**BIOE xxxx HW xxx**

**Technical Stock Analysis with Analysis of Variance (ANOVA)**

Closing stock prices a for few biotech and pharmaceutical companies is provided in the attached data set. Here we will perform analysis to determine if these stocks differed in their technical performance (return) from the time period January 1st, 2012 to December 31st, 2015 using ANOVA.

Preliminaries

Price changes are equal to returns. We’re not interested in the absolute changes in stock prices only relative percent price changes from close to close. To convert to percent price change, we used the formula,

$$Δ\_{day}≜DailyPercentChange$$

$$Δ\_{day}=\frac{Today^{'}sClosingPrice}{Yesterday^{'}sClosingPrice}=\frac{Price[n]}{Price[n-1]}$$

A long term investor is may be more interest in yearly return rather than daily returns. A yearly percent return is a product of the changes from each day, which is approximately equal to the product of the mean value of $Δ\_{day}$ taken to the same power.

$$YearlyReturn=Π\_{n=1}^{N\_{sessions}}Δ\_{day}[n]≊mean(Δ\_{day}[n])^{N\_{sessions}}$$

Where $N\_{sessions}$ is the number of trading sessions in a calendar year.

We care about if the daily percent changes are the same or different amongst the stocks included in the dataset. If one has higher mean daily percentage changes we may wish to invest in it.

1. If we were to use anova for our analysis of this stock data what would the Null hypothesis be? State the null hypothesis. (1 Sentence)

Bonus: Can you think of any other questions that we could ask from this data with statistical hypothesis testing. State the null hypothesis for your question.

2. To use Anova we need to determine if $Δ\_{day}$ has a normal distribution. The template matlab script multiVariableComparsion.m has been included in the .zip file associated with this assignment. This script will load the data.

Work with the data by creating plots and code in section 2. qqplots and boxplots are possible graphical descriptions. Include some of these plots in your assessment of if $Δ\_{day}$ is normally distributed for these stocks.

>> help qqplot

>>help boxplot

3. If the data is not normally distributed we have other options. A Kruskal Wallis test is a non parametric version of an anova. Perform appropriate (based on answer to question 2) statistical test to determine if these the daily price changes are the same or different. Select different combinations of the stocks to include by changing the 1:end index to different subsets of the variables. Which stocks different in mean daily return? State your conclusions with p values. “We reject the null hypothesis that Your Text Here with *Your Text Here*.”

Bonus: What is different about those/that company?

4. What is the problem (error type?) in following the anova up with 1:1 t-test comparisons? What is **post hoc** analysis? Search the web and provide short description what post hoc analysis is relative to ANOVA. Should we perform any of these **post hoc** techniques in addition to what we did in #3? If so what would be the hypotheses to test?

5. In one or 2 sentences, what do you conclude from this technical stock analysis? Would one of these stocks be better to invest in than another?

6. Summarize what ANOVA is and for what kind of data you would need to use it for analysis.

7. Bonus: Repeat the Matlab Analysis in R. Completion will be assessed by executable R script. You are welcome to use your favorite statistical package for this entire homework exercise if you don’t want to use Matlab.

Further Reading

[1] - B. G. Malkiel, *A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing*, Ninth Edition edition. New York: W. W. Norton & Company, 2007.