STATISTICAL LITERACY, NUMERACY AND THE FUTURE

Peter Holmes, Senior Consultant,
RSS Centre for Statistical Education.
Nottingham Trent University, Nottingham England

1. BACKGROUND

The following is taken from a transcript of an address given on 31 March, 2003 at Augsburg College in Minneapolis, Minnesota. As an informal presentation, it is not expected to be as complete or as deep as a written paper might be. Peter has reviewed these materials and made those changes he felt necessary and/or appropriate for a written version without losing the flavour of the original oral presentation.

2. PERSONAL INTRODUCTION

Introduction by Milo Schield: "*Peter Holmes is a Chartered Statistician, a Council member of the Royal Statistical Society, and a Senior Consultant at the RSS Centre for Statistical Education at Nottingham Trent University in Nottingham England. I first met Peter in fall 1995, when I spent my sabbatical at the Centre. Peter seemed receptive to some of my different ideas and we've stayed in touch. As the Director of the W. M. Keck Statistical Literacy Project at Augsburg College, I've asked Peter to review the materials being used to teach Statistical Literacy at Augsburg College and to make whatever comments he felt appropriate*.'

3. INTRODUCTION

I first met Milo some years ago when he came to Nottingham and was working on sabbatical with us at the RSS Centre for Statistical Education and came to know that he had some unique interests. I was not sure that they were the sort of interests I ought to have until he said, "Read this table." I thought I could read tables. He said well does it mean the percentage of mothers with babies are this? Or does it mean the percentage of mothers with this are babies? I thought, if I think hard I can do that. Then he showed me one that I could not tell which way around it was. Then I thought I'm supposed to be a Statistician. I'm supposed to know these things. I'm supposed to be able to do this sort of stuff. I even teach other people how to do it. Then he said, "It helps to think about the language that you use and then look at the difference between what's a whole and what's a part?" We spent a lot of time talking about this and it led to some interesting discussions. There is a saying that England and the United States are two countries divided by a common language. We found that was really true. We weren't using words in quite the same way, but it was quite crucial how you did use the words when trying to communicate information from the tables

My brief tonight is to look at statistical literacy at Augsburg and say what it is. Milo has described the work that he's doing earlier So I'm taking my brief more to say how does what is being done at Augsburg link with other aspects of statistical literacy and the work other people are doing in statistical literacy. Where does it fit in with the general statistical process? Where does it fit in with the sort of thinking that you need if you're going to think statistically, particularly if you're going to think statistically as somebody who's going into in business or commerce, or somebody who's going generally into everyday life. Now there are overlaps, but there are very special specific things in the work that Milo is doing here at Augsburg and I want to point them out to you. But I do want to show that they link in with things that other people are doing.

4. STATISTICAL LITERACY & NUMERACY

Let's start by thinking about statistical literacy. This is another one of those things, which divides England from America with the way that we actually speak and use words. This is a growing area of investigation. There are more and more people who are trying to say, "What is statistical literacy?" The literature about it is growing. You have some of the biggest experts in the world here in the United States who are spending time looking at it: Joan Garfield from the University of Minnesota is one of them. Beth Chance from the Cal-Poly in California is another. Another major person is Ido Gal in Israel. But there are many people who are trying to look at statistical literacy and saying: what does it actually mean to be statistically literate?

I think the whole thing started in England. Brits do start some things. We started with a word. We had a word that you didn't have. In 1959, there was a government report in England that talked about the numeracy problem. In the context it was talking about the education of 16-year-olds saying that they needed to be literate. There was a literacy strand, but they also needed to be numerate. So there was a numeracy strand. So from 1959, we have had a very good English word called numeracy. I came over here in 1980 and I was talking to Bob Hogg as part of the National Council of Teachers of Mathematics and the ASA Joint Committee on Statistical Education. They were putting in a proposal for what they called the Quantitative Literacy Project. I said, "That sounds a bit like Numeracy." And they said, "What?" I said, "It is an English word. They said,

"It's not an American word." So that's why you have a Quantitative Literacy Project and we have Numeracy. The logo on my shirt that's about the RSS Centre says "Promoting Statistical Numeracy". Milo's says "Promoting Statistical Literacy".

