Final

Arthur Hernandez

5/3/2021

For this Final, I referred to course notes, previous problem sets, and course lecture videos.

Question 1:

p = Most Beautiful People have a 48.5% chance of being girls

H0:p ≤ 0.485

H1:p > 0.485

part a:

```
1 - pbinom(156, 329, 0.485)
```

```
## [1] 0.6321467
```

part b:

v(x) = npq

```
xbar = 156 / 329
# Lower
xbar - qnorm(0.975) * sqrt(xbar * (1 - xbar) / 329)
```

```
## [1] 0.4202082
```

```
# Upper
xbar + qnorm(0.975) * sqrt(xbar * (1 - xbar) / 329)
```

```
## [1] 0.5281201
```

part c:

The probability of a girl might be anywhere from 42% and 53%. Therefore, because our null hypothesis that children of the Most Beautiful People have a 48.5% chance of being girls falls between these two bounds (42%, 53%), it is satisfied. With more information we could better test this hypothesis to have a smaller confidence interval that could change our result.

Question 2:

part a:

u1 = traditional ward mean percentage weight loss

u2 = experimental ward mean percentage weight loss

Null Hypothesis: H0: u1 = u2

Alternative Hypothesis: H1: u1 ≠ u2

This would be a Students two sample t-test because our standard deviations are the same so the variances are as well.

part b:

```
Delta.hat = .051 - .06
print(Delta.hat)
```

```
## [1] -0.009
```

```
se =sqrt(.02^2/393 + .02^2 / 388)
print(se)
```

```
## [1] 0.001431342
```

```
t.Welch = Delta.hat/se
print(t.Welch)
```

```
## [1] -6.287806
```

```
nu = (.02^2/393+.02^2/388)^2 / ((.02^2/393)^2/392 + (.02^2/388)^2 / 387)
print(nu)
```

```
## [1] 778.872
```

```
P.value <- 2*(1- pt(abs(t.Welch), df = nu))
P.value
```

```
## [1] 5.361236e-10
```

part c:

```
q <-qt(0.975, df=nu)
lower <- Delta.hat-q * se
upper <- Delta.hat+q * se
lower
```

```
## [1] -0.01180974
```

```
upper
```

```
## [1] -0.006190255
```

part d:

Since the p-value is less than .05 we can reject the null hypothesis and can assume that these two populations came from different mean percentages.

Question 3:

Heights of sample: mean: $\bar{X} = 48$, standard deviation: sX = 3.75 Math scores of sample: mean $\bar{Y} = 50$, standard deviation sY = 15 Sample correlation: r = 0.6

part a:

```
xbar = 48
ybar = 50
sx = 3.75
sy = 15
r = .6
b = r * sy / sx
a = ybar - b * xbar
print(a)
```

```
## [1] -65.2
```

```
print(b)
```

```
## [1] 2.4
```

reg = -65.2 + 2.4x

part b:

```
x = 50
reg = a + b*x
print(reg)
```

```
## [1] 54.8
```

part c*: Yes, this is what the confidence interval means The interval has a probability of containing the value of of all samples that could be drawn in that interval, the confidence interval will cover its true value. Question 4:

part a:

```
gm <- (98 * .49 + 94 * -0.56 + 98* 0.04) / 291
SSB <- 98 * (.49 - gm)^2 * + 94 * (-0.56 - gm)^2 + 98 * (.04 - gm)^2
SSW <- 98 * 1.34^2 + 94 * -.56^2 + 98 * 1.05^2
SST <- SSB + SSW
dfb <- 3
dfw <- 98 + 94 + 98 - 3
dft <- 98 + 94 + 98 -1
MSB <- SSB / dfb
MSW <- SSW / dfw
F.stat <- MSB / MSW
P.value <- 1 - pf(F.stat, dfb, dfw)
#Between
#Sum of squares
SSB
```

```
## [1] 694.6193
```

```
#DF
dfb
```

```
## [1] 3
```

```
#Mean Square
MSB
```

```
## [1] 231.5398
```

```
#F
F.stat
```

```
## [1] 261.0714
```

```
#P-value
P.value
```

```
## [1] 0
```

```
#Within
#Sum of squares
SSW
```

```
## [1] 254.5354
```

```
#DF
dfw
```

```
## [1] 287
```

```
#Mean Square
MSW
```

```
## [1] 0.8868829
```

```
#total
#Sum of squares
SST
```

```
## [1] 949.1547
```

```
#DF
dft
```

```
## [1] 289
```

part b:

i. Is there a significant difference in average authoritarianism between 300 viewers and V for Vendetta viewers?

ii. Is there a significant difference in average authoritarianism between 300 viewers and 21 Jump Street viewers?

iii. Is there a significant difference in average authoritarianism between V for Vendetta and 21 Jump Street viewers?

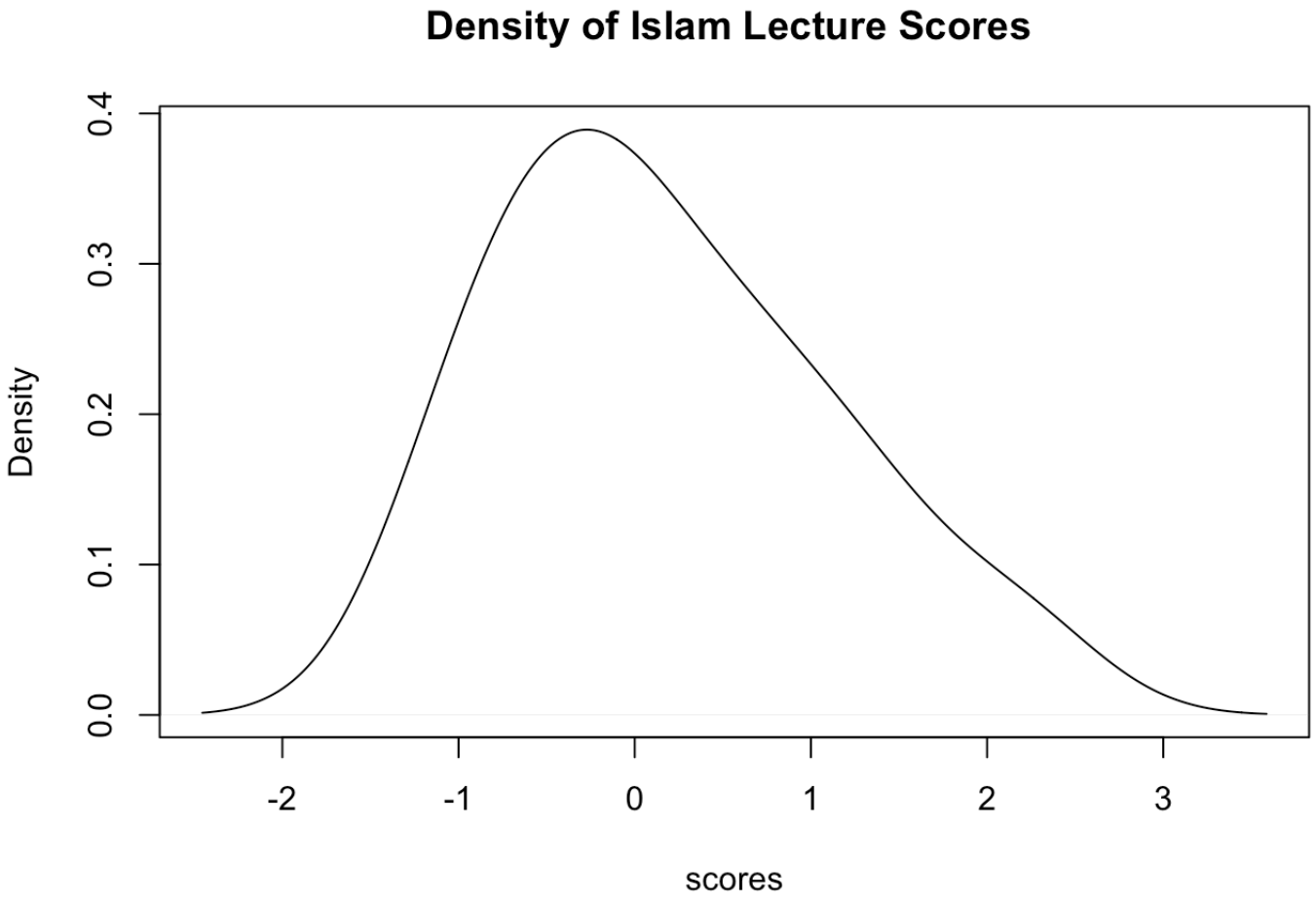
Question 5:

65 students either laptop or paper too many differences in the way they were randomized P -value of 0.03 null hypothesis of no difference between laptops and paper notes on the open questions. p = difference between laptops and paper notes on the open questions. H0 = 0 H1 ≠ 0

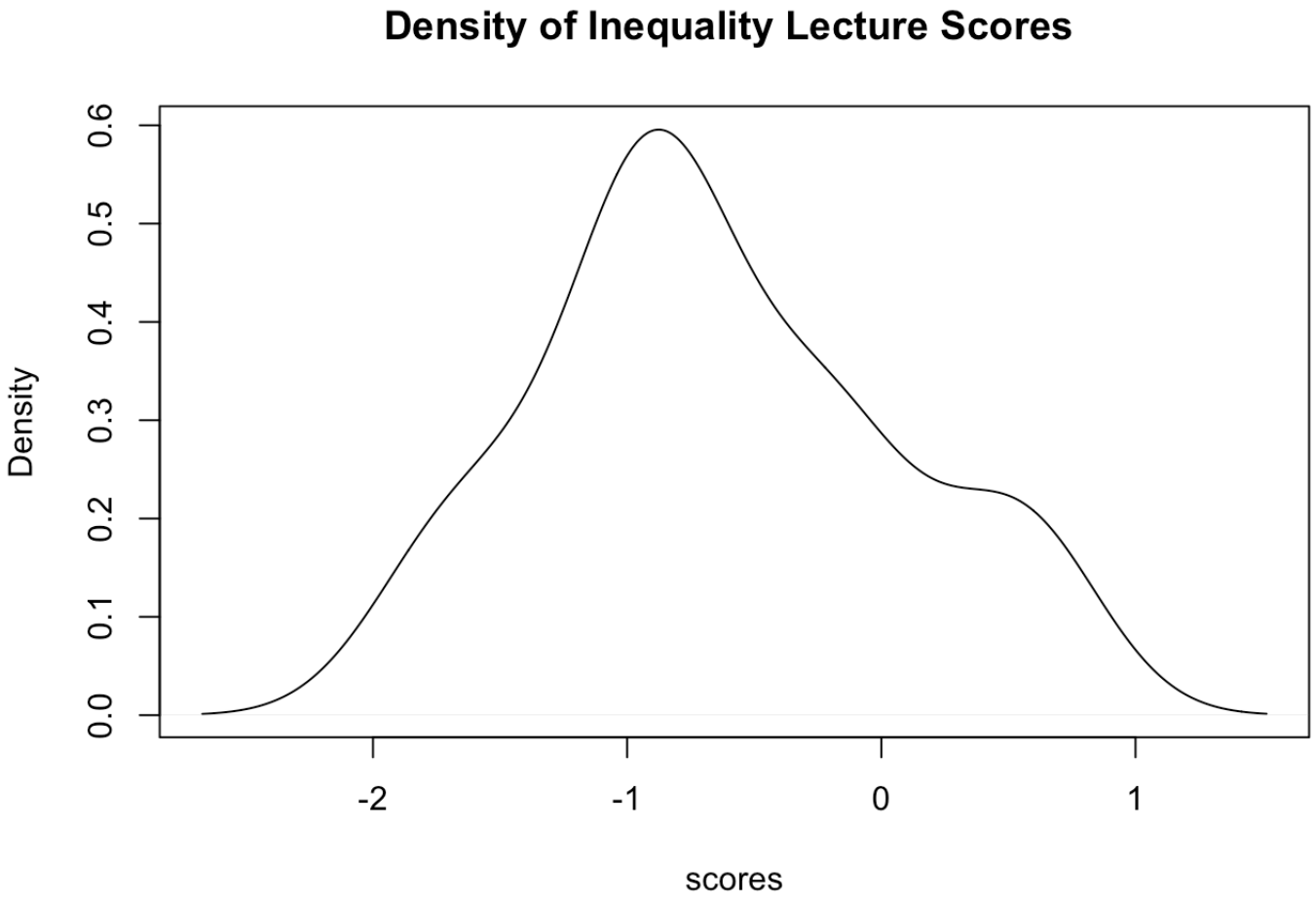
```
lap <- read.table("laptopstudy1.txt", header = TRUE)
```

part a:

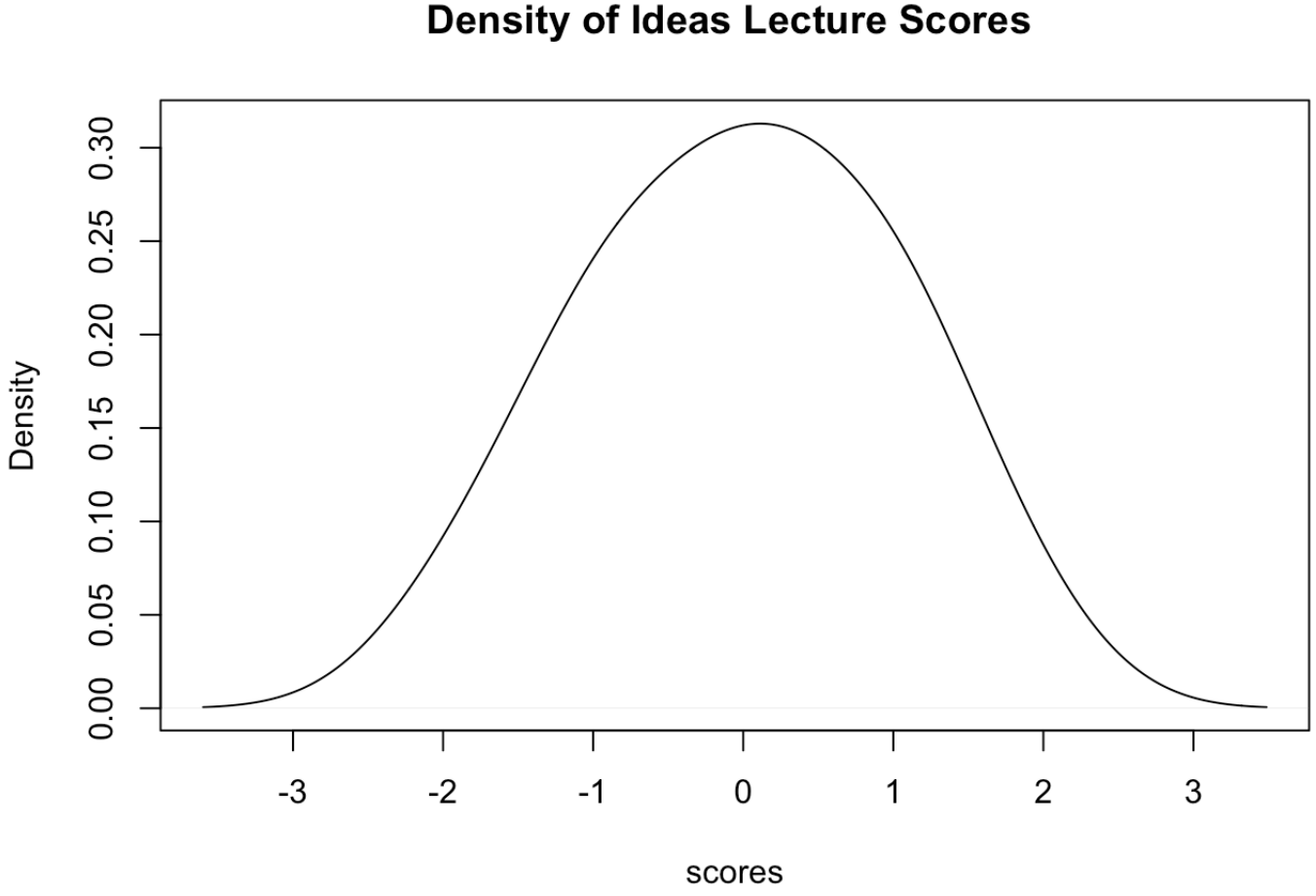
```
library(ggplot2)
lap$whichtalk <- unclass(lap$whichtalk)
lap$condition = factor(lap$condition )

islam <- subset(lap, lap$whichtalk == 1 )
inequality <- subset(lap, lap$whichtalk == 2 )
ideas <- subset(lap, lap$whichtalk == 3 )
indus <- subset(lap, lap$whichtalk == 4 )
algorithms<- subset(lap, lap$whichtalk == 5 )

plot(density(islam$objectiveZ), main = "Density of Islam Lecture Scores ",xlab = "sco
res")
```



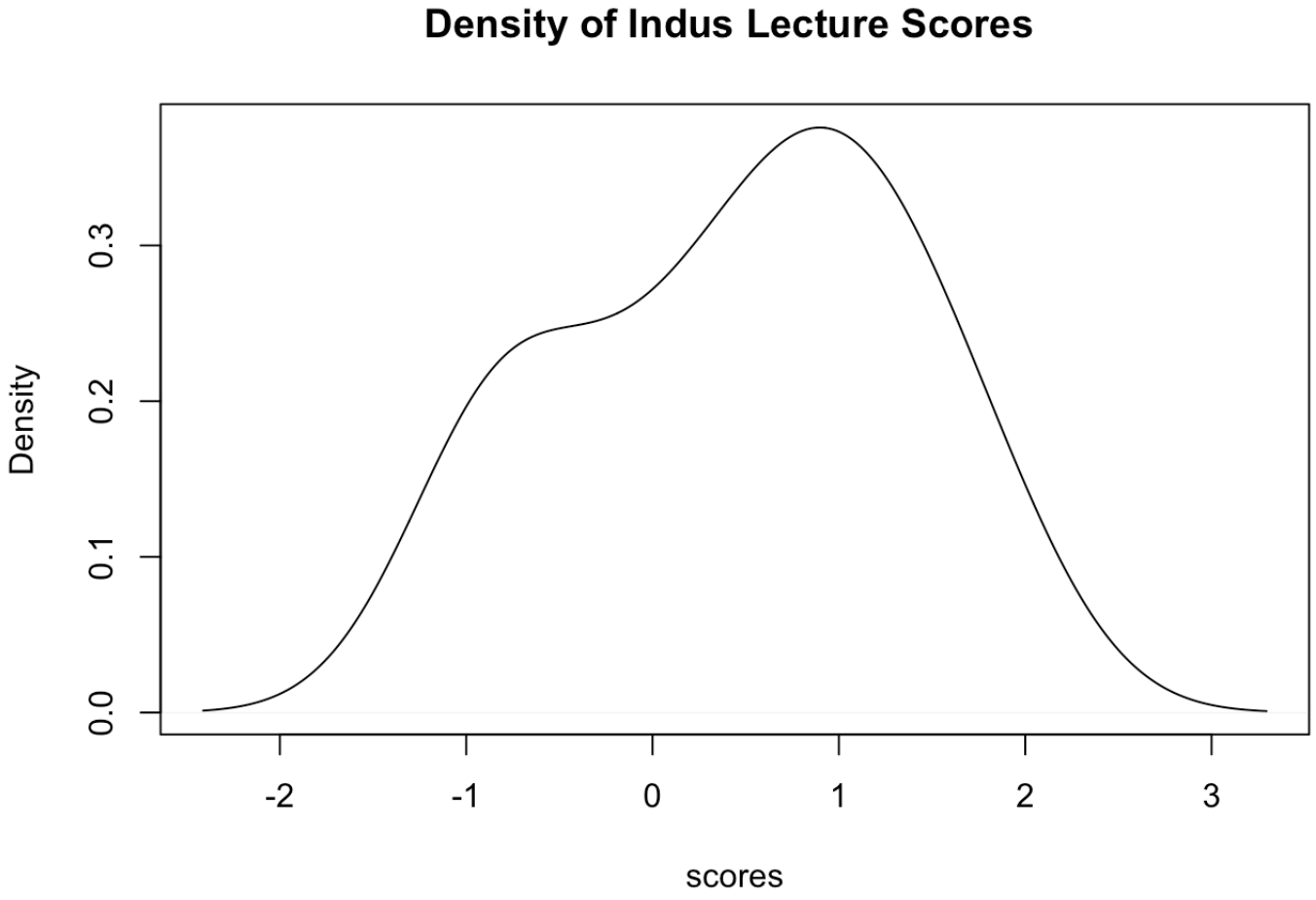

```
plot(density(inequality$objectiveZ), main = "Density of Inequality Lecture Scores ",x
lab = "scores")
```



