# LOG RATIO METHODOLOGY FOR ANALYSIS OF COMPOSITIONAL DATA: A CASE OF OLFACTOMETER BIOASSAY DATA FROM INSECT BEHAVIOURAL STUDIES

**ANTHONY RAYMOND EPEL**

**MASTER OF SCIENCE**

**(Research methods)**

**JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY**

**2013**

# Log Ratio Methodology for Analysis of Compositional Data: A case of Olfactometer Bioassay Data from Insect Behavioural Studies

**Anthony Raymond Epel**

**A thesis submitted to the Department of Agriculture in partial fulfillment of the requirement for the award of the Degree of Master of Science in Research Methods of Jomo Kenyatta University of Agriculture and Technology**

**2013**

## DECLARATION

This thesis is my original work and has not been presented for award of a degree in any other University.

Signature……. Date …….

**Anthony Raymond Epel**

This thesis has been submitted with our approval as University Supervisor.

Signature……. Date ……

**Dr John Mwibanda Wesonga**

JKUAT, Kenya

Signature……. Date ……

**Dr Daisy Salifu**

ICIPE, Kenya

## DEDICATION

To my wife **Esther Akello** and children: **Ernest Epel**, **Michelle Akol,** and **Louis Magnus Epel**.

## ACKNOWLEDGEMENT

Firstly, I thank the Almighty God that I have reached this level of the academic ladder; it is all because of His mercy and grace. May His purpose for this favour be fulfilled. I am very grateful to the following people for their support in this study: Doctors John Mwibanda Wesonga and Daisy Salifu, my supervisors at university and International Centre of Insect Physiology and Ecology (*ICIPE*), respectively. Their guidance and encouragement kept me inspired throughout this study. Special thanks go to Dr Charles Midega, research scientist push-pull project at Thomas Odhiambo, Mbita. His intellectual guidance helped so much in this study. Appreciations are extended to Mr. Cyrus Oyiko, senior technician at Thomas Odhiambo, Mbita for his practical input during laboratory bioassays. Further, I am indebted to the Capacity Building and Institutional Development (CB&ID) Programme at *ICIPE* for effectively handling all my training matters.

I also thank all members of staff and fellow students of the Department of Horticulture at JKUAT for their assistance in various ways during the course of this research work. Finally, I recognize the Regional University Forum for Capacity Building in Agriculture, for sponsoring my studies.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

**LIST OF APPENDICES**

## LIST OF ABBREVIATIONS/ ACRONYMS

**ALR:** Additive Log ratio Transformation

**ANOVA**: Analysis of Variance

**CLR:** Centered Log ratio Transformation

**CoDa:** Compositional Data

**DRIP:** Dissertation Research Internship Programme

**GPL:** General Public License

**HCL:** Hydrochloric Acid

**HIPVs:** Herbivore Induced Plant Volatiles

**ICIPE:** International Center of Insect Physiology and Ecology

**ILR:** Isometric Log ratio Transformation

**JKUAT:** Jomo Kenyatta University of Agriculture and Technology

**KPa:** Kilo Pascals

**LSD:** Least Significant Difference

**$NH_4OH$:** Ammonium Hydroxide

**OIPVs:** Oviposition Induced Plant Volatiles

**OPVs:** Open Pollinated Varieties

**RUFORUM:** Regional University Forum for Capacity Building in Agriculture

**R.H:** Relative Humidity

**VOCs:** Volatile Organic Compounds

## ABSTRACT

Researchers are often times confronted with compositional data in insect choice studies. The choice of a statistical method to model this type of data is always not obvious. In this study, three approaches for analysis of compositional data from choice tests made by the predatory parasitoid *Cotesia sesamaie* Cameron in a four-arm olfactometer was explored using centered, additive and isometric log ratio transformations. Oviposition induced plant volatiles (OIPVs), herbivore induced plant volatiles (HIPVs), and two controls were tested in the four-arm olfactometer. The response variable measured was time spent in each field of the olfactometer, when the time to observe the insect was restricted to 12 minutes. The data generated in this study was compositional, thus it conveys exclusively relative information and has a constant sum constraint such that standard statistical methods of analysis (ANOVA, t-test), cannot be used on this data. This study therefore explored the log ratio methodology advocated by Aitchison (1986). *CLT, ALT,* and *ILR* log ratio transformations were then performed using CoDaPack statistical software. Using this methodology, mean differences in olfactometer response of female parasitic wasp, *Cotesia sesamiae* to OIPVs, HIPVs, and control were computed. These findings imply that the *CLR* transformation is probably the best choice for processing raw compositional data prior to analysis by standard statistical methods. These results revealed that the, parasitic wasps spent much time in olfactometer arm with OIPVs, followed by the olfactometer arm with HIPVs and lastly spent least time in the control arm of the olfactometer. More studies need to be conducted using the log-ratio methodology on olfactometer bioassay data from a different species of parasitic wasps.

# CHAPTER 1

# INTRODUCTION

## 1.1 Background to the study

Demonstration of responses of insects to their hosts and host habitats is the subject matter in insect behavioral studies (Chen-Cheng 2010). Insects respond to different olfactory cues like volatiles (chemicals secreted by plants to attract pollinators, predators, repel pests and pathogens), host odour and pheromones for mate searching and aggregation (Vet & Dicke 1992, Reed *et al* 1995, and Bruce *et al* 2009) among others. In insect behavioural studies, the researchers measure responses to odours in the laboratory using olfactometers and flight tunnels (Noldus *et al* 1990). There are various forms of olfactometers used to detect and measure insect responses to multiple odour cues, such as Y-shaped glass tube olfactometer (Janssen *et al* 1997), four-arm olfactometer (Vet *et al* 1983), T-shaped linear tract olfactometer (Sakuma and Fukami, 1985), and six-arm olfactometer (Ted *et al*, 2004).

The Y-shaped olfactometer contains Y-shaped glass tube at its base. During operation of this olfactometer, the experimenter introduces the test odour to the insects released at the opposite side of the olfactometer using the Y-shaped glass tube. When the insect walks towards a particular odourised arm of the olfactometer, it indicates its preference for that odour. The T-shaped olfactometer has the mechanism of operation as that of Y-shaped olfactometer. The four-arm olfactometer has four distinct odourized fields are created in its chamber. The test insect is introduced into the olfactometer chamber through the

centre of the odourised fields. In principle, four different odour sources can be tested simultaneously, but usually two types of odours are the norm (Tamiru *et al*., 2011; Said *et al*., 2006). The six-arm olfactometer is similar to the four-arm olfactometer, except that one can compare larger number of odour sources and insects can be released in groups (Turlings *et al*., 2004).

The response variables collected in olfactometer experiments are mainly two; the number of insects that choose an odourized field at the beginning of the bioassay (first choice made by insect) and time spent in each field, where the time to observe the insect vary from 1 to 30 minutes (Suazo *et al* 2003, Gohole *et al.,* 2003, Shepherd *et al* 2005, Oluwafemi *et al*., 2011, Ukeh, 2008, Addesso *et al*, 2010, Jonfia-Essien *et al* 2007), and Ninkovic *et al*, 2011).

When analyzing insect count data from Y-shaped or T-shaped olfactometer bioassays, One would use a binomial test to compare the probability of choice in the two arms assuming that each insect response is entirely independent of one another and that the probabilities are constant across all replicates with the same experimental conditions (Sullivan *et al* 2000). The same principle applies for the four and six-arm olfactometer bioassays. In this situation counts data would be assumed to follow a multinomial distribution with probabilities $P_1$, $P_2$, $P_3$, $P_4$ and $P_1$, $P_2$, $P_3$, $P_4$, $P_5$, $P_6$ respectively (Turlings *et al.,* 2004). Such a model accounts automatically for the dependence between the numbers of insects choosing the different arms. Its underlying assumptions are that

individual insect act entirely independently of one another and that the $P_i$ are constant across all replicates with the same experimental conditions. If these assumptions are true, it is possible to model how the probabilities depend on the experimental treatments (Turlings *et al*., 2004). However, when the response is time spent (a continuous variable) by insect in each arm, the appearance of the data is that of real numbers. The data collected by olfactometer assays is compositional in nature, because the data collected from each insect is simply a partition of the total time allocated for each observation (Pawlowsky-Glahn *et al* 2012).

Compositional data (CoDa) consist of vectors whose components are the proportions or percentages of some whole. Their peculiarity is that their sum is constrained to be some constant, equal to 1 for proportions, 100 for percentages or possibly some other constant say *k.* In other words, any increase in the value of a data point automatically requires the other data points to decrease, demonstrating the "constant sum constraint" of such data (Aitchison and Egozcue, 2005). This non-independence of data points restricts the application of standard, statistical techniques such as analysis of variance (ANOVA), t-test, Principal component analysis among others. This is because all these procedures implicitly assume a data distribution, independence of data points, as well as absence of interactions between data points (Aebischer *et al* 1993).

Unfortunately these standard statistical methods are being used for analyzing raw compositional data, in disregard of the fact that the sample space (a set of all possible outcomes of a variable) is different from the common Euclidean sample space in which

those techniques operate (Aitchison & Egozcue, 2005). The results and inferences made from such techniques of analysis though some times interpretable may be incorrect (Pawlowsky-Glahn *et al* 2012. To avoid such problems Aitchison (1986), suggested that the data be transformed using appropriate transformations, which preserve the geometry of compositional data on the simplex (the support space of compositional data). These transformations are additive log ratio transformation (*ALR*), isometric log ratio transformation (*ILR*) and centered log ratio transformation *(CLR*) the most often used (Aitchison 2008).

Compositional data frequently occur in many disciplines. Examples are compositions of pollutants in envirometrics, household budget compositions in economics, blood and urine compositions in medicine, food composition in food science and time compositional data in insect behavioural studies. The intuition of compositional data from olfactometer assays in insect behavioural studies is described below.

Consider a four-arm olfactometer bioassay in which an insect is given 12 minutes in total to spend around the arms of an olfactometer. Assume the insect spends 3 minutes in one arm, and then only 9 minutes left. The second time the insect 'makes a decision'; the number of minutes available to be spent during the second visit will depend on how many minutes were spent during the first visit. The fewer minutes spent during the first visit, the more minutes available for subsequent visits, and vice versa, giving rise to dependency over visits or time spent. This means the total time T is broken down into components $t_1$, $t_2$, $t_3$, $t_4$ given the four-arm olfactometer and this translates into

proportion time spent of the total time; $p_1$, $p_2$, $p_3$, $p_4$ to give a total of 1 or 100% (Pawlowsky-Glahn & Buccianti 2011). Despite the guiding principles provided by Aitchison (1986) for handling compositional data, the application of such principles for compositional data analysis especially in the biological sciences is still patchy and many researchers are unaware of the appropriate concepts for analysis of compositional data.

### 1.2 Statement of the Problem

The choice of appropriate statistical method for the evaluation of the response "time spent" from olfactometer studies is not always straightforward. In two choice instances, the raw data had mistakenly been analyzed using paired t-test or its non-parametric counterparts, or analyzed as percentages using Chi-square Addesso *et al.,* (2010). While Jan de Kogel *et al., (*1999) used a binomial test for analysis of the same data type. For multiple choice instances, Reed (1995), Said *et al* (2006), and Ninkovic *et al* (2011), used Kruskal-Wallis / Wilcoxon rank test, Freidman ANOVA and ANOVA respectively on the raw compositional data. While Ukeh (2008) used a pair wise t-test to compare the means. Such standard statistical methods designed for unconstrained data are not appropriate for raw compositional data, owing to their sample space and would undoubtedly lead to inappropriate inferences (Aebischer *et al* 1993). This study explored the log ratio methodology developed by Aitchison, (1986) for analysis of compositional data from olfactometer bioassays.

