

# PHP 2530 Assignment #3

Anthony Sisti

3/18/2020

## Question 5.3

a) We use methods demonstrated in class to sample from the posterior distribution of  $\theta$  for each school. Once we have samples from the posterior, we compute the probability that each school was the “best” of the eight by assessing the frequency of having the largest  $\theta$  value within each sample. A table of results is shown below.

Table 1: Posterior Probability of being the 'Best' School

|             | A     | B     | C     | D     | E     | F     | G     | H     |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Probability | 0.257 | 0.106 | 0.089 | 0.096 | 0.054 | 0.068 | 0.192 | 0.138 |

Using the same set of posterior  $\theta$  values, we compute the probability that each school had a larger positive effect than the others.

Table 2: Posterior Probability of each School Being Better than Another

| School | Probability of being better than |       |       |       |       |       |       |       |
|--------|----------------------------------|-------|-------|-------|-------|-------|-------|-------|
|        | A                                | B     | C     | D     | E     | F     | G     | H     |
| A      | -                                | 0.639 | 0.678 | 0.647 | 0.735 | 0.702 | 0.531 | 0.613 |
| B      | 0.361                            | -     | 0.565 | 0.516 | 0.624 | 0.577 | 0.385 | 0.478 |
| C      | 0.322                            | 0.435 | -     | 0.445 | 0.549 | 0.508 | 0.338 | 0.424 |
| D      | 0.353                            | 0.484 | 0.555 | -     | 0.608 | 0.563 | 0.368 | 0.472 |
| E      | 0.265                            | 0.376 | 0.451 | 0.392 | -     | 0.453 | 0.271 | 0.369 |
| F      | 0.298                            | 0.423 | 0.492 | 0.437 | 0.547 | -     | 0.318 | 0.408 |
| G      | 0.469                            | 0.615 | 0.662 | 0.632 | 0.729 | 0.682 | -     | 0.582 |
| H      | 0.387                            | 0.522 | 0.576 | 0.528 | 0.631 | 0.592 | 0.418 | -     |

b) When we set  $\tau = \infty$ , we are estimating each of schools' means separately, meaning  $\theta_j \sim N(\bar{y}_j, \sigma_j^2)$ . In order to obtain the probability that any one school has the largest positive effect on scores, we can sample  $\theta$ 's for each school and examine the frequency that each school has the largest  $\theta$  values in each sample. A table of the results can be found below.

Table 3: Posterior Probability of being the 'Best' School (Independent Estimates)

|             | A     | B     | C     | D     | E     | F     | G    | H     |
|-------------|-------|-------|-------|-------|-------|-------|------|-------|
| Probability | 0.555 | 0.034 | 0.026 | 0.033 | 0.004 | 0.011 | 0.17 | 0.166 |

Then, we can compute  $P(\theta_j > \theta_i)$  for each  $\theta_i$  to determine the probability that school  $j$  has a larger positive effect than the others. From our sample of  $\theta$ 's we compare each school to the others to obtain the following.

Table 4: Posterior Probability of each School Being Better than Another (Independent Estimates)

| School | Probability of being better than |       |       |       |       |       |       |       |
|--------|----------------------------------|-------|-------|-------|-------|-------|-------|-------|
|        | A                                | B     | C     | D     | E     | F     | G     | H     |
| A      | -                                | 0.864 | 0.922 | 0.885 | 0.948 | 0.931 | 0.714 | 0.758 |
| B      | 0.136                            | -     | 0.71  | 0.527 | 0.729 | 0.686 | 0.24  | 0.417 |
| C      | 0.078                            | 0.29  | -     | 0.311 | 0.457 | 0.442 | 0.141 | 0.27  |
| D      | 0.115                            | 0.473 | 0.689 | -     | 0.693 | 0.663 | 0.225 | 0.396 |
| E      | 0.052                            | 0.271 | 0.543 | 0.307 | -     | 0.475 | 0.095 | 0.267 |
| F      | 0.069                            | 0.314 | 0.558 | 0.337 | 0.525 | -     | 0.129 | 0.284 |
| G      | 0.286                            | 0.76  | 0.859 | 0.775 | 0.905 | 0.871 | -     | 0.621 |
| H      | 0.242                            | 0.583 | 0.73  | 0.604 | 0.733 | 0.716 | 0.379 | -     |

c) When we conduct the standard Bayesian analysis, there appears to be a shrinkage effect toward an overall mean value when compared to independent estimation. When considering the probabilities of each school having the most positive effect, we notice that schools B, C, D, E and F all have slightly larger probabilities of being the best in the standard Bayesian analysis compared to the independent estimation, while the probability that A is the best is reduced by more than half. When we consider the probabilities of one school being better than another, we see a similar pattern. In the independent analysis, school A is better than school E with a probability of nearly 0.95, while the standard Bayesian approach tells us that school A is better than school E with probability 0.735. The standard Bayesian method appears to have taken the extreme values and shrunk them toward more reasonable estimates, accounting for similarities between schools that the independent estimation could not.

d) If  $\tau = 0$ , we are saying that there is no variability between schools, and each  $\theta_j$  comes from the same distribution. In this case, the probability of any one of the schools being the best is  $1/8$ . Also, the probability that any one school is better than any other school is  $1/2$ . These values come from the properties of a random sample from the same distribution.

## Question 5.4

a) We know that half the of the  $\theta$ 's are from a  $N(1, 1)$  and half are from  $N(-1, 1)$ . Since they are independently sampled, the true joint density we would have the form

$$\prod_{i=1}^{J_{\mu=1}} N(\theta_i | 1, 1) \prod_{k=1}^{J_{\mu=-1}} N(\theta_k | -1, 1)$$

Where  $J_{\mu=1}$  and  $J_{\mu=-1}$  represent the number of  $\theta$ 's that are from the  $N(1, 1)$  and  $N(-1, 1)$  respectively. We also let  $N(\theta | \mu, 1)$  denote the normal density with mean  $\mu$  and variance 1. We have to take into account the fact that we do not know which set of  $\theta$ 's come from which distribution. In order to do this, we need to take

into account each possible partition of the  $2J$   $\theta$ 's into two equal groups. The total number of partitions is then  $\binom{2J}{J}$ . Let  $J_{\mu=1}^q$  be the set of  $J$  theta values that were selected to be from  $N(1, 1)$  distribution in the  $q^{th}$  partition, similarly for  $J_{\mu=-1}^q$ . Then we can write the density as,

$$\frac{\sum_{q=1}^{\binom{2J}{J}} \left( \prod_{i \in J_{\mu=1}^q} N(\theta_i | 1, 1) \prod_{k \in J_{\mu=-1}^q} N(\theta_k | -1, 1) \right)}{\binom{2J}{J}}$$

Since we have accounted for all possible partitions of the  $\theta$ 's into the two groups, swapping one  $\theta$  index with another will not influence the density function, it will maintain the same form. Given this fact, we see that the  $\theta_j$  values are exchangeable.

b) Let  $\mu$  be a vector with length  $2J$ . Assume half the entries are 1's and the other half are -1's, but each entry can take either value. This vector represents all the possible mean values that can be assigned to the  $\theta$ 's where

$$E[\theta_i | \mu_i] = \mu_i$$

By the law of total covariance, we have

$$\text{cov}(\theta_i, \theta_j) = E[\text{cov}(\theta_i, \theta_j | \mu)] + \text{cov}[E(\theta_i | \mu), E(\theta_j | \mu)]$$

Notice that,

$$E[\text{cov}(\theta_i, \theta_j | \mu)] = 0$$

since  $\theta_i$  and  $\theta_j$  are conditionally independent single draws from a  $N(\mu_i, 1)$  and  $N(\mu_j, 1)$  distribution respectively. We also have that

$$\begin{aligned} \text{cov}[E(\theta_i | \mu), E(\theta_j | \mu)] &= \text{cov}(\mu_i, \mu_j) \\ &= E[\mu_i \mu_j] - E[\mu_i]E[\mu_j] \\ &= E[\mu_i \mu_j] \\ &= 1 * \text{Pr}(u_i = 1 | u_j = 1) \text{Pr}(\mu_j = 1) + 1 * \text{Pr}(u_i = -1 | u_j = -1) \text{Pr}(\mu_j = -1) \\ &\quad + (-1) * \text{Pr}(u_i = 1 | u_j = -1) \text{Pr}(\mu_j = -1) + (-1) * \text{Pr}(u_i = -1 | u_j = 1) \text{Pr}(\mu_j = 1) \\ &= \left( \frac{J-1}{2J-1} \right) \left( \frac{1}{2} \right) + \left( \frac{J-1}{2J-1} \right) \left( \frac{1}{2} \right) - \left( \frac{J}{2J-1} \right) \left( \frac{1}{2} \right) - \left( \frac{J}{2J-1} \right) \left( \frac{1}{2} \right) \\ &= \frac{-1}{2J-1} \end{aligned}$$

Thus,  $\text{cov}(\theta_i, \theta_j) = \frac{-1}{2J-1} \neq 0$ , so this density cannot be written as a mixture of i.i.d components.

c) We cannot take the limit as  $J$  goes to infinity and get a counter example because, as is seen above, the covariance between  $\theta_i$  and  $\theta_j$  goes to zero in the limit. So as  $J$  nears infinity, the joint distribution will approach an i.i.d distribution which would satisfy deFinetti's theorem.

## Question 5.11

In order to determine the posterior density, we can write

$$\begin{aligned} p(\theta, \mu, \tau|y) &\propto p(y|\theta, \mu, \tau)p(\theta, \mu, \tau) \\ &\propto p(y|\theta, \mu, \tau)p(\theta|\mu, \tau)p(\mu, \tau) \\ &\propto p(y|\theta)p(\theta|\mu, \tau)p(\mu, \tau) \end{aligned}$$

From the problem, we know

$$p(y|\theta) \propto \prod_{i=1}^N \theta^{y_i} (1 - \theta)^{n_i - y_i}.$$

Since the problem is concerned with  $\theta$  rather than  $\text{logit}(\theta)$ , we have.

$$\begin{aligned} p(\theta|\mu, \tau) &\propto \prod_{i=1}^N p(\text{logit}^{-1}(\theta_i)|\mu, \tau) \left| \frac{d\text{logit}^{-1}(\theta_i)}{d\theta_i} \right| \\ &\propto \prod_{i=1}^N \frac{1}{\tau \theta_i (1 - \theta_i)} \exp \left\{ -\frac{1}{2\tau^2} (\text{logit}^{-1}(\theta_i) - \mu)^2 \right\} \end{aligned}$$

We then have

$$p(\theta, \mu, \tau|y) \propto p(\mu, \tau) \prod_{i=1}^N \theta^{y_i} (1 - \theta)^{n_i - y_i} \prod_{i=1}^N \frac{1}{\tau \theta_i (1 - \theta_i)} \exp \left\{ -\frac{1}{2\tau^2} (\text{logit}^{-1}(\theta_i) - \mu)^2 \right\}$$

b) Integral (5.4) in the context of this problem takes the form

$$\begin{aligned} p(\mu, \tau|y) &= \int p(\theta, \mu, \tau|y) d\theta \\ &= \int p(\mu, \tau) \prod_{i=1}^N \theta^{y_i} (1 - \theta)^{n_i - y_i} \prod_{i=1}^N \frac{1}{\tau \theta_i (1 - \theta_i)} \exp \left\{ -\frac{1}{2\tau^2} (\text{logit}^{-1}(\theta_i) - \mu)^2 \right\} d\theta_1 \dots d\theta_N \\ &= p(\mu, \tau) \prod_{i=1}^N \int \frac{\theta^{y_i - 1} (1 - \theta)^{n_i - y_i - 1}}{\tau} \exp \left\{ -\frac{1}{2\tau^2} (\text{logit}^{-1}(\theta_i) - \mu)^2 \right\} d\theta_1 \dots d\theta_N \end{aligned}$$

Although we have the product of integrals that look as though they could be of a familiar form, even with a  $u$ -substitution of  $u = \text{logit}^{-1}(\theta_i)$ , we still would not have the kernel of any known density. Therefore, we cannot determine a closed form for  $p(\mu, \tau|y)$  via this integral.

c) Expression (5.5) cannot help us here because in order to use it, as step (2) on page 108 states, we first need to analytically determine  $p(\theta|\mu, \tau, y)$ . However, since this is a non-conjugate model,  $p(\theta|\mu, \tau, y)$  has no closed form expression without resorting to proportionality. We ideally would want to write,

$$p(\mu, \tau|y) \propto \frac{p(\theta, \mu, \tau|y)}{p(\theta|\mu, \tau, y)}$$

But since we cannot obtain the denominator in its complete, closed form, we cannot substitute it into (5.5) to determine  $p(\mu, \tau|y)$ . We would need to determine all components of  $p(\theta|\mu, \tau, y)$  to ensure  $\frac{p(\theta, \mu, \tau|y)}{p(\theta|\mu, \tau, y)}$  is the correct function of  $\mu$  and  $\tau$ .

## Question 5.13

a) Since we are dealing with binomial data, where our binomial parameter is distributed as a Beta, we have that

$$\begin{aligned} y_j|\theta_j &\sim \text{Binom}(n_j, \theta_j) \\ \theta_j|\alpha, \beta &\sim \text{Beta}(\alpha, \beta) \end{aligned}$$

Setting a non-informative prior as in the rat tumor example of section 5.3 we have

$$p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$$

We can then write the joint posterior as

$$\begin{aligned} p(\theta, \alpha, \beta|y) &\propto p(y|\theta, \alpha, \beta)p(\theta|\alpha, \beta)p(\alpha, \beta) \\ &\propto (\alpha + \beta)^{-5/2} \prod_{j=1}^{10} \theta_j^{y_j} (1 - \theta_j)^{n_j - y_j} \prod_{j=1}^{10} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1} (1 - \theta_j)^{\beta-1} \\ &\propto (\alpha + \beta)^{-5/2} \prod_{j=1}^{10} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha+y_j-1} (1 - \theta_j)^{\beta+n_j-y_j-1} \end{aligned}$$

b) We can rewrite the joint posterior density as follows.

$$\begin{aligned} p(\theta, \alpha, \beta|y) &\propto (\alpha + \beta)^{-5/2} \prod_{j=1}^{10} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha+y_j-1} (1 - \theta_j)^{\beta+n_j-y_j-1} \\ &\propto (\alpha + \beta)^{-5/2} \prod_{j=1}^{10} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)}{\Gamma(\alpha + \beta + n_j)} \prod_{j=1}^{10} \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)} \theta_j^{\alpha+y_j-1} (1 - \theta_j)^{\beta+n_j-y_j-1} \\ &\propto p(\alpha, \beta|y)p(\theta|\alpha, \beta, y) \end{aligned}$$

Based on this formulation, and since we can write  $p(\theta|\alpha, \beta, y)$  exactly, we have

$$p(\theta|\alpha, \beta, y) = \prod_{j=1}^{10} \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)} \theta_j^{\alpha+y_j-1} (1 - \theta_j)^{\beta+n_j-y_j-1}$$

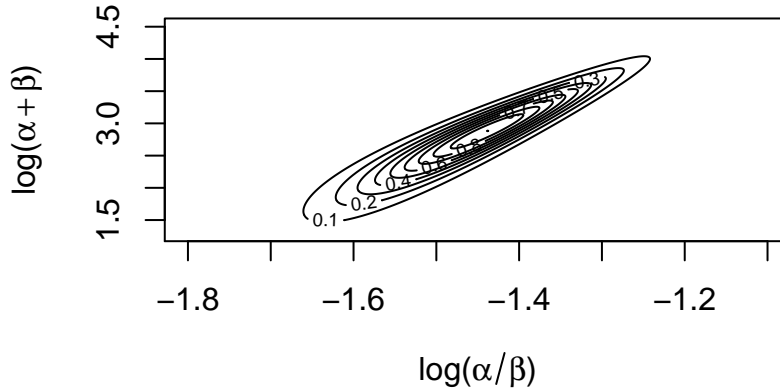
$$p(\alpha, \beta|y) \propto (\alpha + \beta)^{-5/2} \prod_{j=1}^{10} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)}{\Gamma(\alpha + \beta + n_j)}$$

We first sample over the marginal posterior distribution for the hyper-parameters. As is advised in the textbook, we simulated 1000 draws of  $(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta))$  using the prior on the natural transformed scale, where

$$p(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta)) \propto \alpha\beta(\alpha + \beta)^{-5/2}$$

Our grid is centered around the method of moments estimates for  $(\alpha, \beta)$ , which are 2.89 and 11.86 respectively. On the natural transformed scale, this corresponds to  $(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta))$  being -1.41 and 2.69. Below we plot a contour of the marginal posterior density.

Marginal Posterior of  $\alpha$  and  $\beta$  on Transformed Scale



We numerically sample over our grid to obtain 3000 pairs of  $(\alpha, \beta)$ . For each of these pairs, we sample one value for  $\theta_j$ ,  $j=1, \dots, 10$ . This leaves us with 3000 values for each  $\theta_j$  sampled from their respective Beta distributions.

