선형혼합모형 III - 최대가능도 추정법

서울시립대 통계학과 이용희

FALL 2019 학부

1 선형회귀모형과 최대가능도 추정

1.1 선형회귀모형

반응변수가 y이고 k개의 독립변수 $(x_1, x_2, \cdots, x_k)$가 있다고 가정하고 표본의 크기 n인 자료가 얻어지면 선형회귀식을 행렬로 다음과 같이 표현할 수 있다.

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i \\
&= \boldsymbol{x}_i^t \boldsymbol{\beta} + e_i, \quad i = 1, 2, \cdots, n
\end{aligned}
$$

위의 식을 다시 표현하면 다음과 같이 쓸 수 있다. 이와 같은 회귀모형을 선형중회귀모형이라 부르며. 각 개체에 대한 모형의 방정식을 하나의 식으로 표현하면 다음과 같다.

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1, & x_{11}, & \cdots & x_{1k} \\ 1, & x_{21}, & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1, & x_{n1}, & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_n \end{bmatrix}
$$

위의 모형식을 벡터와 행렬을 이용하여 표시하면 다음과 같다.

$$
\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{1}
$$

여기서 회귀분석의 오차항의 가정을 살펴보면 오차항이 서로 독립이고 동일한 분산을 갖는다. 즉, 오차항은 다음의 분포를 따른다. 즉, $\boldsymbol{e} \sim (0, \sigma_e^2 \boldsymbol{I}_n)$. 관측값 벡터 $\boldsymbol{y}$의 평균과 분산을 보면

$$
E(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}, \quad Var(\boldsymbol{y}|\boldsymbol{X}) = \sigma_e^2 \boldsymbol{I}_n
$$

여기서 오차항이 정규분포를 따른다면 ($\boldsymbol{e} \sim N(0, \sigma_e^2 \boldsymbol{I}_n)$ 관측값 벡터 $\boldsymbol{y}$ 또한 정규분포를 따른다

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma_e^2 \boldsymbol{I}_n)$$

1.2 최소제곱 추정

위의 선형 모형 가정하에서, 최소제곱 추정량 $\hat{\boldsymbol{\beta}}$ (least square estimator)는 다음과 같이 오차제곱합 (Error Sum of Sqaures) SSE 를 최소로 하는 추정량이다.

$$\begin{aligned} SSE(\boldsymbol{\beta}) &= \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^t \boldsymbol{\beta})^2 \\ &= (\boldsymbol{y} - \boldsymbol{X\beta})^t (\boldsymbol{y} - \boldsymbol{X\beta}) \\ \hat{\boldsymbol{\beta}} &= \arg\min_{\boldsymbol{\beta}} SSE(\boldsymbol{\beta}) \end{aligned}$$

따라서 $\hat{\boldsymbol{\beta}}$는 오차제곱합을 최소로 하는 계수 벡터이다. 오차제곱합에 최소제곱 추정량을 사용하면 이를 잔차제곱합(Residual Sum of Squares)라고 하며 이를 $SSE(\hat{\boldsymbol{\beta}})$로 표시한다.

$$SSE(\hat{\boldsymbol{\beta}}) = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^t (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

오차항의 분산에 대한 추정은 정규분포 가정을 오차항에 대한 정규분포를 가정하고 다음과 같은 잔차제곱합의 분포에 대한 결과를 이용하면

$$SSE(\hat{\boldsymbol{\beta}}) \sim \sigma_e^2 \chi^2 (n-k-1)$$

오차항의 분산 σ_e^2의 불편추정량 S^2을 구할 수 있다.

$$S^2 = \frac{SSE(\hat{\boldsymbol{\beta}})}{(n-k-1)} = \frac{\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^t \hat{\boldsymbol{\beta}})^2}{(n-k-1)}$$

여기서 자유도 $n-k-1$은 자료의 개수 n에서 절편을 포함한 회귀계수의 개수 $k+1$를 뺀 수이다.

1.3 최대가능도 추정

위에서 오차항에 대한 정규분포를 가정하면 반응변수가 정규분포를 따른다. 이러한 분포가정을 이용하여 회귀계수와 오차항의 분산에 대한 최대가능도 추정법(Maximum likelihoodestimation)을 고려할 수 있다. 선형모형과 정규분포 가정 하에서

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma_e^2 \boldsymbol{I}_n)$$

i번째 관측치 y_i는 정규분포 $N(\boldsymbol{x}_i^t\boldsymbol{\beta}, \sigma_e^2)$를 따르고 서로 독립이므로 관측치의 가능도함수(likelihood function) $L = L(\boldsymbol{\beta}, \sigma_e^2|\boldsymbol{y})$는 다음과 같다.

$$\begin{aligned}
L &= L(\boldsymbol{\beta}, \sigma_e^2|\boldsymbol{y}) \\
&= \prod_{i=1}^{n} f(y_i) \\
&= \prod_{i=1}^{n} (2\pi\sigma_e^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_e^2}(y_i - \boldsymbol{x}_i^t\boldsymbol{\beta})^2\right] \\
&= (2\pi\sigma_e^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_e^2}(\boldsymbol{y} - \boldsymbol{X\beta})^t(\boldsymbol{y} - \boldsymbol{X\beta})\right]
\end{aligned}$$

모든 관측값은 독립적이기 때문에 그들의 결합확률밀도는 그들의 주변 밀도의 곱이다. 위의 식으로부터 로그가능도함수(Log likelihood function) $l = l(\boldsymbol{\beta}, \sigma_e^2|\boldsymbol{y})$는 다음과 같이 된다.

$$l = \log L = const - \frac{n}{2}\log\sigma_e^2 - \frac{1}{2\sigma_e^2}(\boldsymbol{y} - \boldsymbol{X\beta})^t(\boldsymbol{y} - \boldsymbol{X\beta})$$

로그가능도함수 l을 최대화하는 회귀계수 $\boldsymbol{\beta}$와 분산 σ_e^2을 찾아야한다. 따라서 로그가능도함수 l을 $\boldsymbol{\beta}$와 σ_e^2으로 미분하여 0이 되는 값을 찾으면 그 값이 최대가능도 추정량이다. 여기서 주목할 점은 $(y - \boldsymbol{X\beta})^t(y - \boldsymbol{X\beta})$가 최소제곱법에서 나타나는 목표함수(오차제곱합)와 동일하다. 로그가능도함수 l을 미분하여 얻은 방정식은 다음과 같다.

