General tips for writing Free Response Answers: Understand your obligation as a test taker. You are being evaluated not only on the correctness of your answers, but also on your ability to communicate the methods you used to reach them. The answer is everything you write down, not just the last line or number at the end. Convince the reader that you understand the key concepts in the question. Don’t just give them the numbers and hope they will assume you understand the concepts.

1. A newspaper in Germany reported that the more semesters needed to complete an academic program at the university, the greater the starting salary in the first year of a job. The report was based on a study that used a random sample of 24 people who had recently completed an academic program. Information was collected on the number of semesters each person in the sample needed to complete the program and the starting salary, in thousands of euros, for the first year of a job. The data are shown in the scatterplot below.

```
   70
   65
   60
   55
   50
   45
   40
   35
   30
   25

   Starting Salary (1,000 euros)

   5  10  15  20
   Number of Semesters
```

a.) Does the scatterplot support the newspaper report about the number of semesters and starting salary? Justify your answer.

The Scatterplot supports the newspaper report because there is a positive, linear association between number of semesters and starting salary.

The table below shows computer output from a linear regression analysis on the data.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>34.018</td>
<td>4.455</td>
<td>7.64</td>
<td>0.000</td>
</tr>
<tr>
<td>Semesters</td>
<td>1.1594</td>
<td>0.3482</td>
<td>3.33</td>
<td>0.003</td>
</tr>
</tbody>
</table>

S = 7.37702  R-Sq = 33.5%  R-Sq(adj) = 30.5%

b.) Identify the slope of the least-squares regression line, and interpret the slope in context.

Slope = 1.1594  \[
\frac{Y}{x} = \frac{\text{Starting salary}}{\text{# of semesters}} = 1.1594
\]

For every one semester needed to complete program, salary increases by 1.1594 euros (in thousands).
An independent researcher received the data from the newspaper and conducted a new analysis by separating the data into three groups based on the major of each person. A revised scatterplot identifying the major of each person is shown below.

![Scatterplot showing salary vs. number of semesters by major](image)

**Major**
- Business
- Physics
- Chemistry

(c.) Based on the people in the sample, describe the association between starting salary and number of semesters for the business majors.

*Linear*

Business majors have a strong, negative, linear correlation between number of semesters to finish a program and starting salaries.
The more semesters it takes, the less the starting salary.

(d.) Based on the people in the sample, compare the median starting salaries for the three majors.

Business majors have the lowest median salary at approx. $38,000 euros, followed by physics majors at approx. $48,000 euros, and then chemistry majors at approx. $58,000 euros.

(e.) Based on the analysis conducted by the independent researcher, how could the newspaper report be modified to give a better description of the relationship between the number of semesters and the starting salary for the people in the sample?

The newspaper report could be modified to account for the major. Overall, majors that take the longest to complete tend to have a higher starting salary. Also, within the majors, students that take longer to complete the program tend to make less.
2.) A plot of the number of defective items produced during 20 consecutive days at a factory is shown below.

<table>
<thead>
<tr>
<th>Number of Defective Items</th>
<th>Tally of defective items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 1 1 = 3</td>
</tr>
<tr>
<td>2</td>
<td>1 1 1 1 = 4</td>
</tr>
<tr>
<td>3</td>
<td>1 1 1 1 1 = 5</td>
</tr>
<tr>
<td>4</td>
<td>1 1 1 1 = 4</td>
</tr>
<tr>
<td>5</td>
<td>1 1 = 2</td>
</tr>
</tbody>
</table>

(a) Draw a histogram that shows the frequencies of the number of defective items. Ignore day number 1998.

(b) Give one fact that is obvious from the histogram but is not obvious from the scatterplot.

The number of defective items has a symmetric distribution.

(c) Give one fact that is obvious from the scatterplot but is not obvious from the histogram.

The number of days and number of defective items are negatively correlated, meaning as days increase, the number of defects decrease.

3.) The Great Plains Railroad is interested in studying how fuel consumption is related to the number of railcars for its trains on a certain route between Oklahoma City and Omaha. A random sample of 10 trains on this route has yielded the data in the table below.

<table>
<thead>
<tr>
<th>Number of Railcars</th>
<th>Fuel Consumption (gallons/mile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>56</td>
</tr>
<tr>
<td>20</td>
<td>52</td>
</tr>
<tr>
<td>27</td>
<td>94</td>
</tr>
<tr>
<td>31</td>
<td>80</td>
</tr>
<tr>
<td>47</td>
<td>114</td>
</tr>
<tr>
<td>23</td>
<td>98</td>
</tr>
<tr>
<td>39</td>
<td>87</td>
</tr>
<tr>
<td>51</td>
<td>122</td>
</tr>
<tr>
<td>29</td>
<td>100</td>
</tr>
<tr>
<td>39</td>
<td>70</td>
</tr>
</tbody>
</table>
A scatterplot, a residual plot, and the output from the regression analysis for these data are shown below.

(a) Is the linear model appropriate for modeling these data? Clearly explain your reasoning.
   Yes, a linear model is appropriate for these data. The scatterplot shows a strong, positive, linear association, and the residual plot shows a reasonably random scatter of points.

(b) Suppose the fuel consumption cost is $25 per unit. Give a point estimate (single value) for the change in the average cost of fuel per mile for each additional railcar attached to a train. Show your work.
   \[ \text{Slope} = \frac{\text{Fuel Consumption}}{\text{Number of Railcars}} = 2.15 \]
   Each additional car increases fuel consumption by 2.15 units per mile.

(c) If cost is 8.25 per unit, then average cost per car added will be approx. 25(2.15) = $53.75

(d) Would it be reasonable to use the fitted regression equation to predict the fuel consumption for a train on this route if the train had 65 railcars? Explain.
   No, the data set does not contain any information about fuel consumption for any trains with more than 50 cars. Using the LSRL to predict fuel consumption for a train with 65 cars is extrapolation and is not reasonable.