International Journal of Scientific Research and Management (IJSRM)
||Volume||12||Issue||03||Pages||1730-1741||2024||
Website: https://ijsrm.net ISSN (e): 2321-3418
DOI: 10.18535/ijsrm/v12i03.sh03

My Data Are Ready, How Do I Analyze Them: Navigating Data Analysis in Social Science Research

Fodouop Kouam Arthur William

https://orcid.org/0009-0009-3030-1094
School of Management, Hebei University
ZIP Code 071000, Baoding City, Hebei Province, China

Abstract
This paper offers a practical guide for researchers to effectively analyze their data in social science research, addressing the challenges and pitfalls commonly encountered in data analysis. By exploring various data analysis techniques, highlighting key challenges such as data quality issues and statistical assumptions violations, and providing practical tips and guidelines, this study fills a gap in the existing literature by offering a comprehensive approach to navigating data analysis in social science research. The significance of this study lies in its potential to improve the quality and reliability of research findings in the social sciences, equipping researchers with the necessary knowledge and skills to conduct robust data analysis. This study is a valuable resource for researchers seeking to enhance their analytical skills, avoid common pitfalls, and advance knowledge in their field of study.

Keywords: Data analysis, data quality, research findings, research methods, social science research, statistical techniques

1. Introduction

Social science research often involves collecting and analyzing vast amounts of data to answer complex research questions. Tindall et al. (2022) highlight the role of big data and computational social science in social network analysis, particularly in data collection, statistical inference, and ethical considerations. Aravindh and Thirupathi (2019) compare the use of quantitative and qualitative research methods in social science, emphasizing the depth and scope of qualitative research. Hou et al. (2020) discuss the practical applications of data analysis in social media, including topic analysis, time series analysis, sentiment analysis, and network analysis, and their impacts on disaster management, healthcare, and business. Stukal et al. (2021) explore the use of data science methods in political science research, focusing on analyzing protest activity in social media and applying supervised and unsupervised learning techniques.

In recent years, there has been a growing emphasis on the importance of data analysis in social science research. Manian and Vadivel (2021) highlight the importance of data analytics in the modern era, particularly in fields like big data analytics, artificial intelligence, and machine learning. Toker (2022) emphasizes the significance of qualitative data analysis in understanding and interpreting the social world, providing a guide for novice researchers. With the advent of new technologies and tools for data collection, researchers have access to unprecedented amounts of data.

However, the sheer volume of data can be overwhelming, and researchers often need help understanding it all. Despite the increasing data availability, many researchers need more knowledge and skills to analyze their data effectively. Zwilling (2023) underlines the need for students to acquire the necessary skills for big data analysis, suggesting the design of appropriate academic programs. It often results in incomplete or inaccurate analyses, leading to flawed research findings. As a result, there is a pressing need for a practical guide to help researchers navigate the data analysis process.

Previous studies have highlighted the importance of data analysis in social science research. However, these studies have often focused on specific techniques or methods without providing a comprehensive guide for researchers. Additionally, many of these studies still need to address the challenges and pitfalls researchers often encounter during data analysis. In light of the above, this paper's research question is: **How can researchers effectively navigate the data analysis process in social science research?** The primary objective of this study is to provide a practical guide to help researchers analyze their data effectively, avoiding common pitfalls and challenges.

This study is significant as it fills a gap in the existing literature by providing a comprehensive guide to data analysis in social science research. It aims to improve the quality and reliability of research findings in the social sciences by equipping researchers with the necessary knowledge and skills to analyze their data effectively.

The paper is organized as follows: In the next section, we overview the various data analysis techniques commonly used in social science research. We then discuss the challenges and pitfalls researchers often encounter during data analysis. Subsequently, we offer practical tips and guidelines for effectively navigating the data analysis process. Finally, we summarize the study, including its limitations and future research avenues.

2. Data analysis in social science research: Exploring commonly used techniques

Data analysis techniques are crucial in social science research, enabling researchers to make sense of complex datasets and draw meaningful conclusions. This section explores commonly used data analysis techniques, from descriptive statistics to network analysis, highlighting their applications and significance in social science research.

