Original Article

Optimizing Linear Models via Sinusoidal Transformation for Boosted Machine Learning in Medicine.

DOI: https://doi.org/ 10.32007/jfacmedbagdad.613,41713

Ahmed M. Al-Imam* **MBChB, PG-Diploma, MSc, M.Res.**

J Fac Med Baghdad 2019; Vol.619, No.3,4
Received: Dec., 2019
Accepted: April ,2020
Published:April, 2020

Abstract:

Background: Machine learning relies on a hybrid of analytics, including regression analyses. There have been no attempts to deploy a sinusoidal transformation of data to enhance linear regression models.

Objectives: We aim to optimize linear models by implementing sinusoidal transformation to minimize the sum of squared error.

Methods: We implemented non-Bayesian statistics using SPSS and MatLab. We used Excel to generate 30 trials of linear regression models, and each has 1,000 observations. We utilized SPSS linear regression, Wilcoxon signed-rank test, and Cronbach's alpha statistics to evaluate the performance of the optimization model.

Results: The sinusoidal transformation succeeded by significantly reducing the sum of squared errors (*P-value*<0.001). Inter-item reliability testing confirmed the robust internal consistency of the model (Cronbach's alpha=0.999).

Conclusion: Our optimization model is valuable for high-impact research based on linear regression. It can reduce the computational processing demands for powerful real-time and predictive analytics of big data.

Keywords: Artificial Intelligence; Data Transformation; Machine Learning; Predictive Analytics; Regression.

Introduction:

Although data science and statistical modeling have been evolving for centuries, most data analytic models are not entirely accurate [1]. The British statistician, George EP Box, coined the epigram "*All models are wrong, but some are useful*" [1, 2]. The famous aphorism of statistics appeared in a paper published by George Box [1, 3]. Linear models describe a continuous response variable as a function of predictor variables that can predict the behavior of complex systems [4, 5]. Data scientists implement linear regression to model the causality relationships within data between explanatory and outcome variables [6]. However, these methods are not sufficiently "bulletproof" in terms of statistical precision [7, 8]. Sir Ronald Fisher, a British data scientist and geneticist, introduced the modern regression model in 1922 [9-11]. Ronald Fisher followed the footsteps of contemporary statisticians, including Karl Pearson, a 19th-century English mathematician [12]. Pearson introduced many statistical tests, including Pearson's correlation, which relates to Fisher's linear regression models [13]. The simple linear regression examines the relationship between one predictor variable and one outcome variable. In contrast, the multiple regression tests a multitude of explanatory variables for a higher predictive power [14-16].

**Department of Anatomy, College of Medicine, University of Baghdad. Email: ahmed.lutfi@uob.edu.iq*

Machine learning, an application of artificial intelligence, relies on several methods, including regression models, neural networks, and classification trees [17]. The optimization of those analytics can positively influence machine learning technologies [18-20]. As mentioned earlier, linear regression exploits the least-squares method to extrapolate a line that best fits the causality association [21, 22]. Perfecting the least-squares is critical for a rigorous statistical inference and predictive modeling [23-27]. There are several data transformation techniques that scientists are using to boost a spectrum of statistical tests, particularly for non-Bayesian parametric tests, including the Fourier transformation, the Log Base-10 (Log10) transformation, the natural logarithm (Ln) transformation and inverse transformation as well as the square root and cubic root transformers [28, 29]. The sinusoidal optimization of data can capitalize on unprecedented powerful and economic computational processing for real-time analyses and predictive models [27]. Researchers and data analysts can integrate sinusoidally-optimized linear models in combination with Hill's criteria to infer robust data that possess the least prediction error and the highest statistical power while sparing the human resources and the computational infrastructure to a minimum [30, 31]. Our primary objective is to optimize linear models, principally for analytics that are dependent on correlation and regression statistics, by implementing a trigonometry-based sinusoidal transform function that significantly reduces the error of residuals by minimizing the sum of squared errors (SSE). Hence, achieving more powerful and externally valid linear models that apply for the predictive models in machine learning that are necessary for high-impact research based on big data. These optimized models will not only be most accurate in terms of statistical inference but also will be economical in terms of the computational processing demands for analyses of big data.

