

ANALYSIS OF ORDERED CATEGORICAL DATA

By

Janis Chang

B. Sc., (Biochemistry) University of British Columbia, 1985

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF STATISTICS

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 1988

© Janis Chang, 1988

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Statistics
The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date:

Oct. 17, 1988

Abstract

Methods of testing for a location shift between two populations in a longitudinal study are investigated when the data of interest are ordered, categorical and non-linear. A non-standard analysis involving modelling of data over time with transition probability matrices is discussed. Next, the relative efficiencies of statistics more frequently used for the analysis of such categorical data at a single time point are examined. The Wilcoxon rank sum, McCullagh, and 2 sample t statistic are compared for the analysis of such cross sectional data using simulation and efficacy calculations. Simulation techniques are then utilized in comparing the stratified Wilcoxon, McCullagh and chi squared-type statistic in their efficiencies at detecting a location shift when the data are examined over two time points. The distribution of a chi squared-type statistic based on the simple contingency table constructed by merely noting whether a subject improved, stayed the same or deteriorated is derived. Applications of these methods and results to a data set of Multiple Sclerosis patients, some of whom were treated with interferon and some of whom received a placebo are provided throughout the thesis and our findings are summarized in the last Chapter.

Table of Contents

Abstract	ii
List of Tables	iv
List of Figures	v
Acknowledgement	vi
1 Introduction	1
2 Markov Analysis	3
2.1 Tests of Order and Stationarity	4
2.1.1 Tests of Order of Markov Chain	6
2.1.2 Tests of Stationarity	8
2.2 Modelling of the Tridiagonals	10
2.2.1 Modelling Incorporating \mathcal{P}^2	11
2.2.2 Modelling Without \mathcal{P}^2	12
3 Analysis At One Time Point	19
3.1 Simulation	20
3.2 Efficacy Calculations	22
3.2.1 Efficacy of the Wilcoxon	23
3.2.2 Efficacy of the T test	31
3.3 Comparison of the Efficacies	34

4	Analysis at Two Time Points	36
4.1	Description of the Simulation	36
4.1.1	Stratified Wilcoxon	37
4.1.2	Chi-squared Statistic	38
4.2	Simulation Results	46
4.3	Application of Tests to the MS Data	51
5	Conclusions	53
A	Sample Transition Matrices	55
	Bibliography	60

List of Tables

2.1	Comparison of Treatment and Control Groups	9
2.2	Tests of 2 nd Order vs. 1 st Order Markov Process	9
2.3	Tests of 1 st Order vs. Independence	16
2.4	Results of Test for Stationarity	16
2.5	Treatment vs. Control (general tridiagonal modelling)	16
2.6	Fit of General Tridiagonal Models	17
2.7	Treatment vs. Control (general tridiagonal - no P^2)	17
2.8	Fit of General Model - no P^2	17
2.9	Comparison of Specific and General Models (treatment)	17
2.10	Comparison of Specific and General Models	18
4.1	Results of McCullagh Analysis	51

List of Figures

3.1	Simulation of Data at One Time Point	21
4.1	Simulation Run 1 ($\beta = 0.1, \rho = 0.8$)	47
4.2	Simulation Run 2 ($\beta = 0.1, \rho = 0.6$)	48
4.3	Simulation Run 3 ($\beta = 0.3, \rho = 0.8$)	49
4.4	Simulation Run 4 ($\beta = 0.3, \rho = 0.6$)	50

Acknowledgement

I would like to thank my supervisor, Dr. N.E. Heckman for her continual guidance, support and many invaluable suggestions. Also, I am grateful to Dr. A. J. Petkau advice and careful reading of this work. In addition, the encouragement of Dr. J. Berkowitz and expert computing advice of Peter Schumacher are gratefully acknowledged. I also with to thank the MS Clinic at the UBC Hospital for their support.

This research was funded by the Multiple Sclerosis Society of Canada through the UBC MS Clinic.

Chapter 1

Introduction

One hundred Multiple Sclerosis patients from the patient population of the MS Clinic at the UBC Hospital participated in a randomized double-blind clinical trial to determine the effectiveness of treatment with interferon. Multiple Sclerosis is a progressive disease which attacks the nervous system and often results in loss of vision, motor coordination and/or sensory perception. The severity of the symptoms varies among patients. The subjects in this study are chronic progressive, that is, their condition deteriorates progressively over time. In this trial, fifty patients were assigned randomly to control and treatment groups. Subjects in the control group were given injections of a placebo and those in the treatment group were given injections of interferon for six months. The patients were monitored during the six months of treatment and for eighteen months of follow-up to the subsequent termination of the treatment.

No standard quantitative method of measuring the level of the disease exists. In this study, measurements of symptoms such as mobility or numbness were used to assess the severity of the subject's condition and this information was used to produce the Kurtzke extended disability status scale (EDSS). The Kurtzke EDSS, referred to here as Kurtzke score, was chosen as the means of tracing the subjects' conditions over time. The Kurtzke score is ordered and categorical, taking on values of 0 (normal) to 10 (dead) in increments of 0.5. It is also nonlinear, so that, for example, a change in score of 1 to 2 is not as severe as a change of 5 to 6.

The nonlinearity and categorical nature of the scores makes it inappropriate to treat

them as continuous variables in the assessment of the extent of the disease. In particular, any statistic which requires the assumption of normally distributed observations (such as the 2 sample t statistic) is not suitable for analyzing this type of data. On the other hand, standard categorical data analysis could be done ignoring the ordinal nature of the data. This type of analysis is not appropriate here as information on the degree of improvement or deterioration of the patient's condition would be lost.

One method of analysis, described in Chapter 2, is to consider the categories as states and the movements of the subjects from category to category over time as transitions from state to state. The treatment and control groups were modelled by different transition probability matrices which were compared to determine the effect of the interferon injections. Models assuming stationarity, the Markov property and other restrictions on the transition probabilities were fit to the data to determine whether a simple matrix could be used to describe the transitions of the patients.

In Chapter 3, the Wilcoxon statistic, McCullagh model and 2 sample t test were compared to assess their relative efficiency in detecting a location shift between the two groups when the scores are assumed to have an underlying continuous distribution. The McCullagh model, a modification of the logistic regression model, incorporates the ordinal nature of the scores. In this analysis the data at only one time point was used. The comparison of the tests was based on asymptotic efficacy calculations and simulations.

In Chapter 4, the stratified Wilcoxon, McCullagh and chi squared-type statistic were compared using simulation when analyzing the data between two time points. The chi squared-type statistic calculated was the usual analysis involving contingency tables where the cells were the number of patients in treatment and control who improved, stayed the same or worsened. The distribution of this statistic under the hypothesis of no difference between the groups is discussed.

Chapter 2

Markov Analysis

Since patients in this study moved from state to state over time, the movements from one state to another were modelled for each group using transition probability matrices. Comparison of the two groups then involved analyzing the matrices estimated for each group. This approach does not require the assumption of a parametric form for the transition probabilities, however some models assuming a specific form for the probabilities were fit to the data to determine if the number of parameters required to describe the data could be reduced.

Initially, the data was analyzed in its original form. Transition probability matrices were produced for each group (treatment and control) from 0 months to each of the other time points (1,3,6,9,12,18,24 months). Most observations were on or near the diagonals, indicating that the patients' scores did not change much from one time period to the next. As the dimensions of the matrices were 21×21 and there were only fifty patients in each group, many cells were empty. For this reason, the Kurtzke scores were grouped into five categories, chosen to ensure at least four people in each category at 0 months. The categories were scores of 0-4, 4.5-5.5, 6.0, 6.5, and 7-10, with 0-4 becoming state 1, 4.5-5.5 becoming state 2, etc. Transition matrices of the patients from 0 months to 24 months are displayed in Appendix A. All further calculations in this chapter are based upon this collapsed data.

In Section 2.1, likelihood ratio statistics were used to determine if a second order Markov structure fit appreciably better than a first order model. A test for stationarity

was also applied to the data. Some modelling of the data in the tridiagonal positions is discussed in Section 2.2.

2.1 Tests of Order and Stationarity

The following tests were based on those described by Bhat [1]. Some notation used in the remainder of this chapter is:

z = number of states

t = time points studied, $t = 0, 1, \dots, 7$ corresponding to the times
0, 1, 3, 6, 9, 12, 18, 24 months

X_t = Kurtzke score at time t .

Using the collapsed data, transition probability matrices were calculated for transitions from 0 months to each of the other time points studied. Maximum likelihood estimates for P^T and P^C , the transition matrices of the treatment and control group were computed, as well as for P , the transition matrix under $H_0 : P^T = P^C$. A likelihood ratio statistic was computed separately for each time point $t = 0, 1, \dots, 7$ to test whether the two groups' transition probabilities were similar. Under the null hypothesis that the two groups can be modelled using the same transition probability matrix, it can be shown [6] that the log likelihood function $\ln L(P)$ is:

$$\ln L(P) = B + \sum_{i=1}^z \sum_{j=1}^z (n_{ij}^T \ln p_{ij} + n_{ij}^C \ln p_{ij}). \quad (2.1)$$

where

B = a term independent of the p_{ij} 's

n_{ij}^T = number of observations where $X_t = j$ and $X_0 = i$ in the treatment group

n_{ij}^C = number of observations in the control group where $X_t = j$ and $X_0 = i$

p_{ij} = $\Pr[X_t = j \mid X_0 = i]$.

Under the alternate hypothesis, $H_1 : P^T \neq P^C$, the log likelihood function is:

$$\ln L(P^C, P^T) = B + \sum_{i=1}^z \sum_{j=1}^z (n_{ij}^T \ln p_{ij}^T + n_{ij}^C \ln p_{ij}^C) \quad (2.2)$$

where

p_{ij}^T = $\Pr[X_t = j \mid X_0 = i \text{ in treatment group}]$

p_{ij}^C = $\Pr[X_t = j \mid X_0 = i \text{ in control group}]$.

The log likelihood statistic, G^2 , is then

$$\begin{aligned} G^2 &= -2 [\ln L(\hat{P}) - \ln L(\hat{P}^C, \hat{P}^T)] \\ &= 2 \sum_{i=1}^z \sum_{j=1}^z [(n_{ij}^T \ln \hat{p}_{ij}^T + n_{ij}^C \ln \hat{p}_{ij}^C) - (n_{ij}^T + n_{ij}^C)(\ln \hat{p}_{ij})]. \end{aligned}$$

where all the probabilities were estimated using maximum likelihood techniques:

$$\begin{aligned} \ln L(\hat{P}) &= \text{the log likelihood at } \hat{P} \\ \hat{p}_{ij} &= \frac{n_{ij}^T + n_{ij}^C}{n_{i\cdot}^T + n_{i\cdot}^C} \\ \hat{p}_{ij}^C &= \frac{n_{ij}^C}{n_{i\cdot}^C} \\ \hat{p}_{ij}^T &= \frac{n_{ij}^T}{n_{i\cdot}^T} \\ n_{i\cdot}^C &= \sum_{j=1}^z n_{ij}^C \end{aligned}$$

$$n_{i\cdot}^T = \sum_{j=1}^z n_{ij}^T.$$

Under the null hypothesis, G^2 has an asymptotic chi-squared distribution with degrees of freedom equal to $z(z-1) - y^P$ where y^P is the number of zero entries in \hat{P} .

Pearson's chi squared statistic, X^2 , was also calculated:

$$X^2 = \sum_{i=1}^z \sum_{j=1}^z \left[\frac{(n_{ij}^T - m_{ij}^T)^2}{m_{ij}^T} + \frac{(n_{ij}^C - m_{ij}^C)^2}{m_{ij}^C} \right] \quad (2.3)$$

where $m_{ij}^T = n_{i\cdot}^T \hat{p}_{ij}$, and $m_{ij}^C = n_{i\cdot}^C \hat{p}_{ij}$ are the expected number of observations in the treatment and control group who moved from state i to state j between time 0 and time t .

Results from the seven tests are given in Table 2.1 and the transition matrices between time 0 and 24 months are displayed in Appendix A. The p-values reported were based on the χ^2 approximation.

The p-values were smallest (although not significant) at six and nine months, which is when the treatment was discontinued. This suggests that the treatment may have had an effect at this time, which wore off at eighteen months. Since the statistics calculated only measure absolute differences it was not possible to determine from them whether the treatment group did better or worse than the control group. No obvious trend could be seen from examination of the estimated transition matrices, but previous work [7] indicated that in fact the treatment group regressed (relative to the controls) from 0 to 6 months.

2.1.1 Tests of Order of Markov Chain

Second order and first order Markov Chain models were fit to the data and the resultant estimates were compared. Each group was tested separately. A likelihood ratio statistic

was used to test H_0 : chain is 1st order Markov vs. H_1 : chain is 2nd order Markov. This test was carried out separately for every three consecutive time points in the study, i.e. at 0-1-3 months, 1-3-6 months, etc. If the data appeared to act as a 2nd order Markov chain then that would imply that a patient's score would depend on his score both at the previous time point and the time point before that. In this problem, the G^2 statistic becomes:

$$G_2^2 = 2 \sum_{i=1}^z \sum_{j=1}^z \sum_{k=1}^z n_{ijk}^t \ln \left[\frac{n_{ijk}^t \times n_{j.}^t}{n_{ij.}^t \times n_{.jk}^t} \right]$$

where

$$\begin{aligned} n_{ijk}^t &= \text{number of observations in which } X_{t-2} = i, X_{t-1} = j, X_t = k \\ n_{jk}^t &= \text{number of observations in which } X_{t-1} = j, X_t = k \\ n_{j.}^t &= \sum_{k=1}^z n_{jk}^t \\ n_{ij.}^t &= \sum_{k=1}^z n_{ijk}^t \\ t &= 2, \dots, z. \end{aligned}$$

G_2^2 has an approximate chi squared distribution with $\sum_{j=1}^z (z - r_j - 1) \times (z - c_j - 1)$ degrees of freedom where z is the number of categories, c_j and r_j are the number of zero rows and columns respectively in the two dimensional transition matrix consisting of the transition probabilities p_{ijk}^t ($i = 1, \dots, z$, j is fixed and $k = 1, \dots, z$).

The results from these tests are in Table 2.2. Only the statistic for the treatment group in the time interval six to twelve months was significant at an α level of 0.05, which indicated that modelling with a 1st order Markov chain was reasonable for most of the time periods. The counts in these matrices were very low, so the statistics calculated may be misleading.

