Chapter 3 Beta Coefficient, R Squared and Regression Analysis

"Regression is a powerful tool for forecasting. Economists using it successfully predicted ten out of the last two recessions" Anonymous

The last two chapters covered the risk and return characteristics of the portfolio over time and over various economic scenarios, as well as compared to other assets classes. The next level of analysis is to assess the portfolio as compared to other bench markets such as the market. This chapter will specifically focus on the comparative analysis of a portfolio versus the market. Stocks go up and down based on information such as earnings or technical data such as the 200-day average movement of the stock. This chapter will assess the risk or volatility of a portfolio of stocks in order to understand such movements versus market movements and the relationship versus market movements.

Learning Objectives

After reading this chapter, students will be able to:

- Calculate the beta and understand how these betas are used for quantifying return expectations
- Use regression analysis to compare a given portfolio of stocks as compared to the market
- Understand how to use historical data to assess if the movement of asset values are caused by other factors.
- Understand how to differentiate between dependent and independent variables in a linear regression in order to set the return expectation
- Explain the assumptions underlying a linear regression, the regression coefficients, and discuss the limitations of regression analysis.
- Understand the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret an F-statistic.

Regression Analysis – An Analytical Introduction

[Insert boxed text here

AUTHOR'S NOTES:

We love to compare everything. It's our nature. It puts things in perspective. I once asked an analyst that worked for me at the bank if I awarded him a \$1 million bonus would he be happy. Of course, only one year out of college who would not be. He said: "are you serious?". I said, "let's assume

I gave you a \$1 million bonus, but then how would you feel if I told you I gave everyone else at your level \$2 million?". This is the point: He would usually be ecstatic to get such large and unexpected sum of money as his first-year bonus. In his mind, he compared it to his very low expectation of what a first-year analyst, just out of school, should receive. Once however, I introduced the fact that his \$1 million would have been the lowest bonus paid compared to what his colleagues were receiving, his excitement and his expectation automatically changed dramatically. I was trying to show that analysts, statisticians, physicists, chemists, mathematicians and even Albert Einstein talked about relativity and the comparison of one event to another. Therefore, statisticians began providing methods to predict if the movement of one element is dependent on the movement of another element and to what extent. I always wondered if I ran a regression analysis of my employees' productivity and bonus received and ran another after their expectation changed what results would be. Anyway, this mathematical formula was developed in the early 1800s from one of my favorite mathematicians, Carl Friedrich Gauss. Gauss was brilliant. He introduced method called Least of Squares which is used extensively by investment bankers today to calculate the market beta coefficient.

End boxed text here]

[Insert boxed text here

KEY TAKEAWAYS:

- A portfolio analyst often needs to examine the relationship between two or more variables.
- A regression analysis is used for estimating when one variable causes the other or when one is dependent on the other. Such an estimate can be useful for predicting investment value of one variable based on the value of the other variable such as the market
- At perfect negative correlation basically negative 1 between two asset classes or two portfolios there will almost definitely be opportunities for efficiencies if the investor combines these assets or portfolios. In this case, the investor could benefit from the average return between the two asset classes or portfolios and at the same time minimize the standalone standard deviations to a lower combined one therefore reducing risk
- The relationship between average return and combined standard deviation is the basis for introducing the concept of the Sharpe Ratio that will be discussed in later chapters.

End boxed text here]

Introduction to Linear Regression & Beta Coefficient

The objective for using a linear regression model in statistics is to analyze the straight-line relationship between two variables. The relationship can be explained as follows:

$$y = \beta_0 + \beta_1 x + e$$

Where y is the dependent variable or response variable, x is the independent variable or explanatory variable, and β_0 is the y-intercept. The y-intercept is the value of y when x=0, and β_1 is the slope of the line. β_1 gives the amount of change in y for every unit change in x. Finally, *e* is the random error. The random error is included under the premise that analysts are never

perfectly precise within their analysis of predicting a future outcome. Many financial models that use regression analysis to predict future values or returns do not include the error calculation. The regression analysis on excel uses the statistical concept of least squares. A method that seeks the coefficients β_0 and β_1 such as that:

$$y = \beta_0 + \beta_1 x$$

The regression analysis estimates the average value of the dependent variable (y), such as the value of the portfolio, when the independent variable (x), such as the market index, is fixed. Let's say that the average market index had a monthly average return of 1% and you are convinced that the market movements, up or down, affect the value of the portfolio, which is usually explained as a multiple. Let's assume 2%, which means that every time the market goes up or down by 1.0%, the value of the portfolio or the specific stock moves twice that amount in the same direction (a multiple 2.0x). We will run a regression analysis to calculate first if there is indeed a strong dependent relationship and attempt to estimate how much of is affected by what is called the Beta (β_1) . The beta coefficient (β_1) , called in statistics the slope, can be calculated by using linear regression. The calculation of is demonstrated in figure 3.1 below.

