RESEARCH ARTICLE OPEN ACCESS

Data Analysis Regression with CSR Files in Data Science

Kotadi Chinnaiah[1], Akash Saxena[2]
[1](Rescarch Scholor, Depaptment Cse, Sunrise University, Alwar, Rajasthan, India)
[2](Rescarch Superviser, Depaptment Cse, Sunrise University, Alwar, Rajasthan, India)

Abstract:

Statistics is an important part in big data because many statistical methods are used for big data analysis. The aim of statistics is to estimate population using the sample extracted from the population, so statistics is to analyze not the population but the sample. But in big data environment, we can get the big data set closed to the population by the advanced computing systems such as cloud computing and high-speed internet. According to the circumstances, we can analyze entire part of big data like the population of statistics. But we may be impossible to analyze the entire data because of its huge data volume. So, in this paper, we propose a new analytical methodology for big data analysis in regression problem for reducing the computing burden. We call this a divided regression analysis. To verify the performance of our divided regression model, we carry out experiment and simulation.

Keywords — Divided regression analysis, statistics, population, sample, big data analysis.

1. Introduction

We believe that some components of data science and business analytics have been around for a long time,but there are significant new questions and oppor-tunities created by the availability of big data and major advancements in machine intelligence.2 While the notion that analytical techniques can be used to make sense of and derive insights from data is as old as the field of statistics, and dates back to the 18th century, one obvious difference today is the rapid pace at which economic and social transactions are moving online, allowing for the digital capture of big data. The ability to understand the structure and con-tent of human discourse has considerably expanded the dimensionality of data sets available. As a result, the set of opportunities for inquiry has exploded exponentially with readily available large and com-plex data sets related to any type of phenomenon researchers want to study, ranging from deconstruct-ing the human genome, to understanding the pathol-ogy of Alzheimer's disease across millions of patients, to observing consumer response to different market-ing offers in large scale field experiments. And, easy (and relatively inexpensive) access to computational capacity and user-friendly analytical software have democratized the field of data science allowing many more scholars (and practitioners) to participate in the opportunities enabled by big data.

In some ways in could be argued that the nature of inquiry has also changed, turbocharged by machines becoming a lot smarter through better algorithms, and by information technologies that enable people and things to be inherently instrumented for observation and interaction that feeds the algorithms. Increasingly, data are collected not with the aim of solely testing a human-generated hypotheses or essential record-keeping, but to the extent that data torrents are cap-tured inexpensively, often for the possibility of testing hypotheses that have not yet been envisioned at the time of collection. When such data are gathered on a scale that observes every part of the joint distributions of the observed variables (behaviors, demographics, etc.), the computer becomes an active question ask-ing machine as opposed to a pure analytic servant. By initiating interesting questions and refining them without active human intervention, it becomes capa-ble of creating new knowledge and making discover-ies on its own (Dhar 2013). It can, for example, dis-cover automatically from a large swath of

healthcare system data that younger people in a specific region of the world are becoming increasingly diabetic and then conjecture and test whether the trend is due to specific habits, diet, specific types of drugs, and a

2 Arguably, for the first time in history, a machine passed the famous Turing test by defeating human champions at Jeopardy where topics are not known in advance and questions are posed in natural language of considerable complexity and nuance.

range of factors we may not have hypothesized as humans. This is powerful. As scientists, we have not seriously entertained the possibility of theory origi-nating in the computer, and as science-fiction-like as that may sound, we are in principle already there.

New challenging problems and inquiry also lead to research on better algorithms and systems. Since the torrent of data being generated is increasingly unstructured and coming from networks of people or devices, we are seeing the emergence of more pow-erful algorithms and better knowledge representation schemes for making sense of all of this heteroge-neous and fragmented information. Text and image processing capability are one frontier of research, with systems such as IBM's Watson being on the cutting edge in natural language processing, albeit with a long way to go in terms of their capability for ingest-ing and interpreting big data across the Internet.

Networks, such as those created by connections between individuals and/or products, further create significant and unique challenges at a fundamental level such as how we sample them or infer treat-ment effects. For example, in A/B testing, a "standard approach" for estimating the average treatment effect of a new feature or condition by exposing a sam-ple of the overall population to it, the treatment of individuals can spill over to neighboring individu-als along the structure of the underlying network. To address this type of "social interference," newer algorithms are required that support valid sampling and estimation of treatment effects (Ugander et al. 2013). This is but one example of how "relational" and "networked" data necessitates new development in algorithms. Developments may emerge not only from computer science but also from IS or other disciplines where researchers are "closer" to the problem being studied than pure methods researchers tend to be.

