Denis Ostroushko - PUBH 7440 - HW1

Problem 1

Problem 2

2 - A

According to given parameters, prior distribution of θ is given by $\theta \sim N(0, 2)$

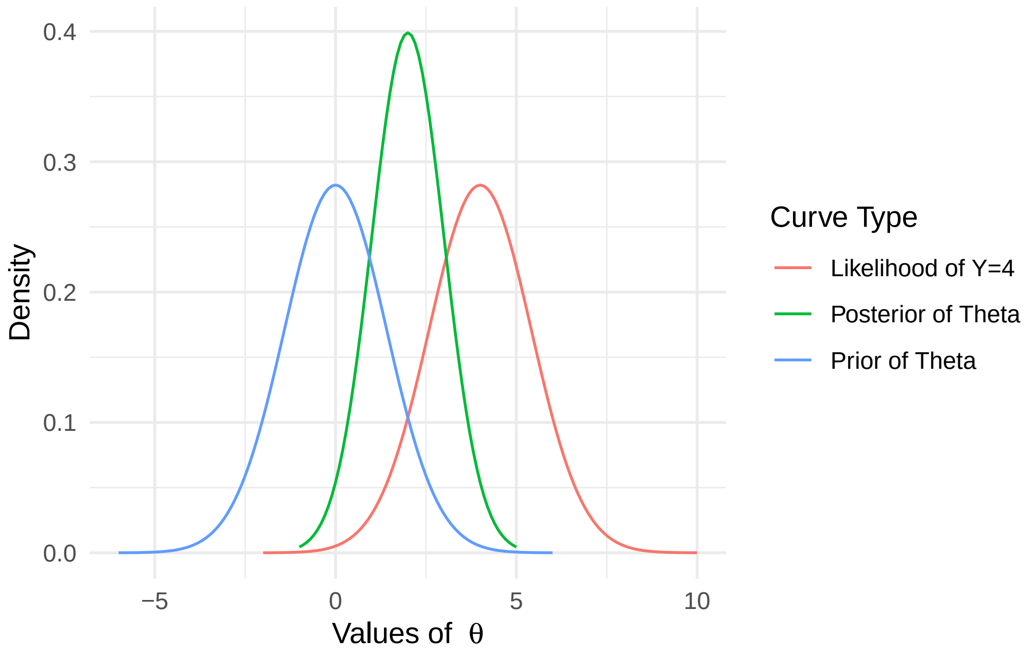
And the distribution of data, random variable Y is also normal, $y|\theta \sim N(\theta, 2)$

Using given derivation of posterior distribution, we obtain the distribution of θ given the observed values of data y.

Going forward, we fix $y = 4$, we let $B = \frac{2}{2+2} = \frac{1}{2}$. $\theta|y \sim N(B \times \mu + (1-B) \times y, (1-B) \times \sigma^2) = N(\frac{1}{2} \times 0 + \frac{1}{2} \times 4, \frac{1}{2} \times 2) = N(2, 1)$.

So, $\theta|y \sim N(2, 1)$

To plot likelihood of observing y= 4, I will use the distribution of data $y|\theta$, I will vary the value of θ and keep variance of Y fixed at 2. Since θ is centered at 0 with variance 2, I will obtain values that correspond to prior distribution of θ



