HW (10 pts bonus)

February 28, 2018

```
BIOSTAT 705 Spring 2018
Due on March 2nd
(Have all your answers on this sheet)
```

In a simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$:
Prove the following statements without using matrix form:

1 Part a (2 pts) $\bar{e} = \frac{1}{n}\sum e_i = 0$

Let the sum of squared residual be Q, where $Q = \sum e_i^2$

By minimizing Q with respect to β_0

$\frac{\partial Q}{\partial \beta_0} = 0 \Rightarrow -\sum 2e_i = 0 \Rightarrow \bar{e} = \frac{1}{n}\sum e_i = 0$

2 Part b (3 pts) $E[\hat{\beta}_0] = \beta_0$; $E[\hat{\beta}_1] = \beta_1$

(1)

$E[\hat{\beta}_1]$
$= E\left[\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}\right]$
$= \frac{\sum(x_i-\bar{x})E[(y_i-\bar{y})]}{\sum(x_i-\bar{x})^2}$
$= \frac{\sum(x_i-\bar{x})E[(\beta_0+\beta_1 x_i+\epsilon_i)-(\beta_0+\beta_1\bar{x})]}{\sum(x_i-\bar{x})^2}$
$= \frac{\sum(x_i-\bar{x})E[\beta_1(x_i-\bar{x})+\epsilon_i]}{\sum(x_i-\bar{x})^2}$
$= \frac{\sum(x_i-\bar{x})\beta_1(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$ (since $E[\epsilon_i] = 0$)
$= \beta_1$

(2)

$E[\hat{\beta}_0]$
$= E[\bar{y} - \hat{\beta}_1\bar{x}]$
$= E[\bar{y}] - E[\hat{\beta}_1\bar{x}]$
$= \beta_0 + \beta_1\bar{x} - \beta_1\bar{x}$
$= \beta_0$

3 Part c (5 pts) $\mathbf{var}(e_i) = \mathbf{var}(y_i - \hat{y}_i) = \sigma^2\left[1 - \frac{1}{n} - \frac{x_i - \bar{x}}{\sum_j (x_i - \bar{x})^2}\right]$

On one hand, $\beta_0, \beta_1, x_i, \bar{x}$ are given and thus are treated as constants. On the other hand, $\hat{\beta}_0, \hat{\beta}_1, y_i, \bar{y}$ are treated as variables in the model.

(1)

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
$\epsilon_i \sim N(0, \sigma^2)$
$\Rightarrow y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

(2)

$\bar{y} = \frac{1}{n}\sum y_i = \frac{1}{n}\sum(\beta_0 + \beta_1 x_i + \epsilon_i) = \beta_0 + \beta_1 \bar{x} + \frac{1}{n}\sum \epsilon_i$
$\Rightarrow \mathrm{var}(\bar{y}) = \frac{\sigma^2}{n}$

(3)

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
$\hat{\beta}_1$
$= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$
$= \sum c_i (y_i - \bar{y})$
$= \sum c_i y_i - \bar{y}\sum c_i$
$= \sum c_i y_i$
where $c_i = \frac{x_i - \bar{x}}{\sum(x_i - \bar{x})^2}$ and $\sum c_i = 0$

(4)

$\mathrm{var}(\hat{\beta}_0) = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right]$
$\mathrm{var}(\hat{\beta}_1) = \sigma^2 \frac{1}{\sum(x_i - \bar{x})^2}$
$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{y} + \hat{\beta}_1(x_i - \bar{x})$
$\Rightarrow var(\hat{y}_i) = \sigma^2\left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right]$

(5)

$\mathrm{var}(e_i)$
$= \mathrm{var}(y_i - \hat{y}_i)$
$= \mathrm{var}(\beta_0 + \beta_1 x_i + \epsilon_i - \hat{y}_i)$ (by (1))
$= \mathrm{var}(\epsilon_i - \hat{y}_i)$ (since β_0, β_1, x_i are treated as constants)
$= \mathrm{var}(\epsilon_i) + \mathrm{var}(\hat{y}_i) - 2\mathrm{cov}(\epsilon_i, \hat{y}_i)$
$= \sigma^2 + \sigma^2\left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right] - 2\mathrm{cov}(\epsilon_i, \hat{y}_i)$ (by (4))

(6)

$\mathrm{cov}(\epsilon_i, \hat{y}_i)$
$= E[\epsilon_i \hat{y}_i] - E[\epsilon_i]E[\hat{y}_i]$
$= E[\epsilon_i \hat{y}_i]$ (since $E[\epsilon_i] = 0$)
$= E\left[\epsilon_i\left(\bar{y} + \hat{\beta}_1(x_i - \bar{x})\right)\right]$
$= E\left[\epsilon_i \bar{y} + \epsilon_i \hat{\beta}_1 (x_i - \bar{x})\right]$
$= E\left[\epsilon_i \bar{y}\right] + (x_i - \bar{x})E\left[\epsilon_i \hat{\beta}_1\right]$

(7)

