5

T-TEST,F-TEST, TEST OF EQUAL MEAN, TEST OF EQUAL VARIANCE

In finance, we could estimate the mean return for a given ticker (stock) easily. For example, we could download the daily stock price data from Yahoo!Finance for Wal-Mart. If based on the last 10 years daily price data (from 3/23/2007 to 3/24/2017), the mean daily return is 0.032%. Is this estimate statistically different from zero? From Chapter 5: Risk vs. Return, we have learnt that standard deviation of stock returns could be used to measure the total risk. Is this type of risk, expressed by the standard deviation of stock returns, a constant over time? How to answer a simple question: what is the expected annual return for IBM next year with certain confidence interval? Assume that we have last 50 years annual returns for the company. Could we simply offer the mean as our answer? How to claim that two stocks have different risk levels? In many cases, we have to conduct certain types of tests to confirm or reject our hypotheses. In other words, we should not simply offer an answer before conducting such a test. In this chapter, the following topics will be covered:

- Size matter: sample vs. population
- Two important values for T-value, p-value
- How to estimate standard error
- What is the expected return for IBM next year?
- descriptive statistics, covariance and correlation
- Benford Law: the first digit law
- How to run a linear regression?
- F-test for equal variances
- T-test: two samples assuming equal variances
- T-test: two samples assuming unequal variances
- How to generate various types of random numbers

5.1 INTRODUCTION

In the real world, researchers and students could ask many questions. For example, if the estimated mean return is 0.1%, could we claim this value is the same as zero or statistically different from a zero? In other words, how to test the hypothesis that the volatility of a stock's returns is constant? To conduct a test, usually we have a prior, or hypothesis. Below is a typical example.

Hypothesis: IBM's monthly returns follow a normal distribution (1)

This is called null hypothesis. Most of times, the alternative hypothesis is offered as well. For the above hypothesis the opposite is that the IBM's monthly stock returns don't following a normal distribution. If our testing results support the null hypothesis, we would accept the null hypothesis, i.e., the null hypothesis is supported. Otherwise, we would reject the null hypothesis and accept the alternative one. In the following table, a partial list is offered for several most frequently asked questions.

Table 5.1 A partial list for several most frequently asked questions

Questions
What is the expected annual return of S&P500 next year?
How to measure the total risk for a given stock?
What do they mean: sample and population?
If the average daily return is 0.01% per day, does it the same as zero?
What does it mean that the expected annual return is 10% with a range of -2% to 10%?
How to conduct a normality test?
What is the critical T-value we should remember?
What is the critical p-value we should remember?
What is F-distribution?
How to test the equal-means
How to run a regression for a single factor model?
What does it mean "reject a null hypothesis"?
What is the definition of standard error?
How to use the Excel solver() to find an x value for 95% confidence interval?

5.2 IMPACT OF SIZE, SAMPLE VS. POPULATION, SAMPLE STATISTICS

For empirical research, researchers start with a data set or sample. If the sample size is too small, the conclusion reached might not be that reliable since the potential bias in the data. A natural question is: what is the required minimum number of observations? Many researchers would have different answers or opinions. Based on our research experience, the minimum sample data set should be at least 30. Fortunately for most financial problems this condition is easily satisfied.

In Chapter 6: Risk vs. Return, we have learnt that standard deviation would be used to represent the total risk. Assume that n returns are available. The following formula is used to estimate their sample variance.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(R_i - \bar{R})^2}{n-1} \tag{2}$$

where σ^2 is the variance, R_i is the i^{th} return, $\bar{R}$ is the mean and n is the number of returns. Note that the denominator is n-1, instead of n, since this is for a sample. Many textbooks use s^2 for a sample variance while use σ^2 for the variance for a population. In this book, we don't make such a distinction. The standard deviation is the square root of variance.

$$\sigma = \sqrt{\sigma^2} \tag{3}$$

For a population, we have the following formula to estimate variance.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(R_i - \bar{R})^2}{n} \tag{4}$$

Note that for a population, the denominator is n instead of n-1. Obviously, when n is large enough, there is virtually no difference between those two definitions. The standard error is defined below.

$$stdError = \frac{\sigma}{\sqrt{n}} \tag{5}$$

For example, what is the standard error for S&P500 monthly returns? First, we download monthly price data from Yahoo!Finnance by using ^GSPC which is the ticker for S&P500, see the following image.

N26

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	1/1/1950	16.66	17.09	16.65	17.05	17.05	42570000
3	2/1/1950	17.05	17.35	16.99	17.22	17.22	33430000
4	3/1/1950	17.24	17.61	17.07	17.29	17.29	40410000
5	4/1/1950	17.34	18.07	17.34	18.07	18.07	48250000
6	5/1/1950	18.22	18.78	18.11	18.78	18.78	45080000
7	6/1/1950	18.77	19.4	17.44	17.69	17.69	45660000

Then we need to estimate annul returns, i.e., converting monthly returns to annual ones. For this reason, we estimate log returns.

