The Statistician (2002)
51, *Part* 1, *pp.* 99–104

A new angle on the *t*-test

Graham R. Wood

Massey University, Palmerston North, New Zealand

and David J. Saville

AgResearch, Lincoln, New Zealand

[Received December 2000. Final revision December 2001]

Summary. R. A. Fisher's early work on linear models relied heavily on his knowledge of n-dimensional geometry. To illustrate this, we present an elementary, complete, modern day account of Fisher's geometric approach for the simple case of a paired samples t-test for a sample size of 3. A natural consequence of this approach is a surprisingly simple and explicit expression for the p-value.

Keywords: Linear algebra; Paired samples t-test; p-value; Vector geometry

1. Introduction

R. A. Fisher's early work on linear models relied heavily on his knowledge of n-dimensional geometry (Box (1978), pages 122–129). Unfortunately his geometric approach was not understood by most of his colleagues, partly because of his tendency to 'let too much be clear or obvious', to quote the words of his colleague Gosset, better known to statisticians as 'Student' (Box (1978), page 122). Because of this difficulty, Fisher derived an algebraic approach which was more universally accepted (Fisher (1925) and Box (1978), page 129). In this paper we provide an elementary, complete, modern day account of Fisher's geometric approach for the simple case of a paired samples t-test for a sample size of $n = 3$. Surprising insights into the theory of linear models can be gained by working through this simple example.

2. Paired samples *t*-test for $n = 3$

For a paired samples data set we use the heights of males M and females F in three mixed sex twin pairs of adult humans, from Saville and Wood (1996). In the first twin pair, John's height was 185 cm whereas Janet's height was 166 cm (the data are real, but all names are falsified). In the second twin pair, Alistair's height was 185.4 cm whereas Joanna's was 177.8 cm. In the third twin pair, Bill's height was 182.9 cm whereas Mary's was 160 cm. The three differences in height $(M - F)$ are 19.0, 7.6 and 22.9 cm. We treat these three differences as a sample of size $n = 3$ drawn independently from a single normally distributed population $N(\mu, \sigma^2)$, consisting of differences in height between the male and female in mixed sex twin pairs of adult humans.

Address for correspondence: David J. Saville, Statistics Group, AgResearch, PO Box 60, Lincoln, Canterbury, New Zealand.
E-mail: dave.saville@agresearch.co.nz

 0039–0526/02/51099

2.1. Geometric route

For an analysis of the above data using the geometric route, we immediately ask 'How do we derive the chance of obtaining a sample "as extreme as, or more extreme than" the above sample (19.0, 7.6, 22.9) under the two-sided test hypothesis H_0: $\mu = 0$?'. The key, discovered by Fisher, is to think of the data as a vector in three-dimensional space, the vector $(19.0, 7.6, 22.9)^T$. This leads naturally to a measure of 'extremeness'.

To understand how such data vectors behave in 3-space, consider Fig. 1 where in each picture we display the data vectors resulting from many repetitions of a study. Firstly, if $\mu = 0$ each sample comes from an uncorrelated trivariate normal distribution. The spherical symmetry of this distribution ensures that all directions in 3-space are equally likely (Fig. 1(a)). Secondly, if $\mu \neq 0$ the tips of the data vectors are distributed around the tip of the vector $(\mu, \mu, \mu)^T$, so the directions of the data vectors are closer to the direction $(1, 1, 1)^T$ (Figs 1(b) and 1(c)). Lastly, the larger μ is in relation to σ, the smaller the angle between each data vector and the direction $(1, 1, 1)^T$ (Fig. 1(c) *versus* 1(b)).

This suggests that we use the angle θ between our particular data vector $(19.0, 7.6, 22.9)^T$ and the direction $(1, 1, 1)^T$ as a measure of how extreme our sample is under hypothesis H_0: $\mu = 0$. If the angle is 'large' then our data are consistent with H_0: $\mu = 0$, whereas if the angle is 'small' then our data are consistent with the alternative hypothesis H_1: $\mu \neq 0$.

For our particular sample, we calculate θ by using a standard formula from linear algebra which involves the dot product of two vectors:

$$\cos(\theta) = \frac{\begin{pmatrix} 19.0 \\ 7.6 \\ 22.9 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}{\surd(19.0^2 + 7.6^2 + 22.9^2)\surd 3} = \frac{19.0 + 7.6 + 22.9}{\surd 2829.51} = 0.93057.$$

Thus the angle is $\theta = 21.476°$ (or 0.3748 radians), as shown in Fig. 2.

