Valeria Duran

Math 6359

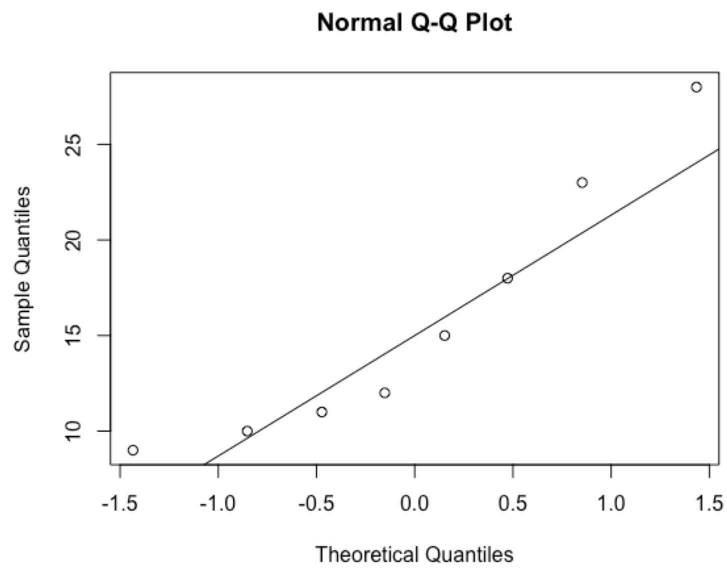
Exam 1 Take Home

Problem 1.

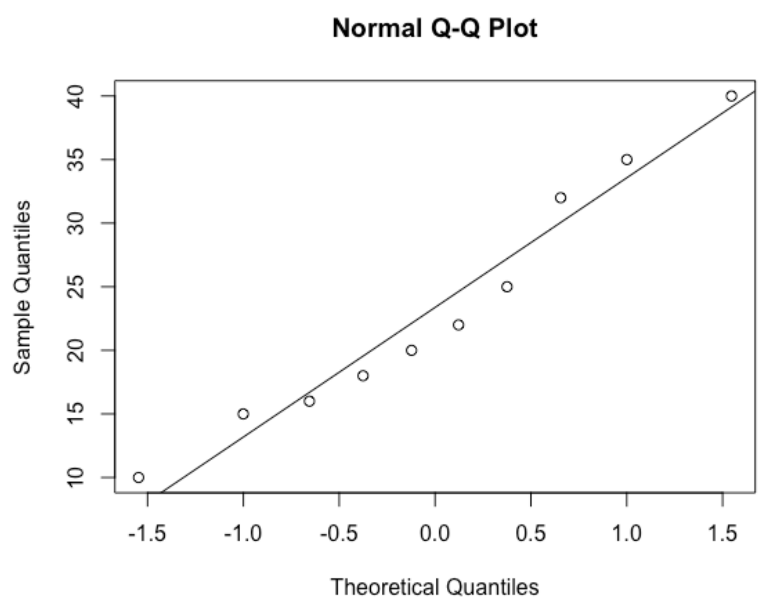
a) > x <- c(15, 23, 12, 18, 9, 28, 11,10)
> y <- c(25, 20, 35, 15, 40, 16, 10, 22, 18, 32)
> #part a
> mean(x) #mean is 15.75
[1] 15.75
> mean(y) #mean is 23.3
[1] 23.3
> var(x) #sample variance is 46.21429
[1] 46.21429
> var(y) #sample variance is 92.67778
[1] 92.67778
> #estimator is var(x)/var(y)
> var(x)/var(y) #0.4986555
[1] 0.4986555

b) > #part b
> #chi squared CI 95%
> qf(1-.025,7,9) #4.197047
[1] 4.197047
> qf(1-.025,9,7) #4.823217
[1] 4.823217
> lcl <- var(x)/var(y) / qf(1-.025,7,9) #0.1188111
> ucl <- var(x)/var(y) * qf(1-.025,9,7) #2.405124
> c(lcl, ucl)
[1] 0.1188111 2.4051238

