MA50260 Statistical Modelling

Lecture 16: Mixed Effects Models - Nested and Crossed Designs

Ilaria Bussoli

March 29, 2024

Mixed Effect Models

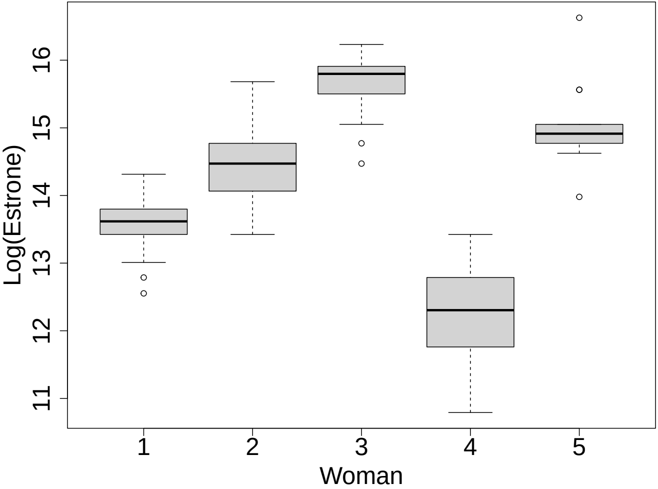
A mixed effects (normal) linear model is written as

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}\mathbf{b} + \underline{\epsilon}$$

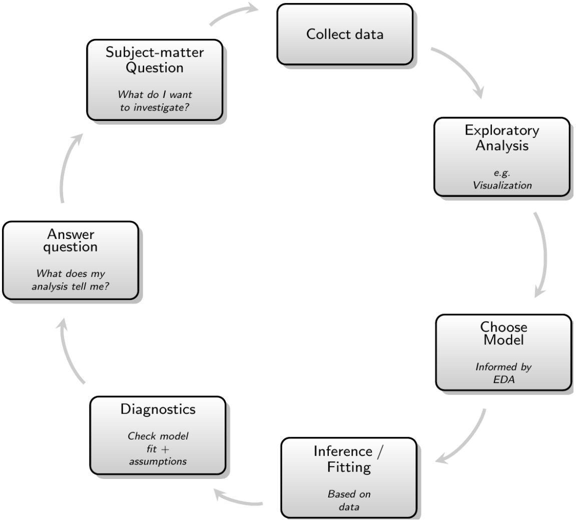
where

- $\mathbf{Y}$ is the vector of responses
- $\mathbf{X}$ is the design matrix
- $\underline{\beta}$ are the regression coefficients
- $\mathbf{Z}$ is the matrix of covariates associated to the q (unknown) random effects $\mathbf{b}$.
- $\underline{\epsilon}$ is the vector of residuals

Example - Analysis of Estrone Levels



Summary of Last Lecture



Nested and Crossed Effects

In several applications, we have multiple random effects.

How can we extend the introduced linear mixed model?

Nested and Crossed Effects

In several applications, we have multiple random effects.

How can we extend the introduced linear mixed model?

We generally distinguish between two types of random effects:

- **Nested effects:** levels of one factor are **completely contained** within the levels of another factor
- **Crossed effects:** levels of a factor vary across the levels of another

Example - Setup (I)

Consider a study with

- N machines
- P types of moulds that can be used on any machine
- n components being produced for each mould and each machine

Example - Setup (II)

Let $Y_{i,j,s}$ be the measurement for the s-th component from the j-th mould for the i-th machine.

$\Rightarrow$ $P \times n$ measurements for each machine.

Example - Setup (II)

Let $Y_{i,j,s}$ be the measurement for the s-th component from the j-th mould for the i-th machine.

$\Rightarrow P \times n$ measurements for each machine.

We are interested in the influence of machine and mould on the measurements:

- treat machine effect as random effect
- treat mould effect as random effect

Example - Nested Random Effects

Suppose that a different set of moulds is used on each machine.

We then model

$$Y_{i,j,s} = \mu + b_{1,i} + b_{12,ij} + \epsilon_{i,j,s},$$

where

- $b_{1,i} \sim \mathrm{Normal}\left(0, \sigma_M^2\right)$
- $b_{12,ij} \sim \mathrm{Normal}\left(0, \sigma_P^2\right)$
- $\epsilon_{i,j,s} \sim \mathrm{Normal}\left(0, \sigma_\epsilon^2\right)$

Note, the mould factor appears **only** within a particular level of the machine factor.

Example - Crossed Random Effects

Assume that the same moulds are used on each machine.

This is modelled as

$$Y_{i,j,s} = \mu + b_{1,i} + b_{2,j} + \epsilon_{i,j,s},$$

where

- $b_{1,i} \sim \mathrm{Normal}\left(0, \sigma_M^2\right)$
- $b_{2,j} \sim \mathrm{Normal}\left(0, \sigma_P^2\right)$
- $\epsilon_{i,j,s} \sim \mathrm{Normal}\left(0, \sigma_\epsilon^2\right)$

Note, mould effects are no longer specific to machines, but each mould effect b_j remains the same for all machine effects b_i.

