Practice and Evidence of Scholarship of Teaching and Learning in Higher Education
Special Issue: Threshold Concepts and Conceptual Difficulty
Vol. 12, No.2 April 2017, pp. 430-443

Identifying threshold concepts in postgraduate statistical teaching to non-statisticians

Andrew K Wills
Faculty of Health Sciences,
University of Bristol. UK
a.k.wills@bristol.ac.uk

Abstract

In this enquiry I discuss and investigate potential threshold concepts in the postgraduate statistics curriculum for medics and scientists (non-statisticians). In doing so, I also reflect on the elements of statistics these students need to learn and how they learn. Threshold concepts are a useful framework for thinking about the content of a curriculum. Having made large changes to the Units that I teach, I wanted to assess whether these changes were justified and whether there was rationale for further modifications, for example; emphasising particular elements of the curriculum and/or deemphasising or dropping others. I used a questionnaire, administered to staff and students to try to identify potential threshold concepts. My hypothesis was that sampling variation (i.e. random/ sampling error, uncertainty) as coded by the sampling distribution is a threshold concept. The results, while preliminary, support this hypothesis and stimulated deeper thinking about the pedagogy of my teaching to these audiences.

Keywords: Threshold concepts, introductory statistics, teaching

Background

Statistics teaching on postgraduate science and medicine programs

All postgraduate taught programmes in science and medicine include at least one unit in statistics. Statistics is a natural and necessary compliment to both science- science being the process by which information about the world is gained, and Western medicine - with its emphasis on using the best available evidence to inform practice.

Despite this, students are often surprised and anxious by the amount of statistics included in such programmes.

Statistical literacy (& illiteracy)

Statistical literacy is often a tacit learning objective of postgraduate courses in statistics. It can be defined as the ability to interpret and critically evaluate statistical information and stochastic phenomena which one may encounter in diverse contexts, and an ability to discuss such information with respect to its implications and any concern for the acceptability of conclusions that are drawn (Gal, 2002). The importance of statistical literacy was encapsulated over a century ago in a piece written by HG Wells; it has since been distilled into the following quote'

> "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read or write" HG Wells (Tankard, 1979)

Over the years, this sentiment has been widely echoed by political, educational and health leaders (Gal, 2002). It seems particularly prescient given today's data-centric world where citizens are provided with ever more information and access to information, and where there is a culture of individual choice and decision making. With respect to science, it is well acknowledged that a misunderstanding of basic statistical ideas (or lack of statistical literacy) is a contributory cause of poor science, wasted research effort and misguided translation of research findings (eg, Altman, 1994; Chalmers & Glasziou, 2009). Several articles have described the statistical principles that underpin some of these misunderstandings, the most highly cited is that by Ioannidis (2005) which explains why most published findings are false. A good understanding of fundamental statistical concepts is one way of tackling this endemic problem.

There is a clear need for statistical literacy in both the immediate and wider context in which higher education operates. Given the statistical content of MSc programmes, one might expect such programmes to support this through the supply of statistically literate postgraduates. Unfortunately, there is plenty of evidence to suggest otherwise. For example, poor levels of statistical literacy and understanding have even been found among highly educated senior medical doctors (Anderson et al., 2013; Girgerenzer, 2003; Girgerenzer et al., 2008). And anecdotally, many teachers bemoan their student's statistical misperceptions and misunderstandings and failure to grasp key concepts (Dunne, 2003; Garfield, 1995; Nikiforidou, 2010). Regardless of when and where the roots of these misunderstandings lie, statistical teachers in higher education have an opportunity to correct this, or at the very least, offer a curriculum that promotes statistical literacy.

Are we teaching students the right things? Threshold concepts?

Given all this, one question is whether the content and balance of content in statistical units is restricting deeper levels of understanding. A cookbook analogy has been used to describe the curriculum of many units, where students are taught, in sequential fashion, a battery of statistical tests for different scenarios, each akin to a recipe. In such curricula, which were perhaps designed to cover the vast swathe of statistical tests typically applied in each field, the underlying concepts that link all the methods and key aspects of interpretation and comprehension are given insufficient attention. So rather than statistical learning empowering students with new insights and a framework for critical thinking, we often end up with a curriculum that is "stuffed with content" (Cousin, 2006), rote learned and quickly forgotten. However, it is the underlying concepts and ways of thinking that contribute to statistical literacy. And as Garfield (1995) points out, if teachers were asked what they would like their students to know one year after completing a course in statistics, most would probably describe a capacity to think statistically and to be able to apply some key statistical ideas rather than an ability to look up probabilities in a statistical table. Despite this, these desired learning outcomes almost always remain tacit in the list of learning outcomes of a unit.

If the tacit outcomes of statistical literacy are made explicit and our statistical teaching is focused more towards understanding concepts, and to depth rather than breadth then a key question is whether we are focusing on the right concepts. One theory to help us think about this is the so-called "threshold concept", proposed by Land and Meyer (2003). According to Land and Meyer, a threshold concept is (a) Transformative - it changes the way a learner thinks about a subject, (b) Irreversible - due to its transformative nature it is unlikely to be forgotten, (c) Integrative - it enables students to make connections to other concepts that were previously hidden, (d) Troublesome - it may appear counterintuitive or alien until learnt, and (e) Bounded - there are borders separating it from other conceptual areas. Understanding which concepts in a subject may be "threshold" allows one to think more clearly about the whether we are teaching the right things to achieve the intended learning outcomes, and in particular to make decisions on the content of the course, emphasising those important topics (concepts) and de-emphasising or even removing others, thereby de- stuffing the curriculum.

Background to the statistical courses for non-statisticians

Two years ago I made substantial changes to the curricula of two post-graduate statistics units that I lead. These taught units are part of the MSc Reproduction & Development and MRes Health Sciences Research programs at the University of Bristol. The changes I made reflect my description above, a shift from breadth to depth, from rote learning tests to understanding concepts. More attention is now placed on appreciating the role of statistics in quantifying uncertainty in research, how ideas such as p-values and confidence intervals allow us to estimate this uncertainty, on understanding these ideas in more depth, and on exercising good judgement when interpreting p-values and confidence intervals. One key change I made to achieve this was to cover the concept of a sampling distribution in much more depth than other courses. I was unaware of Land & Meyer's theory of threshold concepts at the time I revised the course but looking back, I had ascribed the sampling distribution with similar pedagogical status to a threshold concept. Despite the changes I made, there is still scope for reducing the content in my course, as supported by student feedback and peer review.

Objective

The objective of this enquiry is to investigate potential threshold concepts within the learning of statistics. Given the changes to my course I was particularly interested in exploring whether there is any evidence for the sampling distribution being a potential threshold concept. My enquiry reviews the literature in this area, and reports and discusses the results from a questionnaire that I designed and administered to a group of students who had taken one of my statistics modules and to staff with a postgraduate qualification in statistics.

Previous research and evaluative reports on threshold concepts in statistics

There have been relatively few empirical attempts to identify threshold concepts in statistics among students (Bulmer et al., 2007; Dunne et al., 2003; Khan, 2014), none to trained statisticians or academic staff, and to the best of my knowledge only two reports have discussed threshold concepts in statistics from a more theoretical or epistemological perspective (Norton, 2015; MacDougall, 2010).

Bulmer et al (2007) asked over 500 biology students who had undertook a statistics module to describe one or two important concepts that they had learnt. P-values, confidence intervals, hypothesis testing and analysis of variance were most frequently cited. The students were also asked about concepts that they found difficult, and the most frequent response concerned how to select a particular statistical technique. Khan (2014) searched for threshold concepts in statistics among business students undertaking a statistics module. The focus was on identifying troublesome aspects, and was done through proxy by identifying questions from exam scripts that students performed poorly on. Under this criteria, the chi-squared test was identified, and the authors reasoned that the calculation of expected frequencies and determining trends in data were threshold concepts. Dunne et al. (2003) also focused on aspects students found most difficult and also identified the chi-squared test.

Leaving aside the conflation of difficult and troublesome, there are several potential limitations and points of discussion. First applying almost all weight to the criterion of troublesomeness is not in concordance with Land and Meyers description of a threshold concept. There are many troublesome concepts in statistics that are not threshold, deriving estimators for instance. Land and Meyer (2003) also state that a threshold concept is likely to be troublesome, implying that it is not a necessary characteristic for threshold status. Second, a student’s response is limited by the curriculum they’ve been exposed to. So if the focus is on rote learning statistical tests rather than a deeper understanding of concepts, then unless a student has read outside the curriculum, they will almost certainly cite a particular test as most difficult. Lastly and related to this, understanding a threshold concept is possibly a prerequisite for describing a concept as such, a study that surveys statisticians or staff as well as students may thus offer a more comprehensive search.

With respect to theoretical justification, Norton (2015) describes the sampling distribution as a threshold concept in some detail; he notes that understanding the underlying principles [ie; sampling distribution] that allow the calculation [of p-values & confidence intervals] is the key to unlocking understanding of the entire premise of inferential statistics. Kennedy 1998 (cited by Meyer, 2016) also wrote about the sampling distribution in a similar light. And lastly, MacDougal (2010) described the difficulty of getting medical students to tolerate ideas of uncertainty and in particular confidence intervals - as stated by Norton (2015), both these concepts are underpinned by the sampling distribution.

Survey methods

Design

A questionnaire was designed and administered to two groups of students from the University of Bristol who had completed my statistics course within the last 4 months, and to a group of University of Bristol academics with postgraduate qualifications in statistics and/or epidemiology. The survey was anonymous.

The questionnaire was administered using the Bristol Online Survey tool. Students on the MRes Health Sciences Research (n=5) and MSc Reproduction & Development (n=22) were invited to participate through notifications on Blackboard and by email. Two follow up requests were sent and 15/27 (56%) responded. Staff were invited via an email circulated on the Applied Statistics email forum. This forum is predominantly comprised of statisticians and epidemiologists, 15/33 (45%) participated, of which 13 reported having a statistical or epidemiological background.

Questionnaire

The following two questions were asked to elicit information about potential threshold concepts:

> From the list of statistical concepts below, please select up to three items that you found most difficult to learn and understand, ranking them in order of difficulty.
>
> Did any particular concept deepen your understanding of statistics? Think about those which may have helped you link statistical ideas together or interpret data and read papers in a different way. Please select up to three items.

The list of options for respondents to select was compiled based on the way the course is organised and referenced so that the terminology was familiar to the students (see figure 1 for the list of items). The list of items was the same for the staff because the purpose is to find out about threshold concepts in postgraduate statistics to medical and science students not a postgraduate course in statistics for statisticians. Nonetheless, an 'other' option was included. Many items could have been grouped but they were listed individually to (a) widen access to the potential choices, (b) to avoid imposing a conceptual structure on the participant, and (c) given the enquiry was explorative, it was thought that the pattern of responses may be informative.

The first question attempts to capture the criterion of troublesomeness and the second targets the criteria of being transformative and integrative. The characteristic of being irreversible was not possible to assess in our sample, and the characteristic of being bounded is an epistemological endeavour. According to Land and Meyer (2003), the characteristics of being transformative, integrative and bounded are necessary constructs for a threshold concept, while the concepts of irreversibility and troublesomeness are described as likely to be seen. The questionnaire was thus designed to capture information on as many of these dimensions as possible and in as parsimonious a way as possible.

Analysis

The analysis was descriptive. The results were plotted and given the small sample size, are presented as frequencies.

Survey results & interpretation

Among students, 9/15 (60%) described their background as medical, and the rest as scientific. For staff, 12/13 (92%) had a background in statistics, the other was an epidemiologist. Figures 1 a and 1b & 2a and 2b plot the responses from the two main questions that screened for threshold concepts.

Figure 1a. Topics (concepts) that staff found most difficult.

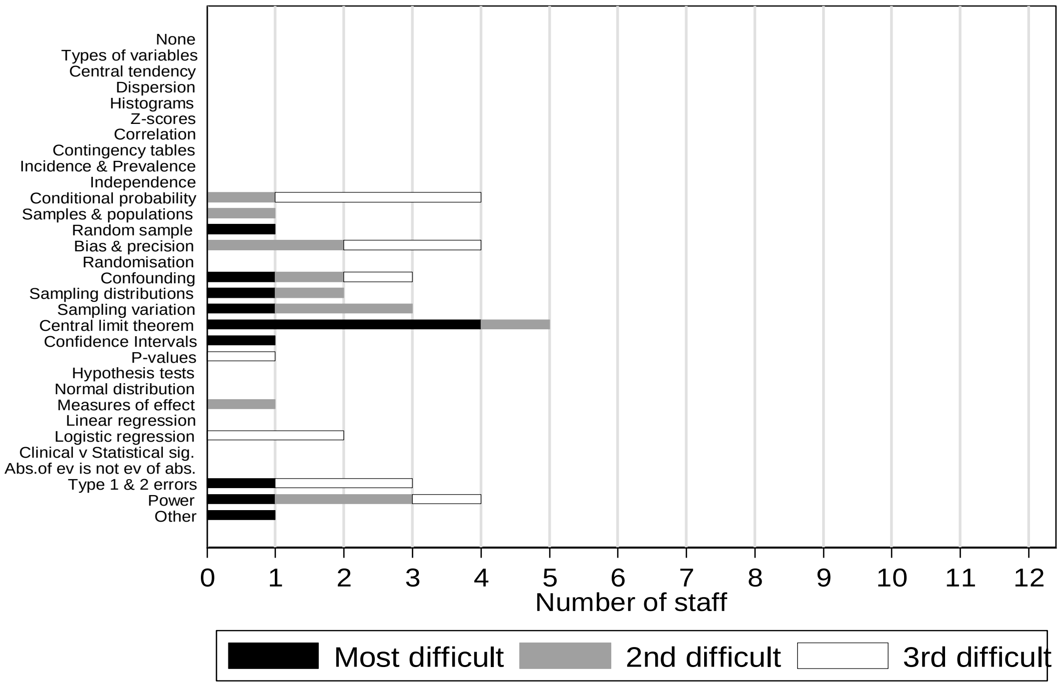


Figure 1b. Topics (concepts) that students (bottom) found most difficult.

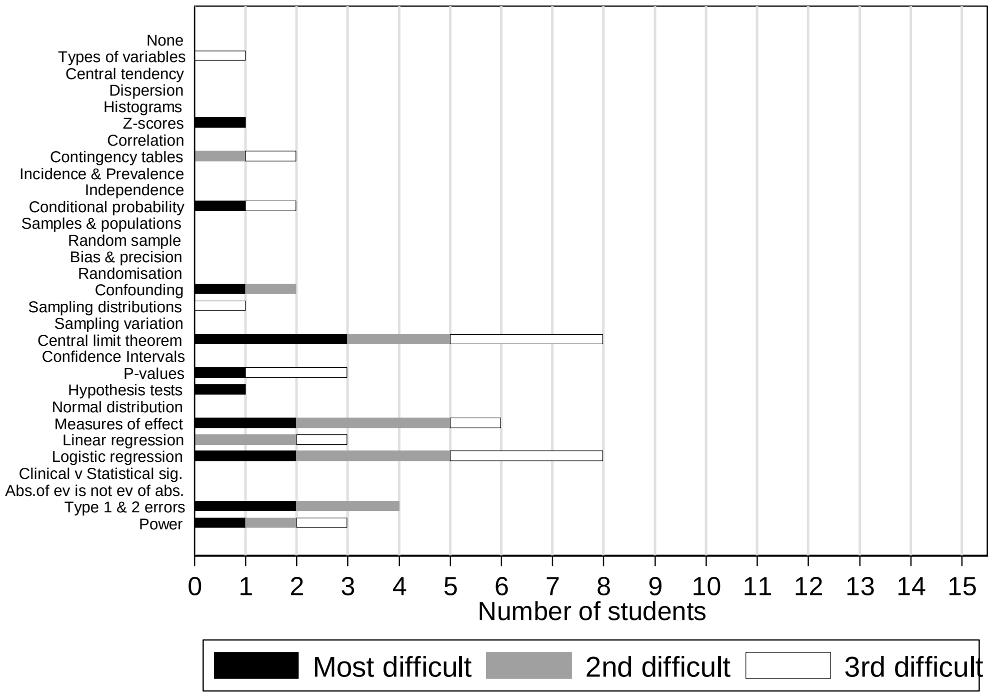


Figure 2a. Topics (concepts) that) found most transformative.

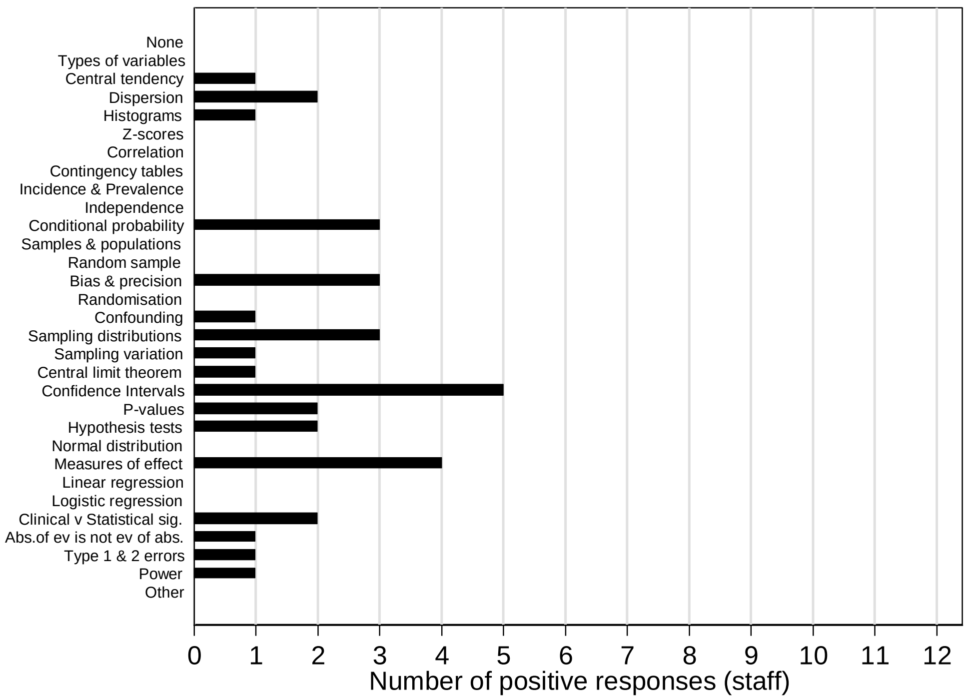
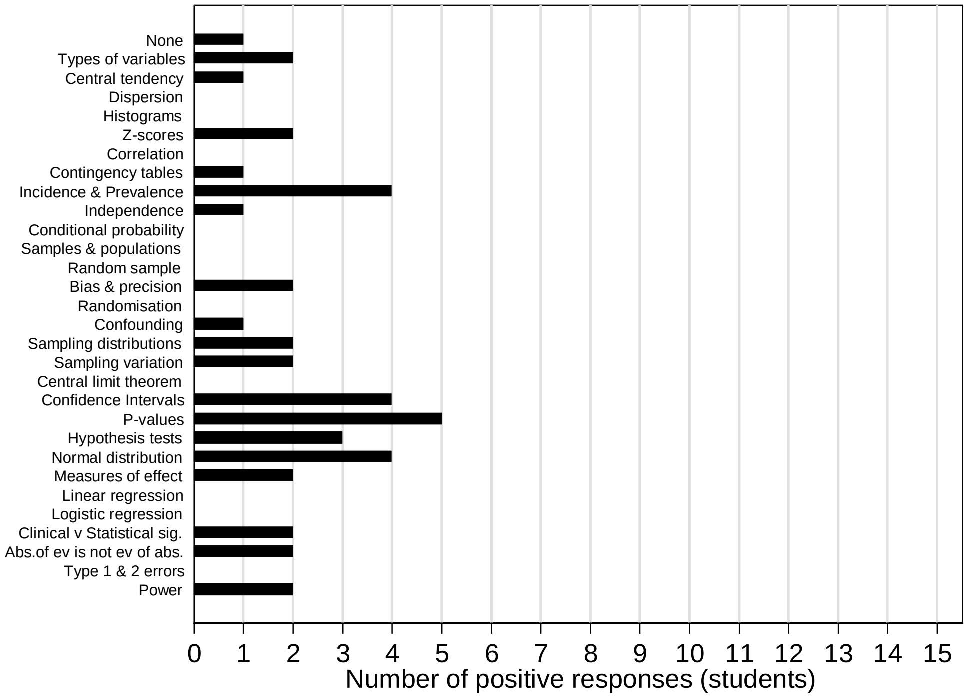


Figure 2b. Topics (concepts) that students found most transformative.



Statistical inference, understanding uncertainty and sampling distributions?

Central limit theorem was noted as the most difficult concept among both staff and students. Staff reported sampling distributions, bias and precision and sampling variation as difficult, while only one student did. In contrast, students reported p-values and hypothesis tests as difficult while only one staff member did. Both staff and students reported Type 1 & 2 errors and power as difficult topics. The concept of sampling variation and its sampling distribution is integrative to all of these concepts. All of the concepts just described were also identified by both staff and students in the question that targeted transformative concepts with the addition of clinical versus statistical significance and absence of evidence is not evidence of absence. The latter two are practical heuristics for critical appraisal but again are underpinned by the concept of sampling variation.

Other concepts?

A few students and staff flagged descriptive concepts such as central tendency, dispersion, types of variables, histograms, incidence and prevalence and contingency tables as potentially transformative. These topics are important foundation topics and are all part of the topic of understanding and describing data. Several staff also noted conditional probability and one student flagged independence. These items concern probability and as with the former are foundation topics.

Discussion

A range of concepts were flagged by students and staff as being either transformative, or transformative and troublesome. These two conditions reflect Land and Meyer's (2003) criteria for a threshold concept. The list of potential concepts that were presented to participants contains overlap, as such the concepts are not separate entities. One interpretation of the concepts that were identified is that they reflect three underlying themes. The first theme that had most support concerns the nature of statistical inference and the ideas of uncertainty and sampling variation; the second concerns basic probability and the third which had least support concerns describing and visualising data (central tendency, dispersion).

As stated, with respect to the first theme, Norton (2015), and Kennedy (1998, cited by Land & Meyer, 2003) directly cite the sampling distribution, and MacDougall (2010) also indirectly talks about similar ideas. Exploring the pedagogy of the concepts in this theme with other experts, in particular identifying those that are integrative to the others, may be useful to identify the concept(s) that are potentially threshold and provide useful information on how to teach these ideas.

The second theme is interesting because the concepts of independence and conditional probability often get muddled in our everyday thinking, leading to fallacious reasoning. Psychologists consider it a form of cognitive bias (Garfield, 1995). There are many examples of this type of fallacious reasoning, one being the Sally Clarke case, which resulted in a mother being wrongfully convicted of the murder of her two children, both of whom died tragically but unsuspiciously from sudden infant death syndrome (https://en.wikipedia.org/wiki/Prosecutor%27s_fallacy). In this example the prosecutor's evidence was misrepresented by mixing up conditional and independent probabilities. The ideas of independence and conditional probability are also linked to many other aspects of my curriculum, and are certainly bounded. The reports that cite the chi-squared test as a threshold concept (Dunnet et al, 2003; Khan, 2014) also supports this finding since the key to understanding the calculations for a chi-squared test requires understanding independence and conditional probability. It is perhaps not surprising then that these topics were raised as being transformative. Again though, it is worth interrogating the pedagogy of these topics.