Table 5: Observed Proportions Against Posterior  $\theta$  distributions

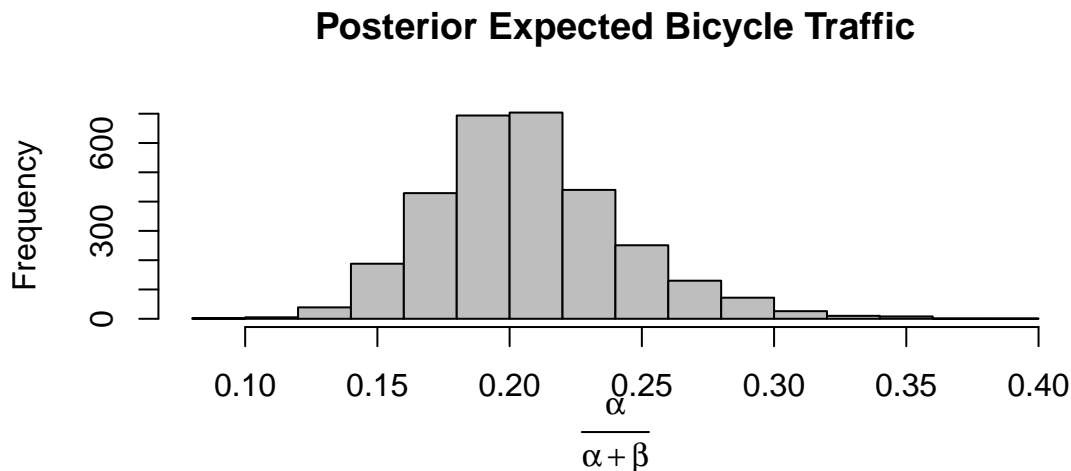
| Location $j$ | Observed Porportion | Posterior Distribution Summary |       |       |       |       |       |       |
|--------------|---------------------|--------------------------------|-------|-------|-------|-------|-------|-------|
|              |                     | Mean                           | SD    | 0.05  | 0.25  | 0.5   | 0.75  | 0.95  |
| 1            | 0.216               | 0.214                          | 0.043 | 0.148 | 0.184 | 0.212 | 0.241 | 0.287 |
| 2            | 0.091               | 0.107                          | 0.029 | 0.064 | 0.085 | 0.105 | 0.126 | 0.157 |
| 3            | 0.172               | 0.178                          | 0.044 | 0.110 | 0.147 | 0.176 | 0.206 | 0.256 |
| 4            | 0.186               | 0.187                          | 0.041 | 0.124 | 0.159 | 0.185 | 0.213 | 0.259 |
| 5            | 0.156               | 0.161                          | 0.031 | 0.112 | 0.139 | 0.159 | 0.180 | 0.213 |
| 6            | 0.260               | 0.250                          | 0.045 | 0.180 | 0.218 | 0.248 | 0.279 | 0.329 |
| 7            | 0.173               | 0.178                          | 0.035 | 0.122 | 0.153 | 0.177 | 0.201 | 0.238 |
| 8            | 0.132               | 0.140                          | 0.028 | 0.096 | 0.120 | 0.137 | 0.158 | 0.189 |
| 9            | 0.114               | 0.118                          | 0.018 | 0.090 | 0.105 | 0.117 | 0.129 | 0.148 |
| 10           | 0.462               | 0.428                          | 0.045 | 0.353 | 0.398 | 0.428 | 0.459 | 0.502 |

The posterior distributions for each  $\theta_j$  reflect what we saw in the raw observed proportions, the data that was collected for each location, and the underlying grand mean. Each posterior is approximately centered around the observed proportion, except for location 10, where the distribution is shifted slightly downward. The standard deviations are a reflection of the amount of observed data at each location. Location 9 had the smallest standard deviation, corresponding to the most observed data collected. Locations with smaller observed proportions have posterior distributions that are skewed right and locations with larger observed proportions have posterior distributions that are skewed left. This lends itself to the idea that the posteriors are being pulled toward an underlying grand mean.

d) We can obtain a 95% credible interval for the underlying proportion of of bicycle traffic by determining this interval for

$$\begin{aligned} E[\theta|y] &= E[E[\theta|\alpha, \beta]|y] \\ &= E\left[\frac{\alpha}{\alpha + \beta}|y\right] \end{aligned}$$

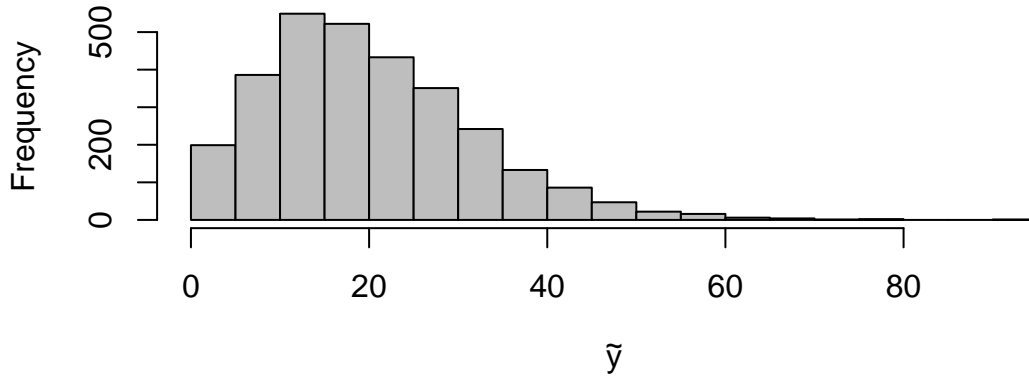
We can sample from  $p(\alpha, \beta|y)$  and determine this interval numerically. A histogram of the samples is displayed below.



The corresponding 95% credible interval is  $[0.145, 0.291]$ .

e) In order to predict the number of of vehicles out of 100 on a new city block that are bicycles, we sample  $(\alpha, \beta)$  from their posterior distribution, draw a new  $\theta$  from the population distribution  $p(\theta|\alpha, \beta)$  and then draw  $\tilde{y}$ , the number of bicycles that would appear out of 100 from a the data distribution,  $\text{Binom}(100, \theta)$ .

## Posterior Predictions



The 95% predictive interval is  $[3, 47]$ . This is a wide interval, and in practice, it doesn't offer very specific information. What it does tell us is that there is very high probability that less than half of the vehicles observed in the new location will be bikes. Assuming that the next location is from the same underlying population, we can be confident that less than half of the vehicles will be bikes, based on the 10 locations we have observed.

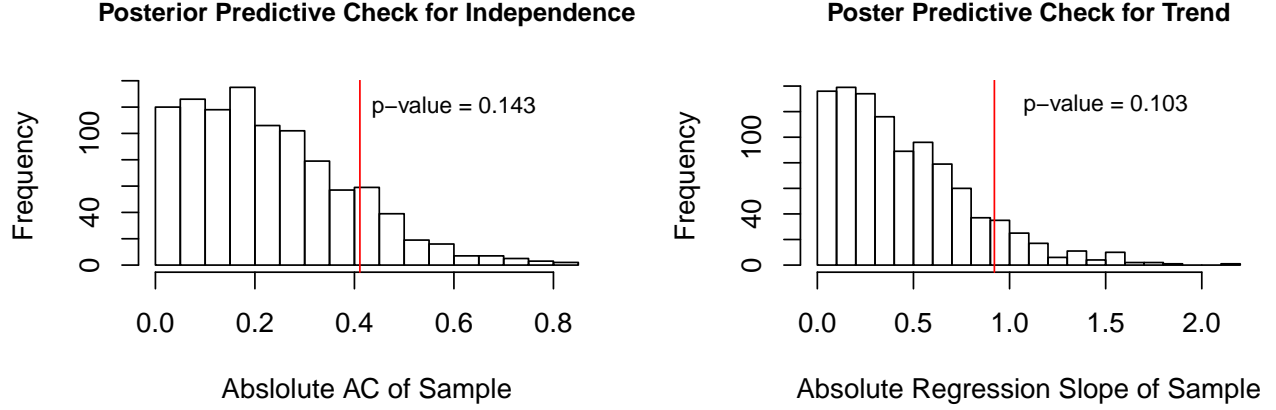
f) The beta distribution of the  $\theta_j$ 's was reasonable. We were modeling a proportion, and the Beta distribution restricted the range of each  $\theta_j$  to be between 0 and 1. We were able to use the data to infer a distribution on the hyper parameters  $(\alpha, \beta)$  that informed the average underlying proportion of bicycle traffic. The beta distribution also made intuitive sense as it was paired with the binomial distribution, where we were raising a the probability of observing a bicycle to the number of bicycles actually observed.

## Question 6.2

a) Our goal is to derive a posterior predictive check for (1) independent Poisson distributions and (2) no trend over time for each model that was used in Exercise 2.13. In order to check for independence, we can sample a set of 10 predicted data points and use the sample auto-correlation between them as the test quantity. The  $p$ -value will represent the probability you see a sample auto-correlation with a larger absolute value than what was seen in the actual data. To check for no trend over time, we can use our posterior for each model to predict 10 new years worth of observations, and run a linear regression on these values as a function of time. We will use the slope of this regression as our test quantity and the  $p$ -value will represent the probability we see a slope with a larger absolute value than what we saw in the data.

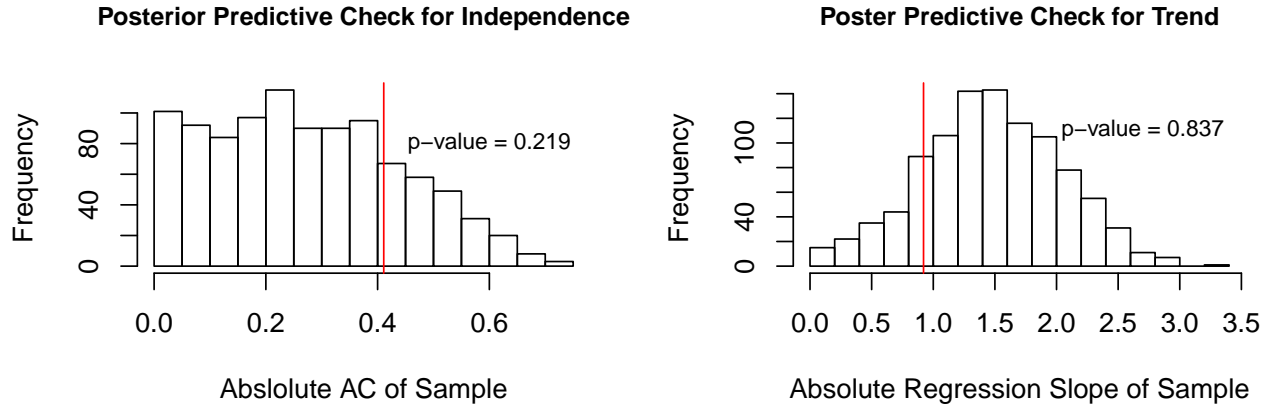
b) The first model we study is from 2.13(a), where  $\theta|y \sim \text{Gamma}(238, 10)$ . The posterior is not dependent on any data additional data from a given year, so we can sample theta 10 times and draw one observation from  $y^{\text{pred}} \sim \text{Pois}(\theta)$  for each sampled  $\theta$  to obtain one set of 10 years of predicted fatal accidents. From here we can compute the two test quantities and compare them to our data. The observed sample auto-correlation was 0.411 and the sample regression slope was 0.921.





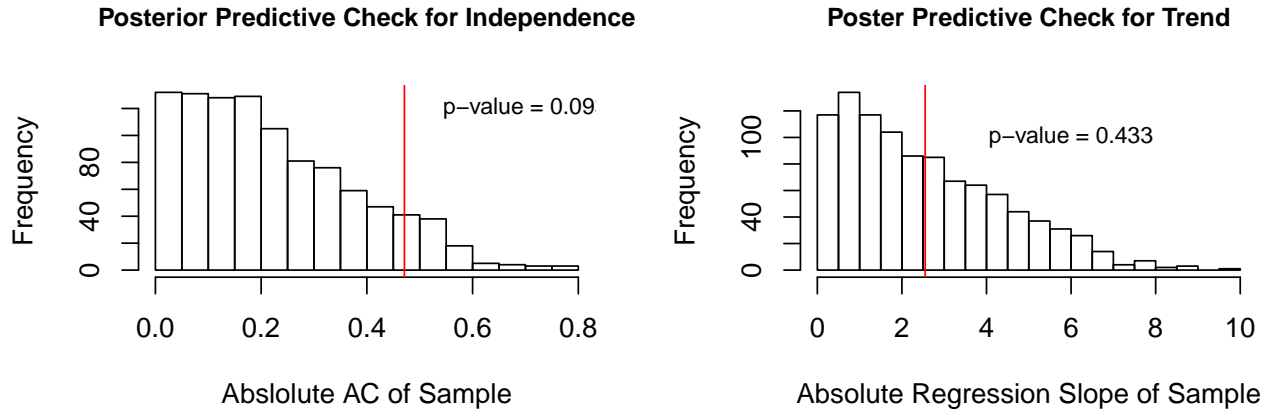
In our check for independence we find that 14% of 10-year samples had as large an observed auto-correlation as we observed in our actual data. If we assume that the observed data is made up of independent Poisson random variables, we can remain confident that our predicted values are from independent Poisson random variables also. We find a similar result when checking for linear trend. Only 12% of samples had a larger absolute value regression slope than the observed data. Assuming no linear trend in the data, we can be confident our predicted values do not have a linear trend over time. Since both  $p$ -values are over 0.1, we feel comfortable saying our model is capturing our data.

We conduct a similar study for the model in 2.13(b). Now  $y_i|x_i, \theta \sim \text{Poisson}(x_i\theta)$  where  $x_i$  is the number of passenger miles flow in year  $i$ . Recall then,  $\theta|y \sim \text{Gamma}(\sum y_i, \sum x_i)$ . In order to perform our posterior predictive check, we sample 10  $\theta$  values from the posterior and then sample 10 years worth of fatal accidents corresponding to the 10 years we observed in the data. We repeat this process 1000 times and calculate the distribution of our test quantities.



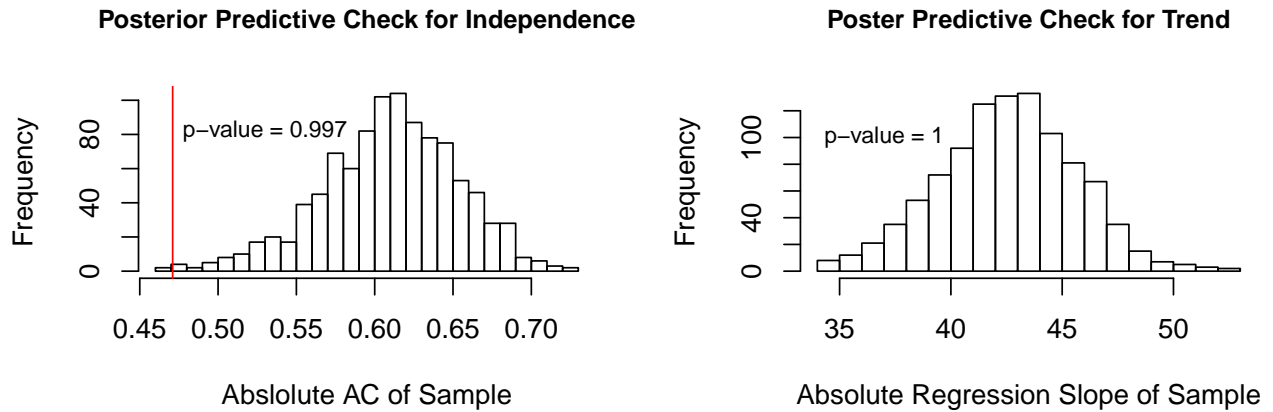
In our check for independence, we find that 23% of predicted values had an auto-correlation of larger absolute value than what we observed in our actual data. If we assume the data we observed was independent then since a large majority of our predictions had even less auto-correlation than that of the observed sample, it appears that our model is generating independent Poisson values. In our check for no linear trend, we find that the model generated predictions which had a larger slope than what was observed in the data 86% of the time. If we assume there is no linear trend in the observed data, we would have a tougher time believing that our model is generating predictions with no trend over time than in the first model.

Now we assess the model in 2.13(c), which was used for analyzing passenger deaths rather than fatal accidents. In this model,  $y^{\text{pred}}|\theta \sim \text{Poisson}(\theta)$  and the posterior  $\theta|y \sim \text{Gamma}(6919, 10)$ . We follow a similar procedure as above to generate the posterior predictive checks. For the data on passenger deaths, the observed absolute value auto-correlation is 0.471 and regression slope is 2.55.



We find that 9% of our samples had a larger absolute auto-correlation than what we saw in the data. Again, if we assume our data is independent, we can feel confident that we are generating independent Poisson predictions. With a  $p$ -value of 0.09 it might be cause for concern that our data is not being represented in the best way possible, but this is not egregious. In our check for trend, we find that 43% of predictions had a larger absolute regression slope than what we saw in the data. In this case our data is represented well and we can be confident there is no trend over time.

