Student Seminar #9 (2018/06/26)

Bayesian Data Analysis CHAPTER 17: Metric Predicted Variable with One Metric Predictor

M2 Yuki Kajihara

Contents

Overview of Chapter17	1
17.1 Simple Linear Regression	2-3
17.2 Robust Linear Regression	4-9
17.3 Hierarchical Regression on Individuals Within Groups	10-12
17.4 Quadratic Trend and Weighted Data	13-16
17.5 Procedure and Perils for Expanding a Model	17-19

Overview of Chapter 17

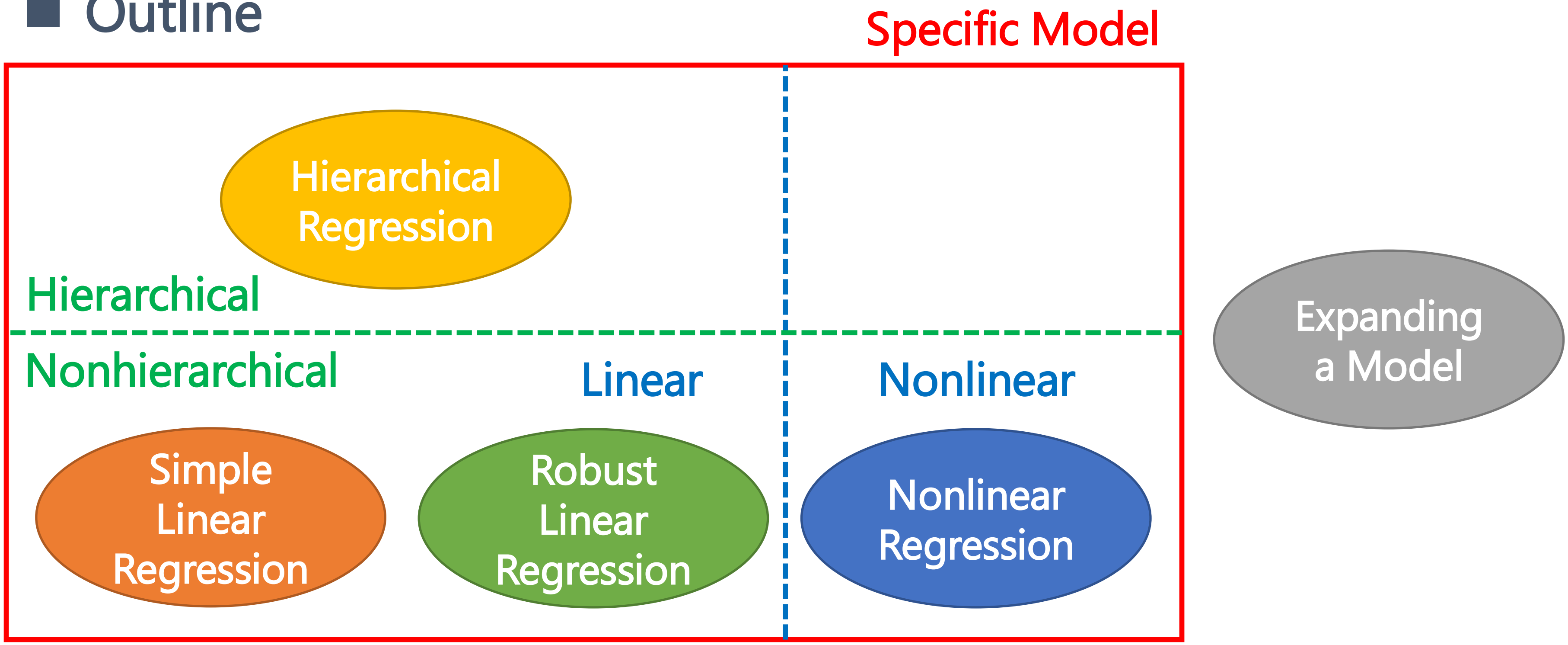
Purpose

✓ Predict <u>one metric variable</u> from <u>one metric predictor</u>

y x

ex.) weight height

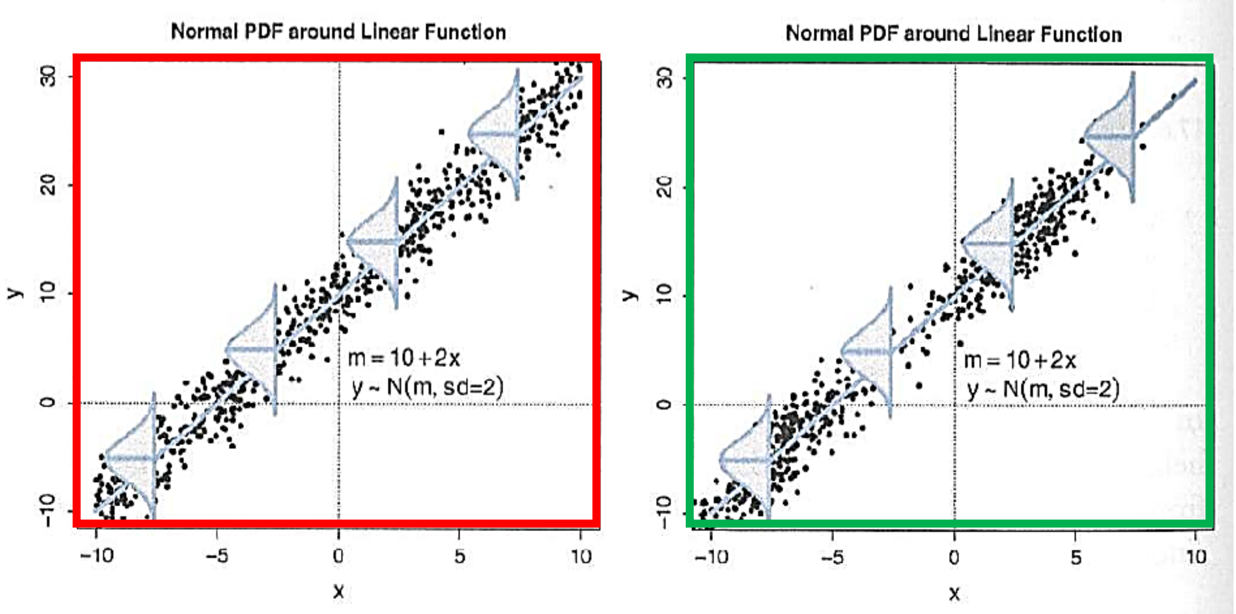
Outline



Step

- ✓ Generate any random x
