Sample Size for MLMs

Sample Size for MLMs

May, 2021

Multilevel Model (MLM)

$$Y_{ij} = X_{ij}\beta + Z_{ij} b_i + \varepsilon_{ij}$$

$$j = 1, 2, \cdots, n_i$$
$$i = 1, 2, \cdots, m$$

$$\underset{q\times 1}{b_i} \sim MVG(0, \underset{q\times q}{D}) \; ; \; \underset{n_i\times 1}{\underset{\sim}{\varepsilon}} \sim G(\underset{\sim}{0}, \sigma^2 R_i(\theta))$$

$$\varepsilon_i \perp b_i$$

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \sim G(X_i\beta, Z_i D Z_i' + \sigma^2 R_i)$$

- $$\hat{\beta}(D, \sigma^2, \theta) = \left(\sum_{i=1}^{m} X_i' V_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{m} X_i' V_i^{-1} Y_i\right)$$

$$\text{where } V_i = Z_i D Z_i' + \sigma^2 R_i(\theta)$$

- $$\operatorname{Var} \hat{\beta} = \left(\sum_{i=1}^{m} X_i' V_i^{-1} X_i\right)^{-1}$$

$$\hat{\beta}_j / se_{\hat{\beta}_j} = t_j$$

$$= [A_1 \, A_{2j} \, Y] / [A_1]_{jj}^{1/2}$$

where $A_1 = \left(\sum X_i' V_i^{-1} X_i \right)^{-1}$

$$A_{2j} = \sum_{i=1}^{n} [X_i' V_i^{-1}]_{j\cdot}$$

Goal of sample size considerations

- Choose m, n_i's so that

$Pr(t_j > c) = 1-\beta$ for given values of β, σ^2, $\hat{\theta}$, and D for given matrices X_i, Z_i, $i = 1, \cdots, m$.

Steps

1. Specify model for sample size calculations.

CARTOON of actual model

Ex 1. ACTUAL

$$Y \sim \underline{s(t)} + \text{Covariates}$$

CARTOON

$$Y \sim t$$

estimate trend

EX 2. ACTUAL

$Y \sim s(t) * G + \text{Covariates}$

CARTOON

$Y \sim t * G$

estimate different trends across groups

Ex 3. ACTUAL

$$Y \sim (X - \overline{X}) + \overline{X} + \text{Covariates}$$

CARTOON

$$Y \sim (X - \overline{X}) + \overline{X}$$

estimate contextual effect of $\overline{Z}$

2. Steps to CALCULATE $t_j(m, n_2's)$ given $\beta, \sigma^2, D, \Theta$

2.1 Specify $\beta, \sigma^2, D, \Theta$

2.2 Repeat for several values of m, $n_2's$

(i) specify m and $F(n, \phi)$

(ii) simulate iid $n_i \sim F(n, \phi)$, $i = 1, \ldots, m$

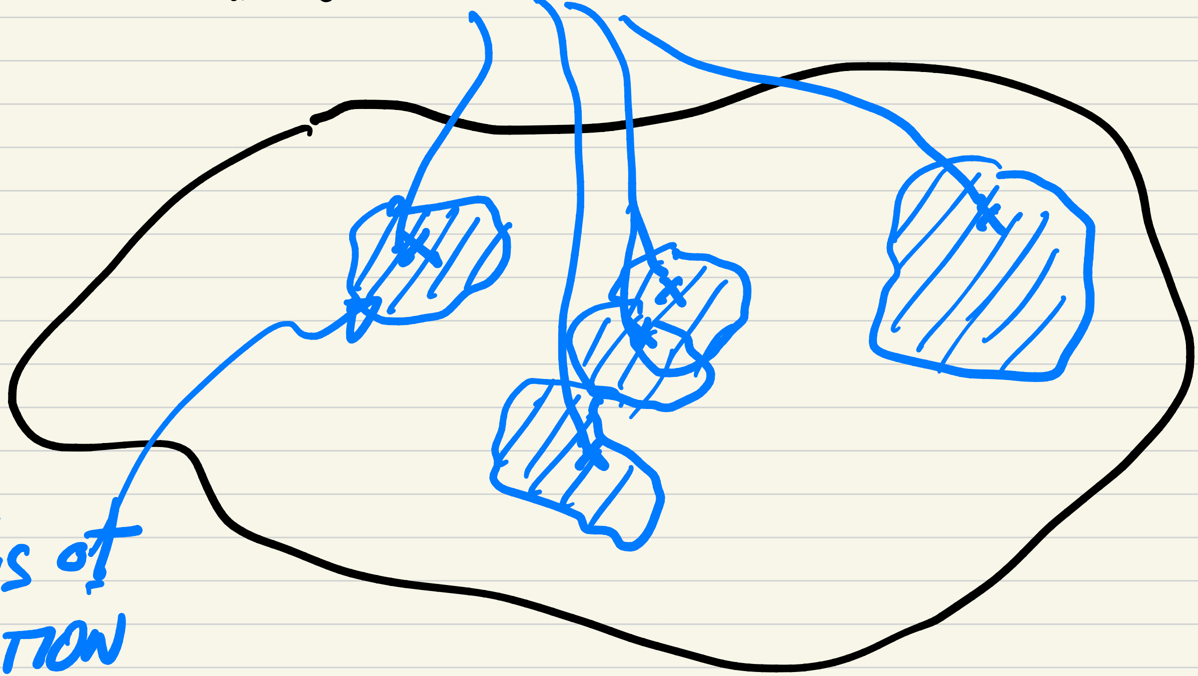
(iii) Calcute $V_i = Z_i D Z_i' + \sigma^2 R_i(\theta)$