Methods:

Mathematical Simulations**:** We made multiple simulations based on a random number generator that follows a normal distribution [mean= 0, standard deviation= 1]. We created thirty trials (models) of linear regression [k= 30]. Each has a sample size of one thousand observations [n= 1000] for two variables as a predictor and an outcome, X and Y, respectively, thereby summing to a grand sample size of 30000 [n_{total}= 30000] for a robust simulation and hypothesis testing. We transformed the two variables using the "sin" function in Excel 2016, thereby changing each variable to the range of -1 to 1. Within each linear model, we calculated correlation and regression statistics, including the sum of squares (SS), mean of squares (MS), F statistic [ANOVA], *P-value* [regression]. We calculated the sum of squared errors using the formula $SSE=\sum (y-\hat{y})^2$ to fulfill the regression equation $\hat{y}=b0+b1X$. We conducted the calculations twice, before [pre-optimization] and after deploying the sinusoidal transformation [post-optimization]. We statistically tested the performance of the sinusoidal optimization model using the Wilcoxon signed-rank test for non-parametric within-subjects statistical inference by comparing the pre-optimization versus post-optimization statistics. Ultimately, we further examined the optimization efficacy of our model by implementing Cronbach's alpha as a measure of the internal consistency of the summative optimization model. Statistical analysis, level of evidence and ethics: We implemented the Statistical Package for the Social Sciences [IBM-SPSS version 24] and Excel [Microsoft Office 2016] with integrated Data Analysis ToolPak. We tabulated some of the raw data with Word [Microsoft Office 2016] and exported it afterward to Excel. We made descriptive statistics using Excel and GNU-Octave version 5.1.0 [GNU's Not Unix Project]. We implemented MatLab high-level programming language (HLL) version R2019a [MathWorks] for two-dimensional array transposition before importing the data via SPSS for Cronbach's alpha calculations. We carried out an elaborate set of parametric as well as non-parametric models of non-Bayesian models of statistics, including linear regression, Fisher' ANOVA, Wilcoxon signed-rank test for within-subjects study design, and Cronbach's alpha analytics for assessing the reliability of our proposed statistical model based on the sinusoidal transformation of data. Our study is of level-1c, which belongs to the top tier [Level-1, Grade-A], in compliance with the categorization scheme by the Oxford Centre for Evidence-Based Medicine [32, 33]. The authors made a mathematical model superimposed onto simulated statistical calculations. Hence, there were neither patients nor animal models that mandate ethical permission.The systematic review of the literature: During the second half of September 2019, we conducted a pragmatic review of the databases of peer-reviewed literature, including the Cochrane Library [the Cochrane Database of Systematic Reviews | the Cochrane Collaboration], PubMed [the United States National Library of Medicine], and Embase [Elsevier]. We implemented an exhaustive set of keywords based on medical subject headings (MeSH) in addition to generic terms while using Boolean expression operators and truncations. Keywords represented five main themes, including 1) machine learning and artificial intelligence, 2) real-time and predictive analytics, 3) real-time analytics and epidemiology, 4) data transform functions, and 5) an amalgamation of the previous four themes. We aimed to explore the existing literature for prior attempts of using sinusoidal data transformation for enhancing and optimizing linear models.

Results*:*

For the optimization model, we generated a sinusoidal transform for 30 trials, i.e., simulations of linear regression analyses (Table 1). The model was triumphant in attaining a significant reduction of the sum of squared errors (SSE) for each trial following the application of the sinusoidal transform [absolute Z-score= 4.782, *P-value*<0.001, Wilcoxon signed-rank test] (Table 2). We utilized a non-parametric alternative of the Student's *t*-test for pair-wise within-subjects study design due to the violation of all of the t-test assumptions, including the absence of statistical outliers, homoscedasticity and the normality of distribution [Shapiro-Wilk test] (Table 2). On the other hand, there was no significant change in the coefficient of determination (R^2 score) for the pre-optimized versus post-optimized trials, as we created each simulation using a random number generator function by using the Data Analysis ToolPak plug-in in Excel. A randomly selected linear model, the seventh trial, manifested with a sum of squared errors of 1.03E+11 [pre-optimization] and 499.797 [post-optimization], confirming a significant SSE reduction and a better predictive model fitting (Figures 1 and 2). The sinusoidal transform had a centrifugal-like effect on the scattered correlates of the tested variables, representing some degree of data deformation. Lastly, Cronbach's alpha analysis yielded collateral evidence and further confirmed the internal consistency of the optimization model (Cronbach's alpha= 0.999) (Table 3). Deleting any trial from the optimization model had no effect on the inter-item reliability with an exception for one simulation, the 16th trial, the deletion of which imperceptibly increases the internal consistency.

Table (1) Optimization Model Analytics.

