

UM-HSRI-80-59

A PROCEDURE FOR STATISTICAL ANALYSIS
OF THE LOGIT MODEL
ON MTS

Phyllis A. Gimotty
Thipatai Chirachavala

The University of Michigan
HIGHWAY SAFETY RESEARCH INSTITUTE
Ann Arbor, Michigan 48109

August 1980

Prepared under contract No. DOT-HS-8-01944

National Highway Traffic Safety Administration
Department of Transportation
Washington, D.C. 20590

Prepared for the Department of Transportation, National Highway Traffic Safety Administration, under Contract No. DOT-HS-8-01944. This Document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or the use thereof.

Technical Report Documentation Page

1. Report No.		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle A Procedure for Statistical Analysis of the Logit Model on MTS				5. Report Date August 1980	
				6. Performing Organization Code	
7. Author(s) Phyllis A. Gimotty, Thipatai Chirachavala				8. Performing Organization Report No. UM-HSRI-80-59	
9. Performing Organization Name and Address Highway Safety Research Institute 2901 Baxter Road Ann Arbor, Michigan 48109				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No. DOT-HS-8-01944	
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration Department of Transportation Washington, D. C.				13. Type of Report and Period Covered Special	
				14. Sponsoring Agency Code	
15. Supplementary Notes					
16. Abstract The logit model is used to model injury severity in the National Crash Severity Study. A statistical analysis that includes logit model estimation and evaluation is described. The procedure for doing this analysis using software available on the Michigan Terminal System (MTS) is given. Programs to produce coordinate data to plot the estimated logit curves and associated confidence intervals are also described and documented.					
17. Key Words NCSS, Logit Model, Injury Severity, Model Estimation, Model Evaluation			18. Distribution Statement		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 40	22. Price

This page is reserved for the Metric Conversion Chart.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.	1
1 INTRODUCTION.	2
2 DESCRIPTION OF MODELLING AND MODEL EVALUATION TECHNIQUES. . .	3
3 A COMPUTER RUN.	5
4 DISCUSSION OF THE COMPUTER COMMANDS	15
5 EXPLANATION OF MIDAS OUTPUT	25
6 FURTHER DISCUSSION OF THE COMPUTER RUN.	30
7 GRAPHICAL DISPLAY OF THE LOGIT MODEL.	32

ACKNOWLEDGMENTS

The authors wish to acknowledge Mr. David W. Smith at the Statistical Research Laboratory for his cooperation in the development of this procedure. He made available necessary software for this logit analysis. The authors would also like to thank Ms. Sharon Derry for typing this document.

2 DESCRIPTION OF MODELLING AND MODEL EVALUATION TECHNIQUES

The logit model specifies the probability that an individual receives injuries with an OAIS code of 0, 1 or 2 (NEWOAIS3 non-severe) as a function of variables such as the crash-severity variables, the occupant variables and the vehicle-characteristic variables. The model can be expressed mathematically as follows:

$$(2-1) \quad p_i = F(c - a_1 X_{1i} - a_2 X_{2i} - \dots - a_n X_{ni})$$

where:

p_i is the probability that the i^{th} individual receives injuries with an OAIS code of 0, 1 or 2 (NEWOAIS3 non-severe).

F is the logistic distribution such that

$$(2-2) \quad p_i = (1 + \text{EXP}(-(c - a_1 X_{1i} - a_2 X_{2i} - \dots - a_n X_{ni})))^{-1}$$

a, c_1, c_2, \dots, c_n are the parameters to be estimated.

$X_{1i}, X_{2i}, \dots, X_{ni}$ are the values of the independent variables pertaining to the i^{th} individual.

The estimated model can be evaluated by substituting into Equation (2-2) the estimated values of a, c_1, c_2, \dots, c_n and calculating \hat{p}_i for the i^{th} individual. This \hat{p}_i value will be bounded by the values of zero and one. A \hat{p}_i value greater than 0.5 will predict that the i^{th} individual is likely to sustain a maximum overall injury of low or negligible severity. On the other hand, a \hat{p}_i value less than 0.5 will predict the maximum overall injury to be severe. After the model has been estimated and the \hat{p}_i values calculated for all cases which have valid values of X_1, X_2, \dots, X_n , the predicted severity and the observed severity can be compared using the information available from a twoway table that could be described by Table 2.1.

R_1 is the number of cases observed and predicted to be non-severe and therefore represents the number of correct prediction of non-severe injuries; the percent correct prediction of non-severe injuries is defined as R_1/N_1 where N_1 is the total number of observed non-severe injuries. W_1 is the number of cases which were observed to be non-severe but are being predicted severe by the model. Therefore, W_1

TABLE 2.1

Observed and Predicted Injury Severity

Observed	Predicted	
	No. of Cases Predicted to be severe \hat{p}_i less than 0.5	No. of Cases Predicted to be non-severe \hat{p}_i greater than 0.5
No. of Cases Observed Non-Severe (N_1)	W_1	R_1
No. of Cases Observed Severe (N_2)	R_2	W_2

represents the number of cases of misprediction of non-severe injuries; the percent misprediction is defined as W_1/N_1 .

R_2 is the number of cases observed and predicted to be severe and therefore represents the number of correct prediction of severe injuries; the percent correct prediction of severe injuries is R_2/N_2 where N_2 is the total number of injuries observed to be severe. W_2 is the number of cases observed to be severe but predicted by the model to be non-severe. Therefore, W_2 represents the number of misprediction of severe injuries and the percent misprediction of severe injuries is defined as W_2/N_2 . Finally, the overall correct prediction is $(R_1 + R_2)/(N_1 + N_2)$. The three terms, R_1/N_1 , R_2/N_2 and $(R_1 + R_2)/(N_1 + N_2)$ are measures of goodness of fit used to evaluate the predictive capability of the estimated model.

3 A COMPUTER RUN

There are three steps necessary to accomplish the logit analysis on MTS.

1. The user must establish access to MTS.
2. A program, called ADAAS, is used to mount a tape where the data is stored.
3. MIDAS is used to read the data and perform the statistical analyses.

This section contains an example of a computer run that performs a logit analysis. This analysis was run on the first fifteen months of NCSS data and the variable numbers are those given by the ADAAS codebook that describes this data set.

The model estimation involves estimating the values of c , a_1 , a_2 , ..., a_n while the model evaluation involves calculating the percent correct prediction as defined by R_1/N_2 , R_2/N_2 and $(R_1 + R_2)/(N_1 + N_2)$. The computer commands to carry out these tasks are listed below. These commands illustrate the procedure for the modelling of the injury severity as a function of Delta V, Age and the Rural/Urban variable for front-impact accidents based on the Phase 1 Data. Delta V and Age are the independent variables which had shown significant influence in predicting injury severity in all models (Sections 3.2 to 3.6 of Analysis of The National Crash Severity Study: Final Report). Rural/Urban, which did not appear to be a significant variable, is included here for the purpose of illustrating the use of dummy variables in the modelling. In modelling the front-impact accidents, six homogeneous subsets had been defined and the model for each of these subsets were estimated (see Section 3.3 of Analysis of The National Crash Severity Study: Final Report).