H3 =LOG(F3/F2)

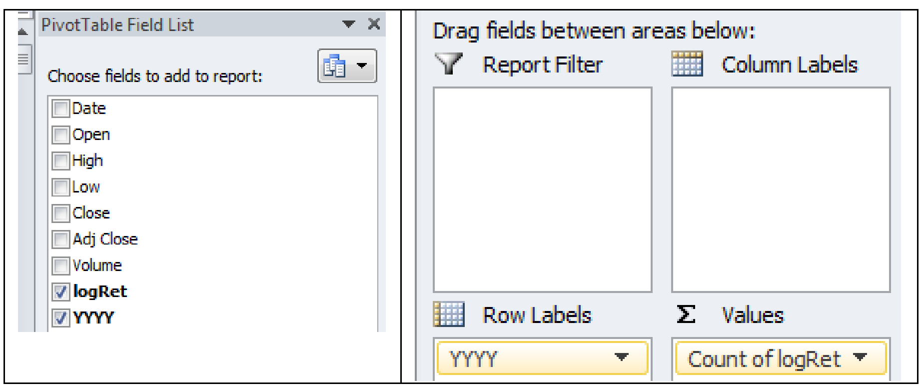
	A	B	C	D	E	F	G	H
1	Date	Open	High	Low	Close	Adj Close	Volume	logRet
2	1/1/1950	16.66	17.09	16.65	17.05	17.05	42570000	
3	2/1/1950	17.05	17.35	16.99	17.22	17.22	33430000	0.004309
4	3/1/1950	17.24	17.61	17.07	17.29	17.29	40410000	0.001762
5	4/1/1950	17.34	18.07	17.34	18.07	18.07	48250000	0.019163
6	5/1/1950	18.22	18.78	18.11	18.78	18.78	45080000	0.016737

Then by using the Excel year() function, we generate a YYYY column, see column I.

I3 =YEAR(A3)

	A	B	C	D	E	F	G	H	I
1	Date	Open	High	Low	Close	Adj Close	Volume	logRet	YYYY
2	1/1/1950	16.66	17.09	16.65	17.05	17.05	42570000		
3	2/1/1950	17.05	17.35	16.99	17.22	17.22	33430000	0.004309	1950
4	3/1/1950	17.24	17.61	17.07	17.29	17.29	40410000	0.001762	1950
5	4/1/1950	17.34	18.07	17.34	18.07	18.07	48250000	0.019163	1950
6	5/1/1950	18.22	18.78	18.11	18.78	18.78	45080000	0.016737	1950

The, highlight our data set and clicking “Insert” on the Menu bar, then “`PivoTable`” on the left-hand side, we have the following image.



We would see the following output.

	A	B
1		
2		
3	Row Labels	Count of logRet
4	1950	11
5	1951	12
6	1952	12
7	1953	12
8	1954	12

	A	B
1		
2		
3	Row Labels	Sum of logRet
4	1950	0.078118647
5	1951	0.066186177
6	1952	0.048362373
7	1953	-0.029764808

Since we wanted to get the log annual return instead of numbers of monthly returns for each year, we click the downward arrow beside "`Count of logRet`". Click "Value Field Setting" and choose "Sum", we could see the annual log return, see the image above on the right. To convert log annual returns to percentage annual returns, we apply the following formula.

$$R = e^{R^{logret}} - 1 \quad (6)$$

The related formula is shown in Cell C2.

C2 fx =EXP(B2)-1

	A	B	C	D
1	year	log annual return	R	
2	1950	0.078118647	0.081251	
3	1951	0.066186177	0.068426	
4	1952	0.048362373	0.049551	

Then we could estimate mean, standard deviation and standard error based on those annual returns, see below.

I2 fx =F2-2*F4

	A	B	C	D	E	F	G	H	I
1	year	log annual return	R						
2	1950	0.078118647	0.081251		mean=	0.0347999		mean - 2stdError	0.017778
3	1951	0.066186177	0.068426		std	0.0701822		mean + 2stderror	0.051822
4	1952	0.048362373	0.049551		steError	0.0085108	=E3/SQRT(COUNT(C2:C69))		
5	1953	-0.029764808	-0.02933						

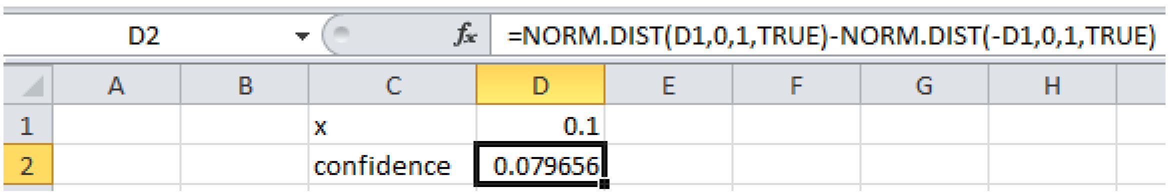
Based on the last 58-years' annual returns, the mean is 3.5%. If using historical mean as our prediction of the future, the expected value in 2018, is 3.5%. Since the mean plus and minus two standard errors would include about 90% all future possible outputs, we have a range of 1.78% and 6.18%, see the above image.