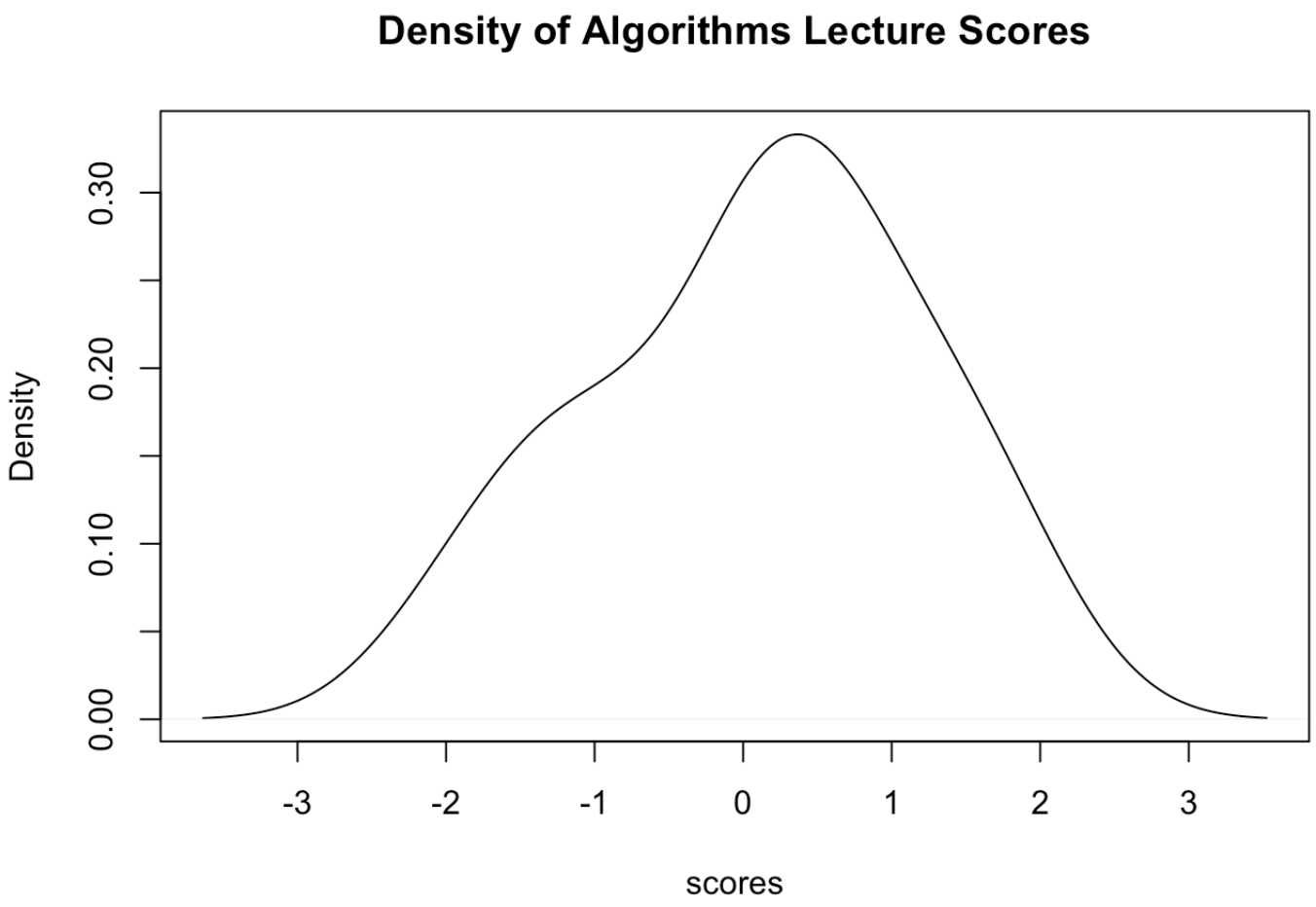

```
plot(density(ideas$objectiveZ), main = "Density of Ideas Lecture Scores ",xlab = "sco
res")
```




```
plot(density(indus$objectiveZ), main = "Density of Indus Lecture Scores ",xlab = "sco
res")
```



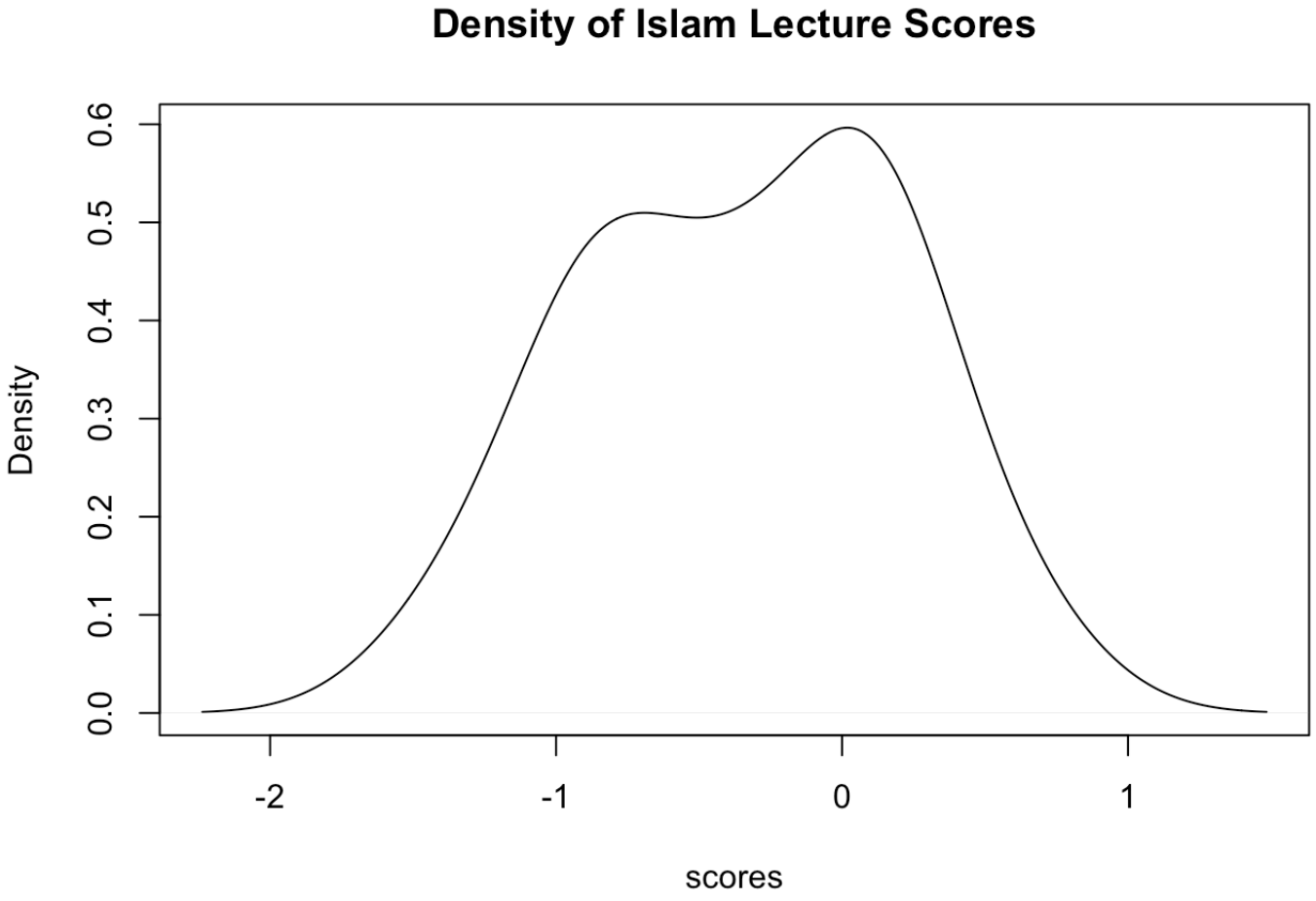

```
plot(density(algorithms$objectiveZ), main = "Density of Algorithms Lecture Scores",xl
ab = "scores")
```



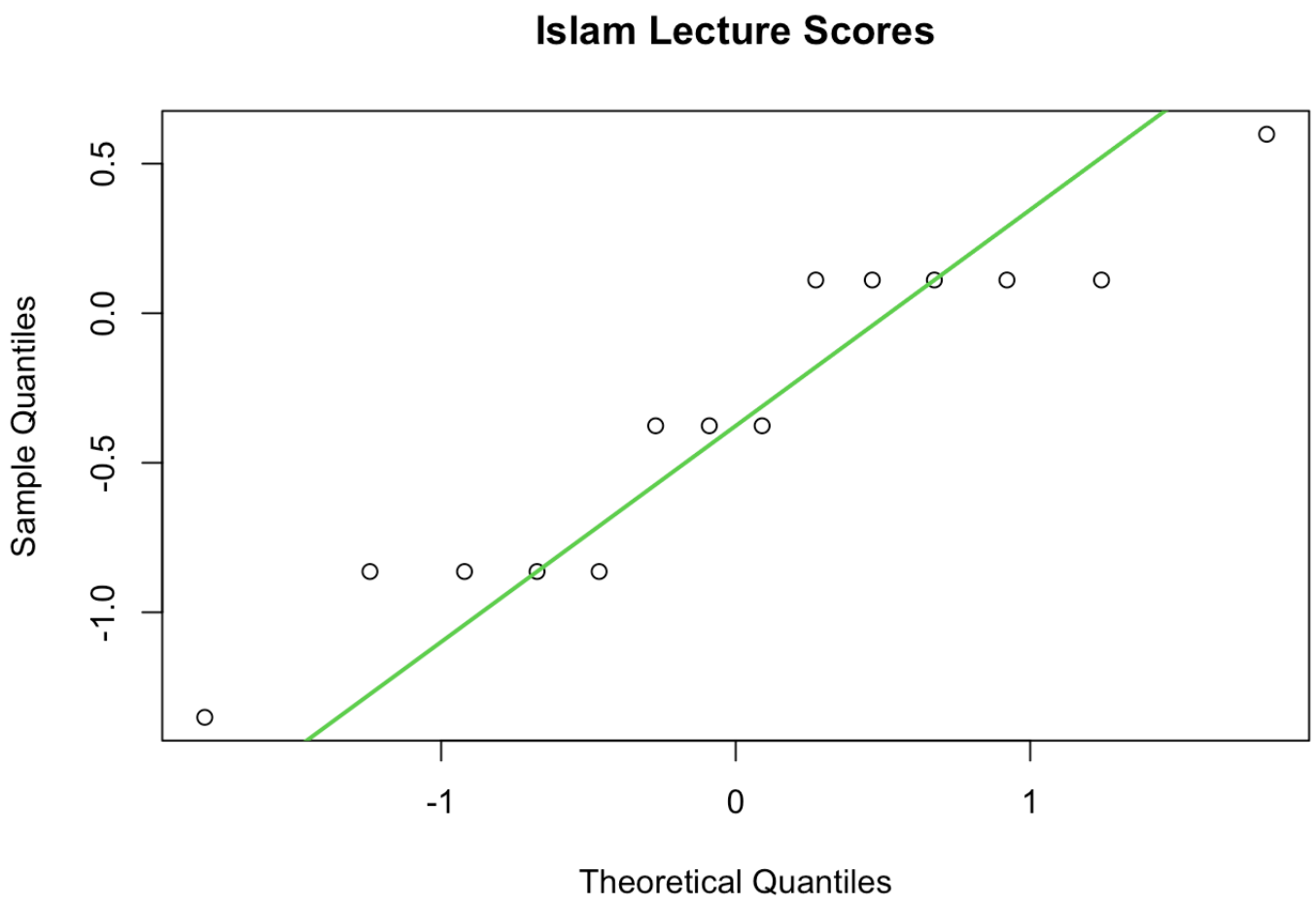
Yes, the distributions for the different lectures are off.

part b:

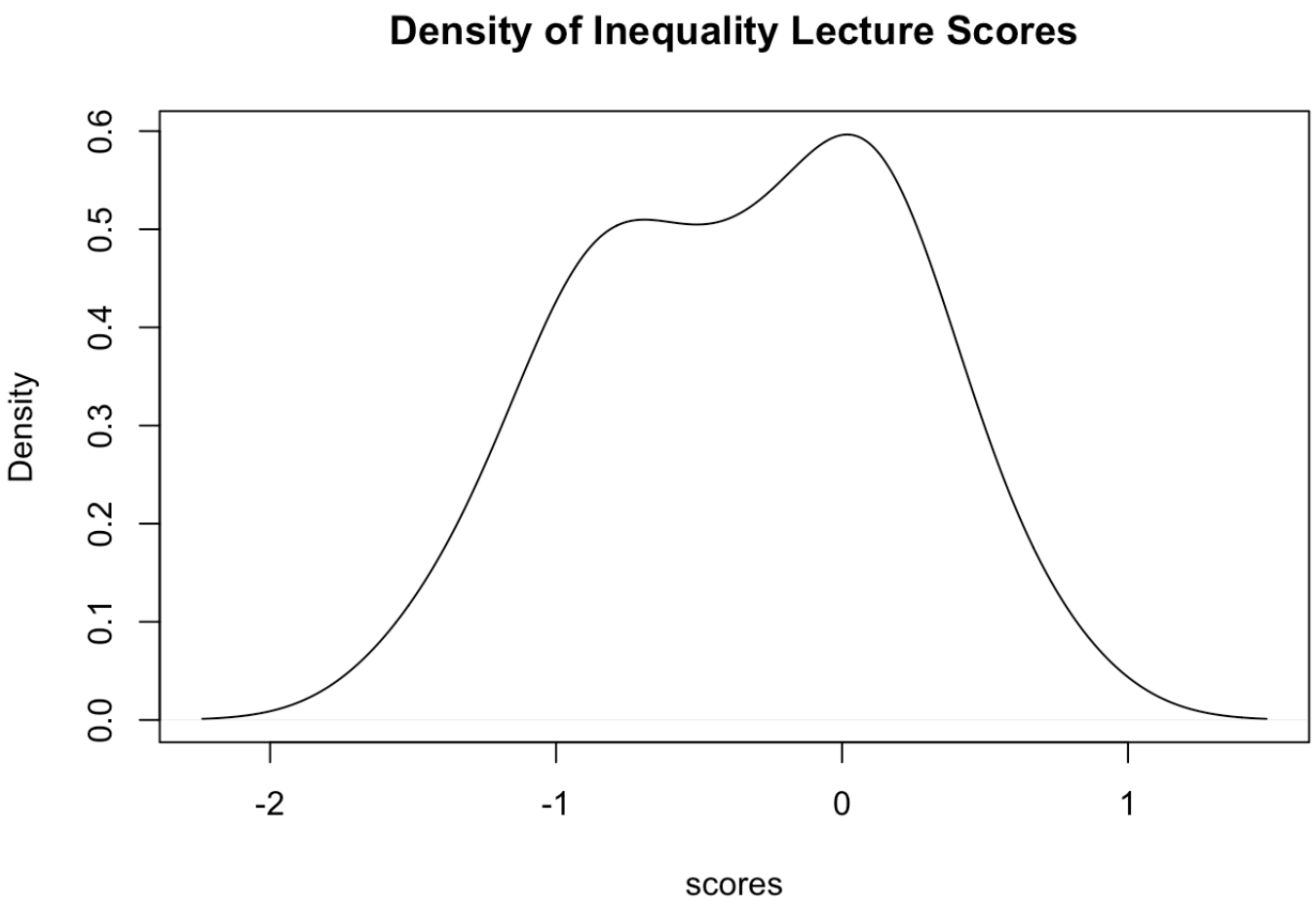
```
plot(density(islam$openZ), main = "Density of Islam Lecture Scores",xlab = "scores")
```



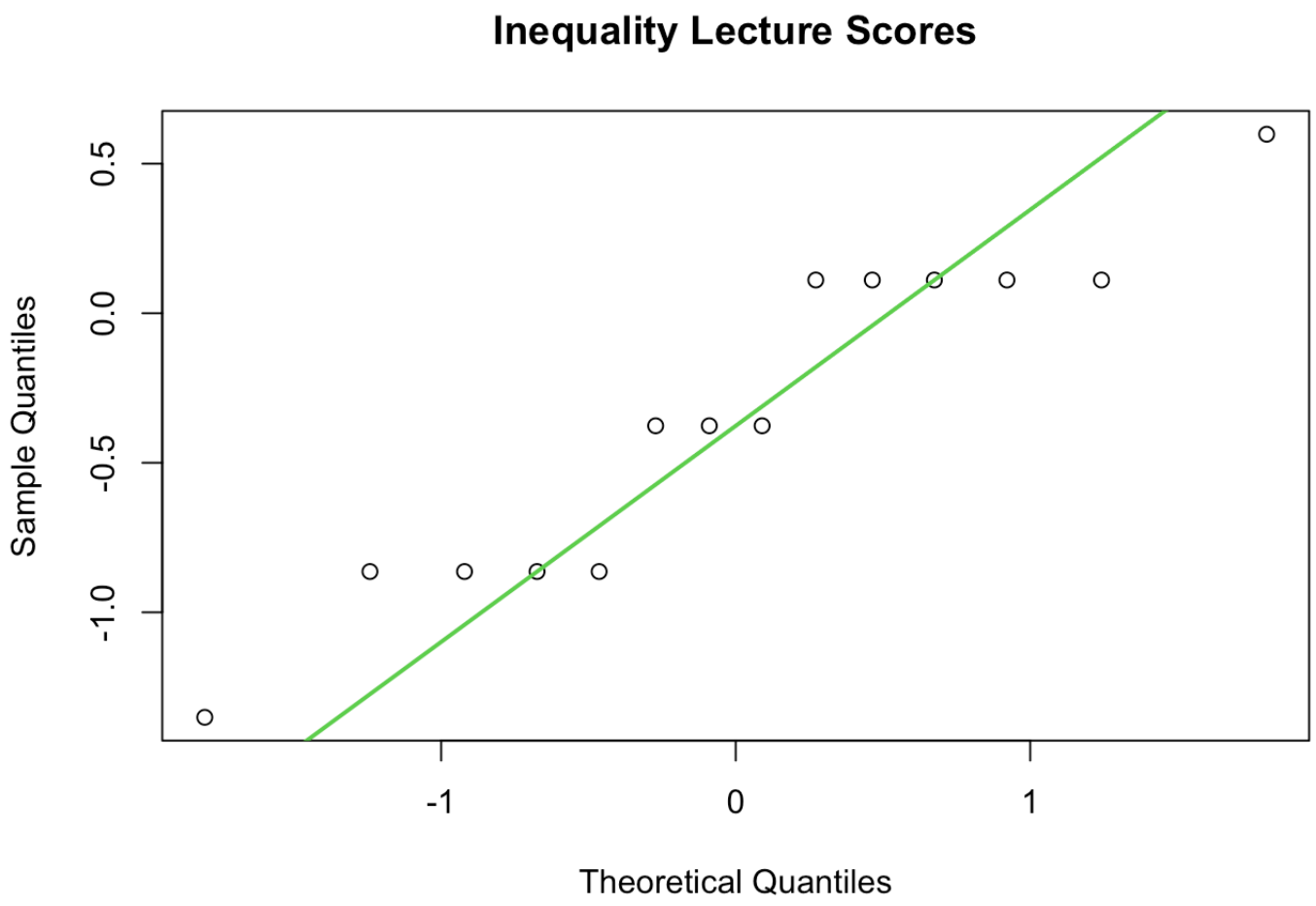

```
qqnorm(islam$openZ, main = 'Islam Lecture Scores')
qqline(islam$openZ, col = 3, lwd = 2)
```



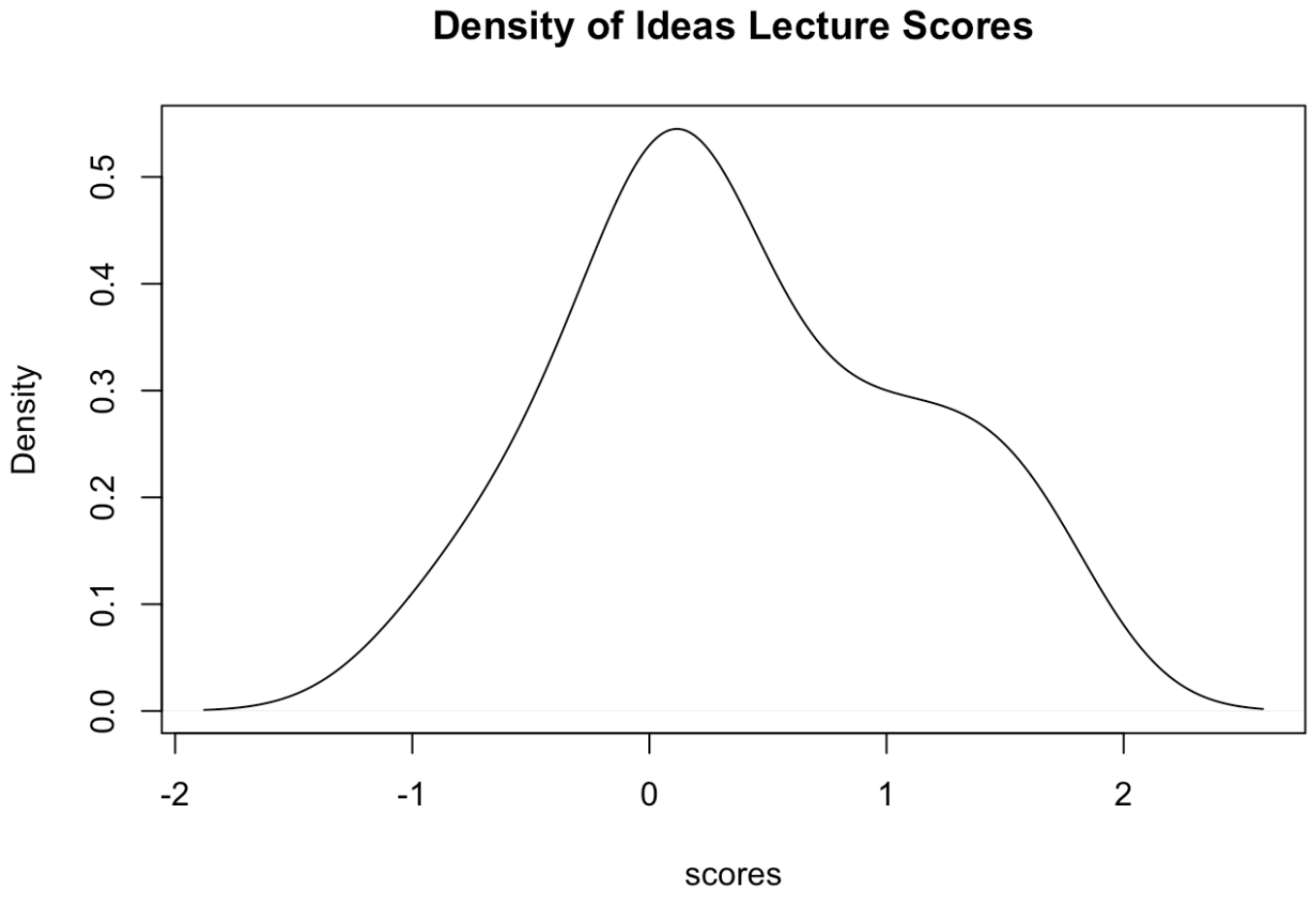

```
plot(density(islam$openZ), main = "Density of Inequality Lecture Scores",xlab = "scor
es")
```



