

Lecture 5

- Wei Wu, The University of Southern Mississippi
 - December, 2016

Acknowledgement: SESYNC Bayesian modeling for socio-environmental data workshop

Design Matrices and Experimental Data

Experimental designs expressed as joint distributions



$$[\alpha, \beta, \sigma | y] \propto \prod_{i=1}^N \text{normal}(y_{i,j} | g(\alpha, \beta, x_{i,j}), \sigma^2) \times \\ \text{normal}((\alpha | 0, 1000) \times \text{normal}((\beta_1 | 0, 1000) \times \\ \text{normal}((\beta_2 | 0, 1000) \times \text{normal}((\sigma | 0, 1000)$$

$$g(\alpha, \beta, x_{i,j}, w_{i,j}) = \alpha + \beta_1 x_{i,j} + \beta_2 w_{i,j}$$

$j = \text{group}(i)$ and it takes the value of 1 to M

Design matrix: What is this?

$$\begin{array}{c} E[y] \\ \left[\begin{array}{c} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \\ \vdots \\ \hat{y}_N \end{array} \right] \end{array} \quad \begin{array}{c} \text{Design Matrix} \\ = \left[\begin{array}{cc} 1 & 1.2 \\ 1 & 3.4 \\ 1 & 1.7 \\ 1 & 7.9 \\ & \vdots \\ 1 & 4.3 \end{array} \right] \end{array} \quad \begin{array}{c} \left[\begin{array}{c} \alpha \\ \beta \end{array} \right] \\ = \left[\begin{array}{c} \alpha + \beta \times 1.2 \\ \alpha + \beta \times 3.4 \\ \alpha + \beta \times 1.7 \\ \alpha + \beta \times 7.9 \\ \vdots \\ \alpha + \beta \times 4.3 \end{array} \right] \end{array}$$

Great! But how do we handle categorical experimental treatments?

Parameterize a model with categorical predictors

unique solution
cell means model

$$g(\boldsymbol{\alpha}, \mathbf{x}) = \alpha_j x_{i,j}^{(j)}$$

over-parameterized
effects model

$$g(\mu, \boldsymbol{\alpha}, \mathbf{x}) = \mu + \sum_{j=1}^M \alpha_j x_{i,j}^{(j)}$$

impose constraint

multilevel model

$$\alpha_j \sim N(0, \sigma^2)$$

set to zero

$$\alpha_M = 0$$

sum to zero

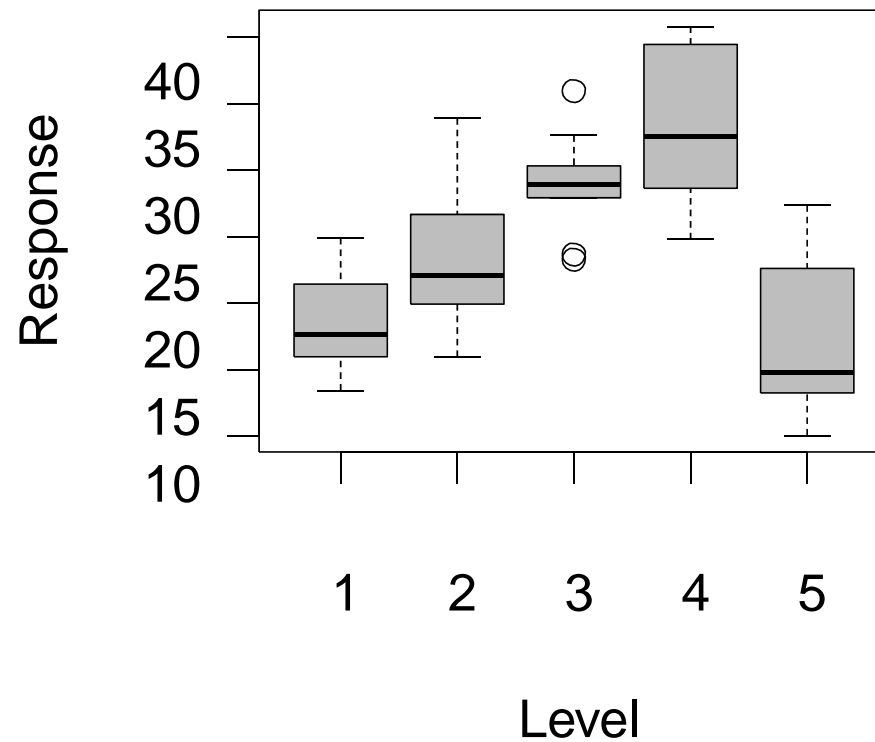
$$\alpha_M = -(\alpha_1 + \alpha_2 + \cdots + \alpha_{M-1})$$

