International Doctoral School
Medical University of Lodz

Joanna Kostanek
International Doctoral School, Medical University of Lodz

Atypical Distribution Analysis with Bootstrap

Statistical models of multiple comparisons for small samples using the bootstrap method in the analysis of selected epidemiological factors in cardiovascular diseases.

Introduction

The Gaussian distribution, commonly known as the normal distribution, serves as a fundamental concept in statistics. Assessing the shape of this distribution and its deviations from normality involves the examination of two critical parameters: skewness and kurtosis. Skewness measures the asymmetry of the distribution, assuming a value of 0 for symmetric distributions, a characteristic feature of the normal distribution. Positive values indicate right-skewness, while negative values suggest left-skewness. Kurtosis characterizes the peakedness of the distribution, taking a value of 0 for a Gaussian distribution. Values higher than 0 indicate distributions with sharper peaks, whereas negative values suggest flatter distributions [1]. Figures 1 and 2 visually illustrate the differences between distributions, emphasizing these two parameters. In fact, data in various fields rarely follow a normal distribution [2]. Therefore, it is crucial to find methods of data analysis that ensure the interpretation of results is as reliable and accurate as possible.

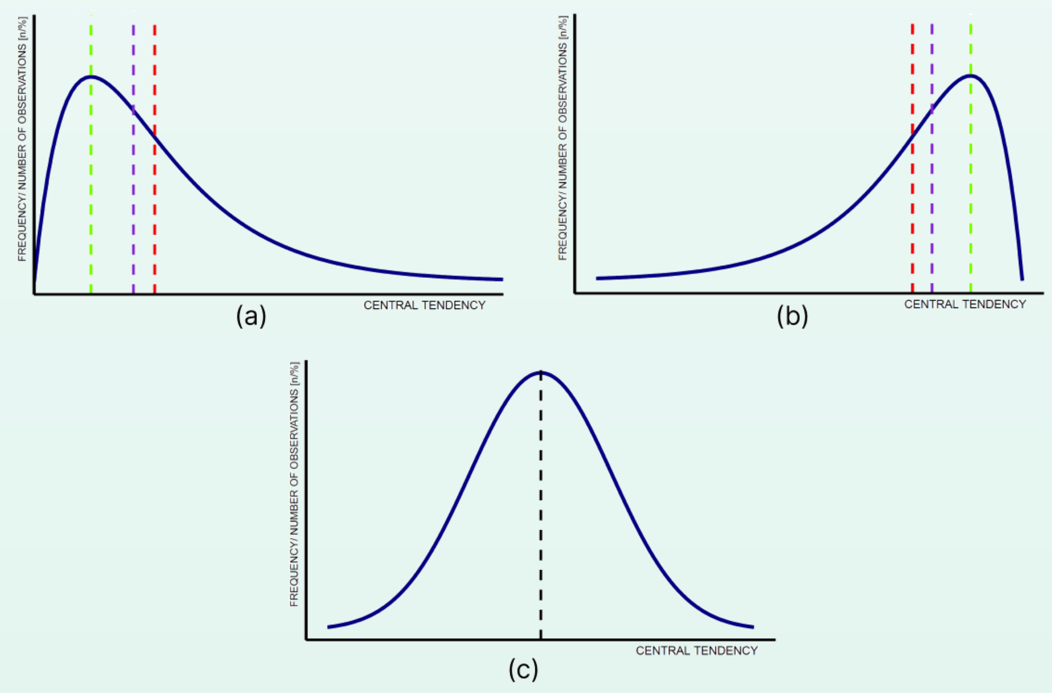

Fig. 1. Three distributions varying in skewness. a) A right-skewed distribution where $\bar{x}$>Me>Mo. The red line represents the mean ($\bar{x}$), the purple line indicates the median (Me), and the green line represents the mode (Mo). b) A left-skewed distribution where $\bar{x}$<Me<Mo. The red line represents the mean ($\bar{x}$), the purple line indicates the median (Me), and the green line represents the mode (Mo). c) symmetric distribution with no skewness, black line represents mean, median and mode which are equal $\bar{x}$=Me=Mo.