Now of course when things start they can grow. What was essentially the same thing to start with had more and more people working on it and picking it apart. There's now "Statistical Numeracy," or there's "Statistical Literacy," or in amongst all that, there's "Statistical Reasoning". And then there's "Statistical Thinking". Now all of these are to me part of the same thing and picking them apart is to try to say, 'This bit is more important. – we must make sure this bit is done and this bit is done. But they're all in the same ballpark. The word numeracy when it was first introduced was in the context of the ability to use numbers in practice. So it was particularly in the context of statistics that you might have to read and interpret. In fact in that first use of it in 1959, it was in terms of reading tables. So it is very good history if you talk about reading tables as part of statistical numeracy. That's where it actually started in 1959.

5. STATISTICS IN PRACTICE

If you're talking about statistical literacy in any form, you have to say that it has to reflect the way statistics is used in practice. I ran a major project in the United Kingdom for curriculum development for statistical education in schools. I had a Steering Committee that said to me, "Well what is statistics? I said, "Well you're statisticians. – you're my Steering Committee. – you tell me what is statistics." We decided that we mustn't draw the definition too narrowly to rule out anybody. The first definition we came up with and it's still my working definition, is *statistics is what statisticians do*. So you look at what statisticians actually do do, from the government statistics people to those in pharmaceuticals to people working in marketing. There is a wide range of people using a wide range of things in statistics. So if you're talking about statistical literacy you start by saying it's being numerically literate in all those areas. If you're trying to do cover the whole range of uses, you have to say there are going to be different levels of statistical literacy. There is perhaps a basic statistical literacy that you'd like everybody coming out of high school to have or perhaps some of it coming into University courses. But then there might be certain add-ons. The person who is statistically literate in a pharmaceutical world would have different add-ons to make him or her statistically literate than the add-ons for the person who's going into business and who would not necessarily be coming across significance testing, hypothesis testing, in the same sort of way. So my view on this sort of basic statistical literacy would include the aspects of statistics that all sorts of people use, and then you'd build bits onto that that are relevant to different groups and different communities.

There is actually a very important difference I think between statistical numeracy and statistical literacy and that's shown by a very early definition that came from the UK about statistical numeracy. It is from a government report on the teaching of mathematics in schools. It talked about statistical numeracy requiring a feel for numbers, an appreciation of levels of accuracy, making sense of estimates, a common sense approach to data in supporting an argument, an awareness of the variety of interpretation of figures, a judicious understanding of widely used concepts such as the mean and percentages, all these are a part of everyday living. That's statistical numeracy. Statistical literacy would add to that the ability to read and communicate in those areas. That makes it literate as opposed to just numerate. It adds words in as well, so you need to be able to know what the words mean when you are communicating. It isn't just about the figures.

6. FOCUS ON CONFOUNDING

I looked at the literature from Milo when it first came out and said, "What is this on this flyer that says 'Statistical Literacy at Augsburg'?" Milo was talking about the strength of statistics. I think, from what he said earlier, that this means the strength of arguments that you can get based on statistics. The strength of the argument depends on what you're doing. If you're doing small and well-designed experiments, then the strength of the argument depends on chance. Is it a small sample? In which case there is a wider variability and you use chance to measure confidence in your argument about the real world. In your small well-designed experiments you're really concerned with controlling chance to make sure you know the chance, you know the probability so can draw valid inferences.

If you're looking at poorly designed studies, whether they are experiments or surveys, then there's a real problem about bias and you want to make sure you haven't got bias. This is the second point in the flyer. Finally Milo's saying that if you look at large well-designed observational studies, there's a problem of confounding.

That drove me back to saying: What do I understand about the statistical process and how has it changed over the years? I want to go though with you what I had always thought of as the statistical process and I still do think it's the basic statistical process. But let me show how over the last 10 years or 15 years particu-

larly, one thing has change dramatically. It's changed because of the large amount of data, the large data sets that we now have. But we'll come to that later.

7. THE STATISTICAL PROCESS

Figure 1 illustrates what I consider the basic statistical process.