	Re	turns		ations age Return	Product from Deviation	Product from Return
	y	x	$y-\overline{y}$	$x-ar{x}$	$(y-\bar{y})\cdot(x-\bar{x})$	$(x-\bar{x})^2$
-	Stocks	S&P Index	Stocks	S&P Index		
	%	%	%	%	%	%
Year -12	-6.50	-4.60	-18.18	-9.82	178.42	96.37
Year -11	-13.20	-11.30	-24.88	-16.52	410.85	272.80
Year -10	-8.90	-5.00	-20.58	-10.22	210.21	104.38
Year -9	25.00	12.00	13.33	6.78	90.39	46.01
Year -8	48.50	23.00	36.83	17.78	654.87	316.25
Year -7	37.60	25.00	25.93	19.78	512.88	391.38
Year -6	10.50	1.00	-1.18	-4.22	4.95	17.78
Year -5	7.20	4.50	-4.48	-0.72	3.21	0.51
Year -4	-5.60	1.00	-17.28	-4.22	72.84	17.78
Year -3	17.50	3.00	5.83	-2.22	-12.91	4.91
Year -2	21.50	12.00	9.83	6.78	66.65	46.01
Year -1	6.50	2.00	-5.18	-3.22	16.65	10.35
					2209.01	1324.54
Average Return	11.68	5.22				
Standard Deviation	19.20	10.97			Beta (β)	1.6678
				$\frac{\sum[(y - \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{j=1}^{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{j=1}$	$\overline{y}) \cdot (x - \overline{x})]$ $x - \overline{x})^2$	-

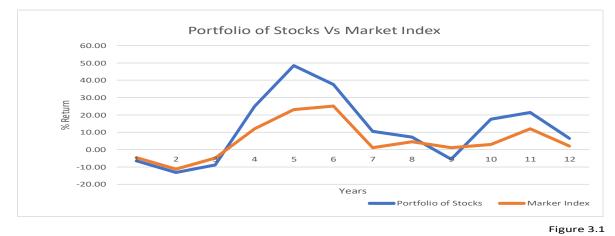
HISTORICAL ANALYSIS

EXCEL FORMULAS Slope (6)= Forecast = Standard Error =

1.6678 Rela 4.6426 Prec

Relationship between Dependent Y with Indepent X Predicts value of y given a value of x=1%

6.0923 Predicts the standard error y-value for each x in the regression



[Insert boxed text here

Bloomberg Definition of Beta

Adjusted Beta vs. Raw Beta - The beta of a stock can be presented as either an Adjusted Beta or a Raw Beta. A Raw Beta is obtained from the linear regression of a stock's historical data. Raw Beta, also known as Historical Beta, is based on the observed relationship between the security's return and the returns of an index.

The Adjusted Beta is an estimate of a security's future Beta. Adjusted Beta is initially derived from historical data but modified by the assumption that a security's true Beta will move towards the market average, of 1, over time. The formula used to adjust Beta is: $(0.67) \times \text{Raw Beta} + (0.33) \times 1.0$.

Bloomberg Equity Codes:

• BETA – Historical Beta

• *COMP – Comparative Total Returns*

End boxed text here]

Regression Analysis, R square and the Analysis of Variance (ANOVA) – Using Excel

Figure 3.2 shows the regression analysis using Excel application for comparing the portfolio returns vs the market index returns as illustrated previously in Figure 3.1. The Excel Regression analysis yields 3-part outputs:

- 1. The Regression Statistics: R square and the Standard Error;
- 2. Analysis of Variance (ANOVA) between the two variables. The ANOVA shows the degrees of freedom (df), calculates the Sum of Squares (SS), the Mean Squares, the F-Statistic and Significance F. The ANOVA shows whether there are any statistically significant differences between the means of two or more variables.
- 3. **Regression Coefficients**: The slope or Beta Coefficient, the t-test and upper/lower 95% limits "within" the group variability and "between" the group variability. The t-test is another measurement for comparing the samples. The t-test is more effective if there is a comparison of two variables.

