MATH43515: Multilevel Modelling

Lecture 7: Generalized Linear Models

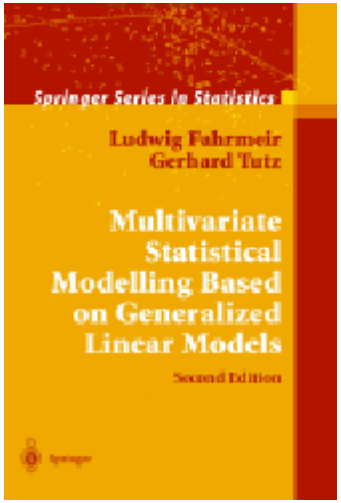
Module Convenor / Tutor: Andy Golightly

Durham University

Literature

This presentation follows essentially the developments in

Fahrmeir, L. and Tutz, G. Multivariate Statistical Modelling based on Generalized Linear Models, Springer, 2001.



... also available via the library and hereafter denoted by [FT].

Outline (Lecture 7)

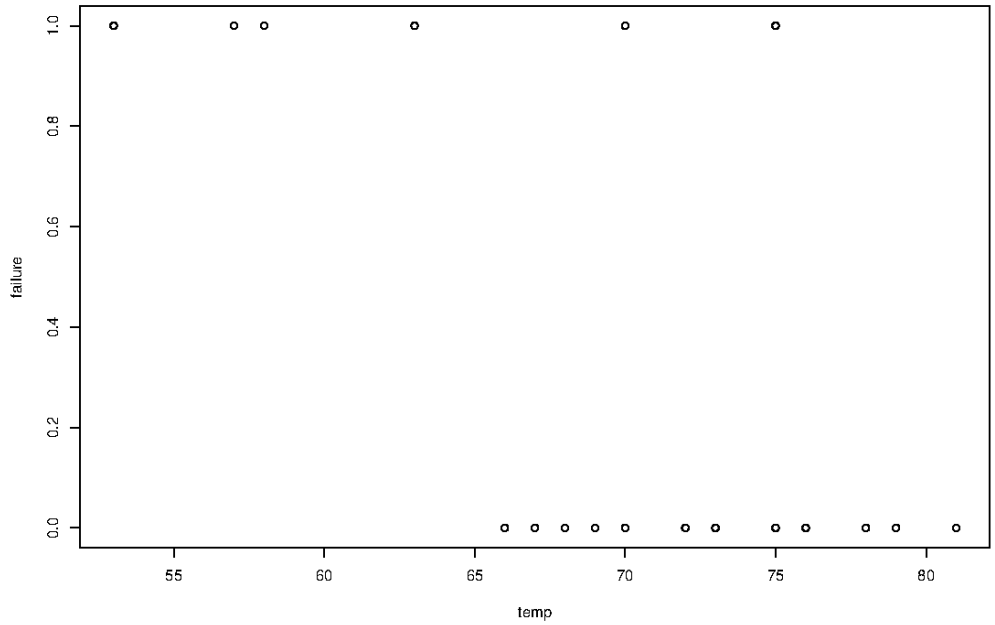
- Examples for data sets requiring generalized linear models
- Limitations of the linear model
- The exponential family of distributions
- Definition of the GLM
- Link functions, and the natural link
- Logistic regression, with R example
- Poisson regression, with R example
- Considerations on model fitting

General context: "Regression" (one level)

- One or more predictors $\boldsymbol{x} = (x_1, \ldots, x_p)$; one univariate response variable y.
- y and $x_1, \ldots, x_p$ may be continuous or categorical.
- Given: set of data $(\boldsymbol{x}_i, y_i)$, $i = 1, \ldots, n$.
- Aim: Find a model between predictors and responses
- Well known: The linear model, $y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i$ or equivalently, $E(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$ with usual assumptions $y_i \sim N(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma^2)$.
- Is the linear model capable to deal with all "one-level" problems?

