Module 4 Homework

Cody S

Module 4 Peer Review Assignment

Problem 1

A continuous random variable with cumulative distribution function F has the median value m such that $F(m) = 0.5$. That is, a random variable is just as likely to be larger than its median as it is to be smaller. A continuous random variable with density f has the mode value x for which $f(x)$ attains its maximum. For each of the following three random variables, (i) state the density function, (ii) compute the median, mode and mean for the random variable, and (iii) Provide at least one graph for the density function using values of the parameter(s) that you select. Indicate the median, mode, and mean values on your graph. (The purpose of this problem is to see the relative locations of the median, mode, and mean for the different random variables).

a) W which is uniformly distributed over the interval $[a, b]$, for some value $a, b \in \mathbb{R}$.

(i) state the density function

$$f(w) = \frac{1}{b-a} \quad \text{for} \quad a \leq w \leq b$$

(ii) compute the median, mode and mean for the random variable

For this example I'll use $[a, b] = [100, 110]$

$$Mean = E(W) = \frac{a+b}{2} = \frac{100+110}{2} = 105$$

$Median(W)$ is the value t at which $P(W \leq t) = 0.5$ and is the integral of the PDF.

$$\int_a^t \frac{1}{b-a} dx = \Big|_a^t \frac{x-a}{b-a} = \frac{t-a}{b-a}$$

