

**ANALYSIS OF DENSITY AND
DIVERSITY DATA OF MARINE
NEMATODE ALONG THE KENYAN
COAST**

By

David Njauini Gitahi

I56/80167/2012

A project submitted to the School of Mathematics in
partial fulfillment for the degree of Master of Science in
Biometry

University of Nairobi

June 2014

DECLARATION

This project is my own work carried out at University of Nairobi during the 2013-2014 academic year and has not been presented for award of any other degree.

David Njauini Gitahi

Signature: Date.....

This project has been submitted for examination with approval of my supervisors

Mr.J.N.Mwangi
Chief Biometrician
Kenya Agricultural Research Institute

Signature: Date:

Prof M.M.Manene
School of Mathematics
University of Nairobi

Signature: Date.....

Dr Agnes Muthumbi
School of Biological Sciencies
University of Nairobi

Signature: Date.....

DEDICATION

I dedicate this project to my Late father Gitahi for the encouragement and support to further my Education. To my mother Joyce, wife Ruth for being there for me when I needed their support

ACKNOWLEDGMENTS

I would wish to acknowledge everyone who assisted me and contributed to the successful completion of this study.

To start with are my supervisors Mr.J.N. Mwangi and Prof Manene for their continuous support, contribution and guidance throughout the study.

I wish to express my special thanks to Dr A.Muthumbi for providing the data and guiding me through the study.

To my classmate Sheilla, Mary, and Tifow, who gave moral support and encouragement receive my heartfelt thanks.

ABSTRACT

The data used in this study was obtained from Muthumbi, 1998 (PhD thesis) and Muthumbi et al, (2004) based on the Western Indian Ocean off Kenyan coast. The effect of the depth, locality and seasons were examined. The physical factors affecting the productivity in the sea i.e. oxygen concentration, food availability and graunometry were measured. Analysis of Variance and Analysis of Covariance were run to determine the factors that had significant effect on the Nematodes density. The composition and diversity of the nematode was examined by use of Hierarchical agglomerative clustering. Dendrogram was used to represent the information diagrammatically.

The effect of Transect and Depth was highly significant on the Nematode densities with seasons not having a significant effect. The nematode density varied form 112-1350 ind/cm². The density decreased with increase in water depth up to 1338m and then there was slight increase. This increase was a result of the oxygen concentration increasing after the 1000m depth. The *Monhyestra* genus dominated the Western Indian Ocean and it dominated the deep slope. The upper slope was dominated by *Terschellingia* genus and the middle slope was dominated by *Sabatieria*, *Deptonema* and *Halalaimus* genus where the oxygen concentration was at minimal. Most of the genera were identified in the upper slope between 50m-200m.

Table of Contents

CHAPTER 1	6
INTRODUCTION.....	6
1.1 BACKGROUND INFORMATION.....	6
1.2 Statement of the problem	7
1.3 Objectives.....	8
1.4 Justification of the study	8
1.5 METHODOLOGY	8
1.5.1 Source of the data.....	8
1.5.2 Study area	9
1.5.3 Sampling.....	9
1.6 DATA ANALYSIS.....	10
CHAPTER 2	11
LITERATURE REVIEW	11
CHAPTER 3	15
EXPLORATORY DATA ANALYSIS	15
3.1 Introduction	15
3.2 Non – graphical procedure	15
3.3 Graphical procedure.....	17
3.4 Data transformation	19
3.5 Significance test for normality.....	21
CHAPTER 4	22
GENERAL LINEAR MODELS	22
4.1 Introduction.....	22
4.2 Parameter Estimation	23
4.2.1 Ordinary least squares method.....	23
4.3 Analysis of Variance.....	25

4.3.1 Analysis of variance without interaction	26
4.3.2 Analysis of variance with interaction	28
4.3.3 Fitting a quadratic model	30
4.4 Testing the correlations of the covariates	33
4.5 Environmental and physical factors	34
4.6.1 ANACOVA model.....	37
CHAPTER 5	39
CLUSTER ANALYSIS	39
5.2 Dissimilarity measure.....	39
5.3 Similarity measure	41
5.4 Agglomerative Hierarchal Clustering Method	43
5.5 Wald's Method	44
5.6 Dedograms.....	45
5.7 Nematode density in relative to the depth	47
CONCLUSION.....	48
REFERENCES	50
APPENDICES	53

List of Tables

TABLE 1: SUMMARY STATISTICS FOR THE NEMATODE DENSITIES BY SEASONS	15
TABLE 2: SUMMARY STATISTICS FOR THE NEMATODE DENSITIES BY DEPTH	16
TABLE 3: SUMMARY STATISTICS FOR THE NEMATODE DENSITIES BY TRANSECTS	17
TABLE 4: SIGNIFICANCE TEST FOR KURTOSIS AND SKEWNESS.....	21
TABLE 5: ANALYSIS OF VARIANCE.....	27
TABLE 6: TRANSECT COMPARISON AT 0.05 LEVEL OF SIGNIFICANCE	27
TABLE 7: ANALYSIS OF VARIANCE WITH INTERACTION OF SEASONS AND DEPKLAS	30
TABLE 8: VARIABLES SELECTED TO BE INCLUDED IN THE MODEL	33
TABLE 9: ANALYSIS OF COVARIANCE TABLE	38

List of Figures

FIGURE 1: BOX PLOT FOR THE NEMATODE DENSITY	17
FIGURE 2: HISTOGRAM FOR THE NEMATODE DENSITIES	18
FIGURE 3: NORMAL QUARTILE PLOT FOR THE NEMATODE DENSITIES	19
FIGURE 4: BOX PLOT FOR THE TRANSFORMED NEMATODE DENSITIES	20
FIGURE 5: HISTOGRAM FOR THE TRANSFORMED NEMATODE DENSITIES	20
FIGURE 6: NORMAL QUARTILE PLOT FOR THE TRANSFORMED DATA	21
FIGURE 7: INTERACTION PLOT FOR SEASONS AND DEPTHCLASS	29
FIGURE 8: CURVE FIT FOR NEMATODES SEASON1	31
FIGURE 9: CURVE FIT FOR NEMATODES SEASON2	32
FIGURE 10: CURVE FIT FOR THE NEMATODE DENSITIES VERSUS THE DEPTH	32
FIGURE 11: OXYGEN CONCENTRATION IN RELATION TO DEPTH	34
FIGURE 12: SEDIMENT PHAEOPIGMENTS IN RELATION TO DEPTH, TRANSECT AND SEASONS.....	35
FIGURE 13: DNA-RNA RATIO IN RELATION TO DEPTH, TRANSECT AND SEASONS.....	35
FIGURE 14: SAND AND SILT COMPOSITION IN RELATION TO DEPTH, TRANSECT AND SEASONS	36

CHAPTER 1

INTRODUCTION

The data used in this study focuses on the composition and density of the free living nematodes that are found in the sea. It was done in Western Indian Ocean (WIO) of Kenya.

1.1 BACKGROUND INFORMATION

Free living marine nematodes are considered the most abundant meiofauna taxa, with estimate that 80% of all meiofauna are nematodes (Bongers, 1988). According to (Lamshead, 1993) it is estimated that there are 1×10^8 nematode species in the deep sea, but the numbers of described species are only 20,000. They can be classified into two classes of selective feeders and non-selective feeders. The non-selective feeders density increases with increase in sediment depth (Ingels, 2011). There is also known to decrease with decrease in food availability (Vincx et al., 1994) and hydrodynamic disturbances (Ingels et al., 2011)

Due to their short lifecycles, high turnover rates, lack of planktonic stages, nematodes are particularly sensitive to changes in environmental parameters and thus useful in assessing environmental disturbances (Kennedy and Jacoby, 1999). Nematodes help in breakdown of detrital organic matter and recycling of nutrients, thereby enriching the coastal waters to support marine benthic production. They also participate in energy transfer through the ecosystem and are important link between primary producers and higher trophic levels in benthic systems (Giere 2009).

Western Indian Ocean (WIO) region include the western part of the Indian Ocean and the Red sea, Persian Gulf and Arabian Sea. The ocean continental shelves are narrow ranging between 7-80 kilometers in width. The average depth of the the Indian Ocean is 3890m. It is known to be oligotrophic area (Semeneh et al., 1995) with oxygen levels declining with depth along the upper

slope down to a minimum at 1000m. The climate is affected by the monsoon winds. Strong southern eastern monsoon winds blows from April to October and northern eastern monsoon winds blows from November to March. In the months of May and November there is an inter monsoon period. The water circulation is mainly influenced by inflow from the Red sea and Atlantic Ocean. It is characterized by minimum surface temperature of 22 °C and a maximum of 28 °C. The surface salinity ranges between 32 to 37 PSU. The deep water is characterized with low temperature of 0.1 – 2 °C, low salinity (34.7 PSU) and highly dissolved oxygen concentration (4.7 ml/l) (Rao and Griffith, 1998)

Western Indian Ocean (WIO) is a home to fascinating range of marine life, including whales, dolphin, crustaceans, fish, sponges and many others. These marine resources are of significant economic importance. With a clear understanding of the productivity potential of WIO, it will help in formulation of necessary policies that will enhance the productivity.

1.2 Statement of the problem

The coastal region accounts for a population 3,325,307 which is about 10% of the Kenyan population and most of the population depend on the WIO either directly or indirectly. Lack of enough relevant studies on WIO has caused poor documentation on factors affecting its productivity. Besides the marine life found in the deep sea, recent discovery of oil and gas reserves about 100km off the Mombasa coastline has drawn attention from many investors all over the world. A significant increase in industry interest offshore Kenya is expected. This has increasingly made the WIO to be a major potential source of revenue for the coastal people and Kenya as a whole. Thus more studies to investigate the physical and environmental affecting its productivity is required. The Nematodes density and diversity is a good indicator of how the temperature, granulometry, oxygen concentration and food availability affect the deep sea

productivity. Analysis of variance procedure was used to investigate the density and Cluster Analysis procedure was used to determine the diversity of the Nematode densities.

1.3 Objectives

The general objective is to establish how the benthic fauna distribution in Western Indian Ocean is affected by physical factors, seasons and food availability with a focus on nematodes.

The specific objectives are;

1. To determine the effect of physical features on nematode density in WIO of Kenyan shores
2. To determine the Nematode clustering in the Western Indian Ocean using Wald's linkage procedure

1.4 Justification of the study

Western Indian Ocean (WIO) biodiversity is subject that has not been fully exhausted. Oil and gas reserves are the latest discovery along the Kenyan coastal line. More information is required to ensure that the WIO biodiversity is not compromised during the oil exploration and exploitation. The study of nematode density and diversity will provide a proxy for biodiversity trends and factors that influence it in the WIO region. The information generated from this study will also form a basis of for future audit on the environment following the economic activities of oil and gas exploration.

1.5 METHODOLOGY

1.5.1 Source of the data

Data and information for this study was obtained from Muthumbi, 1998 PhD thesis and Muthumbi et al 2004. In these two sources data was analyzed using analysis of variance (ANOVA) to examine the effect of physical factors and environmental factors on Nematode

densities. Two Way Indicator Species Analysis (TWINSPAN) was used in studying the diversity of the Nematodes

In the current study we intend to analyze the same data using Analysis of covariance (ANACOVA) so as to establish the combined effect of physical factors and environmental factors. In this study Agglomerative clustering using the Wald's linkage procedure was used in to determine the clustering of the Nematodes

1.5.2 Study area

Four depth transects were sampled off the Kenyan coast in the Western Indian Ocean (WIO).

From North to South these transects were named Kiwayu , Tana , Sabaki and Gazi . At

each transect up to six stations were sampled from the continental shelf (20 or 50 m) to the continental slope up to a maximum of 2000 m depth . Sampling was performed during the southeast monsoon between 20th June and 4th July 1992 for season1 and repeated during the onset of the Northeast monsoon between 20th November and 8th December 1992 for season2. During the second campaign, the Tana transect was not sampled. Two extra stations were sampled within the vicinity of Gazi during a training programme conducted in November/December period referred here as Training transects.

1.5.3 Sampling

Sediment samples were taken from the RV Tyro during the cruises season1 and season2 (Netherlands Indian Ocean Program, 1992). Samples were collected with a modified box corer with a closing lid on top taking virtually undisturbed surface sediment. From each box core two sub-samples were taken up to a depth of 5 cm with a plastic core of 2.1 cm internal diameter and pooled. At most stations this procedure was repeated in order to get two replicate samples from separate box cores. Samples were preserved in 4% buffered formaldehyde solution. In the

laboratory they were sieved over 1 mm and collected on a 32 μm sieve. They were centrifuged twice in Ludox and the supernatant stained overnight in rose bengal (Heip *et al.*,1985).

