Parabola Volume 48, Issue 2 (2012)

History of Mathematics: The Vectors of Mind

Michael A B Deakin[1]

My title for this column is that of an influential book, first published in 1935 by the American psychologist L. L. Thurstone. It is almost self-explanatory: although we are accustomed to see intelligence measured as a simple scalar, the IQ, most of us realize that this single number has to be a gross oversimplification of the underlying reality. Different people exhibit different mental capacities. Here is an obvious example: there are those of us who (like *Parabola*'s readers) are good at Mathematics, but there are also otherwise highly intelligent people who have no aptitude for it at all!

This was the sort of situation that Thurstone wanted to address. He believed that different aptitude tests could respond in different degrees to underlying intelligence "factors", and that a suitable mathematical analysis could determine what these "factors" are. *The Vectors of Mind* itself is essentially a rather dry methodological work, detailing the mathematical techniques he hoped to use in uncovering the "factors", these being seen as more fundamental than the raw scores attained on the different tests themselves.[2]

The mathematical techniques Thurstone hoped to use in this ambitious program of research are lumped together (along with others) and given the name "factor analysis". It remains a somewhat controversial subject, in part because there are many different ways of doing it and these can lead to very different answers. I will illustrate the underlying principles by following a specific example. However, before entering on this illustration, I need to develop a number of mathematical preliminaries. Some readers will have encountered this material, but others not, so I provide brief descriptions here.

First we need some concepts from descriptive statistics. Suppose we have a set of measurements $\{x_1, x_2, \ldots, x_n\}$. The *mean* of this set is defined as

$$\overline{x} = (x_1 + x_2 + \cdots + x_n)/n.$$

Its *variance* is defined as

$$\mathrm{Var}(x) = \left[(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2\right]/n.$$

[1]Dr Michael Deakin is an Adjunct Senior Research Fellow in the School of Mathematical Sciences at Monash University.

[2]The word "vector" used here perhaps requires explanation. In this context, it means a set of n ordinary numbers (scalars). In the terminology to be developed later in this article, a vector is a $1 \times n$ (or $n \times 1$) matrix. Some readers may have met the term "vector" in the context of Theoretical Mechanics. Specifically these are "3−vectors", and they find exemplification in momentum, force, etc. 3−vectors possess a number of special properties, which are not of interest for our present purposes.

This provides a measure of the extent to which the measurements scatter about the mean (or average). The square root of the variance is known as the *standard deviation* and is designated by the symbol σ_x; this may be a more familiar concept to readers than is the variance. However, the variance is also important in our present context.

For another set $\{y_1, y_2, \ldots, y_n\}$, the analogous quantities may be defined, with the symbols changing in the obvious way. For the two different sets, we need also a quantity known as the *covariance* and defined as follows:

$$\begin{aligned}\mathrm{Cov}(x, y) = [&(x_1 - \overline{x})(y_1 - \overline{y}) + \\ &(x_2 - \overline{x})(y_2 - \overline{y}) + \cdots + (x_n - \overline{x})(y_n - \overline{y})].\end{aligned}$$

Another important idea for our purpose is that of a *correlation coefficient*. This is a statistic defined as $\mathrm{Cov}(x, y)/(\sigma_x \sigma_y)$. It relates the two different sets of measurement: $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_n\}$, and measures how closely the points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

come to lying on a straight line. A correlation coefficient consists of a real number with an attached sign. It may be shown that the actual number always lies between 0 and 1; the sign is that of the slope of the best fitting straight line. It is common, if somewhat inaccurate, to regard a correlation coefficient near ± 1 as indicating some underlying connection between the two sets of measurements.

The other mathematical preliminary we need is some basic matrix algebra. A *matrix* (plural *matrices*) is a rectangular array of numbers in a pattern like this:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Note the systematic way in which the different entries are numbered; this is standard. This particular matrix has m (horizontal) rows and n (vertical) columns. Such a matrix is described as $m \times n$. If $m = n$, the matrix is square; all the matrices to be discussed in the rest of this article will be square. Square matrices exhibit a number of useful properties. An $n \times n$ matrix (A, let us call it) is associated with a polynomial of the nth degree, known as its *characteristic polynomial*. This will have n (real or complex) roots, which are known as the *eigenvalues* of the matrix A. To each eigenvalue corresponds an ordered set of n numbers, constituting the associated *eigenvector*. It is usual to scale the size of the components of an eigenvector so that their squares add up to 1.

