STP 429 - Regression Analysis

Arizona State University

Takahiro Wada

Lab #3

Executive Summary

Whether you work for a hedge fund, investment bank, or even being a retail trader. You need to anticipate what the future price of a stock will be in order to be profitable in your trades or investments. Hence, it is important to understand what determines the stock's prices of the day. We will be focusing on the stock Centene, Ticker Symbol: CNC. Centene is categorized in the Managed Health Care GICS Sub-Industry. Understanding what determines Centene's daily closing price allows the trader to become more profitable by identifying which key stocks influence Centene's closing price of the day.

Many stocks influence the daily closing price of a stock, so a study was conducted to try and predict which stocks have the best correlation to Centene. Stocks that are considered critical when predicting Centene's closing price are Anthem's (ANTM), Cigna's (CI), Humana's (HUM), and UnitedHealth Group's (UNH) closing prices. A statistical analysis was performed to determine which stocks are most important when building a model to predict Centene's daily closing price.

Using multiple regression techniques to help analyze each of the stocks closing price and their relevance, I can conclude that Anthem, Cigna, Humana, Anthem times Cigna, Anthem times Humana, Anthem^2, and Cigna^2 can be used to predict Centene's daily closing price. The following report will include details of the analysis. The model that was developed will help traders determine which stocks are best suited for determining Centene's daily closing price.

Introduction

This study was generated in order to determine if there are stocks that predict Centene's daily closing price so we can build a prediction model for traders. We have 6 independent variables that may have a strong correlation to predicting Centene's daily closing price such as Anthem, Cigna, Humana, UnitedHealth Group, AbbVie, and Abiomed which are included for our prediction model in our study. Traders rely on identifying key stocks that predict Centene's daily closing price in order to maximize their profits.

Analysis

To determine the best model for predicting Centene's daily closing price, I first analyzed each variables' scatter plots. The scatter plots allow us to see the relationship between independent variables vs Centene's daily closing price and determine whether the independent variable has any association to our dependent variable. The scatter plot alone does not tell us which independent variable is significant to our prediction model, but does give us an idea. So, I used a correlation matrix which allows us to determine which independent variables have a strong association to our dependent variable. The chart allows us to see the strength for each variable, also to see if we can identify any outliers within our data set. Once we decide which independent variables are acceptable to be used, I used all the selection criteria methods to determine our best 1st order model. I then next tested for multicollinearity from our 1st order model and analyzed the data. Next, I tested our model to see whether an interaction term or quadratic term will help better predict Centene's daily closing price. Finally, I used a stepwise method, so the analysis will stop when all independent variables have been confirmed and are acceptable to be used for the model.

Data Section

From Table 8, the descriptive analysis for our 7 variables used for our model:

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
CNC	CNC	250	65.3701600	4.6657318	57.9700010	74.6999970
ANTM	ANTM	250	356.9464798	36.1691132	278.2500000	436.2399900
CI	CI	250	224.4849200	19.3444671	174.8399960	266.9100040
HUM	HUM	250	419.5595192	24.5753870	375.1499940	471.2200010
UNH	UNH	250	382.3649202	34.6687737	312.1000060	455.4400020
ABBV	ABBV	250	109.5558801	5.8323296	87.9599990	120.7799990
ABMD	ABMD	250	316.0664791	28.3462383	254.8500060	376.2000120

Centene's daily closing price ranges from \$57.97 to \$74.69. Anthem's daily closing price ranges from \$278.25 to \$436.24. Cigna's daily closing price ranges from \$174.84 to \$266.91. Humana's daily closing price ranges from \$375.15 to \$471.22. UnitedHealth Group's daily closing price ranges from

$312.10 to $455.44. AbbVie's daily closing price ranges from $87.96 to $120.78. Abiomed's daily closing price ranges from $254.85 to 376.20. I chose Anthem, Cigna, Humana, and UnitedHealth group for this study since they are all in the same GICS sub-industry category, Managed Health Care. AbbVie and Abiomed were included in the study to observe if different GICS sub-industry categories will better determine Centene's daily closing price. AbbVie is part of the pharmaceuticals sub-industry and Abiomed is part of the Health Care Equipments sub-industry. We are given a total of 6 independent variables. At the end, I only used Anthem, Cigna, and Humana for our 3 variables. So to identify which variables are viable for our model, I analyzed both the scatter plots and the correlation matrix of each variable to determine if there is a relationship between the independent variables and dependent variables. I first identified which variables displayed a clear relationship between the independent variables and the dependent variable from their respective scatter plots. I determined that Anthem, Humana, and UnitedHealth Group have a moderate association with Centene's daily closing price. To ensure I have got all the viable variables for our model, I ran a correlation matrix to look at each p-value for each variable. I only accepted variables that had a p-value less than 0.05 and linear correlation coefficient greater than 0.50 which were Anthem, Cigna, Humana, and UnitedHealth Group. I rejected AbbVie and Abiomed which is explained in the results section.

Results

Looking at our scatter plots for each variable from Table 1 - Table 6, we see that Anthem, Humana, and UnitedHealth Group have a moderate, positive, and a linear relationship, with a little bit of variation towards the middle, top-right of the graph, while the other scatter plots show a weak or no relationship with Centene's daily closing price. Scatter plots alone are not viable for choosing our variables for our prediction model and we do not want to miss any important variables that will improve our model.

So I next looked at our correlation matrix in Table 7, we see that Anthem, Cigna, Humana, and UnitedHealth Group are best correlated with Centene's daily closing price. I chose not to accept AbbVie and Abiomed because their linear correlation coefficient was below .50 which I determined is not significant enough. Also their respective scatter plots, Table 5 & 6, showed weak or no relationship with Centene's daily closing price. I also noticed that the most significantly correlated variable with Centene's daily closing price is Humana with a linear correlation coefficient of 0.78205. This is further supported by the scatter plot in Table 3, showing a moderate, positive and linear association as seen as the data points are moderately close together and follow a positive trend.

