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.
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:

$$\text{spurters: } X_1, \ldots, X_{n_S} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_S, \lambda_S^{-1})$$
$$\text{controls: } Y_1, \ldots, Y_{n_C} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_C, \lambda_C^{-1}).$$

   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 $\mu_S > \mu_C$:

$$\mathbb{P}(\boldsymbol{\mu}_S > \boldsymbol{\mu}_C \mid x_{1:n_S}, y_{1:n_C}).$$

Let's use independent NormalGamma priors:

$$\text{spurters: } (\boldsymbol{\mu}_S, \boldsymbol{\lambda}_S) \sim \text{NormalGamma}(m, c, a, b)$$
$$\text{controls: } (\boldsymbol{\mu}_C, \boldsymbol{\lambda}_C) \sim \text{NormalGamma}(m, c, a, b)$$

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

- $m = 0$ (Don't know whether students will improve or not, on average.)

[1] The 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.

- $c = 1$ (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

$$
\begin{aligned}
&\text{for spurters:} \\
&\qquad M = \frac{1 \cdot 0 + 7 \cdot 27.43}{1+7} = 24.0 \\
&\qquad C = 1 + 7 = 8 \\
&\qquad A = 1/2 + 7/2 = 4 \\
&\qquad B = 100/2 + \tfrac{1}{2} \cdot 7 \cdot 11.66^2 + \tfrac{1}{2} \frac{1 \cdot 7}{1+7}(27.43 - 0)^2 = 855.0 \\
&\text{for controls:} \\
&\qquad M = \frac{1 \cdot 0 + 48 \cdot 12.04}{1+48} = 11.8 \\
&\qquad C = 1 + 48 = 49 \\
&\qquad A = 1/2 + 48/2 = 24.5 \\
&\qquad B = 100/2 + \tfrac{1}{2} \cdot 48 \cdot 16.10^2 + \tfrac{1}{2} \frac{1 \cdot 48}{1+48}(12.04 - 0)^2 = 6344.0
\end{aligned}
$$

which implies that

$$
\begin{aligned}
\boldsymbol{\mu}_S, \boldsymbol{\lambda}_S \mid x_{1:n_S} &\sim \text{NormalGamma}(24.0, 8, 4, 855.0) \\
\boldsymbol{\mu}_C, \boldsymbol{\lambda}_C \mid y_{1:n_C} &\sim \text{NormalGamma}(11.8, 49, 24.5, 6344.0).
\end{aligned}
$$

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?
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:

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

and obtain the approximation

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

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 2. 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 2 and report your findings.

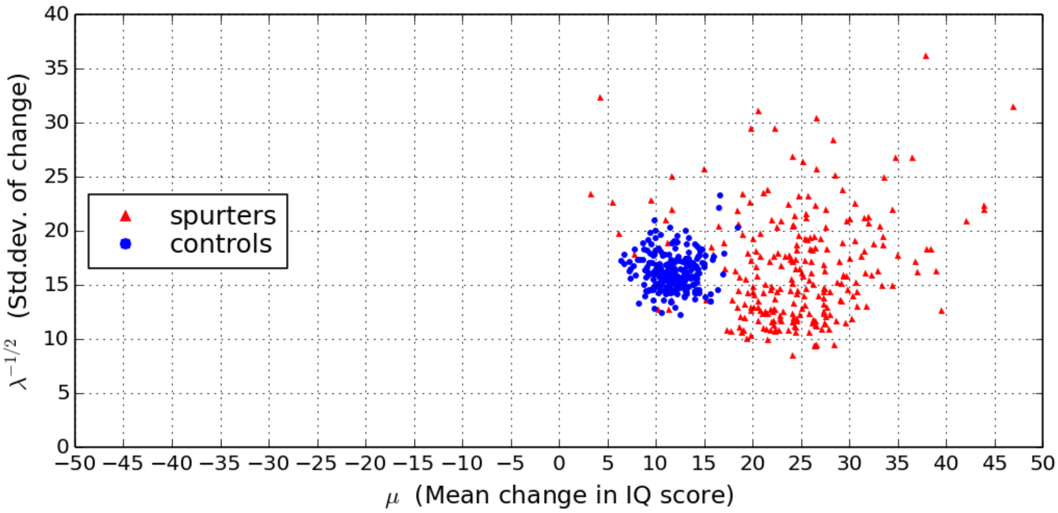


Figure 1: Samples of (μ, σ) from the posteriors for the two groups.

