WEEK 2: FITTING BAYESIAN (GENERAL) LINEAR MODELS

STA465: Theory and Methods for Complex Spatial Data

Instructor: Dr. Vianey Leos Barajas

QUICK ANNOUNCEMENT:

- Homework 1 will be posted later today

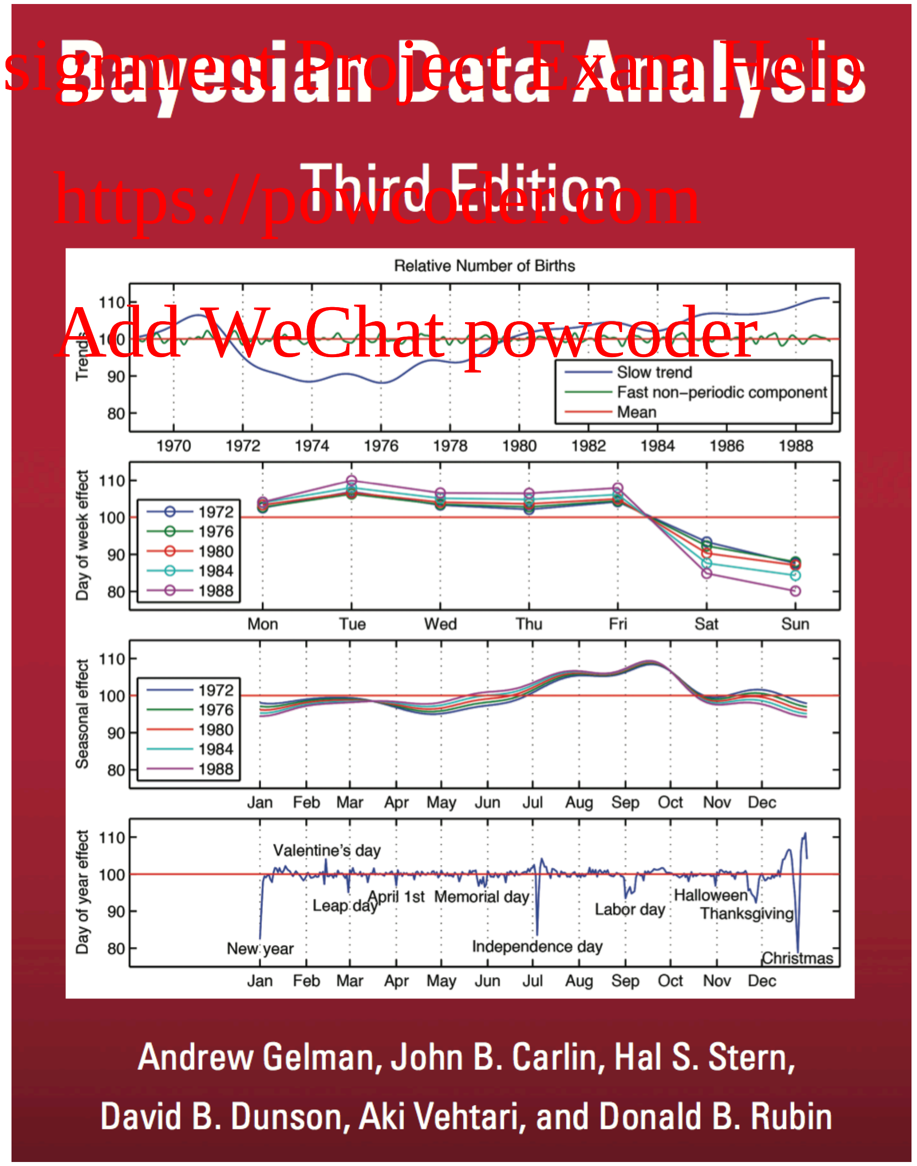
- Due Friday, January 29th, 23:59 EST (Time in Toronto)

- You'll run simulations, fit basic Bayesian models, and more simulations — most of the code will be provided, you'll primarily have to modify it to suit your needs

BAYESIAN DATA ANALYSIS

Book Website: *http://www.stat.columbia.edu/~gelman/book/*

PDF of book: *http://www.stat.columbia.edu/~gelman/book/BDA3.pdf*



“VISUALIZATION IN A BAYESIAN WORKFLOW”

PM2.5 Paper: *https://rss.onlinelibrary.wiley.com/doi/full/10.1111/rssa.12378*

FITTING A BAYESIAN LINEAR REGRESSION MODEL

BACK TO THE SIMPLE LINEAR REGRESSION MODEL

- Consider a linear regression model

$$y_i \mid \beta, \sigma \sim N(X\beta, \sigma^2 I)$$

- To fit this model in a Bayesian framework, we have to specify prior distributions for all parameters of the model: β, σ

- Three general categories of prior distributions:
  - Informative
  - Weakly informative
  - *"Noninformative"*

SPECIFYING PRIORS

- We always include information about the parameter values through the prior.
- So-called ‘noninformative’ priors can actually be quite informative in practice…
- We can assess the information we include in our prior through simulation!

