21. Dichotomous Predicted Variable

Ben Blayney

14/05/2020

21.1. Multiple Metric Predictors

- Looked at this in earlier chapters but now have the framework of the generalized linear model.
- Traditional treatment of these sorts of data structure is “logistic regression”
- Easy in Bayesian to generalize so:
    - robust to outliers
    - allow different variances within levels of a nominal predictor
    - have hierarchical structure to share information across levels or factors as appropriate
- Inverse-link function is logistic
- Bernoulli distribution for describing noise in the data

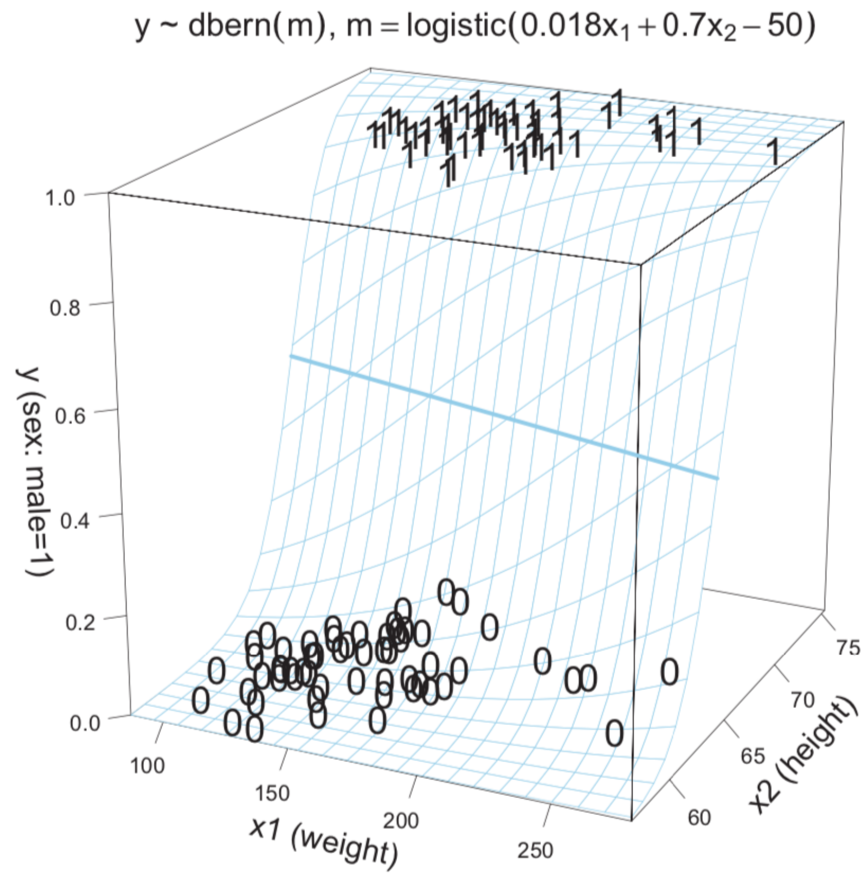

Figure 21.1: Data show gender (arbitrarily coded as male=1, female=0) as a function of weight (in pounds) and height (in inches). All 0's are located on the bottom plane of the cube, and all 1's are located on the top plane of the cube. Logistic surface shows maximum-likelihood estimate. Heavy line shows 50% threshold.

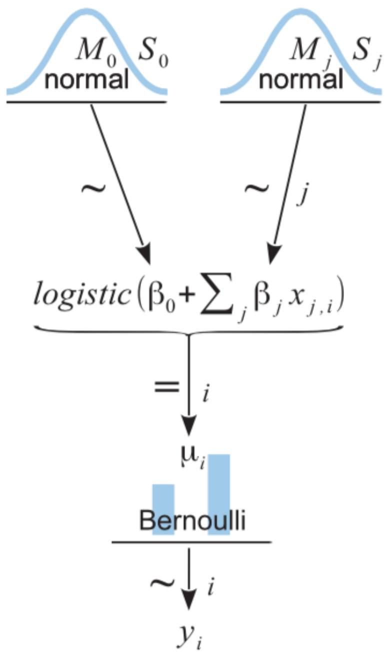

Figure 21.2: Dependency diagram for multiple logistic regression. Compare with the diagram for robust multiple linear regression in Figure 18.4 (p. 498). Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

