| EEEB UN3005/GR5005 | Homework - Week 14 - Due 07 May 2019

Xun Zhao, xz2827

Homework Instructions: Complete this assignment by writing code in the code chunks provided. If required, provide written explanations below the relevant code chunks. Replace "USE YOUR NAME HERE" with your name in the document header. When complete, knit this document within RStudio to generate a pdf. Please review the resulting pdf to ensure that all content relevant for grading (i.e., code, code output, and written explanations) appears in the document. Rename your pdf document according to the following format: hw_week_14_firstname_lastname.pdf. Upload this final homework document to CourseWorks by 5 pm on the due date.

This week's homework problems will use a historical human ecology/demography dataset from 1988 on nearly 2,000 Bangladeshi women. Access this dataset, which includes information on the women's households and behavior, from the `rethinking` package using `data("bangladesh")`.

Problem 1 (1.5 points)

Import the `bangladesh` data and **filter the data to only contain information on women from administrative districts 1 through 50 (the `district` variable).** Use this filtered dataset for all problems in this assignment. If you don't, you're sure to run into some modeling trouble.

First, fit a binomial generalized linear model using `district` to predict a woman's decision to use contraception (the `use.contraception` variable). Note, treating `district` as an index variable to generate a vector of intercept parameters (one corresponding to each district) will be the most effective approach here.

After fitting the model, use `precis()` to report the 97% PIs of the fit model parameters.

```
data("bangladesh")
d = filter(bangladesh, district >= 1 & district <= 50)
model.1 = map(
    alist(
        use.contraception ~ dbinom(1, p),
        logit(p) <- a[district],
        a[district] ~ dnorm(0, 5)),
    data = d)
precis(model.1, 0.97)
```

```
##        Mean StdDev  5.5% 94.5%
```

```
## a[1]  -1.06   0.21 -1.40 -0.72
## a[2]  -0.61   0.47 -1.36  0.13
## a[3]   2.82   2.61 -1.36  6.99
## a[4]   0.00   0.36 -0.58  0.58
## a[5]  -0.58   0.33 -1.11 -0.05
## a[6]  -0.88   0.27 -1.32 -0.45
## a[7]  -0.94   0.52 -1.78 -0.11
## a[8]  -0.49   0.34 -1.03  0.05
## a[9]  -0.82   0.45 -1.54 -0.10
## a[10] -2.39   0.98 -3.95 -0.82
## a[11] -4.71   2.10 -8.07 -1.35
## a[12] -0.64   0.39 -1.26 -0.02
## a[13] -0.33   0.41 -0.99  0.33
## a[14]  0.52   0.19  0.22  0.82
## a[15] -0.56   0.44 -1.26  0.15
## a[16]  0.20   0.45 -0.52  0.91
## a[17] -0.88   0.45 -1.59 -0.17
## a[18] -0.66   0.31 -1.15 -0.17
## a[19] -0.47   0.40 -1.11  0.17
## a[20] -0.40   0.52 -1.24  0.44
## a[21] -0.45   0.48 -1.22  0.32
## a[22] -1.37   0.55 -2.25 -0.49
## a[23] -1.00   0.58 -1.92 -0.07
## a[24] -2.46   0.97 -4.02 -0.91
## a[25] -0.21   0.25 -0.60  0.18
## a[26] -0.46   0.57 -1.37  0.44
## a[27] -1.49   0.39 -2.12 -0.87
## a[28] -1.12   0.33 -1.65 -0.59
## a[29] -0.93   0.39 -1.56 -0.31
## a[30] -0.03   0.26 -0.44  0.38
## a[31] -0.18   0.35 -0.74  0.38
## a[32] -1.32   0.50 -2.12 -0.53
## a[33] -0.28   0.54 -1.14  0.57
## a[34]  0.65   0.36  0.08  1.21
## a[35]  0.00   0.29 -0.46  0.46
## a[36] -0.60   0.50 -1.41  0.21
## a[37]  0.15   0.55 -0.73  1.04
## a[38] -0.90   0.59 -1.84  0.03
## a[39]  0.00   0.39 -0.63  0.62
## a[40] -0.15   0.31 -0.65  0.35
## a[41]  0.00   0.39 -0.63  0.62
## a[42]  0.18   0.60 -0.78  1.14
## a[43]  0.13   0.30 -0.34  0.61
## a[44] -1.24   0.46 -1.98 -0.51
## a[45] -0.69   0.34 -1.23 -0.15
```

```
## a[46]  0.09   0.22 -0.25  0.44
## a[47] -0.13   0.51 -0.95  0.69
## a[48]  0.09   0.31 -0.40  0.59
## a[49] -3.36   2.43 -7.24  0.52
## a[50] -0.10   0.46 -0.84  0.63
```

