Even more hierarchical models

FW 891

Click here to view presentation online

Christopher Cahill

25 October 2023

Quantitative Fisheries Center
MICHIGAN STATE UNIVERSITY

Purpose

- Today we introduce a powerful extension of mixed effects models
  - Random slopes (aka varying effects models)
  - Repent for our earlier sins
- Adventures in covariance a la McElreath (2023)
  - Cover some math necessary for working with covariance matrices
- Simulate a varying effects problem
- Develop both centered and noncentered varying effects models in Stan

Some references

- This lecture is drawing heavily on McElreath (2023), and much of the code and analyses are adapted from information in that text

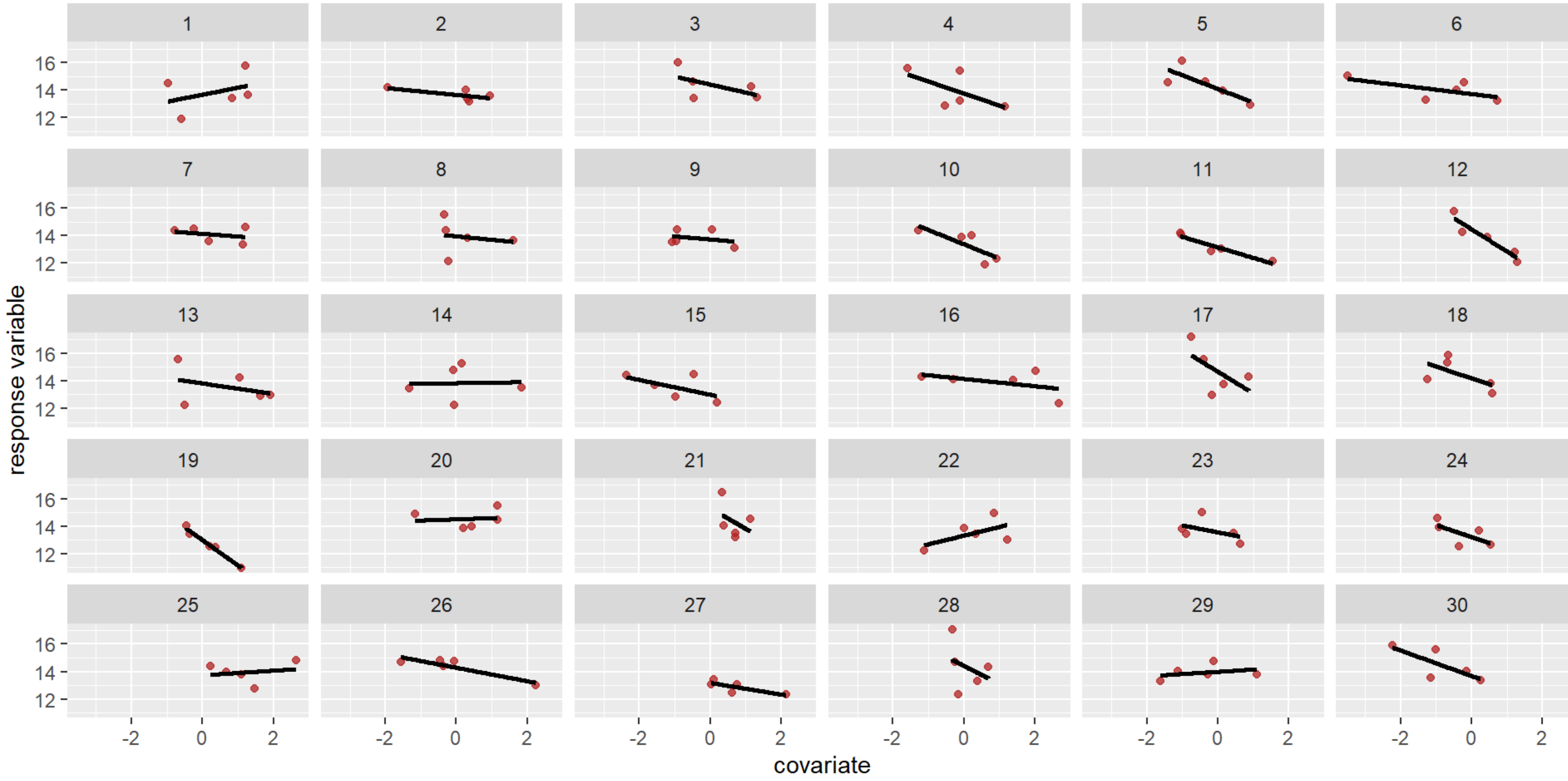
Some references

- This lecture is drawing heavily on McElreath (2023), and much of the code and analyses are adapted from information in that text
  - If you want more information, check “Adventures in Covariance” chapter of this book

Some references

- This lecture is drawing heavily on McElreath (2023), and much of the code and analyses are adapted from information in that text
  - If you want more information, check “Adventures in Covariance” chapter of this book
- See also Gelman and Hill (2007)
  - Specifically chapter 13