Simulate data for CRD Design

Completely randomized design (CRD) with 1 factor and 5 levels

Simulate data for a factor with 5 levels

10 replicates per level, 50 replicates overall



Parameterize a model with categorical predictors

unique solution
cell means model

$$g(\boldsymbol{\alpha}, \mathbf{x}) = \alpha_j x_{i,j}^{(j)}$$

over-parameterized
effects model

$$g(\mu, \boldsymbol{\alpha}, \mathbf{x}) = \mu + \sum_{j=1}^M \alpha_j x_{i,j}^{(j)}$$

impose constraint

multilevel model

$$\alpha_j \sim N(0, \sigma^2)$$

set to zero

$$\alpha_M = 0$$

sum to zero

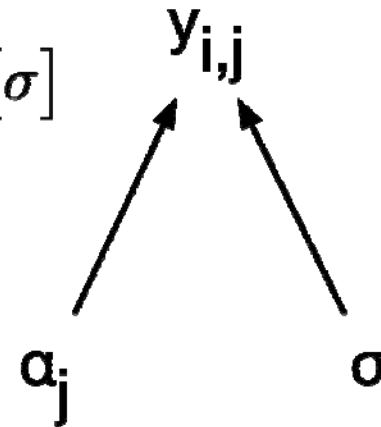
$$\alpha_M = -(\alpha_1 + \alpha_2 + \cdots + \alpha_{M-1})$$

Cell Means Model: Joint and DAG

$$[\boldsymbol{\alpha}, \sigma \mid \mathbf{y}] \propto \prod_{i=1}^N \left[y_{i,j} \mid g(\boldsymbol{\alpha}, \mathbf{x}), \sigma^2 \right] [\alpha_j] [\sigma]$$

$$g(\boldsymbol{\alpha}, \mathbf{x}) = \alpha_j x_{i,j}^{(j)}$$

$j = \text{group}(i)$ and it takes the value of 1 to M



Interest in group means and not effects

Have prior information for group means

Lack prior information for group means - use vague priors

Recover effects or grand mean as derived quantities

Cell Means Model: Design Matrix

$$\begin{array}{c} E[y] \\ \left[\begin{array}{c} \widehat{y_{n,1}} \\ \widehat{y_{n,2}} \\ \widehat{y_{n,3}} \\ \widehat{y_{n,4}} \\ \widehat{y_{n,5}} \end{array} \right] \end{array} = \begin{array}{c} \text{Design Matrix} \\ \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right] \end{array} \begin{array}{c} \text{Parameters} \\ \left[\begin{array}{c} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{array} \right] \end{array} = \left[\begin{array}{c} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{array} \right]$$

Cell Means Model: JAGS

```
#priors
for (i in 1:5) {
  alpha[i] ~ dnorm(0, 0.001)
}
sigma ~ dunif(0, 100)
tau <- 1 / ( sigma * sigma)

# Likelihood
for (i in 1:50) {
  y[i] ~ dnorm(alpha[x[i]], tau)
}

# Derived quantities
effect.2.1 <- mu[2] - mu[1]
effect.3.1 <- mu[3] - mu[1]
grandMean <- mean(mu[])
```

Compute effects and grand mean as derived quantities.
mu = alpha

Parameterize a model with categorical predictors

unique solution
cell means model

$$g(\boldsymbol{\alpha}, \mathbf{x}) = \alpha_j x_i^{(j)}$$

over-parameterized
effects model

$$g(\mu, \boldsymbol{\alpha}, \mathbf{x}) = \mu + \sum_{j=1}^M \alpha_j x_i^{(j)}$$

impose constraint

multilevel model

$$\alpha_j \sim N(0, \sigma^2)$$

set to zero

$$\alpha_M = 0$$

sum to zero

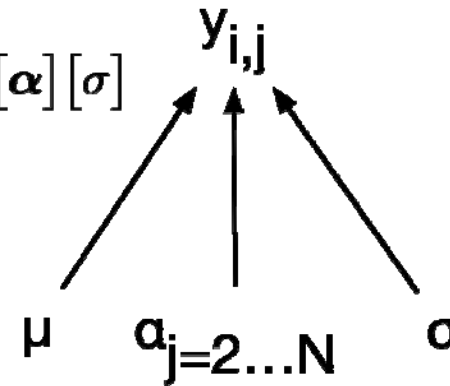
$$\alpha_M = -(\alpha_1 + \alpha_2 + \cdots + \alpha_{M-1})$$

