BIAR Publisher
Matondang Journal
e-ISSN : 2828-1942
p-ISSN : 2828-979x

Selection of Appropriate Statistical Methods for Data Analysis: A Review

Agbonyinma Mary Igharo[1], Adetayo Olorunlana[2], Nnenna Okorie-Eugene[3]
[1]Director of Nursing Services Edo State Hospitals Management Agency, Benin City. An adjunct lecturer with Edo State College of Nursing Sciences
[2] College of Nursing and Basic Medical Sciences, Caleb University, Lagos, Nigeria
[3]Department of Nursing, College of Nursing and Basic Medical Sciences,Caleb University, Lagos, Nigeria
Email: m.igharo@edostate.gov.ng, adetayo.olorunlana@calebuniversity.edu.ng

Abstract: *Choosing the right statistical techniques for analyzing data is essential to obtain accurate and trustworthy results from research efforts. This article gives a detailed summary of the factors that affect the selection of statistical techniques, such as data characteristics, study goals, sample size, data spread, measurement level, assumptions, software access, and researchers' skills. Using case studies and examples, we demonstrate how statistical methods can be applied in real-life situations, showcasing both effective techniques and obstacles faced. Practical advice is given to researchers and practitioners, highlighting the significance of comprehensive exploratory data analysis, validation of assumptions, clear communication of findings, and ongoing learning. Opportunities for future research and development in statistical methodology are pinpointed, such as the advancement of methods for managing intricate data structures and the improvement of software tools and resources. Following proven methods and welcoming new ideas, researchers and professionals can improve the accuracy and dependability of their studies and help advance statistical techniques.*
Keywords: *Accuracy, Assumption, Data, Methods, Research, Statistical techniques*

I. Introduction

In the constantly growing field of research and data-informed decision-making, choosing the right statistical techniques for analyzing data is crucial for ensuring the quality and credibility of study results (Okoye & Hosseini, 2024a; Adeleke et al., 2024; Ravid, 2024). With data collection becoming more widespread in various fields like healthcare, economics, social sciences, and engineering, there is a growing demand for strong and informed statistical methods (Gawali, 2023). The significant increase in data accessibility, known as "big data," offers benefits and difficulties for researchers and professionals (Furstenau et al., 2023). In the field of healthcare, advanced statistical methods are necessary for examining complicated biological information and enhancing patient results (Faizi & Alvi 2023). Likewise, in the field of economics, being able to analyze and understand extensive amounts of data is essential for making informed decisions and creating effective policies (Veldkamp & Chung, 2024). Sophisticated statistical tools are utilized by the social sciences to comprehend human behavior and societal trends, and engineering fields use these methods to enhance systems and create innovative solutions. Choosing appropriate statistical techniques is important as it greatly affects the accuracy, consistency, and understandability of research results (Cooksey, 2020; Schober et al., 2021; Gawali, 2023).

Inappropriate statistical methods may result in inaccurate findings, causing misguided design with significant consequences on final output (Barkow, 2019; Pickles et al., 2023; Davis-Stober et al., 2024). On the other hand, using suitable statistical techniques enables

DOI: *https://doi.org/10.5281/zenodo.18669360*

researchers to extract valuable insights, reveal concealed patterns, and draw informed conclusions from intricate datasets (Kharakhash, 2023; Khan et al., 2024). Likewise, the selection of statistical approaches is closely connected with the particular goals and features of the research project (Alaimo, 2022; Mukherjee et al., 2024). Misuse of statistic occurs when a statistic argument asserts a falsehood, negligence, and deliberate deception for favor, accidental misapplied data, design, misinterpretation, misleading readers and others (Calzon, 2021; Pasquetto et al., 2024).

Similarly, Obilor & Ikpa (2022) observed that misuse of statistics could also be unintentional due to inexperience, lack of statistical skills and inappropriate knowledge. Some common causes of misuse of statistics in research include the following: bias, inappropriate tests, ignoring important features, overgeneralization, assuming causation from correlation, abuse of data collection and others. Selecting appropriate statistical techniques that closely match the research goals is crucial for achieving reliable results, whether the purpose is describing phenomena, comparing groups, predicting outcomes, or testing hypotheses (Kraimer et al., 2023; Abdumalikovich, 2023; McDermott, 2023; Airaoje et al., 2024).

Moreover, the increasing use of complex statistical methods requires a detailed comprehension of their assumptions, constraints, and relevance to the study of interest (Spanos, 2023; Haddad et al., 2023; Thakur, 2024). Scientists need to understand and utilize a wide range of statistical methods, including traditional inferential techniques and contemporary machine learning algorithms, which are designed to tackle various analytical obstacles and data formats.

Choosing the right statistical techniques is essential because it can greatly influence the accuracy and dependability of research results. For example, conventional methods such as regression analysis or ANOVA are frequently appropriate for relatively straightforward and well-known datasets with distinct assumptions about the underlying distributions (Haines et al., 2023). Yet, for scenarios with high-dimensional data or intricate, non-linear connections, advanced methods like random forests, support vector machines, or neural networks might be required to reveal significant patterns and insights (Saber & Yousof, 2023). Moreover, the incorporation of these sophisticated techniques demands a thorough evaluation of overfitting, computational efficiency, and interpretability, which are crucial aspects in guaranteeing reliable and widely applicable outcomes (Pavlov et al., 2024).

Hence, researchers must have a thorough knowledge of both traditional and modern statistical methods to make well-informed decisions that improve the quality and significance of their analyses. Apart from the technical factors, researchers need to take into consideration the background and objectives of their research when choosing statistical techniques. During clinical trials, the main goal is frequently to assess the effectiveness of a treatment, which requires the application of strict hypothesis testing and confidence interval estimation to confirm that findings are both statistically valid and medically important (Gao et al., 2023). Conversely, in the field of exploratory data analysis in the social sciences, the emphasis may be on discovering patterns and developing hypotheses, which may include using methods such as cluster analysis or principal component analysis to identify hidden patterns in the data (Johnson & Wichern, 2023). Moreover, within the scope of big data, where datasets may be extensive and diverse, techniques like data mining and predictive analytics are crucial for uncovering valuable insights (Liu & Zhang, 2023).

The suitability of a statistical method is closely connected to the research question, the type of data, and the intended results. Researchers need to skillfully align their methods of analysis with the particular requirements of their research to enhance the credibility and influence of their results (Smith & Roberts et al., 2024). Given these factors, this review seeks to thoroughly analyze the influences on the selection of the right statistical methods for analyzing data.

