

**International Journal Devoted to All Aspect of
Health and Disease of Animals in the Tropics**

Published by the Faculty of Veterinary Medicine,
University, Ibadan (Nigeria)

Editor-in-Chief

M.O. Abatan

Associate Editors

G.A.T. Ogundipe

V.O. Taiwo

F.O. Olayemi

J.O. Olopade

Omolade Oladele

R.E. Antia

A.A. Adedapo

Editorial Advisory Board

T.A. Aire, St Georges, Grenada

R.O.A. Arowolo, Baeokuta

M.O. Olowookorun, Ibadan

J.B. Adeyanju, Ilorin

Helen Nottidge, Ibadan

Joru Chen, Atlanta, GA, USA

J.O. Ajanusi, Zaria

J.K.P. Kwaga, Zaria

F.O. Ayanwale, Tuskegee AL, USA

M.O. Makinde, Ouagadougou

O. Tomori, Ibadan

V. Shonekan, Nsukka

Gisela F. Erf, Fayetteville, AR, USA

E.O. Fayemi, Umudike

A.O. Oyekan, Houston, TX, USA

Business Manager

Comfort Aiki-Raji

Publication Data:

Subscription Rates:

Personal Subscription:

Institutional Subscription:

Airmail Postage:

**Single Issues and Back
volumes:**

Subscription Orders:

Tropical Veterinarian is published four times annually. Subscriptions run for a full calendar year. Prices are given per volume, along with postage rate. N1,000.00, US\$80.00 (must be in the name of, billed to, and paid by an individual. Order must be marked personal subscription).

N3,500.00, US\$120.00 (Regular rate).

N650.00; US\$60.00 (Extra per Volume).

Information on availability and prices can be obtained through the Business Manager.

Orders should be placed with the:

Business Manager,

Tropical Veterinarian,

Faculty of Veterinary Medicine,

University of Ibadan, Ibadan, Nigeria.

FAX: 02-8103043

E-mail: **tropivet@gmail.com**

tropivet@ui.edu.ng

WEBSITE: www.tropivet.com

ISSN 0794-4845

**CONSTRUING SIMPLE CORRELATION AND REGRESSION USING
STUDENTS' TEST SCORES AND THREE INDEPENDENT MATRICES
IN AN EPIDEMIOLOGY CLASS IN NIGERIA**

*Oludairo, O. O.¹, Aiyedun, J. O.¹, Daodu, O. B.², Kwaga, J. K. P.³

¹Department of Veterinary Public Health and Preventive Medicine, University of Ilorin, ²Department of Veterinary Microbiology, University of Ilorin, ³Department of Veterinary Public Health and Preventive Medicine, Ahmadu Bello University

*Corresponding author; Oludairo, O. O.

Department of Veterinary Public Health and Preventive Medicine, University of Ilorin

oludairo@hotmail.com, 234 7032131111

Key Words: Correlation, regression, scores, age, height, sex

Abstract

Correlation and regression are important statistical tools used in the analysis of relationships between continuous dependent and independent variables. The understanding of the basic principles and behind the scene calculations will therefore go a long way in the correct interpretation and application of this statistics. This study analysed the strength and nature of relationship between test scores and age, weight and height of students of Veterinary Public Health in a Veterinary school in Nigeria. Data were collected from the students at the time of writing a particular test, the data were thereafter used to calculate the Pearson correlation coefficient and to arrive at the regression equation using established formulae and calculating machines. Correlation and regression statistical analysis were used to analyse the data from the study. Other descriptive statistics were done using Fast Statistics version 2 and IBM SPSS Statistics 19. The three independent variables were directly related to the students' test scores. Weight and height had intermediate relationship while age is weakly related to test scores. The test scores were predicted for two randomly chosen extreme values of the 3 matrices used for the study. The predicted test scores indicated that the higher the weight, height and age of students the higher their test scores. The right application of correlation and regression could be a veritable tool in Veterinary research to statistically evaluate

the relationship between variables and predict dependent variables for the purpose of preparation, projection and planning.

