IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 11, Issue 1 Ver. V (Jan - Feb. 2015), PP 35-41
www.iosrjournals.org

Estimation of Metric Distance of Observations from the Average Score

Oyeka I.C.A. and Nwankwo Chike H.
Department of Statistics, Nnamdi Azikiwe University, Awka. Nigeria

Abstract: *Under the assumption that there exist a subject whose performance, observed value or score may be regarded as an average or standard value or score to be used as reference, standard, average value or score for comparison with the performance or score of other subjects on a condition in a population, this paper developed what is here referred to as metric distance measured in terms of the number of subjects to determine how many subjects' performance or scores a randomly selected subject is above (better, worse), the same or lower (worse, better), than their own relative to the standard with respect to a condition of research interest in a population. The farther away the so-called metric distance, expressed in terms of number of subjects, a randomly selected subject is from the average or standard subjects' performance value or score, the better or less serious (the worse or more serious) than that subjects' condition relative to, that is in comparison with the average, reference or standard subjects' condition relative to, the average, reference or standard subjects' condition, that is in comparison with the conditions of some other subjects in the study population. Test statistics are developed for use in testing the statistical significance of the metric distance of a randomly selected subject as well as the statistical significance of the difference between the metric distances of any two randomly selected subjects in the population. The proposed method is illustrated with some sample data.*
Keywords: *Metric Distance, Score, Relative, Standard value distance, Measure.*

I. Introduction

Sometimes a researcher or policy implementer may have some data on the distribution of a population by some conditions present in such a population. These conditions may include diseases, various types of illnesses, injury levels, viral loads, poverty levels, educational status, employment levels etc by gender of subject or some other demographic characteristics.

In each of these and other such conditions a subject's situation may range from high (good, least serious) through average (fair, moderate) to low (bad, most serious).

The researcher or policy implementers problem may be to determine how many subjects or patients who are in a situation whose condition is considered above (better, higher, less serious) than the situation of a subject whose condition is considered average (fair, normal, moderate) as well as to also determine the number of subject in a situation whose condition is considered lower (worse, more serious) than that of a subject whose condition is considered average (moderate, fair, normal) in the population being studied [Siegel, 1956; Mahalanobis, (1936); Mclachlan (1992), Dodge (2003)]. The average subject may here be considered as the index subject in the population that may be used as reference or standard subject in the population as a guide in the formulation and implementation of any interventionists measures aimed at the management of the condition of interest in the population when resources or opportunities are limited or scarce.

This paper proposes to develop a method for the estimation of the so-called Metric Distance of Observations from the Average Score in a population as a measure of how many subjects' conditions, the condition of a randomly selected subject is better (less serious, more favourable) than in comparison with, that is relative to the condition of an average subject in a population as well as how many subjects' conditions, the condition of a randomly selected subject is worse (more serious, less favourable) than, when compared with, that is relative to the condition of the average, moderate or standard subject in the same population.
Test statistics are also developed for testing any desired null hypotheses.

II. The Proposed Method

Let x_i be the observation or score by the ith randomly selected subject in a random sample of size n drawn from population X, for $i = 1,2,\dots n,$ where population X may be measurements on as low as the ordinal scale and need not be continuous or even numeric. Research interest is to determine the metric distance of a given observation from the average score or value in the sampled population, where the average value or score may be either the population mean or population median of the sampled population. Here the 'metric distance' of a given observation from the average score or value is measured or defined as the number of subjects in the sampled population whose scores are lower (or higher) than the score or observation by the index subject in question relative to that, in comparison with, and different from the average score in the population.

Now to estimate the metric distance of the given subject from the average score in the population, let x_l and x_j be respectively the observations on the lth and jth subjects randomly drawn from population X that are each different, that is are not the same observations as a well determined average score or value of observations in the sampled population, for $l, j = 1,2, \dots n - t$, where $t (0 \le t \le n)$, is the total number of observations or scores in the random sample that have the same value that is that are tied with the average score or value, including the average score itself.

