Statistical Modeling SDS 383D: Excercise 5

Khai Nguyen - kbn397
April 19, 2022

Price elasticity of demand :
We denotes $i = 1, \ldots, n$ (for stores), $j = 1, \ldots, N_i$ (for observations of the store i), the price at store i and observation j is P_{ij}, the display at store i and observation j is $disp_{ij}$, and the quantity demand at store i and observation j is Q_{ij} The model is

$$\log Q_{ij} = \beta_{1i} + \beta_{2i} \log P_{ij} + \beta_{3i} 1_{disp_{ij}=1} + \beta_{4i} \log P_{ij} 1_{disp_{ij}=1} + \epsilon_{ij},$$

where $\epsilon_{ij} \sim \mathcal{N}(0, 1)$. In the matrix form, let

$$\begin{aligned}
y_{ij} &= \log Q_{ij} \\
X_{ij} &= \left(1, \log P_{ij}, 1\left(\text{disp}_{ij}\right), \log P_{ij} \cdot 1\left(\text{disp}_{ij}\right)\right)^T \\
\beta_i &= \left(\beta_{1i}, \beta_{2i}, \beta_{3i}, \beta_{4i}\right)^\top
\end{aligned}$$

then our model is

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

Putting priors to our parameters, we have the following hierarchical model

$$\begin{aligned}
&p\left(y_{ij} \mid \beta_i, \sigma_i^2\right) = \mathcal{N}\left(X_{ij}^T \beta_i, \sigma_i^2\right) \\
&p\left(\beta_i \mid \mu_\beta, \Sigma\right) = \mathcal{N}\left(\mu_\beta, \Sigma = \operatorname{diag}\left(s_1^2, \ldots, s_4^2\right)\right) \\
&p\left(\mu_\beta\right) \propto 1 \\
&p\left(s_p^2\right) = \text{Inv-Ga}\left(\frac{1}{2}, \frac{1}{2}\right) \\
&p\left(\sigma_i^2\right) = \text{Inv-Ga}\left(\frac{a}{2}, \frac{b}{2}\right) \\
&p(a) = \text{Gamma}(2, 1) \\
&p(b) = \text{Gamma}(2, 1)
\end{aligned}$$

I skip the derivation of the complete conditional posteriors here since we have done it many times. The conditional posteriors are:

$$p\left(\beta_i \mid -\right) = \mathcal{N}\left(\left(\frac{X_i^T X_i}{\sigma_i^2} + \Sigma^{-1}\right)^{-1}\left(\frac{X_i^T Y_i}{\sigma_i^2} + \Sigma^{-1}\mu_\beta\right), \left(\frac{X_i^T X_i}{\sigma_i^2} + \Sigma^{-1}\right)^{-1}\right)$$

and

$$p\left(\mu_\beta \mid \cdot\right) = \mathcal{N}\left(\frac{1}{n}\sum_{i=1}^n \beta_i, \frac{1}{n}\Sigma\right)$$

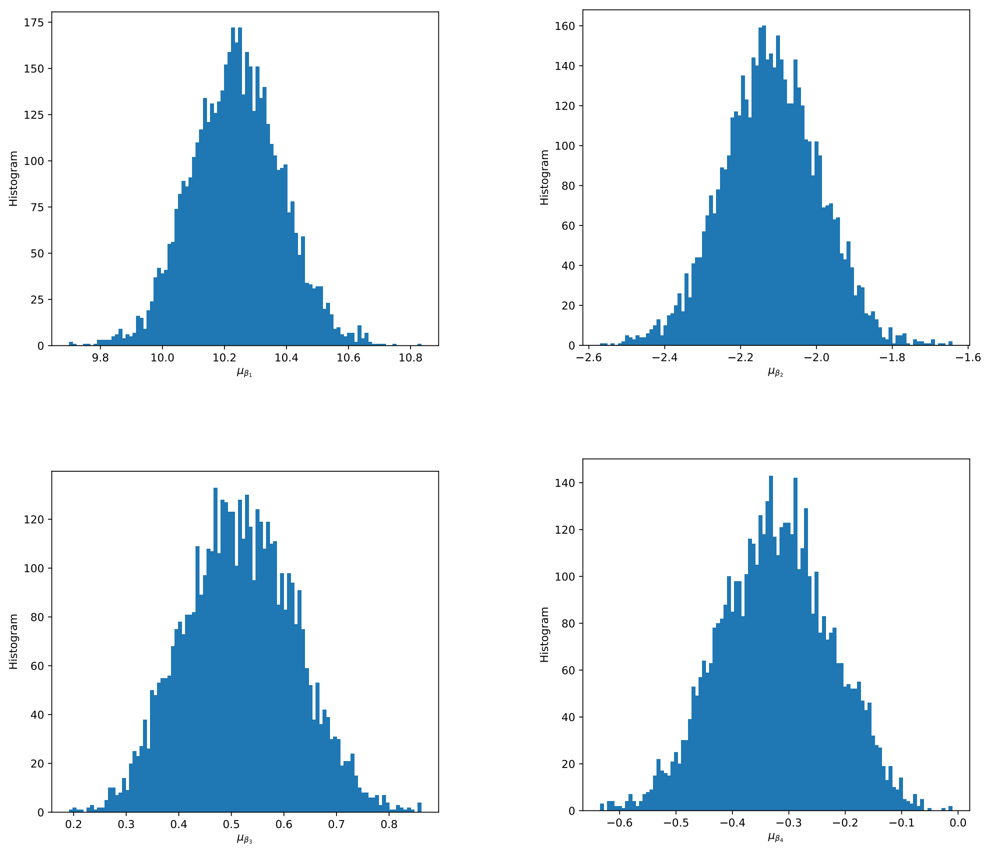

Figure 1: Histograms of Gibbs samples for $\mu_{\beta_1}, \mu_{\beta_2}, \mu_{\beta_3}, \mu_{\beta_4}$.

