AN ORDINAL RESPONSE MODEL FOR THE CONTROL OF EUCALYPTUS *GRANDIS* CUT STUMPS OF MULTIPLE STEM ORIGIN

by

Lynn Beresford Broadfoot

Submitted for the Degree of Master of Science
in the Department of Statistics, University of Glasgow,
September 2000

to my Grandmother

CONTENTS

SUMMARY

It has been suggested that an effective predictor of the survival of Eucalyptus *grandis* cut stumps of multiple stem origin in trials using selected herbicide, may be found from measurements taken on the dimensions of the cut stumps (or stools). To examine this conjecture, data on stool diameter, sum of cut surface diameters and highest height of coppice growth was collected from 714 cut stumps during a study of cut stumps to assess the efficacy of five different types of herbicide in combination with three methods of application, carried out in the KwaZulu-Natal Midlands region of South Africa. Although the measurements were available they were not used in the final analysis of the trial. Little, Maxfield and Kritzenger (1997), found that those stumps treated with herbicide using a basal frill method of application were more efficient at killing stumps on the first and second applications. It was reported that there were no statistical differences between the herbicides. The work described in this thesis follows on from, and extends their analysis. By considering the continuous measurements of stool dimensions in addition to the treatment factors and modelling the total number of applications until a kill is achieved, as an ordinal response it was hoped to construct an accurate predictor of cut stump survival.

In Chapter 1, an outline is given of the forestry background to the control of the Eucalyptus *grandis* cut-stumps problem, and the original KwaZulu-Natal experiment is more fully described. A resume of Little, Maxfield and Kritzenger's results follows.

A preliminary examination of the data is described in Chapter 2. One outlier from the continuous variates was identified and changed to a more meaningful value. Stool diameter, sum of cut surface diameters and highest height were considered to be potential predictors of cut stumps survival because of varying degrees of linearity when plotted against the cumulative sample logits with stool diameter showing the strongest linear trend. More complicated functions of the continuous variates were assessed in the same way but with little success. One of the new variables was derived from stool

diameter and sum of cut surface diameters to produce the ratio of the stool diameter to sum of cut surface diameters, by seeking to account for the poor performance of the cut surface method of application in cut-stumps of multiple stem origin. This poor performance is in contrast to the relatively largely successful results seen in trials of a similar nature on cut-stumps of single stem origin. No evidence was found from this sample that the ratio of stool diameter to the sum of cut surface diameters had a systematic effect on the total number of applications until a kill is achieved. However, for thoroughness and continuity this variable was assessed in a best subsets approach to selecting the best linear predictor in Chapter 4 and then later used in an additional analysis. This meant that there were 3 strong possible predictors for modelling the survival of cut-stumps in this study: stool diameter, highest height and sum of cut surface diameters. As a small proportion of the trial had been terminated before its completion the response contained some 29 missing values. The occurrence of missing values was scrutinised. The pattern of missing data appeared to be random.

The present case is an example of the general polytomous data problem with 4 response categories (representing total number of applications until a tree stump is killed). So, in Chapter 3, approaches to ordinal response data in general use are discussed including the use of cumulative logit models. An outline of the benefits of using cumulative logit models for ordinal response data is given which leads to the presentation of the proportional odds model as a suitable model for analysing the cut-stump data. An introduction to generalised linear models is given with a close look, in the general case, at the estimation of parameters using the method of maximum likelihood. The likelihood functions for the multinomial distribution are derived in the final section.

In Chapter 4, the best subsets procedure for obtaining the best linear predictor is explained and further features of the proportional-odds model are discussed. Testing for differences among factors is discussed and the corner-point style of re-parameterisation applied is explained. The log likelihood ratio statistic is presented as the method used in the analysis to examine the adequacy of the models being assessed. Details of the fitting procedure used in the Minitab software are also given. The 601 sample cases for which complete data vectors were available were used to fit each of the ten linear predictors resulting from the best subsets approach to selecting the best linear predictor, using the proportional odds model. It was found that stool diameter and highest height in addition to the constants and treatment factors formed the best linear predictor. The results from testing this model showed that stool diameter, highest height, method of application, type of herbicide and an interaction term for method of application and type of herbicide were all statistically significant in explaining some of the behaviour of the ordinal response variable. Fitting the proportional odds model to these terms revealed that the presence of stool diameter in the model showed that larger stool diameters are associated with a higher number of attempts to kill. Highest height was not as strongly significant in the model but nevertheless was significant at the 5 % level. Of the three methods of application used in the trial basal frill outperformed the cut surface method. The Chopper and Brush-off types of herbicide performed better than the Timbril herbicide. Prediction using the results from Chapter 4, were used to produce graphs of the probabilities, given stool diameter and highest height, of a cut-stump requiring follow up operations.

The results from the additional analysis using the proportional odds model that was performed exclusively on the data from cut-stumps that had been treated using the cut surface method of application, showed that the ratio of stool diameter to the sum of cut surface diameter was not useful in predicting the number of applications of herbicide until a kill is achieved.

It is concluded in Chapter 5 that stool diameter and highest height are useful predictors in modelling the survival of cut-stumps of multiple stem origin. It is also concluded that the method used to analyse the ordinal response in this study is far more efficient than an analysis of variance. It is recommended that the proportional odds model is used as a suitable framework for data of a similar nature. For possible further work on this data set an ordinal response model for nested or hierarchical response data is suggested in Chapter 6.

ACKNOWLEDGEMENTS

I would like to thank Dr. Keith Little of the Institute for Commercial Forestry Research, Pietermaritzburg, Republic of South Africa, who provided the data from his research and explained the background to the problem. I am also very grateful to everyone who has offered comments, criticisms and suggestions at the proof reading stages of this work. Most of all I would like to thank my supervisor for his advice and encouragement throughout.

CHAPTER 1

INTRODUCTION

1.1 Tree poisoning

Unlike most other commercially grown tree species, *Eucalyptus grandis* has the ability to produce new growth via epicormic buds situated in the live bark, or cambium, when felled. These buds originate from meristematic tissue, present in the leaf axils of small seedlings, which have grown outwards. Usually these buds are prevented from sprouting by the presence of growth hormones called auxins, which are produced in the leaves. When the crown is removed, as occurs during felling operations, the inhibitory effect, known as apical dominance, of the auxin on the growth of these epicormic buds is removed and new growth is produced, known as coppice shoots. In the early stages after felling a mass of new coppice shoots are produced. If left, one or two of the coppice shoots will become dominant and suppress the remaining shoots. These coppice shoots can be selectively thinned over time and managed as a coppice stand for the production of pulp wood.

One problem is that the genetic advantages obtained from tree breeding means that larger gains, in terms of volume of timber, may be obtained through the replanting of old eucalyptus stands with genetically superior cuttings (clones) rather than allowing for coppice regrowth. Consequently, if the eucalypt stools are not killed before replanting, the coppice regrowth competes with the newly planted trees, resulting in a significant growth reduction as well as highly variable tree performance in the new trees. Therefore there is a clear desire to control coppice regrowth by killing the tree stools. The problem of killing stools is not just specific to coppicing eucalypts but is an established difficulty for foresters working with hardwoods.

Researchers have sought to kill tree stools by means other than the manual bashing method, which is expensive whether carried out by hand or by heavy equipment such as bulldozers. Of the alternative methods of killing stools the most widely researched is the use of herbicides and arboricides on various types of species of hardwood. In experiments carried out by the South African Department of Forestry (Marsh, 1963) to determine the cheapest method of controlling eucalyptus stools (*Eucalyptus saligna*), the cheapest method found was to strip the bark from the stools to a level about 6 in. below that of the soil. However, this method was not found to be 100 percent effective and so interest over the years has centered on the next cheapest method reported, which was to poison the stools. (The economics still hold today.) In the original experiments sodium arsenite was used and was wholly effective but since it is extremely poisonous, both to mankind and wildlife, other cost-effective poisons have been sought.

The treatment of hardwood stools by chemicals has been practiced for at least the last sixty years. Historically, various inorganic chemicals were used, such as salts of arsenic, sodium chlorate and hydrocarbon oils, which were sprayed or painted on the stools. Today, herbicides such as phenoxyacids, glyphosate and triclopyr are used in addition to ammonium sulphate, which are applied by a variety of methods, the most common of which are the cut surface, basal frill and foliar spray methods (See section 1.2.).

In practice, it has been difficult to find a herbicide that is 100 percent effective on a hardwood species at the first application and so consequently follow-up operations are required. The present day problem has shifted slightly to that of finding a herbicide that will kill tree stools with the minimum number of follow up operations (for both efficiency and economic reasons). In the Institute for Commercial Forestry Research (ICFR) Annual Report 1998 Little *et al.* showed that it is possible to obtain an almost complete kill to *Eucalyptus grandis* single stem stools using selected herbicides applied using the cut surface method of application.

The distinction between single stem and multiple stem eucalyptus stools is as follows. Where there is coppice re-growth the coppice stems are selectively thinned to leave the most robust on the stool (original stem). Generally the coppice shoots are reduced to the best two and then to a single stem. The stools of multiple stem origin are those stools that have been allowed to coppice for 2-5 rotations and so have several dominant stems. When a forester fells a compartment of trees with single stems the stems will be felled at a uniform height. If allowed to coppice, the problem that then arises with the stools of multiple stem origin is that each time they are re-felled, the discrepancy between felling height above ground becomes greater and is dependent on how high the bowl of the stem is. This results in a larger stool in terms of height and diameter, yet with similar cut surface area. As the reserves (sinks) are larger it is thought that the ability of these larger stools to withstand efforts to kill them is that much greater. Imagine that the stool is an upside down bowl with pencils sticking out of it (the pencils being the stems that are to be harvested). The stems would be cut as close to the bowl as possible such that the height of the cut-surface from the ground would equate to the height of the stool.

1.2. <u>The sample</u>

Of the 714 tree stools involved in this study only 630 were included in the final analysis; this was due to a number of reasons explained below. The study was designed as a 3 x 5 factorial design with two additional controls replicated three times. The trial was initiated on stands of *Eucalyptus grandis* that had been coppiced four times, covering a total area of 6480m^2 in the KwaZulu-Natal Midlands region of South Africa. The trial design allows for 48 plots arranged in three blocks of 17, to which the treatments were applied in a completely randomised fashion. Each plot comprised of 14 stools with the treatments being applied to those stools that were living prior to being felled. Five types of herbicide were tested in combination with three methods of application. The two additional controls did not receive herbicide and any coppice re-growth was manually removed with a bush knife.

The distinction between the two types of control lies in the addition of a basal frill so that a basal frill was cut in to the stumps of one set of the controls. This was used in comparisons with the basal frill treatments. The other control was used as a comparison for the foliar spray and cut surface treatments in the initial analysis. (See section 1.3.) Unfortunately, due to circumstances outwith the researcher's control, the trial was terminated prematurely. At some time between the researcher's fifth and sixth visit to apply the treatments and record the assessments of the stools, the forester had cut all the coppice down (where present) and had also used a stump grinder to remove stools to allow for access for timber extraction and planting.

On the fifth assessment 100 percent of the stools that were manually bashed were still alive. However, these data cannot be used in the analysis. Due to the forester's actions an artificial ceiling has been forced on the control data and consequently, owing to the nature of the results from the first five assessments, the data as it stands holds no information, and even less distributional information. Hence all 84 control units were removed from the analysis and are not discussed additionally here. Furthermore, the treatment information recorded on the fifth assessment was deemed to be unreliable as the treatment effects at this stage were now confounded with the effects of the forester's actions. This meant that a further 15 records were replaced with missing values in the analysis to maintain balance, so that in the final analysis the sample size was reduced from 714 to 630 with 29 missing values in the response.

On inspection of the data, it was clear that although the data was right censored at the fourth application, the treatment effects were well established by that point. In support of this 28.1 percent of those that received a foliar spray were still alive at this stage, 11 percent of those where the cut surface was sprayed were still alive and only 0.5 percent of the stools that received a basal frill were still alive. Losing the control data does mean that there is an absence of a control level for use in comparisons with the various treatment combinations which in turn means that care is needed when interpreting the results of the analysis (See Chapter 4.)

Stools of multiple stem origin are acknowledged to be more difficult to kill than single stem origin stools. The *Eucalyptus grandis* coppice re-growth draws on a large reserve in terms of the underground portion of the stem and roots and for this reason, the use of soil active or systemic herbicides has proven to be more effective than the use of contact herbicide. Five types of herbicides were to be evaluated in this trial along with three types of method of application. As mentioned previously, the trial was a 3 x 5 factorial experimental design and the types of herbicide applied are given below in Table 1.1.

Table 1.1.

Description of Herbicides used in the trial

Type of Herbicide	Commercial Name
Triclopyr (amine)	Timbril
Triclopyr (ester)	Garlon
Imazypyr	Chopper
Metsulfuron-methyl	Brush-off
Triclopyr (ester) + glyphosate (isopropylamine salt)	Nomix + Garlon 6

The first method of application of herbicides to the stools to be discussed here is the cut surface method. The herbicides were applied to the cut surfaces of the stools immediately after cutting, as it is generally accepted that the sooner after felling the herbicide is applied, the greater the degree of success. The herbicide was applied to the cut surface according to label recommendations within 15 minutes of felling. Herbicides were applied with a marker dye to the cambium region of each stool using a hand held sprayer dosing gun. Care was taken during application to ensure a directed spray on to the stool in order to minimise drift. An important aspect considered when using this method was the residual activity of the herbicide used in the soil. Any lasting effects of the herbicide in the soil can consequently mean that the period before planting has to be increased, as some commercially planted species may be susceptible to soil active herbicides. If shown to be successful, this method could prove to be one of the most cost-effective, but there are indications that larger diameter stools and stools of multiple stem origin require follow-up applications.

The basal frill method relies on the application of herbicides into cuts made into the cambium at the base of the stool. An axe was used to make horizontal, downward-angled cuts into the sapwood of the stool such that run-off of the herbicide onto the soil was limited. These cuts were made as complete rings and the herbicide was applied within 15 minutes of the cuts being made. The herbicide was applied to the cuts as a directed spray using a solid stream tip nozzle with a backpack sprayer. The initial treatments were applied the day prior to felling so that the presence of any plantation residue would not hinder operations. This method is particularly effective in the killing of stools of multiple stem origin. Although this method is extremely effective in achieving a one-off kill, it is very time consuming and therefore may not be considered as cost-effective as other methods.

Herbicides can also be applied as a foliar spray onto coppice re-growth when it is approximately 1.5m (0.5-2.0m) in height. Spraying of the herbicides may take place from the air or on the ground with pressurised backpack sprayers, depending on the product used. Nowadays, compartments are seldom allowed to remain unplanted for any length of time, resulting in the aerial application of herbicide as a foliar spray being limited to temporarily unplanted areas as a pre-plant spray. The reason being that the planted seedlings amongst the coppice regrowth could be equally susceptible to the herbicides used for this method of application. In this trial the herbicides were applied according to label recommendations when the dominant height of the coppice regrowth was 1.5m. Treatments were applied using a backpack sprayer. To avoid over-application, the herbicide was applied in a circular motion from the bottom to the top of the coppice re-growth, wetting the entire canopy to the point of run-off. The foliar applied treatments were sprayed 86 days after felling when the coppice re-growth was approximately 1.5m tall.

After the initial treatments were applied, assessments were made at the same time that the treatments were re-applied. In other words, if coppice was present it was re-treated with the initial treatment. This was recorded per stool as a cumulative value and indicates the initial application of the treatment added to the number of return times until the stool was killed. At each assessment the presence/absence of coppice re-growth was noted and a record of survival/kill registered.

Table 1.2. contains a full list of the variables recorded in the study, and the number of cases for which each was missing. Altogether there were 20 variables and as mentioned previously 630 observations were used. Of these observations at least 12 were missing for each variate, and it was believed that a measurement was missing independently of the combination of treatments applied. This assertion, however, remains to be investigated.

Table 1.2.

Variables recorded during trial

Variable	Type	No. of missing values
Plot number	Identifier	-
Stool number	Identifier	-
Treatment number (method*herbicide) (1 – 18)	Factor	-
Replication (1 – 3)	Factor	-

Type of herbicide applied (1 – 5)	Factor	-
Method of application (1=cut surface, 2=basal frill, 3=foliar spray)	Factor	-
Additional control	Factor	-
Total number of applications	Response	29
Number of times coppice controlled	Variable	12
Stool diameter at ground level (cm)	Variable	12
Total number of coppice stems per stool that were felled when harvested	Variable	12
Total number of coppice stems per stool with cut surface diameter > 5cm (Never more than three)	Variable	12
Diameter of cut surface 1 > 5cm	Variable	12
Diameter of cut surface 2 > 5cm	Variable	493
Diameter of cut surface 3 > 5cm	Variable	618
Sum of cut surface diameters > 5cm	Variable	12
Height 1 above ground of each of those cut surfaces that is over 5cm. (cm)	Variable	12
Height 2 above ground(cm)	Variable	492
Height 3 above ground(cm)	Variable	618
Height of the highest cut surface (cm)	Variable	12

1.3. Previous analysis

The above data was collected in 1996, and the statistical analysis then carried out was published in Little, Maxfield and Kritzinger (1997). There follows a brief description of their results.

All of the 714 observations were used in their analysis. The response data, total number of applications until the stump is killed, was originally analysed as binomial data and an arc sine transformation was performed for the purpose of equalising the variances within treatment groups (Snedecor & Cochran, 1980). Statistically speaking, replacing the proportions $\hat{p}_{ij}$ by the angle whose sine is $\sqrt{\hat{p}_{ij}}$ allows the proportions near zero or one to spread out so as to increase their variance. In effect, when successfully transforming the response in this way the data can be said to be 'Normalised'. The purpose of doing this was to perform an analysis of variance on the factorial experiment.

In the ICFR bi-annual report the authors reported "In comparison to manual coppice control, the use of herbicides proved to be more effective. Of the three methods used the basal frill method proved to be the most effective with little need for follow up operations.....no significant differences occurred between the various herbicides." The results presented in the report were not transformed and were in the form of a histogram with a bar indicating the mean number of operations for each treatment. The report went on to conclude that the manual bashing of coppice regrowth on stools of multiple stem origin was not as effective as the use of herbicides. There were no significant differences between the various herbicides used. Of the three methods of killing stools that were tested, the basal frill proved to be the most successful with little need for follow-up operations.

The methods employed to analyse this data are unsatisfactory for the following reasons. The most obvious is the authors do not use the ordinal nature of the response in their analysis. Transforming the data into the arcsine scale only allows the standard analysis to be used as an adequate approximation and performing an analysis of variance has the additional complication of back transforming the results. Also, the method does not incorporate the individual explanatory variables such as stool diameter, sum of cut surface diameters and height. If the data was modelled using generalised linear model, normalising the response by transforming to the arcsine scale would mean that the analysis is restricted to using the identity link function.

The work described below follows on from, and extends, the above analysis. By considering more complicated functions of the variables, and a generalised linear (proportional-odds) model, it was hoped to find a framework that would adequately model the error structure of the data whilst taking the opportunity to use the cumulative information within the response itself. The proportional-odds model is more commonly used in the modelling of survival data in medical research where in general it is used to model the odds of an individual surviving beyond some time point t.

Of more theoretical interest, the present study also gave an opportunity to introduce a suitable model to analyse the performance of new herbicides by modeling the odds of response below any given number of applications on a real set of data. At present the statistical methods employed within the forestry industry for this type of data analysis are rather weak.

CHAPTER 2

PRELIMINARY EVALUATION OF THE DATA

2.1. The purpose of data-exploration

Data from 630 felled eucalyptus trees monitored during the study described above were available for analysis. As previously stated records of the control plots were disregarded and 15 records relating to the results of a fifth application of herbicide were deemed redundant due to exceptional circumstances, and were recorded as missing to maintain balance in the designed experiment. In total, values were missing from 618 data vectors. Excluding the variables collecting information on the diameters and heights of stems > 5cm (dependent on number of stems > 5cm), values were missing from 29 data vectors. These missing values were mostly due to the 15 redundant records and 12 trees that were dead prior to felling.

Before this data set was used in formal procedures using a proportional-odds model, it was explored so that desirable features, (such as apparently good predictors) might be highlighted and, where possible, undesirable ones ameliorated. The question of whether values of the various variables were missing "at random" was examined. Extreme observations, which were thought to be wrongly recorded or which, even if correctly noted, would have had undue influence on the formal procedures to follow, were identified and, where necessary, eliminated. To cut out excessive computation and to avoid over-fitting of the model, the explanatory variables were assessed and the number was reduced. In some cases, variables that appeared to have little predictive power were excised. New composite variables were created and their relationship with other variables explored and their ability to account for the probability that a stump was killed with at most j treatments assessed.

2.2. Missing values

Of the data vectors, 29 (4.6%) were incomplete. The number of cases for which the response (total number of applications) was missing in each group is shown in Table 2.1.

Table 2.1.

The pattern of missing data

Method	Herbicide					
	1	2	3	4	5	All
1	4	2	3	2	2	13
	(9.52%)	(4.76%)	(7.14%)	(4.76%)	(4.76%)	(6.19%)
2	4	1	0	1	3	9
	(9.52%)	(2.38%)	(0.00%)	(2.38%)	(7.14%)	(4.28%)
3	5	0	0	2	0	7
	(11.9%)	(0.0%)	(0.0%)	(4.76%)	(0.0%)	(3.34)
All	13	3	3	5	5	29
	(10.3%)	(2.38%)	(2.38%)	(3.96%)	(3.96%)	(4.6%)

Although it is possible that this relatively small number of missing values could have been non-random due to treatment or location specific effects Table 2.1. gives no consistent evidence that such a relationship exists.

2.3. Identification of outliers

With the exception of those details recorded on the factorial design of the experiment, there were 9 additional measurements recorded on each of the stools. The sample distribution of the values of each of these 9 variables was examined to identify outliers. In this context, an outlier was any observation sufficiently different in character from the rest of the sample to have an excessive influence on the calculation of summary statistics for the whole sample. Such observations were investigated as they arose in the preliminary inquiry.

2.4. The relationships between stool dimensions

It was primarily intended to find, from the values of those variables on the dimensions of each stool, additional variables that might further explain the survival of stools to different treatments. It was thought that any relationship between number of times the coppice re-growth occurred and the amount of herbicide applied would be quite important. Since there was no information on the actual amounts of herbicide applied, investigation found that the total number of applications of herbicide until a kill was achieved could in itself be a measure of 'how much' herbicide was applied to a stool. However, on closer inspection it was discovered that on every assessment visit, each stool was always controlled for coppice re-growth until the stool was killed. Therefore an anomaly arose here, since the coppice control was effectively the total number of applications until death of each stool minus 1: not a very useful predictor!