1.6 DATA ANALYSIS

In order to establish whether transects, seasons and depth had an effect on the nematode density analysis of variance was carried out. The nematode density which is the response variable was transformed by the use log transformation to ensure that it fulfilled the normality assumption. The interaction effect of the seasons and depth was also investigated. The effect of the environmental factors was investigated by the use of analysis of covariance model. Correlation analysis was done on the environmental factors with an aim of ensuring that only those that are not highly related (below 0.5) were included in the model. The nematode composition was analyzed by use of cluster analysis. The nematodes were clustered by their dominance in relation to the depth. Data analysis was done using Stata/SE 11.2 and SAS computer packages.

CHAPTER 2

LITERATURE REVIEW

Soetaert et al. (1994) found out that the nematode assemblages in Mediterranean Sea significantly differ along the depth. Deeper sediment layers are dominated by fewer and larger species. It is argued that spatial segregation in the vertical plane can explain coexistence of several species belonging to the genus *Sabatieria*, while food resource partitioning (as witnessed by different buccal morphology) can explain coexistence of species belonging to the genus *Acantholaimus*. The Nematode community structure was analysed by means of the similarity coefficient, and a dendrogram was made by group-average clustering (Heip.*et al.*, 1988).

In a study of biodiversity of meiofauna in Vietnam by Quang et al. (2007) nematodes taxa were the most dominant. Eighty nematode genera belonging to 24 families with Comesomatidae having the highest abundance 33.8 % were found. *Theristus* and *Neochromadora* decreased in densities from the lower water line towards the mangrove forest edge, while Paracomesoma and Hopperia are typical and more abundant at the middle of the mudflat. Halalaimus increased from high on the mudflat to the low water line. There was a significant increase in densities from the mangrove forest edge towards the low water line. The densities of meiofauna ranged from 1156 inds/10cm² to 2082 inds/10cm². Data were analyzed using univariate and multivariate techniques. The significant differences in univariate measures between sites were tested using one-way ANOVA. In order to test the assumption of homogeneity of variances, Levene's tests were applied and data were log transformed. Tukey's multiple comparison tests were used when significant differences were detected.

Joanna Pawłowska et al. (2011) suggested that strong environmental seasonality has an effect in the nematode densities and composition. In a study seasonal variability of meiobenthic and macrobenthic in an Arctic fjord four seasons were examined. The seasonality observed in benthic biota was related to the pelagic processes, primarily the seasonal fluxes of organic and inorganic particles. The highest abundance, biomass and richness of benthic fauna occurred in the spring and during the summer, when a high load of glacial mineral material was transported to the fjord, the number of both meio- and macrobenthic individuals decreased remarkably. The data were double-root transformed to reduce the influence of numerically dominant species. Multivariate analysis and calculation of diversity indices was performed. The forward selection of environmental variables was used to quantify and rank the importance of variables in determining the species composition.

It is evident that the meiofaunal distributions are highly affected by environmental factors e.g steep environmental gradients of sedimentation, organic matter content, and salinity. A total of 12 higher meiofaunal taxa were recorded at in Hornsund fjord (77°N), with nematodes predominating at all stations. Non-parametric multivariate analyses demonstrated clear differences in meiofaunal abundance and composition between stations in the glacial bay and in the outer part of the fjord. Meiofaunal abundance increased with increasing distance from the source of disturbance, which in this study is tidal glaciers. Therefore, study demonstrates that the spatial structure of meiofauna is affected by the natural environmental disturbance, and analysis of meiofaunal assemblages can be used to assess the effect of such disturbances (Katarzyna Grzelak and Lech Kotwicki , 2011).

Thilagavathi et al. (2011) in a study of benthic meiofaunal composition and community structure in the Sethukuda mangrove area and adjacent open Sea, East Coast of India investigated the following environmental factors, surface water salinity water pH and dissolved oxygen concentration. Analysis was done using methods of Strickland and Parsons (1972). Comparatively, species density in the open sea was higher than the mangrove creek. The two way ANOVA showed significant variations between seasons and stations. Among the two study sites, maximum diversity index was observed in the mangrove creek, followed by open sea. The two way ANOVA showed significant difference between sites but not between seasons. The study area was subjected to a wide range of temperature fluctuation (23.5°C - 31.8°C) which might be the reason for higher density in early post monsoon and low density during monsoon. Maximum percentage composition observed in two stations in early winter may be due to enrichment of organic materials.

According to Boucher et al. (1994) diversity of nematodes is significantly different along different biotopes that were classified according to latitude and depth. A non-linear relationship was established between the depth and diversity, with biotopes with depth between 200 m – 6km displaying the highest diversity. Dinet & Viver, (1979) suggested a parabolic curve for the relationship between species richness and depth with a peak at 4000m. Analysis of variance was used to determine whether biotopes were significantly different according to environmental factors. Multivariate Analysis of Variance was used to determine among the factors affecting the diversity, which one was important. Depth, sample size, core penetration and abundance were found to be important in affecting the diversity of nematode but granulometry was not.

In the study of nematode community structure (Muthumbi. et.al, 2004) nematode densities in the Western Indian Ocean were found to be correlated with the oxygen concentration in the overlying water, with the lowest density at mid –depth (500-1000m). Sediment composition, water depth and oxygen levels had significant effect on the nematode composition. The highest diversity value was at mid –depth and the diversity assumed a unimodal trend along the gradient. Analysis of Variance was used to determine the effect of environmental factors along the sampled transects and Multivariate analysis was conducted on the nematode genera distribution data.

CHAPTER 3

EXPLORATORY DATA ANALYSIS

3.1 Introduction

Exploratory data analysis (John Tukey, 1978) is a scientific approach that is used to detect outliers, errors and hunt for new or unexpected patterns in the data and determine the relationship among the explanatory variables. Exploratory Data Analysis can be classified into two categories; graphical and non-graphical procedures. Non- graphical procedures entail calculation of summary statistics, while graphical method involves summarizing the data in diagrammatic or pictorial way. Exploratory Data Analysis is crucial in determining whether the data conforms to the underlying assumptions e.g. normality. This will help in determining whether to transform the data or to run non-parametric tests. Exploratory data analysis is carried out using histogram, normal quartile plot and box plot. Non-graphical procedures were also used e.g. calculating the means, skewness and kurtosis.

3.2 Non – graphical procedure

The means for the nematode densities was calculated as per transect, depth and the season. To check on normality, Skewness and Kurtosis (Mardia,1970) of the Nematode density were examined.

Table1: Summary statistics for the nematode densities by seasons

SEASON	Mean	p50	se(mean)	Max	Min	Skewness	Kurtosis
1	420.5556	297	73.34101	1350	112	1.755464	5.586946
2	354.8667	332	35.85777	661	189	0.741148	2.769644
Total	390.697	317	42.98589	1350	112	2.147165	8.370568

From table1 the means of the two seasons are different with season 1 having higher average number of nematodes than in season 2. The maximum and minimum number of nematodes was both in season 1.

The data in both seasons is generally skewed to the right since the value of skewness is greater than zero. Season 1 is leptokurtick since the kurtosis measure is greater than 3.

Table 2: Summary statistics for the nematode densities by depth

DEPKLAS	Mean	p50	se(mean)	max	Min	skewness	Kurtosis
20	838.3333	927	324.0537	1350	238	-0.28297	1.5
50	482.375	500	69.70062	669	112	-0.68661	2.440913
200	327	327	28	355	299	0	1
500	323.1429	295	30.52511	445	222	0.52194	1.941737
1000	257.8571	215	35.76805	443	189	1.268855	3.031676
2000	299.6667	272	49.33806	488	176	0.55069	1.918579
Total	390.697	317	42.98589	1350	112	2.147165	8.370568

From table 2 the nematodes density means declined with increase in depth with the lowest density recorded at 1000 meter and then again the number increased at the 2000m depth. The highest number of nematodes was found at 20m depth which was 1350 ind./10cm². The data for 500m, 1000m and 2000m was skewed to the right with rest assuming normality. Except for the data at 1000m which the kurtosis measure satisfied the normality requirement of 3 the rest was platykurtick i.e. the kurtosis measure was less than 3.

From the table 3 Tana and Training transect was not included because most of the depth classes were not sampled. Also data from 20m depth and 200 meter was not considered since most of the

stations were not sampled at those depth levels. The highest nematodes density recorded at Kiwayu and the lowest at Gazi i.e. 421.5 ind./10cm² and 317.5 ind./10cm² respectively. Sabaki and Training transect had data that was skewed to the right and data from Kiwayu was platykurtick.

Table3: Summary statistics for the nematode densities by transects

TRANSECT	Mean	p50	se(mean)	max	Min	skewness	Kurtosis
Kiwayu	421.5	388.5	62.34495	669	200	0.341693	1.74479
Sabaki	330.625	326.5	28.88891	443	189	-0.31199	2.213797
Gazi	317.5	218.5	61.37211	667	176	1.094406	2.900995
Total	356.5417	324.5	30.88818	669	176	0.824105	2.796039

3.3 Graphical procedure

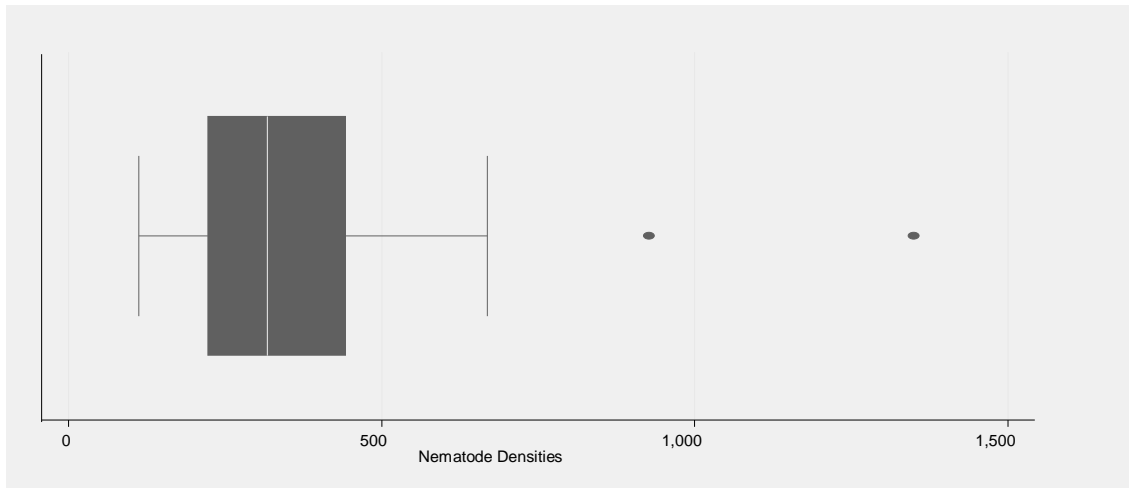


Figure 1: Box plot for the nematode density

The box plot displays the quartiles. The upper whisker is the 90th percentile (upper percentile), the lower whisker is the 10th percentile (lower percentile) and the line cutting across stand for the median. Thus for a normal data it is expected that median divides the data into two equal halves. From the above plot it can also be observed that there are two outliers and the median is not located at the center of the data thus the data is not normal

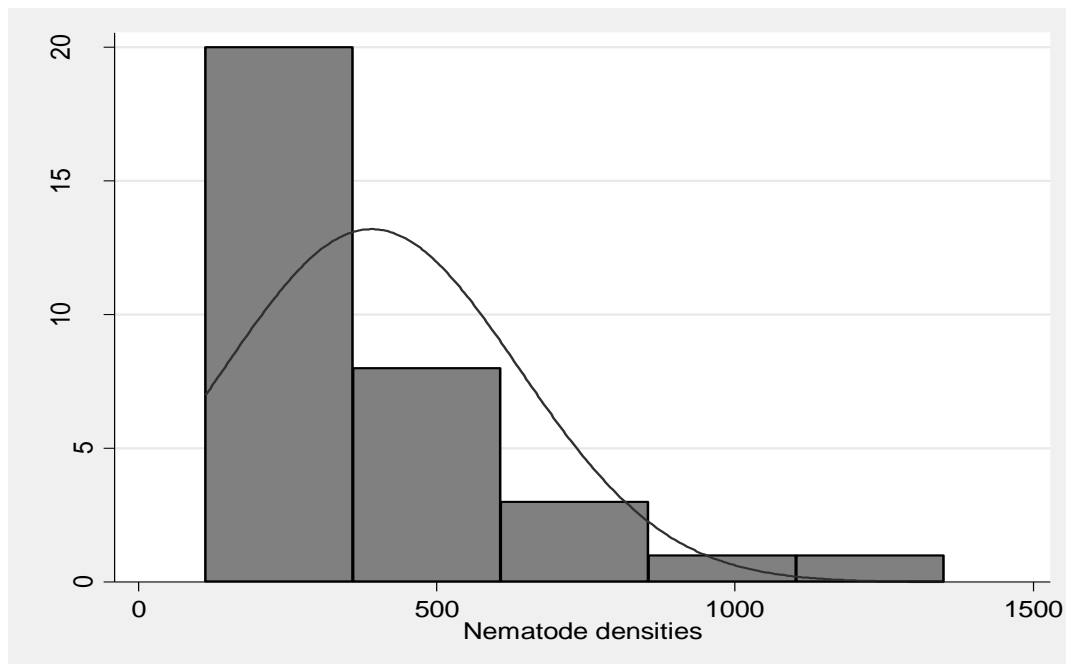


Figure 2: Histogram for the nematode densities

Histogram was plotted for the nematode densities against frequency. For a normally distributed data the histogram is expected to be bell shaped. From the above figure it is clear the data is skewed to the right.

