ISSN 2348-313X (Print)
International Journal of Life Sciences Research ISSN 2348-3148 (online)
Vol. 5, Issue 4, pp: (37-44), Month: October - December 2017, Available at: www.researchpublish.com

Validation of Hierarchical Cluster Solutions Based on Quantitative Pollen Characters of Some *Indigofera* Linn Species

Philip Onyango[1], Regina Nyunja[2], George Opande[3]

[1,3]Department of Botany, School of Physical and Biological Sciences, Maseno University, Kisumu, Kenya.
[2]Department of Biological Sciences, School of Physical and Biological Sciences, Jaramogi Oginga Odinga University of Science and Technology, Kisumu, Kenya

Abstract: **This study broadens the scope of representation of the genus *Indigofera* and provides information to facilitate the reexamination of the phenetic grouping of the *Indigofera* species. Quantitative pollen data were collected from acetolysed pollen of annotated herbarium specimens: *I. tanganyikensis, I. denroides, I. spicata, I. sabulata, I. volkensii, I. fulvopilosa, I. hirsuta, I. drepanocarpa, I. paracapitata, I. asparagoides, I. ambelacensis, I. vohemarensis, I. subargentea, I. circinella, I. zenkeri, I. arrecta,* and *I. vicioides*. Hierachical cluster algorithm validates both 3 and 4 phenons as significantly separated by the Polar axis, Equatorial axis and P/E. ANOVA (α=0.05, p=0, $F(2,14)$=48.67), (α=0.05, p=0, $F(3,13)$=64.41) validates a 3 and 4-cluster solution based on Polar axis respectively. ANOVA (α=0.05, p=0, $F(2,14)$=44.61); (α=0.05, p=0, $F(3,13)$=37.99) validates 3 and 4-cluster solution based on Equatorial axis respectively. The P/E had the lowest F statistic in both 3 and 4 cluster solutions. Tukey HSD post hoc analysis revealed significant differences in all the phenons based on the Polar axis and the Equatorial axis and not in the P/E ratio.The findings of this study are congruent with previous 3-phenon taxonomic treatments coalescing *I. ambelacensis, I. fulvipilosa* and *I. dendroides*.**

Keywords: **Cluster solutions, Polar axis, Equatorial axis, P/E,Validation, ANOVA, *Indigofera*.**

I. INTRODUCTION

Indigofera is the third largest genus in legumes (Schrire, 2005). There has been confusion in the estimation of the number of taxa in this group. This could be attributed to perceived similarities in their structural and reproductive biology and probably due to scarcity of palynological data on the genus despite the fact that pollen data could be very relevant in the delimitation and proper understanding of the *Indigofera* (Edeoga and Nwachukwu, 2006). The palynological studies in this research focuses on the pollen polar axis, equatorial axis and the P/E of the pollen of 17 species of the genus *Indigofera* Linn. Although the *Indigofera* has been subjected to phylogenetic, morphological, and molecular analysis, only 12% of the genus *Indigofera* has been sampled in the past taxonomic analyses. This leaves many taxonomic and biogeographical findings in need of reexamination (Lewis *et al*, 2005). Palynology generates data that have the potential to help reevaluate the past classifications of *Indigofera*. This study intends to broaden the scope of representation of the genus *Indigofera* and, the information obtained through cluster analysis will be of great use to plant taxonomists, as it will facilitate the reexamination of the past classifications, and thus help resolve the affinities of the *Indigofera* species affinities in conflict. The palynological data will also be available for integration with other data in order to come up with much more natural classification of the *Indigofera* species.

Factor and cluster analysis has been used to show similarity among pollen grain quantitatively in the *Saxifragaceae* (Hideux and Ferguson, 1977) and, this can bypass the need for qualitative descriptors. Cluster analysis has the advantage

of quick review of data, making specific purpose classification and also provides for a measure of dissimilarity as one of the most successful coefficients. Determination, profiling and interpretation and assessment of validity of the clusters is an intergral aspect of cluster analysis (Mooi and Sarstedt, 2011). Determination of cluster numbers in a classification is one of the biggest problems in taxonomy (Sneath and Sokal, 1973).

