SAMUEL NEAMAN INSTITUTE
FOR NATIONAL POLICY RESEARCH

The Israel Statistics and Data Science Association
الرابطة الإسرائيلية للإحصاء وعلم البيانات
האיגוד הישראלי לסטטיסטיקה ומדע הנתונים

A one-day conference

On the foundations of applied statistics

11/4/2024

10:00AM – 17:00

12 invited speakers and discussants

4 sessions

Each session: 2 speakers, 1 discussant

Each speaker and discussant gets 15 minutes to present

1 discussant is commenting on 2 speakers

Following the discussant comments, an open discussion with participants

The conference is dedicated to the memory of **Prof Edna Schechtman**

Following the conference, a special issue of the journal *Applied Stochastic Models in Business and Industry* will be published

PROGRAM

Opening remarks: 10:00AM – 10:10AM

Professor Marom, Director SNI, and Professor Manor, President ISDSA

Remembering Prof Edna Schechtman: 10:10AM – 10:20AM

Yoav Benjamini

Session 1: 10:20AM – 11:30AM

1. Daniel Lakens – Eindhoven University of Technology, Holland

2. Stephen Senn – University of Sheffield and the Medical University of Vienna

Discussant: Clelia DiSerio – San Raffaele, Milano, Italy

Coffee Break: 11:30AM – 11:45AM

Session 2: 11:45AM – 13:00

3. Christian Hennig – University of Bologna, Italy

4. Bernard Francq – GSK, Belgium

Discussant: David Zucker – Hebrew University

Lunch Break: 13:00AM – 14:00AM

Session 3: 14:00 – 15:15

5. Yoav Benjamini – Tel Aviv University

6. David Steinberg – Tel Aviv University

Discussant: Philip Tzvi Reiss – University of Haifa

Coffee Break: 15:15AM – 15:30AM

Session 4: 15:30 – 16:45

7. Yair Goldberg - Technion

8. Yisrael Parmet – Ben Gurion University

Discussant: Tal Oron – Ben Gurion University

Concluding remarks: 16:45 – 17:00

Ron Kenett

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.

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 an 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" does not succeed 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.