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:

<table>
<thead>
<tr>
<th>ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Years of Education (x)</td>
<td>8</td>
<td>10</td>
<td>14</td>
<td>3</td>
<td>9</td>
<td>12</td>
<td>12</td>
<td>16</td>
<td>16</td>
<td>6</td>
<td>11</td>
<td>15</td>
<td>18</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>Hours of Wage (y)</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>17</td>
<td>15</td>
<td>9</td>
<td>11</td>
<td>15</td>
<td>18</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Years of Education (x)</td>
<td>7</td>
<td>9</td>
<td>14</td>
<td>5</td>
<td>10</td>
<td>11</td>
<td>13</td>
<td>17</td>
<td>15</td>
<td>5</td>
<td>12</td>
<td>14</td>
<td>17</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>Hours of Wage (y)</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>18</td>
<td>14</td>
<td>9</td>
<td>10</td>
<td>16</td>
<td>18</td>
<td>25</td>
<td>10</td>
</tr>
</tbody>
</table>

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 \( x \) 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:

- \texttt{wage}: hourly wage
- \texttt{age}: Age 16 or older
- \texttt{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 \texttt{wage}, \texttt{age}, and \texttt{educ}
2. Compute Q1, Q2, and Q3 of \texttt{wage}, \texttt{age}, and \texttt{educ}
3. Draw a histogram of \texttt{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 \( \rho_{\text{wage, age}} \) and \( \rho_{\text{wage, educ}} \). (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 \( P(-0.83235 < z < 1.21532) \)

2. Compute \( P(2.5321 < z) \)

3. Compute \( P(z < -1.6523583) \)