

Statisticians have a Word for it

Derek Bissell
Abergavenny, Wales.

◆INTRODUCTION ◆

STATISTICAL terminology contains many words with interesting histories. Some have been hi-jacked from other disciplines, others have undergone strange twists in meaning. Personal names appear in eponyms, combinations of initial letters in acronyms. The origins of many words can be traced to the fore-runners of Latin, Greek and Sanskrit languages of 6000 years ago, but others are of recent invention, often using classical elements as abase. The article reviews these mechanisms with examples of each.

As an example of the antiquity of some word-ideas', consider the basic notion of **counting**, so essential to statistics. It is readily traceable to the Old French 'conter', meaning to add up or reckon. This in turn developed from Latin 'computare', to calculate, an expansion of putare' which could have either the physical meaning 'to prune' (eg a tree) or to consider. Even this can be traced to an Indo-European proto-language root *peue, to cut or dig into (the * indicates a hypothetical reconstruction).

Similarly **number** is traceable via Old French 'nombre' and Germanic 'nummer' to Latin 'numerus', Greek 'nemein' and Indo-European *em, to divide or allocate. The Indo-Europeans were among the earliest farmers and traders, and so needed basic numeracy for controlling these activities. Even the number-systems of most European languages can be traced to the Indo-European oines, duwo, trejes, kuetur, pengke, seks, septm, okto, newn, dekm, despite various consonant shifts which account for the development of four from kuetur, five from pengke, eight from okto. etc.

We shall only trace a few basic terms back to this early language, but even where Greek, Latin, French are cited as origins, it is worth remembering that these in turn developed from earlier sources.

◆MEASUREMENT ◆ AND AVERAGES

Statistics is concerned with measurement as well as counting, and a remarkable number of scientific measurement terms descend from the Indo-European root *me, to measure (especially time) and *meu, to move (in the sense of to push). *Mei and *mel are further members of this set with modern descendants like remuneration, migrate, ameliorate and multitude.

Table 1 shows how the Indo-European roots developed into the Greek, Latin and Germanic forbears of scientific, mathematical and statistical terms. 'Moment', of course, needs a final sidestep from the 'leverage' sense of mechanics to its application to statistical distributions - verbal hi-jack.

The original meaning of **mean** (sorry about the pun!) is interesting, as it was formerly, in modern notation, $(x_{\max} + x_{\min})/2$ This would now be called the mid-point or mid-range. 'Mean' could also indicate the central terms b and c in the ratio

$a:b::c:d$, or $a/b = c/d$ whence $ad = bc$.

What is now the arithmetic mean was the 19th century average, which has a fascinating history of its own. Starting perhaps 6000 years ago with the element *op = work, which made its way into Latin 'habere' = to have, we find medieval French has the word 'aver' = property. Meanwhile the Arabic 'arawiya', which meant goods damaged by sea water (!) reached Middle English in the 15th century as 'averays', partly influenced by the French 'aver'; at this time, averays could denote either lost cargo or a form of feudal service. The 'lost cargo' sense then develops thus:

duty charged for surety → financial loss from damage
C.XVI

→ equitably shared loss C.XVII

→ the 'fair share' or arithmetic mean $(\sum x/n)$ C.XIX

This also clarifies the activities of the 'average adjuster', not as one who massages the statistics but one who estimates reasonable compensation.

Subsequently, 'average'; has gone rather down-market, firstly by a lack of precision as to which measure

(mode, mean, median, mid-value, etc) it actually refers to, then as a term meaning ‘vaguely typical’, and eventually to ‘ordinary’ or even ‘mediocre’ (a recent assessment form that came my way had boxes to be ticked for quality of service: excellent, very good, good, average, poor - average is no longer in the middle, but towards the lower end).