## 1.3 Objectives

### 1.3.1 General Objective

To investigate log ratio methodology in the analysis of olfactometer compositional data and provide guidance in data transformation and analytical procedures, to researchers engaged in insect behavioural studies.

### 1.3.2 Specific Objectives

i. To investigate the log-ratio transformations in the analysis of compositional data from olfactometer bioassays

ii. To determine the mean time spent by *Cotesia sesamiae* Cameron in the OIPVs, HIPVs, and control arms of the olfactometer bioassay

## 1.4 Hypotheses

i. $H_{0:}$ There is no difference between the mean times spent in the olfactometrer arms containing OIPVs, HIPVs and Control.

## 1.5 Significance of the Study

This study acts as a benchmark for researchers encountering compositional data. It has teased out Centered log-ratio transformation for processing compositional data as advocated by Aitchison. It has also brought attention to CoDapack, freely available statistical software for compositional data analysis.

### 1.6 Outline of the thesis

Chapter 1 looked at the background of the study, statement of the problem, research objectives, and hypotheses and stated the significance of the study. Chapter 2 presents history of compositional data analysis, principles, and sample space of compositional data, Aitchison's log ratio methodology, and a basic description of a four-arm olfactometer. Chapter 3 discusses the sources of insects' parasitoids and sources of volatiles used in the study, describes four-arm olfactometer bioassay and the statistical methods used in the study. Chapter 4 presents and discusses the results; finally, in Chapter 5 conclusions and recommendations are given.

# CHAPTER 2

# REVIEW OF LITERATURE

## 2.1. History of compositional data analysis

Compositional data is defined as the data which consists of vectors whose components are proportions or percentages of some whole Aitchison (1986). Summation of these data usually results in 1 or 100 for proportions and percentages respectively. Aitchison & Egozcue (2005) paper shows that such vector of compositional data have mistakenly been taken as real vectors in statistical analysis with disregard of nature that this data is constrained and hence not amenable for analysis by statistical techniques designed for an unconstrained data. The result of this inappropriate analysis may have led to inappropriate inferences as well. The inappropriateness of standard statistical methods designed for unconstrained data stems from the fact that the sample space for compositional data is not the same with that of Euclidean space associated with real vectors (Pawlowsky-Glahn *et al* 2011). Despite of the warnings by Chayes (1960) in Aitchison and Egozcue (2005), regarding the use of standard statistical methods designed for Euclidean space for analysis of compositional data, the trend has continued unabated. Below is a break down into four phases the historical attempts to model compositional data.

In the phase before 1960, statisticians and scientists were using standard multivariate statistical methods of analyses, which by default are tailored for statistical analysis of data with real sample space/ Euclidean sample space (Aitchison 2003. These standard

multivariate techniques of analysis were employed indiscriminately for analysis of any data set regardless of whether its sample space was appropriate for the method or not. A case in point was the use of standard correlation analysis for compositional data vectors. Such methods has no inbuilt features to model constrained/compositional data, therefore were not suitable for analysis of compositional data in its raw form.

In Phase two, the use of standard multivariate statistical methods of analysis came under criticism. Mainly due to the challenge in interpretation of product-moment correlation between components of a geochemical composition, especially in regard to the negative bias; a distorting factor which occurred during interpretation of results Chayes (1960) and Sarmonov (1959) in Pawlowsky-Glahn *et al* (2011). These critisms were directed to those applying standard multivariate statistical methods for analysis of compositional data in the field of geology. While in the field of biology, Mosimann (1963) in Pawlowsky-Glahn *et al* (2011) extended the criticism of Chayes and Sarmonov for biological applications. It is however regrettable that in spite of all these critisms the application of standard statistical methods of analysis on constrained data has continued with the approach of distorting standard statistical methods; rather than adopting an appropriate method of analysis (Aitchison & Egozcue 2005).

Phase three is characterized by the work by Aitchison (1982) who noted that compositions provide information about relative, not absolute, values of components. In other wards every statement about a composition can be stated in terms of ratios of

components. During this phase log ratios were perceived to be easy to handle mathematically as compared to ratios and therefore transformations using additive log ratio and centred log ratio transformation methodology became immensely popular in the analysis of constrained / closed data (Aitchison 1986). What was unique in those log ratio transformations was that, they were capable of converting compositional data from the simplex sample space into real / Euclidean sample space. This provided the possibility of using standard statistical techniques for analysis of log ratio transformed data (Aitchison 2003). The advantages of the log ratio methodology of processing compositional data before analysis by standard statistical methods of analysis were twofold: Inferences obtained would be translatable to their original compositional statements, and it converted compositional data from simplex sample space to real Euclidean space (Pawlowsky-Glahn *et al* 2011).

The fourth phase in compositional data analysis began after the paper by Billheimer *et al*., (1997) as cited in Pawlowsky-Glahn *et al*., (2011). In this phase compositional data is analysed in the raw form by applying operations such as perturbing, powering. The authors noted that when the above operations are applied on the raw vectors of compositional data the analysis would precede without the use of any log ratio transformations as advocated in the third phase above. For the sake of this study the log ratio methodology advocated by Aitchison (19860 was used in the analysis of compositional data. In the sections below are given brief descriptions of the principles

for compositional data analysis, the sample space for compositional data analysis and log ratio transformations for processing raw compositional data.

**2.2 Principles of Compositional data analysis**

These are conditions, which must be fulfilled during analysis of compositional data by any statistical method. They include the following:

Principle of scale invariance: According to (Aitchison 2003), this principle states that, ‘‘when a problem is compositional, the sizes of the specimens are irrelevant’’. Aitchison has illustrated this principle with an example with two specimen vectors $w = (1.6, 2.4, 4.0)$ and $W = (3.0, 4.5, 7.5)$. These vectors represent weights of the three parts (a, b, c) of two specimens of total weight 8 grams and 15 grams, respectively. This principle means that, the above vectors have the same composition with their differing weights taken account of by the scaling relationship given as $W = \left(\frac{15}{8}\right)w$.

Principle of sub compositional coherence: This principle states that sub compositions should behave as orthogonal projections do in conventional real analysis (Aitchison 2003). Below is an example from Aichison to illustrate this principle. Consider scientist A and B working on soil samples divided into aliquots. Scientist A records a 4-part composition (animal, vegetable, mineral, water); while scientist B first dries each aliquot without recording the water content and obtains a 3-part composition (animal, vegetable,

and mineral). According to the above principle the 3-part composition of scientist B $[s_1, s_2, s_3]$for an aliquot should be a sub composition of scientist A's 4-part composition$[x_1, x_2, x_3, x_4]$. This implies that any compositional statements that A and B make about the common parts; animal, vegetable, and mineral must agree. Therefore, it is possible for one to analyze data from a sub-composition and then extend the inferences to the whole composition. This has advantages because it helps in optimizing resources especially if obtaining samples for the whole composition is expensive.

**2.3 Compositional data and sample space**

In any statistical analysis it is important to identify the sample space which will represent the data (Aitchison 2008). Therefore in order to analyse compositional data there is need to recognise its sample space. Below is a typical compositional data vector.

$$x = [x_1, x_2, ..., x_D], \tag{2.1}$$

This vector is defined as a D-part composition when all its components are strictly positive real number, and carry only relative information (Chayes, 1971) as cited in Pawlowsky-Glahn *et al* (2011). This implies that the sample space for compositional data occurs in a simplex as seen in equation (2.2).

$$S^D = \{x = [x_1, x_2, ..., x_D] / x_i > 0, i = 1, 2, ..., D; \Sigma_{i=1}^{D} x_i = 1\} \tag{2.2}$$

where $S^D$ = parts.

The vector of D real positive component is given by:

$$Z = [z_1, z_2, ..., z_D] \in \Re_+^D \tag{2.3}$$

where ($z_i$>0 for all i= 1, 2, D)

The closure of Z is defined as:

$$C(Z)=\left[\frac{K.z_1}{\sum_{i=1}^{D} z_i},\frac{K.z_2}{\sum_{i=1}^{D} z_i},...,\frac{K.z_D}{\sum_{i=1}^{D} z_i}\right]. \quad (2.4)$$

The results of these equations are the same vectors rescaled so that all the components sum to a constant (*C*). This therefore calls for log ratio transformations which have the ability to open this type of data from its closed sample space (Equation 2.4) to the Euclidean sample space (Equation 2.5).

$$D=\sqrt{\sum_{i=1}^{s}(a_i-c_i)^2} \quad 2.5$$

**2.4 Log-ratio statistical method for compositional data analysis**

The method of log-ratio analysis for compositional data problems arose in the 1980's out of the realization of the importance of the principle of scale invariance Pawlowsky-Glahn *et al* (2011). The initial approach followed when analyzing compositional data in regards to the principle of scale invariance was to use ratios; but due to the awareness that logarithms of ratios are mathematically more tractable than ratios led to the advocacy of a transformation technique-involving log-ratios of components (Aitchison 1986). The three log ratio transformations appropriate for compositional data are briefly described below:

**Additive Log Ratio (*ALR*) Transformation:** This transforms raw compositional data from simplex to real space/Euclidean space. This log ratio transformation is also capable of performing its inverse transformation that is from real space/Euclidean space back to the simplex, with its inverse *ALR*-1 (Aitchison 2003). The distinguishing feature of *ALR* transform from the other log ratio transformations is that, it maps a composition in the D-part Aitchison simplex none isometrically to a D-1 dimensional Euclidean vector. As it maps, the last part is treated as a common denominator of the others. This has implications in that if the part in the denominator is changed; different *ALR* transformations would be obtained (Equation 2.6)

$$y = alr(x) = \left[ \log\left(\frac{x_1}{x_D}\right) \log\left(\frac{x_2}{x_D}\right) ... \log\left(\frac{x_{D-1}}{x_D}\right) \right], \tag{2.6}$$

in equation (2.6), the ratios involve the division of each of the first D-1 components by the final component.

The inverse transformation $ALR^{-1} : R^{D-1} \rightarrow S^D$ is

$$x = alr^{-1}(y) = C[\exp(y_1)\exp(y_2)...\exp(y_{D-1})1], \tag{2.7}$$

where $C$ denotes the closure operation (Aitchison 2003).

In the transformed state (Equation 2.6), the data can then be analyzed by all standard statistical methods of analysis not relying on a distance. This therefore means that statistical methods such as ANOVA and t-test should not be used for analysis of *ALR* transformed data. The interpretation of the results from *ALR* transformed data is relatively simple, since the relation to the original D-1 first parts is preserved. Another weakness of *ALR* transformation is that, it is not an isometric transformation from the

simplex. This weakness could be solved by use of an appropriate metric with oblique coordinates in real *ALR*-space, but that is not a standard practice (Aitchison *et al* 2005). In case of situations were distance is a vital concept in the analysis a *CLR* or *ILR* transformations described below should be applied before analysis.

**Centered log ratio (*CLR*) transformation:**

$CLR : S^D \rightarrow U^D$ , unlike the *ALR,* transformation can map a composition in the D-part Aitchison simplex isometrically to a D-1 dimensional Euclidean vector (equation 2.8).

$$z = clr(x) = \left[ \log\left\{ \frac{x_1}{g(x)} \right\} - \log\left\{ \frac{x_D}{g(x)} \right\} \right], \tag{2.8}$$

where $U^D = \{[u_1 ... u_D] : u_1 + ... u_D = 0\}$ (2.9)

The inverse of *CLR* is $CLR^{-1} : U^D \rightarrow S^D$ which takes the form of $x = C[\exp(z_1) ... \exp(z_D)]$ just like the *ALR* above, the *CLR* takes the composition to the Euclidean sample space hence opening up the possibility of using standard unconstrained statistical methods of analysis Aitchison, (2003). A prominent weakness of the *CLR* transformation is that, the orthogonal references in its subspace are not obtained in a straightforward manner Fernandez *et al* (2011) in Egozcue *et al* (2003).

