Statistics for Biologists I  
STT 464 Section II  
Homework # 1  
Due Monday September 9, 2002

Question 1: Classify each of the following variables using each of the schemes: (i) qualitative or quantitative (ii) discrete or continuous, and (iii), nominal, ordinal, ratio or interval.

(a) The number of cancer deaths in Iowa during 2000.
(b) The letter grade (A,B,C,D, or F) received in a senior level statistics course
(c) The systolic blood pressure of a child
(d) The sex of a turtle
(e) The annual milk production of a Holstein
(f) The age of a sea lamprey
(g) The breed of a sheep
(h) Temperament scores (scored from 1-4) of goats during milking time.

Question 2: Given that

\[
x_1 = 11 \quad x_2 = 12 \quad x_3 = 13 \quad x_4 = -10 \\
y_1 = 9 \quad y_2 = 7 \quad y_3 = 5 \quad y_4 = 3
\]

find:

a) \( \sum_{i=1}^{4} x_i \)  
\( \sum_{i=1}^{4} x_i \)  
\( \left( \sum_{i=1}^{4} x_i \right)^2 \)  
\( \left( \sum_{i=1}^{4} x_i \right)^2 \)  
\( \sum_{i=1}^{4} x_i y_i \)  
\( \sum_{i=1}^{4} x_i \left( \sum_{j=1}^{4} y_j \right) \)  
\( \left( \sum_{i=1}^{4} x_i \right) \left( \sum_{j=1}^{4} y_j \right) \)  
\( \sum_{i=1}^{4} x_i \left( \sum_{j=2}^{4} y_j \right) \)

Question 3:

The following data are the ages of 48 tennis players from a community recreation center who suffer from lateral humeral epicondylitis, also known as "tennis elbow".

<table>
<thead>
<tr>
<th>12</th>
<th>18</th>
<th>21</th>
<th>24</th>
<th>26</th>
<th>28</th>
<th>28</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>34</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>38</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>41</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>42</td>
<td>43</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>47</td>
<td>47</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>49</td>
<td>51</td>
<td>53</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
</tbody>
</table>
(a) Compute the sample mean, sample median and sample standard deviation from this data set.
(b) Use SAS PROC CHART to plot a histogram. Would you characterize the data as appearing to be "normal"? Explain.
(c) Use SAS PROC UNIVARIATE to provide a stem-and-leaf plot and box-and-whisker plot. Comment on the information provided.
(d) Determine the 25th and 75th percentiles for this sample.
(e) Does the "empirical rule" apply well to this data?
(f) Does the "range approximation" apply well to this data?

**Question 4:** In general, a linear transformation of a variable $x$ to the variable $y$ can be written as $y = ax + b$ for some constants $a$ and $b$. For example, if $x$ represents the temperature in the Celsius scale, and $y$ is the temperature in the Fahrenheit scale, then the linear transformation relating $x$ and $y$ can be written:

$$y = \frac{9}{5} x + 32$$

(a) Given that the temperatures in the Celsius scale of five patients are $x_1 = 38.2$, $x_2 = 41.1$, $x_3 = 37.5$, $x_4 = 40.8$ and $x_5 = 39.2$, obtain the mean and standard deviation of the temperatures in Fahrenheit degrees.

(b) Without transforming the temperatures from Celsius to Fahrenheit degrees, confirm the results obtained in (a) with the formulas:

$$\bar{y} = \frac{9}{5} \bar{x} + 32 \quad \text{and} \quad s_y = \frac{9}{5} s_x$$