Answer Key for Problem Set 2

남 준우 교수

1.

(1) $\frac{1}{8}\sum_{i=1}^{8} GPA_i = 3.2125 = \overline{GPA}$, $\frac{1}{8}\sum_{i=1}^{8} STUDY_i = 25.875 = \overline{STUDY}$,

$\sum_{i=1}^{8}(STUDY_i - \overline{STUDY})^2 = 56.875$, $\sum_{i=1}^{8}(STUDY_i - \overline{STUDY})(GPA_i - \overline{GPA}) = 5.8125$

임을 표에서 구할 수 있다. 따라서 β_1, β_2의 추정치 b_1, b_2를 구하면,

$$b_2 = \frac{\sum_{i=1}^{8}(STUDY_i - \overline{STUDY})(GPA_i - \overline{GPA})}{\sum_{i=1}^{8}(STUDY_i - \overline{STUDY})^2} = \frac{5.8125}{56.875} \approx 0.102,\ b_1 \approx 0.568$$

(2) $5 \times b_2 = 0.510$

(3) $\hat{GPA}_i = 0.568 + 0.102 STUDY_i$, $e_i = GPA_i - \hat{GPA}_i$

GPA_i	$\hat{GPA}_i$	e_i
2.8	2.714	0.086
3.4	3.021	0.379
3	3.225	-0.225
3.5	3.327	0.173
3.6	3.532	0.068
3	3.123	-0.123
2.7	3.123	-0.423
3.7	3.634	0.066

$\sum_{i=1}^{8} e_i = 0$.

2. (1) 주어진 자료에서 $\bar{X} = \frac{1}{30}\sum_{i=1}^{30} X_i = 15$, 마찬가지로 $\bar{Y} = 80$.

$$b_2 = \frac{\sum_{i=1}^{30}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{30}(X_i - \bar{X})^2} = \frac{3600}{900} = 4,\ b_1 = \bar{Y} - b_2\bar{X} = 80 - 150 \times b_2 = 20.$$

따라서 $\hat{Y}_i = 20 + 4X_i$.

(2) 학습시간 이외에 성적에 영향을 주는 변수로 IQ, 결석횟수, 통계학 성적, 수업시간 집중도 등.

3. $\varphi \equiv \sum_{i=1}^{n}\left(Y_i - \hat{\mu}\right)^2$ 이라 하면 $\frac{d\varphi}{d\mu} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\mu}\right) = 0 \Rightarrow \hat{\mu} = \bar{Y}$ 이다.

(보충 설명)

여기서 기억해야 할 사실은 표본평균도 일종의 최소자승추정량이라는 것이다. 즉 독립변수가 상수항 하나 밖에 없는 모형에서 표본평균은 잔차항 제곱합을 최소화시키는 값이다.

이를 모집단의 개념으로 전환시키면 우리가 어떤 집단의 대표치로 모집단 평균(표본평균)을 선택한다면

(*) 모집단 평균(표본평균)은 $E\left(Y_i - c\right)^2$ 의 squared error 를 최소화시키는 값이다.

(**) 마찬가지로 표본평균은 $\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(Y_i - \hat{\mu}\right)^2$ 의 squared error 를 최소화시키는 값이다.(여기서 $1/n$ 은 없어도 괜찮음).

이와 같이 모집단 개념과 표본 개념은 1 대 1 의 관계가 있는데 모집단의 행위를 그대로 답습하여 표본추정량을 구축하는 이러한 행위를 analogue estimation 이라 한다.

4. $Y_i = \beta X_i + \varepsilon_i$

(1) β 의 추정량 b 를 구하라

$$Min\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - bX_i)^2 \equiv \phi$$

(*) $\frac{\partial\phi}{\partial b} = -2\sum_{i=1}^{n}(Y_i - bX_i)X_i = 0$ => $\sum_{i=1}^{n} e_i X_i = 0$: Normal Equation

$$\sum_{i=1}^{n} e_i X_i = \sum_{i=1}^{n}(Y_i - bX_i)X_i = \sum_{i=1}^{n} Y_i X_i - b\sum_{i=1}^{n} X_i^2 = 0$$

$$\therefore b = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} \neq \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

여기서는 정규방정식이 (*)식 하나만 도출된다는 점에 유의할 것.

(2) 절편이 없는 모형에서 normal equation 에서 $\sum_{i=1}^{n} e_i = 0$ 의 조건이 도출되지 않으므로, 일반적으로 $\sum_{i=1}^{n} e_i \neq 0$ 이다.

5. 수업시간에 도출한 정규방정식:

1) $\sum_{i=1}^{n} e_i = 0$ 2) $\sum_{i=1}^{n} X_i e_i = 0$

(1) $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} (a + bX_i + e_i) = \sum_{i=1}^{n} (a + bX_i) + \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (a + bX_i) = \sum_{i=1}^{n} \hat{Y}_i$

$$\therefore \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

(2)

$$\sum_{i=1}^{n} X_i Y_i = \sum_{i=1}^{n} X_i (a + bX_i + e_i) = \sum_{i=1}^{n} X_i (a + bX_i) + \sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i (a + bX_i) = \sum_{i=1}^{n} X_i \hat{Y}_i$$

$$\therefore \sum_{i=1}^{n} X_i Y_i = \sum_{i=1}^{n} X_i \hat{Y}_i$$