Assignment 3

Yakup Tezcan

```
knitr::opts_chunk$set(echo = TRUE)
assn3 <- read.csv(file="~/assignment.csv")
```

Question 1

A

Since we want to estimate the mean score of Science and Technology students on Academic Dishonesty Scale, using a confidence interval for the mean would be the most appropriate method to use.

B

First of all, we need to evaluate the Randomization of our study.Our sample has to be a Simple Random Sample(SRS), however the data was collected from those who agreed to participate after seeing the post on the websites of twenty-eight Polish universities.This means that online questionnaires prepared by researchers are filled by participants with higher or lower motives.(voluntary response bias) Also, a post on the website of a university is more likely to be skipped by majority of students which is an undercoverage in this case.Therefore, the sample is not gathered by using randomization.

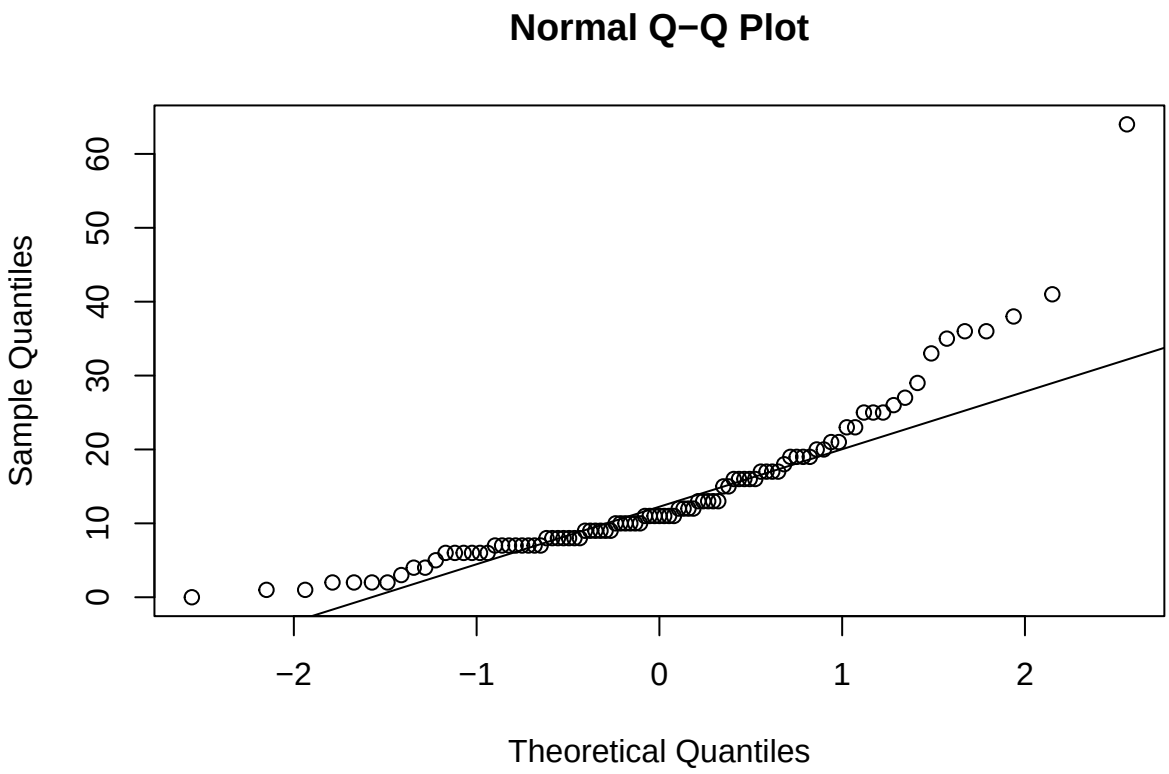
Secondly, we need to determine whether the sample size is no more than 10% of the population of interest or not, since the sample is drawn without replacement. Population of interest in our study is the students in the world, so our sample is no more than 10% of the population of interest. Therefore, we can conclude that our sample satisfies this condition.

```
#Shows the number of participants. (Sample size)
#n=390
max(assn3$participant)
```

