J. Appl. Environ. Biol. Sci., 2(10)500-507, 2012

ISSN 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Diversity Indices in Combining Classifiers for Cancer Diseases Data

[1]Gebriel M. Shamia, [2]Mufid M. EL-Mistiri, [3]Mohamed A. Bosheha

[1]Associate Professor at Department of Statistics, Faculty of Science, University of Benghazi, Libya
[2]Assistant Professor at Department of Medicine (Oncology), Faculty of Medicine, University of Benghazi
[3]Assistant Lecture at Department of Health Informatics, Faculty of Public Health, University of Benghazi

ABSTRACT

The behaviour of cancer diseases data from *the Benghazi Cancer Registry* as far as ecology through *Broken-stick* model for the spread of different species was used to study how diversity of the distribution of species of cancers is reflected in clusteringof zones and in different diversity measures related to such clustering.

The aim of this paper is to study the dissimilarity in cancer distribution in the departmental divisions, among which are evenness component by $J(\alpha,\beta)$ *Sp* and species richness component by $R(\alpha,\beta)$utilizing the generalized diversity index (*GDI*), $H(\alpha,\beta)$, at the actual values (α,β).

The result from this registry data showed that an interrelating in diversity among divisions in the areas (zones) of rare and common of species is exist in *five* clusters. It is indicated that the most powerful tools for discriminating the spatial variations of cancer species diversity are in the multivariate category and its superimposition with cluster analysis was recommended in order to obtain more information regarding the relationship between sites (zones). However, ordination by diversity measures is a more informative summary than cluster analysis for such data.

KEYWORDS : Environmental heterogeneity, Species diversity, MacArthur model.

INTRODUCTION

The environment which we live in, should be livable and habitable for human, and should be protected from pollution and the outbreak of diseases by all means. More attention has been paired in community to the measurement of species diversity then to almost any other parameter.

The environmental methods, studies and statistics are many, for example in this study we use biodiversity to study the cancer epidemic utilizing data from *Benghazi Cancer Registry* during the year 2003 where the registered cancer cases reached 1003 cases among males and females in 13 zones.

Diversity indices attempt to summarize both the total number of species in a population and the degree of evenness of the species relative abundances. There are two types of measurements of diversity by means of heterogeneity indices. Type-I indices are most sensitive to changes in the rare species in the community. However, type-II indices are most sensitive to changes in the more abundant species (common species). We can verify the behaviour of cancer data as far as ecology through *Broken-stick model*due to [13,14] and compare it to the probability values of cancer data at each departmental division. For more details see [19] and [1,2].

In this study we implement, another approach, one of the clustering analysis methods on the 13 zones to obtain a division through this analysis to identify the similar and dissimilar areas in the numbers and different species of cancers. See [3] and [4].

Cancer prevalence is the proportion of individuals in a population who at some stage during their lifetime have been diagnosed with cancer, irrespective of the date of diagnosis. Cancer registration in Northern Africa is still limited. See [5].

- Objectives of the Study

The aim of this paper is to study the behaviour of data from ecology locality and measuring diversity of cancer between areas (zones) by using heterogeneity measures of diversity for determining variation in species diversity. In addition to study the similarity and dissimilarity between these areas.

MATERIAL AND METHODS

- Study Area

The recording of information connected with new cases from 1.1.2003 to 31.12.2003 during one year. These cases were distributed between *Benghazi, Almarj, Abiar, Guba, Shahat, Baida, Darna, Tubruk, Ejdabia, Jalo, Tazrbo, Kufra*and*Brega.*

***Correspondence Author:** Gebriel M. Shamia, Department of Statistics, Faculty of Science, University of Benghazi, Libya. Email: gmshamia@yahoo.com.

These inflected cases (1003) from it (979, non-melanoma skin cancer). The distribution in both genders, according to recoding office visits to east areas which the record books cover; *theBenghazi Cancer Registry*. The estimated population in 2003 was approximately 1,632,051 people with a high proportion of children and young adults, according to the national information center estimations.

After visiting death department of civil record for gathering dead cases of cancer in 2003, nothing that recording models prepared on the base of the international criteria were vise made accurate made thedecreasing data completed. Made these date corresponded with the cases from different resources to avoid the doubling of the recording. Then codifying and putting them in the system, so the outside cases were expelled and also the cases which diagnoses before 2003. For more details see [5].