Effects Models - Set to Zero: Joint and DAG

$$[\mu, \alpha, \sigma \mid \mathbf{y}] \propto \prod_{i=1}^N [y_{i,j} \mid g(\mu, \alpha, \mathbf{x}), \sigma^2] [\mu] [\alpha] [\sigma]$$

$$g(\mu, \alpha, \mathbf{x}) = \mu + \sum_{j=2}^M \alpha_j x_{i,j}^{(j)}$$

$j = \text{group}(i)$ and it takes the value of 1 to M



Interest in effects and not means

Have prior information for effect sizes

Lack prior information for effect sizes - can estimate conservatively

Recover group means as derived quantities

Effects Model- Set to Zero: Design Matrix

$$\begin{array}{c} \text{E[y]} \\ \widehat{y_{n,5}} \\ \widehat{y_{n,1}} \\ \widehat{y_{n,2}} \\ \widehat{y_{n,3}} \\ \widehat{y_{n,4}} \end{array} = \begin{array}{c} \text{Design Matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \text{Parameters} \\ \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} \end{array} = \begin{array}{c} \begin{bmatrix} \mu \\ \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_3 \\ \mu + \alpha_4 \end{bmatrix} \end{array}$$

CRD with 1 factor and 5 levels

Remove parameter by setting $\alpha_5 = 0$

Group 5 is now represented by intercept μ

α_j represent deviations from this baseline/control group

Effects Models - Set to Zero: JAGS

```
# Priors
for (i in 1:4){
  alpha[i] ~ dnorm(0, 0.001)
}
mu ~ dnorm(0, 0.001)
sigma ~ dunif(0, 100)
tau <- 1 / ( sigma * sigma)

# Likelihood
for (i in 1:50) {
  y[i] ~ dnorm(yhat[i], tau)
  yhat[i] <- mu + alpha[1]*treatment1[i] + alpha[2]*treatment2[i] + alpha[3]*treatment3[i] +
    alpha[4]*treatment4[i]
}

# Derived quantities
cell[5] <- mu
for (i in 1:4){
  cell[i] <- mu + alpha[i]
}
grandMean <- mean(cell[])
```

Compute cell and grand means as derived quantities

Parameterize a model with categorical predictors

unique solution
cell means model

$$g(\boldsymbol{\alpha}, \mathbf{x}) = \alpha_j x_i^{(j)}$$

over-parameterized
effects model

$$g(\mu, \boldsymbol{\alpha}, \mathbf{x}) = \mu + \sum_{j=1}^M \alpha_j x_i^{(j)}$$

impose constraint

multilevel model

$$\alpha_j \sim N(0, \sigma^2)$$

set to zero

$$\alpha_M = 0$$

sum to zero

$$\alpha_M = -(\alpha_1 + \alpha_2 + \cdots + \alpha_{M-1})$$

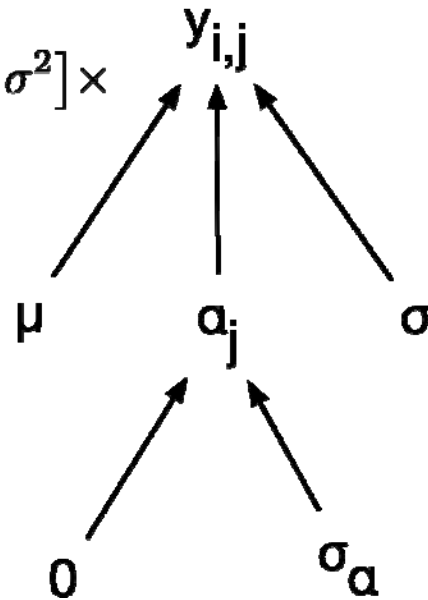
Effects Model - Multi-level: Joint and DAG

$$[\mu, \alpha, \sigma, \sigma_\alpha | \mathbf{y}] \propto \prod_{i=1}^N \left[y_{i,j} | g(\mu, \alpha, \mathbf{x}), \sigma^2 \right] \times$$

