

# STA2201 ASSIGNMENT 2 Alice Huang

Alice Huang

05/03/2023

## Question 1

### Question 1a

We know that if  $\mu \sim N(\mu_0, \sigma_{\mu_0}^2)$  and  $y_i|\mu, \sigma^2 \sim N(\mu, \sigma^2)$  and  $\sigma^2$  is known, then

$$\mu|y, \sigma^2 \sim N\left(\frac{\mu_0/\sigma_{\mu_0}^2 + n\bar{y}/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}, \frac{1}{1/\sigma_{\mu_0}^2 + n/\sigma^2}\right)$$

We are given that

$$\mu_0 = 100, \sigma_{\mu_0} = 15, n = 10, \sigma = 15, \bar{y} = 113$$

So

$$\mu|y, \sigma^2 \sim N\left(\frac{100/15^2 + 10(113)/15^2}{1/15^2 + 10/15^2}, \frac{1}{1/15^2 + 10/15^2}\right) = N\left(\frac{1230}{11}, \frac{15^2}{11}\right) = N(111.8182, 20.45455)$$

The Bayesian point estimate is given by the posterior mean  $\hat{\mu} = E(\mu|y) = \frac{1230}{11} = 111.8182$ .

The 95% credible interval is given by  $(\hat{\mu} - q_{0.975}\sigma, \hat{\mu} + q_{0.975}\sigma)$

## [1] 102.9539 120.6825

$$(111.8182 - 1.96\sqrt{20.45455}, 111.8182 + 1.96\sqrt{20.45455}) = (102.9539, 120.6825)$$

So the 95% credible interval is (102.9539, 120.6825).

### Question 1b

$$\begin{aligned} E((\hat{\mu} - \mu^*)^2|\mu^*) &= E((\hat{\mu} - E(\hat{\mu}|\mu^*) + E(\hat{\mu}|\mu^*) - \mu^*)^2|\mu^*) \\ &= E((\hat{\mu} - E(\hat{\mu}|\mu^*))^2|\mu^*) + 2E((\hat{\mu} - E(\hat{\mu}|\mu^*))(E(\hat{\mu}|\mu^*) - \mu^*)|\mu^*) + E((E(\hat{\mu}|\mu^*) - \mu^*)^2|\mu^*) \\ &= Var(\hat{\mu}|\mu^*) + 2E[\hat{\mu}E(\hat{\mu}|\mu^*) - \hat{\mu}\mu^* - E(\hat{\mu}|\mu^*)^2 + \mu^*E(\hat{\mu}|\mu^*)|\mu^*] + E[E(\hat{\mu}|\mu^*)^2 - 2\mu^*E(\hat{\mu}|\mu^*) + \mu^*|\mu^*] \\ &= Var(\hat{\mu}|\mu^*) + E[2\hat{\mu}E(\hat{\mu}|\mu^*) - 2\hat{\mu}\mu^* - 2E(\hat{\mu}|\mu^*)^2 + 2\mu^*E(\hat{\mu}|\mu^*) + E(\hat{\mu}|\mu^*)^2 - 2\mu^*E(\hat{\mu}|\mu^*) + \mu^*|\mu^*] \\ &= Var(\hat{\mu}|\mu^*) + 2E(\hat{\mu}|\mu^*)E(\hat{\mu}|\mu^*) - 2E(\hat{\mu}|\mu^*)\mu^* - E(\hat{\mu}|\mu^*)^2 + \mu^* \\ &= Var(\hat{\mu}|\mu^*) + E(\hat{\mu}|\mu^*)^2 - 2E(\hat{\mu}|\mu^*)\mu^* + \mu^* \\ &= Var(\hat{\mu}|\mu^*) + (Bias(\hat{\mu}|\mu^*))^2 \end{aligned}$$

### Question 1c

We derive the distribution of  $\hat{\mu}_{BAYES} = \frac{\mu_0/\sigma_{\mu_0}^2 + n\bar{y}/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}$

Since  $y_i|\mu, \sigma^2 \sim N(112, 15^2)$  we have that

$$E[\hat{\mu}_{BAYES}] = E\left[\frac{\mu_0/\sigma_{\mu_0}^2 + n\bar{y}/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}\right] = \frac{\mu_0/\sigma_{\mu_0}^2 + n\mu/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}$$

$$Var[\hat{\mu}_{BAYES}] = Var\left[\frac{\mu_0/\sigma_{\mu_0}^2 + n\bar{y}/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}\right] = \frac{(n/\sigma^2)^2(\sigma^2/n)}{(1/\sigma_{\mu_0}^2 + n/\sigma^2)^2} = \frac{n/\sigma^2}{(1/\sigma_{\mu_0}^2 + n/\sigma^2)^2}$$

So  $\hat{\mu}_{Bayes} \sim N(\frac{\mu_0/\sigma_{\mu_0}^2 + n\mu/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}, \frac{n/\sigma^2}{(1/\sigma_{\mu_0}^2 + n/\sigma^2)^2})$

We substitute

$$\mu_0 = 100, \sigma_{\mu_0} = 15, n = 10, \sigma = 15, \mu = 112$$

## [1] 110.9091

## [1] 18.59504

Then we get

$$\hat{\mu}_{Bayes} \sim N\left(\frac{100/(15^2) + 10(112)/15^2}{1/15^2 + 10/15^2}, \frac{10/15^2}{(1/15^2 + 10/15^2)^2}\right) \sim N(110.9091, 18.59504)$$

So

$$MSE(\hat{\mu}_{Bayes}|\mu^* = 112) = Var(\hat{\mu}|\mu^*) + (Bias(\hat{\mu}|\mu^*))^2 = 18.59504 + (110.9091 - 112)^2 = 19.7851$$

The bias for  $\hat{\mu}_{BAYES}$  is  $110.9091 - 112 = -1.0909$ . The variance of  $\hat{\mu}_{BAYES}$  is 18.59504.

We have  $\mu^* = 112$ . The ML estimator is the sample mean, which is  $\bar{y}$ .  $y \sim N(112, 15^2) \implies \bar{y} \sim N(112, \frac{15^2}{10})$ .

The bias of the ML estimator is  $E(\bar{y}|\mu^*) - 112 = E(y) - 112 = 112 - 112 = 0$

The variance of the ML estimator is  $\frac{15^2}{10} = 22.5$ .

The MSE of the ML estimator is  $MSE = Var(\bar{y}) + (Bias(\bar{y}))^2 = \frac{Var(y)}{10} + 0 = \frac{15^2}{10} = 22.5$

$\hat{\mu}_{BAYES}$  has larger magnitude of bias, lower variance, and lower MSE.

The ML estimator has lower bias, higher variance, and higher MSE.

### Question 1d

We have  $\mu^* = 112$ . The ML estimator is the sample mean, which is  $\bar{y}$ .

$y \sim N(112, 15^2) \implies \bar{y} \sim N(112, \frac{15^2}{10})$ .

So the sampling distribution of the ML estimate  $\bar{y}$  is  $N(112, \frac{15^2}{10})$ .

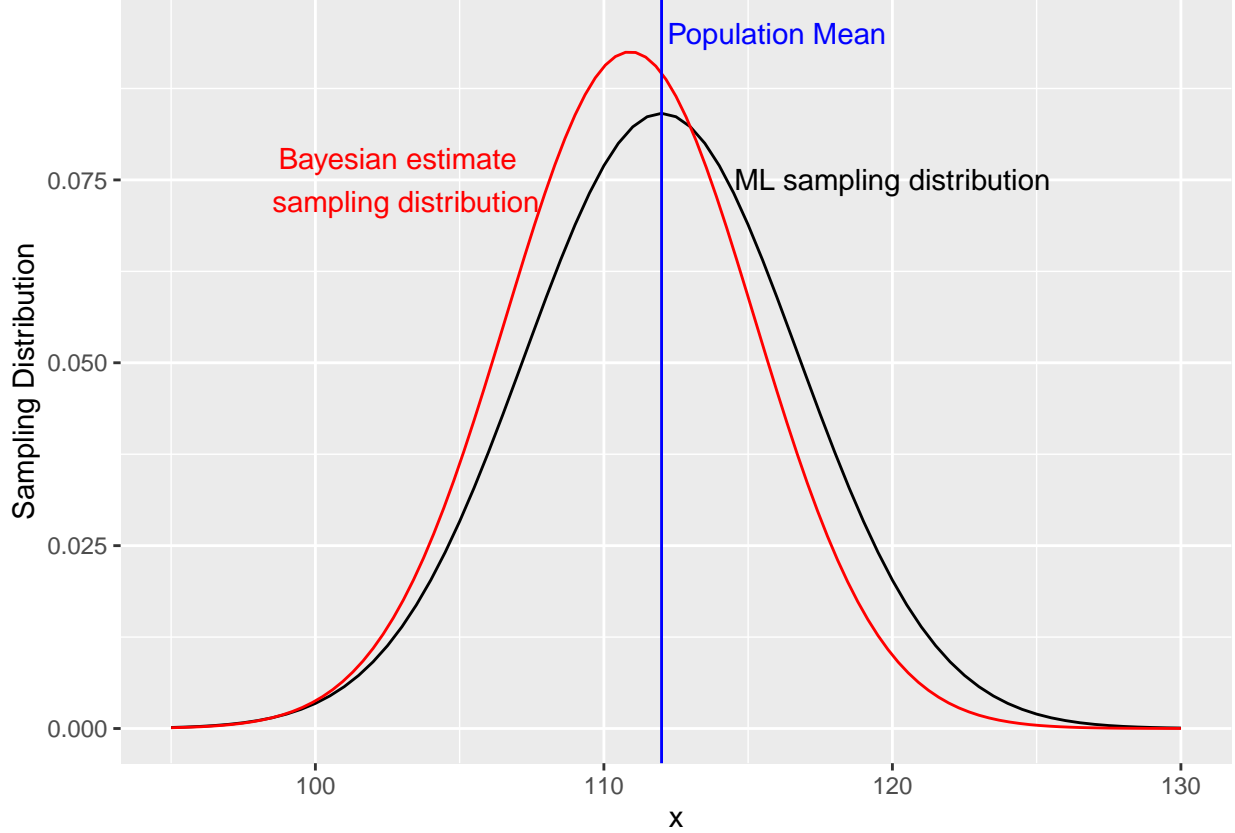
Now we derive the distribution of  $\hat{\mu}_{BAYES} = \frac{\mu_0/\sigma_{\mu_0}^2 + n\bar{y}/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}$

Since  $y_i|\mu, \sigma^2 \sim N(\mu, \sigma^2)$  we have that

$$E[\hat{\mu}_{BAYES}] = E\left[\frac{\mu_0/\sigma_{\mu_0}^2 + n\bar{y}/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}\right] = \frac{\mu_0/\sigma_{\mu_0}^2 + n\mu/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}$$

$$Var[\hat{\mu}_{BAYES}] = Var\left[\frac{\mu_0/\sigma_{\mu_0}^2 + n\bar{y}/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}\right] = \frac{(n/\sigma^2)^2(\sigma^2/n)}{(1/\sigma_{\mu_0}^2 + n/\sigma^2)^2} = \frac{n/\sigma^2}{(1/\sigma_{\mu_0}^2 + n/\sigma^2)^2}$$

