Linear Regression Analysis of Title Word Count and Article Time Cited using R

Alireza Mohebbi*, Yousef Douzandegan

Student Research Committee, School of Medicine, Golestan University of Medical Sciences, Gorgan, IRAN.

ABSTRACT

There is a common idea, that title variables like article title length would influence article citations. The aim of the present study was to investigating possible relationship between size of article title and number of article citations by minimizing scientometric variable biases. A dataset containing ~100,000 virological literatures was obtained from Web of Science InCite™ database from 1997 to 2016. Variables, Title size (TWC), Year (YoP), Source (JS), and Publisher were selected. In addition number of times article is cited 'Time Cited' (TC) was retrieved from Web of Science InCite™. Linear regression analysis was performed between variables and TC using R for a possible prediction model for TC. Result has shown a robust standard error corrected linear regression with only 30.6% power of predictability. Furthermore, it was found that TC, YoP, and JS have meaningful potential in the linear model. Moreover, TC is negatively correlated with YoP, JS, and positively with TWC. As a result, size of article title, years passed since publication and the journal in which article accepted are good but not sufficient predictor of article citations. In addition, article is a multi-characteristic subject and other predictors can be supposed. However, we think that finding an efficient statistical linear predication model for TC, by increase of articles citation, is overwhelming.

Keywords: Scientometric, Article Time Cited, Title Word Count, Linear Regression, R statistic.

INTRODUCTION

Effects of article title length on article impact are controversial. Studies have shown that an article title length may have a positive, negative, our neutral influence on articles citations.[1,2,3] However, many factors may affect study's outcome. Importantly, sample size, statistical methods, journals or topics by which articles are retrieved, and time-spans are major factors. For example, popular journals with more attentions may cause a bias toward those in which articles title are in a pre-defined format by the journal.[1]

Studies about this subject had different results due to different data selection, statistical analysis strategies. In a study on more than 9000 article from 22 different journals, authors conclude that articles in journals with higher impact factors tend to have large word counts in title and get more citations.[5] In later study, authors have chosen more variables from titles of 423 Articles divide into two separate results-describing and methods-describing titles groups. With different statistical analysis as well as logistic regression they have shown that titles with less characters would bring more citations.[12] Results similar to those of Paiva *et al* were obtained in a study of seven journals from PloS publication. It was found that though each journal had different scope short titles had more downloads and citations.[2]

These studies clearly show that design of study may lead to either positive or negative correlation between title size and citations. Therefore to minimize bias due to different data sources, in this paper we have selected a dataset

***Address for correspondence:**

Alireza Mohebbi, Student Research Committee, School of Medicine, Golestan University of Medical Sciences, Gorgan, IRAN.

Email: Mohebbi-a@goums.ac.ir

Phone no: +98-0935-4674593

Fax: 0098 17 32430564

Access this article online	
Official Publication of Phcog.Net	**Website:** www.jscires.org
	DOI: 10.5530/jscires.6.1.3

from uniform topic and research area within a certain time period.

R is a free statistical tool with over 2,000 cutting-edge, user-contributed packages available on CRAN. Additionally, we preferred R to other statistical tools because of, in addition to its availability, accessibility to routinely updated advanced packages incorporating recent developments in mathematics making it a comprehensive tool to carry out different types of data analysis, use of data presentation packages, and it's capability to incorporate and analyzing various types of data formats.

As a result, it was found that article word length has a potential impact on article citation. In addition, it was concluded that, along with title size, other scientometrics variables would have influence on article citations.

MATERIALS AND METHODS

Data Retrieve and Article Title Word/Character Counting

One thousand scientific article records in virology research area (SU=Virology) were retrieved from Web of Science InCite™ database from 1997 to 2016 as on September 27, 2016. After that, data was merged into CSV file format using Publish or Perish 4.0 software package (Harzing, A.W. 2007, http://www.harzing.com/pop.htm). Microsoft (MS) Excel formulae were used for data manipulation and title word counting.