SAMPLE COMPUTER RUN

%LF03: AJ832

MTS Ann Arbor (AJ832,LF03-HSRI-MP72-CD01,00261)

*Procedure for
obtaining MTS
access*

#signon sj06

#ENTER USER PASSWORD.

?

#TERM,NORMAL,UNIV

***LAST SIGNON WAS: 16:36:41 TUE AUG 05/80

USER "SJ06" SIGNED ON AT 08:27:38 ON THU AUG 21/80

#run hsri:adaas

#EXECUTION BEGINS

@ A D A A S

@HSRI - UNIVERSITY OF MICHIGAN

@08:28:02

@AUG 21, 1980

@VERSION 07-23-79

@Conference - "\$SOURCE SKD7:ADAAS"

*ADAAS used to
mount the tape
containing the
NCSS data.*

@COMMAND

?data keyword=ncsslcoc

@ "NCSSLCOG": A magnetic tape will be mounted next

#C4475A 9TP *HSRI* SIZE=16000 LP=OFF BLK=OFF VOL=NCSSF6 'HS0120'

##*HSRI* (C4475A): Mounted on T901

@ "NCSSLCOG": Access is established

@COMMAND

?midas

@The MIDAS/OSIRIS command "FILE" keyword for "NCSSLCOG" is:

@ FILE=*HSRI* NCSSLCCDIC;*HSRI* CVCCDAT

#EXECUTION TERMINATED

#run stat:midas

#EXECUTION BEGINS

M I D A S

Statistical Research Laboratory

The University of Michigan

08:30:37

AUG 21, 1980

*MIDAS is used
to read the data
and do the
analysis of the
logit model.*

This is a copy of the printout at a terminal for an actual computer run that does an analysis of the logit model. Note that user inputs are in lower case characters and computer responses are in upper case characters. A description of the various steps are shown at the right margin.

SAMPLE COMPUTER RUN (Continued)

COMMAND
?osiris file=*hsri* ncssl0ccdi;*hsri* cvoccdat maximum=&
?15000 option=none cases=all variables=18(9),36(9),109(99) &
?,115(9),116(10),118(9),136(2),371,1002(99),1004,1109(9)

*Eleven variables
are read from
the tape.*

OSIRIS "*HSRI* NCSSL0CCDI" INPUT ALL
FROM DATA FILE "*HSRI* CVOCCDAT"

14652 CASES READ FOR 11 VARIABLES

COMMAND
?\$release *hsri*
#T901 RELEASED.

*The tape is
released.*

COMMAND
?missing variable=371 value=0

*These two
MISSING commands
define MIDAS
missing data.*

MISSING DATA SUMMARY

VARIABLE	#MISSING	#LEVELS
371.1-TOTAL	6787	
TOTAL MISSING	6787	
TOTAL # CASES	14652	

COMMAND
?missing v=1004 val=99

MISSING DATA SUMMARY

VARIABLE	#MISSING	#LEVELS
1004.OCCUPANT	217	
TOTAL MISSING	217	
TOTAL # CASES	14652	

SAMPLE COMPUTER RUN (Continued)

COMMAND
?code v9000=1 case=v115:1*v136:1*v118:(1-3,8,9) label=front

An indicator is created for front impacts.

VARIABLE CATEGORIZATION CASES=1-GEN AR:1*CASE VEH:1*1-DAMAGE:(1-3,8,9)

VARIABLE	TOTAL	VALID	MISS LEVELS
9000.FRONT	7374	7374	0 1

COMMAND
?code v5003=1 case=v18:1*v9000:1*v116:(5,6)*v1002:(11-13) &
?label=subset

The next six CODE commands create the six front subsets.

VARIABLE CATEGORIZATION CASES=NO OF TO:1*FRONT:1*1-HORZ L:(5,6)*OCC SEAT:(11-13)

VARIABLE	TOTAL	VALID	MISS LEVELS
5003.SUBSET	510	510	0 1

COMMAND
?code v5003=2 case=v18:1*v9000:1*v116:(2,4,8,9)*v1002:11 &
?label=subset

VARIABLE CATEGORIZATION CASES=NO OF TO:1*FRONT:1*1-HORZ L:(2,4,8,9)*OCC SEAT:11

VARIABLE	TOTAL	VALID	MISS LEVELS
5003.SUBSET	789	789	0 2*

* CASES CHANGED IN EXISTING VARIABLE

COMMAND
?code v5003=3 case=v18:1*v9000:1*v116:(2,4,8,9)*v1002:(12,13) &
?label=*

VARIABLE CATEGORIZATION CASES=NO OF TO:1*FRONT:1*1-HORZ L:(2,4,8,9)*OCC SEAT:(12,13)

VARIABLE	TOTAL	VALID	MISS LEVELS
5003.SUBSET	337	337	0 3*

* CASES CHANGED IN EXISTING VARIABLE

SAMPLE COMPUTER RUN (Continued)

COMMAND
?code v5003=4 case=v18:2*v9000:1*v116:(5,6)*v1002:(11-13) &
?label=subset

VARIABLE CATEGORIZATION CASES=NO OF TO:2*FRONT:1*1-HORZ L:(5,6)*OCC
SEAT:(11-13)

VARIABLE	TOTAL	VALID	MISS LEVELS
5003.SUBSET	1448	1448	0 4*

* CASES CHANGED IN EXISTING VARIABLE

COMMAND
?code v5003=5 case=v18:2*v9000:1*v116:(2,4,8,9)*v1002:11 &
?label=subset

VARIABLE CATEGORIZATION CASES=NO OF TO:2*FRONT:1*1-HORZ L:(2,4,8,
9)*OCC SEAT:11

VARIABLE	TOTAL	VALID	MISS LEVELS
5003.SUBSET	1838	1838	0 5*

* CASES CHANGED IN EXISTING VARIABLE

COMMAND
?code v5003=6 case=v18:2*v9000:1*v116:(2,4,8,9)*v1002:(12,13) &
?label=*

VARIABLE CATEGORIZATION CASES=NO OF TO:2*FRONT:1*1-HORZ L:(2,4,8,
9)*OCC SEAT:(12,13)

VARIABLE	TOTAL	VALID	MISS LEVELS
5003.SUBSET	769	769	0 6*

* CASES CHANGED IN EXISTING VARIABLE

COMMAND
?code v9109=strata strata=v1109:0,1 cases=all label=ois

*The dichotomous
dependent variable
(injury severity)
is created.*

STRATA CATEGORIZATION STRAT=NEWOAIS3:MISS,1

VARIABLE	TOTAL	VALID	MISS LEVELS
9109.OAIS	11689	11689	0 2

SAMPLE COMPUTER RUN (Continued)

COMMAND
?code v9036=strata strata=v36:1,2 label=rural

The urbanization variable is recoded.