The study of matrices began with the question of solving sets of simultaneous equations; eigenvalues and eigenvectors play a key role in this aspect of the work. The determination of eigenvalues and eigenvectors was once an extremely onerous task, but nowadays computer packages manage it routinely.

With these preliminaries behind us, we are now in a position to begin a discussion of factor analysis. I here discuss an actual example. In the $1920s$, 140 seventh grade children were subjected to a battery of tests measuring:

1. reading speed 2. reading "power"

3. arithmetic speed 4. arithmetic "power".

Their scores were analyzed statistically and a table of correlation coefficients r_{ij} was constructed as a 4×4 matrix, the "correlation matrix", which in this case turned out to be:

	$Test\ 1$	$Test\ 2$	$Test\ 3$	$Test\ 4$
$Test\ 1$	1	0.698	0.264	0.081
$Test\ 2$	0.698	1	−0.061	0.092
$Test\ 3$	0.264	−0.061	1	0.594
$Test\ 4$	0.081	0.092	0.594	1

You should notice several things about this 4×4 array of numbers. In the first place, the entries in the main diagonal (from top left to bottom right) are all 1. This is because a correlation of each test with itself will, of course, reflect perfect agreement. In terms of the notation used above, $y_1 = x_1$, $y_2 = x_2$, etc. and so the pairs of points all line up exactly on the line $y = x$. Next notice that the pattern is symmetric; this is because the correlation between, say, Test 1 and Test 2 is necessarily the same as the correlation between Test 2 and Test 1, and so on. Because of this property of symmetry, the eigenvalues of the correlation are all real numbers; indeed in the practical case, they are all positive. Furthermore the eigenvectors have an important property known as orthogonality. If one eigenvector is say (x_1, x_2, x_3, x_4) and another is (y_1, y_2, y_3, y_4), then $x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 = 0$. These properties follow from theorems proved in the course of systematic matrix analysis.

The basic idea is to try to approximate the entries in the table by simple products. So we seek numbers (x_1, x_2, x_3, x_4) such that $x_1^2 \approx 1$, $x_1x_2 \approx 0.698$, $x_1x_3 \approx 0.264$, etc. The idea is to make the approximation as good as can be achieved, and it can be shown that this is achieved by means of eigenvectors, specifically the eigenvector corresponding to the largest eigenvalue.

I had the package *Scientific Workplace* (which I use to prepare these columns) calculate the eigenvalues and eigenvectors of this matrix, which it does at the touch of a button. This gave:

an eigenvector of $(0.81350,\ 0.68915,\ 0.61098,\ 0.58085)$ corresponding to an eigenvalue of 1.84740,

an eigenvector of $(0.44284,\ 0.62390,\ -0.66966,\ -0.65604)$ corresponding to an eigenvalue of 1.46420,

an eigenvector of $(0.29083,\ -028788,\ 0.37629,\ -0.46156)$ corresponding to an eigenvalue of 0.52209,

an eigenvector of $(0.23984,\ -0.23011,\ -0.19149,\ 0.13853)$ corresponding to an eigenvalue of 0.16643.

Having got so far with *Scientific Workplace*, I took the values it generated and resorted to Excel for the rest of the calculation. (Excel actually provides an add-on that would have done the whole calculation, but my system doesn't have it.)

