ISSN 2348 – 8395

RESEARCH ARTICLE

ADVANCED DATA MINING PROCEDURES TO HANDLE BIG DATA SET: RESEARCH AND RESEARCHERS' DIRECTION

Dr. T. Saravanan
Assistant Professor, Department of Library & Information Science,
Annamalai University, Chidambaram, Tamil Nadu, India
sarlib@live.in

ABSTRACT

The aim of this paper is to brief the advanced research methods (Saravanan, T 2013a) for the LIS researchers as well as data mining procedures when the researchers face a big data set. The opted sample data (5114) comprises the research output of the country GERMANY, one of the developed nations in the field of Artificial Intelligence a subfield of Computer Science during the period 2002-2006. Germany research contributions are showing its status in scientific research. Sufficient records related to the Artificial Intelligence discipline have been captured from INSPEC a product of IEEE to carry out the research. The Sources Journal Articles and Conference Articles have been brought into the circle of the research by using the appropriate key terms. Study helps the researchers to incorporate new experimental designs in research and also keep their research on right track. Further, this study explores the progress of opting country research status in the field of AI. A well sophisticated data mining procedures have been applied to reach this target. This paper may pull the researchers into the boundaries of advanced level data mining, and also keep their research in the right direction.

Keywords : Data Mining; Artificial Intelligence; Germany; INSPEC; IEEE.

1. INTRODUCTION

Any research requires a clear introduction about the opted research field. Hence, researchers need to take care to render the relevant details concerning with their research area, and should walk within their research track (Saravanan, T 2010). The given introduction part may enable the researchers to realize the structure of this section, where relevant data need to be provided.

Encyclopedia Britannica (2010) defines AI as “Ability of a machine to perform tasks thought to require human intelligence. Typical applications include game playing, language translation, expert systems, and robotics. Although pseudo-intelligent machinery dates back to antiquity, the first glimmerings of true intelligence awaited the development of digital computers in the 1940s. AI, or at least the semblance of intelligence, has developed in parallel with computer processing power, which appears to be the main limiting factor. Early AI projects, such as playing chess and solving mathematical problems, are now seen as trivial compared to visual pattern recognition, complex decision making, and the use of natural language”. The centrality of knowledge and the production of codified knowledge in particular, have made it a subject of interest for governments and the private sector alike (Nonaka, I et al, 2003). As a result, there is an increased production of reports focusing on research and development (R&D) that aim at assessing the knowledge, ability and potential of international, national and regional levels (Hagstrom, W.O. ,1965). A competitive spirit has crept in, among the nations of the world striving for excellence in research output through publications which serve as an indicator for assessing the progress. This trend has brought nations into a Science & Technology race, namely the knowledge race.

2. RESEARCH DESIGN

A good research should satisfy its basic requirements. It should have a well structured research question. Hence, it is must for the researchers to formulate a few research questions in their research. In this study, the given research questions may be structured to conduct the further study. The possibilities are; Is there any difference in overall model? ; Is there a difference between the sources? ; Is there a difference between the years wise output? ; Do authors differ in their contributions? ; Do interactions differ between the sources and years? ; Do interactions differ between the sources and authors? ; Do interactions differ between the years and authors? ; Is there a difference between group comparisons means among years? ; Is there a difference between group comparisons means among authors? ; Is there a difference between group comparisons means of years within each source? ; Is there a difference between group comparisons means of sources within each year? ; Is there a difference between group comparisons means of sources within each author? etc. The objectives and hypotheses are playing major role in any research. So, keeping those aspects in the mind enough hypotheses need to be structured by the researchers.

The given hypotheses may be formulated to conduct the present research as complex research questions have been structured here. A big data set always helps the researchers to explore the research gates where they can play well.

3. OBJECTIVES OF THE STUDY

The objectives of this study are centred on appraise the research output performance related to Artificial Intelligence research from Germany during 2002 -2006 in multiple dimensions.

- To explore the Country Germany research output related to Artificial Intelligence from 2002-2006.
- To trace the significance between Journal articles and Conference articles output.
- To trace the significance between Single Author, Double Authors and Collaborative Authors contributions in Journal articles and Conference articles.
- To trace the significance between the interaction effects among Journal articles, Conference articles, Years and authors (Single Author, Double Authors and Collaborative Authors).
- To reveal the major findings and also trace the exact places, where further research would be possible.

4. HYPOTHESES

As you know that a decision maker always depends on the results of the hypothetical research to solve any kinds of problems. So, the hypotheses would be framed in all possible aspects that make the researchers to reach their targets. However, don't forget that these kinds of complex research hypotheses need to be tested with the right statistical tests as well as proved with more cautions. There would be a chance to frame a few hypotheses in this research as shown below.

H1 There would be no statistically significant differences between the means of all classes.

H2 There would be no statistically significant differences between the means of sources.

H3 There would be no statistically significant differences between the means of years wise output.

H4 There would be no statistically significant differences between the means of authors.

H5 There would be no statistically significant differences between the means of interactions of sources and years.

H6 There would be no statistically significant differences between the means of interactions of sources and authors.

H7 There would be no statistically significant differences between the means of interactions of years and authors.

H8 There would be no statistically significant differences between the groups comparisons means among the years.

H9 There would be no statistically significant differences between the groups comparisons means among the authors.

H10 There would be no statistically significant differences between the groups of means among the years within each source.

H11 There would be no statistically significant differences between the groups of means among the sources within each year.

H12 There would be no statistically significant differences between the groups of means among the sources within each author.

5. SAMPLE CHARACTERISTICS

Information regarding the samples may be discussed as shown below. It should be clear, short form and should not be presented in confusing ways. “First, you see your data for what they seem to be -Then you ask them for the truth-Are you what you seem to me?” (Forrest W, Y. ,2003).

This investigation takes into account all the entries of records of research publications on AI from the country of Germany only as found enumerated in the INSPEC database, a product from the IEEE. In Total 5114 records have been traced in this study by way of applying appropriate key terms. The Study encompasses the research output of Journals and Conference Articles only during the period 2002-2006, and didn't include other documents, Countries, etc. Table 1 shows the frequency distribution of a number of groups across the authors and the total number of observations for the opted years.

6. EXPERIMENTAL DESIGN

Researchers should design their experimental designs carefully as this part takes the research into the next level. The observations need to be analyzed first to capture the Gaussian distribution. If not, they need to be carefully treated and transformed using log or Box-Cox test, etc. This study comprises multi factor design, which is entirely a complex process rather than any other bivariate model. In this experimental design each factor and levels have got an opportunity to interact themselves that enable the researchers to move in depth to trace the significance for different pairs. In n-way ANOVA the analysis includes several comparisons of the means since there are several classifications.