STRATA CATEGORIZATION STRAT=RURAL/UR:1,2

VARIABLE	TOTAL	VALID	MISS	LEVELS
9036.RURAL	14651	14651	0	2

COMMAND
?trans v=9036 result=3601,3602 function=dummy strata=none &
?label=same

Dummy variables for urban/rural are created.

DUMMY TRANSFORMATION

VARIABLE	TOTAL	VALID	MISS
3601.RURAL	14652	14651	1
3602.V3602	14652	14651	1

COMMAND
?probit option=matrix max=5 function=logistic rel=* strata=v9109 &
?case=v5003:1 var=1004,3601,371

Logit analysis

PROBIT ANALYSIS CASES=SUBSET:1

NO. ITERATIONS= 3 FUNCTION= LOGISTIC

NO. CASES= 243 SUM WEIGHTS= 243.00

MODEL:	MARGINAL	FULL	DIFF	DF=	SIG=
-2LOG (LIKELIHOOD)	297.05	251.50	45.553	3	.0000
FRACTION EXPLAINED	.5427	.5960	.0533		

FULL MODEL ESTIMATES	COEFF	STD ERR	RATIO
OAIS: (1) (2)	3.2241	.52672	6.1211
1004.OCCUPANT	.16797 -1	.10267 -1	1.6361
3601.RURAL	.27696	.31631	.87558
371.1-TOTAL	.91097 -1	.17282 -1	5.2711

SAMPLE COMPUTER RUN (Continued)

STEP -2*LOG(L) CONVERGENCE CRITERION

*Logit analysis
(Continued)*

0	251.83		
1	251.50	.32894	
2	251.50	.34773	-3
3	251.50	.67338	-9

CORRELATION MATRIX OF PARAMETER ESTIMATES

VARIABLE				
0.OAIS	1.0000			
1004.OCCUPANT	.6771	1.0000		
3601.RURAL	.3503	.1671	1.0000	
371.1-TOTAL	.6843	.1071	-.0373	1.0000
	0.	1004.	3601.	371.
	OAIS	OCCUPANT	RURAL	1-TOTAL

COVARIANCE MATRIX OF PARAMETER ESTIMATES

VARIABLE				
0.OAIS	.27743			
1004.OCCUPANT	.36613	-2	.10540	-3
3601.RURAL	.58358	-1	.54272	-3
			.10005	
371.1-TOTAL	.62294	-2	.19004	-4
				-.20393
				-3
				.29868
				-3
	0.	1004.	3601.	371.
	OAIS	OCCUPANT	RURAL	1-TOTAL

COMMAND

?trans v2000=linear v=1004,3601,371 rel=-.016797,-.27696,&
?-.091097.3.2241 cases=same strata=none label=*

*The following trans and
code commands create
the predicted injury
categories.*

LINEAR TRANSFORMATION CASES=SUBSET:1

VARIABLE	TOTAL	VALID	MISS
2000.V2000	510	304	206

SAMPLE COMPUTER RUN (Continued)

COMMAND

?code v2001=cuts(v2000) points=,.000001, case=same label=*

CUTS CATEGORIZATION CASES=SUBSET:1

VARIABLE	TOTAL	VALID	MISS	LEVELS
2001.V2001	510	304	206	2

COMMAND

?twoway var=2001,9109 case=same option=row%

TWOWAY CROSS-TABULATION CASES=SUBSET:1

9109.	2001.V2001			
OAIS	MISS	(1)	(2)	
N=	243			
TOTAL=	510	152	37	206
ROW%			15.2	84.8
MISS	61	54	3	58
ROW%				
(1)	170	120	11	159
ROW%			6.5	93.5
(2)	73	32	26	47
ROW%			35.6	64.4

The observed and predicted injury category are compared in the twoway table.

COMMAND

?trans v2002=exp(v2000) case=same label=*

EXP TRANSFORMATION CASES=SUBSET:1

VARIABLE	TOTAL	VALID	MISS
2002.V2002	510	304	206

The next three TRANS commands calculate the predicted probabilities of only non-severe injuries from the logit analysis.

SAMPLE COMPUTER RUN (Continued)

COMMAND

?trans v2003=1+v2002 case=same label=*

ADDITION TRANSFORMATION CASES=SUBSET:1

VARIABLE	TOTAL	VALID	MISS
2003.V2003	510	304	206

COMMAND

?trans v2004=1/v2003 case=same label=predprob

DIVIDE TRANSFORMATION CASES=SUBSET:1

VARIABLE	TOTAL	VALID	MISS
2004.PREDPROB	510	304	206

COMMAND

?histogram bystrata strata=v9109 var=2004 case=same &
?interval=21:(0,1.0)

These are histograms of the predicted probabilities of only non-severe injuries by injury classification.

HISTOGRAM <1> OASIS:1 CASES=SUBSET:1

MIDPOINT COUNT FOR 2004.PREDPROB (EACH X= 1)

0.	0	+
.50000	-1	1 +X
.10000	24	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
.15000	37	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
.20000	36	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
.25000	35	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
.30000	7	+XXXXXXX
.35000	10	+XXXXXXXXXXX
.40000	2	+XX
.45000	6	+XXXXXX
.50000	2	+XX
.55000	1	+X
.60000	4	+XXXX
.65000	1	+X
.70000	3	+XXX
.75000	0	+
.80000	0	+
.85000	0	+
.90000	0	+
.95000	0	+
1.0000	1	+X
MISSING	120	
TOTAL	290	(INTERVAL WIDTH= .50000 -1)

SAMPLE COMPUTER RUN (Continued)

HISTOGRAM <2> OAS:2 CASES=SUBSET:1

MIDPOINT COUNT FOR 2004.PREDPROB (EACH X= 1)

0.	0	+	
.50000	-1	0	+
.10000		1	+X
.15000		8	+XXXXXXXX
.20000		4	+XXXX
.25000		11	+XXXXXXXXXXXX
.30000		5	+XXXXX
.35000		8	+XXXXXXXX
.40000		4	+XXXX
.45000		3	+XXX
.50000		5	+XXXXX
.55000		4	+XXXX
.60000		5	+XXXXX
.65000		3	+XXX
.70000		2	+XX
.75000		3	+XXX
.80000		0	+
.85000		1	+X
.90000		0	+
.95000		4	+XXXX
1.0000		2	+XX

MISSING 32
 TOTAL 105 (INTERVAL WIDTH= .50000 -1)

