Assignment 1- Statistical Analysis and Report Writing

Introduction

The objective of t and compare the net assets ($m) of a sample of companies of four g sectors using descriptive statistical techniques. The four manufac vestigation are furniture, electrical, whitegoods and clothing respectiv

Data Analysis a

The sample sizes ectors are as follows:

Sector 1: 47
Sector 2: 97
Sector 3: 36
Sector 4: 14

Before beginning to analyse and interpret the data, it is crucial to take note of the sample sizes. All sectors have relatively small sample sizes, in particular sector 4 is extremely small comprising of only 14 samples. The small number of samples in this sector will lead to a decrease in accuracy of the descriptive statistics calculated in comparison with the calculation in the other sectors, in particular to Sector 2 which has the largest sample space. The sample space is also illustrated in Graph 1, the Box Plot below. The sample sizes are reflected in the size of the actual box sections of the graphs, and it becomes clearly evident that Sector 4 has the smallest sample size and Sector 3 the largest.

Figures 1,2,3, and 4 in the appendix are stem and leaf plots which show a visual representation of the samples sizes and are useful for getting a general overview of the data distribution before beginning the analysis.

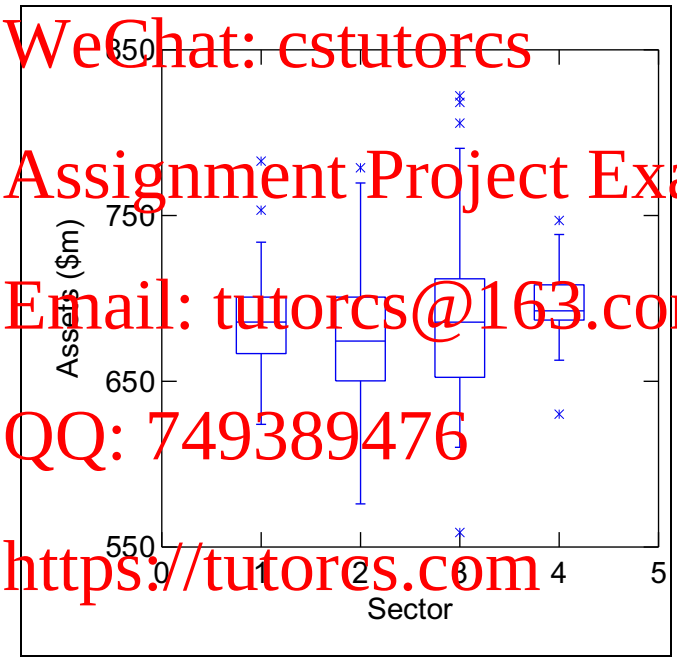
	Sector 1: Furniture	Sector 2: Electrical	Sector 3: Whitegoods	Sector 4: Clothing
Number of companies in sector:	47	97	36	14
Mean:	679.19	668.04	696.22	694.75
Standard Deviation:	47.11	78.35	67.37	29.86
Interquartile Range:	35.56	51.04	71.19	20.59
Minimum:	437.26	218.74	558.78	630.17
10th Percentile:	642.56	619.23	630.54	664.00
Quartile 1:	664.67	649.78	652.97	687.20
Quartile 2 (median):	684.06	672.74	685.75	692.53
Quartile 3:	700.23	700.82	724.16	707.79
90th Percentile:	713.39	731.88	798.16	731.71
Maximum:	782.84	865.17	892.65	747.04

Table 1: Descriptive Statistics of the Four Sectors

Table 1 illustrates a number of descriptive statistics of the four different sectors. On examining the means of the various sectors, it is interesting to note that sectors 3 and 4 have very similar means, 696.22 and 694.75 respectively. Therefore one would infer that the average net values of the assets ($m) of manufacturing sectors 3 and 4 are alike in value.

Sectors 1 and 2 have a lower mean of 679.19 and 668.04 respectively, indicating that the companies within these sectors have lower average net asset values ($m) than the sectors 3 and 4. This however may not necessarily by the case. Outliers can distort the mean considerable since it is an average value calculated using all data values, even the ones that do not fit the trend of the data.

Since the mean ca by outliers, we use the median for a more accurate measure of the ce the central value of the assets when arranged in ascending order. ove it can be seen that the medians of Sector 1, (684.06) and Sect r values than their means. This indicates that there are more compani hat have above average net asset values ($m). The opposite is true fo ing medians below the mean, indicating that more companies in the ues ($m) below the average.