- **The Behaviour of Cancer Data**

A number of assumption and decisions regarding the data to be analyzed are required. The measurement of diversity by means of heterogeneity (dual-concept diversity) indices has proceeded along with two relatively distinct paths. The first approach is to use statistical sampling theory to investigate how communities are structured. The second approach is to information theory for appropriate measures of diversity. Both approaches are widely used in diversity studies. In fact, the measurement of diversity is not as simple as might be expected.

One model which suitably characterize the population may be required to investigate effects of evolutionary or environmental change is *Broken-stick* model due to [13], which defined as

$$\pi_i = \frac{1}{s}\sum_{r=i}^{s}\frac{1}{r} \;;\quad i = 1,2,3,\ldots,s.$$

Where π_i is the unequal relative abundances in the community.In practice π_i is replaced by:

$$p_i = \frac{n_i}{n} \;;$$

where

n_i: the frequency of every type.

n : total number of cancer cases in the quadrate (area).

Diversity Measures

The concept of diversity known as heterogeneity and their indices attempt to summaries both total number of species and the degree of evenness of relative abundances defined as heterogeneity indices [16].For more details see[11].

The first heterogeneity index is Shannon's index [22] which gives greater weight to rare species based on the information theory and other most popular one Simpson's index [23] which gives weight to common species .

The most satisfactory measure includes both Shannon's and Simpson's indices seems to be that measure suggested by [7].

$$H(\alpha,\beta) = \sum_{i=1}^{s}\pi_i^{\alpha}\left(-\ln\pi_i\right)^{\beta} \;,$$

for positive integer values of (α,β), where π_i is the *i*threlative abundance of population consisting of *s* species.

This measure have been investigated for (α,β)can take values in the real plane by [21] as a generalized diversity index (*GDI*).

Such index should satisfy two properties (P1) and (P2) due to [18] :

(P1) for given *s*, the index should be maximum when the π_i are equal.

(P2) if π_i are equal, the index should be an increasing function of *s*.

According to these two properties, the combined acceptable regions are obtained.The behaviour of $H(\alpha,\beta)$ according to the changes in the number of species, in the relative abundances or in both of these criteria is considered.

- **The Evenness Component of Diversity**

The equitability concept is based on the assumption that the measure of equitability partition total diversity into an effect of the species count and an effect of variations in relative abundances. For example, [12] introduced an equitability measure which was intended to partition the Shannon species diversity into a component depending on the species count and a component of equitability depending on the distribution of the relative abundances among the species in the sample.

To assess the extent to which $H(\alpha,\beta)$ is sensitive to evenness (equitability) define the *J-ratio*, $J(\alpha,\beta)$.

$$J(\alpha,\beta)=\frac{H_{Bs}(\alpha,\beta)}{H_{Ep}(\alpha,\beta)};$$

where *J* is an equitability measure as introduced by [17] for examining Shannon's index as a measure of the relative abundances of species in the community.([15], p.79) refers to this as Shannon's evenness index.

Notice that $J\leq 1$ for (α,β) in the acceptable region for rare species. The Shannon's index $\alpha=\beta=1$ would be one such an example. For the acceptable region for common species$J\geq 1$, and (α,β) =(2,0) is one such an example, corresponding to Simpson's index.

Sensitivity to evenness is indicated by the ratio *J*. If $J\cong 1$, for some (α,β) and given *s*, then $H(\alpha,\beta)$ is not a good discriminator between the equiprobable model and the broken-stick model. Accordingly small values of*J* in region rare species or large values of *J*in region common species indicate optimal choices of (α,β). In practice, of course, the estimated $H(\alpha,\beta)$ would be evaluated from the data.

[21] showed that, the optimal regions for different values of (α,β) as follows :

i)if $\alpha\cong 1$ andβlarge positive in the region for rare species.

ii) if$\alpha$is large positive and β is large negative in the region for common species.

As *s* increases the optimal region (i) moves along the upper curve boundary. By contrast, the optimal region (ii) remains unchanged. Thus the optimal values of α and β are affected by the change of equitability values as *s* increases.

Generally, *J*increases as *s* increases within regions for rare species and decreases for common species for fixed values of α and β.

It is known that as more samples are taken rare species will be added. [16] showed that*J*is more affected by species richness for Shannon's index than for Simpson's index for only a few species. This reflects the fact that an index based on the Shannon formula gives more weight to the equitability of the rare species than one based on Simpson's index.