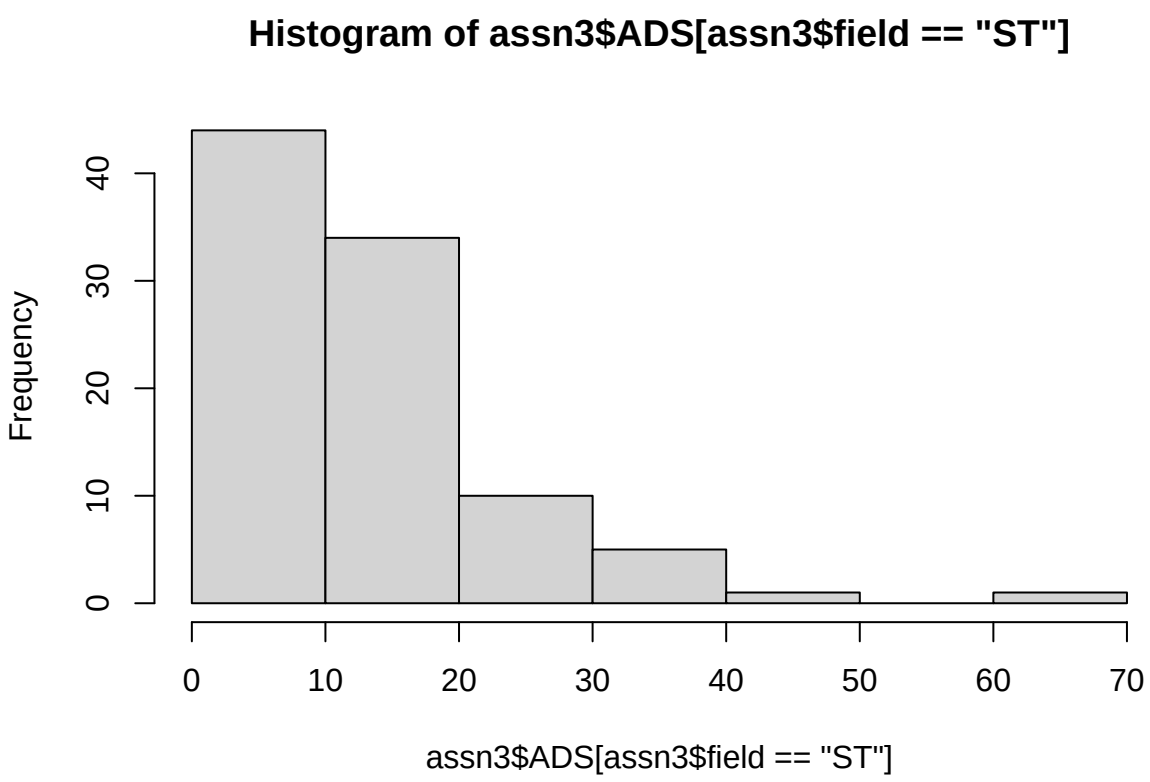
```
## [1] 390
```

Moreover, we need to ensure that the sample size is sufficiently large. As we have the knowledge of sample size, n=390, this condition is validated.

```
qqnorm(assn3$ADS[assn3$field=="ST"])
qqline(assn3$ADS[assn3$field=="ST"])
```




```
hist(assn3$ADS[assn3$field=="ST"])
```



As it can be noticed from the QQ plots and histogram(Right-skewed) above, the distribution does not seem to be Normal which is another condition that we need to confirm to build confidence intervals. However, since the sample is sufficiently large, we can use confidence intervals.

Finally, to build a confidence interval we need our data to be independent. If we used a random sampling method while collecting the data, each observation in the data would be independent of every other observation. However, a random sampling method is not used during the data collection process, so it can be said that the data does not meet the independence requirements for building a confidence interval.

Question 2

A

Since we want to estimate the mean score of Science and Technology students on Academic Dishonesty Scale, using a confidence interval for the mean would be the most appropriate method to use.

B

```
#Builds one sample t-test
t.test(x=assn3$ADS[assn3$field=="ST"], conf.level = 0.92)
```

```
##
##  One Sample t-test
##
## data:  assn3$ADS[assn3$field == "ST"]
## t = 13.21, df = 94, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 92 percent confidence interval:
##  12.01489 15.73248
## sample estimates:
## mean of x
##  13.87368
```

```
#Shows confidence interval
confint(lm(assn3$ADS[assn3$field=="ST"]~1), level=0.92)
```

```
##                   4 %     96 %
## (Intercept) 12.01489 15.73248
```

T-test for one sample of Academic Dishonesty Scale score of Science and Technology students. 92% confidence interval is used with 94 degrees of freedom. Mean score is 13.87.

C

The mean score of Science and Technology students on the Academic Dishonesty Scale is 12.01 - 15.73(with %92 confidence, df=94).