*FINISH terminates
 the analysis in
 MIDAS*

COMMAND
 ?finish

\$5.83 CPU= 7.986 VM= 267 ELAPSED= 31.1
 #EXECUTION TERMINATED
 #sig \$
 #SJ06 08:27:38-09:16:34 THU AUG 21/80

*Command to
 terminate
 computer run.*

4 DISCUSSION OF THE COMPUTER COMMANDS

The statements contained in the above listing are comprised of MTS commands, ADAAS commands and MIDAS commands. The lower case characters are commands input by the user and the upper case characters the computer responses. In the terminal mode, MTS commands are prompted by a #. ADAAS commands are prompted for by a ?. Data from the appropriate tape can be retrieved by issuing an ADAAS command to mount the tape and issuing a MIDAS command to read the data from the mounted tape. The MIDAS commands, like the ADAAS commands, are prompted for by a ?. Details of MTS commands can be found in Introduction to MTS.³ Details of ADAAS commands are documented in Automated Data Access and Analysis System (ADAAS),⁴ while that of MIDAS commands are in Documentation for MIDAS.⁵

This section explains details of all the statements listed in the computer run.

4.1 Access to MTS

In terminal mode, in order to use MTS, the user must first issue the SIGNON command to permit access to MTS. The user enters an account identification number after the word SIGNON and returns the carriage control. Next the user is required to enter a PASSWORD. These will establish the access into the MTS.

4.2 Retrieval of Data from Tape

In order to mount the tape it is first necessary to access the ADAAS program. This is done with the MTS command:

```
#RUN HSRI:ADAAS
```

This is followed by a sequence of ADAAS commands, the first of which is

³Introduction to MTS: Introductory User Guide Series (Ann Arbor: The University of Michigan, Computing Center, September 1978).

⁴ADAAS Program Document (Ann Arbor: The University of Michigan, Highway Safety Research Institute).

⁵Daniel J. Fox and Kenneth E. Guire, Documentation for MIDAS, 3rd ed. (Ann Arbor: The University of Michigan, Statistical Research Laboratory, September 1976).

```
?DATA KEYWORD=NCSS1COC
```

This statement requests the tape containing the Phase 1 case-vehicle-occupant data (NCSS1COC) to be mounted. For the Phase 2 case-vehicle-occupant data, NCSS1COC is replaced by another appropriate tape name for the Phase 2 data⁶. There will be a response from ADAAS that will indicate whether or not it had been able to mount the requested tape. Once the tape is mounted, the next statement to enter is

```
?MIDAS
```

This statement produces the names of the files on the tape from which the data are to be read. When the names of the files have been given, the user is automatically returned to MTS.

4.3 Reading Data into MIDAS

The next task is to read the data from the files given above into MIDAS. MIDAS is accessed by issuing the MTS command:

```
#RUN STAT:MIDAS
```

The variables are read into MIDAS by the following MIDAS statement:

```
?OSIRIS FILE=*HSRI* NCSS1OCCDIC; *HSRI* CVOCCDAT MAXIMUM=&7  
?20000 CASES=ALL8 OPTION=NONE VARIABLE=...
```

In this command, the FILE keyword specifies the file name obtained after running ADAAS to mount the tape, (*HSRI* NCSS1OCCDIC;*HSRI* CVOCCDAT), the MAXIMUM keyword specifies the maximum number of cases to be read, the CASES keyword specifies the particular subset of cases to be read, i.e. it acts as a filter, the keyword OPTION=NONE specifies the option in which the data read will have the same missing data values as on the tape, and the VARIABLE keyword specifies the variables to be read from the files, the variable numbers are separated by commas. Normally, each of the variables read will be analytical. However, if the variable is immediately followed by a number contained in

⁶These keywords can be found in the ADAAS documentation.

⁷& is the line continuation indicator in MIDAS.

⁸CASES=ALL causes all cases to pass the filter.

parentheses, it will be read as a categorical variable⁹. For example, V18(9) will be read as a categorical variable with 9 levels. Any levels above 9 will be made missing with a value of zero.

The tape can now be released as the data has been read into core by issuing:

```
?$RELEASE *HSRI*
```

4.4 Model Estimation

The modelling task can be performed once the required data are read into MIDAS by a sequence of MIDAS statements. A typical MIDAS statement has the following format:

```
COMMAND MODIFIER KEYWORD1=VALUE1 KEYWORD2=VALUE2 ...
```

As an example, consider the following MIDAS statement,

```
MISSING VARIABLE=371 VALUE=0
```

Here the command is MISSING, the keywords are VARIABLE and VALUE. Each keyword appears with its value. There is no modifier in this statement. This statement will recode those cases having V371 coded as zero with the MIDAS missing data value.

MIDAS distinguishes two types of variables when used in the MIDAS statements, namely, analytical and categorical variables. The normal convention of designating variables as being analytical or categorical is that continuous measurements should be designated as analytical and discrete variables as categorical. The designation of variables is made when reading the data into MIDAS as mentioned in Section 4.3. MIDAS stores categorical variables as positive integers. Creation of additional analytical variables by transformations of categorical or analytic variables can be achieved by the TRANS command. Creation of additional categorical variables or conversion of variables originally designated as analytical to categorical can be done by the CODE command. It should be remembered that the CODE command will truncate values to make them integers and make coded zeros into the MIDAS missing

⁹Any variable can be declared categorical as long as it is coded with positive integers. Variables with valid zeros, negative, or noninteger values must be read as analytical variables.

data code. The variables appearing in the following keywords must always be categorical:

CASES keyword
STRATA keyword
VARIABLE keyword of the TWOWAY command

MIDAS excludes all cases that have the MIDAS missing data values on a variable being used from the analysis. This is done by recoding the cases with coded missing data to the MIDAS missing data code. For the model having Delta V, Age and Rural/Urban as the independent variables, the coded missing data of Delta V (V371) and Age (V1004) are made missing in MIDAS by the MISSING commands:

```
?MISSING VARIABLE=371 VALUE=0
```

```
?MISSING VARIABLE=1004 VALUE=99
```

MISSING is the MIDAS command name; VARIABLE and VALUE are keywords and each is accompanied by its "value" (e.g., 371 for the VARIABLE keyword and 0 for the VALUE keyword). The VALUE keyword specifies what value or range of values of the specified variable are to be made missing in MIDAS.