With the 4 independent variables I have determined to be significant to our model, I used all the selection criteria methods to select my primary variables for our model. With the stepwise selection technique, Table 9, the output given is a 3 variable model containing Humana, Anthem, and Cigna. With the R-Squared selection technique, Table 10, I chose the 3 variable model containing Anthem, Cigna, and Humana with an R-Square of 0.6406. I chose the 3 variable model, because the R-Squared only improved a little bit when it was a 4 variable model and the 3 variable model was more than sufficient. With the R-Square Adjusted selection technique, Table 11, I chose the 3 variable model containing Anthem, Cigna, and Humana with an R-Square Adjusted of 0.6362 and R-Square of 0.6406. The 3 variable model is a little bit better than the 4 variable model, but the difference is so miniscule that it does not really matter. So I chose the parsimonious 3 variable model instead. With the Cp selection technique, Table 12, I chose the 3 variable model containing Anthem, Cigna, and Humana as it has the lowest Cp value and the highest R-Square value than any of the other models. With the PRESS selection technique, Table 13, the output given is a 2 variable model containing Anthem and Humana. From all the selection techniques, I noticed that Anthem and Humana always appear in every model and Cigna appears the majority of the time. On the other hand, UnitedHealth Group was never included and was removed during the selection process. Since the stepwise selection method gave a 3 variable model and the majority of the selection technique also gave the same 3 variable model, I believe the 3 variable model is our best model. Next I will be testing for any multicollinearity from our 3 variable model.

Since I have determined potentially our best model, I decided to test if multicollinearity exists within our model. Looking at Table 7, we see that Anthem and Humana have a linear correlation coefficient of 0.71561 which is moderately high. To check if it is multicollinearity, I ran a regression analysis. Looking at Table 14, we have F-Value = 146.17 and the P-value = <0.0001 which are very significant. The Root MSE = 2.81404 and Coefficient of Variation = 4.30478 are also both very significant. When looking at the R^2 Adjusted we have 0.6362 which is moderate and would be better if it were a larger value. Since it's not a very high value, we can say this is evidence against multicollinearity existing within our model. Looking at each of the VIF values of the independent variables, each variable is below VIF of 10 which is very significant. When looking at Table 15, we have moderate distribution which is decent. Looking at Table 16, we see the majority of the data points on the line with a couple outliers on the bottom left of the graph. Looking at Table 17, we see a clear normal distribution pattern which is significant. Looking at Table 18, we see that the residual is longer than the fit-mean on the bottom of the graph, but those are outliers and the mass majority of the data points are within the length of the fit-mean data points. Some of the graphs show weak signs of multicollinearity, but the overall model and the VIF are very significant so it is safe to conclude that there is no multicollinearity within our model..

I finally decided to test interaction terms and quadratic terms to see if they help improve our model for predicting Centene's daily closing price. First, I made an interaction plot of each independent variable to see if any interaction terms are necessary for this model. Looking at Table 19 - 21, there is clearly interaction between all the independent variables, hence interaction terms are necessary for this model. For the interaction term: Ho: B4=B5=B6=0; Ha: B4=B5=B6 does not equal 0. When we look at Table 22, we see that the interaction between Anthem and Cigna our t-value = 4.85 and p-value = <.0001. We can see that the p-value is less than the acceptable value of 0.05. It is safe to conclude that we reject the null hypothesis and the interaction term is significant in the model for predicting Centene's daily closing price. The interaction between Anthem and Humana our t-value = -1.45 and p-value = .1477. We can see that the p-value is greater than the acceptable value of 0.05. It is safe to conclude that we fail to reject the null hypothesis and the interaction term has no significance in the model for predicting Centene's daily closing price. The interaction between Cigna and Humana our t-value = -4.57 and p-value = <.0001. We can see that the p-value is less than the acceptable value of 0.05. It is safe to conclude that we reject the null hypothesis and the interaction term is significant in the model for predicting Centene's daily closing price. For the quadratic terms: Ho: B7=B8=B9= 0; Ha: B7=B8=B9 does not equal 0. When we look at Table 23, the quadratic term for Anthem our t-value = -2.00 and p-value = 0.0464. We can see that the p-value is less than the acceptable value of 0.05. It is safe to conclude that we reject the null hypothesis and the quadratic term is significant in the model for predicting Centene's daily closing price. The quadratic term for Cigna our t-value = -8.26 and p-value = <.0001. We can see that the p-value is less than the acceptable value of 0.05. It is safe to conclude that we reject the null hypothesis and the quadratic term is significant in the model for predicting Centene's daily closing price. The quadratic term for Humana our t-value = -0.44 and p-value = 0.6580. We can see that the p-value is greater than the acceptable value of 0.05. It is safe to conclude that we fail to reject the null hypothesis and the quadratic term has no significance in the model for predicting Centene's daily closing price.

The stepwise regression method resulted in a model which included the best 3 variables out of the 6 variables and included both interaction & quadratic terms of the independent variables. This is because compared to all other variables and models, they have the most significant values which will best predict Centene's daily closing price. The final model is the following, Table 24:

Centene's Daily Closing Price (y) = -177.16483 - 0.79200(Anthem, x1) + 1.61720(Cigna, x2) + 0.80402(Humana, x3) + 0.00482(Anthem*Cigna, x4) - 0.00206(Anthem*Humana, x5) + 0.00093211(Anthem^2, x6) - 0.00734(Cigna^2, x7)

The interpretation of each coefficient of the final model is as follows: B0 = -177.16483 represents the estimated mean Centene's daily closing price when all independent variables are equal to 0. B1 = -0.79200: We estimate the mean Centene's daily closing price to decrease by 0.79200 for every 1 unit increase in Anthem's closing price, x1, when all other variables are held fixed. B2 = 1.61720: We estimate the mean Centene's daily closing price to increase by 1.61720 for every 1 unit increase in Cigna's closing price, x2, when all other variables are held fixed. B3 = 0.80402: We estimate the mean Centene's daily closing price to increase by 0.80402 for every 1 unit increase in Humana's closing price, x3, when all other variables are held fixed. B4 = 0.00482: We estimate the mean Centene's daily closing price to increase by 0.00482 for every 1 unit increase in Anthem's times Cigna's closing price, x4, when all other variables are held fixed. B5 = -0.00206: We estimate the mean Centene's daily closing price to decrease by 0.00206 for every 1 unit increase in Anthem's times Huamana's closing price, x5, when all other variables are held fixed. B6 = 0.00093211: We estimate the mean Centene's daily closing price to increase by 0.00093211 for every 1 unit increase in Anthem's squared closing price, x6, when all other variables are held fixed. B7 = -0.00734: We estimate the mean Centene's daily closing price to decrease by 0.00734 for every 1 unit increase in Cigna's squared closing price, x6, when all other variables are held fixed.

