Bayesian Tools for EDA and Model Building: A Brainy Study

Steven N. MacEachern
Mario Peruggia

Department of Statistics

The Ohio State University

Columbus, OH 43210-1247

www.stat.ohio-state.edu/∼peruggia/papers/brainy.pdf
www.stat.ohio-state.edu/∼peruggia/papers/JSM00.pdf

snm@stat.ohio-state.edu
peruggia@stat.ohio-state.edu

Outline

- Strategy for Bayesian model building
- Numerical and graphical exploratory tools
- Application to an allometric study
- Conclusions and moral of the story

Background

- Jerison's theory:

  (brain mass) $= c$ (body mass)$^{2/3}$

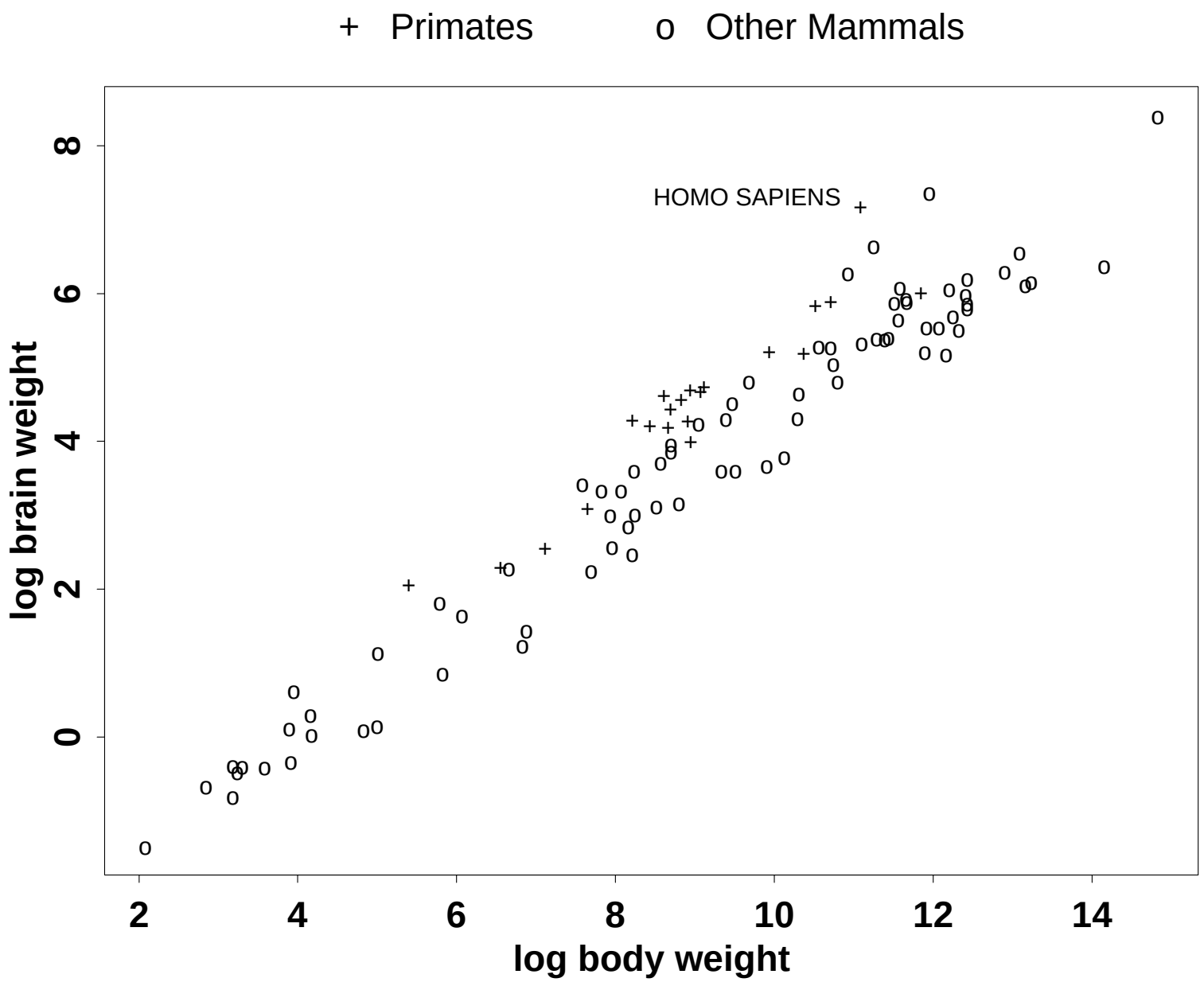
- S.J. Gould's conjecture:

  Brain mass proportional to body surface because body surfaces serve as end points for nerve channels

Data

- Brain weights and body weights for 100 species of placental mammals (Sacher and Staffeldt, 1974)

- Taxonomy of species based on order and sub-order


+ Primates
o Other Mammals
HOMO SAPIENS
log brain weight
log body weight
0
2
4
6
8
2
4
6
8
10
12
14

Bayesian SLR Model

- log(brain weight) =
  $\alpha + \beta$ log(body weight) + error

- Diffuse normal prior on α

- Normal prior centered at 2/3 on β, giving tiny mass to negative values

- Diffuse IG prior on error variance

- $\widehat{\beta} = .72$

- 95% post. prob. interval for β is $(.68, .76)$, casting doubt on Jerison's theory

- Define **case-deletion weights** as case-deleted posterior over full posterior evaluated at Gibbs draws from full posterior

- Sample correlation btw sets of log-weights corresponding to two different cases, y_j and y_k, measures **synergistic** impact of the two cases on the posterior

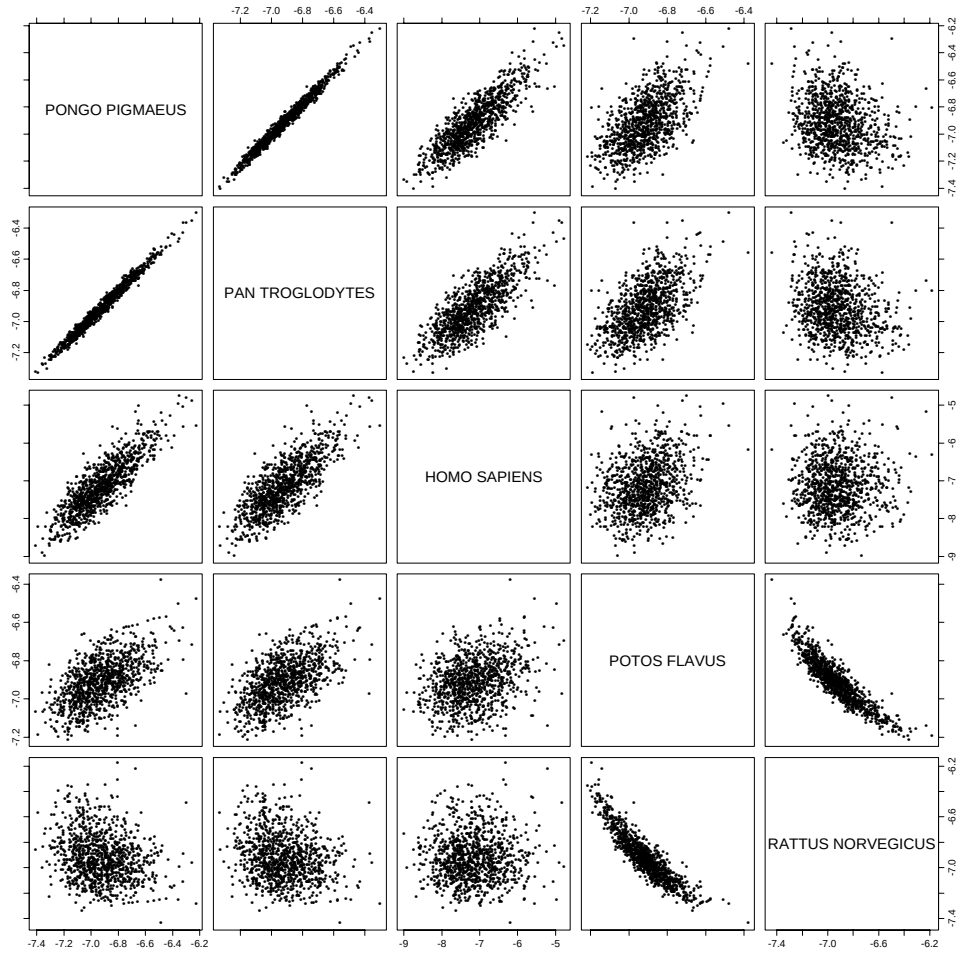
- **Bayesian EDA Strategy:**

  Examine scatterplots of log-weights

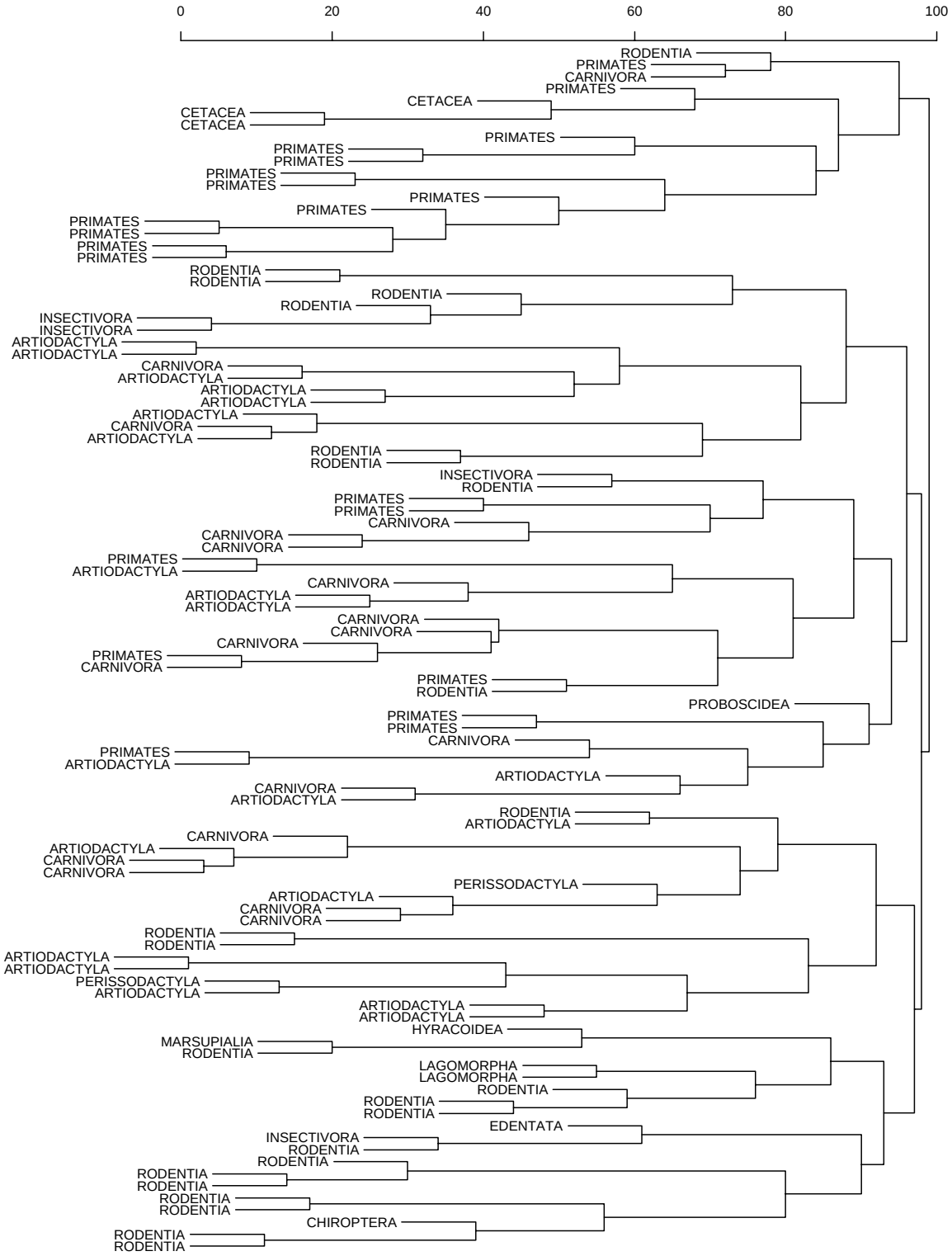
  Cluster observations based on synergy:

  dist(y_j,y_k) = 1 - corr. btw sets of log-weights

Matrix Scatterplot of Selected Sets of Case-Deletion Log-Weights

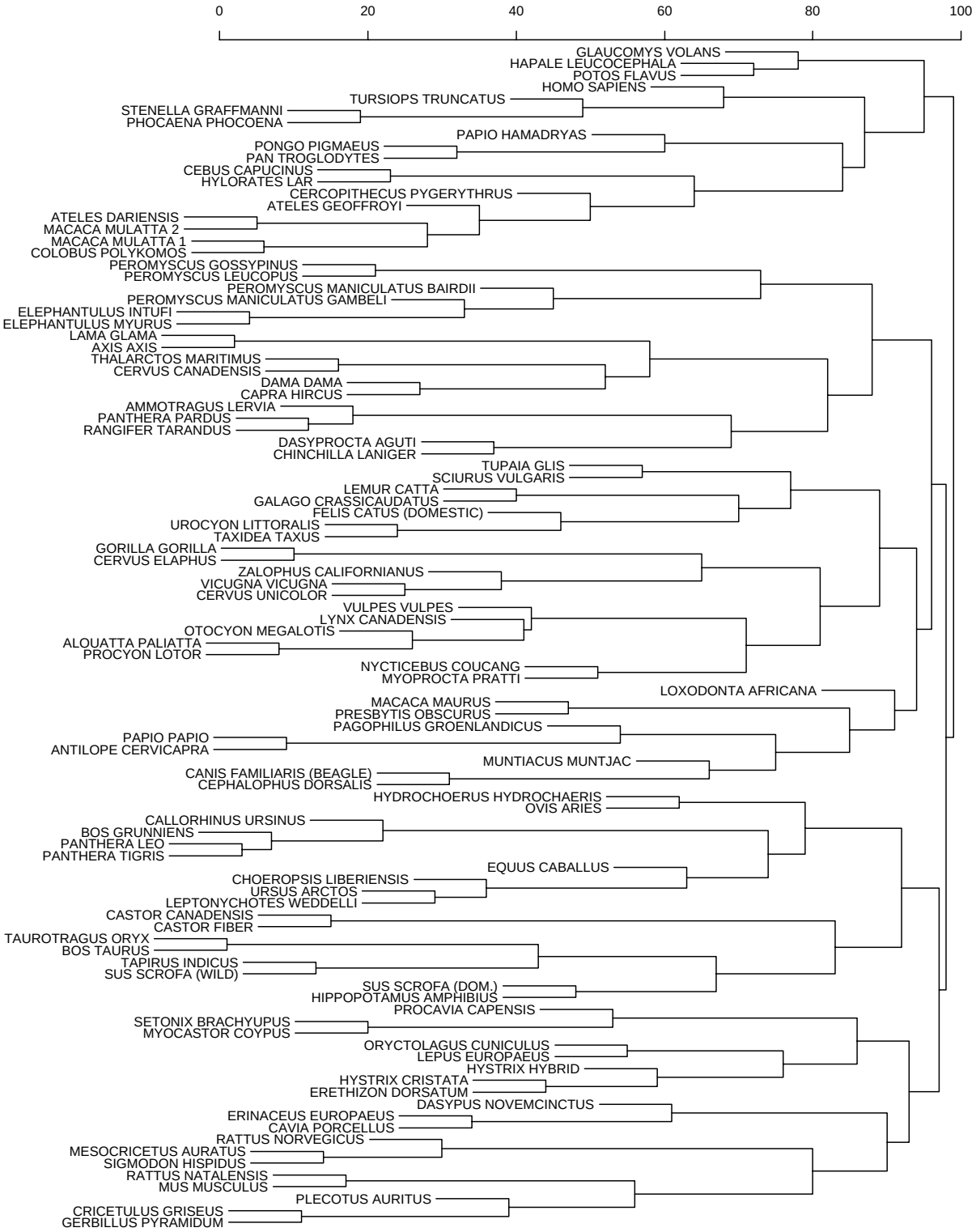


Labeled by Order



Simple Linear Regression

are Primates



Simple Linear Regression

Bayesian VC Model

- EDA shows clustering based on order and sub-order: SLR model is inadequate

- Alternative variance components model:

  log(brain weight) =
  $\alpha + \beta$ log(body weight) +
  γ(order) + δ(sub-order) + error