Note that if there are no observations or scores in the given random sample that are exactly equal to (the same as) the average score value, then $t = 0$.

Now let

$$U_{lj} = \begin{cases} 1, \text{if } x_j \text{ is greater (better, higher) than } x_l; \text{or } x_j > x_l \\ 0, \text{if } x_j \text{ is equal to (the same as) } x_l; \text{or } x_j = x_l \\ -1, \text{if } x_j \text{ is smaller (worse, lower)than } x_l, \text{or } x_j < x_l \end{cases}$$

For $l, j = 1,2, \dots, \quad n - t, \quad l \neq j$

Note that actual observations on the scores by subject rather than the deviations of these scores from the average score are used in the construction of equation (1). This approach is sufficient because if the actual score by a given subject is, say, higher (better, larger) than the score by some other subject, then this same pattern of relationship would also exist between the deviations of these scores from some corresponding measure of central tendency such as the average score by subjects in the sampled population.

Let

$$\Pi_j^{\ +} = P(U_{lj} = 1); \quad \Pi_j^0 - P(U_{lj} = 0); \ \Pi_j^- = P(U_{lj} = -1) \qquad - \quad - \quad - \quad 2$$

Where

$$\Pi_j^+ + \Pi_j^0 + \Pi_j^- = 1 \qquad - \quad - \quad - \quad 3$$

Let

$$W_j = \sum_{\substack{l=1 \\ l \neq j}}^{n-t} ulj \qquad - \quad - \quad - \quad 4$$

Now the expected value and variance of U_{lj} are respectively

$$E(U_{lj}) = \Pi_j^+ - \Pi_j^- \ and \ \mathrm{var}(U_{lj}) = \Pi_j^+ - \Pi_j^- - (\Pi_j^+ - \Pi_j^-)^2 \qquad - \ - \ - \ 5$$

Similarly, the expected value and variance of W_j are respectively

the score by the jth subject, relative to, that is n comparison with

$$\sum_{\substack{l=1 \\ l \neq j}}^{n-t} E(U_{lj}) = (n-t-1)(\Pi_j^+ - \Pi_j^-); Var(W_j) = \sum_{\substack{l=1 \\ l \neq j}}^{n-t} Var(U_{lj}) = (n-t-1)(\Pi_j^+ + \Pi_j^- (\Pi_j^+ - \Pi_j^-)^2) \qquad - \ - \ - \ 6$$

Now Π_j^+, Π_j^0 and Π_j^- are respectively the probabilities that or the proportions of subjects whose scores are lower (worse, smaller), equal to (the same as) or greater (better, higher) than the score by the jth subject, relative to, that is in comparison with the average score in the sampled population. Their sample estimates are respectively

$$\Pi_j^+ = \frac{f_j^+}{n-t-1} \ ; \ \Pi_j^0 = \frac{f_j^0}{n-t-1} \ ; \ \Pi_j^- = \frac{f_j^-}{n-t-1} \qquad - \quad - \quad - \quad 7$$

Where f_j^+, f_j^0 *and* f_j^- are respectively the total number of subjects whose scores, the observation or scores by the jth subject is found to be greater (better, higher), the same as (equal to) or smaller (worse, lower) than, relative to, that is in comparison with the average score of the given random sample regarded as the fulcrum, reference or standard score in the sample. In other words f_j^+, f_j^0 *and* f_j^- are respectively the total numbers of 1s, 0s and -1s in U_{lj} for all $l = 1,2,..n-t,$ and for some $j = 1,2,..n \ ; \ l \neq j$.