Finally, the Internet has fueled our ability to con-duct large scale experiments on social phenomena. As of this writing, Facebook researchers conducted a massive study to determine whether the mood of users could be manipulated and found that it could (Kramer et al. 2014). By conducting controlled exper-iments in large numbers of people such studies can extract the causal structure among variables. While the study raised important questions about privacy and the ethical implications of conducting experi-ments without informed consent, the broader point is that researchers now have a medium for theory development through massive experimentation in the social, health, urban, and other sciences.

2. A Divided Regression Analysis

In big data era, we can get huge data closed to population, and gain an opportunity to analyze the population. But traditional statistics needs much time for data analysis, and it is focused on the sample data analysis for inference of population. To settle this burden of big data analysis, we propose an approach to divide big data into some sub data sets as follow.

Figure 1. Divided Data Sets for BIG Data science

We divide the big data closed to population into some sub data sets with small size closed to sample. These divided data sets are proper to statistical analysis. In this paper, we select regression analysis as a statistical method for big data analysis because regression is a popular model for data analysis including big data analysis. The regression method has the computing burden because it is also one of statistical methods. To overcome the computing problem in big data regression, we propose a divided regression analysis, which splits whole data into n sub data sets. The whole data are regarded as the population in statistics, and the sub data set stands for a sample. In addition, we apply this data partitioning to estimate the parameters in regression model. In traditional regression analysis, the estimating process of the regression parameters is performed by the following figure.

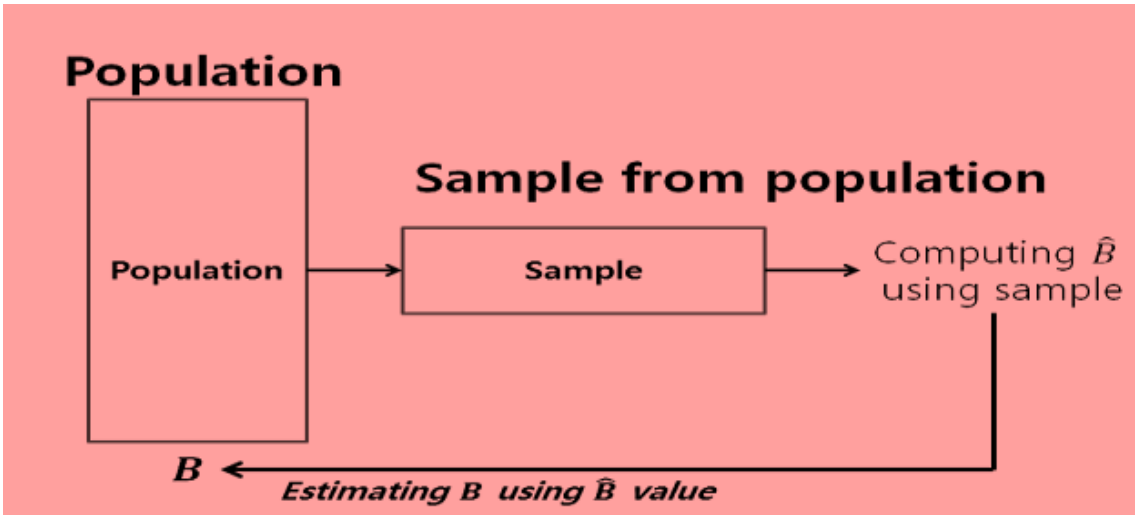

Figure 2. Traditional Regression Analysis

The multiple linear regression model is represented by the following [18].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

Where Y is dependent variable, and $_1, _2, \ldots,$ are independent variables. Also $_0, _1, \ldots,$ are regression parameters, and ε is an error of the model. To estimate the regression parameter vector B = ($_0, _1, _2, \ldots,$) for the population, we extract a sample data set from the population, and compute an estimate parameter vector$\hat{}$ = ($\hat{}_0, \hat{}_1, \hat{}_2, \ldots, \hat{}$) using the sample. Hence, we can estimate the regression function as follow.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k \quad (2)$$

This is very standard approach in statistical analysis. But, in the big data analysis, we have new problem, which is different to the traditional statistics. This is to use and analyze whole data according to circumstances. In this paper, we consider big data to the population of statistics, and separate the population into sub-populations as follow.