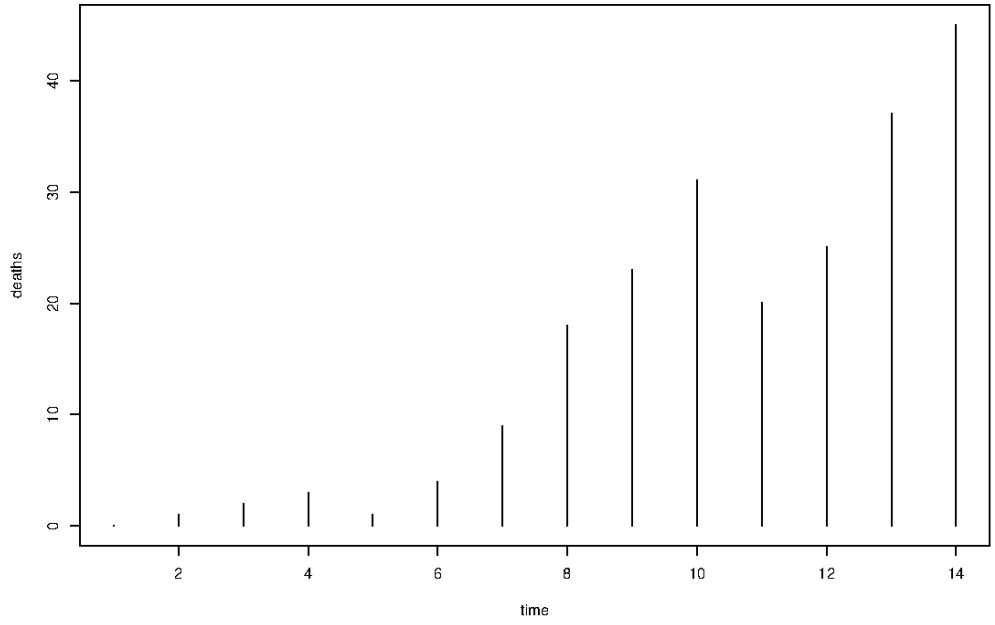
Example 1

Temperature ($^\circ F$) at time of take-off and occurrence/non-occurrence (1=yes, 0=no) of an "O-ring"-failure at the 23 U.S. space shuttle flights prior to the Challenger disaster of January 20, 1986.



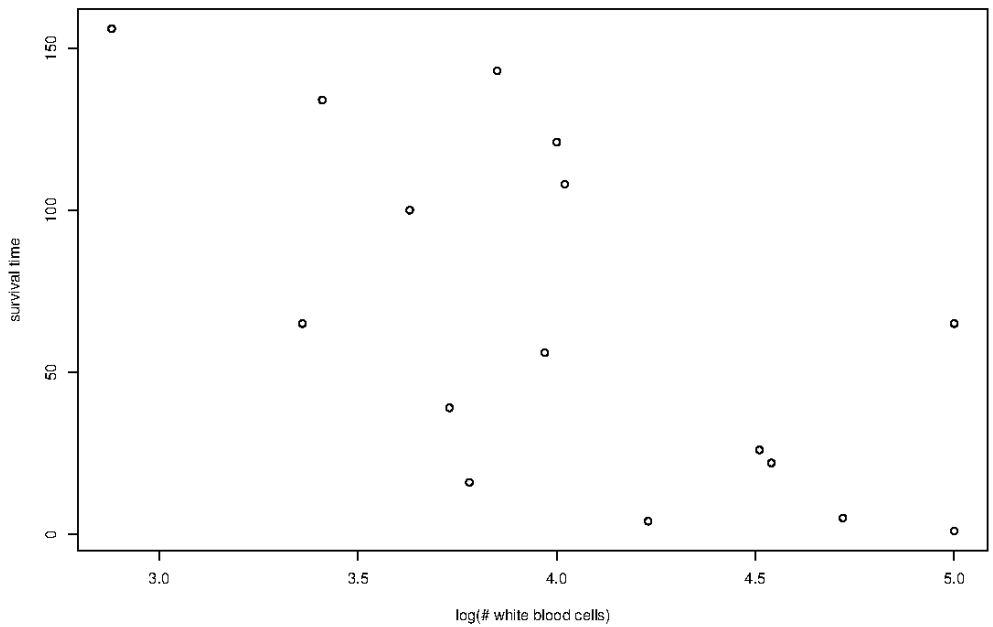
Example 2

Cases of death due to AIDS in Australia in 14 successive quarters from Jan 1983 to Jun 1986.