In particular, comparisons are made for the classification themselves and for the interactions of each pair of classifications. As this study discusses about the complex data the statistical tools, namely Three-Way Anova with two way interaction, and two tailed T-Test may be applied to test the formulated hypotheses.The Researcher may conduct the said analysis as the objects are needed to be measured under three different classes, namely sources, years & authors in the level of significance at alpha 0.05 to test the formulated hypotheses of no effect (Saravanan, T 2013c). In addition to that Two-way interactions between the classes may be applied to test the hypotheses that no pair of sample means is different. Further a complete Post-Hoc comparison method, namely Tukey-HSD Test can be applied to trace the significance among the group means at alpha 0.05 levels in order to test the null hypotheses that group means are equal. Many current multivariate techniques and parametric estimations could not be realized without the widespread availability of sophisticated computer technology (Nunnally,J.C & Bernstein,I.H. 1994). This kind of experimental design has an advantage of being able to examine not only the "main" effects of variables hypothesized to affect the dependent variable, but also to be able to examine the interaction effects of these variables on the dependent variable. The n-way design has three grouping factors and one observed value. The model for the analysis can be realized from the given design (George, A. Marcoulides,2000 & Forrest W, Y. ,1992).+

OBS = A B C, A*B, A*C, B*C, A*B*C

Where A, B, and C are the main effects of the three factors. A*B, A*C and B*C are the two way interactions and A*B*C is the three way interaction. OBS is the observed variable. The matrix model could be traced from the given expressions.

Sources*authors*years
ie. 1*3=3 -> 3*5= 15(N)

Sources*authors*years (Level wise)

ie. (Journal=1*3, Conference=1*3 ->3+3=6)-> 6*1=6(N)

Sources*authors*years (Level wise)
ie. (Journal (SA) =1*5, Conference (SA) =1*5 ->5+5=10)-> 10(N) and so on.

Researchers should structure their table in a nice manner, where the possibilities are available to extract the required information. The samples may be structured as shown in the Table 1 that gives multilevel percent distributions across the factors (Rojer Stern et al., 2002). Any kind of huge data set can be handled like this way as it enables the researchers and the users too capture enough details regarding its distribution. The matrix design of Table 1 offers the flexibility to conduct an advanced analysis like n-way model to test the formulated hypotheses. There would be no need to split and increase the table numbers as they may consume more number of pages of this article. However, the Table 1 must be viewed from all directions.

6.1. Germany-Research Output Break Up: Sample Dispersion

Table. 1
Germany-Research Output Break Up: Sample Dispersion

Sources	Authors	2002	2003	2004	2005	2006	TOTAL
Jrl. Articles	SA	9	2	8	3	3	25
	Tot%	0.2	0	0.2	0.1	0.1	0.5
	Row%	36	8	32	12	12	
	Col%	0.9	0.4	0.5	0.2	0.4	
	DA	10	10	10	10	11	51
	Tot%	0.2	0.2	0.2	0.2	0.2	1
	Row%	19.6	19.6	19.6	19.6	21.6	
	Col%	1	1.9	0.7	0.7	1.5	
	CA	14	19	13	19	18	83
	Tot%	0.3	0.4	0.3	0.4	0.4	1.6
	Row%	16.9	22.9	15.7	22.9	21.7	
	Col%	1.5	3.7	0.9	1.4	2.5	
Conf.Articles							
	SA	214	102	268	204	78	866
	Tot%	4.2	2	5.2	4	1.5	16.9
	Row%	24.7	11.8	30.9	23.6	9	
	Col%	22.3	19.8	17.6	14.7	10.7	
	DA	318	163	515	438	219	1653
	Tot%	6.2	3.2	10.1	8.6	4.3	32.3
	Row%	19.2	9.9	31.2	26.5	13.2	
	Col%	33.1	31.6	33.9	31.5	30.1	
	CA	396	220	706	715	399	2436
	Tot%	7.7	4.3	13.8	14	7.8	47.6
	Row%	16.3	9	29	29.4	16.4	
	Col%	41.2	42.6	46.4	51.5	54.8	
TOTAL		961	516	1520	1389	728	5114
Tot%		18.8	10.1	29.7	27.2	14.2	100

An interpretation needs to be structured for better capturing of the Descriptive Statistics-Least-Squares Means for transformed data as shown below where statistic test results along with the discussions, graphs are explored from Table 2 on onwards:

Table. 2
Interactions: Sources*Years

GROUP		N	MEAN	VARIANCE	STD.DEV.
Jrl.	1	15	2.246	0.522	0.723
Conf.	2	15	6.015	0.538	0.733
2002	1	6	4.276	4.085	2.021
2003	2	6	3.694	3.950	1.988
2004	3	6	4.489	5.493	2.344
2005	4	6	4.307	6.044	2.458
2006	5	6	3.889	4.187	2.046
SA	1	10	3.414	4.604	2.146
DA	2	10	4.252	4.002	2.001
CA	3	10	4.726	3.899	1.975
TOTAL		30	4.131	4.185	2.046

CONF*2=6.33....CONF*3=5.6....CONF*4=6.6....CONF*5=6.5....CONF*6=5.8
JRL*2=2.5.....JRL*3=1.9.....JRL*4=2.3.....JRL*5=2.00.....JRL*6=2.00

Table 2 shows the least-squares means for the sources, years and authors. Of the sources of conference articles are traced with the highest mean (6.015) rather than the journal contributions (2.246). The Year 2004 has scored the greater mean value (4.489) and followed by 2005, 2002, 2006 and 2003 in that order. Collaborative authors are traced with the maximum mean (4.726) followed by double authors (4.252), and single authors (3.414). Mean distributions for the Table 2 are graphically visualized in 3D format towards the Screen Shots i-iii.

Table. 3

GROUP	Row	Col.	N	MEAN	VARIANCE	STD.DEV.
Cell	1	1	3	2.448	0.060	0.244
Cell	1	2	3	2.038	1.461	1.209
Cell	1	3	3	2.380	0.066	0.257
Cell	1	4	3	2.176	0.964	0.982
Cell	1	5	3	2.190	0.939	0.969
Cell	2	1	3	6.104	0.127	0.356
Cell	2	2	3	5.349	0.190	0.436
Cell	2	3	3	6.598	0.324	0.570
Cell	2	4	3	6.437	0.528	0.727
Cell	2	5	3	5.588	0.868	0.932
Jrl.	1		15	2.246	0.522	0.723
Conf.	2		15	6.015	0.538	0.733
2002	1		6	4.276	4.085	2.021
2003	2		6	3.694	3.950	1.988
2004	3		6	4.489	5.493	2.344
2005	4		6	4.307	6.044	2.458
2006	5		6	3.889	4.187	2.046
TOTAL			30	4.131	4.185	2.046

Tables 3 and 4 explore the descriptive statistics for each combination of factors in the model. Interactions between the sources and years (Screen Shot-iv) present the mean value 6.6, which is the highest value of conference articles for the year 2004. Least mean value (5.6) is found in the same source during the year 2003. Interactions between the sources and years present the mean value 2.5, which is the highest value of journal articles for the year 2002. Least mean value (1.9) is found in the year 2003. Interactions between the sources and authors present the mean value 6.5, which is the highest value for the collaborative authors, who have contributed more conference articles. Least mean value (5.8) is found for the single authors in the same source. Interactions between the sources and authors (Screen Shot-v) present the mean value 2.6, which is the highest value for the collaborative authors, who have contributed more in journal articles. Like as conference measures least mean value (1.48) is identified for the single authors. In both the sources collaborative authors are appeared with more domination. Interactions between the years and authors (Screen Shot-vi) present the mean value 5.8, which is the highest value for year 2005 and the collaborative authors. On the other hand year 2006 is found with the least mean value (4.5) for the collaborative authors. Concern with the double authors during the year 2005 highest mean value (5) is identified, whereas the year 2006 is falling in the least mean value 3.5. Single authors have received highest mean value 4.5 during the period 2004 and the same authors fall with the least value 3.0 during the year 2005. The mean distributions for various interactions have been graphically visualized in 3 dimensions from the Screen Shots a-j (See Annexure).

