

# Bayesian Analysis of Designed Experiments

ESS 575

N. T. Hobbs and C. Che-Castaldo

April 16, 2019



# Housekeeping

- No lecture on next two Thursdays (4/18, 4/26)
- Lab as usual until end of semester
- Remember class party 4/27
- Poll food habits

# Nitrous oxide lab

## Debrief

# Individual projects

How to “figure it out.”

# Analyzing experimental data: Why Bayes?

KEY TO STATISTICAL METHODS

	Design or Purpose	Measurement Variables	Ranked Variables	Attributes
1 variable 1 sample	Examination of a single sample	Procedure for grouping a frequency distribution, Box 2.1; stem-and-leaf display, Section 2.5; testing for outliers, Section 13.4 Computing median of frequency distribution, Box 4.1 Computing arithmetic mean: unordered sample, Box 4.2; frequency distribution, Box 4.3 Computing standard deviation: unordered sample, Box 4.2; frequency distribution, Box 4.3 Setting confidence limits: mean, Box 7.2; variance, Box 7.3 Computing $g_1$ and $g_2$ , Box 6.2		Confidence limits for a percentage, Section 17.1. Runs test for randomness in dichotomized data, Box 18.3
	Comparison of a single sample with an expected frequency distribution	Normal expected frequencies, Box 6.1 Goodness of fit tests: parameters from an extrinsic hypothesis, Box 17.1; from an intrinsic hypothesis, Box 17.2 Kolmogorov-Smirnov test of goodness of fit, Box 17.3 Graphic "tests" for normality: large sample sizes, Box 6.3; small sample sizes (rankit test), Box 6.4 Test of sample statistic against expected value, Box 7.4		Binomial expected frequencies, Box 5.1 Poisson expected frequencies, Box 5.2 Goodness of fit tests: parameters from an extrinsic hypothesis, Box 17.1; from an intrinsic hypothesis, Box 17.2
1 variable $\geq 2$ samples	Single classification	Single classification anova: unequal sample sizes, Box 9.1; equal sample sizes, Box 9.4 Planned comparison of means in anova, Box 9.8, single degree of freedom comparisons of means, Box 14.10 Unplanned comparison of means: T-method, equal sample sizes, Box 9.9; T <sub>1</sub> , G <sub>12</sub> , and Tukey-Kramer, unequal sample sizes, Box 9.10; Welch step-up, Box 9.11; STP test, Section 9.7; contrasts using Scheffé, T, and GT2, Box 9.12; multiple confidence limits, Section 14.10 Estimate variance components: unequal sample sizes, Box 9.2; equal sample sizes, Box 9.3 Setting confidence limits to a variance component, Box 9.3 Tests of homogeneity of variances, Box 13.1 Tests of equality of means when variances are heterogeneous, Box 13.2	Kruskal-Wallis test, Box 13.5 Unplanned comparison of means by a nonparametric STP, Box 17.5	G-test for homogeneity of percentages, Boxes 17.5 and 17.8 Comparison of several samples with an expected frequency distribution, Box 17.4; unplanned analysis of replicated tests of goodness of fit, Box 17.5
	Nested classification	Two level nested anova: equal sample sizes, Box 10.1; unequal sample sizes, Box 10.4 Three-level nested anova: equal sample sizes, Box 10.3; unequal sample sizes, Box 10.5		
	Two-way or multi-way classification	Two-way anova: with replication, Box 11.1; without replication, Box 11.2; unequal but proportional subclass sizes, Box 11.4; with a single missing observation, Box 11.5 Three-way anova, Box 12.1 More-than-three way classification, Section 12.3 and Box 12.2 Test for nonadditivity in a two-way anova, Box 13.4	Friedman's method for randomized blocks, Box 13.9	Three-way log-linear model, Box 17.9 Randomized blocks for frequency data (repeated testing of the same individuals), Box 17.11

# Analyzing experimental data: Why Bayes?



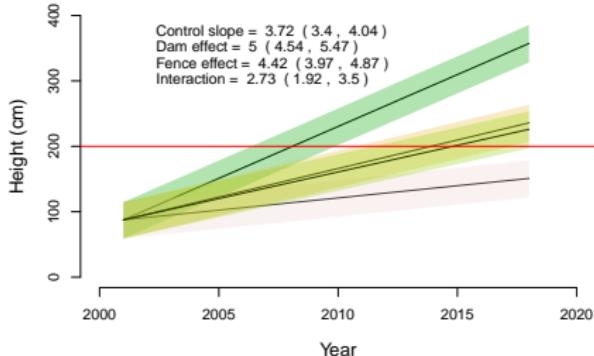
- Probabilistic interpretation of effects
- Contrasts easy to construct
- No “error” terms
- Can accommodate errors in responses (and predictors for ANCOVA)
- Can accommodate missing data
- Multiple comparisons of means handled sensibly
- Derived quantities handled easily

Photo c/o of the Minnesota Agricultural Experiment Station at <HTTP://WWW.meas.mun.ed>.