(iv) Calculate $se_{\hat{\beta}} = \sqrt{\left[\left(\sum_{i=1}^{m} X_i' V_i^{-1} X_i\right)^{-1}\right]_{jj}}$

(v) Calculate $\hat{\beta}_j / se_{\hat{\beta}_j}$

JOHN W. TUKEY CONCEPT of "LEADING CASES"

SET of Problems



DOMAINS of APPLICATION

Simplest LEADING CASE (#1)

$$Y_{ij} = \beta_{0_i} + \varepsilon_{ij} \quad \Big| \quad , j = 1, \ldots, n_i; \; i = 1, \ldots, m$$

$$\beta_{0_i} = \beta_0 + b_i \quad \Big| \quad b_i \overset{iid}{\sim} G(0, \tau^2); \; \varepsilon_{ij} \overset{iid}{\sim} G(0, \sigma^2)$$

$$\underset{n_i \times n_i}{\operatorname{Var} \underset{\sim}{Y_i}} = \underset{n_i \times n_i}{\begin{pmatrix} \sigma^2 + \tau^2 & & \sigma^2 \\ & \ddots & \\ \sigma^2 & & \sigma^2 + \tau^2 \end{pmatrix}} = \tau^2 I + \sigma^2 \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \ddots & & \vdots \\ 1 & \cdots & \cdots & 1 \end{pmatrix}$$

$$= \tau^2 I_{n_i} + \sigma^2 \underset{\sim}{1}_{n_i} \underset{\sim}{1}_{n_i}'$$

$$\hat\beta_0 = \frac{1}{W}\sum_{i=1}^{m} w_i \bar Y_i \,, \quad W = \sum_{i=1}^{m} w_i$$

where $w_i = \left(\tau^2 + \sigma^2/n_i\right)^{-1} = 1/\mathrm{var}\,\bar Y_{i\cdot}$

$\tau^2 = 0$: $\quad \hat\beta = \dfrac{1}{\sum_{i=1}^{m} \frac{1}{\tau^2+\sigma^2/n_i}} \sum_{i=1}^{m} \dfrac{1}{\tau^2+\sigma^2/n_i} \bar Y_{i\cdot}$

$$= \frac{1}{\sum \frac{n_i}{\sigma^2}} \sum_{i=1}^{m} \frac{n_i}{\sigma^2} \bar Y_i$$

$$= \bar Y_{\cdot\cdot}$$

equal weight to each observation

$\tau^2 \gg \sigma^2/n_i$: $\hat{\beta} = \dfrac{1}{\sum_{i=1}^{m} \frac{1}{\tau^2}} \sum_{i=1}^{m} \frac{1}{\tau^2} \bar{Y}_i$

$$= \frac{1}{m} \sum_{i=1}^{m} \bar{Y}_{i \cdot}$$

equal weight to each cluster mean

$$\operatorname{Var}\hat{\beta}_0 = (X'V^{-1}X)^{-1}$$

$$= \frac{1}{W} = \frac{1}{\sum_{i=1}^{m} \frac{1}{\tau^2 + \sigma^2/n_i}}$$

Suppose $n_i = n$ for all i

$$\operatorname{Var}\hat{\beta}_0 = \left(\frac{1}{W}\right)^2 \sum_{i=1}^{m} w_i^2 \operatorname{Var}\bar{Y}_{i\cdot}$$

$$= \left(\frac{1}{W}\right)^2 \sum_{i=1}^{m} w_i^2 \cdot \frac{1}{w_i} = \left(\frac{1}{W}\right)$$

$$= \frac{1}{\sum_{i=1}^{m} \frac{1}{\tau^2 + \sigma^2/n_i}} = \frac{1}{\sum_i \left(\frac{n_i}{n_i \tau^2 + \sigma^2} \right)}$$

$$\tau^2 = 0 : \quad Var\,\hat{\beta}_0 = \frac{1}{\sum_i \frac{n_i}{\sigma^2}} = \frac{\sigma^2}{\sum n_i}$$

$$\tau^2 > 0 : \quad n_i = n$$

$$Var\,\hat{\beta}_0 = \frac{n\tau^2 + \sigma^2}{nm}$$

$$= \frac{\tau^2+\sigma^2}{mn}\left(\frac{n\tau^2+\sigma^2}{\tau^2+\sigma^2}\right) = \frac{\tau^2+\sigma^2}{mn}\left(\frac{(\tau^2+\sigma^2)+(n-1)\tau^2}{\tau^2+\sigma^2}\right)$$

$$= \frac{\tau^2+\sigma^2}{mn}\left(1+(n-1)\frac{\tau^2}{\tau^2+\sigma^2}\right)$$

$$\boxed{= \frac{\operatorname{Var} Y_{ij}}{m \cdot n}\underbrace{\left(1+(n-1)\operatorname{Corr}(Y_{ij}, Y_{ik})\right)}_{\text{Design effect for 2-level random intercept model}}}$$

