**연구조사방법론 10장 과제**

2021310001 이성우

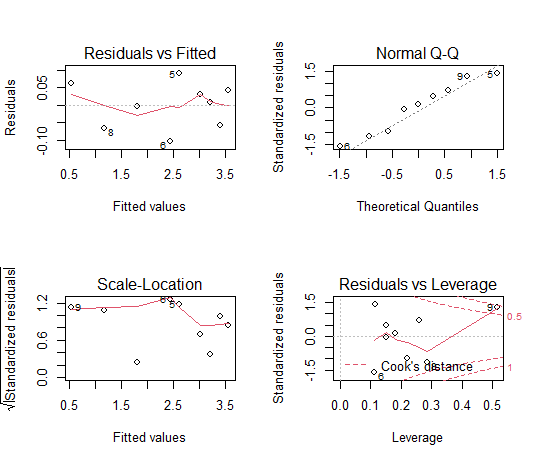
1. one\_comp.csv 를 이용하여 log(conc)를 종속변수, time 을 독립 설명 변수로 회귀분석한 결과의 회귀진단을 해보아라.

종속변수를 log scale변환하여 선형회귀분석을 해본 결과는 다음과 같다.

텍스트이(가) 표시된 사진

자동 생성된 설명

Multiple R-squared: 0.9962, Adjusted R-squared: 0.9957으로 1에 매우 가깝기 떄문에 선형성을 가진다고 말할 수 있다.



모형의 회귀진단 결과는 위와 같이 나타난다. 잔차도 퍼져있으면서 선형성을 나타내고 있고, Normal Q-Q line에 근접하게 결과가 나타나며 정규성을 띈다고 말할 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

정규성 검정을 위한 Shapiro test p-value가 0.796으로 정규성을 가진다고 말할 수 있다.

Birthwt 데이터를 회귀분석 하여라.

F-test를 이용한 Backward elimination으로 모형을 선택

𝑅^2 조정된 결정계수 로 모형을 선택

최종선택한 모형의 회귀 진단

최종 모형에 대한 회귀분석 결과 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

Bwt를 종속변수로 두고 나머지 변수를 독립변수로 두고 다중회귀식을 돌린결과는 위와 같다.

Multiple R-squared: 0.2424, Adjusted R-squared: 0.2087으로 그리 좋은 모델은 아닌 것 같다. Age와 ptl의 p-value가 0.05보다 높아 두 변수를 제거하는 Backward elimination을 진행한다.

텍스트이(가) 표시된 사진

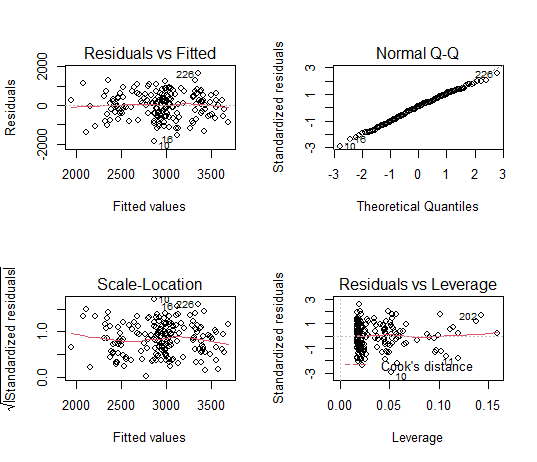
자동 생성된 설명

두 모델의 anova test의 p-value가 0.792로 큰 차이가 없음을 알 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

두번째 모델의 다중회귀 결과이다. Multiple R-squared: 0.2404로 첫번째 모델과 차이가 없지만Adjusted R-squared: 0.2154으로 첫번쨰 모델보다 소폭 상승하였다. 모든 독립변수가 의미있는 p-value를 가지므로 두번째 모델을 채택한다.



회귀진단의 결과이다. 잔차가 고르게 분포되어 있고 선형성을 띈다. Normal Q-Q line에 따라 잘 분포해 있고 Cook’s distance를 넘는 점도 없다.

텍스트이(가) 표시된 사진

자동 생성된 설명

Shapiro test결과 p-value가 0.05보다 커서 변수들의 분포가 정규성을 띈다고 말할 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명vif값이 모두 10을 넘지 않으므로 공선성을 가지지 않는다.