Exercises 4 · SSC 383D
Exercises 5: Hierarchical Linear Models

1 Price elasticity of demand

The data in "cheese.csv" are about sales volume, price, and advertisting display activity for packages of Borden sliced "cheese." The data are taken from Rossi, Allenby, and McCulloch's textbook on *Bayesian Statistics and Marketing.* For each of 88 stores (store) in different US cities, we have repeated observations of the weekly sales volume (vol, in terms of packages sold), unit price (price), and whether the product was advertised with an in-store display during that week (disp = 1 for display).

Your goal is to estimate, on a store-by-store basis, the effect of display ads on the demand curve for cheese. A standard form of a demand curve in economics is of the form $Q = \alpha P^{\beta}$, where Q is quantity demanded (i.e. sales volume), P is price, and α and β are parameters to be estimated. You'll notice that this is linear on a log-log scale,

$$\log Q = \log \alpha + \beta \log P$$

which you should feel free to assume here. Economists would refer to β as the price elasticity of demand (PED). Notice that on a log-log scale, the errors enter multiplicatively.

There are several things for you to consider in analyzing this data set.

- The demand curve might shift (different α) and also change shape (different β) depending on whether there is a display ad or not in the store.
- Different stores will have very different typical volumes, and your model should account for this.
- Do different stores have different PEDs? If so, do you really want to estimate a separate, unrelated β for each store?
- If there is an effect on the demand curve due to showing a display ad, does this effect differ store by store, or does it look relatively stable across stores?
- Once you build the best model you can using the log-log specification, do see you any evidence of major model mis-fit?

Propose an appropriate hierarchical model that allows you to address these issues, and use Gibbs sampling to fit your model.

$\star$ Denote we have data for $i = 1, \ldots, n$ stores and for each store we have observations $j = 1, \ldots, N_i$. The quantity demand is denoted as $Q_{i,j}$ and the price is $P_{i.j}$, and $disp_{i,j}$ is the

indicator that product is displayed or not. I use the model as follows,

$$\log Q_{i,j} = \beta_{0,i} + \beta_{1,i} \log P_{i,j} + \beta_{2,i} 1(disp_{i,j} = 1) + \beta_{3,i} \log P_{i,j} \cdot 1(disp_{i,j} = 1).$$

This can be rewritten as a linear regression,

$$y_{ij} = x_{ij}^T \beta_i + \epsilon_{ij},$$

where $y_{ij} = \log Q_{i,j}$, $x_{ij} = (1, \log P_{i,j}, 1(disp_{i,j} = 1), \log P_{i,j} \cdot 1(disp_{i,j} = 1))^T$, and $\beta_i = (\beta_{0,i}, \ldots, \beta_{3,i})^T$. I propose a hierarchical model as follows,

$$\begin{aligned}
y_{ij} \mid \beta_i, \sigma_i^2 &\sim N(x_{ij}^T \beta_i, \sigma_i^2) \\
\beta_i \mid \beta, \Sigma &\sim N(\beta, \Sigma = diag(s_1^2, \ldots, s_4^2)) \\
\sigma_i^2 &\propto \text{Inv-Ga}(\frac{a}{2}, \frac{b}{2}) \\
a &\sim U(1, 10) \\
b &\sim U(1, 10) \\
\beta &\propto 1 \\
s_j^2 &\sim \text{Inv-Ga}(\frac{a_0}{2}, \frac{b_0}{2}) = \text{Inv-Ga}(\frac{1}{2}, \frac{1}{2}) \quad j = 1, \ldots, 4
\end{aligned}$$

Denote $X_i = \begin{bmatrix} x_{i1}^T \\ \vdots \\ x_{iN_i}^T \end{bmatrix}$, then the sampling model can also be written as

$$Y_i \mid \beta_i, \sigma_i^2 \sim N_{N_i}(X_i \beta_i, \sigma_i^2 I).$$

We can use the Gibbs sampler to obtain posterior samples based on the full conditionals as follows.

The joint posterior:

$$\begin{aligned}
p(Y_{1:n}, \beta_{1:n}, \beta, \Sigma, \sigma_{1:n}^2) \propto \prod_{i=1}^{n} \{ &(\sigma_i^2)^{-\frac{N_i}{2}} \exp\{-\frac{1}{2\sigma_i^2}(Y_i - X_i\beta_i)^T(Y_i - X_i\beta_i)\} \\
&\cdot |\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\beta_i - \beta)^T \Sigma^{-1} (\beta_i - \beta)\} \cdot (\sigma_i^2)^{-\frac{a}{2}-1} \exp[-\frac{b}{2\sigma_i^2}]\} \\
&\cdot \prod_{j=1}^{4} (s_j^2)^{-\frac{1}{2}-1} \exp[-\frac{1}{2s_j^2}] 1(a \in (1, 10)) 1(b \in (1, 10))
\end{aligned}$$

The full conditionals:

$$p(\beta_i \mid -) = N(\tilde{\mu}_i, \tilde{\Sigma}_i),$$

where

$$\tilde{\Sigma}_i = \left(\frac{X_i^T X_i}{\sigma_i^2} + \Sigma^{-1} \right)^{-1} \text{ and } \tilde{\mu}_i = \tilde{\Sigma}_i \left(\frac{X_i^T Y_i}{\sigma_i^2} + \Sigma^{-1} \beta \right).$$

$$p(\beta \mid -) = N(\bar{\beta}, \frac{1}{n}\Sigma),$$

where

$$\bar{\beta} = \frac{1}{n}\sum_{i=1}^{n}\beta_i.$$

$$p(\sigma_i^2 \mid -) = \text{Inv-Ga}(\frac{a}{2} + \frac{N_i}{2}, \frac{b}{2} + \frac{1}{2}(Y_i - X_i\beta_i)^T(Y_i - X_i\beta_i)).$$

$$p(s_j^2 \mid -) = \text{Inv-Ga}(\frac{1}{2} + \frac{n}{2}, \frac{1}{2} + \frac{1}{2}\sum_{i=1}^{n}(\beta_{ij} - \beta_j)^2).$$

$$p(a \mid -) \propto \prod_{i=1}^{n}(\sigma_i^2)^{-\frac{a}{2}-1}1(a \in (0.5, 10)).$$

$$p(b \mid -) \propto \exp[-\sum_{i=1}^{n}\frac{b}{2\sigma_i^2}]1(b \in (0.5, 10)).$$

After implementing the MCMC, I got the posterior mean estimates of the parameters, and here the β_i for each stores are shown in Figure 3.

