'X)^{-1}X'Xa = (X'X)^{-1}X'y$$

$$\Rightarrow a = b$$

That is $y = Xb$, exactly

In scalar form,

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



All points are exactly on the least squares plane

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

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

What's point in V that is closest to y? Euclidean distance is

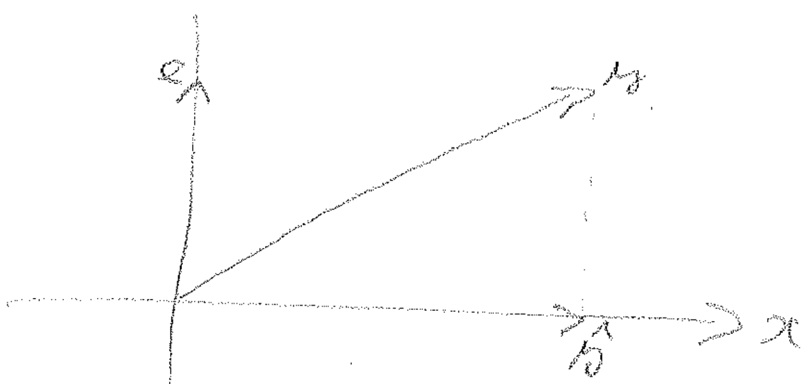
$$\sqrt{(y_1-p_1)^2+(y_2-p_2)^2+\cdots(y_n-p_n)^2} \qquad \underset{\sim}{p} = Xa \text{, some } a$$

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

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

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

So $\hat{y}$ is the point in V that is <u>closest</u> to y



$e = y - \hat{y}$

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

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

$\hat{y} = Hy$, H 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 Hz is the point in V closest to z

Projections

2.12

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

$$\hat{y}'e = (Xb)'e = b'\underbrace{X'e}_{0} = 0$$

- 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 $v \in \mathcal{V}$.

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

yes

Expected value and variance-covariance

2.13

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

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

$cov(b) = cov\left((X'X)^{-1}X'y\right) = (X'X)^{-1}X'cov(y)\left((X'X)^{-1}X'\right)'$

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

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

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

- $\hat{y} = Xb$

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

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

$cov(\hat{y}) = cov(Xb) = X\,cov(b)X'$

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

2.14

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

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

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

<u>Book has</u> $M = (I - H)$

$$\begin{aligned} e &= y - \hat{y} = y - Hy = (I-H)y \\ &= (I-H)(X\beta + \varepsilon) = (I-H)X\beta + (I-H)\varepsilon \\ &= X\beta - HX\beta + (I-H)\varepsilon \\ &= X\beta - X\underbrace{(X'X)^{-1}X'X}_{I}\beta + (I-H)\varepsilon = M\varepsilon \end{aligned}$$

And could have done

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

$$\text{or } \sigma^2 (I-H)$$

M is useful it seems

Estimation of σ^2

2.15

Parameters of the model are (β, σ^2)
Unbiased estimate of β is b.
How should we estimate σ^2?
Seek an UNBIASED ESTIMATE
$E(\hat{\sigma}^2) = \sigma^2$

Use $tr(AB) = tr(BA)$

For example, $tr(H) = tr(\underbrace{X}_{\substack{A \\ n \times k+1}} \underbrace{(X'X)^{-1}X'}_{\substack{B \\ k+1 \times n}})$

$= tr(\underbrace{(X'X)^{-1}X'X}_{(k+1)\times(k+1)}) = tr(I_{k+1}) = k+1$

So $tr(H) = k+1$

Start with spread of points around the least-squares plane

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

Compare text

$$E(\underset{1\times 1}{e'e}) = E(tr(e'e)) = E(tr((M\varepsilon)'M\varepsilon))$$

$$= E(tr(\varepsilon' M' M \varepsilon)) = E(tr(\varepsilon' M M \varepsilon))$$

$$= E(tr(\underset{\substack{A \\ 1\times n}}{\varepsilon'} \underset{\substack{B \\ n\times 1}}{M\varepsilon})) = E(tr(M\varepsilon\varepsilon'))$$

$$= tr(M\, E(\varepsilon\varepsilon'))$$

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

$$= tr(M \sigma^2 I_n) = \sigma^2 tr(M)$$

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

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

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 the we'll see that the usual sample variance is a special case.
- $e'e$ is often called SSE &
  s^2 is called MSE