Insert Figure 3.2

REGRESSION SUMMARY OUTPUT (EXCEL)

Regression Statistics	
Multiple R	0.953139
R Square	0.908474
Adjusted R Square	0.899322
Standard Error	6.092286
Observations	12

	df	SS	MS	F	Significance F
Regression	1	3684.083055	3684.083055	99.25877166	1.64467E-06
Residual	10	371.1594445	37.11594445		
Total	11	4055.2425			

Slope (beta)	Coefficients	Standard Error	t-Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	2.974868822	1.963560699	1.515037871	0.160712968	-1.40021706	7.349954704	-1.40021706	7.349954704
X Variable 1 (Beta Slope)	1.667756775	0.167397229	9.96286965	1.64467E-06	1.294772505	2.040741046	1.294772505	2.040741046

Figure 3.2

Regression Statistics

Multiple R: The multiple R which is measured between 0 to 1 reflects the correlation coefficient of the two variables. It tells the analyst how strong the linear regression is. Figure 3.2 above shows a relatively strong correlation between the portfolio and the market index at 0.95. Value 1 represents a perfect correlation and a value of zero means no relationship at all.

R-Squared: The R-Squared (r²), also called Coefficient of Determination, measures how many

points fall on the regression line between the two variables. Figure 13.2 shows $r^2 = 0.908$. This means that approximately 91% of the variation of the portfolio values (y values) can be explain by the market index (x-values). The r^2 is relatively strong at 90% because it means that more than 90% of the data fits the model – one of the first numbers the analyst looks to determine if the data can be compared more accurately. A low r^2 might make the regression analysis irrelevant, so using a beta to determine the expected return, for example, could cause a problem for the analyst.

Adjusted R-Square: This number reflects a more accurate r^2 if the number of variables used are more than one.

Standard Error: The standard error in a regression analysis measures how precise the regression coefficient is. In other words, it tells the analyst how spread out the y variables are around the mean (average). Looking at the standard error in figure 3.2 above it reflects that the data is widely spread showing that the average deviation from the regression line is estimated at 6.1%. The smaller the standard error tha closer the values are to the regression line.

Observations: Number of observations in the analysis. Figure 3.2 shows 12 representing the 12 years of historical performance for both the portfolio and market index annual returns

<u>ANOVA</u>

Degrees of Freedom (df): When calculating historical data, the average standard deviation of each of the variables is based on the number of observations minus one (n-1).

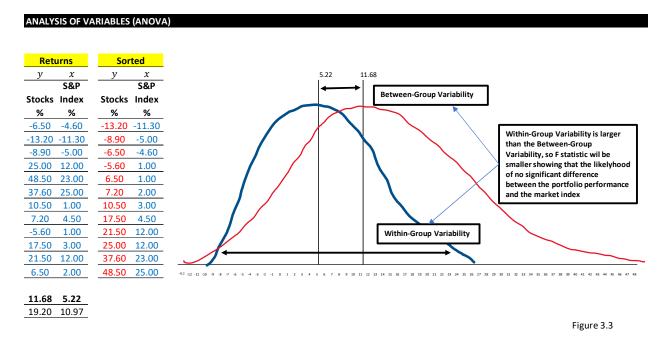
Between Group Variability and Sum of Squares: When looking for comparing two samples – in this case, the portfolio value and market index – there could be an overlap in the date. This is where the portfolio could have had the same returns as the index at fa ew specific times. If the analyst plots the data into a normal distribution and compares each variable to the grand mean between the two variables (the average of both variables), it could calculate the Between-Group Variability, which is the distance between the two means as illustrated in figure 3.3 below.

Within Group Variability and Mean Squares: If the analyst plots the data into a normal distribution and compares each variable to the grand mean within the two variables it could calculate the Within-Group Variability, which is the distance between the two means as illustrated in figure 3.3 below.