# Yellowstone experiment



# Analysis of the joint distribution of data and parameters



$$\mu_{ijt} = \beta_{0j} + (\beta_1 + \beta_2 x_{1,ij} + \beta_3 x_{2,ij} + \quad (1)$$

$$\beta_4 x_{1,ij} x_{2,ij}) t \quad (2)$$

$$y_{ijt} \sim \text{lognormal}(\log(\mu_{ijt}), \sigma_j^2)$$

$$\beta_{0j} \sim \text{normal}(\mu_{\beta_0}, \sigma_{\beta_0}^2)$$

$$\mu_{\beta_0} \sim \text{normal}(0, 10000)$$

$$\sigma_{\beta_0}^2 \sim \text{uniform}(0, 5) \quad (3)$$

$$\beta_{i \in 1, \dots, 3} \sim \text{normal}(0, 10000) \quad (4)$$

## Learning objectives

- Understand design matrix notation for models with nominal (qualitative) variables.
- Be able to compose Bayesian models for simple experimental designs.
- Build a foundation of knowledge needed for developing models appropriate for your specific research.
- Learn a spectacular trick for composing design matrices.

# Topics

- Review of matrix algebra
- Design matrices
- Specifying models including design
- A general approach to model building
  - ▶ Means models vs effects models
  - ▶ Constraints on parameters
- Inference
  - ▶ Effect sizes
  - ▶ Contrasts
  - ▶ Multiple comparisons

## Matrix notation for linear models

Remember matrix multiplication?

Example of matrix multiplication for  $n$  observations using 2 predictor variables  $x_{i,1}$  and  $x_{i,2}$  and an intercept.

$$\begin{pmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ 1 & x_{3,1} & x_{3,2} \\ 1 & \cdot & \cdot \\ 1 & \cdot & \cdot \\ 1 & \cdot & \cdot \\ 1 & x_{n,1} & x_{n,2} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} \\ \beta_0 + \beta_1 x_{2,1} + \beta_2 x_{2,2} \\ \beta_0 + \beta_1 x_{3,1} + \beta_2 x_{3,2} \\ \cdot \\ \cdot \\ \cdot \\ \beta_0 + \beta_1 x_{n,1} + \beta_2 x_{n,2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \cdot \\ \cdot \\ \cdot \\ \mu_n \end{pmatrix}$$

Define inner product (aka dot product.)

## Matrix notation for linear models

You will often see models written using something like

$$y_i \sim \text{normal}(\mathbf{x}'_i \boldsymbol{\beta}, \sigma^2)$$

or

$$y_i \sim \text{normal}(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$$

or (incorrectly, in my view)

$$y_i \sim \text{normal}(\mathbf{X}_i \boldsymbol{\beta}, \sigma^2)$$

or

$$\mathbf{y} \sim \text{multivariate normal}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I)$$

Note that  $\mathbf{X}$  is a matrix with ones in column 1 and values of covariates in other columns. Thus,  $\mathbf{X}\boldsymbol{\beta}$  returns a vector.

## Matrix notation for linear models

You also see models written using something like

$$y_i \sim \text{normal}(\beta_0 + \mathbf{x}'_i \boldsymbol{\beta}, \sigma^2)$$

or

$$y_i \sim \text{normal}(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$$

Note that in this case  $\mathbf{X}$  is a matrix of values of covariates in columns. It does not have a ones in column one. I like this form because it is easy to create multi-level models by simply subscripting  $\beta_0$  to represent groups. Often you see the' or the  $T$  superscript omitted.

## Design matrix

$E[y]$  Design Matrix

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \\ \vdots \\ \hat{y}_N \end{bmatrix} = \begin{bmatrix} 1 & 1.2 \\ 1 & 3.4 \\ 1 & 1.7 \\ 1 & 7.9 \\ \vdots \\ 1 & 4.3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \alpha + \beta \times 1.2 \\ \alpha + \beta \times 3.4 \\ \alpha + \beta \times 1.7 \\ \alpha + \beta \times 7.9 \\ \vdots \\ \alpha + \beta \times 4.3 \end{bmatrix}$$

- Fine for quantitative x's, but how do we handle qualitative predictor variables, i.e., different treatments and treatment levels in an experiment, qualitative variables in descriptive (non-experimental) models?
- These mean the same thing in this context: categorical = non-metric = nominal = qualitative

## Note

The next several completely random designs but could be easily modified to incorporate restrictions on the randomization, for example blocking. The  $\beta_0$  term can be interpreted as the mean in the control or “reference” condition. If there is no control, then the  $\beta_0$  term is viewed as the “grand mean.” The slope terms represent “effects”, the changes in the control attributable to different treatments and treatment levels. These are called “effects models.” We will talk about more complex designs, “means models”, and models that lack controls subsequently. We will also show how the slopes need to be constrained to prevent problems with identifiability.

