1 Prediction & Fit

- Once we have values for $\widehat{a}$ and $\widehat{b}$, we can use them to predict values for Y: $\widehat{Y}_i = a + bX_i$, for some given value of X: X_0.
- Difference between sample and population regression line for $E(Y|X_0)$ is the forecasting or prediction error.
- Note that $\widehat{Y}_0$ is a point estimator and therefore has a sampling distribution:

$$E(Y|X_0) = a + bX_0$$

$$\widehat{\sigma}^2_{Y_0} = \widehat{\sigma}^2_e \left[\frac{1}{n-2} + \frac{(X_0 - \overline{X})^2}{\sum_i x_i^2} \right]$$

- We can also estimate a confidence interval:

$$P \begin{array}{l} [\left(\widehat{a} + \widehat{b}X_0 - t_{\alpha/2}\widehat{\sigma}(Y_0)\right) \leqslant a + bX_0 \leqslant \\ \left(\widehat{a} + \widehat{b}X_0 + t_{\alpha/2}\widehat{\sigma}(Y_0)\right)] = (1-\alpha) \end{array}$$

- We can also look at the Predictive Ability of an estimated equation using the Chow test.
- The Chow test is a 'within sample' prediction test. Test is whether the observations in the sample, and in a truncated part of the sample are from the same population.
- If a sample of size n is drawn, you select a sub-sample of size n_1. Estimate the same model: $\widehat{Y}_i = a + bX_i$ with both n and n_1 and compute $\sum_i e_i^2$ for both models. Compute the F-test:

$$F^* = \frac{\left(\sum_i e^2_{ni} - \sum_i e^2_{n_1 i}\right)/(n - n_1)}{\sum_i e^2_{n_1 i}/(n_1 - k)}$$

- $F^* \sim F_{\alpha,(n-n_1,n_1-k)}$
- Predictive ability and general 'goodness of fit' for the regression equation can be summarized by the Coefficient of determination (R^2):

$$R^2 = \frac{\text{Explained Sum of Squares}}{\text{Total Sum of Squares}}$$

$$R^2 = \frac{\widehat{b}\sum_i x_i y_i}{\sum_i x_i^2}$$