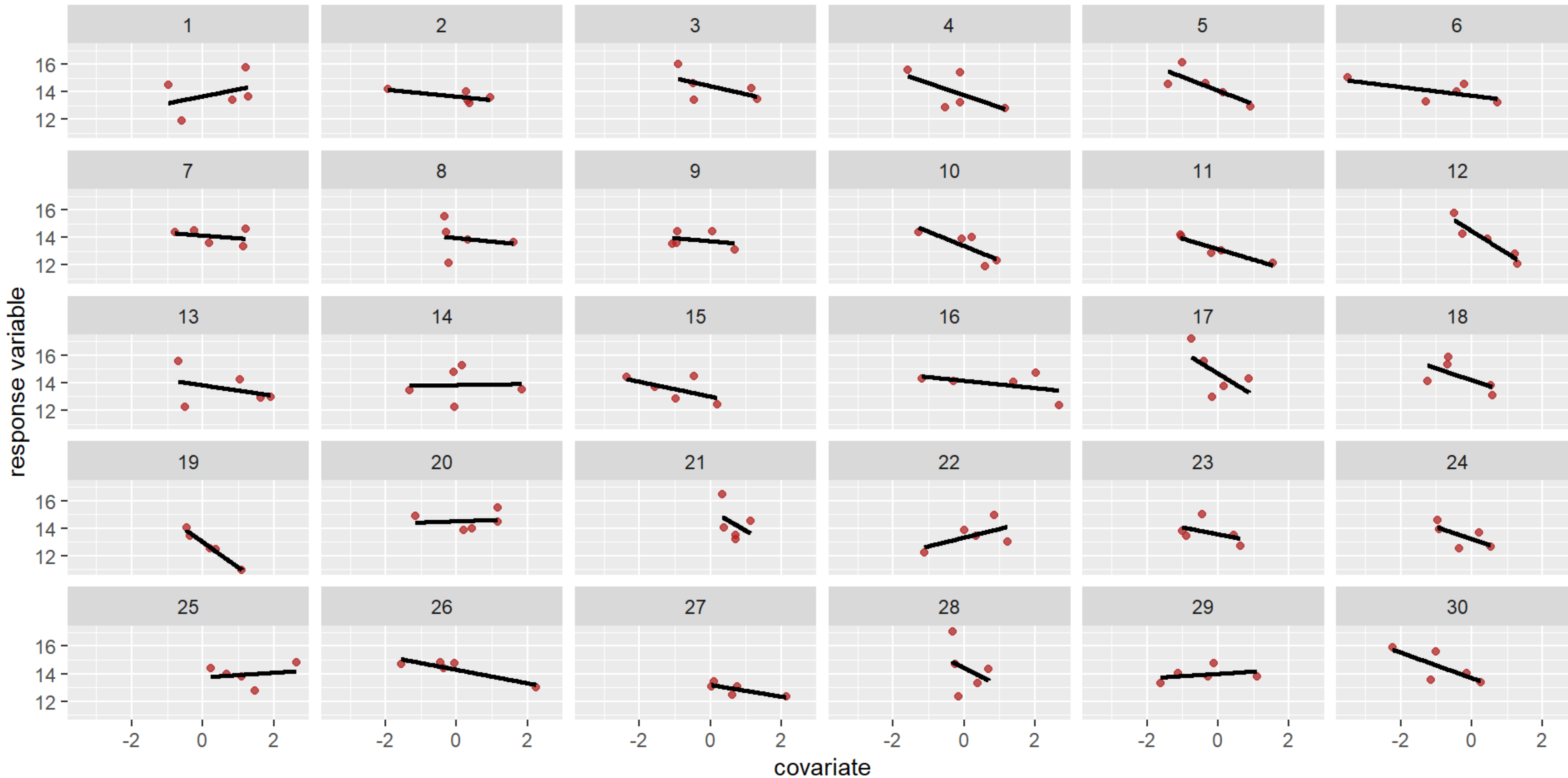
Thinking about variability in ecological systems



