Assignment 4

(1) $$\frac{\partial Q}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2$$

$$= \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})(-1) \overset{\text{set}}{=} 0$$

$$\Rightarrow n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{i1} + \beta_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i$$

$$\frac{\partial Q}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2$$

$$= \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})(-x_{i1})$$

$$= -2 \sum_{i=1}^{n} (x_i y_i - \beta_0 x_{i1} - \beta_1 x_{i1}^2 - \beta_2 x_{i1} x_{i2}) \overset{\text{set}}{=} 0$$

$$\Longrightarrow \beta_0 \sum_{i=1}^{n} x_{i1} + \beta_1 \sum_{i=1}^{n} x_{i1}^2 + \beta_2 x_{i1} x_{i2} = \sum_{i=1}^{n} x_{i1} y_i$$

$$\frac{\partial Q}{\partial \beta_2} = \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})(-x_{i2})$$

$$= -2 \sum_{i=1}^{n} (x_{i2} y_i - \beta_0 x_{i2} - \beta_1 x_{i1} x_{i2} - \beta_2 x_{i2}^2 \overset{\text{set}}{=} 0$$

$$\Rightarrow \beta_0 \sum_{i=1}^{n} x_{i2} + \beta_1 \sum_{i=1}^{n} x_{i1} x_{i2} + \beta_2 \sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} x_{i2} y_i$$

In matrix form,

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1} x_{i2} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i1} x_{i2} & \sum_{i=1}^{n} x_{i2}^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \sum_{i=1}^{n} x_{i2} y_i \end{pmatrix}$$

$$X'X \qquad \beta \quad = \quad X'y$$

(2) $X'X\beta = X'y \Rightarrow (X'X)^{-1}X'X\beta = (X'X)^{-1}X'y$

$\Rightarrow \beta = (X'X)^{-1}X'y$, call the solution $\vec{\beta}$

(3) (a) $\frac{\partial Q}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2$

$= 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})(-1) \overset{set}{=} 0$

$\Rightarrow \sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{i1} + \cdots + \beta_k \sum_{i=1}^{n} x_{ik}$

(b) $\sum_{i=1}^{n} y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}$

(c) $= \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k) = \sum_{i=1}^{n} \hat{y}_i$

(d) $\sum_{i=1}^{n} \hat{\varepsilon}_i = \sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{y}_i = 0$

(e) From (b)

$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}(n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik})$

$\Rightarrow \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \cdots + \hat{\beta}_k \bar{x}_k$

(f) LS plane passes through $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k, \bar{y})$

(4) See Lecture Unit 6 (Least Squares), slides 57-59.

(5) (a) $\sum_{i=1}^{n}(y_i - 0)^2 = \sum_{i=1}^{n}(y_i - \bar{y} + \bar{y} - 0)^2$

$= \sum_{i=1}^{n}(y_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \bar{y})\bar{y} + \sum_{i=1}^{n}(\bar{y} - 0)^2$

$= \sum_{i=1}^{n}(y_i - \bar{y})^2 + 2\bar{y}\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})}_{0} + \sum_{i=1}^{n}(\bar{y} - 0)^2$

$= \sum_{i=1}^{n}(y_i - \bar{y})^2 + \sum_{i=1}^{n}(\bar{y} - 0)^2$

(b) $R^2 = \dfrac{n\bar{y}^2}{\sum_{i=1}^{n} y_i^2}$ Thinking of $\sum_{i=1}^{n}(y_i - \bar{y})^2$ as still unexplained

6 (a) $E(y_i) = \beta_0 + \beta_1 (x_i - \bar{x})$, so minimize

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 (x_i - \bar{x}))^2$$

$$\frac{\partial Q}{\partial \beta_0} = 2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 (x_i - \bar{x}))(-1)$$

$$= -2\left(\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{0}\right) \overset{set}{=} 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i = n\beta_0 \Rightarrow \beta_0 = \bar{y}$$

$$\frac{\partial Q}{\partial \beta_1} = 2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 (x_i - \bar{x}))(-(x_i - \bar{x}))$$

$$= -2\left(\sum_{i=1}^{n} (x_i - \bar{x}) y_i - \beta_0 \underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{0} - \beta_1 \sum_{i=1}^{n} (x_i - \bar{x})^2\right)$$

$$= 2\left(\beta_1 \sum_{i=1}^{n}(x_i - \bar{x})^2 - \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right) \overset{set}{=} 0$$

