1. The following ordered data represent the amount of cadmium observed in a sample of 40 sea scallops:

   2.7 3.7 4.5 4.5 5.0 5.1 5.1 5.5 5.5 5.7
   6.4 6.5 7.5 7.7 7.9 7.9 8.0 8.4 8.5 8.9
   9.5 9.6 9.6 10.1 10.8 10.8 11.4 12.1 12.4 12.7
  13.1 13.1 14.1 14.4 14.7 16.9 17.1 18.0 18.9 27.0

   a. Find the lower (Q₁) and upper quartile (Q₃) and calculate the interquartile range for the cadmium concentration data. Show workings.

      \(Q₁ = \) 
      \(Q₃ = \) 
      \(IQR = \)

   b. Find the median of the data.

      Median = 

   c. Using your answer in part (b) and knowing that the mean of the data is 10.03, what can you say about the shape of the distribution of cadmium concentration in sea scallops? Explain.

      Circle one: fairly symmetric / skewed to the left / skewed to the right

      Explain: 

   d. Knowing that the mean and standard deviation for the sample are 10.03 and 4.99 respectively, give the standardized version of \(x\), where \(x\) denotes the amount of cadmium found in sea scallops.

      \(Z = \)

   e. Using your answer in part (d), determine the z-score for a concentration of cadmium in sea scallops of 27 (largest observation in sample).

      Answer:

   f. Interpret the z-score in part (e). 
      Interpretation: 

g. If the largest observation in the data (27) is removed from the sample, what effect would it have on the mean? Explain

The mean for the 39 observations would be (circle one) smaller / same / larger than 10.03.

Because: 

(16 Marks)

2. A nutritionist studied the nutritional status of children in two villages in Ghana. Among other things she recorded the weight (in kg) of each child. The boxplot for the weight of children in the two villages follow.

a. By looking at the boxplot diagram above, comment on the shape of the distribution of weight of children in the 2nd village.

Circle one: fairly symmetric / skewed to the left / skewed to the right

Because: 

b. Get an approximate value for the range of weight of children in both villages. Show workings

Approx. range for children's weight in 1st village:

Approx. range for children's weight in 2nd village:

Because: 

c. In which village is children's weight more variable? Explain.

The children’s weight in the 1st village display (circle one) more / same / less variability than the children’s weight in the 2nd village.

Because: 

-2-
d. In which village is the children’s average weight higher?

Approx. average weight for children in 1st village: __________

Approx. average weight for children in 2nd village: __________

The average weight of children in 1st village is (circle one): smaller / same / larger than the average weight of children in the 2nd village.

(10 Marks)

3. Many studies have been done that indicate the maximum heart rate an individual can reach during intensive exercise decreases with age. A physician decided to do his own study and recorded the ages and peak heart rates of 10 randomly selected people. The results are shown in the following table where x denotes age, in years, and y denotes peak heart rate.

<table>
<thead>
<tr>
<th>x</th>
<th>30</th>
<th>38</th>
<th>41</th>
<th>38</th>
<th>29</th>
<th>39</th>
<th>46</th>
<th>41</th>
<th>42</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>186</td>
<td>183</td>
<td>171</td>
<td>177</td>
<td>191</td>
<td>177</td>
<td>175</td>
<td>176</td>
<td>171</td>
<td>196</td>
</tr>
</tbody>
</table>

a. Plot the data.

b. In looking at the plot in part (a) what type of correlation (negative / positive / no correlation) would you expect? Explain.

Answer (circle one): positive / negative / no correlation

Because: __________________________________________

c. Calculate the correlation coefficient for the sample. Show workings.

(Hint: \( \sum x = 368, \quad \sum x^2 = 13,968, \quad \sum y = 1,803, \quad \sum y^2 = 325,723, \quad \sum xy = 65,865 \))
4. The following frequency table shows the classification of 58 landfills in a state according to their concentration of the two hazardous chemicals, arsenic and mercury.

<table>
<thead>
<tr>
<th>Mercury</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arsenic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Low</td>
<td>14</td>
<td>26</td>
</tr>
</tbody>
</table>

Define the events A and M as follows:
A : high concentration of arsenic
M : high concentration of mercury.

If a landfill is selected at random find the probability that it has

a. high concentration of mercury. Answer:

b. high concentration of mercury given that it has a high concentration of arsenic. Answer:

c. low concentrations in any of the two chemicals. Answer:

d. high concentration in both chemicals. Answer:

e. Are the events A and M independent? Explain. Circle one: Yes / No

Because:

T H E   E N D
FORMULAE

The following formulae may be useful:

\[
\bar{x} = \frac{\sum x_i}{n} \quad \quad \quad \quad \mu = \frac{\sum x_i}{N}
\]

\[
\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} \quad \quad \quad \quad \quad s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}
\]

\[
\sigma^2 = \frac{\sum x_i^2}{N} - \mu^2 \quad \quad \quad \quad \quad s^2 = \frac{\sum x_i^2 - (\sum x_i)^2}{n - 1}
\]

Range = Max - Min

IQR = Q_3 - Q_1

\[
S_{xx} = \frac{\sum x_i^2 - (\sum x_i)^2}{n} \quad \quad \quad \quad \quad S_{yy} = \frac{\sum y_i^2 - (\sum y_i)^2}{n}
\]

\[
S_{xy} = \frac{\sum x_i y_i - (\sum x_i)(\sum y_i)}{n} \quad \quad \quad \quad \quad r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}
\]

\[
Z = \frac{X - \mu}{\sigma} \quad \quad \quad \quad \quad P(A) + P(\bar{A}) = 1
\]

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \quad \quad \quad \quad P(A \cap B) = P(B) \times P(A|B)
\]

\[
P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \quad \quad \quad \quad P(A|B) = P(A)
\]