21.1.1. The model and implementation in JAGS

$$\mu = logistic(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$
$$y \sim Bernoulli(\mu)$$
$$logistic(x) = \frac{1}{1 + e^{-x}}$$

Preliminary details:

- We will try to reduce autocorrelation in the MCMC chains by standardizing the data; helps efficiency of MCMC
- y values must be 0s and 1s and therefore not standardized.
- In JAGS logistic function is called ilogit (inverse logit)

21.1.2. Example: height, weight, and gender

Weight coefficient less important once height included. Most of the predictive work done by height, weight and height are correlated, so independent predictiveness of weight is small once height included.

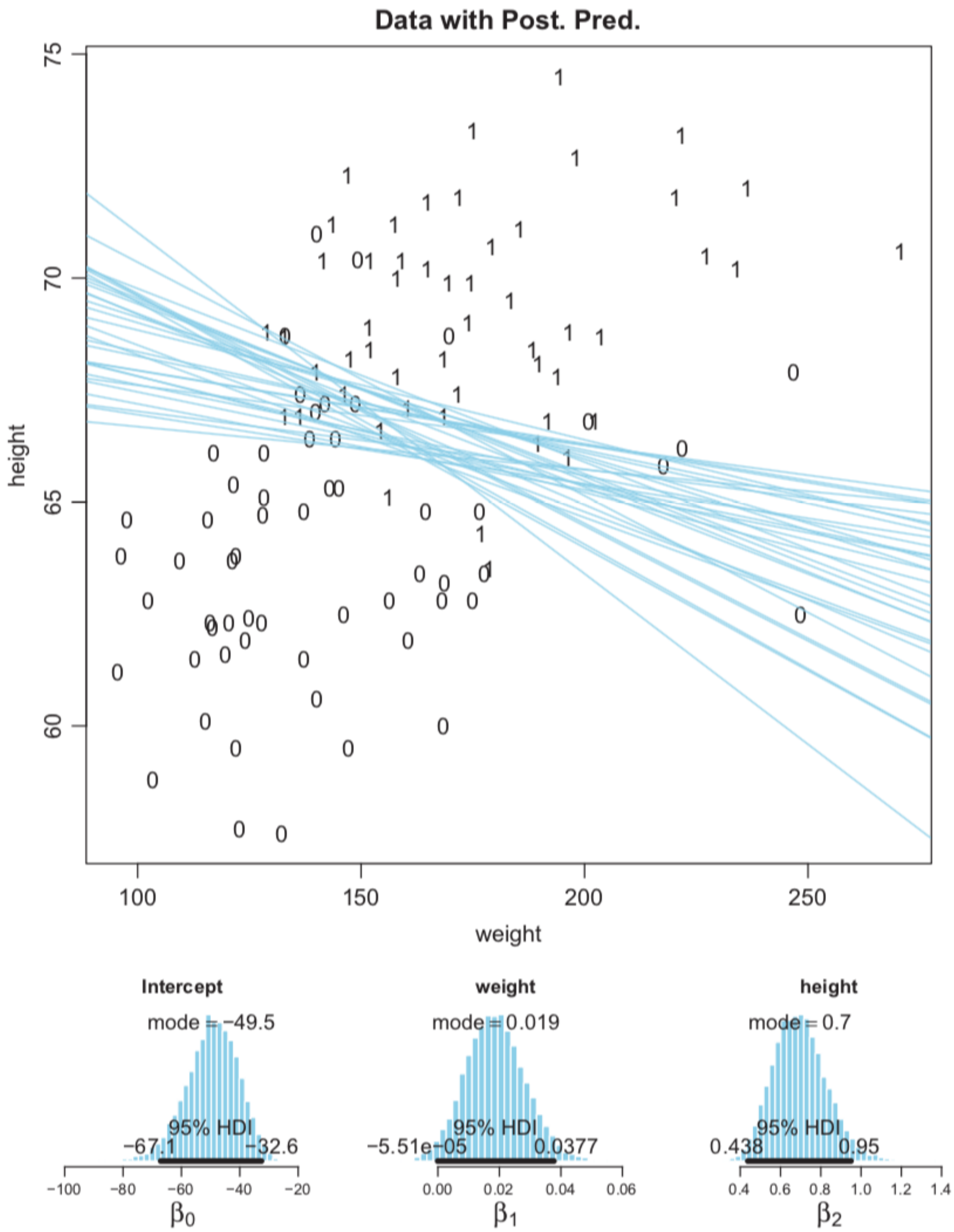

Figure 21.4: Predicting gender (arbitrarily coded as male=1, female=0) as a function of weight (in pounds) and height (in inches), using logistic regression. Upper panel: Data are indicated by 0's and 1's. Lines show a random sample from the MCMC posterior of thesholds at which the probability of male is 50%. Lower panels: Marginal posterior distribution on baseline and slopes. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

21.2. Interpreting the Regression Coefficients

21.2.1. Log odds

- $logit(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ implies that when x_1 goes up by 1 unit then $logit(\mu)$ goes up by amount β_1. The regression coefficients are telling us about increases in $logit(\mu)$.