Problem 2 (2.5 points)

Now, formulate and fit this same model as a multilevel binomial generalized linear model. Don't worry about any warning messages you may receive regarding divergent iterations during sampling.

After fitting the model, use `precis()` to report the 97% HPDIs of the fit model parameters.

Using posterior samples from the model, visualize the implied probability of contraception use for a woman occupying an *average* district.

```
model.2 = map2stan(
    alist(
        use.contraception ~ dbinom(1, p),
        logit(p) <- a[district],
        a[district] ~ dnorm(mu, sigma),
        mu ~ dnorm(0, 5),
        sigma ~ dcauchy(0, 5)),
    chains = 2,
    data = d)
```

```
## Warning: Variable 'use.contraception' contains dots '.'.
## Will attempt to remove dots internally.

## Warning: Variable 'living.children' contains dots '.'.
## Will attempt to remove dots internally.

## Warning: Variable 'age.centered' contains dots '.'.
## Will attempt to remove dots internally.

## Warning: There were 1 divergent transitions after warmup. Increasing adapt_delta abov
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Warning: Examine the pairs() plot to diagnose sampling problems

## Computing WAIC

## Constructing posterior predictions
```

```
precis(model.2, 0.97)
```

```
##        Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
## a[1]  -0.99   0.20      -1.31      -0.67  2237 1.00
## a[2]  -0.59   0.34      -1.15      -0.05  2646 1.00
```

