STAT 902: Advanced Data Analysis

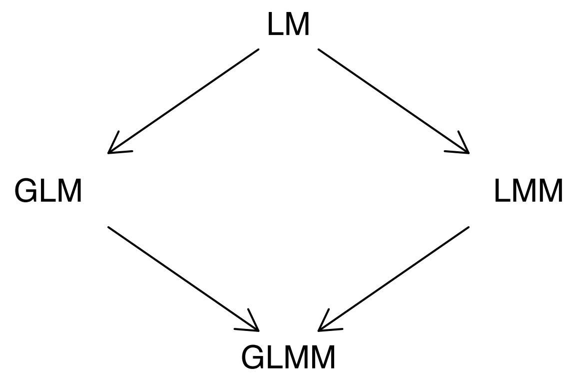
Generalised Linear Mixed Models

Major Statistical Model Families

Linear Model	(LM)
Generalised Linear Model	(GLM)
Linear Mixed Model	(LMM)
Generalised Linear Mixed Model	(GLMM)

The following graphic shows the diamond-type relationship between these families.

Arrow ($\longrightarrow$) means 'generalisation'.



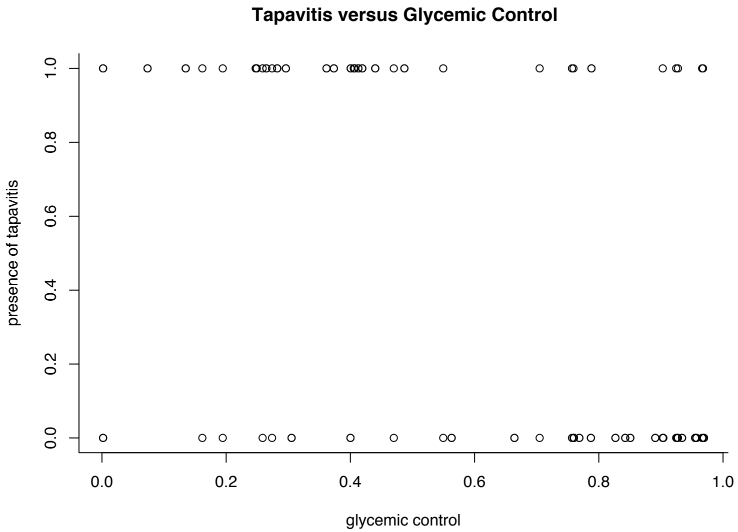
Motivating Example (albeit fictitious)

Tapavitis is an eye disease that is thought to be related to diabetes.

The following scatterplot shows data on 100 human eyes for the variables:

$$y = \begin{cases} 0 & \text{tapavitis absent} \\ 1 & \text{tapavitis present} \end{cases}$$

$$x = \text{glycemic control of patient}$$

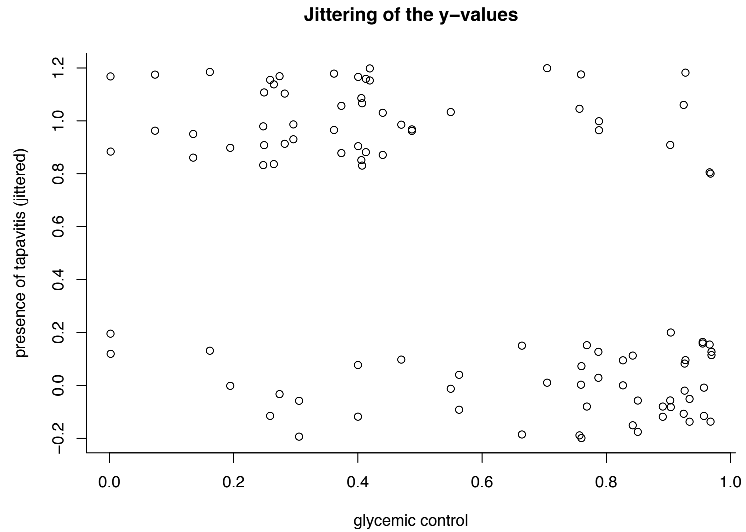


Enhancement via Jittering

We next apply

jittering

to the y-values to enhance visualisation.