Is this angle large or small? To answer this, we calculate the probability p of observing a data vector with angle θ as small as, or smaller than, 21.476° under the null hypothesis H_0: $\mu = 0$. This is the probability of a data vector lying within the (infinite) double cone formed by rotating the data vector about the $(1, 1, 1)^T$ direction (Fig. 2). To evaluate p, we first calculate the probability under hypothesis H_0 that a data vector of a *fixed length*, such as $r_0 = \surd(19.0^2 + 7.6^2 + 22.9^2) = 30.711$, the length of the data vector, lies in the double cone.

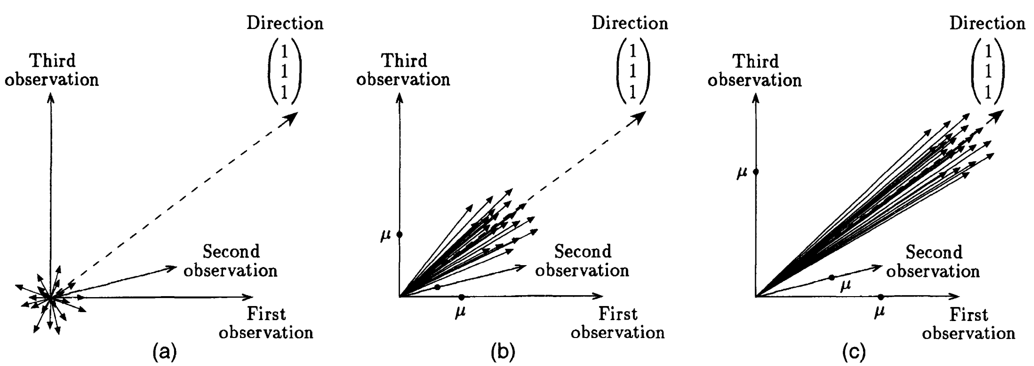

Fig. 1. Data vectors resulting from many repetitions of a study in the cases (a) $\mu = 0$, (b) $\mu \neq 0$ and (c) $\mu \neq 0$ with μ larger than in (b): in general, the angle between a data vector and the $(1, 1, 1)^T$ direction grows smaller as μ grows larger; this suggests the use of the angle as a test statistic for the hypothesis H_0: $\mu = 0$

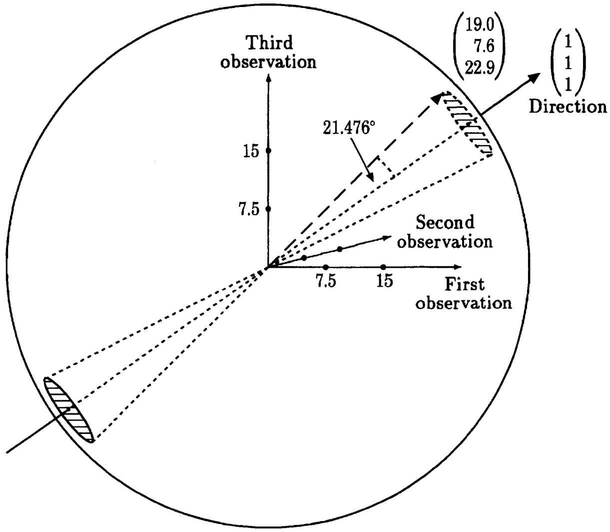

Fig. 2. Data vector $(19.0, 7.6, 22.9)^T$, associated sphere of radius $r_0 = \sqrt{(19.0^2 + 7.6^2 + 22.9^2)} = 30.711$ and double cone whose intersection with the sphere is shaded: the smaller the ratio of the surface area of the shaded intersection to that of the sphere, the stronger the evidence that μ is non-zero

From the spherical symmetry of the distribution of data vector directions under hypothesis H_0, this probability is

$$\frac{\text{surface area(shaded intersection in Fig. 2)}}{\text{surface area(sphere of radius } r_0)} = \frac{4\pi r_0^2\{1 - \cos(\theta)\}}{4\pi r_0^2} = 1 - \cos(\theta)$$

where the surface areas were obtained by elementary calculus, and where the answer is independent of the r_0-value that is used in the calculation. To complete the calculation of the p-value, we now integrate over all possible lengths of data vector r, weighting the above conditional probability by the probability that a data vector is of length r. This yields

$$p = \int_0^\infty \{1-\cos(\theta)\} f(r)\mathrm{d}r = \{1-\cos(\theta)\} \int_0^\infty f(r)\mathrm{d}r = 1-\cos(\theta) = 1-\cos(21.476) = 0.07$$

where $f(r)$ is the probability density function for the length of a data vector.

To conclude our analysis, we note that the p-value of 0.07 is larger than 0.05, so our data are 'not unusual' under hypothesis H_0: $\mu = 0$ if we use a 0.05 cut-off point as our criterion. We conclude that we do not have strong evidence of a non-zero true mean difference in height between the male and female in mixed sex twin pairs.