Validation analysis provides information to justify or not a given cluster solution. Cluster validation for quantitative variables may include an anlysis of variance (ANOVA). Calculating an ANOVA may indicate whether there is a significant variation of a given quantitative character among the clusters, but not between which groups the analysis occurs. A posthoc analysis is therefore done to determine where the differences lie, look for data patterns that were not specified apriori, and see relationships between the subgroups that would otherwise go undetected (Mooi and Sarstedt , 2011; Everitt and Landau, 2001). The Tukey HSD posthoc analysis of the ANOVA based on the quantitative variables accurately maintain the alpha levels, model assumptions of normality homogeneity, and independence. Although Turkey HSD was developed for equal samples or groups (Stevens, 1999) it is able to adapt unequal sample sizes using harmonic mean in the formula;

HSD=q√MSE/n*{q=Critical value for the studentised t range statistics, n*=Number of the scores used for calculating the group means}.

Given the true alpha level, multiple comparisons among the clusters on the quantitative data is estimated as $1-(1-\alpha)^{c}$, for LSD the true value would deviate from 0.05 α giving inaccurate values (Stevens, 1999; Everitt and Landau 2001).Use of LSD is prone to type I errors and P-values associated with multiple posthoc LSD tests are inaccurate (Howell, 2002). Its use in cluster validation would have requires a Bonferroni correction which maintains a group cluster differences as statistically significant when the P value is below α/k where k is the number of posthoc test to be carried out (Adkins *et al*, 2010). This study numerically investigated the validity of two different cluster solutions (three and four clusters) and comparability of results of hierarchical cluster analysis to an external *Indigofera* groupings based on continous pollen characters.

II. MATERIALS AND METHODS

Materials:

Pollen characteristics of the following 17 *Indigofera* species deposited in Maseno University Herbarium have been

studied: *I. tanganyikensis* Baker F, *I. hirsuta* L, *I. viciodes* Jaub, *I. dendroides* Jacq, *I. arrecta* Hotchst ex A.Rich, *I. drepanocarpa* Taub, *I. vohemarensis* Baill, *I. aspargoides* Taub, *I. ambelacensis* Schweinf, *I. paracapitata* Gillett, *I. fulvopilosa* Brenan, *I. volkensii* Taub, *I. spicata* Forsk, *I. circinella* Baker F., *I. zenkeri* Baker F, *I. subargentea De Wild*, *I. subulata* Vahl.

The major equipments used are: Nikon Type-102 Microscope,Water bath (Type: W600 DINI 2877-KI, GERMANY Model).

Chemicals and reagents:

All chemicals and reagents were of analytical grade.

Acetolysis and Microscopy:

Acetolysis and microscopy according to according to Reitsima (1969) method. Light microscopy was carried out according to Perveen and Qaiser (1998) method.

Measurement of pollen characteristics:

Measurement of pollen was done according to Faegri and Eversen (1989). Ocular micrometer was callibrated at 10 units= 25**µm**

Statistical analysis:

Hierarchical cluster algorithm according to IBM SPSS base 20 was used.

ISSN 2348-313X (Print)
International Journal of Life Sciences Research ISSN 2348-3148 (online)
Vol. 5, Issue 4, pp: (37-44), Month: October - December 2017, Available at: www.researchpublish.com

