

Regression Analysis

Assignment #2

타이어 마모도 자료

다음 자료는 실험실에서 수행된 타이어 마모에 관한 자료이다. 조사하고자 하는 특정 브랜드의 타이어를 동일한 차종에 장착하여 타이어의 마모도를 조사하였다. 분석의 목표는 달린 거리('Mileage', 단위 1000 마일)가 증가하면서 마모도('Depth', 단위 1/1000 인치)의 증감의 패턴을 파악하는 것이다. 아래는 그 결과이다.

Mileage (단위: 1000 miles)	Depth (단위: 1/1000 inches)
0	394.33
4	329.50
8	291.00
12	255.17
16	229.33
20	204.83
24	179.00
28	163.83
32	150.33

위 자료는 이클래스에 'tire.csv' 파일로 저장되어 있다. SAS를 이용하여 분석을 하고 아래의 질문에 답하라. 단, 분석에 사용된 SAS 코드 및 결과를 함께 제출토록 하라.

1. 변수 'Mileage'와 'Depth' 간의 산점도(scatterplot)를 그려라. 두 변수간의 관계가 어떠한지 적절히 설명하라.
2. 아래의 선형회귀분석을 수행하라. 추정된 추정식을 SAS의 결과로부터 기술하라.

$$(\text{Depth}) = \beta_0 + \beta_1(\text{Mileage}) + e$$

3. 위에서 적합한 식으로부터 얻은 표준화된 잔차(standardized residuals)을 이용한 잔차그림을 그려라. 이때, 세로축은 잔차, 가로축은 예측값(\hat{y}_i)으로 하여 그려라. 잔차그림을 통해 회귀식이 적당한지, 부적당하다면 이유를 설명하라.

4. 설명변수의 2차항을 넣어 아래의 회귀식을 적합하라. 추가적으로 넣은 2차항이 필요한지 여부를 t 검정을 통해 결과를 해석하라.

$$(\text{Depth}) = \beta_0 + \beta_1(\text{Mileage}) + \beta_2(\text{Mileage}^2) + e$$

5. 문제 4에서 수행한 회귀식으로부터의 잔차그림을 그려라. 이로부터 회귀식의 적절성을 판단하라.

시멘트 자료

다음 자료는 시멘트의 경도(hardness)에 대한 자료이다. 설명변수로는 ‘tricalcium aluminate (x_1)’, ‘tricalcium silicate (x_2)’, ‘tetracalcium alumino ferrite (x_3)’, ‘dicalcium silicate (x_4)’ 4개가 고려되었으며 각기 시멘트 조성에 차지하는 퍼센트로 측정되었다. 자료는 아래와 같다.

x_1	x_2	x_3	x_4	y
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

위 자료는 이클래스에 ‘cement.csv’ 파일로 저장되어 있다. SAS를 이용하여 분석을 하고 아래의 질문에 답하라. 단, 분석에 사용된 SAS 코드 및 결과를 함께 제출토록 하라.

1. 아래의 회귀분석을 수행하라. 다중공선성(multicollinearity)이 있는지 여부를 분산팽창계수(variance inflation factor; VIF)를 출력하여 확인하라.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + e$$

2. 다음의 회귀분석을 수행하여 아래 회귀식에서 얻은 R^2 값을 $1/(1 + R^2)$ 로 변환한 값이 위에서 얻은 첫번째 변수의 VIF 값과 동일함을 보여라.

$$x_1 = \alpha_0 + \alpha_1 x_2 + \alpha_2 x_3 + \alpha_3 x_4 + e$$

3. 위의 회귀결과를 이용하여 표준화된 잔차를 이용한 잔차그림을 그려라. 이때, 세로축은 잔차, 가로축은 예측값으로 하여 그려라. 잔차그림을 통해 회귀식이 적당한지, 부적당하다면 이유를 설명하라.
4. 아래의 회귀분석을 수행하라. 설명변수 x_3 과 x_4 가 모형에서 필요한지 여부를 1번의 결과를 함께 이용한 부분F검정(partial F test)을 수행하고 그 결과를 해석하라.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$$