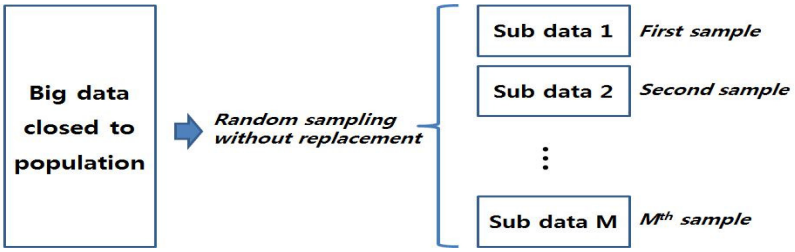

Figure 3. Dividing Big Data using Sampling Method

We use statistical sampling methods to dividing big data into sub samples. There are many sampling techniques from statistics such as simple random sampling, stratified sampling, systematic sampling, and cluster sampling [19]. These diverse sampling techniques were important to big data analysis [20, 21]. We should select sampling techniques carefully according to the aim of study and the characteristic of given data set. The main goal of all sampling methods is to get a representative sample from the population, and all sampling techniques should be based on the random sampling. In the random sampling, all elements of population have equal chances to be selected to sample. In our research, we use simple random sampling without replacement as a sampling method for dividing big data. Next figure shows the proposed method for entire regression analysis.

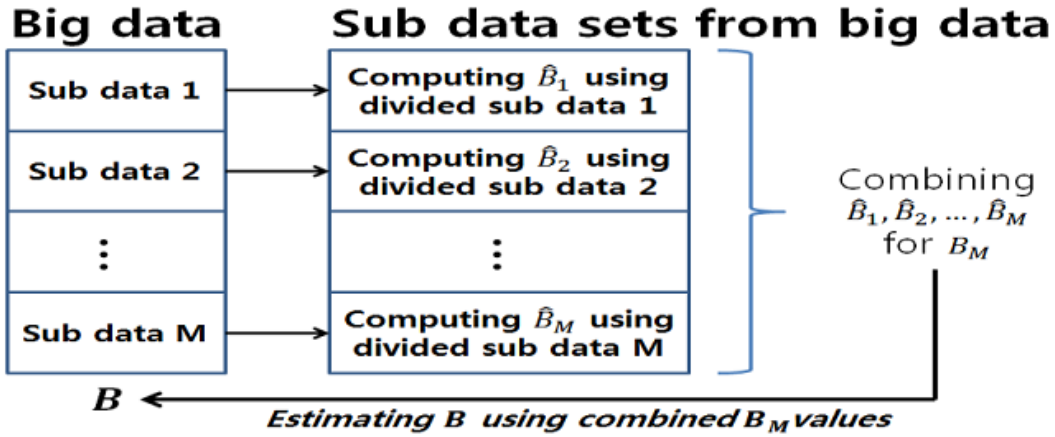

Figure 4. Divided Regression Analysis

Using all data with M sub sets, we compute the parameters which are ̂1, ̂2, … , ̂ respectively. We combine the M parameters to decide ̂ for estimating B as follow.

Where is a combine function for combining the results from sub-data 1 to sub-data M. We can consider diverse functions for combining the sample results. In our research, we use mean value for combining the estimated parameters as follow.

$$\hat{B}_C = f_C(\hat{B}_1, \hat{B}_2, \dots, \hat{B}_M) \quad (3)$$

$$\hat{B}_C = \frac{1}{M}\sum_{i=1}^{M} \hat{B}_i \quad (4)$$

So, using we can get same result of population analysis. To validate the statistical significance between and , where is the regression parameter of population, we can compute the significance interval of regression parameters. If the combined parameter of our model is included in the confidence interval, the performance of the estimated parameter by our work will be validated. Also, in this paper, we use mean squared error (MSE) for verifying regression result. This is defined as follow [18].

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (5)$$

The MSE measures the size and importance of forecasting error. Next we show the process of our study.

(Step1) Dividing big data

(1.1) Separating big data (population) into M sub data sets (samples) by simple random sampling without replacement

(1.2) Preparing sub data sets for regression analysis

(Step2) Performing multiple linear regression analysis (2.1)
Computing regression parameters for M sub data sets
(2.2) Averaging M regression parameters for estimating the regression parameters of all big data

(Step3) Evaluating model
(3.1) Comparing the regression results of M sub data sets with all big data
(3.3) Computing confidence interval of regression parameter
(3.3) Checking whether the averaged parameter is included in the confidence interval

To verify the performance our research, and to discover the potential of our model for real fields, we consider simulation and experiment in next section.

4. Experimental Results

To evaluate the proposed model, we performed simulation study, and made experiment using data set from UCI machine learning repository [9]. We also checked the confidence intervals of the regression results between divided and full data sets.