2 - B

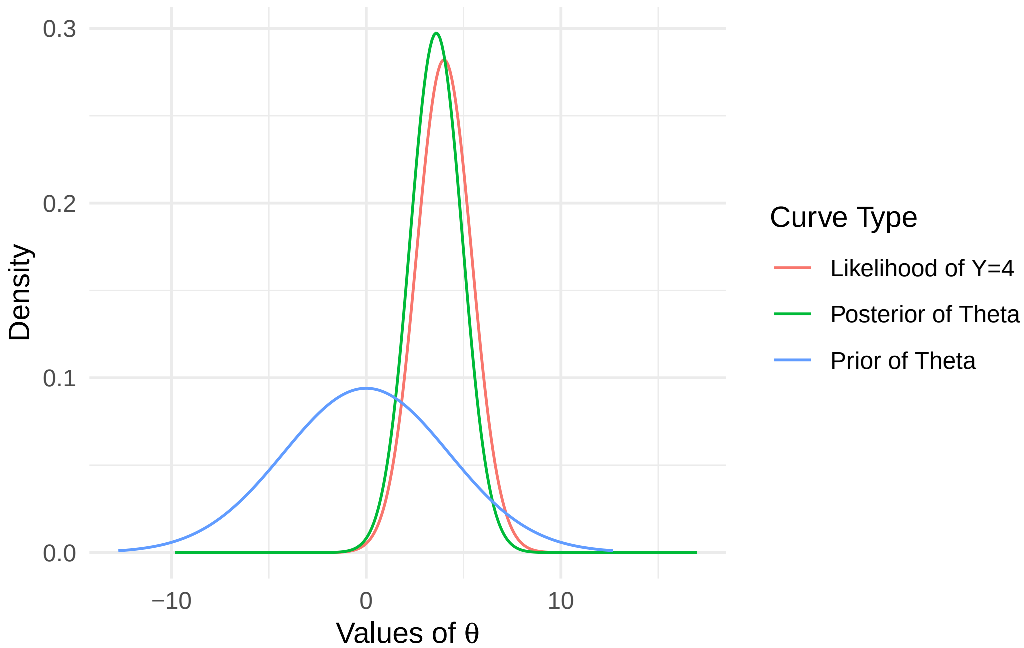
When $\tau^2 = 18$, $B = \frac{2}{2+18} = \frac{1}{10}$. The resulting distributions are then

$\theta \sim N(0, 18)$

$y|\theta \sim N(\theta, 2)$

Assuming y= 4, $\theta|y \sim N(0.1 * 0 + 0.9 * y, 0.9 * 2) = N(3.6, 1.8)$

Prior, marginal likelihood, and posterior distributions take on these shapes:



Problem 3

3 - A

The chance from picking a ball from bucket 2 is governed by a fair six-sided die. We pick from bucket 2 when we roll 5 or 6, so, $P(Draw\ from\ 2) = \frac{1}{3}$

3 - B

Probability of drawing a blue ball is a weighted average of probabilities that correspond to the bucket that we draw from. $P(Blue) = P(Draw\ from\ 1) \times P(Blue|Draw\ from\ 1) + P(Draw\ from\ 2) \times P(Blue|Draw\ from\ 2) = \frac{2}{3} \times \frac{17}{17+35} + \frac{1}{3} \times \frac{37}{37+23} = 0.42$

3 - C

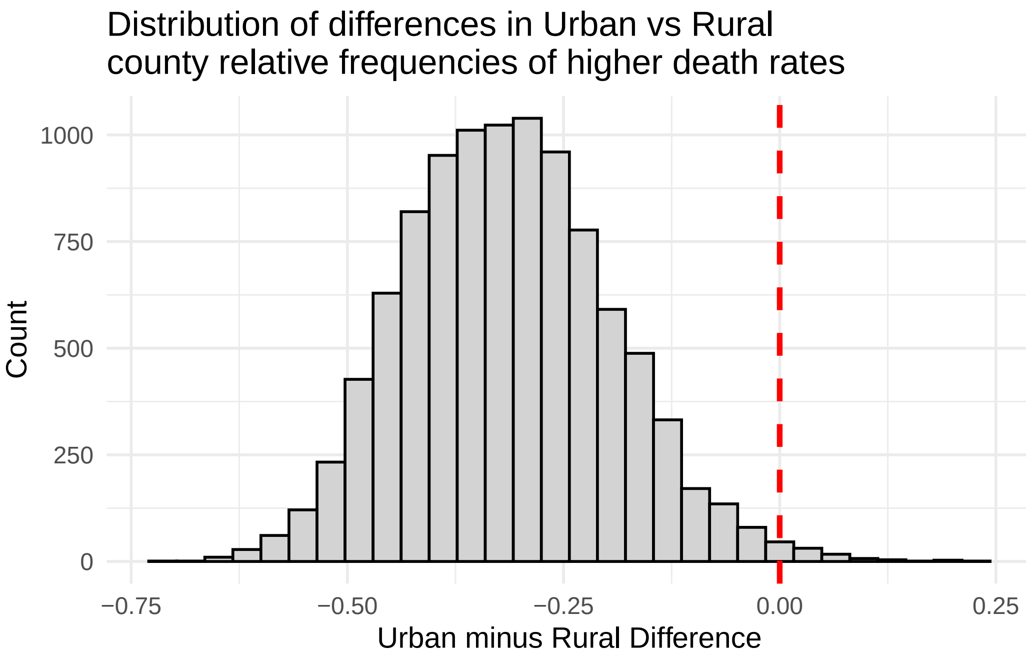
We need to use Bayes Rule to find this probability from available data