$$\frac{\partial l(\boldsymbol{\beta}, \sigma_e^2)}{\partial\boldsymbol{\beta}} = -\frac{1}{2\sigma_e^2}(2\boldsymbol{X}^t\boldsymbol{X\beta} - 2\boldsymbol{X}^t\boldsymbol{y}) = 0$$

$$\frac{\partial l(\boldsymbol{\beta}, \sigma_e^2)}{\partial\sigma_e^2} = -\frac{n}{2\sigma_e^2} + \frac{1}{2\sigma_e^4}(\boldsymbol{y} - \boldsymbol{X\beta})^t(\boldsymbol{y} - \boldsymbol{X\beta}) = 0$$

편미분함수를 0으로 놓고 최대가능도 추정량 $\hat{\beta}$와 $\hat{\sigma}_e^2$방정식을 풀면 다음과 같이 쓸 수 있다.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$

$$\hat{\sigma}_e^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^t(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \frac{1}{n}SSE(\hat{\boldsymbol{\beta}})$$

그러므로 최대가능도 추정량 $\hat{\boldsymbol{\beta}}$는 최소제곱 추정량과 같다. 여기서 오차 분산의 최대가능도추정량 $\hat{\sigma}_e^2$는 편향되어진다(biased estimator).

$$E(\hat{\sigma}_e^2) = \frac{n-k-1}{n}\sigma_e^2 \neq \sigma_e^2$$

결과적으로 불편 추정량 $S^2 = \frac{SSE(\hat{\boldsymbol{\beta}})}{(n-k-1)}$과 유사한 것을 선호한다. 그러나, n이 증가할 때, 그 $\hat{\sigma}_e^2$ 편향은 0으로 감소하게된다.

여기거 주목할 점은 가능도함수에 최대가능도추정량을 대입하면 그 값이 $SSE(\hat{\boldsymbol{\beta}})$의 함수로 나타난다.

$$
\begin{aligned}
L(\hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2) &= (2\pi\hat{\sigma}_e^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\hat{\sigma}_e^2}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^t(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})\right] \\
&= (2\pi\hat{\sigma}_e^2)^{-\frac{n}{2}} \exp\left[-\frac{n}{2}\right] \\
&= \exp\left[-\frac{n}{2}\right]\left(2\pi\frac{SSE(\hat{\boldsymbol{\beta}})}{n}\right)^{-\frac{n}{2}} \\
l(\hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2) &= \text{constant} - \frac{n}{2}\log SSE(\hat{\boldsymbol{\beta}})
\end{aligned}
$$

따라서 잔차제곱함 $SSE(\hat{\boldsymbol{\beta}})$ 작아지면 가능도함수는 커진다.

2 선형혼합효과모형의 구조

이번 절에서는 일반적인 선형모형에 임의효과가 추가되는 선형혼합효과모형의 예제를 통하여 선형혼합모형의 구조를 살펴보고자 한다.

2.1 예제 1: 일원배치모형

일원배치모형(one-way random effect model)은 선형혼합효과모형의 가장 단순한 구조로서 다음과 같은 모형으로 나타낼 수 있다.

$$
y_{ij} = \mu + b_i + e_{ij}, \quad i = 1, 2, \ldots, I \text{ and } j = 1, 2, \ldots, J \tag{2}
$$

여기서 $b_i, i = 1, 2, \ldots, I$는 첫 번째 계층의 표본 효과(예를 들어 학교 효과)를 나타내는 임의효과이며 서로 독립이고 $N(0, \sigma_b^2)$를 따른다. 또한 계층내에서 관측값(예를 들어 학생)에 대한 오차항 $e_{ij}, j = 1, 2, \ldots, J$는 서로 독립이며 $N(0, \sigma_e^2)$를 따른다. 임의효과 b_i와 오차항 e_{ij}는 서로 독립이다. 고정효과 μ는 전체 평균을 나타내는 모수이다. 첫 번째 계층에서 표본(학교)을 추출할 때 σ_b^2를 가지는 정규모집단에서 추출한다고 가정하는 것이다. 보통 이러한 표본은 하나의 군집(cluster)로 보고 군집간의 변동을 σ_b^2 으로 설명하고 군집내의 관측치들간의 변동은 σ_e^2으로 설명한다.

위와 같은 혼합효과모형(mixed effects model)을 각 개체 i에 대하여 행렬식으로 표시하면 다음과 같다.

$$
\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i b_i + \boldsymbol{e}_i, \quad i = 1, 2, \ldots, J
$$

여기서

$$\boldsymbol{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iJ} \end{bmatrix},\ \boldsymbol{X}_i = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \end{bmatrix},\ \boldsymbol{Z}_i = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},\ \boldsymbol{b}_i = \begin{bmatrix} b_i \end{bmatrix},\ \boldsymbol{e}_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iJ} \end{bmatrix}$$

위의 각 개체에 대한 혼합효과모형을 합쳐서 아래와 같은 행렬식으로 표현할 수 있다.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{e}$$

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_I \end{bmatrix} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1J} \\ \hdashline y_{21} \\ y_{22} \\ \vdots \\ y_{2J} \\ \hdashline \vdots \\ \hdashline y_{I1} \\ y_{I2} \\ \vdots \\ y_{IJ} \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_I \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \hdashline 1 \\ 1 \\ \vdots \\ 1 \\ \hdashline \vdots \\ \hdashline 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \end{bmatrix}$$

$$
\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & 0 & \dots & 0 \\ 0 & \boldsymbol{Z}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \boldsymbol{Z}_I \end{bmatrix} = \left[\begin{array}{cccc} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ \hdashline 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 1 & \dots & 0 \\ \hdashline \vdots & \vdots & \dots & \vdots \\ \hdashline 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{array}\right] \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_I \end{bmatrix} \quad \boldsymbol{e} = \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_I \end{bmatrix} = \left[\begin{array}{c} e_{11} \\ e_{12} \\ \vdots \\ e_{1J} \\ \hdashline e_{21} \\ e_{22} \\ \vdots \\ e_{2J} \\ \hdashline \vdots \\ \hdashline e_{I1} \\ e_{I2} \\ \vdots \\ e_{IJ} \end{array}\right]
$$

위의 행렬식에서 $\boldsymbol{y}$는 $IJ \times 1$의 관측값을 가지는 반응변수벡터이고 $\boldsymbol{X}$는 전체평균 μ에 대한 모형벡터로 모두 1인 값을 가지는 $IJ \times 1$ 벡터이다. 여기서 반응변수벡터 $\boldsymbol{y}$와 고정효과 $\boldsymbol{\beta}$에 대한 계획행렬 X는 각 개인의 반응변수벡터 $\boldsymbol{y}_i$와 $\boldsymbol{X}_i$를 행으로 쌓아놓은 것으로 표현된다. 오차항에 대한 벡터 $\boldsymbol{e}$도 동일한 형식의 벡터이다. 임의효과 벡터 $\boldsymbol{b}$는 각 개인에 대한 임의효과 b_i를 행으로 쌓아놓은것과 같고 임의효과에 대한 계획행렬 $\boldsymbol{Z}$는 각 개인의 계획행렬 $\boldsymbol{Z}_i$를 대각원소로 같은 행렬이다.

