STAT 2606 ASSIGNMENT # 1 SOLUTIONS
OUT OF 36 MARKS TOTAL

1. [1] (c)
2. [1] (c)
3. [1] (b)
4. [1] (c)
5. [1] (a)

6. (4) a) [2] Population of interest –set of 25,000 registered voters.
Sample – set of 200 voters interviewed.
b) [2] Population of interest –set of computer chips that are manufactured.
Sample – set of 1000chips selected for checking.

7. (8) a) [1] Quantitative, Ratio scale
b) [1] Categorical/qualitative, Ordinal scale
c) [1] Quantitative, Ratio scale
d) [1] Categorical/qualitative, Nominal scale
e) [1] Quantitative, Ratio scale
f) [1] Categorical/qualitative, Ordinal scale
g) [1] Categorical/qualitative, Ordinal scale
i) [1] Quantitative, Ratio scale

8. (7) [1] a) The measurements are take on 28 randomly selected dowels not on each and every dowel produced. Therefore the measurements are not a population but a sample.

[1] b) i) mean= $\frac{\sum x_i}{n} = \frac{181.11}{28}$ =6.468

[2] b) ii) Stem-leaf plot with stems 50, 51, ….., 58 and leaf unit 0.01

60 | 7
61 | 0
62 | 6 7 9 9
63 | 0 4 4 6 9
64 | 2 4 6 7
65 | 0 3 5 7 8
66 | 1 2 3 5 8
67 | 5 9
68 | 5

[1] b) iii) From the stem-leaf plot the 14-th and the 15-th observation are 6.46 and 6.47. So the median is $\frac{6.46+6.47}{2} = 6.465$

[1] c) The shape of the distribution is symmetric and there are no obvious outliers.

[1] d) Since the shape of the distribution is symmetric i.e. not skewed, mean is a suitable measurement to represent the central location.

9. (5 – 2 for the mean, median, mode, and quartiles, 1 for the boxplot, 1 for the histogram, 1 for concluding statement about the shape of the distribution)

```
                 The SAS System   11:51 Wednesday, September 26, 2007

                          Basic Statistical Measures

                Location                     Variability

            Mean     6.468214     Std Deviation            0.19558
            Median   6.465000     Variance                 0.03825
            Mode     6.290000     Range                    0.78000
                                  Interquartile Range      0.29500

       NOTE: The mode displayed is the smallest of 2 modes with a count of 2.

                             Quantiles (Definition 5)
                              Quantile      Estimate

                              100% Max         6.850
                              99%              6.850
                              95%              6.790
                              90%              6.750
                              75% Q3           6.615
                              50% Median       6.465
                              25% Q1           6.320
                              10%              6.260
                              5%               6.100
                              1%               6.070
                              0% Min           6.070

                              The UNIVARIATE Procedure
                                   Variable:  x

                               Extreme Observations
                          ----Lowest----        ----Highest---

                          Value      Obs        Value      Obs

                           6.07        9         6.65       27
                           6.10       22         6.68       23
                           6.26       12         6.75       20
                           6.27       25         6.79       19
                           6.29       21         6.85       26
```

```
Stem Leaf                     #             Boxplot
  68 5                        1                |
  68                                           |
  67 59                       2                |
  67                                           |
  66 58                       2                |
  66 123                      3             +-----+
  65 578                      3             |     |
  65 03                       2             |     |
  64 67                       2             *--+--*
  64 24                       2             |     |
  63 69                       2             |     |
  63 044                      3             +-----+
  62 6799                     4                |
  62                                           |
  61                                           |
  61 0                        1                |
  60 7                        1                |
     ----+----+----+----+
   Multiply Stem.Leaf by 10**-1
```

```
    Frequency
    7 +             ***** *****
      |             ***** *****
    6 +             ***** *****
      |             ***** *****
    5 +             ***** ***** *****
      |             ***** ***** *****
    4 +       ***** ***** ***** *****
      |       ***** ***** ***** *****
    3 +       ***** ***** ***** ***** *****
      |       ***** ***** ***** ***** *****
    2 + ***** ***** ***** ***** ***** *****
      | ***** ***** ***** ***** ***** *****
    1 + ***** ***** ***** ***** ***** *****
      | ***** ***** ***** ***** ***** *****
      --------------------------------------

        6.075 6.225 6.375 6.525 6.675 6.825
```

The shape of the distribution is reasonable symmetric.

10. (7) a) i) [1] $\sum_{i=1}^{4} 2 = 8$

ii) [1] $\sum_{i=1}^{4} 2x_i = 2(1+5+8+11) = 50$

iii) [1] $\sum_{i=1}^{4} (2x_i)^2 = \sum_{i=1}^{4} 4x_i^2 = 4\sum_{i=1}^{4} x_i^2 = 4(1+25+64+121) = 844$

b) [1] Taking the values used in 10. a)

$$\sum_{i=1}^{3} x_i^2 = (1+25+64+121) = 211$$

$$\left(\sum x_i\right)^2 = (1+5+8+11)^2 = 25^2 = 625$$

Therefore, we cannot say in general that $\sum x_i^2 = \left(\sum x_i\right)^2$ □

c) i) [1] $\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = n\bar{x} - n\bar{x} = 0$ □

ii) [2] $\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} \left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right) = \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2$

$$= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\left(\frac{\sum x_i}{n}\right)^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$ □