The final model has an R-Square Adjusted = 0.7648 which means that 76.48% of the variation in Centene's daily closing price are explained by Anthem, Cigna, Humana, Anthem times Cigna, Anthem times Humana, Anthem^2, and Cigna^2. The R-Squared Adjusted is greater than the acceptable value for R-Squared Adjusted of 50%. The F-Statistic = 116.67 and P-Value = <0.0001, Root MSE = 2.26277, Coefficient of Variation = 3.46147, Table 24, shows our model is significant. I also noticed when we removed the non-significant interaction and quadratic term, our overall model and the independent variables t-test statistics have significantly improved, so it is safe to say we have found our best prediction model.

The residual plot in Table 25 shows that there is a random disbursement of data points on the plot which is the appropriate graph outcome. The normal probability plot in Table 26 shows that points are generally close to the line. There are a few outliers on the bottom left corner of the graph.

In conclusion, I am able to create a statistically significant model to predict Centene's daily closing price using Anthem, Cigna, Humana, Anthem times Cigna, Anthem times Humana, Anthem^2, and Cigna^2.

Future Work

Given that I am only given a year's worth of data of daily closing price, it is difficult to determine whether the selected stocks are capable of predicting Centene's daily closing price. In order for this study to be accurate, I would require more than 1 year of data to be collected and more stocks with the same GICS sub-industry category. Also, the data was collected during the Covid-19 pandemic which means the closing price may be hyperinflated or the stocks are worth less than they really are.

