

An Examination of Changes in Analysis Techniques across Fields

Bzdok Nomi*

Department of Biomedical Engineering, McConnell Brain Imaging Centre (BIC), Montreal Neurological Institute (MNI), Faculty of Medicine, McGill University, Montreal, Canada, Mila—Quebec Artificial Intelligence Institute, Montreal, Canada

Abstract

The biomedical, life, and social (BLS) sciences are experiencing a rapid development of data analysis techniques. Meanwhile, there is growing worry that students are not being sufficiently prepared for modern research through quantitative techniques teaching. Demands for changes to undergraduate and graduate quantitative research method courses have been sparked by these trends. We contend that such reform should be founded on data-driven insights concerning the application of analytical techniques within and across disciplines. About 1.3 million publically accessible research articles were examined as part of our assessment of the peer-reviewed literature to track the interdisciplinary mentions of analytical methodologies over the previous ten years. In order to detect trends in analytic method mentions shared across disciplines as well as those specific to each discipline, we used data-driven text mining analyses to the “Methods” and “Results” sections of a significant subset of this corpus. We discovered that the most frequently mentioned statistical techniques in research articles in the fields of biomedicine, life science, and social science are the t test, analysis of variance (ANOVA), linear regression, and chi-squared test. Between 2009 and 2020, however, the proportion of published literature that mentioned these techniques fell. On the other hand, the overall share of scientific publications has significantly increased for multivariate statistical and machine learning approaches, including artificial neural networks (ANNs). Additionally, we discovered distinct clusters of analytical techniques related to each BLS research field, such as the application of structural equation modelling (SEM) in psychology, survival models in cancer, and multiple learning in ecology. We talk about how these results affect research techniques and statistics education, as well as disciplinary and inter-disciplinary collaboration.

Keywords: Biomedical; Analysis Techniques; Methodological

Introduction

The biomedical, biological, and social sciences’ (BLS) methodological environment is getting more complicated. The emergence of open-source science, the accessibility of substantial, complex datasets, and rising computational power are the primary causes of this rising complexity. When it comes to preparing researchers for the era of big data, machine learning, and open-source software, it is sometimes felt that the traditional statistical tools (such as the t test, analysis of variance (ANOVA), and linear regression) taught in introductory statistics courses are insufficient. Many researchers and statisticians have argued for educational change to introductory research methods and statistics courses because they are concerned that BLS [1-3] sciences educational programme is finding it difficult to stay up with these trends. We contend that a deeper comprehension of current patterns in the application of analytical methods throughout BLS sciences is an essential first step in this direction. Such an understanding will provide useful insights into the methodological abilities and information required to prepare early career scientists for success in their fields and in multidisciplinary collaborations in the future. Analytical techniques created for one subject are increasingly being successfully applied to another. Biologists have successfully employed deep learning, a machine learning approach created by the artificial intelligence community, to predict the three-dimensional structure of proteins, for instance. The rapid adoption of neural networks in biology and many other disciplines demonstrates the importance of educational preparation in recognising these trends and keeping up with the requirement for proficiency in these new advanced analytical techniques. In this work, we carried out an organised temporal analysis of the application of analytical methods across BLS disciplines. We processed a sizable corpus of open-access, peer-reviewed literature using natural language processing methods. The objective of our study was to sketch out the methodological terrain of the BLS disciplines and identify recent changes (2009 to 2020). The phrase “analytic methods”

is used here to refer broadly to any quantitative or qualitative method for data analysis, including any algorithms, statistics, or models used to describe, summarise, or interpret a sample of data. The components of research technique involved in data collecting, experiment design, or study design are intended to be excluded by this definition. A peer-reviewed quantitative or qualitative analysis of measured data points, including experimental, observational, or meta-analytic research, is also included in the general definition of “study.” We tracked historical and disciplinary changes in analytical techniques. We identified analytical techniques that have gained or lost popularity over the last ten years across BLS disciplines from a temporal viewpoint. We identified analytic approaches that are distinctively prevalent within each BLS field and the similarity or dissimilarity of BLS disciplines in terms of their employment of analytic methods from a cross-disciplinary perspective. According to our assessment, the analytical techniques that are frequently taught in beginning research methods and statistics courses—such as the t test and the ANOVA—remain the techniques that have been discussed the most in BLS research publications during the past ten years. Over the past ten years, these techniques have, however, mostly lost favour or stayed constant. On the other side, from 2009 to 2020, mentions have consistently increased, sometimes exponentially, according to multivariate [4-6] statistics and machine

