Anything but simple: Inappropriate use of Euclidean distance in Govinder et al. (2013)

AUTHOR:
Victor M.H. Borden[1]

AFFILIATION:
[1]Department of Educational Leadership and Policy Studies, Indiana University, Bloomington, Indiana, USA

CORRESPONDENCE TO:
Victor Borden

EMAIL:
vborden@iu.edu

POSTAL ADDRESS:
Department of Educational Leadership and Policy Studies, Indiana University, 201 N. Rose Avenue, Bloomington, IN 47405-1006, USA

KEYWORDS:
Equity Index; higher education; South Africa; benchmarking

HOW TO CITE:
Borden VMH. Anything but simple: Inappropriate use of Euclidean distance in Govinder et al. (2013). S Afr J Sci. 2014;110(5/6), Art. #a0065, 4 pages. http://dx.doi.org/10.1590/sajs.2014/a0065

The 'Equity Index' (EI) introduced by Govinder et al.[1] has stimulated critiques addressing a variety of flaws in the use of this allegedly 'simple and objective' measure of racial and gender equity among South African higher education institutions. Dunne[2] noted that the use of a mathematical formula and the resultant numerical result provides a false sense of validity and precision. He further described in great technical detail why measures of distance are not as simple as they may seem when portrayed, for explanatory purposes, as the distance between points in a two-dimensional space. Dunne also addresses several issues of substantive validity including the stochastic nature of social measures for which dynamic probabilistic models are required as compared to the mathematical models that serve physical phenomena like measuring distance between objects in space. Moultrie and Dorrington[3] extend this critique, examining other mathematical (double counting) and conceptual (suitability of benchmark) problems.

As a long-time institutional research practitioner within the US context, I was intrigued by the publication of the index and the ensuing critiques as they touch upon the long-standing institutional research practices of peer institution benchmarking.[4-6] Because of the diversity of the US higher education landscape, with over 7000 post-secondary institutions ranging from for-profit, single programme vocational institutions and 2-year community colleges to 4-year regional and comprehensive universities and both public and private research universities, it is not common for us to think of a single measure that can be applied equally to all institutions, or even to those that are internationally competitive for students and staff. Because of this complexity, we are well versed in comparing institutions across a variety of measures and dimensions, including the demographic and academic profile of students, the mix of academic programmes, the types of instructional and non-instructional staff, and revenue sources and expenditure targets. One thing we have learned from this vast experience is that there is no such thing as either a simple or objective measure of institutions in relation to a target (whether that be another institution or a regional or national benchmark).

In the remainder of this critique, I will illustrate the lack of reliability (and therefore questionable validity) of employing a Euclidean distance measure on the concatenated distribution of two sets of proportions (race and gender). Rather than explore the mathematical and technical dimensions of these problems, I will illustrate how the comparison of the 23 South African higher education institutions changes depending on what type of distance measure is used and whether it is used on race and gender separately or combined.

When comparing the 'position' of an institution relative to other institutions or to criterion benchmarks like the national representation among racial and gender groups, one must take into account the scale characteristics of the measurement variables (nominal, ordinal, interval, ratio), as well as the statistical relationship (association) among the variables. If one is simply considering race and gender as distinct variables, then it may be suitable to describe these as independent measures (the likelihood of being male or female is not contingent, at least conceptually, on the racial group). However, when the values of a proportional representation variable are portrayed as the values upon which comparisons are based, then, as Moultrie and Dorrington pointed out, there is redundancy. That is, the percentage of males is linearly dependent on the percentage of females (percentage males = 100 – percentage females). Thus, the values of the variable gender have only one degree of freedom. Moreover, as race entails four categories and gender two, if we assume equal probability of each category, the race factor has three times the weight in the characterisation of the position (because race is four groups, there are three degrees of freedom, compared to one for gender). However, race is not uniformly distributed (that is, the general probability for each category is not one divided by the number of categories), so one must take into account the non-linear qualities of proportions across the range values. More prosaically, a 5% point difference has different substantive meaning when an event is rare (e.g. 5%), or more common (e.g. 60%).

There is a wide variety of ways to calculate similarity or difference for use in a positioning (nearest neighbour) analysis. Even if one would like to use a Euclidean-based measure, there are several to choose from. Govinder et al. use the 'RSSD' version, that is, the root of the sum of squared differences. If the variables are on notably different scales in terms of variation, it is advisable to first transform the measures to their standardised form (value minus mean, divided by standard deviation). When using percentages, the Chord form of Euclidean distance is recommended, where the values are first subject to a square root transformation. There are several derivatives of the Euclidean form, such as a City Block metric and Minkowski metric that vary the root to which the difference between coordinate points is raised. In addition to Euclidean-based measures, there are correlation-based distance measures (Pearson and Spearman) and the Mahalanobis measure, which takes into account both Euclidean distance and covariance among the variables.

