U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

DOT HS 807 985

June 1993

NHTSA Technical Report

Imputation in the NASS General Estimates System

The United States Government does not endorse products or manufactures. Trade or manufacturer's names appear only because they are considered essential to the object of this report.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 807 985		

4. Title and Subtitle	5. Report Date
Imputation in the NASS General Estimates System	June 1993
	6. Performing Organization Code
	8. Performing Organization Report No.

7. Author(s)	
Terry S.T. Shelton	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Mathematical Analysis Division National Center for Statistics and Analysis 400 7th St. S.W. Washington, D.C. 20590	11. Contract or Grant No.
	13. Type of Report and Period Covered
	NHTSA Technical Report

12. Sponsoring Agency Name and Address	14. Sponsoring Agency Code
Research and Development National Highway Traffic Safety Administration 400 7th St. S.W. Washington, D.C. 20590	

15. Supplementary Notes

16. Abstract

This report documents a statistical procedure, called imputation, that was implemented in the National Accident Sampling System General Estimates System (GES). Imputation is a method of replacing unknown data with known data on a case by case basis. This document gives an overview of the imputation process, why it was implemented, how imputed values are best used, and some cautionary advice for when the imputed values are best not used. This report provides users of the GES datafile with the information necessary for better understanding of the GES data.

17. Key Words	18. Distribution Statement
Imputation, Hot Deck Imputation	Document is available to the public through the National Technical Information Service Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	126	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

Introduction

Unknown data are a common problem whenever data are coded in any system, and the National Accident Sampling System General Estimates System (GES) data are no different. In an attempt to increase the utility of the GES data, an imputation process has been implemented as part of the GES file preparation. This document gives an overview of the imputation process, why it was implemented, how imputed values are best used, and some cautionary advice for when the imputed data are best not used. For more information on the GES sample design and estimation procedures refer to the "National Accident Sampling System General Estimates System Technical Note, 1988 to 1990"[1/] . This document on imputation is not intended as a reference for missing data techniques, but many good references are available such as Compensating for Missing Survey Data.[2/]

Unknowns in GES

GES data are centrally coded from a sample of police accident reports from across the country. For almost all GES variables, the data coder can enter "unknown" as a valid code. Unknown data usually result from these situations:

- The police officer failed to fill out the information on the police accident report (PAR), or the information was not legible and the information could not be derived from the diagram, the narrative, or other variables; or,
- The PAR does not have a block to fill in the data and the information could not be derived from the diagram, the narrative, or other variables; or,
- The police officer checks the unknown box for a variable on the PAR.

The percentage of unknowns for particular variables can be a function of the data coding protocol. For example, if police do not fill out the box on the PAR for sex of the driver, the GES data coder must code

the variable SEX unknown. However, when no information exists on the PAR for a variable such as EJECTION, in most instances the coding default is "not ejected."

For a few GES variables, unknowns can never be coded: "yes/no" variables, such as FIRE OCCURRENCE, and computer-assigned variables, such as SAMPLING WEIGHT. This report, though, is only concerned with the variables where unknowns can and do occur. One way to deal with unknown data is the process of imputation.

What is Imputation?

Imputation has been defined by Kalton as, "the process of assigning values for the missing values to produce a complete data set." Imputation fabricates data when data are unknown. For example, suppose a surveyor called 100 people on the phone and asked them if they drive a car. Eighty said yes, ten said no and the other ten hung up on the surveyor. Instead of reporting 80 yes, 10 no and 10 unknown, the surveyor decides to guess what the ten people who hung up would have said. In this case, with no other data to go on, he might guess that since 80 out of 90 respondents were drivers, the people that didn't respond were probably drivers. The guess as to what the nonrespondents might have said is one example of imputing data for unknown data.

Imputation can be accomplished in many ways, and three different methods have been used for GES data. The evolution from summary level imputation to univariate imputation and hot-deck imputation will be described later in this report.

Perhaps the most sophisticated imputation currently used in GES is called hot-decking or hot-deck imputation. This specific type of imputation makes use of current survey data to replace unknown data and has been defined as, "a procedure where an incomplete response is completed by using values from one or more other records on the same file (i.e., from the same survey), and the choice of these records varies with the record requiring imputation.[3]" The adjective "hot" refers to an imputation process that uses current survey data. On the other hand, cold-deck imputation uses historical data or data from previous surveys. The most important characteristic of hot-decking is that related variables, called classification variables, are used to determine what the unknown data should be replaced with. Again consider the simple phone interview example above. Assume that information on location was available for all the potential respondents and the people who refused to answer and hung up lived in the city. If the 80 yes respondents lived in the suburbs and the 10 no respondents lived in the city, the surveyor may use this location information to better represent the ten refusals as no's.

Several methods of imputation are used in GES, but first the issue of why imputation should be used at all is explored.

Why Impute in GES?

The three primary reasons that data are imputed in GES are:

- historical precedence;
- convenience and consistency; and,
- potential reduction in bias.

Each reason is discussed in more detail below.

Historical Precedence:

Historically, unknown data from the GES predecessor, the National Accident Sampling System (NASS), have been published only for selected tables. Since its inception in 1979, unknown NASS data have been distributed proportionally to the known values for the annual reports. (See, for example, NASS 1986.[4]) The decision to continue this tradition with GES was made by NCSA statisticians prior to producing the first annual GES report.

Convenience and Consistency:

Another reason for imputation is convenience and consistency for the users. GES data are supposed to answer questions such as, "How many crashes occurred on Saturday night?" The GES user usually wants a definitive answer, such as, 500,000 crashes occurred on Saturday night. Most users do not know how to interpret statistics such as, 500,000 crashes on Saturday and another 10,000 crashes occurred on an unknown day. Without a definitive answer, tracking statistics from year to year or across crash types is very difficult. The best way to meet our customers' needs is to give them our best estimates with the unknowns accounted for.

Similarly, imputation allows every data user to get the same answer when using GES data and allows users to compare their results with published GES data. Without imputed values on the file, analysts would have to choose whether or not to include unknown data and what to do with the unknowns if these data are included. Not all analyses should use imputed data, though, and this is discussed later in this paper.

Potential Reduction in Bias:

Without imputation, certain estimates may be misleading or biased. For example, assume a sample of 100 crashes had this distribution:

70	No Alcohol Involved
10	Alcohol Involved
20	Unknown Alcohol Involvement
100	

One way to interpret the data is to eliminate the unknowns when making statements about alcohol involvement. One such statement could be, 10 out of 80 or 12.5 percent of the crashes involve alcohol. However, if in fact 15 of the 20 crashes with unknown alcohol involvement actually did involve alcohol then the true percentage of alcohol involvement would be 25 out of 100 or 25 percent. The 12.5 percent estimate would be a biased estimate if the true number were actually 25 percent.

Imputation is one way of trying to reduce this bias. By making intelligent "guesses" about the unknowns, the bias may very well be reduced. Comparisons of imputed GES data with other sources, such as data from 1986 NASS, have shown that the imputed data appear reasonable. However, without a followup study of the unknowns for that survey year, we cannot make a definitive statement that the bias was reduced.

Evolution of Imputation in GES

Since the decision to impute GES data was made in 1988, the process has evolved from a simple, tedious process to a complex, computerized process.

Summary Level Imputation

For the first two annual GES reports, 1988 and 1989, a manual procedure of distributing unknowns (summary level imputation) for each table was employed. This tedious process added up to a month to the report production schedule. Control totals or benchmarks were used to ensure table-to-table consistency, but the manual imputation process still required some judgement on which tables should be used as the controls. In addition, these imputed values were produced at the summary level on a table by table basis, and no imputed values were added to the file. If someone tried to reproduce the published values, they were unable to do so without following the same manual imputation procedure. No individual unknown values on the file were replaced with known values.

Univariate Imputation

Prior to the 1990 report production, a new way of distributing unknowns was introduced for most of the key report variables. Imputation was done on a case by case basis and not on the summary level. This process was univariate imputation and the imputed values were added to the GES files so users could reproduce published values. Univariate imputation refers to the process of randomly substituting known values for one variable for unknown values for the same variable. For example, assume that there are 120 people on the GES file and that the variable SEX has these values:

MALE	60
FEMALE	40
UNKNOWN SEX	20
TOTAL	120

To distribute the unknowns on a univariate basis, 60/100 or 60 percent of the unknowns would be assigned to the male category and 40/100 or 40 percent of the unknowns would be assigned to the female category. The imputed variable SEX would have these values:

MALE	72
FEMALE	48
UNKNOWN SEX	0
TOTAL	120

Univariate imputation works very well for variables that have very few unknown values or variables with very little relationship to other variables on the file. The univariate imputation was implemented in the 1988 through 1991 GES files for several variables where it was assumed that the unknowns were distributed in the same proportion as the known values. However, even though univariate imputation is still being used for some variables in GES, the procedure is not appropriate for other variables.

For example, the univariate imputation process was performed on the GES variable AGE. All persons with unknown ages, about ten percent of the people on the file, were assigned ages based on the distribution of the persons where ages were known - no other variables on the file were considered. As soon as several tables using this imputed age variable were created, the problem with using univariate imputation for age became apparent. For example, a table of drivers by vehicle type and age showed a significant percentage of medium and heavy truck drivers under the age of 18. Needless to say, a variable such as age is very dependent on other related variables.

Hot-deck Imputation

With the realization that unknowns for some variables simply cannot be replaced using univariate imputation, a more sophisticated imputation approach had to be developed. A review of the available missing data correction techniques was conducted and the approach chosen was hot-deck imputation. The next section gives a detailed description of this process.

How Hot-decking Works in GES

In summary, hot-decking replaces unknown values with known values from similar records. The process is performed on a case by case basis and values are added to the GES file. Hot-deck imputation in GES was accomplished in several steps:

1. Identification of classification variables.
2. Random assignment of a known value for the unknown value from another unit in the classification group.
3. Repetition of step 2 at a less exact level until all (or almost all) unknowns are gone.
4. Univariate imputation as a last resort for remaining unknowns.

Classification Variables

The first step, identification of the classification variables, is the key to the success of the hot-decking process. For each variable to be hot-decked, a set of correlated variables, called classification variables, were identified. Individual chi-square tests of independence were conducted between each variable to be imputed and almost all other GES variables. Some variables were left out of the chi-square testing because they didn't make logical sense. For example, for the hot-decked variable SEATING POSITION, PEDESTRIAN LOCATION was not tested as a potential classification variable. Variables that correlated well with the variable to be imputed and made logical sense as classification variables were selected as the classification variables. For example, for age, the classification variables were: SAMPLING STRATUM, HOUR OF CRASH, DAY OF WEEK, VIOLATIONS CHARGED, PERSON TYPE, SEATING POSITION, SEX, DRUG INVOLVEMENT, ALCOHOL INVOLVEMENT, NUMBER OF OCCUPANTS, and NUMBER OF VEHICLES. STRATUM was used as a classification variable for all hot-decked variables to limit imputation to records within the same stratum. This variable has four possible choices and is used at the time of sample selection to identify the severity and type of vehicles involved in the crash. For more information, see the GES Technical Note.

When identifying the classification variables, the temptation is great to use too many classification variables. Every time another variable is used, the imputed data may be more accurate, but the chance of not finding a matching record increases. Each classification variable that is added means the cells of data available to find a match contain fewer

and fewer records. Imagine cutting a pie in half, then in half again, then in half again, and so on. Eventually, the pieces are so thin, nothing is left to eat. Finding an appropriate number of classification variables as well as the proper variables themselves is crucial.

Random Sort and Match

The next step is randomly sorting all records within each cell of the n-dimensional table that is created by the n classification variables. Since n-space is not an easy concept to imagine, consider a simplified example of imputing for unknown age using only SEX and ALCOHOL. All records on the GES person file (those with known age and unknown age) fit somewhere into a two-way table created by the classification variables:

	SEX	
ALCOHOL	Male	Female
Yes	A	B
No	C	D

The records in each cell, such as males that had been drinking (cell A), are randomly sorted. If imputation is being done for a person with an unknown age, but who falls into cell A, that person is assigned an age from another record in that cell that has a known age. In other words, if you don't know someone's age, but you do know that person was male and had been drinking, chances are that his age is more like other males that had been drinking. This example has been simplified for illustrative purposes, because in the actual hot-decking process for AGE, eleven classification variables were used.

Several Iterations

The hot-decking process continues until all possible matches and assignments are made within the n-dimensional classification cells. In many cases, though, matches cannot be found within the classification cell. For example, if all the males that had been drinking also had unknown ages, no match could be found. If after the first iteration, more unknowns remain, one of the classification variables is dropped. Unknowns are then replaced using an (n-1)-dimensional classification table. For all hot-decked variables, the majority of unknowns were replaced with known values on the first iteration. However, most

variables required several iterations using one less classification variable at each iteration.

A test was done to determine what effect the order of elimination of the classification variables had on the imputation. No significant effect was discovered. Since the majority of unknowns were replaced at the first iteration, the order in which the classification variables were removed prior to subsequent iterations was not crucial.

Last Resort: Univariate Imputation

If unknown values still remained after several iterations of the hot-decking procedure, a simple univariate imputation process was used. For all hot-decked variables, the majority of unknowns were replaced during the hot-deck procedure. As a last resort, though, the few remaining unknowns were assigned in the same proportion as the known values for that one variable. The univariate imputation had little effect on the final distribution of the variable.

Speed Limit: An Example

The only variable chosen from the crash level variables for hot-decking was SPEED LIMIT. Speed Limit is a variable of great interest to many GES users, but, unfortunately, is unknown for 27 percent of the 42,600 records on the 1991 GES crash file. With that high number of unknowns, univariate imputation would not be appropriate. Eight variables, all from the crash file, were used as the classification variables for the hot-decking process. The number of attributes for each variable is given in parentheses:

1. STRATUM (4)
2. LAND USE (5)
3. ROADWAY SURFACE CONDITION (7)
4. TRAFFIC CONTROL DEVICE (21)
5. NUMBER OF TRAVEL LANES (8)
6. RELATION TO JUNCTION (9)
7. TRAFFICWAY FLOW (4)
8. INTERSTATE HIGHWAY (3)

(For definitions of these variables see Appendix C.)

These classification variables form an 8-dimensional "box" with:

4 X 5 X 7 X 21 X 8 X 9 X 4 X 3 = 423,360 potential cells.

Not all cells will be filled with data, but this example certainly exhibits the importance of not selecting too many classification variables.

In the first iteration of the hot-deck imputation, all eight classification variables were used and matches were found for 8,217 records (out of the original 11,396 unknown records). The first iteration is the most precise in terms of matching as many classification variables as possible. Unfortunately, not all unknowns could be matched and replaced with known values on this iteration so six more iterations in the hot-decking process were completed using one less classification variable at each iteration. The frequency distribution of SPEED LIMIT after each successive iteration of the hot-decking process is given in Table 1.

For the second iteration, only classification variables 1 through 7 were used and an additional 119 records were imputed. At the third iteration, classification variables 1 through 6 were used and an additional 1,162 records were imputed. At the fourth iteration, classification variables 1 through 5 were used and an additional 913 records were imputed. At the fifth iteration, the first four classification variables were used and an additional 643 records were imputed. At the sixth iteration, the first three classification variables were used and an additional 270 records were imputed. At the seventh iteration, the first three classification variables were used and an additional 61 records were imputed. This last iteration left only 11 records with unknown values and the univariate imputation procedure was used to eliminate them. At this last stage only SPEED LIMIT itself was used in the univariate process. The frequency distribution of SPEED LIMIT before and after imputation is presented in Appendix A.

TABLE 1: HOT-DECKING FOR SPEED LIMIT (1991)

Iteration	# of Class. Variables	Known		Unknown	
		Frequency	%	Frequency	%
Prior to Imputation	*	31,204	73.2	11,396	26.8
1	8	39,421	92.5	3,179	7.5
2	7	39,540	92.8	3,060	7.2
3	6	40,702	95.5	1898	4.5
4	5	41,615	97.7	985	2.3
5	4	42,258	99.2	342	0.8
6	3	42,528	99.8	72	0.2
7	2	42,589	100.0	11	0.03

Which GES Variables Have Imputed Values?

Some variables on each of the three GES files- crash, vehicle and person- have variables with imputed values. The variables chosen for imputation were the most commonly used variables or ones that appeared in the annual publication. Future plans for expanding the imputation process are discussed in the last section of this paper.

A total of 32 variables out of the 90 GES variables have imputed values and they can be placed into three categories:

- Hot-decked Imputed Variables (9 variables)
- Univariate Imputed Variables (19 variables)
- Variables Derived from the Imputed Variables (4 variables)

The last category are variables that are not directly imputed, but are calculated from imputed data. For example, the crash variable, ALCOHOL INVOLVED IN CRASH, is derived by reviewing all persons involved in the crash to determine whether anyone was using alcohol. To determine whether or not someone was using alcohol, the imputed person variable, ALCOHOL INVOLVEMENT, is used.

The hot-decked variables can be identified by a variable label ending in "_h." For example, the hot-decked imputed version of BODY_TYP (body type) is labelled, BDYTYP_H. The other imputed variables, both the univariate and derived, have labels ending in "_i." For example, imputed ATMOSPHERIC CONDITIONS is labelled WEATHR_I.

Tables 2, 3 and 4 show which variables have imputed values, which imputation process was used, and the original percentage of unknown values (from the 1991 file) for each of these variables. Appendix A shows the frequency distributions for the imputed variables before and after imputation.

TABLE 2: IMPUTED CRASH VARIABLES

	Variable	Imputation Type *	1991 % Unknown
1	Alcohol Involved in Crash	D	2.6
2	Atmospheric Conditions	U	2.3
3	Day of Week	U	0.0
4	First Harmful Event	U	0.4
5	Hour	U	0.9
6	Light Condition	U	2.4
7	Manner of Collision	U	0.4
8	Maximum Injury Severity	D	2.1
9	Minute	U	0.9
10	Relation to Junction	U	0.7
11	Roadway Alignment	U	2.7
12	Roadway Profile	U	30.5
13	Roadway Surface Condition	U	1.7
14	Speed Limit	H	26.8
15	Traffic Control Device	U	2.3

* D = Derived From Imputed Data
H = Hot-deck Imputation
U = Univariate Imputation

TABLE 3: IMPUTED VEHICLE VARIABLES

	Vehicle Variables	Imputation Type *	1991 % Unknown
1	Body Type	H	2.4
2	Driver Alcohol Involvement	D	6.0
3	Driver Violations Charged	U	0.3
4	Hit and Run	U	0.0
5	Initial Point of Impact	H	5.3
6	Maximum Injury Severity	D	3.8
7	Model Year	U	4.1
8	Most Harmful Event	H	5.1
9	Vehicle Maneuver	U	1.1
10	Vehicle Role	U	0.6

* D = Derived From Imputed Data
H = Hot-deck Imputation
U = Univariate Imputation

TABLE 4: IMPUTED PERSON VARIABLES

	Person Variables	Imputation Type *	1991 % Unknown
1	Age	H	10.5
2	Alcohol Involvement	H	3.4
3	Ejection	U	0.1
4	Injury Severity	H	3.4
5	Non-Motorist Location	U	0.0
6	Seating Position	H	2.9
7	Sex	H	8.4

* H = Hot-deck Imputation
U = Univariate Imputation

The hot-decked variables and their classification variables are listed below. The classification variables can be taken from any of the three GES files and are not necessarily on the file where the imputation is being done. The file location of each classification variable is given in parentheses.

	Crash File
Hot-decked Variable	Classification Variables
SPEED LIMIT	Stratum (crash) Land Use (crash) Roadway Surface Condition (crash) Traffic Control Device (crash) Number of Travel Lanes (crash) Relation to Junction (crash) Trafficway Flow (crash) Interstate Highway (crash)

Vehicle File

Hot-decked Variable	Classification Variables
BODY TYPE	Stratum (vehicle) Vehicle Model (vehicle) Number of Occupants (vehicle) Jackknife (vehicle) Vehicle Trailing (vehicle) Special Use (vehicle) Make (vehicle)
INITIAL POINT OF IMPACT	Stratum (vehicle) First Harmful Event (crash) Most Harmful Event (vehicle) Manner of Collision (crash) Vehicle Maneuver (vehicle) Number of Vehicles (crash) Accident Type (vehicle)
MOST HARMFUL EVENT	Stratum (vehicle) Number of Non-motorists (crash) Number of Vehicles (crash) Accident Type (vehicle) Vehicle Role (vehicle) First Harmful Event (crash)

Person File

AGE	Stratum (person) Restraint System Use (person) Number of Occupants (vehicle) Number of Vehicles (crash) Alcohol Involvement (person) Drug Involvement (person) Sex (person) Seating Position (person) Person Type (person) Violations Charged (person) Day of Week (crash) Hour (crash)

Hot-decked Variables	Classification Variables
SEX	Stratum (person) Restraint System Use (person) Number of Vehicles (crash) Alcohol Involvement (person) Drug Involvement (person) Age (person) Seating Position (person) Person Type (person) Violations Charged (person) Day of Week (crash) Hour (crash)
INJURY SEVERITY	Stratum (person) Accident Type (vehicle) Ejection (person) Taken to Hospital (person) Person Type (person) Restraint System Use (person) Manner of Leaving Scene (vehicle) Damage Severity (vehicle)
SEATING POSITION	Stratum (person) Restraint System Use (person) Body Type (Imputed, vehicle) Age (person) Person Type (person)
ALCOHOL INVOLVEMENT	Stratum (person) Body Type (Imputed, vehicle) Person Type (person) Violations Charged (person) Age (person) Sex (person) Day of Week (crash) Hour (crash)

No Hot-decking for Restraint Use

Another person level variable, RESTRAINT SYSTEM USE, was put through the hot-deck imputation process using the following classification variables: STRATUM, SEX, EJECTION, INJURY SEVERITY, PERSON TYPE, AGE, and SEATING POSITION. Restraint use prior to imputation was 63 percent, non-use was 21 percent, and unknown use was 16 percent. Although the imputation process worked properly, the meaning of the results were questioned. After imputation the use rate rose to more than 85 percent, a figure that simply was not realistic. The accuracy of police-reported restraint use has often been questioned, and if the original values for use were too high to begin with, imputation only exaggerates the problem. The decision was made to not impute for this variable and publish the unknown data as collected.

Both the hot-decking and univariate imputations were accomplished using SAS computer programs- a separate program for each of the three GES files. A copy of each of these programs can be found in Appendix B. The programs were written for the mainframe computing environment, but have been revised for a PC SAS computing environment. Testing has shown that the programs for all imputed variables currently take over 40 hours to process on a 486/33 PC.

Caution: When to Use Imputed Values and When Not To

Imputed data should be treated carefully, as Kalton states, "... researchers may falsely treat the completed data set for n respondents as if it were from a straightforward sample of size n.... [Imputation] does not increase the amount of information available, and indeed the use of a random subsampling of donors serves to reduce, not to increase, the effective sample size."

For every variable with imputed values on the GES file, the original variable was also left on the file. For example, both the original variable BODY_TYP (body type) and the imputed variable BDYTYP_H exist on the vehicle file. The analyst can choose when to use imputed values and when not to.

Although an analyst may be tempted to use the imputed variables for all types of analyses, it is not recommended. Imputed variables are recommended for overall size assessment questions: How many passenger cars versus light trucks are involved in side-impact crashes? How many pedestrians are injured in motor vehicle crashes at night?

More detailed analyses using imputed variables may lead to erroneous conclusions. Even though an attempt has been made to substitute the best known value possible in the imputation process, these new values have not been thoroughly edited. The original variables, prior to imputation, are put through a more rigorous editing and review process than the imputed variables. For example, a vehicle with an unknown vehicle maneuver may have an imputed vehicle maneuver of "stopped in traffic." However, this same vehicle may have a non-zero travel speed since travel speed was not used in the imputation process for this variable. Increasing the quality of the imputed data is one of the future considerations discussed in the next section.

Future Considerations

Improving the GES imputation process is a continuing process, and many issues still are being addressed:

- *Expanding hot-decking to other variables*

  For example, the variables that are currently imputed on a univariate basis could be added to the hot-decking procedure. In addition, as the GES annual report changes and new tables are added, new variables will need imputed values.

- *Using different or altered classification variables*

  Values within a classification variable could be grouped prior to imputation. Instead of using every possible value for age, for example, ages could be grouped to define the cells.

- *Additional data editing after imputation*

  Currently when GES data are entered and after coding is completed, the GES file is put through a rigorous editing process. This year we tried putting the imputed data through this same editing process and ended up with more failed edits than could be reasonably reviewed. The reason is that hot-decking does not take into account every possible relationship with other variables on the file, while editing tries to accomplish this task. More work needs to be done in this quality control area.

- *Effect of imputation on variances*

  In general, imputed data tend to have smaller variances than unimputed data, because unknown data are made similar to known data. Hot-decking effectively reduces the sample size because data from a "good" sample unit is used in place of data from "bad" sample unit. However, from preliminary testing of the effect of imputation on variances, very little effect on our generalized variance models has been discovered. This area, though, needs to be explored further.

- *Measuring bias increase or decrease*

  This task, although important, is easier said than done. One suggestion was made to compare data from GES cases with data from the same cases in the Fatal Accident Reporting System or the NASS Crashworthiness Data System. Unfortunately, the overlap of GES and these other systems is slight and there is no guarantee that the missing GES data will not be missing in these other systems. However, this issue is worth exploring.

References

1/ National Highway Traffic Safety Administration, National Center for Statistics and Analysis, *National Accident Sampling System General Estimates System Technical Note, 1988 to 1990*, (December, 1991), DOT HS 807 796.

2/ Kalton, Graham, *Compensating for Missing Survey Data*, Research Report Series, Institute for Social Research, The University of Michigan, (1983).

3/ Sande, Innis G., "A Personal View of Hot-Deck Imputation," Survey Methodology (1979).

4/ National Highway Traffic Safety Administration, National Center for Statistics and Analysis, *National Accident Sampling System 1986: A report on traffic crashes and injuries in the United States*, (July, 1988), DOT HS 807 296.

