Lab 4

Aasha Reddy   Shirley Mathur   Bo Liu   Marc Brooks

9/17/2021

EDA

According to the histogram, the distribution of netincome is relatively normal (we restrict the x axis in the below graph). The distribution is somewhat right-skewed. We do not believe a log transformation of netincome is appropriate given this would eliminate all of the negative values. The QQ plot is roughly normal (in the below plot we trim off the top and bottom 5% of values).
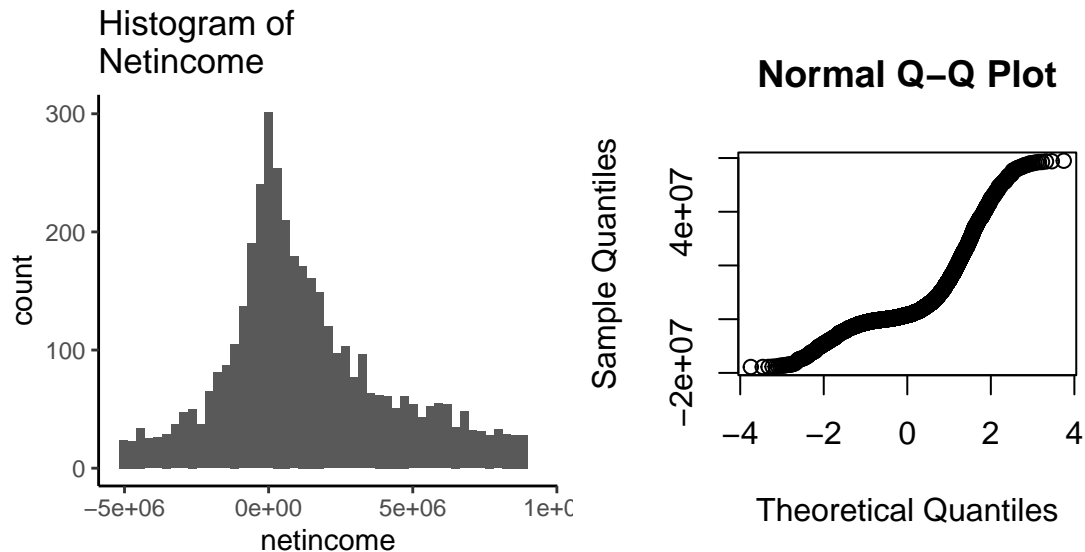

We can see below that the distribution of the mean net income by state is also relatively normal which is an assumption we care about for our model. This justifies a state specific intercept.
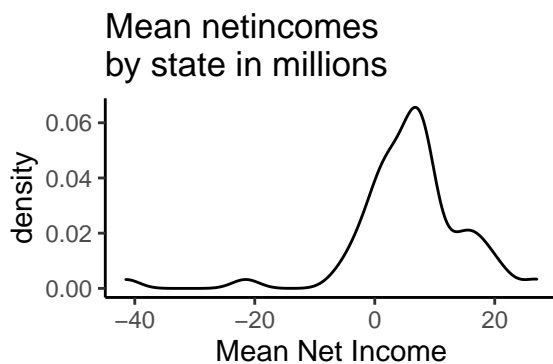
The boxplots below demonstrate that netincome varies by state, which again justifies a random intercept.

Sample size and mean net income by state


Distribution of netincome by state
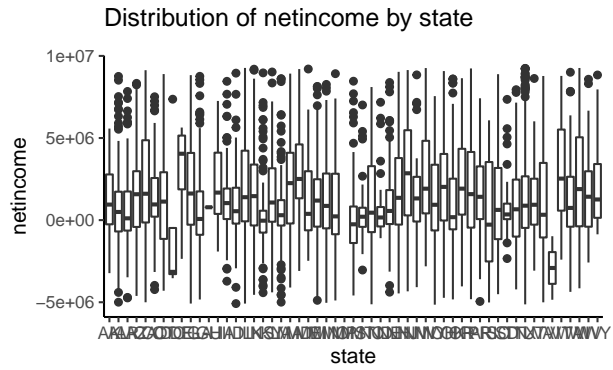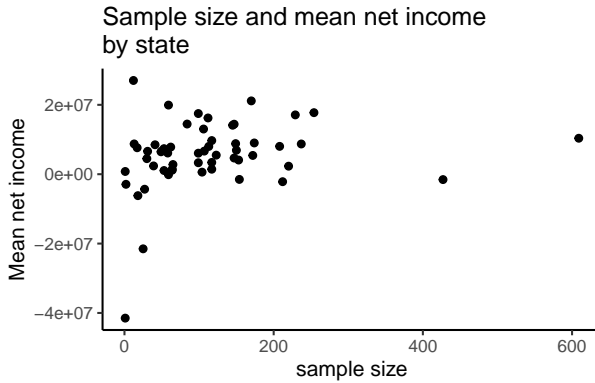
Although the scatter plot above shows that some states with smaller sample sizes have more extreme values, we decide not to disregard these potential outliers in our model.