SOME READING ON PRIOR DISTRIBUTIONS

Article

The Prior Can Often Only Be Understood in the Context of the Likelihood

http://www.stat.columbia.edu/~gelman/research/published/entropy-19-00555-v2.pdf

PRIOR FOR β :

$$y_i \mid \beta, \sigma \sim N(X\beta, \sigma^2 I)$$

- Let's consider three priors:

  - Normal(0.4, 0.2)
  - Normal(0, 10)
  - Normal(0, 1000)

FROM LAST WEEK:

```r
#-----------------------------------------------
# Setting the values of the parameters
#-----------------------------------------------

beta0  <- 1
beta1 <- 0.5
sigma <- 1

#-----------------------------------------------
# Simulating covariate values
#-----------------------------------------------
set.seed(17)

x <- runif(n = 100, min = 1, max=5)
y.mean <- beta0 + beta1*x
y <- rnorm(n = 100,
           mean = y.mean,
           sd = sigma)

sim.data <- tibble(x,y, y.mean)
```

SIMULATING DATA FROM THE THREE PRIORS: N(0.4, 0.2)

SIMULATING DATA FROM THE THREE PRIORS: N(0, 10)

SIMULATING DATA FROM THE THREE PRIORS: N(0, 1000)

SCALE OF X:

- Prior distributions will depend on the values of X!

- Say you have two covariates,

x_1 = maximum daily temperature in Toronto (in F)

x_2 = the amount of snow that falls in Toronto (in m)

- What do equal priors imply about the process *a priori*?

READING MORE ABOUT PRIORS:

- https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

5 levels of priors

- Flat prior (not usually recommended);
- Super-vague but proper prior: normal(0, 1e6) (not usually recommended);
- Weakly informative prior, very weak: normal(0, 10);
- Generic weakly informative prior: normal(0, 1);
- Specific informative prior: normal(0.4, 0.2) or whatever. Sometimes this can be expressed as a scaling followed by a generic prior: theta = 0.4 + 0.2*z; z ~ normal(0, 1);

PRIORS CAN BE USED IN LOTS OF WAYS:

- To induce 'sparsity'
- To impose structure in the model

- To include information in data-limited settings

- We're open and direct about the *bias* we include in our models through specification of the prior

GENERALIZED LINEAR MODELS

GENERALIZED LINEAR REGRESSION MODEL EXAMPLE:

- Consider a generalized linear regression model

$$y_i \mid \lambda \sim Pois(\lambda)$$

- What values does λ take on? What's the parameter space?

- How do we include covariates?

POISSON DISTRIBUTION:

- With no covariates, we just have that our observations are generated from a Poisson distribution:

$$y_i | \lambda \sim Pois(\lambda)$$

- What are some sensible prior distributions for λ?

  - What about N(0, 1)?
  - What else?

DISTRIBUTIONS WITH 'RESTRICTED' PARAMETER SPACES

$$\lambda > 0$$

TRANSFORMING λ

- To include covariates, we do not generally do:

$$\lambda = \beta_0 + \beta_1 x$$

- Can we think of why?

TRANSFORMING λ

- How to include covariates then?

- We transform λ using a function that maps the positive real values onto the real line.

  — One option is to use log() [implied natural log]

  — Any function that maps positive real values onto the entire real line works!

LOG(λ)