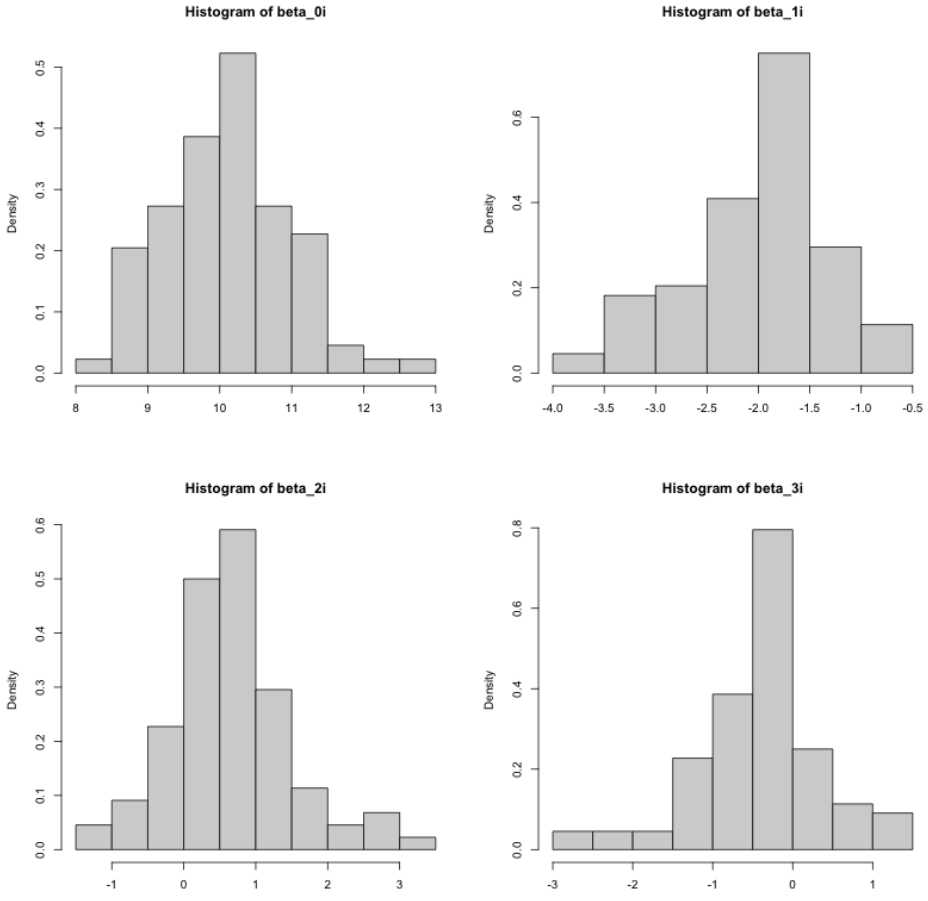


Figure 1: Histogram of the posterior mean estimates of the $beta_i$ over the stores.

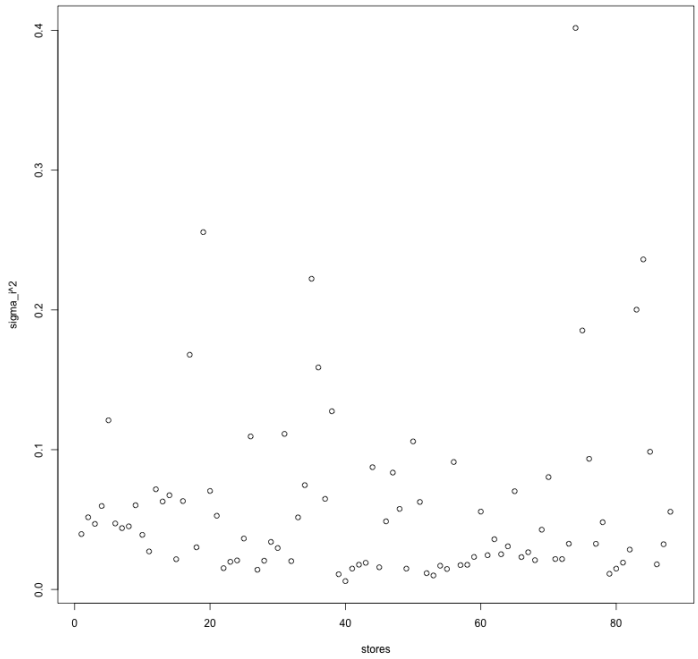


Figure 2: Scatter plot of σ_i^2.

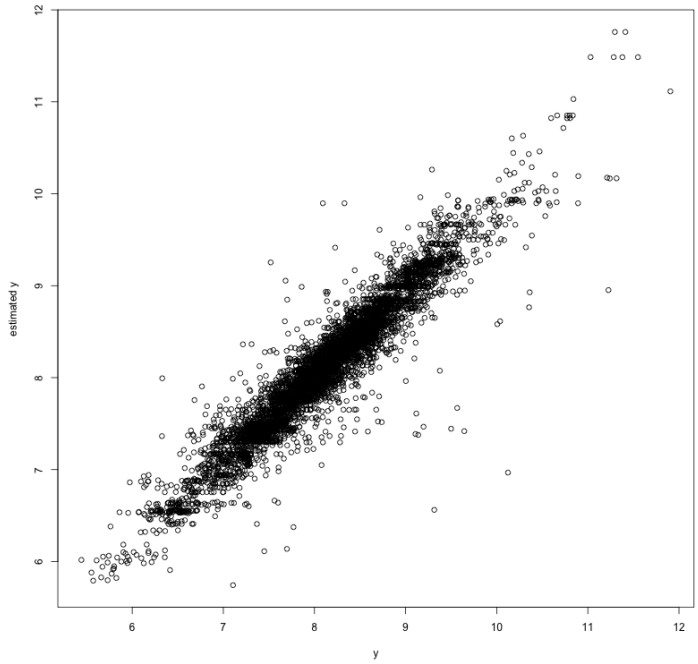


Figure 3: True y against estimated y.