Despite the fact that misuse of statistics has caused damages to programmes in higher institutions, companies Political groups, medical researchers who depends on viable and reliable data that could benefit them. Several scholars (Calzon, 2021; Baldwin, et al., 2022; Isha et al., 2022; Airaoje et al., 2023) noted that data abuses include the incorrect application of statistical tests, lack of transparency and disclosure about decisions that are made to obtain multiple measures for latent variable modeling, test simpler models, and incorporate error correction methods, incomplete or incorrect multivariate model building or exclusion of outliers. Therefore, proper, efficient, and transparent use of statistics is necessary. Referencing previous research and practical instances, we explore the intricate relationship among data features, research goals, sample size deliberations, and software accessibility when determining statistical methods. This review aims to provide researchers and practitioners with useful insights and advice to help them effectively navigate the complex world of statistical analysis when choosing appropriate statistical methods for data analysis.

1.1 Objectives

The review objectives are as follows:

1. To examine the factors that influence the selection of appropriate statistical methods for data analysis.
2. To use existing literature and real-world examples to explore how these factors interact and guide the choice of statistical techniques.
3. To provide practical insights and recommendations to help researchers and practitioners choose the right statistical methods.

II. Review of Literatures

2.1 Factors influencing the selection of appropriate statistical methods

Several main factors influence the choice of suitable statistical methods for data analysis. Researchers must understand these factors to make well-informed decisions on statistical techniques (Dehalwar & Sharma, 2023). The type of statistical techniques used is greatly impacted by the type of data being examined, which can be broadly classified as either categorical or numerical (Ravid, 2024). The selection of statistical methods is influenced by the nature of the data being examined, such as whether it is categorical or numerical. Various techniques are appropriate for examining diverse types of data; for instance, chi-square tests are used for categorical data, while regression analysis is employed for numerical data (Lester et al., 2020; Azen & Walker, 2021). Data that is categorical is made up of clear categories or groups and is generally qualitative, such as gender, ethnicity, or types of products (Grodal et al., 2021; Idika et al., 2023). Conversely, numerical data encompasses quantities or measurements and can be categorized as either discrete or continuous, with age, income, or temperature serving as examples (Lock et al., 2020; Cressie & Moores, 2023; Airaoje et al., 2024). The selection of statistical techniques is impacted by the unique features of categorical and numerical data, which have varying characteristics and properties. Commonly utilized methods for analyzing relationships and testing associations between categorical variables include chi-square tests, logistic regression, and contingency tables in the case of categorical data (Miola & Miot, 2022). To analyze numerical data, techniques like correlation analysis, t-tests, ANOVA, and regression analysis are used to investigate connections, compare averages, and make forecasts based on numerical variables (Bensken et al., 2021; Langenberg et al., 2023).

In his study, Abdi (2023) uses Chi-square tests, including Pearson's chi-square test and Fisher's exact test, to assess the association or independence among categorical variables. Correlation analysis, which includes both the Pearson correlation coefficient and Spearman's rank correlation coefficient, evaluates the strength and direction of the relationship between two numeric variables (Yu & Hutson 2024). Comparing averages among different groups is done through T-tests and ANOVA (Murphy, 2024) while showcasing the relationship between predictor variables and a numeric outcome variable is achieved through regression analysis like linear regression and logistic regression (Kamel, & Abonazel, 2023). The specific research goals and questions determine the appropriate statistical methods (Althubaiti, 2023). Different methods are selected to efficiently reach objectives like recognizing patterns, contrasting groups, making forecasts, or conducting hypothesis tests (Rahnenführer et al., 2023). Determining the proper statistical techniques for data analysis is crucial based on the research objectives. These objectives consist of description, comparison, prediction, and hypothesis testing, as specified by Rahnenführer et al. (2023). According to Gong, et al. (2023), objectives might involve explaining and summarizing the main elements of a dataset. This may include calculating central tendency measures (like mean, and median) and dispersion (such as standard deviation) as well as portraying data using graphs or charts (Contreras, et al., 2023). Making comparisons allows for the assessment of differences or similarities between groups or variables. This can involve using t-tests or ANOVA to compare averages for numerical data, or employing chi-square tests for categorical data (Murphy, 2024). Regarding predicting outcomes, the process involves using available data to anticipate future results (Riley & Collins, 2023). Regression analysis is better suited for this purpose, with linear regression used for continuous outcomes and logistic regression used for categorical outcomes (Pate et al., 2023; Boe et al., 2024).

Conversely, hypothesis testing aims to evaluate specific hypotheses or research questions by analyzing sample data (Lakens, 2022; Jing et al., 2024). This might include performing t-tests or ANOVA for mean comparisons (Lee, 2022), exploring correlations between variables using correlation analysis (Chatterjee, 2021), or analyzing proportions using chi-square tests (Roldán-Nofuentes et al., 2024). Barroga et al. (2023) and Ackermann (2024) highlighted the importance of utilizing customized statistical methods that are suited to particular research goals and hypotheses. Descriptive statistics show how data is spread out and average values, while inferential statistics conclude a population using sample data. Likewise, regression analysis is vital for predicting outcomes, while statistical tests are used to evaluate research hypotheses in hypothesis testing (Rahnenführer et al., 2023; Ambrose et al., 2024). Analyzing data frequently includes utilizing descriptive statistics and visual aids like average, middle value, most frequent value, data spread, bar graphs, and box plots. To evaluate variations in means, standard deviations, or proportions among various groups or scenarios, researchers utilize T-tests, ANOVA, and chi-square tests (Babu R, 2020; Pérez-Guerrero et al., 2024). Predictors in regression analysis are utilized to predict continuous outcomes via linear regression and categorical outcomes through logistic regression (Ma et al., 2023). Different statistical tests are used to test hypotheses based on the type of research question and data, including t-tests, ANOVA, correlation analysis, and chi-square tests (Vandever, 2020; Dar et al., 2023).