Three key takeaways

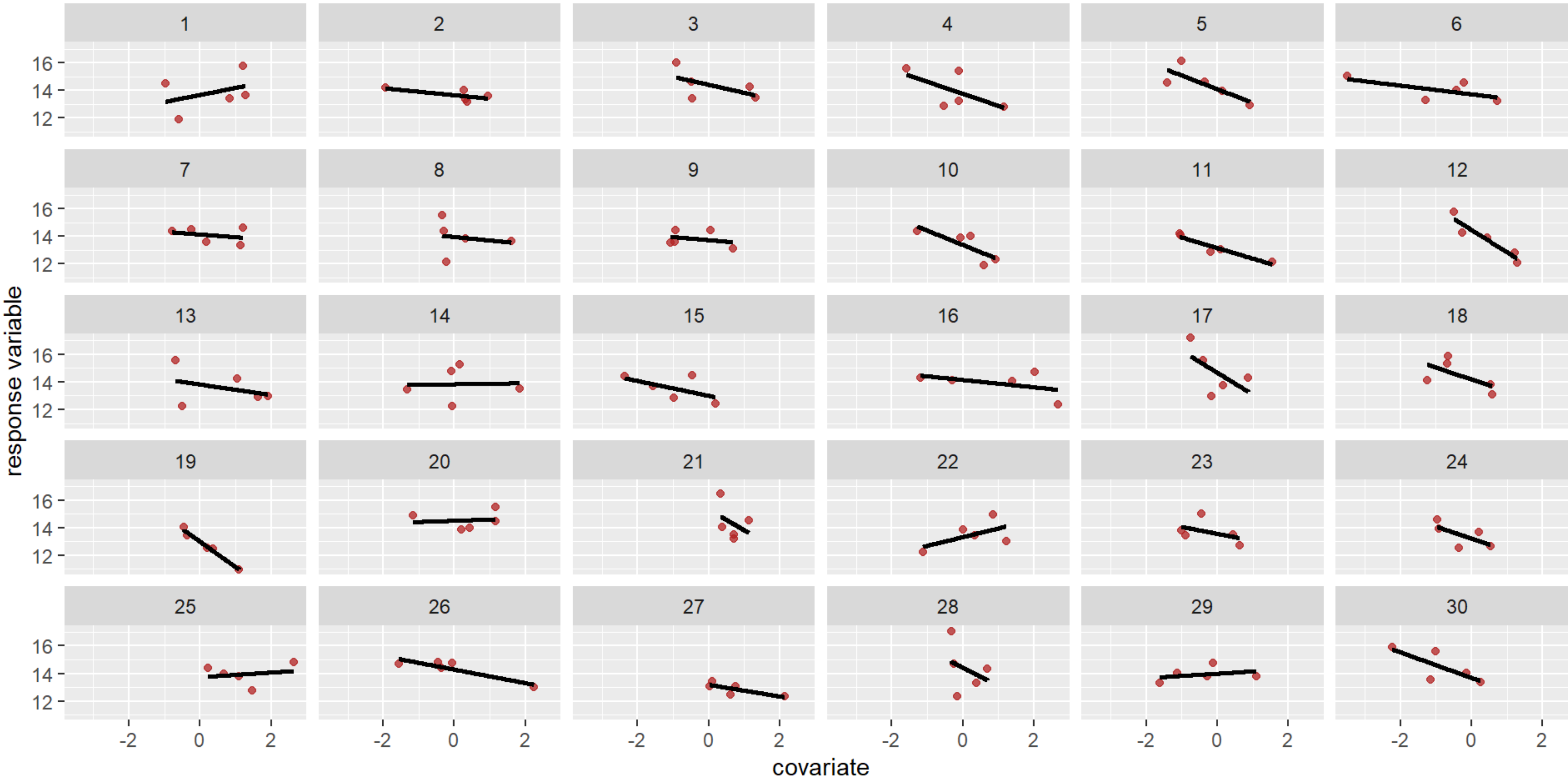
Point #1

- variability in both intercepts *and* slopes among replicates



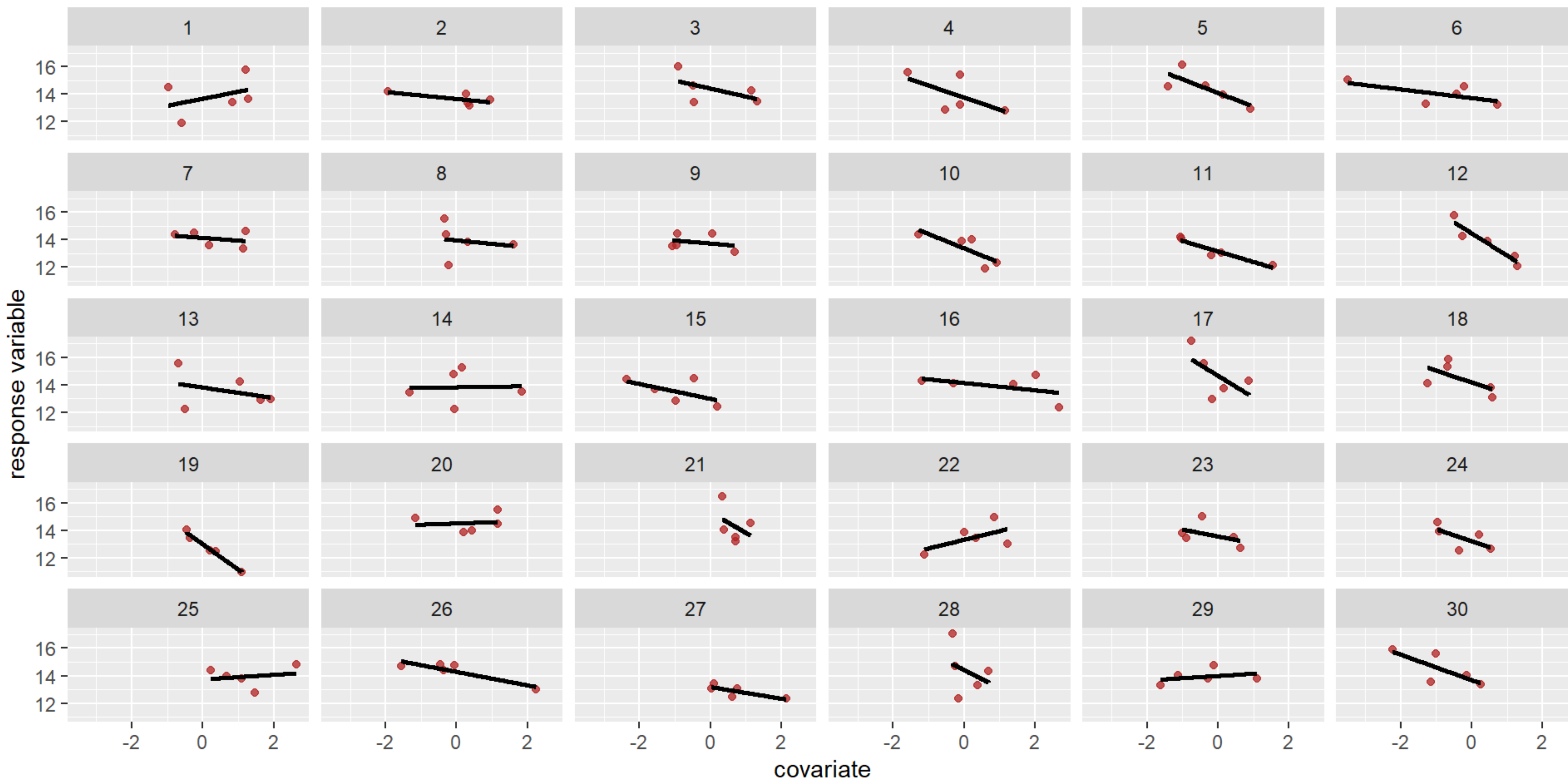
Point #2

- slopes get steeper as intercepts get bigger



Point #3

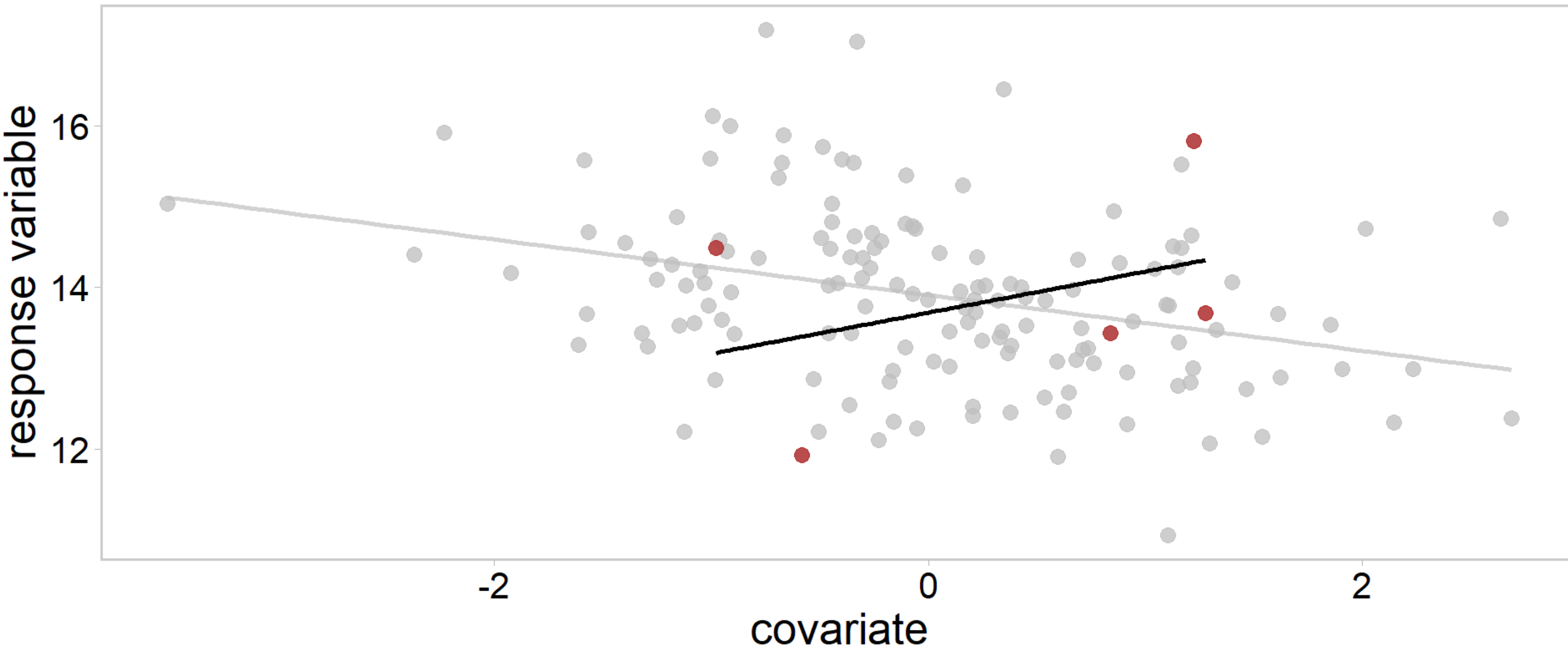
- some groups display Simpson's paradox



Simpson's paradox

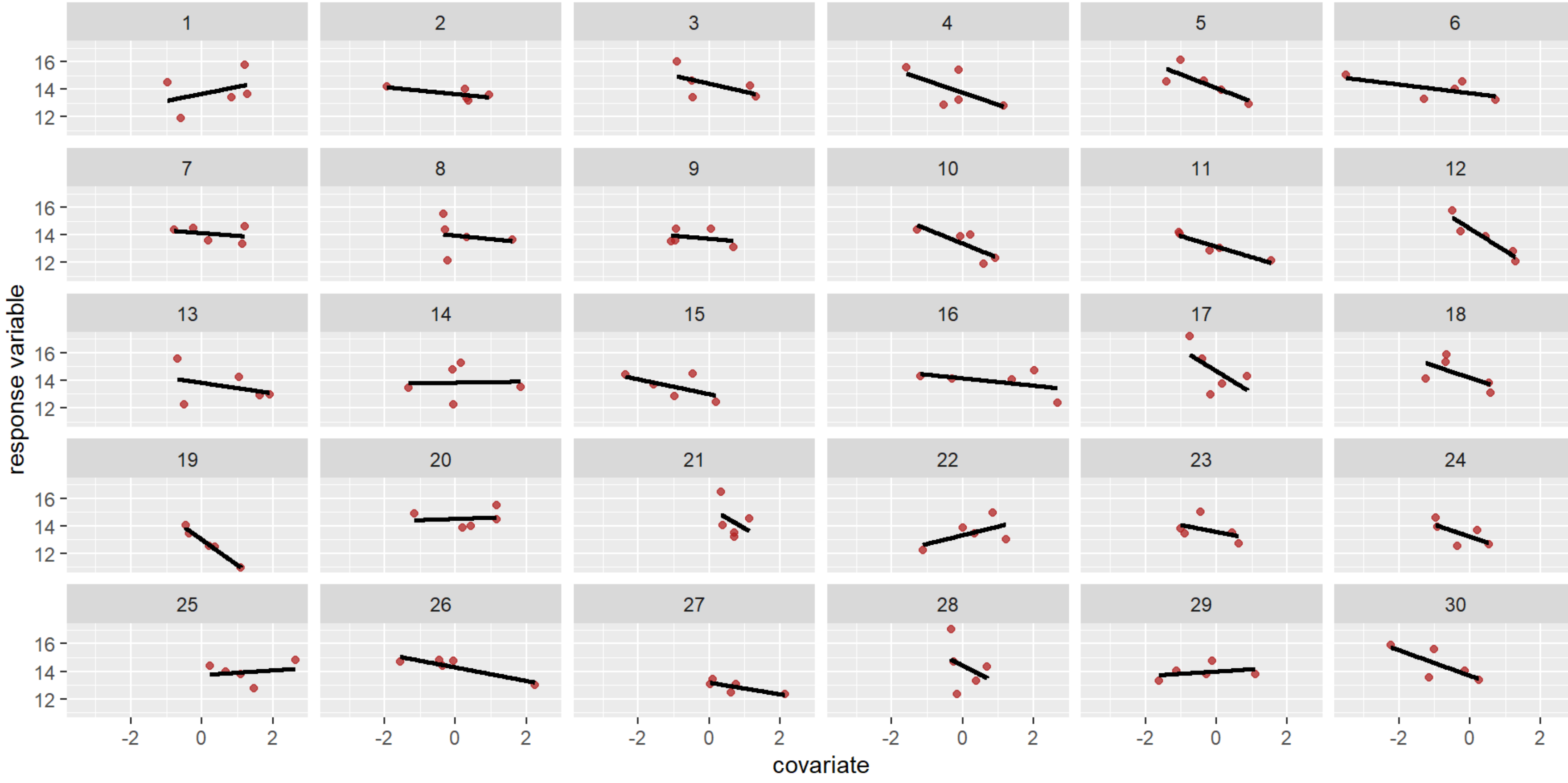
Colored points are data from replicate 1

Grey points are data from all other replicates



Group question

- what is a question in your field of study that might show a similar pattern?



Varying effects

- Generalization of standard multilevel regression
  - Specifically, models that allow slopes and intercepts to vary by group
- Several ways to write, here's one for the model we just visualized

Varying effects maths

$$y_i \sim \mathrm{N}(\mu_i, \sigma) \qquad \text{[likelihood]}$$

$$\mu_i = \beta_{0[group]} + \beta_{1[group]} x_{1[i]} \qquad \text{[linear model]}$$

adapted from McElreath 2023

Varying effects maths

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\mu_i, \sigma) && \text{[likelihood]} \\
\mu_i &= \beta_{0[group]} + \beta_{1[group]} x_{1[i]} && \text{[linear model]}
\end{aligned}
$$

Then comes the matrix of varying intercepts and slopes, with it's covariance matrix $\mathbf{\Sigma}$:

adapted from McElreath 2023

Varying effects maths

