



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Mahalanobis Distance: A Study on Achievement of Science and Mathematics

Subir Sen¹ and Indranil Pal²

1. Department of Education, Sidho-Kanho-Birsha University, Purulia, W.B. India

2. Independent Scholar

Abstract. Present work is dealt with the technique; how Mahalanobis Distance can be applied to analyze achievement related issues. The problem addressing the group achievement in mathematics, Physical Science and Life Science for two groups of students viz. seventh grade and eighth grade for three types of schools. Mahalanobis Distance is considered to test the dynamical nature of achievement of two groups of learner in three subjects. It is found that for most of the cases distances are significant. It represents the nature of the achievement for the group of subjects between different set of students differs significantly.

Keywords: Mahalanobis Distance, Mathematics, Physical Science, Life Science, Seventh Grade, Eighth Grade, Achievement Analysis, Formative and Summative Evaluation.

1. Introduction

In Mahalanobis Distance a measure of divergence or distance between groups in terms of multiple characteristics is used. Mahalanobis proposed this measure in 1936 (Mahalanobis, 1936) in the context of his studies on racial likeness. It has played a fundamental and important role in statistics and data analysis with multiple measurements. For last few decades Mahalanobis Distance playing a vital role in differentiating characteristics in different field like anthropology, clustering, classification, image processing, Neurocomputing, physics, Precision Medicine etc. Following works such as Xiang et al. (2008), Bedrick et al. (2000), Cochran and Rubin (1973), Rubin (1976, 1979, 1980), Mclachlan (1999), Rosenbaum (2015), Diedrichsen, Provost and Zareamoghaddam (2016), Cristani and Murino (2018), Toma (2019), Imani (2019) and Etherington (2019) are some evidence from such fields.

But in the field of education no sufficient documents are found regarding the use of Mahalanobis Distance. It may be used as a strong measure in terms of a single dimensionless number for comparing several set of data taken together as a unit.

Ahmed et al (2020) used Mahalanobis Distance to address the nature of achievement in mathematics for two groups of higher secondary level students. Most of the cases they found significant difference between achievements in different grades for two group of students. Present work is based on the work of Sen and Kar (2014). Their work is dealt with the study of the achievements in Unit Test (Formative) and Annual Examination (Summative) for seventh and eighth grade students for three different type of schools viz. Boys, Girls and Co-educational in Kalna subdivision of Purba Bardhaman, West Bengal, India. Some statistical measures like coefficient of correlation and t-test are used by them to analysis the data.

Present work is an attempt to study the difference in dynamical nature of the achievement in three subjects, Mathematics, Physical Science and Life Science taken together as a bunch of achievement terms of Mahalanobis Distance.

Objectives of the Study

Firstly, our objective is to apply MD to find out the difference in achievement for two groups where each group contains several set of data (here scores in mathematics, physical science and life science).

Secondly, another objective of this study is to find out the difference in achievement in mathematics and Science subjects for different learners in different grades and different type of schools i.e. boys, girls and co-educational.

Mahalanobis Distances are considered for more generalized reflection of difference of their results. The Distance is used to measure the distance between dynamical nature of wholistic achievement between groups.

Methodology

Step I

Two sets of data for three schools are prepared for calculating the distance.

A. For first set of data, learners of seventh grade are considered. Here, scores in three different subjects i.e. mathematics, physical science and life science are considered for formative and summative tests.

B. Similar data sets for eighth grade are also prepared.

Step II

Mean, Covariance etc. are calculated for achievement of each subject for different groups because those are required for Mahalanobis Distance.

Step III

Mahalanobis Distance is calculated by the following formula:

$$\Delta^2 = (X - Y)^T \Sigma^{-1} (X - Y)$$

Where X and Y are column vectors of means and Σ is pooled covariance matrix of two groups of data.

$$\text{Mahalanobis Distance} = \left[(X - Y)^T \Sigma^{-1} (X - Y) \right]^{\frac{1}{2}}$$

Pooled Covariance Matrix Σ can be defined as

$$\Sigma = [(n_1 - 1)\Sigma_1 + (n_2 - 1)\Sigma_2] / N$$

Where Σ_1 and Σ_2 be the Covariance Matrices, n_1 and n_2 be the sample size for first and second group respectively and $N = n_1 + n_2 - 2$.

As the covariances and pooled covariances are used to calculate Mahalanobis Distance (MD), it is more effective and also valid measure to compare two groups of data. The distance is represented by a single dimensionless number.

Step IV

It may be assumed from the distribution that;

1. If $0 < MD < 1$, distance is insignificant. It may be concluded that there is no significant difference between the dynamical natures of the groups. Here group means a set of data for different subjects.
2. If $1 \leq MD < 2$, distance is significant. It represents that there is a significant difference between the dynamical natures of the groups.
3. $MD \geq 2$: represents the distances are strongly significant. This result shows that there is a very strong difference between the dynamical natures of the groups.