$$\int_{110}^{t} \frac{1}{110-100}dx = \Big|_{100}^{t} \frac{x-100}{110-100} = \frac{t-100}{110-100}$$

$$\frac{t-100}{110-100} = 0.5 \Rightarrow t - 100 = 0.5 \cdot 10 \Rightarrow t = 105$$

$$Median(W) = 105$$

Mode - Since all values [a,b] are equally likely in a uniform random distribution there is no mode that can be calculated, the actual mode in a specific distribution is equally likely to be any value in [a,b] or there may be multiple modes.

(iii) Provide at least one graph for the density function using values of the parameter(s) that you select

```
a <- 100
b <- 110
samples <- 1000000
data <- runif(samples, a, b)
mean_unif <- mean(data)
median_unif <- median(data)
mode_unif <-  density(data)$x[which.max(density(data)$y)]

hist(data,
     breaks = 100,
     main = paste("Uniform distribution of" ,samples, "samples from", a, "to", b),
     xlab = "Value")

abline(v = mean_unif, col = "red", lwd = 2)
text(mean_unif, 10500, labels = "Mean", pos = 4)

abline(v = median_unif, col = "blue", lwd = 2)
text(median_unif, 10500, labels = "Median", pos = 2)

abline(v = mode_unif, col = "orange", lwd = 2)
text(mode_unif, 10500, labels = "Mode", pos = 2)

print(paste("Mean =",mean_unif))
print(paste("Median =",median_unif))
print(paste("Mode =", mode_unif))
```
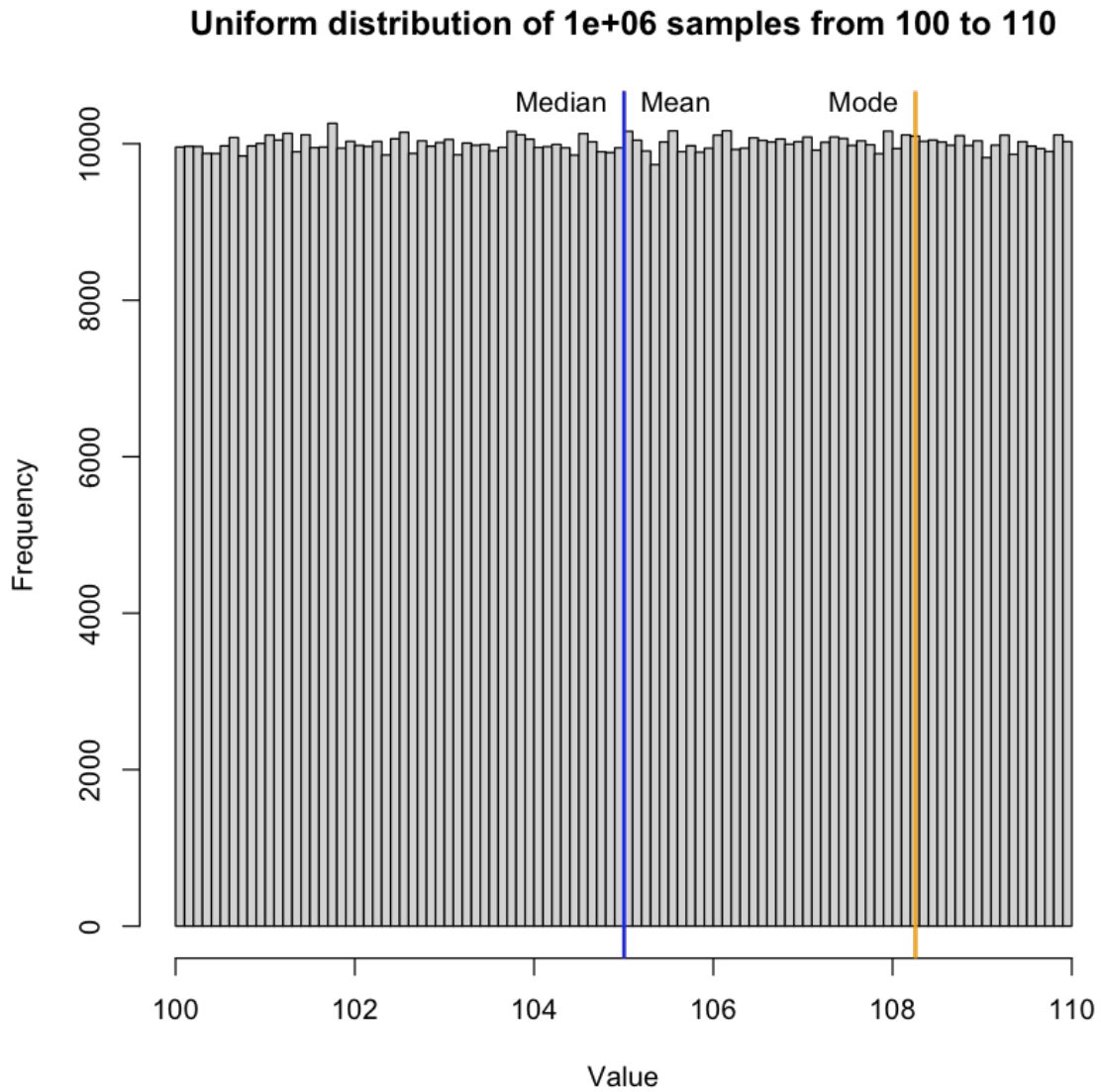
```
[1] "Mean = 105.001950105408"
[1] "Median = 105.003646860132"
[1] "Mode = 108.256314603815"
```