- $logit(\mu) = log(\mu/(1-\mu))$
- μ is the probability that $\gamma = 1$.
- So: $logit(\mu) = logit(p(\gamma = 1)) = log(p(\gamma = 1)/p(\gamma = 0))$ which is the log odds of outcome 1 to outcome 2.
- A log odds of 0 indicates a 50% chance. A coefficient value then indicates how much the log odds increases if the variable increases by 1 unit. So you get most change in the probability when the probability is already close to 50%.
- A regression coefficient of 1.0 corresponds to a rate of probability change of about 24.4 percentage points per x-unit at the threshold x value.

21.2.2. When there are few 1s or 0s in the data

Makes coefficient estimates more uncertain.

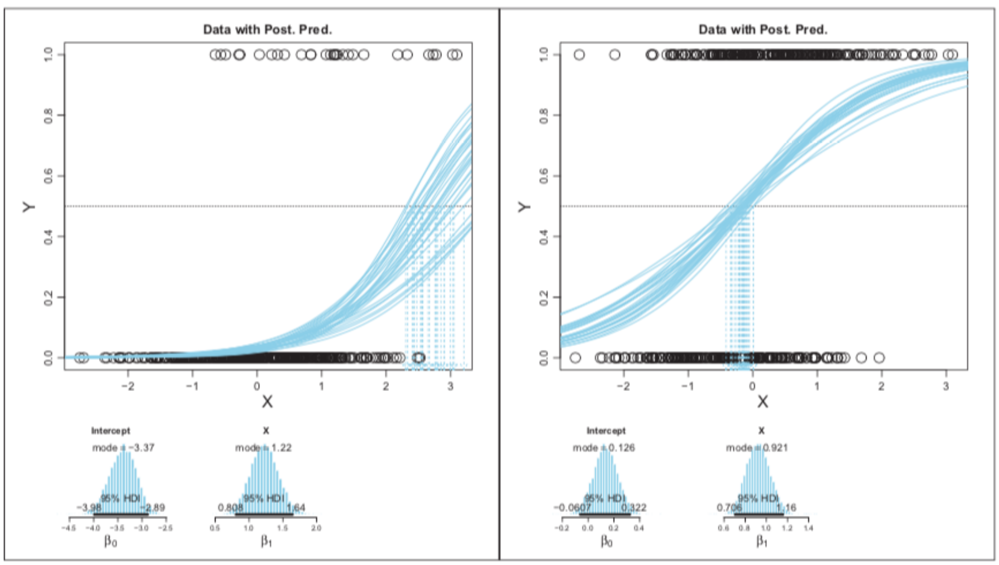

Figure 21.5: Parameter estimates are more uncertain when there are few 0's or 1's in the data. The main panels have data with the same x values, and y values randomly generated by logistic functions with the same slope ($\beta_1 = 1$) but different intercepts ($\beta_0 = -3$ and $\beta_0 = 0$). Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

21.2.3. Correlated predictors

- Get very ambiguous threshold lines (see Figure 21.6), and anticorrelation of credible β values.
- As with linear regression, very important to consider the correlations of predictors when interpreting the parameters of logistic regression.