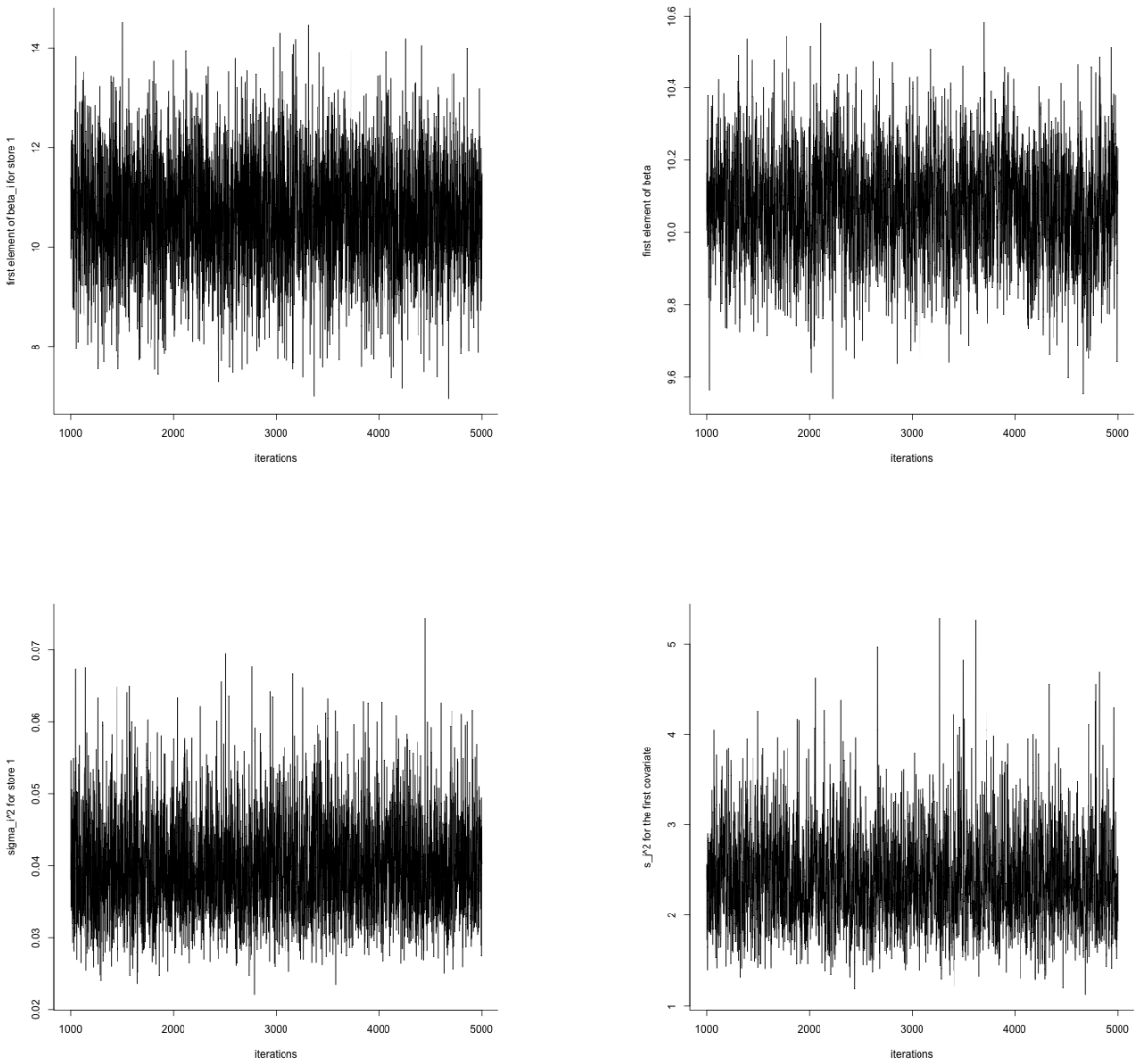


Figure 4: Trace plots.

2 A hierarchical probit model

Read the following paper (or read some distillation of the paper in a book/blog post/slides/etc).

> "Bayesian Analysis of Binary and Polychotomous Response Data." James H. Albert and Siddhartha Chib. *Journal of the American Statistical Association*, Vol. 88, No. 422 (Jun., 1993), pp. 669-679

The paper describes a Bayesian treatment of probit regression (similar to logistic regression) using the trick of *data augmentation*—that is, introducing "latent variables" that turn a hard problem into a much easier one. Briefly summarize your understanding of the key trick proposed by this paper. Then see you if you can apply the trick in the following context, which is more complex than ordinary probit regression.

In "polls.csv" you will find the results of several political polls from the 1988 U.S. presidential election. The outcome of interest is whether someone plans to vote for George Bush (senior, not junior). There are several potentially relevant demographic predictors here, including the respondent's state of residence. The goal is to understand how these relate to the probability that someone will support Bush in the election. You can imagine this information would help a great deal in poll re-weighting and aggregation (ala Nate Silver).

Use Gibbs sampling, together with the Albert and Chib trick, to fit a hierarchical probit model of the following form:

$$\begin{aligned} \Pr(y_{ij} = 1) &= \Phi(z_{ij}) \\ z_{ij} &= \mu_i + x_{ij}^T \beta_i \,. \end{aligned}$$

Here y_{ij} is the response (Bush=1, other=0) for respondent j in state i; $\Phi(\cdot)$ is the probit link function, i.e. the CDF of the standard normal distribution; μ_i is a state-level intercept term; x_{ij} is a vector of respondent-level demographic predictors; and β_i is a vector of regression coefficients for state i.

Notes:

1. There are severe imbalances among the states in terms of numbers of survey respondents. Following the last problem, the key is to impose a hierarchical prior on the state-level parameters.

