Models for Binary Responses and Proportions
Binary Response Data

\( y_{ij} = 0 \) or \( 1 \) such as whether someone is or is not overweight

\[
\pi_{ij} = \Pr(y_{ij} = 1)
\]

\[
y_{ij} \sim \text{Binomial}(n_{ij}, \pi_{ij})
\]

\[
\text{Var}(y_{ij}) = \pi_{ij}(1 - \pi_{ij})/n_{ij}, \text{ where } n_{ij} = 1
\]

for binary data (Bernoulli trials).
Modelling Proportions

E.g. $y_{ij}$ is proportion of individuals overweight in electoral ward $i$ in constituency $j$ (i.e. aggregate data)

Denominator $n_{ij}$ could be number of adults in electoral ward $i$ in constituency $j$

After defining $n_{ij}$ we assume $y_{ij}$ has a binomial distribution with mean $\pi_{ij}$
Binary data: 2-Level Random Intercept Logit Model

\[ y_{ij} = \pi_{ij} + e_{ij}, \]  
where \( e_{ij} \) has mean 0 and variance \( \pi_{ij}(1 - \pi_{ij}) \);  
\( i = 1...n_j \) (level 1); \( j = 1..J \) (level 2)

\( x_{ij} \) is age of individual \( i \) in area \( j \), for example.

\[
\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 x_{ij} + u_j
\]

\( u_j \sim \text{Normal}(0, \sigma_u^2) \)
Other Link Functions

More generally, a model for binary responses can be written:

\[ f(\pi_{ij}) = \beta_0 + \beta_1 x_{ij} + u_j \]

where \( f(.) \) is logit (see above), probit, or complementary log-log function.
Interpretation: Fixed Part

\[ \log\left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \beta_0 + \beta_1 x_{ij} + u_j, \quad u_j \sim N(0, \sigma_u^2) \]

- \( \beta_0 \) is the log-odds that \( y = 1 \) when \( x = 0 \) and \( u = 0 \)
- \( \beta_1 \) is the effect on log-odds of a one unit increase in age for individuals in same area (same value of \( u \))
- \( \beta_1 \) often referred to as a cluster-specific effect of \( x \)
Interpretation: Random Part

\[
\log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \beta_0 + \beta_1 x_{ij} + u_j, \quad u_j \sim N(0, \sigma_u^2)
\]

- \(u_j\) is the effect of being in area \(j\) on the log-odds that \(y = 1\); also known as the level 2 residual
- As for continuous \(y\), we can obtain estimates and confidence intervals for \(u_j\)
- \(\sigma_u^2\) is the level 2 (residual) variance, or the between-area variance in the log-odds that \(y = 1\) after accounting for \(x\)
Response Probabilities from Logit Model

Response probability for individual $i$ in area $j$ calculated as:

$$\pi_{ij} = \frac{\exp(\beta_0 + \beta_1 x_{ij} + u_j)}{1 + \exp(\beta_0 + \beta_1 x_{ij} + u_j)}$$

Substitute estimates of $\beta_0$, $\beta_1$ and $u_j$ to get predicted probability $\hat{\pi}_{ij}$.

We can also make predictions for ‘typical’ individuals with particular values for $x$, but need to decide what to substitute for $u_j$ (see later).
Example: US Voting Intentions

\( y_{ij} = 1 \) if vote Bush in 2004 election, 0 other

Individuals (at level 1) within states (at level 2)

Results from null logit model (no x):

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 ) (intercept)</td>
<td>-0.107</td>
<td>0.049</td>
</tr>
<tr>
<td>( \sigma_u^2 ) (between-state variance)</td>
<td>0.091</td>
<td>0.023</td>
</tr>
</tbody>
</table>

- Should we accept or reject the hypothesis that \( \sigma_u^2 = 0 \)?
- Does \( \hat{\sigma}_u^2 = 0.091 \) represent a large state effect?
Testing the Level 2 Variance

- **Likelihood ratio test.** Only available if model estimated via maximum likelihood (not in MLwiN)

- **Wald test** (equivalent to t-test), but only approximate because variance estimates do not have normal sampling distributions

- **Bayesian credible intervals.** Available if model estimated using Markov chain Monte Carlo (MCMC) methods
Example of Wald Test: US Voting Data

Wald statistic = \left( \frac{\hat{\sigma}_u^2}{se} \right)^2 = \left( \frac{0.091}{0.023} \right)^2 = 15.65

Compare with $\chi^2_1$? reject $H_0$ that $\sigma_u^2 = 0$ and conclude there are state differences.