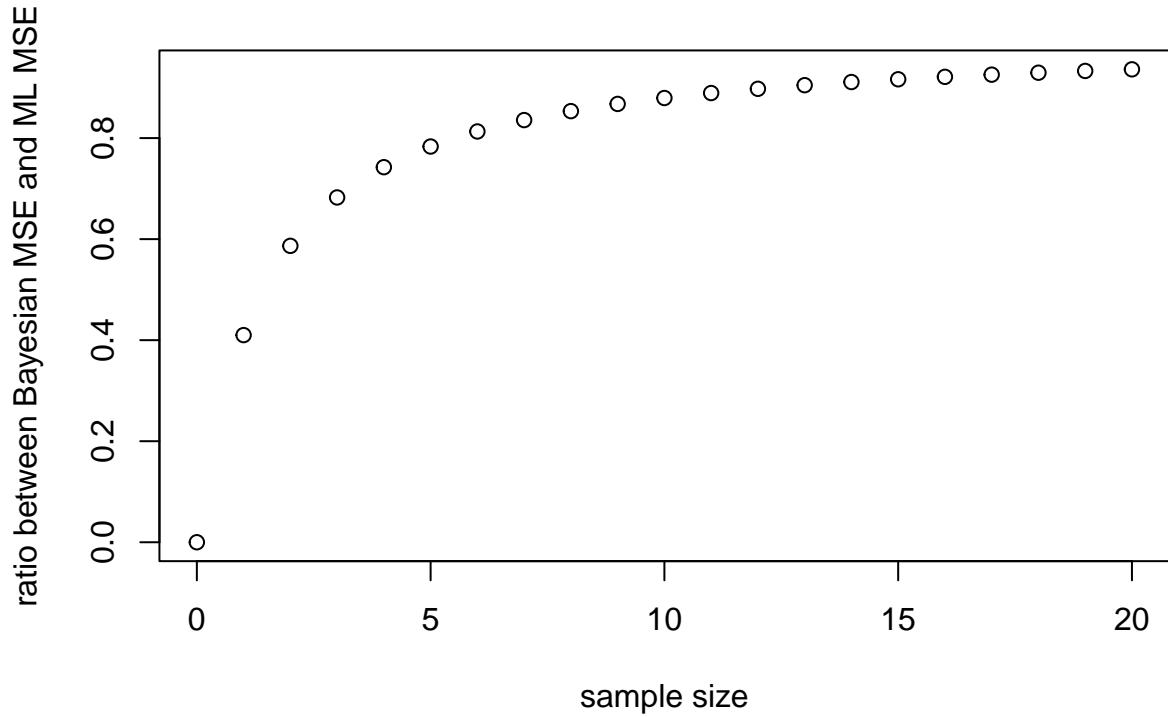
$$\text{So } \hat{\mu}_{Bayes} \sim N\left(\frac{\mu_0/\sigma_{\mu_0}^2 + n\mu/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2}, \frac{n/\sigma^2}{(1/\sigma_{\mu_0}^2 + n/\sigma^2)^2}\right)$$



The curve in red represents the Bayesian estimate. We can see that the mean of the Bayesian estimate sampling distribution is farther from the true mean than the mean of the ML estimate sampling distribution. The variance of the ML estimate sampling distribution seems to be slightly larger than the variance of the Bayesian estimate sampling distribution. The bias of the ML estimate is smaller than the bias of the Bayesian estimate because the peak of the Bayesian estimate sampling distribution is at the true mean. The variance of the Bayesian estimate is smaller but the bias of the Bayesian estimate is larger.

$$MSE(\bar{y}|\mu) = Var(\bar{y}|\mu) + [E(\bar{y}|\mu) - \mu]^2 = \frac{\sigma^2}{n}$$

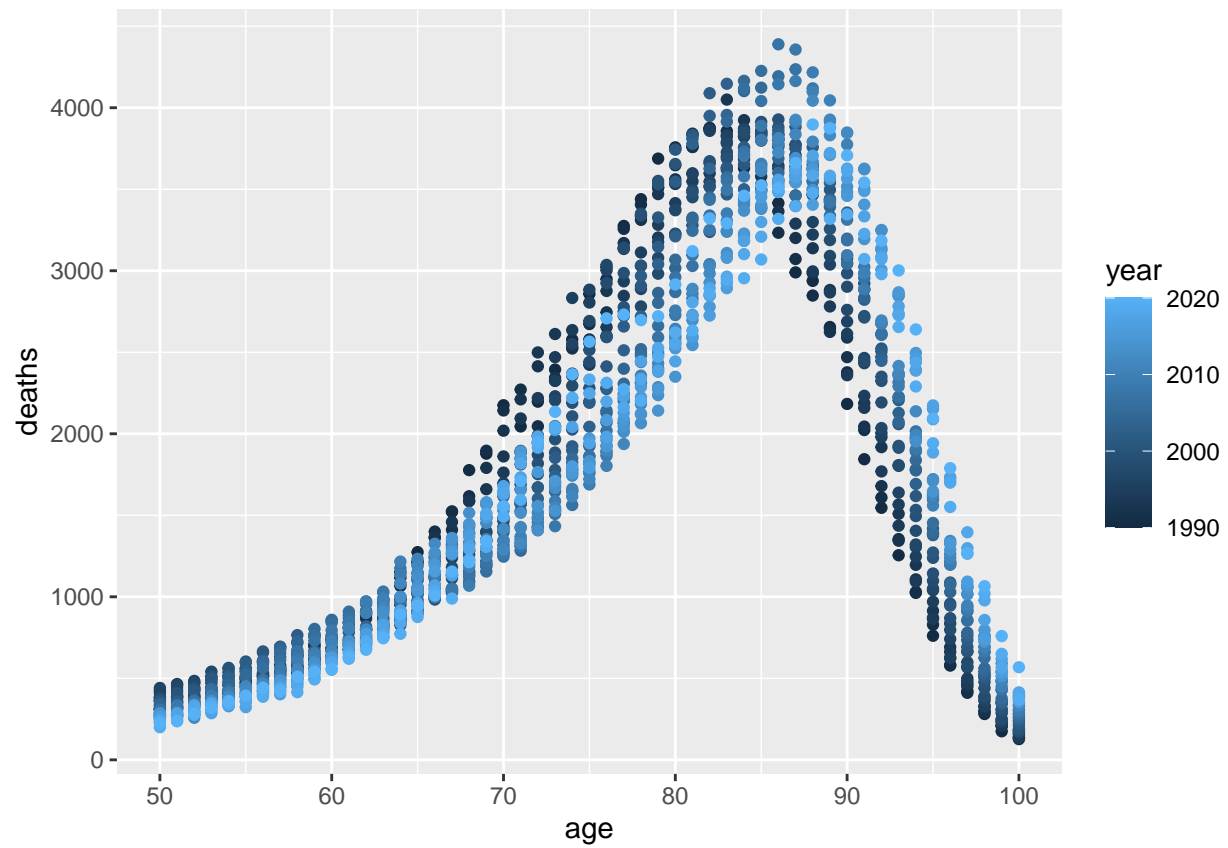
$$MSE(\hat{\mu}_{Bayes}) = Var(\hat{\mu}_{Bayes}|\mu) + [E(\hat{\mu}_{Bayes}|\mu) - \mu]^2 = \frac{n/\sigma^2}{(1/\sigma_{\mu_0}^2 + n/\sigma^2)^2} + \left[\frac{\mu_0/\sigma_{\mu_0}^2 + n\mu/\sigma^2}{1/\sigma_{\mu_0}^2 + n/\sigma^2} - \mu\right]^2$$



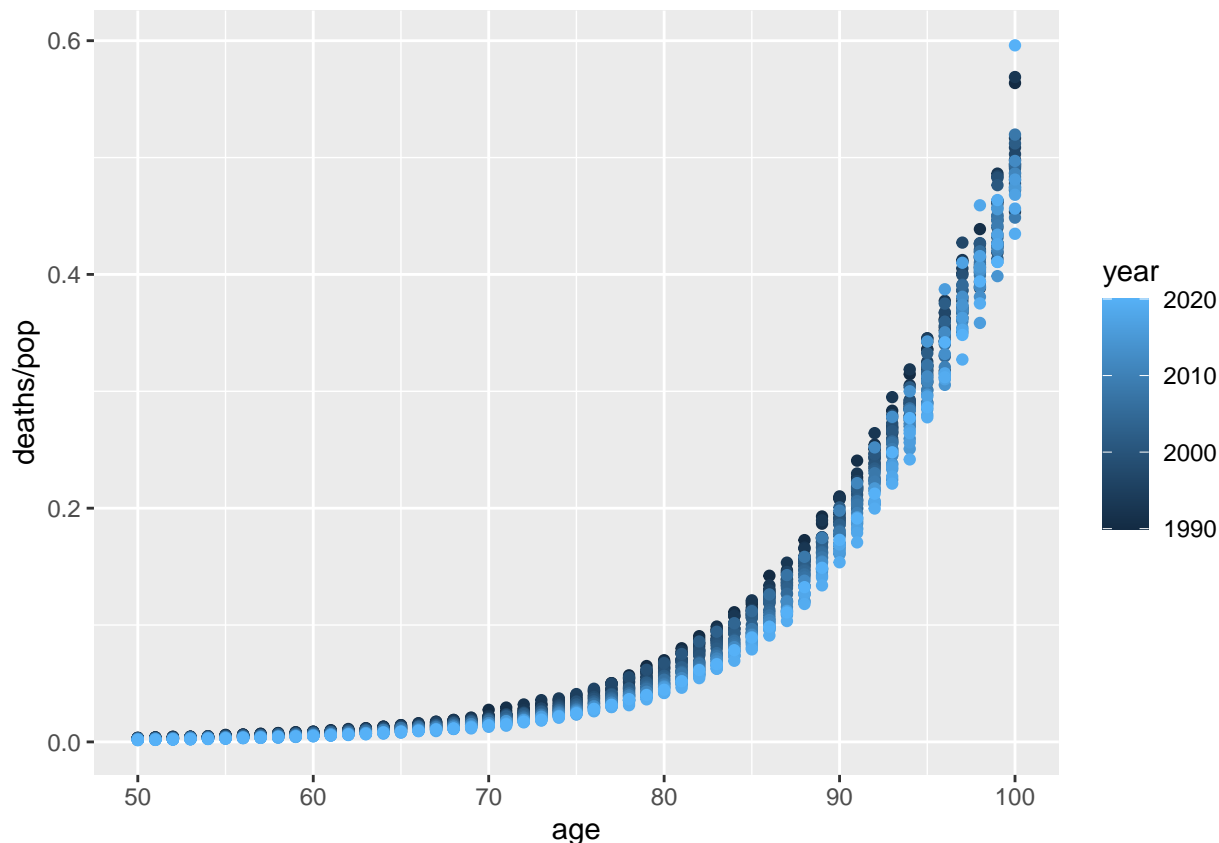
In general the ratio between the Bayesian estimate's MSE and the ML estimate's MSE is less than or equal to 1. This is expected because the peak of the Bayesian estimate's sampling distribution was to the left of the peak of the ML estimate's sampling distribution. We can see that as sample size increases, the ratio of the MSE of the Bayes estimate to the MSE of the ML estimate increases towards 1. As our sample size increases, our Bayes estimate MSE gets closer to our ML estimate MSE, because we reduce the variance in the ML estimates. As our sample sample size increases, our Bayes estimate also has lower bias.

## Question 2

### Question 2a



Over time, the number of deaths for people ages 50-80 has decreased but the number of deaths for people ages 80-100 has risen. It seems that the age with the most deaths has increased over time. This suggests that people are living longer over time.



We see that as age increases, the mortality rate (number of deaths/population) increases exponentially.

## Question 2b

Note that  $\mu_0 = \alpha \exp(\beta(0)) = \alpha$  so  $\alpha$  should be the baseline mortality rate. We will later demean the age covariate before fitting our model so  $\alpha$  will be the mortality rate for the age 75 group. In 2019, the median mortality rate across all age groups was around 0.0237. I expect the mortality rate for 2019 to be close to the mortality rate for 2020. So I set my prior for  $\alpha$  to be normally distributed with mean 0.02.

The rate at which mortality increases for every increase in age should be  $\frac{\alpha \exp(\beta(x+1))}{\alpha \exp(\beta x)} = \exp(\beta)$ . We expect this to be positive, because as people age, they should have a higher chance of dying.