Because the modelling is on front impact accidents involving occupants in the case-vehicle passenger cars, a variable needs to be created for this class of accidents, namely, V9000. This variable is created using the CODE command:

```
?CODE V9000=1 CASES=V115:1*V136:1*V118:(1-3,8,9) LABEL=FRONT
```

V9000 is a categorical variable created from the categorical variables, V115, V136 and V118 and it has one level. This CODE statement will make all cases have a value of 1 in V9000 when and only when they have V115=1, and V136=1 and V118=1 or 2 or 3 or 8 or 9. The value of the CASES keyword in this statement is made up of the appropriate categorical variables. Each of these three categorical variables appears with its appropriate category (or categories). The variable is specified first followed by a colon and then the categories of that variable, e.g., V115:1 designates those cases of variable 115 which are coded 1 and V118:(1-3, 8, 9) designates those cases of variable 118 which are coded 1 or 2 or 3 or 8 or 9. The *'s in the CASES keyword

represent "and", while commas represent "or". The LABEL keyword has the name "FRONT" as its value. If the user prefers not to give a name to the variable, the LABEL keyword LABEL=* can be used instead.

The six independent subsets of the front impacts can now be created by six CODE commands. The first command,

```
?CODE V5003=1 CASES=V18:1*V9000:1*V116:(5,6)*V1002:(11-13) &  
?LABEL=SUBSET
```

will create the first category (to be coded 1) of Variable 5003, which corresponds to the Single Vehicle/Center Impact/All Occupants subset or CIA-1VEH.

Likewise, the next five CODE commands,

```
?CODE V5003=2 CASES=V18:1*V9000:1*V116:(2,4,8,9)*V1002:1 &  
?LABEL=SAME
```

```
?CODE V5003=3 CASES=V18:1*V9000:1*V116:(2,4,8,9)*V1002:&  
?(12,13) LABEL=SAME
```

```
?CODE V5003=4 CASES=v18:2*V9000:1*V116:(5,6)*V1002:(11-13) &  
?LABEL=SAME
```

```
?CODE V5003=5 CASES=V18:2*V9000:1*V116:(2,4,8,9)*V1002:11 &  
?LABEL=SAME
```

```
?CODE V5003=6 CASES=V18:2*V9000:1*V116:(2,4,8,9)*V1002:&  
?(12,13) LABEL=SAME
```

will create the corresponding categories (to be coded 2 to 6) of V5003 to designate the Single Vehicle/Off-Center Impact/Driver subset (OID-1VEH), the Single Vehicle/Off-Center Impact/Passenger subset (OIP-1VEH), the Two-Vehicle/Center Impact/All Occupants subset (CIA-2VEH), the Two-Vehicle/Off-Center Impact/Driver subset (OID-2VEH) and the Two-Vehicle/Off-Center Impact/Passenger subset (OIP-2VEH) respectively. Note that the newly created V5003 will be a categorical variable with six levels. The variables appearing in the CASES keywords must all be categorical. The cases which qualify to be the first category of V5003 (i.e. the CIA-1VEH subset) are those where V18=1 and V9000=1 and V116=5 or 6 and V1002=11 or 12 or 13. The LABEL keyword gives a name SUBSET to V5003. The keyword LABEL=SAME will leave the variable name unchanged.

The dependent variable of the model is a dichotomous variable designating injury severity. One category of this variable designates occupants with OAIS codes 0 through 2 and the other occupants with OAIS codes greater than 2. V1109 in the NCSS file was coded such that OAIS 0-2 is given a value of 0 and OAIS 3-6 a value of 1. Since MIDAS treats 0 as a missing value, V1109 needs to be modified before it can be used in the modelling, to give OAIS 0-2 a value of 1 and OAIS 3-6 a value of 2. This can be achieved by a MIDAS CODE command:

```
?CODE V9109=STRATA STRATA=V1109:0,1 CASES=ALL LABEL=OAIS
```

Here V9109 is the new OAIS variable which will contain two levels of injury severity, namely, OAIS 0-2 and OAIS 3-6. The former level is now coded 1 in V9109 (i.e., V9109=1) and the other level coded 2 in V9109 (i.e., V9109=2). The STRATA keyword appears with the variable (which must always be categorical) to be re-stratified, a comma is used to separate the strata to be recoded. The stratum closest to the colon will be assigned a value of 1 in V9109, the next stratum a value of 2.

The Rural/Urban variable is to be incorporated into the modelling as a dummy variable such that Rural is given a value of 1 and zero otherwise. This can be done by the following commands:

```
?CODE V9036=STRATA STRATA=V36:1,2 LABEL=RURAL
```

```
?TRANS VARIABLE=9036 RESULT=3601,3602 FUNCTION=DUMMY &  
?LABEL=SAME
```

The CODE command creates a new variable (V9036) which stratifies the variable V36 into 2 categories, Rural and Urban. V36 originally has 4 levels, 0,1,2, and 9, designating missing data, rural, urban and unknown. Variable V9036 is a two-level categorical variable. The label of this new variable is RURAL as specified by the LABEL keyword. The STRATA keyword appears with a categorical variable to be re-stratified (V36). The TRANS command takes the two categories of V9036 which have just been created and creates two dummy variables V3601 and V3602 by means of the keyword FUNCTION=DUMMY. The RESULT keyword specifies the dummy variables to be created. These two dummy variables are:

```
V3601 = 1 if rural  
        0 otherwise
```

and

V3602 = 1 if urban
0 otherwise

The LABEL keyword LABEL=SAME gives the same label-name to v3601 as V9036. Since V3601 and V3602 are the compliments of one another, only one can be included in the model.

The user now has in MIDAS two analytical variables, namely, V371 (Delta V) and V1004 (Age), two dummy variables designating Rural and Urban (V3601 and V3602), a two-level severity variable (V9109) and V5003 with six levels designating the six front-impact subsets. The modelling of injury severity as a function of Delta V, Age and Rural/Urban for each of the six subsets can be done by the PROBIT command:

```
?PROBIT OPTION=MATRIX MAXIMUM=5 FUNCTION=LOGISTIC RELATION=* &  
?STRATA=V9109 CASES=V5003:1 VARIABLE=1004,3601,371
```

In the PROBIT command, the dependent variable is indicated by the value of the STRATA keyword and must be categorical. The CASES keyword specifies the cases in a particular subset to be used in the modelling. The variable used in the CASES keyword must also be categorical. The independent variables to be included in the model are specified by the VARIABLE keyword and can be either analytical or categorical. The variables are separated by commas. The keyword OPTION=MATRIX requests the correlation and covariance matrices for the estimated coefficients to be printed with the modelling results. The keyword MAXIMUM=5 specifies the maximum number of iterations to be five. The keyword FUNCTION=LOGISTIC specifies the logistic distribution¹⁰ and the keyword RELATION=* indicates that no initial parameter values for the iterations are to be specified by the user and therefore the computer will choose a set of parameter values to use in the first iteration, and will continue until either the specified maximum number of iterations is met or the program determines further iteration will produce essentially no change in the magnitude of the estimated parameters. The logit model for other subsets can be carried out by giving the CASES keyword the appropriate values for the required subset; for example CASES=V5003:2 for OI-D-1VEH, etc. Other keywords in the statement remain unchanged.