**Isometric log ratio transformation:** due to the above shortcomings posed by the *ALR* and *CLR* transformations, Egozcue *et al* (2003) came up with isometric log ratio *ILR* transformation to counter those shortcomings. Like *ALR* and *CLR,* the *ILR* can transform the data from simplex to real space according to isometric log ratio transformation or its

inverse. The *ILR* transform like *CLR* maps a composition in the D-part Aitchison-simplex isometrically to a D-1 dimensional Euclidian vector (equation 2.10). Once the data is *ILR* transformed it can then be analyzed in this transformed state by all standard statistical methods of analysis. The only problem will be the difficulty involved during the interpretation of the results, since there is no one-to-one relation between the original parts and the transformed variables Egozcue *et al,* (2003).

$$y = ilt(x) = (y_1, ...y_{D-1}) \in R^{D-1}, \tag{2.10}$$

Where:-

$$y_i = \frac{1}{\sqrt{i(i+1)}} \ln\left(\frac{\prod_{j=1}^{i} x_j}{(x_i + 1)i}\right), \tag{2.11}$$

and its inverse is:-

$$x = ilt^{-1}(y) = (x_1, ..., X_D)R^D, \tag{2.12}$$

where:-

$$[x_1, x_2, ..., x_D] = C \exp[z_1, z_2, ..., z_D], z_j = \sum_{j-1}^{D-1} \psi_{ijyi,} C \tag{2.13}$$

Equation 2.13 stands for the closure operation Aitchison, (1986)

and

$$C[z_1, z_2, ..., z_D] = \left[\frac{z_1}{\sum_{j=1}^{D}, z_j}, \frac{z_2}{\sum_{J=1}^{D} z_j}, ..., \frac{z_D}{\sum_{j=1}^{D} z_j}\right], \tag{2.13}$$

Having looked at the principles, sample space and log ratio transformations, a basic description of the four- arm olfactometer that is used to generate the raw compositional data is given below.

## 2.5 Four-arm Olfactometer

A typical four-arm olfactometer is used for measuring responses of insects towards a potential host or host habitat. In a four-arm olfactometer, test insects are given an opportunity to choose between several different odours simultaneously. When a four-arm olfactometer is being operated, it creates clearly distinct contiguous odour fields that can be easily entered, left, and re-entered by walking insects Vet *et al* (1983). Figure 2.1 on the next page shows the four numbered arms of olfactometer chamber.

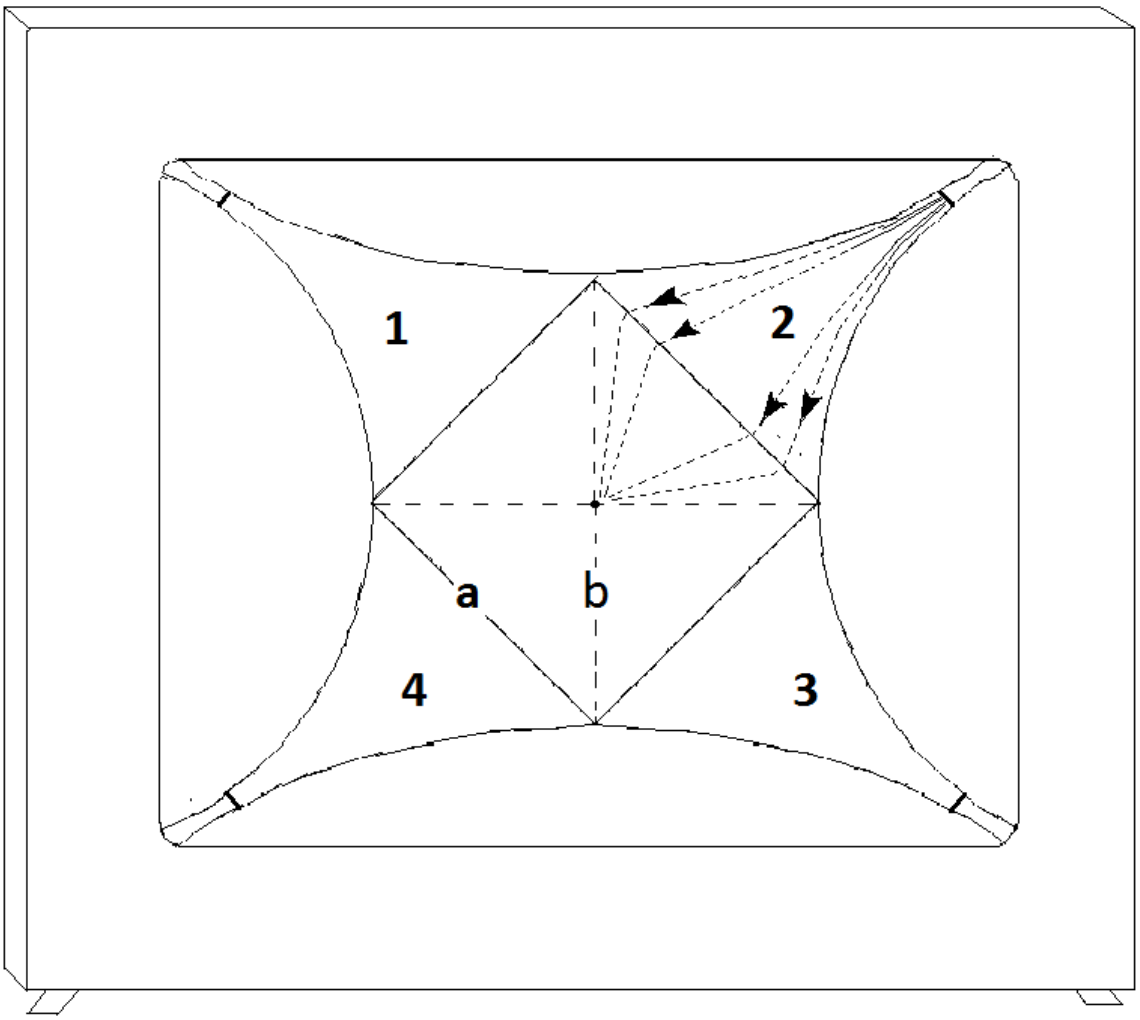


**Figure 2.1: A four-arm olfactometer chamber showing odour field and first choice line.**

Source: Vet *et al* (1983)

Note: **a** = first choice line, **b** = borderlines odour fields, → direction of flow. 1, 2, 3 and 4 are olfactometer arms.

The freedom of movement given to insects is important because it allows them to explore the odour fields presented, sampling freely between areas containing different 'attractants' separated by sharp boundaries. A four-arm olfactometer is fitted with an inlet system for solid or fluid odour sources, catching jar to facilitate the testing of individual insects, and a sensitive airflow control system which creates sharp boundaries between the odour fields Vet *et al* (1983); (Figure 2.2)

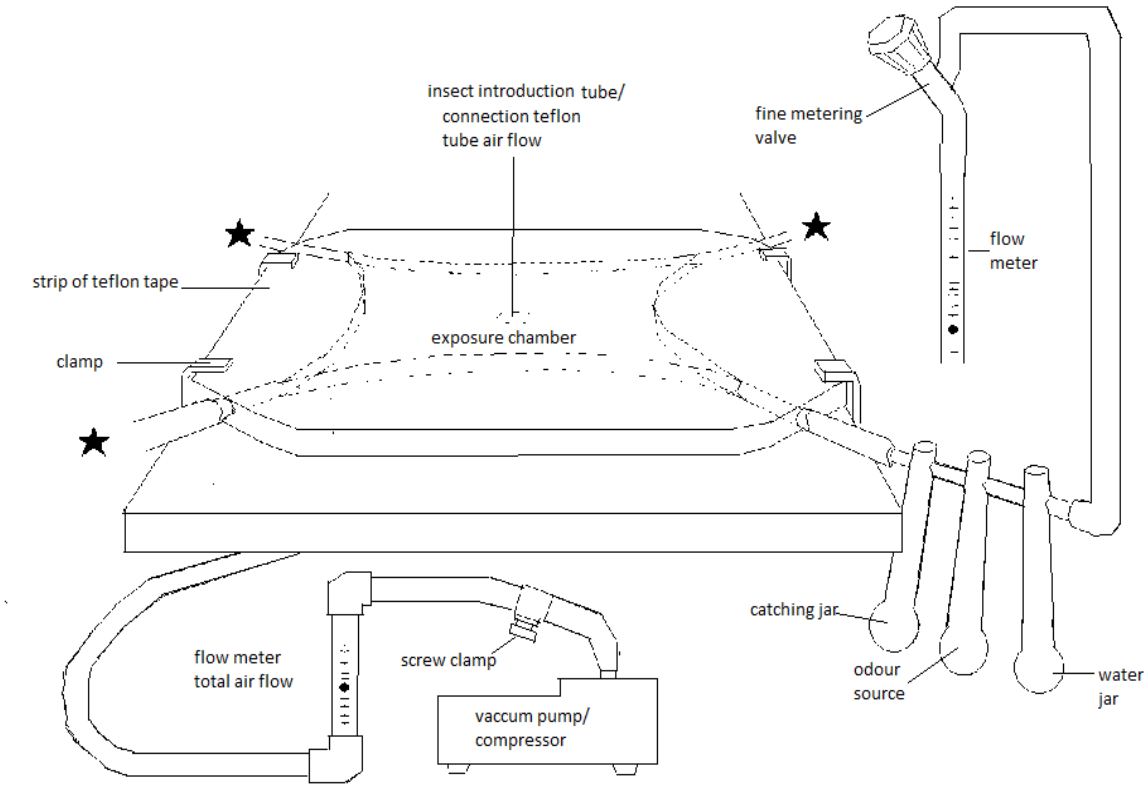


**Figure 2.2: A four-arm olfactometer**

Source: Vet *et al* (1983)

When using a four arm olfactometer bioassays, the insects are tested individually by setting up experimental odour fields up in the chamber at a specified overall flow rate of 180 mL/min (Ninkovic *et al* 2011), 250 mL/min (Suazo *et al* 2003), 300 mL/min (Vet *et al* 1983), 400 mL/min (Jonfia-Essien *et al* 2007) and 800 mL/min (Ukeh 2008). The insect to be tested is introduced through a disconnected 50 mm long Teflon tube leading up to the hole in the centre of the olfactometer chamber (Vet *et al* 1983). The experimenter then reconnects the extractor tube and re-starts the airflow. In this set up a vertical entry, tube will expose the test insect to odours in the olfactometer chamber. When the test insect reaches the floor of the chamber, it will walk towards one of the four odourized fields. Therefore, it will move at freedom over the floor or on the ceiling and either stay in the same odour field, or leave it, sample the others and select one of them. When the insect crosses one of the lines of the arbitrary first choice square (Fig 2.1) a record of first choice made is taken. The first choice normally correlates positively to the sector of the entry tube via which the insect first approached the chamber floor. Once the experimenter has finished recording the first choice, the parasitoid will be allowed to spend a given amount of time in the olfactometer, such as 1 minute Suazo (2003), 10 minutes (Ninkovic *et al*.,2011; Ukeh 2008; Vet *et al.,* 1983), 20 minutes Jonfia-Essien *et al* (2007). During the bioassay period, data were recorded and analyzed with a computer software package (OLFA 33100 Udine, a computer program OLFA (33100 Udine, Italy) Ukeh (2008).

The four-arm olfactometer is usually used to test, 10 insects (Ninkovic *et al*., 2011, Ukeh 2008), 40 insects Vet *et al* (1983) different insects in each odour situation. During

time of experimentation, the whole system is usually rotated at 90$^{o}$C after every ten to fifteen insects, and at that point, the chamber is cleaned out with 96% ethanol. In between each individual odour test, the whole process is dismantled, thoroughly washed with hot detergent, and swabbed with ethanol.

