Lecture 3

Monday, May 13, 2019 3:16 PM

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

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

$$= \frac{Cov(X,Y)}{Var(X)}$$

$$* \; Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)\,SD(Y)}$$

$$= \frac{Corr(X,Y) \times SD(Y)}{SD(X)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\rightarrow \hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x$$ ← Eqn of line of best fit.

What about an estimate of σ?

- Recall $\sigma = SD(\varepsilon_i)$
- We can't observe the $\varepsilon_i = y_i - \mu_i$, but we can observe $e_i = y_i - \hat{\mu}_i$, which we call a residual. We estimate σ by quantifying dispersion in the residuals:

$\bar{e} = 0$ "sum of squared residuals"

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(e_i-\bar{e})^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)^2}{n-2}}$$

"residual error"

PGA Example:

$\hat{\beta}_1 = -0.39175$

$\hat{\beta}_0 = 176.5841$

$\hat{\sigma} = 4.215$

How do we interpret this?

$$E[y] = \mu = \beta_0 + \beta_1 x$$

- If $x=0 \Rightarrow E[y|x=0] = \beta_0$ = expected response when $x=0$
- $E[y|x=a+1] - E[y|x=a] = [\beta_0 + \beta_1(a+1)] - [\beta_0 + \beta_1 a] = \beta_1$

  = expected change in y for a unit increase in x

- σ is loosely interpretted as the average deviation between an observed response and our predicted value for it.