How is statistics different from mathematics? assignment

Not my own work but it is fun to read. How is Statistics Different from Mathematics? And Why Is the Distinction Important? Isadora P. David Acronyms: ACHED-Commission on Higher Education; INHERE-National Higher Education Research Agenda; HE-Higher Education Institution; AS- Philippine Statistical Association; SORT-Statistical Research and Training Center; UPPED-AS- University of the Philippines School of Statistics; PULP-IS- University of the Philippines at Los Banes Institute of Statistics; GUISE- Guidelines for the Assessment and Instruction in Statistics Education. . Introduction 1. On 9 October 2009 this writer was invited by the Commission on Higher Education (ACHED) to Join a group " to discuss the state of research, research gaps and priority research areas in the mathematical and statistical sciences". The group's deliberations would be inputs to the agency's National Higher Education Research Agenda Phase 2 (NEARER) for 2009-2018.

Aside from the group's leader who is a mathematical statistician and yours truly who many acquaintances would consider an applied statistician, the other attendees are mathematicians from universities across the country. The mathematicians pointed out that since they needed only encircle and paper in their research, the low budgets of their proposals submitted for funding under the previous INHERE led to very high rejection rates.

With the possible exception of a few, we take this observation as true for most Filipino mathematicians. 2. The situation in statistics is Just the opposite. Statisticians who rely exclusively on pencil and paper tend to be found in academe and are fast becoming relics. The large majority practice

their profession with the aid of vast computing power, collaborate actively with researchers in substantive fields, or employ large numbers f persons.

Statisticians in academe do model simulations, resembling techniques and data mining on computers as well as consulting in government and private sample surveys that require supervising large groups of people who collect and process the data, and write voluminous reports on the results; biostatistics design and analyze drug screening protocols; market research and opinion polling companies employ statisticians who design, analyze and interpret the results of their surveys; and so on. None of these activities come close to pen and paper research work by mathematicians. The many important differences between what the majority of mathematicians do on the one hand, and what most statisticians do on the other hand, should make it plain that their respective disciplines have become sufficiently distinct and should be treated as such. However, very few of the 1700 higher educational institutions (HE's) under Shed's orbit recognize the distinction and act accordingly, e. G. Have separate departments and curricula for the two disciplines; some are very slowly getting there and have statistician positions in their mathematics departments; but many have either positions nor statistics graduates in their faculties.

In fact, from its creation by R. A. 7722 in 1994 until very recently, ACHED had operated on the premise that statistics was not a separate discipline but treated it as part of mathematics. The question, is statistics different from mathematics? Came up again during the aforementioned October 9 meeting. Indeed, Shed's use of the phrase " mathematical and statistical sciences" in place of mathematical sciences is very new. 4.

The rest of this note reviews how mathematics and statistics are different at a more fundamental level, speculates on reasons why this difference has not been ideal acknowledged in Philippine HE's, and dwells on the importance for those concerned, namely mathematicians, mathematics departments, HESS administrators and ACHED management, to understand and acknowledge the difference – and act accordingly. 2. Statistics is not Mathematics 5. Both disciplines deal with numbers. However, as pointed out by many (e. G. Cob and Moore, 1997) and emphasized in newer basic texts (e. G. De Beaux et. Al. 006), statistics actually deals with data, which are numbers with context – the what, why, when, where and how behind the numbers. What do the numbers represent, family incomes in pesos, fasting blood sugar levels of a group of persons 20 years or older, rice stem borer morbidity rates from different insecticides, or voters' preferences among presidential candidates in the next election? The why has to do with reasons for collecting the data in the first place, a careful reflection of which lead to a systematic statement of the study, formulation of hypotheses, and exploration of possible methods of testing these hypotheses.

When were the data collected? WSDL the inferences be confined to the same period or is the intention to forecast the true, such as when the data represent a time series? 6. In most cases the numbers are a subset (sample) oaf much bigger but when applied to the sample enable the researcher to make inferences about the unobserved population and, additionally, measure the likely margin of error in the inference. To be able to do so, it is important to be clear where the sample came from (the population), and necessary to know how it was collected.