Articles database was cleaned for any duplication and articles with missing data on any of the selected variables mentioned in 2.2 were deleted.

Variable Selection

Information on the following variables was then tabulated from the above articles database:

Article title word count (TWC), year of publication (YoP), publishers, and journal sources (JS) in which article is published and number of article citations (TC)retrieved from Science In Cite™ as on September 27, 2016. Journals (JS) were grouped into high impact or low impact factor using Web of Science Journal Citation Reports®. Data mining was performed in less than one hour.

Statistical Analysis

We have selected chosen high enough subjects per variable (SPV) to prevent R^2 biases.[2] R statistic software was used for data analysis.[4] Packages used were : lmtest, rms, Hmisc, and ggplot2 was used for diagnosing the heteroscedasticity of the regression model, correct standard errors, correlation, and drawing scatter plot.[6],[7],[8-13]

RESULTS

After data retrieval and trimming, only 99,838 articles were left with desired information. As shown in Figure 1, articles were from 37 publishers and 56 journals. Based on sources and publishers, mean of citations and total number of articles are shown in Tables 1 and 2. Sum of citations in 99,838 articles was 2,542,056. Figure 1a shows American Social Microbiology with largest number of papers to have most citations (40.55 ± 49.950) and TWCs with subjects AIDS Research and Human Retroviruses having largest total citations in years between s1997-2016. Linear regression resulted in a model with a prediction ability of 13.68% (y-intercept = 62.325, slope = 0.545, adjusted R-squared = 0.1368, and p-value < 2.2e-16). All Predictors had significant same < 2.2e-16 p-value in regression modeling, but with TWC p-value equal to 1.12e-06. This model with three TWC, YoP, and Publisher predictors was better than that models based on each variable alone. Adjusted R-squared with the Year predictor was 0.1228, 0 with the Journal Source variable, 0.021 with the Publisher predictor, and 0 with TWC. However, predictors TWC (p-value = 0.012), Year (p-value: < 2.2e-16), and Publisher (p-value < 2.2e-16) have a synergic potential to prediction of TC ($Adj.R^2$=0.1368). By removing Journal Source from the model, the power of prediction has changed inconsiderably ($Adj.R^2$= 0.1357, p-value < 2.2e-16). Figure 3 shows diagnostic plot of the predicted model. In addition, there was no multi-colinearity in any of predictors. Moreover, heteroscedasticity was evaluated to check for hetero dispersion within variable if any using studentized Breusch-Pagan test. Result showed existence of non-acceptable heteroscedasticity (BP = 527.89 [df: 4], p-value < 2.2e-16). Standard errors were corrected to take care of this. The correction was found to change R-squared value to 0.306 (y-intercept: 4.19, S.E:0.02, p-value < 0.0001).

Figure 4 shows observed vs. predicted values. Based on journal source, R^2 was significantly higher but only for those journals in with few articles. Predicting equation model obtained with these journal sources were not able to predict observed citations (data is not shown). We hypothesized that significant negative correlation between citations and Publisher could be because of inclusion of large number of journals with low impact factors (IF) in the dataset. To answer this question, data was split into

Table 1: Information of data based on publishers.