APPENDIX A

CRASH FREQUENCIES

HOUR OF CRASH				IMPUTED HOURS			
HOUR	Frequency	Percent	Cumulative	HOUR_I	Frequency	Percent	Cumulative
0	957	2.2	957	0	965	2.3	965
1	894	2.1	1851	1	901	2.1	1866
2	894	2.1	2745	2	901	2.1	2767
3	563	1.3	3308	3	567	1.3	3334
4	407	1.0	3715	4	410	1.0	3744
5	533	1.3	4248	5	537	1.3	4281
6	946	2.2	5194	6	954	2.2	5235
7	2125	5.0	7319	7	2143	5.0	7378
8	1907	4.5	9226	8	1925	4.5	9303
9	1637	3.8	10863	9	1651	3.9	10954
10	1800	4.2	12663	10	1816	4.3	12770
11	2130	5.0	14793	11	2149	5.0	14919
12	2428	5.7	17221	12	2451	5.8	17370
13	2438	5.7	19659	13	2460	5.8	19830
14	2775	6.5	22434	14	2800	6.6	22630
15	3259	7.7	25693	15	3289	7.7	25919
16	3249	7.6	28942	16	3280	7.7	29199
17	3347	7.9	32289	17	3378	7.9	32577
18	2525	5.9	34814	18	2548	6.0	35125
19	1845	4.3	36659	19	1862	4.4	36987
20	1581	3.7	38240	20	1595	3.7	38582
21	1468	3.4	39708	21	1481	3.5	40063
22	1312	3.1	41020	22	1323	3.1	41386
23	1169	2.7	42189	23	1179	2.8	42565
24	35	0.1	42224	24	35	0.1	42600
99	376	0.9	42600				

MINUTE OF CRASH				IMPUTED MINUTE			
MINUTE	Frequency	Percent	Cumulative	MINUTE_I	Frequency	Percent	Cumulative
0	4518	10.6	4518	0	4560	10.7	4560
1	274	0.6	4792	1	276	0.6	4836
2	303	0.7	5095	2	306	0.7	5142
3	275	0.6	5370	3	277	0.7	5419
4	250	0.6	5620	4	252	0.6	5671
5	1537	3.6	7157	5	1551	3.6	7222
6	219	0.5	7376	6	220	0.5	7442
7	277	0.7	7653	7	280	0.7	7722
8	309	0.7	7962	8	312	0.7	8034
9	240	0.6	8202	9	242	0.6	8276
10	2130	5.0	10332	10	2149	5.0	10425
11	255	0.6	10587	11	257	0.6	10682
12	330	0.8	10917	12	333	0.8	11015
13	274	0.6	11191	13	276	0.6	11291
14	291	0.7	11482	14	294	0.7	11585

CRASH FREQUENCIES

MINUTE	Frequency	Percent	Cumulative	MINUTE_1	Frequency	Percent	Cumulative
15	2935	6.9	14417	15	2961	7.0	14546
16	257	0.6	14674	16	259	0.6	14805
17	256	0.6	14930	17	259	0.6	15064
18	292	0.7	15222	18	294	0.7	15358
19	228	0.5	15450	19	230	0.5	15588
20	2271	5.3	17721	20	2291	5.4	17879
21	234	0.5	17955	21	236	0.6	18115
22	251	0.6	18206	22	253	0.6	18368
23	282	0.7	18488	23	285	0.7	18653
24	250	0.6	18738	24	252	0.6	18905
25	1538	3.6	20276	25	1551	3.6	20456
26	258	0.6	20534	26	260	0.6	20716
27	262	0.6	20796	27	265	0.6	20981
28	277	0.7	21073	28	279	0.7	21260
29	216	0.5	21289	29	218	0.5	21478
30	4274	10.0	25563	30	4315	10.1	25793
31	214	0.5	25777	31	215	0.5	26008
32	245	0.6	26022	32	248	0.6	26256
33	264	0.6	26286	33	266	0.6	26522
34	252	0.6	26538	34	254	0.6	26776
35	1524	3.6	28062	35	1538	3.6	28314
36	244	0.6	28306	36	246	0.6	28560
37	270	0.6	28576	37	272	0.6	28832
38	311	0.7	28887	38	314	0.7	29146
39	233	0.5	29120	39	235	0.6	29381
40	2147	5.0	31267	40	2166	5.1	31547
41	225	0.5	31492	41	227	0.5	31774
42	255	0.6	31747	42	257	0.6	32031
43	288	0.7	32035	43	291	0.7	32322
44	252	0.6	32287	44	254	0.6	32576
45	2873	6.7	35160	45	2898	6.8	35474
46	227	0.5	35387	46	230	0.5	35704
47	264	0.6	35651	47	266	0.6	35970
48	306	0.7	35957	48	308	0.7	36278
49	222	0.5	36179	49	224	0.5	36502
50	2449	5.7	38628	50	2471	5.8	38973
51	214	0.5	38842	51	215	0.5	39188
52	291	0.7	39133	52	294	0.7	39482
53	248	0.6	39381	53	250	0.6	39732
54	268	0.6	39649	54	271	0.6	40003
55	1526	3.6	41175	55	1539	3.6	41542
56	255	0.6	41430	56	257	0.6	41799
57	282	0.7	41712	57	284	0.7	42083
58	288	0.7	42000	58	291	0.7	42374
59	224	0.5	42224	59	226	0.5	42600
99	376	0.9	42600				

CRASH FREQUENCIES

RELATION TO JUNCTION

REL_JCT	Frequency	Percent	Cumulative
0	19258	45.2	19258
1	12575	29.5	31833
2	4906	11.5	36739
3	356	0.8	37095
4	4220	9.9	41315
5	506	1.2	41821
6	83	0.2	41904
8	389	0.9	42293
9	307	0.7	42600

IMPUTED RELATION TO JUNCTION

RELJCT_I	Frequency	Percent	Cumulative
0	19390	45.5	19390
1	12666	29.7	32056
2	4947	11.6	37003
3	358	0.8	37361
4	4254	10.0	41615
5	509	1.2	42124
6	84	0.2	42208
8	392	0.9	42600

ROADWAY ALIGNMENT

ALIGN	Frequency	Percent	Cumulative
1	36691	86.1	36691
2	4761	11.2	41452
9	1148	2.7	42600

IMPUTED ROADWAY ALIGNMENT

ALIGN_I	Frequency	Percent	Cumulative
1	37720	88.5	37720
2	4880	11.5	42600

ROADWAY PROFILE

PROFILE	Frequency	Percent	Cumulative
1	21698	50.9	21698
2	7149	16.8	28847
3	633	1.5	29480
8	128	0.3	29608
9	12992	30.5	42600

IMPUTED ROADWAY PROFILE

PROFIL_I	Frequency	Percent	Cumulative
1	31363	73.6	31363
2	10151	23.8	41514
3	903	2.1	42417
8	183	0.4	42600

ROADWAY SURFACE CONDITION

SUR_COND	Frequency	Percent	Cumulative
1	30963	72.7	30963
2	8344	19.6	39307
3	314	0.7	39621
4	2034	4.8	41655
5	63	0.1	41718
8	173	0.4	41891
9	709	1.7	42600

IMPUTED SURFACE CONDITION

SURCON_I	Frequency	Percent	Cumulative
1	31479	73.9	31479
2	8489	19.9	39968
3	320	0.8	40288
4	2072	4.9	42360
5	64	0.2	42424
8	176	0.4	42600

CRASH FREQUENCIES

TRAFFIC CONTROL DEVICE

TRAF_CON	Frequency	Percent	Cumulative
0	27454	64.4	27454
1	7029	16.5	34483
4	231	0.5	34714
8	171	0.4	34885
9	439	1.0	35324
21	4517	10.6	39841
22	367	0.9	40208
23	5	0.0	40213
28	251	0.6	40464
29	1	0.0	40465
40	39	0.1	40504
41	44	0.1	40548
42	37	0.1	40585
43	6	0.0	40591
49	117	0.3	40708
51	83	0.2	40791
61	44	0.1	40835
62	29	0.1	40864
97	415	1.0	41279
98	346	0.8	41625
99	975	2.3	42600

IMPUTED TRAFFIC CONTROL DEVICE

TRFCON_I	Frequency	Percent	Cumulative
0	28093	65.9	28093
1	7198	16.9	35291
4	236	0.6	35527
8	175	0.4	35702
9	449	1.1	36151
21	4627	10.9	40778
22	377	0.9	41155
23	5	0.0	41160
28	256	0.6	41416
29	1	0.0	41417
40	40	0.1	41457
41	45	0.1	41502
42	38	0.1	41540
43	6	0.0	41546
49	119	0.3	41665
51	85	0.2	41750
61	45	0.1	41795
62	30	0.1	41825
97	424	1.0	42249
98	351	0.8	42600

LIGHT CONDITIONS

LGHT_CON	Frequency	Percent	Cumulative
1	27906	65.5	27906
2	5343	12.5	33249
3	6784	15.9	40033
4	532	1.2	40565
5	993	2.3	41558
6	2	0.0	41560
9	1040	2.4	42600

IMPUTED LIGHT CONDITIONS

LGTCON_I	Frequency	Percent	Cumulative
1	28615	67.2	28615
2	5470	12.8	34085
3	6948	16.3	41033
4	544	1.3	41577
5	1021	2.4	42598
6	2	0.0	42600

ATMOSPHERIC CONDITIONS

WEATHER	Frequency	Percent	Cumulative
1	34044	79.9	34044
2	5672	13.3	39716
3	181	0.4	39897
4	1161	2.7	41058
5	266	0.6	41324
6	11	0.0	41335
7	3	0.0	41338
8	282	0.7	41620
9	980	2.3	42600

IMPUTED ATMOSPHERIC CONDITIONS

WEATHR_I	Frequency	Percent	Cumulative
1	34841	81.8	34841
2	5808	13.6	40649
3	184	0.4	40833
4	1192	2.8	42025
5	272	0.6	42297
6	11	0.0	42308
7	3	0.0	42311
8	289	0.7	42600

CRASH FREQUENCIES

MAXIMUM INJURY SEVERITY

MAX_SEV	Frequency	Percent	Cumulative
0	20621	48.4	20621
1	7298	17.1	27919
2	8648	20.3	36567
3	4480	10.5	41047
4	403	0.9	41450
5	186	0.4	41636
6	2	0.0	41638
8	56	0.1	41694
9	906	2.1	42600

IMPUTED MAXIMUM INJURY SEVERITY

MAX_SEV	Frequency	Percent	Cumulative
0	21228	49.8	21228
1	7466	17.5	28694
2	8709	20.4	37403
3	4535	10.6	41938
4	408	1.0	42346
5	191	0.4	42537
6	7	0.0	42544
8	56	0.1	42600

MANNER OF COLLISION

MAN_COL	Frequency	Percent	Cumulative
0	15720	36.9	15720
1	9206	21.6	24926
2	843	2.0	25769
3	4	0.0	25773
4	15204	35.7	40977
5	1132	2.7	42109
6	302	0.7	42411
8	10	0.0	42421
9	179	0.4	42600

IMPUTED MANNER OF COLLISION

MANCOL_I	Frequency	Percent	Cumulative
0	15778	37.0	15778
1	9249	21.7	25027
2	847	2.0	25874
3	4	0.0	25878
4	15271	35.8	41149
5	1137	2.7	42286
6	304	0.7	42590
8	10	0.0	42600

ALCOHOL INVOLVED IN CRASH

ALCOHOL	Frequency	Percent	Cumulative
1	3394	8.0	3394
2	38083	89.4	41477
9	1123	2.6	42600

ALCOHOL INVOLVED IN CRASH

ALCOHOL	Frequency	Percent	Cumulative
1	4041	9.5	4041
2	38559	90.5	42600

DAY OF WEEK

WEEKDAY	Frequency	Percent	Cumulative
1	4621	10.8	4621
2	6045	14.2	10666
3	6245	14.7	16911
4	5927	13.9	22838
5	6377	15.0	29215
6	7436	17.5	36651
7	5948	14.0	42599
9	1	0.0	42600

IMPUTED DAY OF WEEK

WKDY_I	Frequency	Percent	Cumulative
1	4621	10.8	4621
2	6045	14.2	10666
3	6245	14.7	16911
4	5928	13.9	22839
5	6377	15.0	29216
6	7436	17.5	36652
7	5948	14.0	42600

CRASH FREQUENCIES

FIRST HARMFUL EVENT				IMPUTED FIRST HARMFUL EVENT			
EVENT1	Frequency	Percent	Cumulative	EVENT1_I	Frequency	Percent	Cumulative
1	1421	3.3	1421	1	1431	3.4	1431
2	76	0.2	1497	2	77	0.2	1508
3	11	0.0	1508	3	11	0.0	1519
5	82	0.2	1590	5	82	0.2	1601
6	145	0.3	1735	6	146	0.3	1747
8	572	1.3	2307	8	577	1.4	2324
9	25	0.1	2332	9	26	0.1	2350
10	11	0.0	2343	10	11	0.0	2361
21	1610	3.8	3953	21	1617	3.8	3978
22	1320	3.1	5273	22	1326	3.1	5304
23	36	0.1	5309	23	37	0.1	5341
24	903	2.1	6212	24	919	2.2	6260
25	26842	63.0	33054	25	26842	63.0	33102
26	1997	4.7	35051	26	2032	4.8	35134
27	4	0.0	35055	27	4	0.0	35138
28	118	0.3	35173	28	119	0.3	35257
29	43	0.1	35216	29	44	0.1	35301
31	17	0.0	35233	31	17	0.0	35318
32	95	0.2	35328	32	96	0.2	35414
33	5	0.0	35333	33	5	0.0	35419
34	140	0.3	35473	34	141	0.3	35560
35	796	1.9	36269	35	804	1.9	36364
36	394	0.9	36663	36	397	0.9	36761
37	1597	3.7	38260	37	1615	3.8	38376
38	1022	2.4	39282	38	1032	2.4	39408
39	539	1.3	39821	39	545	1.3	39953
40	436	1.0	40257	40	440	1.0	40393
41	343	0.8	40600	41	347	0.8	40740
42	126	0.3	40726	42	127	0.3	40867
43	81	0.2	40807	43	82	0.2	40949
44	41	0.1	40848	44	42	0.1	40991
45	994	2.3	41842	45	1004	2.4	41995
46	70	0.2	41912	46	71	0.2	42066
48	458	1.1	42370	48	463	1.1	42529
49	70	0.2	42440	49	71	0.2	42600
99	160	0.4	42600				

CRASH FREQUENCIES

SPEED LIMIT				IMPUTED SPEED LIMIT			
SPD_LIM	Frequency	Percent	Cumulative	SPDLIM_H	Frequency	Percent	Cumulative
0	98	0.2	98	0	297	0.7	297
5	13	0.0	111	5	16	0.0	313
10	24	0.1	135	10	32	0.1	345
15	206	0.5	341	15	316	0.7	661
20	272	0.6	613	20	425	1.0	1086
25	4241	10.0	4854	25	6661	15.6	7747
30	3278	7.7	8132	30	4715	11.1	12462
35	6765	15.9	14897	35	9096	21.4	21558
40	2594	6.1	17491	40	3381	7.9	24939
45	4337	10.2	21828	45	5492	12.9	30431
50	1160	2.7	22988	50	1521	3.6	31952
55	7422	17.4	30410	55	9556	22.4	41508
60	23	0.1	30433	60	31	0.1	41539
65	771	1.8	31204	65	1061	2.5	42600
99	11396	26.8	42600				

VEHICLE FREQUENCIES

BODY TYPE				IMPUTED BODY TYPE			
BODY_TYP	Frequency	Percent	Cumulative	BDYTYP_H	Frequency	Percent	Cumulative
1	98	0.1	98	1	105	0.1	105
2	13071	17.7	13169	2	13217	17.9	13322
3	1632	2.2	14801	3	1634	2.2	14956
4	13827	18.7	28628	4	13975	18.9	28931
5	765	1	29393	5	768	1	29699
6	1954	2.6	31347	6	2004	2.7	31703
7	103	0.1	31450	7	105	0.1	31808
8	22	0	31472	8	23	0	31831
9	16968	23	48440	9	17746	24	49577
10	110	0.1	48550	10	110	0.1	49687
11	17	0	48567	11	17	0	49704
12	2	0	48569	12	2	0	49706
13	8	0	48577	13	8	0	49714
14	2379	3.2	50956	14	2561	3.5	52275
20	804	1.1	51760	20	806	1.1	53081
21	479	0.6	52239	21	482	0.7	53563
22	15	0	52254	22	17	0	53580
28	13	0	52267	28	14	0	53594
29	1927	2.6	54194	29	2399	3.2	55993
30	2466	3.3	56660	30	2476	3.4	58469
31	2149	2.9	58809	31	2156	2.9	60625
32	8	0	58817	32	8	0	60633
39	3871	5.2	62688	39	4732	6.4	65365
40	29	0	62717	40	40	0.1	65405
41	1	0	62718	41	1	0	65406
42	3	0	62721	42	4	0	65410
48	16	0	62737	48	21	0	65431
49	270	0.4	63007	50	223	0.3	65654
50	222	0.3	63229	58	310	0.4	65964
58	309	0.4	63538	59	19	0	65983
59	14	0	63552	60	722	1	66705
60	577	0.8	64129	63	39	0.1	66744
63	22	0	64151	65	3915	5.3	70659
65	3793	5.1	67944	68	1466	2	72125
68	1148	1.6	69092	70	1470	2	73595
69	1434	1.9	70526	71	95	0.1	73690
70	1343	1.8	71869	72	1	0	73691
71	74	0.1	71943	78	27	0	73718
72	1	0	71944	79	7	0	73725
78	20	0	71964	80	18	0	73743
79	6	0	71970	81	12	0	73755
80	15	0	71985	82	36	0	73791
81	4	0	71989	83	11	0	73802
82	35	0	72024	88	27	0	73829
83	11	0	72035	89	4	0	73833
88	19	0	72054				
89	2	0	72056				
99	1777	2.4	73833				

VEHICLE FREQUENCIES

DRIVER ALCOHOL INVOLVEMENT

VEH_ALCH	Frequency	Percent	Cumulative
1	3258	4.4	3258
2	66910	90.6	70168
9	3665	5	73833

IMPUTED DRIVER ALCOHOL

VEH_ALCH	Frequency	Percent	Cumulative
1	3916	5.3	3916
2	69917	94.7	73833

DRIVER VIOLATIONS CHARGED

VIOLATN	Frequency	Percent	Cumulative
0	51388	69.6	51388
1	1901	2.6	53289
2	2663	3.6	55952
3	80	0.1	56032
4	649	0.9	56681
5	384	0.5	57065
6	3420	4.6	60485
7	1181	1.6	61666
50	2848	3.9	64514
97	1784	2.4	66298
98	7301	9.9	73599
99	234	0.3	73833

IMPUTED DRIVER VIOLATIONS CHARGED

VLTN_I	Frequency	Percent	Cumulative
0	51555	69.8	51555
1	1906	2.6	53461
2	2671	3.6	56132
3	80	0.1	56212
4	651	0.9	56863
5	385	0.5	57248
6	3430	4.6	60678
7	1184	1.6	61862
50	2859	3.9	64721
97	1789	2.4	66510
98	7323	9.9	73833

HIT AND RUN

HIT_RUN	Frequency	Percent	Cumulative
0	69966	94.8	69966
1	3831	5.2	73797
9	36	0	73833

IMPUTED HIT AND RUN

HITRUN_I	Frequency	Percent	Cumulative
0	70000	94.8	70000
1	3833	5.2	73833

INITIAL POINT OF IMPACT

IMPACT	Frequency	Percent	Cumulative
0	3001	4.1	3001
1	29583	40.1	32584
2	11243	15.2	43827
3	11832	16	55659
4	12150	16.5	67809
5	102	0.1	67911
6	282	0.4	68193
7	1720	2.3	69913
9	3920	5.3	73833

IMPUTED INITIAL POINT OF IMPACT

IMPACT_I	Frequency	Percent	Cumulative
0	3001	4.1	3001
1	31447	42.6	34448
2	12144	16.4	46592
3	12577	17	59169
4	12339	16.7	71508
5	114	0.2	71622
6	332	0.4	71954
7	1879	2.5	73833

MAXIMUM INJURY IN VEHICLE

MAX_VSEV	Frequency	Percent	Cumulative
0	48473	65.7	48473
1	9743	13.2	58216
2	7941	10.8	66157
3	4023	5.4	70180
4	319	0.4	70499
5	215	0.3	70714
6	2	0	70716
8	290	0.4	71006
9	2827	3.8	73833

MAXIMUM INJURY IN VEHICLE

MAX_VSE_I	Frequency	Percent	Cumulative
0	50852	68.9	50852
1	9990	13.5	60842
2	8057	10.9	68899
3	4082	5.5	72981
4	324	0.4	73305
5	231	0.3	73536
6	7	0.0	73543
8	290	0.4	73833

VEHICLE FREQUENCIES

MODEL YEAR				IMPUTED MODEL YEAR			
MODEL_YR	Frequency	Percent	Cumulative	MDLYR_I	Frequency	Percent	Cumulative
40	4	0	4	40	4	0	4
42	1	0	5	42	1	0	5
44	2	0	7	44	2	0	7
45	5	0	12	45	5	0	12
46	2	0	14	46	2	0	14
47	2	0	16	47	3	0	17
48	4	0	20	48	4	0	21
50	6	0	26	50	6	0	27
51	4	0	30	51	4	0	31
52	4	0	34	52	4	0	35
53	4	0	38	53	4	0	39
54	2	0	40	54	2	0	41
55	10	0	50	55	11	0	52
56	1	0	51	56	1	0	53
57	2	0	53	57	2	0	55
58	9	0	62	58	9	0	64
59	8	0	70	59	8	0	72
60	11	0	81	60	12	0	84
61	9	0	90	61	9	0	93
62	22	0	112	62	23	0	116
63	41	0.1	153	63	43	0.1	159
64	58	0.1	211	64	60	0.1	219
65	103	0.1	314	65	108	0.1	327
66	122	0.2	436	66	127	0.2	454
67	152	0.2	588	67	158	0.2	612
68	200	0.3	788	68	208	0.3	820
69	232	0.3	1020	69	241	0.3	1061
70	278	0.4	1298	70	289	0.4	1350
71	342	0.5	1640	71	356	0.5	1706
72	515	0.7	2155	72	536	0.7	2242
73	538	0.7	2693	73	560	0.8	2802
74	677	0.9	3370	74	703	1	3505
75	691	0.9	4061	75	719	1	4224
76	1098	1.5	5159	76	1144	1.5	5368
77	1830	2.5	6989	77	1910	2.6	7278
78	2490	3.4	9479	78	2597	3.5	9875
79	3020	4.1	12499	79	3143	4.3	13018
80	2578	3.5	15077	80	2687	3.6	15705
81	2825	3.8	17902	81	2944	4	18649
82	2928	4	20830	82	3057	4.1	21706
83	3330	4.5	24160	83	3472	4.7	25178
84	5081	6.9	29241	84	5298	7.2	30476
85	5697	7.7	34938	85	5943	8	36419
86	6227	8.4	41165	86	6499	8.8	42918
87	6082	8.2	47247	87	6348	8.6	49266
88	6724	9.1	53971	88	7016	9.5	56282
89	6554	8.9	60525	89	6834	9.3	63116
90	5820	7.9	66345	90	6070	8.2	69186

VEHICLE FREQUENCIES

MODEL_YR	Frequency	Percent	Cumulative	MDLYR_I	Frequency	Percent	Cumulative
91	4224	5.7	70569	91	4405	6.0	73591
92	233	0.3	70802	92	242	0.3	73833
99	3031	4.1	73833				

	MOST HARMFUL EVENT				IMPUTED MOST HARMFUL EVENT		
V_EVENT	Frequency	Percent	Cumulative	V_EVNT_H	Frequency	Percent	Cumulative
1	1980	2.7	1980	1	2280	3.1	2280
2	112	0.2	2092	2	124	0.2	2404
3	15	0	2107	3	17	0	2421
4	4	0	2111	4	4	0	2425
5	58	0.1	2169	5	82	0.1	2507
6	143	0.2	2312	6	158	0.2	2665
8	373	0.5	2685	8	386	0.5	3051
10	29	0	2714	10	30	0	3081
21	1580	2.1	4294	21	1614	2.2	4695
22	1316	1.8	5610	22	1326	1.8	6021
23	36	0	5646	23	40	0.1	6061
24	889	1.2	6535	24	909	1.2	6970
25	56164	76.1	62699	25	57735	78.2	64705
26	1952	2.6	64651	26	2051	2.8	66756
27	4	0	64655	27	4	0	66760
28	330	0.4	64985	28	351	0.5	67111
29	36	0	65021	29	41	0.1	67152
31	19	0	65040	31	22	0	67174
32	92	0.1	65132	32	112	0.2	67286
33	5	0	65137	33	5	0	67291
34	118	0.2	65255	34	131	0.2	67422
35	590	0.8	65845	35	752	1	68174
36	274	0.4	66119	36	371	0.5	68545
37	1237	1.7	67356	37	1617	2.2	70162
38	645	0.9	68001	38	846	1.1	71008
39	181	0.2	68182	39	364	0.5	71372
40	226	0.3	68408	40	328	0.4	71700
41	221	0.3	68629	41	320	0.4	72020
42	100	0.1	68729	42	120	0.2	72140
43	66	0.1	68795	43	80	0.1	72220
44	19	0	68814	44	26	0	72246
45	853	1.2	69667	45	1029	1.4	73275
46	48	0.1	69715	46	58	0.1	73333
48	323	0.4	70038	48	432	0.6	73765
49	54	0.1	70092	49	68	0.1	73833
99	3741	5.1	73833				

VEHICLE FREQUENCIES

VEHICLE MANUEVER

MANUEVER	Frequency	Percent	Cumulative
1	39776	53.9	39776
2	3143	4.3	42919
3	1038	1.4	43957
4	8541	11.6	52498
5	724	1	53222
6	209	0.3	53431
7	19	0	53450
8	80	0.1	53530
9	1109	1.5	54639
10	2241	3	56880
11	6779	9.2	63659
12	274	0.4	63933
13	974	1.3	64907
14	2251	3	67158
15	826	1.1	67984
98	5026	6.8	73010
99	823	1.1	73833

IMPUTED VEHICLE MANUEVER

MANEUV_I	Frequency	Percent	Cumulative
1	40207	54.5	40207
2	3178	4.3	43385
3	1051	1.4	44436
4	8648	11.7	53084
5	732	1	53816
6	213	0.3	54029
7	19	0	54048
8	82	0.1	54130
9	1120	1.5	55250
10	2268	3.1	57518
11	6858	9.3	64376
12	277	0.4	64653
13	991	1.3	65644
14	2273	3.1	67917
15	834	1.1	68751
98	5082	6.9	73833

VEHICLE ROLE

VEH_ROLE	Frequency	Percent	Cumulative
0	2522	3.4	2522
1	40877	55.4	43399
2	26371	35.7	69770
3	3642	4.9	73412
9	421	0.6	73833

IMPUTED VEHICLE ROLE

VROLE_I	Frequency	Percent	Cumulative
0	2531	3.4	2531
1	41114	55.7	43645
2	26530	35.9	70175
3	3658	5	73833

PERSON FREQUENCIES

AGE	AGE Frequency	Percent	Cumulative	AGE_H	IMPUTED AGE Frequency	Percent	Cumulative
0	390	0.4	390	0	668	0.6	668
1	547	0.5	937	1	709	0.7	1377
2	646	0.6	1583	2	841	0.8	2218
3	634	0.6	2217	3	792	0.7	3010
4	607	0.6	2824	4	746	0.7	3756
5	606	0.6	3430	5	772	0.7	4528
6	557	0.5	3987	6	594	0.5	5122
7	545	0.5	4532	7	613	0.6	5735
8	529	0.5	5061	8	727	0.7	6462
9	526	0.5	5587	9	667	0.6	7129
10	579	0.5	6166	10	676	0.6	7805
11	507	0.5	6673	11	659	0.6	8464
12	554	0.5	7227	12	699	0.6	9163
13	640	0.6	7867	13	763	0.7	9926
14	772	0.7	8639	14	833	0.8	10759
15	1165	1.1	9804	15	1395	1.3	12154
16	2901	2.7	12705	16	3309	3.0	15463
17	3305	3.0	16010	17	3560	3.3	19023
18	3234	3.0	19244	18	3643	3.3	22666
19	3195	2.9	22439	19	3660	3.4	26326
20	3243	3.0	25682	20	3891	3.6	30217
21	3104	2.8	28786	21	3289	3.0	33506
22	2765	2.5	31551	22	2930	2.7	36436
23	2640	2.4	34191	23	3054	2.8	39490
24	2534	2.3	36725	24	2967	2.7	42457
25	2483	2.3	39208	25	2787	2.6	45244
26	2568	2.4	41776	26	2901	2.7	48145
27	2574	2.4	44350	27	2694	2.5	50839
28	2467	2.3	46817	28	2639	2.4	53478
29	2288	2.1	49105	29	2481	2.3	55959
30	2378	2.2	51483	30	3103	2.8	59062
31	2231	2.0	53714	31	2424	2.2	61486
32	2136	2.0	55850	32	2222	2.0	63708
33	2094	1.9	57944	33	2247	2.1	65955
34	1998	1.8	59942	34	2082	1.9	68037
35	2046	1.9	61988	35	2455	2.3	70492
36	1866	1.7	63854	36	1935	1.8	72427
37	1753	1.6	65607	37	1812	1.7	74239
38	1719	1.6	67326	38	1776	1.6	76015
39	1564	1.4	68890	39	1623	1.5	77638
40	1634	1.5	70524	40	1725	1.6	79363
41	1443	1.3	71967	41	1519	1.4	80882
42	1436	1.3	73403	42	1556	1.4	82438
43	1396	1.3	74799	43	1451	1.3	83889
44	1414	1.3	76213	44	1468	1.3	85357
45	1125	1.0	77338	45	1193	1.1	86550
46	1023	0.9	78361	46	1108	1.0	87658
47	1004	0.9	79365	47	1066	1.0	88724
48	1031	0.9	80396	48	1080	1.0	89804