임의효과 벡터 $\boldsymbol{b}$는 I개의 임의효과를 가지는 벡터이고 그 분포는 평균이 0이고 공분산행렬이 $\sigma_b^2 I$를 가지는 다변량 정규분포를 따른다. 오차항의 벡터는 일반적인 선형모형과 같이 평균이 0 이고 분산이 $\sigma_e^2 I$인 다변량정규분포를 따른다고 가정한다.

$$
\boldsymbol{b} \sim N(\boldsymbol{0}, \sigma_b^2 \boldsymbol{I}_I) \quad \boldsymbol{e} \sim N(\boldsymbol{0}, \sigma_e^2 \boldsymbol{I}_{IJ})
$$

여기서 주의할 점은 오차항의 분산과는 달리 임의효과의 분산 σ_b^2의 가능한 값의 범위는 0을 포함한 양수이다. 따라서 분산 σ_b^2의 추정값은 0의 값을 가질수 있으며 이러한 경우는 첫번 째 계층에서 나오는 변동이 없다는 것을 의미한다. 임의효과에 대한 설계행렬(design matrix) $\boldsymbol{Z}$는 $IJ \times I$의 행렬로서 원소의 값이 1 또는 0인 행렬이며 이 설계행렬은 각각의 관측치 y_{ij}에 해당하는 임의효과 b_i를 지정해주는 역활을 한다.

2.2 예제 2: 반복측정자료

반복측정자료(longitudinal data, repeated measurements)는 관측단위안에서 여러 개의 관측값을 측정한 자료의 형식을 말한다. 혼합모형을 이용하여 반복측정자료를 분석하는 방법을 예를 들어 설명해보자. `lme4` 패키지에 자료인 `spleepstudy`는 화물트럭 운전사들에 대한 수면부족 현상에 대하여 연구한 자료이다. 18명의 운전자들이 매일 3시간의 수면을 하면서 10일동안 매일 동작의 반응시간을 반족적으로 측정한 자료가 있다. 한 운전사에게 10일 동안의 반응에 대한 측정자료 10개가 존재하므로 이는 반복측정 자료이며 이러한 10개의 자료는 독립이 아니다. 따라서 이러한 종속성을 설명할 모형이 필요하며 혼합모형을 이용하여 자료를 적합하려고 한다.

18명에 대한 회귀직선의 절편과 기울기를 보면 개인의 차이에 따른 변동을 볼 수 있으며 이러한 각 개인간의 변동을 임의효과를 이용하여 다음과 같은 모형을 생각해보자.

$$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j + e_{ij} \tag{3}$$

여기서 임의효과 $\boldsymbol{b}_i = (b_{0i}, b_{1i})^t$는 다음과 같은 이변량정규분포를 가진다고 가정한다. 또한 오차항 e_{ij}는 정규분포 $N(0, \sigma_e^2)$를 따르고 임의효과 $\boldsymbol{b}_i$와는 독립이다.

$$\boldsymbol{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b11} & \sigma_{b12} \\ \sigma_{b12} & \sigma_{b22} \end{bmatrix} \right)$$

위의 모형은 각 개인의 회귀직선에서 각 절편과 기울기가 전체평균 β_0와 β_1를 따르며 각 개인의 차이는 전체평균에 임의효과인 b_{0i}와 b_{1i}가 더해져서 나타난다는 것을 의미한다. 이변량 임의효과 b_{0i}와 b_{1i}는 이변량 정규분포를 따르며 각각의 분산과 상관계수가 σ_{b11}, σ_{b22}, $\rho = \sigma_{b12}/\sqrt{\sigma_{b11}\sigma_{b22}}$이다.

다른 개체에 대한 임의효과는 서로 독립이며 임의 효과와 오차항은 독립이다. 여기서 오차항은 서로 독립이며 $N(0, \sigma_e^2)$를 따른다고 가정한다.

$$Cov(\boldsymbol{b}_i, \boldsymbol{b}_j) = \boldsymbol{0} \text{ when } i \neq j, \quad Cov(\boldsymbol{b}_i, e_{jk}) = \boldsymbol{0} \text{ for all } i, j, k$$

위와 같은 혼합효과모형(mixed effects model)을 각 개체 i에 대하여 행렬식으로 표시하면 다음과 같다.

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i, i = 1.2 \cdots, 18$$

여기서

$$\boldsymbol{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i,10} \end{bmatrix}, \ \boldsymbol{X}_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 9 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \ \boldsymbol{Z}_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 9 \end{bmatrix}, \ \boldsymbol{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}, \ \boldsymbol{e}_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{i,10} \end{bmatrix}$$

위의 각 개인에 대한 모형을 모두 합쳐서 하나의 혼합효과모형으로 나타내면 다음과 같이 표현할 수 있다.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{e} \tag{4}$$

여기서 반응변수벡터 $\boldsymbol{y}$와 고정효과 $\boldsymbol{\beta}$에 대한 계획행렬 X는 각 개인의 반응변수벡터 $\boldsymbol{y}_i$와 $\boldsymbol{X}_i$를 행으로 쌓아놓은 것으로 표현된다. 오차항에 대한 벡터 $\boldsymbol{e}$도 동일한 형식의 벡터이다.

$$\boldsymbol{y}_= \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_{18} \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_{18} \end{bmatrix} \ \boldsymbol{e} = \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_{18} \end{bmatrix}$$