and

$$p\left(s_p^2 \mid -\right) = \text{ Inverse-Gamma } \left(\frac{n}{2}+\frac{1}{2}, \frac{1}{2}\left(1+\sum_{i=1}^n\left(\beta_{ip}-\mu_{\beta_p}\right)^2\right)\right)$$

and

$$p\left(\sigma_i^2 \mid \cdot\right) = \text{ Inverse-Gamma } \left(\frac{a}{2}+\frac{1}{2}, \frac{1}{2}\left(b+\left(Y_i-X_i \beta_i\right)^T\left(Y_i-X_i \beta_i\right)\right)\right)$$

Using MCMC for update a and b with the proposal Uniform(1, 8), the acceptance ratio is:

$$\alpha_{a,b} = \frac{Gamma(a_{new}|2,1)Gamma(b_{new}|2,1)\prod_i Inv-Gamma(\sigma_i|a_{new}, b_{new})}{Gamma(a_{old}|2,1)Gamma(b_{old}|2,1)\prod_i Inv-Gamma(\sigma_i|a_{old}, b_{old})}$$

I run the Gibbs sampler for 10000 iterations and thin 5000 samples. I plot the histograms of Gibbs samples for $\mu_{\beta_1}, \mu_{\beta_2}, \mu_{\beta_3}, \mu_{\beta_4}$ in Figure 1. From the figure, we can see that the intercept is about 10 for each store. From μ_{β_2}, we can know that the demand will decrease when the price increases (in

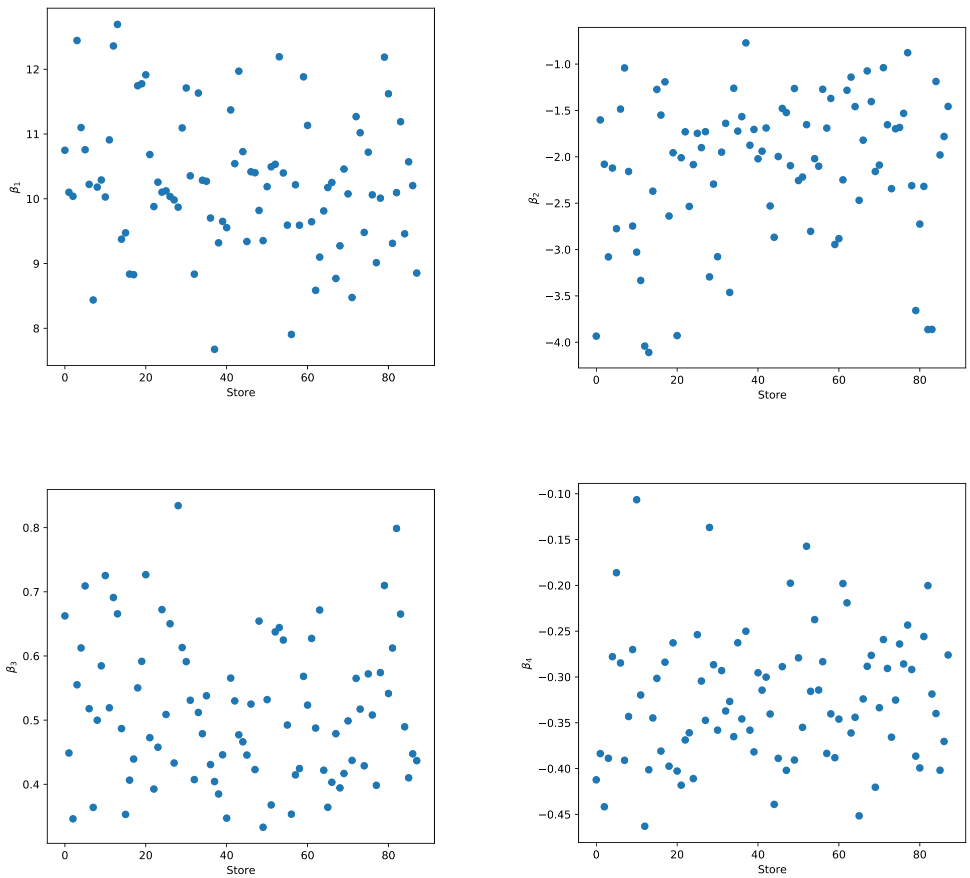

Figure 2: Posterior means of $\beta_1, \beta_2, \beta_3, \beta_4$ of each store.

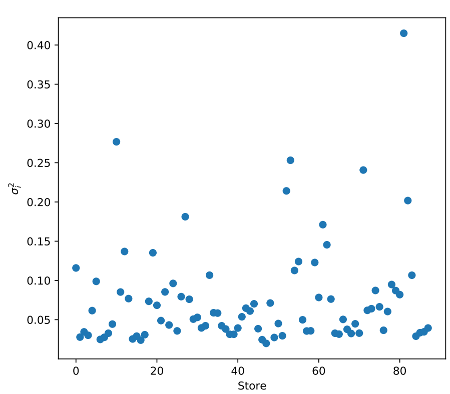

Figure 3: Posterior means of σ of each store.

log-scale). Having advertisement can also help to increase the demand based on μ_{β_3}, however, the advertisement also leads to the lower demand by increasing the price from μ_{β_4}.
Moreover, I plot the posterior means of $\beta_1, \beta_2, \beta_3, \beta_4$ of each store in Figure 2. From the figure, we can see that the effect of using advertisement is different across stores. However, the effect is not very significant since the difference is quite small.
Finally, I plot the posterior mean of σ of each store in Figure 3. We see that the variances of all stores seems to be between 0.05 to 0.15. Some stores have higher variances which might because the number of observations in those stores are relatively small. Overall, using different variances for different stores is a good model specification.

