

Generalised Linear Modelling Cheat Sheet

GLMs

Key Components

1. A set of response data. $E[Y_i] = \mu_i$
2. A set of predictors.
3. A data distribution $P(Y_i|X_i)$ that is the error component of the model
4. A **link** function g , $g(\mu_i) = x_i^T \beta = \eta_i$
 - (a) Monotone
 - (b) Continuous
 - (c) Differentiable
5. Independent observation assumption (observations are uncorrelated)

Common GLMs

1. Gaussian GLM
2. Logistic regression binary data
3. Logistic regression binomial data
4. Poisson regression
5. Gamma GLM
6. Multinomial GLM

Exponential Families

A member of exponential families has a pdf of the form:

$$f(y_i|\theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right\}$$

θ : natural parameter, represents location

ϕ : dispersion parameter, represents scale

$$E[Y] = \mu = b'(\theta)$$

$$Var[Y] = b''(\theta)\phi = V(\mu)\phi$$

Canonical Links

If the canonical link is chosen, then $X^T Y$ is sufficient for estimation of β .

$$\eta_i = g(\mu_i) = \theta = x_i^T \beta$$

Distribution	Link Type	Link Function
Normal(μ_i, σ^2)	identity	$\eta_i = \mu_i$
Poisson(μ_i)	log	$\eta_i = \log(\mu_i)$
Binomial(n, p_i)	logit	$\eta_i = \log\left(\frac{p_i}{1-p_i}\right) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$
Gamma(α, β_i)	reciprocal	$\eta_i = -\frac{1}{\mu_i} = -\frac{\beta_i}{\alpha}$

Family	E[Y]	Var(μ_i)	ϕ
Normal(μ_i, σ^2)	η_i	1	σ^2
Poisson(μ_i)	$\exp(\eta_i)$	μ_i	$\eta_i = \log(\mu_i)$
Binomial(n, p_i)	$n \left(\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right)$	$\frac{\mu_i(n - \mu_i)}{n}$	1
Gamma(α, β_i)	$-\frac{1}{\eta_i}$	μ_i^2	$\frac{1}{\alpha} \equiv z$

Weighted Least Squares(WLS)

High variability \rightarrow low weights.

Regression $\sqrt{w_i}y_i$ on $\sqrt{w_i}x_i$

$$WSS = \sum_{i=1}^n w_i^2 (y_i - X_i \hat{\beta})^2$$

$$\hat{\beta}_{WLS} = (X^T W X)^{-1} X^T W Y$$

Adjusted residual: $\sqrt{w_i} \hat{\epsilon}_i$

Iteratively Reweighted Least Squares(IRWLS)

$$g(y) \approx g(u) + (y - u)g'(u) = \eta + (y - u)\frac{d\eta}{d\mu}$$

$$\text{Delta Method: } Var[\widehat{g(y)}] \equiv \left(\frac{d\eta}{d\mu}\right)^2 V(Y)$$

1. set initial estimate $\hat{\eta}_0$ and $\hat{\mu}_0$
2. $z_0 = \hat{\eta}_0 + (y - \hat{\mu}_0)\frac{d\eta}{d\mu}|_{\hat{\eta}_0}$
3. $w_0^{-1} = \left(\frac{d\eta}{d\mu}\right)^2|_{\hat{\eta}_0} V(y)$
4. Regress $\sqrt{w_0}Z$ on $\sqrt{w_0}X$ and get the new $\hat{\eta}$
5. Iterate steps 2-4 until convergence

$$\widehat{Var}(\hat{\beta}) = \phi(X^T W X)^{-1}$$

An Example: Binomial(n,p)

- $\eta = \log \frac{\mu}{1-\mu}$
- $\frac{d\eta}{d\mu} = \frac{1}{\mu(1-\mu)}$
- $V(Y) = \mu(1-\mu) = np(1-p)$
- $w = n\mu(1-\mu)$

Linear combination of MLEs

c : vector describing the linear combination of β 's.

$$E[c^T \hat{\beta}] = c^T \beta$$

$$Var[c^T \hat{\beta}] = \phi c^T (X^T W X)^{-1} c$$

Confidence Interval for $g^{-1}(x_0^T \beta)$

Construct a confidence interval for $x_0^T \beta$, say (l, u)

Compute the interval: $(g^{-1}(l), g^{-1}(u))$

Make sure the confidence interval is inside the allowable range for the quantity in question