- **The Richness Component of Diversity**

The effect of two different values of *s* for a specific model on the generalized diversity index is considered. Define $R(\alpha,\beta)$ as the ratio of $H_{Bs}(\alpha,\beta)$ with s_1 and $s_2$species for the *broken stick model*.

$$R(\alpha,\beta)=\frac{H_{Bs}(\alpha,\beta;s_1)}{H_{Bs}(\alpha,\beta;s_2)};$$

Good choices of (α,β) would give large values of $R(\alpha,\beta)$ within the region for rare species or small values within the region common species, since these would differentiate best between the two models for different *s*.

Within region for rare species large values of $R(\alpha,\beta)$ result if α is small and β is close to the upper boundary of the acceptance region. Within the region for common species large positive values of $\alpha$and negative va1ues ofβgive small $R(\alpha,\beta)$.

Methodology of the Cluster Analysis

Communities may be more or less similar, and ecologists often wish to express this similarity quantitatively and to classify communities on the basis of this similarity. Similarity measures may be binary based on presence-absence data, or quantitative, based on some measure of importance such as population size, biomass, cover, or productivity. For more details see [10].

- **Similarity Measures**

Fundamental to the use of any clustering technique is the computation of a measure of similarity or distance between the respective objects. The selection of a similarity or distance measure presents an interesting problem in cluster analysis. These measures can be separated into two broad classes based upon the quality of the data available. With data having metric properties, a distance type measure can be used, whereas with data having qualitative components, a matching type measure is appropriate.

The properties of the clustering algorithm are driven mainly by the choice of distance. In this section use Euclidean Distance assume data have been collected on *n* objects or individuals. Each object will be represented by a vector of observations $X' = (X_1, X_2, ..., X_p)$ on the *p* variables. We use the notation $X_i = (X_{i1}, X_{i2}, ..., X_{ip})$ to denote the measurements collected on the *i*thobject or individual. We have the familiar Euclidean distance between objects *i* and*j*:

$$d_{ij} = \left\{ \sum_{k=1}^{p} \left(X_{ik} - X_{jk} \right)^2 \right\}^{\frac{1}{2}} \quad ;$$

where d_{ij} denotes the distance between two objects *i*and*j*.

- **Constructing Linkage Distance**

Cluster analysis should be used to increase our ecological insight and not to baffle the reader, and for this reason simpler methods are often preferable to very complex ones, also cluster analysis is a method for generating classifications from a series of community samples. Many different types of cluster analysis have been developed, and there is no one "correct" or ideal system. Most ecological data can be classified simply by average linkage clustering (*UPGMA*) and this technique is recommended for general usage.

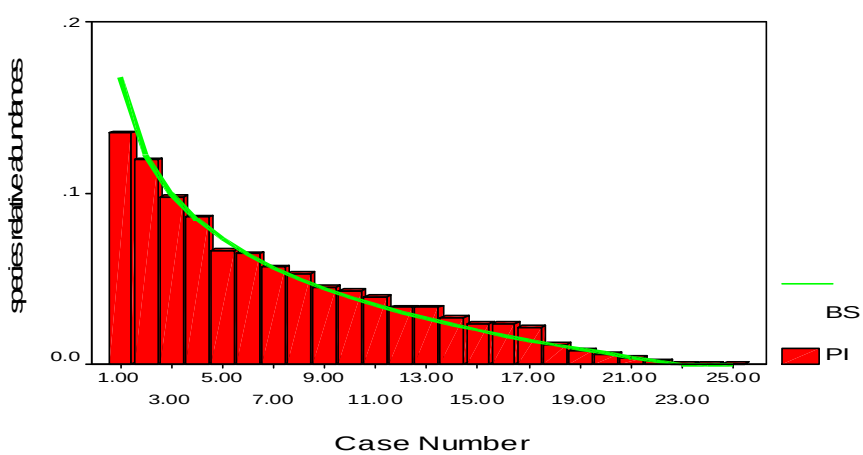
Assume that two observations or clusters, *p* and *q*, are combined in a cluster denoted by $p \cup q$. Let $d(p,q)$ denote the distance between clusters *p* and *q* and n_p and n_q the number of observations belonging to clusters *p* and*q*, respectively. For choosing the method or approach which can be used in clustering, it was defined as the agglomerativecoefficient by[20], which measures the clustering structure of the dataset.