A range of data analysis techniques are commonly used in social science research. Bryman and Cramer (1990) emphasize using the Statistical Package for the Social Sciences (SPSS) for quantitative data analysis. Ball (1965) suggests using computer-oriented techniques for data analysis, and Hall (2010) explores the potential of remote sensing data in social science research, particularly when combined with traditional methods. Hou et al. (2020) and Blank et al. (2021) highlight using social media data, focusing on practical applications and commonly used techniques such as topic analysis, time series analysis, sentiment analysis, and network analysis. Maravelakis (2019) comprehensively reviews statistical methods in the social sciences, including parametric and nonparametric modeling, multivariate methods, time series modeling, and categorical data analysis. Chen et al. (2021) explicitly discuss social media images as data sources, noting the increasing interest in this area and using thematic coding, object recognition, and narrative analysis as common techniques.

2.1. Descriptive statistics

Descriptive statistics are commonly used to summarize and describe the characteristics of a dataset. This method is handy for quantitative data that can be easily quantified and summarized numerically. For example, suppose a researcher collects survey data on the ages of participants. In that case, descriptive statistics like the mean, median, and standard deviation can help summarize the central tendency and variability of the data.

Rogers (1998) further explores descriptive analysis measures such as frequency distributions, central tendency, variability, and correlation. Mair and Kierans (2007) and Rose and Sullivan (1996) both emphasize the importance of rethinking the use of descriptive data in social research, with Mair and Kierans focusing on the use of drawings and text and Rose and Sullivan providing a practical guide to descriptive and inferential data analysis.

For analyzing descriptive statistics, researchers commonly use statistical software such as SPSS (Statistical Package for the Social Sciences), SAS (Statistical Analysis System), or R. These software packages provide tools for calculating and interpreting central tendency measures, variability measures, and other descriptive statistics.

2.2. Inferential statistics

Inferential statistics is crucial in social science research, but its application requires careful consideration. Stoodley (1980) highlights the need for valid quantitative results, particularly in educational research. Adeyemi (2009) emphasizes the importance of using the appropriate test for a particular study, whether parametric or non-parametric. Gailmard (2014) provides a comprehensive overview of the role of inferential statistics in linking data-generating processes and observable data and in assessing claims about these processes. However, Gibbs et al. (2017) warn against inferential techniques' unnecessary and potentially misleading application, particularly in using administrative data in educational research.

Inferential statistics are used to make inferences and conclusions about a population based on a sample of data. This method is appropriate for quantitative data where the researcher wants to generalize findings to a larger population. For example, if a researcher collects data on the effectiveness of a new educational intervention on student performance, inferential statistics like t-tests or analysis of variance (ANOVA) can be used to determine if the intervention had a significant impact.

Statistical software like SPSS, SAS, R, and STATA are widely used for conducting inferential statistics. These tools offer a range of statistical tests and procedures for hypothesis testing, regression analysis, ANOVA, and other inferential analyses.

2.3. Qualitative analysis

Qualitative data analysis is a crucial aspect of social science research, influenced by the researcher's ontological perspective (Toker, 2022). It involves interpreting words and actions, requiring creativity to condense and organize the data without losing meaning (Clarke, 1999). While computer software can be used, manual analysis of interview transcripts is also feasible (Ajagbe et al., 2015). Qualitative research methodology has increased in recent years, with its ability to explore human behavior and improve organizational development (Mohajan, 2018).

Qualitative analysis is ideal for non-numeric data that cannot be easily quantified, such as interview transcripts or open-ended survey responses.

Researchers can use qualitative methods like content analysis or thematic analysis to identify patterns, themes, and relationships within the data. This method is valuable for gaining in-depth insights and understanding complex phenomena.

For qualitative analysis, researchers often use software programs such as NVivo, MAXQDA, or Atlas.ti. These qualitative analysis software tools provide features for coding, organizing, and analyzing non-numeric data like text, images, and videos.

2.4. Data mining

Data mining, a powerful tool for extracting valuable information from large datasets, has been increasingly utilized in social science research. Vedanayaki (2014) and Adedoyin-Olowe et al. (2013) highlight the application of data mining techniques in social network analysis, with the latter emphasizing the complexity of social network data and the need for computational analysis. Cao (2019) and Adedoyin-Olowe et al. (2013) further underscore the significance of data mining in this field, particularly in detecting trends, patterns, and rules.

Data mining suits large, complex datasets where the researcher wants to discover hidden patterns or relationships. This method is often used for quantitative marketing, finance, or healthcare data. Researchers can use algorithms like clustering or classification to extract valuable insights and predictions from the data.