임의효과 벡터 $\boldsymbol{b}$는 각 개인에 대한 임의효과벡터 $\boldsymbol{b}_i$를 행으로 쌓아놓은것과 같고 임의효과에 대한 계획행렬 $\boldsymbol{Z}$는 각 개인의 계획행렬 $\boldsymbol{Z}_i$를 대각원소로 같은 행렬이다.

$$\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_{18} \end{bmatrix}, \ \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & 0 & \dots & 0 \\ 0 & \boldsymbol{Z}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \boldsymbol{Z}_{18} \end{bmatrix}$$

2.3 선형혼합모형의 확률모형

일반적인 선형혼합모형은 다음과 같은 행렬식을 통해 표현할 수 있다.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{e} \tag{5}$$

여기서 $\boldsymbol{y}$는 $n \times 1$ 벡터로 반응변수이다. $\boldsymbol{X}$는 $n \times p$의 설명변수로 이루어진 행렬이며 절편항과 $p-1$개의 독립변수를 포함한다(따라서 절편을 포함한 설명변수의 수는 p개이다). $\boldsymbol{\beta}$는 $p \times 1$의 고정 효과 벡터이다. $\boldsymbol{Z}$는 $n \times q$ 행렬이며 임의효과를 고려한 설명변수 벡터로 특수한 형태를 갖는다. $\boldsymbol{b}$는 $q \times 1$의 임의효과 벡터이고 오차항 $\boldsymbol{e}$은 $n \times 1$ 확률벡터이다.

이 때 임의효과 $\boldsymbol{b}$와 잔차 $\boldsymbol{e}$는 각각 $\boldsymbol{b} \sim N(0, \boldsymbol{G}), \boldsymbol{e} \sim N(0, \sigma_e^2 \boldsymbol{I}_n)$을 따르며 $\boldsymbol{b}$와 $\boldsymbol{e}$는 서로 독립이다. 따라서 $\boldsymbol{y}$의 평균과 분산은 다음과 같이 표현된다.

$$
\begin{aligned}
E(\boldsymbol{y}) &= \boldsymbol{X}\boldsymbol{\beta} \\
Var(\boldsymbol{y}) &= Var(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{e}) \\
&= Var(\boldsymbol{Z}\boldsymbol{b}) + Var(\boldsymbol{e}) \\
&= \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^t + \sigma_e^2 \boldsymbol{I}_n \\
&\equiv \boldsymbol{V}
\end{aligned}
$$

또한 만약 임의효과 $\boldsymbol{b}$가 주어졌을 경우 $\boldsymbol{y}$의 조건부 분포는 $N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}, \sigma^2 \boldsymbol{I}_n)$ 이므로 조건부 평균과 분산은 다음과 같다.

$$
\begin{aligned}
E(\boldsymbol{y}|\boldsymbol{b}) &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} \\
Var(\boldsymbol{y}|\boldsymbol{b}) &= \sigma_e^2 \boldsymbol{I}_n
\end{aligned}
$$

임의효과 $\boldsymbol{b}$의 공분산 행렬은 $\boldsymbol{G}$는 임의효과의 구조에 따라 그 형태가 정해지는 공분산 행렬이며 임의효과들 사이의 상관성을 나타낼 수 있다. 또한 관측벡터의 공분산행렬 $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^t + \sigma_e^2 \boldsymbol{I}$이고 이 행렬은 관측값의 분산과 공분산을 나타낸다.

따라서 선형혼합모형에서 추정해야 할 모수들은 고정효과인 $\boldsymbol{\beta}$, 임의효과의 분산인 $\boldsymbol{G}$, 오차의 분산인 σ_e^2이 된다.

여기서 임의효과 $\boldsymbol{b}$의 공분산 행렬 $\boldsymbol{V}$의 형태를 앞에서 본 두 예제들 경우에 살펴보자.

- 일원배치모형: $I = 3, J = 2$인 경우

  일원배치모형에서 층(그룹)의 수가 3개이고($I = 3$) 각 층에서 관측치가 2개($J = 2$)인 경우 임의효과 $\boldsymbol{b}$와 공분산 행렬 $\boldsymbol{G}$이 다음과 같다. 여기서 유의할 점은 임의 벡터의 공분산행렬 $\boldsymbol{G}$에서 추정해야할 모수는 σ_b^2 하나이다.

$$
\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \quad \boldsymbol{G} = Cov(\boldsymbol{b}) = \begin{bmatrix} \sigma_b^2 & 0 & 0 \\ 0 & \sigma_b^2 & 0 \\ 0 & 0 & \sigma_b^2 \end{bmatrix}
$$

또한 각 개체에 대한 임의효과에 대한 계획행렬 $\boldsymbol{Z}_i$와 전체 계획행렬 $\boldsymbol{Z}$는 다음과 같이 나타나고

$$\boldsymbol{Z}_i = \boldsymbol{Z}_* = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ i = 1,2,3 \quad \boldsymbol{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{Z}_3 \end{bmatrix}$$

또한 $\boldsymbol{G\Sigma Z}^t$은 다음과 같이 나타난다.

$$\begin{aligned} \boldsymbol{ZGZ}^t &= \begin{bmatrix} \boldsymbol{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{Z}_3 \end{bmatrix} \begin{bmatrix} \sigma_b^2 & 0 & 0 \\ 0 & \sigma_b^2 & 0 \\ 0 & 0 & \sigma_b^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{Z}_3 \end{bmatrix}^t \\ &= \begin{bmatrix} \sigma_b^2 \boldsymbol{Z}_1 \boldsymbol{Z}_1^t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_b^2 \boldsymbol{Z}_2 \boldsymbol{Z}_2^t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_b^2 \boldsymbol{Z}_3 \boldsymbol{Z}_3^t \end{bmatrix} \end{aligned}$$

이때 $\boldsymbol{Z}_i\boldsymbol{Z}_i^t$의 형태는 모두 $\boldsymbol{Z}_*\boldsymbol{Z}_*^t$로 모두 동일하고 다음과 같으므로

$$\boldsymbol{Z}_i\boldsymbol{Z}_i^t = \boldsymbol{Z}_*\boldsymbol{Z}_*^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

위에서 행렬 $\boldsymbol{ZGZ}^t$는 3개의 2×2 행렬이 대각(diagonal)으로 배치된 형태로서 다음과 같이 나타낼 수 있다.

$$\boldsymbol{ZGZ}^t = diag\{\sigma_b^2 \boldsymbol{Z}_i \boldsymbol{Z}_i^t\}_{i=1}^3$$

