

12강. LMM 2-수준 군집자료분석

■ 주요용어

용어	해설
하향식 모형 구축	모형을 선택할 때 고정 효과 모수를 가능한 많이 가지는 모형에서 출발하여 변량효과의 선정, 오차의 공분산구조의 선정, 그리고 고정효과 축소 순서로 모형을 구축해 가는 방법이다.
이분산 (heteroscedasticity) 모형	오차의 분산이 범주형 설명 변수의 수준에 따라 다른 모형
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 $H_0 \subset H_1$ 성립한다. 고정효과의 유의성을 검정할 때 사용한다. 검정통계량의 분포는 $LR \sim_0^H \chi^2(df)$ 이며 df 는 카이제곱 분포의 자유도로 $df = (H_1 \text{에서 추정된 모수의 수}) - (H_0 \text{에서 추정된 모수의 수})$.
급 내 상 관 계 수 (ICC:intraclass correlation coefficient)	군집 내 반응변수들 간의 유사성 또는 동질성을 나타내는 척도로 사용된다. j 번째 군집의 i 와 i' 번째 반응변수에 대하여 급내상관계수는 아래와 같이 정의된다. 단 $i \neq i'$, $ICC = Corr(Y_{ij}, Y_{i'j})$.
조건부 잔차와 주변부 잔차	조건부 생잔차: $r_{cj} = y_j - X_j\hat{\beta} - Z_j\hat{u}_j$ 주변부 생잔차: $r_{mj} = y_j - X_j\hat{\beta}$

정리하기

■ 요약하기

1. 자료의 특성 파악하기: 기술통계량 및 그래프 상용
2. 모형의 구축
 - 2.1 하향식(Top-down) 모형 구축(Verbeke and Molenberghs 2000, Ch.9)
 - ▷ 짝 찬 평균모형($E(Y_{ij}) = X_i\beta$)에서 출발
 - 교호작용을 포함하여 가능한 많은 수의 공변량을 모형 포함 시킴
 - 종속변수에 내포된 체계적인 변동성이 잘 설명되도록 함
 - ▷ 변량효과의 구조(D)를 선택
 - 제한우도비 검정으로 선택한 변량효과의 유의성을 검정
 - ▷ 오차의 공분산(R_i) 구조를 선택
 - 제한우도비 검정(검정 모형간에 지분관계 성립할 때)
 - 정보량 기준(지분관계가 성립하지 않을 때)
 - ▷ 평균 모형을 축소
 - F -검정, T -검정, 우도비 검정 등으로 고정효과 모수의 유의성을 검정

2.2 변량절편모형

j 번째 litter의 i 번째 출생체중 Y_{ij} 에 대하여 다음의 변량 절편모형을 가정한다.

$$Y_{ij} = \beta_0 + \beta_1 \times Tr1_j + \beta_2 \times Tr2_j + \beta_3 \times Sexf_{ij} + \beta_4 \times litsize_j + \beta_5 \times Tr1_j \times Sexf_{ij} + \beta_6 \times Tr2_j \times Sexf_{ij} + u_j + e_{ij}$$

여기서, $u_j \sim^{iid} N(0, \sigma_{litter}^2)$, $e_{ij} \sim^{iid} N(0, \sigma^2)$, $cov(u, e) = 0$.

2.3 변량효과의 유의성 검정

$$H_0 : \sigma_{litter}^2 = 0 \text{ vs. } H_1 : \sigma_{litter}^2 > 0$$

- 제한가능도비 검정통계량과 분포

$$LR = -2 \log \left(\frac{L_{H_0}}{L_{H_1}} \right) \sim_0^H 0.5 \times \chi_0^2 + 0.5 \times \chi_1^2$$

여기서 L_{H_0} 와 L_{H_1} 는 각각 귀무가설과 대립가설에서 구한 제한최대가능도 값이고 χ_k^2 는 자유도가 k 인 카이제곱 분포함수임.

2.4 오차분산구조의 선택

2.4.1(모형1) 오차 분산이 등분산인 모형

$$u_j \sim^{iid} N(0, \sigma_{litter}^2), \quad e_{ij} \sim^{iid} N(0, \sigma^2), \quad cov(u, e) = 0.$$

2.4.2(모형2) 오차 분산이 이분산인 모형1

$$\begin{aligned} u_j &\sim^{iid} N(0, \sigma_{litter}^2), \\ \text{High: } e_{ij} &\sim^{iid} N(0, \sigma_h^2), \\ \text{Low: } e_{ij} &\sim^{iid} N(0, \sigma_l^2), \\ \text{Control: } e_{ij} &\sim^{iid} N(0, \sigma_c^2), \\ cov(u, e) &= 0. \end{aligned}$$

2.4.3(모형3) 오차 분산이 이분산인 모형2

$$\begin{aligned} u_j &\sim^{iid} N(0, \sigma_{litter}^2), \\ \text{High/Low: } e_{ij} &\sim^{iid} N(0, \sigma_{h/l}^2), \\ \text{Control: } e_{ij} &\sim^{iid} N(0, \sigma_c^2), \\ cov(u, e) &= 0. \end{aligned}$$

3. 연구가설의 검정

▷ 최종 분석 모형에서 연구가설은

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0 \text{ or } \beta_2 \neq 0$$

▷ 연구가설에 대한 우도비 검정통계량 LR 과 분포

$$LR = -2 \log \left(\frac{L_{H_0}}{L_{H_1}} \right) \sim_0^H \chi_k^2$$

여기서 L_{H_0} 와 L_{H_1} 는 각각 귀무가설과 대립가설에서 구한 최대가능도값이고 χ_k^2 는 자유도가 k 인 카이제곱 분포함수임.