A hierarchical probit model :
I use the Polya Gamma model for Logistic Regression. I uses all the features (and an intercept) for the model. I change the categorical data into indicators of the corresponding value. Finally, I get 12 features. Similar to the previous problem, I denote $i = 1, \ldots, n$ for states, and $j = 1, \ldots, N_i$ for individuals in state i. Let $\sigma(.)$ is the Sigmoid function, the model is

$$\begin{aligned}
&P(Y_{ij} = 1) = \sigma(X_{ij}^\top \beta_i) \\
&p(\beta_i | \mu_{\beta_i}, \Sigma) = \mathcal{N}(\mu_{\beta_i}, \Sigma_i) \\
&p(\mu_{\beta_i}) = \mathcal{N}(0, I) \\
&p(\Sigma_i) = Inv - W(p + 1, I).
\end{aligned}$$

The Gibbs sampler is defined by:

$$\begin{aligned}
&p(\Sigma_i | -) = Inv - W(N_i + p + 1, I + \sum_{i=1}^{n} (\beta_i - \mu_{\beta_i})(\beta_i - \mu_{\beta_i})^\top) \\
&p(\mu_{\beta_i} | -) = \mathcal{N}((\Sigma_i^{-1} + I)^{-1} \beta_i^\top \Sigma_i^{-1}, (\Sigma_i^{-1} + I)^{-1}) \\
&w_{ij} | - \sim PG(1, X_{ij}^\top \beta_i) \\
&\beta_i | - \sim \mathcal{N}(m_{w_i}, V_{w_i}),
\end{aligned}$$

where $V_{\omega_i} = \left(X_i^T \Omega_i X_i + (\Sigma)^{-1}\right)^{-1}$, $m_{\omega_i} = V_{\omega_i}\left(X_i^T \kappa_i + (\Sigma)^{-1} \mu_{\beta_i}\right)$ with $\kappa = \left(\left(y_{i1} - \frac{1}{2}\right), \ldots, \left(y_{1n} - \frac{1}{2}\right)\right)$, $\Omega_i = diag(w_i)$. This comes from James' paper about Poyla Gamma for Logistic regression.
I remove all observations that contain NaN. Next, I run the Gibbs sampler for 4000 iterations and thin 2000 samples. I show the posterior means of $\beta_0, \ldots, \beta_{11}$ across stats in Figure 4- Figure 15. From these figures, we can see that the intercept (β_0) indicates the trend is voting for Bush in almost all states. I do not understand the politics of the US, hence, I hope that the intercept shows the right thing. We can also see that people with Bachelor degree and attending some colleges seem to vote for Bush while No-high-school people do not vote for him. Moreover, young people seem to vote for Bush while blacks do not vote for Bush. The weight does not show any trends.

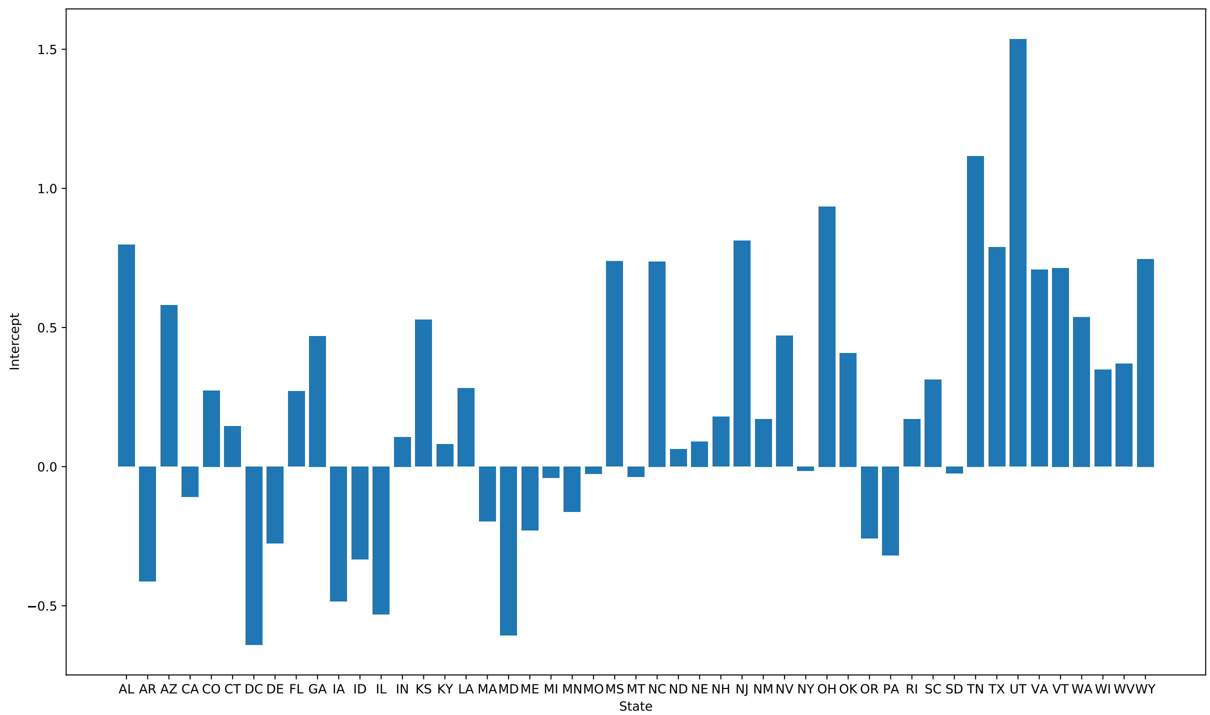

Figure 4: Posterior means of β_0 across stats.