Introduction

Correlation and regression are among the most commonly used methods of statistical analysis in epidemiology, science and social-science researches (Liang et al., 1993; Bewick et al., 2003; Maiwada and Okey, 2015, Kafle, 2019). The methods have gained application in epidemiology, including spatial and spatiotemporal epidemiology, where they have been used to evaluate diverse environmental variables to establish their relationship to the occurrence, distribution, control and prevention of specific diseases of public health importance (Olugasa et al., 2014; Bisht et al., 2022).

Correlation is a statistical technique used to determine the relationship between two variables (Tsumoto and Tsumoto, 2011; UFH, 2017). Scatter diagrams are usually utilized as a pointer to the type of relationship existing between the independent and dependent variables (Dendukuri and Reinhold, 2004; BUSPH, 2013). The relationship between variables could be either non-existent, negative or positive (Van Belle et al., 2004). This leads to the determination of the simple correlation coefficient 'r', which is otherwise known as Pearson correlation

coefficient or Product moment correlation coefficient (Rogers and Nicewander, 1998; Gideon, 2018). This statistic reveals the nature and strength between variables. (Rahbar, 2017). Positive sign indicates direct relationship implying an increase or decrease in one variable is associated with an increase or decrease in the other. Negative sign indicates indirect/inverse relationship which means increase in one of the variable is associated with decrease in the other one (Van Belle et al., 2004).

The value of 'r' varies from -1 to +1. It denotes a measure of the strength of the association between the two variables. The value of -1 means strong/perfect indirect correlation, 0 means no correlation, +1 implies strong/perfect direct correlation while -0.75 to -0.25, -0.25 to 0, +0.75 to +0.25 and +0.25 to 0 indicate intermediate indirect correlation, weak indirect correlation, intermediate direct correlation and weak direct correlation respectively (Le, 2003, Schober et al., 2018).

Regression is the statistical technique used to predict some variables when others are known (Dean and Illowsky, 2013). When independent variable 'x' is known dependent variable 'y' can be

predicted by using the regression equation [Starfield et al., 1991; UFH, 2017; Kaengthong and Domthong, 2017).

Age, weight and height are three variables possessed by every individual including students who are bound to write tests and examinations at one time or the other in schools. Do these three matrices have relationship with students' test score? This study sought to determine the relationship between these 3 variables and students' performance among veterinary students in a Nigerian North Central University, and to predict likely test scores at certain age, weight and height using correlation and regression analysis.

Materials and Methods

Midyear Veterinary students were randomly selected for the study. Measuring scale and stadiometer were provided for the measurement of these variables under the supervision of the researchers. Forms were passed round to be filled by selected members of the class to contain their age, weight and height as at the time a particular test was to be written. Returned forms were carefully reviewed, cases with missing values were excluded from the study. Seventeen fully filled cases were obtained (Table 1).

Table 1: The age, weight, height and test scores of students

STUDENT'S SERIAL NUMBER	AGE	WEIGHT	HEIGHT	SCORE
1	21	78	1.8	50
2	21	64	1.63	45.5
3	23	64	1.62	19
4	22	46	1.62	20
5	20	70	1.8	40
6	20	50	1.64	25
7	19	47	1.64	33
8	22	69	1.7	27.5
9	21	65	1.8	47.5
10	18	67	1.89	42
11	21	45	1.66	12
12	22	47	1.63	21.5
13	21	52	1.7	29
14	27	62	1.71	50

15	20	49	1.64	47.5
16	20	62	1.8	16
17	21	48	1.5	37.5

The data mined, together with the scores of the test written at that time were imputed into Fast Statistics software package 2.0 for descriptive statistics while other calculations were

done with the aid of the calculator. Scatter diagrams (Figure 1-3) and integer numbers were used to present the results.