Results and Discussions:

School	Class	Subject	Mean Formative	Mean Summative	Number of students
Co-Educational School	VII	Mathematics	49.58	39.46	121
		Physical Sc.	43.89	29.58	121
		Life Sc.	43.61	36.77	121
	VIII	Mathematics	42.6	26.79	101
		Physical Sc.	50.2	31.84	101
		Life Sc.	45.93	31.1	101
Boys' School	VII	Mathematics	31.53	17.82	138
		Physical Sc.	30.18	23.75	138
		Life Sc.	50.72	40.23	138
	VIII	Mathematics	46.13	27.63	105
		Physical Sc.	45.95	35.96	105
		Life Sc.	50.93	38.94	105
Girls School	VII	Mathematics	36.63	29.17	226
		Physical Sc.	44.99	39.0	226
		Life Sc.	41.54	28.84	226
	VIII	Mathematics	35.63	16.33	224
		Physical Sc.	42.27	29.34	224
		Life Sc.	35.81	30.42	224

Table.1. Descriptive Statistics for the group of students in different schools.

Our aim is to study the dynamic nature of the group of subjects (here Mathematics, Physical Science and Life Science) as a unit or a branch of achievement. Here t-test or equivalent test cannot be used. Let us construct some groups;

Group No	Class	School type	Formative scores	Summative scores	Subject columns
I	VII	Coeducational	Yes	No	3
II	VII	Coeducational	No	Yes	3
III	VIII	Coeducational	Yes	No	3
IV	VIII	Coeducational	No	Yes	3
V	VII	Boys	Yes	No	3
VI	VII	Boys	No	Yes	3
VII	VIII	Boys	Yes	No	3
VIII	VIII	Boys	No	Yes	3
IX	VII	Girls	Yes	No	3
X	VII	Girls	No	Yes	3
XI	VIII	Girls	Yes	No	3
XII	VIII	Girls	No	Yes	3
XIII	VII	Coeducational	Yes	Yes	6
XIV	VIII	Coeducational	Yes	Yes	6
XV	VII	Boys	Yes	Yes	6
XVI	VIII	Boys	Yes	Yes	6
XVII	VII	Girls	Yes	Yes	6
XVIII	VIII	Girls	Yes	Yes	6

Table 2. Different class-wise and subject-wise groups considered by the authors

groups	G-I	G-II	G-III	G-IV	G-V	G-VI	G-VII	G-VIII	G-IX	G-X	G-XI	G-XII
G-I	X	1.32	0.84									
G-II	1.32	X		1.45								
G-III	0.84		X	3.17								
G-IV		1.45	3.17	X								
G-V					X	1.60	2.81					
G-VI					1.60	X		1.73				
G-VII					2.81		X	1.65				
G-VIII						1.73	1.65	X				
G-IX									X	1.12	0.53	
G-X									1.12	X		1.75
G-XI									0.53		X	2.34
G-XII										1.75	2.34	X

Table 3: MD for achievement in group of three subjects for class VII and VIII of three different schools.

Our target was to study the behaviour of the achievement for different group of learners who are studying in different grades and there is a difference in their achievement. If we consider three subjects as a unit it is more effective than subject-wise analysis and in that case MD may be one of the effective measures. Here we get 3X3 covariance and pooled covariance matrices.

We can classify the calculated MD into three categories as follows:

- MD < 1: represents the distances are insignificant. Only two such cases are found. (vide Table 3.)
- $1 \leq MD < 2$: represents the distances are significant. In six such cases this type of distances are found. (vide Table 3.)
- MD ≥ 2 : represents the distances are strongly significant. In three such cases are found. (vide Table 3.)

When we consider the subjects (here Mathematics, Physical Science and Life Science is considered) as a group and calculate the MD between target groups, the distances represent significant difference between the groups. Strongly significant difference found between 'Group III and Group IV', Group V and Group VII and 'Group XI and Group XII'. This is the opportunity of MD, any number of subjects can be considered as a unit (each unit should have equal number of subjects) and one can determine the distance.

When the achievement of formative and summative are considered for a class as a group of achievement (groups XIII to XVIII in table 2), we get a 6X6 covariance and pooled covariance matrices. MDs are shown below:

Groups	XIII	XIV	XV	XVI	XVII	XVIII
XIII	-----	1.86	3.15		2.29	
XIV		-----		1.47		1.47
XV			-----	2.21	3.24	
XVI				-----		1.53
XVII					-----	1.608
XVIII						-----

Table 4. MDs for six achievements are taken together as a group for class VII and VIII.

In five such distances are significant and four distances are strongly significant. There is a significance difference in achievement when scores for five different tests taken together between class VII and VIII of coeducational and

girls' school. Similar results found for class VIII between (a) coeducational and boys' schools, (b) coeducational and girls' schools and (c) boys' and girls' schools.