Uniform distribution of 1e+06 samples from 100 to 110

b) X which is normal with parameters μ and σ^2, for some value $\mu, \sigma^2 \in \mathbb{R}$.

(i) state the density function

$$f(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)e^{\frac{(\frac{x-\mu}{\sigma})^2}{2}}$$

(ii) compute the median, mode and mean for the random variable

For this example I'll use $\mu = 3$ and $\sigma^2 = 0.16$

$$Mean(given) = \mu = 3$$

$Median(x)$ is when the CDF of $X = 0.5$

$$\Phi(z) = 0.5 \Rightarrow z = 0$$

$$\Phi(\frac{x-\mu}{\sigma}) = 0.5$$

$$\frac{x-\mu}{\sigma} = 0 \Rightarrow x - \mu = 0 \Rightarrow x = \mu$$

$$Median(x) = 3$$

$$Mode = 3$$

(iii) Provide at least one graph for the density function using values of the parameter(s) that you select

```
mean <- 3
variance <- 0.16
samples <- 1000000
data_normal <- rnorm(samples, mean, sd = sqrt(variance))
mean_norm <- mean(data_normal)
med_norm <- median(data_normal)
mode_norm <- density(data_normal)$x[which.max(density(data_normal)$y)]

hist(data_normal,
     breaks = 100,
     main = "Normal Distribution  mean = 3, SD = 0.4",
     xlab = "Value")

abline(v = mean_norm, col = "red", lwd = 2)
```
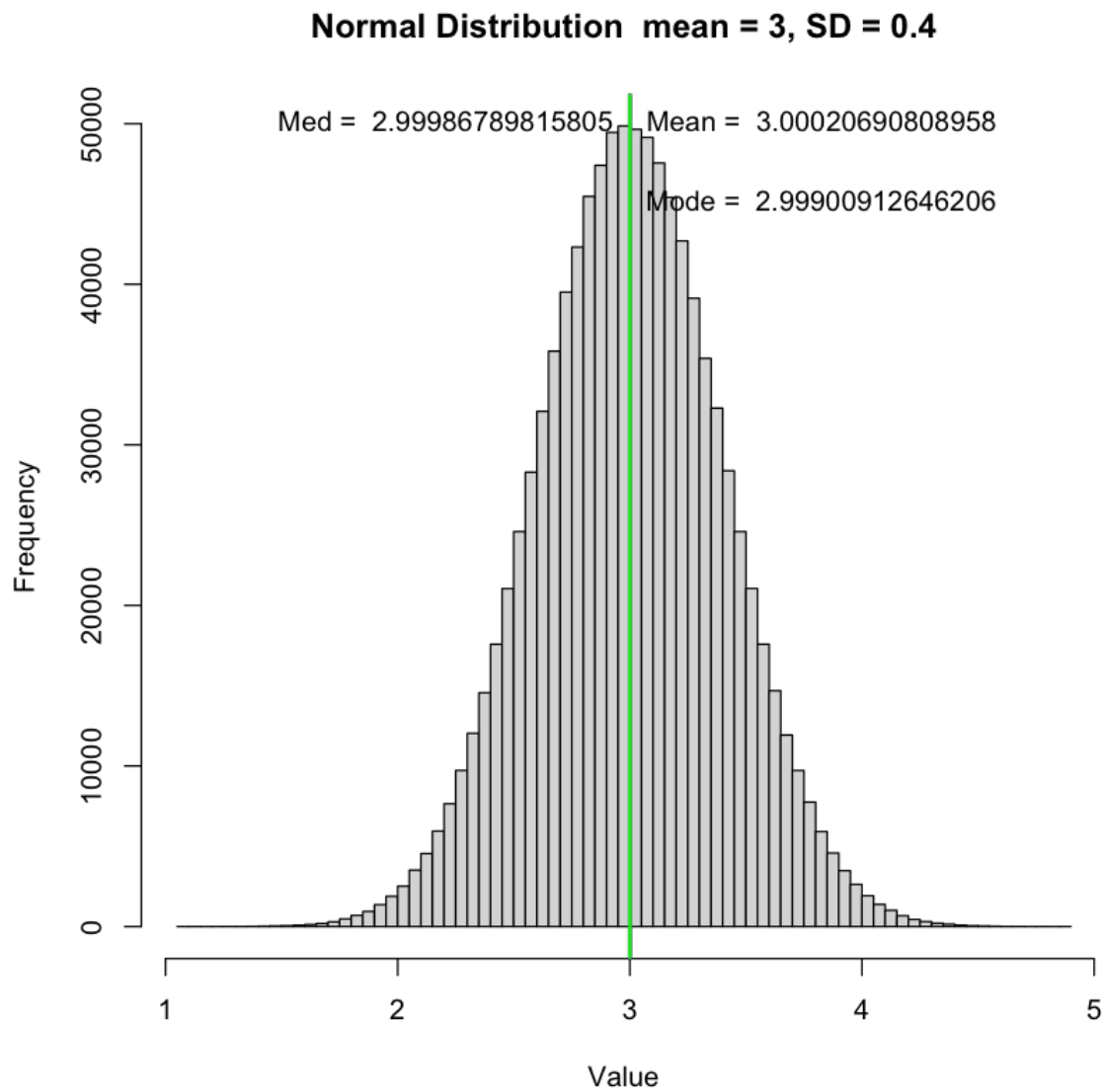
```
text(mean_norm, 50000, labels = paste("Mean = ",mean_norm), pos = 4)

abline(v = med_norm, col = "blue", lwd = 2)
text(med_norm, 50000, labels = paste("Med = ", med_norm), pos = 2)

abline(v = 3, col = "green", lwd = 2)
text(mode_norm, 45000, labels = paste("Mode = ", mode_norm), pos = 4)
```

Normal Distribution mean = 3, SD = 0.4

Med = 2.99986789815805   Mean = 3.00020690808958

Mode = 2.99900912646206

Frequency

Value

c) Y which is exponential with rate $\lambda \in \mathbb{R}$.

(i) state the density function

$$f(y) = \lambda e^{-\lambda y}$$

(ii) compute the median, mode and mean for the random variable

For this example I'll use $\lambda = 2$

$$Mean(Y) = \frac{1}{\lambda} = 0.5$$

$Median(Y)$ is the value at which $F(y) = 0.5$

$$F(y) = \int \frac{1}{\lambda} = 1 - e^{-\lambda y}$$

$$1 - e^{-\lambda y} = 0.5 \Rightarrow -e^{-\lambda y} = -0.5 \Rightarrow -\lambda y = ln(0.5) \Rightarrow y = \frac{ln(0.5)}{-\lambda}$$