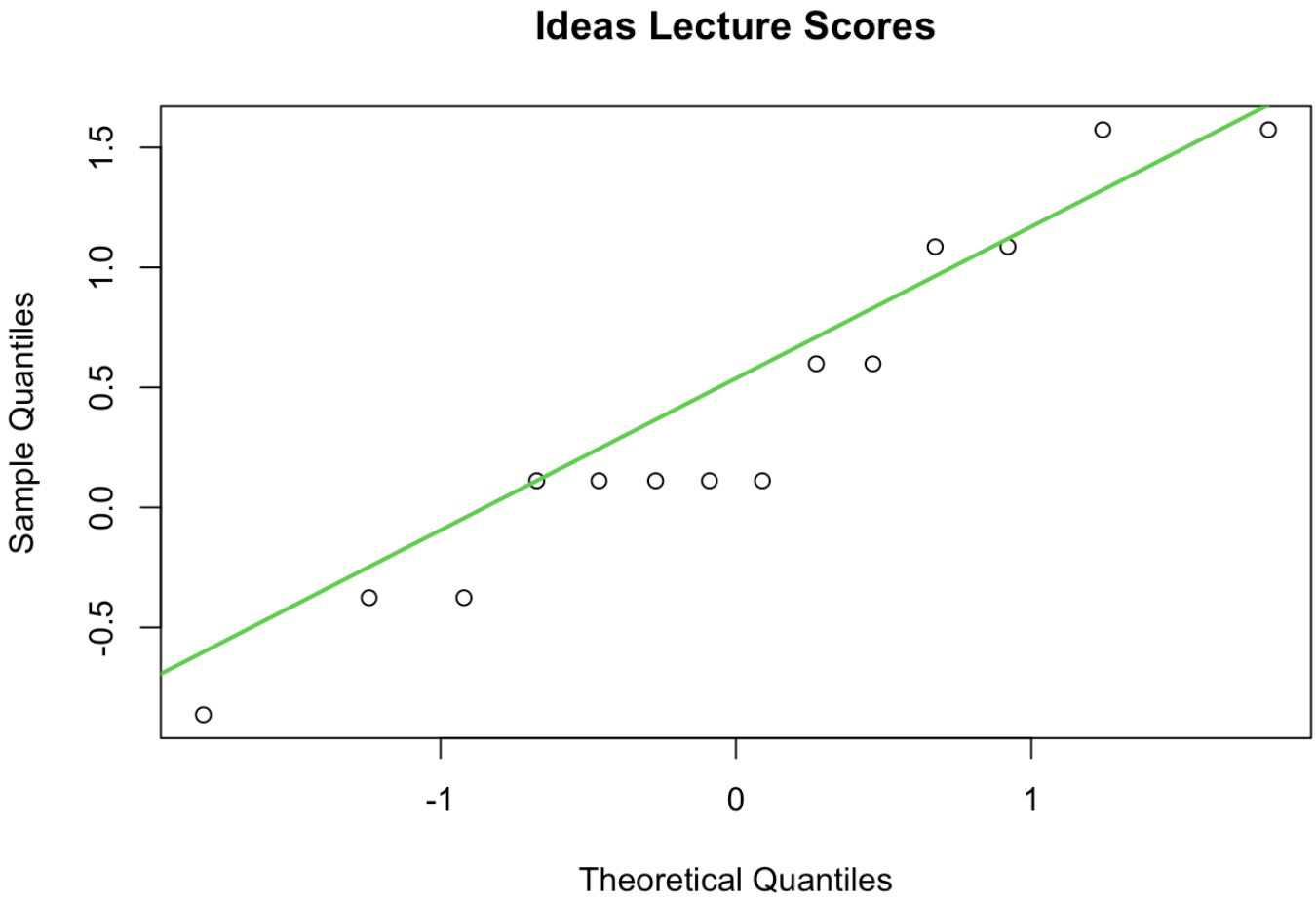

```
qqnorm(islam$openZ, main = 'Inequality Lecture Scores')
qqline(islam$openZ, col = 3, lwd = 2)
```



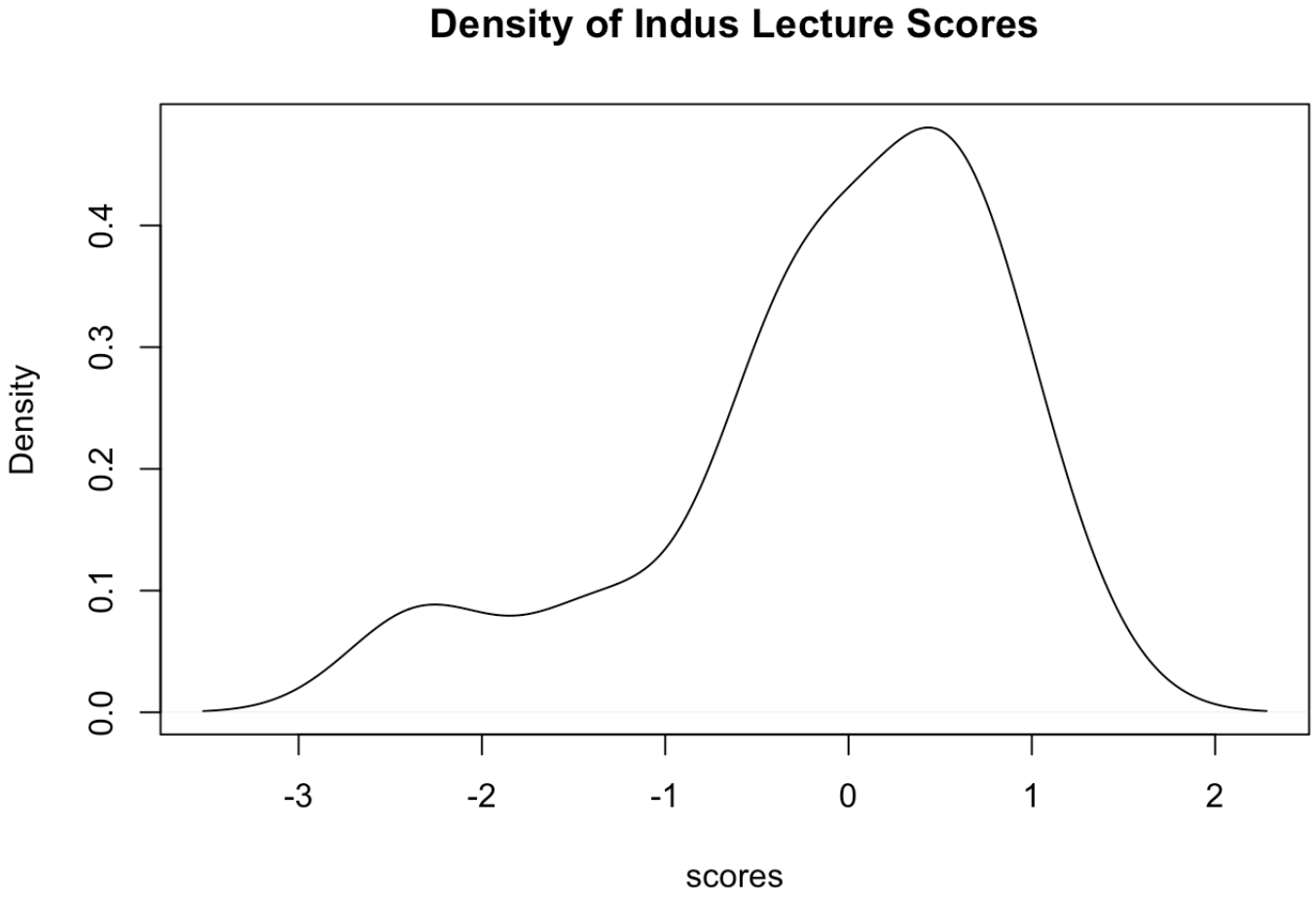

```
plot(density(ideas$openZ), main = "Density of Ideas Lecture Scores",xlab = "scores")
```



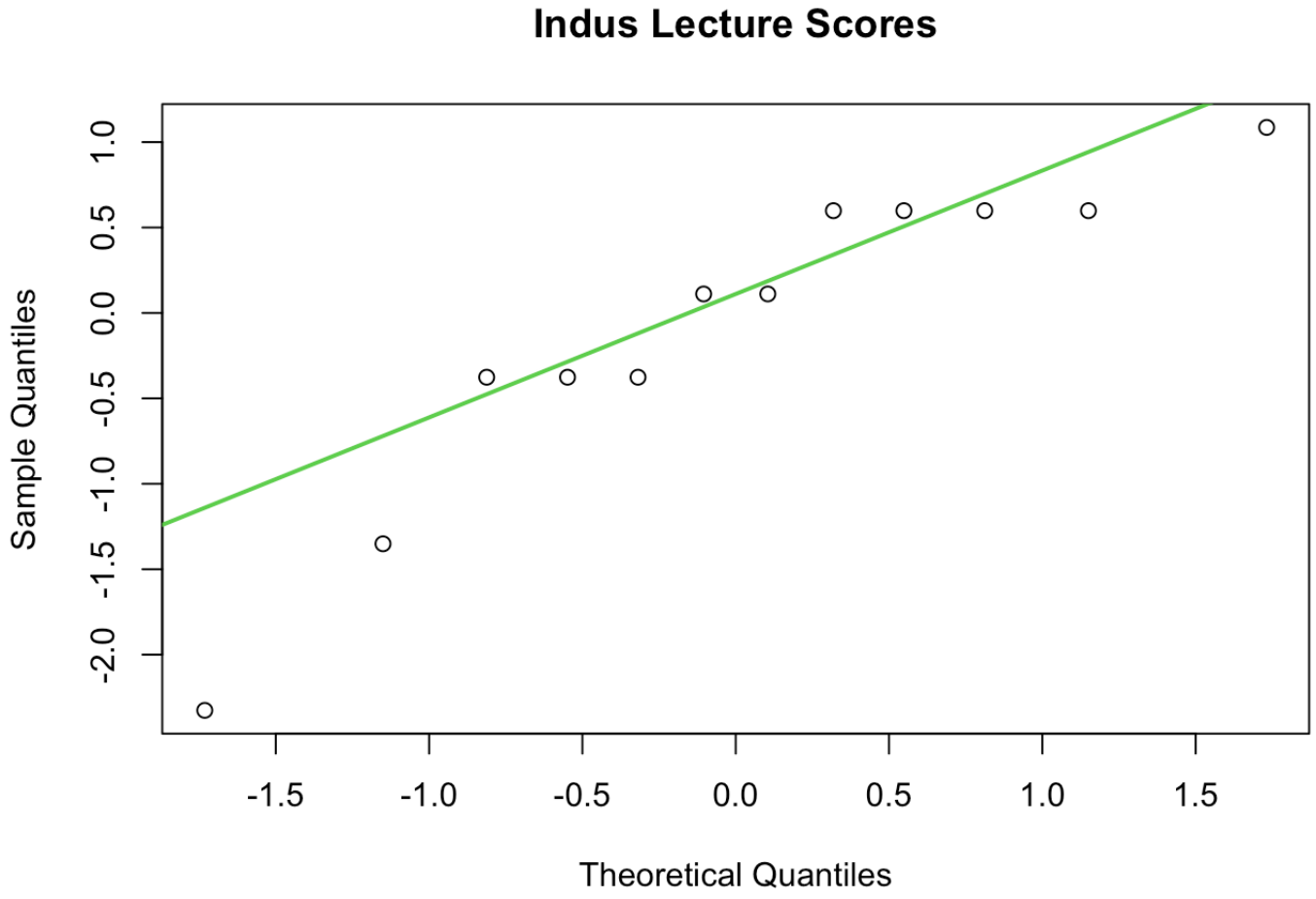

```
qqnorm(ideas$openZ, main = 'Ideas Lecture Scores')
qqline(ideas$openZ, col = 3, lwd = 2)
```



