

Lab 4: Do a teacher's expectations influence student achievement?

Lei Qian and Rebecca C. Steorts

August 24, 2020

Agenda

Do a teacher's expectations influence student achievement? In a famous study, Rosenthal and Jacobson (1968) performed an experiment in a California elementary school to try to answer this question. At the beginning of the year, all students were given an IQ test. For each class, the researchers randomly selected around 20% of the students, and told the teacher that these students were “spurters” that could be expected to perform particularly well that year. (This was not based on the test—the spurters were randomly chosen.) At the end of the year, all students were given another IQ test. The change in IQ score for the first-grade students was:¹

spurters (S)
 $x = (18, 40, 15, 17, 20, 44, 38)$

controls (C)
 $y = (-4, 0, -19, 24, 19, 10, 5, 10, 29, 13, -9, -8, 20, -1, 12, 21, -7, 14, 13, 20, 11, 16, 15, 27, 23, 36, -33, 34, 13, 11, -19, 21, 6, 25, 30, 22, -28, 15, 26, -1, -2, 43, 23, 22, 25, 16, 10, 29)$

1. Plot histograms for the change in IQ score for the two groups. Report your findings.

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v tibble  3.1.6      v dplyr    1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.1      v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

¹The original data is not available. This data is from the `ex1321` dataset of the R package `Sleuth3`, which was constructed to match the summary statistics and conclusions of the original study.

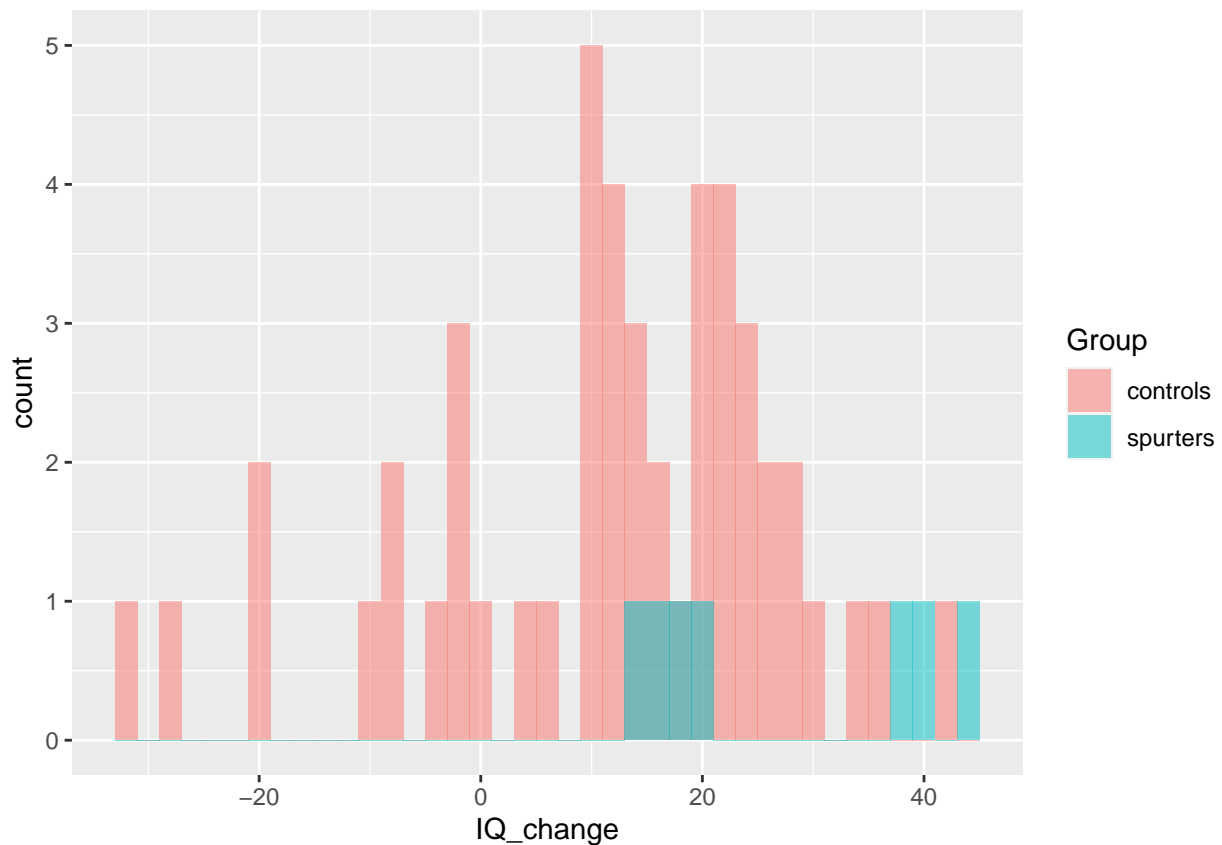
```

# data
x <- c(18, 40, 15, 17, 20, 44, 38)
y <- c(-4, 0, -19, 24, 19, 10, 5, 10, 29, 13, -9, -8, 20, -1, 12, 21, -7, 14,
      13, 20, 11, 16, 15, 27, 23, 36, -33, 34, 13, 11, -19, 21, 6, 25, 30,
      22, -28, 15, 26, -1, -2, 43, 23, 22, 25, 16, 10, 29)

# make data frame with two columns
plotdf <- data.frame(
  IQ_change = c(x, y),
  Group = c(rep('spurters', length(x)), rep('controls', length(y)))
)

# plot histogram of IQ_change for both groups
plotdf %>%
  ggplot() +
  geom_histogram(aes(fill=Group, x=IQ_change),
                 binwidth = 2,
                 position = 'identity',
                 alpha = 0.5) +
  NULL