### 2.6 Transformations and analysis of Compositional Data in Chemical Ecology

Find below a summary of reviewed papers in chemical ecology and their methods of statistical analysis of compositional data:

Bruce *et al* (2009) in their study of parasitic wasps did not transform raw compositional data. The authors used a Y-tube olfactometer to test the response of three different egg parasitoid species; (*Scilionids telenomus* Busseolae Gahan, *Telenomus isis* Polarszek, and *Trichogrammatid trichogramma* Bournieri Pintureau) to calling and non-calling females of the noctuids Busseola fusca (Fuller), *sesamia calamistis* (Hampson), and S*esamia nonagrioides* (Lefelvre). The data was analyzed using Kruskal-Wallis test, Mann-Whitney U-test was used to compare mean onset of calling time between stem borers' species. They also used chi-square test to assess the first choice of each parasitoid. Another study conducted by Jonfia-Essien *et al.,* (2007) to measure the olfactometer responses of *Tribolium casteneum* (Herbst) to six major volatiles of cocoa beans, using a Peterson olfactometer. Data was analyzed using ANOVA, after performing a logarithmic transformation. In the study conducted by Said *et al.,* (2006) while attempting to evaluate the adoption of a four-arm olfactometer for large insects. The authors performed a logarithmic transformation before analyzing data using a one-

way ANOVA to compare mean times spent in each of the four fields, followed by New man-Keuls multiple comparison tests. Ninkovic *et al* (2010) tested the responses of adult *Coccinella septempunctata* (Linnaeus) to odours of barley genotypes, genotype mixtures, and barley genotypes using a two-arm olfactometer. They also used a four-arm olfactometer to test the preference of *C.septempunctata* (Linnaeus) for odours from single genotypes. In their analysis, they used Wilcoxon matched pairs tests to analyze raw data generated from two-arm olfactometer bioassays, and Friedman ANOVA rank test for analyzing data from four-arm olfactometer. Where a significant difference was found, multiple comparisons between treatments were performed using the Wilcoxon-Nemenyi-McDonald-Thompson test. Reed *et al* (1995) tested the response of mated parasitoid *Diaeretiella rapae* (M'Intosh) (Hymenoptera: Aphidiidae) females to odours from wheat, cabbage, and plant-host complexes. They used a four-arm olfactometer and data analysis was done using Kruskal-Wallis test; Wilcoxon matched pair's tests. Ukeh (2008) evaluated the essential oils extracted from *Aframomum melegueta* (K.Schum) seeds and *Zingiber officinale* (Ginga) rhizomes for their repellency against *Rhizopertha dominica* in a four-arm air flow olfactometer. The parameters assessed were time spent and number of entries or visits made by male and female adults into treatment and control arms of the olfactometer. No data transformation was done; a pair-wise t-test was used to test for significant differences between the treated arm and the mean of the control arms.

Apparently, there seems to be no evidence of adaption of log ratio methodology in chemical ecology as advocated in Aitchison (1986). The above studies suggest that, scientists do check their compositional data for normality and homogeneity of variances. Once the data is normal or variances are, homogeneous it would be analyzed using ANOVA, t-tests, or non-parametric tests for data not obeying the assumption of normality and homogeneity of variances. Although transformations are sometimes used, they are not the appropriate ones for changing the sample space from the Simplex to the Euclidean sample space. However, log ratio methodology is slowly gaining ground in other disciplines of biology as seen in Alison *et al* (2009; Korhoňova *et al* (2009) who after finding that, proportions of the different chemical compounds were not independent within a sample, performed a centered log ratio transformation according to Aitchison, (1986), Pawlosky-Glahn and Egozcue, (2006). After log-ratio transformation, the authers did principal components analysis on their data. Veverka *et al*., (2012), also applied log-ratio methodology to make standard statistical analysis by principle components possible.

# CHAPTER 3

# MATERIALS AND METHODS

## 3.1 Introduction

Chapter 3 gives the source of parasitic wasps *C. sesamiae* Cameron, herbivore induced plant volatiles (HIPVs), and oviposition induced plant volatiles (OIPVs). It also gives the description of four-arm olfactometer bioassay, data collection methods, and statistical methods employed in this study.

## 3.2 Source of parasitic wasps *Cotesia sesamiae* Cameron (Hymenoptera: Braconidae)

The *C.sesamiae*, parasitic wasps for *Chilo partellus* used for this study were reared in a bio-containment unit at International Centre of Insect Physiology and Ecology, Mbita field station. Newly emerged adults of *C.sesamiae* were transferred into plastic jars (1.0 litre capacity) fitted with brass mesh for ventilation. A cotton swab soaked in 50% honey was used for feeding adult parasitic wasps. The colony was maintained in a controlled environment chamber at 25$^{o}$C and 65% Relative Humidity (RH) and Day Lengh (LD) 12:12 h, 50000 lux). From this population mated 2-3-day-old females were selected using an aspirator for the bioassays and all test insects were naïve in that they had never encountered a host or plant as an adult.

### 3.3 Odour sources

The odours used for these bioassays were pre-processed in the Department of Chemical Ecology at *icipe* Thomas Odhiambo, Mbita field station. The processed odours sources were contained in jars designed for used in a four-arm olfactometer assay. They consisted of volatiles from accessions of Cuba 91maize leaves (HIPVS) and volatiles from from *C.partellus* eggs obtained from *icipe* Thomas Odhiambo, Mbita field station, plus leaves of Cuba 91maize cultivar complex (OIPVs).

### 3.4 Olfactometer bioassays

The olfactometer used in this bioassays was similar to the one designed by (Vet *et al.,* 1983). This type of olfactometer is designed for testing odour responses of small diptera and hymenoptera. Prior to olfactometer bioassay tests with the test odours, a system bias check was conducted by visualized odour fields using smoke generated by Ammonium hydroxide ($NH_4OH$) and Hydrochloric acid (HCL). During this time, the compressor was adjusted until a uniform field of smoke from Ammonium Hydroxide ($NH_4OH$) and Hydrochloric acid (HCL) was obtained. This was obtained at 40 Kilo Pascal (KPa); and airflow rate was at 300 ml/min at this pressure. The boundaries of each odour field were marked on the external surface of the olfactometer (Figure 2.1). Having attained the optimum pressure and airflow rate for uniform odour fields, another system bias check was conducted on the olfactometer system by testing the responses of 20 parasitoids to the humidified air in all four quadrants. The data collected was analyzed using one-way ANOVA; there was no significant difference between the four means.

After ruling out any system bias, the olfactometer bioassays were conducted using odours from oviposition induced plant volatiles (OIPVs); herbivore induced plant volatiles (HIPVs), and humidified air (Control). The olfactomer was illuminated by providing a 20-watt circular fluorescent lamp (3700 lux) 20 cm above it. An airflow meter (Cole-Parmer: PO4-N11202 G) maintained an airflow rate of 300 ml/min through each of the four quadrants. In order to maintain that flow rate a vacuum pressure pump placed on the floor, pulled room air through charcoal filters before entering the olfactometer chamber and then was vented into a fume hood.

Experiments were conducted from (0900hrs -1400hrs) at room temperature of $22.5^{0}$C for two consecutive days. Parasitoids to be tested were introduced singly into the olfactometer exposure chamber through the insect introduction tube port in the top of the olfactometer chamber (Figure 2.2). The experiment was replicated 20 times and during each bioassay 1 parasitic wasps was used. Once in the chamber, the parasitic wasps were given 12 minutes to make their choices between the four airfields. The amount of time spent in each odour field and the first airfield chosen was recorded for each insect by program Olfa a computerized program for collection and preliminary analysis of insect behavioural data. This summary data from program Olfa was saved in the computer and later used for further processing and statistical analysis.

### 3.5 Statistical analyses

The linear effect model was fitted to the centered log ratio transformed data of *C.sesamiae*. The linear model was

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (3.1)$$

where: $y_{ij}$ = Time spent in Olfactometer arm (log ratio transformed)

$\mu$ = Overall mean

$\varepsilon_{ij}$ = Treatment effect (error)

Prior to statistical analysis, the data from each of the treated arms plus the average of the two control arms generated by program Olfa above were converted to proportions of the total time spent. This was to account for dependence over time spent. In order to bring data to the real sample space amenable to standard statistical analysis techniques, log ratio transformations were conducted using Coda pack software (Cosmas-Cufi *et al*., 2011).

All the analyses were done using the 'R' data analysis software, version 2.12.3 (R Development Core Team, 2011). The numerical variables were checked for normality and equality of variances using formal tests. Graphical displays of response variate data were done using box plots and basic diagnostic plots. ANOVA model was used to evaluate data from centered log ratio transformed data, followed by Tukey multiple comparison tests. The data from additive and isometric log ratio transformations were analyzed using Kruskal-Wallis; Wilcoxon signed rank test. The logical framework (Figure 3.1) below, informed these analyses.

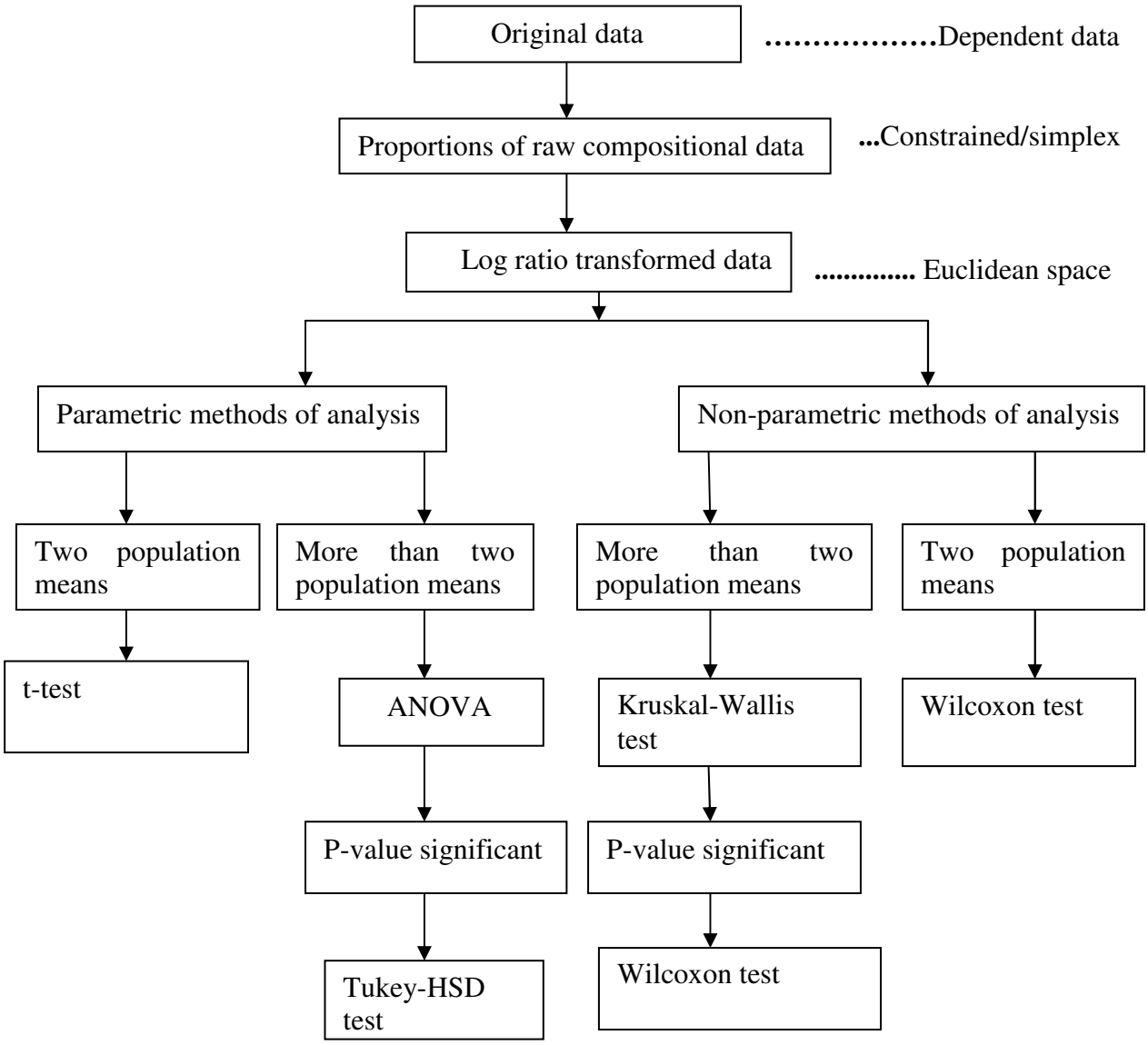


**Figure 3.1: Conceptual model of log ratio methodology for analysis of compositional data**

# CHAPTER 4

# RESULTS AND DISCUSSION

## 4.1 Introduction

In this chapter, the results of log ratio transformations of raw compositional data from four-arm olfactometer bioassay are tabulated; descriptive and inferential statistics of the centered, additive, and isometric log ratio transformed data from four arm olfactometer bioassays are also given.