5.3 SIGNIFICANT LEVEL VS. CONFIDENCE INTERVAL

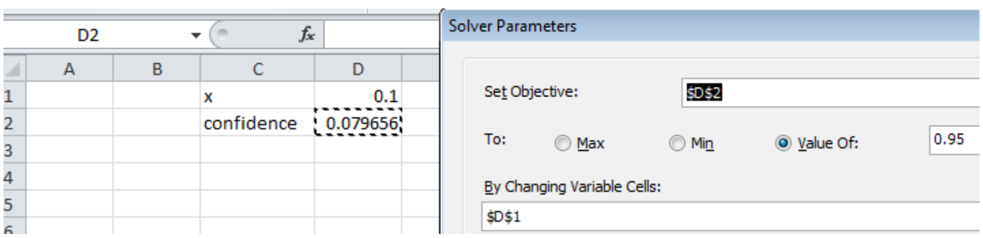
If we choose 95% confidence interval, the corresponding significant level would 5%. On the other hand, if the confidence interval is 99%, then the significant level is 1%. In other words, we have the following formula to convert one to the other.

$$confident\ interval = 1 - significant\ level \qquad (7)$$

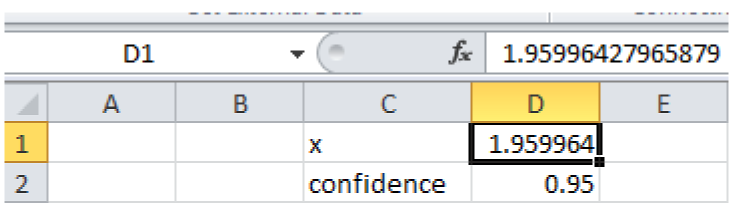
For a standard normal distribution, what are the range around zero represent 95% confidence interval? First, we enter 0.1 for an x value in D1. The probability between -0.1 and 0.1 would the difference of two cumulative density distribution, see the formula in the following image.



Using the Excel Solver() function, we could find the x value is 2.



The final x-value is 1.96, roughly 2, see the image below. This is why we use plus and minus two standard deviations representing a 95% confidence interval.



5.4 TWO VALUES: 2 FOR A T-VALUE, 5% FOR A P-VALUE

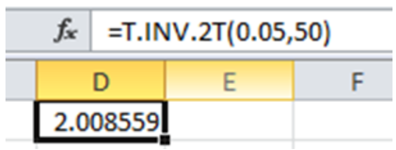
It is quite convenient to remember two critical values for a T-test. The first one is related to the T-value while the second one is related to the P-value. Below are decision rules when we conduct a T-test.

$$\begin{cases} \text{if the absolute T value} > 2, & \text{reject} \\ \text{if the p vlaue is less than 5\%,} & \text{reject} \end{cases} \quad (8)$$

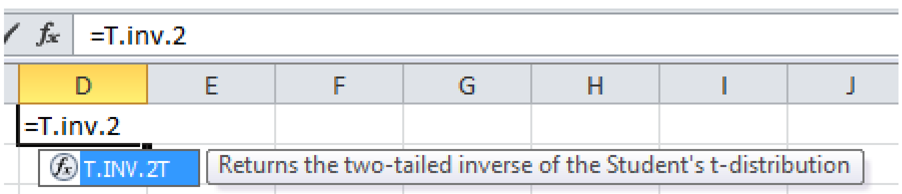
Assume that our null hypothesis that a stock's mean daily return is zero.

Null hypothesis: Returns of our stock follow a normal distribution

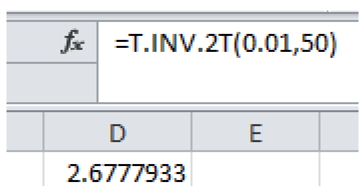
If the estimated T-value is above 2, in terms of the absolute value, we would reject the null hypothesis. Otherwise, we accept it. Alternatively, if the estimated p-value is less than 5%, we reject the null hypothesis. Actually, those two values, p=0.05 and T=2 are associated, see below.



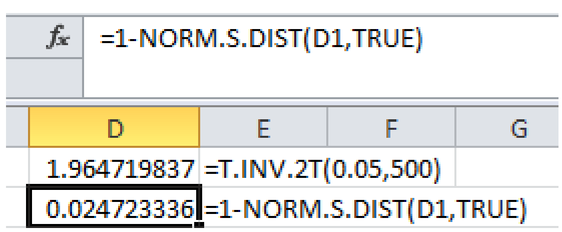
where the Excel `T.INV2T()` is the inverse function of a student's t-distribution (2-tailed), see the following image.



The higher our T-value, the stronger will be our rejection of the null hypothesis. For a 1% significant level, the related T-value will be 2.68, see below.



For the difference between one-tailed (one sided) and two-tailed (two sided) tests, see the next section. On the other hand, when the number of observations increases, student t-distribution will approach a normal distribution.



5.5 ONE-SIDED AND TWO SIDED TESTS

To test whether the returns equal 0, we could have one-side test or two-sided tests. For the one sided test, we might have the following hypothesis.