Pairing Aspect

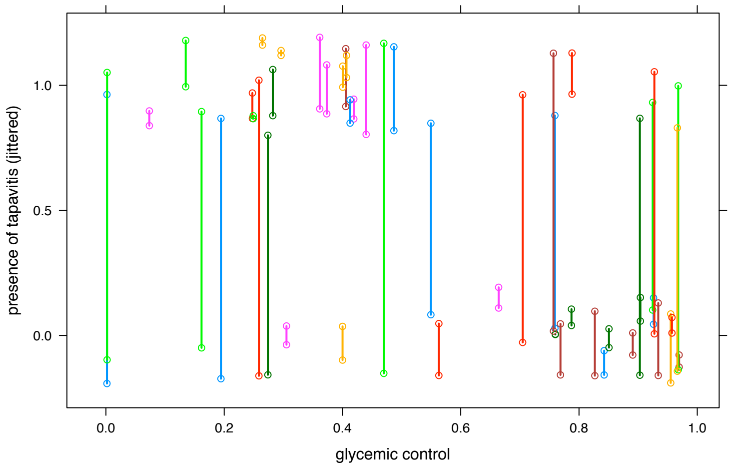
The 100 eyes correspond to only

50 patients

$\Longrightarrow$ pairing!

Pairing due to measurements on same person

(left eye, right eyed)



Notation for Repeated Measures

y_{ij}=jth Tapavitis measurement on ith patient.

$j = 1 \Longleftrightarrow$ left eye $\quad$ $j = 2 \Longleftrightarrow$ right eye

Logistic Regression with Random Intercept

$$\text{logit}\{P(y_{ij} = 1)\} = \beta_0 + \beta_1 x_i + U_i$$

$$U_i \overset{\text{ind.}}{\sim} N(0, \sigma_U^2)$$

GLM Reminder

For exponential family with canonical link:

$$\log[\mathbf{y};\beta,\phi] = \{\mathbf{y}^T\mathbf{X}\beta - \mathbf{1}^T b(\mathbf{X}\beta)\}/\phi + \mathbf{1}^T c(\mathbf{y},\phi)$$

Exponential Family Examples

name	canonical link	$b(\eta)$	$c(y,\phi)$	ϕ
Bernoulli	$\eta =$ logit(μ)	$\log(1+e^\eta)$	0	1
Poisson	$\eta = \ln(\mu)$	e^η	$-\ln(y!)$	1
$N(\mu,\sigma^2)$	$\eta = \mu$	$\eta^2/2$	$(y^2/\sigma^2 - \ln(2\pi\sigma^2))/2$	σ^2

Exponential Family GLM

$$\log[\mathbf{y};\beta,\phi] = \{\mathbf{y}^T\mathbf{X}\beta - \mathbf{1}^T b(\mathbf{X}\beta)\}/\phi + \mathbf{1}^T c(\mathbf{y})$$

GLMM Extension

$$\log[\mathbf{y}|\mathbf{u}] = \{\mathbf{y}^T(\mathbf{X}\beta+\mathbf{Z}\mathbf{u}) - \mathbf{1}^T b(\mathbf{X}\beta+\mathbf{Z}\mathbf{u})\}/\phi + \mathbf{1}^T c(\mathbf{y})$$

$$\mathbf{u} \sim N(\mathbf{0},\mathbf{G})$$

Maximum Likelihood Estimation

$$
\begin{aligned}
\mathcal{L}(\beta, \mathbf{G}, \phi) &= [\mathbf{y}; \beta, \mathbf{G}] \\
&= \int_{\mathbb{R}^q} [\mathbf{y}, \mathbf{u}] \, d\mathbf{u} \\
&= \int_{\mathbb{R}^q} [\mathbf{y}|\mathbf{u}][\mathbf{u}] \, d\mathbf{u} \\
&= (2\pi)^{-q/2} |\mathbf{G}|^{-1/2} \int_{\mathbb{R}^q} \exp[\{\mathbf{y}^T(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}) + \mathbf{1}^T c(\mathbf{y}) \\
&\qquad -\mathbf{1}^T b(\mathbf{X}\beta + \mathbf{Z}\mathbf{u})\}/\phi - \tfrac{1}{2}\mathbf{u}^T \mathbf{G}^{-1} \mathbf{u}\}] \, d\mathbf{u}
\end{aligned}
$$

GLMMs Big Headache

The likelihood involves an
intractable integral
(often high-dimensional).

Ways Around Intractable Integral Problem

- Analytic integral approximations.
  - Most popular has acronym PQL.
  - However often criticised for being 'too' approximate.
- Convert to Bayesianism and use MCMC.
  - Gets around the intractable integral headache.
  - But introduces new MCMC headaches (e.g. implementation, convergence, computing time).
  - However made easier since release of `BRugs` in 2005.

A Hot Topic

In some sense GLMM research started in about 1993.

Has been the topic of great deal of research since then.

Still ongoing...