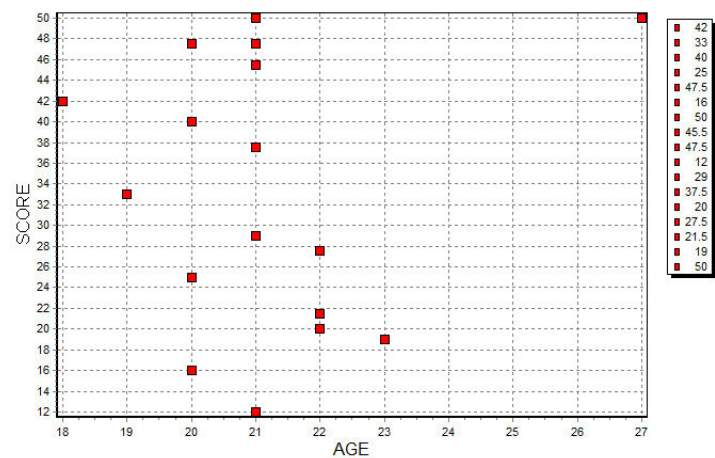


Figure 1: Scatter diagram of age/test score

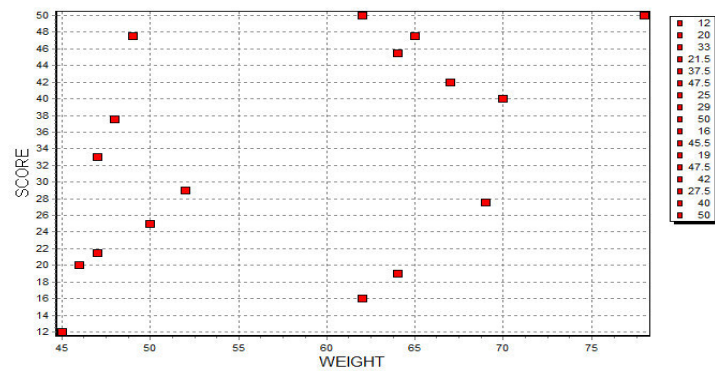


Figure 2: Scatter diagram of weight/test score

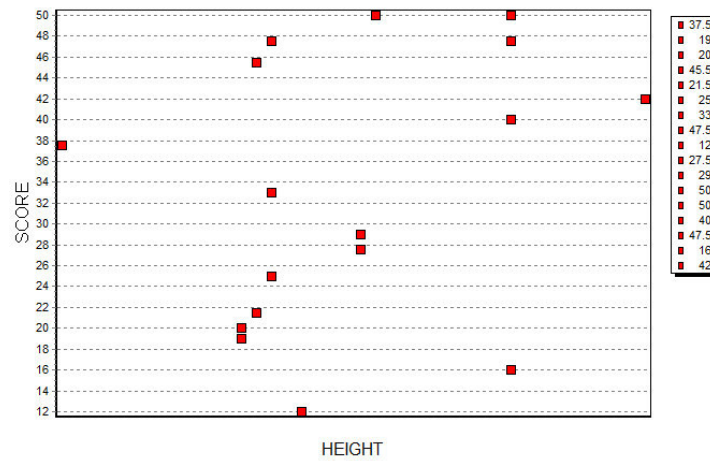


Figure 3: Scatter diagram of height/test score

Test scores were predicted for students with age 15 and 40 years, weight 30 and 90 kg and height 1 and 2m (Figure 4-6).

