Elizabeth Binkina

Dr. Kristi McDuffie

RHET 233

July 09 2023

Statistical Research Methods in Different Genres of Writing

Statisticians work in a wide range of fields and professions conducting various types of research, analyzing data, and writing reports in all forms and genres. While statisticians perform similar statistical tests and use the same methodology to analyze data, the differentiating factor between fields of research lies where they obtain their data and how they present their findings and answer their research question in their final reports. Reports on statistical analysis are written in various genres to cater to different audiences with varying levels of understanding of specific topics. My first artifact is a medical peer reviewed article written by academics that explains how machine learning algorithms and logistic regression, both types of statistical research techniques, can better predict the main contributor to Alzheimer's disease. My second artifact is a government statistical journal that uses government data and their own statistical research methods to predict and analyze global trading and economic patterns. My third artifact is a policy research document that explains how a randomized control experiment, another common statistical technique, is used to identify if a violence prevention program is effective. In my paper, I explore how these authors write their reports on statistical research differently in a variety of fields, even though they use similar statistical methodologies.

My first artifact, titled "*Machine Learning prediction of incidence of Alzheimer's disease using large-scale administrative health data*" , is written by Ji Hwan Park, Han Eol Cho, Jong Hun Kim, Melanie M. Wall, Yaakov Stern, Hyunsun Lim, Shinjae Yoo, Hyoung Seop Kim, and

Jiook Cha. It was published by NPC (nature partner journals) Digital Medicine in 2020. The article focuses on using machine learning to predict Alzheimer's disease based on certain variables such as age, sex, fitness level, etc. Park et al. collected a vast amount of data from health records and utilized machine learning algorithms and logistic regression to identify the variables that contribute the most to the development of Alzheimer's. This article also demonstrates that these methods can be applied to other diseases such as diabetes, suicide death, overdoses, etc. Another important aspect of this article is to inform readers that these statistical techniques lead to the development of "better therapeutic strategies for delaying the onset of AD"(Park et.al 2) As a result, the audience gains a sense of hope that these research methods can help millions of people with these incurable diseases.

It is evident that this is a peer reviewed article because it underwent evaluation by scholarly and accredited professionals in the medical field, as mentioned in the acknowledgments section (Park et al. 6). The article includes thirty-eight references to scholarly articles, data files, published studies, and peer reviewed journals written by accredited doctors and researchers.(Park et al. 6) Park et al. include multiple tables, diagrams, and graphs along with explanations of the data analysis. For instance, Table 3 (Park et al. 4) presents an output table of a logistic regression with ten predictor variables. Logistic regression is a powerful statistical technique used to accurately determine the most effective or statistically significant variables in predicting a result. In this article, the regression led to the conclusion that "hemoglobin level as the most significant predictor" (Park et al. 3) The purpose of this article is to present the results of machine learning algorithms and other statistical techniques in predicting variables that contribute the most to the development of dementia and Alzheimer's. This information is crucial because healthcare professionals reading this paper now know that a patient's hemoglobin level is the most

significant predictor when assessing their likelihood of developing Alzheimer's. The intended audience of this article consists of healthcare professionals, particularly those specializing in research, as Park et al. utilize specific medical terminology and formatting familiar to doctors and researchers. Given the usage of advanced statistical terms like AUC scores, p-values, ICD-10, the article may also cater to individuals familiar with machine learning, data analytics, and statistical methodology.

In this article, Park et al. perform their analysis using another study's database. They acknowledge and give credit to this data by incorporating it into their references list, and referring to it in their writing. They make it clear that their analysis and models are still their own work. The use of these statistical techniques is incredibly beneficial and useful because " Machine learning is an optimal choice of analytics for analyzing the large-scale administrative health data containing thousands of descriptors from hundreds of thousands of individuals…Given the recent rapid growth of the machine learning technology, application of the AI technology to clinical predictive modeling is likely to have a deep impact on medicine" (Park et al. 1) Machine learning algorithms are developing rapidly, becoming more and more efficient every day, while researchers also collect more data. Consequently, results will continuously become more accurate. Therefore, this article does a great job of convincing readers that this statistical technology is the future of medicine.

The authors establish their ethos and credibility by sharing their affiliations and acknowledgments. Some of them are established doctors with degrees from prestigious universities, specializing in rehabilitation, dementia, brain and cognitive sciences, neurology, and psychiatry. Since dementia and Alzheimer's disease are relevant to all of these fields, readers can be assured that Park et al. possess adequate comprehension in this field. Additionally, some

authors are highly experienced data scientists and researchers, further measuring the audience that the regression models and tables are accurately calculated and interpreted. Moreover, several authors have received impressive awards for their work in research and investigation, which adds to the credibility of this research.