Sample size by control group vs. average netincome

Boxplots of netincome by control group
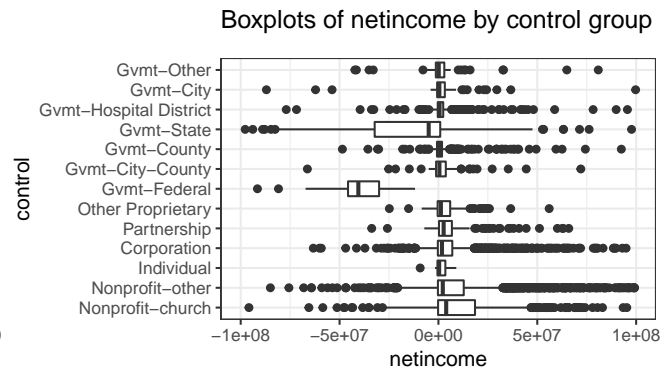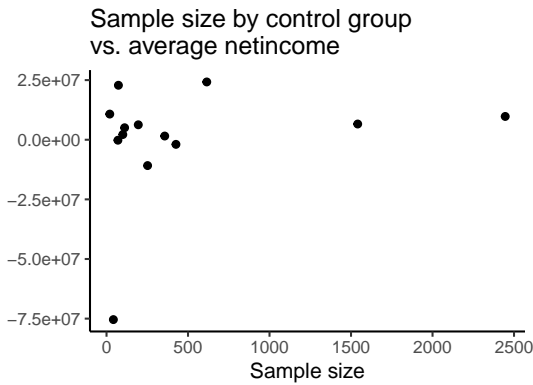
We also would like to examine the relationship between netincome and the control grouping variable. The plot above and on the left suggests that control groups have a sufficient sample size (the smallest being 42 observations). As we are using random effects, it also does not matter much how large our sample sizes are since we expect shrinkage towards the mean.

We can see based on the above boxplots of netincome by control group that most of the groups are centered around a similar mean. For this reason, we could try to include a random intercept.

Netincome vs. numbeds for all states

Netincome vs. numbeds for 4 states with largest sample size
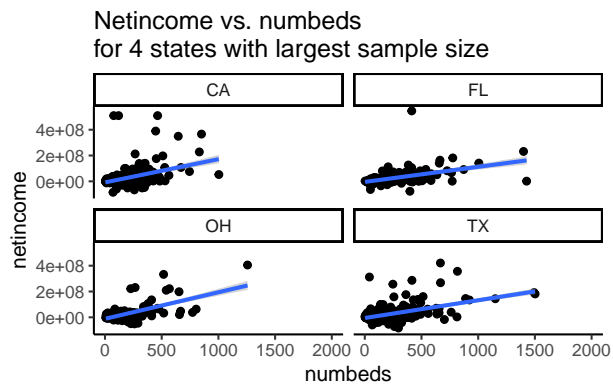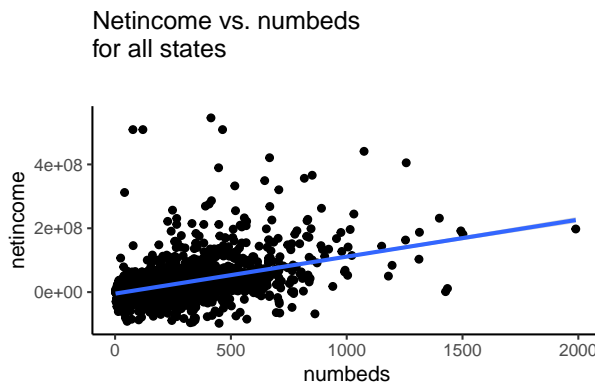
Examining the relationship between netincome and number of beds, we see that the slope for each state is similar to the slope across all the states. This justifies a fixed effect for numbeds.

