The General Linear Model

See Ch 7

Scalar form

For $i = 1, \ldots, n$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad \text{where}$$

β_j are unknown constants

x_{ij} are observable, known constants

$\varepsilon_1, \ldots, \varepsilon_n$ are unobservable random variables with
$E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$, $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$
σ^2 unknown

Y_i are observable random variables

Matrix form

$$Y = X\beta + \varepsilon \quad \text{where}$$

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

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

ε is an $n \times 1$ random vector with
$E(\varepsilon) = 0$, $cov(\varepsilon) = \sigma^2 I_n$
σ^2 unknown

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

β values are called REGRESSION COEFFICIENTS

Meaning of the regression coefficients

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

$$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$ Really $E(Y_i \mid x_{i1}, \ldots x_{ik})$

Imagine a sub-population of Y values at each combination of x values. THIS IS THE population MEAN.

For $k=1$, a straight line
For $k>1$, a (hyper) plane
β_0 is the intercept

$$\frac{\partial E(Y)}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\beta_0 + \beta_1 x_1 + \cdots \beta_j x_j + \cdots + \beta_k x_k\right)$$
$$= \beta_j$$

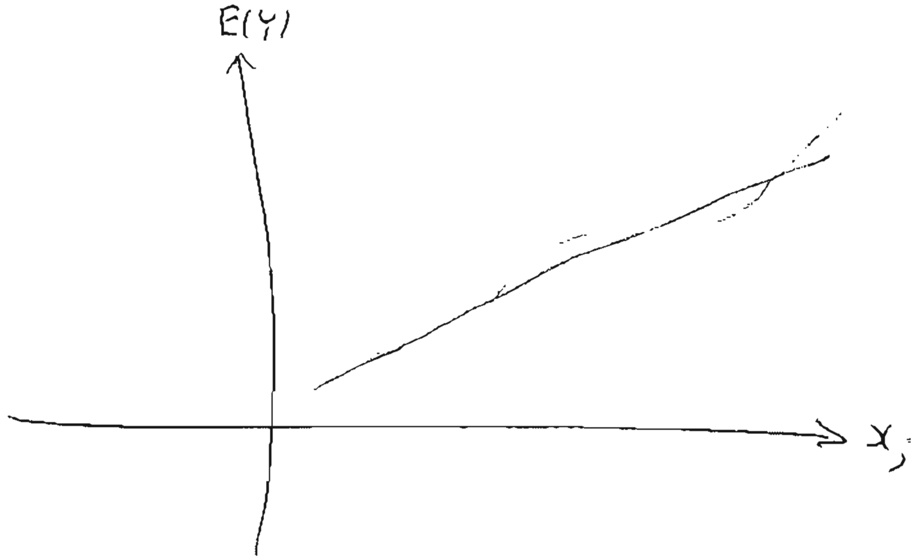
HOLDING CONSTANT = CONTROLLING FOR
"Partial" regression coefficient

It's more: For fixed values of x_ℓ, $\ell \neq j$

$$E(Y) = \left(\beta_0 + \sum_{\ell \neq j} \beta_\ell x_\ell\right) + \beta_j x_j$$

STRAIGHT LINE

- NOT REALLY TRUE, BUT OFTEN CLOSE ENOUGH TO BE USEFUL



- "All models are wrong, but some are useful"

Sometimes not close – try to detect

Say "Controlling for" the other variables

$\beta_j > 0$ x_j positively related to Y

$\beta_j < 0$ x_j neg " " "

$\beta_j = 0$ unrelated

need tests

MODEST GOAL – We really mean x_j related to average Y.

The $\underset{\sim}{X}$ matrix holds the independent variables.

n by k+1

$\uparrow$

Sample size

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

Study $X'X$ You'll see it a lot

- Symmetric: $(X'X)' = X'X'' = X'X$
- Sums of Squares & cross-products (SSCP)

