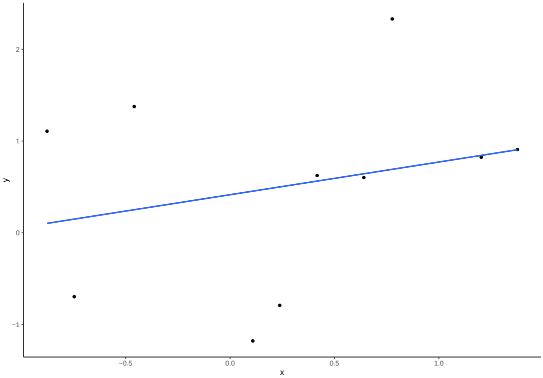
General & Generalized & Multilevel Linear Models

Mark Andrews
Psychology Department, Nottingham Trent University

✉ `mark.andrews@ntu.ac.uk`

Regression models



- Regression models are often introduced as fitting lines to points.
- This is a limited perspective that makes understanding more complex regression models, like generalized linear models, harder to grasp.

Regression models

- Put simply and generally, a regression model is a model of how the probability distribution of one variable, known as the *outcome* variable and other names, varies as a function of other variables, known as the *explanatory* or *predictor* variables.
- The most common or basic type of regression models is the *normal linear* model.
- In normal linear models, we assume that the outcome variable is normally distributed and that its mean varies linearly with changes in a set of predictor variables.
- By understanding the normal linear model thoroughly, we can see how it can be extended to deal with data and problems beyond those that it is designed for.

Normal linear models

- In a normal linear model, we have n observations of an outcome variable:
$$y_1, y_2 \dots y_i \dots y_n,$$
and for each y_i, we have a set of $K \geqslant 0$ explantory variables:
$$\vec{x}_1, \vec{x}_2 \dots \vec{x}_i \dots \vec{x}_n,$$
where $\vec{x}_i = [x_{1i}, x_{2i} \dots x_{ki} \dots x_{Ki}]^\mathsf{T}$.
- We model $y_1, y_2 \dots y_i \dots y_n$ as observed values of the random variables $Y_1, Y_2 \dots Y_i \dots Y_n$.
- Each Y_i, being a random variable, is defined by a probability distribution, which we model as conditionally dependent on $\vec{x}_i$.
- In notation, for convenience, we often blur the distinction between an (ordinary) variable indicating an observed value and, e.g. y_i, and its corresponding random variable Y_i.

Normal linear models

- In normal linear models, we model $y_1, y_2 \ldots y_i \ldots y_n$ as follows:

$$
\begin{aligned}
y_i &\sim N(\mu_i, \sigma^2), \quad \text{for } i \in 1 \ldots n, \\
\mu_i &= \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}
\end{aligned}
$$

- In words, each y_i is modelled a normal distribution, of equal variance σ^2, whose mean is a linear function of $\vec{x}_i$.
- From this model, for every hypothetically possible value of the K predictor variables, i.e. $\vec{x}_{i'}$, there is a corresponding mean $\mu_{i'}$, i.e. $\mu_{i'} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki'}$.
- If we change $x_{ki'}$ by Δ_k, then $\mu_{i'}$ changes by $\beta_k \Delta_k$.

The problem of binary outcome data

- What if our outcome variable is binary, e.g.,

$$y_1, y_2 \cdots y_i \cdots y_n,$$

  with $y_i \in \{0, 1\}$?
- Modelling $y_1, y_2 \ldots y_n$ as samples from a normal distribution is an extreme example of *model misspecifcation*.
- Instead, we should use a more appropriate model.
- The easiest way to do this is to use an extension of the normal linear model.

Logistic regression's assumed model

- For all $i \in 1 \ldots n$,

$$y_i \sim \text{Bernoulli}(\theta_i),$$

$$\text{logit}(\theta_i) = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki},$$

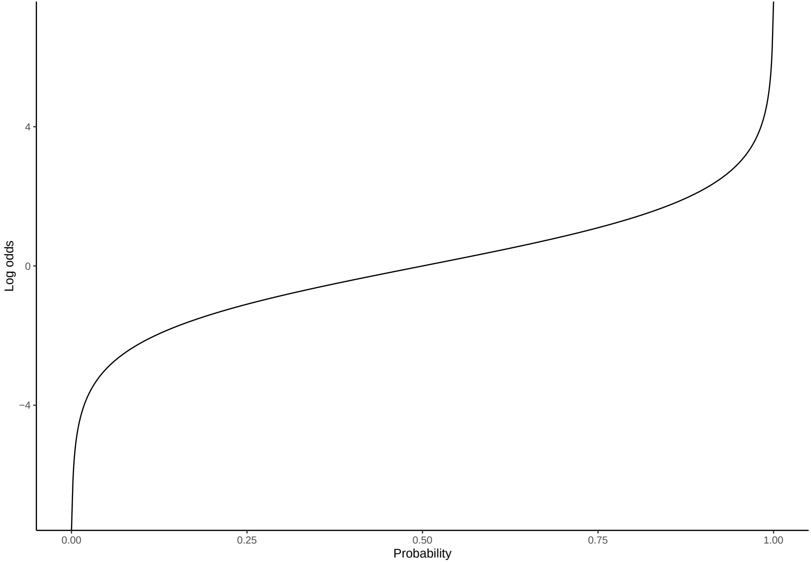
  where

$$\text{logit}(\theta_i) \doteq \log\left(\frac{\theta_i}{1 = \theta_i}\right).$$

- In other word, we are saying that each observed outcome variable value $y_1, y_2 \ldots y_n$ is a sample from a *Bernoulli* distribution with parameter θ_i, and the log odds of θ_i is a *linear* function of the $\vec{x}_i$.