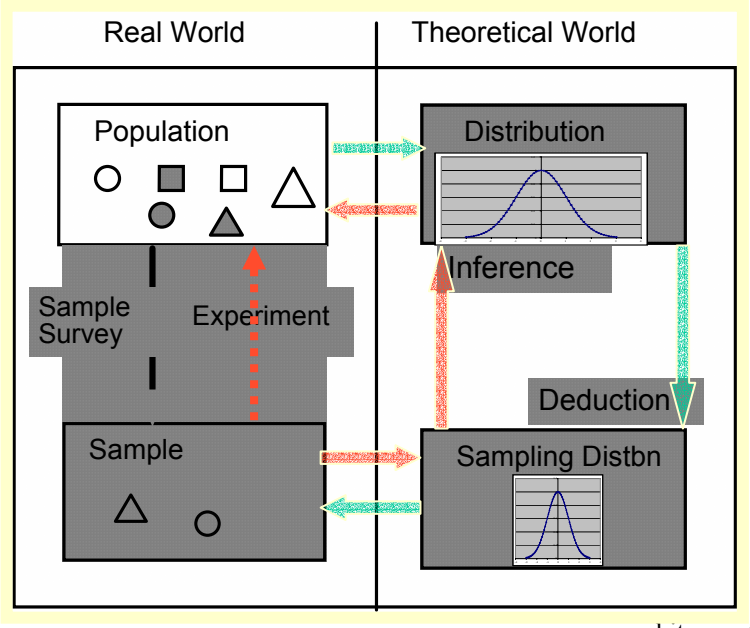

Figure 1

The statistical process it seems to me starts with the real world, and in that real world there's a population. These are real things with real qualities. This is the top left corner of the diagram and I have drawn circles and squares and they're shaded or they're not so shaded. They're real things. The essence of statistics traditionally is you go to that real world and you draw from it some sort of real sample. So that sample is a subset of the population. The sample should in some way be representative of the population. This might be done in one of two ways; either by carrying out a sample survey or by doing an experiment. I've always used the phrase *sample survey*. Milo uses *observational studies*. I think there are pros and cons for using either of those two. They tend to overlap anyhow. Whichever phrase is used it is the same point. You can either design your experiment and get your sample, and that way ensure that it's representative and you have all that goes with sampling in practice.. But now what do you want to do? You want to say; given that sample, what can I say about the population? So you really want to follow the arrow from bottom left to top left. Given the sample, what can I say about the population? The population includes at least one triangle. Very interesting. Not really We are not in the business of deduction. We're trying to say; what can you infer from the sample about the population? The problem is you really can't do much about inference that way without something else, not in this structure.

Do you know the story about the pure mathematician, the statistician, and the social scientist who are travelling on the train? They're travelling on a train in England through Yorkshire. Yorkshire is a big county and it has a lot of fields as well as a lot of cities in it. They were on the train and they looked out of the train window and they saw in this field one sheep, a black sheep. Now I don't want to be rude to the social scientists, the point is not about different subjects as so much about the difference between inference and deduction. So if there's a social scientist here please forgive me. The social scientist said, "Oh look, the sheep in Yorkshire are black." The statistician says, "Well no you couldn't quite say that. You could say there are some black sheep in Yorkshire." And the pure mathematician says, "Well you can't say that. All you can say is that in Yorkshire there is at least one sheep that is black on at least one side." Now to the pure mathematician that was a deduction. The social scientist was perhaps a bit over the top. The statistician may or may not have been right. He said there were some black sheep. He was extrapolating from one to say there may be more. He may have been wrong. It was an inference, and the essence of inference is that you make your statement and then try to quantify the level of your guessing, if you can.

So how do you go from bottom left to top left? It's really interesting for you to try to analyze that. You start by looking at the population and your interest is in some particular measures in that population. Now looking at that population we assume a lot of things about it and model the measures in that population with some sort of distribution. We start moving clockwise round the diagram and move into a theoretical model of the world. I've drawn a normal distribution, but it could be any sort of distribution that you think follows the assumptions you make about the population. That's deducing something. If the population is like that, this is what the distribution will be. Now if that distribution parallels the population, I will sample from that distribution in a way that is theoretically parallel to the practical ways I sampled in the real world. So I then get a sample with a sampling distribution of a particular statistic that I'm looking at. And that is deduction.

Finally, in the clockwise direction, I compare the two across the bottom. That's the sampling distribution. I happen to have gotten one particular sample and I now work anti-clockwise and say, suppose this sample were taken from this sampling distribution, Then I ask what I can infer from that what sort of distribution it came from – or some values of the parameters in it. Once I've gotten back to the distribution I now say; having done that inference, what does that imply about the population? And that's the traditional view of how you would do an inference. You deduce things to get distributions and then you infer back through to get things like confidence interval, significance tests and so on.