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} \\ x_{13} & x_{23} & x_{33} & \cdots & x_{n3} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ \vdots & \vdots & & \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{bmatrix} = \begin{bmatrix} n & \Sigma x_{i1} & \Sigma x_{i2} & \Sigma x_{i3} \\ \Sigma x_{i1} & \Sigma x_{i1}^2 & \Sigma x_{i1}x_{i2} & \Sigma x_{i1}x_{i3} \\ \Sigma x_{i2} & \Sigma x_{i2}x_{i1} & \Sigma x_{i2}^2 & \Sigma x_{i2}x_{i3} \\ \Sigma x_{i3} & \Sigma x_{i3}x_{i1} & \Sigma x_{i3}x_{i2} & \Sigma x_{i3}^2 \end{bmatrix}$$

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

- At least non-negative definite :

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

Xa: $n \times (k+1)$, $(k+1) \times 1$

Save

- Suppose the columns (independent variables) are linearly independent. This means the IVs are not redundant. No IV is a linear combination of the others (6 quiz marks ~~sum~~ two quiz ave)

<u>Def</u> $(X'X)a = 0 \Rightarrow a = 0$ Save

<u>Theorem</u> $(X'X)^{-1}$ exists if and only if the columns of X are linearly independent

<u>Proof</u>: First linear ind $\Rightarrow$ existence of inverse

Suppose cols of X are linearly independent. Since $a'X'Xa \geq 0$, Showing $a'X'Xa = 0 \Rightarrow a = 0$ will show positive definite.

Let $a'X'Xa = 0 \Rightarrow Xa = 0 \Rightarrow a = 0$ by linear independence

Hence $X'X$ is pos def $\Rightarrow$ eigenvalues all pos

$\Rightarrow (X'X)^{-1}$ exists.

Now $(X'X)^{-1}$ hence l.i.

Assume $(X'X)^{-1}$ exists. Show $Xa = 0 \Rightarrow a = 0$. Let $Xa = 0$

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

Estimation by LEAST SQUARES

Estimate $\beta_0, \beta_1, \ldots \beta_k$ by picking values that get the observed Y_i values as close as possible to their expected values.

$$\text{Minimize } \sum_{i=1}^{n} (Y_i - E(Y_i))^2 = Q$$

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

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

Choose a plane that minimizes the sum of squared vertical distances from the points to the plane.

5.7

Try differentiating, setting $k+1$ derivatives to zero. Cleaner in matrix notation but...

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

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

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

$$\frac{\partial Q}{\partial \beta_2} = \sum_{i=1}^{n} 2 (y_i - \beta_0 - \cdots - \beta_k x_{ik}) x_{i2} \stackrel{set}{=} 0$$

$\vdots$

$$\frac{\partial Q}{\partial \beta_j} = \sum 2 (\qquad\qquad) x_{ij} \stackrel{set}{=} 0$$

$\vdots$

$$\frac{\partial Q}{\partial \beta_k} = \sum 2 (\qquad\qquad) x_{ik} = 0$$

Divide by 2, multiply through, rearrange to get the NORMAL EQUATIONS

Normal Equations

$$\sum y_i = n\beta_0 + \beta_1 \sum x_{i1} + \cdots + \beta_k \sum x_{ik}$$

$$\sum x_{i1} y_i = \beta_0 \sum x_{i1} + \beta_1 \sum x_{i1}^2 + \beta_2 \sum x_{i1} x_{i2} + \cdots + \beta_k \sum x_{i1} x_{ik}$$

$$\sum x_{i2} y_i = \beta_0 \sum x_{i2} + \beta_1 \sum x_{i1} x_{i2} + \beta_2 \sum x_{i2}^2 + \cdots + \beta_k \sum x_{i2} x_{ik}$$

$$\vdots$$

$$\underbrace{\sum x_{ik} y_i}_{X'Y} = \underbrace{\beta_0 \sum x_{ik} + \beta_1 \sum x_{i1} x_{ik} + \beta_2 \sum x_{i2} x_{ik} + \cdots + \beta_k \sum x_{ik}^2}_{(X'X)\beta}$$

So in matrix form, the normal equations are

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

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

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

& put a hat on it

Is it really a minimum?