Researchers must consider the sample size when choosing statistical methods for small samples. Olorunlana et al., (2018) suggest utilizing non-parametric tests such as the Wilcoxon rank-sum test or Fisher's exact test for group or variable comparisons. Alternatively, one could employ techniques such as Fisher's exact test or permutation tests to assess relationships or test theories. Parametric techniques such as t-tests, ANOVA, and regression analysis can be suitable for large datasets if assumptions are met and data distributions are relatively normal. Non-parametric methods or precise tests may be essential for smaller sample sizes, whereas parametric methods can be effectively utilized for larger datasets (Zygmonta, 2023; Mokhtar et al., 2023; Aliyu et al., 2023). The sample size plays a key role in selecting the appropriate statistical techniques for data analysis (Althubaiti, 2023; Bhardwaj et al., 2024). Acknowledging the significance of sample size is essential when selecting appropriate methods and ensuring the precision of research results (Rahman, 2023; Zickar & Keith, 2023). The size of the sample significantly affects the accuracy and dependability of estimates, along with the credibility of statistical tests, which in turn impacts the choice of statistical approaches (Zhang & Hartmann, 2023; Bhardwaj et al., 2024). Limited sample sizes can lead to more uncertainty and less statistical power, while extensive datasets can provide more precise estimates and improved statistical power (Zimmer et al., 2023; Nakagawa et al., 2024). Having small sample sizes in statistical analysis can be challenging, as traditional parametric methods may not be reliable if assumptions are not fulfilled or if the data does not adhere to a normal distribution (Althubaiti, 2023; Bhardwaj et al., 2024). Non-parametric methods or exact tests may be better suited for analyzing small samples in certain situations as they do not depend on distributional assumptions and offer reliable results (Park et al., 2024).

III. Result and Discussion

3.1 Interaction of Factors Guiding Statistical Technique Selection

Evaluating the spread of data is essential in statistical analysis as it guides the choice of suitable inferential techniques and guarantees the accuracy of statistical results (Mertler, et al., 2021; Casella & Berger, 2024). Comprehending how data is spread out gives researchers a better understanding of its fundamental traits, aiding in the discovery of trend, anomalies, and biases (Shahbazi et al., 2023; Sladekova & Field, 2024). It also influences the selection of statistical techniques, as varied distributions may need distinctive analytical methods (Hammer & Harper, 2024). The selection of inferential statistical techniques is influenced by the distribution characteristics of the data (Klein, 2024). Parametric techniques like t-tests and ANOVA rely on particular distributions (e.g., normal distribution) and could yield incorrect outcomes if these assumptions are not met (Shatz, 2024). In contrast, non-parametric techniques do not rely on specific distributions and are less affected by deviations from normality (Okoye & Hosseini, 2024b). According to Tikka et al. (2024), parametric techniques like t-tests, ANOVA, and linear regression are suitable for examining relationships or distinctions between groups when the data conforms to a normal distribution. Non-parametric methods such as the Wilcoxon rank-sum test or Kruskal-Wallis test can be utilized to compare groups or variables in situations where distributions are skewed or non-normal (Kumari et al., 2024). Chi-square tests are frequently employed in analyzing categorical data to examine relationships or evaluate hypotheses involving categorical variables (Hile et al., 2022; Vierra et al., 2023; Cuhadar & Kalkan, 2024).

The measurement scale of variables, such as nominal, ordinal, interval, and ratio, impacts the selection of statistical techniques during data analysis (Teli et al., 2023; Okoye & Hosseini, 2024a). Categorical variables depict different groups or categories without any specific order or hierarchy, such as gender, ethnicity, or various types of products (Pérez-Guerrero et al., 2024; Jain & Kamalja, 2024). Ordinal variables possess an innate sequence or hierarchy, however, the gaps between values might vary (Gravel et al., 2021). Some instances are survey replies (such as Likert scales) or educational backgrounds (Robinson, 2024). Interval variables follow a meaningful order with equal intervals between values but lack a true zero point, such as temperatures in Celsius or Fahrenheit (McCaskey, 2020; Baak et al., 2020; Chikendu, 2023). Ratio variables are characterized by a significant order, evenly spaced intervals, and a genuine zero point denoting the lack of specific attributes, such as height, weight, and income (Imbens, 2021).

Table 1. Selection of Appropriate Statistical Methods for Data Analysis

Analysis Type	Variable Characteristics	Recommended Test	Assumptions	Non-Parametric Alternative
One Independent Variable (IV)				
Continuous DV	Interval/Ratio, Normal Distribution	Independent t-test	Normality, Equal variances	Mann-Whitney U test
Ordinal DV or Non-Normal	Ordinal/Non-Normal	-	-	Wilcoxon Rank-Sum test
Categorical DV	Nominal/Binary	Chi-square test	Expected frequencies ≥5	Fisher's exact test
One IV (2 paired groups)				
Continuous DV	Interval/Ratio, Normal Differences	Paired t-test	Normality of differences	Wilcoxon Signed-Rank test
Categorical DV (Binary)	Nominal	-	-	McNemar's test
One IV (≥3 independent groups)				
Continuous DV	Interval/Ratio, Normal Distribution	One-way ANOVA	Normality, Homogeneity of variance	Kruskal-Wallis H test
Categorical DV	Nominal	Chi-square test	Expected frequencies ≥5	Fisher-Freeman-Halton test
One IV (≥3 paired groups)				
Continuous DV	Interval/Ratio, Normal Distribution	Repeated Measures ANOVA	Sphericity	Friedman test
Two IVs (Independent Groups)				
Continuous DV	Interval/Ratio, Normal Distribution	Two-way ANOVA	Normality, Homogeneity of variance	Aligned Rank Transform ANOVA
Categorical DV	Nominal	Log-linear Analysis	Cell frequencies ≥5	Generalized Linear Models
Relationships Between Two Variables				

Both Continuous	Interval/Ratio	Pearson's r	Linearity, Normality	Spearman's rho
One Continuous, One Binary	Interval + Nominal	Point-Biserial Correlation	Normality of continuous variable	Rank-Biserial Correlation
Both Ordinal	Ordinal	Kendall's tau	-	Goodman-Kruskal gamma
Both Categorical	Nominal	Chi-square test	Expected frequencies ≥5	Cramer's V or Phi coefficient

Key Notes:

1. Normality checks: Use Shapiro-Wilk (small samples) or Kolmogorov-Smirnov (large samples) tests.
2. Post-hoc tests: Apply Tukey's HSD (ANOVA) or Dunn's test (Kruskal-Wallis) for pairwise comparisons.
3. Effect sizes: Report Cohen's *d* (t-tests), η^2 (ANOVA), or odds ratios (categorical data).
4. A path model: Depicts the causal relations (Regression equations), between independent variables of interest predicting the (dependent variable) outcome variable and follows specific rules for drawing complicated path models.

This framework integrates parametric efficiency with non-parametric robustness, aligning with variable measurement levels and research designs.

