LHS:

$$\sum_i x_i y_i - n\bar{x}\bar{y}$$

RHS:

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

$$= \sum_i (x_i y_i - \bar{x} y_i - \underline{\bar{y} x_i} + \bar{x}\bar{y})$$

$$= \sum_i x_i y_i - \sum_i \bar{x} y_i - \bar{y}\sum_i x_i + n\bar{x}\bar{y}$$

aside
$$\begin{cases} -(\bar{x}y_1 + \bar{x}y_2 + \cdots + \bar{x}y_n) \\ -\bar{x}\left(\sum_i y_i\right) \\ -\bar{x}(n\bar{y}) \end{cases}$$

$$= \sum_i x_i y_i - n\bar{x}\bar{y} - \cancel{n\bar{x}\bar{y} + n\bar{x}\bar{y}}$$

General:

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

for any $\{(x_i, y_i)\}_{i=1}^n$

Special Case:

$$x_i = y_i \ \forall\, i = 1, \ldots, n$$

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

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

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

Inference & Diagnostics for SLR

First some useful facts:

Recall if $e_i = y_i - \hat{y}_i$ then:

1. $\sum_i e_i^2$ is minimized.

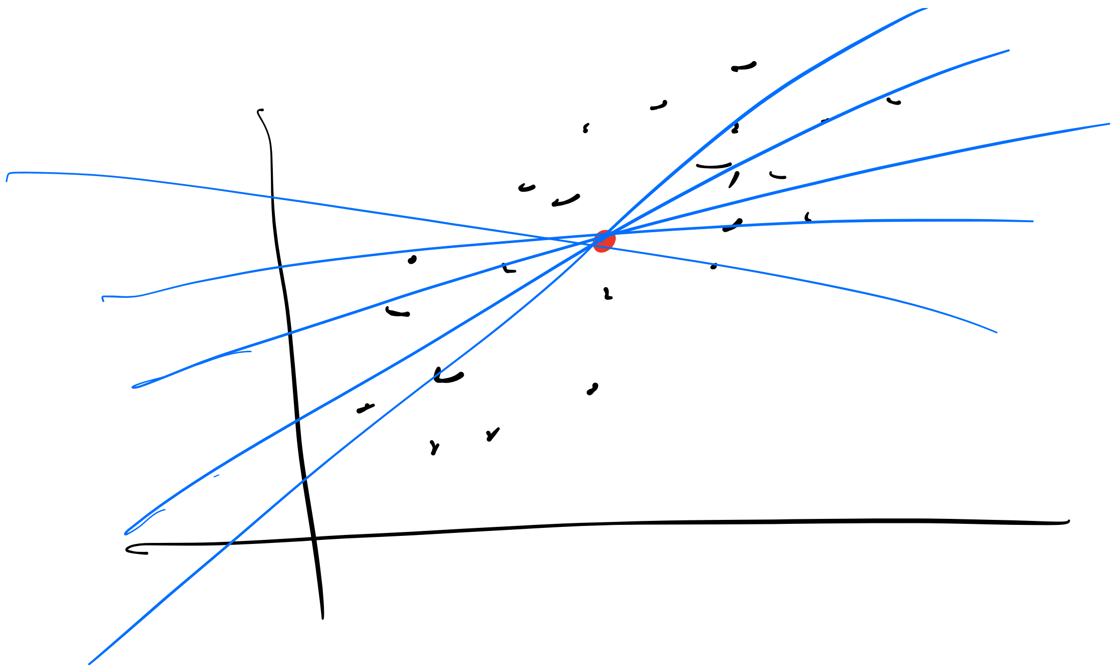
2. $\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i$

3. $\sum_{i=1}^n e_i = 0$

4. $\sum_{i=1}^n x_i e_i = 0$

5. $\sum_{i=1}^n \hat{y}_i e_i = 0$

6. The regression line always goes through $(\bar{x}, \bar{y})$



Pf of 2:

$$\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$= \sum_{i=1}^{n} \hat{\beta}_0 + \sum_{i=1}^{n} \hat{\beta}_1 x_i$$

$$= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i$$

$$= n\hat{\beta}_0 + n\hat{\beta}_1 \bar{x}$$

$$= n(\hat{\beta}_0 + \hat{\beta}_1 \bar{x})$$

$$= n(\bar{y} - \cancel{\hat{\beta}_1 \bar{x}} + \cancel{\hat{\beta}_1 \bar{x}})$$

$$= \sum_{i=1}^{n} y_i$$

We have these estimators

$\rightarrow$ $\hat{\beta}_0$ & $\hat{\beta}_1$ which estimate

$\rightarrow$ β_0 & β_1.

