Generalising the linear model

$$\mathbf{Y} = \alpha + \beta \mathbf{x}$$

Or equivalently:

- $\mathrm{E}[\mathbf{Y}] = \alpha + \beta \mathbf{x}$
- $\mathbf{Y} \sim \mathrm{N}(\mu, \sigma^2)$

The generalisation:

- $\mathbf{h}(\mathrm{E}[\mathbf{Y}]) = \alpha + \beta \mathbf{x}$ **(link function)**
- $\mathbf{Y} \sim \mathbf{f}(\mu, \ldots)$ **(error distribution)**

Generalised linear models

- $h(E[Y]) = \alpha + \beta x$ (Link function)
- $Y \sim f(\mu, \ldots)$ (error distribution)

- choice of link function makes linear dependence on explanatory variables more plausible
- choice of error distribution adds flexibility
- puts a very wide range of statistical methods under a common framework
- encourages open thinking (problem-driven rather than recipe-driven)

Residuals re-visited

Linear model

$$E[Y_i] = \alpha + \beta x_i = \mu_i \qquad Var(Y) = \sigma^2$$

Generalised linear model

$$E[Y_i] = h^{-1}(\alpha + \beta x_i) = \mu_i \qquad Var(Y) = v(\mu_i)$$

Residuals and standardised residuals

	Linear	Generalised linear
residual	$y_i - \hat{\mu}_i$	$y_i - \hat{\mu}_i$
standardised residual	$(y_i - \hat{\mu}_i)/\hat{\sigma}$	$(y_i - \hat{\mu}_i)/\sqrt{v(\hat{\mu}_i}$

The statistical modelling cycle

1. identify the question
2. design the study
3. collect the data
4. build the model
    1. exploration
    2. fitting
    3. diagnostic checking
    4. repeat (a) to (c) as necessary
5. answer the question

Examples of generalised linear models: open-ended count response

- **Poisson log-linear**

$$\log(\mu) = \alpha + \beta \mathbf{x} \qquad \mathbf{Y} \sim \text{Poiss}(\mu)$$

- **Extra-Poisson square-root linear**

$$\sqrt{\mu} = \alpha + \beta \mathbf{x} \qquad \text{Var}(\mathbf{Y}) = \phi \times \mu$$

Examples of generalised linear models: binary response

- **Logistic**

$$\log\{\mu/(1-\mu)\} = \alpha + \beta x \qquad \mu = P(Y=1)$$

- **Complementary log-log**

$$\log\{-\log(\mu)\} = \alpha + \beta x \qquad \mu = P(Y=1)$$