Conference minutes derived from AI companion via Zoom recording, 11/4/2024

Conference web site: https://neaman.org.il/en/On-the-foundations-of-applied-statistics.

The conference opened with remarks by Shimon Maron, Samual Neaman Institute director and Orly Manor, president of the Israel Society for Statistics and Data Science. This was followed by a tribute to Edna Shechtman for whom the conference was dedicated, by Yoav Benjamni.

Recap

The role of statistics and philosophy of science in research, with a focus on statistical approaches, key comparison studies, and the limitations of dichotomous claims based on p-values, were discussed. The application of statistical methods in biomedicine, the challenges and opportunities presented by big data, and the importance of interdisciplinary dialogue between philosophy and statistics were also emphasized. Lastly, the development of computational tools using robust methods, the need for a more sophisticated approach to compare the efficacy of a booster dose with the original two-dose regimen of the Covid-19 vaccine, and the importance of effective communication between statisticians and other disciplines during the pandemic were discussed.

Next steps

A possible follow up to this event is to organize a workshop with experts in the philosophy of science, statistics, and experimental design in order to discuss and establish a coherent and consistent approach to the application of statistical methods. Some examples include: Christian Henning explored methods to assess the alignment between the interpretative and effective hypotheses in statistical tests. David Zucker collaborates with medical professionals to better understand their perspectives and incorporate them into statistical practices. Yair Goldberg promotes the use of quasi-experimental designs in observational studies and advocate for their inclusion in the statistical analysis toolbox. Ron Kenett looked into the lack of reproducibility in biomedical research and proposed strategies to improve it based on alternative representations of findings He will be glad to share his experience and insights, emphasizing the importance of understanding the data generation process. Ron Kenett also encouraged the participants to submit papers to the special issue he is editing with David Stenberg on the foundations of applied statistics in *Applied Stochastic Model in Business and Industry*, highlighting the importance of advancing the field through research and publications. See: https://www.statisticsviews.com/call-for-papers-asmbi-special-issue-on-foundations-of-applied-statistics/

Statistics, Philosophy of Science, and Meta-Analysis

Daniel Lakens discussed the role of statistics and philosophy of science in their research, focusing on the choice of statistical approach, the concept of 'key comparison study', and the limitations of dichotomous claims based on p-values. They also explored the methodology of scientific discovery, the impact of forcing variability in laboratory experiments, and the need for a more coherent set of aims, methods, and claims in certain fields. Stephen Senn presented on the history and development of statistical theory, discussing the works of notable scientists such as Thomas Bayes, Laplace, Gauss, and the author of a 19th-century probability book. These authors discussed the concept of combining observations, the calculation of the standard error of the difference in means, and the use of Bayesian and frequentist approaches in meta-analysis.

Statistical Methods in Biomedicine and Big Data

The discussion focused on the application of statistical methods in biomedicine, with a particular emphasis on the challenges and opportunities presented by big data. Clelia Di Serio, a statistician, highlighted the importance of representativeness, reproducibility, and sampling design, and discussed the significant impact of technology on data. She also pointed out the changing trends in cancer incidence and mortality, and the need for evidence-based methods in research. Hennig suggested that different statistical approaches could offer complementary perspectives on a problem, while Lakens proposed explicitly stating which approach was used to make which claim. Lastly, Stephen Senn and Daniel Lakens discussed the value and challenges of interdisciplinary dialogue between philosophy and statistics, emphasizing the need for mutual understanding and resolving long-standing disagreements.

Post-Traumatic Stress Disorder Analysis and Statistical Models
David Steinberg introduced Christian Hennig, a professor at the University of Bologna, and Bernard Francq, a lead statistician at Glaxo Smith Klein, as the first speakers of the session. Christian presented an analysis of a dataset published in 2017, which included measurements from 13 volunteers with post-traumatic stress disorder. He discussed the correlation between perceived stress and amygdala activity, and the use and interpretation of statistical models. He emphasized the importance of understanding the properties of test statistics, and the need to acknowledge model uncertainty. David Zucker, a professor at the Hebrew University in Jerusalem, was announced as the discussion following the two talks.

Statistical Topics and Robustness Concepts
The talk by Bernard, involved complex discussions on various statistical topics including the new interpretation of the T test, tolerance intervals, success probabilities, and P values. David then presented on the utility of P values, statistical significance, and the importance of model simplicity and goodness of fit, critiquing a study for overstating correlation based on limited data. Stephen raised concerns about the interpretation of tolerance intervals in clinical trials, while Daniel Lakens questioned the preference for certain measures over others. Finally, Christian Hennig discussed the concept of robustness, emphasizing its interpretative and nominal hypothesis nature, and highlighting its importance in processing decisions and the choice of models.

