

6강. 회귀분석과 공분산분석

◆ 담당교수 : 백재욱 교수

■ 정리하기

- 회귀분석은 독립변수와 종속변수 간의 관계를 함수식으로 표현하여 살펴보는 분석방법이다. 독립변수의 수가 1개인 경우 단순회귀가 되고, 2개 이상인 경우 다중회귀가 된다. 독립변수와 종속변수 간의 관계를 선형식으로 나타내면 선형회귀가 된다. 다중선형회귀 모형은 $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \epsilon$ 과 같다.
- 산점도란 n 쌍의 관측값 (x_i, y_i) 를 이차원 좌표평면에 점으로 표시한 것으로 두 변수 간의 관계를 파악하는 데 유용하다.
- 단순선형회귀모형은 $y = \beta_0 + \beta_1x + \epsilon$ 와 같다. 최소제곱법으로 회귀계수를 추정하면 와 같다.
- $R^2 = \frac{SS_R}{SS_r}$ 이 크면 클수록 주어진 자료가 회귀선에 적합하다.
- 회귀계수 β_1 은 $f_0 = \frac{ss_R/1}{SS_E(n-2)} = \frac{MS_R}{MS_E}$ 이면 통계적으로 유의하다. 이 경우 β_1 에 대한 $100(1-\alpha)\%$ 신뢰구간을 추정할 필요가 있다.
- 두 개의 독립변수 A와 B가 이산형이고 종속변수가 연속형인 경우 이원배치를 적용하고 분산분석을 실시한다. 두 개의 독립변수 중 하나(B)가 연속형이면 이를 공변수라고 하고, 공분산분석을 실시한다. 이 경우 모형은 $y_{ij} = \mu' + \alpha_i + \beta x_{ij} + \epsilon_{ij}$ 와 같다. 이 모형은 일원배치법 모형인 $y_{ij} = \mu' + \alpha_i + \epsilon_{ij}$ 와 단순선형회귀모형인 $y_{ij} = \mu' + \alpha_i + \beta x_{ij} + \epsilon_{ij}$ 이 결합된 모형이다.
- 앞의 모형의 경우 데이터의 총변동은 A로 인해 생기는 변동, B로 인해 생기는 변동, 잔차변동의 합으로 분할된다. A의 효과(주효과)가 있는지는 F검정을 통해 확인할 수 있다.