# Graphical Analysis of Simulated Financial Data using R

## Shubhi Gangele[1], Pamila Koli[2], Prof.Nilima Nikam[3]

*Dept. of Computer Engineering, YTIET, Chandhai, Karjat-410201.*

-------------------------------------------------------------------***-------------------------------------------------------------------

**Abstract**-*We have particularly seen in case of Indian markets that they are not efficient. When a good event or bad event happens, the market valuation of a particular product or company is slow to respond. It may take from hours to days for the market price to reach its actual valuation. Even with publicly available information like financial sheets, there seems to be an delay in market adjusting to the actual valuation of the company. Keeping this in mind we have developed a tool which would calculate basic financial ratios to help investors realize the market value of the company with publicly available information.Two of the tools we have developed are DuPont analysis on financial data available from a company in London. Similar we have developed ANOVA analysis on simulated data.Our simulation can generate random numbers in huge amounts & then perform a analysis on it.The following paper demonstrates the process & rationale behind doing it.*

*Key Words:*Dupont analysis, Big data,R,Anova analysis.

## 1.Introduction

Financial analysis refers to an assessment of the viability, stability and profitability of a business, sub-business or project. It is performed by professionals who prepare reports using ratios that make use of information taken from financial statements and other reports. These reports are usually presented to top management as one of their bases in making business decisions.

Financial Statements are used for a Multitude of Different Purposes.Readers of a financial statement are seeking to understand key facts about the performance and disposition of a business. For example, return on assets (ROA) is a common ratio used to determine how efficient a company is at using its assets and as a measure of profitability. This ratio could be calculated for several similar companies and compared as part of a larger analysis.

### 1.1 Graphical Analysis

Graphical analytics is "the science of analytical reasoning facilitated by interactive visual interfaces." It can attack certain problems whose size, complexity, and need for closely coupled human and machine analysis may make them otherwise intractable.Visual analytics advances science and technology developments in analytical reasoning, interaction, data transformations and representations for computation and visualization, analytic reporting, and technology transition.

As a research agenda, visual analytics brings together several scientific and technical communities from computer science, information visualization, cognitive and perceptual sciences, interactive design, graphic design, and social sciences.

### 1.2 About R Programming

R and its libraries implement a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others. R is easily extensible through functions and extensions, and the R community is noted for its active contributions in terms of packages. Many of R's standard functions are written in R itself, which makes it easy for users to follow the algorithmic choices made. For computationally intensive tasks, C, C++, and Fortran code can be linked and called at run time. Advanced users can write C, C++, Java, .NET or Python code to manipulate R objects directly. R is highly extensible through the use of user-submitted packages for specific functions or specific areas of study. Due to its S heritage, R has stronger object-oriented programming facilities than most statistical computing languages.Another strength of R is static graphics, which can produce quality graphs, including mathematical symbols. Dynamic and interactive graphics are available through additional packages.
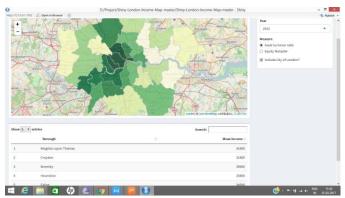
## 2. Analytical Techniques

### 2.1 Dupont Analysis

DuPont analysis is a method of performance measurement that was started by the DuPont Corporation in the 1920s. With this method, assets

are measured at their gross book value rather than at net book value to produce a higher return on equity (ROE). It is also known as DuPont identity.

Return On Asset(ROE) is a closely-watched number among knowledgeable investors. It is a strong measure of how well a company's management creates value for its shareholders. The number can be misleading, however, as it is vulnerable to measures that increase its value while also making the stock more risky. Without a way of breaking down ROE components, investors could be duped into believing a company is a good investment when it's not.