- ✓ Compute the mean predicted value of y by $\underline{\mu = \beta_0 + \beta_1 x}$
- ✓ Generate random variable for datum y from a normal distribution (μ : mean, σ : standard deviation)

Figure



Generate x from
Left :
a uniform distribution
Right :
a bimodal distribution

Both shows data from **the same model**

17.1 Simple Linear Regression

Assumption

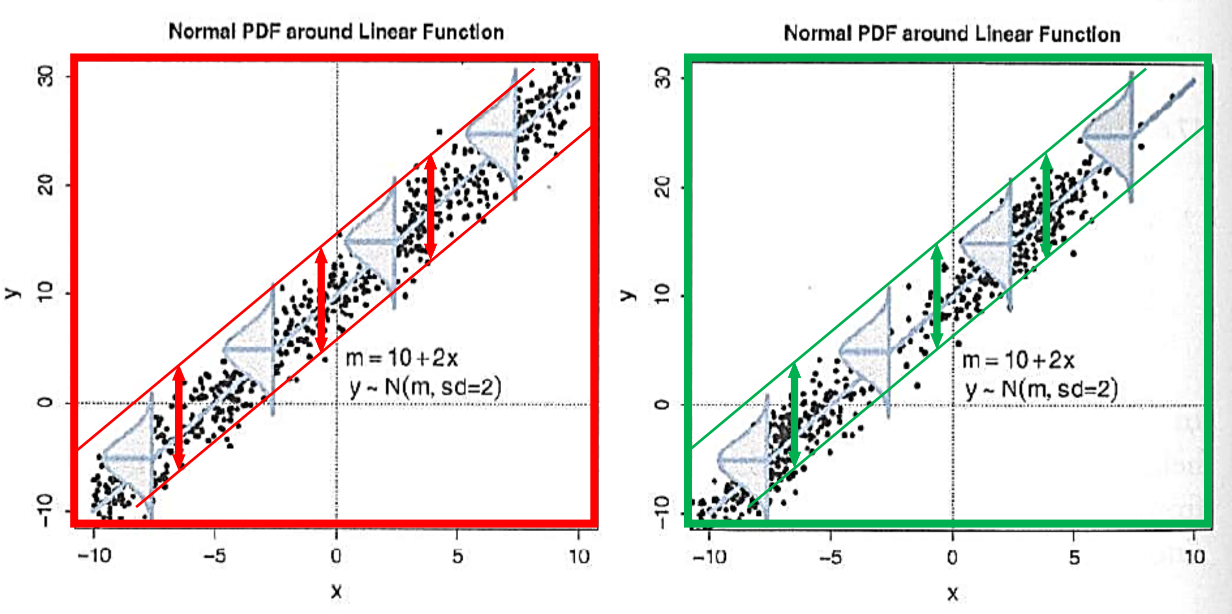
✓ *Homogeneity of variance*

At every value of x, the variance of y is the same

Note

✓ Simple Linear Regression describes **tendencies**, not **causality**

Figure



Both assumes *Homogeneity of variance*

But hard to see it on **the right figure** because there is an area which x is sparse