PERSON FREQUENCIES

AGE	Frequency	Percent	Cumulative	AGE_H	Frequency	Percent	Cumulative
49	891	0.8	81287	49	1014	0.9	90818
50	908	0.8	82195	50	940	0.9	91758
51	788	0.7	82983	51	826	0.8	92584
52	777	0.7	83760	52	796	0.7	93380
53	773	0.7	84533	53	1618	1.5	94998
54	695	0.6	85228	54	743	0.7	95741
55	677	0.6	85905	55	723	0.7	96464
56	651	0.6	86556	56	753	0.7	97217
57	599	0.5	87155	57	620	0.6	97837
58	587	0.5	87742	58	606	0.6	98443
59	611	0.6	88353	59	642	0.6	99085
60	593	0.5	88946	60	683	0.6	99768
61	586	0.5	89532	61	609	0.6	100377
62	535	0.5	90067	62	559	0.5	100936
63	485	0.4	90552	63	492	0.5	101428
64	481	0.4	91033	64	500	0.5	101928
65	499	0.5	91532	65	513	0.5	102441
66	468	0.4	92000	66	543	0.5	102984
67	473	0.4	92473	67	485	0.4	103469
68	423	0.4	92896	68	436	0.4	103905
69	420	0.4	93316	69	573	0.5	104478
70	444	0.4	93760	70	509	0.5	104987
71	406	0.4	94166	71	415	0.4	105402
72	361	0.3	94527	72	372	0.3	105774
73	332	0.3	94859	73	337	0.3	106111
74	312	0.3	95171	74	368	0.3	106479
75	293	0.3	95464	75	360	0.3	106839
76	252	0.2	95716	76	266	0.2	107105
77	242	0.2	95958	77	265	0.2	107370
78	242	0.2	96200	78	246	0.2	107616
79	215	0.2	96415	79	224	0.2	107840
80	190	0.2	96605	80	195	0.2	108035
81	163	0.1	96768	81	165	0.2	108200
82	142	0.1	96910	82	184	0.2	108384
83	131	0.1	97041	83	133	0.1	108517
84	84	0.1	97125	84	84	0.1	108601
85	90	0.1	97215	85	93	0.1	108694
86	76	0.1	97291	86	81	0.1	108775
87	39	0.0	97330	87	39	0.0	108814
88	60	0.1	97390	88	63	0.1	108877
89	37	0.0	97427	89	37	0.0	108914
90	14	0.0	97441	90	14	0.0	108928
91	14	0.0	97455	91	14	0.0	108942
92	4	0.0	97459	92	4	0.0	108946
93	5	0.0	97464	93	5	0.0	108951
94	2	0.0	97466	94	2	0.0	108953
95	1	0.0	97467	95	1	0.0	108954
97	1	0.0	97468	97	1	0.0	108955
99	11487	10.5	108955				

PERSON FREQUENCIES

ALCOHOL INVOLVEMENT

PER_ALCH	Frequency	Percent	Cumulative
0	101780	93.4	101780
1	3134	2.9	104914
7	304	0.3	105218
8	8	0.0	105226
9	3729	3.4	108955

IMPUTED ALCOHOL INVOLVMENT

ALCH_H	Frequency	Percent	Cumulative
0	104714	96.1	104714
1	3912	3.6	108626
7	321	0.3	108947
8	8	0.0	108955

EJECTION

EJECT	Frequency	Percent	Cumulative
0	108596	99.7	108596
1	265	0.2	108861
9	94	0.1	108955

IMPUTED EJECTION

EJECT_I	Frequency	Percent	Cumulative
0	108690	99.8	108690
1	265	0.2	108955

INJURY SEVERITY

INJ_SEV	Frequency	Percent	Cumulative
0	73332	67.3	73332
1	13791	12.7	87123
2	11639	10.7	98762
3	5753	5.3	104515
4	447	0.4	104962
5	325	0.3	105287
6	5	0.0	105292
9	3663	3.4	108955

IMPUTED INJURY SEVERITY

INJSEV_H	Frequency	Percent	Cumulative
0	76359	70.1	76359
1	14132	13.0	90491
2	11821	10.8	102312
3	5833	5.4	108145
4	456	0.4	108601
5	346	0.3	108947
6	8	0.0	108955

NON-MOTORIST LOCATION

LOCATN	Frequency	Percent	Cumulative
0	105580	96.9	105580
1	252	0.2	105832
2	1236	1.1	107068
8	19	0.0	107087
9	2	0.0	107089
11	17	0.0	107106
12	1395	1.3	108501
18	216	0.2	108717
19	8	0.0	108725
98	207	0.2	108932
99	23	0.0	108955

IMPUTED LOCATION

LOCATN_I	Frequency	Percent	Cumulative
0	105603	96.9	105603
1	252	0.2	105855
2	1236	1.1	107091
8	19	0.0	107110
9	2	0.0	107112
11	17	0.0	107129
12	1395	1.3	108524
18	216	0.2	108740
19	8	0.0	108748
98	207	0.2	108955

SEATING POSITION

SEAT_POS	Frequency	Percent	Cumulative
0	3375	3.1	3375
11	73495	67.5	76870
12	1186	1.1	78056
13	17327	15.9	95383
18	73	0.1	95456
19	32	0.0	95488
21	3436	3.2	98924

IMPUTED SEATING POSITION

SEAT_H	Frequency	Percent	Cumulative
0	3375	3.1	3375
11	73843	67.8	77218
12	1294	1.2	78512
13	19344	17.8	97856
21	3834	3.5	101690
22	1589	1.5	103279
23	4467	4.1	107746

PERSON FREQUENCIES

SEAT_POS	Frequency	Percent	Cumulative
22	1418	1.3	100342
23	4092	3.8	104434
28	72	0.1	104506
29	113	0.1	104619
99	3185	2.9	108955

SEAT_H	Frequency	Percent	Cumulative
30	18	0.0	107764
40	1007	0.9	108771
50	1	0.0	108772
60	183	0.2	108955

SEX

SEX	Frequency	Percent	Cumulative
1	60064	55.1	60064
2	39736	36.5	99800
9	9155	8.4	108955

