keep in mind: $E(aX) = a\,E(X)$
$E(\sum a_i X_i) = \sum a_i E(X_i)$
$\sum a X_i = a \sum X_i$
$\sum a_i X_i \neq a_i \sum X_i$

Derivation of OLS estimators

$\hat{\alpha}, \hat{\beta}$ are values that minimize SS deviations from the line defined by $\hat{\alpha}$ & $\hat{\beta}$ & the data (X_i, Y_i)

/ write down SS deviations
/ take derivative w.r.t. α, β (get two derivatives)
/ set derivatives equal to 0
/ solve for $\hat{\alpha}, \hat{\beta}$

$$S(\alpha,\beta) = \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

if X stochastic, can think of normal eqns as sample counterparts of restict^ns on the joint distribut^n of X & Y and, therefore, one could motivate OLS as a M of M estimator

take derivatives

$$\alpha: \frac{\partial S}{\partial \alpha} = \sum -2(Y_i - \alpha - \beta X_i)$$

$$\beta: \frac{\partial S}{\partial \beta} = \sum -2X_i(Y_i - \alpha - \beta X_i)$$

set equal to 0 (can also multiply both sides by constants)

We sometimes write $\hat{e}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$

$$\alpha: \frac{1}{n}\sum (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0$$

$$\beta: \frac{1}{n}\sum X_i(Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0$$

"normal equations"

Solve first normal equation for $\hat{\alpha}$

$$\underbrace{\tfrac{1}{n}\Sigma Y_i}_{\bar{Y}} - \hat{\alpha} - \hat{\beta}\underbrace{\tfrac{1}{n}\Sigma X_i}_{\bar{X}} = 0$$

so $\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$ or $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$

take first normal equation again & multiply by $\bar{X}$

$$\bar{X}\tfrac{1}{n}\Sigma(Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0$$

subtract it from the second normal eqn

$$\tfrac{1}{n}\Sigma X_i(Y_i - \hat{\alpha} - \hat{\beta}X_i) - \tfrac{1}{n}\Sigma\bar{X}(Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0$$

nothing wrong w/ sticking it inside the summation

plug in $\bar{Y} - \hat{\beta}\bar{X}$ for $\hat{\alpha}$ & collect terms

$$\tfrac{1}{n}\Sigma(Y_i - (\bar{Y} - \hat{\beta}\bar{X}) - \hat{\beta}X_i)(X_i - \bar{X}) = 0$$

$$\tfrac{1}{n}\Sigma((Y_i - \bar{Y}) - \hat{\beta}(X_i - \bar{X}))(X_i - \bar{X}) = 0$$

$$\tfrac{1}{n}\Sigma((X_i - \bar{X})(Y_i - \bar{Y})) - \hat{\beta}\tfrac{1}{n}\Sigma((X_i - \bar{X})^2) = 0$$

$$\boxed{\begin{array}{l} \text{so } \hat{\beta} = \dfrac{\frac{1}{n}\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\Sigma(X_i - \bar{X})^2} \\ \text{and } \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \end{array}}$$

expectation of $\hat{\beta}$ & $\hat{\alpha}$

$$\hat{\beta} = \frac{\sum(X_i - \bar{X})(\overbrace{\alpha + \beta X_i + \varepsilon_i}^{Y_i} \overbrace{- \alpha - \beta\bar{X} - \bar{\varepsilon}}^{\bar{Y}})}{\sum(X_i - \bar{X})^2}$$

$$= \frac{\sum(X_i - \bar{X})(\beta(X_i - \bar{X}) + (\varepsilon_i - \bar{\varepsilon}))}{\sum(X_i - \bar{X})^2}$$

$$= \beta\frac{\sum(X_i - \bar{X})^2}{\sum(X_i - \bar{X})^2} + \frac{\sum(X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})}{\sum(X_i - \bar{X})^2}$$

so $E(\hat{\beta}) = \beta + \frac{\sum(X_i - \bar{X})\, E(\varepsilon_i - \bar{\varepsilon})}{\sum(X_i - \bar{X})^2}$

but $E(\varepsilon_i - \bar{\varepsilon}) = 0$ because $E(\varepsilon_i) = 0$ &
$E(\bar{\varepsilon}) = 0$

$= \beta$

$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = \alpha + \beta\bar{X} + \bar{\varepsilon} - \hat{\beta}\bar{X}$

so $E(\hat{\alpha}) = \alpha + \beta\bar{X} + E(\bar{\varepsilon}) - \underbrace{E(\hat{\beta})}_{=\beta}\bar{X}$