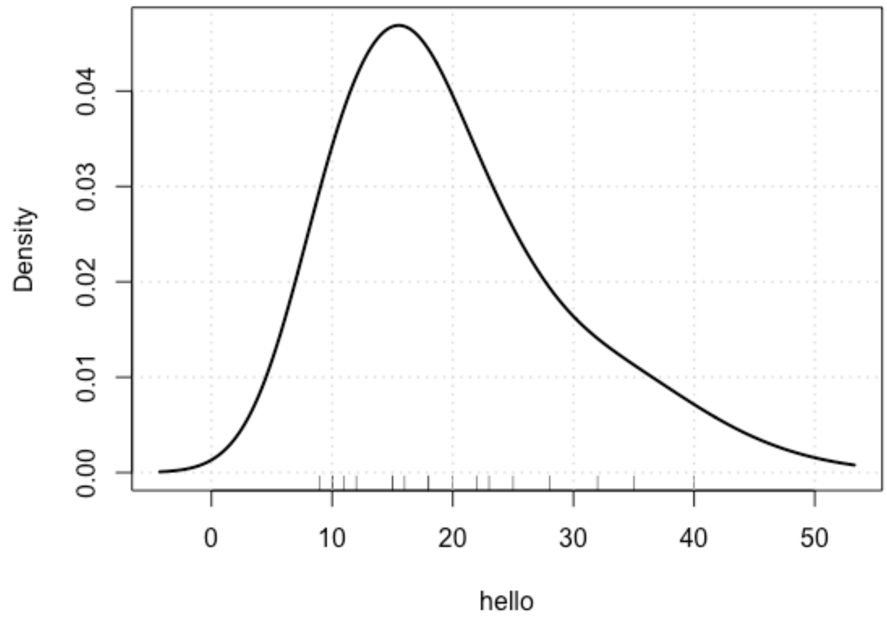
c) > #part c
> qqnorm(x)
> qqline(x)



> qqnorm(y)
> qqline(y)



> densityPlot(hello)

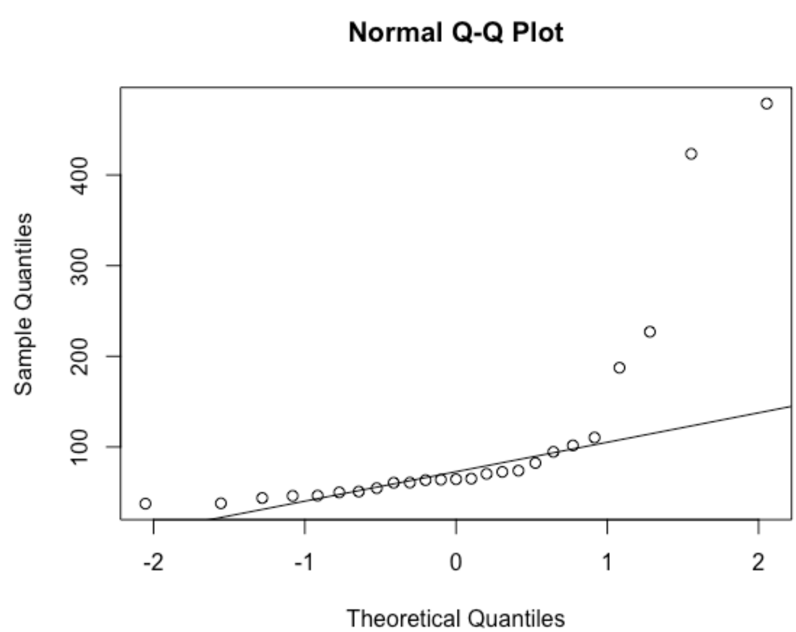


We can assume a normal distribution based on the plots. We can also assume the variances are the same, and the data comes from an SRS.