The largest eigenvalue is the first. Its overall contribution to the pattern in the correlation matrix is measured by the corresponding eigenvector, adjusted by multiplying its four components by the square root of its associated eigenvalue 1.84740. This gives a new vector $(0.81350, 0.68915, 0.61098, 0.58085)$. We may now write a matrix whose values are given by products of these four numbers; thus for example, the entry in the top row and the second column of this new matrix would be $0.81350 \times 0.68915 = 0.560\,62$, etc. This new matrix may now be compared with the original one, and a further matrix computed comprising the discrepancies. This is

$$\begin{pmatrix} 0.33821 & 0.13737 & -0.23304 & -0.39152 \\ 0.13737 & 0.52507 & -0.48206 & -0.30829 \\ -0.23304 & -0.48206 & 0.62670 & 0.23911 \\ -0.39152 & -0.30829 & 0.23911 & 0.66261 \end{pmatrix}$$

The entries in this matrix indicate quite large discrepancies, so the expectation of replacing each entry in the original matrix by a simple product has not been met. However this is the best we can do in this endeavour, and so it is concluded that the first factor alone is insufficient to "explain" the correlation coefficients in the original table. We have to look further. We do this by looking at the second eigenvector (corresponding to the second largest eigenvalue), and doing to the matrix above exactly as we did before to the original one. This produces a further matrix of discrepancies from the matrix just above:

$$\begin{pmatrix} 0.14211 & -0.13891 & 0.06352 & -0.10101 \\ -0.13891 & 0.13582 & -0.06426 & 0.10101 \\ 0.06352 & -0.06426 & 0.17825 & -0.20021 \\ -0.10101 & 0.10101 & -0.20021 & 0.23223 \end{pmatrix}$$

These new discrepancies are still too large to be ignored and so we proceed to look at the third eigenvalue, and again proceed exactly as before. This gives

$$\begin{pmatrix} 0.05752 & -0.05519 & -0.04592 & 0.03323 \\ -0.05519 & 0.05295 & 0.04407 & -0.03187 \\ -0.04592 & 0.04407 & 0.03666 & -0.02653 \\ 0.03323 & -0.03187 & -0.02653 & 0.01919 \end{pmatrix}$$

A further calculation based on the fourth eigenvalue produces a matrix all of whose values are zero, apart from some insignificant rounding errors. The idea is that each eigenvalue corresponds to a "factor", which measures something underlying the original measurements, and seen as being more basic. The contributions of the various "factors" are measured by taking each eigenvalue and expressing it as a percentage of the sum of all the eigenvalues, which in this case is 4. Thus we find that the first factor contributes $1.84740/4 = 46.2\%$, the second $1.46420/4 = 36.6\,\%$, the third $0.52209/4 =$

13.1% and the fourth $0.16643/4 = 4.2\%$. These percentages are those of the contributions of each component to the variance in the data. (Analyzing the variances in this way is a common statistical procedure.) The first "factor" was seen as corresponding to overall general ability, the second to a difference between arithmetic and verbal ability, the third to a measure of "speed versus deliberation" and the last to something not worth bothering about.

But now it is time for me to come clean and make a confession. In a sense, I have been leading you up the garden path. The version of factor analysis I have just demonstrated uses a technique known as "principal component analysis". For some authors, this is one among many approaches to "factor analysis"; for others, "principal component analysis" and "factor analysis" are two quite different things. Nearly forty years ago, when I first expressed an interest in this area, I consulted an eminent statistician, and was surprised to find that he considered factor analysis to be just so much mumbo jumbo, and quite unworthy of serious attention, whereas principal component analysis was an established mathematical technique and eminently respectable.

Principal component analysis was pioneered by Karl Pearson ($1857 - 1936$) in a 1901 paper, but it was developed and popularized by another statistician, Harold Hotelling ($1895 - 1973$). I have taken my example above from Hotelling's first paper on the subject (published in two parts in the *Journal of Educational Psychology, Vol*. **29** (1933), pp. 417-441, 498-520). The data were taken from work by the psychologist Truman Kelly in his book *Crossroads in the Mind of Man* (Stanford University Press, 1928). I have recalculated Hotelling's figures, in the process correcting some very minor inaccuracies. (Remember, I have the advantage of modern computing machinery; Hotelling probably had to do all the calculations by hand, perhaps with the help of some very simple mechanical aids).

Hotelling says straightforwardly that it was discussion with Thurstone that led him to look into the matter. However, Thurstone went his own way and did not use principal component analysis. As a later author, Dennis Child, wrote (in *The Essentials of Factor Analysis*, London: Holt Rinehart and Winston, 1970 and later editions): "[Thurstone's] primary objective was to organize the factor axes ... so that their meaning would make better sense in psychological terms".