3.2 Practical Insights and Recommendations for Choosing Statistical Methods

Several statistical techniques depend on particular assumptions about both the data and the population that it represents (Campbell & Jacques, 2023). It is crucial to comprehend these assumptions and confirm them before utilizing statistical methods to guarantee the accuracy of the outcomes (James et al., 2023; Shatz, 2024). Here we outline three typical beliefs that support statistical techniques. To begin with, most parametric methods rely on the assumption of data being normally distributed (Habibzadeh, 2024; Chen et al., 2024). Moreover, uniformity of variance is crucial as parametric tests such as ANOVA require equal variance among groups (Das et al., 2022; Juarros-Basterretxea, et al., 2024). In various statistical tests, independence is assumed among observations, where they are not influenced by each other (Thams et al., 2023; Zhang & Zhang 2024; Podkopaev & Ramdas, 2024).

Confirming assumptions is important as not doing so can result in skewed estimates, increased Type I error rates, or decreased statistical power (Schielzeth et al., 2020; Knief & Forstmeier, 2021; Chipman et al., 2024). Researchers need to determine if the data fulfill necessary assumptions by examining them visually, conducting diagnostic tests, or making transformations (Jones et al., 2024). Researchers have three choices to rectify errors when assumptions are not met, as mentioned by Shatz (2024). Data can be manipulated through transformations, such as logarithmic or square root adjustments, to normalize distributions or stabilize variances, as stated by Politis (2024). Two, utilizing Non-Parametric Methods, as these tests do not depend on distributional assumptions and may be better suited for skewed or non-normal data (Awuma et al., 2024). Lastly, robust statistical techniques are not as affected by assumptions being violated and can still give accurate results even when dealing with outliers or non-normal data (Wilcox, 2023).

Case studies and examples provide useful perspectives on how statistical methods are used in real-life situations, showcasing both effective approaches and obstacles faced. Through marketing analysis, a marketing company can conduct surveys to evaluate the impact of various advertising campaigns on customer engagement (Vinerean & Opreana 2021; Black, 2023). The company might employ ANOVA to assess the average engagement scores among the advertising groups, and logistic regression to pinpoint factors influencing customer response (van Esch et al., 2020; Sunarya et al., 2024). Using these statistical analyses, the marketing company can accurately interpret the results in the context of the research goals. For example, ANOVA findings will show if there are notable variations in customer involvement among various advertising strategies (Fisher, 2022; McClave et al., 2023). Post-hoc tests can determine the specific campaigns that outperformed others if significant discrepancies are observed. Likewise, the outcomes of logistic regression will identify crucial factors that influence customer response, assisting companies in enhancing their advertising tactics to effectively reach the most important predictors (Sweeney & Wren, 2023; Huber, 2024).

In the case of a medical research team looks into the effectiveness of a new treatment for a particular disease using healthcare research (Im, et al., 2023). Survival analysis techniques like Kaplan-Meier estimation and Cox proportional hazards regression are utilized to examine time-to-event data and evaluate treatment effects (Ameis et al., 2024; Charu et al., 2024; Jiang & Guterman 2024). Survival analysis methods like Kaplan-Meier estimation and Cox proportional hazards regression are important for evaluating the effectiveness of new treatments in healthcare research (Lee & Wang, 2023; Zhang et al., 2024). Kaplan-Meier curves visualize survival probability over time, whereas the Cox model detects factors that impact survival rates. These evaluations assist in comprehending the effects of the treatment and discovering patient traits that could indicate improved results.

The use of these statistical techniques allows marketing companies and medical scientists to base their decisions on data, improve their tactics, and reach their objectives more effectively (Miller & Roberts, 2023; Smith & Johnson et al., 2024). To sum up, the careful choice and use of suitable statistical techniques in marketing and healthcare studies demonstrate their essential contribution to thorough data analysis and well-informed decision-making (Fisher, 2022; McClave et al., 2023; Lee & Wang, 2023; Zhang et al., 2024). These techniques not just confirm the success of advertising plans and evaluate treatment efficiency but also provide detailed insights through the identification of key predictors and comprehension of intricate relationships (Sweeney & Wren, 2023; Miller & Roberts, 2023; Smith & Roberts et al., 2024). The accuracy in statistical analysis extends beyond these fields to include a variety of areas such as finance, education, and public policy.

Through the adoption of statistical rigor, companies in different industries can enhance resource distribution, improve tactics, and accomplish goals more effectively and confidently (Huber, 2024; Zhang et al., 2024). This thorough method of examining data doesn't just improve operational results but also encourages a culture of decision-making based on evidence, leading to ongoing enhancements and innovations (Lee & Wang, 2023; Miller & Roberts, 2023). As businesses and organizations incorporate more advanced statistical methods into their operations, they enhance their competitive advantage and also contribute to overall progress in knowledge and implementation globally (Sweeney & Wren, 2023; Smith & Roberts et al., 2024).

3.3 Further Research

More research is necessary to create innovative statistical techniques capable of managing intricate data formats like high-dimensional data, longitudinal data, and network data. Enhance methods for dealing with missing data, outliers, and non-normal distributions. Improve software tools and resources for statistical analysis, giving attention to ease of use, ability to grow, and compatibility with various systems. Investigate partnerships across different fields to combine statistical techniques with machine learning, artificial intelligence, and data visualization. Examine how new technologies like big data and the Internet of Things are affecting statistical methods and create creative solutions to tackle fresh obstacles.

IV. Conclusion

Various factors such as data nature, research goals, sample size, data distribution, measurement level, assumptions, software accessibility, and researchers' skills affect the choice of statistical methods. Various statistical techniques are needed to analyze diverse types of data. Comprehending the measurement level is essential for choosing suitable techniques. Researchers need to confirm the assumptions on which statistical methods are based to validate the results. Dealing with violations of assumptions can include transformations of data, non-parametric methods, or robust techniques. Factors like sample size, software accessibility, and researchers' knowledge of specific techniques also impact method choice. Perform a comprehensive analysis of the data to gain insights into its features and pinpoint any possible obstacles. Select statistical techniques that match the research goals and data attributes. Consistently enhance statistical knowledge and skills by participating in training, and workshops, and working with experts. Check assumptions before using statistical methods and explore other options if assumptions are not met. Effectively relay statistical results to stakeholders, offering clear explanations and practical recommendations. There is need to improve the efforts of investigators, editors, and reviewers to improve research presentation when data are subject to error for quality data output.

References

Abdi, S. J. (2023). A Comprehensive Guide for Selecting Appropriate Statistical Tests: Understanding When to Use Parametric and Nonparametric Tests. *Open Journal of Statistics*, *13*(4), 464-474.