Appendix

Table 1 - Scatter Plot of Centene vs Anthem

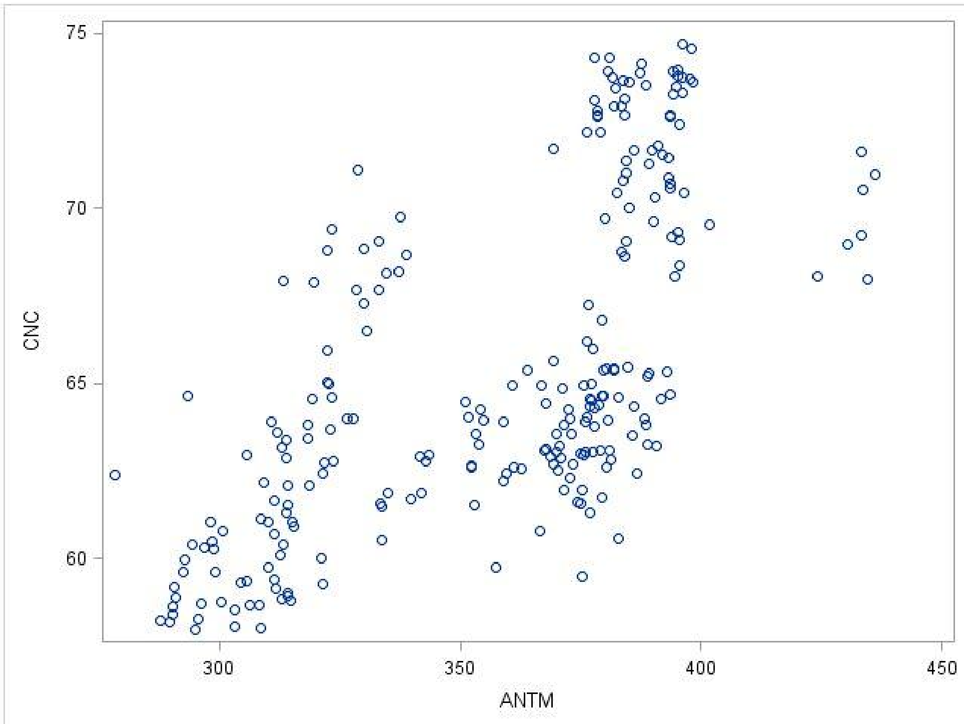


Table 2 - Scatter Plot of Centene vs Cigna

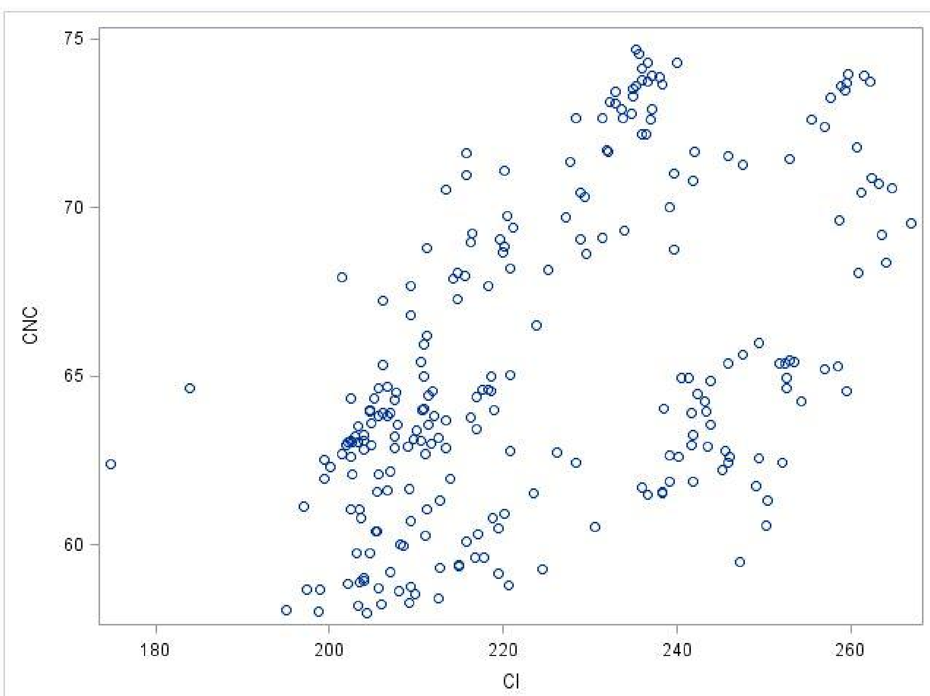


Table 3 - Scatter Plot of Centene vs Humana

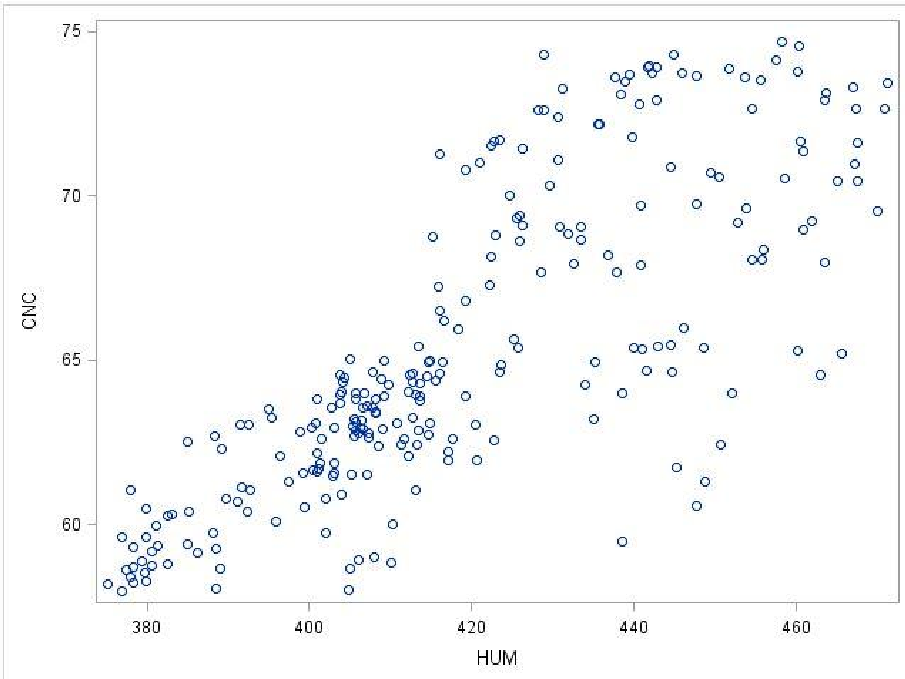


Table 4 - Scatter Plot of Centene vs UnitedHealth Group

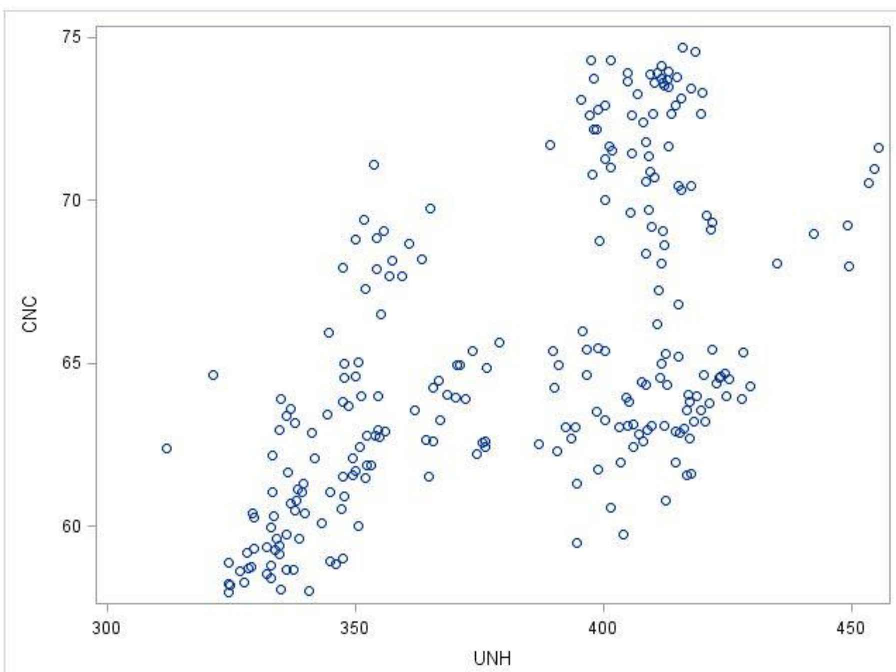


Table 5 -Scatter Plot of Centene vs AbbVie

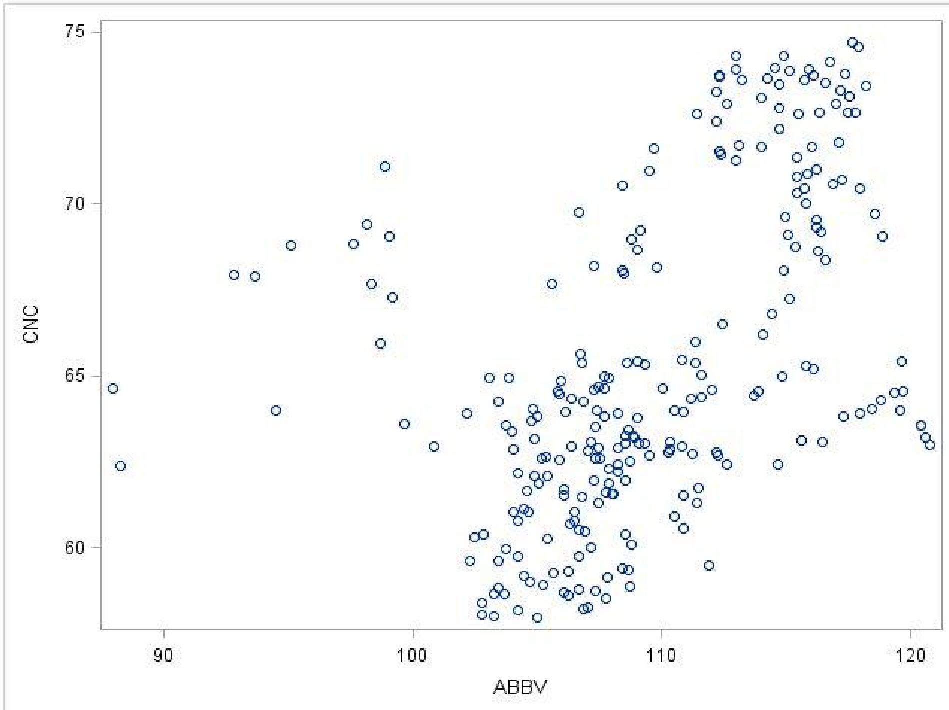


Table 6 - Scatter Plot of Centene vs Abiomed

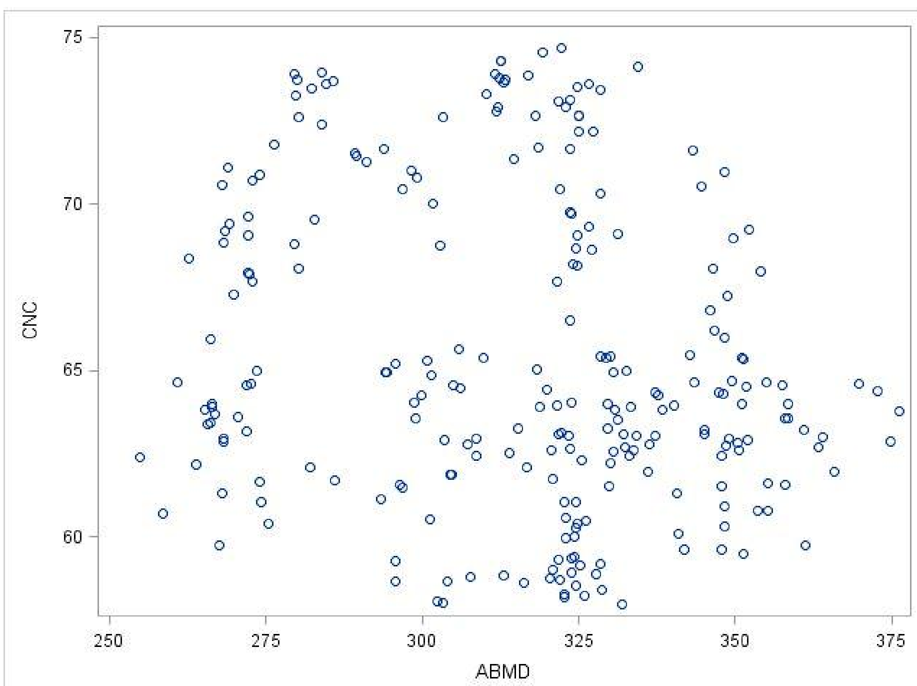


Table 7 - Correlation Matrix

Pearson Correlation Coefficients, N = 250 Prob > |r| under H0: Rho=0							
	CNC	ANTM	CI	HUM	UNH	ABBV	ABMD
CNC CNC	1.00000	0.66366 <.0001	0.52446 <.0001	0.78205 <.0001	0.59460 <.0001	0.47655 <.0001	-0.19872 0.0016
ANTM ANTM	0.66366 <.0001	1.00000	0.47385 <.0001	0.71561 <.0001	0.95953 <.0001	0.64844 <.0001	0.26800 <.0001
CI CI	0.52446 <.0001	0.47385 <.0001	1.00000	0.56508 <.0001	0.30711 <.0001	0.42305 <.0001	-0.23184 0.0002
HUM HUM	0.78205 <.0001	0.71561 <.0001	0.56508 <.0001	1.00000	0.65354 <.0001	0.41880 <.0001	-0.07251 0.2534