```



Based on the histogram, it appears that the spurters on average have more positive changes (greater improvement) in their IQ scores than the controls.

2. How strongly does this data support the hypothesis that the teachers' expectations caused the spurters to perform better than their classmates? IQ tests are purposefully calibrated to make the scores normally distributed, so it makes sense to use a normal model here:

$$\begin{aligned}\text{spurters: } X_1, \dots, X_{n_S} &\stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_S, \lambda_S^{-1}) \\ \text{controls: } Y_1, \dots, Y_{n_C} &\stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_C, \lambda_C^{-1}).\end{aligned}$$

We are interested in the difference between the means—in particular, is $\mu_S > \mu_C$? We don't know the standard deviations $\sigma_S = \lambda_S^{-1/2}$ and $\sigma_C = \lambda_C^{-1/2}$, and the sample seems too small to estimate them very well.

it is easy using a Bayesian approach: we just need to compute the posterior probability that $\mu_S > \mu_C$:

$$\mathbb{P}(\mu_S > \mu_C \mid x_{1:n_S}, y_{1:n_C}).$$

Let's use independent NormalGamma priors:

$$\begin{aligned}\text{spurters: } (\mu_S, \lambda_S) &\sim \text{NormalGamma}(m, c, a, b) \\ \text{controls: } (\mu_C, \lambda_C) &\sim \text{NormalGamma}(m, c, a, b)\end{aligned}$$

with the following hyperparameter settings, based on subjective prior knowledge:

- $m = 0$ (Don't know whether students will improve or not, on average.)
- $c = 1$ (Unsure about how big the mean change will be—prior certainty in our choice of m assessed to be equivalent to one datapoint.)
- $a = 1/2$ (Unsure about how big the standard deviation of the changes will be.)
- $b = 10^2 a$ (Standard deviation of the changes expected to be around 10 = $\sqrt{b/a} = \mathbb{E}(\lambda)^{-1/2}$.)

The updated posterior parameters are

for spurters:

$$\begin{aligned}M &= \frac{1 \cdot 0 + 7 \cdot 27.43}{1 + 7} = 24.0 \\ C &= 1 + 7 = 8 \\ A &= 1/2 + 7/2 = 4 \\ B &= 100/2 + \frac{1}{2} \cdot 7 \cdot 11.66^2 + \frac{1}{2} \frac{1 \cdot 7}{1 + 7} (27.43 - 0)^2 = 855.0\end{aligned}$$

for controls:

$$\begin{aligned}M &= \frac{1 \cdot 0 + 48 \cdot 12.04}{1 + 48} = 11.8 \\ C &= 1 + 48 = 49 \\ A &= 1/2 + 48/2 = 24.5 \\ B &= 100/2 + \frac{1}{2} \cdot 48 \cdot 16.10^2 + \frac{1}{2} \frac{1 \cdot 48}{1 + 48} (12.04 - 0)^2 = 6344.0\end{aligned}$$

which implies that

$$\begin{aligned}\mu_S, \lambda_S \mid x_{1:n_S} &\sim \text{NormalGamma}(24.0, 8, 4, 855.0) \\ \mu_C, \lambda_C \mid y_{1:n_C} &\sim \text{NormalGamma}(11.8, 49, 24.5, 6344.0).\end{aligned}$$

3. Based on the calculations from the previous task, provide a scatterplot of samples from the posterior distributions for the two groups. Your scatterplot should look similar to Figure 1. What are your conclusions?

