ISSN 1361-9802

Manchester Centre
for
Statistical Science

Technical Report 2002/01

Diagnosis of Carpal Tunnel Syndrome using Logistic Regression (2)

Lectures given at the Facultad de Matemáticas Universidad Autónoma de Yucatán Mérida, Yucatán, Mexico

18–22 February, 2002

Stephan M. Rudolfer

April 2002

University of Manchester

Manchester Centre for Statistical Science

Technical Reports

Available by post from the author
Department of Mathematics, University of Manchester, Manchester M13 9PL, U.K.
or from his webpage
http://www.maths.man.ac.uk/DeptWeb/Homepages/smr/Reprints/mexico2.pdf

Contents

1 ACKNOWLEDGEMENTS **0**

2 BACKGROUND **1**

2.1 Introduction 1
2.2 Notation 1
2.3 Carpal Tunnel Syndrome (CTS) Dataset: Rudolfer(2001), Section 2 2
2.3.1 (ORDINAL) Response Variable 3
2.3.2 Predictor Variables 3
2.3.3 Definition of Carpal Tunnel Syndrome 10
2.3.4 Cause of Carpal Tunnel Syndrome 11
2.3.5 Source of the dataset 11
2.4 Models considered so far: Rudolfer(2001), Sections 3.2–3.4 12
2.4.1 Proportional Odds (PO) 12
2.4.2 Continuation Ratio (CR) 12
2.5 Classification procedures: Rudolfer (2001), Section 3.6 13
2.5.1 Highest Probability (HP) 13
2.5.2 Anderson and Philips (AP) for the PO model 14
2.6 General methodology of model fitting: Rudolfer(2001), Section 4 14
2.6.1 Preliminary Steps 14
2.6.2 Data Reduction 14
2.6.3 Verify the model's assumptions 14
2.6.4 Fit the model 15
2.6.5 Compute measures of predictive accuracy 15
2.6.6 Validate the model 15
2.7 Computational Aspects 15
2.7.1 Statistical software used this time: SAS Version 8.2 15
2.7.2 Previous problem: Rudolfer(2001), Section 2.5 15
2.7.3 Solution: Bad News/Good News 15
2.7.4 Ascending and Descending options in proc LOGISTIC are equivalent 17

3 FURTHER MODEL FITTING **17**

3.1 Use of Binary Logistic Regression software for fitting Continuation Ratio (CR) models 17
3.1.1 Underlying Reason why the Method works 18
3.1.2 Details of the Method 18
3.2 Partial Proportional Odds (PPO) models 19
3.3 Extended Continuation Ratio (ECR) models 20
3.4 Equivalence between PO and CR models 20

4 FURTHER VARIABLE SELECTION **20**
4.1 Univariate Wald's test . . . 20
4.2 Score test . . . 21
4.3 Automatic Variable Selection (SAS, 2000) . . . 21
4.3.1 Backward elimination . . . 21
4.3.2 Fast backward elimination . . . 22
4.3.3 Forward selection . . . 22
4.4 The Bonferroni method . . . 22
4.5 Occam's (or Ockham's) razor . . . 23

5 PERFORMANCE EVALUATION/MEASURES OF PREDICTIVE ACCURACY: Rudolfer(2001), Section 4.5 **24**
5.1 Preliminary definitions . . . 25
5.2 Concordant/Discordant pairs of observations . . . 27
5.2.1 Properties of π_c, π_d, π_t . . . 28
5.3 Goodman & Kruskal's Gamma, γ . . . 30
5.3.1 Interpretations of γ . . . 30
5.3.2 Properties of γ . . . 31
5.3.3 G = Estimator of γ . . . 31
5.3.4 Properties of G . . . 31
5.3.5 Estimated Asymptotic Variance of G . . . 32
5.3.6 Estimated Asymptotic Variance of G under the null hypothesis that $\gamma = 0$ . . . 32
5.4 Somers' D rank correlation indices . . . 32
5.4.1 Interpretations of $\Delta(C|R)$ . . . 33
5.4.2 Properties of $\Delta(C|R)$ . . . 33
5.4.3 $D(C|R)$ = Estimator of $\Delta(C|R)$ . . . 34
5.4.4 Properties of $D(C|R)$ . . . 34
5.4.5 Estimated Asymptotic Variance of $D(C|R)$ . . . 34
5.4.6 Estimated Asymptotic Variance of $D(C|R)$ under the null hypothesis that $\Delta(C|R) = 0$ . . . 34
5.5 Concordance index, c: scaled Somers' D . . . 35
5.6 Percentage correct . . . 35
5.7 Cohen's Kappa statistic,$\hat{\kappa}$ . . . 35
5.7.1 Properties of $\hat{\kappa}$ . . . 36
5.7.2 Estimated Asymptotic Variance of $\hat{\kappa}$ . . . 36
5.8 Brier's score, B . . . 36
5.8.1 Properties of B . . . 37

6 RESULTS **37**
6.1 Variable selection . . . 37
6.1.1 History variables . . . 37

6.1.2 Clinical sign variables . . . 38
6.1.3 Nerve conduction variables . . . 38
6.1.4 Combined model: M_1 . . . 38
6.1.5 Six-variable submodel of M_1: M_2 . . . 39
6.1.6 Seven-variable submodel of M_1: M_3 . . . 39
6.1.7 Full model without variable selection: M_4 . . . 39
6.2 Performance evaluation . . . 39
6.2.1 Anderson & Philips' (AP) classification procedure performs badly . . . 40
6.2.2 Which of the two Somers' Ds to use? . . . 40
6.2.3 Performance of M_1 . . . 41
6.2.4 Performance of M_2 . . . 41
6.2.5 Performance of M_3 . . . 42
6.2.6 Performance of M_4 . . . 42
6.3 Conclusions . . . 43

7 OPEN PROBLEMS / FURTHER WORK 43
7.1 Inadequacy of the AP classification rule . . . 43
7.2 Further variable selection . . . 43
7.2.1 Branch & Bound . . . 43
7.2.2 Use of bootstrap to refine automatic variable selection . . . 43
7.3 Valedictory . . . 44

8 REFERENCES 44

1 ACKNOWLEDGEMENTS

Firstly, I would like to thank Dr. Luis A. Rodriguez Carvajal, Director of the Faculty of Mathematics, for inviting me again to give the talks, and the Autonomous University of Yucatán, Mérida, for its financial support.

Secondly, I extend my most grateful thanks to Humberto Cervera Rossado for the excellent and difficult job he did of simultaneous translation of my talks for the second time.

My thanks also go to Dr. Peter Eccles, Senior Lecturer in Mathematics, and Dr. Chris Paul, Systems Administrator, for their invaluable assistance with the preparation of my talks as a technical report and a web page.

I owe a deep debt of gratitude to the late Dr. John L. James, Consultant Physician, St Luke's Hospital, Huddersfield, Yorkshire, England. He introduced me to Carpal Tunnel Syndrome, and gave consistently helpful and encouraging input to my research in this area with him. The dataset used for these talks would never have seen the light of day without his immense cooperation. I am also profoundly grateful to Dr. James' secretary, Mrs Sylvia Hague, and to his technicians, Miss Gill Lockwood, Mrs Ginny Lockwood and Mrs Louise Mullinger, as well as to Martyn Thompson, a more recently appointed technician, for their great patience and cooperation in the acquisition and evaluation of the data.

Lastly, but not leastly, my deepest appreciation goes to my wife, Christine, for her support and encouragement during the preparation and delivery of these lectures.

2 BACKGROUND

2.1 Introduction

This chapter covers the necessary background from Rudolfer (2001) needed to understand the applications to Dr. James' dataset of the further methods described in this report. Sections 2.4 – 2.6 are summaries of the corresponding sections of Rudolfer (2001), which should be referred to for more details. Section 2.3, on the other hand, describes the dataset in detail, being based on Rudolfer (2001), Section 2, since it is not widely known outside the circle of clinical neurophysiologists and is essential to understand the results of the further methods.

2.2 Notation

Muddled notation produces muddled thought. Precise notation produces precise thought.

Hence, we shall adopt the following convention throughout this report:

sample **observed values:** ***small*** Roman letters

population **random variables:** **LARGE** Roman letters

population **parameters:** Greek letters

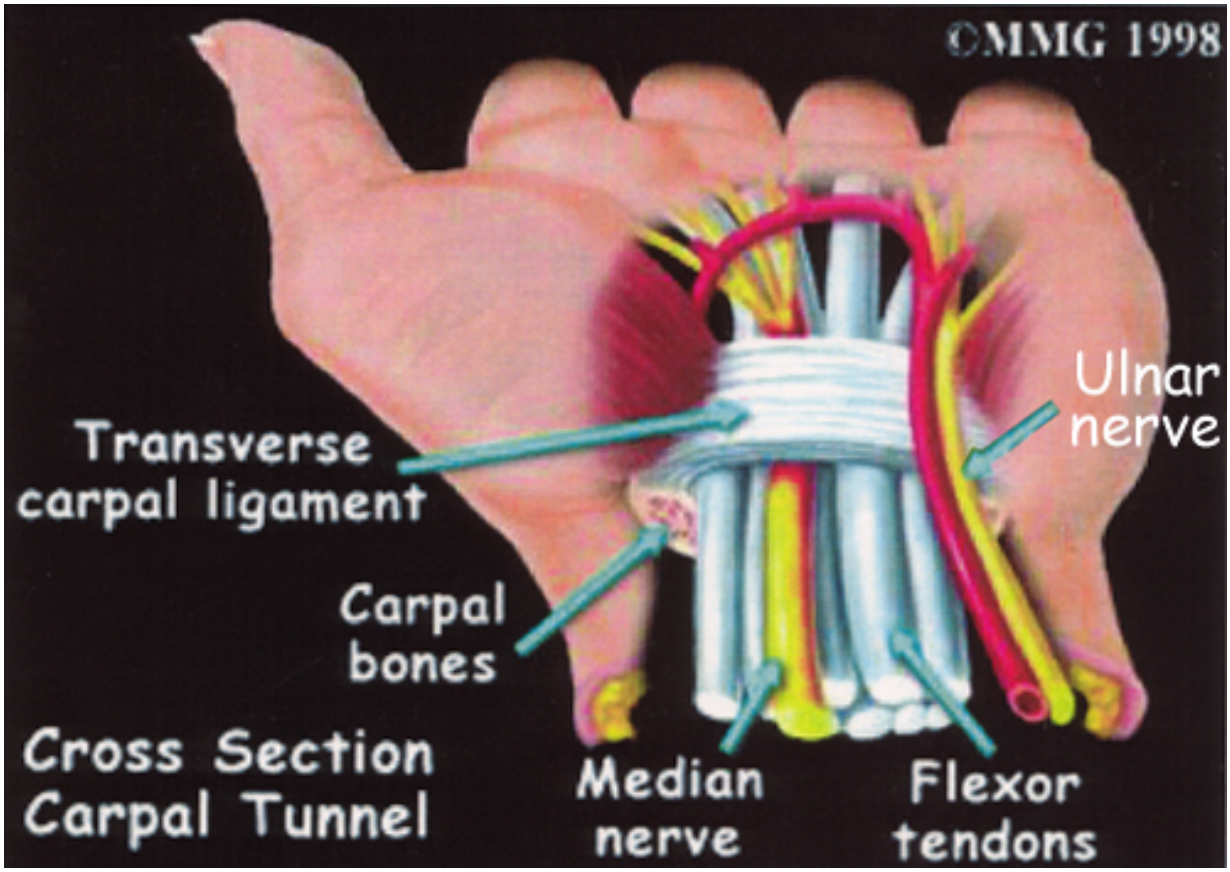
estimat<u>E</u>: observed sample statistic that estimates a population parameter