True because $\bar{y} \sum_{i=1}^{n}(x_i - \bar{x}) = 0$

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

(6b) $$\frac{\partial^2 Q}{\partial \beta_0^2} = \frac{\partial}{\partial \beta_0}\left(2n\beta_0 - 2\sum_{i=1}^{n} y_i\right)$$

$$= 2n > 0 \quad \text{concave up } \cup \text{ minimum}$$

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

$$= 2\sum_{i=1}^{n}(x_i-\bar{x})^2 > 0 \quad \text{concave up } \cup \text{ minimum}$$

(6c) They are the same, as expected.

7 (a) $$\frac{\partial Q}{\partial \beta_0} = \sum_{i=1}^{n} 2\left(y_i - \beta_0 - \beta_1(x_{i1}-\bar{x}_1) - \cdots - \beta_k(x_{ik}-\bar{x}_k)\right)(-1)$$

$$\stackrel{\text{set}}{=} 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i = n\beta_0 - \beta_1 \underbrace{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)}_{=0} - \cdots - \beta_k \underbrace{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)}_{=0}$$

$$= n\beta_0 \quad \text{so} \quad \hat{\beta}_0 = \bar{y}$$

(b) When all the predictor variables are at their sample mean values, the height of the least squares plane is $\bar{y}$.

8 See lecture unit 6 (Least Squares), slides 35-38.

9 $(X'X)^{-1} = (X'X)^{-\frac{1}{2}}(X'X)^{-\frac{1}{2}} = (X'X)^{-\frac{1}{2}\prime}(X'X)^{-\frac{1}{2}}$

so $v'(X'X)^{-1}v = v'(X'X)^{-\frac{1}{2}\prime}(X'X)^{-\frac{1}{2}}v$

$= \left[(X'X)^{-\frac{1}{2}}v\right]'(X'X)^{-\frac{1}{2}}v = z'z \geq 0$

And if $z'z = 0$ then $z = (X'X)^{-\frac{1}{2}}v = 0$

$\Longrightarrow (X'X)^{\frac{1}{2}}(X'X)^{-\frac{1}{2}}v = (X'X)^{\frac{1}{2}}0 = 0$

$\Longrightarrow v = 0$. Thus $v'\Sigma^{-1}v > 0$ for all vectors $v \neq 0$, satisfying the definition of positive definite. ▨

(10) (a) $n \times n$

(b) $H' = \left(X(X'X)^{-1}X'\right)' = X''(X'X)^{-1'}X'$

$= X\left((X'X)'\right)^{-1}X' = X(X'X)^{-1}X' = H$

(c) $H^2 = X(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}X' = X(X'X)^{-1}X' = H$

(d) $tr(H) = tr\left(\underbrace{X(X'X)^{-1}}_{\substack{A \\ n \text{ by } k+1}}\underbrace{X'}_{\substack{B \\ k+1 \text{ by } n}}\right) = tr\left(X'X(X'X)^{-1}\right)$

$= tr\left(I_{k+1}\right) = k+1$

(e) $HH = H \Rightarrow \underbrace{H^{-1}H}_{I}H = \underbrace{H^{-1}H}_{I}$

$\Longrightarrow H = I$

(f) $\text{Rank}(H) = k+1$ (Rank of product is minimum of ranks)