$= \alpha$

variance of $\hat{\beta}$

$$\mathrm{Var}(\hat{\beta}) = E\left[(\hat{\beta}-\beta)^2\right],$$

$$= E\left[\frac{\sum (X_i-\bar{X})(\varepsilon_i-\bar{\varepsilon})}{\sum (X_i-\bar{X})^2}\right]^2$$

from calculation of $E(\hat{\beta})$

$$= \frac{1}{\left[\sum (X_i-\bar{X})^2\right]^2} E\left[\left(\sum (X_i-\bar{X})(\varepsilon_i-\bar{\varepsilon})\right)^2\right]$$

$$\left\{ \begin{aligned} \text{but } \sum (X_i-\bar{X})(\varepsilon_i-\bar{\varepsilon}) &= \sum (X_i-\bar{X})\varepsilon_i - \sum (X_i-\bar{X})\bar{\varepsilon} \\ &= \sum (X_i-\bar{X})\varepsilon_i - \bar{\varepsilon}\underbrace{\sum (X_i-\bar{X})}_{=0 \text{ by construct}^{n}} \\ &= \sum (X_i-\bar{X})\varepsilon_i \end{aligned} \right.$$

$$= \frac{1}{\left[\sum (X_i-\bar{X})^2\right]^2} E\left[\left(\sum (X_i-\bar{X})\varepsilon_i\right)^2\right]$$

$$= \frac{1}{\left[\sum (X_i-\bar{X})^2\right]^2} E\left[\underbrace{\sum (X_i-\bar{X})^2\varepsilon_i^2}_{\text{diagonals}} + \sum_{i\neq j} (X_i-\bar{X})(X_j-\bar{X})\varepsilon_i\varepsilon_j\right]$$

$$= \frac{1}{\left[\sum (X_i-\bar{X})^2\right]^2} \left\{ \sum (X_i-\bar{X}) \underbrace{E(\varepsilon_i^2)}_{=\sigma^2}\right.$$

$$+ \sum_{i \neq j} (x_i - \bar{x})(x_j - \bar{x}) \underbrace{E(\varepsilon_i \varepsilon_j)}_{=0} \Big\}$$

$$= \frac{\sigma^2}{\Sigma (x_i - \bar{x})^2} = \frac{\sigma^2}{n \hat{\sigma}_x^2} \quad \text{where } \hat{\sigma}_x^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2$$

variance of $\hat{\alpha}$

$$Var(\hat{\alpha}) = E(\hat{\alpha} - \alpha)^2$$

$$= E\Big[\big[(\underbrace{\bar{Y} - \hat{\beta}\bar{X}}_{\hat{\alpha}}) - (\underbrace{\bar{Y} - \beta\bar{X} - \bar{\varepsilon}}_{\alpha})\big]^2\Big]$$

$$= E\left[\left[-(\hat{\beta} - \beta)\bar{X} + \bar{\varepsilon}\right]^2\right]$$

$$= E\left[(\hat{\beta} - \beta)^2 \bar{X}^2\right] + E(\bar{\varepsilon}^2) - 2E\left[(\hat{\beta} - \beta)\bar{X}\bar{\varepsilon}\right]$$

$$= \underbrace{\bar{X}^2 \, Var(\hat{\beta})} + \underbrace{Var(\bar{\varepsilon})} - 2\bar{X} \, \underbrace{Cov(\hat{\beta}, \bar{\varepsilon})}$$

$$= \frac{\bar{X}^2 \sigma^2}{n \hat{\sigma}_x^2} + \frac{\sigma^2}{n}$$

note:

$$\left[\hat{\sigma}_x^2 = \frac{1}{n} \Sigma\left[(X_i - \bar{X})^2\right]\right]$$

$$Cov(\hat{\beta}, \bar{\varepsilon}) = E\left[(\hat{\beta} - \beta)\bar{\varepsilon}\right]$$

$$= E\Big[\underbrace{\frac{\Sigma (X_i - \bar{X})\varepsilon_i}{n \hat{\sigma}_x^2}}_{\text{got this from calculat}^n\text{ of variance of }\hat{\beta}} \ \underbrace{\sum_j \frac{\varepsilon_j}{n}}_{\text{by definit}^n}\Big]$$

$$= \frac{1}{n^2 \hat{\sigma}_x^2} E\left[\Sigma (X_i - \bar{X})\varepsilon_i^2 + \sum_{i \neq j} (X_i - \bar{X})\varepsilon_i \varepsilon_j\right]$$

diagonals

off-diagonals

since $E(\varepsilon_i^2) = \sigma^2$ & $E(\varepsilon_i \varepsilon_j) = 0$

$$= \frac{1}{n^2 \hat{\sigma}_x^2} \Sigma (X_i - \bar{X}) \sigma^2 = 0$$

covariance between $\hat{\alpha}, \hat{\beta}$

$$
\begin{aligned}
\operatorname{Cov}(\hat{\alpha}, \hat{\beta}) &= E\left[(\hat{\alpha}-\alpha)(\hat{\beta}-\beta)\right] \\
&= E\left[\left(-(\hat{\beta}-\beta)\bar{x} + \bar{\varepsilon}\right)(\hat{\beta}-\beta)\right] \\
&= \underbrace{E\left[-(\hat{\beta}-\beta)^2 \bar{x}\right]} + \underbrace{E\left[\bar{\varepsilon}\,(\hat{\beta}-\beta)\right]}_{=0 \text{ (see calculation above)}} \\
&= -\bar{x}\operatorname{Var}(\hat{\beta}) \\
&= \frac{-\bar{x}\,\sigma^2}{n\,\hat{\sigma}_x^2}
\end{aligned}
$$