Trial	Pre-Optimization [Randomly-Generated Normal Distribution]				Post-Optimization [Sinusoidal Transform Function]				*P-value*
	R^2 Score	SSE	F Score	*P-value*	R^2 Score	SSE	F Score	*P-value*	
1	0.007	9.74E+10	6.782	0.009	0.001	506.387	0.745	0.388	
2	0.001	9.99E+10	1.451	0.229	0.002	506.864	2.324	0.128	
3	<0.001	1.01E+11	0.424	0.515	0.050	470.628	2.549	0.111	
4	0.001	9.35E+10	0.559	0.455	0.001	491.498	0.689	0.407	
5	0.001	1.02E+11	1.069	0.301	<0.001	490.904	0.100	0.752	
6	<0.001	1.01E+11	0.281	0.596	<0.001	490.793	0.768	0.381	
7	0.002	1.03E+11	1.753	0.186	0.005	499.797	5.411	0.020	
8	<0.001	9.84E+10	0.222	0.638	<0.001	492.525	0.050	0.823	
9	<0.001	1.03E+11	0.177	0.674	0.001	505.400	0.589	0.443	
10	<0.001	1.00E+11	0.007	0.934	0.001	516.569	0.605	0.437	
11	0.004	9.57E+10	4.211	0.040	<0.001	499.195	0.488	0.485	
12	<0.001	1.08E+11	0.002	0.968	0.002	490.659	2.474	0.116	
13	0.004	9.82E+10	3.809	0.051	<0.001	487.855	<0.001	0.988	
14	0.002	9.96E+10	2.178	0.140	0.002	512.479	2.292	0.130	
15	0.003	1.05E+11	2.773	0.096	0.002	492.662	1.585	0.208	<0.001
16	<0.001	1.02E+10	0.076	0.783	<0.001	489.088	0.127	0.722	
17	<0.001	1.04E+11	0.001	0.979	<0.001	510.862	0.401	0.527	
18	<0.001	9.05E+10	0.043	0.835	0.001	513.120	1.073	0.300	
19	0.004	1.03E+11	4.435	0.035	<0.001	500.158	0.079	0.778	
20	0.002	1.04E+11	1.911	0.167	<0.001	494.541	0.044	0.834	
21	0.006	1.03E+11	6.239	0.013	<0.001	493.778	0.490	0.484	
22	<0.001	9.96E+10	0.457	0.499	<0.001	520.466	0.202	0.653	
23	0.002	1.13E+11	2.184	0.140	0.003	474.729	3.035	0.082	
24	<0.001	1.03E+11	0.013	0.911	<0.001	498.553	0.198	0.657	
25	<0.001	9.50E+10	0.281	0.596	<0.001	479.322	0.195	0.659	
26	0.003	1.05E+11	2.587	0.108	0.003	511.846	3.362	0.067	
27	<0.001	1.04E+11	0.181	0.671	<0.001	498.580	0.003	0.955	
28	<0.001	9.80E+10	0.461	0.498	<0.001	497.725	0.276	0.599	
29	<0.001	9.82E+10	0.412	0.521	<0.001	504.515	0.193	0.660	
30	0.001	9.68E+10	0.973	0.324	<0.001	500.986	0.121	0.727	

† Wilcoxon signed-rank test statistics are in connection with calculations of the sum of squared errors (SSE).
†† Linear model-of-interest is in bold font [Random Selection, Trial 7].

Table (2) Optimization Model Statistics: Normality tests and Wilcoxon signed-rank test.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pre-Optimization SSE	.335	30	.000	.410	30	.000
Post-Optimization SSE	.097	30	.200*	.975	30	.694

*: This is a lower bound of the true significance.
[a]: Lilliefors Significance Correction.

Wilcoxon signed-rank test
Ranks

		N	Mean Rank	Sum of Ranks
Post-Optimization SSE versus Pre-Optimization SSE	Negative Ranks	30[a]	15.50	465.00
	Positive Ranks	0[b]	.00	.00
	Ties	0[c]		
	Total	30		

[a]: Post-Optimization SSE < Pre-Optimization SSE
[b]:Post-Optimization SSE > Pre-Optimization SSE
[c]: Post-Optimization SSE = Pre-Optimization SSE

Test Statistics [a]

	Post-Optimization SSE versus Pre-Optimization SSE
Z	-4.782[b]
Asymp. Sig. (2-tailed)	<0.001

[a]: Wilcoxon Signed Ranks Test.
[b]: Based on positive ranks.