Now you have to question whether that is all that there is to the statistical process. Does it represent all that statisticians do? You can say that the diagram was trying to answer the question: What does the sample tell us about the population? If you have well designed experiments, you will avoid confounding, you will avoid bias, and you will use chance. An experiment is nice; you can do those sorts of things. Then you come into things like the social sciences and business and so on, and you realize that you can't allocate people to particular things in real life. Let me illustrate.

I was teaching some 16 year-olds in a secondary school in England and I was talking about the problem of smoking and lung cancer. I said that there are problems, of course, because people choose to smoke and you can't therefore distinguish between whether people are smoking because they're a certain sort of people and it might or might not have caused their lung cancer. That was Fisher's argument, which I'll come back to later. So I asked, "Can you think of anything we might do about that?" And one young lad put his hand up and said," I've got an idea sir! I've got an idea!" So I said, "Come on, what's your idea?" He said, "We can do a sort of experiment." I said, "Yes?" He said, "Well we could sort of put some of us chosen from this class and say; you're going to smoke twenty cigarettes a day for the next six months. And to the others, we're going to say; you're not going to smoke anything for the next six months. At the end of six months we could see how healthy they were." I said, "That's very good. Anything else you want to say about that?" He said, "Yes, I'd like to be in the smoking group." "Oh? Anything else?" He said, "Yes, you can buy the cigarettes." I said, "Well, can you see anything wrong with that?" And one of the others said, "I don't want to smoke." And I said, "That's right you can't force people and you can't make it a proper experiment can you? Not in our society." So what happens if you can't allocate randomly or whatever it is and you have to convince. You are then surveying and if you look at the literature, you find that in a lot of the surveys you're trying to match what's considered the experiment. It is a gold standard. You find some way to find a control so even if you can't randomly allocate it, it's as good as randomly allocating. Your survey is really trying to say: If only these things are okay we can use the same mathematics, we can use the same principles as they do for the gold standard allocating randomly things and it's as good as an experiment.

But then you look at large observational studies and they start to fall outside those categories. That made me start thinking again; what is there about large observational studies? How do they differ? Are they important? I realized that they are actually growing in number. There's huge numbers of them.

8. STUDIES ON POPULATIONS

If you think about going into a local store, you might have a loyalty card. Every time you use that loyalty card in a supermarket or whatever, all the data of what you have bought are collected. You may go to a hospital. There is legislation that says that hospital has to record various things about you. So everyone who goes to the hospital has all this data recorded about them. If you work in any kind of administration, say in education, you know there are requirements that you have to fill in; data of every student, of every faculty member, this, that, and the other. You've got huge, huge amounts of data, which are essentially population data. {We'll come back to that in a minute.} You've got huge amounts of data, which aren't surveys, not in the real sense. It's not even an opportunist survey. It's much more than that. It's all of the data you can get about all of these people who are doing some activity. The data are collected routinely. You have traffic information that is collected on every car. You have, what I call local authorities, you have a different word over here, having to report every crime that's committed or every crime that's reported, and things like that. So there are huge databases all over the place and these are effectively, large observational studies. If you look at any one of them, let's take the hospital one, you will often find that the same group of people, the same subset can be considered a sample and also a population. All the people who come in through hospital 'A', could be considered as a sample of people that might come into the hospital. It would be a poor sample, it wouldn't be a representative, but it is a subset of all those that might come into the hospital. On the other hand it is *all* of those who came to the hospital, so it's the population of all who came to the hospital. From yet another point of view, it's a subset of all the people who've been to hospitals. And then you get into the business of can you compare hospital 'A' with hospital 'B' with hospital 'C' for all the patients who went to them. Now you're in the business of comparing them

you have the confounding sort of problem in the Simpson's paradox Milo gave earlier.

9. INTERPRETING STATISTICS

It's easy to see that that model I gave earlier doesn't cover all the aspects of what you are having to do now in statistics. We want to be able to say; what can we deduce about causality, about the relationship between different things with huge databases where it isn't so much a sample, these *are* the populations. It's not the variability of any estimates that's of prime interest. One of the examples I saw in Milo's class this week was from the mid-term test. There was a table. It said something like this: Fifty percent of married males age 25 to 34 would not remarry their current spouses if they had to go through the process again. Now, you can say that may have been a sample, but suppose it was a sample and that from sample you estimated that 50 percent was the population figure plus or minus 3 percent with the confidence interval. It really isn't important that it's 50 plus and minus three in terms of social consequences. What's important is that it's near 50. You might be interested to know the other thing that went along side of it, if I remember rightly, was that the equivalent figure for females was 65 percent.