Overall, this genre of medical research articles follow a fact and evidence based format. They begin by presenting a theory and a list of possible contributory variables that they hypothesize might affect the overall progression of Alzheimer's, along with a database containing the data with these specific predictors. Then they formulate the evidence by utilizing machine learning and logistic regression algorithms to obtain the results. Next, they present their findings in the form of tables, data, and charts. Lastly, they employ numbers and statistical measures in charts and tables for the analysis and interpretation of these results.

The next artifact is a government statistical trade journal , titled "*World Trade Statistical Review*" that discusses trends in world trade, commercial services, trade growth, and merchandise trade and exports in developing economies. It is directed by Hubert Barbosa and Andreas Maurer who are the WTO (World Trade Organization) Chief Statistician and senior trade official for the WTO statistical program, respectively. The journal was published by the WTO in 2016.

This journal extensively employs statistical language to interpret numerous high-detail graphs and charts that depict trading patterns over several years between different countries and more. It is considered a statistical journal due to its high volume of visual evidence, including detailed, colorful, and descriptive bar graphs and charts containing both numeric and categorical on world trade statistics. The journal's colorful and easily interpretable information, along with descriptions accompanying each visual, enhance reader comprehension of the graphs'

representations and enable interpretation of the results. Such characteristics are a common attribute of journals in this field.

Barbosa et al. present visual graphs and provide explanations regarding their meaning, data sources, usage, and implications. These graphs offer the audience a clear visualization of the variations in trade value among different countries. The journal follows a structured format, including a table of context that describes the specific trade types covered in each section. This allows readers to easily identify and navigate to the trade areas they are mostly interested in. looking at. Additionally, the journal features an introduction with acknowledgements and it is divided into nine different sections, each focusing on different topics. For example, in the Global Value Chains section, the authors were trying to "outline trade flows taking place within the global value chains (GVCs)...that covers all sectors of the economy, including primary products, manufactured goods and services…The chemicals industry is a sector dominated by supply chains. The chemicals market faces strong international competition, and companies make use of supply chains to optimize their production or services and to make cost savings." (Barbosa et al. 41). Global value chains involve intricate processes where different stages of production take place across multiple countries to maximize profitability.

By comprehending this trade flow algorithm within the global economy, the audience of this journal can obtain valuable information regarding countries that can achieve maximum efficiency in product production and generate optimal profits. This journal specifically caters to business owners involving manufacturing, as it guides them in making informed financial decisions and selecting the most suitable countries for exporting their goods. To illustrate the functionality of this flow algorithm, Barbosa et al. use the world chemical market as an example, considering it is highly competitive. They provide a comprehensive bar graph (Barbosa et al. 42)

displaying the domestic value share of chemical exports for each country. The graph reveals that Brazil, the EU, and the US exhibit the largest domestic value added, while Singapore and Korea have comparatively lower values. Through interpreting this graph, readers can discern which countries excel in maximizing profits within the chemical market.

Barbosa et al. establish their ethos by including an acknowledgements section that provides readers with insight into their extensive work and research undertaken in the development of this journal. The team comprises individuals involved in the preparation of tables and charts, report production, statistical research, data compilation, and estimation. These tasks are carried out by highly skilled professionals in the field of economics, trading, and statistics. The directors also acknowledge the utilization of statistics and research obtained from private and national institutions, as well as data sets from UNCTAD and ITC concerning services and trade. Additionally, directors mention that the journal underwent peer review by WTO (World Trade Statistical Review) Statistics users.(Barbosa et al. 4)

Another pattern utilized by Barabosa et al. is the comparison and contrast of data over an extended period to inform readers about currency trends . For example, “The Chinese yuan appreciated along with the dollar, rising 10 per cent on average in 2015 and 13 per cent between June 2014 and December 2015, due to the Chinese currency’s quasi-peg to the US dollar at the time. The appreciation of the yuan may have contributed to the economic slowdown in China to the extent that it made Chinese exports more expensive in foreign markets.”(Barbosa et.at 19). Thus, changes in the currency of one country can significantly impact global markets and the currency value of other countries. This analysis is particularly valuable for business owners, especially those involved in international business and commerce. It allows them to understand the constantly changing nature of the markets and enables them to adjust their business strategies

and production algorithms to remain profitable. By providing these statistics, the journal assists business owners in adapting their trading algorithms and strategies based on the international market, saving them from financial losses. The effectiveness of this journal lies in its ability to analyze a vast amount of government data and predict trends in global markets, thereby aiding individuals in avoiding excessive trading costs in a rapidly changing economy.

The last artifact is a policy research document titled "*Have US-Funded CARSI Programs Reduced Crime and Violence in Central America? An Examination of LAPOP'S Impact Assessment of US Violence Prevention Programs in Central America*" . It is written by David Rosnick, Alexander Main, and Laura Jung. It was published by the CEPR (Center for Economic and Policy Research) in September 2016. The document presents a study on community-based violence prevention programs implemented by the CARSI (US State Department's Central American Regional Security Initiative). The study utilizes survey data to assess public perception of crime in various treatment and control neighborhoods in several countries where this prevention program has been implemented. The study practices a common statistical technique known as a double blind randomized experiment to determine if a variable affects a particular group of people more than the other. In this case, the researchers randomly assign a population into two different groups: control and treatment. The control group does not receive treatment, meaning the prevention program is not implemented, while the treatment group is the only one where the prevention program is.