Deviance

$$deviance = Constant - 2\log(LMAX)$$

Likelihood Ratio Test

(Here the **saturated model** is a model with a parameter for every observation so that the data are fitted exactly.)

$$LRT = 2\log\Lambda = 2(\log L(\hat{\theta}_{sat}, \phi|y) - \log L(\hat{\theta}, \phi|y)) \sim \chi_v^2$$

v : difference in number of parameters between two models

$$LRT = deviance_{reduced} - deviance_{sat}$$

For Exponential Families

The deviance is in the form of:

$$D(Y, \hat{Y}) = 2\phi \log \Lambda = 2 \sum_i (y_i(\hat{\theta}_{sat} - \hat{\theta}) - b(\hat{\theta}_{sat}) + b(\hat{\theta}))$$

$$\text{Scaled Deviance: } D^*(Y, \hat{Y}) = \frac{D(Y, \hat{Y})}{\phi}$$

Scaled deviances are used in hypothesis tests.

Note: Expectation of χ_d is d . Variation is $\sqrt{2d}$

Goodness of Fit

H_0 : The model is an adequate fit to the data.

$$D^*(Y, \hat{Y}) = \frac{D(Y, \hat{Y})}{\phi} \sim \chi_{n-p} \text{ under } H_0$$

H_0 : The smaller model fits the model as well as the saturate model

ϕ must be known. Therefore only applicable to Poisson and Binomial GLMs, nor to quasi-likelihood models.

Not applicable to binary data.

Drop-in Deviance

Let β_L denote the extra parameters in the larger model. Models must have a nested structure.

$$H_0 : \beta_L = 0, H_\alpha : \beta_L \neq 0$$

Dispersion ϕ known

Chi-square Test

$$D(Y, \hat{Y}_S) - D(Y, \hat{Y}_L) \sim \chi_{df_S - df_L}^2$$

Binomial, Poisson models

Dispersion ϕ unknown

F Test

$$\frac{(D(Y, \hat{Y}_S) - D(Y, \hat{Y}_L)) / (df_S - df_L)}{\hat{\phi}_L} \sim F_{df_S - df_L, df_L}$$

Other than normal distributions, F-test is just an approximation.

Normal, Gamma, quasi-Binomial, quasi-Poisson models.

GLM Diagnostics

Response residual: $y_i - \hat{\mu}_i$

Pearson residual: $\frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$

Deviance residual: $sign(y_i - \hat{\mu}_i)\sqrt{\hat{d}_i}$

$$\sum d_i^2 = D(Y, \hat{Y})$$

$d_i = w_i\sqrt{\hat{D}_i}$, w_i^2 is the additional weight.

$$D_i = 2Y_i\{b(Y_i) - b(\hat{Y}_i)\} - 2\{c(Y_i) - c(\hat{Y}_i)\}$$

Normal(μ_i, σ^2)	$D_i = (Y_i - \hat{Y}_i)^2$
Binomial(n_i, π)	$D_i = Y_i \log\left(\frac{Y_i}{\hat{Y}_i}\right) + 2(n_i - Y_i) \log\left(\frac{n_i - Y_i}{n_i - \hat{Y}_i}\right)$
Poisson(λ)	$D_i = 2Y_i \log\left(\frac{Y_i}{\hat{Y}_i}\right) - 2(Y_i - \hat{Y}_i)$
Gamma(α, β)	$D_i = -2 \log\left(\frac{Y_i}{\hat{Y}_i}\right) + 2\left(\frac{Y_i - \hat{Y}_i}{\hat{Y}_i}\right)$

- Deviance Residuals v.s. Linear Predictor $\hat{\eta}$
 - Trends
 - Non-constant Variance
- Deviance Residuals v.s. Fitted values $\hat{\mu} = g^{-1}(\hat{\eta})$
- Response Residuals v.s. Linear Predictor $\hat{\eta}$
 - Variance should be non-constant
- Absolute Deviance Residuals v.s. Linear Predictor $\hat{\eta}$
- $g(Y_i)$ v.s. Linear Predictor $\hat{\eta}$:
Assess the link function
- Partial Residual Plots (Check for predictor transformations)
- Linearized Responses z v.s. Linear Predictor $\hat{\eta}$
 $z = \eta + (y - \mu) \frac{d\eta}{d\mu}$

Outlier Detection

Leverage

$$H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2}$$

$$h_i = H_{ii}$$

Cook's Distance

$$\frac{(\hat{\beta} - \hat{\beta}_{-i})(X^T W X)(\hat{\beta} - \hat{\beta}_{-i})}{p \hat{\phi}}$$

p is the number of parameters in the model

Usual Cutoff: **1** or **4/n**

Studentized Residuals

t distribution with n-p degrees of freedom.