Publisher	Citations (Mean±S.D)	TWC (Mean±S.D)	Yeara	Number of Articles	R^2 (p-value)[b]
ACADEMIC PRESS	26.08± 33.275	16.10± 5.265	1997-2016	9180	0 (0.1)
AEPRESS	2.13± 3.623	15.08± 4.344	1998-2015	256	0.006 (0.1)
AMER SOC MICROBIOLOGY	40.55± 49.950	16.73± 5.371	1997-2016	25699	0 (0.5)
ANNUAL REVIEWS	3.57± 4.233	8.32± 3.369	2014-2015	56	0 (0.8)
AOSIS OPEN JOURNALS	0 .43± 1.042	12.65± 5.117	2008-2015	37	0 (0.5)
BENTHAM SCIENCE PUBL LTD	7.84± 10.574	13.70± 13.70	2003-2016	649	0.015 (0.01)
BIOMED CENTRAL LTD	11.39± 15.280	15.54± 5.047	2004-2016	3834	0.004 (0.0001)
BLACKWELL	26.40± 28.365	16.18± 5.314	1997-2008	792	0 (0.3)
CELL PRESS	50.21± 63.049	13.79± 3.463	2007-2016	876	0 (0.8)
EDITIONS SCIENTIFIQUES MEDICALES ELSE	10.72± 10.792	13.25± 4.149	1997-1998	85	0 (0.7)
ELSEVIER	15.23± 21.274	15.29± 5.102	1997-2016	15001	0.005 (0.0001)
FUTURE MEDICINE LTD	1.73± 3.356	11.60± 4.267	2006-2016	275	0.003 (0.3)
GUSTAV FISCHER VERLAG	9.17± 13.644	13.84± 5.768	1997-2000	264	0 (0.4)
HINDAWI PUBLISHING CORP	0 .24±0.437	14.47± 2.348	2015-2016	17	0 (0.6)
INDIAN VIROLOGICAL SOC	1.97± 2.201	15.15± 4.426	2005-2010	79	0 (0.7)
INT MEDICAL PRESS	17.28± 21.244	15.72± 4.493	1997-2015	1787	0 (0.6)
JOHN WILEY & SONS LTD	27.64± 38.136	8.77± 4.810	1997-2009	47	0.010 (0.2)
KARGER	11.77± 23.213	14.96± 5.654	1997-2015	971	0 (0.4)
KLUWER ACADEMIC	12.63± 14.908	14.85± 5.389	1998-2004	546	0 (0.6)
LIPPINCOTT WILLIAMS & WILKINS	37.40± 52.752	14.75± 4.205	1997-2016	6224	0 (0.4)
MARY ANN LIEBERT	14.12± 19.250	16.80± 5.166	1997-2016	4567	0 (0.9)
MDPI AG	3.62± 5.293	14.56± 4.883	2005-2016	683	0.009 (0.008)
MICROBIOLOGY SOC	3.50± 3.536	18.50± 7.778	2013, 2015	2	-
NATURE PUBLISHING GROUP	27.69± 24.455	13.53± 5.193	2000-2001	127	0.005 (0.2)
NEW YORK ACAD SCIENCES	39.59± 33.431	11.15± 4.258	2001	39	0.13 (0.014)
PLENUM PRESS DIV PLENUM PUBLISHING CO	9.66± 16.864	13.28± 5.102	1998	106	0.017 (0.1)
PUBLIC LIBRARY SCIENCE	31.42± 38.978	14.19± 4.139	2005-2016	4969	0.006 (0.0001)

RAPID SCIENCE PUBLISHERS	52.66± 58.820	14.46± 5.277	1997	215	0 (0.7)
SA HIV CLINICIANS SOC	1.43± 5.222	12.12± 5.069	2007-2014	165	0 (0.9)
SLOVAK ACADEMIC PRESS LTD	6.89± 7.477	14.97± 4.770	1997-2011	530	0.004 (0.08)
SOC GENERAL MICROBIOLOGY	25.72± 31.643	16.29± 5.348	1997-2016	6762	0.001 (0.017)
SPRINGER	10.95± 21.797	15.21± 4.982	1997-2016	7234	0 (0.5)
STOCKTON PRESS	25.49± 36.054	14.39± 5.866	1997-2000	183	0.007 (0.1)
TAYLOR & FRANCIS INC	21.55± 26.840	15.62± 5.488	2001-2010	506	0 (0.6)
URBAN & FISCHER VERLAG	26.14± 31.251	13.55± 5.563	2000-2005	288	0.014 (0.024)
WILEY-BLACKWELL	16.21± 22.819	15.94± 4.836	1997-2016	6775	0.001 (0.005)
WORLD HEALTH ORGANIZATION	0 .00	0 .000	2013	12	-
Total	25.46±38.006	15.77±5.150	-	99838	

[a]Year in which data are published.
[b] Adjusted-R^2 of linear regression analysis of TC and TWC based on publisher.

two categories, one with four quartiles (Cat. 1-4) of journals with IF articles with less than 1.5, 1.501-2.45, 2.451-4.1 and another of journals with IF greater than 4.11. Moreover, data were spilled uniformly in each IF category based on the source of publication. As it is illustrated in Figure 2, high impact journals have more citations as expected. In contrast in my hypothesis, large portions of articles were in to categories with more IF.