$$[\alpha_j | 0, \sigma_\alpha] [\mu] [\sigma] [\sigma_\alpha]$$

$$g(\mu, \alpha, \mathbf{x}) = \mu + \sum_{j=1}^M \alpha_j x_{i,j}^{(j)}$$

$j = \text{group}(i)$ and it takes the value of 1 to M



Interest in effects and not means

Have prior information for effect sizes

Lack prior information for effect sizes - can estimate conservatively

Recover group means as derived quantities

Effects Model - Multi-Level: Design Matrix

$$\begin{array}{c} \text{E[y]} \\ \begin{bmatrix} \widehat{y_{n,1}} \\ \widehat{y_{n,2}} \\ \widehat{y_{n,3}} \\ \widehat{y_{n,4}} \\ \widehat{y_{n,5}} \end{bmatrix} \end{array} = \begin{array}{c} \text{Design Matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \text{Parameters} \\ \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} \end{array} = \begin{array}{c} \begin{bmatrix} \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_3 \\ \mu + \alpha_4 \\ \mu + \alpha_5 \end{bmatrix} \end{array}$$

CRD with 1 factor and 5 levels

Intercept, μ , is the grand mean

α_j represent deviations from the grand mean

α_j are partially pooled allowing us to estimate all of them directly

Effects Models - Multi-level: JAGS

```
# Priors
mu ~ dnorm(0, 0.001)
for (i in 1:2){
  sigma[i] ~ dunif(0, 100)
  tau[i] <- 1 / ( sigma[i] * sigma[i])
}

# Likelihood
for (i in 1:5){
  alpha[i] ~ dnorm (0, tau[2])
}
for (i in 1:50) {
  y[i] ~ dnorm(y.hat[i], tau[1])
  y.hat[i] <- mu + alpha[x[i]]
}

# Derived quantities
for (i in 1:5){
  cell[i] <- mu + mean(alpha[i])
}
```

Compute cell means as derived quantities

Bayesian Approach to Experimental Analysis

Flexible framework

Ease of interpreting effects

Make statements like:

$\Pr(\text{Browsed} > \text{Unbrowsed}) = .8$

CI95: effect of browse = -4.0

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥ 0.1	

xkcd.com

Bayesian ANOVA

A way to summarize the “relative importance of different sources of variation in a dataset.” (Gelman and Hill, 2007)

Uses the finite-population SD and not the superpopulation SD

Can show variation decomposition across multiple levels

Unbalanced data and complex or incomplete designs easily handled

Can still be done with “fixed” effects

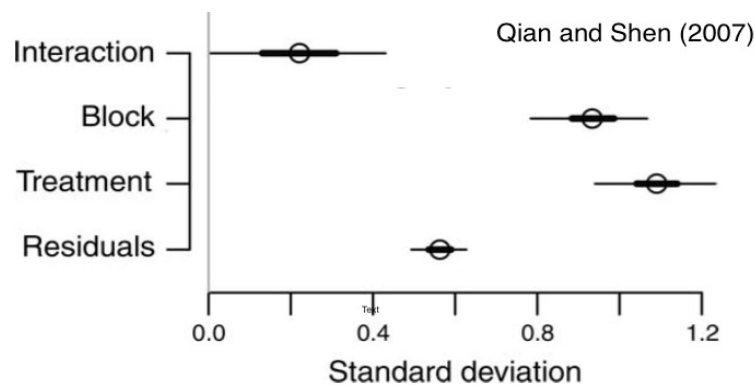


FIG. 1. Seaweed example, with ANOVA display of the estimated standard deviation of the estimated variance components showing a general pattern similar to that of the conventional ANOVA. Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

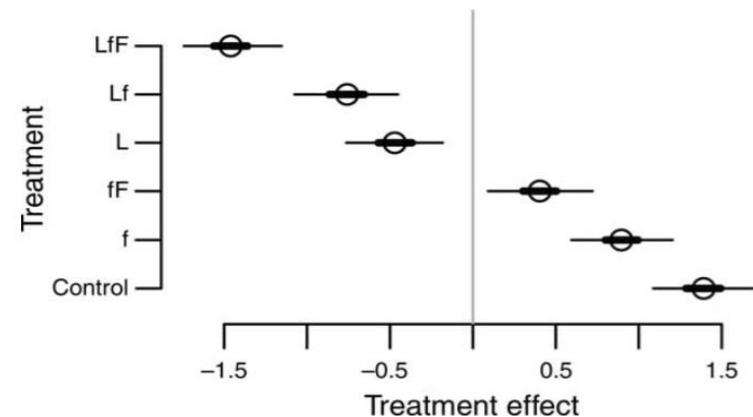


FIG. 2. Estimated treatment main effect of the seaweed grazer example shows that the regeneration rate decreases as grazing pressure increases. The six treatments are: C, control, no grazers allowed; L, only limpets allowed; f, only small fish allowed; Lf, large fish excluded; fF, limpets excluded; and LfF, all grazers allowed. The largest difference between treatments is

