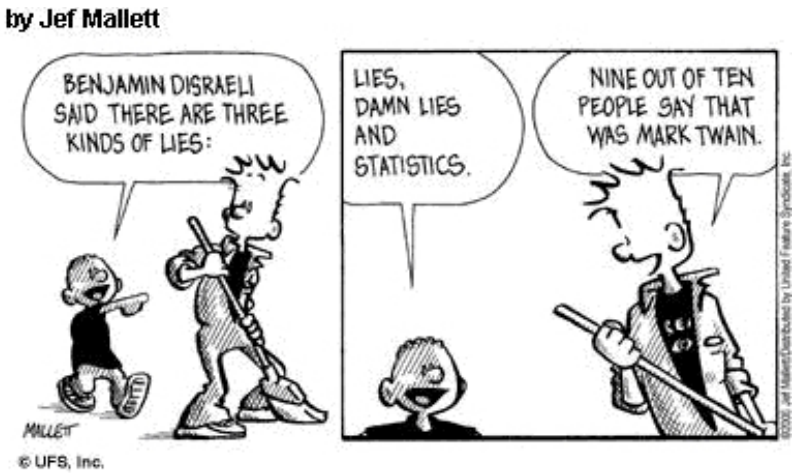
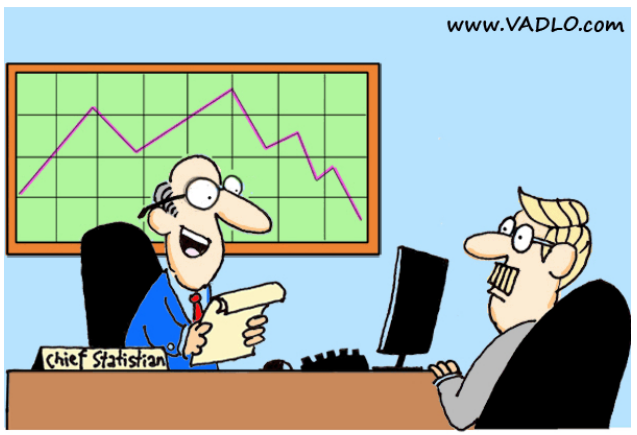
INTRODUCTION: WHY STATS?

Outline

Why Study Statistics?

Sometimes 'Statistics' is treated with scepticism - you may have heard the famous quote: 'There are lies, damn lies, and statistics', attributed to both Mark Twain and Benjamin Disraeli.





"I can prove it or disprove it! What do you want me to do?"

But in a data rich world, statistical literacy is essential for everyone, whether in medicine, science, business, education, engineering, marketing or law, as it allows evidence based reasoning.

For example, has the incidence of brain cancer risen in Australia since the introduction of mobile phones 29 years ago? Mobile phone use in Australia has increased rapidly since its introduction in 1987 with whole population usage being 94% by 2014. There is a popularly hypothesised association between brain cancer incidence and mobile phone use. Is it valid?

Or, does public feeling toward sharks grow negative following shark bites on humans? Media and government responses are often predicated on this presumptive emotional response; but how should public policy be developed?

Statistics has endless interesting applications and is highly collaborative, as John Wilder Tukey said, 'The best thing about being a statistician is that you get to play in everyone's backyard.'

That's why there are so many jobs for 'data scientists'. According to Val Kinsley, Chief economist at Google, 'I keep saying the sexy job in the next ten years will be statisticians. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s?' or a recent Fortune magazine, 'The nerdy-cool job that companies are scrambling to fill: data scientist.'

▸ Cartoon1Source ▸ Cartoon2 Source ▸ Mistaken Data ▸ McKinsey Article ▸ 2016 Top Jobs

▸ Tukey ▸ Data Scientist ▸ Fortune

What is Statistics?

Definition (Statistics)

Statistics is the science of learning from data. It is problem-solving. We pose questions of data and enable data to pose questions.

Definition (Big Data)

Big Data is 'inconveniently large data' (Clair Alston), with characteristics such as high-volume, high-variety, high-velocity, high-variability, high-value and low-veracity.

Definition (Population and Sample)

A population is the full amount of information being studied, of which a sample (or data) is a subset.

We often only have access to a sample of a population. Why?
Hence, the critical question driving much Statistical Research is:
What information does the sample give about the population and how reliable is that information?