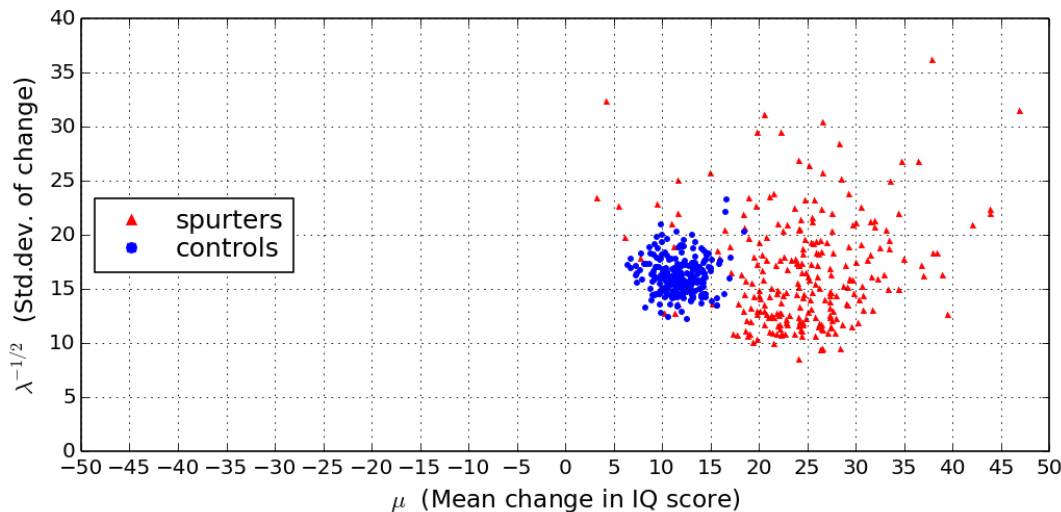


Figure 1: Samples of (μ, σ) from the posteriors for the two groups.

4. Now, we can answer our original question: “What is the posterior probability that $\mu_S > \mu_C$?”

The easiest way to do this is to take a bunch of samples from each of the posteriors, and see what fraction of times we have $\mu_S > \mu_C$. This is an example of a Monte Carlo approximation (much more to come on this in the future).

To do this, we draw $N = 10^6$ samples from each posterior:

$$(\mu_S^{(1)}, \lambda_S^{(1)}), \dots, (\mu_S^{(N)}, \lambda_S^{(N)}) \sim \text{NormalGamma}(24.0, 8, 4, 855.0)$$

$$(\mu_C^{(1)}, \lambda_C^{(1)}), \dots, (\mu_C^{(N)}, \lambda_C^{(N)}) \sim \text{NormalGamma}(11.8, 49, 24.5, 6344.0)$$

and obtain the approximation

$$\mathbb{P}(\mu_S > \mu_C \mid x_{1:n_S}, y_{1:n_C}) \approx \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\mu_S^{(i)} > \mu_C^{(i)}) = ??$$

Interpret the posterior probability that you compute above.

5. Let’s return back to the prior assumptions. There are a few ways that you can check that the prior conforms with our prior beliefs. Let’s go back and checks these. Draw some samples from the prior and look at them—this is probably the best general strategy. See Figure 3. It’s also a good idea to look at sample hypothetical datasets $X_{1:n}$ drawn using these sampled parameter values.

Please replicate a plot similar to Figure 3 and report your findings.