여기서 $diag\{\boldsymbol{A}_i\}_{i=1}^q$ 기호는 행렬 $\boldsymbol{A}_i$들을 i번째 대각위치로 구성하는 행렬을 나타낸다. 또한 관측벡터의 공분산 행렬 $\boldsymbol{V} = \boldsymbol{ZGZ}^t + \sigma_e^2\boldsymbol{I}$는 다음과 같이 주어진다.

$$\boldsymbol{V} = \begin{bmatrix} \sigma_b^2+\sigma_e^2 & \sigma_b^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma_b^2+\sigma_e^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & 0 & 0 \\ 0 & 0 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 \\ 0 & 0 & 0 & 0 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 \end{bmatrix}$$

또는

$$\boldsymbol{V} = diag\{\sigma_b^2 \boldsymbol{Z}_i \boldsymbol{Z}_i^t\}_{i=1}^3 + \sigma_e^2 \boldsymbol{I}_6$$

같은 그룹에 속한 관측값의 공분산은 σ_b^2이고 상관계수는 다음과 같다.

$$cor(y_{ij}, y_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$$

- 반복측정자료: $I = 3, J = 4$인 경우

위의 반복측정자료 예제에서 개체가 3개이고($I = 3$) 반복한 시간의 개수가 4번인 경우($J = 4$) 각 개체에 대한 임의효과에 대한 계획행렬 $\boldsymbol{Z}_i$와 전체 계획행렬 $\boldsymbol{Z}$ 은 다음과 같이 나타난다.

$$\boldsymbol{Z}_i = \boldsymbol{Z}_* = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \; i = 1,2,3 \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z}_3 \end{bmatrix}$$

각 개체에 대한 임의효과 벡터 $\boldsymbol{b}_i$의 분포를 다음과 같이 나타내면

$$\boldsymbol{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b11} & \sigma_{b12} \\ \sigma_{b12} & \sigma_{b22} \end{bmatrix} \right)$$

위에서 한 개체에 대한 임의효과 벡터 $\boldsymbol{b}_i$의 공분산을 $\boldsymbol{\Psi}$라고 하자.

$$\boldsymbol{\Psi} = \begin{bmatrix} \sigma_{b11} & \sigma_{b12} \\ \sigma_{b12} & \sigma_{b22} \end{bmatrix}$$

이제 전체 임의효과 $\boldsymbol{b} = (\boldsymbol{b}_1^t, \boldsymbol{b}_2^t, \boldsymbol{b}_3^t)^t$의 공분산 행렬 $\boldsymbol{G}$은 다음과 같다.

$$\boldsymbol{G} = Cov(\boldsymbol{b}) = \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Psi} \end{bmatrix}$$

따라서 $\boldsymbol{ZGZ}^t$은 다음과 같이 나타난다.

$$\begin{aligned} \boldsymbol{ZGZ}^t &= \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Psi} \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z}_3 \end{bmatrix}^t \\ &= \begin{bmatrix} \boldsymbol{Z}_1 \boldsymbol{\Psi} \boldsymbol{Z}_1^t & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 \boldsymbol{\Psi} \boldsymbol{Z}_2^t & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z}_3 \boldsymbol{\Psi} \boldsymbol{Z}_3^t \end{bmatrix} \end{aligned}$$

이때 $\boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^t = \boldsymbol{Z}_*\boldsymbol{\Psi}\boldsymbol{Z}_*^t$의 형태는 다음과 같이 모두 같으므로

$$
\begin{aligned}
\boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^t \quad &= \boldsymbol{Z}_*\boldsymbol{\Psi}\boldsymbol{Z}_*^t \\
&= \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \sigma_{b11} & \sigma_{b12} \\ \sigma_{b12} & \sigma_{b22} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{b11} & \sigma_{b12} \\ \sigma_{b11}+\sigma_{b12} & \sigma_{b12}+\sigma_{b22} \\ \sigma_{b11}+2\sigma_{b12} & \sigma_{b12}+2\sigma_{b22} \\ \sigma_{b11}+3\sigma_{b12} & \sigma_{b12}+3\sigma_{b22} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{b11} & \sigma_{b11}+\sigma_{b12} & \sigma_{b11}+2\sigma_{b12} & \sigma_{b11}+3\sigma_{b12} \\ \sigma_{b11}+\sigma_{b12} & \sigma_{b22}+2\sigma_{b12}+\sigma_{b22} & \sigma_{b22}+3\sigma_{b12}+2\sigma_{b22} & \sigma_{b22}+4\sigma_{b12}+3\sigma_{b22} \\ \sigma_{b11}+2\sigma_{b12} & \sigma_{b22}+3\sigma_{b12}+2\sigma_{b22} & \sigma_{b22}+4\sigma_{b12}+4\sigma_{b22} & \sigma_{b22}+5\sigma_{b12}+6\sigma_{b22} \\ \sigma_{b11}+3\sigma_{b12} & \sigma_{b22}+4\sigma_{b12}+3\sigma_{b22} & \sigma_{b22}+5\sigma_{b12}+6\sigma_{b22} & \sigma_{b22}+6\sigma_{b12}+9\sigma_{b22} \end{bmatrix}
\end{aligned}
$$

관측벡터의 공분산 행렬 $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^t + \sigma_e^2\boldsymbol{I}$는 다음과 같이 주어진다.

$$\boldsymbol{V} = diag\{\boldsymbol{Z}_i\Psi\boldsymbol{Z}_i^t\}_{i=1}^3 + \sigma_e^2\boldsymbol{I}_{12}$$

같은 그룹에 속한 관측값의 분산과 공분산은 다음과 같이 직접 계산할 수도 있다.

$$var(y_{ij}) = var((\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j + e_{ij}) = \sigma_{b11} + 2t_j\sigma_{b12} + t_j^2\sigma_{b11} + \sigma_e^2$$

$$
\begin{aligned}
cov(y_{ij}, y_{ik}) &= cov[(\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j + e_{ij},\ (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_k + e_{ik}] \\
&= cov(b_{0i}, b_{0i})) + (t_j + t_k)cov(b_{0i}, b_{1i}) + t_jt_kcov(b_{1i}, b_{1i}) + cov(e_{ij}, e_{ik}) \\
&= \sigma_{b11} + (t_j + t_k)\sigma_{b12} + t_jt_k\sigma_{b11} + 0
\end{aligned}
$$

따라서 같은 그룹에 속한 두 관측값의 상관계수는 다음과 같다.

