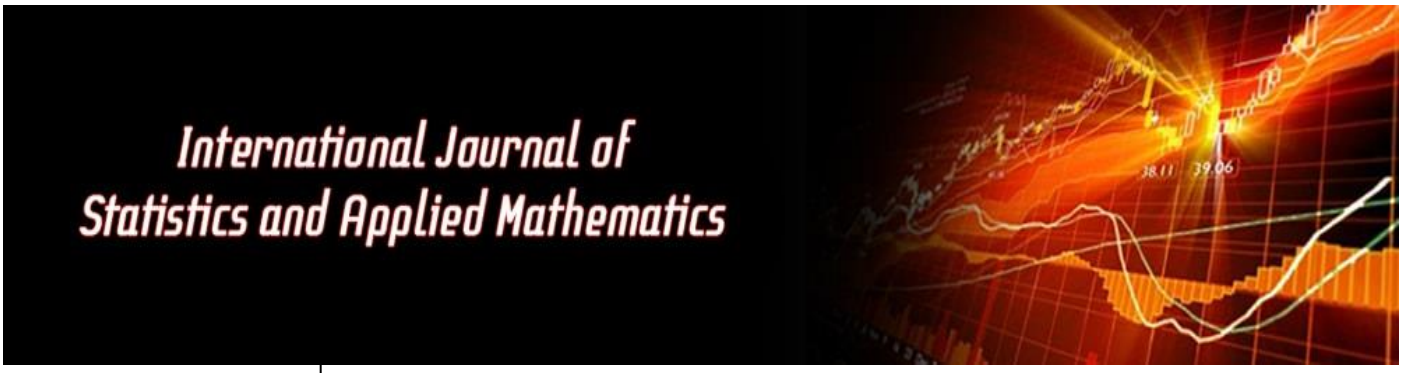
International Journal of Statistics and Applied Mathematics 2018; 3(2): 484-491



ISSN: 2456-1452
Maths 2018; 3(2): 484-491

www.mathsjournal.com
Received: 21-01-2018
Accepted: 22-02-2018

Shukla Jaysan H
Research Scholar at Rai University, Ahmedabad, Gujarat, India

Dr. Pradeep J Jha
Professor of Mathematics at Rai University, Ahmedabad, Gujarat, India

Correspondence
Dr. Pradeep J Jha
Professor of Mathematics at Rai University, Ahmedabad, Gujarat, India

Factors of bio-rhythms and applications to *PCA

Shukla Jaysan H and Dr. Pradeep J Jha

Abstract
In this note, authors target at, using PCA, identifying new variables of Bio-rhythms which shade dominance in the total spectrum of successful orientation. The important feature is the post analysis on the restructured variables and derived data based on PCA* and its orientation the perpendicular frame of Eigen vectors. The important point, at this stage, is to observe and study the impact of data corresponding eradicated variables on the data of existing variables.

Keywords: Bio-rhythms, eigenvalues, eigenvectors, covariance matrix, PCA

1. Introduction
The basic content of this paper target upon identifying the fundamental causes those are responsible for the various grades achieved by different students on taking up the tests on different subjects on different time slots. It took us a long time to identify some of the major causes which were responsible for the so-called variation between the predicted marks and marks received on different examinations given on different times during a week's slot. The Bio-rhythm is, as understood, a mental drive which is a gross of four broad divisions depicted below,

1. Physical Bio-rhythm
2. Emotional Bio-rhythm
3. Intellectual Bio-rhythm and
4. Spiritual Bio-rhythm

As advanced studies reflect that these main divisions, stated above, have also further subdivisions. It is principally hypothesised that, in general, all these main characteristics are the function of the birth date and more precisely the time of birth. To be more precise all these factors experienced as a gross effect on a given time are associated with the total time elapsed after the birth of the individual. Total number of days that the subject, without any major illness, has passed from the time of birth up to or a day before the test is administered is an important factor. It is accepted that mental behaviour and attitude which are gross outcomes of Bio-rhythms directly reflect the individual performance of the subject. Some of the dominating characteristics are immutable but there exists some more characteristics which by imparting regular instilling pre-designed lessons certainly elevates some features to qualify for better standards on the tests to follow.
The content of the first paper on the same line has been and we have focussed total contemplation on collection of primary data on different time frames.