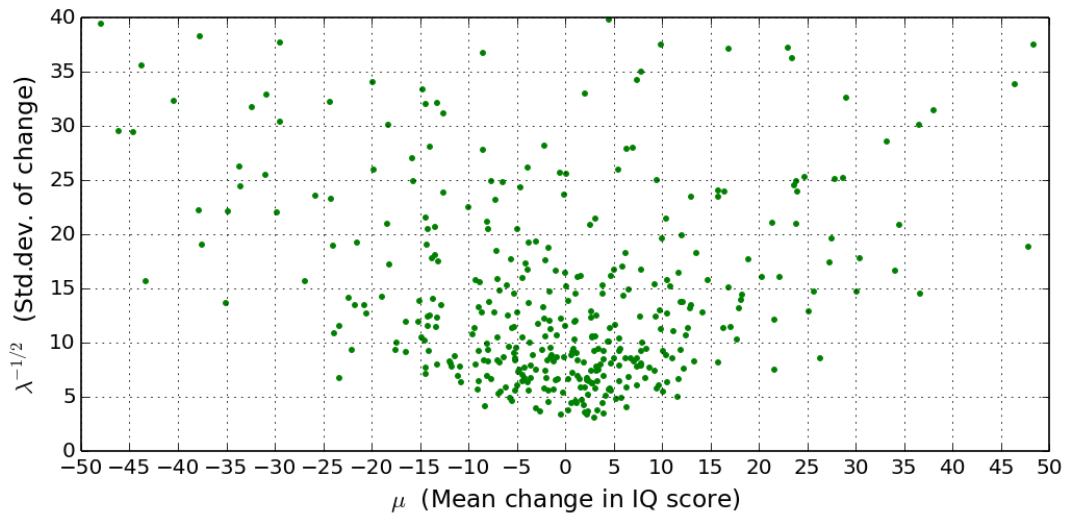


Figure 2: Samples of (μ, σ) from the prior.

Loading Packages

Let's first load packages that we'll need in this assignment and also load the data.

```
set.seed(123)
# input data
# spurters
x = c(18, 40, 15, 17, 20, 44, 38)
# control group
y = c(-4, 0, -19, 24, 19, 10, 5, 10,
      29, 13, -9, -8, 20, -1, 12, 21,
      -7, 14, 13, 20, 11, 16, 15, 27,
      23, 36, -33, 34, 13, 11, -19, 21,
      6, 25, 30, 22, -28, 15, 26, -1, -2,
      43, 23, 22, 25, 16, 10, 29)
# store data in data frame
iqData = data.frame(Treatment = c(rep("Spurters", length(x)),
                                   rep("Controls", length(y))),
                    Gain = c(x, y))
```

Task 1

Plot histograms for the change in IQ score for the two groups. Report your findings.