Plots of stool diameter against the sum of the cut surface diameters are shown, for the sample, in Figures 2.1., 2.2(a) and (b) respectively. It is clear from 2.1. that there exists a positive linear relationship, if somewhat messy, between stool diameter and the sum of cut surface diameters. The sum of cut surface diameters increases with stool diameter so that, in general, the larger the stool diameter the larger will be the sum of the cut surface diameters of that stool. Figure 2.2(a) investigates this relationship further. The points have been colour coded by the response categories, such that each colour point represents a stool's number of follow up operations until a kill was achieved. As suspected those stools with bigger dimensions mostly took a greater number of follow-up operations until killed. There are two points that are the exception to this, where both have large dimensions and were killed after the first application. Figure 2.2(b) plots the same graph but this time with the points colour coded to represent the type of method used to apply the herbicides. (1=cut surface, 2=basal frill, 3=foliar spray.)

Figure 2.1.. Relationship between Stool Diameter and Sum of Cut Surface Diameter.

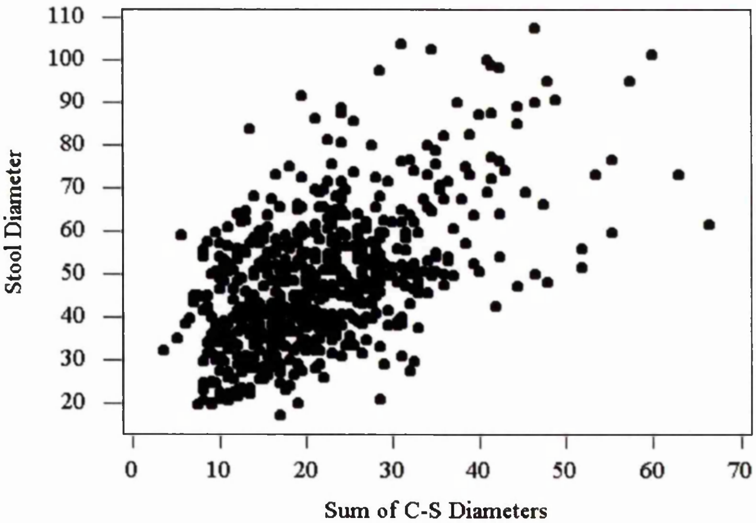


Figure 2.2(a). Relationship between Stool Diameter and Sum of Cut Surface Diameters. (Number of applications until a kill is achieved.)

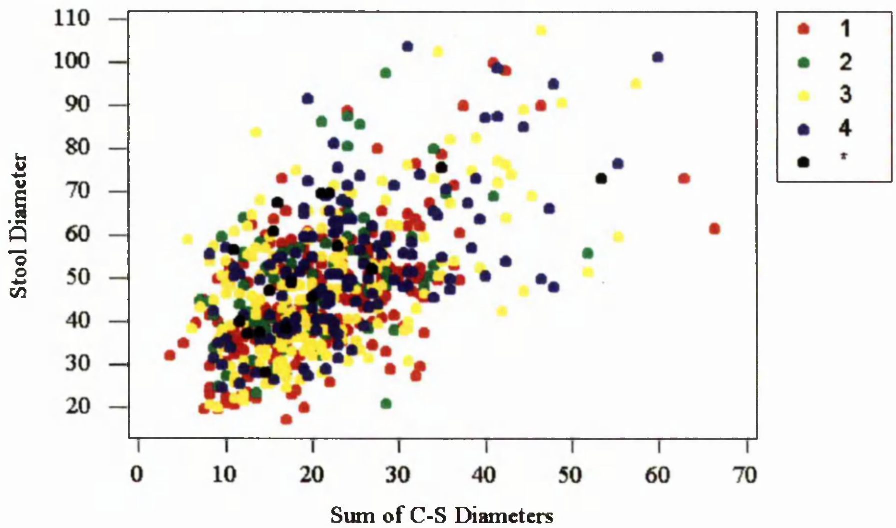
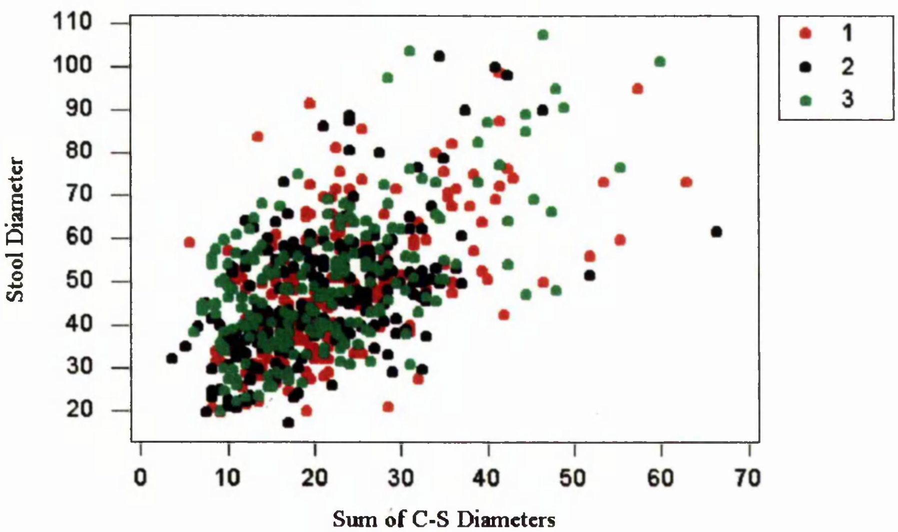


Figure 2.2(b). Relationship between Stool Diameter and Sum of Cut Surface Diameters. (Type of method used to apply herbicide.)



It was thought that those stools with larger cut-surfaces were killed sooner when herbicide is applied using the cut-surface method. However figure 2.2.(a) and figure 2.2.(b) together show the point representing the stool with the largest sum of cut surface diameters was treated using the basal frill method and the other using the cut-surface method of application. From both figure 2.2(a). and 2.2(b). a pattern can be identified in those stools with larger stool diameters killed first time;. stools with large diameters killed first time were largely treated using the basal frill method of application. Where there are instances of stools with large stool diameters having had 3 or 4 follow-up treatments these have almost always been treated with the cut-surface or foliar spray methods of application. Larger sums of cut surface diameters were rarely killed first time regardless of the method used. (This notwithstanding the observation that the distribution of stools with large sums of cut surface diameters is slightly uneven across methods with most being found in the cut surface and foliar spray treatments.)

In an attempt to relate all three variables: stool diameter, sum of cut surface diameters and highest height of coppice re-growth, a new variable, ratio of the stool diameter to the sum of cut surface diameters, was created. A scatter plot of this ratio vs highest height is shown in Figure 2.3(a). In doing this at least one of the dimensions measured on three of the stools were found to be much higher than those of the remainder of the sample. These three outlier measurements were from cases 376, 602, and case 261. Case 376 has an abnormally high measurement of height recorded (601cm) implying that the highest height of the felled coppices was over 6 metres; a highly implausible value. It was concluded that the data in this case was incorrectly recorded and it was replaced with the more reasonable value of 60cm. Cases 602 and 261 had high ratios due to very small sums of cut surfaces relative to their stool diameter measurement. Whilst this is of some concern, these measurements were not excised from the data set.

Figure 2.3(a). Relationship between Highest Height and Stoold:C-SD.

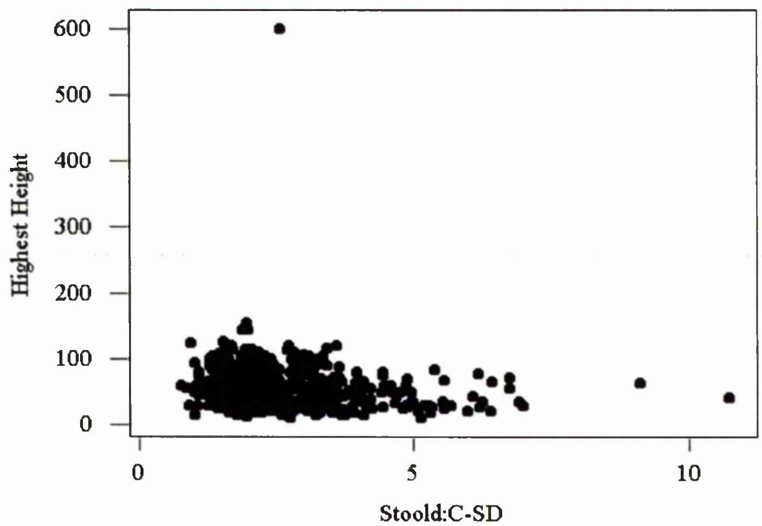


Figure 2.3(b). Relationship between Highest Height and Stoold:C-SD with case 376 replaced.

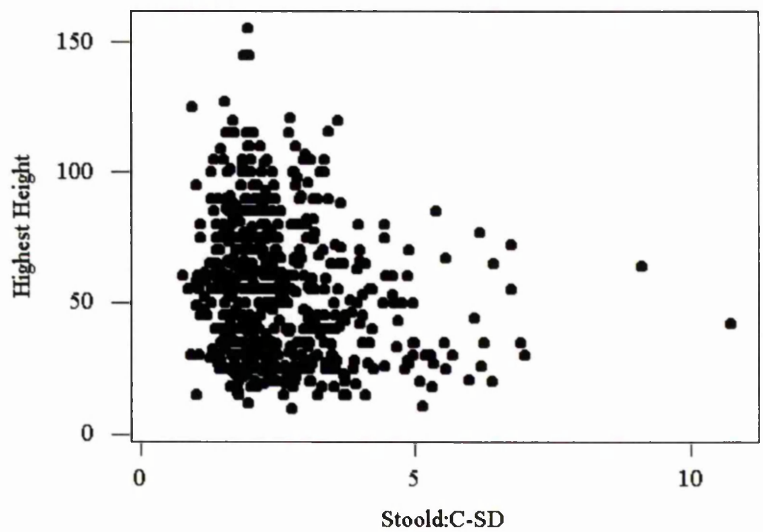


Figure 2.3(c). Relationship between Highest Height and Stool Diameter.

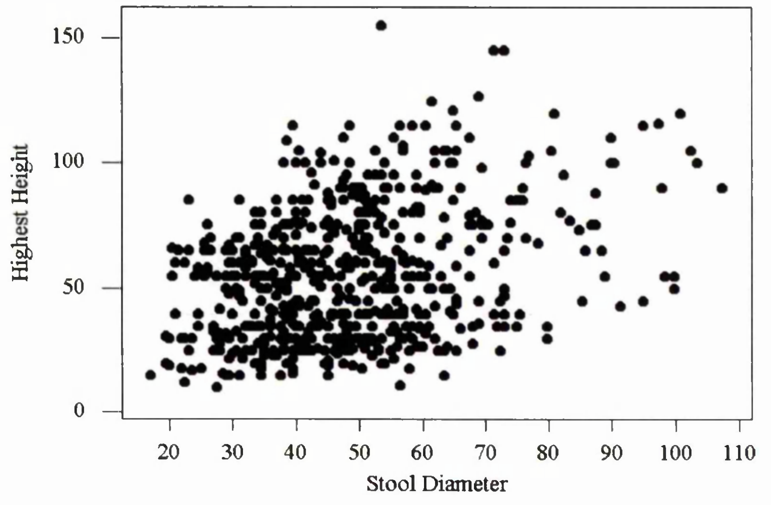
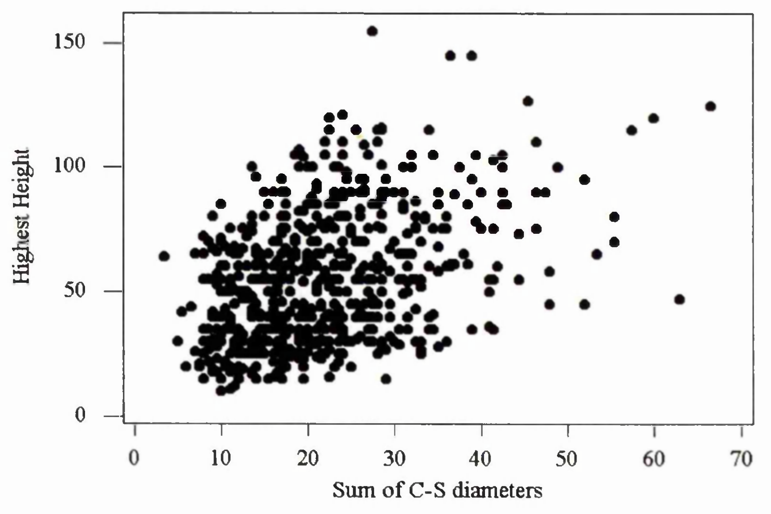


Figure 2.3(d). Relationship between Highest Height and Sum of Cut Surface Diameters



Re-producing the scatter plot (Figure 2.3.(b).) with the ill-fitting observation replaced, reveals that the ratio between stool diameter and cut surface diameters is fairly skewed, i.e., within the range of values 1 – 5 cm for the bulk of the data, with a light spattering of points which lie outwith and to the right of this range. The relationship between these two variables appears to be weak and it is easy to observe that the distribution of the ratio between stool diameter and sum of cut surface diameters lies within the same range whatever the highest height measurement.

Further plots of highest height vs stool diameter and highest height vs sum of cut surface diameters are shown in figures 2.3(c). and 2.3(d). From figure 2.3(c) it is seen that highest height and stool diameter measurements have a weak positive relationship. As stool diameter increases the variability of highest height also increases. Figure 2.3(d) shows a similar relationship.

Table 2.2.

Table of Pearsons Correlation Coefficients.

	Stool Diameter	Sum of C-S Diameters	Stool Diameter: C-S Diameters
Sum of C-S Diameters	$\rho = 0.558$		
Stool Diameter: C-S Diameters	$\rho = 0.177$	$\rho = -0.586$	
Highest Height	$\rho = 0.254$	$\rho = 0.295$	$\rho = -0.132$

Overall, these scatter plots gave no suggestion that highest height, stool diameter, sum of cut surface diameters and the ratio of these latter two variables are strongly correlated. However, stool diameter and sum of cut surface diameters are moderately correlated $(\rho = 0.553)$, as is sum of cut surface diameters to it's ratio (with stool diameter) which is expected. See Table 2.2 above.

2.5. <u>Assessing possible predictors</u>

Due to the nature of the response variable it was necessary in the first instance to consider the proportions within each response category to ascertain the relationship of the response (proportion killed in category j or below) to changing levels of factors and continuous variables. The manner in which the logistic transform of the response probability varied over the levels of the predictor was essential in determining whether any relationship is linear. The existence of such a relationship gives justification for using the predictor in the modelling of the response. This is because the cumulative probability that a stool is killed on the jth application is given by

$$P(Y \leq j) = \pi_1 + \ldots + \pi_j, \qquad j = 1,\ldots,J$$

and the cumulative logit is equal to

$$\operatorname{logit} P(Y \leq j) = \log\left(\frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots + \pi_J}\right), \qquad j = 1,\ldots,J-1$$

In the usual ordinal logistic regression model,

$$\operatorname{logit} P(Y \leq j) = \alpha_j + \beta\mathbf{x},$$

so when plotting the cumulative *sample* logits against a suitable predictor the relationship should be approximately linear.

For each of the continuous variables, a factor was defined whose levels corresponded to consecutive intervals of that variable. These were used to judge the presence of linearity in the empirical cumulative logits in plots for response categories $Y \leq 1$ up to $Y \leq 3$. Figures 2.4(a). – 2.6(d). show the presence or absence of linear trends in the cumulative *sample* logits for stool diameter, highest heights, sums of cut surface diameters and the ratio of stool diameter to sum of cut surface diameters respectively.

Stool diameter at $Y \leq 1$, figure 2.4(a). showss a negative linear trend (although there is slight concavity present) in the logit proportions of those stools killed against their level of stool diameter size. It is clear that as stool diameter increases the proportion that is killed in the first application decreases. Figures 2.4(b) & (c) emulate this linear trend more definitely and so the above can be generalised to state that as stool diameter increases it is thought that the probability that a stool is killed on or before the jth application decreases linearly. Generally, the linear trend applies throughout the response categories and it was noted that although the logits of the proportions killed on or before the jth category decreased with stool diameter, overall the cumulative logits of the sample proportions increased steadily.

Highest height measurements show the same trend initially at $Y \leq 1$, although it is not as strongly linear as the stool diameter measurements. However, this linearity becomes less obvious in figures 2.5(b) & 2.5(c). This is due to those stools killed on the second application largely having smaller height values. On the third application of herbicide a greater number of stools with larger height measurements were killed. In general, the highest height measurements show a weak negative trend in the cumulative logits, although signs of curvature are present.

The sum of cut-surface diameters displays some similar qualities for the three cumulative logit proportions, notably the general negative decrease in the cumulative sample logits with increasing factor levels.. In Figures 2.6(a) - (c), the relationship between the cumulative sample logits and the sum of cut surface diameters shows signs of curvature, indicating that the relationship between the two might be non-linear. That notwithstanding the probability that a stool is killed

Figure 2.4(a). $Y \leq 1$: Relationship between Stool Diameter and Logit of Proportion Killed.

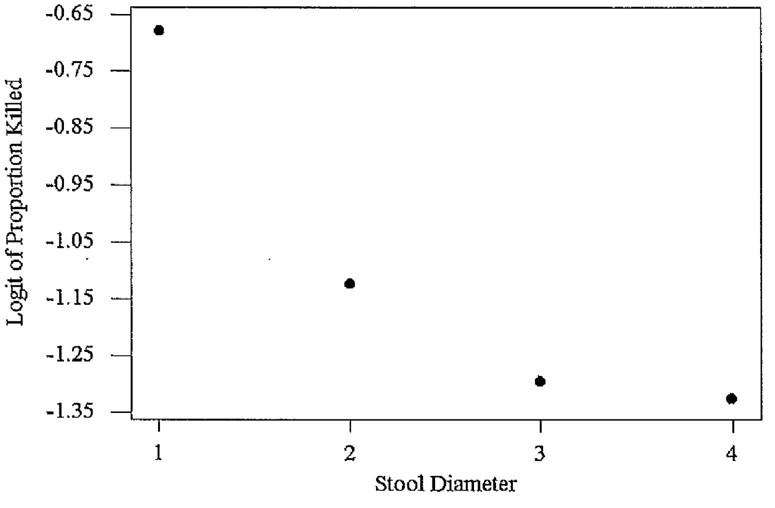


Figure 2.4(b). $Y \leq 2$: Relationship between Stool Diameter and Logit of Proportion Killed.

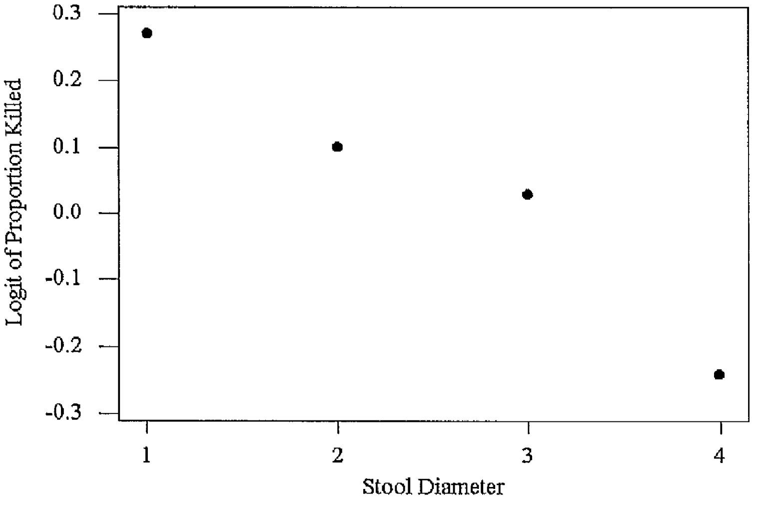


Figure 2.4(c). $P(Y \leq 3)$: Relationship between Stool Diameter and Logit of Proportion Killed.

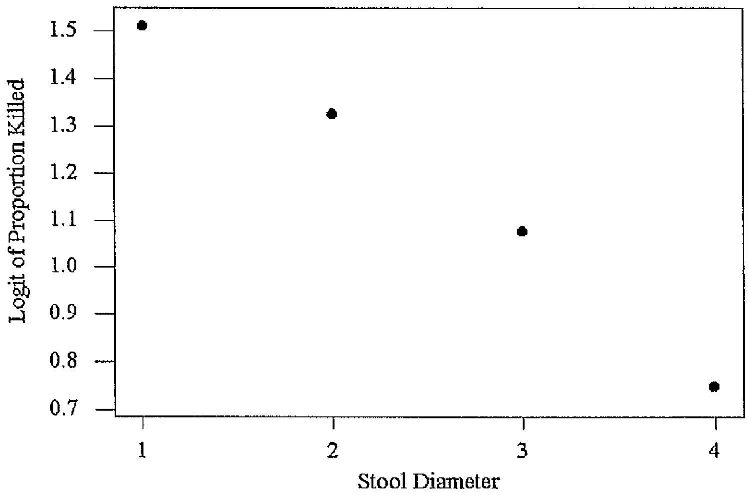


Figure 2.5(a). $Y \leq 1$: Relationship between Highest Height and Logit of Proportion Killed.

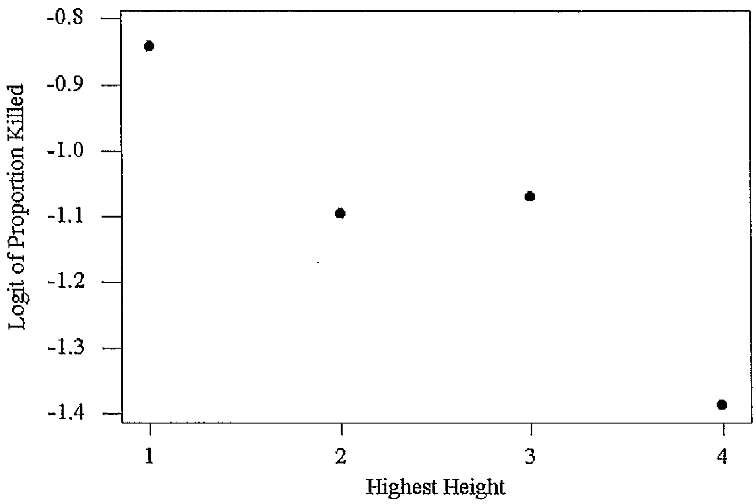


Figure 2.5(b). $Y \leq 2$: Relationship between Highest Height and Logit of Proportion Killed.