$P(Draw\ from\ 2|Blue) = \frac{P(Draw\ from\ 2\ and\ Blue)}{P(Blue)} = \frac{P(Blue|Draw\ from\ 2) \times P(Draw\ from\ 2)}{P(Blue)} = \frac{\frac{37}{37+23} \times \frac{1}{3}}{0.42} = 0.49$

Problem 4

4 - A

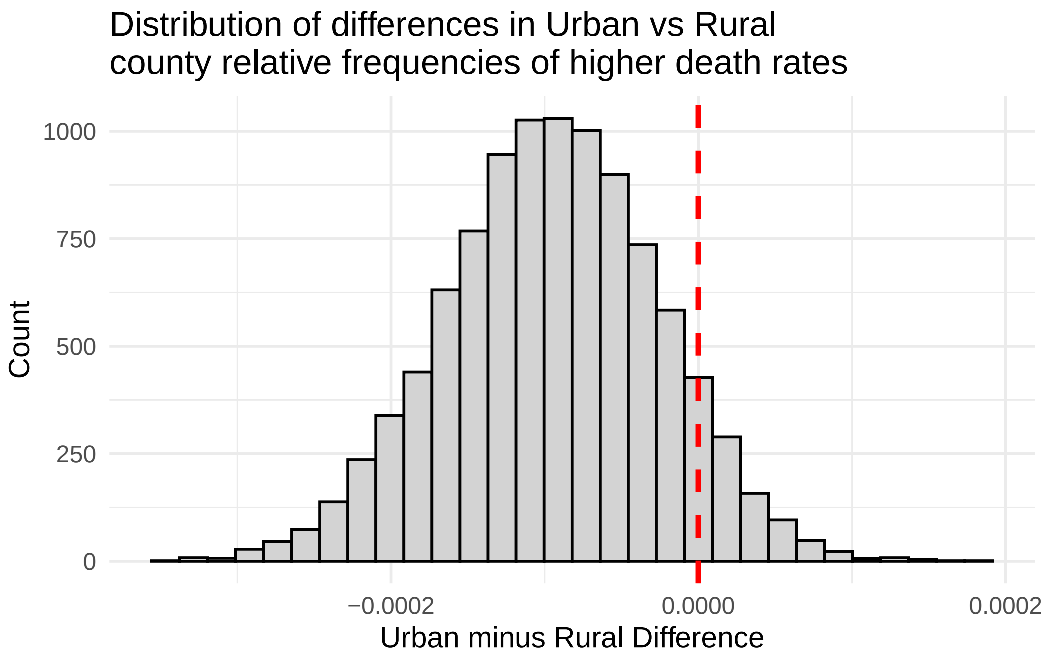
Among 10000 draws from estimated posterior distributions, there were 80 cases where the relative frequency of counties with higher than average mortality rate in the Urban class was higher than the relative frequency of counties in the Rural class. This corresponds to the 0.8% chance that the true rate of counties with higher than average mortality in the urban class is above that or counties in the rural class. At the $\alpha = 0.05$ statistical significance level, we can conclude that the the rate of deaths in urban counties must be consistently lower



4 - B

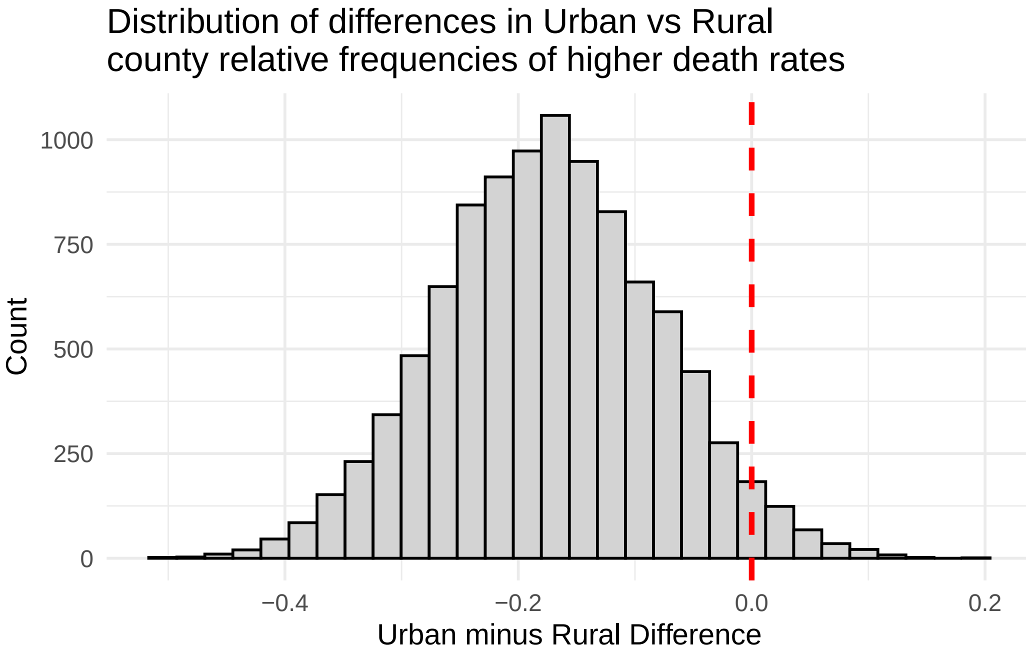
Among 10000 draws from estimated posterior distributions, there were 842 cases where the death rate per 1,000 people in the urban counties was higher than the death rate per 1,000 in the rural counties. This corresponds to the 8.42% chance that the true rate of deaths per 1,000 is higher in the urban counties is higher than that of rural counties.