Graph 1: Box Plot of all sectors

Graph 1 is a box plot of all the manufacturing sectors, which we will use to determine if any outliers are present. Unfortunately when changing the axis of the plot so that all the sectors fit on one plot for comparison, some of the outliers were omitted. Graph 6 in the appendix however shows the full set of data, including outliers and extreme outliers. It can be seen that Sector 1 has an extreme outlier below the median (represented by the vertical line) and Sector 2 has two extreme outliers below the median and one above. It becomes evident that the value of the extreme outlier in Sector 1 is 437.26 as read from the minimum value Table 1 above. The same can be assumed from the smallest and largest outliers in Sector 2, being the minimum and maximum values; 218.74 and 865.17 respectively. The box plot also demonstrated that Sector 3 has quite a few outliers above the median and one below, and Sector 4 has one below and one above the median. Using the table above it can be seen that the largest outlier for sector 3 is 892.65 and the smallest is 558.78, and for sector one the outliers are 747.04 and 630.17.

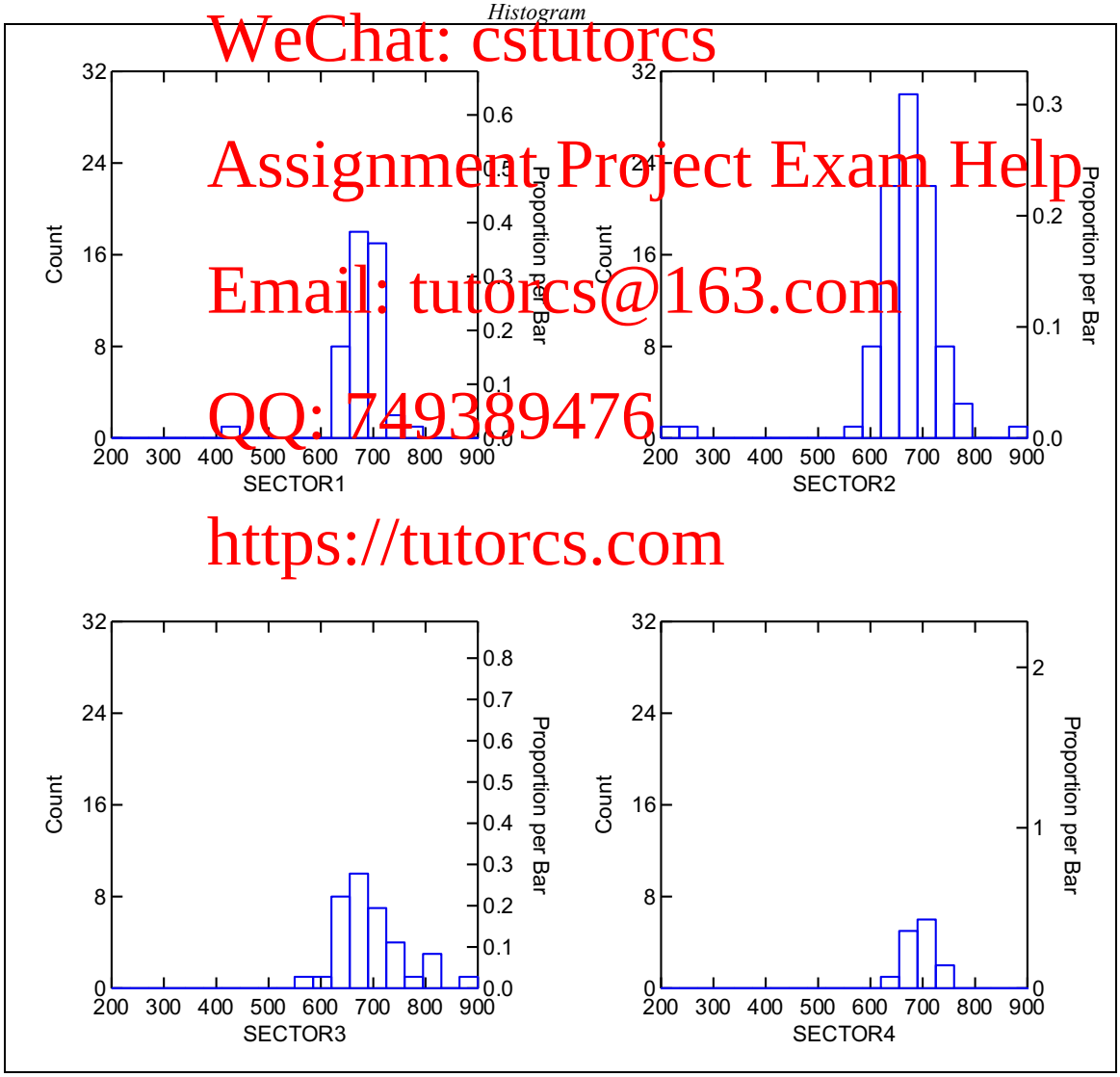
As mentioned previously, outliers can distort the mean making it inaccurate and unusable. Therefore by omitting the extreme outliers mentioned above for Sectors 1 and 2 and by recalculating, we can obtain more precise values for the mean. It was found that the new mean for Sector 1 is 682.27 ($m) and for Sector 2 675.27($m). These values are no longer