$$Median(Y) = \frac{ln(0.5)}{-2} \approx 0.3466$$

(iii) Provide at least one graph for the density function using values of the parameter(s) that you select

```
set.seed(110)

lambda <- 2
samples <- 1000000
data_exp <- rexp(samples, lambda)
mean_exp <- mean(data_exp)
med_exp <- median(data_exp)
mode_exp <-  density(data_exp)$x[which.max(density(data_exp)$y)]

hist(data_exp,
     breaks = 100,
     main = paste("Exponential distribution of" ,samples),
     xlab = "Value",
     xlim = c(0,3))

abline(v = mean_exp, col = "red", lwd = 2)
```

```r
text(mean_exp, 50000, labels = paste("Mean", mean_exp), pos = 4)

abline(v = med_exp, col = "blue", lwd = 2)
text(med_exp, 60000, labels = paste("Median", med_exp), pos = 4)

abline(v = mode_exp, col = "orange", lwd = 2)
text(mode_exp, 85000, labels = paste("Mode", mode_exp), pos = 4)

print(paste("Mean =", mean_exp))
print(paste("Median =", med_exp))
print(paste("Mode =", mode_exp))
```

```
[1] "Mean = 0.500487561428524"
[1] "Median =  0.347011677790655"
[1] "Mode =   0.0505625070439445"
```

Exponential distribution of 1e+06

Mode 0.0505625070439445

Median 0.347011677790655

Mean 0.500487561428524
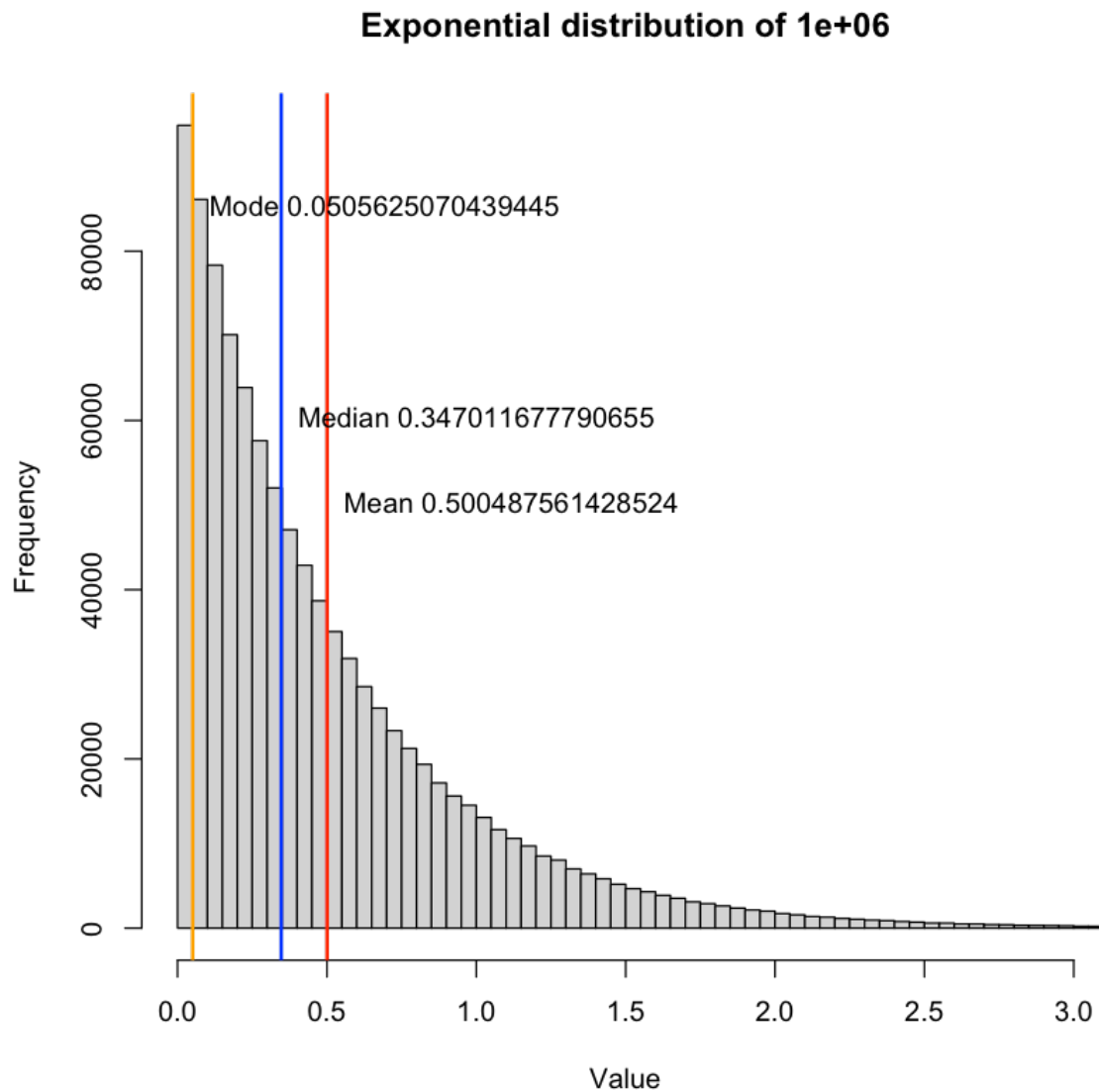
Problem 2

For this problem, we're going to visualize what's happening when we go between different normal distributions.