2. The data-augmentation trick from the Albert and Chib paper above is explained in many standard references on Bayesian analysis. If you want to get a quick introduction to the idea, you can consult one of these. A good presentation is in Section 8.1.1 of "Bayesian analysis for the social sciences" by Simon Jackman, available as an ebook through lib.utexas.edu.

3. You are welcome to use the logit model instead of the probit model. If you do this, you'll need to read the following paper, rather than Albert and Chib: Polson, N.G., Scott, J.G. and Windle, J. (2013). Bayesian inference for logistic models using Polya-Gamma latent variables. J. Amer. Statist. Assoc. 108 1339–1349. You can find a routine for simulation Polya-Gamma random variables in the BayesLogit R package and the pypolyagamma python library.

$\star$ The basic idea of the reference paper is that for models like probit model, we can use data augmentation to use MCMC algorithim to sample from the posterior. Adding some hidden variable may make the inference easier.

The model I am using to solve this problem is as follows.

Likelihood:

$$\begin{aligned} \Pr(y_{ij}=1) &= \Phi(z_{ij}) \\ z_{ij} &= \mu_i + x_{ij}^T \beta_i \end{aligned}$$

rewrite this as (change notation for convenience)

$$z_{ij} = x_{ij}^T \beta_i$$

Priors:

$$\begin{aligned} \beta_i &\sim N(\mu, \Sigma = diag(\sigma_1^2, \ldots, \sigma_p^2)) \\ \sigma_k^2 &\sim \text{Inv-Ga}(\frac{1}{2}, \frac{1}{2}) \\ \mu &\propto 1 \end{aligned}$$

We can use the following data augmentation:

$$\begin{aligned} a_{ij} &\sim N(x_{ij}^T \beta_i, 1) \\ y_{ij} &= 1(a_{ij} > 0) \end{aligned}$$

The full conditionals for Gibbs sampling are:

$$p(\beta_i \mid -) \propto N(\tilde{\mu}_i, \tilde{\Sigma}_i),$$

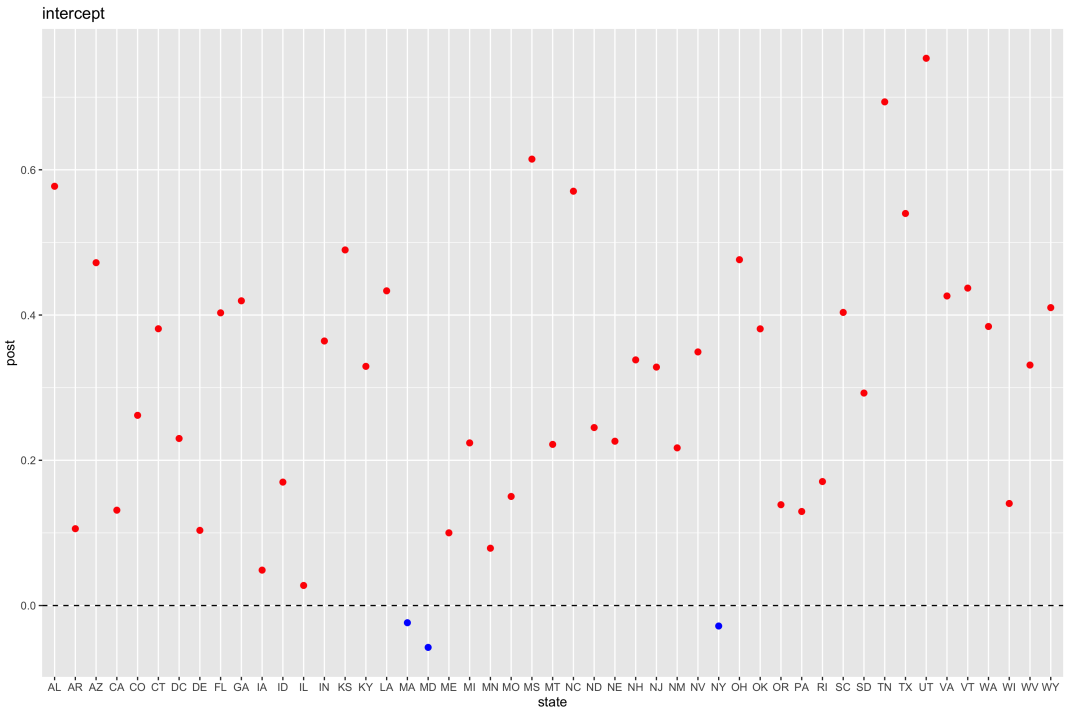
where

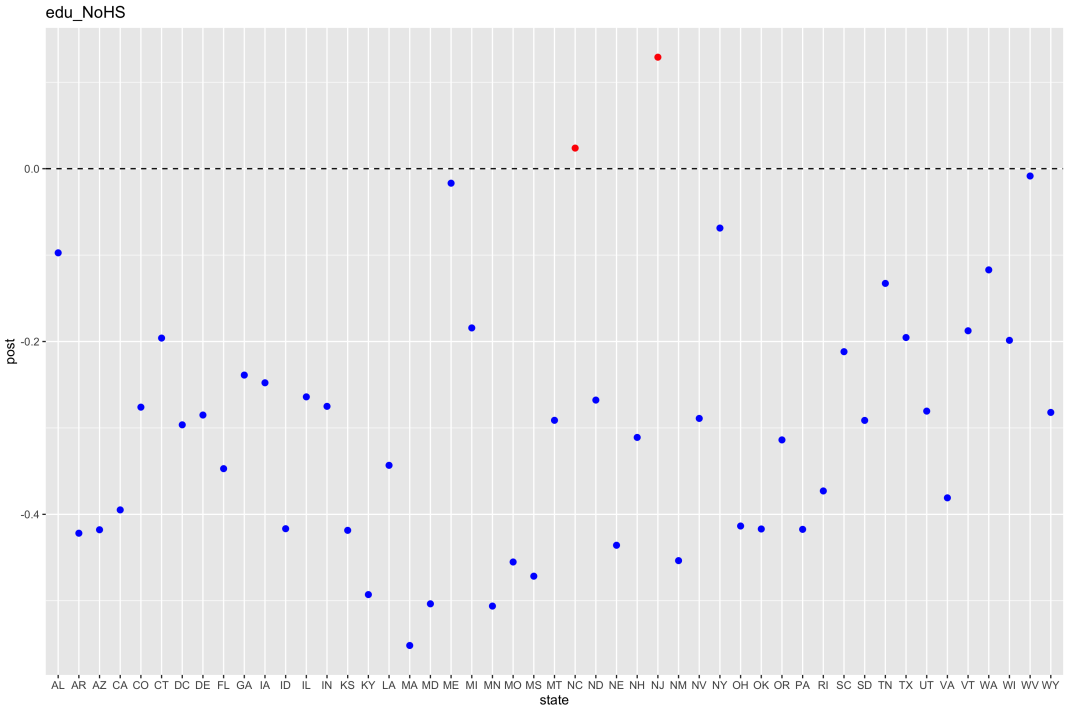
$$\tilde{\Sigma}_i = (X_i^T X_i + \Sigma^{-1})^{-1}, \;\; \tilde{\mu}_i = \tilde{\Sigma}_i (X_i^T a_i + \Sigma^{-1} \mu).$$

