1. Consider the following data on monthly sales ($Y$, $\$1000$s) and years of experience ($X$) for a sample of 10 real estate agents:

<table>
<thead>
<tr>
<th>Agent</th>
<th>$X$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>123</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>117</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>136</td>
</tr>
</tbody>
</table>

The analysis of the complete data set yielded the following quantities:

- $\bar{Y} = 108$
- $\sum Y_i^2 = 2442$
- $\sum X_i = 60$
- $\sum X_i^2 = 632$
- $\sum X_iY_i = 1896$
- $\sum Y_i^2 = 25000$
- $s_x = 3.97$
- $s_y = 16.5$

where lower-case letters indicate variables measured in deviations from their means and upper case indicate original metric. The regression model of interest is

$$Y_i = \alpha + \beta x_i + \epsilon_i$$

where the disturbance $\epsilon_i$ is normally distributed with mean zero and variance $\sigma^2(\epsilon|x)$, and all the assumptions of the classical linear regression model are met.

a. Identify the random variables in the above regression equation.

b. Given the above model and assumptions, write an expression for the so-called population regression function:

c. Find the least-squares estimates of all the parameter(s) determining the population mean of $Y$.

d. Using the metrics of the variables as identified above, give an interpretation of $\hat{\beta}$ and contrast it with the interpretation of $\beta$ itself.

e. According to your estimates, what is the expected difference in the mean sales of agents with 6 years of experience and agents with 12 years of experience.

f. According to your estimates, what does the model predict for the monthly sales of agent "8."

g. Is the fitted value found for question "f" an estimate of a population parameter? Explain.

h. Think about the other forces besides experience that might affect an agent’s monthly sales. For agent "3", would you estimate the average contribution of these forces to agent 3’s sales to be positive or negative or zero?

i. Find the value of the quantity $\sum (\hat{Y}_i - \bar{Y})^2$, and then use it to construct a measure of the goodness-of-fit of the regression.

j. Give the least-squares estimates of the unconditional variance $\sigma^2(y)$ and the conditional variance $\sigma^2(y|x)$ of $Y$. Give an interpretation of the difference in these quantities.

k. What are the values of the following correlations:
ka. $r_{XY}$ (i.e., correlation between observed X and fitted $\hat{Y}$)

kb. $r_{eX}$ (i.e., correlation between residuals and observed X)

c. $r_{eY}^2$ (i.e., square correlation of residuals and observed Y)

1. Find the standard error of the regression, or, equivalently, the standard error of estimate, identify its metric, and interpret it.

j. Estimate the standard error (i.e., $\hat{\sigma}_{\hat{\beta}}$) of $\hat{\beta}$.

end of question 1 pertaining to data on page 1

2. Shown below is a partial listing of Stata output from a multiple regression of some Y on three independent variables, X, W, and V.

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1600</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>Residual</td>
<td>B</td>
<td>180</td>
<td>D</td>
</tr>
<tr>
<td>Total</td>
<td>2800</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Find the values of the missing quantities given by the upper-case letters.

3. Indicate whether the following are true or false:

a. $E(Y|X) = \alpha + \beta X_i + \epsilon$

b. $SS_{Residual} = \sum y_i^2 - \hat{\beta} \sum x_i^2$

c. The least-squares estimator of $\beta$ will be biased if the regressor X is correlated with the disturbances.

d. $\hat{\beta} = \beta + \sum x_i \epsilon_i / \sum x_i^2$ is true only if the disturbances are homoscedastic.

e. Other things equal, a decrease in the disturbance variance $\sigma_\epsilon^2$ will reduce the random error in estimates of $\beta$.

f. The least-squares estimator of $\beta$ can still be unbiased even if the observations on the dependent variable are not independent and are heteroscedastic.

g. SSRegress is a good measure of the effect of a unit change in X on the mean of Y.

i. Too much variation in a regressor $x$ is a bad thing because it reduces the chance that a single randomly selected $\hat{\beta}$ will be close to $\beta$. 

2