STATISTICS 641 - ASSIGNMENT 3

DUE DATE: NOON (CDT), MONDAY, SEPTEMBER 27, 2021

Name Jack Rodani

Email Address jrodani@tamu.edu.

Please TYPE your name and email address. Often we have difficulty in reading the handwritten names and email addresses. Make this cover sheet the first page of your Solutions.

STAT 641 Assign #3: H.O. 4 & 5 / Chps 1 & 4 in book.

1.) Let Y have a double exponential distribution, that is Y has a pdf & cdf in the following form w/ parameters $\theta, \beta > 0$.

$$f_y(y) = (1/2\beta)\, e^{-\left(\left|\frac{y-\theta}{\beta}\right|\right)} \quad ; -\infty < y < \infty$$

$$F_y(y) = \begin{cases} \frac{1}{2} e^{-\left(\frac{\theta - y}{\beta}\right)} & ; y < \theta \\ 1 - \frac{1}{2} e^{-(y-\theta/\beta)} & ; y \geq \theta \end{cases}$$

(a) Derive the quantile function for Y.

For $y < \theta$:

$$Q(y) = F^{-1}(y) \Rightarrow y = \frac{1}{2} e^{-\left(\frac{\theta - Q(y)}{\beta}\right)} \Leftrightarrow 2y = e^{-\left(\frac{\theta - Q(y)}{\beta}\right)}$$

$$\Leftrightarrow \ln(2y) = \frac{-\theta + Q(y)}{\beta} \Leftrightarrow Q(y) = \beta \ln(2y) + \theta \ ; y < \theta$$

For $y \geq \theta$:

$$y = 1 - \tfrac{1}{2} e^{-(Q(y) - \theta)/\beta} \Leftrightarrow 2 - 2y = e^{-(Q(y) - \theta)/\beta}$$

$$\Leftrightarrow \beta \ln(2-2y) = -Q(y) + \theta \Leftrightarrow \theta - \beta\ln(2-2y) = Q(y); \ y \geq \theta$$

$$\boxed{Q(y) = \begin{cases} \beta\ln(2y) + \theta & ; y < \theta \\ \theta - \beta\ln(2-2y) & ; y \geq \theta \end{cases}}$$

(b) Derive the survival function for Y. *(Pg 46 H.O. 3)*

$$S(y) = 1 - F(y)$$

$$\boxed{S(y) = \begin{cases} 1 - \frac{1}{2} e^{-(\theta - y)/\beta} & ; y < \theta \\ \frac{1}{2} e^{-(y-\theta)/\beta} & ; y \geq \theta \end{cases}}$$

(c) Derive the Hazard function for Y.

$$h(y) = \frac{f(t)}{S(t)}$$

for $y < \theta$:

$$h(y) = \frac{\frac{1}{2\beta} e^{-\left(\left|y - \theta/\beta\right|\right)}}{1 - \frac{1}{2} e^{-(\theta - y)/\beta}} = \boxed{\frac{e^{-\left(\left|\frac{y-\theta}{\beta}\right|\right)}}{\beta\left(2 - e^{-(\theta - y)/\beta}\right)}}$$

for $y \geq \theta$:

$$h(y) = \frac{(1/2\beta)\, e^{-\left(\left|\frac{y-\theta}{\beta}\right|\right)}}{\frac{1}{2} e^{-(y-\theta)/\beta}} = \boxed{\frac{1}{\beta}}$$

$$\boxed{h(y) = \begin{cases} \dfrac{e^{\left(\frac{y-\theta}{\beta}\right)}}{\beta\left(2 - e^{-(\theta - y)/\beta}\right)} & ; y < \theta \\ 1/\beta & ; y \geq \theta \end{cases}}$$

2.) (Done in R.) Calculate the estimates of the Quartiles of $Q(0.25)$, $Q(0.5)$, $Q(0.75)$ for just the large Litter size.

** In R **: quantile (LLitterSize, na.rm = True).

$\boxed{\begin{aligned} Q(0.25) &= 3.3525 \\ Q(0.50) &= 7.9300 \\ Q(0.75) &= 16.6525 \end{aligned}}$

3) Using the data frame from (2) for just the Large Litter Size, we want to estimate the pdf $f(y)$ for the relative brain weights of the 44 species of mammal. The kernel density estimate is given by:

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1} K\left(\frac{y - Y_i}{h}\right).$$

& we use the gaussian kernel and a bandwidth of $h=3$.