RESULTS AND DISCUSSION

In this study we divided our research area into 5 departmental divisions as follow: *Benghazi, East Benghazi-A, East Benghazi-B, East Benghazi-C, and Southwest Benghazi-D.*See [8].

(Fig. 1) represents the pattern of the relative abundances for the broken-stick model (*Bs*) along with p_i values for *Benghazi* area as an example.

It is obvious that, the same patterns of broken-stick model (*Bs*) and p_i values for five areas are similar for muse cases. Again it is clear that the similarity is increasing between *Benghazi* and east *Benghazi-A* as will as east *Benghazi-B.*



(Fig. 1) The pattern of p_i and broken-stick model(Bs) in Benghazi area.

- **Comparative Study between Benghazi Quadrate and Others**

In this stage we are attempt to compare the measurements $H(\alpha,\beta)$, $J(\alpha,\beta)$ and $R(\alpha,\beta)$ for the five quadrates (areas). (Table 1) shows the results of using the generalized diversity index in the rare region

for some values of (α,β). we notice that, the value of indices $H(0.25,0.50)$ and $H(1,1)$ are closed to each other for *5* areas, but *Benghazi* area had more diverse for $H(0.25,0.50)$ than the others of cancer types.

For equitability measure by *J*-ratio of these types on cancers in *Benghazi* quadrate was less equitability at (0.25,0.50) and followed by *Benghazi-B,Benghazi-A, Benghazi-C* and *Benghazi-D*.

For the species richness measure by *R*-ratio, we notice that, the nearest quadrates for *Benghazi* was *Benghazi-C, Benghazi-B, Benghazi-A*,and *Benghazi-D* respectively.

(Table 1)Results for using the generalized diversity indexin rare region.

Index	(α,β)	S=22 *Benghazi*	S=21 *A*	S=21 *B*	S=20 *C*	S=19 *D*
H (α,β)	(1.00,1.00)	2.7778	2.806	2.785	2.800	2.833
	(0.25,0.50)	16.859	16.486	16.473	15.806	15.353
J (α,β)	(1.00,1.00)	0.899	0.923	0.915	0.935	0.962
	(0.25,0.50)	0.944	0.963	0.962	0.966	0.983
R (α,β)	(1.00,1.00)	1.026	1.016	1.024	1.018	1.006
	(0.25,0.50)	1.067	1.091	1.092	1.138	1.172

(Table 2) shows the results of using the generalized diversity index in rare region for some values of (α,β). we notice that, the value of indices $H(1.75,-2)$ and $H(2,0)$ are closed to each other for *5* areas, the but *Benghazi-D* area had more diverse than the rest of cancer types.

The *J*-ratio measure which represents the equitability for cancer types more for *Benghazi, Benghazi-B, Benghazi-A, Benghazi-C* and *Benghazi-D*.

The species richness of cancer types which expressed in *R*-ratio was close to each other for *Benghazi* and *Benghazi-B* there others.

(Table 2) Results for using the generalized diversity indexin common region.

Index	(α,β)	S=22 *Benghazi*	S=21 *A*	S=21 *B*	S=20 *C*	S=19 *D*
H (α,β)	(2.00,0.00)	0.074	0.071	0.075	0.068	0.638
	(1.75,-2.0)	0.024	0.022	0.025	0.02	0.018
J (α,β)	(2.00,0.00)	1.634	1.493	1.569	1.359	1.212
	(1.75,-2.0)	2.347	2.045	2.265	1.70	1.402
R (α,β)	(2.00,0.0)	0.904	0.946	0.899	0.988	1.054
	(1.75,-2.0)	0.837	0.900	0.813	1.011	1.139

- **General Comment on Diversity**

Through the aspect of environmental study for cancer disease, we note that, small values of ($0.25 \le \alpha \le 0.5, 0.25 \le \beta \le 1$) indices gave more diverse (heterogeneity) of cancer types, especially for $(\alpha = 0.25, \beta = 0.50)$, than others in the rare region as a rare measures.Notice that, sensitivity to evenness is examined and as it can be seen from results most of *J*-values are closed to one. As a result of this, evenness measures have not been used to much advantage in community analysis.

