FIT1006 Assignment 1

Statistical Analysis and Report Writing

Explanation of data and Assumptions

It was assumed that this data was for the net assets ($m) for a sample of companies in four different manufacturing sectors. Below is a table of a number of descriptive statistics for the 4 different sectors. This table was obtained using the program SYSTAT 12. It has been assumed that each company observed is well established i.e. has not just recently been started, and is not on the verge on bankruptcy

Table 1: Descriptive Statistics acquired from SYSTAT

	Sector 1	Sector 2	Sector 3	Sector 4
Number of Companies	14	102	48	106
Minimum	1178.4	1123.4	1066.9	186.4
Maximum	1425.3	5416.3	2799.0	2054.1
Range	246.9	4292.8	1732.0	1867.7
Median	1272.6	1351.1	1195.2	1121.8
Mean	1281.5	1462.3	1361.1	1134.6
10% Trimmed Mean	1275.5	1394.9	1315.9	1125.6
Standard Deviation	64.0	448.5	330.100	155.0
Variance	4096.6	201226.3	108965.9	24041.8

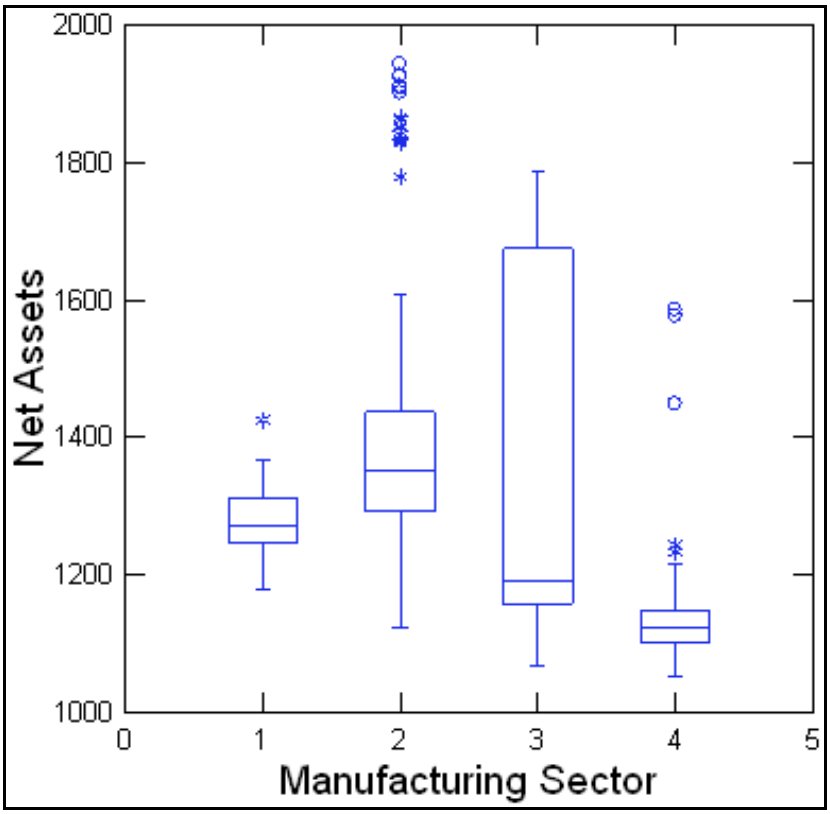
Explanation of Information in Table

There are a few interesting observations we can obtain just by looking at the data in this table. The first of which is the large differences in the number of companies in each sector, with 14 for Sector 1 being the smallest, and 106 in Sector 4 being the largest. This small amount of observations in Sector 1 could lead to the sample not accurately representing the entire population of companies in that Sector, or possibly Sector 1 in its entirety (i.e. not just the sample shown here) could be much smaller than the other three manufacturing sectors. Looking at the Minimum and Maximum values for each Sector, the Maximum for Sector 2 and the Minimum for Sector 4 really stand far away from the bulk of each Sector. These values could affect the mean of the data, but since those two sectors also have the largest number of observations, it doesn't have as large affect as it would have if say, the $5416m was in the Sector 1 group i.e. The more observations in each group, the less affect each individual observation has. To compensate for this, the Trimmed Mean has also been calculated; this removes the outer 10% from each end of the range, and then calculates the mean from the remaining observations. This is so that large outliers are removed so they do not affect the mean so greatly.

The Standard Deviation and Variance that have been calculated can tell us a lot about these sets of data. From the table we can see that Sector 2 and Sector 3 have a very large Standard Deviation. This could be for a number or reasons, possibly because they are bimodal, or maybe because they just have a very spread out distribution. This also means that Sector 1 would have the least spread of net assets

This box plot was obtained using the program SYSTAT 12. Due to the modified Y-Axis scale, a small number of outliers are not shown, but this does not change the basis of my statements, as most of which are derived from the significant features of the box plots which are still shown in figure 1. (See appendix figures 3 and 4 for box plots which include the extreme outliers