F-Statistic or Significance F: The F-Statistic or Significance F and sometimes called the F-Ratio, is the number that measures if the means of different samples are significant different or not. The lower the number, more similar the sample means are. In Figure 3.2 above and illustrated in Figure 3.3 below shows that the Between-Group Variability is smaller than Within-Group Variability by a factor of 10x (37.1 vs 371.2) making the F-Ratio smaller. The 99.25 shown previously in figure 3.2 does not mean much except if you had to compare this portfolio to another portfolio and then compare each to stock market index (independent variable). This will give the analyst a reference on how close the average return performance of each portfolio is to the average market return performance. Unlike the t-distribution (explained later), the F-distribution does not have any

negative values because between and within-group variability are always positive because both are squared.

Insert Figure 3.3



Regression Coefficients:

Beta Coefficient (Slope): As mentioned above, the linear regression analysis is basically the linear relationship between the independent (x) and depended (y) expressed in the equation $y = \alpha + \beta x$ called the Regression Line. The first calculation coefficient is beta coefficient (β) or also known as the slope. In both figures 13.1 and 13.2, the beta (β) shown is calculated at 1.66x. Assuming x represents the market index return and assuming the constant alpha coefficient (α) is zero, then the beta (β) represents the multiplier effect that calculates the portfolio or company specific expected return (y). The beta (β) results from the regression analysis and can be used to predict the portfolio's expected return. The following chapter will closely examine this model for predicting return expectation called the Capital Asset Pricing Model (CAPM). Though CAPM measures a slightly different result, it uses the same regression line concept to predict the expected return of a specific stock or a portfolio of stocks.

Standard Error: The standard error is similar to standard deviation measuring the spread or the deviation from the data's average. Though both are calculated the same, the standard error uses a smaller statistical sample data and the standard deviation uses the entire population data (aka parameters) to calculate the difference from the average.

t-test: The t-test or t-statistic is used when the analyst decides to accept or reject the null hypothesis – basically that the movements or performance of the dependent variable (y) does depend on the movement of the independent variable (x). In our example, the portfolio value is looking to see if

there is systemic risk. The t-test is used for a small sample and is more effective if it compares two variables. The higher the t, the more evidence the analyst has to show that there is significant difference between the combined average of the two variables.

p-value: The p-value, or probability value, is expressed from 0-1 and it measures the percentage difference in the combined average of the variables. Figure 3.2 shows 0.16 or 16.0% variation - relatively small percentage.

Lower and Upper 95% confidence intervals: The confidence intervals are expressed as a percentage – in this case 95%. It means that the results are expected to match 95% of the time. In a normal distribution, the performance of the portfolio that's based on historical returns, is predicted to stay within 95% of the range. The upper represents the positive or right side of the normal distribution graph and the lower represents the negative or the left side of the normal distribution graph – See figure 3.4 below:

Insert Figure 3.4

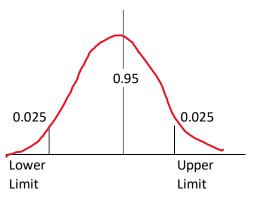


Figure 3.4

Portfolio of Stocks and Bonds – Zeus Fund I - Case Study

The previous two chapters focused primarily on the risk and return of the Zeus Fund I portfolio which includes stocks and bonds and how each of these asset classes have performed against each other. It also attempted to find the right mix in order to achieve efficiency and optimization based on the correlational phenomenon. This chapter will take the Zeus Fund I performance and compare each asset class to the equivalent index such as the S&P 500 Index for stocks and the Bloomberg Barclays U.S. Aggregate Bond Index (Bloomberg Bond Index) for bonds. This comparative analysis should establish the beta coefficient (β) for each asset class. This beta (β) will then be used for basic portfolio analysis in later chapters. Figure 3.5 below shows the 7-month performance of each of the stocks and bonds, as well as the historical return of the stock index benchmark (S&P 500) and bond Index benchmark (Bloomberg Bond Index).