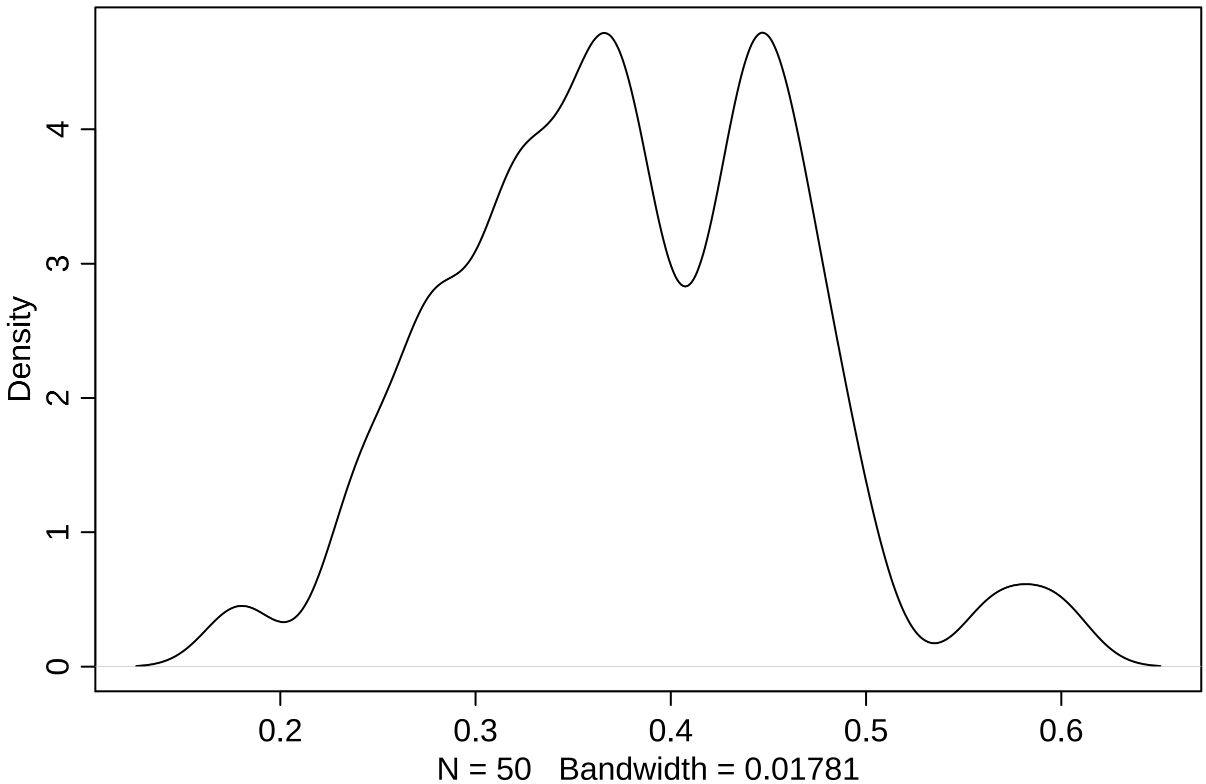
```
## a[3]  -0.22   0.50       -1.03        0.51  2299 1.00
## a[4]  -0.19   0.30       -0.66        0.30  2740 1.00
## a[5]  -0.56   0.27       -0.99       -0.13  2409 1.00
## a[6]  -0.81   0.24       -1.18       -0.40  2392 1.00
## a[7]  -0.74   0.37       -1.29       -0.16  2270 1.00
## a[8]  -0.51   0.29       -0.99       -0.07  2740 1.00
## a[9]  -0.70   0.34       -1.23       -0.13  2320 1.00
## a[10] -1.11   0.43       -1.74       -0.39  1539 1.00
## a[11] -1.52   0.43       -2.23       -0.86  1126 1.00
## a[12] -0.60   0.32       -1.14       -0.11  2589 1.00
## a[13] -0.42   0.32       -0.93        0.11  2894 1.00
## a[14]  0.39   0.18        0.11        0.68  2136 1.00
## a[15] -0.56   0.34       -1.09        0.01  2459 1.00
## a[16] -0.11   0.33       -0.66        0.40  2434 1.00
## a[17] -0.74   0.33       -1.30       -0.28  2785 1.00
## a[18] -0.62   0.27       -1.06       -0.19  2530 1.00
## a[19] -0.49   0.31       -0.97        0.01  2491 1.00
## a[20] -0.47   0.37       -1.12        0.03  2594 1.00
## a[21] -0.50   0.34       -1.00        0.07  2695 1.00
## a[22] -0.94   0.37       -1.47       -0.30  1978 1.00
## a[23] -0.75   0.38       -1.36       -0.13  2468 1.00
## a[24] -1.14   0.43       -1.85       -0.52  1603 1.00
## a[25] -0.27   0.22       -0.64        0.05  2867 1.00
## a[26] -0.50   0.39       -1.13        0.11  3155 1.00
## a[27] -1.18   0.31       -1.62       -0.66  1829 1.00
## a[28] -0.95   0.27       -1.35       -0.49  1872 1.00
## a[29] -0.79   0.32       -1.27       -0.26  2248 1.00
## a[30] -0.13   0.22       -0.49        0.21  2308 1.00
## a[31] -0.29   0.29       -0.77        0.17  2491 1.00
## a[32] -0.97   0.36       -1.55       -0.41  1649 1.00
## a[33] -0.41   0.37       -1.01        0.14  2230 1.00
## a[34]  0.27   0.29       -0.23        0.69  1899 1.00
## a[35] -0.14   0.25       -0.58        0.23  2586 1.00
## a[36] -0.56   0.36       -1.08        0.06  2627 1.00
## a[37] -0.22   0.40       -0.80        0.43  2700 1.00
## a[38] -0.71   0.38       -1.30       -0.12  2219 1.00
## a[39] -0.19   0.32       -0.74        0.28  2148 1.00
## a[40] -0.25   0.28       -0.70        0.18  2464 1.00
## a[41] -0.19   0.32       -0.65        0.36  3250 1.00
## a[42] -0.23   0.40       -0.83        0.42  2514 1.00
## a[43] -0.05   0.26       -0.46        0.37  2339 1.00
## a[44] -0.94   0.33       -1.47       -0.42  1851 1.00
## a[45] -0.65   0.29       -1.14       -0.21  2336 1.00
## a[46]  0.00   0.21       -0.33        0.32  2507 1.00
## a[47] -0.34   0.36       -0.91        0.21  2643 1.00
```

```
## a[48] -0.08   0.27      -0.52       0.33  2530 1.00
## a[49] -0.83   0.47      -1.64      -0.11  1809 1.00
## a[50] -0.30   0.36      -0.82       0.29  2604 1.00
## mu    -0.52   0.09      -0.66      -0.36  1272 1.00
## sigma  0.51   0.09       0.36       0.64   633 1.01
```

```r
sample.2 = extract.samples(model.2, 2000)
```

