Anova Lab Report

Jack Ruder and Tatum Bunnett

Oct 30, 2022

1 Introduction

Our report uses data from the top 2000 companies in the Global 2000 list of 2017 from Forbes.com. The dataset includes information about the 2000 companies, such as the company's country, industry, sales, profits, assets, market value, and sector. This report uses the above data to determine if there is a difference in sales between the multiple sectors. Our report will discuss our analysis and provide our conclusions as well.

2 Analysis

To begin determine if there is a difference in sales between the multiple sectors we first needed explored our data through density plots.
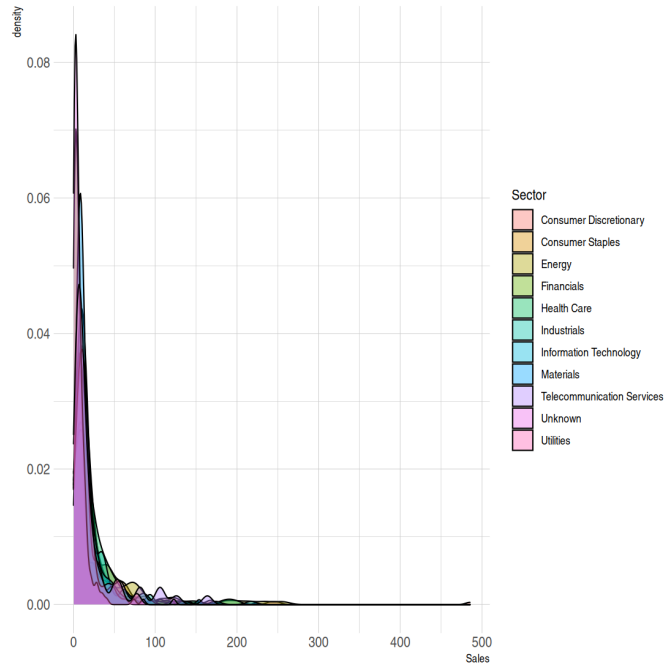
Figure 1: Densities per Sector

The plots in Figure 1 showed that our data was heavily right skewed for all groups, indicating the need for a transformation.

We may check the needed transformation by plotting the log of the group standard deviations versus the log of the group means.
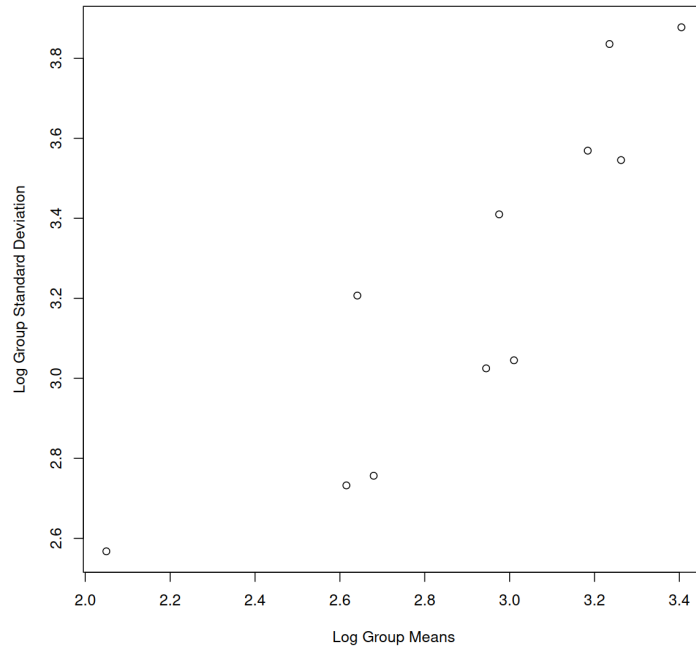
Figure 2: Group Standard Deviations against Means