Bayesian ANOVA: JAGS

Compute finite-population SDs computation as derived quantities

```
# Priors
mu ~ dnorm(0, 0.001)
for (i in 1:2){
  sigma[i] ~ dunif(0, 100)
  tau[i] <- 1 / ( sigma[i] * sigma[i])
}

# Likelihood
for (i in 1:5){
  alpha[i] ~ dnorm (0, tau[2])
}
for (i in 1:50) {
  y[i] ~ dnorm(y.hat[i], tau[1])
  y.hat[i] <- mu + alpha[x[i]]
  s.yerr[i] <- y[i] - y.hat[i]
}

# Derived quantities
for (i in 1:5){
  cell[i] <- mu + mean(alpha[i])
}
s.alpha <- sd(alpha[])
s.y <- sd(y.err[])
```

Multiple Mean Comparison

Fundamentally different approach to mean comparisons
Shrinkage and/or informed priors

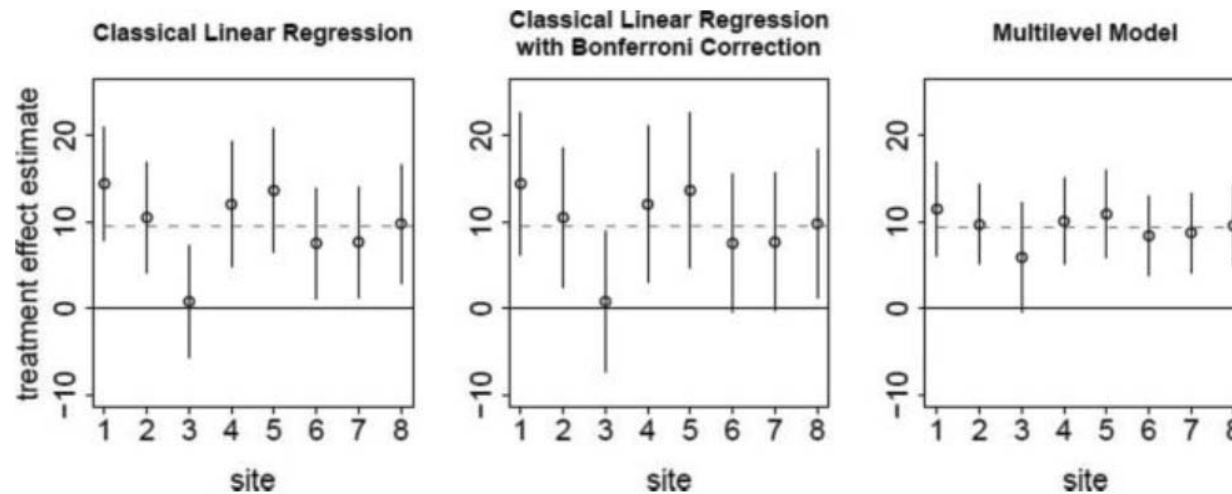


Figure 1. Treatment effect point estimates and 95% intervals across the eight Infant Health and Development Program sites. *Note.* The left panel display classical estimates from a linear regression. The middle panel displays the same point estimates as in the left panel but with confidence intervals adjusted to account for a Bonferroni correction. The right panel displays posterior means and 95% intervals for each of the eight site-specific treatment effects from a fitted multilevel model.

Gelman et al. 2012

Gelman 2013, Gelman et al. 2012

References

- [1] A. Gelman. Analysis of variance – why it is more important than ever. *Annals of Statistics*, 33(1):1–31, 2005.
- [2] A. Gelman and J. Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, Boston, MA, USA, 2007.
- [3] A. Gelman, J. Hill, and M. Yajima. Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, 5(2):189–211, 2012.
- [4] A. Hector, T. Bell, Y. Hautier, F. Isbell, M. Kéry, P. B. Reich, J. van Ruijven, and B. Schmid. BUGS in the analysis of biodiversity experiments: Species richness and composition are of similar importance for grassland productivity. *PLoS ONE*, 6(3):e17434, 2011.
- [5] S. S. Qian and Z. Shen. Ecological applications of multilevel analysis of variance. *Ecology*, 88(10):2489– 2495, 2007.
- [6] A. Gelman and E. Loken. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Department of Statistics, Columbia University, 2013.