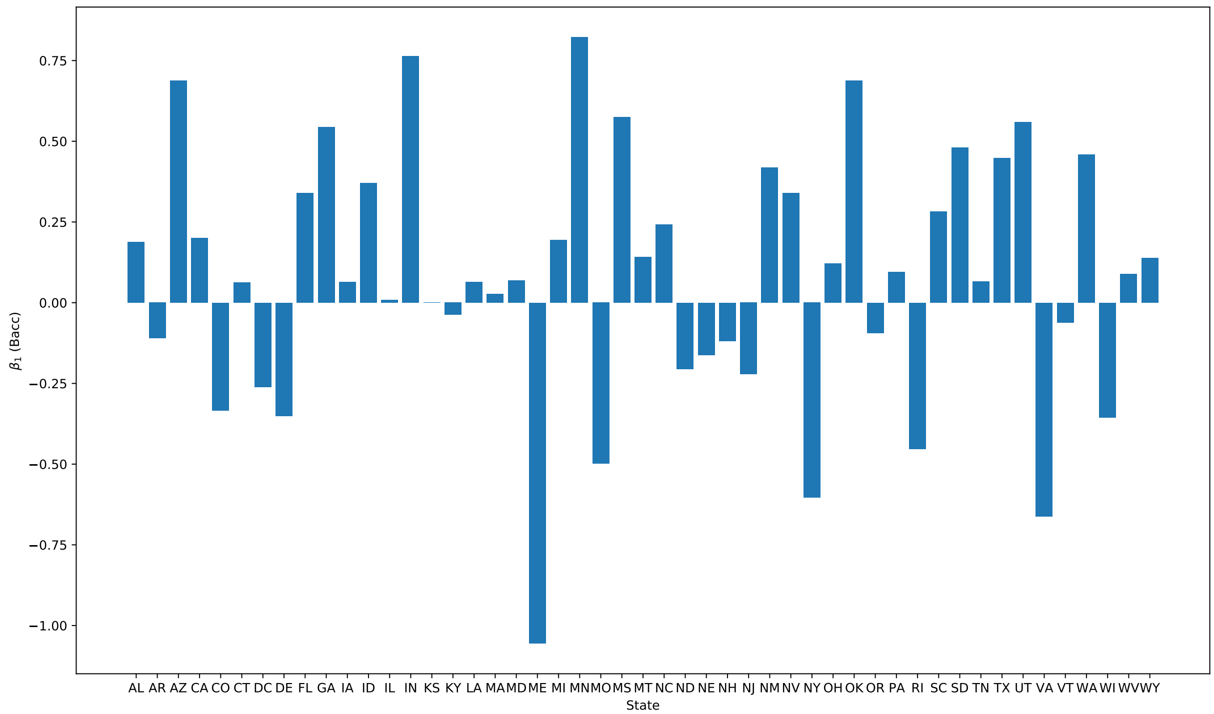

Figure 5: Posterior means of β_1 across stats.

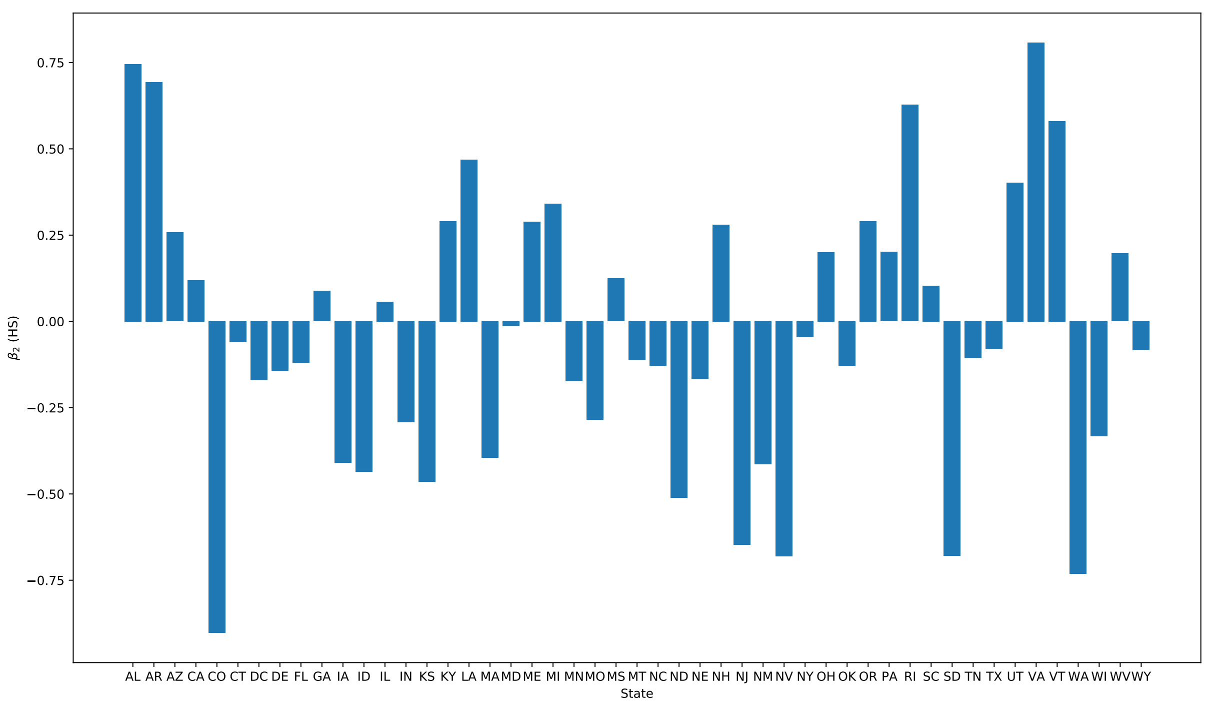

Figure 6: Posterior means of β_2 across stats.

