SOC510 Data Analysis Homework 1
Due October 12 (Thursday)

To get the full credit (30 points), submit the following two documents:
(1) your "typed" and printed answers including graphs, and
(2) a hardcopy of your script file.

A. Using the given data, estimated the following:

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Years of Education (x)	8	10	14	3	9	12	12	16	16	6	11	15	18	22	16
Hourly Wage (y)	5	7	8	5	7	10	9	17	15	9	11	15	18	25	10
ID	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Years of Education (x)	7	9	14	5	10	11	13	17	15	5	12	14	17	21	18
Hourly Wage (y)	5	7	8	5	8	10	9	18	14	9	10	16	18	25	10

1. Compute mean, median, variance, and standard deviation of x and y
2. Compute SS(x); SS(y); SS(xy)
3. Draw a boxplot of of x and y
4. Calculate the correlation coefficient between x and y
5. Do a regression analysis between x and y
6. Interpret the regression result.
7. Draw a scatterplot between x and y with the estimated regression line.

B. Import "soc510hw1.csv" into R and do the following analysis. Note that the dataset, soc510hw1.csv, can be downloaded from our course website.

Below are the descriptions of the variables in the data set:

- `wage`: hourly wage
- `age`: Age 16 or older
- `educ`: 1. Less than high school; 2, High school graduate; 3, Some college; 4. Bachelor degree; 5, Graduate degree

1. Compute mean and standard deviation of `wage`, `age`, and `educ`
2. Compute Q1, Q2, and Q3 of `wage`, `age`, and `educ`
3. Draw a histogram of `wage` with 30 bins
   (a) Add a line of normal curve to the histogram.
   (b) Add a line of mean and a line of median.

(c) Describe the shape of the distribution of `wage`

4. Draw a boxplot of `wage`, `age`, and `educ`
5. Calculate the correlation coefficients of `wage` and `age`
6. Calculate the correlation coefficients of `income` and `educ`
7. Compare $\rho_{\texttt{wage, age}}$ and $\rho_{\texttt{wage, educ}}$. (Note that $\rho_{x,y}$ refers to the linear correlation coefficient between x and y.)
8. Do a regression analysis between `wage` and `age` and interpret the result.
9. Do a regression analysis between `wage` and `educ` and interpret the result.
10. Draw a scatterplot between `wage` and `educ` with the estimated regression line.

C. Standard normal distribution

1. Compute $P(-.83235 < z < 1.21532)$
2. Compute $P(2.5321 < z)$
3. Compute $P(z < -1.6523583)$