2. PCA on Bio-rhythmic Components
Principal component analysis (PCA) has been known to be one of the most important processes derived from application of linear algebra. PCA is used abundantly in all forms of analysis - from neuroscience to computer graphics because it is a simple, non-parametric method of extracting relevant information, a simplified version obtained by bypassing confusing statistical routines on primary data sets. PCA provides a structural routine indicating 'how to reduce a complex multi-variate data set to a data of fewer state / variables' and dimension. At this stage it is noteworthy that these new variable do not conform in nature with those of original variables. So it is really left to the discretion of the studied writer to interpret

and then to apply the new variables to the original primary data. Thus the reduction of variables is liable to cause more profound work of evaluation and interpretation.

We follow the following procedure in order to generate a tabular data.

(1) We take a randomly selected sample of 16 students from a group of 120 students studying in the same standard of one school providing same educational environment. According to the assumption [1] we know their birth dates and approximate birth time.
(2) The students are well informed about the date and the time they will be taking up the examination on a particular subject.
(3) We define a parameter't' as follows. $t = \text{Date of examination} - \text{Birth Date of student}$. ['t' is a real value.]
(4) We convert this real value using the following formula, shown below*, associating 't' value in to scores corresponding to four different components. All these scores are in the interval [-1, 1].

[Note: Negative value is an indication that the subject on that time does not exhibit positive attitude on that features and vice versa.]

2.1 Calculation of Bio-rhythms

$$\text{Physical Bio-rhythm Score} = \sin\left(\frac{2\pi t}{23}\right) \quad (1)$$

$$\text{Emotional Bio} - \text{rhythm Score} = \sin\left(\frac{2\pi t}{28}\right) \quad (2)$$

$$\text{Intellectual Bio} - \text{rhythm Score} = \sin\left(\frac{2\pi t}{33}\right) \quad (3)$$

$$\text{Spiritual Bio} - \text{rhythm Score} = \sin\left(\frac{2\pi t}{53}\right) \quad (4)$$

Tabulated format of these readings is,

Table 1: Bio-rhythms of 16 students.

No. of Student	Date of Experiment	Date of Birth	Bio-rhythms			
			Physical	Emotional	Intellectual	Spiritual
1	01-Jan-18	01 November 1990	-0.52	-0.22	0.87	0.00
2		11 January 1990	-0.94	0.22	1.00	-0.29
3		02 March 1991	-0.82	-0.78	-0.87	-0.98
4		22 January 1989	0.52	0.62	-0.19	0.99
5		03 February 1992	-0.27	-0.43	0.46	0.85
6		24 March 1990	-0.40	-0.62	0.37	0.93
7		12 May 1989	0.94	0.90	-0.76	0.81
8		30 August 1990	-0.82	0.97	1.00	0.93
9		29 October 1990	0.27	0.43	0.46	0.35
10		22 November 1991	0.73	0.90	-0.76	-0.98
11		21 December 1989	-0.63	0.97	-0.69	-0.35
12		31 May 1991	-1.00	0.43	-0.37	0.12
13		12 February 1993	-0.82	0.90	0.37	1.00
14		24 February 1991	-0.52	-0.78	-0.81	-0.88
15		16 July 1989	0.14	-0.78	-0.87	-0.46
16		03 September 1989	0.82	0.62	0.91	0.00

2.2 Calculation of range conversion for Bio-rhythms from [-1, 1] to [1,100]:

For sake of simplicity let us consider range conversion from [-1, 1] to [1,100] using formula,

$$\text{New value} = \frac{(\text{Old Value} - \text{Old Min}) * (\text{New Max} - \text{New Min})}{(\text{Old Max} - \text{Old Min})} + \text{New Min} = \frac{(\text{Old Value} - \text{Old Min}) * (\text{New Range})}{(\text{Old Range})} + \text{New Min} \quad (5)$$

$$\text{New value} = \frac{(\text{Old Value} - (-1)) * (100 - 1)}{(1 - (-1))} + 1 \quad (6)$$

With this conversion data is shifted to,

Table 2: Re scaled Bio-rhythms of 16 students.