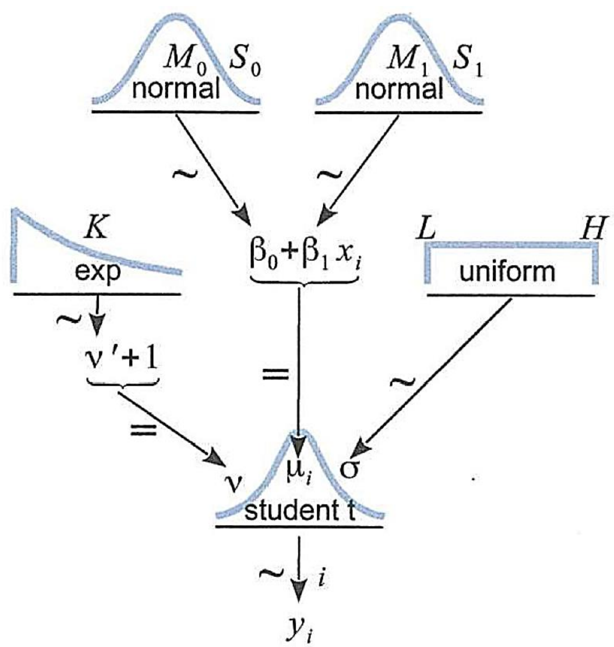
Object

✓ Data which have **outliers**

Assumption

✓ The datum y_i is a **t-distributed** random value around the central tendency $\mu_i = \beta_0 + \beta_1 x_i$

Diagram



Goal

- ✓ Determine what combinations of β_0, β_1, σ, ν are credible, given the data

  The answer (from Bayes' rule) :

$$p(\beta_0, \beta_1, \sigma, \nu | D) = \frac{p(D|\beta_0, \beta_1, \sigma, \nu)\, p(\beta_0, \beta_1, \sigma, \nu)}{\iiiint d\beta_0\, d\beta_1\, d\sigma\, d\nu\, p(D|\beta_0, \beta_1, \sigma, \nu)\, p(\beta_0, \beta_1, \sigma, \nu)}$$

  → Complicated…

  Use JAGS or Stan!

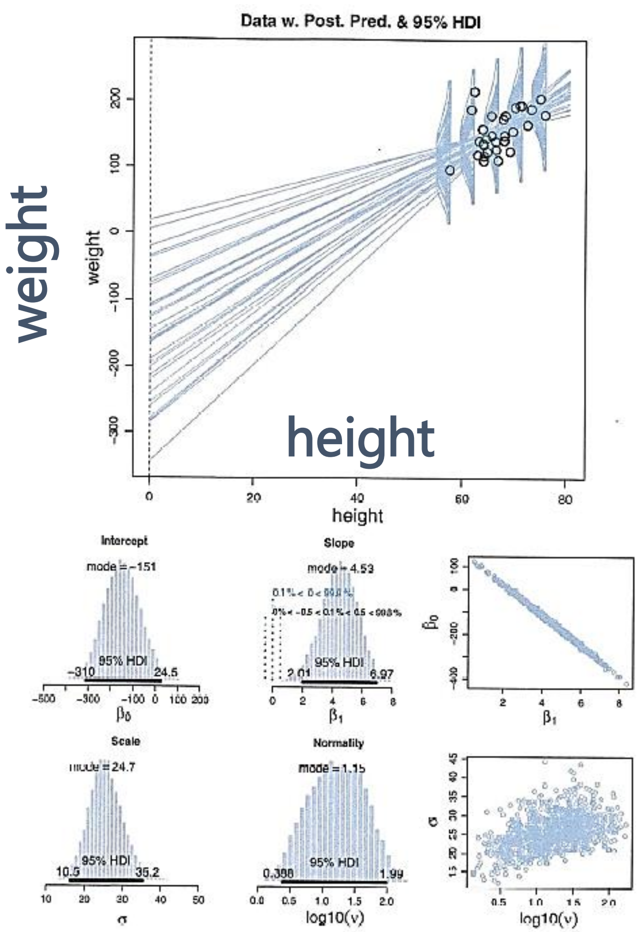
What we have to do

- ✓ Specify sensible priors
- ✓ Make sure that the MCMC process generates a trustworthy sample that is converged and well mixed