Log odds (or logit)

- The log odds, or logit, is simply the logarithm of the odds.



Examples

```
affairs_df <- read_csv(here('data/affairs.csv')) %>%
  mutate(cheater = affairs > 0)

M <- glm(cheater ~ yearsmarried,
         family = binomial(link = 'logit'),
         data = affairs_df)
```

Model Fit with Deviance

- Once we have the estimates of the parameters, we can calculate *goodness of fit*.
- The *deviance* of a model is defined

$$-2 \log L(\hat{\beta}|\mathcal{D}),$$

  where $\hat{\beta}$ are the maximum likelihood estimates.
- This is a counterpart to R^2 for generalized linear models.

Model Fit with Deviance: Model testing

- In a model with K predictors ($\mathcal{M}_1$), a comparison "null" model ($\mathcal{M}_0$) could be a model with a subset $K' < K$ of these predictors.
- The difference in the deviance of the null model minus the deviance of the full model is

$$\Delta_D = D_0 - D_1 = -2 \log \frac{L(\hat{\beta}_0|\mathcal{D})}{L(\hat{\beta}_1|\mathcal{D})},$$

  where $\hat{\beta}_1$ and $\hat{\beta}_0$ are the maximum likelihood estimators of the models $\mathcal{M}_1$ and $\mathcal{M}_0$, respectively.
- Under the null hypothesis, Δ_D is distributed as χ^2 with $K - K'$ degrees of freedom.
- In other words, under the null hypothesis that subset and full models are identical, the difference in the deviances will be distributed as a χ^2 with df equal to the difference in the number of parameters between the two models.

Example

```r
M_1 <- glm(cheater ~ yearsmarried + age + gender,
           family = binomial(link = 'logit'),
           data = affairs_df)

# The "null" model
M_0 <- glm(cheater ~ yearsmarried,
           family = binomial(link = 'logit'),
           data = affairs_df)

anova(M_0, M, test = 'Chisq')
```

Multilevel models

- Multilevel models are a broad class of models that are applied to data that consist of sub-groups or clusters, including when these clusters are hierarchically arranged.
- A number of related terms are used to describe multilevel models: *hierarchical* models, *mixed effects* models, *random effects* models, and more.
- The defining feature of multilevel models is that they are *models of models*.
- In other words, for each cluster or sub-group in our data we create a statistical model, and then model how these statistical models vary across the clusters or sub-groups.

Linear mixed effects models

- A multilevel linear model with a single predictor variable is as follows.
$$\begin{aligned} \text{for } i \in 1 \ldots n, \quad & y_i \sim N(\mu_i, \sigma^2), \\ & \mu_i = \beta_{[s_i]0} + \beta_{[s_i]1} x_i, \\ \text{for } j \in 1 \ldots J, \quad & \vec{\beta}_j \sim N(\vec{b}, \Sigma). \end{aligned}$$
- Note that here the i index ranges over all values in the entire data-set, i.e. $i \in 1, 2 \ldots n$, and each $s_i \in 1, 2 \ldots J$ is an indicator variable that indicates the identity of the grouping variable for observation i.
- Using this new notation, given that $\vec{\beta}_j \sim N(\vec{b}, \Sigma)$, we can rewrite $\vec{\beta}_j$ as $\vec{\beta}_j = \vec{b} + \vec{\zeta}_j$ where $\vec{\zeta}_j \sim N(0, \Sigma)$.

- Substituting $\vec{b} + \zeta_j$ for $\vec{\beta}$, and thus substituting $b_0 + \zeta_{j0}$ and $b_1 + \zeta_{j1}$ for β_{j0} and β_{j1}, respectively, we have the following model.

$$
\begin{aligned}
\text{for } i \in 1 \dots n, \quad & y_i \sim N(\mu_i, \sigma^2), \\
& \mu_i = \underbrace{b_0 + b_1 x_i}_{\text{fixed effects}} + \underbrace{\zeta_{[s_i]0} + \zeta_{[s_i]1} x_i}_{\text{random effects}}, \\
\text{for } j \in 1 \dots J, \quad & \vec{\zeta}_j \sim N(0, \Sigma).
\end{aligned}
$$

- As we can see from this, a multilevel normal linear model is equivalent to a non-multilevel model (the *fixed effects* models) plus a normally distributed random variation to the intercept and slope for each subject (the *random effects*).

Example

```r
library(lme4)
M_ml <- lmer(Reaction ~ Days + (Days|Subject),
             data = sleepstudy)
```