No. of Student	Date of Experiment	Date of Birth	Bio-rhythms			
			Physical	Emotional	Intellectual	Spiritual
1	01-Jan-18	01 November 1990	25	39	93	50
2		11 January 1990	4	62	100	36
3		02 March 1991	10	12	8	2

4		22 January 1989	76	81	41	99
5		03 February 1992	37	29	73	92
6		24 March 1990	31	20	69	96
7		12 May 1989	97	95	13	91
8		30 August 1990	10	99	100	96
9		29 October 1990	64	72	73	68
10		22 November 1991	87	95	13	2
11		21 December 1989	19	99	16	33
12		31 May 1991	1	72	32	56
13		12 February 1993	10	95	69	100
14		24 February 1991	25	12	10	7
15		16 July 1989	57	12	8	28
16		03 September 1989	91	81	96	50

The same has been shown below in concise form for further statistical process.

Table 3: Mean values of Bio-rhythms.

No. of Student	Bio-rhythms			
	Physical	**Emotional**	**Intellectual**	**Spiritual**
1	25	39	93	50
2	4	62	100	36
3	10	12	8	2
4	76	81	41	99
5	37	29	73	92
6	31	20	69	96
7	97	95	13	91
8	10	99	100	96
9	64	72	73	68
10	87	95	13	2
11	19	99	16	33
12	1	72	32	56
13	10	95	69	100
14	25	12	10	7
15	57	12	8	28
16	91	81	96	50
Mean	$\bar{X}_P$ =40	$\bar{X}_E$ =61	$\bar{X}_I$ =51	$\bar{X}_S$ =57

2.3 Shifting data with respect to mean value

In order to make calculation simpler we subtract mean of each bio-rhythm from original figures of the same Bio-rhythm; technically saying we find deviations from mean. The deviations have been tabularized below.

Table 4: Shifted Bio-rhythms with respect to mean of 16 students.

No. of Student	Bio-rhythms(Shift of Origin about Mean)			
	Physical-$\bar{X}_P$	**Emotional-$\bar{X}_E$**	**Intellectual-$\bar{X}_I$**	**Spiritual-$\bar{X}_S$**
1	-15	-22	42	-7
2	-36	1	49	-21
3	-30	-49	-43	-55
4	36	20	-10	42
5	-3	-32	22	35
6	-9	-41	18	39
7	57	34	-38	34
8	-30	38	49	39
9	24	11	22	11
10	47	34	-38	-55
11	-21	38	-35	-24
12	-39	11	-19	-1
13	-30	34	18	43
14	-15	-49	-41	-50
15	17	-49	-43	-29
16	51	20	45	-7

We arrange the array in the form of a matrix denoting it by 'A' whose rows sequentially represent serial number of students and consecutive columns stand for deviations in Bio-rhythms in corresponding order.

$$\mathrm{A} = \begin{bmatrix} -15 & -22 & 42 & -7 \\ -36 & 1 & 49 & -21 \\ -30 & -49 & -43 & -55 \\ 36 & 20 & -10 & 42 \\ -3 & -32 & 22 & 35 \\ -9 & -41 & 18 & 39 \\ 57 & 34 & -38 & 34 \\ -30 & 38 & 49 & 39 \\ 24 & 11 & 22 & 11 \\ 47 & 34 & -38 & -55 \\ -21 & 38 & -35 & -24 \\ -39 & 11 & -19 & -1 \\ -30 & 34 & 18 & 43 \\ -15 & -49 & -41 & -50 \\ 17 & -49 & -43 & -29 \\ 51 & 20 & 45 & -7 \end{bmatrix}$$