ZEUS Fund I

Performance		0	1	2	3	4	5	6	7
Symbol			July 1 20x1	Aug 1 20x1	Sep 1 20x1	Oct 1 20x1	Nov 1 20x1	Dec 1 20x1	Jan 2 20x1
Stocks			-1.0%	6.2%	3.5%	3.5%	3.0%	7.2%	1.2%
Bonds			1.9%	0.5%	0.8%	0.5%	0.6%	-0.1%	-0.4%
Total Stock Value		81,757	82,600	82,200	84,050	85,000	85,450	87,300	92,275
Total Bonds Value		97,495	95,650	93,620	97,035	101,575	102,110	101,410	97,325
Total Portfolio Value		179,253	178,250	175,820	181,085	186,575	187,560	188,710	189,600
Portfolio % change			-0.6%	-1.4%	3.0%	3.0%	0.5%	0.6%	0.5%
Cummulative % Change			-0.6%	-1.9%	1.1%	4.1%	4.6%	5.2%	5.7%
Bench Mark: S&P 500		2,430.1	2,429.0	2,476.4	2,476.6	2,529.1	2,579.4	2,644.2	2,644.2
% Increase / Decrease			0.0%	1.9%	0.0%	2.1%	2.0%	2.5%	0.0%
	ays US Aggreagate Bond Index	2,021.6	2,041.8	2,034.8	2,044.5	2,039.5	2,039.5	2,042.8	2,042.8
% Increase / Decrease			1.0%	-0.3%	0.5%	-0.2%	0.0%	0.2%	0.0%
Weighted Combined Index		2,207.9	2,221.2	2,241.2	2,245.0	2,262.6	2,285.5	2,321.0	2,335.5
% Increase / Decrease			0.6%	0.9%	0.2%	0.8%	1.0%	1.6%	0.6%
Stocks (Weights)		45.6%	46.3%	46.8%	46.4%	45.6%	45.6%	46.3%	48.7%
Bonds (Weights)		54.4%	53.7%	53.2%	53.6%	54.4%	54.4%	53.7%	51.3%
7-month Portfolio Performar	nce								
STOCKS		BONDS				<u>(</u>	COMBINED P	ORTFOLIO	
Average	3.3777%	Average		0.5428%		7	Average		1.8613%
Standard Deviation	2.7976%	Standard Dev	iation	0.7477%		١	/ariance		0.0185%
Average % of Portfolio	46.5076%	Average % of	Portfolio	53.4924%		c	tandard Dev	iation	1.3612%

Figure 3.6 below shows the betas between the stock portfolio versus the stock market benchmark, the bond portfolio versus the bond market benchmark and the combined Zeus Fund I portfolio of stocks and bonds versus the weighted benchmark of the stock and bond index.

ZEUS Fund I

Beta (β) Coefficient Analysis
STOCK AND BOND PORTFOLIO

		ST	OCK PORTFOLI	0		
	y Portfolio	x S&P Index	[y - Avg(y)] Portfolio	[x-Avg(x)] S&P Index	[y-Avg(y)].[x- Avg(x)] Product Deviation	[x-Avg(x)]^2 Product Deviation
Month 1	-1.0%	0.0%	-4.4%	-1.3%	0.05554%	0.01595%
Month 2	6.2%	1.9%	2.9%	0.7%	0.02086%	0.00532%
Month 3	3.5%	0.0%	0.1%	-1.2%	-0.00133%	0.01468%
Month 4	3.5%	2.1%	0.1%	0.9%	0.00076%	0.00816%
Month 5	3.0%	2.0%	-0.3%	0.8%	-0.00265%	0.00588%
Month 6	7.2%	2.5%	3.8%	1.3%	0.04966%	0.01677%
Month 7	1.2%	0.0%	-2.1%	-1.2%	0.02615%	0.01488%
Mean Avg=	3.4%	1.2%			0.14898%	0.08162%
Stand. Dev.=	2.8%	1.2%			Beta=	1.8251175

		BO	ND PORTFOLI	0		
	У	х	[y - Avg(y)]	[x-Avg(x)]	[y-Avg(y)].[x- Avg(x)]	[x-Avg(x)]^2
		Bond		Bond	Product	Product
	Bonds	Index	Bonds	Index	Deviation	Deviation
Month 1	1.9%	1.0%	1.4%	0.8%	0.01178%	0.00721%
Month 2	0.5%	-0.3%	-0.1%	-0.5%	0.00025%	0.00244%
Month 3	0.8%	0.5%	0.2%	0.3%	0.00079%	0.00109%
Month 4	0.5%	-0.2%	-0.1%	-0.4%	0.00028%	0.00158%
Month 5	0.6%	0.0%	0.1%	-0.1%	-0.00015%	0.00022%
Month 6	-0.1%	0.2%	-0.6%	0.0%	-0.00008%	0.00000%
Month 7	-0.4%	0.0%	-1.0%	-0.2%	0.00148%	0.00023%
Mean Avg=	0.5%	0.2%			0.01435%	0.01277%
Stand. Dev.=	0.7%	0.5%			Beta=	1.1238116