Exercise 1

Measurements of NO_X across Bath at 4:00, 10:00, 16:00 and 22:00.

We have to consider

- Location
- Time of day

Exercise 1

Measurements of NO_X across Bath at 4:00, 10:00, 16:00 and 22:00.

We have to consider

- Location
- Time of day

Nested or crossed design?

Exercise 2

Delivery time for pizza service.

We have to consider

- Variation across companies
- Variation across delivery drivers
- Time of Day

Exercise 2

Delivery time for pizza service.

We have to consider

- Variation across companies
- Variation across delivery drivers
- Time of Day

Nested or crossed design?

General Notation

For nested models, we can write

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_{12}\mathbf{b}_{12} + \underline{\epsilon},$$

where we use the double subscript to indicate that the random effects are nested.

General Notation

For nested models, we can write

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_{12}\mathbf{b}_{12} + \underline{\epsilon},$$

where we use the double subscript to indicate that the random effects are nested.

On the other hand, for crossed random effects, we can write

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_2\mathbf{b}_2 + \underline{\epsilon},$$

where $\mathbf{b}_1$ and $\mathbf{b}_2$ are **crossed random effects**.

General Notation

For nested models, we can write

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_{12}\mathbf{b}_{12} + \underline{\epsilon},$$

where we use the double subscript to indicate that the random effects are nested.

On the other hand, for crossed random effects, we can write

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_2\mathbf{b}_2 + \underline{\epsilon},$$

where $\mathbf{b}_1$ and $\mathbf{b}_2$ are **crossed random effects**.

We can also write these models in the general form

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}\mathbf{b} + \underline{\epsilon}.$$

The Story so far (I)

In the past weeks, we considered a range of models, including:

- Normal linear regression
- Logistic regression (Binomial GLM)
- Poisson regression
- Gamma/Exponential regression
- Ordinal regression
- Linear mixed models

The Story so far (II)

Linear models

$$Y_i \sim \mathrm{Normal}(\mathbf{x}_i^{\mathrm{T}}\underline{\beta}, \sigma^2)$$

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \underline{\epsilon}$$

Generalized linear models

$$\mathbb{E}(Y_i) = g^{-1}\left(\mathbf{x}_i^{\mathrm{T}}\underline{\beta}\right)$$

Y_i from exponential family

Linear mixed models

$$Y_{i,j} \sim \mathrm{Normal}\left(\mathbf{x}_i^{\mathrm{T}}\underline{\beta}, \sigma_b^2 + \sigma_\epsilon^2\right)$$

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}\mathbf{b} + \underline{\epsilon}$$

The Story so far (III)

Linear models
$$Y_i \sim \mathrm{Normal}(\mathbf{x}_i^{\mathrm{T}}\underline{\beta}, \sigma^2)$$
$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \underline{\epsilon}$$

Generalized linear models
$$\mathbb{E}(Y_i) = g^{-1}\left(\mathbf{x}_i^{\mathrm{T}}\underline{\beta}\right)$$
Y_i from exponential family

Linear mixed models
$$Y_{i,j} \sim \mathrm{Normal}\left(\mathbf{x}_i^{\mathrm{T}}\underline{\beta}, \sigma_b^2 + \sigma_\epsilon^2\right)$$
$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}\mathbf{b} + \underline{\epsilon}$$

Generalized linear mixed models

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

   $\rightarrow$ Gamma regression

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

   $\rightarrow$ Gamma regression

2. Restaurant reviews

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

   $\rightarrow$ Gamma regression

2. Restaurant reviews

   $\rightarrow$ Ordinal regression

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

   $\rightarrow$ Gamma regression

2. Restaurant reviews

   $\rightarrow$ Ordinal regression

3. Number of faults in a production line

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

   $\rightarrow$ Gamma regression

2. Restaurant reviews

   $\rightarrow$ Ordinal regression

3. Number of faults in a production line

   $\rightarrow$ Poisson regression

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

   $\rightarrow$ Gamma regression

2. Restaurant reviews

   $\rightarrow$ Ordinal regression

3. Number of faults in a production line

   $\rightarrow$ Poisson regression

4. Student performance across units

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

   $\rightarrow$ Gamma regression

2. Restaurant reviews

   $\rightarrow$ Ordinal regression

3. Number of faults in a production line

   $\rightarrow$ Poisson regression

4. Student performance across units

   $\rightarrow$ Linear mixed model

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

   $\rightarrow$ Gamma regression

2. Restaurant reviews

   $\rightarrow$ Ordinal regression

3. Number of faults in a production line

   $\rightarrow$ Poisson regression

4. Student performance across units

   $\rightarrow$ Linear mixed model

5. Wingspan of an albatross

Exercise

Decide which model to use to analyse the following responses:

1. Size of a car insurance claim

   $\rightarrow$ Gamma regression

2. Restaurant reviews

   $\rightarrow$ Ordinal regression

3. Number of faults in a production line

   $\rightarrow$ Poisson regression

4. Student performance across units

   $\rightarrow$ Linear mixed model

5. Wingspan of an albatross

   $\rightarrow$ Normal linear regression