

STATS 205 Final

Damon Bayer & Corey Katz

3/22/2019

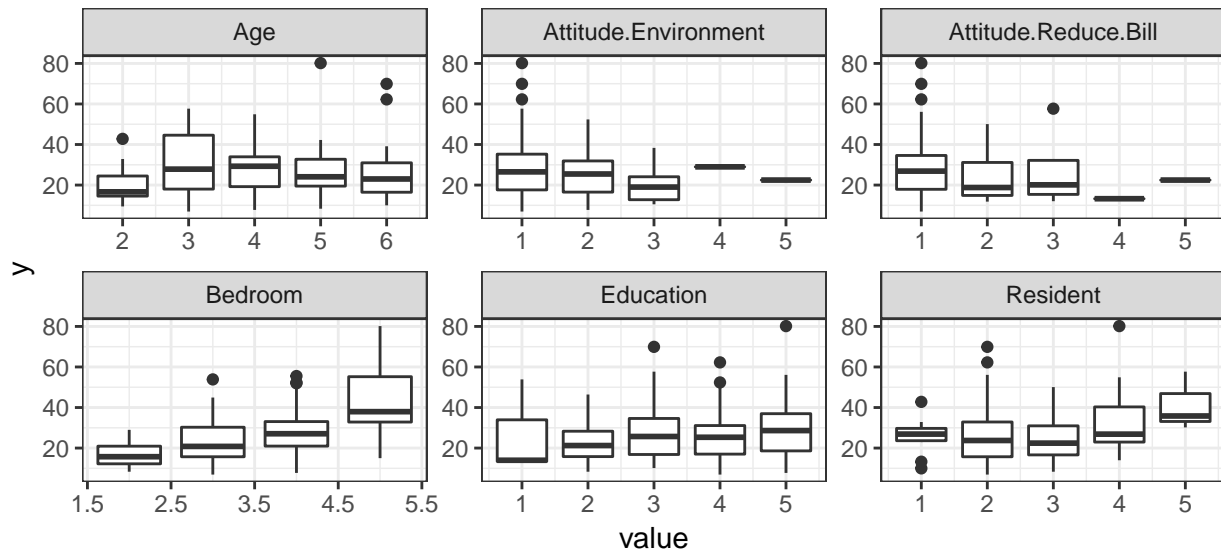
1. For simplicity, we will first focus on the first day of the four months period (November 15, 2009).

(a) In a few short sentences, propose a modeling approach to describe the daily total electricity consumption as a function of a (subset) of the covariates.

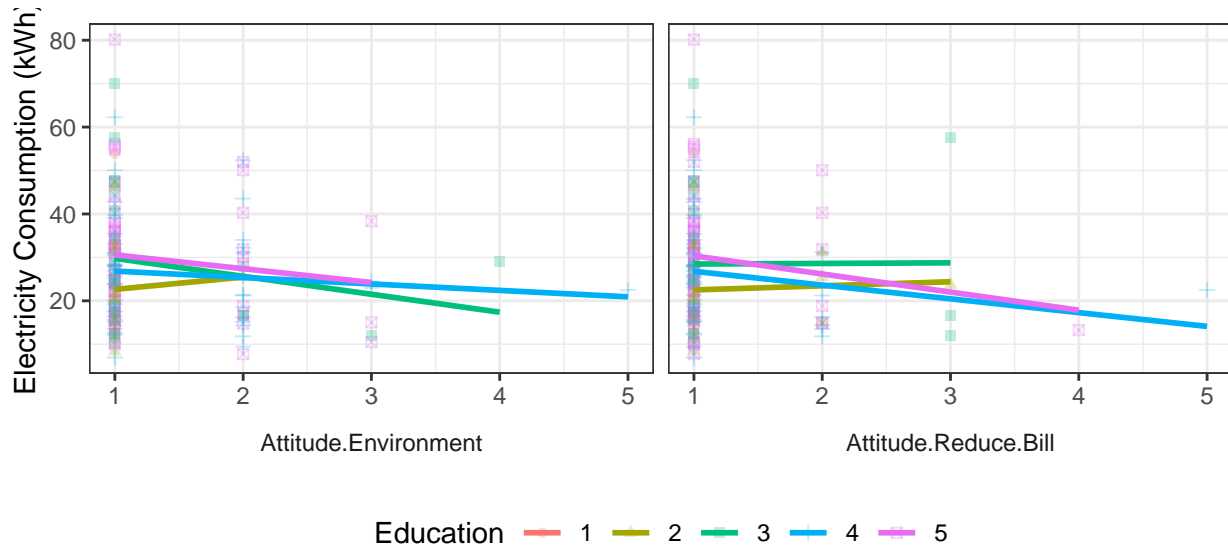
In order to describe the daily total electricity consumption as a function of the covariates, we propose a Bayesian linear regression model using column 7 (daily total electricity consumption on November 15, 2009) as the response variable and the first 6 columns (Age, Attitude-Reduce Bill, Attitude-Environment, Education, Resident, and Room) as predictors. Because Age, Attitude-Reduce Bill, Attitude-Environment, and Education take only 5 discrete values, we may consider treating them as categorical (factor) variables. However, because they are each ordinal values, we choose to treat them as continuous.

(b) Conduct an exploratory data analysis and discuss possible predictors you may consider in the model.

We begin our exploratory data analysis by looking at bi-variate plots of our possible predictors and our response:



Based on this plot, there is some evidence for including **Bedroom** and **Resident** in our model. We now consider some possible interactions. We hypothesize that households with more education are more likely to take action based on their beliefs, thus we look at possible interactions between **Education** and the two **Attitude** variables.



Based on these plots, there is some evidence for interaction between `Attitude.Reduce.Bill` and `Education`.

(c) **This is a methodological question, no data analysis required:** Consider a standard improper prior (SIR) and a prior which assumes $B_j \sim N(0, 10^6)$, $j = 1, \dots, p$, and $\tau \sim \text{Gamma}(0.001, 0.001)$ for inference on p regression coefficients and the precision in a regression model, respectively. What are the pros and cons of each choice? How do they differ?

The standard improper reference prior and vague priors are used when we have very little information regarding our parameters of interest. Using them when information of our prior is available, would be considered a disadvantage because we lose information that could have been incorporated into the model. Both priors share the same kernels, and the posterior means of the parameters will be very similar.

The standard improper reference prior is a flat prior distribution on β 's and on $\log(\tau)$. The use of the SIR prior results in frequentist-like inference on the beta's and tau. This can be considered both a pro and con by different statisticians. Unlike frequentist's, the inference using the SIR prior does not rely on large sample theory. Some programming languages, such as JAGS, cannot handle improper priors, which is a clear disadvantage to using the SIR.