## 4.2 Results

### 4.2.1 Log ratio transformation of original compositional data

Table 4.2.1, 4.2.2 and 4.2.3 Show centered, additive and isometric log ratio transformed datasets of oviposition induced plant volatiles, and herbivore induced plant volatiles and the control. The transformations were made from the original compositional data (Appendix C).

**Table 4.2.1: Original and Centered log-ratio transformed data on time spent for the three treatments in a four-arm olfactometer assay as produced by CoDapack software.**

| Orig (OIPVs) | Orig (HIPVs) | Orig (Control) | *Clr* (OIPVs) | *Clr* (HIPVs) | *Clr* (Control) |
|---|---|---|---|---|---|
| 0.37 | 0.14 | 0.17 | 0.699 | -0.273 | -0.079 |
| 0.41 | 0.16 | 0.14 | 0.833 | -0.108 | -0.242 |
| 0.34 | 0.24 | 0.18 | 0.725 | 0.377 | 0.089 |
| 0.51 | 0.18 | 0.12 | 0.912 | -0.13 | -0.535 |
| 0.36 | 0.28 | 0.12 | 0.634 | 0.382 | -0.465 |
| 0.44 | 0.26 | 0.15 | 0.670 | 0.144 | -0.407 |
| 0.41 | 0.25 | 0.12 | 0.839 | 0.345 | -0.389 |
| 0.27 | 0.24 | 0.11 | 0.283 | 0.166 | -0.615 |
| 0.38 | 0.19 | 0.18 | 0.867 | 0.174 | 0.120 |
| 0.20 | 0.30 | 0.19 | 0.085 | 0.49 | 0.034 |
| 0.19 | 0.26 | 0.21 | 0.033 | 0.347 | 0.133 |
| 0.37 | 0.11 | 0.15 | 0.696 | -0.517 | -0.207 |
| 0.2 | 0.26 | 0.17 | 0.002 | 0.264 | -0.161 |
| 0.17 | 0.23 | 0.20 | -0.131 | 0.172 | 0.032 |
| 0.31 | 0.20 | 0.16 | 0.411 | -0.027 | -0.250 |
| 0.35 | 0.2 | 0.16 | 0.649 | 0.089 | -0.134 |
| 0.27 | 0.25 | 0.19 | 0.445 | 0.368 | 0.093 |
| 0.26 | 0.22 | 0.17 | 0.286 | 0.118 | -0.139 |
| 0.23 | 0.19 | 0.17 | 0.146 | -0.045 | -0.156 |
| 0.31 | 0.14 | 0.20 | 0.526 | -0.269 | 0.087 |

Rows = Proportions of original data and centered log ratio transformed data.

*Clr* (OIPVs, HIPVs, and Control) = Centered log ratio transformed data from oviposition induced plant volatiles, herbivore induced plant volatiles, and control.

Orig (OIPVs, HIPVs, and Control) = Original data from oviposition induced plant volatiles, herbivore induced plant volatiles, and control.

**Table 4.2.2: Original and Additive log ratio transformed data on time spent for the three treatments in a four-arm olfactometer as produced by CoDapack software.**

| Orig (OIPVs) | Orig (HIPVs) | Orig (Control) | *Alr* (OIPVs) | *Alr* (HIPVs) | *Alr* (Control) |
|---|---|---|---|---|---|
| 0.37 | 0.14 | 0.17 | 0.778 | -0.194 | -0.569 |
| 0.41 | 0.16 | 0.14 | 1.075 | 0.134 | 0.047 |
| 0.34 | 0.24 | 0.18 | 0.636 | 0.288 | -0.469 |
| 0.51 | 0.18 | 0.12 | 1.447 | 0.405 | -1.218 |
| 0.36 | 0.28 | 0.12 | 1.099 | 0.847 | -1.955 |
| 0.44 | 0.26 | 0.15 | 1.076 | 0.550 | -1.299 |
| 0.41 | 0.25 | 0.12 | 1.229 | 0.734 | -1.811 |
| 0.27 | 0.24 | 0.11 | 0.898 | 0.780 | -1.959 |
| 0.38 | 0.19 | 0.18 | 0.747 | 0.054 | 1.203 |
| 0.20 | 0.38 | 0.19 | 0.051 | 0.457 | -0.877 |
| 0.19 | 0.26 | 0.21 | -0.100 | 0.214 | -0.017 |
| 0.37 | 0.11 | 0.15 | 0.903 | -0.310 | 0.047 |
| 0.20 | 0.26 | 0.17 | 0.163 | 0.425 | -0.916 |
| 0.17 | 0.23 | 0.20 | -0.163 | 0.140 | 0.358 |
| 0.31 | 0.20 | 0.16 | 0.661 | 0.223 | -0.333 |
| 0.35 | 0.2 | 0.16 | 0.783 | 0.223 | -0.332 |
| 0.27 | 0.25 | 0.19 | 0.351 | 0.274 | -0.368 |
| 0.26 | 0.22 | 0.17 | 0.425 | 0.258 | -0.417 |
| 0.23 | 0.19 | 0.17 | 0.302 | 0.111 | 0.424 |
| 0.31 | 0.14 | 0.20 | 0.438 | -0.357 | 0.368 |

Rows = Proportions of original and Additive log ratio transformed data.

*Alr* (OIPVs, HIPVs, and Control) = Centered log ratio transformed data from oviposition induced plant volatiles, herbivore induced plant volatiles, and control.

Orig (OIPVs, HIPVs, and Control) = Original data from oviposition induced plant volatiles, herbivore induced plant volatiles, and control.

**Table 4.2.3: Original and Isometric log ratio transformed data on time spent for the three treatments in a four-arm olfactometer as produced by CoDapack software.**

| Orig (OIPVs) | Orig (HIPVs) | Orig (Control) | *Ilr* (OIPVs) | *Ilr* (HIPVs) | *Ilr* (Control) |
|---|---|---|---|---|---|
| 0.37 | 0.14 | 0.17 | 0.238 | 0.687 | -0.988 |
| 0.41 | 0.16 | 0.14 | 0.493 | 0.665 | -1.102 |
| 0.34 | 0.24 | 0.18 | 0.377 | 0.246 | -0.222 |
| 0.51 | 0.18 | 0.12 | 0.756 | 0.736 | -1.283 |
| 0.36 | 0.28 | 0.12 | 0.794 | 0.178 | -0.278 |
| 0.44 | 0.26 | 0.15 | 0.664 | 0.372 | -0.642 |
| 0.41 | 0.25 | 0.12 | 0.801 | 0.350 | -0.757 |
| 0.27 | 0.24 | 0.11 | 0.685 | 0.083 | 0.197 |
| 0.38 | 0.19 | 0.18 | 0.327 | 0.490 | -0.708 |
| 0.20 | 0.30 | 0.19 | 0.207 | -0.287 | -0.804 |
| 0.19 | 0.26 | 0.21 | 0.046 | -0.222 | -0.855 |
| 0.37 | 0.11 | 0.15 | 0.242 | 0.858 | -1.233 |
| 0.20 | 0.26 | 0.17 | 0.240 | -0.186 | -0.683 |
| 0.17 | 0.23 | 0.20 | 0.009 | -0.214 | -0.325 |
| 0.31 | 0.20 | 0.16 | 0.361 | 0.310 | -0.467 |
| 0.35 | 0.20 | 0.16 | 0.411 | 0.396 | -0.640 |
| 0.27 | 0.25 | 0.19 | 0.256 | 0.055 | 0.884 |
| 0.26 | 0.22 | 0.17 | 0.279 | 0.118 | 0.257 |
| 0.23 | 0.19 | 0.17 | 0.169 | 0.135 | 0.163 |
| 0.31 | 0.14 | 0.20 | 0.033 | -0.730 | 0.562 |

Row= Proportion of raw and isometric log ratio transformed data.

*Ilr* (OIPVs, HIPVs, and Control) = Centered log ratio transformed data from oviposition induced plant volatiles, herbivore induced plant volatiles, and control.

Orig (OIPVs, HIPVs, and Control) = Original data from oviposition induced plant volatiles, herbivore induced plant volatiles, and control.

The rationale behind the above transformations was to convert the data from its simplex sample space (support space for compositional data) to Euclidean sample space (three-dimensional space). After log ratio transformations subsequent analysis were conducted by parametric or non-parametric methods depending on the results of tests for normality and homogeneity of variances.

### 4.2.2 Descriptive statistics

The centered, additive, and isometric log ratio transformations were explored graphically, with the aid of box plots (Figure 4.2.1, 4.2.2, and 4.2.3). The box plots for the centered log ratio transformed data seem to suggest normality (Figure 4.2.1). While the box plots for additive and isometric log ratio, transformed data show the contrary (Figure 4.2.2 & 4.2.3). All the box plots from the three-log ratio transformations seem to suggest that, the parasitoid *C.sesamiae* Cameron, spent much time in olfactometer arm with OIPVs, followed by the olfactometer arm with HIPVs and lastly spent least time in the control arm of the olfactometer.