Figure 3 is a plot of the quartile of the data against the value of the standard normal distribution that lies at the quartile of the data. For a normal data the points are expected to fall on the straight line. In this case the data is skewed to the right thus the data is not normal.

Since the data does not satisfy the normal assumption we thus do a transformation to the response variable.

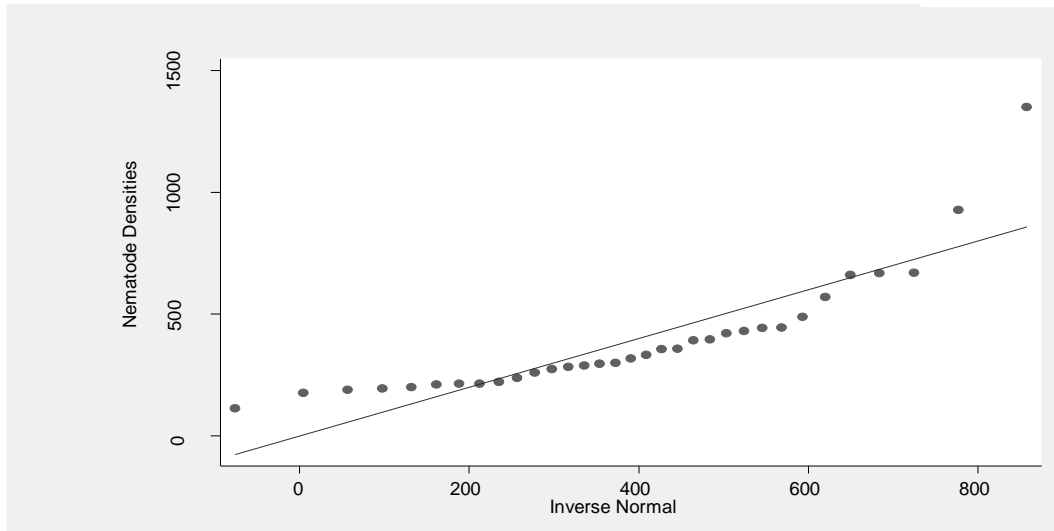


Figure 3: Normal quartile plot for the nematode densities

3.4 Data transformation

Five transformations are commonly used with ecological and environmental data: the logarithmic transformation, square root transformation, the angular transformation, the reciprocal transformation and the Box cox transformation (a premier of ecological statistics by Nicolas J. Gotelli). Since the data is positively skewed and had some outliers the logarithmic transformation (Scotsman J N, 1614) was the most appropriate.

Logarithm to the base of 10 is applied to the nematode densities:

$$Y(\text{transformed nematode densities}) = \log_{10}(\text{nematode density})$$

After transformation using the logarithm transformation the nematode density data was shifted to normal as shown below.

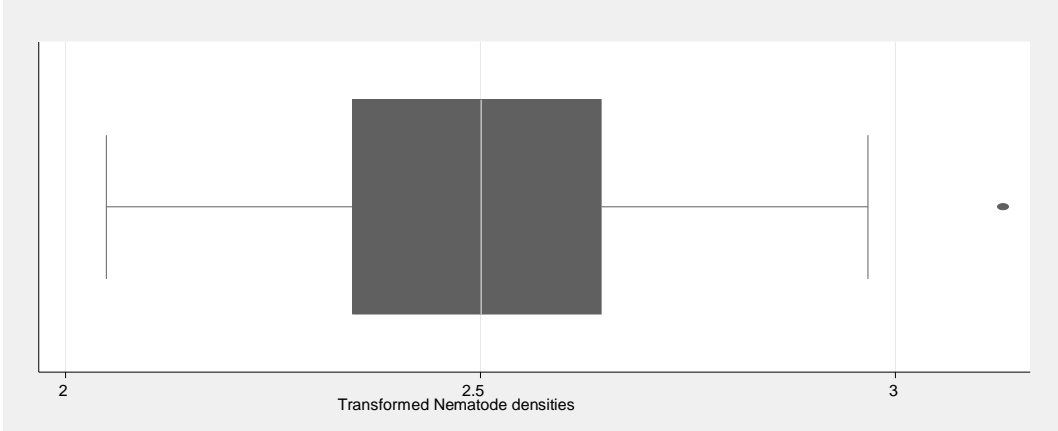


Figure 4: Box plot for the transformed Nematode densities

The box plot Figure 4 indicates the data is normal since the median divides the box into two equal halves although there is one outlier.

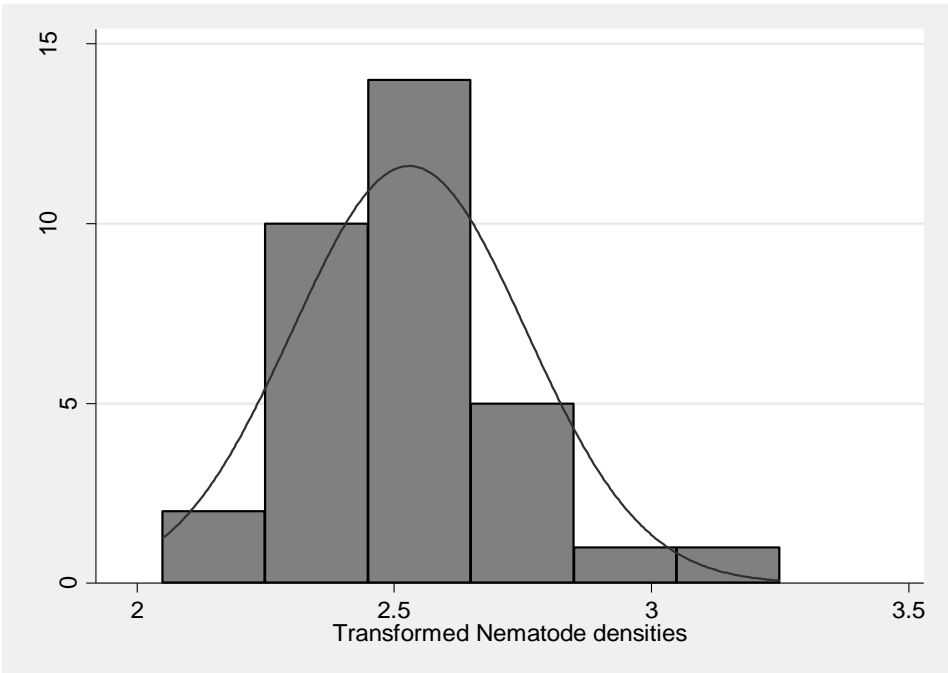


Figure 5: Histogram for the Transformed Nematode densities

The histogram Figure 5 indicates the transformed nematode densities to be normally distributed since it appears to be bell shaped.

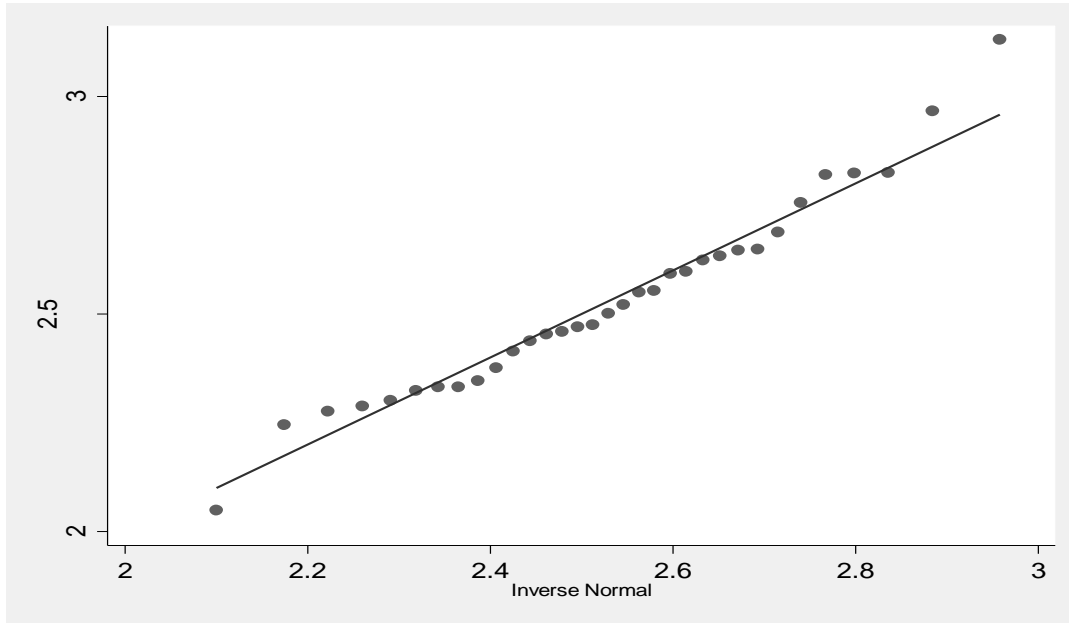


Figure 6: Normal quartile plot for the transformed data

Since most of the points in Figure 6 fall close to the straight line the data is thus considered to be normally distributed.

3.5 Significance test for normality

In testing that an underlying population is normally distributed the skewness and kurtosis statistics have been shown to be powerful and informative test, and a good complete normality analysis would consist of the use of the plot plus the statistics (D’Agostino et al, 2006)

Table 4: Significance test for kurtosis and skewness

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	Joint	
				adj chi2(2)	Prob>chi2
LOGNEM_DEN	33	0.1828	0.3885	2.73	0.2551

The p- value for the skewness is 0.1828 and that of Kurtosis is 0.3885,

The joint p-value for the skewness and kurtosis is 0.2551 and thus the data can be considered to have fulfilled the normality assumption.

CHAPTER 4

GENERAL LINEAR MODELS

4.1 Introduction

Linear models are fundamental to analysis of both univariate and multivariate data. In formulating a linear model we need an observed data vector (matrix) Y and relate the observed to a set of linearly independent fixed variables. The relationship between the dependent variable and independent variables is examined using a set of parameters that are considered to be independent of a vector (matrix) ε of errors. Thus to construct a GLM we need the random vector of dependent variable Y which is related to a vector β of k parameters through a known fixed design matrix X plus a random vector ε of errors. Thus the equation will be as below,

$$Y = X\beta + \varepsilon \tag{4.1}$$

Where

Y is the $N \times 1$ response vector

X is an $n \times k$ matrix of constants ($n < k$, $\text{rank}(X) = k$)

β is a $k \times 1$ vector of parameters

ε is an $n \times 1$ random vector whose elements are independent and all have the normal distribution $N(0, \sigma^2)$

For $i=1,2,\dots,n$ the relationship between the dependent variable Y and the k independent variables x_1, x_2, \dots, x_k is linear in parameters. Furthermore assume that the parameter $\beta_0, \beta_1, \dots, \beta_k$ are free to vary over the entire parameter space so that there is no restriction

on $\beta' = \{ \beta_0, \beta_1, \dots, \beta_k \}$ where $q=k+1$ and error e has a mean zero and unknown variance δ^2 then the equation will be,

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & x_{1k} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & \cdot & \cdot & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix} \quad (4.2)$$

Where the design matrix has a full column rank $r(X) = q$. If the $r(X) < q$ so that X is not a full column rank and X contains indicator variables, we obtain the ANOVA. The design matrix X can be portioned into two sets of independent variables, matrix $A_{n \times q_1}$ that is not full rank and matrix $Z_{n \times q_2}$ that is full rank so that $X=[A Z]$ where $q = q_1 + q_2$. The matrix A is ANOVA design and matrix Z is regression design matrix also called matrix of covariates, the model is then called ANACOVA model

4.2 Parameter Estimation

After data collection we have observations of random variables Y_i that are denoted by y_i . The simultaneous equations formed by the general linear model cannot be solved since the numbers of parameters are usually selected to be more than the number of independent equations. Thus a method that best fit the data is required. This can be achieved by the use of method of ordinary least squares.