Null Hypothesis: the mean return is bigger than 0

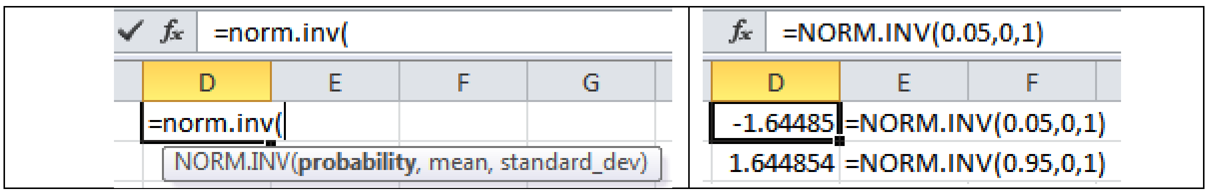
Similarly, we could have the following null hypothesis:

Null Hypothesis: the mean return is smaller than 0

A two-sided test could be viewed as the combination of the above two.

Null hypothesis: mean return is zero,

This is equivalent to the mean returns is not statistically bigger than zero and at the same time it is not statistically smaller than zero. Thus, the T-value for a two-sided test is higher than the one-sided test which is 1.64, see below.



What is the stock Market expected return next year? Assume that based on the log annual returns of the last 56 years, we have a value of 6.27%. Our answer would be that the expected log return next year (in 2018) is 6.27%. However, this is not a complete story. If the standard error is 0.02151, then a better answer would be: our expected next annual log return is 6.27% with a minimum value of 1.99% and a maximum value of 10.57%. Actually, we apply the following formula

$$\begin{cases} \text{min value} = \bar{\text{R}} - 2 * \text{stdError} \\ \text{max value} = \bar{\text{R}} + 2 * \text{stdError} \end{cases} \quad (9)$$

Where $\bar{R}$ is our mean, while stdError is the standard error, see its definition in the previous section.

Is IMB's daily mean return zero? For example, we could download last 5-year daily price data from Yahoo! Finance (3/24/2012 to 3/24/2017). Estimate daily returns, their mean and standard error. Based on the above equations, construct a range (min, max). If zero is located within this range, then we could claim that this mean is not statistically different from zero.

5.6 T-TEST, CRITICAL VALUE, DECISION RULE

A t-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis. It can be used to determine if two sets of data are significantly different from each other. For an one-sample t-test and for the null hypothesis that the mean is equal to a specified value μ0, we use the following statistic:

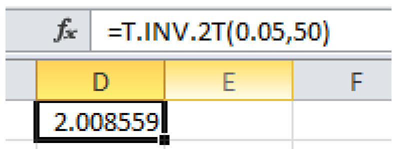
$$\begin{cases} t = \frac{\bar{R}-\mu_o}{S.E} \\ S.E = \frac{\sigma}{\sqrt{n}} \end{cases} \quad (10)$$

where t is the T-value, $\bar{R}$ is the sample mean, μ_0 is the our assumed return mean, σ is the sample standard deviation of the sample, n is the sample size and S.E. is the standard error. The degrees of freedom used in this test are n – 1.

Critical T value is used to accept or reject the null hypothesis. The decision rule is given below.

$$\begin{cases} T > T_{critical} & reject\ the\ null \\ T < T_{critical} & accept\ the\ null \end{cases} \quad (11)$$

To find the critical T-value, the Excel T.INV.2T() function is applied. If our confidence level is 95%, i.e., alpha is 5%, and the degree of freedom is 50, the critical T-value is 2.009, see the image below.



Thus, the first rule of thumb is that we remember a number of 2. With 5% significant level, any T value, absolute value, is bigger than 2, we reject. Otherwise, we accept.

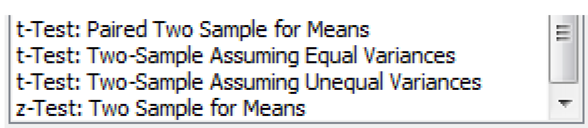
For a given T-value, we could find out its corresponding p-value. The related Excel function is `TDIST(x,deg_freedom,tails)`. The first one is the x input value, `deg_freedom` is the degree of freedom and the last input value indicates whether it is one-tail or two tail test. For the above example, if x=2.008559, deg_fredom=50, tails=2, we have the same alpha of 5% as well.

Application #1: Is IMB's daily mean return zero? You can use last 5-year daily data.

First, download IBM's daily price data from Yahoo!Finance. A better and much way is to down the data is using the following commands

```
> .show_ibmDaily(0)
 Launch Excel and paste
```

After click "Data", then "data analysis", we could see the following image. Note that if "Data Analysis" is not available on the menu bar, see Appendix A to activate it.



Application #2: Are the mean monthly returns for IBM and WMT same?

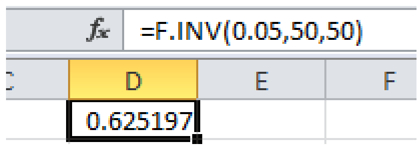
After we download IBM and WMT's monthly price data, we estimate monthly returns. Then we choose "Two-sample assuming Equal variances" or "Two sample assuming Unequal Variances", see the next chapter about how to test equal-variances. Note that we could use the same methodology to download stock data for those two stocks.