In Figure 2, the observed slope of 1.01 indicates that we should transform our Sales response variable using a log transform.
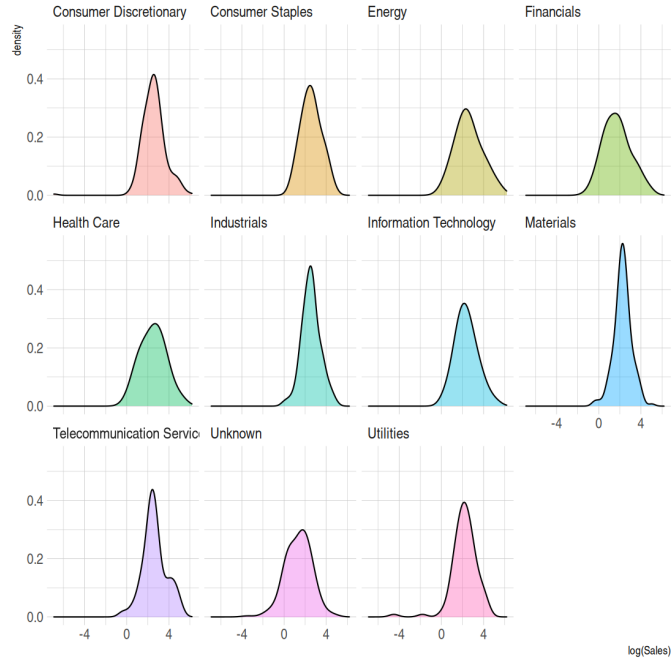
Figure 3: Log Transformed Sales per Sector

The adjusted densities in Figure 3 show a shift towards a normal distribution in each sector as desired. At this stage, it is worth pointing out a sector labeled Unknown, this represents the data where there was no entry for the sector column. We will continue to refer to the sector as Unknown, refraining from interpereting meaning until it becomes apparent.

It appears that the Unknown sector and Financials sectors have peaks in their densities ever-so-slightly shifted from the rest. This is made apparent in Figure 4,
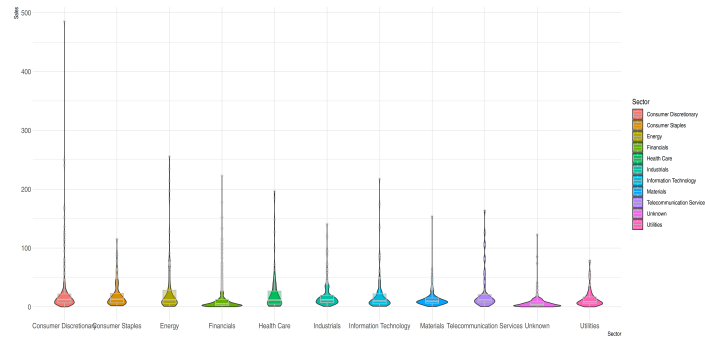
Figure 4: Violin/Box Plot of Untransformed Sales

where it is clear that the distributions are quite different, shifted towards fewer sales.