Similar tests were applied to the 1st order Markov transition matrices to determine whether the final response depended on the previous score (1st order Markov) or not

(independence). The appropriate G^2 statistic becomes:

$$G_3^2 = 2 \sum_{i=1}^z \sum_{j=1}^z n_{ij}^t \ln \left[\frac{n_{ij}^t \times n^t}{n_{i.}^t \times n_{.j}^t} \right] \quad (2.4)$$

where

$$n_{.j}^t = \sum_{i=1}^z n_{ij}^t$$

$$n^t = \sum_{i=1}^z \sum_{j=1}^z n_{ij}^t.$$

The statistics calculated using (2.4) are in Table 2.3. All of the statistics were very large in comparison with the degrees of freedom at all time points which meant that the reduction from a 1st order Markov chain model to an independence model was not reasonable.

2.1.2 Tests of Stationarity

Assuming the data had the Markov Property, another likelihood ratio statistic was computed, testing for stationarity. We assume that the same transition matrix could be used to describe the patients' movements between all of the time intervals measured, independent of the actual real-time length of that time interval. A transition matrix was estimated using data from all time points. This was compared to the seven matrices estimated with the counts for every two consecutive time points in the study (0-1 month, 1-3 months, etc.). The G^2 statistic here takes on the form:

$$G_4^2 = 2 \sum_{i=1}^z \sum_{j=1}^z \sum_{t=1}^7 n_{ij}^t \ln \left[\frac{n_{ij}^t \times n_{i.}}{n_{i.}^t \times n_{ij}} \right]$$

where

$$n_{i.} = \sum_{j=1}^z \sum_{t=1}^7 n_{ij}^t$$

$$n_{ij} = \sum_{t=1}^7 n_{ij}^t.$$

Table 2.1: Comparison of Treatment and Control Groups

Time Period (months)	G^2 statistic	p value (G^2)	X^2 statistic	p value (X^2)	degrees of freedom
0-1	12.386	0.19	10.718	0.30	9
0-3	15.086	0.13	12.745	0.24	10
0-6	22.050	0.08	17.387	0.24	14
0-9	17.985	0.08	14.931	0.19	11
0-12	17.502	0.13	14.239	0.29	12
0-18	12.621	0.32	10.683	0.47	11
0-24	14.072	0.44	11.825	0.62	14

Table 2.2: Tests of 2nd Order vs. 1st Order Markov Process

	Time Period (months)	G_2^2 statistic	degrees of freedom	p-value
Treatment Group	0 \rightarrow 1 \rightarrow 3	25.400	19	0.15
	1 \rightarrow 3 \rightarrow 6	22.141	16	0.14
	3 \rightarrow 6 \rightarrow 9	20.006	15	0.17
	6 \rightarrow 9 \rightarrow 12	21.161	11	0.03
	9 \rightarrow 12 \rightarrow 18	3.589	12	0.99
	12 \rightarrow 18 \rightarrow 24	29.795	24	0.19
Control Group	0 \rightarrow 1 \rightarrow 3	18.429	12	0.10
	1 \rightarrow 3 \rightarrow 6	5.997	12	0.92
	3 \rightarrow 6 \rightarrow 9	10.104	9	0.34
	6 \rightarrow 9 \rightarrow 12	19.101	13	0.12
	9 \rightarrow 12 \rightarrow 18	11.920	11	0.37
	12 \rightarrow 18 \rightarrow 24	11.251	17	0.84

The results are in Table 2.4. The statistics for both groups were nonsignificant (at $\alpha = 0.05$) which implied that this model was not unreasonable for the time period studied.

2.2 Modelling of the Tridiagonals

Specific modelling of the tridiagonals was done as most transitions were made to neighbouring Kurtzke scores. In this section, all transitions made to non-neighboring states were ignored and transition probabilities not in a tridiagonal position were assumed to be zero; that is, $p_{ij} = 0$ if $|i - j| > 2$. In the following calculations, the transition matrices were assumed to be stationary and first order Markov. The observations on the tridiagonals were modelled first using a general form where each row had different entries, then using a more specific form which was suggested by the data.

In this section, the uneven spacing of the time points will be taken into account. The data at the one month time point were omitted so that the remaining time points (0,3,6,9,12,18,24) were separated by intervals which were multiples of three months. This allowed modelling of transitions over a three month period using the matrix, \mathcal{P} . Thus, transitions over a six month period could be modelled using \mathcal{P}^2 . To adjust for this omission, redefine $t = 0, 1, \dots, 6$ corresponding to 0, 3, 6, 9, 12, 18, and 24 months.

Subsection 2.2.1 discusses modelling the tridiagonals using the matrices \mathcal{P} and \mathcal{P}^2 to describe transitions occurring over three and six month periods respectively. In Section 2.2.2, models are fitted assuming that the same transition matrix, P , can be used to model both three and six month periods. The entire time interval and that during treatment (0, 3, 6 months) and after treatment (6, 9, 12, 18, 24 months) were modelled separately so that the estimates could be compared.

Some other observations were omitted from the analysis because the transitions were

larger than those allowed by the model. In this case, data from the subject in question was used until the violation occurred. A subject with a missing observation was dealt with similarly. Approximately 25% of the subjects in each group had a missing value at some time point after six months. Only six patients in the treatment group and three in the control group had transitions which were larger than those allowed by the model.

2.2.1 Modelling Incorporating \mathcal{P}^2

Transition matrices for treatment (\mathcal{P}^T) and control groups (\mathcal{P}^C) were estimated and compared to a matrix estimated under the null hypothesis $\mathcal{P}^C = \mathcal{P}^T$. The data were first modelled using the general matrix, \mathcal{P} :

$$\begin{bmatrix} p_{11} & 1 - p_{11} & 0 & 0 & 0 \\ p_{21} & p_{22} & 1 - p_{21} - p_{22} & 0 & 0 \\ 0 & p_{32} & p_{33} & 1 - p_{32} - p_{33} & 0 \\ 0 & 0 & p_{43} & p_{44} & 1 - p_{43} - p_{44} \\ 0 & 0 & 0 & 1 - p_{55} & p_{55} \end{bmatrix}.$$

The likelihood function $L(\mathcal{P})$ was as follows:

$$L(\mathcal{P}) = K \times \prod_{i,j} p_{ij}^{n_{ij}} \cdot (P_{ij}^2)^{N_{ij}}$$

where p_{ij} was redefined as:

$$p_{ij} = \Pr[X_t = j \mid X_{t-1} = i] \text{ for } t = 1, 2, 3, 4$$

and

$$n_{ij} = \sum_{t=1}^4 n_{ij}^t$$

$$N_{ij} = \sum_{t=5}^6 n_{ij}^t$$

$$P_{ij}^2 = \Pr[X_t = j \mid X_{t-1} = i] \text{ for } t = 5, 6$$

$$K = \text{a constant independent of } p_{ij}'\text{s.}$$

To calculate the m.l.e.'s, it was necessary to solve eight nonlinear equations. This was accomplished using Powell's method, described in [5]. The G^2 statistic calculated to compare the two groups was:

$$G_5^2 = \sum_{i=1}^z \sum_{j=1}^z \left[n_{ij}^T \ln \left(\frac{\hat{p}_{ij}^T}{\hat{p}_{ij}} \right) + n_{ij}^C \ln \left(\frac{\hat{p}_{ij}^C}{\hat{p}_{ij}} \right) + N_{ij}^T \ln \left(\frac{\hat{P}_{ij}^{T2}}{\hat{P}_{ij}^2} \right) + N_{ij}^C \ln \left(\frac{\hat{P}_{ij}^{C2}}{\hat{P}_{ij}^2} \right) \right] \quad (2.5)$$

where $n_{ij}^T, n_{ij}^C, N_{ij}^T, N_{ij}^C$ were defined to correspond to counts in the treatment and control groups in the obvious way. Results of the test are in Table 2.5.

In the time period during the treatment, G_5^2 was large relative to the degrees of freedom. The low p values (calculated using the χ^2 approximation) indicated that, during this time period, the two groups' transition probability matrices were different. In the follow up period after treatment, and the entire two year period, there was no evidence that the two groups behaved differently. G^2 statistics were calculated to determine how well the models fit the data (see Table 2.6). These statistics indicated that the fit of the model was reasonable at all of the time intervals modelled.

2.2.2 Modelling Without \mathcal{P}^2

Since the "stationarity" test seemed to indicate that all consecutive intervals in the study could be modelled using the same matrix, the above analysis was repeated under this assumption with the data gathered at one month again omitted. The period during treatment did not include any six month intervals so estimates are the same as in the previous section. G_6^2 , the log likelihood statistic comparing the two groups, was similar to that calculated in (2.3). The results, displayed in Table 2.7, show that the groups could be modelled reasonably according to the same transition probability matrix after treatment. Over the two year period, the statistics indicated that the transition probability matrices for the two groups were different, due probably to the differences which were detected in the first six months of the study (see Table 2.5).

G^2 statistics for the period after treatment and the whole time studied are displayed in Table 2.8. These show that the predictions from the models agreed reasonably well with the actual data.

After further examination of the data it was noted that most of the diagonal elements were similar. A pattern among the off diagonal elements was noticed also. To determine whether the data could be modelled with four parameters instead of the eight used above, the previous calculations were repeated, this time using a transition probability matrix suggested by the data, namely:

$$\begin{bmatrix} p_{11} & 1 - p_{11} & 0 & 0 & 0 \\ p_{21} & p_{22} & 1 - p_{21} - p_{22} & 0 & 0 \\ 0 & p_{32} & p_{11} & 1 - p_{32} - p_{11} & 0 \\ 0 & 0 & 1 - p_{32} - p_{11} & p_{11} & p_{32} \\ 0 & 0 & 0 & 1 - p_{11} & p_{11} \end{bmatrix}.$$

The estimates computed for the specific and general tridiagonal models were compared for each group separately using G_7^2 :

$$G_7^2 = 2 \sum_{i=1}^z \sum_{j=1}^z n_{ij}^* \ln \left(\frac{\hat{p}_{ij}^G}{\hat{p}_{ij}^S} \right)$$

where

$$n_{ij}^* = \sum_{t=1}^6 n_{ij}^t$$

$$\hat{p}_{ij}^G = \text{m.l.e. of } p_{ij} \text{ under the hypothesis of a} \\ \text{general tridiagonal matrix}$$

$$\hat{p}_{ij}^S = \text{m.l.e. of } p_{ij} \text{ under the hypothesis of the} \\ \text{specific tridiagonal matrix above.}$$

The results are shown in Table 2.9 to 2.10 and estimates for the specific and general matrices are provided in the Appendix. The statistics for the control group indicate that the general model does not fit significantly better (at an α level of 0.05) than the specific model over the two years studied. The specific matrix estimated for the treatment group however, was only reasonable for the follow up period. As it was of interest whether the treatment and control group could be modelled using the same matrix after treatment, this matrix was compared to that for the control group (over the two year period) and they were found to be significantly different. This difference may be due to the fact that the specific matrix for the control group does not fit the data particularly well. A chi-squared statistic for the goodness of fit was 6.056 ($p=0.195$).

From an examination of the matrices, it appears that in the control group, those patients in states other than two were more likely to stay in that state. If they were in state two, however, they were more likely to move to state one. This was also true for patients in the treatment group after six months. During administration of the drug, treatment subjects were more likely to move to a higher state (i.e. to become sicker) than the control group.

All of these analyses indicate that the progress of the disease in the treatment group was not the same as in the control group in the first six months of the study. In the follow up period, there was no evidence to indicate that any of the groups was worse off than the other and over the entire time period studied there seemed to be little difference between the two. A model assuming the transition probability matrix was first order Markov and stationary seemed to fit the data reasonably well. The modelling of the tridiagonal elements revealed that a specific model in which all diagonal elements were the same except for the second, appeared to fit the data. From the matrices estimated using this model, it is seems that patients with scores of 4.5

to 5.5 are more likely to move to an adjacent score than patients in any of the other categories used. However, it should be noted that the collapsed categories used in this analysis were somewhat arbitrary and may produce misleading results. As well, many of the cells in the transition matrices tested were zero, which could produce statistics that do not reflect the data accurately.