The design of experiments, laid on a formal foundation by the father of modern statistics R. A. Fisher (1935) in a book by the same title, now permeates all so-called controlled experiments done in fields and laboratories. Because the experiment can be designed to control many extraneous factors other than what are to be studied, these experiments usually involve very small samples. Also, the population does not physically exist, but is defined by the experimental design and through a statistical model of the researcher's choosing.

In many other situations the population is real but too large to enumerate completely (e. G. All people residing in the Philippines at the moment) or very difficult to capture comprehensively (e. G. Fish and wildlife, water r air quality over wide areas). In some cases the population has yet to exist, as in opinion polls to predict how the electorate will vote in the coming election. Designing sample surveys that use probability-based procedures to choose a subset from the population, develop instruments (e. G. Questionnaires, training manuals for field staff and methods of analyzing the sample to infer about the population, collecting the data, and writing documentations on the survey and its results are activities that survey statisticians train for and do. Since for ethical and other reasons you cannot exercise as much control as you wish with people and the real world, and because the phenomena being measured inherently fluctuate widely (e. G. Income, farm sizes, rice imports, daily dietary calorie consumption), bigger survey samples are required than controlled experiment samples.

The Labor Force Survey done quarterly by the National Statistics Office has 50, 000 sample dwelling units out of a national total of some 18 million.

Market researchers and opinion pollsters use smaller samples because they are interested mainly in proportions and percentages that are subject to less variation. Provided the sample has been drawn following statistically sound and efficient procedures, pollsters often claim that with a sample of 1200-1500 persons they have 19 in 1 odds that their estimate is within 3 percentage points of the unknown true proportion (the oft-heard B±3 percent margin of error). . These are Just some of the activities that statisticians study and do for a living – which mathematicians generally don't. This fixation with data as numbers with context that came from a bigger set which is the real object of interest is what differentiates statistics from mathematics. Bradley Effort, the statistician who did many things but is most known for developing the bootstrap method, put it succinctly when he said, " Statistics is the science of information gathering, especially when the information arrives in little pieces instead of big ones"(Larkin/McGraw Hill).

It is doubtful that anyone would suggest, seriously or in Jest, that mathematics is preoccupied with the science of information gathering. 8. Mathematics is associated with deterministic logic in which a function f(XSL, … , CNN) leads unerringly to one value every time XSL, CNN are replaced by n numbers at arability that one card picked at random from a well shuffled deck of 52 cards will be an ace is 4/52, no plus or minus. The antonyms of deterministic are random, probabilistic, and stochastic.

Statistics recognizes that XSL , CNN represent data from a sample which is only one of many possibilities, so that the value f takes is not fixed but it in fact varies across the space of possible samples (sampling variation). The decision on what experimental or sampling design to use among so many

choices has to do with reducing sampling variation to a minimum, given limited resources. Secondly, the observed data often are not the true values for myriad reasons, e. G. Sighing scale off by a few grams, farmer's recollection of last season's rice output was inaccurate, respondent's preferred candidate could change between survey time and election day, respondent fibs on answering a sensitive question, or interviewer resorts to creative reporting on less accessible respondents. Good experiment and survey practice strives to reduce these sources of variation (from true value) to the extent possible; however, some will remain regardless of which sample is chosen (hence, the term non-sampling variation).

Thirdly, in statistics f is not the absolutely correct function, but is regarded as a model of the relationship between the variables under study. For example, the popularity of linear regression models is not due to their correctness; rather we know from mathematics that most functions can be expressed as convergent Taylor series, hence a linear regression model may be viewed as the first order approximation from the series. The choice of f therefore is subject to error and is replaced by $f + e$, with e representing error (including model misspecification error).

Furthermore, e is not fixed but is assumed to behave randomly. The statement, " Essentially, all models are wrong, but some are useful" is traced to the eminent statistician George Box (1987). 9. Whether a field has accumulated enough body of knowledge upon which it could stand on its own is another criterion used to determine if the field can be considered a separate discipline. The website Static. Org lists 190 core Journals that publish articles in statistics. The International Statistical Institute (http://sis.

