Directional Data

KEYWORDS:
Teaching;
Mean;
Variance;
Vectors.

Gerald Goodall - the Editor

Summary
This article displays some very counter-intuitive problems of "averages" and shows one area of statistics in which vector methods play a natural role.

◆NOTE BY THE EDITOR◆

THIS ARTICLE started life as a comprehension question in an A-level Mathematics examination (UK age 18). The examination candidates are presented with an article about an area of mathematics with which they are unlikely to be familiar and have to answer comprehension questions about it. Several colleagues who have seen this question have commented that it is of such interest, as an introduction to an important but unusual application area of statistics, that it is worthy of wider publication. We are therefore publishing it in *Teaching Statistics* in this spirit. We are not publishing the comprehension questions that were based on it.

The examination was in the MEI Structured Mathematics A-level scheme, administered by the Examinations Board now known as OCR which is based in Cambridge, England. The copyright is held by OCR. *Teaching Statistics* is very grateful to OCR for allowing us to publish the piece. The examiner who wrote it wishes to remain unidentified.

A very enjoyable book that considers many aspects of this application area with a wealth of beautiful examples is "Circular statistics in biology" by E Batschelet, published by Academic Press posthumously in 1981. There is also a monograph by the same author in 1965 published by the American Institute of Biological Sciences entitled "Statistical methods for the analysis of problems in animal orientation and certain biological rhythms"; despite its somewhat forbidding title, any difficulty in locating a copy is more than repaid by the very clear description of the basic concepts and, again, by some excellent examples.

The key reference to the theory of this area is "Statistics of directional data" by K V Mardia, published by Academic Press in 1972. This is a book at an advanced level, but the introductory pages (and the introductory sections to later chapters) are quite accessible and again give several good examples.

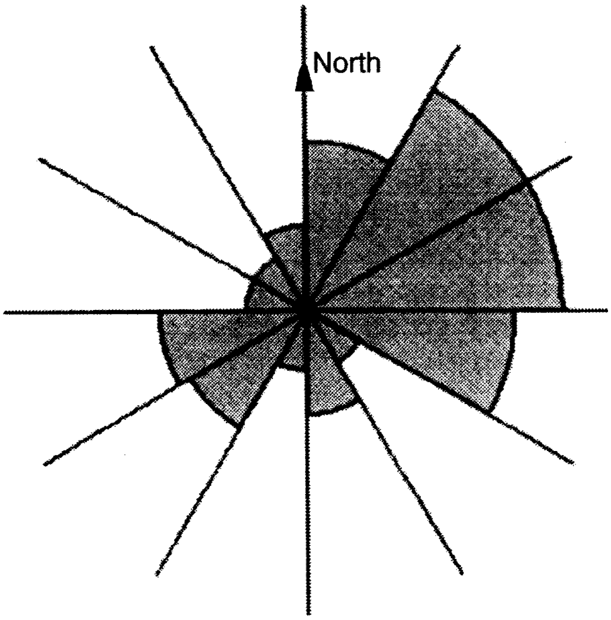
Some years ago, an experiment was carried out on the orientation of a certain type of turtle. 76 of the turtles, considered to be a random sample, were taken to an experimental site at sea and released. The initial directions of their travel were recorded. The data, measured in the conventional way as bearings clockwise from North, were as shown in Table 1. Figure 1 shows a diagrammatic representation of these data.

Inspection of the data shows that the majority of the turtles went in a direction roughly North-East ('homewards'), but a substantial minority went in almost exactly the opposite direction. Also there were a few which went in other directions. One interpretation is that the turtles have a fairly strong sense of orientation but some confuse 'forwards' with 'backwards'.

Table 1.

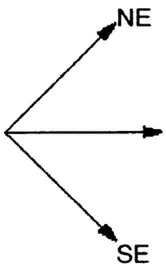
Bearing x (in degrees)	Frequency
$000 \le x < 030$	8
$030 \le x < 060$	18
$060 \le x < 090$	18
$090 \le x < 120$	12
$120 \le x < 150$	1
$150 \le x < 180$	3
$180 \le x < 210$	1
$210 \le x < 240$	5
$240 \le x < 270$	6
$270 \le x < 300$	1
$300 \le x < 330$	1
$330 \le x < 360$	2

Fig 1.



Data of this kind are called *directional data*. Describing and analysing such data statistically presents interesting problems. To simplify the situation, suppose that there are just three turtles and that their directions are NorthEast, East and SouthEast. This is shown diagrammatically by figure 2. It is clear that the ‘average direction’ is East.

Fig 2.