III. RESULTS

Table 1. Quantitative pollen characteristics of studied Indigofera species

Taxon	P(µm)	E(µm)	P/E
I. tanganyikensis	22.5(33.25±2.12)42.5	25(27.5±0.71)32.5	0.77(1.222±0.25)1.72
I. denroides	25(28.34±0.95)32.5	22.5(24.94±0.55)27.5	1(1.15±0.075)1.3
I. spicata	12.5(29.25±2.34)32.5	12.5(25±1.84)30	0.91(1.17±0.14)1.36
I. sabulata	20(28±1.70)32.5	15(24.25±1.78)32.5	0.9(1.17±0.18)1.7
I. volkensii	22.5(25.88±0.81)30	17.5(24±1.10)30	1(1.09±0.11)1.43
I. fulvopilosa	25(30.58±1.4)32.5	22.5(26±0.93)28.75	0.83(1.18±0.13)1.4
I. hirsuta	18.75(25.45±1.1)30	17.5(21.75±0.89)26.5	0.95(1.17±0.11)1.38
I. drepanocarpa	30(37.56±1.98)45	25(34.34±2.43)45	0.92(1.11±0.14)1.4
I .paracapitata	22.5(28.75±0.97)32.5	20(24.31±0.68)27.5	0.95(1.15±0.12)1.44
I. asparagoides	25(36.31±1.34)42.5	30(31.06±0.75)42.5	0.96(1.17±0.12)1.33
I. ambelacensis	20(29.56±2.55)45	16.3(25.13±2.19)37.5	0.96(1.18±0.095)1.4
I. vohemarensis	20(26.81±1.18)32.5	15(23.69±1.2)27.5	1(1.09±0.25)1.33
I. subargentea	21.3(25.56±1.13)31.3	17.5(24±1.31)28.75	0.9(1.08±0.14)1.36
I. circinella	25(27.44±0.82)31.3	17.5(22.25±0.88)26.3	1.1(1.28±0.15)1.67
I. zenkeri	22.5(26.75±0.91)30	21.3(24.69±0.76)28.8	0.87(1.1±0.14)1.43
I. arrecta	21.3(25.19±0.86)31.3	21.25(23.38±0.49)25	0.9(1.08±0.092)1.32
I. vicioides	15(23.63±1.53)25	17.5(21.425±1.41)25	0.85(1.14±0.183)1.67

Key:P=Polar axis; E=Equatorial Axis; P/E=Polar axis:Equatorial axis ratio.

The polar axis ranged from 12.5µm (*I.spicata*) to 45µm (*I.drepanocarpa*). The highest mean polar axis was observed in *I.drepanocarpa* (37.56 µm). Equatorial axis ranged from 12.5µm (*I.spicata*) to 45 µm (*I. drepanocarpa*). Highest P/E ratio was 1.72 observed in *I.tanganyikensis.*