Example 3

For 17 leukaemia patients, the survival time after diagnosis is provided in dependence on of the base 10 logarithm of the white blood cell count (at time of diagnosis).



Limitations of the Linear Model

All three examples have in common...

- The range of the response variable is restricted

  Example 1: $\{0, 1\}$

  Example 2: $\mathbb{N}$

  Example 3: $\mathbb{R}^+$.

- The distribution of the responses is clearly non-normal, but rather

  Example 1: Binomial (Bernoulli)

  Example 2: Poisson

  Example 3: Gamma ?

Data of this type cannot be (adequately) dealt with by the Linear Model!

Exponential family

We wish to be able to model responses from a wide range of distributions. Almost all commonly used distributions are members of the exponential family

$$f(y|\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)\right\} \tag{1}$$

where

θ	is the natural parameter of the family,
ϕ	is a scale or dispersion parameter and
$b(\cdot)$ and $c(\cdot)$	are specific functions corresponding to the type of the family.

Normal:

$$\begin{aligned} f(y) &= \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\} \\ &= \exp\left\{\frac{y\mu-\mu^2/2}{\sigma^2}-\frac{y^2}{2\sigma^2}-\log\left(\sqrt{2\pi}\,\sigma\right)\right\}. \end{aligned}$$

Therefore,

- $\theta = \mu$,
- $b(\theta) = \theta^2/2 = \mu^2/2$,
- $\phi = \sigma^2$.

Bernoulli:

$$\begin{aligned} f(y) &= \pi^y (1-\pi)^{1-y} \\ &= \exp\left\{ y \log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi) \right\} \end{aligned}$$

where $\pi = P(y = 1)$ is the probability for "success". This is an exponential family with

- $\theta = \log(\pi/(1-\pi))$
- $b(\theta) = \log(1 + \exp(\theta)) = -\log(1-\pi)$,
- $\phi = 1$.

Other Exponential family members

- Binomial
- Poisson
- Geometric
- Negative Binomial
- Gamma and Exponential
- Inverse Gauss
- ⋮

See [FT], p. 21.

Table 2.1. Simple exponential families with dispersion parameter

$$f(y|\theta,\phi,\omega) = \exp\left\{\frac{y\theta - b(\theta)}{\phi}\omega + c(y,\phi,\omega)\right\}$$

(a) Components of the exponential family

Distribution		$\theta(\mu)$	$b(\theta)$	ϕ
Normal	$N(\mu,\sigma^2)$	μ	$\theta^2/2$	σ^2
Bernoulli	$B(1,\pi)$	$\log(\pi/(1-\pi))$	$\log(1+\exp(\theta))$	1
Poisson	$P(\lambda)$	$\log\lambda$	$\exp(\theta)$	1
Gamma	$G(\mu,\nu)$	$-1/\mu$	$-\log(-\theta)$	ν^{-1}
Inverse Gaussian	$IG(\mu,\sigma^2)$	$1/\mu^2$	$-(-2\theta)^{1/2}$	σ^2

(b) Expectation and variance

Distribution	$E(y)=b'(\theta)$	var. fct. $b''(\theta)$	$\text{var}(y)=b''(\theta)\phi/\omega$
Normal	$\mu=\theta$	1	σ^2/ω
Bernoulli	$\pi=\frac{\exp(\theta)}{1+\exp(\theta)}$	$\pi(1-\pi)$	$\pi(1-\pi)/\omega$
Poisson	$\lambda=\exp(\theta)$	λ	λ/ω
Gamma	$\mu=-1/\theta$	μ^2	$\mu^2\nu^{-1}/\omega$
Inverse Gaussian	$\mu=(-2\theta)^{-1/2}$	μ^3	$\mu^3\sigma^2/\omega$

Derivatives are denoted by $b'(\theta)=\partial b(\theta)/\partial\theta, b''(\theta)=\partial^2 b(\theta)/\partial\theta^2$. The weight ω is equal to 1 for individual ungrouped observations. For grouped data, y denotes the average of individual responses, the densities are scaled, and the weight ω equals the group size (i.e., the number of repeated observations in a group).