Abdumalikovich, U. A. (2023). Analysis of the Paradigm of the Principles of Overcoming the Problems of Productivity of Researchers Conducting Research in the Field of Physical Culture. *Journal of Positive School Psychology*, 529-534.

Ackermann, J. (2024). From the idea to the research question and hypothesis. *Arthroskopie*, 1-5.

Adeleke, A. K., Montero, D. J. P., Olu-lawal, K. A., & Olajiga, O. K. (2024). Statistical techniques in precision metrology, applications and best practices. *Engineering Science & Technology Journal*, *5*(3), 888-900.

Airaoje, O. K., Aondover, E. M., Obada, A. A., Akin-Odukoya, O. O., & Ridwan, M. (2024). High Incidence of Different Drug Uses and Media Campaign on the Injection Method in Borno State, Nigeria. *Konfrontasi: Jurnal Kultural, Ekonomi dan Perubahan Sosial*, *11*(4), 242-258.

Airaoje, O. K., Obada, A. A., & Msughter, A. E. (2023). A Critical Review on Gender Based Violence in Nigeria: Media Dimension. *Humanities*, *3*(2), 9-16.

Airaoje, O. K., Ogunbola, O., Falobi, F., Obada, A., & Eric, M. (2024). Scoping Review on Factors Associated with Continuity of Treatment among People Living with HIV in Nigeria. *Biomedical Journal of Scientific & Technical Research*, *57*(3), 49283-49292.

Alaimo, L. S. (2022). *Complexity of social phenomena: Measurements, analysis, representations and synthesis* (Vol. 127). Sapienza Università Editrice.

Aliyu, M. A., Msughter, A. E., & Nneka, A. Q. (2023). Comparative Study of National Development Plans in Nigeria and India: Media Dimension. *SIASAT*, *8*(4), 202-212.

Althubaiti, A. (2023). Sample size determination: A practical guide for health researchers. *Journal of General and Family Medicine*, *24*(2), 72-78.

Ambrose, N., Aimufua, G. I., & Nyap, C. D. (2024). A Novel Approach for Developing a Linear Regression Model within Logistic Cluster Using Scikit-Learn. *Journal of Data Analysis and Information Processing*, *12*(3), 348-369.

Ameis, L., Kuss, O., Hoyer, A., & Möllenhoff, K. (2024). A nonparametric proportional risk model to assess a treatment effect in time-to-event data. *Biometrical Journal*, *66*(4), 2300147.

Awuma, W., Amankwa, I. A., Richard, A., & Danso, I. (2024). Parametric versus Non-Parametric Models in Stochastic Frontier Analysis: A Theoretical Review. *European Journal of Science, Innovation and Technology*, *4*(2), 1-12.

Azen, R., & Walker, C. M. (2021). *Categorical data analysis for the behavioral and social sciences*. Routledge.

Baak, M., Koopman, R., Snoek, H., & Klous, S. (2020). A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. *Computational Statistics & Data Analysis*, *152*, 107043.

Babu R, A. M. H. (2020). Deep Data Stat-A Survey Analysis on Impact of Statistics in Data Science for Students. *International Journal of Data Science and Applications*, *3*(1), 17-22.

Baldwin, J. R., Pingault, J. B., Schoeler, T., Sallis, H. M., & Munafò, M. R. (2022). Protecting against researcher bias in secondary data analysis: Challenges and potential solutions. *European Journal of Epidemiology, 37*(1), 1–10. https://doi.org/10.1007/s10654-021-00839-0

Barkow, R. E. (2019). Categorical Mistakes. *Harvard Law Review*, *133*(1), 200-241.

Barroga, E., Matanguihan, G. J., Furuta, A., Arima, M., Tsuchiya, S., Kawahara, C., ... & Izumi, M. (2023). Conducting and writing quantitative and qualitative research. *Journal of Korean Medical Science*, *38*(37).

Bensken, W. P., Ho, V. P., & Pieracci, F. M. (2021). Basic Introduction to Statistics in Medicine, part 2: Comparing Data. *Surgical Infections*, *22*(6), 597-603.

Bhardwaj, R., Agrawal, U., Vashist, P., & Manna, S. (2024). Determination of sample size for various study designs in medical research: A practical primer. *Journal of Family Medicine and Primary Care*, *13*(7), 2555-2561.

Black, K. (2023). *Business statistics: for contemporary decision making*. John Wiley & Sons.

Boe, L., Vingan, P. S., Kim, M., Zhang, K. K., Rochlin, D., Matros, E., ... & Nelson, J. A. (2024). Methods in regression analysis in surgical oncology research-best practice guidelines. *Journal of Surgical Oncology*, *129*(1), 183-193.

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature, 582*(1), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Calzon, B. (2021). Misleading statistics examples – Discover the potential for misuse of statistics & data in the digital age. *Datapine*. Retrieved June 12, 2022, from https://www.datapine.com/blog/misleading-statistics-and-data/

Campbell, M. J., & Jacques, R. M. (2023). *Statistics at square two*. John Wiley & Sons.

Casella, G., & Berger, R. (2024). *Statistical inference*. CRC Press.

Charu, V., Tian, L., Tamura, M. K., & Montez-Rath, M. E. (2024). Using Restricted Mean Survival Time to Improve Interpretability of Time-to-Event Data Analysis. *Clinical Journal of the American Society of Nephrology*, *19*(2), 260-262.

Chatterjee, S. (2021). A new coefficient of correlation. *Journal of the American Statistical Association*, *116*(536), 2009-2022.

Chen, L. T., Chen, Y. K., Yang, T. R., Chiang, Y. S., Hsieh, C. Y., Cheng, C., ... & Peng, C. Y. J. (2024). Examining the normality assumption of a design-comparable effect size in single-case designs. *Behavior Research Methods*, *56*(1), 379-405.

Chikendu, R. E. P. (2023). Statistical variables. *Statistics and data interpretation in educational research*, 20.

Chipman, J. J., Greevy Jr, R. A., Mayberry, L. S., & Blume, J. D. (2024). Sequential monitoring using the Second Generation P-Value with Type I error controlled by monitoring frequency. *The American Statistician*, 1-11.

committee. Norwegian Computing Center.

Contreras, B. P., Tate, S. A., Morris, S. L., & Kahng, S. (2023). A systematic review of the correspondence between descriptive assessment and functional analysis. *Journal of Applied Behavior Analysis*, *56*(1), 146-165.

Cooksey, R. W. (2020). *Illustrating statistical procedures: Finding meaning in quantitative data*. Springer Nature.