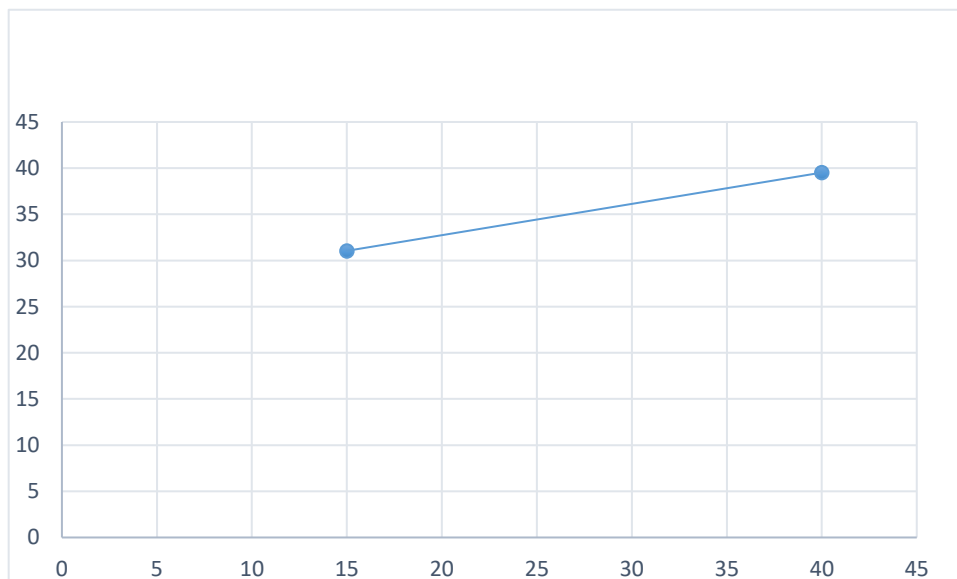


Figure 4: Predicted scores of students of age 15 and 40 years.

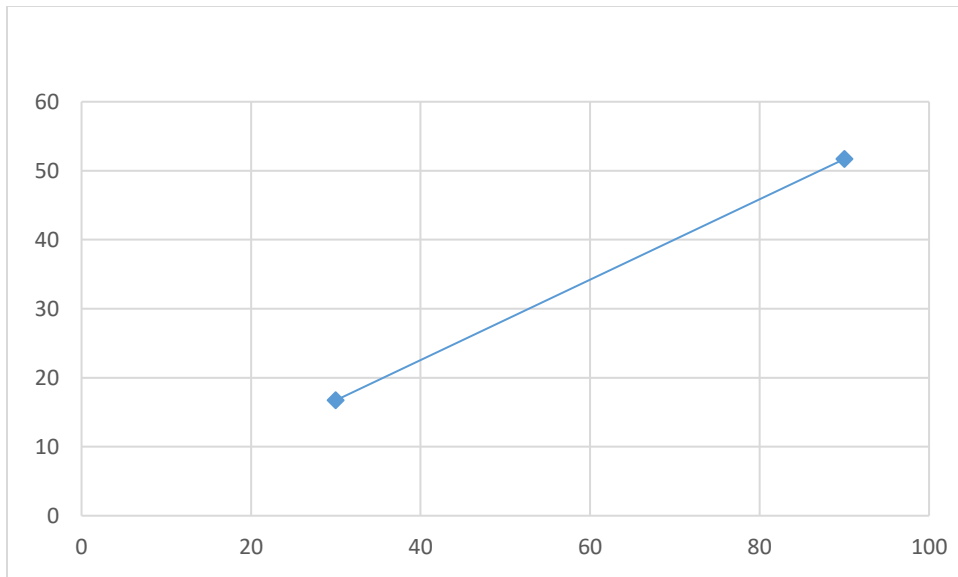


Figure 5: Predicted scores of students with weight 30 and 90Kg

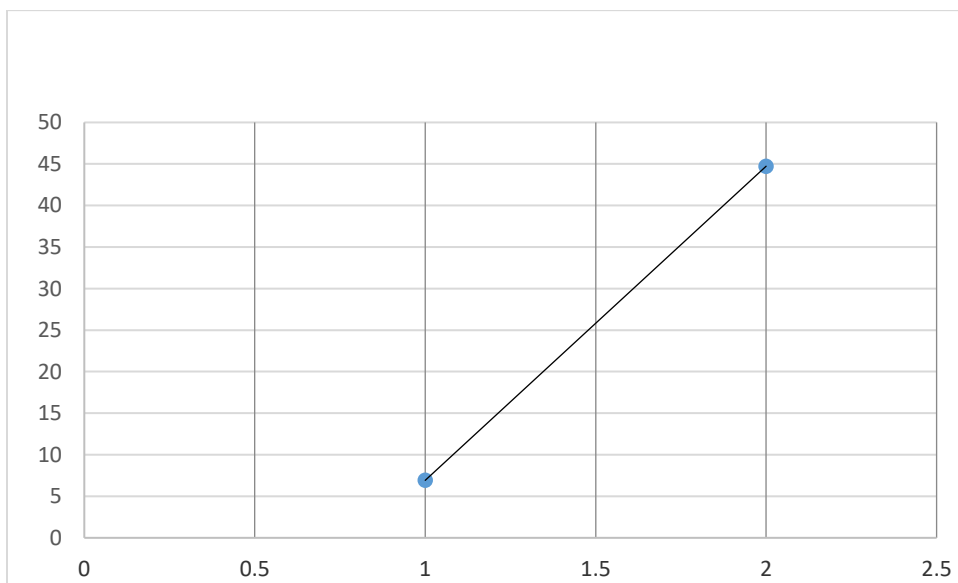


Figure 6: Predicted scores of students at height 1 and 2m.

