1) A) T-test for the slope

B) i) A slope of zero is required for the higher education students who are employed.

ii) We are looking for $\beta_1 \neq 0$

C) Response variable observations are independent; the response of each student is independent of the other so that makes this point valid. Linear relationship between explanatory (X) and response (Y), we do see linearity in the scatter plot of employed vs test grades. The code used was: plot(body$workhours[body$employed !="no"], body$testgrade[body$employed != "no"]). Response variable (Y) is Normally distributed for each X, with the amount of data that is available, we can conclude that the variable y is normally distributed since we see that the qqplot is linear; the r code used was:
>reg <- lm(body$workhours[body$employed !="no"]~body$testgrade[body$employed != "no"])
> summary(reg),
> qqnorm(reg$residuals)
> qqline(reg$residuals)
Constant standard deviation of response variable (and therefore, ε) for each X, we can conclude that the mean is zero with constant variance. The r code used was: plot(reg$residuals)

2) A) T-test for the slope

B) > reg <- lm(body$workhours[body$employed !="no"]~body$testgrade[body$employed != "no"])
> summary(reg)
The work hours of those who are employed is required, as a result, the code included → …!=”no” to eliminate those who don’t work.

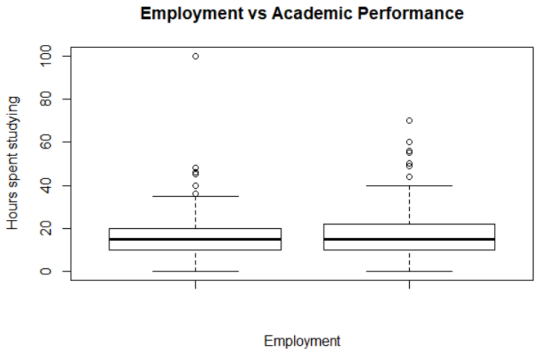
C) The slope of the linear relationship between time taken to complete the task In the simulation and time to complete the real task is -0.6420 (t= -3.992, df = 99, p = 0.000126 ; the relationship is described as: test grade = 15.7265 -0.6420*WorkHours (R-squared: 0.1386)

3) The data is subset using the r code body <- read.table("C:\\Users\\obaida\\Desktop\\Stats\\lab5data.csv",header = T, sep = ","). Then 4 new variables are made which are:

1. employed <- body [body $employed!="no",] → This code makes a new variable, employed but eliminates the responses of those who are not employed.
2. unemployed <- body [body $employed=="no",] → Similarly to the code above, this code makes a new variable but instead only takes the responses of those who are not employed.
3. study<-employed$studyhours → This code makes a variable to include the data points of those who are employed.
4. nostudy<-unemployed$studyhours → This code makes a variable to include the data points of those who are unemployed. This new variable and the one above are made to be compared in the T-test code.

B)

>boxplot(body$studyhours[body$employed!="no"],body$studyhours[body$employed=="no"],main ="Employment vs Academic Performance", xlab = "Employment",ylab= "Hours spent studying")



4A) Researchers investigated the correlation between employment and academic performance of higher education students. Researches focused on the results of the employed students' data to see the effects resulted from employment. Of the survey respondents, 437 responses were considered for this study. Of this number, it was observed that the students who are employed had a median of 14.8 test grade, mean of 14, max of 20 and finally, a minimum of 5.5. On the other hand, those who did not work, who represented 77% of the sample, had a median test grade of 15, mean of 14.4, max of 20 and finally a minimum of 4.5. As a result, those who were employed during the school year had lower test grades than those who did not work. In conclusion, it was observed that those who worked during the school year had lower test results than those who were not employed.