To test this hypothesis, we run a One-Way Anova using R to determine if there is a significant difference between sector sales. On running a One-Way Anova, we recieve an F value of 27.868 and a p-value of $< 2.2e^{-16}$, indicating that we may accept the alternate hypothesis that there exists a difference in Sales between groups.
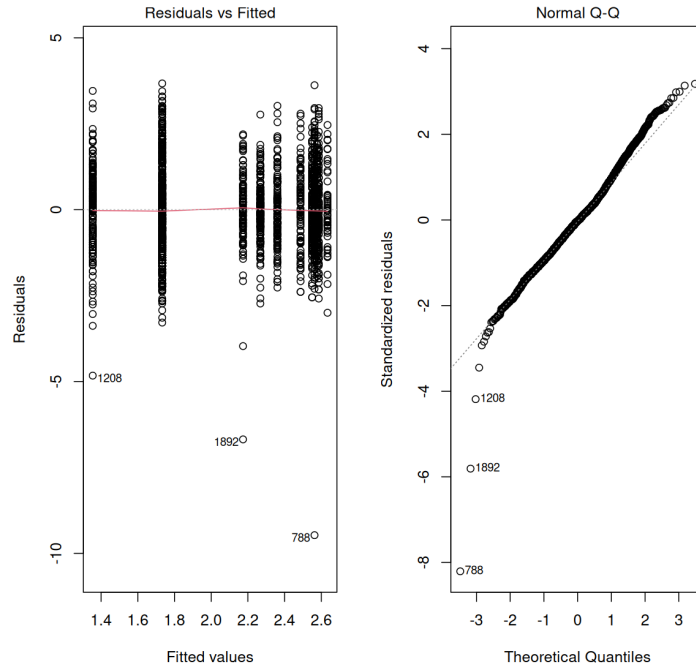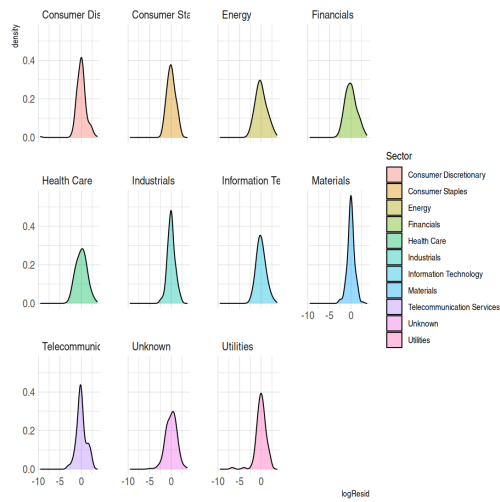
Figure 5: Conditions for Inference

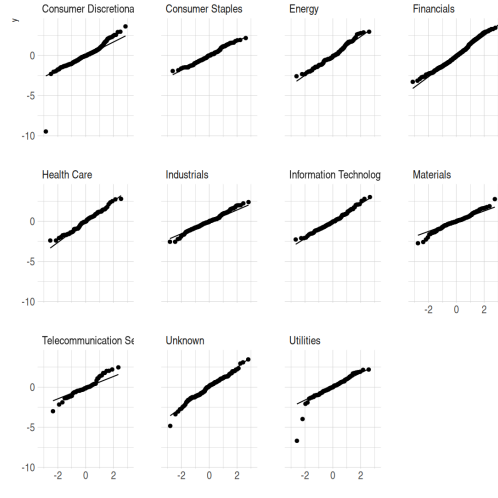Figure 6: Density Plot of Group Residuals

6

Figure 7: QQ plot by Group

In chekcing conditions for inference, we check Figure 5, and confirm that the residuals are random, independent, and normally distributed. In Figures 6 and 7, we see that the residuals within groups have the same variability as well as appearing normal.

A TukeyHSD test shows that the only significant differences between group means lie with either Unknown or Financial sectors. That is, every adjusted p-value for a difference containing either of those two groups is less than 0.05. The adjusted p-value between Unknown and Financials is 0.0035, indicating a significant difference between these two groups. Thus, there are three categories of sectors in terms of sales: Unknown, Financials, and the rest.

With a grand mean of \$8.52 billion, the group effects (in billions of dollars) are

	Group Effects, % change
Consumer Discretionary	152
Consumer Staples	154
Energy	155
Financials	66
Health Care	141
Industrials	150
Information Technology	124
Materials	113
Telecommunication Services	163
Unknown	45
Utilities	103

.

3   Conclusion

In our analysis we determined that there is a significant difference in sales between the different sectors, specifically the financials sector, unknown sector, and the rest. The financials sector posts 66% of the average sales figures, the unknown sector posts 45% of the average sales figures. Likely, this implies that the unknown sector consists of smaller-cap, non-traditional companies who are not as focused on sales. Within the rest of the groups there are not any significant enough differences to claim fundamentally different behavior. As this is purely an observational study we cannot use these results to infer any causation, nor should we extrapolate these figures to any smaller comanies since this is cherrypicked data from Forbes. But, it might be useful in understanding how different sectors among the most successful buisnesses rely on sales for recognition.