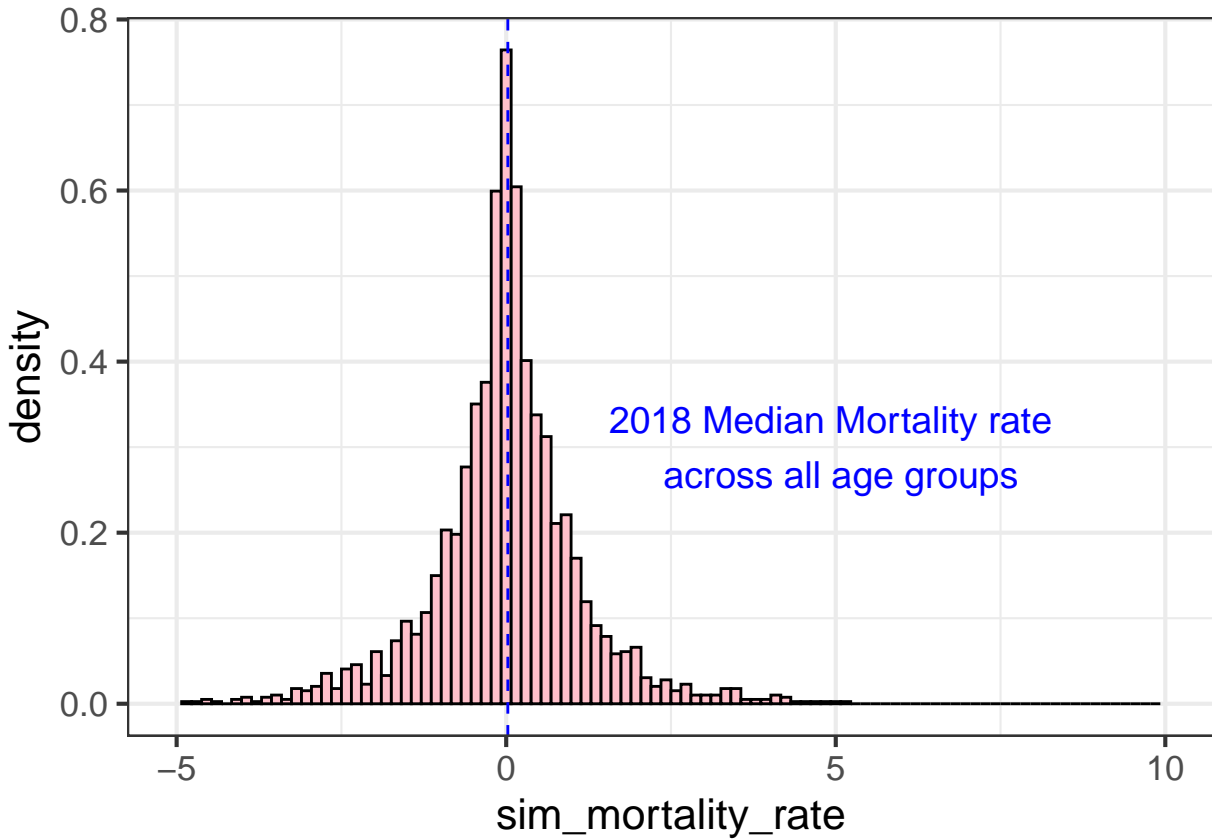
```
## [1] 0.02373468
```

```
## [1] 0.002453453
```

In 2019, the median difference for the mortality rate between consecutive age groups was around 0.0025. We expect differences between consecutive age groups' death rates to get bigger as age increases. We think that setting a prior mean for  $\beta$  as 0.0025 should be reasonable.

Mortality rate should be between 0 and 1 so we use a standard deviation of 0.01 so that alpha estimates within 2 standard deviations of the prior mean will fall in (0,1). I think the difference in mortality rate between consecutive age groups should be positive in general, so I set the prior variance for  $\beta$  to be 0.001.

I consider  $\alpha \sim N(0.02, 0.01)$  and  $\beta \sim N(0.0025, 0.001)$ . Then I will simulate mortality rates given these priors and compare them to the 2018 median mortality rate among all age groups.

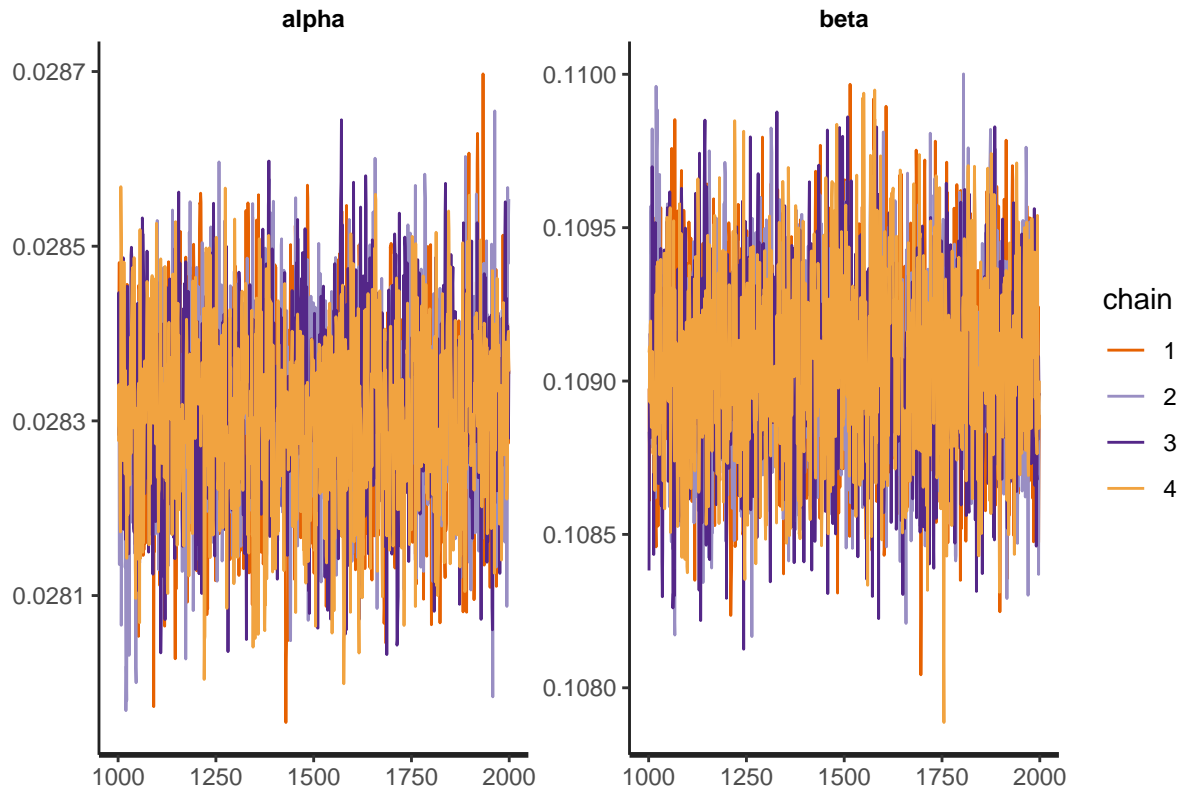


Here we find that most of the simulated mortality rates given  $\alpha \sim N(0.02, 0.01)$  and  $\beta \sim N(0.0025, 0.001)$  priors are close to the 2018 median mortality rate across all age groups.

### Question 2c

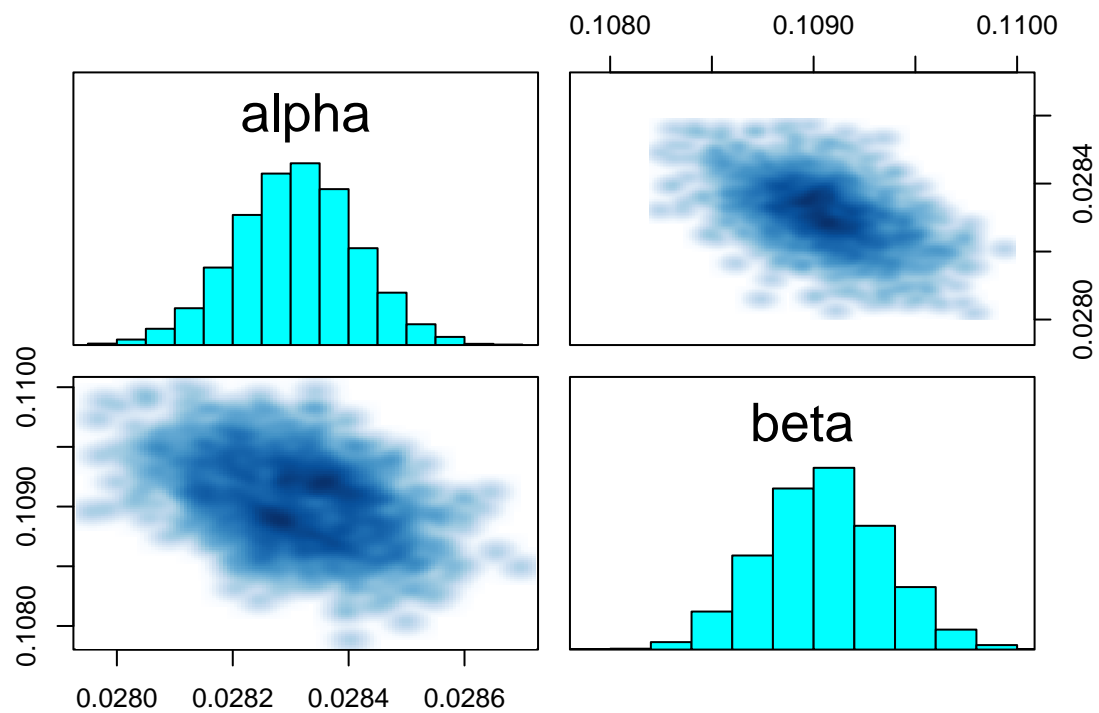
Before fitting the model, we demeaned the age to improve the fit of the model. Hopefully this will speed up convergence and reduce correlation between alpha and beta. Demeaning the population doesn't seem to work because some of the demeaned population values will be negative. We pass population into the model  $\log(\mu_x) = \log(\alpha) + \beta * age_x + \log(pop_x)$  and log can't take negative values.

We see that the model converges as the traceplot shows that the chains mixed well and there are no obvious trends.

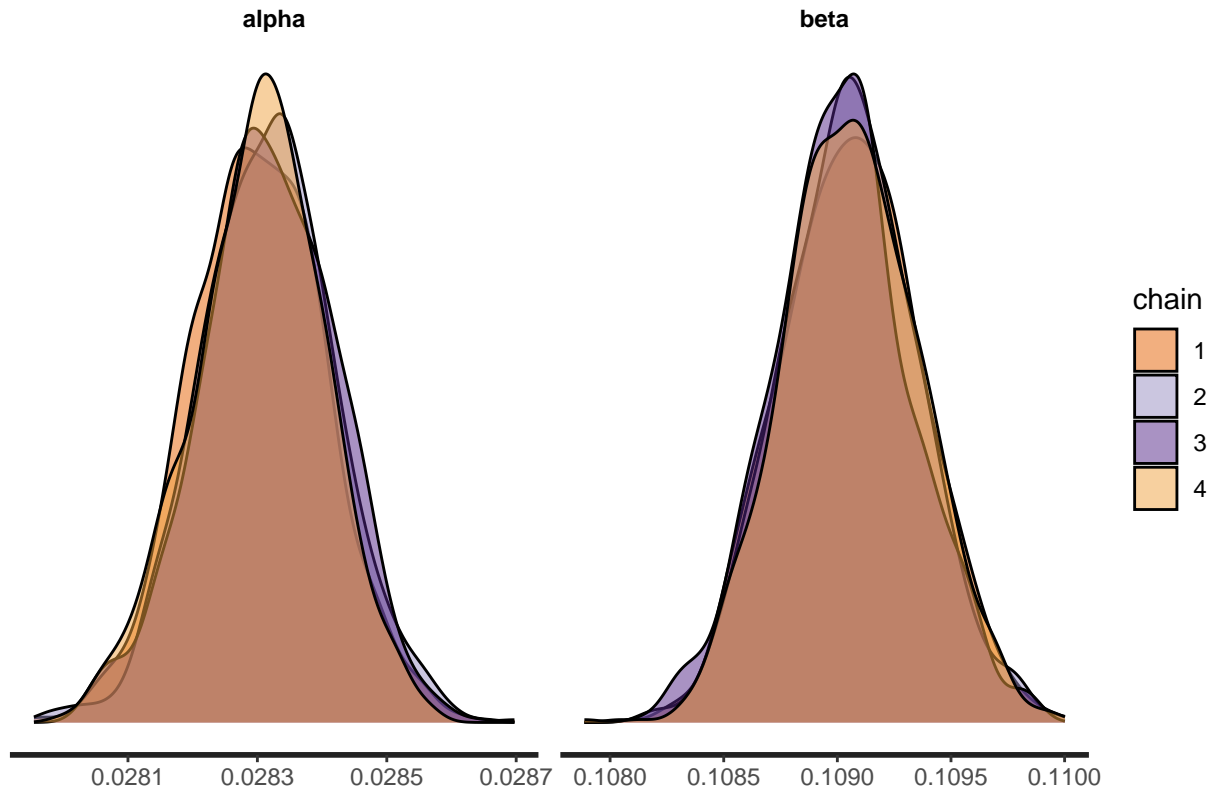


Demeaning the age covariate helped reduce correlation between alpha and beta. If we had not demeaned the age covariate, there would have been very strong correlation between alpha and beta.





We see that the distributions for  $\alpha$  and  $\beta$  generated by each chain are similar.