Data mining software such as IBM SPSS Modeler, RapidMiner, or Weka are commonly used for analyzing large datasets and extracting meaningful patterns and insights. These software packages offer clustering, classification, association analysis, and predictive modeling algorithms.

2.5. Network analysis

Network Analysis is a valuable tool in social science research, particularly in social work (Rice & Yoshioka-Maxwell, 2015; Streeter & Gillespie, 1993). It can be used to identify mechanisms of social change and model these processes, leading to implications for practice and intervention efforts (Rice & Yoshioka-Maxwell, 2015). The technique complements standard analysis methods, broadening and deepening the application of theories in social work (Streeter & Gillespie, 1993). However, the approach has some deficiencies, particularly in science and technology studies (Shrum & Mullins, 1988). An extra application of network analysis includes influence maximization in social networks (Cuomo & Maiorano, 2018).

Bartholomew et al. (2002) and Rupp and Sweet (2011) further underscore the significance of multivariate methods in social science, focusing on correlational relationships and computer software packages. Rupp and Sweet (2011) also provide a comprehensive overview of various multivariate methods, making them accessible to practitioners with limited statistical knowledge.

Network analysis is appropriate for studying relationships and interactions between entities, such as social networks or communication networks. This method is suitable for both quantitative and qualitative data and can help researchers visualize and analyze the structure and dynamics of networks.

Software tools like Gephi, Pajek, or UCINET are commonly used for network analysis. These tools provide features for visualizing, analyzing, and modeling relationships between entities in a network. Scholars can use network analysis software to measure network centrality, identify clusters, and study network information flow.

2.6. Multivariate analysis

Multivariate analysis is practical when analyzing data with multiple variables or factors to understand complex relationships. This method is suitable for quantitative and qualitative data and can help researchers identify underlying patterns and associations among variables. Techniques like factor analysis or structural equation modeling can be used to analyze multivariate data effectively.

Statistical software packages like SPSS, SAS, R, STATA, and SmartPLS also support multivariate analysis techniques. Researchers can use these tools for conducting factor analysis, cluster analysis, structural equation modeling, and other multivariate analysis methods. These software packages provide features for analyzing relationships between multiple variables and identifying underlying patterns in data. SmartPLS uses the partial least squares method for structural equation modeling (SEM). Researchers interested in analyzing relationships between variables, testing complex models, or exploring latent variable relationships may find SmartPLS a valuable tool for data analysis.

2.7. Time series analysis

Time series analysis, particularly the Autoregressive integrated moving average (ARIMA) method, has been increasingly applied in social science research due to its accessibility and potential for impact assessment, causal modeling, and forecasting (Berk et al., 1980). This technique is handy in public opinion research, policy analysis, political science, economics, and sociology (Yaffee & McGeever, 2000). It is also a critical method in longitudinal designs, addressing various research questions and allowing for the analysis of treatment effects over time (Beran, 2003). The application of time series analysis in social science research is further expanded by modeling social dynamics and dynamic regression models (Box-Steffensmeier et al., 2014).

Time series analysis is designed for data collected over time, such as financial, weather, or stock market data. This method identifies trends and seasonal patterns and forecasts future values. Time series analysis techniques like trend analysis or ARIMA models can help researchers analyze and interpret temporal data effectively.

Time series analysis software like R, SAS, MATLAB, or EViews are commonly used for analyzing temporal data. These tools offer time series modeling, trend analysis, seasonal decomposition, and forecasting functions.

In conclusion, data analysis techniques are integral to social science research, providing researchers with the tools to analyze, interpret, and draw insights from their data. By effectively understanding and utilizing these techniques, researchers can generate robust findings, contribute to knowledge advancement in their field, and make informed decisions based on empirical evidence.

3. Challenges and pitfalls in data analysis

Data analysis in social science research presents various challenges and pitfalls that researchers must navigate to ensure the validity and reliability of their findings. Validity can be assimilated as the extent to which a research study reflects or measures the concepts it claims to measure. At the same time, reliability connotes the consistency and stability of the findings gained from a study (William, 2024). Gainey et al. (2021) highlight the need for better support and understanding of data literacy in undergraduate social science courses. Manian and Vadivel (2021) provide a broader perspective, discussing the history and significance of data science and its applications in various fields. Olteanu et al. (2016, 2019) underscore social data biases, methodological limitations, and ethical boundaries, urging a more rigorous approach to these issues. King (2011) emphasizes the need for improved data sharing, management, and statistical methodology in the social sciences.

