Homework 4

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

$$E(\hat{\beta}_1) = E\left(\frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} \right) = \frac{1}{\underbrace{\sum (x_i - \bar{x})^2}_{\text{constant}}} E\left(\sum (x_i - \bar{x}) y_i\right)$$

$$= \frac{1}{\sum (x_i - \bar{x})^2} \sum E\left((\underbrace{x_i - \bar{x}}_{\text{constant}}) y_i\right)$$

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

$$= \frac{1}{\sum (x_i - \bar{x})^2} \sum (x_i - \bar{x}) E(\overbrace{\beta_0 + \beta_1 x_i}^{\text{fixed}} + \epsilon_i)$$

$$= \frac{1}{\sum (x_i - \bar{x})^2} \sum (x_i - \bar{x}) (\beta_0 + \beta_1 x_i + \overbrace{E(\epsilon_i)}^{=0})$$

$$= \frac{1}{\sum (x_i - \bar{x})^2} \sum (x_i - \bar{x}) (\beta_0 + \beta_1 x_i)$$

$$= \frac{1}{\sum (x_i - \bar{x})^2} \left(\overbrace{\sum (x_i - \bar{x})}^{0} \beta_0 + \sum (x_i - \bar{x}) \beta_1 x_i \right)$$

(as $\sum (x_i - \bar{x}) \beta_0 = \beta_0 \sum (x_i - \bar{x}) = \beta_0 (\overbrace{\sum x_i - n\bar{x}}^{0}) = 0$

$$= \frac{\beta_1}{\sum(x_i - \bar{x})^2} \sum (x_i - \bar{x}) x_i = \beta_1$$

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

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

$$= \sum (x_i - \bar{x}) x_i - \bar{x} \underbrace{\sum (x_i - \bar{x})}_{=0}$$

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

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

$$Var(\hat{\beta}_1) = Var\left(\frac{\sum(x_i - \bar{x}) y_i}{\sum(x_i - \bar{x})^2}\right) = \frac{1}{(\sum(x_i - \bar{x})^2)^2} Var\left(\sum(x_i - \bar{x}) y_i\right)$$

$$= \frac{1}{(\sum(x_i - \bar{x})^2)^2} Var\left(\sum (x_i - \bar{x})(\beta_0 + \beta_1 x_i + \epsilon_i)\right)$$

$$= \frac{1}{(\sum(x_i - \bar{x})^2)^2} Var\left(\underbrace{\sum (x_i - \bar{x})(\beta_0 + \beta_1 x_i)}_{\text{constant}} + \sum (x_i - \bar{x})\epsilon_i\right)$$

$$= \frac{1}{(\sum(x_i - \bar{x})^2)^2} Var\left(\sum (x_i - \bar{x})\epsilon_i\right)$$

$$= \frac{1}{(\sum(x_i - \bar{x})^2)^2} \sum \left(Var((x_i - \bar{x})\epsilon_i)\right)$$

(because $Var(\sum(x_i - \bar{x})\epsilon_i) = \sum Var((x_i - \bar{x})\epsilon_i) + 2 \sum_{i<j} \sum (x_i - \bar{x})(x_j - \bar{x}) \underbrace{Cov(\epsilon_i, \epsilon_j)}_{= 0 \text{ by definition}}$

$$= \sum Var((x_i - \bar{x})\epsilon_i) \;)$$

$$= \frac{1}{(\sum(x_i - \bar{x})^2)^2} \sum Var(\underbrace{(x_i - \bar{x})}_{\text{constant}}\epsilon_i)$$

$$= \frac{1}{(\sum(x_i - \bar{x})^2)^2} \left(\sum (x_i - \bar{x})^2 Var(\epsilon_i)\right)$$

$$= \frac{1}{(\sum(x_i - \bar{x})^2)^2} \left(\sum (x_i - \bar{x})^2 \sigma^2\right)$$

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

$\Rightarrow Var(\hat{\beta}_1) = \sigma^2 / \sum (x_i - \bar{x})^2$

$\Rightarrow \hat{\beta}_1 \sim N(\beta_1, \sigma^2 / \sum (x_i - \bar{x})^2)$

• $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$\Rightarrow E(\hat{\beta}_0) = E(\bar{Y} - \hat{\beta}_1 \bar{X}) = E(\bar{Y}) - E(\hat{\beta}_1 \bar{X})$

$= E(Y_i) - \beta_1 x_i$

$= \beta_0 + \beta_1 X_i + \cancelto{0}{E(\epsilon)} - \beta_1 X_i = \beta_0$

$\Rightarrow E(\hat{\beta}_0) = \beta_0$

$Var(\hat{\beta}_0) = Var(\bar{y} - \hat{\beta}_1 \bar{X}) = Var(\bar{y}) + (\bar{X})^2 Var(\hat{\beta}_1)$
$- 2\bar{X} Cov(\bar{y}, \hat{\beta}_1)$

We have $Var(\bar{y}) = var\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(y_i) = \frac{\sigma^2}{n}$

