

답

1. (1) 코딩 가능

(2) Forward 했을 때 (Age, Income, Price) \rightarrow stepwise의 경계로 마찬가지로

Backward 했을 때 (HS, Female, Black) 이 나왔을 때

전체 가능한 모형들을 모두 비교하여 보았을 때 (Age, Income, Price) , (Age, Income, Black, Price)의 MSE가 작게 나왔고 또 (Age, Income, Price)가 더 유리해 보였다. (F값을 보았을 때).

$$(3) e_i = y_i - \hat{y}_i \quad PRESS = \sum_{i=1}^n \frac{e_i^2}{1 - h_{ii}} = \sum_{i=1}^n \left[\frac{e_i}{1 - h_{ii}} \right]^2$$

$$= \frac{e_i}{1 - h_{ii}} \Rightarrow \text{press 변수의 합} \Rightarrow \sum_{i=1}^n e_i = 43.1855$$

(4) 아무것도 제거 안했을 때, $\text{adj } R^2 = 0.2222$

R^2 student 값이 큰 두 개나 관찰치들

(29, 30) 번째 data들은 제외하였을 때 $\text{adj } R^2 = 0.5363$

2. 두외종양을 공변으로 포함하고 등분산성도, 선형성도 만족 못하며 설명력도 최적이다.
두 변수를 각각 log 변환한 뒤 회귀적합을 하면 F값도 유의적인 설명력도 좋다.

3. (1) ~~코딩 가능~~ (2종이면 $x_2=0, x_3=0$ / 3종이면 $x_2=1, x_3=0$ / 4종 이상이면 $x_2=0, x_3=1$)
(4종이면 $x_4=1$, 5종이면 $x_4=0$)

(2) 공변의 회귀계수는 546.18402로 경향이 1인 늘어난다면
공변은 546.18402 증가한다. \therefore 회귀계수의 95% 신뢰구간 구하면 됨.

$$(3) \text{대졸 관리직의 급여} \Rightarrow 8035.59763 + 546.18402 * x + 3144.03521 + 6893.531$$

$$= 546.18402 * x + 18063.16384$$

$$\text{고졸 비관리직의 급여} \Rightarrow 8035.59763 + 546.18402 * x$$

$$\text{회귀계수는 } 16027.56621$$