Table 2.3: Tests of 1st Order vs. Independence

	Time Period (months)	G_3^2 statistic	degrees of freedom
Treatment Group	0 → 1	52.025	16
	1 → 3	42.851	16
	3 → 6	53.937	16
	6 → 9	46.017	16
	9 → 12	59.955	16
	12 → 18	51.898	16
	18 → 24	49.019	16
Control Group	0 → 1	58.278	16
	1 → 3	44.748	16
	3 → 6	55.247	16
	6 → 9	45.811	16
	9 → 12	52.629	16
	12 → 18	51.747	16
	18 → 24	63.797	16

Table 2.4: Results of Test for Stationarity

Group	G_4^2	df	p value
treatment	88.921	84	0.34
control	65.198	78	0.85

Table 2.5: Treatment vs. Control (general tridiagonal modelling)

Time (months)	G_5^2	p-value
0-3-6	19.250	0.01
6-9-12-18-24	4.283	0.83
0-3-6-9-12-18-24	11.415	0.18

Table 2.6: Fit of General Tridiagonal Models

Time (in months)	Treatment Group (G^2)	degrees of freedom	p value	Control Group (G^2)	degrees of freedom	p value
0-3-6	1.22	8	0.99	3.73	8	0.88
6-9-12-18-24	14.75	24	0.93	25.91	24	0.36
0-3-6-9-12-18-24	43.76	40	0.32	42.80	40	0.35

Table 2.7: Treatment vs. Control (general tridiagonal - no P^2)

Time (in months)	G_6^2	p-value
6-9-12-18-24	3.456	0.484
0-3-6-9-12-18-24	9.642	0.047

Table 2.8: Fit of General Model - no P^2

Time (in months)	Treatment Group	degrees of freedom	p value	Control Group	degrees of freedom	p value
6-9-12-18-24	16.74	8	0.86	27.66	24	0.27
0-3-6-9-12-18-24	44.08	40	0.30	43.55	40	0.32

Table 2.9: Comparison of Specific and General Models (treatment)

Time (in months)	G_7^2	Treatment degrees of freedom	p-value
0-3-6	11.815	4	0.018
6-9-12-18-24	3.919	4	0.417
0-3-6-9-12-18-24	14.7211	4	0.005

Table 2.10: Comparison of Specific and General Models

Time (in months)	G_7^2	Control degrees of freedom	p-value
0-3-6	6.531	3	0.088
6-9-12-18-24	6.017	4	0.197
0-3-6-9-12-18-24	5.565	4	0.234

Chapter 3

Analysis At One Time Point

A typical analysis involves comparing the scores of the patients in the treatment and control group at one time point. Commonly used for this are the 2 sample t statistic (although inappropriate) and the Wilcoxon rank sum statistic. Here, these statistics were compared to a McCullagh model statistic described in [4]. The McCullagh model can be used to analyze data with ordered, categorical responses. The form used in this chapter is the proportional odds model:

$$\log \left[\frac{\gamma_j(\vec{x})}{1 - \gamma_j(\vec{x})} \right] = \frac{\theta_j - \vec{\beta}^T \vec{x}}{\kappa} \quad (3.6)$$

where

$$j = 1, 2, \dots, z$$

$$z = \text{total number of categories}$$

$$\vec{x} = \text{vector of covariates}$$

$$\gamma_j(\vec{x}) = \text{probability of being in category } j \text{ or lower given covariate } \vec{x}$$

$$\theta_j = \text{cutpoint } j, \text{ the (unknown) point separating the categories } j \text{ and } j + 1$$

$$\vec{\beta} = \text{a vector of unknown parameters}$$

$$\kappa = \text{a scale parameter.}$$

The relative efficiencies of the three tests were compared using simulation and efficacy calculations. Data for the treatment and control group were simulated by first generating normal data with different location parameters. This continuous data was

used to create ordered categorical data according to a chosen set of cutpoints. Tests for differences in the two groups were then carried out using the 2 sample t, Wilcoxon rank sum and McCullagh statistic. The 2 sample t test was also calculated on the uncategorized data so that a test which did not lose information due to the categorization could be compared to the others. These results are given in Section 3.1. Efficacy calculations were also made to determine how different the tests were asymptotically. The efficacies of the Wilcoxon rank sum test are calculated in Section 3.2.1, and for the t test calculated on the categorical data in Section 3.2.2. These efficacies are compared to the t test calculated on the underlying continuous data in Section 3.3.

3.1 Simulation

A simulation was run to compare the powers of the Wilcoxon rank sum statistic, both 2 sample t tests and McCullagh model estimates. The control and treatment group were simulated using $N(0,1)$ and $N(\Delta,1)$ distributions respectively to produce underlying continuous responses. These responses were then categorized with cut points chosen so that the probability of being in any category was 0.2 if the data came from a $N(0,1)$ distribution. Fifty observations (the same number as were in the Multiple Sclerosis data) were generated for each group. The simulations involved one thousand replications at each value of Δ . The Δ values used were those from 0 to 1 in increments of 0.1. The data were generated without conditioning on the number of observations in each category as this was the form of the Multiple Sclerosis data.

Power curves were calculated for a one sided 0.05 level test of the null hypothesis $H_0 : \Delta = 0$ vs. the alternate hypothesis $H_1 : \Delta > 0$. The results of the simulation appear in Figure 3.1. As expected, the 2 sample t test on the underlying continuous data was the most powerful because no information was lost through collapsing the

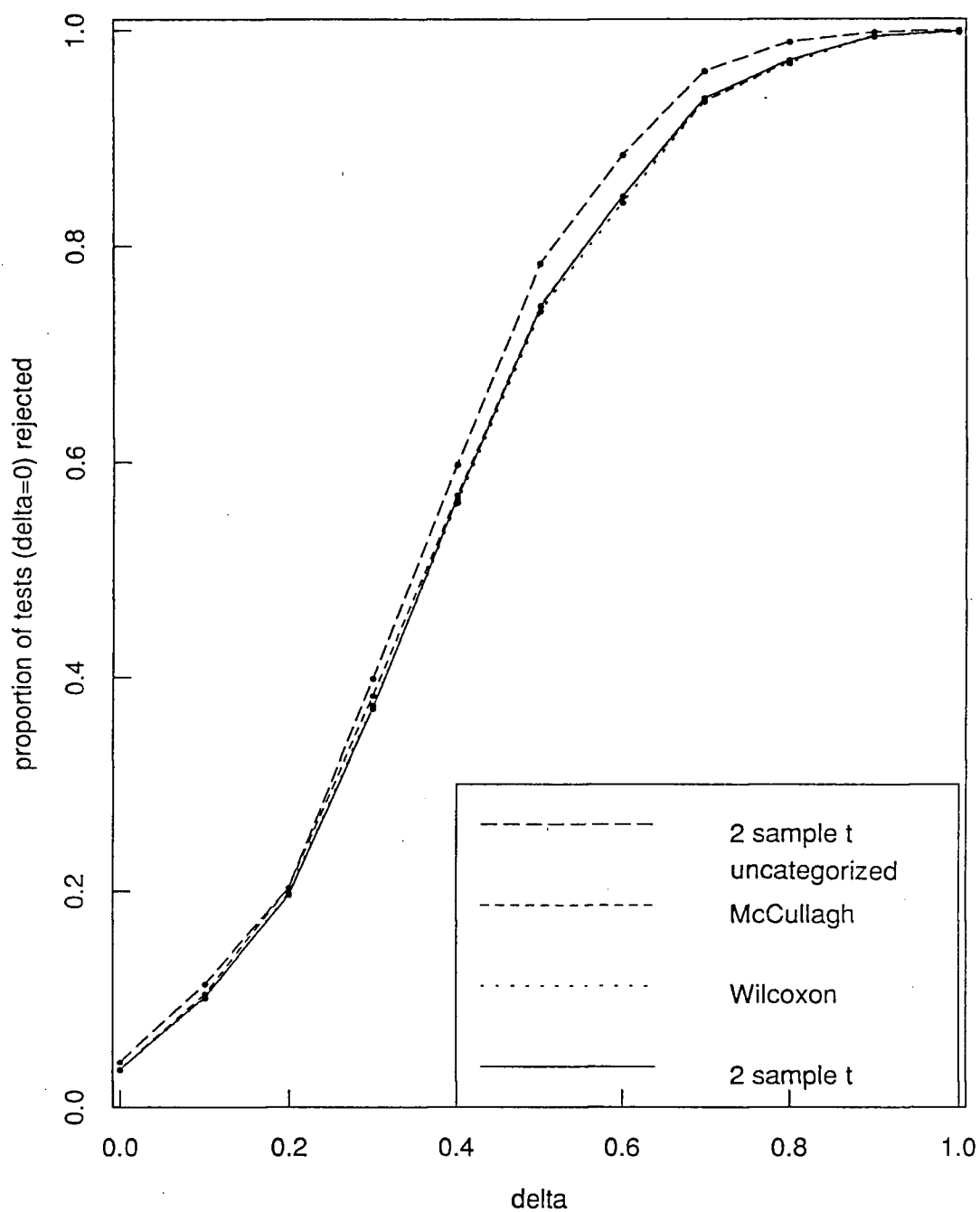


Figure 3.1: Simulation of Data at One Time Point

data. The Wilcoxon rank sum statistic and the 2 sample t test on the categorized data were very similar at all points. For the results summarized in Figure 3.1, κ , the scale parameter for the McCullagh analysis was set to one. The covariate \vec{x} was an indicator variable distinguishing the treatment from the control observations. In this case the test based on the McCullagh statistic was as powerful as the other tests on the categorized data. If κ was included as a parameter depending on covariate \vec{x} , it was not significantly different from 1 \sim 92% of the time. However, including this parameter decreased the power of the McCullagh statistic for values of Δ larger than 1.5.

Differences in the simulated power between any of the models were not very large. The maximum difference was ~ 0.05 between the 2 sample t test calculated using the continuous data and the McCullagh model at $\Delta = 0.7$, where the power of the t test is about 0.95. Given that the standard error of this difference was ~ 0.02 , the t test was significantly more powerful than the McCullagh model at this time point.

3.2 Efficacy Calculations

Another way of comparing the tests is to calculate their asymptotic relative efficiencies. The method of calculation used is that in Lehmann [3].

Let V_N, V'_N be two sequences of statistics based on N observations. Assume the distributions of both V_N and V'_N depend on a real valued parameter θ . Let the null hypothesis be $\theta = \theta_0$ and the alternate hypothesis be $\theta > \theta_0$ as in the simulation. Define β_N to be the power of the test which rejects H_0 if

$$\frac{V_N - \mu(\theta_0)}{\sigma_N(\theta_0)} \geq c_N \quad (3.7)$$

and β'_N to be the power of the test based on V'_N which rejects H_0 if

$$\frac{V'_N - \mu'(\theta_0)}{\sigma'_N(\theta_0)} \geq c'_N \quad (3.8)$$

where $\mu(\theta_0)$, $\sigma_N(\theta_0)$ and $\mu'(\theta_0)$, $\sigma'_N(\theta_0)$ are normalizing constants which may be the expectation and standard deviation of V_N and V'_N respectively, and c_N and c'_N are sequences of critical values.

Assume θ_N is a sequence of alternatives, converging to θ_0 in such a way that $\theta_N = \theta_0 + \frac{\Delta}{\sqrt{N}}$. For most commonly used tests, this condition is sufficient to ensure that $\beta_N(\theta_N) \rightarrow \beta_0$, $0 \leq \beta_0 \leq 1$. Define N' to be the sample size required for $V'_{N'}$ to achieve the same limiting power as V_N against the same sequence of alternatives θ_N . If c_N , $c_{N'} \rightarrow z_\alpha$ (the $(1 - \alpha)^{th}$ quantile of the standard normal distribution) as $N \rightarrow \infty$, then the Pitman efficiency of V_N relative to $V'_{N'}$ is $\lim_{N \rightarrow \infty} (N'/N)$.

Suppose that whenever $\theta_N = \theta_0 + \frac{\Delta}{\sqrt{N}}$,

$$\beta_N(\theta_N) \rightarrow \Phi(c\Delta - z_\alpha)$$

and

$$\beta'_N(\theta_N) \rightarrow \Phi(c'\Delta - z_\alpha).$$

Then c and c' are called the efficacies of the tests based on V_N and $V'_{N'}$ and the Pitman efficiency of V_N relative to $V'_{N'}$ is:

$$(c/c')^2$$

3.2.1 Efficacy of the Wilcoxon

Suppose the ordered categorical variable takes on values $g_1 < g_2 < \dots < g_z$, where z is the number of categories. In this analysis, observations tied in a category were assigned midranks, so the Wilcoxon rank sum statistic (with ties) was:

$$W = T_1(N_1 + 1)/2 + T_2(N_1 + (N_2 + 1)/2) + \dots + T_z\left(\sum_{j=1}^{z-1} N_j + (N_z + 1)/2\right) \quad (3.9)$$

where

T_j = number of observations in the treatment group in category j

C_j = number of observations in the control group in category j

and

$$N_j = T_j + C_j.$$

Let

$q_j = \Pr[\text{in category } j \mid \text{in control group}]$

$p_j = \Pr[\text{in category } j \mid \text{in treatment group}]$

N =total number of observations.

The test based on the Wilcoxon involved normalizing W as follows:

$$\frac{W - E_o(W)}{\sqrt{\text{var}_o^*(W)}},$$

where $E_o(W)$ is the expectation of the Wilcoxon under $H_0 : q_j = p_j \forall j$, and $\text{var}_o^*(W)$ is the variance of the Wilcoxon under H_0 conditional on the number of tied observations in each category.

The power of the test at a specific alternative could have been calculated either unconditionally or conditionally on the tied observations. In this case, unconditional power was used as it simplified the efficacy calculations of the t test and eliminated the need to generate a fixed number of observations in each category for the simulations.

The unconditional power of the test using the Wilcoxon statistic was:

$$\beta_N(\vec{p}, \vec{\delta}) = \Pr \left[\frac{W - E_o(W)}{\sqrt{\text{var}_o^*(W)}} > z_\alpha \right]$$

where

$$\begin{aligned}
E_o^*(W) &= E_o(W) = \frac{n(N+1)}{2} \\
var_o^*(W) &= \frac{mn(N+1)}{12} - \frac{mn}{12N(N-1)} \sum_{j=1}^z (N_j^3 - N_j) \\
\vec{\delta} &= \sqrt{N}(\vec{q} - \vec{p}) \\
m &= \text{total number of observations in the control group} \\
n &= \text{total number of observations in the treatment group} \\
N &= \text{total number of observations in the study} \\
&= n + m.
\end{aligned}$$

Theorem 3.2.1 Suppose $(T_1, T_2, \dots, T_z) \sim \text{multinomial}(n, p_1, p_2, \dots, p_z)$ independent of $(C_1, C_2, \dots, C_z) \sim \text{multinomial}(m, q_1, q_2, \dots, q_z)$ and $N \rightarrow \infty$ in such a way that $n/N \rightarrow a, m/N \rightarrow b, 0 < a < 1$. Then, as $N \rightarrow \infty$,

$$\beta_N(\vec{p}) \rightarrow \Phi(f(\vec{p}, \vec{\delta}) - z_\alpha) \quad (3.10)$$

where

$$f(\vec{p}, \vec{\delta}) = \frac{\sqrt{ab} \left(\sum_{j=1}^{z-1} \delta_j + \sum_{j=1}^{z-1} \sum_{i=1}^{j-1} p_j \delta_i - \sum_{j=1}^{z-1} \sum_{i=j+1}^{z-1} p_j \delta_i \right)}{\sqrt{\sum_{j=1}^{z-1} r_j^2 p_j - \sum_{i=1}^{z-1} \sum_{j=1}^{z-1} r_i r_j p_i p_j}}$$

and

$$r_j = 1 + \sum_{i=j+1}^{z-1} p_j - \sum_{i=1}^{j-1} p_i, \text{ defining } \sum_{i=1}^{l-1} p_i \text{ to be 0 for any integer } l.$$

In the simulation run of Section 3.1 $p_1 = p_2 = \dots = p_z$ and the underlying continuous distribution was known. The following corollaries state the efficacies under those conditions and the proof of Theorem 3.2.1 follows.

