## Unit 1 Capstone by John Foxworthy

https://github.com/sarilacivert/Data-Science/blob/master/Unit1capstone.ipynb

### Executive Summary

What is the behavioral difference between a short term loan and a long term loan? We propose to test one of the behaviors between shorter interest rate debt to longer interest rate debt by capturing the attribute of stability. A hypothesis will surmise the behavioral stability by testing some of the population of the data with present information commonly available to the public. Test results will then be weighted with shortcomings to ascertain further actions.

### Introduction

The dataset that we will be using are the Daily U.S. Treasury Yield Curve Rates and its significance is the global importance in the financial trading industry, data quality and its statistical significance. The figures below have daily dates as rows with maturity points as columns. For example, a ten year is 2.97 percent as of the 1st of May, 2018, which is the amount of the interest rate on a decade long loan. The data goes back to 1990 up until to the current date in 2018.

**Daily Treasury Yield Curve Rates**

Get updates to this content.

XML These data are also available in XML format by clicking on the XML icon.

XSD The schema for the XML is available in XSD format by clicking on the XSD icon.

If you are having trouble viewing the above XML in your browser, click here.

To access interest rate data in the legacy XML format and the corresponding XSD schema, click here.

**Select type of Interest Rate Data**

Daily Treasury Yield Curve Rates Go

**Select Time Period**

2017 Go

| Date | 1 Mo | 3 Mo | 6 Mo | 1 Yr | 2 Yr | 3 Yr | 5 Yr | 7 Yr | 10 Yr | 20 Yr | 30 Yr |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 05/01/18 | 1.68 | 1.85 | 2.05 | 2.26 | 2.50 | 2.66 | 2.82 | 2.93 | 2.97 | 3.03 | 3.13 |
| 05/02/18 | 1.69 | 1.84 | 2.03 | 2.24 | 2.49 | 2.64 | 2.80 | 2.92 | 2.97 | 3.04 | 3.14 |
| 05/03/18 | 1.68 | 1.84 | 2.02 | 2.24 | 2.49 | 2.62 | 2.78 | 2.90 | 2.94 | 3.02 | 3.12 |

https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield

Each of the figures represents hundreds of billions of U.S. dollar trades from participants from dozens of countries and have the impressive data attribute of a complete absence of stale data. Not some, but all figures, on each business day of the year, has at least one person, and likely many people trying to influence the interest rate values to change. Specifically, the strong liquidity, in other words, the quantity of buyers finding sellers and vice versa is very high, in every trading day, in the year going back decades since the U.S. Dollar is the world's reserve currency. According to organizations like the International Monetary Fund (IMF), 6 out of 10 or 7 out of 10 transactions by physical hand or electronically, involve the U.S. Dollar. There are more than 200 countries in the world and each have some type of exporting and importing business that does involve the U.S. Dollar. The more a country exports, the more their currency, not the U.S. Dollar, rises in value because of increased demand, thereby making the value of their exports more expensive. To counteract, countries buy U.S. Dollars to dampen their local currency appreciation and the most common format is in the U.S. loan or fixed income market. Borrow a month at 1.68% or a year at 2.26% in U.S. dollars on the 1st of May for your exporting business from the Daily Treasury Yield Curve, for instance. Not to mention, the U.S. economy is the largest in the world so there are plenty of individuals and institutions that depend on this data to raise money for countless causes from a mortgage loan for a residential property to purchasing a new office building for an expanding company or local government institution.

Lastly, the source of this data is the U.S. Treasury department of the U.S. government. There are no impediments for the use of this data as it public from a legal perspective and officially published by the U.S. government on every business day. The Daily Treasury Yield Curve is widely used by countless people and institutions for many causes. There is little to no data quality issues with the chosen data set.

```
In [1]: import json
        import matplotlib.pyplot as plt
        import pandas as pd
        import numpy as np
        %matplotlib inline

In [2]: sample_json_df = pd.read_json('/Users/lacivert/2014_2017.json')

In [3]: #assign DataFrame
        df = pd.DataFrame(sample_json_df)

In [5]: #list column names
        list(df.columns.values)

Out[5]: ['1 mo', '1 yr', '10 yr', '30 yr', '6 mo']
```

## Problem Statement

There is an assumption that the entire dataset is more similar than different because it is grouped together as one compound object, a yield curve. The eleven maturity points all have figures that may differ, but move together with the same behavior. Perhaps people are overwhelmed by the quantity of data and desire to feel safe with an assumption of similarity along with the large quantities of participants that buy and sell in the global markets. Nonetheless, a deep dive into data may shed some light on behavioral differences such what are the interest rates on page 1 used for in the financial markets. In the U.S., it is common for residential purchases of houses to be accompanied with a 30 year mortgage loan so is that the same as an entrepreneur getting a 1 year business loan for workspace office or factory? The interest rate numbers may differ because the risk taking of a homeowner is less than a business owner, but are they both stable? If we have a week or month window to purchase a loan product, then a stable market will most likely provide a decent interest rate as it will change little. A less stable interest rate would require urgency for loan buyers to get the best value as the daily changes will roller coaster expectations.