Tables 1 and 2 demonstrate how the calculated distance value and the rank of the 23 South African higher education institutions change depending on which proximity measure is used to calculate the distance from the national benchmark. For these tables, the benchmarks were taken from the Govinder et al. article and the proportions of enrolled students from the Department of Higher Education and Training document, *Statistics on Post-School Education and Training in South Africa: 2011*[7]. The first three proximity measures included in Table 1 are three forms of the Euclidean distance: the RSSD version used by Govinder et al., one based on standardised values for each proportion, and the 'Chord' version, which is based on a square root transformation of the original values. In addition, the table shows the results using the Mahalanobis metric, which incorporates the covariance between the variables, and a measure based on the Pearson correlation, which has been reversed (Pearson values range from 1 for the most similar to 0 for the least similar, so the calculated value is subtracted from 1) and multiplied by 1000 to represent the value in integer digits. The rightmost columns of the tables show the rankings among the 23 institutions of the corresponding calculated values.

Table 1 exhibits these various distance measures for the combined race and gender proportions as employed by Govinder et al.[1] The reader is reminded that there are several technical reasons why it is not appropriate to combine these proportions into a single estimation of distance, as noted in the critiques of Dunne[2] and Moultrie and Dorrington[3]. Some of the ramifications for the inappropriateness of doing so are manifest in the variation of calculated distance values and rank in these tables. For example, the Central University of Technology, ranked 2nd using the RSSD calculation, is ranked 11th using the Mahalanobis measure. Durban University of Technology varies considerably by the four measures, as high as 5th using the RSSD and as low as 17th using the Mahalanobis metric.

Table 2 uses the same five measures on the four categories of race. While not suggesting that examining race alone establishes evidence of equity, the benchmarking of distance from the national norms is a cleaner measurement concept than when incorporating race and gender into a single measure. Although the rankings for race alone are not as varied as they are for race and gender combined, they still vary considerably. For example, University of Johannesburg, which is ranked 1st by four measures, is ranked 10th using the Pearson correlation measure. It is also interesting to note that the Chord version of the Euclidean measure, which is generally recommended over RSSD for percentage measures, varies considerably from the RSSD measure.

Establishing equity

Although it is not without controversy, it is instructive to consider how equity is established in other, long-standing methodologies. For example, the US Department of Labor’s Office of Federal Contract Compliance, has required since the early 1970s that organisations and businesses that obtain federal contracts establish the equity in both hiring and compensation of their workforce. The compliance requirements revolve around ‘labour-market availability’ within job groups that are defined

Table 1: Comparison of five distance measures using both race and gender percentages benchmarked against national norms

Institution	Calculated distance value					Rank of distance value				
	Euclidean			Mahal-anobis	Pearson correlation	Euclidean			Mahal-anobis	Pearson correlation
	RSSD	Std	Chord			RSSD	Std	Chord		
University of Johannesburg	9	112	15	25	9	1	2	1	1	3
Central University of Technology, Free State	12	144	17	28	7	2	4	2	11	2
Tshwane University of Technology	17	106	25	26	6	3	1	5	6	1
University of South Africa	17	277	19	27	35	4	16	3	8	12
Durban University of Technology	18	232	33	30	34	5	13	12	17	11
Nelson Mandela Metropolitan University	22	128	22	27	45	6	3	4	9	13
University of Fort Hare	23	192	27	25	11	7	8	7	3	4
Vaal University of Technology	24	207	29	29	14	8	10	8	14	9
University of the Free State	25	248	26	27	73	9	14	6	10	14
University of Limpopo	26	154	37	25	12	10	6	13	2	5
University of Witwatersrand	28	219	32	28	76	11	11	11	12	15
Mangosuthu University of Technology	28	152	45	26	13	12	5	19	7	7
Walter Sisulu University	28	193	42	25	14	13	9	17	4	8
University of Venda	28	159	48	26	12	14	7	22	5	6
University of KwaZulu-Natal	31	426	40	38	113	15	22	16	22	16
North-West University	31	417	30	31	115	16	20	9	19	17
Cape Peninsula University of Technology	33	225	31	32	118	17	12	10	20	18
University of Zululand	33	394	44	29	30	18	19	18	16	10
University of Pretoria	41	283	39	29	209	19	18	15	15	19
Rhodes University	41	281	37	29	210	20	17	14	13	20
University of Western Cape	53	431	47	42	379	21	23	21	23	21
University of Cape Town	54	275	46	31	426	22	15	20	18	22
University of Stellenbosch	84	419	71	38	915	23	21	23	21	23

Source: Republic of South Africa Department of Higher Education and Training[7]

Table 2: Comparison of five distance measures using only race percentages benchmarked against national norms