Corollary 3.2.1 Assume that the multinomial vectors \vec{T} and \vec{C} are generated from $X_{T_1}, X_{T_2}, \dots, X_{T_n} \sim \text{iid } N(\Delta, 1)$ and $X_{C_1}, X_{C_2}, \dots, X_{C_m} \sim \text{iid } N(0, 1)$ in the usual way. That is, an observation is in category i if and only if the X value is between cutpoints θ_i and θ_{i-1} where $\theta_o = -\infty$ and $\theta_z = \infty$. Then under the assumptions of Theorem 3.2.1, and with $\Delta = \frac{\delta}{\sqrt{N}}$,

$$\beta_N(\Delta) \rightarrow \Phi(c \cdot \delta - z_\alpha)$$

where

$$c = \frac{\sqrt{ab} \left[\sum_{j=1}^{z-1} d_j + \sum_{j=1}^{z-1} \sum_{i=1}^{j-1} p_j d_i - \sum_{j=1}^{z-1} \sum_{i=j+1}^{z-1} p_j d_i \right]}{\sqrt{\sum_{j=1}^{z-1} r_j^2 p_j - \sum_{i=1}^{z-1} \sum_{j=1}^{z-1} r_i r_j p_i p_j}}$$

$$d_j = u(\theta_j) - u(\theta_{j-1}) \text{ for } j = 1, \dots, z-1$$

$$u(\theta_j) = \text{probability density function of the standard normal distribution at } \theta_j.$$

Corollary 3.2.2 *In the case where $X_{T_1} - \Delta, X_{T_2} - \Delta, \dots, X_{T_n} - \Delta$ and $X_{C_1}, X_{C_2}, \dots, X_{C_m}$ are iid observations with a continuous density, h , and $p_1 = p_2 = \dots = p_z$ then, under the assumptions of Theorem 3.2.1., the efficacy of the Wilcoxon becomes:*

$$\frac{\sqrt{ab} \sum_{j=1}^{z-1} (z-j) e_j}{\sqrt{(z^2-1)/12}}$$

where

$$\Delta = \delta / \sqrt{N}$$

$$e_j = h(\theta_j) - h(\theta_{j-1}) \text{ } j = 1, 2, \dots, z-1$$

$$\theta_j = \text{cutpoint } j.$$

The above results do not depend upon the scores associated with the categories.

Proof of Theorem 3.2.1:

$$\begin{aligned} \beta_N(\vec{p}, \vec{\delta}) &= \Pr \left[\frac{W - E_o(W)}{\sqrt{\text{var}_o^*(W)}} > z_\alpha \right] \\ &= \Pr \left[\frac{W - E(W)}{\sqrt{mnN}} > z_\alpha \sqrt{\frac{\text{var}_o^*(W)}{mnN}} - \frac{E(W) - E_o(W)}{\sqrt{mnN}} \right] \end{aligned}$$

It will be shown that if $N \rightarrow \infty$ in such a way that $\frac{n}{N} \rightarrow a$, $\frac{m}{N} \rightarrow b$, then

$$\begin{aligned} 1. \quad & \frac{W - E(W)}{\sqrt{mnN}} \xrightarrow{d} N(0, \sigma_W^2) \\ & \sigma_W^2 = \frac{1}{4} \left[\sum_{j=1}^{z-1} r_j^2 p_j - \sum_{i=1}^{z-1} \sum_{j=1}^{z-1} r_i r_j p_i p_j \right] \\ & r_i = 1 + \sum_{j=i+1}^{z-1} p_j - \sum_{j=1}^{i-1} p_j \end{aligned}$$

$$2. \frac{\text{var}_0^*(W)}{mnN\sigma_W^2} \rightarrow 1$$

$$3. \frac{E(W) - E_0(W)}{\sqrt{mnN\sigma_W^2}} \rightarrow f(\vec{p}, \vec{\delta}).$$

1. To simplify the calculations, substitute $m - \sum_{i=1}^{z-1} C_i$ for C_z and $n - \sum_{i=1}^{z-1} T_i$ for T_z in the expression for the Wilcoxon (refeq:wilc) to obtain:

$$\begin{aligned} \frac{W - E(W)}{\sqrt{mnN}} &= \frac{1}{2\sqrt{mnN}} \left[n \sum_{i=1}^{z-1} C_i - m \sum_{j=1}^{z-1} T_j + \sum_{j=1}^{z-1} \sum_{i=1}^{j-1} T_j C_i - \sum_{j=1}^{z-1} \sum_{i=j+1}^{z-1} T_j C_i \right. \\ &\quad \left. - \left(\frac{mn}{\sqrt{N}} \sum_{j=1}^{z-1} \delta_j + \sum_{j=1}^{z-1} \sum_{i=1}^{j-1} p_j \delta_i - \sum_{j=1}^{z-1} \sum_{i=j+1}^{z-1} p_j \delta_i \right) \right] \end{aligned}$$

Let

$$\begin{aligned} C_i^* &= \frac{C_i - mq_i}{\sqrt{m}} \\ T_i^* &= \frac{T_i - np_i}{\sqrt{n}}. \end{aligned}$$

Then,

$$\frac{W - E(W)}{\sqrt{mnN}} = \frac{1}{2} \left[\sqrt{\frac{n}{N}} \sum_{i=1}^{z-1} r_i C_i^* - \sqrt{\frac{m}{N}} \sum_{i=1}^{z-1} t_i T_i^* + \frac{1}{\sqrt{N}} \sum_{j=1}^{z-1} \sum_{i=1}^{j-1} (T_j^* C_i^* - T_i^* C_j^*) \right].$$

Since T_j^*, C_j^* converge in distribution to normal random variables with mean 0 and variance $p_j(1 - p_j)$, as $N \rightarrow \infty$,

$$\frac{1}{\sqrt{N}} \sum_{j=1}^{z-1} \sum_{i=1}^{j-1} (T_j^* C_i^* - T_i^* C_j^*) \rightarrow 0$$

Thus as $N \rightarrow \infty$,

$$\frac{W - E(W)}{\sqrt{mnN}} \sim \frac{1}{2} \left(\sqrt{a} \sum_{i=1}^{z-1} r_i C_i^* - \sqrt{b} \sum_{i=1}^{z-1} s_i T_i^* \right)$$

where

$$r_i = 1 + \sum_{j=i+1}^{z-1} p_j - \sum_{j=1}^{i-1} p_j$$

$$s_i = 1 + \sum_{j=i+1}^{z-1} q_j - \sum_{j=1}^{i-1} q_j$$

defining $\sum_{j=m}^{m-1} q_j = 0$ and $\sum_{j=m}^{m-1} p_j = 0$. Since $\text{cov}(C_i^*, C_j^*) = -q_i q_j$ and $\text{cov}(T_i^*, T_j^*) = -p_i p_j$, $\frac{W - E(W)}{\sqrt{mnN}}$ converges to a normally distributed random variable with mean 0 and variance σ_W^2 :

$$\sigma_W^2 = \frac{1}{4} \left(\sum_{j=1}^{z-1} r_j^2 p_j - \sum_{i=1}^{z-1} \sum_{j=1}^{z-1} r_i r_j p_i p_j \right)$$

2. Rewriting the ratio as

$$\frac{\text{var}_o^*(W)}{mnN\sigma_W^2} = \frac{\text{var}_o^*(W)}{\text{var}_o(W)} \times \frac{\text{var}_o(W)}{mnN\sigma_W^2},$$

it will be shown that both ratios on the right hand side converge to 1.

For the first ratio,

$$\begin{aligned} \text{var}_o(W) &= E_o[\text{var}_o^*(W)] + \text{var}_o[E_o^*(W)] \\ &= E_o[\text{var}_o^*(W)] \end{aligned}$$

since $E_o^*(W)$ is a constant. So

$$\frac{\text{var}_o^*(W)}{mnN\sigma_W^2} = \frac{\text{var}_o^*(W)}{E_o[\text{var}_o^*(W)]} \times \frac{\text{var}_o(W)}{mnN\sigma_W^2}.$$

But

$$\begin{aligned} \text{var}_o^* \left(\frac{W}{\sqrt{N^3}} \right) &= \frac{1}{N^3} \text{var}_o^*(W) \\ &= \frac{1}{N^3} \left[\frac{mn(N+1)}{12} - \frac{mn}{12N(N-1)} \sum_{j=1}^z (N_j^3 - N_j) \right] \\ &\sim \frac{ab}{12} \left[1 - \sum_{j=1}^z \left(\frac{N_j^3}{N^3} \right) \right] \text{ as } N \rightarrow \infty, \text{ and} \\ &= \frac{ab}{12} \left[1 - \sum_{j=1}^z \left(\frac{T_j + C_j}{N} \right)^3 \right] \end{aligned}$$

$$\begin{aligned}
& \rightarrow \frac{ab}{12} \left[1 - \sum_{j=1}^z (ap_j + bq_j)^3 \right] \\
& \rightarrow \frac{ab}{12} \left[1 - \sum_{j=1}^z p_j^3 \right], \text{ under } H_0.
\end{aligned}$$

Thus, $\text{var}_o^* \left(\frac{W}{\sqrt{N^3}} \right)$ converges to a constant, say y , and is bounded and positive. Hence, $E \left[\text{var}_o^* \left(\frac{W}{\sqrt{N^3}} \right) \right]$ also converges to y , and so

$$\frac{\text{var}_o^*(W)}{E[\text{var}_o^*(W)]} \rightarrow 1.$$

Now we need to prove that

$$\frac{\text{var}_o(W)}{mnN\sigma_W^2} \rightarrow 1.$$

Recalling that under H_o , $p_j = q_j \forall j$, and so $r_j = s_j \forall j$ it can be shown that

$$\begin{aligned}
\text{var}_o \left(\frac{W - E(W)}{\sqrt{mnN}} \right) & \rightarrow \frac{1}{4} \left[\sum_{j=1}^{z-1} r_j^2 p_j - \sum_{i=1}^{z-1} \sum_{j=1}^{z-1} r_i r_j p_i p_j \right] \\
& = \sigma_W^2.
\end{aligned}$$

So

$$\frac{\text{var}_o(W)}{mnN\sigma_W^2} \rightarrow 1$$

as $N \rightarrow \infty$.

3. Finally, using parts 1 and 2, the efficacy term can be shown to converge to a constant:

$$\begin{aligned}
& \frac{E(W) - E_o(W)}{\sqrt{mnN\sigma_W^2}} \rightarrow \\
& \frac{\sqrt{ab} \left(\sum_{j=1}^{z-1} \delta_j + \sum_{j=1}^{z-1} \sum_{i=1}^{j-1} p_j \delta_i - \sum_{j=1}^{z-1} \sum_{i=j+1}^{z-1} p_j \delta_i \right)}{\sqrt{\sum_{j=1}^{z-1} r_j^2 p_j - \sum_{i=1}^{z-1} \sum_{j=1}^{z-1} r_i r_j p_i p_j}} \\
& = f(\vec{p}, \vec{\delta}).
\end{aligned}$$

So now we have that

$$\beta_N(\vec{p}, \vec{\delta}) \rightarrow \Pr \left[N(0, 1) < f(\vec{p}, \vec{\delta}) - z_\alpha \right].$$

Proof of Corollary 3.2.1. Note that,

$$q_j = \Phi(\theta_j) - \Phi(\theta_{j-1})$$

$$p_j = \Phi(\theta_j - \Delta) - \Phi(\theta_{j-1} - \Delta).$$

Hence, as $N \rightarrow \infty$,

$$q_j - p_j \sim \Delta [u(\theta_j) - u(\theta_{j-1})] = \frac{d_j \delta}{\sqrt{N}},$$

that is,

$$\delta_j = d_j \delta.$$

Substitution of this expression into (3.10) yields the result of Corollary 3.2.1.

Proof of Corollary 3.2.2. From Theorem 3.2.1, the efficacy term is:

$$f(\vec{p}, \vec{\delta}) = \frac{\sqrt{ab} \left(\sum_{j=1}^{z-1} \delta_j + \sum_{j=1}^{z-1} \sum_{i=1}^{j-1} p_j \delta_i - \sum_{j=1}^{z-1} \sum_{i=j+1}^{z-1} p_j \delta_i \right)}{\sqrt{\sum_{j=1}^{z-1} r_j^2 p_j - \sum_{i=1}^{z-1} \sum_{j=1}^{z-1} r_i r_j p_i p_j}}.$$

For the case $p_j = 1/z \forall j$,

$$\begin{aligned} \sqrt{ab} \left(\sum_{j=1}^{z-1} \delta_j + \sum_{j=1}^{z-1} \sum_{i=1}^{j-1} p_j \delta_i - \sum_{j=1}^{z-1} \sum_{i=j+1}^{z-1} p_j \delta_i \right) &= \\ \sqrt{ab} \left[-\delta_z + \frac{1}{z} \sum_{j=1}^{z-2} \delta_j (z - j - 1) - \frac{1}{z} \sum_{j=2}^{z-1} (j - 1) \delta_j \right] &= \\ \sqrt{ab} \left[-\delta_z + \frac{1}{z} \sum_{j=2}^{z-2} (z \delta_j - 2j \delta_j) + \frac{z-2}{z} (\delta_1 - \delta_{z-1}) \right] &= \\ \frac{1}{z} \sqrt{ab} \left[z \sum_{j=1}^{z-1} \delta_j + \sum_{j=1}^{z-1} (z - 2j) \delta_j \right] \text{ since } \sum_{i=1}^z \delta_i = 0 &= \\ \frac{1}{z} \sqrt{ab} \left[2 \sum_{j=1}^{z-1} (z - j) \delta e_j \right]. \end{aligned}$$

Since it can easily be seen from the above argument that

$$\begin{aligned} \delta_j &\sim \sqrt{N} \Delta [h(\theta_j) - h(\theta_{j-1})] \\ &= \delta e_j \end{aligned}$$

for any continuous density h . Note that when $p_j = 1/z$,

$$\begin{aligned}\sum_{j=1}^{z-1} r_j &= z - 1 \\ \sum_{j=1}^{z-1} r_j^2 &= \frac{2(z-1)(2z-1)}{3z}.\end{aligned}$$

Substitution of these expressions into the denominator of the efficacy term produces:

$$\sqrt{\sum_{j=1}^{z-1} r_j^2 p_j - \sum_{i=1}^{z-1} \sum_{j=1}^{z-1} r_i r_j p_i p_j} = \sqrt{\frac{z^2 - 1}{3z^2}},$$

and it can easily be seen that the efficacy term is:

$$\frac{\sqrt{ab} \sum_{j=1}^{z-1} (z-j) e_j}{\sqrt{(z^2 - 1)/12}}.$$

3.2.2 Efficacy of the T test

Let the response variable be as in Section 3.2.1. Then the 2 sample t statistic can be written as:

$$T = \frac{\sum_{j=1}^{z-1} (g_j - g_z)(mT_j - nC_j)}{nmS\sqrt{1/m + 1/n}},$$

where

S = pooled standard deviation of the 2 groups (treatment and control).