IMPUTED SEX

SEX_H	Frequency	Percent	Cumulative
1	65688	60.3	65688
2	43267	39.7	108955

APPENDIX B

SAS PROGRAMS FOR GES IMPUTATION

```
**   UNIVARIATE AND HOTDECK IMPUTATION FOR 1991 ACCIDENT FILE **;
*************************************************************************;

************************************************************;
OPTION NONUMBER NODATE PAGESIZE=60 LINESIZE=130;
************************************************************;
libname  ges91    'c:\ges\ges91' ;
libname  LIBRARY    'c:\ges\ges91' ;
************************************************************;
PROC FREQ DATA=GES91.ACC91;
TABLES HOUR MINUTE REL_JCT ALIGN PROFILE SUR_COND TRAF_CON
LGHT_CON WEATHER MAX_SEV MAN_COL ALCOHOL WEEKDAY EVENT1 SPD_LIM;
************************************************************;

************************************************************;
*     UNIVARIATE IMPUTATION  - ACCIDENT FILE               *;
*                USING WEIGHT VARIABLE                     *;
*   THIS PROGRAM CALCULATES THE FREQUENCY DISTRIBUTION     *;
*   FOR THE KNOWN VALUES OF THE VARIABLES TO BE IMPUTED    *;
*   AND APPLIES THESE NUMBERS TO THE RECORDS WITH          *;
*   UNKNOWN VALUES, GENERATING A CUMULATIVE FREQUENCY      *;
*   SCALE FOR SUCH RECORDS IN THE ACCIDENT FILE.  THE      *;
*   RESULTS ARE THEN USED TO IMPUTE VALUES FOR THE         *;
*   RECORDS HAVING MISSING VALUES FOR EACH VARIABLE        *;
************************************************************;

*************************************************************************;
*                      VARIABLES IMPUTED                                ;
*         NAME                 SAS NAME     NEW SAS NAME               ;
*    ------------------        --------    -------------               ;
*         HOUR                   HOUR        HOUR_I                    ;
*         MINUTES                MINUTE      MINUTE_I                  ;
*         RELATION TO JUNCTION   REL_JCT     RELJCT_I                  ;
*         ALIGNMENT              ALIGN       ALIGN_I                   ;
*         PROFILE                PROFILE     PROFIL_I                  ;
*         SURFACE CONDITION      SUR_COND    SURCON_I                  ;
*         TRAF CONTROL DEV.      TRAF_CON    TRFCON_I                  ;
*         LIGHT CONDITION        LGHT_CON    LGTCON_I                  ;
*         WEATHER                WEATHER     WEATHR_I                  ;
*         MANNER OF COLLISION    MAN_COL     MANCOL_I                  ;
*         WEEKDAY                WEEKDAY     WKDY_I                    ;
*         FIRST HARMFUL EVENT    EVENT1      EVENT1_I                  ;
*************************************************************************;
*                                                                       ;
**** (1)  UNIVARIATE IMPUTATION OF HOUR, MINUTE, REL_JCT, ALIGN,       ;
*         PROFILE, SUR_COND, AND TRAF_CON                             ;
*                                                                       ;
*************************************************************************;

DATA ONE (KEEP=CASENUM HOUR MINUTE REL_JCT ALIGN PROFILE
              SUR_COND TRAF_CON WEIGHT DUMMY ); SET GES91.ACC91;
DUMMY=1;
```

```
DATA TWO; SET ONE;     IF HOUR NE 99;
PROC FREQ; TABLES HOUR/ OUT=TAB NOPRINT; WEIGHT WEIGHT;

DATA THREE; SET ONE; IF HOUR EQ 99;
PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;

DATA TABLE; IF _N_=1 THEN SET NCOUNT;      SET TAB;
  IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
       CUM_PER=PERCENT+CUM_PER;
      CUM_NO = CUM_PER*NO_UNKN*.01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

*  NOW STORE CUM SCALE INTO ONE RECORD AND MERGE WITH ORIG DATA;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO25;
          RETAIN CUM_NO1-CUM_NO25 I; DROP I;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=12345; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000     DATA=ONE ; BY RAN_NO;

   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_HR K; IF _N_=1 THEN DO; CT_HR=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO25; DROP K CT_HR CUM_NO1-CUM_NO25;

   IF HOUR EQ 99 THEN DO;
      CT_HR=CT_HR +1;

   STOPPER=' NO';
   DO I=K TO 25 WHILE(STOPPER=' NO');
       J=I-1;
      IF CT_HR LE CUM THEN DO;
      HOUR_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
END;
ELSE HOUR_I=HOUR;

DATA ONE (KEEP=CASENUM HOUR_I MINUTE REL_JCT ALIGN PROFILE
               SUR_COND TRAF_CON WEIGHT DUMMY );    SET ONE;
*****************************************************************;
* NEXT STEP -- PROCEED TO NEXT VARIABLE: MINUTES;

DATA TWO ; SET ONE; IF MINUTE NE 99;
PROC FREQ;TABLES MINUTE/OUT=TAB NOPRINT; WEIGHT WEIGHT;
DATA THREE; SET ONE; IF MINUTE EQ 99;
   PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
 IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
 CUM_PER= PERCENT + CUM_PER;
 CUM_NO=CUM_PER * NO_UNKN*.01; CUM_NO=ROUND(CUM_NO);
```

```
DROP _TYPE_ _FREQ_;

*  NOW STORE CUM SCALE INTO ONE RECORD AND MERGE WITH ORIG DATA;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO60;
         RETAIN CUM_NO1-CUM_NO60 I; DROP I;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=23457; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_MIN K; IF _N_=1 THEN DO; CT_MIN=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO60; DROP K CT_MIN CUM_NO1-CUM_NO60;

   IF MINUTE EQ 99 THEN DO;
      CT_MIN=CT_MIN +1;

   STOPPER=' NO';
   DO I=K TO 60 WHILE(STOPPER=' NO');
       J=I-1;
      IF CT_MIN LE CUM THEN DO;
     MINUTE_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
END;
ELSE MINUTE_I=MINUTE;

DATA ONE (KEEP=CASENUM HOUR_I MINUTE_I REL_JCT ALIGN PROFILE
               SUR_COND TRAF_CON WEIGHT DUMMY );   SET ONE;

*****************************************************************;
* NEXT STEP -- PROCEED TO NEXT VARIABLE: REL_JCT;

DATA TWO; SET ONE; IF REL_JCT NE 9;
PROC FREQ; TABLES REL_JCT/OUT=TAB NOPRINT; WEIGHT WEIGHT;
DATA THREE; SET ONE; IF REL_JCT EQ 9;
PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0;
RETAIN CUM_PER;
CUM_PER=PERCENT +CUM_PER;
CUM_NO=CUM_PER * NO_UNKN*.01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

*  NOW STORE CUM SCALE INTO ONE RECORD AND MERGE WITH ORIG DATA;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO8;
         RETAIN CUM_NO1-CUM_NO8 I; DROP I;
   IF _N_=1 THEN I=0;
```

```
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=34567; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;
   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_REL K; IF _N_=1 THEN DO; CT_REL=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO8; DROP K CT_REL CUM_NO1-CUM_NO8;

   IF REL_JCT EQ 9 THEN DO;
      CT_REL=CT_REL +1;

    STOPPER=' NO';
   DO I=K TO 8 WHILE(STOPPER=' NO');
       J=I-1;
       IF I EQ 8 THEN J=8;
      IF CT_REL LE CUM THEN DO;
      RELJCT_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
END;

ELSE RELJCT_I=REL_JCT;

DATA ONE (KEEP=CASENUM HOUR_I MINUTE_I RELJCT_I ALIGN PROFILE
               SUR_COND TRAF_CON WEIGHT DUMMY );   SET ONE;
*****************************************************************;

**** NEXT STEP -- PROCEED TO NEXT VARIABLE: ALIGN;

DATA TWO; SET ONE;  IF ALIGN NE 9;
PROC FREQ; TABLES ALIGN/OUT=TAB NOPRINT; WEIGHT WEIGHT;
 DATA THREE; SET ONE; IF ALIGN EQ 9;
   PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0;
RETAIN CUM_PER;
CUM_PER=PERCENT + CUM_PER;
CUM_NO= CUM_PER * NO_UNKN * .01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

*  NOW STORE CUM SCALE INTO ONE RECORD AND MERGE WITH ORIG DATA;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO2;
          RETAIN CUM_NO1-CUM_NO2 I; DROP I;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=45679; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;
```

```
   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_ALN K; IF _N_=1 THEN DO; CT_ALN=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO2; DROP K CT_ALN CUM_NO1-CUM_NO2;

   IF ALIGN EQ 9 THEN DO;
      CT_ALN=CT_ALN +1;

   STOPPER=' NO';
   DO I=K TO 2 WHILE(STOPPER=' NO');
      IF CT_ALN LE CUM THEN DO;
      J=I;
     ALIGN_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
END;

ELSE ALIGN_I=ALIGN;

DATA ONE (KEEP=CASENUM HOUR_I MINUTE_I RELJCT_I ALIGN_I PROFILE
               SUR_COND TRAF_CON WEIGHT DUMMY );   SET ONE;

*****************************************************************;
* NEXT STEP -- PROCEED TO NEXT VARIABLE: PROFILE;

DATA TWO; SET ONE;
 IF PROFILE NE 9; PROC FREQ; TABLES PROFILE/OUT=TAB NOPRINT; WEIGHT WEIGHT;
DATA THREE; SET ONE; IF PROFILE EQ 9;
PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
CUM_PER=PERCENT + CUM_PER;
CUM_NO=CUM_PER * NO_UNKN * .01 ; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;


   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO4;
         RETAIN CUM_NO1-CUM_NO4 I; DROP I;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
    SEED=56789; RAN_NO=RANUNI(SEED);
    PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_PR K; IF _N_=1 THEN DO; CT_PR=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO4; DROP K CT_PR CUM_NO1-CUM_NO4;

   IF PROFILE EQ 9 THEN DO;
      CT_PR=CT_PR +1;
```

```
   STOPPER=' NO';
   DO I=K TO 4 WHILE(STOPPER=' NO');
       J=I;
      IF I=4 THEN J=8;
      IF CT_PR LE CUM THEN DO;
     PROFIL_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
END;

ELSE PROFIL_I=PROFILE;

DATA ONE (KEEP=CASENUM HOUR_I MINUTE_I RELJCT_I ALIGN_I PROFIL_I
               SUR_COND TRAF_CON WEIGHT DUMMY );    SET ONE;

*****************************************************************;
* NEXT STEP -- PROCEED TO NEXT VARIABLE: SURFACE CONDITION;

DATA TWO; SET ONE; IF SUR_COND NE 9;
PROC FREQ; TABLES SUR_COND/OUT=TAB NOPRINT; WEIGHT WEIGHT;
DATA THREE; SET ONE; IF SUR_COND=9;
    PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
 IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
   CUM_PER=PERCENT + CUM_PER;
    CUM_NO=CUM_PER * NO_UNKN * .01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO6;
          RETAIN CUM_NO1-CUM_NO6 I; DROP I;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=67899; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_SC K; IF _N_=1 THEN DO; CT_SC=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO6; DROP K CT_SC CUM_NO1-CUM_NO6;

   IF SUR_COND EQ 9 THEN DO;
      CT_SC=CT_SC +1;

   STOPPER=' NO';
   DO I=K TO 6 WHILE(STOPPER=' NO');
       J=I; IF I=6 THEN J=8;
      IF CT_SC LE CUM THEN DO;
     SURCON_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
```

```
END;

ELSE SURCON_I=SUR_COND;
DATA ONE (KEEP=CASENUM HOUR_I MINUTE_I RELJCT_I ALIGN_I PROFIL_I
               SURCON_I TRAF_CON WEIGHT DUMMY );   SET ONE;

*****************************************************************;
* NEXT STEP -- PROCEED TO NEXT VARIABLE: TRAF CONTRL DVC'S;

DATA TWO; SET ONE; IF TRAF_CON NE 99;
PROC FREQ; TABLES TRAF_CON/OUT=TAB NOPRINT; WEIGHT WEIGHT;
DATA THREE; SET ONE; IF TRAF_CON EQ 99;
     PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
CUM_PER= PERCENT + CUM_PER;
CUM_NO= CUM_PER * NO_UNKN * .01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO20;
          RETAIN V1-V20  CUM_NO1-CUM_NO20 I; DROP I;
          ARRAY VALUE (I) V1-V20;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   VALUE=TRAF_CON;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=78999; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_TC K; IF _N_=1 THEN DO; CT_TC=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO20; DROP K CT_TC CUM_NO1-CUM_NO20;
   ARRAY VALUE (I) V1-V20; DROP V1-V20;

   IF TRAF_CON EQ 99 THEN DO;
      CT_TC=CT_TC +1;

   STOPPER=' NO';
   DO I=K TO 20 WHILE(STOPPER=' NO');
      IF CT_TC LE CUM THEN DO;
     TRFCON_I=VALUE;
      STOPPER='YES';
      K=I;
      END;
  END;
END;

ELSE TRFCON_I=TRAF_CON;

LABEL
  HOUR_I='IMPUTED HOURS'
  MINUTE_I='IMPUTED MINUTE'
  RELJCT_I='IMPUTED RELATION TO JUNCTION'
  ALIGN_I = 'IMPUTED ALIGN'
```

```
  PROFIL_I='IMPUTED PROFILE'
  SURCON_I='IMPUTED SURFACE CONDITION'
  TRFCON_I='IMPUTED TRAFFIC CONTRL DEVICE';

LENGTH HOUR_I MINUTE_I RELJCT_I ALIGN_I PROFIL_I SURCON_I TRFCON_I 3 ;

KEEP CASENUM HOUR_I MINUTE_I RELJCT_I ALIGN_I PROFIL_I SURCON_I TRFCON_I ;

DATA ACC91_I; SET ONE;

*********************************************************************;
*                                                                   ;
****  (2)  UNIVARIATE IMPUTATION OF LGHT_CON, WEATHER, MAX_SEV,     ;
*          MAN_COL, ALCOHOL, WEEKDAY AND EVENT1                     ;
*                                                                   ;
*********************************************************************;

*********************************************************************;

DATA ONE (KEEP=CASENUM LGHT_CON WEATHER MAX_SEV MAN_COL ALCOHOL WEEKDAY
                EVENT1 WEIGHT DUMMY ); SET GES91.ACC91;
DUMMY=1;
*********************************************************************;

* NEXT STEP PROCEED TO NEXT VARIABLE: LGHT CONDITIONS;

 DATA TWO; SET ONE; IF LGHT_CON NE 9;
 PROC FREQ; TABLES LGHT_CON/OUT=TAB NOPRINT; WEIGHT WEIGHT;
 DATA THREE; SET ONE; IF LGHT_CON EQ 9;
PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
 DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
CUM_PER= PERCENT + CUM_PER;
CUM_NO= CUM_PER *NO_UNKN * .01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO6;
         RETAIN CUM_NO1-CUM_NO6 I; DROP I;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=98765; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

    * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_LC K; IF _N_=1 THEN DO; CT_LC=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO6; DROP K CT_LC CUM_NO1-CUM_NO6;

   IF LGHT_CON EQ 9 THEN DO;
      CT_LC=CT_LC +1;

   STOPPER=' NO';
   DO I=K TO 6 WHILE(STOPPER=' NO');
       J=I;
```

```
      IF CT_LC LE CUM THEN DO;
     LGTCON_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
END;

ELSE LGTCON_I=LGHT_CON;

DATA ONE (KEEP=CASENUM LGTCON_I WEATHER MAX_SEV MAN_COL
          ALCOHOL WEEKDAY EVENT1 WEIGHT DUMMY ); SET ONE;

*****************************************************************;
* NEXT STEP -- PROCEED TO NEXT VARIABLE: WEATHER;

DATA TWO; SET ONE; IF WEATHER NE 9;
PROC FREQ; TABLES WEATHER/OUT=TAB NOPRINT; WEIGHT WEIGHT;
DATA THREE; SET ONE; IF WEATHER EQ 9;
PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
 IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
CUM_PER= PERCENT + CUM_PER;
CUM_NO= CUM_PER * NO_UNKN * 0.01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO8;
         RETAIN CUM_NO1-CUM_NO8 I; DROP I;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=76543; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_WT K; IF _N_=1 THEN DO; CT_WT=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO8; DROP K CT_WT CUM_NO1-CUM_NO8;

   IF WEATHER EQ 9 THEN DO;
      CT_WT=CT_WT +1;

   STOPPER=' NO';
   DO I=K TO 8 WHILE(STOPPER=' NO');
       J=I;
      IF CT_WT LE CUM THEN DO;
     WEATHR_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
END;

ELSE WEATHR_I=WEATHER;
```

```
DATA ONE (KEEP=CASENUM LGTCON_I WEATHR_I MAX_SEV MAN_COL
          ALCOHOL WEEKDAY EVENT1 WEIGHT DUMMY ); SET ONE;

*****************************************************************;
* NEXT STEP -- PROCEED TO NEXT VARIABLE: MANNER OF COLLISION;

DATA TWO; SET ONE;  IF MAN_COL NE 9;
PROC FREQ; TABLES MAN_COL/OUT=TAB NOPRINT; WEIGHT WEIGHT;
 DATA THREE; SET ONE; IF MAN_COL=9;
 PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
CUM_PER =PERCENT + CUM_PER;
CUM_NO=CUM_PER * NO_UNKN * .01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO8;
          RETAIN CUM_NO1-CUM_NO8 I; DROP I;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=54321; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_MC K; IF _N_=1 THEN DO; CT_MC=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO8; DROP K CT_MC CUM_NO1-CUM_NO8;

   IF MAN_COL EQ 9 THEN DO;
      CT_MC=CT_MC +1;

   STOPPER=' NO';
   DO I=K TO 8 WHILE(STOPPER=' NO');
       J=I-1;
       IF I=8 THEN J=8;
      IF CT_MC LE CUM THEN DO;
     MANCOL_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
END;

ELSE MANCOL_I=MAN_COL;

DATA ONE (KEEP=CASENUM LGTCON_I WEATHR_I MAXSEV_I MANCOL_I
          ALCOHOL WEEKDAY EVENT1 WEIGHT DUMMY ); SET ONE;

*****************************************************************;
* NEXT STEP -- PROCEED TO NEXT VARIABLE: WEEKDAY      ;

DATA TWO; SET ONE; IF WEEKDAY NE 9;
PROC FREQ; TABLES WEEKDAY/OUT=TAB NOPRINT; WEIGHT WEIGHT;
DATA THREE; SET ONE; IF WEEKDAY EQ 9;
```

```
PROC SUMMARY ; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;

DATA TABLE; IF _N_=1 THEN SET NCOUNT;
SET TAB;
IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
CUM_PER=PERCENT + CUM_PER;
CUM_NO=CUM_PER*NO_UNKN*0.01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO7;
          RETAIN CUM_NO1-CUM_NO7 I; DROP I;
   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=32113; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_WK K; IF _N_=1 THEN DO; CT_WK=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO7; DROP K CT_WK CUM_NO1-CUM_NO7;

   IF WEEKDAY EQ 9 THEN DO;
      CT_WK=CT_WK +1;

   STOPPER=' NO';
   DO I=K TO 7 WHILE(STOPPER=' NO');
       J=I;
      IF CT_WK LE CUM THEN DO;
      WKDY_I=J;
      STOPPER='YES';
      K=I;
      END;
  END;
END;

ELSE WKDY_I=WEEKDAY;

DATA ONE (KEEP=CASENUM LGTCON_I WEATHR_I MAXSEV_I MANCOL_I
          ALCHL_I WKDY_I EVENT1 WEIGHT DUMMY ); SET ONE;

*****************************************************************;
* NEXT STEP -- PROCEED TO NEXT VARIABLE: FIRST HARMFUL EVNT;

DATA TWO; SET ONE; IF (EVENT1 NE 99) AND (EVENT1 NE 25);
PROC FREQ; TABLES EVENT1/ OUT=TAB NOPRINT; WEIGHT WEIGHT;
DATA THREE; SET ONE; IF EVENT1 EQ 99;
PROC SUMMARY ; VAR DUMMY; OUTPUT OUT=NCOUNT
N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT;
   SET TAB;
IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
CUM_PER=PERCENT +CUM_PER;
CUM_NO=CUM_PER*NO_UNKN*0.01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;
```

```
   DATA CUMDAT; SET TABLE; DUMMY=1;
   DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO37;
          RETAIN V1-V37  CUM_NO1-CUM_NO37 I; DROP I;
          ARRAY VALUE (I) V1-V37;

   IF _N_=1 THEN I=0;
   I=I+1;
   CUM=CUM_NO;
   VALUE=EVENT1;
   IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=21211; RAN_NO=RANUNI(SEED);
   PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

   * NEXT STEP : PERFORM IMPUTATION;

   DATA ONE  ; IF _N_=1 THEN SET CUMDAT; SET ONE;
   RETAIN CT_EV K; IF _N_=1 THEN DO; CT_EV=0; K=1; END;
   ARRAY CUM (I) CUM_NO1-CUM_NO37; DROP K CT_EV CUM_NO1-CUM_NO37;
    ARRAY VALUE (I) V1-V37; DROP V1-V37;

   IF EVENT1 EQ 99 THEN DO;
      CT_EV=CT_EV +1;

   STOPPER=' NO';
   DO I=K TO 37 WHILE(STOPPER=' NO');
      IF CT_EV LE CUM THEN DO;
     EVENT1_I=VALUE;
      STOPPER='YES';
      K=I;
      END;
  END;
END;
ELSE EVENT1_I=EVENT1;
LABEL
  LGTCON_I='IMPUTED LGT CONDITIONS'
  WEATHR_I='IMPUTED WEATHR CONDITION'
  MAXSEV_I='IMPUTED MAXIMUM INJURY'
  MANCOL_I='IMPUTED MANNER OF COLLISION'
  ALCHL_I ='IMPUTED ALCOHOL INVLMNT'
  WKDY_I  ='IMPUTED DAY OF WEEK'
  EVENT1_I='IMPUTED FIRST HARMFUL EVENT'    ;

LENGTH LGTCON_I WEATHR_I MAXSEV_I MANCOL_I ALCHL_I WKDY_I EVENT1_I 3;

KEEP CASENUM LGTCON_I WEATHR_I MAXSEV_I MANCOL_I ALCHL_I WKDY_I EVENT1_I;

*****************************************************************;

PROC SORT SORTSIZE=150000  DATA=ACC91_I; BY CASENUM;
PROC SORT SORTSIZE=150000  DATA=ONE;     BY CASENUM;

DATA ACC91_I; MERGE ACC91_I ONE; BY CASENUM;

*****************************************************************;

PROC FREQ DATA=ACC91_I;
TABLES HOUR_I MINUTE_I RELJCT_I ALIGN_I PROFIL_I SURCON_I TRFCON_I
LGTCON_I WEATHR_I MAXSEV_I MANCOL_I ALCHL_I WKDY_I EVENT1_I
```

```
      /LIST ;

*****************************************************************;

      ***************************************************************;
      *  (3)         HOTDECK IMPUTATION - ACCIDENT FILE              ;
      *                    SPEED LIMIT                               ;
      *           USING STRATUM AND SORTING BY RANDOM NUMBER         ;
      ***************************************************************;

DATA ONE
 (KEEP= LAND_USE SUR_COND TRAF_CON NUM_LAN REL_JCT TRAF_WAY INT_HWY
       SPD_LIM CASENUM STRATUM RAN_NO);  SET GES91.ACC91;
SEED=12345; RAN_NO =RANUNI(SEED);

PROC FREQ; TABLES SPD_LIM;

PROC SORT SORTSIZE=150000   DATA=ONE OUT=TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON NUM_LAN REL_JCT TRAF_WAY
   INT_HWY RAN_NO;

*---------------------------------------*;
*  (1)   INTERSTATE HIGHWAY              *;
*---------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON NUM_LAN REL_JCT TRAF_WAY INT_HWY;

IF FIRST.INT_HWY AND SPD_LIM=99 THEN TRACK=99;
ELSE IF SPD_LIM^=99 THEN TRACK=SPD_LIM;

RETAIN TRACK;

IF NOT(FIRST.INT_HWY) AND SPD_LIM=99 THEN SPDLM_H1=TRACK;
ELSE SPDLM_H1=SPD_LIM;

DROP TRACK;

PROC FREQ; TABLES SPDLM_H1;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON NUM_LAN REL_JCT TRAF_WAY RAN_NO;

*-------------------------------------*;
* (2)   TRAFFICWAY FLOW               *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON NUM_LAN REL_JCT TRAF_WAY;

IF FIRST.TRAF_WAY AND SPDLM_H1=99 THEN TRACK=99;
ELSE IF SPDLM_H1^=99 THEN TRACK=SPDLM_H1;

RETAIN TRACK;

IF NOT(FIRST.TRAF_WAY) AND SPDLM_H1=99 THEN SPDLM_H2=TRACK;
ELSE SPDLM_H2=SPDLM_H1;

DROP TRACK;
```

```
DATA TEST; SET TEST; DROP SPDLM_H1;

PROC FREQ; TABLES SPDLM_H2;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON NUM_LAN REL_JCT RAN_NO;

*-------------------------------------*;
* (3)   RELATION TO JUNCTION          *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON NUM_LAN REL_JCT;

IF FIRST.REL_JCT AND SPDLM_H2=99 THEN TRACK=99;
ELSE IF SPDLM_H2^=99 THEN TRACK=SPDLM_H2;

RETAIN TRACK;

IF NOT(FIRST.REL_JCT) AND SPDLM_H2=99 THEN SPDLM_H3=TRACK;
ELSE SPDLM_H3=SPDLM_H2;

DROP TRACK;

DATA TEST; SET TEST; DROP SPDLM_H2;
PROC FREQ; TABLES SPDLM_H3;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON NUM_LAN RAN_NO;

*-------------------------------------*;
* (4)     NUMBER OF TRAVEL LANES      *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON NUM_LAN;

IF FIRST.NUM_LAN AND SPDLM_H3=99 THEN TRACK=99;
ELSE IF SPDLM_H3^=99 THEN TRACK=SPDLM_H3;

RETAIN TRACK;

IF NOT(FIRST.NUM_LAN) AND SPDLM_H3=99 THEN SPDLM_H4=TRACK;
ELSE SPDLM_H4=SPDLM_H3;

DROP TRACK;

DATA TEST; SET TEST; DROP SPDLM_H3;

PROC FREQ; TABLES SPDLM_H4;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON RAN_NO;

*-------------------------------------*;
* (5)    TRAFFIC CONTROL DEVICE       *;
*-------------------------------------*;
```

```
DATA TEST;
SET TEST;
BY STRATUM LAND_USE SUR_COND TRAF_CON;

IF FIRST.TRAF_CON AND SPDLM_H4=99 THEN TRACK=99;
ELSE IF SPDLM_H4^=99 THEN TRACK=SPDLM_H4;

RETAIN TRACK;

IF NOT(FIRST.TRAF_CON) AND SPDLM_H4=99 THEN SPDLM_H5=TRACK;
ELSE SPDLM_H5=SPDLM_H4;

DROP TRACK;

DATA TEST; SET TEST; DROP SPDLM_H4;

PROC FREQ; TABLES SPDLM_H5;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM LAND_USE SUR_COND RAN_NO;

*-------------------------------------*;
* (6)   ROADWAY SURFACE CONDITION     *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM LAND_USE SUR_COND;

IF FIRST.SUR_COND AND SPDLM_H5=99 THEN TRACK=99;
ELSE IF SPDLM_H5^=99 THEN TRACK=SPDLM_H5;

RETAIN TRACK;

IF NOT(FIRST.SUR_COND) AND SPDLM_H5=99 THEN SPDLM_H6=TRACK;
ELSE SPDLM_H6=SPDLM_H5;

DROP TRACK;

DATA TEST; SET TEST; DROP SPDLM_H5;

PROC FREQ; TABLES SPDLM_H6;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM LAND_USE RAN_NO;

*-------------------------------------*;
* (7)       LAND  USE                 *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM LAND_USE;

IF FIRST.LAND_USE AND SPDLM_H6=99 THEN TRACK=99;
ELSE IF SPDLM_H6^=99 THEN TRACK=SPDLM_H6;

RETAIN TRACK;

IF NOT(FIRST.LAND_USE) AND SPDLM_H6=99 THEN SPDLM_H7=TRACK;
```

```
ELSE SPDLM_H7=SPDLM_H6;
DROP TRACK;

PROC FREQ; TABLES SPDLM_H7;

DATA ACC91_H (KEEP=CASENUM SPDLIM_H);
SET TEST;  IF SPDLM_H7=99 THEN SPDLM_H7=55;

SPDLIM_H=SPDLM_H7;

LABEL SPDLIM_H='HOTDECK IMPUTED SPEED LIMIT';
LENGTH SPDLIM_H 3;
PROC FREQ; TABLES SPDLIM_H/LIST;

*************************************************************;
*   THIS PROGRAM combines three files for ACCIDENT          *;
*     ACCIDENT    FILE CONTAINS UNIMPUTED ACCIDENT          *;
*          VARIABLES AND STRATUM                            *;
*     ACC91_H     FILE CONTAINS HOTDECK IMPUTED            *;
*          ACCIDENT VARIABLES (USING STRATUM AND RANDOM #) *;
