Quiz 4

Please show your work in all the problems to get full credit.

**Problem 1:** The following data set is available:

<table>
<thead>
<tr>
<th>x</th>
<th>-4</th>
<th>2</th>
<th>8</th>
<th>6</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>3</td>
<td>6</td>
<td>12</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

(1) [1 points] Draw a scatterplot of this data set.

(2) [2 points] Calculate the linear correlation coefficient.

(3) [1 points] Set the hypothesis with the correct symbols for testing whether there is a significant linear correlation between the two variables.

(4) [2 points] Use a formal test statistic with a level of significance $\alpha = 0.01$ to determine whether there is a significant linear correlation between $x$ and $y$.

**Problem 2:** MULTIPLE CHOICE. For the next two questions, identify the choice that contains a conclusion with a common correlation error.

(1) Given: There is a significant linear correlation between cordgrass monitored for survival on the left bank of a creek and cordgrass monitored for survival on the right bank of a creek.
   A) Conclusion: $r$ results in a value close to 1.
   B) Conclusion: an increase in cordgrass survival on the left bank is related to an increase in cordgrass survival on the right bank.
   C) Conclusion: a decrease in cordgrass survival on the left bank causes a decrease in cordgrass survival on the right bank.
D) Conclusion: $r$ results in a value close to -1.

(2) Given: There is a linear correlation coefficient very close to 0 between mothers who smoked during pregnancy and the incidence of influenza in their babies.
   A) Conclusion: the frequency of mothers’ smoking is not related in any way to the incidence of influenza in their babies.
   B) Conclusion: an increase in the frequency of mothers’ smoking is not linearly related to an increase in the incidence of influenza in their babies.
   C) Conclusion: there is not a linear relationship between the frequency of mothers’ smoking and the incidence of influenza in their babies.
   D) Conclusion: a decrease in the frequency of mothers’ smoking is not linearly related to a decrease in the incidence of influenza in their babies.

(4.0 pts (2.0 each))

Problem 3: Use the following sample data to answer the next four questions. The paired data consist of the cost of regionally advertising (in thousands of dollars) a certain pharmaceutical drug and the number of new prescriptions written (in thousands).

<table>
<thead>
<tr>
<th>Cost</th>
<th>9</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>2</th>
<th>5</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>85</td>
<td>52</td>
<td>55</td>
<td>68</td>
<td>67</td>
<td>86</td>
<td>85</td>
<td>75</td>
</tr>
</tbody>
</table>

Use: $\Sigma x = 44$, $\Sigma y = 573$, $\Sigma x^2 = 320$, $\Sigma y^2 = 42313$, $\Sigma x.y = 3385$

(1) [2 points] Find the equation of the regression line, letting Cost be the independent ($x$) variable.

(2) [1.5 points] Find the predicted value of the number of new prescriptions written if $6000$ is spent in regional advertising.

(3) [1.5 points] Find the standard error of estimate.

Equations
Linear correlation coefficient: $r = \frac{n(\Sigma x.y) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$

t-test for correlation: $t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$ with $n - 2$ degrees of freedom

Linear regression equation: $\hat{y} = b_0 + b_1 x$ where $b_1 = \frac{n(\Sigma x.y) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$ and $b_0 = \bar{y} - b_1 \bar{x}$

Standard error of estimate: $S_e = \sqrt{\frac{\Sigma(y)^2 - b_0 \Sigma y - b_1 \Sigma x.y}{n-2}}$