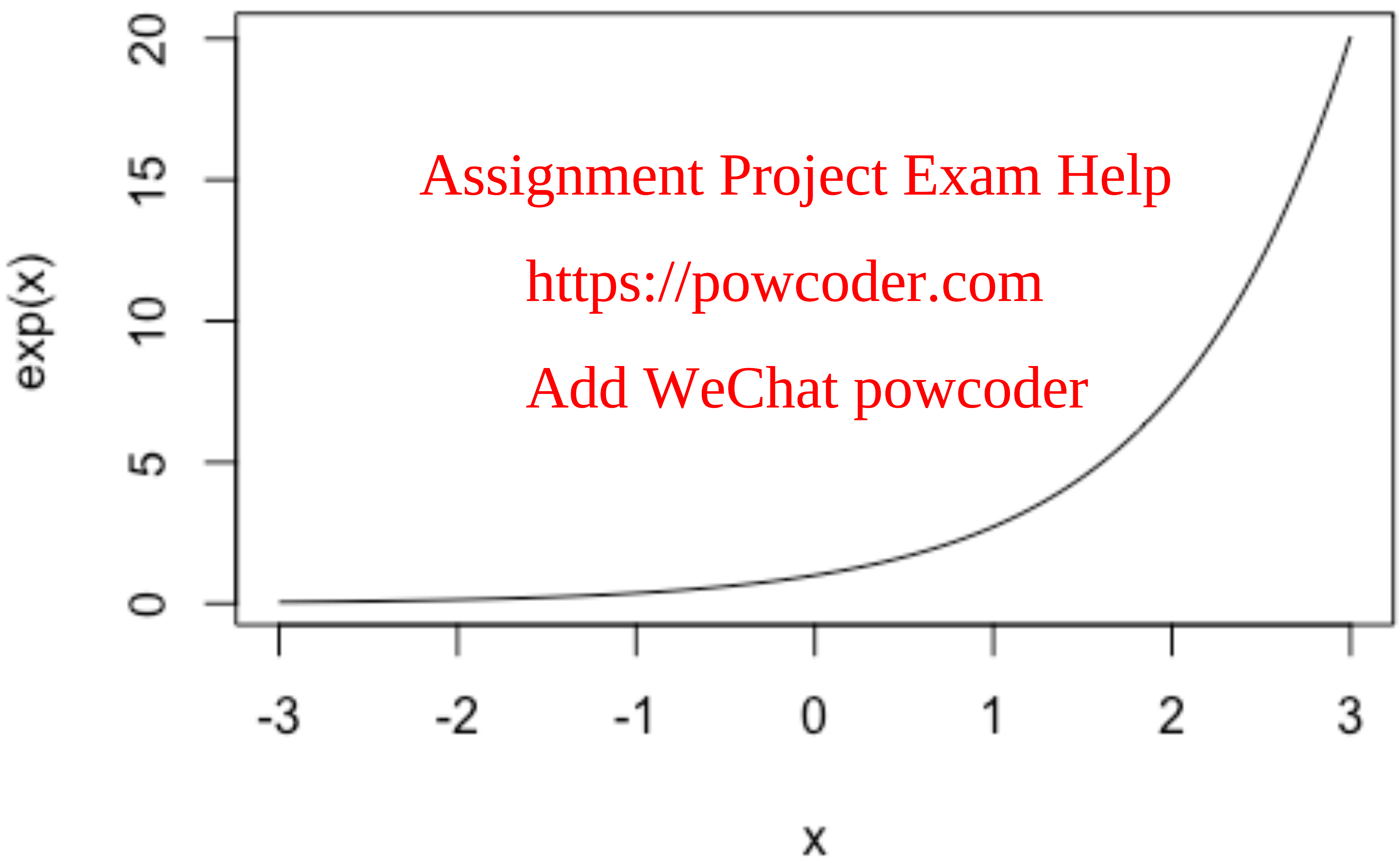
- $log(\lambda) = \beta_0 + \beta_1 x$

- Now we're back in a linear regression framework, we can specify prior distributions for the values of β

- We can consider the following priors:

  - Normal(0, 1)
  - Normal(0, 10)
  - Normal(0, 1000)

N(0,1) UNDER AN EXPONENTIAL FUNCTION



MAPPING λ BACK TO POSITIVE VALUES:

- Let's map λ back to the positive real line:

$$\lambda = exp(\beta_0 + \beta_1 x)$$

$$\lambda = exp(\beta_0)exp(\beta_1 x)$$

- Our specification of prior distributions should now take this into consideration.

If $\beta_0 = 2, \beta_1 = 5, x = 1$, λ will be quite large!

MULTILEVEL MODELS

Hierarchical models
Random effects models

ALLOWING THE MEAN TO VARY BY GROUP J:

- We can allow the mean to vary across groups:

$$y_{ij} \mid \mu, \sigma \sim N(\mu_j, \sigma^2)$$

$$\mu_j \sim N(\nu, \tau^2)$$

- Here we have to specify priors for parameters: ν, τ, σ

- Recall that in this case the value of τ controls how varied the values of μ_j can be.

SIMULATING FROM A MULTILEVEL MODEL:

HIERARCHICAL REGRESSION MODEL:

- We can allow the slope and intercept terms to vary across groups j:

$$y_{ij} \mid \mu, \beta, \sigma \sim N(\mu_j + \beta_j x_{ij}, \sigma^2)$$

$$\mu_j \sim N(\mu, \sigma_\mu^2)$$

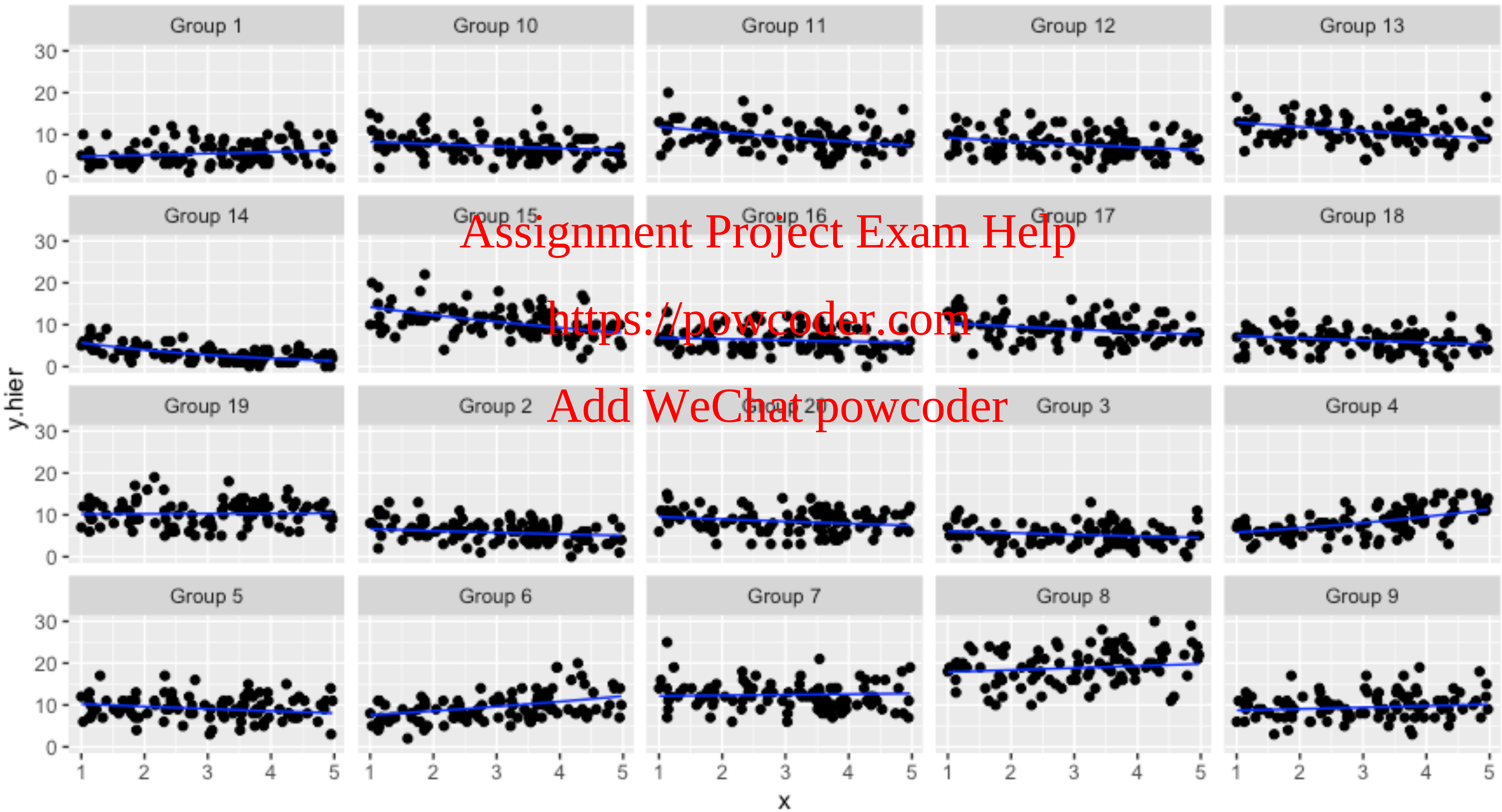
$$\beta_j \sim N(\eta, \sigma_\beta^2)$$

- Our models are starting to get more complex…how do we select appropriate priors?

- Learning about the model and prior through simulations!

SIMULATING DATA FROM A HIERARCHICAL REGRESSION MODEL:

SIMULATING DATA FROM A HIERARCHICAL REGRESSION MODEL:



SIMULATING DATA FROM A HIERARCHICAL REGRESSION MODEL:

PRIOR PREDICTIVE DISTRIBUTION SAMPLING:

POSTERIOR PREDICTIVE DISTRIBUTION SAMPLING:

POSTERIOR PREDICTIVE DISTRIBUTION SAMPLING:

MODEL VALIDATION AND COMPARISON

MODEL VALIDATION AND COMPARISON:

- Prior and posterior predictive checks (simulation):

  — On the homework!

- Residuals and other common metrics for goodness-of-fit

  — homework again!

- Model comparison: cross-validation, information criterion, simulations (week 4/5-ish)

NEXT WEEK:

Spatial maps!

Books for next week: *Geocomputation with R, Geospatial Health Data (online for free — details in syllabus)* *https://juliasilge.com/blog/texas-opioids/*