Take p-value/2 because $H_A$ is that $\sigma_u^2 > 0$ (one-sided).
State Effects on Probability of Voting Bush

Calculate $\hat{\pi}$ for ‘average’ states ($u = 0$) and for states 2 s.d. above and below the average ($u = \pm 2\hat{\sigma}_u$)

$\hat{\sigma}_u = \sqrt{0.091} = 0.302$

$u = -2\hat{\sigma}_u = -0.603 \quad \rightarrow \quad \hat{\pi} = 0.33$

$u = 0 \quad \rightarrow \quad \hat{\pi} = 0.47$

$u = +2\hat{\sigma}_u = +0.603 \quad \rightarrow \quad \hat{\pi} = 0.62$

Under a normal distribution, expect 95% of states to have $\hat{\pi}$ within (0.33, 0.62)
State Residuals with 95% Confidence Intervals
Adding Income as a Predictor

$x_{ij}$ is household income (9 categories) centred at mean of 5.23

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$ (intercept)</td>
<td>-0.099</td>
<td>0.056</td>
</tr>
<tr>
<td>$\beta_1$ (income effect)</td>
<td>0.140</td>
<td>0.008</td>
</tr>
<tr>
<td>$\sigma_u^2$ (between-state variance)</td>
<td>0.125</td>
<td>0.030</td>
</tr>
</tbody>
</table>

- 0.140 is effect on the log-odds of voting Bush of a 1-category increase in income, adjusting for state differences
- odds of voting Bush are $\exp(8 \times 0.14) = 3.1$ times higher for an individual in the highest income band than for individual in same state but lowest income band
- between state variance now higher
Prediction Lines by State: Random Intercepts
Latent Variable Representation of Binary Response Model

Consider latent continuous $y^*$ underlying observed binary $y$:

$$ y_{ij} = \begin{cases} 
1 & \text{if } y_{ij}^* \geq 0 \\
0 & \text{if } y_{ij}^* < 0 
\end{cases} $$

Threshold model:

$$ y_{ij}^* = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}^* $$

- $e_{ij}^* \sim N(0,1) \rightarrow$ probit model
- $e_{ij}^* \sim$ standard logistic (variance $\approx 3.29$) $\rightarrow$ logit model
Variance Partition Coefficient for Binary $y$

From threshold model for latent $y^*$ we obtain:

$$VPC = \frac{\sigma_{u}^2}{\sigma_{e^*}^2 + \sigma_{u}^2}$$

where $\sigma_{e^*}^2 = 1$ for probit model and 3.29 for logit model

In voting intentions example, VPC = 0.037 (for logit). Adjusting for income, about 4% of variance in propensity to vote Bush is attributable to state differences.
Impact of Adding $u_j$ on Coefficients

Single-level model: $y_i^* = \beta_0 + \beta_1 x_i + e_i^*$

$\text{var}(y_i^* \mid x_i) = \text{var}(e_i^*) = 3.29$

Multilevel model: $y_{ij}^* = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}^*$

$\text{var}(y_{ij}^* \mid x_{ij}) = \text{var}(u_j) + \text{var}(e_{ij}^*) = \sigma_u^2 + 3.29$

Adding $u_j$ increases residual variance, so scale of $y^*$ stretched out and $\beta_0$ and $\beta_1$ increase in absolute value.

Coefficients have different interpretation to single-level and other marginal models.
Marginal Model for Clustered y

An alternative to a random effects model is a marginal model, which has 2 components:

- A generalised linear model for relationship between $\pi_{ij}$ and $x_{ij}$

- Specification of structure of correlation between pairs of individuals in same group, e.g. exchangeable (equal correlation) or autocorrelation structure

Estimated using Generalised Estimation Equations (GEE)
Marginal vs. Random Effects Approaches

Marginal

- Clustering regarded as nuisance
- No parameter representing between-group variance
- No distributional assumptions about group effects (but no estimates either)

