Predicting CPU Performance from Hardware Characteristics

by
Tan Tran and Fan Yang
STAT 410/510
Regression Analysis
Professor Hojin Moon

California State University Long Beach
December 18, 2014

Acknowledgments

First and foremost, we would like to thank Dr. Hojin Moon, our course instructor, for his valuable guidance and support for the completion of this project. Despite his busy schedule, he has been very patient and kind, and his excellent teaching helped us be better prepared for this project and things beyond.

Secondly, we are also thankful to our parents. With their encouragement and support, we have been able to complete this project on time.

Thirdly, we would like to thank Shiva Mostaghel for her help in the initial stages of this project.

Last but not the least, we give our thanks to everyone who helped us in some way.

Abstract

A central processor known as the central processing unit (CPU) has always been considered important in determining the performance of the system. However, when it comes to evaluating the performance of central processing units, it can be a really difficult task to find the most effective solution to the problem of performance comparisons. The objective of this project is to illustrate a method for developing a multiple regression model to predict the relative performance of central processing units (CPUs) based on the information of cache memory size, average memory size, and channel capacity. The use of statistical computing package SAS is implemented to perform the tests in the following discussion and analysis.

Introduction

A data set titled "Computer Hardware Data Set" was obtained from the UCI Machine Learning Repository. The data set was compiled by Drs. Phillip Ein-Dor and Jacob Feldmesser during research conducted in 1987 on the fitting of a linear regression model to predict computer CPU performance from hardware characteristics such as machine cycle time, memory size, and number of input/output channels. The research was documented in a journal article titled "Attributes of the Performance of Central Processing Units: A Relative Performance Prediction Model", hereby referred to as [Ein-Dor/Feldmesser].

In this student project, the student authors sought to replicate [Ein-Dor/Feldmesser]'s findings using methods described in Kutner et. al.'s *Applied Linear Regression Models* textbook. The variable transformations described in [Ein-Dor/Feldmesser] were performed on the obtained data set using the SAS (Statistical Analysis Software) software suite. Additional statistical tests and analyses were made independently in order to reinforce understanding of course material. In summary: the model given in [Ein-Dor/Feldmesser] was reproduced, tests for performed for significance of overall model and of individual parameters, residuals were analyzed, and the model was verified on a testing set.

Table 1 summarizes the data set obtained.

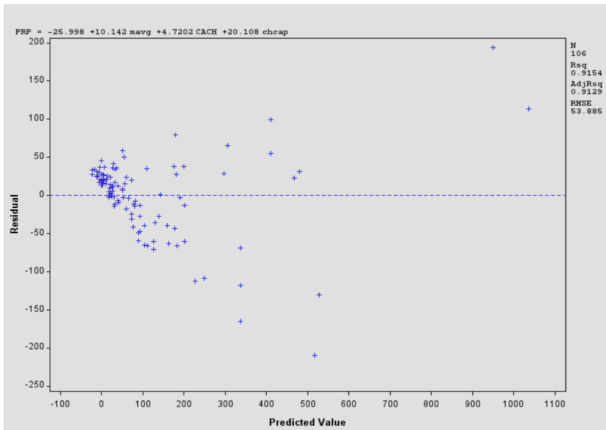
Table 1. Attributes of "Computer Hardware Data Set"

Attribute	**Description**	**Sample data: Min … Max**
Vendor name	String: CPU vendor	'HP', 'IBM', 'Siemens', ...
Model name	String: CPU model	'v8535', 'universe:2203t', ...
MYCT	Integer: Machine cycle time (nanoseconds)	17, 23, …, 1100, 1500
MMIN	Minimum main memory (kilobytes)	64, 128, …, 16000, 32000
MMAX	Maximum main memory (kilobytes)	64, 512, …, 32000, 64000
CACH	Cache memory (kilobytes)	0, 4, …, 160, 256
CHMIN	Minimum I/O channels (# channels)	0, 1, …, 32, 52
CHMAX	Maximum I/O channels (# channels)	0, 1, … 128, 176
PRP	Published relative performance (dimensionless, higher PRP being faster)	6, 7, …, 1144, 1150
ERP	Paper authors' estimated relative performance	15, 17, … 978, 1238

Construction of Regression Model

First, the data set was divided into a training set and a testing set of 106 and 103 observations, respectively. A first-order model predicting the PRP response variable from three calculated predictor variables was fitted to the training set in order to confirm the need for further refinement. The three calculated predictor variables were given by [Ein-Dor/Feldmesser] as MAVG, CACH, CHCAP, and are given in Table 2.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Variance Inflation
Intercept	1	-25.99818	7.81152	-3.33	0.0012	0	0
mavg	1	10.14171	1.24065	8.17	<.0001	0.43574	3.42576
CACH	1	4.72023	1.68373	2.80	0.0061	0.11037	1.86862
chcap	1	20.10825	1.99423	10.08	<.0001	0.49059	2.85400



The Box-Cox procedure was done in order to assess the best transformation on PRP. This procedure resulted in a Box-Cox parameter λ=0.5, which indicated that the best transformation was PRP' = $PRP^{0.5}$. This value for λ agrees with the transformation done in [Ein-Dor/Feldmesser], where the transformed Y-value PRP' is called SQRPERF.