The biggest difference between the SIR prior and the vague reference priors is that the vague priors are proper. Thus, they can also be used in JAGS, which makes them preferable to SIR. Vague priors are also useful for preliminary data analysis and sensitivity analysis. Overall, if we do not have any prior information vague priors are preferred.

(d) **This is a methodological question, no data analysis required:** Define the so-called g-prior for inference on p regression coefficients in a regression model. What are the pros and cons of the g-prior?

Zellner's G-priors, one of the most common prior choices, are of the following form:

$$\beta|\tau \sim N_r(\beta_0, \frac{g}{\tau}(X^T X)^{-1}) \quad \tau \sim \text{Gamma}(a, b)$$

where $\beta_0 = (X^T X)^{-1} X^T Y$ or 0. If β_0 is chosen to be $(X^T X)^{-1} X^T Y$, then the prior seems to be data dependent. For this reason, $\beta_0 = 0$ is the better choice.

Two benefits of using g-priors are that they are invariate and conjugate priors. This allows for easier calculations and more information to be provided to the prior based on how the choice of g . Different choices of g that determine how much weight is put on the prior information.

(e) This is a methodological question, no data analysis required: Consider a g -prior with 3 values of g : a value that puts equal weight to prior and likelihood, a value that has the equivalent weight of one observation, and the value suggested by Fernandez et al (2001), i.e. $g = \max(n, r^2)$. Based on your knowledge of the scientific problem considered here, which of the choices would you suggest?

In this problem, $n = 151$ and $r = 6$, which means that considering a g where the prior would have the equivalent to one observation ($g = n$) is the same as the g suggested by Fernandez. This limits our decision to choosing between $g = n$ and $g = 1$ (the g that gives equal weight to both the prior and the data). As discussed in part d) we want to use a g -prior that is centered at $\beta_0 = 0$, and therefore, do not want to put too much weight on the prior distribution. By setting $g = n$, we allow the prior distribution to have higher variance, which reflects our lack of prior knowledge. Giving too much weight to a prior centered at zero pulls our estimates toward zero. The covariance of the prior is also influenced by the covariance of our predictors. This could cause issues if the elements of the predictor covariance matrix are very small. Using $g = 1$, can result in a prior with a small variance and too much influence over the posterior. Overall, we want put more weight on the our data, so we set $g = n$.

(f) Discuss possible models for the Irish Electricity data, including different sets of predictors, and prior construction. Please, focus on main effects (i.e., no interactions) for now. Which variables would you surely want to include in the model? Which ones would you decide to exclude? You should try deleting terms and use model selection criteria for deciding which variables to keep and which to remove. Please, provide a motivated summary of your exploration. Be sure to explain and motivate all your steps.

Based on our answer in part E, we choose to proceed using the Zellner g -prior with $g = n$. From our exploratory data analysis, we see no reason to exclude any of our possible predictors. We would like to include interaction between Education and Attitude.Reduce.Bill, but will omit them for now.

Our Stan model is presented below.

```
data {
  int<lower=0> N;    // number of data items
  int<lower=0> K;    // number of predictors
  matrix[N, K] x;   // predictor matrix
  vector[N] y;      // outcome vector

  vector[K] mu_beta; // beta prior mean
  matrix[K, K] Sigma_beta; // beta prior covariance
  real<lower=0> a;
  real<lower=0> b;

  int<lower=0> M;    // number of new data items
  matrix[M, K] xnew;
}
parameters {
  vector[K] beta;    // coefficients for intercept & predictors
  real<lower=0> tau;  // error scale
}
transformed parameters{
  vector[N] CP0inv;
  vector[K] BetaMoreThan0;

  for (k in 1:K){
    BetaMoreThan0[k] = step(beta[k]);
  }
}
```

```

    for (n in 1:N){
      CP0inv[n] = sqrt(2*3.14159 / tau) * exp(0.5 * tau * pow(y[n] - x[n,] * beta, 2));
    }
  }
}
model {
  tau ~ gamma(a, b);
  beta ~ multi_normal(mu_beta, Sigma_beta / tau);

  y ~ normal(x * beta, 1 / sqrt(tau)); // likelihood
}
generated quantities {
  vector[M] ynew;
  vector[N] log_lik;

  for(i in 1:M) {
    ynew[i] = normal_rng(xnew[i,] * beta, 1 / sqrt(tau));
  }
  for (n in 1:N){
    log_lik[n] = normal_lpdf(y[n] | x[n] * beta, 1 / sqrt(tau));
  }
}

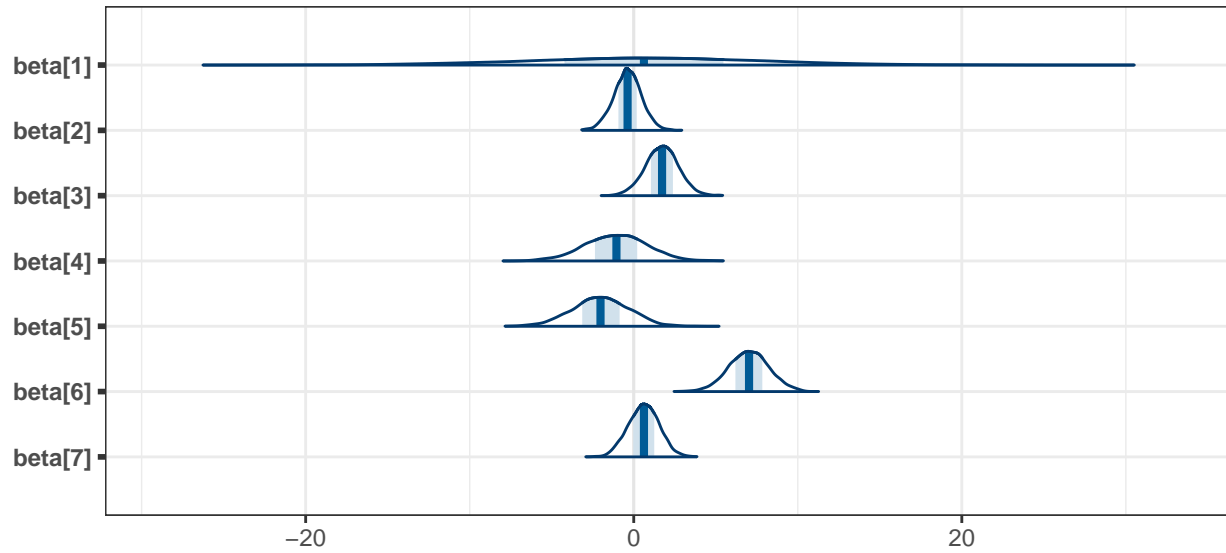
```

All Variables Model

Posterior inference for the All variables model are presented below. Note the betas correspond to **Intercept**, **Age**, **Resident**, **Attitude.Reduce.Bill**, **Attitude.Environment**, **Bedroom**, and **Education**, respectively.