◆SOURCES OF◆ STATISTICAL TERMS

Some well-known ‘stat-words’ are eponymous - they carry the name of someone who first thought of the concept. One of the earliest was the ninth century Arabic mathematician Abu Ja’far Muhammad ibn Musa Al **Khwarisma**. The last part of his name indicates he was a man of Chorasmia, then in Persia but now (as Khiva) in Uzbekistan, part of the former USSR; it gives us **algorithm**. The eighteenth century yielded **Bayes’ Theorem**, first propounded by Rev Thomas Bayes, and Monsieur S D Poisson developed Poisson’s Exponential Limit to the Binomial distribution (to give the **Poisson** Distribution its full and original name). Later came Brownian motion (Robert **Brown**), Pareto Analysis (Wilfredo **Pareto**) and Lorenz curves (M **Lorenz**), and the more obvious twentieth century contributions from R A **Fisher** (F-distribution, although it was finalised by W Snedecor and the ‘F’ used in Fisher’s honour), K **Pearson** (Pearson curves) and M **Bartlett** (Bartlett’s test). **Student’s t** and **studentize** are from the pen-name of W S Gossett, who was at first not permitted by his employer to publish under his own name.

Useful words are often hi-jacked from other disciplines. The **ogive**, or cumulative distribution curve, takes its name from the half-rib shape of a diagonal Gothic arch; ogee-curve is of similar origin - it is not an expression of surprise at the shape of the curve! The graph **grid** comes from the square lattice of the griddle-iron, and **outlier** is stolen from geology - the occurrence of aberrant or alien rocks. (The alternative name maverick for an observation that does not fit the general pattern is attributable to Samuel Maverick, a Texan lawyer and rancher who did not always brand his steers - the unidentifiable beasts became a nuisance among other herds). As a final example of hi-jack, we note J W Tukey’s adoption of the Scout’s jack-knife (or jackknife or jacknife - all three spellings occur in the literature) his useful general-purpose re-sampling technique parameter estimation.

Neo-classical terms are modem words formed from classical roots. Statistical examples include **histogram** literally a ‘standing up graph’, **quota**,

simply the Latin word for ‘how many?’ and **kurtosis**, from the Greek word for ‘arched’, and hence measuring the humpiness (or lack of it) in a distribution curve.

Acronyms use just initial letters, or occasionally first few, of several words, The **BASIC** computer code is variously ascribed to British American Scientific International Commercial or Beginner’s All-purpose Symbolic Instruction Code. Most other statistical acronyms are associated with more advanced methods but ANOVA (Analysis of Variance) and GLIM (General Linear Model) are well-known.

◆SHIFTS AND TWISTS◆

Words in all aspects of life gradually change in form and sense, and statistical jargon (**jargon** itself is from the Old French ‘jargonne’ = twittering of birds, hence unintelligible talk) is no exception. We have already seen how ‘average’ developed over several centuries. Occasionally words seem to develop beyond a logical progression of ideas and make sudden leaps into new meanings.

Among statistical / mathematical terms, **calculation** no longer depends on counting pebbles (Latin ‘calculus’) although cricket umpires are known to use coins or marbles for counting overs. **Analysis** literally means to unlock or unloose (from Greek ‘analysis’). The underlying idea of **random** is to run impetuously, or to gallop, so that firing ‘at random’ conjures up images of arrows shot without real purpose from horseback! The French come nearer to statistical principles with **au hazard**, implying the use of gaming dice. The random element appears again in stochastic; the Greek ‘stoxos’ was a target, and ‘stokhazestai’ was literally ‘to aim at’. In this sense, stochastic was in use in the 17th and 18th centuries (it is still listed, as ‘rare and obsolete’, in the Oxford English Dictionary). It died out, and was revived in the 20th century as ‘randomly determined, not predicted precisely’; it has been further hi-jacked into the terminology of modern music.

Robust methods no longer imply the use of wood though the original sense was strong and hardy, like an oak tree (Latin ‘robus’). You may have wondered why, to quote the adage, ‘the exception proves the rule’. The early meaning of ‘prove’ was to test - and certainly an exception severely tests rule (the ‘proof of the pudding is in the eating’ similarly implies testing). What has this to do with statistics? Our **probability** is, strangely, descended from ‘provable’ in the sense of testing, and something ‘probable’ could be tested and, hopefully, shown to be true. The statistical sense now places much more emphasis on “uncertainly” than on proof!

