# Packet 5: Linear Regression

## Chap 7.6 More Regression

**Confidence intervals for $\alpha$ and $\beta$:**

We see that the variances of $\hat{\alpha}$ and $\hat{\beta}$ depend on the unknown error variance $\sigma^2$. Given $\sigma^2$, under the normality assumption, we have

$\hat{\alpha} \sim N(\alpha, \frac{\sigma^2}{n})$

$\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2})$

Replace the unknown $\sigma^2$ with $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ (MLE).

Therefore a $(1-a) \times 100\%$ CI for $\alpha$ is

$$\left[\hat{\alpha} - t_{\frac{a}{2}, n-2} \sqrt{\frac{\hat{\sigma}^2}{(n-2)}}, \hat{\alpha} + t_{\frac{a}{2}, n-2} \sqrt{\frac{\hat{\sigma}^2}{(n-2)}}\right].$$

Similarly, a $(1-a) \times 100\%$ CI for $\beta$ is

$$\left[\hat{\beta} - t_{\frac{a}{2}, n-2} \sqrt{\frac{n\hat{\sigma}^2}{(n-2)\sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{\beta} + t_{\frac{a}{2}, n-2} \sqrt{\frac{n\hat{\sigma}^2}{(n-2)\sum_{i=1}^n (x_i - \bar{x})^2}}\right].$$

These results allow us to construct confidence intervals for $\alpha$ and $\beta$, and to carry out hypothesis tests using $t$-tests, e.g.,

$$H_0 : \beta = 0 \quad \text{v.s.} \quad H_1 : \text{Otherwise.}$$

Gauss-Markov Theorem: It turns out that under the current assumptions, the least square estimators $\hat{\alpha}$ and $\hat{\beta}$ can be proved to be the

Best Linear Unbiased Estimators (BLUEs) for $\alpha$ and $\beta$.

In other words, the least square estimators are the ones with the smallest variance among all unbiased estimators under the current assumptions.

**Confidence interval for** $E(Y_i|x_i)$: We are interested in knowing the expected value of the response for a given value of the predictor.

**Prediction interval for a new observation** $Y_{n+1}$: We are interested in knowing the possible value of a new observation $Y_{n+1}$ for a given value of the predictor.

It covers a new data point $y_{n+1}$ with probability $1 - a$.