UNH UNH	0.59460 <.0001	0.95953 <.0001	0.30711 <.0001	0.65354 <.0001	1.00000	0.68917 <.0001	0.40564 <.0001
ABBV ABBV	0.47655 <.0001	0.64844 <.0001	0.42305 <.0001	0.41880 <.0001	0.68917 <.0001	1.00000	0.32072 <.0001
ABMD ABMD	-0.19872 0.0016	0.26800 <.0001	-0.23184 0.0002	-0.07251 0.2534	0.40564 <.0001	0.32072 <.0001	1.00000

Table 8 - Descriptive Analysis

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
CNC	CNC	250	65.3701600	4.6657318	57.9700010	74.6999970
ANTM	ANTM	250	356.9464798	36.1691132	278.2500000	436.2399900
CI	CI	250	224.4849200	19.3444671	174.8399960	266.9100040
HUM	HUM	250	419.5595192	24.5753870	375.1499940	471.2200010
UNH	UNH	250	382.3649202	34.6687737	312.1000060	455.4400020
ABBV	ABBV	250	109.5558801	5.8323296	87.9599990	120.7799990
ABMD	ABMD	250	316.0664791	28.3462383	254.8500060	376.2000120

Table 9 - Variable Selection: Stepwise

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	HUM		HUM	1	0.6116	0.6116	19.7099	390.52	<.0001
2	ANTM		ANTM	2	0.0222	0.6338	6.5392	14.96	0.0001
3	CI		CI	3	0.0068	0.6406	3.8601	4.68	0.0314