As $N \rightarrow \infty$, (with $n/N \rightarrow a, m/N \rightarrow b$),

$$\begin{aligned}S^2 \rightarrow & a \left[\sum_{j=1}^z (g_j - g_z)^2 p_j - \sum_{j=1}^{z-1} \sum_{i=1}^{z-1} (g_j - g_z)(g_i - g_z) p_i p_j \right] + \\ & b \left[\sum_{j=1}^z (g_j - g_z)^2 p_j - \sum_{j=1}^{z-1} \sum_{i=1}^{z-1} (g_j - g_z)(g_i - g_z) q_i q_j \right]\end{aligned}$$

and, under H_0 ,

$$S^2 \rightarrow \sigma_o^2 = \sum_{j=1}^{z-1} (g_j - g_z)^2 p_j - \sum_{j=1}^{z-1} \sum_{i=1}^{z-1} (g_j - g_z)(g_i - g_z) p_i p_j.$$

The power of the 2 sample t test, $\Pi_N(\vec{p})$ is

$$\Pi_N(\vec{p}) = \Pr_{\vec{p}, \vec{\delta}}[T > t_\alpha].$$

Theorem 3.2.2 *Under the assumptions of Theorem 3.2.1*

$$\Pi_N(\vec{p}) \rightarrow \Phi[w(\vec{p}, \vec{\delta}) - z_\alpha]$$

where

$$w(\vec{p}, \vec{\delta}) = \frac{\sqrt{ab} \sum_{j=1}^{z-1} (g_j - g_z) \delta_j}{\sqrt{\sum_{j=1}^{z-1} (g_j - g_z)^2 p_j - \sum_{j=1}^{z-1} \sum_{i=1}^{z-1} (g_i - g_z)(g_j - g_z) p_i p_j}}.$$

Corollary 3.2.3 *If X_T and X_C are as in Section 3.2.1 then, as $N \rightarrow \infty$,*

$$\Pi_N(\Delta) = \Phi(c \cdot \delta - z_\alpha)$$

where d_j and δ are as in the previous section and

$$c = \frac{\sqrt{ab} \sum_{j=1}^{z-1} (g_j - g_z) d_j}{\sqrt{\sum_{j=1}^{z-1} (g_j - g_z)^2 p_j - \sum_{j=1}^{z-1} \sum_{i=1}^{z-1} (g_i - g_z)(g_j - g_z) p_i p_j}}$$

where $\Delta = \delta/\sqrt{N}$ and $n/N \rightarrow a$, $m/N \rightarrow b$.

Corollary 3.2.4 *When $p_1 = p_2 = \dots = p_z = \frac{1}{z}$, $(X_{T_1} - \Delta, X_{T_2} - \Delta, \dots, X_{T_z} - \Delta)$ and $(X_{C_1}, X_{C_2}, \dots, X_{C_z})$ are as in Corollary 3.2.2 and $g_i \propto i$, the efficacy term of the 2 sample t test becomes*

$$\frac{\sqrt{ab} \sum_{j=1}^{z-1} (z - j) e_j}{\sqrt{\frac{z^2 - 1}{12}}}.$$

where e_j is defined as previously.

Proof of Theorem 2.2.2:

$$\Pi_N(\vec{p}) = \Pr_{\vec{p}, \vec{\delta}}[T > t_\alpha]$$

Let $\gamma = \sum_{j=1}^{z-1} (g_j - g_z)(mT_j - nC_j)$. Then

$$\begin{aligned}\Pi_N(\vec{p}) &= \Pr_{\vec{p}, \vec{\delta}} \left[\frac{\gamma}{mnS\sqrt{1/m + 1/n}} > t_\alpha \right] \\ &= \Pr_{\vec{p}, \vec{\delta}} \left[\frac{\gamma - E(\gamma)}{\sqrt{\text{var}(\gamma)}} > \frac{mnt_\alpha S\sqrt{1/m + 1/n} - E(\gamma)}{\sqrt{\text{var}(\gamma)}} \right]\end{aligned}$$

In order to reduce this expression to the form in (3.2.2) the following will be shown:

1. $\frac{\gamma - E(\gamma)}{\sqrt{\text{var}(\gamma)}} \rightarrow N(0, 1)$
2. $\text{var}\left(\frac{\gamma}{\sqrt{mnN}}\right) \rightarrow \sigma_o^2$
3. $\frac{mnt_\alpha S\sqrt{1/m + 1/n}}{\sqrt{\text{var}(\gamma)}} \rightarrow z_\alpha$
4. $\frac{E(\gamma)}{\sqrt{\text{var}(\gamma)}} \rightarrow w(\vec{p}, \vec{\delta})$

Part 1 is true by the Central Limit Theorem. To prove part 2, note that:

$$\begin{aligned}\text{var}(\gamma) &= \text{var} \left(\sum_{j=1}^{z-1} (g_j - g_z)(mT_j - nC_j) \right) \\ &= mn \left[\sum_{j=1}^z (g_j - g_z)^2 [mp_j + nq_j] - \sum_{j=1}^{z-1} \sum_{i=1}^{z-1} (g_j - g_z)(g_i - g_z) [mp_i p_j + nq_i q_j] \right].\end{aligned}$$

As $N \rightarrow \infty$, $p_j \rightarrow q_j$ and hence,

$$\begin{aligned}\text{var}(\gamma) &\sim nmN \left(\sum_{j=1}^{z-1} (g_j - g_z)^2 p_j - \sum_{j=1}^{z-1} \sum_{i=1}^{z-1} (g_j - g_z)(g_i - g_z) p_i p_j \right) \\ &\Rightarrow \text{var}\left(\frac{\gamma}{\sqrt{mnN}}\right) \rightarrow \sigma_o^2.\end{aligned}$$

Now

$$\begin{aligned}\frac{mnt_\alpha S\sqrt{1/m + 1/n}}{\sqrt{\text{var}(\gamma)}} &= \frac{t_\alpha S\sqrt{mnN}}{\sqrt{\text{var}(\gamma)}} \\ &= \frac{t_\alpha S}{\sqrt{\text{var}\left(\frac{\gamma}{\sqrt{mnN}}\right)}} \\ &\rightarrow z_\alpha.\end{aligned}$$

The efficacy term can then be found to converge to a constant:

$$\begin{aligned} \frac{E(\gamma)}{\sqrt{\text{var}(\gamma)}} &\rightarrow \\ &\frac{\sqrt{ab} \sum_{j=1}^{z-1} (g_z - g_j) \delta_j}{\sqrt{\sum_{j=1}^{z-1} (g_j - g_z)^2 p_j - \sum_{j=1}^{z-1} \sum_{i=1}^{z-1} (g_i - g_z)(g_j - g_z) p_i p_j}} \\ &= w(\vec{p}, \vec{\delta}). \end{aligned}$$

3.3 Comparison of the Efficacies

The 2 sample t test is often used in analyzing ordered categorical data although it is not appropriate. For this reason, the relative efficiency of the Wilcoxon and the t test is of some interest. Under conditions listed below, as the number of observations becomes large, the efficacy terms for the Wilcoxon and the t tests applied to such data converge to the same expression:

$$\frac{\sqrt{ab} \sum_{j=1}^{z-1} (z - j) e_j}{\sqrt{(z^2 - 1)/12}} \quad (3.11)$$

Sufficient conditions for (3.11) to hold are:

1. the probabilities of the responses are equal i.e. $p_1 = p_2 = \dots = p_z$

and

2. the scores for the categories are proportional to the category number.
3. the scores are generated by a continuous distribution.

The efficacy for the t test applied to the continuous data is calculated in Lehmann [3]. It is

$$c = \frac{\sqrt{ab}}{\sigma}$$

where

σ is the standard deviation of the distribution of the data

$a = n/N$, the proportion of the sample in the treatment group

$b = m/N$, the proportion of the sample in the control group.

The efficiency of the Wilcoxon applied to the categorical data relative to the the t test applied to the underlying continuous data was of interest as this indicated the decrease in power due to the categorization of the continuous data. Under the condition of equal p_j 's and data generated from a continuous distribution, the Pitman efficiency of the Wilcoxon to the t test was:

$$e_{wilc,t(cut)} = \left[\frac{z^2 - 1}{12\sigma^2(\sum_{j=1}^{z-1}(z-j)e_j)^2} \right]$$

So, for example if the control group's scores were uniformly distributed on $[0, 1]$ and the treatment group's scores were uniformly distributed on $[\Delta, 1 + \Delta]$ then the efficiency of the t (uncut) test to the Wilcoxon (on scores categorized as in the simulation) is:

$$c = \frac{z^2 - 1}{(z - 1)^2}$$

(Although the uniform density is not continuous at 0 and 1, Corollary 3.2.2 can be modified for this case.) It is easily seen from this that as z , the number of categories becomes large, $c \rightarrow 1$, which is what would be expected.

The simulation in Section 3.1 was run under conditions 1 and 2 with the normal as its underlying continuous distribution. Although there were only 100 observations in each run, the curves in Figure 3.1 are very close. Using the efficacy terms calculated, the asymptotic Pitman efficiency of the Wilcoxon with respect to the t test on the continuous data was calculated and found to be 0.89. This means that the t test requires $\sim 89\%$ of the observations needed by the Wilcoxon in order to attain the same limiting power under the same alternate hypotheses. These calculations seem to agree with the simulation run in Section 3.1.

Chapter 4

Analysis at Two Time Points

Most patients in the Multiple Sclerosis trials were examined at all of the time points 0, 1, 3, 6, 9, 12, 18, 24 months during the study. A better indication of the differences between the treatment and control group could be gained by analyzing how the scores of the patients change between two time points. The power of some statistics frequently used to analyze ordered data over time were compared using simulation. These statistics were the stratified Wilcoxon, the McCullagh, and a chi squared statistic. Section 4.1 contains a description of the simulation runs. The Wilcoxon and the chi-squared statistic are described more fully in Sections 4.1.1 and 4.1.2. In the latter, the distribution of the chi-squared statistic is derived under the hypothesis of no difference between the treatment and control groups. In Section 4.2, the results of the simulation are discussed.

4.1 Description of the Simulation

An initial and final score were generated for each subject simulating his condition at two time points. The values of the underlying continuous random variables giving rise to these scores were denoted by (R_i^T, Y_i^T) or (R_i^C, Y_i^C) for those in the treatment or control group. These data were categorized using the same cut points as in the previous chapter to produce scores. The distribution of these pairs was:

$$\begin{pmatrix} R_i^T \\ Y_i^T \end{pmatrix} \sim BVN \left(\begin{bmatrix} 0 \\ \beta + \Delta \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

and

$$\begin{pmatrix} R_i^C \\ Y_i^C \end{pmatrix} \sim BVN \left(\begin{bmatrix} 0 \\ \beta \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

The parameters β and Δ represented the changes in the scores due to the progression of the disease and the treatment respectively. Power curves were computed for a level $\alpha=0.05$ test of the null hypothesis $\Delta = 0$ vs. the alternate hypothesis $\Delta \neq 0$. A simulation run generated two groups (fifty observations/group) at Δ values of 0 to 1.2 in increments of 0.1. Each curve was calculated using one thousand simulation runs.

The stratified Wilcoxon, chi-squared and McCullagh model statistics were calculated for each set of data generated. The McCullagh model statistic was computed using the plum software in a way similar to that in Chapter 2, except that the covariate \mathbf{x} had multiple components. One component was a variable indicating whether the observation was from the treatment or control group. The other components were indicator variables for each possible initial score. The scale parameter was set to one.

Since, in the simulation, the continuous observations were known, the continuous response variable was regressed using an indicator for the treatment/control groups and the initial values for covariates. The null hypothesis $H_0 : \Delta = 0$ was rejected if the estimate for the treatment/control variable was significantly different from zero. This statistic was included so that the decrease in power due to the loss of information from categorization could be determined.

4.1.1 Stratified Wilcoxon

The stratified Wilcoxon rank sum statistic, described in [3] (p. 132), is an extension of that employed in the previous chapter. The subjects in the two groups were first stratified according to their initial score. Within each strata, s , the Wilcoxon statistic

W_s was computed, as in Chapter 3, using the final scores as the basis for the ranking procedure and summing the treatment ranks. Each of the statistics, W_s , was normalized by subtraction of its mean to produce W_s^* :

$$W_s^* = W_s - E(W_s)$$

where

$$E(W_s) = \frac{T_s(N_s+1)}{2}$$

T_s = number of observations in treatment group with initial score s

N_s = total number of observations with initial score s .

The overall Wilcoxon, W^* , was then calculated as:

$$W^* = \sum_{j=1}^z \frac{W_j^*}{N_j + 1}.$$

The variance of W^* was:

$$[se(W^*)]^2 = \sum_{j=1}^z \left[\frac{se(W_j^*)}{N_j + 1} \right]^2$$

where

z = number of strata

$se(W_j^*)$ = standard deviation of W_j^* calculated in the usual way.

The test was then rejected or accepted using the normal approximation:

$$\frac{W^*}{se(W^*)} \approx N(0, 1).$$

4.1.2 Chi-squared Statistic

A natural test of the null hypothesis that a subject's progress is independent of treatment received is based on a simple contingency table analysis. The resulting statistic, referred to as χ_{calc}^2 , is based on the number of subjects whose condition improved, deteriorated or remained the same over the two time points. To calculate χ_{calc}^2 , the data were collapsed into a 3×2 contingency table:

Treatment	Control	Totals
T^-	C^-	n^-
T^o	C^o	n^o
T^+	C^+	n^+
n	m	N

where

$$T^- = \sum_{i=1}^{z-1} \sum_{j=i+1}^z T_{ij}$$

$$C^- = \sum_{i=1}^{z-1} \sum_{j=i+1}^z C_{ij}$$

$$T^o = \sum_{j=1}^z T_{jj}$$

$$C^o = \sum_{j=1}^z C_{jj}$$

$$T^+ = \sum_{i=2}^z \sum_{j=1}^{i-1} T_{ij}$$

$$C^+ = \sum_{i=2}^z \sum_{j=1}^{i-1} C_{ij}$$

T_{ij} = the number of observations in the treatment group
with initial score i and final score j

C_{ij} = the number of observations in the control group
with initial score i and final score j

n = number of observations in the treatment group

m = number of observations in the control group

N = total number of observations in the study.