A final model matching that of [Ein-Dor/Feldmesser] was arrived at, and the regression output given in Tables 5. The final model fitted on the training set was: PRP_SQRT = 3.95 + 0.432*MAVG + 0.317*CACH + 0.315*CHCAP. This model indicates that performance increases as any one of MAVG, CACH, or CHCAP increases.

The R^2 value of the model was found to be 0.9281, indicating that the model explains 92.81% of the variance in performance across CPUs. The F statistic of this model was found to be 438.77 with p-value 0+, indicating that the model is significant (that not all of β_1, β_2, and β_3 = 0).

Multicollinearity was examined for by testing for the significance of individual parameters and observing their VIFs (variance inflation factors). Automated tests shown in the regression output indicate that all three predictors are individually significant with p-values 0+. Variance inflation factors were observed to be 3.43, 1.87, and 2.85 for MAVG, CACH, and CHCAP, respectively. Because no VIF nor the mean VIF is greater than 10, it was concluded that the model exhibits no multicollinearity.

Tables 4. Box-Cox Procedure Output

The TRANSREG Procedure

Calculated Variable	Formula	Explanation
MAVG: Average Memory	MAVG = (MMIN+MMAX)/2 * 10^{-3}	Average of CPU's (minimum memory) and (maximum memory)
CACH: Cache in MB	CACH = CACHE * 10^{-1}	Transformation done in order to scale all predictors to magnitude of 10^{1}
CHCAP: Total channel capacity	CHCAP = CHAVG*SPEED*10, where: CHAVG = INT((CHMIN+CHMAX)/2) + 1, and SPEED = 1 / MYCT	Take the average number of channels, multiply by speed per channel, then by 10.

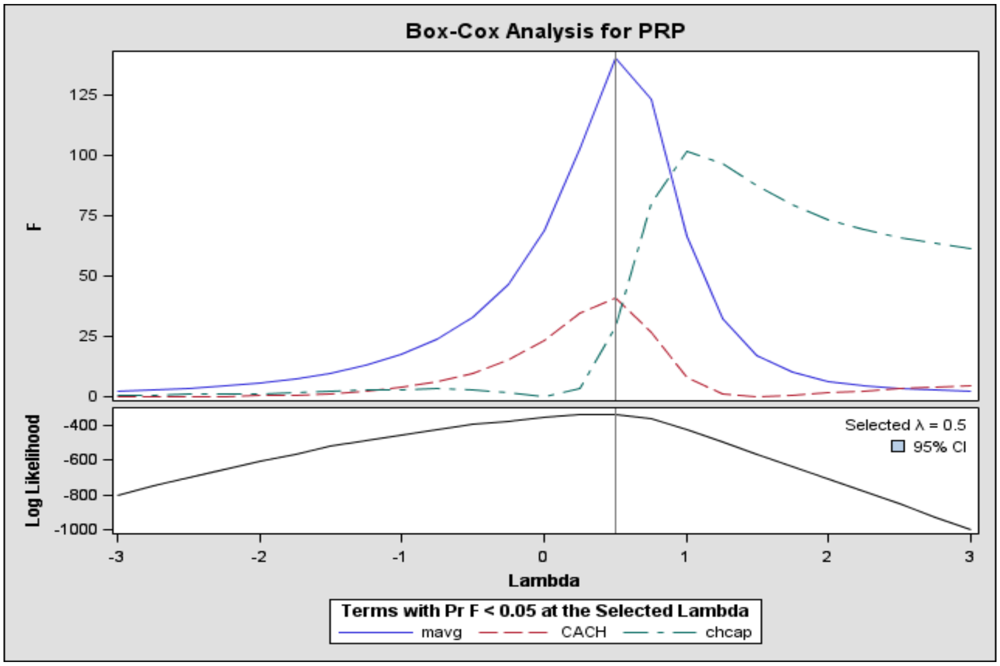
Tables 3 show the output from fitting the PRP response variable from these three predictors. Residuals display a fan-out shape, indicating the need for a transformation on the PRP response variable.

Tables 3. Regression Model *PRP = mavg cach chcap*

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	3204526	1068175	367.88	<.0001
Error	102	296167	2903.59541		
Corrected Total	105	3500692			

Root MSE	53.88502	**R-Square**	0.9154
Dependent Mean	113.64151	**Adj R-Sq**	0.9129
Coeff Var	47.41667		

Parameter Estimates



Dependent Variable BoxCox(PRP)

Number of Observations Read	106
Number of Observations Used	106

The TRANSREG Procedure Hypothesis Tests for BoxCox(PRP)

Univariate ANOVA Table Based on the Usual Degrees of Freedom

Source	DF	Sum of Squares	Mean Square	F Value	Liberal p
Model	**3**	13174.94	4391.648	438.77	>= <.0001
Error	**102**	1020.92	10.009		
Corrected Total	**105**	14195.86			

The above statistics are not adjusted for the fact that the dependent variable was transformed and so are generally liberal.