Table. 3A
Sources*Authors-
Descriptive Statistics

GROUP	Row	Col.	N	MEAN	VARIANCE	STD.DEV.
Cell	1	1	5	1.462	0.476	0.690
Cell	1	2	5	2.386	0.002	0.045
Cell	1	3	5	2.891	0.037	0.193
Cell	2	1	5	5.367	0.355	0.596
Cell	2	2	5	6.118	0.297	0.545
Cell	2	3	5	6.561	0.318	0.564
Jrl.	1		15	2.246	0.522	0.723
Conf.	2		15	6.015	0.538	0.733
SA	1		10	3.414	4.604	2.146
DA	2		10	4.252	4.002	2.001
CA	3		10	4.726	3.899	1.975
TOTAL			30	4.131	4.185	2.046

Interactions: Sources*Authors

CONF*CA=6.5...CONF*DA=6.3...CONF*SA=5.8
JRL*CA=2.6....JRL*DA=2.3....JRL*SA=1.4

David Nicholas et.al used more than 15 graphical displays in their research article in order to explore the data summaries (David nicholas et.al. 2002). "For in pictures we find insight While in numbers we find strength" (Forrest W,Y. 2003). For complex research designs, graphical displays far surpass textual descriptions in their ability to express the full patter of the data. For an Instance, a typical journal article has reported an experiment results something like this:

Results of 2x3x4 factorial ANOVA showed a significant main effect of instructions. F(1,66)= 4.15,p=.046, and a significant main effect of message type, F(2,66)=1056,p=.0001. The main effect of time, the repeated measures factor, had also been significant, F(3,198)=7.13, p=.0001. The interactions of time by instructions was significant, F(3,198)=4.80,p=.003, as was the interaction of time by message type F(6,198)=4.61,p=.0002.

This passage is virtually incomprehensible, but enough graphical displays may nicely depict the patterns in these data (Mark hallahan & Robert Rosenthal, 2000).

Table. 4
Years*Authors -
Descriptive Statistics

GROUP	Row	Col.	N	MEAN	VARIANCE	STD.DEV.
Cell	1	1	2	3.987	6.001	2.450
Cell	1	2	2	4.268	7.238	2.690
Cell	1	3	2	4.573	6.845	2.616
Cell	2	1	2	2.792	8.766	2.961
Cell	2	2	2	3.889	4.637	2.153
Cell	2	3	2	4.400	3.650	1.911
Cell	3	1	2	4.053	7.389	2.718
Cell	3	2	2	4.546	9.504	3.083
Cell	3	3	2	4.868	9.896	3.146
Cell	4	1	2	3.389	10.361	3.219
Cell	4	2	2	4.452	8.706	2.951
Cell	4	3	2	5.078	8.236	2.870
Cell	5	1	2	2.850	6.037	2.457
Cell	5	2	2	4.106	5.375	2.318
Cell	5	3	2	4.711	5.918	2.433
2002	1		6	4.276	4.085	2.021
2003	2		6	3.694	3.950	1.988
2004	3		6	4.489	5.493	2.344
2005	4		6	4.307	6.044	2.458
2006	5		6	3.889	4.187	2.046
SA	1		10	3.414	4.604	2.146
DA	2		10	4.252	4.002	2.001
CA	3		10	4.726	3.899	1.975
TOTAL			30	4.131	4.185	2.046

Interactions: Years*Authors

2*CA=4.60......2*DA=4.50......2*SA=4.00
3*CA=4.57......3*DA=4.00......3*SA=3.50
4*CA=5.00......4*DA=4.80......4*SA=4.50
5*CA=5.80......5*DA=5.00......5*SA=3.00
6*CA=4.50......6*DA=3.50......6*SA=3.50