Model Specification

$$y_{ijk} = \mu + \alpha_j + \gamma_k + \beta * B_{ij} + \epsilon_{ijk}$$

where i indicates the hospital i, j indicates the state j and k is the type of control. B_{ij} is numbeds.

$$\alpha_j \sim \mathcal{N}(0, \tau_\alpha^2), \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2), \quad \gamma_k \sim \mathcal{N}(0, \tau_\gamma^2)$$

Model Specification

We set the below priors. We selected priors that looked appropriate for our model. For the tau's, we are setting very weak priors because we know very little about variance of the random effects. For the grand mean, we assume a normal prior of N(0, 5) since the means were relatively centered around 0. From our intuition, a variance of 5 makes sense for our prior on μ.

$$\mu \sim \mathcal{N}(0, 5), \quad \sigma^2 \sim \mathcal{IG}(0.5, 5), \quad \beta \sim \mathcal{N}(0, 5)$$
$$\tau_{alpha}^2 \sim \mathcal{IG}(0.1, 0.1), \quad \tau_{alpha}^2 \sim \mathcal{IG}(0.1, 0.1)$$

For our model, we scaled the netincome values by dividing them by 1,000,000.

Evaluate diagnostics

We can see from the below plots that the chains seem to converge.
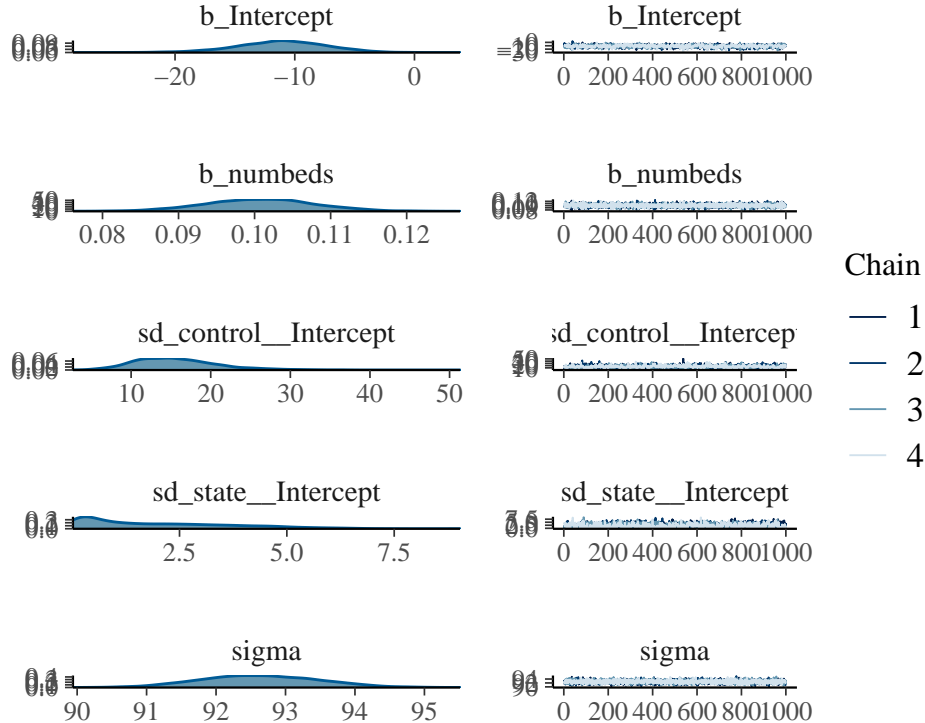
Results

For most of the states, the mean is very similar to each other. The random effects for state are relatively centered around 0. The intervals are also very similar across the states. No point estimates are outside of the

other state intervals, so there is not enough evidence to say that the random effects are different between any two hospita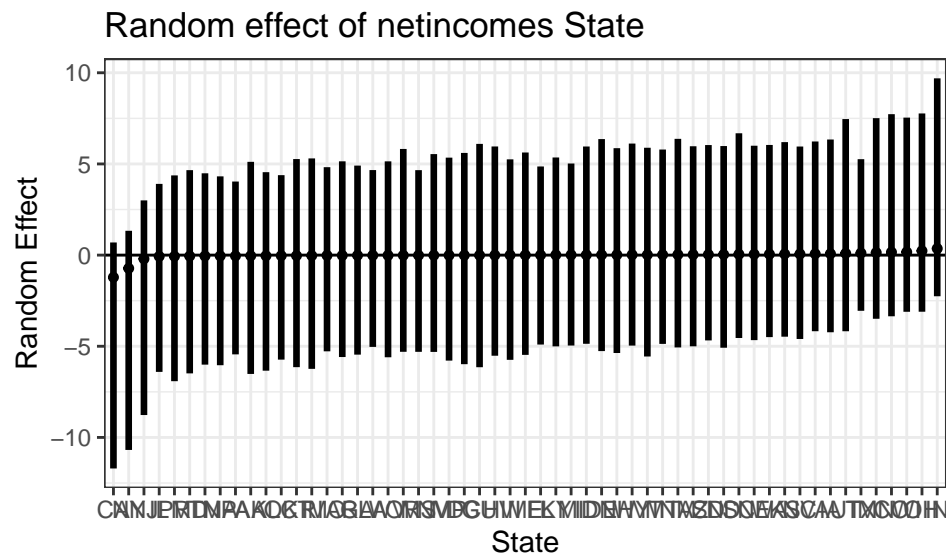ls (not really enough information to provide a ranking). We can provide a ranking based on the point estimate but there is a lot of uncertainty around this.

