

Chapter 1
What is Statistics?
An Intro for the Uninitiated!



I have yet to see any problem, however complicated, which, when you looked at it in the right way, did not become still more complicated.

Poul (William) Anderson

Quoted in William Thorpe, 'Reduction v. Organicism,' *New Scientist*, 25 Sep 1969, **43**, No 66, 638. In Carl C. Gaither, *Statistically Speaking: A Dictionary of Quotations* (1996), 187

Why Statistics?

Up to a few years ago, access to data was something nearing the impossible. Official statistics was available in state and national libraries while research outputs were only available in hard-to-find expensive books, mostly out of reach from most young researchers' pockets.

Three relatively recent upheavals changed it all!

- I) The advent of Personal Computers in the 1980s
- II) The launching of the internet and the worldwide web in the 1990s
- III) The availability of raw real-time data in the 2000s.

From an era of hoarding and dearth of data we now live in a reality where the problem lies with unravelling the most-relevant dataset required for one's study.

Computers are here to stay. They will become smaller and integrated within all kinds of tools and appliances. Possibly, they will become almost invisible and unobtrusive. But in the past (say, for example 25 years ago) life was very different! There was even a time when people feared computers and predicted that these machines of artificial intelligence would replace human employees. Of course, this would have brought about a disastrously, soaring unemployment rate. Thankfully, these technophobic prophets were wrong. Their apocalyptic prophecy proved to be wrong. In fact, not only did not computers replace human workers but the advent of computers even created more jobs as, for example, in sectors such as gaming, social networking, research and academic work.

People today are comfortable with a visualisation module that allows them to interact with their digital tools. We now live in an era of digital networking, virtual lives and on-demand access to information. Away from a text-based operating system, we started getting WYSIWYG¹ awe-inspiring screens. Fast-forward a few years and one is frowned upon if terminologies are questioned: a mouse is not a cheese-guzzler, a monitor is not a warden, a keyboard is not a musical instrument! This reality was brought shockingly to the attention of the authors very recently when a 65-year old described a computer to them as: 'a television with a piano and a box'. Some of us never realise that it is not a box anymore, especially when one can run complex software on a 10cm Smartphone which carries a spatial locator (known as GPS for surveying), a camera for target recording, a microphone for interviewing, all the spatial and statistical tools possible and of course a phone.... pity mobile telephones have yet to start brewing coffees and baking cookies! Thus processing power and miniaturisation process initiated the radical change necessary to ensure that research be made more accessible.

The next step was even more interesting and had an impact on the social, economic and physical lives we live in: the internet! The internet changed it all! Actually the World-Wide-Web and a genius by the name of Tim Berners Lee did! Basing himself on technologies bringing together interconnected computers as far back as the 1960s within military circles, Mr. Lee created something that was not forbidding to the non-tech as well as easy to understand: an interactive and linked interface. Big words but a simple concept lies behind all this; a network of hotlink, which activate once one clicks on them to open new information nodes.

No longer restricted by mere text, users can now create their own online surveys, run online searches, carry out statistical analysis, employ graphing tools and eventually draft reports using such interfaces. The 1990s have yet to tell us their story on how civilisation was changed in a few years. From the fear of immersion as promoted in the 1990s film 'The Lawnmower Man' to today's real-time Social Networks such as Facebook² and Second Life³, the 'new' online world is beckoning researchers to perpetually take the next step.

¹ WYSIWYG – What You See Is What You Get

² <http://www.facebook.com>

³ <http://secondlife.com/>

The third pivot was brought about by such legislative tools as the Data Protection Act⁴, the Aarhus Convention⁵ and the Freedom of Information Act⁶, which have enabled both the solidification of ethical and regulatory issues as well as provided access to information way beyond the dreams of researchers in the 1990s.

This said, the main problem today is one of access to information. The issue is not necessarily the result of sourcing of data, since that has been solved by both the transmission mode (internet) and access issues (legislation), but more of understanding what the data being used stands for: what does its metadata hold, where does the data reside, what is its currency, can it be compared to other data? There is much one needs to know before delving into the usage of said data or information. The problem encountered today is more one of too much data as against a situation where little data was available. Today, it is more a question of how to interpret such data than where to access it from.

The Tower of Babel Syndrome or Valhalla?