Note that $f_j^+ = (n-t-1)\hat{\Pi}_j^+, f_j^0 = (n-t-1)\hat{\Pi}_j^0$, and $f_j^- = (n-t-1). \hat{\Pi}_j^-$ may be actually interpreted respectively as the total number of subjects in the sample, the jth subjects score is higher than their own score relative to the average score; the number of subjects whose scores are the same as the jth subject score, relative to the average score; and the number of subjects whose scores are lower than the jth subject score relative to the average score in the sample, which may provide estimates of the metric distances of the jth subject score above, the same as and below the average score in the population, for some $j = 1,2,..n.$

Note also that for the jth subject $\Pi_j^+ - \Pi_j^-$ is the gap in the proportion of subjects in the sampled population that that jth subject's observation or score is above, greater (better, higher) less the proportion of subjects that the jth subject's observation or score is below, lower (worse, smaller) than, in relation to, that is when juxtaposed against the relative position of the average population value or observation, and may be regarded and used as a measure of the gap in the metric distance of the observation on the jth subject from the average score in the sampled population.

Its sample estimate is

$$MD(j) = \hat{\Pi}_j^+ - \hat{\Pi}_j^- = \frac{W_j}{n-t-1} = \frac{f_j^+ - f_j^-}{n-t-1} \qquad - \quad - \quad - \qquad 8$$

The corresponding estimated sample variance is

$$Var\ \left(MD(j)\right) = Var\,\hat{\Pi}_j^+ - \hat{\Pi}_j^- = \frac{Var\left(W_j\right)}{\left(n-t-1\right)^2} = \frac{\hat{\Pi}_j^+ + \hat{\Pi}_j^-\left(\hat{\Pi}_j^+ - \hat{\Pi}_j^-\right)^2}{n-t-1} \qquad - \ - \ - \quad 9$$

A null hypothesis that may be of research interest would be that the metric distance of some jth subject is some specified value, that is that for a given subject, the jth subject say, the proportion of subjects that subject's score is higher (better, greater) than their scores is at least some constant, θ_0 say, larger than the proportion of subjects that subject's score is lower (worse, smaller) than their own scores in comparison with the relative position of some specified average score in the sampled population. That is the null hypothesis.

$$H_0 : \hat{\Pi}_j^+ - \hat{\Pi}_j^- \geq \theta_0 \quad vs \quad H_1 : \hat{\Pi}_j^+ - \hat{\Pi}_j^- < \theta_0, \left(-1 \leq \theta_0 \leq 1\right) \qquad - \quad - \quad - \qquad 10$$

This null hypothesis is tested using the test statistic

$$\chi^2 = \frac{\left(W_j - (n-t-1)\theta_0\right)^2}{Var\left(W_j\right)} = \frac{\left(n-t-1\right)\left(\left(\hat{\Pi}_j^+ - \hat{\Pi}_j^-\right) - \theta_0\right)^2}{\hat{\Pi}_j^+ + \hat{\Pi}_j^- - \left(\hat{\Pi}_j^+ + \hat{\Pi}_j^-\right)^2} \qquad - \quad - \quad - \qquad 11$$

Which under the null hypothesis of equation 10 Ho has approximately the chi-square distribution with 1 degree of freedom for sufficiently large sample size n.

The null hypothesis Ho of equation 10 is rejected at the $\propto -level$ of significance if $\chi^2 \geq \chi^2_{(1-\alpha\ ;\ 1)}$ - - - 12

Otherwise H_0 is accepted.

A problem that may also be of research interest may be to determine whether metric distances of the jth and lth subject from the average value or score of the subjects in the sampled population are the same. To answer this question we note that the difference between the jth and lth subject's metric distances in terms of the number of subject's these subjects scores are higher (better, larger), less than the number of subjects whose scores these subjects scores are lower (worse, smaller) than, is

$$W_{lj} \ = W_j - \ W_l \qquad - \quad - \quad - \qquad 13$$

which when expressed in the corresponding sample proportions of metric distances becomes

$$\left(\hat{\Pi}_j^+ - \hat{\Pi}_j^-\right) - \left(\hat{\Pi}_l^+ - \hat{\Pi}_l^-\right) \ = \ \frac{W_j - W_l}{n-t-1} \ = \ \frac{\left(f_j^+ - f_j^-\right) - \left(f_l^+ - f_l^-\right)}{n-t-1} \quad - \quad - \quad - \qquad 14$$