(b) Estimate $f(3)$ & $f(16)$ using the kernel density estimator.

** Computed in R **

$$f(3) = \frac{1}{44(3)} \sum_{i=1}^{44} \frac{1}{\sqrt{2\pi}} e^{-0.5\left(\frac{3 - Y_i}{3}\right)^2} \approx 0.0597$$

$$f(16) = \frac{1}{44(3)} \sum_{i=1}^{44} \frac{1}{\sqrt{2\pi}} e^{-0.5\left(\frac{16 - Y_i}{3}\right)^2} \approx 0.0167$$

(b)* Done in R.*

$f(3) \approx 0.08$; $f(16) \approx 0.02$

(c)-(d): *Done in R*

(c) 35.45 (d) 16

4) * Done in R *

(b) For large: Range: [0.94, 35.45]; Location: [μ = 10.392, median: 7.93]

Shape: distribution is bi-modal w/ the largest occurring at $x=4$ and the other mode occurring around $x=20$. The dist is also skewed right.

For Small: Range [0.42, 20]; Location [μ = 6.886, median = 5]

Shape: The distribution is unimodal & skewed right.

(c) Litter size seems to be positively correlated w/ Brain weights

5.) Select the letter of the best answer for each question.

(1) E (All the functions can be derived given any one of the options).

(2). D (see H.O.4 pg 22)

(3) A

(4) D (H.O.4 pg 27)

(5) B (H.O.4 pg 47)

(6) B (see wiki page for Kernel Density Estimation [Bandwidth Selection]

(7) D

(8) D (H.O.5 pg 14)

(9) C (H.O.5 pg. 22 (bottom))

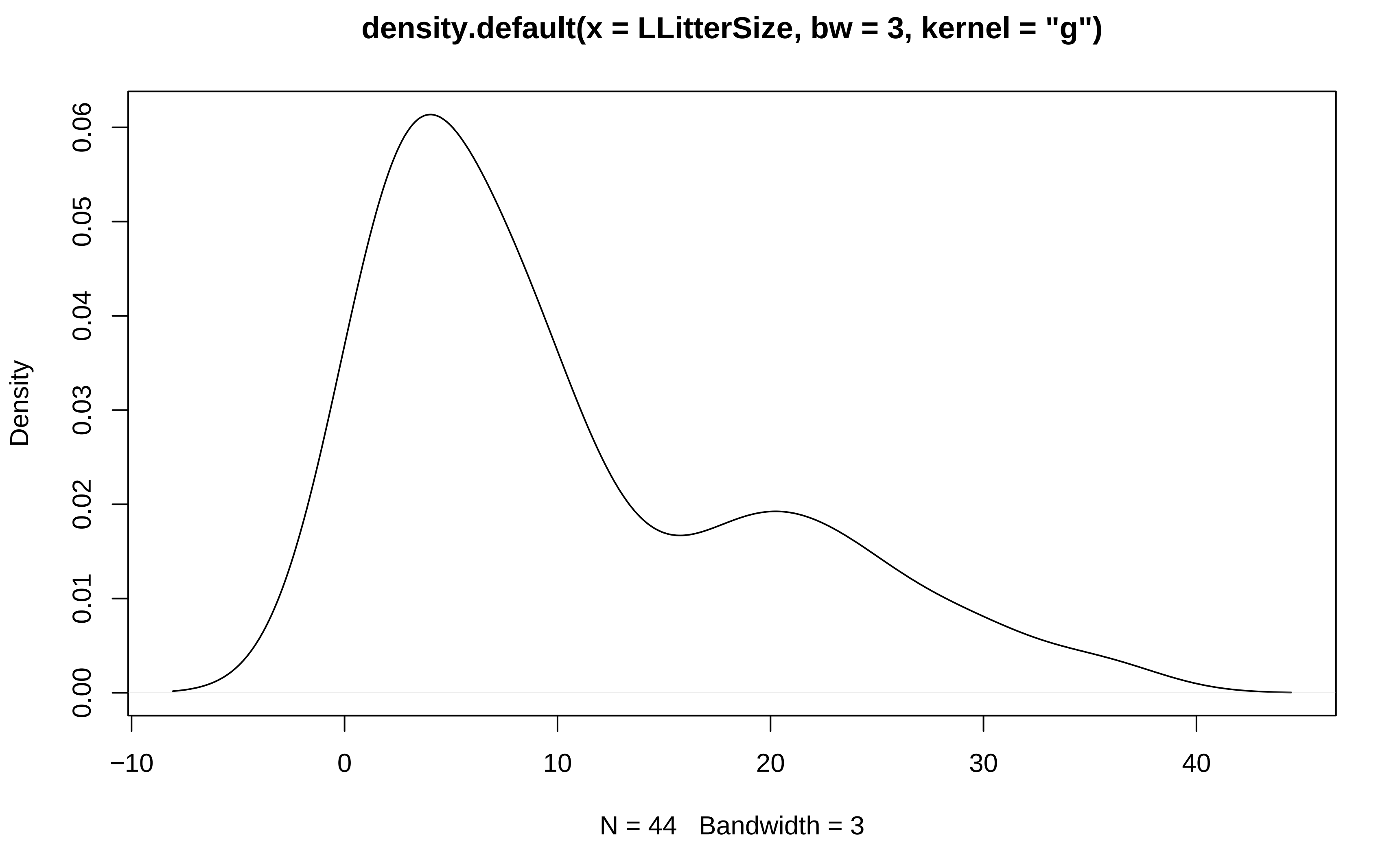
(10) D (H.O.5 pg 27 (top) + Lognormal not symmetric)