$$r_{SD} = \frac{d_i}{\sqrt{\hat{\phi}(1-h_i)}}$$

Usual Cutoff: **2**

Jackknife Residuals

t distribution with n-p-1 degrees of freedom.

$$sign(y - \hat{y})\sqrt{(1-h_i)r_{SD}^2 + h_i r_{SP}^2}$$

$$r_{SP} = z_i / \sqrt{1-h_i} \quad (z_i: \text{Pearson residual})$$

Use half-normal on absolute residuals to find outliers.

Binomial/Binary GLMs

Link Functions

Logistic	$\eta = \text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
Probit	$\eta = \Phi^{-1}(\mu)$
Complementary log-log	$\eta = \log(-\log(1-\mu))$

Fisher Information Matrix

A way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ upon which the probability of X depends.

$$I(\beta) = -E\left[\frac{\partial \log L^2(\beta|y, X)}{\partial \beta \partial \beta^T}\right]$$

$$\widehat{Var}(\hat{\beta}) = I^{-1}(\hat{\beta})$$

Wald's Test

Test a single coefficient against normal distribution with estimated variation from Fisher Information Matrix.

Generally drop-in deviance test is more reliable than Wald's test.

Problematic Situations

- Predictors are collinear
- (Separation) Any linear combination of predictors is perfectly aligned with outcome.

Overdispersion

Overdispersion is a very common feature in applied data analysis because in practice, populations are frequently heterogeneous (non-uniform) contrary to the assumptions implicit within widely used simple parametric models.

How to identify overdispersion:

- Dependent responses.
- High variability for responses with the same values of covariates.
- Too many outliers.
- Goodness-of-fit test rejected. (Important explanatory variables absent)
- Overdispersion cannot exist for ungrouped binary data.
- Plot deviance residuals v.s. linear predictors.

$\hat{\phi}$ represents a multiplicative departure from binomial variation:

$$Var[Y_i|X_i] = \hat{\phi} n_i p_i (1-p_i)$$

Standardized(Pearson) residual

$$z_i = \frac{y_i - \hat{y}_i}{sd(\hat{y}_i)}$$

$$\text{Estimated overdispersion } \hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n z_i^2$$

Compare $\sum_{i=1}^n z_i^2$ to χ_{n-p}^2 to determine if overdispersion exists.

Adjusting inferences for overdispersion:

- Multiply all standard errors by $\sqrt{\hat{\phi}}$**
- The goodness-of-fit test is no longer applicable.
- Use F-test for drop-in deviance test statistics.

$$\frac{(D(Y, \hat{Y}_S) - D(Y, \hat{Y}_L)) / (df_S - df_L)}{\hat{\phi}_L} \sim F_{df_S - df_L, df_L}$$

- Use t-distribution as the reference for tests of significance of individual coefficient estimates.

Expanding binomial counts into binary counts **will not change the inferential results**, but will change the **residual deviance** (the saturated model is different), and the goodness-of-fit test is no appropriate.

Poisson GLM

$$P(Y = y) = \frac{e^{-\mu} \mu^y}{y!} = \exp(y \log(\mu) - \mu - \log(y!)), y = 0, 1, 2, \dots$$

- Rare events (If not rare, use normal distribution!)
- Occurrence of an event in a given time interval (proportional to the length of that time interval)

How to identify overdispersion:

- Check the binomial list of conditions.
- Events clustered or spaced unevenly through time.

Rate Models

$$y_i \sim \text{Poisson}(u_i \lambda_i)$$

u_i is the **exposure variable**. $\log(u_i)$ is called offset.

$$\log(\lambda_i) = \log\left(\frac{\mu_i}{u_i}\right) = X_i^T \beta$$

$$\log(\mu_i) = \log u_i + X_i^T \beta$$

Multinomial Models

Responses: J categories

$$P(Y_i, 1 = y_{i,1}, \dots, Y_{i,J} = y_{i,J}) = \frac{n_i!}{y_{i,1}! \dots y_{i,J}!} \pi_{i,1}^{y_{i,1}} \dots \pi_{i,J}^{y_{i,J}}$$

$$n_i = \sum_j y_{ij} \text{ and } \sum_j \pi_{i,j} = 1$$

Unordered Categories

Categories are nominal.