Design effects

$\rho = \mathrm{corr}(y_{ij}, y_{ik})$

n	0	.1	.5	.9
2	1	1.1	1.5	1.9
10	1	1.9	5.5	9.1
100	1	10.9	50.5	90.1

Sample size $m \cdot n$ must increase by design effect to offset within cluster correlation

LEADING CASE #3

$$Y_{ij} = \beta_{0i} + \beta_{1i} X_{ij} + \varepsilon_{ij}$$

$$\begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix} \overset{iid}{\sim} MVG\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \right)$$

$$\varepsilon_{ij} \quad iid \quad G(0, \sigma^2)$$

Rewrite

$$Y_{ij} = \beta_0 + \beta_1 (X_{ij} - \overline{X}_{i\cdot}) + \beta_1 \overline{X}_{i\cdot} + b_{0i} + b_{1i} X_{ij} + \varepsilon_{ij}$$

$2a: \; D_{12} = D_{22} = 0$ (RANDOM INTERCEPT only)

$$Y_{ij} = \beta_0 + b_{0i} + \beta_2 (X_{ij} - \bar{X}_{i\cdot}) + \beta_1 (\bar{X}_{i\cdot}) + \varepsilon_{ij}$$

$$\hat{\beta}_1 = \gamma \hat{\beta}_{1_W} + (1-\gamma) \hat{\beta}_{1_B}$$

where $$\hat{\beta}_{1_W} = \frac{\sum_{i,j} (X_{ij} - \bar{X}_{i\cdot})(Y_{ij} - \bar{Y}_{i\cdot})}{\sum (X_{ij} - \bar{X}_{i\cdot})^2}$$

$$\operatorname{Var} \hat{\beta}_{1_W} = \frac{\sigma^2}{\sum_{i,j} (X_{ij} - \bar{X}_{i\cdot})^2}$$

depends on σ^2 spread within

$$\hat{\beta}_{1A} = \frac{\sum_i \bar{Y}_{i\cdot}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})}{\sum_i (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2}$$

$$\operatorname{Var}\hat{\beta}_{1A} = \frac{\sum_{i=1}^{m} (\tau^2 + \sigma^2/n_i)}{\left(\sum_{i=1}^{m} (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})\right)^2}$$

depends on τ^2; spread among $\bar{X}_{i\cdot}$s

$$\gamma = \operatorname{Var}\hat{\beta}_{1B} / (\operatorname{Var}\hat{\beta}_{1W} + \operatorname{Var}\hat{\beta}_{1B})$$

LEADING CASE 2 - GENERAL

$$\underset{3\times 1}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_1 \end{pmatrix} ; \quad \underset{n_i\times 3}{X_i} = \begin{pmatrix} 1 & x_{i1} - \bar{x}_{i\cdot} & \bar{x}_{i\cdot} \\ 1 & \vdots & \vdots \\ \vdots & & \\ 1 & x_{in_i} - \bar{x}_{i\cdot} & \bar{x}_{i\cdot} \end{pmatrix}$$

$$\underset{n_i\times 2}{Z_i} = \begin{pmatrix} 1 & x_{i1} \\ 1 & \vdots \\ \vdots & \vdots \\ 1 & x_{in_i} \end{pmatrix}$$

$$\underset{n_i\times n_i}{Var(\underset{\sim}{Y_i})} = \underbrace{\underset{n_i\times 2}{Z_i}\ \underset{2\times 2}{D}\ \underset{2\times n_i}{Z_i'}}_{n_i\times n_i} + \underbrace{\sigma^2 I_{n_i}}_{n_i\times n_i}$$

$$\hat{\beta} = \left(\sum_{i=1}^{m} X_i' V_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{m} X_i' V_i^{-1} Y_i\right)$$

$$\operatorname{Var}\hat{\beta} = \left(\sum_{i=1}^{m} X_i' V_i^{-1} X_i\right)^{-1}$$

Calculate $\operatorname{Var}\hat{\beta}_{1W}$, $\operatorname{Var}\hat{\beta}_{1B}$ for different assumptions about

D, σ^2, $\sum_{i,j}(x_{ij}-\bar{x})^2$, $\sum(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$

e.g

(1) $\bar{X}_{i\cdot} \sim G(0, V_{X_B})$

$X_{ij} - \bar{X}_{i\cdot} \sim G(0, V_{X_W})$

(2) Choose m $(= 50)$

$n_i \sim \text{Poisson}(\mu_n = 4)$

(3) Choose σ^2, $\underset{\sim}{D} = (D_{11}, D_{12}, D_{22})$

(4) Calculate $\mathrm{Var}\,\hat{\beta}$