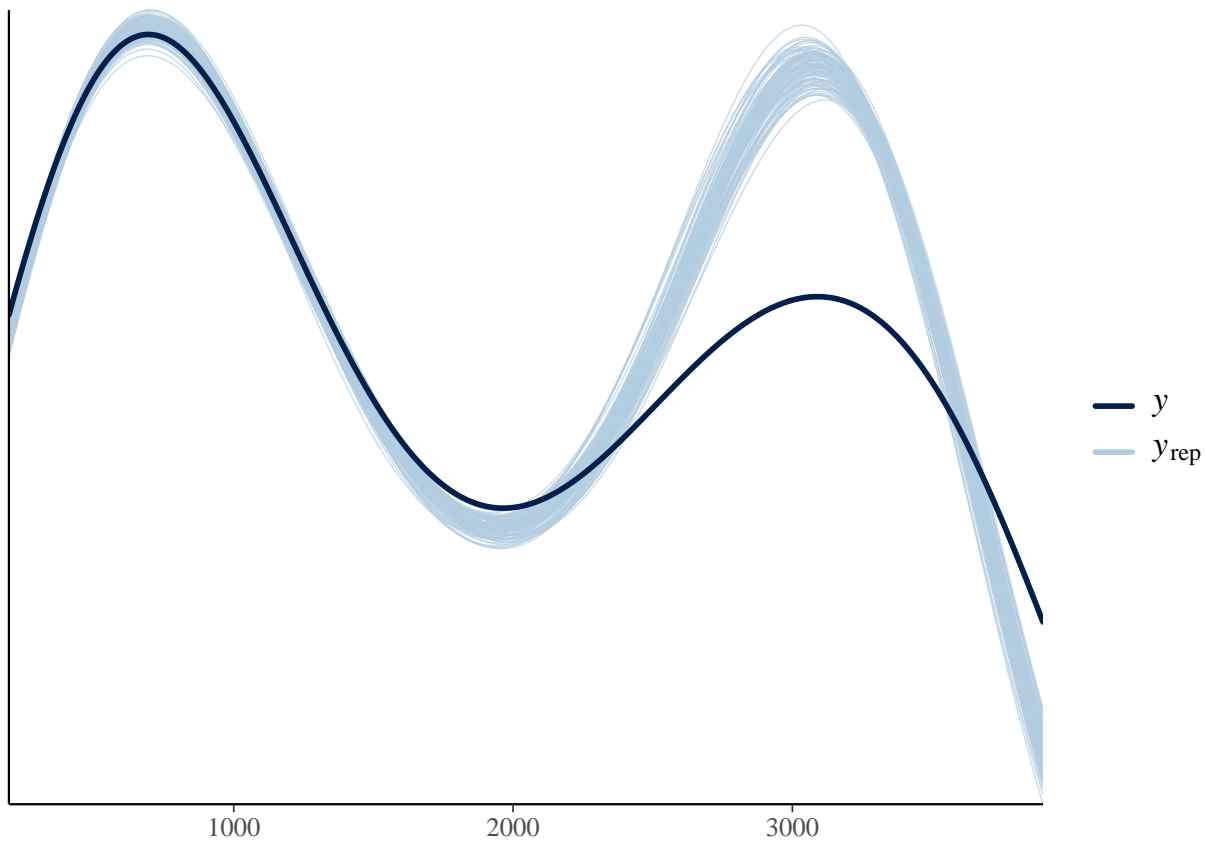
Here is the summary of Model 1.

```
##           mean      se_mean        sd      2.5%      25%      50%
## alpha  0.0283097  2.494200e-06  0.0001031620  0.02810242  0.0282415  0.02831065
## beta   0.1090507  6.821062e-06  0.0002997896  0.10848228  0.1088461  0.10905013
##           75%      97.5%    n_eff    Rhat
## alpha  0.02837947  0.02851045  1710.712  1.002528
## beta   0.10925022  0.10964875  1931.653  1.001479
```

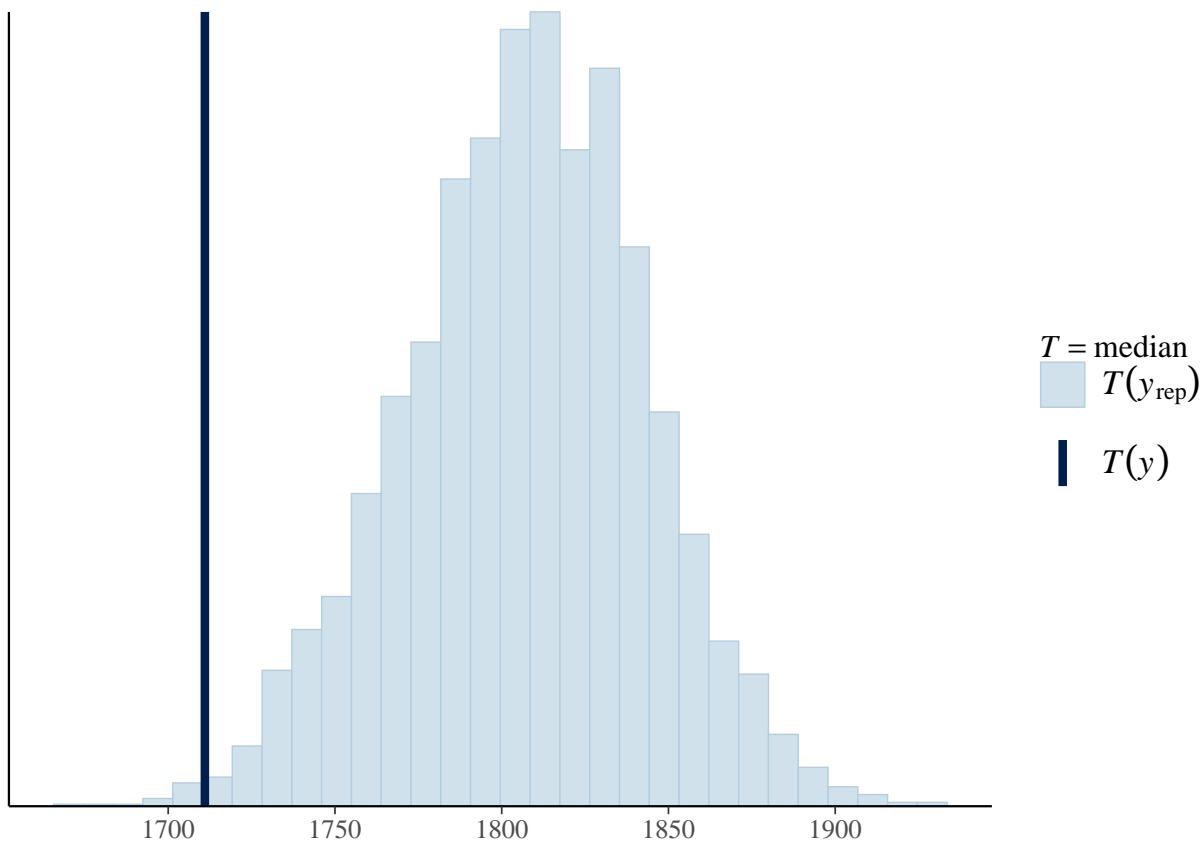
The mortality rate for age group 75 is 0.0283097. The ratio of mortality rate between age groups 1 year apart is  $\exp(0.1090507)$

## Question 2d

We compare the densities of 100 sampled datasets to the actual data. The densities of the sampled datasets based on the model are somewhat close to the true data.



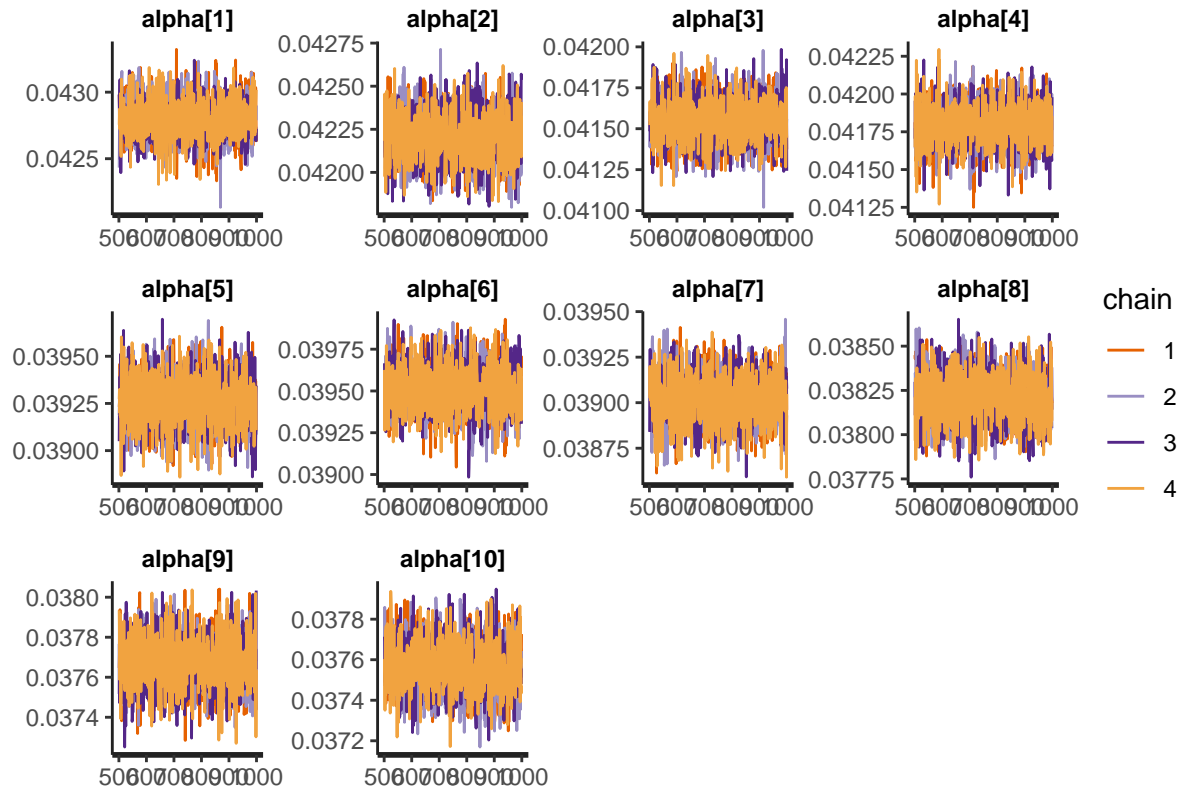
The median deaths and replicated median deaths are somewhat close to each other.