Part A)

Draw at least 10000 samples from the standard normal distribution $N(0, 1)$ and store the results. Make a density histogram of these samples. Set the x-limits for your plot to $[-10, 10]$ and your y-limits to $[0, 0.5]$ so we can compare with the plots we'll generate in **Parts B-D**.

```r
samples <- 10000
mean <- 0
variance_set <- 1

data_standard_norm <- rnorm(samples, mean, sd = sqrt(variance_set))

hist(data_standard_norm,
     breaks = 100,
     freq = FALSE,
     xlim = c(-10, 10),
     ylim = c(0, 0.5))
```

Histogram of data_standard_norm
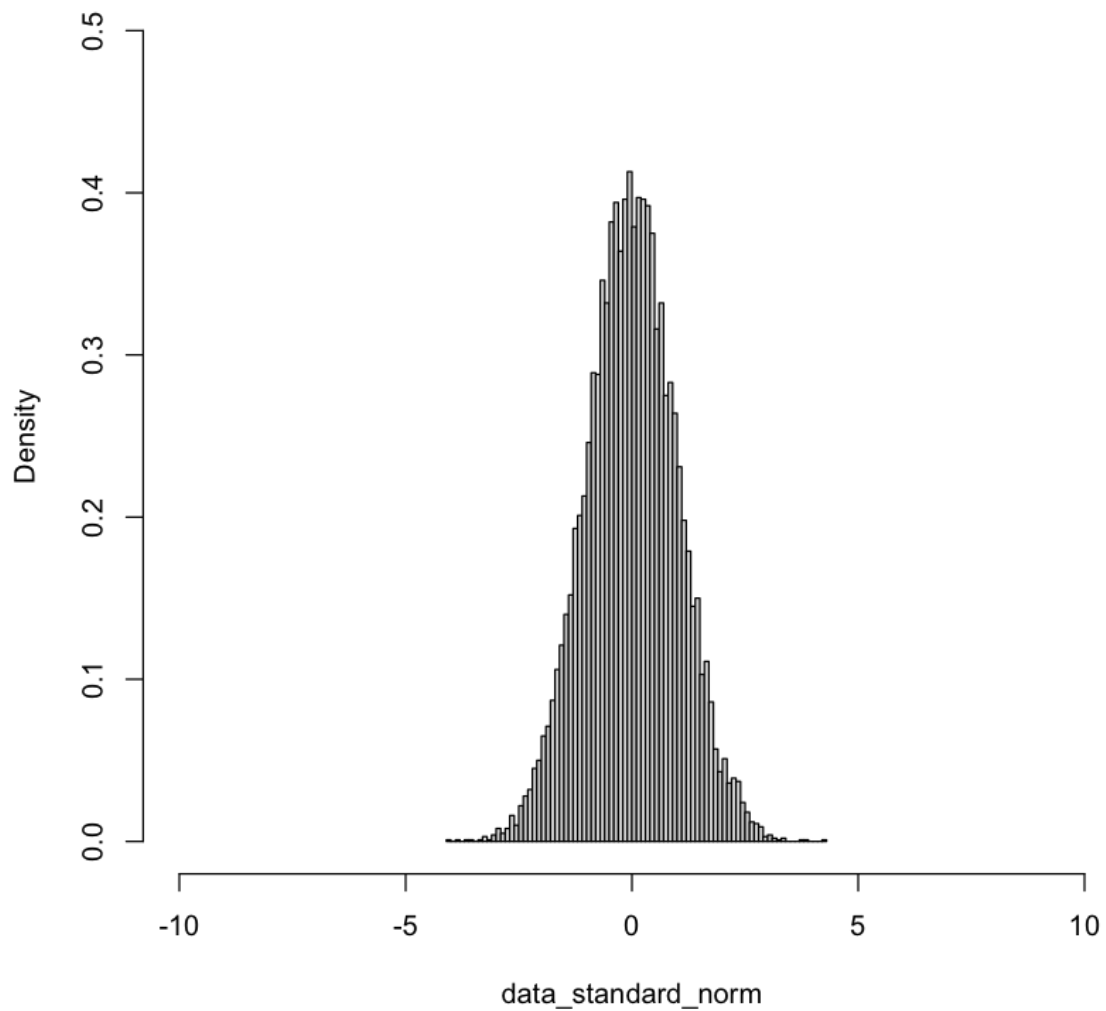
Part b) Now generate 10000 samples from a $N(2, 3)$ distribution and plot a histogram of the results, with the same x-limits and y-limits. Does the histogram make sense based on the changes to parameters?

