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 :

$$\hat{eta}_1 = rac{\sum_{i=1}^n (x_i - ar{x})(y_i - ar{y})}{\sum_{i=1}^n (x_i - ar{x})^2}$$

And

$$egin{split} \sum (x_i-ar{x})(y_i-ar{y}) &= \sum (x_i-ar{x})y_i \ &\sum (x_i-ar{x})^2 &= \sum (x_i-ar{x})x_i \ & ext{only when } \sum (x-ar{x}) &= 0 \end{split}$$

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

$$\hat{eta}_1 = rac{\sum (x_i - ar{x}) y_i}{\sum (x_i - ar{x})^2}$$

Assuming the x are fixed we get :

$$E(\hat{eta}_1) = E\left(rac{\sum (x_i - ar{x})y_i}{\sum (x_i - ar{x})^2}
ight)$$

Since X's are fixed they can be considered constants

$$egin{align} &=rac{1}{\sum(x_i-ar{x})^2}\sum E((x_i-ar{x})y_i)\ &=rac{1}{\sum(x_i-ar{x})^2}\sum(x_i-ar{x})E(y_i)\ &E(y_i)=E(eta_0+eta_1x_i+arepsilon_i)\ &E(y_i)=eta_0+eta_1x_i+E(arepsilon_i) \end{aligned}$$

We also assume that ε is zero

$$egin{split} &=rac{1}{\sum(x_i-ar{x})^2}\sum(x_i-ar{x})(eta_0+eta_1)x_i \ &=rac{1}{\sum(x_i-ar{x})^2}\sum(x_i-ar{x})eta_0+\sum\;(x_i-ar{x})eta_1x_i \end{split}$$

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

$$=rac{eta_1}{\sum (x_i-ar{x})^2}+\sum \;(x_i-ar{x})x_i$$

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

$$E(eta_1)=eta_1$$

• 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):

$$egin{align} SE(\hat{eta}_1)^2 &= Var(\hat{eta}_1) = Var\left(rac{\sum (x_i-ar{x})y_i}{\sum (x_i-ar{x})^2}
ight) \ &= rac{1}{(\sum (x_i-ar{x})^2)^2} \sum Var((x_i-ar{x})y_i) \ &= rac{1}{(\sum (x_i-ar{x})^2)^2} Var(\sum (x_i-ar{x})(eta_0+eta_1x_i+arepsilon_i)) \end{split}$$

 $Var(\sum (x_i - \bar{x})(eta_0 - eta_1 x_i))$ can be canceled since it doesn't effect the variance

$$=rac{1}{(\sum (x_i-ar{x})^2)^2}Var(\sum (x_i-ar{x})arepsilon_i)$$

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

$$egin{aligned} &= rac{1}{(\sum (x_i - ar{x})^2)^2} \sum Var((x_i - ar{x})arepsilon_i) \ &= rac{1}{(\sum (x_i - ar{x})^2)^2} \sum (x_i - ar{x})^2 Var(arepsilon_i) \ &= rac{1}{(\sum (x_i - ar{x})^2)^2} \sum (x_i - ar{x})^2 \sigma^2 \ &= rac{\sigma^2}{(\sum (x_i - ar{x})^2)^2} \sum (x_i - ar{x})^2 \ Var(eta_1) = rac{\sigma^2}{\sum (x_i - ar{x})^2} \end{aligned}$$

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

Normality:

$$\hat{eta}_1 = rac{\sum (x_i - ar{x})y_i}{\sum (x_i - ar{x})^2}$$

It can be written as a linear combination

$$egin{split} &= \sum c_i y_i = ext{where } c_i = rac{\sum (x_i - ar{x}) y_i}{\sum (x_i - ar{x})^2} \ &arepsilon \sim N(0, \sigma^2) \implies yi \sim N(eta_0 + eta_1 x_i + \sigma^2) \end{split}$$

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