		СОМВ	INED PORTE	OLIO		
	У	х			у.х	x^2
Portfolio						
Weight Stocks	Combined Portfolio	Weighted Benchmark	Combined Portfolio	Weighted Benchmark	Product Deviation	Product Deviation
45.6%	0.6%	0.6%	-1.3%	-0.2%	0.00259%	0.00041%
46.3%	3.2%	0.9%	1.3%	0.1%	0.00120%	0.00009%
46.8%	2.0%	0.2%	0.2%	-0.6%	-0.00117%	0.00404%
46.4%	1.9%	0.8%	0.0%	0.0%	0.00000%	0.00001%
45.6%	1.7%	1.0%	-0.1%	0.2%	-0.00027%	0.00042%
46.3%	3.3%	1.6%	1.4%	0.8%	0.01072%	0.00563%
48.7%	0.4%	0.6%	-1.5%	-0.2%	0.00273%	0.00034%
Mean Avg=	1.9%	0.8%			0.01581%	0.01093%
Stand. Dev.=	1.1%	0.4%			Beta=	1.4463175

Figure 3.6

Figure 3.7 below shows the Regression Analysis, ANOVA and Regression Coefficient. The figure below demonstrates that the betas on a standalone basis are 1.12 and 1.82 for the stock and bond portfolios as compared with their corresponding bench marks. By combining the portfolio and adjusting for the weights, the beta is at 1.44 – basically the weighted average of betas between stocks and bonds. The R-squares are relatively low – below 60% suggesting that the analyst should be warned that the data could be not as effective.

ZEUS Fund I

Regression Analysis

STOCK PORTFOLIO						
Regression Statistics	ANOVA					
Multiple R 0.69344121		df	SS	MS	F	Significance F
R Square 0.48086071	Regression	1	0.00016129	0.00016129	4.6313265	0.08402679
Adjusted R 0.37703285	Residual	5	0.00017413	3.4826E-05		
Standard E 0.00590135	Total	6	0.00033542			
Observatio 7						

		Standard						
	Coefficients	Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.00373566	0.00236511	1.57948636	0.17505993	-0.00234405	0.00981536	-0.00234405	0.00981536
X Variable	1.12381161	0.52220474	2.1520517	0.08402679	-0.21855842	2.46618164	-0.21855842	2.46618164

Regression St	atistics						
Multiple R 0.7	6091956	ANOVA					
R Square 0.5	7899858		df	SS	MS	F	Significance F
Adjusted R 0.4	9479829	Regression	1	0.00271897	0.00271897	6.87644448	0.0469674
Standard E 0.0	1988476	Residual	5	0.00197702	0.0003954		
Observatio	7	Total	6	0.00469599			

	Coefficients	Stanaara Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.0115171	0.01133778	1.01581574	0.35632352	-0.0176276	0.0406618	-0.0176276	0.0406618
X Variable	1.82511751	0.6959994	2.62229756	0.0469674	0.03599411	3.61424092	0.03599411	3.61424092

COMBINED	PORTFOLIO							
Regressio	on Statistics							
Multiple R	0.5479264	A	NOVA					
R Square	0.30022334			df	SS	MS	F	Significance F
Adjusted R	0.16026801	F	Regression	1	0.00022865	0.00022865	2.14513689	0.20290804
Standard E	0.01032421	F	Residual	5	0.00053295	0.00010659		
Observatio	7	Т	otal	6	0.00076159			
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%

 Intercept
 0.00696525
 0.00887245
 0.78504219
 0.46797673
 -0.01584212
 0.02977261
 -0.01584212
 0.02977261

 X Variable
 1.44631755
 0.98749783
 1.46462859
 0.20290804
 -1.09212643
 3.98476153
 -1.09212643
 3.98476153

Figure	3.7

CASE STUDY AND PRACTICE CASES

1. Based on the information below, complete the projected spreadsheet. (access spreadsheet <u>www.professordrou.com</u>)

TO BE PROVIDED LATER

References (Chapter 3)

TO BE PROVIDED LATER