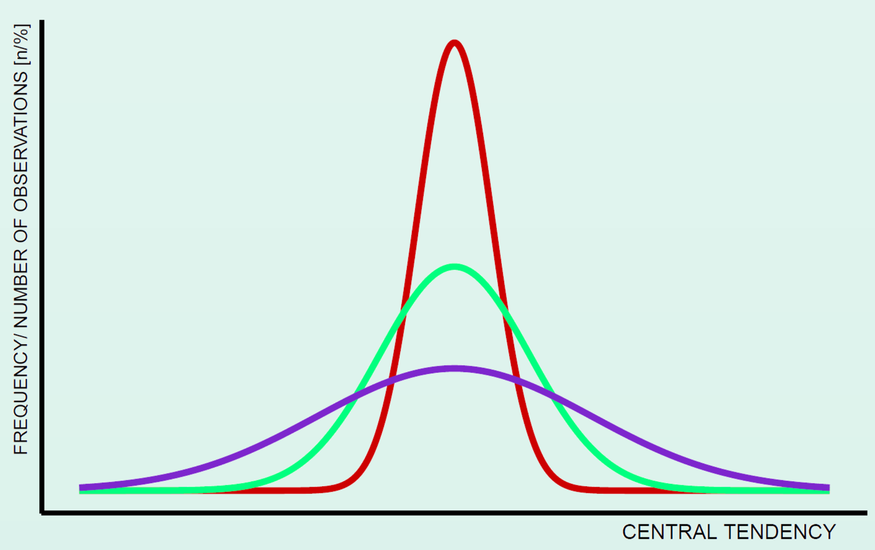

Fig. 2. Three distributions with varying kurtosis. Platykurtic (kurtosis < 0), leptokurtic (kurtosis > 0), and mesokurtic (kurtosis = 0). They were color-coded on the plot for clarity: platykurtic in purple, leptokurtic in red, and mesokurtic in green.

Materials and Methods

Data preparation

In the simulation study the anonymous data obtained from the Department and Clinic of Cardiology at the Medical University of Wroclaw were utilized. In our analysis we focused on a group of male patients (N=486), all diagnosed with cardiovascular disease. For the purpose of our analysis we focused on two variables: serum concentration of triglycerides (TG [mg/dl]) and serum concentration of High-Density Lipoprotein Cholesterol (HDL [mg/dl]).

Statistical Analysis

The kurtosis and skewness were estimated according to Brown [3] and Jar [4]. To compare means among groups, the classical one-way analysis of variance (ANOVA) was employed, followed by the post hoc tests of varying conservativeness (Tukey, Bonferroni, and LSD Fisher). Statistical significance was considered at the values of $P < 0.05$. In all estimations (descriptive statistics incl. Skewness and kurtosis, one-way ANOVA and post hoc tests) we utilized the software of STATIS-TICA (Dell Inc., 2016, Dell Statistica: data analysis software system, version 13, software.dell.com.) and R Platform Software (R-4.2.2 for Windows, with publicly available libraries and self-written scripts).

Results

Bootstrap Efficacy in Handling Skewness and Kurtosis Disparities under Homogeneous Variance Settings

The simulation was conducted to assess how the bootstrap method performs with distributions of specific shapes. The variable TG was examined in its original form (leptokurtic right-skewed distribution) and in the form of randomly generated datasets with specific characteristics (skewness and kurtosis). Datasets were generated to ensure homogeneous variances in each case. Tukey's test results revealed significant differences between Group 1 and 2 as well as between Group 1 and 3. Bootstrap-derived results predominantly exhibited higher P values. For some distributions, the bootstrap results did not show statistical significance in comparison to the classical approach (Figure 3).