5.7 F-TEST FOR EQUAL VARIANCES

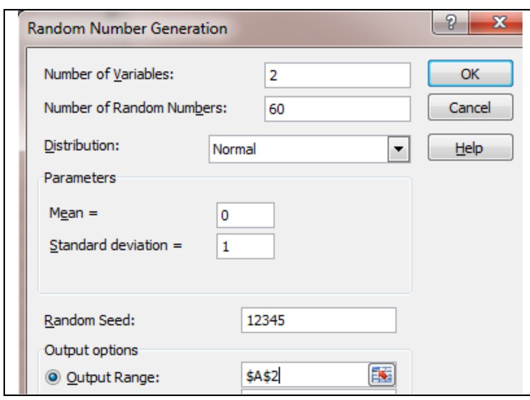
An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis.

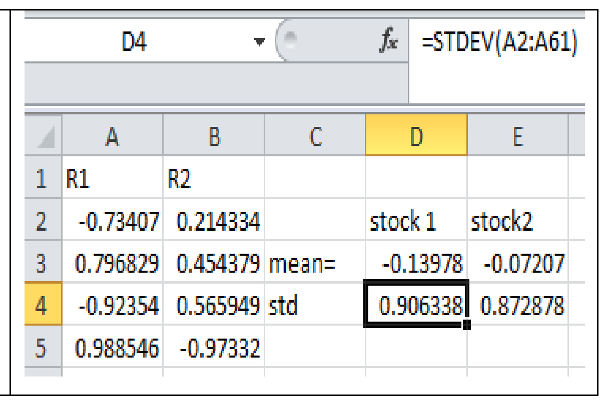
$$\begin{cases} F > F_{critical} & reject\ the\ null \\ F < F_{critical} & accept\ the\ null \end{cases} \tag{12}$$

The Excel function used to estimate a critical F value is `=F.INV(probability, deg_freedom1,deg_freedom2)`. The first input is the probability, or accuracy of our test, `deg_freedom` is the degree of freedom for the numeration.



Below, we generate some returns by using the Excel Random Number Generation. Then, we test whether they have equal variances. After clicking "Data" on the menu bar, "Data Analysis" and "Random Number Generation", we enter the following set of values, see the left image below.

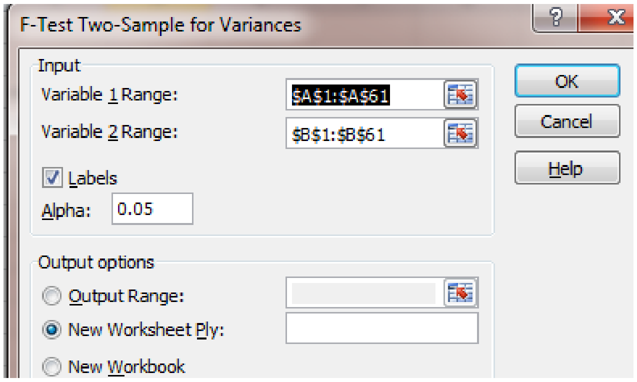




We could estimate mean and standard deviations of those two stock returns, see the right image above. Next we would conduct a test to see it they have the same variance.

Hypothesis: the two returns have the same variances

Clicking "Data" on the menu bar, then "Data Analysis", "F-Test Two Sample for Variances", see below.



The result is shown below.

	A	B	C
1	F-Test Two-Sample for Variances		
2			
3		*R1*	*R2*
4	Mean	-0.139778365	-0.0720714
5	Variance	0.821448948	0.76191567
6	Observations	60	60
7	df	59	59
8	F	1.078136311	
9	P(F<=f) one-tail	0.386793602	
10	F Critical one-tail	1.539956607	

Since the F value of 1.079 is smaller than the critical F-value of 1.54, we could not reject the null hypothesis. Because of this, we conclude that the two "returns" have the same variance.

Do IMB and WMT have the same variances? Assume that we use 10-year monthly data to test.

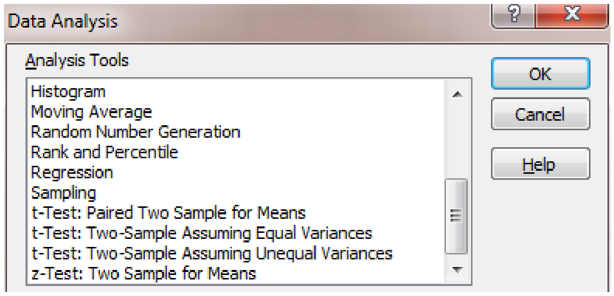
Step 1: download data
Step 2: estimate monthly returns
Step 3: put two returns into the same spread sheet
Step 4: click "Data" -> "Data Analysis" -> "F-test Two Sample for Variance".
Note: you can manually download data or use our efficient way to download monthly data.

5.8 T-TEST FOR EQUAL MEANS

After we download IBM and Apple's monthly price data from Yahoo!Finance, we could estimate their monthly returns and their monthly means. The question is whether those two company have the same means, i.e., the following hypothesis.

Hypothesis: IBM and Apple have the same monthly means

To answer this question, we could conduct a T-test. One more twist is that we have two T-tests, see below.