Finally we examine the model in 2.13(d). As in 2.13(b), the posterior for  $\theta$  is now dependent on passenger-miles for the given year. We have  $\theta|y \sim \text{Gamma}(6919, 5.7 \times 10^{12})$ . We follow a similar procedure to obtain our test quantity distribution.



In both the check for independence and the check for trend over time, nearly every predicted sample had a larger test quantity than what was observed. In this model the data was not well represented and we were not generating independent Poisson observations with no trend over time.

c) Our original intuition in question 2.13(e) was that it would be a good idea to include passenger miles, as we would expect deaths and accidents to increase as more miles were flown. The posterior predictive checks performed above, especially in model 4, tell us that our intuition did not match what we actually observed. By fitting a fatalities model that included passenger miles, we were unable to maintain independence, or represent our data well. This emphasizes the importance of using posterior checks to understand what your model is implying.

## Question 6.6

a) Assume that we choose to stop after 13 zeros were observed in the example on page 147. Then the likelihood is a function of both  $n$  and  $y$ . We would have the likelihood be zero after thirteen 0s are observed, and also require the  $n^{\text{th}}$  observation to be zero. We can formulate this as follows.

$$p(y, n | \theta) \propto \left( \prod_{i=1}^n \theta^{y_i} (1 - \theta)^{1-y_i} \right) \mathbf{1}_{\{(\sum_{i=1}^{n-1} 1-y_i=12) \cap (y_n=0)\}}$$

Since the prior on theta is uniform, the posterior is

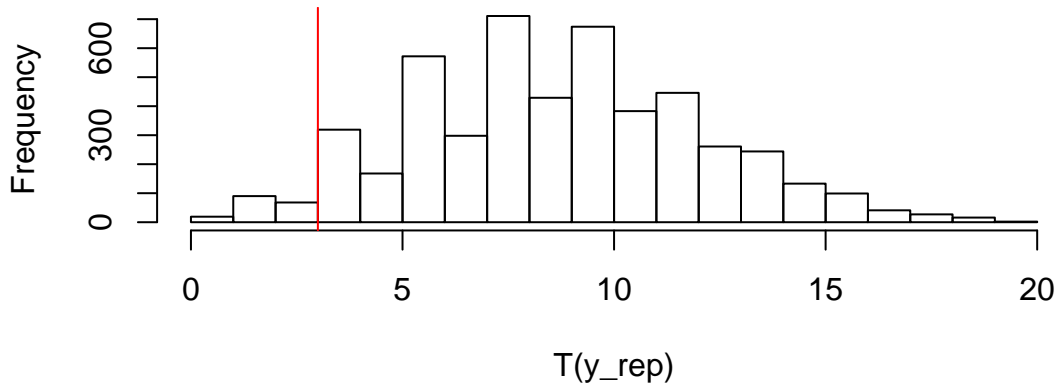
$$p(\theta | y, n) \propto \theta^7 (1 - \theta)^{13}$$

Notice that this is the same posterior as the approach taken on page 147 where they had

$$\begin{aligned} p(\theta | y) &\propto \theta^{\sum y_i} (1 - \theta)^{n - \sum y_i} \\ &\propto \theta^7 (1 - \theta)^{13} \end{aligned}$$

b) Using the new protocol we can perform a posterior predictive check for the number of switches between 0 and 1's observed.

### Posterior Predictive Check for Switches

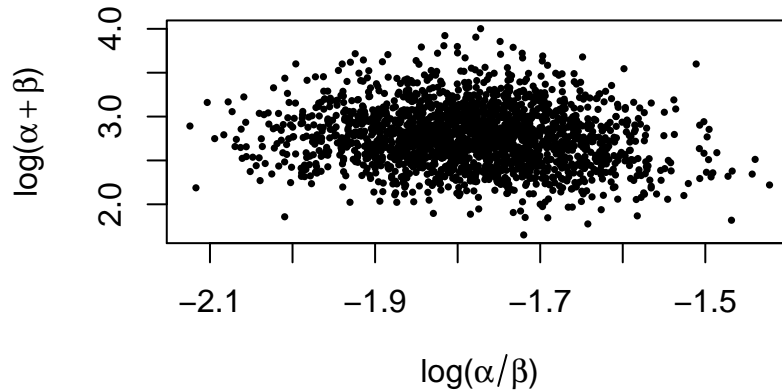


The distribution of switches from 0 to 1's has stark visual differences from what we see in Figure 6.5. Using this protocol, it appears that the number of switches is more often even than odd. We did not see this trend in figure 6.5. Similarly, we find that the range that the number of switches takes is wider in the new protocol, putting more mass on 0 and having observations as large as 20. In figure 6.5 we only see switches of 16 at most. This is to be expected though, since  $n$  is not restricted to be 20, and can grow as large as necessary to meet the conditions of the new protocol. We do find a similar “ $p$ -value” for this simulation, however, as  $Pr(T(y^{\text{rep}}, \theta) \leq T(y, \theta) | y) \approx 0.029$ .

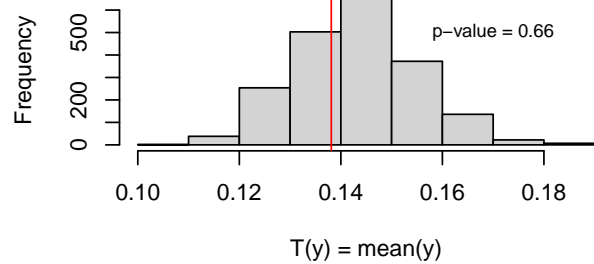
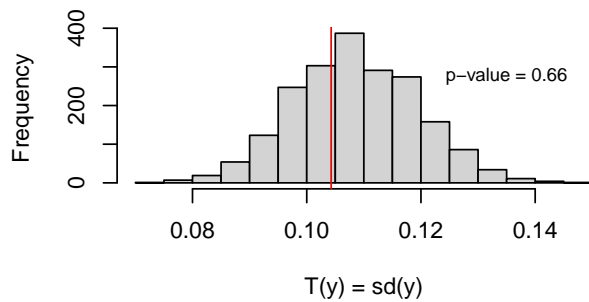
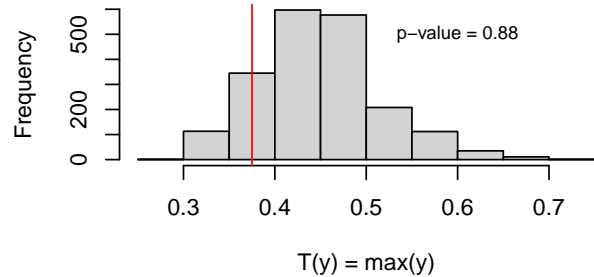
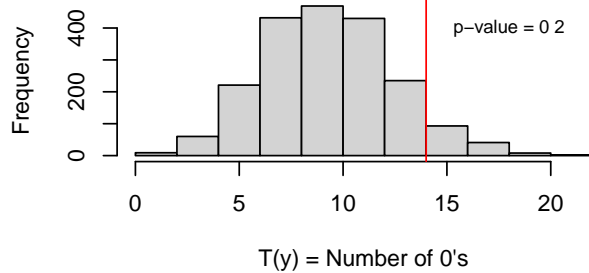
### Question 6.9

In this question, we use build the assumed model for the rat tumor data found in section 5.3. After constructing the posterior, we draw samples and compare them to figure 5.3 to ensure we are correct.

## 2000 Posterior Samples



We see approximately the same sampling distribution and move forward with posterior predictive checking. Using the same set up as the book in figure 6.12, We generate posterior predictive distributions for the the maximum, mean and standard deviation of the predicted proportions. In place of the minimum, we decide to check the number of 0 tumor predictions that we have in each of our samples, as this is a unique feature of the rat tumor data set. Histograms and  $p$ -values, the probability of observing a test quantity larger than what was seen in the data, are presented below.



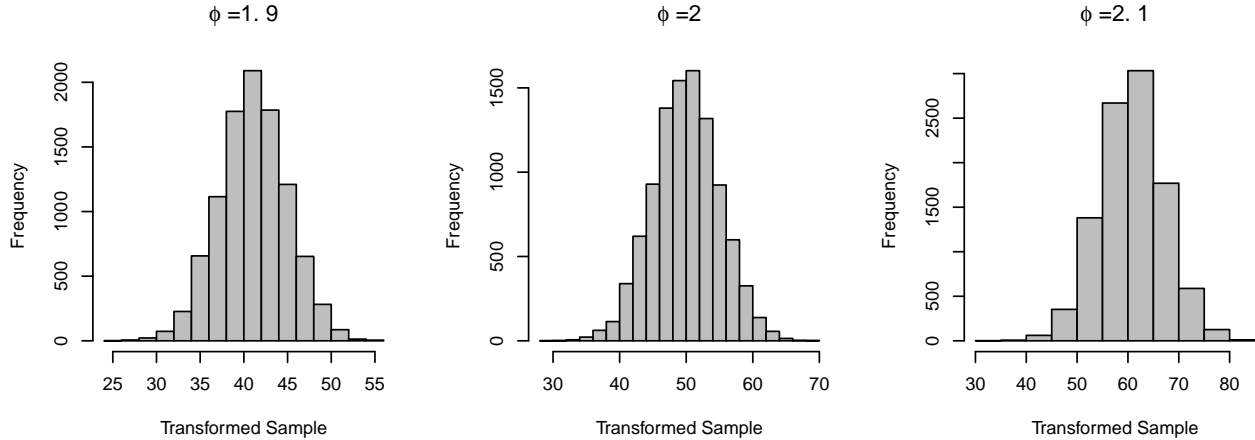
We find that overall, the model is representing the data well. The mean and standard deviation of the predicted proportions both have  $p$ -values of 0.66, indicating that the data was very likely to be seen in posterior draws. The number of zeros in our posterior samples had a  $p$ -value of 0.2. While not as well represented as in the mean and standard deviation, one in every 5 draws had as many or more zeros than what we saw in the data, so we feel comfortable saying that what we observed is feasible within our model. The least well-represented quantity was the maximum. With a  $p$ -value of 0.88, just under 9 of every 10 draws had a larger maximum proportion than what we saw in the data. If we were very concerned with measuring this quantity, we might want to consider the implications of this, but in general, we feel the data we observed is being captured appropriately.

## Question 7.5

a) When we apply a prior of the form

$$p(\mu, \log(\sigma), \phi) \propto p(\phi),$$

we are assuming that the parameters are mutually independent. This is not the case, as the value of  $\mu$  and  $\sigma^2$  are directly related to the value of the parameter  $\phi$  that is chosen. This is intuitive, as the range and location of our data could be drastically changed as a result of our choice of  $\phi$ . Assume we sample 10,000 observations from the square root of a Normal distribution with  $\mu = 50$  and  $\sigma = 5$ . We know the true value of  $\phi = 2$ . Below we display histograms of the data transformed with  $\phi = 1.9, 2$ , and  $2.1$ , respectively.



Even small differences in choice of  $\phi$  result in wide differences in resulting distribution. By visual comparison, it is clear that you would not want to make inference about  $\mu$  or  $\sigma$  before or independent of your knowledge of  $\phi$ , as heuristic guesses for  $\mu$  could range between 40 and 60 depending on  $\phi$  for the above distributions. The prior we have discussed implies that each value of  $\mu$  and  $\sigma$  is apriori equally likely regardless of  $\phi$ , which clearly should not be the case. To remedy this, we must add some dependence structure in our prior.

b) Building off of what we found in part (a), we know that the prior we assign has to ensure that  $\mu$  and  $\sigma$  are dependent on the value of  $\phi$ . In conjunction with this, given a value of  $\phi$ , we would like  $\mu$  and  $\sigma$  to be relatively uniform in a feasible area. In mathematical notation, we want

$$\begin{aligned} p(\mu, \log \sigma, \phi) &= p(\mu_\phi, \log \sigma_\phi | \phi) p(\phi) \\ &\propto f(\phi) p(\phi) \end{aligned}$$

Here, we use the subscript  $\phi$  to emphasize the dependence of  $\mu$  and  $\sigma$  on our choice of  $\phi$ . This formulation ensures that for a given  $\phi$ ,  $p(\mu, \log \sigma | \phi) \propto 1$ , but in the case that  $\phi$  is not known, the density of  $\mu$  and  $\sigma$  remain dependent on the value of  $\phi$  in some way. In the example given in this part of the question,  $f(\phi) = \left( \prod_{i=1}^n y_i^{1/n} \right)^{1-\phi}$ . By including this function of  $\phi$  in our prior, we can ensure some dependence structure on  $\mu$  and  $\sigma$  with respect to our choice of  $\phi$ . This way, we must obtain information on  $\phi$  before we can claim we know how  $\mu$  and  $\sigma$  behave. In their paper *An Analysis of Transformations* (1964), Box and Cox acknowledge the somewhat arbitrary nature of this choice of  $f(\phi)$ , but note that it is necessary for the purposes we have discussed above.

c) The posterior has the following form

$$p(\mu, \sigma, \phi|y) \propto p(\mu, \sigma|\phi)p(y|\mu, \sigma, \phi) \\ \propto \frac{f(\phi)p(\phi)}{\sigma^{n+2}} \exp \left\{ \frac{-\sum_{i=1}^n (y_i^{(\phi)} - \mu)^2}{2\sigma^2} \right\} J(\phi|y)$$

where  $f(\phi) = \left( \prod_{i=1}^n y_i^{1/n} \right)^{1-\phi}$ ,  $p(\phi)$  is the prior on  $\phi$  and  $J(\phi|y) = \prod_{i=1}^n \left| \frac{dy_i^{(\phi)}}{dy_i} \right|$ . We can rewrite the posterior in the following form using the steps of section 3.2.

$$p(\mu, \sigma, \phi|y) \propto \frac{f(\phi)p(\phi)}{\sigma^{n+2}} \exp \left\{ \frac{-\sum_{i=1}^n (y_i^{(\phi)} - \mu)^2}{2\sigma^2} \right\} J(\phi|y) \\ \propto \frac{f(\phi)p(\phi)J(\phi|y)}{\sigma^{n+2}} \exp \left\{ \frac{-(n-1)s_{(\phi)}^2 + n(\bar{y}_{(\phi)} - \mu)^2}{2\sigma^2} \right\}$$

Where  $s_{(\phi)}^2$  and  $\bar{y}_{(\phi)}$  have their standard meanings, except taken for the vector  $y_i^{(\phi)}$ . We then know from 3.2 that when we take the integral of this posterior with respect to  $\mu$  we have the following.

$$p(\sigma^2, \phi|y) \propto \frac{f(\phi)p(\phi)J(\phi|y)}{(\sigma^2)^{(n+1)/2}} \exp \left\{ \frac{-(n-1)s_{(\phi)}^2}{2\sigma^2} \right\}$$

We can then integrate out  $\sigma^2$ ,

$$p(\phi|y) = \int_0^\infty p(\sigma^2, \phi|y) d\sigma^2 \\ \propto \int_0^\infty \frac{f(\phi)p(\phi)J(\phi|y)}{(\sigma^2)^{(n+1)/2}} \exp \left\{ \frac{-(n-1)s_{(\phi)}^2}{2\sigma^2} \right\} d\sigma^2 \\ \propto f(\phi)p(\phi)J(\phi|y) \int_0^\infty (1/\sigma^2)^{(n+1)/2} \exp \left\{ -\frac{1/\sigma^2}{\left( \frac{2}{(n-1)s_{(\phi)}^2} \right)} \right\} d\sigma^2 \\ \propto f(\phi)p(\phi)J(\phi|y) \int_0^\infty (u)^{\frac{(n-1)}{2}-1} \exp \left\{ -\frac{u}{\left( \frac{2}{(n-1)s_{(\phi)}^2} \right)} \right\} du \quad \text{Sub: } u = 1/\sigma^2$$

Note, the final integral has the kernel of a Gamma  $\left( \frac{(n-1)}{2} - 1, \frac{2}{(n-1)s_{(\phi)}^2} \right)$  and so we can write,

$$p(\phi|y) \propto f(\phi)p(\phi)J(\phi|y) \Gamma\left(\frac{n-1}{2}\right) \left( \frac{2}{(n-1)s_{(\phi)}^2} \right)^{\frac{n-1}{2}} \\ \propto \frac{f(\phi)J(\phi|y)}{(s_{(\phi)}^2)^{(n-1)/2}} p(\phi)$$

where  $f(\phi) = \left( \prod_{i=1}^n y_i^{1/n} \right)^{1-\phi}$ ,  $p(\phi)$  is the prior on  $\phi$ , and  $J(\phi|y) = \prod_{i=1}^n \left| \frac{dy_i^{(\phi)}}{dy_i} \right|$ .

d) The fact that the prior distribution depends on the data is problematic in the sense that the prior is no longer a quantification of what you knew before the study, but rather a blend of what you knew before and what you know as a result of the study. If you are operating under the philosophy that the prior represents your uncertainty before the study, and the posterior represents your uncertainty afterward, by constructing a prior as we have in this question, we lose that separation. It becomes more difficult to discuss how the data modified our prior beliefs, and less clear how seeing different data would have shifted us away from prior uncertainty.

e) By assuming that negative values of  $y^{(\phi)}$  are not possible, we are limiting the potential normal distributions that can result from transforming the data. If we assume  $y^{(\phi)} \sim N(\mu, \sigma)$ , then by truncating the possible data values at 0, the mean and standard deviation of this distribution must take values in tandem so that it would be unreasonable to see a sampled value of  $y^{(\phi)}$  that is negative. Consider the case when the transformed data has a  $\mu = 3$ , the corresponding standard deviation of the normal distribution would have to be no more than 1 so that  $y^{(\phi)}$  would (essentially) always be positive. This changes drastically if we were to say the transformed data has  $\mu = 1000$ . In this case the standard deviation could range from near 0 to as high as 333 before we had to consider the possibility of obtaining negative values of  $y^{(\phi)}$ . This interplay of the mean and standard deviation that necessary to deal with the truncation at zero is something to be aware of when applying the transformation.