Least Squares

Minimize $Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots \beta_k x_{ik})^2$ over $\underset{\sim}{\beta}$

$$= \sum_{i=1}^{n} (y_i - \underset{\sim}{x}_i' \beta)^2 = (Y - X\beta)'(Y - X\beta)$$

$$= (\underbrace{Y - X\hat{\beta}}_{A} + \underbrace{X\hat{\beta} - X\beta}_{B})'(\underbrace{Y - X\hat{\beta}}_{A} + \underbrace{X\hat{\beta} - X\beta}_{B}) = A'A + A'B + B'A + B'B$$

Look at $A'B = (Y' - \hat{\beta}'X')(X\hat{\beta} - X\beta)$

$$= Y'X\hat{\beta} - Y'X\beta - \hat{\beta}'X'X\hat{\beta} + \hat{\beta}'X'X\beta$$

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

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

$$= Y'X(X'X)^{-1}X'Y - Y'X\beta - Y'X(X'X)^{-1}X'Y + Y'X\beta = 0$$

So $B'A = 0' = 0$, and

$$Q = (Y - X\hat{\beta})'(Y - X\hat{\beta}) + [X(\hat{\beta} - \beta)]'X(\hat{\beta} - \beta)$$

$$= \underbrace{(Y - X\hat{\beta})'(Y - X\hat{\beta})}_{SSE} + (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)$$

Since $X'X$ is positive definite (see reverse side), the second term is non-negative, and zero only when $\hat{\beta} - \beta = 0 \iff \beta = \hat{\beta}$. This minimizes the function Q over all β.

$$a'X'Xa = \underset{1\times n}{(Xa)'}\underset{n\times 1}{(Xa)} \qquad SS, \geq 0$$

Suppose $= 0$.

$$\Longleftrightarrow \quad Xa = 0 \Longrightarrow X'Xa = X'0 = 0$$

$$\Longrightarrow (X'X)^{-1}XX'a = (X'X)^{-1}0 = 0$$

$$\Longrightarrow a = 0 \qquad \text{So } X'X \text{ is P.D.}$$

Now Least squares with Ch Unit 6

Properties of the Least-Squares Estimators

We can do it, but is it any good?

$\hat{\beta} \neq \beta$ obviously

$\hat{\beta}$ is random, has a probability distribution

$\hat{\beta}_j$

β_j

Unbiased means $E(\hat{\theta}) = \theta \quad \forall\, \theta \in \Theta$.

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

so $E(\hat{\beta}_j) = \beta_j$ for $j = 0, \ldots, k$

How about the variance? Smaller variance means more accurate estimation.

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

$\hat{\beta}_j$ is a linear combination of the Y values

Row j of $A = (X'X)^{-1}X'$ times Y

Gauss-Markov theorem says $\hat{\beta}$ is

BLUE

Gauss - Markov Thm

7.2

It's in the book in Ch , but
Proof is not exactly the same. I go
directly to Cor
on P.

IDEA

- $\hat{\beta} = (X'X)^{-1}X'Y$ is a random vector used to estimate β
- It has a sampling distribution.
- It's unbiased: $E(\hat{\beta}) = \beta$
- There are ∞ many other possible ways to estimate β, even if we insist on unbiased estimation
- Smaller variance of the sampling distribution means more precise estimation, and is better
- $\hat{\beta}$ is B.L.U.E.

Best Linear Unbiased Estimator

7.3

Model: $Y = X\beta + \varepsilon$

X fixed, columns linearly independent
β fixed & unknown

$E(\varepsilon) = 0$, $\quad Cov(\varepsilon) = \sigma^2 I$

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

Want to
ESTIMATE $a'\beta$, a linear combination of the true β values

For ex

$$\underset{a'}{(0\,0\,1\,0\,0)} \underset{\beta}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}} = \beta_2$$

etc

~~Natural candidate is~~ 7.4

Natural choice for estimating $a'\beta$ is $a'\hat{\beta}$: $E(a'\hat{\beta}) = a'E(\hat{\beta}) = a'\beta$ unbiased.

It's also a Linear Combination of the Y values

$$\underset{1\times(k+1)}{a'}\ \underset{(k+1)\times 1}{\hat{\beta}} = \underbrace{\underset{1\times k+1}{a'}\ \underset{(k+1)\times(k+1)}{(X'X)^{-1}}\ \underset{(k+1)\times n}{X'}}_{1\times n}\ \underset{n\times 1}{Y}$$

A LINEAR ESTIMATOR of the form $c'Y$

The Gauss-Markov theorem says that $a'\hat{\beta}$ has the smallest variance of ANY estimator of the form $c'Y$ with $E(c'Y) = a'\beta$ (Linear unbiased estimator)

