Math 455 Class 13 Feb. 19

Q: Wackerly 11.89 (c): is the answer in the back of the textbook right?

A: No.

Q: But I got a copy of the Instructors' Solution Manual from an E-Book site and it has the same answer!

A: I think the ISM is wrong too.

Q: What if you are wrong?

A: I've been wrong many times!

I will make my best judgement at the time I am grading.

Midterm 1: Proposed date: Monday March 3.

Sample Midterm is posted – homework!

Please try to do this midterm in 80 minutes without book/notes.

After you try this, then you can take all the

time you want to produce correct solutions and turn them in.

Recall the Gauss-Markov Theorem: (Ch. 2)

the OLS estimators are BLUE.

Note L means linear in the Y_s not linear in the X_s.

One way to see that the OLS estimators are linear in the formula (Faraway p. 33)

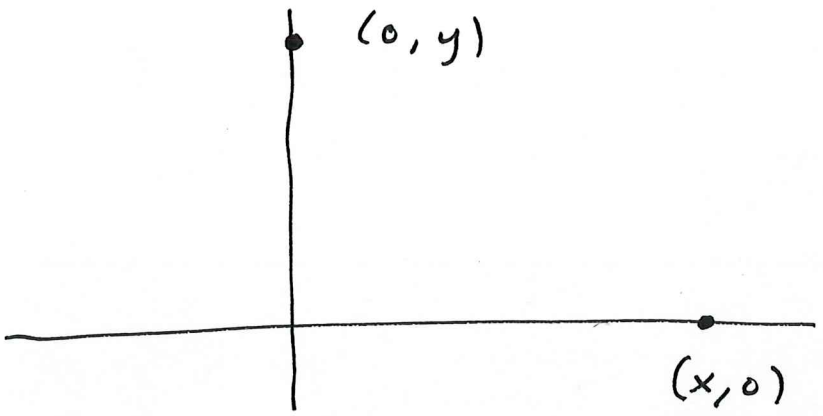
$$\hat{\beta} = \underbrace{(X^TX)^{-1}X^T}_{\text{matrix}} \underbrace{Y}_{\text{vector}}.$$

matrix mult. is linear:

$$A(u+v) = Au + Av.$$

Q: I looked at the proof of Gauss-Markov in Ch. 2 and it is incomprehensible.

A: Some remarks on the ideas of that proof. First a very simple case of a proof based on the same idea.



What point on the x-axis is closest to $(0,y)$? (With proof.)

Answer: the origin $(0,0)$.

Why? the distance from $(0,y)$ to an arbitrary point $(x,0)$ on the x-axis is

$$d = \sqrt{(0-x)^2 + (y-0)^2} = \sqrt{x^2+y^2}$$

Minimizing d is the same as minimizing d^2 (You'll get the same point.

$$d^2 = y^2 + x^2 = d^2((0,y),(0,0)) + x^2$$

$$\geq d^2((0,y),(0,0))$$

since $x^2 \geq 0$.

So the origin is the closest point.

This is the basic idea in the proof of Gauss-Markov: We want to show that the OLS estimators are "best". (among unbiased linear estimators)

"best" means <u>minimum variance</u>

Remember: variance is a lot like distance squared.

Analogy between probability & linear algebra:

random variables	$\longleftrightarrow$	vectors
covariance	$\longleftrightarrow$	inner product.
correlation	$\longleftrightarrow$	cosine of angle.

(Analogy not perfect, need to subtract of the mean)

Idea of proof: We have OLS estimator $= AY$ (for $A = (X^TX)^{-1}X^T$)

alternate estimator $= (A+D)Y$.

Use the unbiasedness of the alternate estimator to write

Variance of Alt. Estimator

= Variance of OLS Est

+ Other Variance

$\geq$ Variance of OLS Est.

In order to do this need a combination of probability and linear algebra!

We can study <u>vectors</u> of RVs and much of math 447 is still true, with appropriate modifications.

Example: if X is a vector of RVs.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \qquad E[X] \overset{\text{def}}{=} \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{bmatrix}$$

Or a matrix of RVs.