4.1 Simulation data

First we carried out an experiment using simulation data set for showing the performance of our research. Assume represents the value of dependent variable and represents the value of independent variable in the th trial. We used the following simple linear regression model.

The error terms , = 1,2, ... , are assumed to be independent random variables having a normal distribution with mean E( ) = 0 and constant variance Var( ) = ^{2}. In this experiment, we simulated simple linear regression model having the regression parameters $_0$ = 5 and $_1$ = 3. In this simulation, we set the variance is one, ^{2} = 1,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (6)$$

that is, the error terms , = 1, 2, ... , were all drawn from a normal distribution with mean 0 and variance 1 denoted by N(0,1). In this simulation, we determined the population size N was 1,000,000. Also, we divided the population into 10 sub-populations, so the size of each sub-population was 100,000. First of all, we got the following regression model using entire simulation data.

$$\hat{Y} = 4.998612 + 2.998295X \quad (6)$$

Next we estimated the regression parameters for 10 sub-populations, respectively. Table 1 shows the simulation result of our divided regression model.

Table 1. Simulation Data Set Result for Divided Regression Model

Data set	β_0	β_1	MSE
Sub.1	5.003847	2.996239	1.004636
Sub.2	4.994696	2.998008	0.996257
Sub.3	4.997074	2.996837	1.002991
Sub.4	4.998384	3.001986	1.000697
Sub.5	5.000347	2.999503	0.995672
Sub.6	5.002670	3.001851	0.998701
Sub.7	4.998995	3.002176	0.997655
Sub.8	4.997011	2.997172	1.002975
Sub.9	4.995441	2.996445	0.999974
Sub.10	4.997664	2.992761	0.996550
Mean	4.998613	2.998298	0.999611
Overall	4.998612	2.998295	0.999609

The size of each sub-population was 100,000, and we estimated β_0, β_1, and MES(mean squared error). In the table, "mean" of data set column is mean value of 10 sub-populations, and overall is the result by using the population. We knew the value between mean and overall was very similar, so we verified the performance of our study. Next we show 95% confidence intervals for regression parameters. The following figure shows the 95% confidence interval of intercept $_0$.

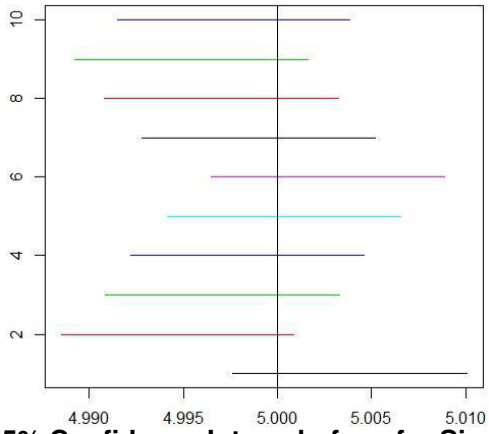

Figure 5. 95% Confidence Interval of for Simulation Set

We found all sub-populations included the parameter $_0$ in their confidence intervals. Next figure shows the 95% confidence interval of the regression parameter .

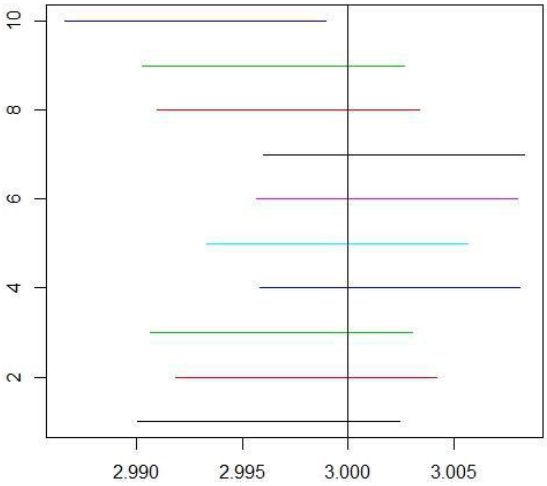

Figure 6. 95% Confidence Interval of for Simulation Set

Similar to the case of $_0$, all confidence intervals of 10 sub-populations also includes the parameter $_1$ of the population. We knew that all confidence intervals of β_0 and β_1 contained their parameters, so we confirmed the validity of our research. To verify the performance of our study clearly, we performed additional simulations by 100 iterations of the previous simulation as follow.