Table 10 - Variable Selection: R-Square

Number in Model	R-Square	Variables in Model
1	0.6116	HUM
1	0.4404	ANTM
1	0.3536	UNH
1	0.2751	CI
2	0.6338	ANTM HUM
2	0.6238	HUM UNH
2	0.6216	CI HUM
2	0.4973	ANTM CI
2	0.4826	CI UNH
2	0.4629	ANTM UNH
3	0.6406	ANTM CI HUM
3	0.6397	ANTM HUM UNH
3	0.6362	CI HUM UNH
3	0.4974	ANTM CI UNH
4	0.6419	ANTM CI HUM UNH

Table 11 - Variable Selection: R-Square Adjusted

Number in Model	Adjusted R-Square	R-Square	Variables in Model
3	0.6362	0.6406	ANTM CI HUM
4	0.6360	0.6419	ANTM CI HUM UNH
3	0.6353	0.6397	ANTM HUM UNH
3	0.6318	0.6362	CI HUM UNH
2	0.6308	0.6338	ANTM HUM
2	0.6207	0.6238	HUM UNH
2	0.6185	0.6216	CI HUM
1	0.6100	0.6116	HUM
2	0.4932	0.4973	ANTM CI
3	0.4913	0.4974	ANTM CI UNH
2	0.4784	0.4826	CI UNH
2	0.4586	0.4629	ANTM UNH
1	0.4382	0.4404	ANTM
1	0.3509	0.3536	UNH
1	0.2721	0.2751	CI

Table 12 - Variable Selection: Capability Potential

Number in Model	C(p)	R-Square	Variables in Model
3	3.8601	0.6406	ANTM CI HUM
3	4.4889	0.6397	ANTM HUM UNH
4	5.0000	0.6419	ANTM CI HUM UNH
2	6.5392	0.6338	ANTM HUM
3	6.8645	0.6362	CI HUM UNH
2	13.3835	0.6238	HUM UNH
2	14.8618	0.6216	CI HUM
1	19.7099	0.6116	HUM
2	99.8992	0.4973	ANTM CI
3	101.8319	0.4974	ANTM CI UNH
2	109.9711	0.4826	CI UNH
2	123.4334	0.4629	ANTM UNH
1	136.7997	0.4404	ANTM
1	196.2448	0.3536	UNH
1	249.9424	0.2751	CI

Table 13 - Variable Selection: PRESS

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	2	3435.38620	1717.69310	213.73
Error	247	1985.10798	8.03687	
Corrected Total	249	5420.49418		

Root MSE	2.83494
Dependent Mean	65.37016
R-Square	0.6338
Adj R-Sq	0.6308
AIC	775.99192
AICC	776.15518
PRESS	2027.48375
SBC	534.55630

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	5.411944	3.131211	1.73
ANTM	1	0.027501	0.007111	3.87
HUM	1	0.119510	0.010466	11.42

Table 14 - Proc Reg: Variance Inflation Factor

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3472.45976	1157.48659	146.17	<.0001
Error	246	1948.03442	7.91884		
Corrected Total	249	5420.49418			

Root MSE	2.81404	R-Square	0.6406
Dependent Mean	65.37016	Adj R-Sq	0.6362
Coeff Var	4.30478		

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	4.33222	3.14794	1.38	0.1700	0
ANTM	ANTM	1	0.02565	0.00711	3.61	0.0004	2.07984
CI	CI	1	0.02435	0.01126	2.16	0.0314	1.49077
HUM	HUM	1	0.11063	0.01117	9.90	<.0001	2.36944