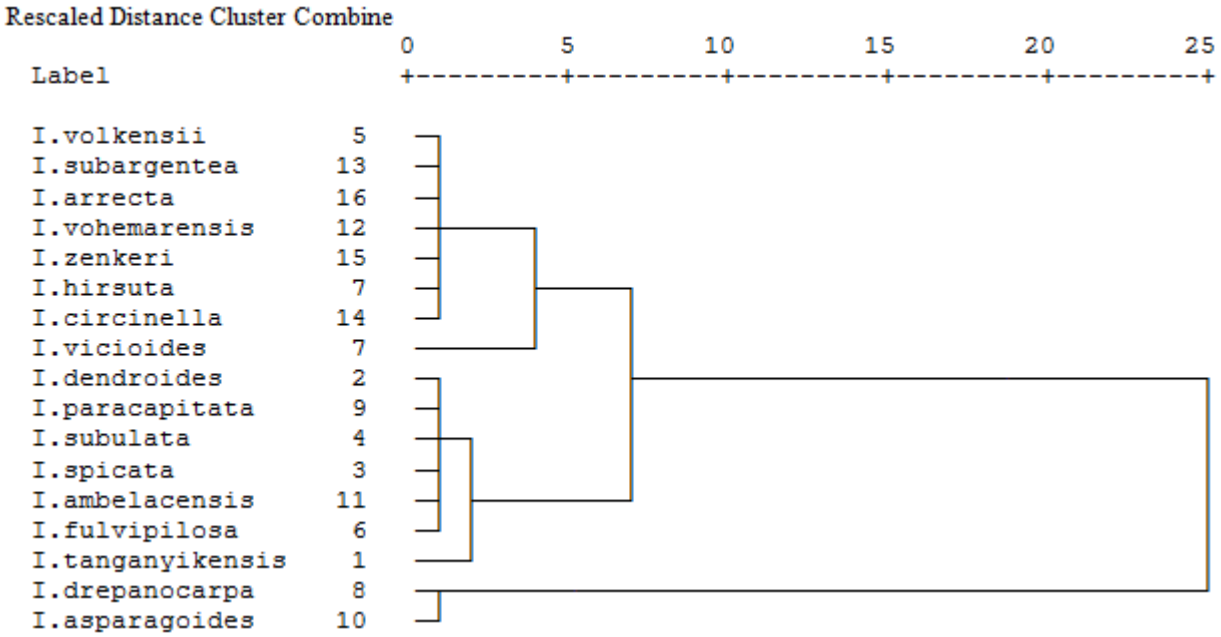

Fig. 1. Dendrogram for the 17 *Indigofera* species studied

In the dendogram above, the 3-cluster solution gave three clusters were formed with 7, 8, and 2 species in cluster 1, 2, and 3 respectively. Cluster 2 was the first do be formed followed by cluster 1 and finally cluster 3. The phenons formed were as follows: Cluster 2 (First to be formed) is compossed of 8 species: *I. volkensii, I.subargentea, I. arrecta I. vohemarensis, I. zenkeri, I. hirsuta, I. circinella, I. vicioides.* Cluster 1(The second to be formed) is composed of 7

ISSN 2348-313X (Print)
International Journal of Life Sciences Research ISSN 2348-3148 (online)
Vol. 5, Issue 4, pp: (37-44), Month: October - December 2017, Available at: www.researchpublish.com

species: *I. dendroides, I. paracapitata, I. subulata, I. spicata, I. ambelacensis I. fulvipilosa I. tanganyikensis.* Cluster 3 (The last to be formed) is composed of 2 species: *I. drepanocarpa, I. asparagoides.*

In the 4-cluster solution, the first cluster formed was cluster 3, composed of 8 species (*I. volkensii, I.subargentea, I. arrecta,I. vohemarensis, I. zenkeri, I. hirsuta, I. circinella,* and *I.vicioides*). The second cluster to be formed was cluster 2 composed of : *I. dendroides, I. paracapitata, I. subulata I. spicata, I. ambelacensis I. fulvipilosa.* Cluster 2 has a weight of 6 OTUs. The third cluster to be formed is cluster 1. Cluster 1 is made of *I.tanganyikensis* only.Cluster 4 is composed of only two species, *I. drepanocarpa,* and *I. asparagoides.*

Table 2. Cluster descriptives, 3 cluster solution

Character	Cluster	N	Mean	Std Dev	Std Error	95% interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Polar axis	1	7	29.68	1.79	0.68	28.02	31.33	28.00	33.25
	2	8	25.84	1.19	0.42	24.85	26.83	23.63	27.44
	3	2	36.94	0.88	0.62	29.00	44.88	36.31	37.56
	Total	17	28.72	3.85	0.93	26.74	30.71	23.63	37.56
Equatorial axis	1	7	25.30	1.13	0.43	24.26	26.35	24.25	27.5
	2	8	23.15	1.19	0.42	22.15	24.14	21.43	24.69
	3	2	32.7	2.32	1.64	11.86	53.54	31.06	34.34
	Total	17	25.16	3.25	0.79	23.49	26.83	21.43	34.34
P/E ratio	1	7	1.17	0.02	0.01	1.15	1.20	1.15	1.22
	2	8	1.13	0.07	0.02	1.07	1.19	1.08	1.28
	3	2	1.14	0.04	0.03	0.76	1.52	1.11	1.17
	Total	17	1.15	0.05	0.01	1.12	1.18	1.08	1.28

Table 2 above shows that cluster 1 of the 3-cluster solution is characterized by polar axis, equatorial axis, P/E ratio as follows: (29.68±1.79, 28-33.25), (25.30±1.13, 24.25-27.5) and (1.17±0.02, 1.15-1.22) respectively. Cluster 2 is made up 8 species: *I. volkensii, I. vicioides, I. sabulata, I. arrecta, I. vohemarensis, I. zenkerii, I. circinella,* and *I. hirsuta.* This cluster is characterized with a polar axis (25.84±1.19, 23.63-27.44), (23.15±1.19, 21.43-24.69), and P/E (1.13±0.07, 1.08-1.28). This is the cluster with the highest standard deviation in its polar axis, lowest equatorial axis mean. Cluster 3 is composed of only two species, *I. drepanocarpa,* and *I. asparagoides*. The cluster has the largest mean polar axis value (36.94±0.88, 36.31-37.56), equatorial axis (32.7±2.32, 31.06-34.34); the equatorial axis is the most varied.