Root MSE	3.16370	**R-Square**	0.9281
Dependent Mean	15.90650	**Adj R-Sq**	0.9260
Coeff Var	19.88933	**Lambda**	0.5000

Tables 5. Regression Model *PRP_SQRT = MAVG CACH CHCAP*

The REG Procedure
Model: MODEL1
Dependent Variable: prp_sqrt

Number of Observations Read	106
Number of Observations Used	106

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	3293.73622	1097.91207	438.77	<.0001
Error	102	255.22891	2.50224		
Corrected Total	105	3548.96513			

Root MSE	1.58185	**R-Square**	0.9281
Dependent Mean	8.95325	**Adj R-Sq**	0.9260
Coeff Var	17.66787		

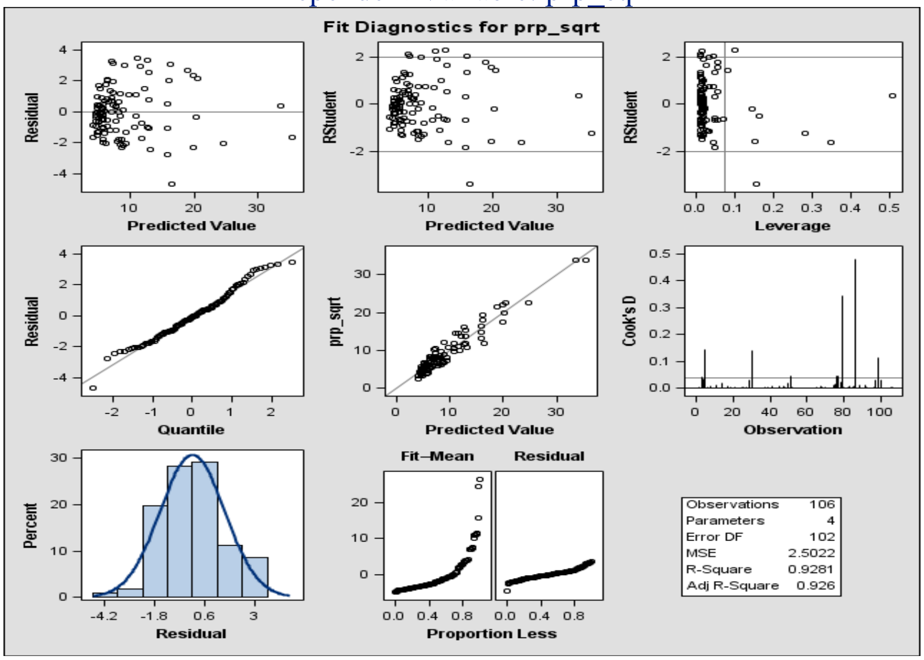
Parameter Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	**1**	3.94828	0.22931	17.22	<.0001	0
mavg	**1**	0.43193	0.03642	11.86	<.0001	3.42576
CACH	**1**	0.31677	0.04943	6.41	<.0001	1.86862
chcap	**1**	0.31494	0.05854	5.38	<.0001	2.85400

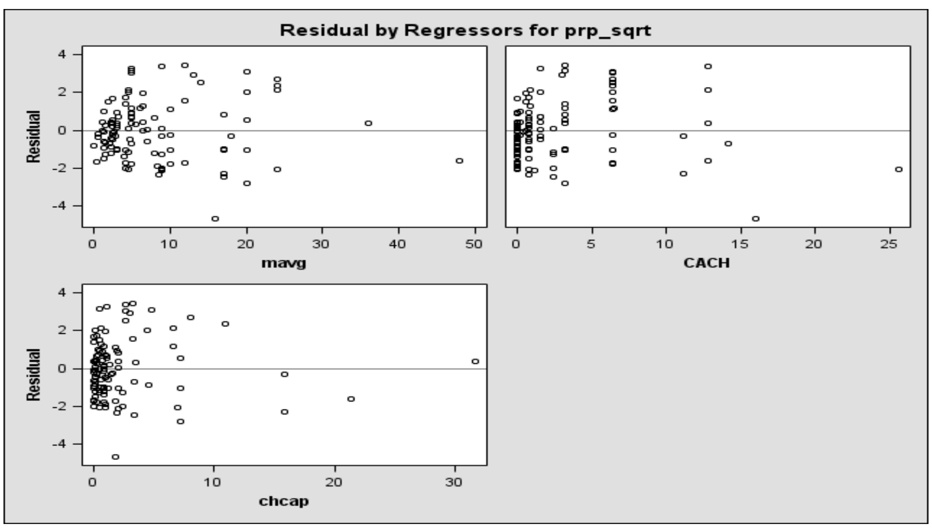
Residual Analysis

Residual analysis was performed visually using the following diagnostic plots (Tables 6): residual vs. predicted value, residual against each of three predictors, normal distribution plot of residuals, and box plot of residuals. Residuals were concluded to be normally distributed, but to display crowding around 0 for low values of the predicted response variable PRP_SQRT.