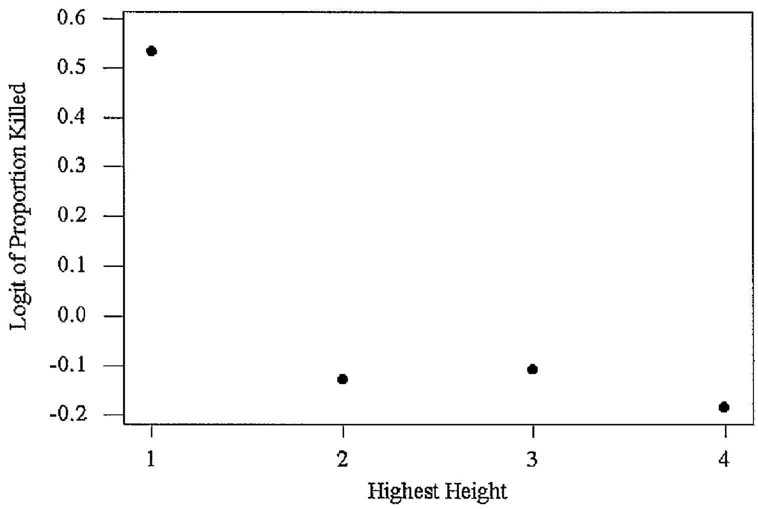


Figure 2.5(c). $Y \leq 3$: Relationship between Highest Height and Logit of Proportion Killed.

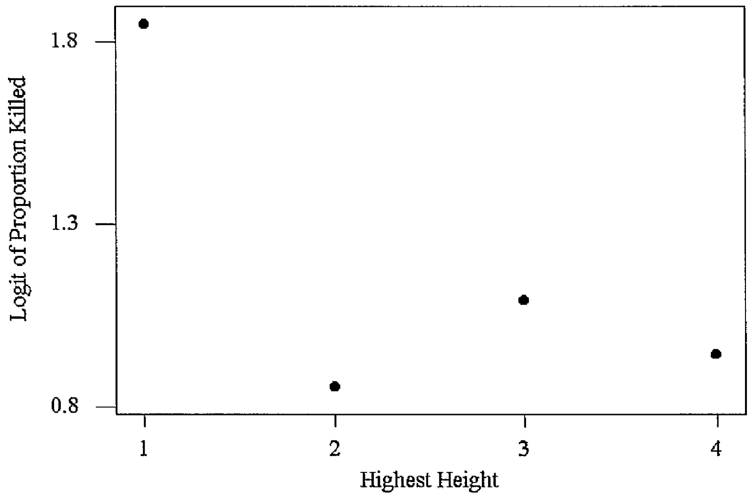


Figure 2.6(a). $Y \leq 1$: Relationship between Sum of C-S Diameters and Logit of Proportion Killed.

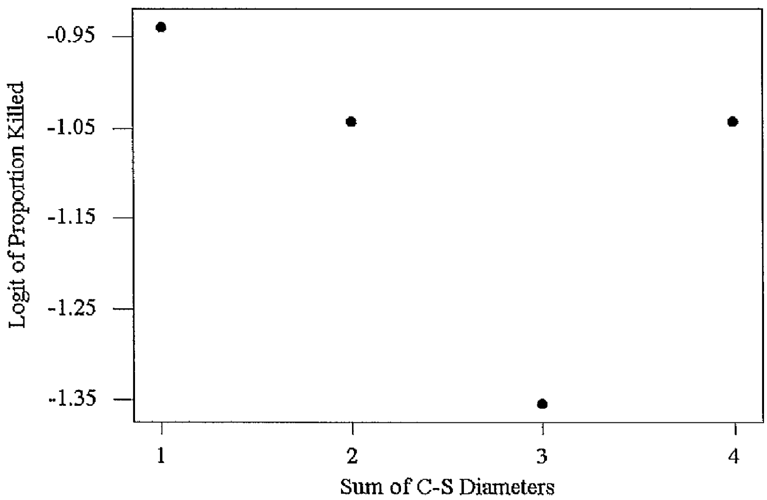


Figure 2.6(b). $Y \leq 2$: Relationship between Sum of C-S Diameters and Logit of Proportion Killed.

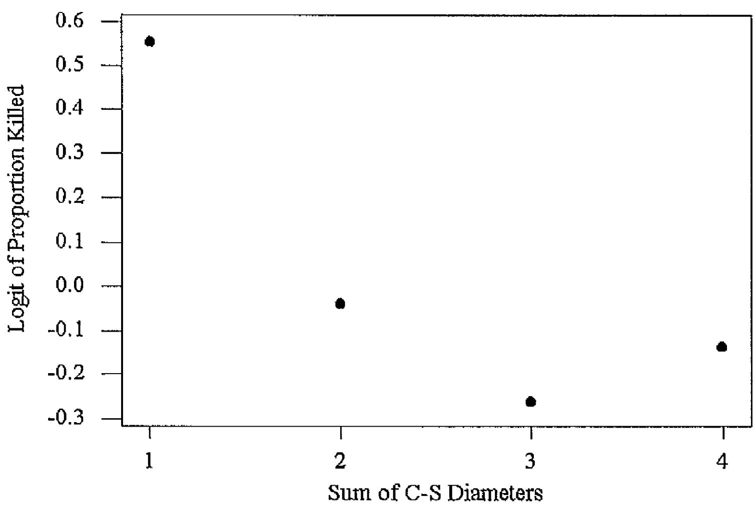


Figure 2.6(c). $Y \leq 3$: Relationship between Sum of C-S Diameters and Logit of Proportion Killed.

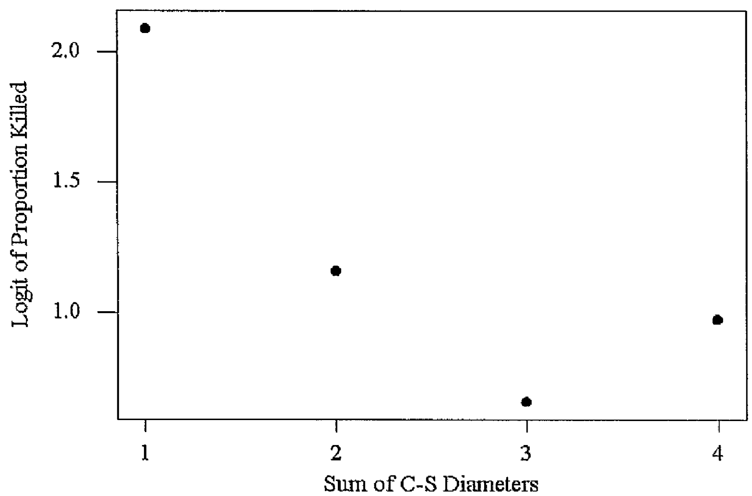


Figure 2.7(a). $P(Y \leq 1)$: Relationship between Stool Diameter to Sum of Cut Surface Diameters Ratio and Logit of Proportion Killed.

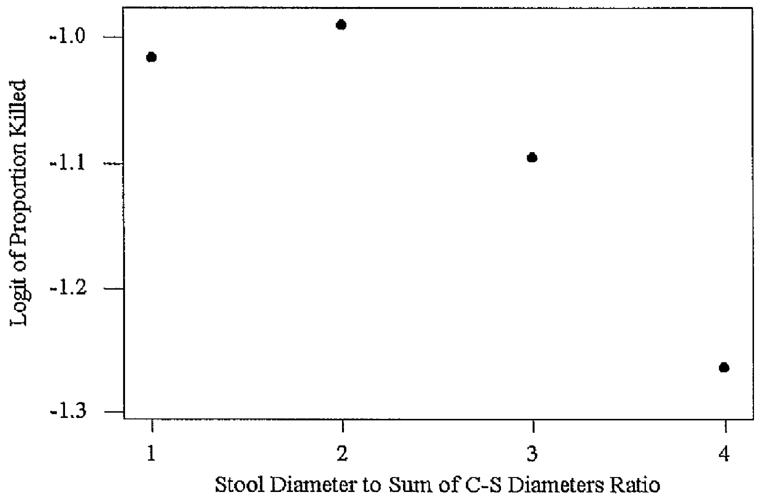


Figure 2.7(b). $P(Y \leq 2)$: Relationship between Stool Diameter to Sum of Cut Surface Diameters Ratio and Logit of Proportion Killed.

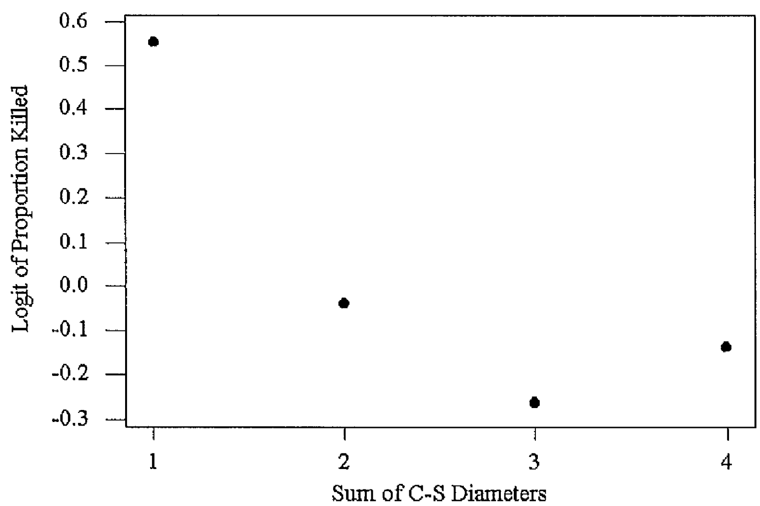
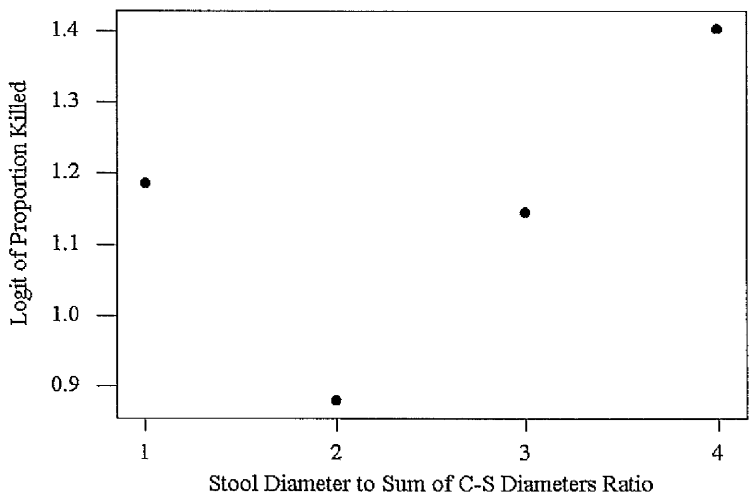


Figure 2.7(c). $P(Y \leq 3)$: Relationship between Stool Diameter to Sum of Cut Surface Ratio and Logit of Proportion Killed.



on or before the jth application decreases as the sum of cut surface diameters increases - the exception being the very largest sums of cut surface diameters.

A new variable stool diameter to sum of cut surface ratio was created from two of the original measurements taken at the time of the trial. This was done in an attempt to formularise the belief mentioned previously, that smaller cut-surfaces on larger stools may inhibit the effect of the herbicides applied. In figures 2.7(a) – 2.7(c), the relationship between the cumulative logits and the ratio of stool diameter to sum of cut surface diameters appears to be non-linear. The proportion of stools killed on or before the first application of herbicide is higher for smaller ratios of stool diameter and sums of cut surface diameter and lower for larger ratios. This is reversed for those stools killed on or before the second application of herbicide suggesting that a greater number of stools with higher ratios were killed on the second application of herbicides. Figure 2.7(c) shows a shifting to the left of the trend seen in figure 2.7(b) so that behaviour in the cumulative sample logits is now increasing with stool diameter to sum of cut surface diameters ratio. The trend here seems to be inconsistent and would be difficult to model.

Of the four continuous measurements studied here stool diameter showed the strongest linear trend. Highest height also showed a degree of linearity but it was thought that of the three measurements assessed stool diameter showed the best linear behaviour. As previously stated, the experiment was originally designed as a factorial experiment and so it is the intention of this study to model the additional information on stool dimension in the systematic component of the generalised linear model whilst staying true to the original factorial design. It was strongly thought (by the researcher) that the ratio of stool diameters to the sum of cut surface diameters was the cause for a poor performance in using the method of cut surface application when treating cut-stumps of multiple stem origin. It is for this reason, although the variable shows little sign of being a useful predictor at this stage, that it was used in a best subsets approach along with the other variables assessed here to find the best linear predictor. It was also used in an additional analysis using the proportional-odds model to illustrate the result more conclusively.

During the course of the preliminary analysis other new variables were created using the stool diameter, sum of cut surface diameters and highest height variables assessed above. In attempts to determine the dimensions of stools that might give some clear insight into the killing of cut–stumps of multiple stem origin the following variables were created and assessed exactly as above, these variables were cut surface area, sum of cut surface diameters squared, sum of cut surface diameters squared multiplied by highest height, stool diameter squared, and stool diameter squared multiplied by highest height (a measurement used in yield of timber calculations). None of these functions when plotted against the cumulative sample logits showed convincing linearity or provided additional information that might be useful in the ordinal response model. The usefulness of the above functions (including cut surface and stool diameter to sum of cut surface diameters ratio) in nonlinear models remains to be investigated.

CHAPTER 3

GENERALISED LINEAR MODELS

3.1 The general problem

The general polytomous data problem may be described as follows. When the response of an individual or item in a study is restricted to one of a discrete set of possible values, such a response is known as polytomous. This extends to include the fixed set of possible values that individually reduce to a simple dichotomy. In such a situation the binary response variables measured on each individual or item collectively form a multivariate response.

A fixed set of categories is said to be ordinal when the categories are ordered much like the ordinal numbers 'first', 'second', 'third' etc. and where it does not make sense to talk of 'distance' or 'spacing' between categories. More distinctly, when the categories are ordered the extreme categories must be treated differently from the intermediate ones.

The number of applications of herbicides it takes to a kill tree stump clearly fits into this general framework. The response categories in this case consist of the number of applications of herbicide until a stump is killed, namely, one, two, three or four. The categories are obviously ordered with the first time and fourth time kills being more extreme than the intermediate numbers of applications. At each application category a success (kill) or failure (survival) is noted. Therefore, the response may be considered to be at least ordinal. In fact, the response is discrete interval scaled which is treated as ordinal due to the way in which the extreme categories are interpreted.

In practice a number of broad ways are employed when dealing with this type of ordinal data. The most basic and simplistic approach is to ignore the structure of the response completely and perform logistic regression analysis for each of the separate binary response categories. In some situations this may be all that is needed but it is not considered further here.

A second approach is to treat the responses as nominal where the ordering is ignored in the first instance. If the resulting conclusions are invariant with respect to the ordering then additional assurance can be taken in this. Caution is advised in this approach as it is unsuitable for response data where there is a 'large' number of categories, since each category in this case has its own parameter vector β_k, so the nominal response model contains many more parameters than a single model for ordinal data.

A third approach is to assign conventional numerical values to the scale points, typically equally spread, i.e., -2, -1, 0, 1, 2 and proceed as for quantitative variables. It is wise to check that the conclusions are not critically dependent on the scoring system used and additional care is needed if there is some tendency to cluster around the extreme points of the scale. Often the use of integer scores is unlikely to satisfy completely the theoretical assumptions. Arbitrary scores of 0, 1, 2, 3,... are often used and analysed by analysis of variance methods on the assumption either that the necessary criteria are satisfied or that the list is sufficiently robust for it not to matter. In such a situation there are methods for deriving scores.

Snell 1964 outlines a method for deriving category scores assuming an underlying logistic distribution. For a scale of k categories, points x_j ; j = (1, ... k) are defined such that category s_j corresponds to the interval x_{j-1} to x_j. Denoting the underlying continuous distribution by $P_i(x_j)$ the probability of an observation i in category s_j is equal to $P_i(x_j) - P_i(x_{j-1})$, i = 1, 2,....; j = 1, 2, ..., k. That is,

$$P_i(x_j) = [1 + e^{-(a_i + x_j)}]^{-1} \tag{3.1.1}$$

where $a_i + x_j$ represents the logit of the proportion $P_i(x_j)$. Equations for the maximum likelihood estimates of the parameters a_i and x_j, are determined and from these approximate estimates of x_j are obtained by substituting $\hat{P}_i(x_j)$ with the observed proportions.

Hence values of s_j are calculated by the mid class points

$$s_j = (x_j + x_{j-1})/2, \qquad j = 2, \quad ,k\text{-}1 \qquad (3.1.2)$$

and the average distance $x - (\log P)/Q$, where P denotes the probability of a value less than x and Q = 1 – P, is added to x_1 and x_k for the extreme scores s_1 and s_k.

Finally a more satisfactory approach is to centre interest on modelling the multivariate ordinal response using a multivariate generalised linear model. McCullagh's 1980 paper has been highly influential in advancing this type of model's use. In his paper McCullagh states that 'An appealing requirement for any model is that it should apply under varying conditions and should, as far as possible, be consistent with known physical or biological laws. This means, for example, that to measure the difference between two proportions, the logistic scale is preferable to the probability scale since a constant difference is a logical possibility on the logistic scale but is logically impossible on the probability scale.'

A general property of all log-linear models that do not use scores is that they are permutation invariant. A cumulative logit model has this and the above properties. It often results in simpler interpretations than some of the procedures listed above and has potentially greater power than ordinary logit models.

3.2 Cumulative Logit Models for Ordinal Responses

If there are J response categories with $J \geq 2$, there are many ways of forming logits (A. Agresti, 1984). For example, for a polytomous variable having response probabilities $(\pi_1, ..., \pi_J)$ at a certain combination of levels of explanatory variables a conditional logit could be formed.

$$\log\left(\frac{\pi_j \mid \pi_j + \pi_k}{\pi_k \mid \pi_j + \pi_k}\right) = \log\left(\frac{\pi_j}{\pi_k}\right) \qquad (3.2.1)$$

This is the log odds of classification in category j instead of category k, given that an observation falls in one of those two categories. When the response categories are ordered it makes sense to work with the cumulative response probabilities. The cumulative probabilities can be defined as the probabilities that the response Y falls in the category j or below, for each possible j where j = 1,...,J (J = 4 in the present study.) Let $P(Y \leq j)$ be the probability that Y belongs to a category less than or equal to j, then the jth cumulative probability is given by

$$P(Y \leq j) = \pi_1 + \cdots + \pi_j, \qquad j = 1,\ldots,J. \quad (3.2.2)$$

The cumulative probabilities reflect the ordering of the categories, with $P(Y \leq 1) \leq P(Y \leq 2) \leq \ldots \leq P(Y \leq J) = 1$. This cumulative probability model does not include the final probability, $P(Y \leq J)$, since it necessarily equals 1. The logits of the first J - 1 cumulative probabilities are formed in a way that takes account of the ordering of the categories.

$$\text{logit}[[P(Y \leq j)]] = \log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) \qquad j = 1,\ldots, J\text{-}1$$

$$= \log\left(\frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots + \pi_J}\right), \qquad (3.2.3)$$

Each cumulative logit uses all of the J response categories. For each cumulative logit the response categories collapse into two distinct categories, namely, those categories 1 to j and j + 1 to J so that the jth cumulative logit is effectively the binary response logit $P(Y \leq j)/ P(Y > j +1)$. In effect, the jth cumulative logit is the log odds that Y belongs to category j or below.

An ordinal model generally needs to provide structure for the J-1 cumulative logits. This was illustrated by Clayton (1974) in his use of the below model to calculate a common ratio θ of the odds for $x \leq x$, where x is a random variable measured in two populations. If, for samples from each of these populations, the frequency distributions are available stratified at the points $\{x_i\}(i = 1,\ldots,M)$ the data form a (M+1) x 2 contingency table, with ordered structure in the row classification. Let $F_1(x)$ and $F_2(x)$ be the cumulative distribution functions for x then the model proposed by Clayton (1974) is

$$\log\left(\frac{F_1(x_i)}{1-F_1(x_i)}\right)=\log\theta+\log\left(\frac{F_2(x_i)}{1-F_2(x_i)}\right) \qquad i=1,\ldots,M$$

(3.2.3)

In his paper Clayton showed that the above model assumes that for each of the M possible ways of collapsing the table into a 2 x 2 table the odds ratio is the same proportionality constant between corresponding logits. This is a fundamental feature of the proportional odds model that is desirable in modelling complex structured explanatory variables since the results are easily interpreted.

For the tree cut-stump study, J = 4 and so the chosen model will refer to $\log[\pi_1/(\pi_2+\pi_3+\pi_4)]$, $\log[\pi_1+\pi_2/(\pi_3+\pi_4)]$ and $\log[(\pi_1+\pi_2+\pi_3)/\pi_4]$.

3.3 Proportional Odds Model

Let **X** be a predictor, then the model

$$\text{logit}[P(Y\leq j)]=\alpha_j+\beta^{T}\mathbf{x} \qquad j=1,\ldots,J-1 \qquad (3.3.1)$$

has a linear parameter β describing the effect of **x** on the log odds of response Y belonging to category j or below. This model assumes **X** has an identical effect for all the J-1 collapsings of the response into binary outcomes. In other words the linear equations are parallel.