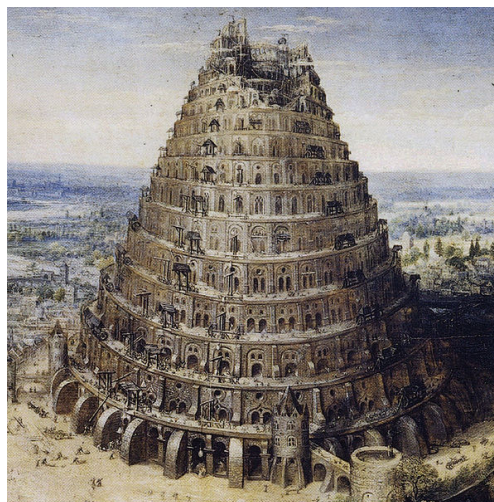
Should researchers today face the reality of having to deal with vast amounts of data? This is a reality that cannot be ignored in information processing. Society is today facing a situation unprecedented in history, where data and information is readily available, it is easy to decipher and is accessible to all...

Or is it?

Are we going down the Babel way?

Technologies used by researchers today face the same problems as the ancient architects who got together and created the brightest inventions and plans for the tallest tower. There is a problem that data may not be readable, not comparable, not of a reliable format, not current and/or does not follow standard regulation. If the data is not decipherable, not comparable, not of a consistent format, not up-to-date and does not adhere to standard research rules, then research would crash as its output would not be credible.

The tower of Babel was not concluded since at the end there were too many languages and they could not communicate between themselves, which situation eventually killed the project. Data faces a similar dilemma... too much data, easily-abused choices of statistical measures, and over-reliance on online data and technologies can all lead to non-credible outputs.



Tower of Babel by Lucas van Valckenborch in 1594⁷
Source: http://commons.wikimedia.org/wiki/Main_Page

⁴ <http://idpc.gov.mt/> (The Data Protection Act of 2001 was enacted in Malta on the 14th December 2001)

⁵ <http://ec.europa.eu/environment/aarhus/>

⁶ http://europa.eu/legislation_summaries/environment/general_provisions/128091_en.htm

⁷ http://en.wikipedia.org/wiki/File: Tower_of_Babel_cropped_square.jpg

The ‘fear of stats’

The study called statistics calls for red-eyed rimmed nights for many a student who has few notions that numbers can be fun! Having seen troops of students calling on “San Pawl tal-iStatistika ghini” (Saint Paul of Statistics – if ever there was one – help me) to “Madonna tal-hlas ehlisti” (Our Lady of Parturition grant me release) just before an open book exam always mystified us. The *fear of stats* was there 20 years ago in our student days and still seems to be one of the major phobias afflicting current students reading for the social sciences. The same idea that students can run statistical analysis has yet to filter through in some disciplines, particularly due to the entrenched schools of thought leaning towards the qualitative, which unfortunately has not helped students to realise that the difference between the two is steadily growing leaner, particularly since the import of new technologies made available tools that help students understand qualitative aspects through quantitative ones, giving a better understanding of the movement towards triangulation studies, where both methods are employed, aimed at enhancing the final study result. One could argue that the fear of stats is unfounded since the main target should not be centred on the numeric and statistical measures (forming the dreaded equations!) but on the process of getting there.

The scope of this publication is not to review those same equations (though a few have been inserted for ease of reference) – there are numerous books to this effect. Some research methods books are even area-specific. For example, one finds ‘statistics for sociologists or psychologists’ with others taking the further step of concentrating on the topic and the particular software for that same theme. Echoing the notorious ancient Chinese curse: researchers today live in highly interesting times! This publication strives to take an abstract view of the whole research process as well as to aid students in their research journey.

Myths and Realities

There are:

LIES DAMNED LIES and STATISTICS⁸

.... so the adage goes and one need only take a look at the phenomenon of media and bloggers’ interpretation of what numbers represent. Statistics has fallen in the same dilemma as that perhaps faced by a person holding a glass of wine half full: “What makes a glass half full or half empty?” Statistics has been driven into a situation where numbers have been given a life of their own and have ended up representing themselves rather than the thematic target. Society has become inundated with huge volumes of raw numbers, data and information snippets. Overload is surely due! And in turn, this situation has led to the use and abuse of statistics.

Mythology has become the fulcrum for statistical review with numbers isolated from the reality within which they reside, percentages are drafted without any reference to the absolute numbers, surveys are misquoted, in cases emphasising the importance of a minority over the majority (of a 10% having more weight than a 90%), of a sample stating that 50 questionnaires represent the whole nation or that it is easy to run a survey, often starting from the end rather than the beginning (actually drafting a questionnaire before conceptualising the whole study)...

The realities are highly different: research methodology follows strict rules and that is why it is termed a science. Banking on such steps across a wide range of disciplines, research methodology has become the foundation of solid physical and social dimensions. From Durkheim’s Rules of Sociological Method to Codd’s database rules, from simple summing equations to complex modeling structures, from base data to action processes; such has been the journey travelled.

⁸ attributed by Mark Twain to the British Prime Minister Benjamin Disraeli (1804–1881)