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	0.5964	0.1604	7.3039	-13.7805	-4.2168	0.6166	5.4665	14.7451	2075	1.0007
beta[2]	-0.3801	0.0163	0.8496	-2.0781	-0.9336	-0.3756	0.1796	1.2719	2707	1.0013
beta[3]	1.7209	0.0175	1.0105	-0.2860	1.0481	1.7262	2.3882	3.6845	3330	1.0012
beta[4]	-1.0725	0.0372	1.8928	-4.8380	-2.3635	-1.0561	0.2110	2.5957	2589	1.0001
beta[5]	-2.0188	0.0337	1.7269	-5.3985	-3.1393	-2.0252	-0.8617	1.2477	2630	0.9994
beta[6]	7.0213	0.0220	1.2254	4.6059	6.1988	7.0277	7.8372	9.4557	3111	1.0011
beta[7]	0.6103	0.0175	0.9375	-1.2035	-0.0297	0.6227	1.2494	2.4277	2881	0.9998
tau	0.0073	0.0000	0.0009	0.0057	0.0068	0.0073	0.0079	0.0091	3628	0.9999

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
BetaMoreThan0[1]	0.5347	0.0100	0.4989	0	0	1	1	1	2503	1.0004
BetaMoreThan0[2]	0.3230	0.0084	0.4677	0	0	0	1	1	3122	1.0009
BetaMoreThan0[3]	0.9555	0.0039	0.2062	0	1	1	1	1	2750	1.0027
BetaMoreThan0[4]	0.2870	0.0083	0.4524	0	0	0	1	1	2949	1.0010
BetaMoreThan0[5]	0.1230	0.0062	0.3285	0	0	0	0	1	2803	0.9994
BetaMoreThan0[6]	1.0000	NaN	0.0000	1	1	1	1	1	NaN	NaN
BetaMoreThan0[7]	0.7405	0.0079	0.4384	0	0	1	1	1	3113	0.9996



From the plots and tables, we note that there is a high probability of beta 2 and beta 7 being near 0. We now consider removing the corresponding variables (**Age** and **Education**) from our model. We proceed by fitting a model without **Age**, a model without **Education** and a model with neither. Posterior inference tables for each are presented below. In the future models, we consider a coefficient to have “a high probability of being near 0” if mean `BetaMoreThan0` $\in (0.25, 0.75)$.

No Age Model

Note that betas correspond to `Intercept`, `Resident`, `Attitude.Reduce.Bill`, `Attitude.Environment`, `Bedroom`, and `Education`, respectively.

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	-1.2256	0.1355	6.1727	-13.1852	-5.3405	-1.2381	2.8564	11.2129	2077	1.0002
beta[2]	1.6913	0.0189	1.0382	-0.3980	0.9832	1.7127	2.3746	3.6892	3029	0.9997
beta[3]	-1.1934	0.0359	1.8845	-4.7748	-2.4446	-1.1704	0.0856	2.4407	2756	0.9999
beta[4]	-1.8555	0.0303	1.6895	-5.1178	-3.0103	-1.8590	-0.7296	1.5124	3104	0.9999
beta[5]	7.0255	0.0244	1.1992	4.7217	6.2248	7.0196	7.8640	9.2751	2410	1.0012
beta[6]	0.6885	0.0169	0.9114	-1.0881	0.1024	0.7010	1.3271	2.4421	2913	1.0001
tau	0.0073	0.0000	0.0009	0.0058	0.0067	0.0073	0.0079	0.0091	3582	0.9992

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
BetaMoreThan0[1]	0.4145	0.0099	0.4927	0	0	0	1	1	2464	0.9997
BetaMoreThan0[2]	0.9468	0.0043	0.2246	0	1	1	1	1	2736	1.0001
BetaMoreThan0[3]	0.2642	0.0082	0.4410	0	0	0	1	1	2882	1.0001
BetaMoreThan0[4]	0.1350	0.0063	0.3418	0	0	0	0	1	2916	0.9997
BetaMoreThan0[5]	1.0000	NaN	0.0000	1	1	1	1	1	NaN	NaN
BetaMoreThan0[6]	0.7735	0.0076	0.4186	0	1	1	1	1	3004	0.9995

We may consider `Attitude.Reduce.Bill` for removal if we were to continue with a step-wise selection process.

No Education Model

Note that betas correspond to `Intercept`, `Age`, `Resident`, `Attitude.Reduce.Bill`, `Attitude.Environment`, and `Bedroom`, respectively.

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	3.0139	0.1410	6.1315	-9.0005	-0.9779	3.1596	7.0055	14.900	1892	1.0006
beta[2]	-0.4754	0.0156	0.8000	-2.0239	-1.0324	-0.4826	0.0599	1.110	2626	1.0003
beta[3]	1.6884	0.0173	1.0346	-0.3894	1.0090	1.6959	2.3748	3.696	3588	1.0011
beta[4]	-1.1729	0.0315	1.8463	-4.8723	-2.3970	-1.1622	0.0711	2.357	3434	0.9998
beta[5]	-1.9068	0.0281	1.6635	-5.2341	-3.0254	-1.8566	-0.7863	1.326	3493	0.9997
beta[6]	7.1481	0.0258	1.1943	4.7664	6.3642	7.1823	7.9329	9.499	2145	1.0010
tau	0.0073	0.0000	0.0008	0.0058	0.0067	0.0073	0.0079	0.009	3694	0.9999

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
BetaMoreThan0[1]	0.6978	0.0099	0.4593	0	0	1	1	1	2166	0.9995
BetaMoreThan0[2]	0.2742	0.0084	0.4462	0	0	0	1	1	2846	0.9999
BetaMoreThan0[3]	0.9478	0.0044	0.2226	0	1	1	1	1	2516	1.0011
BetaMoreThan0[4]	0.2602	0.0078	0.4388	0	0	0	1	1	3195	0.9996
BetaMoreThan0[5]	0.1215	0.0060	0.3267	0	0	0	0	1	2935	1.0014
BetaMoreThan0[6]	1.0000	NaN	0.0000	1	1	1	1	1	NaN	NaN

We may consider `Attitude.Reduce.Bill` for removal if we were to continue with a step-wise selection process.

No Age or Education Model

Note that betas correspond to `Intercept`, `Resident`, `Attitude.Reduce.Bill`, `Attitude.Environment`, and `Bedroom`, respectively.