## Notation for models: design matrices

- data matrix
$$\begin{pmatrix} y_1 & 0 & 1 \\ y_2 & 1 & 0 \\ y_3 & 1 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ y_n & 0 & 1 \end{pmatrix}$$
- deterministic model
$$\mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$
- design matrix
$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & 0 & 1 \end{pmatrix}$$
- or, equivalently
$$\boldsymbol{\mu} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$
- $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$
- likelihood
- $y_i \sim [y_i | \mu_i, \sigma^2]$
- e.g.,
- $y_i \sim \text{normal}(\mu_i, \sigma^2)$

# Notation for models: design matrices with repeated measures

data matrix

$$\begin{pmatrix} y_1 & 0 & 1 & \overbrace{1}^{\text{time}} \\ y_2 & 1 & 0 & 1 \\ y_3 & 1 & 1 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ y_n & 0 & 1 & T \end{pmatrix}$$

design matrix

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 1 & \end{pmatrix}$$

deterministic model

$$\mu_i = \beta_0 + (\underbrace{\beta_1}_{\text{time slope}} + \beta_1 x_{1,i} + \beta_2 x_{2,i})t$$

or, equivalently

$$\begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \mu_{3,t} \\ \cdot \\ \cdot \\ \mu_{n,t} \end{pmatrix} = \beta_0 + \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_2 \end{pmatrix} \times t$$

$$\boldsymbol{\mu} = \beta_0 + \mathbf{X}\boldsymbol{\beta}$$

likelihood

$$y_i \sim [y_i | \mu_i, \sigma^2]$$

e.g.,

$$y_i \sim \text{normal}(\mu_i, \sigma^2)$$

## Notation for models: design matrices with interactions

data matrix

$$\begin{pmatrix} y_1 & 0 & 1 \\ y_2 & 1 & 0 \\ y_3 & 1 & 1 \\ \vdots & \vdots & \vdots \\ y_n & 0 & 1 \end{pmatrix}$$

design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 & \overbrace{0}^{\text{product}} \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & .. & \vdots \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

deterministic model

$$\mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i} x_{2,i}$$

or, equivalently

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

likelihood

$$y_i \sim [y_i | \mu_i, \sigma^2]$$

e.g.,

$$y_i \sim \text{normal}(\mu_i, \sigma^2)$$

Both approaches are easily related to rectangular structures for storing your data.

## Tricks for specifying design matrices

Look into `model.matrix` in R or, if you can specify a model for `lme` you can get the design and the random effects matrices using

```
#get design and random effects matrices for split plot design
library(MASS)
library(lme4)
fit.lme = lmer(Y ~ B + V * N + (1 | B:V), data= oats)
X = as.matrix(getME(fit.lme, "X"))
Z = as.matrix(getME(fit.lme, "Z"))
#Code for JAGS likelihood:
#beta is vector of coefficients, X is fixed effects model matrix
#gammma is vector of ranom effects, Z is random effects model matrix
for (i in 1:n) {
  y[i] ~ dnorm(mu[i],tau.res)
  mu[i] <- inprod(beta[],X[i,]) + inprod(gamma[],Z[i,])
}
```

# Notation for models: subscripting parameters

Coefficient	Interpretation	
$\beta_{1,1}$	Effect of level one of treatment $\beta_1$	
$\beta_{1,2}$	Effect of level two of treatment $\beta_1$	
$\beta_{2,1}$	Effect of level one of treatment $\beta_2$	
$\beta_{2,2}$	Effect of level two of treatment $\beta_2$	

Data matrix

Response	Index for $\beta_1$	Index for $\beta_2$
$y_1$	1	1
$y_2$	1	2
$y_3$	2	1
$y_4$	2	2
.	.	.
.	.	.
$y_n$	2	2

deterministic model

$$\mu_{jk} = \beta_0 + \beta_{1,j} + \beta_{2,k}$$

likelihood

$$y_{ijk} \sim [y_{ijk} | \mu_{jk}, \sigma^2]$$

Coding: Use the index trick.

```
mu[i] <- b0 + b1[b1.index[i]] + b2[b2.index[i]]  
y[i] ~ dnorm(mu[i], tau)
```

Note: Level one must be set to zero if there is a control. See example code. Implementing these models requires a thorough understanding of using the index trick to subscript parameters in code.