Random effects

- Clustering of substantive interest
- Estimate of between-group variance $\sigma_u^2$ and group effects
- Can allow between-variance to depend on $x$ (via random coefficients)
Marginal and Random Effects Models

Marginal $\beta$ have population-averaged (PA) interpretation
Random effects $\beta$ have cluster-specific (CS) interpretation

Random intercept logit model

$$\text{logit}(\pi_{ij}) = \beta_0^{CS} + \beta_1^{CS}x_{ij} + u_j, \text{ where } u_j \sim N(0, \sigma_u^2)$$

Marginal logit model

$$\text{logit}(\pi_{ij}) = \beta_0^{PA} + \beta_1^{PA}x_{ij}$$

with specification of structure of within-cluster covariance matrix
Interpretation of CS and PA Effects

Cluster-specific
- $\beta_i^{CS}$ is effect of $x$ on log-odds for a given cluster, i.e. holding constant (or conditioning on) cluster-specific unobservables
- Compares two individuals in same cluster with $x$-values 1 unit apart

Population-averaged
- $\beta_i^{PA}$ is effect of $x$ on log-odds in the study population, i.e. averaging over cluster-specific unobservables
Contraceptive Use in Bangladesh

The data are a sub-sample from the 1989 Bangladesh Fertility Survey (Huq & Cleland, 1990).

The binary response variable that we consider refers to whether a woman was using contraception at the time of the survey.

The aim of the analysis is to identify the factors associated with use of contraception and to examine the extent of between-district variation in contraceptive use. The data have a two-level hierarchical structure, with 2867 women nested within 60 districts.
Contraceptive Use in Bangladesh

• 2867 women nested in 60 districts

• $y=1$ if using contraception at time of survey, $y=0$ if not using contraception

• Covariates: age at level one (mean centred at 30), urban residence (vs. rural)
## Random Intercept Model

<table>
<thead>
<tr>
<th></th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed</strong></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-0.69 (0.08)</td>
</tr>
<tr>
<td>$\beta_1$ (urban)</td>
<td>0.71 (0.10)</td>
</tr>
<tr>
<td>$\beta_2$ (age)</td>
<td>0.015 (0.004)</td>
</tr>
<tr>
<td><strong>Random (between-district)</strong></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{u0}$</td>
<td>0.21 (0.06)</td>
</tr>
</tbody>
</table>
Interpretation of Random Part

$$\hat{\sigma}_{u_0} = \sqrt{0.21} = 0.46$$ and $$\exp(0.46) = 1.58$$, so for 1 SD increase in district-level unobservables we expect odds of using contraceptives to increase by 58%.

District effects (for a 30-year-old woman in a rural area):

<table>
<thead>
<tr>
<th></th>
<th>$$u_0 = -\hat{\sigma}_{u_0}$$</th>
<th>$$u_0 = 0$$</th>
<th>$$u_0 = \hat{\sigma}_{u_0}$$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“below average” district</td>
<td>“average”</td>
<td>“above average” district</td>
</tr>
<tr>
<td>$$\hat{\pi}$$</td>
<td>0.24</td>
<td>0.33</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Random Coefficient for Urban

\[
\logit(\pi_{ij}) = \beta_0 + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + u_{0j} + e_{ij}
\]

\[
\beta_{1j} = \beta_{11} + u_{1j}
\]
Random Coefficient for Urban

A subscript \( j \) has been added to the coefficient of urban, indicating that the coefficient depends on district. The average effect of urban is \( \beta_1 \), but the effect for district \( j \) is \( \beta_1 + u_{ij} \) where \( u_{ij} \) is a Normally distributed random effect with mean zero and variance \( \sigma_{u1}^2 \). Allowing the coefficient of urban to vary across districts also introduces a parameter representing the covariance between \( u_{0j} \) and \( u_{ij} \).
Random Coefficient for Urban: District-Level Covariance Matrix

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.297 (0.084)</td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>-0.205 (0.102)</td>
<td>0.390 (0.178)</td>
</tr>
</tbody>
</table>

District-level variance = \( \sigma^2_{u0} + 2\sigma_{u01} URBAN_{ij} + \sigma^2_{u1} URBAN_{ij}^2 \)

Rural: 0.297
Urban: 0.297 + 2(-0.205) + 0.390 = 0.277