# 제5장 회귀모형의 선택

## 5.1 모형선택

■ 회귀분석: 주어진 설명변수들과 반응변수 간의 관계를 구하는 것

■ 회귀분석의 목적

• 적합된 회귀모형으로부터 설명변수와 반응변수 간의 관계를 파악

• 적합된 회귀모형에 새롭게 주어진 설명변수 값을 대입하여 반응변수값을 예측

■ 모형선택의 필요성

• 회귀분석의 역설

상기 회귀분석의 목적을 달성하기 위해 이용할 수 있는 모든 설명변수들을 전부 다 사용하면 가장 정확한 회귀모형을 얻을 수 있을 것으로 생각되지만 실제로는 사용할 수 있는 설명변수들 중 일부만 사용하여 구축한 회귀모형이 더 좋은 경우가 흔히 있다.

• "어떤 모형이 다른 모형보다 더 좋은가?"에 대한 기준

→ 흔히 "예측오차"라는 값으로 나타낸다.

## 5.2 변수선택의 기준

■ 예비모형의 구축 단계에서는 고려된 설명변수 전부가 사용된다.
총 설명변수의 개수$=k-1$ → $\{X_1, \cdots, X_{k-1}\}$

• 최대모형(maximal/full/saturated model): $k-1$개 설명변수 전부 사용

• 현재모형(current/postulated model): $p-1$개 설명변수 사용. $(p \le k)$

• 최소모형[영모형](minimal/null model)

: 설명변수가 전혀 사용되지 않고 반응변수 자체의 변동만을 고려한 모형

■ 변수선택 기준

• $R_p^2$ (결정계수) • $R_{ap}^2$ (수정된 결정계수)

• Mallows' $C_p$ • Allen's $PRESS_p$