It isn't the inference from the sample to the population here that is of prime importance. What's of prime importance is those figures themselves. It really doesn't matter whether it's 50 or 52 or 48, or 65 or 67. Those are big numbers. What is it saying about society? Why did those things happen? Those are the interesting questions, not the sampling variability. When you have huge data sets, which are essentially populations, it isn't the sampling variability that's important. It is the actual figures themselves and what are the connections between them. Hence, I think that this is an important part of what I would now put into statistical literacy which I wouldn't have put in 20 or 30 years ago, because there wasn't so much of this sort of data around.

The whole balance has changed. A whole lot of data mining seems to be in this sort of area. The interest then is what caused the data to be as they are and what are the underlined mechanisms to make the data as they are. You're getting out of the realm of statistics here, and getting into the realm of interpreting statistics in different subject areas.

The Augsburg material that Milo sent me describes two major areas for statistical literacy: One is on interpreting data and one is on identifying relationships and causes. I'm trying to show how these link with other ideas of statistical literacy. There is another one that you may not want to take on board, but it exists, a third major area. This is on how to make decisions based on data. Although this isn't listed, listening to Milo and seeing the sort of things he's doing, it's implicitly there and you might want to think about making it explicit. Making decisions based on data is a part of statistical literacy.

10. READING STATISTICS IN TABLES

When you are interpreting data from large studies; then typically the output is in tables. It's important to be able to the tables. This isn't new. I remember seeing it first in a presentation by a Professor of Business at the London Business School called Andrew Ehrenberg. In 1982, he gave a talk at the First International Conference on Teaching Statistics where he said we must preach what is practised. And what is practised is that a whole lot of data are collected and put into tables. You have to infer from them which particular marketing methods are working and which aren't. He was saying, "We don't actually teach this. – we ought to." We must preach what is practised was what he was saying. He came up with a whole lot of techniques which I'll mention later

This is not a skill that even the best newly graduated statisticians have. I act as an interviewer to sort out applicants who are applying for posts in, the Government Statistical Service. They're applying to go on what's call the Statisticians Fast Track. The Fast Track means they're expected to make very quick progress to become fairly Senior Civil Servants in the statistical area. They come with good first or second degrees in statistics or with a major statistics component. They come as very intelligent people. It's the top group of people that you might be expected to see coming out of UK Universities with statistics qualifications. One of the exercises we give them is present them with something like six different tables from government publications all in the same area and say: 'Look at those'. Start thinking about what recommendations you would make to the minister in that particular area for how you might change government policy or how you think government policy is working in …. The area could be crime, or education, things like that. They have real difficulties. They work together, they talk to each other, and they misunderstand and misinterpret the tables time and time again. Very often they will misread a row as a column percentage. By very often, I mean about 20-25 percent of the time. Now admittedly they're working under pressure they're working under a lot of stress, but it's a much higher proportion than I would like to see. If they misinterpret these things, - the percentage of people who are using cars to go shopping, or the percentage of people who are shopping in their cars – it can make a big difference to what you are

going to tell the transport minister. It's worrying and I would like to make sure within our statistics courses in our universities that sort of thing is covered as part of it.

So what do you do when you're reading tables? When you look at the sort of thing that is actually published, it's quite clear that there are all sorts of different tables that are being published in different ways over time. And somebody thought, "I've got these data and wouldn't it be nice if I could make it more concise by missing this out or I could contract in this way and all the information is there, nicely, very densely concise. Nobody sat down and said, "What are the rules that we must obey? What are we going to do?" Things have just grown and developed in this way. What Milo has done it seems to me, is to look at those and say: Let's see what underlying rules we can see in these tables as they've developed. It's more like science isn't it? You look at biology and you say, what are the underlying rules given the evidence that we've seen? And he's come up with some quite interesting rules that would help you interpret the tables. The ones I always found difficult were like the example he gave earlier about mothers and babies where you've actually got to get the 'whole' from the row title and the column title and you get something like 19 percent of mothers with babies in that age range have - whatever it was. It really is important to distinguish being row and column conditions. Consider these invented figures: saying 25 percent of men are smokers would hold very different health implications from saying 25 percent of smokers are men. With the second one, you'd be saying, "Why are so many more women smoking than are men?" In the first one, you'd say, "Oh well, is that going up or down? Is 25 percent good or is it bad? What are the trends in it?" If you want to link that with more traditional statistics, it's very important to distinguish between the condition of 'A' given 'B,' or 'B' given 'A'. This is something we major on quite a lot of in traditional statistics courses.