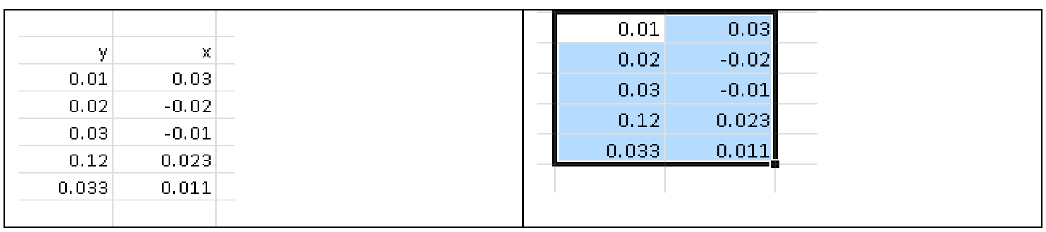
For is "t-Test: Two-Sample Assuming Equal Variances" and "t-Test: Two Sample Assuming Unequal Variances". Because of this, we have to conduct a variance test first, i.e., test if IBM and Apple have the same variances. If the answer is a yes, we apply "t-Test: Two-Sample Assuming Equal Variances". Obviously, is they have the difference variances, then we conduct the "t-Test: Two Sample Assuming Unequal Variance.

5.9 HOW TO RUN A SINGLE FACTOR LINEAR REGRESSION?

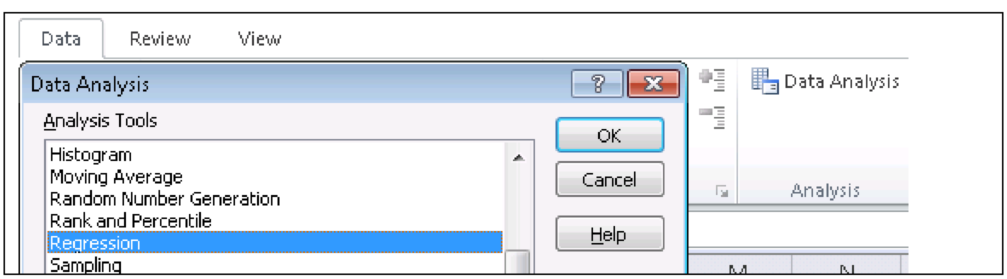
First, let's look at the general formula for a single-linear regression.

$$y = \alpha + \beta * x \qquad (13)$$

We could randomly enter two columns of y and x, see the left panel below.

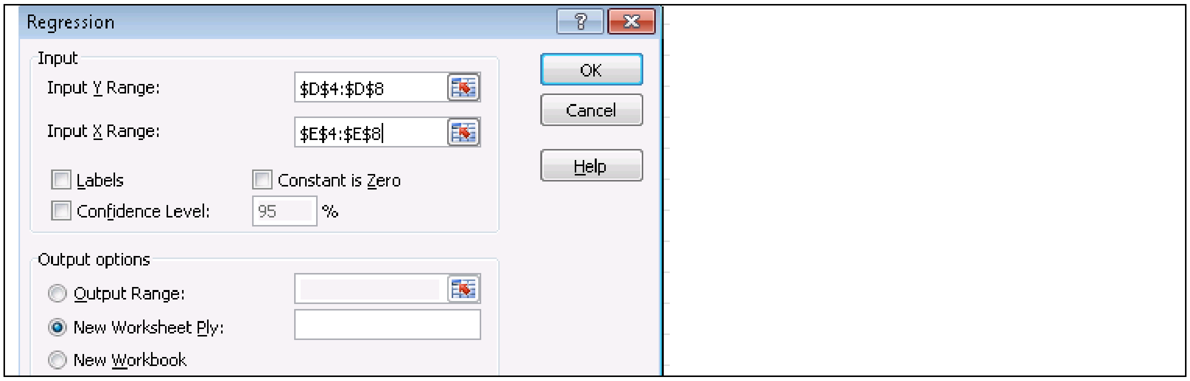


We have the following steps to run a linear aggression.
Step 1: Highlight data set, see below. (note that the labels of y and x are included)
Step 2: Click "Data", then "Data Analysis",[1] then choose "Regression"



[1] If "Data Analysis" is not available on the menu bar, see V) on page 9.

Step 3: choose y and x data ranges, see below.



Below is the final result. The beta will be 0.7

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple F	0.338074							
5	R Square	0.114294							
6	Adjusted I	-0.18094							
7	Standard E	0.048035							
8	Observati	5							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*gnificance F*			
12	Regressio	1	0.000893	0.000893	0.387128	0.577897			
13	Residual	3	0.006922	0.002307					
14	Total	4	0.007815						
15									
16		*Coefficients*	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
17	Intercept	0.037835	0.022806	1.658959	0.195706	-0.03475	0.110414	-0.03475	0.110414
18	X Variable	0.700792	1.12632	0.622196	0.577897	-2.88366	4.285245	-2.88366	4.285245

If you choose label in Step 1, then you have to click "Labels" in Step 3. You could choose to include the output within the same spreadsheet (output options).