Table 2. Simulation Result of 100 Iterations

iteration	β_0	β_1	MSE	iteration	β_0	β_1	MSE
1	4.998613	2.998298	0.999610	51	5.000129	3.001609	0.998443
2			7	52			9
3	4.999802	2.997924	1.001704	53	4.999243	3.000334	1.001651
4			7	54			3
5	4.999922	3.000928	0.999121	55	4.999808	2.999247	0.999774
6	5.000292	3.000897	7	56			8
7		3.000229	1.002215	57	4.999433	2.998897	1.000790
8	4.999059		5	58			5
9		2.998958	0.998992	59	4.999343	3.000346	0.999230
10	5.001468		9	60	5.000246		5
11	5.002022	3.000479	1.000405	61		2.999866	0.998855
12		3.000919	1.000022	62	4.999529		4
13	4.999755		2	63		2.996379	1.001982
14	4.999283	2.998926	0.997984	64	4.998619		9
15			2	65		3.001135	1.000260
16	5.000375	3.000725	1.002252	66	5.000734		1
17			8	67		2.999514	1.000943
18	4.999411	3.000641	1.001533	68	4.999366		1
19		2.999207	0.997547	69		3.001206	1.000133
20	5.000593		5	70	4.999356		6
21		3.001179	1.002407	71		3.001746	0.999657
22	5.000292		8	72	5.000855		6
23		2.998396	1.000664	73		2.999797	0.998683
24	4.999598		5	74	4.999062		3
25	4.999249	3.002291	1.001462	75		2.999517	0.997797
26			3	76	5.000612		4
27	5.000283	2.999421	1.000786	77		3.000631	1.002215
28		3.000159	9	78	4.999097		9
29	4.998499		0.99775	79		3.000088	0.998546
30		3.000815	0.999341	80	5.000843		9
31	5.000165		1	81		3.000342	0.999627
32		2.999771	1.001093	82	4.999825		5
33	4.999861		2	83		3.000202	1.001381
34		3	0.999262	84	4.998616		2
35	4.998522	2.999518	0.998081	85		3.000796	0.999283
36			2	86	5.000076		8
37	4.999722	2.999931	0.999585	87		2.998757	1.000629
38			1	88	5.000459		5
39	5.000019	2.997546	0.999859	89		3.000205	1.003816
40			2	90	5.000275		8
41	5.000529	2.999379	1.000318	91		3.002274	0.997839
42			2	92	4.999822		4
43	5.001169	3.000759	1.001588	93		2.999616	0.999085
44			8	94	4.999693		6
45	4.998981	3.000865	1.001365	95		3.000315	1.000650
46			1	96	5.000283		5
47	4.999158	3.001395	0.998801	97		3.000141	0.998072
48			7	98	5.000058		4
49	5.000116	2.999317	0.997960	99		2.999823	1.001486

50	5.000519 4.999477 5.000743 5.000318 4.998239 4.998937 4.999285 4.999635.0005 4.999614.999372 5.001531 5.000014 4.999631 4.999764 5.000036 5.000094 4.999013 5.000097 4.998531 5.000957 4.998223 4.998936	4 2.998408 2.999901 2.999843 2.999275 2.999802 3.000453.000704 3.001771 3.001388 3.000977 2.999343 3.001104 2.999774 3.000128 2.999337 3.000088 3.000162 2.999652 3.000963.00044 2.999018 2.999668	2 0.9983631 0.9987005 1.0007689 0.9961941 1.0007328 1.0025224 1.0033361 0.9999068 0.9995341 1.0001628 1.0019517 1.0001503 1.0025164 1.0000653 0.9989867 0.9996047 1.0006604 0.998651 0.9989312 0.9967413 0.9976865 1.0017268 0.9981292	100	4.997727 5.000191 4.999135 5.000374 5.000146 5.000957 4.998635 4.999392 5.000161 5.000418 4.999465 5.000595 5.001411 5.000346 4.997907 5.000156 4.99988 4.999924.999392 4.999958 5.001767 4.999299 4.999767	1 2.999489 2.999379 3.000627 3.000478 2.999974 2.99882 2.998569 3.00007 2.998612 3.000752 3.001357 2.999916 3.000627 2.999729 3.000521 2.998586 2.999398 3.000343 3.000112 2.999362.99900 8 3.000042.99966 4 2.999545 2.99739	0.9986342 0.9990676 1.0001244 1.0009958 1.003555 0.9996925 1.0000858 1.0007833 1.0006379 1.0018623 0.9991689 0.9987259 1.0012172 0.9996501 0.9980508 0.9988638 0.9992054 1.0011135 1.0002026 0.9987681 1.0014414 1.0003697 0.9997982 0.9978502 1.0014147

We knew this result was similar to the previous simulation result. Next two figures show the 95% confidence intervals of $_0$ and $_1$ from the 100 iterative simulations.