In this section, we explore common challenges and pitfalls researchers often encounter during data analysis, such as data quality issues, sample size concerns, statistical assumption violations, and misinterpreting results.

3.1. Data quality issues

Data quality issues in social science research are multifaceted and can significantly impact the validity and reproducibility of findings. It may include missing data, outliers, data entry errors, or inconsistencies in the data. Srivastava and Mishra (2021) highlight data quality challenges in social media research, emphasizing the need for a comprehensive framework to address these issues. Krejčí (2010) underscores the importance of a holistic approach to survey data quality, which includes controlling the entire survey process and considering both sampling and nonsampling errors. Shanks and Corbitt (1999) extend this discussion by emphasizing social and cultural aspects of data quality, particularly in data warehousing and enterprise resource planning systems. Cook et al. (2021) further explore the specific challenges of social determinants of data quality in the electronic health record, including plausibility and bias issues. Researchers need to carefully clean and preprocess the data to address these issues before proceeding with the analysis. Some respondents may have skipped specific questions in a survey study, leading to missing data. Researchers must decide whether to impute missing values or exclude incomplete responses before analysis.

3.2. Sample size and representativeness

Researchers may encounter challenges related to sample size and representativeness of their data. Small sample sizes can limit the generalizability of the findings, while non-representative samples may introduce bias and affect the validity of the results. Zhao (2020) highlights the problem of non-representative samples, particularly over-reliance on undergraduate students from Euro-American institutions. It is further complicated by the methodological challenges in assessing sample quality, as Bartholomew et al. (2021) discussed in the context of phenomenological research. Nanjundeswaraswamy and Divakar (2021)

emphasize the importance of precise sampling techniques and appropriate sample size estimation, particularly in health science research. Tufekci (2014) adds a layer of complexity by discussing the challenges of representativeness and validity in big data analyses of social media.
Researchers should carefully consider the sample size’s adequacy and the sample population's representativeness. A researcher conducting a study on consumer behavior with a small sample size of 30 individuals may struggle to generalize the findings to the entire population of interest, such as all consumers in a specific demographic.

3.3. Statistical assumptions violation

The violation of statistical assumptions in social science research can have significant implications for the validity of the results. Yu (2002) highlights the common occurrence of these violations in social science data and discusses remedial tools to address them. Silver (2021) further emphasizes the need to understand the impact of these violations on statistical estimates, using simulated demonstrations to illustrate their effects. Osborne and Waters (2002) specifically focus on the assumptions of multiple regression, underscoring the importance of testing these assumptions and discussing those that could be more robust to violation.
Many statistical analysis techniques are based on assumptions that must be met for the results to be valid. Researchers may encounter challenges when these assumptions are violated, leading to inaccurate or misleading results. Common assumptions include normality, homoscedasticity, independence, and linearity.
In linear regression analysis, researchers may violate the assumption of homoscedasticity if the residuals exhibit a pattern of increasing variance with higher predicted values (William, 2024), indicating that the variance of the errors is not constant.

3.4. Overfitting and underfitting

Overfitting occurs when a statistical model is too complex and captures noise in the data, leading to poor generalizability. Conversely, underfitting occurs when a model is too simple and fails to capture the underlying patterns in the data. Finding the right balance between overfitting and underfitting is a common challenge in data analysis.
Overfitting and underfitting are significant issues in social science research, particularly in regression-type models (Babyak, 2004). Overfitting can be addressed through early stopping, network reduction, data expansion, and regularization (Ying, 2019). However, these issues are also influenced by the prevailing scientific research paradigm, which can be overly restrictive and hinder the exploration of essential questions and data (Heineman, 1981). Predictive overfitting can be countered in immunological applications through model complexity reduction, reliable model evaluation, and data diversity (Gygi et al., 2023).
A researcher developing a machine learning model to predict stock prices may overfit the historical data by including too many predictors, resulting in a model that performs well on training data but fails to generalize to new, unseen data.