Let $\mu_i = E(y_i|\boldsymbol{x}_i)$. A generalized linear model is determined by

- the type of the exponential family which specifies the distribution of $y_i|\boldsymbol{x}_i$,
- the form of the linear predictor $\boldsymbol{x}_i^T\boldsymbol{\beta}$, i.e. the selection and coding of covariates,
- the transformation

$$\mu_i = h(\boldsymbol{x}_i^T\boldsymbol{\beta}), \quad \text{equivalently} \quad g(\mu_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}, \tag{2}$$

where

h	is a known one-to-one *response function*,
g	is the *link function*, i.e. the inverse of h.

The natural link

Several choices of the link function can be considered, but the most common choice is the natural or canonical link

$$g(\cdot) = \theta(\cdot)$$

which is uniquely determined by the respective exponential family, and can be directly read from the exponential family density. For instance, for Poisson response one has $g(\cdot) = \log(\cdot)$.

Use of the canonical link simplifies GLM methodology and theory tremendously.

For Gaussian response, $\theta(\cdot)$ is the identity function and so the natural link is the identity link. In other words, the usual Gaussian linear model is a special case of the generalized linear model!

For data $(\boldsymbol{x}_i, y_i)$, with $\{0, 1\}$-valued response, we are interested in modelling the probability of "success", $\pi_i = P(y_i = 1|\boldsymbol{x}_i)$.

One has

$$\pi_i = P(y_i = 1|\boldsymbol{x}_i) = E(y_i|\boldsymbol{x}_i) = \mu_i$$

so that the binary regression model can be generally formulated as

$$\pi_i = h(\boldsymbol{x}_i^T \boldsymbol{\beta})$$

Recall that for the Bernoulli distribution

$$f(y) = \exp\left\{y \log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi)\right\}$$

Hence, the natural link for the Bernoulli distribution is the so-called logit link

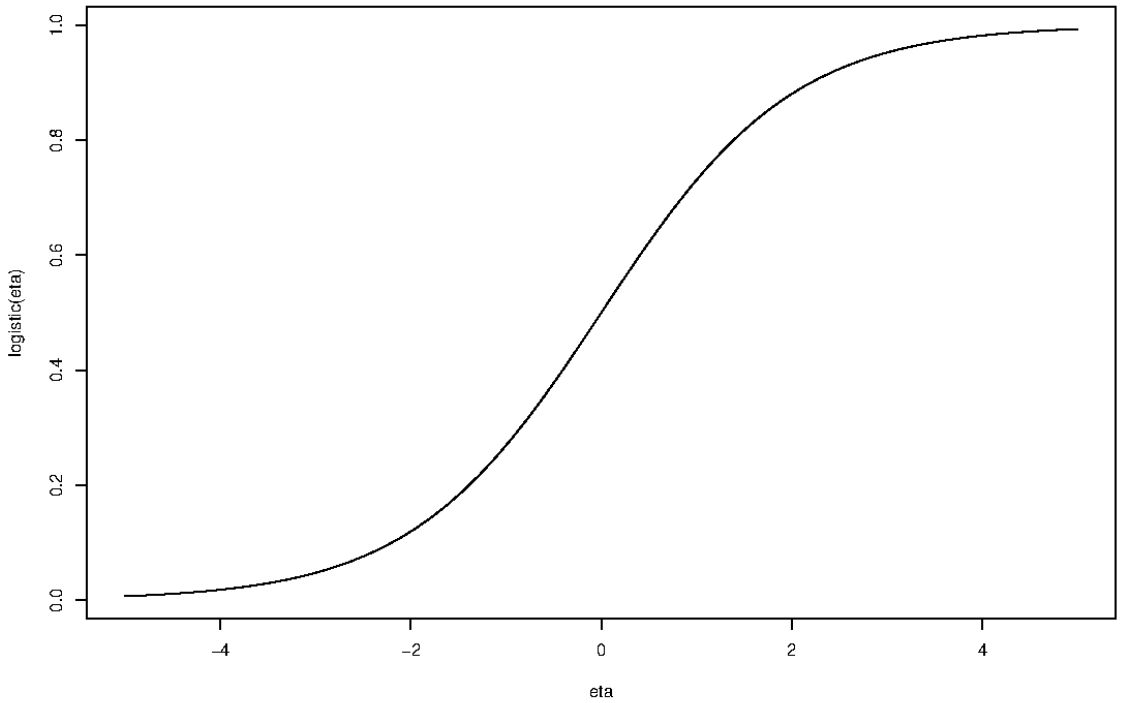
$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right),$$