Preparation

7.5

If $u'x = v'x \quad \forall\, x \in \mathbb{R}^4$, then $u = v$ LEAVE UP

$$(u_1\ u_2\ u_3\ u_4)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = (v_1\ v_2\ v_3\ v_4)\begin{pmatrix} \\ \\ \end{pmatrix}$$

zeros & ones

Proof of G-M

Let $c'Y$ be an unbiased estimator of $a'\beta$, so

$$E(c'Y) = a'\beta = c'E(Y) = c'X\beta$$

So $a'\beta = c'X\beta$ for All Beta, hence

$$a' = c'X \iff a = X'c$$ Leave up

Now $Var(c'Y) = c'\,cov(Y)\,c$

$$= c'\sigma^2 I_n c = \sigma^2 c'c$$

So to minimize $Var(c'Y)$, choose c to minimize $c'c$ subject to

$$a = X'c$$

7.6

What "should" c be?

$$c'Y = a'\hat{\beta} = \underline{a'(X'X)^{-1}X'}\,Y$$

That's our c', so

$$c = X(X'X)^{-1}a$$

$$c'c = \left(\left(c - X(X'X)^{-1}a + X(X'X)^{-1}a\right)' \left(\underbrace{c - X(X'X)^{-1}a}_{A} + \underbrace{X(X'X)^{-1}a}_{B}\right) \right)$$

$$= A'A + A'B + B'A + B'B$$

Look at $A'B$, substituting $a = X'c$

$$\left(c - X(X'X)^{-1}a\right)' X(X'X)^{-1}a$$

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

$$= c'X(X'X)^{-1}X'c - c'X\cancel{(X'X)^{-1}X'X}(X'X)^{-1}X'c$$

$$= 0$$

Old

Gauss - Markov

Let $E(Y) = X\beta$ & $\operatorname{cov}(Y) = \sigma^2 I_n$, and $\hat\beta = (X'X)^{-1}X'Y$

Then $a'\hat\beta$ has the smallest variance of any linear unbiased estimator of the form $c'Y$ with $E(c'Y) = a'\beta$

In particular if $a_j = 1$ and $a_i = 0$ for all $i \neq j$, $\operatorname{Var}(\hat\beta_j)$ is less than the variance of any other linear unbiased estimator of β_j.

<u>Proof</u> $E(c'Y) = c'E(Y) = \underset{1\times n}{c'}\,\underset{n\times p}{X}\,\underset{(p+1)\times 1}{\beta} = a'\beta \quad \forall \beta$ one ele

$\Rightarrow c'X = a' \iff a = X'c$

$\operatorname{Var}(c'Y) = c'\operatorname{cov}(Y)\,c = c'\sigma^2 I_n c = \sigma^2 c'c$

Seek $c \in \mathbb{R}^n$ that makes this smallest subject to $X'c = a$.

What "should" c be? $a'\hat\beta = (X'X)^{-1}X'Y$ should be $c'Y$

So try $c' = a'(X'X)^{-1}X' \iff c = X(X'X)^{-1}a$. Add & subtract

$$c'c = [\underbrace{c - X(X'X)^{-1}a}_{A} + \underbrace{X(X'X)^{-1}a}_{B}]'[\underbrace{c - X(X'X)^{-1}a}_{A} + \underbrace{X(X'X)^{-1}a}_{B}]$$

$$= A'A + A'B + B'A + B'B$$

Look at $A'B$, substituting $a = X'c$

$$A'B = [c - X(X'X)^{-1}X'c]' X(X'X)^{-1}X'c$$

$$= c'X(X'X)^{-1}X'c - c'X(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}X'c = 0$$

So $B'A = 0' = 0$ also

In $B'B$, don't substitute

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

Free of c

So $c'c = A'A$ plus a positive term free of c

$$A'A = \underset{1 \times n}{[c - X(X'X)^{-1}a]'}\ \underset{n \times 1}{[c - X(X'X)^{-1}a]}$$

It's a sum of squares, ≥ 0, & $= 0$ only if all terms are zero.

$$c - X(X'X)^{-1}a = 0 \iff c = X(X'X)^{-1}a$$

Thus $a'\hat{\beta}$ is BLUE of $a'\beta$

$\blacksquare$