The statistic was calculated as:

$$\chi_{calc}^2 = \frac{\left(T^- - \frac{nn^-}{N}\right)^2}{\frac{nn^-}{N}} + \frac{\left(C^- - \frac{mn^-}{N}\right)^2}{\frac{mn^-}{N}} + \frac{\left(T^o - \frac{nn^o}{N}\right)^2}{\frac{nn^o}{N}} + \frac{\left(C^o - \frac{mn^o}{N}\right)^2}{\frac{mn^o}{N}} +$$

$$\frac{\left(T^+ - \frac{nn^+}{N}\right)^2}{\frac{nn^+}{N}} + \frac{\left(C^+ - \frac{mn^+}{N}\right)^2}{\frac{mn^+}{N}}.$$

The statistic, χ_{calc}^2 , is usually assumed to have a limiting χ^2 distribution with two degrees of freedom. It was not clear if this approximation was valid so the limiting null distribution of χ_{calc}^2 was derived. Two methods of sampling which produce data in a fashion similar to that in the Multiple Sclerosis study were considered, and led to distinct limiting null distributions for χ_{calc}^2 . A description of each of these sampling schemes and the distribution of χ_{calc}^2 is preceded by some notation used in this analysis:

$p_{ij} = \Pr[\text{final score} = j \mid \text{initial score} = i, \text{in treatment group}]$

$q_{ij} = \Pr[\text{final score} = j \mid \text{initial score} = i, \text{in control group}]$

$p_{ij}^u = \Pr[\text{final score} = j, \text{initial score} = i, \text{in treatment group}]$

$q_{ij}^u = \Pr[\text{final score} = j, \text{initial score} = i, \text{in control group}]$

$n_i = \text{number of treatment observations in strata } i$

$m_i = \text{number of control observations in strata } i$

Unconditional Sampling

In unconditional sampling, subjects are randomly assigned to groups and then are stratified according to the chosen covariate. In this case, the numbers of observations in each strata is random. This type of sampling is appropriate when it is not important if strata are empty, or there are so many subjects that the experimenter is assured that it is unlikely that any of the strata will be empty.

Suppose that $(T_{11}, T_{12}, \dots, T_{zz}) \sim \text{multinomial}(n, p_{11}^u, p_{12}^u, \dots, p_{zz}^u)$, independent of $(C_{11}, C_{12}, \dots, C_{zz}) \sim \text{multinomial}(m, q_{11}^u, q_{12}^u, \dots, q_{zz}^u)$ under unconditional sampling. If we define \mathbf{T} to be a vector with the components (T^+, T^-, T^o) and \mathbf{C} to be (C^+, C^-, C^o)

then \mathbf{T} and \mathbf{C} are independent with distributions:

$$\mathbf{T} \sim \text{multinomial}(n, p^+, p^-, p^\circ)$$

$$\mathbf{C} \sim \text{multinomial}(m, q^+, q^-, q^\circ)$$

where

$$p^+ = \sum_{i=2}^z \sum_{j=1}^{i-1} p_{ij}^u$$

$$p^- = \sum_{i=1}^{z-1} \sum_{j=i+1}^z p_{ij}^u$$

$$p^\circ = \sum_{j=1}^z p_{jj}^u$$

$$q^+ = \sum_{i=2}^z \sum_{j=1}^{i-1} q_{ij}^u$$

$$q^- = \sum_{i=1}^{z-1} \sum_{j=i+1}^z q_{ij}^u$$

$$q^\circ = \sum_{j=1}^z q_{jj}^u.$$

In this case, the usual asymptotic theory for independent multinomials holds and under the hypothesis $p^+ = q^+$, $p^\circ = q^\circ$, $p^- = q^-$, (and thus also under the more restrictive null hypothesis $p_{ij}^u = q_{ij}^u \forall i, j$), χ_{calc}^2 is asymptotically distributed as a χ^2 with two degrees of freedom.

Conditional Sampling

In conditional sampling, the number of subjects in each strata is fixed at the beginning of the experiment. This type of sampling could be used when it is expensive to allow many subjects to take part. It ensures that there will be observations in each strata, although the total number of subjects may be relatively small. This type of sampling is common so it is natural to calculate the distribution of χ_{calc}^2 conditional on n_i and m_i . In this case, the distribution of the counts becomes $(T_{i1}, T_{i2}, \dots, T_{iz}) \sim \text{multinomial}(n_i, p_{i1}, p_{i2}, \dots, p_{iz})$ and $(C_{i1}, C_{i2}, \dots, C_{iz}) \sim \text{multinomial}(m_i, q_{i1}, q_{i2}, \dots, q_{iz})$ where $i = 1, 2, \dots, z$ and all multinomials are independent. Some assumptions made to

facilitate computations were that $n_i = m_i \forall i$ and that $N \rightarrow \infty$ in such a way that $n_i/N \rightarrow a_i$. Note that χ_{calc}^2 can be rewritten as:

$$\chi_{calc}^2 = K_1^2 + K_2^2 + K_3^2$$

where

$$\begin{aligned} K_1 &= \sqrt{\frac{N^3}{mnn^-}} \left[\frac{1}{\sqrt{N}} \left(T^- - \frac{nn^-}{N} \right) \right] \\ K_2 &= \sqrt{\frac{N^3}{mnn^o}} \left[\frac{1}{\sqrt{N}} \left(T^o - \frac{nn^o}{N} \right) \right]^2 \\ K_3 &= \sqrt{\frac{N^3}{mnn^+}} \left[\frac{1}{\sqrt{N}} \left(T^+ - \frac{nn^+}{N} \right) \right]^2. \end{aligned}$$

Theorem 4.1.1 *If $n_i = m_i \forall i$, then, under $H_o : p_{ij} = q_{ij}, \forall i, j$ as $N \rightarrow \infty$ and $n_i/N \rightarrow a_i$,*

$$(K_1, K_2, K_3) \xrightarrow{d} MVN(\mathbf{0}, \Sigma)$$

with

$$\begin{aligned} \Sigma_{11} &= 1 - \frac{\sum_{i=2}^z \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} a_i p_{ij} p_{ik}}{\sum_{i=2}^z \sum_{j=1}^{i-1} a_i p_{ij}} \\ \Sigma_{22} &= 1 - \frac{\sum_{j=1}^z a_j p_{jj}^2}{\sum_{j=1}^z a_j p_{jj}} \\ \Sigma_{33} &= 1 - \frac{\sum_{i=1}^{z-1} \sum_{j=i+1}^z \sum_{k=i+1}^z a_i p_{ij} p_{ik}}{\sum_{i=1}^{z-1} \sum_{j=i+1}^z a_i p_{ij}} \\ \Sigma_{12} &= \frac{-\sum_{i=2}^z \sum_{j=1}^{i-1} a_i p_{ii} p_{ij}}{\left[\left(\sum_{i=2}^z \sum_{j=1}^{i-1} a_i p_{ij} \right) \left(\sum_{j=1}^z a_j p_{jj} \right) \right]^{1/2}} \\ \Sigma_{13} &= \frac{-\sum_{i=2}^z \sum_{j=1}^{i-1} \sum_{k=i+1}^z a_i p_{ij} p_{ik}}{\left[\left(\sum_{i=2}^z \sum_{j=1}^{i-1} a_i p_{ij} \right) \left(\sum_{i=1}^{z-1} \sum_{j=i+1}^z a_i p_{ij} \right) \right]^{1/2}} \\ \text{and} \\ \Sigma_{23} &= \frac{-\sum_{i=1}^{z-1} \sum_{j=i+1}^z a_i p_{ii} p_{ij}}{\left[\left(\sum_{i=1}^{z-1} \sum_{j=i+1}^z a_i p_{ij} \right) \left(\sum_{j=1}^z a_j p_{jj} \right) \right]^{1/2}} \end{aligned}$$

Thus,

$$\chi_{calc}^2 \sim \left(\sqrt{\lambda_1}Z_1\right)^2 + \left(\sqrt{\lambda_2}Z_2\right)^2 + \left(\sqrt{\lambda_3}Z_3\right)^2$$

where

Z_1, Z_2, Z_3 are iid $N(0, 1)$ and

$\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues of Σ .

Proof of Theorem 4.1.1. Let

$$\begin{aligned} X_1 &= \frac{1}{\sqrt{N}} \left(T^- - \frac{nn^-}{N} \right) \\ X_2 &= \frac{1}{\sqrt{N}} \left(T^\circ - \frac{nn^\circ}{N} \right) \\ X_3 &= \frac{1}{\sqrt{N}} \left(T^+ - \frac{nn^+}{N} \right). \end{aligned}$$

To determine the limiting distribution of χ_{calc}^2 , first find the limiting distributions of X_1, X_2, X_3 . If

$$T_{ij}^* = \frac{T_{ij} - n_i p_{ij}}{\sqrt{n_i}}$$

and

$$C_{ij}^* = \frac{C_{ij} - m_i q_{ij}}{\sqrt{m_i}}$$

then

$$\begin{aligned} X_1 &= \frac{1}{\sqrt{N}} \left(T^- - \frac{nn^-}{N} \right) \\ &= \frac{1}{\sqrt{N}} \left(\sum_{i=2}^z \sum_{j=1}^{i-1} T_{ij} - \frac{nn^-}{N} \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{N}} \left[\left(\frac{m}{N} \right) \sum_{i=2}^z \sum_{j=1}^{i-1} T_{ij} - \frac{n}{N} \sum_{i=2}^z \sum_{j=1}^{i-1} C_{ij} \right] \\
&= \frac{1}{\sqrt{N}} \left[\left(\frac{m}{N} \right) \sum_{i=2}^z \sum_{j=1}^{i-1} T_{ij}^* \sqrt{n_i} - \frac{n}{N} \sum_{i=2}^z \sum_{j=1}^{i-1} C_{ij}^* \sqrt{m_i} + \right. \\
&\quad \left. \sum_{i=2}^z \sum_{j=1}^{i-1} \left(\left(\frac{m}{N} \right) n_i p_{ij} - \frac{n}{N} m_i q_{ij} \right) \right].
\end{aligned}$$

Therefore, under H_0 ,

$$X_1 \sim \frac{1}{2} \sum_{i=2}^z \sum_{j=1}^{i-1} (T_{ij}^* - C_{ij}^*) \sqrt{a_i}$$

since $n_i = m_i$ and $p_{ij} = q_{ij}$ under H_0 .

Under H_0 , T_{ij}^* and C_{ij}^* converge to normal random variables with mean zero, and thus X_1 is asymptotically normal with mean zero and variance:

$$\begin{aligned}
\text{var}(X_1) &\sim \text{var} \left[\frac{1}{2} \left(\sum_{i=2}^z \sum_{j=1}^{i-1} (T_{ij}^* - C_{ij}^*) \sqrt{a_i} \right) \right] \\
&= \frac{1}{4} \left[2 \left(\sum_{i=2}^z \sum_{j=1}^{i-1} a_i p_{ij} (1 - p_{ij}) - 2 \sum_{i=2}^z \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} a_i p_{ij} p_{ik} \right) \right] \\
&= \frac{1}{2} \left[\sum_{i=2}^z \sum_{j=1}^{i-1} a_i p_{ij} - \sum_{i=2}^z \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} a_i p_{ij} p_{ik} \right].
\end{aligned}$$

Under H_0 ,

$$\begin{aligned}
\frac{n^-}{N} &= \frac{\sum_{i=2}^z \sum_{j=1}^{i-1} (T_{ij} + C_{ij})}{N} \\
&= \sum_{i=2}^z \sum_{j=1}^{i-1} \left(\frac{T_{ij}}{n_i} \times \frac{n_i}{N} + \frac{C_{ij}}{m_i} \times \frac{m_i}{N} \right) \\
&\rightarrow \sum_{i=2}^z \sum_{j=1}^{i-1} (p_{ij} a_i + p_{ij} a_i) \\
&= 2 \sum_{i=2}^z \sum_{j=1}^{i-1} a_i p_{ij},
\end{aligned}$$

and thus K_1 is asymptotically normal with mean zero and variance,

$$\frac{\sum_{i=2}^z \sum_{j=1}^{i-1} a_i p_{ij} - \sum_{i=2}^z \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} a_i p_{ij} p_{ik}}{\sum_{i=2}^z \sum_{j=1}^{i-1} a_i p_{ij}}.$$

Similar calculations show that K_2 and K_3 are asymptotically normal. The variance/covariance calculations are straightforward. Now

$$(K_1, K_2, K_3) \sim MVN(0, \Sigma).$$

If Σ is a positive semidefinite matrix, \exists a diagonal matrix D , with the eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ of Σ as its elements, and a matrix P , such that $P^T P = I$, and $P \Sigma P^T = D$.

Let $Y = PK$, then $Y \sim N(0, D)$ and

$$\begin{aligned} K^T K &= Y^T Y \\ &\sim \lambda_1 Z_1^2 + \lambda_2 Z_2^2 + \lambda_3 Z_3^2 \end{aligned}$$

In most cases, the eigenvalues of Σ are difficult to find explicitly. However, one simple case is considered in the following:

Corollary 4.1.1 *Suppose that $n_1 = n_2 = \dots = n_z$ and $p_{ij} = 1/z \forall i, j$. Under the null hypothesis,*

$$\chi_{calc}^2 \xrightarrow{d} Z_1^2 + \left(\sqrt{\frac{2}{3}} - \frac{1}{3z} Z_2 \right)^2$$

where

$$Z_1, Z_2 \sim N(0, 1).$$

Thus, χ_{calc}^2 does not have a limiting chi-squared distribution with two degrees of freedom in this case.