4.2.1 Ordinary least squares method

We seek $\beta_0, \beta_1, \dots, \beta_k$ that will minimize

$$\sum_{i=1}^n \varepsilon^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{4.3}$$

$$= \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_k x_{ik} \right)^2$$

This equation can be solved by the use of matrix notion using the following theorem

Theorem:

If $Y = X\beta + \varepsilon$ where X is $n \times k+1$ of rank $k+1 < n$, then the value of $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ that

minimizes (4.3) is $\hat{\beta} = (x'x)^{-1} x'y$

Proof:

Equation (4.3) can be written as

$$\hat{\varepsilon}'\hat{\varepsilon} = (y_i - x'_i\hat{\beta}')^2 = (y - x\hat{\beta})' (y - x\hat{\beta}) \tag{4.4}$$

After expanding we have

$$\hat{\varepsilon}'\hat{\varepsilon} = y'y - 2y'x\hat{\beta} + \hat{\beta}'x'x\hat{\beta} \tag{4.5}$$

To find $\hat{\beta}$ that minimizes $\hat{\varepsilon}'\hat{\varepsilon}$ we differentiate with respect to $\hat{\beta}$

$$\frac{\partial \hat{\varepsilon}'\hat{\varepsilon}}{\partial \hat{\beta}} = 0 - 2x'y + 2x'x\hat{\beta} = 0 \tag{4.6}$$

$$= x'y = x'x\hat{\beta} \tag{4.7}$$

$$\hat{\beta} = (x'x)^{-1} (x'y)$$

4.3 Analysis of Variance

The ANOVA technique extends what an independent-sample t test can do to multiple means. If more than two means are compared, repeated use of the independent-samples t test will lead to a higher Type I error rate. A better approach than the t test is to consider all means in one null hypothesis—that is, examining the plausibility of the null hypothesis with a single statistical test. In doing so, researchers not only save time and energy, but more important, they can exercise a better control of the probability of falsely declaring significant differences among means. In this case we use generalized linear model ANOVA since we have unequal groups.

The test statistic is developed using the idea of portioning of the total sum of squares (TSS) of the measurement about the mean. Total sum of squares is divided into variability due to treatments (SST) and within variability which in other words is referred to as error term (SSE).

$$TSS = \sum_{ij} \left(y_{ij} - \bar{y}_{..} \right)^2 = \sum_{ij} \left(y_{ij} - \bar{y}_{i.} \right)^2 + \sum_{ij} \left(\bar{y}_{i.} - \bar{y}_{..} \right)^2$$

$$SST = \sum_{ij} \left(y_{ij} - \bar{y}_{i.} \right)^2$$

$$SSE = \sum_{ij} \left(\bar{y}_{i.} - \bar{y}_{..} \right)^2$$

Where: y_{ij} is the observed value

$\bar{y}_{..}$ is the grand mean

$\bar{y}_{i.}$ is the treatment means

The test hypothesis is defined as below;

$$H_0 = \mu_1 = \mu_2 = \mu_3 = \dots = \mu_t$$

vs

H_1 : Atleast one of the means is different from the rest

To calculate the F statistic we require the mean sum of squares (MSS) which is the sum of squares divided by its degree of freedom. Thus we have MST mean square between samples and MSE mean square within samples. The numbers of degrees for SST are $t-1$ and similarly the numbers of degree for SSE are $n-t$.

The null hypothesis of equality of population mean is rejected if $F = \frac{MST}{MSE}$ exceeds the tabulated value of F for $\alpha = \infty$, $df_1 = t-1$ and $df_2 = n-t$

4.3.1 Analysis of variance without interaction

To test whether there is difference in the nematodes means we run analysis of variance.

The model for a three way ANOVA

$$y_{ijk} = u + \alpha_i + \beta_j + \eta_k + \varepsilon_{ijkl}$$

$$i = 1, 2$$

$$j = 1, 2, 3, 4, 5$$

$$k = 1, 2, 3, 4, 5, 6$$

y_{ijk} the nematode densities from the combination of the i^{th} season, j^{th} transect and k^{th} depth

μ the overall (constant) population mean of nematode densities

α_i the main effect of i^{th} transect on the nematode densities, holding all other factors constant

β_j the main effect of the j^{th} season on the nematode densities, holding all other factors constant

η_k the main effect of the k^{th} depth on the nematode densities, holding all other factors constant

Table 5: Analysis of variance

Source	Partial SS	df	MS	F	Prob > F
Model	1.104925	10	0.110493	4.48	0.0016
TRANSECT	0.568645	4	0.142161	5.76	0.0025
SEASON	0.004217	1	0.004217	0.17	0.6832
DEPCLASS	0.663137	5	0.132627	5.38	0.0022
Residual	0.542529	22	0.02466		
Total	1.647455	32	0.051483		

From table 5 the overall model is significant at $\alpha=0.05$

Transect and depth class effect is significant at $\alpha=0.05$ but the season's effect is not significant.

To determine which transects is significantly different, lsd test was done on transects as shown in table 6. Tana is significantly different from Kiwayu, Training and Sabaki, while Gazi is significantly different from Kiwayu. Tana had the minimum mean nematode density of 187 ind./10cm² and Kiwayu had the highest(477 ind./10cm²) followed by Sabaki(422 ind./10cm²), though the difference between Kiwayu and Sabaki was not significant.

Table 6: Transect comparison at 0.05 level of significance

TRANSECT Comparison	Difference Between Means	95% Confidence Limits		
k - tr	0.1039	-0.5389	0.7468	
k - s	0.1735	-0.1961	0.5431	
k - g	0.449	0.0494	0.8486	*
k - t	0.9172	0.369	1.4654	*
tr - k	-0.1039	-0.7468	0.5389	
tr - s	0.0696	-0.5626	0.7017	
tr - g	0.3451	-0.305	0.9952	
tr - t	0.8133	0.0626	1.564	*
s - k	-0.1735	-0.5431	0.1961	
s - tr	-0.0696	-0.7017	0.5626	
s - g	0.2755	-0.1066	0.6576	
s - t	0.7437	0.2081	1.2793	*
g - k	-0.449	-0.8486	-0.0494	*
g - tr	-0.3451	-0.9952	0.305	
g - s	-0.2755	-0.6576	0.1066	
g - t	0.4682	-0.0885	1.0249	
t - k	-0.9172	-1.4654	-0.369	*
t - tr	-0.8133	-1.564	-0.0626	*
t - s	-0.7437	-1.2793	-0.2081	*
t - g	-0.4682	-1.0249	0.0885	

Comparisons significant at the 0.05 level are indicated by *

4.3.2 Analysis of variance with interaction

To start with we run a interaction plot. This helped in determining whether there is interaction between seasons and depth

From the Figure7 below it is clear that there is interaction between the season and the depth class. We can proceed and test whether the interaction is significant using a three way ANOVA with interaction.

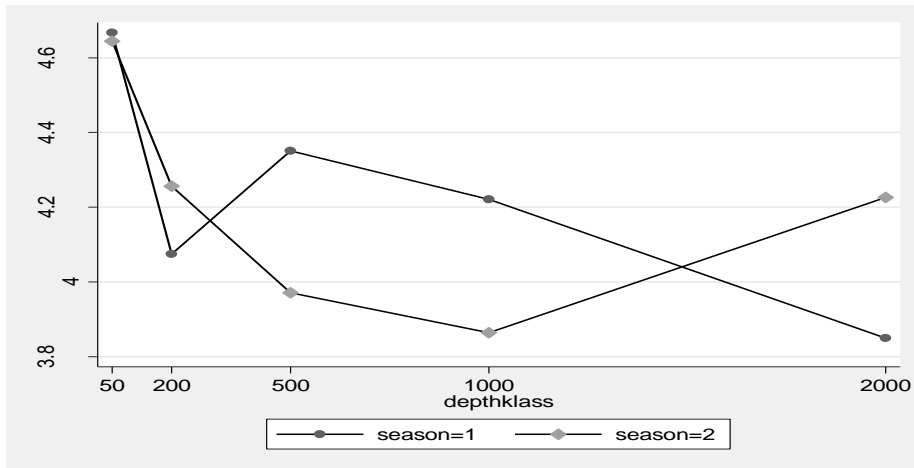


Figure 7: Interaction plot for seasons and depthclass

The model for a three way ANOVA

$$y_{ijkl} = u + \alpha_i + \beta_j + \eta_k + (\alpha\beta)_{ij} + \varepsilon_{ijkl}$$

y_{ijkl} The nematode densities from l^{th} replicate from the combination of the i^{th} season, j^{th} transect and k^{th} depth

u The overall (constant) population mean of nematode densities

α_i The main effect of i^{th} transect on the nematode densities, holding all other factors constant

β_j The main effect of the j^{th} season on the nematode densities, holding all other factors constant

η_k The main effect of the k^{th} depth on the nematode densities, holding all other factors constant

$(\alpha\beta)_{ij}$ The effect on nematode densities of the interaction of the seasons and depth

Table 7: Analysis of variance with interaction of seasons and depth class

Source	Partial SS	Df	MS	F	Prob > F
Model	1.210903	14	0.086493	3.57	0.0064
TRANSECT	0.570412	4	0.142603	5.88	0.0033
DEPKLAS	0.669131	5	0.133826	5.52	0.003
SEASON	0.000704	1	0.000704	0.03	0.8667
DEPCLAS#SEASON	0.105978	4	0.026494	1.09	0.39
Residual	0.436551	18	0.024253		
Total	1.647455	32	0.051483		

R-squared = 0.7350

Root MSE = .155733

Generally the model was highly significant at alpha= 0.05 .

The effect of transect and depth on the nematode densities are significant at alpha 0.05 The Seasons effect is not significant. Though there exists an interaction between the seasons it is not significantly affecting the nematode densities.

4.3.3 Fitting a quadratic model

For qualitative data our interest would be to determine the minimum or the maximum point depending on how the data behaves. This brings in the aspect of fitting non-linear models especially for ecological data due to the range of explanatory variables.

The means for the nematode densities was inversely proportional to the depth up to a given depth and then thereafter nematode density was proportional to the depth. Thus we can fit polynomial to determine the depth where we have the minimum nematode densities. Wald test was used to determine the number of coefficients to be included in the polynomial

The means for the depth seem to decrease up to a given level and then start to increase again. From the scatter diagram it is apparent that the data fits a quadratic model with a minimum level.

The quadratic model to be fitted is given as mean density = $E(Y) = a + bX + cX^2$ (4.7)

Where Y stands for density and X for depth

From equation (4.7) we get

$$\frac{dE(y)}{dx} = b + 2cX$$

$$\frac{d^2E(y)}{d^2x} = 2c$$

For $\frac{dE(y)}{dx} = 0$, the value of $X = X_0 = -\frac{b}{2c}$

The value of X_0 is at maximum if $\frac{d^2E(y)}{d^2x} < 0$ and at minimum if $\frac{d^2E(y)}{d^2x} > 0$

The Nematode density data was fitted into a quadratic curve in order to determine the depth at which the minimum density of nematode was recorded as shown below