The correlation between TC and other parameters was investigated. Results have shown negative correlations for TC and Yop (-0.35, p=0.0001), Publisher (-0.14, p=0.0001), a positive correlation with TWC (0.01, p=0.0121), and no correlation with journal source (0, p>0.05) (Figure 5).

DISCUSSION AND CONCLUSION

A p-value less than 0.05 is considered sufficient for assigning a variable into a predicting linear model. Linear regression results obtained here also indicate effect of TWC on response variable, TC. However in this paper we have examined in detail if TWC-based linear model for predicting response variable TC is reliable or not.

We have conducted a linear regression analysis on a database containing Virological papers. Interestingly, using TWC variable, we found that in case of low TC in sets of data containing small number of articles, a linear model can be assigned (Table 2). However, results do not show a reliable linear model for prediction of TC irrespective of number of articles and high TC . It is likely that in articles that receive higher number of citations, readers pay attention to many more variables than simply TWC, making it harder to model a regression.

Having checked relationship between TWC and TC, to show no linear relations (only 30.6% predicting ability with standard error corrections) we have then incorporated, in addition to TWC (article word size), YoP (year of publications), and JS (journal source) and searched f for a meaningful predictors of TC (article time cited). We find that TC is negatively correlated to YoP and JS (Publisher,) and positively with TWC (P<0.05). Negative correlation of JS and TC, is shown in Figure 2, thus TC of articles in high impact factors journals during the years 1997-2016 are less predictable.

We note that Scientometric and Bibliometrics studies-employ varied ways of data collection and analysis. However, a scientific paper also has descriptive and reflective contents.[1] Falahati *et al*, have observed that title length and subject of article are both relevant to article citations, but they did not find correlation between title length and citations,[3] implicating other factors from bibliometrics materials may be involved. Article citation may be influenced by research area, topics, words size, characters, punctuations etc. Also some topics, in a certain time period may attractmore interest than other subjects. Therefore analysis based on, different time period segments may minimize biases either in variety of published