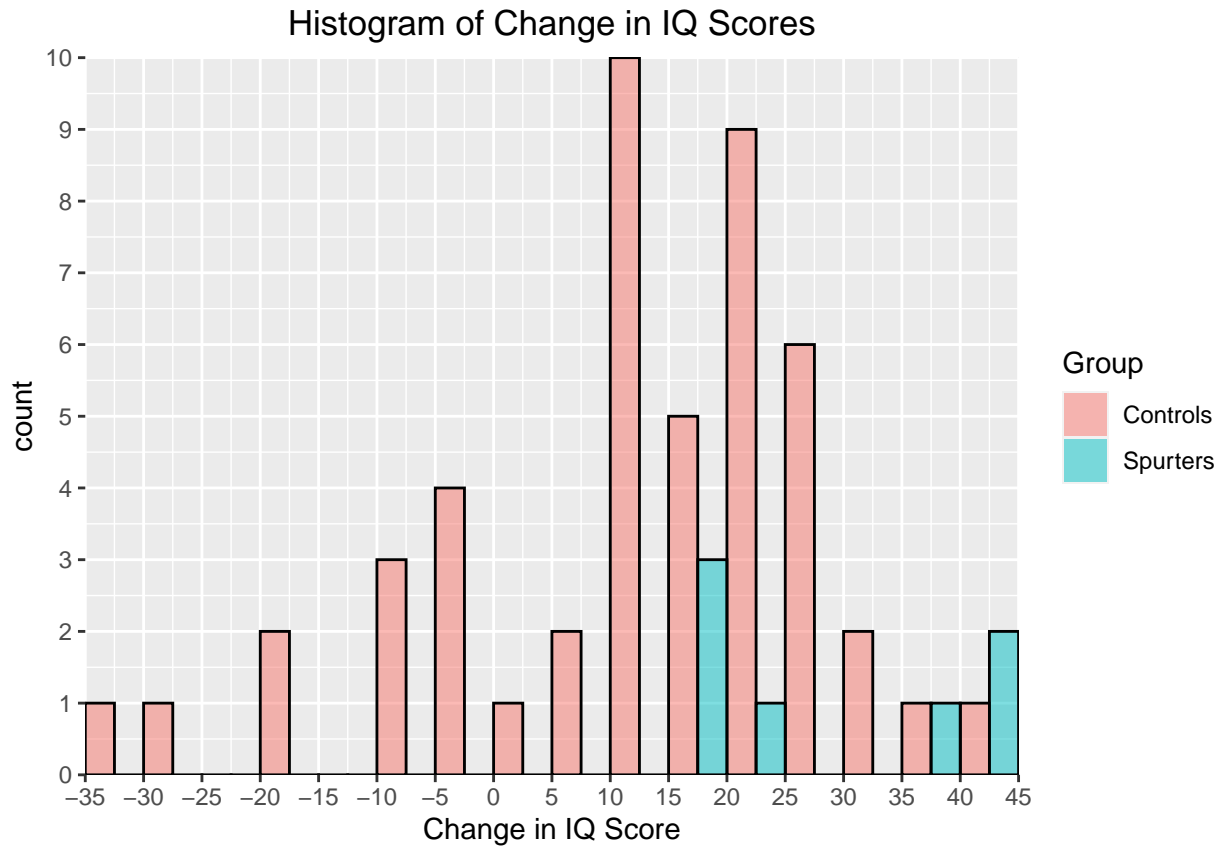
```
xLimits = seq(min(iqData$Gain) - (min(iqData$Gain) %% 5),
              max(iqData$Gain) + (max(iqData$Gain) %% 5),
              by = 5)

ggplot(data = iqData, aes(x = Gain, fill = Treatment, colour = I("black"))) +
```

```

geom_histogram(position = "dodge", alpha = 0.5, breaks = xLimits, closed = "left")+
scale_x_continuous(breaks = xLimits,
                   expand = c(0,0))+
scale_y_continuous(expand = c(0,0),
breaks = seq(0, 10, by = 1))+
ggtitle("Histogram of Change in IQ Scores") + labs(x = "Change in IQ Score",
                                                  fill = "Group") +
theme(plot.title = element_text(hjust = 0.5))

```



From the histograms, I know that the randomly selected “spurters” group has a different distribution than the “controls” group. This could indicate that teachers being told that a specific group of students is expected to perform particularly well will pay more attention and time on that group and resulting in more improvement over the year.

Task 2

How strongly does this data support the hypothesis that the teachers’ expectations caused the spurters to perform better than their classmates? IQ tests are purposefully calibrated to make the scores normally distributed, so it makes sense to use a normal model here:

$$\begin{aligned}
 X_1, \dots, X_{n_s} &| \mu_s, \lambda_s^{-1} \stackrel{iid}{\sim} \text{Normal}(\mu_s, \lambda_s^{-1}) \\
 Y_1, \dots, Y_{n_c} &| \mu_c, \lambda_c^{-1} \stackrel{iid}{\sim} \text{Normal}(\mu_c, \lambda_c^{-1}).
 \end{aligned}$$

We are interested in the difference between the means—in particular, is $\mu_s > \mu_c$? We don’t know the

standard deviations $\sigma_S = \lambda_S^{-1/2}$ and $\sigma_C = \lambda_C^{-1/2}$, and the sample seems too small to estimate them very well.

It is easy using a Bayesian approach. We just need to compute the posterior probability that $\mu_S > \mu_C$:

$$\mathbb{P}(\mu_S > \mu_C \mid x_{1:n_S}, y_{1:n_C}).$$

Let's assume independent Normal-Gamma priors:

$$\text{sputters: } (\mu_S, \lambda_S) \sim \text{NormalGamma}(m, c, a, b)$$

$$\text{controls: } (\mu_C, \lambda_C) \sim \text{NormalGamma}(m, c, a, b)$$

with the following hyperparameter settings, based on subjective prior knowledge:

- $m = 0$ (Don't know whether students will improve or not, on average.)
- $c = 1$ (Unsure about how big the mean change will be—prior certainty in our choice of m assessed to be equivalent to one datapoint.)
- $a = 1/2$ (Unsure about how big the standard deviation of the changes will be.)
- $b = 10^2 a$ (Standard deviation of the changes expected to be around 10 = $\sqrt{b/a} = E(\lambda)^{-1/2}$.)