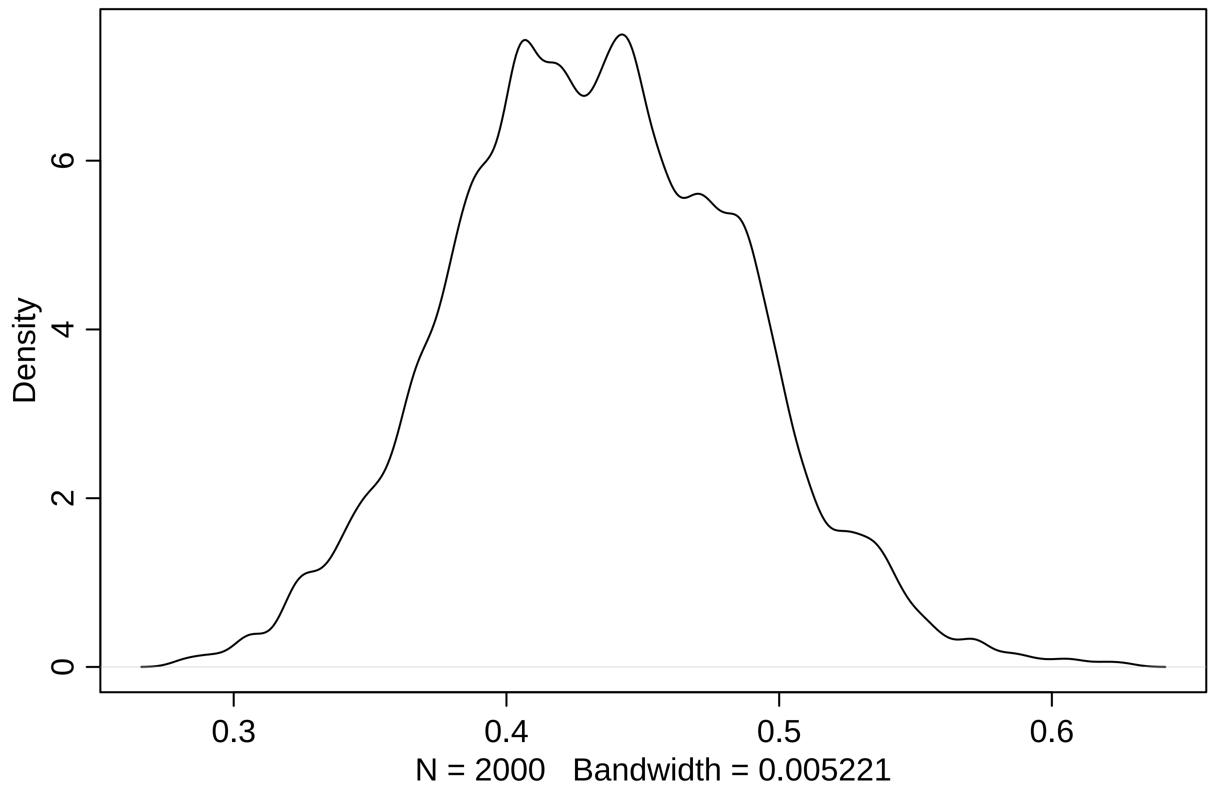
If *average* means the average of all distrcts:

```r
dens(logistic(colMeans(sample.2$a)))
```