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	-1.2256	0.1355	6.1727	-13.1852	-5.3405	-1.2381	2.8564	11.2129	2077	1.0002
beta[2]	1.6913	0.0189	1.0382	-0.3980	0.9832	1.7127	2.3746	3.6892	3029	0.9997
beta[3]	-1.1934	0.0359	1.8845	-4.7748	-2.4446	-1.1704	0.0856	2.4407	2756	0.9999
beta[4]	-1.8555	0.0303	1.6895	-5.1178	-3.0103	-1.8590	-0.7296	1.5124	3104	0.9999
beta[5]	7.0255	0.0244	1.1992	4.7217	6.2248	7.0196	7.8640	9.2751	2410	1.0012
beta[6]	0.6885	0.0169	0.9114	-1.0881	0.1024	0.7010	1.3271	2.4421	2913	1.0001
tau	0.0073	0.0000	0.0009	0.0058	0.0067	0.0073	0.0079	0.0091	3582	0.9992

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
BetaMoreThan0[1]	0.4145	0.0099	0.4927	0	0	0	1	1	2464	0.9997
BetaMoreThan0[2]	0.9468	0.0043	0.2246	0	1	1	1	1	2736	1.0001
BetaMoreThan0[3]	0.2642	0.0082	0.4410	0	0	0	1	1	2882	1.0001
BetaMoreThan0[4]	0.1350	0.0063	0.3418	0	0	0	0	1	2916	0.9997
BetaMoreThan0[5]	1.0000	NaN	0.0000	1	1	1	1	1	NaN	NaN
BetaMoreThan0[6]	0.7735	0.0076	0.4186	0	1	1	1	1	3004	0.9995

We may consider `Attitude.Reduce.Bill` for removal if we were to continue with a step-wise selection process.

(g) Justify your final choice of model. For example, you can present posterior inference for regression parameters and for sub-population means in appropriately designed tables or figures.

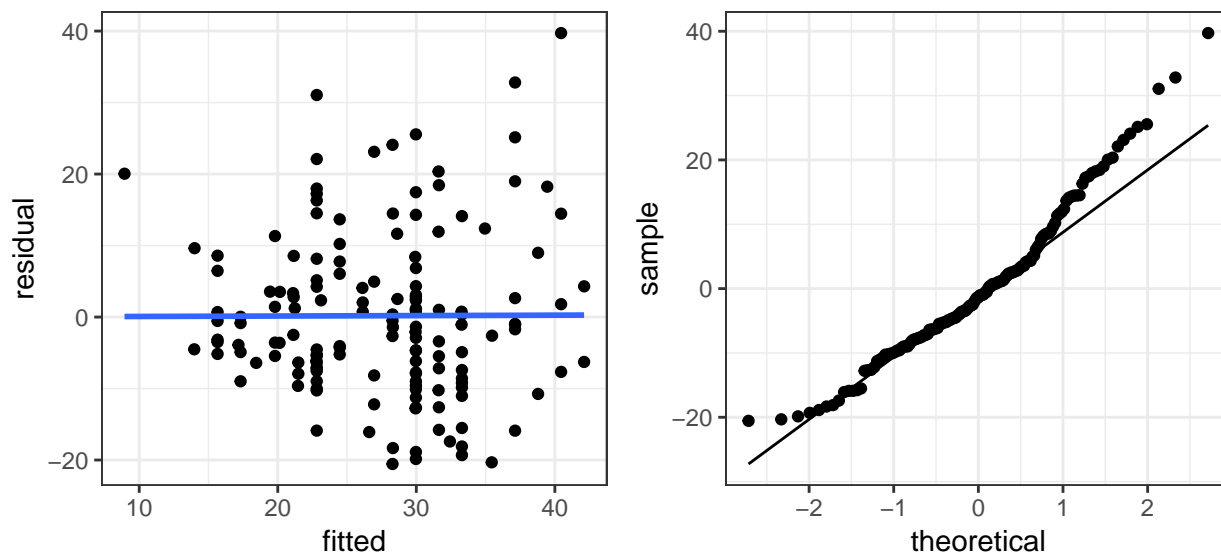
Model	LPML	WAIC
All	-591.1	1182
No Age	-589.9	1180
No Edu	-589.8	1179
Neither	-588.9	1178

While all models achieve roughly equivalent LPML and WAIC values, we choose the neither model as our preferred model because it is the simplest and achieves the highest LPML and lowest WAIC. Posterior inference for betas is available above.

Posterior inference for the predicted values of 5 randomly selected points are presented below, along with the actual observed y value. In 4 of the 5 samples, the actual y is captured in the 50% credible interval.

Age	Res	Bill	Env.	Bed	Ed	y	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%
3	2	1	1	3	5	15.270	22.77	0.1968	11.93	-0.8573	14.681	22.71	31.00	46.21
2	2	1	2	2	4	9.457	13.95	0.1946	12.25	-9.8823	5.677	14.06	22.26	38.08
3	3	1	1	3	2	20.409	24.39	0.1825	11.81	1.1469	16.537	24.23	32.32	47.40
5	2	1	1	5	5	36.152	37.24	0.1892	11.95	13.7491	29.118	37.29	45.07	61.24
3	2	2	3	5	5	15.023	32.51	0.1928	12.06	8.8495	24.264	32.70	40.51	56.22

Residual plot for the posterior means are is presented below:



The upper right of the qq plot may be cause concern, but, overall, we are satisfied with our residual plot.

(h) Based on your analysis, is the level of education associated with energy consumption? Motivate your answer, using Bayesian hypothesis testing.

We conduct a Bayesian hypothesis test using the pseudo Bayes Factor to compare the full model to the model without education.

The calculated 2LBF is 2.6288. Because this value is near 2, the data weakly supports the hypothesis that education is associated with energy consumption.

2. Now consider a model that contains the following predictors: Age, number of residents, level of education and the interaction between level of education and number of residents.

(a) This is a methodological question, with a data analysis question: Discuss the log marginal pseudo likelihood (LPML) as a criterion for model selection. What are the main features, advantages and possible disadvantages of this model selection criterion? Run a regression model with the interaction variable versus a model without. Compare the two models using the LPML criterion.

The log pseudo-marginal likelihood criteria uses pseudo-marginal likelihoods, instead of marginal likelihoods, which may be impossible or impractical to compute. This criterion allows for evaluation of a model through the model's predictive accuracy. Calculating the components of the LMPL are much easier with MCMC sampling, one of the biggest advantages of this criteria. This criterion is very easily implemented in model, but may still be computationally expensive. Models with the higher LPML are preferred. This method only allows for comparison between nested models, which is a disadvantage.