3.5. Misinterpretation of results

Researchers may misinterpret the results of their analysis due to a lack of statistical knowledge or understanding of the data. It is essential to carefully interpret the results, consider the limitations of the analysis, and avoid drawing unwarranted conclusions based on the findings.
The misinterpretation of results in social science research is a complex issue with various contributing factors. Lehrer et al. (2007) highlight the need for a platform to disseminate negative results, which are often overlooked but can provide valuable insights. Brenner (1981) emphasizes the importance of considering the social and psychological conditions in data collection, as these can introduce biases. Boruch and Rui (2008) discuss the challenges in evidence-based practice, including the potential for misinformation and the need for users to understand and use evidence effectively. Edwards (2005) underscores the divide between

researchers and policy practitioners, with differing perspectives and expectations often leading to misinterpretation of research findings.

A study examining the relationship between exercise and weight loss may erroneously conclude that exercise causes weight loss when the correlation observed is merely associative and does not imply causation.

3.6. Multiple comparisons

Conducting multiple statistical tests increases the likelihood of obtaining false-positive results. Researchers need to adjust for multiple comparisons or use appropriate corrections to control the family-wise error rate and minimize the risk of Type I errors.

Type I errors are a significant concern, with Lachlan and Spence (2005) highlighting the widespread neglect of corrections for Type I errors in complex designs. O'Keefe (2003) argues against adjusting the alpha level for the number of tests, suggesting that it is not a principled approach and reduces statistical power. However, Keselman et al. (2002) and Holland et al. (2009) emphasize the importance of controlling the Type I error rate, particularly in multiple tests. Keselman et al. (2002) recommend using false discovery rate methods, while Holland et al. (2009) suggest a compromise approach for accounting research.

A researcher testing the effectiveness of different marketing strategies on sales may conduct multiple t-tests without adjusting for multiple comparisons, leading to an increased risk of finding significant results by chance.

3.7. Data visualization and communication

Data visualization and communication are crucial in social science research, particularly in analyzing social media data (Jensen et al., 2021). Reyes (2022) emphasizes the importance of combining creativity and technique in producing compelling visualizations, providing a practical guide for researchers. Healy and Moody (2014) highlight the need for a higher standard in the graphical display of sociological insights, particularly in the context of easier data sharing. Lastly, Andrienko and Andrienko (1997) underscore the role of intelligent software in generating appropriate data presentations, focusing on the different design principles for data exploration and communication.

Presenting data analysis results clearly and concisely can be a challenge for researchers. Effective data visualization techniques and clear communication of the findings are essential for conveying the results to a broader audience and ensuring the impact of the research.

A researcher presenting complex statistical results in a research paper without using appropriate visual aids, such as graphs or tables, may make it difficult for readers to interpret and understand the findings effectively.

In sum, researchers must be aware of the challenges and pitfalls that can arise during data analysis in social science research. By addressing issues such as data quality, sample size considerations, statistical assumptions, and result interpretation, researchers can enhance the credibility and rigor of their research findings. Understanding and mitigating these challenges is essential for producing valid and reliable conclusions in social science research.

4. Practical tips and guidelines for effective data analysis

Data analysis is a critical component of the research process that requires attention to detail, analytical skills, and careful consideration of various factors to ensure the validity and reliability of the findings. This section provides helpful tips and guidelines to help researchers successfully navigate the data analysis process. By following these recommendations, researchers can enhance the quality and validity of their data analysis and findings.

4.1. Define clear research objectives

Clearly define the research objectives and specific research questions to guide the data analysis process. A clear understanding of the research goals will help researchers select appropriate analytic techniques and interpret the results effectively.

4.2. Plan the data analysis approach

Develop a data analysis plan outlining the steps, procedures, and techniques. Consider the type of data collected, the research design, and the desired outcomes to ensure a systematic and organized approach to data analysis.

4.3. Familiarize yourself with the data

Before conducting the analysis, thoroughly review and familiarize yourself with the data. Understand the variables, data structure, and any missing values or outliers in the dataset to make informed decisions during the analysis.

4.4. Use descriptive statistics

Start the data analysis process by performing descriptive statistics to summarize and explore the data. Utilize measures such as means, frequencies, and standard deviations to gain insights into the distribution and characteristics of the variables.

4.5. Apply statistical techniques appropriately

Choose the statistical techniques and tests most appropriate for the research questions and the data analysis type. Consider factors such as the level of measurement, assumptions of the tests, and the nature of the relationships being examined.

4.6. Document the analysis process

Maintain detailed documentation of the data analysis process, including the steps taken, decisions made, and any transformations or manipulations applied to the data. Documenting the analysis process ensures transparency and reproducibility of the results.

4.7. Interpret and communicate the results

Interpret the analysis results in the context of the research objectives and hypotheses. Communicate the findings using visual aids, tables, and figures to make the results accessible and understandable to a broader audience.