Again, we note that, values of ($1.8 \le \alpha \le 2, -2 \le \beta \le 0$) indices gave more diverse of cancer types, especially for $(\alpha = 1.75, \beta = -2)$, than others in the common region as common measures.Sensitivity to evenness is examined and as it can be seen from results the *J*-values are greater than one.These results will be generalized to all departmental divisions.

From the comparison between departmental divisions for rare types, through (*GDI*) $H(\alpha,\beta)$, *J*-ratio evenness measure, and *R*-ratio richness measure. We noted that in the rare region at value $(\alpha = 0.25, \beta = 0.50)$, the index given more diversity (heterogeneity) since we given the divisions according to index value respectively *Benghazi, Benghazi-A, Benghazi-B, Benghazi-C* and *Benghazi-D*.While from sensitivity to evenness measures *J*-ratio values close to one although it cannot given homogeneity in p_i values for rare types. But when compare the divisions in common types, the index given more diversity at $(\alpha = 1.75, \beta = -2)$. We noted that evenness in *J*-ratio between divisions are which be respectively *Benghazi, Benghazi-B, Benghazi-A, Benghazi-C* and *Benghazi-D*.

Again the richness values of *R*-ratio are close to each other. This means that the cancer diseases are distributed with the same number of species of cancer.

- **Analytical Cluster Results**

For the value of agglomerative coefficient, the Wards error sum of squares method were a best technique, agglomerative coefficient equal to 0.88, can be used for clustering method after choosing the best choice of cluster number in this study. See [24].

In this study we deal with all 13 areas (zones) on the bases each of this area can be revealing in a cluster. Deeming the appropriates number of cluster, otherwise areas which can be combined in clusters which every cluster contains a group of areas according to similarly within groups depends on the number of cancer cases in each one.

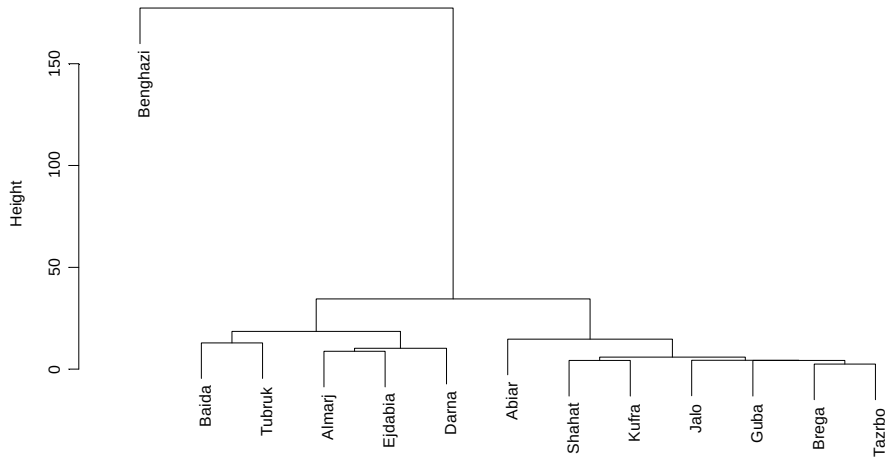
Before using the previous methods (R^2 andHartigan's Criterion) for determent the number of clusters, we can assume that from the logical view in our study we can apply these methods on 2 to 7 cluster as in (Table 3). See [6].

(Table 3)Results of applying the stopping rules to choose the number of clusters, m.

Number of clusters	*R*-square	|ln (*SS(B)/SS(W)*)|
7	0.422	0.315
6	0.405	0.386
5	0.371	0.528
4	0.298	0.856
3	0.246	1.119
2	0.010	4.569

(Table 3) shows the obtained values of R-square and Hartigan's Criterion method on the number of suggested clusters.We obtained that the number of appropriate cluster is *5* cluster which can be used by using Wards error sum of squares method.

Then from previous figures it can be obtained that the number of appropriate cluster is *5* cluster which can be used by wards error sum of squares method.



(Fig.2)Tree diagram resulting from the cluster analysis.

(Fig.2) shows the Algorithm statistic resulted through the usage of cluster analysis to divide the areas of study according to their similarity in the number of cases of each cancer species, which requires medical comments for these results and interpretation thereof, where the resulted divisions.

The first cluster consists of *Benghazi* area only due to high ratio of cases and its life style as a big city and its population comparison to other areas.