Table 3: Internal Consistency Analysis: Cronbach's Alpha.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	No. of Items
.999	1.000	30

Item-Total Statistics

	Cronbach's Alpha if Item Deleted
Trial 1	.999
Trial 2	.999
Trial 3	.999
Trial 4	.999
Trial 5	.999
Trial 6	.999
Trial 7	.999
Trial 8	.999
Trial 9	.999
Trial 10	.999
Trial 11	.999
Trial 12	.999
Trial 13	.999
Trial 14	.999
Trial 15	.999
Trial 16	1.000
Trial 17	.999
Trial 18	.999
Trial 19	.999
Trial 20	.999
Trial 21	.999
Trial 22	.999
Trial 23	.999
Trial 24	.999
Trial 25	.999
Trial 26	.999
Trial 27	.999
Trial 28	.999
Trial 29	.999
Trial 30	.999

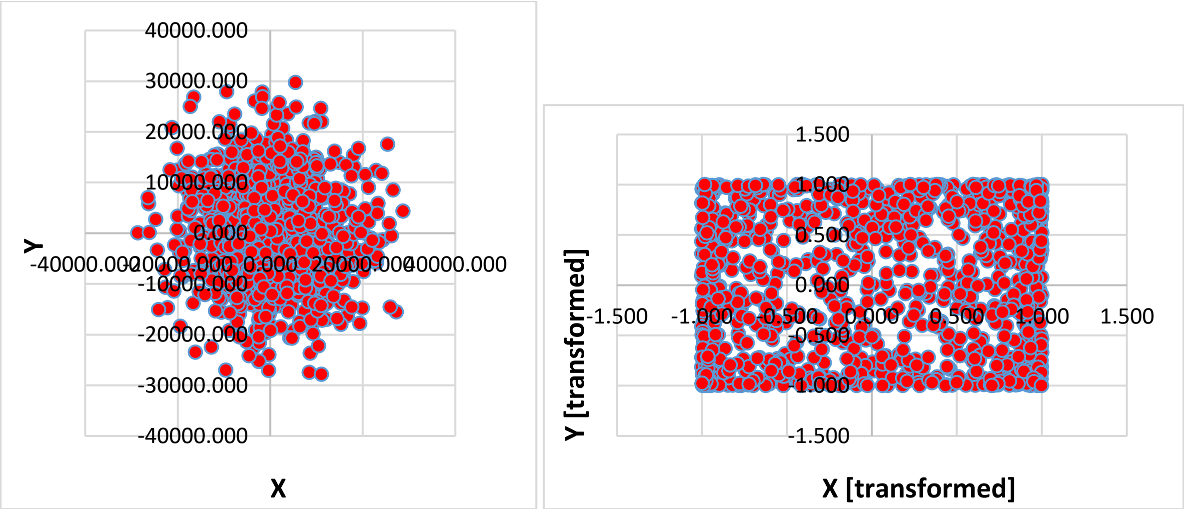

Figure (1) The "Centrifugal" Effect of the Sinusoidal Transformation: Pre and Post-Optimization [Trial 7].

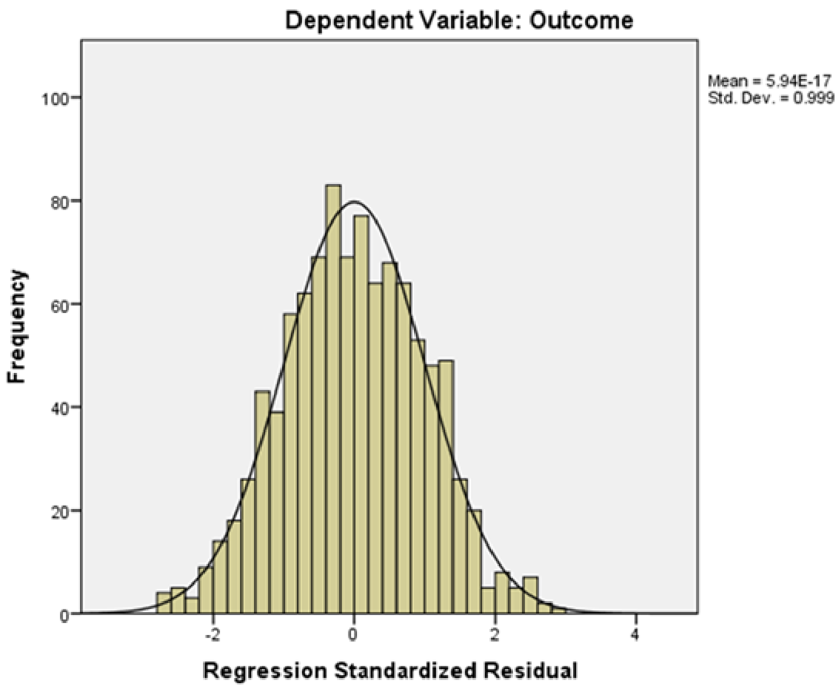

Figure (2) Frequency Distribution of the Linear Model's Residuals [Trial 7].