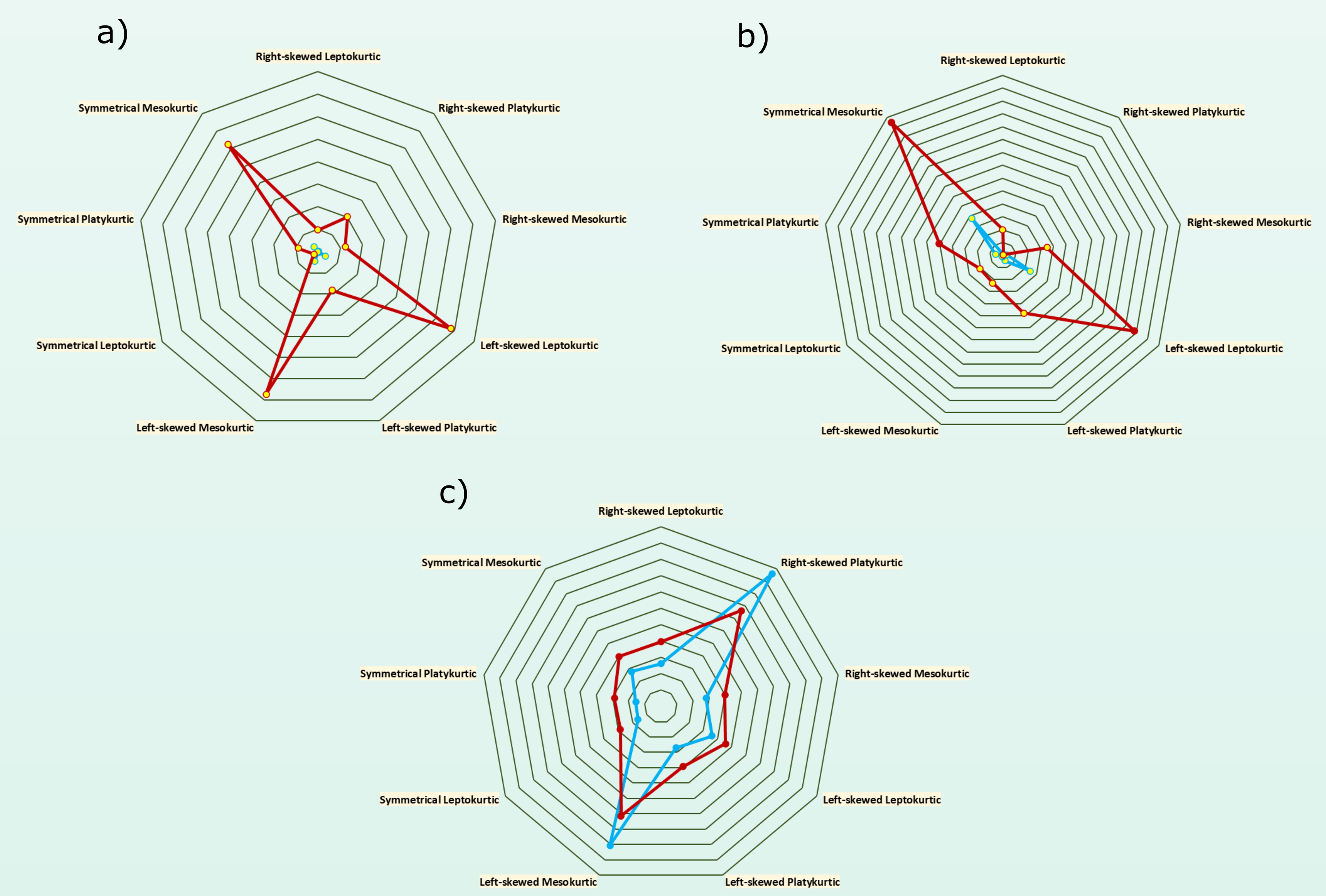

Fig.3. The radar chart of the results of Tukey's test and Bootstrap Boosted Tukey Test for various comparison types: a) Group 1 and group 3 comparison; b) Group 1 and group 2 comparison; c) Group 2 and group 3 comparison based on the distribution characteristics (right-skewed, left-skewed, symmetrical) along with their kurtosis and skewness.The closer to the center, the smaller the p-values for both tests, indicating statistical significance of the results. Conversely, the farther from the center, the larger the p-values, indicating lack of statistical significance. The blue line represents Tukey's test, while the red line represents Bootstrap Boosted Tukey Test. Yellow points denote p-values less than 0.05, indicating statistical significance for those comparisons.

Conclusion

The study showed how the bootstrap method enhances multiple comparison analyses across various distribution scenarios. We employed ANOVA and post hoc tests to scrutinize differences between groups based on homogeneity of variances and distribution characteristics. Despite slight inflation of P values, bootstrap analyses provided robust results across both homogeneous and heterogeneous variance scenarios. Table 1 presents a summary of the performance of bootstrap and its effectiveness across various distributions, as well as its application in multiple group analyses.

Table 1 Applications of the bootstrap method: examples of situations and contexts in data analysis.

Independence from Population Distribution Assumptions	Bootstrap does not require advanced assumptions about the population distribution, making it universal and versatile
Handling Heterogeneity	Bootstrap effectively handles cases regardless of whether the variance of a variable is heterogeneous or homogeneous
Efficiency in Post Hoc Tests	Bootstrap is efficient in the multi comparison analysis with post hoc tests, demonstrating effectiveness at different levels of conservativeness