Improving Replicability and Generalizability in Research
Ron Kenett chaired the next session where Yoav Benjamini, a co-developer of the widely used False Discovery Rate, and David Steinberg, an expert in experimental design, were introduced as speakers. Yoav discussed his research on multiple comparison procedures, the importance of standardization in animal behaviour experiments, and the development of computational tools using robust methods to address these issues. He emphasized the need to consider interaction variability when evaluating results and proposed a replicability adjustment tool to improve the replicability rate in multiple lab studies. Yoav also highlighted the significance of identifying relevant sources of variability in data analysis, the limitations of traditional statistical tools, and the importance of assessing generalizability in scientific research. David suggested a new approach to experiment design that prioritizes generalizability over precision and proposed the concept of 'blocking' to achieve this.

AI Insulin Dosage Study and Covid-19 Quasi-Experiments
Yair Goldberg introduced a study from Schneider Medical Center on an AI system for adjusting insulin doses in young diabetes patients, which showed promising results across various factors. The discussion also included the introduction of several experts, including professors Leon and Gilad, who have specific interests in data and decision science, human factors engineering, and enhancing human-machine interactions. Additionally, Yair discussed the quasi-experiments conducted regarding Covid-19, focusing on the challenges faced and lessons learned, and the implications of a potential third dose of the Covid vaccine. He emphasized the need for a more sophisticated approach, such as matching and regression discontinuity, to compare the efficacy of a booster dose with the original two-dose regimen. The results showed that the booster provided 9.21 times better protection compared to those who didn't receive it.

Statistical Specialists, Collaboration, and Innovation
Yisrael Parmet shared his insights on the role of a statistician in various fields and the importance of collaborations, consultation, and simplifying data analysis methods. He also discussed the need for statistical specialists to be involved in the research process from design to analysis and the importance of promoting innovation and breaking conventions. Issues with reviewing research papers, the complexity of quasi-experimental designs, and the importance of distinguishing between populations were also discussed. The conference ended with Ron Kenett presenting a list of foundational principles in applied statistics, emphasizing the importance of effective communication between statisticians and other disciplines like during the Covid-19 pandemic or other chaotic situations.

Appendix 1: Introduction to conference speakers and discussants

Session 1

1. Daniel Lakens is an Associate Professor of Metascience and chair of the Ethical Review Board at the Human-Technology Interaction group at Eindhoven University of Technology in The Netherlands. Lakens' work focuses on improving research methods and statistical inferences in the social sciences. He has worked on rewarding replication studies, introducing Registered Reports and study preregistration, improving hypothesis testing and sample size justification by developing free educational resources on these topics.
2. Stephen Senn is a Swiss and British statistician who has worked in England, Scotland, Switzerland, Luxembourg & France in academia, the pharmaceutical industry & public health. His research interests are in statistics applied to drug development.

Discussant: Clelia diSerio is Full Professor in medical statistics and epidemiology at San Raffaele University, Milan, Professor of Statistics at the State Swiss University (Lugano) and Director of University Centre for Statistics in the Biomedical Sciences (CUSSB)

Session 2

3. Christian Hennig is a full professor at the university of Bologna. His work focuses on cluster analysis, multivariate data analysis, classification and data visualisation, robust statistics, foundations and philosophy of statistics, statistical modelling and applications
4. Bernard G Francq is lead statistician with GSK, driving statistical research & innovation for CMC projects. His research interests include DOE, non-clinical statistics, tolerance intervals and data visualization. He regularly offers talks and webinars in international conferences and biopharma industry.

Discussant: David Zucker is full professor at the Department of Statistics and Data Science, Hebrew University. Previously David worked for the US FDA and the US NIH. He served as Co-Editor of Biometrics. His fields of interest include biostatistics, survival analysis, measurement error models, and clinical trials.

Session 3

5. Yoav Benjamini is Professor Emeritus of Applied Statistics at the Department of statistics and Operations Research at Tel Aviv University, and a member of the Sagol School of Neuroscience and the Edmond Safra Bioinformatics Center. Yoav is a co-developer of the widely used False Discovery Rate concept and methodology. His other research topics are replicability and reproducibility and applications across diverse areas of science. He is a member of the Israel and US Academies of Sciences, and received the Israel Prize in Statistics and Economics and the Pearson Prize of the International Statistical Institute.
6. David Steinberg is an Emeritus Professor of Statistics at Tel Aviv University. His primary area of research is the design of experiments, and he has been involved in many applications.

Discussant: Philip Tzvi Reiss earned his PhD in biostatistics at Columbia University and was on the faculty of the New York University School of Medicine. He made aliyah (relocated to Israel) in 2015 and is now Professor of Statistics at the University of Haifa.