Table 2: Information of data based on publishers

Journal Source	Citations (Mean±S.D)	TWC (Mean±S.D)	Yeara	Number of Article	R^2 (p-value)[b]
ACTA VIROLOGICA	5.34± 6.850	15.01± 4.633	1997-2015	786	0.005 (0.031)
ADVANCES IN VIROLOGY	0 .24± 0.437	14.47± 2.348	2015-2016	17	0 (0.6)
ADVANCES IN VIRUS RESEARCH	0.00	7.00	2000	1	-
AIDS	38.20± 53.328	14.70± 4.209	1997-2016	6345	0 (0.2)
AIDS RESEARCH AND HUMAN RETROVIRUSES	15.12± 20.425	16.95± 5.218	1997-2016	3680	0 (0.8)
ANNUAL REVIEW OF VIROLOGY	3.57±4.233	8.32±3.369	2014-2015	56	0 (0.8)
ANTIVIRAL CHEMISTRY & CHEMOTHERAPY	15.57±17.729	14.83±5.156	1997-2001	184	0.001 (0.3)
ANTIVIRAL CHEMISTRY & CHEMOTHERAPY CLINICAL A	6.96±9.998	6.78±3.384	1999	23	0.084 (0.1)
ANTIVIRAL RESEARCH	17.57±22.755	15.23±4.826	1997-2016	2027	0.009 (0.0001)
ANTIVIRAL THERAPY	17.47±21.607	15.82±4.401	1998-2015	1603	0 (0.8)
ARCHIVES OF VIROLOGY	12.70±24.843	15.45±5.053	1998-2015	4894	0 (0.2)
BIOLOGY OF EMERGING VIRUSES: SARS, AVIAN	28.20±24.917	9.20±5.534	2007	10	0.372 (0.036)
BULLETIN DE L INSTITUT PASTEUR	3.00±4.359	10.67±2.517	1997-1998	3	0 (0.9)
CELL HOST & MICROBE	50.21±63.049	13.79±3.463	2007-2016	876	0 (0.8)
CLINICAL AND DIAGNOSTIC VIROLOGY	21.46±20.877	15.01±4.930	1997-1998	70	0.005 (0.2)
CORONAVIRUSES AND ARTERIVIRUSES	9.66±16.864	13.28±5.102	1998	106	0.017 (0.1)
CURRENT HIV RESEARCH	7.84±10.574	13.70±4.776	2003-2016	649	0.015 (0.001)
CURRENT OPINION IN VIROLOGY	13.79±17.029	8.89±3.191	2011-2016	513	0.003 (0.1)
FOOD AND ENVIRONMENTAL VIROLOGY	5.48±9.023	14.08±4.589	2009-2016	217	0.015 (0.039)
FUTURE VIROLOGY	1.73±3.356	11.60±4.267	2006-2016	275	0.003 (0.2)
GASTROENTERITIS VIRUSES	25.00±28.154	6.29±3.312	2001	17	0 (0.9)
HIV INTERACTIONS WITH DENDRITIC CELLS: INFECT	6.70±4.855	8.80±2.044	2013	10	0 (0.7)
INDIAN JOURNAL OF VIROLOGY	1.70±1.914	15.69±4.664	2009-2013	162	0 (1)
INFLUENZA AND OTHER RESPIRATORY VIRUSES	6.58±11.464	15.31±4.864	2009-2016	742	0 (0.5)
INTERNATIONAL JOURNAL OF MEDICAL MICROBIOLOGY	16.32±22.870	14.52±5.067	2000-2016	1056	0.013 (0.0001)
INTERVIROLOGY	11.69±23.543	15.11±5.639	1997-2015	934	0 (0.4)
JAAGSIEKTE SHEEP RETROVIRUS AND LUNG CANCER	32.50±12.581	9.75±4.950	2003	8	0.474 (0.035)