No Interaction Model

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	13.7772	0.1481	6.6172	0.5028	9.4815	13.8019	18.2170	26.6042	1996	0.9997
beta[2]	-0.0196	0.0167	0.8857	-1.7432	-0.6236	-0.0069	0.5676	1.7482	2810	0.9996
beta[3]	2.9632	0.0191	1.0997	0.8150	2.2348	2.9704	3.6689	5.1441	3313	0.9992
beta[4]	1.5447	0.0209	1.0237	-0.4754	0.8775	1.5436	2.2260	3.5366	2400	1.0000
tau	0.0058	0.0000	0.0007	0.0046	0.0054	0.0058	0.0063	0.0073	3265	1.0002

Interaction Model

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	25.1130	0.4174	12.9789	0.2359	16.1223	25.2722	34.0598	51.0139	967.1	1.0013
beta[2]	-0.1335	0.0187	0.8570	-1.8200	-0.7179	-0.1393	0.4386	1.5271	2111.6	1.0006
beta[3]	-1.1688	0.1340	4.2979	-9.7153	-4.1339	-1.0711	1.7613	7.1391	1028.4	1.0004
beta[4]	-1.1583	0.0936	2.9013	-6.8158	-3.1526	-1.1196	0.8472	4.3187	961.7	1.0011
beta[5]	1.0339	0.0326	1.0430	-0.9409	0.3097	1.0210	1.7462	3.0739	1024.6	1.0004
tau	0.0059	0.0000	0.0007	0.0046	0.0054	0.0059	0.0063	0.0073	2088.2	0.9995

Model	LPML	WAIC
Interaction	-606.4	1213
No Interaction	-605.9	1212

Based on the LMPL, the No Interaction model is preferred.

(b) This is a methodological question, with a data analysis question: Discuss the Bayes Factor as a criterion for model selection. What are the main features, advantages and possible disadvantages of this model selection criterion? What rule of thumb has been suggested to determine strength of evidence in favor of one model by using the Bayes Factor? Run a regression model with the interaction variable versus a model without. Compare the two models using the Bayes Factor.

Bayes Factors allow for comparisons of models that may not share the same parameters. This is a distinct advantage over the LPML criterion, which only allowed for the comparison of nested models. The Bayes factor between 2 models is derived from the posterior odds. After the application of Bayes' Theorem, Bayes Factor takes the form, $\frac{p(y|M_0)}{p(y|M_1)}$. Multiplying the Bayes Factor by the prior odds, yields the posterior odds. More commonly, 2 times the log of the Bayes Factor is used to compare two models. Using 2LBF, the strength of evidence towards one model/hypothesis can be determined. The LBF is also more stable and interpretable when the BF is very large or very small.

In many cases it is easy to calculate the regular Bayes Factor from posterior samples, but we can also use LPML of two models to calculate a pseudo Bayes Factor. One advantage of Bayes Factors is its transitivity. Knowing how Model A compares to Model B and how Model B compares to Model C allows us to determine how Model A compares to Model C. Bayes Factors also have some disadvantages. They cannot be used with improper priors. Additionally, use of the Bayes Factor implicitly assumes that one of the two models is correct. Bayes Factors can also be computationally expensive.

Based on our calculated 2LBF of 1.1143, we conclude that the interaction model is not worth considering. Thus, we prefer the no interaction model.

(c) Discuss a sensitivity analysis on the best model you have selected from the previous two points (with or without interaction). Comment on the results from the sensitivity analysis.

In this section, we compare different priors: $\tau \sim \Gamma(0.01, 0.01)$, $\tau \sim \Gamma(0.001, 0.001)$ and $\tau \sim \Gamma(0.0001, 0.0001)$.

$\tau \sim \Gamma(0.01, 0.01)$

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	13.7710	0.1460	6.5625	1.3785	9.2306	13.7536	18.2550	26.4316	2021	1.0036
beta[2]	0.0014	0.0179	0.8999	-1.7992	-0.5831	0.0047	0.5986	1.7092	2527	1.0016
beta[3]	2.9498	0.0193	1.1232	0.7843	2.2081	2.9359	3.6799	5.2097	3398	1.0012
beta[4]	1.5422	0.0202	0.9931	-0.3765	0.8582	1.5413	2.2228	3.4785	2428	1.0014
tau	0.0059	0.0000	0.0007	0.0046	0.0054	0.0058	0.0063	0.0072	3327	0.9995

$\tau \sim \Gamma(0.001, 0.001)$

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	13.7772	0.1481	6.6172	0.5028	9.4815	13.8019	18.2170	26.6042	1996	0.9997
beta[2]	-0.0196	0.0167	0.8857	-1.7432	-0.6236	-0.0069	0.5676	1.7482	2810	0.9996
beta[3]	2.9632	0.0191	1.0997	0.8150	2.2348	2.9704	3.6689	5.1441	3313	0.9992
beta[4]	1.5447	0.0209	1.0237	-0.4754	0.8775	1.5436	2.2260	3.5366	2400	1.0000
tau	0.0058	0.0000	0.0007	0.0046	0.0054	0.0058	0.0063	0.0073	3265	1.0002

$\tau \sim \Gamma(0.0001, 0.0001)$

Variable	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	13.7772	0.1481	6.6172	0.5028	9.4815	13.8019	18.2170	26.6042	1996	0.9997
beta[2]	-0.0196	0.0167	0.8857	-1.7432	-0.6236	-0.0069	0.5676	1.7482	2810	0.9996
beta[3]	2.9632	0.0191	1.0997	0.8150	2.2348	2.9704	3.6689	5.1441	3313	0.9992
beta[4]	1.5447	0.0209	1.0237	-0.4754	0.8775	1.5436	2.2260	3.5366	2400	1.0000
tau	0.0058	0.0000	0.0007	0.0046	0.0054	0.0058	0.0063	0.0073	3265	1.0002

Posterior estimates for each beta are approximately equal in all cases. In some cases, the posterior mean for Beta 2 is positive, while in others, it is negative. This is not concerning, as the 95% credible intervals are approximately equal.

(d) Now consider a household with median age, number of residents, and level of education. What is the median level of energy consumption and a range of values that you can predict for this household based on that information? How would you change your answer if the household had four residents instead?

Data	Age	Resident	Education	2.5%	50%	97.5%
Median	4	2	4	0.3346	25.99	52.43
Modified	4	4	4	5.5382	31.78	57.96

The estimated median is given by the 50% column. A reasonable range of values is given by the 2.5% and 97.5% columns. We note that the estimated consumption for the median house is about 6 units less than the estimated consumption for the house with 4 residents.

3. Now we want to describe the patterns of household electricity consumption over time, that is considering all the daily consumption data. In this regard, we want to propose a hierarchical model that is able to conduct inference about both population level usage as well as the household level usage.

