

Homework 5

Gabriel Lahman

2/22/2019

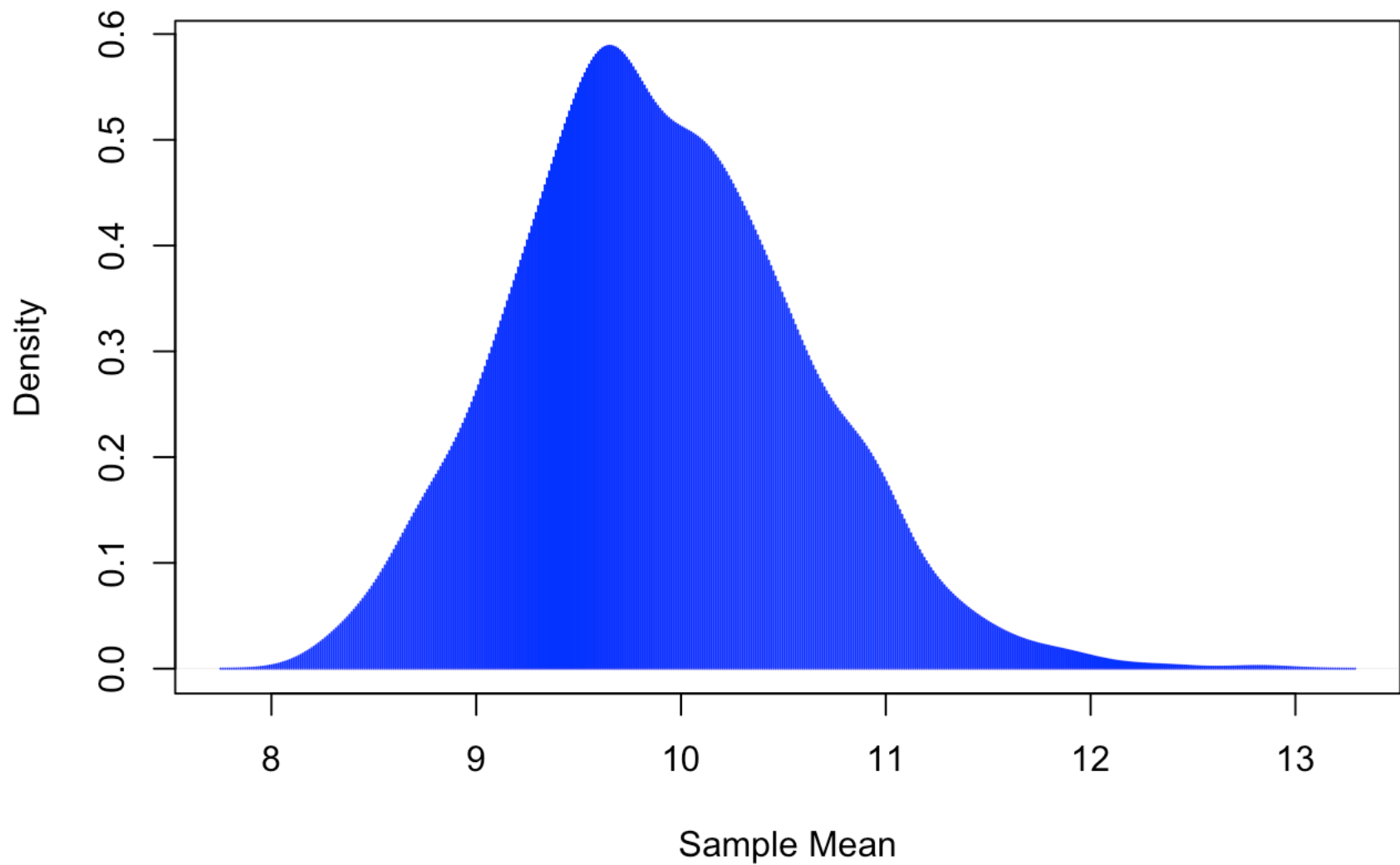
Question 1

```
N = 10000
theta = 10
rate = 1/theta
exp_pop = rexp(N, rate = rate)
```