The asset turnover ratio is an efficiency ratio that measures a company's ability to generate sales from its assets by comparing net sales with average total assets. In other words, this ratio shows how efficiently a company can use its assets to generate sales.
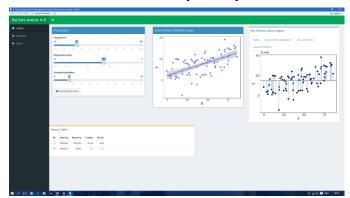
## 2.2. Anova Analysis

Analysis of variance (ANOVA) is an analysis tool used in statistics that splits the aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, but the random factors do not. Analysts use the analysis of the variance test to determine the result independent variables have on the dependent variable amid a regression study.
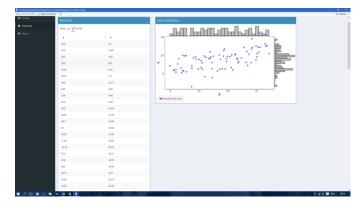
The two-way ANOVA compares the mean differences between groups that have been split on two independent variables (called factors). The primary purpose of a two-way ANOVA is to understand if there is an interaction between the two independent variables on the dependent variable. For example, you could use a two-way ANOVA to understand whether there is an interaction between gender and

educational level on test anxiety amongst university students, where gender (males/females) and education level (undergraduate/postgraduate) are your independent variables, and test anxiety is your dependent variable. Alternately, you may want to determine whether there is an interaction between physical activity level and gender on blood cholesterol concentration in children, where physical activity (low/moderate/high) and gender (male/female) are your independent variables, and cholesterol concentration is your dependent variable.



The interaction term in a two-way ANOVA informs you whether the effect of one of your independent variables on the dependent variable is the same for all values of your other independent variable (and vice versa). For example, is the effect of gender (male/female) on test anxiety influenced by educational level (undergraduate/postgraduate)? Additionally, if a statistically significant interaction is found, you need to determine whether there are any "simple main effects", and if there are, what these effects are.

## 3.Methodology

There some assumptions to do Two way ANOVA or we can say that these are the conditions for Two way ANOVA

- Assumption #1: Your dependent variable should be measured at the continuous level (i.e., they are interval or ratio variables).
- Assumption #2: Your two independent variables should each consist of two or more categorical, independent groups.
- Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves.
- Assumption #4: There should be no significant outliers. Outliers are data points within your data that do not follow the usual pattern
- Assumption #5: Your dependent variable should be approximately normally distributed for each combination of the groups of the two independent variables.
- Assumption #6: There needs to be homogeneity of variances for each combination of the groups of the two independent variables.

Step 1) Identify Hypothesis to be tested

Step 2) Formula for calculation of SumofSquare(within) is:

Y(ij) are the elements in the groups.Y¯(ij) is mean of combination

$$SS_{within} = \sum_{i=1}^{2}\sum_{j=1}^{3}\sum_{k=1}^{4}\left(Y_{ijk} - \bar{Y}_{ij.}\right)^2$$

Step 3) Calculate the df(within):

df(within) = (r-1)*a*b = 3*2*3 = 18

Step 4)Calculate MeanSquare(within):

MS(within) = SS(within)/df(within) = 38/18 = 2.1111

Step 5) Calculate  F value

F variable is the ratio of two independent chi-square variables divided by their respective degrees of freedom. The degrees of freedom for the numerator are the degrees of freedom for the between group (k-1) and the degrees of freedom for the denominator are the degrees of freedom for the within group (N-k).

$$F = \frac{S_b^2}{S_w^2}$$

## 4. Conclusion

Our study on simulated data & DuPont analysis have helped us to develop a sound understanding of how the financial ratios behave. The results from simulated data are not limited to Indian markets; can be applied on various markets  & seen in the current stock market to determine how inefficient it is. We believe investors who will invest in long term portfolios will benefit from the methodology which we have proposed.

### REFERENCES
[1] 2013 IEEE Third International Conference on Cloud and Green Computing. Available at: http://ieeexplore.ieee.org/abstract/document/6686088/?reload=true
[2] Improving Decision Making in the World of Big Data. Available at:

http://www.forbes.com/sites/christopherfrank/2012/03
/25/improvingdecision-making-in-the-world-of-big-data/
[Access Date: August 11, 2013]
[3] How Big Is Facebook's Data? Available at:
http://techcrunch.com/2012/08/22/how-big-is-
facebooks-data-2-5-
billion-pieces-of-content-and-500-terabytes-ingested-
every-day/ [AccessDate: August 11, 2013]
[4] Big Data Makes a Big Difference in Marketing. Available
at:
http://www.spiral16.com/blog/2012/09/big-data-
makes-a-big-difference-
in-marketing/ [Access Date: August 11, 2013]
[5] Unleash the power of open source R. Available at:
http://www.revolutionanalytics.com/what-is-open-
source-r/
[AccessDate: August 11, 2013]