and the corresponding response function is

$$h(\cdot) = g^{-1}(\cdot) = \frac{\exp(\cdot)}{1+\exp(\cdot)}.$$

Binary regression using the logit link is known as logistic regression.

Logistic response function $h(\boldsymbol{x}_i^T\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}$.



It shows clearly how the linear predictor range $\boldsymbol{x}_i^T\boldsymbol{\beta} \in (-\infty, \infty)$ is mapped to $(0, 1)$.

Back to Example 1 (Shuttle data)

Here $\boldsymbol{x}_i^T\boldsymbol{\beta} = \beta_0 + \beta_1 \text{temp}_i$, $y_i = 1$ if failure and 0 otherwise.

```
shuttle <-
  read.table("https://andygolightly.github.io/teaching/MATH43515/
  shuttle.asc", header=TRUE)
```

```
head(shuttle)
```

```
##   flight temp td
## 1      1   66  0
## 2      2   70  1
## 3      3   69  0
## 4      4   68  0
## 5      5   67  0
## 6      6   72  0
```

Back to Example 1 (Shuttle data)

```
glm1<- glm(td~temp, family=binomial(link=logit), data = shuttle)
glm1


##
## Call:  glm(formula = td ~ temp, family = binomial(link = logit)
##
## Coefficients:
## (Intercept)         temp
##     15.0429      -0.2322
##
## Degrees of Freedom: 22 Total (i.e. Null);  21 Residual
## Null Deviance:     28.27
## Residual Deviance: 20.32  AIC: 24.32
```

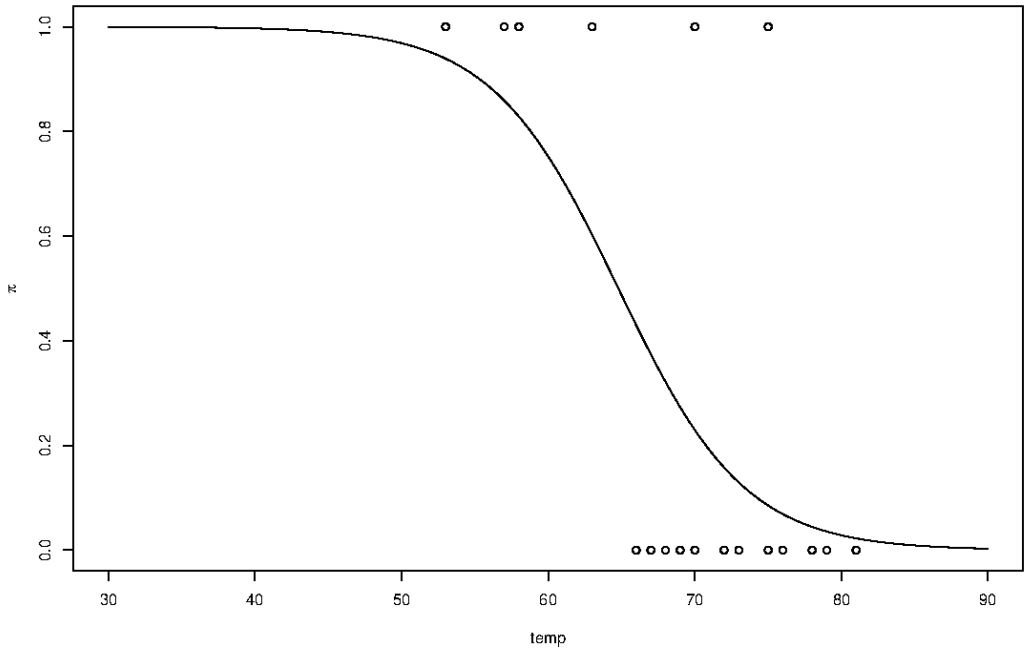
i.e. $\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}} = 15.0429 - 0.2322\text{temp}_i$.