Interpretations for this model are widely referenced and refer to the *odds ratios* for the collapsed response scale, for any fixed j. Agresti (1996) points out that for two values $\mathbf{x}_1$ and $\mathbf{x}_2$, the odds ratio utilises the cumulative probabilities and their complements,

$$\frac{P(Y\leq j\mid \mathbf{x}_2)/P(Y>j\mid \mathbf{x}_2)}{P(Y\leq j\mid \mathbf{x}_1)/P(Y>j\mid \mathbf{x}_1)}. \qquad (3.3.2)$$

The log of this odds ratio is the difference between the cumulative logits at those two values of **X**. This equals $\beta(\mathbf{x}_2-\mathbf{x}_1)$, which is proportional to the distance between the **X** values. The same proportionality constant (β) applies for each possible point j for the collapsing. That is, the odds ratios depend only on the difference ($\mathbf{x}_2-\mathbf{x}_1$), but are independent of the choice of category. Note that this is a strong restriction which has to be relaxed in models for interval scales. (McCullagh & Nelder, 1989; pp155)

An interesting feature of the proportional odds model is the way in which the response probabilities vary with **x** for the single variable case in which $\beta > 0$. As **x** increases, the response on Y is more likely to fall at the low end of the ordinal scale. The probability for the highest numbered category decreases with **x**. For intermediate categories the probability decreases with **x** up to a certain point and thereafter increases. Over certain ranges of **x**, the probability for some of the intermediate categories is almost constant: over the same range the probabilities for the extreme categories may change quite appreciably.

The proportional odds model is appealing with respect to the context of killing tree stumps as it allows for modelling of the multivariate response data without loss of information, with potentially greater power and simpler interpretations through the use of linear predictors. The proportional odds model outlined above can be adjusted to allow for the incorporation of the design structure of the experiment in addition to the continuous variates in the tree stumps data. So the proportional odds model for the killing tree stumps data can now be written as

$$\log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) = \alpha_j + \beta^{T}\mathbf{x} \qquad j = 1, \ldots, 3 \quad (3.3.3)$$

where β^{T} is the vector of unknown parameters associated with the covariate vector **x**. Explanatory variables in the cumulative logit models can be continuous, categorical, or of both types, which is clearly suited to the killing of cut-stumps study as both types are prominent.

3.4 Generalised Linear Models

The proportional odds model is just one of a collection of models otherwise known as generalised linear models. In practical terms generalised linear models extend linear models to allow the random component to have a distribution other than Normal, and to allow for modelling some function of the mean. Broadly speaking, generalised linear models accommodate both non-normal response distributions and transformations to linearity. Firth (1991) explains that this extension yields a class of models of the form

$$f_Y(y) = f(y; \beta^T \mathbf{x}) \qquad (3.4.1)$$

in which $\mathbf{x}$ continues to appear only through the linear predictor, $\eta = \beta^T \mathbf{x}$, and where $f_Y(y)$ is the density of the response variable Y. Nelder and Wedderburn (1972) first identified this class of generalised linear models and provided the following theory. Let $E(Y) = \mu$, then μ is determined by η and so

$$g(\mu) = \eta \qquad (3.4.2)$$

where g is the link function between the mean and linear predictor. If f is a density function or probability function that belongs to the exponential family, then parameter estimates based on the likelihood enjoy special properties. One of these properties is that they satisfy enough regularity conditions to ensure that the global maximum of the log likelihood function $\ell(\theta;\ \mathbf{y})$ is given uniquely by the solution of the equations $\partial\ell / \partial\theta = 0$ or equivalently $\partial\ell / \partial\beta = 0$. Particular special instances of generalised linear models with these properties are the probit and logit link functions for binomial distributions. Note that Section 3.3. has centred attention on the proportional odds model using the logit link function.

The following general formulation for generalised linear models is based on N independent observations. This theory can be extended to the multinomial response where the nature of the response categories are non-independent and will be dealt with later in this chapter. Let

$$E(Y) = \mu(\beta_1, \ldots, \beta_p) \qquad (3.4.3)$$

be the parametric model for the mean of a response Y, where $\beta_1, \ldots, \beta_p$ ($p<N$) are unknown parameters and $\mu(\cdot)$ is a known function. The model is said to be linear if $\mu(\cdot)$ is a linear function of $\beta_1, \ldots, \beta_p$,

$$\mu = \sum_{r=1}^{p} x_r \beta_r \qquad (3.4.4)$$

for some explanatory variables $x_1, \ldots, x_p$ associated with the response Y.
A generalised linear model is of the form

$$\mu = g^{-1}\left(\sum_{r=1}^{p} x_r \beta_r\right) \tag{3.4.5}$$

where $g(\cdot)$ is a one-to-one differentiable function and is termed the link function. The link function $g(\mu)$ maps the response interval on to $(-\infty,\infty)$, or if the response interval is restricted to positive values only $g(\mu)$ maps $(0,\infty)$ to the whole real line. In particular, the logit link function maps the interval $(0,1)$ to $(-\infty,\infty)$. In the case where the link function is the inverse of the distribution function of μ, simplification of the log-likelihood function allows for a general formula that can be applied to a number of the exponential distribution families.

A general example is as follows. Assuming independence in N observations, $y_1,\ldots,y_N$, where each y_i is from an exponential family with parameters (θ_i,ϕ) (where ϕ is the dispersion parameter), the log-likelihood for the sample is

$$l = \sum_{i=1}^{N} \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi), \tag{3.4.6}$$

(the function $a(\phi)$ has the form $a(\phi) = \phi/\omega_i$ for known weight ω_i)

and so let

$$g(\mu_i) = g\{b'(\theta_i)\} = \sum_{r=1}^{p} x_{ir}\beta_r = \mathbf{x}_i^{\mathrm{T}}\beta = \eta_i \qquad i = 1,\ldots,N \tag{3.4.7}$$

be the structure imposed on the parameters, θ_i, through a generalised linear model. If $g(\cdot)$ is the canonical link function then the likelihood for the regression parameters may be generalised to the following formula

$$\sum_{r=1}^{p} \beta_r \sum_{i=1}^{N} \frac{y_i x_{ir}}{a(\phi)} - \sum_{i=1}^{N} \left\{ \frac{b(\theta_i)}{a(\phi)} - c(y_i,\phi) \right\}. \tag{3.4.8}$$

This general form along with the known quantity ϕ leads to further simplification of the estimating equations for the parameters $\beta_1,\ldots,\beta_p$.

The likelihood functions for some of the most common families of distributions can be written in the general form

$$L(\theta,\phi;\mathbf{y}) = \exp(\frac{\mathbf{y}^{\mathrm{T}}\theta - b(\theta)}{\phi} - c(\mathbf{y},\phi)) \qquad (3.4.9))$$

for a single observation y. If the value of ϕ is known, the family is referred to as a linear exponential family indexed by the natural parameter θ.

It is appropriate at this point to illustrate that the Multinomial distribution can be written in the exponential family form, so for example

$$f_y(\mathbf{y}) = P(Y_1 = y_1,...Y_J = y_J) = \binom{m}{\mathbf{y}} \pi_1^{y_1}...\pi_J^{y_J}, \qquad (3.4.10)$$

(where J is the number of attributes or response categories.), can be written as

$$f_y(\mathbf{y}) = \exp((y_1 \log(\pi_1) + ... + y_J \log(\pi_J) + \log\binom{m}{\mathbf{y}}). \quad (3.4.11)$$

where ***y*** is the vector of responses and

$$\binom{m}{\mathbf{y}} = \frac{m!}{y_1!,...,y_J!}$$

which, by comparison with equation (3.4.9) is seen to belong to the exponential family,

$$f(y_i;\theta_i,\phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i,\phi)\} \qquad (3.4.12)$$

where i in this instance refers to the *ith* observation vector and where $\theta = \log(\pi_1),...,\log(\pi_J)$, $c(y_i,\phi) = m!/y_1!...y_J!$, $b(\theta) = 0$ and $a(\phi) = 1$.

3.5 Maximum Likelihood

The maximum likelihood method operates through the likelihood equation, which is obtained by differentiating the log-likelihood function for the sample. When the log-likelihood is of the linear exponential form some elementary properties follow

$$\begin{aligned} &E(\partial\ell/\partial\theta) = 0 \\ -&E(\partial^2\ell/\partial\theta^2) = E(\partial\ell/\partial\theta)^2 \end{aligned} \qquad (3.5.1)$$

where $\ell = \log \mathrm{L}\ (\theta,\phi;y)$. These properties, although not restricted to the exponential form but hold in general, are used in maximising the likelihood equation. By setting the likelihood equation equal to zero and finding the minimal sufficient statistics for the unknown parameters which satisfy (3.5.1), these minimal sufficient statistics when substituted in (3.5.2) maximise the log-likelihood function.

For N independent observations, $Y_1,\ldots,Y_N$, the log likelihood is

$$\ell(\theta_i,\phi;y_i)=\sum_i[y_i\theta_i-b(\theta_i)]/a(\phi)+c(y_i,\phi). \qquad (3.5.2)$$

i = 1,…,N. It follows then, that the first and second derivatives of this log-likelihood are

$$\partial\ell/\partial\theta_i=[y_i-b'(\theta_i)]/a(\phi), \text{ and}$$

$$\partial^2\ell/\partial\theta_i^2=-b''(\theta_i)/a(\phi) \qquad (3.5.3)$$

respectively, where $b'(\theta_i)$ and $b''(\theta_i)$ denote the first two derivatives of b evaluated at θ_i. In light of the above, equations (3.5.1) imply

$$E(Y_i)=\mu_i=b'(\theta_i),$$

and $$\text{var}(Y_i)=b''(\theta_i)a(\phi). \qquad (3.5.4)$$

To obtain the likelihood equations, it is first important to recognise that θ depends on the model parameter β, such that

$$\ell(\beta)=\sum_{i=1}^{N}\log f(y_i;\theta_i,\phi)=\sum_{i=1}^{N}\ell_i,$$

where $\ell_i=\ell(\theta_i,\phi;y_i)$; i = 1,…,N, and so the likelihood equations are given by

$$\frac{\partial\ell_i}{\partial\beta_r}=\frac{\partial\ell_i}{\partial\theta_i}\frac{\partial\theta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_r}. \qquad (3.5.5)$$

Since $\partial\ell/\partial\theta_i=[y_i-b'(\theta_i)]/a(\phi)$, and since $\mu_i=b'(\theta_i)$ and $\text{var}(Y_i)=b''(\theta_i)a(\phi)$,

$$\partial\ell/\partial\theta_i=[y_i-\mu_i]/a(\phi),$$

$$\partial\mu_i/\partial\theta_i=b''(\theta_i)=\text{var}(Y_i)/a(\phi) \qquad (3.5.6)$$

Also, given (3.4.7) where $\eta_i=\sum_{r=1}^{p}x_{ir}\beta_r$, where r = 1,..p and p<N is the number of model parameters,

$$\frac{\partial\eta_i}{\partial\beta_r}=x_{ir}. \qquad (3.5.7)$$

Finally, since $g(\mu_i) = \eta_i$, $\partial\mu_i / \partial\eta_i$ depends on the link function of the model. So substituting these results in (3.5.5) yields,

$$\frac{\partial \ell_i}{\partial \beta_r} = \frac{(y_i - \mu_i)}{a(\phi)} \frac{a(\phi)}{\text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ir},$$

which reduces to

$$\frac{\partial \ell_i}{\partial \beta_r} = \frac{(y_i - \mu_i) x_{ir}}{\text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i}, \tag{3.5.8}$$

and so the likelihood equations are given by

$$\sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ir}}{\text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \qquad r = 1,\ldots,p \tag{3.5.9}$$

In general the likelihood equations are non-linear and they have to be solved by numerical iteration methods. Dobson (1990, p40) illustrates the use of the Newton-Raphson method. The mth approximation to the true parameters β is given by

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} - \left[\frac{\partial^2 \ell(\beta)}{\partial \beta_r \partial \beta_s}\right]^{-1}_{\beta = \mathbf{b}^{(m-1)}} \Psi^{(m-1)} \tag{3.5.10}$$

where $\left[\frac{\partial^2 \ell(\beta)}{\partial \beta_r \partial \beta_s}\right]^{-1}_{\beta = \mathbf{b}^{(m-1)}}$

is the matrix of second derivatives of $\ell(\beta)$ evaluated at $\beta = \mathbf{b}^{(m-1)}$ and $\Psi^{(m-1)}$ is the vector of first derivatives $\partial \ell(\beta) / \partial \beta_r$ evaluated at $\beta = \mathbf{b}^{(m-1)}$. The Newton-Raphson method can be simplified by using the method of scoring. Effectively this is the Newton-Raphson method with matrix of second derivatives in (3.5.10) replaced with the matrix of expected values

$$E\left[\frac{\partial^2 \ell_i}{\partial \beta_r \partial \beta_s}\right] = -E\left[\frac{\partial \ell_i}{\partial \beta_r \partial \beta_s}\right]$$

$$= -E\left[\frac{(Y_i - \mu_i) x_{ir}}{\text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{(Y_i - \mu_i) x_{is}}{\text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i}\right]$$

$$= \frac{-x_{ir} x_{is}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

and so

$$E\left[\frac{\partial^2 \ell(\beta)}{\partial \beta_r \partial \beta_s}\right] = -\sum_{i=1}^{N} \frac{x_{ir} x_{is}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) \tag{3.5.11}$$

Equation (3.5.11) is also known as the information matrix **Inf**. The information matrix can be written in matrix notation so that

$$\mathbf{Inf}=\mathbf{X}^{T}\mathbf{W}\mathbf{X} \tag{3.5.12}$$

where **W** is the diagonal N by N matrix with elements

$$w_i = (\partial\mu_i / \partial\eta_i)^2 / \mathrm{var}(Y_i) \tag{3.5.13}$$

and **X** is the N x p design matrix.

The actual iterative method used to fit generalised linear models is called Fisher Scoring. The initial values for the weights w_i are calculated by the initial linear predictor, at iteration m a new approximation to the estimate of β is found by a weighted regression of 'working' values z_i on X with weights w_i. To identify these working values consider the Fisher Scoring algorithm with the mth approximation to the true parameters β

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + (\mathbf{X}^T\hat{\mathbf{W}}^{(m-1)}\mathbf{X})^{-1}\mathbf{q}^{(m-1)}, \text{ or}$$

$$(\mathbf{X}^T\hat{\mathbf{W}}^{(m-1)}\mathbf{X})\mathbf{b}^{(m)} = (\mathbf{X}^T\hat{\mathbf{W}}^{(m-1)}\mathbf{X})\mathbf{b}^{(m-1)} + \mathbf{q}^{(m-1)} \tag{3.5.14}$$

where $(\mathbf{X}^T\hat{\mathbf{W}}^{(m-1)}\mathbf{X})$ is the m-1 approximation for the estimated information matrix. The right hand side of (3.5.14) is the vector having elements

$$\sum_{r=1}^{p}\left[\sum_{i=1}^{N}\frac{x_{ir}x_{is}}{\mathrm{var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\beta_r^{(m-1)}\right] + \sum\frac{(y_i-\mu_i^{(m-1)})}{\mathrm{var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right).$$

Therefore, (3.5.15)

$$(\mathbf{X}^T\hat{\mathbf{W}}^{(m-1)}\mathbf{X})\mathbf{b}^{(m-1)} + \mathbf{q}^{(m-1)} = \mathbf{X}^T\hat{\mathbf{W}}^{(m-1)}\hat{\mathbf{z}}^{(m-1)}$$

$$(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})\mathbf{b}^{(m)} = \mathbf{X}^T\hat{\mathbf{W}}^{(m-1)}\hat{\mathbf{z}}^{(m-1)} \tag{3.5.16}$$

where $\hat{\mathbf{z}}^{(m-1)}$ has been defined as $\hat{\mathbf{z}}^{(m-1)} = \boldsymbol{X}\mathbf{b}^{(m)}$, which has elements

$$z_i^{(m-1)} = \sum_{r=1}^{p}x_{ir}\beta_r^{(m-1)} + (y_i - \mu_i^{(m-1)})\left(\frac{\partial\eta_i^{(m-1)}}{\partial\mu_i^{(m-1)}}\right)$$

$$z_i^{(m-1)} = \eta_i^{(m-1)} + (y_i - \mu_i^{(m-1)})\left(\frac{\partial\eta_i^{(m-1)}}{\partial\mu_i^{(m-1)}}\right) \tag{3.5.17}$$

Finally, (3.5.16) can be arranged to find the form of the Fisher Scoring equations for generalised linear models

$$\hat{\mathbf{b}}^{(m)} = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{W}}^{(m-1)}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{W}}^{(m-1)}\hat{\mathbf{z}}^{(m-1)}, \qquad (3.5.18)$$

The vector $\mathbf{z}$ is known as the working variable because at each mth cycle of the iterative method $\hat{\mathbf{z}}^{(m-1)}$ is regressed on $\mathbf{X}$ with weight $\mathbf{W}^{(m-1)}$ to obtain a new estimate of $\hat{\mathbf{b}}^{(m)}$. This provides a new linear predictor $\hat{\eta}^{(m)} = \mathbf{X}\hat{\mathbf{b}}^{(m-1)}$ and $\hat{\mu}^{(m)} = g^{-1}(\hat{\eta}^{(m)})$. The maximum likelihood estimator is the limit of $\hat{\mathbf{b}}^{(m)}$ as m tends to infinity. The iteration procedure continues producing new values until a suitable convergence criterion is satisfied. This procedure is more generally termed iterative weighted least squares (IWLS).

A neat result of the above procedure is that the asymptotic covariance matrix of $\hat{\beta}$ is the inverse of the information matrix, estimated by

$$\hat{Cov}(\hat{\beta}) = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{W}}^{(m)}\mathbf{X})^{-1} \qquad (3.5.19)$$

where $(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{W}}^{(t)}\mathbf{X})$ is a by-product of the final IWLS iteration and $\hat{\beta}$ is $\hat{\mathbf{b}}^{(m)}$ at convergence.

Finally, the IWLS scheme depends on the response distribution only through its mean and variance functions which is why it is fairly straight forward to extend the generalised linear modelling theory to include multi-response models.

3.6 The Multinomial Distribution

In the tree stumps data the tree stumps possess one and only one of the k attributes $A_1,\ldots,A_J$, namely, killed after the first application of herbicide, killed after second application and so on, up to the kth application of herbicide. Theoretically (McCullagh & Nelder, 1989), if the population is infinitely large and a simple random sample of size m is taken, the probability of an observation to have been killed after the jth application of herbicide is given by

$$pr(Y_1 = y_1,\ldots,Y_J = y_J; m,\pi) = \frac{m!}{y_1!y_2!\ldots y_J!}\pi_1^{y_1}\ldots\pi_J^{y_J}, \qquad (3.6.1)$$

otherwise known as the multinomial distribution. The probability vector π is related to explanatory variables, through a link function. In this case the cumulative probabilities are modelled using the logit link function. As was shown in section 3.4, the multinomial model is a member of the linear exponential family. The variance function of the multinomial distribution is now the covariance matrix Σ with elements

$$\Sigma_{rs} = \begin{cases} m\pi_i(1-\pi_i) \ldots\ldots\ldots\ldots\ldots \mathrm{i}=\mathrm{j} \\ -m\pi_i\pi_j \ldots\ldots\ldots\ldots\ldots\ldots i \neq \mathrm{j} \end{cases}, \qquad (3.6.2)$$

and the expectation is given by

$$E(Y_i) = m\pi_i = \mu_i = \frac{m.\exp(\eta_i)}{1+\exp(\eta_i)}, \qquad (3.6.3)$$

for the proportional odds model with the logit link function. Therefore, it follows that the variance of Y_i can be written as

$$\mathrm{var}(Y_i) = m\pi_i(1-\pi_i) = \frac{m.\exp(\eta_i)}{1+\exp(\eta_i)}\left(1-\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right). \qquad (3.6.4)$$

To obtain the likelihood functions, consider the following. Each stool Y_i, say, is independently observed of another and each Y_i falls in one and only one of the J categories. However, due to the presence of the continuous predictor in the linear component of the proportional odds model, the probability with which Y_i can be expected to be in the jth category differs with each $Y_i \mid x_i$. To clarify this further suppose that $\mathbf{Y}_i$ is the vector $\left(y_{1(i)},\ldots,y_{J(i)}\right)$ where $y_{j(i)} = 1$ if Y_i falls in category j and $y_{j(i)} = 0$, then

$$\mathbf{Y}_i \sim Mu(1,\pi_{1(i)},\pi_{2(i)},\ldots,\pi_{J(i)})$$

where $m = 1$ and $\pi_{j(i)}$ indicates the different category probabilities for the ith observation given x_i.