The corresponding sample variance is

$$Var\left(\hat{\Pi}_j^+ - \hat{\Pi}_j^-\right) - \left(\hat{\Pi}_l^+ - \hat{\Pi}_l^-\right) = \frac{Var\left(W_j - W_l\right)}{\left(n-t-1\right)^2} = \frac{\left(\left(\hat{\Pi}_j^+ - \hat{\Pi}_j^-\right) - \left(\hat{\Pi}_j^+ - \hat{\Pi}_j^-\right)^2 + \left(\hat{\Pi}_l^+ - \hat{\Pi}_l^-\right) - \left(\hat{\Pi}_l^+ - \hat{\Pi}_l^-\right)^2\right)}{n-t-1}$$

- - - 15

A null hypothesis that may be tested would be that the metric distances between the jth and lth subjects scores in the sampled population is zero. That is the null hypothesis

$$H_0 : \left(\Pi_j^+ - \Pi_j^-\right) - \left(\Pi_l^+ - \Pi_l^-\right) = 0 \quad \text{Vs} \quad H_1 : \left(\Pi_j^+ - \Pi_j^-\right) - \left(\Pi_j^+ - \Pi_j^-\right) \neq 0 \qquad - \quad - \quad 16$$

$$\chi^2 = \frac{\left(W_j - W_l\right)^2}{\operatorname{var}\left(W_j - W_l\right)} = \frac{\left(\left(\hat{\Pi}_j^+ - \hat{\Pi}_j^-\right) - \left(\hat{\Pi}_l^+ - \hat{\Pi}_l^-\right)^2\right)}{(n-t-1)\left(\hat{\Pi}_j^+ + \hat{\Pi}_j^- - \left(\hat{\Pi}_j^+ - \hat{\Pi}_j^-\right)^2\right) + \left(\hat{\Pi}_l^+ + \hat{\Pi}_l^- - \left(\hat{\Pi}_l^+ - \hat{\Pi}_l^-\right)^2\right)} \qquad - \quad - \quad - \quad 17$$

for

$$for\, l, j = 1,2,..n; \quad l \neq j$$

The null hypothesis H_0 of equation 16 is rejected at the $\propto$ level of significance if the chi-square value of equation 17 satisfies equitation 12 otherwise H_0 is accepted.

Illustrative Example I

We here use data on the scores in letter grades by a random sample of 13 Health Education Students in an introductory course in health statistics as shown in table 1 below.

Table 1: Scores in Letter Grades by a random sample of 13 health education students in Health Statistics and their ranks

$S/N_{(j)}$	1	2	3	4	5	6	7	8	9	10	11	12	13
Score (x_j)	C^-	A	C^+	B^+	C^-	C	E	D	E	B^-	C^+	A^+	A
Rank of score (y_j)	9.5	2.5	6.5	4	9.5	8	12.5	11	12.5	5	6.5	1	2.5

It is seen from the above results that the median score is a C^+ grade, which may here be taken as the average score by the student.

Applying equation 1 to the above data on students' scores with an average or median score of C^+, we obtain values of U_{lj} as shown in table 2 below.

Note that two students, student numbers 3 and 11 each has a score of C^+ which coincides with the median score by the students. Hence $t = 2$,
so that $n - t = 13 - 2 = 11$.