Proof of Corollary:

Using Theorem 4.1.1, the asymptotic covariance matrix of (K_1, K_2, K_3) is:

$$\begin{bmatrix} \frac{1}{3} \left(1 + \frac{1}{z} \right) & -\frac{\sqrt{2(z-1)}}{2z} & -\frac{1}{3} \left(1 - \frac{2}{z} \right) \\ -\frac{\sqrt{2(z-1)}}{2z} & 1 - \frac{1}{z} & -\frac{\sqrt{2(z-1)}}{2z} \\ -\frac{1}{3} \left(1 - \frac{2}{z} \right) & -\frac{\sqrt{2(z-1)}}{2z} & \frac{1}{3} \left(1 + \frac{1}{z} \right) \end{bmatrix}.$$

The eigenvalues for this covariance matrix are $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = \frac{2}{3} - \frac{1}{3z}$.

4.2 Simulation Results

The power curves computed by the simulations appear in Figures 4.1 to 4.4. In all of the figures, the most powerful test was the one based on the regression parameter. This result was expected as no information due to categorization was lost in the calculation of this statistic. The next most powerful was the McCullagh statistic which had a similar power curve to the Wilcoxon. In all cases both of these statistics were much more powerful than the chi-squared statistic. Because the chi-squared statistic did not take into account the initial score or the size of the difference, this was not surprising.

The powers of the tests increased with increasing ρ . This probably occurred because increasing ρ would decrease the variation in differences between the initial and final scores at fixed values of β and Δ . Changes in β did not affect the power curves of the Wilcoxon, McCullagh or regression statistics, but increasing it did decrease the power of the chi-squared statistic. As β became large, most patients' scores increased, and the difference between the treatment and control groups were in the sizes of these increases. Since the Wilcoxon and the McCullagh statistics compared the treatment and control groups based on the differences of the final scores given the initial scores, an increase in β would not be expected to change the power of these statistics. However, very large changes in β would result in all patients moving to the largest score, in which case, none of the tests would detect any difference between the groups. The chi-square statistic computed here was unconditional as the initial number of observations in the strata were not fixed. It only recorded whether or not the patients' scores increased, decreased or remained the same, so its power was expected to decrease as β increased.

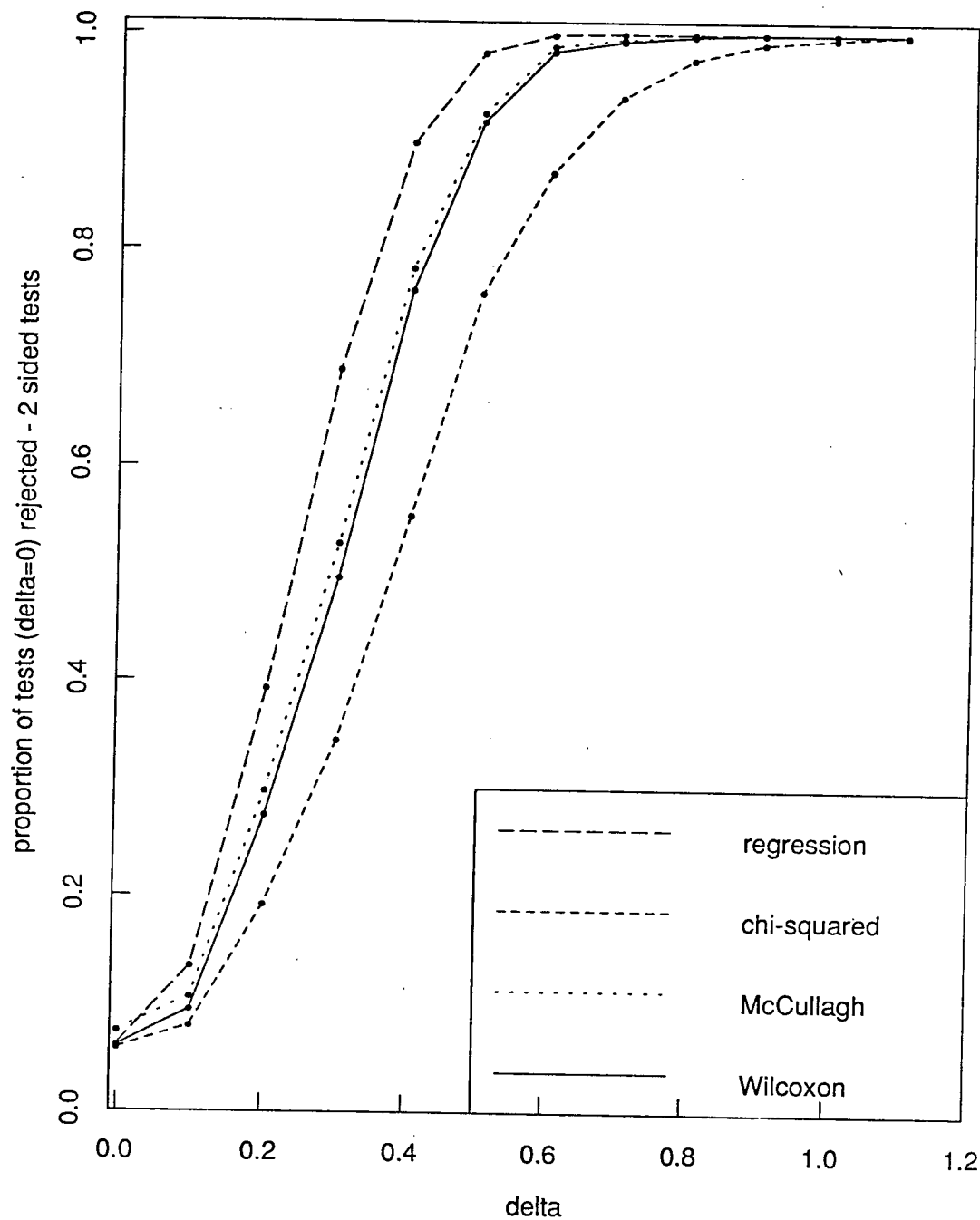
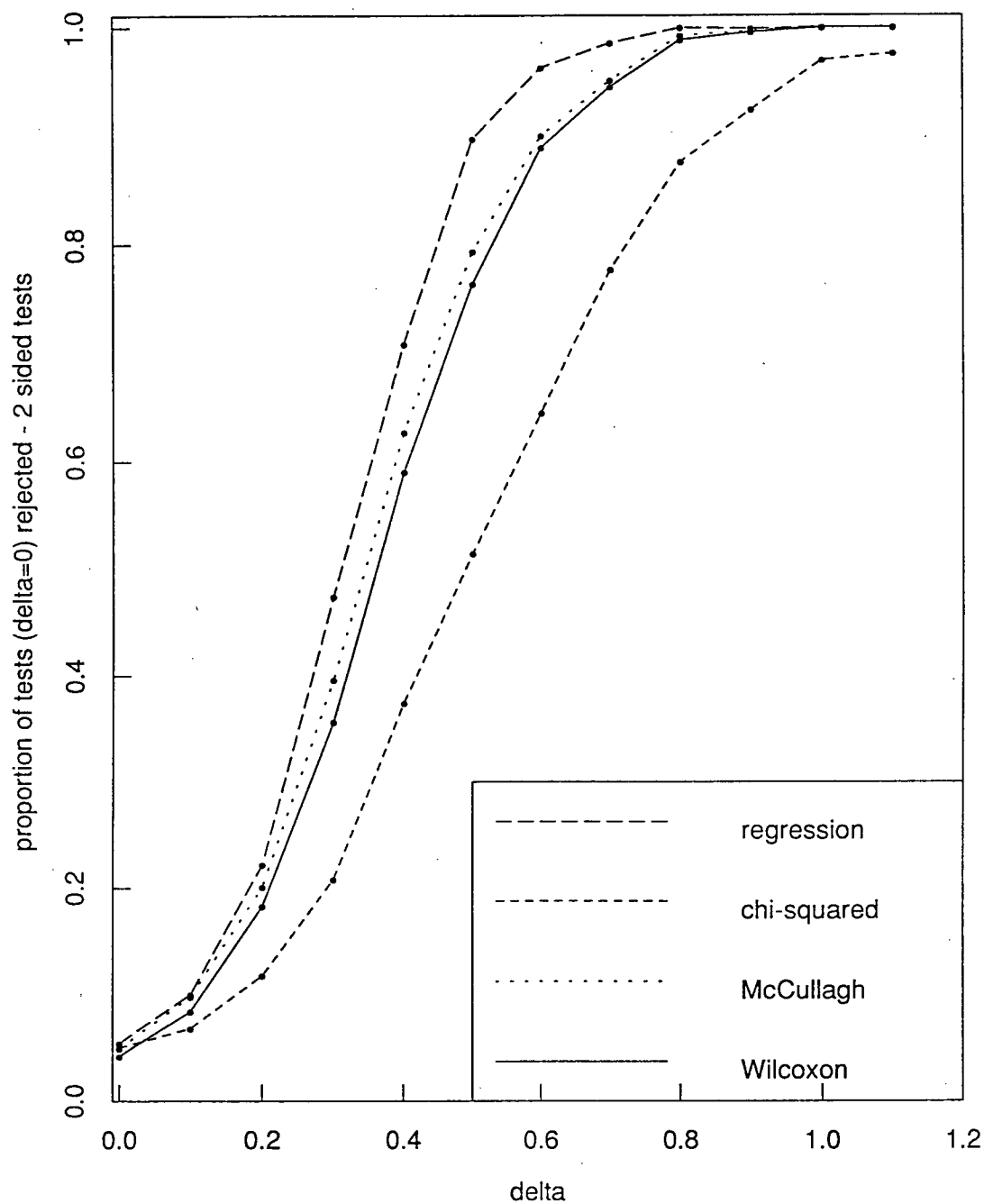
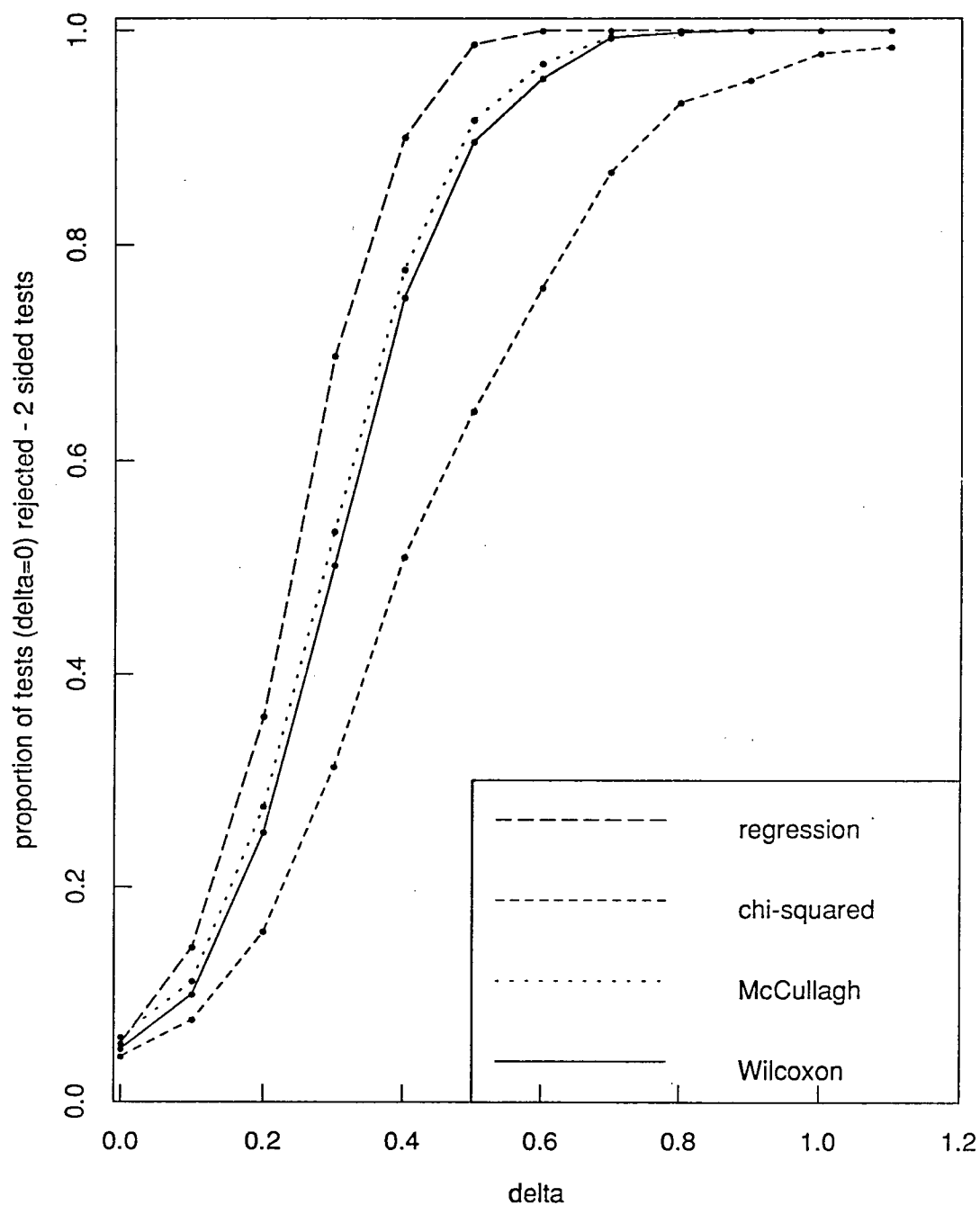


Figure 4.1: Simulation Run 1 ($\beta = 0.1, \rho = 0.8$)

Figure 4.2: Simulation Run 2 ($\beta = 0.1, \rho = 0.6$)

Figure 4.3: Simulation Run 3 ($\beta = 0.3, \rho = 0.8$)