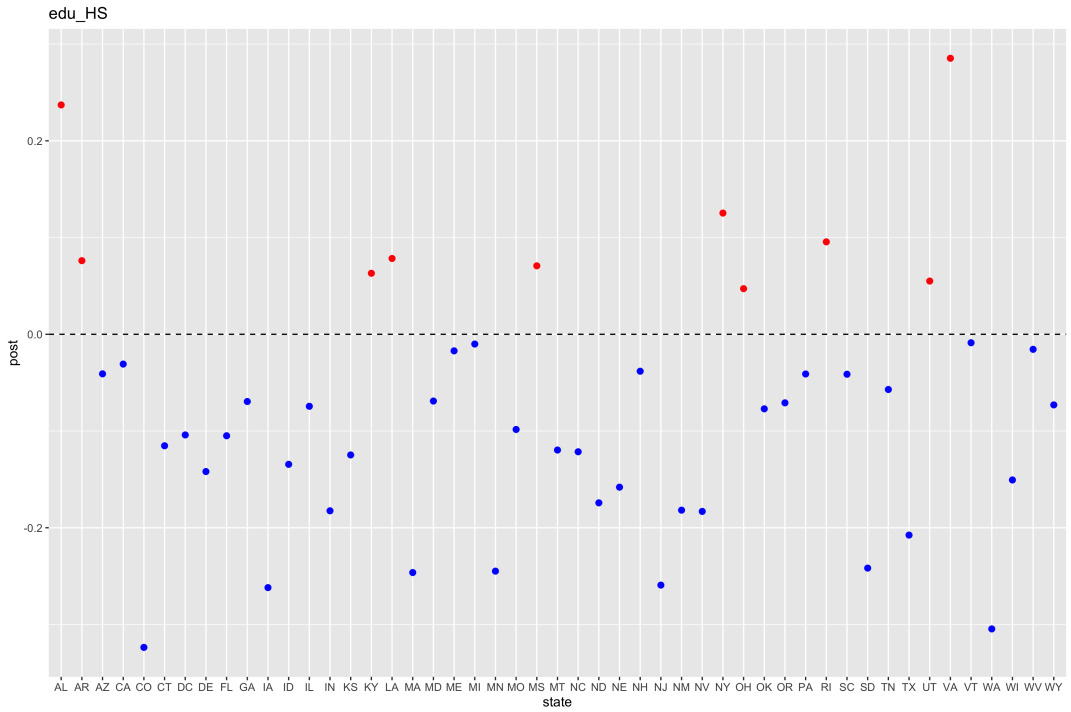
$$p(\mu \mid -) = N(\bar{\beta}, \frac{1}{n}\Sigma).$$

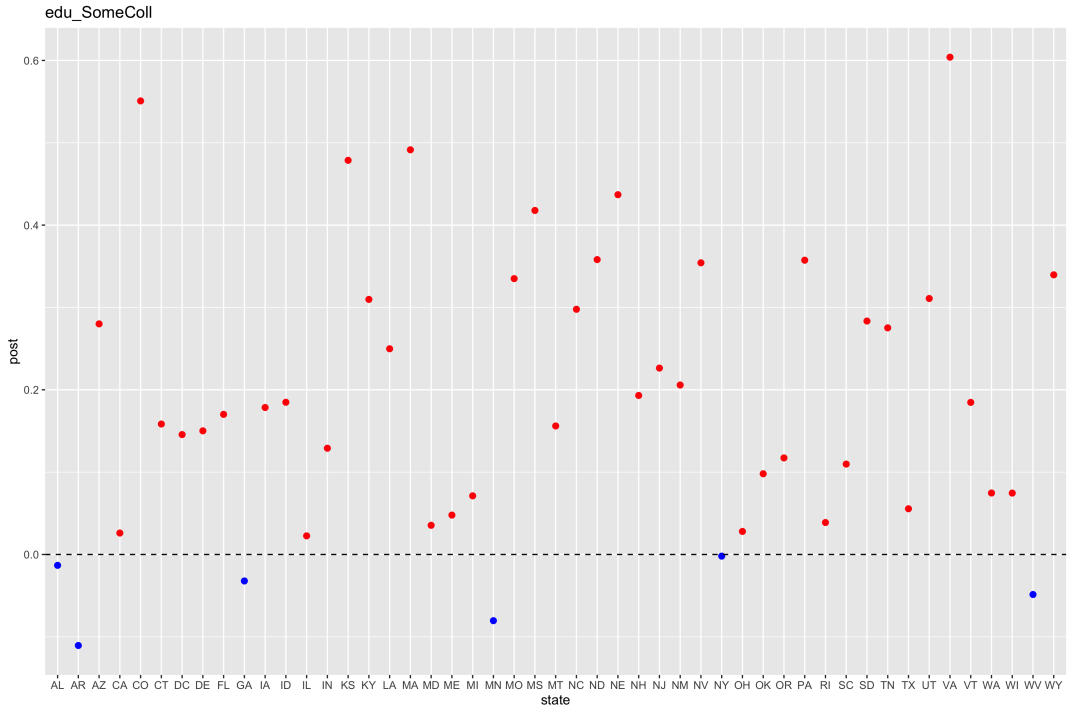
$$p(\sigma_k^2 \mid -) = \text{Inv-Ga}(\frac{1}{2} + \frac{1}{n}, \frac{1}{2} + \frac{1}{2}\sum_{i=1}^{n} (\beta_{ik} - \mu_k)^2).$$