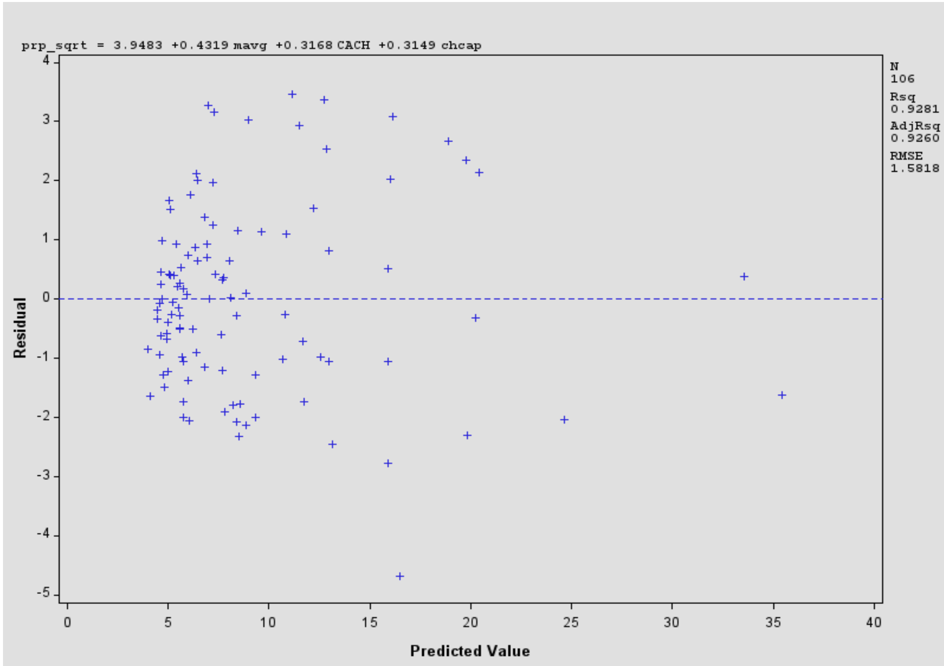
Tables 6. Diagnostic Plots of Regression Model PRP_SQRT = MAVG CACH CHCAP

The REG Procedure
Model: MODEL1
Dependent Variable: prp_sqrt





The REG Procedure



The following tests on residuals were done: correlation test for normality, the Shapiro-Wilk test for normality, lack-of-fit test, and the Breusch-Pagan test for constant error variance.

Conclusions of normality of residuals were confirmed using the test for correlation of residuals with expected normal values. The correlation coefficient was determined to be 0.99099, which was greater than the threshold value of 0.98, thus indicating that residuals were normally distributed. Correlation output is shown in Tables 7.

Tables 7. Correlation of Residuals with Expected Values under Normal Assumption

The CORR Procedure

2 Variables:	ev residual

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
ev	106	0	1.56555	0	-3.98487	3.98487	
residual	106	0	1.55909	0	-4.67539	3.46011	Residual

Pearson Correlation Coefficients, N = 106 Prob > |r| under H0: Rho=0		
	ev	**residual**

Pearson Correlation Coefficients, N = 106 Prob > |r| under H0: Rho=0		
	ev	residual
ev	1.00000	0.99099 <.0001
residual Residual	0.99099 <.0001	1.00000

The Shapiro-Wilk test was performed in order to test the null hypothesis that the observations came from a normally distributed population. The test p-value was 0.1541, which indicated the null hypothesis. Output is shown in Table 8.

Table 8. Shapiro-Wilk Test for Normal Distribution of Population
The UNIVARIATE Procedure
Variable: residual (Residual)

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	W	0.981765	Pr < W	0.1541
Kolmogorov-Smirnov	D	0.056604	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.059039	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.490226	Pr > A-Sq	0.2239

Residual analysis concluded with the Breusch-Pagan test for homoscedasticity of residuals. The X^2*(3 df) statistic was calculated to be (SSR*)/2 ÷ $(SSE/106)^2$ = (299.7)/2 ÷ $(255/106)^2$ = 25.85 with p-value 0+, indicating the alternative hypothesis that residuals are hetereoscedastic.

Table 9. ANOVA Results for *ressqr* used for Breusch-Pagan Test

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	299.72203	99.90734	11.19	<.0001
Error	102	910.64699	8.92791		
Corrected Total	105	1210.36903			

An attempt to remedy unequal variances in residuals was made using the method of weighted least squares. Linear regression output and ressqr ANOVA used to re-calculate Breusch-Pagan are shown in Tables 10 and 11, respectively. The X^2*(3 df) statistic was calculated to be (SSR*)/2 ÷ $(SSE/106)^2$ = (785)/2 ÷ $(147/106)^2$ = 255.23 with p-value 0+, indicating the alternative hypothesis that residuals remained heteroscedastic, such that the method of least squares failed. The unweighted model was kept.