CBS. Nil/), established in 1885, is one of the oldest scientific associations operating in the modern world.

The SIS'S more than 2000 elected members are supplemented by additional 3000 members of its eight specialized sections, biz. Bernoulli Society for Mathematical Statistics and Probability, International Association for Statistical Computing, International Society for Business and Industrial Statistics, International Association for Statistics Education, International Association of Survey Statisticians, International Association for Official Statistics, International Environments Society, and Irving Fisher Committee on Central Bank Statistics.

Countries have their own statistics societies. The Philippine Statistical Association (www. Phillips. Org. PH) which as established in 1952 currently has 960 individual members, 53 institutional members and six active regional chapters. The AS publishes its own Journal The Philippine Statistician and newsletter, AS Editions. 3. Why the continuing confusion in the Philippines between the two disciplines reveals certain patterns. In England, Karl Pearson, who gave us the correlation coefficient and chi-square test in 1900, got his PhD in mathematics under Francis Gallon. Who may have strongly influenced the farmer's research interest in evolution and eugenics . K Pearson founded Biometrics, the Journal devoted to statistical applications in the life sciences, which survives today. E. S. Pearson of the Newman-Pearson lemma for testing statistical hypotheses fame, did graduate work in astronomy, but eventually succeeded his father as professor of statistics at University College London and editor of Biometrics. R. A. Fisher (1890-1962)

had his formal training in mathematics, but became famous as statistician, evolutionary biologist and geneticist.

He had been described as " a genius who almost single-handedly created the foundations for modern statistical sciences" and " the greatest of Darning's successors". He did most of his pioneering work, including his two books Statistical Methods for Research Workers and The Design of Experiments while working at an agricultural experiment station. 11. P. C. Inhalations, chanced upon Biometrics while studying physics in England, came back to India as physics professor and did research on anthropometry on the side; this led him to statistics' role in discriminate analysis and in developing what is now known as Inhalations' DO.

He founded the world famous Indian Statistical Institute in 1931 and was the driving force in the establishment of the Indian national statistical system. The development of large scale sample surveys is considered by many as his greatest contribution to statistics. C. R. Raw, who succeeded Inhalations as Director of the Indian Statistical Institute, had Mass in mathematics and statistics before he went to Cambridge England for his PhD in statistics under R. A. Fisher.

There are numerous theorems and methods bearing his name in statistics and mathematics (e. G. Raw-Blackwell theorem, Cramer-Raw bound), he had been named by The Times of India as one of the top 10 Indian scientists of all time, recipient of the US National Medal of Science, and a CRY Raw Advanced Institute of Mathematics, Statistics and Computer Science has been established in Hydrated to honor him. 12. In the US, George Consider,

whose graduate training was in physics, was professor of mathematics at Iowa State University.

His interest in statistics grew from his interaction with researchers in agriculture and biology, and in the late sass he began teaching a statistics course and organized seminars in the department of mathematics. In the early sass the first department of statistics and first statistical laboratory in the United States were established in that university. Many SIS autistics graduates fanned out to start statistics departments and programs in other universities. Jerry Newman, he of the Newman-Pearson lemma, earned a PhD in mathematics from his native Poland, with a thesis on application of probability to agricultural experimentation.

He had extensive experiences working in experiment stations and collaborating with mathematicians and statisticians in Europe prior to moving to US Berkeley in 1938 as a professor teaching probability and statistics in the mathematics department. It took him 17 years to finally convince his university to establish a separate statistics department in 1955. W. Edwards Deeming (1900-1993) mathematics and physics got introduced to applied statistics at the US Census Bureau and assignments to improve the quality of war materials during World War II.

His work on statistical quality control helped significantly in the Japanese post-war industrial revival and was regarded by many as the leading quality guru in the United States. He was a very successful statistics consultant in business and industry. 13. It may be noted that the statistics pioneers (a) had advanced training in mathematics or fields like physics that required

advanced mathematics and (b) most were professors in mathematics departments. However, these two attributes were not sufficient for statistics to blossom and in time grow out of the mathematics department.