(g) $\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Hy$

(h) $(I - H)y = y - Hy = y - \hat{y} = \hat{\varepsilon}$

(10 i) $(I-H)' = I' - H' = I - H$

(j) $(I-H)(I-H) = (I-H) - (I-H)H$

$= (I-H) - H + HH = I - H - H + H$

$= I - H$

(k) $tr(I-H) = tr(I) - tr(H) = n - (k+1)$

$= n - k - 1$

(l) $(I-H)y = (I-H)(X\beta + \varepsilon)$

$= (I-H)X\beta + (I-H)\varepsilon$

$= X\beta - HX\beta + (I-H)\varepsilon$

$= X\beta - X\underbrace{(X'X)^{-1}X'X}_{I}\beta + (I-H)\varepsilon$

$= X\beta - X\beta + (I-H)\varepsilon$

$= (I-H)\varepsilon$

(11) $X'\hat{\varepsilon} = X'(I-H)y$

$$= (X' - X'H)y$$

$$= \left(X' - \underbrace{X'X(X'X)^{-1}}_{I}X'\right)y$$

$$= (X' - X')y = 0$$

Another way is

$$X'\hat{\varepsilon} = X'(y - \hat{y}) = X'y - X'\hat{y}$$

$$= X'y - X'X\hat{\beta}$$

$$= X'y - \underbrace{X'X(X'X)^{-1}}_{I}X'y$$

$$= X'y - X'y = 0$$

maybe this way is better.

12 (a) $\hat{\beta} = (X'X)^{-1}X'y = X^{-1}\underbrace{(X')^{-1}X'}_{I}y = X^{-1}y$

(b) $H = X(X'X)^{-1}X' = \underbrace{XX^{-1}}_{I}\underbrace{(X')^{-1}X'}_{I} = I_n$

(c) $\hat{y} = Hy = Iy = y$

(d) $\hat{\varepsilon} = y - \hat{y} = y - y = 0$

(e) The residuals are the differences between the height of the y points and the height of the regression plane. All the differences are zero, so the points are all on the plane.

(f) $n = k+1 = 1+1 = 2$

(g) Yes. Two points determine a line, & the best fitting line goes through both points.

13 See lecture unit 6 (least squares) slides 50-51.

14 (a) $$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

(b) $$X'X = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

(c) $$X'y = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

(d)

$$(X'X)^{-1} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{pmatrix}$$

15 See lecture unit 6 (Least Squares) slides 61-63.

16

a) $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

b) $X'X = \sum_{i=1}^{n} x_i^2$

c) $X'y = \sum_{i=1}^{n} x_i y_i$

d) $(X'X)^{-1} = \dfrac{1}{\sum_{i=1}^{n} x_i^2}$

e) $\hat{\beta} = (X'X)^{-1} X'y = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$

17 (a) $\frac{d}{d\beta_0} \sum_{i=1}^{n} (y_i - \beta_0)^2 = \sum_{i=1}^{n} 2(y_i - \beta_0)(-1) \overset{\text{set}}{=} 0$

$\Longrightarrow \sum_{i=1}^{n} y_i = n\beta_0 \Longrightarrow \beta_0 = \frac{\sum_{i=1}^{n} y_i}{n}$ and

$\hat{\beta}_0 = \bar{y}$

(b) $Q = \sum_{i=1}^{n} (y_i - \beta_0)^2 = \sum_{i=1}^{n} (y_i - \bar{y} + \bar{y} - \beta_0)^2$

$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + 2\sum_{i=1}^{n} (y_i - \bar{y})(\bar{y} - \beta_0) + n(\bar{y} - \beta_0)^2$

$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + 2(\bar{y} - \beta_0) \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})}_{=0} + n(\bar{y} - \beta_0)^2$

$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \beta_0)^2$

The first term does not involve β_0. The second term is non-negative, and equals zero if and only if $\beta_0 = \bar{y}$. So the unique minimum of Q over all β_0 occurs when $\beta_0 = \bar{y}$.

c) $X = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ (d) $X'X = n$

e) $X'y = \sum_{i=1}^{n} y_i$ (f) $(X'X)^{-1} = \frac{1}{n}$

(17g) $\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'y = \dfrac{\sum_{i=1}^{n} y_i}{n} = \bar{y} = \hat{\beta}_0$

(h) $\hat{y} = X\hat{\beta} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \bar{y} = \begin{pmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix}_{n \times 1}$

(18) (a) See lecture unit 6 (Least Squares) slide 71.

(b) This s^2 is a generalization of the usual s^2. It reduces to the usual s^2 in the special case of Question 17, where there are no predictor variables.