Discussion:

We reviewed the literature on the 20th of September 2019 (Figure 3). Based on the combination of thematic keywords search, there were 55,288 publications. Those indexed in the Cochrane Library were 117 (0.21%), in the PubMed, there were 40 publications (0.07%), while those indexed in the Embase were 55,131 publications (99.71%). Following a full-text retrieval of papers of interest, only fifteen publications (0.03%) indexed in the national library of medicine were found relevant to the primary objective of the current study. However, none of these studies implemented a sinusoidal data transform for boosting linear regression models. Since the last decade, there have been several attempts in the existing peer-reviewed literature to implement linear models as well as other machine-learning methods in combination with data transform function, including logistic regression, regression trees and Fourier transform, logistic regression with Log10 transformation, logistic regression with Ln transformation, multiple linear regression with log10 transformation, cycling regression model with Fourier transform, proportional hazards Cox regression model, time-series analytics regression with Fourier transform, logistic regression with square root and log10 transformation and proportional hazards model in combination with logistic regression [34-36]. Machine learning relies upon the analyses of big data using a plethora of well-established techniques of mathematical and data science models, including artificial neural networks, regression analysis, and classification trees [17]. Artificial intelligence techniques attempt to reach the lowest possible error of mathematically interpreted predictions for causality associations [19]. Machine learning is mandatory for unwitnessed benefits when it comes to applications related to spatio-temporal description and prediction of phenomena of interest, including epidemiological and digital epidemiological investigations [17, 19]. The infrastructure of big data upon which machine learning algorithms operate is the same as those for classical epidemiology and digital epidemiological research [37]. Researchers can retrieve data using survey tools, internet snapshots, longitudinal studies, cross-sectional studies, in addition to analyses of web-based social networks and electronic commerce website analytics of the surface web as well as the deep web, including the infamous Darknet hypermarket [37-40]. The main limitation of our study is the implementation of a limited number of simulations, and the fact that we based these simulations on random number generators rather than factual data. Besides, the sinusoidal transform function may introduce some degree of deformation of data. Nevertheless, our sinusoidal transformation model was successful. The model applies to anticipated high-impact research that requires linear model analyses, including anatomical sciences, dermatology as well as medical research and practice, as in the case of psychoactive and novel psychoactive substances research [41-45]. Optimized regression analytics is precious when it comes to applications of big data analytics and bioinformatics, comprehensive genomic analyses, and analytics based on extracting information from open-source deposits of big data, for instance, Google Trends and Google Analytics databases. Optimum linear models not only will reinforce the hypothesis testing for inferences that are more powerful but will also lessen the computational processing power and the human resources allocated for demanding real-time and predictive analyses. When integrated with the anticipated quantum computing, the benefits will be monumental concerning the precision of analytics and the efficacy of computational processing.

