Confidence And Prediction Intervals Derivations

For Mean Response Y_0:

We know the matrix form of multiple regression :

$$Y = X\beta + \varepsilon$$

With the assumption of :

$$\varepsilon \sim N(0, \sigma^2 I)$$

Goal: Constructing a confidence interval for $E(Y_0)$

$$Y_0 = X_0^T \beta + \varepsilon_0$$

- $Y_0 \to$ New response (Not one of the observed reposes)
- $X_0 \to$ New unseen predictor

$$X_0 = \begin{bmatrix} 1 \\ x_{1,0} \\ x_{2,0} \\ \vdots \\ x_{p,0} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

- That's why we use the transpose of X_0

We also Know:

$$E(Y_0) = E(X_0^T \beta + \varepsilon_0)$$

This is what we constructing the confidence interval for :

$$E(Y_0) = X_0^T \beta$$

- We don't know β
- So we need an interval of where our expected value of unobserved response will fall

Also :

$$\hat{Y}_0 \text{ is the estimated value of } \to Y_0$$

$$\hat{Y}_0 = X_0^T \hat{\beta}$$

And from Ordinary Least Squares and Standard Error Derivation we know :

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

Note:

$$\mathrm{Var}(ay) = a\mathrm{Var}(y)a^T$$

Now we start the derivation :

$$\mathrm{Var}(\hat{Y}_0) = \mathrm{Var}(X_0^T \hat{\beta})$$

$$\mathrm{Var}(\hat{Y}_0) = X_0^T \mathrm{Var}(\hat{\beta}) X_0$$

$$\mathrm{Var}(\hat{Y}_0) = \sigma^2 X_0^T (X^T X)^{-1} X_0$$

- σ^2 is mostly estimated in practice with the sample standard deviation $S^2(RSE)$

$$S^2 = \mathrm{MSE} = \frac{\sum e_i^2}{n-p} = \frac{e^T e}{n-p}$$

  So we get :

$$\mathrm{Var}(\hat{Y}_0) = S^2 X_0^T (X^T X)^{-1} X_0$$

$$\hat{\mathrm{SE}}(\hat{Y}_0) = S\sqrt{X_0^T (X^T X)^{-1} X_0}$$

- Our Confidence Interval is :

$$1 - \alpha \times 100\% \text{ Confidence Interval for } E(Y_0)$$

$$\hat{Y}_0 \pm t_{\frac{\alpha}{2},n-p}\hat{\text{SE}}(\hat{Y}_0)$$

Prediction Interval for New Response Y_0

The matrix form for Multiple Linear Regression:

$$Y = X\beta + \varepsilon$$

a New Response :

$$Y_0 = X_0^T\beta$$

We estimate it with

$$\hat{Y}_0 = X_0^T\hat{\beta}$$

We know From Ordinary Least Squares and Standard Error Derivation :

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

$$\text{Var}(\hat{\beta}) = \sigma^2(X^TX)^{-1}$$

Now we calculate

$$\text{Var}(Y_0 - \hat{Y}_0) = \text{Var}(Y_0) - \text{Var}(\hat{Y}_0)$$

$$\text{Var}(Y_0 - \hat{Y}_0) = \text{Var}(X_0^T\beta + \varepsilon_0) + \text{Var}(X_0^T\hat{\beta})$$

Note : Every constant can be ignored

$$\text{Var}(Y_0 - \hat{Y}_0) = \text{Var}(\varepsilon_0) + X_0^T\text{Var}(\hat{\beta})X_0$$

$$\text{Var}(Y_0 - \hat{Y}_0) = \text{Var}(\varepsilon_0) + X_0^T\sigma^2(X^TX)^{-1}X_0$$

Note :$\varepsilon = \sigma^2 I$

$$\text{Var}(Y_0 - \hat{Y}_0) = \sigma^2 + X_0^T\sigma^2(X^TX)^{-1}X_0$$

$$\text{Var}(Y_0 - \hat{Y}_0) = \sigma^2(1 + X_0^T(X^TX)^{-1}X_0)$$

- σ^2 is mostly unknown practice so we use the sample standard deviation S^2

$$S^2 = \text{MSE} = \frac{\sum e_i^2}{n-p} = \frac{e^Te}{n-p}$$

$$\text{Var}(Y_0 - \hat{Y}_0) = S^2(1 + X_0^T(X^TX)^{-1}X_0)$$

$$\hat{\text{SE}}(Y_0 - \hat{Y}_0) = S\sqrt{1 + X_0^T(X^TX)^{-1}X_0}$$

- Now we can construct the Prediction interval for a new response Y_0

$$(1-\alpha) \times 100\% \text{ Prediction Interval For } Y_0 \text{ is}$$

$$\hat{Y}_0 \pm t_{\frac{\alpha}{2},n-p}\hat{\text{SE}}(Y_0 - \hat{Y}_0)$$

- This Prediction Interval is wider than the Confidence Interval