2.4 Calculation of Covariance for matrix A

The next one in the order is to find covariance matrix of the given matrix A. Each entry of this matrix shows the index of associativity between the variables. The rule is higher the value of this index indicates strong associativity in context of comparison. Covariance matrix of data matrix A is defined as,

$$Cov(A) = \begin{bmatrix} var(P) & cov(P,E) & cov(P,I) & cov(P,S) \\ cov(E,P) & var(E) & cov(E,I) & cov(E,S) \\ cov(I,P) & cov(I,E) & var(I) & cov(P,E) \\ cov(S,P) & cov(S,E) & cov(E,P) & var(S) \end{bmatrix}$$

Where $Cov(X,Y) = \frac{\sum_{i=1}^{16}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$ and $Cov(X,X) = Var(X)$. We extract the entries from above matrix A and on application of the above formula we get the following matrix Cov.(A).

$$Cov(A) = \begin{bmatrix} 1113.994 & 273.179 & -219.562 & 73.573 \\ 273.179 & 1191.222 & 202.874 & 382.902 \\ -219.562 & 202.874 & 1343.416 & 608.077 \\ 73.573 & 382.902 & 608.077 & 1312.834 \end{bmatrix}$$

We find latent values associated with this matrix and corresponding latent (Eigen) vectors. The Eigen vectors will find some invariant direction and the Eigen values giving the contraction factor for the vector in connection to the one of original matrix. These are found using characteristic equation derived as $|cov(A) - \lambda \cdot I| = 0$

Denoting λ_i as Eigen values and X_i as Eigen vectors where i=1, 2, 3 and 4, we have

$$\lambda_1 = 2122.42;\ \mathrm{X}_1 = \begin{bmatrix} 0.0262574 \\ 0.416102 \\ 0.619876 \\ 0.664774 \end{bmatrix}$$

$$\lambda_2 = 1435.05;\ \mathrm{X}_2 = \begin{bmatrix} 0.740039 \\ 0.527649 \\ -0.416075 \\ 0.0284713 \end{bmatrix}$$

$$\lambda_3 = 773.474;\ \mathrm{X}_3 = \begin{bmatrix} 0.540995 \\ -0.740568 \\ 0.050125 \\ 0.395436 \end{bmatrix}$$

$$\lambda_4 = 630.527;\ \mathrm{X}_4 = \begin{bmatrix} -0.398719 \\ 0.00191294 \\ -0.663418 \\ 0.633163 \end{bmatrix}$$

2.5 Energy associated to Eigen values of Cov. (A)

At this stage we define energy of each Eigen value as its own percentage in the 'less than type of cumulative contribution' on the

scale of a centage scale. To achieve this we have two criteria.

(1) Selection of variables will follow decreasing order of Eigen values.
(2) The cut off criteria will be about 70 % in accumulation.

Table 5: Energy calculation of Eigen vectors.

Order of Eigen values	Eigen Value	Cumulative Energy of Eigen Value	Percentage of Energy	Cumulative Percentage of Energy
λ_1	2122.42	2122.42	42.78%	42.78%
λ_2	1435.05	3557.47	28.92%	71.70%
λ_3	773.474	4330.944	15.59%	87.29%
λ_4	630.527	4961.471	13%	100.00%

As a virtue of these two criteria, we are able to fix up the first two leading variates who dominates in the set of given four Variates; which are X_1 and X_2.

To define Principal component we will consider Feature vector as,

$$F = [X_1\ X_2] = \begin{bmatrix} 0.0262574 & 0.740039 \\ 0.416102 & 0.527649 \\ 0.619876 & -0.416075 \\ 0.664774 & 0.0284713 \end{bmatrix}$$

This feature matrix F such defined will transform original data matrix A in 2 dimensional subspace re-oriented Data Matrix A_{New} defined as the regular product of two matrices A and F taken in the order as they appear.