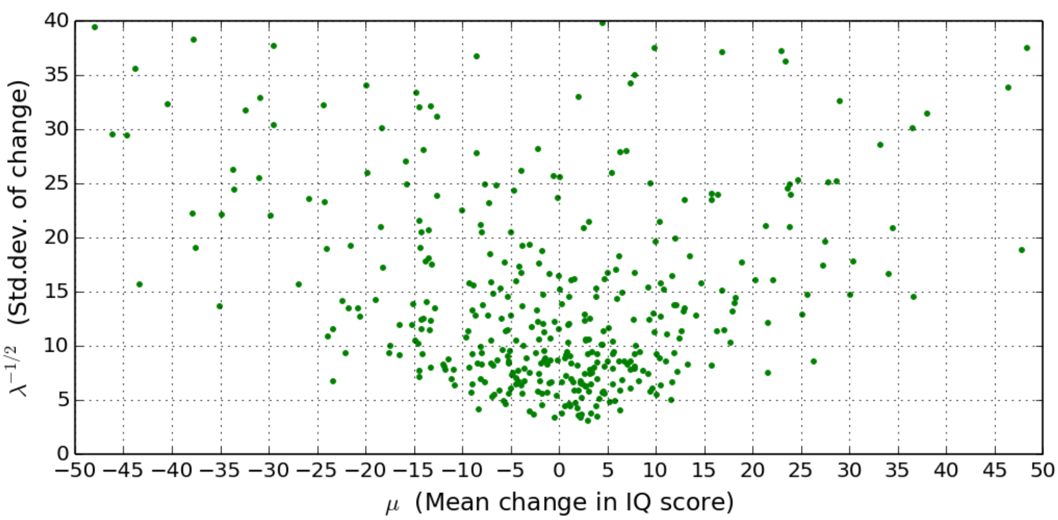


Figure 2: Samples of (μ, σ) from the prior.

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:

$$X_1, \ldots, X_{n_s} \mid \mu_s, \lambda_s^{-1} \overset{iid}{\sim} \text{Normal}(\mu_S, \lambda_S^{-1})$$
$$Y_1, \ldots, Y_{n_C} \mid \mu_c, \lambda_c^{-1} \overset{iid}{\sim} \text{Normal}(\mu_C, \lambda_C^{-1}).$$

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:

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

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

$$p(X_1...X_{n_s}|\mu_s, \lambda_s^{-1}) = \prod_{i=1}^{n_s} \frac{1}{\sqrt{2{\sigma_s}^2\pi}} e^{-\frac{(x_i-\mu_s)^2}{2{\sigma_s}^2}} = \prod_{i=1}^{n_s} \frac{1}{\sqrt{2{\lambda_s}^{-1}\pi}} e^{-\frac{\lambda_s(x_i-\mu_s)^2}{2}}$$

$$p(Y_1...Y_{n_c}|\mu_c, \lambda_c^{-1}) = \prod_{i=1}^{n_c} \frac{1}{\sqrt{2{\sigma_c}^2\pi}} e^{-\frac{(y_i-\mu_c)^2}{2{\sigma_c}^2}} = \prod_{i=1}^{n_c} \frac{1}{\sqrt{2{\lambda_c}^{-1}\pi}} e^{-\frac{\lambda_c(y_i-\mu_c)^2}{2}}$$

Priors

Note that the priors are Normal-Gammas.