JOURNAL OF CLINICAL VIROLOGY	16.17±23.983	15.22±5.193	1998-2016	3087	0.008 (0.0001)
JOURNAL OF GENERAL VIROLOGY	25.72±31.640	16.29±5.348	1997-2016	6764	0.01 (0.017)
JOURNAL OF HUMAN VIROLOGY	18.31±17.149	17.06±5.740	1999-2002	94	0 (0.5)
JOURNAL OF MEDICAL VIROLOGY	18.58±24.732	15.83±4.842	1997-2016	4999	0.001 (0.007)
JOURNAL OF NEUROVIROLOGY	18.48±27.584	14.79±5.419	1997-2016	1222	0.001 (0.2)
JOURNAL OF VIRAL HEPATITIS	18.04±23.277	16.67±4.923	1997-2016	1810	0.002 (0.033)
JOURNAL OF VIROLOGICAL METHODS	14.33±20.025	16.10±4.807	1997-2016	4791	0.001 (0.016)
JOURNAL OF VIROLOGY	40.55±49.950	16.73±5.371	1997-2016	25699	0 (0.5)
NIDOVIRUSES (CORONAVIRUSES AND ARTERIVIRUSES)	3.92±5.611	13.53±4.643	2001	102	0 (0.6)
NIDOVIRUSES: TOWARD CONTROL OF SARS AND OTHER	3.74±3.726	10.67±3.836	2006	111	0 (0.5)
PLOS PATHOGENS	31.42±38.978	14.19±4.139	2005-2016	4969	0.006 (0.0001)
POLYOMAVIRUSES AND HUMAN DISEASES	25.96±25.740	8.38±3.487	2006	24	0.078 (0.1)
RESEARCH IN VIROLOGY	11.45±12.727	13.07±4.258	1997-1998	89	0 (0.4)
RESPIRATORY VIROLOGY AND IMMUNOGENICITY	0.88±0.991	12.50±3.625	2015	8	0.535 (0.024)
RETROVIROLOGY	16.40±19.993	15.34±5.049	2005-2016	1061	0.004 (0.029)
REVIEWS IN MEDICAL VIROLOGY	25.00±40.541	10.14±4.673	1997-2013	17	0.037 (0.1)
SEMINARS IN VIROLOGY	22.00±17.297	8.19±3.731	1997-1998	26	0.195 (0.014)
SIMIAN VIRUS 40 (SV40): POSSIBLE HUMAN POLYOM	13.92±12.203	11.30±4.795	1998	37	0.012 (0.2)
SOUTHERN AFRICAN JOURNAL OF HIV MEDICINE	1.25±4.753	12.22±5.069	2007-2015	202	0 (0.9)
VIRAL IMMUNOLOGY	9.95±12.479	16.18±4.897	1998-2016	887	0 (0.3)
VIROLOGICA SINICA	0.54±0.691	14.68±2.897	2015-2016	37	0.063 (0.07)
VIROLOGY	26.10±33.310	16.12±5.251	1997-2016	9153	0 (0.1)
VIROLOGY JOURNAL	9.47±12.519	15.61±5.046	2004-2016	2773	0.003 (0.002)
VIRUS GENES	9.28±11.538	15.39±4.742	1998-2016	1857	0 (0.2)
VIRUS RESEARCH	14.95±20.696	15.30±5.171	1997-2016	3738	0.008 (0.0001)
VIRUSES-BASEL	3.62±5.293	14.56±4.883	2009-2016	683	0.009 (0.008)
WEST NILE VIRUS: DETECTION, SURVEILLANCE, AND	39.59±33.431	11.15±4.258	2001	39	0.13 (0.014)
WHO EXPERT CONSULTATION ON RABIES: SECOND REP	0.00±0.000	4.75±1.815	2013	12	-

ZENTRALBLATT FUR BAKTERIOLOGIE-INTERNATIONAL	9.17±13.644	13.84±5.768	1997-2000	264	0 (0.4)
Total	25.46±38.006	15.77±5.150	-	99838	

[a] Year in which data are published.

[b] Adjusted-R^2 of linear regression analysis of TC and TWC based on publisher.

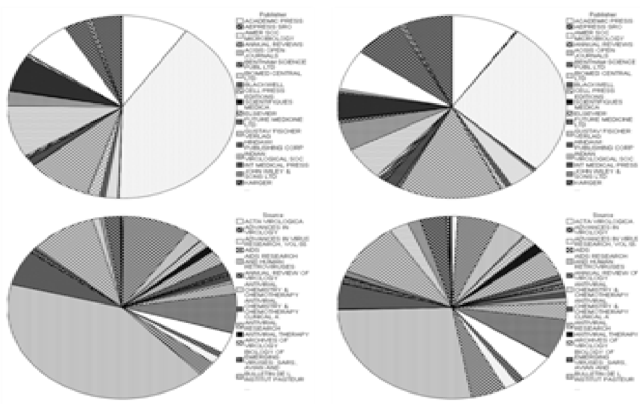

Figure 1: Schematic representation of TWC and TC based on publishers and journal source. a) Shows total TC in each publisher, b) illustrates TWCs within publishers, c) is TC related to journal sources, and d) demonstrated number of TWC used in each journal source.