$$cor(y_{ij}, y_{ik}) = \frac{\sigma_{b11} + (t_j + t_k)\sigma_{b12} + t_jt_k\sigma_{b11}}{\sqrt{(\sigma_{b11} + 2t_j\sigma_{b12} + t_j^2\sigma_{b11} + \sigma_e^2)(\sigma_{b11} + 2t_k\sigma_{b12} + t_k^2\sigma_{b11} + \sigma_e^2)}}$$

3 최대가능도 추정법

3.1 분포의 가정

선형혼합모형에서 모수를 추전하는 방법은 앞의 강의 노트에서 보인 것처럼 모형이 단순한 경우 적률법(method of moments)를 이용하여 평균과 분산성분을 추정할 수 있다 (선형혼합모형 I-혼합모형의 소개, p7에 일원배치법 참고). 하지만 일반적인 모형에 대하여 각각 적률법을 적용할 제곱합을 설정하고 그 기대값을 구하는것은 어려운 일이다. 따라서 선형혼합모형에서는 정규분포를 가정하고 모수의 추정을 최대가능도추정법을 사용한다.

잎에서 본 예제들과 같이 선형혼합모형은 각 개체 또는 그룹안에서 관측된 확률변수들은 독립이 아니지만 다른 그룹에서 나온 관측 변수들은 독립을 가정한다. 위와 같은 혼합효과모형(mixed effects model)을 각 개체 i에 대하여 행렬식으로 표시하면 다음과 같다.

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, i = 1.2\cdots, I$$

일반적으로 그룹간에서 반복된 관측치를 가지는 모형은 다음과 같은 분포를 고려할 수 있다.

$$\boldsymbol{b}_i \sim_{iid} N(\boldsymbol{0}, \sigma_e^2\boldsymbol{\Psi}), \quad \boldsymbol{e}_i \sim N(\boldsymbol{0}, \sigma_e^2\boldsymbol{I})$$

여기서 $\boldsymbol{b}_i$와 $\boldsymbol{e}_i$는 독립이다. 위의 가정에서 주의할 점은 임의효과의 공분산행렬을 오차항의 분산 σ_e^2과 행렬 $\boldsymbol{\Psi}$의 곱으로 표현한 것이며 그 이유는 최대가능도 추정함수를 좀 더 간결하게 표시하기 위해서이다.

3.2 가능도함수

가능도 함수는 반응 벡터 $\boldsymbol{y}_i$의 분포를 이용하여 표현할 수 있다. $\boldsymbol{y}_i$의 분포는 다변량 정규분포를 따르고

$$E(\boldsymbol{y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}, \quad V(\boldsymbol{y}_i) = V_i = \sigma_e^2(\boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^t + \boldsymbol{I})$$

이므로 관측벡터 $\boldsymbol{y}_i$의 차원이 J라면

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{y}_i) &= f(\boldsymbol{y}_i|\boldsymbol{\theta}) \\ &= (2\pi)^{-J/2}|V_i|^{-1/2}\exp\left[-\frac{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^t V_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})}{2}\right] \\ &= (2\pi\sigma_e^2)^{-J/2}|\boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^t + \boldsymbol{I}|^{-1/2}\exp\left[-\frac{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^t[\boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^t + \boldsymbol{I}]^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})}{2\sigma_e^2}\right] \end{aligned} \tag{6}$$

여기서 모수벡터 $\boldsymbol{\theta}$는 회귀계수 $\boldsymbol{\beta}$, 임의효과의 공분산 Ψ을 결정하는 분산성분에 대한 모수들 $\boldsymbol{\psi}$, 오차항의 분산 σ_e^2를 모아 놓은 벡터이다 (부록참조)

이제 반응값들의 벡터들 $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_I$들이 모두 독립이므로 모든 관측값을 모아 놓은 관측벡터 $\boldsymbol{y} = (\boldsymbol{y}_1^t, \boldsymbol{y}_2^t, \ldots, \boldsymbol{y}_I^t)^t$에 의한 최대가능도 함수는 각각의 최대가능도 함수들을 곱한것과 같다.

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{I} L(\boldsymbol{\theta}|\boldsymbol{y}_i) \tag{7}$$

위에서 구한 가능도함수 (6)는 일반적으로 계산하기가 쉽지 않다. 왜냐하면 관측벡터 $\boldsymbol{y}_i$의 공분산 행렬 $\boldsymbol{V}_i$에 대한 역행렬을 구하는 것이 단순한 모형을 제외하면 일반적으로 쉽지 않다.

3.3 일원배치모형에서의 가능도 함수

선형혼합모형의 가장 기본적인 형태는 앞에서 공부한 일원배치모형이다. 하나의 요인이 반응변수에 미치는 영향을 알 수 있는 모형으로 다음과 같이 정의할 수 있다.

$$y_{ij} = \mu + b_i + e_{ij}, \quad i = 1, 2, \ldots, I,\ j = 1, 2, \ldots, J$$

이 때, y_{ij}는 i번째 그룹의 j번째 관측값이고 μ는 고정효과이다. 임의효과 b_i와 오차항 e_{ij}는 각각 $b_i \sim N(0, \sigma_b^2), e_{ij} \sim N(0, \sigma_e^2)$을 따르며 서로 독립이다.

일원배치모형의 가능도 함수를 얻기 위해서 i번째 관측벡터 $\boldsymbol{y}_i$의 분포를 알아야 한다. $\boldsymbol{y}_i$의 분포는 다음과 같다.

$$\boldsymbol{y}_i \sim_{ind} \ N(\mu\mathbf{1}, \boldsymbol{V}_i),\ \boldsymbol{V}_i = \sigma_e^2 \boldsymbol{I} + \sigma_b^2 \mathbf{1}\mathbf{1}^t, \quad i = 1, 2, \ldots, I$$

여기서 $\mathbf{1}$은 모든 원소가 1이고 차원이 J 인 벡터, $\boldsymbol{I}$은 차원이 $J \times J$ 인 항등행렬이다. 또한 모수들은 $\boldsymbol{\theta} = (\mu, \sigma_b^2, \sigma_e^2)^t$으로 구성된다.

따라서 정규분포의 가능도함수를 이용하여 식을 표현하면 다음과 같다.