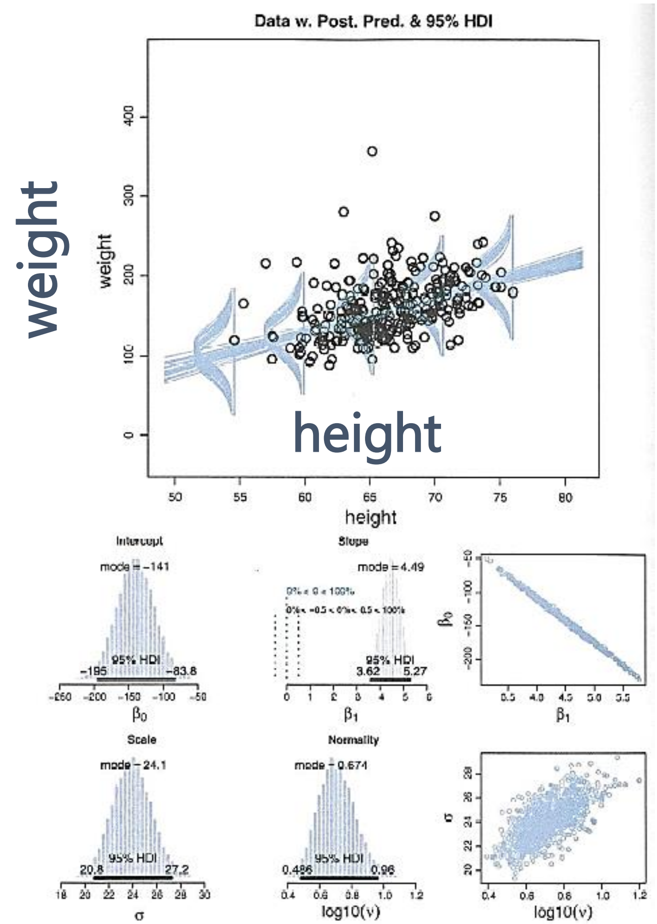
  → Talk about it later

■ Figure

N = 30



N = 300



■ Problem with using raw data

✓ Parameter-correlation problem

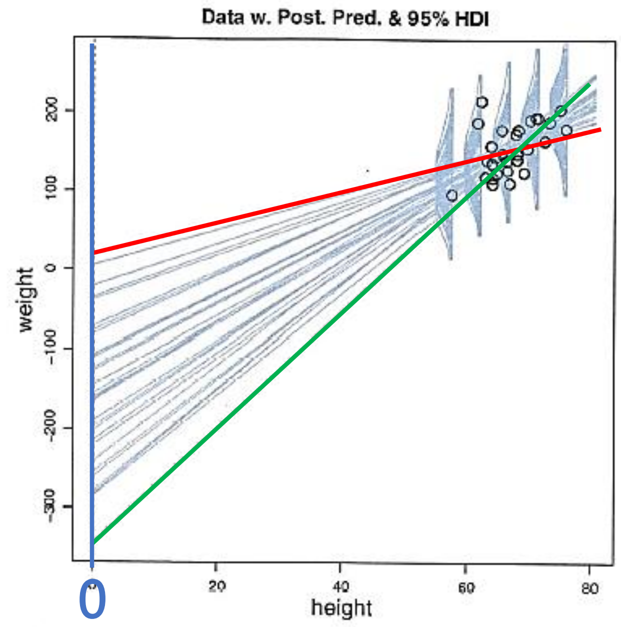
The credible slopes and intercepts trade off

When the slope is small, the intercept is big

When the slope is big, the intercept is small

→ MCMC sampling is difficult

Two parameter values change slowly



■ Ways to make the sampling more efficient

✓ Change the sampling algorithm → Stan : HMC

✓ Transform the data → **JAGS : Standardization**

17. 2 Robust Linear Regression

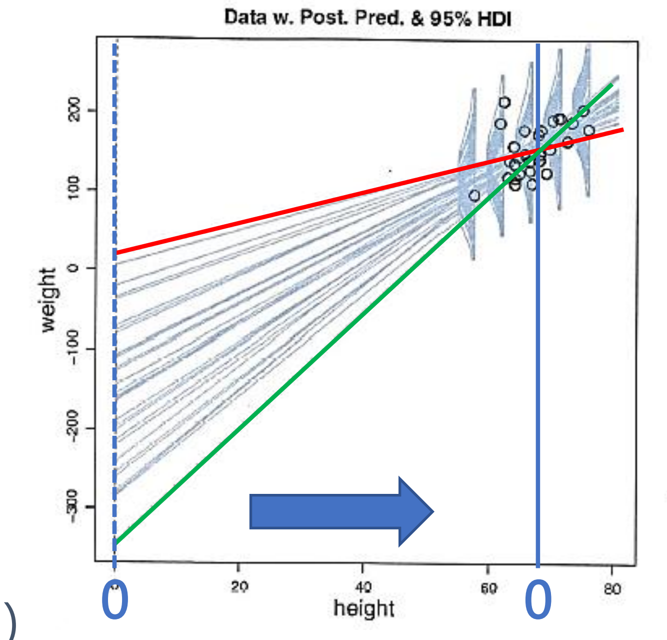
Mean centering

✓ Slide the axis

so that zero falls under the mean

→ The Slope changes without any big changes on the intercept

→ Solve parameter-correlation problem(???)