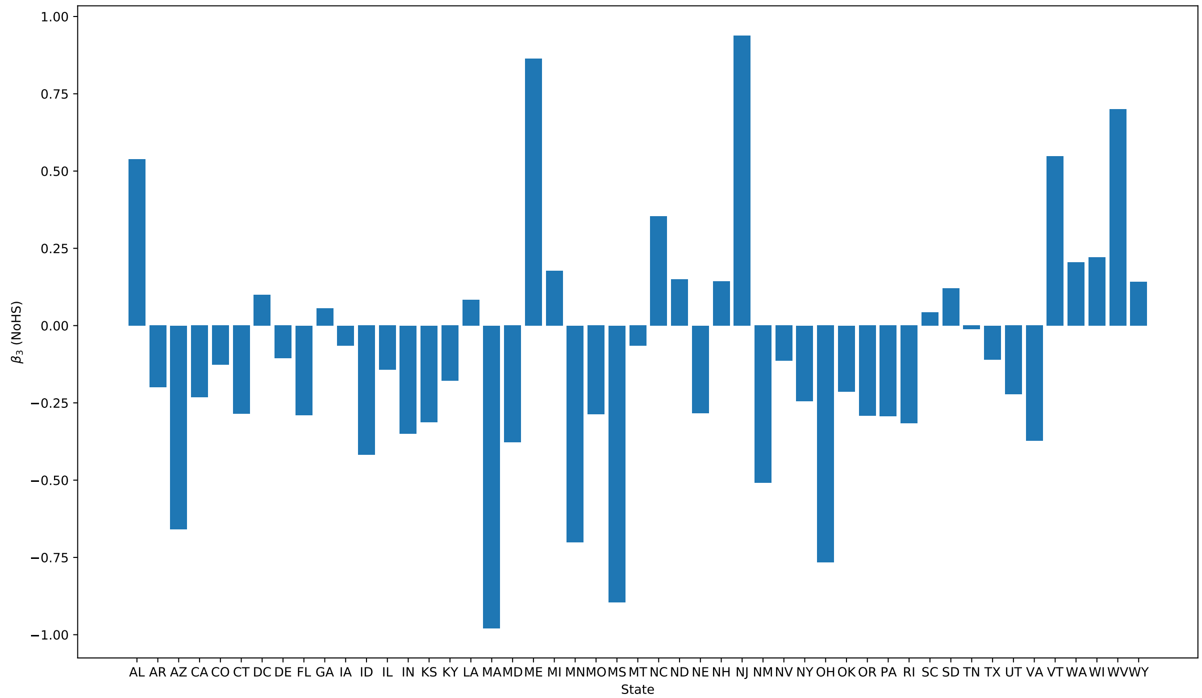

Figure 7: Posterior means of β_3 across stats.

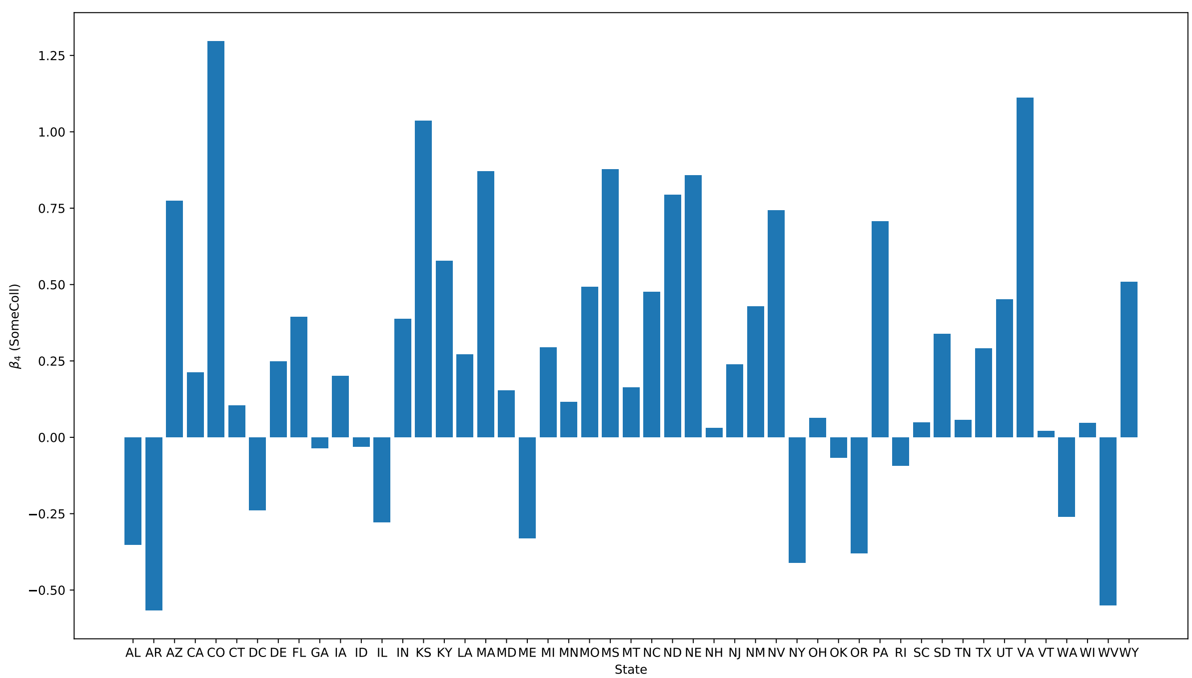

Figure 8: Posterior means of β_4 across stats.