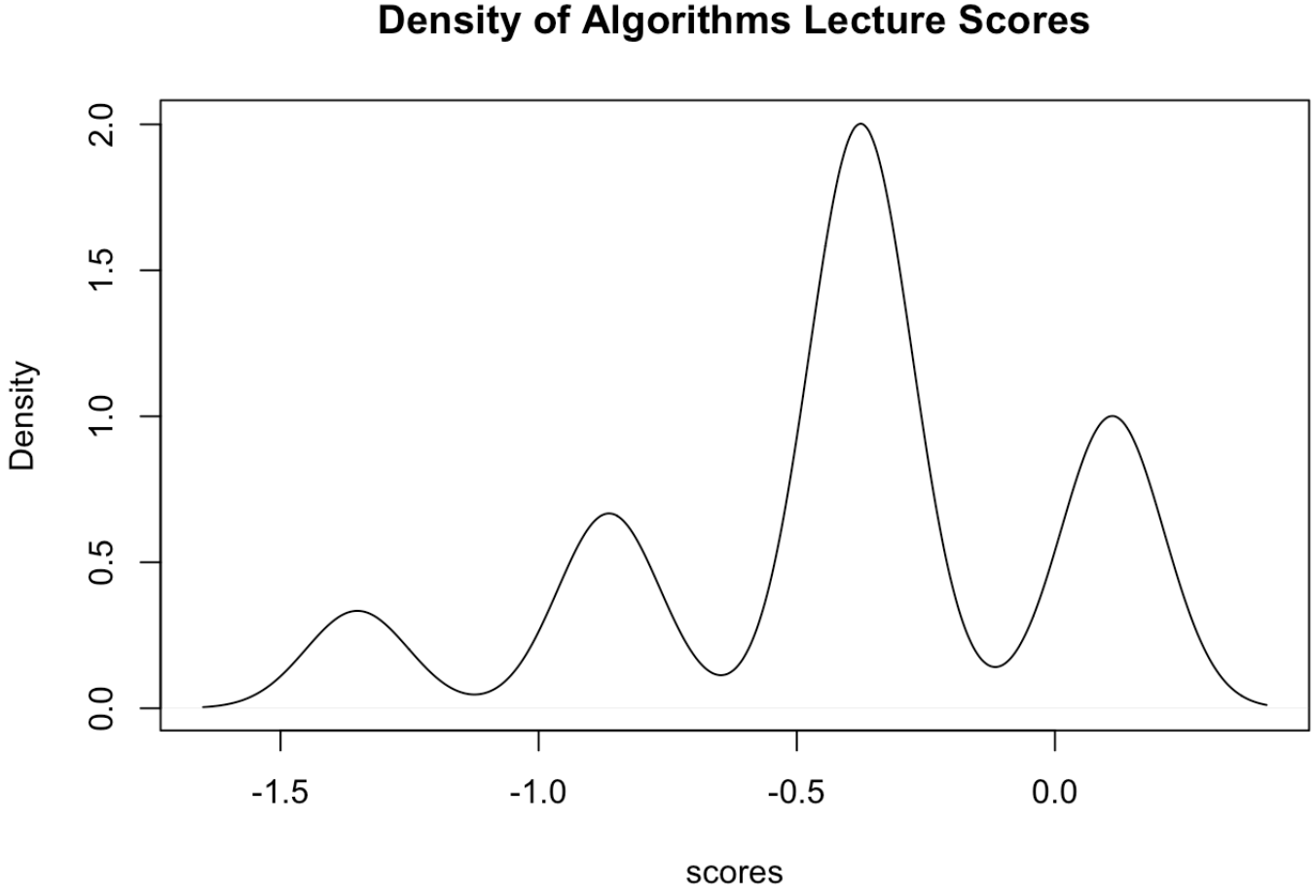

```
plot(density(indus$openZ), main = "Density of Indus Lecture Scores",xlab = "scores")
```



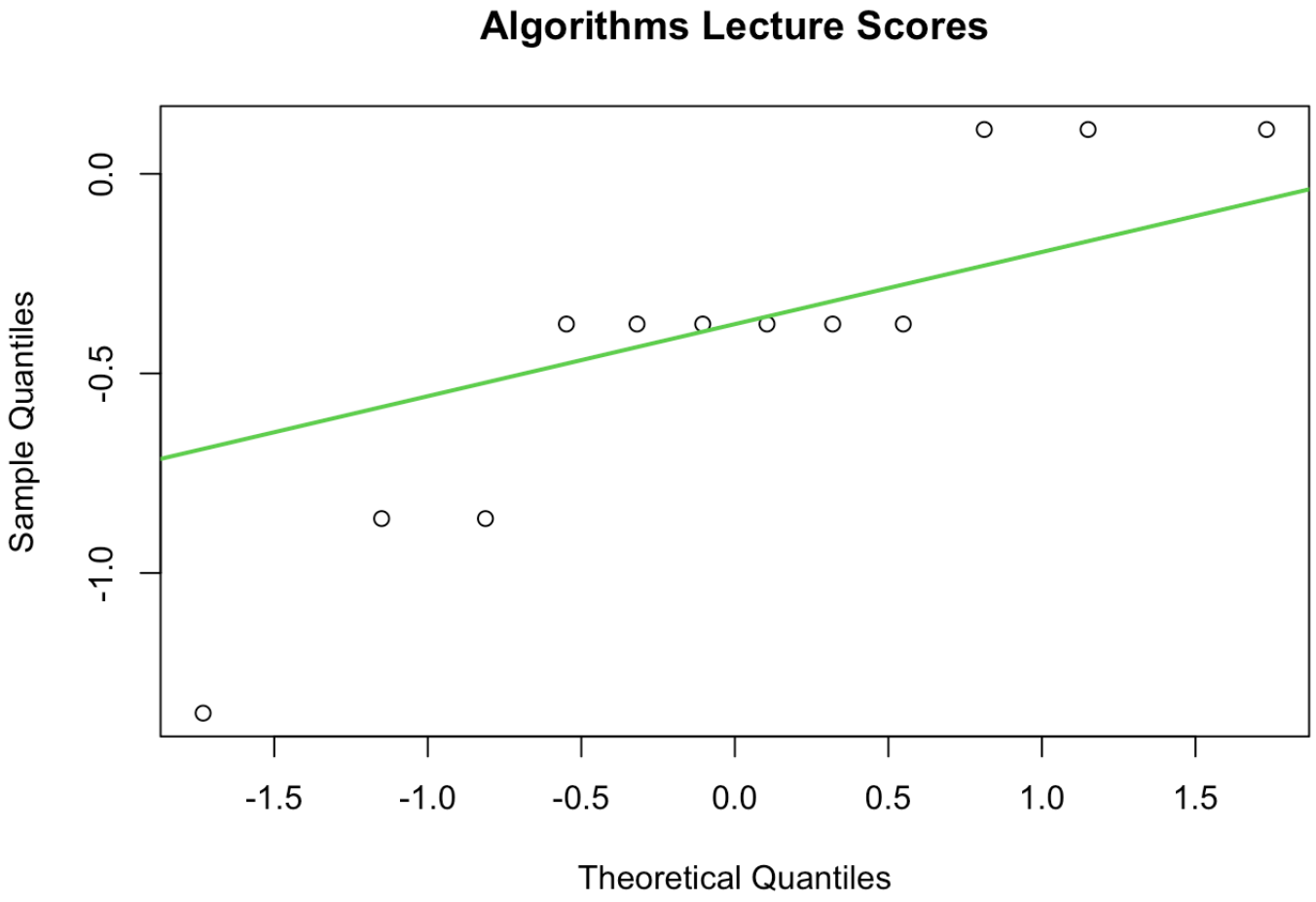

```
qqnorm(indus$openZ, main = 'Indus Lecture Scores')
qqline(indus$openZ, col = 3, lwd = 2)
```




```
plot(density(algorithms$openZ), main = "Density of Algorithms Lecture Scores",xlab =
"scores")
```




```
qqnorm(algorithms$openZ, main = 'Algorithms Lecture Scores')
qqline(algorithms$openZ, col = 3, lwd = 2)
```



Yes, they were successful at making standard normal distributions

part c:

u1 = objectiveZ paper scores u2 = objectiveZ laptop scores H0: u1 = u2 H1: u1 ≠ u2

```
paper <- subset(lap, lap$condition == 1 )
laptop <- subset(lap, lap$condition == 0 )

summary(paper$objectiveZ)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -1.920836 -0.683671  0.068574  0.008599  0.671614  2.182323
```

```
summary(laptop$objectiveZ)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1.17481 -0.80179 -0.30444  0.02125  1.00111  1.80931
```

```
xbarpaper <- mean(paper$objectiveZ)
q <- qnorm(.95)
spaper <- sd(paper$objectiveZ)
npaper <- length(paper)

#Lower
xbarpaper - q * spaper / sqrt(npaper)
```

```
## [1] -0.2877813
```

```
#Upper
xbarpaper + q * spaper / sqrt(npaper)
```

```
## [1] 0.3049803
```

```
xbarlaptop <- mean(laptop$objectiveZ)
slaptop <- sd(laptop$objectiveZ)
nlaptop <- length(laptop)

#Lower
xbarlaptop - q * slaptop/ sqrt(nlaptop)
```

```
## [1] -0.2685245
```

```
#Upper
xbarlaptop + q * slaptop / sqrt(nlaptop)
```

```
## [1] 0.3110158
```

Yes, there is a strong difference betweeen the

part d:

u1 = openZ paper scores u2 = openZ laptop scores H0: u1 = u2 H1: u1 ≠ u2

```
summary(paper$openZ)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.3260 -0.3763  0.1111  0.1541  0.5986  2.5483
```

```
summary(laptop$openZ)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.3511 -0.8637 -0.3763 -0.1562  0.1111  2.5483
```

```
xbarpaper <- mean(paper$openZ)
q <- qnorm(.95)
spaper <- sd(paper$openZ)
npaper <- length(paper)

#Lower
xbarpaper - q * spaper / sqrt(npaper)
```

```
## [1] -0.1586531
```

```
#Upper
xbarpaper + q * spaper / sqrt(npaper)
```

```
## [1] 0.4669347
```

```
xbarlaptop <- mean(laptop$openZ)
slaptop <- sd(laptop$openZ)
nlaptop <- length(laptop)

#Lower
xbarlaptop - q * slaptop/ sqrt(nlaptop)
```