Suppose that the ith observation lies in category j(i) such that $\mathbf{y}_i = (0,1,0,\ldots,0)$, e.g. $j(i) = 2$, then

$$Lik(\pi_i;\mathbf{y}_i) = \frac{1!}{1!0!\ldots0!}\pi_1^{y_{1(i)}},\ldots,\pi_J^{y_{J(i)}} \qquad (3.6.5)$$

$$= \pi_{2(i)}\,.$$

which for the proportional odds model is equivalent to

$$= \exp(\alpha_{2(i)} + \beta^T\mathbf{x})\,.$$

In general, the likelihood is proportional to

$$Lik(\boldsymbol{\pi}_i ; \mathbf{y}_i) \propto \prod_{j=1}^{n} \pi_{j(i)}^{y_{j(i)}} . \tag{3.6.6}$$

(Note that the $y_{j(i)}$ is zero except for the category in which the ith observation falls.) Then from the ith observation $\mathbf{y}_i$, the contribution to the log likelihood is

$$\ell(\pi_i ; \mathbf{y}_i) = \sum_j y_{j(i)} \log \pi_{j(i)} \qquad j = 1,\ldots,J \tag{3.6.7}$$

The total log likelihood is the sum of contributions one from each of the n observations. Thus,

$$\ell(\pi;\mathbf{y}) = \sum_{ij} y_{j(i)} \log \pi_{j(i)} \tag{3.6.8}$$

Differentiation of the log-likelihood with respect to $\pi_{j(i)}$ subject to the constraints $\sum_j \pi_{j(i)} = 1$ and $\sum_j y_{j(i)} = m_i$ where $m_i = 1$ in this case, gives

$$\frac{\partial \ell(\pi;\mathbf{y})}{\partial \pi_{j(i)}} = \frac{y_{j(i)} - \pi_{j(i)}}{\pi_{j(i)}} . \tag{3.6.9}$$

At first glance, it is not immediately obvious how McCullagh & Nelder (1989, pp171) arrive at equation (3.6.9). However, it can be seen that differentiating (3.6.8) with respect to $\pi_{j(i)}$, subject to the constraints, yields equations equivalent to (3.6.9) at the maximum likelihood estimators of $\pi_{j(i)}$. Let

$$L_0 = \ell(\pi;\mathbf{y}) = \sum_{ij} y_{j(i)} \log \pi_{j(i)} , \tag{3.6.10}$$

be a function of the same form as the likelihood but without the constraints and let $L = L_0$ with the constraint $\sum_j \pi_{j(i)} = 1$. Then L is a function of

$\pi_{1(i)},\ldots,\pi_{J-1(i)}$ for all i, plus

$$\pi_{J(i)} = 1 - \sum_{j=1}^{J-1} \pi_{j(i)} \text{ and where } \frac{\partial \pi_{J(i)}}{\partial \pi_{j(i)}} = -1 . \tag{3.6.11}$$

Therefore using the chain rule,

$$\begin{aligned} \frac{\partial L}{\partial \pi_{j(i)}} &= \frac{\partial L_0}{\partial \pi_{j(i)}} + \frac{\partial L_0}{\partial \pi_{J(i)}} \frac{\partial \pi_{J(i)}}{\partial \pi_{j(i)}} \\ &= \frac{\partial L_0}{\partial \pi_{j(i)}} - \frac{\partial L_0}{\partial \pi_{J(i)}} . \end{aligned} \tag{3.6.12}$$

Differentiating L_0 with respect to $\pi_{j(i)}$ yields

$$\frac{\partial L_0}{\partial \pi_{j(i)}} = \frac{y_{j(i)}}{\pi_{j(i)}} \tag{3.6.13}$$

such that

$$\frac{\partial L}{\partial \pi_{j(i)}} = \frac{y_{j(i)}}{\pi_{j(i)}} - \frac{y_{J(i)}}{\pi_{J(i)}} \tag{3.6.14}$$

It is at this point that McCullagh & Nelder set $\pi_{J(i)}$, in the second term to the maximum likelihood estimate y.

$$y_{J(i)} = m_i \hat{\pi}_{J(i)} \tag{3.6.15}$$

and so

$$\frac{\partial L_0}{\partial \pi_{j(i)}} = \frac{y_{j(i)}}{\pi_{j(i)}} - m_i$$

$$= \frac{y_{j(i)} - m_i \pi_{j(i)}}{\pi_{j(i)}} \tag{3.6.16}$$

which is equivalent to equation (3.6.9). So as has been shown, it is possible using the above method to derive the maximum likelihood equation from the log-likelihood. Equation (3.6.9) follows from 3.6.14 if the second (and only the second) term is replaced by the maximum likelihood term. This seems odd at first sight. To understand the implications and justification of this approximation, let us look at an alternative derivation, using the formalism of Lagrange multipliers.

The constrained maximum may be found using the method of Lagrange multipliers (Finney & Thomas, 1990). In general terms, this method states that the extreme values of a function $f(x, y)$ whose variables are subject to a constraint $g(x, y) = 0$, are found to be on the surface $g = 0$ at the points where

$$\nabla f = \lambda \nabla g \tag{3.6.17}$$

for some scalar λ (called a Lagrange multiplier) and where ∇f and ∇g are the derivatives of $f(x, y)$ and $g(x, y)$ respectively.

Using this method as a general framework for finding the maximum likelihood equation and given (3.6.10)

$$L_0 = \ell(\pi;\mathbf{y}) = \sum_{ij} y_{j(i)} \log \pi_{j(i)}$$

and

$$C_i = 0. \tag{3.6.18}$$

Now define,

$$L = L_0 - \sum_i \lambda_i C_i. \tag{3.6.19}$$

Note that if equation (3.6.18) is satisfied, the second term disappears, and L becomes equal to the log likelihood. That is,

$$\frac{\partial L}{\partial \pi_{j(i)}} = \frac{\partial L_0}{\partial \pi_{j(i)}} - \lambda_i \frac{\partial C}{\partial \pi_{j(i)}} = 0 \tag{3.6.20}$$

and finding the values of λ_i, $\pi_{j(i)}$ and $y_{j(i)}$ that satisfy

$$\frac{\partial L_0}{\partial \pi_{j(i)}} = \lambda_i \frac{\partial C}{\partial \pi_{j(i)}} \tag{3.6.21}$$

and

$$C = \sum_j \pi_{j(i)} - 1 = 0.$$

Now,

$$\frac{\partial L}{\partial \pi_{j(i)}} = \frac{\partial L_0}{\partial \pi_{j(i)}} - \lambda_i$$

$$= \frac{y_{j(i)}}{\pi_{j(i)}} - \lambda_i, \tag{3.6.22}$$

implying that $\lambda_i = y_{j(i)} / \pi_{j(i)}$ or $\pi_{j(i)} = y_{j(i)} / \lambda_i$. That is, the Lagrange multipliers which are determined from the constraint equation (3.6.18). By setting (3.6.22) equal to zero, solving for λ_i, and inserting the resulting expression into equation (3.6.18) this leads to

$$\frac{1}{\lambda_i} \sum_j y_{j(i)} = 1 \tag{3.6.23}$$

which along with

$$\sum_j y_{j(i)} = m_i, \tag{3.6.24}$$

yields

$$\frac{1}{\lambda_i} \sum_j y_{j(i)} = \frac{m_i}{\lambda_i} = 1 \tag{3.6.25}$$

$$\Rightarrow \lambda_i = m_i . \tag{3.6.26}$$

Thus,

$$\frac{\partial L}{\partial \pi_{j(i)}} = \frac{y_{j(i)}}{\pi_{j(i)}} - m_i . \tag{3.6.27}$$

which is exactly (3.6.16) and reduces to (3.6.9) when $m_i = 1$.

Now, consider the parameter likelihood equations for the proportional odds model, where for neatness $\gamma_{j(i)} = \pi_{1(i)} + \ldots + \pi_{j(i)}$ then our proportional odds model is of the general form

$$\log it(\gamma_{j(i)}) = \sum_r x_{j(i)r} \beta_r \tag{3.6.28}$$

where $x_{j(i)r}$ are the components of the design matrix X of order nJ x p^* where $p^* = p + k - 1$ and $\beta = (\alpha_1, \ldots, \alpha_{k-1}, \beta_1, \ldots \beta_p)$. The (i,j) row of $\mathbf{X}$ has components $(0,\ldots,1,\ldots,0,\mathbf{x}_i)$ with the unit value in position j and the *i*th block of J-1 rows is

$$[\mathbf{I}_{J-1} : \mathbf{1x}_i]$$

Differentiation with respect to β yields

$$\frac{\partial \ell}{\partial \beta_r} = \sum_{ij} \frac{\partial \ell}{\partial \gamma_{j(i)}} \frac{\partial \gamma_{j(i)}}{\partial \beta_r}$$

$$= \sum_{ij} x_{j(i)} \gamma_{j(i)} (1 - \gamma_{j(i)}) \frac{\partial \ell}{\partial \gamma_{j(i)}} \tag{3.6.29}$$

where

$$\frac{\partial \ell}{\partial \gamma_{j(i)}} = \frac{y_{j(i)} - \pi_{j(i)}}{\pi_{j(i)}} - \frac{y_{j-1(i)} - \pi_{j-1(i)}}{\pi_{j-1(i)}} . \tag{3.6.30}$$

The maximum likelihood estimates are obtained by setting the above equation equal to zero and using the iterative weighted least squares procedure.

In the above illustration the category probabilities are for the *i*th observation are conditional on the corresponding *i*th value of x, thus the conditional expectation of $\boldsymbol{Y}_i$ has changed and can be determined by the following equation.

$$E(Y_i \mid x) = P(Y = 1; \mathbf{x}) + 2 * P(Y = 2; \mathbf{x}) + 3 * P(Y = 3; \mathbf{x}) + 4 * P(Y = 4; \mathbf{x})$$

which for the proportional odds model is equivalent to (3.6.31)

$$E(Y_i \mid x) = 4 - \frac{\exp(\alpha_{1(i)} + \beta^T \mathbf{x})}{1 + \exp(\alpha_{1(i)} + \beta^T \mathbf{x})} - \frac{\exp(\alpha_{2(i)} + \beta^T \mathbf{x})}{1 + \exp(\alpha_{2(i)} + \beta^T \mathbf{x})} - \frac{\exp(\alpha_{3(i)} + \beta^T \mathbf{x})}{1 + \exp(\alpha_{3(i)} + \beta^T \mathbf{x})} \qquad (3.6.32)$$

To summarise, the method described above comprises the iterative weighted least squares for a multivariate generalised linear model using the Newton-Raphson method with Fisher Scoring as described in section 3.5. (McCullagh,1980) gives further details of the general fitting methods for multivariate generalised linear models in his Appendix.

CHAPTER 4

RESULTS OF GENERALISED LINEAR MODEL ANALYSIS

4.1 Overview

This chapter describes the results obtained from fitting a generalised linear model to the tree stump data. Data from 629 tree stumps, from each of which records of the stump's dimensions were taken (See Chapter 2.), and assessed as covariates in addition to the original treatment factors in the generalised linear model. These variates were used in a best subset procedure to select the best linear predictor equation. For reference, Table 4.1. lists the level numbers and descriptions for both the method of application and type of herbicide factors.

Table 4.1.

Reference table for factor levels

Method of application (Factor a)	Level Number
Cut surface	Level 1
Basal frill	Level 2
Foliar spray	Level 3
Type of herbicide (Factor b)	
Timbril	Level 1
Garlon	Level 2
Chopper	Level 3
Brush-off	Level 4
Nomix + Garlon	Level 5

N.B. Combinations of the two factors are indicated by levels (method*herbicide).

4.2 Linear Predictor

Consider again a linear model in which there are both continuous and qualitative variates present. Within the context of a generalised linear model, interest was focused on how the continuous variable affects the response variable in the presence of the qualitative factors. In practice a factor may influence the relationship between X and Y in various ways.

Consider the model in which the intercept varies with the jth factor level, but where the slope is constant over levels,

$$\eta = \alpha_j + \beta \mathbf{x} \qquad j = 1,\ldots,J \quad (4.2.1)$$

in such a case there is no interaction between the effects of **x** and the factor. If the slope does change with the factor level then the model is written as

$$\eta = \alpha_j + \beta_j \mathbf{x}. \qquad J = 1, \ .J \quad (4.2.2)$$

In this situation an interaction exists between the effects of **x** and the factor. The changing slope and intercept terms with the levels of factor means that there is a different linear predictor for the different levels of factor. In accordance with this result, modelling the tree stump data with continuous and qualitative variables in the model means that differing effects between the levels of a factor are measured by the coefficient in the linear predictor.

To maintain the original factorial design of the experiment and incorporate the inclusion of covariates, the linear predictor for the cut stump data was of the form

$$\eta_{jm\ell}^{(k)} = \alpha_j + Q\mathbf{X} + F \qquad (4.2.3)$$

where $\eta_{jm\ell}^{(k)}$ is the kth possible linear predictor, $k = 1,\ldots,10$ (see equation 4.2.5); $j = 1,\ldots,3$; $m = 1,\ldots,3$; $\ell = 1,\ldots,5$ and where α_j is the response factor that represents the J-1 collapsings of the total number of applications into binary outcomes (kill or survival). Since the purpose of this analysis was to explore the relevance of a stool's physical dimensions in predicting the number of applications until a kill is achieved, the best subset approach applied to finding the best linear predictor involved only those continuous variates representing the stool dimensions. To make the linear predictor useful for predictive purposes the model should include as many variables as necessary to keep bias errors small, so that reliable values can be determined. In addition to the continuous variates being assessed, the nature of the design of the experiment dictated that method, herbicide and the interaction term method*herbicide be included in the equation. In equation 4.2.3 , Q represents the matrix of coefficients for the possible subsets of the continuous explanatory variates

$\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$, where $\mathbf{x}_0,..., \mathbf{x}_3$ denotes the variables stool diameter, sum of cut surface diameters, highest height and ratio of stool diameter to sum of cut surface diameters respectively,

$$Q = \begin{pmatrix} 0 & +0 & +0 & +0 \\ q_0 & +0 & +0 & +0 \\ 0 & +q_1 & +0 & +0 \\ 0 & +0 & +q_2 & +0 \\ q_0 & +q_1 & +0 & +0 \\ q_0 & +0 & +q_2 & +0 \\ 0 & +q_1 & +q_2 & +0 \\ q_0 & +q_1 & +q_2 & +0 \\ \hline 0 & +0 & +q_2 & +q_3 \\ 0 & +0 & +0 & +q_3 \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{pmatrix}. \tag{4.2.4}$$

Note that because $\mathbf{x}_3$ represents the ratio of stool diameter and sum of cut surface diameters it was considered on its own and in conjunction with highest height to avoid introducing collinearity into the model. Adopting this approach meant fitting ten possible linear predictors to the data. The choice of which linear predictor was best to use was then made by assessing the patterns observed.

For this model the quantities α_j provide quantification of the difference between successive categories on the scale of the logit function, however, a necessary restriction of the proportional odds model is that it must avoid obtaining probabilities that are negative and thus $\alpha_1 \leq \alpha_2 \leq \alpha_3$. The cumulative sum of all response category probabilities is equivalent to 1, so the number of cut-points is equal to J-1, or in the case of the tree stump data, three. This at first seems incomplete (since only three equations are estimated for four categories), but the probability of a tree being killed at the fourth application is easily found by calculating

$$P(Y = 4) = 1 - P(Y \leq 3).$$

In general, to find the probability that a tree stump is killed after the jth application for j = 1,...,4

$$P(Y = j) = P(Y \leq j) - P(Y \leq j-1), \qquad (4.2.5)$$

and where

$$P(Y \leq j) = \frac{\exp(\eta_{jm\ell}^{(k)})}{1+\exp(\eta_{jm\ell}^{(k)})} \qquad (4.2.5)$$

As stated previously, effects of the factors on the the log odds of success may vary among the levels of each factor in F,

$$F = \beta_{1m}a_m + \beta_{2\ell}b_\ell + \beta_{3m\ell}ab_{m\ell} \qquad (4.2.6)$$

this is represented by β_{1m} for factor a (method of application), where $m =$ 1,...,3; by $\beta_{2\ell}$ for factor b (type of herbicide) where $\ell = 1,\ldots,5$ and by $\beta_{3m\ell}$ for the interaction of factors a and b. In short, β_{1m} and $\beta_{2\ell}$ measure the difference between the effects of the first and the m or ℓ th levels of the corresponding factor. As a consequence of this the possible number of parameters used in the model is high, as potentially there is a different parameter to be estimated for each level of each factor. Even so, the parallel lines assumption, born from using the proportional odds model, is adhered to. The model constrains the J-1 response curves to have the same shape ensuring that the model cannot be fitted by forcing separate logit models for each cut-point. (See Chapter 3, 3.3)

Clearly model (4.2.3) is over parameterised due to the number of parameters being greater than the number of independent equations, because of this the linear predictors need to be constrained. Dobson (1990) illustrates in detail the style of re-parameterisation applied here known as corner-point constraints. The corner-point constraints for this model are:

$$\beta_{11} = 0,\ \beta_{21} = 0,\ \beta_{311} = 0,\ \beta_{313} = 0,\ \beta_{314} = 0,\ \beta_{315} = 0,$$
$$\beta_{312} = 0,\ \beta_{321} = 0,\ \beta_{331} = 0. \qquad (4.2.7)$$

These constraints, although different from the more usual sum-to-zero constraints, correspond to the same alternative hypothesis, namely, that the response mean for some level of a factor may differ, compared with the null hypothesis that the means are all equal. Thus, the parameters β, where

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix},$$

is estimated sequentially in such a way that the redundant corner-point parameters, i.e., those parameters set to zero (β_2), are identified as the baseline effects of their associated factors.

To illustrate this further consider

(1) $$\alpha_j + \beta_0 x \quad (4.2.8)$$

to be the baseline model for the tree stump data, where $\beta_0 x$ is a covariate and the unknown parameters for level one of each factor and factor combination, are taken to be the redundant corner-point parameters. To test the hypothesis that the response mean for level two of the factors are equal to the baseline level, the following formal test is conducted. So, without loss of generality, assume that the linear predictor is

$$\alpha_j + \beta_0 x + \beta_{12} a_2 + \beta_{22} b_2 + \beta_{322}(a_2 b_2) \quad (4.2.9)$$

the differences can then be tested using (4.2.8) and the models derived from (4.2.9) to give

(2) $$\alpha_j + \beta_0 x + \beta_{12} a_2$$

(3) $$\alpha_j + \beta_0 x + \beta_{12} a_2 + \beta_{22} b_2$$

(4) $$\alpha_j + \beta_0 x + \beta_{12} a_2 + \beta_{22} b_2 + \beta_{322}(a_2 b_2). \quad (4.2.10)$$

The null hypothesis in each case respectively, is:

H_A: Mean response for method of application, level 2 is the same as the mean response for the method of application, level 1. (i.e. mean difference equals zero.)

H_B: Mean response for herbicide, level 2 is the same as the mean response for herbicide, level 1

H_{AB}: Mean response for level 2 (method)*level 2 (herbicide) interaction is the same as the mean response for the level 1 (method)*level 1 (herbicide) interaction.

To find the differences between levels of method of application, type of herbicide and method by herbicide interactions, the above process was repeated systematically for each linear predictor fitted.

4.3. Log-Likelihood Ratio Statistic

Assessing the adequacy of the model of interest is done by comparing the likelihood under the proposed model with the likelihood under the maximal model. In general, the maximal model uses the same distribution and link function as the proposed model but the number of parameters in the maximal model is equal to the number of observations, the benefit of this is that the maximal model provides a complete description of the data. The likelihood functions for the maximal model and the model of interest are evaluated at the respective maximum likelihood estimates $\hat{\beta}_{max}$ and $\hat{\beta}$ to obtain values $L(\hat{\beta}_{max};\mathbf{y})$ and $L(\hat{\beta};\mathbf{y})$ respectively. If the model of interest describes the data well then $L(\hat{\beta};\mathbf{y})$ will be approximately equal to $L(\hat{\beta}_{max};\mathbf{y})$. Hence, this is used to produce a measure of goodness of fit of (4.2.3) to the tree stump data. Now, let

$$\lambda = \frac{L(\hat{\beta}_{max};\mathbf{y})}{L(\hat{\beta};\mathbf{y})} \text{ or equivalently } \log\lambda = l(\hat{\beta}_{max};\mathbf{y}) - l(\hat{\beta};\mathbf{y})$$

$l(\hat{\beta}_{max};\mathbf{y})$ and $l(\hat{\beta};\mathbf{y})$ are the log likelihood functions evaluated at their respective maximum likelihood estimates, so that large values of $\log\lambda$ suggest that the model of interest is a poor description of the data. The critical region for $\log\lambda$ is calculated using the test statistic

$$D = 2\log\lambda = 2[l(\hat{\beta}_{max};\mathbf{y}) - l(\hat{\beta};\mathbf{y})] \sim \chi^2_{N-p} \qquad (4.3.1)$$

which has the χ^2_{N-p} distribution and where N is the set of observations and p is the number of parameters to be estimated in the model of interest. The test statistic D is compared with the associated Chi-squared distribution. If the model is a good fit then D is expected to lie in the middle of the distribution. Therefore, if D does not lie in the middle of the distribution but is to be found in the upper 5% tails of the distribution it is concluded that there is sufficient evidence (at the 5% significance level) to reject the null hypothesis.

4.4 Fitting the model

An Ordinal Logistic Regression procedure is contained in the Minitab Software package. This procedure fits a regression model (proportional odds) to ordinal data using an underlying logistic distribution and a logit link function. Iterative weighted least squares was used to calculate the maximum likelihood estimates and standard errors for the unknown parameters. (See Chapter 3.) These are given in Table 4.5.4 on page 62. Interestingly, it is convention in other statistical packages to state the general proportional odds model as

$$\operatorname{logit}(P \leq j) = \theta_j - \beta^{\mathrm{T}}\mathbf{x} \tag{4.4.1}$$

which is different in sign to that of model (3.3.3). This has important implications for the interpretation of the results since the negative sign ensures that as $\mathbf{x}$ increases the probability of the response lying in the higher category also increases. However, the ordinal logistic regression procedure in Minitab uses (3.3.3) which conversely ensures that as $\mathbf{x}$ increases the probability that the response will lie in the higher category decreases.

The maximum likelihood parameter estimates are tested for equivalence to zero – in the case of the baseline levels (first levels) of each factor, the cut-points are calculated and the baseline parameters are taken to be equal to zero. That is, the response is modelled by the systematic component

$$\alpha_j + \beta_0\mathbf{x}. \tag{4.4.2}$$

where, for generality, $\beta_0\mathbf{x}$ represents the single continuous variable case.

To test

$$\mathrm{H}_0\text{: } \beta = 0 \tag{4.4.3}$$

for β symbolising any of the non-redundant factor parameters specified in F (see equation (4.2.6)), against

$$\mathrm{H}_1\text{: } \beta \neq 0 \tag{4.4.4}$$

the test statistic is:

$$z = \hat{\beta} / ASE \tag{4.4.5}$$

which has approximately a standard normal distribution when $\beta = 0$ and where ASE is the asymptotic (large sample Normal distribution) standard error, obtained from taking the inverse of the information matrix. z was referred to the standard Normal table to get two-sided P-values.

4.5. Results

To establish the best linear predictor each of the ten possible regressions were fitted to the data using the proportional odds model. As previously stated, the predictor variables here are $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$ representing stool diameter, sum of cut surface diameters, highest height and the ratio of stool diameter to sum of cut surface diameters, respectively. An α_j term and the treatment factors F are always included. Thus there are $k = 2^3 + 2 = 10$ possible linear predictors to be fitted. The first of these is the fit of

$$\eta_{jm\ell}^{(1)} = \alpha_j + F \qquad (4.5.1)$$

The other 9 fits appear in Appendix A. Table 4.5.1 looks at the significance of the subsets of $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$ in the model. The subsets are divided into three sets A, B and C.

Table 4.5.1

Significance /non significance for all possible subsets

(P-value given in brackets)

A (1-variable)	B (2-variable)	C (3-variable)
$\mathbf{x}_0$ (0.000)	$\mathbf{x}_0, \mathbf{x}_1$ (0.002; 0.390)	$\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$ (0.005; 0.731; 0.057)
$\mathbf{x}_1$ (0.002)	$\mathbf{x}_0, \mathbf{x}_2$ (0.001; 0.040)	-
$\mathbf{x}_2$ (0.001)	$\mathbf{x}_1, \mathbf{x}_2$ (0.042; 0.018)	-
$\mathbf{x}_3$ (0.801)	$\mathbf{x}_2, \mathbf{x}_3$ (0.001; 0.777)	-

Table 4.5.1 reveals that $\mathbf{x}_0$, $\mathbf{x}_1$, and $\mathbf{x}_2$ are all statistically significant in the model when entered as single covariates in equation 4.5.1. The addition of further variables to these shows a consistent pattern in the behaviour of $\mathbf{x}_1$. It can be seen that in the presence of $\mathbf{x}_0$, $\mathbf{x}_1$ is no longer significant in the regression and because $\mathbf{x}_0$ ($p < 0.0001$) is much more significant than $\mathbf{x}_1$ ($p < 0.002$) it can be concluded that stool diameter ($\mathbf{x}_0$) is a stronger predictor variable than that of sum of cut surface diameters ($\mathbf{x}_1$). So $\mathbf{x}_1$, although significant when entered into the model on its own, removes very little of the unexplained variation in comparison to $\mathbf{x}_0$. The examination of all possible 2-

variable subsets entered into the model does not provide a clear cut solution to finding the best linear predictor. Since the combinations of $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$; $\mathbf{x}_0, \mathbf{x}_1$ and $\mathbf{x}_2, \mathbf{x}_3$ can be disregarded due to the non significance of one or more of the variables, the remaining possible linear predictors include either $\mathbf{x}_0, \mathbf{x}_2$ or $\mathbf{x}_1, \mathbf{x}_2$. The question is, which of these two subsets should be selected for further attention? Both of these equations in Table 4.5.1, column B are acceptable models but if

$$\eta^{(7)}_{jm\ell} = \alpha_j + q_1\mathbf{x}_1 + q_2\mathbf{x}_2 + F \tag{4.5.2}$$

is chosen then there will be some inconsistency because the best single-variable equation involves $\mathbf{x}_0$. It can also be seen that the p-values for $\mathbf{x}_0, \mathbf{x}_2$ are more significant than those for $\mathbf{x}_1, \mathbf{x}_2$ suggesting that $\mathbf{x}_0$ and $\mathbf{x}_2$ are stronger predictors. This is seen more clearly in Table 4.5.2 which lists the log likelihood ratio statistics, in order of the highest to lowest values, for a model with only the constant terms α_j and the fitted model. This is given for each model fitted. (See section 4.3.) This was used as an informal way of assessing which combination of predictors performed best. Those models with larger likelihood ratio statistics (when considered with the degrees of freedom in the model) may be taken to explain more of the variation in the data.

Table 4.5.2

Likelihood ratio statistics for all possible regressions

(*df* = Degrees of freedom)

A (1-variable)	B (2-variable)	C (3-variable)
518.426 ($\mathbf{x}_0$)	522.448 ($\mathbf{x}_0, \mathbf{x}_2$)	522.560 ($\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$)
df = 15	*df* = 16	*df* = 17
510.585 ($\mathbf{x}_2$)	519.140 ($\mathbf{x}_0, \mathbf{x}_1$)	-
df = 15	*df* = 15	
509.261 ($\mathbf{x}_1$)	514.561 ($\mathbf{x}_1, \mathbf{x}_2$)	-
df = 15	*df* = 15	
499.564 ($\mathbf{x}_3$)	510.668 ($\mathbf{x}_2, \mathbf{x}_3$)	-
df = 15	*df* = 15	

Table 4.5.2 was examined to determine the existence of any consistent pattern of variables in the leading equations in each column. It can be seen that those models with the same degrees of freedom but with higher log likelihood ratio statistics - and so consequently can be said to explain more variation in the data - are those that contain $\mathbf{x}_0$ and $\mathbf{x}_2$. To summarise, the best linear predictor presented in these investigations is the linear predictor which includes the continuous variables stool diameter, $\mathbf{x}_0$ and highest height, $\mathbf{x}_2$. To test whether this model was significantly different from the model with the single continuous variable $\mathbf{x}_0$, a goodness of fit statistic was calculated with the result that the difference in deviance statistics 4.062, when tested on 1 degree of freedom was significant at the 0.05 level of significance. So the best linear predictor is of the form

$$\eta_{jm\ell} = \alpha_j + q_0\mathbf{x}_0 + q_2\mathbf{x}_2 + F \qquad (4.5.3)$$

For thoroughness and to determine the adequacy of (4.5.3) to model the tree stump data, each component was assessed systematically and the maximum likelihood estimates and log-likelihood ratio statistics for the proportional odds model calculated. To assess the significance of the introduction of each factor to the proportional odds model the following models were compared.

Model 1: $logit(\pi_j) = \alpha_j + q_0\mathbf{x}_0 + q_2\mathbf{x}_2$

Model 2: $logit(\pi_{jm}) = \alpha_j + q_0\mathbf{x}_0 + q_2\mathbf{x}_2 + \beta_{1m}a_m$

Model 3: $logit(\pi_{jml}) = \alpha_j + q_0\mathbf{x}_0 + q_2\mathbf{x}_2 + \beta_{1m}a_m + \beta_{2\ell}b_\ell$

Model 4: $logit(\pi_{jml}) = \alpha_j + q_0\mathbf{x}_0 + q_2\mathbf{x}_2 + \beta_{1m}a_m + \beta_{2\ell}b_\ell + \beta_{3m\ell}(a_m b_\ell)$

The table overleaf gives the results of the comparisons of models 1 and 2, 2 and 3, and finally 3 and 4, respectively. The procedure used here (also used to assess the goodness of fit test on above) to compare the different models was simplified somewhat in recognising that the maximal model is the same for all models. (The maximal model has the number of parameters equal to the number of observations.) Therefore, calculating the difference in deviance statistics did not involve the maximal model at all and so the deviance statistic (rewritten) reduces to:

$$D_2 - D_1 = 2log\lambda_2 - 2log\lambda_1 =$$

$$2\{[\ell(\hat{\beta}_2;\mathbf{y}) - \ell(\hat{\beta}_{2max};\mathbf{y})] - [\ell(\hat{\beta}_1;\mathbf{y}) - \ell(\hat{\beta}_{1max};\mathbf{y})] + [\ell(\hat{\beta}_{2max};\mathbf{y}) - \ell(\hat{\beta}_{1max};\mathbf{y})]\}$$

$$= 2[-\ell(\hat{\beta}_2;\mathbf{y}) + \ell(\hat{\beta}_1;\mathbf{y})]$$

$$= 2[\ell(\hat{\beta}_2;\mathbf{y}) - \ell(\hat{\beta}_1;\mathbf{y})]. \qquad (4.5.4)$$

Thus to test the null hypothesis that the models of interest are the same, the difference in deviance of the two models was calculated as above and the probability value was obtained from the $\chi^2_{p_2-p_1}$ statistical table. The results are given in Table 4.5.3. For example, the column titled 'Model 2' gives the result for the comparison of 'Model 2' with 'Model 1'. The null hypothesis for this test was

$$H_0: \; logit(\pi_j) = \alpha_j + \beta_0 x \; = \; logit(\pi_{jm}) = \alpha_j + \beta_0 x + \beta_{1m} a_m$$

and the alternative was

$$H_1: \; logit(\pi_j) = \alpha_j + \beta_0 x \; \neq \; logit(\pi_{jm}) = \alpha_j + \beta_0 x + \beta_{1m} a_m .$$

H_0 is rejected in favour of H_1 if the difference in deviance statistics is higher than a critical value. This was tested at the 5% significance level.

Table 4.5.3.

Results of goodness of fit statistics

Parameters	Model 1	Model 2	Model 3	Model 4
Log-likelihood	-793.089	-565.602	-553.014	-539.733
No. of Parameters	2	4	8	16
Deviance Statistic (Difference)	-	454.974	25.176	26.534
No. of Parameters (Difference)	-	2	4	8
Significance Level	-	<0.001	<0.001	<0.05

From Table 4.5.3, there is clear evidence that the best fitting model is 'Model 4'. The final conclusion that 'Model 4' is the better model, was due to the reduction in the differences between the log-likelihood values as more terms are added. It was of interest to assess the inclusion of a stool diameter by herbicide or method of application interaction term but further investigations did not uncover any of these combinations to significantly improve on Model 4.

In summary, Table 4.5.3. shows that there exists both a method of application effect and a type of herbicide effect in the response. Furthermore, there was statistically significant evidence of an interaction between both of these factors.

The results of fitting

$$logit(\pi_{jml}) = \alpha_j + q_0\mathbf{x}_0 + q_2\mathbf{x}_2 + \beta_{1m}a_m + \beta_{2\ell}b_\ell + \beta_{3m\ell}(a_m b_\ell)$$

to the data are given in Table 4.5.4. The *z*-statistics are printed alongside the associated probabilities. The estimated response factors α_1, α_2, and α_3 (-2.2216, -0.7968 and 1.4030, respectively) quantify the difference between successive categories on the logit scale. Therefore the difference between a kill at the first application of treatment and one at the second application is quantified by -2.2216 and the difference between a kill at the second application and the third application is –0.7968. By the same token the difference between a kill at the fourth application and one at the third is 1.4030. As expected the difference between the extreme categories and the intermediate categories is larger than the difference between the intermediate categories, with the biggest difference existing between category 1 and 2.

From Table 4.5.4 stool diameter (q_0 = -0.020289; z = -3.42, p = < 0.001) is statistically significant at the 5% level. The negative coefficient, and an odds ratio that is just below one indicates that larger stool diameters are associated with a higher number of attempts to kill. Although stool diameter is statistically significant, the odds ratio is very close to one, indicating that a one centimetre increase in diameter minimally effects a tree stump's survival. A more meaningful difference is found by comparing tree stumps with a larger diameter difference. Moreover, the confidence interval for the odds ratio is narrow, suggesting that the odds ratio is precise. Highest

Table 4.5.4

Ordinal logistic regression parameter estimates

Parameter	Estimate	St. err.	z-statistic	p-value	Odds ratio	95% CI	
α_1	-2.2216	0.4734	-4.69	0.000			
α_2	-0.7968	0.4542	-1.75	0.079			
α_3	1.4030	0.4548	3.08	0.002			
q_0	-0.0202	0.0059	-3.42	0.001	0.98	0.97	0.99
q_2	-0.0072	0.0035	-2.05	0.040	0.99	0.99	1.0
Method							
β_{12}	5.2303	0.567	9.22	0.00	186.84	61.49	567.7
β_{13}	0.2075	0.4507	0.46	0.645	1.23	0.51	2.9
Herbicide							
β_{22}	0.8058	0.4394	1.83	0.067	2.24	0.95	5.3
β_{23}	1.1366	0.4470	2.54	0.011	3.12	1.30	7.4
β_{24}	1.3729	0.411	3.11	0.002	3.95	1.66	9.3
β_{25}	-0.5154	0.4551	-1.13	0.257	0.60	0.24	1.4
Interaction							
β_{322}	-1.7536	0.7016	-2.50	0.012	0.17	0.04	0.60
β_{323}	1.069	1.183	0.90	0.366	2.91	0.29	29.6
β_{324}	-0.3145	0.8701	-0.36	0.718	0.73	0.13	4.0
β_{325}	0.7053	0.7885	0.89	0.371	2.02	0.43	9.50
β_{332}	0.0943	0.6126	0.15	0.878	1.10	0.33	3.63
β_{333}	-0.4083	0.6172	-0.66	0.508	0.66	0.20	2.21
β_{334}	-0.2063	0.6132	-0.34	0.737	0.81	0.24	2.71
β_{335}	1.1189	0.6314	1.77	0.076	3.06	0.89	10.55

height (q_2 = -0.007244; z = -2.05; p < 0.040) is statistically significant at the 5% level. Again, the negative coefficient, and an odds ratio that is just below one indicates that higher highest heights are associated with a higher number of attempts to kill. As before, even though highest height is statistically significant a one centimetre increase in height minimally affects a tree stump's survival. Larger differences in highest height measurements will have a more pronounced effect.

If at this stage the constant terms in the model are interpreted as the cut-points when the factors are at level one, then the parameter β_{12} is the difference between the cut-points for the cut surface and basal frill methods, (i.e., basal frill is $e^{\beta_{12}}$ times the cut surface method.). For the levels of both factors the parameters represent differences from the first level, making it easy to compare each level of a factor with its base level. The estimate for basal frill $(\beta_{12} = 5.2303; z = 9.22; p < 0.0001)$ shows that the difference between the cut-points for the levels of method of application is significant at the 5% level. The positive coefficient and the odds ratio substantially greater than one (186.19) indicates that those stools that have herbicide applied in this manner tend to require a lower number of applications. The 95% confidence interval for the odds ratio is wide (61.49, 567.7). In contrast, the estimate for the foliar spray method is not significant (p = 0.645) and so it was concluded that the foliar spray method is not significantly different to that of the cut surface method of application.

Recall that there are five types of herbicide in all, namely, Timbril, Garlon, Chopper, Brush-off and Nomix + Garlon. Of these only Chopper $(\beta_{23} = 1.1366, z = 2.54, p = 0.011)$ and Brush-off $(\beta_{24} = 1.3729, z = 3.11, p = 0.002)$ performed significantly differently to the Timbril herbicide. The estimate for Garlon was not statistically significant $(\beta_{22} = 0.8058, z = 1.83, p = 0.067)$ neither was the estimate for the Nomix+Garlon herbicide $(\beta_{25} = -0.5154, z = -1.13, p = 0.257)$. It was also of interest to determine whether the statistically significant herbicides were performing better or worse than Timbril. The odds ratio for Brush-off (OR = 3.71) and Chopper (OR = 2.88), are both greater than one. Meaning that tree stumps treated with Brush-off tend to require a lower number of applications than those treated with Timbril.

As previously stated parameter estimates for the levels of herbicide and method of application reflect the estimated change in

$$\log\left(\frac{\pi_{1m\ell}+\ldots+\pi_{(j-1)m\ell}}{1-(\pi_{1m\ell}+\ldots+\pi_{(j-1)m\ell})}\right),$$

The presence of a statistically significant interaction factor suggests that the two factors were not acting independently of each other and that the factors are behaving multiplicatively.

The combination of applying the Garlon herbicide using the basal frill method (level 2*2) produces a significant change in the cumulative logits (when compared against the base level interactions). It is difficult to understand what is actually happening here as the negative coefficient means that relatively more probability mass at this level falls at the higher end of the response scale. Listed in Table 4.5.5 are the number of kills, the mean stool diameter of the stools and its standard deviation. Means and standard deviations are given only for stool diameter as it is the strongest continuous predictor in the model. Table 4.5.5 was useful at this stage for exploring the raw data but is not adjusted for main effects and as such is not presented here as meaningful in the predictive sense. The results here seem to contradict the regression results in showing that the highest instances of kills are found in the lower end of the response scale (see row 2*2). An explanation for this behaviour at this stage is offered in the simple fact that the trends are weakly defined in some of the base interaction levels and a number of these have higher instances of kills in the intermediate response categories. In addition it was thought that the p-value could be considered close enough to the 5% level to warrant caution because a lot of combinations are being independently contrasted with the base levels of the interaction term. In such circumstances it

Table 4.5.5

Table of number of applications until a kill by treatment combination

Method*Herbicide	Number of applications 1	2	3	4	All
1*1					
	1	4	12	21	38
	65.500	56.750	52.208	56.762	55.553
	--	22.824	19.911	15.217	17.104
1*2					
	4	8	13	15	40
	48.625	49.813	41.308	54.400	48.650
	27.654	16.551	10.301	21.763	18.483
1*3					
	5	8	14	12	39
	43.000	47.438	40.679	45.000	43.692
	13.852	12.397	10.263	11.308	11.334
1*4					
	5	5	22	8	40
	46.200	41.400	53.136	47.375	49.650
	13.989	15.278	17.790	8.043	15.627
1*5					
	5	1	8	26	40
	30.800	39.500	44.813	48.462	45.300
	9.115	--	12.983	12.292	13.045
2*1					
	32	1	5	0	38
	47.141	53.000	61.700	--	49.211
	15.041	--	26.035	--	16.970
2*2					
	28	10	2	1	41
	40.036	51.300	49.500	58.500	43.695
	14.267	17.375	7.778	--	15.422
2*3					
	41	0	1	0	42
	50.817	--	55.000	--	50.917
	14.695	--	--	--	14.529
2*4					
	38	3	0	0	41
	46.211	68.167	--	--	47.817
	16.233	26.269	--	--	17.656
2*5					
	34	4	1	0	39
	41.574	44.750	51.500	--	42.154
	16.169	7.643	--	--	15.328

Table 4.5.5 Contd.

Method*Herbicide	Number of applications				
3*1	0	5	17	15	37
	--	47.700	42.353	51.933	46.959
	--	8.715	15.196	14.970	14.786
3*2					
	0	6	32	4	42
	--	37.833	47.156	66.750	47.690
	--	12.910	9.010	24.659	13.270
3*3					
	1	9	20	12	42
	35.000	53.667	47.825	51.917	49.940
	--	20.788	22.929	20.903	21.378
3*4					
	2	11	19	8	40
	48.000	44.227	49.868	52.750	48.800
	2.828	10.036	17.057	11.032	13.851
3*5					
	1	5	24	12	42
	51.000	51.800	51.167	62.250	54.405
	--	9.846	14.637	18.251	15.630
All					
	197	80	190	134	601
	45.302	48.688	48.082	52.862	48.317
	15.654	15.456	15.945	16.201	16.044

Counts of kills

Mean of stool diameter

Standard deviation of stool diameter

would have been preferable to use a multiple comparison test because as the risk of making a Type II error conclusion was increased.

On further investigation Table 4.5.5.led to the production of Table 4.5.6. which ranks each herbicide within the levels of method of application according to the percentage of first time kills in each factor combination.

Table 4.5.6

Ranks of Herbicide (% killed after first/second application)

Type of Herbicide ranked highest to lowest

Method					
1	**3** (12%)	**4** (12%)	**5** (12%)	**2** (10%)	**1** (3%)
		Y=2 (12%)	Y=2 (2%)		
2	**3** (97%)	**4** (93%)	**5** (87%)	**1** (84%)	**2** (68%)
3	**4** (5%)	**3** (2.3%)	**5** (2.3%)	**2** (0%)	**1**(0%)
		Y=2 (21%)	Y=2 (12%)	Y=2 (14%)	Y=2 (13%)

Table 4.5.6 confirms that both herbicides 3 and 4 consistently out-perform the other types of herbicide regardless of the method used to apply the herbicide. There is evidence of a slight interaction between method 3 and herbicide 3 and herbicide 4, but this has not proved to be significant in the analysis. Level 5 of type of herbicide performs consistently regardless of the method of application used and better than herbicides 2 and 1. In the last two columns of Table 4.5.6 an interaction is suggested at level 2 of method of application. The application of herbicide 2 using method 1, produces a slightly higher percentage kill on the first application than when applying herbicide 1 using the same method. A slimmer difference is seen in method 3 (although a first kill is not obtained until the second application). In contrast when these herbicides are applied using method 2 herbicide 1 has quite a considerably higher percentage of first time kills than herbicide 2 yeilding the interaction between levels 2 of method and herbicide in the analysis. So when the Garlon herbicide is applied using the basal frill method of application this combination

produces a performance 0.18 times as much as the base level of interactions set by the constrained model.

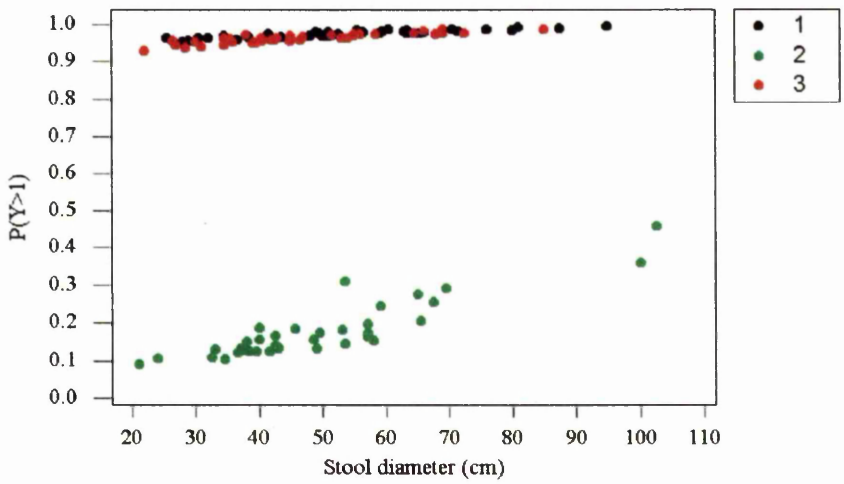
4.6 Prediction

Using the cut-point and coefficient parameter estimates from Table 4.6.4, predicted logits were calculated and from these predicted values of P(Y>1) for each of the five types of herbicides were also calculated. Figures 4.6.1 (a). - 4.6.5 (a) plot predicted values of (P>1) as a function of the stool diameter at the three levels of method of application. Figures 4.6.1 (b) – 4.6.5 (b) plot predicted values of (P>1) as a function of highest height. The constant difference between the three methods is displayed in all ten graphs. In each graph the lower line represents the basal frill method of application, and the upper lines, often intermingled, represent the cut surface and foliar spray methods of application.

For the basal frill method, as stool diameter increases it can be expected that the probability that follow-up operations will be needed increases. This probability increases most quickly when using the Garlon herbicide seen in Figure 4.6.2 (a). This pertains to the interaction term in the model and reflects the findings that Garlon when applied using the basal frill method of application performs poorer in terms of a first time kill as stool diameter increases. Figures 4.6.3 (a) & 4.6.3 (b) show that the probability of a follow up operation, as stool diameter and highest height increase, remains roughly constant for the Chopper and Brush-off herbicides when applied using the basal frill method of application. The cut surface and foliar spray methods show a small but steady increase across all herbicides but neither method really performs better than the other. Figures 4.6.5 (a) & (b) reveal the largest difference between these two methods; the Garlon+Nomix herbicide performs slightly better when applied with the cut surface method rather than with the foliar spray method.

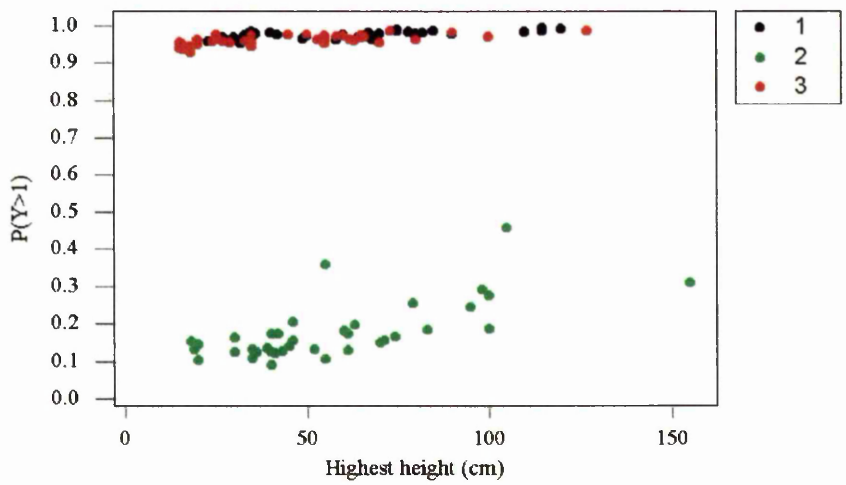
Overall, Figures 4.1 (a). – 4.1 (e), show that as stool diameter and highest height increase, the probability that further applications are needed increases too. The rate of probability increase, as stool diameter and highest height

Figure 4.6.1 (a). Predicted probabilities of P(Y>1) vs stool diameter for the Timbril herbicide.



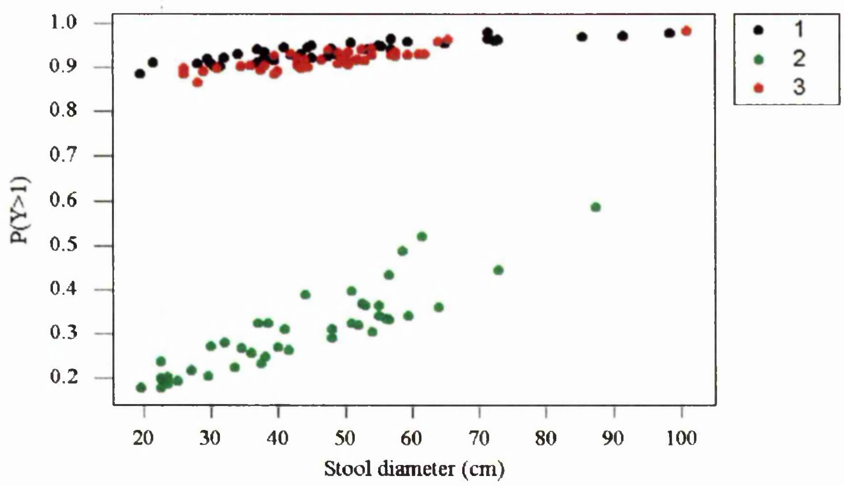
(1= cut surface; 2 = basal frill; 3 = foliar spray)

Figure 4.6.1 (b). Predicted probabilities of P(Y>1) vs highest height for the Timbril herbicide.



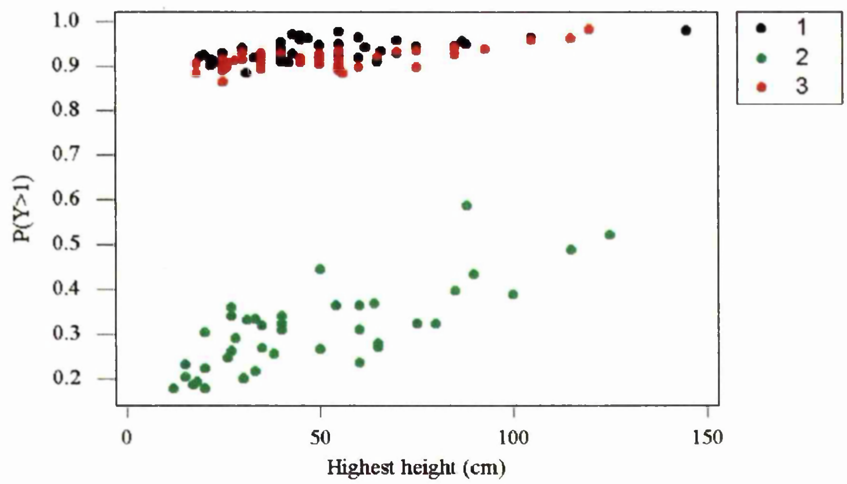
(1= cut surface; 2 = basal frill; 3 = foliar spray)

Figure 4.6.2 (a). Predicted probabilities of P(Y>1) vs stool diameter for the Garlon herbicide.



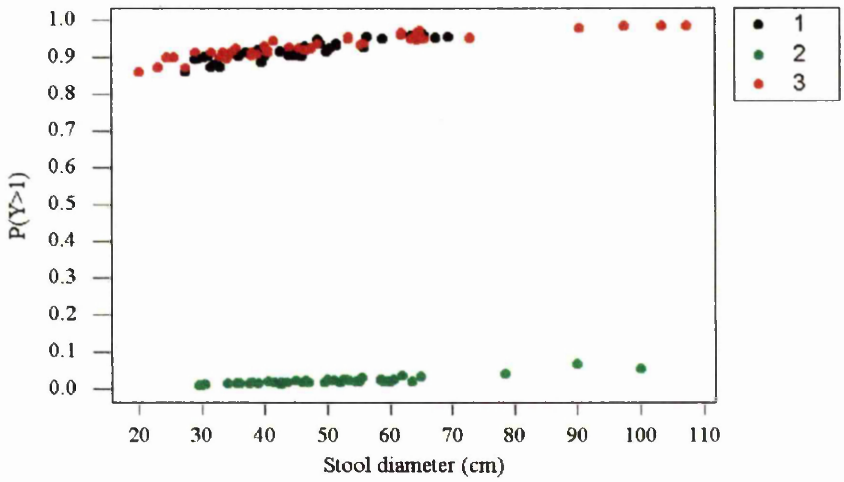
(1= cut surface; 2 = basal frill; 3 = foliar spray)

Figure 4.6.2 (b). Predicted probabilities of P(Y>1) vs highest height for Garlon herbicide.



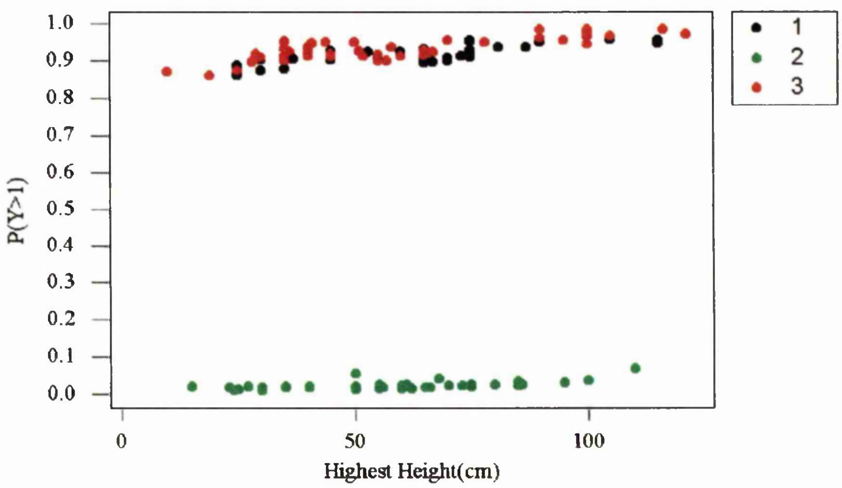
(1= cut surface; 2 = basal frill; 3 = foliar spray)

Figure 4.6.3 (a). Predicted probabilities of P(Y>1) vs stool diameter for Chopper herbicide.



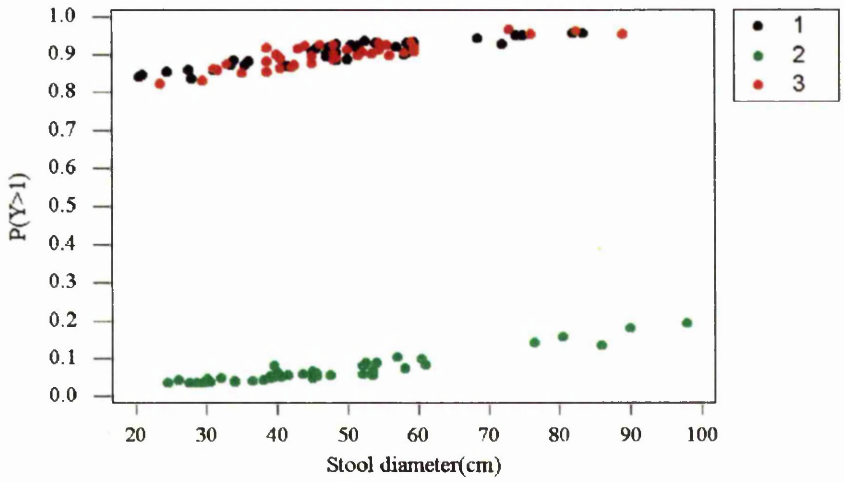
(1= cut surface; 2 = basal frill; 3 = foliar spray)

Figure 4.6.3 (b). Predicted probabilities of P(Y>1) vs highest height for Chopper herbicide



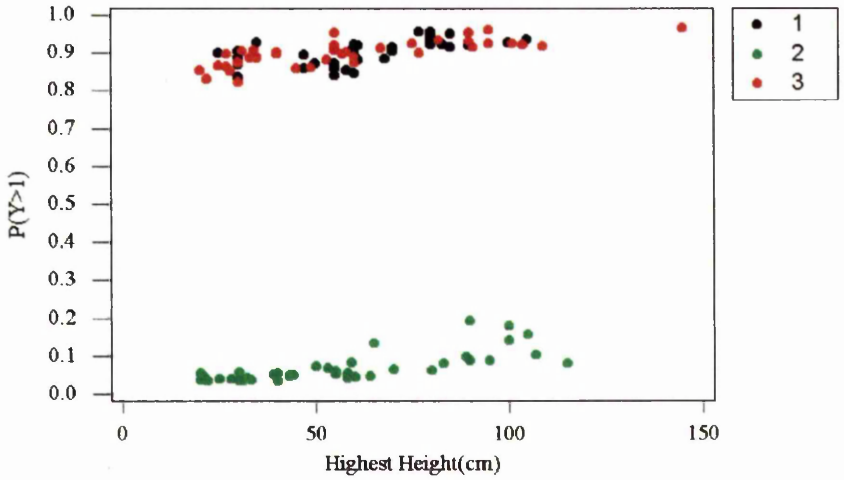
(1= cut surface; 2 = basal frill; 3 = foliar spray)

Figure 4.6.4 (a). Predicted probabilities of P(Y>1) vs stool diameter for Brush-off herbicide



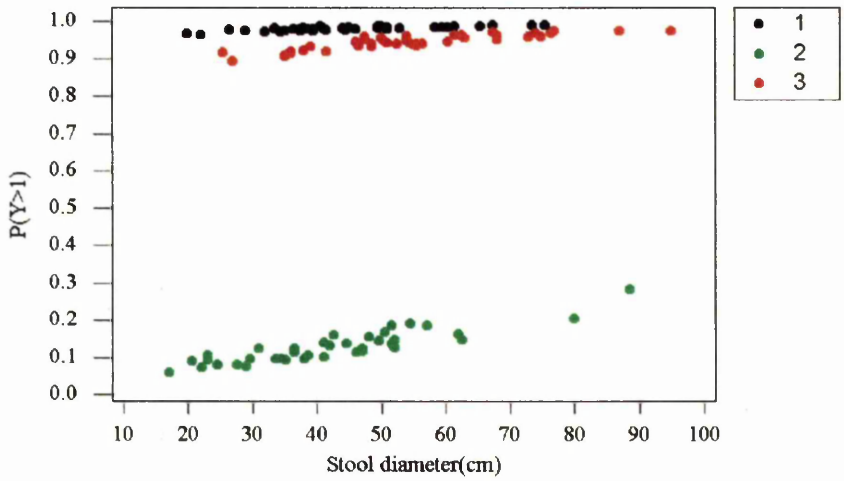
(1= cut surface; 2 = basal frill; 3 = foliar spray)

Figure 4.6.4 (b). Predicted probabilities of P(Y>1) vs highest height for Brush-off herbicide



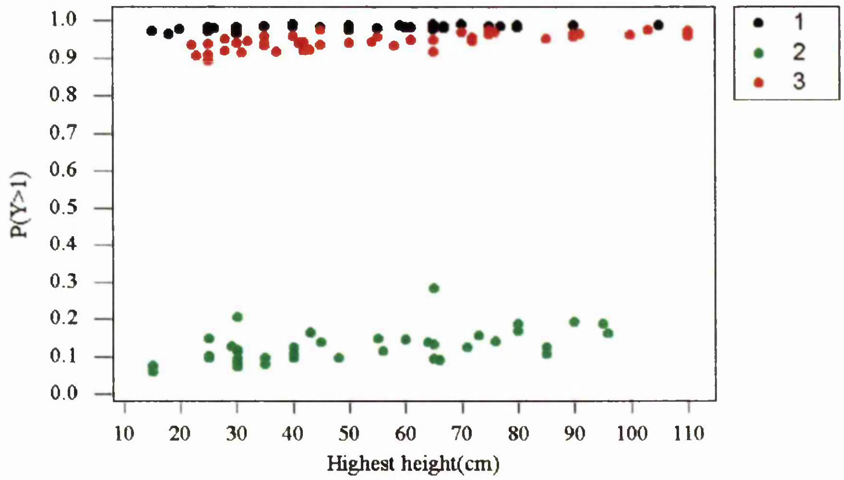
(1= cut surface; 2 = basal frill; 3 = foliar spray)

Figure 4.6.5 (a). Predicted probabilities of P(Y>1) vs stool diameter for Nomix+Garlon herbicide



(1= cut surface; 2 = basal frill; 3 = foliar spray)

Figure 4.6.5 (b). Predicted probabilities of P(Y>1) vs highest height for Nomix+Garlon herbicide



(1= cut surface; 2 = basal frill; 3 = foliar spray)

increase, differs for each of the methods depending on the type of herbicide used. The basal frill method of application consistently performs better than the other methods. The probability of further applications is much lower than the other two methods of application, even s stool diameter and highest height increase. This effect is more defined in the Brush-off and Chopper herbicides. The probability of further applications when herbicide is applied using either the cut surface or foliar spray method of application is at the higher end of the probability scale with this probability slowly approaching one with increases in stool diameter and highest height.

4.7 An additional analysis

An additional analysis was performed on a subset of the data set, using the same model as above. It is thought that in some instances an individual stool's take up of herbicide might be inhibited when applied using the cut surface method of application. This is because of the way in which the herbicide is applied using this method. Recall (Chapter 1, page 5) that the herbicide is applied to the cut surface area of the stools immediately after felling. This cut surface is variable across stools and there was concern that those stools with small cut surface areas and large sinks (large stool diameters and so hence greater stool volumes) were receiving a disproportionate amount of herbicide to their size therefore making them harder to kill. Although in the trial no record of the amount of herbicide applied to a stool was recorded it was considered reasonable that those stools with greater sums of cut surface diameters would receive more herbicide than others.

The subset of data on which the analysis was carried out involved only those stools that had herbicide applied using the cut surface method of application. The subset contained 210 data vectors of which 13 had missing values in the response, total number of application until a kill. The analysis did not use any of the treatment factors in the linear predictor since the purpose of this analysis was primarily to determine if the ratio of stool diameter to sum of cut surface diameters had any effect on the survival of the tree stumps. The linear predictor used in this model was of the form,

$$\eta = \alpha_j + q_3\mathbf{x}_3$$

where α_j is as before and $\mathbf{x}_3$ represents the ratio of stool diameter to sum of cut surface diameters. The full results of the analysis are given in Appendix B. The results from the analysis revealed that the ratio of stool diameter to the sum of cut surface diameters ($q_3 = 0.01442$; $z = 1.07$; $p = 0.285$) was not useful in predicting the number of applications of herbicide to a tree stump until a kill is achieved.

CHAPTER 5

CONCLUSIONS

Of the many continuous variables assessed as predictors for the systematic component of the generalised linear model, stool diameter (cm) was the most promising in the preliminary analysis with the strongest linear relationship with the cumulative logits. Highest height and sum of cut surface diameters also showed some signs of linearity with the cumulative logits and so accordingly, these three variates were assessed, using a best subsets approach, as predictors in the proportional odds model. Stool diameter and highest height were found to be significant predictors of the number of applications required until a stool is killed. The ratio of stool diameter to sum of cut surface diameter was also assessed as a predictor in the model in the above way. This was later used in an additional analysis on a subset of the data to establish if a relationship between amount of herbicide, size of stool sink and stool survival could be modelled. The experiment was originally designed as a factorial experiment and so the study modelled these additional variates whilst staying true to the original factorial design. In addition to the continuous variable the proportional odds model made use of two factors: method of application and type of herbicide. An interaction term for the two factors was also included in the model.

For these predictors, estimates and standard errors of the estimates were calculated using the iterative weighted least squares method for multivariate generalised linear models. The results show that stool diameter, highest height, method of application, type of herbicide and an interaction term for method of application and type of herbicide were all statistically significant in explaining some of the behaviour of the ordinal response variable.

The presence of stool diameter in the model showed that larger stool diameters are associated with a higher number of attempts to kill. A one centimetre increase in stool diameter minimally effects a stool's survival, but the effect is cumulative. Highest height was not as strongly significant in the model but nevertheless was significant at the 5 % level. The presence of highest height in the model meant that higher highest height measurements are associated with a higher number of attempts to kill. As with stool diameter a small increase in highest height minimally effects a stool's survival, but again the effect is cumulative. Of the three methods of application used in the trial basal frill outperformed the cut surface method. Quantification of this performance was given by the odds ratio statistic. The odds that a stool is killed below any fixed level was estimated to be 183.19 times as much using the basal frill method as using the cut surface method: a clear and strong result.

The Chopper and Brush-off types of herbicide outperformed the Timbril herbicide, where the odds that a stool is killed below any fixed level was estimated to be 2.88 times as much using the Chopper herbicide as using the Timbril herbicide. Similarly, the odds ratio for Brush-off showed that the odds of a kill was estimated to be 3.71 times as much in comparison to using the Timbril herbicide. Out of all the contrasts with the Timbril herbicide the Brush-off herbicide performed best. As the estimates given were for paired contrasts, the same model specified with different corner-point constraints will produce new parameter estimates and hence new paired contrasts are calculated.

The model reveals that there exists an interaction effect in the data. The interaction occurs when the Garlon herbicide is applied using the basal frill method. The result was that the odds that a stool is killed below any fixed level was estimated to be 0.18 times as much than using a combination with the Timbril herbicide or the cut surface method of application.

An additional analysis using the proportional odds model was performed exclusively on the data from stools that had been treated using the cut surface method of application. It was concluded from the analysis that the ratio of stool diameter to the sum of cut surface diameter is not useful in predicting the number of applications of herbicide until a kill is achieved.

The recommendations resulting from this study with respect to the type of herbicide to be applied using one of the methods of application are based on the comparisons made in the ordinal logistic regression. Given that all of the parameter estimates for the factor levels arise due to comparative calculations involving the cut surface method of application and the Timbril herbicide the recommendations to be given are conditional on the base line levels. Had the experiment not originally been prematurely terminated these recommendation could have been based on comparisons with the control treatments and as a result would have been much stronger. It is recommended that the basal frill method is used in preference to the cut surface methods and it is also recommended that the Chopper and Brush-off herbicides are used in preference to the Timbril herbicide.

Statistically the method of analysing ordinal response data in this study has proved to be far more efficient and informative than an analysis of variance on the set of transformed data set (variance stabilizing). In addition this method has enabled the effects of factor levels and increments in continuous measurements to be quantified through the proportional odds. Furthermore, the proportional odds model used in this study is more sophisticated statistically, than the previous methods used. For instance it has the property that the response categories can be analysed on their natural scale by being thought of as contiguous intervals on a continuous scale, which in turn means the response can be measured more finely and the estimation of cut-point parameters can be produced. As a consequence of the findings in this study the use of the proportional odds model is advocated as a suitable framework for data of a similar nature.

CHAPTER 6
FURTHER WORK

6.1 Hierarchical/nested response models

This chapter gives a brief introduction to an alternative model that could be used to model the cut-stump data. This particular type of model is known as a nested or hierarchical response model and is discussed in McCullagh & Nelder (1989) on pages 160 – 164 and again on pages 180 – 181.

Consider the total number of applications until a kill is achieved as having response categories that can be broken up into four levels/stages and at each stage reduced to a dichotomy. The response at stage one is the dichotomy between killed and not killed, whereas the response at stage two is the dichotomy between killed and not killed given that the $m - y_1$ stools were not killed on the first application of treatment. Here m represents the total number of stools in the study, and y_1 represents the number of stools killed on the first application of herbicide. At stage three the response is the dichotomy between killed and not killed given that $m - y_1 - y_2$, where y_2 is the number of stools killed on the second application of herbicide, and so on. These responses are very similar but technically different. The response is broken down into conditional factors such as the proportion of stools being killed at the second stage given that they survived the first application. If used to analyse the cut-stump data this approach would allow for an expected incline/decline in the mean susceptibility to the herbicides at the successive stages whilst at the same time the effect of particular predictors in the model affecting stool survival would also be apparent.

The model for this type of ordinal response is as follows:

$$g(\pi_1) = \alpha_1 + \beta^T \mathbf{x}$$

$$g(\pi_2 /(1 - \gamma_1)) = \alpha_2 + \beta^T \mathbf{x}$$

$$g(\pi_3 /(1 - \gamma_3)) = \alpha_3 + \beta^T \mathbf{x}$$

$$g(\pi_4 /(1 - \gamma_4)) = \alpha_4 + \beta^T \mathbf{x}$$

where g must be the same link function. If the logistic link function were used the model would be

$$\log(\pi_j/(1-\gamma_j)) = \alpha_j + \beta^T \mathbf{x} \qquad (6.1.1)$$

where α_j are the parameters allowing for incline/decline in susceptibility as mentioned above and $\gamma_j = \pi_1 + ... + \pi_j$. $\beta^T \mathbf{x}$ is an explanatory variable that can be qualitative or continuous. If **x** is a treatment factor then the result from the analysis is similar to that of the proportional odds model: the odds in favour of success in category j is exp($\alpha_j + \beta^T \mathbf{x}$).

Model (6.1.1) draws many parallels with the proportional-odds model but it has the added benefit of allowing for the unusual case where the ordinal response is to be considered at each stage of the trial, i.e. in the cut stump case at each return visit to re-treat the stools. It is also a good way of modelling a response scale that could in theory have an indefinite number of stages. Further work on this data set might consider fitting (6.1.1) to the data and comparing the results reported in this study.

REFERENCES

1. Agresti, A.: 1984. "Categorical Data Analysis". (John Wiley and Sons, New York).
2. Agresti, A.: 1996. "An Introduction to Categorical Data Analysis". (John Wiley and Sons, New York).
3. Clayton, D.G.: 1974. "Some odds ratio statistics for the analysis of Ordered categorical data". Biometrika, Vol. 61 pp 525-531.
4. Dobson, A.J.: 1990. "An Introduction to Generalised Linear Models". (Chapman and Hall, London).
5. Finney, R.L. and Thomas, G.B.: 1990. "Calculus". (Addison-Wesley Publishing Company, Massachusetts).
6. Little, K.M. and Eccles, N.S.: 1998. "Control of Eucalyptus *grandis* cut-stumps of single stem origin". Institute for Commercial Forestry Research bulletin.
7. Little, K.M., Maxfield, M. and Kritzenger, J.: 1998. "Control of Eucalyptus *grandis* cut-stumps". Institute for Commercial Forestry Research bulletin.
8. Marsh, E.K.: 1963. "Methods of killing eucalyptus stumps". Forestry in South Africa, Vol. 2 pp 111-114.
9. McCullagh, P.: (1980). "Regression models for ordinal data (with Discussion)". Journal of the Royal Statistical Society Series B, Vol. 42 pp109-142.
10. McCullagh, P. and Nelder, J.: (1989). "Generalised Linear Models (Second Edition)". (Chapman and Hall, London)
11. Nelder, J.and Wedderburn, R.W.M.: (1972). "Generalised linear models". Journal of the Royal Statistical Society Series A, Vol. 135 pp370-384.
12. Snedecor, G.W. and Cochran, W.G.: (1980). "Statistical Methods". (Iowa State University Press, Iowa)
13. Snell, E.J.: (1964). "A scaling procedure for ordered categorical data". Biometrics, Vol. 20 pp599-607.

APPENDICES

APPENDIX A

All possible regressions

The variables shown in the following Minitab analysis outputs are:

$\mathbf{x}_0$ = Stool diameter

$\mathbf{x}_1$ = Sum of cut surface diameters

$\mathbf{x}_2$ = Highest height

$\mathbf{x}_3$ = Ratio of stool diameter to sum of cut surface diameters

The ten possible linear predictors are:

$$\eta^{(1)}_{jm\ell} = \alpha_j + F$$

$$\eta^{(2)}_{jm\ell} = \alpha_j + q_0\mathbf{x}_0 + F$$

$$\eta^{(3)}_{jm\ell} = \alpha_j + q_1\mathbf{x}_1 + F$$

$$\eta^{(4)}_{jm\ell} = \alpha_j + q_2\mathbf{x}_2 + F$$

$$\eta^{(5)}_{jm\ell} = \alpha_j + q_0\mathbf{x}_0 + q_1\mathbf{x}_1 + F$$

$$\eta^{(6)}_{jm\ell} = \alpha_j + q_0\mathbf{x}_0 + q_2\mathbf{x}_2 + F$$

$$\eta^{(7)}_{jm\ell} = \alpha_j + q_1\mathbf{x}_1 + q_2\mathbf{x}_2 + F$$

$$\eta^{(8)}_{jm\ell} = \alpha_j + q_0\mathbf{x}_0 + q_1\mathbf{x}_1 + q_2\mathbf{x}_2 + F$$

$$\eta^{(9)}_{jm\ell} = \alpha_j + q_2\mathbf{x}_2 + q_3\mathbf{x}_3 + F$$

$$\eta^{(10)}_{jm\ell} = \alpha_j + q_3\mathbf{x}_3 + F$$

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
totalops	1	197
	2	80
	3	190
	4	134
	Total	601

601 cases were used
29 cases contained missing values

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-3.6714	0.3672	-10.00	0.000			
Const(2)	-2.2710	0.3361	-6.76	0.000			
Const(3)	-0.1361	0.3157	-0.43	0.666			
method							
2	5.2426	0.5619	9.33	0.000	189.16	62.89	568.97
3	0.4737	0.4426	1.07	0.284	1.61	0.67	3.82
herbicid							
2	0.9927	0.4339	2.29	0.022	2.70	1.15	6.32
3	1.3034	0.4374	2.98	0.003	3.68	1.56	8.68
4	1.3748	0.4351	3.16	0.002	3.95	1.69	9.28
5	-0.3115	0.4479	-0.70	0.487	0.73	0.30	1.76
method*herbicid							
2*2	-1.7621	0.6947	-2.54	0.011	0.17	0.04	0.67
2*3	0.821	1.175	0.70	0.485	2.27	0.23	22.77
2*4	-0.3867	0.8592	-0.45	0.653	0.68	0.13	3.66
2*5	0.6688	0.7839	0.85	0.394	1.95	0.42	9.07
3*2	-0.1927	0.6050	-0.32	0.750	0.82	0.25	2.70
3*3	-0.7372	0.6074	-1.21	0.225	0.48	0.15	1.57
3*4	-0.3752	0.6072	-0.62	0.537	0.69	0.21	2.26
3*5	0.6766	0.6174	1.10	0.273	1.97	0.59	6.60

Log-likelihood = -551.209
Test that all slopes are zero: G = 499.497, DF = 14, P-Value = 0.000

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
totalops	1	197
	2	80
	3	190
	4	134
	Total	601

601 cases were used
29 cases contained missing values

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-2.4289	0.4653	-5.22	0.000			
Const(2)	-1.0100	0.4451	-2.27	0.023			
Const(3)	1.1785	0.4434	2.66	0.008			
x_0	-0.024134	0.005596	-4.31	0.000	0.98	0.97	0.99
method							
2	5.2105	0.5658	9.21	0.000	183.19	60.44	555.25
3	0.2813	0.4495	0.63	0.531	1.32	0.55	3.20
herbicid							
2	0.8404	0.4393	1.91	0.056	2.32	0.98	5.48
3	1.0586	0.4442	2.38	0.017	2.88	1.21	6.88
4	1.3114	0.4392	2.99	0.003	3.71	1.57	8.78
5	-0.4952	0.4551	-1.09	0.277	0.61	0.25	1.49
method*herbicid							
2*2	-1.7312	0.7007	-2.47	0.013	0.18	0.04	0.70
2*3	1.149	1.181	0.97	0.331	3.16	0.31	31.95
2*4	-0.2836	0.8681	-0.33	0.744	0.75	0.14	4.13
2*5	0.7071	0.7892	0.90	0.370	2.03	0.43	9.53
3*2	0.0050	0.6113	0.01	0.993	1.01	0.30	3.33
3*3	-0.4008	0.6164	-0.65	0.516	0.67	0.20	2.24
3*4	-0.2270	0.6126	-0.37	0.711	0.80	0.24	2.65
3*5	1.0531	0.6305	1.67	0.095	2.87	0.83	9.86

Log-likelihood = -541.744
Test that all slopes are zero: G = 518.426, DF = 15, P-Value = 0.000

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
totalops	1	197
	2	80
	3	190
	4	134
	Total	601

601 cases were used
29 cases contained missing values

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Const(1)	-3.0762	0.4129	-7.45	0.000			
Const(2)	-1.6674	0.3879	-4.30	0.000			
Const(3)	0.5056	0.3782	1.34	0.181			
x_1	-0.028008	0.008871	-3.16	0.002	0.97	0.96	0.99
method							
2	5.2653	0.5650	9.32	0.000	193.51	63.93	585.69
3	0.4478	0.4457	1.00	0.315	1.56	0.65	3.75
herbicid							
2	0.9338	0.4366	2.14	0.032	2.54	1.08	5.99
3	1.3610	0.4399	3.09	0.002	3.90	1.65	9.24
4	1.5056	0.4391	3.43	0.001	4.51	1.91	10.66
5	-0.3724	0.4509	-0.83	0.409	0.69	0.28	1.67
method*herbicid							
2*2	-1.8018	0.6982	-2.58	0.010	0.17	0.04	0.65
2*3	0.850	1.178	0.72	0.471	2.34	0.23	23.55
2*4	-0.5300	0.8633	-0.61	0.539	0.59	0.11	3.20
2*5	0.6863	0.7863	0.87	0.383	1.99	0.43	9.28
3*2	-0.1705	0.6081	-0.28	0.779	0.84	0.26	2.78
3*3	-0.7506	0.6102	-1.23	0.219	0.47	0.14	1.56
3*4	-0.4974	0.6106	-0.81	0.415	0.61	0.18	2.01
3*5	0.7452	0.6217	1.20	0.231	2.11	0.62	7.13

Log-likelihood = -546.327
Test that all slopes are zero: G = 509.261, DF = 15, P-Value = 0.000

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
totalops	1	197
	2	80
	3	190
	4	134
	Total	601

601 cases were used
29 cases contained missing values

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-3.0346	0.4057	-7.48	0.000			
Const(2)	-1.6212	0.3802	-4.26	0.000			
Const(3)	0.5439	0.3706	1.47	0.142			
x_2	-0.011304	0.003326	-3.40	0.001	0.99	0.98	1.00
method							
2	5.2623	0.5644	9.32	0.000	192.93	63.82	583.27
3	0.3068	0.4464	0.69	0.492	1.36	0.57	3.26
herbicid							
2	0.8984	0.4358	2.06	0.039	2.46	1.05	5.77
3	1.3601	0.4397	3.09	0.002	3.90	1.65	9.23
4	1.4516	0.4375	3.32	0.001	4.27	1.81	10.07
5	-0.3966	0.4502	-0.88	0.378	0.67	0.28	1.63
method*herbicid							
2*2	-1.7901	0.6976	-2.57	0.010	0.17	0.04	0.66
2*3	0.786	1.178	0.67	0.504	2.20	0.22	22.09
2*4	-0.4338	0.8626	-0.50	0.615	0.65	0.12	3.51
2*5	0.6954	0.7854	0.89	0.376	2.00	0.43	9.34
3*2	-0.0030	0.6087	-0.00	0.996	1.00	0.30	3.29
3*3	-0.6527	0.6098	-1.07	0.285	0.52	0.16	1.72
3*4	-0.3090	0.6095	-0.51	0.612	0.73	0.22	2.42
3*5	0.8813	0.6228	1.42	0.157	2.41	0.71	8.18

Log-likelihood = -545.665
Test that all slopes are zero: G = 510.585, DF = 15, P-Value = 0.000

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
totalops	1	197
	2	80
	3	190
	4	134
	Total	601

601 cases were used
29 cases contained missing values

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-2.3962	0.4671	-5.13	0.000			
Const(2)	-0.9773	0.4469	-2.19	0.029			
Const(3)	1.2166	0.4457	2.73	0.006			
x_0	-0.021013	0.006713	-3.13	0.002	0.98	0.97	0.99
x_1	-0.00917	0.01068	-0.86	0.390	0.99	0.97	1.01
method							
2	5.2230	0.5665	9.22	0.000	185.49	61.11	562.98
3	0.3007	0.4501	0.67	0.504	1.35	0.56	3.26
herbicid							
2	0.8422	0.4396	1.92	0.055	2.32	0.98	5.50
3	1.1098	0.4477	2.48	0.013	3.03	1.26	7.30
4	1.3654	0.4438	3.08	0.002	3.92	1.64	9.35
5	-0.4884	0.4557	-1.07	0.284	0.61	0.25	1.50
method*herbicid							
2*2	-1.7493	0.7011	-2.49	0.013	0.17	0.04	0.69
2*3	1.118	1.182	0.95	0.345	3.06	0.30	31.02
2*4	-0.3455	0.8704	-0.40	0.691	0.71	0.13	3.90
2*5	0.7048	0.7892	0.89	0.372	2.02	0.43	9.50
3*2	-0.0149	0.6120	-0.02	0.981	0.99	0.30	3.27
3*3	-0.4512	0.6186	-0.73	0.466	0.64	0.19	2.14
3*4	-0.2907	0.6163	-0.47	0.637	0.75	0.22	2.50
3*5	1.0226	0.6314	1.62	0.105	2.78	0.81	9.58

Log-likelihood = -541.387
Test that all slopes are zero: G = 519.140, DF = 16, P-Value = 0.000

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
totalops	1	197
	2	80
	3	190
	4	134
	Total	601

601 cases were used
29 cases contained missing values

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-2.2216	0.4734	-4.69	0.000			
Const(2)	-0.7968	0.4542	-1.75	0.079			
Const(3)	1.4030	0.4548	3.08	0.002			
x_0	-0.020289	0.005927	-3.42	0.001	0.98	0.97	0.99
x_2	-0.007244	0.003533	-2.05	0.040	0.99	0.99	1.00
method							
2	5.2303	0.5670	9.22	0.000	186.84	61.49	567.70
3	0.2075	0.4507	0.46	0.645	1.23	0.51	2.98
herbicid							
2	0.8058	0.4394	1.83	0.067	2.24	0.95	5.30
3	1.1366	0.4470	2.54	0.011	3.12	1.30	7.48
4	1.3729	0.4411	3.11	0.002	3.95	1.66	9.37
5	-0.5154	0.4550	-1.13	0.257	0.60	0.24	1.46
method*herbicid							
2*2	-1.7536	0.7016	-2.50	0.012	0.17	0.04	0.68
2*3	1.069	1.183	0.90	0.366	2.91	0.29	29.60
2*4	-0.3145	0.8701	-0.36	0.718	0.73	0.13	4.02
2*5	0.7053	0.7885	0.89	0.371	2.02	0.43	9.50
3*2	0.0943	0.6126	0.15	0.878	1.10	0.33	3.65
3*3	-0.4083	0.6172	-0.66	0.508	0.66	0.20	2.23
3*4	-0.2063	0.6132	-0.34	0.737	0.81	0.24	2.71
3*5	1.1189	0.6314	1.77	0.076	3.06	0.89	10.55

Log-likelihood = -539.733
Test that all slopes are zero: G = 522.448, DF = 16, P-Value = 0.000

Ordinal Logistic Regression

```
Link Function:  Logit

Response Information

Variable  Value       Count
totalops  1             197
          2              80
          3             190
          4             134
          Total         601

  601 cases were used
   29 cases contained missing values
```

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-2.7799	0.4278	-6.50	0.000			
Const(2)	-1.3643	0.4050	-3.37	0.001			
Const(3)	0.8188	0.3990	2.05	0.040			
x_1	-0.019419	0.009532	-2.04	0.042	0.98	0.96	1.00
x_2	-0.008487	0.003581	-2.37	0.018	0.99	0.98	1.00
method							
2	5.2734	0.5659	9.32	0.000	195.07	64.34	591.48
3	0.3309	0.4478	0.74	0.460	1.39	0.58	3.35
herbicid							
2	0.8845	0.4370	2.02	0.043	2.42	1.03	5.70
3	1.3872	0.4408	3.15	0.002	4.00	1.69	9.50
4	1.5212	0.4397	3.46	0.001	4.58	1.93	10.84
5	-0.4160	0.4514	-0.92	0.357	0.66	0.27	1.60
method*herbicid							
2*2	-1.8160	0.6992	-2.60	0.009	0.16	0.04	0.64
2*3	0.815	1.180	0.69	0.490	2.26	0.22	22.81
2*4	-0.5152	0.8646	-0.60	0.551	0.60	0.11	3.25
2*5	0.6946	0.7863	0.88	0.377	2.00	0.43	9.35
3*2	-0.0385	0.6101	-0.06	0.950	0.96	0.29	3.18
3*3	-0.6857	0.6112	-1.12	0.262	0.50	0.15	1.67
3*4	-0.4079	0.6119	-0.67	0.505	0.67	0.20	2.21
3*5	0.8764	0.6245	1.40	0.160	2.40	0.71	8.17

```
Log-likelihood = -543.677
Test that all slopes are zero: G = 514.561, DF = 16, P-Value = 0.000
```

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
totalops	1	197
	2	80
	3	190
	4	134
	Total	601

601 cases were used
29 cases contained missing values

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-2.2174	0.4738	-4.68	0.000			
Const(2)	-0.7928	0.4546	-1.74	0.081			
Const(3)	1.4084	0.4553	3.09	0.002			
x_0	-0.019173	0.006810	-2.82	0.005	0.98	0.97	0.99
x_1	-0.00378	0.01099	-0.34	0.731	1.00	0.97	1.02
x_2	-0.006920	0.003637	-1.90	0.057	0.99	0.99	1.00
method							
2	5.2345	0.5672	9.23	0.000	187.63	61.73	570.32
3	0.2186	0.4512	0.48	0.628	1.24	0.51	3.01
herbicid							
2	0.8086	0.4395	1.84	0.066	2.24	0.95	5.31
3	1.1544	0.4491	2.57	0.010	3.17	1.32	7.65
4	1.3920	0.4444	3.13	0.002	4.02	1.68	9.61
5	-0.5117	0.4554	-1.12	0.261	0.60	0.25	1.46
method*herbicid							
2*2	-1.7608	0.7017	-2.51	0.012	0.17	0.04	0.68
2*3	1.060	1.183	0.90	0.370	2.89	0.28	29.35
2*4	-0.3384	0.8719	-0.39	0.698	0.71	0.13	3.94
2*5	0.7042	0.7885	0.89	0.372	2.02	0.43	9.49
3*2	0.0816	0.6132	0.13	0.894	1.08	0.33	3.61
3*3	-0.4287	0.6190	-0.69	0.489	0.65	0.19	2.19
3*4	-0.2331	0.6167	-0.38	0.706	0.79	0.24	2.65
3*5	1.1036	0.6324	1.75	0.081	3.02	0.87	10.41

Log-likelihood = -539.677
Test that all slopes are zero: G = 522.560, DF = 17, P-Value = 0.000

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
totalops	1	197
	2	80
	3	190
	4	134
	Total	601

601 cases were used
29 cases contained missing values

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-2.9635	0.4701	-6.30	0.000			
Const(2)	-1.5498	0.4484	-3.46	0.001			
Const(3)	0.6150	0.4412	1.39	0.163			
x_3	-0.02207	0.07775	-0.28	0.777	0.98	0.84	1.14
x_2	-0.011468	0.003364	-3.41	0.001	0.99	0.98	1.00
method							
2	5.2540	0.5645	9.31	0.000	191.33	63.28	578.50
3	0.2964	0.4467	0.66	0.507	1.35	0.56	3.23
herbicid							
2	0.8903	0.4360	2.04	0.041	2.44	1.04	5.73
3	1.3424	0.4426	3.03	0.002	3.83	1.61	9.11
4	1.4385	0.4396	3.27	0.001	4.21	1.78	9.97
5	-0.4058	0.4515	-0.90	0.369	0.67	0.28	1.61
method*herbicid							
2*2	-1.7765	0.6979	-2.55	0.011	0.17	0.04	0.66
2*3	0.800	1.178	0.68	0.497	2.23	0.22	22.43
2*4	-0.4120	0.8653	-0.48	0.634	0.66	0.12	3.61
2*5	0.7034	0.7857	0.90	0.371	2.02	0.43	9.43
3*2	0.0182	0.6105	0.03	0.976	1.02	0.31	3.37
3*3	-0.6371	0.6108	-1.04	0.297	0.53	0.16	1.75
3*4	-0.2894	0.6121	-0.47	0.636	0.75	0.23	2.49
3*5	0.9024	0.6257	1.44	0.149	2.47	0.72	8.40

Log-likelihood = -545.623
Test that all slopes are zero: G = 510.668, DF = 16, P-Value = 0.000

Ordinal Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
totalops	1	197
	2	80
	3	190
	4	134
	Total	601

601 cases were used
29 cases contained missing values

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-3.7260	0.4228	-8.81	0.000			
Const(2)	-2.3256	0.3958	-5.88	0.000			
Const(3)	-0.1897	0.3779	-0.50	0.616			
x_3	0.01934	0.07659	0.25	0.801	1.02	0.88	1.18
method							
2	5.2501	0.5624	9.34	0.000	190.58	63.30	573.82
3	0.4805	0.4428	1.09	0.278	1.62	0.68	3.85
herbicid							
2	0.9982	0.4341	2.30	0.021	2.71	1.16	6.35
3	1.3199	0.4408	2.99	0.003	3.74	1.58	8.88
4	1.3877	0.4376	3.17	0.002	4.01	1.70	9.44
5	-0.3044	0.4491	-0.68	0.498	0.74	0.31	1.78
method*herbicid							
2*2	-1.7735	0.6953	-2.55	0.011	0.17	0.04	0.66
2*3	0.808	1.176	0.69	0.492	2.24	0.22	22.50
2*4	-0.4059	0.8619	-0.47	0.638	0.67	0.12	3.61
2*5	0.6624	0.7845	0.84	0.398	1.94	0.42	9.02
3*2	-0.2081	0.6065	-0.34	0.732	0.81	0.25	2.67
3*3	-0.7498	0.6084	-1.23	0.218	0.47	0.14	1.56
3*4	-0.3916	0.6096	-0.64	0.521	0.68	0.20	2.23
3*5	0.6606	0.6197	1.07	0.286	1.94	0.57	6.52

Log-likelihood = -551.175
Test that all slopes are zero: G = 499.564, DF = 15, P-Value = 0.000

APPENDIX B

An additional analysis

The variable shown in the following Minitab analysis output is:

$\mathbf{x}_3$ = Ratio of stool diameter to sum of cut surface diameters

The linear predictor used in this analysis is:

$$\eta_{jm\ell} = \alpha_j + q_3 \mathbf{x}_3$$

Ordinal Logistic Regression: totalops versus stoold:csd

```
Link Function:  Logit

Response Information

Variable  Value        Count
totalops  1              197
          2               80
          3              190
          4              134
          Total          601

  601 cases were used
   29 cases contained missing values

Logistic Regression Table
                                                     Odds       95% CI
Predictor        Coef     SE Coef        Z      P   Ratio    Lower    Upper
Const(1)      -0.7238      0.1821    -3.98 0.000
Const(2)      -0.1623      0.1797    -0.90 0.366
Const(3)       1.2429      0.1875     6.63 0.000
stoold:c      0.00220     0.06396     0.03 0.973     1.00     0.88     1.14

Log-likelihood = -800.957
Test that all slopes are zero: G = 0.001, DF = 1, P-Value = 0.971
```