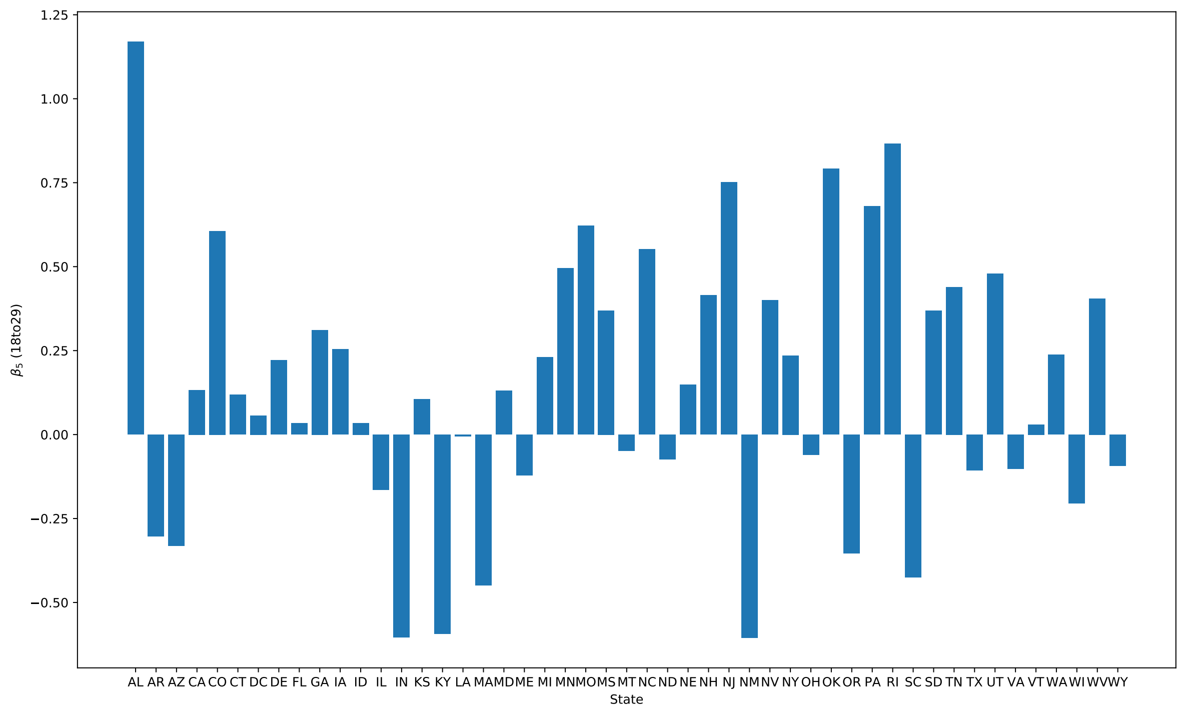

Figure 9: Posterior means of β_5 across stats.

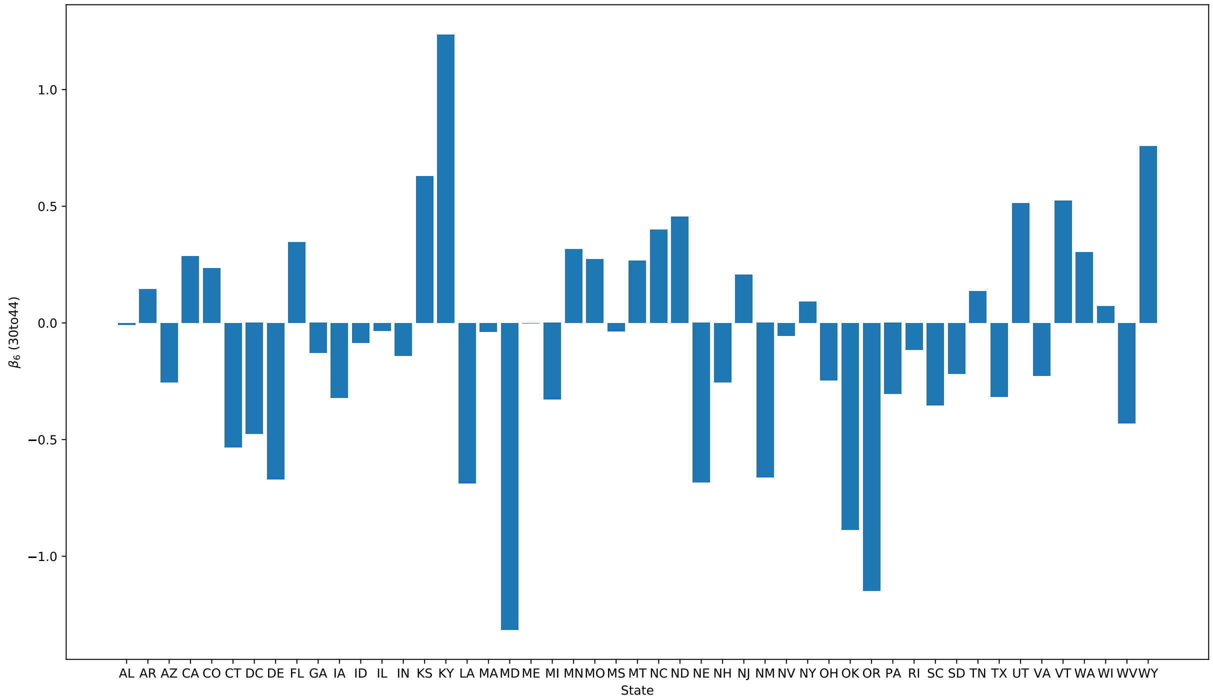

Figure 10: Posterior means of β_6 across stats.

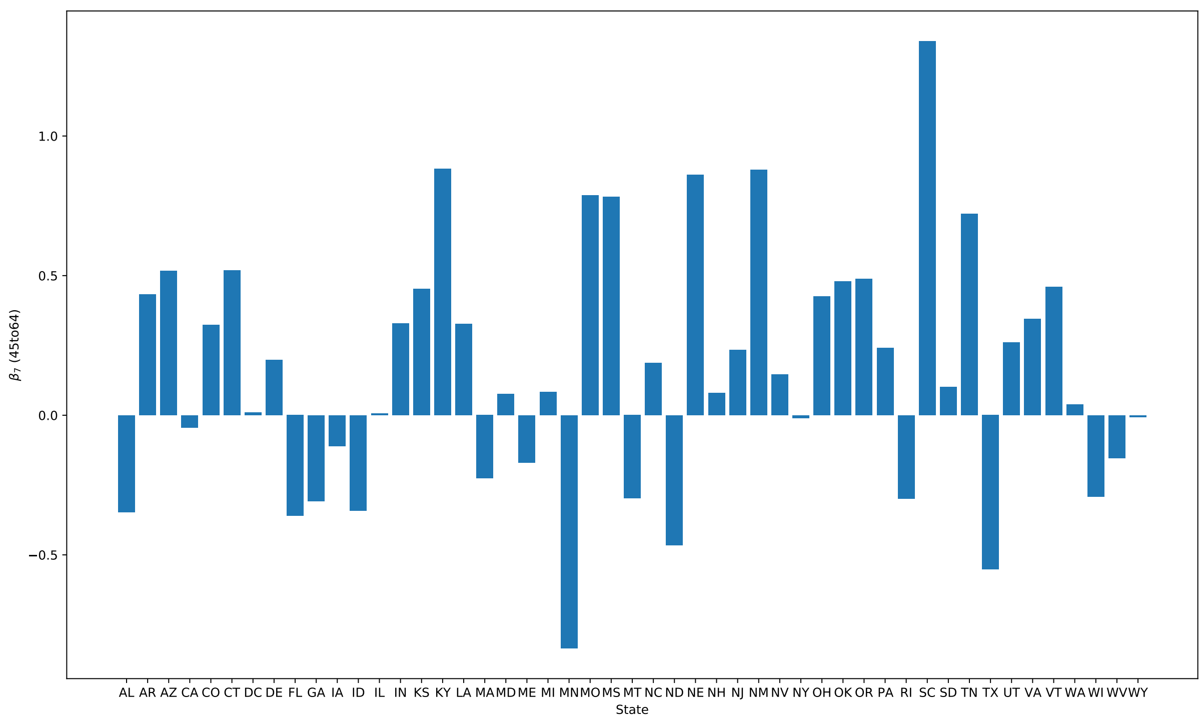

Figure 11: Posterior means of β_7 across stats.