In order to solve this task, let's recall that the likelihoods are gaussians:

Likelihood Functions

$$p(X_1 \dots X_{n_S} \mid \mu_S, \lambda_S^{-1}) = \prod_{i=1}^{n_S} \frac{1}{\sqrt{2\sigma_S^2\pi}} e^{-\frac{(x_i - \mu_S)^2}{2\sigma_S^2}} = \prod_{i=1}^{n_S} \frac{1}{\sqrt{2\lambda_S^{-1}\pi}} e^{-\frac{\lambda_S(x_i - \mu_S)^2}{2}}$$

$$p(Y_1 \dots Y_{n_C} \mid \mu_C, \lambda_C^{-1}) = \prod_{i=1}^{n_C} \frac{1}{\sqrt{2\sigma_C^2\pi}} e^{-\frac{(y_i - \mu_C)^2}{2\sigma_C^2}} = \prod_{i=1}^{n_C} \frac{1}{\sqrt{2\lambda_C^{-1}\pi}} e^{-\frac{\lambda_C(y_i - \mu_C)^2}{2}}$$

Priors

Note that the priors are Normal-Gammas.

$$p(\mu_S, \lambda_S \mid m, c, a, b) = \frac{b^a \sqrt{c}}{\Gamma(a) \sqrt{2\pi}} \lambda_S^{a-0.5} e^{-b\lambda_S} e^{-\frac{c\lambda_S(\mu_S - m)^2}{2}}$$

$$p(\mu_C, \lambda_C \mid m, c, a, b) = \frac{b^a \sqrt{c}}{\Gamma(a) \sqrt{2\pi}} \lambda_C^{a-0.5} e^{-b\lambda_C} e^{-\frac{c\lambda_C(\mu_C - m)^2}{2}}$$

Posterior Distribution

Recall that from the course notes, that the posterior distribution of the likelihood and prior is an updated Normal Gamma, which has the following form:

$$(\mu_S, \lambda_S) \mid x_{1:n_S} \sim \text{NormalGamma} \left(m' = \frac{cm + n_S \bar{x}}{c + n_S}, c' = c + n_S, a' = a + \frac{n_S}{2}, b' = b + \frac{1}{2} \sum_{i=1}^{n_S} (x_i - \bar{x})^2 + \frac{n_S c}{c + n_S} \frac{(\bar{x} - m)^2}{2} \right)$$

$$= \text{NormalGamma}(24, 8, 4, 855)$$

$$(\mu_C, \lambda_C) \mid y_{1:n_C} \sim \text{NormalGamma} \left(m^* = \frac{cm + n_C \bar{y}}{c + n_C}, c^* = c + n_C, a^* = a + \frac{n_C}{2}, b^* = b + \frac{1}{2} \sum_{i=1}^{n_C} (y_i - \bar{y})^2 + \frac{n_C c}{c + n_C} \frac{(\bar{y} - m)^2}{2} \right)$$

$$= \text{NormalGamma}(11.8, 49, 24.5, 6344)$$

Corresponding Code

```
prior = data.frame(m = 0, c = 1, a = 0.5, b = 50)
findParam = function(prior, data){
  postParam = NULL
  c = prior$c
  m = prior$m
  a = prior$a
  b = prior$b
  n = length(data)
  postParam = data.frame(m = (c*m + n*mean(data))/(c + n),
                        c = c + n,
                        a = a + n/2,
                        b = b + 0.5*(sum((data - mean(data))^2)) +
                          (n*c *(mean(data)- m)^2)/(2*(c+n)))
  return(postParam)
}
postS = findParam(prior, x)
postC = findParam(prior, y)
```

% latex table generated in R 4.1.2 by xtable 1.8-4 package % Sun Jan 23 00:25:23 2022

	m	c	a	b
prior	0.00	1.00	0.50	50.00
Spurters Posterior	24.00	8.00	4.00	855.00
Controls Posterior	11.80	49.00	24.50	6343.98