2.2. *Link between geometric and traditional routes*

In the preceding subsection we completed the paired samples t-test of H_0: $\mu = 0$ *without mentioning* the t-value. How does the geometric route to the p-value relate to the more traditional route?

Now the familiar t-test statistic for paired samples is

$$t = \frac{\bar{y}}{s/\sqrt{n}} = \frac{\bar{y}\sqrt{n}}{s} = \frac{\bar{y}\sqrt{n}}{\sqrt{\left\{\sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)\right\}}} = \frac{16.5\sqrt{3}}{\sqrt{[\{2.5^2 + (-8.9)^2 + 6.4^2\}/2]}} = 3.595$$

where $y_1, \ldots, y_n$ are the data values (differences between the paired sample values), $\bar{y}$ is the sample mean, s is the sample standard deviation and n is the sample size.

To relate this to θ, we display in Fig. 3 a right-angled triangle which was implicit in Fig. 2. The triangle is obtained by projecting the data vector onto the $(1, 1, 1)^T$ direction and represents the vector sum

$$\begin{pmatrix} 19.0 \\ 7.6 \\ 22.9 \end{pmatrix} = \begin{pmatrix} 16.5 \\ 16.5 \\ 16.5 \end{pmatrix} + \begin{pmatrix} 2.5 \\ -8.9 \\ 6.4 \end{pmatrix}$$

with sides of length $A = \surd(3 \times 16.5^2) = 16.5\surd 3$ and $B = \surd\{2.5^2 + (-8.9)^2 + 6.4^2\} = \surd 126.42$. Thus

$$t = \frac{16.5\surd 3}{\surd[\{2.5^2 + (-8.9)^2 + 6.4^2\}/2]} = \frac{A}{B/\surd 2} = \surd 2 \cot(\theta) = \surd 2 \cot(21.476) = 3.595,$$

establishing the relationship $t = \surd 2 \cot(\theta)$ between t and θ for the case of a sample of size 3. As an aside, note that the sample mean and standard deviation arise naturally within the triangle, with $A = \bar{y}\surd 3$ and $B = s\surd 2$.

2.3. *Transformation from θ to t*

In the above development, the angle θ has been implicitly restricted to the range 0–90° for ease of explanation. This is entirely adequate for two-sided tests, but not for one-sided tests. More generally, θ is defined to be the angle between the data vector and the *positive* sense of the direction $(1, 1, 1)^T$, so $0° \leqslant \theta \leqslant 180°$.

If we also express θ in radians $(0 \leqslant \theta \leqslant \pi)$, we can write the probability density function for θ in its simpler form as $\sin(\theta)/2$, since

$$\int_0^{\theta} \frac{\sin(u)}{2} \, du = \frac{1 - \cos(\theta)}{2}.$$

In summary, the distribution of θ (in radians) under hypothesis H_0: $\mu = 0$ (Fig. 4(a)) is converted via the transformation $t = \surd 2 \cot(\theta)$ (Fig. 4(b)) to the t_2-distribution (Fig. 4(c)). Note that values of θ close to 0 or π are transformed to large positive and negative values of t respectively, whereas values of θ close to $\pi/2$ are transformed to small values of t.

2.4. *Case H_0: $\mu = \mu_0$*

We note that the more general case H_0: $\mu = \mu_0$ $(\neq 0)$ can be rewritten in the current form as H_0: $\mu - \mu_0 = 0$, with data analysis carried out using the transformed variable $y - \mu_0$.

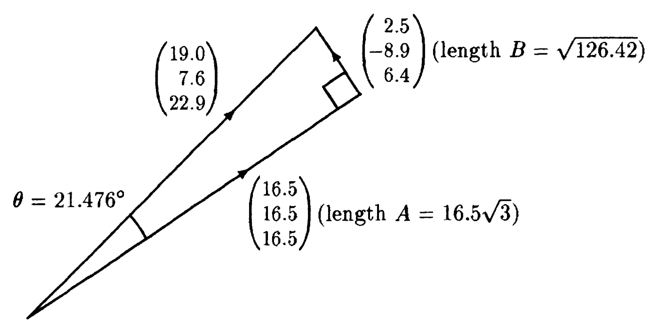

Fig. 3. Vector decomposition of the data vector: the familiar t-statistic is equal to $\surd 2 \cot(\theta)$