$$A_{New} = A \cdot F$$

$$A_{New} = \begin{bmatrix} -15 & -22 & 42 & -7 \\ -36 & 1 & 49 & -21 \\ -30 & -49 & -43 & -55 \\ 36 & 20 & -10 & 42 \\ -3 & -32 & 22 & 35 \\ -9 & -41 & 18 & 39 \\ 57 & 34 & -38 & 34 \\ -30 & 38 & 49 & 39 \\ 24 & 11 & 22 & 11 \\ 47 & 34 & -38 & -55 \\ -21 & 38 & -35 & -24 \\ -39 & 11 & -19 & -1 \\ -30 & 34 & 18 & 43 \\ -15 & -49 & -41 & -50 \\ 17 & -49 & -43 & -29 \\ 51 & 20 & 45 & -7 \end{bmatrix} \cdot \begin{bmatrix} 0.0262574 & 0.740039 \\ 0.416102 & 0.527649 \\ 0.619876 & -0.416075 \\ 0.664774 & 0.0284713 \end{bmatrix} \Rightarrow A_{New} = \begin{bmatrix} 12.590 & -40.429 \\ 15.673 & -47.436 \\ -84.663 & 11.056 \\ 35.988 & 42.863 \\ 23.874 & -27.208 \\ 19.814 & -34.974 \\ 14.604 & 77.012 \\ 71.437 & -21.477 \\ 26.081 & 14.522 \\ -44.607 & 66.743 \\ -22.096 & 18.321 \\ -8.596 & -15.140 \\ 52.951 & 10.393 \\ -79.315 & -21.658 \\ -66.250 & 4.014 \\ 33.091 & 29.730 \end{bmatrix}$$

2.6 Defining Principal Components

In context of the given *4 – variate bio-rhythm matrix A*, its corresponding *2-variate (new- different from original) reduced matrix* is the matrix 'A' new. In terms of these two Eigen vectors we have new variables - principal Components are defined by linear combinations,

$$* PC1 = (0.0262574)\text{Physical} + (0.416102)\text{Emotional} + (0.619876)\text{Intellectual} + (0.664774)\text{Spiritual} \quad (7)$$

$$**PC2 = (0.740039)\text{Physical} + (0.527649)\text{Emotional} + (-0.416075)\text{Intellectual} + (0.0284713)\text{Spiritual} \quad (8)$$

A_{New} is matrix where rows represents students serial number and columns represents principal component 01 and principal component 02 respectively as,

$$A_{New} = \begin{bmatrix} 12.590 & -40.429 \\ 15.673 & -47.436 \\ -84.663 & 11.056 \\ 35.988 & 42.863 \\ 23.874 & -27.208 \\ 19.814 & -34.974 \\ 14.604 & 77.012 \\ 71.437 & -21.477 \\ 26.081 & 14.522 \\ -44.607 & 66.743 \\ -22.096 & 18.321 \\ -8.596 & -15.140 \\ 52.951 & 10.393 \\ -79.315 & -21.658 \\ -66.250 & 4.014 \\ 33.091 & 29.730 \end{bmatrix}$$

2.7 Statistical analysis using Paired t-test

Next goal is to test data statistically with help of paired t-test, for which we will consider marks predicted by bio rhythms and marks predicted by Principal component Analysis. Converting these two set of marks in same range of [1,100] using range conversion formula,

$$\text{New value} = \frac{(\text{Old Value} - \text{Old Min}) * (\text{New Max} - \text{New Min})}{(\text{Old Max} - \text{Old Min})} + \text{New Min.}$$

Predicted marks by Average of Bio-rhythms,

$$\text{Marks} - \text{I} = \frac{(\text{Average of biorhythm} - (-44)) * (100-1)}{(27-(-44))} + 1 \quad (9)$$

Predicted marks by PCA,

$$\text{Marks} - \text{II} = \frac{(\text{Average of PC1and PC2} - (-50)) * (100-1)}{(46-(-50))} + 1 \quad (10)$$

†All these values are derived by using formula of Marks-I as shown by result (9)
‡ All these values are derived by using formula of Marks-II as shown by result (10)

Table 6: Comparison of Expected Marks.

