# Assignment 2

**The assignment is due Thursday May 19 in your tutorial session**
For the following questions, use the data file sleep75 from the package wooldridge. Restrict the data to married individuals with good health living in the Standard Metropolitan Statistical Area (SMSA). It means that you need 3 conditions in the subset() function. In order to know which variables to restrict, read the help file if you get the data from the wooldridge package.

All time variables are expresses in minutes per week. Just to avoid huge numbers when you will be computing variances, convert the variable "sleep" in hours per day. Here is how you do it:

```
data$sleep <- data$sleep/60/7
```

a) Print the summary statistics of age, educ, male and sleep using stargazer, and plot the histogram of sleep. Using the table and graph, what can you tell about the sleep habits in our sample? Does it look normal?

b) Using the results from the table, construct a 95% confidence interval for the average number of hours of sleep per day. Compare the intervals using the t-distribution and the normal. Interpret. (Do it without using the t.test() function)

c) Test the hypothesis that the average number of hours of sleep per day is 8 against the alternative that it is less than 8, at 5%. Use the critical value and p-value approach, based on the t-distribution. Do it manually and verify your result with t.test()

d) Create a dataset "male" for males (male=1) and a dataset "female" for females, using the subset() function. Print the summary statistics of the same four variables that you showed in question a) for the two groups using stargazer. Interpret the difference.

e) Test the hypothesis that the variance of sleep for the two groups are the same at 5%. Use the critical value and p-value approach.

f) Test the hypothesis that males and females sleep the same number of hours per week, against the alternative that females sleep more. Do it manually using the appropriate formula for the standard error based on the result of the previous question. Verify with t.test (results may not be identical). (use the t-distribution)

g) Using the whole sample (sample for males and females), test the hypothesis that the variance of sleep is 1 against the alternative that it is greater than 1, at 5%

h) Using the whole sample, construct a 95% confidence interval for the variance of sleep.

i) We saw in class that the test (questions g and h)) on the variance is only valid if the observations are normally distributed. If not, we can use the expression:

$$test = \frac{S^2 - 1}{SD_{s^2}},$$

which is N(0,1) in large samples. We can also use the same formula to construct a 95% confidence intervals:

$$\left[S^2 - 1.96 \times SD_{s^2}, \quad S^2 + 1.96 \times SD_{s^2}\right]$$

where 1.96 is the 97.5% quantile of the N(0,1). Use the bootstrap method to get $SD_{s^2}$, perform the test that the variance is 1 against the alternative that it is greater at 5% and construct a

95% confidence interval for the variance. Use 1000 samples. This method works in large samples (it approaches the N(0,1) when n goes to infinity). Do you think the sample size is large enough here?