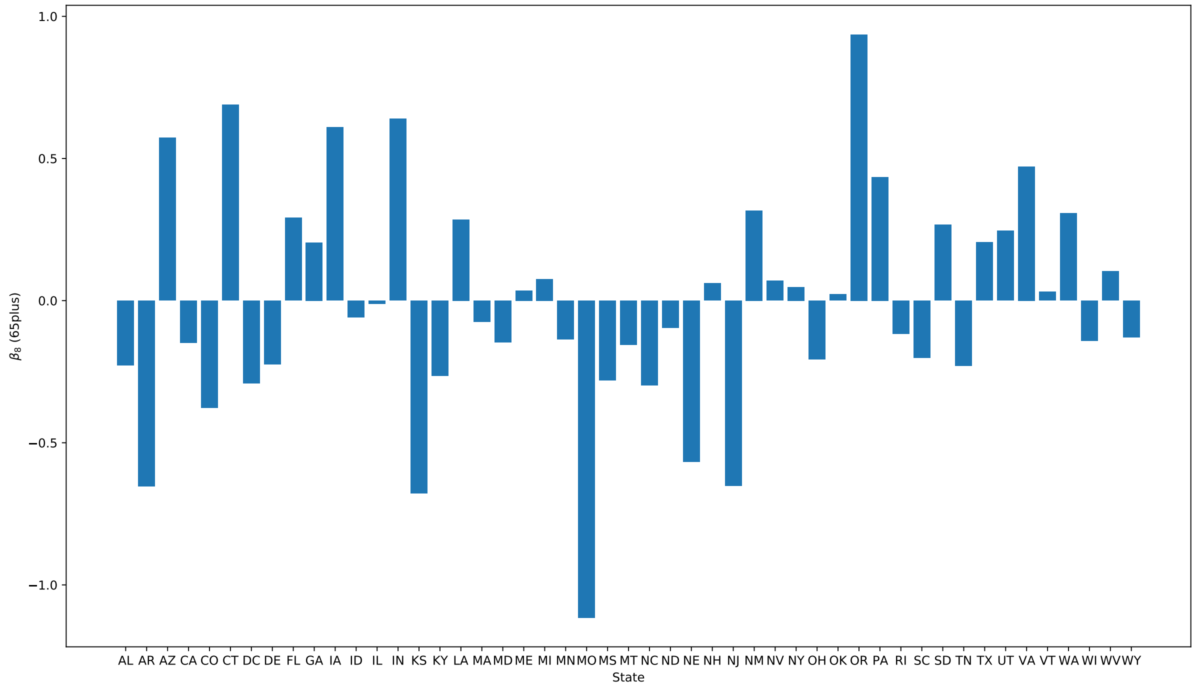

Figure 12: Posterior means of β_8 across stats.

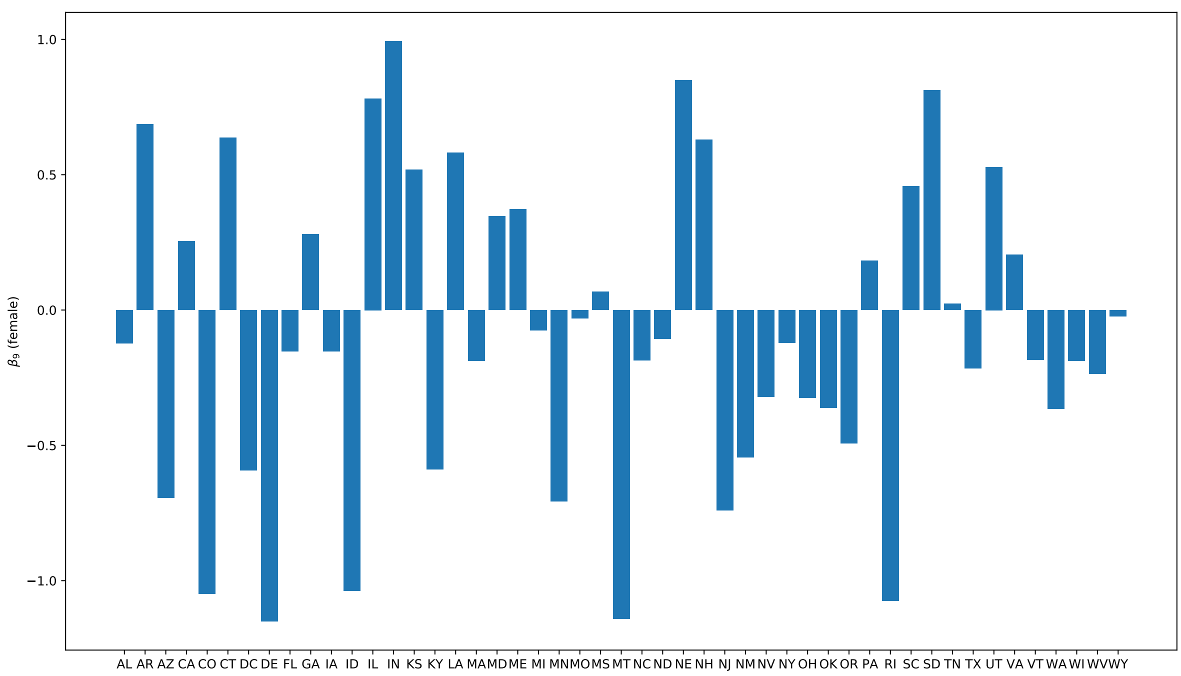

Figure 13: Posterior means of β_9 across stats.