While this chance is higher than the traditional 5% frequentist cut off rate for the significance level, estimated chance is still fairly slim.


Distribution of differences in Urban vs Rural
county relative frequencies of higher death rates
Count
1000
750
500
250
0
−0.0002
0.0000
0.0002
Urban minus Rural Difference

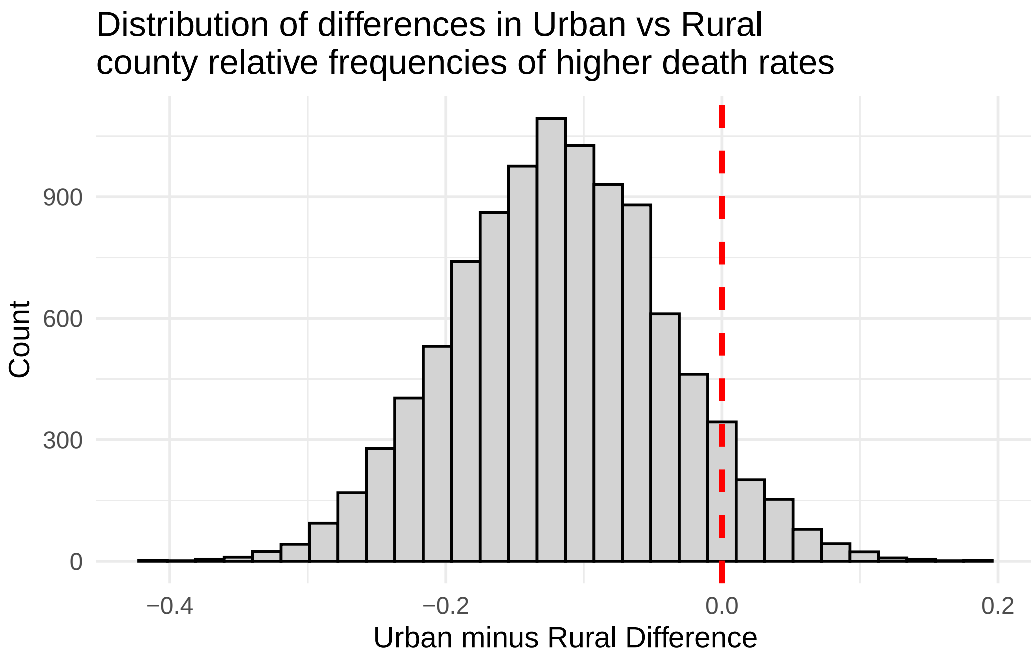
Appendix

Increasing A, B to 12, 12 for 4-A



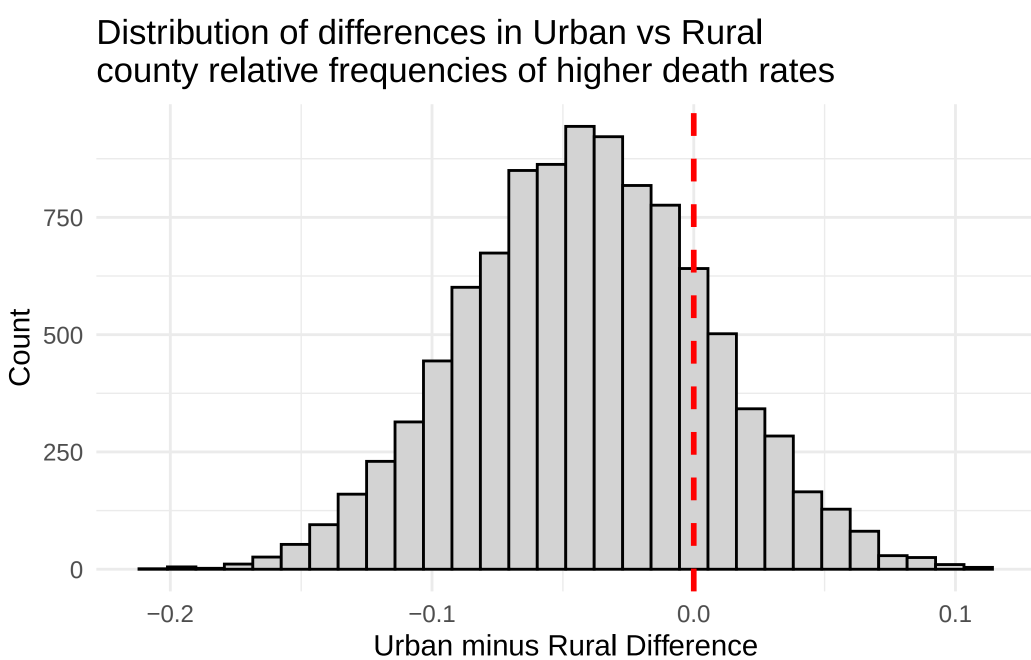
- New chance: 3.37%