4.8. Seek feedback and validation

Involve peers, mentors, or collaborators in reviewing and validating the data analysis process and results. Seeking feedback helps identify potential errors, biases, or alternative interpretations that enhance the rigor of the analysis.

In summary, incorporating practical tips and guidelines for data analysis is crucial for researchers to ensure the robustness and accuracy of their research outcomes. By defining clear research objectives, planning the analysis approach, using appropriate statistical techniques, documenting the process, interpreting results, seeking validation, and engaging in effective communication, researchers can conduct data analysis effectively and contribute to advancing knowledge in the social sciences.

5. Conclusion

This paper provides a practical guide to help researchers effectively analyze their data in social science research. Researchers can confidently and precisely navigate the process by exploring commonly used data analysis techniques, highlighting challenges and pitfalls, and offering valuable tips and guidelines.

The significance of this study lies in its contribution to filling a gap in the existing literature by providing a comprehensive guide to data analysis in social science research. This study aims to improve the quality and reliability of research findings in the social sciences by equipping researchers with the necessary knowledge and skills to analyze their data effectively.

However, it is essential to acknowledge the limitations of this study. The guide provided in this paper is by no means exhaustive, and researchers may encounter other data analysis techniques and challenges. Future research could focus on exploring advanced data analysis methods, addressing specific data quality issues, and delving deeper into the nuances of interpreting the results of complex analytical procedures.

Overall, this study lays the groundwork for enhancing the rigor and quality of data analysis in social science research, ultimately advancing knowledge and understanding in various fields of study. Researchers can successfully navigate the data analysis process and produce robust research findings by continually refining analytical skills, adopting best practices, and seeking expert guidance.

Acknowledgment

The author is grateful to each person who supported the redaction of this work.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflicts of interest.

References

1. Adedoyin-Olowe, M., Gaber, M.M., & Stahl, F.T. (2013). A Survey of Data Mining Techniques for Social Media Analysis. *J. Data Min. Digit. Humanit., 2014*.
2. Adeyemi, T.O. (2009). Inferential Statistics for Social and Behavioural Research.
3. Ajagbe, A.M., Sholanke, A.B., Isiavwe, D.T., & Oke, A.O. (2015). Qualitative Inquiry for Social Sciences.
4. Andrienko, G.L., & Andrienko, N.V. (1997). Research issues in intelligent data visualisation for exploration and communication. *CHI '97 Extended Abstracts on Human Factors in Computing Systems*.
5. Aravindh, R., & Thirupathi, S. (2019). Data Analysis In Social Sciences: Comparison between Quantitative and Qualitative Research. *International Journal of Management Research and Social Science*.
6. Babyak, M.A. (2004). What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. *Psychosomatic Medicine, 66*, 411-421.
7. Ball, G.H. (1965). Data analysis in the social sciences: what about the details? *AFIPS '65 (Fall, part I)*.
8. Bartholomew, D.J., Moustaki, I., Galbraith, J.I., & Steele, F. (2002). The Analysis and Interpretation of Multivariate Data for Social Scientists.
9. Bartholomew, T.T., Joy, E.E., Kang, E., & Brown, J. (2021). A choir or cacophony? Sample sizes and quality of conveying participants' voices in phenomenological research. *Methodological Innovations, 14*.
10. Beran, J. (2003). Time series analysis.
11. Berk, R.A., McCleary, R., & Hay, R.J. (1980). Applied Time Series Analysis for the Social Sciences.