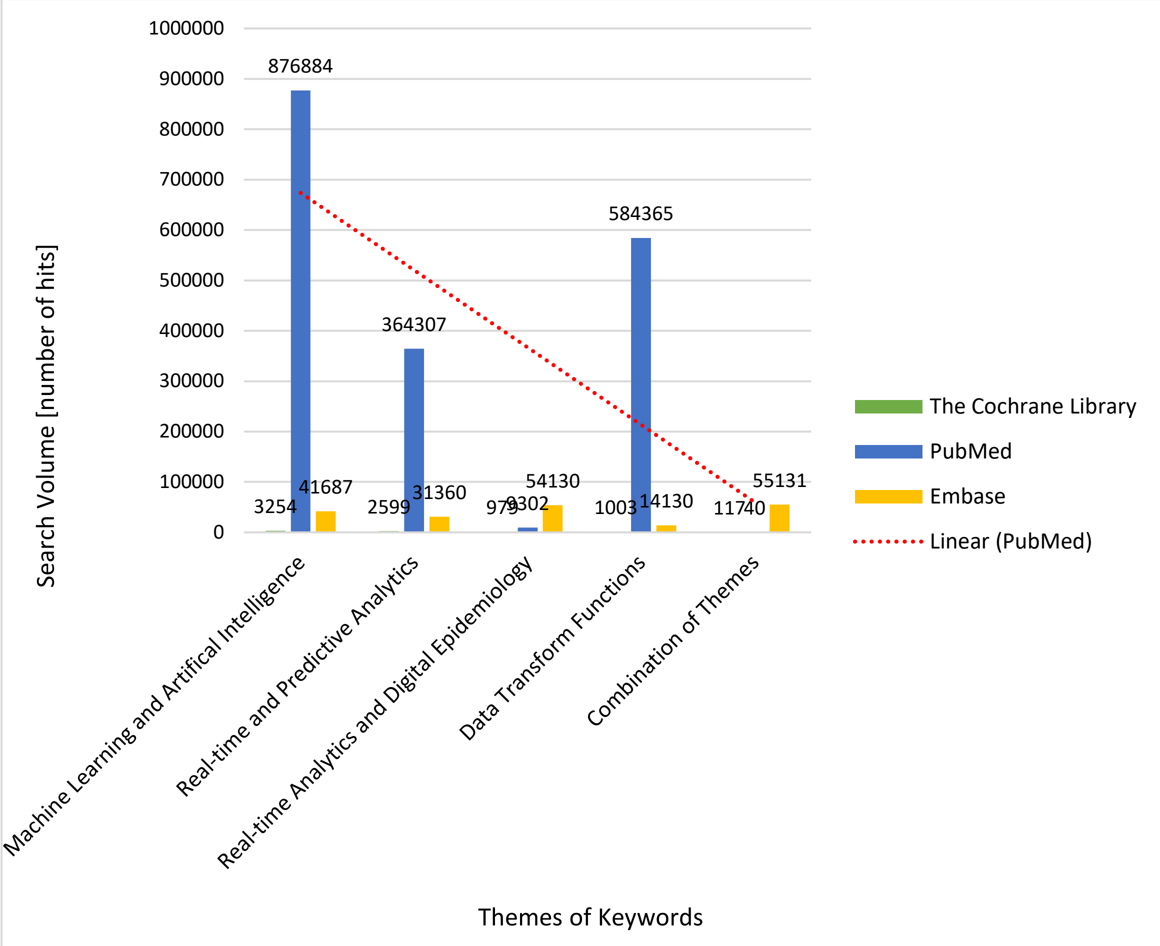

Figure (3) Keywords-Based Systematic Review of the Databases of Literature.

Conclusion:

The sinusoidal optimization of linear models serves three primary purposes: 1) Reducing the total sum of squared errors (SSE), which will provide a better line of best fit, i.e., a "trend" line. 2) The sinusoidal transformation will scale down any variable to the range of -1 to 1, thereby significantly reducing the computational processing demands for mathematical calculations for big data with an extensive list of variables as well as an extended number of observations within a variable, that is tangible in multiple regression analyses. 3) Real-time processing of correlations and regression among exhaustive multidimensional arrays of data will be more demanding in terms of the requirement for computational processing power that can burden supercomputers. The optimization will transform all variables to be into a narrower range, with limited decimal places, which is “economical” for mathematical and computational processing.

Acknowledgments: The authors would like to thank Rand R. Wilcox, professor of statistics and psychology at the University of Southern California, for his insightful remarks concerning our optimization method for linear models.

Availability of data: Our data are available upon request from the corresponding author.

Conflict of interest: The authors declare that they have no conflict of interest.

Source of funding: The authors self-funded this study.

References:

1. *Box GE. Science and statistics. Journal of the American Statistical Association 1976; 71(356): 791-9.*
2. *Field EH. All models are wrong, but some are useful. Seismological Research Letters 2015; 76(2A): 291-293.*
3. *Berro J. “Essentially, all models are wrong, but some are useful”—a cross-disciplinary agenda for building useful models in cell biology and biophysics. Biophysical Reviews 2018; 10(6): 1637-47.*
4. *Dawes RM, Corrigan B. Linear models in decision making. Psychological Bulletin. 1974; 81(2): 95.*
5. *Searle SR, Gruber MH. Linear models. New York: Wiley; 1971.*

