

# 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  $\beta_1$  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

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

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

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

So 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 :

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

Since X's are fixed they can be considered constants

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

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

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

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

We also assume that  $\varepsilon$  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  $\hat{\beta}_1$  is  $\beta_1$   
which means its an unbiased estimator

## Deriving The Variance (Standard Error ):

$$\begin{aligned} SE(\hat{\beta}_1)^2 &= Var(\hat{\beta}_1) = Var\left(\frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}\right) \\ &= \frac{1}{(\sum(x_i - \bar{x})^2)^2} \sum Var((x_i - \bar{x})y_i) \\ &= \frac{1}{(\sum(x_i - \bar{x})^2)^2} Var(\sum(x_i - \bar{x})(\beta_0 + \beta_1 x_i + \varepsilon_i)) \end{aligned}$$

$Var(\sum(x_i - \bar{x})(\beta_0 - \beta_1 x_i))$  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 )

$$\begin{aligned} &= \frac{1}{(\sum(x_i - \bar{x})^2)^2} \sum Var((x_i - \bar{x})\varepsilon_i) \\ &= \frac{1}{(\sum(x_i - \bar{x})^2)^2} \sum (x_i - \bar{x})^2 Var(\varepsilon_i) \\ &= \frac{1}{(\sum(x_i - \bar{x})^2)^2} \sum (x_i - \bar{x})^2 \sigma^2 \\ &= \frac{\sigma^2}{(\sum(x_i - \bar{x})^2)^2} \sum (x_i - \bar{x})^2 \\ \hat{SE}(\hat{\beta}_1)^2 &= Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2} \end{aligned}$$

- 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

$$\begin{aligned} &= \sum c_i y_i = \text{where } c_i = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} \\ \varepsilon &\sim N(0, \sigma^2) \implies y_i \sim N(\beta_0 + \beta_1 x_i + \sigma^2) \end{aligned}$$

If the error are normally distributed and  $\beta_0, \beta_1, X_i$  are fixed means  $y_i$  is normally distributed