Table 1: Parameters

Task 3

Based on the calculations from the previous task, provide a scatterplot of samples from the posterior distributions for the two groups. What are your conclusions?

```
# sampling from two posteriors

# Number of posterior simulations
sim = 1000

# initialize vectors to store samples
mus = NULL
lambdas = NULL
muc = NULL
lambdac = NULL

# Following formula from the NormalGamma with
# the update paramaters accounted accounted for below

lambdas = rgamma(sim, shape = postS$a, rate = postS$b)
lambdac = rgamma(sim, shape = postC$a, rate = postC$b)
```



```

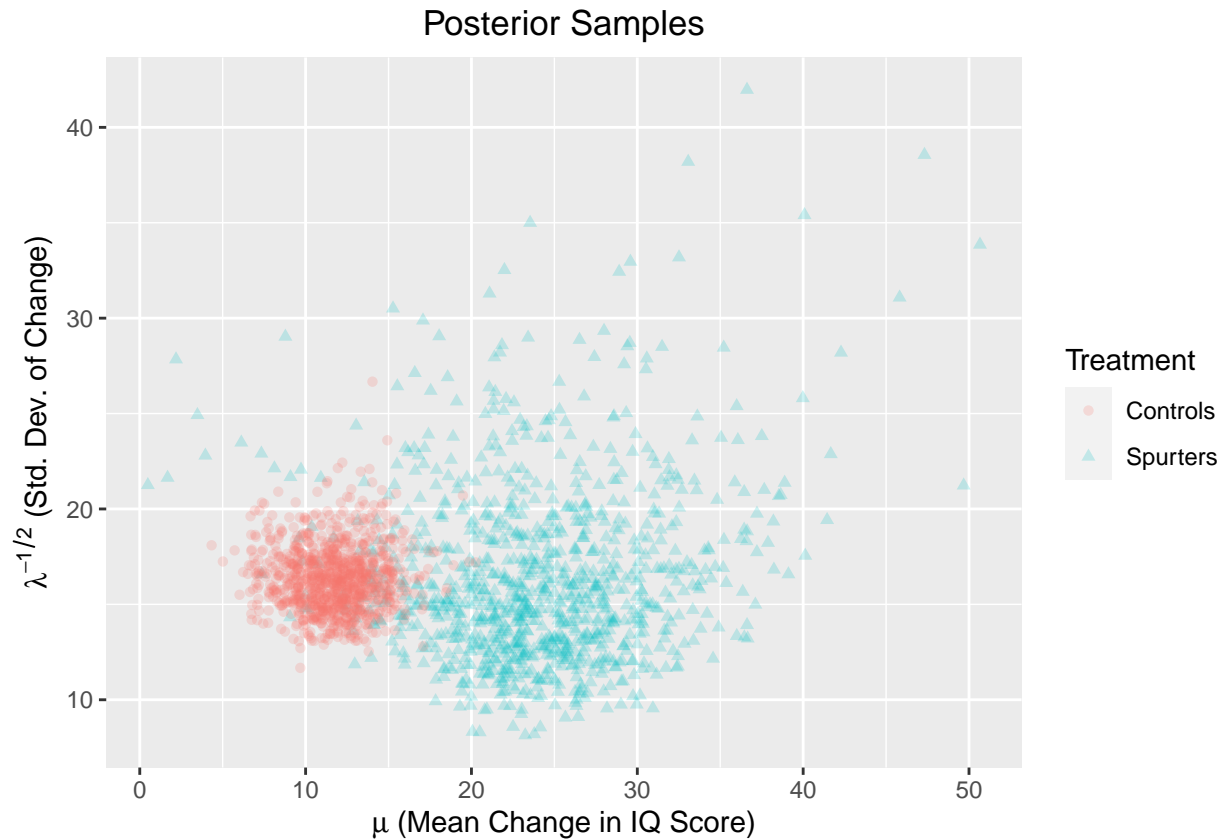
mus = sapply(sqrt(1/(postS$c*lambdaS)),rnorm, n = 1, mean = postS$m)
muc = sapply(sqrt(1/(postC$c*lambdaC)),rnorm, n = 1, mean = postC$m)

# Store simulations
simDF = data.frame(lambda = c(lambdaS, lambdaC),
                    mu = c(mus, muc),
                    Treatment = rep(c("Spurters", "Controls"),
                                   each = sim))

simDF$lambda = simDF$lambda^{-0.5}

# Plot the simulations
ggplot(data = simDF, aes(x = mu, y = lambda, colour = Treatment, shape = Treatment)) +
  geom_point(alpha = 0.2) +
  labs(x = expression(paste(mu, " (Mean Change in IQ Score)")),
       y = expression(paste(lambda^{-1/2}, " (Std. Dev. of Change)"))) +
  ggtitle("Posterior Samples")+
  theme(plot.title = element_text(hjust = 0.5))