REFERENCES

Wikipedia, Student's t-test, https://en.wikipedia.org/wiki/Student's_t-test
Wikipedia, Student's F-test, https://en.wikipedia.org/wiki/F-test

Appendix A: 20 questions after typing ".c5"

```
> .c5
function(i){
" i  Chapter 5: T-test, F-test, tests for equal mean (variances)
  -  ------------------------------------
  1  If 'data analysis' is not available on your menu bar
  2  sample vs. population
  3  T-value, p-value
  4  Descriptive statistics
  5  Rank and percentile
  6  estimate correlation
  7  estimate covariance
  8  How to run a linear regression?
  9  F-test: two samples for equal variances
 10  T-test: paired two samples for means
 11  T-test: two samples assuming   equal variances
 12  T-test: two samples assuming unequal variances
 13  z-test: two samples for means
 14  Is the mean of a set of values equal zero?
 15  Generate random numbers between 0 and 1
 16  F-distribution table
 17  function criticalFvalue()
 18  Fourier Analysis
 19  Regression with multiple independent variables
 20  Links

 Example #1:>.c5      # find out the list
 Example #2:>.c5(1)   # see the first explanation
```

Appendix B: Data Case #6: Which political party manages the economy better?

According to the web page of http://www.enchantedlearning.com/history/us/pres/list.shtml, we could find to which party the US presidents belong.

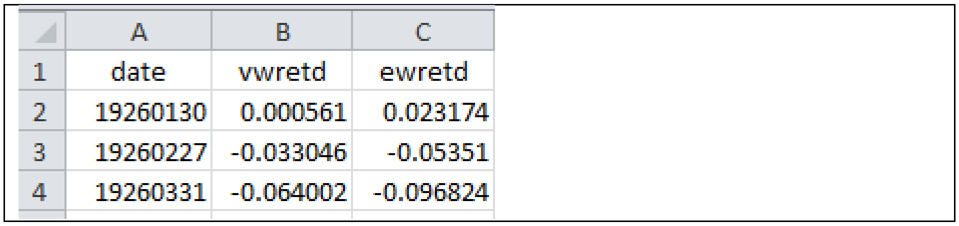
President	which party	time period
30. Calvin Coolidge (1872-1933)	Republican	1923-1929
31. Herbert C. Hoover (1874-1964)	Republican	1929-1933
32. Franklin Delano Roosevelt (1882-1945)	Democrat	1933-1945
33. Harry S Truman (1884-1972)	Democrat	1945-1953
34. Dwight David Eisenhower (1890-1969)	Republican	1953-1961
35. John Fitzgerald Kennedy (1917-1963)	Democrat	1961-1963
36. Lyndon Baines Johnson (1908-1973)	Democrat	1963-1969
37. Richard Milhous Nixon (1913-1994)	Republican	1969-1974
38. Gerald R. Ford (1913- 2006)	Republican	1974-1977
39. James (Jimmy) Earl Carter, Jr. (1924-)	Democrat	1977-1981
40. Ronald Wilson Reagan (1911- 2004)	Republican	1981-1989
41. George H. W. Bush (1924-)	Republican	1989-1993
42. William (Bill) Jefferson Clinton (1946-)	Democrat	1993-2001
43. George W. Bush (1946-)	Republican	2001-2009
44. Barack Obama (1961-)	Democrat	2009-

Thus, we could generate the following table. The PARTY and RANGE variables are from the web page. YEAR2 is the second number of RANGE minus 1, except the last row.

Table 1: Parties and Presidents since 1923

PARTY	RANGE	YEAR1	YEAR2
Republican	1923-1929	1923	1928
Republican	1929-1933	1929	1932
Democrat	1933-1945	1933	1944
Democrat	1945-1953	1945	1952
Republican	1953-1961	1953	1960
Democrat	1961-1963	1961	1962
Democrat	1963-1969	1963	1968
Republican	1969-1974	1969	1973
Republican	1974-1977	1974	1976
Democrat	1977-1981	1977	1980
Republican	1981-1989	1981	1988
Republican	1989-1993	1989	1992
Democrat	1993-2001	1993	2000
Republican	2001-2009	2001	2008
Democrat	2009-2012	2009	2014

Step 1: Open an Excel file called which_party_data.xlsx which has 3 columns: date, VWRETD (value-weighted market index) and EWRETD (equal-weighted market index), see a few observations below.



	A	B	C
1	date	vwretd	ewretd
2	19260130	0.000561	0.023174
3	19260227	-0.033046	-0.05351
4	19260331	-0.064002	-0.096824

Step 2: Classify VWRETD (returns) into two groups according to YEAR1 and YEAR2: under Republican and under Democratic

Step 3: Test the null hypothesis: two group means are equal.

$$\bar{R}_{Democratic} = \bar{R}_{Republican} \quad (1)$$

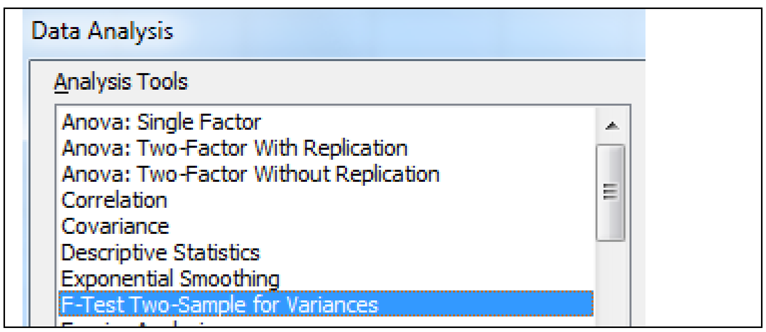
Step 4: Discuss your results and answer the following question: are the monthly mean returns under both parties equal?