Table 2: values of U_{lj} (equation 1) for students scores and other statistics

		1	2	3	4	5	6	7	8	9	10	11	12	13	f_l^+	f_l^0	f_l^-	$\hat{\Pi}_l^+$	$\hat{\Pi}_l^0$	$\hat{\Pi}_l^-$	W_l $(f^+ - f^-)$	$(n-t-1)$	MD (l)
S/N	Score	C^-	A	C^+	B^+	C^-	C	E	D	E	B^-	C^+	A^+	A									
1	C^-		-1	-	-1	0	-1	1	1	1	-1	-	-1	-1	3	1	6	$3/10$	$1/10$	$6/10$	-3	10	-0.30
2	A	1		-	1	1	1	1	1	1	1	-	-1	0	8	1	1	$8/10$	$1/10$	$1/10$	7	10	0.70
3	C^+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	B^+	1	-1	-		1	1	1	1	1	1	-	-1	-1	7	0	3	$7/10$	$0/10$	$3/10$	4	10	0.40
5	C^-	0	-1	-	-1		-1	1	1	1	-1	-	-1	-1	3	1	6	$3/10$	$1/10$	$6/10$	-3	10	-0.30
6	C	1	-1	-	-1	1		1	1	1	-1	-	-1	-1	5	0	5	$5/10$	$0/10$	$5/10$	0	10	0.00
7	E	-1	-1	-	-1	-1	-1		-1	0	-1	-	-1	-1	0	1	9	$0/10$	$1/10$	$9/10$	-9	10	-0.90
8	D	-1	-1	-	-1	-1	-1	1		1	-1	-	-1	-1	2	0	8	$2/10$	$0/10$	$8/10$	-6	10	-0.60
9	E	-1	-1	-	-1	-1	-1	0	-1		-1	-	-1	-1	0	1	9	$0/10$	$1/10$	$9/10$	-9	10	-0.90
10	B^-	1	-1	-	-1	1	1	1	1	1		-	-1	-1	6	0	4	$6/10$	$0/10$	$4/10$	2	10	0.20
11	C^+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	A^+	1	1	-	1	1	1	1	1	1	1	-		1	10	1	0	$10/10$	$1/10$	$0/10$	10	10	1.00
13	A	1	0	-	1	1	1	1	1	1	1	-	-1		8	1	1	$8/10$	$1/10$	$1/10$	7	10	0.70
f_j^+		6	1	-	3	6	5	9	8	9	4	-	0	1									
f_j^0		1	1		0	1	0	1	0	1	0	-	0	1									

Note that $MD(j) = \dfrac{f_j^+ - f_j^-}{n-t-1}$

If research interest is to test the hypothesis that the metric distance between the proportion of students whose grades are greater than 'A' (j=2) from the proportion of students whose grades are less than 'A' relative to the median score C^+ is at least 0.6, the null and alternative hypotheses are;

$Ho: \Pi_2^+ - \Pi_2^- \geq 0.6$ Vs $H_1: \Pi_2^+ - \Pi_2^- \geq 0.6$

The null hypothesis is tested using equation 11, thus

$$\chi^2 = \frac{(n-t-1)\left[\left(\hat{\Pi}_2^+ - \hat{\Pi}_2^-\right) - 0.6\right]^2}{\left[\hat{\Pi}_2^+ + \hat{\Pi}_2^- - \left(\hat{\Pi}_2^+ - \hat{\Pi}_2^-\right)\right]} = \frac{10\left[\left(\frac{8}{10} - \frac{1}{10}\right) - 0.6\right]^2}{\left[\frac{8}{10} - \frac{1}{10} - \left(\frac{8}{10} - \frac{1}{10}\right)^2\right]} = 0.476$$

For $\propto = 0.05$ the null hypothesis is rejected at $\chi^2_{(0.95,1)}$ which is 3.841. The null hypothesis is here accepted, leading to a conclusion that the metric differences between the proportion of students whose grades are greater than A and the proportion of students whose grades are less than A relative to the median score C^+ is at least 0.6.

Illustrative Example II

We here also illustrate the present method with the following data on the cholesterol levels of a random sample of 11 male school teachers from a certain community.