6. Grizzle JE, Starmer CF, Koch GG. Analysis of categorical data by linear models. Biometrics 1969; 25(3): 489-504.
7. Cohen J. Statistical power analysis. Current Directions in Psychological Science 1992; 1(3): 98-101.
8. Hlaváč LM, Krajcarz D, Hlaváčová IM, Spadło S. Precision comparison of analytical and statistical-regression models for AWJ cutting. Precision Engineering 2017; 50: 148-59.
9. Hald A. A History of Mathematical Statistics from 1750 to 1930. Wiley; 1998.
10. Edwards AW. Mathematizing darwin. Behavioral ecology and sociobiology 2011; 65(3): 421-30.
11. Efron B. RA Fisher in the 21st century. Statistical Science 1998; 13(2): 95-114.
12. Norton BJ. Karl Pearson and statistics: The social origins of scientific innovation. Social studies of Science 1978; 8(1): 3-4.
13. Sedgwick P. Pearson's correlation coefficient. The British Medical Journal 2012; 345: e4483.
14. Freedman DA. Bootstrapping regression models. The Annals of Statistics 1981; 9(6): 1218-28.
15. Zou KH, Tuncali K, Silverman SG. Correlation and simple linear regression. Radiology 2003; 227(3): 617-28.
16. Godfrey K. Simple linear regression in medical research. In Medical uses of statistics NEJM Books, Boston; 1992. pp. 201-232.
17. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015; 349(6245): 255-60.
18. Schneider S. Science fiction and philosophy: from time travel to super-intelligence. John Wiley & Sons; 2016.
19. Everitt T, Goertzel B, Potapov A. Artificial general intelligence. Lecture Notes in Artificial Intelligence. Heidelberg: Springer; 2017.
20. Pueyo S. Growth, degrowth and the challenge of artificial super-intelligence. Journal of Cleaner Production 2018; 197: 1731-6.
21. Chevreuil M, Lebrun R, Nouy A, Rai P. A least-squares method for sparse low rank approximation of multivariate functions. SIAM/ASA Journal on Uncertainty Quantification 2015; 3(1): 897-921.
22. Van Dao T, Chaitusaney S, Nguyen HT. Linear least-squares method for conservation voltage reduction in distribution systems with photovoltaic inverters. IEEE Transactions on Smart Grid 2016; 8(3): 1252-63.
23. Al-Imam, A. Inferential Analysis of Big Data in Real-Time: One Giant Leap for Spatiotemporal Digital Epidemiology in Dentistry. Odontostomatology Research Anatomy Learning and Implantology 2019; 12(1): 1-14.
24. Al-Imam A. Monitoring and Analysis of Novel Psychoactive Substances in Trends Databases, Surface Web and the Deep Web, with Special Interest and Geo-Mapping of the Middle East. info:eu-repo/semantics/masterThesis [dissertation on the Internet]. United Kingdom: University of Hertfordshire; 2017. DOI: 10.13140/RG.2.2.27636.24961.
25. Bressoud DM. Historical Reflections on Teaching Trigonometry. Mathematics Teacher 2010; 104(2): 106-12.
26. Al-Imam A, Khalid U, Al-Hadithi N, Kaouche D. Real-time Inferential Analytics Based on Online Databases of Trends: A Breakthrough Within the Discipline of Digital Epidemiology of Dentistry and Oral-Maxillofacial Surgery. Modern Applied Science 2019; 13(2): 81-94.
27. Schilling M, Maiwald T, Bohl S, Kollmann M, Kreutz C, Timmer J, Klingmüller U. Computational processing and error reduction strategies for standardized quantitative data in biological networks. The FEBS Journal 2005; 272(24): 6400-11.
28. O'hara RB, Kotze DJ. Do not log-transform count data. Methods in Ecology and Evolution 2010; 1(2): 118-22.
29. Takeda M, Ina H, Kobayashi S. Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. Journal of the Optical Society of America 1982; 72(1): 156-60.
30. Phillips CV, Goodman KJ. The missed lessons of sir Austin Bradford Hill. Epidemiologic Perspectives and Innovations 2004; 1(1): 1-5.
31. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerging themes in Epidemiology 2015; 12(1): 1-9.
32. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? The British Medical Journal 2014; ; 348: 3725.
33. OCEBM Levels of Evidence. https://www.cebm.net/2016/05/ocebm-levels-of-evidence/ (accessed 2019).
34. Lorenz MW, Abdi NA, Scheckenbach F, Pflug A, Bülbül A, Catapano AL, Agewall S, Ezhov M, Bots ML, Kiechl S, Orth A. Automatic identification of variables in epidemiological datasets using logic regression. BMC Medical Informatics and Decision Making 2017; 17(1): 40.
35. Menotti A, Puddu PE, Lanti M. The estimate of cardiovascular risk. Theory, tools and problems. Annali Italiani di Medicina Interna: Organo Ufficiale Della Societa Italiana di Medicina Interna 2002; 17(2): 81-94.
36. Shaban-Nejad A, Michalowski M, Buckeridge DL. Health intelligence: How artificial intelligence transforms population and personalized health. npj Digital Medicine 2018; 1: 53.
37. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
38. Al-Imam A. Monitoring and Analysis of Novel Psychoactive Substances in Trends Databases,

