HW2 Solutions
Problem A.

<table>
<thead>
<tr>
<th>y</th>
<th>x</th>
<th>x-bar</th>
<th>y-bar</th>
<th>(x-bar)(y-bar)</th>
<th>(x-bar)^2</th>
<th>Y^</th>
<th>Y-Y^</th>
<th>(Y-Y^)^2</th>
<th>(Y-Y-bar)^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>7</td>
<td>3</td>
<td>-3</td>
<td>-9</td>
<td>9</td>
<td>3.6</td>
<td>0.4</td>
<td>0.2</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>7.0</td>
<td>-1.0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>-2</td>
<td>3</td>
<td>-6</td>
<td>4</td>
<td>9.3</td>
<td>0.7</td>
<td>0.5</td>
<td>9</td>
</tr>
</tbody>
</table>

Y = 11.6 - 1.1X

\[ -1.1 \] SSR = 1.7 \[ R^2 = 0.91 \]

\[ 11.6 \] SST = 20.0 \[ 0.09 \]

\[ 1 - 1.7 / 20 \]

\[ -0.96 \]

If X=10 Y^ = 11.6 - 1.1 * 10 = 0.6

```
. reg Y X

 Source | SS       df | MS             Number of obs = 4
---------|-----------|-----------------|-------------------|
 Model   | 18.2857143 1 | 18.2857143 | Prob > F = 0.0438
 Residual | 1.71428571 2 | .857142857 | R-squared = 0.9143
---------|-----------|-----------------|-------------------|
 Total    | 20        | 3 | 6.66666667 | Root MSE = 0.92582
---------|-----------|-----------------|-------------------|

 Y | Coef.  Std. Err. | t | P>|t| | [95% Conf. Interval]|
   |         |         |   |     |                 |
 X  | -1.142857 | .2474358 | -4.62 | 0.044 | -2.207488 | -.0782267 |
 _cons | 11.57143 | 1.092647 | 10.59 | 0.009 | 6.870148 | 16.27271 |
```


2.6 (i) Yes. If living closer to an incinerator depresses housing prices, then being farther away increases housing prices. 

(ii) If the city chose to locate the incinerator in an area away from more expensive neighborhoods, then log(dist) is positively correlated with housing quality. This would violate SLR.4, and OLS estimation is biased.

(iii) Size of the house, number of bathrooms, size of the lot, age of the home, and quality of the neighborhood (including school quality), are just a handful of factors. As mentioned in part (ii), these could certainly be correlated with dist [and log(dist)].

2.11 (i) We would want to randomly assign the number of hours in the preparation course so that hours is independent of other factors that affect performance on the SAT. Then, we would collect information on SAT score for each student in the experiment, yielding a data set \{(sat, hours) : i = 1,...,n\}, where n is the number of students we can afford to have in the study. From equation (2.7), we should try to get as much variation in hours, as is feasible.

(ii) Here are three factors: innate ability, family income, and general health on the day of the exam. If we think students with higher native intelligence think they do not need to prepare for the SAT, then ability and hours will be negatively correlated. Family income would probably be positively correlated with hours, because higher income families can more easily afford preparation courses. Ruling out chronic health problems, health on the day of the exam should be roughly uncorrelated with hours spent in a preparation course.

(iii) If preparation courses are effective, \( \beta_1 \) should be positive: other factors equal, an increase in hours should increase sat.

(iv) The intercept, \( \beta_0 \), has a useful interpretation in this example: because E(u) = 0, \( \beta_0 \) is the average SAT score for students in the population with hours = 0.

C2.2 (i) Average salary is about 865.864, which means $865,864 because salary is in thousands of dollars. Average ceoten is about 7.95.

(ii) There are five CEOs with ceoten = 0. The longest tenure is 37 years.

(iii) The estimated equation is

\[
\log(\text{salary}) = 6.51 + .0097 \text{ceoten}
\]

\[n = 177, \quad R^2 = .013.\]

We obtain the approximate percentage change in salary given \( \Delta \text{ceoten} = 1 \) by multiplying the coefficient on ceoten by 100, 100(.0097) = .97%. Therefore, one more year as CEO is predicted to increase salary by almost 1%.
C2.4 (i) Average salary is about $957.95 and average IQ is about 101.28. The sample standard deviation of IQ is about 15.05, which is pretty close to the population value of 15.

(ii) This calls for a level-level model:

\[ wage = 116.99 + 8.30 \times IQ \]

\[ n = 935, \ R^2 = .096. \]

An increase in IQ of 15 increases predicted monthly salary by 8.30(15) = $124.50 (in 1980 dollars). IQ score does not even explain 10% of the variation in wage.

(iii) This calls for a log-level model:

\[ \log(wage) = 5.89 + .0088 \times IQ \]

\[ n = 935, \ R^2 = .099. \]

If ΔIQ = 15 then Δlog(wage) = .0088(15) = .132, which is the (approximate) proportionate change in predicted wage. The percentage increase is therefore approximately 13.2.

C2.5 (i) The constant elasticity model is a log-log model:

\[ \log(rd) = \beta_0 + \beta_1 \log(sales) + u, \]

where \( \beta_1 \) is the elasticity of \( rd \) with respect to \( sales \).

(ii) The estimated equation is

\[ \log(rd) = -4.105 + 1.076 \log(sales) \]

\[ n = 32, \ R^2 = .910. \]

The estimated elasticity of \( rd \) with respect to \( sales \) is 1.076, which is just above one. A one percent increase in \( sales \) is estimated to increase \( rd \) by about 1.08%.
(i) It seems plausible that another dollar of spending has a larger effect for low-spending schools than for high-spending schools. At low-spending schools, more money can go toward purchasing more books, computers, and for hiring better qualified teachers. At high levels of spending, we would expend little, if any, effect because the high-spending schools already have high-quality teachers, nice facilities, plenty of books, and so on.

(ii) If we take changes, as usual, we obtain
\[ \Delta \text{math10} = \beta \Delta \log(\text{expend}) \approx \left( \beta_i / 100 \right)(\% \Delta \text{expend}), \]
just as in the second row of Table 2.3. So, if \( \% \Delta \text{expend} = 10 \), \( \Delta \text{math10} = \beta_i / 10 \).

(iii) The regression results are
\[ \text{math10} = -69.34 + 11.16 \log(\text{expend}) \]
\[ n = 408, \quad R^2 = .0297 \]

(iv) If \( \text{expend} \) increases by 10 percent, \( \text{math10} \) increases by about 1.1 percentage points. This is not a huge effect, but it is not trivial for low-spending schools, where a 10 percent increase in spending might be a fairly small dollar amount.

(v) In this data set, the largest value of \( \text{math10} \) is 66.7, which is not especially close to 100. In fact, the largest fitted values is only about 30.2.