1. Are employee sick days for a particular company equally likely to be used on the five days of the workweek (Monday – Friday)? To investigate this the company’s manager analyzes 250 employee sick days and records the day of the week for each.

a) Identify the observational units and variable in this study.

b) Is this a categorical (also binary?) or quantitative variable?

c) Specify (with symbols) the null hypothesis to be tested.

d) Identify the appropriate test procedure for the manager to use.

e) If the manager obtains a p-value of .314, can she legitimately conclude that the sample data provide very strong evidence that employee sick days are not equally likely to be used on the five days of the workweek? Explain.

f) If the manager obtains a p-value of .314, can she legitimately conclude that the sample data provide very strong evidence that employee sick days are equally likely to be used on the five days of the workweek? Explain.

2. In a recent study, researchers investigated possible biochemical mechanisms that could be involved in the early stages of romantic love. They measured plasma level of neurotrophins for a sample of 58 subjects who had recently fallen in love. They also asked each subject to rate his/her level of passionate love feelings on a numerical scale. Researchers calculated the correlation coefficient between the level of passionate love and plasma level of neurotrophins to be \( r = 0.34 \).

Conduct the appropriate test of whether this sample provides strong evidence (at the \( \alpha = .05 \) level) of a positive correlation between these variables in the population. Report the relevant hypotheses, test statistic, p-value, test decision, and conclusion.

3. A recent study investigated whether seat position within a bus may be related to whether a passenger experiences motion sickness. The following table classifies each person in a random sample of bus riders by the location of his or her seat and whether nausea was reported:

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Middle</th>
<th>Rear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nausea</td>
<td>58</td>
<td>166</td>
<td>193</td>
</tr>
<tr>
<td>No nausea</td>
<td>870</td>
<td>1163</td>
<td>806</td>
</tr>
</tbody>
</table>

a) What proportion of those who rode in the front experienced nausea?

b) Determine the expected count for the (middle, no nausea) cell of the table. (Show your work.)

Consider the following output (notice that one entry has been replaced by \( \text{xxx} \)): 
Chi-Square Test: Front, Middle, Rear

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Middle</th>
<th>Rear</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>118.85</td>
<td>170.21</td>
<td>127.94</td>
<td>417</td>
</tr>
<tr>
<td>2</td>
<td>809.15</td>
<td>xxx</td>
<td>871.06</td>
<td>2839</td>
</tr>
<tr>
<td>Total</td>
<td>928</td>
<td>1329</td>
<td>999</td>
<td>3256</td>
</tr>
</tbody>
</table>

Chi-Sq = 73.789, DF = 2, P-Value = 0.000

c) What conclusion would you draw from the above chi-square output? Explain briefly.

d) Which seating position seems to be the worst with respect to nausea? Explain how you conclusion follows from the chi-square analysis.

4. Researchers studied heart rates after engaging in physical exercise for adults who were also classified according to whether and how much they smoke. Data were collected to investigate whether there are differences in mean heart rates among various smoking classifications (heavy, light, moderate, non-smoker). The output below pertains to an ANOVA analysis addressing this issue:

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>smoking status</td>
<td>3</td>
<td>1464.1</td>
<td>488.0</td>
<td>6.08</td>
<td>0.004</td>
</tr>
<tr>
<td>Error</td>
<td>20</td>
<td>1604.8</td>
<td>80.2</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>3069.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a) State the appropriate null hypothesis in symbols. Also explain what the symbols mean.

b) Summarize the conclusion that you would draw from the ANOVA F-test.

5. In July 2006 I selected a sample of 87 used Honda Civics listed for sale on the web. The data led to the following Minitab output:
The regression equation is
Price = 18971 - 0.0983 Mileage

<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>18970.7</td>
<td>500.5</td>
<td>37.90</td>
<td>0.000</td>
</tr>
<tr>
<td>Mileage</td>
<td>-0.098348</td>
<td>0.009071</td>
<td>-10.84</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Obs  Fit  SE Fit  95% CI 95% PI
1  14053  256  (13545, 14562)  (9303, 18803)

Obs Mileage
1  50000

a) Identify the explanatory variable and the response variable.

b) Report the value of the slope coefficient, and interpret what this value means in this context.

c) Use the output to determine the value of the correlation coefficient.

d) What proportion of the variability in prices is explained by the least squares line with mileage?

e) Report the test statistic and p-value for testing whether the data provide evidence that mileage is of some use in predicting price. Also summarize the conclusion that you draw from this test.

f) Does the output include a confidence interval for the population slope? If so, report the interval, but do not bother to interpret it. If not, just answer no.

g) Does the output include a prediction interval for the price of a used Honda Civic with 50,000 miles? If so, report the interval, but do not bother to interpret it. If not, just answer no.

For the next two questions, think about how a prediction interval for the price of a used Honda Civic with 50,000 miles would compare to a prediction interval for the price of a used Honda Civic with 100,000 miles.

i) How would their midpoints compare? (Circle one. Do not bother to explain.)

Larger for 50,000 mile car  
Same midpoints  
Larger for 100,000 mile car

j) How would their widths compare? (Circle one. Do not bother to explain.)

Wider for 50,000 mile car  
Same width  
Wider for 100,000 mile car

6. Short answer:

a) It can be shown that the sum of the residuals from a least squares regression line equals zero. (Do not bother to try to justify this. Just accept this fact.) Does it follow that the mean of the residuals must equal zero? Explain briefly.
b) Does it follow that the median of the residuals must equal zero? Explain briefly.

c) Suppose that every student in a class scores 120 points for their combined score on the first two exams. What would be the value of the correlation coefficient between exam 1 score and exam 2 score? Explain briefly.