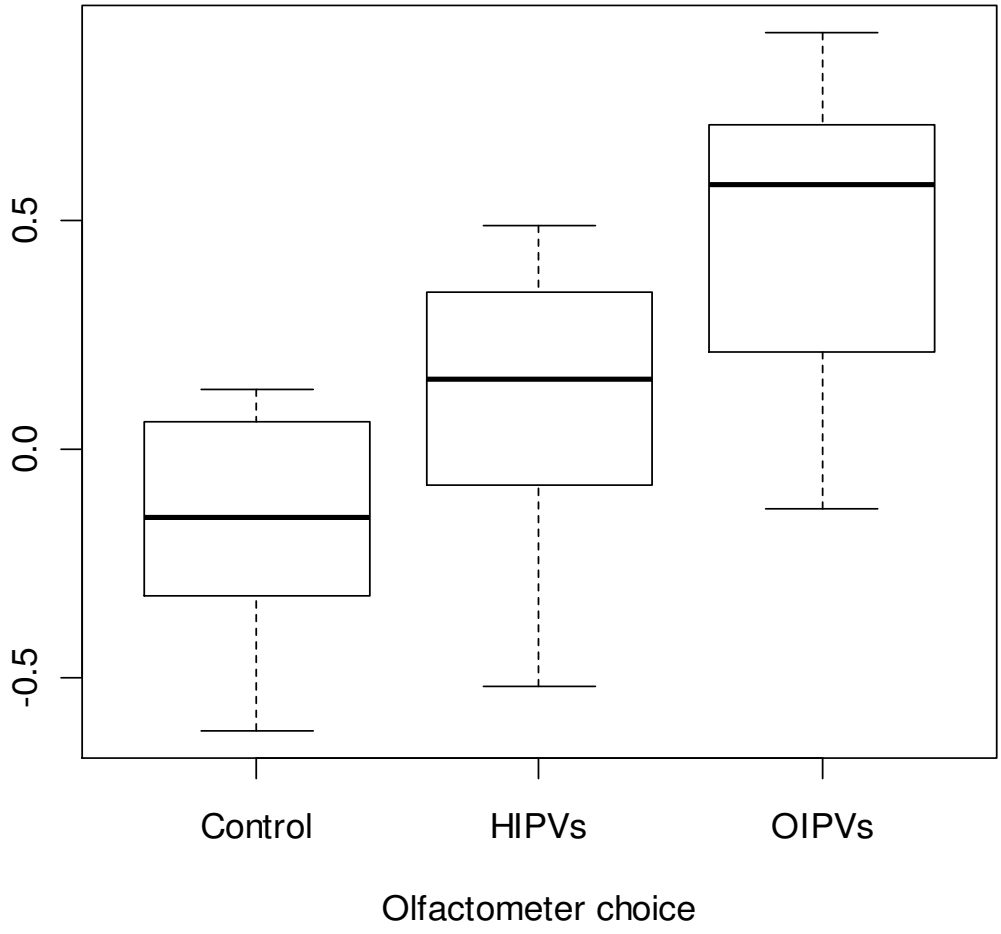


**Figure 4.2.1: Box plots for time spent by *Cotesia sesamiae* Cameron in the olfactometer bioassay, as produced by R software: Centered log-ratio transformed data**

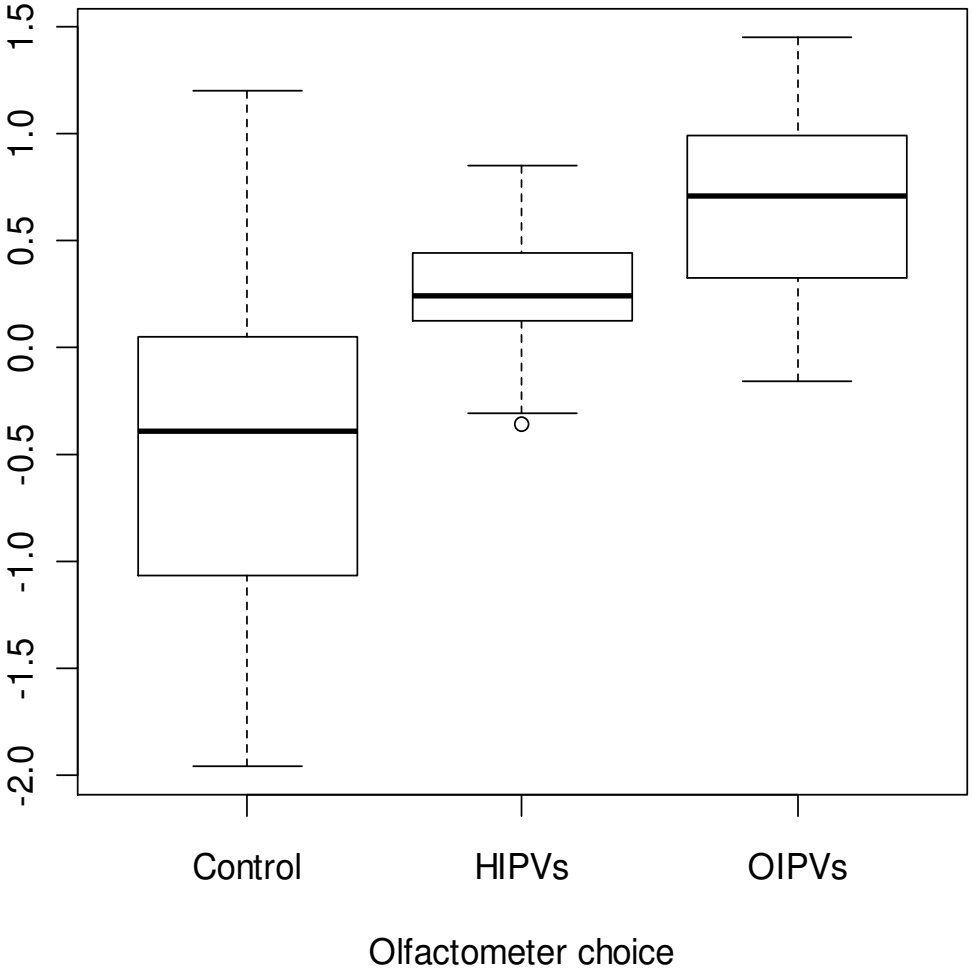


**Figure 4.2.2: Box plots for time spent by *Cotesia sesamiae* Cameron in the olfactometer bioassay, as produced by R software: Additive log-ratio transformed data**

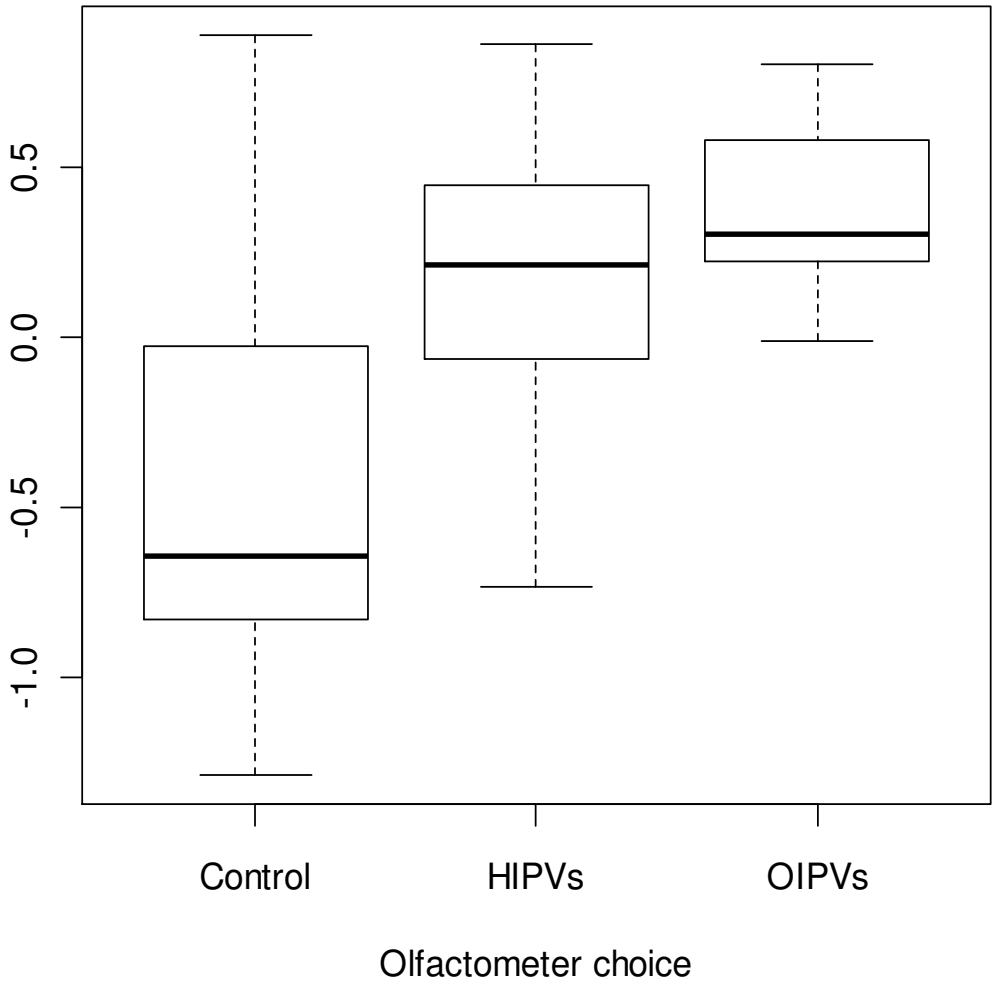


**Figure 4.2.3: Box plots for time spent by *Cotesia sesamiae* Cameron in the olfactometer bioassay, as produced by R software: Isometric log-ratio transformed data**

### 4.2.3: Inferential statistics for olfactometer bioassay data from *cotesia sesamiae*

The log ratio transformed datasets were checked for normality and homogeneity of variances. The results indicate that only the *CLR* transformation was able to normalize the data and stabilize the variances (Table 4.2.4). Based on these results subsequent analysis were conducted using one-way ANOVA for the *CLR* transformed data, Kruskal Wallis and Wilcoxon sign rank tests for *ALR* and *ILR* transformed data..

**Table 4.2.4: Results of tests for assumption of ANOVA as produced by R software**

| Test | Test-statistic | Statistic | P.values |
|---|---|---|---|
| ***Original data*** | | | |
| Normality | Shapiro-Wilk | 0.9158 | 0.0005*** |
| Homogeneity | Bartlett's | 16.5719 | 0.0003*** |
| ***CLR transformed data*** | | | |
| Normality | Shapiro-Wilk | 0.9778 | 0.3433 |
| Homogeneity | Bartlett's | 2.1965 | 0.3335 |
| ***ALR transformed data*** | | | |
| Normality | Shapiro-Wilk | 0.9297 | 0.0019*** |
| Homogeneity | Bartlett's | 17.649 | 0.0001*** |
| ***ILR transformed data*** | | | |
| Normality | Shapiro-Wilk | 0.9378 | 0.0043*** |
| Homogeneity | Bartlett's | 12.7565 | 0.0017*** |