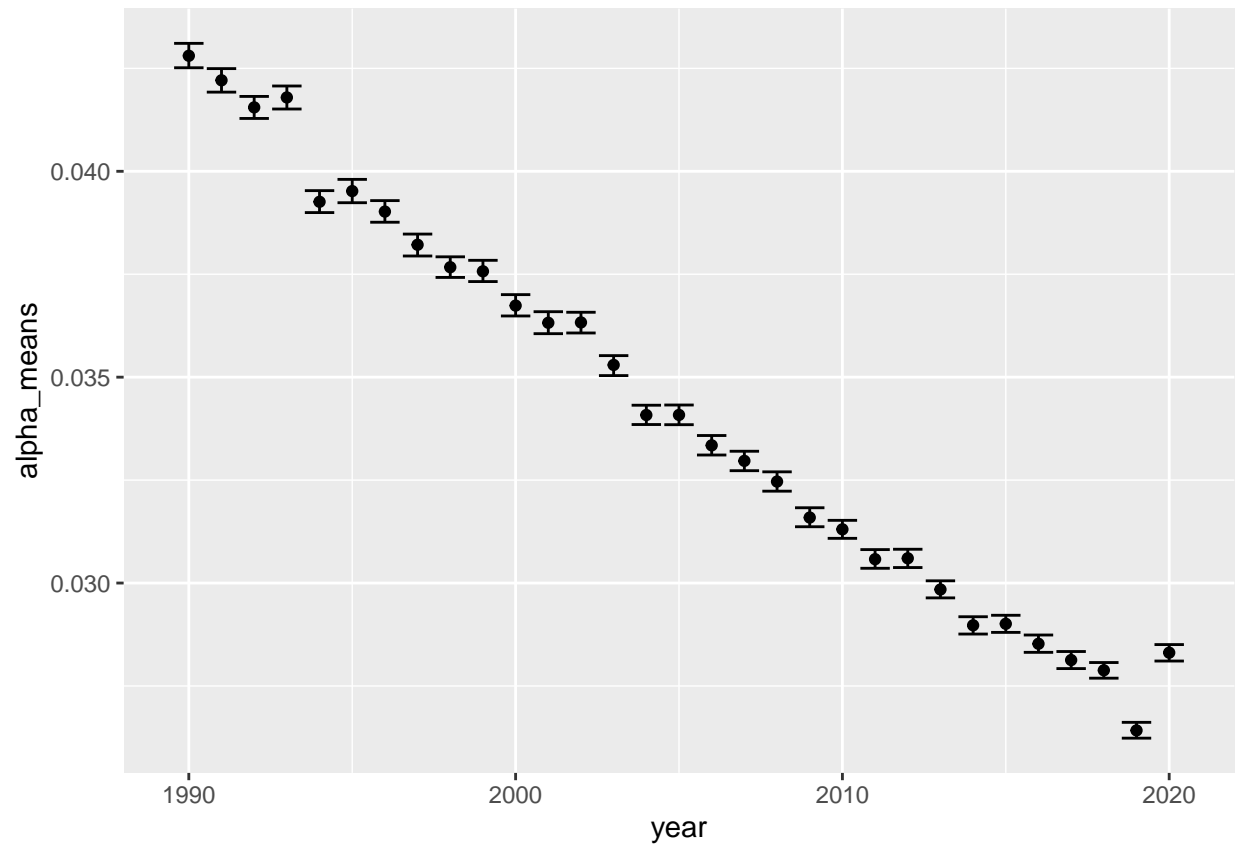


### Question 2e

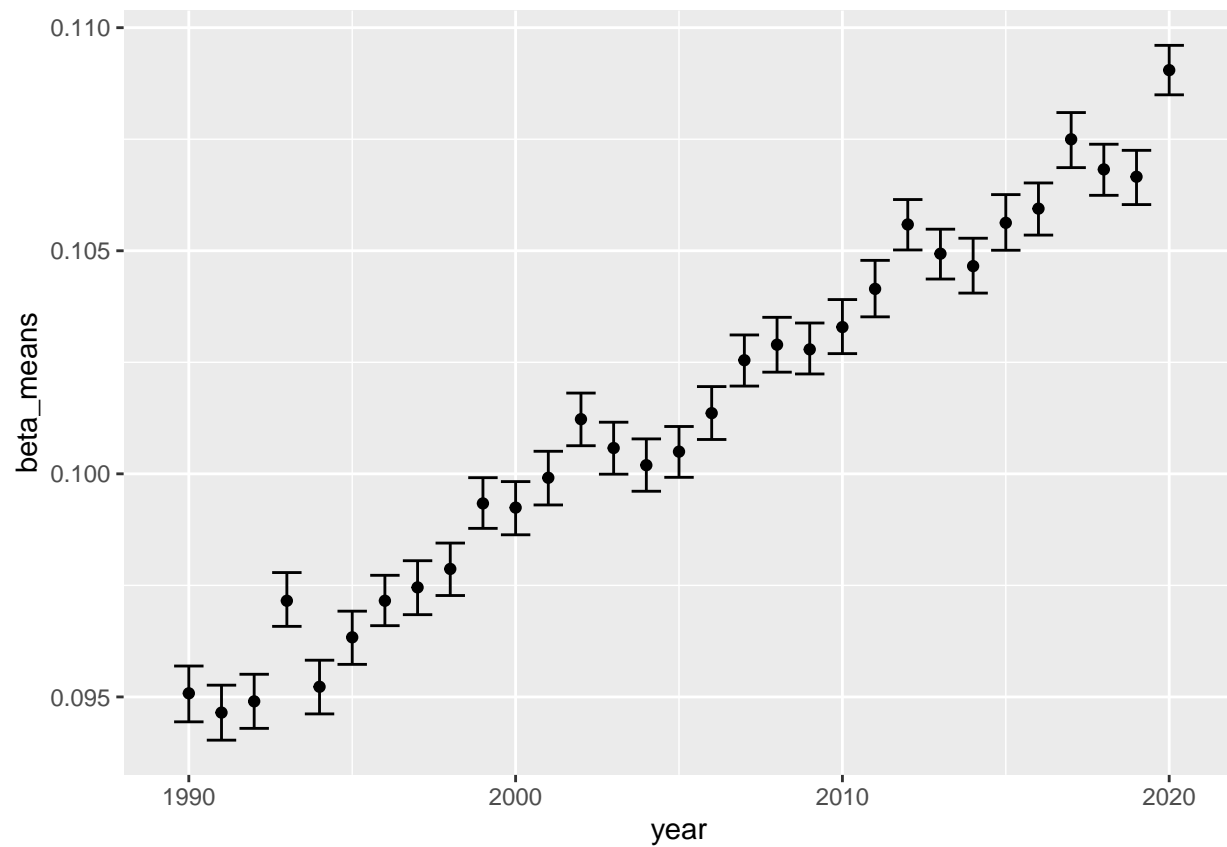
Again, we demean the age covariate for the entire Sweden mortality dataset, in order to speed up convergence and reduce correlation between  $\alpha, \beta$ . We create a matrix of death counts where the  $ij$ -th element corresponds to the  $i$ -th age and  $j$ -th year. We extend the previous model to this matrix. We use the same alpha and beta priors.

The model seems to have converged.



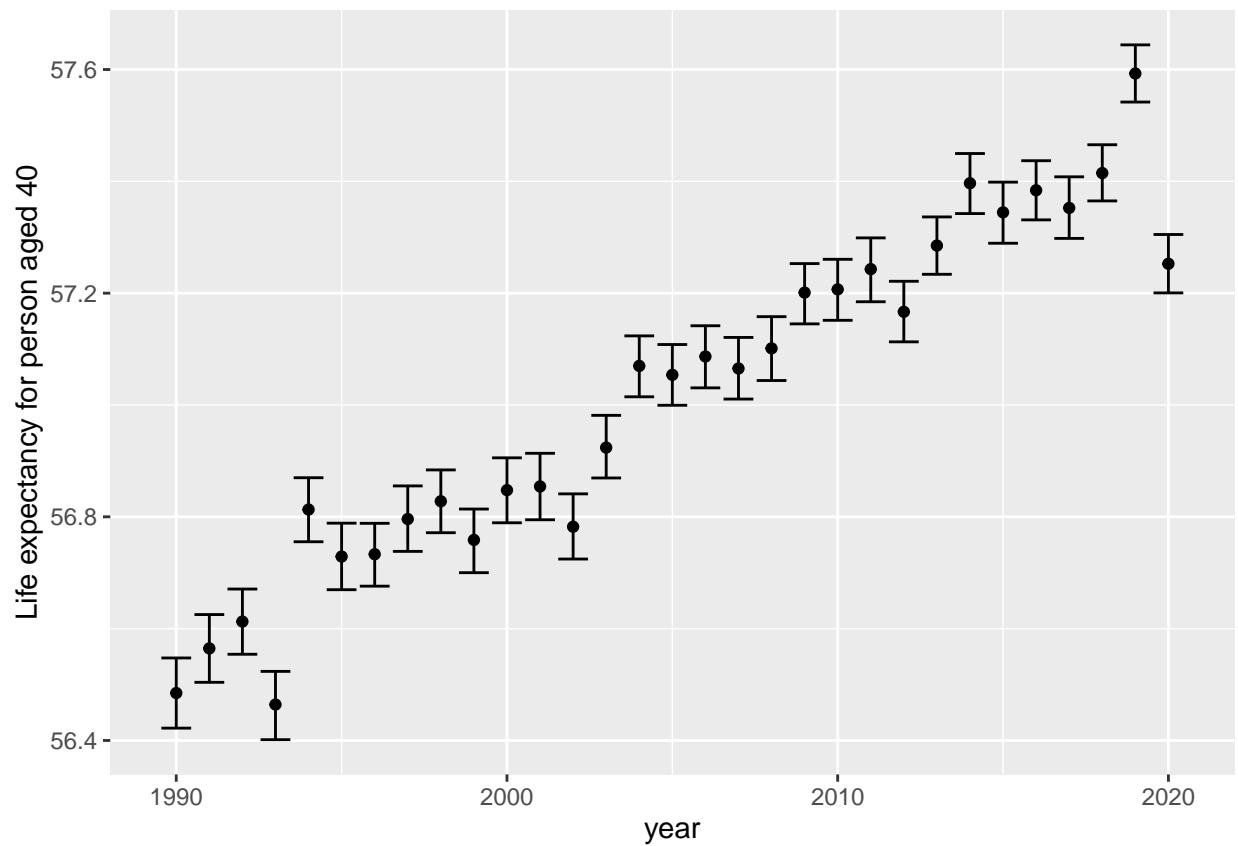


It seems that over time, the model estimated lower  $\alpha$  coefficients. This suggests that the mortality rate for the age 75 group decreased over time.



Over time, the model estimated higher  $\beta$  coefficients. This suggests that the ratio between mortality rates of age groups 1 year apart increased.

## Question 2f

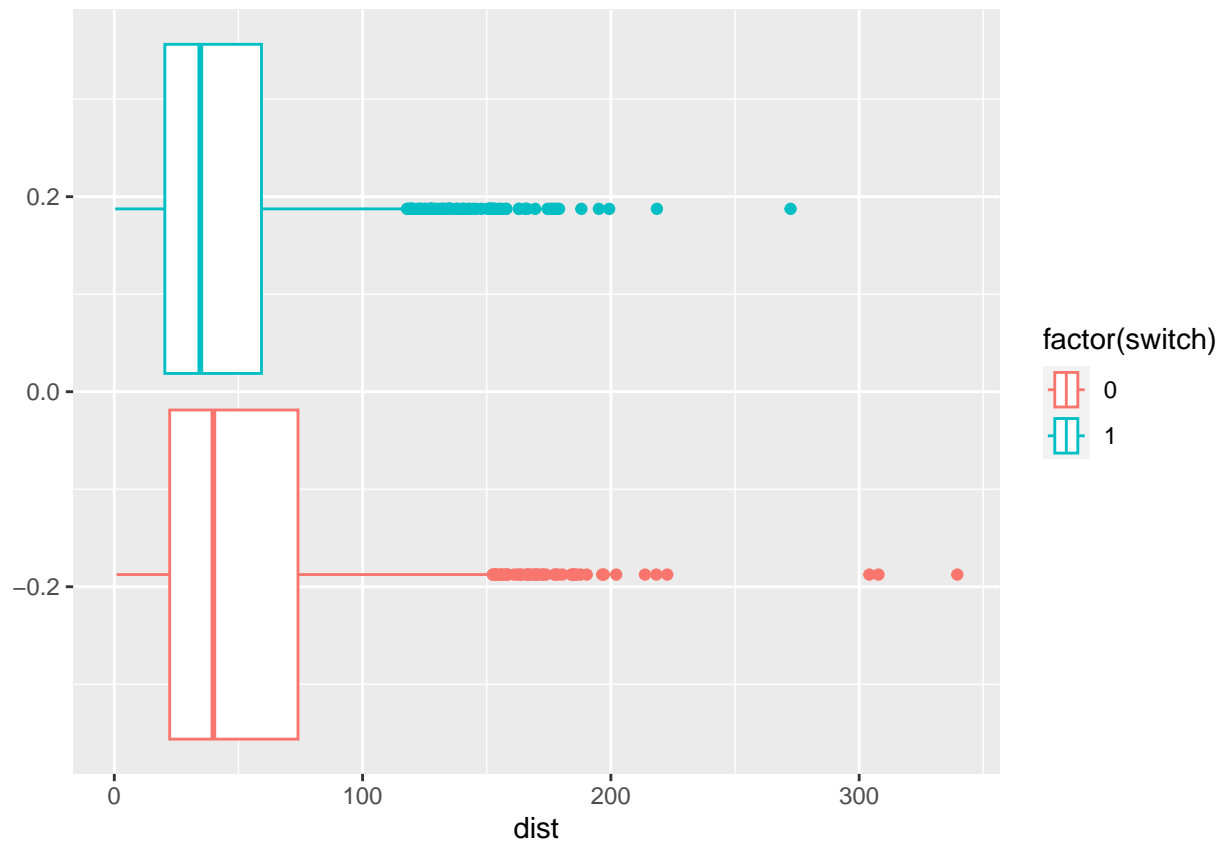


We can see that over time, the predicted life expectancy for a person aged 40 increases. The life expectancy estimates are rather high however, because the model predicts someone aged 40 living until their late 90s (between 56-58 years of life left at age 40).