No. of Student	Bio-rhythms (Shift of Origin about Mean)				Average of Bio-rhythms.	Predicted marks by Average of Bio-rhythms [Marks-I]†	PC1	PC2	Average of PC1 And PC2	Predicted marks by PCA [Marks-II]‡	Difference of Marks-I and Marks-II
	[P]	[E]	[I]	[S]							
1	-15	-22	42	-7	0	61	12.59	-40.43	-14	38	23
2	-36	1	49	-21	-2	59	15.67	-47.44	-16	36	23
3	-30	-49	-43	-55	-44	1	-84.66	11.06	-37	15	-14
4	36	20	-10	42	22	92	35.99	42.86	39	93	-1
5	-3	-32	22	35	6	69	23.87	-27.21	-2	51	18
6	-9	-41	18	39	2	64	19.81	-34.97	-8	45	19
7	57	34	-38	34	22	91	14.60	77.01	46	100	-9
8	-30	38	49	39	24	95	71.44	-21.48	25	78	17
9	24	11	22	11	17	85	26.08	14.52	20	73	12
10	47	34	-38	-55	-3	57	-44.61	66.74	11	64	-7
11	-21	38	-35	-24	-10	47	-22.10	18.32	-2	51	-4
12	-39	11	-19	-1	-12	45	-8.60	-15.14	-12	40	5
13	-30	34	18	43	16	84	52.95	-10.39	21	75	9
14	-15	-49	-41	-50	-39	8	-79.31	-21.66	-50	0	8
15	17	-49	-43	-29	-26	26	-66.25	4.01	-31	20	6
16	51	20	45	-7	27	99	33.09	29.73	31	85	14

[P]-Physical Bio-rhythm, [E]-Emotional Bio-rhythm, [I]-Intellectual Bio-rhythm, [S]-Spiritual Bio-rhythm

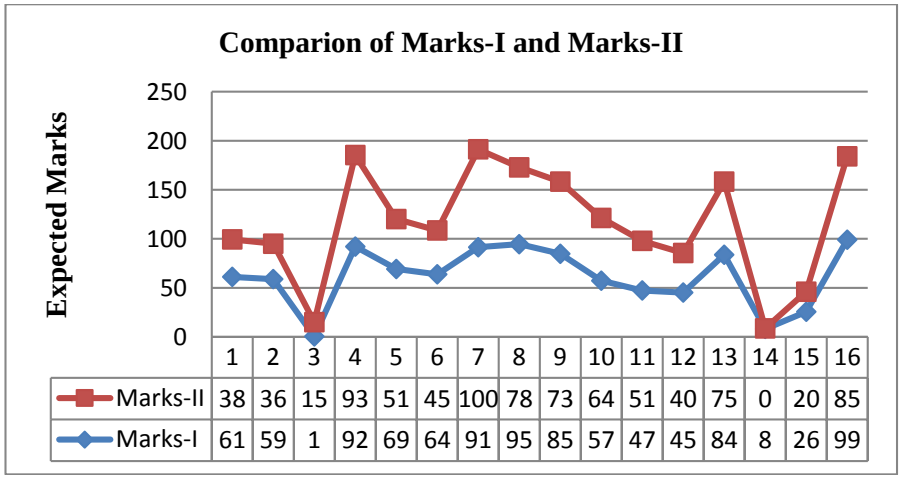


Fig 1: Comparison of Expected Marks.

2.7.1 Comment and Application

Now the next important question arises about the reliability of the marks derived by using the new parameters. To justify this situation we need to apply 'paired 't' test'. We apply it on the data that we have cited above and in order to generalize our conclusion we have verified the effectiveness in many cases by taking different samples and of different sizes. A few of such

Cases have been cited in the annexure and even some more in thesis that follows. To the height of that acceptance and rejection cases have been strictly examined at 0.01 level of significance. All these facts are duly equipped in the thesis. We proceed to justify our assumption statistically as follows

2.7.2 Paired t-test on Bio rhythm Marks-I and PCA Marks-II

In the context of our data we have the total routine for application of 'paired't' test. We define the null hypothesis as follows.