(a) Propose (and motivate) a model (in formulas!) to describe the daily consumption pattern as a function of the predictors.

We will use a hierarchical linear regression model to describe daily consumption methods.

$$Y_i^j | \underline{\beta}^j, \tau^j \sim N(X_i \underline{\beta}^j, \tau^j)$$

where $i = 1:151$, $j = 1:121$, X_i is the row of the predictor matrix corresponding to the i^{th} house, $\underline{\beta}^j$ is the vector of coefficients for the j^{th} day, and τ^j is the precision for the j^{th} day.

We choose to use g -priors, with $g = n$, on $\underline{\beta}^j$ and vague gamma priors on τ^j . This allows us to obtain a separate regression line for each day, with its own precision. The following are the distributions:

$$\underline{\beta}^j | \tau^j \sim N\left(\underline{\beta}_p, \frac{n}{\tau^j} CO^{-1}\right) \quad \text{and} \quad \tau^j \sim \text{Gamma}(0.001, 0.001)$$

where $\underline{\beta}_p$ is a population level regression line and CO^{-1} is the covariance matrix for the predictors.

The model contains a separate regression line for each day and a regression line for the population. The last step is to put priors on the population level regression line. We again use g -priors for the population level regression coefficients and a vague prior on the precision at the population level.

$$\underline{\beta}_p | \tau_p \sim N\left(\underline{0}, \frac{n}{\tau_p} CO^{-1}\right) \quad \text{and} \quad \tau_p \sim \text{Gamma}(0.001, 0.001)$$

This allows for population level inference as well as daily inference. We will be able to obtain predictive posterior distributions for each household at both the daily level and the population level (overall).

(b) Run the previous model in Jags and provide a plot to describe the time-varying effect of the number of residents on the daily consumption both at the population level and also for households # 30, # 83 and # 91, and comment on their time patterns. Motivate and comment on your inference. (I believe you should be able to propose a model to conduct at least population level inference. I would recognize partial credit so solutions that allow only population level inference. In any event, credit will be recognized only if you appropriately justify your modeling choices)

Our stan model is presented below:

```
data {
  int<lower=0> N;    // number of data items
  int<lower=0> K;    // number of predictors
  int<lower=0> D;    // number of days
  matrix[N, K] x;   // predictor matrix
  matrix[N,D] y;    // outcome vector

  int<lower=0> M;    // number of new households
  matrix[M,K] x_new; // predictor matrix

  vector[K] beta_0_pop; // beta prior mean for population
  matrix[K, K] inv_covariance; // covariance matrix
  real<lower=0> a;
  real<lower=0> b;
}

parameters {
  vector[K] beta_pop; // coefficients for intercept & predictors at population level
  matrix[K,D] beta_day; //coefficients for intercept & predictors at day level
  real<lower=0> tau_pop; // error scale
  real<lower=0> tau_day[D];
}

model {
  tau_pop ~ gamma(a, b);
  tau_day ~ gamma(a,b);

  beta_pop ~ multi_normal(beta_0_pop, (N*inv_covariance)/tau_pop);

  for (d in 1:D){
    beta_day[,d] ~ multi_normal(beta_pop, (N*inv_covariance)/tau_day[d]);
    y[,d] ~ normal(x * beta_day[,d], 1 / sqrt(tau_day[d])); // likelihood
  }
}

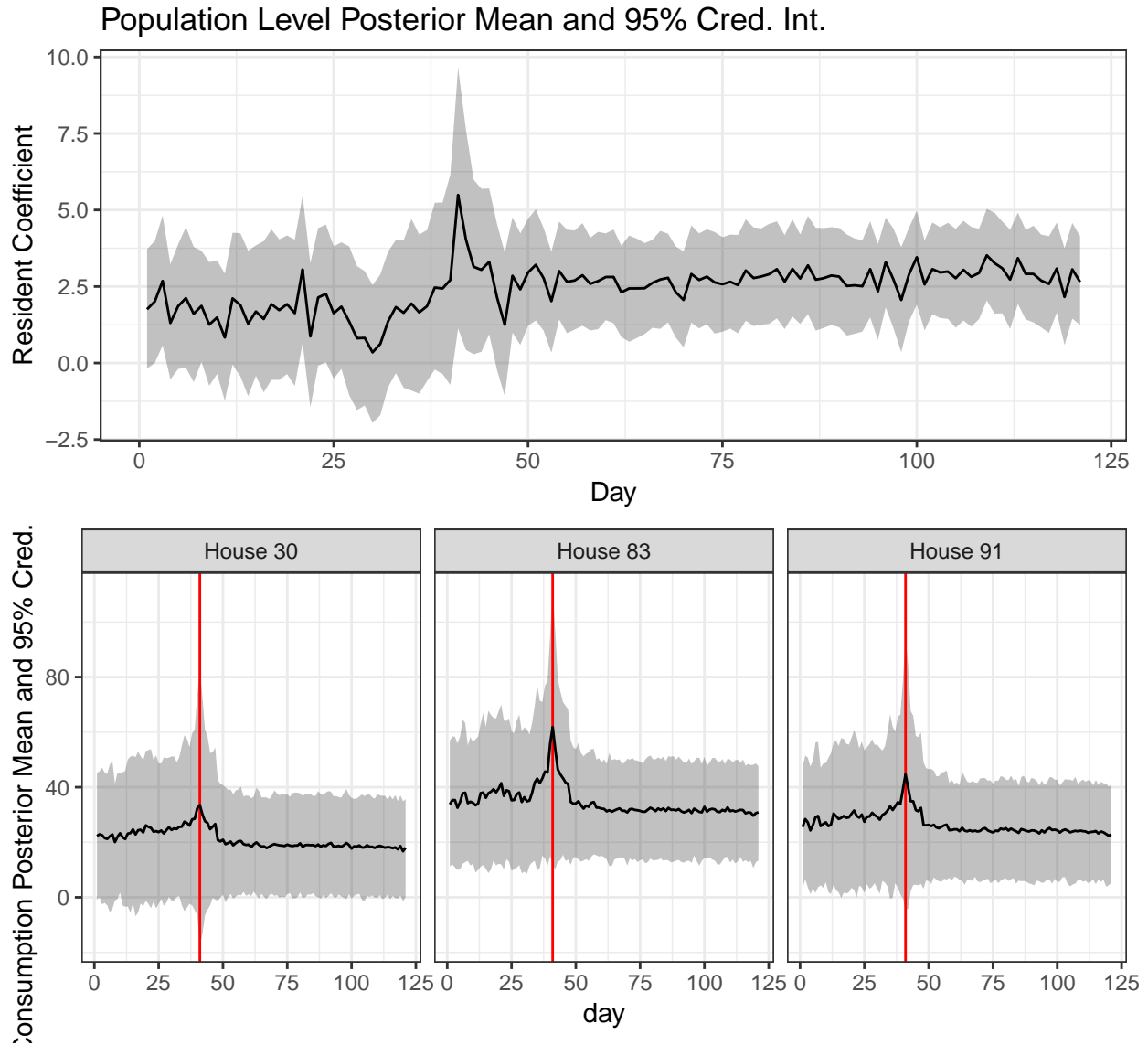
generated quantities{
  matrix[M,D] ynew;
  vector[M] ynew_p;

  for(d in 1:D){
    for(i in 1:M) {
      ynew[i,d] = normal_rng(x_new[i,] * beta_day[,d], 1 / sqrt(tau_day[d]));
    }
  }
}
```

```

    }
  }
  for(i in 1:M) {
    ynew_p[i] = normal_rng(x_new[i,] * beta_pop, 1 / sqrt(tau_pop));
  }
}