If *average* means the district with average label number (25, the mean of 1 and 50):

```r
dens(logistic(sample.2$a[,25]))
```



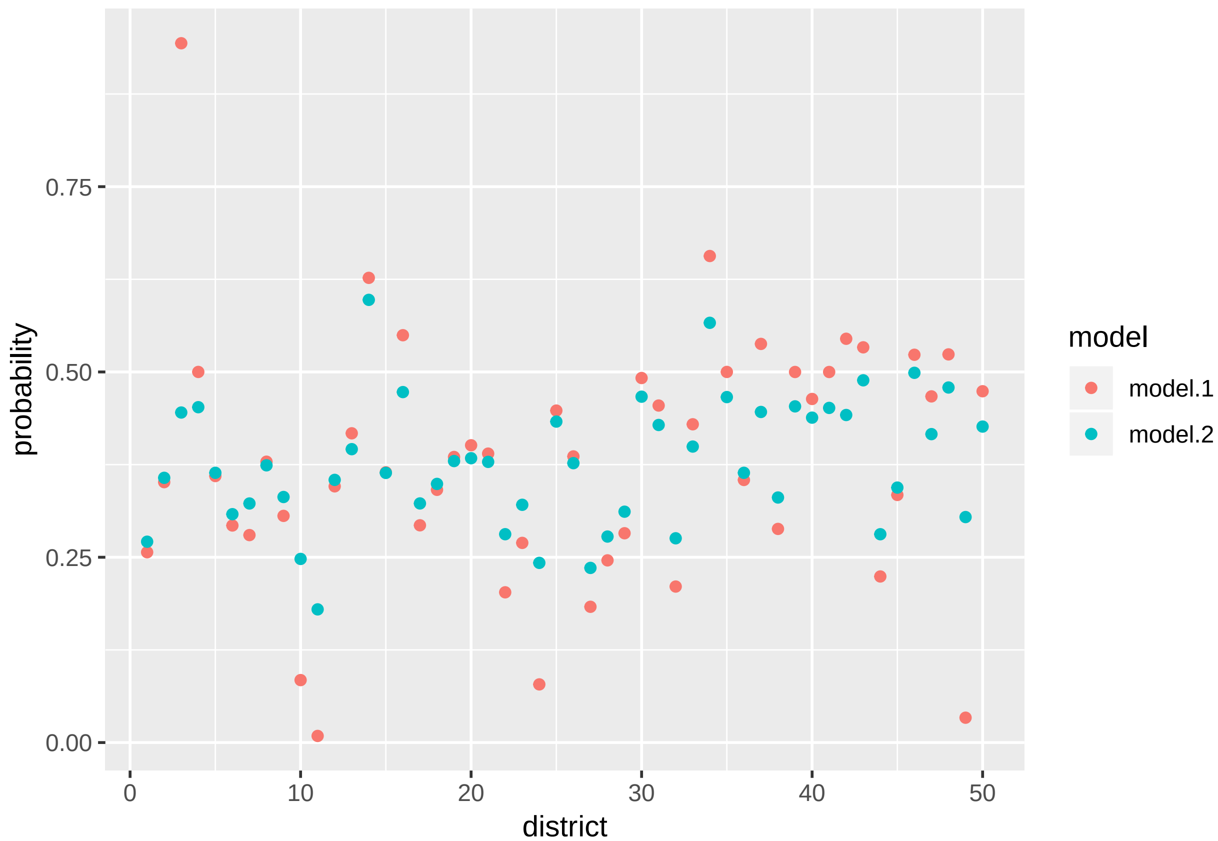
Problem 3 (4 points)

On one plot, show the implied probability of contraception use values for each district from each of the two models you've fit so far. District ID should appear on the x-axis of this plot, while implied probability of contraception use values should appear on the y-axis. It will likely help to distinguish the estimates from each model using different colors. For this plot, you can ignore uncertainty in the posterior estimates for each district, and instead simply plot the mean value for each district.