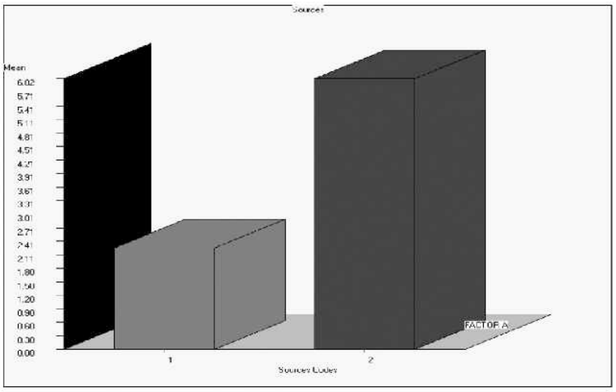

Figure :1. Screen Shot-i : 3D - Mean Distribution for Source

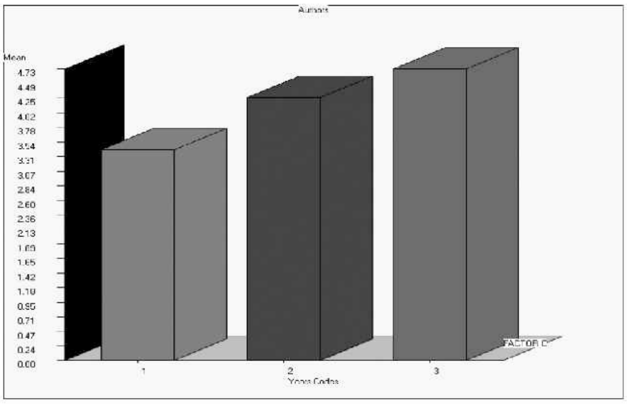

Figure :2. Screen Shot-ii : 3D - Mean Distribution for Authors

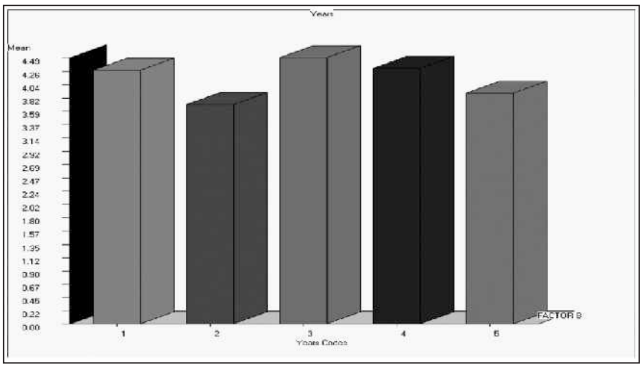

Figure :3. Screen Shot-iii : 3D - Mean Distribution for Years

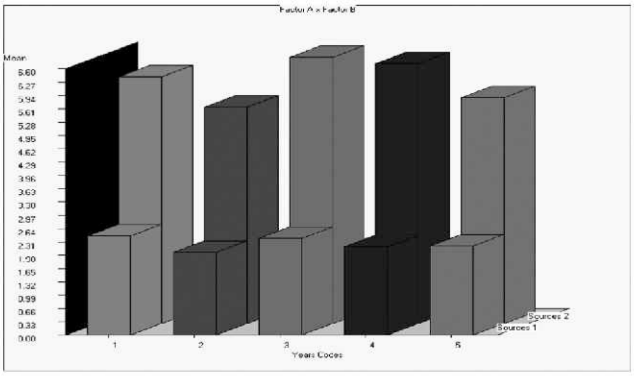

*Figure: 4. Screen Shot-iv : 3D - Source * Years*

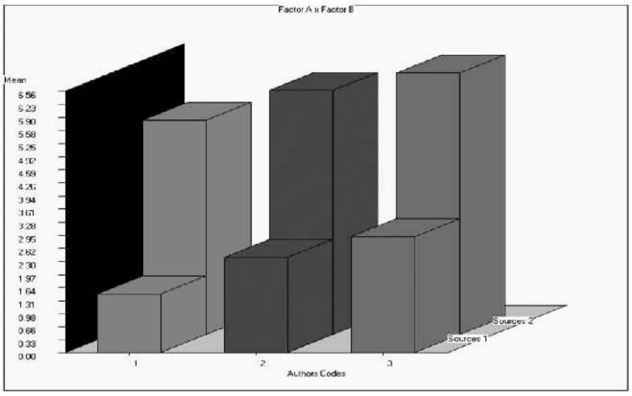

Figure :5. Screen Shot-i : 3D - Sources Authors*

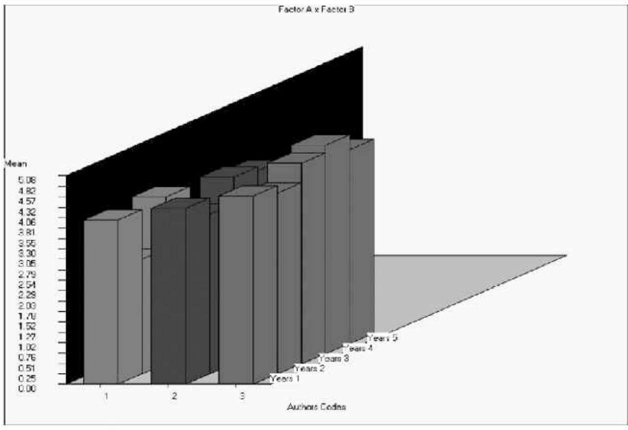

Figure :6. Screen Shot-i : 3D - Years Authors*

Table 5
Parameter Estimates (Least Squares)
With Two-Tailed T-Tests

Term	Estimate	Std. Error	t-Ratio	P-Value
Constant	4.10	0.11	37.51	<.0001
Sources[C-J]	1.97	0.11	17.99	<.0001
Years[2-6]	0.23	0.22	1.07	0.3169
Years[3-6]	0.23	0.22	1.07	0.3169
Years[4-6]	-0.10	0.22	-0.46	0.6595
Years[5-6]	-0.10	0.22	-0.46	0.6595
Authors[C-S]	0.10	0.15	0.65	0.5358
Authors[D-S]	0.20	0.15	1.29	0.2318
Sources[C-J]*Years[2-6]	0.03	0.22	0.15	0.8826
Sources[C-J]*Years[3-6]	0.03	0.22	0.15	0.8826
Sources[C-J]*Years[4-6]	0.03	0.22	0.15	0.8826
Sources[C-J]*Years[5-6]	0.03	0.22	0.15	0.8826
Sources[C-J]*Authors[C-S]	0.03	0.15	0.22	0.8346
Sources[C-J]*Authors[D-S]	0.13	0.15	0.86	0.4134
Years[2-6]*Authors[C-S]	0.07	0.31	0.22	0.8346
Years[2-6]*Authors[D-S]	-0.03	0.31	-0.11	0.9168
Years[3-6]*Authors[C-S]	0.57	0.31	1.83	0.1041
Years[3-6]*Authors[D-S]	-0.53	0.31	-1.73	0.1228
Years[4-6]*Authors[C-S]	-1.10	0.31	-3.56	0.0074
Years[4-6]*Authors[D-S]	0.30	0.31	0.97	0.3602
Years[5-6]*Authors[C-S]	-0.10	0.31	-0.32	0.7546
Years[5-6]*Authors[D-S]	0.80	0.31	2.59	0.0322

The two-tailed t-test is applied to estimate the observed classes, namely sources, years, authors, which would not indicate possible significance except the classes constant (p<.0001); conference*journal sources (p<.0001); the years 2004*2006 and collaborative*single authors (p=.0074); the years 2005*2006 and double*single authors (p=.0322); The estimations between other classes are not identified as significant. Along with the T-test the complete post hoc tests also were applied to estimate the interactions for testing the significance between the means of classes in brief. (In "sources" letter 'C' refers conference articles whereas in "authors" it denotes the collaborative authors). Anova Assumptions (Modified version of Levene's test-(as described by Glantz et al. in "Primer of applied regression & analysis of variance", ch. 7)): Sources (2 Groups)-Equal Variance: P=0.57978/ Year (5 Groups)-Equal Variance: P=0.63985/Authors (3 Groups) -Equal Variance: P= 0.88326.

Summary of fit

R Squared (Total Effect Strength)	: 0.98
Adjusted R Squared	: 0.92
Sigma hat (RMS error)	: 0.60
Number of cases	: 30
Degrees of freedom	: 8

Table 6
Analysis Of Variance: Model Test

Source	Sum-of-Squares	df	Mean-Square	F-Ratio	P-Value	R-Square
Model	127.83	21	6.09	16.99	0.0002	0.98
Error	2.87	8	0.36			
Total	130.70	29				

Table 7
Analysis Of Variance: Effects Tests

Source	Sum-of-Squares	df	Mean-Square	F-Ratio	P-Value	R-Square
Sources	116.03	1	116.03	323.81	<.0001	0.89
Years	1.20	4	0.30	0.84	0.5384	0.01
Authors	1.40	2	0.70	1.95	0.2038	0.01
Sources*Years	0.13	4	0.03	0.09	0.9819	0.00
Sources*Authors	0.47	2	0.23	0.65	0.5470	0.00
Years*Authors	8.60	8	1.07	3.00	0.0706	0.07
All Sources	127.83	21	6.09	16.99	0.0002	0.98
Error	2.87	8	0.36			
Total	130.70	29				

Researchers should conduct the Tests for full model and each model to prove the hypotheses as given below. Enough graphs need to be generated to explore the determinations for better understanding.

6.2 Full Model Test

Each term in the model is tested for its ability to account for variation in the dependent variable. For these data an observed F-Test having 21 and 8 degrees of freedom. This test resulted in F=16.99, p=0.0002. This means that when the null hypothesis of no effect for any source in the model is true, the probability is p=0.0002 of obtaining a sample of data with an effect that is as extreme or is more extreme. Thus, of many repetitions of this sampling design, the probability is p=0.0002 that we would be wrong when we reject the null hypothesis of no effect in favour of the alternative hypothesis that at least one pair of treatment means is different. Results produced from t-test (p<.0002) also confirms the same and based on these enough statistical evidences there would be a possible significance identified at the significance level of 0.05, and hence author was not able to support the formulated hypothesis H1 in favour of the alternative. The significance could be visually observed with the help of the partial regression plot (Screen Shot-1.1). The 95% confidence boundaries also support the same inferences. The residuals plot ((Screen Shot-1.2) shows the observations and their dispersions of the full model.