¹⁰ MIDAS has two distributions available in the Probit command, namely, the logistic distribution and the Normal distribution.

The model for a particular subset is now estimated and the values of the estimated c , a_1 , a_2 , ..., a_n are printed.

4.5 Model Evaluation

The next task is to evaluate the goodness of fit of the estimated model. This can be achieved by three MIDAS commands:

```
?TRANS V2000=LINEAR VARIABLE=1004,3601,371 &
?RELATION=- $\hat{a}_1$ , - $\hat{a}_2$ , - $\hat{a}_3$ ,  $\hat{c}$  CASE=V5003:1 STRATA=NONE LABEL=*

?CODE V2001=CUTS(V2000) POINTS= ,.000001, CASES=SAME LABEL=*

?TWOWAY VARIABLE=2001,9109 CASES=SAME OPTION=ROW%
```

The TRANS command creates a variable V2000 (an analytical variable) which is a linear combination of \hat{c} , $-\hat{a}_1X_1$, $-\hat{a}_2X_2$ and $-\hat{a}_3X_3$. It has the following form:

$$(4-1) \quad V2000 = \hat{c} - \hat{a}_1X_1 - \hat{a}_2X_2 - \hat{a}_3X_3$$

where:

\hat{c} is the estimated threshold value (a constant).

\hat{a}_1 , \hat{a}_2 and \hat{a}_3 are the estimated coefficients of Age, the dummy variable designating Rural and Delta V.

X_1 , X_2 and X_3 are Age, the dummy variable designating Rural and Delta V.

The VARIABLE keyword specifies the variables from which the new variable (V2000) is created. The RELATION keyword specifies the constant by which each independent variable is multiplied; the last constant which has no matching variable is added to the linear combination. The order to be specified is \hat{a}_1 , \hat{a}_2 , \hat{a}_3 , \hat{c} where \hat{a}_1 , \hat{a}_2 and \hat{a}_3 are in the same order as the variables in the VARIABLE keyword. Note that \hat{a}_1 , \hat{a}_2 , and \hat{a}_3 in the RELATION keyword always have the opposite signs to \hat{a}_1 , \hat{a}_2 , and \hat{a}_3 values given in the PROBIT output. The CASES keyword specifies the particular subset being analysed. The variable in the CASES keyword must be categorical with the appropriate level designated after a colon. STRATA=NONE means that no stratification is involved; it is required here because the statement immediately preceding the TRANS statement had a STRATA keyword and MIDAS requires notification that the previous

STRATA keyword is to be disabled. The keyword LABEL=* gives no name to V2000.

The CODE command creates a categorical variable (V2001) from the analytical V2000. $V2001=CUTS(V2000)$ partitions V2000 into 2 groups according to the values of V2000. The POINTS keyword gives the cut-off value for this regrouping, the cases with V2000 less than zero or zero form one group and those with V2000 greater than zero form the other group. Here .000001 is used to approximate zero. This partitioning is equivalent to the one which divides \hat{p}_i into two groups: the first with \hat{p}_i less than or equal to .5, the second with \hat{p}_i greater than .5. The CASES keyword specifies the particular subset being analysed; if this were the same as that in the immediately preceding statement, CASE=SAME would suffice. The created variable, V2001, is categorical with 2 levels.

The TWOWAY command will give a two-way tabulation of V2001 and V9109 which has the format as shown in Table 1. The VARIABLE keyword in this statement must always appear with categorical variables. The CASES keyword again designates the subset being analysed. The OPTION keyword which appears with its value ROW% requests printing of the row percentages of the two-way table.

The histograms of the estimated probability of a non-severe injury (\hat{p}_i) can be obtained by the following five MIDAS statements.

```
?TRANS V2002=-1*V2000 CASES=SAME LABEL=*
?TRANS V2003=EXP(V2002) CASES=SAME LABEL=*
?TRANS V2004=1+V2003 CASES=SAME LABEL=*
?TRANS V2005=1/V2004 CASES=SAME LABEL=*
?HISTOGRAM BYSTRATA STRATA=V9109 VARIABLE=2005 CASES=SAME &
?INTERVAL=21:(0,1.0)
```

The four TRANS commands create a variable, V2005, which is expressed mathematically as:

$$(4-2) \quad V2005 = \hat{p}_i = (1 + \text{EXP}(-(\hat{c} - \hat{a}_1 X_1 - \hat{a}_2 X_2 - \hat{a}_3 X_3)))^{-1}$$

Notice that V2002, V2003, V2004 and V2005 are analytical variables. The CASES keywords in the above TRANS statements designate the subset being analysed.

The HISTOGRAM command plots two histograms of the \hat{p}_i values, one for non-severe and the other for severe injuries. Here, HISTOGRAM is the command name. BYSTRATA is the modifier to indicate that the histograms are to be plotted by strata specified by the STRATA keyword. The STRATA keyword is the dependent variable (V9109) with 2 levels (i.e., OAIS 0-2 and OAIS 3-6) and is categorical. The variable to be plotted is specified by the VARIABLE keyword. The CASES keyword specifies the subset being analysed and the INTERVAL keyword specifies the number of intervals to be plotted plus one, followed by a colon and then the lower and the upper values of the plotted variable in parentheses.

The user has now estimated the model for a particular subset, calculated the goodness of fit measures of the model and plotted the histograms of the estimated probability of a non-severe injury for both non-severe injuries and severe injuries. To terminate the MIDAS run, the user enters the MIDAS command:

```
?FINISH
```

This statement will tell the computer that MIDAS is no longer required by the user and it will return the user to MTS. The user then can sign off by issuing an MTS command:

```
#Signoff
```

5 EXPLANATION OF MIDAS OUTPUT

The PROBIT command, once executed, will yield the model estimation results shown below:

PROBIT ANALYSIS CASES=SUBSET:1

NO. ITERATIONS= 3 FUNCTION= LOGISTIC

NO. CASES= 243 SUM WEIGHTS= 243.00

MODEL:	MARGINAL	FULL	DIFF	DF=	SIG=
-2LOG (LIKELIHOOD)	297.05	251.50	45.553	3	.0000
FRACTION EXPLAINED	.5427	.5960	.0533		

FULL MODEL ESTIMATES	COEFF	STD ERR	RATIO
OAIS: (1) (2)	3.2241	.52672	6.1211
1004.OCCUPANT	.16797 -1	.10267 -1	1.6361
3601.RURAL	.27696	.31631	.87558
371.1-TOTAL	.91097 -1	.17282 -1	5.2711