Standardize data

✓ Re-scaling the data relative to their mean(*M*) and standard deviation(*SD*): $z_x = \frac{(x - M_x)}{\mathrm{SD}_x}$ and $z_y = \frac{(y - M_y)}{\mathrm{SD}_y}$

✓ Linear Regression using standardized data

$$z_{\hat{y}} = \boxed{\zeta_0} + \boxed{\zeta_1} z_x \qquad \text{by definition of the model}$$

$$\frac{(\hat{y} - M_y)}{\mathrm{SD}_y} = \zeta_0 + \zeta_1 \frac{(x - M_x)}{\mathrm{SD}_x} \qquad \text{from Equation 17.1}$$

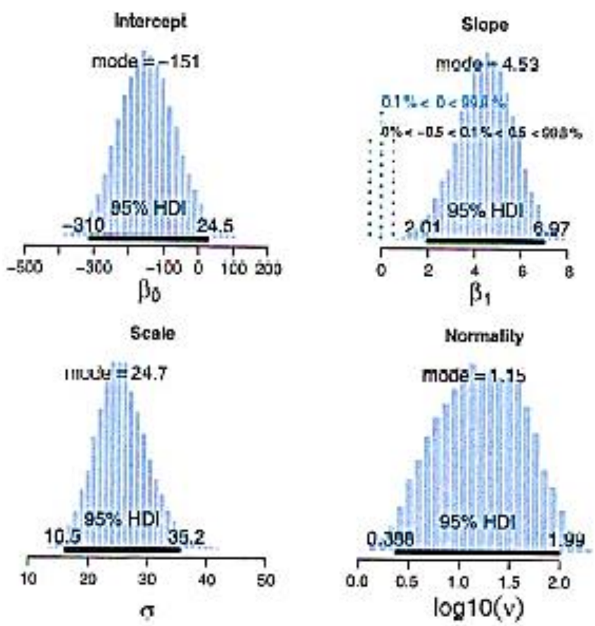
$$\hat{y} = \underbrace{\zeta_0 \mathrm{SD}_y + M_y - \zeta_1 M_x \mathrm{SD}_y / \mathrm{SD}_x}_{\beta_0} + \underbrace{\zeta_1 \mathrm{SD}_y / \mathrm{SD}_x}_{\beta_1} x \qquad (17.2)$$

$\boxed{\zeta 0}$: the slope with the data

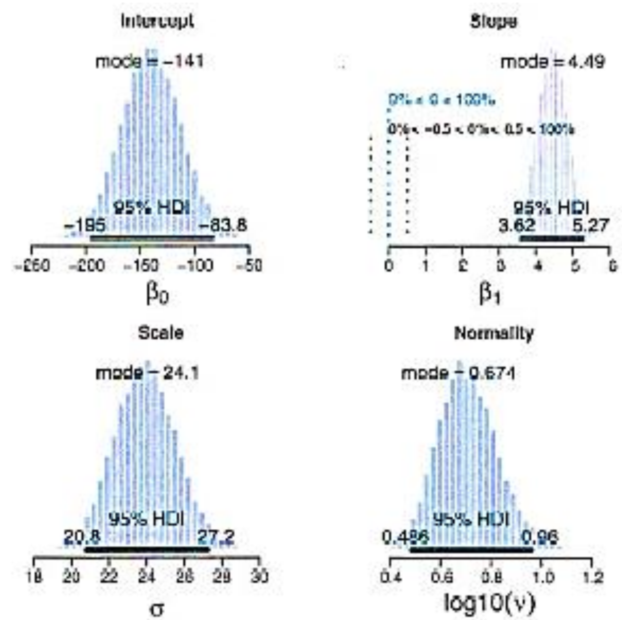
$\boxed{\zeta 1}$: the intercept with the data

Interpreting the posterior distribution

- ✓ Compare N=30 regression and N=300 one
- ✓ **The slope, intercept and scale** are about the same
- ✓ **The certainty** of the estimate for N=300 is tighter than for N=30
- ✓ **The normality parameter** for N=300 is bigger than for N=30