Part 1

```
num_samples = 1000
sample_size = 200
sample_1 = rep(0, num_samples)
for (i in 1:num_samples) {
  sample_1[i] = mean(sample(exp_pop, sample_size))
}
sample_1_mean_dens = density(sample_1)
plot(sample_1_mean_dens,
      xlab = "Sample Mean",
      main = "Density of Sample Means, Resampling Population",
      type = "h",
      col = "blue")
```

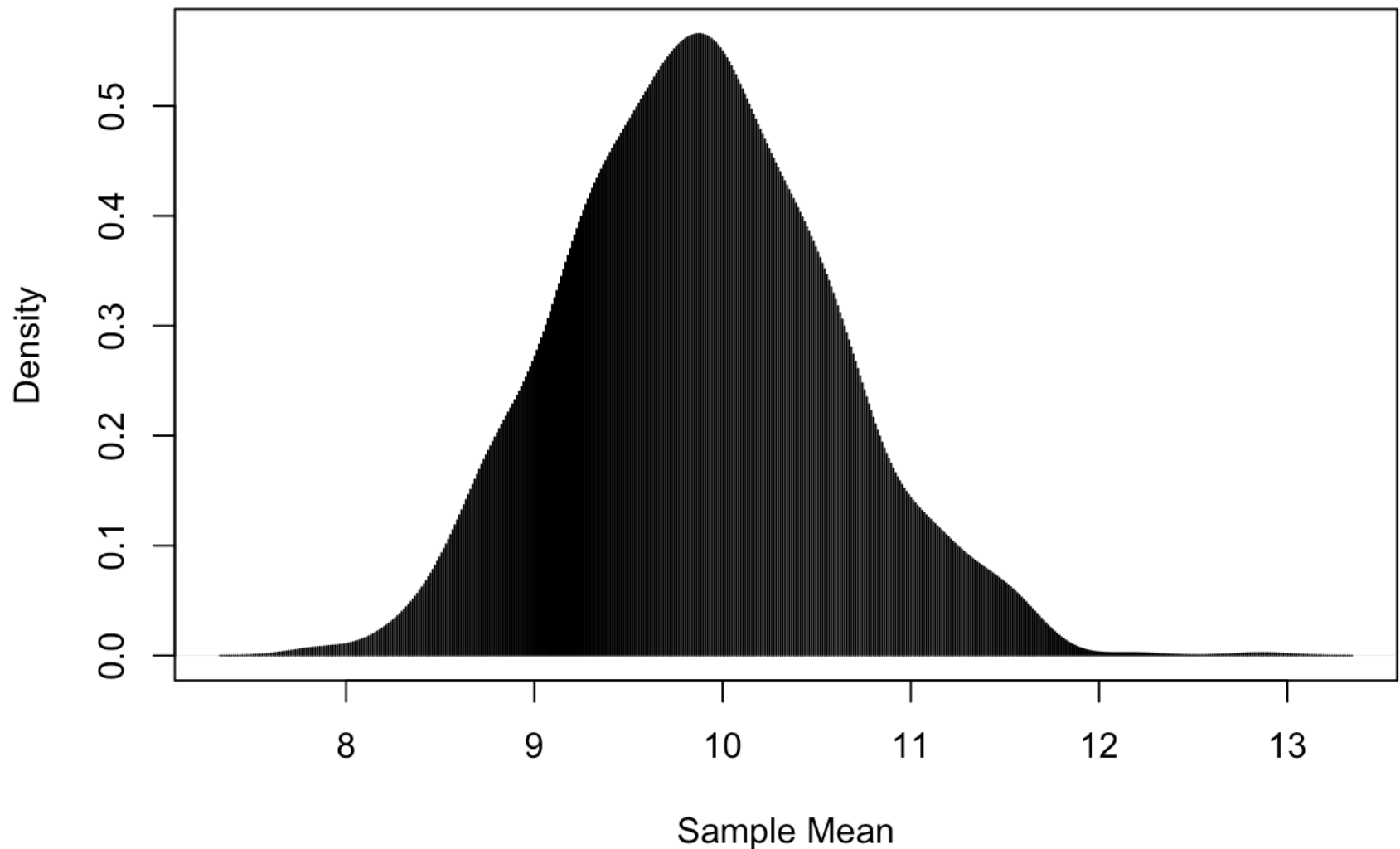
Density of Sample Means, Resampling Population