Session 3

7. Yair Goldberg is an associate professor of statistics in the Faculty of Data and Decision Sciences at the Technion and the Vice Dean for Graduate Studies. His lab uses statistical and machine-learning tools to address theoretical and applied statistical questions. During 2021-2023, he was part of a research team advising Israel's Ministry of Health on COVID-19 policymaking.
8. Yisrael Parmet is Professor with the Department of Industrial Engineering and Management at Ben Gurion University. He specializes in areas of Statistical learning and modeling, Machine learning, Design of experiments and Human Factors.

Discussant: Tal Oron-Gilad is a full professor in the Department of Industrial Engineering and Management at Ben-Gurion University. Her expertise is in human factors engineering and her specific interests are geared towards promoting user-centred design of Human-Machine Interactions, Improving Human's ability to interact with autonomous systems.

Concluding remarks

Ron S. Kenett is Senior Research Fellow at the Samuel Neaman Institute, Technion, Chairman of the KPA Group and of the Israel Data Science Society. He is an applied statistician who authored over 250 publications and 18 books. Recipient of the Greenfield Medal by the Royal Statistical Society, the Box Medal by the European Network for Business and Industrial Statistics and the life achievement award by the Israel Society for Quality.

Appendix 2: Abstracts

1. There is nothing as practical as a good philosophy
Daniel Lakens, Eindhoven University of Technology, Holland
An empirical researcher naïve enough to ask a statistician how they should analyze their data are almost certain to receive the answer 'it depends'. Some specific century-old disagreements among statisticians about what to do might be only a mild embarrassment for the statistics community, but a blessing for researchers who do not have to change their ways because 'even statisticians don't agree on what is best'. Although pluralism and pragmatism have their virtues, so do principles. Yes, how data should be analyzed depends, but in some cases, it primarily depends on your philosophy of science. Diverging advice can often be explained by different philosophies of science, but statisticians rarely ground their views in philosophy. This is regrettable, as a strong foundation in philosophy of science provides practical guidance. For example. decisions about whether violations of the likelihood principle are problematic or not can only be answered if one has committed to a specific philosophy of science. Another example is the decision to correct for multiple comparisons or not is not a statistical question, but a philosophical one. I will argue that a stronger connection between statistics and philosophy of science is a requirement for improvements in statistical practice. Declaring one's philosophical views should be a prerequisite when making recommendations on which statistical inferences empirical researchers should 'want to know'. Researchers can rely on multiple philosophies when analyzing their data, as long they reason from basic principles, and prevent incoherent courses of action when designing experiments and analyzing data.

2. Déjà eu. How we keep on losing and re-inventing statistical theory
Stephen Senn, University of Sheffield and the Medical University of Vienna

"*No scientific discovery is named after its original discoverer.*" Stigler's Law of Eponymy[1]
Student's t-distribution of 1908[2] was, in fact, derived by Jakob Luroth in 1876[3, 4] but neither paper is cited in the first (1917)[5] or second (1931)[6] edition of the book by David Brunt on what we would now call meta-analysis, which itself has been completely ignored in the modern development of that subject. Brunt himself was merely developing what Airy had done more than half a century earlier[7] but what Airy did is also largely forgotten today. Much of what has recently been developed in network meta-analysis was established earlier, better and deeper in the theory of incomplete block designs. Many more examples could be given. Statisticians might blame others for not knowing what statisticians have already invented but we ourselves are also guilty of ignoring what our colleagues have done. To give a simple example, claims frequently made in the biostatistics literature about representativeness of clinical trials could not possibly survive if viewed from the perspective of sampling theory, a field about which many biostatisticians appear to be completely ignorant. I speculate that part of the problem is failure to ally theory to practice. The more statistical theory is used, the more likely it is to survive. The collaboration of applied and theoretical statisticians is much to be desired. Good understanding of applications will lead to more robust theory and more relevant theory will improve application.
References:
1. Stigler, S.M., *Stigler's law of eponomy.* Transactions of the New York Academy of Sciences, 1980. 2nd series, 39: p. 147-157.
2. Student, *The probable error of a mean.* Biometrika, 1908. 6: p. 1-25.
3. Lüroth, J., *Vergleichung von zwei Werten des warscheinlichen Fehlers.* Astronomische Nachrichten, 1876(87): p. 209-220.
4. Pfanzagl, J. and O. Sheynin, *Studies in the history of probability and statistics .44. A forerunner of the t-distribution.* Biometrika, 1996. 83(4): p. 891-898.
5. Brunt, D., *The combination of observations*. 1917: University Press.
6. Brunt, D., *The Combination of Observations*. 1931, Cambridge: Cambridge. 239.
7. Airy, G.B., *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations*. 1862, London: MacMillan and Co.