Surface Web and the Deep Web, with Special Interest and Geo-Mapping of the Middle East. info:eu-repo/semantics/masterThesis [dissertation on the Internet]. United Kingdom: University of Hertfordshire; 2017. DOI: 10.13140/RG.2.2.27636.24961.

39. Motyka MA, Al-Imam A. Musical Preference and Drug Use among Youth: An Empirical Study. Research and Advances in Psychiatry 2019; 6(2): 50-57.

40. Al-Imam A, Al-Shalchi A. Ekbom's Delusional Parasitosis: A Systematic Review. Egyptian Journal of Dermatology and Venerology 2019; 39(1): 5-13.

41. Al-Imam A, Al-Lami F. Machine Learning for Potent Dermatology Research and Practice. Journal of Dermatology and Dermatologic Surgery. 2020; 24(1): 1-4 .

42. Al-Imam A, Sahai A, Al-Derzi AR, Al-Shalchy A, Abdullah F. "All Models Are Wrong, But Some Are Useful": On the Non-Bayesian Statistical Robustness of Hilton's Law. The European Journal of Anatomy. 2019; 24(1): 75-78.

43. Al-Imam A, Motyka MA. On the Necessity for Paradigm Shift in Psychoactive Substances Research: The Implementation of Machine Learning and Artificial Intelligence. Alcoholism and Drug Addiction/Alkoholizm i Narkomania. 2019; 32(3): 1-6.

44. Al-Imam A. Novel Psychoactive Substances Research: On the Necessity of Real-time Analytics and Predictive Modelling. Research and Advances in Psychiatry. 2020; 7(1): [in press].

45. *Al-Imam A. A Novel Method for Computationally Efficacious Linear and Polynomial Regression Analytics of Big Data in Medicine. Modern Applied Science. 2020 15(4): [in press].*

تحسين النماذج الخطية عن طريق التحويل الجيبي لتعلم الآلة المعزز في الطب

الدكتور أحمد محمد الامام *

الخلاصة:

خلفية البحث: يعتمد التعلم الآلي على مزيج من التحليلات ، بما في ذلك تحليلات الانحدار, والتراجع الخطي. لا توجد هناك هناك محاولات مسبقة لأستخدام المحولات الجيبية للبيانات لتعزيز نماذج الانحدار الخطي.

الأهداف: نحن نهدف إلى تحسين النماذج الخطية من خلال تطبيق التحويل الجيبي لتقليل العدد الإجمالي للمربعات للبيانات لتعزيز نماذج الانحدار الخطي.

المنهجية: قمنا بتطبيق إحصاءات غيربايزي باستخدام SPSS و MatLab تم استخدام Excel لإنشاء ثلاثين تجربة لنماذج الانحدار الخطي ، ولكل منها ألف ملاحظة (عينة). تم استخدام برنامج SPSS من أجل الانحدار الخطي ، واختبار ويلكوكسون ، وإحصائيات كرونباخ ألفا لتقييم أداء نموذج التحسين (التحويل الجيبي).

النتائج: كان التحويل الجيبي ناجحًا عن طريق تقليل إجمالي المربعات و بقيمة $P<0.001$ بشكل أحصائي ملحوظ. أكد اختبار كرونباخ ألفا الثبات الداخلي للنموذج المستخدم (معامل كرونباخ ألفا = 0.999)

الإستتنتاج: يعد نموذج التحويل الجيبي ذو أهمية في الأبحاث عالية التأثير التي تعتمد على الانحدار و التراجع الخطي. حيث يمكن أن تقلل من متطلبات المعالجة الحسابية لتحليلات قوية في الوقت الحقيقي والتنبؤات الأحصائية.

الكلمات المفتاحية: الذكاء الاصطناعي؛ تحويل البيانات؛ تعلم الآلة؛ التحليلات التنبؤية؛ تحليل الانحدار.