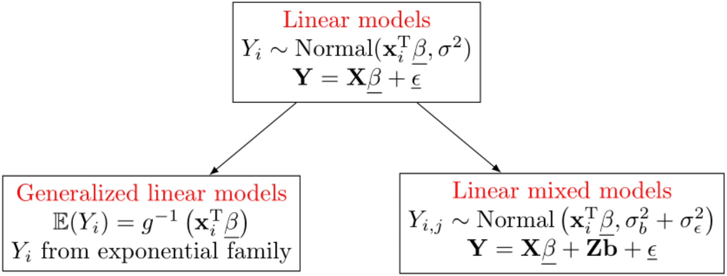
MA50260 Statistical Modelling

Lecture 17: Generalised Mixed Effects Models (GLMMs) - Introduction

Ilaria Bussoli

April 16, 2024

The Story So Far



What are the assumptions of the different types of models?

How would you model. . . ?

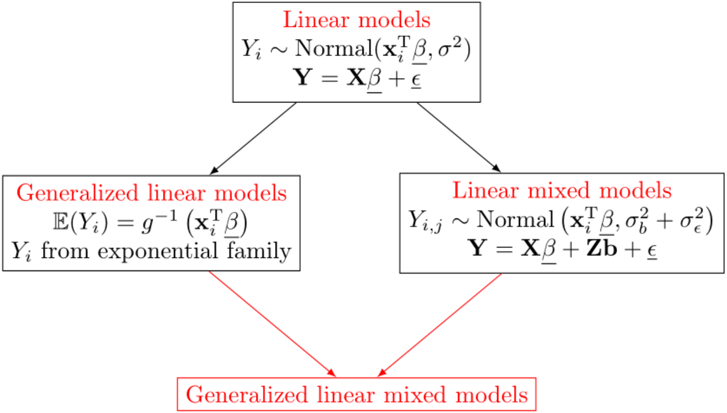
Decide which of the introduced models may be best for modelling

a. 5-year survival rate for a cancer treatment

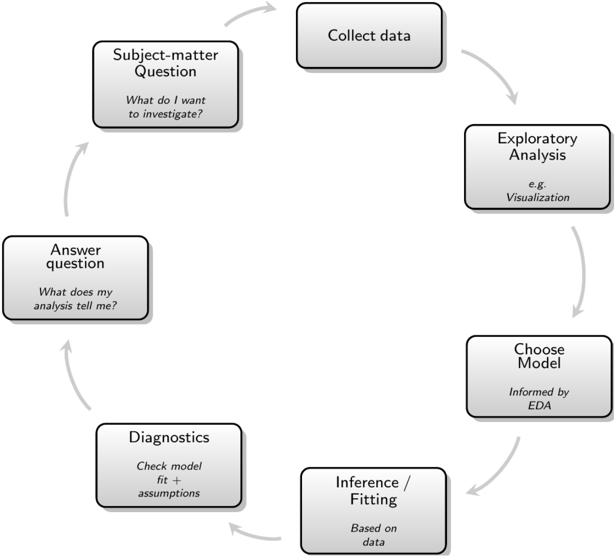
b. Daily temperatures across multiple UK cities

c. Covid-19 vaccine effectiveness across countries

The Next Step



Philosophy of Statistical Modelling



Recap: The Exponential Family

A random variable Y has a distribution in the exponential family if

$$f(y \mid \theta, \phi) = \exp\left\{ \frac{y\theta - d(\theta)}{a(\phi)} + c(y, \phi) \right\},$$

where θ and $\phi > 0$ are parameters; $a(\cdot)$, $d(\cdot)$ and $c(\cdot)$ are functions.

The exponential family includes:

- Normal(μ, σ^2)
- Poisson(μ)
- Binomial(m, p)

Generalised Linear Mixed Models (GLMMs)

Suppose we have I groups and J observations per group.

A GLMM generally comprises three components:

- **Linear predictor** $\eta_{i,j} = \mathbf{x}_{i,j}^{\mathrm{T}}\underline{\beta} + \mathbf{z}_{i,j}^{\mathrm{T}}\mathbf{b}$
- **Link function** $g(\mu_{i,j}) = \eta_{i,j}$ linking $\eta_{i,j}$ to $\mu_{i,j} = \mathbb{E}(Y_{i,j})$
- **Probability distribution** $Y_{i,j} \sim F\left(\mu_{i,j}\right)$ from the exponential family

Example 1 - Vaccine Effectiveness

Let's again consider Covid-19 vaccine effectiveness across countries.