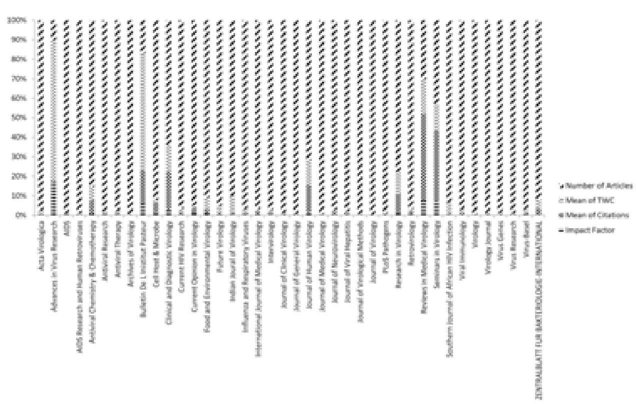

Figure 2: Journal Sources with their respect Scientometrics information's. Journal sources in Cat1 comprised of TC, number of articles and TWC mean of 6.5, 424.67, and 13.59, respectively. In Cat 2, data have changed to 10.45 for TC, 1907.36 for number of articles, and 14.68 for TWC. Mean of citations in Cat 3 and 4 was 17.67 and 27.19, 2843.70 and 4596.20 for mean number of articles, 15.30 and 13.55 for TWC, respectively.

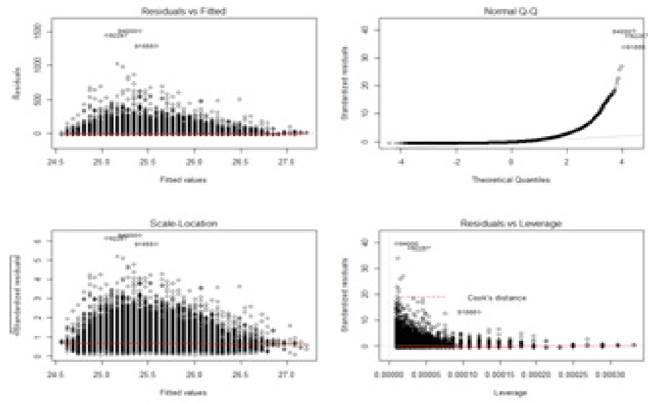

Figure 3: Diagnostic plot of adjusted linear model. Upper left plot shows residuals versus fitted values; data are in the regression line as red line is laid on the dotted line. Normal Q-Q plot shows residual are normally distributed; outline data with higher citations are shown as well. Lower left plot is used to measure square root of standardize residuals against the fitted value; as the red line is flat, it is assumed that the variance of residuals does not change the distribution. Fourth plot shows how each data point influences the regression; as it is shown, outliers- high leverage and large residuals data (point over 0.5 Cook's distances)- may affect linear regression fit, as the red line is not leaving dashed line, it indicate good regression fit.

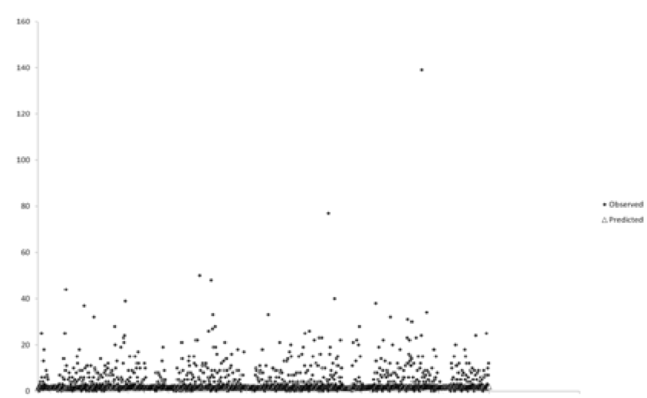

Figure 4: 1/100 random sample predicted by adjusted, standard error corrected linear model.

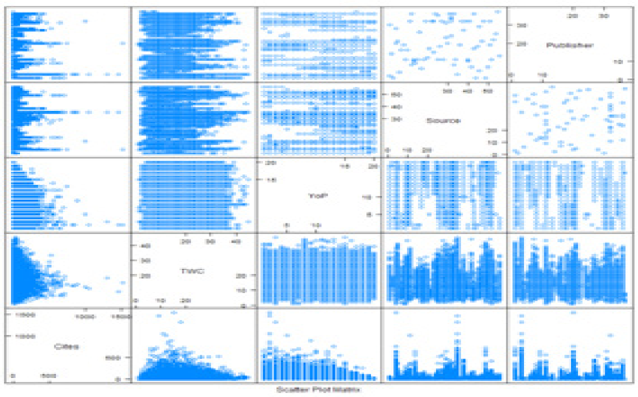

Figure 5: Scatter plot matrix of data variables correlation.

articles or time variable itself. For this, other methods or data retrieve strategies need to be taken.