Here we have the key to the disputes that rage around factor analysis. Essentially, Thurstone wanted to use his psychological insight to guide his choice of factors. Any mathematical analysis (and there was a lot of it) was directed to his attempts to confirm his initial hunches. Compare this with Hotelling's example outlined above; there the factors were *entirely* determined by the data matrix, and it was only *after that analysis was complete* that Hotelling attempted to assign meanings to them. This is a philosophical point that will need to be looked at later, but there are also mathematical questions involved.

Because the data matrix is symmetric, its eigenvectors are orthogonal. This implies that they are independent of one another. The correlation between any pair is zero. With other approaches, which do not necessarily use eigenvalues, we lack this assurance and need to go to some length to remove from (e.g.) Factor 2 any residual component of Factor 1, and so on. So the technique of principal component analysis automatically deals with a potentially worrying complication.

Essentially, principal component analysis is a (relatively) straightforward mathematical technique, absolutely devoid of any value judgments, at least until the very end when we try to *name* the contributions from the different eigenvectors. This is why my eminent statistician colleague of all those years ago dismissed other approaches as lacking in rigor (and also why I chose this example to illustrate this column). Nonetheless, this need not be the last word on the matter. Just because the *Mathematics* is above reproach, does this mean that the components discovered in the course of the analysis have any real meaning at all in psychological terms? The answer to this question is almost certainly "No!", and this can most readily be demonstrated by using the same data, but applying it to the matrix of covariances rather than of the correlations.

This would seem to be equally mathematically "respectable" and one might think that it would result in similar conclusions, but this is not always the case. Indeed there are some reasons for preferring the use of a covariance matrix to a correlation matrix. In particular, the sum of the eigenvalues of a correlation matrix must equal the number of variables. (Recall that in Hotelling's example discussed above, it came to 4, also the number of tests involved; this was not a coincidence.) This constitutes a constraint imposed by the use of the correlation matrix.

In a discussion of Hotelling's analysis, a later author, J. Edward Jackson (*A User's Guide to Principal Component Analysis*, New York: Wiley, 1991) finds other reasons for possible disquiet:

> [It is implied] that only reading and arithmetic variables are of concern. If one were really trying to measure *overall* ability, other criteria besides these two would have to be considered. . . . The inclusion of just one more reading variable would change all the coefficients . . .

But beyond these mathematical questions, there are fundamental philosophical ones. There is really no reason at all to suppose that a formal mathematical procedure will result in a psychologically significant construct. This is a matter discussed in great detail by an advocate of other forms of factor analysis, Stanley Mulaik (in a lengthy article published in the technical journal *Multivariate Behavioral Research, Vol.* **22**, 1987). Essentially, the argument boils down to a distinction between *objective* and *subjective* evidence. The *objective* factors are produced by firmly based mathematical algorithms such as Hotelling's; we can't dispute their values, but we are at liberty to query their significance. The *subjective* factors on the other hand have clear significance, but they have no claim to be anything other than personal opinion. Thurstone preferred the subjective approach because "he suspected [principal components] were arbitrary and artifactual". Mulaik agrees. For such authors, one need not be shy of advancing one's own carefully considered opinions; it is quite reasonable to do so.

Now Thurstone was by no means the only nor even the first researcher to investigate these questions, but I pay particular attention to him for three reasons. First he has provided me with a catchy title for this column, secondly he was certainly extremely influential, and thirdly because I have a (very minor and distant) personal involvement with this part of the very complex history.

In a sequel to *The Vectors of Mind* (*Primary Mental Abilities*, University of Chicago Press, 1938), Thurstone claimed to have isolated seven fundamental components of intelligence: verbal comprehension, word fluency, number facility, spatial visualization, associative memory, perceptual speed and reasoning. Note that Thurstone's list does not include a factor for "general ability", whereas Hotelling's did. It is widely held that such a factor exists and is real; it is even given a name: g. One school of thought holds that this is what IQ tests actually measure.

However, the question facing all factor analyses is ultimately: "How can we be sure that the factors you identify are truly fundamental?" This is where another researcher, Herluf Strandskov, enters the story. He proposed a test of Thurstone's seven in a quest to answer exactly this.

