Data I: Vartanian

1. You hypothesize that there is a positive relationship between the level of education of an individual and the income that individual earns. You collect a sample of individuals to test your theory and get the following information:

<table>
<thead>
<tr>
<th>X (years of education)</th>
<th>Y (Income, in thousands of $s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

A. What are your research and null hypotheses?
B. How many tails does your test have?
C. You’ve determined that the correlation coefficient=.60. What information does the correlation coefficient give you?
D. How much of the variation in the dependent variable (Y) is being explained by the independent variable (x)?
E. You’ve determined the slope coefficient, b, is equal to .60, the SE_b=.462, and the intercept, a, is equal to 2.6. What is the regression equation for this sample? Draw this line out on a graph. What information do the a and b coefficients give you?
F. Can you reject the null hypothesis at the .05 level of significance?
G. Through some miracle, you are able to find out that in the population there is not a relationship between x and y. Have you made any errors?

2. You hypothesize that there is a negative relationship between the number of times you eat out per week and the stress level with your significant other. You sample yourself and your other for 6 weeks and get the following sample information:

<table>
<thead>
<tr>
<th>Times eating out</th>
<th>Stress Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
</tr>
</tbody>
</table>
Answer questions A-F above with the following information.
\[ r = -0.78, \ b = -0.35, \ a = 6.8, \ SE_b = 0.140 \]

G. While digging a very deep hole in your backyard to relieve some stress, you come across a secret document that informs you that in the population there is a relationship between the number of times you eat out and the stress level of couples. Have you made any errors?

3. You hypothesize that there is a negative relationship between the number of questions asked in class by students and the amount of material covered on the exam. You sample your classes and find the following information:

<table>
<thead>
<tr>
<th>Number of Questions Asked</th>
<th>Pages (*100) of Material Covered on Exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
</tr>
</tbody>
</table>

Answer questions A-F from question 1 with the following information:
\[ r = -0.99, \ b = -0.48795, \ a = 8.93, \ SE_b = 0.031 \]

G. After analyzing your data you have a heart attack and almost die. Fortunately, you have an out-of-body experience where you learn the secrets of the population. One of the lessons you learn is that there is no relationship between the number of questions asked in class and the amount of material covered on the exam in the population. Because of this near-death experience, you come to appreciate all that your classes have to offer. This makes you wonder whether or not you've made an error with your sample. Have you?
Answer Sheet for Regression and Correlation Analysis.

In each of the answers I calculate the r, b, and a coefficients even though I’ve already given them to you in the problems.

Answers

1.

A. Research Hypothesis: $H_a$: Education $> 0$ or $B > 0$
   Null Hypothesis: $H_0$: Education $= 0$ or $B = 0$.

B. 1 tailed test since you predicted the direction.

C. The correlation coefficient tells you the strength and direction of the relationship between your independent and dependent variables.

Correlation coefficient:

$$N=5 \quad \Sigma XY=106 \quad \Sigma X=20 \quad \Sigma Y=25 \quad \Sigma X^2=90 \quad \Sigma Y^2=135$$

$$(\Sigma X)^2=400 \quad (\Sigma Y)^2=625$$

$$r_{xy}=\frac{(5*106)-(20*25)}{\sqrt{(5*90-400)(5*135-625)}}=.60$$

D. 36%

E. The a coefficient or the intercept gives you the value of Y when the independent variable is equal to 0. In this question, a gives you the value of income when the level of education is equal to 0. The b coefficient gives you the slope between X and Y. The slope tells you how much your dependent variable changes when the independent variable increases by a single unit. In this question the slope or b coefficient is telling you how much income changes when education increases by a single unit (or simply increases by 1).

Regression line:

$$Y=a + bX$$

To determine b:
Therefore, to determine a:

\[ \begin{align*}
  b &= \frac{(5 \times 106) - (20 \times 25)}{(5 \times 90) - 400} = 0.60 \\
  a &= \frac{25 - 0.60 \times 20}{5} = 2.6 
\end{align*} \]

Thus, the least squares line is

\[ Y = 2.6 + 0.60 \times X \]

F. \[ t_{5-1-1} = \frac{0.60}{0.462} \]

\[ t_3 = 1.2987 \]

Because the t value is less than the critical value, 2.353, you would fail to reject the null hypothesis.

G. You did not make an error. Your sample shows no relationship between the variables and this is in fact the case for the population.

2.

A. Research Hypothesis: \( H_r: \) Eating Out < 0 or B < 0
B. Null Hypothesis: \( H_0: \) Eating Out = 0 or B = 0.

B. 1 tailed test.
C. Correlation Coefficient:

\[ r_{xy} = \frac{(6 \times 77) - (48 \times 24)}{\sqrt{(6 \times 712 - 2304)(6 \times 162 - 576)}} = -0.78 \]
D. The percentage of the variation of the dependent variable explained by the independent variable is \((-0.78)^2 = 0.61\) or 61%.

E. Look up the critical value of \(t\) in the table in the book. This critical value for a one-tailed test at the .05 level of significance is 2.132. Because the absolute value of your \(t\) (2.5) is greater than the critical value in the book, you will reject the null hypothesis.

F. The regression equation:

\[
b = \frac{(6*77)-(48*24)}{(6*712)-2304} = -0.35
\]

Therefore, to determine \(a\):

\[
a = \frac{24-(-0.35)*48}{6} = 6.8
\]

The regression line is:

\[Y = 6.8 - 0.35X\]

G. You did not make an error. Because you have found a relationship at the 5% level of significance in your sample and there was truly a relationship in the population, no error was made.

3.

A. \(H_1: \) Questions Asked < 0 or B < 0
   \(H_0: \) Questions Asked = 0 or B = 0.

B. One tailed test.

C. Correlation coefficient.

\[
r_{xy} = \frac{(5*99)-(30*30)}{\sqrt{[5*346-900][5*220-900]}} = -0.99
\]
D. \((-0.99)^2 = 98.8\%\) of the variation in Y is explained by X.

E. The t value in the book shows that the critical value of t for a df=3 is 2.353. Because the absolute value of the t you have calculated is greater than the critical value, you will reject the null hypothesis.

F. The Regression Equation:

\[ b = \frac{(5\times99) - (30\times30)}{(5\times346) - 900} = -0.48795 \]

Therefore, to determine a:

\[ a = \frac{30 - (-0.48795)\times30}{5} = 8.93 \]

Regression equation:

Y = 8.93 - 0.48795 X

F. You made an error. Because your sample shows that there is a relationship between the variables when in fact there is not a relationship, you have made a type I error.