Decomposition of SS

(See Thm 2.4 & Cor 2.2)

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

Leave up $\longrightarrow \underset{(k+1)\times n}{X'} \underset{n\times 1}{e} = \underset{(k+1)\times 1}{0}$, $\sum_{i=1}^{n} e_i = 0$.

EXPLAIN

Theorem (Cor 2.2 in the text)

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

$$\text{SSTO} = \text{SSE} + \text{SSR}$$

Total variation to explain | Unexplained | Must be explained

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

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

$$= \text{SSE} + \text{SSR} + 2\Big(\underbrace{\sum_{i=1}^{n} e_i\hat{y}_i}_{e'\hat{y}=\hat{y}'e} - \bar{y}\underbrace{\sum_{i=1}^{n} e_i}_{0}\Big)$$

$$= \text{SSE} + \text{SSR} + 2\underbrace{(Xb)'e}_{0}$$

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

Proportion Explained

Thm

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

2.18

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

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

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

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

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

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

$$= \frac{b_1^2 \sum (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \left(r \frac{s_y}{s_x} \right)^2 \frac{\sum (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

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

$$= r^2$$

18

The Gauss-Markov Theorem

2.19

Idea is that b is the "best" estimate of β

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

For example, any β_i:

$$\underbrace{(0\ 0\ 1\ 0)}_{l'} \underbrace{\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{pmatrix}}_{\beta} = \beta_2$$

or difference, or average, or $E(y_0)$ for a new set of x values

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

Natural choice is $l'b$

- unbiased $E(l'b) = l'E(b) = l'\beta \quad \forall \beta$
- Linear combination of y

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

G-M Theorem says

2.20

$\ell' b$ is better than any other unbiased linear combination $c'y$ in the sense of having smaller variance

Say it's B. L. U. E.

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

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

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

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

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

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

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

$$(\pi_1\ \pi_2\ \pi_3\ \pi_4)\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{pmatrix} = (\ell_1\ \ell_2\ \ell_3\ \ell_4)\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{pmatrix}$$

Erase & Put zeros & ones, so $\ell' = c'X$

Theorem 2.5

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

Proof

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

$$\boxed{c'X = \ell' \iff \ell = X'c} \quad **$$

Now $Var(c'y) - Var(\ell' b) = c'\sigma^2 I_n c - \sigma^2 \ell'(X'X)^{-1}\ell$

$$\overset{**}{=} \sigma^2\left(c'c - c'X(X'X)^{-1}X'c\right)$$

$$= \sigma^2(c'c - c'Hc) = \sigma^2(c'Ic - c'Hc)$$

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

$$= \sigma^2 \underbrace{c'(I-H)'}_{z'}\underbrace{(I-H)c}_{z} \geq 0 \text{ and } Var(c'y) \geq Var(\ell' b)$$

If the variances are equal,

$$(I-H)c = 0 \iff$$

$$c = Hc = X(X'X)^{-1}\underbrace{X'c}_{\ell} = X(X'X)^{-1}\ell \overset{(*)}{=} c_0$$

So $c_0' y = \ell' b$ is the unique linear combination with the smallest variance ∎

The geometry is remarkable

2.22

Because $\ell = X'c$, we've found that

$$c_0 = Hc$$

For any $c \in \mathbb{R}^n$ with $E(c'y) = \ell'\beta$ $\quad \infty$ many

That is, c_0 is the projection of c on $\mathcal{V} = \{ v \in \mathbb{R}^n : v = Xa,\ a \in \mathbb{R}^{k+1} \}$

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

c is like a y

c_0 is like a $\hat{y}$

$c - c_0$ is like an e

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

HW?