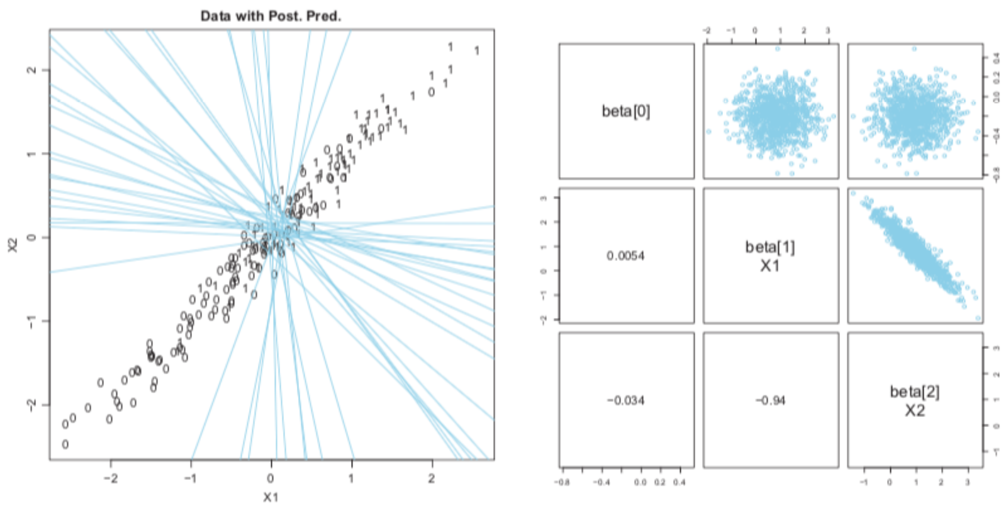

Figure 21.6: Estimates of slope parameters trade off when the predictors are correlated. Left panel shows credible 50% level contours superimposed on data. Right panel shows strong anti-correlation of credible β_1 and β_2 values. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

21.2.4. Interaction of metric predictors

- Might be the case that additive combinations of predictors not very accurate.
- See figure 21.7
- When no interaction the 50% level contour is always straight; with an interaction it can be curved.
- **Remember that when using an interaction term the regression coefficients on individual predictors only indicate the slope when all other predictors are set to 0 (analogously to linear regression).**

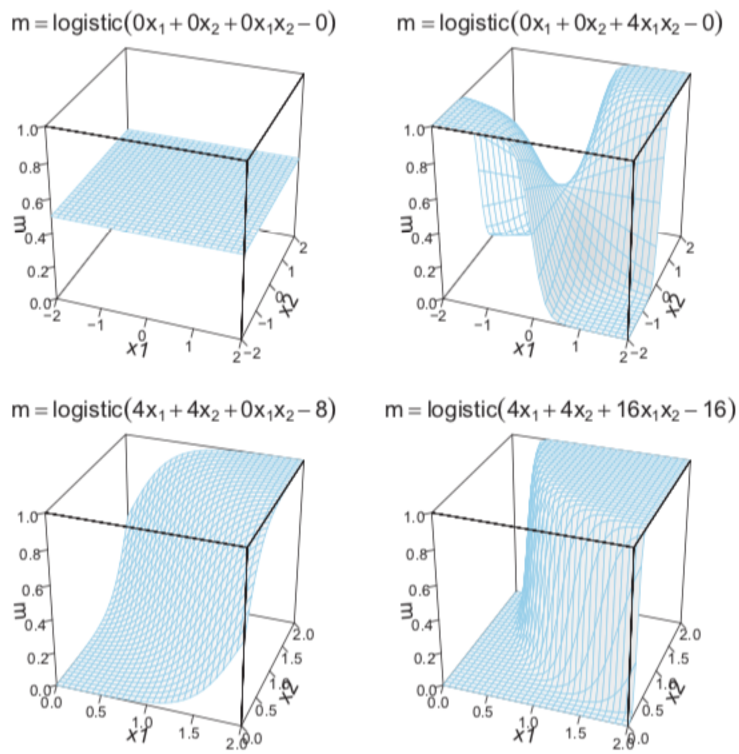

Figure 21.7: Multiplicative interaction of metric predictors in logistic regression. Left column shows examples of no interaction. Right column shows corresponding logistic surfaces with interaction. Title of each plot shows the coefficients on the predictors. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