6.3 Each Model Test

For sources an observed F-Test having 1 and 8 degrees of freedom. This test resulted in F=323.81, p=<.0001. This means that when the null hypothesis of no effect for sources is true, then the probability is p=<.0001 of obtaining a sample of data with an effect that is as extreme or is more extreme. Thus, over many repetitions of this sampling design, the probability is p=<.0001 that we would be wrong when we reject the null hypothesis of no effect in favour of the alternative hypothesis that at least one pair of sources means is different. Results produced from t-test(p<.0001)also confirms the same and based on these enough statistical evidences there would be a possible significance identified at the significance level of 0.05, and hence we fail to claim support to the formulated hypothesis H2 in favour of the alternative. The significance could be visually observed with the help of the partial regression plot (Screen Shot-2.1). The 95% confidence boundaries also support the same inferences. The residuals plot (Screen Shot-2.2) shows the observations and their dispersions of the source model. The LS means for the sources could be clearly observed towards the profile plot (Screen Shot-2.3).

For years the F-Test explores 4 and 8 degrees of freedom. This test resulted in F=0.84, p=0.5384. This means that when the null hypothesis of no effect for the year is true, then the probability is p=0. 5384 of obtaining a sample of data with an effect that is as extreme or is more extreme. Thus, over many repetitions of this sampling design, the probability is p=0. 5384 that we would be wrong when we reject the null hypothesis of no effect in favour of the alternative hypothesis that at least one pair of years'means is different. With the help of enough statistical evidences possible significance is not identified, which would be the reason to claim support to the formulated hypothesis H3 against the alternative at the significance level of 0.05. The significance could be visually observed with the help of the partial regression plot (Screen Shot-3.1). The 95% confidence boundaries also support the same inferences. The residuals plot (Screen Shot-3.2) shows the observations and their dispersions of the year's model. The LS means for the years could be clearly observed towards the profile plot (Screen Shot-3.3).

For authors an observed F-Test indicates 2 and 8 degrees of freedom. This test resulted in F=1.95, p=0.2038. This means that when the null hypothesis of no effect for an author is true, then the probability is p=0. 2038 of obtaining a sample of data with an effect that is as extreme or is more extreme. Thus, over many repetitions of this sampling design, the probability is p=0. 2038 that we would be wrong when we reject the null hypothesis of no effect in favour of the alternative hypothesis that at least one pair of authors' means is different. With the help of enough statistical evidences possible significance is no identified, which would be the reason to claim support to the formulated hypothesis H4 against the alternative at the significance level of 0.05. The significance could be visually observed with the help of the partial regression plot (Screen Shot-4.1). The 95% confidence boundaries also support the same inferences. The residuals plot (Screen Shot-4.2) shows the observations and their dispersions of the authors' model. The LS means for the author could be clearly observed towards the profile plot (Screen Shot-4.3).

For sources*years an observed F-Test having 4 and 8 degrees of freedom. This test resulted in F=0.09, p=0.9819. This means that when the null hypothesis of no effect for sources*years is true, then the probability is p=0. 9819 of obtaining a sample of data with an effect that is as extreme or is more extreme. Thus, over many repetitions of this sampling design, the probability is p=0. 9819 that we would be wrong when we reject the null hypothesis of no effect in favour of the alternative hypothesis that at least one pair of sources*years means is different. With the help of enough statistical evidences possible significance is not identified, which would be the reason to claim support to the formulated hypothesis H5 against the alternative at the significance level of 0.05. The significance could be visually observed with the help of the partial regression plot (Screen Shot-5.1). The 95% confidence boundaries also support the same inferences. The residuals plot (Screen Shot-5.2) shows the observations and their dispersions of the sources*years model. The LS means for the sources*years could be clearly observed towards the profile plot (Screen Shot-5.3).

For sources*authors the F-Test shows 2 and 8 degrees of freedom. This test resulted in F=0.65, p=0.5470. This means that when the null hypothesis of no effect for sources* authors is true, then the probability is p=0. 5470 of obtaining a sample of data with an effect that is as extreme or is more extreme. Thus, over many repetitions of this sampling design, the probability is p=0. 5470 that we would be wrong when we reject the null hypothesis of no effect in favour of the alternative hypothesis that at least one pair of sources*authors means is different. With the help of enough statistical evidences possible significance is not identified, which would be the reason to claim support to the formulated hypothesis H6 against the alternative at the significance level of 0.05. The significance level could be visually observed with the help of the partial regression plot (Screen Shot-6.1). The 95% confidence boundaries also support the same inferences. The residuals plot (Screen Shot-6.2) shows the observations and their dispersions of the sources*authors model. The LS means for the sources*authors could be clearly observed towards the profile plot (Screen Shot-6.3), where the lines are parallel.

For years*authors the F-Test shows 8 and 8 degrees of freedom. This test resulted in F=3, p=0.0706. This means that when the null hypothesis of no effect for years*authors is true, then the probability is p=0.0706 of obtaining a sample of data with an effect that is as extreme or is more extreme. Thus, over many repetitions of this sampling design, the probability is p=0.0706 that we would be wrong when we accept the null hypothesis of no effect against the alternative hypothesis that at least one pair of years*authors means is different. With the help of enough statistical evidences it could be inferred that there would not be a possible significance identified, which would be the reason to claim support to the formulated hypothesis H7 against the alternative at the significance level of 0.05.The class wise interactions between the years*authors also have been traced using the two tailed t-test in order to measure the significance. Years 2004*2006 and Collaborative*Single authors (p=0.0074) ; Years 2005*2006 and double*single authors (p=0.0322) identify with the significance. The generated statistical evidences also support the same inferences. The significance could be visually observed with the help of the partial regression plot (Screen Shot-7.1). The 95% confidence boundaries also support the same inferences. The residuals plot (Screen Shot-7.2) shows the observations and their dispersions of the years*authors model. The LS means for the years*authors could be clearly observed towards the profile plot (Screen Shot-7.3), where the lines are roughly parallel.

Strength of Relationship-Full Model

Based on this sample of data, it could be estimated from the outcome that 98% of the variability in sample is related to variability in all sources. This estimate would vary over repeated samples of data. It has been clearly visualized towards the Screen Shot-1 for better capturing of sample dispersion

6.4 Strength of relationship-each model

Based on this sample of data, it could be estimated that 89% of the variability in observation is related to variability in sources (Screen Shot-2), 0.1% of the variability in observation is related to variability in years (Screen Shot-3), and authors (Screen Shot-4), Null variability in observation is related to variability in sources*years (Screen Shot-5) and sources*authors (Screen Shot-6), 0.07% of the variability in observation is related to variability in years*authors (Screen Shot-7). These estimates would vary over repeated samples of data.

6.5 Comparisons Among Years

Table 8
Comparisons among Years
(Tukey HSD Test (Differences between Means alpha selected = 0.05))

Groups	Difference	Statistic	Probability
1 - 2	0.582	q = 4.674	0.0614
1 - 3	-0.213	q = 1.710	0.7473
1 - 4	-0.030	q = 0.244	0.9998
1 - 5	0.387	q = 3.104	0.2716
2 - 3	-0.796	q = 6.384	0.0125
2 - 4	-0.613	q = 4.918	0.0485
2 - 5	-0.196	q = 1.569	0.7977
3 - 4	0.183	q = 1.466	0.8323
3 - 5	0.600	q = 4.815	0.0536
4 - 5	0.417	q = 3.349	0.2178

The comparisons among years (2002-2006) have been analyzed and explored here to test the hypothesis H8. The given analyses have been made based on years comparisons. (1=2002 / 2=2003 and so on). The Tukey HSD Test was adopted to measure the significance between the group means at alpha level 0.05 and observed that there would be possible significance statistically identified in the groups 2-3 and 2-4 while the rest of the groups are not identified as significant. Based on these enough evidences author can claim support to the formulated hypothesis H8 for the groups, which are not identified as significant and in contrary the same hypothesis would not be supported for the groups that are identified as significant.

