Lecture 10

Monday, June 10, 2019 2:50 PM

Recall:

- $\vec{\varepsilon} \sim MVN(\vec{0}, \sigma^2 I)$
- $\vec{y} \sim MVN(\vec{\mu} = X\vec{\beta}, \sigma^2 I)$
- $\hat{\vec{\beta}} \sim MVN(\vec{\beta}, \sigma^2 (X^TX)^{-1})$

<u>Related Quantities</u>: $\underset{\text{fitted values}}{\hat{\vec{\mu}}} = X\hat{\vec{\beta}}$ and $\underset{\text{residuals}}{\vec{e}} = \vec{y} - \hat{\vec{\mu}} = \vec{y} - X\hat{\vec{\beta}}$

We want to derive the distributions of $\hat{\vec{\mu}}$ and $\vec{e}$. It will become useful to define the "<u>hat</u>" matrix $H = X(X^TX)^{-1}X^T$.

$$\hat{\vec{\mu}} = X\hat{\vec{\beta}} = X(X^TX)^{-1}X^T\vec{y} = H\vec{y}$$

$$\vec{e} = \vec{y} - X\hat{\vec{\beta}} = \vec{y} - H\vec{y} = (I-H)\vec{y}$$

Notice that $\hat{\vec{\mu}}$ and $\vec{e}$ can both be written as a matrix multiple of the response vector $\vec{y}$, which MVN distribution. Therefore, we know $\hat{\vec{\mu}}$ and $\vec{e}$ both follow MVN distributions also.

Let's derive the mean vector and variance-covariance matrices for both of these distributions. It will be useful to recognize that both H and $I-H$ are <u>symmetric</u> $(A^T = A)$ and <u>idempotent</u> $(AA = A)$.

- $E[\hat{\vec{\mu}}] = E[H\vec{y}] = HE[\vec{y}] = X\underline{(X^TX)^{-1}}\,\underline{X^TX}\vec{\beta} = X\vec{\beta} = \vec{\mu}$

- $$\begin{aligned} Var[\hat{\vec{\mu}}] = Var[H\vec{y}] = H\,Var[\vec{y}]\,H^T &= H\sigma^2 I H^T \\ &= \sigma^2 HH^T \\ &= \sigma^2 H \end{aligned}$$

$$\therefore \hat{\vec{\mu}} \sim MVN(\vec{\mu}, \sigma^2 H)$$

- $$\begin{aligned} E[\vec{e}] = E[(I-H)\vec{y}] = (I-H)E[\vec{y}] &= (I-H)X\vec{\beta} \\ &= X\vec{\beta} - \underline{HX\vec{\beta}} \\ &= \vec{\mu} - \vec{\mu} \\ &= \vec{0} \end{aligned}$$

$\hookrightarrow X(X^TX)^{-1}X^TX\vec{\beta} = X\vec{\beta} = \vec{\mu}$

- $$\begin{aligned} Var[\vec{e}] = Var[(I-H)\vec{y}] &= (I-H)\,Var[\vec{y}]\,(I-H)^T \\ &= \sigma^2(I-H) \end{aligned}$$

$$\therefore \vec{e} \sim MVN(\vec{0}, \sigma^2(I-H))$$

* one consequence of this result for the residuals is that $\dfrac{\hat{\sigma}^2(n-p-1)}{\sigma^2} \sim \chi^2_{(n-p-1)}$.

<u>Confidence intervals for μ_0 and Prediction intervals for y_0</u>

Given a set of values for the explanatory variables, we wish to estimate μ and predict y. We typically want to accompany these point estimates and point predictions with the appropriate interval estimate.

We organize the relevant explanatory variable values as follows:

$$\underset{1\times(p+1)}{\vec{x}_0} = [1 \quad x_{01} \quad x_{02} \quad x_{03} \quad \cdots \quad x_{0p}]$$

where x_{0j} is the particular value of x_j that we care about.

