Statistical Analysis of Stock Prices in John Wiley & Sons
Rawiyah Alraddadi
rawiyah-2013@hotmail.com

ABSTRACT
John Wiley & Sons is internationally known and is abound in recent years. It is an academic publishing company for books, magazines, encyclopedias, cultural & others and is famous in US. The aim of this paper is to summarize and interpret the given data about stock prices of John Wiley & Sons Company over a period time (almost one year). In the data set, the stock prices such as Open, High, Low, Closing, Volume, and Adjusted Close prices are given.

Keywords: Price, Standard Deviation, Histogram, Stock, Open, Time

1. INTRODUCTION
For many years, the main question which led to controversy is how the past history of a particular stock price helps to make meaningful predictions concerning the present behavior, the future price of the stock? Answer to this question can be given partially by analyzing the descriptive statistics of the stock, such as five number summary, analysis of graphs (Bar chart, Q-Q plot, time series plot etc).

According to the famous theories, it is assumed that the past behavior of a stock price is rich in information concerning its future behavior. According to a proverb “History repeats”, the same way, the patterns of past prices will give some information about the future trends. Thus, if through careful analysis of the stock prices over some periods, one can develop a theory for understanding of these patterns. The more interesting and more unique alerts, however, make use of detailed statistical analysis to watch the market for useful patterns, and to decide which patterns are interesting enough to report.

Particularly, the purpose of this paper is to understand the general behavior of John Wiley stock prices over a period of time and the sudden fluctuations involved along with possible reasons through different descriptive statistics (Mean, five number summary, graphs etc.). It might also be useful to have one or two numbers that capture relevant characteristics of a stock's behavior. Mean and variance are two descriptive statistics often used to summarize data.

Apart from the basic statistics, graphical summaries like, histogram, box plot etc. will give the overall summary of the data and its reliability and trend over some period.

Histogram generally gives the nature of the data and information on the symmetric nature of the data. If the data does not show symmetric nature, it will also give information on which side the data is distributed so that the spread of the data can be understood. Similarly, the stem-and-leaf plots also give the same information. Apart from the nature, the summary of the data graphically as well quantitatively can be found using box-plots. The median, quartiles and the maximum and minimum values can be obtained using box-plot. The time series analysis will be the best tool for forecast and also to predict the trend. The trend chart will provide adequate guideline for the investor. Some time it may not address or forecast the variations or steady flow of the market. It may forecast only on particular season, but it is not adequate for long term decision making.

2. METHODOLOGY
- Using the descriptive statistics, the nature and the shape of the distribution is estimated.
- Using the five-number summary of different stock prices, the maximum and the minimum along with the median and mean values are obtained.
- The shape and symmetric nature of different prices are checked using box-plots and histograms.
- The normality of the data is checked using Q-Q plot and histograms.
- The time series plots are used to identify the trend of the stock prices.

3. DESCRIPTION OF DATA
- The data given is from 3rd June, 2013 to 1st May, 2015.
- The Open stock price indicates the price of the stock at the opening trade.
- The High price indicates the highest price that a stock has attained over the day.
- The Low stock price indicates the lowest price of the stock over the entire day.
- According to trade wave, the close price is the closing price of the stock on a particular day.
- The total of all trade amounts is noted as Volume price.
- According to the historical returns analysis, the price of the stock is adjusted and the same is denoted by "Adjusted closing price" according to Investopedia
4. ANALYSIS OF DATA

4.1 Measures of Central Tendency

<table>
<thead>
<tr>
<th>Table 1: Measures of Central Tendency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
</tr>
<tr>
<td>1st Q.</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>3rd Qu.</td>
</tr>
<tr>
<td>Max.</td>
</tr>
</tbody>
</table>

- The five point summary along with the mean of each stock price is tabulated. From the table, it can be observed that the mean and median for all stock prices are approximately equal except for volume price.
- For volume price, the mean value is more than median indicating that the volume price data is right-skewed.
- Observing the values for all types of prices, it can be concluded that the difference between various prices seems to be almost similar.