*     ACC91_I     FILE CONTAINWS UNIVERIATE IMPUTED        *;
*          ACCIDENT VARIABLES (USING WEIGHT)                *;
*                                                           *;
*************************************************************;

PROC SORT SORTSIZE=150000 DATA= GES91.ACC91; BY CASENUM ;
PROC SORT SORTSIZE=150000 DATA= ACC91_H; BY CASENUM ;
PROC SORT SORTSIZE=150000 DATA= ACC91_I; BY CASENUM ;

DATA GES91.ACC91_HI; MERGE GES91.ACC91 ACC91_H ACC91_I;
      BY CASENUM ;

LENGTH HOUR MINUTE REL_JCT ALIGN PROFILE SUR_COND TRAF_CON
LGHT_CON WEATHER MAX_SEV MAN_COL ALCOHOL WEEKDAY EVENT1 SPD_LIM
INT_HWY LAND_USE MONTH NON_INVL NUM_INJ NUM_LAN PED_ACC PSU PSU_RCD
REGION REL_JCT REL_RWY RUR_URB SCHL_BUS STRATUM TRAF_WAY
YEAR VEH_INVL 3;

PROC CONTENTS DATA=GES91.ACC91_HI;

DATA ONE; SET GES91.ACC91_HI;
IF SPD_LIM NE 99;
S=2; IF SPD_LIM=SPDLIM_H THEN S=1;
PROC FREQ; TABLES S/LIST;
```

```
********** IMPUTATION FOR 1991 VEHICLE FILE **************;

OPTION NONUMBER NODATE PAGESIZE=60 LINESIZE=130 ;
*********************************************************************;
libname  ges91     'c:\ges\ges91' ;
libname  LIBRARY   'c:\ges\ges91' ;
*********************************************************************;
PROC FREQ DATA=GES91.VEH91 ; TABLES MODEL_YR MANEUVER VIOLATN
               HIT_RUN VEH_ROLE MAX_VSEV VEH_ALCH
               BODY_TYP IMPACT V_EVENT/LIST;

*********************************************************************;

*************************************************************;
*       UNIVARIATE IMPUTATION-VEHICLE FILE                  *;
*           USING VARIABLE WEIGHT                           *;
*   THIS STEP CALCULATES THE FREQUENCY DISTRIBUTION         *;
*   FOR THE KNOWN VALUES OF THE VARIABLES TO BE IMPUTED     *;
*   AND APPLIES THESE NUMBERS TO THE RECORDS WITH           *;
*   UNKNOWN VALUES, GENERATING A CUMULATIVE FREQUENCY       *;
*   SCALE FOR SUCH RECORDS IN THE VEHICLE FILE.  THE        *;
*   RESULTS ARE THEN USED TO IMPUTE FOR MISSING VALUES      *;
*   FOR VARIABLES IN THE  VEHICLE FILE.                     *;
*************************************************************;
*                  VARIABLES IMPUTED                          ;
*                                                             ;
*        NAME              OLD SAS NAME      NEW SAS NAME     ;
*        ----              ------------      ------------     ;
*     MODEL YEAR             MODEL_YR          MDLYR_I        ;
*     MANEUVER               MANEUVER          MANEUV_I       ;
*     VIOLATIONS             VIOLATN           VLTN_I         ;
*     HIT AND RUN            HIT_RUN           HITRUN_I       ;
*     VEHICLE ROLE           VEH_ROLE          VROLE_I        ;
*     MAX. INJ. SEV          MAX_VSEV          MAXSEV_I       ;
*     VEHICLE ALCOHOL        VEH_ALCH          V_ALCH_I       ;
*************************************************************;

DATA VEH91_I (KEEP=MODEL_YR MANEUVER VIOLATN
               HIT_RUN VEH_ROLE MAX_VSEV VEH_ALCH
               CASENUM VEHNO WEIGHT DUMMY); SET GES91.VEH91 ;
DUMMY=1;
*********************************************************************;
************       IMPUTATION OF MODEL_YR       *****;

DATA TWO ; SET VEH91_I;
IF MODEL_YR NE 99; PROC FREQ;TABLES MODEL_YR/OUT=TAB NOPRINT;
WEIGHT WEIGHT;
DATA THREE; SET VEH91_I; IF MODEL_YR EQ 99;
   PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
 IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
 CUM_PER= PERCENT + CUM_PER;
 CUM_NO=CUM_PER * NO_UNKN*.01; CUM_NO=ROUND(CUM_NO);

   DATA CUMDAT; SET TABLE; DUMMY=1;
    DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO51;
       RETAIN V1-V51 CUM_NO1-CUM_NO51 I; DROP I;
       ARRAY VALUE (I) V1-V51;
 IF _N_=1 THEN I=0;
 I=I+1;
```

```
 CUM=CUM_NO;
VALUE=MODEL_YR;
IF LAST.DUMMY THEN OUTPUT;

 DATA VEH91_I; SET VEH91_I;
SEED=23457; RAN_NO=RANUNI(SEED);
PROC SORT SORTSIZE=150000   DATA=VEH91_I ; BY RAN_NO;

* NEXT STEP  PERFORM IMPUTATION;

 DATA VEH91_I; IF  _N_=1 THEN SET CUMDAT; SET VEH91_I;
 RETAIN CT_MY K; IF _N_=1 THEN DO; CT_MY=0; K=1; END;
   ARRAY CUM(I) CUM_NO1-CUM_NO51; DROP K CT_MY CUM_NO1-CUM_NO51;
    ARRAY VALUE (I) V1-V51; DROP V1-V51;

  IF MODEL_YR EQ 99 THEN DO;
     CT_MY=CT_MY+1;

   STOPPER=' NO';
     DO I=K TO 51 WHILE(STOPPER=' NO');
        IF CT_MY LE CUM THEN DO;
          MDLYR_I=VALUE;
          STOPPER='YES';
           K=I;
           END;
       END;
    END;
  ELSE MDLYR_I=MODEL_YR;
  DATA VEH91_I (KEEP=MDLYR_I MANEUVER VIOLATN
               HIT_RUN VEH_ROLE MAX_VSEV VEH_ALCH
               CASENUM VEHNO WEIGHT DUMMY); SET VEH91_I;

*********************************************************************;

 * PROCEED TO NEXT VARIABLE  MANEUVER;

DATA TWO; SET VEH91_I;
IF MANEUVER NE 99; PROC FREQ; TABLES MANEUVER/OUT=TAB NOPRINT;
WEIGHT WEIGHT;
 DATA THREE; SET VEH91_I; IF MANEUVER EQ 99;
   PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF  _N_=1 THEN SET NCOUNT; SET TAB;
IF  _N_=1 THEN CUM_PER=0;
RETAIN CUM_PER;
CUM_PER=PERCENT + CUM_PER;
CUM_NO= CUM_PER * NO_UNKN * .01; CUM_NO=ROUND(CUM_NO);

   DATA CUMDAT; SET TABLE; DUMMY=1;
    DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO16;
        RETAIN V1-V16 CUM_NO1-CUM_NO16 I; DROP I;
        ARRAY VALUE (I) V1-V16;
 IF  _N_=1 THEN I=0;
 I=I+1;
 CUM=CUM_NO;
VALUE=MANEUVER;
IF LAST.DUMMY THEN OUTPUT;

DATA VEH91_I; SET VEH91_I;
SEED=45679; RAN_NO=RANUNI(SEED);
PROC SORT SORTSIZE=150000   DATA=VEH91_I ; BY RAN_NO;
```

```
* NEXT STEP  PERFORM IMPUTATION;

 DATA VEH91_I; IF _N_=1 THEN SET CUMDAT; SET VEH91_I;
 RETAIN CT_MN K; IF _N_=1 THEN DO; CT_MN=0; K=1; END;
   ARRAY CUM(I) CUM_NO1-CUM_NO16; DROP K CT_MN CUM_NO1-CUM_NO16;
    ARRAY VALUE (I) V1-V16; DROP V1-V16;

  IF MANEUVER EQ 99 THEN DO;
     CT_MN=CT_MN+1;

   STOPPER=' NO';
     DO I=K TO 16 WHILE(STOPPER=' NO');
        IF CT_MN LE CUM THEN DO;
          MANEUV_I=VALUE;
          STOPPER='YES';
           K=I;
           END;
       END;
    END;
  ELSE MANEUV_I=MANEUVER;

DATA VEH91_I (KEEP=MDLYR_I MANEUV_I VIOLATN
               HIT_RUN VEH_ROLE MAX_VSEV VEH_ALCH
               CASENUM VEHNO WEIGHT DUMMY); SET VEH91_I;

*****************************************************************;

* NEXT STEP -- PROCEED TO NEXT VARIABLE VIOLATIONS;

DATA TWO; SET VEH91_I;
IF VIOLATN NE 99; PROC FREQ; TABLES VIOLATN/OUT=TAB NOPRINT;
WEIGHT WEIGHT;
DATA THREE; SET VEH91_I; IF VIOLATN EQ 99;
PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
CUM_PER=PERCENT + CUM_PER;
CUM_NO=CUM_PER * NO_UNKN * .01 ; CUM_NO=ROUND(CUM_NO);


   DATA CUMDAT; SET TABLE; DUMMY=1;
    DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO11;
        RETAIN V1-V11 CUM_NO1-CUM_NO11 I; DROP I;
        ARRAY VALUE (I) V1-V11;
 IF _N_=1 THEN I=0;
 I=I+1;
 CUM=CUM_NO;
VALUE=VIOLATN;
IF LAST.DUMMY THEN OUTPUT;

 DATA VEH91_I; SET VEH91_I;
SEED=56789; RAN_NO=RANUNI(SEED);
PROC SORT SORTSIZE=150000   DATA=VEH91_I ; BY RAN_NO;

* NEXT STEP  PERFORM IMPUTATION;

 DATA VEH91_I; IF _N_=1 THEN SET CUMDAT; SET VEH91_I;
 RETAIN CT_VL K; IF _N_=1 THEN DO; CT_VL=0; K=1; END;
   ARRAY CUM(I) CUM_NO1-CUM_NO11; DROP K CT_VL CUM_NO1-CUM_NO11;
    ARRAY VALUE (I) V1-V11; DROP V1-V11;
```

```
  IF VIOLATN EQ 99 THEN DO;
     CT_VL=CT_VL+1;

   STOPPER=' NO';
     DO I=K TO 11 WHILE(STOPPER=' NO');
        IF CT_VL LE CUM THEN DO;
          VLTN_I=VALUE;
          STOPPER='YES';
           K=I;
           END;
       END;
    END;
  ELSE VLTN_I=VIOLATN;
DATA VEH91_I (KEEP=MDLYR_I MANEUV_I VLTN_I
                HIT_RUN VEH_ROLE MAX_VSEV VEH_ALCH
                CASENUM VEHNO WEIGHT DUMMY); SET VEH91_I;

*****************************************************************;

* NEXT STEP -- PROCEED TO NEXT VARIABLE  VEH. ROLE ;

 DATA TWO; SET VEH91_I;
IF VEH_ROLE NE 9; PROC FREQ; TABLES VEH_ROLE/OUT=TAB NOPRINT;
 WEIGHT WEIGHT;
 DATA THREE; SET VEH91_I; IF VEH_ROLE EQ 9;
PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
 DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
CUM_PER= PERCENT + CUM_PER;
CUM_NO= CUM_PER *NO_UNKN * .01; CUM_NO=ROUND(CUM_NO);

   DATA CUMDAT; SET TABLE; DUMMY=1;
    DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO4;
        RETAIN V1-V4 CUM_NO1-CUM_NO4 I; DROP I;
        ARRAY VALUE (I) V1-V4;
 IF _N_=1 THEN I=0;
 I=I+1;
 CUM=CUM_NO;
VALUE=VEH_ROLE;
IF LAST.DUMMY THEN OUTPUT;

 DATA VEH91_I; SET VEH91_I;
SEED=98765; RAN_NO=RANUNI(SEED);
PROC SORT SORTSIZE=150000   DATA=VEH91_I ; BY RAN_NO;
* NEXT STEP  PERFORM IMPUTATION;

 DATA VEH91_I; IF _N_=1 THEN SET CUMDAT; SET VEH91_I;
 RETAIN CT_VR K; IF _N_=1 THEN DO; CT_VR=0; K=1; END;
   ARRAY CUM(I) CUM_NO1-CUM_NO4; DROP K CT_VR CUM_NO1-CUM_NO4;
    ARRAY VALUE (I) V1-V4; DROP V1-V4;

  IF VEH_ROLE EQ 9 THEN DO;
     CT_VR=CT_VR+1;

   STOPPER=' NO';
     DO I=K TO 4 WHILE(STOPPER=' NO');
        IF CT_VR LE CUM THEN DO;
          VROLE_I=VALUE;
          STOPPER='YES';
           K=I;
```

```
          END;
        END;
      END;
  ELSE VROLE_I=VEH_ROLE;

DATA VEH91_I (KEEP=MDLYR_I MANEUV_I VLTN_I
                V_ALCH_I MAXSEV_I VROLE_I HIT_RUN
                CASENUM VEHNO WEIGHT DUMMY); SET VEH91_I;

*****************************************************************;

* NEXT STEP -- PROCEED TO NEXT VARIABLE HIT AND RUN ;

DATA TWO; SET VEH91_I; IF HIT_RUN NE 9;
PROC FREQ; TABLES HIT_RUN/OUT=TAB NOPRINT; WEIGHT WEIGHT;
DATA THREE; SET VEH91_I; IF HIT_RUN EQ 9;
PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
   CUM_PER= PERCENT + CUM_PER;
   CUM_NO= CUM_PER * NO_UNKN * 0.01 ; CUM_NO=ROUND(CUM_NO);


   DATA CUMDAT; SET TABLE; DUMMY=1;
    DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO2;
        RETAIN V1-V2 CUM_NO1-CUM_NO2 I; DROP I;
        ARRAY VALUE (I) V1-V2;
 IF _N_=1 THEN I=0;
 I=I+1;
 CUM=CUM_NO;
VALUE=HIT_RUN;
IF LAST.DUMMY THEN OUTPUT;

 DATA VEH91_I; SET VEH91_I;
SEED=76543; RAN_NO=RANUNI(SEED);
PROC SORT SORTSIZE=150000   DATA=VEH91_I ; BY RAN_NO;

* NEXT STEP  PERFORM IMPUTATION;

 DATA VEH91_I; IF _N_=1 THEN SET CUMDAT; SET VEH91_I;
 RETAIN CT_HR K; IF _N_=1 THEN DO; CT_HR=0; K=1; END;
   ARRAY CUM(I) CUM_NO1-CUM_NO2;
    DROP K CT_HR CUM_NO1-CUM_NO2;
    ARRAY VALUE (I) V1-V2; DROP V1-V2;

  IF HIT_RUN EQ 9 THEN DO;
     CT_HR=CT_HR+1;

   STOPPER=' NO';
     DO I=K TO 2 WHILE(STOPPER=' NO');
        IF CT_HR LE CUM THEN DO;
          HITRUN_I=VALUE;
          STOPPER='YES';
           K=I;
           END;
       END;
    END;
  ELSE HITRUN_I=HIT_RUN;

 LABEL
```

```
    MDLYR_I='UNIVERIATE IMPUTED MODEL YR'
    MANEUV_I='UNIVERIATE IMPUTED VEHICLE MANEUVER'
    VLTN_I='UNIVERIATE IMPUTED VIOLATIONS CHARGED'
    V_ALCH_I='UNIVERIATE IMPUTED ALCOHOL INVOLVED'
    MAXSEV_I='UNIVERIATE IMPUTED MAX SEVERITY'
    VROLE_I='UNIVERIATE IMPUTED VEHICLE ROLE'
    HITRUN_I='UNIVERIATE IMPUTED HIT AND RUN' ;

  KEEP MDLYR_I MANEUV_I VLTN_I V_ALCH_I MAXSEV_I VROLE_I  HITRUN_I
               CASENUM VEHNO ;

 PROC FREQ; TABLES MDLYR_I MANEUV_I VLTN_I V_ALCH_I MAXSEV_I VROLE_I
                   HITRUN_I/LIST;
*************************************************************************;

*************************************************************;
*            HOTDECK IMPUTATION - VEHICLE FILE              *;
*                   BODY TYPE (BODY_TYP)                    *;
*       USING STRATUM AND SORTING BY RANDOM NUMBER          *;
*************************************************************;

 DATA VEHICLE  (KEEP=MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE
        WEIGHT RAN_NO MAKE BODY_TYP CASENUM VEHNO STRATUM);
 SET GES91.VEH91;
 SEED=2468; RAN_NO=RANUNI(SEED);

  DATA PART1 ; SET VEHICLE;

 IF (1 LE BODY_TYP LE 9) OR (14 LE BODY_TYP LE 49) ;

 DATA PART2 ; SET VEHICLE;

 IF (10 LE BODY_TYP LE 13) OR (50 LE BODY_TYP LE 99);

*=========================================================*;
*---------------------------------------------------------*;
* (49)   AUTO, VAN, OR LIGHT TRUCKS CLASS/MAKE            *;
*---------------------------------------------------------*;

PROC SORT SORTSIZE=150000   DATA=PART1 OUT=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE MAKE RAN_NO;
DATA TEST;  SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE MAKE ;

IF FIRST.MAKE AND BODY_TYP=49 THEN TRACK=49;
ELSE IF BODY_TYP^=49 THEN TRACK=BODY_TYP;
RETAIN TRACK;
IF NOT(FIRST.MAKE) AND BODY_TYP=49 THEN BDTYP_H1=TRACK;
ELSE BDTYP_H1=BODY_TYP;
DROP TRACK BODY_TYP;

PROC FREQ; TABLES BDTYP_H1/LIST;
*---------------------------------------------------------*;
* (49)   AUTO, VAN, OR LIGHT TRUCKS CLASS/SPEC_USE        *;
*---------------------------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE ;
```

```
IF FIRST.SPEC_USE AND BDTYP_H1=49 THEN TRACK=49;
ELSE IF BDTYP_H1^=49 THEN TRACK=BDTYP_H1;
RETAIN TRACK;
IF NOT(FIRST.SPEC_USE) AND BDTYP_H1=49 THEN BDTYP_H2=TRACK;
ELSE BDTYP_H2=BDTYP_H1;
DROP TRACK;

DATA TEST; SET TEST; DROP BDTYP_H1;

PROC FREQ; TABLES BDTYP_H2/LIST;
*-------------------------------------*;
* (49)   AUTO, VAN, OR LIGHT TRUCKS CLASS/TRAILER      *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER ;

IF FIRST.TRAILER AND BDTYP_H2=49 THEN TRACK=49;
ELSE IF BDTYP_H2^=49 THEN TRACK=BDTYP_H2;
RETAIN TRACK;
IF NOT(FIRST.TRAILER) AND BDTYP_H2=49 THEN BDTYP_H3=TRACK;
ELSE BDTYP_H3=BDTYP_H2;
DROP TRACK;

DATA TEST; SET TEST; DROP BDTYP_H2;
PROC FREQ; TABLES BDTYP_H3/LIST;

*-------------------------------------*;
* (49)   AUTO, VAN, OR LIGHT TRUCKS CLASS/JACKNIFE     *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE ;

IF FIRST.JACKNIFE AND BDTYP_H3=49 THEN TRACK=49;
ELSE IF BDTYP_H3^=49 THEN TRACK=BDTYP_H3;
RETAIN TRACK;
IF NOT(FIRST.JACKNIFE) AND BDTYP_H3=49 THEN BDTYP_H4=TRACK;
ELSE BDTYP_H4=BDTYP_H3;
DROP TRACK;
RUN;

DATA TEST; SET TEST; DROP BDTYP_H3;
PROC FREQ; TABLES BDTYP_H4/LIST;

*-------------------------------------*;
* (49)   AUTO, VAN, OR LIGHT TRUCKS CLASS/OCC_INV      *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL ;

IF FIRST.OCC_INVL AND BDTYP_H4=49 THEN TRACK=49;
ELSE IF BDTYP_H4^=49 THEN TRACK=BDTYP_H4;
RETAIN TRACK;
IF NOT(FIRST.OCC_INVL) AND BDTYP_H4=49 THEN BDTYP_H5=TRACK;
ELSE BDTYP_H5=BDTYP_H4;
```

```
DROP TRACK;

DATA TEST; SET TEST; DROP BDTYP_H4;
PROC FREQ; TABLES BDTYP_H5/LIST;

*-------------------------------------*;
* (49)   AUTO, VAN, OR LIGHT TRUCKS CLASS/MODEL        *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL ;

IF FIRST.MODEL AND BDTYP_H5=49 THEN TRACK=49;
ELSE IF BDTYP_H5^=49 THEN TRACK=BDTYP_H5;
RETAIN TRACK;
IF NOT(FIRST.MODEL) AND BDTYP_H5=49 THEN BDTYP_H6=TRACK;
ELSE BDTYP_H6=BDTYP_H5;
DROP TRACK;

DATA PART1; SET TEST; BODY_TYP=BDTYP_H6;
IF BODY_TYP=49 THEN BODY_TYP=99;
DROP BDTYP_H5;

PROC FREQ; TABLES BDTYP_H6/LIST;
*=============================================================*;
DATA VEHICLE; SET PART1 PART2;
*=============================================================*;

DATA PART1; SET VEHICLE;
IF (14 LE BODY_TYP LE 48) OR (60 LE BODY_TYP LE 69) ;

DATA PART2; SET VEHICLE;
IF (BODY_TYP LE 13) OR (50 LE BODY_TYP LE 59)
    OR (70 LE BODY_TYP LE 99);

*-----------------------------------------------*;
* (69)   MEDIUM/HEAVY TRUCKS CLASS/MAKE         *;
*-----------------------------------------------*;

PROC SORT SORTSIZE=150000   DATA=PART1 OUT=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE MAKE RAN_NO;
DATA TEST;  SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE MAKE ;

IF FIRST.MAKE AND BODY_TYP=69 THEN TRACK=69;
ELSE IF BODY_TYP^=69 THEN TRACK=BODY_TYP;
RETAIN TRACK;
IF NOT(FIRST.MAKE) AND BODY_TYP=69 THEN BDTYP_H1=TRACK;
ELSE BDTYP_H1=BODY_TYP;
DROP TRACK BODY_TYP;

PROC FREQ; TABLES BDTYP_H1/LIST;

*-------------------------------------*;
* (69)   MEDIUM/HEAVY TRUCKS CLASS/SPEC_USE     *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE RAN_NO ;
DATA TEST; SET TEST;
```

```
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE ;

IF FIRST.SPEC_USE AND BDTYP_H1=69 THEN TRACK=69;
ELSE IF BDTYP_H1^=69 THEN TRACK=BDTYP_H1;
RETAIN TRACK;
IF NOT(FIRST.SPEC_USE) AND BDTYP_H1=69 THEN BDTYP_H2=TRACK;
ELSE BDTYP_H2=BDTYP_H1;
DROP TRACK BDTYP_H1;

PROC FREQ; TABLES BDTYP_H2/LIST;

*-------------------------------------*;
* (69)   MEDIUM/HEAVY TRUCKS CLASS/TRAILER      *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER ;

IF FIRST.TRAILER AND BDTYP_H2=69 THEN TRACK=69;
ELSE IF BDTYP_H2^=69 THEN TRACK=BDTYP_H2;
RETAIN TRACK;
IF NOT(FIRST.TRAILER) AND BDTYP_H2=69 THEN BDTYP_H3=TRACK;
ELSE BDTYP_H3=BDTYP_H2;
DROP TRACK   BDTYP_H2;
PROC FREQ; TABLES BDTYP_H3/LIST;

*-------------------------------------*;
* (69)   MEDIUM/HEAVY TRUCKS CLASS/JACKNIFE     *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE ;

IF FIRST.JACKNIFE AND BDTYP_H3=69 THEN TRACK=69;
ELSE IF BDTYP_H3^=69 THEN TRACK=BDTYP_H3;
RETAIN TRACK;
IF NOT(FIRST.JACKNIFE) AND BDTYP_H3=69 THEN BDTYP_H4=TRACK;
ELSE BDTYP_H4=BDTYP_H3;
DROP TRACK BDTYP_H3;

PROC FREQ; TABLES BDTYP_H4/LIST;
*-------------------------------------*;
* (69)   MEDIUM/HEAVY TRUCKS CLASS/OCC_INV      *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL;

IF FIRST.OCC_INVL AND BDTYP_H4=69 THEN TRACK=69;
ELSE IF BDTYP_H4^=69 THEN TRACK=BDTYP_H4;
RETAIN TRACK;
IF NOT(FIRST.OCC_INVL) AND BDTYP_H4=69 THEN BDTYP_H5=TRACK;
ELSE BDTYP_H5=BDTYP_H4;
DROP TRACK BDTYP_H4;

PROC FREQ; TABLES BDTYP_H5/LIST;
*-------------------------------------*;
```

```
* (69)   MEDIUM/HEAVY TRUCKS CLASS/MODEL        *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL RAN_NO  ;
DATA TEST; SET TEST;
BY STRATUM MODEL ;

IF FIRST.MODEL AND BDTYP_H5=69 THEN TRACK=69;
ELSE IF BDTYP_H5^=69 THEN TRACK=BDTYP_H5;
RETAIN TRACK;
IF NOT(FIRST.MODEL) AND BDTYP_H5=69 THEN BDTYP_H6=TRACK;
ELSE BDTYP_H6=BDTYP_H5;
DROP TRACK BDTYP_H5;

DATA PART1I; SET TEST; BODY_TYP=BDTYP_H6;
IF BODY_TYP=69 THEN BODY_TYP=99;

PROC FREQ; TABLES BDTYP_H6/LIST;

*============================================================*;
DATA VEHICLE; SET PART1I PART2;
*============================================================*;
*-------------------------------------*;
* (99)   OTHER BODY CLASS/MAKE          *;
*-------------------------------------*;

PROC SORT SORTSIZE=150000   DATA=VEHICLE OUT=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE MAKE RAN_NO;
DATA TEST;  SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE MAKE ;

IF FIRST.MAKE AND BODY_TYP=99 THEN TRACK=99;
ELSE IF BODY_TYP^=99 THEN TRACK=BODY_TYP;
RETAIN TRACK;
IF NOT(FIRST.MAKE) AND BODY_TYP=99 THEN BDTYP_H1=TRACK;
ELSE BDTYP_H1=BODY_TYP;
DROP TRACK BODY_TYP;

PROC FREQ; TABLES BDTYP_H1/LIST;

*-------------------------------------*;
* (99)   OTHER BODY CLASS/SPEC_USE     *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER SPEC_USE ;

IF FIRST.SPEC_USE AND BDTYP_H1=99 THEN TRACK=99;
ELSE IF BDTYP_H1^=99 THEN TRACK=BDTYP_H1;
RETAIN TRACK;
IF NOT(FIRST.SPEC_USE) AND BDTYP_H1=99 THEN BDTYP_H2=TRACK;
ELSE BDTYP_H2=BDTYP_H1;
DROP TRACK BDTYP_H1;

PROC FREQ; TABLES BDTYP_H2/LIST;

*-------------------------------------*;
* (99)   OTHER BODY CLASS/TRAILER      *;
*-------------------------------------*;
```

```
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE TRAILER ;

IF FIRST.TRAILER AND BDTYP_H2=99 THEN TRACK=99;
ELSE IF BDTYP_H2^=99 THEN TRACK=BDTYP_H2;
RETAIN TRACK;
IF NOT(FIRST.TRAILER) AND BDTYP_H2=99 THEN BDTYP_H3=TRACK;
ELSE BDTYP_H3=BDTYP_H2;
DROP TRACK BDTYP_H2;

PROC FREQ; TABLES BDTYP_H3/LIST;

*-------------------------------------*;
* (99)   OTHER BODY CLASS/JACKNIFE     *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL JACKNIFE ;

IF FIRST.JACKNIFE AND BDTYP_H3=99 THEN TRACK=99;
ELSE IF BDTYP_H3^=99 THEN TRACK=BDTYP_H3;
RETAIN TRACK;
IF NOT(FIRST.JACKNIFE) AND BDTYP_H3=99 THEN BDTYP_H4=TRACK;
ELSE BDTYP_H4=BDTYP_H3;
DROP TRACK BDTYP_H3;

PROC FREQ; TABLES BDTYP_H4/LIST;
*-------------------------------------*;
* (99)   OTHER BODY CLASS/OCC_INV      *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL OCC_INVL RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL OCC_INVL;

IF FIRST.OCC_INVL AND BDTYP_H4=99 THEN TRACK=99;
ELSE IF BDTYP_H4^=99 THEN TRACK=BDTYP_H4;
RETAIN TRACK;
IF NOT(FIRST.OCC_INVL) AND BDTYP_H4=99 THEN BDTYP_H5=TRACK;
ELSE BDTYP_H5=BDTYP_H4;
DROP TRACK BDTYP_H4;

PROC FREQ; TABLES BDTYP_H5/LIST;

*-------------------------------------*;
* (99)   OTHER BODY CLASS/MODEL        *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM MODEL RAN_NO ;
DATA TEST; SET TEST;
BY STRATUM MODEL ;

IF FIRST.MODEL AND BDTYP_H5=99 THEN TRACK=99;
ELSE IF BDTYP_H5^=99 THEN TRACK=BDTYP_H5;
RETAIN TRACK;
IF NOT(FIRST.MODEL) AND BDTYP_H5=99 THEN BDTYP_H6=TRACK;
ELSE BDTYP_H6=BDTYP_H5;
```

```
DROP TRACK ;

DATA TEST; SET TEST; BDTYP_H=BDTYP_H6; DROP BDTYP_H6 BDTYP_H5;
PROC FREQ; TABLES BDTYP_H/LIST;

* ============================================================*;
* NEXT STEP PERFORM UNIVARIATE IMPUTATION;
*============================================================*;

DATA VEHICLE; SET TEST;
DUMMY=1;

DATA TWO ; SET VEHICLE; IF BDTYP_H  NE 99;
PROC FREQ ; TABLES BDTYP_H /NOPRINT OUT=TAB; WEIGHT WEIGHT;
DATA THREE; SET VEHICLE; IF BDTYP_H EQ 99;
PROC SUMMARY ; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
   DATA TABLE; IF _N_=1 THEN SET NCOUNT;
        SET TAB;
  IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
       CUM_PER=PERCENT+CUM_PER;
       CUM_NO = CUM_PER*NO_UNKN*.01; CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

   DATA CUMDAT; SET TABLE; DUMMY=1;
    DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO49;
        RETAIN V1-V49 CUM_NO1-CUM_NO49 I; DROP I;
       ARRAY VALUE (I) V1-V49;
 IF _N_=1 THEN I=0;
 I=I+1;
 CUM=CUM_NO;
VALUE=BDTYP_H ;
IF LAST.DUMMY THEN OUTPUT;

 DATA VEHICLE; SET VEHICLE;
 SEED=12345; RAN_NO1=RANUNI(SEED);
 PROC SORT SORTSIZE=150000   DATA=VEHICLE;BY RAN_NO;

 DATA VEHICLE; IF _N_=1 THEN SET CUMDAT; SET VEHICLE;
 RETAIN CT_BT K; IF _N_=1 THEN DO; CT_BT=0; K=1; END;
   ARRAY CUM(I) CUM_NO1-CUM_NO49; DROP K CT_BT CUM_NO1-CUM_NO49;
    ARRAY VALUE (I) V1-V49; DROP V1-V49;

  IF BDTYP_H  EQ 99 THEN DO;
     CT_BT=CT_BT+1;

   STOPPER=' NO';
     DO I=K TO 49 WHILE(STOPPER=' NO');
        IF CT_BT LE CUM THEN DO;
          BDYTYP_H=VALUE;
          STOPPER='YES';
           K=I;
           END;
       END;
    END;
  ELSE BDYTYP_H=BDTYP_H ;

PROC FREQ ; TABLES BDYTYP_H/LIST;

  DATA BDYTYP_H (KEEP=CASENUM VEHNO BDYTYP_H); SET VEHICLE;
```

```
*************************************************************;
*      HOTDECK IMPUTATION  INITIAL POINT OF IMPACT( IMPACT) *;
*       USING STRATUM AND SORTING BY RANDOM NUMBER          *;
*************************************************************;

DATA VEHICLE (KEEP=CASENUM VEHNO ACC_TYPE MANEUVER V_EVENT
            WEIGHT RAN_NO IMPACT STRATUM); SET GES91.VEH91 ;

SEED=2468; RAN_NO=RANUNI(SEED);

DATA ACCIDENT (KEEP=CASENUM VEH_INVL MAN_COL EVENT1);
SET GES91.ACC91_HI;

PROC SORT SORTSIZE=150000   DATA=VEHICLE;  BY CASENUM;
PROC SORT SORTSIZE=150000   DATA=ACCIDENT; BY CASENUM;

DATA VEHICLE; MERGE VEHICLE (IN=A) ACCIDENT ; BY CASENUM; IF A;

DATA TEST; SET VEHICLE; IF IMPACT GT 0;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM EVENT1 V_EVENT MAN_COL MANEUVER VEH_INVL ACC_TYPE RAN_NO;

*-------------------------------------*;
*  (1)   ACCIDENT TYPE                *;
*-------------------------------------*;

DATA TEST; SET TEST;
BY STRATUM EVENT1 V_EVENT MAN_COL MANEUVER VEH_INVL ACC_TYPE;

IF FIRST.ACC_TYPE AND IMPACT=9 THEN TRACK=9;
ELSE IF IMPACT^=9 THEN TRACK=IMPACT;

RETAIN TRACK;

IF NOT(FIRST.ACC_TYPE) AND IMPACT=9 THEN IMPCT_H1=TRACK;
ELSE IMPCT_H1=IMPACT;

DROP TRACK;

RUN;

DATA TEST;
SET TEST;
DROP IMPACT;
PROC FREQ; TABLES IMPCT_H1/LIST;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM EVENT1 V_EVENT MAN_COL MANEUVER VEH_INVL RAN_NO;

*-------------------------------------*;
* (2)  NUMBER OF VEHICLES INVOLVED    *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM EVENT1 V_EVENT MAN_COL MANEUVER VEH_INVL;

IF FIRST.VEH_INVL AND IMPCT_H1=9 THEN TRACK=9;
ELSE IF IMPCT_H1^=9 THEN TRACK=IMPCT_H1;
```

```
RETAIN TRACK;

IF NOT(FIRST.VEH_INVL) AND IMPCT_H1=9 THEN IMPCT_H2=TRACK;
ELSE IMPCT_H2=IMPCT_H1;

DROP TRACK;

RUN;


DATA TEST;
SET TEST;
DROP IMPCT_H1;
PROC FREQ; TABLES IMPCT_H2/LIST;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM EVENT1 V_EVENT MAN_COL MANEUVER RAN_NO;

*-------------------------------------*;
* (3)  VEHICLE MANEUVER               *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM EVENT1 V_EVENT MAN_COL MANEUVER;

IF FIRST.MANEUVER AND IMPCT_H2=9 THEN TRACK=9;
ELSE IF IMPCT_H2^=9 THEN TRACK=IMPCT_H2;

RETAIN TRACK;