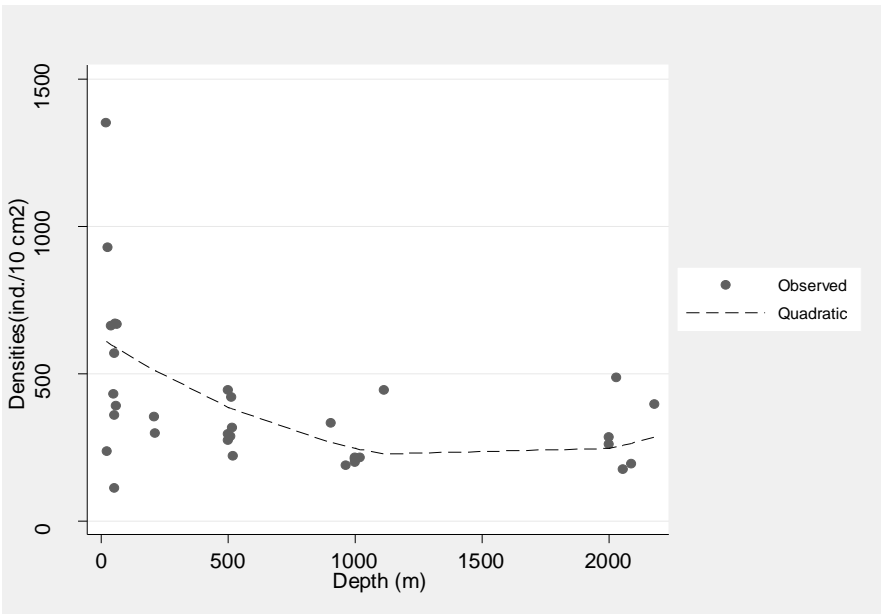


Figure 8: curve fit for nematodes season1

$$density = 619.90858 - .55880722 \text{ depth} + 0.00018612 \text{ depth}^2$$

Minimum number of nematodes is found at 1501.20 meters for season 1

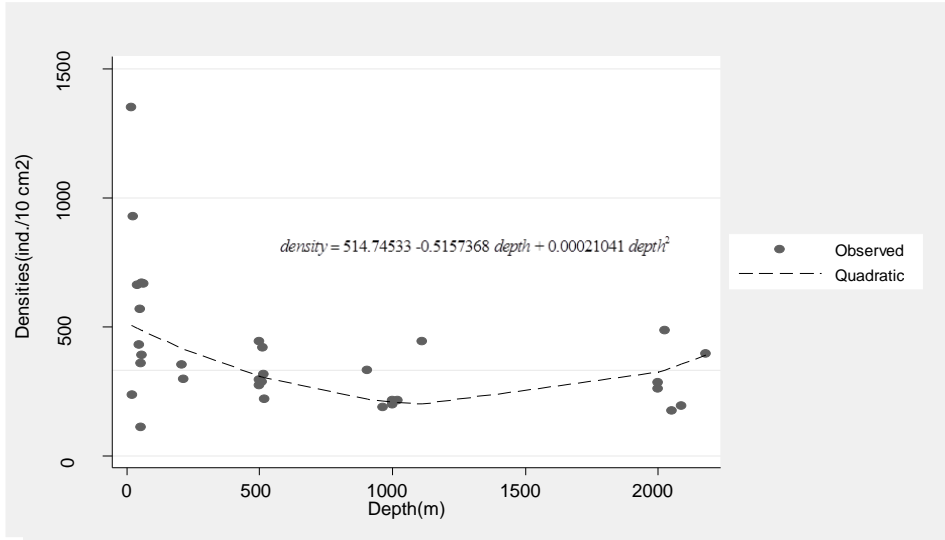


Figure 9: curve fit for nematodes season2

$$density = 514.74533 - 0.5157368 \text{ depth} + 0.00021041 \text{ depth}^2$$

Minimum density of the nematodes is found at 1225.552017 meters for season 2

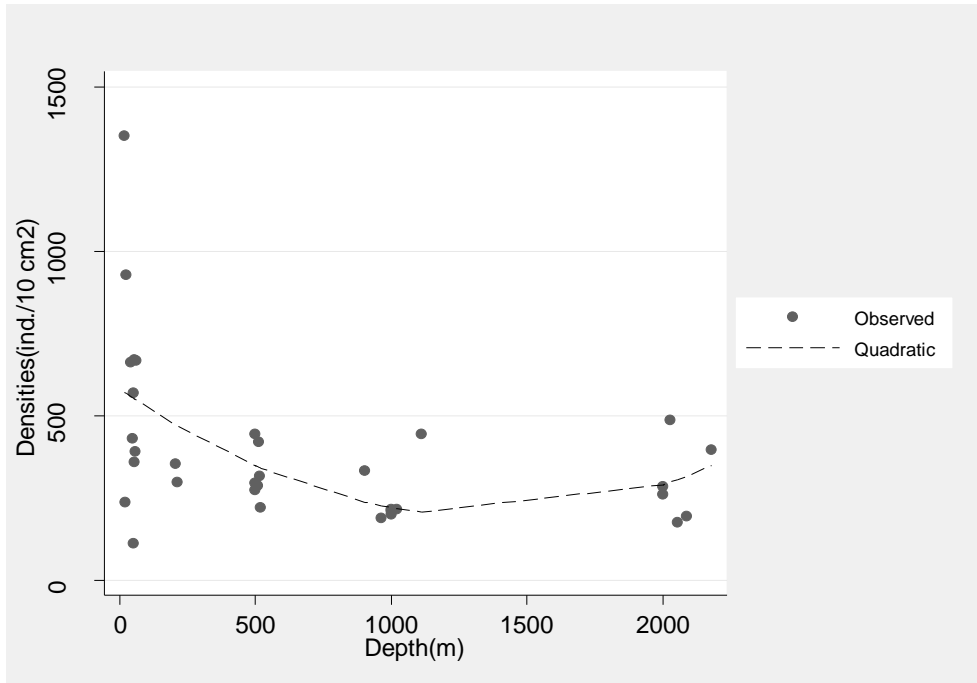


Figure 10: Curve fit for the nematode densities versus the depth

$$density = 581.48782 - 0.57419797 \text{ depth} + 0.00021442 \text{ depth}^2$$

Minimum number of nematodes is found at 1338.956 meters

4.4 Testing the correlations of the covariates

If the predictor variables are highly correlated this may result to multicollinearity in the model.

Multicollinearity is when two or more predictor variables are highly correlated and one can be linearly predicted from the other, this results to unstable model estimate due to high standard errors. To eliminate the effect of multicollinearity the highly correlated predictor variables are presented by one of them. A correlation analysis was done to determine the highly correlated variables. A correlation of 0.5 was considered as high and either of the predictor variables was included in the model.

Table 8: Variables selected to be included in the model

	SFINES	SD50MUC	SED__PHA	ORGC	DNA_RNA
SFINES	1				
SD50MUC	0.2692	1			
SED__PHA	0.4074	-0.0051	1		
ORGC	-0.4977	-0.5704	0.4283	1	
DNA_RNA	0.3186	0.1257	0.4674	0.2791	1

The variables fine sand, sediment phaeopigments, organic matter, dna-rna were selected and included in the ANOVA model.

4.5 Environmental and physical factors

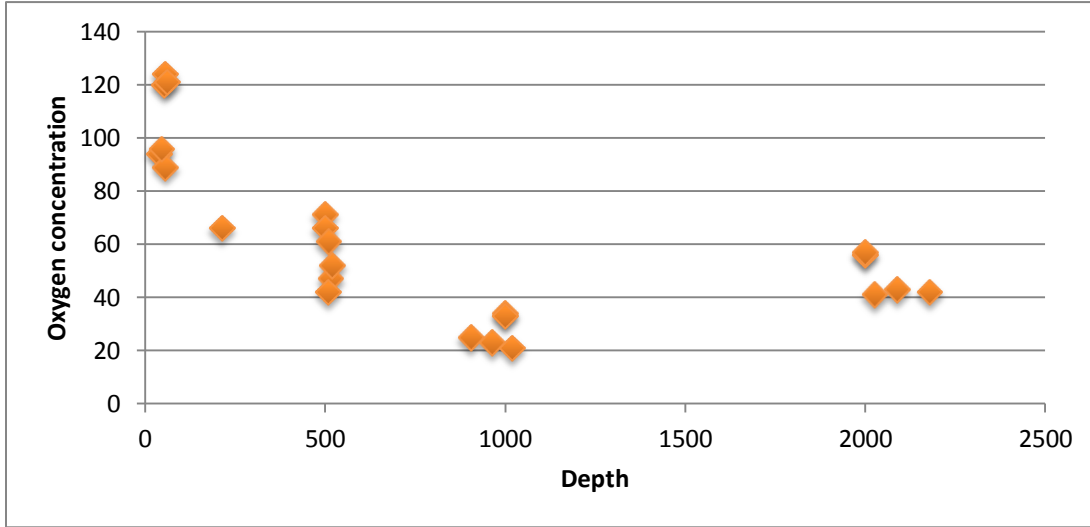


Figure 11: Oxygen concentration in relation to depth

From Figure 11 the oxygen level decreases with increase in depth up to around 1000m. At 2000m depth the oxygen concentration starts to increase again.

From figure 12 below the sediment phaeopigment is highest in the shallow depth and it decreased with increasing depth. With the highest amount recorded at Kiwayu station during the first season. Generally the amount of sediment phaeopigments was lower in the second season as compared to the first season for all transects

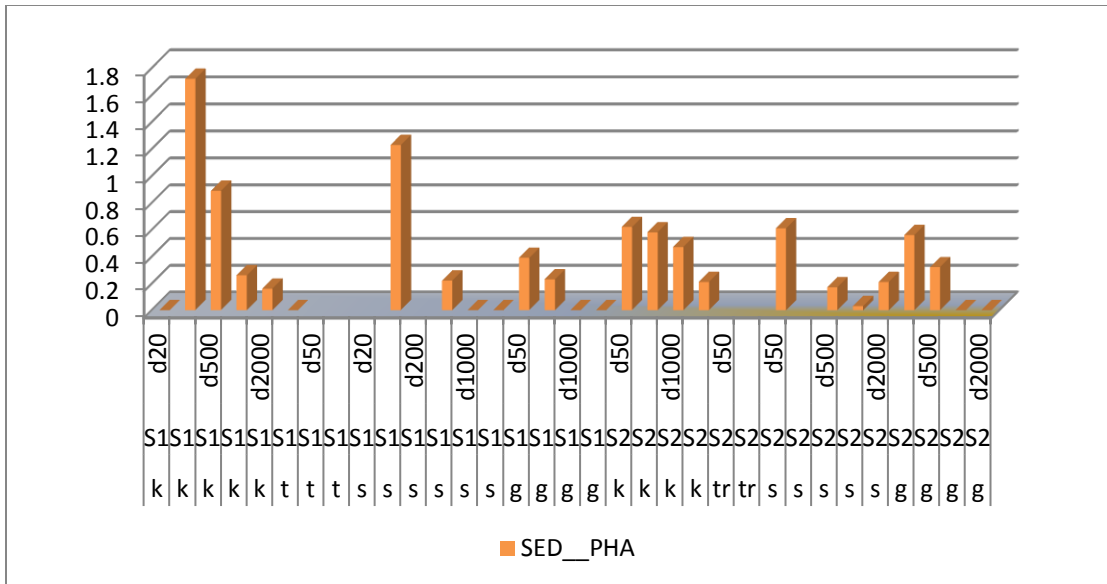


Figure 12: Sediment phaeopigments in relation to Depth, Transect and Seasons

From figure 13 the amount of DNA-RNA ratio was highest in Kiwayu in the second season. The amount decreased with increase in depth and a slight increase was observed at 1000m depth along all transects except for sabaki where the highest was recorded at 2000m depth.