Part 2

```
solo_sample = sample(exp_pop, sample_size)
sample_2 = rep(0, num_samples)
for (i in 1:num_samples) {
  sample_2[i] = mean(sample(exp_pop, sample_size, replace = TRUE))
}
sample_2_mean_dens = density(sample_2)
plot(sample_2_mean_dens,
      xlab = "Sample Mean",
      main = "Density of Sample Means, Resampling Sample",
      type = "h")
```

Density of Sample Means, Resampling Sample



Part 3

The densities are incredibly similar. Both appear to be relatively normal, which is expected due to the central limit theorem. The most apparent difference between them is that the density of the samples that used resampling of the population, as opposed to the resampling of one initial sample, looks as if it has more variance. This is explainable because it is resampling from the whole population of 10000 values instead of only an initial sample of 2000.

Question 2

```
data("beavers")

y = beaver1$temp

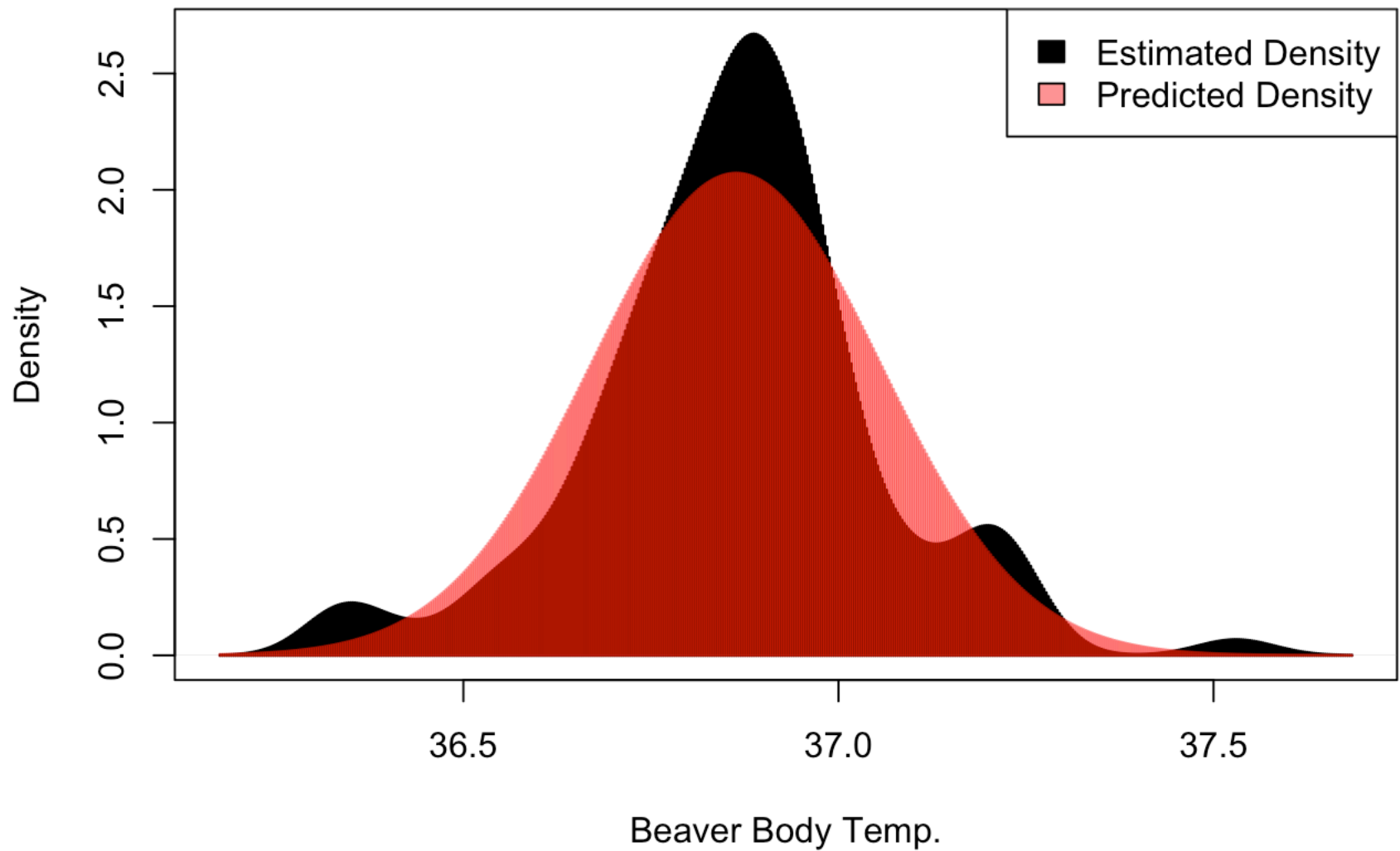
# Fitting is done using the generalized gamma function from HW3, however
# I hid the code as it is quite verbose

MLE = MLE_fun_gen_gamma(y)

density_est = density(y)
density_pred = pred_dens_gen_gamma(density_est$x, MLE)
plot(density_est,
     xlab = "Beaver Body Temp.",
     main = "Beaver Body Temp. Density",
     lwd = 1.5,
     type = "h")
lines(x=density_est$x,
      y = density_pred,
      type="h",
      lwd=1.5,
      col=rgb(1,0,0,alpha = .5))

legend("topright",
      c("Estimated Density", "Predicted Density"),
      fill=c("black",rgb(1,0,0,alpha = .5) ))
```