The purpose of this document is to provide evidence of the study's success through the use of statistical research techniques. According to the LAPOP survey, which assessed the communities perception of crime in both treated and control groups across 4 different countries, the prevention programs were found to be successful. According to their study, fifty-one percent

less community members of the treated group in the community suggested they were aware of murder and extortion during the past year, and nine-teen percent fewer community members have reported hearing about robberies in their neighborhoods. (Rosnick et al. 2). Rosnick highlights these findings to support the argument that the prevention programs effectively reduced crime rates. Rosnick et al. uses this statistical survey data to convince the audience, likely consisting of other governments and policymakers interested in implementing similar programs, of the program's benefits in reducing crime rates. Another pattern observed in this document is the inclusion of several statistical regression plots, accompanied by explanations of their meaning and relevance to the study. For instance, in Figure 2 (Rosnick 8) displays the output of a ggplot in the statistical software R, predicting the post-treatment outcomes based on the pre-treatment data. The light blue points represent the treatment group, with approximately ten out of fourteen points lying below the estimated regression equation. This indicates a significantly lower robbery rate than predicted , providing evidence of the program effectiveness. On the other hand, about half of the dark blue points, representing the control group, lie above the regression equation, suggesting that the crime rate did not improve as much as it did in the treatment group. By utilizing this graph, the authors present a clear visual distinction between the groups, thereby persuading the audience of the program's efficiency. Additionally, Rosnick et al. provide simplified explanations of the graph's interpretation, making it accessible to policymakers who may have no expertise in statistical analysis and allow for them to see and understand the evidence presented.

It is clear that this is a policy research document due to its format. The format is a theoretical framework that discusses the effectiveness of a government implemented program, chart or table presenting data from the statistical experiment, analysis and description of each

figure, and ending with an explanation. Another strong indicator that this is a research document focused on policy is the fact that it is published and written by the CEPR, which is a private, non-profit organization that specifically publishes to policy-makers or government employees so they can read, learn, and potentially implement these practices and programs. The authors, David Rosnick, Alexander Main, and Laura Jung, establish their ethos by stating their credentials on the bottom of the title page (Rosnick et al. 1) David Rosnick is an economist at the CEPR in D.C, Alexander Main is a Senior Associate for International Poly, and Laura Jung is an International Program Intern at CEPR. This provides the audience with a sense of credibility and trust, suggesting that Rosnick et al. are presenting accurate information based on properly implemented statistical research models and graphs.

Overall, these three artifacts exhibit both differences and similarities.While each genre utilizes data from different sources, they all employ statistical research methods to conduct their analysis. The authors of the peer reviewed article used data from other published papers, medical documents, and databases, using it to conduct their own analysis. The journal used government data on trading and economic trends to perform their own analysis using their unique set of skills. Similarly, the policy research document relied on government data to perform their analysis. In both the scholarly article and the policy research document, researchers used logistic regression to identify the most influential variables in their respective research areas. All three used extensive amounts of bar graphs, charts, and tables that each had a correlated explanation and analysis. Thus, although researchers in all different fields collect data differently, they rely on similar statistical research techniques and methods to address their research questions. Another notable similarity among the genres is the presence of evidence-based arguments. Each genre presents statistical analysis tests along with thorough explanations, demonstrating how the

findings contribute to answering the research question. The most significant difference between these genres lies in their use of language and terminology tailored to specific audiences in each respective field. In the genres explored, these include the medical field, trading and international business field, and the government and policy field.

Work Cited

1. Park, J.H., Cho, H.E., Kim, J.H. *et al.* Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *npj Digit. Med.* 3, 46 (2020). https://doi.org/10.1038/s41746-020-0256-0

International Trade Statistics (World Trade Organization). Geneva: World Trade Organization, 2000. Print. *https://www.wto.org/english/res_e/statis_e/wts2016_e/wts16_toc_e.htm*

ROSNICK, D.; MAIN, A.; JUNG, L. Have US-Funded CARSI Programs Reduced Crime and Violence in Central America? An Examination of LAPOP'S Impact Assessment of US Violence Prevention Programs in Central America. *[s. l.]*, p. 22 pages, 2016. Disponível em: https://search-ebscohost -com.proxy2.library.illinois.edu/login.aspx?direct=true&db=eoh&AN=1599368&site=eds-live& scope=site. Acesso em: 27 jun. 2023. https://cepr.net/images/stories/reports/carsi-2016-09.pdf