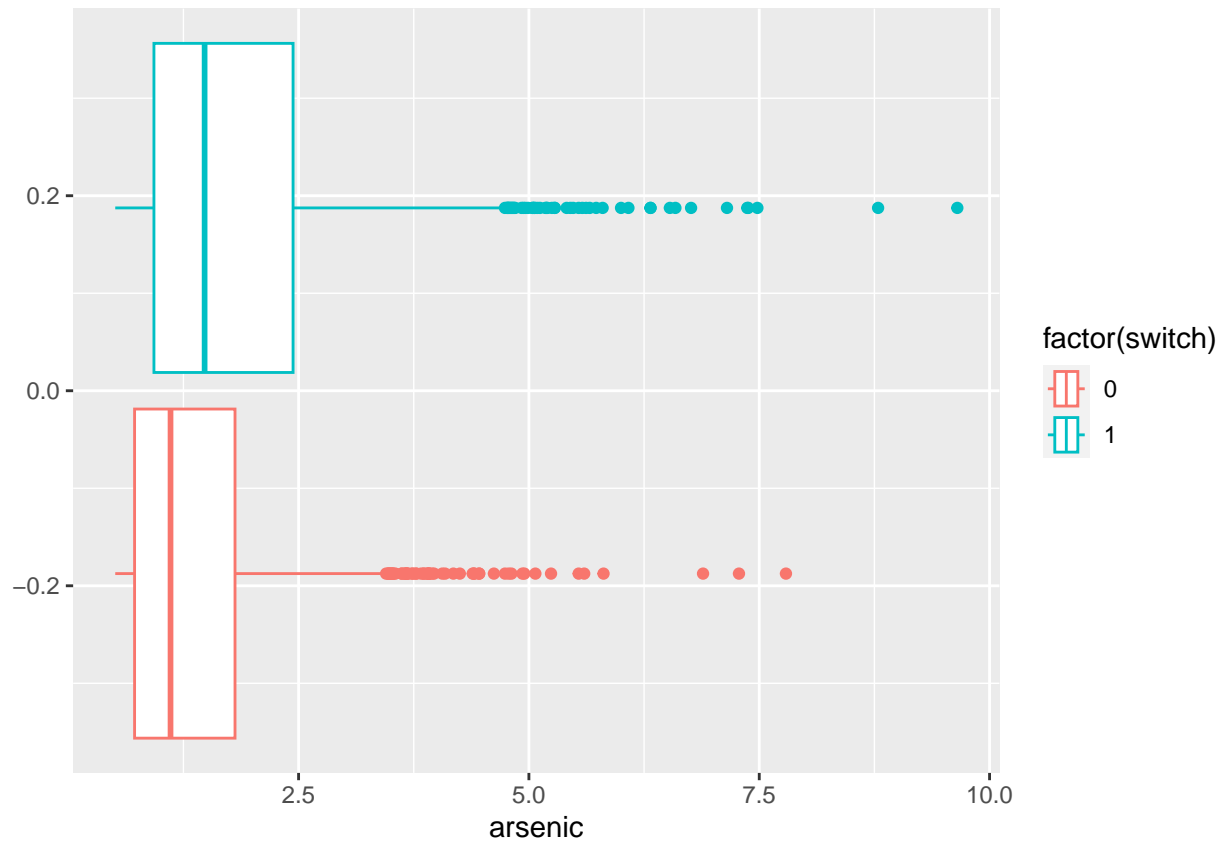


### Question 3

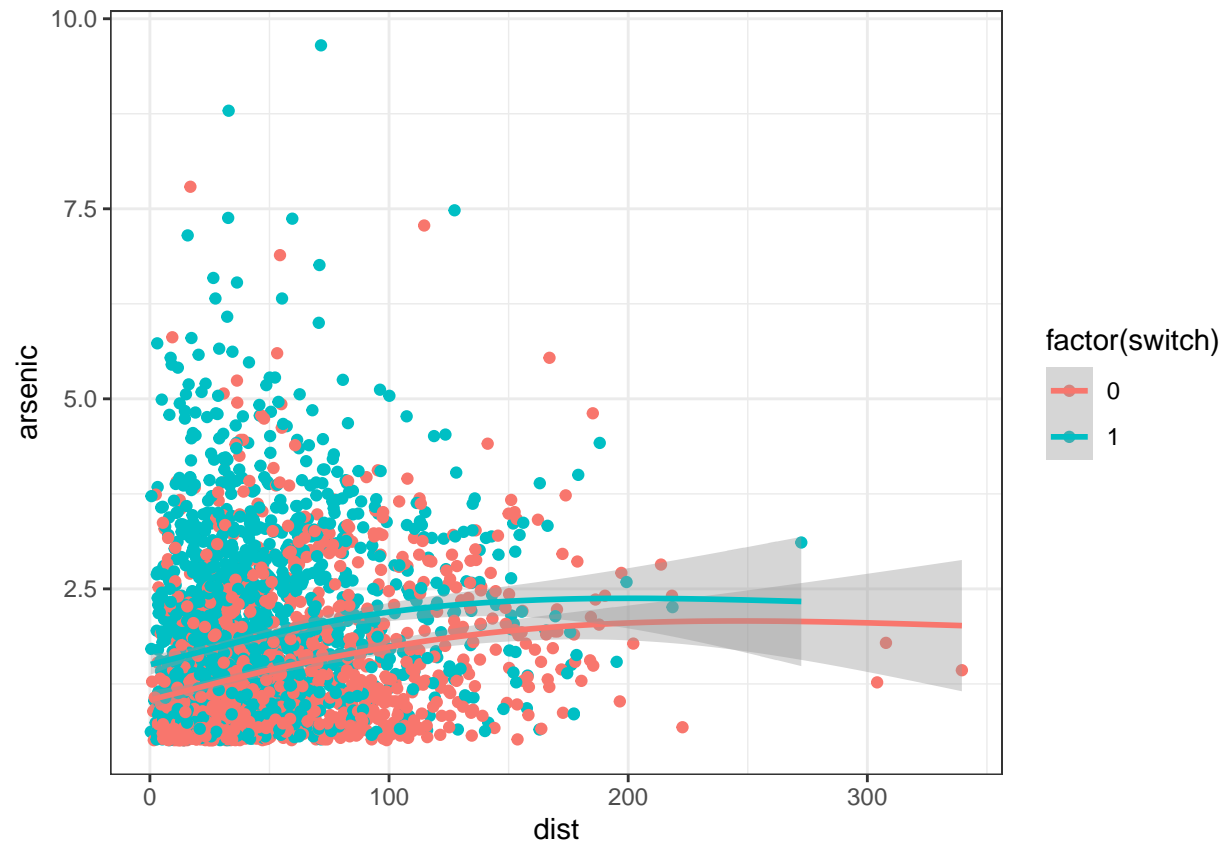
#### Question 3a



It seems that the median distance of people who switched their well was slightly lower than median distance of people who did not switch their well.



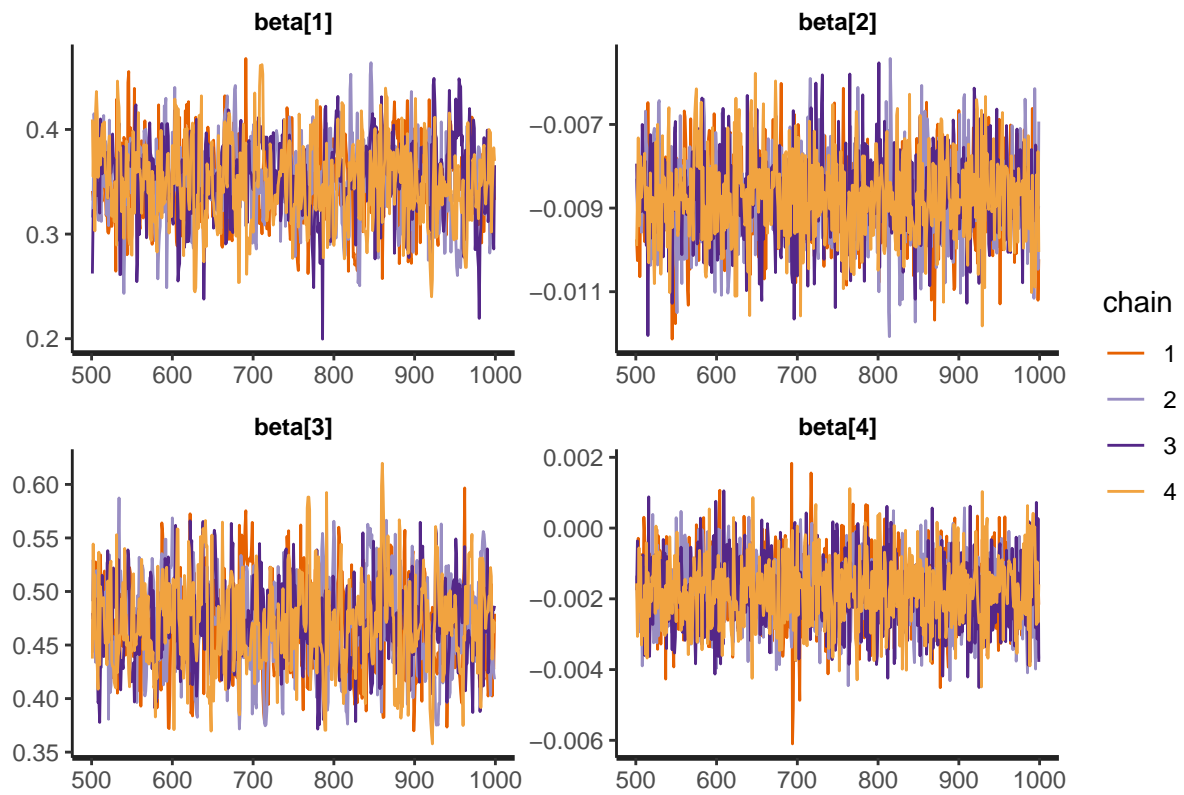
The median arsenic level was higher for people who switched their wells than for people who did not switch their wells.



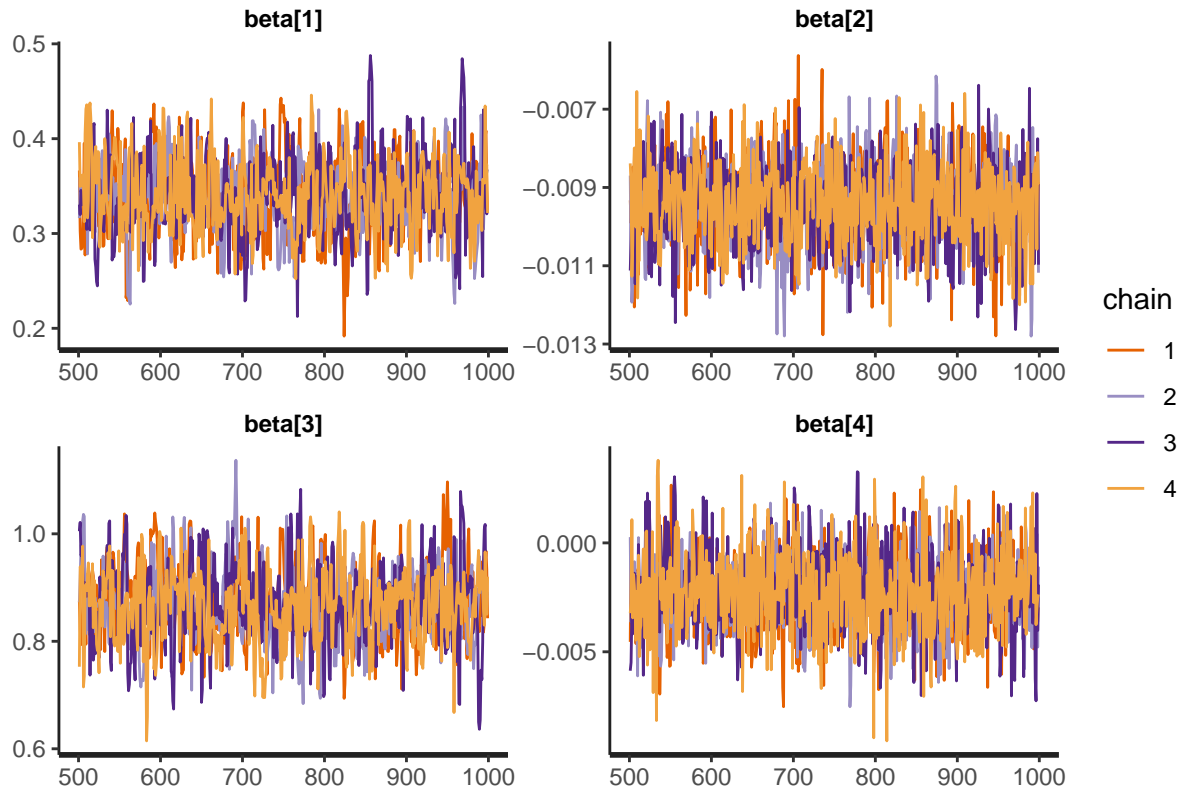
Distance and arsenic did not appear to be heavily correlated. However it seems that when arsenic levels were higher, more people switched their wells. When distance was greater, more people did not switch their well.