6.6 Comparisons Among Authors

Table 9
Comparisons among Authors (Tukey HSD Test (Differences between Means alpha selected = 0.05))

Groups	Difference	Statistic	Probability
1 - 2	-0.838	q = 8.681	0.0008
1 - 3	-1.312	q = 13.587	0.0001
2 - 3	-0.474	q = 4.907	0.0206

The comparisons among authors have been analyzed and explored here to test the hypothesis H9. The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would be a possible significance statistically identified among authors and these evidences would not help to retain the formulated hypothesis H9 against the alternative

6.7 Journal Articles Comparisons

Table 10
Journal Articles Comparisons
Comparisons Among Years Within Each Source (Tukey HSD Test (Differences between Means alpha selected = 0.05)

Groups	Difference	Statistic	Probability
1 - 2	0.410	q = 0.406	0.9979
1 - 3	0.068	q = 0.067	1.0000
1 - 4	0.272	q = 0.269	0.9996
1 - 5	0.258	q = 0.255	0.9997
2 - 3	-0.342	q = 0.339	0.9990
2 - 4	-0.138	q = 0.137	1.0000
2 - 5	-0.152	q = 0.151	1.0000
3 - 4	0.204	q = 0.202	0.9999
3 - 5	0.190	q = 0.188	0.9999
4 - 5	-0.014	q = 0.014	1.0000

The comparisons among years (2002-2006) within each source (Jrl. Articles) is analyzed and explored here to test the hypothesis H10. The given analyses have been made based on row 1 comparisons.

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would be no significance statistically identified among years within each source (Journal articles) and based on these enough evidences the formulated hypothesis H10 is retained against the alternative.

6.8 Conference Articles Comparisons

Table 10 a
Conference Articles Comparisons

Groups	Difference	Statistic	Probability
1 - 2	0.755	q = 0.747	0.9794
1 - 3	-0.494	q = 0.489	0.9957
1 - 4	-0.333	q = 0.329	0.9991
1 - 5	0.516	q = 0.511	0.9949
2 - 3	-1.249	q = 1.235	0.8935
2 - 4	-1.088	q = 1.076	0.9298
2 - 5	-0.239	q = 0.236	0.9997
3 - 4	0.161	q = 0.159	0.9999
3 - 5	1.010	q = 0.999	0.9446
4 - 5	0.849	q = 0.840	0.9689

The comparisons among years (2002-2006) within each source (Conf. Articles) have been analyzed and explored here to test the hypothesis H10. The given analyses have been made based on row 2 comparisons.

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would be no significance statistically identified among the groups, which would be the reason for retaining the formulated hypothesis H10 against the alternative.

6.8 Comparisons among sources within each year

Table 11
Comparisons - 2002
Comparisons among Sources within each Year
(Tukey HSD Test (Differences between Means alpha selected = 0.05)

Groups	Difference	Statistic	Probability
1 - 2	-3.656	q = 3.617	0.0628

The comparisons between sources within each year have been analyzed and explored here to test the hypothesis *H11*. The given analyses have been made based on 2002 comparisons.

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would not be a possible significance statistically identified among sources within each year (2002) and based on these enough evidences the formulated hypothesis *H11* would be retained against the alternative.

6.10. Comparisons-2003

Table 11 a
Comparisons-2003

Groups	Difference	Statistic	Probability
1 - 2	-3.311	q = 3.276	0.0815

The comparisons between sources within each year have been analyzed and explored here to test the hypothesis *H11*. The given analyses have been made based on 2003 comparisons.

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would not be a possible significance statistically identified among sources within each year (2003) and based on these enough evidences the formulated hypothesis *H11* would be retained against the alternative.

6.11. Comparisons - 2004

Table 11 b
Comparisons-2004

Groups	Difference	Statistic	Probability
1 - 2	-4.218	q = 4.173	0.0419

The comparisons between sources within each year have been analyzed and explored here to test the hypothesis *H11*. The given analyses have been made based on 2004 comparisons.

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would be possible significance statistically identified among sources within each year (2004) and based on these enough evidences the formulated hypothesis *H11* would not be retained in favour of the alternative.

6.12 Comparisons-2005

Table 11 c
Comparisons-2005

Groups	Difference	Statistic	Probability
1 - 2	-4.261	q = 4.216	0.0407

The comparisons between sources within each year have been analyzed and explored here to test the hypothesis *H11*. The given analyses have been made based on 2005 comparisons.

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would be possible significance statistically identified among sources within each year (2005) and based on these enough evidences the formulated hypothesis *H11* would not be retained in favour of the alternative.

6.13 Comparisons-2006

Table 11 d
Comparisons-2006

Groups	Difference	Statistic	Probability
1 - 2	-3.398	q = 3.362	0.0763

The comparisons between sources within each year have been analyzed and explored here to test the hypothesis *H11*. The given analyses have been made based on 2006 comparisons.

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would not be a possible significance statistically identified among sources within each year (2006) and based on these enough evidences the formulated hypothesis *H11* would be retained against the alternative.

Comparisons between sources within each author (tukey hsd test (differences between means alpha selected = 0.05))

6.14 Comparisons-Single Author

Table 12
Comparisons-Single Author

Groups	Difference	Statistic	Probability
1 - 2	-3.904	q = 5.034	0.0707

The comparisons between sources within each author have been analyzed and explored here to test the hypothesis H12. The given analyses have been made based on single author comparisons.

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would not be possible significance statistically identified among sources within each author (SA) and based on these enough evidences the formulated hypothesis H12 would be retained against the alternative.

6.15 Comparisons-Double Authors

The comparisons between sources within each author have been analyzed and explored here to test the hypothesis H12. The given analyses have been made based on double authors comparisons.

Table 12
Comparisons-Double Authors

Groups	Difference	Statistic	Probability
1 - 2	-3.732	q = 4.812	0.0766

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would not be possible significance statistically identified among sources within each author (DA) and based on these enough evidences the formulated hypothesis H12 would be retained against the alternative.

6.16 Comparisons-Collaborative Authors

The comparisons between sources within each author have been analyzed and explored here to test the hypothesis H12. The given analyses have been made based on collaborative authors comparisons.

Table 12 b
Comparisons-Collaborative Authors

Groups	Difference	Statistic	Probability
1 - 2	-3.670	q = 4.731	0.0790

The Tukey HSD Test was adopted to measure the significance among the group means at alpha level 0.05 and observed that there would not be possible significance statistically identified among sources within each author (CA) and based on these enough evidences the formulated hypothesis H12 would be retained against the alternative.

6.17 Determinations

The conclusion part needs to be structured in a brief manner as it explores the findings of the study by which the decision makers may find the suitable solution that they actually looking for to solve the problems. Further, the next stage of the research also could be traced towards the hypothetical results.