# Notation for models: subscripting parameters with interactions

Coefficient	Interpretation
$\beta_{1,1}$	Effect of level one of treatment $\beta_1$
$\beta_{1,2}$	Effect of level two of treatment $\beta_1$
$\beta_{2,1}$	Effect of level one of treatment $\beta_2$
$\beta_{2,2}$	Effect of level two of treatment $\beta_2$

Data matrix

Response	Index for $\beta_1$	Index for $\beta_2$
$y_1$	1	1
$y_2$	1	2
$y_3$	2	1
$y_4$	2	2
.	.	.
.	.	.
$y_n$	2	2

deterministic model

$$\mu_{jk} = \beta_0 + \beta_{1,j} + \beta_{2,k} + \underbrace{\beta_{3,jk}}_{\text{interaction}}$$

likelihood

$$y_{ijk} \sim [y_{ijk} | \mu_{jk}, \sigma^2]$$

Coding: Use the index trick, e.g,

```
mu[i] <- b0 + b1[b1.index[i]] + b2[b2.index[i]] +
           b3[b1.index[i], b2[b2.index[i]]]
y[i] ~ dnorm(mu[i], tau)
```

Must set interaction = 0 if either  
index = 1.

# Pseudo code for implementing subscripted models with controls and interactions

```
beta0 ~ dnorm(0,.00001) #mean in the absence of treatment
sigma ~ dunif(0,100)
tau <- 1/sigma^2
#Set first coefficient to 0 when there are controls. This is indexed as 1 in the data.
beta1[1] <- 0
beta2[1] <- 0
for(i in 2 : n.levels.beta1){
  beta1[j] ~ dnorm(0,.000001)
  beta3[i,1] <- 0
}
for(j in 1 : n.levels.beta2){
  beta2[j] ~ dnorm(0,.000001)
}
#priors for interactions
for(i in 1 : n.levels.beta1{
  for(j in 1 : n.levels.beta2){
    beta3[i,j] ~ dnorm(0,.000001)
  }
}
#likelihood
for(i in 1:length(y)){
  #x.index[i] is data vector set to zero if either coefficient
  #index = 1, so that observation is from control level of
  #either treatment
  mu[i] <- beta0 + beta1[beta1.index[i]] + beta2[beta2.index[i]] +
    beta3[beta1.index[i], beta2.index[i]] * x.index[i]
prior
  y[i] ~ dnorm(mu[i], tau)
}
```

# Pseudo code for implementing subscripted models without controls and interactions

```
beta0 ~ dnorm(0, .00001) #grand mean
sigma ~ dunif(0, 100)
tau <- 1/sigma^2
sigma.beta1 ~ dunif(0, 100)
sigma.beta2 ~ dunif(0, 100)
sigma.beta3 ~ dunif(0, 100)
#note multilevel structure for coefficients
for(j in 1:n.levels.beta1){
  beta1[j] ~ dnorm(0, sigma.beta1^-2)
}
for(j in 1:n.levels.beta2){
  beta2[j] ~ dnorm(0, sigma.beta2^-2)
}
#priors for interactions
for(i in 1:n.levels.beta1{
  for(j in 1:n.levels.beta2){
    beta3[i,j] ~ dnorm(0,sigma.beta3^-2)
  }
}
#likelihood
for(i in 1:length(y)){
  mu[i] <- beta0 + beta1[beta1.index[i]] +
    beta2[beta2.index[i]] +
    beta3[beta1.index[i], beta.2.index[i]]]
  #Order of index for beta3 must match prior
  y[i] ~ dnorm(mu[i], tau)
}
```

# Which notation to use?

- Design matrix
  - ▶ Clear interpretation analogous to regression
  - ▶ Coefficients drop out for controls because  $\beta x_i = 0$
  - ▶ Easily adapted for selecting among alternative models without modifying JAGS code using R's `model.matrix()` function
  - ▶ Particularly well suited to single levels of treatment with control
- Subscripting parameters
  - ▶ Easier to write for complex models
  - ▶ Most models in texts written this way. Never invent what you can imitate.
  - ▶ Usually easier to code, but be careful about interactions when there are controls
  - ▶ Must set one slope parameter to zero if you have a control or reference.