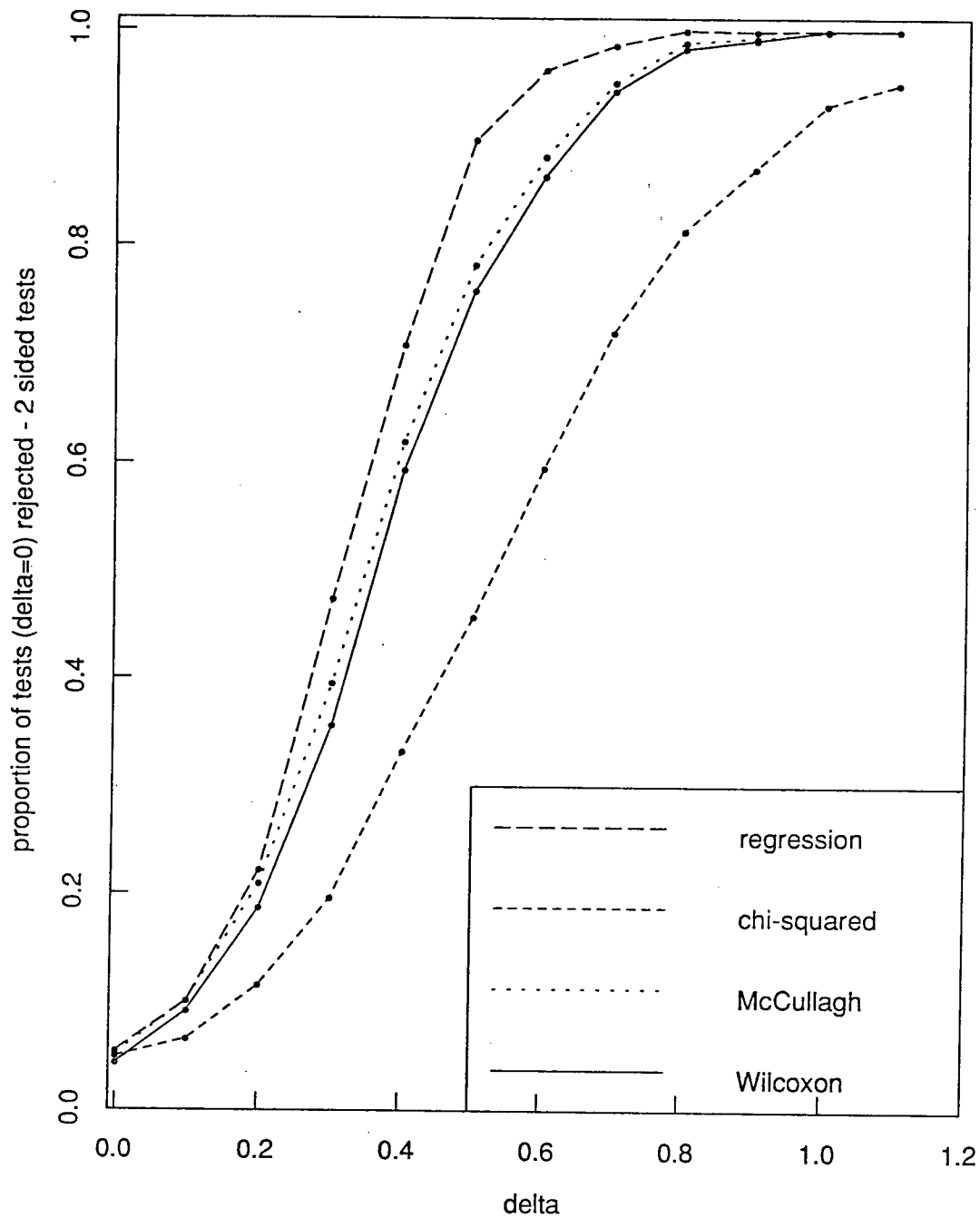
Figure 4.4: Simulation Run 4 ($\beta = 0.3, \rho = 0.6$)

Table 4.1: Results of McCullagh Analysis

Time (in months)	estimate of treatment parameter	standard error	z-score
0-1	-0.82	0.418	-1.95
1-3	-0.40	0.403	-1.00
3-6	-0.17	0.424	-0.40
6-9	0.98	0.466	2.10
9-12	0.07	0.51	0.14
12-18	0.31	0.452	0.69
18-24	-0.16	0.48	-0.33

4.3 Application of Tests to the MS Data

The stratified Wilcoxon and χ^2 tests have been calculated on the MS data set previously [2]. Both statistics were calculated on the data between 0 months and each of the other times the patients were observed. The results of the test based on the Wilcoxon statistic showed that the treatment group regressed ($p \sim 0.05$) relative to the control group in the time periods 0-1 months and 0-3 months. The remaining statistics calculated were nonsignificant. However, the change in sign in those time periods larger than six months indicates that patients in the control group may have fared less well than those in the treatment group in the follow up period. Since the χ^2 is less powerful than the Wilcoxon, it is not surprising that none of the tests based on it were significant.

The data were fit to a McCullagh model. Unlike the previous analysis, the collapsed scores were used here since the data were sparse. The covariates in this analysis were indicator variables for each of the possible initial scores and one for treatment/ control group. The data were modelled between consecutive time points rather than from baseline to the other scores. The results are shown in Table 4.1.

The variable for the treatment group parameter was set up so that a negative value

implied that the control group regressed with respect to the treatment group. Collapsing the data has produced results which are contradictory to the previous Wilcoxon analysis in the 0-1 month time period.

The Wilcoxon and χ^2 analysis agree with the analysis using the Markov techniques in that all three show that patients in the treatment group may not have progressed at the same rate as those in the control group during the first six months of the study. In the case of the Markov analysis, it is not possible to determine which group improved relative to the other.

Chapter 5

Conclusions

As the data set was small, the number of observations which had any particular Kurtzke score was low. For this reason, the data were collapsed into five categories chosen only to ensure at least four observations in each category at zero months. Results from the Markov analysis indicate that, during the administration of interferon, the treatment group regressed relative to the controls. After this period, subjects in the treatment group appeared to return to a state in which their transition probabilities were not significantly different from the controls. Models which did not incorporate information about the length of time intervals between observations fit the data just as well as those that did. That is, a model which assumed that transitions between any two consecutive time points could be modelled using the same transition matrix fit as well as a model which allowed a different matrix for each time interval. Modelling with a specific tridiagonal matrix with all diagonal elements equal, except for the second, proved to be reasonable for the control group over the entire time period. The treatment group could be modelled reasonably by this form only in the eighteen month follow up period. The major differences between the two groups were in those patients that started at Kurtzke scores in the range 4.5 to 5.5. Control patients with initial scores in this range fared better than treatment patients, during the period when the interferon was administered.

Statistics commonly used to compare the groups at one time point were examined

using simulation techniques. If data for the control group were generated using underlying $N(0, 1)$ distribution and those for the treatment group by a $N(\Delta, 1)$ distribution, and the probability of being in each category was equal for $\Delta = 0$, then the Wilcoxon, McCullagh and 2 sample t test were found to have similar relative efficiencies. Theoretical asymptotic calculations yielded expressions for Pitman efficiencies of these statistics for general shift models. Under the conditions of the simulations, the efficacies of the Wilcoxon and 2 sample t test calculated on the categorized data were found to be equal. The Pitman efficiency of these two statistics with respect to the t test calculated on the underlying continuous data was 0.89.

A comparison of statistics used to compare the two groups' progression of disease between two time points was then carried out. Simulation results showed that when the categorical data were generated using a bivariate normal distribution, with equal initial probabilities of being in any category, the Wilcoxon and McCullagh statistics were much more efficient than the chi squared statistic. The asymptotic distribution of the chi-squared statistic was derived under the hypothesis of no treatment effect. It was determined that the chi-squared statistic did not necessarily have a limiting χ^2 distribution with two degrees of freedom if, in the sample of patients, the initial number of subjects in each category was fixed. However, if the number was not fixed, the chi squared statistic did have an asymptotic χ^2 distribution with two degrees of freedom.

Appendix A

Sample Transition Matrices

Transition Matrix for Control Group from 0 to 24 months

$$\begin{bmatrix} 21 & 7 & 1 & 0 & 0 \\ 8 & 7 & 3 & 0 & 0 \\ 0 & 2 & 67 & 29 & 1 \\ 0 & 0 & 23 & 49 & 7 \\ 0 & 0 & 1 & 5 & 8 \end{bmatrix}$$

Transition Matrix for Treatment Group from 0 to 24 months

$$\begin{bmatrix} 23 & 6 & 1 & 0 & 0 \\ 4 & 2 & 7 & 0 & 0 \\ 1 & 3 & 59 & 20 & 1 \\ 0 & 0 & 20 & 40 & 14 \\ 0 & 0 & 0 & 12 & 16 \end{bmatrix}$$

Control Group - General Tridiagonal Modelling

0 months to 6 months

$$\begin{bmatrix} 0.67 & 0.33 & 0.00 & 0.00 & 0.00 \\ 0.57 & 0.29 & 0.14 & 0.00 & 0.00 \\ 0.00 & 0.03 & 0.65 & 0.32 & 0.00 \\ 0.00 & 0.00 & 0.35 & 0.59 & 0.06 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \end{bmatrix}$$

6 months to 24 months

$$\begin{bmatrix} 0.81 & 0.19 & 0.00 & 0.00 & 0.00 \\ 0.36 & 0.45 & 0.18 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.71 & 0.28 & 0.00 \\ 0.00 & 0.00 & 0.24 & 0.64 & 0.11 \\ 0.00 & 0.00 & 0.00 & 0.20 & 0.80 \end{bmatrix}$$

0 months to 24 months

$$\begin{bmatrix} 0.75 & 0.25 & 0.00 & 0.00 & 0.00 \\ 0.44 & 0.39 & 0.17 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.68 & 0.30 & 0.00 \\ 0.00 & 0.00 & 0.29 & 0.62 & 0.09 \\ 0.00 & 0.00 & 0.00 & 0.38 & 0.62 \end{bmatrix}$$

Treatment group - General Modelling

0 months to 6 months

$$\begin{bmatrix} 0.75 & 0.25 & 0.00 & 0.00 & 0.00 \\ 0.13 & 0.13 & 0.75 & 0.00 & 0.00 \\ 0.00 & 0.07 & 0.57 & 0.37 & 0.00 \\ 0.00 & 0.00 & 0.35 & 0.32 & 0.32 \\ 0.00 & 0.00 & 0.00 & 0.58 & 0.42 \end{bmatrix}$$

6 months to 24 months

$$\begin{bmatrix} 0.82 & 0.18 & 0.00 & 0.00 & 0.00 \\ 0.60 & 0.20 & 0.20 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.81 & 0.17 & 0.00 \\ 0.00 & 0.00 & 0.21 & 0.70 & 0.09 \\ 0.00 & 0.00 & 0.00 & 0.31 & 0.69 \end{bmatrix}$$

0 months to 24 months

$$\begin{bmatrix} 0.79 & 0.21 & 0.00 & 0.00 & 0.00 \\ 0.31 & 0.15 & 0.54 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.72 & 0.24 & 0.00 \\ 0.00 & 0.00 & 0.27 & 0.54 & 0.19 \\ 0.00 & 0.00 & 0.00 & 0.43 & 0.57 \end{bmatrix}$$

Control Group - Specific Tridiagonal Modelling

0 months to 6 months

$$\begin{bmatrix} 0.61 & 0.39 & 0.00 & 0.00 & 0.00 \\ 0.57 & 0.29 & 0.14 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.61 & 0.35 & 0.00 \\ 0.00 & 0.00 & 0.35 & 0.61 & 0.04 \\ 0.00 & 0.00 & 0.00 & 0.39 & 0.61 \end{bmatrix}$$

6 months to 24 months

$$\begin{bmatrix} 0.71 & 0.29 & 0.00 & 0.00 & 0.00 \\ 0.36 & 0.45 & 0.18 & 0.00 & 0.00 \\ 0.00 & 0.05 & 0.71 & 0.24 & 0.00 \\ 0.00 & 0.00 & 0.24 & 0.71 & 0.05 \\ 0.00 & 0.00 & 0.00 & 0.29 & 0.71 \end{bmatrix}$$

0 months to 24 months

$$\begin{bmatrix} 0.67 & 0.33 & 0.00 & 0.00 & 0.00 \\ 0.44 & 0.39 & 0.17 & 0.00 & 0.00 \\ 0.00 & 0.05 & 0.67 & 0.29 & 0.00 \\ 0.00 & 0.00 & 0.29 & 0.67 & 0.05 \\ 0.00 & 0.00 & 0.00 & 0.33 & 0.67 \end{bmatrix}$$

Treatment Group - Tridiagonal Modelling

0 months to 6 months

$$\begin{bmatrix} 0.48 & 0.52 & 0.00 & 0.00 & 0.00 \\ 0.13 & 0.13 & 0.75 & 0.00 & 0.00 \\ 0.00 & 0.18 & 0.48 & 0.33 & 0.00 \\ 0.00 & 0.00 & 0.33 & 0.48 & 0.18 \\ 0.00 & 0.00 & 0.00 & 0.52 & 0.48 \end{bmatrix}$$

6 months to 24 months

$$\begin{bmatrix} 0.76 & 0.24 & 0.00 & 0.00 & 0.00 \\ 0.60 & 0.20 & 0.20 & 0.00 & 0.00 \\ 0.00 & 0.05 & 0.76 & 0.19 & 0.00 \\ 0.00 & 0.00 & 0.19 & 0.76 & 0.05 \\ 0.00 & 0.00 & 0.00 & 0.24 & 0.76 \end{bmatrix}$$

0 months to 24 months

$$\begin{bmatrix} 0.65 & 0.35 & 0.00 & 0.00 & 0.00 \\ 0.31 & 0.15 & 0.54 & 0.00 & 0.00 \\ 0.00 & 0.11 & 0.65 & 0.25 & 0.00 \\ 0.00 & 0.00 & 0.25 & 0.65 & 0.11 \\ 0.00 & 0.00 & 0.00 & 0.35 & 0.65 \end{bmatrix}$$

Bibliography

- [1] Bhat, Narayan U., (1984). *Elements of Applied Stochastic Processes* 2nd edition. John Wiley & Sons.
- [2] Kastrukoff, L., et. al., (1988). unpublished manuscript. UBC Hospital.
- [3] Lehmann, E.L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day Inc., San Francisco, California.
- [4] McCullagh, P. (1980). Regression Models For Ordinal Data. *J.R. Statist. Soc. B*, **42**, No. 2, pp. 109-142.
- [5] Walsh, G.R. (1975). *Methods of Optimization*. John Wiley & Sons.
- [6] Whittle, P. (1955). Some Distribution and Moment Formulae For the Markov Chain. *J.R. Statist. Soc. B*, **17**, pp. 235-242.
- [7] The Statistical Advisory Service (1988). Multiple Sclerosis Dataset. unpublished manuscript, University of Manitoba.