$$\log\left(\frac{\pi_{ij}}{\pi_{i1}}\right) = \eta_{ij} = x_i^T \beta_j \text{ for } j = 2, \dots, J$$

$$\pi_{ij} = \exp(\eta_{i,j} \pi_{i,1})$$

$$\pi_{i,1} = 1 - \sum_{j=2}^J \pi_{i,j} = 1 - \sum_{j=2}^J \exp(\eta_{i,j}) \pi_{i,1} = \frac{1}{1 + \sum_{j=2}^J \exp(\eta_{i,j})}$$

$$\pi_{i,j} = \frac{\exp(\eta_{i,j})}{1 + \sum_{j=2}^J \exp(\eta_{i,j})}$$

Ordered Categories

Categories are ordinal.

$$\gamma_{i,j} = p(y_i \leq j) \text{ where } \gamma_{i,J} = 1$$

$$g(\gamma_{i,j}) = \theta_j - x'_i \beta$$

Vector x_i does not include an intercept. β s do not depend on j
Let z_j to be a continuous latent(unobserved) variable.

$$\begin{aligned} \gamma_{i,j} &= P(y_i \leq j) = P(z_i \leq \theta_j) \\ &= P(z_i - x'_i \leq \theta_j - x'_i \beta) = F(\theta_j - x'_i \beta) \end{aligned}$$

Logistic Model / Proportional Odds Model

F follows logistic distribution.

$$\gamma_{i,j} = \frac{\exp(\theta_j - x'_i \beta)}{1 + \exp(\theta_j - x'_i \beta)}$$

$$\left(\frac{\gamma_{1,j}(x_1)}{1 - \gamma_{1,j}(x_1)} \right) / \left(\frac{\gamma_{2,j}(x_2)}{1 - \gamma_{2,j}(x_2)} \right) = \exp(-(x_1 - x_2)^T \beta)$$

Does not depends on j .

$$\text{logit}(\gamma_{i,j}) - \text{logit}(\gamma_{i,k}) = \theta_j - \theta_k$$

Should be near constant and does not have any trends.

Probit Model

F follows normal distribution.

$$\gamma_{i,j} = \Phi(\theta_j - x'_i \beta)$$

Proportional Hazards Model

Used in insurance industry.

$$\log(-\log(1 - \gamma_j(x_i))) = \theta_j - x_i^T \beta$$

$$\begin{aligned} \text{Hazard}(j) &= P(Y_i = j | Y_i \geq j) = \frac{P(y_i = j)}{Y_i \geq j} \\ &= \frac{\pi_{ij}}{1 - \gamma_{i,j-1}} = \frac{\gamma_{i,j} - \gamma_{i,j-1}}{1 - \gamma_{i,j-1}} \end{aligned}$$

$$F(\theta_j - x_i^T \beta) = 1 - \exp(-\exp(\theta_j - x_i^T \beta))$$

Gamma GLM

$$P(X = x) = \frac{\beta_i^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta_i x}$$

$\alpha > 0$: Shape $\beta_i > 0$: Rate

$$E[X_i] = \frac{\alpha}{\beta_i}, \text{Var}(X_i) = \frac{\alpha}{\beta_i^2}$$

Common Link Functions

$$\begin{array}{ll} \text{reciprocal} & \eta_i = -\frac{1}{\mu_i} = -\frac{\beta_i}{\alpha} \\ \text{log} & \eta_i = \log \mu_i = \log \frac{\alpha}{\beta_i} \\ \text{identity} & \eta_i = \mu_i = \frac{\alpha}{\beta_i} \end{array}$$

$$\hat{\phi} = CV = \frac{1}{n-p} \sum_{i=1}^n \left(\frac{Y_i - \hat{Y}_i}{\hat{Y}_i} \right)^2$$

CV: estimated coefficient of variation.

Contingency Table

- Used to show cross-classified categorical data on two or more variables.
- The variables can be **nominal** or **ordinal**.

Sampling Schemes for $R \times C$ Tables:

- Poisson** Distribution: None of the marginal totals are known.
 $Y_{ij} \sim \text{Poisson}(\lambda_{ij})$
- Multinomial** Distribution: Total sample size is known in advance.
 $P(Y_{ij} = y_{ij}) = \frac{n!}{\prod_i \prod_j y_{ij}!} \prod_i \prod_j p_{ij}^{y_{ij}}$
- Product Multinomial** Distribution: Row/column totals are fixed.