4. 모수 추정치의 해석

4.1 고정효과 모수의 해석

▷ 고정효과 모수추정치(REML추정방법 사용)와 해석

① 성별과 배의 크기가 보정되었을 때 폭로약물의 용량이 증가할 수록 출생체중의 감소도 커지는 것으로 나타남.

- 세 군간의 상호비교를 위한 사후 검정을 실시할 수 있음.
- 모형의 단순화(parsimony) 측면에서 용량 수준을 연속형 변수로 하여 분석을 시도할 수 있음.

② 용량과 배의 크기가 보정되었을 때 암컷은 수컷에 비하여 출생체중이

유의하게 작음.

③ 용량과 성별이 보정되었을 때 배의 크기가 증가할 수록 출생체중은 유의하게 감소함.

▷용량군에 대한 사후 검정

결과 해석: 성별과 배의 크기를 보정하였을 때,

① 저용량군(Low)과 고용량군(High)의 평균출생체중은 대조군(Control)과 비교하여 각각 0.43g, 0.86g 적으며, 이들 차이는 유의수준 5%에서 유의함.(각각 유의확률은 0.0238, 0.0003).

② 고용량군의 평균출생체중은 저용량과 비교하여 0.43g 적었고, 이들 차이는 유의수준 5%에서 경계수준에서 유의하였음(유의확률은 0.0569).

4.2 공분산 모수의 해석

▷ 급내상관계수(ICC:intraclass correlation coefficient)

- 군집내 반응변수들 간의 유사성 또는 동질성을 나타내는 측도로 사용된다. j 번째 군집의 i 와 i' 번째 반응변수에 대하여 급내상관계수는 아래와 같이 정의된다.단 $i \neq i'$,

$$ICC = Corr(Y_{ij}, Y_{i'j}) = \frac{\sigma_{litter}^2}{\sigma_{litter}^2 + \sigma^2} .$$

- 성질

- $0 \leq ICC \leq 1$

- ICC 값이 크면 급내 반응변수들의 동질성이 큼.

5. 모형의 진단

5.1 잔차 진단

▷ 조건부 잔차

$$r_{ij}^c = Y_{ij} - X_{ij}\hat{\beta} - Z_j\hat{u}_j$$

▷ 주변부 잔차

$$r_{ij}^m = Y_{ij} - X_{ij}\hat{\beta}$$

▷조건부 잔차를 이용한 오차분포 가정에 대한 검토

- 정규성에 대한 검토
- 등분산성에 대한 검토
- 이상치에 대한 검토

▷조건부 잔차의 종류

- 조건부 생잔차(raw conditional residuals)

$$r_{ij}^c = Y_{ij} - X_{ij}\hat{\beta} - Z_j\hat{u}_j$$

- 조건부 스튜던트화 잔차(conditional studentized residuals)

$$\frac{r_{ij}^c}{\sqrt{\widehat{Var}(r_{ij}^c)}}$$

- 피어슨 잔차(Pearson type residuals)

$$\frac{r_{ij}^c}{\sqrt{\widehat{Var}(y_j|u_j)}}$$

과제하기	
구분	내용
과제 주제	<p>1. 생쥐의 출생체중 자료분석에서 모형1과 모형3을 제한최대우도 검정법으로 비교하고자 한다. 다음 물음에 답하시오.</p> <p>① 모형1과 모형3을 기술하시오.</p> <p>② 모형1과 모형3을 비교하는 가설과 제한가능도비 검정통계량을 기술하고 검정통계량의 분포를 제시하시오.</p> <p>③ 제한가능도비 검정통계량의 관측값을 제시하고, 유의확률을 구하고 유의수준 5%에서 검정결과를 기술하시오.</p> <p>2. Model 3.1 에 대하여 다음 물음에 답하시오.</p> <p>① Model 3.1을 기술하시오.</p> <p>② treatment (용량군)의 유의성을 Type III F-검정을 사용하여 검정하시오. 이 때 귀무가설과 대립가설을 무엇인가?</p> <p>③ sex (성)의 유의성을 Type III F-검정을 사용하여 검정하시오. 이 때 귀무가설과 대립가설을 무엇인가?</p> <p>3. 다음 기술이 타당하면 O 틀리면 X로 표시하시오</p> <p>① 고정효과 모수의 유의성검정을 가능도비 검정법으로 할 때 제한가능도함수를 사용한다.</p> <p>② 변량절편모형에서 급내상관계수는 0과 1사이의 값을 가진다.</p> <p>③ 급내상관계수의 값이 클수록 급내 이질성은 증가한다.</p> <p>④ 변량효과 u_j의 최량선형불편예측량(BLUP)은 u_j를 고정효과로 가정하고 추정된 값보다 작게 추정하는 경향이 있다.</p> <p>⑤ 변량효과 모형의 오차분포에 대한 진단에는 조건부 잔차보다 주변부 잔차를 사용하는 것이 합리적이다.</p> <p>4. R package 에서 제공하는 Pearson-type 조건부 잔차를 사용하여 오차분포 가정에 대한 잔차진단을 실시하고 결과를 해석하시오.</p>
목적	12주차 강의 내용을 복습하고, 변량절편 모형의 구축과 모형의 선택에 대한 이해를 심화 하기 위함. 또한 모수 추정치의 의미와 설명법에 대한 이해도를 높이기 위함.
제출 기간	12주차 강의 후 1주 후 토요일 밤 10시까지
참고 자료	교재와 강의자료를 참고하기 바람