## Recall ignorability and research design

- Designs that are ignorable require no indices other than an index for the individual observations (i.e.  $y_i$ ) and the covariates ( $x_i$ ). Simple random sampling and completely randomized experiments have ignorable designs. In these cases  $[I | x, y, \phi] = [I]$ .
- Designs that are not completely random, for example, randomized complete block experiments, stratified random sampling, cluster, and others are not ignorable and must include information on the design in the analysis. Usually, proper indexing and information about the sample sizes specifies all of the needed information. In these cases,  $[I | x, y, \phi] = [I|x]$ .

# Widely used designs in ecology illustrated with Yellowstone treatments

DF: dam and fence

F: fence, no dam

D: dam, no fence

C: no dam, no fence

0: no dam or no fence

Completely Random

DF	D	C
F		D
	C	D
DF	F	C
C		C
D	F	D
		DF

Randomized Complete Block

DF	C
F	D
DF	C
F	D
DF	

Split plot

D	0
F	0
0	F
F	0
0	F
D	0
0	0
F	F
0	0
F	F

Replications within each cell

### Completely Random

$$\mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i} x_{2,1}$$

$$y_{ikm} \sim \text{lognormal}(y_{ikm} | \log(\mu_{km}), \sigma^2)$$

$$\beta_0 \sim \text{normal}(0, 10000)$$

### Randomized Complete Block

$$\mu_i = \beta_{0,j} + \beta_1 x_{1,ij} + \beta_2 x_{2,ij} + \beta_3 x_{1,ij} x_{2,1}$$

$$\beta_{0,j} \sim \text{normal}(\mu_{\beta_0}, \sigma_{\beta_0}^2)$$

$$y_{ij} \sim \text{lognormal}(y_{ij} | \log(\mu_{jkm}), \sigma^2)$$

### Split Plot

$$\mu_{ijk} = \beta_{0,j} + \beta_1 x_{1,ijk} + \beta_2 x_{2,ijk} + \beta_3 x_{1,ijk} x_{2,ijk} + \varepsilon_{jk}$$

$$\beta_{0,j} \sim \text{normal}(\mu_{\beta_0}, \sigma_{\beta_0}^2)$$

$$\varepsilon_{jk} \sim \text{normal}(0, \sigma_{jk}^2) \quad (5)$$

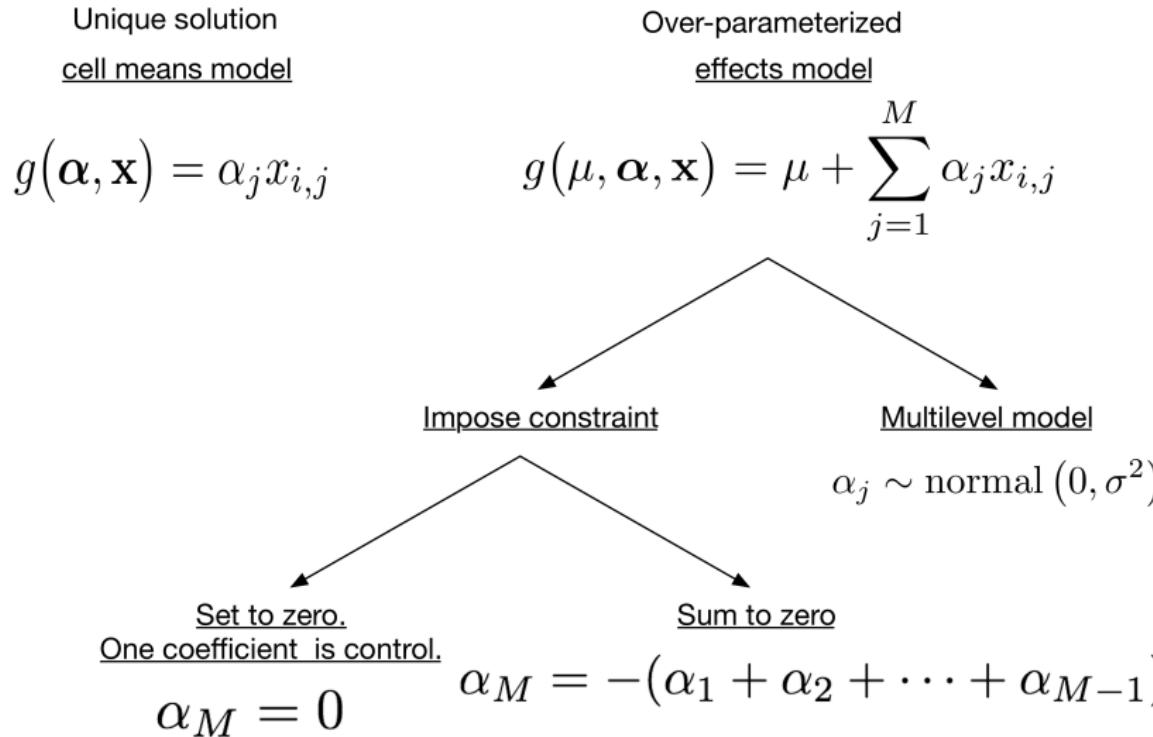
$$y_{ijk} \sim \text{lognormal}(y_{ijk} | \log(\mu_{jk}), \sigma^2)$$

