

제 2 문 (20점)

어느 지역에서 소득에 대하여 알아보기 위하여 소득(Y), 연령(X_1), 교육기간(X_2)을 조사하였다. 조사한 자료에서 Y 를 반응변수(종속변수), X_1 , X_2 를 설명변수(독립변수)로 하여 회귀분석을 실시한 결과가 다음과 같다. 각 회귀모형에서 $\epsilon_i \sim iid N(0, \sigma^2)$ 라고 가정한다.

모형 1 : $y_i = \beta_0 + \beta_1 x_1 + \epsilon_i$

Source	자유도	제곱합	평균제곱합	F-값
X_1	1	19.9558	19.95586	20.79
오 차	98	94.0783	0.95998	
전 체	99	114.0343		
	Estimate	Std. Error	t value	Pr > t
Intercept	1.86987	0.40242	4.65	<.0001
X_1	0.37845	0.08301	4.56	<.0001

모형 2 : $y_i = \beta_0 + \beta_2 x_2 + \epsilon_i$

Source	자유도	제곱합	평균제곱합	F-값
X_2	1	24.0222	24.0222	26.15
오 차	98	90.0121	0.9185	
전 체	99	114.0343		
	Estimate	Std. Error	t value	Pr > t
Intercept	0.8294	0.16217	5.11	<.0001
X_2	0.6505	0.04624	14.07	<.0001

모형 3 : $y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon_i$

Source	자유도	제 곱합	평균제 곱합	F-값
X_1, X_2	2	112.8308	56.41539	4547.07
오 차	97	1.2035	0.01241	
전 체	99	114.0343		

	Estimate	Std. Error	t value	Pr > t	VIF
Intercept	3.8133	0.05097	74.82	<.0001	0
X_1	-1.0114	0.01863	-54.29	<.0001	3.89775
X_2	1.4172	0.01638	86.52	<.0001	3.89775

- (1) Y 의 변동에 대하여 X_2 가 모형에 있을 때 X_1 이 추가됨으로써 설명이 증가하는 부분은 얼마인가?
- (2) 모형 3에서 독립변수(또는 설명변수) X_1 이 종속변수(또는 반응변수) Y 에 미치는 영향이 유의한지를 검정하고 X_1 이 Y 에 미치는 영향을 설명하라.
- (3) 모형 1에서의 β_1 추정값과 모형 3에서의 β_1 추정값의 부호가 다른 이유를 설명하라.
- (4) 추후 모형 3에서 성별을 추가하여 소득에 미치는 영향을 보려고 한다. 이를 위한 모형을 설정하고 모형을 설명하라.