$\Big(var(\Sigma y_i) = var(\Sigma(\underbrace{\beta_0 + \beta_1 x_i}_{constant} + \epsilon_i)) = Var(\Sigma \epsilon_i)$

$= \sum_{i=1}^{n}\sum_{j=1}^{n} cov(\epsilon_i, \epsilon_j) = \sum_{i=1}^{n} cov(\epsilon_i, \epsilon_i)$

$= \sum_{i=1}^{n} var(\epsilon_i) = \sum_{i=1}^{n} var(\beta_0 + \beta_1 x_i + \epsilon_i)$

$= \sum_{i=1}^{n} var(y_i) \Big)$

From previous step, we have $(\bar{x})^2 Var(\hat{\beta}_1) = \frac{\bar{x}^2 \sigma^2}{\Sigma(x_i - \bar{x})^2}$

Let's look at $Cov(\bar{y}, \hat{\beta}_1) = Cov\left(\frac{1}{n}\Sigma y_i, \frac{\Sigma(x_i - \bar{x})y_i}{\Sigma(x_i - \bar{x})^2}\right)$

$= \frac{1}{n}\frac{1}{\Sigma(x_i - \bar{x})^2} Cov(\Sigma y_i, \Sigma(x_j - \bar{x}) y_j)$

$= \frac{1}{n\Sigma(x_i - \bar{x})^2} \Sigma(x_j - \bar{x}) \Sigma Cov(y_i, y_j)$

$= \frac{1}{n\Sigma(x_i - \bar{x})^2} \underbrace{\Sigma(x_j - \bar{x})}_{=0} \sigma^2 = 0$

$\Rightarrow Var(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{x}^2}{\Sigma(x_i - \bar{x})^2} = \frac{\sigma^2 \Sigma(x_i - \bar{x})^2}{n\Sigma(x_i - \bar{x})^2} + \frac{n\sigma^2 \bar{x}^2}{n\Sigma(x_i - \bar{x})^2}$

$= \frac{\sigma^2}{n\Sigma(x_i - \bar{x})^2}\left(\Sigma(x_i - \bar{x})^2 + n\bar{x}^2\right)$

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

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

$\Rightarrow Var(\hat{\beta}_0) = \frac{\sigma^2}{n\sum(x_i-\bar{x})^2}\left(\sum x_i^2 - n\bar{x}^2 + n\bar{x}^2\right)$

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

$\Rightarrow \hat{\beta}_0 \sim \left(\beta_0, \frac{\sigma^2 \sum x_i^2}{n\sum(x_i-\bar{x})^2}\right)$

b) For a new covariate X_{n+1}, the ~~p~~ 99% confidence interval for $E(Y_{n+1} \mid X_{n+1})$ is

$\hat{Y}_{n+1} \pm t\left(1 - \frac{0.01}{2}, n-2\right) s\{\hat{Y}_{n+1}\}$

$= \hat{Y}_{n+1} \pm t(0.995, n-2) s\{\hat{Y}_{n+1}\}$

We have $s^2\{\hat{Y}_{n+1}\} = var(\bar{Y}) + Var(\hat{\beta}_1(X_{n+1} - \bar{X}))$

(as $\hat{Y}_{n+1} = \bar{Y} + \hat{\beta}_1(X_{n+1} - \bar{X})$)

$\Rightarrow s^2\{\hat{Y}_{n+1}\} = \frac{\sigma^2}{n} + (\overbrace{X_{n+1} - \bar{X}}^{\text{fixed}})^2 \, Var(\hat{\beta}_1)^2$

$= \frac{\sigma^2}{n} + (X_{n+1} - \bar{X})^2 \frac{\sigma^2}{\sum(x_i-\bar{x})^2}$

$\Rightarrow$ 99% CI is $\hat{Y}_{n+1} \pm t(0.995, n-2) \cdot \sqrt{\left(\frac{\sigma^2}{n} + (X_{n+1}-\bar{X})^2 \frac{\sigma^2}{\sum(x_i-\bar{x})^2}\right)}$

sigma known so can use z-score for CI of part b

c) From previous homework, we have proved that the ~~OLS~~ $\hat{\sigma}^{(OLS)}$ is unbiased

$\Rightarrow E(\hat{\sigma}^{(OLS)2}) = \sigma^2$

if sigma unknown, part c we use t-dist

So to replace σ^2 with $(\hat{\sigma}^{OLS})^2$ in the interval derived in part ~~a~~ b, we are still going to get the same valid interval, which means if we repeat the experiment many times, the CI will contain $E(Y_{n+1} \mid X_{n+1})$ approximately 99% of the time.

d) 99% PI for Y_{n+1} assuming σ^2 is known is

$$\hat{Y}_{n+1} \pm z(0.995).\sigma$$

This interval is smaller than the one in b), and it agrees with my intuition because it is only predicting a single point, in contrast to b where it gives the description for the mean response $E(Y_{n+1} \mid X_{n+1})$.

Wider, the multiplier is z-score the same with part (b)
the standard error now is sigma, > s(Y^h)