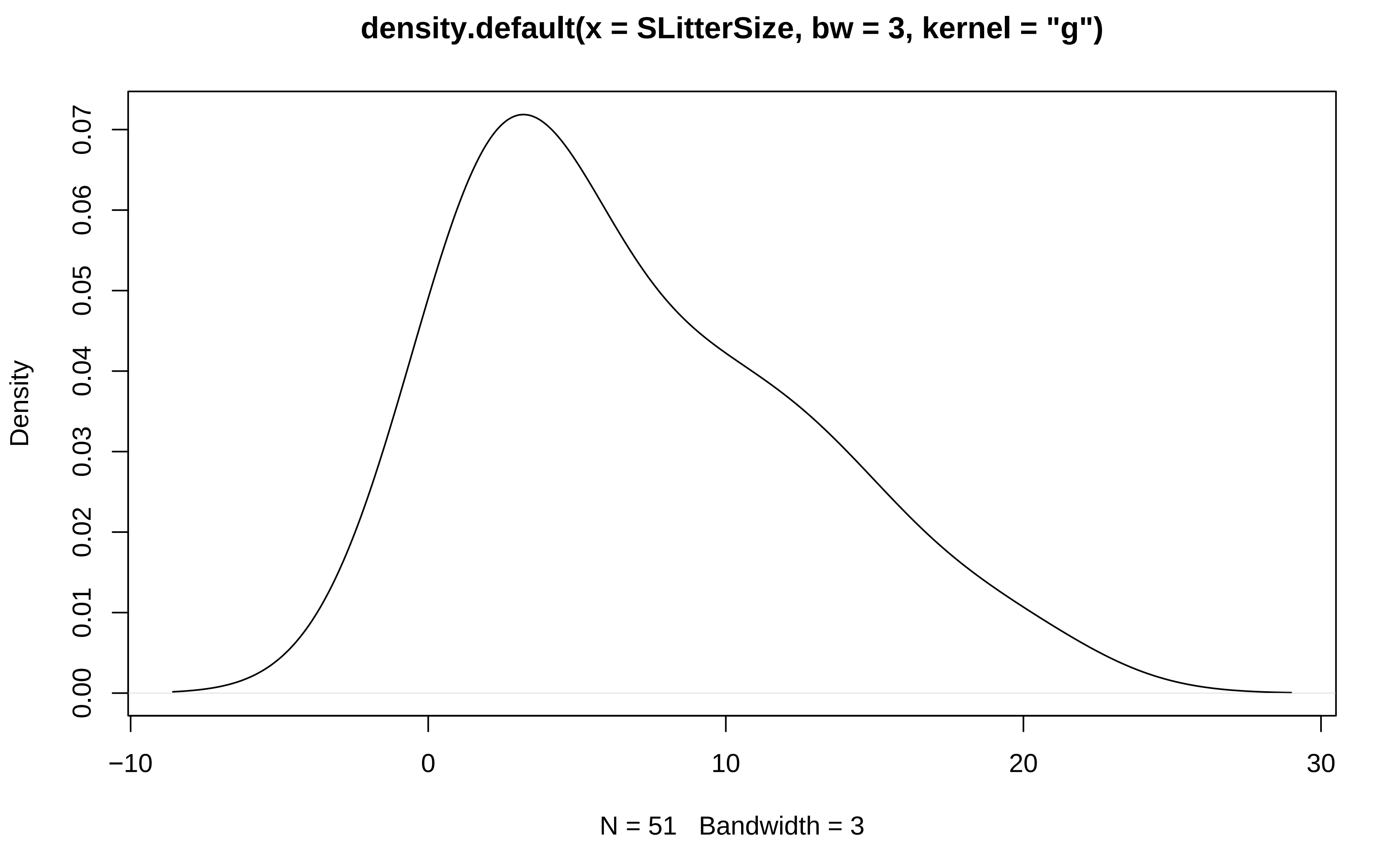
(11) E (H.O.5 pg 14)

