STAT 350: Lecture 9

Reading: Chapter 6.1–6.

Distribution Theory for Least Squares

Assume:

$$\mathrm{E}(\epsilon_i) = 0 \qquad \text{and} \qquad Y = X\beta + \epsilon$$

Then

1. $\mathrm{E}(\hat\beta) = \beta$
2. $\mathrm{E}(\hat\mu) = X\mathrm{E}(\hat\beta) = X\beta = \mu$
3. $\hat\epsilon = (I - X(X^TX)^{-1}X^T)\epsilon \equiv M\epsilon$ where $M = I - X(X^TX)^{-1}X^T$.
4. $\mathrm{E}(\hat\epsilon) = M\mathrm{E}(\epsilon) = 0$

Define: $H = X(X^TX)^{-1}X^T$, the **hat** matrix so that $M = I - H$.
If also

$$\mathrm{Var}(\epsilon) = \sigma^2 I$$

(as will be the case for instance if the ϵ_i are iid) then

1. $\mathrm{Var}(\hat\beta) = \sigma^2(X^TX)^{-1}$
2. $\mathrm{Var}(\hat\mu) = \sigma^2 X(X^TX)^{-1}X^T = \sigma^2 H$
3. $\mathrm{Var}(\hat\epsilon) = M\mathrm{Var}(\epsilon)M^T = \sigma^2 MM^T$

If also $\epsilon_i \sim N(0, \sigma^2)$ are independent then

$$\hat\beta \sim MVN(\beta, \sigma^2(X^TX)^{-1}) \qquad \text{and} \qquad \hat\mu \sim MVN(\mu, \sigma^2 H)$$

Notice $\hat\mu = HY$ and that H is $n \times n$.
Some algebraic simplification of the variances above is possible.

$$\begin{aligned}
\mathrm{Var}(\hat\epsilon) &= \sigma^2 MM^T \\
&= \sigma^2(I-H)(I-H)^T \\
&= \sigma^2(II^T - HI^T - IH^T + HH^T)
\end{aligned}$$

But

$$I = I^T$$

so $II^T = I$ and

$$\begin{aligned} H^T &= \left[X(X^TX)^{-1}X^T\right]^T \\ &= X\left[(X^TX)^{-1}\right]^T X^T \\ &= X\left[(X^TX)^T\right]^{-1} X^T \\ &= X\left[X^TX\right]^{-1} X^T \\ &= H \end{aligned}$$

ASIDE: as you read that sequence of formulas you will see that I expect you to remember a number of algebraic facts about matrices:

1. $(A^T)^{-1} = (A^{-1})^T$
2. $(AB)^T = B^TA^T$
3. $(A^T)^T = A$

So

$$M^T = M \qquad \text{and} \qquad H^T = H$$

Finally

$$HH^T = HH = H^2 = X(X^TX)^{-1}\underbrace{X^TX(X^TX)^{-1}}_{\text{Identity}}X^T = X(X^TX)^{-1}X^T = H$$

Thus

$$\begin{aligned} H^2 &= H \\ M^2 &= I - H - H + H^2 \\ &= I - H \\ &= M \end{aligned}$$

Definition: A matrix Q is idempotent if

$$QQ \equiv Q^2 = Q$$

So What?

1. Distribution theory of Sums of Squares in ANOVA tables uses this.
2. F tests for hypotheses about parameters are justified using these matrix ideas.
3. t-tests, and confidence intervals for $c^T\beta$ (where c is a vector of length p can be derived using these ideas.

Example: Estimation of σ

Estimation of σ^2 is based on the error sum of squares defined by

$$\begin{aligned}
\text{ESS} &= ||\hat{\epsilon}||^2 \\
&= \sum \hat{\epsilon}_i^2 \\
&= (M\epsilon)^T(M\epsilon) \\
&= \epsilon^T M^T M\epsilon \\
&= \epsilon^T M\epsilon
\end{aligned}$$

Now note that

$$\begin{aligned}
\text{E[ESS]} &= \text{E}(\epsilon^T M\epsilon) \\
&= \text{E}[\sum_{ij} M_{ij}\epsilon_i\epsilon_j] \\
&= \sum_{ij} M_{ij}\text{E}[\epsilon_i\epsilon_j]
\end{aligned}$$

But $\text{E}[\epsilon_i\epsilon_j] = 0$ for $i \neq j$ and $\text{E}[\epsilon_i^2] = \sigma^2$ so

$$\begin{aligned}
\text{E[ESS]} &= \sigma^2 \sum_i M_{ii} \\
&= \sigma^2 \text{trace}(M)
\end{aligned}$$

Definition: The trace of a square matrix Q is defined by

$$\text{trace}(Q) = \sum_i Q_{ii}$$

Marvelous Matrix Identity (cyclic invariance of the trace). Suppose that $A_{m\times n}$ and $B_{n\times m}$ so that $(AB)_{m\times m}$ and $(BA)_{n\times n}$ are both square. Then

$$\begin{aligned}
\text{trace}(AB) &= \sum_i (AB)_{ii} \\
&= \sum_i (\sum_j A_{ij}B_{ji}) \\
&= \sum_{ij} A_{ij}B_{ji} \\
&= \sum_j (\sum_i B_{ji}A_{ij}) \\
&= \sum_j (BA)_{jj} \\
&= \text{trace}(BA)
\end{aligned}$$

The same idea works with more than two matrices provided the product is square so, e.g.,

$$\text{trace}(ABCD) = \text{trace}(DABC) = \text{trace}(CDAB) = \text{trace}(BCDA)$$

Another algebraic identity for the trace:

$$\begin{aligned}
\text{trace}(A+B) &= \sum (A+B)_{ii} \\
&= \sum (A_{ii} + B_{ii}) \\
&= \text{trace}(A) + \text{trace}(B)
\end{aligned}$$

SO:

$$\begin{aligned}
\text{trace}(M) &= \text{trace}(I - H) \\
&= \text{trace}(I) - \text{trace}(H) \\
&= n - \text{trace}(\underbrace{X}_{A}\underbrace{(X^TX)^{-1}X^T}_{B}) \\
&= n - \text{trace}(\underbrace{(X^TX)^{-1}X^TX}_{I_{p\times p}}) \\
&= n - \text{trace}(I_{p\times p}) \\
&= n - p
\end{aligned}$$

Notice that p is the number of columns of X **including** the column of 1's if present.

Summary of result

$$\mathrm{E}\left[\frac{\mathrm{ESS}}{n-p}\right] = \sigma^2$$

So the Mean Squared Error, $ESS/(n-p)$ is an unbiased estimate of σ^2.

Extras

Here is an example of some of the matrix algebra I was doing in class. Consider the weighing design where two objects of weights β_1 and β_2 are weighed individually and together. The resulting design matrix is

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

So

$$X^t X = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

and

$$(X^T X)^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

The hat matrix is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

which is

$$\begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Notice that the trace of H is 2 which is the number of parameters in β.

You can also check that H is idempotent.