

06.01 모형진단과 수정

- 제대로 된 모델인지, 아니면 모델을 어떻게 수정할 것인지

모형 진단 : 사용된 데이터가 사용된 모형의 가정을 제대로 만족하고 있는지 확인 (모델의 가정을 진짜 가져가도 되는건지..)

1) 잔차 정규성 => 잔차가 정규성이 없다면, 기본 가정을 위배

1. 잔차가 정규분포가 아니다. (데이터 분포가 비선형일 경우, 발생 가능한 현상)
(1-2page, a,b)

2) 잔차와 독립 변수의 관계 => 관계가 있다면, 기본 가정을 위배

1. 잔차가 정규분포이더라도 (데이터가 비선형임에도), 잔차 - 독립변수 간 관계 다시 살펴봐야 함
(3-5 page, a-d)

3) 이분산성 => *epsilon*의 분산이 데이터에 따라 다르다면, 기본 가정을 위배

(6page)

4) 자기 상관 계수 => y_1, y_2, y_3 간 상관관계 시, 시계열 모델을 활용해야 함

5) 비선형 변형 => 종속-독립 간 비선형이면, 관계를 선형으로 바꿀 수 있도록, "**독립 자체를 비선형으로**" 변환! => 비선형 변환 방법 : 1) x, y log취하기 2) 제곱, 세제곱으로 feature 추가해 비선형으로 변환

(7-11p, a-d)

6) 범주형을 사용한 비선형성 => 독립변수를 강제로 범주형으로 끊어서 종속의 '비선형 변화'를 '비선형 상수항'으로 모형화해 모델 성능 향상 (11-15p, a-e)

7) 시간 독립변수의 변형 => 1) epoch 기준 지나온 시간(실수형)으로 변형 (시간값 => 반드시 스케일링!) => 2) 연/월/일/요일 등을 개별 feature column으로 분리! (상관관계 발견 가능) (16-17p, a-b)