Adaptation to Skewness and Kurtosis	Bootstrap adapts to distributions with varying levels of skewness and kurtosis, enabling analysis of more complex data

Scientific achievements & further research

Authorship in two research articles:

1. Karolczak, K., Guligowska, A., Sołtysik, B. K., Kostanek, J., Kostka, T., & Watala, C. (2024). Estimated Intake of Potassium, Phosphorus and Zinc with the Daily Diet Negatively Correlates with ADP-Dependent Whole Blood Platelet Aggregation in Older Subjects. Nutrients, 16(3), 332. https://doi.org/10.3390/nu16030332
2. Gurubaran, I. S., Watala, C., Kostanek, J., Szczepanska, J., Pawlowska, E., Kaarniranta, K., & Blasiak, J. (2024). PGC-1α regulates the interplay between oxidative stress, senescence and autophagy in the ageing retina important in age-related macular degeneration. Journal of cellular and molecular medicine, 28(8), e18051. https://doi.org/10.1111/jcmm.18051

Participation in the creation of two original scientific articles that are currently being prepared and the original scientific article with my first authorship, which has been submitted to a journal. Currently I am preparing he second article, of which I am the primary author. This work will be a part of my doctoral dissertation and will focus on comparing lipid profiles among women and men afflicted with cardiovascular disease, examining whether they differ significantly based on BMI classification. All analyses will be conducted using the bootstrap method.

REFERENCES

1. Blanca MJ, Arnau J, López-Montiel D, Bono R, Bendayan R. Skewness and kurtosis in real data samples. Methodology (Gott) 2013, 9, 78–84. doi.org/10.1027/1614-2241/a000057.
2. Bono R, Blanca MJ, Arnau J, Gómez-Benito J. Non-normal distributions commonly used in health, education, and social sciences: A systematic review. Front. Psychol. 2017, 8, 1602. doi.org/10.3389/fpsyg.2017.01602.
3. Brown S. Measures of Shape: Skewness and Kurtosis, 2008–2016, Oak Road Systems, Available from: brownmath.com (accessed on 12.04.2024).
4. Jar JH. Biostatistical Analysis, 4th ed. Prentice Hall International, Inc., Upper Saddle River, NJ., 1999; pp. 67-73.