

변량 모형

서울시립대 통계학과

2021-03-22

차 례

변량 모형	vii
제 1 절 고정효과	vii
제 2 절 임의효과 (Random Effects)	viii
Literature	xi

서문

이번 강의에서는 변량 모형(random effect models)에 대하여 알아봅니다.

변량 모형

제 1 절 고정효과

앞 장에서 하나의 요인있는 일원배치 모형에 대한 추론에 대하여 알아보았다.

$$x_{ij} = \mu + \alpha_i + e_{ij} \text{ where } e_{ij} \sim N(0, \sigma_e^2) \quad (1)$$

여기서 오차항 e_{ij} 는 모두 독립이다.

일원배치 모형 (??) 에서 전체 평균 μ 와 처리수준의 효과를 나타내는 $\alpha_1, \alpha_2, \dots, \alpha_a$ 는 모두 고정된 값을 가지는 모수(parameter)이다. 식 (??) 의 오른쪽 항들 중에서 확률변수는 오차항 e_{ij} 이 유일하다.

처리수준의 효과 α_i 들이 모수라는 것은 다음과 같은 의미를 가진다.

만약 새로운 실험에서 동일한 실험단위(experiment unit)에 동일한 처리를 적용하면 평균 효과는 α_i 로 일정하다.

예를 들어 예제 3.1에서 수행한 실험을 다른 회사에서 동일한 납품업체의 원단(동일한 실험 단위와 처리)을 가지고 새로운 실험을 하면 평균적인 효과는 예제 3.1과 동일하다는 가정을 할 수 있다. 또한 예제 4.1 에 대한 실험도 동일한 돼지 품종과 사료를 사용하여 새로운 실험을 수행할 수 있고 처리 효과도 전의 실험과 다르지 않다고 가정할 수 있다. 즉, 처리라는 것이 기술적인 의미를 지니고 있어 반복하여 재현할 수 있는 효과이다. 이러한 고정된 모수로서의 효과를 **고정 효과(fixed effect)**라고 부른다.

더 나아가 고정효과를 가지는 모형에서는 고정효과를 추정하는 것이 주 목적이다.

제 2 절 임의효과 (Random Effects)

이제 효과가 다른 의미를 가지는 실험을 생각해 보자.

2.1 예제 3.3(교과서)

화학약품 회사에서는 매년 원자재의 수백 개의 배치(batch)를 정제하여 순도가 높은 화학약품을 만든다. 품질 관리를 위하여 수백 개의 배치들 중에서 5개를 랜덤하게 선택하고 배치당 3개의 시료를 채취한 후에 순도를 측정하였다.

실험의 목적은 품질 관리이며 배치마다 순도가 크게 다르면 문제가 생긴다. 따라서 실험의 목적은 배치 간의 변동과 배치 내의 변동을 알아보는 것이다.

2.2 Test-Retest

새로 개발된 CT 로 만든 영상에 근거하여 의사들이 암의 단계를 파악하는 실험을 진행하였다. 일단 5명의 암환자들에서 CT 영상을 촬영하였다. 다음으로 15명의 의사를 임의로 추출하고 5명의 CT 영상을 본 후 암의 진행 단계를 판단할 수 있는 점수를 매기도록 하였다.

실험의 목적은 CT 영상에 근거한 진단이 의사들간에 잘 일치하는지를 알아보는 실험이다.

2.3 학교간의 성적 비교

학교 간에 성적의 차이를 알아보기 위하여 서울에 있는 603개의 학교중 20개의 학교를 임의로 추출하고 추출된 학교에 속한 모든 6학년 학생들에게 과학시험을 보게하여 점수를 얻었다.

이러한 자료에서 학생들의 성적은 모두 같지 않을 것이 당연하며 가장 점수가 낮은 학생부터 높은 학생까지 점수의 변동(variation)이 존재한다. 변동의 요인은 무엇일까? 학생의 개인의 차이(예:학생의 지능, 노력 정도, 학습 환경)도 변동의 요인이지만 또한 학교의 차이(교사, 거주 환경)도 변동의 요인이 될 수 있다.

위의 예제에서 **배치**, **의사**, **학교**는 고정 효과를 가정한 실험에서 고려하는 요인과는 성격이 틀리다. 5개의 배치들은 수백 개의 배치들에서 임의로 추출 되었으며 5명의 의사들은 다수의 의사들 중 임의로 추출되었다. 603개 초등학교의 모집단에서 20개의 학교가 임의로 추출되었다. 이 때 배치 간의 차이, 의사간의 차이 또는 학교

간의 차이는 잘 설계된 실험의 처리에 대한 고정 효과와는 다르다. 위의 예제에서 요인간의 차이는 모집단의 구성 단위들의 변동이라고 할 수 있다.

위에서 언급한 3개 예제는 실험의 목적이 선택된 수준들의 효과의 기술적인 비교가 아니라 모집단이 가지고 있는 여러 가지 변동(variance)에 대하여 추론하는 것이다.



같은 학교에 다니는 학생들은 주거 환경, 교사 등 공통적인 요인에 의하여 영향을 받는다고 가정할 수 있다. 따라서 같은 학교에 다니는 학생들의 성적이 독립이 아닐 수도 있다. 동일한 의사가 판단한 5명의 환자에 의 평가도 독립이라고 하기 어렵다.

이렇게 요인의 변동이 고정효과처럼 기술적인 변동이 아니라 모집단의 단위들에 대한 변동 효과를 **임의효과(random effect, 변량)**이라고 한다. 임의효과를 가진 일원배치 변량모형은 다음과 같이 나타낼 수 있다.

$$x_{ij} = \mu + \alpha_i + e_{ij} \text{ where } \alpha_i \sim N(0, \sigma_A^2), e_{ij} \sim N(0, \sigma_E^2) \quad (2)$$

위의 식에서 $\alpha_1, \alpha_2, \dots, \alpha_a$ 를 임의 효과라고 부르며 서로 독립인 확률 변수로서 분포는 $N(0, \sigma_A^2)$ 을 따른다. 또한 임의 효과 α_i 와 오차항 e_{ij} 은 서로 독립이다.

Literature

Here is a review of existing methods.