N=30



N=300

17. 3 Hierarchical Regression on Individuals Within Groups

■ Object

✓ Data that each individual contributes multiple observations

ex.) Reading-ability scores of children across several years

Family income for different size of the family, for different regions

■ Assumption

✓ Each individual is representative of the group

→ Every individual informs the estimate of the group slope and intercept

Get sharing of information across individuals

■ Goal

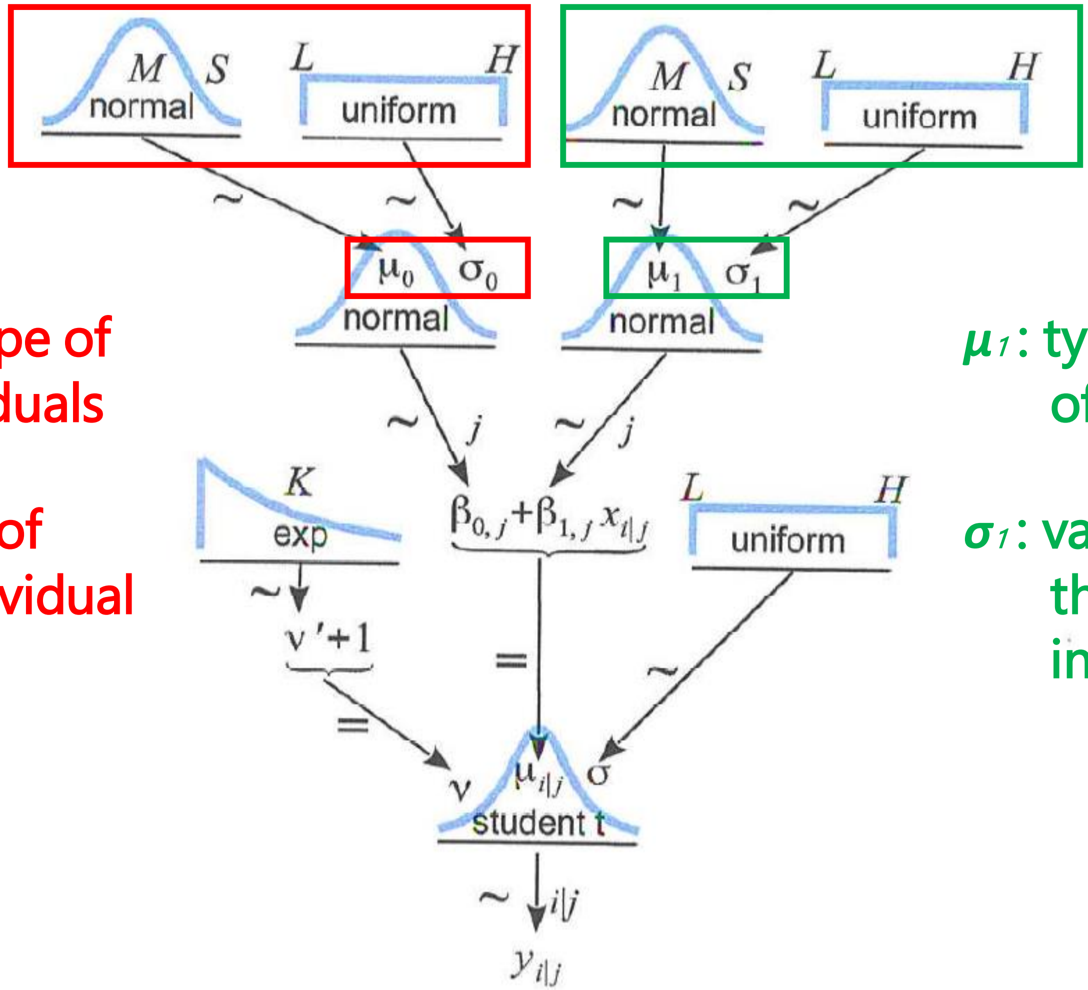
✓ describe each individual with a linear regression

✓ Estimate the typical slope and intercept of the group overall

17. 3 Hierarchical Regression on Individuals Within Groups

■ The model and implementation in JAGS

✓ Diagram



μ_0: typical slope of the individuals

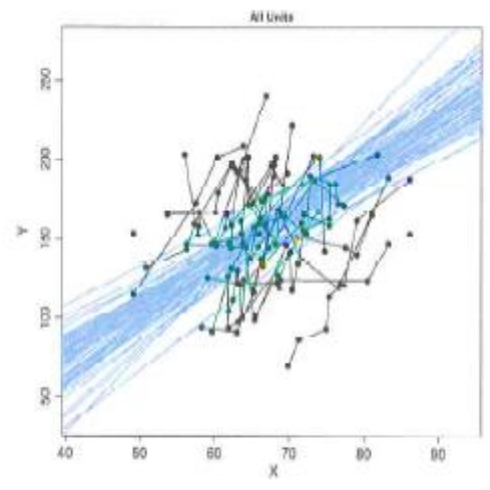
σ_0: variability of those individual slopes

μ_1: typical intercepts of the individuals

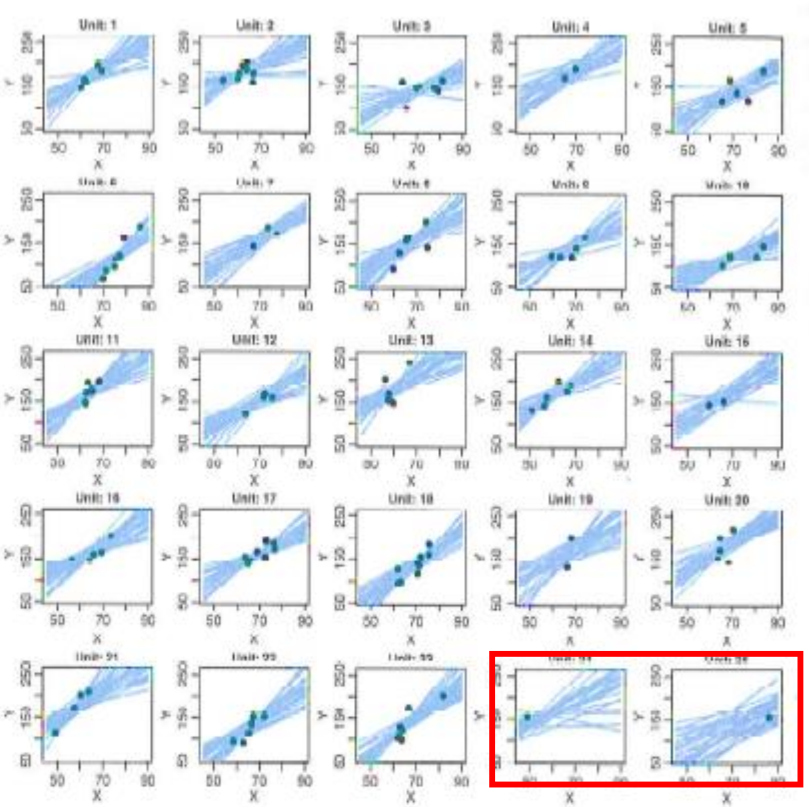
σ_1: variability of those individual intercepts