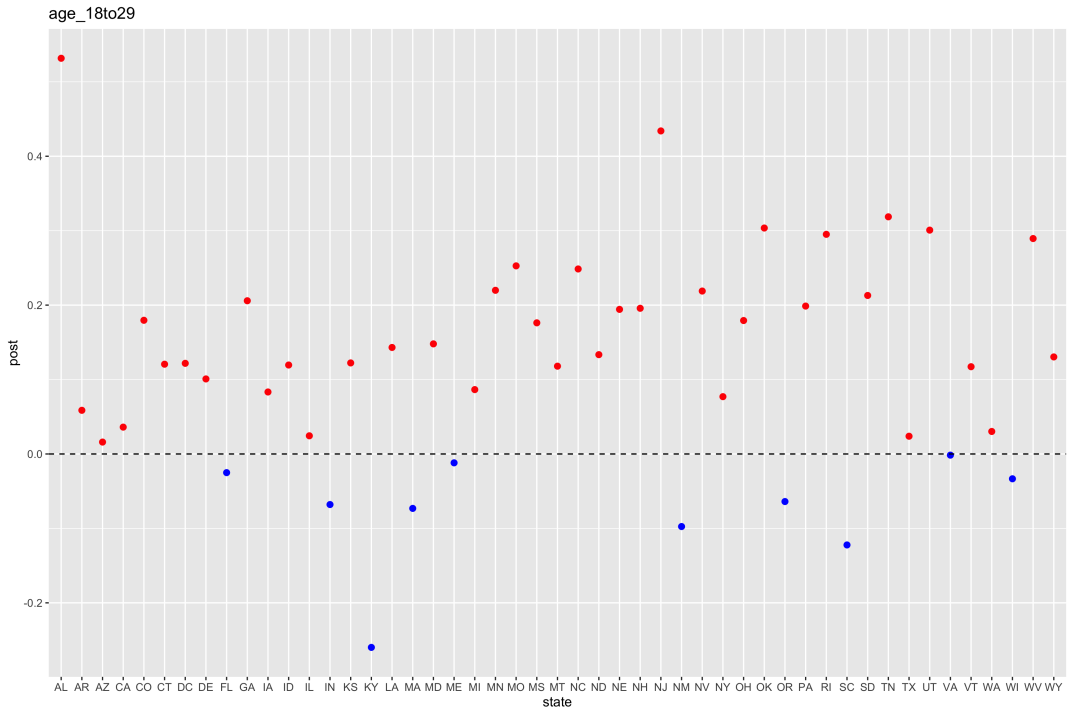
$$p(a_{ij} \mid -) = N(a_{ij} \mid x_{ij}^T \beta_i, 1) \begin{cases} 1(a_{ij} > 0), & y_{ij} = 1 \\ 1(a_{ij} \leq 0), & y_{ij} = 0 \end{cases}$$

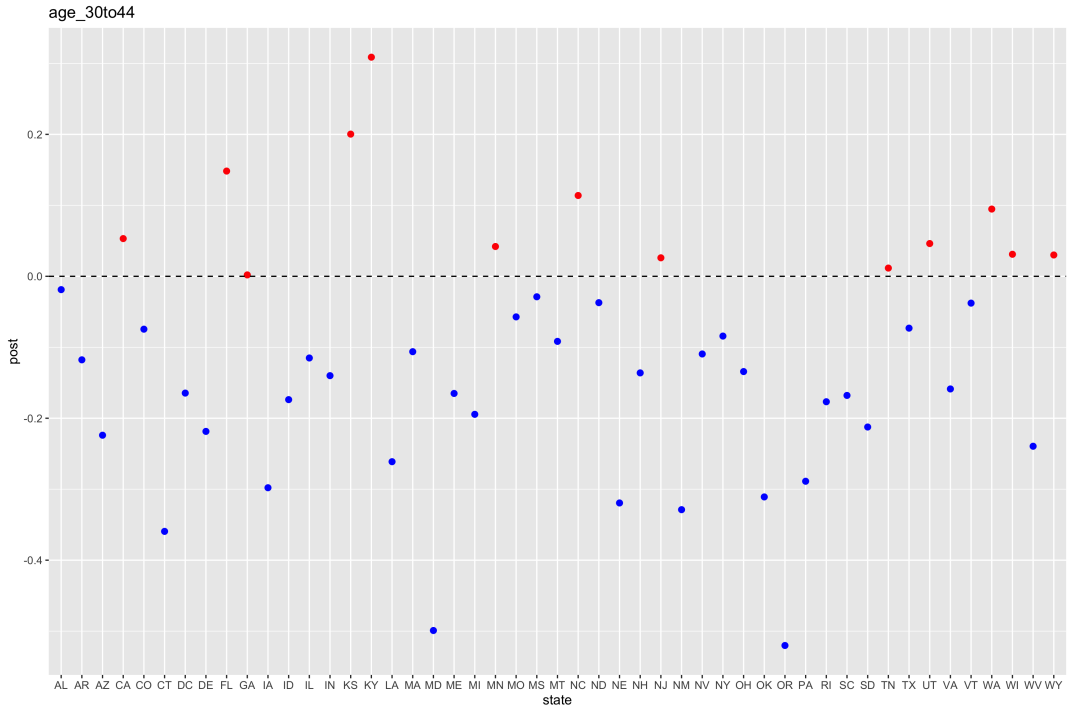

intercept
post
0.6
0.4
0.2
0.0
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state