## Question 7.6

a) We apply the prior on our parameters from the question above so that,

$$p(\mu, \log \sigma, \phi) \propto \left( \prod_{i=1}^n y_i^{1/n} \right)^{1-\phi}$$

Note, this indicates that  $p(\phi) \propto 1$ . We can write the posterior for the power transformed model as follows.

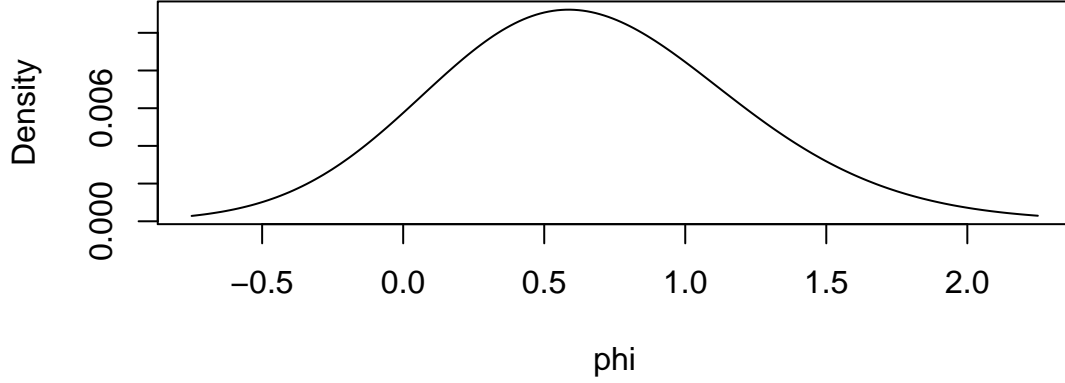
$$\begin{aligned} p(\mu, \sigma^2, \phi|y) &\propto p(\mu, \sigma^2, \phi, y) \\ &\propto p(\mu|\sigma^2, \phi, y)p(\sigma^2|\phi, y)p(\phi|y) \end{aligned}$$

We know the components of this posterior from Question 7.5, and can pull samples for  $\mu$  and  $\sigma^2$  using techniques derived in section 3.2. Recall that under this prior,

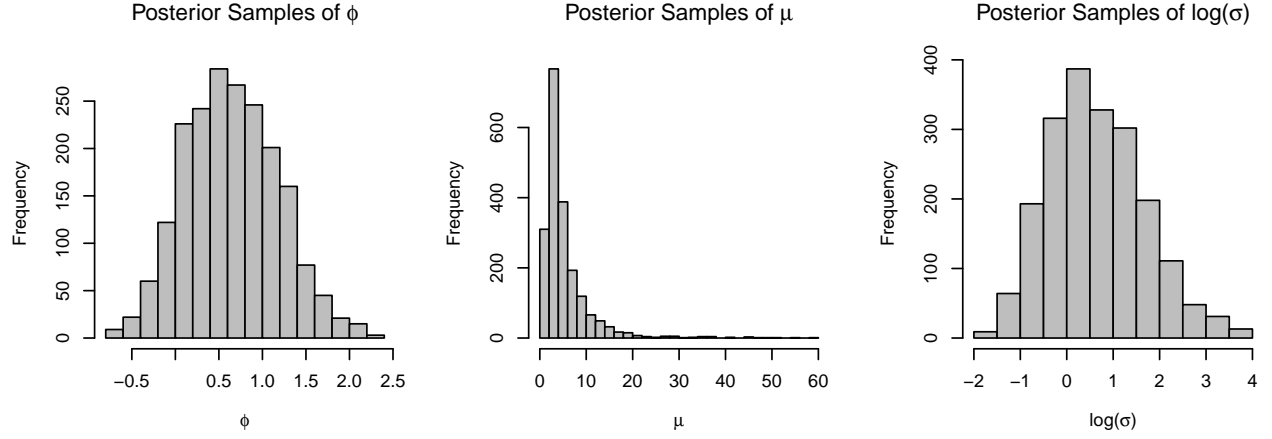
$$p(\phi|y) \propto \frac{\left( \prod_{i=1}^n y_i^{1/n} \right)^{1-\phi} \prod_{i=1}^n \left| \frac{dy_i^{(\phi)}}{dy_i} \right|}{(s_{(\phi)}^2)^{(n-1)/2}}$$

Below, we plot the calculated posterior density values of  $\phi$ .

## Posterior Density of Phi



Most of the density is concentrated around  $\phi = 1/2$ . This happens to coincide with the MLE. We now sample from this distribution to obtain estimates for  $\mu$  and  $\sigma$  (we use  $\log \sigma$  for the visual). Histograms of the samples are presented below.



As the density plot showed, most samples of  $\phi$  from the posterior are concentrated slightly above 0.5. The sampling distribution of  $\mu$  is skewed right. This makes sense as  $\mu$  is bounded below by 0, but for large values sampled from  $p(\phi|y)$ , we expect to see large estimates for  $\mu$  in its tail. The range of  $\log \sigma$  is large, with most of the density concentrated between -1 and 2. This indicates a wide range of potential standard deviation estimates. We summarize each of these distributions in the table below.

Table 6: Posterior Distribution Summary

|          | Mean  | SD    | 5%     | 25%   | 50%   | 75%   | 95%    |
|----------|-------|-------|--------|-------|-------|-------|--------|
| $\phi$   | 0.647 | 0.534 | -0.177 | 0.245 | 0.622 | 1.029 | 1.541  |
| $\mu$    | 5.412 | 5.546 | 1.507  | 2.415 | 3.711 | 6.313 | 14.347 |
| $\sigma$ | 3.398 | 5.077 | 0.414  | 0.891 | 1.725 | 3.805 | 11.203 |

The wide range in summary statistics for  $\mu$  and  $\sigma$  demonstrate the impact that the variation in  $\phi$  can have on their estimates. For smaller  $\phi$  values,  $\mu$  and  $\sigma$  samples will generally be smaller as the range and location of the data  $y$  is shifted toward 0. An opposite effect occurs for larger values of  $\phi$ . This dependence adds to the uncertainty of  $\mu$  and  $\sigma$  as they are being estimated.



b) Now we apply the same model, but allow means and variances to change by county while  $\phi$  remains fixed. Let  $\mu = (\mu_1, \mu_2, \mu_3)$  and  $\sigma = (\sigma_1, \sigma_2, \sigma_3)$  corresponding to the mean and variance for each county in alphabetical order (as displayed in table 7.3). Let  $n_1 + n_2 + n_3 = N$  where  $n_i$  corresponds to the total sample size in the  $i^{th}$  county. We apply a prior of the form

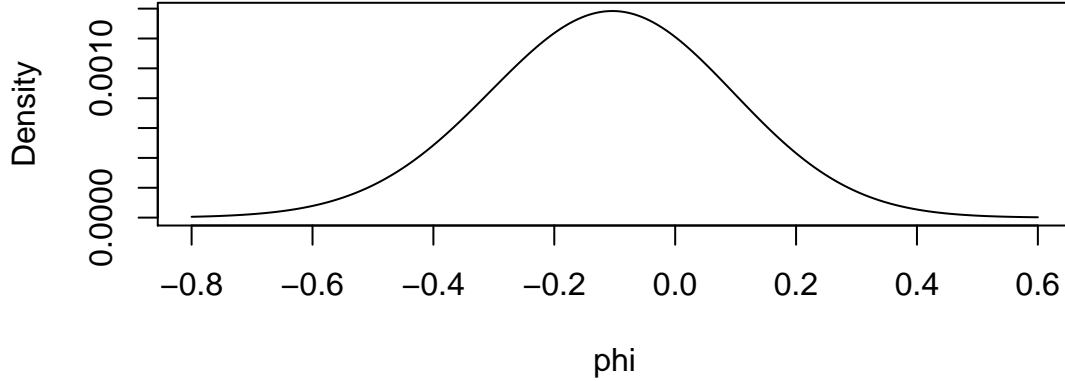
$$p(\mu, \log \sigma, \phi) \propto \left( \prod_{i=1}^3 \prod_{j=1}^{n_i} y_{ij}^{1/N} \right)^{1-\phi}$$

Note, this is the same prior as in 7.5, but applied to all the data found in Table 7.3. We can write the posterior as follows,

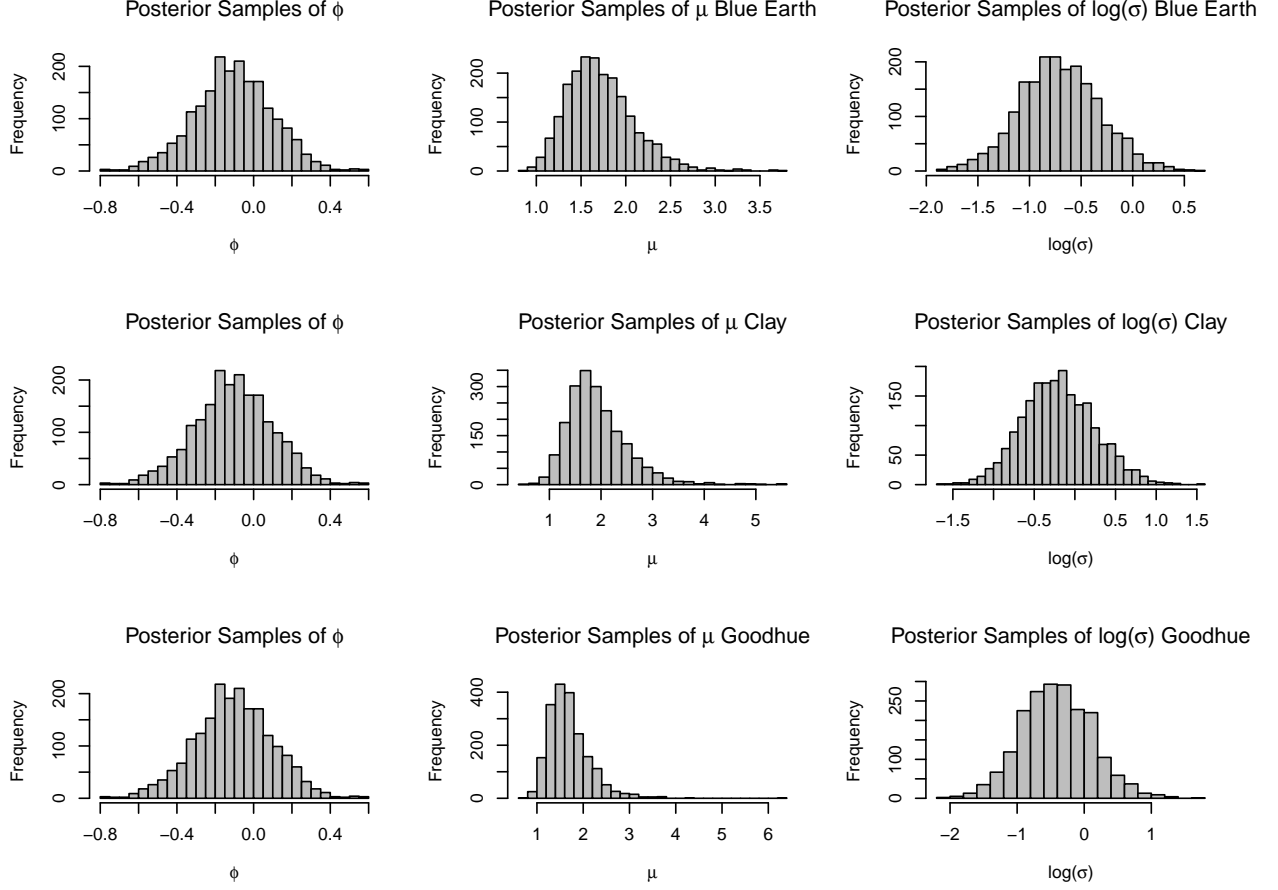
$$\begin{aligned} p(\mu, \sigma^2, \phi | y) &\propto p(\mu, \sigma^2, \phi, y) \\ &\propto p(\phi | y) \prod_{i=1}^3 p(\mu_i | \sigma_i^2, \phi, y) p(\sigma_i^2 | \phi, y) \end{aligned}$$

In order to draw samples from the posterior, we start by examining the posterior density for  $\phi$

### Posterior Density of Phi



This posterior density is centered closer to zero than what we observed when we were only considering the data from Blue Earth. We sample from this distribution and use the same  $\phi$  to estimate the mean and standard deviation of the three counties' transformed normal distributions individually. Sampling for  $\mu$  and sigma again relies on techniques from section 3.2. Below, we plot our posterior samples. Note, the sampling distribution of  $\phi$  is the same for each county.



While relatively similar in shape and range, there are subtle differences in the sampling distributions observed for each parameter. It is worth noticing that the distributions of  $\log \sigma$  for Goodhue and Clay take larger values than that for Blue Earth. When we refer back to the data, this makes sense, as Goodhue had an extremely high Radon measurement of 43.5 and Clay had two measurements above 19. The highest included measurement from Blue Earth was only 13. We also notice that the distribution for the transformed mean of Clay is shifted toward slightly higher values than that for Goodhue and Blue Earth. This is again reflected in our observed data, as Clay has 6 observations above 9.0 while Blue Earth and Goodhue have 6 combined. A table examining the distributions further is presented below.

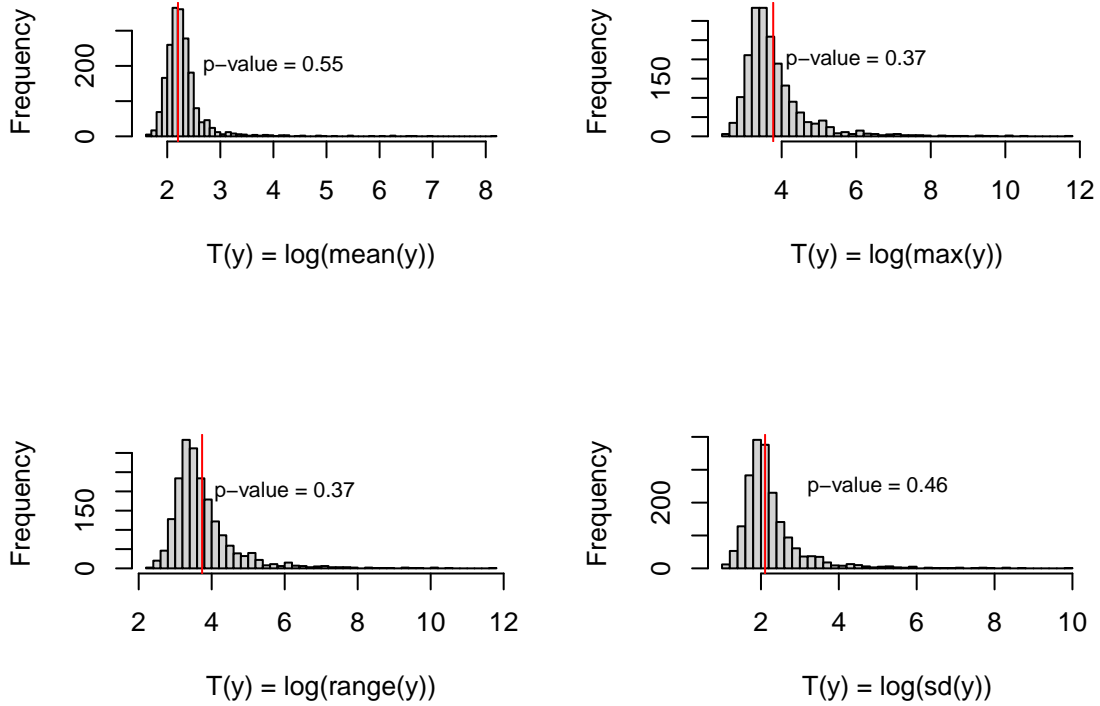
Table 7: Posterior Distribution Summary

|      | $\phi$ | Posterior Means |         |            | Posterior SDs |            |               |
|------|--------|-----------------|---------|------------|---------------|------------|---------------|
|      |        | $\mu_{BE}$      | $\mu_C$ | $\mu_{GH}$ | $\sigma_{BE}$ | $\sigma_C$ | $\sigma_{GH}$ |
| Mean | -0.104 | 1.720           | 1.921   | 1.692      | 0.533         | 0.907      | 0.765         |
| SD   | 0.203  | 0.376           | 0.576   | 0.453      | 0.220         | 0.439      | 0.459         |
| 5%   | -0.444 | 1.197           | 1.171   | 1.122      | 0.258         | 0.409      | 0.284         |
| 25%  | -0.233 | 1.452           | 1.533   | 1.387      | 0.380         | 0.606      | 0.460         |
| 50%  | -0.104 | 1.670           | 1.823   | 1.616      | 0.488         | 0.811      | 0.653         |
| 75%  | 0.033  | 1.928           | 2.209   | 1.904      | 0.636         | 1.105      | 0.950         |
| 95%  | 0.227  | 2.385           | 2.948   | 2.497      | 0.944         | 1.724      | 1.610         |

