

CHAPTER 11

Statistics



Terry Speed
*Professor of Bioinformatics
at the Walter and Eliza
Hall Institute*



John Henstridge
*Adjunct Professor of
Statistics at The University
of Western Australia*



David Warton
*Associate Professor in
Statistics at University
of New South Wales*

Just as it is difficult to quantify the benefit of having a workforce with an adequate knowledge of written and spoken language, so it is difficult to quantify the benefit of having a workforce who can adequately communicate about variability and uncertainty.

Statistics is more than a numbers game — it underpins all sciences

Terry Speed

Walter and Eliza Hall Institute

Contributors: John Henstridge¹ and David Warton²

¹*The University of Western Australia*

²*University of New South Wales*

We are all familiar with many instances of statistics in everyday life: the statistics of sport, weather, population, the stock market ... the sort of thing that might appear in the ‘number crunch’ column in *The Age* and *The Sydney Morning Herald*.

The central feature of statistics is data: designing ways to collect, summarise, visualise, present and draw inferences from data.

Fundamental concepts are *variability* and *uncertainty*. As a way of making sense of the great variety of ways in which we meet statistics, I will give a simple breakdown into three broad categories. The first category is statistical facts.

What was the proportion of Australians 15 years of age or older in 2001, who have never married? The answer is 32%.

Such facts are often interesting in themselves, or in comparison with the same fact for different times or places. They might seem like data, but in reality they are summaries of data. Behind the 32% figure is the 2001 Australian Census. As a summary of census data, it has no statistical uncertainty.

More complex are data collected in survey sampling of populations based on randomly selected units. Variability enters the equation, as different samples will generally give different results.

Sample survey data are typically summarised as proportions, percentages or averages, and the results are presented as inferences concerning the population. They will be uncertain inferences, because of sampling variability ('error') and other types of error.

The 2007 National Survey of Mental Health and Wellbeing conducted by the Australian Bureau of Statistics (ABS) found that an estimated 3.2 million Australians (20% of the population aged between 16 and 85) had a mental disorder in the 12 months prior to the survey. (Presenting the uncertainty in estimates like this one and explaining how that is interpreted is beyond the scope of this chapter.)

Let's call this sample survey statistics — the second category. Survey results such as those in our example are clearly important if we are to make informed decisions towards realising our aspiration, as outlined by the Chief Scientist in his introductory article.

Now let me turn to my third category: the enormous range of other uses of statistics. It includes the use of statistics by professionals ('statisticians') and people whose principal activity is something other than statistics.

Statistical subcultures

Sectors of the economy such as agriculture make use of statistics in characteristic ways.

Yet, when you search the web with key phrases such as 'statistics in agriculture', you will most likely find statistical facts about agriculture, not how statistics can be used to improve agriculture practice. That takes a more directed search, but the uses of statistics in this way are many and varied.

Over a long career in the CSIRO, Helen Newton Turner (1908–1995) applied statistics to animal production, and later to sheep breeding, and modern versions of her research continue to this day.

Statisticians work on comparative crop variety testing, which is a source of information on variety performance for Australian farmers. Similarly, the mining and manufacturing industries, and even the service industry, make use of statistics.

It is hard to name an area of science, technology, business, industry or government, even the humanities, which does not have its own statistical subculture — its own types of data, and collection of questions and models and methods used to answer these questions by analysing the data, alongside a statistical literature.

It seems evident that statistics pervades every field of human endeavour, either through the use of statistical facts, some form of sample-population inference, or other informal or formal methods of addressing questions of interest. As I will explain, this statistical activity need not, and in general will not, be carried out by people calling themselves statisticians.

A role in all aspects of health

We might start with summary facts concerning mortality and morbidity, which show where we are, as a prelude to where we would like to be. We can move to surveys to elicit more detailed information on particular topics than we have from routine data collections.

The mental health survey mentioned above is one example. Epidemiology works hand in hand with biostatistics in seeking to understand the patterns and causes of disease, and to inform policy decisions by identifying risk factors.

Biostatistics is used in the design, conduct and analysis of clinical trials to assess the efficacy of vaccines, drugs, devices, treatments or interventions.