Hypotheses in interest:

- homogeneity**: Column totals are fixed.
 $H_0 : p_{i1} = p_{i2} = \dots = p_{iC} = p_i$ if column totals are the same.
- independence**: Row categorization is independent of the column categorization. ($p_{ij} = p_i \cdot p_j$)

Cannot test for independence with product multinomial models.

Multinomial Sampling Model

$$\mu_{ij} = E[Y_{ij} = np_{ij}]$$

$$\log L = \sum_i \sum_j y_{ij} \log p_{ij} + d(Y) = \sum_i \sum_j y_{ij} \log \frac{u_{ij}}{n} + d(Y)$$

Under independence $p_{ij} = p_i p_j$

$$\hat{p}_i = \sum_j \frac{y_{ij}}{n} = \frac{y_{i\bullet}}{n} \text{ and } \hat{p}_j = \sum_i \frac{y_{ij}}{n} = \frac{y_{\bullet j}}{n}$$

The fitted values for the independence model are

$$\mu_{ij}^* = np_i p_j = \frac{y_{i\bullet} y_{\bullet j}}{n}$$

$$D = 2 \sum_i \sum_j y_{ij} \log\left(\frac{y_{ij}}{\mu_{ij}^*}\right) = 2 \sum_i \sum_j O_{ij} \log\left(\frac{O_{ij}}{E_{ij}}\right)$$

O_{ij} : Observed count; E_{ij} : Expected count;

p-value: $O(\chi^2_{(R-1)(C-1)} \geq D)$

$$\log(1 + \delta) \approx \delta - \frac{1}{2}\delta^2$$

$$\frac{O_{ij}}{E_{ij}} = 1 + \delta_{ij} \Rightarrow \delta_{ij} = \frac{(O_{ij} - E_{ij})}{E_{ij}}$$

Pearson's χ^2 statistic:

$$2 \sum_i \sum_j O_{ij} \log\left(\frac{O_{ij}}{E_{ij}}\right) \approx \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \chi^2$$

Degrees of freedom: **(R-1)(C-1)**

Caution: The approximation is unreliable if a large majority of E_{ij} 's are less than 5.

Test of Homogeneity

$$p_i j = \frac{\mu_{ij}}{y_{\bullet j}} - p_i$$

$$E_{ij} = y_{\bullet j} \hat{p}_{i\bullet} = \frac{y_{i\bullet} y_{\bullet j}}{n}$$

Same E_{ij} 's from the independence model. \rightarrow test for homogeneity is the same as the test for independence!

Poisson Sampling Model

$$E[Y_{ij}] = n\pi_{ij} \text{ Pearson residuals:}$$

$$z_{ij} = \frac{Y_{ij} - \hat{Y}_{ij}}{\sqrt{V(\hat{Y}_{ij})}} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}}$$

$$\text{Pearson's } \chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_i \sum_j z_{ij}^2$$

Limitations of Pearson's χ^2

- Expected cell counts larger than 5**
- Not very informative. Only obtain p-value as output. Does not describe degree of dependence.
- The alternative hypothesis is very general.

Higher-Dimensional Tables of Counts

- Mutual Independence: $p_{ijk} = p_i p_j p_k$
- Joint Independence: $p_{ijk} = p_{ij} p_k$
- Conditional Independence:
 $p_{ij|k} = p_{i|k} p_{j|k} \Rightarrow p_{ijk} = p_{ik} p_{jk} / p_k$

Fit Poisson linear regression models, run drop-in deviance tests.

Multilevel Models

a.k.a. hierarchical linear models, nested models, mixed models

- Data are structured as observations within groups(i.e. nested data)
- Coefficient estimates in models may vary by group
- Account for variation within groups and across groups
- Account for dependencies in data due to grouping beyond what can be explained by group level predictors.
- hard to estimate group level variation for a small number of groups (e.g. < 5)
- Even one or two observations per group is acceptable.

Notation

Groups $j = 1, \dots, J$

n_j = number of observations in group j .

$j[i]$ codes group membership.

σ_y are data-level errors and σ_α are group-level errors.

Assumptions

- **Linearity:** The assumption of linearity states that there is a rectilinear (straight-line, as opposed to non-linear or U-shaped) relationship between variables.
- **Normality:** The assumption of normality states that the error terms at every level of the model are normally distributed.
- No homoscedasticity: units of observations in the same group are more similar than those in different groups.
- No independence of observations: while groups are independent of each other, observations within a group share values on variables, and thus, they are not independent.
- **Intraclass correlation:** assumes that data from the same context are more similar than data from different contexts.