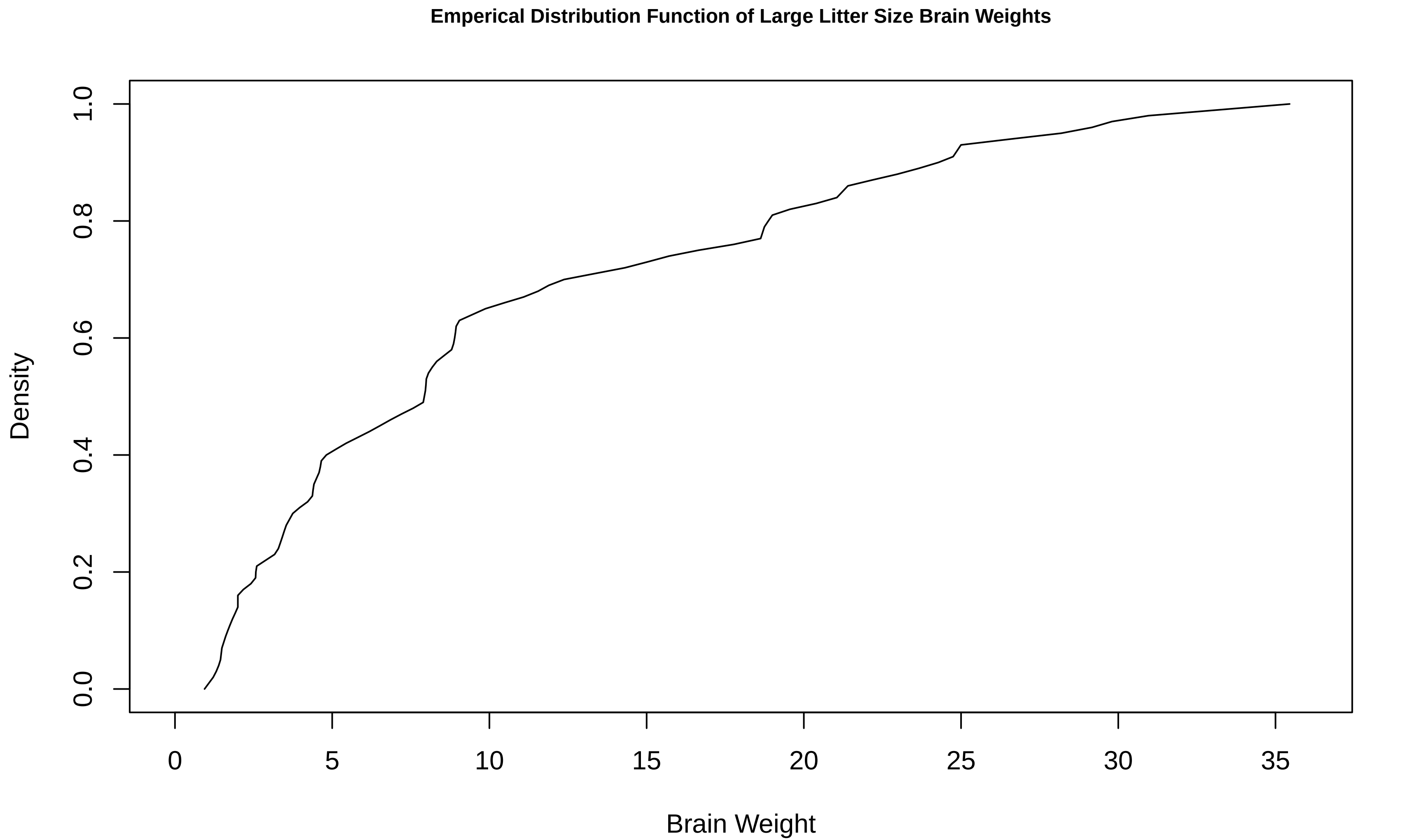
(12) E (H.O.5 pg 27 (5))