Tables 10. Linear Regression Model after Performing Method of Least Squares

The REG Procedure
Model: MODEL1
Dependent Variable: prp_sqrt

Number of Observations Read	106
Number of Observations Used	106

Weight: wt

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	2469.22856	823.07619	570.48	<.0001
Error	102	147.16451	1.44279		
Corrected Total	105	2616.39307			

Root MSE	1.20116	**R-Square**	0.9438
Dependent Mean	7.35200	**Adj R-Sq**	0.9421
Coeff Var	16.33790		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	3.77991	0.16943	22.31	<.0001	3.44384	4.11597
mavg	1	0.43692	0.03865	11.31	<.0001	0.36027	0.51358
CACH	1	0.41328	0.06932	5.96	<.0001	0.27577	0.55078
chcap	1	0.27114	0.04666	5.81	<.0001	0.17859	0.36369

Table 11. ANOVA Results for *ressqr* used for Breusch-Pagan Test after Method of Least Squares was Performed

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	785.45077	261.81692	20.86	<.0001
Error	102	1280.16335	12.55062		
Corrected Total	105	2065.61412			

The procedures to identify outlying Y-values, and to identify outlying X-observations, were performed. No outliers were found in either case. Table 12 in the appendix contains studentized deleted residuals and hat matrix diagonal values for each of the 106 observations. Outlying observations were searched for using rules of thumb:

- |RStudent| $> t_{n-1-p}(1-\alpha/2n)$
  $= t_{106-1-4}(1-.1/(2*106))$
  $= t_{101}(0.9995)$
  $\cong$ **3.40**
- |Hat Diag H| $< 2p/n$
  $= 2*4/106$
  = **0.075**

No outlying observations were identified using either rule of thumb.

Model validation

Model validation was performed using two procedures. First, a test for lack-of-fit was performed. Output is shown in Table 13. The p-value of this test was 0.1489, indicating that the model is reasonably accurate.

Second, the model was applied to the testing set. Mean squared prediction error (MSPR) was calculated to be 2.67. As this value was close to the model MSE of 2.50, the model was concluded to have significant predictive ability. Output is shown in Table 14.

Table 13. PROC RSREG Output Used in Lack-of-Fit Testing

Residual	DF	Sum of Squares	Mean Square	F Value	Pr > F
Lack of Fit	93	252.984334	2.720262	2.28	0.1489
Pure Error	6	7.163335	1.193889		
Total Error	99	260.147670	2.627754		

Table 14. Calculation of MSPR on Testing Set

The MEANS Procedure

Analysis Variable : devsq	
Sum	Mean
275.5031069	2.6747874

Conclusion

This project focuses on how to develop a multiple regression model to predict the relative performance of central processing units on the basis of system components. Based on the preceding analysis, the overall conclusion is that there is evidence of a relationship between the relative performance of central processing units and cache memory size, average memory size, and channel capacity. Although this linear prediction model generally has high predictive ability, it has one failure in having unequal variances of residuals.

The unequal variance of residuals is accepted by [Ein-Dor/Feldmesser] as a consequence of synergistic effects in CPU performance. It is explained that, in order for a CPU to perform well, it must have all three of: high memory, high cache, and high channel capacities. CPUs that perform poorly are bottlenecked by the lack of one or more of these factors, and perform worse than the linear regression model would predict. The linear regression model performs highly when these low-performing CPUs are excluded. That is, the predictive accuracy of the model will be improved as the relative performance of central processing units grows.

Appendix

Table 12. Diagnostics for Outlying Values

The REG Procedure
Model: MODEL1
Dependent Variable: prp_sqrt

Output Statistics

Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS			
						Intercept	mavg	CACH	chcap
1	0.5201	0.3356	0.0484	1.0883	0.0757	-0.0146	0.0521	-0.0429	-0.0034
2	-1.0487	-0.6778	0.0484	1.0735	-0.1529	0.0295	-0.1052	0.0866	0.0070
3	-2.7662	-1.8127	0.0484	0.9618	-0.4090	0.0788	-0.2814	0.2317	0.0186
4	2.3482	1.5395	0.0578	1.0061	0.3812	-0.0883	0.1750	-0.1237	0.0807
5	-1.6220	-1.2136	0.2828	1.3688	-0.7621	0.3745	-0.4538	0.2182	-0.0575
6	1.1623	0.7456	0.0331	1.0524	0.1378	0.0640	-0.0578	0.1133	-0.0181
7	-0.8406	-0.5352	0.0211	1.0506	-0.0785	-0.0785	0.0464	0.0016	-0.0258
8	-1.8974	-1.2146	0.0202	1.0018	-0.1743	-0.0546	-0.0996	0.1052	0.0594
9	-0.5765	-0.3663	0.0183	1.0540	-0.0501	-0.0497	0.0301	-0.0031	-0.0167
10	-0.2877	-0.1826	0.0172	1.0570	-0.0241	-0.0228	0.0104	0.0063	-0.0100
11	-0.7109	-0.4892	0.1622	1.2299	-0.2153	-0.0590	0.1288	-0.2030	-0.0233
12	-0.4985	-0.3159	0.0136	1.0503	-0.0371	-0.0295	0.0003	0.0161	-0.0021
13	-0.0533	-0.0338	0.0139	1.0548	-0.0040	-0.0038	0.0011	0.0005	-0.0003
14	2.9282	1.8937	0.0202	0.9233	0.2718	0.0022	0.1964	-0.0900	-0.1142
15	-0.1483	-0.0940	0.0150	1.0556	-0.0116	-0.0114	0.0055	0.0004	-0.0032
16	-0.4974	-0.3155	0.0155	1.0524	-0.0396	-0.0389	0.0201	0.0017	-0.0129
17	-2.0794	-1.3323	0.0190	0.9890	-0.1854	-0.0467	-0.1164	0.0916	0.0819
18	1.2549	0.7971	0.0130	1.0277	0.0913	0.0512	0.0302	-0.0314	-0.0296
19	2.1188	1.3528	0.0116	0.9795	0.1466	0.1235	-0.0114	-0.0270	-0.0077
20	-0.1880	-0.1194	0.0195	1.0603	-0.0168	-0.0166	0.0102	-0.0022	-0.0043
21	1.7553	1.1178	0.0121	1.0024	0.1235	0.1013	-0.0043	-0.0216	-0.0163