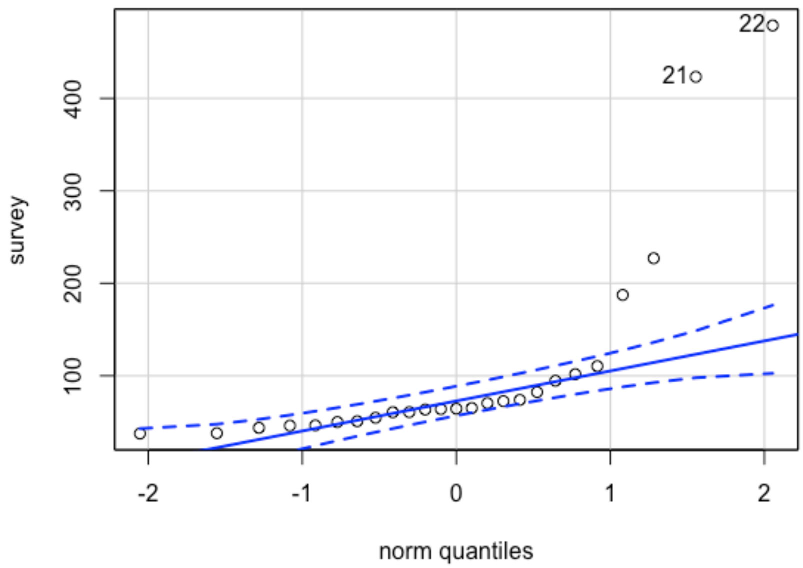
Problem 2.

a) H0: the sample mean is 63.688.
Ha: the sample mean is greater than 63.688

b) > #part b
> qqnorm(survey)
> qqline(survey)

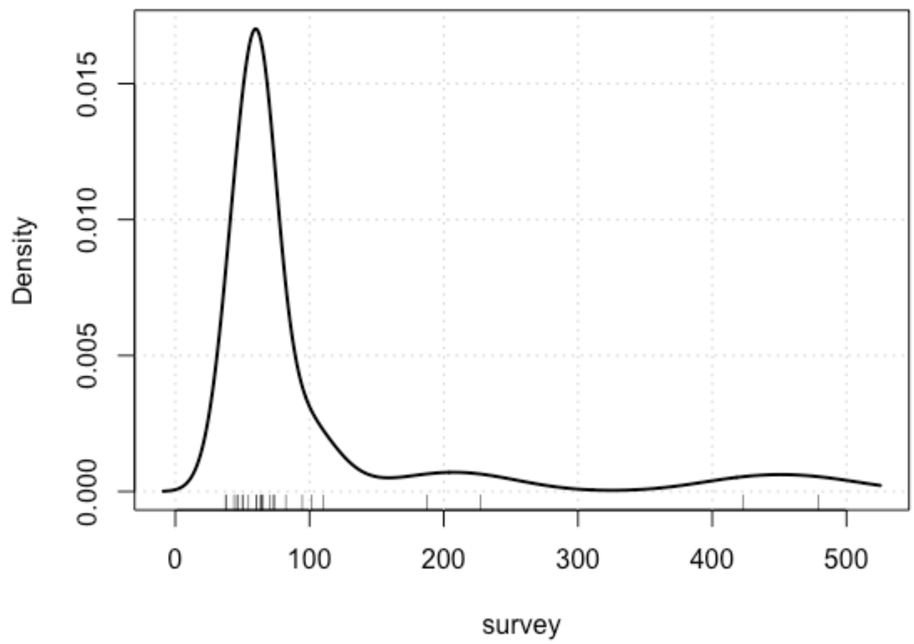


> qqPlot(survey)



There are outliers making the data non normal (cannot assume a normal distribution). We have to use a non-parametric test in order to test the hypothesis.