STEP -2*LOG(L) CONVERGENCE CRITERION

0	251.83	
1	251.50	.32894
2	251.50	.34773 -3
3	251.50	.67338 -9

CORRELATION MATRIX OF PARAMETER ESTIMATES

VARIABLE				
0.OAIS	1.0000			
1004.OCCUPANT	.6771	1.0000		
3601.RURAL	.3503	.1671	1.0000	
371.1-TOTAL	.6843	.1071	-.0373	1.0000
	0.	1004.	3601.	371.
	OAIS	OCCUPANT	RURAL	1-TOTAL

COVARIANCE MATRIX OF PARAMETER ESTIMATES

VARIABLE					
0.OAIS	.27743				
1004.OCCUPANT	.36613 -2	.10540 -3			
3601.RURAL	.58358 -1	.54272 -3	.10005		
371.1-TOTAL	.62294 -2	.19004 -4	-.20393 -3	.29868 -3	
	0.	1004.	3601.	371.	
	OAIS	OCCUPANT	RURAL	1-TOTAL	

The output of the model estimation contains:

1. The subset being analysed (CASES=SUBSET:1).
2. The number of cases on which the estimation is based (N=243).
3. The Likelihood Ratio Statistic (LRS), designated by DIFF = 45.553.

In general, given the same sample size, the models with higher values of LRS are likely to fit the data better than those with lower values of LRS. The LRS has an asymptotic chi-square distribution with the degrees of freedom equal to the number of independent variables in the model.

4. The degrees of freedom, designated by DF=3.
5. The estimated parameter values (COEFF), their associated standard errors (STD ERR) and the t-statistics (RATIO). Here,

$$\begin{aligned} \hat{c} &= 3.2241 & \text{S.E.} &= 0.52672 & \text{RATIO} &= 6.1211 \\ \hat{a}_1 &= 0.01680 & \text{S.E.} &= 0.01027 & \text{RATIO} &= 1.6361 \\ \hat{a}_2 &= 0.27096 & \text{S.E.} &= 0.31731 & \text{RATIO} &= 0.8756, \text{ and} \\ \hat{a}_3 &= 0.09110 & \text{S.E.} &= 0.01728 & \text{RATIO} &= 5.2711. \end{aligned}$$

The estimated model can be expressed as follows:

$$\hat{p}_i = 3.2241 - 0.0168\text{AGE} - 0.27696\text{RURAL} - 0.0911\text{DELTA V.}$$

where

\hat{p}_i is the estimated probability of non-severe injuries,

F is the logistic distribution, and

RURAL has a value of 1 if Rural, 0 otherwise.

6. The convergence criterion for the iterations which should be of the order of 10^{-5} . Here it is 0.67338×10^{-9} .
7. The correlation matrix for the estimated parameters.
8. The covariance matrix of the estimated parameters. The covariance matrix is required for the calculation of the confidence intervals for the estimated model.

The output of the model evaluation is a two-way table:

TWOWAY CROSS-TABULATION CASES=SUBSET:1

9109.	2001.V2001			
OAIS	MISS	(1)	(2)	
N=	243			
TOTAL=	510	152	37	206
ROW%			15.2	84.8
MISS	61	54	3	58
ROW%				
(1)	170	120	11	159
ROW%			6.5	93.5
(2)	73	32	26	47
ROW%			35.6	64.4

In this example:

$$N_1 = 170$$

$$N_2 = 73$$

The percent correct prediction of non-severe injuries is 93.5

The percent correct prediction of severe injuries is 35.6, and

The percent overall correct prediction is 80.0.

The histograms can be viewed as another way of presenting the evaluation results. The histograms of the model estimated are shown below. The model for a specific subset will produce two histograms, one

for non-severe injuries and the other for severe injuries. The axes of both histograms are identical; one axis represents the estimated probability of a non-severe injury (\hat{p}_i) at a .05 interval and the other the number of cases having the particular \hat{p}_i values.

HISTOGRAM <1> OAIS:1 CASES=SUBSET:1

MIDPOINT COUNT FOR 2004.PREDPROB (EACH X= 1)

0.	0	+
.50000	-1	1 +X
.10000	24	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
.15000	37	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
.20000	36	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
.25000	35	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
.30000	7	+XXXXXXX
.35000	10	+XXXXXXXXXX
.40000	2	+XX
.45000	6	+XXXXXX
.50000	2	+XX
.55000	1	+X
.60000	4	+XXXX
.65000	1	+X
.70000	3	+XXX
.75000	0	+
.80000	0	+
.85000	0	+
.90000	0	+
.95000	0	+
1.0000	1	+X

MISSING 120

TOTAL 290 (INTERVAL WIDTH= .50000 -1)

HISTOGRAM <2> OAS:2 CASES=SUBSET:1

MIDPOINT COUNT FOR 2004.PREDPROB (EACH X= 1)

0.	0 +
.50000 -1	0 +
.10000	1 +X
.15000	8 +XXXXXXXX
.20000	4 +XXXX
.25000	11 +XXXXXXXXXXXX
.30000	5 +XXXXX
.35000	8 +XXXXXXXX
.40000	4 +XXXX
.45000	3 +XXX
.50000	5 +XXXXX
.55000	4 +XXXX
.60000	5 +XXXXX
.65000	3 +XXX
.70000	2 +XX
.75000	3 +XXX
.80000	0 +
.85000	1 +X
.90000	0 +
.95000	4 +XXXX
1.0000	2 +XX

MISSING 32

TOTAL 105 (INTERVAL WIDTH= .50000 -1)

6 FURTHER DISCUSSION OF THE COMPUTER RUN

The computer commands listed in this document will estimate the model and then evaluate the estimated model for the Single Vehicle/Center-Impact/All Occupant subset (CIA-1VEH) of frontal collisions using the Phase 1 data. If the user wishes to analyse the Phase 2 data, an appropriate change must be made to the "value" of the KEYWORD in the ADAAS DATA command, that is

```
?DATA KEYWORD=name of the tape containing the Phase 2 data
```

If the user wishes to analyse Side Impacts instead of Front Impacts, a change must be made to the CODE command creating V9000 in MIDAS. For Side Impacts, this statement becomes:

```
?CODE V9000=1 CASES=115:(2,4)*V18:2*V136:1*V118:(1-3,8,9) &  
?LABEL=SIDES
```

Note the change in the value of the CASES keyword for the Side Impacts. The analysis of Side Impacts will produce different subsets from that of the Front Impacts and therefore changes must be made to the set of CODE commands which created the Front-Impact subsets. This will require changing the values of the CASES keywords to reflect the Side-Impact subsets.

Any number of independent variables may be added to the models by specifying them in the VARIABLE keyword of the PROBIT command. The independent variables can be either analytical or categorical or dummy variables (a special form of categorical variables). If a variable is made into a set consisting of "k" dummy variables, only (k-1) dummy variables may be included in the model and not k dummy variables.