estimat<u>OR</u>: random variable of which the estimate is an observed value

2.3 Carpal Tunnel Syndrome (CTS) Dataset: Rudolfer(2001), Section 2

CTS = cluster of certain hand symptoms (to be specified later)

cause = entrapment of the median nerve in the Carpal Tunnel at the wrist



An excellent and comprehensive account of CTS is given in Rosenbaum & Ochoa (1993).

2.3.1 (ORDINAL) Response Variable

$$Y = \begin{cases} 1: \text{ No Abnormality Detected (NAD)} \\ 2: \text{ Mild CTS} \\ 3: \text{ Moderate CTS} \\ 4: \text{ Severe CTS} \end{cases}$$

Important property of ordinal Y: the event $\{Y \leq j\}$ is defined. For a non-ordinal Y, the statement "$Y \leq j$" is meaningless.

2.3.2 Predictor Variables

These are contained in the vector $\mathbf{x} = (x_1, \ldots, x_p)^T$, which divides into three types of variables:

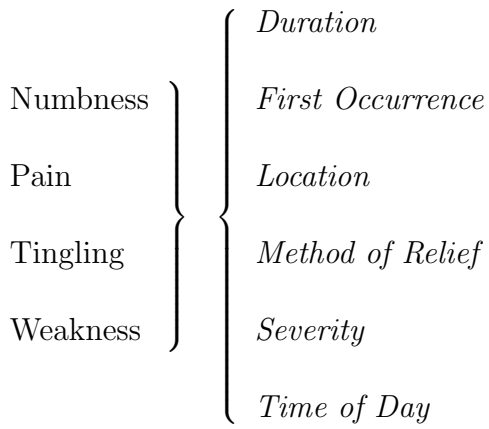
$$\mathbf{x} = \begin{cases} \textit{history} \\ \textit{clinical signs} \\ \textit{nerve conduction studies} \end{cases}$$

- HISTORY
  - Age (*in years*)
  - Sex (*Male/Female*)

  together with the **symptoms**
  - Numbness
  - Pain
  - Tingling
  - Weakness

  qualified by their *descriptors* (defined on the next page).

Descriptors of Symptom Variables



Coding of Descriptors for Numbness, Pain, Tingling and Weakness

Descriptor	Coding (0 = symptom absent)	
Duration	1	at most 10 minutes
	2	over 10 minutes
First	1	less than 3 months
	2	from 3 months to one year
	3	from 1 to 5 years
	4	from 6 to 10 years
	5	over 10 years
Location	1	first to third fingers
	2	fourth and fifth fingers
	3	all five fingers
	4	other
Relief	1	shaking hand
	2	other
	3	none
Severity	1	mild
	2	moderate
	3	severe
Time	1	daytime episodes
	2	nocturnal episodes
	3	episodes day and night
	4	continuous symptom

- CLINICAL SIGNS

Coding of Variables

Variable	Coding (0 = symptom absent)	
Sensory Loss		
Location	1	first to third fingers
	2	fourth and fifth fingers
	3	all fingers
	4	other
Wasting		
Location	1	Thenar Eminence
	2	Hypothenar Eminence
	3	other
Severity	1	mild
	2	moderate
	3	severe
Weakness		
Location	1	Thenar Eminence
	2	Hypothenar Eminence
	3	other
Severity	1	mild
	2	moderate
	3	severe

- NERVE CONDUCTION STUDIES

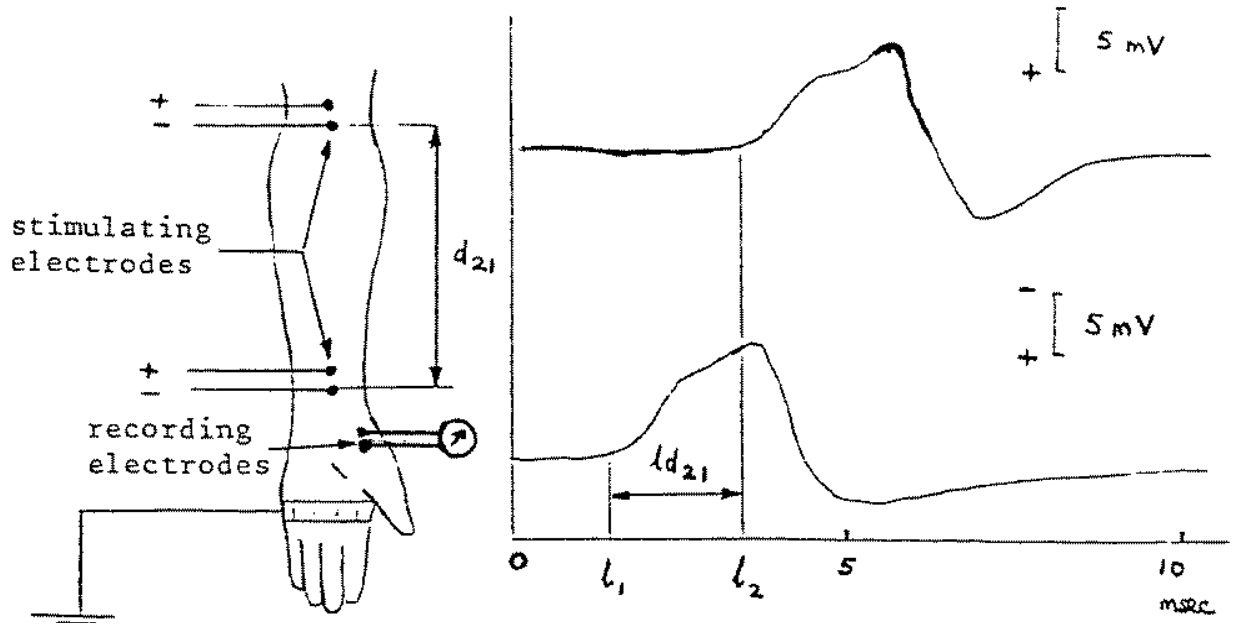
Variables

Nerve	Measurement
Median	Motor Latency at the Wrist
	Motor Latency at the Elbow
	Motor Rate, Elbow to Wrist
	Sensory Latency
	Sensory Amplitude
	Sensory Duration
Ulnar	Motor Latency at the Wrist
	Motor Latency at the Elbow
	Motor Rate, Elbow to Wrist

Units

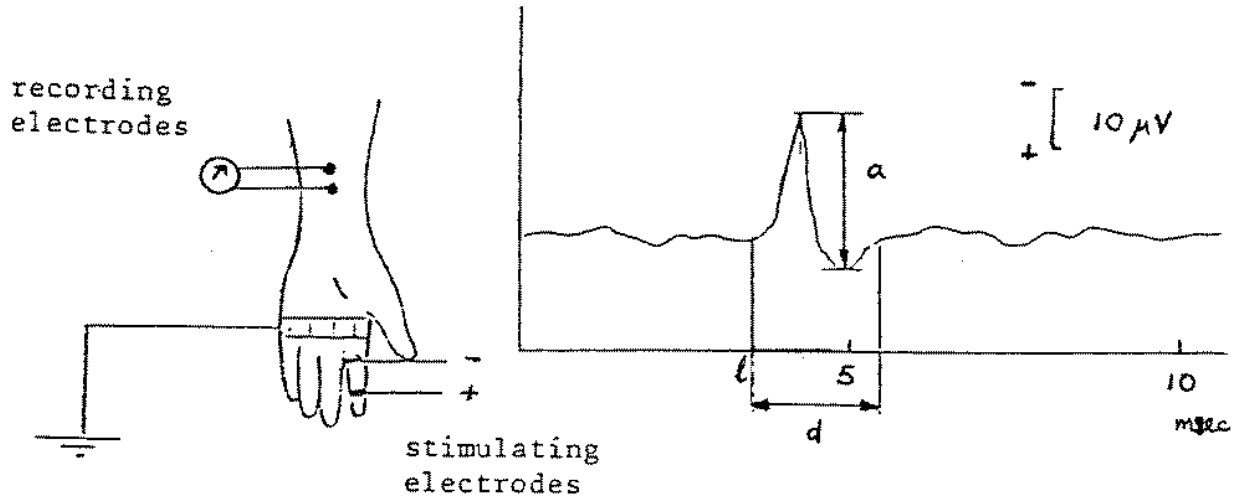
Amplitudes	microvolts
Durations	milliseconds
Latencies	milliseconds
Rates	metres per second