> densityPlot(survey)



Since data is skewed, must use sign test since Wilcox test can only be used on symmetric data.

c) > SIGN.test(survey, mu= 63.688, alternative = "greater")

One-sample Sign-Test

data: survey
s = 25, p-value = 2.98e-08
alternative hypothesis: true median is greater than 0
95 percent confidence interval:
59.8109 Inf
sample estimates:
median of x
64.41

Achieved and Interpolated Confidence Intervals:

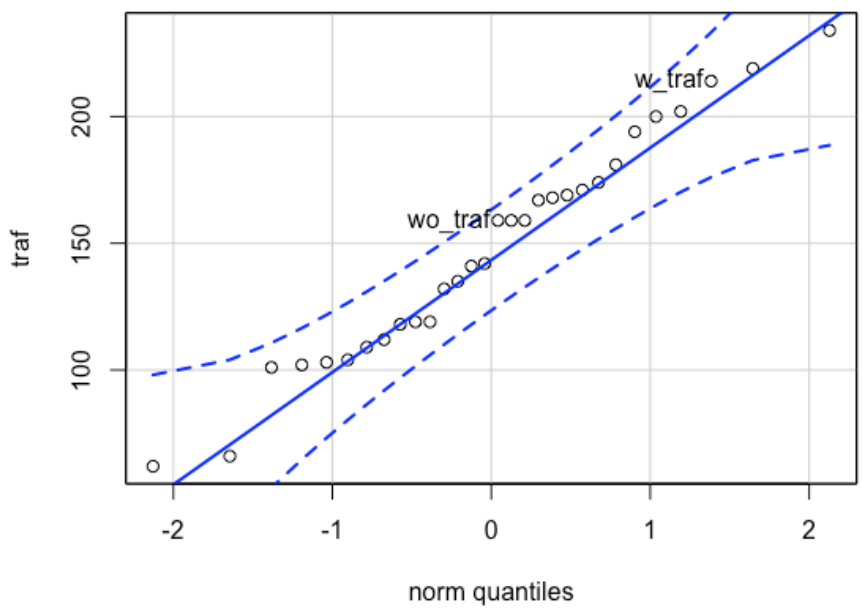
	Conf.Level	L.E.pt	U.E.pt
Lower Achieved CI	0.9461	60.5300	Inf
Interpolated CI	0.9500	59.8109	Inf
Upper Achieved CI	0.9784	54.5500	Inf

Since the p-value is < 0.05, we reject the null hypothesis and conclude that the sample mean is greater than 63.688.

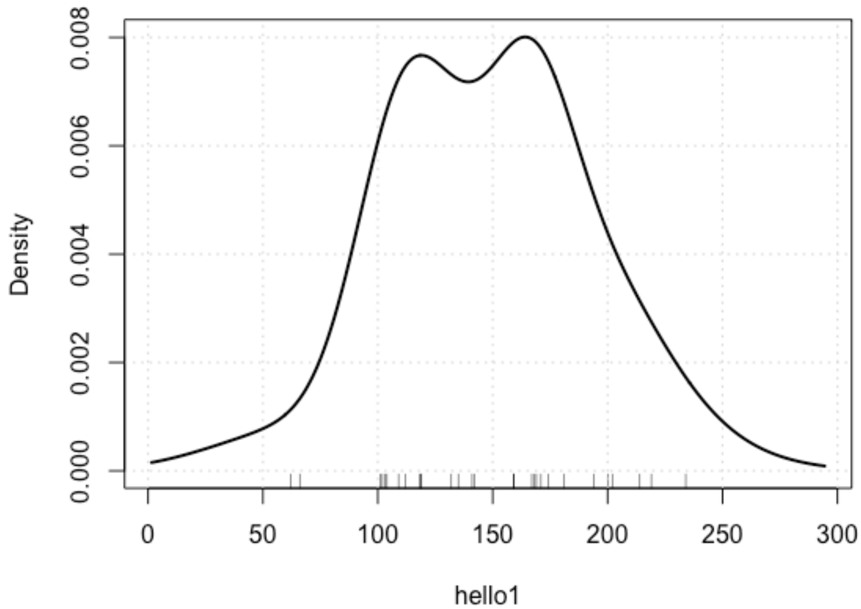
Problem 3.