### Question 3b

```
## [1] "Traceplots for Model 1 Coefficients"
```



```
## [1] "Traceplots for Model 2 Coefficients"
```



##		mean	se_mean	sd	2.5%	25%
##	beta[1]	0.350705008	1.405035e-03	0.038658177	0.274939237	0.323957894
##	beta[2]	-0.008739881	2.301971e-05	0.001044260	-0.010791446	-0.009460373
##	beta[3]	0.469091474	1.547105e-03	0.042051947	0.389150730	0.440073733
##	beta[4]	-0.001775180	2.312529e-05	0.001016146	-0.003708608	-0.002477051
##		50%	75%	97.5%	n_eff	Rhat
##	beta[1]	0.351805445	0.378021798	0.4235259036	757.0223	1.001478
##	beta[2]	-0.008770970	-0.008046386	-0.0066583541	2057.8687	1.000670
##	beta[3]	0.468831630	0.497625820	0.5533361148	738.8096	1.000787
##	beta[4]	-0.001766262	-0.001090100	0.0001656812	1930.8037	1.000656

The odds of switching a well for someone with mean arsenic levels and mean distance is  $\exp(0.350705008) \approx 1.42$ .

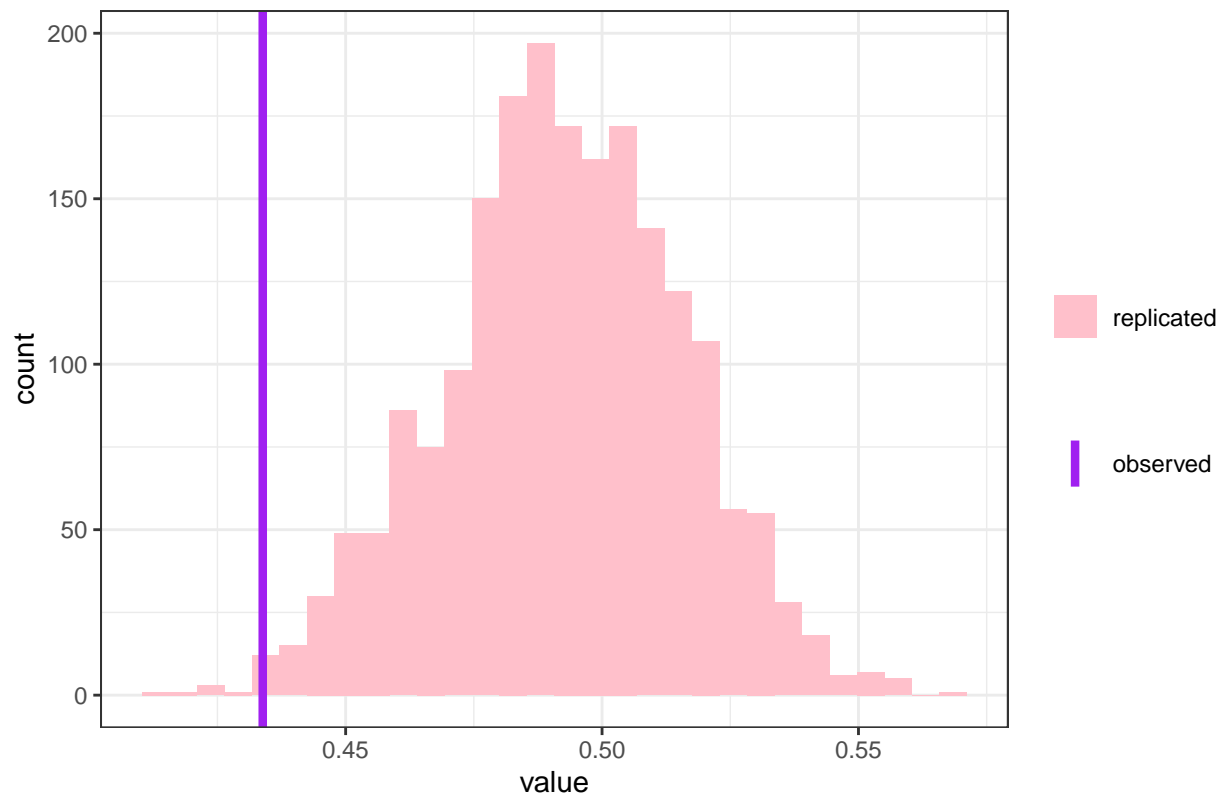
For every unit increase in distance, the odds of switching a well are multiplied by  $\exp(-0.008739881) \approx 0.99$ , assuming all other factors are held constant. Since this is close to 1, someone with a unit increase in distance is probably unlikely to switch their well, assuming other factors held constant.

For every unit increase in arsenic levels, the odds of switching a well are multiplied by  $\exp(0.469091474) \approx 1.60$ , assuming all other factors are held constant. If there is a unit increase in arsenic, someone is 59.8% more likely to switch their well, assuming other factors held constant.

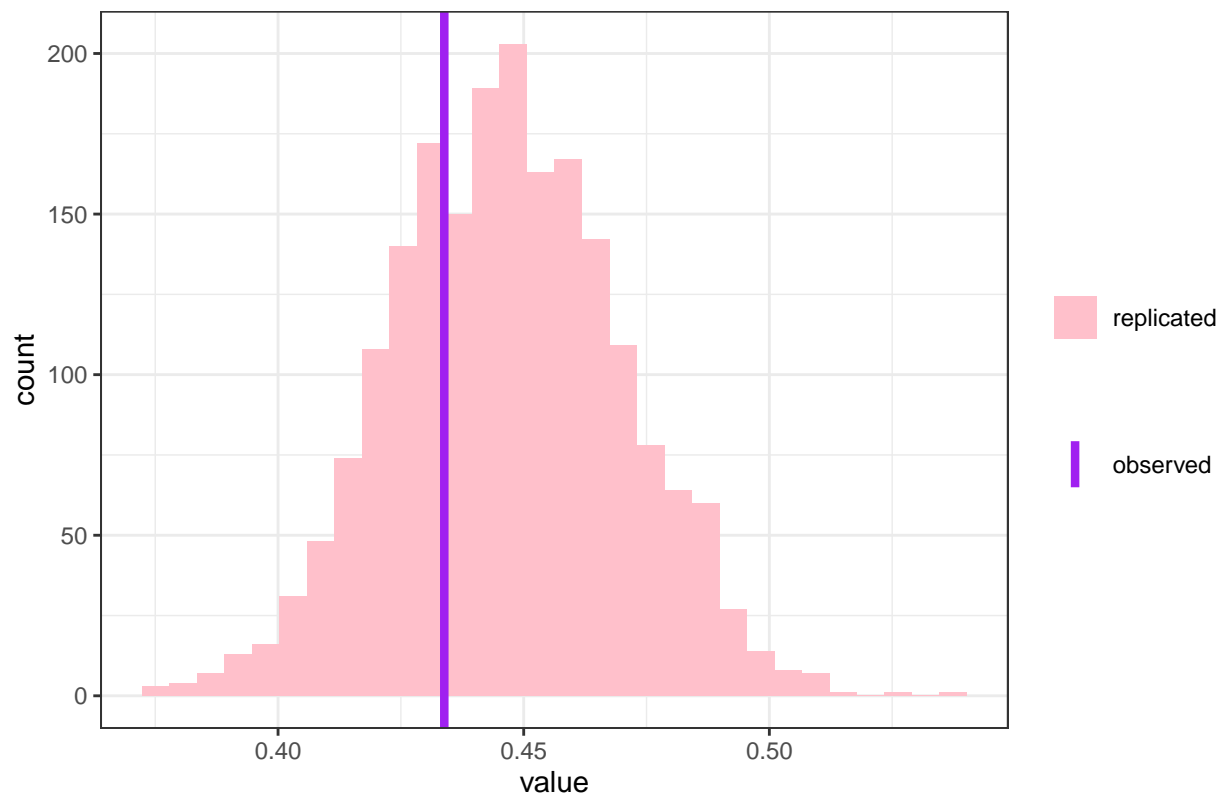
For every unit increase in demeaned distance multiplied by demeaned arsenic levels, the odds of switching a well are multiplied by  $\exp(-0.001775180) \approx 0.998$ , assuming all other factors are held constant. As seen in the previous graph of arsenic against distance, there isn't much correlation between arsenic and distance. So arsenic and distance joined together don't have much influence on switching wells.

### Question 3c

Model 1: proportion of households that switched wells with arsenic level be



Model 2: proportion of households that switched wells with arsenic level be



Model 2 was better, because most of the replicated test statistics are closer to the observed test statistic. We note that the generated test statistics for model 1 and model 2 follow a normal distribution, so we use `pnorm()` to compute the probability that the generated test statistics are less than the true test statistic.

```
## [1] 0.006085676
```

```
## [1] 0.2936043
```

For model 1,  $P(t(\mathbf{y}^{rep}) < t(\mathbf{y})) = 0.006085676$ . This means that the true value of the test statistic is in the tails of the distribution of the generated test statistics.

For model 2,  $P(t(\mathbf{y}^{rep}) < t(\mathbf{y})) = 0.2936043$ , which is closer to 0.5. This means that the true value of the test statistic is closer to the median of the distribution of the generated test statistics. This confirms that Model 2 is better.

### Question 3d

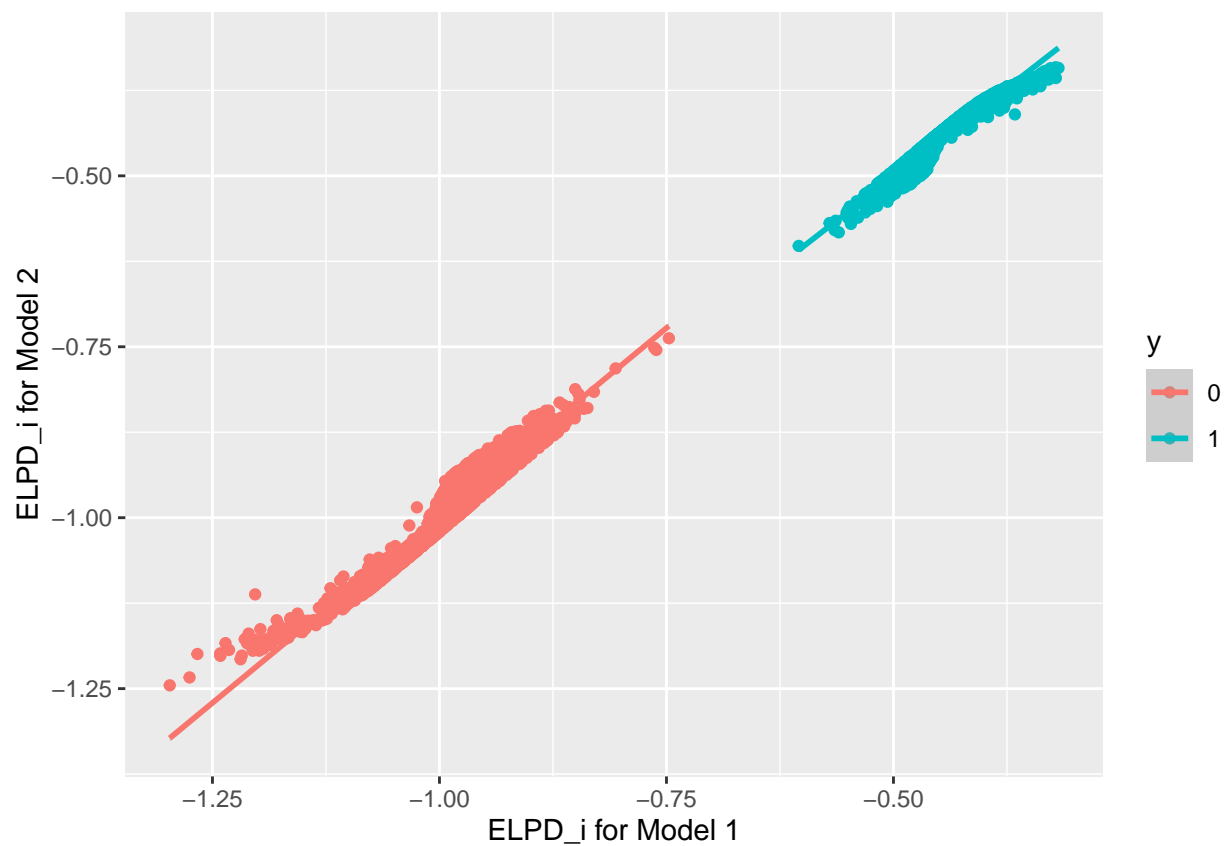
```
##
## Computed from 2000 by 3020 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo -2046.2 15.6
## p_loo      0.2  0.0
## looic      4092.4 31.1
## -----
## Monte Carlo SE of elpd_loo is 0.0.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.

##
## Computed from 2000 by 3020 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo -2039.8 15.6
## p_loo      0.2  0.0
## looic      4079.7 31.2
## -----
## Monte Carlo SE of elpd_loo is 0.0.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.

##           elpd_diff se_diff
## model2  0.0         0.0
## model1 -6.3         0.9
```