Second cluster consists of *Baida* and *Tubruk* areas due to the similarity between the two cities in the number of cases despite the large difference in population which could be a result of imperfection in gathering the register data; there is also an advanced central hospital in *Tubruk* which attributed to the existence of more cases. This result need re-evaluation or another parallel study to be done at the same time, to be sure from data, because apparently no clear or real explanation for increasing the incidence of cancer in *Tubruk*, hence it is similar to the incidence in *Baida* city also the similarity of the simple life style between each other.

Regarding the third cluster which includes three areas (*Almarj, Ejdabia, Darna*) and by more focusing in the cluster, we notice that the similarity between *Almarj* and *Ejdabia* is more due to close numbers of population and similar simple life styles, but *Darna* which is in the same geographical area looks less than *Almarj* and *Ejdabia* in the similarity of number of cancer cases could be due to the size of population in *Darna*which is less than *Almarj* and *Ejdabia*.

The fourth cluster consists of *Abiar*, while the fifth cluster consists of six similar areas (*Shahat, Kufra, Jalo, Guba, Brega,* and*Tazrbo*), this similarity is because these cities are considered small (villages) that have smaller numbers of cases and the simplicity of life style in these areas is differ from the big cities.

In the division of *Benghazi*, *East Benghazi-A*, and *East Benghazi-B*, a similarity was found in the behaviour between the probability values of cancer in these divisions and the probability values of broken-stick model, with dissimilarity in divisions that contains small values of cancer data.

Cancer data had been studied through the clarification of the probability values pattern *Broken-stick Model* (*Bs*) with the probability values (p_i) of the five departmental divisions to determine the data behaviour of these divisions from the environmental side, as the results were showing more similarity between *Benghazi* and *East Benghazi-A* and *East Benghazi-B*. One can also notice that the *SouthwestBenghazi-D* does not give the same behaviour (*Bs*-Model) as in other areas; this was due to the small number of cases of cancer species.

Through the measurements of heterogeneity we notice that there is aheterogeneity inside each division with a similarity of heterogeneity among the divisions. The diversity of cancer varies from country to another and from city to city due to its tie to environmental factors and different dieting habits of different people. Due to importance of knowing of whither there are differences in the diversity of different cancer species among the society within the research areas, and the utilization of data recorded by *Benghazi Register of Cancer* during the year 2003.

In this study, the heterogeneity in cancer diversity through the generalized diversity index (*GDI*) is considered to be as a measure to the heterogeneity rare and common living species.For the cancerdata in the region of rare species,the indices for small values of ($0.25 \le \alpha \le 0.5,\ 0.25 \le \beta \le 1$)showed more diverse of cancer types, but when measuring the sensitivity to evenness is most of *J*-values are closed to one. As a result of this, evenness measures have not been used to much advantage in community analysis.

Studying the common species within the departmental divisions at values ($1.8 \le \alpha \le 2, -2 \le \beta \le 0$), the indices gave more diverse of cancer types, but in reviewing the *J*-ratio value for measuring the equitability which gives results higher than one.

As for studying the species richness by *R*-ratio, this measure is not considered to be reliable, due to the interrelating of cancer species, this is due to homogeneity in the *Libyan* society as far as environmentally and behaviour of life style.

Regarding the comparison between the departmental divisions in both cases (rare and common), the index value shows an interrelating in the areas between each other as far as diversities, but as far as comparisons that concern *J*-ratio of equitability, *Benghazi* area gives a minor difference from the rest areas.

Diversity between areas were studied using a statistical method which is the cluster analysis to reveal the similar and dissimilar areas as far as species and number of cancer within through the resulted division of cluster analysis; as a result we have obtained five clusters.The similarity is because these cities are considered small (villages) that have smaller numbers of cases and the simplicity of life style in these areas is differ from the big cities.

Therefore, these resulted diversities are due to diversity of reasons, the environment causing the cancer (directly or indirectly), this is very important in following the coming years which will be rely on as a basis for treatment and prevention plan.

CONCLUSION

Through the comparison between divisions, we notice that there is an interrelating in diversity among divisions in the area of rare and common of species, this is due to the diversity of the number of infected cases among each division despite the similarity in their cancer species from the descriptive aspect.

Also from the aspect of environmental study, we notice the diversity of rare and common species with unevenness in the rare species, but as far as the richness measure is concern, it is unreliable due to the interrelating of the number of cancer species among the divisions.