21.3. Robust Logistic Regression

- Outliers relative to the logistic function can only be accomodated by reducing the magnitude of the estimated slope.
- We can describe the data as being a mixture of two different sources:
  - the logistic function of the predictors
  - randomness
- We suppose that every data point has a small chance α of being generated by the guessing process but usually (with probability $1-\alpha$) the γ value comes from the logistic function of the predictor.
- The predicted probability that $\gamma = 1$ is $\mu = \alpha\frac{1}{2} + (1-\alpha)logistic(\beta_0 + \sum(\beta_j x_j))$
  - When guessing coefficient is 0 then get conventional logistic model. When guessing coefficient is 1 then get completely random data values.
- We need to estimate α as well as the logistic parameters, so lets also have a prior on α.
- This ends up looking like curves that asymptote at 0.1 and 0.9 instead of 0 and 1 for example.
- There are other ways to model outliers (see chapter 23 exercise 23.2)

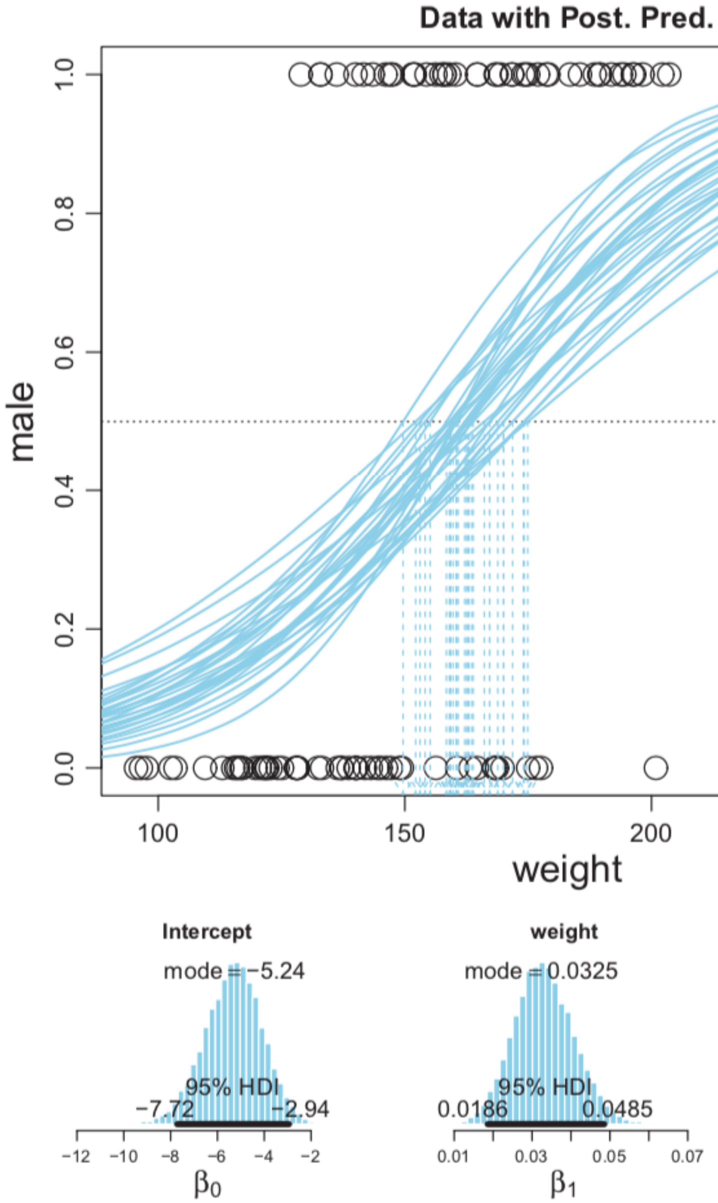

Figure 21.3: Predicting gender (arbitrarily coded as male=1, female=0) as a function of weight (in pounds), using logistic regression. Upper panel: Data are indicated by dots. Logistic curves are a random sample from the MCMC posterior. Descending arrows point to threshold weights at which the probability of male is 50%. Lower panels: Marginal posterior distribution on baseline and slope. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