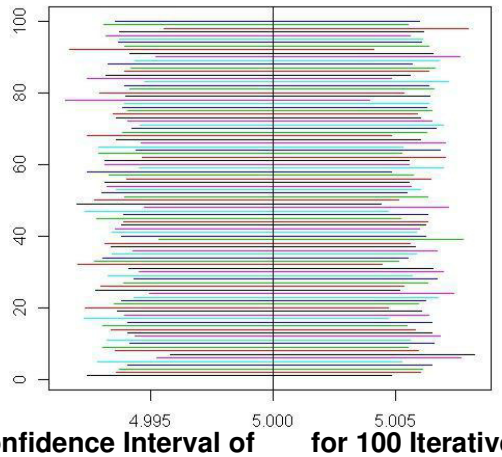

Figure 7. 95% Confidence Interval of for 100 Iterative Simulation Sets

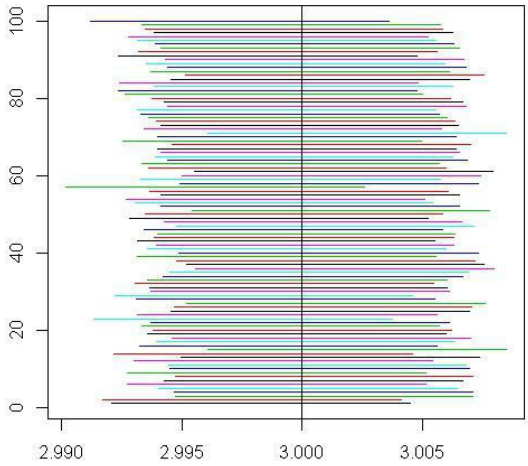

Figure 8. 95% Confidence Interval of for 100 Iterative Simulation Sets

In the two figures for 95% confidence interval, all 100 intervals included the regression parameters of 0 and 1. Therefore we showed the validity of our research. In the next section, we considered another experiment using example data set.

4.2. A case Study using the Bike Data

We made another experiment using example data from UCI machine learning repository [9]. The data set was Bike sharing data set including the time information of bike rental system, and the numbers of attributes and instances were 16 and 17,379 respectively. From the data set, we used three variables. The dependent variable was “cnt” and the independent variables were “temp” and “hum”. The “cnt" represents the number of total rental bikes, and the “temp” and “hum” show temperature and humidity respectively. So we modeled the multiple regression equation as follow.

$$\text{cnt} = \beta_0 + \beta_1 \text{temp} + \beta_2 \text{hum} + \varepsilon \quad (7)$$

In this model, we found the influence of temperature and humidity to bike rental. In addition, we divided the Bike sharing data into ten sub data sets for performing our divided regression analysis. The following table shows the regression result.

Table 3. Bike Data Set Result for Divided Regression Model

Data set	β_0	β_1	β_2	MSE
Sub.1	188.8797	369.6033	-295.8221	24535.77
Sub.2	159.6991	370.4749	-250.7093	22948.67
Sub.3	181.4024	373.1526	-281.6638	23589.74
Sub.4	193.5926	353.8599	-288.9963	24233.32
Sub.5	210.175	338.5803	-296.6555	24942.94
Sub.6	180.7998	387.5178	-288.3613	24826.61
Sub.7	194.0239	349.3915	-277.8862	26746.23
Sub.8	162.5819	360.5018	-246.7723	25335.49
Sub.9	184.3552	361.4922	-275.6302	24875.45
Sub.10	185.0865	354.8126	-279.4127	24575.58
Mean	184.0596	361.9387	-278.1910	24659.18
Overall	184.2446	361.8051	-278.3578	24643.72

We knew there are the slight differences between the regression parameters of sub-populations and overall population, but mean value of the regression parameters of sub-populations were similar to the parameter value of entire population. Next we computed the confidence intervals of the regression parameters in ten sub-populations. We show the 95% confidence interval of $_0$ is shown in the following figure.