```



Aside from a peak at day 41 (Christmas - indicated by a red vertical line), the estimate of the Resident coefficient appears relatively stable, and oscillates around 2.

For each house, there is a clear peak Consumption on Christmas. Aside from this, the consumption levels appear to be relatively stable, with a possible decrease after the new year. We note that some of our 95% credible intervals fall below 0 consumption in House 30 and House 91.

4. On January 1, 2010, there was a major change in the tariff structure on electricity consumption. Is there a difference in energy consumption pattern before versus after the policy change? During the 2.5 months after the change does the pattern of energy consumption at the population level tend to return to the pattern before the policy change? Hint: there are many ways to answer this question. Some are OK, some are not. Please, a) justify your modeling choices b) Bayesian hypothesis testing is always rooted on probabilistic statements c) be brief, to the point but also make sure I can follow what you are doing. As always, credit will be recognized only if you appropriately justify your choices.

We will use a hierarchical linear regression model to describe daily consumption methods before and after January 1 when the new tariffs were implemented.

The Y 's (consumption) follow the distribution:

$$Y_i^j | \underline{\beta}^j, \tau^j \sim N(X_i \underline{\beta}^j, \tau^j)$$

where $i=1:151$, $j=1:121$, X_i is the row of the predictor matrix corresponding to the i^{th} house, $\underline{\beta}^j$ is the vector of coefficients for the j^{th} day, and τ^j is the precision for the j^{th} day.

Now we need prior distributions on $\underline{\beta}^j$ and τ^j . We choose to use g -priors, with $g = n$, on the $\underline{\beta}^j$ and vague gamma priors on τ^j . This allows us to obtain a separate regression line for each day, with its own precision. This time we break up the days into two time intervals; before and after January 1st. The following are the distributions for the day level β 's and the day level τ 's:

For $j=1:47$:

$$\underline{\beta}^j | \tau^j \sim N(\underline{\beta}_{BJ1}, \frac{n}{\tau^j} CO^{-1}) \quad \text{and} \quad \tau^j \sim \text{Gamma}(0.001, 0.001)$$

where $\underline{\beta}_{BJ1}$ is a regression line for days before January 1st and CO^{-1} is the covariance matrix for the predictors.

For $j=48:121$:

$$\underline{\beta}^j | \tau^j \sim N(\underline{\beta}_{AJ1}, \frac{n}{\tau^j} CO^{-1}) \quad \text{and} \quad \tau^j \sim \text{Gamma}(0.001, 0.001)$$

where $\underline{\beta}_{AJ1}$ is a regression line for days after January 1st and CO^{-1} is the covariance matrix for the predictors.

Now we have a model with a separate regression line for each day, and a regression line for each time period. Now each time period should come from an overall population level β . This allows us to connect the two time periods together. We also consider different precisions for each time period. The following are the distributions on the two sets of parameters for before and after January 1st:

$$\underline{\beta}_{BJ1} | \tau_{BJ1} \sim N\left(\underline{\beta}_p, \frac{n}{\tau_{BJ1}} CO^{-1}\right) \quad \text{and} \quad \tau_{BJ1} \sim \text{Gamma}(0.001, 0.001)$$

where $\underline{\beta}_p$ is a regression line for all days and CO^{-1} is the covariance matrix for the predictors.

and

$$\underline{\beta}_{AJ1} | \tau_{AJ1} \sim N\left(\underline{\beta}_p, \frac{n}{\tau_{AJ1}} CO^{-1}\right) \quad \text{and} \quad \tau_{AJ1} \sim \text{Gamma}(0.001, 0.001)$$

where $\underline{\beta}_p$ is a regression line for all days and CO^{-1} is the covariance matrix for the predictors.

Finally, we put priors on the population level regression line. We again use g -priors for the population level regression coefficients and a vague prior on the precision at the population level.

$$\underline{\beta}_p | \tau_p \sim N\left(\underline{0}, \frac{n}{\tau_p} CO^{-1}\right) \quad \text{and} \quad \tau_p \sim \text{Gamma}(0.001, 0.001)$$

This allows for population level inference as well as daily inference. We now can investigate differences in regression lines for the two time periods. We are able to obtain predictive posterior distributions for each household as well at the daily level, time period level, and the population level (overall).