Random effect of netincomes State

Government-federal and government-state owned hospitals seems to have on average a negative effect on netincome. Non-profit church owned hospitals seems to have on average a positive effect on netincome.

Random effect of netincomes Control
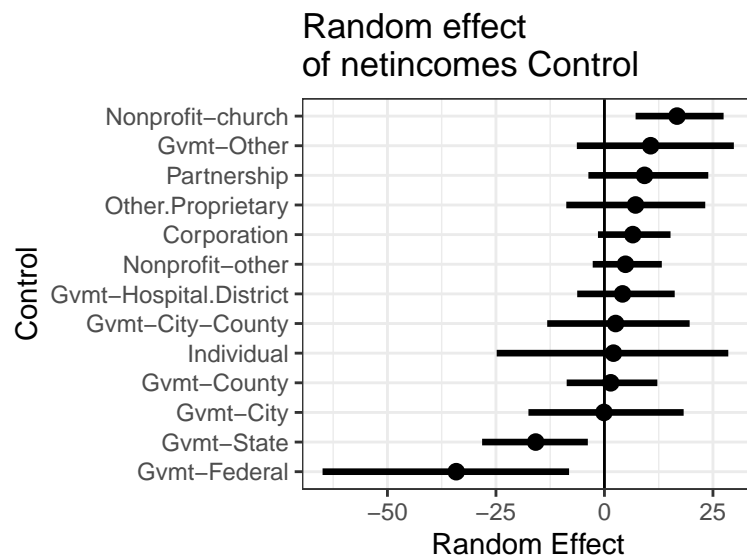
Below are our estimates of the parameters, including the 95% credible intervals. We can see that our estimate of μ, the grand mean is -11.321 million (credible interval of [-18.885, -4.311]). This suggests that the hospitals at baseline (no beds) are expected to on average lose money.

Our estimate of numbeds is 0.101 (credible interval of [0.087, 0.114]). We expect netincome to increase by 0.101 million on average for each additional bed.

This is evidence that health care costs too much money!

Table 1: Estimated posterior parameters

	Est	Lwr	Upr
Intercept (grand mean)	-11.103	-18.766	-4.209
numbeds	0.101	0.087	0.115
sigma	92.632	91.073	94.242
sd_control	15.657	7.006	27.866
sd_state	1.969	0.067	5.846

Table 2: Estimated group means of netincome by state (10 randomly selected)

state	Group Means	.lower	.upper
CT	-11.187	-20.465	-2.838
GU	-11.141	-20.291	-2.356
KY	-10.938	-20.065	-2.835
MI	-11.028	-20.318	-2.720
NH	-10.943	-20.279	-2.401
NM	-10.849	-20.156	-2.204
NY	-12.761	-24.234	-4.688
OK	-11.283	-19.990	-3.107
OR	-11.040	-20.158	-2.583
PR	-11.347	-21.240	-3.167

Table 3: Estimated group means of netincome by control

control	Group Means	.lower	.upper
Corporation	-4.450	-9.365	0.292
Gvmt-City	-11.115	-28.886	6.635
Gvmt-City-County	-8.399	-23.768	8.024
Gvmt-County	-9.533	-18.111	-1.267
Gvmt-Federal	-45.329	-77.442	-17.678
Gvmt-Hospital.District	-6.743	-15.801	2.444
Gvmt-Other	-0.419	-17.171	18.231
Gvmt-State	-26.920	-38.826	-15.559
Individual	-9.046	-37.120	17.229
Nonprofit-church	5.698	-2.120	13.615
Nonprofit-other	-6.160	-10.355	-1.867
Other.Proprietary	-3.769	-18.979	11.030
Partnership	-1.658	-14.033	10.762