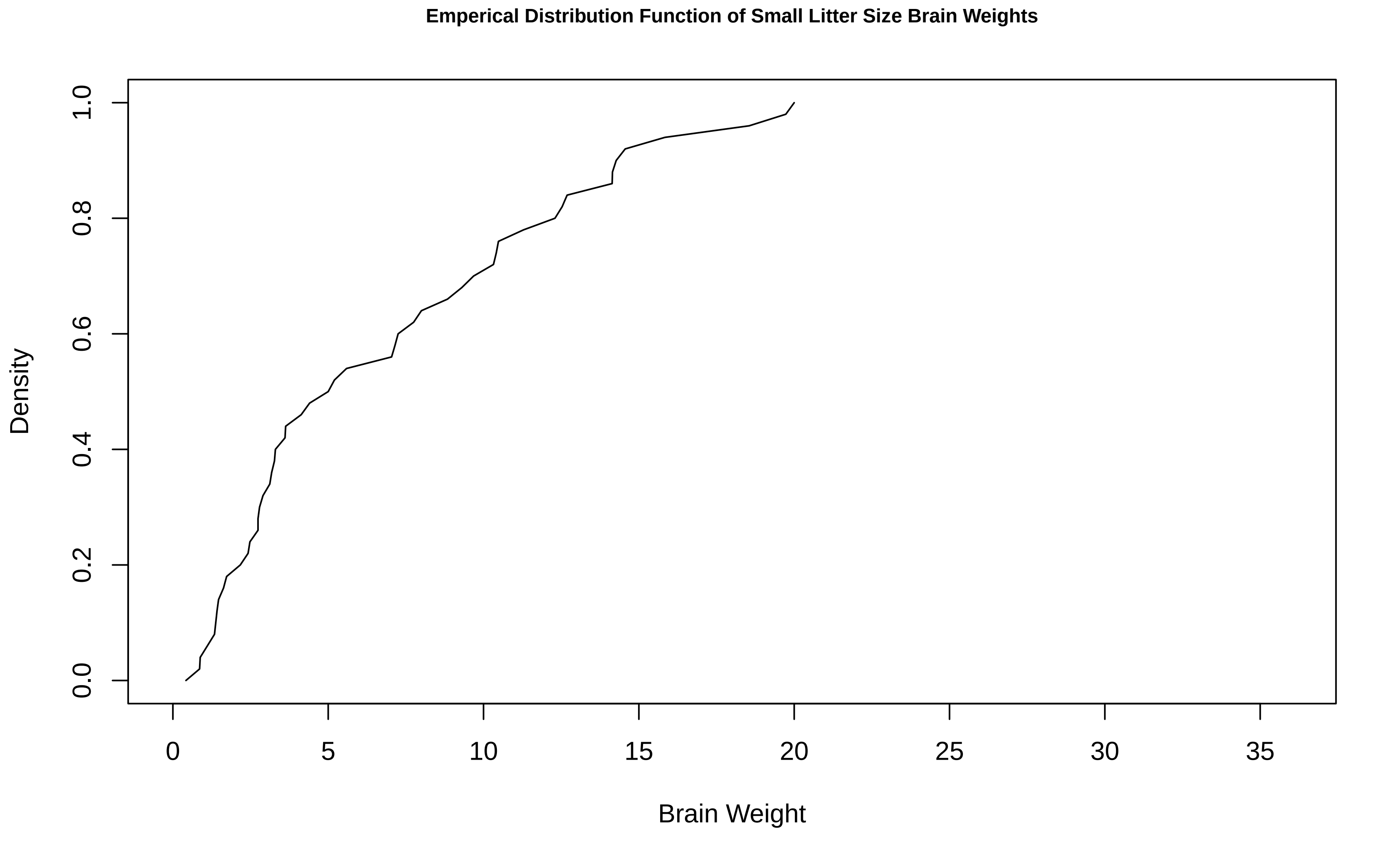
(13) B (H.O.5 pg 32)