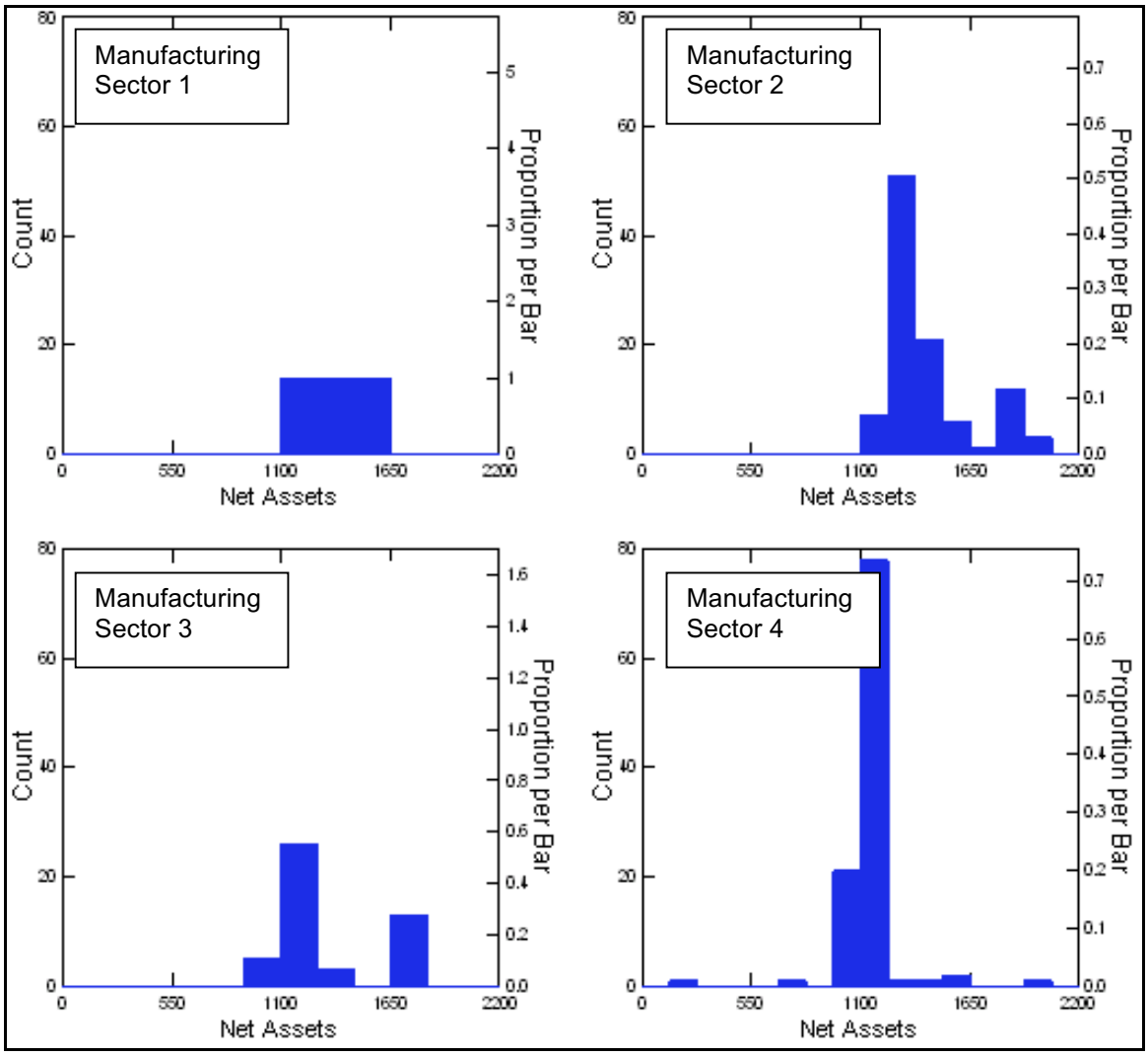
Figure 1: Box Plot of Net Assets for each sector



Explanation of Box Plot

Figure 1 shows a box plot of the net assets of the four sectors of manufacturing companies. From this box plot we can see that Sector 2 has the significantly highest median, with its lower-quartile even being higher than the median of all the other sectors. As seen before with the Standard Deviation, Sectors 1 and 4 had a very low spread, so the inter-quartile range is quite small, as shown by the box plots. Sector 2 has a large amount of outliers and extreme-outliers in the 1800-2000 area. This could be that the distribution for this sector is bimodal. There are also likely to be more observations in the lower peak of the bimodal distribution, causing this to contain the median, and the values from 1800-200 being considered outliers. Sector 3 has a very large inter-quartile range; this could suggest that this Sector's distribution is also bimodal. You will also see that the median of the data is very close the lower-quartile, but quite far away from the upper-quartile, meaning that there are probably more companies situated in the lower area (lower, meaning, under 1200 in net assets in this case). Companies above the median are most likely quite far above it, which would cause the large difference between the upper-quartile range and the median.

Figure 2: Histogram of Net Assets of each Manufacturing sector



Explanation of the Histograms

From these histograms we can sum up everything we have gathered from the descriptive statistics and the box plots. We were able to tell that Sector 2 was bimodal but with a larger number of observations in the lower peak, which as we can see from the histogram, there peaks at ~1300 and ~1900, with the ~1300 peak being much larger, so this makes it also possible to be described as skewed with a long upper tail. We were also able to tell that Sectors 1 and 4 had the majority of their observations around the one point. We can also confirm that Sector 3 was bimodal, with two peaks at ~1200 and ~1750; these peaks could possibly represent sub-groups of companies in the manufacturing sector, but it is uncertain what has caused this sector to be bimodal. If compared back to the table of descriptive statistics, the mean for Sector 3 is quite misleading, as there are very few observations around 1361.

Appendix:

Figure 3 & 4: Box plots with larger Y-Axis scale to show extreme outliers