Some hints may be useful here. First, you can use any plotting approach you'd like, but base R plotting techniques will avoid the extra step of having to package together estimates from the two different models into one data frame. Second, note that you can access the mean parameter estimates from a model using the `coef()` function like so: `coef(my_model)`. For a binomial model with a logit link, these estimates will of course appear on the log-odds scale, so keep that in mind. Finally, note that the `points()` function will allow you to add data points to an existing plot. So perhaps the easiest strategy here is to set up the plot with the estimates from one model (using a base R plot), then layer on top the estimates from the second model using `points()`.

Using your visualization to help with interpretation, how do these two models, the binomial GLM and the binomial GLMM, disagree on the implied probability of contraception use among districts? In which districts is the disagreement most extreme? Can you explain why?

```
coef.1 = coef(model.1)
coef.2 = coef(model.2)[1:50]
df = data.frame(district = c(1:50, 1:50),
                prob = logistic(c(coef.1, coef.2)),
                model = c(rep('model.1', 50), rep('model.2', 50)))
graph = ggplot(df, aes(x = district, y = prob, color = model)) +
    geom_point() +
    xlab("district") + ylab("probability") +
    labs(fill = "model")
plot(graph)
```



Answer:

The biggest difference is that GLMM's reuslts are closer to each other, and none of them is extremely high or extremely low. However, In GLM's results, some points are scattered far away from the center line.

In the 3rd district, GLM gives out a result over 0.8, but GLMM's result is less that 0.5. This is because that the GLMM use the hyper-parameter to restrict the model-parameter, making sure that they follow certain distribution.

Problem 4 (2 points)

Adapt the model you constructed in Problem 2 to also consider the impact of whether or not a woman lives in an urban area as a predictor for contraception use. Note, `urban` is already a dummy variable, with a value of 1 indicating a woman who lives in a city and a value of 0 indicating a woman who lives in a rural area.

After fitting the new model, use `precis()` to report the 97% HPDIs of the fit model parameters.