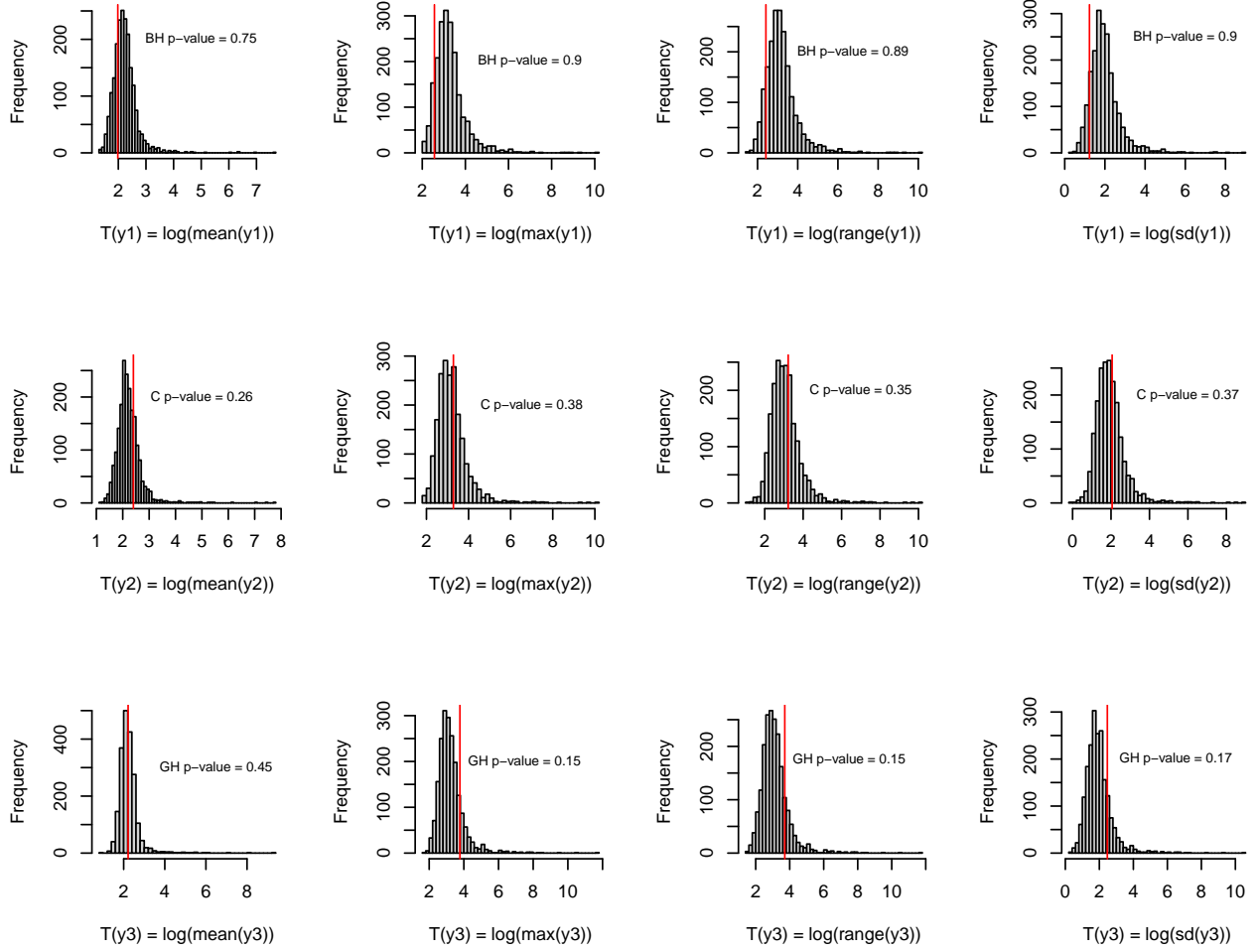
The distribution summaries reiterate what was discussed for the plots. We learn that the mean value of  $\phi$  from posterior sampling is -0.104 with a standard deviation of 0.203. It is worth mentioning that this is close

to 0, which would indicate a log transformation.

c) We run posterior predictive checks for this model using mean, max, range (max-min), and standard deviation as test quantities. We take the log of each of these values due to the long tails exhibited for each of these statistics. In order to obtain our posterior predictive samples, we first draw a value  $\phi$  from  $p(\phi|y)$ . Then, since we had 3 counties, we draw 3 pairs from  $p(\mu, \sigma|\phi, y)$ . Then, we sample a vectors of length 12, 10 and 11 from  $N(\mu_1, \sigma_1)$ ,  $N(\mu_2, \sigma_2)$ , and  $N(\mu_3, \sigma_3)$  respectively to mimic the data we have in table 7.3. Once we have these vectors, we transform them back to standard data form using the inverse of the transformation corresponding to the original  $\phi$  we sampled. We found it useful to look at the overall max, mean, standard deviation, and range (max - min) for all the data (combining the three counties) as a basic way to determine if what we see in table 7.3 is likely to be seen from our model.



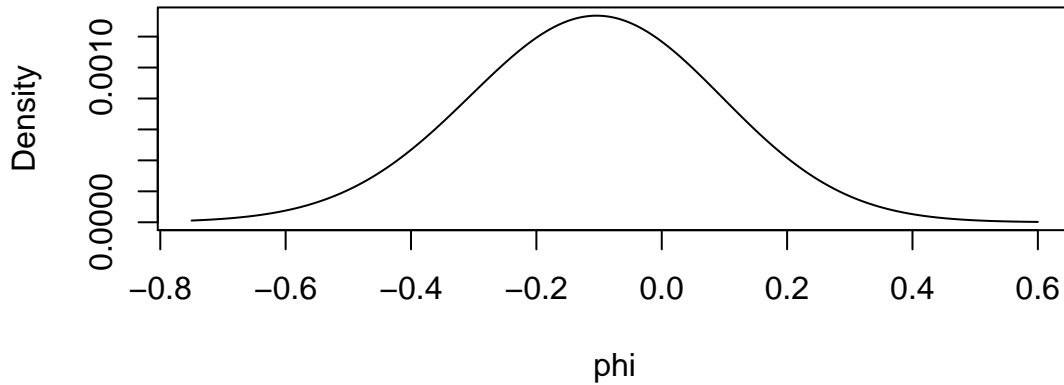
Based on the plots above, our data is being well represented. For each of the test quantities, we have  $p$ -values larger than 0.3 and smaller than 0.6. However, it is worth noting the massive variation in the test quantities that we see in sampling. This is likely a result of the wide range of estimates that are possible for  $\mu$  and  $\sigma$  for the different  $\phi$  values that can be sampled. This variation is translated in extremes for each of the test quantities above. It is possible that we should be concerned about this because the data we observe does not have very extreme characteristics. The most extreme values are very unlikely in sampling, but depending on the scientific question, this is something to be wary of. We now look at the same test statistics for the three different counties of data we generated data separately, comparing them to the county that they correspond to in table 7.3.



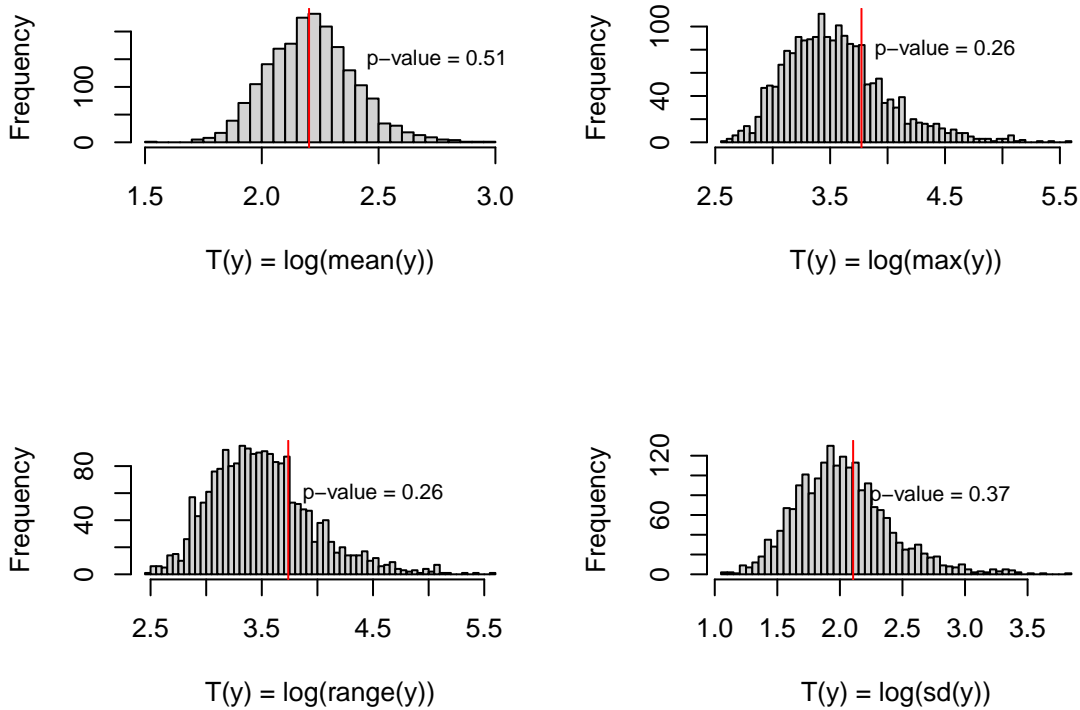
When looking at individual counties, the  $p$ -values are slightly less convincing than when we had all the generated data pooled together. Despite this, none of the  $p$ -values ranged below 0.09 or above 0.89. Based on this, we can still say that our data would not be out of the ordinary if it were to be generated by our model based on all of these statistics. However, we are gain confronted with very extreme values in the test quantity sampling distribution. While we are inclined to say the model is fitting the data well, we must proceed with caution, knowing it has the capabilities to generate out of the ordinary samples.

d) In order to answer whether a lognormal model would be appropriate for this data, we reference our posterior marginal density for  $\phi$ .

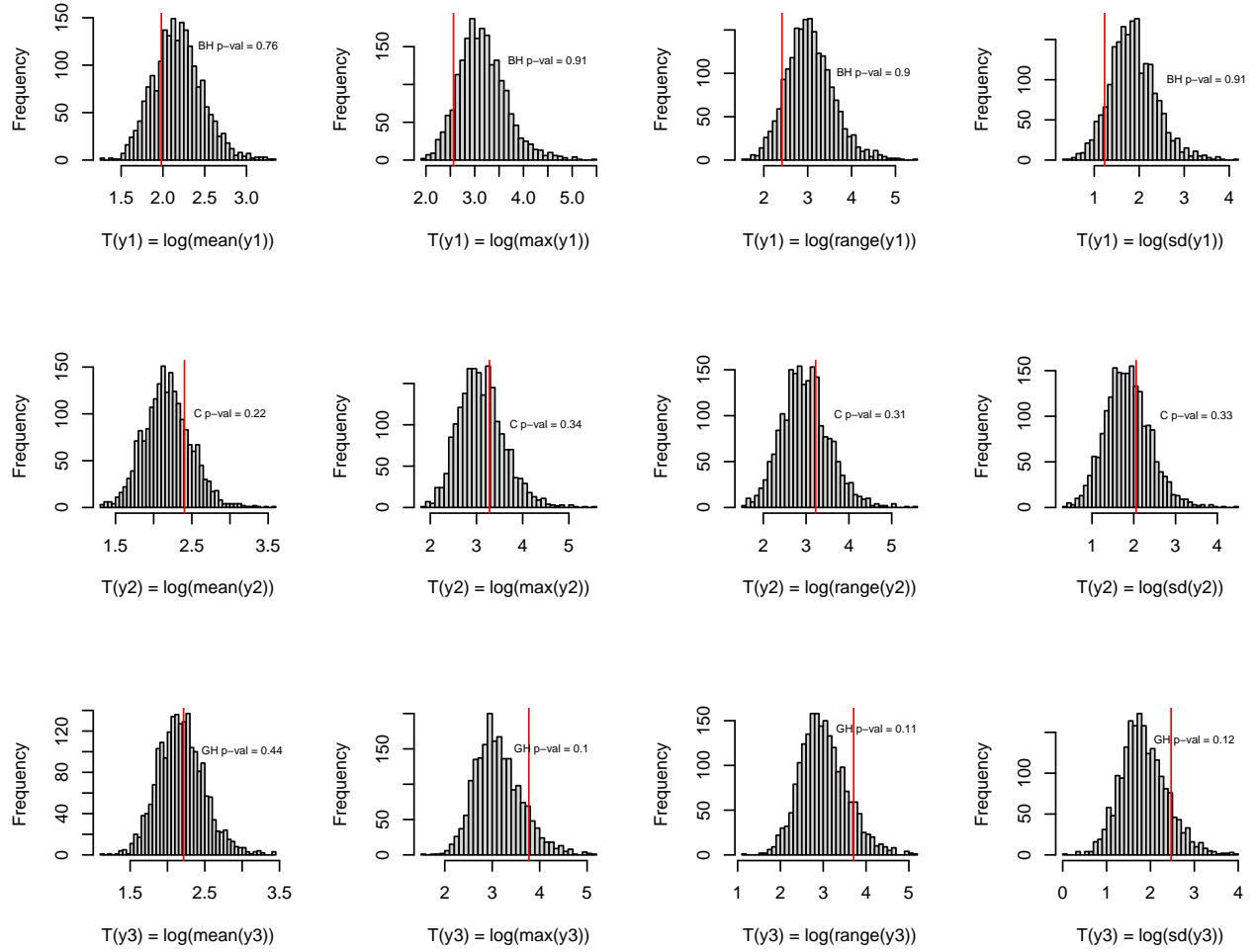
## Posterior Density of Phi



As can be seen from the plot, the density is nearly centered around 0, which indicates a log transformation. It would also be useful to reference past literature about how toxic chemical concentrations are modeled. Has lognormal been used before, and was it successful? Intuitively, and based on viewing the data and this posterior, it seems as though a standard lognormal model would be simpler, be more stable, and possibly offer more reasonable interpretability. We generate similar posterior predictive checks as in part (c) to examine the result under the log-normal model and present them below. We first look the counties pooled test quantities.



The  $p$ -values similar to those in part (c), and there are less extreme values in the test quantity sampling distribution. Now we look at each generated county individually and compare it to its corresponding county in table 7.3.



We again find a very similar result to our original model but reduce the extreme values in our sampling distribution. Based on these results, we would claim a lognormal model is reasonable for this data.