Output Statistics									
Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS			
						Intercept	mavg	CACH	chcap
22	0.0766	0.0485	0.0140	1.0548	0.0058	0.0042	0.0008	-0.0022	-0.0010
23	-0.4932	-0.3127	0.0145	1.0514	-0.0379	-0.0355	0.0167	-0.0073	-0.0024
24	0.4261	0.2699	0.0132	1.0511	0.0312	0.0159	0.0106	-0.0042	-0.0149
25	1.9987	1.2750	0.0119	0.9876	0.1398	0.1085	-0.0061	0.0053	-0.0345
26	-2.0014	-1.2780	0.0138	0.9892	-0.1514	-0.1154	-0.0117	0.0542	0.0227
27	-1.7431	-1.1109	0.0138	1.0048	-0.1316	-0.1003	-0.0101	0.0471	0.0197
28	-1.0527	-0.6683	0.0138	1.0363	-0.0792	-0.0603	-0.0061	0.0283	0.0119
29	3.0249	1.9725	0.0335	0.9253	0.3674	0.2120	-0.2428	0.2787	0.0889
30	3.3634	2.2870	0.0998	0.9442	0.7614	0.1179	-0.2723	0.7235	-0.0963
31	-0.3414	-0.2172	0.0216	1.0612	-0.0323	-0.0319	0.0198	0.0025	-0.0142
32	0.4536	0.2882	0.0193	1.0571	0.0404	0.0400	-0.0221	-0.0045	0.0153
33	0.2088	0.1326	0.0176	1.0581	0.0177	0.0169	-0.0082	-0.0042	0.0075
34	0.9248	0.5865	0.0127	1.0393	0.0665	0.0583	-0.0161	-0.0194	0.0199
35	0.3309	0.2102	0.0188	1.0582	0.0291	0.0235	-0.0119	-0.0094	0.0181
36	0.002964	0.001883	0.0198	1.0612	0.0003	0.0003	-0.0002	-0.0000	0.0001
37	1.5132	0.9640	0.0161	1.0191	0.1231	0.1219	-0.0612	-0.0035	0.0328
38	0.007333	0.004645	0.0137	1.0547	0.0005	0.0003	0.0002	-0.0002	-0.0002
39	1.9719	1.2583	0.0129	0.9902	0.1436	0.0880	0.0384	-0.0618	-0.0271
40	0.5260	0.3337	0.0159	1.0524	0.0425	0.0417	-0.0224	-0.0019	0.0154
41	-2.3176	-1.4846	0.0147	0.9683	-0.1812	-0.0782	-0.0771	0.1003	0.0277
42	-0.6153	-0.3906	0.0164	1.0512	-0.0505	-0.0494	0.0205	0.0066	-0.0096
43	0.9230	0.5858	0.0143	1.0411	0.0706	0.0651	-0.0168	-0.0195	0.0101
44	0.1711	0.1084	0.0137	1.0542	0.0128	0.0124	-0.0050	-0.0010	0.0027
45	-0.9763	-0.6387	0.0717	1.1027	-0.1776	0.0510	-0.1622	0.0462	0.1339
46	0.3726	0.2361	0.0136	1.0523	0.0278	0.0236	-0.0148	0.0099	0.0072