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{I} (2\pi)^{-\frac{J}{2}} |\boldsymbol{V}_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}_i - \mu\mathbf{1})^t \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \mu\mathbf{1})\right]$$

여기서 V_i의 행렬식와 역행렬은 각각 $|\boldsymbol{V}_i| = (\sigma_e^2 + J\sigma_b^2)(\sigma_e^2)^{J-1}$

$$\boldsymbol{V}_i^{-1} = \frac{1}{\sigma_e^2}\left[\boldsymbol{I} - \frac{\sigma_b^2}{\sigma_e^2 + J\sigma_b^2}\boldsymbol{J}\right]$$

이고 (부록 참조) 이를 가능도함수에 대입하여 로그 가능도함수를 구해보면 다음과 같다.

$$\log L = -\frac{1}{2}(IJ)\log 2\pi - \frac{1}{2}\sum_i \log(\sigma_e^2 + J\sigma_b^2) - \frac{1}{2}I(J-1)\log\sigma_e^2$$
$$-\frac{1}{2\sigma_e^2}\sum_i\sum_j (y_{ij}-\mu)^2 + \frac{J^2\sigma_b^2}{2\sigma_e^2}\sum_i \frac{(\bar{y}_{i\cdot}-\mu)^2}{\sigma_e^2 + J\sigma_b^2}$$

여기서 $\bar{y}_{i\cdot} = \sum_{j=1}^{J} y_{ij}/J$이다. 또한 여기서 $\lambda = \sigma_e^2 + J\sigma_b^2$로 놓으면

$$\begin{aligned}\log L = &-\frac{1}{2}(IJ)\log 2\pi - \frac{1}{2}I(J-1)\log\sigma_e^2 - \frac{1}{2}I\log\lambda \\ &-\frac{SSE}{2\sigma_e^2} - \frac{SSA}{2\lambda} - \frac{IJ(\bar{y}_{\cdot\cdot}-\mu)^2}{2\lambda}\end{aligned} \tag{8}$$

위 식에서 $SSA = \sum_i J(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$, $SSE = \sum_i\sum_j (y_{ij} - \bar{y}_{i\cdot})^2$ 이다.

로그가능도함수를 각 모수 $(\mu, \sigma_e^2, \lambda)$에 대해 미분하면 다음과 같은 식이 주어진다.

$$\begin{aligned}\frac{\partial \log L}{\partial \mu} &= \frac{IJ(\bar{y}_{\cdot\cdot}-\mu)}{\lambda} \\ \frac{\partial \log L}{\partial \sigma_e^2} &= -\frac{J(I-1)}{2\sigma_e^2} + \frac{SSE}{\sigma_e^4} \\ \frac{\partial \log L}{\partial \sigma_b^2} &= \frac{-I}{2\lambda} + \frac{SSA}{2\lambda^2} + \frac{IJ(\bar{y}_{\cdot\cdot}-\mu)^2}{2\lambda^2}\end{aligned}$$

위의 방정식들을 0으로 놓고 풀면 다음과 같은 추정량의 근을 구할 수 있다.

$$\tilde{\mu} = \bar{y}_{\cdot\cdot}, \quad \tilde{\sigma}_e^2 = MSE = \frac{SSE}{I(J-1)}, \quad \tilde{\lambda} = \frac{SSA}{I}$$

모수 $\lambda = \sigma_e^2 + J\sigma_b^2$임을 이용하면

$$\tilde{\sigma}_b^2 = \frac{\hat{\lambda} - \hat{\sigma}_e^2}{J} = \frac{SSA/I - MSE}{J}$$

이다.

위에서 구한 가능도함수의 미분 방정식의 근 $(\tilde{\mu}, \tilde{\sigma}_e^2, \tilde{\sigma}_b^2)$ 중에서 평균과 오차항의 추정량은 최대가은도 추정량으로 설정할 수 있지만 임의효과의 분산의 근은 0보다 작은 값이 나올 수 있으므로 다음과 같은 추정량이 최대가능도 추정량이다.

$$\begin{aligned}\hat{\mu} &= \bar{y}_{\cdot\cdot} \\ \hat{\sigma}_e^2 &= \frac{SSE}{I(J-1)} = MSE \\ \hat{\sigma}_b^2 &= \begin{cases} \frac{SSA/I-MSE}{J} & \text{if } \frac{SSA/I-MSE}{J} \geq 0 \\ 0 & \text{otherwise} \end{cases}\end{aligned} \tag{9}$$

위의 최대가능도 추정량을 적률법으로 구한 추정량과 비교하면 (강의노트 선형혼합모형 I,page 8) 평균과 오차항의 추정량은 동일하고 임의효과의 분산 추정량이 다른 점을 알 수 있다.

3.4 임의효과의 예측

임의효과는 관측할 수 없는 확률변수이지만 관측값이 주어진 경우 임의효과에 대한 기대값으로 그 값을 예측할 수 있다. 임의효과는 모수가 아니라 확률변수이므로 이에 대한 추론을 추정(estimation)이라고 하지 않고 예측(prediction)이라고 한다.

앞에서 공부한 일원배치모형에서 임의효과 b_i와 그룹의 평균 $\bar{y}_{i.}$의 공분산 구해보자.

$$\begin{aligned}
cov(b_i, \bar{y}_{i.}) &= cov(b_i, \mu + bi + \bar{e}_{i.}) \\
&= cov(b_i, b_i) \\
&= var(b_i) \\
&= \sigma_b^2
\end{aligned}$$

임의효과 b_i와 그룹의 평균 $\bar{y}_{i.}$는 각각 다음과 같은 정규분포를 따르므로

$$b_i \sim N(0, \sigma_b^2), \quad \bar{y}_{i.} \sim N(\mu, \sigma_b^2 + \sigma_e^2/J)$$

임의효과 b_i와 그룹의 평균 $\bar{y}_{i.}$는 다음과 같은 이변량 정규분포를 따른다.

$$\begin{bmatrix} b_i \\ \bar{y}_{i.} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma_e^2/J \end{bmatrix} \right)$$

따라서 그룹의 평균 $\bar{y}_{i.}$ 이 주어진 경우 임의효과 b_i의 조건부 분포는 정규분포이며 조건부 기대값은 다음과 같이 주어진다 (강의노트 이변량 정규분포 참조). 는 다음과 같다.

$$E(b_i|\bar{y}_{i.}) = E(b_i) + \frac{cov(b_i, \bar{y}_{i.})}{var(\bar{y}_{i.})}(\bar{y}_{i.} - E[\bar{y}_{i.}])$$

따라서

$$E(b_i|\bar{y}_{i.}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2/J}(\bar{y}_{i.} - \mu)$$

로 주어진다. 위의 식에서 모르는 모수를 최대가능도 추정량으로 대체해주면 임의료과에 대한 예측값을 구할 수 있다.

$$\hat{b}_i = \hat{E}(b_i|\bar{y}_{i.}) = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_e^2/J}(\bar{y}_{i.} - \bar{y}_{..}) \tag{10}$$

위의 식 (10)을 보면 고정효과를 고려한 일원배치모형(강의노트, 선형혼합모형 I의 7번 모형)에서 각 그룹의 효과 $_i$의 추정식 $\hat{\alpha}_i$인 다음의 식과 차이가 나는 것을 볼 수 있다.

