ANALYSIS OF LIGHT RESPONSE DATA

by

John N. Joshua Nasila

B.Sc.(Statistics), Univ. of Madras (India), 1981

PROJECT SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Mathematics and Statistics

©John N. Joshua Nasila
SIMON FRASER UNIVERSITY

July, 1992

Approval

Name: John N. Joshua Nasila

Degree: Master of Science (Statistics)

Title of project: Analysis of Light Response Data

Examining Committee:

Chairman: *Dr. A.H. Lachlan*

Dr. T. Swartz
Senior Supervisor

Dr. D. M. Eaves

Dr. R. Lockhart

Dr. C. Dean
External Examiner
Department of Mathematics and Statistics
Simon Fraser University

Date Approved: July 23, 1992

Abstract

In this project I analyse a data set brought to the Statistical Consulting Service. The data set arises from a biological experiment conducted to investigate the response of nematodes to light treatments. The data consists of 3 response variables measured over 30 treatments (seconds) in an experiment conducted on various nematodes. The analysis reveals no such light response and further suggests ways of improving the experimental methodology.

Acknowledgements

I would like to take this chance to express my appreciation to Dr. Tim Swartz for his guidance towards the preparation of this project and for his assistance towards the successful completion of my program. I must mention that I have gained a tremendous amount of knowledge and useful skills during my preparation of this project.

I would also like to extend my heart-felt gratitude to all my professors for the valuable advice and comments that they have given me during my time here as a student. I wish them the best of luck in their endeavours and promise to utilize the knowledge that they have given me for gainful needs. At this point, I must mention that high praise and appreciation go to the spouses of my professors who have treated me and my other fellow students to dinners, birthday parties and Christmas parties on several occassions. Words are not enough to express my gratitude for this kindness. All it has meant to me is that our professors' successes and good work have been contributed to by their spouses. I wish them a good living and high achievements in life.

Warm thanks go to my wife and our children for the patience and endurance they have exercised for the time I have been away. I recognize that it has been a very difficult task for my wife to have taken care of the children as a single parent. She deserves praise and comment.

I must mention that the Statistical Consulting Service was very helpful to me and I am very grateful for the technical and analytical advice I received from Mr. Francois Bellavance. Further, I must thank the Department of Biology for making available the data for the analysis and for their unreserved cooperation whenever I needed information. In this regard, I thank Dr. A. H. Burr and his graduate student Ms. Deming Liu.

Warm thanks to my brothers, sisters, relatives and friends for their courage and moral support.

I wish to thank my parents Mr. Joshua Nasila Wopicho and Mrs. Rebeccah Lubisia Nasila for their support, care, best wishes and long time parental sacrifices.

Lastly, I must thank my sponsor (The Canadian International Development Agency) without whom my dream of getting further training as a research scientist would not have been achieved by now. To my sponsors, I say "increase the budget for training of people from other countries". Besides, I am very grateful to the office of the Kenya High Commission for the continuous services that they have rendered to me in particular and all Kenyans in general.

Dedication

To my wife Mrs. Loice Nakhumicha Nasila and my children

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

It is part of the policy of the Department of Mathematics and Statistics that Statistics graduate students attend/participate in sessions directed by the Statistical Consulting Service. The Statistical Consulting Service is a small statistical unit whose objectives cover a wide range of areas such as:

- resource unit for statistical expertise
- practical statistical education and
- an exposure for graduate students to real life statistical situations.

As a graduate student, I have had the golden opportunity of coming across situations pertaining to statistical data analyses. During the fall semester of 1991, under the directorship of Monsieur François Bellavance, a statistics problem was brought to the Statistical Consulting Service from the Department of Biology by a biology graduate student, a Ms. Deming Liu. Deming and her senior supervisor Professor A.H. Burr are interested in understanding the photo-receptive behaviour of Caenorhabditis Elegans, a worm of great importance in the field of neural systems research. These worms (nematodes) are widely used in genetic

and developmental studies.

Previous studies have enabled biologists to completely understand the full anatomy and the developmental lineage of all of the 302 neurons found in Caenorhabditis Elegans (here after referred to as C. Elegans). Other researchers have shown that this nematode shows a response to chemical, thermal and mechanical stimuli. It has also been hypothesized that a behavioural response is elicited by an increase in light. And thus the primary objective of the experiment is to understand the photo-movement characteristics of the nematode. In the experiment the worm is subjected to a cycle of different intensities of light and darkness. The response variable which is described in Chapter 2 measures nematode activity during each second of a 30 second sequence of treatments involving the presence and absence of light.

In this project

- Chapter 2 gives a description of the experiment followed by comments
- Chapter 3 provides an analysis of the data and
- Chapter 4 presents the conclusions.

Three statistical packages are used for the analysis of the data. The statistical packages are:

- SAS
- Minitab and
- the S_Plus language.

Chapter 2

The Experiment

2.1 Description of the Experiment

As mentioned in the introduction, the data originated from the Department of Biology through Ms. Deming Liu, a biology graduate student whose senior supervisor is Professor A.H. Burr. "Understanding the photo-receptive behaviour of C. Elegans" is a subject Dr. Burr has been pursuing for a number of years and from which a number of in-depth publications have been produced. For example, see Burr(1985). Other response studies have been made by other researchers and their results have been made known. For example, see Hedgecock and Russell(1975).

One may reasonably wonder as to why C. Elegans has been selected instead of other species. The reason is that C. Elegans is an animal with a simple neural system with only 302 nerves and all of them have been fully understood by scientists. Mammals have about 10^{11} nerves while insects have about 10^6 nerves and in both of these cases, studying all of the nerves is too complex. As a result scientists have chosen to study C. Elegans and infer the results to the larger animals.

The experiment begins by transferring about 100 nematodes from a culture plate containing a food source to a behaviour plate. Each nematode is carefully washed during the process to ensure that the presence of food is not confounded in the experiment. Figure 2.1 shows the behaviour plate with the circular viewing area which is monitored by a sensitive computer tracking system. The circular viewing area contains approximately 20-25 nematodes on average. The computer tracking system has been introduced as the nematodes are too numerous and small (only 1.00 mm in length) to monitor by eye. The behaviour plate is then subjected to three treatments where each treatment period is referred to as a "phase". Thus a "cycle" consists of three 10 second phases where the plate is exposed to:

- light for the first 10 seconds
- darkness for the next 10 seconds and
- light for the last 10 seconds.

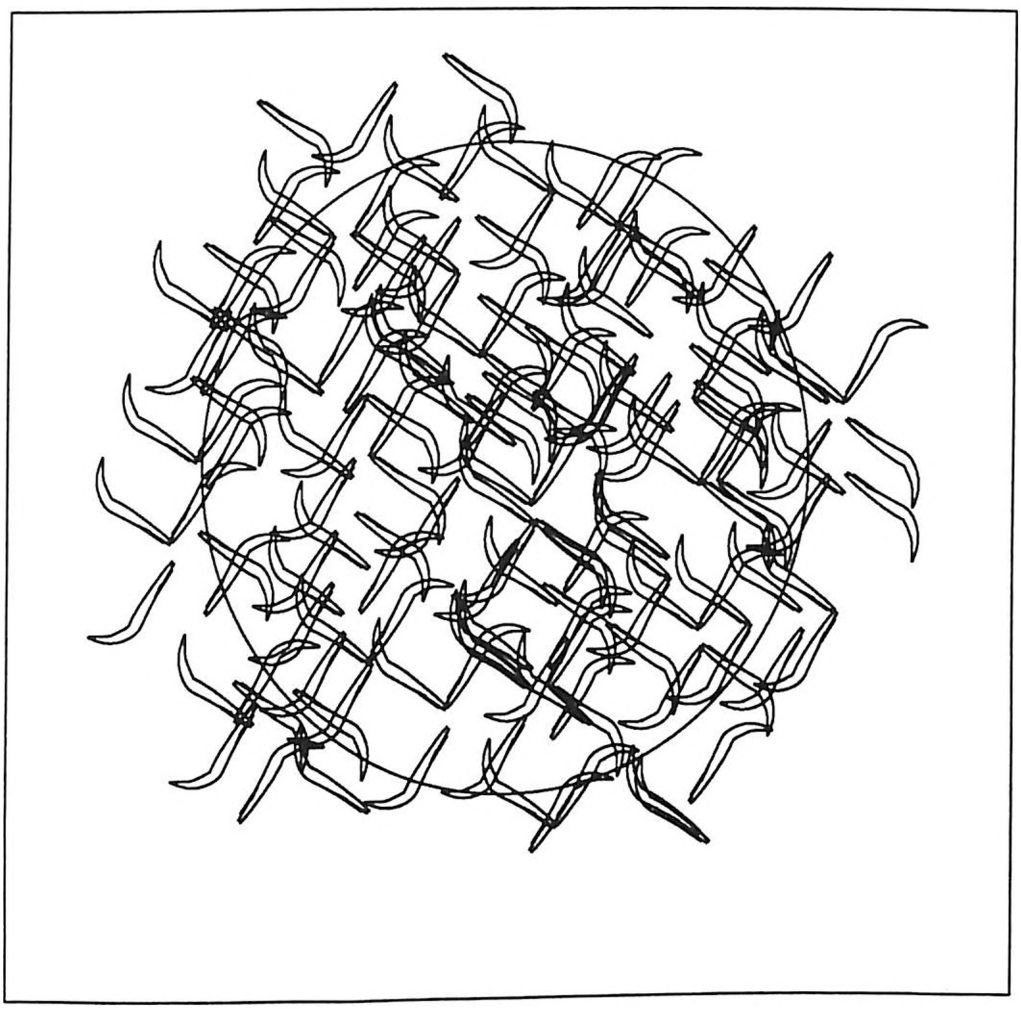

Figure 2.1: A picture of nematodes on the behaviour plate

Data is collected for each nematode in the viewing area during each second of the 30 second experiment. Three types of data are collected and we refer to these response variables as:

- speed data
- stop data and
- reversal data

where:

[**speed**] The speed data is a continuous variate which measures the speed of the nematode during each second of the 30 second experiment.

[**stop**] The stop data is a binary response taking the value 1 if the nematode is still during a particular second and the value 0 if the nematode is moving.

[**reversal**] The reversal data is a binary response taking the value 1 during a particular second if the nematode reverses its previous direction of travel. Otherwise the value 0 is recorded.

For example, consider the following data corresponding to the first 10 seconds of the experiment for a particular nematode.

Speed	0.0	0.0	0.0	1.0	1.5	0.0	0.8	1.9	2.1	0.0
Stop	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
Reversal	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0

Here the nematode is motionless during the first three seconds, begins travel on the 4th second, continues in the same direction on the 5th second and then stops on the 6th second. On the 7th second the nematode resumes travel but in the opposite direction and then

reverses direction back to the original direction on the 8th second. The nematode continues in the same direction on the 9th second before stopping on the 10th second.

This monitoring continues for as long as the nematode is in the field of view. One cycle is completed after 30 seconds of the experiment. If the worm remains in the region of view, data is again collected for the second cycle of 30 seconds, third cycle of 30 seconds and so on. This applies to all the worms in the region of view and the data is collected simultaneously. Data is collected on worms which come back to the region of view after being out of the region of view for a period of time. Data is only collected for full cycles. That is to say if a nematode enters/leaves the region of view sometime during a 30 second cycle then no data will be collected for that cycle. The experiment normally runs for a period of five minutes and the computer signals when this time is over.

2.2 Complications with the Experiment

A number of difficulties are encountered while carrying out the light response experiment. Though the nematodes are washed before transfer from the culture plate to the behaviour plate, some dirt is bound to remain and may confuse the computer tracking system. In these cases dirt may be interpreted as a motionless worm. For nematodes which remain motionless for more than 20 seconds, no data is collected for that particular cycle. This data collection rule was introduced by the experimenters to deal with problems of dirt and dead worms. I am told that approximately 1% to 6% of living worms are motionless during a particular 20 second period. However the occurrence of dead worms is very rare. When a motionless worm becomes active, data is collected as usual.

A second complication is that Ms. Liu and Professor Burr have repeated the experiment many times. On each run of the experiment they alter the experimental conditions slightly hoping to discover a situation where there exists a light response. Examples of subtle changes

to the experimental conditions include:

1) varying the salt level of agar on the behaviour plate

2) changing the light intensity and

3) having different numbers of worms (in the region of view) for different experiments.

We have only been given data for one of the many experiments conducted. The fact that many experiments are conducted has an impact on the statistical analysis. We discuss this in the conclusion where recommendations are offered to the experimenter.

A third complication of the experiment is that the computer tracking program currently collects data on cycles but can not determine which nematode corresponds to a particular cycle. This lack of identification has a potential impact on the assumption of cycle independence and is discussed in Section 3.3.

Finally, a close look at the speed data shows that speed can only assume certain values, namely: 0.00, 1.00, 1.41, 2.00, 2.24, 2.83, 3.00, 3.16 or 5.00. These values are obtained as the square root of the sum of two squared integers. Such a calculation is the result of applying the Pythagorean theorem to obtain the length of the hypotenuse using an integer grid. Thus despite the apparent continuous appearance of the speed data, it is actually discrete in nature. Nevertheless, we treat it as continuous data in the analysis.

2.3 The Data

Having gone through the above information, you are now in a position to understand the data structure and its components. The experiment is monitored by a computer which utilizes three programmes to collect the data on the response variables of interest. The response variables of interest as mentioned earlier are:

- the speed data

- the stop data and
- the reversal data.

In general, the computer keeps track of the movement of the worm by noting the change in coordinates from one time point to the next. If there is no change in coordinates it means that the worm is not moving.

The following are the first eight lines of each data set that I received from Deming by electronic mail. One line of data sent by electronic mail corresponds to 1 cycle for a single worm. Therefore each of the abridged data sets contains $8(30) = 240$ entries. Note that the entire data set consists of 30 trivariate measurements for each of 175 cycles.

SPEED DATA

```
2.00  1.00  1.00  1.00  1.41  3.00  2.24  1.00  0.00  0.00
1.00  0.00  0.00  0.00  2.24  3.16  1.41  2.00  1.00  0.00
1.00  0.00  0.00  0.00  1.00  1.00  1.41  1.00  0.00  0.00

3.16  2.24  1.41  1.00  1.00  1.00  0.00  1.00  1.00  1.00
1.41  0.00  1.00  0.00  0.00  0.00  1.00  1.00  1.00  1.00
2.24  1.00  2.24  1.41  0.00  1.00  1.00  1.00  0.00  0.00

1.00  1.00  1.00  1.00  1.00  1.00  1.00  0.00  0.00  0.00
0.00  1.00  1.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00

2.00  2.24  1.00  1.41  3.16  0.00  2.24  1.00  0.00  1.00
1.00  1.00  1.41  0.00  0.00  0.00  0.00  0.00  1.00  1.00
```

```
 0.00  1.00  1.00  0.00  0.00  0.00  0.00  0.00  0.00  2.24

 1.00  1.00  1.00  1.00  1.00  0.00  1.00  1.00  1.00  1.00
 1.00  0.00  3.00  1.00  1.00  1.00  1.00  0.00  2.24  1.41
 1.41  1.00  2.00  1.41  1.00  0.00  2.00  1.00  2.83  5.00

 1.00  1.00  0.00  0.00  1.41  1.41  1.00  1.00  1.00  0.00
 1.00  2.24  1.41  1.41  2.00  1.41  0.00  1.00  1.41  1.00
 1.00  1.00  1.00  2.00  1.00  1.00  1.41  1.00  1.00  1.41

 1.00  1.00  2.00  0.00  1.00  0.00  0.00  1.00  1.00  1.00
 0.00  1.00  1.00  0.00  0.00  1.00  0.00  1.41  1.00  0.00
 1.00  1.00  0.00  1.00  1.00  0.00  0.00  0.00  1.00  1.00

 1.00  1.00  0.00  0.00  0.00  0.00  1.00  1.41  1.00  1.41
 0.00  1.00  1.41  1.00  1.00  1.41  2.24  1.00  2.24  1.00
 1.00  0.00  1.00  2.00  0.00  1.00  0.00  0.00  1.00  0.00

STOP DATA


 0 0 0 0 0 0 0 0 1 1 0 1 1 1 0 0 0 0 0 1 0 1 1 1 0 0 0 0 1 1
 0 0 0 0 0 0 1 0 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0 1 0 0 0 1 1
 0 0 0 0 0 0 0 0 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 0 0 0 0 0 1 0 0 1 0 0 0 0 1 1 1 1 1 0 0 1 0 0 1 1 1 1 1 1 0
 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
```

0 0 0 1 0 1 1 0 0 0 1 0 0 1 1 0 1 0 0 1 0 0 1 0 0 1 1 1 0 0

0 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 1 1 0 1

REVERSAL DATA

0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Chapter 3

Data Analysis

3.1 Exploratory Data Analysis

When presented with the data set I found it a good idea to "play" with the data set for a period of time. By producing simple plots and computing descriptive statistics one can often uncover features of a data set before more formal and time-consuming statistical methods are attempted. Exploratory data analyses also sometimes suggest further questions to be asked of the experimenter.

We begin by producing some simple plots keeping in mind the experimental inquiry "Do nematodes respond to light?". Figures 3.1 and 3.2 consider the speed component of the data set. The speed data is summarized by summing the speeds of the nematode over all cycles for each second. When plotting the total speed over cycles, the idea is that a pattern may present itself if in fact there is a response due to light. For example, we might expect that the total speed might be greater/less during the treatment seconds 11-20 than during the control seconds 1-10 and 21-30. We also kept our eyes open for "delayed" patterns where the treatment effect is not seen immediately and may not last for the full 10 seconds. For example, if there is a light response, one might observe that total speed is greater/less

during the seconds 14-18 than during the seconds 1-13 and 19-30. Under close scrutiny, Figure 3.1 provides some evidence of decreasing speed over time. This could be due to a light response during the 3^{rd} treatment or possibly it may be the case that worms tire over time. The smoothed plot in Figure 3.1 is obtained by applying the smooth function lowess in S-plus. Figures 3.3 and 3.5 are similar to Figure 3.1 in that they show smoothed plots of totals for the reversal data and the stop data respectively. In both of these plots there seems to be a slight upward trend in the 3^{rd} phase.

However it can be misleading to simply look at the smoothed plots in figures 3.1, 3.3, and 3.5 to discover trends. We must look for trends in the context of the variability of the data. Therefore we have reproduced the plots of Figures 3.1, 3.3, and 3.5 in Figures 3.2, 3.4, and 3.6 respectively and have included two standard error limits. We recall from normal theory that 2 standard error limits should include approximately 95% of the data. From these plots we see that there is no strong evidence of a light effect. The data appear spurious.

It came to our attention that the previous plots may not be entirely adequate. One problem concerns a potential lack of independence between cycles. This issue is explored in more detail in Section 3.3.

It is also possible that the previous plots may not reveal a treatment pattern when only a few of the nematodes experience an effect due to light. For example, suppose that only 5 of the nematodes experience an effect due to light. This is valid information to the experimenter but is the type of information that may get "lost" in the previous plots. When data is presented as in Figures 3.1 - 3.6, the data values for the 5 nematodes are added to the data values corresponding to the remaining cycles in the 175 cycle experiment. The overall sums may possibly "damp out" the effect seen in the case of the 5 nematodes. To deal with this problem, we have tried to identify any cycle which appears to demonstrate a treatment effect. This has been done using a cluster analysis to distinguish the 175 cycles

where each cycle has 30 measurements. The cluster analysis for the speed data is presented in Figures 3.7 and 3.8. Here, to keep the dendrograms from becoming too messy, the cluster plots shown are based on the first 30 and 60 cycles of the experiment. The dendrogram based on the entire data set of 175 cycles exhibits a similar pattern. Cluster analyses were not carried out for the reversal data and stop data due to its binary form. As can be seen from Figures 3.7 and 3.8, there are no cycles which clearly distinguish themselves. This is further evidence that no nematodes experience an effect due to light. In addition to the cluster analysis, a visual inspection of the data in its raw form did not reveal any groups of worms that seemed to indicate a light response pattern.

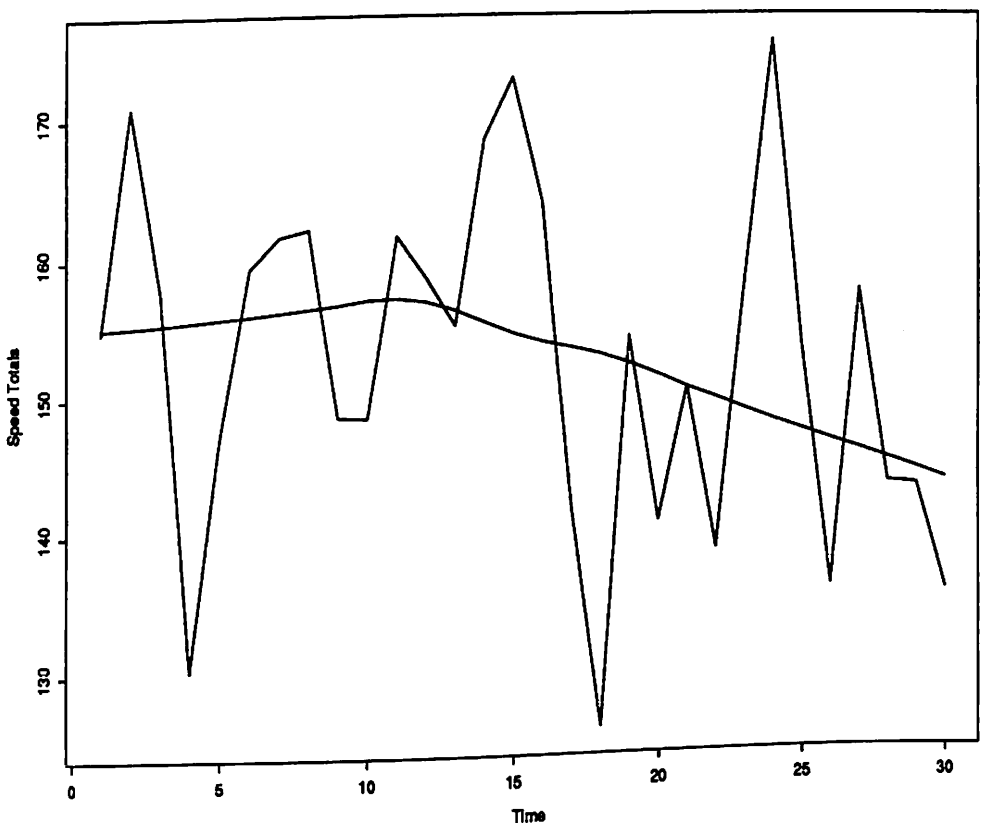

Figure 3.1: A smoothed plot of speed totals for each second

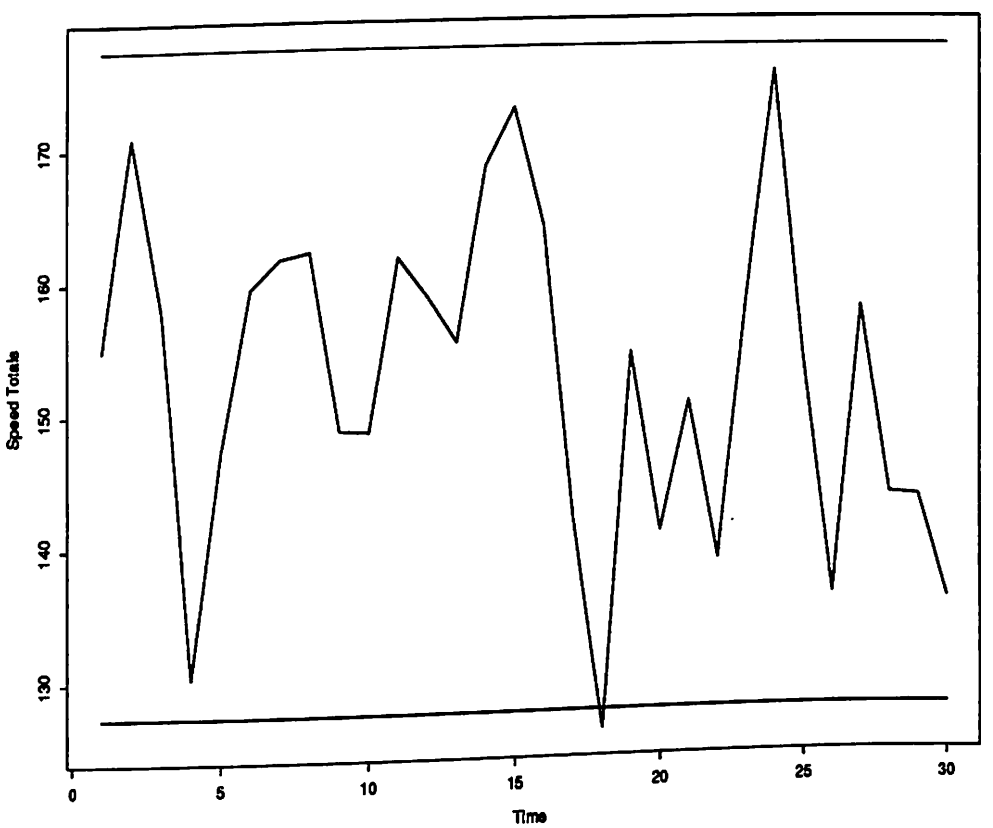

Figure 3.2: A plot of speed totals and the 2 s.e. limits for each second

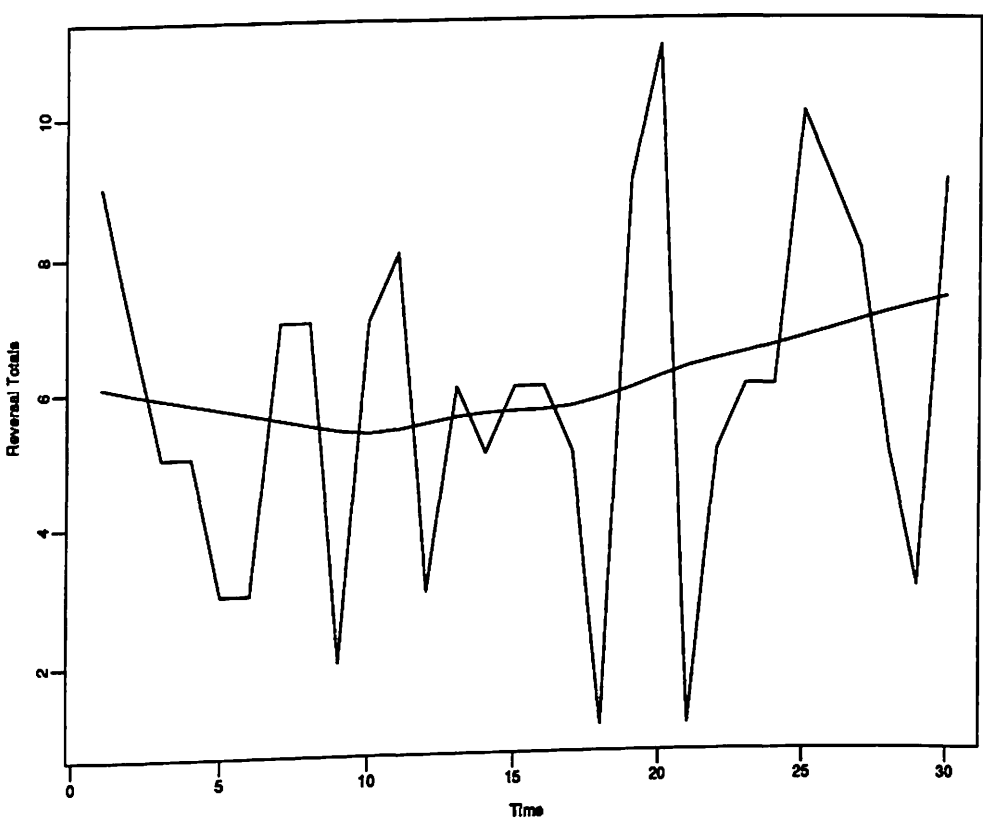

Figure 3.3: A smoothed plot of reversal totals for each second

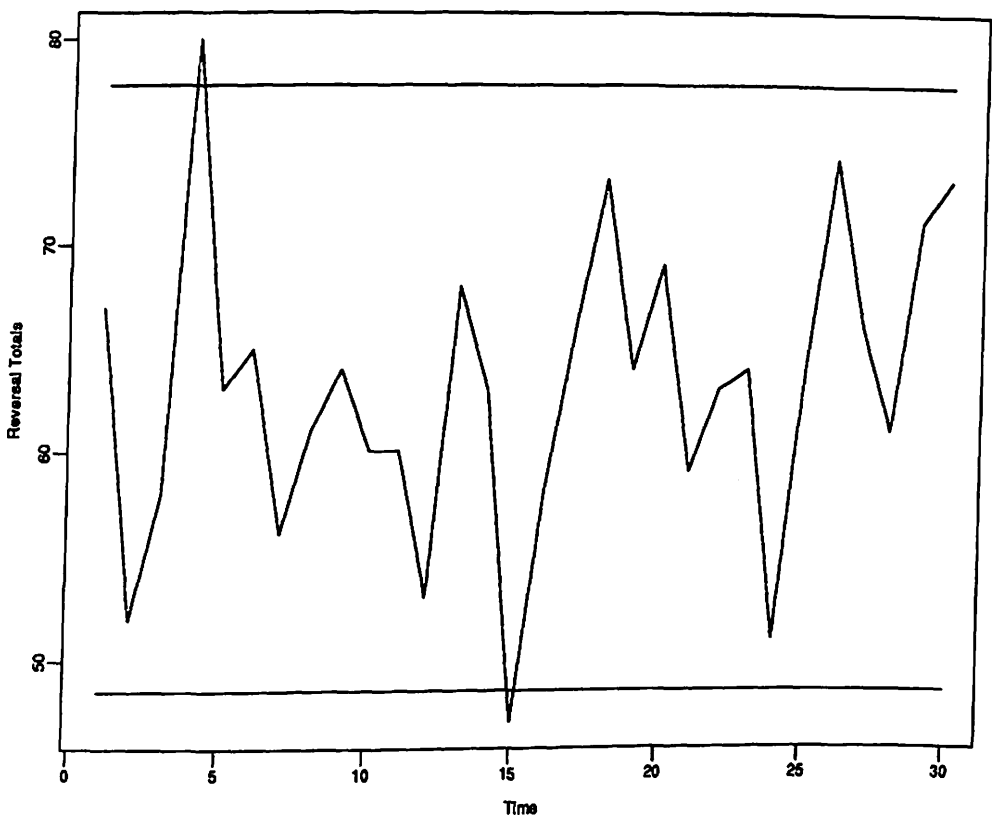

Figure 3.4: A plot of reversal totals and the 2 s.e. limits for each second

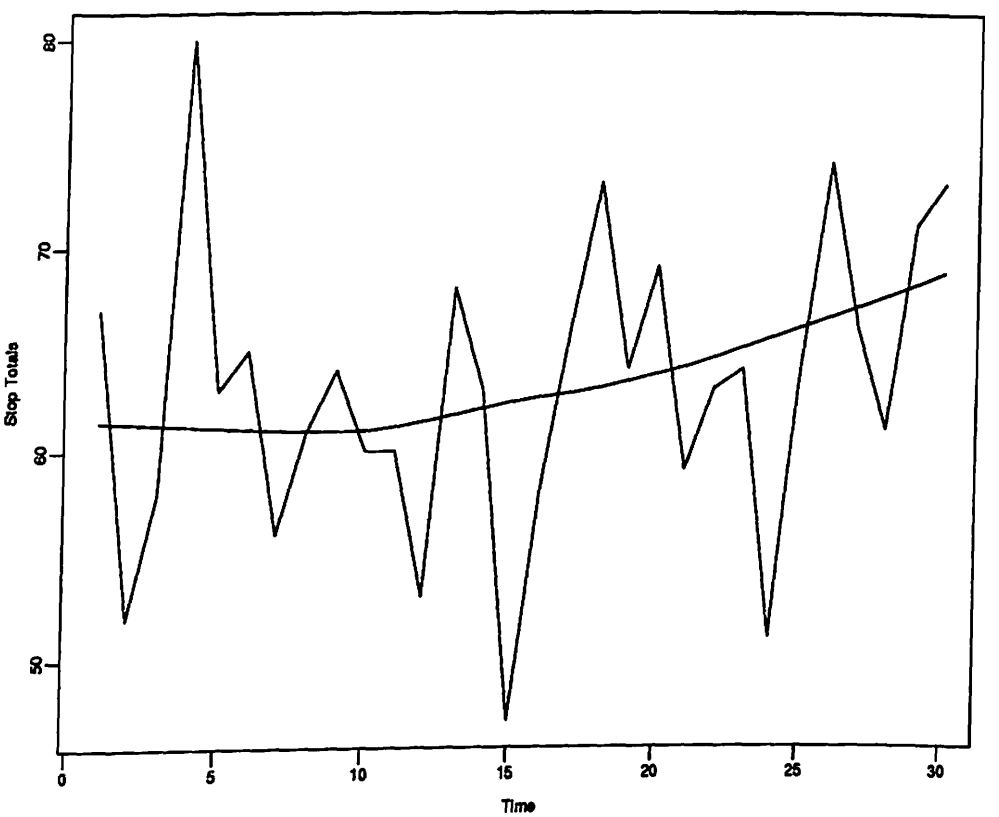

Figure 3.5: A smoothed plot of stop totals for each second

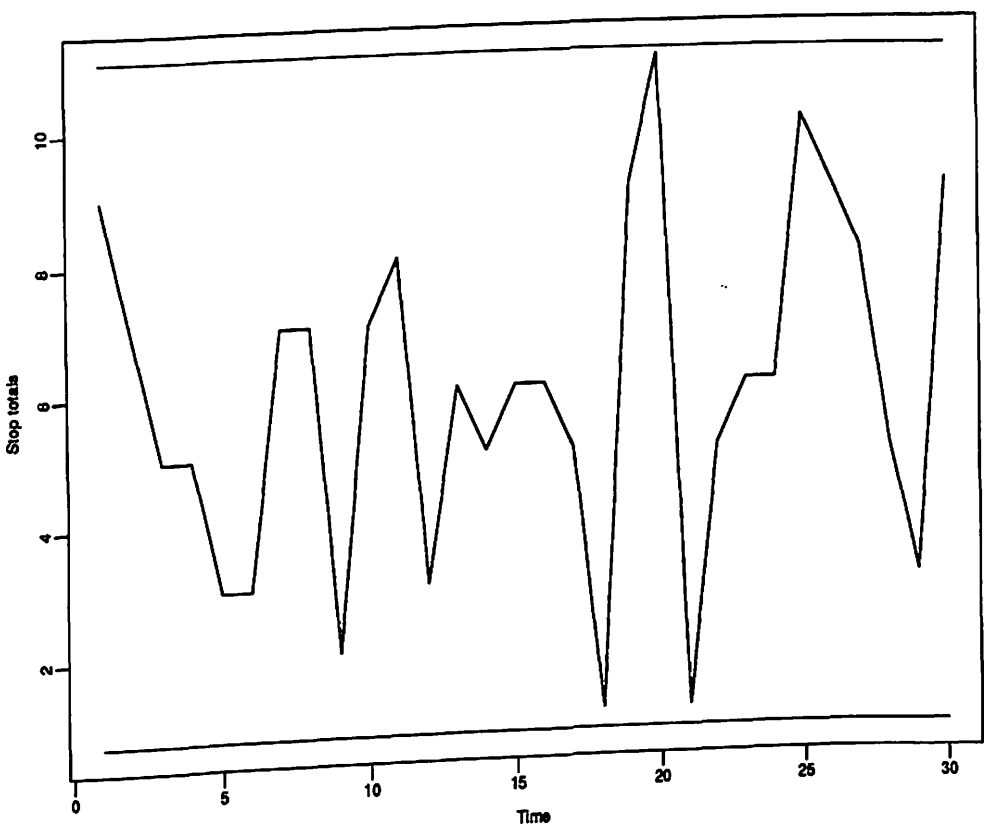

Figure 3.6: A plot of stop totals and the 2 s.e. limits for each second

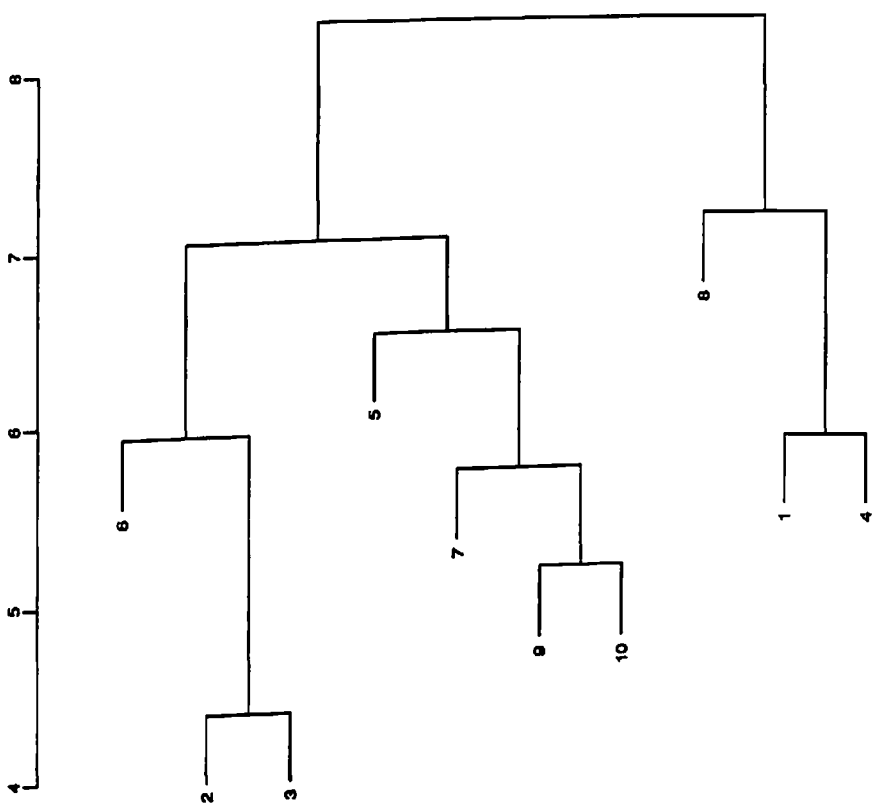

Figure 3.7: A cluster plot of speed based on the first 30 cycles

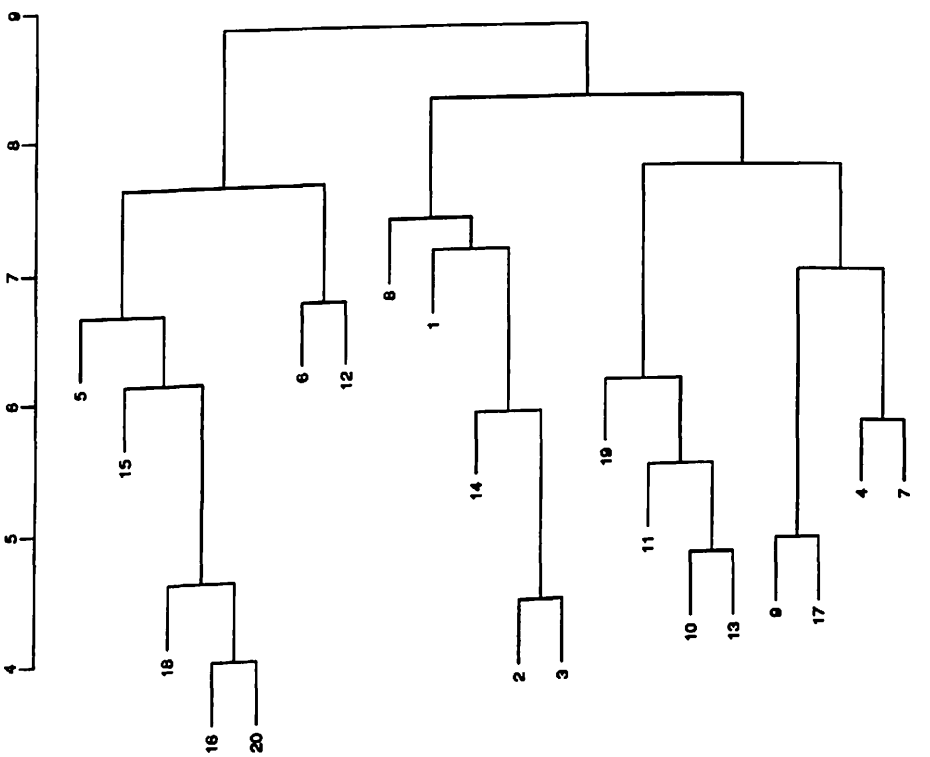

Figure 3.8: A cluster plot of speed based on the first 60 cycles

3.2 Formal Statistical Analysis

In the previous section, descriptive plots of the data suggest that there is no light effect on the nematodes. In this section we support those findings through formal statistical tests. It is always a good idea to carry out formal statistical tests as they are not subjective in the sense that a p-value is reported and they are sometimes able to detect significances which are undetectable with the eye.

We begin by considering the stop and reversal data as they each have the same binary form. We will assume that each cycle is independent so that X_{ij} is Bernoulli(p_j), $i = 1, \ldots, 175$ and $j = 1, \ldots, 30$ where X_{ij} is the observation corresponding to the j^{th} second in the i^{th} cycle. The cycle independence assumption is of course highly questionable, and will be discussed in more detail in Section 3.3. This model seems reasonable as we might expect a different probability of success for each treatment (second). It then follows that $X_{.j} = \sum_{i=1}^{175} X_{ij}$ is Binomial($175, p_j$) and our test for an effect due to light is a test of H_0:$p1 = p2 = \ldots = p30$ versus H_1: not all p_i's are the same. The data can therefore be summarized conveniently in the following 2 by 30 contingency tables.

second	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
success	9	7	5	5	3	3	7	7	2	7	8	3	6	5	6
failure	166	168	170	170	172	172	168	168	173	168	167	172	169	170	169
second	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
success	6	5	1	9	11	1	5	6	6	10	9	8	5	3	9
failure	169	170	174	166	164	174	170	169	169	165	166	167	170	172	166

Table 3.1: A 2 by 30 contingency table for reversals

second	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
success	67	52	58	80	63	65	56	61	64	60	60	53	68	63	47
failure	108	123	117	95	112	110	119	114	111	115	115	122	107	112	128
second	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
success	58	66	73	64	69	59	63	64	51	64	74	66	61	71	73
failure	117	109	102	111	106	116	112	111	124	111	101	109	114	104	102

Table 3.2: A 2 by 30 contingency table for stops

The total number of (successes,failures)=(T_1, T_2) for the reversal and stop data are (177,5073) and (1893,3357) respectively. The standard contingency table test then involves calculating the expected cell entries which for the reversal data are

$e_{1j} = 175(\frac{T_1}{T_1+T_2}) = 5.90$ and $e_{2j} = 175(\frac{T_2}{T_1+T_2}) = 169.10, j = 1, \ldots, 30$

For the stop data, the expected cell entries are

$e_{1j} = 175(\frac{T_1}{T_1+T_2}) = 63.10$ and $e_{2j} = 175(\frac{T_2}{T_1+T_2}) = 111.90, j = 1, \ldots, 30$

The chi-square test statistic is then

$\sum_{j=1}^{30} \frac{(x_{.j}-e_{1j})^2}{e_{1j}} + \sum_{j=1}^{30} \frac{((175-x_{.j})-e_{2j})^2}{e_{2j}} = 34.50$ for the reversal data and

$\sum_{j=1}^{30} \frac{(x_{.j}-e_{1j})^2}{e_{1j}} + \sum_{j=1}^{30} \frac{((175-x_{.j})-e_{2j})^2}{e_{2j}} = 38.13$ for the stop data.

Comparing these against a chi-square distribution on 29 degrees of fredom gives p-values of 0.22 and 0.12 respectively. Thus again there is no evidence of an effect due to light.

Now it is perhaps worthwhile to collapse the data to more clearly delineate between the first 10 seconds, the middle 10 seconds and the last 10 seconds of a cycle. We do this as there may be a small effect due to light which is not detectable over the individual seconds but when the seconds are collapsed, the small effect adds up to a significant effect. This leads to the following 2 by 3 contingency tables.

period	sec(1-10)	sec(11-20)	sec(21-30)	Total
success	62	48	67	177
failure	1688	1702	1683	5073
Total	1750	1750	1750	5250

Table 3.3: A 2 by 3 contingency table for reversals

period	sec(1-10)	sec(11-20)	sec(21-30)	Total
success	626	621	646	1893
failure	1124	1129	1104	3357
Total	1750	1750	1750	5250

Table 3.4: A 2 by 3 contingency table for stops

With the collapsed contingency tables, we observe chi-square statistics of 3.40 and 0.87 respectively for the collapsed reversal and stop data. Comparing these against a chi-square statistic with 2 degrees of freedom gives corresponding p-values of 0.18 and 0.65. Thus, again, there does not seem to be any evidence of an effect due to light.

With the speed data, we have a different scenario since the underlying data is no longer binary but instead continuous. We recognize the data as that of a repeated measures design where the subjects are the cycles and the repeated measurements are taken over the thirty seconds. This simple design is analysed in the same way as a 1-way anova with 30 treatments and 1-blocking variable taking 175 levels. The corresponding anova is given below.

Source	DF	Type III SS	Mean SS	F Value	P value
Cycle	174	473.61	2.72	4.76	0.00
Treatment	29	26.21	0.90	1.58	0.03
Error	5046	2891.33	0.57		

Table 3.5: An analysis of variance table for speed

Here the p-value is found to be 0.03 which is barely significant at the 5% level of significance. However it is not clear whether this demonstrates a significant effect due to light. Recall that Figures 3.1 and Figure 3.2 indicate a pattern of decreasing speed but only during the third phase. We might expect a significant effect due to light to result in similar speeds during the 1^{st} and 3^{rd} phases.

Besides, as was done for the reversal and stop data, an analysis performed on the collapsed data is given in Table 3.6. The anova on the collapsed data shows a p-value of 0.57 which accepts the null hypothesis of equality of phase means.

Thus, though the initial analyses on speed indicate significance, a less specific hypothesis is clearly rejected. The three phase means with corresponding standard errors in parentheses are 0.881(0.030), 0.881(0.029) and 0.851(0.029) for seconds $1-10$, $11-20$ and $21-30$ respectively. The following is the anova table for the collapsed data.

Source	Df	Type III SS	Mean Square	F Value	P value
Cycle	174	35.7188	0.2053	2.3040	0.0000
Treatment	2	0.1014	0.0507	0.5700	0.5663
Error	348	30.9974	0.0890		

Table 3.6: An analysis of variance table for collapsed speed data

3.3 A Test for Cycle Independence

An assumption that has been made so far in the analysis is the assumption of cycle independence. This is an assumption that should be tested. It is highly plausible that cycles are not independent as a given nematode can give rise to several cycles. It is possible, for example, that a nematode may get tired over time. In that case we would expect earlier cycles to exhibit more activity than in subsequent cycles. The tracking program currently used for data collection by the experimenter does not include facilities for identifying the worm corresponding to a given cycle. The identification is, however, an important step as it could help us in determining whether cycle independence exists. The experimenter has been advised of the importance of the cycle independence assumption and has agreed to rewrite the computer tracking program in such a way as to identify the worms corresponding to given cycles. Assuming that such an identification is available, statistical nonparametric procedures such as Kendall's tau or the well known Spearman's rho could be utilized for this kind of analysis in the following way:

For the i^{th} worm in the experiment, we let x_i be the sum of the observations for (either speed, reversal, or stop data) over 30 seconds for its 1^{st} cycle and let y_i be the sum of the corresponding observations for its 2^{nd} cycle. This is done for all the worms giving data $(x_1, y_1), ..., (x_n, y_n)$. The two nonparametric tests mentioned above and described in Conover(1980) can then be used to test whether the correlation between the x's and the y's

is zero. An alternative, of course, would involve testing whether the correlation coefficient is equal to zero. This test is more powerful but also relies on the assumption of normality. Recall that in the case of normality, zero correlation implies independence.

Chapter 4

Conclusions

We now conclude with 2 recommendations which we offer to the experimenters and a summary of our findings.

As mentioned throughout the project, the independence of cycles was assumed but was not established because of difficulities arising from the failure to identify worms and their corresponding cycles. It is, however, ethical that "independence" should be established when dealing with such data if results obtained should have a well founded statistical basis. Owing to this fact, I have advised the experimenter against the casual assumption of independence between cycles and have therefore suggested that the experimenter collect only a single cycle per worm. Alternatively, I have suggested to the experimenter that the computer tracking program be rewritten so that it identifies worms with thier corresponding cycles. In this case a test for cycle independence has been proposed and is described in Section 3.3.

It was mentioned earlier that the department of Biology has been carrying out a number of experiments with similar objectives but varying experimental conditions. Changes in experimental conditions such as different salt levels on the behaviour plate and different light intensities were intended to create all possible conditions in which a light response

might be detected. Now it is possible that in some future experiment, a light response may be detected. This effect may not be due to an actual response but rather due to the fact that so many experiments have been performed. This is analogous to the situation where someone flips a coin many times. You do not expect 10 heads to turn up in a row but eventually this may happen. It is not due to an unfair coin but rather due to the many flips of the coin. In view of this, any evidence of an effect due to light should be confirmed with subsequent experiments.

Our analysis has consisted of a number of steps. Figures 3.1-3.6 show plots of the data for the three variables: speed, reversals and stops. These plots do not reveal any response due to light. In addition the formal analyses of Section 3.2 do not reveal any light response. From the above analyses, the only conclusion to be drawn is "Given the current conditions and procedures for the light response experiment, there is NO evidence that nematodes respond to light".

Finally, it was brought to my attention by Dr. D. M. Eaves (a member of my advisory committee) that an entirely different analysis could be conducted. Such an analysis would be based on studying the sample paths (trajectories) travelled by each nematode. This analysis would have the advantage of comparing the direction of travel, a characteristic not captured by the speed, stop and reversal data. The analyst would translate the points of origin to be the same for all worms and then consider the length of the path and the directions followed by the various worms. Such an analysis would require the availability of the original data in the form of coordinates in order that the sample paths could be plotted. I have comfirmed with the experimenters that the coordinates are available.

References

Burr, A.H. (1985). The photomovement of caenorhabditis elegans, a nematode which lacks ocelli. Proof that the response is to light not radiant heating. Photochemistry and Photobiology Vol. 41, No. 5, pp. 577-582.

Hedgecock, E.M. and Russel, R.L. (1975). Proc. Natl. Acad. Sci. USA 72, 4061-4065.

Koch, G.G. and Landis, J.R. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics 33, 133-158.

Conover, W.J. (1980). Practical Nonparametric Statistics, (2nd edition). Wiley, New York.