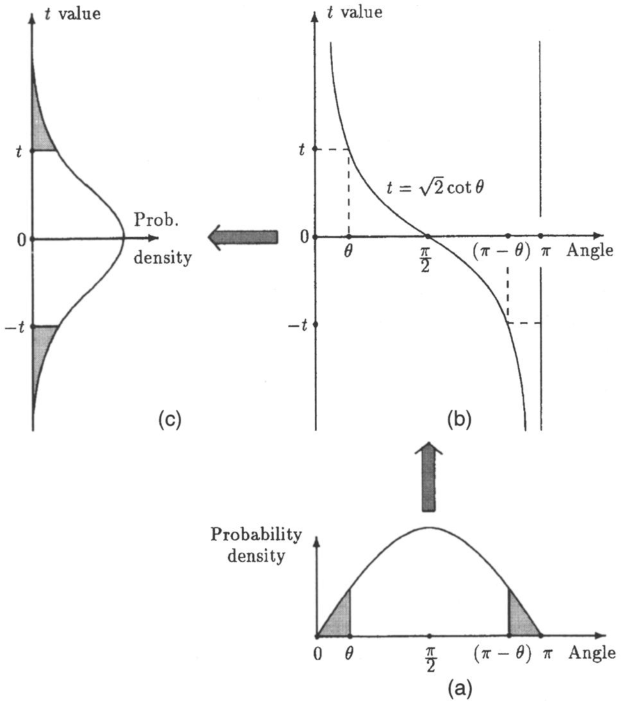

Fig. 4. Transformation $t = \sqrt{2}\cot(\theta)$ (shown in (b)) which links the probability density function for θ to that for t, shown in (a) and (c) respectively: the shaded areas indicate values that are more extreme than θ or t

3. Discussion

In this paper we have conveyed the flavour of Fisher's geometric intuition and have related his 'angle' ideas to the more traditional ideas involving the t-distribution. These angle ideas extend to the paired samples t-test for a general sample size, and to any single degree of freedom hypothesis test in the general linear model. Examples of this extension to the paired samples t-test for a general sample size, the independent samples t-test, analysis-of-variance single degree of freedom contrast tests and the test of slope in a simple regression are given in appendix D of Saville and Wood (1996). In all these cases, the surprising result is that there is a direct computational formula for the p-value which does not rely on an approximation to a reference t-distribution.

The approach outlined in this paper has been partly, but not wholly, described by other workers, i.e. the pieces of the 'jigsaw' are all present in the literature, but nowhere, to the knowledge of the authors, has the jigsaw been completed. In a book co-authored by Fisher's son-in-law, George Box, the p-value is given as the ratio of surface areas shown in Fig. 2, following a reference to the spherical symmetry of the distribution of data vector directions (Box *et al.* (1978), page 202). Box *et al.* (1978) pointed out that a small angle is associated with a large t-value and a small p, and described the t-value as a measure of the size of the angle θ. Together with Fisher's biographer and daughter, Joan Fisher Box (Box (1978), pages 126–127), Box *et al.* (1978) implied that the angle is the basic quantity, with the t-value being simply a means to an end. Heiberger (1989), pages 150–168, also discussed the relationship between θ and t for a range of linear model tests. Chance (1986) gave an explicit expression for the p-value for the case of the correlation coefficient which involves a ratio of volumes. In Saville and Wood (1996), we elucidated the relationships between θ, t, F, the correlation coefficient $r = \cos(\theta)$ and the p-value for a range of linear model tests; the contribution of the current paper is to present concisely a complete analysis for a single example and to place the work in the proper historical context.

Acknowledgements

P. A. Carusi is thanked for stimulating this work by asking how our geometric approach (Saville and Wood, 1991) related to the approach taken in, for example, Chance (1986) and Box *et al.* (1978). Fellow statisticians Michael Ryan, Karen Baird, Lesley Hunt, Andrew Wallace and K. Govindaraju, agricultural scientists Gilbert Wells, Shona Lamoureaux and John Kean, and the journal referees are thanked for helpful suggestions. Aaron Knight is thanked for assisting with the graphic art.

References

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978) *Statistics for Experimenters: an Introduction to Design, Data Analysis and Model Building.* New York: Wiley.
Box, J. F. (1978) *R.A. Fisher, the Life of a Scientist.* New York: Wiley.
Chance, W. A. (1986) A geometric derivation of the distribution of the correlation coefficient $|r|$ when $\rho = 0$. *Am. Math. Mnthly*, **93**, 94–98.
Fisher, R. A. (1925) Applications of "Student's" distribution. *Metron*, **5**, 90–104.
Heiberger, R. M. (1989) *Computation for the Analysis of Designed Experiments.* New York: Wiley.
Saville, D. J. and Wood, G. R. (1991) *Statistical Methods: the Geometric Approach.* New York: Springer.
———(1996) *Statistical Methods: a Geometric Primer.* New York: Springer.