It could be inferred from the statistical results that the highest productivity has been traced in conference articles with the domination of collaborative authors during the year 2005 rather than the productivity of the journal articles with the rest of the authors. Screen Shot-2 reveals that the conference articles output observations appear to have the same variability of the journal articles output. Screen Shot-3 reveals that the observations for all the years appear to have same variability. Screen Shot-4 reveals that the observations for the collaborative and double authors appear to have greater variability. On the other hand, single authors seem to have the smallest variability. Screen Shot-5 reveals that the observations of the conference articles output during the year 2004, 2005 appear to have greater variability than those from the other sources and years. The observations for the journal article output during the years 2004-2006 appear to have greater variability than those from the other sources and years. Screen Shot-6 reveals that the observations for the journal articles and conference articles output along with the collaborative authors appear to have greater variability than those from the other sources and authors. Screen Shot-7 reveals that the observations for the collaborative authors' contributions during the year 2002 appear to have greater variability than those from the other authors and years. The observations for the double authors' contributions during the year 2004 appear to have greater variability than those from the other authors and years. The observations for the single authors' contributions during the years 2004, 2006 appear to have greater variability than those from the other authors and years (Saravanan, T 2013b). As many of the distributions found in these figures depict the equal variances, and hence the assumptions of homogeneity of variance would not be violated.

R2 is the strength of the model fit, which shows that about 98 per cent of variation is explained in all sources by the overall model. Among the two rows the highest mean value 6.015 is identified for the conference articles output. whereas the journal articles output is just 2.246. R2 shows that approximately 89 per cent of variation for sources. It is inferred that authors are interested to contribute more in conferences rather than journal sources. Year wise analysis shows that during the year 2004 the highest mean is traced as 4.489 with StdDev 2.344 followed by 2005, 2002, 2006 and 2003, which is the least one when compared to other's mean scores. R2 shows 0.01 per cent of variation for years. It is inferred that authors' contributions are high during the year 2004 compared to rest of the years. Analysis for authors explore that collaborative authors are in the top slot with the mean 4.726 followed by double (4.252) and single author (3.414) in that order. R2 shows that just 0.01 per cent of variation is traced for the model authors. It is inferred that papers with the collaborative authors show the inclining trend that should be a healthy sign in AI research field. TheAuthors prefer to contribute in collaboration mode rather than alone or double. Conference articles with Collaborative mode are in top while journal articles with collaborative output have got the next slot. It is inferred that in both the sources collaborative authors' contributions are dominated.

(This stage may be a right place, where further research is required). R2 is the strength of the model fit, which shows that just 0.07 per cent of variation in years*authors is explained by the model. Collaborative contributions during the year 2005 are higher than the rest of the authors. (This stage may be a right place, where further research is required).

The three-way anova with 2 way interaction test results reflect that there would be a possible significance exist in sources(P=<0.0001), which would be the reason for not accepting the hypothesis (H2) in favour of the alternative. On the other hand, there would not be a possible significance identified for the remaining classes, which would be the reason for not rejecting the hypotheses H2-H7 in favour of the alternative at the significance level of alpha 0.05. Tables: 13-18 explore the hypotheses test results status for better understanding.

The complete post hoc comparisons tests at the significance level of 0.05 among the years have been discussed below.

The Tukey HSD test (Table 8) for differences between means for the groups has not identified as significant except the pairs 2-3, 2-4 (p value 0.0125, 0.0485). Hence, it could be inferred that there would not be an enough evidences to support the formulated hypothesis H8 that the observed group means among the years are equal. However, statistic test results for a few pairs show an opposite inference.

The complete post hoc comparisons tests at the significance level of 0.05 among the authors have been discussed below.

The Tukey HSD test (Table 9) for differences between means for the groups has identified as significant for all the pairs. Hence, it could be inferred from the test results that there would not be enough evidences to support the formulated hypothesis H9 that the observed group means among the authors are equal.

The complete post hoc comparison tests at the significance level of 0.05 among the years within the journal articles have been discussed below.

The Tukey HSD test (Table 10) for differences between means for the groups has not identified as significant for all the pairs. Hence, it could be inferred from the test results that there would be enough evidence to support the formulated hypothesis H10 that the observed group means among the years within journal sources are equal.

The complete post hoc comparison tests at the significance level of 0.05 among the years within the conference articles have been discussed below.

The Tukey HSD test (Table 10.1) for differences between means for the groups has not identified as significant. Hence, it could be inferred from the test results that there would be enough evidence to support the formulated hypothesis H10 that the observed group means among the years within conference sources are equal.

The complete post hoc comparison tests at the significance level of 0.05 among the sources within the year 2002 have been discussed below.

The Tukey HSD test (Table 11) for differences between means for the groups is identified as insignificant. Hence, it could be inferred from the test results that there would be enough evidence to support the formulated hypothesis H11 that the observed group means among the sources within 2002 are equal.

The complete post hoc comparison tests at the significance level of 0.05 among the sources within the year 2003 have been discussed below.

The Tukey HSD test (Table 11.1) for differences between means for the groups is identified as insignificant. Hence, it could be inferred from the test results that there would be enough evidences (P=0.0815) to support the formulated hypothesis H11 that the observed group means among the sources within 2003 are equal.

The complete post hoc comparison tests at the significance level of 0.05 among the sources within the year 2004 have been discussed below.

The Tukey HSD test (Table 11.2) for differences between means for the groups has identified as significant. Hence, it could be inferred from the test results that there would not be enough evidences (P=0.0419) to support the formulated hypothesis H11 that the observed group means among the sources within 2004 are equal.

The complete post hoc comparison tests at the significance level of 0.05 among the sources within the year 2005 have been discussed below.

The Tukey HSD test (Table 11.3) for differences between means for the groups has identified as significant. Hence, it could be inferred from the test results that there would not be enough evidences (P=0.0407) to support the formulated hypothesis H11 that the observed group means among the sources within 2005 are equal.

The complete post hoc comparison tests at the significance level of 0.05

among the sources within the year 2006 have been discussed below.

The Tukey HSD test (Table 11.4) for differences between means for the groups is identified as insignificant. Hence, it could be inferred from the test results that there would be enough evidences (P=0.0763) to support the formulated hypothesis H11 that the observed group means among the sources within 2006 are equal.

The complete post hoc comparison tests at the significance level of 0.05 among the sources within single author have been discussed below.

The Tukey HSD test (Table 12) for differences between means for the groups has not identified as significant. Hence, it could be inferred from the test results that there would be enough evidences (P=0.0707) to support the formulated hypothesis H12 that the observed group means among the sources within single author are equal.

The complete post hoc comparison tests at the significance level of 0.05 among the sources within double authors have been discussed below.

The Tukey HSD test (Table 12.1) for differences between means for the groups has not identified as significant. Hence, it could be inferred from the test results that there would be enough evidences (P=0.0766) to support the formulated hypothesis H12 that the observed group means among the sources within double authors are equal.

The complete post hoc comparison tests at the significance level of 0.05 among the sources within collaborative authors have been discussed below.

The Tukey HSD test (Table 12.2) for differences between means for the groups has not identified as significant. Hence, it could be inferred from the test results that there would be enough evidences (P=0.0790) to support the formulated hypothesis H12 that the observed group means among the sources within collaborative authors are equal.

However, these estimates may vary over repeated samples of data.

Apart from the above discussed possibilities for further research, a few more chances are there to analyze the reasons behind the variations among the variables that have been opted for the post hoc comparisons. But, this model would be a complicated one like the present study, and requires an advanced analysis with the right interpretations. For instance, comparisons among the conference and journal article output within 2004 show the significance with q=4.173 and p=.0.0419. This location enables the researchers to formulate a research question. The research question may be, why does source wise productivity differ during the year 2004? However, this kind of study requires a good experimental design. Researchers, who are keen to conduct the research in this way may further, alter the shape of the research in a multiple dimensions.