## Proposal

Let us provide some type of outline of the yield curve behavior into short and long term maturities and applying some analysis for any differences. The buyers of interest rates may drive the level of stability of interest rates by taking a look into its averages, distributions and how a slice of the data may describe the overall population. We begin by selecting the data set from page 1 as there are roughly 250 business days in a given year and there are about 6,750 business days for this single column only from 1990 to 2017. The 11 columns from the 1 month to the 30 year would be more than 75,000 interest rate daily data from 1990 to the current day. For this report, we will use daily yield curve rates from 2014 to 2017, which is 1,000 data points per column maturity point for a total of 5,000 of the 1 month, 6 month, 1 year, 10 year and 30 year. We propose the data set will provide an outcome of findings on a hypothesis that long term maturities of 10 year and 30 year, for example, are more stable than the short term maturities. In other words, the entire yield curve is not one compound object and behavior. Conversely, the null hypothesis will be long term maturities are not more stable than the short term maturities because the entire yield curve is one compound object and behavior. Lastly, the proposed analysis has no test version, i.e. no A / B testing as the environment does not require any intervention of the interest rate products in a government controlled product set with A / A testing only.

## Evaluation

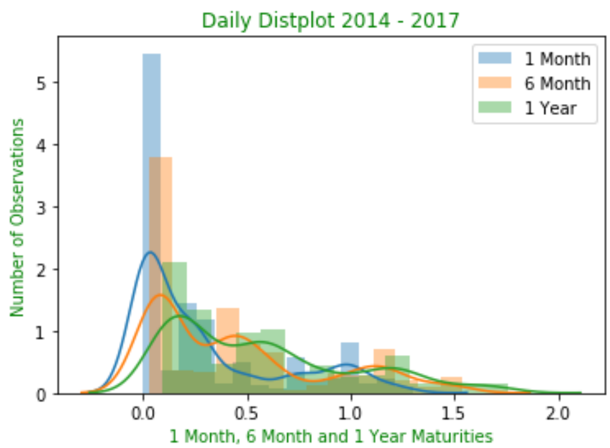
If we take the average of each of the 1 month, 6 month, 1 year, 10 year and 30 year maturity points, then we see a progression below as loan length increases. Later, we will use mean values to ascertain the sample to the population, but for now the distribution adds to the expectation of short and long term maturity differences.

```
In [37]: np.mean(one_month),np.mean(six_month),np.mean(one_year),np.mean(ten_year),np.mean(thirty_year)

Out[37]: (0.2924599999999999,
          0.4409699999999998,
          0.5654500000000011,
          2.2103299999999972,
          2.9158199999999996)
```

In [42]:
```
sns.distplot(one_month,label='1 Month')
sns.distplot(six_month,label='6 Month')
sns.distplot(one_year, label='1 Year')
plt.xlabel('1 Month, 6 Month and 1 Year Maturities',color='green')
plt.ylabel('Number of Observations',color='green')
plt.title('Daily Distplot 2014 - 2017', color='green')
plt.legend()
```

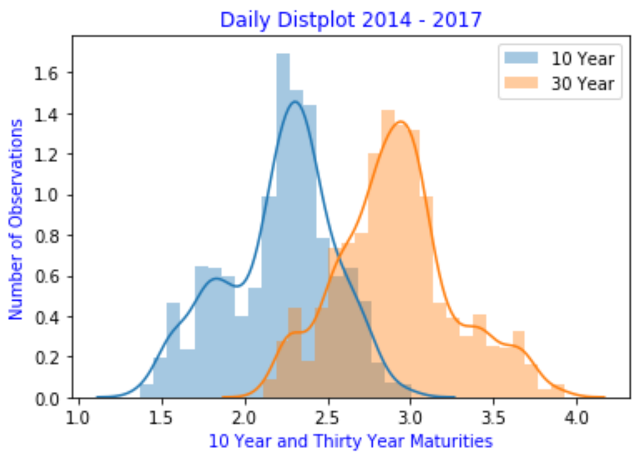
Out[42]:
```
<matplotlib.legend.Legend at 0x1144da2b0>
```