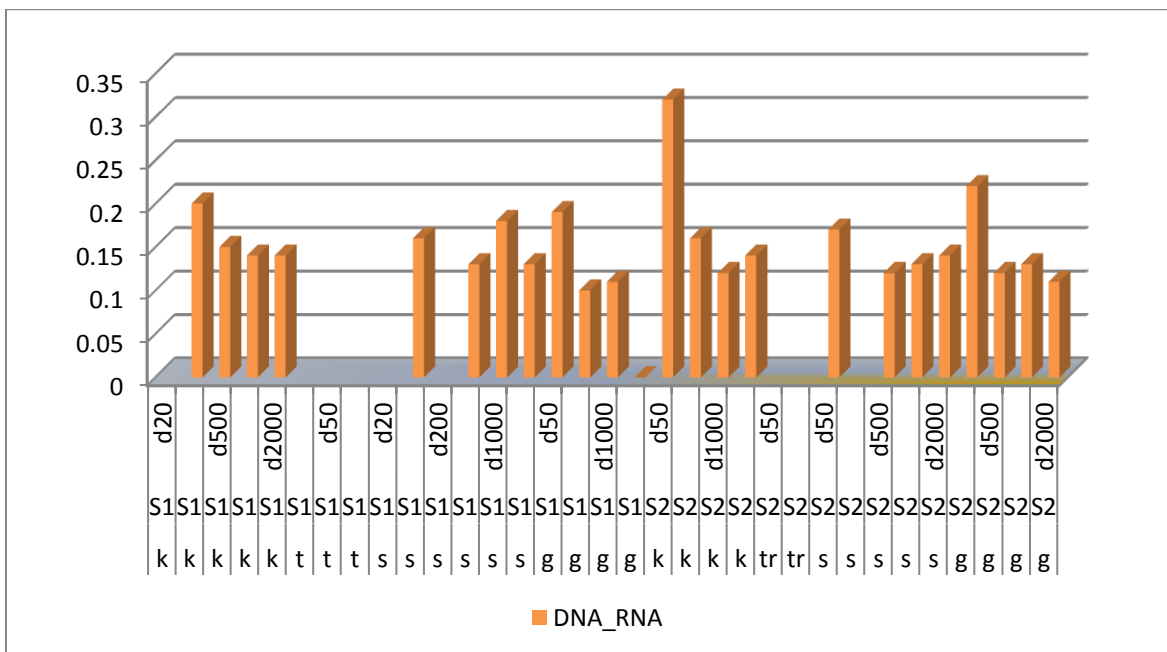


Figure 13: DNA-RNA ratio in relation to Depth, Transect and Seasons

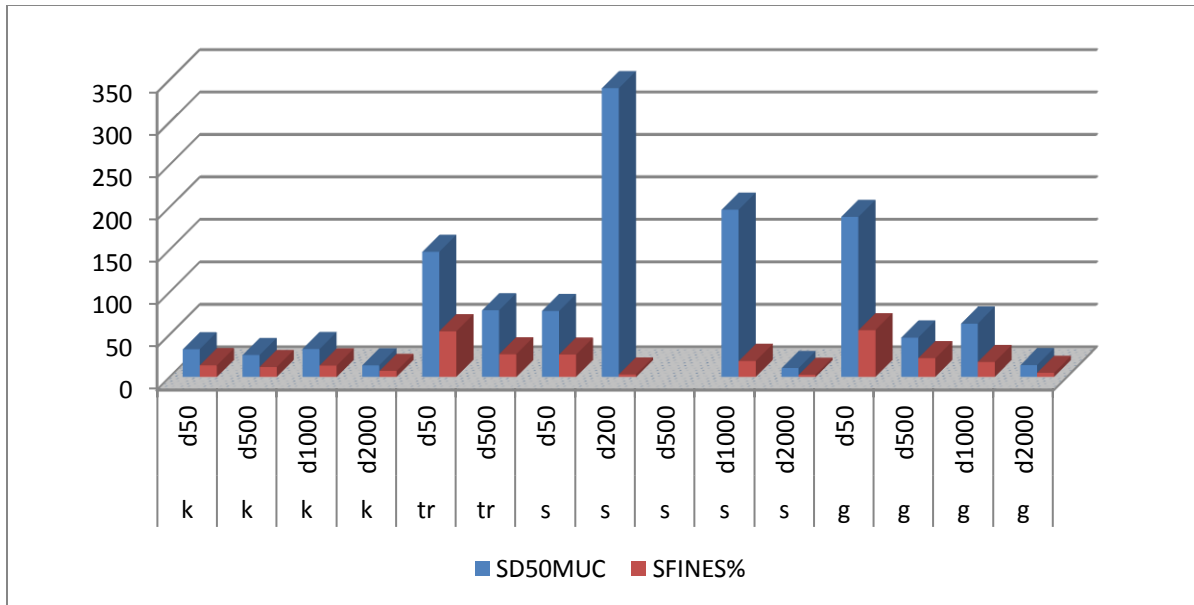


Figure 14: Sand and Silt composition in relation to Depth, Transect and Seasons

From figure 14 above the percentage of sand was higher than the silt in all the stations and the various depths. The highest difference was observed in Sabaki transect 200m depth.

4.6 Analysis of Covariance

This is an extension of analysis of variance by including one or more continuous variables known as covariates. Though covariates are not the main effect they have an effect on the response variable thus their inclusion in the model helps in improving the precision of treatment comparison. The inclusion of covariates that have an effect results to reduction of the sum of squares due to error, and our test for treatment difference will be more powerful.

One of the major applications of ANACOVA is aiding in interpretation of the research results. In this study ANACOVA is used in the interpretation and characterization of transect, season and depth effect on the nematode densities. By examining the nematode densities together with the

other characters whose functional relationships to the nematode densities are known, the biological process governing transect, season and depth effects on the nematode density can be characterized.

In this study the aim is to measure whether the main effect i.e. transect, depth and season has significant effect on the nematode densities with a precision. This can be done by including the covariates that are expected to have an effect to the nematode densities. Composition of sand particles, sediment phosphides, and organic matter and RNA-DNA ratio were measured and included in the model.

ANCOVA is important in error control. Proper blocking reduces experimental error by maximizing the difference between blocks and thus minimizing the difference within blocks. However in some cases blocking may not adequately reduce the experimental error and thus measuring the covariates known to be linearly related with nematode density will reduce the experimental error.

4.6.1 ANACOVA model

$$y_{ij} = u + A_i(X_{ij} - \bar{X}_i) + \varepsilon_{ij}$$

Where

u is the overall mean

A_i is the effect of treatment a

X_{ij} is the covariate measured for observation y_{ij}

\bar{X}_i is the average value of the covariate for treatment group i

ε_{ij} is the error term

Table 9: Analysis of covariance table

Source	Partial SS	DF	df MS	F	Prob > F
Model	0.30071	9	0.033412194	1.7	0.537
TRANSECT	0.006145	2	0.0030726	0.16	0.8727
DEPKLAS	0.041686	3	0.013895454	0.71	0.6799
DNA_RNA	0.020113	1	0.020112757	1.02	0.4962
SED__PHA	0.017305	1	0.017304597	0.88	0.5201
SFINES	0.000935	1	0.00093478	0.05	0.8632
ORGC	0.007427	1	0.007426526	0.38	0.649
Residual	0.019632	1	0.019632393		
Total	0.320342	10	0.032034214		

Root Mse= 0.14

R-squared= 0.9386

The Main factors and the cofactors had no significant effect on the nematode density. The model R-squared is 0.93 which means the factors explained 93% of the variability in the nematode density.

CHAPTER 5

CLUSTER ANALYSIS

5.1 Introduction

Cluster analysis is concerned with group identification. The goal of cluster analysis is to partition a set of observation into distinct number of groups or clusters in such a manner that all observations within the group are similar while observations in different groups are not similar.

If data is represented as $n \times p$ matrix $Y = [y_{ij}]$

$$Y_{n \times p} = \begin{bmatrix} y_1' \\ y_2' \\ \cdot \\ \cdot \\ y_n' \end{bmatrix}$$

Where

the goal of cluster analysis is to develop a classification scheme that will partition the rows of Y into k distinct clusters. The rows usually represent items or objects.

To uncover the grouping in the data a measure of nearness, also called a proximity measure needs to be defined. The two natural measure of nearness are degree of distance (dissimilarity measure) or degree of association (similarity measure). The choice depends on the subject matter, scale measurement and type of the variable being analyzed.

5.2 Dissimilarity measure

Given two objects y_r and y_s in a p - dimensional space, a dissimilarity measure satisfies the following condition

1. $d_{rs} \geq 0$ for all objects y_r and y_s
2. $d_{rs} = 0$ if and only if $y_r = y_s$
3. $d_{rs} = d_{sr}$

Condition (3) implies that the measure is symmetric. Condition (2) requires the measure to be zero whenever object r equal object s, the objects are identical only if $d_{rs} = 0$ and under no other situation. Finally (1) implies that the measure is never negative. For continuous variable the most common dissimilarity measure is Euclidean distance between two objects. The calculation of the Euclidean distance is based on the Pythagoras theorem.

Given $(n \times p)$ matrix Y with $(1 \times p)$ row vectors y_i' the square of the Euclidean distance between two rows y_r and y_s is defined as

$$d_{rs}^2 = (y_r - y_s)'(y_r - y_s) = \|y_r - y_s\|^2$$

The $n \times n$ data matrix $D = [d_{rs}]$ is called the Euclidean distance matrix. Because the variables are not commensurate some variable may dominate the ranking of distance thus weighting is applied on the squared differences by

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_{.j})^2, j = 1, 2, \dots, p$$

Where S_j^2 and $\bar{y}_{.j}$ represent estimates of the mean and variance of variable j

$$d_{rs}^2 = (y_r - y_s)' (\text{diag } S)^{-1} (y_r - y_s)$$

The process eliminates the dependence of the analysis on unit of measurement.

Because Euclidean distance is a special case of the Minkowski metric (Lp-norm) the dissimilarity measure may be represented as

$$d_{rs} = \left(\sum_{j=1}^p |y_{rj} - y_{sj}|^\lambda \right)^{1/\lambda}$$

$\lambda = 2$ for Euclidean distance.

5.3 Similarity measure

Given two objects y_r and y_s in a p- dimensional space, a dissimilarity measure satisfies the following condition;

1. $0 \leq s_{rs} \leq 1$ for all objects y_r and y_s
2. $s_{rs} = 1$ if and only if $y_r = y_s$
3. $s_{rs} = s_{sr}$

Condition 1 and 2 ensure that it is always positive and identical only if objects r and s are identical

Given a similarity measure satisfying all the three conditions one can get a dissimilarity measure by using the relation that $d_{rs} = 1 - s_{rs}$.

The common measure of similarity is the Pearson product correlation between

y_r and y_s $r, s = 1, 2, 3, \dots$

$$q_{rs} = \frac{\sum_{j=1}^p (y_{rj} - \bar{y}_r)(y_{sj} - \bar{y}_s)}{\left[\sum_{j=1}^p (y_{rj} - \bar{y}_r)^2 \sum_{j=1}^p (y_{sj} - \bar{y}_s)^2 \right]^{1/2}}$$

However since $-1 \leq q_{rs} \leq 1$ it does not satisfy condition (1)

To correct that we standardize the matrix Y so that $\tilde{y}_{ij} = \frac{(y_{rj} - \bar{y}_i)}{s_i}$

Where; $s_i = \frac{\sum_{j=1}^p \bar{y}_{sj}^2}{(p-1)}$ for $i=1, 2, 3, \dots, n$ objects, now from the above q_{rs} can be related to the

Euclidean distance as follows;

$$\begin{aligned} d_{rs}^2 &= \sum_{j=1}^p (\tilde{y}_{rj} - \tilde{y}_{sj})^2 \\ &= \sum_{j=1}^p \tilde{y}_{rj}^2 + \sum_{j=1}^p \tilde{y}_{sj}^2 - 2 \sum_{j=1}^p (\tilde{y}_{rj} \tilde{y}_{sj}) \\ &= 2(1 - q_{rs}) \\ \text{then } d_{rs} &= \sqrt{2(1 - q_{rs})} \end{aligned}$$

To perform cluster analysis one has to have proximity matrix. The proximity matrix represents the strength of the relationship between pairs of rows in the data matrix $Y_{n \times p}$.

Methods of cluster analysis are usually divided into two classes i.e. hierarchal and non-hierarchal clustering methods. Hierarchal methods are used for clustering variables and items only. Thus in this study agglomerative hierarchal clustering method is applied.

5.4 Agglomerative Hierarchal Clustering Method

The agglomerative hierarchal clustering method use the element of proximity matrix to generate a dendogram . It generate a sequence of clustering solution beginning with cluster containing a single cluster.

Steps for agglomerative hierarchal clustering

1. Begin with a cluster each containing only a single object.
2. Search the dissimilarity matrix D for the most similar pair, let the pair chosen be associated with element d_{rs} so that object r and s are selected.
3. Combine objects r and s into a new cluster(rs) employing some criterion and reduce the number of cluster by 1 by deleting the row and column for objects r and s. Calculate the dissimilarity between cluster(rs)and all remaining clusters, using the criterion and add the row and column to the new dissimilarity matrix.
4. Repeat step 2 and 3, (n-1) times until all objects form a single cluster. At each step, identify the merged cluster and value of dissimilarity at which the cluster are merged.

The criterion in step 3 can be changed to suite the analysis. The common criterions in agglomerative hierarchal clustering are single linkage, complete linkage, and average linkage, centroid and Wald's method.

Many studies conclude that the Wald's method and the average linkage method are the most suitable. After several trials the Wald's method stood out to be best for this study.

5.5 Wald's Method

It is referred to as the incremental sum of squares method. It uses the within cluster squared distances and between cluster squared distances.

$$SSE_A = \sum_{i=1}^{n_A} (y_i - \bar{y}_A)' (y_i - \bar{y}_A)$$

$$SSE_B = \sum_{i=1}^{n_B} (y_i - \bar{y}_B)' (y_i - \bar{y}_B)$$

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (y_i - \bar{y}_{AB})' (y_i - \bar{y}_{AB})$$

$$\text{where } \bar{y}_{AB} = \frac{(n_A \bar{y}_A + n_B \bar{y}_B)}{(n_A + n_B)}$$

n_A , n_B and n_{AB} are the number of points in A , B and AB respectively.

The Wald's method joins the two clusters A and B that minimize SSE defined as,

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B)$$

It can be shown that the increase I_{AB} is equivalent to the following

$$= \frac{n_A n_B}{n_A + n_B} (\bar{y}_A - \bar{y}_B)' (\bar{y}_A - \bar{y}_B)$$

Thus by minimizing the increase in SSE is equivalent to minimizing the between cluster distances.

5.6 Dendograms

Dendograms graphically present the information concerning which observations are grouped together at various levels of (dis)similarity. At the bottom of the dendrogram, each observation is considered its own cluster. Vertical lines extend up for each observation, and at various (dis)similarity values, these lines are connected to the lines from other observations with a horizontal line. The observations continue to combine until, at the top of the dendrogram, all observations are grouped together. The height of the vertical lines and the range of the (dis)similarity axis give visual clues about the strength of the clustering. Long vertical lines indicate more distinct separation between the groups. Long vertical lines at the top of the dendrogram indicate that the groups represented by those lines are well separated from one another. Shorter lines indicate groups that are not as distinct.