IF NOT(FIRST.MANEUVER) AND IMPCT_H2=9 THEN IMPCT_H3=TRACK;
ELSE IMPCT_H3=IMPCT_H2;

DROP TRACK;

RUN;


DATA TEST;
SET TEST;
DROP IMPCT_H2;
PROC FREQ; TABLES IMPCT_H3/LIST;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM EVENT1 V_EVENT MAN_COL RAN_NO;

*-------------------------------------*;
* (4)  MANNER OF COLLISION            *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM EVENT1 V_EVENT MAN_COL;

IF FIRST.MAN_COL AND IMPCT_H3=9 THEN TRACK=9;
ELSE IF IMPCT_H3^=9 THEN TRACK=IMPCT_H3;

RETAIN TRACK;
```

```
IF NOT(FIRST.MAN_COL) AND IMPCT_H3=9 THEN IMPCT_H4=TRACK;
ELSE IMPCT_H4=IMPCT_H3;

DROP TRACK;

DATA TEST; SET TEST; DROP IMPCT_H3;
PROC FREQ; TABLES IMPCT_H4/LIST;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM EVENT1 V_EVENT RAN_NO;

*-------------------------------------*;
* (5)  MOST HARMFUL EVENT              *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM EVENT1 V_EVENT;

IF FIRST.V_EVENT AND IMPCT_H4=9 THEN TRACK=9;
ELSE IF IMPCT_H4^=9 THEN TRACK=IMPCT_H4;

RETAIN TRACK;

IF NOT(FIRST.V_EVENT) AND IMPCT_H4=9 THEN IMPCT_H5=TRACK;
ELSE IMPCT_H5=IMPCT_H4;

DROP TRACK;

DATA TEST; SET TEST; DROP IMPCT_H4;
PROC FREQ; TABLES IMPCT_H5/LIST;

PROC SORT SORTSIZE=150000   DATA=TEST;
BY STRATUM EVENT1 RAN_NO;

*-------------------------------------*;
* (6)  FIRST HARMFUL EVENT             *;
*-------------------------------------*;

DATA TEST;
SET TEST;
BY STRATUM EVENT1;

IF FIRST.EVENT1 AND IMPCT_H5=9 THEN TRACK=9;
ELSE IF IMPCT_H5^=9 THEN TRACK=IMPCT_H5;

RETAIN TRACK;

IF NOT(FIRST.EVENT1) AND IMPCT_H5=9 THEN IMPCT_H6=TRACK;
ELSE IMPCT_H6=IMPCT_H5;

DROP TRACK;

DATA TEST ;  SET TEST; IMP_H=IMPCT_H6;
DUMMY=1;
KEEP WEIGHT IMP_H CASENUM VEHNO DUMMY;

* ============================================================*;
* NEXT STEP PERFORM UNIVARIATE IMPUTATION;
*============================================================*;
```

```
DATA TWO ; SET TEST; IF IMP_H  NE 9;
PROC FREQ ; TABLES IMP_H /NOPRINT OUT=TAB; WEIGHT WEIGHT;
DATA THREE; SET TEST; IF IMP_H EQ 9;
PROC SUMMARY ; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
   DATA TABLE; IF _N_=1 THEN SET NCOUNT;
        SET TAB;
  IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
        CUM_PER=PERCENT+CUM_PER;
        CUM_NO = CUM_PER*NO_UNKN*.01; CUM_NO=ROUND(CUM_NO);
        DROP _TYPE_ _FREQ_;

   DATA CUMDAT; SET TABLE; DUMMY=1;
    DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO9;
         RETAIN V1-V9 CUM_NO1-CUM_NO9 I; DROP I;
         ARRAY VALUE (I) V1-V9;
 IF _N_=1 THEN I=0;
 I=I+1;
 CUM=CUM_NO;
 VALUE=IMP_H ;
 IF LAST.DUMMY THEN OUTPUT;

 DATA TEST; SET TEST;
 SEED=23456; RAN_NO=RANUNI(SEED);
 PROC SORT SORTSIZE=150000   DATA=TEST;BY RAN_NO;

 DATA TEST; IF _N_=1 THEN SET CUMDAT; SET TEST;
 RETAIN CT_BT K; IF _N_=1 THEN DO; CT_BT=0; K=1; END;
   ARRAY CUM(I) CUM_NO1-CUM_NO9; DROP K CT_BT CUM_NO1-CUM_NO9;
    ARRAY VALUE (I) V1-V9; DROP V1-V9;

  IF IMP_H  EQ 9 THEN DO;
     CT_BT=CT_BT+1;

   STOPPER=' NO';
     DO I=K TO 9 WHILE(STOPPER=' NO');
        IF CT_BT LE CUM THEN DO;
          IMPACT_H=VALUE;
          STOPPER='YES';
           K=I;
           END;
       END;
    END;
  ELSE IMPACT_H=IMP_H ;
  KEEP IMPACT_H CASENUM VEHNO;
*=========================================================;

  DATA ZERO; SET VEHICLE; IF IMPACT=0;
  IMPACT_H=IMPACT;
  KEEP IMPACT_H CASENUM VEHNO;

  DATA IMPACT_H ; SET ZERO TEST;

  PROC FREQ ; TABLES IMPACT_H/LIST;
  RUN;
*************************************************************;
*    HOTDECK IMPUTATION -  MOST HARMFUL EVENT (V_EVENT)     *;
*       USING STRATUM AND SORTING BY RANDOM NUMBER          *;
*************************************************************;

DATA VEHICLE (KEEP=RAN_NO CASENUM VEHNO ACC_TYPE VEH_ROLE V_EVENT
```

```
              WEIGHT STRATUM); SET GES91.VEH91 ;
SEED=2468; RAN_NO=RANUNI(SEED);

DATA ACCIDENT (KEEP=CASENUM EVENT1 VEH_INVL NON_INVL);
SET GES91.ACC91_HI;

PROC SORT SORTSIZE=150000   DATA=VEHICLE; BY CASENUM;
PROC SORT SORTSIZE=150000   DATA=ACCIDENT; BY CASENUM;

DATA VEHICLE;  MERGE VEHICLE(IN=A) ACCIDENT; BY CASENUM; IF A;

PROC SORT SORTSIZE=150000   DATA=VEHICLE;
BY STRATUM NON_INVL VEH_INVL ACC_TYPE VEH_ROLE EVENT1 RAN_NO;

*---------------------------------------*;
*  (1)   FIRST HARMFUL EVENT             *;
*---------------------------------------*;

DATA VEHICLE; SET VEHICLE;
BY STRATUM NON_INVL VEH_INVL ACC_TYPE VEH_ROLE EVENT1;

IF FIRST.EVENT1 AND V_EVENT=99 THEN TRACK=99;
ELSE IF V_EVENT^=99 THEN TRACK=V_EVENT;

RETAIN TRACK;

IF NOT(FIRST.EVENT1) AND V_EVENT=99 THEN V_EV_H1=TRACK;
ELSE V_EV_H1=V_EVENT;

DROP TRACK;

DATA VEHICLE; SET VEHICLE; DROP V_EVENT;
PROC FREQ; TABLES V_EV_H1/LIST;

PROC SORT SORTSIZE=150000   DATA=VEHICLE;
BY STRATUM NON_INVL VEH_INVL ACC_TYPE VEH_ROLE RAN_NO;

*---------------------------------------*;
* (2)        VEHICLE ROLE               *;
*---------------------------------------*;

DATA VEHICLE;   SET VEHICLE;
BY STRATUM NON_INVL VEH_INVL ACC_TYPE VEH_ROLE;

IF FIRST.VEH_ROLE AND V_EV_H1=99 THEN TRACK=99;
ELSE IF V_EV_H1^=99 THEN TRACK=V_EV_H1;

RETAIN TRACK;

IF NOT(FIRST.VEH_ROLE) AND V_EV_H1=99 THEN V_EV_H2=TRACK;
ELSE V_EV_H2=V_EV_H1;

DROP TRACK;

DATA VEHICLE;  SET VEHICLE; DROP V_EV_H1;
PROC FREQ; TABLES V_EV_H2/LIST;

PROC SORT SORTSIZE=150000   DATA=VEHICLE;
BY STRATUM NON_INVL VEH_INVL ACC_TYPE RAN_NO;
```

```
*--------------------------------------*;
* (3)    ACCIDENT TYPE                 *;
*--------------------------------------*;

DATA VEHICLE;
SET VEHICLE;
BY STRATUM NON_INVL VEH_INVL ACC_TYPE;

IF FIRST.ACC_TYPE AND V_EV_H2=99 THEN TRACK=99;
ELSE IF V_EV_H2^=99 THEN TRACK=V_EV_H2;

RETAIN TRACK;

IF NOT(FIRST.ACC_TYPE) AND V_EV_H2=99 THEN V_EV_H3=TRACK;
ELSE V_EV_H3=V_EV_H2;

DROP TRACK;

DATA VEHICLE; SET VEHICLE; DROP V_EV_H2;
PROC FREQ; TABLES V_EV_H3/LIST;

PROC SORT SORTSIZE=150000   DATA=VEHICLE;
BY STRATUM NON_INVL VEH_INVL RAN_NO;

*--------------------------------------*;
* (4)  NUMBER OF VEHICLES INVOLVED     *;
*--------------------------------------*;

DATA VEHICLE;
SET VEHICLE;
BY STRATUM NON_INVL VEH_INVL;

IF FIRST.VEH_INVL AND V_EV_H3=99 THEN TRACK=99;
ELSE IF V_EV_H3^=99 THEN TRACK=V_EV_H3;

RETAIN TRACK;

IF NOT(FIRST.VEH_INVL) AND V_EV_H3=99 THEN V_EV_H4=TRACK;
ELSE V_EV_H4=V_EV_H3;

DROP TRACK;

DATA VEHICLE; SET VEHICLE; DROP V_EV_H3;
PROC FREQ; TABLES V_EV_H4/LIST;

PROC SORT SORTSIZE=150000   DATA=VEHICLE;
BY STRATUM NON_INVL RAN_NO;

*--------------------------------------*;
* (5)  NUMBER OF NONMOTOREST INVOLVED *;
*--------------------------------------*;

DATA VEHICLE;  SET VEHICLE; BY STRATUM NON_INVL;

IF FIRST.NON_INVL AND V_EV_H4=99 THEN TRACK=99;
ELSE IF V_EV_H4^=99 THEN TRACK=V_EV_H4;

RETAIN TRACK;

IF NOT(FIRST.NON_INVL) AND V_EV_H4=99 THEN V_EV_H5=TRACK;
```

```
ELSE V_EV_H5=V_EV_H4;

VEV_H=V_EV_H5;
DUMMY=1;
PROC FREQ; TABLES V_EV_H5/LIST;

*================================================================*;
* NEXT STEP   PERFORM UNIVARIATE IMPUTATION;
*================================================================*;

DATA TWO ; SET VEHICLE; IF VEV_H  NE 99;
PROC FREQ ; TABLES VEV_H /NOPRINT OUT=TAB; WEIGHT WEIGHT;
DATA THREE; SET VEHICLE; IF VEV_H EQ 99;
PROC SUMMARY ; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
   DATA TABLE; IF _N_=1 THEN SET NCOUNT;
        SET TAB;
  IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
        CUM_PER=PERCENT+CUM_PER;
        CUM_NO = CUM_PER*NO_UNKN*.01; CUM_NO=ROUND(CUM_NO);
        DROP _TYPE_ _FREQ_;

   DATA CUMDAT; SET TABLE; DUMMY=1;
    DATA CUMDAT; SET CUMDAT; BY DUMMY; ARRAY CUM (I) CUM_NO1-CUM_NO39;
        RETAIN V1-V39 CUM_NO1-CUM_NO39 I; DROP I;
        ARRAY VALUE (I) V1-V39;
 IF _N_=1 THEN I=0;
 I=I+1;
 CUM=CUM_NO;
 VALUE=VEV_H ;
 IF LAST.DUMMY THEN OUTPUT;

 DATA VEHICLE; SET VEHICLE;
 SEED=34567; RAN_NO=RANUNI(SEED);
 PROC SORT SORTSIZE=150000   DATA=VEHICLE;BY RAN_NO;

 DATA VEHICLE; IF _N_=1 THEN SET CUMDAT; SET VEHICLE;
 RETAIN CT_BT K; IF _N_=1 THEN DO; CT_BT=0; K=1; END;
   ARRAY CUM(I) CUM_NO1-CUM_NO39; DROP K CT_BT CUM_NO1-CUM_NO39;
    ARRAY VALUE (I) V1-V39; DROP V1-V39;

  IF VEV_H  EQ 99 THEN DO;
     CT_BT=CT_BT+1;

   STOPPER=' NO';
     DO I=K TO 39 WHILE(STOPPER=' NO');
        IF CT_BT LE CUM THEN DO;
          V_EVNT_H=VALUE;
          STOPPER='YES';
           K=I;
          END;
       END;
    END;
  ELSE V_EVNT_H=VEV_H ;

  KEEP V_EVNT_H CASENUM VEHNO;

DATA MHARM_H (KEEP=CASENUM VEHNO V_EVNT_H); SET VEHICLE;

PROC FREQ; TABLES V_EVNT_H/LIST;
```

```
* COMBINING VEHICLE FILES FOR 1991*;
PROC SORT SORTSIZE=150000 DATA=GES91.VEH91 ; BY CASENUM VEHNO;
PROC SORT SORTSIZE=150000 DATA=VEH91_I;      BY CASENUM VEHNO;

PROC SORT SORTSIZE=150000  DATA=IMPACT_H;       BY CASENUM VEHNO;
PROC SORT SORTSIZE=150000  DATA=BDYTYP_H;       BY CASENUM VEHNO;
PROC SORT SORTSIZE=150000  DATA=MHARM_H ;        BY CASENUM VEHNO;

DATA GES91.VEH91_HI;
MERGE IMPACT_H BDYTYP_H MHARM_H  GES91.VEH91 VEH91_I;
BY CASENUM VEHNO;

LENGTH IMPACT_H 3 V_EVNT_H 3 BDYTYP_H 3 MDLYR_I 3 MANEUV_I 3
       VLTN_I 3 V_ALCH_I 3 MAXSEV_I 3 VROLE_I 3 HITRUN_I 3;

LABEL   BDYTYP_H='HOTDECK IMPUTED BODY TYPE'
        IMPACT_H='HOTDECK IMPUTED INITIAL POINT OF IMPACT'
        V_EVNT_H='HOTDECK IMPUTED MOST HARMFUL EVENT' ;

PROC CONTENTS DATA=GES91.VEH91_HI;

DATA VEHICLE; SET GES91.VEH91_HI;
IF BODY_TYP NE 49;
IF BODY_TYP NE 69;
IF BODY_TYP NE 99;
S1=2; IF BODY_TYP=BDYTYP_H THEN S1=1;
PROC FREQ; TABLES S1/LIST;

DATA VEHICLE; SET GES91.VEH91_HI;
IF IMPACT NE 9;
S2=2; IF IMPACT=IMPACT_H THEN S2=1;
PROC FREQ; TABLES S2/LIST;

DATA VEHICLE; SET GES91.VEH91_HI;
IF V_EVENT NE 99;
S3=2; IF V_EVENT=V_EVNT_H THEN S3=1;
PROC FREQ; TABLES S3/LIST;
```

```
************* IMPUTATION FOR 1991 PERSON FILE ************;

libname  ges91    'c:\ges\ges91' ;
libname  LIBRARY  'c:\ges\ges91' ;

OPTION  NONUMBER NODATE PS=70 LS=100;

*************************************************************;
*      UNIVARIATE IMPUTATION - PERSON FILE                  *;
*             USING WEIGHT VARIABLE                         *;
*   THIS STEP CALCULATES THE FREQUENCY DISTRIBUTION         *;
*   FOR THE KNOWN VALUES OF THE VARIABLES TO BE IMPUTED     *;
*   AND APPLIES THESE NUMBERS TO THE RECORDS WITH           *;
*   UNUNKNOWN VALUES, GENERATING A CUMULATIVE FREQUENCY     *;
*   SCALE FOR SUCH RECORDS IN THE PERSON FILE.  THE         *;
*   RESULTS ARE THEN USED TO IMPUTE FOR MISSING VALUES      *;
*   FOR PARTICULAR VARIABLES IN THE PERSON FILE             *;
*************************************************************;

PROC FREQ DATA=GES91.PER91;
        TABLES EJECT LOCATN REST_SYS AGE SEX INJ_SEV PER_ALCH
        SEAT_POS     /LIST;

*************************************************************;
*      VARIABLES IMPUTED:          LOCATION                 *;
*************************************************************;
DATA ONE (KEEP=CASENUM VEHNO PERNO EJECT LOCATN WEIGHT DUMMY);
SET GES91.PER91;  DUMMY=1;

DATA TWO ; SET ONE;
 IF LOCATN NE 99; PROC FREQ ;TABLES LOCATN/OUT=TAB NOPRINT;
 WEIGHT WEIGHT;
DATA THREE; SET ONE; IF LOCATN EQ 99;
   PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
 IF _N_=1 THEN CUM_PER=0; RETAIN CUM_PER;
 CUM_PER= PERCENT + CUM_PER;
 CUM_NO=CUM_PER * NO_UNKN*.01;  CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

DATA CUMDAT; SET TABLE; DUMMY=1;

  DATA CUMDAT; SET CUMDAT;BY DUMMY;ARRAY CUM (I) CUM_NO1-CUM_NO11;
              RETAIN V1-V11  CUM_NO1-CUM_NO11 I ; DROP I;
              ARRAY VALUE (I) V1-V11;
     IF _N_=1 THEN I=0;
     I=I+1;
     CUM=CUM_NO;
     VALUE=LOCATN;
     IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
   SEED=98765; RAN_NO=RANUNI(SEED);
    PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

    DATA ONE ;  IF _N_=1 THEN SET CUMDAT; SET ONE;
       RETAIN CT_LOC K; IF _N_=1 THEN DO;CT_LOC=0;K=1;END;
      ARRAY CUM (I) CUM_NO1-CUM_NO11 ; DROP K CT_LOC CUM_NO1-CUM_NO11;
      ARRAY VALUE (I) V1-V11; DROP V1-V11;
```

```
        IF LOCATN EQ 99  THEN DO;
              CT_LOC=CT_LOC+1;

    STOPPER=' NO';
        DO I=K TO 10 WHILE(STOPPER=' NO');
                  IF CT_LOC LE CUM THEN DO;
                       LOCATN_I=VALUE;
                       STOPPER='YES';
                       K=I;
                       END;
                END;
   END;
   ELSE LOCATN_I=LOCATN;

*************************************************************;
*      VARIABLES IMPUTED:  EJECT                            *;
*************************************************************;

  DATA TWO; SET ONE;
 IF EJECT NE 9; PROC FREQ ; TABLES EJECT/OUT=TAB NOPRINT;
 WEIGHT WEIGHT;
 DATA THREE; SET ONE; IF EJECT EQ 9;
PROC SUMMARY; VAR DUMMY; OUTPUT OUT=NCOUNT
N=NO_UNKN;
DATA TABLE; IF _N_=1 THEN SET NCOUNT; SET TAB;
IF _N_=1 THEN CUM_PER=0;
RETAIN CUM_PER;
CUM_PER=PERCENT +CUM_PER;
CUM_NO=CUM_PER * NO_UNKN*.01;  CUM_NO=ROUND(CUM_NO);
DROP _TYPE_ _FREQ_;

DATA CUMDAT; SET TABLE; DUMMY=1;

 DATA CUMDAT; SET CUMDAT;BY DUMMY;ARRAY CUM (I) CUM_NO1-CUM_NO2;
              RETAIN V1-V2  CUM_NO1-CUM_NO2 I ; DROP I;
              ARRAY VALUE (I) V1-V2;
     IF _N_=1 THEN I=0;
     I=I+1;
     CUM=CUM_NO;
     VALUE=EJECT;
     IF LAST.DUMMY THEN OUTPUT;

   DATA ONE; SET ONE;
     SEED=87654; RAN_NO=RANUNI(SEED);
     PROC SORT SORTSIZE=150000   DATA=ONE ; BY RAN_NO;

    DATA PER91_I; IF _N_=1 THEN SET CUMDAT; SET ONE;
       RETAIN CT_EJT K; IF _N_=1 THEN DO;CT_EJT=0;K=1;END;
      ARRAY CUM (I) CUM_NO1-CUM_NO2; DROP K CT_EJT CUM_NO1-CUM_NO2;
      ARRAY VALUE (I) V1-V2; DROP V1-V2;

        IF EJECT EQ 9  THEN DO;
              CT_EJT=CT_EJT+1;

    STOPPER=' NO';
        DO I=K TO 2 WHILE(STOPPER=' NO');
                  IF CT_EJT LE CUM THEN DO;
                       EJECT_I=VALUE;
                       STOPPER='YES';
                       K=I;
```

```
                    END;
                END;
    END;
  ELSE EJECT_I=EJECT;

  LABEL
     EJECT_I='IMPUTED EJECTION'
     LOCATN_I='IMPUTED LOCATION';

  KEEP CASENUM VEHNO PERNO EJECT_I LOCATN_I ;

PROC FREQ; TABLES EJECT_I LOCATN_I/LIST ;

*****************************************************************;
*            HOTDECK IMPUTATION - 1991 PERSON FILE              *;
*            USING STRATUM AND SORTING BY RANDOM NUMBER          *;
*                   PART ONE - HD IMPUTING AGE, SEX,             *;
*                   RESTRAINT SYSTEM USE AND INJURY SEVERITY     *;
*****************************************************************;

*===============================================================;
*            HOTDECK IMPUTATION OF AGE IN PERSON FILE       *;
*****************************************************************;
DATA ONE (KEEP=CASENUM VEHNO AGE REST_SYS SEX PER_TYPE SEAT_POS
         STRATUM RAN_NO PERNO PER_ALCH PER_DRUG); SET GES91.PER91 ;
SEED=2468; RAN_NO=RANUNI(SEED);

DATA TWO (KEEP=CASENUM VEHNO VIOLATN OCC_INVL); SET GES91.VEH91_HI ;

PROC SORT SORTSIZE=150000  DATA=ONE; BY CASENUM VEHNO;

PROC SORT SORTSIZE=150000  DATA=TWO; BY CASENUM VEHNO;

DATA ONE;  MERGE ONE (IN=A) TWO; BY CASENUM VEHNO; IF A;

PROC SORT SORTSIZE=150000  DATA=ONE;  BY CASENUM;

DATA THREE (KEEP=CASENUM WEEKDAY HOUR VEH_INVL);  SET GES91.ACC91_HI;

PROC SORT SORTSIZE=150000  DATA=THREE; BY CASENUM;

DATA ONE  ;  MERGE ONE (IN=A) THREE; BY CASENUM;       IF A;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG  SEX SEAT_POS
  PER_TYPE VIOLATN WEEKDAY HOUR RAN_NO;

*-------------------------------------*;
*  (1)  HOUR                          *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX SEAT_POS PER_TYPE
   VIOLATN WEEKDAY HOUR;

IF FIRST.HOUR AND AGE=99 THEN TRACK=99;
ELSE IF AGE^=99 THEN TRACK=AGE;
RETAIN TRACK;
IF NOT(FIRST.HOUR) AND AGE=99 THEN AGE_H1=TRACK;
ELSE AGE_H1=AGE;
```

```
DROP TRACK;

DATA ONE; SET ONE; DROP AGE;

PROC FREQ; TABLES AGE_H1/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX SEAT_POS PER_TYPE
 VIOLATN WEEKDAY RAN_NO;

*-------------------------------------*;
* (2)  WEEKDAY                        *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX SEAT_POS PER_TYPE
  VIOLATN WEEKDAY;

IF FIRST.WEEKDAY AND AGE_H1=99 THEN TRACK=99;
ELSE IF AGE_H1^=99 THEN TRACK=AGE_H1;
RETAIN TRACK;
IF NOT(FIRST.WEEKDAY) AND AGE_H1=99 THEN AGE_H2=TRACK;
ELSE AGE_H2=AGE_H1;
DROP TRACK;

DATA ONE; SET ONE; DROP AGE_H1;

PROC FREQ; TABLES AGE_H2/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX SEAT_POS PER_TYPE
VIOLATN RAN_NO;

*-------------------------------------*;
* (3) VIOLATION CHARGED               *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX SEAT_POS PER_TYPE
 VIOLATN;
IF FIRST.VIOLATN AND AGE_H2=99 THEN TRACK=99;
ELSE IF AGE_H2^=99 THEN TRACK=AGE_H2;
RETAIN TRACK;
IF NOT(FIRST.VIOLATN) AND AGE_H2=99 THEN AGE_H3=TRACK;
ELSE AGE_H3=AGE_H2;
DROP TRACK;

DATA ONE;  SET ONE;  DROP AGE_H2;

PROC FREQ; TABLES AGE_H3/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX SEAT_POS PER_TYPE
 RAN_NO;

*-------------------------------------*;
* (4)    PERSON TYPE                  *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX SEAT_POS PER_TYPE;
```

```
IF FIRST.PER_TYPE AND AGE_H3=99 THEN TRACK=99;
ELSE IF AGE_H3^=99 THEN TRACK=AGE_H3;
RETAIN TRACK;
IF NOT(FIRST.PER_TYPE) AND AGE_H3=99 THEN AGE_H4=TRACK;
ELSE AGE_H4=AGE_H3;
DROP TRACK;

DATA ONE; SET ONE; DROP AGE_H3;

PROC FREQ; TABLES AGE_H4/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX SEAT_POS RAN_NO;

*-------------------------------------*;
* (5)        SEAT POSTION             *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX SEAT_POS;

IF FIRST.SEAT_POS AND AGE_H4=99 THEN TRACK=99;
ELSE IF AGE_H4^=99 THEN TRACK=AGE_H4;
RETAIN TRACK;
IF NOT(FIRST.SEAT_POS) AND AGE_H4=99 THEN AGE_H5=TRACK;
ELSE AGE_H5=AGE_H4;
DROP TRACK;

DATA ONE; SET ONE; DROP AGE_H4;

PROC FREQ; TABLES AGE_H5/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX RAN_NO;

*-------------------------------------*;
* (6)           SEX                   *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX;

IF FIRST.SEX AND AGE_H5=99 THEN TRACK=99;
ELSE IF AGE_H5^=99 THEN TRACK=AGE_H5;
RETAIN TRACK;
IF NOT(FIRST.SEX) AND AGE_H5=99 THEN AGE_H6=TRACK;
ELSE AGE_H6=AGE_H5;
DROP TRACK;

DATA ONE; SET ONE; DROP AGE_H5;
PROC FREQ; TABLES AGE_H6/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG SEX;

*-------------------------------------*;
* (7) POLICE REPORTED DRUG INVL.      *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG;
```

```
IF FIRST.PER_DRUG AND AGE_H6=99 THEN TRACK=99;
ELSE IF AGE_H6^=99 THEN TRACK=AGE_H6;
RETAIN TRACK;
IF NOT(FIRST.PER_DRUG) AND AGE_H6=99 THEN AGE_H7=TRACK;
ELSE AGE_H7=AGE_H6;
DROP TRACK;

DATA ONE; SET ONE; DROP AGE_H6;

PROC FREQ; TABLES AGE_H7/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH PER_DRUG;

*-------------------------------------*;
* (8) POLICE REPORTED ALCOHOL INVL.   *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH;

IF FIRST.PER_ALCH AND AGE_H7=99 THEN TRACK=99;
ELSE IF AGE_H7^=99 THEN TRACK=AGE_H7;
RETAIN TRACK;
IF NOT(FIRST.PER_ALCH) AND AGE_H7=99 THEN AGE_H8=TRACK;
ELSE AGE_H8=AGE_H7;
DROP TRACK;

DATA ONE;  SET ONE;  DROP AGE_H7;

PROC FREQ; TABLES AGE_H8/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL PER_ALCH;

*-------------------------------------*;
* (9) NUMBER OF VEHICLES INVOLVED     *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL;

IF FIRST.VEH_INVL AND AGE_H8=99 THEN TRACK=99;
ELSE IF AGE_H8^=99 THEN TRACK=AGE_H8;
RETAIN TRACK;
IF NOT(FIRST.VEH_INVL) AND AGE_H8=99 THEN AGE_H9=TRACK;
ELSE AGE_H9=AGE_H8;
DROP TRACK;

DATA ONE;  SET ONE; DROP AGE_H8;

PROC FREQ; TABLES AGE_H9/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL VEH_INVL;

*-------------------------------------*;
* (10) NUMBER OF OCCUPANTS INVOLVED   *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY STRATUM REST_SYS OCC_INVL ;
```

```
IF FIRST.OCC_INVL AND AGE_H9=99 THEN TRACK=99;
ELSE IF AGE_H9^=99 THEN TRACK=AGE_H9;
RETAIN TRACK;
IF NOT(FIRST.OCC_INVL) AND AGE_H9=99 THEN AGE_H10=TRACK;
ELSE AGE_H10=AGE_H9;
DROP TRACK;

DATA ONE;  SET ONE; DROP AGE_H9;

PROC FREQ; TABLES AGE_H10/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS OCC_INVL ;

*-------------------------------------*;
* (11) RESTRAINT SYSTEM USED          *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS ;

IF FIRST.REST_SYS AND AGE_H10=99 THEN TRACK=99;
ELSE IF AGE_H10^=99 THEN TRACK=AGE_H10;
RETAIN TRACK;
IF NOT(FIRST.REST_SYS) AND AGE_H10=99 THEN AGE_H11=TRACK;
ELSE AGE_H11=AGE_H10;
AGE_H=AGE_H11;
DROP TRACK  AGE_H10;

PROC FREQ; TABLES AGE_H11/LIST;
DATA AGE_H  (KEEP=CASENUM VEHNO PERNO AGE_H); SET ONE;

IF AGE_H= 99 THEN AGE_H=20;

PROC FREQ; TABLES AGE_H/LIST;

*=========================================================*;
*           Hotdeck Imputation of Sex in Person file      *;
***********************************************************;
DATA ONE (KEEP=CASENUM VEHNO AGE REST_SYS SEX PER_TYPE PERNO
          STRATUM RAN_NO SEAT_POS PER_ALCH PER_DRUG);
SET GES91.PER91 ;
SEED=2468; RAN_NO=RANUNI(SEED);

PROC SORT SORTSIZE=150000  DATA=ONE; BY CASENUM VEHNO;

DATA TWO (KEEP=CASENUM VEHNO VIOLATN ); SET GES91.VEH91_HI ;
PROC SORT SORTSIZE=150000  DATA=TWO; BY CASENUM VEHNO;

DATA ONE; MERGE ONE (IN=A) TWO; BY CASENUM VEHNO; IF A;

PROC SORT SORTSIZE=150000  DATA=ONE; BY CASENUM;

DATA THREE (KEEP=CASENUM WEEKDAY HOUR VEH_INVL); SET GES91.ACC91_HI ;
PROC SORT SORTSIZE=150000  DATA=THREE; BY CASENUM;

DATA ONE; MERGE ONE (IN=A) THREE; BY CASENUM; IF A;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS PER_TYPE VIOLATN
 WEEKDAY HOUR RAN_NO;
```

```
*-------------------------------------*;
*  (1)  HOUR                          *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS PER_TYPE VIOLATN
 WEEKDAY HOUR;

IF FIRST.HOUR AND SEX=9 THEN TRACK=9;
ELSE IF SEX^=9 THEN TRACK=SEX;

RETAIN TRACK;

IF NOT(FIRST.HOUR) AND SEX=9 THEN SEX_H1=TRACK;
ELSE SEX_H1=SEX;

DROP TRACK;

DATA ONE; SET ONE; DROP SEX;

PROC FREQ ; TABLES SEX_H1/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS PER_TYPE VIOLATN
 WEEKDAY RAN_NO;

*-------------------------------------*;
* (2)  WEEKDAY                       *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS PER_TYPE VIOLATN
 WEEKDAY;

IF FIRST.WEEKDAY AND SEX_H1=9 THEN TRACK=9;
ELSE IF SEX_H1^=9 THEN TRACK=SEX_H1;

RETAIN TRACK;

IF NOT(FIRST.WEEKDAY) AND SEX_H1=9 THEN SEX_H2=TRACK;
ELSE SEX_H2=SEX_H1;

DROP TRACK;

DATA ONE;  SET ONE;  DROP SEX_H1;

PROC FREQ; TABLES SEX_H2/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS PER_TYPE
VIOLATN RAN_NO;

*-------------------------------------*;
* (3) VIOLATION CHARGED              *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS PER_TYPE VIOLATN;

IF FIRST.VIOLATN AND SEX_H2=9 THEN TRACK=9;
ELSE IF SEX_H2^=9 THEN TRACK=SEX_H2;
```

```
RETAIN TRACK;

IF NOT(FIRST.VIOLATN) AND SEX_H2=9 THEN SEX_H3=TRACK;
ELSE SEX_H3=SEX_H2;

DROP TRACK;

DATA ONE;  SET ONE;  DROP SEX_H2;

PROC FREQ; TABLES SEX_H3/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS PER_TYPE RAN_NO;

*-------------------------------------*;
* (4)    PERSON TYPE                  *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS PER_TYPE;

IF FIRST.PER_TYPE AND SEX_H3=9 THEN TRACK=9;
ELSE IF SEX_H3^=9 THEN TRACK=SEX_H3;

RETAIN TRACK;

IF NOT(FIRST.PER_TYPE) AND SEX_H3=9 THEN SEX_H4=TRACK;
ELSE SEX_H4=SEX_H3;

DROP TRACK;

DATA ONE;  SET ONE; DROP SEX_H3;
PROC FREQ; TABLES SEX_H4/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS RAN_NO;

*-------------------------------------*;
* (5)        SEAT POSTION             *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE SEAT_POS;

IF FIRST.SEAT_POS AND SEX_H4=9 THEN TRACK=9;
ELSE IF SEX_H4^=9 THEN TRACK=SEX_H4;

RETAIN TRACK;

IF NOT(FIRST.SEAT_POS) AND SEX_H4=9 THEN SEX_H5=TRACK;
ELSE SEX_H5=SEX_H4;

DROP TRACK;

DATA ONE; SET ONE; DROP SEX_H4;

PROC FREQ; TABLES SEX_H5/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE RAN_NO;

*-------------------------------------*;
```

```
* (6)            AGE                          *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG AGE;

IF FIRST.AGE AND SEX_H5=9 THEN TRACK=9;
ELSE IF SEX_H5^=9 THEN TRACK=SEX_H5;

RETAIN TRACK;

IF NOT(FIRST.AGE) AND SEX_H5=9 THEN SEX_H6=TRACK;
ELSE SEX_H6=SEX_H5;

DROP TRACK;

DATA ONE; SET ONE; DROP SEX_H5;

PROC FREQ; TABLES SEX_H6/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG RAN_NO;

*-------------------------------------*;
* (7) POLICE REPORTED DRUG INVL.      *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH PER_DRUG;

IF FIRST.PER_DRUG AND SEX_H6=9 THEN TRACK=9;
ELSE IF SEX_H6^=9 THEN TRACK=SEX_H6;

RETAIN TRACK;

