**BIOS 543 GRADUATE RESEARCH METHODS I**

**HOMEWORK 2**

**DUE TUESDAY, OCTOBER 4, 2022**

Directions: Your completed homework is due at the beginning of class on the date listed above. Compose your responses to the problems listed below electronically (e.g., Microsoft Word) and print. This assignment is subject to the VCU honor code: you may collaborate with other students, but your homework must be the product of your own efforts and understanding.

The data for this assignment are stored in an excel file entitled “HW2\_Data.csv”, which contains measures from a study comparing health outcomes between cancer patients treated with or without antithymocyte gobulin (ATG). Use the data within this file to conduct the following:

1. Say we have the following research question: is mean CD34 count greater than 2?

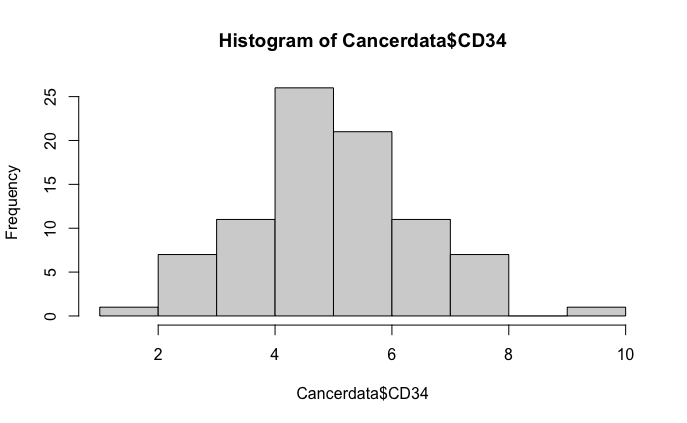
**Mean CD34 = 4.998471**

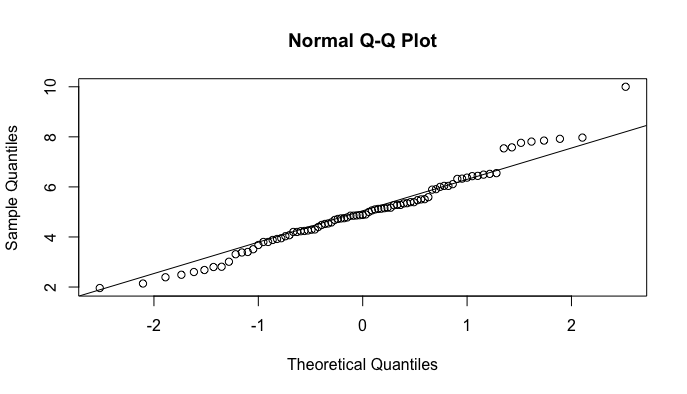
* 1. Translate this research question into a set of testable hypotheses in symbolic form.

**H0: Mean CD34 2**

**H1: Mean CD34 > 2**

1. Assess the assumptions needed to test this hypothesis.
   1. Visualize the distribution of CD34 values.





* 1. Comment on whether the assumptions are satisfied.

**H0 False**

**H1 True**

* 1. Based on your assessment, state unequivocally which test you are conducting.

**mean(Cancerdata$CD34) > 2**

**result: TRUE**

* 1. State which functions in R/RStudio you’re using for each task.

**rm(list=ls())**

**setwd("~/Documents/Ehsan/homework 2/2 oct")**

**getwd()**

**Cancerdata <- read.csv("HW2\_Data.csv")**

**str(Cancerdata)**

**mean(Cancerdata$CD34)**

**mean(Cancerdata$CD34) > 2**

**hist(Cancerdata$CD34)**

**#assessing the normality of the column CD34:**

**summary(Cancerdata$CD34)**

**class(Cancerdata$CD34)**

**# normality test:**

**hist(Cancerdata$CD34)**

**qqnorm(Cancerdata$CD34)**

**qqline(Cancerdata$CD34)**

**#mean and Standard deviation:**

**mean(Cancerdata$CD34)**

**sd(Cancerdata$CD34)**

1. Summarize the variable CD34.
   1. Use appropriate measures of center and variability.

**The CD34 is distributed normally.**

**Mean SD: 5.00 1.48**

**Sample size: 85**

* 1. Calculate a 95% confidence interval.

**Mean 1.96\*SD: [ (5.00- (1.96\*1.48)) , (5.00+(1.96\*1.48))] = [2.10 , 7.90]**

1. Perform the hypothesis test.
   1. Report the relevant statistics from performing that hypothesis test.

**H1 is accepted. Hence, the mean of CD34 is higher than 2 with a 95%confidence interval. (95%CI: 2.10 -7.90)**

* 1. State your conclusion from these findings.

**In conclusion, I would say that the mean of CD34 was higher than 2 in this population. (95%CI: 2.10 -7.90)**

1. Extra Credit
   1. Calculate the association (and 95% confidence interval) between Age and CD34

**Introduction:** Numerous people all around the world are suffering from cancer. Our research will help us know if a certain treatment works for people with cancer. We performed a statistical analysis to compare the health outcomes of cancer patients who received antithymocyte gobulin (ATG) to the proportions of patients who did not.

**Methods:** The quantitative (CD34) levels are reported in this study for each racial group. The sample size for patients treated with ATG is n=57 and n=30 for non-treated with ATG.

The R statistical software was used for all measurement analysis.