If X is a matrix (or vector) of RVs,

we defined $E[X] = (E[X_{i,j}])$

$E[X]$: matrix of RVs; $(E[X_{i,j}])$: matrix of numbers

Expectation is linear: if A, B, C are matrices of real numbers and sizes are compatible, then:

$$E[AXB + C] = A\,E[X]\,B + C$$

and if X, Y are vectors of RVs and A, B matrices of numbers

$$E[\underbrace{AX + BY}_{\text{vector of RVs}}] = \underbrace{A\,E[X] + B\,E[Y]}_{\text{vector of numbers}}$$

If X, Y are vectors of RVs define

$$\text{Cov}(X, Y) = (\text{cov}(X_i, Y_j))$$

$\text{Cov}(X, Y)$: vectors of RVs; $(\text{cov}(X_i, Y_j))$: matrix of numbers.

variance-covariance matrix.

$$\mathrm{Cov}(X,Y) = E\left[(X-\alpha)(Y-\beta)^T\right]$$

where $\alpha = E[X]$ and $\beta = E[Y]$

and Cov is still bilinear:

$$\mathrm{Cov}(AX, BY) = A\,\mathrm{Cov}(X,Y)\,B^T$$

and $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.

(X: vector of RVs; $\mathrm{Var}(X)$: variance-cov. matrix of X.)

Many results have matrix analogs:

Recall that if W is a RV and $a \in \mathbb{R}$, then $V[aW] = a^2 V[W]$

If A is a matrix of numbers and X is a vector of RVs:

$$V[AX] = A\,V[X]\,A^T.$$

Also: $\mathrm{Cov}(X+Y, Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$

$\mathrm{Var}(X+Y) = \mathrm{Cov}(X+Y, X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$

We study a simple linear regression

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

You are given $E[Y] = \frac{1}{n}(Y_1 + \cdots + Y_n)$

also $E[X]$, $V[X]$, $V[Y]$, and $Cov[X, Y]$.

How much of the regression output can we recover? Can we find $\hat{\beta}_0$, $\hat{\beta}_1$, the OLS estimators?

A: We learned that $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$

$$= \frac{Cov(X, Y)}{Var(X)} \quad \begin{array}{l} \longleftarrow \text{ given} \\ \longleftarrow \text{ given} \end{array}$$

So we can get $\hat{\beta}_1$.

What about $\hat{\beta}_0$? Remember that $(\bar{X}, \bar{Y})$ is on the regression line.

Thus $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$ and we know $\bar{X}, \bar{Y}$ (given) and $\hat{\beta}_1$ (above) so we know $\hat{\beta}_0$.

More linear algebra: if $X \in M_{n\times p}(\mathbb{R})$

$X: \mathbb{R}^p \longrightarrow \mathbb{R}^n$ $\qquad$ $X^T: \mathbb{R}^n \longrightarrow \mathbb{R}^p$

$$\mathbb{R}^p \xrightarrow{X} \mathbb{R}^n \xrightarrow{X^T} \mathbb{R}^p$$

So $\quad X^TX: \mathbb{R}^p \longrightarrow \mathbb{R}^p$

$$\underbrace{\text{row}(X) \oplus \text{null}(X)}_{\uparrow \text{ orthogonal direct sum}} = \mathbb{R}^p \xrightarrow{X} \mathbb{R}^n = \underbrace{\text{col}(X)}_{= \text{range}(X)} \oplus \text{null}(X^T)$$

(Note: requires real numbers)

From this, it follows that $\text{range}(X^TX) = \text{range}(X^T)$

Remember: $\text{range}(X^T) = \text{col}(X^T) = \text{row}(X)$.

Need to show if $z \in \mathbb{R}^p = X^Ty$, then $\exists\, v \in \mathbb{R}^p$ such that $X^TXv = z$

$$\begin{array}{ccccc} \mathbb{R}^p & \xrightarrow{X} & \mathbb{R}^n & \xrightarrow{X^T} & \mathbb{R}^p \\ \psi & & \psi & & \psi \\ v & & y & & z \end{array}$$

Q: What do we actually have to know?

A: The statement of Gauss-Markov and what it means.

Note that the standard assumptions of linear regression.

Linear relationship (approximate)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Where ε_i are independent identically distributed RVs.

With $\varepsilon_i \sim N(0, \sigma^2)$.

Gauss-Markov only requires

that the ε_i be uncorrelated

not independent.

It does NOT require that they be normal.

Example of a conceptual question that wasn't on the midterm but could have been.