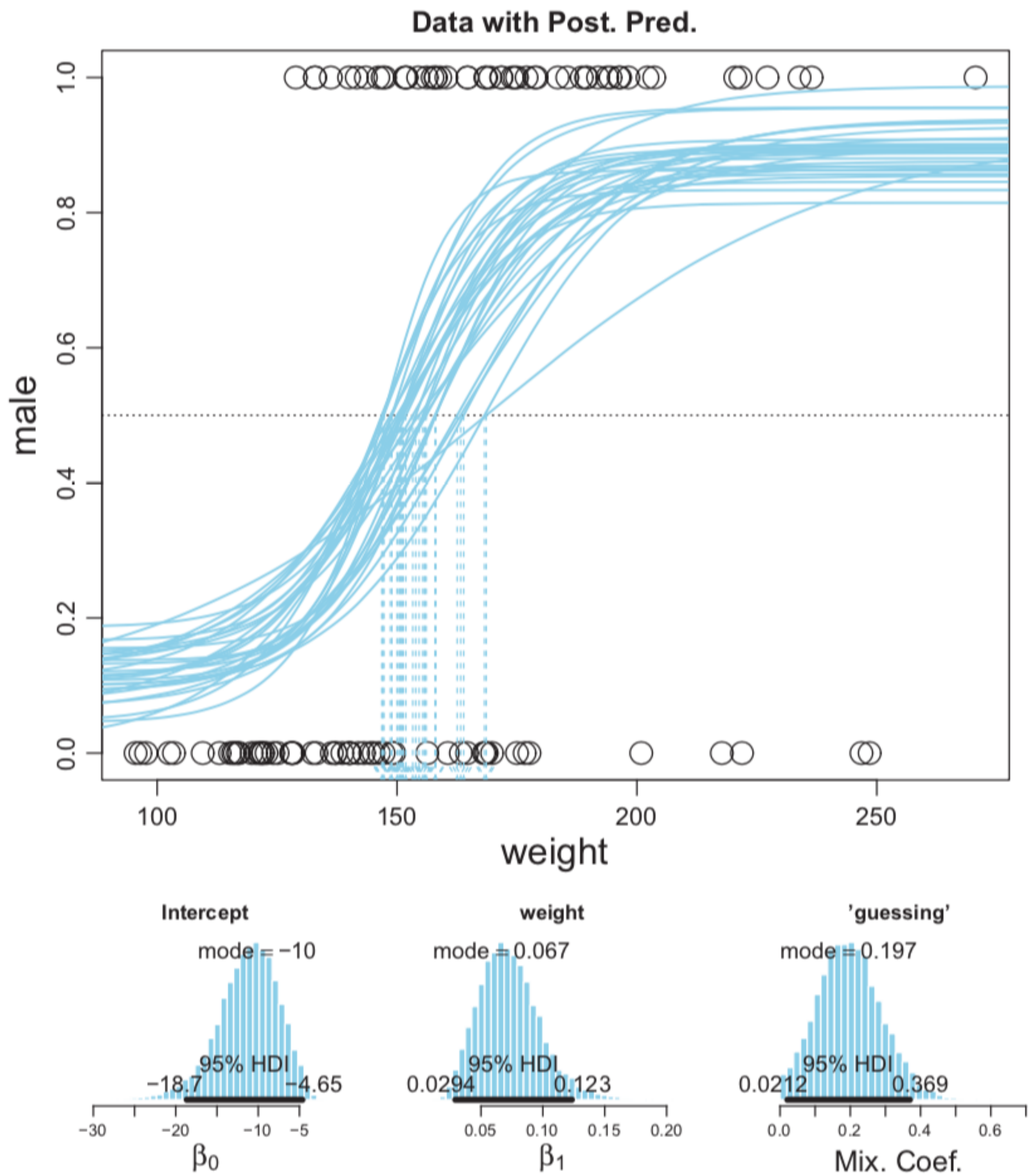

Figure 21.8: Predicting gender (arbitrarily coded as male=1, female=0) as a function of weight (in pounds), using *robust* logistic regression. Upper panel: Data are indicated by dots, the same as in Figure 21.3. Curves are a random sample from the MCMC posterior; notice asymptotes away from 0,1 limits. Descending arrows point to threshold weights at which the probability of male is 50%. Lower panels: Marginal posterior distribution on baseline, slope, and guessing coefficient. Pairwise plots are shown in Figure 21.9. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

21.4. Nominal Predictors

21.4.1. Single group

$y \sim Bernoulli(\mu)$, $\mu = logistic(\beta_0)$, with prior $\beta_0 \sim normal(M_0, S_0)$; example is working out the bias of a coin by observing flips.