***Corresponding author:** Bzdok Nomi, Department of Biomedical Engineering, McConnell Brain Imaging Centre (BIC), Montreal Neurological Institute (MNI), Faculty of Medicine, McGill University, Montreal, Canada, Mila—Quebec Artificial Intelligence Institute, Montreal, Canada, E-mail: iomi@gmail.com

Received: 03-Jan-2023, Manuscript No: jabt-23-86930, **Editor assigned:** 05-Jan-2023, Pre QC No: jabt-23-86930(PQ), **Reviewed:** 19-Jan-2023, QC No: jabt-23-86930, **Revised:** 23-Jan-2023, Manuscript No: jabt-23-86930(R), **Published:** 30-Jan-2023, DOI: 10.4172/2155-9872.1000493

Citation: Nomi B (2023) An Examination of Changes in Analysis Techniques across Fields. J Anal Bioanal Tech 14: 493.

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

learning methodologies. Furthermore, we discovered that analytical techniques are not equally distributed throughout BLS fields but rather favour some over others.

Materials and Method

Recognition of named entities using analytical approach entities

It would be impossible to manually tag each analytic technique term in each section text in the final corpus of methodology and outcomes section texts. So, in order to find analytic technique phrases, we choose an automated recognition approach. We coupled a phrase matching/rule-based technique with machine learning. First, we created a lengthy phrase list (1,129 terms total) related to frequently employed analytical techniques throughout BLS fields. These were employed as part of a rule-based matching strategy to find references to analytical methods in the section texts. We chose 20,000 randomly chosen method and results section texts with tagged terms from our rule-based approach and fed them as training samples to a statistical NER algorithm in order to generalise detection of analytic method phrases outside of our manually produced list. The NER algorithm's goal is to use the rule-based training samples as context "clues" for more widely identifying analytical technique phrases (i.e., those outside the original phrase list). We used the convolutional [7-10] neural network (CNN) technique from the free and open-source spaCy python package to run NER on our entire corpus (100 iterations, 0.2 dropout, and mini-batch training). To create the final list of detected analytic technique words or entities, the trained NER model was applied to the whole corpus of methodology and results section texts. Over half of the texts in the corpus of 2,292,668 methodology section texts ($N_{\text{text}} = 1,438,077$) had at least one analytic technique entity. The discovered analytic method phrases from the trained NER model are referred to as analytic method entities or simply analytic methods in the main text. The total number of distinct method entities that were found using the NER algorithm was $N_{\text{entity}} = 15,016$ (before pre-preprocessing).

Processing of analytical method entities

An exclusive list of analytical technique entities was produced by the NER algorithm and was represented as a distinctive string of characters. The same analytical technique may be referred to by different names. For instance, a t test may be referred to as a "t test," "t-tests," "ttest," etc. Each entity string for the analytic technique underwent a series of preprocessing processes for harmonisation to try to correct for these minor spelling variations: First, lowercase letters; second, non-alphanumeric or non-Greek characters (such as hyphens, quotations, and commas); and third, lemmatization of tokenized words. We first manually grouped analytical methods into larger superordinate categories of conceptually related techniques (analytic method categories—for example, t test/ANOVA, GLIMs, and survival analysis) ($N = 34$) in order to examine temporal variations in analytic method mentions throughout BLS sciences. Then, using 12 time points total ($N = 12$), we calculated annual counts for each category of analytical procedure. For the following explanations, we decided on an annual time resolution: (1) Some journals in our corpus are published only sometimes (e.g., annually or biannually), and (2) It was hypothesised that significant trends in the use of analytical methods within a discipline happen infrequently (i.e., over years, as opposed to months).