## Code Appendix

```
knitr::opts_chunk$set(message = F)
knitr::opts_chunk$set(warning = F)
knitr::opts_chunk$set(fig.height = 3)
knitr::opts_chunk$set(fig.width = 6)
knitr::opts_chunk$set(fig.align="center")
knitr::opts_chunk$set(echo = F)
`%notin%` <- Negate(`%in%`)
library(kableExtra)
library(latex2exp)
set.seed(53)

library(lattice)

# output: one sample from p(theta | mu, tau, y)
conditional.theta=function(ybar,mu,tau,sigma){
  theta=rep(0,nschools)
  theta.hat=rep(0,nschools)
  V.hat=rep(0,nschools)
  for(j in 1:nschools) {
    V.hat[j]=1/(1/sigma[j]^2+1/(tau^2))
    theta.hat[j]=(ybar[j]/sigma[j]^2+mu/tau^2)*V.hat[j]
    theta[j]=rnorm(1,theta.hat[j],sqrt(V.hat[j]))
  }
  theta
}

# output: nsample samples from p(mu | tau, y)
sample.mar.mu=function(ybar, tau, sigma,nsample){
  V.mu.inv=sum(1/(sigma^2+tau^2))
  mu.hat=sum((1/(sigma^2+tau^2))*ybar)/V.mu.inv
  mu.sample=rnorm(nsample,mu.hat,sqrt(1/V.mu.inv))
  mu.sample
}

# evaluates p(tau | y)
marginal.tau=function(ybar,tau,sigma){
  V.mu.inv=sum(1/(sigma^2+tau^2))
  mu.hat=sum((1/(sigma^2+tau^2))*ybar)/V.mu.inv
  eval=exp(-(ybar-mu.hat)^2/(2*(sigma^2+tau^2)))
  eval=eval/sqrt(sigma^2+tau^2)
  eval=sqrt(1/V.mu.inv)*prod(eval)
  eval
}

ybar=c(28.39,7.94,-2.75,6.82,-0.64,0.63,18.01,12.16)
nschools=length(ybar)
sigma=c(14.9,10.2,16.3,11.0,9.4,11.4,10.4,17.6)
```

```

# Grid to evaluate p(tau | y)
x.tau=seq(0.00001,40,length=10000)
# evaluate p(tau | y) at 1000 points in the
# interval [0.00001,40]
post.tau=apply(t(x.tau),2,marginal.tau, ybar=ybar, sigma=sigma)
sample.tau=sample(x.tau,10000,replace=TRUE,prob=post.tau)
# draw 200 samples from p(mu | tau, y)
sample.mu=apply(t(sample.tau),2,sample.mar.mu,ybar=ybar, sigma=sigma,nsample=1)
# draw 200 samples from p(theta | mu, tau, y)
sample.theta=matrix(0,ncol=nschools,nrow=10000)
for (i in 1:10000){
  sample.theta[i,]=conditional.theta(ybar, sample.mu[i], sample.tau[i],sigma)
}

best_school <- rep(0, nschools)

for (i in 1:nrow(sample.theta)){
  best <- which(sample.theta[i,] == max(sample.theta[i,]))
  best_school[best] <- best_school[best] + 1
}

best_school <- best_school/nrow(sample.theta)

best_school <- as.data.frame(t(round(best_school,3)), row.names = c("Probability"))

colnames(best_school) <- c("A","B","C", "D", "E", "F" ,"G", "H")

kable_styling(kable(best_school, format = "latex", booktabs= TRUE, align = 'c',
                    caption = "Posterior Probability of being the 'Best' School", row.names = T), latex=TRUE)

better_school <- matrix("-", nrow = length(best_school), ncol = length(best_school))

for (i in 1:ncol(sample.theta)) {
  for (j in 1:ncol(sample.theta)) {
    if(i != j){
      better_school[i,j] <- round(sum(sample.theta[,i] > sample.theta[,j])
                                / nrow(sample.theta),3)
    }
  }
}

better_school <- cbind(c("A","B","C", "D", "E", "F" ,"G", "H"), better_school)

better_school <- as.data.frame(better_school)
colnames(better_school) <- c("School", "A","B","C", "D", "E", "F" ,"G", "H")

better_school_table <- kable_styling(kable(better_school, format = "latex", booktabs= TRUE, align = 'c',
                    caption = "Posterior Probability of each School Being Better than Another", row.names = T), latex=TRUE)

add_header_above(better_school_table, c("", "Probability of being better than" = 8))
set.seed(53)

```



```

ybar=c(28.39,7.94,-2.75,6.82,-0.64,0.63,18.01,12.16)
nschools=length(ybar)
sigma=c(14.9,10.2,16.3,11.0,9.4,11.4,10.4,17.6)

School_a <- rnorm(5000, mean = ybar[1], sd = sigma[1])
School_b <- rnorm(5000, mean = ybar[2], sd = sigma[2])
School_c <- rnorm(5000, mean = ybar[3], sd = sigma[3])
School_d <- rnorm(5000, mean = ybar[4], sd = sigma[4])
School_e <- rnorm(5000, mean = ybar[5], sd = sigma[5])
School_f <- rnorm(5000, mean = ybar[6], sd = sigma[6])
School_g <- rnorm(5000, mean = ybar[7], sd = sigma[7])
School_h <- rnorm(5000, mean = ybar[8], sd = sigma[8])

Indep_schools <- cbind(School_a,School_b,School_c,School_d,
                      School_e,School_f,School_g,School_h)

best_school_indep <- rep(0, nschools)

for (i in 1:nrow(Indep_schools)){
  best <- as.numeric(which(Indep_schools[i,] == max(Indep_schools[i,])))
  best_school_indep[best] <- best_school_indep[best] + 1
}

best_school_indep <- round(best_school_indep/nrow(Indep_schools),3)

best_school_indep
best_school_indep <- as.data.frame(t(best_school_indep), row.names = c("Probability"))

colnames(best_school_indep) <- c("A","B","C", "D", "E", "F" ,"G", "H")

kable_styling(kable(best_school_indep, format = "latex", booktabs= TRUE, align = 'c',
                    caption = "Posterior Probability of being the 'Best' School (Independent Estimates)

better_school_indep <- matrix("-", nrow = length(best_school_indep),
                             ncol = length(best_school_indep))

for (i in 1:ncol(Indep_schools)) {
  for (j in 1:ncol(Indep_schools)) {
    if(i != j){
      better_school_indep[i,j] <- round(sum(Indep_schools[,i] > Indep_schools[,j])
                                       / nrow(Indep_schools),3)
    }
  }
}

better_school_indep <- cbind(c("A","B","C", "D", "E", "F" ,"G", "H"), better_school_indep)

better_school_indep <- as.data.frame(better_school_indep)
colnames(better_school_indep) <- c("School", "A","B","C", "D", "E", "F" ,"G", "H")

```

```

better_school_indep_table <- kable_styling(kable(better_school_indep, format = "latex", booktabs= TRUE,
caption = "Posterior Probability of each School Being Better than Another (Independent)"))

add_header_above(better_school_indep_table, c("", "Probability of being better than" = 8))
y_props <- c(16/(16+58), 9/(9+90), 10/(10+48), 13/(13+57),
            19/(19+103), 20/(20+57), 18/(18+86), 17/(17+112),
            35/(35+273), 55/(55+64))

y_bikes <- c(16, 9, 10, 13, 19, 20, 18, 17, 35, 55)
y_v <- c(58, 90, 48, 57, 103, 57, 86, 112, 273, 64)
ns <- y_bikes+y_v

Mom_a <- 2.89
Mom_b <- 11.86

logaoverb <- log(Mom_a/Mom_b)
logapusb <- log(Mom_a + Mom_b)

X <- seq(-1.8, -1.1, length.out = 500)
Y <- seq(1.3, 4.5, length.out = 500)

alphs <- exp(X+Y)/(1+exp(X))
bets <- exp(Y)-exp(X+Y)/(1+exp(X))
z <- matrix(0,nrow=500,ncol=500)

for(i in 1:500){
  for(j in 1:500){
    z[i,j] <- log(alphs[i])+log(bets[j])-(5/2)*log(alphs[i] + bets[j]) -
      10*lbeta(alphs[i],bets[j]) + sum(lbeta(alphs[i]+y_bikes, bets[j]+ns-y_bikes))
  }
}

Z <- exp(z-max(z))

contour(X,Y,Z)
title(xlab=TeX("$\\log(\\alpha / \\beta)$"), ylab=TeX("$\\log(\\alpha + \\beta)$"), main = TeX("Marginal"))
pulls <- 3000
sampleZ <- apply(Z,1,sum)
alphaI <- sample(1:length(alphs), pulls, replace=T,
                prob = sampleZ)

alphas <- alphs[alphaI]
betas <- rep(NA, pulls)
for (i in 1:pulls) {
  betas[i] <- sample(bets, 1, prob = Z[alphaI[i],],replace = F)
}

Thetas <- matrix(0, nrow = pulls, ncol = 10)

for(i in 1:10){

```

```

  Thetas[,i] <- rbeta(pulls, alphas + y_bikes[i], betas + ns[i] - y_bikes[i])
}

Theta_summary <- matrix(NA, nrow = 10, ncol = 9)

for(i in 1:nrow(Theta_summary)){
  Theta_summary[i,] = round(c(i, y_props[i], mean(Thetas[,i]), sd(Thetas[,i]),
    quantile(Thetas[,i], c(0.05,0.25,0.5,0.75,0.95))), 3)
}

Theta_summary <- as.data.frame(Theta_summary)

colnames(Theta_summary) <- c("Location $j$", "Observed Porportion", "Mean", "SD", "0.05",
  "0.25", "0.5", "0.75", "0.95")

Theta_table <- kable_styling(kable(Theta_summary, format = "latex", booktabs= TRUE, align = 'c',
  caption = "Observed Proportions Against Posterior  $\theta$  distributions", row.names = NULL),
  add_header_above(Theta_table, c("", "" , "Posterior Distribution Summary" =7 ))

hist(alphas/(alphas+betas), breaks = 20 , col = "grey",
  main = "Posterior Expected Bicycle Traffic", xlab = TeX(" $\frac{\alpha}{\alpha + \beta}$ ") )

quantile(alphas/(alphas+betas), c(0.025, 0.975))
pulls <- 3000
sampleZ <- apply(Z,1,sum)
alphaI <- sample(1:length(alphas), pulls, replace=T,
  prob = sampleZ)

alphas <- alphas[alphaI]
betas <- rep(NA, pulls)
for (i in 1:pulls) {
  betas[i] <- sample(betas, 1, prob = Z[alphaI[i],],replace = F)
}

pred_thetas <- rbeta(pulls, alphas, betas)

y_preds <- rbinom(pulls, 100, pred_thetas)

quantile(y_preds, c(0.025, 0.975))
hist(y_preds, main = "Posterior Predictions", col = "grey", breaks =25,
  xlab = TeX(" $\tilde{y}$ "))

observed_fatal <-data.frame( fatal = c(24,25,31,31,22,21,26,20,16,22),
  time = 1:10)

fatal_acf <- abs(acf(observed_fatal$fatal,plot = F, lag.max = 1)$acf[2])

```

```

fatal_slope <- lm(fatal~time, observed_fatal)
fatal_slope <- abs(fatal_slope$coefficients[2])
set.seed(62)

pulls <- 1000

new_fatalts <- matrix(NA, pulls, 10)

for (i in 1:pulls) {
  new_fatalts[i,] <- rpois(10, rgamma(10, 238)/10)
}

acfs_fatal <- rep(NA, pulls)
for (i in 1:pulls) {
  acfs_fatal[i] <- abs(acf(new_fatalts[i,],plot = F, lag.max = 1)$acf[2])
}

slope_fataldf <-data.frame( fatal = c(24,25,31,31,22,21,26,20,16,22),
                             time = 1:10)

slopes_fatal <- rep(NA, pulls)
for (i in 1:pulls) {
  slope_fataldf$fatal <-new_fatalts[i,]
  fatal_slope_new <- lm(fatal~time, slope_fataldf)
  slopes_fatal[i] <- abs(fatal_slope_new$coefficients[2])
}

sum(slopes_fatal > fatal_slope)/pulls
sum(acfs_fatal > fatal_acf)/pulls
par(mfrow = c(1,2))
hist(acfs_fatal,breaks = 20, main = "Posterior Predictive Check for Independence", cex.main = 0.9,
     xlab = "Abslolute AC of Sample")
abline(v = 0.411, col = "red")
text(0.6, 120, label = paste("p-value =",sum(acfs_fatal > fatal_acf)/pulls), cex= 0.8 )
hist(slopes_fatal, breaks = 20, main = "Poster Predictive Check for Trend", cex.main = 0.9,
     xlab = "Absolute Regression Slope of Sample")
abline(v =0.921, col = "red")
text(1.5, 125, label = paste("p-value =", sum(slopes_fatal > fatal_slope)/pulls), cex = 0.8)
pass_mi <- c(3.9e11, 4.3e11, 5.0e11, 5.5e11,
            5.8e11, 6.0e11, 5.9e11, 6.2e11,
            7.4e11, 7.1e11)

pulls <- 1000

new_fatalts_x <- matrix(NA, pulls, 10)
y_sum <- sum(c(24,25,31,31,22,21,26,20,16,22))
x_sum <- sum(pass_mi)

for (i in 1:pulls) {
  new_fatalts_x[i,] <- rpois(10, pass_mi*(rgamma(10, y_sum)/x_sum))
}

```

```

acfs_fatal_x <- rep(NA, pulls)
for (i in 1:pulls) {
  acfs_fatal_x[i] <- abs(acf(new_fatal_x[i,],plot = F, lag.max = 1)$acf[2])
}

slope_fataldf_x <-data.frame(fatal = c(24,25,31,31,22,21,26,20,16,22),
                           time = 1:10)

slopes_fatal_x <- rep(NA, pulls)
for (i in 1:pulls) {
  slope_fataldf_x$fatal <-new_fatal_x[i,]
  fatal_slope_new_x <- lm(fatal~time, slope_fataldf_x)
  slopes_fatal_x[i] <- abs(fatal_slope_new_x$coefficients[2])
}

sum(slopes_fatal_x > fatal_slope)/pulls
sum(acfs_fatal_x > fatal_acf)/pulls

par(mfrow = c(1,2))
hist(acfs_fatal_x,breaks = 20, main = "Posterior Predictive Check for Independence", cex.main = 0.9,
     xlab = "Absolute AC of Sample")
abline(v = 0.411, col = "red")
text(0.6, 80, label = paste("p-value =",sum(acfs_fatal_x > fatal_acf)/pulls), cex= 0.8 )
hist(slopes_fatal_x, breaks = 20, main = "Poster Predictive Check for Trend", cex.main = 0.9,
     xlab = "Absolute Regression Slope of Sample")
abline(v =0.921, col = "red")
text(2.7, 110, label = paste("p-value =", sum(slopes_fatal_x > fatal_slope)/pulls), cex = 0.8)
observed_death <-data.frame( death = c(734,516,754,877,814,362,764,809,223,1066),
                           time = 1:10)

death_acf <- abs(acf(observed_death$death,plot = F, lag.max = 1)$acf[2])
death_slope <- lm(death~time, observed_death)
death_slope <- abs(death_slope$coefficients[2])
set.seed(62)

pulls <- 1000

new_deaths <- matrix(NA, pulls, 10)

for (i in 1:pulls) {
  new_deaths[i,] <- rpois(10, rgamma(10, 6919)/10)
}

acfs_death <- rep(NA, pulls)
for (i in 1:pulls) {
  acfs_death[i] <- abs(acf(new_deaths[i,],plot = F, lag.max = 1)$acf[2])
}

slope_deathdf <-data.frame(death =c(734,516,754,877,814,362,764,809,223,1066),
                          time = 1:10)

slopes_death <- rep(NA, pulls)

```

```

for (i in 1:pulls) {
  slope_deathdf$death <- new_deaths[i,]
  death_slope_new <- lm(death~time, slope_deathdf)
  slopes_death[i] <- abs(death_slope_new$coefficients[2])
}

sum(slopes_death > death_slope)/pulls
sum(acfs_death > death_acf)/pulls
par(mfrow = c(1,2))
hist(acfs_death,breaks = 20, main = "Posterior Predictive Check for Independence", cex.main = 0.9,
     xlab = "Absolute AC of Sample")
abline(v = 0.471, col = "red")
text(0.7, 120, label = paste("p-value =",sum(acfs_death > death_acf)/pulls), cex= 0.8 )
hist(slopes_death, breaks = 20, main = "Posterior Predictive Check for Trend", cex.main = 0.9,
     xlab = "Absolute Regression Slope of Sample")
abline(v = 2.55, col = "red")
text(6, 100, label = paste("p-value =", sum(slopes_death > death_slope)/pulls), cex = 0.8)
pulls <- 1000

new_deaths_x <- matrix(NA, pulls, 10)
y_sum <- sum(observed_death$death)
x_sum <- sum(pass_mi)

for (i in 1:pulls) {
  new_deaths_x[i,] <- rpois(10, pass_mi*(rgamma(10, y_sum)/x_sum))
}

acfs_death_x <- rep(NA, pulls)
for (i in 1:pulls) {
  acfs_death_x[i] <- abs(acf(new_deaths_x[i,],plot = F, lag.max = 1)$acf[2])
}

slope_deathdf_x <- data.frame(death = c(734,516,754,877,814,362,764,809,223,1066),
                             time = 1:10)

slopes_death_x <- rep(NA, pulls)
for (i in 1:pulls) {
  slope_deathdf_x$death <- new_deaths_x[i,]
  death_slope_new_x <- lm(death~time, slope_deathdf_x)
  slopes_death_x[i] <- abs(death_slope_new_x$coefficients[2])
}

sum(slopes_death_x > death_slope)/pulls
sum(acfs_death_x > death_acf)/pulls

par(mfrow = c(1,2))
hist(acfs_death_x,breaks = 20, main = "Posterior Predictive Check for Independence", cex.main = 0.9,
     xlab = "Absolute AC of Sample")
abline(v = 0.471, col = "red")
text(0.53, 82, label = paste("p-value =",sum(acfs_death_x > death_acf)/pulls), cex= 0.8 )
hist(slopes_death_x, breaks = 20, main = "Posterior Predictive Check for Trend", cex.main = 0.9,
     xlab = "Absolute Regression Slope of Sample")