Stimulation of Median Nerve Motor Fibres



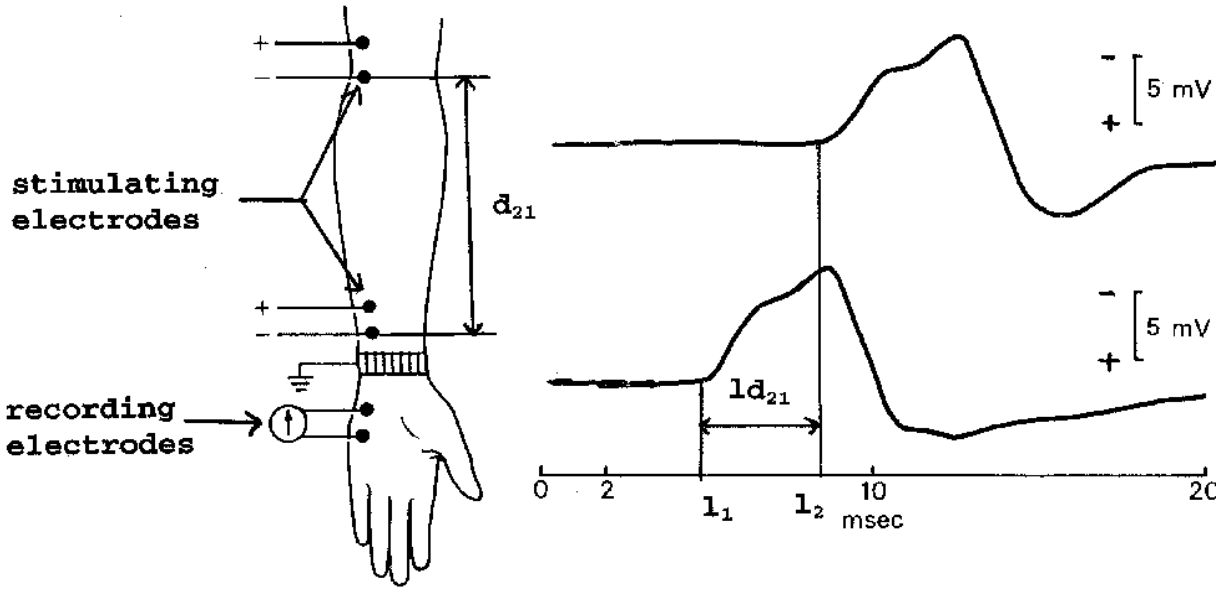
$$
\begin{aligned}
l_1 &= \text{latency at the wrist (milliseconds)} \\
l_2 &= \text{latency at the elbow (milliseconds)} \\
d_{21} &= \text{distance, elbow to wrist (centimeters)} \\
ld_{21} &= l_2 - l_1 \\
&= \text{latency difference, elbow to wrist} \\
r_{21} &= 10d_{21}/ld_{21} \\
&= \text{rate, elbow to wrist (meters/second)}
\end{aligned}
$$

Stimulation of Median Nerve Sensory Fibres



a = median sensory amplitude (microvolts)
d = median sensory duration (milliseconds)
l = median sensory latency (milliseconds)

Stimulation of Ulnar Nerve Motor Fibres



l_1 = latency at the wrist (milliseconds)

l_2 = latency above elbow (milliseconds)

d_{21} = distance, above elbow to wrist (centimeters)

$ld_{21} = l_2 - l_1$

= latency difference, above elbow to wrist

$r_{21} = 10d_{21}/ld_{21}$

= rate, above elbow to wrist (meters/second)

Non-Response to electrical stimulation
occurs in the median motor and sensory measurements

Table 1: Distribution of non-responses for the whole dataset

	NAD	MILD CTS	MOD CTS	SEV CTS	ALL
Median Motor Wrist	0	0	2	16	18
Median Motor Elbow	0	1[a]	2[a]	17	20
Median Sensory	0	17	128	85	230
Total	0	18	132	118	268

[a]These are likely to have been technical errors (inability to elicit a response rather than non-excitability of the nerve).

Note: sensory fibres are thinner than motor fibres, hence are damaged more easily.

Coding of non-responses: *Pseudo-values*

These were taken as 99.9 for latencies and durations, 0 for amplitudes and motor rates. These aren't actual physical measurements, but represent plausible codings: if there is no response, then the amplitude of the waveform will be zero, as will be the motor rate; similarly, the time to response (latency) will be "infinite". We have taken the largest possible number available in the given format (F4.1 in FORTRAN notation), namely, 99.9. The choice of 99.9 for non-response duration is somewhat arbitrary, but seems to fit in with the overall pattern.

2.3.3 Definition of Carpal Tunnel Syndrome

- varies among doctors
- most doctors would accept that CTS results from a median nerve lesion at the wrist (see section 2.3.4)
- "typical" hand symptoms: some of *numbness, pain,tingling, weakness*
  - lasting at most ten minutes
  - at night (waking patient)
  - in the first to third fingers
  - relieved by shaking hand

2.3.4 Cause of Carpal Tunnel Syndrome

is agreed by most doctors to be the entrapment of (pressure on) the median nerve at the wrist, with resulting damage to the nerve at that point (median nerve lesion at the wrist).

There are many and varied reasons for this entrapment: for example,

- fracture of the wrist
- rheumatoid arthritis of the wrist
- fluid retention, as in pregnancy

The symptoms described in section 2.3.3 can also be caused by damage to

- median nerve at the elbow or shoulder
- nerves in the neck

`ONLY NERVE CONDUCTION STUDIES CAN FIND OUT THE EXACT LOCATION OF THE NERVE DAMAGE`

2.3.5 Source of the dataset

994 patients referred with suspected Carpal Tunnel Syndrome (CTS)

- to the Electromyography Clinics of the late Dr. John L James, Consultant Physician, St Luke's Hospital, Huddersfield, Yorkshire, England
- between March 1991 and March 1994

Dr. James' diagnoses of the examined hands:

- NAD
- Mild CTS
- Moderate CTS
- Severe CTS
- Non-CTS Abnormality (possibly with some severity of CTS)

Non-CTS Abnormality class was omitted from the study, since it was very *inhomogeneous*.

Table 2: Distribution of hand diagnoses for the whole dataset

NAD	MILD CTS	MOD CTS	SEV CTS	TOTAL
777	621	294	102	1794

If you compare this table with the corresponding one in Rudolfer (2001), you will see that it contains 20 more hands. The reason for this is that, because of Section 2.7.3, the 20 hands with non-responses in their median motor readings can now be included in the dataset.

2.4 Models considered so far: Rudolfer(2001), Sections 3.2–3.4

2.4.1 Proportional Odds (PO)

It assumes

- cutpoints $\alpha_0,\ \alpha_1,\ \ldots,\ \alpha_J$ satisfying the condition

$$-\infty\ =\ \alpha_0\ <\ \ldots\ <\ \alpha_{J-1}\ <\ \alpha_J\ =\ \infty, \tag{1}$$

- a linear form for the logit of the cumulative probabilities $\gamma_j(\mathbf{x})$ $(j = 1, \ldots, J-1)$, which is equivalent to

$$\gamma_j(\mathbf{x}) = \frac{\exp(\alpha_j - \boldsymbol{\beta}^T\mathbf{x})}{1 + \exp(\alpha_j - \boldsymbol{\beta}^T\mathbf{x})}, \tag{2}$$

  where

$$\gamma_j(\mathbf{x}) = P(Y \leq j|\ \mathbf{x})$$

  is the cumulative probability of $Y \leq j$ given $\mathbf{x}$.

2.4.2 Continuation Ratio (CR)

For $j = 1, \ldots, J$, let

$$\pi_j(\mathbf{x}) = P(Y = j|\mathbf{x})$$

and the (backward) *Continuation Ratio (CR)*

$$\begin{aligned}\delta_j(\mathbf{x}) &= P(Y = j|Y \leq j, \mathbf{x}) \\ &= \frac{\pi_j(\mathbf{x})}{\pi_1(\mathbf{x}) + \ldots + \pi_j(\mathbf{x})}.\end{aligned}$$

Then

$$\begin{aligned}\delta_1(\mathbf{x}) &= 1, \\ \delta_J(\mathbf{x}) &= \pi_J(\mathbf{x}),\end{aligned}$$

and for $j = 2, \ldots, J-1$,

$$\pi_j(\mathbf{x}) = \frac{\delta_j(\mathbf{x})}{1-\delta_j(\mathbf{x})}[\pi_1(\mathbf{x}) + \ldots + \pi_{j-1}(\mathbf{x})].$$

The CR model is given by

$$\text{logit}\{\delta_j(\mathbf{x})\} = \alpha_j - \boldsymbol{\beta}^T\mathbf{x}, \tag{3}$$

and the corresponding $\pi_j(\mathbf{x})$ is

$$\pi_j(\mathbf{x}) = \begin{cases} \frac{1}{\prod_{t=2}^{J}[1+\exp(\alpha_t - \boldsymbol{\beta}^T\mathbf{x})]} & (j=1), \\ \frac{\exp(\alpha_j - \boldsymbol{\beta}^T\mathbf{x})}{\prod_{t=j}^{J}[1+\exp(\alpha_t - \boldsymbol{\beta}^T\mathbf{x})]} & (2 \le j \le J). \end{cases} \tag{4}$$

($\prod_{t=j}^{J}$ denotes the **product** from $t = j$ to J.)

$\pi_j(\mathbf{x})$ can also be expressed directly in terms of the δ_js:

$$\pi_j(\mathbf{x}) = \begin{cases} \prod_{t=2}^{J}[1-\delta_t(\mathbf{x})] & (j=1), \\ \delta_j(\mathbf{x})\prod_{t=j+1}^{J}[1-\delta_t(\mathbf{x})] & (1<j<J), \\ \delta_J(\mathbf{x}) & (j=J) \end{cases} \tag{5}$$

2.5 Classification procedures: Rudolfer (2001), Section 3.6

Let $d(\mathbf{x})$ be the classification function evaluated at the covariate vector $\mathbf{x}$:

$$d(\mathbf{x}) \text{ is the index of the group into which } \mathbf{x} \text{ is classified.}$$

Let $\hat{d}(\mathbf{x})$ denote the estimate of $d(\mathbf{x})$ obtained by substituting the maximum likelihood estimates for the corresponding parameters in the model.

2.5.1 Highest Probability (HP)

$$\hat{d}(\mathbf{x}) \;=\; \underset{1\le k\le J}{\text{argmax}}\; \hat{P}(Y=k \mid \mathbf{x}), \tag{6}$$

where $\hat{d}(\mathbf{x})$ is chosen to be the smallest index if several groups have the same maximum conditional probability.

(6): classify $\mathbf{x}$ to the group k which *maximises* $\hat{P}(Y = k \mid \mathbf{x})$: **classify to the most probable group**.