Cressie, N., & Moores, M. T. (2023). Spatial statistics. In *Encyclopedia of Mathematical Geosciences* (pp. 1362-1373). Cham: Springer International Publishing.

Cuhadar, I., & Kalkan, Ö. K. (2024). Performance of estimation methods in bifactor models with ordered categorical data. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*(2), 329-339.

Dar, N. A., Tali, T. A., Ganie, B. A., Sofi, M. A., Sofi, S. R., Khan, N. A., ... & Ahmad, S. N. (2023). Statistical Methods Used in Medical Research and Cancer Registries: A Review. *Journal of Radiation and Cancer Research*, *14*(3), 111-115.

Das, B. K., Jha, D. N., Sahu, S. K., Yadav, A. K., Raman, R. K., & Kartikeyan, M. (2022). Analysis of variance (ANOVA) and design of experiments. In *Concept Building in Fisheries Data Analysis* (pp. 119-136). Singapore: Springer Nature Singapore.

Davis-Stober, C. P., Dana, J., Kellen, D., McMullin, S. D., & Bonifay, W. (2024). Better accuracy for better science... through random conclusions. *Perspectives on Psychological Science*, *19*(1), 223-243.

Dehalwar, K., & Sharma, S. N. (2023). *Fundamentals of Research Writing and Uses of Research Methodologies*. Edupedia Publications Pvt Ltd.

Faizi, N., & Alvi, Y. (2023). *Biostatistics manual for Health Research: A practical guide to data analysis*. Elsevier.

Fisher, R. A. (2022). Statistical methods for research workers. Oliver & Boyd.

Furstenau, L. B., Leivas, P., Sott, M. K., Dohan, M. S., López-Robles, J. R., Cobo, M. J., ... & Choo, K. K. R. (2023). Big data in healthcare: Conceptual network structure, key challenges and opportunities. *Digital Communications and Networks*, *9*(4), 856-868.

Gao, F., Wang, J., & Chen, Y. (2023). Statistical methods in clinical trials: Ensuring efficacy and safety. *Clinical Research Statistics*, 12(2), 145-162.

Gawali, R. B. (2023). *Research Methodology And Statistical Methods*. Academic Guru Publishing House.

Gong, Y., Liu, G., Xue, Y., Li, R., & Meng, L. (2023). A survey on dataset quality in machine learning. *Information and Software Technology*, *162*, 107268.

Gravel, N., Magdalou, B., & Moyes, P. (2021). Ranking distributions of an ordinal variable. *Economic Theory, 71*, 33-80.
Grodal, S., Anteby, M., & Holm, A. L. (2021). Achieving rigor in qualitative analysis: The role of active categorization in theory building. *Academy of Management Review*, *46*(3), 591-612.
Gronli, K.S. (2022). Data misuse – or seven ways to fail with statistics. National Ethics
Habibzadeh, F. (2024). Data distribution: normal or abnormal? *Journal of Korean medical science*, *39*(3).
Haddad, J., Silva, F., & Nunes, R. (2023). Understanding statistical limitations in data analysis. *Applied Statistics Review*, 60(3), 223-240.
Haddad, R., Al Maali, S., & Saadeh, M. (2023). The translation of research central tendencies in clinical practice: power and limitations of statistics. In *Seminars in Orthodontics* (Vol. 29, No. 3, pp. 317-323). WB Saunders.
Haines, N., Sullivan-Toole, H., & Olino, T. (2023). From classical methods to generative models: Tackling the unreliability of neuroscientific measures in mental health research. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*(8), 822-831.
Haines, P., Miller, S., & Donovan, M. (2023). Classical inferential methods in modern research. *Journal of Statistical Methods*, 45(2), 98-112.
Hammer, Ø., & Harper, D. A. (2024). *Paleontological data analysis*. John Wiley & Sons.
Hanel, P. H. P., Maio, G. R., & Manstead, A. S. R. (2019). A new way to look at the data:
Hile, M. M., Msughter, A. E., & Babale, A. M. (2022). A Public Health Communication: Towards Effective Use of Social Marketing for Public Health Campaigns in Nigeria. *Ann Community Med Prim Health Care*, *5*(1), 1002.
Huber, P. J. (2024). Logistic regression and its applications in marketing. Statistical Science, 39(2), 112-125.
Idika, D. O., Owan, V. J., & Agama, V. U. (2023). The application of the nominal scale of measurement in research data analysis.
Im, D., Pyo, J., Lee, H., Jung, H., & Ock, M. (2023). Qualitative Research in Healthcare: data Analysis. *Journal of Preventive Medicine and Public Health*, *56*(2), 100.
Imbens, G. W. (2021). Statistical significance, p-values, and the reporting of uncertainty. *Journal of Economic Perspectives*, *35*(3), 157-174.
Isha, S. R., Azam, M. N. D., & Daud, M. N. (2022). A systematic review: Identifying research gaps in emotional intelligence development using storybook as a learning tool. *International Journal of Academic Research in Business and Social Sciences, 12*(10), 2496–2507.
Jain, R. R., & Kamalja, K. K. (2024). Variants of non-symmetric correspondence analysis for nominal and ordinal variables. *Journal of the Korean Statistical Society*, *53*(2), 435-478.
James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Resampling methods. In *An Introduction to Statistical Learning: with Applications in Python* (pp. 201-228). Cham: Springer International Publishing.
Jiang, F., & Guterman, E. (2024). Survival analysis. *Statistical Methods in Epilepsy*, 124-142.
Jing, X., Cimino, J. J., Patel, V. L., Zhou, Y., Shubrook, J. H., De Lacalle, S., ... & Liu, C. (2024). Data-driven hypothesis generation among inexperienced clinical researchers: A comparison of secondary data analyses with visualization (VIADS) and other tools. *Journal of Clinical and Translational Science*, *8*(1), e13.

Johnson, R. A., & Wichern, D. W. (2023). *Applied Multivariate Statistical Analysis*. Pearson.

Jones, L., Barnett, A., & Vagenas, D. (2024). Common misconceptions held by health researchers when interpreting linear regression assumptions, a cross-sectional study. *medRxiv*, 2024-02.

Juarros-Basterretxea, J., Aonso-Diego, G., Postigo, Á., Montes-Álvarez, P., Menéndez-Aller, Á., & García-Cueto, E. (2024). Post-hoc tests in one-way ANOVA: The case for normal distribution. *Methodology*, *20*(2), 84-99.

