투빅스 과제

11기 강수민

1.선형회귀분석

#1-1. 선형 회귀분석의 전제조건

1. 선형성 : 독립변수(x)와 종속변수(y)가 선형 관계를 띄어야 한다.
2. 정규성 : 종속변수(y)의 측정값들이 정규형 분포를 띄어야 한다. 🡪 하지만 이것을 분석하기 쉽지 않기 때문에 잔차를 이용. **잔차가 정규분포를 따른다. (plot그렸을 때 점들이 선 위에 있으면 정규성 가정을 만족한다고 하는 것!/**plotting으로 확인 🡪 plot\_model(regression1\_1, type = “diag”**)** #R언어 코드
3. 등분산성 : 잔차(오차의 분산)가 랜덤하게 흩어져 있는 것. (일정한 규칙이 없는 것 / plotting으로 확인 🡪 plot\_model(regression1\_1, type = “diag”)) #R언어 코드
4. 독립성 : 측정값들간에 자기상관관계가 없이 독립적이어야 한다. (Durbin-watson statistics=2 정도 나오면 독립적이라고 측정/dwtest) #R언어 코드
5. 비상관성 : 측정값들간에 상관관계가 없어야 한다. (VIF 10정도) #R언어 코드

#1-2. 회귀분석 결과 해석 - 모형 유의성, 적합성, 회귀계수 유의성 및 신뢰수준, 설명력

모형 유의성 측면에서는 p값을 먼저 참고해 보니 0.05보다 작은 값으로 유의하다고 할 수 있겠다.그리고 R2의 값에 따라 이 모델을 통해서 측정값들의 51.8%를 설명할 수 있다고 해석할 수 있다. 적합성 측면에서는 상단의 왼쪽 표에서 covariance 의 type이 nonrobust인 것으로 보아, 이 통계량은 이상치 값에 영향을 크게 받는다고 해석할 수 있겠다.

현재로서 어느 변수가 종속변수인지는 모르겠지만(const로 추정된다..!), 회귀계수들을 볼 때 유의한 변수로는 t통계량으로 계산한 p값을 보고(모델 전체의 유의성이 아니라 (이때는 F통계량) 경로계수 하나하나를 살펴보기 때문에 T통계량 사용) 유의한지 안한지를 참고하면 찾을 수 있겠다. 회귀계수들 중 p값이 0.05보다 작은 x2,x3,x4,x9가 유의한 회귀계수임을 알 수 있다.

또한 회귀계수들의 등분산성을 파악하기 위해서 durbin watson의 값을 보니 약 2정도로 측정값들이 독립적이라고 할 수 있다.