The resulted cluster analysis division which divides the areas according to the similarity in the number of cancer cases for each species, and this is what we have obtained through clustering where the results thereof revealed and exemplary division from the statistical and medical aspects.

Also the homogeneity in Libya society from the structure of the population point of view, eating habits and not rarity infectious diseases in one city than other, makes the number of population is the main reason for the similarity between the regions regarding the number of cancer cases.

REFERENCES

[1] Baczkowski, A.J., Joanes, D.N. &Shamia, G.M. (1997). Properties of Generalized Diversity Index. *J.Theor. Biology,*188, 207-213.

[2] Baczkowski, A.J., Joanes, D.N. &Shamia, G.M. (2001). *Effect of Pooling Species on Good's Generalized.* Internal Report Stat 01/01, Department of Statistics, University of Leeds.

[3] Biliotti, D., Antonini, G. & Thiran, J.P. (2005).*Multi-layer Hierarchical Clustering of Pedestrian Trajectories for Automatic Counting of People in Video Sequences*. wacv-motion, pp. 50-57.

[4] Desrochers, R.E. &Anand, M. (2004). From Traditional Diversity Indices TO Toxonomic Diversity Indices. *Intern. J.Ecology and Environ.Sn.* 30, 85-92.

[5] El-Mistiri, M., Verdecchia, A., Rashid, I., El Sahli, N., El Mangush, M. & Federico, M.(2006). Cancer Incidence in Eastern Libya: The First Report from the Benghazi Cancer Registry, 2003. *Wiley-Liss.*120, 392–397.

[6] Fernandez, A. & Gomez, S. (2006). *Solving Non-Uniqueness in Agglomerative Hierarchical Clustering Using Multi-dendrograms.*arXiv: cs/0608049v1.

[7] Good, I.J. (1953).The population frequencies of species and the estimation of population parameters. *Biometrika*, 40, 237-64.

[8] Horn, H.S (1966). Measurement of (Overlap) in Comparative Ecological Studies. *Amer.Natur.*, 100:419-424.

[9] Kohn, A.J. (1959). The Ecology of Conus in Hawaii. *Ecolog*. Monoger, 29:47-90.

[10] Krebs, C.J. (1989). *Ecological Methodology*. Harper Collins Publishers.

[11] Laxton, R.R. (1977).The Measure of Diversity. *J. Theor. Biol*.,70, 51-67.

[12] Lloyd, M. &Ghelardi, R.J. (1964). A Table For Calculating The "Equitability" Component Of Species Diversity. *J. Ecology*, 33, 217-25.

[13] MacArthur, R.H. (1957).Onthe relative abundance of Bird species.*Proc. Nation. Acad. of Sc, USA*, 43, 293-5.

[14] MacArthur, R.H. (1965). Patterns of Species Diversity. *Biological Reviews of the Cambridge Philosophical Society*, 40:510-533.

[15] Magurran, A. E. (1988). *Ecological Diversity and Its Measurement*. Croom Helm, London.

[16] Peet, R.K. (1974).The Measurement of Species Diversity. Ann.Rev.Ecolog. and Syst., 5, 285-307.

[17] Pielou, E.C.(1966).The Measurements of Diversity in Different Types of Biological Collections . *J.Theor. Biology,* 13, 131-44.

[18] Pielou, E.C.(1975). *Ecological Diversity.* Wiley, London.

[19] Rejmanek, M. (1981). Corrections To The Indices of CommunityDissimilarity Based on Species Diversity Measures. *Oecologia (Berl.).* 48,290-291.

[20] Rousseeuw, P.J, visual display for hierarchical classification, in: E. Diday, Y. Escoufier, L. Lebart, J. Pag_es, Y. Schektman, & R. Tomassone (Eds.) (1986). *Data Analysis and Informatics*. North- Holland, Amsterdam, 743-48.

[21] Shamia, G.M. (1997). *Statistical Properties of the Generalized Diversity Index.* PhD Dept. of Statistics, University of Leeds.

[22] Shannon,C.E. (1948).A Mathematical Theory of Communication.*Bell Syst. Tech. Journal*, 27, 379-423.

[23] Simpson, E. H. (1949).Measurement of Diversity. *Nature*, 163, 688.

[24] Zorn, C. (2003). *Agglomerative clustering of Rankings Data, With an Application to Prison Rodeo Events.*Unpublished Working Paper. Emory University.