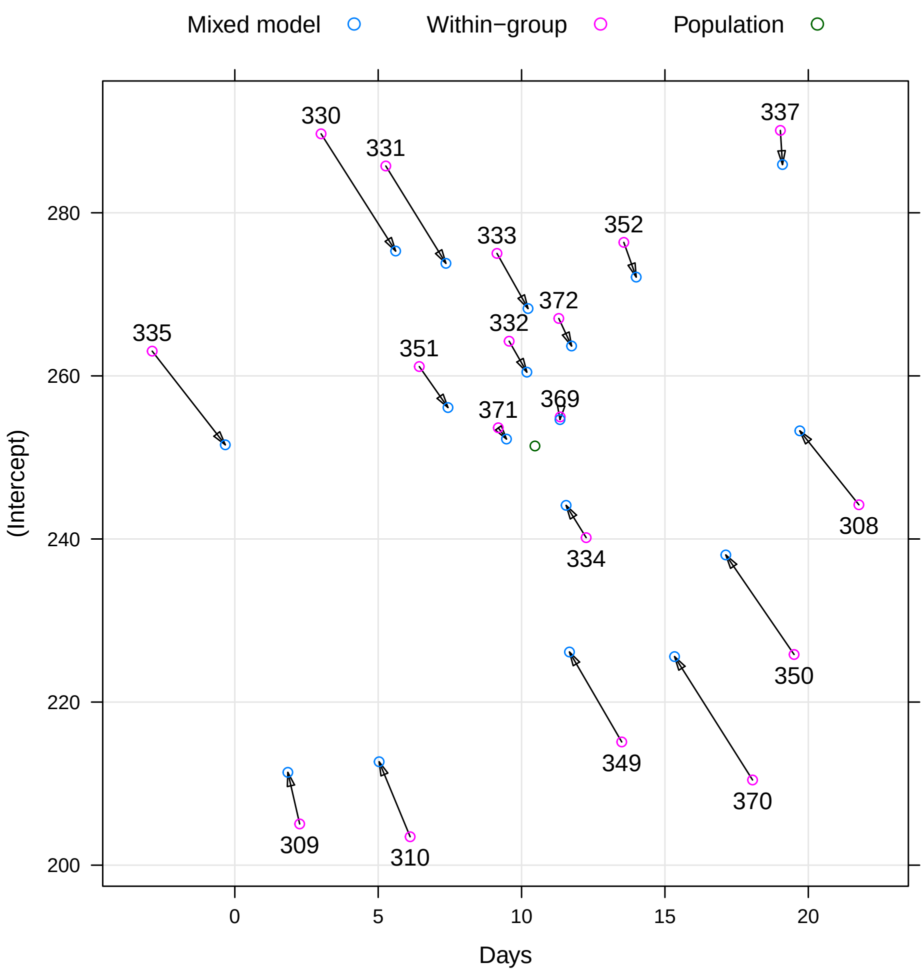
$$\hat{\alpha}_i = y_{i.} - \bar{y}_{..}$$

임의효과를 이용하면 그룹의 효과가 고정효과를 이용한 모형보다 그 절대값이 작게 나오는 것을 알 수 있다.

$$\left| \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_e^2/J} (\bar{y}_{i.} - \bar{y}_{..}) \right| \leq |y_{i.} - \bar{y}_{..}|$$

이러한 현상을 전체 평균으로의 축소현상(shrinkage to grand mean)이라고 부른다.

앞에서 공부한 반복측정자료 sleepstudy에서 혼합모형을 통해서 얻은 각 개인의 절편과 기울기에 대한 예측값과 각각의 개인에 대해서 회귀직선을 따로 적합하여 얻은 절편과 기울기의 관계를 그림으로 그려보면 다음과 같다. 즉 혼합모형을 통해서 얻은 각 개인의 절편과 기울기는 절편과 기울기의 전체평균값 방향으로 축소되는 경향(shrinkage)을 볼수있다.



4 부록: 행렬대수

4.1 다변량 정규분포

k 차원 다변량 정규분포 $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$의 확률밀도함수가 다음과 같다.

$$f(\boldsymbol{y}) = (2\pi)^{-k/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right]$$

여기서 $|\boldsymbol{A}|$는 행렬 $\boldsymbol{A}$의 행렬식이다.

4.2 Woodbury formula

$$(\boldsymbol{A}+\boldsymbol{UCV})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}$$

$$(\boldsymbol{I}+\boldsymbol{UCV})^{-1} = \boldsymbol{I} - \boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{VU})^{-1}\boldsymbol{V}$$

$$(\boldsymbol{A}+\boldsymbol{u}\boldsymbol{v}^t)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^t\boldsymbol{A}^{-1}}{1+\boldsymbol{v}^t\boldsymbol{A}^{-1}\boldsymbol{u}}$$

$$(a\boldsymbol{I}_n + b\boldsymbol{1}_n\boldsymbol{1}_n^t)^{-1} = \frac{1}{a}\left[\boldsymbol{I}_n - \frac{b}{a+nb}\boldsymbol{1}\boldsymbol{1}^t\right]$$

4.3 행렬식(Determinant)

$$|\boldsymbol{A}^t| = |\boldsymbol{A}|$$

$$|c\boldsymbol{A}| = c^k|\boldsymbol{A}|$$

여기서 k,$\boldsymbol{A}$의 차원이다.

$$|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$$

$$|\boldsymbol{A}^{-1}| = |\boldsymbol{A}|^{-1}$$

$$|a\boldsymbol{I}_n + b\boldsymbol{1}_n\boldsymbol{1}_n^t| = a^{n-1}(a+nb)$$