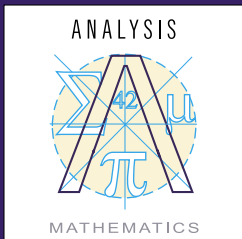


"The past six months has seen major growth at Data Analysis Australia, with talented new staff and exciting new projects. What remains constant is our attention to the highest professional standards. Our first article presents one aspect of this."

Dr John Henstridge  
Managing Director



## When Can You Ignore It?

Everyone has to face information that they would rather not see sometimes. It might be a speeding ticket or their weight. It might be the result of a survey that gave results that they are not comfortable with. One course of action is to ignore it but that is not always the right or the best action. Sometimes the unpleasant information is a critical warning that must be heeded.

More importantly, ignoring facts can damage credibility. In Court, the requirement is to give "the truth, the whole truth and nothing but the truth". This is also the rule in science where all the results of an experiment must be presented.

When analysing data, these problems arise in the form of values that were not expected or do not fit well with your model of how things should be. The statistician calls these outliers. Can those data values be ignored as "obviously in error"? Or are they actually warnings that contain important information?

The Depression of 1929-30 is an example of economic time series data. It did not fit the pattern of any previous economic data and clearly conflicts with many of the mathematical models currently used by econometricians. But the Depression was real and perhaps taught us more about the world economy than any other event of the last century.

There are situations where some data must be ignored, or at least downplayed. For example, when statistics on house prices by suburb are given, it is usual to give the median price, not the average. This recognises that averages can be seriously affected by one or two extreme values and hence not give an accurate indication of the typical price. The median, which is simply a value that has half the data on each side of it, uses the information that these high prices exist but ignores their actual values.

Every time information is ignored, it is valid to ask whether it was appropriate to do so. Simply that it was inconvenient is not a good excuse and detracts from the credibility of the analysis. Not surprisingly, it can raise questions of objectivity and integrity. There is no substitute for being methodical and documenting why decisions are made the way they are. If it is believed that the data is in error, it is good to have some idea of how the errors may have been made.

Again, the Depression is an example of this. Every time an economic forecaster ignores it - saying in effect "it will never happen again" - they are understating the risks or uncertainties in their forecasts.

These are issues that statisticians grapple with every day and tools have been developed to

help. For a start, statisticians have clear ideas of random variation and thus ways of deciding whether a data value is reasonable or unreasonable in terms of the context. Some of these methods have become highly systematic, including ones that build a natural caution into the analysis while minimising how much information is lost - this is the area of robust statistics. The median is just one example of a robust statistic.

Another strength of statisticians is in recognising the bias that might be introduced by leaving out data. For example, the distribution of the original data might be skewed and removing extreme data points can lead to the distribution of the refined data set to be centred.

However, these methods do not remove the responsibility to ensure that it is appropriate to ignore some information. If anything, the use of robust methods increases the risk that poor decisions are made since they are automatic. In the end it is the professional responsibility of the statistician to say what data must not be ignored, even when such a statement is unpopular.



Two aspects of the Depression - the financial cost (Dow Jones Index) and the human cost.

## Company News

Data Analysis Australia in partnership with CATALYSE® was recently appointed to the Water Corporation's Panel for the provision of market research consultancy services. We have worked with the Water Corporation on more than 40 projects since 1991 and look forward to many more over the term of this panel and beyond.

Data Analysis Australia will be supporting the next Australian Statistical Conference to be held in Auckland New Zealand in July 2006 by providing travel sponsorship in addition to sending five delegates. **Anna Munday** and **April Rutkay** have taken up committee positions for 2006 for the Western Australian Branch of the Statistical Society of Australia.

**Greg Wogan-Browne** and **James Andrewartha** attended the annual Linux conference in New Zealand in January. They are both looking forward to attending next year's Linux conference in Sydney.



On 16th November, **Cheryl Praeger**, a Director of Data Analysis Australia, received the award of Doctor Honoris Causa from the Faculty of Science of the Université Libre de Bruxelles (Free University of Brussels), in recognition of her international contributions to mathematics research.

We welcome a number of arrivals. Consultant Statisticians **Rhiannon Marchant** and **Aimee**

**Roche** have recently completed Honours in Applied Statistics and Statistics respectively (both from the University of Western Australia). **Miriam Maclean** joins us as our first Consultant Survey Researcher. She has first class Honours in Psychology and received the Australian Psychological Society Prize as the most outstanding psychology graduate from Edith Cowan University for 2005. **Scott Brown** joined us in March as a Consultant Mathematician after submitting his PhD in Pure Mathematics at the University of Western Australia.

97 Broadway  
Nedlands WA 6009  
Telephone: +61 8 9386 3304  
Fax: +61 8 9386 3203  
Web: [www.daa.com.au](http://www.daa.com.au)  
Email: [daa@daa.com.au](mailto:daa@daa.com.au)



Many surveys are one-off surveys that simply aim to measure a single point in time. However, sometimes the aim is to measure and understand change, which leads to far more complex survey design issues that we explore in this issue's Analytical Idea.

This paper can be found at:

[www.daa.com.au/analyticalideas/measuringchange.html](http://www.daa.com.au/analyticalideas/measuringchange.html).

## Staff Profile

Dr Linda Eaton is our "newest" Senior Consultant and completed her PhD in mathematics at Murdoch University. Linda has been employed in various aspects of mathematical research and applications. In particular, Linda taught mathematics and statistics at Murdoch University in Perth and the University of Canterbury in New Zealand.



Linda is the project manager of one of our largest undertakings, a survey on behalf of the Department of the Premier and Cabinet to investigate the retirement intentions of the Western Australian public service. This project involves conducting both an E-Survey and a paper based survey of around 68,000 employees of the state government. Linda is also working on a diverse range of other projects including the statistical analysis of data from the impact study of the desalination plant in Cockburn Sound, an investigation of data relating to honey and honeybees, and on the certification of standards in the gold industry.

Linda enjoys rock climbing, yoga and just playing in nature. She studies Japanese tea ceremony and Sufism and is focused on the union of cultures through mutual understanding and experience.

## Classic Quote

*"Fewer names appear in the top 100 than ten years ago"*

From report on most popular names given to babies born in Scotland in 2005.

**The Scotsman (Scotland's National Newspaper)**  
**- 24 December 2005.**