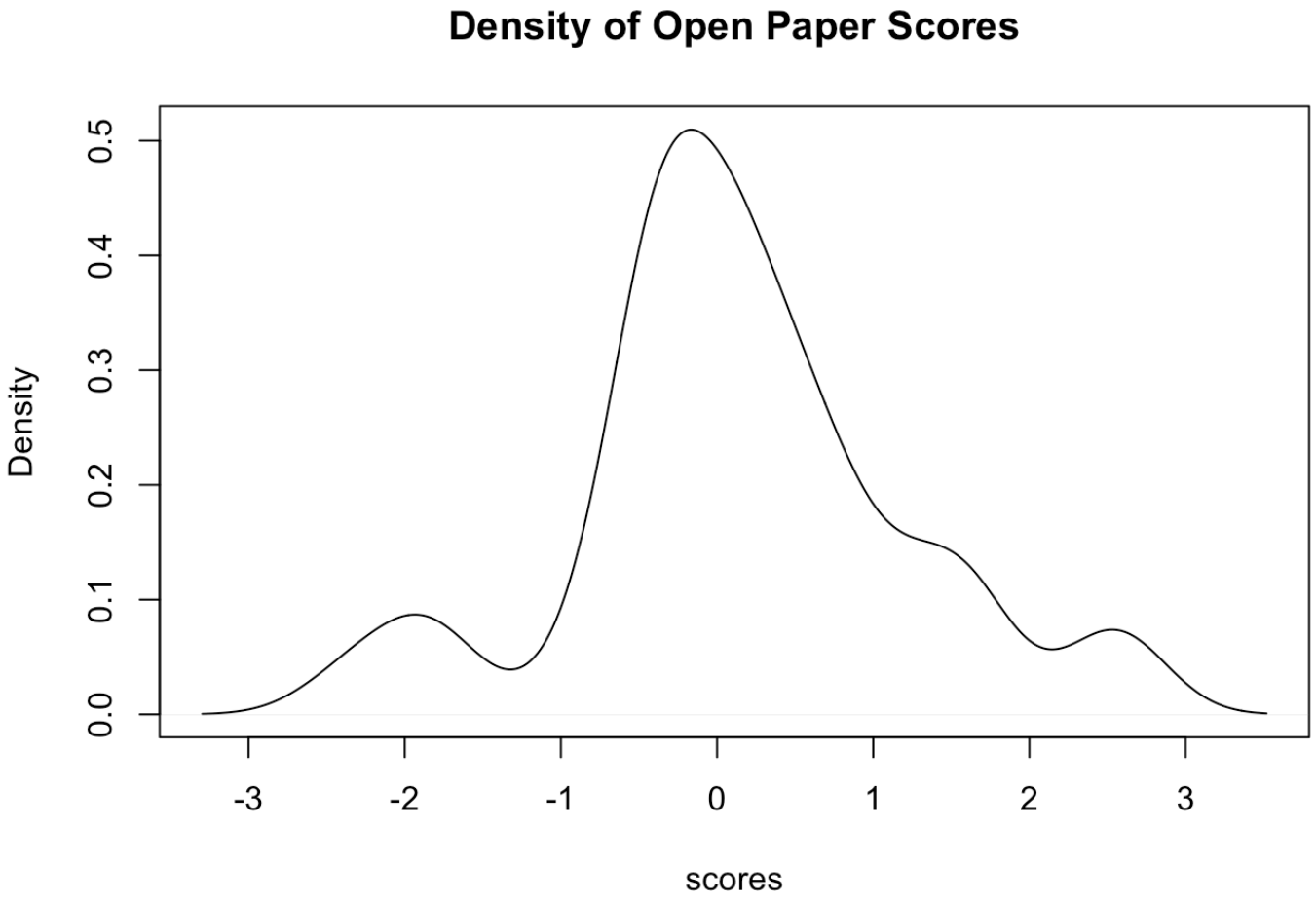
```
## [1] -0.4221697
```

```
#Upper
xbarlaptop + q * slaptop / sqrt(nlaptop)
```

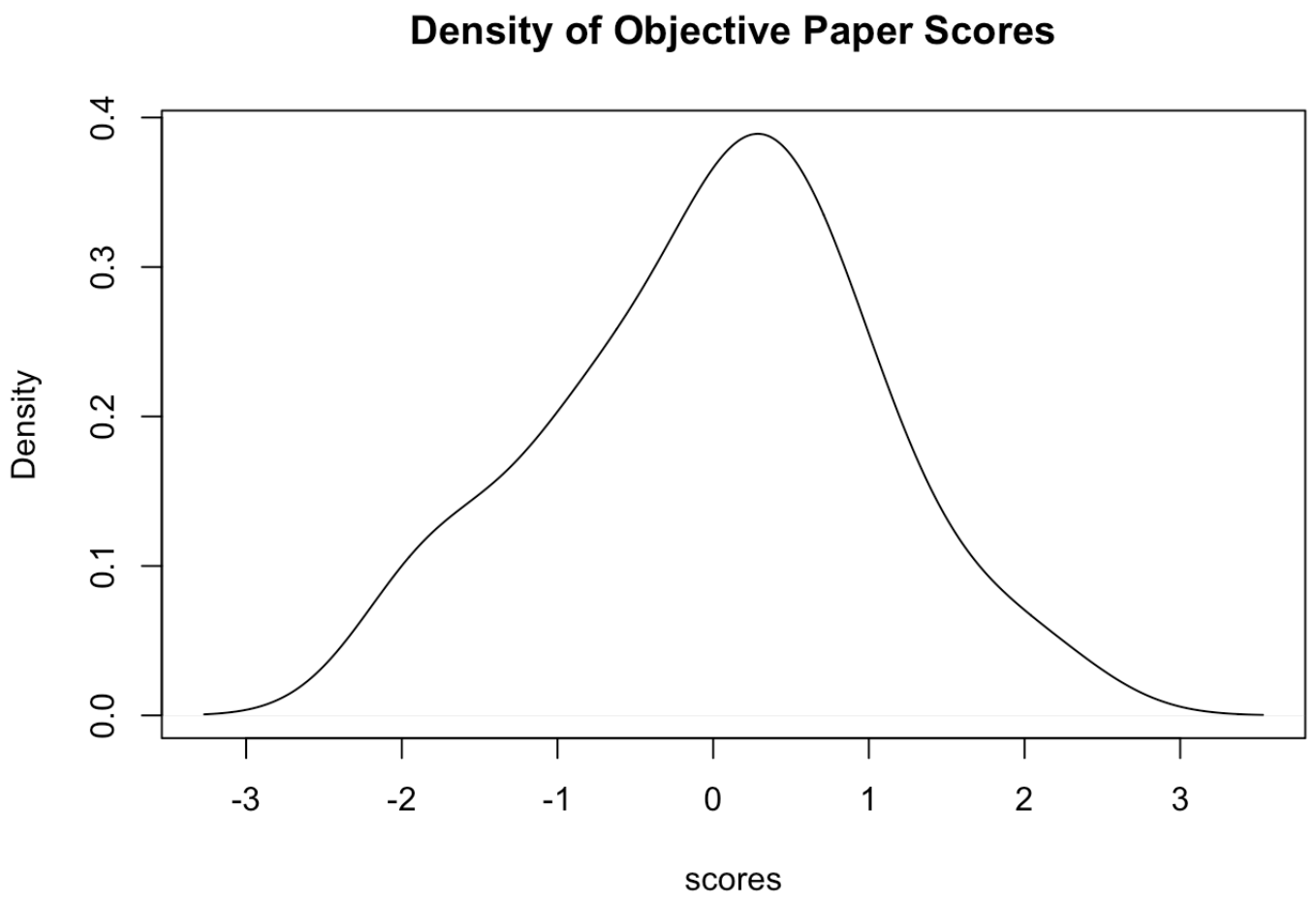
```
## [1] 0.1098406
```

part e:

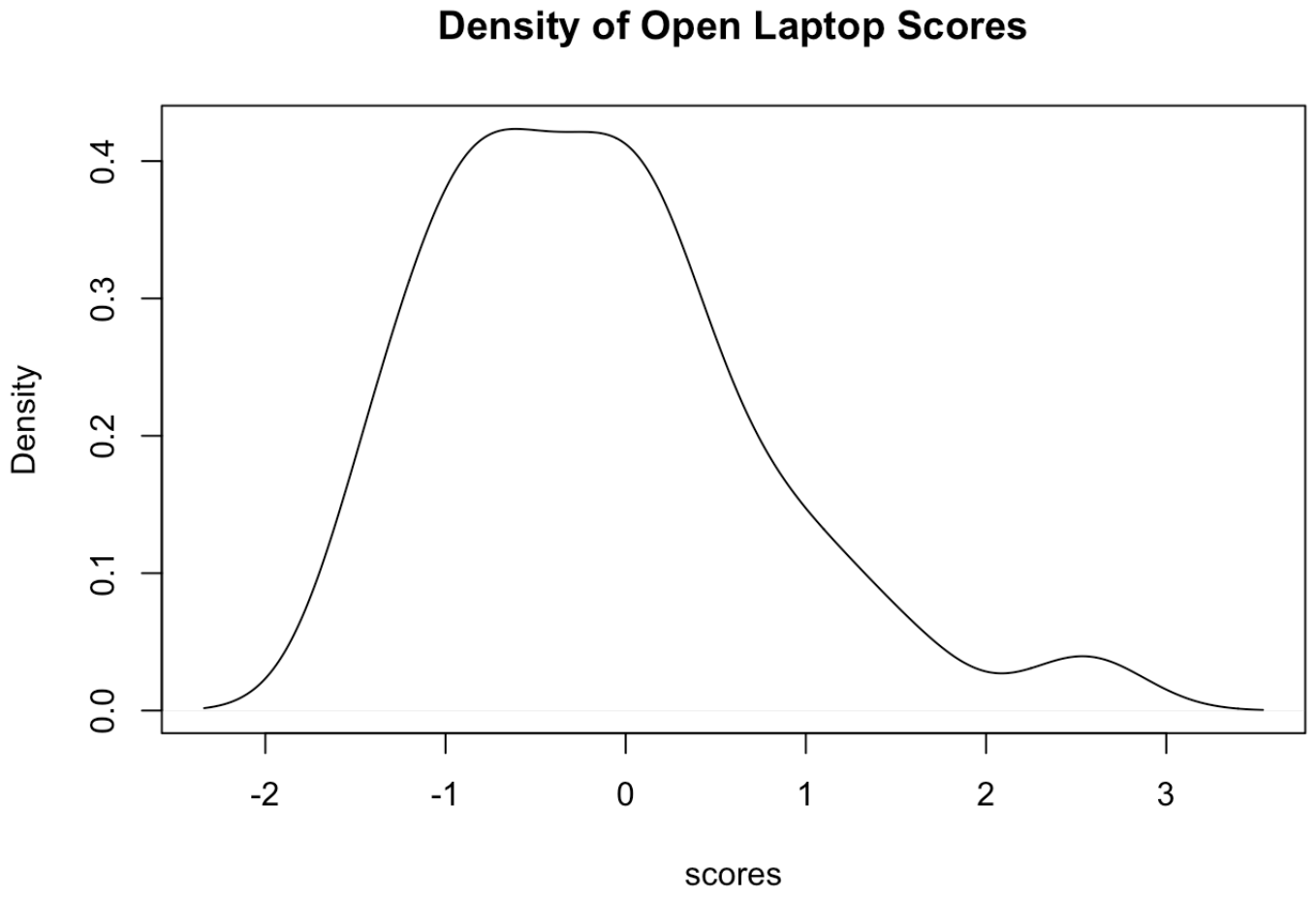
```
plot(density(paper$openZ), main = "Density of Open Paper Scores",xlab = "scores")
```



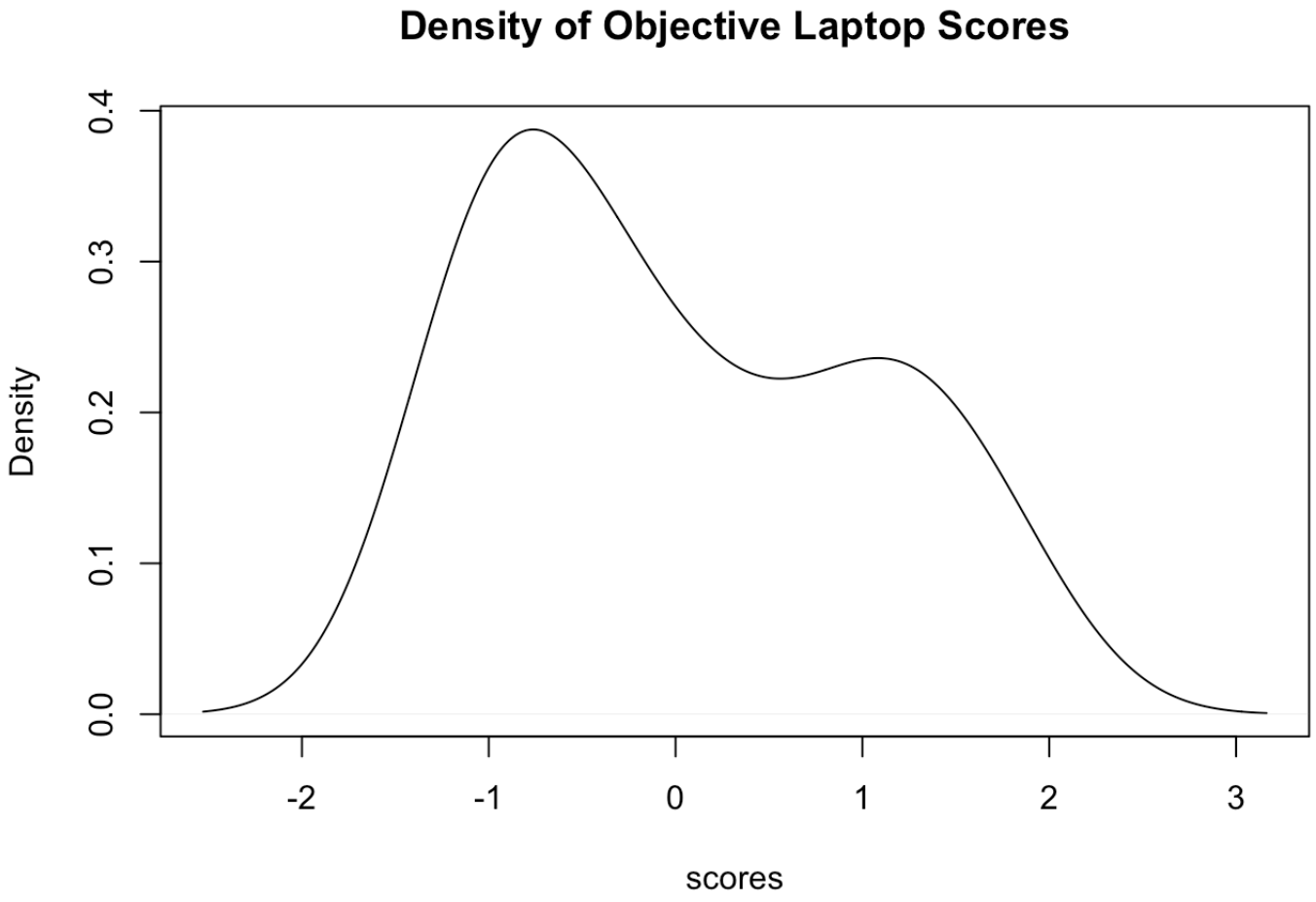

```
plot(density(paper$objectiveZ), main = "Density of Objective Paper Scores",xlab = "sc
ores")
```



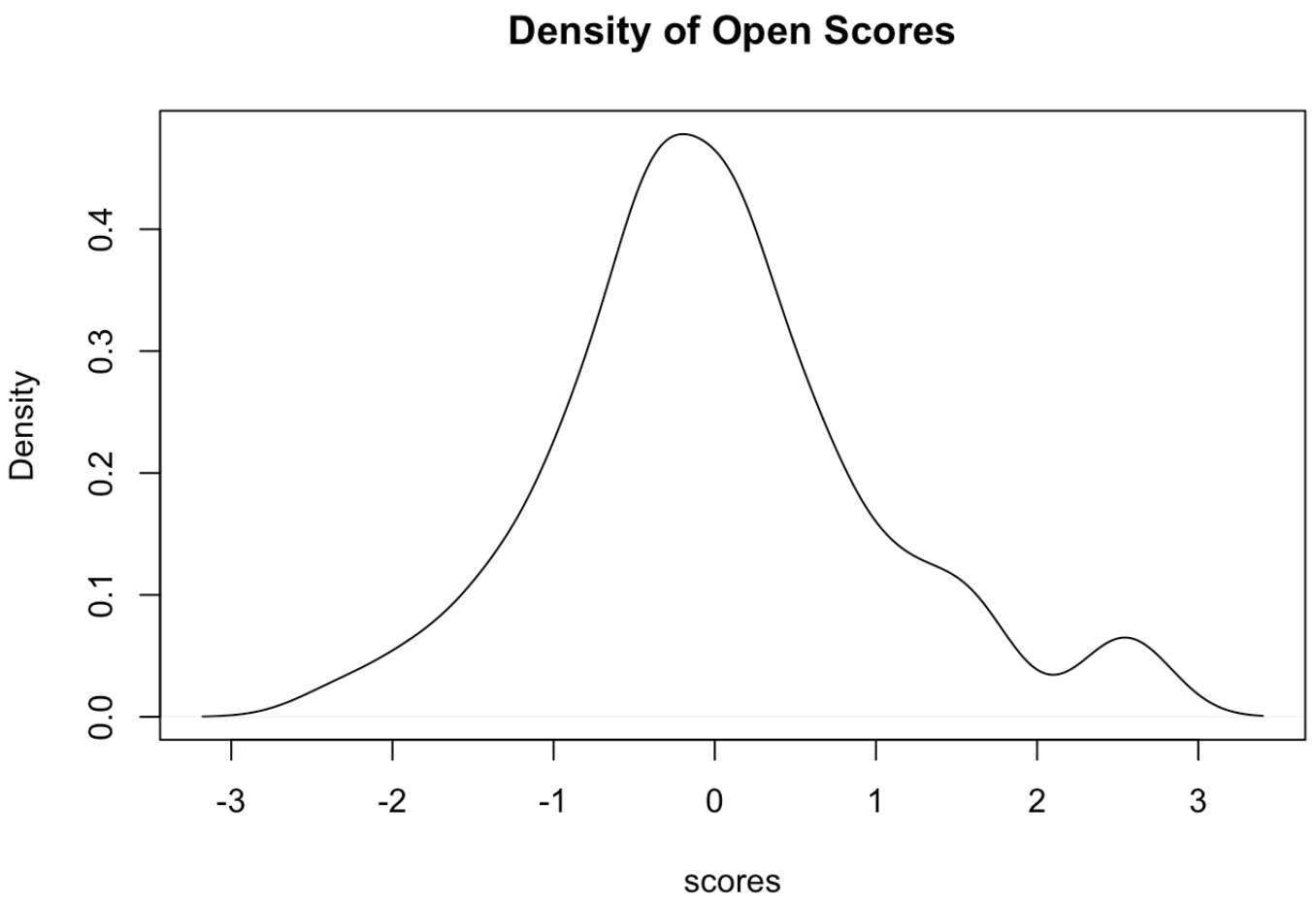

```
plot(density(laptop$openZ), main = "Density of Open Laptop Scores",xlab = "scores")
```



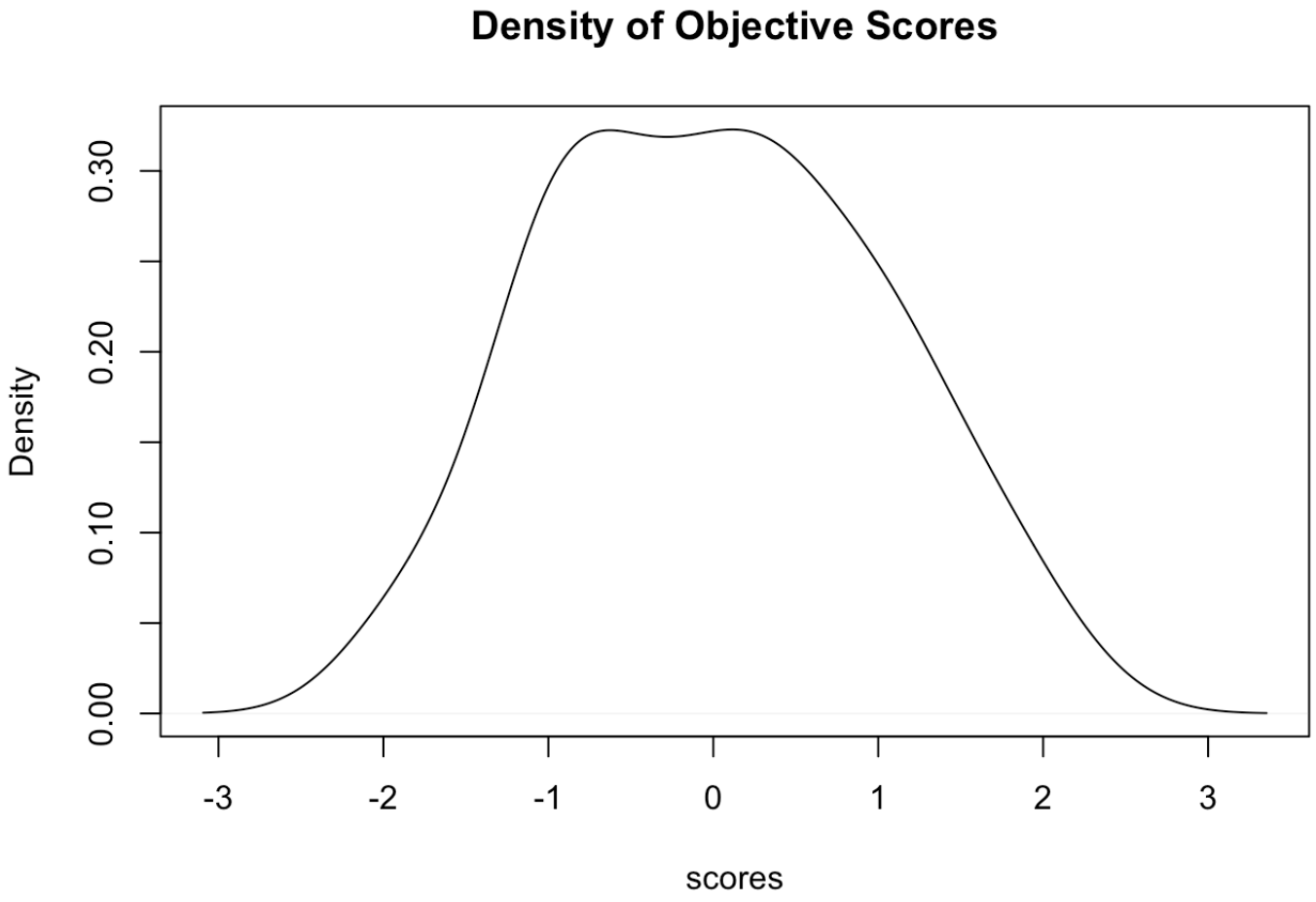

```
plot(density(laptop$objectiveZ), main = "Density of Objective Laptop Scores",xlab = "
scores")
```




```
plot(density(lap$openZ), main = "Density of Open Scores",xlab = "scores")
```




```
plot(density(lap$objectiveZ), main = "Density of Objective Scores",xlab = "scores")
```



There is relative relationship between the objectiveZ and openZ scores, mainly where their is the most equivalent scores.

part f:

The study could have had all students watch one video that had many different subjects and different questions correlating to each.

part g:

I do not think that professors should discourage the use of laptops in class. Laptops are very helpful tools in any class increasing productivity. I personally write notes better on a laptop and helps me have all my documents in one place allowing me to refer back very simple. What professors should instead discourage is the misuse of laptops as distractions.