The seaborn distplot distribution has a pattern of shorter maturities clustering together with more observations skewed to one side then smoothing out away from zero. All three short term maturities cluster together and different from the longer maturities on the next page. This is a visual description to evaluate the hypothesis by counting the number of observations, the frequency or the density of the column maturity vectors on page 1 by sampling the population. If there are more observations repeating, clustering and skewing the distribution to one side, then it implies a different form of stability level than a distribution that resembles a normal distribution. The longer maturities of 10 year and 30 year interest rates of the U.S. yield curve resemble two tails and therefore imply more stability, plus they cluster together with an overlap in the middle of the seaborn distplot. Visually, there appears to be more statistical noise in short term maturities than long term maturities.

```
In [44]: sns.distplot(ten_year,label='10 Year')
         sns.distplot(thirty_year,label='30 Year')
         plt.xlabel('10 Year and Thirty Year Maturities',color='blue')
         plt.ylabel('Number of Observations',color='blue')
         plt.title('Daily Distplot 2014 - 2017', color='blue')
         plt.legend()

Out[44]: <matplotlib.legend.Legend at 0x1166ce470>
```




Continuing in the analysis, if we take a look at means of each maturity, then we can use the t - statistic to analytically demonstrate the statistical noise pattern to see if the sample yields a difference of the yield curve population. The calculated difference in units of standard error, the t - statistic is formally the following.

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_1^2/N_1 + s_2^2/N_2}}$$

```
In [12]: import scipy.stats as stats

In [13]: t_1mo_6mo, p_1mo_6mo = stats.ttest_ind(one_month,six_month)
         print (t_1mo_6mo, p_1mo_6mo)

         -8.551737599903504 2.3722861361784687e-17

In [14]: t_1mo_1yr, p_1mo_1yr = stats.ttest_ind(one_month,one_year)
         print (t_1mo_1yr, p_1mo_1yr)

         -15.351486855574468 2.3087699628424497e-50

In [15]: t_6mo_1yr, p_6mo_1yr = stats.ttest_ind(six_month,one_year)
         print (t_6mo_1yr, p_6mo_1yr)

         -6.516784278003866 9.067179116853322e-11

In [16]: t_10yr_30yr, p_10yr_30yr = stats.ttest_ind(ten_year,thirty_year)
         print (t_10yr_30yr, p_10yr_30yr)

         -46.54774142628523 5.306e-321
```

The findings of the t – statistics yields a progression away from zero as the loan length increases. The 1 month and 6 month t – statistic is about – 8.55 and the 10 year and 30 year t – statistic is about – 46.54. The magnitude of the t – statistic grows from short to long term interest rate loans in the four year daily sample of the yield curve population. In addition, the probability value or p – value, of each of data sets are all near zero, which is less than 0.05 so we can reject the null hypothesis that the entire yield curve is one compound object of behavior and the attribute of stability is the same. Short maturities and long maturities are different products in terms of stability, even though all maturities are part of the same yield curve. A car loan with a ten year debt or a house for a family with a thirty mortgage are stable products compared to shorter term loans by small business owners. Cars and houses are conservatives investments compared to an entrepreneurs' ideas. What the interest rates are used for, contributes to driving stability differences on both sides of the yield curve.

## Resulting Actions

A four year sample slice gives us some picture with the 5,000 data points from 2014 - 2017, but there are limitations. We could follow a frequentist approach and sample again to see if we have come to the same conclusion, but there is a selection bias. If you go back to history on the link on page 1, to the year of 1997, then you will see none of the data is near zero, but instead greater than five.

**Daily Treasury Yield Curve Rates**

Get updates to this content.

XML These data are also available in XML format by clicking on the XML icon.

XSD The schema for the XML is available in XSD format by clicking on the XSD icon.

If you are having trouble viewing the above XML in your browser, click here.

To access interest rate data in the legacy XML format and the corresponding XSD schema, click here.

**Select type of Interest Rate Data**

Daily Treasury Yield Curve Rates Go

**Select Time Period**

1997 Go

| Date | 1 Mo | 3 Mo | 6 Mo | 1 Yr | 2 Yr | 3 Yr | 5 Yr | 7 Yr | 10 Yr | 20 Yr | 30 Yr |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 01/02/97 | N/A | 5.19 | 5.35 | 5.63 | 5.97 | 6.13 | 6.30 | 6.45 | 6.54 | 6.85 | 6.75 |
| 01/03/97 | N/A | 5.17 | 5.34 | 5.60 | 5.95 | 6.11 | 6.28 | 6.42 | 6.52 | 6.84 | 6.74 |

Does the level of interest rates contribute to the attribute of stability? When taking out a loan as a homeowner or entrepreneur is more expensive, then does it alter our hypothesis and findings? Further work is required to strengthen the hypothesis expectation that each column maturity vector is a single compound object of behavior or not in one of the most commonly used financial markets in the world.