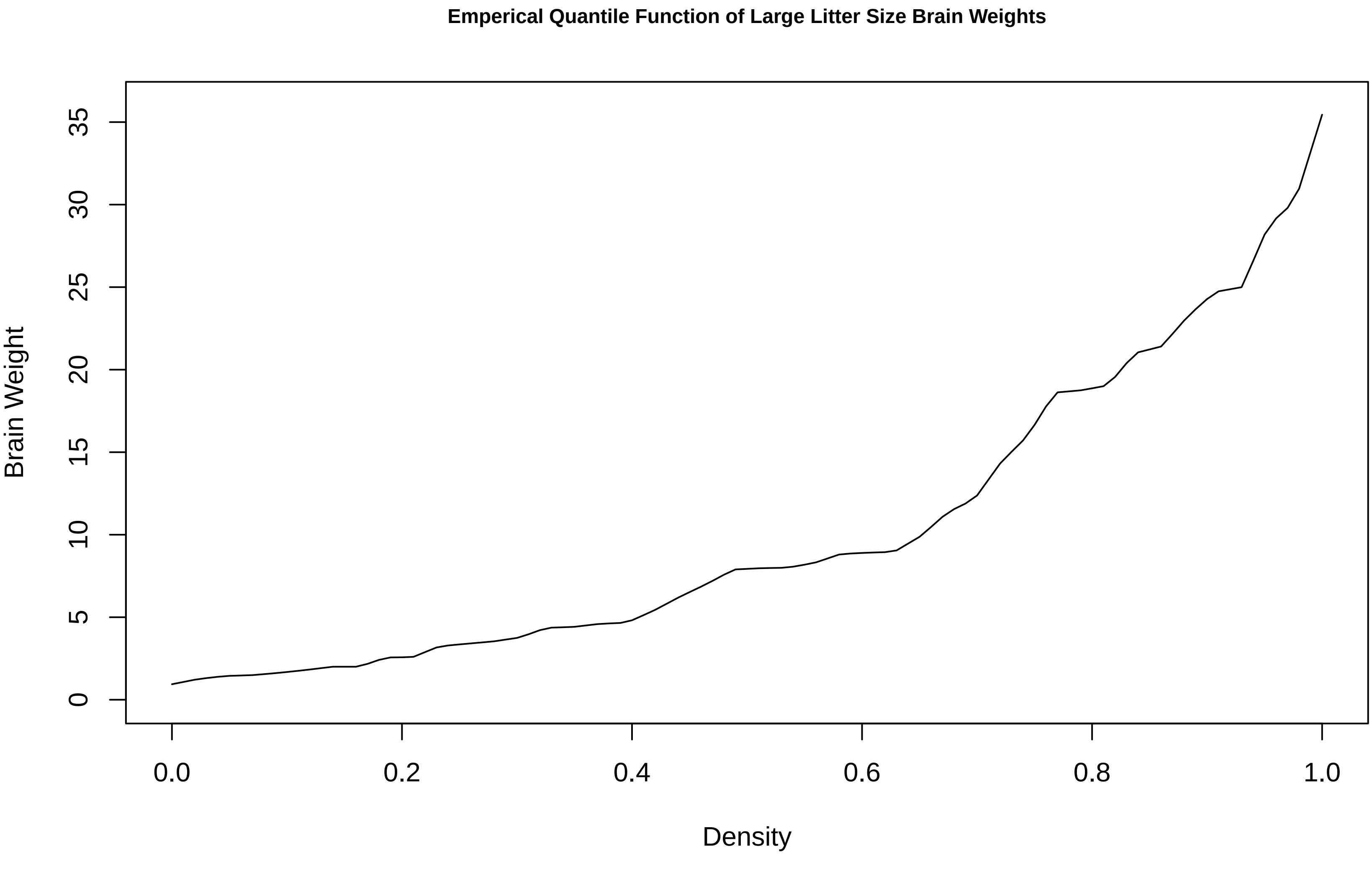

Histogram of LLitterSize
Density
0.00
0.02
0.04
0.06
0.08
0
10
20
30
40
LLitterSize


density.default(x = LLitterSize, bw = 3, kernel = "g")
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
−10
0
10
20
30
40
N = 44   Bandwidth = 3