Bayesian Multi-level Regression

Bayesian, multilevel models for grouped data

- group level intercepts

- group level intercepts with group level covariate

- group level slopes and intercepts

The simple, Bayesian set-up

Deterministic model:

$$g(\boldsymbol{\theta}, x_i)$$

Stochastic model:

$$\underbrace{[\boldsymbol{\theta}, \sigma^2 | y_i]}_{\text{posterior}} \propto \overbrace{[y_i | g(\boldsymbol{\theta}, x_i), \sigma^2]}^{\text{joint}} \underbrace{[\boldsymbol{\theta}]}_{\text{priors}}$$

likelihood

Draw the DAG.

Recall that

$$\underbrace{[\boldsymbol{\theta}, \sigma^2 | y_i]}_{\text{posterior}} \propto \underbrace{[y_i, \boldsymbol{\theta}, \sigma^2]}_{\text{joint}}$$

Hierarchical models: “modeling parameters”

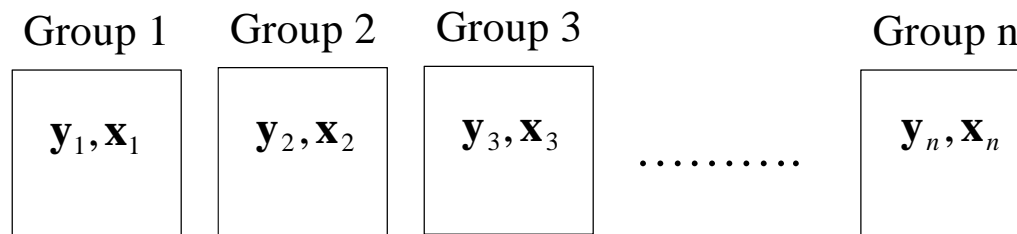
$$\begin{aligned} [\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\alpha}, y_i, \sigma^2] &\propto [y_{ij} | g(\theta_1, \theta_{2,j}, x_{ij}), \sigma_1^2] \\ &\times [\theta_{2,j} | h(\alpha_1, \alpha_2, u_j), \sigma_2^2] \\ &\times [\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\alpha}, \boldsymbol{\sigma}^2] \end{aligned}$$

Draw the DAG.

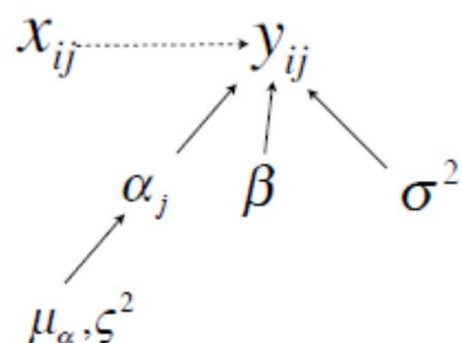
Steps in Bayesian analysis

1. Compose joint distribution of observed and unobserved quantities.
2. Factor joint distribution into sensible parts.
3. Use factored joint distribution to write:
 1. JAGS code *or*
 2. Own MCMC sampler
 1. Write full-conditional distributions
 2. Choose sampling method for each full-conditional
4. Check model
5. Make inference