2.5.2 Anderson and Philips (AP) for the PO model

(Anderson & Philips, 1981)

$$\hat{d}(\mathbf{x}) = j \text{ if } \hat{\alpha}_{j-1} < \hat{\boldsymbol{\beta}}^T \mathbf{x} \leq \hat{\alpha}_j \quad (j = 1, ..., J).$$

This is a valid definition, since the α_j satisfy condition (1).

The AP method assumes that the continuous unobserved latent variable Z has *logistic density*: for $-\infty < z < +\infty$

$$f_Z(z) = \frac{\exp(z - \boldsymbol{\beta}^T \mathbf{x})}{\{1 + \exp(z - \boldsymbol{\beta}^T \mathbf{x})\}^2},$$

with

$$E(Z \mid \mathbf{x}) = \boldsymbol{\beta}^T \mathbf{x}.$$

2.6 General methodology of model fitting: Rudolfer(2001), Section 4

2.6.1 Preliminary Steps

- Define response variable Y and predictor variable $\mathbf{x}$, after consultation with medical expert.
- Choose **interactions** carefully.

2.6.2 Data Reduction

- Methods ignoring Y (Principal Component Analysis, Cluster Analysis)
- Variable Selection

2.6.3 Verify the model's assumptions

- Linearity of continuous predictors
- Additivity of predictors (no interaction)
- Predictors' distribution
- Influential observations

2.6.4 Fit the model

2.6.5 Compute measures of predictive accuracy

2.6.6 Validate the model

using one or more of the following methods:

- Resubstitution
- Data splitting
- Cross-validation
- Bootstrap

2.7 Computational Aspects

2.7.1 Statistical software used this time: SAS Version 8.2

2.7.2 Previous problem: Rudolfer(2001), Section 2.5

Floating point zero division overflow occurred with SAS Version 8 when

- the descending option was selected and
- – **median motor latencies** at the wrist or elbow or
  – **median sensory latency** or **duration**

  equalled their pseudo-value 99.9 for non-response

in proc LOGISTIC (occurred in computing Pearson residuals).

2.7.3 Solution: Bad News/Good News

Bad News: SAS have not yet corrected the error with the *Descending Option* in version 8.2.

Good News: they *have* corrected the error with the *Ascending Option*.

Using the equivalence of ascending and descending options (see section 2.7.4), we *can* compute the model.

Proc LOGISTIC works with **Ordered Values**, Y^*, which are related to the Recorded Values, Y, in one of two ways:

- Ascending Option: $Y^* = Y$
- Descending Option: $Y^* =$ **reversal** of $Y = J + 1$ - Y (Rudolfer, 2001, p. 13)

The distributions of Y and Y^* for the option selected are given by proc LOGISTIC in its **Response Profile**.

The Y column gives both

- the numerical value (in brackets) and
- the formatted value (NAD, *etc.*)

For both options, the probabilities modeled are $P(Y^* \leq j|\ \mathbf{x})$ cumulated over the lower Ordered Values.

For Dr. James' data, the Response Profiles are given in Tables 3 and 4.

Table 3: Response Profile: Ascending option

Ordered Value, Y^*	Recorded Value DIAGNOSIS, Y	Total Frequency
1	(1) NAD	777
2	(2) Mild CTS	621
3	(3) Moderate CTS	294
4	(4) Severe CTS	102

Table 4: Response Profile: Descending option

Ordered Value, Y^*	Recorded Value DIAGNOSIS, Y	Total Frequency
1	(4) Severe CTS	102
2	(3) Moderate CTS	294
3	(2) Mild CTS	621
4	(1) NAD	777

2.7.4 Ascending and Descending options in proc LOGISTIC are equivalent

Define the conditional probabilities, under the ascending and descending options, of $Y = j$ given $\mathbf{x}$, $P^{(A)}(Y = j|\ \mathbf{x})$ and $P^{(D)}(Y = j|\ \mathbf{x})$, respectively, as follows: for $j = 1, 2, \ldots, J$,

$$\begin{aligned} P^{(A)}(Y = j|\ \mathbf{x}) &= P(Y = j|\ \mathbf{x}), \\ P^{(D)}(Y = j|\ \mathbf{x}) &= P(Y^* = j|\ \mathbf{x}) \\ &= P(Y = J - j + 1|\ \mathbf{x}). \end{aligned}$$

Then, for $j = 1, 2, \ldots, J$,

$$P^{(A)}(Y = j|\ \mathbf{x}) = P^{(D)}(Y = j|\ \mathbf{x}). \tag{7}$$

To see (7), note that

$$\begin{aligned} & \text{logit}\{P^{(D)}(Y \leq j|\ \mathbf{x})\} \\ = \ & \text{logit}\{P(Y^* \geq J - j + 1|\ \mathbf{x})\} \\ = \ & \text{logit}\{P(Y^* > J - j|\ \mathbf{x})\} \\ = \ & -\text{logit}\{P(Y^* \leq J - j|\ \mathbf{x})\} \\ = \ & -\{-\alpha_j + \boldsymbol{\beta}^T\mathbf{x}\} \qquad (8) \\ = \ & \text{logit}\{P(Y \leq j|\ \mathbf{x})\} \\ = \ & \text{logit}\{P^{(A)}(Y \leq j|\ \mathbf{x})\}. \end{aligned}$$

Equation (8) follows from Rudolfer(2001), p. 14.

3 FURTHER MODEL FITTING

3.1 Use of Binary Logistic Regression software for fitting Continuation Ratio (CR) models

Restructure the original dataset appropriately by repeatedly including

- corresponding data subsets and
- two new variables
  - the cutpoint CP and
  - the binary response BR at that cutpoint

Then Binary Logistic Regression software can be used to fit CR models.

3.1.1 Underlying Reason why the Method works

The continuation ratio

$$\delta_j(\mathbf{x}) \;=\; P(Y = j | Y \leq j, \mathbf{x})$$

corresponds to the subset

$$\{Y = j\}$$

of the set

$$S_j = \{Y \leq j\}.$$

The complementary subset of $\{Y = j\}$ in S_j is

$$\{Y < j\}.$$

This leads to the *binary situation* of $\{Y = j\}$ *versus* $\{Y < j\}$ within the set S_j:

$$\delta_j(\mathbf{x}) \;=\; 1 - P(Y < j | Y \leq j, \mathbf{x}).$$

3.1.2 Details of the Method

This is described explicitly by Bender & Benner (2000), following less detailed discussions in

- Armstrong & Sloan (1989)
- Berridge & Whitehead (1991)
- Scott *et al.* (1997)

Bender & Benner (2000) also give SAS and S-Plus code to achieve this coding.

(1) Start with the whole dataset

$$S_J \;=\; \{Y \leq J\},$$

and define the cutpoint CP and binary response BR as follows:

$$CP \;=\; J$$

$$BR \;=\; \begin{cases} 0 & \text{if } Y < J \\ 1 & \text{if } Y = J \end{cases}$$

(2) Consider the dataset

$$S_{J-1} = \{Y \leq J-1\},$$

and define the cutpoint CP and binary response BR as follows:

$$CP = J-1$$

$$BR = \begin{cases} 0 & \text{if } Y < J-1 \\ 1 & \text{if } Y = J-1 \end{cases}$$

(3) Continue in this way until the last dataset

$$S_2 = \{Y \leq 2\},$$

and define the cutpoint CP and binary response BR as follows:

$$CP = 2$$

$$BR = \begin{cases} 0 & \text{if } Y = 1 \\ 1 & \text{if } Y = 2 \end{cases}$$

(4) If n_j observations have value $Y = j$, then the restructured dataset has

$$J(n_1 + n_2) + (J-1)n_3 + (J-2)n_4 + \ldots + n_J$$

observations.

(5) Fit binary logistic regression on the restructured dataset to the variable BR using the cutpoint CP as additional covariate to $\mathbf{x}$ and the relationship

$$\delta_j(\mathbf{x}) = P(BR = 1 \mid CP = j,\ \mathbf{x}). \tag{9}$$

3.2 Partial Proportional Odds (PPO) models

These were introduced by Peterson & Harrell (1990), and allow for the Common Slopes assumption to fail for some of the covariates in the PO model.

The model is thus

$$\gamma_j(\mathbf{x}) = \frac{\exp(\alpha_j - \boldsymbol{\beta}_j^T \mathbf{x})}{1 + \exp(\alpha_j - \boldsymbol{\beta}_j^T \mathbf{x})}. \tag{10}$$

It is possible to fit PPO models using

(1) Generalized Estimating Equations (Stokes *et al.*, 2000),

(2) a SAS macro provided by Scott *et al.* (1997).

3.3 Extended Continuation Ratio (ECR) models

These models were introduced by Harrell *et al.* (1998), and allow for the Common Slopes Assumption to fail for some of the covariates in the CR model. In some cases, the Common Slopes Assumption is too restrictive, so it is very useful to have more flexibility.

ECR models are easily fitted using the method of 3.1 by introducing an interaction term between the cutpoint for the response level j and the covariates $\mathbf{x}$ (Bender & Benner, 2000).

3.4 Equivalence between PO and CR models

Laara & Matthews (1985) have shown that if a *complementary log-log link function* is used instead of the logit link in the PO model, then the corresponding CR and PO models are equivalent.

With the complementary log-log link function, equation (2) of the PO model becomes

$$\gamma_j(\mathbf{x}) = 1 - \exp\left\{(-\exp(\alpha_j - \boldsymbol{\beta}^T\mathbf{x})\right\}, \tag{11}$$

while equation (3) of the CR model becomes

$$\delta_j(\mathbf{x}) = 1 - \exp\left\{(-\exp(\alpha_j - \boldsymbol{\beta}^T\mathbf{x})\right\}. \tag{12}$$

Since the logit and complementary log-log link functions are quite similar, at least for small probabilities (McCullagh & Nelder, 1989), we can expect that in general PO and CR models will produce similar fits to data.

4 FURTHER VARIABLE SELECTION

4.1 Univariate Wald's test

This is used to test the null hypothesis $H_0 : \beta = 0$ for a population parameter β. The test statistic is

$$\chi^2_{W_1} = \left\{\hat{\beta}/SE(\hat{\beta})\right\}^2, \tag{13}$$

which under mild assumptions has an asymptotic chi-squared distribution with 1 degree of freedom if H_0 is true.