It is also widely used in pre-clinical biomedical research, where cell lines or model organisms are analysed, and in the studies necessary before a drug goes to human clinical trial and beyond.

Industry quality

What about the role of statistics in knowledge-based industries and high value goods and services — aspects of the economy that many (including me) would like to see greatly expanded in Australia in the coming decade?

Statistical methods were first used widely in the manufacturing industry in the 1930s, building on the pioneering 1920s work of Walter Shewhart, who is known as the father of statistical quality control.

These activities expanded rapidly following World War II, particularly in Japan, where American engineers and statisticians joined their Japanese counterparts in making important contributions to Japan's reputation for innovative, high-quality products.

These developments broadened the role of statistics, moving it beyond the realm of statisticians. Statistical thinking entered the minds of both management and workers through concepts such as quality circles. Problem identification, data collection and basic statistical analyses became important aspects of the job.

In the decades since the 1950s and 1960s, the value of statistical thinking in industries of all kinds has become widely acknowledged.

Developments such as Total Quality Management, Six Sigma and the series of standards from the International Organisation for Standardisation (ISO), known as ISO 9000, have all contributed to the near-universal awareness of the value of statistics in helping 'define, establish, and maintain an effective quality assurance system for manufacturing and service industries' and in improving the quality of products and services.

It was understood from the earliest days that methods from manufacturing industries were also effective, with appropriate modifications, in the service industry, but uptake in that area has definitely been slower.

One thing is certain

Statistics, in one or more of the modes that I have described, has entered almost every aspect of human endeavour. We can use it for better planning, more efficient delivery of services, and increased productivity.

The value of statistics is only limited by a lack of awareness and the severe shortage of adequately trained statisticians. This can be remedied.

Just as it is difficult to quantify the benefit of having a workforce with an adequate knowledge of written and spoken language, so it is difficult to quantify the benefit of having a workforce who can adequately communicate about variability and uncertainty.

Commentary by John Henstridge

Terry has illustrated the way in which statistics pervades almost any area where science, industry and government work to improve Australia. I must agree with this — even though I have spent decades working as a statistician, I remain surprised at just how useful this set of tools is and how widely it can be applied. It really is a mathematical Swiss Army penknife.

Terry presents good reasons to be optimistic about statistics and what it can contribute to our aspirations for Australia. With such a track record, it is difficult to imagine anything else. But statistics is facing challenges in Australia.

With notable exceptions, statistics, like all the mathematical sciences, is in decline in Australian universities. This is partly due to declining standards of mathematics in secondary schools. Statistics is also challenged by a lack of clear recognition, in part due to the all-pervasive nature of its application, as Terry highlights.

People in many areas do not realise that statisticians have developed tools that they need, and often view statistics through the somewhat dated way it used to be taught. Statisticians share some of the responsibility for what is frequently poor teaching of their subject.

Despite these concerns, I have to be optimistic like Terry. One of the reasons is that statisticians form a passionate community who want to understand our world through data. That is exactly what Australia needs to make the right decisions.

Commentary by David Warton

At university I planned to specialise in ecology, but turned to a subject I found more interesting — statistics! An ecologist might spend his or her whole career studying only plants, seals, invertebrates, frogs, birds or microbes, but I have studied all of these and more (although mostly from a computer screen).

Statistical research has a key role in helping ecologists find structure in the face of so much uncertainty. It has a similar role as an enabling science in all disciplines in focus in this book.

While statistics is a rewarding career choice, a challenge is to ensure that sufficient students make this choice — so that the discipline continues to have the capacity and expertise to meet the demands of science and industry, and challenges of the digital age, such as big data and the study of networks.

But recent budget decisions are an immediate threat, with risks that the path to a postgraduate degree in statistics will become financially out of reach for many, and with cuts to graduate destinations such as CSIRO and the ABS.

A second challenge is improving standards of statistical literacy among scientists, journalists, doctors, patients and the community at large, so all can make informed decisions in the face of uncertainty.

The increased role of statistics in the national mathematics curriculum can help improve statistical literacy, provided that teachers have sufficient resources and support to implement curriculum changes effectively.

Universities can also help by ensuring full-time staffing of statistical consulting services, invaluable resources for their research communities.