Discussion

The data analysis environment in the BLS sciences is dynamic. In

the 21st century, the democratisation and commoditization of tools for quantitative analysis have risen dramatically, and in the last ten years, the process has only sped up. The availability of [8] computer power, open-source software, and an abundance of big data in more spheres of human activity has caused this tectonic upheaval. The scientist in training has a difficult hill to climb while learning how to conduct data analysis. Graduate and undergraduate education must take into account the most recent trends and methods in data analysis in order to make this climb simpler. To describe the data analytics environment in the BLS sciences, we provide an automated 12-year study of almost 1.3 million open research papers. This study sought to give a quick overview of the ongoing methodological changes taking place in various scientific groups. We discover that the analytical techniques that are frequently covered in introductory research methods and statistics courses, such as the t test and ANOVA, continue to be the techniques that are most frequently referenced in the "Methods" and "Results" sections of research publications. However, despite being widely used, these techniques have mostly lost ground or stayed constant throughout the course of the study (2009 to 2020). On the other hand, over the course of the study, mentions of multivariate statistics and machine learning methods have consistently increased, occasionally exponentially. Furthermore, we discover that some analytical techniques are not equally prevalent throughout BLS fields but rather enjoy special importance in a few. In order to satisfy the pressing demand for educating a new generation of quantitatively literate scientists, we believe our results offer insightful information on how university curricula should be created. The phrase "multivariate statistics and machine learning approaches" covers a wide range of analytical techniques. These techniques, which include PCA, regression subset selection, PLS, support vector machines, random forest algorithms, and ANNs, are frequently covered in advanced statistics and computer science courses. While some of these techniques date back quite a while (for instance, PCA was created in the early 20th century), others are continuously being explored today (e.g., ANNs have only seen broad use in the past decade). The potential causes of the patterns that were noticed were not attempted to be investigated in this study. We hypothesise that they may be caused by a number of factors, including: (1) the gathering of larger and more complex datasets; (2) the recent rise in popularity of data science as a tool in academia and industry; or (3) a growing awareness among researchers that manuscripts containing advanced analytics are more likely to be well received by reviewers and editors.

Results

Summary of preprocessing and analysis

This study's main objective is to describe and comprehend how the utilisation of analytical methodologies has changed over time within several BLS disciplines. To do this, we looked at over 1.3 million articles that were published throughout a ten-year period of research. From the "Procedures and materials" and "Results" sections of a sizable corpus of peer-reviewed journals, we extracted mentions of/adoptions of analytical methods (PubMed Central Open Access Subset, PMC OAS). In order to accomplish this, we used a named entity recognition (NER) algorithm that was specially trained. These text-extracted references to certain data analysis methods are referred to as "analytic technique entities"—unique sequences of alphanumeric characters.

Author Contributions

The diagnosis and treatment of this cat were handled exclusively by Jennifer Weng and Harry Cridge. This report was written by Jennifer

Weng, and Harry Cridge gave it a critical appraisal. The final draught of the manuscript has received the approval of both Jennifer Weng and Harry Cridge.

Conflict of Interest

According to the authors, there are no conflicts of interest that might be thought to compromise the objectivity of the research presented.

Ethics Statement

The case described in this report was handled as part of the regular clinical caseload at the university teaching hospital; an IACUC or other ethical approval was not necessary. All facets of this patient's care had the owner's consent.

References

1. Burlikowska K, Stryjak I, Bogusiewicz J (2020) Comparison of metabolomic profiles of organs in mice of different strains based on SPME-LC-HRMS. *Metabolites* 10:1-10.
2. Beutler E, Waalen J (2006) The definition of anemia: what is the lower limit of normal of the blood hemoglobin concentration? *Blood* 107: 1747-1750.
3. US Department of Health and Human Services. (n.d.). Blood tests - blood tests. National Heart Lung and Blood Institute.
4. Sundermann FW (1956) Status of clinical hemoglobinometry in the United States. *Am J Clin Pathol* 43: 9-15.
5. Wolf HU, Lang W, Zander R (1984) Alkaline haematin D-575, a new tool for the determination of haemoglobin as an alternative to the cyanhaemoglobin method. II. Standardisation of the method using pure chlorohaemin. *Clin Chim Acta* 136: 95-104.
6. Shah VB, Shah BS, Puranik GV (2011) Evaluation of non cyanide methods for hemoglobin estimation. *Indian J Pathol Micr* 54: 764-768.
7. Karakochuk CD, Hess SY, Moorthy D, Namaste S, Parker ME, et al. (2019) Measurement and interpretation of hemoglobin concentration in clinical and field settings: a narrative review. *Ann NY Acad Sci* 1450: 126-146.
8. Kang SH, Kim HK, Ham CK, Lee DS, Cho HI (2008) Comparison of four hematology analyzers, CELL-DYN Sapphire, ADVIA 120, Coulter LH 750, and Sysmex XE-2100, in terms of clinical usefulness. *Int J Lab Hem* 30: 480-486.
9. Whitehead Jr RD, Zhang M, Sternberg MR, Schleicher RL, Drammeh B, et al. (2017) Effects of preanalytical factors on hemoglobin measurement: A comparison of two HemoCue point-of-care analyzers. *Clin Biochem* 50: 513-520.
10. Ingram CF, Lewis SM (2000) Clinical use of WHO haemoglobin colour scale: validation and critique. *J Clin Pathol* 53: 933-937.