Table 3. Cluster descriptives, 4-cluster solution

Character	Cluster	N	Mean	Std. Dev	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Polar axis	1	1	33.25	------	------	------	------	33.25	33.25
	2	6	29.08	0.93	0.38	28.10	30.06	28.00	30.58
	3	8	25.85	1.19	0.42	24.85	26.83	23.63	27.44
	4	2	36.94	0.88	0.63	28.99	44.88	36.31	37.56
	Total	17	28.72	3.85	0.93	26.74	30.71	23.63	37.56
Equatorial axis	1	1	27.50	------	------	------	------	27.50	27.5
	2	6	24.94	0.64	0.26	24.27	25.61	24.25	26
	3	8	23.15	1.19	0.42	22.15	24.14	21.43	24.69
	4	2	32.7	2.32	1.64	11.86	53.54	31.06	34.34
	Total	17	25.16	3.25	0.79	23.49	26.83	21.43	34.34
P/E ratio	1	1	1.22	------	------	------	-------	1.22	1.22
	2	6	1.17	0.02	0.01	1.15	1.18	1.15	1.18
	3	8	1.13	0.07	0.02	1.07	1.19	1.08	1.28
	4	2	1.14	0.04	0.03	0.76	1.52	1.11	1.17
	Total	17	1.1489	0.05401	0.0131	1.1212	1.1767	1.08	1.28

ISSN 2348-313X (Print)
International Journal of Life Sciences Research ISSN 2348-3148 (online)
Vol. 5, Issue 4, pp: (37-44), Month: October - December 2017, Available at: www.researchpublish.com

In the Cluster descriptives table above, cluster 1 does not have standard deviation, standard error, confidence intervals for all the parameters and was composed of only one OTU. The mean for the polar axis was 28.72µm and the standard deviation was 0.93, with a standard error of 0.63. The polar axes ranges from 23.63µm to 37.56µm. The highest mean polar axis is found in cluster 4. The highest standard deviation in polar axis was observed in cluster 2. The lowest polar axis was 23.63µm while the highest polar axis was 37.56µm observed in cluster 3 and 4 respectively.The mean equatorial axis was 25.16 µm, a standard deviation of 3.25, and a standard error of 0.79 respectively. Cluster 2 had the highest mean equatorial axis (32.7µm) and the most varied equatorial axis (Standard deviation, 2.32). The minimum equatorial axis was observed in cluster 3, the maximum equatorial axis was observed in cluster 4.The P/E ratio was the least varied in all the quantitaive parameters (Standard deviation, 0.05) and the lowest range between the minimum and maximum values(1.28-1.08=0.20). The lowest standard deviation in P,E and P/E was observed in cluster 2.

Table 4. ANOVA table for 3-cluster solution

Character	Variation	Sum of Squares	df	Mean Square	F	Sig.
P	Between Groups	207.779	2	103.889	48.67	0
	Within Groups	29.886	14	2.135		
	Total	237.665	16			
E	Between Groups	146.23	2	73.115	44.61	0
	Within Groups	22.947	14	1.639		
	Total	169.177	16			
P/E	Between Groups	0.008	2	0.004	1.45	0.267
	Within Groups	0.039	14	0.003		
	Total	0.047	16			

There was significant difference in all the parameters between the three groups of the 3-cluster solution using quantitative variables in Table 4. The polar axis had the highest F statistic (48.667). The variation between groups was higher than the variation within groups. The highest mean square between groups was observed in the polar axis (103.889) while the lowest mean square between groups was observed in the P/E ratio.

Table 5. ANOVA, 4-cluster solution.