Beaver Body Temp. Density



```
num_samples = 2000
sample_size = floor(length(y)/3)
samples = rep(0, num_samples)
for (i in 1:num_samples) {
  samples[i] = mean(sample(y, sample_size))
}

mean_of_samples = mean(samples)
alpha = .05

# Naive
quants = quantile(samples, c(alpha/2, 1-alpha/2))
naive_CI = 2*mean_of_samples - quants[2:1]
names(naive_CI) = c("Lower", "Upper")

# Pivot
y_bar = mean(y)
piv_int = (2*y_bar) - naive_CI[2:1]
names(piv_int) = names(naive_CI)
```

```
## Naive 95% Confidence Interval
## Lower      Upper
## 36.81233 36.91392
```

```
## -----
## Pivot Interval
## Lower      Upper
## 36.81046 36.91205
```

The intervals are incredibly close. However, I would be more inclined to use the pivot interval. Since a pivot parameter uses weaker assumptions than the naive interval and is effected by any parameters, it gives us a much stronger conclusion. It is generally better to use less assumptions on random data.

Question 3

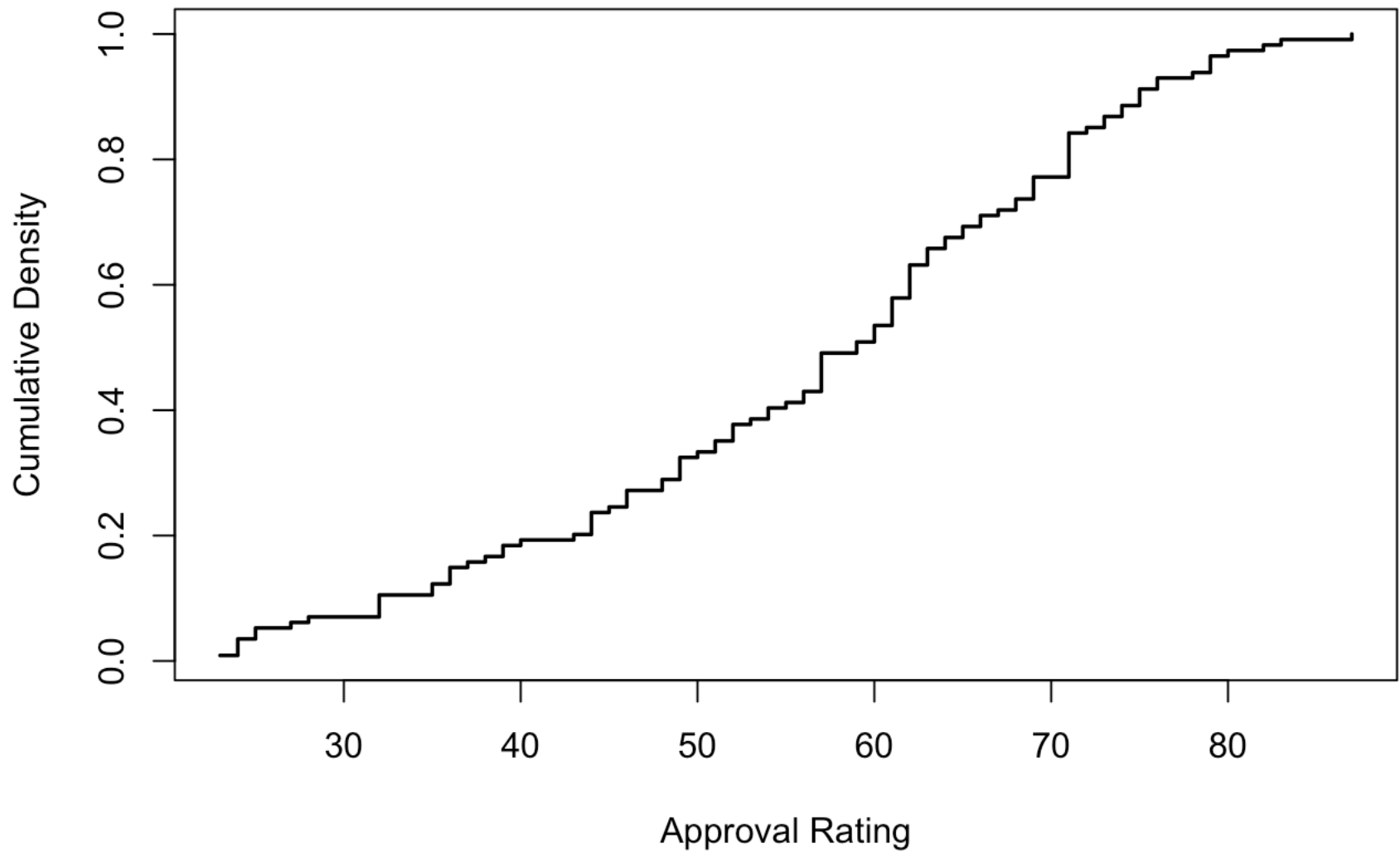
```
data(presidents)
y = as.vector(presidents)
y = y[!is.na(y)]
y = sort(y)
x = unique(y)

n = length(y)
nx = length(x)
counts = rep(0, nx)

val_idx = 1
for (i in 1:n) {
  if (y[i] != x[val_idx]) {
    val_idx = val_idx + 1
  }
  counts[val_idx] = counts[val_idx] + 1
}
emp_dens = cumsum(counts) * (1/n)

plot(x,
      emp_dens,
      type="s",
      lwd=1.75,
      xlab = "Approval Rating",
      ylab = "Cumulative Density",
      main = "Empirical CDF of Presidential Approval Ratings"
)
```

Empirical CDF of Presidential Approval Ratings



What I find very interesting about the CDF is that there does not appear to be any plateau like areas. The line is not too far from being linear, which shows that the approval ratings might be close to a uniform or mostly uniform distribution from a lower bound of 23 to and upper bound of 87.