Interpret the urban living effect. Based on this dataset and model, are urban women more or less likely to use contraception?

```
model.3 = map2stan(
    alist(
        use.contraception ~ dbinom(1, p),
        logit(p) <- a[district] + b[district] * urban,
        a[district] ~ dnorm(mua, sigmaa),
        mua ~ dnorm(0, 5),
        sigmaa ~ dcauchy(0, 5),
        b[district] ~ dnorm(mub, sigmab),
        mub ~ dnorm(0, 5),
        sigmab ~ dcauchy(0, 5)),
    chains = 2,
    data = d)
```

```
## Warning: Variable 'use.contraception' contains dots '.'.
## Will attempt to remove dots internally.

## Warning: Variable 'living.children' contains dots '.'.
## Will attempt to remove dots internally.

## Warning: Variable 'age.centered' contains dots '.'.
## Will attempt to remove dots internally.

## Warning: There were 11 divergent transitions after warmup. Increasing adapt_delta abo
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Warning: There were 2 chains where the estimated Bayesian Fraction of Missing Informa
## http://mc-stan.org/misc/warnings.html#bfmi-low

## Warning: Examine the pairs() plot to diagnose sampling problems

## Warning: There were 1 divergent transitions after warmup. Increasing adapt_delta abov
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Warning: Examine the pairs() plot to diagnose sampling problems

## Computing WAIC
```

```
## Constructing posterior predictions
```

```
## Warning in map2stan(alist(use.contraception ~ dbinom(1, p), logit(p) <- a[district] +
## Check the chains (trace plots, n_eff, Rhat) carefully to ensure they are valid.
```

```r
precis(model.3, 0.97)
```

```
## Warning in precis(model.3, 0.97): There were 11 divergent iterations during sampling.
## Check the chains (trace plots, n_eff, Rhat) carefully to ensure they are valid.
```

```
##          Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
## a[1]    -1.42   0.27      -1.86      -1.02  1203 1.00
## a[2]    -0.68   0.32      -1.18      -0.15  3061 1.00
## a[3]    -0.54   0.49      -1.32       0.20  2718 1.00
## a[4]    -0.59   0.32      -1.11      -0.11  1348 1.00
## a[5]    -0.67   0.27      -1.13      -0.26  1477 1.00
## a[6]    -0.94   0.24      -1.36      -0.58  1557 1.00
## a[7]    -0.83   0.36      -1.40      -0.28  1993 1.00
## a[8]    -0.62   0.28      -1.07      -0.17  3000 1.00
## a[9]    -0.84   0.32      -1.37      -0.33  2699 1.00
## a[10]   -1.16   0.40      -1.77      -0.53  1737 1.00
## a[11]   -1.54   0.43      -2.21      -0.89  1021 1.00
## a[12]   -0.74   0.30      -1.24      -0.30  2432 1.00
## a[13]   -0.61   0.33      -1.15      -0.09  2188 1.00
## a[14]   -0.49   0.30      -1.00      -0.04   481 1.00
## a[15]   -0.77   0.34      -1.33      -0.26  2103 1.00
## a[16]   -0.29   0.33      -0.79       0.26  1882 1.00
## a[17]   -0.81   0.34      -1.37      -0.32  1612 1.00
## a[18]   -0.86   0.28      -1.28      -0.39  2732 1.00
## a[19]   -0.67   0.33      -1.19      -0.13  1158 1.00
## a[20]   -0.57   0.34      -1.10      -0.04  2776 1.00
## a[21]   -0.65   0.37      -1.23      -0.07  1032 1.00
## a[22]   -1.01   0.35      -1.60      -0.46  1787 1.00
## a[23]   -0.84   0.34      -1.38      -0.30  2729 1.00
## a[24]   -1.21   0.40      -1.82      -0.57  1719 1.00
## a[25]   -0.43   0.24      -0.81      -0.05  1337 1.00
## a[26]   -0.62   0.38      -1.23      -0.04  2264 1.00
## a[27]   -1.26   0.31      -1.73      -0.75  1707 1.00
## a[28]   -1.04   0.26      -1.45      -0.63  1929 1.00
## a[29]   -0.98   0.32      -1.44      -0.44  1501 1.00
## a[30]   -0.41   0.25      -0.80      -0.03  2338 1.00
## a[31]   -0.46   0.29      -0.95      -0.02  1866 1.00
## a[32]   -1.01   0.33      -1.49      -0.43  1919 1.00
## a[33]   -0.77   0.40      -1.38      -0.12  1196 1.00
## a[34]    0.13   0.32      -0.38       0.66   689 1.01
## a[35]   -0.41   0.28      -0.87      -0.02  1577 1.00
```

```