If for a given unknown parameter θ we have an estimator $\hat{\theta}$ which satisfies:

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

then we say $\hat{\theta}$ is <u>unbiased</u> for θ.

Claim: $\hat{\beta}_0$ & $\hat{\beta}_1$ are unbiased for β_0 & β_1, resp.

Def:

$$\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Pf for $\hat{\beta}_1$:

WTS:

$$E(\hat{\beta}_1) = \beta_1$$

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

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

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

$$= E\left(\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{SSX}\right)(y_i - \bar{y})\right)$$

$$= \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{SSX} \underbrace{E(y_i - \bar{y})}_{A}$$

$$A \quad E(y_i - \bar{y}) = E(y_i) - E(\bar{y})$$

$$= \beta_0 + \beta_1 x_i - E\left(\frac{1}{n}\sum_i y_i\right)$$

$$= \beta_0 + \beta_1 x_i - \frac{1}{n}\sum_{i=1}^{n} E(y_i)$$

$$= \beta_0 + \beta_1 x_i - \left(\frac{1}{n}\sum_{i=1}^{n} (\beta_0 + \beta_1 x_i)\right)$$

$$= \beta_0 + \beta_1 x_i - \left(\beta_0 + \beta_1 \left(\frac{1}{n}\sum_i x_i\right)\right)$$

$$= \cancel{\beta_0} + \beta_1 x_i - \cancel{\beta_0} - \beta_1 \bar{x}$$

$$= \beta_1 (x_i - \bar{x})$$

$$= \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{SSX} \left(\beta_1 (x_i - \bar{x}) \right)$$

$$= \sum_{i=1}^{n} \frac{\beta_1 (x_i - \bar{x})^2}{SSX}$$

$$= \frac{\beta_1}{SSX} \left(\sum_{i=1}^{n} (x_i - \bar{x})^2 \right)$$

$$= \frac{\beta_1}{\cancel{SSX}} \cdot \cancel{SSX} = \beta_1$$

EX: Show that $E(\hat{\beta}_0) = \beta_0$

What about variance?

The key is to notice that $\hat{\beta}_0$ & $\hat{\beta}_1$ are both linear estimators (in y_i's).

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

k_i

Define $k_i = \frac{x_i - \bar{x}}{SSX}$

$$Var(\hat{\beta}_1) = Var\left(\sum_{i=1}^{n} k_i y_i\right)$$

because $y_i \perp\!\!\!\perp y_j \ \forall \ i \neq j$

$$= \sum_{i=1}^{n} Var(k_i y_i)$$

$$= \sum_{i=1}^{n} k_i^2 \, Var(y_i)$$

$$= \sum_{i=1}^{n} k_i^2 \, (\sigma^2)$$

$$= \sigma^2 \underbrace{\sum_{i=1}^{n} k_i^2}_{\text{algebra}} = \sigma^2 / SSX$$

$$\boxed{= \sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\sum_{i=1}^{n} k_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{SSX} \right)^2$$

$$= \sum_{i=1}^{n} \left(\frac{(x_i - \bar{x})^2}{(SSX)^2} \right)$$

$$= \frac{1}{(SSX)^2} \boxed{\sum_i (x_i - \bar{x})^2}$$

$$= \frac{1}{(SSX)^2} \cdot SSX = \frac{1}{SSX}$$

HW:

$$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{SSX} \right)$$

Strategy:

Express $\hat{\beta}_0 = \sum_{i=1}^{n} c_i y_i$

(Figure out what c_i is.)

Argue that

$$Var(\hat{\beta}_0) = \sigma^2 \sum_{i=1}^{n} c_i^2$$

Calculate $\sum_i c_i^2$.