S/N (i)	1	2	3	4	5	6	7	8	9	10	11
Cholesterol Level (x_i)	267	194	168	250	182	185	185	180	200	205	182
Rank (r_i)	1	5	11	2	8.5	6.5	6.5	10	4	3	8.5

It is seen from the above data that the median cholesterol level is 185 which may now be taken as the average cholesterol level of the school teachers.

Applying equation 1 to the above data on cholesterol levels with an average or median cholesterol level of 185, we obtain values of U_{lj} and other statistics as shown in table 3 below.

Note that two teachers, teacher numbers 6 and 7 each has the cholesterol level of 185 which coincide with the median cholesterol level of the teachers. Hence t=2, so that n-t-1 is 11-2-1=8.

Table 3: Values of U_{lj} (Eqn 1) for 11 male school teachers from a certain community

	S/N (j)	1	2	3	4	5	6	7	8	9	10	11
S/N (l)	Cholesterol level	267	194	168	250	182	185	185	180	200	205	182
1	267		1	1	1	1	-	-	1	1	1	1
2	194	-1		1	-1	1	-	-	1	-1	-1	1
3	168	-1	-1		-1	-1	-	-	-1	-1	-1	-1
4	250	-1	1	1		1	-	-	1	1	1	1
5	182	-1	-1	1	-1		-	-	1	-1	-1	0
6	185	-	-	-	-	-		-	-	-	-	-
7	185	-	-	-	-	-	-		-	-	-	-
8	180	-1	-1	1	1	1	-	-		-1	-1	-1
9	200	-1	1	1	-1	1	-	-	1		-1	1
10	205	-1	1	1	-1	1	-	-	1	1		1
11	182	-1	-1	1	-1	0	1	1	1	-1	-1	

Table 4: teachers' cholesterol levels and other statistics

Total (n-t-1)	f_i^+ (subjects above)	f_i^0 (same)	f_i^- (Subject below)	$\hat{\Pi}_i^+$ (Prop above)	$\hat{\Pi}_i^0$ (Prop same)	$\hat{\Pi}_i^-$ (Prop. Below)	W_i (GMD)	Rank (r_i)
8	8	0	0	1.00	0.00	0.00	8	1
8	4	0	4	0.50	0.00	0.50	0	5
8	0	0	8	0.00	0.00	1.00	-8	9
8	7	0	1	0.875	0.00	0.125	6	2
8	2	1	5	0.25	0.125	0.625	-3	6.5
8	1	0	7	0.125	0.00	0.875	-6	8
8	5	0	3	0.625	0.00	0.375	2	4
8	6	0	2	0.75	0.00	0.25	4	3
8	2	1	5	0.25	0.125	0.625	-3	6.5

It is seen from table 3 that school teacher number 1 with cholesterol level of 267 and $f^+ = 8,\ f^- = 0, \hat{\Pi}^+ = l.0\ and\ \hat{\Pi}^- = 0.0,$ has the highest cholesterol level, more than those of all other subjects sampled. Thus with $f^+ = 8, and\ f^- = 0,$ this subject, under normal circumstances, would be expected to experience more serious problems with cholesterol level than 8 other subjects relative to whatever problems a subject, that is a fellow school teacher, with a median cholesterol level of 185 would be normally have.

School teacher number 2 with a cholesterol level of 194 with $f^+ = f^- = 4,\ \hat{\Pi}^+ = \hat{\Pi}^- = 0.50\ and\ W = 0,$ ranked 5 is above (worse than) 4 and below (better than) 4 other subjects in the likelihood of experiencing problems with cholesterol levels, all thins being equal, relative to school teacher numbers 6 or 7 with the median cholesterol level of 185.