Equation (13) is a special case of the **multivariate Wald's test** of $H_0 : \boldsymbol{\beta} = \mathbf{0}$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$. Its test statistic is

$$\chi^2_{W_p} = \hat{\boldsymbol{\beta}}^T \left[\hat{V}\left(\hat{\boldsymbol{\beta}}\right)\right]^{-1} \hat{\boldsymbol{\beta}}, \tag{14}$$

where $\hat{V}\left(\hat{\boldsymbol{\beta}}\right)$ is the estimated variance-covariance matrix of $\hat{\boldsymbol{\beta}}$. Under mild assumptions, $\chi^2_{W_p}$ has an asymptotic χ^2 distribution with p degrees of freedom if H_0 is true.

4.2 Score test

Let $\boldsymbol{\beta}^{(p)} = (\beta_1, \ldots, \beta_p)^T$ be the parameter of the model being fitted. The **score**, $U(\boldsymbol{\beta}^{(p)})$, of $\boldsymbol{\beta}^{(p)}$ is the vector of partial derivatives of the log likelihood, l, with respect to the components of $\boldsymbol{\beta}^{(p)}$:

$$U(\boldsymbol{\beta}^{(p)}) = (\partial l/\partial\beta_1, \ldots, \partial l/\partial\beta_p)^T. \tag{15}$$

If I denotes the **Fisher information** (variance-covariance) matrix of U, then asymptotically (Rao,1973)

$$S\left(\boldsymbol{\beta}^{(p)}\right) = U^T I^{-1} U \sim \chi^2_p, \tag{16}$$

provided the inverse I^{-1} of I exists (I is non-singular). Let $\boldsymbol{\beta}^{(p+1)}$ denote the parameter vector $\boldsymbol{\beta}^{(p)}$ with an extra component β_{p+1} at the end.

Then the **score chi-square statistic** for β_{p+1} is

$$S\left(\boldsymbol{\beta}^{(p)}\right) - S\left(\boldsymbol{\beta}^{(p+1)}\right) \sim \chi^2_1,$$

asymptotically.

4.3 Automatic Variable Selection (SAS, 2000)

4.3.1 Backward elimination

(1) Parameters for the complete model are estimated.

(2) Results of the Wald test for individual parameters are examined.

(3) The least significant variable that does not meet the level for staying in the model is removed.

Once a variable is removed from the model, it remains excluded.

(4) Continue until no other effect in the model meets the specified level for removal.

4.3.2 Fast backward elimination

(1) This method uses a computational algorithm of Lawless and Singhal (1978) to compute a first-order approximation to the remaining slope estimates for each subsequent elimination of a variable from the model.

(2) Variables are removed from the model based on these approximate estimates.

Note: Fast backward elimination is extremely efficient because the model is not refitted for every variable removed.

4.3.3 Forward selection

(1) Initial parameters for the model are estimated. These parameters are usually the intercepts alone.

(2) At each stage, the score chi-square statistics for the variables not in the model are computed.

(4) The largest of these statistics is examined. If it is significant at the specified entry level, then the corresponding variable is added to the model.

Once a variable is entered in the model, it is never removed from the model.

(5) The process is repeated until none of the remaining variables meets the specified level for entry.

4.4 The Bonferroni method

This is a conservative method, which gives an upper bound for the simultaneous probability of rejecting several null hypotheses.

It is based on

Bonferroni's inequality (Fisher & van Belle, 1993)

The probability of occurrence of one or more of a set of n events is at most the sum of their probabilities:

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i). \tag{17}$$

Equality occurs in (17) if and only if $A_1, \ldots, A_n$ are disjoint.

Bonferroni's Method:

Suppose that n simultaneous tests are to be performed, with an *overall significance level* α. That is, if the null hypothesis is true in all n situations, then the probability of incorrectly rejecting one or more of the null hypotheses is at most α.

Then perform each test at a significance level α/n.

Proof

Let A_i be the event of incorrectly rejecting the ith null hypothesis.

Then the probability of incorrectly rejecting one or more of the null hypotheses is at most

$$\sum_{i=1}^{n}(\alpha/n) = \alpha.$$

4.5 Occam's (or Ockham's) razor

A philosophical principle attributed to the 14th century logician and Franciscan friar, William of Occam (or Ockham).

Ockham was the village in the English county of Surrey where he was born.

The principle states:

Entities should not be multiplied unnecessarily.

The statistical application of Occam's Razor is (Aitkin *et al.*, 1989)

Never fit a more complex model than adequately describes the data:

if two models fit a dataset about equally well, then select the simpler model.

In Latin, Occam's razor is:

(1) *Pluralitas non est ponenda sine neccesitate.*

(2) *Frustra fit perplura quod potest fieri per pauciora.*

(3) *Entia non sunt multiplicanda praeter necessitatem.*

In fact, only forms (1) and (2) appear in his surviving works; form (3) was written by a later scholar.

William used the principle to justify many conclusions, including the statement that

"God's existence cannot be deduced by reason alone."

That one did not make him very popular with the Pope!

The most useful statement of the principle for scientists is,

"When you have two competing theories which make exactly the same predictions, the one that is simpler is the better."

Stephen Hawking explains in A Brief History of Time (Hawking, 1988):

"We could still imagine that there is a set of laws that determines events completely for some supernatural being, who could observe the present state of the universe without disturbing it.

However, such models of the universe are not of much interest to us mortals.

It seems better to employ the principle known as Occam's razor and cut out all the features of the theory which cannot be observed."

5 PERFORMANCE EVALUATION/MEASURES OF PREDICTIVE ACCURACY: Rudolfer(2001), Section 4.5

- Concordant/discordant pairs of observations
- Symmetric measure of association: Goodman & Kruskal's Gamma
- Asymmetric measures of association: Somers' D
- Scaled Somers' D: Concordance index
- Accuracy index: Percentage correct
- Chance-corrected measure of agreement: Cohen's Kappa statistic
- Calibration index: Brier's score

Measures of predictive accuracy indicate how well a diagnostic algorithm can distinguish between classes based on the covariates.

5.1 Preliminary definitions

All measures but the Brier score are based on **contigency tables**.

The cross-classification of observations (hands) is expressed by an $r \times c$ **Contingency Table** of *observed counts* for data cross-classified by

- *ordinal* row (R)
- *ordinal* column (C)

classifications.

For Dr. James' data, $r = c = 4$, and

R = Dr. James' diagnosis
C = Predicted diagnosis

Table 5: **General $r \times c$ Contingency Table**

		C					Row
		1	...	j	...	c	$Totals$
	1	n_{11}	...	n_{1j}	...	n_{1c}	n_{1+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
R	i	n_{i1}	...	n_{ij}	...	n_{ic}	n_{i+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	r	n_{r1}	...	n_{rj}	...	n_{rc}	n_{r+}
	$Column$ $Totals$	n_{+1}	...	n_{+j}	...	n_{+c}	n

where

n_{ij} = number of observations falling in the (i,j)th cell of the table
= number with $R = i$ and $C = j$
= observed value of a random variable N_{ij} obtained by taking random samples of size n from the population of all possible hands seen by Dr. James,

n = total sample size.

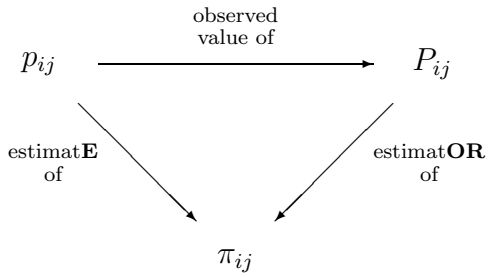
Observed and random proportions p_{ij} and P_{ij} are defined, respectively, by

$$p_{ij} = n_{ij}/n \quad \text{and} \quad P_{ij} = N_{ij}/n.$$

These are related to

π_{ij} = the probability of a randomly selected observation falling in the (i, j)th cell of the table

by the relationships



p_{ij} gives rise to

Table 6: **Observed Table of Proportions**

		C					*Row*
		1	...	j	...	c	*Totals*
	1	p_{11}	...	p_{1j}	...	p_{1c}	p_{1+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
R	i	p_{i1}	...	p_{ij}	...	p_{ic}	p_{i+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	r	p_{r1}	...	p_{rj}	...	p_{rc}	p_{r+}
	Column Totals	p_{+1}	...	p_{+j}	...	p_{+c}	1

Table 6 is an observed value of the random table

Table 7: **Random Table of Proportions**

		C					*Row*
		1	...	j	...	c	*Totals*
	1	P_{11}	...	P_{1j}	...	P_{1c}	P_{1+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
R	i	P_{i1}	...	P_{ij}	...	P_{ic}	P_{i+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	r	P_{r1}	...	P_{rj}	...	P_{rc}	P_{r+}
	Column Totals	P_{+1}	...	P_{+j}	...	P_{+c}	1

Table 6 is an estimate of the probabilities' table

Table 8: **Table of Probabilities**

		C					Row
		1	...	j	...	c	$Totals$
	1	π_{11}	...	π_{1j}	...	π_{1c}	π_{1+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
R	i	π_{i1}	...	π_{ij}	...	π_{ic}	π_{i+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	r	π_{r1}	...	π_{rj}	...	π_{rc}	π_{r+}
	$Column$ $Totals$	π_{+1}	...	π_{+j}	...	π_{+c}	1

5.2 Concordant/Discordant pairs of observations

An observed pair of observations $(r_1, c_1), (r_2, c_2)$ is called (Agresti, 1990)

$$\begin{array}{lll} \textbf{concordant} & \text{if} & (r_1 - c_1)(r_2 - c_2) > 0 \\ \textbf{discordant} & \text{if} & (r_1 - c_1)(r_2 - c_2) < 0 \\ \textbf{tied} & \text{if} & (r_1 - c_1)(r_2 - c_2) = 0 \end{array}$$

Thus, the pair is *concordant* if the subject ranking higher on r also ranks higher on c; it is *discordant* if the subject ranking higher on r ranks lower on c; it is *tied* otherwise.