Note: Be careful with the parameters for `rnorm`. It may help to check the documentation.

```
samples <- 10000
mean <- 2
variance_set <- 3
```

```
data_norm23 <- rnorm(samples, mean, sd = sqrt(variance_set))

hist(data_norm23,
     breaks = 100,
     freq = FALSE,
     xlim = c(-10, 10),
     ylim = c(0, 0.5))
```

Histogram of data_norm23

data_norm23
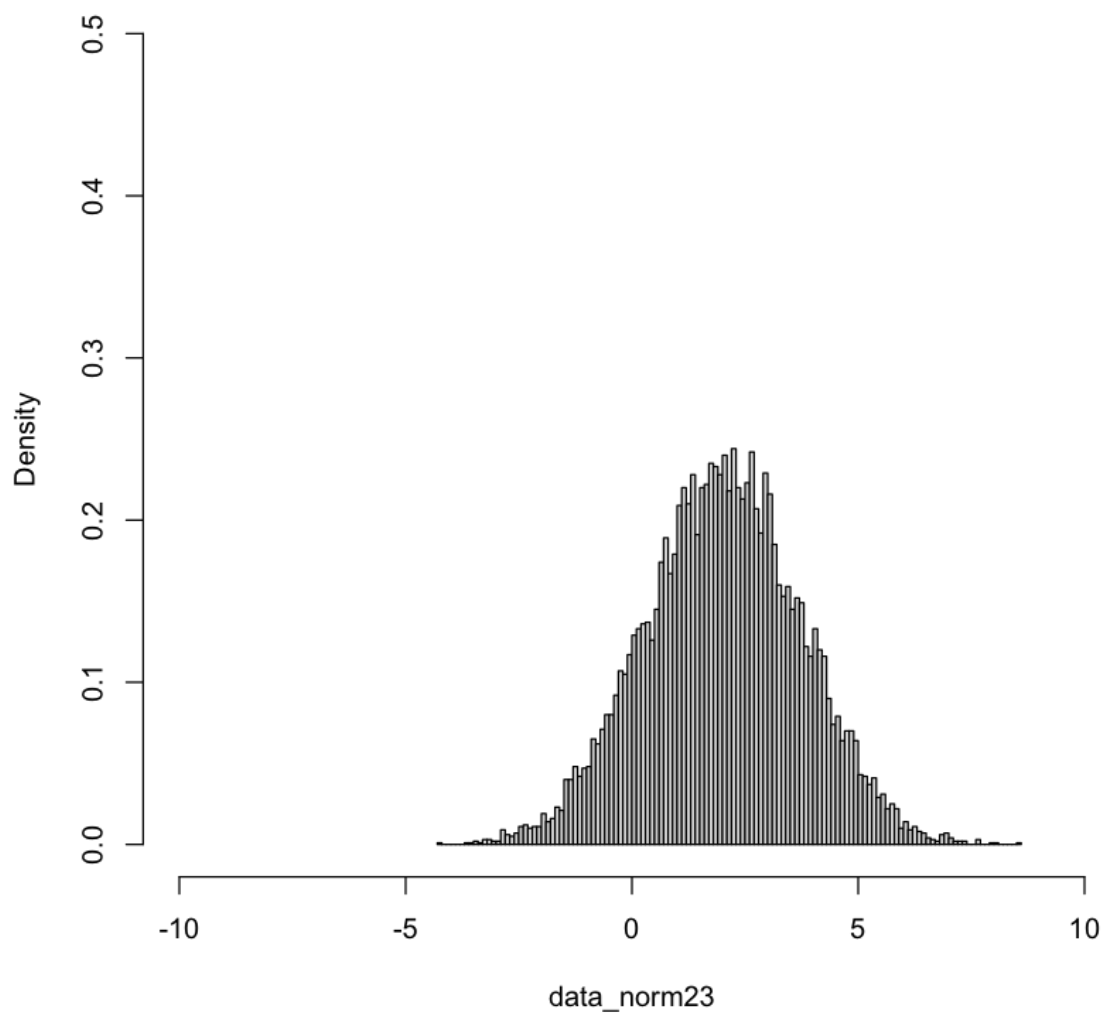
This plot does look correct based on the given parameters for $N(2,3)$. A mean of 2 should shift the entire distribution 2 units to the right to be centered on $x = 2$. Changing the variance to 3 should spread the distribution out by approximately $\sqrt{3}$, or the standard deviation.