Hence, the probability of failure in dependence of the temperature is

$$\hat{\pi} = \frac{\exp(15.0429 - 0.2322\text{temp})}{1 + \exp(15.0429 - 0.2322\text{temp})}.$$

Back to Example 1 (Shuttle data)

On the day of the accident, the temperature was 31F.




```
predict(glm1, newdata= data.frame(temp=31),
type="response")
##         1
## 0.9996088
```

The model fit would have told us that

$$\hat{P}(y = 1|x = 31) = \hat{\pi} = 0.9996088 \qquad (!!!)$$

Poisson response (count data)

For Poisson-distributed response (Example 2), one has density function

$$f(y) = \frac{\lambda^y}{y!} e^{-\lambda} = \exp\{y \log(\lambda) - \lambda - \log(y!)\},$$

where λ corresponds just to the expectation, μ, of f.

Therefore, one has the natural link $g(\lambda) = \log(\lambda)$, and the model takes the shape

$$\log(\lambda_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}.$$

or equivalently

$$\lambda_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$$

Example: Australian Aids data (Poisson response)

```
aids <- read.table("https://andygolightly.github.io/teaching/MATH43515/
aids.asc", header = TRUE)
```

```
summary(glm(deaths~time, family=poisson(link=log), data=aids))
```

```
##
## Call:
## glm(formula = deaths ~ time, family = poisson(link = log), data = aids)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.21008  -1.02032  -0.69704   0.04028   2.70758
##
## Coefficients:
```

Example: Australian Aids data (Poisson response)

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.33963    0.25119   1.352    0.176
## time         0.25652    0.02204  11.639   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 207.272  on 13  degrees of freedom
## Residual deviance:  29.654  on 12  degrees of freedom
## AIC: 86.581
##
## Number of Fisher Scoring iterations: 5
```

Guidelines for choice of distributions and link functions in a GLM

Response type	Exp. family	natural link	other links used
continuous	Normal	μ	
cont., positive	Gamma	$1/\mu$	$\log(\mu)$
count data	Poisson	$\log(\mu)$	
0-1 data	Bernoulli	$\log(\mu/(1-\mu))$	
proportions	Binomial	$\log(\mu/(1-\mu))$	

Notes:

- The `glm` function in R will by default use the natural link unless you tell it otherwise (using the `family` argument).
- Special case of the Gamma: The natural link $g(\mu) = 1/\mu$ for this distribution is often impractical as this restricts the range of the linear predictor to $\boldsymbol{x}_i^T\boldsymbol{\beta} > 0$ or $\boldsymbol{x}_i^T\boldsymbol{\beta} < 0$.

Model fitting for GLMs

Least Squares is not applicable (except for Gaussian response).

Estimation of β is based on the Maximum Likelihood (ML) method.

Mathematically, one works out the score equations (derivatives of the log-likelihood wrt β), and sets these equal to 0.

Generally not solvable analytically.

Therefore, a version of the Newton-Raphson algorithm is used for this, called Fisher Scoring.

One can show that this can be expressed as an iteratively weighted least squares problem [FT, p. 42].

Summary

GLMs are an extremely powerful device to fit models to a wide range of response types.

In practice, GLMs are not more difficult to fit than LMs — just use the `glm` function in R.

However, a bit of knowledge is needed to understand what the fitted models actually mean, and how the parameters needs to be interpreted.

Together with random effects, they constitute the most important development in statistical methodology in the last two decades of the 20th century.