IF NOT(FIRST.PER_DRUG) AND SEX_H6=9 THEN SEX_H7=TRACK;
ELSE SEX_H7=SEX_H6;

DROP TRACK;

DATA ONE; SET ONE; DROP SEX_H6;

PROC FREQ; TABLES SEX_H7/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH RAN_NO;

*-------------------------------------*;
* (8) POLICE REPORTED ALCOHOL INVL.   *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY STRATUM REST_SYS VEH_INVL PER_ALCH;

IF FIRST.PER_ALCH AND SEX_H7=9 THEN TRACK=9;
ELSE IF SEX_H7^=9 THEN TRACK=SEX_H7;

RETAIN TRACK;

IF NOT(FIRST.PER_ALCH) AND SEX_H7=9 THEN SEX_H8=TRACK;
ELSE SEX_H8=SEX_H7;
```

```
DROP TRACK;

DATA ONE;  SET ONE; DROP SEX_H7;

PROC FREQ; TABLES SEX_H8/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS VEH_INVL RAN_NO;

*-------------------------------------*;
* (9) NUMBER OF VEHICLES INVOLVED     *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_SYS VEH_INVL;

IF FIRST.VEH_INVL AND SEX_H8=9 THEN TRACK=9;
ELSE IF SEX_H8^=9 THEN TRACK=SEX_H8;

RETAIN TRACK;

IF NOT(FIRST.VEH_INVL) AND SEX_H8=9 THEN SEX_H9=TRACK;
ELSE SEX_H9=SEX_H8;

DROP TRACK;

DATA ONE; SET ONE; DROP SEX_H8;
PROC FREQ; TABLES SEX_H9/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_SYS RAN_NO;

*-------------------------------------*;
* (10) RESTRAINT SYSTEM USED          *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY STRATUM REST_SYS;

IF FIRST.REST_SYS AND SEX_H9=9 THEN TRACK=9;
ELSE IF SEX_H9^=9 THEN TRACK=SEX_H9;

RETAIN TRACK;

IF NOT(FIRST.REST_SYS) AND SEX_H9=9 THEN SEX_H10=TRACK;
ELSE SEX_H10=SEX_H9;

DROP TRACK;

DATA ONE;  SET ONE; DROP SEX_H9;
PROC FREQ; TABLES SEX_H10/LIST;

DATA SEX_H  (KEEP=CASENUM VEHNO PERNO SEX_H); SET ONE;
SEX_H=SEX_H10;
IF SEX_H=9 THEN SEX_H=1;
PROC FREQ; TABLES SEX_H/LIST;
*============================================================;
*       Hotdeck Imputation of REST_SYS in Person file        *;
*************************************************************;

DATA ONE (KEEP=SEX EJECT INJ_SEV PER_TYPE AGE SEAT_POS STRATUM
```

```
            RAN_NO REST_SYS CASENUM VEHNO PERNO)  ;SET GES91.PER91 ;
SEED=2468; RAN_NO=RANUNI(SEED);
IF REST_SYS GT 0;

PROC SORT SORTSIZE=150000  DATA=ONE ;
BY STRATUM SEX EJECT INJ_SEV PER_TYPE AGE SEAT_POS RAN_NO;

*-------------------------------------*;
* (1)  SEAT POSTION                   *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM SEX EJECT INJ_SEV PER_TYPE AGE SEAT_POS;

IF FIRST.SEAT_POS AND REST_SYS=9 THEN TRACK=9;
ELSE IF REST_SYS^=9 THEN TRACK=REST_SYS;

RETAIN TRACK;

IF NOT(FIRST.SEAT_POS) AND REST_SYS=9 THEN R_SYS_H1=TRACK;
ELSE R_SYS_H1=REST_SYS;

DROP TRACK;

DATA ONE; SET ONE; DROP REST_SYS;
PROC FREQ ; TABLES R_SYS_H1/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM SEX EJECT INJ_SEV PER_TYPE AGE RAN_NO;

*-------------------------------------*;
* (2) AGE                             *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM SEX EJECT INJ_SEV PER_TYPE AGE;

IF FIRST.AGE AND R_SYS_H1=9 THEN TRACK=9;
ELSE IF R_SYS_H1^=9 THEN TRACK=R_SYS_H1;

RETAIN TRACK;

IF NOT(FIRST.AGE) AND R_SYS_H1=9 THEN R_SYS_H2=TRACK;
ELSE R_SYS_H2=R_SYS_H1;

DROP TRACK;

DATA ONE; SET ONE; DROP R_SYS_H1;
PROC FREQ ; TABLES R_SYS_H2/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM SEX EJECT INJ_SEV PER_TYPE RAN_NO;

*-------------------------------------*;
* (3) PERSON TYPE                     *;
*-------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM SEX EJECT INJ_SEV PER_TYPE;
```

```
IF FIRST.PER_TYPE AND R_SYS_H2=9 THEN TRACK=9;
ELSE IF R_SYS_H2^=9 THEN TRACK=R_SYS_H2;

RETAIN TRACK;

IF NOT(FIRST.PER_TYPE) AND R_SYS_H2=9 THEN R_SYS_H3=TRACK;
ELSE R_SYS_H3=R_SYS_H2;

DROP TRACK;

DATA ONE; SET ONE; DROP R_SYS_H2;

PROC FREQ ; TABLES R_SYS_H3/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM SEX EJECT INJ_SEV RAN_NO;

*--------------------------------------*;
* (4) INJURY SEVERITY                  *;
*--------------------------------------*;
DATA ONE;   SET ONE;
BY STRATUM SEX EJECT INJ_SEV;

IF FIRST.INJ_SEV AND R_SYS_H3=9 THEN TRACK=9;
ELSE IF R_SYS_H3^=9 THEN TRACK=R_SYS_H3;

RETAIN TRACK;

IF NOT(FIRST.INJ_SEV) AND R_SYS_H3=9 THEN R_SYS_H4=TRACK;
ELSE R_SYS_H4=R_SYS_H3;

DROP TRACK;

DATA ONE;   SET ONE; DROP R_SYS_H3;

PROC FREQ ; TABLES R_SYS_H4/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM SEX EJECT INJ_SEV;

*--------------------------------------*;
* (5) EJECT                            *;
*--------------------------------------*;
DATA ONE; SET ONE;
BY STRATUM SEX EJECT;

IF FIRST.EJECT AND R_SYS_H4=9 THEN TRACK=9;
ELSE IF R_SYS_H4^=9 THEN TRACK=R_SYS_H4;

RETAIN TRACK;

IF NOT(FIRST.EJECT) AND R_SYS_H4=9 THEN R_SYS_H5=TRACK;
ELSE R_SYS_H5=R_SYS_H4;

DROP TRACK;

DATA ONE; SET ONE; DROP R_SYS_H4;
PROC FREQ ; TABLES R_SYS_H5/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
```

```
BY STRATUM SEX EJECT;

*-------------------------------------*;
* (6) SEX                             *;
*-------------------------------------*;
DATA ONE;    SET ONE; BY STRATUM SEX;

IF FIRST.SEX AND R_SYS_H5=9 THEN TRACK=9;
ELSE IF R_SYS_H5^=9 THEN TRACK=R_SYS_H5;

RETAIN TRACK;

IF NOT(FIRST.SEX) AND R_SYS_H5=9 THEN R_SYS_H6=TRACK;
ELSE R_SYS_H6=R_SYS_H5;
DROP TRACK;

PROC FREQ ; TABLES R_SYS_H6/LIST;

DATA ONE;  SET ONE;   RESTSY_H=R_SYS_H6;
IF RESTSY_H=9 THEN RESTSY_H=8;
DROP R_SYS_H5 R_SYS_H6;
KEEP CASENUM VEHNO PERNO RESTSY_H;

DATA TWO (KEEP=CASENUM VEHNO PERNO RESTSY_H REST_SYS) ; SET GES91.PER91 ;
IF REST_SYS EQ 0;
RESTSY_H=REST_SYS;

DATA RESTSY_H  (KEEP=CASENUM VEHNO PERNO RESTSY_H); SET ONE TWO;
IF RESTSY_H = 9 THEN RESTSY_H = 1;
  PROC FREQ ; TABLES RESTSY_H/LIST;

*=========================================================*;
*       Hotdeck Imputation of INJ_SEV  in Person file     *;
***********************************************************;

DATA ONE (KEEP=CASENUM VEHNO REST_SYS PER_TYPE EJECT HOSPITAL
              INJ_SEV PERNO STRATUM RAN_NO ); SET GES91.PER91 ;
SEED=2468; RAN_NO=RANUNI(SEED);

PROC SORT SORTSIZE=150000  DATA=ONE; BY CASENUM VEHNO;

DATA TWO (KEEP=CASENUM VEHNO TOWED ACC_TYPE VEH_SEV );
SET GES91.VEH91_HI ;

PROC SORT SORTSIZE=150000  DATA=TWO; BY CASENUM VEHNO;

DATA ONE; MERGE ONE (IN=A) TWO; BY CASENUM VEHNO; IF A;

PROC SORT SORTSIZE=150000  DATA=ONE;

BY  STRATUM ACC_TYPE EJECT HOSPITAL PER_TYPE REST_SYS TOWED VEH_SEV RAN_NO;

*---------------------------------------*;
*  (1)   DAMAGE SEVERITY                *;
*---------------------------------------*;
DATA ONE; SET ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL PER_TYPE REST_SYS TOWED VEH_SEV;

IF FIRST.VEH_SEV AND INJ_SEV=9 THEN TRACK=9;
ELSE IF INJ_SEV^=9 THEN TRACK=INJ_SEV;
```

```
RETAIN TRACK;

IF NOT(FIRST.VEH_SEV) AND INJ_SEV=9 THEN INJSV_H1=TRACK;
ELSE INJSV_H1=INJ_SEV;

DROP TRACK;

PROC FREQ ; TABLES INJSV_H1/LIST;

DATA ONE;  SET ONE;  DROP INJ_SEV;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL PER_TYPE REST_SYS TOWED VEH_SEV;

*-------------------------------------*;
* (2)  MANNER OF LEAVING THE SCENE    *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL PER_TYPE REST_SYS TOWED;

IF FIRST.TOWED AND INJSV_H1=9 THEN TRACK=9;
ELSE IF INJSV_H1^=9 THEN TRACK=INJSV_H1;

RETAIN TRACK;

IF NOT(FIRST.TOWED) AND INJSV_H1=9 THEN INJSV_H2=TRACK;
ELSE INJSV_H2=INJSV_H1;

DROP TRACK;

DATA ONE; SET ONE; DROP INJSV_H1;
PROC FREQ ; TABLES INJSV_H2/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL PER_TYPE REST_SYS TOWED;

*-------------------------------------*;
* (3)    RESTRAINT SYSTEM USED        *;
*-------------------------------------*;

DATA ONE;  SET ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL PER_TYPE REST_SYS;

IF FIRST.REST_SYS AND INJSV_H2=9 THEN TRACK=9;
ELSE IF INJSV_H2^=9 THEN TRACK=INJSV_H2;

RETAIN TRACK;

IF NOT(FIRST.REST_SYS) AND INJSV_H2=9 THEN INJSV_H3=TRACK;
ELSE INJSV_H3=INJSV_H2;

DROP TRACK;

DATA ONE;  SET ONE; DROP INJSV_H2;
PROC FREQ ; TABLES INJSV_H3/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL PER_TYPE REST_SYS;
```

```
*--------------------------------------*;
* (4)          PERSON TYPE             *;
*--------------------------------------*;
DATA ONE; SET ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL PER_TYPE;

IF FIRST.PER_TYPE AND INJSV_H3=9 THEN TRACK=9;
ELSE IF INJSV_H3^=9 THEN TRACK=INJSV_H3;

RETAIN TRACK;

IF NOT(FIRST.PER_TYPE) AND INJSV_H3=9 THEN INJSV_H4=TRACK;
ELSE INJSV_H4=INJSV_H3;

DROP TRACK;

DATA ONE; SET ONE; DROP INJSV_H3;

PROC FREQ ; TABLES INJSV_H4/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL RAN_NO;

*--------------------------------------*;
* (5)          HOSPITAL                *;
*--------------------------------------*;
DATA ONE; SET ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL;

IF FIRST.HOSPITAL AND INJSV_H4=9 THEN TRACK=9;
ELSE IF INJSV_H4^=9 THEN TRACK=INJSV_H4;

RETAIN TRACK;

IF NOT(FIRST.HOSPITAL) AND INJSV_H4=9 THEN INJSV_H5=TRACK;
ELSE INJSV_H5=INJSV_H4;

DROP TRACK;

DATA ONE; SET ONE;  DROP INJSV_H4;

PROC FREQ ; TABLES INJSV_H5/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY  STRATUM ACC_TYPE EJECT HOSPITAL;

*--------------------------------------*;
* (6)           EJECTION               *;
*--------------------------------------*;
DATA ONE; SET ONE;
BY  STRATUM ACC_TYPE EJECT;

IF FIRST.EJECT AND INJSV_H5=9 THEN TRACK=9;
ELSE IF INJSV_H5^=9 THEN TRACK=INJSV_H5;

RETAIN TRACK;

IF NOT(FIRST.EJECT) AND INJSV_H5=9 THEN INJSV_H6=TRACK;
ELSE INJSV_H6=INJSV_H5;
```

```
DROP TRACK;

DATA ONE;  SET ONE; DROP INJSV_H5;
PROC FREQ ; TABLES INJSV_H6/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY  STRATUM ACC_TYPE EJECT;

*--------------------------------------*;
* (7)      ACCIDENT TYPE               *;
*--------------------------------------*;
DATA ONE; SET ONE;
BY  STRATUM ACC_TYPE;

IF FIRST.ACC_TYPE AND INJSV_H6=9 THEN TRACK=9;
ELSE IF INJSV_H6^=9 THEN TRACK=INJSV_H6;

RETAIN TRACK;

IF NOT(FIRST.ACC_TYPE) AND INJSV_H6=9 THEN INJSV_H7=TRACK;
ELSE INJSV_H7=INJSV_H6;

DROP TRACK;

PROC FREQ ; TABLES INJSV_H7/LIST;

DATA INJSV_H  (KEEP=CASENUM VEHNO PERNO INJSEV_H); SET ONE;
INJSEV_H=INJSV_H7;
IF INJSEV_H=9 THEN INJSEV_H=0;
PROC FREQ; TABLES INJSEV_H/LIST;

*************************************************************;
*                 Hotdeck Imputation - Person File          *;
*           USING STRATUM AND SORTING BY RANDOM NUMBER      *;
*                  PART TWO - HD IMPUTING SEATING POSITION  *;
*                  AND POLICE REP. ALCHOL INV.              *;
*************************************************************;

*=============================================================;
*        Hotdeck Imputation of Seating Position             *;
*      IMPUTING FOR VALUES 18, 19, 28, 29 AND 99            *;
*                        in Person file                     *;
*           USING STRATUM AND SORTING BY RANDOM NUMBER      *;
*************************************************************;

DATA ONE (KEEP=REST_TYP AGE PER_TYPE SEAT_POS CASENUM
          RAN_NO PERNO VEHNO STRATUM);  SET GES91.PER91 ;
SEED=2468; RAN_NO=RANUNI(SEED);
PROC SORT SORTSIZE=150000  DATA=ONE; BY CASENUM VEHNO;

DATA TWO (KEEP=CASENUM VEHNO BDYTYP_H ); SET GES91.VEH91_HI;

PROC SORT SORTSIZE=150000  DATA=TWO; BY CASENUM VEHNO;

DATA ONE; MERGE ONE (IN=A) TWO; BY CASENUM VEHNO; IF A;

 DATA PART0; SET ONE; IF (SEAT_POS EQ 0);
 SEAT_H=SEAT_POS;

 DATA PART3; SET ONE; IF (SEAT_POS GT 29);
```

```
 SEAT_H=SEAT_POS;
*-------------------------------------*;
* (1) FRONT SEAT/PERSON TYPE          *;
*-------------------------------------*;
DATA PART1; SET ONE;
IF 0 LT SEAT_POS LE 19;
IF SEAT_POS=18 THEN SEAT_POS=19;

PROC SORT SORTSIZE=150000  DATA=PART1 ;
BY STRATUM REST_TYP BDYTYP_H AGE PER_TYPE RAN_NO;

DATA PART1;
SET PART1;
BY STRATUM REST_TYP BDYTYP_H AGE PER_TYPE;

IF FIRST.PER_TYPE AND SEAT_POS=19 THEN TRACK=19;
ELSE IF SEAT_POS^=19 THEN TRACK=SEAT_POS;

RETAIN TRACK;

IF NOT(FIRST.PER_TYPE) AND SEAT_POS=19 THEN SEAT_H1=TRACK;
ELSE SEAT_H1=SEAT_POS;

DROP TRACK;

PROC FREQ; TABLES SEAT_H1/LIST;

PROC SORT SORTSIZE=150000  DATA=PART1;
BY STRATUM REST_TYP BDYTYP_H AGE RAN_NO;

*-------------------------------------*;
* (2)   FRONT SEAT/AGE                *;
*-------------------------------------*;
DATA PART1; SET PART1;
BY STRATUM REST_TYP BDYTYP_H AGE;

IF FIRST.AGE AND SEAT_H1=19 THEN TRACK=19;
ELSE IF SEAT_H1^=19 THEN TRACK=SEAT_H1;

RETAIN TRACK;

IF NOT(FIRST.AGE) AND SEAT_H1=19 THEN SEAT_H2=TRACK;
ELSE SEAT_H2=SEAT_H1;

DROP TRACK;

DATA PART1; SET PART1; DROP SEAT_H1;
PROC FREQ; TABLES SEAT_H2/LIST;

PROC SORT SORTSIZE=150000  DATA=PART1;
BY STRATUM REST_TYP BDYTYP_H RAN_NO;

*-------------------------------------*;
* (3)     FRONT SEAT/BODY TYPE        *;
*-------------------------------------*;
DATA PART1; SET PART1;
BY STRATUM REST_TYP BDYTYP_H;

IF FIRST.BDYTYP_H AND SEAT_H2=19 THEN TRACK=19;
ELSE IF SEAT_H2^=19 THEN TRACK=SEAT_H2;
```

```
RETAIN TRACK;

IF NOT(FIRST.BDYTYP_H) AND SEAT_H2=19 THEN SEAT_H3=TRACK;
ELSE SEAT_H3=SEAT_H2;

DROP TRACK;

DATA PART1; SET PART1; DROP SEAT_H2;
 PROC FREQ; TABLES SEAT_H3/LIST;

PROC SORT SORTSIZE=150000  DATA=PART1;
BY STRATUM REST_TYP RAN_NO;

*----------------------------------------*;
* (4) FRONT SEAT/RESTRAINT SYSTEM TYPE   *;
*----------------------------------------*;
DATA PART1; SET PART1;
BY STRATUM REST_TYP;

IF FIRST.REST_TYP AND SEAT_H3=19 THEN TRACK=19;
ELSE IF SEAT_H3^=19 THEN TRACK=SEAT_H3;

RETAIN TRACK;

IF NOT(FIRST.REST_TYP) AND SEAT_H3=19 THEN SEAT_H4=TRACK;
ELSE SEAT_H4=SEAT_H3;

DROP TRACK;

PROC FREQ; TABLES SEAT_H4/LIST;

DATA PART1; SET PART1; SEAT_H=SEAT_H4;
DROP SEAT_H3 SEAT_H4;
IF SEAT_H=19 THEN SEAT_H=99;

*-------------------------------------*;
* (1) SECOND SEAT/PERSON TYPE         *;
*-------------------------------------*;
DATA PART2; SET ONE;
IF 21 LE SEAT_POS LE 29;
IF SEAT_POS=28 THEN SEAT_POS=29;

PROC SORT SORTSIZE=150000  DATA=PART2 ;
BY STRATUM REST_TYP BDYTYP_H AGE PER_TYPE RAN_NO;

DATA PART2;   SET PART2;
BY STRATUM REST_TYP BDYTYP_H AGE PER_TYPE;

IF FIRST.PER_TYPE AND SEAT_POS=29 THEN TRACK=29;
ELSE IF SEAT_POS^=29 THEN TRACK=SEAT_POS;

RETAIN TRACK;

IF NOT(FIRST.PER_TYPE) AND SEAT_POS=29 THEN SEAT_H1=TRACK;
ELSE SEAT_H1=SEAT_POS;

DROP TRACK;

PROC FREQ; TABLES SEAT_H1/LIST;
```

```
PROC SORT SORTSIZE=150000  DATA=PART2;
BY STRATUM REST_TYP BDYTYP_H AGE RAN_NO;

*-------------------------------------*;
* (2)   SECOND SEAT/AGE                *;
*-------------------------------------*;
DATA PART2;   SET PART2;
BY STRATUM REST_TYP BDYTYP_H AGE;

IF FIRST.AGE AND SEAT_H1=29 THEN TRACK=29;
ELSE IF SEAT_H1^=29 THEN TRACK=SEAT_H1;

RETAIN TRACK;

IF NOT(FIRST.AGE) AND SEAT_H1=29 THEN SEAT_H2=TRACK;
ELSE SEAT_H2=SEAT_H1;

DROP TRACK;

PROC FREQ; TABLES SEAT_H2/LIST;

DATA PART2; SET PART2; DROP SEAT_H1;

PROC SORT SORTSIZE=150000  DATA=PART2;
BY STRATUM REST_TYP BDYTYP_H RAN_NO;

*-------------------------------------*;
* (3)    SECOND SEAT/BODY TYPE         *;
*-------------------------------------*;
DATA PART2;  SET PART2;
BY STRATUM REST_TYP BDYTYP_H;

IF FIRST.BDYTYP_H AND SEAT_H2=29 THEN TRACK=29;
ELSE IF SEAT_H2^=29 THEN TRACK=SEAT_H2;

RETAIN TRACK;

IF NOT(FIRST.BDYTYP_H) AND SEAT_H2=29 THEN SEAT_H3=TRACK;
ELSE SEAT_H3=SEAT_H2;

DROP TRACK;

PROC FREQ; TABLES SEAT_H3/LIST;

DATA PART2;       SET PART2; DROP SEAT_H2;

PROC SORT SORTSIZE=150000  DATA=PART2;
BY STRATUM REST_TYP RAN_NO;

*-----------------------------------------*;
* (4) SECOND SEAT/RESTRAINT SYSTEM TYPE    *;
*-----------------------------------------*;
DATA PART2; SET PART2;
BY STRATUM REST_TYP;

IF FIRST.REST_TYP AND SEAT_H3=29 THEN TRACK=29;
ELSE IF SEAT_H3^=29 THEN TRACK=SEAT_H3;

RETAIN TRACK;
```

```
IF NOT(FIRST.REST_TYP) AND SEAT_H3=29 THEN SEAT_H4=TRACK;
ELSE SEAT_H4=SEAT_H3;

DROP TRACK;

PROC FREQ; TABLES SEAT_H4/LIST;
DATA PART2; SET PART2;
SEAT_H=SEAT_H4;
DROP SEAT_H3 SEAT_H4;
IF SEAT_H=29 THEN SEAT_H=99;

 DATA ONE; SET PART1 PART2 PART3;

*-------------------------------------*;
* (1)   ALL SEAT/PERSON TYPE          *;
*-------------------------------------*;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_TYP BDYTYP_H AGE PER_TYPE RAN_NO;

DATA ONE;  SET ONE;
BY STRATUM REST_TYP BDYTYP_H AGE PER_TYPE;

IF FIRST.PER_TYPE AND SEAT_H=99 THEN TRACK=99;
ELSE IF SEAT_H^=99 THEN TRACK=SEAT_H;

RETAIN TRACK;

IF NOT(FIRST.PER_TYPE) AND SEAT_H=99 THEN SEAT_H1=TRACK;
ELSE SEAT_H1=SEAT_H;

DROP TRACK;

DATA ONE;  SET ONE; DROP SEAT_H;

PROC FREQ; TABLES SEAT_H1/LIST;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_TYP BDYTYP_H AGE RAN_NO;

*-------------------------------------*;
* (2)   ALL SEAT/AGE                  *;
*-------------------------------------*;
DATA ONE; SET ONE;
BY STRATUM REST_TYP BDYTYP_H AGE;

IF FIRST.AGE AND SEAT_H1=99 THEN TRACK=99;
ELSE IF SEAT_H1^=99 THEN TRACK=SEAT_H1;

RETAIN TRACK;

IF NOT(FIRST.AGE) AND SEAT_H1=99 THEN SEAT_H2=TRACK;
ELSE SEAT_H2=SEAT_H1;

DROP TRACK;

PROC FREQ; TABLES SEAT_H2/LIST;

DATA ONE; SET ONE; DROP SEAT_H1;

PROC SORT SORTSIZE=150000  DATA=ONE;
```

```
BY STRATUM REST_TYP BDYTYP_H RAN_NO;

*-------------------------------------*;
* (3)     ALL SEAT/BODY TYPE          *;
*-------------------------------------*;
DATA ONE;  SET ONE;
BY STRATUM REST_TYP BDYTYP_H;

IF FIRST.BDYTYP_H AND SEAT_H2=99 THEN TRACK=99;
ELSE IF SEAT_H2^=99 THEN TRACK=SEAT_H2;

RETAIN TRACK;

IF NOT(FIRST.BDYTYP_H) AND SEAT_H2=99 THEN SEAT_H3=TRACK;
ELSE SEAT_H3=SEAT_H2;

DROP TRACK;

PROC FREQ; TABLES SEAT_H3/LIST;

DATA ONE;   SET ONE; DROP SEAT_H2;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM REST_TYP RAN_NO;

*---------------------------------------*;
* (4) ALL SEAT/RESTRAINT SYSTEM TYPE    *;
*---------------------------------------*;

DATA ONE; SET ONE;
BY STRATUM REST_TYP;

IF FIRST.REST_TYP AND SEAT_H3=99 THEN TRACK=99;
ELSE IF SEAT_H3^=99 THEN TRACK=SEAT_H3;

RETAIN TRACK;

IF NOT(FIRST.REST_TYP) AND SEAT_H3=99 THEN SEAT_H4=TRACK;
ELSE SEAT_H4=SEAT_H3;

DROP TRACK;
SEAT_H=SEAT_H4 ;

PROC FREQ; TABLES SEAT_H4/LIST;

DATA SEAT_H  (KEEP=CASENUM VEHNO PERNO SEAT_H); SET ONE PART0;

IF SEAT_H=99 THEN SEAT_H=11;
PROC FREQ; TABLES SEAT_H/LIST;

*============================================================*;
*       Hotdeck Imputation of PER_ALCH in Person file        *;
**************************************************************;

DATA ONE (KEEP=CASENUM VEHNO PERNO AGE SEX PER_TYPE PER_ALCH
               STRATUM RAN_NO); SET GES91.PER91 ;
SEED=2468; RAN_NO=RANUNI(SEED);
PROC SORT SORTSIZE=150000  DATA=ONE; BY CASENUM VEHNO;

DATA TWO (KEEP= CASENUM VEHNO BDYTYP_H VIOLATN); SET GES91.VEH91_HI;
```

```
PROC SORT SORTSIZE=150000 DATA=TWO;      BY CASENUM VEHNO;

DATA ONE;  MERGE ONE (IN=A) TWO ; BY CASENUM VEHNO; IF A;

PROC SORT SORTSIZE=150000  DATA=ONE; BY CASENUM ;

DATA THREE (KEEP=CASENUM WEEKDAY HOUR ); SET GES91.ACC91_HI ;
PROC SORT SORTSIZE=150000  DATA=THREE ; BY CASENUM;

DATA ONE; MERGE ONE (IN=A) THREE ; BY CASENUM; IF A;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN AGE SEX WEEKDAY HOUR RAN_NO;

*-----------------------------------------*;
*  (1)  HOUR                              *;
*-----------------------------------------*;
DATA ONE;  SET ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN AGE SEX WEEKDAY HOUR;

IF FIRST.HOUR AND PER_ALCH=9 THEN TRACK=9;
ELSE IF PER_ALCH^=9 THEN TRACK=PER_ALCH;

RETAIN TRACK;

IF NOT(FIRST.HOUR) AND PER_ALCH=9 THEN PERAL_H1=TRACK;
ELSE PERAL_H1=PER_ALCH;

DROP TRACK;

PROC FREQ ; TABLES PERAL_H1/LIST;

DATA ONE; SET ONE; DROP PER_ALCH;

PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN AGE SEX WEEKDAY RAN_NO;

*-------------------------------------*;
* (2)  WEEKDAY                        *;
*-------------------------------------*;
DATA ONE;  SET ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN AGE SEX WEEKDAY;

IF FIRST.WEEKDAY AND PERAL_H1=9 THEN TRACK=9;
ELSE IF PERAL_H1^=9 THEN TRACK=PERAL_H1;

RETAIN TRACK;

IF NOT(FIRST.WEEKDAY) AND PERAL_H1=9 THEN PERAL_H2=TRACK;
ELSE PERAL_H2=PERAL_H1;

DROP TRACK;

DATA ONE; SET ONE; DROP PERAL_H1;

PROC FREQ ; TABLES PERAL_H2/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN AGE SEX RAN_NO;

*-------------------------------------*;
```

```
* (3)           SEX                      *;
*-----------------------------------------*;
DATA ONE;  SET ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN AGE SEX;

IF FIRST.SEX AND PERAL_H2=9 THEN TRACK=9;
ELSE IF PERAL_H2^=9 THEN TRACK=PERAL_H2;

RETAIN TRACK;

IF NOT(FIRST.SEX) AND PERAL_H2=9 THEN PERAL_H3=TRACK;
ELSE PERAL_H3=PERAL_H2;

DROP TRACK;

DATA ONE; SET ONE; DROP PERAL_H2;

PROC FREQ ; TABLES PERAL_H3/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN AGE SEX;

*-----------------------------------------*;
* (4)            AGE                      *;
*-----------------------------------------*;
DATA ONE;  SET ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN AGE;

IF FIRST.AGE AND PERAL_H3=9 THEN TRACK=9;
ELSE IF PERAL_H3^=9 THEN TRACK=PERAL_H3;

RETAIN TRACK;

IF NOT(FIRST.AGE) AND PERAL_H3=9 THEN PERAL_H4=TRACK;
ELSE PERAL_H4=PERAL_H3;

DROP TRACK;

DATA ONE; SET ONE; DROP PERAL_H3;

PROC FREQ ; TABLES PERAL_H4/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN RAN_NO;

*-----------------------------------------*;
* (5)    VIOLATIONS CHARGED               *;
*-----------------------------------------*;
DATA ONE;  SET ONE;
BY STRATUM BDYTYP_H PER_TYPE VIOLATN;

IF FIRST.VIOLATN AND PERAL_H4=9 THEN TRACK=9;
ELSE IF PERAL_H4^=9 THEN TRACK=PERAL_H4;

RETAIN TRACK;

IF NOT(FIRST.VIOLATN) AND PERAL_H4=9 THEN PERAL_H5=TRACK;
ELSE PERAL_H5=PERAL_H4;

DROP TRACK;

DATA ONE; SET ONE;  DROP PERAL_H4;
```

```
PROC FREQ ; TABLES PERAL_H5/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM BDYTYP_H PER_TYPE RAN_NO;

*--------------------------------------*;
* (6)       PERSON TYPE                *;
*--------------------------------------*;
DATA ONE;  SET ONE;
BY STRATUM BDYTYP_H PER_TYPE;

IF FIRST.PER_TYPE AND PERAL_H5=9 THEN TRACK=9;
ELSE IF PERAL_H5^=9 THEN TRACK=PERAL_H5;

RETAIN TRACK;

IF NOT(FIRST.PER_TYPE) AND PERAL_H5=9 THEN PERAL_H6=TRACK;
ELSE PERAL_H6=PERAL_H5;

DROP TRACK;

DATA ONE;   SET ONE; DROP PERAL_H5;

PROC FREQ ; TABLES PERAL_H6/LIST;
PROC SORT SORTSIZE=150000  DATA=ONE;
BY STRATUM BDYTYP_H RAN_NO;

*--------------------------------------*;
* (7)         BODY TYPE                *;
*--------------------------------------*;

DATA ONE;  SET ONE;
BY STRATUM BDYTYP_H;

IF FIRST.BDYTYP_H AND PERAL_H6=9 THEN TRACK=9;
ELSE IF PERAL_H6^=9 THEN TRACK=PERAL_H6;

RETAIN TRACK;

IF NOT(FIRST.BDYTYP_H) AND PERAL_H6=9 THEN PERAL_H7=TRACK;
ELSE PERAL_H7=PERAL_H6;

DROP TRACK;

PROC FREQ ; TABLES PERAL_H7/LIST;

DATA PERALC_H  (KEEP=CASENUM VEHNO PERNO ALCH_H);
SET ONE; ALCH_H=PERAL_H7;
IF ALCH_H=9 THEN ALCH_H=0;
PROC FREQ; TABLES ALCH_H/LIST;

* PERSON HOTDECKED IMPUTED FILES *;

PROC SORT SORTSIZE=150000  DATA=AGE_H ;       BY CASENUM VEHNO PERNO;
PROC SORT SORTSIZE=150000  DATA=SEX_H ;       BY CASENUM VEHNO PERNO;
PROC SORT SORTSIZE=150000  DATA=RESTSY_H ;    BY CASENUM VEHNO PERNO;
PROC SORT SORTSIZE=150000  DATA=INJSV_H ;     BY CASENUM VEHNO PERNO;
PROC SORT SORTSIZE=150000  DATA=SEAT_H ;      BY CASENUM VEHNO PERNO;
PROC SORT SORTSIZE=150000  DATA=PERALC_H ;    BY CASENUM VEHNO PERNO;

PROC SORT SORTSIZE=150000  DATA=PER91_I ;     BY CASENUM VEHNO PERNO;
```

```
PROC SORT SORTSIZE=150000  DATA=GES91.PER91 ; BY CASENUM VEHNO PERNO;

DATA GES91.PER91_HI;
MERGE  AGE_H  SEX_H  INJSV_H  RESTSY_H  SEAT_H
PERALC_H   PER91_I  GES91.PER91 ;   BY CASENUM VEHNO PERNO;

LABEL
      SEAT_H    ='HOTDECK IMPUTED SEATING POSITION'
      AGE_H     ='HOTDECK IMPUTED AGE'
      SEX_H     ='HOTDECK IMPUTED SEX'
      ALCH_H    ='HOTDECK IMPUTED POLICE REP. ALCOHOL INV.'
      RESTSY_H ='HOTDECK IMPUTED RESTRAINT SYSTEM USE'
      INJSEV_H ='HOTDECK IMPUTED INJURY SEVERITY'       ;

LENGTH  EJECT_I LOCATN_I AGE_H SEX_H SEAT_H ALCH_H RESTSY_H INJSEV_H 3 ;

PROC CONTENTS ; RUN;

DATA ONE; SET GES91.PER91_HI; IF AGE NE 99;
S=2; IF AGE=AGE_H THEN S=1;
PROC FREQ; TABLES S/LIST;

DATA ONE; SET GES91.PER91_HI; IF SEX NE 9;
S=2; IF SEX=SEX_H THEN S=1;
PROC FREQ; TABLES S/LIST;

DATA ONE; SET GES91.PER91_HI; IF REST_SYS NE 9;
S=2; IF REST_SYS=RESTSY_H THEN S=1;
PROC FREQ; TABLES S/LIST;

DATA ONE; SET GES91.PER91_HI; IF INJ_SEV NE 9;
S=2; IF INJ_SEV=INJSEV_H THEN S=1;
PROC FREQ; TABLES S/LIST;

DATA ONE; SET GES91.PER91_HI;
IF SEAT_POS NE 18;
IF SEAT_POS NE 19;
IF SEAT_POS NE 28;
IF SEAT_POS NE 29;
IF SEAT_POS NE 99;
S=2; IF SEAT_POS=SEAT_H THEN S=1;
PROC FREQ; TABLES S/LIST;

DATA ONE; SET GES91.PER91_HI; IF PER_ALCH NE 9;
S=2; IF PER_ALCH=ALCH_H THEN S=1;
PROC FREQ; TABLES S/LIST;
```

