Timeline of statistics

ASA AMERICAN STATISTICAL ASSOCIATION

ROYAL STATISTICAL SOCIETY

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

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.


Photo: Matthias Kabel

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.

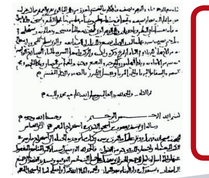
400 BC In the Indian epic the *Mahabharata*, King Rtuparna estimates the number of fruit and leaves (2095 fruit and 50 000 000 leaves) on two great branches of a vibhitaka tree by counting the number on a single twig, then multiplying by the number of twigs. The estimate is found to be very close to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


iStock/Thinkstock

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.

AD 7 Census by Quirinus, governor of the Roman province of Judea, is mentioned in Luke's Gospel as causing Joseph and Mary to travel to Bethlehem to be taxed.


Roman Gudyma/iStock/Thinkstock



840 Islamic mathematician Al-Kindi uses frequency analysis – the most common symbols in a coded message will stand for the most common letters – to break secret codes. Al-Kindi also introduces Arabic numerals to Europe.

10th century The earliest known graph, in a commentary on a book by Cicero, shows the movements of the planets through the zodiac. It is apparently intended for use in monastery schools.

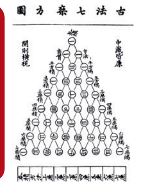
1069 Domesday Book: survey for William the Conqueror of farms, villages and livestock in his new kingdom – the start of official statistics in England.

1150 Trial of the Pyx, an annual test of the purity of coins from the Royal Mint, begins. Coins are drawn at random, in fixed proportions to the number minted. It continues to this day.


Anton Snarikov/iStock/Thinkstock

1188 Gerald of Wales completed the first population census of Wales.

1303 A Chinese diagram entitled "The Old Method Chart of the Seven Multiplying Squares" shows the binomial coefficients up to the eighth power – the numbers that are fundamental to the mathematics of probability, and that appeared five hundred years later in the west as Pascal's triangle.



1346 Giovanni Villani's *Nuova Cronica* gives statistical information on the population and trade of Florence.


byggarn79/iStock/Thinkstock

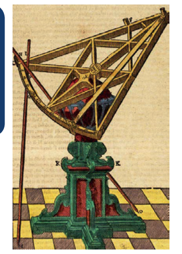
500	400	300	200	100	0	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	

Mathematical foundations

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


iStock/Thinkstock

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

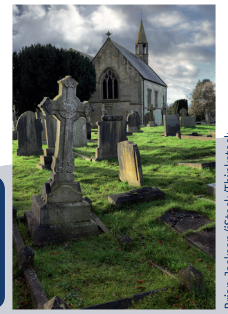


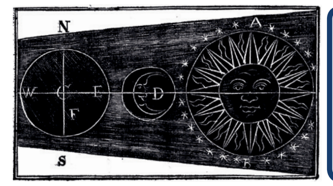
1644 Michael van Langren draws the first known graph of statistical data that shows the size of possible errors. It is of different estimates of the distance between Toledo and Rome.

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

1657 Huygens's *On Reasoning in Games of Chance* is the first book on probability theory. He also invented the pendulum clock.

1663 John Graunt uses parish records to estimate the population of London.


Brian Jackson/iStock/Thinkstock

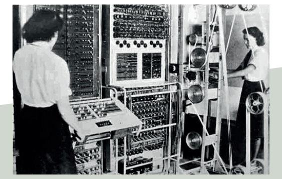


1693 Edmund Halley prepares the first mortality tables statistically relating death rates to age – the foundation of life insurance. He also drew a stylised map of the path of a solar eclipse over England – one of the first data visualisation maps.

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.

1728 Voltaire and his mathematician friend de la Condamine spot that a Paris bond lottery is offering more in prize money than the total cost of the tickets; they corner the market and win themselves a fortune.

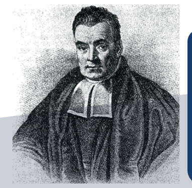
1749 Gottfried Achenwall coins the word "statistics" (in German, *Statistik*); he means the information you need to run a nation state.



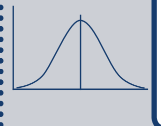
1757 Casanova becomes a trustee of, and may have had a hand in devising, the French national lottery.

1761 The Rev. Thomas Bayes proves Bayes' theorem – the cornerstone of conditional probability and the testing of beliefs and hypotheses.

1786 William Playfair introduces graphs and bar charts to show economic data.



1789 Gilbert White and other clergymen-naturalists keep records of temperatures, dates of first snowdrops and cuckoos, etc; the data is later useful for study of climate change.



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

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

1805 Adrien-Marie Legendre introduces the method of least squares for fitting a curve to a given set of observations.

1808 Gauss, with contributions from Laplace, derives the normal distribution – the bell-shaped curve fundamental to the study of variation and error.

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.

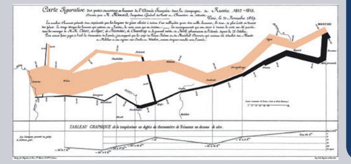
1835 Belgian Adolphe Quetelet's *Treatise on Man* introduces social science statistics and the concept of the "average man" – his height, body mass index, and earnings.