$$y_i \sim \mathrm{N}(\mu_i, \sigma) \qquad \text{[likelihood]}$$

$$\mu_i = \beta_{0[group]} + \beta_{1[group]} x_{1[i]} \qquad \text{[linear model]}$$

Then comes the matrix of varying intercepts and slopes, with it's covariance matrix $\mathbf{\Sigma}$:

$$\begin{bmatrix} \beta_{0_{group}} \\ \beta_{1_{group}} \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \mathbf{\Sigma} \right) \text{[population of varying effects]}$$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{\beta_0} & 0 \\ 0 & \sigma_{\beta_1} \end{pmatrix} \mathbf{\Omega} \begin{pmatrix} \sigma_{\beta_0} & 0 \\ 0 & \sigma_{\beta_1} \end{pmatrix} \text{[construct covariance matrix]}$$

adapted from McElreath 2023

Varying effects maths

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\mu_i, \sigma) && \text{[likelihood]} \\
\mu_i &= \beta_{0[group]} + \beta_{1[group]} x_{1[i]} && \text{[linear model]}
\end{aligned}
$$

Then comes the matrix of varying intercepts and slopes, with it's covariance matrix $\mathbf{\Sigma}$:

$$
\begin{bmatrix} \beta_{0_{group}} \\ \beta_{1_{group}} \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \mathbf{\Sigma} \right) \text{[population of varying effects]}
$$

$$
\mathbf{\Sigma} = \begin{pmatrix} \sigma_{\beta_0} & 0 \\ 0 & \sigma_{\beta_1} \end{pmatrix} \mathbf{\Omega} \begin{pmatrix} \sigma_{\beta_0} & 0 \\ 0 & \sigma_{\beta_1} \end{pmatrix} \text{[construct covariance matrix]}
$$