Results

Determination of Pearson's correlation coefficient,

Since,

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

Score Vs Age,

$$r = \frac{11909.5 - \frac{359 \times 563}{17}}{\sqrt{\left[7641 - \frac{(359)^2}{17}\right] \left[21287.5 - \frac{(563)^2}{17}\right]}}$$

$r = +0.051$.

Score Vs Weight,

$$r = \frac{33626 - \frac{985 \times 563}{17}}{\sqrt{\left[58807 - \frac{(985)^2}{17}\right] \left[21287.5 - \frac{(563)^2}{17}\right]}}$$

$r = +0.46945$.

Score and Height,

$$r = \frac{958.81 - \frac{28.78 \times 563}{17}}{\sqrt{\left[48.8732 - \frac{(28.78)^2}{17}\right] \left[21287.5 - \frac{(563)^2}{17}\right]}}$$

$r = +0.28518$.

Regression

To obtain the regression equation,

$y = a + bx$ was used.

$$\text{Where, } b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$a = \bar{y} - b\bar{x}$$

x = known independent variable for which y is to be predicted (Dean and Illowsky, 2013).

For age, to predict test score at age 15 and 40, using the equation, $y = a + bx$.

$$b = \frac{11909.5 - \frac{359 \times 563}{17}}{7641 - \frac{(359)^2}{17}} = 0.339$$

$$a = \frac{563}{17} - (0.339 \times \frac{359}{17}) = 25.959$$

The regression equation therefore is

$$y = 25.959 + 0.339x$$

At age 15 and 40 therefore, predicted test scores will be 31.044 and 39.519.

For weight, to predict test scores at 30 Kg and 90 kg.

$$b = \frac{33626 - \frac{985 \times 563}{17}}{58807 - \frac{(985)^2}{17}} = 0.5793$$

$$a = \frac{563}{17} - (0.5793 \times \frac{985}{17}) = -0.4477$$

The regression equation therefore is:

$$y = -0.4477 + 0.5793x$$

With weight 30 Kg and 90 Kg, the predicted test scores will be 16.7313 and 51.6893.

For height, to predict test scores at 1 m and 2 m.

$$b = \frac{958.81 - \frac{28.78 \times 563}{17}}{48.8732 - \frac{(28.78)^2}{17}} = 37.8052$$

$$a = \frac{563}{17} - (37.8052 \times \frac{28.78}{17}) = -30.8843$$

The regression equation therefore is

$$y = -30.8843 + 37.8052x$$

With height 1 m and 2 m, the predicted test scores will be 6.9209 and 44.7261.

Discussion and Conclusion

From the scatter diagram, it is possible to draw a wrong inference in relation to the relationship between the variables considered, especially with age. This is because apparently the diagram does not exhibit distinct correlation features. The calculation of the product moment correlation coefficient, statistically establish the actual relationship

between these variables. The positive values obtained for the three independent variables in comparison with the dependent variable confirmed the direct relationship between all the matrices and students' test scores. As observed from the scatter diagram however, the strength of the relationship that existed between age and test score was weak while that of the other two matrices were intermediate. Increase or decrease in any of the independent variables should lead to corresponding increase or decrease of the test scores in a weak and intermediate way.

The manual calculations with formulae and statistical packages, laid bare the basic principles behind the results obtained when statistical packages are used in the analysis of data with step by step analysis of the procedure employed for better interpretation, understanding and inference making from the results obtained.