Model: No predictors

Level 1: $y_i \sim N(\alpha_{j[i]}, \sigma_y^2) \ i = 1, \dots, n$

Level 2: $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \ j = 1, \dots, J$

$$\hat{\alpha}_j \approx \frac{\frac{n_j}{\sigma_y^2} \bar{y}_j + \frac{1}{\sigma_\alpha^2} \bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

$$n_j \rightarrow \infty \Rightarrow \hat{\alpha}_j \rightarrow \bar{y}_j$$

$$n_j \rightarrow 0 \Rightarrow \hat{\alpha}_j \rightarrow \bar{y}_{all}$$

$$\sigma_\alpha^2 / \sigma_y^2 \rightarrow \infty \Rightarrow \hat{\alpha}_j \rightarrow \bar{y}_j$$

$$\sigma_\alpha^2 / \sigma_y^2 \rightarrow 0 \Rightarrow \hat{\alpha}_j \rightarrow \bar{y}_{all}$$

Pooling

1. Complete Pooling:

All groups are identical ($\sigma_\alpha^2 = 0, \hat{\alpha}_j \rightarrow \bar{y}_{all}$)

2. No Pooling:

All groups are different ($\sigma_\alpha^2 \rightarrow \infty, \hat{\alpha}_j \rightarrow \bar{y}_j$)

Multilevel models is most important when it is close to complete pooling, at least for some of the groups. That is, σ_α is relatively small, and groups can borrow information from each other

Intraclass correlation: $\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_y^2}$

Value ranges from 0 (**grouping conveys no information**) to 1 (**all group members identical**).

$\frac{\sigma_\alpha^2}{\sigma_y^2}$ is also often used.

Model: Individual Level Predictors

$y_i \sim N(\alpha_{j[i]} + \beta_{X_i}, \sigma_y^2) \ i = 1, \dots, n$

$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \ j = 1, \dots, J$

`lmer(formula = y ~ x + (1 | group_factor))`

Model: Group Level Predictors

$y_i \sim N(\alpha_{j[i]} + \beta_{X_i}, \sigma_y^2) \ i = 1, \dots, n$

$\alpha_j \sim N(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2) \ j = 1, \dots, J$

Using group level predictors makes partial pooling more effective.

`lmer(formula = y ~ x + u + (1 | group_factor))`

Varying Intercepts and Slopes

$y_i \sim N(\alpha_{j[i]} + \beta_{j[i]X_i}, \sigma_y^2) \ i = 1, \dots, n$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right), j = 1, \dots, J$$

`lmer(formula = y ~ x + (1 + x | group_factor))`

Varying Slopes with Group Level Predictors

$y_i \sim N(\alpha_{j[i]} + \beta_{j[i]X_i}, \sigma_y^2) \ i = 1, \dots, n$

$$\begin{pmatrix} \gamma_0^\alpha + \gamma_1^\alpha u_j \\ \gamma_0^\beta + \gamma_1^\beta u_j \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right), j = 1, \dots, J$$

`lmer(formula = y ~ x + u + x:u + (1 + x | group_factor))`

Statistical Power

Power for level 1 effects is dependent upon the number of individual observations, whereas the power for level 2 effects is dependent upon the number of groups.

However, the number of individual observations in groups is not as important as the number of groups in a study. In order to detect cross-level interactions, given that the group sizes are not too small, recommendations have been made that **at least 20 groups** are needed.

	ϕ	w	$b(\theta)$	$c(y, \phi)$	$\mu = b'(\theta)$	$b''(\theta)$
Normal $Y \sim N(\theta, \phi)$	$ \phi $	1	$\theta^2/2$	$-(y^2/\phi + \log(2\pi\phi))/2$	θ	1
Poisson $Y \sim Po(e^\theta)$	1	1	e^θ	$-\log(y!)$	e^θ	μ
Binomial: $nY \sim Bi(n, e^\theta/(1 + e^\theta))$	1	n	$\log(1 + e^\theta)$	$\log \binom{n}{ny}$	$e^\theta/(1 + e^\theta)$	$\mu(1 - \mu)$
Gamma $Y \sim Ga(\nu, \lambda)^\dagger$	ϕ	1	$-\log(-\theta)$	$\nu \log \nu + (\nu - 1) \log y$ $-\log \Gamma(\nu)$	$-1/\theta$	μ^2

[†]pdf $f(y) = \lambda^\nu y^{\nu-1} e^{-\lambda y} / \Gamma(\nu)$ where $\lambda = -\theta/\phi$ and $\nu = 1/\phi$.