$$p(\mu_s, \lambda_s|m, c, a, b) = \frac{b^a\sqrt{c}}{\Gamma(a)\sqrt{2\pi}} \lambda_s^{a-0.5} e^{-b\lambda_s} e^{-\frac{c\lambda_s(\mu_s-m)^2}{2}}$$

$$p(\mu_c, \lambda_c|m, c, a, b) = \frac{b^a\sqrt{c}}{\Gamma(a)\sqrt{2\pi}} \lambda_c^{a-0.5} e^{-b\lambda_c} e^{-\frac{c\lambda_c(\mu_c-m)^2}{2}}$$

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:

$$\begin{aligned}(\mu_s, \lambda_s)|x_{1:n_s} &\sim \text{NormalGamma}\left(m' = \frac{cm + n_s\bar{x}}{c+n_s}, c' = c + n_s, a' = a + \frac{n_s}{2}, b' = b + \frac{1}{2}\sum_{i=1}^{n_s}(x_i - \bar{x})^2 + \frac{n_s c}{c+n_s}\frac{(\bar{x}-m)^2}{2}\right) \\ &= \text{NormalGamma}(24, 8, 4, 855) \\ (\mu_c, \lambda_c)|y_{1:n_c} &\sim \text{NormalGamma}\left(m^* = \frac{cm + n_c\bar{y}}{c+n_c}, c^* = c + n_c, a^* = a + \frac{n_c}{2}, b^* = b + \frac{1}{2}\sum_{i=1}^{n_c}(y_i - \bar{y})^2 + \frac{n_c c}{c+n_c}\frac{(\bar{y}-m)^2}{2}\right) \\ &= \text{NormalGamma}(11.8, 49, 24.5, 6344)\end{aligned}$$

Corresponding Code

```r
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 4.0.2 by xtable 1.8-4 package % Sat Sep 11 00:56:23 2021

	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

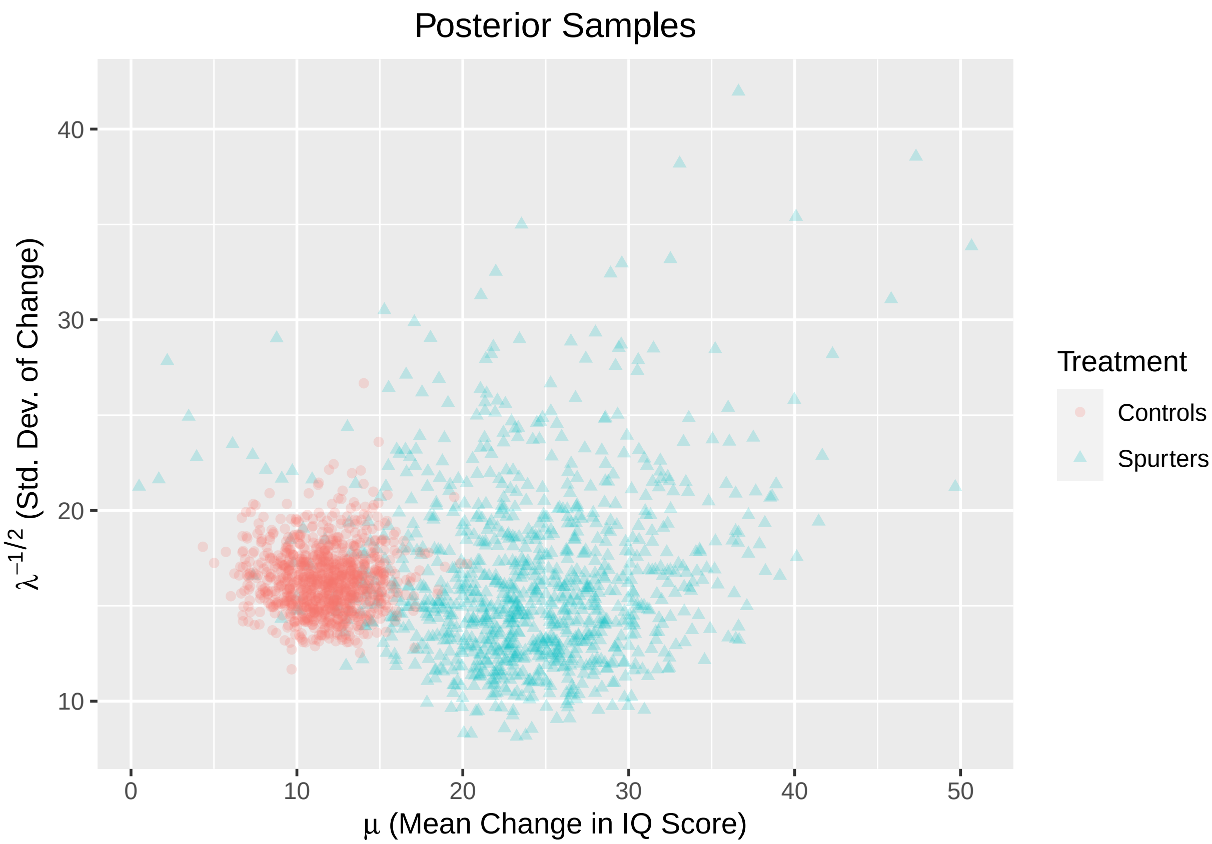
Table 1: Parameters

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)")))  +
```

```r
  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:

```r
set.seed(123)

# sampling from two posteriors

# Number of posterior simulations
sim = 1000000

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

```r
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
sim_means <- data.frame(mus = mus, muc = muc, diff = mus - muc)
sim_means$mus_greater = case_when(sim_means$diff > 0 ~ "yes",
                                  sim_means$diff <= 0 ~ "no")
head(sim_means)
```

```
##        mus      muc      diff mus_greater
## 1 34.08041 11.90184 22.178568         yes
## 2 15.72703 12.42221  3.304816         yes
## 3 16.29141 12.36786  3.923546         yes
## 4 26.23529 15.46697 10.768323         yes
## 5 18.11225  9.90077  8.211480         yes
## 6 18.80054 11.49981  7.300734         yes
```

```r
counts <- sim_means %>%
  count(mus_greater)

c_mus_greater <- counts %>%
  filter(mus_greater == "yes") %>%
  pull(n)

c_mus_greater / sim
```

```
## [1] 0.970673
```

$$\mathbb{P}(\boldsymbol{\mu}_S > \boldsymbol{\mu}_C \mid x_{1:n_S}, y_{1:n_C}) \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left(\mu_S^{(i)} > \mu_C^{(i)}\right) = \quad 0.971$$

The posterior probability that the mean spurter improvement is greater than the mean non-spurter improvement is 0.971. Therefore, it is highly likely that students labeled spurters improve more than students who are not labeled spurters.

