


Early beginnings


**450 BC** Hippias of Elis uses the average value of the length of a king's reign (the mean) to work out the date of the first Olympic Games, some 300 years before his time.




**400 BC** In the Indian epic the *Mahabharata*, King Rtuarna estimates the number of fruit and leaves (2095 fruit and 50 000 000 leaves) on two great branches of a vibhitaka tree by counting the number on a single twig, then multiplying by the number of twigs. The estimate is found to be very close to the actual number. This is the first recorded example of sampling – “but this knowledge is kept secret”, says the account.



**431 BC** Attackers besieging Plataea in the Peloponnesian war calculate the height of the wall by counting the number of bricks. The count was repeated several times by different soldiers. The most frequent value (the mode) was taken to be the most likely. Multiplying it by the height of one brick allowed them to calculate the length of the ladders needed to scale the walls.




**AD 2** Chinese census under the Han dynasty finds 57.67 million people in 12.36 million households – the first census from which data survives, and still considered by scholars to have been accurate.




Mathematical foundations


**1560** Gerolamo Cardano calculates probabilities of different dice throws for gamblers.



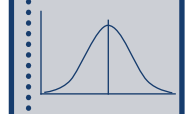
**1654** Pascal and Fermat correspond about dividing stakes in gambling games and together create the mathematical theory of probability.




**1713** Jacob Bernoulli's *Ars conjectandi* derives the law of large numbers – the more often you repeat an experiment, the more accurately you can predict the result.



**1791** First use of the word “statistics” in English, by Sir John Sinclair in his *Statistical Account of Scotland*.



**1886** Philanthropist Charles Booth begins his survey of the London poor, to produce his “poverty map of London”. Areas were coloured black, for the poorest, through to yellow for the upper-middle class and wealthy.

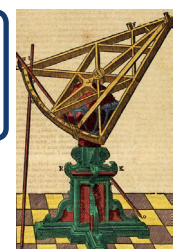


Statistics is about gathering data and working out what the numbers can tell us. From the earliest farmer estimating whether he had enough grain to last the winter to the scientists of the Large Hadron Collider confirming the probable existence of new particles, people have always been making inferences from data. Statistical tools like the mean or average summarise data, and standard deviations measure how much variation there is within a set of numbers. Frequency distributions – the patterns within the numbers or the shapes they make when drawn on a graph – can help predict future events. Knowing how sure or how uncertain your estimates are is a key part of statistics.

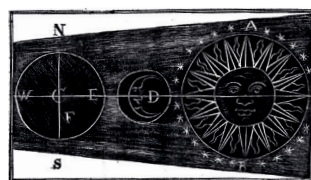
Today vast amounts of digital data are transforming the world and the way we live in it. Statistical methods and theories are used everywhere, from health, science and business to managing traffic and studying sustainability and climate change. No sensible decision is made without analysing the data. The way we handle that data and draw conclusions from it uses methods whose origins and progress are charted here.

Julian Champkin  
*Significance* magazine


**1570** Astronomer Tycho Brahe uses the arithmetic mean to reduce errors in his estimates of the locations of stars and planets.



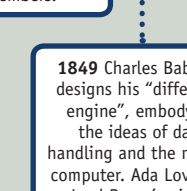
**1644** Michael van Langren draws the first known graph of statistical data that shows the size of possible errors. It is of different estimates of the distance between Toledo and Rome.



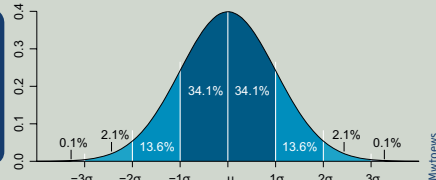
**1790** First US census, taken by men on horseback directed by Thomas Jefferson, counts 3.9 million Americans.



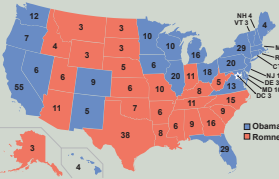
**1833** The British Association for the Advancement of Science sets up a statistics section. Thomas Malthus, who analysed population growth, and Charles Babbage are members. It later becomes the Royal Statistical Society.



**1894** Karl Pearson introduces the term “standard deviation”. If errors are normally distributed, 68% of samples will lie within one standard deviation of the mean. Later he develops chi-squared tests for whether two variables are independent of each other.




**1900** Louis Bachelier shows that fluctuations in stock market prices behave in the same way as the random Brownian motion of molecules – the start of financial mathematics.



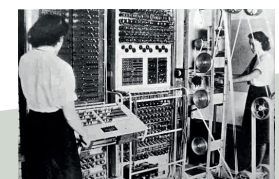
**1908** William Sealy Gosset, chief brewer for Guinness in Dublin, describes the t-test. It uses a small number of samples to ensure that every brew tastes equally good.




**1916** During the First World War car designer Frederick Lancaster develops statistical laws to predict the outcomes of aerial battles: if you double their size land armies are only twice as strong, but air forces are four times as powerful.




**1940-45** Alan Turing at Bletchley Park cracks the German wartime Enigma code, using advanced Bayesian statistics and Colossus, the first programmable electronic computer.



**1948** Claude Shannon introduces information theory and the “bit” – fundamental to the digital age.



**1972** David Cox's proportional hazard model and the concept of partial likelihood.




**2002** Paul DePodesta uses statistics – “sabermetrics” – to transform the fortunes of the Oakland Athletics baseball team; the film *Moneyball* tells the story.

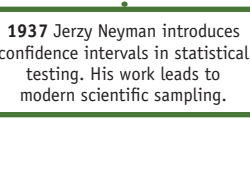


Modern era


**1911** Herman Hollerith, inventor of punch-card devices used to analyse data in US censuses, merges his company to form what will become IBM, pioneers of machines to handle business data and of early computers.




**1935** George Zipf finds that many phenomena – river lengths, city populations – obey a power law so that the largest is twice the size of the second largest, three times the size of the third, and so on.



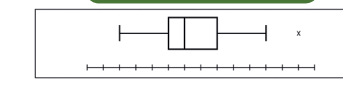
**1944** The German tank problem: the Allies desperately need to know how many Panther tanks they will face in France on D-Day. Statistical analysis of the serial numbers on gearboxes from captured tanks indicates how many of each are being produced. Statisticians predict 270 a month; reports from intelligence sources predict many fewer. The total turned out to be 276. Statistics had outperformed spies.



**1958** The Kaplan-Meier estimator gives doctors a simple statistical way of judging which treatments work best. It has saved millions of lives.



**1977** John Tukey introduces the box-plot or box-and-whisker diagram, which shows the quartiles, medians and spread of data in a single image.



**2012** The Large Hadron Collider confirms existence of a Higgs boson-like particle with probability of five standard deviations – around one chance in 3.5 million that all they are seeing is coincidence.