Table 15 - Residual vs Predicted Value Plot

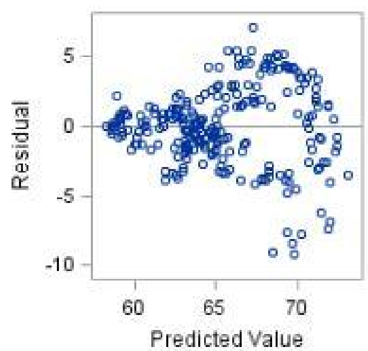


Table 16 - Residual vs Quantile Plot

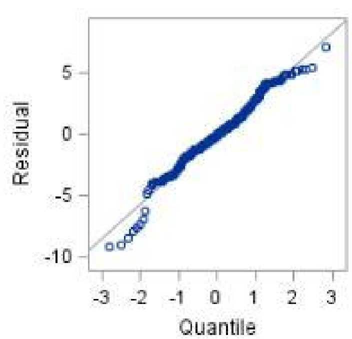


Table 17 - Distribution Plot

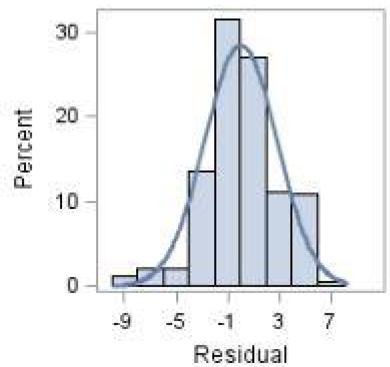


Table 18 - Fit-Mean and Residual vs Proportion Less Plot

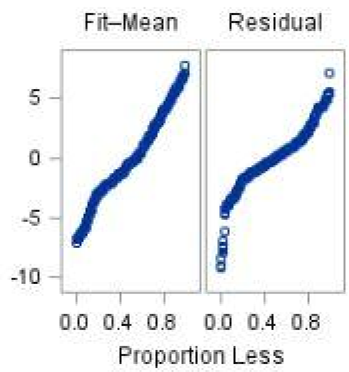


Table 19 - Interaction Between Anthem and Cigna

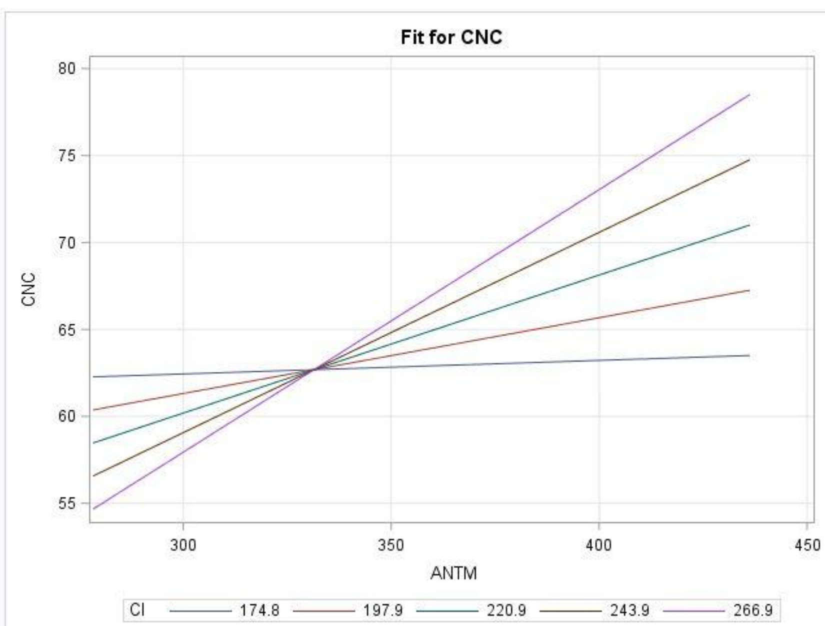


Table 20 - Interaction Between Anthem and Humana

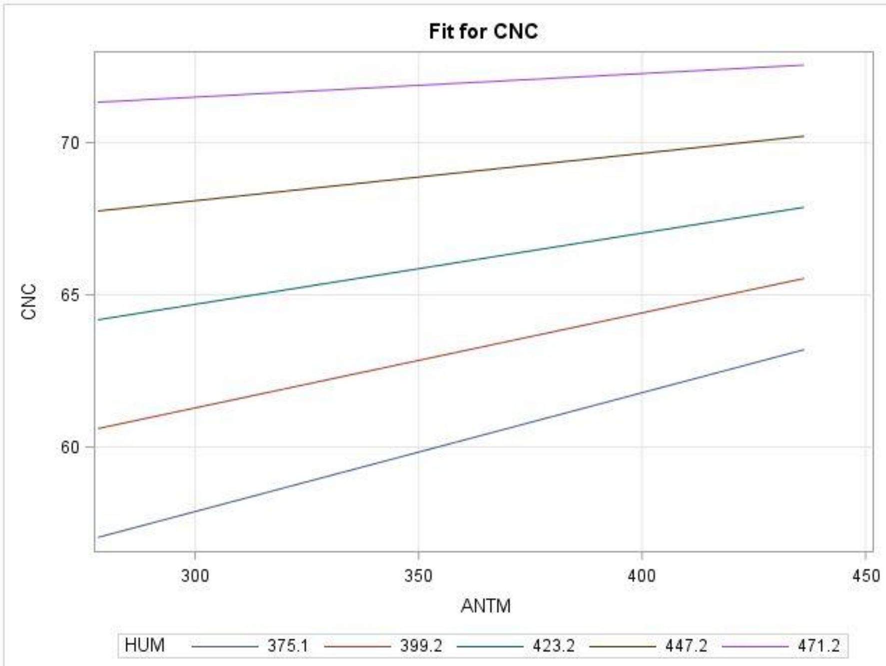


Table 21 - Interaction Between Cigna and Humana

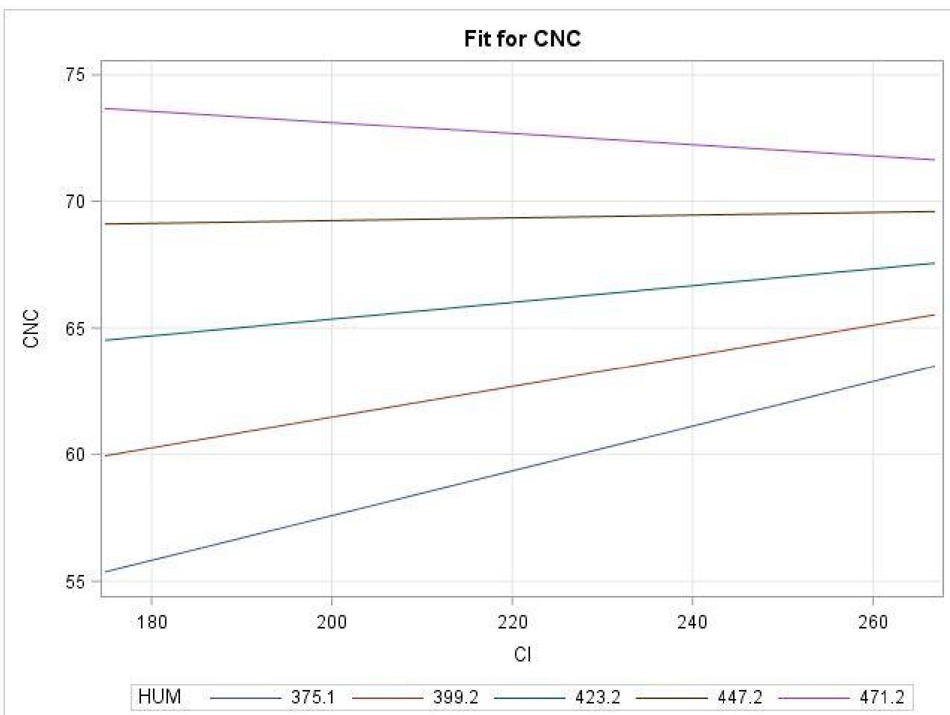


Table 22 - Proc Reg: Interaction Terms

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	3688.93032	614.82172	86.28	<.0001
Error	243	1731.56386	7.12578		
Corrected Total	249	5420.49418			

