

Quantitative Data Standardization and Normalization Techniques for Variability Reduction and Parametric Distribution

Mutifa Aslam*

Department of Biostatistics, Science and Technology of New York, New York, USA

Introduction

Normalization is the process of transforming data to have a common scale or range, usually between 0 and 1, by adjusting the data's values based on certain statistics. This process is used to eliminate the effects of gross influences or to compare different datasets with heterogenic data. The decimal scaling method is a normalization technique that involves moving the decimal point of the data's values. This method divides each data value by the maximum absolute value to normalize the data. This technique results in a scaled version of the data that retains the original data's distribution and shape. The minimum-maximum (Min-Max) data normalization method is a linear transformation of the original data to a common scale. This method subtracts the minimum value of the data and divides the result by the range of the data, which is the difference between the maximum and minimum values. This technique also results in a scaled version of the data that preserves the original distribution and shape [1].

Description

The z-score data normalization procedure standardizes the data by subtracting the mean and dividing by the standard deviation. This method transforms the data into a standard normal distribution with a mean of 0 and a standard deviation. Overall, both normalization and standardization procedures are used to transform data to a common scale, but they differ in terms of the statistical parameters used and their ability to reduce data variability. Normalization techniques preserve the original data's distribution and shape, whereas standardization techniques transform the data into a standard normal distribution, making it easier to compare data across different scales. Quantitative data used for the present study were drawn from previous experiments as described. Briefly, collected data included four growth parameters (diameter, plant height, leaf length and leaf number) of two maize varieties, treated by both rhizobacteria and foliar bio-fertilizing. The same survey displayed smaller bias transformation by using the Box-Cox transformation as opposite to logarithm transformation. The same study revealed that the mean squared error of estimation is smaller with the Box-Cox transformation; and as well, the Box-Cox transformation leads to systematically higher estimated values than Logarithmic transformation. Hence, the Box-Cox transformation should be considered as a viable alternative in statistical modelling if the transformation of variables is required. Low aptitude with regard Exponential and Inverse data transformation in reducing data variability as well as in adjusting data normality could be due to processed positive value of analysed data. Indeed, our analysis suspected Exponential data transformation as a potential source of transformed data variability [2,3].

Further, collected data for each treatment were summarized in a matrix including four columns describing variables parameters (two maize varieties growth parameters) and ninety-six rows corresponding to the observation number. Next, we submitted the above-mentioned data matrix to Box-Cox, Logarithm, Square Root, Inverse and Z-score, Minimum, Exponential and Minimum-Maximum quantitative data standardization as well as normalization

Address for Correspondence: Mutifa Aslam, Department of Biostatistics, Science and Technology of New York, New York, USA, E-mail: aslam93@edu.in

Copyright: © 2023 Aslam M. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: 27 March, 2023, Manuscript No. Jbmbms-23-95425; **Editor assigned:** 29 March, 2023, Pre QC No. P-95425; **Reviewed:** 12 April, 2023, QC No. Q-95425; **Revised:** 17 April, 2023, Manuscript No. R-95425; **Published:** 25 April, 2023, DOI: 10.37421/2155-6180.2023.14.153

(data transformation) procedures. Biometric verification is a method for checking a person's personality by using a piece of their identity, like their finger impression, facial features, or iris design. These features contain unique information that can't be duplicated. Despite their numerous benefits, certain biometrics, particularly facial recognition, have recently come under fire for being an infringement on privacy. Considering everything, your "face print" is your information, and many people don't like the idea that their face prints could be used or shared without their consent. This may eliminate the obscurity that many people anticipate in open areas, such as online. Even the idea of "connecting" a person's face to yet another source of personal data has been floated [4,5].

Conclusion

Above-mentioned data transformation systems was applied to the same data matrix (collected data) generating a new data set for each standardization and/or normalization methods. The present study provided a systematic comparative study that highlighted difference as well as similitude between eight quantitative data standardization methodologies providing useful tool to researchers, in choosing adequately data transformation methodologies that well fitting for their investigations. We focused on eight quantitative data transformation systems in the present comparative study. Processed quantitative data standardization and/or normalization procedures are as following Box-Cox (Box), Exponential (Expo), Inverse, Logarithmic normalization, Maximum, Minimum-Maximum, Square Root and Z-score.

Acknowledgement

We thank the anonymous reviewers for their constructive criticisms of the manuscript. The support from ROMA (Research Optimization and recovery in the Manufacturing industry), of the Research Council of Norway is highly appreciated by the authors.

Conflict of Interest

The authors declare that there was no conflict of interest in the present study.

References

1. Stamler, Jeremiah, Rose Stamler and James D. Neaton. "Blood pressure, systolic and diastolic, and cardiovascular risks: US population data." *Arch Intern Med* 153 (1993): 598–615.
2. Lawes, Carlene MM, Stephen Vander Hoorn and Anthony Rodgers. "Global burden of blood-pressure-related disease, 2001." *JAMA* 371 (2008):1513–1518.
3. Carson, AP, Howard G, Burke GL and Carson April Pet al. "Ethnic differences in hypertension incidence among middle-aged and older adults: The multi-ethnic study of atherosclerosis." *Hypertension* 57 (2011):1101–1107.
4. Egan, BM, Zhao Y, Axon RN. and Egan Brent M. "US trends in prevalence, awareness, treatment, and control of hypertension, 1988-2008." *JAMA* 303 (2010): 2043–2050.
5. Hajjar, Ihab and Theodore A. Kotchen. "Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988-2000." *JAMA* 290 (2003): 199–206.

How to cite this article: Aslam, Mutifa. "Quantitative Data Standardization and Normalization Techniques for Variability Reduction and Parametric Distribution." *J Biom Biosta* 14 (2023): 153.