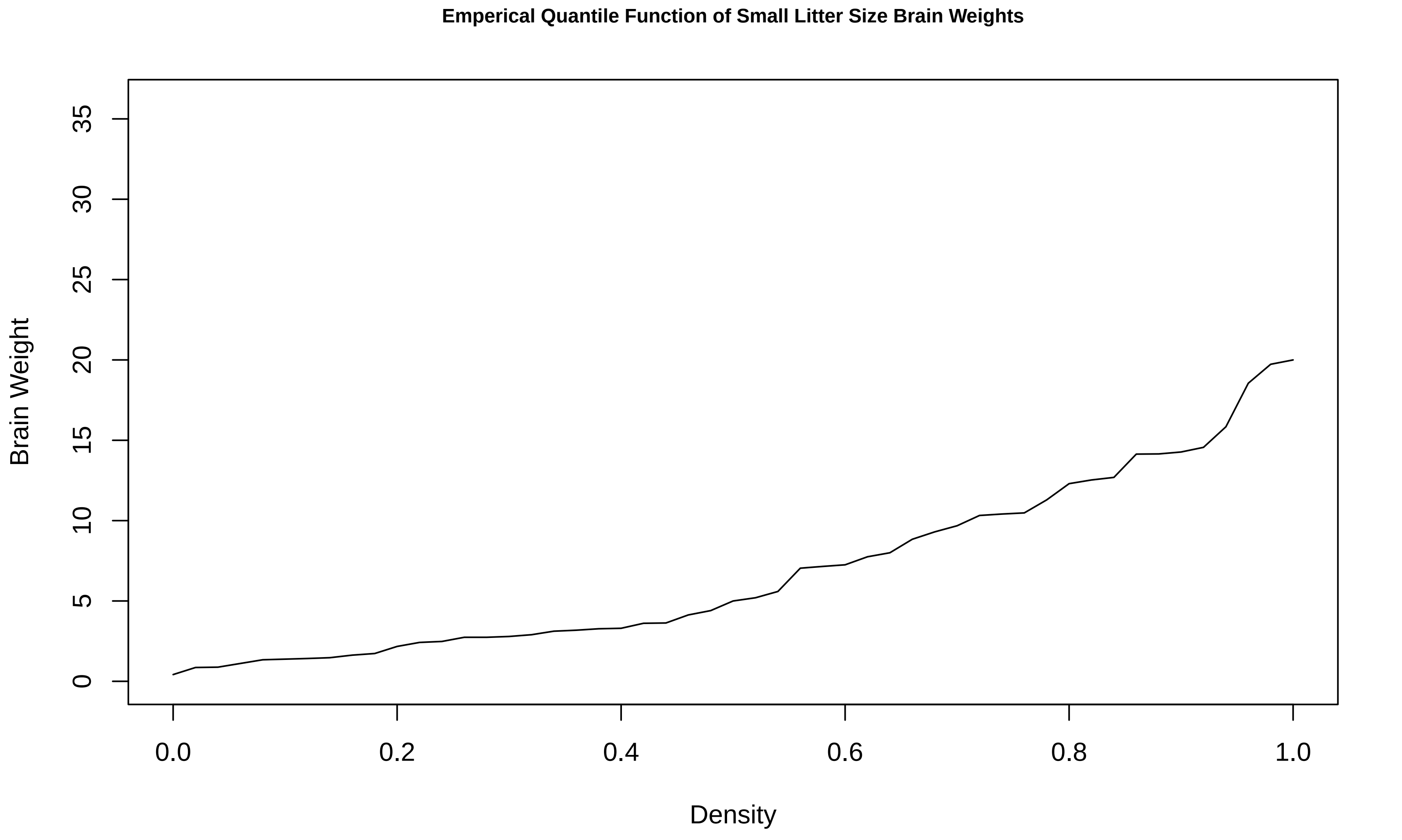

density.default(x = SLitterSize, bw = 3, kernel = "g")
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
−10
0
10
20
30
N = 51   Bandwidth = 3


Emperical Distribution Function of Large Litter Size Brain Weights
Density
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
35
Brain Weight


Emperical Distribution Function of Small Litter Size Brain Weights
Density
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
35
Brain Weight



Emperical Quantile Function of Large Litter Size Brain Weights
Brain Weight
Density
0
5
10
15
20
25
30
35
0.0
0.2
0.4
0.6
0.8
1.0