$E\left[\epsilon_i \bar{y}\right]$
$= E\left[\epsilon_i \left(\beta_0 + \beta_1 \bar{x} + \frac{1}{n} \sum_j \epsilon_j\right)\right]$ (by (2))
$= E\left[\epsilon_i \beta_0 + \epsilon_i \beta_1 \bar{x} + \frac{1}{n} \epsilon_i \sum_j \epsilon_j)\right]$
$= E\left[\epsilon_i \beta_0\right] + E\left[\epsilon_i \beta_1 \bar{x}\right] + E\left[\frac{1}{n} \epsilon_i \sum_j \epsilon_j)\right]$
$= \beta_0 E\left[\epsilon_i\right] + \beta_1 \bar{x} E\left[\epsilon_i\right] + \frac{1}{n} E\left[\epsilon_i \sum_j \epsilon_j\right]$
$= \frac{1}{n} E\left[\epsilon_i \sum_j \epsilon_j\right]$
Since $\epsilon_i \perp \epsilon_j$ for $i \neq j$ and $\epsilon_i \sim N(0, \sigma^2)$ (by (1))
$\Rightarrow E\left[\epsilon_i \epsilon_j\right] = E[\epsilon_i] E[\epsilon_j] = 0$ (for $i \neq j$)
$\Rightarrow E\left[\epsilon_i \sum_j \epsilon_j\right]$
$= E\left[\epsilon_i \epsilon_1 + \cdots + \epsilon_i \epsilon_i + \cdots + \epsilon_i \epsilon_n\right]$
$= E\left[\epsilon_i \epsilon_1\right] + \cdots + E\left[\epsilon_i \epsilon_i\right] + \cdots + E\left[\epsilon_i \epsilon_n\right]$
$= E\left[\epsilon_i \epsilon_i\right]$
$= \text{var}(\epsilon_i) + (E[\epsilon_i])^2$
$= \text{var}(\epsilon_i)$
$= \sigma^2$
Therefore, $\frac{1}{n} E\left[\epsilon_i \sum_j \epsilon_j\right] = \frac{\sigma^2}{n}$

(8)

$E\left[\epsilon_i \hat{\beta}_1\right]$
$= E\left[\epsilon_i \sum_j c_j y_j\right]$ (by (3))
$= E\left[\epsilon_i \sum_j c_j (\beta_0 + \beta_1 x_j + \epsilon_j)\right]$ (by (1))
for $i \neq j$
$E\left[\epsilon_i c_j (\beta_0 + \beta_1 x_j + \epsilon_j)\right]$
$= E\left[\epsilon_i c_j \beta_0 + \epsilon_i c_j \beta_1 x_j + \epsilon_i c_j \epsilon_j)\right]$
$= c_j \beta_0 E\left[\epsilon_i\right] + c_j \beta_1 x_j E\left[\epsilon_i\right] + c_j E\left[\epsilon_i \epsilon_j\right]$
$= 0$
for $i = j$
$E\left[\epsilon_i c_i (\beta_0 + \beta_1 x_i + \epsilon_i)\right]$
$= E\left[\epsilon_i c_i \beta_0 + \epsilon_i c_i \beta_1 x_i + \epsilon_i c_i \epsilon_i)\right]$
$= c_i \beta_0 E\left[\epsilon_i\right] + c_i \beta_1 x_i E\left[\epsilon_i\right] + c_i E\left[\epsilon_i \epsilon_i\right]$
$= c_i E\left[\epsilon_i \epsilon_i\right]$
$= c_i \sigma^2$
$\Rightarrow E\left[\epsilon_i \hat{\beta}_1\right] = c_i \sigma^2$
$\Rightarrow (x_i - \bar{x}) E\left[\epsilon_i \hat{\beta}_1\right]$
$= (x_i - \bar{x}) c_i \sigma^2$
$= (x_i - \bar{x}) \frac{x_i - \bar{x}}{\sum_j (x_j - \bar{x})^2} \sigma^2$ (by (3))
$= \frac{(x_i - \bar{x})^2}{\sum (x_j - \bar{x})^2} \sigma^2$

(9) combining (5)~(8)

$\text{var}(e_i)$
$= \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right] - 2\text{cov}(\epsilon_i, \hat{y}_i)$ (by (5))

$$
\begin{aligned}
&= \sigma^2 + \sigma^2\left[\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right] - 2\left(E\left[\epsilon_i \bar{y}\right] + (x_i - \bar{x})E\left[\epsilon_i \hat{\beta_1}\right]\right) \text{ (by (6))} \\
&= \sigma^2 + \sigma^2\left[\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right] - 2\left(\tfrac{\sigma^2}{n} + \tfrac{(x_i-\bar{x})^2}{\sum(x_j-\bar{x})^2}\sigma^2\right) \text{ (by (7), (8))} \\
&= \sigma^2 - \sigma^2\left[\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right] \\
&= \sigma^2\left[1 - \tfrac{1}{n} - \tfrac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]
\end{aligned}
$$