A third necessary element was (c) involvement in research in other disciplines that needed developing methodologies for gathering information and for dealing with the variation in the information. This is vital because, as C. R. Raw once noted, statistics is unique in that it develops methodologies not for itself but for other fields. 14. In the Philippines mathematics departments had not and are not likely to spawn statistics departments in big numbers because, as mentioned, mathematicians still o mostly pen and paper research and have limited chances for collaborating with researchers in other fields.

This environment does not expose mathematicians, department heads and university management to situations where statistics stands out as truly different from mathematics. In fact, the histories of the two oldest academic statistics units in the country, University of the Philippines at Dilemma School of Statistics (UPPED-AS) and UP Los Banns Institute of Statistics (PULP-IS) are remarkable for the absence of substantive roles by either mathematicians or mathematics departments. UPPED-AS used to be the UP Statistical Center which was established in the sass with help from a United Nations Development Program grant.

The main objective of the Center was to train government statisticians for the Asia-Pacific countries. Its faculty comprised mostly of UNDO expatriate experts and the first offerings were a master's and shorter non-degree

programs. Undergraduate and PhD programs were added much later after termination of the UNDO assistance and the Center was absorbed completely in UP. Although statistics in Los Banns started at about the same time, its development followed a different path. The need for statistics grew out of the active research programs in the campus's only two agriculture and forestry colleges.

A Bachelor of Science in Agriculture major in statistics was instituted which was administered first in the Department of Agricultural Engineering and later in a newly created Department of Applied Mathematics. When UP Los Banns became an autonomous university, BBS and MS in statistics were instituted in early sass under the Department of Mathematics, Statistics and Physics of the College of Arts and Sciences. The three disciplines were split in the sass, with statistics becoming the resent PULP-IS which also added a PhD in its curricular offerings.

A few have started statistics programs, e. G. MS-IT in ' lagan City and CLAUS in Munson City by graduates of UPPED-AS and PULP-IS respectively. After these two, however, it is hard to find more success stories. Quite often, statistics graduates work in isolation or in mathematics departments. A greater number go into non-teaching posts in government and industry or emigrate abroad. Lately, even UPPED-AS and PULP-IS have been having difficulty keeping their own faculty and graduates away from industry, particularly information technology, communication and market research impasse. . Why it is important to treat statistics as a separate discipline 16. So, with the dearth of statistics curricula and graduates, who teaches statistics in Philippine colleges and universities (never mind high schools for

the moment)? 17. In 2007-2008 the AS, with support from the Statistical

Research Training Center (SORT) and ACHED, put 300+ participants through

a " Training Course for College Teachers of Basic Statistics" (AS website,

pop. Cit.).

Round table discussions on the teaching of statistics were held during Saga's

2008 and 2009 Annual Research Conferences. Pending availability of more

accurate data from a comprehensive study, experiences from the above-

mentioned activities indicate that those who teach statistics are mostly BBS

in mathematics, BASE major in mathematics teaching, and a smattering of

engineers, accountancy, economics and social science majors. 18. How do

these teachers approach the teaching of statistics?

Concerning mathematicians, the experience abroad is described by Garfield

and Ben-Xvi (2007), who noted that many who find themselves teaching

statistics are uncomfortable with the messiness of data and the different

possible interpretations depending on the ethos used in collecting the data

and the different assumptions made. Without additional statistics training

they would tend to focus on numbers, computations and formulas and will

have difficulty progressing to statistical inference.

And their students would equate statistics with mathematics or worse,

arithmetic. 19. In order to develop statistical reasoning, Moore (1998)

suggests that students must experience firsthand the process of data

collection and data exploration. These experiences should include

discussions of how data are produced, how and why appropriate statistical

summaries are selected, and how conclusions can be drawn ND supported.

These suggest, at the very least, discouraging a purely lecture type basic statistics course, in favor of one with hours of laboratory.

It follows also that teachers should possess similar and deeper first hand experience. However, with the exception of non-mathematicians who may have been drawn to teaching statistics because of actual research experience, basic statistics teachers in the Philippines generally do not possess adequate hands on involvement in collecting, handling and and Ben-Xvi described, namely a basic statistics course which at semester's end overfed mostly descriptive statistics and very little statistical inference.