A randomly selected pair of observations

$$(R_1, C_1), (R_2, C_2)$$

is called

$$\begin{array}{lll} \textbf{concordant} & \text{if} & (R_1 - C_1)(R_2 - C_2) > 0 \\ \textbf{discordant} & \text{if} & (R_1 - C_1)(R_2 - C_2) < 0 \\ \textbf{tied} & \text{if} & (R_1 - C_1)(R_2 - C_2) = 0 \end{array}$$

Let

$$\begin{array}{rcl} n_c & = & \text{observed number of concordant pairs} \\ n_d & = & \text{observed number of discordant pairs} \\ n_t & = & \text{observed number of tied pairs} \end{array}$$

based on Table 5.

Then (SAS, 2000)

$$n_c = \sum_i \sum_j n_{ij} \left\{ \sum_{k>i} \sum_{l>j} n_{kl} + \sum_{k<i} \sum_{l<j} n_{kl} \right\}, \tag{18}$$

$$n_d = \sum_i \sum_j n_{ij} \left\{ \sum_{k>i} \sum_{l<j} n_{kl} + \sum_{k<i} \sum_{l>j} n_{kl} \right\}, \tag{19}$$

$$n_t = \sum_i \sum_j n_{ij} \left\{ n_{ij} + \sum_{l \neq j} n_{il} + \sum_{k \neq i} n_{kj} \right\}. \tag{20}$$

Let a pair of observations be randomly selected, and

$$\begin{aligned} \pi_c &= \text{probability the pair is concordant} \\ \pi_d &= \text{probability the pair is discordant} \\ \pi_t &= \text{probability the pair is tied} \end{aligned}$$

Then (Agresti, 1990)

$$\pi_c = \sum_i \sum_j \pi_{ij} \left\{ \sum_{k>i} \sum_{l>j} \pi_{kl} + \sum_{k<i} \sum_{l<j} \pi_{kl} \right\}, \tag{21}$$

$$\pi_d = \sum_i \sum_j \pi_{ij} \left\{ \sum_{k>i} \sum_{l<j} \pi_{kl} + \sum_{k<i} \sum_{l>j} \pi_{kl} \right\}, \tag{22}$$

$$\pi_t = \sum_i \sum_j \pi_{ij} \left\{ \pi_{ij} + \sum_{l \neq j} \pi_{il} + \sum_{k \neq i} \pi_{kj} \right\}. \tag{23}$$

5.2.1 Properties of π_c, π_d, π_t

(1) The total number of pairs of observations, $n_c + n_d + n_t = n(n-1)/2$.

(2) $\pi_c + \pi_d + \pi_t = 1$.

(3) For $i = J$ or $j = J$, the first pairs of inner summations in equations (18)–(19) and (21)–(22) do not exist.

(4) For $i = 1$ or $j = 1$, the second pairs of inner summations in equations (18)–(19) and (21)–(22) do not exist.

(5) Replacing n_{ij} in equations (18)–(20) by its corresponding random variable N_{ij}, we get the random numbers of concordant, discordant and tied pairs of observations N_c, N_d and N_t, respectively.

(6) If the classifications R and C are independent, that is, for $i, j = 1, \ldots, J$,

$$\pi_{ij} = \pi_{i+}\pi_{+j}, \tag{24}$$

then

$$\pi_c = \pi_d. \tag{25}$$

This is an intuitive result, since if R and C are independent, then pairs of observations are just as likely to be concordant as discordant.

Proof of (25)

If (24) holds, then

$$\begin{aligned}
\pi_c - \pi_d &= \sum_i \sum_j \pi_{i+}\pi_{+j} \\
&\quad \left\{ \sum_{k>i} \sum_{l>j} \pi_{k+}\pi_{+l} + \sum_{k<i} \sum_{l<j} \pi_{k+}\pi_{+l} \right\} \\
&- \sum_i \sum_j \pi_{i+}\pi_{+j} \\
&\quad \left\{ \sum_{k>i} \sum_{l<j} \pi_{k+}\pi_{+l} + \sum_{k<i} \sum_{l>j} \pi_{k+}\pi_{+l} \right\} \\
&= \left[\sum_i \pi_{i+} \left\{ \sum_{k>i} \pi_{k+} - \sum_{k<i} \pi_{k+} \right\} \right] \times \\
&\quad \left[\sum_j \pi_{+j} \left\{ \sum_{l>j} \pi_{+l} - \sum_{l<j} \pi_{+l} \right\} \right]
\end{aligned}$$

Now

$$\sum_i \pi_{i+} \left\{ \sum_{k>i} \pi_{k+} - \sum_{k<i} \pi_{k+} \right\} = 0, \tag{26}$$

since each product term $\pi_{i_0+}\pi_{i_1+}$ on the left-hand side occurs exactly twice, with opposite signs.

Similarly,

$$\sum_j \pi_{+j} \left\{ \sum_{l>j} \pi_{+l} - \sum_{l<j} \pi_{+l} \right\} = 0. \tag{27}$$

Hence, $\pi_c = \pi_d$.

5.3 Goodman & Kruskal's Gamma, γ

Gamma was introduced by Goodman & Kruskal in their important paper Goodman & Kruskal (1954).

Gamma is a *symmetric* measure of association between the variables R and C of a contigency table: R and C are treated **symmetrically**, not as independent and dependent variables.

It is only based on concordant and discordant pairs, **ignoring tied pairs to avoid ambiguity.**

The population parameter γ is defined by

$$\gamma = \frac{\pi_c - \pi_d}{\pi_c + \pi_d} \tag{28}$$

$$= \frac{\pi_c - \pi_d}{1 - \pi_t}. \tag{29}$$

5.3.1 Interpretations of γ

It is

(1) the difference between the conditional probabilities

P(concordant pairs of observations|no ties)$-P$(discordant pairs of observations|no ties).

(2) "how much more probable it is to get like than unlike orders in the two classifications, when two (untied) individuals are chosen at random and independently from the population." (Goodman & Kruskal, 1979).

(3) "the proportionate excess of concordant over discordant pairs among all pairs which are fully discriminated , or fully ranked (it omits from consideration pairs which are tied on R or C or both.)" (Somers', 1962).

5.3.2 Properties of γ

(1) $-1 \leq \gamma \leq 1$.

(2) $\gamma = 1$ if and only if $\pi_d = 0$:
there is probability one that $C_1 \leq C_2$ for randomly selected observations (R_1, C_1) and (R_2, C_2) with $R_1 < R_2$.

(3) $\gamma = 1$ if the population is concentrated on an upper-left to lower-right diagonal of Table 8 (all other π_{ij}s are zero).

(4) $\gamma = -1$ if and only if $\pi_c = 0$:
there is probability one that $C_1 \geq C_2$ for randomly selected observations (R_1, C_1) and (R_2, C_2) with $R_1 < R_2$.

(5) $\gamma = -1$ if the population is concentrated on a lower-left to upper-right diagonal of Table 8 (all other π_{ij}s are zero).

(6) $\gamma = 0$ if the classifications R and C are independent [see equation (24) and property (6) of π_c, π_d, π_t], but not conversely in general except in the 2×2 case.

5.3.3 G = Estimator of γ

$$G = \frac{N_c - N_d}{N_c + N_d} \tag{30}$$

$$= \frac{N_c - N_d}{n - N_t}, \tag{31}$$

replacing π_c by N_c/n, π_d by N_d/n, and π_t by N_t/n in equations (28) and (29)

5.3.4 Properties of G

(1) $-1 \leq G \leq 1$.

(2) $G = 1$ if and only if $N_d = 0$:
there are no randomly selected discordant pairs of observations.

(3) $G = 1$ if the population is concentrated on an upper-left to lower-right diagonal of Table 7 (all other P_{ij}s are zero).

(4) $G = -1$ if and only if $N_c = 0$:
there are no randomly selected concordant pairs of observations.

(5) $G = -1$ if the population is concentrated on a lower-left to upper-right diagonal of Table 7 (all other P_{ij}s are zero).

(6) If the classifications R and C are independent, then G should be close to 0.

5.3.5 Estimated Asymptotic Variance of G

(SAS,2000) is

$$\frac{16}{(n_c+n_d)^4}\left\{\sum_i\sum_j n_{ij}\left(n_d a_{ij}-n_c d_{ij}\right)^2\right\},$$

where

$$\begin{aligned} a_{ij} &= \sum_{k>i}\sum_{l>j} n_{kl}+\sum_{k<i}\sum_{l<j} n_{kl} \\ &= \text{number of pairs } (r,c) \text{ agreeing (concordant) with}(i,j), \\ d_{ij} &= \sum_{k>i}\sum_{l<j} n_{kl}+\sum_{k<i}\sum_{l>j} n_{kl} \\ &= \text{number of pairs}(r,c) \text{ disagreeing (discordant) with } (i,j). \end{aligned}$$

5.3.6 Estimated Asymptotic Variance of G under the null hypothesis that $\gamma = 0$

(SAS,2000) is

$$\frac{4}{(n_c+n_d)^2}\left\{\sum_i\sum_j n_{ij}\left(a_{ij}-d_{ij}\right)^2-\left(n_c-n_d\right)^2/n\right\}.$$

5.4 Somers' D rank correlation indices

These were introduced by Somers (Somers, 1962) as asymmetric measures of association for ordinal variables.

They are an asymmetric modification of section 5.3, and treat rows (R) and columns (C) asymmetrically.

There are two population Somers' D measures of association: $\Delta(C|R)$ and $\Delta(R|C)$. $\Delta(C|R)$ treats R as the independent variable and C as the dependent variable; in $\Delta(R|C)$, it is the other way round.

$$\Delta(C|R)=\frac{\pi_c-\pi_d}{1-\sum_{i=1}^{J}\pi_{i+}^2} \tag{32}$$

$1-\sum_{i=1}^{J}\pi_{i+}^2$ is the probability that two randomly selected observations are not tied on R.

5.4.1 Interpretations of $\Delta(C|R)$

It is

(1) the difference between the conditional probabilities
P(concordant pairs of observations given no ties on R)
$- P$(discordant pairs of observations given no ties on R).

(2) "how much more probable it is to get like than unlike orders in the two classifications, when two individuals (untied on R) are chosen at random and independently from the population." (Goodman & Kruskal, 1979)

(3) "the proportionate excess of concordant pairs over discordant pairs among pairs not tied on the independent variable." (Somers, 1962).

5.4.2 Properties of $\Delta(C|R)$

(Goodman & Kruskal, 1979)

(1) $-1 \leq \Delta(C|R) \leq 1$.