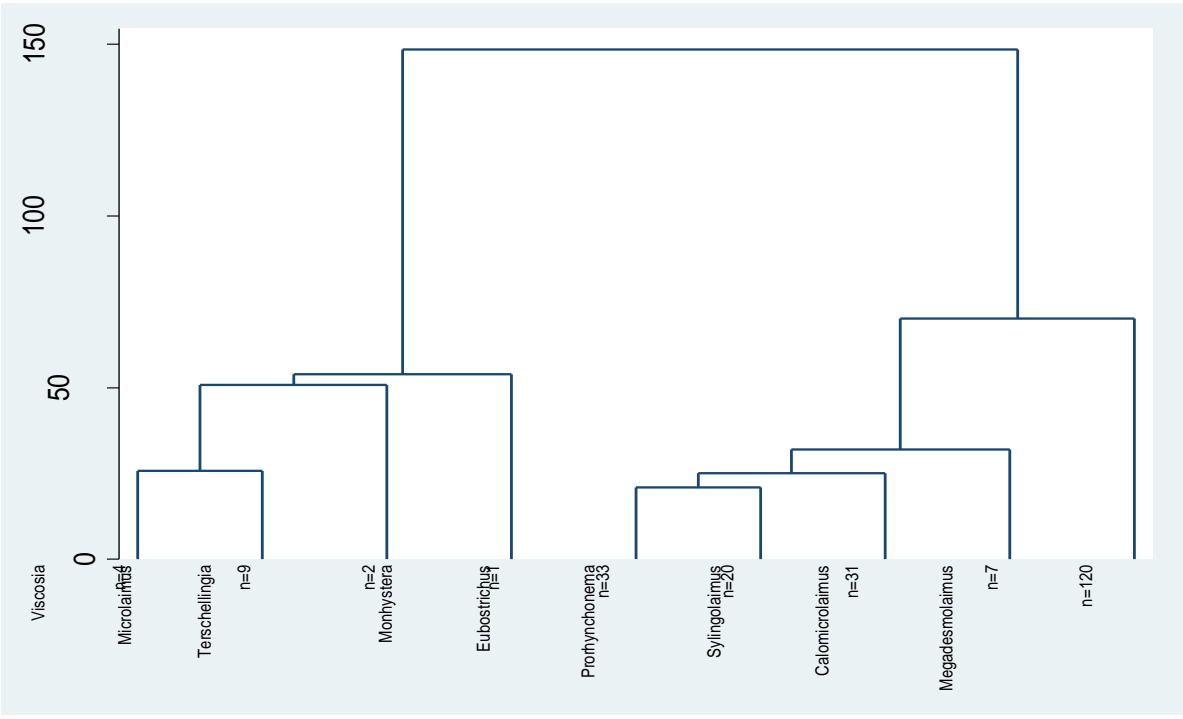


Figure15: Dendrogram for Nematode Composition based on the Depth using wald,s linkage

From the figure 15 above the Nematodes are clustered into two distinct groups in reference to their abundance in relative to the depth. From the dendrogram, *Monhystera* is in its own category and from the Figure 16 below it is shown the genus *Monhystera* is mostly found in deep sea. Thus it has a unique characteristic with most of the other genus. A tabulation of all the clusters is attached in appendix 2

5.7 Nematode density in relative to the depth

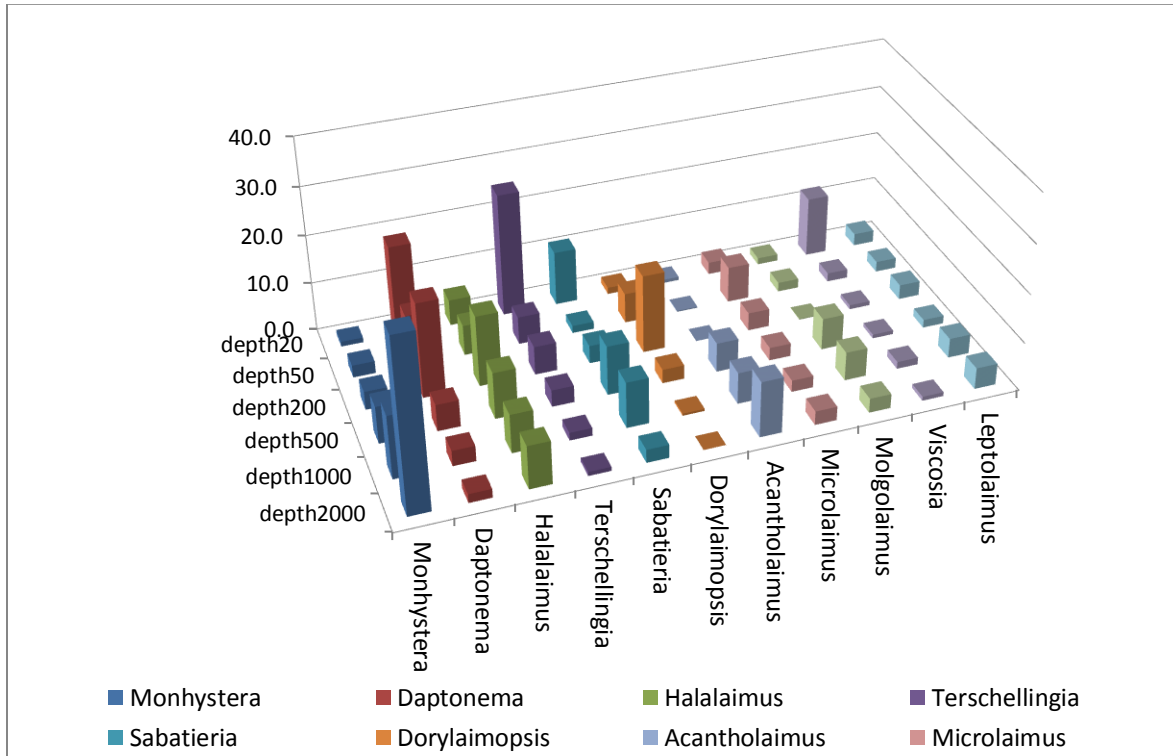


Figure 16: A graphical presentation of the nematodes abundance by depth

The above figure 16 represents eleven selected Nematodes in relation to their abundance. In the 2000m depth *Monhystera* dominated followed by *Acantholaimus*. The 500-1000m depth level was dominated by *Sabatieria* and *Halalaimus*. *Daptonema*, *Microlaimus* and *Dorylaimopsis* dominated the 50-200m depth. The 20 meter depth is dominated by *Terschellingia*, *Daptonema* and *Viscosia*

CONCLUSION

Nematode Density

Nematode density decreased with increase in water depth up to 1338m and then there was a slight increase. This result concurs with Muthumbi et al, 2004. By the use of general linear models procedure of fitting a curve the exact depth at which the nematode density was minima was determined as opposed to regression method used in Muthumbi et al, 2004. Food availability, oxygen concentration and the sediment composition are some of the factors that contributed to the variation in the nematode density. Oxygen concentration had a negative correlation with depth up to maximum depth of 1000m and then there was a positive correlation. This explains the slight increase of the nematode density after the 1338m depth. The most dominant genus is *Monhystera* which is mostly found in deep sea this also explains the increase of the density at 1338 meters. *Sabatieria* and *Halalaimus* dominated the poorly oxygenated depth level between 500m -1000m.

Transects effect on the nematode was significant with Kiwayu having the highest mean nematodes. Although the transect effect on the nematode density was significant, the mean difference between transect was small. Kiwayu which is on the north had the highest density of 421ind/cm² and Tana had the least density of 317.5ind/cm².

Nematode Composition

The most dominant genus in the deep sea was *Monhystera* which was also the most dominant genus in the Western Indian Ocean. Middle depth level between 200m-1000m was dominated by *Halalaimus* and *Daptonema*. The shallow continental shelf was dominated by *Terschellingia* and

Daptonema. The genus composition basically varied by the mode of feeding and the ability to tolerate physical factors like oxygen concentration, pH levels and water temperatures. Also the Sand –silt composition affected the composition with high percentage of sand found on the low continental shelf and more of silt in the deep slope. Two Way Indicator Species Analysis (TWINSpan) is a better procedure of classifying the Nematode (Muthumbi et al,2004) than the hierarchical agglomerative procedure of clustering. TWINSpan procedure includes both the variable (depth) and the genera in the dendrogram and hence it easier to classify the diversity of the Nematodes in regard to different depth classes.

REFERENCES

- Bongers, T. 1988. De Nematoden Van Nederland. Koninklijke Nederlandse Natuurhistorische Vereniging, Utrecht, The Netherlands.
- Boucher, G. & Lamshead, P.J.D. (1995) 'Ecological Biodiversity of Marine Nematodes in Samples from Temperate, Tropical, and Deep-Sea'. *Conservation Biology*, 9,(6) pp. 1594-1604
- Duineveld, G.C.A., De Wilde, P.A.W.J., Berghuis, E.M, Kok, A., Tahey, T., & Kromkamp, J. (1997) 'Benthic Respiration and Standing Stock on two contrasting continental margins in Western Indian Ocean; the Yemen-Somali upwelling region and the margin off Kenya'. *Deep-sea research II* 44, pp. 1293-1317
- Falcaro, M., and A. Pickles. (2010) 'riskplot: A graphical aid to investigate the effect of multiple categorical risk factors'. *Stata Journal* 10 pp. 61-68.
- Field, A.P. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* {4th ed.}. London: Sage.
- Giere, O. (2009). *Meiobenthology: The Microscopic motile fauna of aquatic sediments*, Second edition. Springer-verlag, Berlin.
- Gotelli, N.J. & Ellison, A.M. (2004) *A Primer of Ecological Statistics*. Massachusetts. Sunderland
- Grzelak, K. & Kotwicki, L. (2012) 'Meiofaunal distribution in Hornsund fjord, Spitsbergen'. *Polar Biol*, 35 pp. 269-280
- Heip, C., Vincx, M., & Vranken, G. (1985) 'The Ecology of Marine Nematode'. *Oceanogr. Mar. Biol. Ann. Rev.*, 23 pp. 399-489
- Heip, C. & Soetaert, K. (1995) 'Nematode assemblages of deep-sea and shelf break sites in the North Atlantic and Mediterranean Sea'. *Marine Ecology Progress Series*, 125 pp. 171-183
- Higgins, R.P., Thiel, H. (1988) *Introduction to the study of Meiofauna*, Smithsonian Institute press
- Kennedy, A.D., & Jacoby, C.A. (1999) 'Biological Indicators of marine environment health: Meiofauna a neglected benthic component'. *Environment monitoring and Assessment*, 54, pp 47-68
- Lamshead, P.J.D. (1993) 'Recent development in marine benthic biodiversity research'. *Oceanis*, 19 pp. 5-24

- Lambshead, P.J.D, Tietjen, J, Ferrero, T & Jensen, P. (2000) 'Latitudinal diversity gradients in the deep sea with special reference to North Atlantic nematodes'. *Marine Ecology Progress Series*, 194 pp. 159-167
- Lorenzens, S. (1981). 'Entwurf eines phylogenetischen systems der freilebenden Nematoden' *Recherches de l'Institut pour l'océanographie de Bremerhaven* 7, pp1-142
- Mardia, K. V. 1970. Measures of multivariate skewness and kurtosis with applications *Biometrika* 57: 519-530.
- Miller, G.A, & Chapman, J.P.(2001) ' Misunderstanding analysis of covariance'. *Journal of Abnormal Psychology*, 110(1), 40-48.
- Muthumbi, A.W., Vanreusel, A., Duineveld, G., Soetaert, K., & Vincx, M. (2004) 'Nematode Community Structure along the Continental Slope off the Kenyan Coast, Western Indian Ocean'. *Int. Revue ges. Hydrobiol*, 89(2) pp. 188-205
- Pawlowska, J, Wlodarska-Kowalczyk, M, Zajaczkowski, M, Nygard, H, & Berge, J. (2011) 'Seasonal Variability of Meio and Macrobenthic Standing Stocks and Diversity in an Arctic fjord (Adventfjorden, Spitsbergen)' *Polar Biol*, 34 pp. 833-845
- Plat, H.M. (1982) 'Revision of Ethohaimade (Nematoda: Chromadorida)'. *Bulletin of British Museum* 43 pp 185-252
- Rencher, Alvin. C (2002), *Methods of Multivariate Analysis* 2nd ed. John Wiley & Sons, Inc. New York
- Semeneh, M, Dehairs, M, Elskens, M, Baumann, M. E.M, Kopczynska, E.E, Lancelot, C, Geoyens, L. (1988) 'Nitrogen uptake regime and phytoplankton community structure in the Atlantic and Indian sector of the southern' *Journal of marine system* 17 pp. 159-177
- Soetaert, K., Vincx, M., & Heip, C. (1995) 'Nematode Community Structure Along a Mediterranean Shelf-slope Gradient'. *Marine Ecology*, 16(3) pp. 189-206
- Stickland, J.D.H & Parsons, T.R (1972). *A practical Handbook of seawater Analysis*. Second edition, Bulletin 167. Fisheries Research Board of Canada, Ottawa
- Thilagavathi, B, Das, B, Saravanakumar, A, Raja, K. (2011) 'Benthic Meiofaunal Composition and Community Structure in the Sethukuda Mangrove Area and Adjacent Open Sea, East Coast India'. *Ocean Science Journal*, 46(2) pp. 63-72
- Timm, N.H (2002) *Applied Multivariate Analysis*. Springer-Verlag, Inc. New York