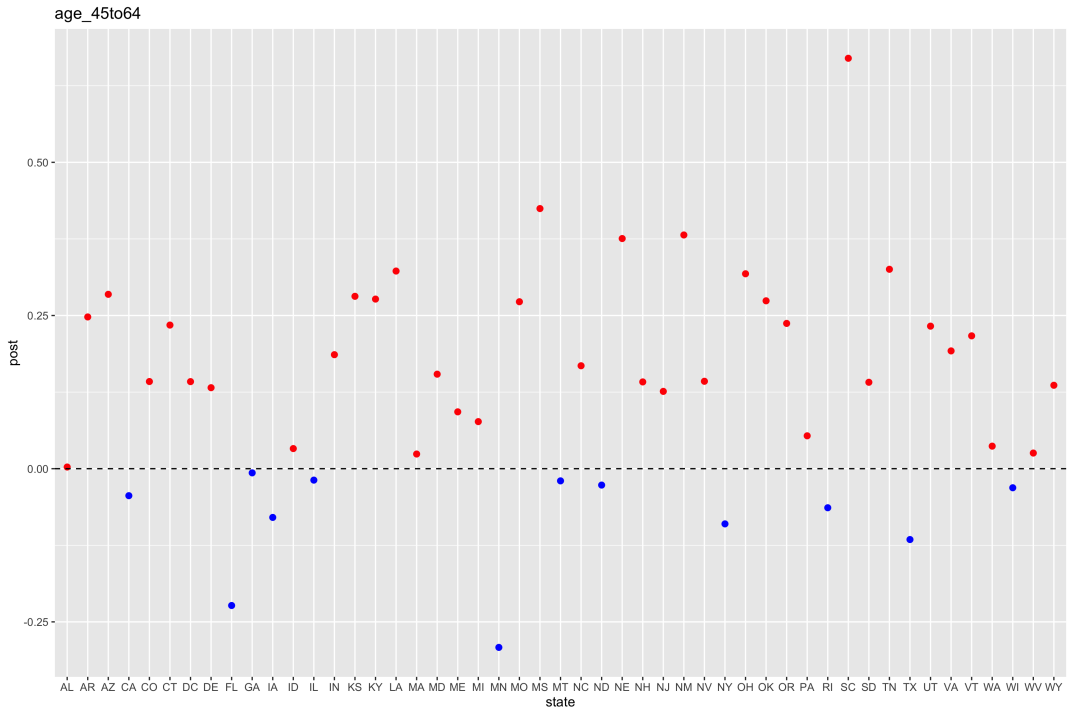

edu_NoHS
post
0.0
-0.2
-0.4
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state

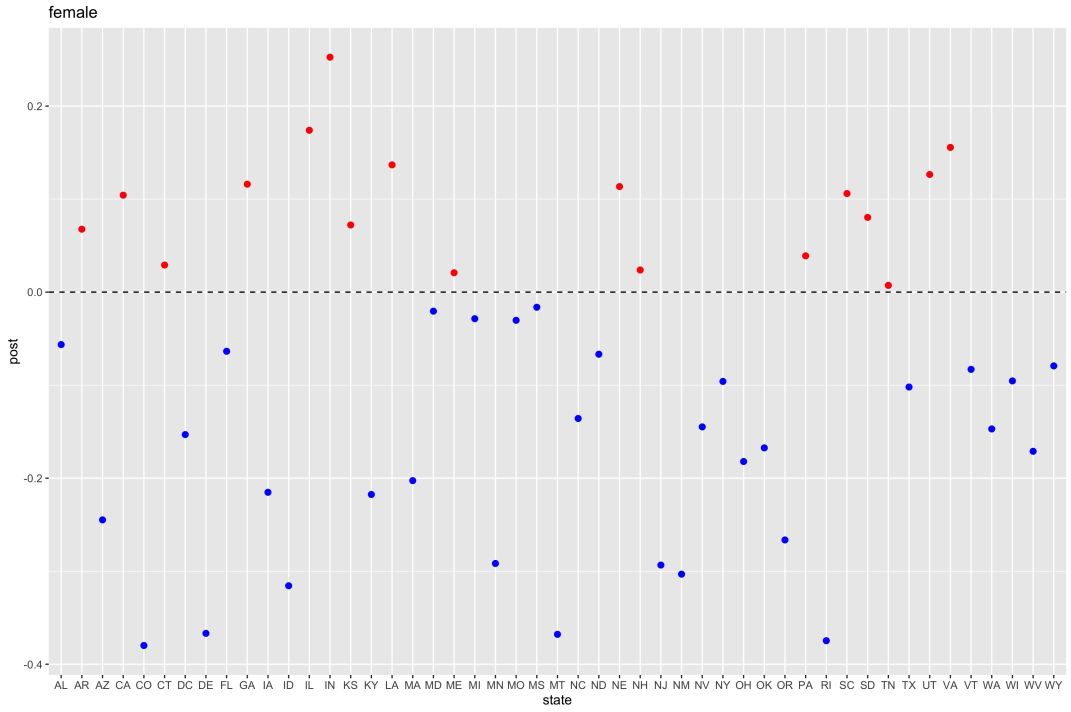

edu_HS
post
0.2
0.0
-0.2
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state


edu_SomeColl
post
0.6
0.4
0.2
0.0
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state


age_18to29
post
0.4
0.2
0.0
-0.2
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state


age_30to44
post
0.2
0.0
-0.2
-0.4
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state


age_45to64
post
0.50
0.25
0.00
-0.25
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state


female
post
0.2
0.0
-0.2
-0.4
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state