Output Statistics									
Obs	**Residual**	**RStudent**	**Hat Diag H**	**Cov Ratio**	**DFFITS**	**DFBETAS**			
						Intercept	**mavg**	**CACH**	**chcap**
47	-1.0602	-0.6879	0.0556	1.0810	-0.1669	0.0444	-0.1515	0.0528	0.1112
48	0.8131	0.5271	0.0556	1.0894	0.1279	-0.0340	0.1161	-0.0405	-0.0852
49	-1.0545	-0.6694	0.0137	1.0362	-0.0790	-0.0763	0.0312	0.0062	-0.0168
50	2.5434	1.6432	0.0266	0.9616	0.2715	-0.0271	0.1451	0.0865	-0.1627
51	2.6783	1.7623	0.0579	0.9782	0.4370	-0.1600	0.3278	-0.1212	-0.0953
52	-1.6432	-1.0499	0.0200	1.0163	-0.1500	-0.1499	0.0839	0.0054	-0.0448
53	0.2407	0.1527	0.0167	1.0569	0.0199	0.0196	-0.0085	-0.0027	0.0045
54	0.7372	0.4673	0.0129	1.0448	0.0535	0.0491	-0.0187	0.0071	0.0010
55	-0.9462	-0.6013	0.0166	1.0428	-0.0782	-0.0767	0.0325	0.0094	-0.0149
56	0.3970	0.2516	0.0140	1.0524	0.0300	0.0266	-0.0044	-0.0079	0.0001
57	0.2635	0.1669	0.0131	1.0529	0.0192	0.0179	-0.0054	-0.0012	0.0007
58	0.6397	0.4053	0.0123	1.0463	0.0452	0.0325	0.0057	-0.0107	-0.0111
59	-0.2884	-0.1834	0.0207	1.0608	-0.0267	-0.0047	-0.0177	0.0074	0.0171
60	-0.2625	-0.1663	0.0144	1.0543	-0.0201	-0.0190	0.0060	0.0035	-0.0023
61	-0.6031	-0.3821	0.0128	1.0476	-0.0435	-0.0246	-0.0136	0.0191	0.0086
62	0.0156	0.009835	0.0103	1.0510	0.0010	0.0007	0.0000	-0.0003	0.0000
63	-0.0779	-0.0494	0.0166	1.0577	-0.0064	-0.0063	0.0027	0.0008	-0.0012
64	0.0965	0.0612	0.0157	1.0567	0.0077	0.0020	0.0043	-0.0011	-0.0045
65	-2.0494	-1.3090	0.0136	0.9859	-0.1534	-0.1186	-0.0079	0.0627	0.0084
66	-1.1528	-0.7315	0.0120	1.0309	-0.0808	-0.0655	0.0188	-0.0202	0.0107
67	-2.1374	-1.3667	0.0142	0.9806	-0.1641	-0.0635	-0.0773	0.0821	0.0348
68	-2.0033	-1.2770	0.0105	0.9860	-0.1316	-0.0645	-0.0390	0.0305	0.0186
69	-1.2895	-0.8182	0.0105	1.0238	-0.0843	-0.0413	-0.0250	0.0195	0.0119
70	-0.8942	-0.5773	0.0473	1.0776	-0.1287	-0.0993	0.0879	0.0250	-0.1067
71	-1.0186	-0.6506	0.0258	1.0500	-0.1058	-0.0137	-0.0210	-0.0636	0.0589

Output Statistics									
Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS			
						Intercept	mavg	CACH	chcap
72	-0.2631	-0.1675	0.0232	1.0637	-0.0258	-0.0044	-0.0035	-0.0161	0.0124
73	1.0948	0.6981	0.0221	1.0434	0.1050	0.0199	0.0111	0.0663	-0.0464
74	3.2823	2.1217	0.0107	0.8832	0.2207	0.1913	-0.0402	-0.0103	0.0065
75	3.4601	2.2466	0.0144	0.8684	0.2716	0.0436	0.1529	-0.0706	-0.0739
76	3.0915	2.0302	0.0450	0.9281	0.4406	-0.1485	0.3518	-0.0314	-0.2267
77	2.1420	1.4208	0.0825	1.0475	0.4261	-0.1605	0.1989	0.2065	-0.1923
78	2.0324	1.3218	0.0483	1.0205	0.2977	-0.1029	0.2425	-0.0184	-0.1671
79	-2.0452	-1.6146	0.3487	1.4423	-1.1814	0.2177	0.0893	-1.0754	0.3152
80	-1.2820	-0.8163	0.0174	1.0312	-0.1087	-0.1073	0.0514	0.0152	-0.0327
81	-1.3746	-0.8740	0.0138	1.0234	-0.1034	-0.0892	0.0135	0.0390	-0.0141
82	-1.2042	-0.7672	0.0194	1.0365	-0.1080	-0.0366	-0.0594	0.0643	0.0360
83	-1.7664	-1.1346	0.0286	1.0180	-0.1948	-0.0233	-0.1413	0.1178	0.0895
84	1.1448	0.7390	0.0451	1.0660	0.1605	0.1051	-0.1166	0.0102	0.1394
85	-1.7389	-1.1148	0.0254	1.0164	-0.1800	-0.0018	-0.0688	-0.0841	0.1080
86	-4.6754	-3.3785	0.1565	0.8036	-1.4554	0.2081	-0.0989	-1.2349	0.7018
87	0.4019	0.2548	0.0146	1.0529	0.0311	0.0290	-0.0080	-0.0069	0.0031
88	-1.7951	-1.1564	0.0338	1.0214	-0.2162	-0.0906	0.0694	-0.1751	0.0553
89	0.4124	0.2615	0.0149	1.0531	0.0321	0.0309	-0.0111	-0.0049	0.0050
90	0.9846	0.6261	0.0177	1.0426	0.0840	0.0836	-0.0431	-0.0059	0.0247
91	3.1613	2.0430	0.0133	0.8966	0.2376	0.1684	-0.0504	0.0979	-0.0515
92	-1.4801	-0.9433	0.0172	1.0219	-0.1247	-0.1233	0.0674	-0.0097	-0.0279
93	-0.9870	-0.6263	0.0134	1.0381	-0.0730	-0.0698	0.0257	0.0052	-0.0107
94	0.7011	0.4438	0.0107	1.0433	0.0462	0.0392	-0.0066	-0.0019	-0.0013
95	0.6432	0.4080	0.0151	1.0492	0.0505	0.0205	0.0246	-0.0243	-0.0164
96	1.5384	0.9813	0.0183	1.0201	0.1338	0.0174	0.0260	0.0682	-0.0442