So when you start reading tables, a key to reading them is to be able to recognize which is a part and which is a whole. Is it the row or is it the column or is it much more complicated than that? And then you have to say, "Well how can I communicate that?" Because with literacy it's not just about reading numbers it's also being able to communicate them, or to read what other people say about them. You then have to say, "Let's look at a careful study of the grammar that will help us write and correct and use unambiguous statements." It's essential to be able to write and say them correctly so that you'll be not misunderstood. Milo has come up with a whole lot of examples, and this is where we've had a lot of our conversations. I would say things like, "Well it may be true Milo, but I would never say it that way." And he'd say, "Well you're just English." So within the stuff that Milo is doing there's a lot about how you can interpret a cell within a table and that's quite hard, and it can be very hard in some of the tables.

I don't think you ought to overlook though that reading tables isn't just about looking at cells. There are a lot of other aspects that you need to think of. One of them is; can you help people who are responsible for constructing tables to make them readable in the first place - instead of contracting them to the maximum number of figures in the minimum space. Would it help if you just gave a bit more space, and instead of missing out the hundred percents, actually put the hundred percents in, things like that. Some of our students will go on and will have to write their own tables and you might get them to write so that people can read them in the first place.

Other things you forget quite often are the title and footnotes. Footnotes are often very important in terms of reading what the table has said. There is often an effect of changing definitions if you're looking at trends. I remember looking at our own employment figures in the UK from 1972 to 1981 and there were sort of big jumps every so often.

I looked underneath at the footnotes and found that over a period of nine years there had been thirteen changes in definition. Now I don't want to say anything about Margaret Thatcher, but it was in her time and of those thirteen changes, changes in definition, eleven of them had the effect of making the numbers smaller. Now that's not a political statement, is it?

11. IDENTIFYING TRENDS

Another aspect of reading tables is to try to identify trends. You may want to identify errors or outliers, (things which have been put in that table but shouldn't be there) they're just wrong. I heard a lovely story from the chief statistician of the Government's statistician service, Sir Claus Moser, from about 15 - 20 years ago He said, "If you read any government statistics tables and you find any figure that is interesting, it is almost certainly wrong." His point was this: If you read the table for this month and the table for next month and the table for the next month, basically you should be seeing the same patterns and nothing should change greatly It so happened that I was looking at the changes in population in towns around Sheffield. The figures were on my desk and showed that every town over the past ten years had grown by between 3 and 5 percent, except one that had dropped by 10 percent. And I thought that's interesting. Then I thought about what Sir Claus Moser had said and I thought it's probably wrong. And sure enough it was. But it was a very interesting mistake. The population at the second

stage, instead of saying 13,100 had said 11,300. The one and the three had been turned around. So instead of what they calculate as a ten-percent drop, it happened to be roughly a four-percent rise. It's easily spotted once you know it. I wrote to the government statistical service and said, "Did you know this and had you compared it with what Sir Claus Moser said. You know, this is your boss. Don't you listen to him?" I got a very nice letter back from some junior statistician in the civil service saying, "Yes we do understand what our boss says is correct, but it's completely impractical to do anything about it."

12. CORRELATION AND CAUSATION

Interpreting the figures back into the context is very important and the statistician, if he's statistically literate, should be doing something about that as well, and helping others to do it. Suppose you have a whole lot of data and you are trying to identify relationships. In standard statistics courses, you often come across the phrase that that correlation is not causation. Essentially that's a negative message? You' have this lovely correlation but you say, oh it's not causation, so you just forget it. One of the nice things, I think that Milo has done is go to other people and find that we don't need to stop there. You can investigate the possibility of underlying causes. There is a growing amount of literature about how you can identify cause; cause that is identified by changing conditional probabilities. Milo didn't mention that in his talk, but there's some of that in the work that he's doing; how you can identify a cause within a correlation. It's not just to do with confounding. There are new approaches to the definition of cause. Judea Pearl is at the forefront of this. Those of you who read the International Statistical Review, will have seen a series of articles in the issue of about three years ago, looking very hard into the statistician's view of what is meant by cause. This follows up the work that Jerome Cornfield did on whether smoking causes lung cancer. How can you rebut Fisher's argument? Fisher said; well there may be something genetic that both causes you to smoke and would also make you more liable to have lung cancer. How can you answer that? Cornfield came up with some fairly simple things of how you could, in his paper of about 1959. It must have been before 1962, because that's when Fisher died.