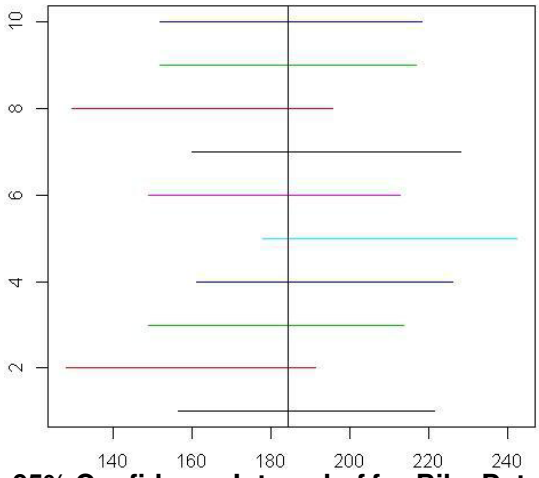

Figure 9. 95% Confidence Interval of for Bike Data Set

All confidence intervals of ten sub-populations contained the regression parameter $_0$ of the population. So, we confirmed the validity of our research. Next two figure show the 95% confidence interval of $_1$ and $_2$.

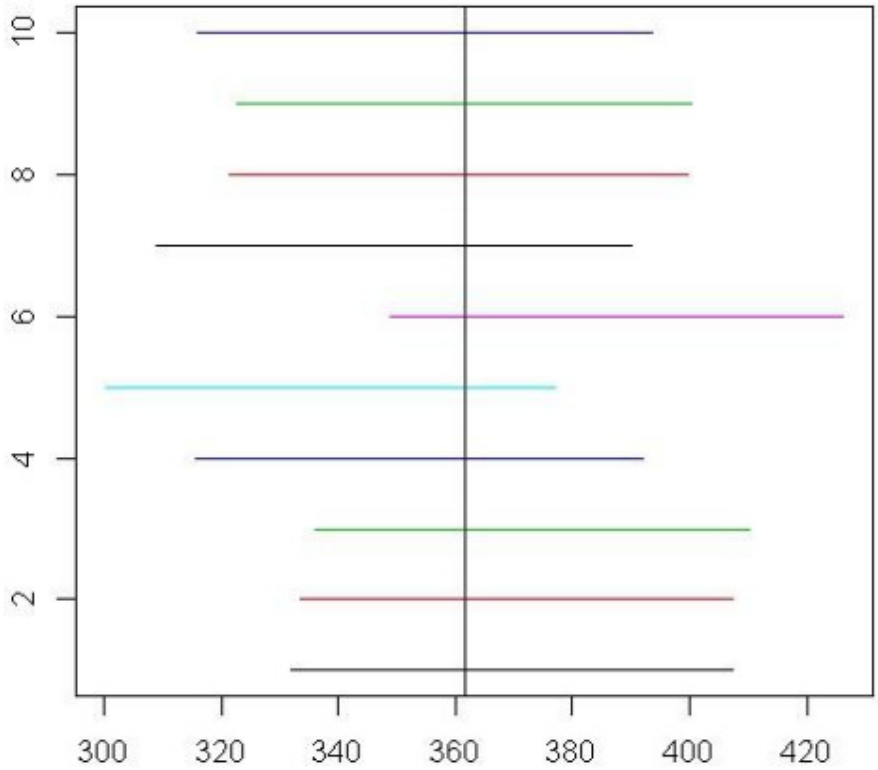

Figure 10. 95% Confidence Interval of for Bike Data Set

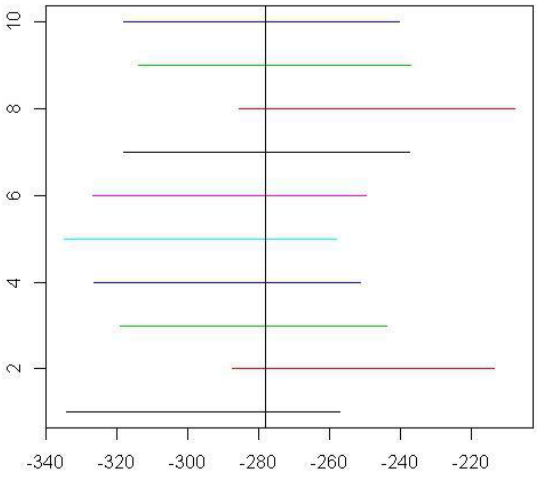

Figure 12. 95% Confidence Interval of for Bike Data Set

All confidence intervals of sub-populations include the regression parameters of the population for 1 and 2. Therefore, we can verify the performance of our regression approach.