- Estimated Difference: -0.17
- Estimated SD of distribution: 0.09

Increasing A, B to 25, 25 for 4-A



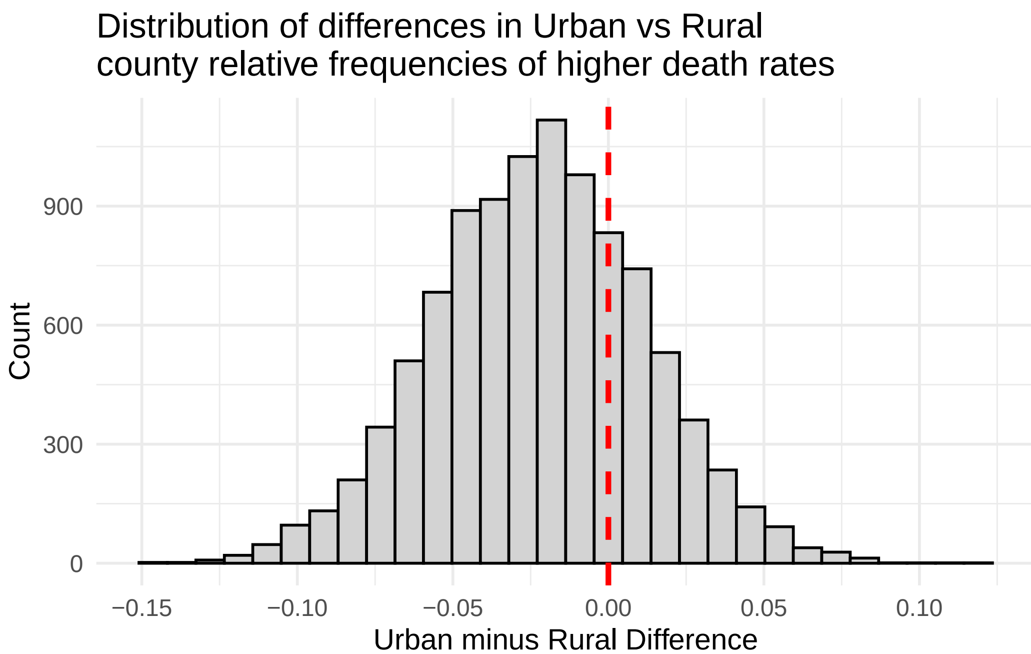
- New chance: 6.7%
- Estimated Difference: -0.12
- Estimated SD of distribution: 0.08

Increasing A, B to 100, 100 for 4-A



- New chance: 18.72%
- Estimated Difference: -0.04
- Estimated SD of distribution: 0.05

Increasing A, B to 200, 200 for 4-A



- New chance: 25.93%
- Estimated Difference: -0.02
- Estimated SD of distribution: 0.03