Surveying both staff (i.e. statisticians/ so-called experts!) and students, and asking about both troublesome and transformative knowledge possibly increases the sensitivity for detecting potential threshold concepts. For example there was some discordancy between the troublesome and transformative questions (eg, central limit theorem) and some discordancy between staff and students (eg, conditional probability). Nonetheless, my enquiry was limited due to time constraints and so results are only preliminary. Replication with a larger sample size is required, and also the use of other approaches to validate and understand these results and their implications in more detail. With respect to the latter, qualitative methods could be used such as staff and student focus groups, and exam results could be interrogated with respect to troublesomeness.

Summary and conclusions

This preliminary enquiry has flagged three potential areas where threshold concepts in statistics may exist – the nature of statistical inference and uncertainty; probability; and descriptive statistics. Further thought and investigation is needed to identify potential candidates for threshold status within these topics. Nonetheless, it is clear that greater statistical literacy and understanding among non-statisticians will most likely be achieved by ensuring that curricula accommodate time for learning concepts in depth rather than rote-learning a great many statistical tests.

Finally, if we are serious about improving levels of statistical literacy, then it is important that we translate this goal into explicit learning objectives and outcomes and seek to understand more about the pedagogy of the subject for non-statistical learners. The theory of threshold concepts will hopefully provide a useful framework for these investigations.

> There is nothing as practical as a good theory. – Kurt Lewin

Acknowledgements

Thank you to all the staff and student at the University of Bristol who completed the survey. Thanks also to my students for igniting and stimulating a passion for helping you to learn statistics.

References

Altman, D, G. (1994). The scandal of poor medical research. BMJ. 308;203.

Anderson B, L., Williams S., & Schulkin J. (2013). Statistical Literacy of Obstetrics-Gynecology Residents. Journal of Graduate Medical Education, 5(2):272-275.

Cousin, G. (2006) An introduction to threshold concepts, Planet No 17, December 2006, pp 4-5.

Bulmer, M., O’Brien, M., & Price, S. (2007). Troublesome concepts in statistics: a student perspective on what they are and how to learn them, UniServe Science, Proceedings of the Assessment in Science Teaching and Learning Symposium, University of Sydney, September 28-29, 2007, pp 9 – 15

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. Lancet. 374:86-89.

Dunne, T., Low, T., & Ardington, C. (2003). Exploring Threshold concepts in basic Statistics, using the Internet, The International Association for Statistical Education, Statistics & the Internet, Berlin, Germany, 2003, (IASE/ISI Satellite, 2003)

Gal, I. (2002). Adults’ statistical literacy: Meanings, components, responsibilities. International Statistical Review. 70,1,1-51.

Garfield, J. (1995). How students learn statistics. International Statistical Review. 63, 1, 25-34.

Gigerenzer, G (2003). Reckoning with risk: learning to live with uncertainty. London, UK: Penguin

Khan, R. N. (2014). Identifying Threshold Concepts in First-Year Statistics, Educational Research and Perspectives 41, 217-231

Ioannidis J. P. A. (2005). Why Most Published Research Findings Are False. PLoS Med 2(8): e124. doi:10.1371/journal.pmed.0020124

Meyer, J. H. F., & Land, R. (2003). Threshold concepts and troublesome knowledge: linkages to ways of thinking and practising, In: Rust, C. (ed.), Improving Student Learning - Theory and Practice Ten Years On. Oxford: Oxford Centre for Staff and Learning Development (OCSLD), pp 412-424.

MacDougall, M. (2010). Threshold concepts in statistics and online discussion as a basis for curriculum innovation in undergraduate medicine. HEA Maths, Stats & OR Network: MSOR Connections Vol 10 No 3 Autumn Term 2010, pp 21-24 & 41.

Nikiforidou, Z., Lekka, A., & Pange, J. (2010). Statistical literacy at University level: the current trends. Procedia Social and Behavioural Sciences. 9: 795-799

Norton, S. (2015). Sampling distributions as a threshold concept in learning classical statistical inference: an evaluative case study report. King's Learning Institute Higher Education Research Network Journal (HERN J): Prizewinning Essays 2015, pages 58-71:

Tankard, J, W .(1979). The H.G. Wells quote on statistics: a question of accuracy. Historia Mathematics. 9. 33-36.