12. Blank, C., McBurney, M., Morgan, M., & Seetan, R.I. (2021). A Survey of Big Data Techniques for Extracting Information from Social Media Data. *Advances in Science, Technology and Engineering Systems Journal, 6*, 189-204.
13. Boruch, R.F., & Rui, N. (2008). From randomized controlled trials to evidence grading schemes: current state of evidence-based practice in social sciences. *Journal of Evidence-Based Medicine, 1*.
14. Box-Steffensmeier, J. M., Freeman, J. R., Hitt, M. P., & Pevehouse, J. C. (2014). *Time series analysis for the social sciences*. Cambridge University Press.
15. Brenner, M. (1981). Problems in collecting social data: a review for the information researcher. *Social Science Information Studies, 1*, 139-151.
16. Bryman, A.E., & Cramer, D. (1990). Quantitative data analysis for social scientists.
17. Cao, K.S. (2019). Data mining for social network data.
18. Chen, Y., Sherren, K., Smit, M., & Lee, K.Y. (2021). Using social media images as data in social science research. *New Media & Society, 25*, 849 - 871.
19. Clarke, A.M. (1999). Qualitative research: data analysis techniques. *Professional nurse, 14 8*, 531-3.
20. Cook, L.A., Sachs, J., & Weiskopf, N.G. (2021). The quality of social determinants data in the electronic health record: a systematic review. *Journal of the American Medical Informatics Association: JAMIA, 29*, 187 - 196.
21. Cuomo, S., & Maiorano, F. (2018). Social network data analysis and mining applications for the Internet of Data. *Concurrency and Computation: Practice and Experience, 30*.
22. Edwards, M. (2005). Social Science Research and Public Policy: Narrowing the Divide1. *Australian Journal of Public Administration, 64*, 68-74.
23. Gailmard, S. (2014). Statistical Modeling and Inference for Social Science.
24. Gainey, M.A., Gunderman, H.C., Slayton, E., & Splenda, R. (2021). Understanding the Practices and Challenges of Teaching with Data in Undergraduate Social Science Courses at Carnegie Mellon University.
25. Gibbs, B.G., Shafer, K., & Miles, A.R. (2017). Inferential statistics and the use of administrative data in US educational research. *International Journal of Research & Method in Education, 40*, 214 - 220.
26. Gygi, J.P., Kleinstein, S.H., & Guan, L. (2023). Predictive overfitting in immunological applications: Pitfalls and solutions. *Human Vaccines & Immunotherapeutics, 19*.
27. Hall, O. (2010). Remote Sensing in Social Science Research. *The Open Remote Sensing Journal, 3*, 1-16.
28. Healy, K., & Moody, J. (2014). Data Visualization in Sociology. *Annual review of sociology, 40*, 105-128.
29. Heineman, M.B. (1981). The Obsolete Scientific Imperative in Social Work Research. *Social Service Review, 55*, 371 - 397.
30. Holland, B.S., fang, S., & Basu, S. (2009). Neglect of Multiplicity When Testing Families of Related Hypotheses*. *Research Methods & Methodology in Accounting eJournal*.
31. Hou, Q., Han, M., & Cai, Z. (2020). Survey on data analysis in social media: A practical application aspect. *Big Data Min. Anal., 3*, 259-279.
32. Jensen, J.B., Singh, L., Davis-Kean, P.E., Abraham, K., Beatty, P.C., Bode, L., Chau, D., Eliassi-Rad, T., Gonzalez, R., Hamilton, R.W., Kim, I.S., Kuchler, T., Ladd, J.M., Lerman, K., Levenstein, M.C., Mneimneh, Z., Nguyen, Q.C., Pasek, J., Raghunathan, T.M., Ryan, R., Soroka, S., Tadesse, M.G., & Traugott, M.W. (2021). Analysis and Visualization Considerations for Quantitative Social Science Research Using Social Media Data.
33. Keselman, H.J., Cribbie, R.A., & Holland, B.S. (2002). Controlling the rate of Type I error over a large set of statistical tests. *The British journal of mathematical and statistical psychology, 55 Pt 1*, 27-39.