4.2 Measures of Variability

<table>
<thead>
<tr>
<th>Table 2: Measures of Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Open Price</td>
</tr>
<tr>
<td>High Price</td>
</tr>
<tr>
<td>Low Price</td>
</tr>
<tr>
<td>Close Price</td>
</tr>
<tr>
<td>Volume Price</td>
</tr>
<tr>
<td>Adj.Close Price</td>
</tr>
</tbody>
</table>

- From the table, observe that the ranges of all prices are approximately same ranging from 19.46 to 20.63. The standard deviation of volume price is highest with a value of 149654.6
- All the measures of dispersion (range, standard deviation and variance) give the same information that the variations in all kinds of prices are almost same except for volume price.
- The measures of variability give the basic information on distance/difference from the mean price.
- Comparing the highest and lowest prices, observe that the difference between High and Low prices seem to be very less. The variation between these prices is not very high. This indicates that the fluctuation in the prices of the stock are not too erratic.
4.3 Plots

4.3.1 Box Plots

Fig 1: Box plots
• The box-plot of open and close prices are almost similar with approximately equal median value.
• The box-plot of High price and Low price are seem to be symmetric, but the median of low price stocks is less than that of high price.
• The box-plot of Adjusted closed price of stocks are absolutely symmetric.
• The variability of volume price is very high
• The box-plot of volume price contains more outliers towards the right side. Also, the area after median line is slightly more than the area below median. Both these indicates that the data of volume price is right-skewed.
• As per the 1.5IQR rule, there exists almost 5 outliers in the data.

4.3.2 Stem-and-Leaf Plots:

![Fig 2.1: Stem-and-Leaf of Open Price](image)

![Fig 2.2: Stem-and-Leaf of High Price](image)

![Fig 2.3: Stem-and-Leaf of Low Price](image)

![Fig 2.4: Stem-and-Leaf of Close Price](image)

![Fig 2.5: Stem-and-Leaf of Volume Price](image)

![Fig 2.6: Stem-and-Leaf of Adj.Close Price](image)
Basing on the stem-and-leaf plots of all stock prices, it can be observed that the nature of all stem plots is similar except for volume price.
- The stem plot of volume price seems to be right-skewed.
- Also, observe that almost all prices are bimodal.

### 4.3.3 Q-Q Plots

![Normal Q-Q Plot of Open Price](image1)

Normal Q-Q Plot of Open Price

![Normal Q-Q Plot of High Price](image2)

Normal Q-Q Plot of High Price

![Normal Q-Q Plot of Low Price](image3)

Normal Q-Q Plot of Low Price

![Normal Q-Q Plot of Close Price](image4)

Normal Q-Q Plot of Close Price
• The Q-Q plots for all stock prices are almost similar except of volume price. In all Q-Q plots, majority of the points are close to straight line. But, few outliers are also visible.
• In Q-Q plot of volume price, very few points are close to straight line and majority of the points are towards the upper side of the straight line which indicates that the data is right-skewed.

4.3.4 Histograms
4.3.5 Time Series Plots

- The histograms also give the same conclusion as Q-Q plots.
From the time series plots, it can be observed that except volume price, remaining all prices have shown increasing trend.

The volume price time series plot, shows some constant trend.

5. FINDINGS

- On average, the open stock price was at $49.87. The minimum open stock price was $38.98 whereas the maximum open stock price was $59.00.
- The high stock price has an average value of $50.39 whereas the maximum and the minimum
highest stock prices were $59 and $39.54 respectively.

- The low stock price has an average value of $49.45 with lowest value $37.42 and highest value of $58.04.
- The mean of the volume price was $212,575 with highest volume price as $1,421,300 and the lowest volume price as $45,400.
- The adjusted close price of stocks has a mean value of $49.55 with lowest value as $38.2 and the highest as $58.83.
- The box-plot of open and close price seems to be similar. But the maximum price of open stock is slightly higher than close price.
- The box-plot of Adjusted close price of stocks is also symmetric.
- The box-plot of volume price indicates that there exists many outliers.

6. CONCLUSION

- Compare the mean of the adjusted close price of stocks with mean of high price and low price. It can be observed that the adjusted value is between high and low stock prices.
- The standard deviations of all stock prices are almost similar and symmetric except for volume price.
- The median prices of open and close stocks are approximately similar.
- The box plots of all stock prices are seem to be symmetrical, (that is mean=median) except for volume price. Because of the outliers, the volume price can be concluded as right-skewed.
- By comparing the Q-Q plot and histograms of all prices, it seems that all prices are symmetrical and follows Normal distribution except for Volume price.

REFERENCES