Part c)

Suppose we are only able to sample from the standard normal distribution $N(0,1)$. Could we take those samples and perform a simple transformation so that they're samples from $N(2,3)$? Try this, and plot another histogram of the transformed data, again with the same axes. Does your histogram based of the transformed data look like the histogram from **Part B**?

```
# take the standard normal data and
# spread it out by the new sd and shift by the new mean
data_shift <- (data_standard_norm * sqrt(3) + 2)

hist(data_shift,
     breaks = 100,
     freq = FALSE,
     xlim = c(-10, 10),
     ylim = c(0, 0.5))
```

Histogram of data_shift
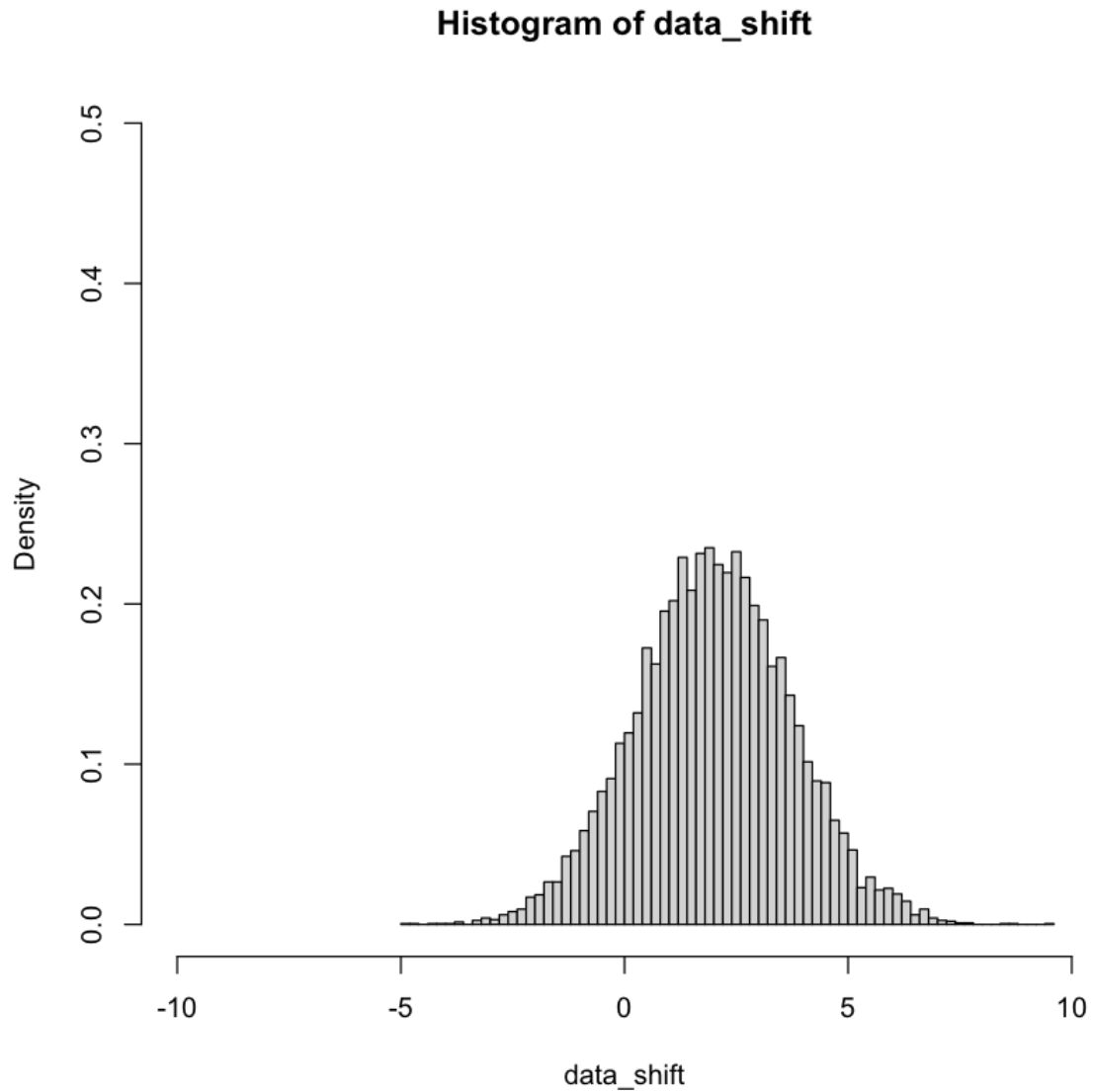
We can create any other normal distribution from the standard normal distribution by spreading the values out by a factor equal to the square root of the variance (the standard deviation, which in this example is $\sqrt{3}$) and then shifting by the new mean μ, which in this example is 2. (Order of operations matters in this process.)

$$N(2,3) = (N(0,1) \cdot \sqrt{3}) + 2$$

Part d)

