1 Inference in Classical Regression

- Start to put together pieces of regression and statistical inference to understand the significance of a regression line.
- Pieces we should be concerned with:
  1. Estimation of the regression line:
  $$b = \frac{\sum_i XY - n\overline{XY}}{\sum_i X^2 - n(\overline{X})^2} = \frac{\sum_i xy}{\sum_i x^2}$$
  remember I am taking $\widehat{b}$ as the estimate of the true regression parameter b.
  2. Assumptions on the randomness of the dependent variable and the error: $E(e_i) = 0, E(e_i X_i) = 0, E(Y) = a + bX, V(Y) = \sigma^2$, $C(Y_h, Y_i)_{h \neq i} = 0$. Values of X are identical and non-stochastic.
  3. Hypothesis testing and confidence intervals.
- Sum of squares identity:
$$\sum_i (Y_i - \overline{Y})^2 = \sum_i (\widehat{Y}_i - \overline{Y})^2 + \sum_i (Y_i - \widehat{Y}_i)^2$$
- $\sum_i (Y_i - \overline{Y})^2 = \sum_i y_i^2$ and is the total variation in Y from its mean value.
- $\sum_i (\widehat{Y}_i - \overline{Y})^2 = \widehat{b} \sum_i x_i y_i$ and is the explained variation in Y from the mean value.
- $\sum_i (Y_i - \widehat{Y}_i)^2 = \sum_i y_i^2 - \widehat{b} \sum_i x_i y_i = \sum_i e_i^2$ and is the unexplained variation in Y.
- Corresponding degrees of freedom identity:
$$(n-1) = 1 + (n-2)$$
- Residual variance of the regression line:
$$\widehat{\sigma}_{YX}^2 = \frac{\sum_i (Y_i - \widehat{Y}_i)^2}{(n-2)} = \frac{\sum_i e_i^2}{(n-2)}$$
- Can now investigate the sampling distribution of the regression line:
- Variance of $\widehat{a}$ (intercept):
$$\widehat{\sigma}_a^2 = \frac{\sum_i X_i^2}{n \sum_i x_i^2} . \widehat{\sigma}_{YX}^2$$

- Standard deviation of $\widehat{a}$ (intercept):

$$\widehat{\sigma}_a = \sqrt{\widehat{\sigma}_a^2}$$

- Variance of $\widehat{b}$ (slope):

$$\widehat{\sigma}_b^2 = \frac{\widehat{\sigma}_{YX}^2}{\sum_i x_i^2}$$

- Standard deviation of $\widehat{b}$ (slope):

$$\widehat{\sigma}_b = \sqrt{\widehat{\sigma}_b^2}$$

- Confidence interval for b:

$$b = \widehat{b} \pm t_{\alpha/2,df}\widehat{\sigma}_b$$

- Hypothesis test: $H_0 : b = 0$ against $H_1 : b \neq 0$. Test statistic (also known as a t-ratio):

$$t = \frac{\widehat{b} - b}{\widehat{\sigma}_b}$$