TALK IN SANTANDERs

I am deeply flattered by the honour you are bestowing on me today. Not only is the University of Cantabria a rising force in the academic and intellectual life of Europe, but the region is a cradle of civilisation in this part of our world. It is arguably the most archeologically rich region anywhere, with evidence of human occupation since the Old Stone Age, and with many cultural icons from other periods of human settlement.

Formally, the Province of Cantabria dates from 1778, a decade before the first white settlers arrived from England in my own country, and during a period of considerable international engagement by Spain.

Despite fierce resistance from the independent people of Cantabria, the Romans established a foothold here in Santander, but you gained independence as long ago as the year 409. In contrast, my home city of Melbourne was founded only in 1842, although native people had inhabited the area for perhaps 40,000 years. So, as you can see, in some respects there are considerable differences between us.

However, we also have much in common, especially the very high regard in which we hold intellectual achievement. This, after all, is at the heart of the development of the economies of both our nations. The economy of Spain is based significantly on export of high-quality manufactured goods; for Australia, the fourth largest export is education (behind the three main mineral exports of iron ore, coal and gold).

In providing services to many areas of industry, and of course to science and education, the mathematical sciences are playing an increasing part in both our nations. We can single out Statistics for a special role.

The UCLA historian Theodore Porter wrote in 1986 that:

> Statistics has become known in the twentieth century as the mathematical tool for analysing experimental and observational data. Enshrined by public policy as the only reliable basis for judgements such as the efficacy of medical procedures or the safety of chemicals, and adopted by business for such uses as industrial quality control, it is evidently among the products of science whose influence on public and private life has been most pervasive.

Almost thirty years have passed since these words were written, and during that period the involvement of Statistics in our industries and our economies has deepened even further.

One of the most obvious changes in our world over the past three decades is the amount of data being collected. You cannot get away from it–on our roads, in the environment, when we go to the doctor, on our credit cards, in our institutions, as part of our scientific experiments.

And the sort of data we collect has changed too. Instead of making a few measurements on a large number of things, we now take masses of measurements of many characteristics of the same thing. Seldom today are data recorded individually by humans. Instead, they are recorded in massive quantities by machines, often with a great many components for each data point.

Such data are a huge resource, but it is often hard to extract the information we need. A great deal of effort in Statistics is directed at the problem of getting the most out of the data, of finding ways of squeezing it for the most clear insights we can get.

Carly Fiorina, a former Hewlett Packard CEO, in a speech titled "Information: The currency of the digital age", argued that:

> The goal is to transform data into information, and information into insight.

That is, indeed, the goal of much statistical resesarch today.

IBM, which today earns much of its income as a provider of consulting services, advertises itself online as a source of statistical advice about converting Big Data into information. As IBM writes:

> The volume, velocity and variety of data has grown exponentially, providing exciting new opportunities for analyzing financial, production, and customer activities. However, data alone has limited value. The real contribution to your bottom line occurs when you can turn data into information, and information into insight.

Hal Varian, Chief Economist at Google and emeritus Professor of Economics at UC Berkeley, remarked a few years ago that:

> The sexy job in the next ten years will be statisticians. And I'm not kidding.

And he wasn't. Statistical science today continues to be vibrant and vital, because it is in a constant state of change, developing new technologies for converting numbers into information and advice.

My own involvement in Statistics research started at about the time that significant interactive computing power began to become available in university Statistics departments, in the late 1970s.

Up to that point, those of us using powerful electronic computers in universities generally were at the mercy of punchcards operating mainframe computers, typically at relatively distant locations. This severely hindered the use of computers for assessing the performance of statistical methodology, and particularly for developing new techniques. However, once computational experiments could be performed from one's desk, and parameter settings adjusted as the results came in, vast new horizons opened up for methodological development.

The new statistical approaches to which this led were able, by virtue of powerful statistical computing, to do relatively complex things to data. For many of us, David Cox's regression model (Cox, 1972), and Bradley Efron's bootstrap (Efron, 1979), became feasible only in the late 1970s and early 1980s.

In 1979 Efron gave a remarkably prescient account of the future relationship between theory and computation in modern Statistics, noting that:

> The need for a more flexible, realistic, and dependable statistical theory is pressing, given the mountains of data now being amassed. The prospect for success is bright, but I believe the solution is likely to lie [with] a blend of traditional mathematical thinking combined with the numerical and organizational aptitude of the computer.

Critically, Efron saw theory and computation working together to ensure the development of future statistical methodology, meeting many different demands. And, of course, that is what happened, despite the arguments of some that advances in computing would replace theoretical statistical
arguments.

Developing theory for Statistics, in conjunction with computational methods, has occupied much of my career. I began life as a mathematician, working in probability theory, but over time moved into Statistics.

I remember a senior colleague advising me, in the early 1980s, that in the future statistical science would be developed through computer experimentation, and that the days of theoretical work in Statistics were numbered. He advised me to abandon my interests in theory and focus instead on simulation. Stubborn as usual, I ignored him, and time has shown that that was the correct course.

Indeed, the demand for mathematical theory, to help us understand what our methodology is doing to our data, and in the process to convert the data to information, has turned out to be much greater than many had anticipated.

In the development of modern statistical methods, theory has played a role that computation really could not. Theoretical arguments point authoritatively to the advantages of some computer-intensive techniques, and to the drawbacks associated with others. It is hard to imagine that numerical methods, on their own, will ever have the capacity to deliver the level of intuition and insightful analysis, with such breadth and clarity, that theory can provide.

I have had an extraordinary time, during my career, working on the development of new statistical methodology, largely from the view-point of mathematical theory. I am very fortunate to have had this opportunity.

Thank you.