a) This is a paired sample data since the pollution levels are being tested in the same streets (before and after road closure).
b) H0: there is not a difference between pollution levels by closing streets to car traffic.
Ha: there is a difference between pollution levels by closing streets to car traffic

> qqPlot(traf)



> densityPlot(hello1)



We can assume it follows a normal distribution, so use a paired t-test.

```
> t.test(traf$w_traf, traf$wo_traf, alternative = "two.sided", paired = TRUE)

        Paired t-test

data:  traf$w_traf and traf$wo_traf
t = 1.5128, df = 14, p-value = 0.1526
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.22361  64.95694
sample estimates:
mean of the differences
               26.86667
```

Test statistic is t = 1.5128 and p-value is 0.1526.

c) > abs(qnorm(0.1526/2)) #0.512673
[1] 1.430408
> 1.430408 / sqrt(15) #0.7204002
[1] 0.3693298

There is a moderate effect size from this test.

d) > #part d
> lcl <- 26.86667 - qt(1-0.05/2,14)* 68.78213/ sqrt(15) #lcl
> ucl <- 26.86667 + qt(1-0.05/2,14)* 68.78213/ sqrt(15) #ucl

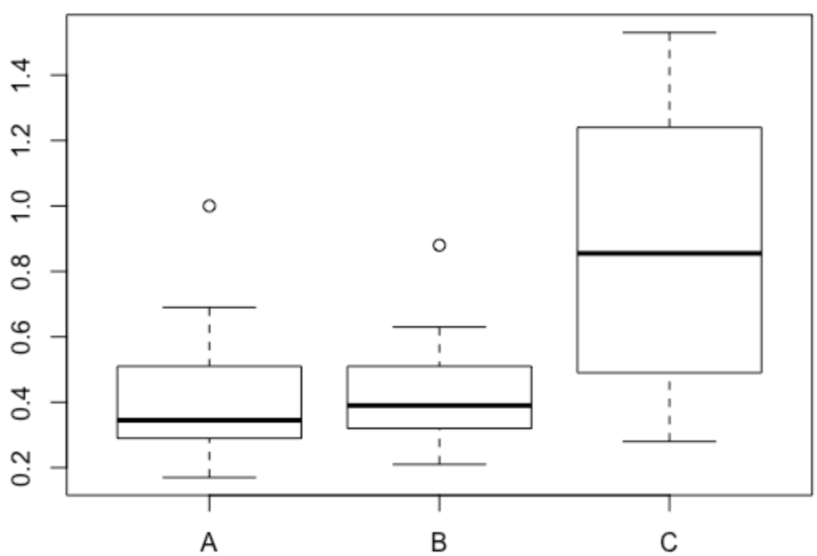
```
> c(lcl, ucl)
[1] -11.22360  64.95694
```

Since p-value >0.05, and CI includes 0 at 95% level, we fail to reject null hypothesis and conclude that there is no difference between pollution levels.

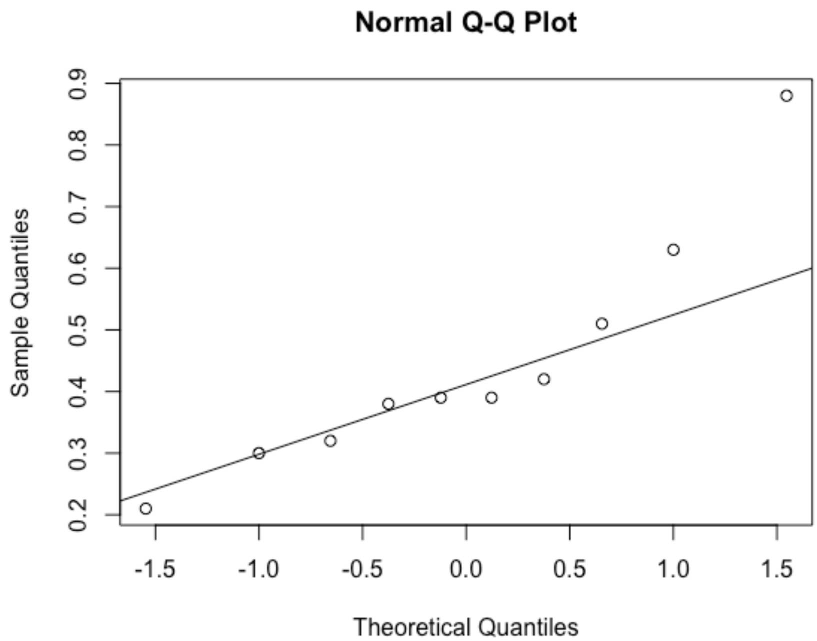
Problem 4.

a) H0: There is no anesthetic effect on concentration
 Ha: There is anesthetic effect on concentration

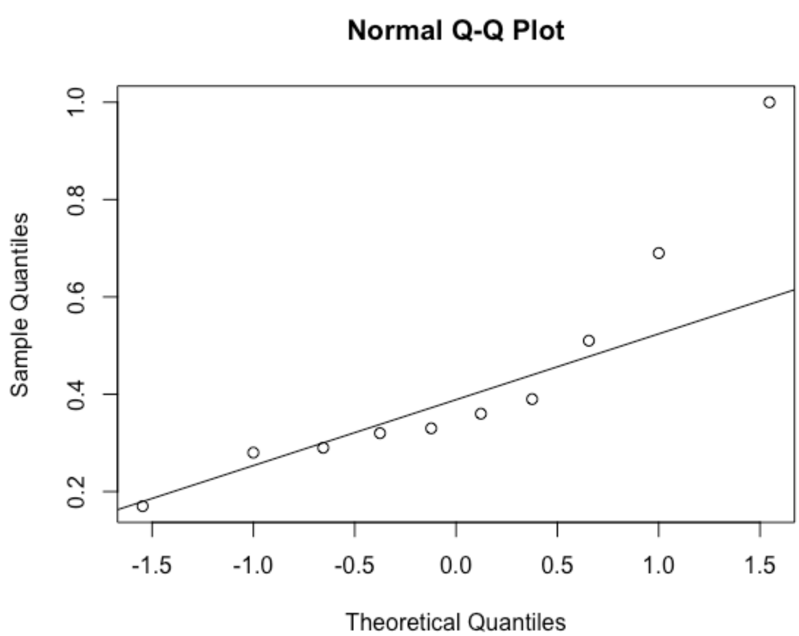
```
> boxplot(conc$Insoflurane, conc$Halothane, conc$Cyclopropane, names = c("A", "B", "C"))
```



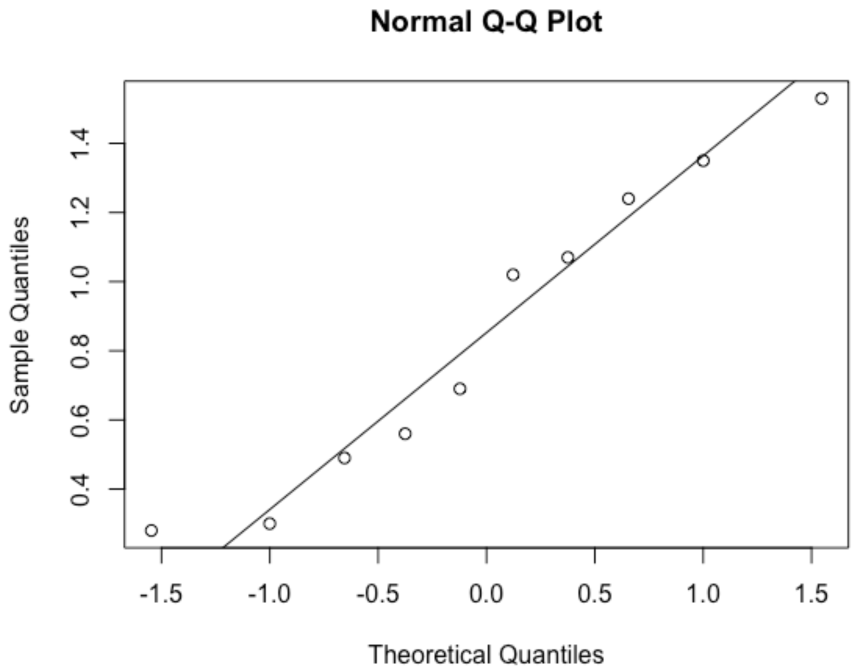
```
>qqnorm(conc$Insoflurane)
> qqline(conc$Insoflurane)
```



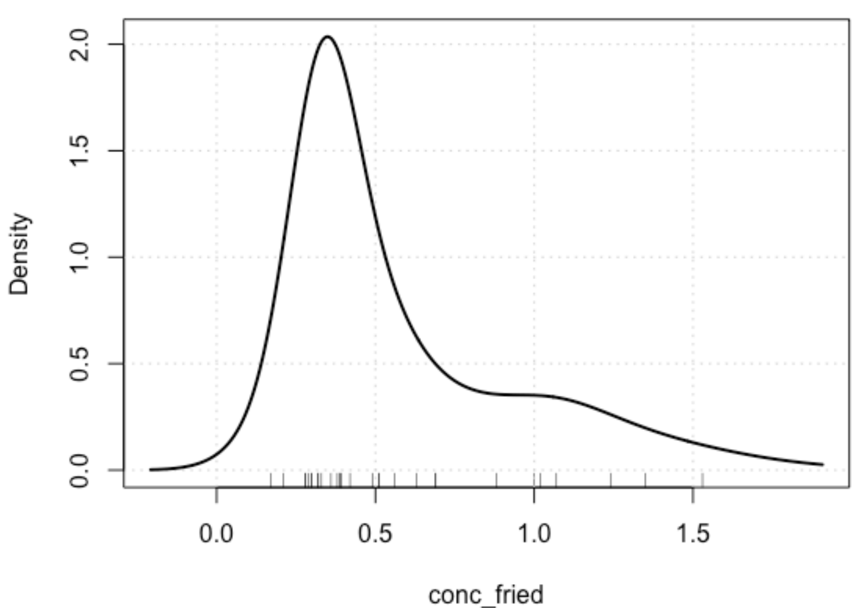

```
> qqnorm(conc$Halothane)
> qqline(conc$Halothane)
```



>qqnorm(conc$Cyclopropane)
> qqline(conc$Cyclopropane)



> densityPlot(conc_fried)



The data does not seem to follow a normal distribution (major outliers in Insoflurane and Halothane).

We can’t do ANOVA since it does not follow a normal distribution. We must use the Friedman test, since it is the non-parametric version.

```
> conc_fried <- as.matrix(conc[,-1]) #drop the dog factor
> friedman.test(conc_fried)

        Friedman rank sum test

data:  conc_fried
Friedman chi-squared = 2.6, df = 2, p-value = 0.2725
```

p-value is > 0.05, therefore we fail to reject null hypothesis. There is no anesthetic effect on concentration.

b) Since there is no anesthetic effect on concentration, none of the anesthetic treatments affects concentration differently.

Problem 5.

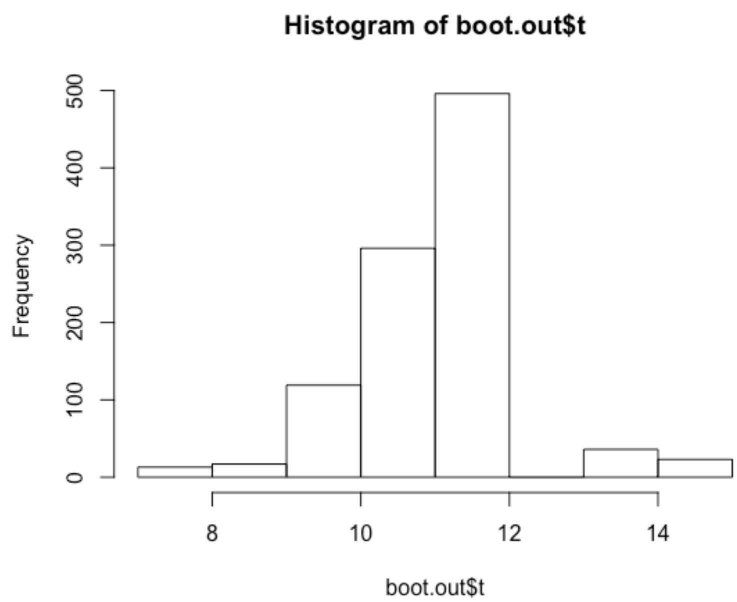
```
a) > #part a
> library(boot)
> median.fun <- function(dat, idx) median(dat[idx], na.rm = TRUE)
> boot.out = boot(data = dat,statistic = median.fun,R = 1000)
> boot.out

ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = dat, statistic = median.fun, R = 1000)


Bootstrap Statistics :
   original   bias    std. error
t1*    11.25 -0.03225     1.16555
```

```
b) > boot.ci(boot.out, type = "norm")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates
```

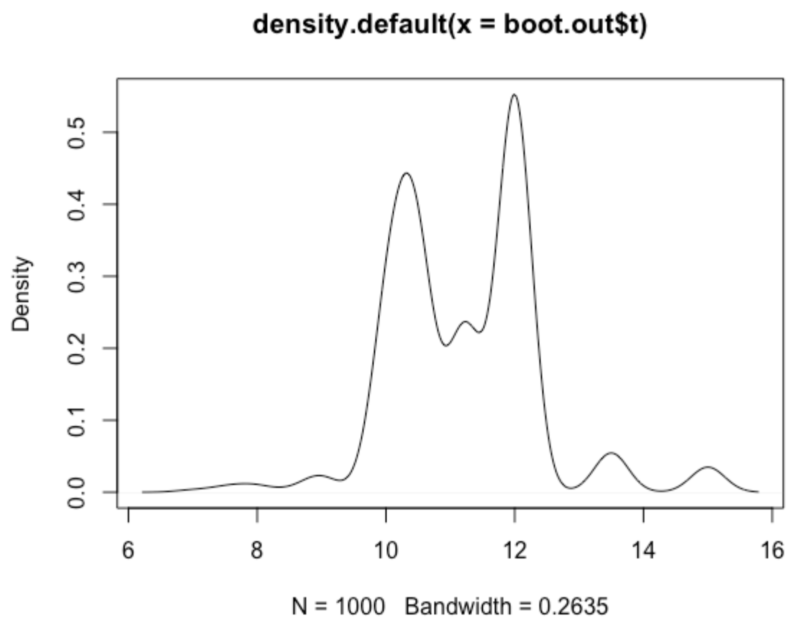
CALL :
boot.ci(boot.out = boot.out, type = "norm")

Intervals :
Level   Normal
95%   (9.00, 13.57)
Calculations and Intervals on Original Scale

c)> #part c
> hist(boot.out$t)



> plot(density(boot.out$t))



The distribution essentially only takes on a few values (majority of values are 10-12). The 95% CI in part b goes from (9.00, 13.57), which seems to be outside the majority range since majority of the values should be within (10, 12).

d) > boot.ci(boot.out, type = "perc")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

CALL :
boot.ci(boot.out = boot.out, type = "perc")

Intervals :
Level   Percentile
95%   (9.0, 13.5)
Calculations and Intervals on Original Scale

e) Since there are no extreme scores in the data, mean is the better estimate for central tendency in this case. Median would be better if there were major outliers, however, since the data seems to be densely centered on 11, then mean would be a better estimate.