- $\mathbf{\Omega}$ is a correlation matrix

adapted from McElreath 2023

Let's break that down

- Math on the previous slide says that each group has a $\beta_{0[group]}$ and $\beta_{1[group]}$ with a prior distribution defined by the two dimensional Gaussian distribution with means β_0 and β_1 and covariance matrix $\mathbf{\Sigma}$

adapted from McElreath 2023

Let’s break that down

- Math on the previous slide says that each group has a $\beta_{0[group]}$ and $\beta_{1[group]}$ with a prior distribution defined by the two dimensional Gaussian distribution with means β_0 and β_1 and covariance matrix $\mathbf{\Sigma}$
  - This is a multivariate normal distribution

adapted from McElreath 2023

Let's break that down

- Math on the previous slide says that each group has a $\beta_{0[group]}$ and $\beta_{1[group]}$ with a prior distribution defined by the two dimensional Gaussian distribution with means β_0 and β_1 and covariance matrix $\mathbf{\Sigma}$
  - This is a multivariate normal distribution
- The final line defines how we construct our covariance matrix $\mathbf{\Sigma}$ by factoring it into a diagonal matrix of σ_{β_0} and σ_{β_1} and a correlation matrix $\mathbf{\Omega}$

adapted from McElreath 2023

Let's break that down

- Math on the previous slide says that each group has a $\beta_{0[group]}$ and $\beta_{1[group]}$ with a prior distribution defined by the two dimensional Gaussian distribution with means β_0 and β_1 and covariance matrix $\mathbf{\Sigma}$
  - This is a multivariate normal distribution
- The final line defines how we construct our covariance matrix $\mathbf{\Sigma}$ by factoring it into a diagonal matrix of σ_{β_0} and σ_{β_1} and a correlation matrix $\mathbf{\Omega}$
  - Several ways to construct $\mathbf{\Sigma}$, but splitting it into standard deviations, σ_{β_0} and σ_{β_1}, and a correlation matrix $\mathbf{\Omega}$ helps with understanding

adapted from McElreath 2023

Let's break that down

- Math on the previous slide says that each group has a $\beta_{0[group]}$ and $\beta_{1[group]}$ with a prior distribution defined by the two dimensional Gaussian distribution with means β_0 and β_1 and covariance matrix $\mathbf{\Sigma}$
  - This is a multivariate normal distribution
- The final line defines how we construct our covariance matrix $\mathbf{\Sigma}$ by factoring it into a diagonal matrix of σ_{β_0} and σ_{β_1} and a correlation matrix $\mathbf{\Omega}$
  - Several ways to construct $\mathbf{\Sigma}$, but splitting it into standard deviations, σ_{β_0} and σ_{β_1}, and a correlation matrix $\mathbf{\Omega}$ helps with learning
- Compare this with a standard normal distribution which takes a mean and a standard deviation

adapted from McElreath 2023

The correlation matrix

- For this simple example, the correlation matrix looks like

$$\mathbf{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

adapted from McElreath 2023

The correlation matrix

- For this simple example, the correlation matrix looks like

$$\mathbf{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

- where ρ is the correlation between β_0 and β_1

adapted from McElreath 2023

The correlation matrix

- For this simple example, the correlation matrix looks like

$$\mathbf{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

- where ρ is the correlation between β_0 and β_1
- $\mathbf{\Omega}$ gets more complicated for models with more varying slopes

adapted from McElreath 2023

Cholesky decomposition

- Note that we can take any arbitrary symmetric, positive-definite matrix $\mathbf{A}$, and factor or decompose it into

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T$$

Cholesky decomposition

- Note that we can take any arbitrary symmetric, positive-definite matrix $\mathbf{A}$, and factor or decompose it into

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T$$

- where $\mathbf{L}$ is a lower triangular matrix with real and positive diagonal entries and $\mathbf{L}^T$ is a transpose of $\mathbf{L}$

Cholesky Decomposition

- If we visualize a Cholesky decomposition

$$\begin{bmatrix} A_{00} & A_{01} & A_{02} \\ A_{10} & A_{11} & A_{12} \\ A_{20} & A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{00} & 0 & 0 \\ L_{10} & L_{11} & 0 \\ L_{20} & L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} L_{00} & L_{10} & L_{20} \\ 0 & L_{11} & L_{21} \\ 0 & 0 & L_{22} \end{bmatrix}$$

Cholesky Decomposition

- If we visualize a Cholesky decomposition

$$\begin{bmatrix} A_{00} & A_{01} & A_{02} \\ A_{10} & A_{11} & A_{12} \\ A_{20} & A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{00} & 0 & 0 \\ L_{10} & L_{11} & 0 \\ L_{20} & L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} L_{00} & L_{10} & L_{20} \\ 0 & L_{11} & L_{21} \\ 0 & 0 & L_{22} \end{bmatrix}$$

- This is helpful from a numerical perspective, particularly with noncentered parameterizations

Cholesky factors continued

- Note that there is a lot of convenient linear algebra that can be done with Cholesky factors of covariance matrices $\mathbf{L}$ or of correlation matrices $\mathbf{L_{corr}}$
  - For example,

$$\mathbf{L} = \begin{pmatrix} \sigma_{\beta_0} & 0 \\ 0 & \sigma_{\beta_1} \end{pmatrix} \mathbf{L_{corr}}$$

- See this link for a useful review

Cholesky factors

```r
# create a correlation matrix and declare sigmas
OMEGA <- matrix(c(1, 0.7, 0.7, 1), nrow = 2)
sigmas <- c(1, 2) # sd_b0, sd_b1

OMEGA
```

```
     [,1] [,2]
[1,]  1.0  0.7
[2,]  0.7  1.0
```

```r
sigmas
```

```
[1] 1 2
```

```r
# note also
diag(sigmas) # diagonal matrix
```

```
     [,1] [,2]
[1,]    1    0
[2,]    0    2
```

Cholesky factors

```r
# calculate covariance matrix:
SIGMA <- diag(sigmas) %*% OMEGA %*% diag(sigmas)
SIGMA
```

```
     [,1] [,2]
[1,]  1.0  1.4
[2,]  1.4  4.0
```

Cholesky factors

```r
# convert Cholesky factor of correlation matrix to covariance Cholesky fact
L_corr <- t(chol(OMEGA)) # note chol() returns upper triangular matrix
diag(sigmas) %*% L_corr
```

```
     [,1]     [,2]
[1,]  1.0 0.000000
[2,]  1.4 1.428286
```

```r
t(chol(SIGMA)) # L of SIGMA
```

```
     [,1]     [,2]
[1,]  1.0 0.000000
[2,]  1.4 1.428286
```

Cholesky factors

```r
# Cholesky factor of correlation matrix to
# covariance matrix Cholesky factor:
L_corr <- t(chol(OMEGA)) # note chol() returns upper tri
Lambda <- diag(sigmas) %*% L_corr

t(chol(SIGMA)) # L of SIGMA
```

