LECTURE 34

Agenda:

(a) Confidence intervals for β_0, β_1
(b) Confidence intervals for future predictions

In the last few lectures, we have derived the form of the least squares estimates in the linear model, and also some of their useful properties. Today, we proceed to derive confidence intervals for β_0, β_1 and "future predicted values". Here is the standard linear model that we have been studying.

Data: $(X_1, Y_1), (X_2, Y_2), ---, (X_n, Y_n)$ (X-values are fixed)

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, ---, n$$

$\varepsilon_1, \varepsilon_2, ---, \varepsilon_n$ are independent and have a common distribution with $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2$

ADDITIONALLY, we will assume that the common distribution for $\varepsilon_1, \varepsilon_2, ---, \varepsilon_n$ is Normal.

In this scenario, we had proved that

$\hat{\beta}_1$ is Normal$\left(\beta_1, \dfrac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)$

$\hat{\beta}_0$ is Normal$\left(\beta_0, \dfrac{\sigma^2 \bar{x}^2}{\sum (x_i - \bar{x})^2} + \dfrac{\sigma^2}{n} \right)$

$\dfrac{(n-2)\hat{\sigma}^2}{\sigma^2}$ is χ^2 with $n-2$ degrees of freedom.

Using these facts, it follows that,

$\dfrac{\hat{\beta}_1 - \beta_1}{\hat{\sigma} / \sqrt{\sum (x_i - \bar{x})^2}}$ has a t-distribution with $n-2$ degrees of freedom

$\dfrac{\hat{\beta}_0 - \beta_0}{\hat{\sigma} \sqrt{\dfrac{\bar{x}^2}{\sum (x_i - \bar{x})^2} + \dfrac{1}{n}}}$ has a t-distribution with $n-2$ degrees of freedom

Hence, both these quantities are useful as pivotal quantities for constructing confidence intervals for β_1 and β_0 respectively.

<u>RESULT:</u> The $(1-\alpha)$- confidence interval for β_1 is given by

$$\left[\hat{\beta}_1 - t_{n-2, 1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \bar{x})^2}}, \; \hat{\beta}_1 + t_{n-2, 1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \bar{x})^2}} \right]$$

Denotes the $1-\frac{\alpha}{2}$ th quantile of the t-distribution with $n-2$ degrees of freedom.

RESULT: The $(1-\alpha)$-confidence interval for β_0 is given by

$$\left[\hat{\beta}_0 - t_{n-2,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{\frac{\bar{x}^2}{\Sigma(x_i-\bar{x})^2}+\frac{1}{n}},\ \hat{\beta}_0 + t_{n-2,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{\frac{\bar{x}^2}{\Sigma(x_i-\bar{x})^2}+\frac{1}{n}}\right].$$

Let us now proceed to the following situation. We have analyzed the linear model using the n pairs of observations, and we wish to predict the average Y-value for a new value of x that comes up in future, say x^*. In terms of the model, we want to estimate

$$E[Y \mid X = x^*] = E[\beta_0 + \beta_1 x^* + \varepsilon] = \beta_0 + \beta_1 x^*.$$

Clearly, an obvious estimator is given by $\hat{\beta}_0 + \hat{\beta}_1 x^*$.

The question is, can we obtain a confidence interval for $\beta_0 + \beta_1 x^*$? The answer is yes. It can be shown that

(i) $\hat{\beta}_0 + \hat{\beta}_1 x^*$ has a $\text{Normal}\left(\beta_0 + \beta_1 x^*,\ \frac{\sigma^2}{n} + \frac{\sigma^2 (x^* - \bar{x})^2}{\Sigma(x_i-\bar{x})^2}\right)$.

(ii) $\dfrac{(\hat{\beta}_0 + \hat{\beta}_1 x^*) - (\beta_0 + \beta_1 x^*)}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}}}$ has a t-distribution with $n-2$ degrees of freedom.

<u>RESULT:</u> The $(1-\alpha)$-confidence interval for $\beta_0+\beta_1 x^*$, which denotes the average Y-value at the x-value x^*, is given by

$$\left[\hat{\beta}_0+\hat{\beta}_1 x^* - t_{n-2,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x^*-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}}\,,\right.$$

$$\left.\hat{\beta}_0+\hat{\beta}_1 x^* + t_{n-2,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x^*-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}}\right].$$