Emperical Quantile Function of Small Litter Size Brain Weights
Brain Weight
0
5
10
15
20
25
30
35
0.0
0.2
0.4
0.6
0.8
1.0
Density

```
#### STAT 641 Assignment 3

# 2.)
Assign3_BrainSize <- read.csv("C:/Users/jackr/OneDrive/Desktop/Graduate School
Courses/STAT 641 - Methods of STAT I/RawData/Assign3_BrainSize.csv")
Assign3_BrainSize

LLitterSize = Assign3_BrainSize$Large.Litter.Size[1:44]
SLitterSize = Assign3_BrainSize$ï..Small.Litter.Size

quantile(LLitterSize, na.rm = TRUE)

# 3.)
# a)
# probably most appropriate way to calculate the kernel density in R
d = density(LLitterSize, kernel = "g", bw = 3, na.rm = TRUE)
dd = approxfun(d$x, d$y)
dd(3)
dd(16)

# Gaussian kernel density estimate for a hard coded Y value
n = length(LLitterSize)
h = 3
y = 3
kern_Density_estimate = NULL
for(i in seq_along(LLitterSize)) {
  kern_Density_estimate[i] = (1/(n*h))*(1/sqrt(2*pi))*exp(-0.5*((y-
LLitterSize[i])/h)^2)
}
sum(kern_Density_estimate)




n = length(LLitterSize)
h = 3
y = 16
kern_Density_estimate = NULL
for(i in seq_along(LLitterSize)) {
  kern_Density_estimate[i] = (1/(n*h))*(1/sqrt(2*pi))*exp(-0.5*((y-
LLitterSize[i])/h)^2)
}
sum(kern_Density_estimate)


# kernel density estimate for multiple x values.
# Y-values here are chosen just as a sequence from 0 to the max value of our
dataset (rounded up to the nearest integer)
# n = length(LLitterSize)
# h = 3
# Y_vals = seq(from = 0, to = max(ceiling(LLitterSize)), by=1)
# kern_Density_estimate = NULL
# kern_Density_temp = NULL
```

```
# for(i in seq_along(Y_vals)) {
#   for(j in seq_along(LLitterSize)){
#     kern_Density_temp[j] = (1/(n*h))*(1/sqrt(2*pi))*exp(-0.5*((Y_vals[i]-
LLitterSize[j])/h)^2)
#   }
#   kern_Density_estimate[i] = sum(kern_Density_temp)
#   if(i == max(Y_vals)){ # if statement is fucking up this loop.. idk why..
figure out later, not needed for hw
#     dataf = cbind(Y_vals, kern_Density_estimate)
#   }
# }
# kern_Density_estimate


ceiling(length(LLitterSize)/5)

#b)
hist(LLitterSize, freq=FALSE, breaks = ceiling(length(LLitterSize)/5))
lines(d, col="red", lwd=2)
# f(3) approx .08
# f(16) approx .02

#c-d)
LLitterSize[which(LLitterSize == min(abs(16-LLitterSize), na.rm = TRUE) + 16)]
LLitterSize[which(LLitterSize == max(abs(16-LLitterSize), na.rm = TRUE) + 16)]

# or can use our loop above with y = 16

LLitterSize[which(kern_Density_estimate == min(kern_Density_estimate))]
LLitterSize[which(kern_Density_estimate == max(kern_Density_estimate))]


# 4.)
# (a)
# pdf estimate for Large Litter Size
plot(density(LLitterSize, kernel = "g", bw = 3))
# cdf estimate for Large litter size
QLarge = quantile(LLitterSize, probs = seq(0,1,0.01))
plot(QLarge, y = seq(0,1,0.01), xlim = c(0,36),
     type = "l", xlab = "Brain Weight", ylab = "Density",
     main = "Emperical Distribution Function of Large Litter Size Brain
Weights", cex.main = .75)
# quantile estimate for Large litter size
plot(y = QLarge, x = seq(0,1,0.01), ylim = c(0,36),
     type = "l", xlab = "Density", ylab = "Brain Weight",
     main = "Emperical Quantile Function of Large Litter Size Brain Weights",
cex.main = .75)



#pdf estimate for Small Litter Size
plot(density(SLitterSize,  kernel = "g", bw = 3))
# cdf estimate for Small litter size
QSmall = quantile(SLitterSize, probs = seq(0,1,0.01))
```

```
plot(QSmall, y = seq(0,1,0.01), xlim = c(0,36),
     type = "l", xlab = "Brain Weight", ylab = "Density",
     main = "Emperical Distribution Function of Small Litter Size Brain 
Weights", cex.main = .75)
# quantile estimate for Large litter size
plot(y = QSmall, x = seq(0,1,0.01), ylim = c(0,36),
     type = "l", xlab = "Density", ylab = "Brain Weight",
     main = "Emperical Quantile Function of Small Litter Size Brain Weights", 
cex.main = .75)
```