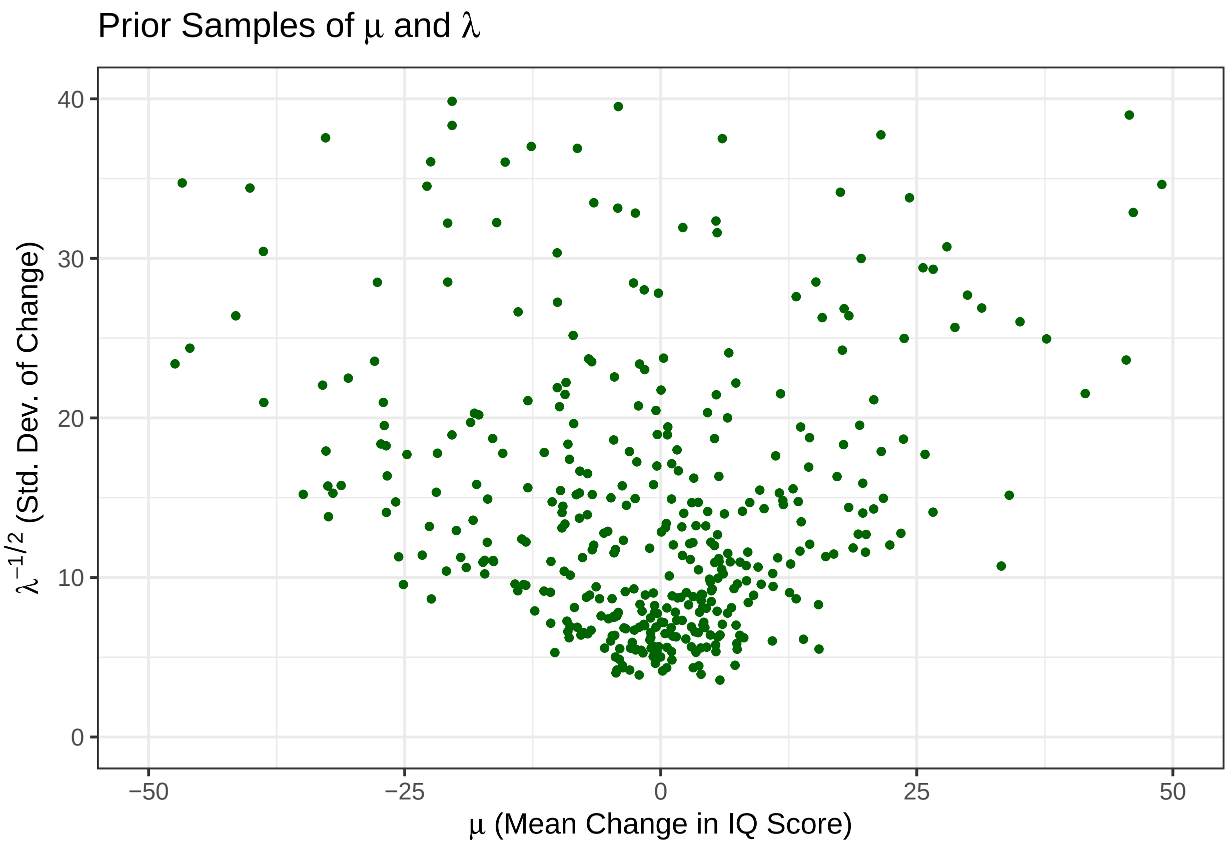
Task 5 (Finish for Homework)

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 2. 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 2 and report your findings.

```r
prior = data.frame(m = 0, c = 1, a = 0.5, b = 50)

# Number of prior simulations
sim2 = 500

# initialize vectors to store samples
mu_prior = NULL
lambda_prior = NULL

# Following formula from the NormalGamma

lambda_prior = rgamma(sim2, shape = prior$a, rate = prior$b)

mu_prior = sapply(sqrt(1/(prior$c*lambda_prior)),rnorm, n = 1, mean = prior$m)

# Store simulations
prior_sim = data.frame(lambda = lambda_prior, mu = mu_prior)

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

# Plot the simulations
ggplot(data = prior_sim, aes(x = mu, y = lambda)) +
  geom_point(size = 1, color = "dark green") +
  xlim(-50, 50) +
  ylim(0, 40) +
  labs(x = expression(paste(mu, " (Mean Change in IQ Score)")),
       y = expression(paste(lambda^{-1/2}, " (Std. Dev. of Change)")))  +
  ggtitle(expression(paste("Prior Samples of ", mu, " and ", lambda))) +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_bw()
```

```
## Warning: Removed 101 rows containing missing values (geom_point).
```



This graph shows that when the standard deviation of change is close to 0, the observed mean changes tend to be close to 0 as well. When the standard deviation of change is large, the observed mean changes tend to vary more. In this graph, we can also see that the observed standard deviations never reach below a 3. These findings from the graph are consistent with our prior beliefs that we don't know if students will improve or not on average (mean change centered around 0), and we wouldn't expect the standard deviation of change to be very small (very unlikely that the standard deviation is less than 3).