Standard Error Derivation

Simple Regression

The assumed model :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Using :

$$E(a + bY) = a + bE(Y)$$

$$Var(a + bY) = b^2 Var(Y)$$

Deriving the Mean

From Ordinary Least Squares we know the estimator for β_1 is :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

And

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - \bar{x})y_i$$

$$\sum(x_i - \bar{x})^2 = \sum(x_i - \bar{x})x_i$$

$$\text{only when } \sum(x - \bar{x}) = 0$$

So $\text{the slope } \hat{\beta}_1$ can be written as :

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}$$

Assuming the x are fixed we get :

$$E(\hat{\beta}_1) = E\left(\frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}\right)$$

$$\text{Since X's are fixed they can be considered constants}$$

$$= \frac{1}{\sum(x_i - \bar{x})^2}\sum E((x_i - \bar{x})y_i)$$

$$= \frac{1}{\sum(x_i - \bar{x})^2}\sum(x_i - \bar{x})E(y_i)$$

$$E(y_i) = E(\beta_0 + \beta_1 x_i + \varepsilon_i)$$

$$E(y_i) = \beta_0 + \beta_1 x_i + E(\varepsilon_i)$$

$\text{We also assume that } \varepsilon \text{ is zero}$

$$= \frac{1}{\sum(x_i - \bar{x})^2}\sum(x_i - \bar{x})(\beta_0 + \beta_1 x_i)$$

$$= \frac{1}{\sum(x_i - \bar{x})^2}\sum(x_i - \bar{x})\beta_0 + \sum\ (x_i - \bar{x})\beta_1 x_i$$

and Since we assume the $\sum(x_i - \bar{x}) = 0$

$$= \frac{\beta_1}{\sum(x_i - \bar{x})^2} + \sum\ (x_i - \bar{x})x_i$$

$\sum(x_i - \bar{x})^2 = \sum(x_i - \bar{x})x_i$

$$E(\beta_1) = \beta_1$$

- which means that the expected value or the mean of β_1 is β_1
  which means its an unbiased estimator

Deriving The Variance (Standard Error):

$$SE(\hat{\beta}_1)^2 = Var(\hat{\beta}_1) = Var\left(\frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}\right)$$

$$= \frac{1}{(\sum(x_i - \bar{x})^2)^2}\sum Var((x_i - \bar{x})y_i)$$

$$= \frac{1}{(\sum(x_i-\bar{x})^2)^2} Var(\sum(x_i-\bar{x})(\beta_0+\beta_1 x_i+\varepsilon_i))$$

$Var(\sum(x_i-\bar{x})(\beta_0-\beta_1 x_i))$ can be canceled since it doesn't effect the variance

$$= \frac{1}{(\sum(x_i-\bar{x})^2)^2} Var(\sum(x_i-\bar{x})\varepsilon_i)$$

- independence implies zero covariance but zero covariance doesn't imply independence, since our error's are uncorrelated (they don't effect each other)

$$= \frac{1}{\left(\sum(x_i-\bar{x})^2\right)^2} \sum Var((x_i-\bar{x})\varepsilon_i)$$

$$= \frac{1}{\left(\sum(x_i-\bar{x})^2\right)^2} \sum(x_i-\bar{x})^2 Var(\varepsilon_i)$$

$$= \frac{1}{\left(\sum(x_i-\bar{x})^2\right)^2} \sum(x_i-\bar{x})^2\sigma^2$$

$$= \frac{\sigma^2}{\left(\sum(x_i-\bar{x})^2\right)^2} \sum(x_i-\bar{x})^2$$

$$\hat{SE}(\hat{\beta}_1)^2 = Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum(x_i-\bar{x})^2}$$

- We only assume that our errors are uncorrelated and the $X's$ are fixed

Normality :

$$\hat{\beta}_1 = \frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}$$

It can be written as a linear combination

$$= \sum c_i y_i = \text{where } c_i = \frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}$$

$$\varepsilon \sim N(0,\sigma^2) \implies yi \sim N(\beta_0+\beta_1 x_i+\sigma^2)$$

If the error are normally distributed and β_0, β_1, X_i are fixed means y_i is normally distributed

Multiple Regression

Deriving The mean of the Coefficients $\hat{\beta}$

Multiple Regression formula can be written as :

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

Lets assume :

$$E(\varepsilon) = 0$$

$$E(Y) = E(X\beta + \varepsilon)$$

$$E(Y) = E(X\beta) + E(\varepsilon)$$

So the Expected Value of The Response is :

$$E(Y) = X\beta$$

We also Know that :

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

The Estimated Coefficient is Unbiased proof:

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

$$E(\hat{\beta}) = (X^TX)^{-1}X^TE(Y)$$

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

Note : $(X^TX)^{-1}(X^TX) = 1$
Which Means that $\hat{\beta}$ is Unbiased Estimator for β

$$E(\hat{\beta}) = \beta$$

Deriving The Variance of Least Squares Coefficient $\hat{\beta}$:

We know : If a is a vector

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

Also That :

$$\mathrm{Var}(Y) = \mathrm{Var}(X\beta + \varepsilon) = \mathrm{Var}(\varepsilon) = \sigma^2 I$$

So :

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

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

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

Note : $(X^TX)^{-1}$ is symmetric so its the same Transpose

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

Note : $(X^TX)^{-1}(X^TX) = 1$

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

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

$$S^2 = \frac{\sum e_i^2}{n-p-1} = \frac{e^Te}{n-p-1} = MSE$$

Interpretation :

- σ^2 Represent the noise which increase the uncertainty in the data
- Larger Gram and Design Matrix means that the Predictors are highly **correlated**