affected by outliers are therefore are more accurate, in addition to being closer to their corresponding medians in their sectors.

The first quartile (Q1) is represented by the bottom of the box, third quartile (Q3) is represented by the top of the box, and the horizontal line within the box depicts the median (Q2) for each sect… …ange (IQR) is the length of the box itself, which is Q3-Q1. When looking … …ent that the medians are all fairly similar, Sector 2 being the only exc… …ntly smaller median. The smaller the IQR, the less spread the value … …ectors are and thus the higher the net asset value ($m). It can be se… …Table 1 that Sector 4 has the smallest IQR, 29.50, indicating that 50… …hin this condensed range. In comparison Sector 3 has the largest IQ… …e can infer that Sectors 1 and 4 have the highest net asset value ($m) … …allest degree of spread around a similar median.

Graph 2: Sector1 Histogram

Graph 3: Sector 2 Histogram



Graph 4: Sector 3 Histogram

Graph 5: Sector 4 Histogram

Graphs 2,3,4 and 5 above are histograms of all four manufacturing sectors with the same vertical and horizontal scale. Having the graphs on the same scale facilitates comparison between the sectors, however it also excludes details of individual the distributions. Thus included in the appendix are Graphs 7,8,9, and 10 the histograms of each sector with their own individual scale of best fit.

By looking at Graphs 2 and 3 we can see that sectors 1, 2 and 4 are unimodal and have a fairly symmetrical distribution ignoring the few outliers present. This inference is supported by the fact that their means and medians in each sector are of similar values, for example the mean for sector 4 is 694.75 which is similar to its median 692.53. Thus these sectors are said to be normally dis[illegible] the approximately 68% of companies in the samples have net asset va[illegible]andard deviation of the mean, 95% of companies in the samples have [illegible] + or – 2 standard deviations of the mean and practically all of th[illegible]mples have net asset values within + or – 3 standard deviations of the r[illegible]003) This means that a large amount of the company's net ass[illegible] sectors lie concentrated around their mean values, and thus the net a[illegible]ctor 1, 2 and 4 are higher than that of sector 3; which is not symm[illegible] contrast to the other sectors, Sector 3 is skewed to the right, positively distributed indicating that there are a larger number of companies in the sample with lower asset values ($m). This inference is reinforced by the fact that the mean is larger than the median in this sector, as seen in Table 1.

As Moore (1997) concisely illustrates "The standard deviation and its square, the variance, measures spread by looking at how far the observations are from their mean." The smaller the standard deviation the smaller the spread of the asset values and thus the more concentrated the assets are around the mean. It is visually obvious by looking at the distribution curves of the sectors in the appendix that Sector 4 has the smallest spread followed by sector 1,3 and 2. This point is reinforced by looking at the standard deviations presented in Table 1, which are 29.86, 47.11, 67.37 and 78.35; for sectors 4,1,3 and 2 respectively. Thus we can deduce that the companies in Sector 4 have a largest proportion of asset values close to the mean, followed by Sector 1,3 and 2. Indicating that the net assets in Sector 4 have the highest values overall.

Conclusion:

With the exception of Sector 2, all of the manufacturing sectors display similar mean and median values indicating that they have approximately the same average net asset value ($m). Sector 4 is normally distributed and demonstrates the lowest degree of spread, followed by Sector 1,3 and 2. We can therefore conclude that manufacturing Sector 4, Clothing, has the largest net asset base ($m).