(2) $\Delta(C|R) = 1$ if and only if $\pi_d = 0$ and

$$\sum_{i=1}^{J}\sum_{i=1}^{J} \pi_{ij}\left(\pi_{+j} - \pi_{ij}\right) \;=\; 0. \tag{33}$$

(3) $\Delta(C|R) = 1$ if and only if each column has at most one non-zero cell.

(4) $\Delta(C|R) = 1$ if and only if after removing all-zero columns, the non-zero cells descend in staircase fashion, perhaps with treads of unequal width.

(5) $\Delta(C|R) = -1$ if and only if $\pi_c = 0$ and

$$\sum_{i=1}^{J}\sum_{i=1}^{J} \pi_{ij}\left(\pi_{i+} - \pi_{ij}\right) \;=\; 0. \tag{34}$$

(6) $\Delta(C|R) = -1$ if and only if each row has at most one non-zero cell.

(7) $\Delta(C|R) = -1$ if and only if after removing all-zero rows, the non-zero cells ascend in staircase fashion, perhaps with treads of unequal height.

(8) $\Delta(C|R) = 0$ if the classifications R and C are independent [see equation (24) and property (6) of π_c, π_d, π_t], but not conversely in general except in the 2×2 case.

5.4.3 $D(C|R)$ = Estimator of $\Delta(C|R)$

is obtained from $\Delta(C|R)$ (equation (32)) by replacing π_c by N_c/n, π_d by N_d/n, and π_{i+} by N_{i+}/n:

$$\begin{aligned} D(C|R) &= \frac{N_c/n - N_d/n}{1 - \sum_{i=1}^{J} N_{i+}^2/n^2} \\ &= \frac{n\,(N_c - N_d)}{n^2 - \sum_{i=1}^{J} N_{i+}^2}. \end{aligned} \tag{35}$$

5.4.4 Properties of $D(C|R)$

(1) $-1 \le D(C|R) \le 1$.

(2) $D(C|R) = 1$ if and only if $N_d = 0$ and

$$\sum_{i=1}^{J}\sum_{j=1}^{J} N_{ij}\,(N_{+j} - N_{ij}) = 0. \tag{36}$$

(3) $D(C|R) = 1$ if the population is concentrated on an upper-left to lower-right diagonal of Table 7 (all other P_{ij}s are zero).

(4) $D(C|R) = 1$ if and only if each column has at most one non-zero cell.

(5) $D(C|R) = 1$ if and only if after removing all-zero columns, the non-zero cells descend in staircase fashion, perhaps with treads of unequal width.

5.4.5 Estimated Asymptotic Variance of $D(C|R)$

(SAS,2000) is

$$\frac{4}{w_r^4}\sum_i\sum_j n_{ij}\left\{w_r(a_{ij} - d_{ij}) - (n_c - n_d)\,(n - n_{i+})\right\}^2,$$

where

$$w_r = n^2 - \sum_i n_{i+}^2.$$

5.4.6 Estimated Asymptotic Variance of $D(C|R)$ under the null hypothesis that $\Delta(C|R) = 0$

(SAS,2000) is

$$\frac{4}{w_r^2}\left\{\sum_i\sum_j n_{ij}\,(a_{ij} - d_{ij})^2 - (n_c - n_d)^2/n\right\}.$$

5.5 Concordance index, c: scaled Somers' D

For those who prefer their indices to lie in the interval [0,1] instead of in the interval [-1,1], c is for them!

c is defined by the equation

$$c(C|R) = \frac{1}{2}\left\{d(C|R) + 1\right\}.$$

5.6 Percentage correct

This is given by the formula

$$\text{percent correct} = 100\left(\sum_i n_{ii}/n\right),$$

and gives a rough indication of the accuracy of the diagnostic method.

Note: $\sum_i n_{ii}$ is the total number of agreements between R and$C$.

Percentage correct is the other side of the coin of *error rates,* in that

$$\text{percent correct} = 100 \times (1 - \text{error rate}).$$

For a recent survey of developments in error rate research over the last ten years, see the article by Schiavo & Hand (2000).

5.7 Cohen's Kappa statistic,$\hat{\kappa}$

Cohen (1960) defined the **simple kappa coefficient** as a measure of chance-corrected interrater agreement:

$$\hat{\kappa} = \frac{P_0 - P_e}{1 - P_e},$$

where

$$\begin{array}{lcll} P_0 & = & \sum_i p_{ii} & = \text{ estimated probability of agreement between } R \text{ and C,} \\ P_e & = & \sum_i p_{i+}p_{+i} & = \text{ estimated probability of agreement between } R \text{ and } C, \\ & & & \text{assuming independence of } R \text{ and } C. \end{array}$$

5.7.1 Properties of $\hat{\kappa}$

(1) $\hat{\kappa} = 1$ when there is complete agreement between the raters.

(2) $\hat{\kappa} > 0$ indicates more agreement between the raters than could be explained by chance alone.

(3) $\hat{\kappa} < 0$ indicates less agreement between the raters than could be explained by chance alone.

(4) The minimum value of $\hat{\kappa}$ is between -1 and 0, depending on the marginal proportions.

5.7.2 Estimated Asymptotic Variance of $\hat{\kappa}$

(Fleiss *et al.*, 1969) is

$$\frac{A + B - C}{n(1 - P_e)^2},$$

where

$$\begin{aligned} A &= \sum_i p_{ii} \left[1 - (p_{i+} + p_{+i})(1 - \hat{\kappa})\right]^2, \\ B &= (1 - \hat{\kappa})^2 \sum \sum_{i \neq j} p_{ij} (p_{i+} + p_{+j})^2, \\ C &= \left[\hat{\kappa} - P_e (1 - \hat{\kappa})\right]^2. \end{aligned}$$

5.8 Brier's score, B

This was proposed by Brier (1950) as a means of measuring a weather forecaster's skill in predicting the weather - still an important topic 50 years later!

Suppose that there are J possible weather conditions that the forecaster can predict, that he does this on n occasions by giving his predicted probability p_{ij} of weather class j occurring on the ith occasion, and, for $i = 1, \ldots, n$, and $j = 1, \ldots, J$, let

$$E_{ij} = \begin{cases} 0 & \text{if the weather condition is not } j \text{ on the } i\text{th occasion} \\ 1 & \text{if the weather condition is } j \text{ on the } i\text{th occasion} \end{cases}$$

Then

$$B = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J} (p_{ij} - E_{ij})^2.$$

5.8.1 Properties of B

(1) The minimum of B is 0 for perfect forecasting ($p_{ij} = E_{ij}$).

(2) The maximum of B is J for worst possible forecasting ($p_{ij} = 1 - E_{ij}$). It seems a bit strange that Brier did not have the extra divisor J, since that would have made the maximum 1, but we use his definition anyway.

(3) B is purely a function of the predicted probabilities, and does not involve any decision rule (as do the other performance indicators).

6 RESULTS

6.1 Variable selection

Each of the three groups of variables

- History
- Clinical Signs
- Nerve Conduction Studies

was considered separately, and a best set of variables for each group selected. The variables selected from the three groups were then combined, and further selection was done.

The methods of variable selection used were

- forward selection
- fast backward elimination
- none

followed by the application of Wald's test using a significance level of 0.001, following Bonferroni's method.

6.1.1 History variables

The five variables selected were (comments as to why these selections are reasonable are given in brackets)

- age (incidence of CTS increases with age)
- sex (females are more susceptible to CTS than are males)
- pain at night (a typical CTS symptom)

- tingling at most ten minutes
- tingling over ten minutes

Tingling is another typical CTS symptom, and a duration of at most ten minutes is very indicative of CTS.

6.1.2 Clinical sign variables

Only wasting was found significant (in fact, it indicates more severe nerve damage).
The four variables selected were

- wasting in the thenar eminence (corresponds to the median nerve)
- wasting in the hypothenar eminence (corresponds to the ulnar nerve, a rather strange selection, since the ulnar nerve is not usually implicated in CTS)
- mild wasting
- moderate wasting

The last two indicate the severity of median nerve damage.

6.1.3 Nerve conduction variables

The four variables selected were

- median motor latency at the wrist
- median sensory latency
- median sensory amplitude
- median sensory duration

These variables are accepted by most neurophysiologists as being relevant to the diagnosis of CTS.

6.1.4 Combined model: M_1

This is the 13-variable model formed from the variables selected in sections 5.1.1 – 5.1.3.

6.1.5 Six-variable submodel of M_1: M_2

Forward selection and fast backward selection with Wald's test both gave the 6-variable model, in which the first two variables are history, and the last four are nerve conduction study:

- tingling at most ten minutes
- tingling over ten minutes
- median motor latency at the wrist
- median sensory latency
- median sensory amplitude
- median sensory duration

Note that no clinical sign variables have been selected here. Also, age and sex were omitted, although selected for the history-only model (this is an example of locally important variables being swamped in a larger set of variables).

6.1.6 Seven-variable submodel of M_1: M_3

Wald's test alone on the combined model, M_1, gave the seven-variable model, M_3, which had the additional clinical sign variable *wasting of the thenar eminence*, which indicates more serious CTS.

6.1.7 Full model without variable selection: M_4

This was included purely for comparison purposes, although it is quite unwieldy and almost impossible to interpret.

M_4 contains 111 variables, some of which are **ALIASED** (linear combinations of other variables). Their coefficients are set to zero by SAS, as they provide no further information.

6.2 Performance evaluation

Performance was evaluated by

- resubstitution and
- bootstrapping 200 times

6.2.1 Anderson & Philips' (AP) classification procedure performs badly

As the following table for the 7-variable model shows, there were no AP diagnoses of moderate or mild CTS.

Doctor's Diagnosis	Predicted Diagnosis NAD	Severe CTS	Total
NAD	149	628	777
Mild CTS	579	42	621
Mod CTS	292	2	294
Severe CTS	102	0	102
Total	1122	672	1794

This is confirmed by considering the range of values of $\hat{\boldsymbol{\beta}}^T\mathbf{x}$ as well as the cutpoints of the PO model: either $\hat{\boldsymbol{\beta}}^T\mathbf{x} \leq \hat{\alpha}_1$ or $\hat{\boldsymbol{\beta}}^T\mathbf{x} > \hat{\alpha}_4$.

Probable reason: poor fit of the PO model.

Because of this, only the HP results will be given.