Non-intuitive mapping of how this prior looks in terms of a distribution on μ

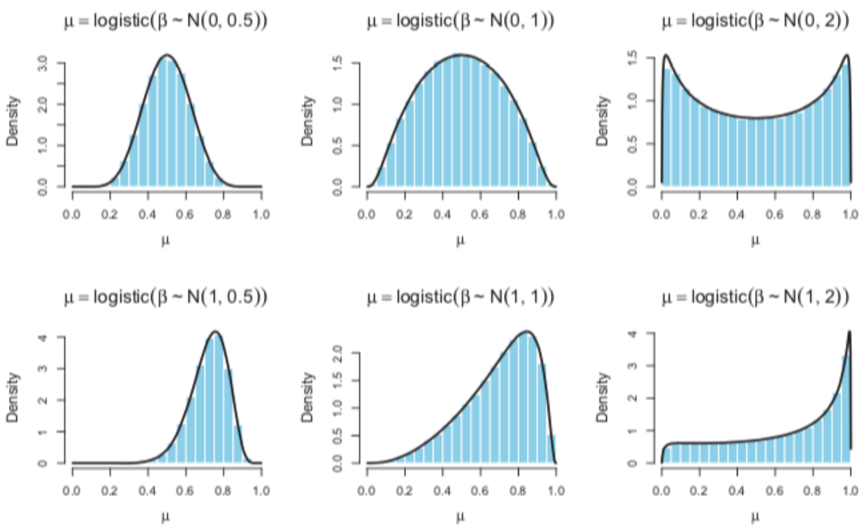

Figure 21.11: Prior distributions on μ for different choices of M_0 and S_0 in Figure 21.10. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

21.4.2 Multiple groups

- Baseball example

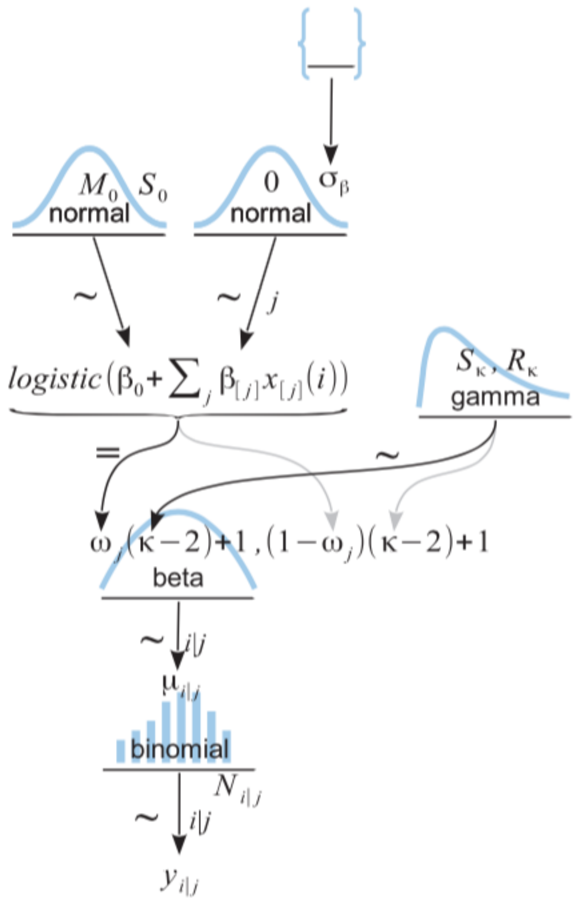

Figure 21.12: Hierarchical diagram for logistic ANOVA-like model. The top part of the structure is based on the ANOVA-like model of Figure 19.2 (p. 529). The lower part of the structure is based on the models of Figure 9.7 (p. 229) and Figure 9.13 (p. 235). Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