Output Statistics									
Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS			
						Intercept	mavg	CACH	chcap
97	-2.4450	-1.5950	0.0466	0.9878	-0.3528	0.0754	-0.3094	0.1719	0.1675
98	-2.2935	-1.5875	0.1534	1.1134	-0.6759	-0.1623	0.4102	-0.2069	-0.5633
99	-0.3220	-0.2189	0.1430	1.2115	-0.0894	-0.0191	0.0508	-0.0261	-0.0735
100	0.3862	0.3458	0.5059	2.0953	0.3500	0.0104	-0.1083	-0.0383	0.3007
101	-1.2335	-0.7843	0.0150	1.0307	-0.0969	-0.0920	0.0295	0.0193	-0.0130
102	-0.6786	-0.4306	0.0153	1.0486	-0.0536	-0.0513	0.0174	0.0101	-0.0080
103	-0.3926	-0.2489	0.0150	1.0535	-0.0308	-0.0292	0.0094	0.0061	-0.0041
104	0.8764	0.5560	0.0137	1.0418	0.0654	0.0480	0.0067	-0.0298	-0.0034
105	1.3786	0.8772	0.0153	1.0247	0.1092	0.0763	-0.0262	0.0502	-0.0275
106	1.6735	1.0665	0.0145	1.0093	0.1296	0.1197	-0.0301	-0.0282	0.0077

Code Used in Project

```
* untransformed prp model;
proc reg data=sasuser.training3 simple;
model pRP= mavg cach chcap/vif stb;
output out=results r=residual;
plot r.*p.;
run;

*box-cox;
proc transreg data=sasuser.training3 TEST;
model boxcox(prp)=identity(mavg cach chcap);
run;

*prp_sqrt;
proc reg data=sasuser.training3 simple;
model pRP_sqrt= mavg cach chcap/vif stb;
output out=results r=residual;
plot r.*p.;
run;

*prp_log;
proc reg data=sasuser.training3 simple;
model pRP_log= mavg cach chcap/vif stb;
output out=results r=residual;
plot r.*p.;
run;

*model validation;
data sasuser.validation;
set sasuser.testing3;
m1 = 3.94828+(0.43193*mavg)+(0.31677*cach)+(0.31494*chcap);
devsq = (prp_sqrt-m1)**2;
proc means sum mean;
var devsq;
run;

*model validation (Standardized Estimate);
data sasuser.validation2;
set sasuser.testing3;
m1 = (0.58285*mavg)+(0.23262*cach)+(0.24132*chcap);
devsq = (prp_sqrt-m1)**2;
proc means sum mean;
var devsq;
run;

*lack of fit test;
proc rsreg data=sasuser.testing3;
```

```
model pRP_log = mavg cach chcap/ covar=3 lackfit ;
run;

*normal test;
proc rank data=results out=step3;
var residual;
ranks ranke;
run;
data new;
set step3;
ev = sqrt(2.50224)*Probit((ranke-.375)/(106+.25));
proc corr; * correlation test for normality on Sec 3.5;
var ev residual;
run;
proc univariate data=results normal plot;
var residual;
run;

*bp test;
data step2;
set results;
ressqr = residual**2;
proc reg data=step2;
model ressqr=mavg cach chcap;
run;

*weighted LS;
proc reg data=sasuser.training3;
model pRP_sqrt= mavg cach chcap/r;
output out=results1 r=residual;
data step2;
set results1;
absresid=abs(residual);
proc reg;
model absresid=mavg cach chcap/p; * fitted value;
output out=results2 p=yhat;
run;
data step3;
set results2;
wt=1/(yhat**2); ** weight;
proc reg;
model pRP_sqrt= mavg cach chcap/clb;
output out=results3 r=residual2;
weight wt; ** Wegithed LS regression;
plot r.*p.;
run;
quit;
```

```
*bp test;
data step4;
set results3;
ressqr = residual2**2;
proc reg data=step4;
model ressqr=mavg cach chcap;
run;
```

References

Ein-Dor, P. & Feldmesser, J. (1987). Attributes of the Performance of Central Processing Units: A Relative Performance Prediction Model. *Communications of the ACM, 30*(4), 308-317.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Linear Regression Models* (4th ed.). New York, NY: McGraw-Hill/Irwin.