Strandskov was a geneticist, specializing in human genetics. In 1963, he gave a course of lectures on this topic and I was one of his students. For reasons that I detail below, I am forced to rely on my memory of what he said about Thurstone's work in one of these lectures. Now memory (in general) is fallible, and this rule necessarily applies in the particular to mine, especially as the lecture was delivered almost half a century ago. So the usual caveats apply!

But Strandskov claimed to have said to Thurstone that if these seven components were *truly* fundamental, then they should demonstrate a clear pattern of heritability. The pair set out to test this idea. The heritability of a human trait can be estimated by comparing the extent to which it is shared by identical twins, who have a full genome in common, with the same figure as demonstrated by fraternal twins (whose genomes are identical in only 50% of sites). Strandskov, off the top of his head, nominated seven different possible components as statistical controls. A twin study was commenced, and the idea was to see it into publication in the technical literature.

It was to appear in two parts, and the first of these was undoubtedly written. It appeared as a report from the Psychometric Laboratory of the University of North Carolina, Chapel Hill, co-written by Thurstone himself, his wife Thelma and Strandskov. Its title was "A psychological study of twins. 1. Distributions of absolute twin differences for identical and fraternal twins".

This seems to be the only part of the study ever to see the light of day (in 1953). Technically it is not a publication in the strict sense of the word (it would not have been "peer-reviewed"); nevertheless some subsequent researchers have been able to access it, although I have not. A second part was listed by one of Thurstone's biographers[3]as scheduled for publication in the *American Journal of Human Genetics* in 1955 or shortly thereafter. This was "A Psychological Study of Twins: 2. Scores of One Hundred and Twenty-five Pairs of Twins on Fifty-Nine Tests".

But it never appeared! Which is why I am forced back on my memory of all those years ago. As I recollect things, of the 14 "factors" (Thurstone's 7 and Strandskov's 7 controls), only one showed a statistically significant heritability. This one was Thurstone's "spatial visualization", the ability to imagine in correct detail a 3-dimensional situation.[4] The next two in the order of statistical significance were both controls, but

[3]J. P. Guildford, writing for the (U. S.) *National Academy of Sciences* (1957).

[4]I can believe this. When I taught spherical trigonometry in Papua New Guinea, I found that the

even these were already in that realm where no heritability can reliably be claimed.

Possibly as a result of this, or perhaps for some other reason, nobody nowadays espouses Thurstone's 7-factor theory of mind. Although, it is clear that there are different and independent components of intelligence, there is still no general agreement as to what or how many they are. Nor is there any agreed methodology among the proponents of various schools of factor analysis. Child summed up the matter quite well (on p. 58):

> Uniqueness [as sought by Thurstone] occurs when the resulting model is singularly appropriate for the description of the underlying causes of a factor. If the same domain is investigated, identical configuration of the factors should be present. The problem facing Thurstone was to find a mathematical design which enabled him to discover the 'unique solution' . The problem still exists!

Appendix

The study of the interaction of heredity and intelligence has more recently fallen into disrepute. In 1969, the educational psychologist Arthur Jensen claimed influentially that African Americans were intellectually behind white Americans because of an intrinsic genetic factor. In this view he was supported by a number of others, most notably the psychologist Hans Eysenck. A fierce controversy ensued, in the course of which it was shown that the Jensen–Eysenck position held a number of statistical flaws. One, an obvious one, was that it depended on a scalar measurement, the IQ, when intelligence is clearly a vector. Jensen claimed in response that IQ tests measured a scalar factor g, which he saw as the only significant factor. He claimed that his analysis showed this to be strongly heritable. This is not really a soundly based claim. (Recall that in Hotelling's example it accounted for only 46.2% of the overall variance.) But another more powerful objection to his proposition emerged when it was discovered that the twin data he used came from the British author Sir Cyril Burt, whose results (to be as charitable as possible) were very dubious.

Manus Islanders in my classes (who came from what is likely to be a relativley homogeneous genetic base) all excelled in this skill. Such an observation is not, of course, a convincing proof (for a start the sample was too small for statistical significance), but it led me to think that the Thurstone-Strandskov finding might have some merit.