Kamel, A. R., & Abonazel, M. R. (2023). A simple introduction to regression modeling using R. *Computational Journal of Mathematical and Statistical Sciences*, *2*(1), 52-79.

Khan, R., Usman, M., & Moinuddin, M. (2024). From raw data to actionable insights: navigating the world of data analytics. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(4), 142-166.

Kharakhash, O. (2023). Data visualization: transforming complex data into actionable insights. *Automation of technological and business processes*, *15*(2), 4-12.

Klein, N. (2024). Distributional regression for data analysis. *Annual Review of Statistics and Its Application*, *11*.

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, *53*(6), 2576-2590.

Kraimer, M. L., Martin, X., Schulze, W., & Seibert, S. E. (2023). What does it mean to test theory? *Journal of Management Scientific Reports*, *1*(1), 8-17.

Kumari, P., Sathish Kumar, M., Vekariya, P., & Ghetiya, R. (2024). 14 Analytical Approaches. *AI in Agriculture for Sustainable and Economic Management*, 179.

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267.

Langenberg, B., Janczyk, M., Koob, V., Kliegl, R., & Mayer, A. (2023). A tutorial on using the paired t-test for power calculations in repeated measures ANOVA with interactions. *Behavior Research Methods*, *55*(5), 2467-2484.

Lee, E. T., & Wang, J. W. (2023). Statistical methods for survival data analysis (4th ed.). Wiley.

Lee, S. W. (2022). Methods for testing statistical differences between groups in medical research: statistical standard and guideline of Life Cycle Committee. *Life Cycle*, *2*.

Lester, J. N., Cho, Y., & Lochmiller, C. R. (2020). Learning to do qualitative data analysis: A starting point. *Human resource development review*, *19*(1), 94-106.

Liu, H., & Zhang, M. (2023). Data mining techniques for big data analytics. *Big Data Research Journal*, 8(1), 23-35.

Lock, R. H., Lock, P. F., Morgan, K. L., Lock, E. F., & Lock, D. F. (2020). *Statistics: Unlocking the power of data*. John Wiley & Sons.

Ma, J., Dhiman, P., Qi, C., Bullock, G., van Smeden, M., Riley, R. D., & Collins, G. S. (2023). Poor handling of continuous predictors in clinical prediction models using logistic regression: a systematic review. *Journal of Clinical Epidemiology*.

McCaskey, J. P. (2020). History of 'temperature': maturation of a measurement concept. *Annals of science*, *77*(4), 399-444.

McClave, J. T., Benson, P. G., & Sincich, T. (2023). Statistics for Business and Economics (14th ed.). Pearson.

McDermott, R. (2023). On the scientific study of small samples: Challenges confronting quantitative and qualitative methodologies. *The Leadership Quarterly*, *34*(3), 101675.

Mertler, C. A., Vannatta, R. A., & LaVenia, K. N. (2021). *Advanced and multivariate statistical methods: Practical application and interpretation*. Routledge.
Miller, M. E., & Roberts, S. (2023). Understanding survival analysis in clinical research. Journal of Medical Statistics, 12(2), 145-158.
Miola, A. C., & Miot, H. A. (2022). Comparing categorical variables in clinical and experimental studies. *Jornal Vascular Brasileiro*, *21*, e20210225.
Mokhtar, S. F., Yusof, Z. M., & Sapiri, H. (2023). Confidence intervals by bootstrapping approach: a significance review. *Malaysian Journal of Fundamental and Applied Sciences*, *19*(1), 30-42.
Mukherjee, A., Abraham, S., Singh, A., Balaji, S., & Mukunthan, K. S. (2024). From Data to Cure: A Comprehensive Exploration of Multi-omics Data Analysis for Targeted Therapies. *Molecular Biotechnology*, 1-21.
Murphy, S. (2024). Analysis of variance. In *Translational Orthopedics* (pp. 151-154). Academic Press.
Nakagawa, S., Lagisz, M., Yang, Y., & Drobniak, S. M. (2024). Finding the right power balance: Better study design and collaboration can reduce dependence on statistical power. *Plos Biology*, *22*(1), e3002423.
Obilor, E. I., & Ikpa, A. I. (2022). Misuses of statistics in research. *International Journal of Innovative Education Research*, *10*(3), 92–99. SEAHI Publications.
Okoye, K., & Hosseini, S. (2024a). Understanding the Different Types of Statistical Data Analysis and Methods. In *R Programming: Statistical Data Analysis in Research* (pp. 109-127). Singapore: Springer Nature Singapore.
Okoye, K., & Hosseini, S. (2024b). Choosing Between Parametric and Non-parametric Tests in Statistical Data Analysis. In *R Programming: Statistical Data Analysis in Research* (pp. 87-98). Singapore: Springer Nature Singapore
Olorunlana, A. Igharo, M.A. & Mark, D.O. (2018). Basics of Research Designs. In *Research Methodology in Behavioural Science and Law A Symbiosis* (pp.16-37). University of Fort Hare Press.
Park, C., Park, S. Y., Kim, H. J., & Shin, H. J. (2024). Statistical Methods for Comparing Predictive Values in Medical Diagnosis. *Korean Journal of Radiology*, *25*(7), 656.
Pasquetto, I. V., Cullen, Z., Thomer, A., & Wofford, M. (2024). What is research data “misuse”? And how can it be prevented or mitigated? *Journal of the Association for Information Science and Technology*, *75*(12).
Pate, A., Riley, R. D., Collins, G. S., Van Smeden, M., Van Calster, B., Ensor, J., & Martin, G. P. (2023). Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. *Statistical Methods in Medical Research*, *32*(3), 555-571.
Pavlov, A., Ivanov, P., & Sidorov, D. (2024). Advanced machine learning techniques for high-dimensional data. *Computational Statistics Journal*, 53(1), 31-49.
Pérez-Guerrero, E. E., Guillén-Medina, M. R., Márquez-Sandoval, F., Vera-Cruz, J. M., Gallegos-Arreola, M. P., Rico-Méndez, M. A., ... & Gutiérrez-Hurtado, I. A. (2024). Methodological and Statistical Considerations for Cross-Sectional, Case-Control, and Cohort Studies. *Journal of Clinical Medicine*, *13*(14), 4005.
Pickles, A., Edwards, D., Horvath, L., & Emsley, R. (2023). Research Reviews: Advances in methods for evaluating child and adolescent mental health interventions. *Journal of Child Psychology and Psychiatry*, *64*(12), 1765-1775.
Podkopaev, A., & Ramdas, A. (2024). Sequential predictive two-sample and independence testing. *Advances in neural information processing systems*, *36*.