5. Conclusions

In this paper, we proposed an approach to overcome the computing burden in big data analysis because most statistical methods were focused on small sample data. Also in big data analysis, we should analyze entire data which are considered as population in statistics, and this data set is so huge. Our research divided the big data closed to population into sub data set like sample for solving the computing cost in big data analysis. In addition, we applied this approach to regression problem in statistics. We

applied the divided method of big data to multiple regression analysis, and used simple random sampling for big data dividing. To verify the performance of our research, we used two data sets from simulation and UCI machine learning repository. In our experimental results, we knew that the regression parameters estimated by the big data were not different to the parameters by sub data sets. This research contributes to avoid the computing problem in many fields for big data analysis. We will apply our approach to more diverse methods in statistics such as factor analysis and clustering. More diverse methods of big data sampling are needed in our future works. We also will study more advanced combining methods for merging the results of sun data sets.

References

[1] W. Hu and N. Kaabouch, "Big Data Management, Technologies, and Applications", Information Science Reference, IGI Global, **(2014)**.

[2] S. Jun and D. Uhm, "A Predictive Model for Patent Registration Time Using Survival Analysis", Applied Mathematics & Information Sciences, vol. **7**, no. 5, **(2013)**, pp. 1819-1823.

[3] S. Jun, "A Technology Forecasting Method Using Text Mining and Visual Apriori Algorithm", Applied Mathematics & Information Sciences-An International Journal, vol. 8, no. (1L), **(2014)**, pp. 35-40.

[4] S. Jun, "Technology Forecasting by Big Data Learning", Proceedings of 2013 NIMS Hot Topic Workshops on Prediction using Fuzzy Theory, vol. 44, **(2013)**.

[5] J. Han, M. Kamber and J. Pei, "Data Mining: Concepts and Techniques", Third Edition, Morgan Kaufmann, **(2012)**.

[6] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute, **(2011)**.

[7] J. J. Berman, "Principle of Big Data", Morgan Kaufmann, **(2013)**.

[8] D. Vesset, B. Woo, H. D. Morris, R. L. Villars, G. Little, J. S. Bozman, L. Borovick, C. W. Olofson, S. Feldman, S. Conway, M. Eastwood and N. Yezhkova, "Worldwide Big Data Technology and Services 2012-2015 Forecast", IDC #233485, vol. 1, **(2013)**.

[9] UCI ML Repository, the UC Irvine Machine Learning Repository, http://archive.ics.uci.edu/ml, **(2014)**.

[10] S. M. Ross, "Introduction to Probability and Statistics for Engineers and Scientists", Elsevier, **(2012)**.

[11] B. Chun and S. Lee, "A Study on Big Data Processing Mechanism & Applicability", International Journal of Software Engineering and Its Applications, vol. 8, no. 8, **(2014)**, pp. 73-82.

[12] S. Ha, S. Lee and K. Lee, "Standardization Requirements Analysis on Big Data in Public Sector based on Potential Business Models", International Journal of Software Engineering and Its Applications, vol. 8, no. 11, **(2014)**, pp. 165-172.

[13] S. Jeon, B. Hong, J. Kwon, Y. Kwak and S. Song, "Redundant Data Removal Technique for Efficient Big Data Search Processing", International Journal of Software Engineering and Its Applications, vol. 7, no. 4, **(2014)**, pp. 427-436.

[14] J. Stanton, "An Introduction to Data Science", Ver. 3, Syracuse University, **(2013)**.

[15] S. M. Ross, "Introductory Statistics", McGraw-Hill, **(1996)**.

[16] P. Vincent, L. Badri and M. Badri, "Regression Testing of Object-Oriented Software: Towards a Hybrid Technique", International Journal of Software Engineering and Its Applications, vol. 7, no. 4, **(2013)**, pp. 227-240.

[17] V. Gupta, D. S. Chauhan and K. Dutta, "Regression Testing based Requirement Prioritization of Desktop Software Applications Approach", International Journal of Software Engineering and Its Applications, vol. 7, no. 6, **(2013)**, pp. 9-18.

[18] B. L. Bowerman, R. T. O'Connell and A. B. Koehler, "Forecasting, Time Series, and Regression", An Applied Approach, Brooks/Cole, **(2005)**.

[19] R. Scheaffer, W. Mendenhall III, R. L. Ott and K. G. Gerow, "Elementary Survey Sampling", 7th Edition, Duxbury, **(2011)**.

[20] M. Riondato, "Sampling-based Randomized Algorithms for Big Data Analytics", PhD dissertation in the Department of Computer Science at Brown University, **(2014)**.

[21] J. Lu and D. LiBias, "Correction in a Small Sample from Big Data", IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 11, **(2013)**, pp. 2658-2663.