```

```

abline(v = 2.55, col = "red")
text(37, 100, label = paste("p-value =", sum(slopes_death_x > death_slope)/pulls), cex = 0.8)
switches <- rep(NA, 5000)

for(i in 1:5000){
  pcheck <- c()
  while(sum(pcheck==0) != 13){
    theta <- rbeta(1, 8, 14)
    pcheck <- c(pcheck, rbinom(1, 1, theta))
  }
  switches[i] <- sum(diff(pcheck) != 0)
}

hist(switches, breaks = 25, xlab = "T(y_rep)", main = "Posterior Predictive Check for Switches")
abline(v = 3, col = "red")

# sum(switches <= 3)/ length(switches)
tums <- c(0,0,0,0,0,0,0,0,0,0,0,
          0,0,0,0,1,1,1,1,1,1,1,
          1,1,2,2,2,2,2,2,2,2,2,
          2,1,5,2,5,3,2,7,7,3,3,
          3,2,9,10,4,4,4,4,4,4,4,
          4,10,4,4,4,5,11,12,5,5,5,
          6,5,6,6,6,6,16,15,15,9,
          4)

tot_rat <- c(20,20,20,20,20,20,20,19,19,19,
            19,18,18,17,20,20,20,20,19,19,
            18,18,25,24,23,20,20,20,20,20,
            20,10,49,19,46,27,17,49,47,20,
            20,13,48,50,20,20,20,20,20,20,
            20,48,19,19,19,22,46,49,20,20,
            23,19,22,20,20,20,52,47,46,24,
            14)

set.seed(69)
ns <- tums+(tot_rat-tums)

# Mom_a <- 2.89
# Mom_b <- 11.86
#
# logaoverb <- log(Mom_a/Mom_b)
# logaplusb <- log(Mom_a + Mom_b)

X <- seq(-2.3, -1.3, length.out = 1000)
Y <- seq(1, 5, length.out = 1000)

alphs <- exp(X+Y)/(1+exp(X))
bets <- exp(Y)-exp(X+Y)/(1+exp(X))

```

```

z <- matrix(0,nrow=1000,ncol=1000)

for(i in 1:1000){
  for(j in 1:1000){
    z[i,j] <- log(alphs[i])+log(bets[j])-(5/2)*log(alphs[i] + bets[j]) -
      71*lbeta(alphs[i],bets[j]) + sum(lbeta(alphs[i]+tums, bets[j]+ns-tums))
  }
}

Z <- exp(z-max(z))

# contour(X,Y,Z, nlevels = 10)
# title(xlab="log(a/b)",ylab="log(a+b)", main = "Marginal Posterior of Alpha and Beta")

set.seed(4)
pulls <- 2000
sampleZ <- apply(Z,1,sum)
alphaI <- sample(1:length(alphs), pulls, replace=T,
  prob = sampleZ)

alphas <- alphs[alphaI]
betas <- rep(NA, pulls)
for (i in 1:pulls) {
  betas[i] <- sample(bets, 1, prob = Z[alphaI[i],],replace = F)
}

# mean(alphas)
# mean(betas)

plot(log(alphas/betas), log(alphas+betas), pch = 16, cex = .5,
  xlab=TeX("$\\log(\\alpha / \\beta)$"), ylab=TeX("$\\log(\\alpha + \\beta)$"), main = "2000 Posterior")

Thetas <- matrix(0, nrow = pulls, ncol = 71)

for(i in 1:71){
  Thetas[,i] <- rbeta(pulls, alphas + tums[i], betas + ns[i] - tums[i])
}

NewTums <- matrix(0, nrow = pulls, ncol = 71)

for(i in 1:71){
  for(j in 1:pulls){
    NewTums[j,i] <- rbinom(1, ns[i], Thetas[j,i])
  }
}

NewProps <- matrix(0, nrow = pulls, ncol = 71)
for (i in 1:pulls) {

```



```

NewProps[i,] <- NewTums[i,]/ns
}

numZers <- rep(NA, pulls)
for (i in 1:pulls) {
  numZers[i] <- sum(NewTums[i,] == 0)
}

mean_props <- apply(NewProps, 1, mean)

# hist(mean_props)
# abline(v = mean(tums/tot_rat))

max_props <- apply(NewProps, 1, max)

# hist(max_props)
# abline(v = max(tums/tot_rat))

sd_props <- apply(NewProps, 1, sd)

# hist(sd_props)
# abline(v = sd(tums/tot_rat))

numZers <- rep(NA, pulls)
for (i in 1:pulls) {
  numZers[i] <- sum(NewTums[i,] == 0)
}
par(mfrow = c(2,2))

hist(numZers, main = "", xlab = "T(y) = Number of 0's" , col = "lightgrey")
abline(v = sum(tums == 0), col = "red", cex = 1.5)
text(18, 400,
     label = paste("p-value =",round(sum(numZers>sum(tums == 0))/pulls,2), cex= 0.85 )

hist(max_props, main = "", xlab = "T(y) = max(y)", col = "lightgrey")
abline(v = max(tums/tot_rat), col = "red", cex = 1.5)
text(0.6, 500,
     label = paste("p-value =",round(sum(max_props > max(tums/tot_rat))/pulls,2)), cex= 0.85)

hist(sd_props, main = "", xlab = "T(y) = sd(y)", col = "lightgrey")
abline(v = sd(tums/tot_rat), col = "red", cex = 1.5)
text(0.135, 275,
     label = paste("p-value =",round(sum(sd_props > sd(tums/tot_rat))/pulls,2)), cex= 0.85)

hist(mean_props, main = "", xlab = "T(y) = mean(y)", col = "lightgrey")
abline(v = mean(tums/tot_rat), col = "red", cex = 1.5)
text(0.17, 500,
     label = paste("p-value =",round(sum(mean_props > mean(tums/tot_rat))/pulls,2)), cex= 0.85)

```

```

check <- sqrt(rnorm(10000, 50, 5))

par(mfrow= c(1,3))
hist(check^1.9, main = TeX("$\\phi$ =1.9"), xlab = "Transformed Sample", col = "grey")
hist(check^2, main = TeX("$\\phi$ =2"), xlab = "Transformed Sample", col = "grey")
hist(check^2.1, main = TeX("$\\phi$ =2.1"), xlab = "Transformed Sample", col = "grey")
BE <- c(5.0, 13.0, 7.2, 6.8, 12.8, 9.5, 6.0,
        3.8, 1.8, 6.9, 4.7, 9.5)

center_phi <- MASS::boxcox(BE~1 ,plot = F)
which(center_phi$y == max(center_phi$y))
center_phi$x[26]

jacob_phi <- function(p, data){
  if(p == 0){
    J <- prod(abs(1/data))
  }else{
    J <- prod(abs(data^(p-1)))
  }
  return(J)}

prior_phi <- function(p, data){
  n <- length(data)
  pri <- prod(data^(1/n))^(1-p)
  return(pri)
}

s2phi<- function(p, data){
  if(p == 0){
    yp <- log(data)
  }else{
    yp <- (data^p-1)/p
  }
  n <- length(data)
  mup <- mean(yp)
  s2p <- sum((yp-mup)^2)/(n-1)
  fin <- s2p^((n-1)/2)
  return(fin)}

post_phi <- function(p, data){
  jacob <- jacob_phi(p, data)
  pri <- prior_phi(p,data)
  s2 <- s2phi(p,data)

  dens <- jacob*pri/s2

  return(dens)
}

phis <- seq(-0.75, 2.25, length.out = 200)
phi_post_dens <- sapply(phis, post_phi, data= BE)
plot(phis,phi_post_dens/sum(phi_post_dens), type = "l",
     main = "Posterior Density of Phi", ylab = "Density", xlab = "phi")

```

```

posterior_mu_sig = function (theta, p, data)
{
  mu = theta[1]
  sig2 = theta[2]
  if(p == 0){
    yp <- log(data)
  }else{
    yp <- (data^p-1)/p
  }
  logf = function(y, mu, sig2) -(y - mu)^2/2/sig2 - log(sig2)/2
  z = sum(logf(yp, mu, sig2))
  # z = z - log(sig2)
  return(z)
}

set.seed(7676)
pulls <- 2000
SampledBox <- matrix(data = NA, ncol = 3, nrow = pulls)
SampledBox[, 1] <- sample(phis, pulls ,
                          prob = phi_post_dens, replace = T)

n <- length(BE)
for (i in 1:pulls) {
  sampled_phi <- SampledBox[i,1]
  if(sampled_phi == 0){
    yp <- log(BE)
  }else{
    yp <- (BE^sampled_phi-1)/sampled_phi
  }

  sampled_sigma2 <- geoR::rinvchisq(1,n-1, sum((yp-mean(yp))^2)/(n-1))

  sampled_mu <- rnorm(1, mean = mean(yp), sd = sqrt(sampled_sigma2/n))

  SampledBox[i, 2] <- sampled_mu
  SampledBox[i, 3] <- log(sqrt(sampled_sigma2))
}

par(mfrow = c(1,3))
hist(SampledBox[,1], breaks = 20, col = "grey",
     main = TeX("Posterior Samples of  $\phi$ "),xlab = TeX(" $\phi$ "))
hist(SampledBox[,2], breaks = 25, col = "grey",
     main = TeX("Posterior Samples of  $\mu$ "),xlab = TeX(" $\mu$ "))
hist(SampledBox[,3], breaks = 20, col = "grey",
     main = TeX("Posterior Samples of  $\log(\sigma)$ "),xlab = TeX(" $\log(\sigma)$ "))

samp_phi <- SampledBox[,1]
samp_mus <- SampledBox[,2]
samp_sig <- exp(SampledBox[,3])

```

```

summary_NR <- as.data.frame(rbind(c(mean(samp_phi), sd(samp_phi),
                                   quantile(samp_phi, c(0.05,0.25,0.5,0.75,0.95))),
                                c(mean(samp_mus),sd(samp_mus),quantile(samp_mus, c(0.05,0.25,0.5,0.75,0.95))),
                                c(mean(samp_sig),sd(samp_sig),quantile(samp_sig, c(0.05,0.25,0.5,0.75,0.95)))))

summary_NR <- round(summary_NR, 3)

colnames(summary_NR) <- c("Mean", "SD", "5\\%", "25\\%", "50\\%", "75\\%", "95\\%")
rownames(summary_NR) <- c("$\\phi$", "$\\mu$", "$\\sigma$")

kable_styling(kable(summary_NR, format = "latex", booktabs= TRUE, align = 'c',
                           caption = "Posterior Distribution Summary", row.names = T, escape = F), latex_options = "auto")
BE <- c(5.0, 13.0, 7.2, 6.8, 12.8, 9.5, 6.0,
        3.8, 1.8, 6.9, 4.7, 9.5)

C <- c(12.9,2.6,26.6,1.5,13.0,8.8,19.5,9.0,
       13.1,3.6)

GH <- c(14.3,7.6,2.6, 43.5,4.9,3.5,4.8,5.6,
       3.5,3.9,6.7)

Y <- c(BE,C,GH)

center_phi2 <- MASS::boxcox(Y~1 ,plot = F)
which(center_phi2$y == max(center_phi2$y))
center_phi$x[20]

phisb <- seq(-0.8, 0.6, length.out = 2000)
phi_post_densb <- sapply(phisb, post_phi, data= Y)
plot(phisb,phi_post_densb/sum(phi_post_densb), type = "l",
     main = "Posterior Density of Phi", ylab = "Density", xlab = "phi")

set.seed(767676)
pulls <- 2000
SampledBoxBE <- matrix(data = NA, ncol = 3, nrow = pulls)
SampledBoxC <- matrix(data = NA, ncol = 3, nrow = pulls)
SampledBoxGH <- matrix(data = NA, ncol = 3, nrow = pulls)

SampledBoxBE[, 1] <- sample(phisb, pulls ,
                           prob = phi_post_densb, replace = T)

SampledBoxC[, 1] <- SampledBoxBE[, 1]

SampledBoxGH[, 1] <- SampledBoxBE[, 1]

nBE <- length(BE)
nC <- length(C)
nGH <- length(GH)

```

```

for (i in 1:pulls) {
  sampled_phi <- SampledBoxBE[i,1]
  if(sampled_phi == 0){
    ypBE <- log(BE)
    ypC <- log(C)
    ypGH <- log(GH)
  }else{
    ypBE <- (BE^sampled_phi-1)/sampled_phi
    ypC <- (C^sampled_phi-1)/sampled_phi
    ypGH <- (GH^sampled_phi-1)/sampled_phi
  }

  sampled_sigma2BE <- geoR::rinvchisq(1,nBE-1, sum((ypBE-mean(ypBE))^2)/(nBE-1))
  sampled_sigma2C <- geoR::rinvchisq(1,nC-1, sum((ypC-mean(ypC))^2)/(nC-1))
  sampled_sigma2GH <- geoR::rinvchisq(1, nGH-1, sum((ypGH-mean(ypGH))^2)/(nGH-1))

  sampled_muBE <- rnorm(1, mean = mean(ypBE), sd = sqrt(sampled_sigma2BE/nBE))
  sampled_muC <- rnorm(1, mean = mean(ypC), sd = sqrt(sampled_sigma2C/nC))
  sampled_muGH <- rnorm(1, mean = mean(ypGH), sd = sqrt(sampled_sigma2GH/nGH))

  SampledBoxBE[i, 2] <- sampled_muBE
  SampledBoxBE[i, 3] <- log(sqrt(sampled_sigma2BE))

  SampledBoxC[i, 2] <- sampled_muC
  SampledBoxC[i, 3] <- log(sqrt(sampled_sigma2C))

  SampledBoxGH[i, 2] <- sampled_muGH
  SampledBoxGH[i, 3] <- log(sqrt(sampled_sigma2GH))
}

par(mfrow = c(3,3))

# par(mfrow = c(1,3))
hist(SampledBoxBE[,1], breaks = 20, col = "grey",
     main = TeX("Posterior Samples of  $\phi$ "),xlab = TeX(" $\phi$ "))
hist(SampledBoxBE[,2], breaks = 25, col = "grey",
     main = TeX("Posterior Samples of  $\mu$  Blue Earth"),xlab = TeX(" $\mu$ "))
hist(SampledBoxBE[,3], breaks = 25, col = "grey",
     main = TeX("Posterior Samples of  $\log(\sigma)$  Blue Earth"),xlab = TeX(" $\log(\sigma)$ "))

# par(mfrow = c(1,3))
hist(SampledBoxC[,1], breaks = 20, col = "grey",
     main = TeX("Posterior Samples of  $\phi$ "),xlab = TeX(" $\phi$ "))
hist(SampledBoxC[,2], breaks = 25, col = "grey",
     main = TeX("Posterior Samples of  $\mu$  Clay"),xlab = TeX(" $\mu$ "))
hist(SampledBoxC[,3], breaks = 25, col = "grey",
     main = TeX("Posterior Samples of  $\log(\sigma)$  Clay"),xlab = TeX(" $\log(\sigma)$ "))

```

```

# par(mfrow = c(1,3))
hist(SampledBoxGH[,1], breaks = 20, col = "grey",
     main = TeX("Posterior Samples of  $\phi$ "), xlab = TeX(" $\phi$ "))
hist(SampledBoxGH[,2], breaks = 25, col = "grey",
     main = TeX("Posterior Samples of  $\mu$  Goodhue"), xlab = TeX(" $\mu$ "))
hist(SampledBoxGH[,3], breaks = 25, col = "grey",
     main = TeX("Posterior Samples of  $\log(\sigma)$  Goodhue"), xlab = TeX(" $\log(\sigma)$ "))

samp_phi <- SampledBoxBE[,1]

samp_musBE <- SampledBoxBE[,2]
samp_sigBE <- exp(SampledBoxBE[,3])

samp_musC <- SampledBoxC[,2]
samp_sigC <- exp(SampledBoxC[,3])

samp_musGH <- SampledBoxGH[,2]
samp_sigGH <- exp(SampledBoxGH[,3])

phi_dist <- c(mean(samp_phi), sd(samp_phi),
              quantile(samp_phi, c(0.05,0.25,0.5,0.75,0.95)))

mean_distBE <- c(mean(samp_musBE), sd(samp_musBE),
                 quantile(samp_musBE, c(0.05,0.25,0.5,0.75,0.95)))
sig_distBE <- c(mean(samp_sigBE), sd(samp_sigBE),
                quantile(samp_sigBE, c(0.05,0.25,0.5,0.75,0.95)))

mean_distC <- c(mean(samp_musC), sd(samp_musC),
                 quantile(samp_musC, c(0.05,0.25,0.5,0.75,0.95)))
sig_distC <- c(mean(samp_sigC), sd(samp_sigC),
                quantile(samp_sigC, c(0.05,0.25,0.5,0.75,0.95)))

mean_distGH <- c(mean(samp_musGH), sd(samp_musGH),
                  quantile(samp_musGH, c(0.05,0.25,0.5,0.75,0.95)))
sig_distGH <- c(mean(samp_sigGH), sd(samp_sigGH),
                 quantile(samp_sigGH, c(0.05,0.25,0.5,0.75,0.95)))

summary_NR2 <- cbind(phi_dist, mean_distBE, mean_distC, mean_distGH,
                     sig_distBE, sig_distC, sig_distGH)

summary_NR2 <- as.data.frame(summary_NR2)
colnames(summary_NR2) <- c(" $\phi$ ", " $\mu_{\text{BE}}$ ", " $\mu_{\text{C}}$ ",
                           " $\mu_{\text{GH}}$ ", " $\sigma_{\text{BE}}$ ", " $\sigma_{\text{C}}$ ",
                           " $\sigma_{\text{GH}}$ ")

rownames(summary_NR2) <- c("Mean", "SD", "5%", "25%", "50%", "75%", "95%")

summary_NR2 <- round(summary_NR2, 3)