Predictions were made for test scores at 15 and 40 years, 30 and 90 Kg weights including 1 and 2m heights. The test score prediction results of 31.0 and 39.5, 16.7 and 51.7 with 6.9 and 44.7 further laid credence to the correlation statistics of direct proportionality. Informed decisions and policy statements can be made from the statistics results on the age and height

of students that will possibly perform better in future test scores. Students can also be advised on the weight to maintain for maximal academic performance.

Other published work have compared other parameters. Shankar et al. carried out a correlation studies of human height with feet length in Indian individuals. Height reportedly had a statistically significant positive correlation ($r = +0.196$) with left foot length in both male and female. Linear regression equation was also derived to predict height from feet length in both male and female (Gogtay et al., 2017; Shankar et al., 2018).

Shaban in the analysis of the correlation and regression coefficients of the interaction between yield and some parameters of snap beans plants reported strong positive relationship between the parameters tested ($r = +0.839$, $+0.774$ and $+0.996$) except yield/weed interaction and snap bean plant ($r = -0.627$) (Shaban, 2005).

In a study to determine the causal relationship model between variables using linear regression to improve professional commitment of 135 lecturers, intermediate positive relationships were reported ($r = +0.663$) (Setyaningsih, 2017).

Positive relationship were reported between loneliness, depression and internet addiction (r ranged between $+0.273$ and $+0.396$) while regression analysis predicted depression and internet addiction (Demir and Kutlu, 2017).

Kumar et al. carried out correlation and regression studies in mango trees (*Mangifera indica L.*) to establish a relationship between growth, leaf nutrient content, fruit yield, and quality attributes analyses indicated variability in fruit yield due to leaf nutrient contents and tree volume to the extent of 66–73% and 35–68% in high and low productive orchards, respectively (Kumar et al., 2013).

Oludairo et al. reported positive strong relationship between the temperature and pH of water samples from 3 local government in Kwara State and predicted the pH of water at 6 different temperatures using regression equation line (Oludairo et al., 2020).

Other researchers reported the use of correlation and regression in the comparison of different matrices and prediction making (Bailey et al., 2000; Knight et al., 2006; Nagata et al., 2009; Oren et al., 2014; Zhao and Li, 2014; McCambridge et al., 2014; Haberman et al., 2015; Carrol and Schuster, 2015).

The regression correlation analysis was performed to investigate the effect of rainfall on surface air pollutants. The concentration of air pollutants are found to decrease with increase of rainfall. Most of the pollutants in this study showed statistically significant negative correlations between their concentrations and rainfall (Sharma and Sharma, 2018).

Linear regression methods try to determine the best linear relationship between data points while correlation coefficients assess the association (as opposed to agreement) between the two methods. Linear regression and correlation play an important part in the interpretation of quantitative method comparison studies. Their major strength is that they are widely known and as a result both are employed in the vast majority of method comparison studies (Twoney and Krol, 2008).

Correlation and regression analyses have been used by researchers in the determination of relationships between variables in addition to prediction of the values of 'y' variable based on known of fixed values of the explanatory variables (Fechete and Nedelcu, 2014).

Regression analysis have been particularly found to be of immense importance in epidemiological studies (Rigby, 2000; Bhaskaran et al., 2013).

Correlation and regression are credible tools for use in making projections and informed decisions and predictions. The statistic could be used to scientifically establish the nature and strength of the relationship between variables. They give reliable statistical figures on which present and future actions can be taken. This is especially so in epidemiological researches where the nature and strength of different variables may need to be determined and where large amount of data may be involved. Different probable scenarios could also be simulated with regression statistics to arrive at figures that could be useful in future planning, budgeting, training and other level of preparedness especially in the area of disease prevention.

References

- Bailey, J. E., Yackle, K. A., Yuen, M. T. and Voorhees, L. I. (2000). Preoptometry and Optometry School Grade Point Average and Optometry Admissions Test Scores as Predictors of Performance on the National Board of Examiners in Optometry Part I (Basic Science) Examination. *Optometry and Vision Science*. 77: 188-193.