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of best practices for evidence for prediction: a review. *JAMA psychiatry*, *77*(5), 534-540.
Politis, D. N. (2024). Studentization Versus Variance Stabilization: A Simple Way Out of an Old Dilemma. *Statistical Science*, *39*(3), 409-427.
Rahman, M. M. (2023). Sample size determination for survey research and non-probability sampling techniques: A review and set of recommendations. *Journal of Entrepreneurship, Business and Economics*, *11*(1), 42-62.
Rahnenführer, J., De Bin, R., Benner, A., Ambrogi, F., Lusa, L., Boulesteix, A. L., ... & topic group "High-dimensional data" (TG9) of the STRATOS initiative. (2023). Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges. *BMC medicine*, *21*(1), 182.
Ravid, R. (2024). *Practical statistics for educators*. Rowman & Littlefield.
Riley, R. D., & Collins, G. S. (2023). Stability of clinical prediction models developed using statistical or machine learning methods. *Biometrical Journal*, *65*(8), 2200302.
Robinson, J. (2024). Likert scale. In *Encyclopedia of quality of life and well-being research* (pp. 3917-3918). Cham: Springer International Publishing.
Roldán-Nofuentes, J. A., Sheth, T. S., & Vera-Vera, J. F. (2024). Hypothesis Test to Compare Two Paired Binomial Proportions: Assessment of 24 Methods. *Mathematics*, *12*(2), 190.
Saber, M. M., & Yousof, H. (2023). Bayesian and classical inference for generalized stress-strength parameter under generalized logistic distribution. *Statistics, Optimization & Information Computing*, *11*(3), 541-553.
Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., ... & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in ecology and evolution*, *11*(9), 1141-1152.
Schober, P., Mascha, E. J., & Vetter, T. R. (2021). Statistics from A (agreement) to Z (z score): a guide to interpreting common measures of association, agreement, diagnostic accuracy, effect size, heterogeneity, and reliability in medical research. *Anesthesia & Analgesia*, *133*(6), 1633-1641.
Shahbazi, N., Lin, Y., Asudeh, A., & Jagadish, H. V. (2023). Representation bias in data: A survey on identification and resolution techniques. *ACM Computing Surveys*, *55*(13s), 1-39.
Shatz, I. (2024). Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. *Behavior Research Methods*, *56*(2), 826-845.
Similarities between groups of people are large and important. Journal of Personality and
Sladekova, M., & Field, A. P. (2024). Psychology researchers' self-reported knowledge of sources of bias in general linear models and how it affects their analytic practice.
Smith, P. L., Johnson, A. D., & Martinez, F. (2024). Statistical models in health research: A guide for practitioners. Health Data Science, 18(1), 88-102.
Smith, T., Roberts, L., & Nguyen, P. (2024). Tailoring statistical methods to research contexts. *Journal of Applied Statistics*, 50(1), 78-95.
Social Psychology, 116(4), 541–562.
Spanos, A. (2023). Assumptions and limitations in statistical modeling. *Statistical Science Journal*, 38(2), 195-209.
Sunarya, P. A., Rahardja, U., Chen, S. C., Lic, Y. M., & Hardini, M. (2024). Deciphering digital social dynamics: A comparative study of logistic regression and random

forest in predicting e-commerce customer behavior. *Journal of Applied Data Sciences*, *5*(1), 100-113.
Sweeney, J. C., & Wren, B. M. (2023). Predictive analytics in marketing: Applications and implications. Journal of Marketing Analytics, 11(1), 33-47.
Teli, A., Nayaka, R., & Ghatanatti, R. (2023). Data analysis–preference of pertinent statistical method in research. *National Journal of Physiology, Pharmacy and Pharmacology*, *13*(10), 2010-2014.
Thakur, R. (2024). Nuances of statistical methodologies in diverse fields. *International Journal of Advanced Statistics*, 47(1), 56-78.
Thams, N., Saengkyongam, S., Pfister, N., & Peters, J. (2023). Statistical testing under distributional shifts. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *85*(3), 597-663.
Tikka, S., Kopra, J., Heinäniemi, M., López-Pernas, S., & Saqr, M. (2024). Introductory statistics with R for educational researchers. In *Learning Analytics Methods and Tutorials: A Practical Guide Using R* (pp. 121-150). Cham: Springer Nature Switzerland.
van Esch, P., von der Heidt, T., Frethey-Bentham, C., & Northey, G. (2020). The effect of marketing simulations on student engagement and academic outcomes. *Marketing Education Review*, *30*(1), 43-56.
Vandever, C. (2020). Introduction to Research Statistical Analysis: An Overview of the Basics. *HCA Healthcare Journal of Medicine*, *1*(2), 71.
Veldkamp, L., & Chung, C. (2024). Data and the aggregate economy. *Journal of Economic Literature*, *62*(2), 458-484.
Vierra, A., Razzaq, A., & Andreadis, A. (2023). Categorical variable analyses: chi-square, Fisher Exact, and mantel–haenszel. In *Translational Surgery* (pp. 171-175). Academic Press.
Vinerean, S., & Opreana, A. (2021). Measuring customer engagement in social media marketing: A higher-order model. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(7), 2633-2654.
Wilcox, R. R. (2023). *A guide to robust statistical methods*. New York: Springer.
Yu, H., & Hutson, A. D. (2024). A robust Spearman correlation coefficient permutation test. *Communications in Statistics-Theory and Methods*, *53*(6), 2141-2153.
Zhang, J., & Zhang, Z. (2024). A normal test for independence via generalized mutual information. *Statistics & Probability Letters*, *210,* 110113.
Zhang, X., & Hartmann, P. (2023). How to calculate sample size in animal and human studies. *Frontiers in Medicine*, *10*, 1215927.
Zhang, Z., Reinikainen, J., & Hänninen, T. (2024). Survival analysis: Methods and applications in health research. Journal of Biomedical Research, 64(1), 1-12.
Zickar, M. J., & Keith, M. G. (2023). Innovations in sampling: Improving the appropriateness and quality of samples in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior*, *10*(1), 315-337.
Zimmer, F., Henninger, M., & Debelak, R. (2023). Sample size planning for complex study designs: A tutorial for the mlpwr package. *Behavior research methods*, 1-18.
Zygmonta, C. S. (2023). Managing the Assumption of Normality within the General Linear Model with Small Samples: Guidelines for Researchers Regarding If, When and How. *The Quantitative Methods for Psychology*.