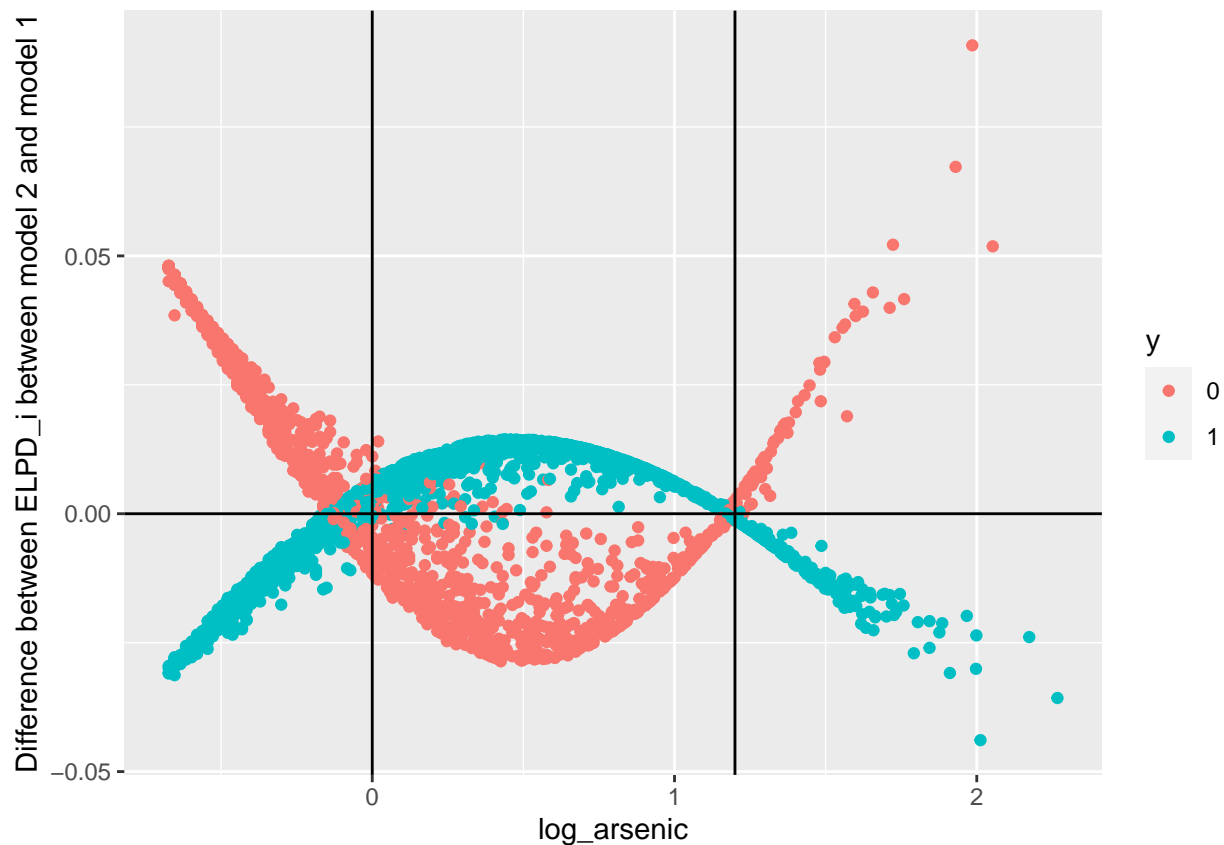
Model 2 is better because it has a higher `ELPD_loo`.

### Question 3e



The slope through the points appears to be close to 1. We see that  $ELPD_i$ 's are for model 2 and model 1 are close to each other, regardless of whether people switch their wells or not. So Model 2 and Model 1 are similar among people who switched wells and among people who did not switch wells.



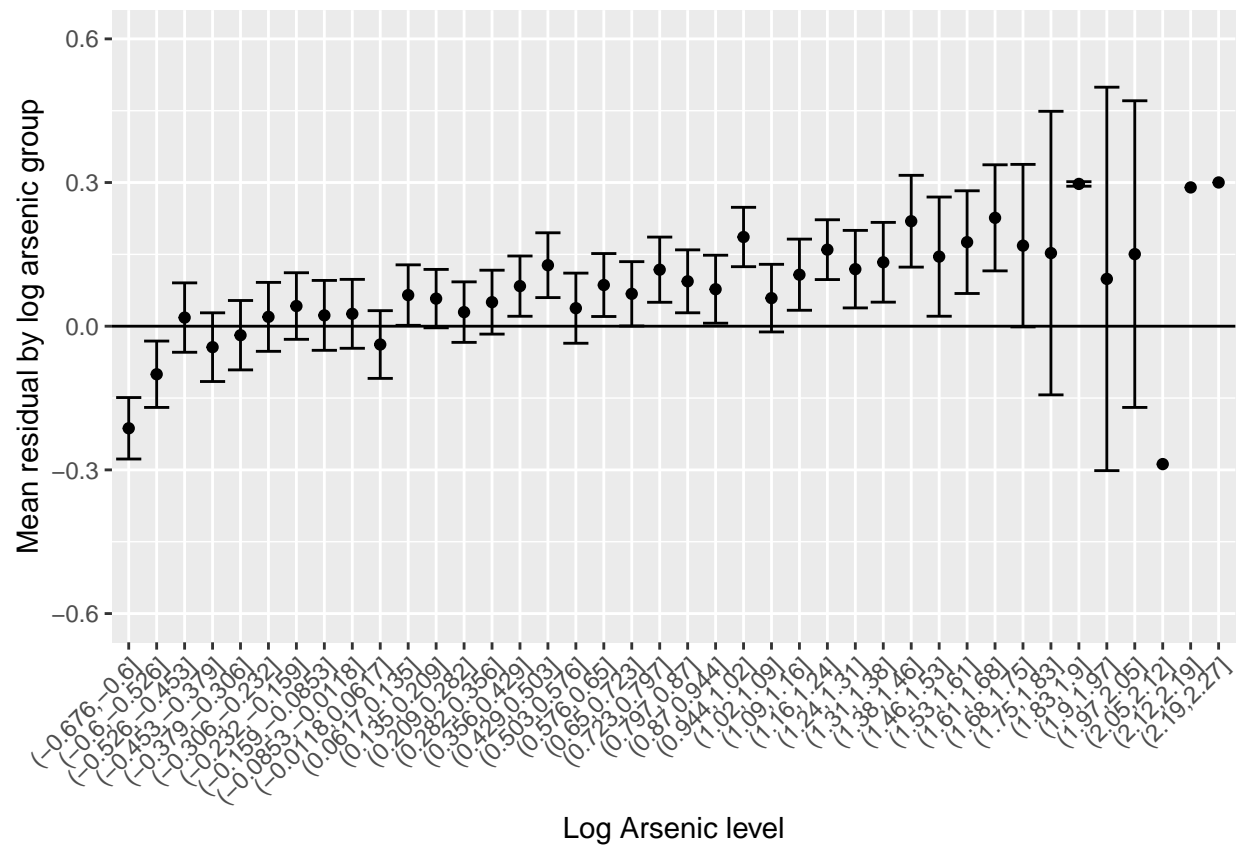
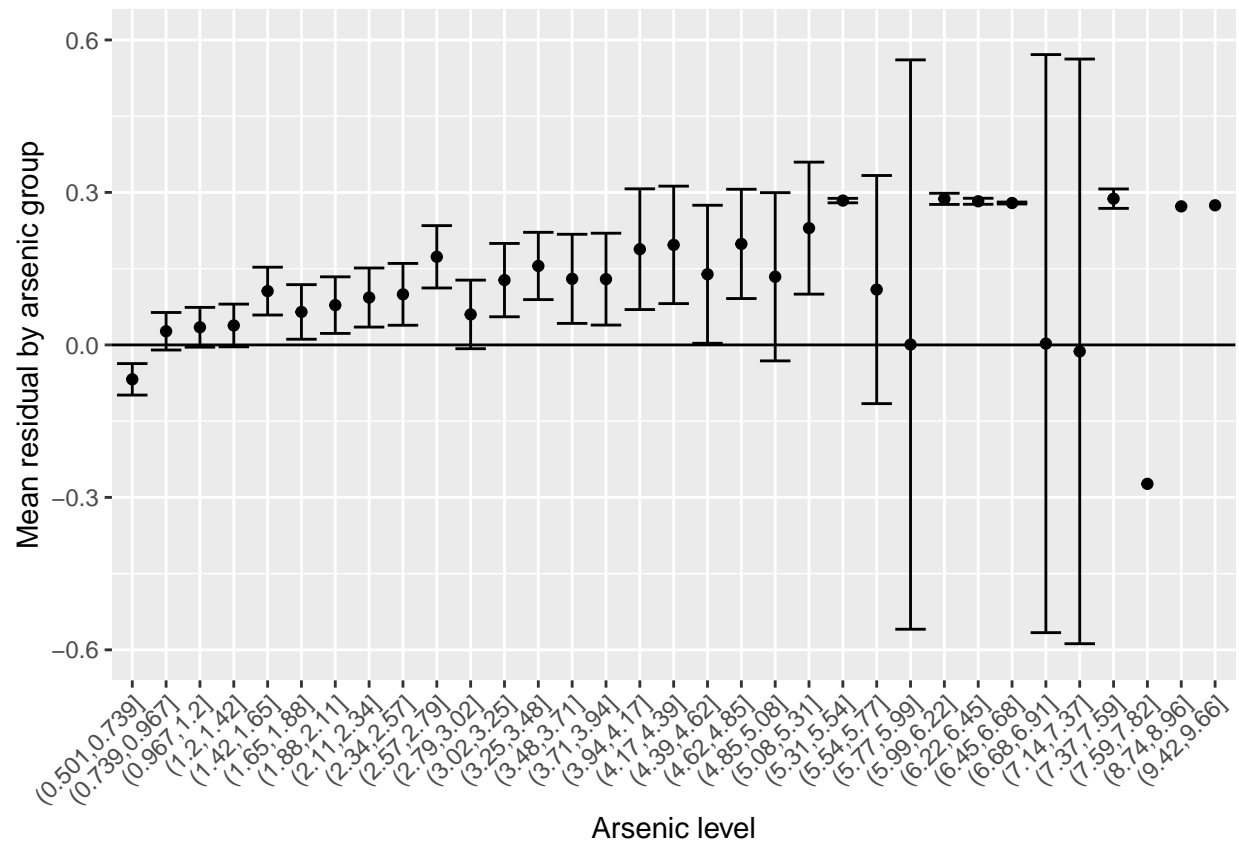


For households that did not switch wells and had extremely high arsenic ( $\log(\text{arsenic}) > 1.2$  or  $\text{arsenic} > \exp(1.2)$ ) or extremely low arsenic ( $\text{arsenic} < 1$  or  $\log(\text{arsenic}) < 0$ ), the difference between  $ELPD_i$  for model 2 and model 1 is positive, so Model 2 is better. Model 2 is also better for households who switched wells and had log-arsenic level in  $(0, 1.2)$ . For the other groups, Model 1 is better.

### Question 3f

$\exp(ELPD_i) = \exp(\log(P(y_i|y_{-i}))) = P(y_i|y_{-i})$  is the expected leave one out probability of switching wells for household  $i$ .

Question 3g



We can see that for model 2, the mean residuals for most of the arsenic level groups are closer to 0 than they are for model 1. This suggests that model 2 predicts better for most arsenic ranges. The standard errors of the residuals for model 2 also seem to be lower than they are for model 1, so model 2 also predicts with more certainty. For both models, as  $\log(\text{arsenic})$  level increases, the mean residuals are further from 0 so the predictions are not as accurate.