- Bewick, V., Cheek, L. and Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical Care*, 7:451-459.
- Bhaskaran, K., Gasparrini, A, Hajat, S., Smeeth, L. and Armstrong, B. (2013). Time series regression studies in environmental epidemiology. *International Journal of Epidemiology*, 42: 1187–1195.
- Bisht, L., Gupta, V., Singh, A., Gautam, A. S., Gautam, S. (2022). Heavy metal concentration and its distribution analysis in urban road dust: A case study from most populated city of Indian state of Uttarakhand. *Spatial and Spatio-temporal Epidemiology*, 40:100470. <https://doi.org/10.1016/j.sste.2021.100470>.
- BUSPH, Boston University School of Public Health (2013). Introduction to correlation and regression analysis. Available at www.sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Correlation-Regression Accessed on 20th November, 2017.
- Carroll, A. M. and Schuster, G. M. (2015). Correlation between Students' Dental Admission Test Scores and Performance on a Dental School's Competency Exam. *Journal of Dental Education*, 79: 11325-1329.
- Dean, S. and Illowsky, B. (2013). Introductory Statistics, 2nd Ed. Adapted by Beintema, M. and Casper, N. of Openstax College. Available at <http://www.clcillinois.edu/programs/mth/introductorystatistics2ndedition.pdf> Accessed 17th November, 2017.
- Demir, Y. and Kutlu, M. (2016). The Relationship between Loneliness and Depression: Mediation Role of Internet Addiction. *Educational Process: International Journal*, 5: 97-105.
- Dendukuri, N. and Reinhold, C. (2004). Correlation and Regression. *American Journal of Roentgenology*, 185:3–18.
- Fechete, F. and Nedelcu, A. (2014). Analysis of the economic performance of an organization using multiple regression. International Conference of Scientific Paper (AFASES). Available online at http://www.afahc.ro/ro/afases/2014/manag/Fechete_Nedelcu_b

- [.pdf](#). Accessed 22nd February, 2018.
- Gideon, R. A. (2018). Correlation estimation system minimization compared to least squares minimization in simple linear regression. Available on line at <https://arxiv.org/pdf/1802.02649.pdf>. Accessed 13th April, 2018.
- Gogtay, N. J., Deshpande, S. P. and Thatte, U. M. (2017). Principles of Regression Analysis. *Journal of the Association of Physicians of India*, 65: 48-52.
- Haberman, S. J., Yao, L. and Sinharay, S. (2015). Prediction of true test scores from observed item scores and ancillary data. *British Journal of Mathematics and Statistics Psychology*, 68: 363-385.
- Kaengthong, N. and Domthong, U. (2017). Modified Regression Correlation Coefficient for Poisson Regression Model. *Journal of Physics: Conference Series*, 890: doi :10.1088/1742-6596/890/1/012155.
- Kafle, S. C. (2019). Correlation and Regression Analysis Using SPSS. *OCEM Journal of Management, Technology and Social Sciences*. 126-132. Available at https://www.academia.edu/43736232/Correlation_and_Regression_Analysis_Using_SPSS. Accessed February 1st 2022.
- Knight, R. G., McMahon, J., Green, T. J., Skeaff, C. M. (2006). Regression equations for predicting scores of persons over 65 on the rey auditory verbal learning test, the minimal state examination, the trail making test and semantic fluency measures. *British Journal of Clinical Psychology*, 45: 393-402.
- Kumar, P., Sharma, S. D. and Yadav, S. K. (2013). Correlation and regression studies in mango (*Mangifera indica* L.), *Journal of Plant Nutrition*, 36; 929-947.
- Le, C. T. (2003). *Introductory Biostatistics*. Published by John Wiley and Sons Inc. Hoboken, New Jersey, DOI: 10.1002/0471308889.
- Liang, K. and Zenger, S. L. (1993). Regression analysis for correlated data. *Annual Review of Public Health*, 14: 43-68.
- Maiwada, S. and Okey, L. E. (2015). The relevance and significance

