ANSWERS TO TEST NUMBER 1

Question 1: (60 points)

Fire up XLISPSTAT and enter the following command to enter your data

```
> (def data (list  3 1 0 2 24 4 1 0 5 8 6 3 10 4 2))
```

To obtain the mean, first sum the elements

```
> (def sum (sum data))
```

and then divide that sum by the number of elements

```
> (def meanvalue (/ sum 15))
```

To calculate the median we need to sort the data.

```
> (def sorted-data (sort-data data))
```

After entering the command

```
> sorted-data
```

it will be easily seen that the median is 3. The mean is greater than the median because the distribution is skewed to the right due to the influence of the observations 8, 10, and especially 24.

To calculate the range and interquartile range, enter the commands

```
> (quantile data .75)
```

and

```
> (quantile data .25)
```

and notice that the difference between these is 4. Note that XLISPSTAT sets the 75th quantile half-way between observations 11 and 12 from the bottom. Three-quarters of the observations—i.e., 11.25/15—will lie below 5.5 (which is half-way between 5 and 6). And the lower quartile is symmetric with the upper one—four observations are below 1.5. Actually, it would seem more reasonable to set the upper quartile at 6 and the lower quartile at 1. Measurement differences of this sort become inconsequential when the number of elements is large.

To obtain the range we subtract the maximum value from the minimum values to obtain the number 24.

To obtain the variance, we need to calculate the deviation from the mean for each element in the data set.

```
> (def deviations (- data meanvalue))
```

Then we need to square these, sum them, and divide by one less than the number of elements.

```
> (def sqddev (^ deviations 2))
> (def var (/ (sum sqddev) 14))
```

The standard deviation is then the square-root of the variance.

```
> (def std (sqrt var))
```

You should get a value of approximately 6.01 for the standard deviation. It should be obvious that the observation value 24 drags the mean above the median and makes the biggest contribution to the sum of squared deviations. You can see this by entering the command

```
> sqddev
```

and looking for the largest squared deviation, which will be 366.084. Obviously the observation is an outlier and is distorting our measures of central tendency.

The coefficient of variation—100 times the standard deviation divided by the mean—can be shown to be approximately 123.5.

To calculate the standardized values we give XLISPSTAT the following command.

```
> (def Z (/ (- data meanvalue) std))
```

After entering the command

```
> Z
```

You should be able to verify that the largest observation is 3.18 standard deviations above the mean and the smallest is .81 standard deviations below the mean. To show that the standardized values have mean zero and variance equal to unity you can apply the procedures used above to calculate the mean and variance of X. Alternatively, you can apply the XLISPSTAT functions **mean** and **standard-deviation** to `Z`.

Question 2: (10 points)

Given that the distribution of length of battery lives is hump shaped, 95% percent of the life-lengths should lie within 2 standard deviations of the mean. If the manufacturer's claim that the mean is 60 months is true, then only about 2.5% of the battery-lives should be below 40 months. That is, 97.5% of the batteries will have lives longer than 40 months (and hence, longer than the 36 month guarantee). You could infer that your battery is in the bottom 2.5% of the distribution. The manufacturer's claim may well be exaggerated but you cannot infer that from your sample of one item.

Question 3: (5 points)

The province numbers are a qualitative variable since they do not fall into a naturally occurring scale. Since the province picked as 1 or 2 or 5, etc., is chosen arbitrarily the scale is arbitrary.

Question 4: (15 points)

a) Populations and samples: The population is all the elements of interest while a sample is a selection of elements from a population.
b) Populations and processes: A process is a mechanism that generates an infinite population.
c) Elements, observations and variables: An element is a selected member of a population or sample of a population. A variable is a characteristic of elements that varies from element to element. An observation is the value of a variable for a particular element of the population or sample.

Question 5: (10 points)

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
2	4	-1.333	2.333	1.778	5.444	-3.11
3	1	-0.333	-0.667	0.111	0.444	0.22
5	0	1.667	-1.667	2.778	2.778	-2.78
Sum ==> 10	5	0.000	0.000	4.667	8.667	-5.67

Mean X = 10/3 = 3.333 Mean Y = 5/3 = 1.667
Std. Deviation X = sqrt(4.667/2) = 1.528
Std. Deviation Y = sqrt(8.667/2) = 2.082
Covariance = -5.67/2 = -2.83
Correlation Coefficient = -2.83/(1.528 x 2.082) = -0.891