The posterior distribution : Shrinkage and prediction

- ✓ Overall : Clearly positive by integrating each individual slope
- ✓ Individual : Notable shrinkage of the estimates of the individuals
- ✓ The estimates are tightly constrained by each data



Overall



Individual

Object

✓ Data that y appears to have a **nonlinear** trend as x increases

Example

✓ Family size and family income for each state in the U.S

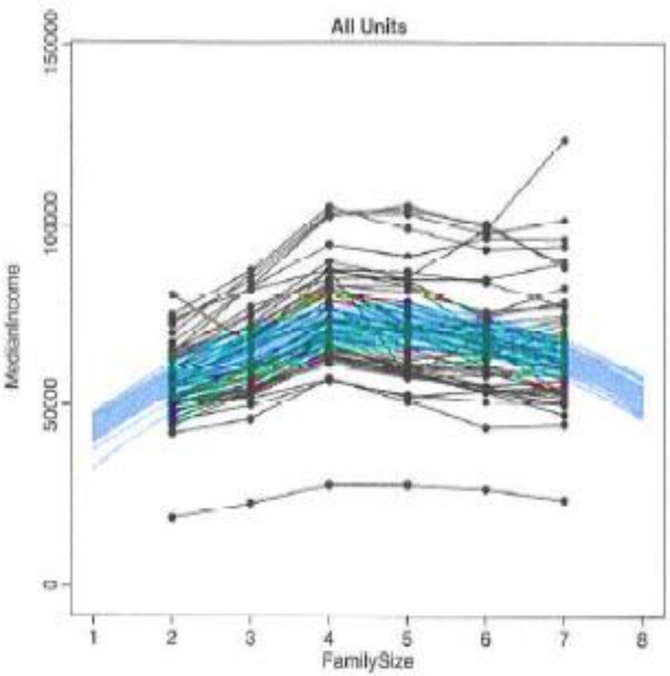
Model

✓ $\underline{\mu_i = b_0 + b_1 x + b_2 x^2}$

✓ If $b_2 = 0$ → linear model

✓ If $|b_2|$ is big

→ nonlinear model is reasonable



Median income

Family size

Nonlinear Regression using standardized data

$$z_{\hat{y}} = \boxed{\zeta_0} + \boxed{\zeta_1} z_x + \boxed{\zeta_2} z_x^2 \qquad \text{by definition of the model}$$

$$\frac{(\hat{y} - M_y)}{SD_y} = \zeta_0 + \zeta_1 \frac{(x - M_x)}{SD_x} + \zeta_2 \frac{(x - M_x)^2}{SD_x^2} \qquad \text{from Equation 17.1}$$

$$\hat{y} = \underbrace{\zeta_0 SD_y + M_y - \zeta_1 M_x SD_y / SD_x + \zeta_2 M_x^2 SD_y / SD_x^2}_{\beta_0}$$

$$+ \underbrace{\left(\zeta_1 SD_y / SD_x - 2\zeta_2 M_x SD_y / SD_x^2\right)}_{\beta_1} x + \underbrace{\zeta_2 SD_y / SD_x^2}_{\beta_2} x^2 \qquad (17.3)$$

$\boxed{\zeta 0}, \boxed{\zeta 1}, \boxed{\zeta 2}$: coefficients using the data

Weighting data

✓ The data report the median income based on different numbers of families at each size

→ every median has a different amount of sampling noise

✓ Consider "margin of error"

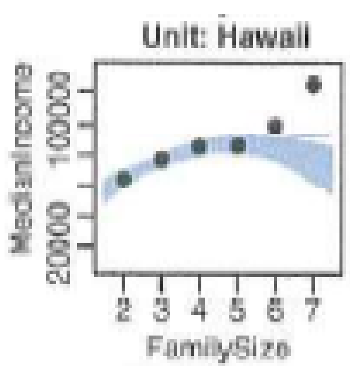
If margin of error is **high**, noise parameter should be **increased**