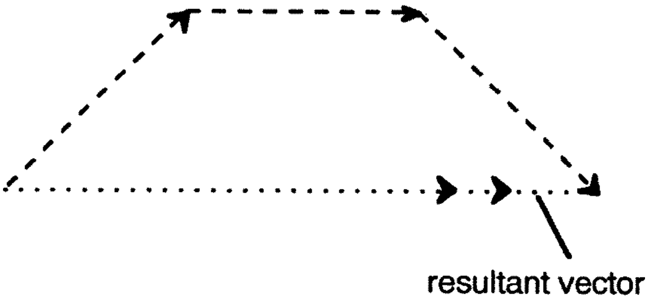
If we represent these three directions conventionally as bearings measured clockwise from North, they are 045, 090 and 135. Working out the mean in the usual way gives (045 + 090 + 135)/3 = 090, which agrees with our intuition from figure 2.

However, mathematicians usually measure angles anti-clockwise from East. If we do this, the directions of these three turtles (in degrees) are 0, 45 and 315. The mean direction is then (0 + 45 + 315)/3 = 120 degrees, which is fairly near North-West. This is clearly wrong!

The problem is that there is no natural ‘starting point’ for measuring directions; any direction can be chosen, arbitrarily, as the ‘base-line’. It is also completely arbitrary as to whether to work clockwise or anti-clockwise. However, this can be overcome by representing the directions as vectors.

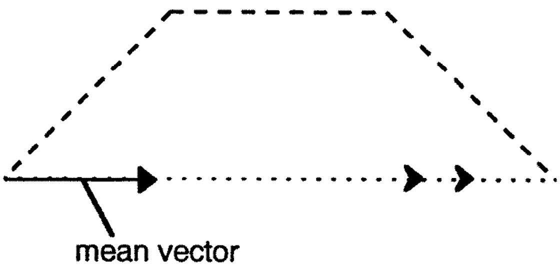
Referring again to figure 2 and representing the directions by vectors of unit length, addition of these vectors in the usual way for vector addition gives a vector pointing East. This is shown in figure 3. A vector obtained in this way, by adding together two or more other vectors, is called the *resultant vector* of the other vectors.

Fig 3.



This resultant vector was obtained by adding three other vectors, so if we divide its length by 3 but leave its direction unchanged we get a vector that can sensibly be regarded as the *mean vector* of the original three vectors. This is shown in figure 4.

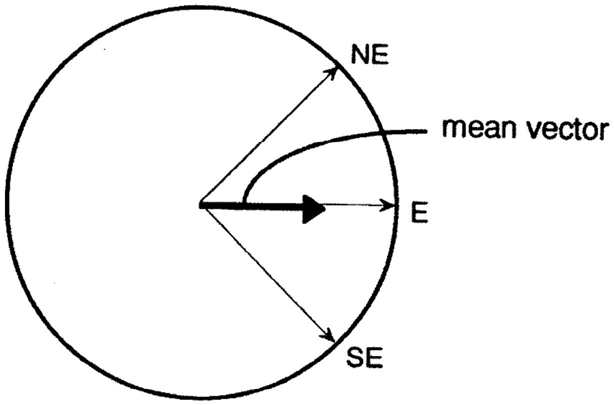
Fig 4



The direction of this mean vector would be recorded as 090 if measured in the conventional way as a bearing or 0 degrees if measured in the usual way as an angle. However, *both of these mean the same thing*-the direction is East. Thus the mean vector obtained in this way gives unambiguously the ‘average direction’.

Not only does the mean vector give the average direction, but also it gives a measure of how variable the directions are. To see how this is so, imagine that the three original unit vectors start from the centre of a circle of unit radius, with their tips on the circumference of the circle, as shown in figure 5. The mean vector is also drawn in figure 5. The length of this mean vector is 0.8047, so its tip falls short of the circumference of the circle by 0.1953.

Fig 5.



Now suppose the three original vectors were all in directions very near East – say with bearings 085, 090 and 095. The mean vector would point East and would have length very near 1 so that its tip would fall only very slightly short of the circumference of the circle.

If, on the other hand, the three original vectors still had an average direction of East but were much more spread out – say with bearings 005, 090 and 175 – the mean vector would point East but would have only a short length. Thus its tip would fall a long way short of the circumference of the circle.

In this way, the *shortfall* of the tip of the mean vector from the circumference of the circle gives a direct measure of how variable the original directions are.

These ideas can be applied to the turtle data in the original table. Taking the mid-points of the ranges in the usual way, the mean direction has bearing 065.5 if measured conventionally as a bearing, or 24.5 degrees if measured in the usual way as an angle, and its length is 0.484. This gives a summary of the diagrammatic display in figure 1 – the mean direction (bearing 065.5 is roughly North-East but the mean vector is fairly short indicating that there is a considerable amount of variability.

Reference

The turtle data are due to E Gould of Johns Hopkins School of Hygiene. They were first cited by M A Stephens (1969) in ‘Techniques for directional data’ (Technical report No 150, Department of Statistics, Stanford University) and have been cited by many authors subsequently.