- Priors on α and β as before

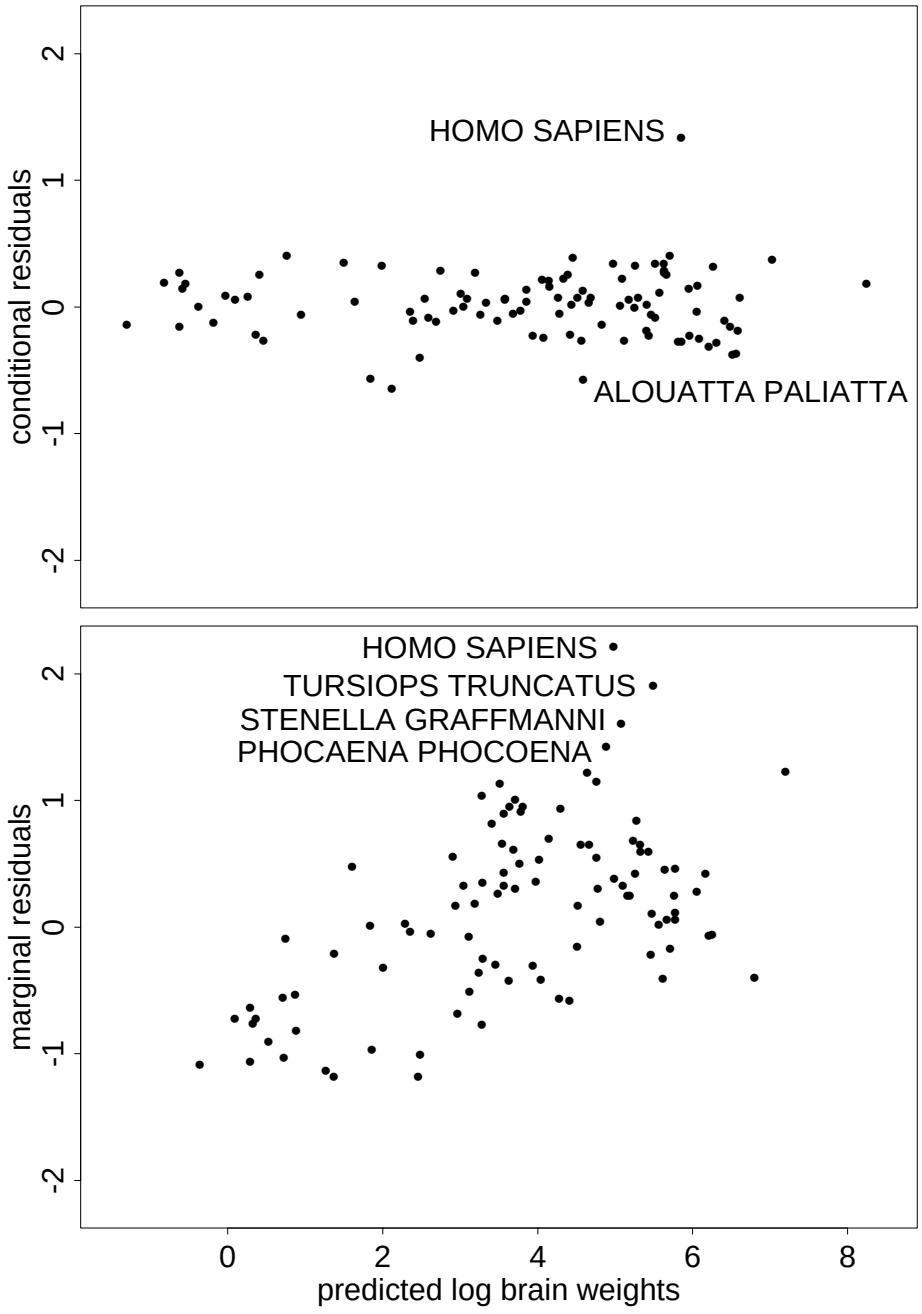
- Normal priors centered at zero on γ's and δ's

- Diffuse IG priors on variance components and error variance

- $\widehat{\beta} = .59$

- 95% post. prob. interval for β is $(.55, .64)$

Variance Components Regression



Outliers

- Two types of residuals in VC model:

  conditional $= y_i - \widehat{y}_i$
  (adjusts for order and sub-order)

  marginal $= y_i - \widehat{\alpha} - \widehat{\beta} x_i$
  (does not adjust for order and sub-order)

- Two questions:

  a. Do we have a big brain for a primate?

  b. Do we have a big brain for a mammal?

- Two answers:

  Conditional residuals say "yes" to a.

  Marginal residuals say "maybe (not)" to b.

- Residuals from SLR model are irrelevant (the model is bad)

Conclusions

- Jerison's theory is tenable because **neither** model yields an interval estimate of the slope that includes 2/3

- Interval from SLR model lies **above** 2/3

- Interval from VC model lies **below** 2/3

- Covariance matrix of VC model only approximates dependencies among species

- More refined covariance structure based on evolutionary tree is of interest

Moral of the Story

Use **DAC** to build models!