6.2.2 Which of the two Somers' Ds to use?

Dr. James' diagnosis, considered as the row variable, in a sense influences the models' predicted diagnoses, since they are designed using Dr. James' diagnosis.

For this reason, we recommend using $D(C|R)$.

For completeness, however, we give the results for both $D(C|R)$ and $D(R|C)$, also for the symmetric measure γ.

6.2.3 Performance of M_1

	Resubstitution	Bootstrap
Gamma	0.96573	0.96535
$D(C\|R)$	0.80602	0.80544
$c(C\|R)$	0.90301	0.90272
$D(R\|C)$	0.82781	0.82742
$c(R\|C)$	0.913905	0.91371
Kappa	0.68323	0.68206
Brier	0.31826	0.31913
%correct	79.2642	79.16945

6.2.4 Performance of M_2

	Resubstitution	Bootstrap
Gamma	0.95765	0.95824
$D(C\|R)$	0.79010	0.79061
$c(C\|R)$	0.89505	0.895305
$D(R\|C)$	0.81218	0.81320
$c(R\|C)$	0.90609	0.90660
Kappa	0.66356	0.66281
Brier	0.32303	0.32318
%correct	77.9822	77.93478

6.2.5 Performance of M_3

	Resubstitution	Bootstrap
Gamma	0.95429	0.95384
$D(C\|R)$	0.79135	0.79099
$c(C\|R)$	0.895675	0.895495
$D(R\|C)$	0.81040	0.80983
$c(R\|C)$	0.90520	0.904915
Kappa	0.65716	0.65749
Brier	0.33022	0.33052
%correct	77.5362	77.55964

6.2.6 Performance of M_4

	Resubstitution	Bootstrap
Gamma	0.96955	0.96919
$D(C\|R)$	0.81502	0.81418
$c(C\|R)$	0.90751	0.90709
$D(R\|C)$	0.83460	0.83365
$c(R\|C)$	0.91730	0.916825
Kappa	0.69383	0.69245
Brier	0.29981	0.30071
%correct	79.9331	79.8450

6.3 Conclusions

(1) The four models' performances are very similar:

- all measures, except % correct, differ only in the second decimal place
- % correct differ by at most 2%

(2) For most measures, Resubstitution performs better than Bootstrap: this is as usual (Resubstitution results tend to be overoptimistic).

(3) $D(C|R) < D(R|C)$: the proportion of tied pairs of Dr. James' diagnoses is less than the proportion of tied pairs of predicted diagnoses.

(4) The Brier's score B is very low (about 0.3), reflecting the fact that the predicted probabilities are rarely zero or one.

(5) M_1 and M_4 are the **best models** in terms of all the measures except Brier's score.

(6) M_3 is the best model in terms of **Brier's score** (but only in the second decimal place).

(7) Kappa is around 0.6 in all cases, indicating a fairly strong chance-corrected agreement between predicted diagnosis and Dr. James' diagnosis.

7 OPEN PROBLEMS / FURTHER WORK

7.1 Inadequacy of the AP classification rule

This rather strange result needs further investigation: could indicate poor fit of the PO model.

7.2 Further variable selection

7.2.1 Branch & Bound

This is a technique which is

- adapted from the field of *Pattern Recognition*.
- implemented in SAS (SAS, 2000).

7.2.2 Use of bootstrap to refine automatic variable selection

See Sauerbrei & Schumacher (1992).

7.3 Valedictory

- Diagnostic modelling, also called *prognostic modelling*, is an exciting, active and ever-expanding area of Medical Statistics: see *Statistica Neerlandica,* Vol.55, No.1 (2001), which is devoted to this topic.
- Statisticians can never replace doctors in their goal of accurate and robust diagnosis, but they can work together with that aim.
- At present, I am working with Dr. Jeremy D P Bland, FRCP, Consultant Neurophysiologist, Kent & Canterbury Hospital, UK, who has an ever-increasing database of over 12,000 patients.
- Carpal Tunnel Syndrome is a very active area of medical research, with many papers published in medical journals.

8 REFERENCES

Agresti, A. (1990). *Categorical Data Analysis.* John Wiley & Sons, New York, USA.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis.* John Wiley & Sons, New York, USA.

Aitkin, M., Anderson, D., Francis, B. & Hinde, J. (1989). *Statistical Modelling in GLIM.* Oxford Statistical Science Series. Oxford University Press.

Anderson, J. A. & Philips, P. R. (1981). Regression, discrimination and measurement models for ordered categorical variables. *Applied Statistics,* **30**, 22-31.

Armstrong, B. G. & Sloan, M. (1989). Ordinal regression models for epidemiologic data. *American Journal of Epidemiology,* **129**, 191-204.

Bender, R. & Benner, A. (2000). Calculating ordinal regression models in SAS and S-Plus. *Biometrical Journal,* **42**,677-699.

Berridge, D. M. & Whitehead, J. (1991). Analysis of failure time data with ordinal categories of response. *Statistics in Medicine,* **14**, 1191-1203.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**, 1-3.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**, 37-46.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, London. **Reprinted 1998 by CRC Press.**

Fisher, L. D. & van Belle, G. (1993). *Biostatistics: A Methodology for the Health Sciences*. John Wiley & Sons, New York, USA.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions, 2nd Edition*. John Wiley & Sons, New York, USA.

Fleiss, J. L., Cohen, J., & Everitt, B.S. (1969). Large-sample standard errors of Kappa and weighted Kappa. *Psychological Bulletin*, **72**, 323 -327.

Goldstein, H. (1995). *Multilevel Statistical Models*. Arnold, London.

Goodman, L. A. & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, **49**, 732-764.

Goodman, L. A. & Kruskal, W. H. (1979). *Measures of Association for Cross Classifications*. Springer-Verlag, New York-Heidelberg-Berlin.

Greenland, S. (1994). Alternative models for ordinal logistic regression. *Statistics in Medicine*, **13**, 1665-1677.

Harrell, F. E. (2001). *Regression Modelling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag, New York-Heidelberg-Berlin.

Harrell, F. E., Lee, K. L. & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, **15**, 361-387.

Harrell, F. E., Margolis, P. A., Gove, S., Mason, K. E., Mulholland, E. K., Lehmann, D., Muhe, L., Gatchalian, S., Eichenwald, H. F., & The WHO/ARI Young Infant Multicentre Study Group (1998). Development of a clinical prediction model for an ordinal outcome: The World Health Organization multicentre study of clinical signs and etiological agents of pneumonia, sepsis and meningitis in young infants. *Statistics in Medicine*, **17**, 909-944.

Hawking, S. W. (1988). *A brief history of time : from the big bang to black holes.* Bantam, London.

Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression.* John Wiley & Sons, New York, USA.

Hosmer, D. W. & Lemeshow, S. (2001). *Applied Logistic Regression, 2nd Edition.* John Wiley & Sons, New York, USA.

Laara, E. & Matthews, J. N. S. (1985). The equivalence of two models for ordinal data. *Biometrika,* **72**, 206-207.

Lawless, J. F. & Singhal, K. (1978). Efficient Screening of Nonnormal Regression Models. *Biometrics*, **34**, 318 -327.

Liebetrau, A. M. (1983). *Measures of Association, Quantitative Application in the Social Sciences, Vol. 32.* Sage Publications, Beverly Hills, California, USA.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B,* **42**, 109-142.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models, 2nd Edition.* Chapman & Hall, New York, USA.

Peterson, B. & Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, **39**, 205-217.

Rao, C. R. (1973). *Linear Statistical Inference & its Applications.* John Wiley & Sons, New York, USA.

Rosenbaum, R. B. & Ochoa, J. L. (1993). *Carpal Tunnel Syndrome and Other Disorders of the Median Nerve.* Butterworth-Heinemann, Stoneham, MA, USA.

Rudolfer, S. M. (2001). *Diagnosis of carpal tunnel syndrome using logistic regression: Lectures given at the Facultad de Matemáticas, Universidad Autónoma de Yucatán, Mérida, Yucatán, Mexico, 12-15 February, 2001.* Technical Report 2001/01 (March 2001), Manchester Centre for Statistical Science, University of Manchester. Published on the Web at
http://www.maths.man.ac.uk/DeptWeb/Homepages/smr/Reprints/mexico.pdf

Rudolfer, S. M. & Peers, I. (1999). Statistical modelling of the diagnosis of carpal tunnel syndrome. Proceedings of the 14th International Workshop on Statistical Modelling, Graz, Austria, July 19-23, 1999 (Friedl, H., Berghold, A. & Kauermann, G. (Editors)), 638-642.

Rudolfer, S. M., Paliouras, G. & Peers, I. (1996). Diagnostic strategies for carpal tunnel syndrome. Abstract Book, Sixth Biennial Conference of the European Society for Medical Decision Making (June 16-18, 1996, Torino, Italy), 65.

Rudolfer, S. M., Paliouras, G. & Peers, I. (1999). A comparison of logistic regression to decision tree induction in the diagnosis of carpal tunnel syndrome. *Computers and Biomedical Research*, **32**, 391-414.

Rudolfer, S. M., Watson, P. C. & Lesaffre, E. (1995). Are ordinal models useful for classification? A revised analysis. *Journal of Statistical Computing and Simulation*, **52**, 105-132.

SAS (2000). Statistical Computations. The FREQ Procedure. *SAS Online Document Version 8.* SAS Institute Inc., Cary, North Carolina, USA.

Sauerbrei, W. & Schumacher, M. (1992). A bootstrap procedure for model building: application to the Cox regression model. *Statistics in Medicine,* **11**, 2093-2109.

Schiavo, R. A. & Hand, D. J. (2000). Ten more years of error rate research. *International Statistical Review*, **68**, 295-310.

Scott, S. C., Goldberg, M. S. & Mayo, N. E. (1997). Statistical assessment of ordinal outcomes in comparative studies. *Journal of Clinical Epidemiology*, **50**, 45-55.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review*, **27**, 799-811.

Stokes, M. E., Davis, C. S., & Koch, G. G. (1995). *Categorical Data Analysis Using the SAS System.* SAS Institute Inc., Cary, North Carolina, USA.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical Data Analysis Using the SAS System, 2nd Edition.* SAS Institute Inc., Cary, North Carolina, USA.

Walker, S. H. & Duncan, D. B. (1967). Estimation of the probability of an event as function of several independent variables. *Biometrika*, **54**, 167-179.