Significance codes: ***' 0.001 '**' 0.01 '*' 0.05. Degree of freedom=2

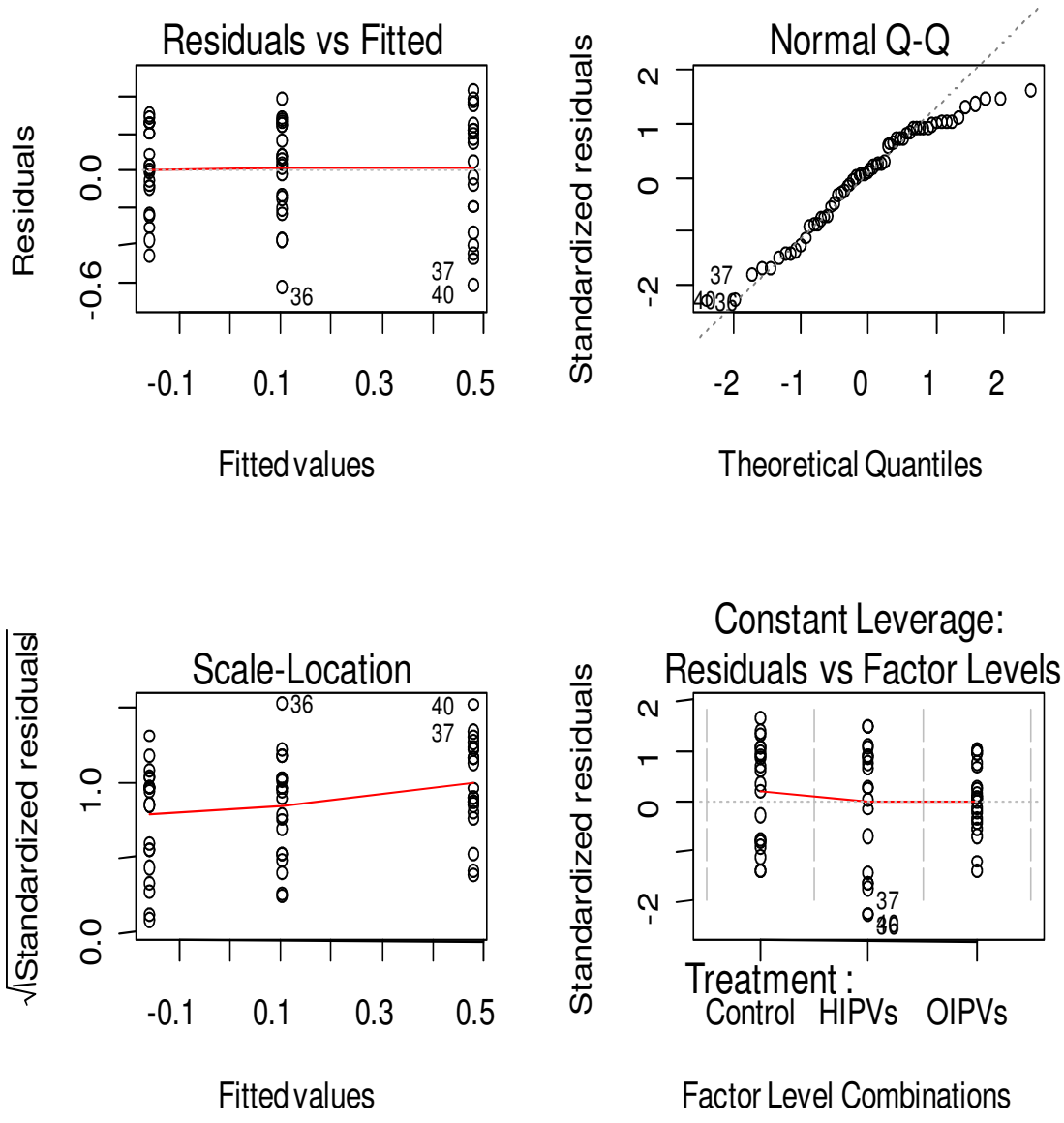


**Figure 4.2.4: Basic diagnostic plots for ANOVA model on CLR transformed data as produced by R software**

Based on the results of the normality and homogeneity of variances (Table 4.2.4) performed on the centered log ratio transformed data; and the basic diagnostic plots for ANOVA model (Figure 4.2.4). It was justifiable to use a one-way ANOVA for analysis of centered log ratio transformed data. Having obtained a significant result from the ANOVA, Tukey HSD tests were conducted on the centered log ratio transformed data at $P<0.05$. The results for Tukey pair wise comparisons were all significant (Table 4.2.5).

**Table 4.2.5: Multiple comparisons of means: Tukey contrasts**

| | Estimate | Std. Error | t value | Pr (>\|t |
|---|---|---|---|---|
| OIPVs - Control | 0.63994 | 0.08712 | 7.346 | <0.001 *** |
| HIPVs - Control | 0.26281 | 0.08712 | 3.017 | 0.0106 * |
| OIPVs - HIPVs | 0.37713 | 0.08712 | 4.329 | <0.001 *** |

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05

(Adjusted p values reported -- single-step method)

The additive and isometric log ratio transformed datasets were analyzed using Kruskal-Wallis test. The results were all significant (Table 4.6).

**Table 4.2.6: Results for Kruskal Wallis test for additive and isometric log ratio transformed data**

| Test | Statistic | P.values |
|---|---|---|
| ***ALR transformed data*** | | |
| Kruskal-Wallis | 24.5928 | 4.568e-06*** |
| ***ILR transformed data*** | | |
| Kruskal-Wallis | 19.3341 | 6.334e-05*** |

Significance codes: ***' 0.001 '**' 0.01 '*' 0.05. Degree of freedom=2

Owing to the significant test of Wilcoxon sign rank test, the post hoc tests were conducted using Wilcoxon sign rank test. The pair wise tests of Wilcoxon sign rank test were all significant at $P < 0.05$ (Table 4.2.7).

**Table 4.2.7: Results for Wilcoxon signed rank test for additive and isometric log ratio transformed data**

| Comparison | Statistic | P.values |
|---|---|---|
| ***ALR transformed data*** | | |
| OIPVs-HIPVs | 179 | 0.0042 |
| OIPVs-Control | 196 | 0.0002 |
| HIPVs-Control | 175 | 0.0095 |
| ***ILR transformed data*** | | |
| OIPVs-HIPVs | 163 | 0.0296 |
| OIPVs-Control | 198 | 0.0001 |
| HIPVs-Control | 175 | 0.0073 |

## 4.3 Discussion

The results for the transformations conducted suggest that, the centered, additive, and isometric log ratio transformations were all able to transform raw compositional data (Appendix C), to Euclidean sample space (Table 4.2.1, 4.2.2 and 4.2.3). Therefore, standard statistical methods of analysis designed to operate in the Euclidean sample space could be also be used for analysis of these data in accordance to (Aitchison 2005). The same methodology was used by Korhoňova *et al*., (2009) and Veverka et al., (2012) to analyse compositional data from coffee aroma and wine volatiles respectively.

From the box plots (4.2.1, 4.2.2, and 4.2.3), only the centered log-ratio transformed data appears to be normal (Figure 4.2.1). Whereas the box plots for additive and isometric log ratio, transformed data appear not to show normality (Figure 4.2.2 & 4.2.3). The box plots also seem to suggest that the parasitoid *C.sesamiae*, spent much time in olfactometer arm with oviposition induced plant volatiles (OIPVs), followed by the olfactometer arm with herbivore induced plant volatiles (HIPVs) and lastly spent least time in the control arm of the olfactometer.

The formal tests for ANOVA were consistent with those obtained by descriptive statistics; only the centered log-ratio transformed data showed normality. The formal test for homogeneity of variances showed that the variances for the centered log-ratio transformed data were homogenous. Therefore, this data was analyzed using

a One-way Analysis of Variance (ANOVA). The justification for use of ANOVA was informed by the results of the formal test for ANOVA model (Table 4.2.4); the centered log-ratio transformation a part from changing the sample space of the data was also able to normalise the data and stabilise the variances. This would be crucial in processing compositional data before subsequent analysis by parametric standard statistical methods. Centered log ratio transformation does allow the transformed data to be analyzed with methods that depend on distance such as ANOVA (Aitchison 2003). This is consistent with the works of (Korhoňova *et al* 2009 and Aebischer *et al* 1993) in analyzing compositional data.

While for the additive and isometric log ratio transformed data from oviposition induced plant volatiles, herbivore induced plant volatiles and that from the control were analyzed using Kruskal-Wallis test. The decision to use a nonparametric test was informed by the results of tests for normality and homogeneity of variances (Table 4.2.4). Veverka *et al*., (2012) has also noted that isometric log ratio transformation is an effective method of processing compositional data, before principal components analysis.

The ANOVA results showed that the data from oviposition induced plant volatiles, herbivore induced plant volatiles and the control were significant, with more time spent in the OIPVs and HIPVs than in the control arms. This meant that *Cotesia sesamiae* female parasitoid wasps were more attracted to Cuba 91 maize variety with and without *C.partellus* oviposition than the control.

The additive and isometric log ratio transformed data were analyzed using Kruskal-Wallis test due to non-adherence of the transformed data to ANOVA assumptions. The result showed a significant difference in result at 0.001 level of significance (Table 4.2.6). The Kruskal-Wallis test was followed by a *post hoc* test using Wilcoxon signed rank test. When additive log ratio transformed data was analyzed by Wilcoxon test, it was able to detect the significant difference between OIPVs and HIPVs at 0.001 level of confidence (Table 4.2.7). The same data transformed by isometric log ratio and tested by Wilcoxon test detected a significant difference between OIPVs and HIPVs at 0.05 level of confidence (Table 4.2.7). These results indicate that the additive log ratio transformation is better than isometric log ratio transformation in processing compositional data.

# CHAPTER 5

# CONCLUSION, AND RECOMMENDATIONS

## 5.1 Conclusion

The results for the transformations conducted show that, the centered, additive, and isometric log ratio transformations were all able to transform raw compositional data to Euclidean sample space.

Inferential statistic confirmed what was suggested by the descriptive statistics that only the centered log-ratio transformed data was normal and the parasitic wasps spent much time in olfactometer arm with OIPVs, followed by the olfactometer arm with HIPVs and lastly spent least time in the control arm of the olfactometer.

## 5.2. Recommendations

The study recommends Centered log-ratio transformation for processing compositional data from olfactometer bioassays in insect behavioural studies, before employing standard statistical methods of analysis and Log ratio transformations should be conducted using CoDapack statistical software.

There is need to do further study on the log-ratio methodology on olfactometer bioassay data from a different species of parasitic wasps.

REFERENCES

Addesso M. Karla (2010). Attraction of pepper weevil to volatiles from damaged pepper plants. *The Netherlands Entomological Society* **138:** 1-11.

Aebischer N.J, Robertson P.A and Kenward R.E (1993). Compositional Analysis of Habitat use from Animal Radio Tracking Data. *Ecology* **74**, 1313-1325.

Aitchison J. (1982). The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **44**(**2**), 139–177.

Aitchison J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.

Aitchison J. (2003). *A Concise Guide to Compositional Data Analysis.* In CDA workshop.Vol.29.No.3. Down loaded at http://www.mendeley.com/research/biofeature-concise-guide-cdna-microarray-biofeature-1/.

Aitchison J, Egozcue JJ. (2005). Compositional data analysis: where are we and where should we be heading? *Math Geol*.**37**:829–850.

Aitchison J. (2008). The single principle of compositional data analysis, continuing fallacies, Confusions, misunderstanding, and some suggested remedies, http://hdl.handle.net/10256/70.

Alison E., Faye M.W., Edmund D.B., Formica V.A. (2009). Differences in defensive volatiles of the forked fungus beetle, *Bolitotherus cornutus*, living on two species of fungus.*J.Chem.Ecol* **35**:1302-1308.

Billheimer, D., Guttorp, P., and Fagan W., (1997). Statistical analysis and interpretation of discrete compositional data: Technical report, NRCSE technical report 11: University of Washington,Seattle, Washington, 48 p.

Bruce T.J.A., Midega C.A.O., Birkett M.A., Pickett J.A. and *Khan Z.R.* (2010). Is quality more important than quantity? Insect behavioural responses to changes in a volatile blend after stemborer oviposition on an African grass. *Biology Letters* **6**: 314–317.

Bruce Y. A., Schulthess F., Mueke J, and Calatayud P.A. (2009). *BioControl* **54**: 763-772.

Chabi-olayi A., Schulthess.F. Poehling.H.M. and Borgemeister.C. (2001). Host location and host discrimination behavior of *Telenomus isis*, an egg parasitoid of the African cereal stem borer *Sesamia calamisti, J.Chem Ecol Vol*. **27**, No 4.

Chayes F (1960) On correlation between variables of constant sum: J. Geophys. Res., v. **65**, no. **12**, p. 4185–4193. Chen-Cheng (2010) Behavioural Ecology of Insect Cognition: Universals and Particulars, *Behavioural and Chemical Ecology* pp 173-193.Nova Science Pub Inc. Australia

Chayes, F. (1971). *Ratio Correlation*. University of Chicago Press.

Comas-Cufí M, Thió-Henestrosa S. (2011). CoDaPack 2.0: stand-alone, multi-platform compositional software. In: Egozcue JJ, Tolosana-Delgado R, Ortego MI, eds. CoDaWork'11: 4th International Workshop on Compositional Data Analysis. Sant Feliu de Guíxols.

Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, and Barcel´o-Vidal C. (2003). Isometric log ratio transformations for compositional data analysis. *Mathematical Geology* **35(3)**, 279–300.