To ‘esteem’ implies, of course, to value. A natural step would be to assess just how valuable the esteemed object

was leading to the verb **estimate**, later picked up by various technologies including surveying, geology and statistics, while ‘an estimate’ uses the verb as a noun. **Samples** are used for estimation - they literally provide an ‘example’ of the wider population. The root can be traced to Indo-European *em = to take, thence via Latin ‘ex-emere’ (to take out = to buy) which became ‘eximere’ and passed into Old French as ‘essample’ - a case of try it before you buy it!

Some of the more spectacular twists of meaning include empirical, hierarchy and optimise, all of which have strong statistical connections. **Empirical** is related to a Greek school of physicians who learned by experience (the dogmatists worked to a set of rules), **Hierarchic** designs in experimentation have no religious connection, though the base is Greek hierarkhes = chief priest; hierarchy later referred to divisions of angles before broadening to mean any system of ordered levels. **Optimise** was coined in 1737 to denote the doctrine of Leibnitz that ‘this is the best of all possible worlds that could have been created’. It was later associated with various shades of hoping, or looking, for the best.

The final place is reserved for the strange statistical use of **regression**. In its other applications - geology, psychology, astronomy, etc, it retains its basic meaning of ‘stepping back’ or ‘reversion to an earlier state’. Indeed, it was first applied statistically (or more strictly, genetically) in this sense; it was noted that the offspring of extreme individuals in a population tend to have characteristics nearer to the overall mean than those of their parents. Thus over several generations there is a regression to the average’. It is not known who first applied (or mis-applied) the term to estimation of parameters of a model (no longer solely genetic models) generally by the use of least-squares techniques.

detailed account, and a comprehensive monograph is in course of preparation, covering mathematical as well as statistical terms. For the enthusiastic reader, a browse through a good dictionary (some are listed in the Bibliography) or one of the books on word origins will reveal other curious and interesting etymologies.

Acknowledgement

I wish to thank the Editor and a referee for useful suggestions which assisted in preparing this article, and the authors of the books listed below for their information and inspiration.

Reference and Bibliography

- Bissell A.F. (1990). The eccentrics of political twittering. *Journal of Applied Statistics*. **17**, 211-217
- Ayto, J. *Dictionary of Word Origins*. Bloomsbury, London.
- Hoad, T. *The Oxford Library of Words and Phrases*, University Press, Oxford.
- Heller, L, Humes, A. and Dror, M. *Th₁ Private Lives of English Words*. Routledge and Kegan Paul, London.
- Heller, A., Weller, N. and Maguire, J. *Zero to Lazy Eight*. Simon Schuster, New York.
- Marriott, F.H.C. *A Dictionary of Statistical Terms*. Longman. (Note: previous editions were edited by Kendall, M.G. and Buckland, W.R.).
- Partridge, E. *Origins*. Routledge, London. Shipley, J.T. *The Origins of English Words: a discursive dictionary of Indo-European roots*. John Hopkins University Press, Baltimore.
- Standard Dictionaries: Chambers, Collins, Oxford, Random House, Webster’s, etc.

◆CONCLUSION◆

This article has merely scratched the surface of a fascinating subject. Bissell (1990) gives a more

Table 1. Measures, mids and means

Proto-Indo-European	*me	*med	*medhi	*menu
Greek	mene (moon)		mesos (middle)	
Latin	mensus (month)	mod- (measure)	medianu (middle)	movere (move)
Germanic	meal (time to eat)			
English derivatives	measure mensurate menstruate mete meet (=apt) metre, meter	modicum modulus modern meditate medic- mediocre MODE	mesne mitten moiety mid- meso- MEAN MEDIAN	motor mobile mutiny promote remote Momentum MOMENT