The problem

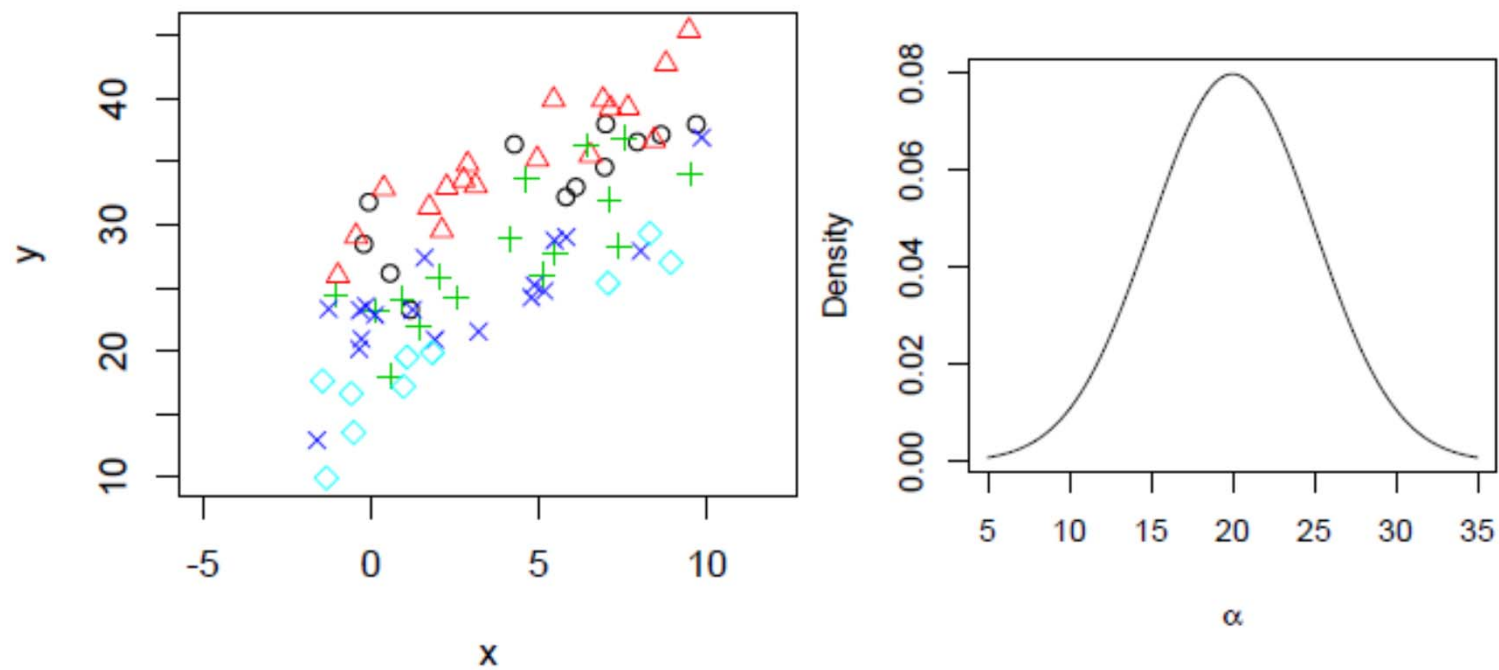


We can model the intercept:



$$\begin{aligned} [\beta, \boldsymbol{\alpha}, \sigma^2, \mu_{\alpha}, \zeta^2, |y] &\propto \prod_{i=1}^{n_j} \prod_{j=1}^J \text{normal}(y_{ij} | \alpha_j + \beta x_{ij}, \sigma^2) \\ &\times \text{normal}(\alpha_j | \mu_{\alpha}, \zeta^2) \\ &\times \text{normal}(\beta | 0, 10000) \text{normal}(\mu_{\alpha} | 0, 1000) \\ &\times \text{inverse gamma}(\sigma^2 | .001, .001) \text{inverse gamma}(\zeta^2 | .001, .001) \end{aligned}$$

We seek to understand the distribution of intercepts.



Some notation

$$\mu_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_j$$

$$y_{ij} \sim \text{normal}(\mu_{ij}, \sigma^2)$$

$$\alpha_j \sim \text{normal}(0, \varsigma^2)$$

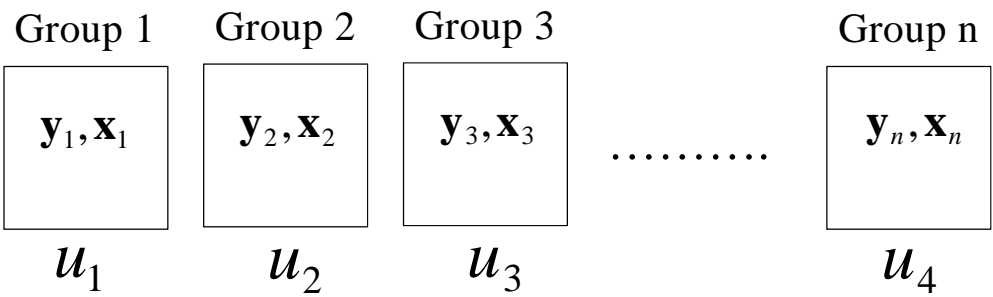
is identical to:

$$\mu_{ij} = \alpha_j + \beta_1 x_{ij}$$

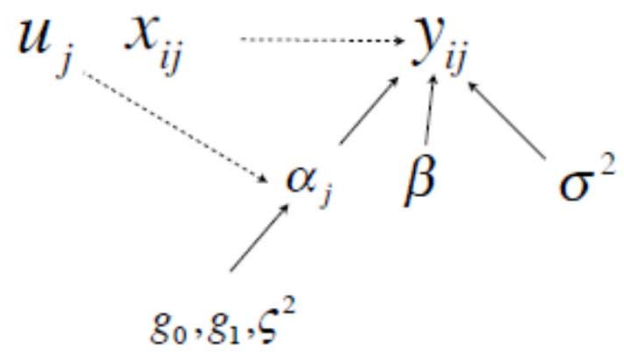
$$y_{ij} \sim \text{normal}(\mu_{ij}, \sigma^2)$$

$$\alpha_j \sim (\mu_\alpha, \varsigma^2)$$

Include data on groups.



We can model the intercept as a function of group level data:

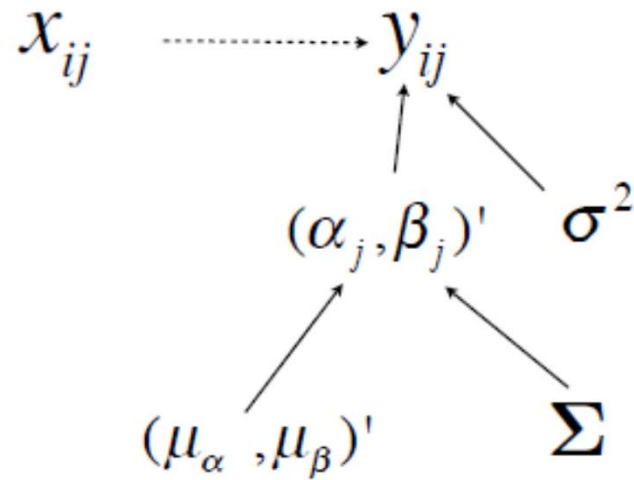


$$\begin{aligned}
 [\boldsymbol{\alpha}, \beta, \sigma^2, \mathbf{g}, \zeta^2, \mathbf{y}] &\propto \prod_{i=1}^{n_j} \prod_{j=1}^J \text{normal}(y_{ij} | \alpha_j + \beta x_{ij}, \sigma^2) \\
 &\times \text{normal}(\alpha_j | g_0 + g_1 u_j, \zeta^2) \\
 &\times \text{normal}(\beta | 0, .001) \text{normal}(g_0 | 0, 1000) \text{normal}(g_1 | 0, 1000) \\
 &\times \text{inverse gamma}(\sigma^2 | .001, .001) \text{inverse gamma}(\zeta^2 | .001, .001)
 \end{aligned}$$

Indexing groups

```
model{
  beta ~ dnorm(0,.0001)
  sigma ~ dunif(0,50)
  tau.p <- 1/sigma^2
  g0 ~ dnorm(0,.0001)
  g1 ~ dnorm(0,.0001)
  varsigma ~ dunif(0,50)
  tau.g <- 1/varsigma^2
  for (i in 1:length(y)){
    mu[i] <- alpha[group[i]]+ beta*x[i]
    y[i] ~ dnorm(mu[i],tau.p)
  }
  for(j in 1:n.group){
    mu.g[j] <- g0 + g1*u[j]
    alpha[j]~dnorm(mu.g[j],tau.g)
  }
}
```

Modeling intercepts *and* slopes



$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim \text{multivariate normal} \left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \Sigma \right)$$

$$\Sigma = \begin{pmatrix} \zeta_\alpha^2 & \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \zeta_\beta^2 \end{pmatrix}, \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \rho \zeta_\alpha \zeta_\beta$$

Modeling intercepts *and* slopes

$$\begin{aligned} [\boldsymbol{\alpha}, \boldsymbol{\beta}, \mu_{\alpha}, \mu_{\beta}, \sigma^2, \varsigma_{\alpha}, \varsigma_{\beta}, \rho | \mathbf{y}] &\propto \prod_{j=1}^J \prod_{i=1}^{n_j} \text{normal}(y_{ij} | \alpha_j + \beta_j x_{ij}, \sigma^2) \\ &\times \text{MVN} \left(\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \middle| \begin{pmatrix} \mu_{\alpha} \\ \mu_{\beta} \end{pmatrix}, \boldsymbol{\Sigma} \right) \\ &\times \text{priors on } \mu_{\alpha}, \mu_{\beta}, \sigma^2, \varsigma_{\alpha}, \varsigma_{\beta}, \rho \end{aligned}$$