H_0: That there is no difference between the Bio rhythm Marks-I(d_{1i}) and PCA Marks-II(d_{2i}) [if there is a little difference then it is attributed to sampling fluctuation only.]

H_1: PCA Marks-II (d_{2i}) ≥ Bio rhythm Marks-I(d_{1i}) [PCA Marks-II (d_{2i}) are closer to actual marks then that of Bio rhythm Marks-I(d_{1i}) were.]

Level of Significance = α = 0.05

Data Tabularized:

Table 7: Paired 't' test data.

No. of Student 'i'	Predicted marks by Average of Bio-rhythms [Marks-I] (d_{1i})	Predicted marks by PCA [Marks-II] (d_{2i})	Difference of Marks-I and Marks-II, $d_i = (d_{1i}) - (d_{2i})$
1	61	38	23
2	59	36	23
3	1	15	-14
4	92	93	-1
5	69	51	18
6	64	45	19
7	91	100	-9
8	95	78	17
9	85	73	12
10	57	64	-7
11	47	51	-4
12	45	40	5
13	84	75	9
14	8	0	8
15	26	20	6
16	99	85	14

Calculation with necessary formula:

Based on above data we have following derivations.

Mean of differences = $\bar{d}$ = 7.437 units

Standard deviation S.D. = 11.280 units

No. of students n = 16

Degree of Freedom = n - 1 = 15

$$\text{S. E.}(\bar{\text{d}}) = \frac{\text{S. D.}}{\sqrt{\text{n}}} = \frac{11.280}{\sqrt{16}} = 2.280$$

$$t_{Calc} = \frac{\bar{d}}{S.E.(\bar{d})} = \frac{7.437}{2.280} = 2.637$$

Now for comparison purpose we look for't' value at 5% level from standard't' table.

$t_{Tab} = 2.131$

Conclusion of 'paired't' test'

At this point t_{Calc}(= 2.637) > t_{Tab} (=2.131). This falls in rejection Region. We reject H_0.We accept H_1 and conclude that in this case, PCA Marks-II (d_{2i}) is fairly closer to the actual marks obtained by the students. This establishes fairness in associating PCA to contraction of original variables.

3 Conclusion

All these facts and figures that we have generated by experimental work on different groups of the students on different time interval assert to astounding fact that application of PCA to transferring multiple variables of Bio-rhythmic characteristics, which are basically subtle in context to physical gestures and expressions, play a dominating role. New variables generated during the process prove justifying and apportioned quantum of original variables.

4. Vision

With some surpassing comments we would like to throw this problem to open mindedness with little reservation on our side.

1. Should we incorporate more bio-rhythms for explicit results?
2. We have a conjecture that 'Do we tend to accuracy as we include more bio-rhythms?'

5. References

1. Jaysan H, Shukla, Dr. Pradeep Jha J. Chi square test Applied on beta average square matrices of biorhythmic trends observed in classified groups. International Journal of Applied Mathematics & Statistical Sciences (IJAMSS) 2016; 5(6):71-78. ISSN (Print):2319-3972; ISSN (Online): 2319-3980, 2016; 5(6):71-78.
2. Jaysan H, Shukla, Dr. Pradeep Jha J. A study of spiritual factor in enhancing basic reflexes. International Journal of Advanced Science and Research. ISSN: 2455-4227, 2017; 2(4):13-18.
3. Hines TM. Comprehensive review of biorhythm theory. Psychology Department, Pace University, Pleasantville, NY. Psychol Rep. 1998; 83(1):19-64.
4. Pearson K. On Lines and Planes of Closest Fit to Systems of Points in Space. (PDF). Philosophical Magazine. 1901; 2(11):559-572. doi:10.1080/14786440109462720.
5. Miranda AA, Le Borgne YA, Bontempi G. New Routes from Minimal Approximation Error to Principal Components, Neural Processing Letters, Springer, 2008, 27(3).
6. Giannotti, Louis John. A Statistical Evaluation of the Theory of Biorhythms. Master's thesis, naval postgraduate school monterey calif