		Sum of Squares	df	Mean Square	F	Sig.
P	Between Groups	222.683	3	74.228	64.41	0
	Within Groups	14.981	13	1.152		
	Total	237.665	16			
E	Between Groups	151.855	3	50.618	37.99	0
	Within Groups	17.322	13	1.332		
	Total	169.177	16			
P/E	Between Groups	0.011	3	0.004	1.28	0.322
	Within Groups	0.036	13	0.003		
	Total	0.047	16			

In Table 5 above, there was a significant difference in all the three parameters among the four groups in the 4-cluster solution. The polar axis had the highest F value in the three and four cluster solutions. P/E ratio had the lowest F statistic.

Table 6. Tukeys HSD post hoc analysis for ANOVA 3-cluster solution.

Variable	Cluster I	Cluster J	Mean difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
P	1	2	3.83696*	0.75617	0	1.8578	5.8161
		3	-7.25929*	1.17146	0	-10.3253	-4.1933
	2	1	-3.83696*	0.75617	0	-5.8161	-1.8578
		3	-11.09625*	1.15507	0	-14.1194	-8.0731
	3	1	7.25929*	1.17146	0	4.1933	10.3253
		2	11.09625*	1.15507	0	8.0731	14.1194
E	1	2	2.15616*	0.6626	0.015	0.422	3.8904

		3	-7.39571*	1.02649	0	-10.0823	-4.7091
	2	1	-2.15616*	0.6626	0.015	-3.8904	-0.422
		3	-9.55188*	1.01213	0	-12.2009	-6.9028
	3	1	7.39571*	1.02649	0	4.7091	10.0823
		2	9.55188*	1.01213	0	6.9028	12.2009
P/E	1	2	0.04582	0.02719	0.245	-0.0253	0.117
		3	0.03457	0.04213	0.697	-0.0757	0.1448
	2	1	-0.04582	0.02719	0.245	-0.117	0.0253
		3	-0.01125	0.04154	0.96	-0.12	0.0975
	3	1	-0.03457	0.04213	0.697	-0.1448	0.0757
		2	0.01125	0.04154	0.96	-0.0975	0.12

*The mean difference is significant at the 0.05 level.

Table 6 shows a post hoc Tukey test to determine whether the significance in the ANOVA test are true for the 3-cluster solution.The mean comparison between any two clusters based on polar and equatorial axis is significant but not the other comparisons. There was no comparison for a 4-cluster solution.There was a significant mean difference (at 0.05 level) in the comparison of polar and equatorial axis. There was no significant comparison for the P/E ratio in all the pairs of clusters, in the 3-cluster solution.The mean differences in polar and equatorial axis were bigger than the *t* critical value. The mean differences in all P/E ratio comparisons were less than the *t* value.

IV. DISCUSSION

Different aspects of cluster validation from the dendogram that have been explored here include:

How well the cluster fits the data without reference to external indices, comparison of two different cluster solutions (three and four clusters) to determine which one is better, determination of the correct number of clusters and comparing of results of cluster analysis to an externally known result. While attempt has been made to validate the cluster solution as reflected by the dendogram in Fig. 1, the following facts were put in to consideration: that there is no perfect classification, that species must not be grouped in a certain way as there does not exist in nature groups of individuals which must be grouped in only one way as objective uncontestable species (Schuh-Randall and Andrew, 2009). Dendrograms are provisional groupings or classifications (Wheeler, 2000).

Because dendrograms are provisional groupings there was need for dendogram solution validation. The ANOVA tables suggest that each of the clusters developed from the dendrogram are significantly different from each other based on the three quantitative characteristics (polar axis, equatorial axis, and the P/E ratio). The ANOVA as an internal index suggested in early studies (Everitt and Landau, 2001; Mooi and Starstedt, 2011) show that for a 3-cluster solution, the F value is highest for polar axis-64.412, 37.988 for the equatorial axis, and 1.280 for the P/E ratio. According to San *et al*, (2004) the higher the F value, the more the cluster is homogenous and the more distinct the cluster is from the others. The error sum of squares is a good internal measure for goodness of clustering since it does not refer to external information while determining the level of cohesion and separation of clusters (Zhao *et al*, 2009).