Vanreusel, A., Vincx, M., Bett, B. J., & Rice, A. L. (1995) 'Nematode Biomass Spectra at Two Abyssal Sites in the NE Atlantic with a Contrasting Food Supply'. *Int. Revue ges. Hydrobiol.*, 80(2) pp.287-296

Xuan, Q.N., Vanreusel, A., Thanh, V.N., & Smol, N. (2007) 'Biodiversity of Meiofauna in the Intertidal Khe Nhan Mudflat, Can Gio Mangrove Forest, Vietnam with Special Emphasis on Free Living Nematodes'. *Ocean Science Journal*, 42(3) pp. 135-152

APPENDICES

Appendix 1

STATA codes

**Descriptive Statistics

```
tabstat NEM_DEN, by( DEPKLAS) stat(mean median semean max min skewness kurtosis)
```

```
tabstat NEM_DEN, by ( TRANSECT) stat(mean median semean max min skewness kurtosis)
```

```
tabstat NEM_DEN, by ( SEASON) stat(mean median semean max min skewness kurtosis)
```

**Checking for the normality assumption

```
graph hbox NEM_DEN, title (Box Plot) scheme(s2mono) ytitle(Nematode Densities)
```

```
hist NEM_DEN, normal title (Hitogram Nematode Densities ) xtitle (Nematode  
densities) scheme(s2mono)
```

```
qnorm NEM_DEN, title (Normal QQ Plot) scheme(s2mono) ytitle(Nematode Densities)
```

**Transforming the data

```
gen LOGNEM_DEN= log10(NEM_DEN)
```

** Checking the normality assumption after transforming the data

```
graph hbox LOGNEM_DEN, title (Box Plot Transformed Nematode Densities) scheme(s2mono)  
ytitle(Transformed Nematode densities)
```

```
hist LOGNEM_DEN , width(0.2) normal title (Hitogram Transformed Nematode Densities ) xtitle  
(Transformed Nematode densities) scheme(s2mono)
```

```
qnorm LOGNEM_DEN, title (Normal QQ Plot for Transformed Nematode Densities) scheme(s2mono)  
ytitle(Transformed Nematode Densities)
```

```
sktest LOGNEM_DEN
```

**Testing the corration of the covariates

```
pwcorr ORGC C_N__MOL SED__PHA SED__CHLO SCOC__DE DNA_RNA OXYGEN__ v18  
SURF_CHL BOTT_CHL
```

```
pwcorr ORGC C_N__MOL SED__PHA SED__CHLO SCOC__DE DNA_RNA OXYGEN__ v18  
SURF_CHL BOTT_CHL if SEASON==2
```

```
pwcorr SD50MUC SMEDIUM SFINES SVFINES SSILT16 SSILT50 SSILT63 if SEASON==2
```

```
pwcorr SD50MUC SMEDIUM SFINES SVFINES SSILT16 SSILT50 SSILT63 ORGC C_N__MOL
SED__PHA SED__CHLO SCOC__DE DNA_RNA OXYGEN__ v18 SURF_CHL BOTT_CHL if
SEASON==2
```

**Running analysis of variance

```
anova LOGNEM_DEN TRANSECT SEASON DEPKLAS
anova LOGNEM_DEN TRANSECT DEPKLAS SEASON DEPKLAS# SEASON
anova LOGNEM_DEN TRANSECT DEPKLAS c.DNA_RNA c.SED__PHA c.SFINES c.ORG
tab DEPKLAS SEASON ,summarize( NEM_DEN) means
```

**TEST THE SIGNIFICANCE DIFFERENCE IN MEANS

```
tukeyhsd TRANSECT
```

**Fitting a curve.

```
gen depth2 = DEPTH^2
gen depth3 = DEPTH^3
gen depth4 = DEPTH^4
nestreg, quietly: reg NEM_DEN DEPTH(depth2 depth3 depth4)
curvefit NEM_DEN DEPTH if SEASON == 1, f(4) scheme(sj)
curvefit NEM_DEN DEPTH if SEASON == 2, f(4) scheme(sj)
curvefit NEM_DEN DEPTH , f(4)scheme(sj)
anova SQRT TRANSECT DEPKLAS SEASON DEPKLAS# SEASON
margin DEPKLAS#SEASON if DEPKLAS !=20
matrix list r(b)
matrix b=r(b)'
matrix list b
matrix dpth=(50\200\500\1000\2000)#(1\1) /* # is the kronecker product operator */
matrix list dpth
```

```

matrix sesn=(1\1\1\1\1)#(1\2) /* # is the kronecker product operator */
matrix list sesn
matrix c=dpth,sesn,b
matrix list c
svmat c, names(c)
clist c1-c3 in 1/7
graph twoway (connect c3 c1 if c2==1 & c1 != 20)(connect c3 c1 if c2==2 & c1 != 20), ///
    xlabel(50 200 500 1000 2000) legend(order(1 "season=1" 2 "season=2")) ///
    xtitle(depthklass) ytitle(ajdusted cell means)///
    name(depth_by_seasons, replace)

```

Cluster Analysis

```

cd "E:\masters\project"
insheet using clusteranalysis.csv,clear case names
log using cluster.log, replace
cluster wardslinkage depth20 depth50 depth500 depth1000 depth2000, measure(L2)
cluster dendrogram _clus_1, labels( genera)
cluster dendrogram _clus_1,cutnumber(9) labels( genera)showcount
log cl

```

Appendix 2

Table of the clusters

Case	Clusters	Cases	Clusters
1:Acantholaimus	1	153:Theristus	2
22:Cervonema	1	156:Xyala	2
89:Molgolaimus	1	157:Xyalid	2
99:Leptolaimus	1	161:Sphaerolaimus	2
130:Eumonhystera	1	163:Astomonema	2
131:Monhysterid	1	167:Disconema	2
136:Amphimonhystrella	1	172:Metalinhomoeus	2
2:Actinonema	2	177:Paradontophora	2
7:Dichromadora	2	178:Parareolaimus	2
11:Hypodontolaimus	2	180:Campylaimus	2
16:Ptycholaimellus	2	181:Diplopeltula	2
20:Trochamus	2	200:Sylingolaimus	2
26:Laimella	2	204:Litinium	2
29:Paracomesoma	2	205:Oxystomina	2

35:Comesa	2	209:Viscosia	2
41:Neotonchus	2	221:Rhabdocoma	2
46:Maryllynnia	2	3:Chromadora	3
47:Metacyatholaimus	2	4:Denticulella	3
52:Paralongicyatholaimus	2	5:Chromadorella	3
58:Gammanema	2	6:Chromadorita	3
61:Richtersia	2	8:Endeolophos	3
65:Chromaspirina	2	9:Euchromadora	3
66:Desmodora	2	10:Graphonema	3
67:Desmodorella	2	15:Parapinanema	3
78:Psammonema	2	17:Parachromadorita	3
82:Aponema	2	19:Spiliphera	3
83:Bobolaimus	2	27:Metacomesoma	3
84:Calomicrolaimus	2	31:Pierrickia	3
98:Leptolaimoides	2	33:Setosabatieria	3
105:Cyartonema	2	34:Vasostoma	3
106:Diplopetloides	2	36:Filitonchus	3
113:Pselionema	2	37:Filoncholaimus	3
126:Quadricoma	2	38:Gophionchus	3
132:Monhystrella	2	39:Nannolaimus	3
137:Cobbia	2	40:Nannolaimoides	3
140:Elzalia	2	43:Cyatholaimus	3
141:Gnomoxyala	2	44:Kraspedonema	3
145:Paramonhystera	2	48:Minolaimus	3
150:Rhynchonema	2	51:Paracyatholaimus	3
72:Notochaetosoma	3	56:Choanolaimus	3
73:Metepsilonema	3	57:Demonema	3
75:Paradraconema	3	60:Latronema	3
76:Parallelocoides	3	62:Synonchiella	3
77:Polysigma	3	63:Synonchium	3
80:Sigmophoranema	3	68:Draconema	3
86:Crassonema	3	69:Epsilonema	3
90:Spirobbolaimus	3	71:Leptolaimella	3
93:Cricolaimus	3	128:Diplolaimella	3
94:Dagda	3	129:Diplolaimelloides	3
95:Deontolaimus	3	134:Ammotheristus	3
96:Diodotolaimus	3	139:Echinotheristus	3
97:Halaphanolaimus	3	143:Manganonema	3
101:procamacolaimus	3	144:Metadesmolaimus	3
107:Southernia	3	148:Retrotheristus	3
108:Chitwoodia	3	149:Rhinema	3
109:Tubolaimoides	3	151:Scaptrella	3
117:Desmogerlachia	3	152:Stylotheristus	3
120:Domorganus	3	154:Valvalaimus	3
121:Gerlachius	3	155:Xenolaimus	3
122:Greeffiella	3	160:Parasphaerolaimus	3
123:Hapalomus	3	162:Subsphaerolaimus	3
124:Meylia	3	164:Siphonolaimus	3
218:Tripyloides	3	165:Desmolaimus	3
219:Pandolaimus	3	166:Didelta	3
220:Halanonchus	3	168:Eleutherolaimus	3
222:Trefusialaimus	3	169:Eumorpholaimus	3
223:Trefusia	3	171:Megadesmolaimus	3
224:Lauratonema	3	173:Paralinhomoeus	3
225:Rhabdolaimus	3	175:Axonolaimus	3
226:Rhabdonema	3	176:Odontophora	3
12:Innocuonema	4	183:Coninckia	3
13:Neochromadora	4	184:Enoplaimus	3
14:Prochromadorella	4	186:Epicanthion	3
18:Rhips	4	187:Mesacanthion	3
21:Trichromadora	4	188:Mesacanthoides	3
24:Hopperia	4	189:Oxynchulus	3
25:Kenyanema	4	191:Anoplostoma	3
28:Metasabatieria	4	192:Chaetosoma	3
30:Paramesonchium	4	193:Capsula	3
42:Acanthonchus	4	194:Crenopharynx	3

45:Longicyatholaimus	4	195:Micoletzkyia	3
49:Paracanthochus	4	196:Phanoderma	3
50:Paracyatholaimoides	4	197:Anticoma	3
53:Pomponema	4	198:Dolicholaimus	3
54:Praeacanthochus	4	199:Parironus	3
55:Cheironchus	4	201:Thalassironus	3
59:Halichoanolaimus	4	206:Paroxystoma	3
64:Catanema	4	210:Filoncholaimus	3
70:Eubostrichus	4	211:Oncholaimus	3
74:Onyx	4	213:Belbolla	3
79:Pseudonchus	4	214:Eurystomina	3
81:Spirinia	4	215:Pareurystomina	3
85:Cinctonema	4	216:Polygastrophora	3
87:Ixonema	4	217:Bathylaimus	3
91:Antimicron	4	125:Pareudesmoscolex	4
92:Camacolaimus	4	127:Tricoma	4
100:Onchium	4	135:Amphimonhystera	4
102:Setoplectus	4	142:Linhystera	4
103:Tarvaia	4	146:Promonhystera	4
104:Aegiololaimus	4	147:Prorhynchonema	4
110:Ceramonema	4	158:Doliolaimus	4
111:Dasynemoides	4	159:Metasphaerolaimus	4
112:Metadasynemella	4	170:Linhomoeus	4
114:Pterygonema	4	179:Areolaimus	4
115:Paramicrolaimus	4	182:Southerniella	4
116:Calligyus	4	185:Enoploides	4
118:Desmorolenzenia	4	190:Paramesacanthion	4
119:Desmoscolex	4	202:Trissonchulus	4
88:Microalaimus	5	207:Thalassoalaimus	4
138:Daptonema	5	208:Wieseria	4
174:Terschellingia	5	212:Bathyeurystomina	4
203:Halalaimus	5	23:Dorylaimopsis	5
133:Monhystera	6	32:Sabatieria	5