Our stan model is presented below

```
data {
  int<lower=0> N;    // number of data items
  int<lower=0> K;    // number of predictors
  int<lower=0> D;    // number of days
  int<lower=0> T_n;  // number of time intervals
  matrix[N, K] x;   // predictor matrix
  matrix[N,D] y;    // outcome vector
  int<lower=0> D_before;

  vector[K] beta_0_pop; // beta prior mean for population
  matrix[K, K] inv_covariance; // covariance matrix
  real<lower=0> a;
  real<lower=0> b;
}

parameters {
  vector[K] beta_pop; // coefficients for intercept & predictors at population level
  matrix[K,T_n] beta_ba; // coefficients for the two time intervals
  matrix[K,D] beta_day; //coefficients for intercept & predictors at day level
  real<lower=0> tau_pop;
  real<lower=0> tau_ba[T_n];
  real<lower=0> tau_day[D];
}

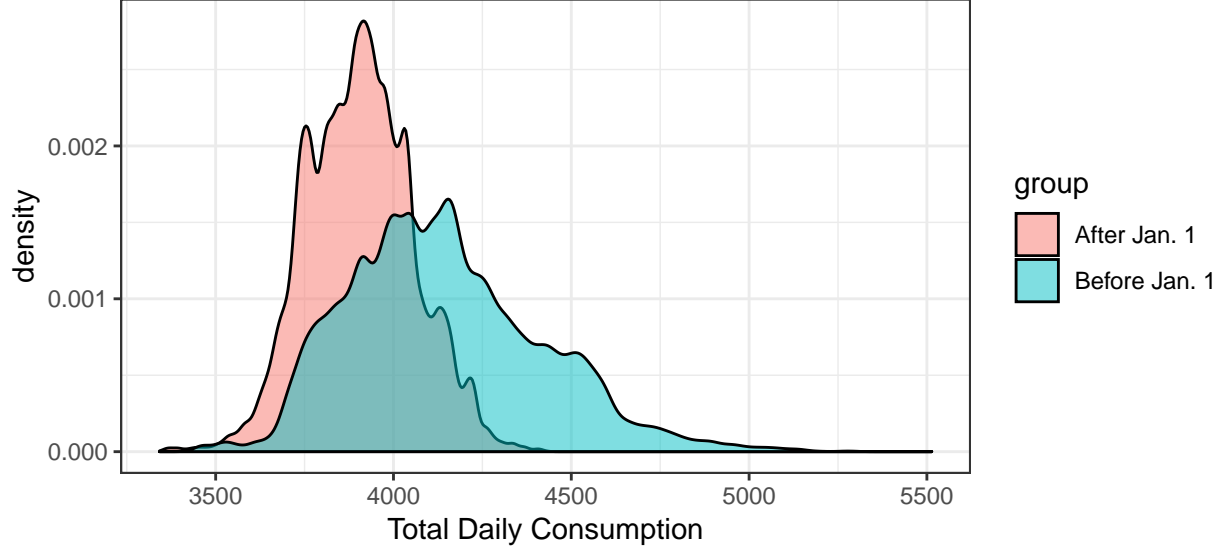
model {
  tau_pop ~ gamma(a, b);
  tau_day ~ gamma(a,b);
  tau_ba ~ gamma(a,b);

  beta_pop ~ multi_normal(beta_0_pop, (N*inv_covariance)/tau_pop);

  for (t in 1:T_n){
    beta_ba[,t] ~ multi_normal(beta_pop, (N*inv_covariance)/tau_ba[t]);
  }

  for (d in 1:D_before){
    beta_day[,d] ~ multi_normal(beta_ba[,1], (N*inv_covariance)/tau_day[d]);
    y[,d] ~ normal(x * beta_day[,d], 1 / sqrt(tau_day[d])); // likelihood
  }

  for (j in (D_before+1):D){
    beta_day[,j] ~ multi_normal(beta_ba[,2], (N*inv_covariance)/tau_day[j]);
    y[,j] ~ normal(x * beta_day[,j], 1 / sqrt(tau_day[j])); // likelihood
  }
}
```



Group	mean	sd	2.5%	50%	97.5%
Before	4143	271.8	3710	4114	4740
After	3906	150.0	3630	3906	4211

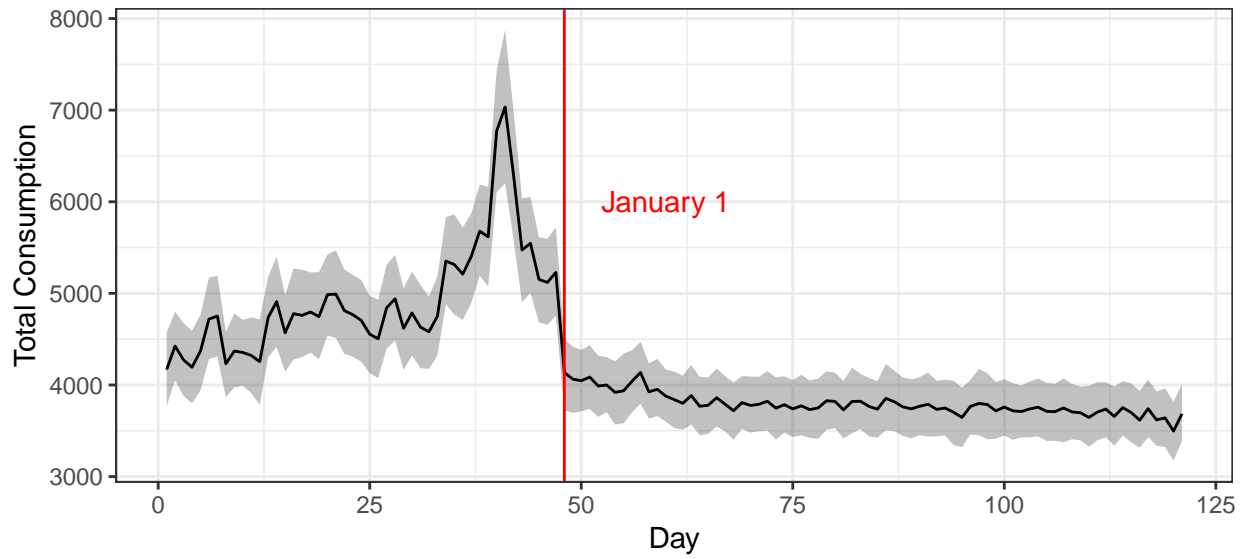
To evaluate whether or not there was a change in daily consumption after the tariff change, we examined the predictive posterior distributions on total daily consumption. To calculate the estimated total daily consumption, before and after the tariff, we used the before and after regression lines to estimate the daily consumption for each house. We then summed over all houses for each day in both groups to obtain a sample of total daily consumption. The distributions are compared in the plot and table above. Based on these, we can see that consumption is typically higher for the days before Jan 1. We also conducted a test to find the probability that the posterior mean before Jan 1 is higher than the posterior mean after Jan 1:

$$H_0 : \mu_{before} = \mu_{after} \quad H_a : \mu_{before} > \mu_{after}$$

We found the probability ($P(\mu_{before} > \mu_{after} | data)$) to be 0.7708. This is high enough to conclude that the total daily consumption of all houses was higher before the tariff than after.

We can also see a difference in the pattern of energy usage before and after the change. The total daily consumption varies day by day prior to January 1st ($sd = 272$), while the day-to-day levels are more consistent after January 1st ($sd = 150$). This is also visible in the plots above and below.

Thus, there is clearly a change in daily consumption following the tariff change. Now, we aim to evaluate whether or not energy consumption eventually returned to its original level. The following plot shows the total daily consumption for all houses with the red line indicating January 1st.



We note that after January 1st the total daily consumption across all houses clearly drops and does not return to its previous level. This change may be due to the end of the holiday season. The spike in consumption on Christmas day is evidence that the holidays may be related to energy consumption. The warming weather may also account for the difference in energy consumption. Because of these possible confounding factors, this data may not be suitable for investigating the affect of the tariff. The question would be better answered by historical data from prior years, so we could evaluate whether or not the change is due to the tariff or one of the possible confounding variables.