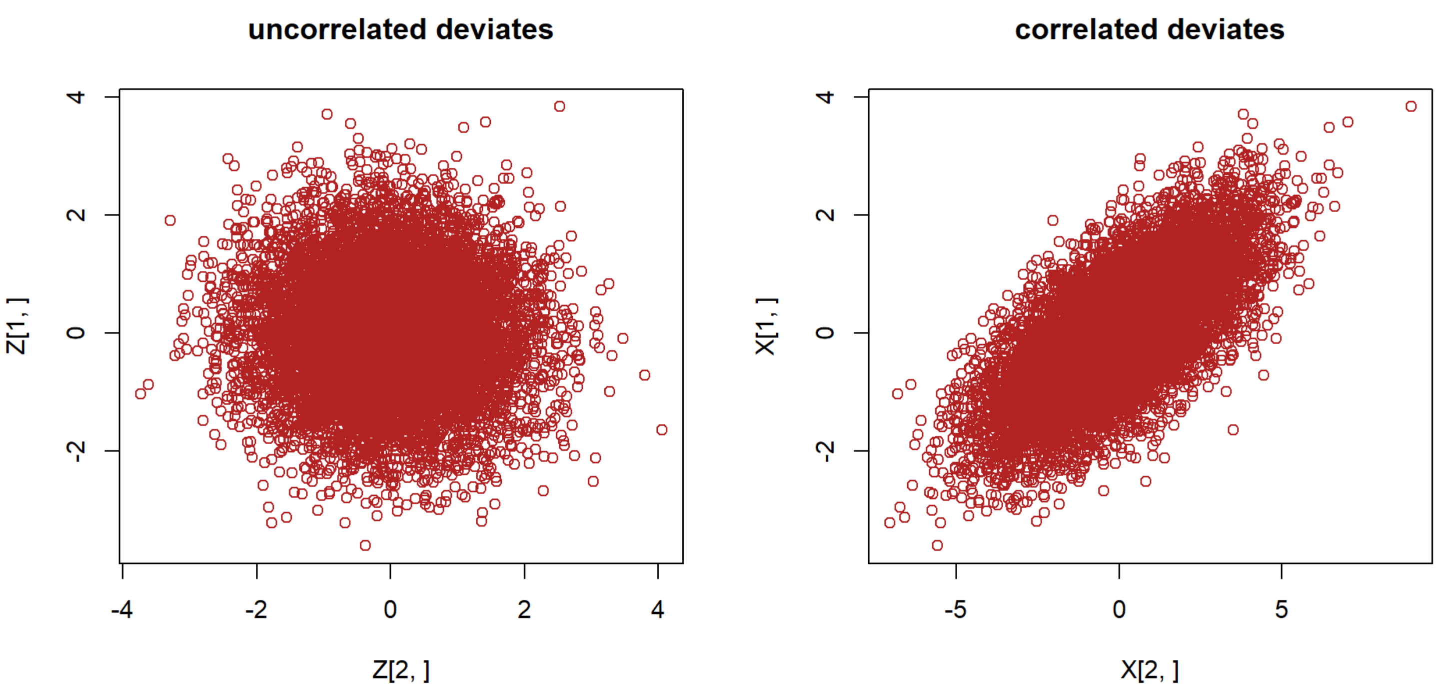
```
     [,1]     [,2]
[1,]  1.0 0.000000
[2,]  1.4 1.428286
```

```r
Lambda
```

```
     [,1]     [,2]
[1,]  1.0 0.000000
[2,]  1.4 1.428286
```

Cholesky factors

```r
# generate random values with desired covariance
Z <- rbind(rnorm(1e4),rnorm(1e4))
X <- Lambda %*% Z
par(mfrow=c(1,2))
plot(Z[1,]~Z[2,], main = "uncorrelated deviates", col = "firebrick")
plot(X[1,]~X[2,], main = "correlated deviates", col = "firebrick")
```



A problem

- People want to know the extent to which juvenile walleye growth rate is density dependent
    - Has implications for both basic ecology and management
- DNR Biologists go to a collection of lakes and measure length of age-0 walleye in fall as a proxy of juvenile growth rate
- Each year, the biologists attempt to go to 30 lakes in total (weather pending)
    - They also conduct surveys to get an estimate of juvenile density
- Let’s simulate some fake data representing this problem, and then build some Stan models to recover
- go to the `varying_effects.r` script

see also Cahill et al. 2020

Hyperpriors for varying effects model

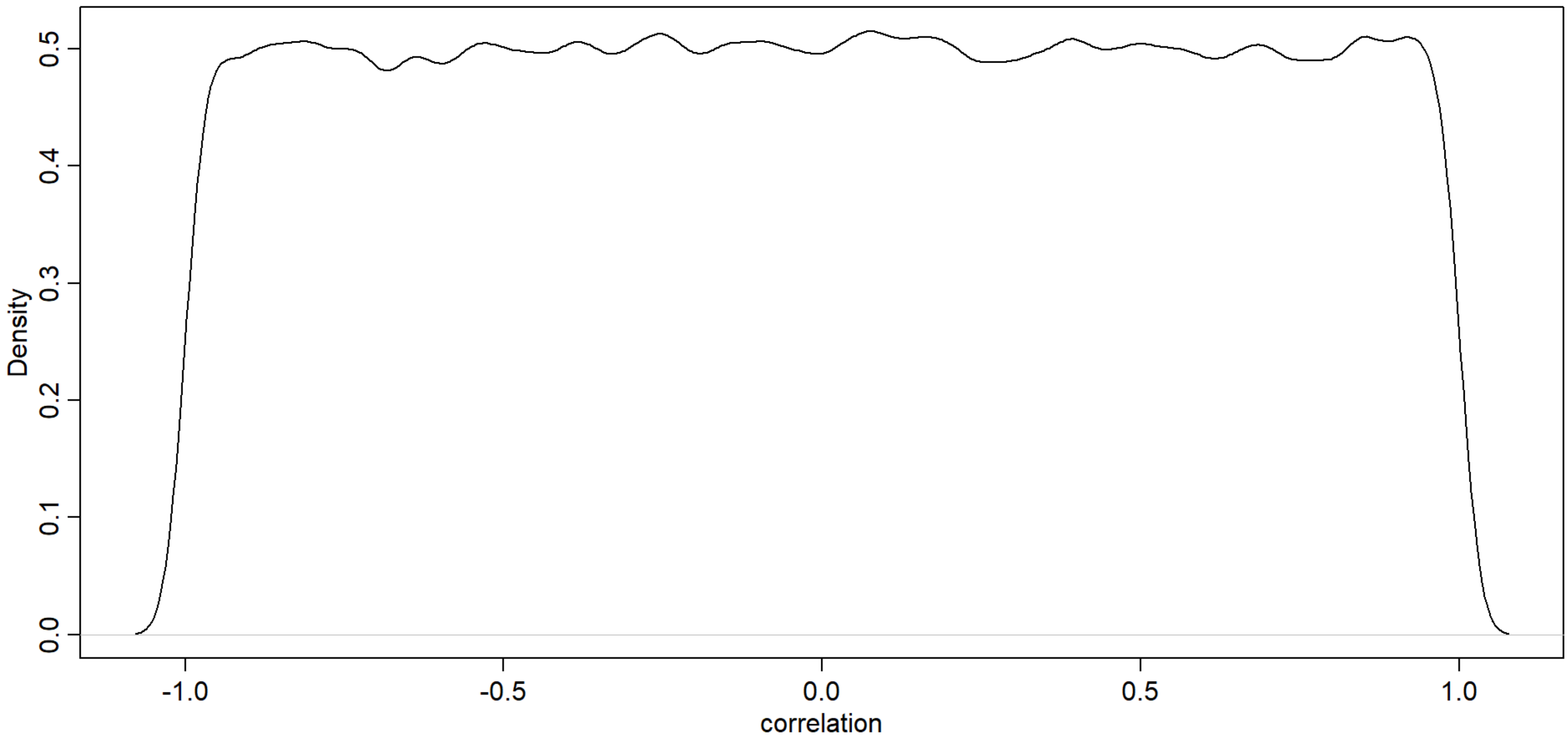
$$
\begin{aligned}
\beta_0 &\sim \text{Normal}(0, 25) && \text{[prior for average intercept]} \\
\beta_1 &\sim \text{Normal}(0, 25) && \text{[prior for average slope]} \\
\sigma &\sim \text{Exponential}(0.01) && \text{[prior for stddev within group]} \\
\sigma_{\beta_0} &\sim \text{Exponential}(0.01) && \text{[prior stddev among intercepts]} \\
\sigma_{\beta_1} &\sim \text{Exponential}(0.01) && \text{[prior stddev among intercepts]} \\
\mathbf{\Omega} &\sim \text{LKJ corr}(2) && \text{[prior for correlation matrix]}
\end{aligned}
$$