If margin of error is **small**, noise parameter should be **decreased**

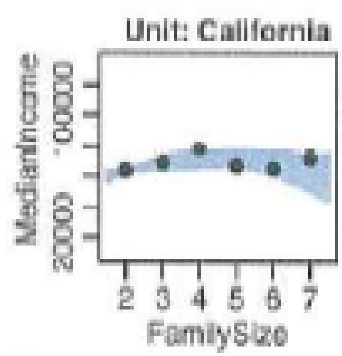
Results and interpretation

- ✓ The quadratic coefficient is -2200 ~ -1700
  → Nonlinear model is reasonable
- ✓ Hawaii (the amount of data is not big)
  → The trend is **upward**, but the curve is **downward curvature**
  → Shrinkage from the group
- ✓ California
  → a **narrow** spread at family size 2
  a **large** spread at family size 7
  → the most of the data for large
  family sizes have
  large standard errors

Hawaii



California



- Further extensions

	An example of family income and family size	Extensions
Trend	linear quadratic	higher-order polynomial Sinusoidal exponential
Noise distribution	a single lying noise for all individuals	vary among individuals
Distribution for parameters	normal distribution	t distribution
Considering Covariation	No	Use a multivariate normal prior on the intercept and slope

Posterior predictive check

✓ Visualize the data and the posterior predictions

→ If the prediction doesn't seem to fit the data, change the model

→ New model should be both **meaningful** and **computationally tractable**

✓ Create a posterior predictive sampling distribution

→ measure of discrepancy between the predictions and the data

Ways to extend a model

✓ Add a parameter

ex.) $\mu_i = \beta_0 + \beta_1 x$ → $\mu_i = \beta_0 + \beta_1 x + \beta_2 x^2$

You can check the validity of the model by considering β_2

✓ Try a completely different model

→ Compare models by **Bayesian model comparison**

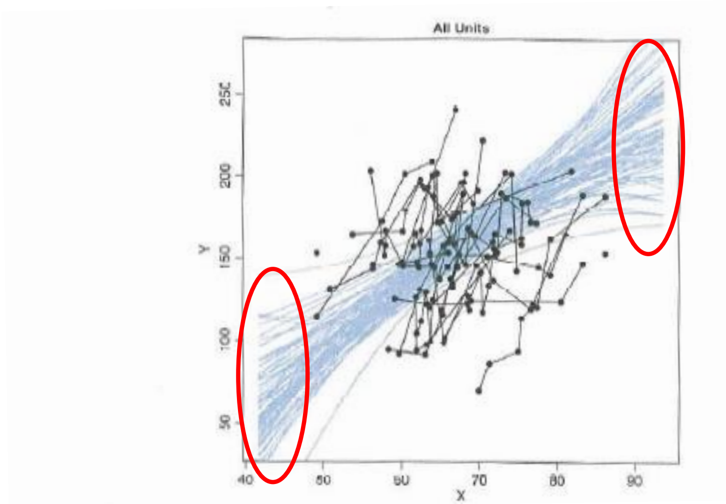
✓ "double dipping" : data are used to change **the prior distribution**

Steps to extend a JAGS or Stan model

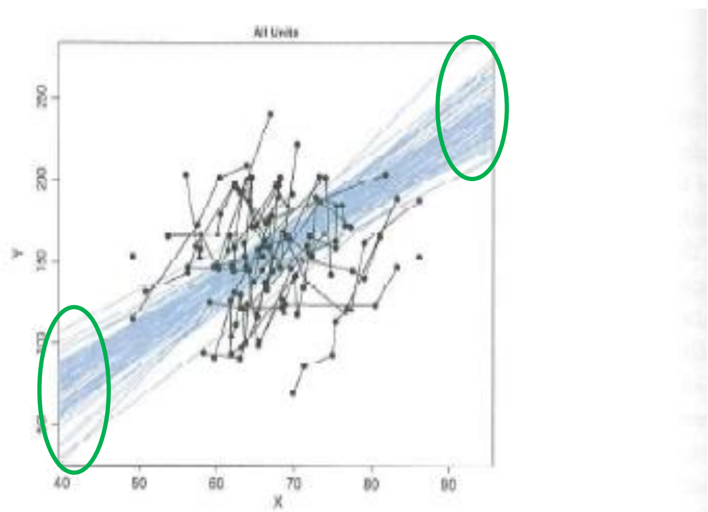
- ✓ Carefully specify the model with its new parameters
  - → Draw a diagram
- ✓ Be sure all the new parameters have sensible priors
- ✓ Define initial values for all the new parameters
  - → You can let JAGS initialize parameters automatically
- ✓ Tell JAGS to track the new parameters
  - → Stan automatically tracks
- ✓ Modify the summary and graphics output to properly display the extended model
  - → You should write **R code** because graphics are displayed by R

Perils of adding parameters

- ✓ Increase in uncertainty of a parameter estimate
  - → The curvature and slope trade-off strongly
  - → Even if curvature is 0, the certainty of slope decreases
- ✓ The one of the ways to solve the problem is standardizing the data



Quadratic trend



Linear trend