Gohole L.S., Overholt.W.A. Khan.Z.R., Pickett.J.A. In addition, Vet.L.E. (2003). Effects of molasses, *Melinis minutiflora* volatiles on the foraging behavior of the cereal stem borer parasitoid, cotesia sesamiae, *J. Chem. Ecol, Vol*. **29**: No. 3.

Jan de kogel W., Kaschier.H. E and Visser H. J. (1999). Y-tube olfactometer to determine the attractiveness of plant volatiles to western flower thrips. *Practical, Experimental and Applied, New Amsterdam, Vol* **10**.

Janssen, A., Bruin, J., Jacobs, G., Schraag, R., Sabelis, M.W. 1997 Predators use volatiles to avoid prey patches with conspecifics. Journal of Animal Ecology, **66**, 223-232.

Jonfia-Essien W.A., Anderson P.G., Tucker.G. linforth R and West.G. (2007). *Pakistan Journal of Biological Sciences* **10,** 212-219

Ken Cheng (2010). Behavioural Ecology of Insects Cognition: Universals and Particulars, *Bebavioural and Chemical Ecology* PP 173-193.Australia, Nova Science Pub Inc.

**(20)**:3549-3556.

Korhoňova M., Hron K., Klimčikova D., Muller L.,Bednař P., Bartak P. (2009). Coffee aroma – statistical analysis of compositional data. *Talanta*, **80**: 710–715.

McIndoo N.E., (1926). An insect olfactometer. *Journal of Economic Entomology*, **19**: 545–571.

Ninkovic V., Abassi S. A., Elham A., Glinwood R, and Jan P. (2011). Effects of within-species plant genotype mixing on habitat preference of a polyphagous insect predator. *Oecologia* **166**:391-400.

Oluwafemi S., Bruce.T.J.A. Pickett.J.A. Ton.J. In addition, Birkett.M.A. (20011). Behavioral responses of the leafhopper, *Cicadulina storey* China, a major vector of maize streak virus, to volatile cues from intact and leafhopper-damaged maize, *J. Chem. Ecol* **37**:40-48.

Pawlowsky-Glahn V and Egozcue J. J. (2006). Compositional data and their analysis: An introduction. *Geological Society, London, Special Publication V*: **264**; p1-10.Down loaded at http://sp.lyellcollection.org on January 2012.

Pawlosky-Glahn. V, Buccianti A. (2011). *Compositional Data Analysis:*Theory and Applications.Willey.com

Pettersson, J. (1970). An aphid sex attractant. I. Biological studies. *Entomologica Scandinavica,* **1**: 63–73.

R Development Core Team. (2011). Software R. Available at http://cran.r-project.org (accessed January 2012).

Reed H.C., Tan S.H., Hapanen K., Killmon M. Reed D.K. and Elliott N.C. (1995) Olfactory responses of the parasitoid Diaeretiella rapae (Hymenoptera: Aphidiidae) to the odor of plants, aphids, and aphid plant complexes. *J. Chem. Ecol*. **21**: 407-418.

Said I. Aldana de la T. R., Morin J P and Rochat D. (2006). Adaptation of a four-arm olfactometer for behavioural bioassays of large beetles. *Chemoecology* **16**:9–16.

Sarmanov, O. V., and Vistelius, A. B., 1959, On the correlation of peercentage values: Dokl. *Akad. Nauk*. SSSR, v. **126**, p. 22–25.

Sakuma M. and Fukami H. (1985). The linear track olfactometer _an assay device for taxes of the german cockroach, Blattella germanica (L) (Dictyoptera, Blattellidae) toward their aggregation pheromone. *Applied Entomology and Zoology*, **20**: 387–402.

Shamshev I.V., Selytskaya.O.G., Chermenskaya.T.D., Burov.V.N. In addition, Roditakis.N. (2003). Behavioural responses of western flower thrips (*frankliniella occidentalis*

(pergande)): laboratory and field bioassays, *Archives of Phytopathology and Plant Protection*, Vol **2**: 111-118.

Shepherd W., Sullivan B.T., Goyer R.A. and Klepzig K.D. (2005) Electrophysiological and olfactometer responses of two histerid predators to three pine bark beetle pheromones, *J.Chem Ecol, Vol*. **31**, No. **5**.

Suazo A., Torto B., Teal P.E.A., and Tumlinson.J.H. (2003). Response of the small hive beetle (*Aethina tumida*) to honey bee (*Apis mellifera*) and beehive-produced volatiles, *Apidologie* **34**: 525-533.

Sullivan, B.T., E.M. Pettersson, K.C. Seltmann, and C.W. Berisford. 2000. Attraction of the bark beetle parasitoid *Roptrocerus xylophagorum* (Hymenoptera: Pteromalidae) to host-associated olfactory cues. Env. Entom. **29**: 1138-1151.

Tamiru A, *et al*. (2012).Oviposition induced volatile emission from African smallholder farmers' maize varieties.*J.Chem. Ecol.Vol*, **38**, Issue **3** pp 231-234.

Tamiru A., Bruce T. J. A, Woodcock C. M., Caufield J. C, Midega C. A., Ogol K. P., Mayon P., Birkett M. A., Pickett J. A and Khan Z. R. (2011). Maize landraces recruit egg and larval parasitoids in response to egg deposition by a herbivore. *Ecology Letters* **14**: 1075–1083.

Turlings.T.C. J.., Davison and Tamo C. (2004). A six-arm olfactometer permitting simultaneous observation of insect attraction and odour trapping. *Physiological Entomology* **29**: 45–55.

Torto B., Arbogast.R.T. Alborn.H., Suaro.A., Engelsdorp.D.V.,Boucias.D., Tumlinson.J.H. and Teal.P.E.A. (2007). Composition of volatiles from fermenting pollen dough and

attractiveness to the small hive beetle *Aethina tumida*, aparasite of the honeybee *Apis mellifera, Apidologie* **38**: 380-389.

Ukeh A. D. (2008). Bioactivities of Essential oils of *Aframomum melegueta* and *Zingiber officinale* (Zingiberaceae) against *Rhyzopertha dominica* (Fabricius).*Journal of Entomology* **5**(**3**):193-199.

Vet L.E.M., Van Lenteren J.C. Heymans M. and Meelis E. (1983). An airflow olfactometer for measuring olfaction of hymenopterous parasitoids and other small insects. *Physiological Entomology* **8**: 97–106.

Veverka L., Marcela Jelinkova M., Hron K., Balik J., Stavek J., and Bartak P. (2012)**.** Chemical markers in the aroma profiles of South Moravian red wine distillates. *Czech J. Food Sci. Vol.* **30**, No. **4**: 369–376.

# APPENDICES

## Appendix A: R. functions used for data analysis using one-way ANOVA

In this appendix an outline of the functions that were used to generate results in chapter 4 is given.

```
#MVolatileclr <- read.table("C:/Users/Anthony/Desktop/October/MVolatilesclr.csv",
  header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
#showData(MVolatileclr, placement='-20+200', font=getRcmdr('logFont'),
  maxwidth=80, maxheight=30)
#boxplot(Time.m.~Olfactometer.choice, ylab="Time.m.",
 xlab="Olfactometer.choice", data=MVolatileclr)
#AnovaModel.2 <- aov(Time.m. ~ Olfactometer.choice, data=MVolatileclr)
summary(AnovaModel.2)
numSummary(MVolatileclr$Time.m. , groups=MVolatileclr$Olfactometer.choice,
  statistics=c("mean", "sd"))
.Pairs <- glht(AnovaModel.2, linfct = mcp(Olfactometer.choice = "Tukey"))
summary(.Pairs)
 # pairwise tests confint(.Pairs)
 # confidence intervals cld(.Pairs)
 # compact letter display old.oma <- par(oma=c(0,5,0,0))
plot(confint(.Pairs)) par(old.oma) remove(.Pairs) oldpar <- par(oma=c(0,0,3,0),
mfrow=c(2,2))
plot(AnovaModel.2) par(oldpar)
plotMeans(MVolatileclr$Time.m., MVolatileclr$Olfactometer.choice,
  error.bars="se")
```

**Appendix B: R. functions used for data analysis using Kruskal Wallis and Wilcoxon sign rank test**

```
#Mvolatilesalr <- read.table("C:/Users/Anthony/Desktop/Sept/MVolatilesalr.csv",
header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
#library(relimp, pos=4)
#showData(Mvolatilesalr, placement='-20+200', font=getRcmdr('logFont'),
 maxwidth=80, maxheight=30)
#tapply(Mvolatilesalr$Time.spent, Mvolatilesalr$Treatment, median,
 na.rm=TRUE)
#kruskal.test(Time.spent ~ Treatment, data=Mvolatilesalr)
#data1a <- read.table("C:/Users/Anthony/Desktop/Nov-2012/data1a.csv",
header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
# library(relimp, pos=4)
# showData(data1a, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
   maxheight=30)
# wilcox.test(data1a$OIPVs, data1a$HIPVs, alternative='two.sided', exact=TRUE,
  Paired=TRUE)
#wilcox.test(data1a $OIPVs, data1a $Control, alternative='two.sided', exact=TRUE,
   paired=TRUE)
# wilcox.test(data1a $HIPVs, data1a $Control, alternative='two.sided', exact=TRUE,
  paired=TRUE)
#Ilrdata <- read.table("C:/Users/Anthony/Desktop/Nov-2012/Ilrdara.csv",
header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
```

```
# library(relimp, pos=4)
#showData(Ilrdata, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80, maxheight=30)
# wilcox.test(Ilrdata$OIPVs, Ilrdata$HIPVs, alternative='two.sided', exact=TRUE, paired=TRUE)
# wilcox.test(Ilrdata$OIPVs, Ilrdata$Control, alternative='two.sided', exact=TRUE, paired=TRUE)
# wilcox.test(Ilrdata$HIPVs, Ilrdata$Control, alternative='two.sided', exact=TRUE, paired=TRUE)
```

**Appendix C: Raw compositional data from olfactometer bioassay**

| OIPVs | HIPVs | Control |
|---|---|---|
| 0.37 | 0.14 | 0.17 |
| 0.41 | 0.16 | 0.14 |
| 0.34 | 0.24 | 0.18 |
| 0.51 | 0.18 | 0.12 |
| 0.36 | 0.28 | 0.12 |
| 0.44 | 0.26 | 0.15 |
| 0.41 | 0.25 | 0.12 |
| 0.27 | 0.24 | 0.11 |
| 0.38 | 0.19 | 0.18 |
| 0.02 | 0.03 | 0.19 |
| 0.19 | 0.26 | 0.21 |
| 0.37 | 0.11 | 0.15 |
| 0.02 | 0.26 | 0.17 |
| 0.17 | 0.23 | 0.02 |
| 0.31 | 0.02 | 0.16 |
| 0.35 | 0.02 | 0.16 |
| 0.27 | 0.25 | 0.19 |
| 0.26 | 0.22 | 0.17 |
| 0.23 | 0.19 | 0.17 |
| 0.31 | 0.14 | 0.17 |

**Appendix D: Analysis of variance results of four- arm olfactometer responses of female *Cotesia sesemiae* to different odour sources.**

| | Df | Sum Sq | Mean Sq | F value | Pr (>F) |
|---|---|---|---|---|---|
| Olfactometer.choice | 2 | 4.1388 | 2.0694 | 27.266 | 4.914e-09 *** |
| Residuals | 57 | 4.3260 | 0.0759 | | |
| Total | 59 | 8.4648 | | | |

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Appendix E: Multiple comparisons of means: Tukey contrasts**

| | Estimate | Std. Error | t value | Pr (>\|t |
|---|---|---|---|---|
| OIPVs - Control | 0.63994 | 0.08712 | 7.346 | <0.001 *** |
| HIPVs - Control | 0.26281 | 0.08712 | 3.017 | 0.0106 * |
| OIPVs - HIPVs | 0.37713 | 0.08712 | 4.329 | <0.001 *** |