The issue of confounding variables is a major problem with all observational studies. You are trying to say does something cause something else and how much does it cause it. This is a nice piece of work that Milo has picked up and is running with. Again you can go beyond identifying confounding factors, which again is essentially negative, to measure the effect of that confounding factor. This takes you a bit further. If you can say; this is still there, allowing for this, then you've travelled even further. If you can do that without going into high-powered mathematical multi-variable regression, than you have a chance at getting the message across to people who are not mathematically profound. And that's most of us. By allowing for confounding factors, it's possible to make stronger claims about cause and effect. As Milo pointed out, Simpson's Paradox is just an extreme case of the effect of a confounding variable. The diagram that Milo used he got from Chance magazine. I also saw it there and that was the first time I've really understood Simpson's Paradox. But I hadn't identified, which Milo does, that the horizontal axis is the confounding variable. I use to go around saying, "I have these data that say that if I'm a male I should go to hospital 'A,' if I'm a female I should go to hospital 'A,' but if I don't know what I am, I should go to hospital 'B.' So which one do you go to? Because I may not know whether I am, male or female, but I know I'm one of them. Ah well!!

13. CONCLUSIONS

So, some conclusions. The Augsburg Course is different. It has a different emphasis from many other courses to establish statistical literacy. It comes from a different background, but that it has a lot of overlaps is what I've been trying to share. In many ways it reflects better the amount of the data that come as part of every day life, certainly from large observational studies.

The traditional statistical literacy courses that are being developed are much more geared and have much more emphasis on the traditional sample to population (confidence intervals, significance tests and similar things). These are getting to be less and less a part of the sort of statistics that a lot of people meet in their everyday life. It reflects that balance better. It does include important material based on what is used. It contains some genuine statistical insight and teaches some very difficult statistical ideas with some insightful diagrams. At first sight it seems to include little on chance, but in fact if you look at some of the proportional reasoning and a lot of the inference, it is essentially probabilistic reasoning. A lot of the times you talk about proportion it can be called probability. Talking to Milo and seeing some of the material he is working on, I expect more of this as the course develops.

In his approach, what he puts together is unique. That's not to say that the individual things are necessarily unique, but the package as a whole comes off as a very different package. Because it draws an existing work from different places, and from places that reputable statisticians rarely visit, he forced me to go there.

I still think I'm fairly reputable and find that the course draws on ideas from areas which have not been in the traditional mainstream of statisticians. But they are there and they are statistical and we should be drawing on them.

As with all new courses, the approach reflects the developer's interests and insights; sometimes some topics get overemphasised. My own personal view of the present course is that there's a bit too much emphasis on all the different ways you can write phrases that mean the same thing. Most of us, I think eventually settle for one phrase which we use with percentage, another which we use with rates, etc. As long as we're happy that we can unambiguously write it in one way, then that's fine, but we should also be able to read the other ways so we can understand them. I think that's one place where at the moment there might be a bit of a overemphasis because I know Milo spent hours and hours when he was at Nottingham looking at the way people said things. It's a great insight, but I think some of that might have to be toned down a bit, reducing the number of them in due course. Please don't take that as a negative. I'm trying to say that any new course has its rough edges and needs refining and so because of that we should be continually refining it in the light of experience. Generally when new courses are developed, other people start taking what they feel they can fit into their work. That's the way most curriculum development seems to work. Some people try something that is different and if it works other people take it on board.

I'm convinced that the standard first course in statistics, which focuses on getting to significance testing and confidence intervals, hasn't been a good grounding for a lot of people in statistics and I don't think it's an appropriate aim for a lot of students. They seem to get very mixed up with it, particularly on confidence intervals and what a significance test does or does not show you. So I think this course's emphasis is much more in line with the sort of statistical literacy needed by most people in everyday life so you can read the news, those who are in business commerce or management and for policy makers.

So I look forward to seeing how the course develops and I hope you will encourage Milo in these new things he is putting together. It's really very exciting. I was impressed by hearing students say: Yes, we think it's a very important thing. This is far better as a reason for doing a course than the more cynical and vindictive reason I have heard about other courses: I had to suffer through my first course so my students ought to suffer through theirs.

Acknowledgments: This talk was given under a grant from the W. M. Keck Foundation to Augsburg College "to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts."