Published by Oxford University Press on behalf of the International Epidemiological Association
© The Author 2010; all rights reserved.
International Journal of Epidemiology 2010;**39**:1151–1153
doi:10.1093/ije/dyq183

Commentary: Charles Spearman and correlation: a commentary on 'The proof and measurement of association between two things'

Sandy Lovie[1] **and Pat Lovie**[2]*

[1]School of Psychology, University of Liverpool, Liverpool L69 3BX, UK and [2]School of Computing and Mathematics, Keele University, Staffordshire ST5 5BG, UK

*Corresponding author. School of Computing and Mathematics, Keele University, Staffordshire ST5 5BG, UK.
E-mail: p.lovie@maths.keele.ac.uk

Accepted 6 September 2010

In 1904, a pair of remarkable papers appeared in the *American Journal of Psychology*.[1,2] The author, Charles Edward Spearman, a 40-year-old former army officer, was at that time still a doctoral student at Germany's (and Europe's) leading centre of experimental psychology, the Leipzig laboratory, founded and run by Wilhelm Wundt. Spearman had joined the army on leaving school and served for almost 15 years, mainly in India. During these years of military service he became increasingly attracted to philosophy and psychology, and shortly after completing a course at the Army Staff College in 1897, from which he gained the coveted qualification of 'psc' ('Passed Staff College'), Spearman resigned his commission in order to study in Leipzig. Spearman still had no formal academic qualification in 1904.

The genesis of these linked papers is also extraordinary. In 1900, at the onset of the Boer War, Spearman was recalled to military duty and it was during a couple of years away from Germany that he seems to have become interested in intelligence and mental testing. Just before his return to Leipzig, and apparently entirely off his own bat, Spearman carried out a series of such studies on some Berkshire schoolchildren and adults. This work provided the empirical stimulus and rationale for these two articles. The first of Spearman's 1904 papers, entitled 'The proof and measurement of association between two things',[1] is concerned with measures (and problems) of correlation, i.e. with the basic tools used in the second paper for his pioneering efforts to develop what we now call factor analysis.[2]

Before turning directly to the former paper, it is worth providing some context for Spearman's work. The problem that he claimed to have both discovered and solved within 19th century literature on intelligence was how to detect and then measure intelligence, where intelligence was taken to be a rather more abstract and higher level concept than specific abilities to, say, solve problems in arithmetic or to memorize lists of countries or important dates. Spearman was, in other words, seeking what later became known as general ability (or g) as distinct from the less interesting special ability (or s). The former was the more abstract part of all abilities, the latter the specific and unique portion of a particular ability. Notice that, in 1904, Spearman was advocating a single unitary concept that was common to all abilities, not the more complex hierarchical structures of ability championed by later writers. Such an essentially minimalist system, therefore, allowed Spearman to represent intelligence (or something very like it) as a property underlying performance on all abilities. The problem posed by all 19th century tests of intelligence was that they were unable to detect or measure intelligence separately from the specific ability that the particular tests were assessing. Spearman's solution was not to change the tests as such, but to improve the statistical methods used to measure performance. In essence, this meant substituting partial correlations for the uncorrected measures of association then available. By doing this, he hoped to correct for the intercorrelation of all abilities, thus yielding a more refined and less contaminated measure of association between tests and performance. Later and more sophisticated versions of this approach are now known as latent variable analysis, thus making more explicit the hitherto rather hidden, rather abstract nature of intelligence. It also meant that intelligence was rather more difficult to define in anything other than somewhat abstract terms: Spearman himself had several attempts at pinning it

Downloaded from http://ije.oxfordjournals.org/ by guest on May 31, 2014

down, referring to it early on as a plastic ability. Eventually, however, he stuck with *g* and resisted any attempts to unambiguously equate this with intelligence *per se*.

One can therefore appreciate the nature of Spearman's rather delicate balancing act between too concrete, and hence too specific a definition of intelligence, and the more intangible, more generalized definition that might easily disappear as a meaningless abstraction.

See Lovie and Lovie[3] and also Spearman's autobiographical essay[4] for further material about his life and work. A discussion of his legacy to test theory can be found in Levy.[5]

In his introductory remarks to 'The proof and measurement of association between two things',[1] Spearman observes, with some exasperation, that studies of the connections between two events or attributes are commonplace within psychology and yet investigators have shied away from using the available correlational measures that would have allowed them to make precise evaluations of such associations. His mission, therefore, would be to 'remedy this deficiency of scientific correlation' (p. 73) by providing formulae that could be used even by those with only elementary mathematical skills.

In the first major section of the article, Spearman discusses the requirements for a good measure of correlation, such as ease of use and accuracy (in terms of the probable error). He then considers several possibilities, starting with what he terms the standard method. At first sight, this appears to be the familiar Bravais–Pearson product moment formula and indeed is cited as such. On further inspection, however, we see that the squared deviations are from medians instead of means. The reason for this remains a mystery though, interestingly, in a subsequent article published 2 years later medians are changed to means without any explanation.[7]

Although paying lip service to other possible measures based on counts, Spearman's real aim is to promote the notion of using ranks instead of actual measurements. While apparently arguing both the pros and cons of such a substitution, his many claims for the advantages, such as mitigating the effects of accidental error or of distributional disparities, make his defence of ranks quite plain. The first simple measure that he proposes (in effect, the straw man), which could be used with either measurements or ranks, is based on differences between class averages (p. 85), but is let down by its accuracy (probable error) and thus discounted. Then Spearman presents his *pièce de résistance*, the method of rank differences (p. 86). One might quite reasonably expect this new index to be what we now know as the Spearman's rank correlation coefficient which, of course, straightforwardly substitutes ranks for measurements in the product moment formula, but instead it actually uses unsigned differences between paired ranks. For all of his advocacy of this method, and a claim that it was just as good as the product moment coefficient for one of the worked examples, even Spearman has to acknowledge that it still had quite serious disadvantages. In fact, Spearman never pursued this particular version and his next foray into finding an easy and accurate rank method in 1906 came up with the so-called footrule.[6] Ironically, he also derives the now familiar formula from the product moment version but dismisses it as being both more difficult to calculate and more susceptible to extreme values than is the footrule. See Lovie[7] for a more complete account of the evolution of the Spearman rank correlation coefficient.

In Part II of the article, Spearman tackles the thorny issue of how to deal with what he terms 'systematic deviations' from some 'real general tendency' (p. 88). These errors are not merely accidental due to sampling variation (and thus quantifiable by probable error), but are errors of measurement that could have a systematic effect on correlations. The first problem identified by Spearman is that of attenuation of correlation. Here, errors in measurement or observation are assumed to attenuate or reduce the true correlation that would be obtained if the two variables could be measured with perfect reliability.

As we discover from brief mentions in the second 1904 paper[2] and Spearman's autobiographical essay of 1930,[4] his concerns about the weakened association due to such errors did not come out of the blue. While Spearman had found substantial correlations for the tests carried out on his Berkshire schoolchildren, a study published in 1901 by Wissler[8] on American college students, which was probably the first ever within psychology in which Pearson's product moment method had been applied, reported virtually no associations between a range of psychophysical tests and academic grades or even between the tests themselves. Quite what it was that Spearman saw as faulty about the American data that gave him his brainwave is not spelt out. It could be argued that without Spearman's insight and perseverance in trying to account for the discrepancy between those results and his own, the idea of a scientific basis for intelligence testing might well have remained in the doldrums.

Now, Spearman's idea was simple: provided that one could make at least two independent series of observations from each variable, then the errors of measurement could be estimated and thus corrected for by assessing the size of the discrepancies between the series (pp. 90–91). Nonetheless, Spearman identifies two possible problems with this approach: the effect of irrelevant factors and the difficulty of obtaining completely independent series. To counter these problems, he suggests a second method, which he claims would avoid both of the above defects. This formula shows explicitly the relationship of

Downloaded from http://ije.oxfordjournals.org/ by guest on May 31, 2014

the length of the tests to their intercorrelation (pp. 91–92). And from this, after some twist and turns and the not-always-friendly intervention of William Brown, we end up with what is generally known today as the Spearman–Brown reliability or prophesy formula.[9,10]

Spearman concludes his homily on the perils of systematic deviations with a discussion of how poor sample selection and the presence of irrelevant variables might cause other problems such as constricting, dilating or distorting correlations. Notable here is that Spearman gives a formula for partial correlation which, although acknowledging Yule's priority, he implies had been derived independently (p. 95).