```



The simulated scatterplot does look similar to Figure 1 in that the control group is more concentrated with a smaller average mean change in IQ score, while the spurters group has a larger average mean change in IQ score.

Task 4 (Finish for Homework)

Now, we can answer our original question: “What is the posterior probability that $\mu_S > \mu_C$?”

The easiest way to do this is to take a bunch of samples from each of the posteriors, and see what fraction of times we have $\mu_S > \mu_C$. This is an example of a Monte Carlo approximation (much more to come on this in the future).

To do this, we draw $N = 10^6$ samples from each posterior:

$$\begin{aligned}(\mu_S^{(1)}, \lambda_S^{(1)}), \dots, (\mu_S^{(N)}, \lambda_S^{(N)}) &\sim \text{NormalGamma}(24.0, 8, 4, 855.0) \\ (\mu_C^{(1)}, \lambda_C^{(1)}), \dots, (\mu_C^{(N)}, \lambda_C^{(N)}) &\sim \text{NormalGamma}(11.8, 49, 24.5, 6344.0)\end{aligned}$$

and obtain the approximation

$$\mathbb{P}(\mu_S > \mu_C \mid x_{1:n_S}, y_{1:n_C}) \approx \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\mu_S^{(i)} > \mu_C^{(i)}) = ??$$

Interpret the posterior probability that you compute above.

Task 5 (Finish for Homework)

Let’s return back to the prior assumptions. There are a few ways that you can check that the prior conforms with our prior beliefs. Let’s go back and check these. Draw some samples from the prior and look at them—this is probably the best general strategy. See Figure 3. It’s also a good idea to look at sample hypothetical datasets $X_{1:n}$ drawn using these sampled parameter values.

Please replicate a plot similar to Figure 3 and report your findings.

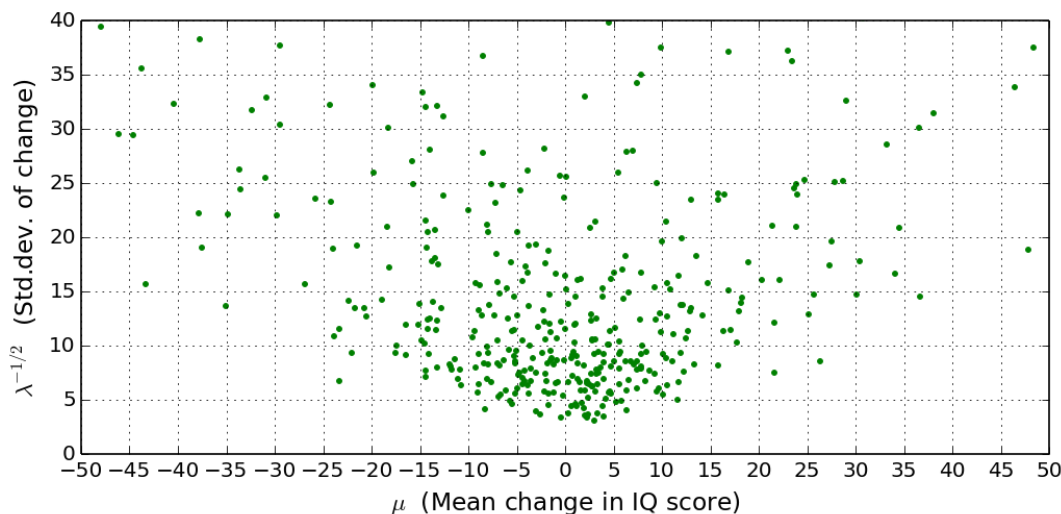


Figure 3: Samples of (μ, σ) from the prior.