34. King, G. (2011). Ensuring the Data-Rich Future of the Social Sciences. *Science, 331*, 719 - 721.
35. Krejčí, J. (2010). Approaching Quality in Survey Research: Towards a Comprehensive Perspective. *Sociologicky Casopis-czech Sociological Review, 46*.
36. Lachlan, K.A., & Spence, P.R. (2005). Corrections for Type I Error in Social Science Research: A Disconnect between Theory and Practice. *Journal of Modern Applied Statistical Methods, 5*, 23.
37. Lehrer, D., Leschke, J., Lhachimi, S.K., Vasiliu, A.M., & Weiffen, B. (2007). negative results in social science. *European Political Science, 6*, 51-68.
38. Mair, M., & Kierans, C. (2007). Descriptions as data: developing techniques to elicit descriptive materials in social research. *Visual Studies, 22*, 120 - 136.
39. Manian, V., & Vadivel, P. (2021). Challenges and Applications of Data Analytics in Social Perspectives: Introduction of Data Science. *Advanced Deep Learning Applications in Big Data Analytics*, 51-67.
40. Maravelakis, P.E. (2019). The use of statistics in social sciences.
41. Mohajan, H.K. (2018). Qualitative Research Methodology in Social Sciences and Related Subjects. *Journal of Economic Development, Environment and People, 7*, 23-48.
42. Nanjundeswaraswamy, T. S., & Divakar, S. (2021). Determination of sample size and sampling methods in applied research. *Proceedings on engineering sciences*, *3*(1), 25-32.
43. O'Keefe, D.J. (2003). Colloquy: Should Familywise Alpha Be Adjusted? Against Familywise Alpha Adjustment.
44. Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data, 2*.
45. Osbourne, J.W., & Waters, E. (2002). Four Assumptions of Multiple Regression That Researchers Should Always Test. *Practical Assessment, Research and Evaluation, 8*, 2.
46. Reyes, J. M. M. (2022). *Data Visualization for Social and Policy Research: A Step-by-step Approach Using R and Python*. Cambridge University Press.
47. Rice, E., & Yoshioka-Maxwell, A. (2015). Social Network Analysis as a Toolkit for the Science of Social Work. *Journal of the Society for Social Work and Research, 6*, 369 - 383.
48. Rogers, B. (1998). Descriptive Analysis of Research Data. *Workplace Health & Safety, 46*, 266 - 267.
49. Rose, D.F., & Sullivan, O. (1996). Introducing data analysis for social scientists.
50. Rupp, A.A., & Sweet, S.J. (2011). Analysis of Multivariate Social Science Data (2nd ed.). *Structural Equation Modeling: A Multidisciplinary Journal, 18*, 686 - 693.
51. Shanks, G., & Corbitt, B. (1999, December). Understanding data quality: Social and cultural aspects. In *Proceedings of the 10th Australasian conference on information systems* (Vol. 785). Victoria University of Wellington, New Zealand.
52. Shrum, W., & Mullins, N.C. (1988). NETWORK ANALYSIS IN THE STUDY OF SCIENCE AND TECHNOLOGY.
53. Silver, I.A. (2021). The Violating Assumptions Series: Simulated demonstrations to illustrate how assumptions can affect statistical estimates. *arXiv: Methodology*.
54. Srivastava, A.K., & Mishra, R. (2021). Analyzing Social Media Research: A Data Quality and Research Reproducibility Perspective. *IIM Kozhikode Society & Management Review, 12*, 39 - 49.
55. Stoodley, K.D. (1980). Statistical Inference in the Social Sciences. *Educational Research, 23*, 51-56.
56. Streeter, C.L., & Gillespie, D.F. (1993). Social Network Analysis. *Journal of Social Service Research, 16*, 201-222.
57. Stukal, D.K., Belenkov, V.E., & Philippov, I. (2021). Data science methods in political science research: analyzing protest activity in social media. *Political Science (RU)*.

58. Tindall, D.B., McLevey, J., Koop-Monteiro, Y., & Graham, A.V. (2022). Big data, computational social science, and other recent innovations in social network analysis. *Canadian review of sociology = Revue canadienne de sociologie*.
59. Toker, A. (2022). A GUIDE FOR QUALITATIVE DATA ANALYSIS IN SOCIAL SCIENCES. *Pamukkale University Journal of Social Sciences Institute*.
60. Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *ArXiv, abs/1403.7400*.
61. Vedanayaki, M. (2014). A Study of Data Mining and Social Network Analysis. *Indian journal of science and technology, 7*, 185-187.
62. William, F.K.A. (2024). Mastering Validity and Reliability in Academic Research: Meaning and Significance. *International Journal of Research Publications, 144*(1), 287-292.
63. William, F.K.A. (2024). Understanding Endogeneity, Exogeneity, Heterogeneity, Homogeneity, Homoskedasticity, Heteroskedasticity in Statistical Analysis: Avoiding Misinterpretations in Social Science Research. *International Journal of Research Publications*, *143*(1).
64. Yaffee, R.A., & McGee, M. (2000). Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS.
65. Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*.
66. Yu, C. H. (2002). An overview of remedial tools for violations of parametric test assumptions in the SAS system. In *Proceedings of 2002 western users of sas software conference* (Vol. 172178).
67. Zhao, K. (2020). Sample representation in the social sciences. *Synthese*, 1-19.
68. Zwilling, M. (2023). Big Data Challenges in Social Sciences: An NLP Analysis. *Journal of Computer Information Systems, 63*, 537 - 554.