On the other hand, four strongly significant differences are found for class VII between (a) co-educational and boys' schools, (b) co-educational and girls' schools and (c) boys' and girls' schools. Also similar result is found for class VII and VIII for boys' school.

So, MD can be used for evaluating distance between two groups of scores where each group contains several numbers of attribute.

Concluding Remarks

The study which is carried out to analyze the difference in achievement of seventh and eighth grade learners of different types of schools where MD is considered as a measure of the difference between their performances (here achievement is considered) . Salient features are listed below:

1. When group of subjects are considered, MDs denotes significant differences in dynamic nature of achievements in three subject groups.
2. Comparisons between formative and summative tests are done. There are 2 insignificant, 6 significant and 3 strongly significant differences found.
3. When six tests are taken together as a group for each class is considered, 5 significant and 4 strongly significant distances are found.

Therefore, summing all the results where the group of subjects considered together as a unit, MDs are significant for most of the cases. Performance in achievement decreases remarkably for summative tests compared to formative tests. Also there are significant and strongly significant differences in achievement between different classes of same school as well as different school.

This work shows how MD can be calculated and a use of this distance in terms of achievement is shown. Researchers may use the distance to measure the difference between two bunches of responses recorded. This is the major advantage of MD.

References

Ahmed, E. A., Banerjee, M., Sen, S. & Chatterjee, P. (2020). Application of Mahalanobis Δ^2 on Achievement Tests on Mathematics: A Study on Higher Secondary Level Students. *Indian Journal of Psychology and Education*, 10(1), 36- 40.

Bedrick, E. J., Lapidus, J. & Powell, J. F. (2000). Estimating the Mahalanobis Distance from Mixed Continuous and Discrete Data, *Biometrics*, 56 (2), 394-401.

Cochran, W. G., and Rubin, D. B. (1973), "Controlling Bias in Observational Studies: A Review. *Sankya Ser. A*, 35, 417-446.

Cristani, M. & Murino, V. (2018). Chapter 10 - Person re-identification. *Image and Video Processing and Analysis and Computer Vision. Academic Press Library in Signal Processing, Volume 6*, 365-394.

Diedrichsen, J., Provost, S. & Zareamoghaddam, H. (2016). On the distribution of cross-validated Mahalanobis distances. ArXiv:1607.01371v1[stat.AP] 5 jul 2016.

Etherington, T. R. (2019). Mahalanobis distances and ecological niche modelling: correcting a chi-squared probability error. *PeerJ* 7:e6678 <http://doi.org/10.7717/peerj.6678>

Imani, M. (2019). Difference-based target detection using Mahalanobis distance and spectral angle. *International Journal of Remote Sensing*, 40(3), 811-831.

Mahalanobis, P.C. (1936). On the Generalized Distance in Statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 49 – 55.

Mclachlan, G. J. (1999). Mahalanobis Distance, *Resonance*, June, 20-26.

Rosenbaum, P. R. (2015). Observational Studies: Overview. James D. Wright (eds.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed. pp.107-112). Oxford: Elsevier.

Rubin, D. B. (1976). Multivariate Matching Methods That are Equal Percent. *Biometrics*, 32, 185-203.

Rubin, D. B. (1979). Using Multivariate Matched Sampling and Regression Studies. *Journal of the American Statistical Association*,

Rubin, D. B. (1980). Bias Reduction Using Mahalanobis Metric Matching. *Biometrics*, 36, 293-98.

Sen, S. & Kar, S. (2014). Comparison Between the Achievements in Unit Tests and Annual Examinations: A Study of Seventh and Eighth Grade Students on Science Subjects. *Indian Streams Research Journal* , 4(7), 1-5.

Toma, E. (2019). Analysis of motor fan radiated sound and vibration waveform by automatic Pattern recognition technique using “Mahalanobis distance”. *Journal of Industrial Engineering International*, 15(1), 81-92.

Xiang, S., Nie, F & Zhang, C. (2008). Learning a Mahalanobis distance metric for data clustering and classification. *Pattern Recognition*, 41, 3600 – 3612.



Filename: 2007236_196866
Directory: C:\Users\Subir\Desktop
Template: C:\Users\Subir\AppData\Roaming\Microsoft\Templates\Normal.dotm
Title: Paper Template for Pre-Olympic Congress
Subject: Sports Science, Sports Engineering, Computer Science in Sport, Management,
Statistics, 体育科学, 体育工程, 体育统计, 体育管理
Author: World Academic Union (World Academic Press)
Keywords:
Comments:
Creation Date: 7/17/2020 12:21:00 PM
Change Number: 4
Last Saved On: 9/9/2020 4:50:00 PM
Last Saved By: Subir
Total Editing Time: 10 Minutes
Last Printed On: 9/9/2020 6:27:00 PM
As of Last Complete Printing
Number of Pages: 6
Number of Words: 2,174 (approx.)
Number of Characters: 12,396 (approx.)