```

```

NR_2Table <- kable_styling(kable(summary_NR2, format = "latex", booktabs= TRUE, align = 'c',
                                caption = "Posterior Distribution Summary", row.names = T, escape = F), latex_options = "math")

add_header_above(NR_2Table, c("", "", "Posterior Means" = 3, "Posterior SDs" = 3))
set.seed(76)
pulls <- 2000

BE <- c(5.0, 13.0, 7.2, 6.8, 12.8, 9.5, 6.0,
        3.8, 1.8, 6.9, 4.7, 9.5)

C <- c(12.9, 2.6, 26.6, 1.5, 13.0, 8.8, 19.5, 9.0,
        13.1, 3.6)

GH <- c(14.3, 7.6, 2.6, 43.5, 4.9, 3.5, 4.8, 5.6,
        3.5, 3.9, 6.7)

Y <- c(BE, C, GH)

NewData <- matrix(NA, nrow = pulls, ncol = length(Y))

n <- length(Y)
for (i in 1:pulls) {
  sampled_phi <- sample(phisb, 1,
                      prob = phi_post_densb, replace = T)
  if(sampled_phi == 0){
    yp <- log(Y)
  }else{
    yp <- (Y^sampled_phi-1)/sampled_phi
  }

  sampled_sigma2 <- geoR::rinvchisq(3, n-1, sum((yp-mean(yp))^2)/(n-1))

  sampled_mu <- rnorm(3, mean = mean(yp), sd = sqrt(sampled_sigma2/n))

  TYrep1 <- rnorm(12, sampled_mu[1], sd=sqrt(sampled_sigma2[1]))
  TYrep2 <- rnorm(10, sampled_mu[2], sd=sqrt(sampled_sigma2[2]))
  TYrep3 <- rnorm(11, sampled_mu[3], sd=sqrt(sampled_sigma2[3]))

  TYrep <- c(TYrep1, TYrep2, TYrep3)

  if(sampled_phi == 0){
    NewData[i,] <- exp(TYrep)
  }else{
    NewData[i,] <- (TYrep*sampled_phi+1)^(1/sampled_phi)
  }
}

```

```

spread <- function(x){return(max(x)-min(x))}

log_meansData <- na.omit(log(apply(NewData, 1,mean)))
log_maxData <- na.omit(log(apply(NewData, 1,max)))
log_spreadData <- na.omit(log(apply(NewData, 1,spread)))
log_sdData <- na.omit(log(apply(NewData, 1,sd)))
samps <- length(log_meansData)

par(mfrow = c(2,2))

hist(log_meansData, breaks = 50, col = "lightgrey", main = "",
      xlab = "T(y) = log(mean(y))")
abline(v = log(mean(Y)), col = "red")
text(4, 200,
      label = paste("p-value =",round(sum(log_meansData > log(mean(Y)))/samps,2)), cex= 0.85)

hist(log_maxData, breaks = 50, col = "lightgrey", main = "",
      xlab = "T(y) = log(max(y))")
abline(v = max(log(Y)), col = "red")
text(6, 200,
      label = paste("p-value =",round(sum(log_maxData > log(max(Y)))/samps,2)), cex= 0.85)

hist(log_spreadData, breaks = 50, col = "lightgrey", main = "",
      xlab = "T(y) = log(range(y))")
abline(v = log(spread(Y)), col = "red")
text(6, 200,
      label = paste("p-value =",round(sum(log_spreadData > log(spread(Y)))/samps,2)), cex= 0.85)

hist(log_sdData, breaks = 50, col = "lightgrey", main = "",
      xlab = "T(y) = log(sd(y))")
abline(v = log(sd(Y)), col = "red")
text(5, 250,
      label = paste("p-value =",round(sum(log_sdData > log(sd(Y)))/samps,2)), cex= 0.85)

spread <- function(x){return(max(x)-min(x))}

log_meansDataBE <- na.omit(log(apply(NewData[,1:length(BE)], 1,mean)))
log_maxDataBE <- na.omit(log(apply(NewData[,1:length(BE)], 1,max)))
log_spreadDataBE <- na.omit(log(apply(NewData[,1:length(BE)], 1,spread)))
log_sdDataBE <- na.omit(log(apply(NewData[,1:length(BE)], 1,sd)))
sampsBE <- length(log_meansDataBE)

log_meansDataC <- na.omit(log(apply(NewData[,13:22], 1,mean)))
log_maxDataC <- na.omit(log(apply(NewData[,13:22], 1,max)))
log_spreadDataC <- na.omit(log(apply(NewData[,13:22], 1,spread)))
log_sdDataC <- na.omit(log(apply(NewData[,13:22], 1,sd)))
sampsC <- length(log_meansDataC)

log_meansDataGH <- na.omit(log(apply(NewData[,23:33], 1,mean)))
log_maxDataGH <- na.omit(log(apply(NewData[,23:33], 1,max)))
log_spreadDataGH <- na.omit(log(apply(NewData[,23:33], 1,spread)))
log_sdDataGH <- na.omit(log(apply(NewData[,23:33], 1,sd)))

```



```

sampsGH <- length(log_meansDataGH)

par(mfrow = c(3,4))

hist(log_meansDataBE, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y1) = log(mean(y1))")
abline(v = log(mean(BE)), col = "red")
text(5, 200,
     label = paste("BH p-value =", round(sum(log_meansDataBE > log(mean(BE)))/sampsBE, 2)), cex= 0.8)

hist(log_maxDataBE, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y1) = log(max(y1))")
abline(v = max(log(BE)), col = "red")
text(7, 200,
     label = paste("BH p-value =", round(sum(log_maxDataBE > log(max(BE)))/sampsBE, 2)), cex= 0.8)

hist(log_spreadDataBE, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y1) = log(range(y1))")
abline(v = log(spread(BE)), col = "red")
text(6.75, 200,
     label = paste("BH p-value =", round(sum(log_spreadDataBE > log(spread(BE)))/sampsBE, 2)), cex= 0.8)

hist(log_sdDataBE, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y1) = log(sd(y1))")
abline(v = log(sd(BE)), col = "red")
text(6, 250,
     label = paste("BH p-value =", round(sum(log_sdDataBE > log(sd(BE)))/sampsBE, 2)), cex= 0.8)

hist(log_meansDataC, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y2) = log(mean(y2))")
abline(v = log(mean(C)), col = "red")
text(5, 200,
     label = paste("C p-value =", round(sum(log_meansDataC > log(mean(C)))/sampsC, 2)), cex= 0.8)

hist(log_maxDataC, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y2) = log(max(y2))")
abline(v = max(log(C)), col = "red")
text(7, 200,
     label = paste("C p-value =", round(sum(log_maxDataC > log(max(C)))/sampsC, 2)), cex= 0.8)

hist(log_spreadDataC, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y2) = log(range(y2))")
abline(v = log(spread(C)), col = "red")
text(7, 200,
     label = paste("C p-value =", round(sum(log_spreadDataC > log(spread(C)))/sampsC, 2)), cex= 0.8)

hist(log_sdDataC, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y2) = log(sd(y2))")
abline(v = log(sd(C)), col = "red")
text(6, 200,
     label = paste("C p-value =", round(sum(log_sdDataC > log(sd(C)))/sampsC, 2)), cex= 0.8)

```

```

hist(log_meansDataGH, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y3) = log(mean(y3))")
abline(v = log(mean(GH)), col = "red")
text(6.5, 300,
     label = paste("GH p-value =", round(sum(log_meansDataGH > log(mean(GH)))/sampsGH, 2)), cex= 0.8)

hist(log_maxDataGH, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y3) = log(max(y3))")
abline(v = max(log(GH)), col = "red")
text(7.5, 200,
     label = paste("GH p-value =", round(sum(log_maxDataGH > log(max(GH)))/sampsGH, 2)), cex= 0.8)

hist(log_spreadDataGH, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y3) = log(range(y3))")
abline(v = log(spread(GH)), col = "red")
text(7.5, 175,
     label = paste("GH p-value =", round(sum(log_spreadDataGH > log(spread(GH)))/sampsGH, 2)), cex= 0.8)

hist(log_sdDataGH, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y3) = log(sd(y3))")
abline(v = log(sd(GH)), col = "red")
text(6.5, 200,
     label = paste("GH p-value =", round(sum(log_sdDataGH > log(sd(GH)))/sampsGH, 2)), cex= 0.8)

Y <- c(BE,C,GH)
phisb <- seq(-0.75, 0.6, length.out = 2000)
phi_post_densb <- sapply(phisb, post_phi, data= Y)
plot(phisb,phi_post_densb/sum(phi_post_densb), type = "l",
     main = "Posterior Density of Phi", ylab = "Density", xlab = "phi")
set.seed(76)
pulls <- 2000

BE <- c(5.0, 13.0, 7.2, 6.8, 12.8, 9.5, 6.0,
        3.8, 1.8, 6.9, 4.7, 9.5)

C <- c(12.9,2.6,26.6,1.5,13.0,8.8,19.5,9.0,
        13.1,3.6)

GH <- c(14.3,7.6,2.6, 43.5,4.9,3.5,4.8,5.6,
        3.5,3.9,6.7)

Y <- c(BE,C,GH)

NewData <- matrix(NA, nrow = pulls, ncol = length(Y))

n <- length(Y)
for (i in 1:pulls) {
  sampled_phi <- 0
  if(sampled_phi == 0){
    yp <- log(Y)

```

```

}else{
yp <- (Y^sampled_phi-1)/sampled_phi
}

sampled_sigma2 <- geoR::rinvchisq(3,n-1, sum((yp-mean(yp))^2)/(n-1))

sampled_mu <- rnorm(3, mean = mean(yp), sd = sqrt(sampled_sigma2/n))

TYrep1 <- rnorm(12, sampled_mu[1], sd=sqrt(sampled_sigma2[1]))
TYrep2 <- rnorm(10, sampled_mu[2], sd=sqrt(sampled_sigma2[2]))
TYrep3 <- rnorm(11, sampled_mu[3], sd=sqrt(sampled_sigma2[3]))

TYrep <- c(TYrep1, TYrep2, TYrep3)

if(sampled_phi == 0){
  NewData[i,] <- exp(TYrep)
}else{
  NewData[i,] <- (TYrep*sampled_phi+1)^(1/sampled_phi)
}
}

spread <- function(x){return(max(x)-min(x))}

log_meansData <- na.omit(log(apply(NewData, 1,mean)))
log_maxData <- na.omit(log(apply(NewData, 1,max)))
log_spreadData <- na.omit(log(apply(NewData, 1,spread)))
log_sdData <- na.omit(log(apply(NewData, 1,sd)))
samps <- length(log_meansData)

par(mfrow = c(2,2))

hist(log_meansData, breaks = 50, col = "lightgrey", main = "",
      xlab = "T(y) = log(mean(y))")
abline(v = log(mean(Y)), col = "red")
text(2.75, 150,
      label = paste("p-value =",round(sum(log_meansData > log(mean(Y)))/samps,2)), cex= 0.85)

hist(log_maxData, breaks = 50, col = "lightgrey", main = "",
      xlab = "T(y) = log(max(y))")
abline(v = max(log(Y)), col = "red")
text(4.5, 80,
      label = paste("p-value =",round(sum(log_maxData > log(max(Y)))/samps,2)), cex= 0.85)

hist(log_spreadData, breaks = 50, col = "lightgrey", main = "",
      xlab = "T(y) = log(range(y))")
abline(v = log(spread(Y)), col = "red")
text(4.5, 60,
      label = paste("p-value =",round(sum(log_spreadData > log(spread(Y)))/samps,2)), cex= 0.85)

hist(log_sdData, breaks = 50, col = "lightgrey", main = "",
      xlab = "T(y) = log(sd(y))")

```

```

abline(v = log(sd(Y)), col = "red")
text(2.8, 80,
     label = paste("p-value =", round(sum(log_sdData > log(sd(Y)))/samps,2)), cex= 0.85)

spread <- function(x){return(max(x)-min(x))}

log_meansDataBE <- na.omit(log(apply(NewData[,1:length(BE)], 1,mean)))
log_maxDataBE <- na.omit(log(apply(NewData[,1:length(BE)], 1,max)))
log_spreadDataBE <- na.omit(log(apply(NewData[,1:length(BE)], 1,spread)))
log_sdDataBE <- na.omit(log(apply(NewData[,1:length(BE)], 1,sd)))
sampsBE <- length(log_meansDataBE)

log_meansDataC <- na.omit(log(apply(NewData[,13:22], 1,mean)))
log_maxDataC <- na.omit(log(apply(NewData[,13:22], 1,max)))
log_spreadDataC <- na.omit(log(apply(NewData[,13:22], 1,spread)))
log_sdDataC <- na.omit(log(apply(NewData[,13:22], 1,sd)))
sampsC <- length(log_meansDataC)

log_meansDataGH <- na.omit(log(apply(NewData[,23:33], 1,mean)))
log_maxDataGH <- na.omit(log(apply(NewData[,23:33], 1,max)))
log_spreadDataGH <- na.omit(log(apply(NewData[,23:33], 1,spread)))
log_sdDataGH <- na.omit(log(apply(NewData[,23:33], 1,sd)))
sampsGH <- length(log_meansDataGH)

par(mfrow = c(3,4))

hist(log_meansDataBE, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y1) = log(mean(y1))")
abline(v = log(mean(BE)), col = "red")
text(2.9, 120,
     label = paste("BH p-val =", round(sum(log_meansDataBE > log(mean(BE)))/sampsBE,2)), cex= 0.65)

hist(log_maxDataBE, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y1) = log(max(y1))")
abline(v = max(log(BE)), col = "red")
text(4.5, 130,
     label = paste("BH p-val =", round(sum(log_maxDataBE > log(max(BE)))/sampsBE,2)), cex= 0.65)

hist(log_spreadDataBE, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y1) = log(range(y1))")
abline(v = log(spread(BE)), col = "red")
text(4.5, 100,
     label = paste("BH p-val =", round(sum(log_spreadDataBE > log(spread(BE)))/sampsBE,2)), cex= 0.65)

hist(log_sdDataBE, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y1) = log(sd(y1))")
abline(v = log(sd(BE)), col = "red")
text(3.5, 100,
     label = paste("BH p-val =", round(sum(log_sdDataBE > log(sd(BE)))/sampsBE,2)), cex= 0.65)

```

```

hist(log_meansDataC, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y2) = log(mean(y2))")
abline(v = log(mean(C)), col = "red")
text(3, 100,
     label = paste("C p-val =", round(sum(log_meansDataC > log(mean(C)))/sampsC, 2)), cex= 0.65)

hist(log_maxDataC, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y2) = log(max(y2))")
abline(v = max(log(C)), col = "red")
text(4.5, 100,
     label = paste("C p-val =", round(sum(log_maxDataC > log(max(C)))/sampsC, 2)), cex= 0.65)

hist(log_spreadDataC, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y2) = log(range(y2))")
abline(v = log(spread(C)), col = "red")
text(4.5, 100,
     label = paste("C p-val =", round(sum(log_spreadDataC > log(spread(C)))/sampsC, 2)), cex= 0.65)

hist(log_sdDataC, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y2) = log(sd(y2))")
abline(v = log(sd(C)), col = "red")
text(3.5, 100,
     label = paste("C p-val =", round(sum(log_sdDataC > log(sd(C)))/sampsC, 2)), cex= 0.65)

hist(log_meansDataGH, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y3) = log(mean(y3))")
abline(v = log(mean(GH)), col = "red")
text(3, 100,
     label = paste("GH p-val =", round(sum(log_meansDataGH > log(mean(GH)))/sampsGH, 2)), cex= 0.65)

hist(log_maxDataGH, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y3) = log(max(y3))")
abline(v = max(log(GH)), col = "red")
text(4.25, 150,
     label = paste("GH p-val =", round(sum(log_maxDataGH > log(max(GH)))/sampsGH, 2)), cex= 0.65)

hist(log_spreadDataGH, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y3) = log(range(y3))")
abline(v = log(spread(GH)), col = "red")
text(4.25, 140,
     label = paste("GH p-val =", round(sum(log_spreadDataGH > log(spread(GH)))/sampsGH, 2)), cex= 0.65)

hist(log_sdDataGH, breaks = 50, col = "lightgrey", main = "",
     xlab = "T(y3) = log(sd(y3))")
abline(v = log(sd(GH)), col = "red")
text(3, 140,
     label = paste("GH p-val =", round(sum(log_sdDataGH > log(sd(GH)))/sampsGH, 2)), cex= 0.65)

```