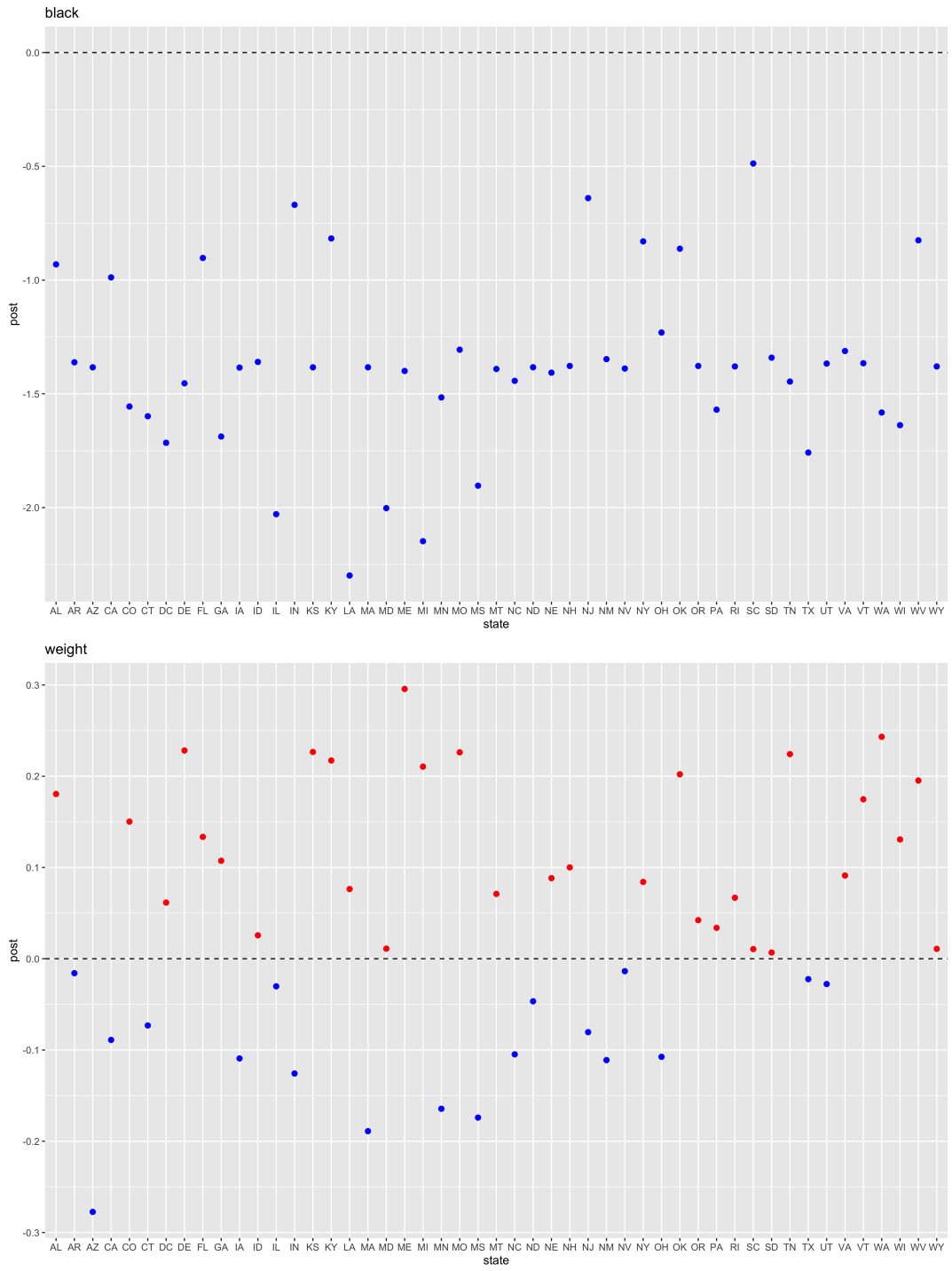


Figure 9: Posterior estimates of the covariates for each state.

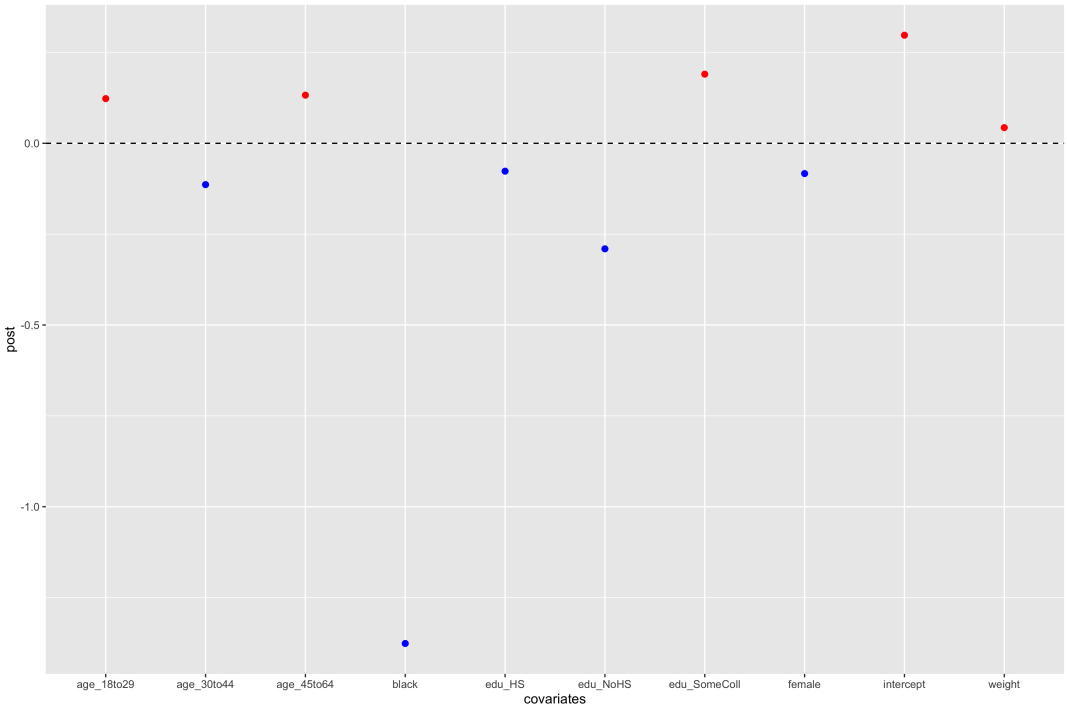


Figure 10: Posterior estimates of the covariates.