## a[36]  -0.72   0.35      -1.28      -0.17  2418 1.00
## a[37]  -0.38   0.40      -1.04       0.22  1436 1.00
## a[38]  -1.01   0.41      -1.68      -0.40  1184 1.00
## a[39]  -0.34   0.31      -0.85       0.14  1902 1.00
## a[40]  -0.65   0.33      -1.14      -0.12  1517 1.00
## a[41]  -0.32   0.32      -0.82       0.16  1975 1.00
## a[42]  -0.44   0.41      -1.11       0.18  2083 1.00
## a[43]  -0.32   0.28      -0.78       0.10  2057 1.00
## a[44]  -1.00   0.35      -1.57      -0.46  1312 1.00
## a[45]  -0.81   0.28      -1.20      -0.33  2635 1.00
## a[46]  -0.15   0.21      -0.50       0.17  1962 1.00
## a[47]  -0.59   0.39      -1.21       0.02  2394 1.00
## a[48]  -0.33   0.27      -0.77       0.10  1748 1.00
## a[49]  -0.96   0.45      -1.65      -0.22  2041 1.00
## a[50]  -0.54   0.33      -1.06       0.00  2272 1.00
## mua    -0.70   0.09      -0.84      -0.55  1265 1.00
## sigmaa  0.47   0.10       0.31       0.61   375 1.02
## b[1]    0.86   0.31       0.41       1.37  1392 1.00
## b[2]    0.83   0.58      -0.04       1.73  2943 1.00
## b[3]    1.05   0.58       0.22       2.05   806 1.00
## b[4]    1.42   0.59       0.57       2.32   151 1.01
## b[5]    0.80   0.51       0.05       1.59  2538 1.00
## b[6]    1.13   0.49       0.41       1.90   363 1.00
## b[7]    0.82   0.61      -0.22       1.67  2293 1.00
## b[8]    1.05   0.54       0.26       1.95   683 1.00
## b[9]    0.95   0.50       0.22       1.82  1378 1.00
## b[10]   0.82   0.55      -0.05       1.69  2478 1.00
## b[11]   0.80   0.58      -0.07       1.74  2171 1.00
## b[12]   0.61   0.47      -0.06       1.39   532 1.00
## b[13]   0.58   0.47      -0.24       1.25   588 1.00
## b[14]   1.14   0.33       0.65       1.66   340 1.00
## b[15]   0.64   0.44      -0.09       1.29   776 1.00
## b[16]   1.01   0.57       0.09       1.83  1071 1.00
## b[17]   0.83   0.58      -0.07       1.72  1975 1.00
## b[18]   0.83   0.39       0.14       1.39  2107 1.00
## b[19]   1.00   0.51       0.31       1.91  1239 1.00
## b[20]   0.83   0.57      -0.08       1.65  2310 1.00
## b[21]   0.24   0.61      -0.66       1.12   134 1.01
## b[22]   0.83   0.58      -0.09       1.76  2494 1.00
## b[23]   0.82   0.61      -0.04       1.79  2359 1.00
## b[24]   0.83   0.58      -0.03       1.79  2278 1.00
## b[25]   0.51   0.41      -0.09       1.16   261 1.00
## b[26]   0.81   0.60      -0.09       1.71  2452 1.00
## b[27]   0.83   0.50       0.04       1.60  2455 1.00
## b[28]   0.63   0.51      -0.15       1.44   794 1.00
```

```
## b[29]   0.96   0.47      0.22      1.67  1468 1.00
## b[30]   1.13   0.40      0.55      1.78   387 1.00
## b[31]   0.71   0.47      0.01      1.48  1437 1.00
## b[32]   0.82   0.61     -0.19      1.68  2052 1.00
## b[33]   1.08   0.49      0.33      1.87   395 1.00
## b[34]   0.24   0.59     -0.66      1.10   145 1.01
## b[35]   0.72   0.37      0.12      1.28  1011 1.00
## b[36]   0.67   0.50     -0.11      1.42  1242 1.00
## b[37]   0.81   0.58     -0.01      1.78  2376 1.00
## b[38]   0.97   0.47      0.24      1.70  1362 1.00
## b[39]   0.75   0.50     -0.04      1.49  2201 1.00
## b[40]   0.68   0.39      0.00      1.21   954 1.00
## b[41]   0.38   0.62     -0.49      1.32   225 1.00
## b[42]   0.42   0.58     -0.49      1.24   212 1.01
## b[43]   0.75   0.38      0.13      1.35  1868 1.00
## b[44]   0.81   0.60     -0.18      1.70  2457 1.00
## b[45]   1.12   0.51      0.41      1.98   403 1.00
## b[46]   0.83   0.40      0.22      1.49  2251 1.00
## b[47]   0.76   0.47      0.04      1.50  2176 1.00
## b[48]   0.71   0.39      0.05      1.32  1812 1.00
## b[49]   0.81   0.60     -0.17      1.68  2412 1.00
## b[50]   1.21   0.59      0.36      2.19   308 1.01
## mub     0.82   0.17      0.57      1.10   586 1.00
## sigmab  0.51   0.26      0.10      0.85    60 1.02
```

Answer:

The results show that all the `b` parameter is positive. For certain distrct, the probability of a woman to use contraception is `logistic(a + b * urban)`. So, when the b is positive, that is, the urban women are more likely to use contraception, since `logistic(a + b)` is larger than `logistic(a)`.