- of correlation in social science research. *International Journal of Sociology and Anthropology Research*, 1: 22 – 28.
- McCambridge, J., Kypri, K. and McElduff, P. (2014). Regression to the mean and alcohol consumption: a cohort study exploring implications for the interpretation of change in control groups in brief intervention trials. *Drug and Alcohol Dependence*, 135: 156–159.
- Nagata, T., Ishii, K., Ito, T., Aoki, K., Ehara, Y., Kada, H., Furukawa, H., Tsumura, M., Shinagawa, S., Kasahara, H. and Nakayama, K. (2009). Correlation between a reduction in Frontal Assessment Battery scores and delusional thoughts in patients with Alzheimer's disease. *Psychiatry and Clinical Neurosciences*, 63: 449–454.
- Oludairo, O. O., Aiyedun, J. O., Olorunshola, I. D., Bale, J. O. O. and Akintola, O. O. (2020). Physical Properties, Correlation and Regression Analyses of Potable Water in Ilorin, Nigeria. *Journal of Research in Forestry, Wildlife and Environment*, 12(2): 225 – 230.
- Olugasa, B. O., Dogba, J. B., Ogunro, B., Odigie, E. A., Nykoi, J., Ojo, J. F., Taiwo, O., Kamara, A., Mulbah, C. K., Fasunla, A. J. (2014). The rubber plantation environment and Lassa fever epidemics in Liberia, 2008-2012: a spatial regression. *Spatial and Spatio-temporal Epidemiology*, 11:163-74. doi: 10.1016/j.sste.2014.04.005. <https://doi.org/10.1016/j.sste.2014.04.005>.
- Oren, C., Kennet-Cohen, T., Turvall, E. and Allalouf, A. (2014). Demonstrating the validity of three general scores of PET in predicting higher education achievement in Israel. *Psicothema*, 26: 117-126.
- Rahbar, M. H. (2017). Correlation and regression. Available at www.pitt.edu/~super7/19011-20001/19051 Accessed on May, 2017.
- Rigby, A. S. (2000). Statistical methods in epidemiology VI; Correlation and regression: the same or different? *Disability and Rehabilitation*, 22: 813-819.
- Rodgers, J. L. and Nicewander, W. A. (1998). Thirteen Ways to Look at the Correlation Coefficient.

- The American Statistician*, 42: 59-66.
- Schober, P., Boer, C. and Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, 126. 1. 10.1213/ANE.00000000000002864.
- Setyaningsih, S. (2017). Causal relationship model between variables using linear regression to improve professional commitment of lecturer. *IOP Conference Series: Materials Science and Engineering*, 166: doi:10.1088/1757-899X/166/1/012004 012004.
- Shaban, N. (2005). Analysis of the Correlation and Regression Coefficients of the Interaction between yield and some parameters of snap beans plants. *Trakia Journal of Sciences*, 3: 27-31.
- Shankar, G. S., Shankar, V. V., Shetty, S. and Radhika, K. (2018). Correlation of human height with foot length in Indian individuals. *International Journal of Anatomy and Research*, 6: 4943-46.
- Sharma, R. C. and Sharma, N. (2018). Statistical Investigation of Effect of Rainfall on Air Pollutants in the Atmosphere, Haryana State, Northern India. *American Journal of Environmental Protection*, 6: 14-21.
- Starfield, B., Shapiro, S., Weiss, J., Liang, K. Y. and Ra, K. (1991). Race, family, income and low birthweight. *American Journal of Epidemiology*, 134: 1 1 67-74.
- Tsumoto, Y. and Tsumoto, S. (2011). Correlation and Regression Analysis for Characterizations of a University Hospital. *The Review of Socionetwork Strategies*, 5: 43–55.
- Twomey P. J. and Krol M. H. (2008): How to use linear regression and correlation in quantitative method comparison studies. *International Journal of Clinical Practice*, 62: 529-538.
- UFH, University of Florida Health (2017). Linear relationships – correlation, regression: In Biostatistics open learning textbook. Available at <http://oli.cmu.edu>. Accessed 17th November, 2017.

- Van Belle, G., Fisher, L. D., Heagerty, P. J. and Lumley, T. (2004). *Biostatistics- A methodology for the Health Sciences*. 2nd Ed. Published by John Wiley and Sons, Inc., Hoboken, New Jersey.
- Zhao, S. D. and Li, Y. (2014). Score test variable screening. *Biometrics*, 70: 862-871.