But can you go back the other way? Take the $N(2, 3)$ samples from **Part B** and transform them into samples from $N(0, 1)$? Try a few transformations and make a density histogram of your transformed data. Does it look like the plot of $N(0, 1)$ data from **Part A**?

```r
renorm_data <- ((data_norm23 - 2) * 1 / sqrt(3))

hist(renorm_data,
     breaks = 100,
     freq = FALSE,
     xlim = c(-10, 10),
     ylim = c(0, 0.5))
```

Histogram of renorm_data
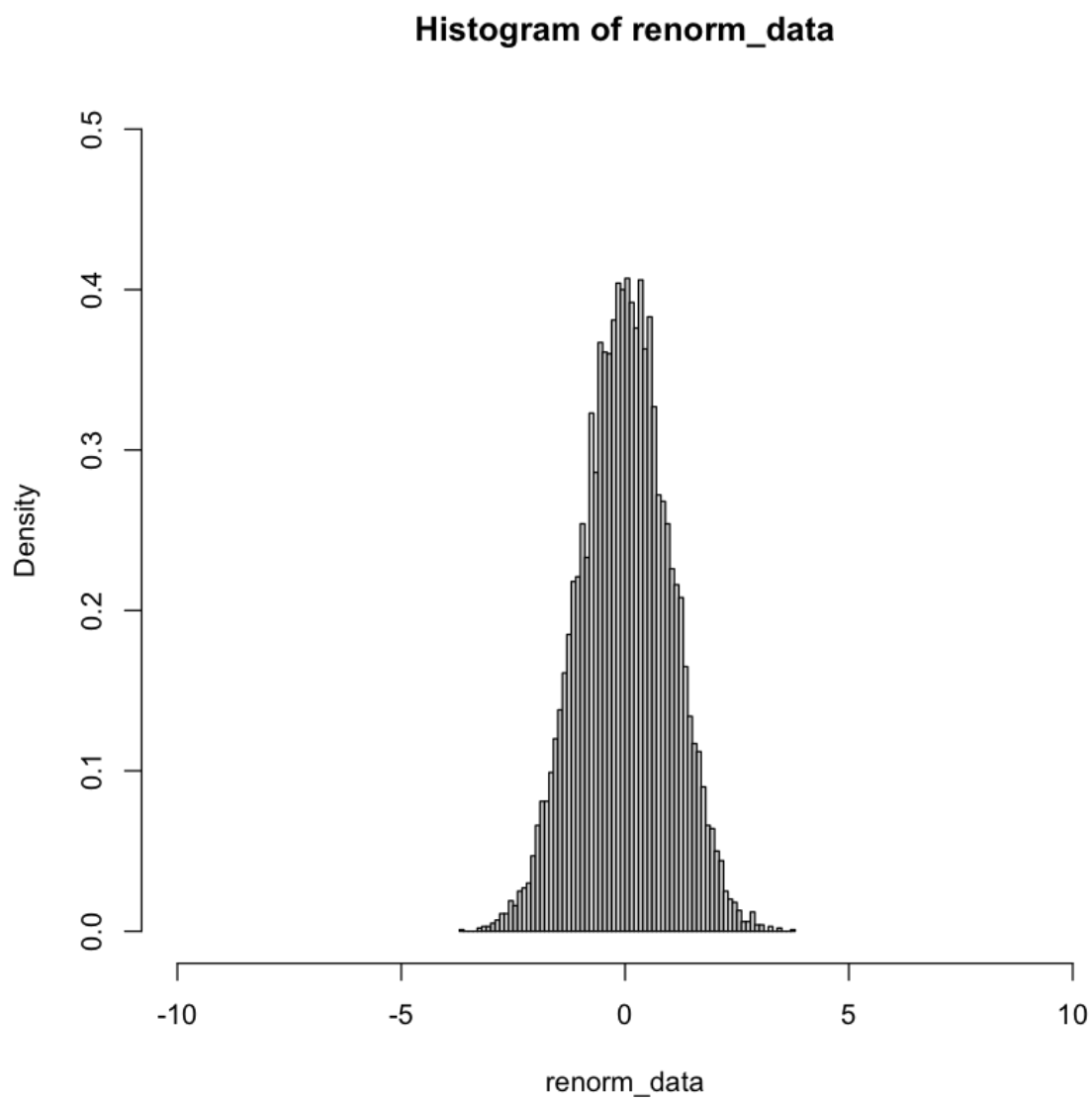
We can reverse the process and convert a non-standard normal distribution to a normal distribution by subtracting the mean to shift the distribution back to be centered on zero, and then by "unspreading" by dividing by the standard deviation.

$$N(0, 1) = (N(2, 3) - 2) \cdot \frac{1}{\sqrt{3}}$$