The final section, 'Criticism of prevalent working methods', is a masterclass on how to make enemies. Spearman again rebukes (unnamed) psychologists for not embracing 'the brilliant work . . . by the Galton-Pearson School' (p. 96), but then goes on to assert that even the latter (by which he actually means Pearson, of course) is guilty of turning a blind eye to the problems of systematic deviations due to measurement error. If his earlier complaints that the Pearson product moment correlation coefficient was too difficult to calculate and the assertion that his own method was just as good, might just have gone unanswered, this later attack did not: Pearson's responses in 1904[11] and 1907[12] were blistering. Spearman had started a bitter and lifelong feud with Pearson. But that, and Spearman's many other battles with his critics, is yet another story.

From this relatively short maiden article of 1904, we have the beginnings of rank correlation, of assessing test reliability and of correcting for attenuation. Not a bad start for a partially self-taught student with no academic qualifications apart from 'Passed Staff College'.

Conflict of interest: None declared.

References

1. Spearman C. The proof and measurement of association between two things. *Am J Psychol* 1904;**15**:72–101. Reprinted. *Int J Epidemiol* 2010;**39:**1137–50.
2. Spearman C. ''General intelligence,'' objectively determined and measured. *Am J Psychol* 1904;**15:**202–93.
3. Lovie P, Lovie AD. Charles Edward Spearman, F.R.S. (1863–1945). *Notes Rec R Soc Lond* 1996;**50:**1–14.
4. Spearman C. In: Murchison C (ed.). *A History of Psychology in Autobiography*. Vol. 1. Worcester, MA: Clark University Press, 1930, pp. 299–331.
5. Levy P. Charles Spearman's contributions to test theory. *Br J Math Stat Psychol* 1995;**48:**221–35.
6. Spearman C. 'Footrule' for measuring correlation. *Br J Psychol* 1906;**2:**89–108.
7. Lovie AD. Who discovered Spearman's rank correlation? *Br J Math Stat Psychol* 1995;**48:**255–69.
8. Wissler C. The correlation of mental and physical tests. *Psychol Rev* 1901;**3(Suppl 6):**1–62.
9. Spearman C. Correlation calculated from faulty data. *Br J Psychol* 1910;**3:**271–95.
10. Brown W. Some experimental results in the correlation of mental abilities. *Br J Psychol* 1910;**3:**296–322.
11. Pearson K. On the laws of inheritance in Man. II. On the inheritance of the mental and moral characters in Man, and its comparison with the inheritance of the physical characters. *Biometrika* 1904;**3:**131–90.
12. Pearson K. On further methods of determining correlation. *Mathematical Contributions to the theory of evolution – XVI*. Drapers' Company Research Memoirs. Biometric Series IV. London: Dulau, 1907.

Downloaded from http://ije.oxfordjournals.org/ by guest on May 31, 2014

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2010;**39**:1153–1155
doi:10.1093/ije/dyq182
© The Author 2010; all rights reserved.

Commentary: 'The next trick is impossible.'

Julie Perks

Department of Psychology and Mental Health, Staffordshire University, College Road, Stoke-on-Trent Staffs, Staffordshire ST4 2DE, UK. E-mail: j.perks@staffs.ac.uk

Accepted 2 June 2010

In 1906, Wilhelm Wundt wrote a reference for his departing student, Charles Spearman, in which he described the latter as an 'investigator who is thoroughly versed in the physical mathematical sciences auxiliary to psychology'.[1] Experimental Psychology had progressed quickly from its infancy in the 1860s to its first schools at the end of the 19th century when, Leipzig based, Wundt was its most highly