3. Understanding statistical inference based on models that aren't true
Christian Hennig, University of Bologna, Italy
Statistical inference is based on probability models, and most of the theory behind it assumes these models to be true. But models are idealisations, and it makes little sense to postulate that they are literally true in reality. Models are however required to analyse the behaviour of statistical methods in any generality. In order to explore the implications of running statistical inference based on models that aren't true, it is helpful to look at more general supermodels that allow for violation of the supposedly assumed models. I will present a framework for how to think about statistical inference based on models that aren't true, conditions under which such inference can be useful or misleading, and what impact this has on the interpretation of the results in practical settings.

4. Tolerance intervals and probability indexes: A new interpretation of the t-test and its p-value
Bernard Francq, CMC Statistical Sciences, GSK, Belgium and Ron S. Kenett, the KPA group and the Samuel Neaman Institute, Technion, Israel

In medical research, statistical significance is often based on confidence intervals (CIs) and p-values, the reporting of which is included in publications in most top-level medical journals. However, recent years have seen ongoing debates on the usefulness of these parameters, leading to a significance crisis. Misinterpretations of CIs and p-values leads to misleading conclusions and nonreproducable claims. The lower the p-value does not necessarily mean the better the treatment. Index probabilities, the s-value, the B-value (namely the probability that a patient under treatment A ends up with a better clinical outcome compared to another patient under treatment B) or the Generalized Pairwise Comparison have been proposed in the literature as alternative solutions. Here, we promote the use of tolerance intervals which allow a generalized definition by calculating the individual success probabilities (ISP). The ISP allows a clear interpretation following both frequentist and Bayesian paradigms. Using synthetic examples with the 1-sample, paired t-test, 2-samples t-test and two one-sided tests (TOST), we show that the ISP is a one-to-one function of the p-value with enhanced interpretability properties. The ISP is the confidence bound of the probability index with a default cut-off value of 50% whatever the type I error that avoids the common pitfalls of the CIs and p-values. The ISP offers enhanced insights in reviewing statistical analysis in medical research from such a perspective. We argue that the ISP should be preferred by researchers and considered by journal editors.

5. Hunting out the relevant variability
Yoav Benjamini, Department of Statistics and Operations Research, Tel Aviv University, Israel
A central challenge in applied statistics is to quantify the level of uncertainty in our conclusions,
be it by standard error, p-value, confidence interval and prediction error or by their Bayesian counterparts.
Underlying them all is the variability which should be relevant to the user of our analysis. Too often the variability is restricted to that of sampling the population of interest, and sometimes only to the population at hand. I'll discuss various strategies (with examples):
The choice of the way to split a dataset for training and validation or in cross validation (in the analysis of length of stay in neonate intensive care unit). The choice of mixed versus fixed analysis - (in the multi-lab study of animal behaviour and its impact on the replicability of results and in meta-analysis). Practical compromises may be needed, yet this problem is too often ignored.

6. Design of Experiments for Generalizability
David M Steinberg, Department of Statistics and Operations Research, Tel Aviv University, Israel
Controlled experiments are the gold standard to learn what happens when we intervene in a system. They are used to compare treatment protocols in medicine, competing social policies, reactions to stimuli in psychology, alternative formulations in chemistry, different production schemes in industry and in many other contexts. The statistical design of experiments has developed valuable methods for limiting bias in these comparisons and improving precision. In particular, blocking plays a major role here. The result is both reliable conclusions and efficient use of resources – improving experimental precision for a fixed budget. Often, though, more is needed. Scientists and engineers want to generalize the findings from their experiments. Much less has been written about how to design experiments when one of the goals is generalization. We discuss some ideas to fill that gap, drawing on relevant literature from the design of clinical trials, the design of engineering experiments and quality by design in the pharmaceutical industry.

7. When the control group gradually becomes the treatment group
Yair Goldberg, Technion, Israel
Estimating vaccine effectiveness (VE) and the waning of the immunity obtained by vaccines is a challenging task. It is even more challenging when performing a retrospective study based on a large cohort of individuals who get the vaccination over time. In such studies, the typical differentiation between the control arm and the treatment arm no longer exists as individuals are dynamically moving from the control arm to the treatment arm. In this talk, I will discuss the theoretical and practical challenges of statistical estimation in such a dynamic environment in the context of the estimation of the COVID-19 vaccination effectiveness and waning.

8. Communication between the researcher and the statistician is essential for research
Yisrael Parmet, Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Israel
Data analysis is an essential step in any empirical study. More than once, the conversation between the researcher and the "statistician" fails because one of the parties fails to understand and communicate with the other. This dialogue between the researchers and a statistician is very important for the success of the research and the optimal production of the insights that can be derived from it. In this presentation, we discuss the critical role of the statistician in this dialogue and how a statistician should behave and manage it to maximize the data analysis to address research questions.