The corresponding estimate of μ_0 (equivalently, the prediction y_0) is:

$$\hat{\mu}_0 = \hat{y}_0 = \vec{x}_0\hat{\vec{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \cdots + \hat{\beta}_p x_{0p}$$

This estimator $\hat{\mu}_0$ has the following normal distribution

$$\hat{\mu}_0 \sim N(\underset{\mu_0}{\vec{x}_0\vec{\beta}}, \sigma^2 \vec{x}_0(X^TX)^{-1}\vec{x}_0^T)$$

$$\Rightarrow \frac{\hat{\mu}_0 - \mu_0}{\sigma\sqrt{\vec{x}_0(X^TX)^{-1}\vec{x}_0^T}} \sim N(0,1)$$

$$\Rightarrow \frac{\hat{\mu}_0 - \mu_0}{\hat{\sigma}\sqrt{\vec{x}_0(X^TX)^{-1}\vec{x}_0^T}} \sim t_{(n-p-1)}$$

And so a $(1-\alpha)\times 100\%$ CI for μ_0 is:

$$\underbrace{\hat{\mu}_0}_{\text{est.}} \pm \underbrace{t_{(n-p-1)}(1-\alpha/2)}_{\text{critical value}} \underbrace{\hat{\sigma}\sqrt{\vec{x}_0(X^TX)^{-1}\vec{x}_0^T}}_{\text{st. error}}$$

When we're interested in predicting y at $\vec{x}_0$ our point prediction is $\hat{y}_0 = \hat{\mu}_0 = \vec{x}_0\hat{\vec{\beta}} \in \mathbb{R}$. To derive the PI for y_0 we must consider the prediction error:

$$y_0 - \hat{y}_0 = (\mu_0 + \varepsilon_0) - \hat{\mu}_0$$

$$= (\mu_0 - \hat{\mu}_0) + \varepsilon_0$$

Since $\hat{y}_0 = \hat{\mu}_0 \sim N(\mu_0, \sigma^2\vec{x}_0(X^TX)^{-1}\vec{x}_0^T)$ then

$$y_0 - \hat{y}_0 \sim N(0, \sigma^2(1 + \vec{x}_0(X^TX)^{-1}\vec{x}_0^T))$$

$$\Rightarrow \frac{y_0 - \hat{y}_0}{\sigma\sqrt{1 + \vec{x}_0(X^TX)^{-1}\vec{x}_0^T}} \sim N(0,1)$$

$$\Rightarrow \frac{y_0 - \hat{y}_0}{\hat{\sigma}\sqrt{1 + \vec{x}_0(X^TX)^{-1}\vec{x}_0^T}} \sim t_{(n-p-1)}$$

And so a $(1-\alpha)\times 100\%$ PI for y_0 is

$$\underbrace{\hat{y}_0}_{\text{est}} \pm \underbrace{t_{(n-p-1)}(1-\alpha/2)}_{\text{c.v.}} \underbrace{\hat{\sigma}\sqrt{1 + \vec{x}_0(X^TX)^{-1}\vec{x}_0^T}}_{\text{s.e.}}$$

* Notice that since $\hat{\mu}_0 = \hat{y}_0$ that the CI for μ_0 and PI for y_0 are centered at the same value, but their widths differ. In particular the PI will always be wider than the CI.

<u>Example</u> (Sales)

Lets estimate the expected response and calculate the predicted response when:

- $x_1 = 3$ (\$3000 in promotional expenditures)
- $x_2 = 45$ (45 stores in the district)
- $x_3 = 10$ (10 competitors in the district)
- $x_4 = 10$ (district potential score is 10)

The estimate of μ_0 / prediction of y_0 is given by

$$\hat{\mu}_0 = \hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \hat{\beta}_3 x_{03} + \hat{\beta}_4 x_{04}$$

$$= \vec{x}_0\hat{\vec{\beta}} = [1 \quad 3 \quad 45 \quad 10 \quad 10]\begin{bmatrix} 177.23 \\ 2.17 \\ 3.54 \\ -22.16 \\ 0.20 \end{bmatrix}$$

$$= (1)(177.23) + (3)(2.17) + (45)(3.54) + (10)(-22.16) + (10)(0.20)$$
$$= 123.4018$$

So the estimated expected sales / predicted sales is \$12,340,180

The corresponding CI is (118.7986, 128.0051). So we're 95% confident that the true <u>expected</u> sales is between \$11,879,860 and \$12,800,510.

The corresponding PI is (111.1015, 135.7022). So we're 95% confident that the true sales is between \$11,110,150 and \$13,570,220.