In scientific view, number of times an article cites is major impact of the article. Therefore, finding factors influencing article citation needs further research in the future. Accordingly, those factors with high impact on article time cited can be used for reconstruction of statistical predictive model(s).

ACKNOWLEDGMENT

We thank Doctor Mohammad Ali Vakili for unsparing assistance in statistical concepts.

FUNDING SOURCES

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

ABBREVIATION USED

TWC: Title Word Count; YoP: Year of Publication; JS: Journal Source; TC: Time Cited; SU: Subject Area; CSV: Comma Separated Value; MS: Microsoft; SPV: Subject Per Variable; IF: Impact Factor.

REFERENCES

1. Aleixandre-Benavent R, Montalt-Resurecció V, Valderrama-Zurián JC. A descriptive study of inaccuracy in article titles on bibliometrics published in biomedical journals. Scientometrics. 2014;101(1):781-91. https://doi.org/10.1007/s11192-014-1296-5.
2. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. Journal of Clinical Epidemiology. 2015;68(6):627-36. https://doi.org/10.1016/j.jclinepi.2014.12.014; PMid:25704724.s.
3. Falahati Qadimi Fumani MR, Goltaji M, Parto P. The impact of title length and punctuation marks on article citations. Annals of Library and Information Studies (ALIS). 2015;s62(3):126-32.
4. Fox J, Weisberg S, Adler D, Bates D, Baud-Bovy G, Ellison S, Firth D, Friendly M, Gorjanc G, Graves S, Heiberger R. Package 'car'.2016.
5. Habibzadeh F, Yadollahie M. Are shorter article titles more attractive for citations? Crosssectional study of 22 scientific journals. Croatian Medical Journal. 2010;51(2):165-70. https://doi.org/10.3325/cmj.2010.51.165; PMid:20401960 PMCid:PMC2859422
6. Harrell Jr FE. Hmisc: Harrell miscellaneous. R package version 3.12-2. Computer software]. Available from http://cran. R-project. Org/web/packages/Hmisc. 2013.
7. Harrell Jr FE. rms: Regression Modeling Strategies. R package version 4.0-0. City 2013.
8. Hothorn T, Zeileis A, Fare brother RW, Cummins C, Millo G, Mitchell D, Zeileis MA. Package 'lmtest'. Testing linear regression models. https://cran. R-project. Org/web/packages/lmtest/lmtest.pdf. Accessed. 2015 Jun 6;6.
9. Jacques TS, Sebire NJ. The impact of article titles on citation hits: an analysis of general and specialist medical journals. JRSM Short Reports. 2010;1(1):2. https://doi.org/10.1258/shorts.2009.100020;PMid:21103094 PMCid:PMC2984326.
10. Jamali HR, Nikzad M. Article title type and its relation with the number of downloads and citations. Scientometrics. 2011;88(2):653-61. https://doi.org/10.1007/s11192-011-0412-z.
11. Letchford A, Moat HS, Preis T. The advantage of short paper titles. Royal Society Open Science. 2015;2(8):150266. https://doi.org/10.1098/rsos.150266; PMid:26361556 PMCid:PMC4555861.
12. Paiva CE, Lima JP, Paiva BS. Articles with short titles describing the results are cited more often. Clinics. 2012;67(5):509-13. https://doi.org/10.6061/clinics/2012(05)17.
13. Wickham H. ggplot2: elegant graphics for data analysis. Springer New York. 2009;1(2):3. https://doi.org/10.1007/978-0-387-98141-3.

How to cite this article: Mohebbi A, Douzandegan Y. Linear Regression Analysis of Title Word Count and Article Time Cited using R. J Scientometric Res. 2017;6(1):15-22.