SYSTEMS 302
LECTURE 9

- **LEAST-SQUARES ESTIMATION**
  - **Unbiasedness of Beta Estimates**
  - **Efficiency of Beta Estimates**

- **VIOLATIONS OF REGRESSION MODEL ASSUMPTIONS**
  - **Nonlinear Model Specifications**
  - **Coefficient of Determination**

- **For next time:**
  - **Devore, Section 12.1-12.2,13.2**

METHOD OF LEAST SQUARES

Recall that for any random variable, Y, with mean μ it must be true that for any other value, a,

$$E\left[(Y-\mu)^2\right] < E\left[(Y-a)^2\right]$$

So if $E(Y\,|\,x) = \beta_0 + \beta_1 x$ for all x, then for any other linear relation, $a_0 + a_1 x$, it must be true that

$$E\left[\left(Y_i - (\beta_0 + \beta_1 x_i)\right)^2\right] \le E\left[\left(Y_i - (a_0 + a_1 x_i)\right)^2\right],\ i = 1,..,n$$

$$\Rightarrow \sum_{i=1}^{n} E\left[\left(Y_i - (\beta_0 + \beta_1 x_i)\right)^2\right] \le \sum_{i=1}^{n} E\left[\left(Y_i - (a_0 + a_1 x_i)\right)^2\right]$$

$$\Rightarrow E\left[\sum_{i=1}^{n}\left(Y_i - (\beta_0 + \beta_1 x_i)\right)^2\right] \le E\left[\sum_{i=1}^{n}\left(Y_i - (a_0 + a_1 x_i)\right)^2\right]$$

Hence it is natural to estimate (β_0, β_1) by minimizing

$$\boxed{S(\beta_0, \beta_1) = \sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2}$$

LEAST SQUARES ESTIMATES

Minimizing the objective function:

$$S(\beta_0,\beta_1) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

yields the following two first-order conditions:

(1) $$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

(2) $$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)x_i = 0$$

NORMAL EQUATIONS

Dividing (1) by n **simplifies this relation to**

(3) $$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$$

where $\bar{x} = \frac{1}{n}\Sigma_i x_i$ **and** $\bar{y} = \frac{1}{n}\Sigma_i y_i$**. Substitution of (3) in (2) then yields the following solution for** $\hat{\beta}_1$**:**

(4) $$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})x_i}{\sum_{i=1}^{n}(x_i - \bar{x})x_i}}$$

Finally, substitution of (4) into (3) yields $\hat{\beta}_0$**.**

LINEARITY OF BETA ESTIMATES

We focus on β_1. To see that $\hat{\beta}_1$ is a linear estimator of β_1, observe first that since $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$,

(1)

$$\hat{\beta}_1 = \frac{\sum_i (y_i - \overline{y})(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2} = \frac{\sum_i (x_i - \overline{x}) y_i - \overline{y} \sum_i (x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}$$

$$= \frac{\sum_i (x_i - \overline{x}) y_i}{\sum_i (x_i - \overline{x})^2} = \sum_i \left(\frac{x_i - \overline{x}}{\sum_j (x_j - \overline{x})^2} \right) y_i$$

Hence by letting

(2)
$$w_i = \frac{x_i - \overline{x}}{\sum_j (x_j - \overline{x})^2}, \quad i = 1, .., n$$

and replacing the data values, y_i, by the corresponding random variables, Y_i, we see that

$$\boxed{\hat{\beta}_1 = \sum_{i=1}^{n} w_i \cdot Y_i}$$

UNBIASEDNESS OF BETA ESTIMATES

To establish unbiasedness of $\hat{\beta}_1$, observe first that the weights, w_i, by definition satisfy

(1)
$$\sum_i w_i = \frac{\sum_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = 0$$

(2)
$$\sum_i w_i (x_i - \bar{x}) = \frac{\sum_i (x_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} = 1$$

so that

(3)
$$1 = \sum_i w_i (x_i - \bar{x}) = \sum_i w_i x_i - \bar{x} \sum_i w_i = \sum_i w_i x_i$$

Hence the expected value of $\hat{\beta}_1$ is given by

$$\begin{aligned} E\left(\hat{\beta}_1\right) &= \sum_i w_i E(Y_i) = \sum_i w_i E(\beta_0 + \beta_1 x_i + \varepsilon_i) \\ &= \sum_i w_i (\beta_0 + \beta_1 x_i) = \beta_0 \sum_i w_i + \beta_1 \sum_i w_i x_i \\ &= (0) + \beta_1 \cdot 1 \end{aligned}$$

$$\Rightarrow \quad \boxed{E(\hat{\beta}_1) = \beta_1}$$

VARIANCE OF BETA ESTIMATES

The linearity of the beta estimates allows us to compute their variance in closed form. For $\hat{\beta}_1$ we see that

(1) $$\operatorname{var}\left(\hat{\beta}_1\right) = \sum_i w_i^2 \operatorname{var}\left(Y_i\right)$$

But since $\beta_0 + \beta_1 x_i$ is nonrandom

(2) $$\operatorname{var}\left(Y_i\right) = \operatorname{var}\left(\beta_0 + \beta_1 x_i + \varepsilon_i\right) = \operatorname{var}(\varepsilon_i) = \sigma^2$$

Hence the variance of $\hat{\beta}_1$ is computable as

(3) $$\begin{aligned} \operatorname{var}\left(\hat{\beta}_1\right) &= \sum_i w_i^2 \operatorname{var}\left(\varepsilon_i\right) = \sigma^2 \sum_i w_i^2 \\ &= \sigma^2 \sum_i \frac{\left(x_i - \overline{x}\right)^2}{\left[\sum_j \left(x_j - \overline{x}\right)^2\right]^2} = \sigma^2 \frac{\sum_i \left(x_i - \overline{x}\right)^2}{\left[\sum_j \left(x_j - \overline{x}\right)^2\right]^2} \end{aligned}$$

which yields

$$\boxed{\operatorname{var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_i \left(x_i - \overline{x}\right)^2}}$$

EFFICIENCY OF BETA ESTIMATES

The key optimality property of least-squares estimators is given by the following fundamental result:

GAUSS-MARKOV THEOREM. For any linear function $L = a_0\beta_0 + a_1\beta_1$ of (β_0, β_1), the least-squares estimator $\hat{L} = a_0\hat{\beta}_0 + a_1\hat{\beta}_1$ has minimum variance among all possible linear unbiased estimators of L.

$\Rightarrow$ $\left(\hat{\beta}_0, \hat{\beta}_1\right)$ **are Best Linear Unbiased (BLU) estimators of the regression parameters** $\left(\beta_0, \beta_1\right)$

$\Rightarrow$ $\hat{y}(x) \equiv \hat{\beta}_0 + \hat{\beta}_1 x$ **is a Best Linear Unbiased (BLU) estimator of the conditional mean** $\mu_{Y|x} \equiv E(Y \mid x) = \beta_0 + \beta_1 x$

NOTE: This also generalizes our earlier result that $\overline{Y}_n$ is a BLU estimator of the mean $E(Y)$ since:

$$(Y_i = \beta_0 + \varepsilon_i) \Rightarrow [\, E(Y_i) = \beta_0 \,] \; and \; [\, \hat{\beta}_0 = \overline{Y}_n \,]$$