- $\text{LKJ corr}(2)$ defines a weakly informative prior on ρ that is skeptical of extreme correlations near -1 or 1

adapted from McElreath 2023

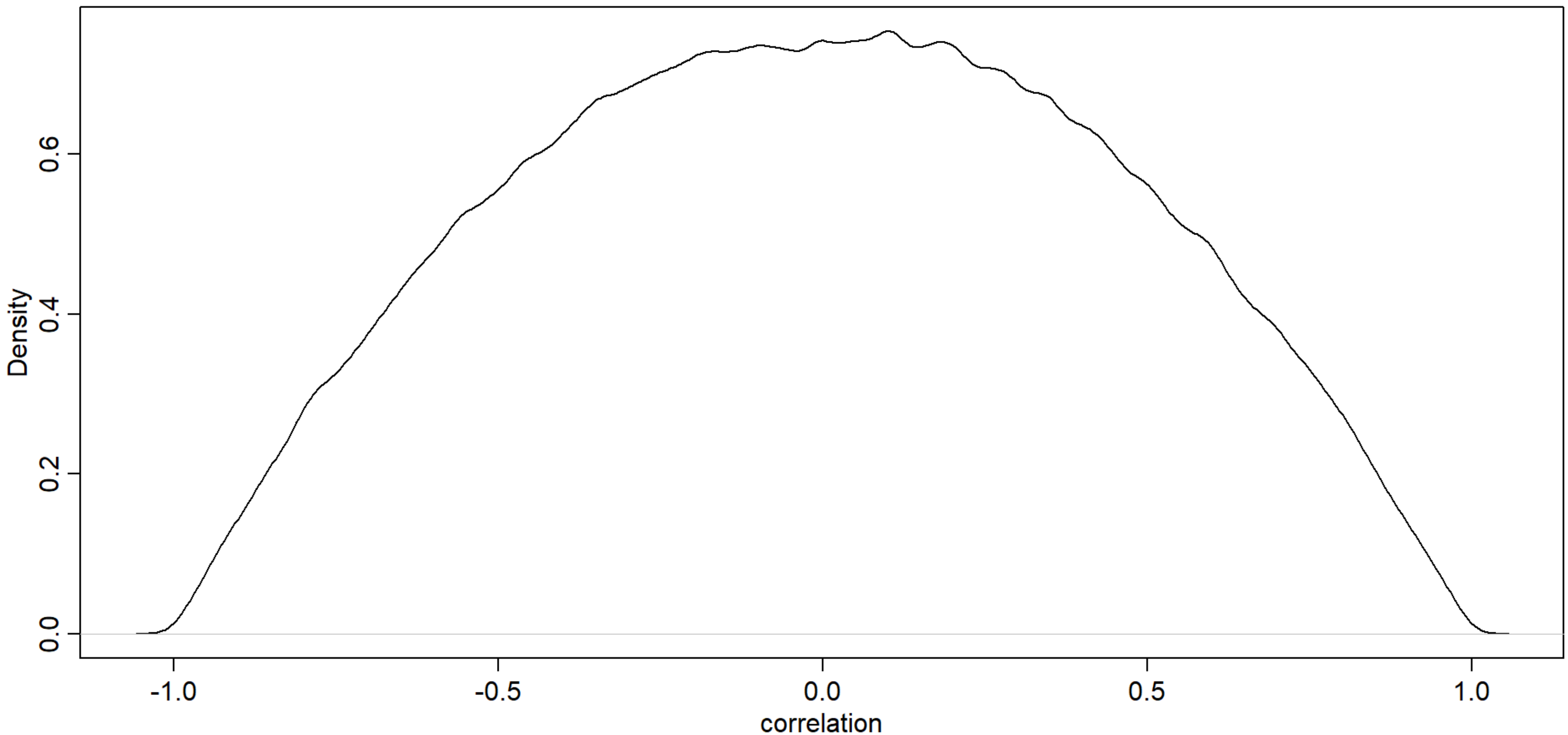
Visualizing the LKJcorr prior

- $\text{LKJ corr}(1)$



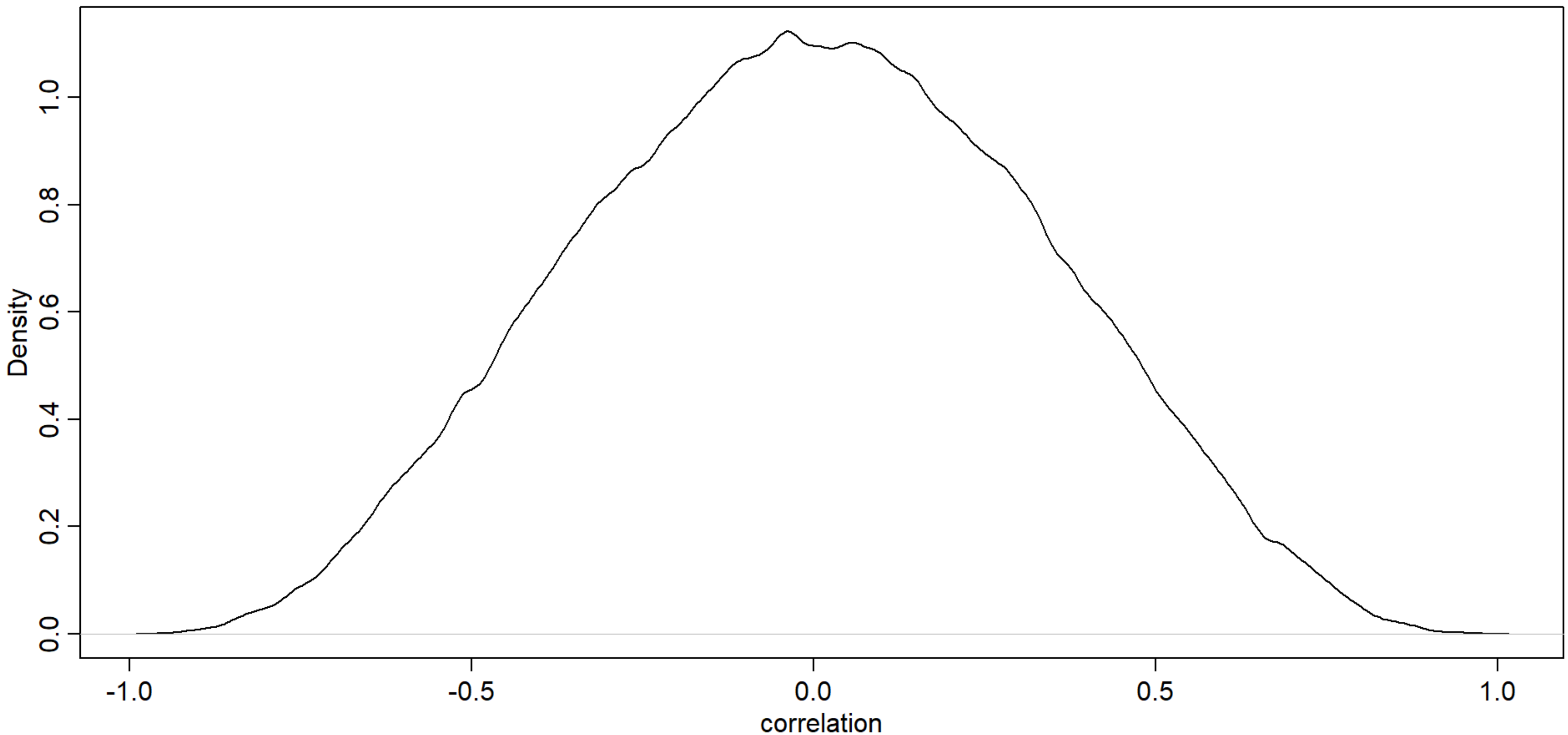
Visualizing the LKJcorr prior

- $\mathrm{LKJ\ corr}(2)$



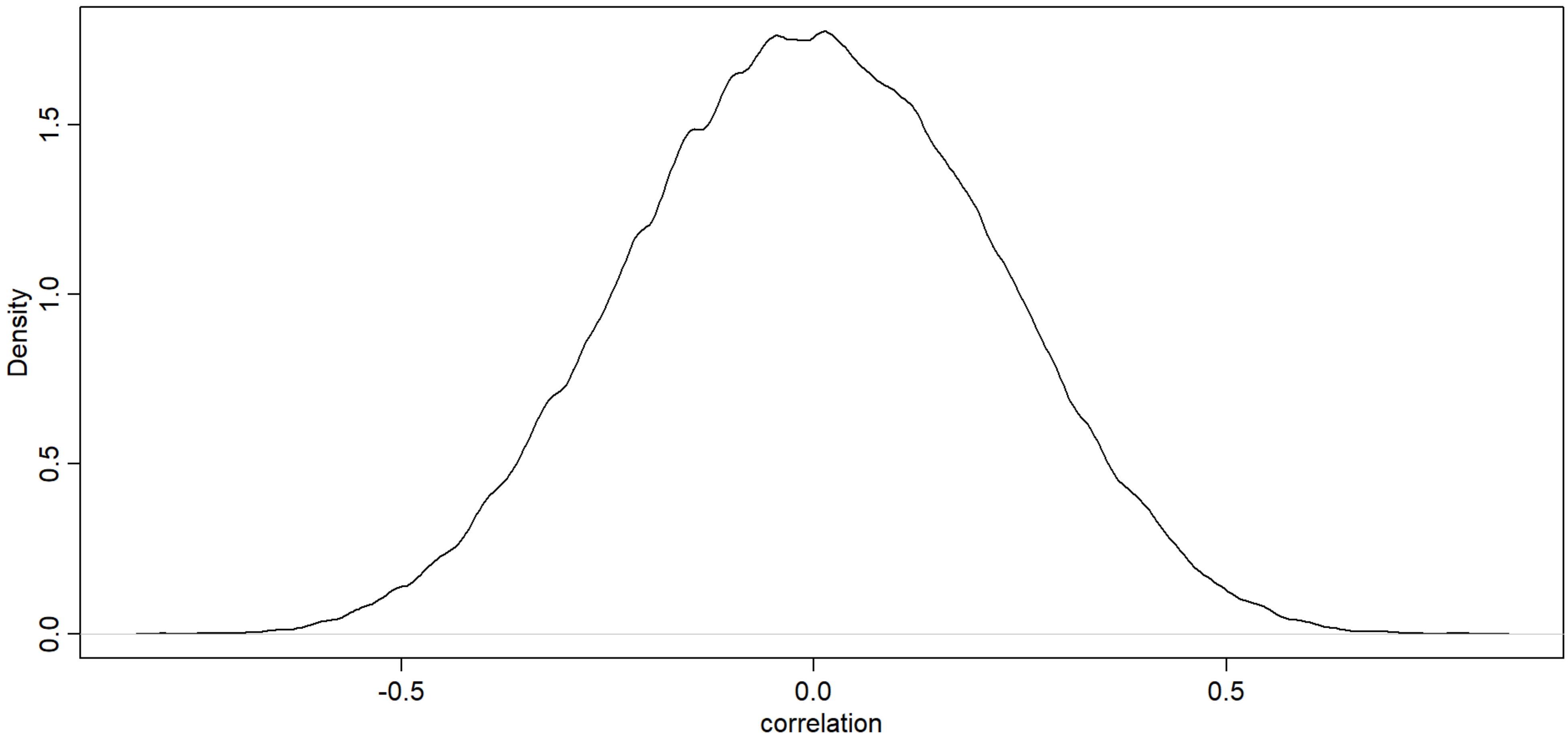
Visualizing the LKJcorr prior

- $\text{LKJ corr}(4)$



Visualizing the LKJcorr prior

- $\text{LKJ corr}(10)$



Wrap up

- Introduced a powerful extension to mixed effects models: varying effects
- Went through a bunch of math to show how to play with multivariate normal distributions
- Simulated an example, conducted prior predictive checks, estimated the model
- Showed a non-centered version of this model as well
- Up next, spatial random effects

References

Cahill et al. 2020. A spatial-temporal approach to modeling somatic growth across inland fisheries landscapes. CJFAS.

Gelman, A. and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models

McElreath 2023. Statistical Rethinking.

Simpson's paradox Wikipedia: https://en.wikipedia.org/wiki/Simpson%27s_paradox

https://mlisi.xyz/post/simulating-correlated-variables-with-the-cholesky-factorization/