Attention should be paid to the CASES and STRATA keywords, both of which must appear with categorical variables only. A TWOWAY command also requires the variables in the VARIABLE keyword to be categorical.

Creating new variables can be achieved in the following ways:

1. To create an analytical variable from another analytical variable or a number of analytical variables, the TRANS command is used.

2. To create a categorical variable from an analytical variable, the CODE command is used.
3. To create a categorical variable from another categorical variable or a set of categorical variables, the CODE command is used.

Before estimating the model using the PROBIT command, the cases with missing data must be given MIDAS missing data values by using the MISSING command. Failure to do this will result in the generation of erroneous results.

In order to reduce the cost of subsequent runs it is recommended that a MIDAS internal dataset be created to save the data that have been read into MIDAS, so that they can be used in later runs without trouble and expense of mounting the same NCSS tape in ADAAS. The WRITE INTERNAL command, to write a permanent MIDAS dataset, is described on page 182 of Documentation for MIDAS, while the READ INTERNAL command, to read that dataset into MIDAS, is described on page 134 of the same document.

7 GRAPHICAL DISPLAY OF THE LOGIT MODEL

This section will describe two APL procedures. Both of these procedures were used to obtain coordinate data for plotting the estimated logit curve and associated confidence intervals as a function of one of the independent variables in the model. This coordinate data was then used in a graphics computer program to produce plots of the logit curve and confidence intervals. These procedures will produce coordinates that can be used in most graphical programs.

Using MIDAS the logit model can be estimated. The output from the PROBIT command produces the estimated coefficients for the model specified. The commands to do this were discussed in Section 4 and the specific output from the PROBIT command was described in Section 5. To graph the estimated logistic curve as a function of one of the independent variables, coordinates for this curve need to be calculated. The procedure PROB produces a vector of values of the estimated logistic curve for specified values of the independent variable. This procedure is given below.

```
PX←FIX PROB PAR
X←(3X((¯1)X121))+2
PAR2←(¯1)↑PAR
PAR3←(¯1)↑PAR
CON←+/FIXXPAR2
PX←1÷1+*((¯1)X(CON+XXPAR3))
```

The most important variable in the model considered is Delta V. All of the graphs produced were plotted with Delta V (or Lateral Delta V) on the X-axis. The procedure PROB is currently written to evaluate the estimated logistic function for specified values of Delta V. The values currently are set at 1, 4, 10, 13, ..., 58, 61. For each of these values the predicted value is calculated as in Equation 4-2.

There are two input parameters associated with PROB. The first parameter is PAR. It is the vector of estimated parameters obtained from the PROBIT command. It is necessary that the last estimated parameter in this vector belong to the variable represented on the X-axis, in this case Delta V. (If these procedures are to be used it is easier if, in the variable list in the PROBIT command, the last variable corresponds to the variable to be plotted on the X-axis.) For the estimated model that was presented in Section 4 it is necessary to assign PAR such that

PAR←3.2241 0.0168 0.2769 .0911.

Since this model has two other variables that can vary, it is necessary, for a graph of the estimated logistic curve, to fix the value of all other independent variables. The input parameter FIX specifies the values for all the independent variables except the last variable. In the above example, the variable omitted would be Delta V. More specifically, the first value of FIX is always 1. This 1 represents the constant term in the equation. The remaining values in FIX are the negatives of the values at which the model is to be evaluated. For example, if the logistic curve for 30 year old occupants in rural areas was to be plotted as a function of Delta V, the parameter FIX would be

FIX←1 ^30 ^1.

If the estimated probabilities for 65 year old occupants in urban areas were to be plotted as a function of Delta V then

FIX←1 ^65 0.

The output from the APL procedure PROB generates 21 values for the estimated logistic equation at the specified Delta V values of 1, 4, 7, 10, 13, ..., 58, 61.

The procedure VARIANCE calculates the half-width of the confidence interval of the predicted probability given by the estimated logit model at a particular Delta V value. This procedure is given below. This procedure is run immediately after PROB to obtain the associated half-widths for confidence intervals for the predicted probabilities calculated by the procedure PROB. Confidence intervals may then be obtained by calculating the upper and lower bounds using the following:

PX+SX (Upper Bound)

PX-SX (Lower Bound)

where PX is the vector of predicted probabilities from PROB and SX is the resultant vector from VARIANCE.

These define the approximate half-width of the 95% confidence interval for the predicted probabilities. The theory used in obtaining these results is discussed in Section 3.1 of Analysis of the National Crash Severity Study: Final Report¹¹.

The procedure VARIANCE has one input parameter, SIGMA. This parameter stores the covariance matrix for the estimated parameters in the logistic model. For the particular example shown in Section 4, SIGMA has the following value:

$$\text{SIGMA} = \begin{bmatrix} .27743 & .0036613 & .058358 & .0062294 \\ .0066613 & .0001054 & .00054272 & .000019004 \\ .058358 & .00054272 & .10005 & .00029868 \\ .0062294 & .000019004 & .00020393 & .00029868 \end{bmatrix}.$$

¹¹Phyllis A. Gimotty, Kenneth L. Campbell, Thipatai Chirachavala, Oliver Carsten, and James O'Day, Statistical Analysis of the National Crash Severity Study. Sponsored by the National Highway Traffic Safety Administration, Department of Transportation, Washington, D.C., under contract No. DOT-HS-01944 (Ann Arbor: Highway Safety Research Institute, The University of Michigan, June 1980). Report No. UM-HSRI-80-38 (release pending).

$SX \leftarrow \text{VARIANCE } \Sigma$
 $\text{DIM} \leftarrow \rho \Sigma$
 $F \leftarrow (\bar{1}), (\bar{1})$
 $\text{SIG1} \leftarrow F \dagger \Sigma$
 $\text{SIG2} \leftarrow F \dagger \Sigma$
 $\text{SIG3} \leftarrow ((\text{DIM}[1] - 1), 0) \dagger \Sigma$
 $\text{SIG3} \leftarrow, \text{SIG3}$
 $\text{SIG4} \leftarrow (\bar{1}) \dagger \text{SIG3}$
 $\text{VAR1} \leftarrow \text{FIX} + .x \text{SIG1} + .x \text{FIX}$
 $\text{VAR2} \leftarrow \text{SIG2} \times X^{\#2}$
 $\text{VAR3} \leftarrow 2 \times (\text{FIX} + . \text{SIG4}) \times X$
 $\text{V1} \leftarrow \text{VAR1} + \text{VAR2} + \text{VAR3}$
 $\text{VX} \leftarrow (\text{PX} \times 2) \times \text{V1} \times (1 - \text{PX})^{\#2}$
 $\text{S} \leftarrow \text{VX} \times 0.5$
 $\text{SX} \leftarrow 1.96 \times \text{S}$