This is most unfortunate since nearly all undergraduate students will have Just this one statistics course in their curricula. 20. With what do statistics teachers teach? Foreign books are very dear relative to third world incomes. What some schools like UPPED-AS and PULP-IS do in their introductory courses is prepare syllabi, guides, manuals and exercises and sell these at cost to their students. Some teachers have written books published by local companies.

In 2005 the AS created an ad-hoc committee to review nine locally tutored books prescribed as introductory statistics course texts in many universities. The committee found these books were authored mostly by teachers with no advanced degrees in statistics, and sometimes no undergraduate statistics degree as well; some were packed with logic and technical errors; used very few real world data and examples; and most did not encourage nor require the use of computers (David and Malignly, 2006). 21.

Furthermore, majority of Philippine colleges and universities do not have computing centers with open access to students. Statistical packages, e. G. AS, SPAS, STATS tend to be costly to lease or purchase, and what are often available to a select few faculties or colleges, but not to undergraduate students, would be pirated copies. Beyond packages bundled with operating systems, e. G. EXCEL, students get little opportunity to be exposed to statistical analysis software that can be downloaded free from the internet. 22. Who take statistics courses?

Pending a more thorough inventory, some relevant parts are gleaned from this writer's notes taken during September 2009 AS Conference's Roundtable Discussion on Education Issues which was attended by mom of our universities' statistics teachers and administrators. 23. There is no one in the Canaan State University faculty with a statistics degree. Inference is hardly covered in the elementary statistics course, where the main aim in requiring it is " it will be used for thesis". There is an obvious contradiction here; but it is understood that the emphasis is to present the tools, e. G. Owe to compute the t-statistic and perform the test, and little else. At the University of SST. Losable – Backlog, all programs have one course in introductory statistics, except psychology and education which have two each. Likewise, it was mentioned that " all courses are focused on thesis writing". All undergraduate curricula at the Polytechnic University of the Philippines-Manila, which has 60, 000+ students, require a statistics course. One-fourth of Central Philippines University-Oldies 11, 000 students take an elementary statistics course each year and they struggle through " statistics teacher shortage and bad books".

CPU is one of few universities with an officially licensed statistical software (SPAS) which cost 600, 000 pesos. Nave Viscera State University president, who has postgraduate training in statistics, confirms that most curricula in ere campus have at least one statistics course; consequently, her main problem is 24. Thus, there may not be much difference between what is happening now and a proposal by a recent national statistical system review committee (2008) to require a basic statistics course in all undergraduate curricula in Philippine HE's.

The problem is the low competency of the teachers currently assigned to teach the course. The cause is clear: while Bachelor's degrees in mathematics, chemistry, physics, zoology, etc. Abound whose graduates go on to teach courses in these disciplines, there are very few HESS offering BBS in statistics. The solution seems clear: more HESS should have BBS statistics in their curricular offerings. There is a catch, however: where to get the teachers to handle BBS statistics courses. This problem goes beyond the scope of this note. 25.

We do not know how it came to pass that majority of undergraduate curricula in the country HESS require a thesis. It seems that the consequences of such a requirement have not been sufficiently anticipated or studied by ACHED and university administrators, including the severe shortage of teachers who can advise competently on the statistical aspects of the theses. In general, the same introductory autistics teachers are also the theses advisers; however, one who is not trained enough to teach basic statistics cannot possibly be good enough to dispense statistical advice, especially if (s)he has not been actively involved in research.

In a desperate effort to solve this very serious problem, introductory statistics courses are " used or focused on thesis writing. " This may lead too lose-lose, not a win-win outcome. On the one hand, it is uncertain that the learning curve for statistical concepts and methods needed to plan and conduct research – albeit at the undergraduate level – can be compressed in one semester. On the other hand, focusing on thesis writing would most certainly deviate from the learning goal of an introductory statistics course in an undergraduate education, which is for students to develop an ability to think and reason statistically.