Institution	Calculated distance value					Rank of distance value				
	Euclidean			Mahal-anobis	Pearson correlation	Euclidean			Mahal-anobis	Pearson correlation
	RSSD	Std	Chord			RSSD	Std	Chord		
University of Johannesburg	8	65	15	25	8.1	1	1	1	1	10
Central University of Technology, Free State	11	73	17	26	1.6	2	2	3	3	1
University of South Africa	11	90	16	25	11.8	3	3	2	2	11
Durban University of Technology	17	222	32	29	39.5	4	18	12	19	12
Tshwane University of Technology	17	105	25	26	2.6	5	4	6	6	2
University of Fort Hare	21	117	27	26	3.0	6	6	7	8	3
Vaal University of Technology	22	121	28	26	3.1	7	7	9	9	4
Nelson Mandela Metropolitan University	22	116	22	26	45.3	8	5	4	4	13
University of the Free State	22	124	24	26	69.1	9	8	5	5	14
North-West University	24	135	27	26	80.7	10	9	8	7	16
University of Limpopo	26	139	37	26	4.8	11	10	14	10	5
University of Zululand	27	145	42	27	5.2	12	11	17	11	9
Walter Sisulu University	27	146	42	27	5.2	13	12	18	12	8
University of Witwatersrand	28	211	32	28	79.2	14	17	11	16	15
Mangosuthu University of Technology	28	150	45	27	5.0	15	13	19	13	7
University of Venda	28	151	48	27	5.0	16	14	22	14	6
University of KwaZulu-Natal	29	382	40	38	123.0	17	22	16	22	17
Cape Peninsula University of Technology	33	223	31	30	136.4	18	19	10	20	18
University of Pretoria	39	211	39	28	250.2	19	16	15	17	19
Rhodes University	39	207	37	27	250.6	20	15	13	15	20
University of Western Cape	52	371	46	40	479.5	21	21	21	23	21
University of Cape Town	54	275	46	28	584.7	22	20	20	18	22
University of Stellenbosch	84	418	71	34	1168.8	23	23	23	21	23

Source: Republic of South Africa Department of Higher Education and Training[7]

according to the wages, job duties and responsibilities, and training requirements. Specifically, the requirements (http://www.dol.gov/ofccp/scaap.htm) note[8]:

> *…federal contractors must conduct availability analyses to determine the percentage of women and minorities who have the skills required to perform the jobs within each job group… Availability involves calculation of minorities and women who are 'available' to work in the job from both external sources (i.e., hired from outside the company) and internal sources (e.g., transfer or promotion of existing employee in the company)…For calculating 'external' availability, you want to consider who is qualified for the job within 'the reasonable recruitment area' for that job. The 'reasonable recruitment area' represents the area from which a contractor usually seeks or reasonably could seek workers for a particular job group.*

Assessing equity in academic programmes can be considered as analogous. To be admitted to an academic programme, students must meet certain basic requirements, such as having completed a secondary education credential and having basic skills suited to a specific programme of study (for example, higher order math skills for engineering and higher order writing skills for communications). Students must also live within commuting distance (except perhaps for UNISA). Comparing proportions of women and racial groups enrolled at a particular university to a generic national benchmark masks all of the availability issues, which are at the root of establishing equity. Throughout my 30 years of experience in using evidence and analysis to address educational access issues, I have found that it is far more constructive to confront directly and as complexly as possible the root causes of inequity, such as those revealed through the many aspects of 'availability'. Conversely, reducing to a single measure such complex phenomena tends to shift attention away from the root causes and can be used by various groups and individuals to absolve the responsibility that we all share in addressing such issues. Establishing equity is anything but simple.

References

1. Govinder KS, Makgoba MW. An Equity Index for South Africa. S Afr J Sci. 2013;109(5/6), Art. #a0020, 2 pages. http://dx.doi.org/10.1590/sajs.2013/a0020
2. Dunne T. Mathematical errors, smoke and mirrors in pursuit of an illusion: Comments on Govinder et al. (2013). S Afr J Sci. 2014;110(1/2), Art. #a0047, 6 pages. http://dx.doi.org/10.1590/sajs.2014/a0047

3. Moultrie TA, Dorrington RE. Flaws in the approach and application of the Equity Index: Comments on Govinder et al. (2013). S Afr J Sci. 2014;110(1/2), Art. #a0049, 5 pages. http://dx.doi.org/10.1590/sajs.2014/a0049

4. James GW. Developing institutional comparisons. In: Howard RD, McLaughlin GW, Knight WE, editors. The handbook of instituitonal research. San Francisco, CA: Jossey-Bass; 2012. p. 644–655.

5. Terinzini PT, Hartmark L, L'Orange Jr. WG, Shirley RC. A conceptual and methodological approach to the identification of peer institutions. Res High Educ. 1980;12(4):347–364.

6. McCormick AC, Zhao CM. Rethinking and reframing the Carnegie classification. Change. 2005;37(5):51–57.

7. Department of Higher Education and Training, Republic of South Africa. Statistics on post-school education and training in South Africa. Pretoria: Department of Higher Education and Training; 2011.

8. US Department of Labor. Office of Federal Contract Compliance Programs [homepage on the Internet]. No date [cited 2014 May 09]. Available from: http://www.dol.gov/ofccp/scaap.htm