Note 1: repeat the above process by using EWRETD (equal-weighted market index).

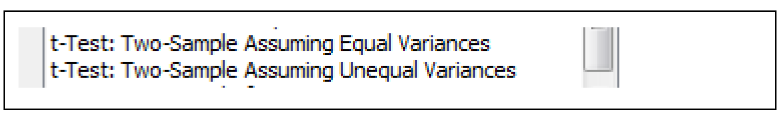
Note 2: How to test whether two groups have the same means?

Click "Data" on the menu bar then == > Data Analysis

Step A: test if the variances of two monthly returns are the same.

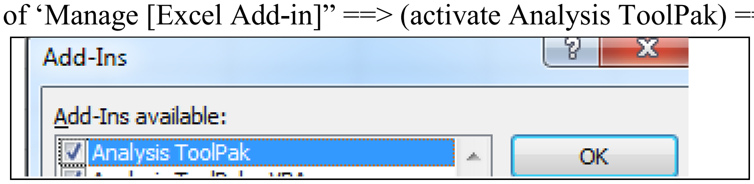


Step B: Depend on the result of Step A, we have the following choices.



Note 3: if "Data Analysis" is not available after clicking "Data" on the menu bar

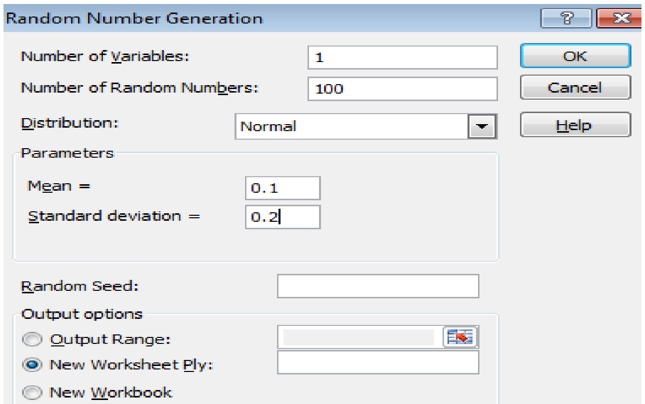
Click "File" ==> "Options" ==> "Add ins" (on the left-hand side) ==> "Go" (on the right-hand side of 'Manage [Excel Add-in]" ==> (activate Analysis ToolPak) ==>"OK"



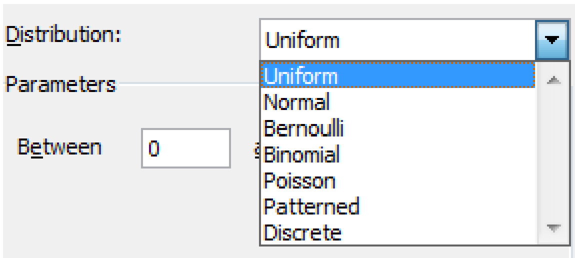
Appendix C: How to generate n random numbers?

Assume that we want to generate 100 random numbers draw from a normal distribution with mean=0.1 and standard deviation = 0.2

Step 1: Click "Data"
Step 2: Click "Data Analysis", note if "Data Analysis" is not available, see Appendix B.
Step 3: click "Random Number Generation", "OK"
Step 4: choose appropriate values, see the image below.



Note #1: if we want to place those random umber numbers in the same spread sheet, then choose a range on the right of "Output Range".
Note #2: if everyone wants to have the same set of random numbers, then enter the same seed, such as 123, on the right of "Random Seed:".
Note #3: if we want to generate 2 sets of n random numbers, just enter 2 on the right of "Number of Variables".
Note #4: There are several different distribution available, see the image below.



Exercises

5.1 What are two values of 2 and 5% for T-value and p-value? Why they are so important?
5.2 What is the critical value of a T-test?
5.3 What is the critical value for a p-value?
5.4 What is the T-test?
5.5 How to test equal variances?

5.6 How to use Excel to test whether a mean stock return is zero?
5.6 How to use Excel to test whether two stocks have the same mean returns?
5.8 Are the mean stock returns for IBM and City Group are the same? You can use the last 5 years daily return as a sample.
5.9 Do IMB's monthly returns follow a normal distribution? For example, we would use latest 10-year data to conduct the test.

Step 1: manually download IBM monthly price data or use the following efficient way to download the data

```
> x=getDailyPrice("ibm")
> y=head(x,121)
> saveYan(y,"c:/temp/ibmDaily.csv")
   [1] "Your saved file is ==>c:/temp/ibmDaily.csv"
>
```

Step 2: estimate IBM's monthly returns
Step 3: conduct the test

5.10 Confirm or reject the result in the chapter about the expected annual return next year. To download the daily price data for S&P500, you can use the following efficient way to retrieve data from Yahoo!Finance.

```
> x=getDailyPrice("^GSPC")
> saveYan(x,"c:/temp/sp500d.csv")
[1] "Your saved file is ==>c:/temp/sp500d.csv"
>
```

5.11 What is the x-value around zero representing 99% confidence interval for a standard normal distribution?