Hypotheses Test Results:

Though the researchers have briefed everything in their findings part, there would be a need to brief the hypothesis test results as shown below. This kind of presentation may enable one to get a clear picture about the research questions and hypothesis test results.

Table 13
Hypotheses 1-7

Table No.	Hypotheses	Research Questions	Answer	Status
7	1	Is there a difference in overall model?	Yes	Not Accepted
7	2	Is there a difference between the sources?	Yes	Not Accepted
7	3	Is there a difference between the years wise output?		
		2002-2006	No	Accepted
		2003-2006	No	Accepted
		2004-2006	No	Accepted
		2005-2006	No	Accepted
7	4	Do authors differ in their contributions?		
		Collaborative Authors-Single Authors	No	Accepted
		Double Authors-Single Authors	No	Accepted
7	5	Do interactions differ between the sources and years?		
		Conference Articles & Journal Articles*2002-06	No	Accepted
		Conference Articles & Journal Articles*2003-2006	No	Accepted
		Conference Articles & Journal Articles*2004-2006	No	Accepted
		Conference Articles & Journal Articles*2005-2006	No	Accepted
7	6	Do interactions differ between the sources and authors?		
		Conf. &Jrl.*CA-SA	No	Accepted
		Conf. &Jrl.*DA-SA	No	Accepted
7	7	Do interactions differ between the years and authors?		
		2002-2006*CA-SA	No	Accepted
		2002-2006*DA-SA	No	Accepted
		2003-2006*CA-SA	No	Accepted
		2003-2006*DA-SA	No	Accepted
		2004-2006*CA-SA	Yes	Not Accepted
		2004-2006*DA-SA	No	Accepted
		2005-2006*CA-SA	No	Accepted
		2005-2006*DA-SA	Yes	Not Accepted

Table 14
Research Question: Is there a difference between group means among years?

Groups	Table	Answer	Hypothesis 8
1 - 2	8	No	Accepted
1 - 3	8	No	Accepted
1 - 4	8	No	Accepted
1 - 5	8	No	Accepted
2 - 3	8	Yes	Not Accepted
2 - 4	8	Yes	Not Accepted
2 - 5	8	No	Accepted
3 - 4	8	No	Accepted
3 - 5	8	No	Accepted
4 - 5	8	No	Accepted

Table 15
Research Question: Is there a difference between group means among authors?

Groups	Table	Answer	Hypothesis 9
1 - 2	9	Yes	Not Accepted
1 - 3	9	Yes	Not Accepted
2 – 3	9	Yes	Not Accepted

Table 16
Research Question: Is there a difference between means of years within each source?

Journal Articles:				Conference Articles:			
Groups	Table	Answer	Hypothesis 10	Groups	Table	Answer	Hypothesis 10
1 - 2	10	No	Accepted	1 - 2	10.1	No	Accepted
1 - 3	10	No	Accepted	1 - 3	10.1	No	Accepted
1 - 4	10	No	Accepted	1 - 4	10.1	No	Accepted
1 - 5	10	No	Accepted	1 - 5	10.1	No	Accepted
2 - 3	10	No	Accepted	2 - 3	10.1	No	Accepted
2 - 4	10	No	Accepted	2 - 4	10.1	No	Accepted
2 - 5	10	No	Accepted	2 - 5	10.1	No	Accepted
3 - 4	10	No	Accepted	3 - 4	10.1	No	Accepted
3 - 5	10	No	Accepted	3 - 5	10.1	No	Accepted
4 - 5	10	No	Accepted	4 - 5	10.1	No	Accepted

Table 17
Research Question: Is there a difference between means of sources within each year?

Groups	Table	Answer	Hypothesis 12
1 - 2	12	No	Accepted
1 - 2	12.1	No	Accepted
1 – 2	12.2	No	Accepted

Table 18
Research Question: Is there a difference between means of sources within each author?

Groups	Table	Answer	Hypothesis 11
1 - 2	11	No	Accepted
1 - 2	11.1	No	Accepted
1 - 2	11.2	Yes	Not Accepted
1 - 2	11.3	Yes	Not Accepted
1 - 2	11.4	No	Accepted

REFERENCES

1. Atificial Intelligence (AI). (2010). *Encyclopaedia Britannica 2010 Ready Reference CD.* Chicago: Encyclopaedia Britannica.
2. David nicholas et.al.(2002). NHS Direct online: Its users and their concerns. *Journal of Information Science,*28, 305-319.
3. Friendly, M.(1994). Mosaic displays for multi way contingency tables. *Journal of the American Statistical Association.* 89. p.190-200.
4. George, A. Marcoulides.(2000) “Generalizability Theory”. *Applied multivariate statistics.*Howard E.A. Tinsley and Stenven D.Brown(Eds.).(pp.536-539).
5. Hagstrom, W.O. (1965). *The scientific community,* Basic, New York.
6. Mark hallahan & Robert Rosenthal. (2000)Interpreting and Reporting Results. Howard,E.A & Steven D,Brown(Eds.),*Applied Multivariate Statistics and Mathematical Modeling.* San Diego; Academic Press.
7. Miller, W. G.(2009). Openstat Reference Manual. *Analysis of Variance.*
8. Nonaka, I, Sasaki, K, and Ahmed, M.(2003). Continuous innovation in Japan: the power of tacit knowledge. *The International Handbook on Innovation.* L.V. Shavinina, (Ed.) Elsevier, St. Louis.
9. Nunnally,J.C & Bernstein,I.H.(1994).*Psychometric Theory.* NY; McGraw-Hill.
10. Rojer Stern et al.(2002).Introduction to Instat plus, Retrieved on Nov,03,2013 from www.rdg.ac.uk
11. Saravanan, T.(2005). Performance of Individual Researcher and GNP per Capita in Astronomy Science Research Output at International Level: A Study. *SRELS, 42, 213-216.*
12. Saravanan, T.(2006).Report on the Potential Aspects of Research in Astronomy in G7 Countries: A Bibliometric Analysis. *IASLIC, 51, 169-177.*
13. Saravanan, T et al. (2009). Research output in Artificial Intelligence: A Performance analysis towards G8 Countries. *Library Progress (International),*29, 247-270.
14. Saravanan, T.(2010). Research Violations: *Are you in the right Track?,* (pp.63-65.) Trichy: LIST & AALIS, Bishop Heeber College.
15. Saravanan, T. (2013a). A dialogue on LIS research issues between Prof. Srimurugan.A and Saravanan.T: A Simulation. *AJMR,* 2, 25-32.
16. Saravanan, T. (2013b). Assessing Single Authors Research Contributions in Artificial Intelligence: A Comprehensive analysis. *AJMR,* 2, 31-51.
17. Saravanan, T. (2013c). Are there differences in USA Research Output in Artificial Intelligence?. *IJILIS,* 26, 119-131.
18. Young, F. W; Bann C.(1992). Analysis of Variance. *The Visual Statistics System.* Retrieved on April, 17, 2012) from www.visualstats.org

Young, F. W.(2003).Vista Statistics Reference Guide. Retrieved on April, 17, 2012) from www.visualstats.org

●●●