Introduction to IPM42 at the 56th ISI conference in Lisbon

The title of this session (IPM 42) is *Observational studies, Confounding and Multivariate Thinking*.

Over 100 years ago, Karl Pearson said "correlation does not imply causation." He also said that "science should not deal with unobservable entities" by which he meant the natures and causes of things.

Karl Pearson introduced a "non-causal" epistemology into statistics. This non-causal approach to knowledge has been adopted by researchers in the social sciences.

Karl Pearson's non-causal epistemology was successfully opposed by Ronald Fisher in the case of experiments involving random assignment.

Karl Pearson's non-causal epistemology is under active review by those statisticians who analyze the results of observational studies.

Observational studies are common. They are relatively cheap compared to a well-controlled experiment with random assignment.

Observational studies are much more common than randomized experiments in the daily news and in technical journals such as the Journal of the American Medical Association (JAMA).

In the US, an estimated 60% of college graduates take statistics. Of these, an estimated 80% are in majors where observational studies are more common than randomized experiments. Majors such as business, economics, sociology, social work, political science, history, geology and journalism focus primarily on observational studies.

Even in majors where randomized assignment is possible (majors such as education and psychology), the lack of control may introduce problems of confounding and bias found in observational studies.

Observational studies are the neglected step child of statistical education.

The biggest problem in observational studies is confounding -- the influence of a related factor that was unobserved and may be unknown or unknowable (at a particular state of technology).

Understanding the influence of a confounder on an association requires a minimum of three variables. Understanding confounding requires multivariate thinking.

Most introductory statistics courses deal with one and two variables. Few -- if any -- have time to deal with three variables necessary to illustrate confounding. . Historically it has required a second course in statistics to introduce multivariate regression or ANOVA.

I want to thank the IASE and Alan Rossman for supporting a session on this topic.

The primary goal of this session is to highlight the importance of observational studies in statistics education. A secondary goal is to focus on the problem of confounding in teaching and in theory.

Our first speaker is John Harraway from New Zealand. The title of his paper is "*Study Design and Confounder Control in Observational Studies: two Cases.*"

Our second speaker is James Nicholson. He is representing his co-authors Jim Ridgway and Sean McKusker from the University of Durham in England. The title of their paper is "*Using multivariate data as a focus for multiple curriculum perspectives at secondary level*.

Our third speaker is Donald Rubin from Harvard University in the US. The title of his paper is "*Dealing with Multivariate Outcomes in Studies for Causal Effects*"

Our fourth speaking group consists of David Cox (Nuffield College Oxford) and Nanny Wermuth (Gothenburg University in Sweden). The title of their paper is "*Some interpretational issues connected with observational studies"*

Questions:

1. Donald. You have taught an introductory statistics course at Harvard to undergraduates who have had no prior training in statistics. I believe that this course focused more on observational studies and on confounding. What can you tell us about that course and your experience in teaching it?
2. David and Nanny. What advice would you have to statistical educators who are trying to educate students on observational studies and on confounding?