While in the 3-cluster solution the F ratio is 64.412, in a 4-cluster solution the F value is 48.667 for the polar axis. This could be attributed to increased homogeneity within the groups and decreased heterogeneity between groups in a 4-cluster model. It is also interesting to note that the P/E ratios, between group and within group sum of squares remain nearly the same 0.04 and 0.03, for both the 3 and 4 cluster solutions respectively. This could be attributed to proportionate increase in between and within group mean squares, and also the low standard deviation of the P/E ratio.The large F values observed are also evidence against null hypothesis (Wuensch, 2007) .The relatively higher F values in ANOVA suggest that the three quantitative parameters sufficiently contribute towards distinction and profiling of the clusters in both the 3 and 4 cluster solutions. The P/E contribution is much lower in the distinction of the clusters based on the low F ratios.

The Tukey HSD post hoc analysis revealed significant differences in all the groups based on the polar axis and the equatorial axis and not in the P/E ratio and the highest mean differences observed in cluster 2 and 3, polar and equatorial axis (11.09625 and 9.55188, respectively) and cluster 3 and 1, polar and equatorial axis, mean differences of 7.25929 and 7.39571, respectively. Although the use of ANOVA as an internal index for cluster validation gave a conclusion that, the parameters are sufficiently able to distinguish the subgroups, the turkey HSD indicate that the ability of polar axis and

equatorial axis to distinguish the clusters occurs in the following order: Cluster 2 and 3 (Highest ability), Cluster 1 and 3 (Higher ability), and Cluster 2 and 1 (High ability).

The tukeys HSD multiple comparison results are based on the magnitude of difference between the means of the clusters and the tukeys HSD critical value (Everitt and Landau, 2001; Mooi and Starstedt, 2011). The lack of significant differences between the groups on the P/E ratio is in consonance with the low F ratios in the ANOVA tabulation, and the low standard errors in the Tukey HSD tables. Because standard error are standard deviations of the sampling distribution of a statistic for example the means (Everitt and Landau , 2001) as in the means of three cluster solution, it implies therefore that the Low SE values for the P/E ratio among the clusters is an indication of small variation of the P/E among the clusters. The standard deviation of the P/E had the lowest values in all the clusters in the 3-cluster solution, 0.02, cluster one; 0.07-cluster two, and 0.04-cluster three. This suggests that all the P/E ratios are within the 95% confidence interval. Since SE is the most useful in the calculation of the confidence interval, for the 4-cluster solution, it could not be calculated as one of the clusters had one observation, and therefore according to Adkins *et al*, (2010), the standard error cannot be calculated.

The validation of the cluster solution was also done in comparison with the past phenetic and phylogenetic studies. According to Wu and Huang (1995) study of 15 *Indigofera* species, based on varied parameters, showed that characters have equal weight. However this study supports the separation of *I. hirsuta* and *I. spicata* into members of pollen type II and IV, respectively. The study also suggests the separation *I. subargentia* and *I. zenkeri* based on the polar axis, equatorial axis and the P/E ratio. These two species had been placed in Section Viscosae Rydb, one of the 47 sections in to which the *Indigofera* were classified by Schrire and Sims (1997). Even in a 3-cluster solution in this study, the two species would not agglomerate since cluster three and two in which *I. zenkeri* and *I. subargentia* belong have striking differences in the ranges and variances in the parameters used in their characterization. Although Ferguson and Strachan (1982) pollen classification into types 3, 4A, 4B and 4C included secondary characters of tectal and wall thickness, the ANOVA for 4-cluster solution validates four grouping system, even though a posthoc for location of differences could not be done.

V. CONCLUSION

The study finds that the grouping of the 17 *Indigofera* species based on P, E and the P/E ratio into three or four clusters is validated. Three and four cluster solutions are significantly separated by the Polar axis, Equatorial axis and P/E. ANOVA (α=0.05, p=0, $F(2,14)$=48.67), (α=0.05, p=0, $F(3,13)$=64.41) validates a 3 and 4-cluster solution based on Polar axis respectively. ANOVA (α=0.05, p=0, $F(2,14)$=44.61); (α=0.05, p=0, $F(3,13)$=37.99) validates both three and four cluster solution based on Equatorial axis respectively. The P/E had the lowest F statistic in both 3 and 4 cluster solutions. Tukey HSD post hoc analysis revealed significant differences in all the phenons based on the Polar axis and the Equatorial axis and not in the P/E ratio. The grouping of *I. ambelacensis*, *I. fulvipilosa* and *I. dendroides* in to a single group based on pollen charactersistics agrees with previous studies by Schrire and Sims (1997) on the *Indigofera* three phenon system.