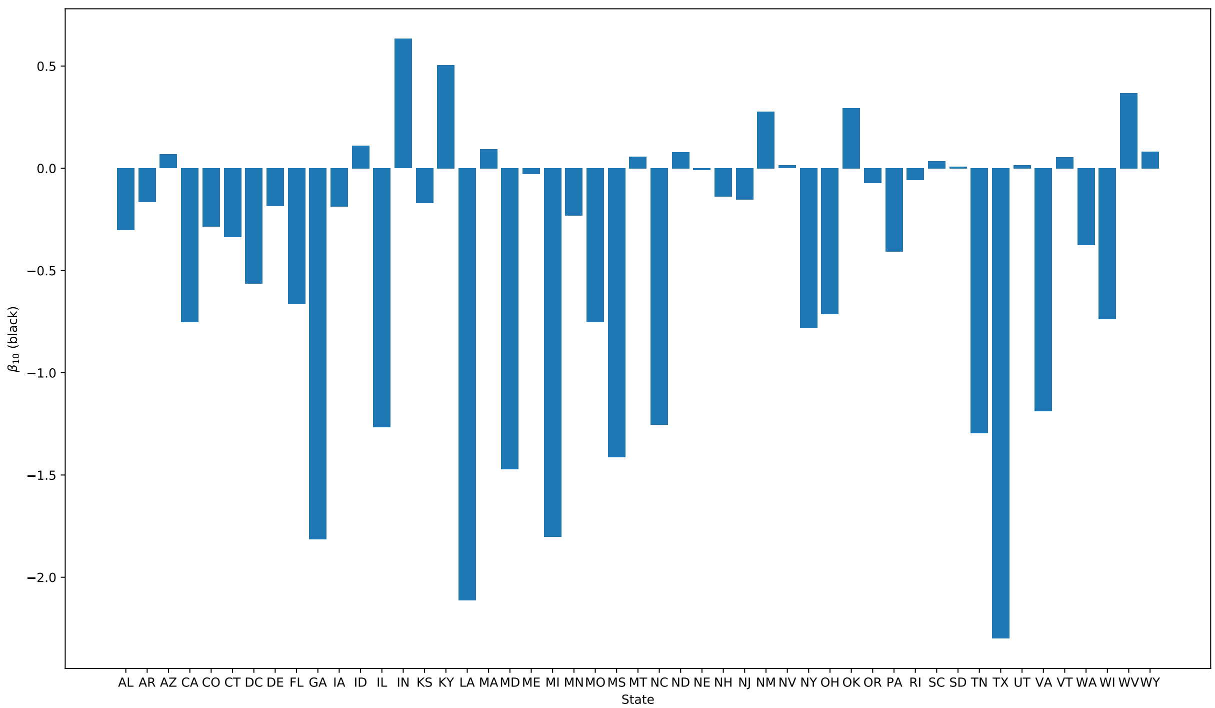

Figure 14: Posterior means of β_{10} across stats.

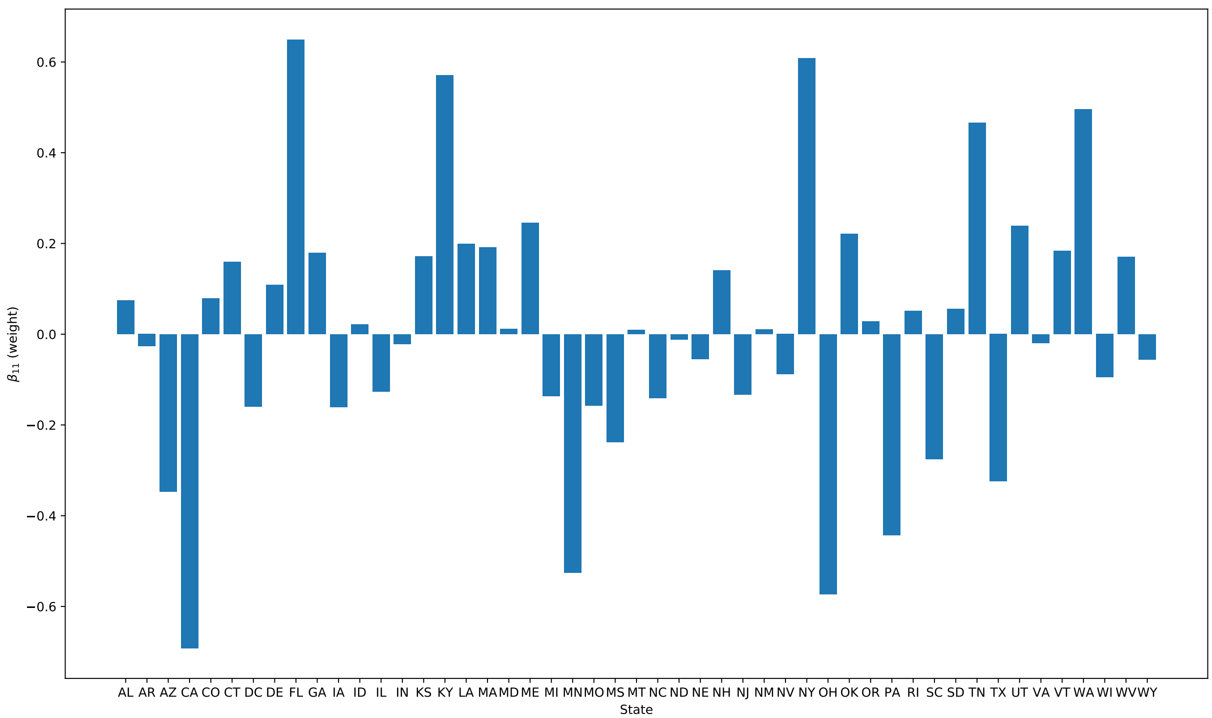

Figure 15: Posterior means of β_{11} across stats.