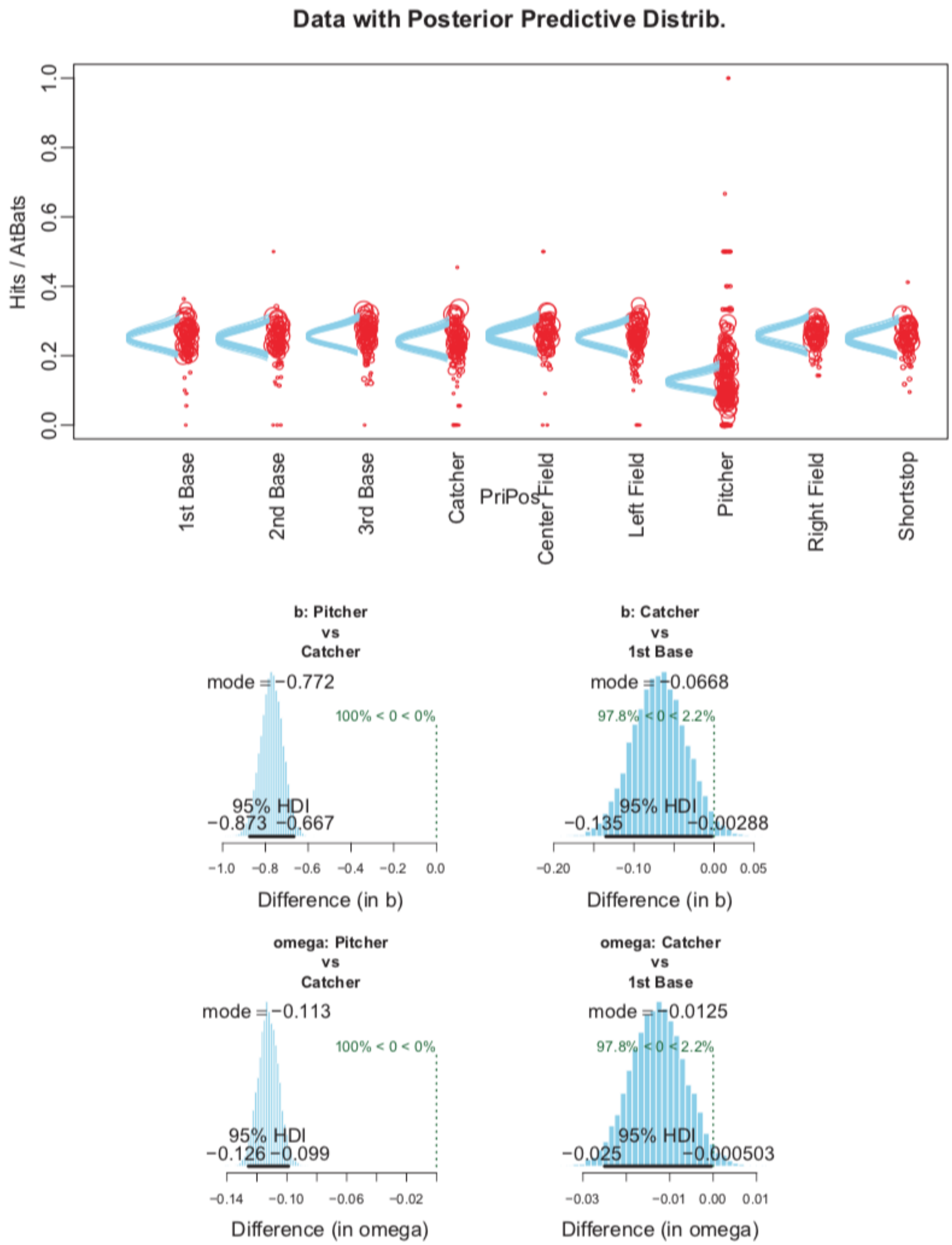

Figure 21.13: Baseball batting data are shown in the upper panel with dot size proportional to number of at-bats. The posterior predictive distributions are credible beta distributions assuming homogeneous concentration across positions. Lower panels show selected contrasts, which can be compared with the contrasts shown in Figure 9.14 (p. 236). Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

21.5 Exercises