# Topics

- A general approach to model building
  - ▶ Means models vs effects models
  - ▶ Constraints on parameters
- Inference
  - ▶ Effect sizes
  - ▶ Contrasts
  - ▶ Multiple comparisons

A general approach to building models with categorical predictors

# Choices in specifying models



## Cell Means Model: Joint and DAG

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

```
graph TD; alpha_j[alpha_j] --> y_ij[y_{i,j}]; sigma[sigma] --> y_ij;
```

- Interest in group means and not effects
- Have prior information for group means
- Lack prior information for group means - use vague priors
- Number of parameters = number of unknowns
- Recover effects or grand mean as derived quantities

## Cell Means Model: Design Matrix

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

# 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

mu = X %*% alpha #X is 50 x 5 design matrix
for (i in 1:50) {
  y[i] ~ dnorm(mu[i], tau)
}
# or, equivalently the likelihood could be
# for (i in 1:50) {
#   y[i] ~ dnorm(alpha[x[i]], tau) #x[i] is index, 1 - 5
# }

# Derived quantities
diff.2.1 <- alpha[2] - alpha[1]
diff.3.1 <- alpha[3] - alpha[1]
grandMean <- mean(alpha)
effects = alpha - grandMean
```

- Use the index trick implements subscripted parameter model.
- Matrix multiplication implements design matrix model.
- Compute effects and contrast as derived quantities

## Effects Models - Set to Zero: Joint and DAG

$$[\mu, \boldsymbol{\alpha}, \sigma | \mathbf{y}] \propto \prod_{i=1}^N \prod_{j=2}^M [y_{i,j} | g(\mu, \boldsymbol{\alpha}, \mathbf{x}), \sigma^2] [\mu] [\boldsymbol{\alpha}] [\sigma]$$
$$g(\mu, \boldsymbol{\alpha}, \mathbf{x}) = \mu + \sum_{j=2}^M \alpha_j x_{i,j}$$

A directed acyclic graph (DAG) illustrating the relationships between variables. The nodes are labeled  $\mu$ ,  $\alpha_{j=2\dots M}$ , and  $\sigma$  at the bottom, and  $y_{i,j}$  and  $x_{i,j}$  at the top. Arrows point from  $\mu$  and  $\alpha_{j=2\dots M}$  to  $y_{i,j}$ . Arrows point from  $y_{i,j}$  and  $\sigma$  to  $x_{i,j}$ . A dashed arrow points from  $x_{i,j}$  back to  $y_{i,j}$ .

- Interest in effects and not means
- Have prior information for effect sizes
- Lack prior information for effect sizes - can estimate conservatively
- Number of parameters > number of unknowns requires constraint
- Recover group means as derived quantities

## Effects Model- Set to Zero: Design Matrix

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

- CRD with 1 factor and 5 levels
- Remove parameter by setting  $\alpha_5 = 0$
- Group 1 is now represented by intercept  $\mu$
- $\alpha_5$  represent deviations from this baseline/control group

# Effects Models - Set to Zero: JAGS