1839: The American Statistical Association is formed. Alexander Graham Bell, Andrew Carnegie and President Martin Van Buren will become members.

1840 William Farr sets up the official system for recording causes of death in England and Wales. This allows epidemics to be tracked and diseases compared – the start of medical statistics.

1849 Charles Babbage designs his "difference engine", embodying the ideas of data handling and the modern computer. Ada Lovelace, Lord Byron's niece, writes the world's first computer program for it.

1854 John Snow's "cholera map" pins down the source of an outbreak as a water pump in Broad Street, London, beginning the modern study of epidemics.



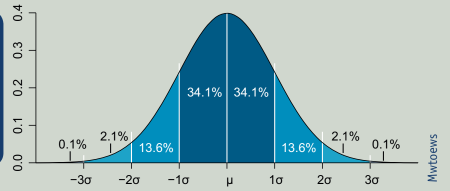
1859 Florence Nightingale uses statistics of Crimean War casualties to influence public opinion and the War Office. She shows casualties month by month on a circular chart she devises, the "Nightingale rose", forerunner of the pie chart. She is the first woman member of the Royal Statistical Society and the first overseas member of the American Statistical Association.

1868 Minard's graphic diagram of Napoleon's March on Moscow shows on one diagram the distance covered, the number of men still alive at each kilometre of the march, and the temperatures they encountered on the way.

1877 Francis Galton, Darwin's cousin, describes regression to the mean. In 1888 he introduces the concept of correlation. At a "Guess the weight of an Ox" contest in Devon he describes the "Wisdom of Crowds" – that the average of many uninformed guesses is close to the correct value.

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.

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.



Mwtoews

1898 Von Bortkiewicz's data on deaths of soldiers in the Prussian army from horse kicks shows that apparently rare events follow a predictable pattern, the Poisson distribution.

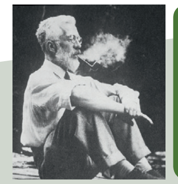
1560	1580	1600	1620	1640	1660	1680	1700	1720	1740	1760	1780	1800	1820	1840	1860	1880	1900	1920	1940	1960	1980	2000	

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.

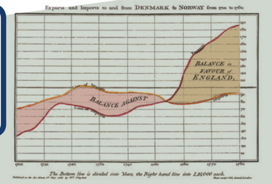

Monkey Business/Thinkstock

1916 During the First World War car designer Frederick Lanchester 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.

1924 Walter Shewhart invents the control chart to aid industrial production and management



1935 R. A. Fisher revolutionises modern statistics. His *Design of Experiments* gives ways of deciding which results of scientific experiments are significant and which are not.



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

1946 Cox's theorem derives the axioms of probability from simple logical assumptions.

1948-53 The Kinsey Report gathers objective data on human sexual behaviour. A large-scale survey of 5000 men and, later, 5000 women, it causes outrage.

1950s Genichi Taguchi's statistical methods to improve the quality of automobile and electronics components revolutionise Japanese industry, which far overtakes western European rivals.

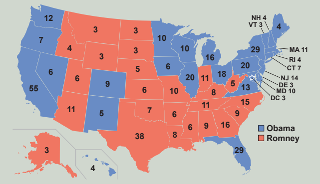

Rainer Plendl/iStock/Thinkstock

1979 Bradley Efron introduces bootstrapping, a simple way to estimate the distribution of almost any sample of data.

1997 The term "Big Data" first appears in print.


Peter Kim/iStock/Thinkstock

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



2012 Nate Silver, statistician, successfully predicts the result in all 50 states in the US Presidential election. He becomes a media star.

1900	1904	1908	1912	1916	1920	1924	1928	1932	1936	1940	1944	1948	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992	1996	2000	2004	2008	2012	

Modern era


Photo: Sami Keinänen

1908 William Sealy Gossett, 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.

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.

1937 Jerzy Neyman introduces confidence intervals in statistical testing. His work leads to modern scientific sampling.

Bundesarchiv, Bild 101I-783-0110-12/ Dörner/CC-BY-SA

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.

1948 Claude Shannon introduces information theory and the "bit" – fundamental to the digital age.

1950 Richard Doll and Bradford Hill establish the link between cigarette smoking and lung cancer. Despite fierce opposition the result is conclusively proved, to huge public health benefit.

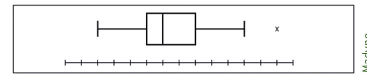

Les Cunliffe/iStock/Thinkstock

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


Catherine Yeulet/iStock/Thinkstock

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

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.


Madyno

1982 Edward Tufte self-publishes *The Visual Display of Quantitative Information*, setting new standards for graphic visualisation of data.

1988 Margaret Thatcher becomes the first world leader to call for action on climate change.


Alejandro Catalá Rubio/iStock/Thinkstock

1993 The statistical programming language "R" is released, now a standard statistical tool.

2002 The amount of information stored digitally surpasses non-digital.

2004 Launch of *Significance* magazine.

2008 Hal Varian, chief economist at Google, says that statistics will be "the sexy profession of the next ten years".

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.