

11강. 선형혼합모형소개

■ 주요용어

용어	설명
변량요인	요인의 모든 수준으로 구성된 모집단에서 랜덤추출된 것으로 생각될 수 있는 인자이다. 변량요인의 대표적인 예로 군집자료에서 각 군집을 들 수 있다. 변량요인의 모든 수준은 자료에 포함되어 있지 않을 수 있어 변량요인의 수준 간 비교는 연구의 주요 관심이 아니다. 연구자는 변량요인 모든 수준으로 구성된 모집단에 대하여 연구의 주요 가설을 추론하고자 한다.
고정효과	선형혼합모형에서 종속변수와 설명변수(범주형 또는 연속형 설명변수)의 관계를 나타내는 회귀계수로 미지의 상수값으로 가정한다.
변량효과	변량요인의 수준 값을 나타내는 관측 불가능한 확률변수로 표현되는 값으로 선형혼합모형에서 군집(또는 개체)내 상관성과 고정효과와의 변동성을 모형화 한다. 변량효과가 있는 대표적인 모형으로 변량절편(random intercept) 모형과 변량기울기(random coefficient) 모형이 있다.
가능도함수	가정한 모형의 확률밀도함수에서 표본관측값을 대입한 모수의 함수
가능도비검정통계량 (LR)	$LR = -2\log\left(\frac{L_{H_0}}{L_{H_1}}\right)$, 여기서 L_{H_0} 와 L_{H_1} 는 각각 귀무가설과 대립가설에서의 최대가능도 값이다. 단 $H_0 \subset H_1$ 성립. $LR \sim_0^H \chi^2(df)$ 이며 df 는 카이제곱 분포의 자유도로 $df = (H_1 \text{에서 추정된 모수의 수}) - (H_0 \text{에서 추정된 모수의 수})$.

정리하기

■ 요약하기

1. 자료의 유형

1) 군집자료(clustered data)

▷ 각 개체에서 한 번씩 관측되고 이 개체들은 집단 구성하거나 또는 상위 집단에서 지분된다.

2) 반복측정자료(repeated-measures data)

▷ 반응변수가 반복측정요인의 수준에 따라 한 개체에서 두 번 이상 측정된 자료

3) 경시적 자료(longitudinal data)

▷ 반응 변수가 한 개체에서 시간의 흐름에 따라 두 번 - 상대적으로 시간 간격이 길다- 이상 측정된 자료

2. 요인과 관련 효과들

1) 변량요인

▷ 인자의 모든 수준으로 구성된 모집단에서 랜덤추출된 것으로 생각될 수 있는 인자.

2) 고정효과

▷ 종속변수와 설명변수의 관계를 나타내는 회귀계수

3) 변량효과

▷ 변량요인의 수준 값을 나타내는 관측 불가능한 확률변수로 표현되는 값

3. 선형혼합모형(LMM: Linear Mixed Models)의 기술

$$Y_{ti} = \beta_1 X_{ti}^{(1)} + \dots + \beta_p X_{ti}^{(p)} \} \text{ 고정효과 모형화} \\ + u_{1i} Z_{ti}^{(1)} + \dots + u_{qi} Z_{ti}^{(q)} + \epsilon_{ti} \} \text{ 변량효과모형화}$$

$$u_i = (u_{i1}, \dots, u_{qi})' \sim N(0, D)$$

$$\epsilon_i = (\epsilon_{1i}, \dots, \epsilon_{n_i})' \sim N(0, R_i)$$

$$\text{cov}(u_i, \epsilon_i) = 0.$$

4. LMM의 추정

1) 최대가능도추정

▷ 로그가능도함수

$$l(\beta, \theta) = -\frac{1}{2} \sum_{i=1}^m \left\{ (y_i - X_i \beta)' V_i^{-1} (y_i - X_i \beta) + \log [\det(2\pi V_i)] \right\}$$

▷ 최대가능도추정값: $(\hat{\beta}, \hat{\theta})$

- $l(\beta, \theta)$ 를 최대가 되게 하는 (β, θ) 값

2) θ 의 제한최대가능도추정

▷ 제한가능도함수(restricted likelihood function): 가능도함수에서 회귀모수 β 의 추정으로 인한 정보의 손실을 보정한 함수

5. 모형의 선택

1) 회귀모수에 대한 근사 t -검정과 근사 F -검정

2) 지분관계 모형들에서 가능도비 검정

2) 공분산 모수 θ 에 대한 제한가능도비 검정

▷ 귀무가설의 형태에 따라 가능도비 검정통계량의 대표본 근사분포가 다름

- Case 1: 귀무가설의 만족하는 공분산 모수가 모수공간의 경계가 아닐 때

◦ 일반적인 우도비 검정통계량과 같은 방법 근사분포 사용

- Case 2: 귀무가설의 만족하는 공분산 모수가 모수공간의 경계가 아닐 때

◦ Single Random effect: $LR \approx 0.5 + 0.5\chi^2(1)$

◦ Two random effects: $LR \approx 0.5\chi^2(1) + 0.5\chi^2(2)$

◦ LR 은 제한가능도함수에서 구한 가능비 검정 통계량.

	과제하기
--	-------------

구분	내용
과제 주제	<p>1. 다음 기술이 타당하면 O 틀리면 X로 표시하시오</p> <p>① 선형모형을 사용하여 상관된 정규분포 자료를 올바르게 분석할 수 있다.</p> <p>② 선형혼합모형에서는 고정효과와 변량효과를 선형관계식으로 표현한다.</p> <p>③ 변량효과는 미지의 상수값으로 일반적으로 추론의 대상이 된다.</p> <p>④ 선형혼합모형의 분산-공분산 모수의 추정은 편의 측면에서 최대가능도추정보다 제한최대가능도추정이 우수하다.</p> <p>⑤ 변량효과에 대한 가능도비 검정은 제한가능도 함수를 사용한다.</p> <p>2. 선형혼합모형을 개별 관측값 표현식 및 행렬 표현식으로 기술하라.</p> <p>3. 변량효과 검정법을 요약하라.</p>
목적	11주차 강의 내용을 복습하고, 선형혼합모형의 특성과 기술방법, 모형 추정과 검정법에 대한 이해를 심화한다.
제출 기간	11주차 강의 후 1주 후 토요일 밤 10시까지
참고 자료	교재와 강의자료를 참고하기 바람
기타 유의사항	