## Handout 1: Deriving least squares estimates of $B_0$ and $B_1$ in simple regression

Mission: Minimize the residual sum of squares (RSS) with respect to our model parameters, $B_0$ and $B_1$. We use partial derivatives.

(1) $$\sum_{i=1}^{n} E_i^2 = \sum E_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - (B_0 + B_1 X_{1i}))^2 = \sum (Y_i - B_0 - B_1 X_{1i})^2$$

1. Differentiate Eq. 1 with respect to $B_0$

(2) $$\frac{\partial \sum E_i^2}{\partial B_0} = \sum (-1)(2)(Y_i - B_0 - B_1 X_{1i}) = -2\sum (Y_i - B_0 - B_1 X_{1i})$$

2. Differentiate Eq. 1 with respect to $B_1$

(3) $$\frac{\partial \sum E_i^2}{\partial B_1} = \sum (-X_{1i})(2)(Y_i - B_0 - B_1 X_{1i}) = -2\sum (X_{1i})(Y_i - B_0 - B_1 X_{1i})$$

3. Find the point on the function where the slope of the line tangent to the function (RSS) is zero. Thus, set Eqs. 2 and 3 to zero and solve simultaneously.

(4) $$0 = -2\sum (Y_i - B_0 - B_1 X_{1i})$$

(5) $$0 = -2\sum (X_i)(Y_i - B_0 - B_1 X_{1i})$$

4. Divide both sides of Eqs. 4 and 5 by -2

(4a) $$0 = \sum (Y_i - B_0 - B_1 X_{1i})$$

(5a) $$0 = \sum (X_{1i})(Y_i - B_0 - B_1 X_{1i})$$

5. Distribute summation operator in both equations and distribute $X_i$ in Eq. 5a

(6) $$0=\sum Y_i - nB_0 - B_1 \sum X_{1i}$$

(7) $$0=\sum X_{1i}Y_i - B_0 \sum X_{1i} - B_1 \sum X_{1i}^2$$

6. Rearrange Eqs. 6 and 7, and we get the "least squares normal equations."

(8) $$\sum Y_i = nB_0 + B_1 \sum X_{1i}$$

(9) $$\sum X_{1i}Y_i = B_0 \sum X_{1i} + B_1 \sum X_{1i}^2$$

7. Multiply Eq. 8 by $-\sum X_{1i}$ ; multiply Eq. 9 by N

(10) $$-\sum X_{1i} \sum Y_i = -nB_0 \sum X_{1i} - B_1 (\sum X_{1i})^2$$

(11) $$n\sum X_{1i}Y_i = nB_0 \sum X_{1i} + nB_1 \sum X_{1i}^2$$

8. Add Eqs. 10 and 11 together.

(12) $$n\sum X_{1i}Y_i - \sum X_{1i} \sum Y_i = nB_0 \sum X_{1i} - nB_0 \sum X_{1i} + nB_1 \sum X_{1i}^2 - B_1 (\sum X_{1i})^2$$

9. Reduce Eq. 12

$$n\sum X_{1i}Y_i - \sum X_{1i} \sum Y_i = nB_1 \sum X_{1i}^2 - B_1 (\sum X_{1i})^2$$

$$n\sum X_{1i}Y_i - \sum X_{1i} \sum Y_i = B_1 \left[ n\sum X_{1i}^2 - (\sum X_{1i})^2 \right]$$

10. Solve for $B_1$

$$B_1 = \frac{n\sum X_{1i}Y_i - \sum X_{1i}\sum Y_i}{n\sum X_{1i}^2 - (\sum X_{1i})^2} =$$

$$\frac{\sum (X_{1i} - \overline{X})(Y_i - \overline{Y})}{\sum (X_{1i} - \overline{X})^2} = \frac{Cov(X_1, Y)}{Var(X_1)} = r_{Y,X_1}\frac{s_Y}{s_{X_1}}$$

11. Solve for $B_0$; use Eq. 8 from the normal equations

$$\sum Y_i = nB_0 + B_1\sum X_{1i}$$

$$\sum Y_i - B_1\sum X_{1i} = nB_0$$

$$\frac{\sum Y_i}{n} - \frac{B_1\sum X_{1i}}{n} = B_0$$

$$B_0 = \overline{Y} - B_1\overline{X}$$