Similarly school teachers numbers 5 and 11 each with a cholesterol level of 182, $f^+ = 2\quad f^0 = 1,\quad f^- = 5,\quad \hat{\Pi}^+ = 0.25,\quad \hat{\Pi}^0 = 0.125,\quad \hat{\Pi}^- = 0.625\ and\ W = -3,$ ranked 6.5 each is above (worse) 2, had the same level as 1 and below (better than) 5 other school teachers in the sample in their relative performance when juxtaposed against, that is relative to their fellow school teachers numbers 6 and 7 each with the sample median cholesterol level of 185.

Finally school teacher number 3 with a cholesterol level of 168, with $f^+ = 0, f^0 = 0, f^- = 8,\quad \hat{\Pi}^+ = \hat{\Pi}^0 = 0.00,\quad \hat{\Pi}^- = 1.00, and\ W = -8,$ ranked 9, under normal circumstances would be expected to experience no more problem with cholesterol levels than any other fellow school teacher, but would likely experience less problem than 8 fellow school teachers relative, that is in comparison to any such problems that would normally be experienced by fellow school teachers numbers 6 and 7 with medium cholesterol level of 185.

If of research interest one may wish to statistically compare the relative performance indices of subjects in the sampled population of school teachers. For instance one may wish to test whether subject number 4 with cholesterol level of 250 and subject number 8 with the cholesterol level of 180 differ statistically in their relative performance indices, that is in their metric distances from the standard, normal or average subject with a median cholesterol level of 185. To do this we have from eqn. 17 and table 3 that

$$\chi^2 = \frac{(6-4)^2}{8\left((0.875+0.125-(0.875-0.125)^2)+(0.125+0.875-(0.125-0.875)^2\right)}$$

$$= \frac{4}{8(0.4375+0.4375)} = \frac{4}{8(0.875)} = \frac{4}{7.00} = 0.571, (R\ value = 0.0243),$$

which at 1 degree of freedom is statistically significant, showing that school teacher number 4 with a cholesterol level of 250 and school teacher number 8 with a cholesterol level of 180 would normally be expected to experience highly significant differences in any problems associated with cholesterol levels when each of them is compared in terms of the corresponding problems school teachers numbers 6 and 7 with median cholesterol level of 185 would ordinarily be expected to experience, all things being equal.

Results from these analyses would enable researchers, policy makers and implementers more rationally and systematically formulate subject targeted remedial measures for subject specific case management of any diseases assuming that there is a reference standard for comparisons of subjects in a study population.

III. Summary And Conclusion

Under the assumption that there exists a subject whose performance, observed value or score may be regarded as an average or standard value or score to be used as a reference, standard, average value or score for comparison with other performance or scores of other subjects on a condition in a population, this paper developed what is here referred to as metric distance measured in terms of the number of subjects to determine by how many subjects or scores the score or performance by a randomly selected subject is above (better, worse), the same as, or lower (worse, better), than their own, relative to the standard with respect to a condition of research interest in a population.

The further away the so-called metric distance, expressed in terms of number of subjects, of a randomly selected subject is from the average or standard subjects performance value or score the better or less serious (the worse or more serious) than that subjects' condition relative to, that is in comparison with the average, reference or standard subjects' condition and hence also relative to, that is in comparison with the conditions of some other subjects in the study population.

Tests statistics are developed for use in testing the statistical significance of the metric distance of a randomly selected subject as well as the statistical significance of the difference between the metric distances of any two randomly selected subjects in the population.

The proposed method is illustrated with some sample data.

References

[1]. Siegel S. (1956): Nonparametric Statistics for the Behavioral Sciences. McGraw Hill Kogakusha Ltd. Tokyo.

[2]. Mahalanobis, P.C. (1936), "On the Generalized Distance in Statistics". Proceedings of the National Institute of Sciences of India 2(1):49-55.

[3]. Machlachlan, G.J. (1992) Discriminant Analysis and Statistical Pattern Recognition. Wiley Intersciences.

[4]. Dodge, Y. (2003): Oxford Dictionary of Statistical Terms, Oxford University Press.