```
# Priors
for (i in 1:5){
  alpha[i] ~ dnorm(0, 0.001)
}

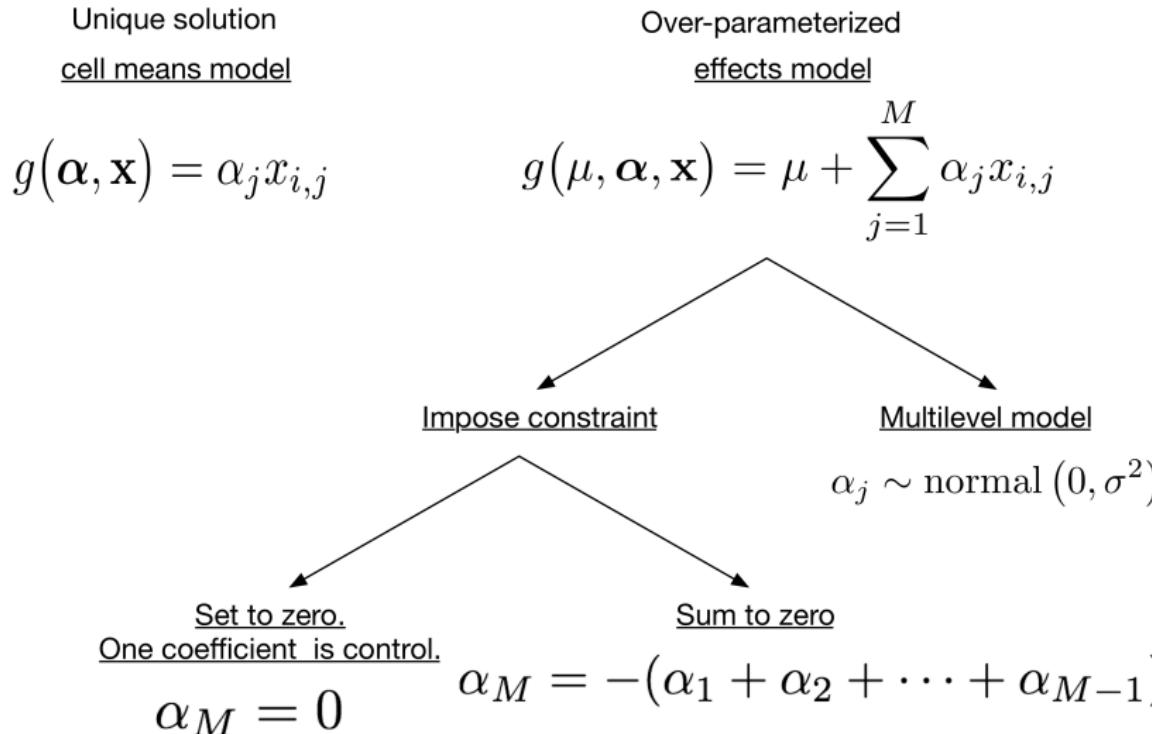
sigma ~ dunif(0, 100)
tau <- 1 / ( sigma * sigma)

# Likelihood, design matrix model
mu = X %*% alpha #X is design 50 x 5 design matrix with 1's in column one
for (i in 1:50) {
  y[i] ~ dnorm(mu[i] , tau)
}

# Derived quantities
cell[1] <- alpha[1]
for (i in 2:4){
  cell[i] <- alpha[1] + alpha[i]
}
grandMean <- mean(cell[])
```

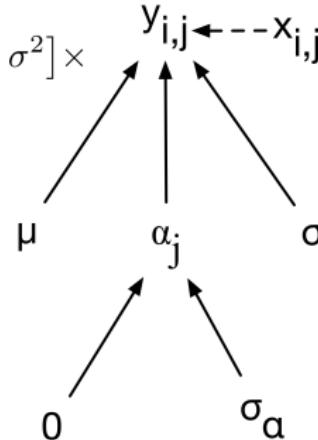
- Compute cell and grand means as derived quantities

# Parameterize a model with categorical predictors



## Effects Model - Multi-level: Joint and DAG

$$[\mu, \boldsymbol{\alpha}, \sigma, \sigma_\alpha | \mathbf{y}] \propto \prod_{i=1}^N \prod_{j=1}^M [y_{i,j} | g(\mu, \boldsymbol{\alpha}, \mathbf{x}), \sigma^2] \times [\alpha_j | 0, \sigma_\alpha] [\mu] [\sigma] [\sigma_\alpha]$$
$$g(\mu, \boldsymbol{\alpha}, \mathbf{x}) = \mu + \sum_{j=1}^M \alpha_j x_{i,j}$$



- Interest in effects and not means
- Have prior information for effect sizes
- Lack prior information for effect sizes - can estimate conservatively
- Number of parameters > number of unknowns is ok! Why?
- Recover group means as derived quantities

## Effects Model - Multi-Level: Design Matrix

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

- CRD with 1 factor and 5 levels
- Intercept,  $\mu$ , is the grand mean
- $\alpha_j$  represent deviations from the grand mean
- $\alpha_j$  are partially pooled allowing us to estimate all of them directly

# Effects Models - Multi-level: JAGS

```
# Priors
mu0 ~ 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 1:6){
  alpha[i] ~ dnorm (0, tau[2])
}
mu ~ dnorm(0,10000)