References:

- Selvanathan, A. et al. (2003) *Australian Business Statistics, abridged 3rd edition.* Victoria, Thomson. (p105)
- Moore, D. (1997) *Statistics: Concepts and Controversies, 4th edition.* New York, W.H Freeman and Company (p 252)
- Spatz, C. (1993) *Basic Statistics: Tales of Distributions, 5th edition.* California, Brooks/Cole Publishing Company
- Silver, M. (1997) *Business Statistics, 2nd edition.* London, McGraw Hill Publishing
- Moore, D. (1999) *The Basic Practice of Statistics, 2nd edition.* New York, W.H Freeman and Company

Appendix:

```
Stem and Leaf Plot of variable: SECTOR1, N = 47
         Minimum: 437.260
         Lower hinge: 665.360
         Median: 684.060
         Upper hinge: 700.225
         Maximum: 782.840

              43   7
```

```
      * * * Outside Values * * *
                62   4
                63   15
                64   13579
                65
                66 H 0336
                6
                6
                6
                7
                7
                7
                7
      * * * Out
                7
                7
50 cases with missing values excluded from plot.
```


Figure 1: Sector1 Stem and leaf Plot

```
Stem and Leaf Plot of variable: SECTOR2, N = 97
         Minimum: 218.740
         Lower hinge: 649.780
         Median: 672.740
         Upper hinge: 700.820
         Maximum: 865.170

                2   14
      * * * Outside Values * * *
                5   7
                5   88999
                6   111
                6   22333
                6 H 44444444455555555555
                6 M 66666666667777777
                6   888888888999999
                7 H 000000001111
                7   22223
                7   44445
                7   66
      * * * Outside Values * * *
                7   7
                8   6
```

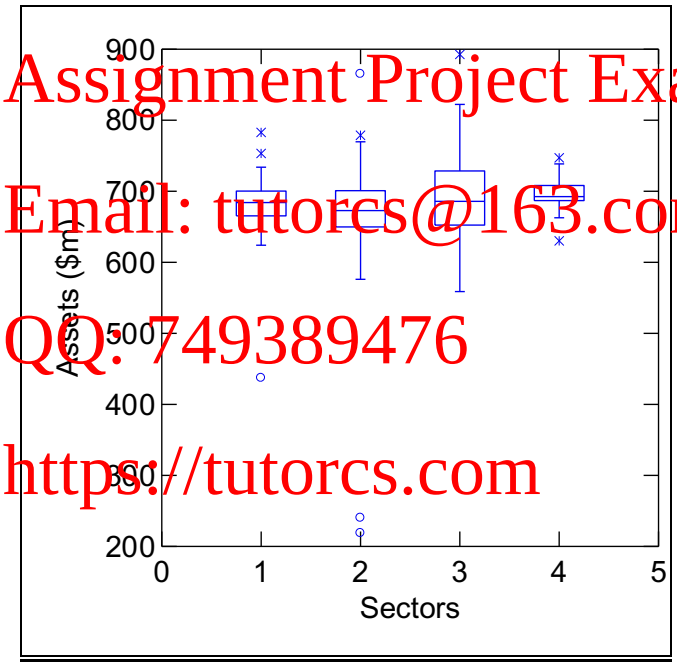

Figure 2: Sector 2 Stem and leaf Plot

```
Stem and Leaf Plot of variable: SECTOR3, N = 36
         Minimum: 558.780
         Lower hinge: 652.415
         Median: 685.745
         Upper hinge: 728.560
         Maximum: 892.650

                5   5
                5
                5
                6   1
                6   2233
                6 H 44555
                6   6677
                6 M 888889999
                7   001
                7 H 3
                7   445
                7
                7   9
                8   01
                8   2
      * * * Outside Values * * *
                8   9
61 cases with missing values excluded from plot.
```

Figure 3: Sector 3 Stem and leaf Plot

```
Stem and Leaf Plot of variable: SECTOR4, N = 14
          Minimum: 630.170
          Lower hinge: 686.930
          Median: 692.590
          Upper hinge: 708.190
          Maximum: 747.040

                6
      * * * Out
                6
                6
                6
                6
                7
                7
                7
                73    8
      * * * Outside Values * * *
                74    7
83 cases with missing values excluded from plot.
```


Figure 4: Sector 4 Stem and leaf Plot



Graph 6: Box Plot showing all outliers.

Graph 7: Sector1 Histogram

Graph 8: Sector 2 Histogram

Graph 9: Sector 3 Histogram

Graph 10: Sector 4 Histogram