Root MSE	2.66942	R-Square	0.6806
Dependent Mean	65.37016	Adj R-Sq	0.6727
Coeff Var	4.08354		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-122.14824	44.27923	-2.76	0.0062
ANTM	ANTM	1	-0.31707	0.12090	-2.62	0.0093
CI	CI	1	0.39422	0.19629	2.01	0.0457
HUM	HUM	1	0.81145	0.13733	5.91	<.0001
antm_CI		1	0.00218	0.00044927	4.85	<.0001
antm_hum		1	-0.00028919	0.00019911	-1.45	0.1477
ci_hum		1	-0.00275	0.00060190	-4.57	<.0001

Table 23 - Proc Reg: Interaction & Quadratic Terms

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	4094.35212	511.79402	93.01	<.0001
Error	241	1326.14205	5.50266		
Corrected Total	249	5420.49418			

Root MSE	2.34578	R-Square	0.7553
Dependent Mean	65.37016	Adj R-Sq	0.7472
Coeff Var	3.58845		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-129.32912	50.72296	-2.55	0.0114
ANTM	ANTM	1	-0.84172	0.13165	-6.39	<.0001
CI	CI	1	1.55774	0.22632	6.88	<.0001
HUM	HUM	1	0.67161	0.26903	2.50	0.0132
antm_CI		1	0.00548	0.00061072	8.97	<.0001
ci_hum		1	-0.00198	0.00079570	-2.49	0.0134
antm2		1	-0.00040973	0.00020467	-2.00	0.0464
ci2		1	-0.00597	0.00072275	-8.26	<.0001
hum2		1	-0.00019548	0.00044102	-0.44	0.6580

Table 24 - Proc Reg: Best Final Model

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	4181.42698	597.34671	116.67	<.0001
Error	242	1239.06720	5.12011		
Corrected Total	249	5420.49418			

Root MSE	2.26277	R-Square	0.7714
Dependent Mean	65.37016	Adj R-Sq	0.7648
Coeff Var	3.46147		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-177.16483	38.35236	-4.62	<.0001
ANTM	ANTM	1	-0.79200	0.11028	-7.18	<.0001
CI	CI	1	1.61720	0.20231	7.99	<.0001
HUM	HUM	1	0.80402	0.12178	6.60	<.0001
antm_CI		1	0.00482	0.00050477	9.54	<.0001
antm_hum		1	-0.00206	0.00034072	-6.05	<.0001
antm2		1	0.00093211	0.00026296	3.54	0.0005
ci2		1	-0.00734	0.00067200	-10.93	<.0001

Table 25 - Residual vs Predicted Value

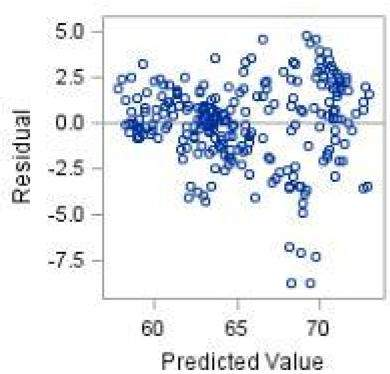
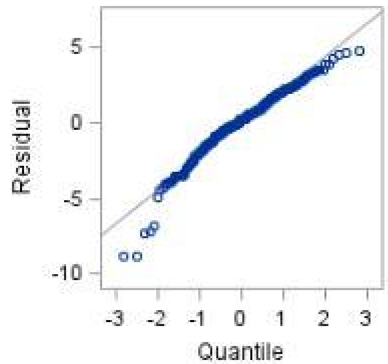


Table 26 - Normal Probability Plot



SAS Code

```
proc contents data=stock; run;

proc sgplot data=stock;
 scatter y=CNC x=antm;
 run;

proc sgplot data=stock;
 scatter y=CNC x=CI;
 run;

proc sgplot data=stock;
 scatter y=CNC x=HUM;
 run;

proc sgplot data=stock;
 scatter y=CNC x=UNH;
 run;

proc sgplot data=stock;
 scatter y=CNC x=abbv;
 run;

proc sgplot data=stock;
 scatter y=CNC x=abmd;
 run;

proc corr data=stock;
 var CNC antm ci hum unh abbv abmd;
 run;

proc means data=stock;
 var cnc antm ci hum unh abbv abmd;
 run;

proc reg data=stock;
 model CNC=antm ci hum unh/selection = stepwise;
 run;

proc reg data=stock;
 model CNC=antm ci hum unh/selection = rsquare;
 run;

proc reg data=stock;
 model CNC=antm ci hum unh/selection = adjrsq;
 run;

proc reg data=stock;
 model CNC=antm ci hum unh/selection = cp;
```

```
 run;

proc glmselect data=stock;
 model CNC=antm ci hum unh/selection = stepwise(choose=press);
 run;

proc reg data=stock;
model CNC=antm ci hum/VIF;
run;

proc glm data=stock;
 model cnc=antm | ci / solution;
 ods select ParameterEstimates ContourFit;
 store GLMModel;
 run;

proc plm restore=GLMModel noinfo;
   effectplot slicefit(x=antm sliceby=ci);
run;

proc glm data=stock;
 model cnc=antm | hum / solution;
 ods select ParameterEstimates ContourFit;
 store GLMModel;
 run;

proc plm restore=GLMModel noinfo;
   effectplot slicefit(x=antm sliceby=hum);
run;

proc glm data=stock;
 model cnc=ci | hum / solution;
 ods select ParameterEstimates ContourFit;
 store GLMModel;
 run;

proc plm restore=GLMModel noinfo;
   effectplot slicefit(x=ci sliceby=hum);
run;

data stock2;
set stock;
antm_CI = antm*ci;
antm_hum = antm*hum;
ci_hum = ci*hum;
antm2 = antm**2;
ci2 = ci**2;
hum2 = hum**2;

proc reg data=stock2;
model cnc = antm ci hum antm_ci antm_hum ci_hum;
```

```
run;

proc reg data=stock2;
model cnc=antm ci hum antm_ci ci_hum antm2 ci2 hum2;
run;

proc reg data=stock2;
model cnc= antm ci hum antm_ci antm_hum antm2 ci2;
run;
```