#Design matrix approach
mu[i] = X %*% alpha
for (i in 1:50) {
  y[i] ~ dnorm(mu[i], tau[1])
}
```

## Not covered: Sum to zero constraint

See

- Kruschke, J. K. 2015. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. Academic Press, Inc. Pages 583- 590
- Ntzoufras, I. 2009. Bayesian Modeling Using WinBUGS John Wiley & Sons, Hoboken, NJ U.S.A. Pages: 169 - 172

# Clarity of interpretation

- We make inference using marginal posteriors of parameters and derived quantities.
- "The probability that the effect of treatment exceed 0 was .95"
- "We can be 90% certain that the dam treatment doubled willow height by year 17."
- $\Pr(\text{Browsed} > \text{Unbrowsed}) = .83$
- CI95: effect of browse = -4.0

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

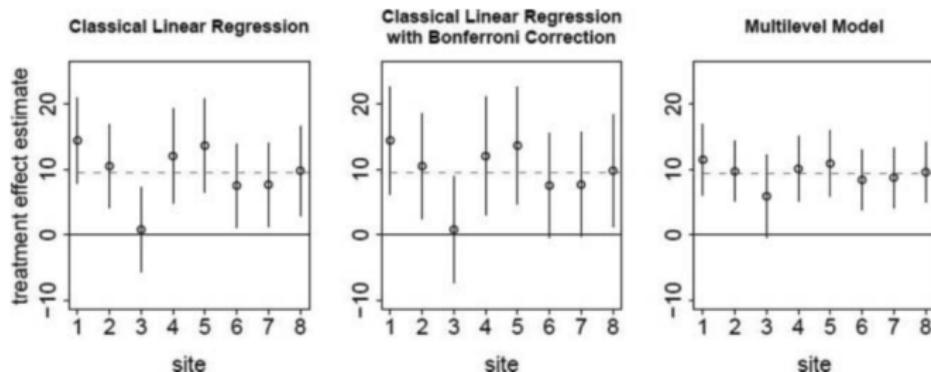
xkcd.com

## Multiple comparisons of cell means

Inference on differences between cell means are calculated directly in a means model or indirectly in an effects model. They are made as the difference posterior distribution of the difference between cell means. These are analogous to single degree of freedom contrasts or Tukeys or the like, but a lot less trouble. But what about the problem of multiple comparisons?

# Mutiple Comparisons of means

- 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

Note that it is more “difficult” to find differences among means in the multi-level case because the are pulled together by partial pooling. Gelman

## Multiple comparisons of cell means

Multiple comparisons are reliable if the model is hierarchical such that means or effects are drawn from a distribution. Illustrating:

$$[\mu, \sigma^2, \alpha, \zeta_\mu^2 | \mathbf{y}] \propto \prod_{i=1}^{n_j} \prod_{j=1}^J [y_{ij} | \mu_j, \sigma^2] [\mu_j | \alpha, \zeta_\mu^2] [\sigma^2] [\zeta^2] [\alpha]$$

Subtract one cell mean from another to get posterior distribution of difference of means. Shrinkage of the distribution of means as the number of means increases assures that it becomes more difficult for the posterior distribution of a difference between to exclude 0. Neat and tidy.

This also holds for effects models where cell means are calculated from effects and the control or grand mean.

## Future study

- ① Hobbs and Hooten, chapters 6.2.3 and 10.2.
- ② A. Gelman, J. B. Carlin, H. S. Stern, D. Dunson, A. Vehtari, and D. B. Rubin. Bayesian data analysis. 2013 Chapman and Hall / CRC, London, UK.
- ③ A. Gelman, and J. Hill. 2009. Data analysis using regression and multilevel / hierarchical models. Cambridge University Press, Cambridge, UK.
- ④ McCarthy, M. A. 2007. Bayesian Methods for Ecology. Cambridge University Press, Cambridge, U. K.

## Take home from this exhausting lecture

- Analysis of designed experiments closely resembles other types of Bayesian modeling, providing enormous flexibility to the experimentalist.
- We can use moment matching and all of the hierarchical tricks we have learned to flexibly create models for analysis of designed experiments:
  - Responses and latent quantities with support 0 or 1, 0 to 1, counts, successes on trials, counts in multiple categories, real strictly non-negative, all real numbers.
  - Errors in quantitative covariates (if we have them).
  - Errors in responses
  - Group level effects (aka random effects) in space and time.
- Results are easy to interpret and communicate.

## Future study

- ① Hobbs and Hooten, chapters 6.2.3 and 10.2.
- ② A. Gelman, J. B. Carlin, H. S. Stern, D. Dunson, A. Vehtari, and D. B. Rubin. Bayesian data analysis. 2013 Chapman and Hall / CRC, London, UK.
- ③ A. Gelman, and J. Hill. 2009. Data analysis using regression and multilevel / hierarchical models. Cambridge University Press, Cambridge, UK.
- ④ McCarthy, M. A. 2007. Bayesian Methods for Ecology. Cambridge University Press, Cambridge, U. K.

## 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.