How would we define a GLMM for this response?

Example 1 - Vaccine Effectiveness

Let's again consider Covid-19 vaccine effectiveness across countries.

How would we define a GLMM for this response?

3. The probability distribution is

$$Y_{i,j} \sim \mathrm{Bernoulli}(\mu_{i,j})$$

Example 1 - Vaccine Effectiveness

Let's again consider Covid-19 vaccine effectiveness across countries.

How would we define a GLMM for this response?

1. In the absence of any patient information, the linear predictor is

$$\eta_{i,j} = \beta_1 + b_i$$

3. The probability distribution is

$$Y_{i,j} \sim \mathrm{Bernoulli}(\mu_{i,j})$$

Example 1 - Vaccine Effectiveness

Let's again consider Covid-19 vaccine effectiveness across countries.

How would we define a GLMM for this response?

1. In the absence of any patient information, the linear predictor is

$$\eta_{i,j} = \beta_1 + b_i$$

2. A possible link function is

$$g(\mu_{i,j}) = \ln\left(\frac{\mu_{i,j}}{1-\mu_{i,j}}\right) = \eta_{i,j}$$

3. The probability distribution is

$$Y_{i,j} \sim \mathrm{Bernoulli}(\mu_{i,j})$$

Example 2 - Seed Germination (I)

In a study, 20 seeds were planted on each of 10 plates, and the number of germinated seeds is counted.

```
##    Plate Germinated Total NotGerminated
## 1      1          6    20            14
## 2      2          3    20            17
## 3      3         10    20            10
## 4      4         11    20             9
## 5      5         16    20             4
## 6      6          5    20            15
## 7      7          9    20            11
## 8      8          9    20            11
## 9      9          4    20            16
## 10    10         10    20            10
```

Example 2 - Seed Germination (II)

A suitable probability distribution is

Example 2 - Seed Germination (II)

A suitable probability distribution is

$$Y_{i,j} \sim \text{Bernoulli}(p_i)$$

If we assume $p_1 = \cdots = p_{10}$, the maximum likelihood estimate for np_i is

```
mean( seeds$Germinated )
```

```
## [1] 8.3
```

Example 2 - Seed Germination (II)

A suitable probability distribution is

$$Y_{i,j} \sim \text{Bernoulli}(p_i)$$

If we assume $p_1 = \cdots = p_{10}$, the maximum likelihood estimate for np_i is

```r
mean( seeds$Germinated )
```

```
## [1] 8.3
```

This suggests that the variance is about $n\hat{p}(1-\hat{p}) = 4.9$.

But:

```r
var( seeds$Germinated )
```

```
## [1] 15.12222
```

Example 2 - Seed Germination (III)

Let's define a GLMM with

$$Y_{i,j} \sim \text{Bernoulli}(p_i)$$

and

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_1 + b_i \qquad (i = 1, \ldots, I).$$

To complete the model, we define

$$\mathbf{b} = (b_1, \ldots, b_{10}) \sim \mathrm{MVN}_{10}\left(0\,,\, \sigma_b^2 \mathbf{I}_{10}\right).$$

Properties of GLMMs

- Conditional on $\mathbf{b}$, $Y_{i,j}$ is independently distributed with mean $\mu_{i,j}$ and variance $\phi\, V(\mu_{i,j})$.
- The **marginal variance** is

$$\begin{aligned} \mathrm{Var}(Y_{i,j}) &= \mathbb{E}\{\mathrm{Var}(Y_{i,j} \mid \mathbf{b})\} + \mathrm{Var}\{\mathbb{E}(Y_{i,j} \mid \mathbf{b})\} \\ &= \phi\, \mathbb{E}\{V(\mu_{i,j})\} + \mathrm{Var}(\mu_{i,j}). \end{aligned}$$

- The random effects $\mathbf{b}$ are usually assumed to be normally distributed for computational tractability.
- Random contributions to the predictor are no longer strictly additive.