REFERENCES

[1] Adkin, E., Jiehua, C., Winecki, J., Koinei,P., and Vijay M., (2010). Energy for sustainable development testing Institutional Biomass cooking stoves in rural schools. Elsevier Pp200-230.

[2] Edeoga, H.O. and Nwachukwu, C.U., (2006). Tannins, starch and crystals in some species of Indigofera L. (Leguminosae-Papilionoideae). International Journal of Botany. 2:159-162.

[3] Everitt, B. S. and Landau, S., (2001). A handbook of statistical analyses using SPSS. Chapman & Hall/CRC.

[4] Faegri, K. and Iversen, J., (1989). Textbook of pollen analysis. Blackwell, Oxford. Pp 295.

[5] Ferguson, I.K. and Strachan, F., (1982). Pollen morphology and taxonomy of the tribe Indigofereae (Leguminosae: Papilionoideae). Pollen Spores, 24: 171-210.

[6] Hideux, M. and Ferguson, I.K., (1977). The stereostructure of the exine and its evolutionary significance in Saxifragaceae sensu lato in: Stuessy, F.T. (2009): The systematic evaluation of comparative data. Columbia University Press. Pp 225-289.

[7] Howell , D.C., (2002). Statistical methods in psychology.Duxbury, Pacific Grove, California, 802pp.

[8] Lewis, G., Schrire, B., Mackinder, B. and Lock, M., (2005). Legumes of the World. Royal Botanic Gardens, Kew, Pp577.

[9] Mooi, E. and Sarstedt, M., (2011). A concise guide to market research. The Process, Data, and Methods Using IBM Statistics. Pp308.

[10] Perveen, A., and Qaiser, M., (1998). Pollen flora of Pakistan –viii Leguminosae (Subfamily-papilioinoideae).Turkish Journal of Botany. 22:73-91.

[11] Reitsima, T.J., (1969). Size modification of recent pollen grains under different treatments. Review of Paleobotany and Palynology. Elsevier publishing Company. Pp 175-200.

[12] San, Ohn Mar, Van-Nam , H. and Yoshiteru, N., (2004). "An alternative extension of the k-means algorithm for clustering categorical data." International Journal of Applied Mathematics and Computer Science, 14 (2):241-248.

[13] Schrire, B. D. and Sims, J.R., (1997). A Re-evaluation of Pollen Morphology and Taxonomy in the Tribe Indigofereae (Leguminosae-Papilionoideae) Kew Bulletin 52 (4): 841-878 URL: http://www.jstor.org/stable/4117816.

[14] Schrire, B. D., (2005). Indigofereae. Pages 361-365 in: Legumes of the world http://tolweb.org

[15] Schuh, Randall T. and Andrew V. Z., (2009). Biological Systematics: principles and applications (2nd edn.) Cornell University Press, ISBN 978-0-8014-4799-0.

[16] Sneath, P.H.A., and Sokal, R.R., (1973). Numerical Taxonomy: The Principles and Practice of Numerical Classification. WH Freeman and Co. USA. Pp: 573.

[17] Wheeler, Q. (2000). Species Concepts and Phylogenetic Theory: A Debate. Columbia University Press. ISBN 0231101430.

[18] Wu, M., and Huang, T., (1995). A Palynological Study of the Genus Indigofera (Leguminosae) in Taiwan. Grana palynologica Scandinavian University press, 34: 160 -181.

[19] Wuensch, K L., (2007). Multivariate Analysis with SPSS. East Carolina University. Available at http://core.ecu.edu/psyc/wuenschk/SPSS/SPSS-MV.htm.

[20] Zhao, Q., Xu, M., & Fränti, P., (2009). Sum-of-squares based cluster validity index and significance analysis. Adaptive and Natural Computing Algorithms, 313-322.