1 Confidence Interval for Mean

In simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim^{iid} N\left(0, \sigma^2\right)$$

its confidence interval for mean $\mu_x = E\left(Y|\, x\right) = \beta_0 + \beta_1 x$ is

$$\hat{\mu_x} \pm t_{\frac{\alpha}{2}}\left(n-2\right) SE\left(\hat{\mu_x}\right),$$

i.e.

$$\left(\hat{\beta}_0 + \hat{\beta}_1 x\right) \pm t_{\frac{\alpha}{2}}\left(n-2\right)\sqrt{\hat{\sigma}^2 \cdot h\left(x\right)}$$

where $\hat{\beta}_0, \hat{\beta}_1$ are least square estimators of β_0 and β_1, and

$$h\left(x\right) = \frac{1}{n} + \frac{\left(x-\bar{x}\right)^2}{\sum\limits_{i=1}^{n}\left(x_i-\bar{x}\right)^2}, \hat{\sigma}^2 = \frac{SSE}{n-2} = MSE.$$

It can be shown that $Var\left(\hat{\beta}_0 + \hat{\beta}_1 x\right) = \sigma^2 \cdot h\left(x\right).$

2 Prediction Interval for Response

Prediction of response Y at a future observation x_0

$$\tilde{Y_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0 + \varepsilon_0, \varepsilon_0 \sim N\left(0, \sigma^2\right).$$

Denote $\hat{Y_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$, then $\tilde{Y_0} = \hat{Y_0} + \varepsilon_0$.

It is known that the future observation $\left(x_0, \tilde{Y_0}\right)$ is independent of the past and current observations $\left\{\left(x_i, Y_i\right), i = 1, ..., n\right\}$, thus $\hat{Y_0}$ and ε_0 are independent of each other. So we have

$$Var\left(\tilde{Y_0}\right) = Var\left(\hat{Y_0}\right) + Var\left(\varepsilon_0\right) = \sigma^2 h\left(x_0\right) + \sigma^2 = \sigma^2\left(h\left(x_0\right)+1\right).$$

Similar to the construction of confidence interval, we obtain the prediction interval for Y at a future observation x_0 as

$$\tilde{Y_0} \pm t_{\frac{\alpha}{2}}\left(n-2\right) SE\left(\tilde{Y_0}\right).$$

Note that $\hat{Y_0}$ is a point estimate of $\tilde{Y_0}$, the prediction interval is given as

$$\left(\hat{\beta}_0 + \hat{\beta}_1 x_0\right) \pm t_{\frac{\alpha}{2}}\left(n-2\right)\sqrt{\hat{\sigma}^2 \cdot \left(h\left(x_0\right)+1\right)}.$$