APPENDIX C

1991 GENERAL ESTIMATES SYSTEM VARIABLES

(* = computer assigned variables)

ON ALL LEVELS

*Case Number
*Primary Sampling Unit (PSU)
*PSU Weight
Police Jurisdiction (PJ)
*PJ Weight
*Region of the Country
*Weight
Sample Stratum
*Stratum Weight
*State

ACCIDENT LEVEL

A1 Month of Accident
A1B Year of Accident
*A1C Day of Week
A2 Hour of Accident
A2A Minute of Accident
A3 Number of Vehicles
*A3B Number of People
A4 Number of Non-Motorists
A5 Land Use
*A5A Percent Rural
A6 First Harmful Event
A7 Manner of Collision
A8 Interstate Highway
A9 Relation to Junction
A10 Relation to Roadway
A11 Trafficway Flow
A12 Number of Travel Lanes
A13 Roadway Alignment
A14 Roadway Profile
A15 Roadway Surface Condition
A16 Traffic Control Device
A18 Speed Limit
A19 Light Condition
A20 Atmospheric Condition
A21 School Bus Related
A22 Police Jurisdiction
A23 Sample Stratum
A24 Ped/Bike Accident Type
*A90 Maximum Injury Severity
*A91 Number Injured in Accident
*A92 Alcohol Involved in Accident

VEHICLE/DRIVER LEVEL

D1 Driver Presence
D2 Violations Charged
D4 Driver's Vision Obscured By
D6 Driver Maneuvered To Avoid
D7 Driver Distracted By
V1 Vehicle Number
V2 Hit and Run
V3 Vehicle Make
V4 Vehicle Model
V5 Body Type
V6 Model Year
V7 Vehicle Identification Number
*V7A Vehicle Identification Number Listed
V8 Special Use
V9 Emergency Use
V10 Number of Occupants
V11 Travel Speed
V12 Vehicle Defect
V13 Vehicle Trailing
V14 Jackknife
V15 Rollover
V16 Fire Occurrence
V18 Damage Severity
V19 Manner of Leaving Scene
V20 Most Harmful Event
V21 Vehicle Maneuver
V22 Vehicle Role
V23 Accident Type
V24 Initial Point of Impact
V25 Damage Areas
*V90 Maximum Injury Severity in Vehicle
*V91 Number Injured in Vehicle
*V92 Driver Drinking in Vehicle

PERSON LEVEL

P1 Vehicle Number
P2 Person Number
P3 Person Type
P4 Seating Position
P6 Ejection
P7 Age
P8 Sex
P9 Injury Severity
P10 Taken to Hospital or Treatment Facility
P11 Police Reported Alcohol Involvement
P13 Non-Motorist's Location
P15 Restraint System Use
P16 Restraint Type
P17 Police Reported Drug Involvement
P18 Person's Physical Impairment
P19 Non-Motorist's Action
P20 Non-Motorist Safety Equipment Use

ACCIDENT LEVEL

A1. Month of the Accident

A1B. Year of the Accident

A1C. Day of the Week

A2. Hour of the Accident

A2A. Minute of the Accident

A3. Number of Vehicles

A3B. Number of People

A4. Number of Non-Motorists

A5. Land Use

1 = Within Area of Population 25,000-50,000
2 = Within Area of Population 50,000-100,000
3 = Within Area of Population 100,000+
8 = Other Area
9 = Unknown

A5A. Percentage Rural

0 = Urban
1 = 10% of Area is Rural
2 = 20% of Area is Rural
3 = 30% of Area is Rural
4 = 40% of Area is Rural
5 = 50% of Area is Rural
6 = 60% of Area is Rural
7 = 70% of Area is Rural
9 = 90% of Area is Rural

A6. First Harmful Event

Noncollision

1 = Rollover/Overturn
2 = Fire/Explosion
3 = Immersion
4 = Gas Inhalation
5 = Jackknife
6 = Noncollision Injury (Injured in Vehicle, or Fell From Veh.)
8 = Other Noncollision
9 = Noncollision - No Details
10 = Thrown or Falling Object

Collision with Object Not Fixed

21 = Pedestrian
22 = Cycle or Cyclist (Pedalcycle or Pedalcyclist)
23 = Railway Train
24 = Animal
25 = Motor Vehicle in Transport
26 = Parked Motor Vehicle (or Other M.V. Not in Transport)
27 = Other Type Non-Motorist
28 = Other Object Not Fixed
29 = Object Not Fixed - No Details

Collision with Fixed Object

31 = Ground
32 = Building
33 = Impact Attenuator/Crash Cushion
34 = Bridge Structure (Bridge Pier/Abutment/Parapet End/Rail)
35 = Guardrail
36 = Concrete Traffic Barrier or Other Longitudinal Barrier Type
37 = Sign Post, Utility Pole or Other Support
38 = Culvert or Ditch
39 = Curb
40 = Embankment
41 = Fence
42 = Wall
43 = Fire Hydrant
44 = Shrubbery or Bush
45 = Tree
46 = Boulder
48 = Other Fixed Object
49 = Fixed Object - No Details

99 = Unknown

A7. Manner of Collision

0 = Not Collision with Motor Vehicle in Transport
1 = Rear-End
2 = Head-On
3 = Rear-to-Rear
4 = Angle
5 = Sideswipe, same direction
6 = Sideswipe, opposite direction
8 = Other
9 = Unknown

A8. Interstate Highway

0 = No
1 = Yes
9 = Unknown

A9. Relation to Junction

0 = Non-Junction
1 = Intersection
2 = Intersection Related
3 = Interchange Area
4 = Driveway, Alley Access, Etc.
5 = Entrance/Exit Ramp
6 = Rail Grade Crossing
8 = Other
9 = Unknown

A10. Relation to Roadway

1 = On Roadway
2 = On Shoulder or Parking Lane
3 = Off Roadway/Shoulder/Parking Lane
4 = On Median
8 = Other
9 = Unknown

A11. Trafficway Flow

1 = Not Physically Divided (Two Way Trafficway)
2 = Divided Highway (Median Strip, Barrier, etc.)
3 = One Way Trafficway
9 = Unknown

A12. Number of Travel Lanes

1 = One Lane
2 = Two Lanes
3 = Three Lanes
4 = Four Lanes
5 = Five Lanes
6 = Six Lanes
7 = Seven or More Lanes
9 = Unknown

A13. Roadway Alignment

1 = Straight
2 = Curve
9 = Unknown

A14. Roadway Profile

1 = Level
2 = Grade
3 = Hillcrest
8 = Other
9 = Unknown

A15. Roadway Surface Condition

1 = Dry
2 = Wet
3 = Snow or Slush
4 = Ice (Use if Snow/Ice Combined)
5 = Sand, Dirt, Oil
8 = Other
9 = Unknown

A16. Traffic Control Device

00 = No Controls

Not at Railroad Grade Crossing

Trafficway Signals

01 = Traffic Control Signal (on colors)
04 = Flashing Traffic Control Signal or Flashing Beacon
08 = Other Traffic Signal
09 = Unknown Traffic Signal

Regulatory, School Zone

21 = Stop Sign
22 = Yield Sign
23 = School Zone Related Sign
28 = Other Sign
29 = Unknown Sign

Warning Signs

40 = Advisory Speed Sign
41 = Warning Sign For Road Conditions (Hill, Steep Grade, etc.)
42 = Warning Sign For Road Construction
43 = Warning Sign For Environment/Traffic (Fog, Wind, Accident Ahead)
49 = Unknown Type Warning

Miscellaneous Not at Railroad Crossing

51 = Officer, Crossing Guard, Flagman, etc.

At Railroad Crossing

61 = Active Device at RR Crossing (e.g., Gates, Flashing Lights)
62 = Passive Device at RR Crossing (e.g., Stop Sign, Cross Bucks)

Other

97 = Traffic Control Present - No Details
98 = Other Traffic Control (whether or not at RR Grade Crossing)
99 = Unknown

A18. Speed Limit

0 = No Statutory Limit
05-65 = Actual Speed Limit (mph in increments of 5)
99 = Unknown

A19. Light Condition

1 = Daylight
2 = Dark
3 = Dark but Lighted
4 = Dawn
5 = Dusk
6 = Dawn or Dusk
9 = Unknown

A20. Atmospheric Conditions

1 = No Adverse Conditions
2 = Rain
3 = Sleet
4 = Snow
5 = Fog
6 = Rain and Fog
7 = Sleet and Fog
8 = Other (Smog, Smoke, Blowing Sand/Dust/Snow,Crosswind, Hail, etc.) 9 = Unknown

A21. School Bus Related

0 = No
1 = Yes

A22. Police Jurisdiction

A23. Stratum

A24. Pedestrian/Bicyclist Accident Type (See Table C1, page 118)

A90. Maximum Injury Severity in Vehicle

0 = No Injury
1 = Possible Injury
2 = Nonincapacitating
3 = Incapacitating
4 = Fatal
5 = Unknown Injury Severity
6 = Died Prior
8 = No Person Coded in the Accident
9 = Unknown

A91. Number Injured in Accident

00 = No Person Injured/Property Damage Only Accident
01-98 = Actual number of people injured
99 = No Person Coded

A92. Alcohol Involved in Accident

1 = Alcohol Involved
2 = No Alcohol Involved
8 = No Person Coded
9 = Unknown

VEHICLE/DRIVER LEVEL

D1. Driver Presence

0 = Unattended Vehicle (Driverless, or No Driver Involved)
1 = Driver Operated Vehicle
2 = Hit and Run (Vehicle and/or Driver Left Scene)
9 = Unknown Driver Presence

D2. Violations Charged

0 = None
1 = Alcohol or Drugs
2 = Speeding
3 = Alcohol or Drugs and Speeding
4 = Reckless Driving
5 = Driving With a Suspended or Revoked License
6 = Failure to Yield Right-of-Way
7 = Running a Traffic Signal or Stop Sign
50 = Hit & Run (and No Information)
97 = Violation Charged - No Details
98 = Other Violation
99 = Unknown if Charged

D4. Driver's Vision Obscured By

00 = No Obstruction
01 = Rain, Snow, Fog, Smoke, Sand, Dust
02 = Reflected Glare, Bright Sunlight, Headlights
03 = Curve, or Hill
04 = Building, Billboard, or Other Design Features (Includes Signs, Embankment)
05 = Trees, Crops, Vegetation
06 = Moving Vehicle (including load)
07 = Parked Vehicle
08 = Splash or Spray of Passing Vehicle
09 = Inadequate Defrost or Defog System
10 = Inadequate Lighting System
11 = Obstruction Interior to the Vehicle
12 = External Mirrors
13 = Head Restraints
14 = Broken or Improperly Cleaned Windshield
50 = Hit-and-Run (and No Information)
97 = Vision Obscured - No Details
98 = Other Obstruction
99 = Unknown Whether Vision was Obscured

D6. Driver Maneuvered To Avoid

00 = Driver did not maneuver to avoid
01 = Object in Road
02 = Poor Road Conditions (puddle, ice, pot hole, etc.)
03 = Animal in Road
04 = Vehicle in Road
05 = Pedestrian, Pedalcyclist, or other Non-motorist in Road
50 = Hit & Run (and no information)
97 = Avoidance Maneuver - No Details
99 = Unknown if Driver Maneuvered to Avoid

D7. Driver Distracted By

00 = Not Distracted or N/A
01 = Passengers, Occupants
02 = Vehicle Instrument Display (Radio, Cassette, CB, Heating/AC)
03 = Phone
04 = Other Internal Distractions
05 = Other Accident ("Rubbernecking")
06 = Other External Distractions
50 = Hit & Run (and no information)
97 = Distractions - No Details
99 = Unknown if Distracted

V1. Vehicle Number

V2. Hit & Run

0 = No, Did Not Leave Scene
1 = Yes, Driver or Car and Driver Left Scene
9 = Unknown

V3. Vehicle Make

V4. Vehicle Model

V5. Body Type (See Table C2, page 121)

V6. Model Year

40 = All vehicles manufactured for 1940 model year and before.
41-91 = (Actual Value)
99 = Unknown

V7. Vehicle Identification Number (VIN)

V8. Special Use

0 = No Special Use
1 = Taxi
2 = Vehicle Used as School Bus
3 = Vehicle Used as Other Bus
4 = Military
5 = Police
6 = Ambulance
7 = Firetruck
8 = Farm or Construction Equipment
9 = Unknown

V9. Emergency Use

0 = No
1 = Yes
9 = Unknown

V10. Number of Occupants

00-30 = Actual Value if Total Known
30 = 30 or More
97 = Unknown - Only Injured Reported
99 = Unknown

V11. Travel Speed

Actual Miles Per Hour

00 = Stopped Vehicle
01-96 = Actual Travel Speed (MPH)
97 = Ninety-Seven MPH or Greater
99 = Unknown

V12. Vehicle Defects

00 = No Vehicle Defects
01 = Tires
02 = Brake System
03 = Steering System - Tie Rod, Kingpin, Ball Joint, etc.
04 = Suspension - Springs, Shock Absorbers, Macpherson Struts, Control Arms, etc.
05 = Power Train - Universal Joint, Drive Shaft, Transmission, etc.
06 = Exhaust System
07 = Headlights
08 = Signal Lights
09 = Other Lights
10 = Wipers
11 = Wheels
12 = Mirrors
13 = Driver Seating and Control
14 = Body, Doors
15 = Trailer Hitch

50 = Hit-and-Run Vehicle
97 = Vehicle Defects - No Details
98 = Other Vehicle Defects
99 = Unknown if Vehicle Has Defects

V13. Vehicle Trailing

0 = No
1 = Yes, One Trailing Unit
2 = Yes, Two Trailing Units
3 = Yes, Three or More Trailing Units
4 = Yes, Number of Trailing Units Unknown
9 = Unknown

V14. Jackknife

0 = No Jackknife Noted on PAR
1 = Jackknife Occurred

V15. Rollover

0 = No Rollover Noted on PAR
1 = Rollover Occurred

V16. Fire Occurrence

0 = No Fire Noted on PAR
1 = Fire Occurred in Vehicle

V18. Damage Severity

0 = None
1 = Minor
2 = Functional (Moderate)
3 = Disabling (Severe)
9 = Unknown

V19. Manner of Leaving Scene

1 = Driven Away
2 = Towed Due to Damage
3 = Towed Not Due to Damage
4 = Abandoned
9 = Unknown if Towed

V20. Most Harmful Event (See First Harmful Event-A6)

V21. Vehicle Maneuver

01 = Going Straight
02 = Slowing or Stopping in Traffic Lane
03 = Starting in Traffic Lane
04 = Stopped in Traffic Lane
05 = Passing or Overtaking Another Vehicle
06 = Leaving a Parked Position
07 = Parked
08 = Entering a Parked Position
09 = Maneuvering to Avoid an Animal, Ped., Object, Vehicle, etc.
10 = Turning Right
11 = Turning Left
12 = Making U-turn
13 = Backing Up (other than parking)
14 = Changing Lanes or Merging
15 = Negotiating a Curve
98 = Other
99 = Unknown

V22. Vehicle Role

0 = Non-Collision
1 = Striking
2 = Struck
3 = Both
9 = Unknown

V23. Accident Type (See Table C3, page 125)

V24. Initial Point of Impact

0 = No Damage/Non-collision
1 = Front
2 = Rightside
3 = Leftside
4 = Back
5 = Top
6 = Undercarriage
7 = Corner
9 = Initial Point of Impact Unknown

V25. Damage Areas (up to five can be coded)

0 = No Damage
1 = Front
2 = Rightside
3 = Leftside
4 = Back
5 = Top
6 = Undercarriage
7 = All Areas Damaged
9 = Damaged Areas Unknown

V90. Maximum Injury Severity in Vehicle

0 = No Injury
1 = Possible Injury
2 = Nonincapacitating Injury
3 = Incapacitating Injury
4 = Fatal Injury
5 = Injured Severity Unknown
6 = Died Prior
8 = No Person Coded
9 = Unknown

V91. Number Injured in Vehicle

V92. Driver Drinking in Vehicle

1 = Alcohol Involved
2 = No Alcohol
8 = No Driver Coded
9 = Unknown

PERSON LEVEL

P1. Vehicle Number

P2. Person Number

P3. Person Type

Motorists

1 = Driver of a Motor Vehicle in Transport
2 = Passenger of a Motor Vehicle in Transport

Non-Motorists - Occupant

3 = Occupant of a Motor Vehicle Not in Transport
4 = Occupant of a Non-Motor Vehicle Transport Device

Non-Motorists - Non-Occupant

5 = Pedestrian
6 = Cyclist (Pedalcyclist)
8 = Other or Unknown Non-Occupant
9 = Unknown Occupant Type in a Motor Vehicle in Transport

P4. Seating Position

00 = Non-motorist
11 = Front Seat - Left Side (Driver's Side)
12 = Front Seat - Middle
13 = Front Seat - Right Side
18 = Front Seat - Other
19 = Front Seat - Unknown
21 = Second Seat - Left Side
22 = Second Seat - Middle
23 = Second Seat - Right Side
28 = Second Seat - Other
29 = Second Seat - Unknown
30 = Sleeper Section of Cab (Truck)
40 = Other Passenger in Passenger or Cargo Area
50 = Trailing Unit
60 = Riding on Vehicle Exterior
99 = Unknown Seating Position

P6. Ejection

0 = Not Ejected
1 = Ejected (Partial or Total)
9 = Unknown

P7. Age

00 = Up to One Year
01-96 = Actual Age
97 = 97 Years or Older
99 = Unknown

P8. Sex

1 = Male
2 = Female
9 = Unknown

P9. Injury Severity

0 = No Injury (O)
1 = Possible Injury (C)
2 = Nonincapacitating Injury (B)
3 = Incapacitating Injury (A)
4 = Fatal Injury (K)
5 = Injured, Severity Unknown (U)
6 = Died Prior to Accident
9 = Unknown if Injured

P10. Taken to Hospital or Treatment Facility

0 = No
1 = Yes
9 = Unknown

P11. Police Reported Alcohol Involvement

0 = Alcohol Not Involved or N/A
1 = Alcohol Involved
7 = Alcohol and/or Drugs Involved
8 = Not Reported
9 = Unknown (Police Reported)

P13. Non-Motorist Location

00 = Not Applicable - Driver or Occupant of M.V. in Transport
01 = Intersection - In Crosswalk
02 = Intersection - On Roadway
08 = Intersection - Other
09 = Intersection - Unknown Location
11 = Non-Intersection - In Crosswalk
12 = Non-Intersection - On Roadway
18 = Non-Intersection - Other
19 = Non-Intersection - Unknown Location
20 = In Crosswalk - Unknown if Intersection
98 = Other Location
99 = Unknown Location

P15. Restraint System Use

0 = None used or N/A
1 = Lap/Shoulder Belt
2 = Lap Belt
3 = Shoulder Belt
4 = Air Bag Deployed
5 = Air Bag Deployed and Lap/Shoulder Belt
6 = Child Safety Seat
7 = Motorcycle Helmet
8 = Restraint Used-Specific Unknown or Other
9 = Unknown if Used

P16. Restraint Type

0 = None or N/A
1 = Automatic (Passive)
2 = Manual (Active)
9 = Unknown Type or Use

P17. Police Reported Drug Involvemnet

0 = Drugs Not Involved or N/A
1 = Drugs Involved
7 = Drugs and/or Alcohol Involved
8 = Not Reported
9 = Unknown (Police Reported)

P18. Person's Physical Impairment

00 = None
01 = Ill, Blackout
02 = Drowsy, Sleepy, Fell Asleep, Fatigued
03 = Requires Cane or Crutches
04 = Paraplegic or Restricted to Wheelchair
05 = Impaired Due to Previous Injury
06 = Deaf
07 = Blind
97 = Physical Impairment - No Details
98 = Other Physical Impairment
99 = Unknown if Physically Impaired

TABLE C1: A24 PEDESTRIAN/CYCLIST ACCIDENT TYPE

0 = No pedestrian/cyclist involved

Bicyclist Rides out from a Driveway, Alley, or Other Midblock Location

1 = Cyclist fails to yield to motorist at a residential driveway or alley; pre-crash path perpendicular to roadway.
2 = Cyclist fails to yield to motorist at a commercial driveway or alley; pre-crash path perpendicular to roadway.
3 = Cyclist turns or merges into the path of motorist from a residential driveway or alley; pre-crash path parallel to roadway.
4 = Cyclist fails to yield to motorist at a midblock location: entry is over curb or shoulder.

Bicyclist Rides out from a Controlled Intersection

5 = Cyclist failed to yield to motorist at am intersection controlled by a stop sign or a flashing red signal.
6 = Cyclist fails to clear intersection controlled by signal before light turns green for cross traffic; motorists' view of cyclist was not obstructed.
7 = Cyclist fails to clear intersection controlled by signal before light turns green for cross traffic; motorists' view of cyclist was obstructed by standing traffic.

Motorist Turns or Drives out in Front of Bicyclist

8 = Motorist exiting from driveway, alley, or other midblock location fails to yield to cyclist.
9 = At an intersection controlled by a stop sign or flashing red light, motorist obeys the sign but fails to yield to cyclist.
10= At an intersection controlled by a signal, motorist obeys signal but fails to yield to cyclist while making right turn on red.
11= Motorist backing from driveway fails to yield to cyclist.
12= Motorist fails to stop at an intersection controlled by a stop sign.

Motorist Overtakes Bicyclist

13= Motorist fails to detect cyclist he/she is overtaking.
14= Motorist loses control of vehicle while overtaking cyclist; in some cases motorist is in uncontrolled slide or spin, but more often, merely loses precise control and veers too far to right.
15= The motorist and the cyclist counteract each other's evasive action.
16= Motorist misjudges space required to pass cyclist.
17= Cyclist's path is obstructed, causing cyclist to strike obstruction or overtaking motorist.

Bicyclist Makes Unexpected Turn or Swerve

18= Cyclist turns left in front of motorist proceeding in the same direction.
19= Cyclist turns left in front of motorist approaching from straight ahead.
20= Cyclist loses control and swerves into the path of a motorist proceeding in the same direction.
21= Cyclist riding on wrong side of street makes right turn in path of approaching motorist.

Motorist Make Unexpected Turn

22= Motorist make left turn in front of cyclist proceeding in the same direction; in some cases cyclist was riding on wrong side of street.
23= Motorist make left turn in front of cyclist approaching from straight ahead.
24= Motorist makes right turn in front of cyclist proceeding in a parallel path; bicyclist either proceeding in same direction or from opposite direction (riding on the wrong side of the street).

Other/Infrequent

25= Vehicles collide at uncontrolled intersection: crossing paths.
26= Vehicles collide head-on: wrong-way bicyclist.
27= Bicyclist overtaking motor vehicle.
28= Vehicles collide head-on; wrong-way motorist.
29= Parking lot, other open area:crossing paths.
30= Vehicles collide head-on: counteractive evasive action.
31= Bicyclist cuts corner when turning left: crossing paths.
32= Bicyclist swing wide when turning right: crossing paths.
33= Motorist cuts corner when turning left: crossing paths.
34= Motorist swing wide when turning right: crossing paths.
35= Motorist driveout from on-street parking.
36= Weird.
37= Insufficient information to classify.
39= Motorist overtaking (Cyclist)
40= Play vehicle (Big wheel, other tricycle, or bicyclist with training wheels).
41= Cyclist struck parked vehicle.
48= Drive out - Intersection (Motorist drove out into or in front of cyclist).
49= Ride out - intersection (Bicyclist)
55= Controlled intersection - other.
98= Paralleled path - unknown.
99= Intersecting path - unknown.

Pedestrian Accident Types

110= Commercial Bus
120= School Bus
130= Vender
140= Mailbox Related
150= Entering/Exiting

210= Driverless Vehicle
220= Backing
230= Hot Pursuit

310= To/from Disabled Vehicle
320= Disabled Vehicle Related
330= Emergency Vehicle Related

410= Working on Roadway
420= Play Vehicle Related
430= Playing in Roadway

510= Hitchhiking
520= Expressway Crossing
531= Walking along Roadway with Traffic
532= Walking along Roadway against Traffic
539= Walking along Roadway Can't Specify

610= Waiting to Cross
620= Not in Roadway

710= Multiple Threat, Intersection
720= Vehicle Turn/Merge
730= Intersection Dash
740= Trapped
750= Pedestrian Walked into Vehicle, Intersection
760= Driver Violation
790= Intersection - other
810= Multiple Threat, Midblock
821= Midblock Dart-out, First half
822= Midblock Dart-out, Second half
829= Midblock Dart-out, Can't specify
830= Midblock dash
840= Pedestrian Walked into Vehicle, Midblock
890= Midblock - other
910= Other - weird
920= Inadequate information

TABLE C2: V5 BODY TYPE

Automobiles

01= Convertible (excludes sun-roof, t-bar)
02= 2-door sedan, hardtop, coupe
03= 3-door/2-door hatchback
04= 4-door sedan, hardtop
05= 5-door/4-door hatchback
06= Station wagon (excluding van and truck based)
07= Hatchback, number of doors unknown
08= Other automobile type
09= Unknown automobile type

Automobile Derivatives

10= Auto based pickup
11= Ambulance
12= Hearse
13= Limousine

Utility Vehicles

14= Utility vehicles

Van Based Light Trucks (< 10,000 lbs GVWR)

20= Minivan
21= Standard Van
22= Step-van or Walk-in van
28= Other van type
29= Unknown van type

Light Conventional Trucks (Pickup style cab, < 10,000 lbs GVWR)

30= Compact pickup
31= Standard pickup
32= Pickup with slide-in camper
39= Unknown (pickup style) light conventional truck

Other Light Trucks (< 10,000 lbs GVWR)

40= Cab chassis based (includes rescue vehicle, light stake, dump, and tow truck)
41= Truck based panel
42= Light truck based motorhome (chassis mounted)
48= Unknown light truck type (pickup, van or other)
49= Unknown light vehicle type (automobile, van, or light truck)

Buses
50= School bus type (designed to carry students, not cross country or transit)
58= Other bus (e.g., transit, intercity, bus based motorhome)
59= Unknown bus type

Medium/Heavy Trucks (>10,000 lbs GVWR)

60= Single unit straight truck
63= Medium/heavy truck based motorhome
65= Truck-tractor (Cab Only, or with any number of trailing units; any weight)
68= Unknown medium/heavy truck type
69= Unknown truck type (light/medium/heavy)

Motored Cycles (Does not include all terrain vehicles/cycles)

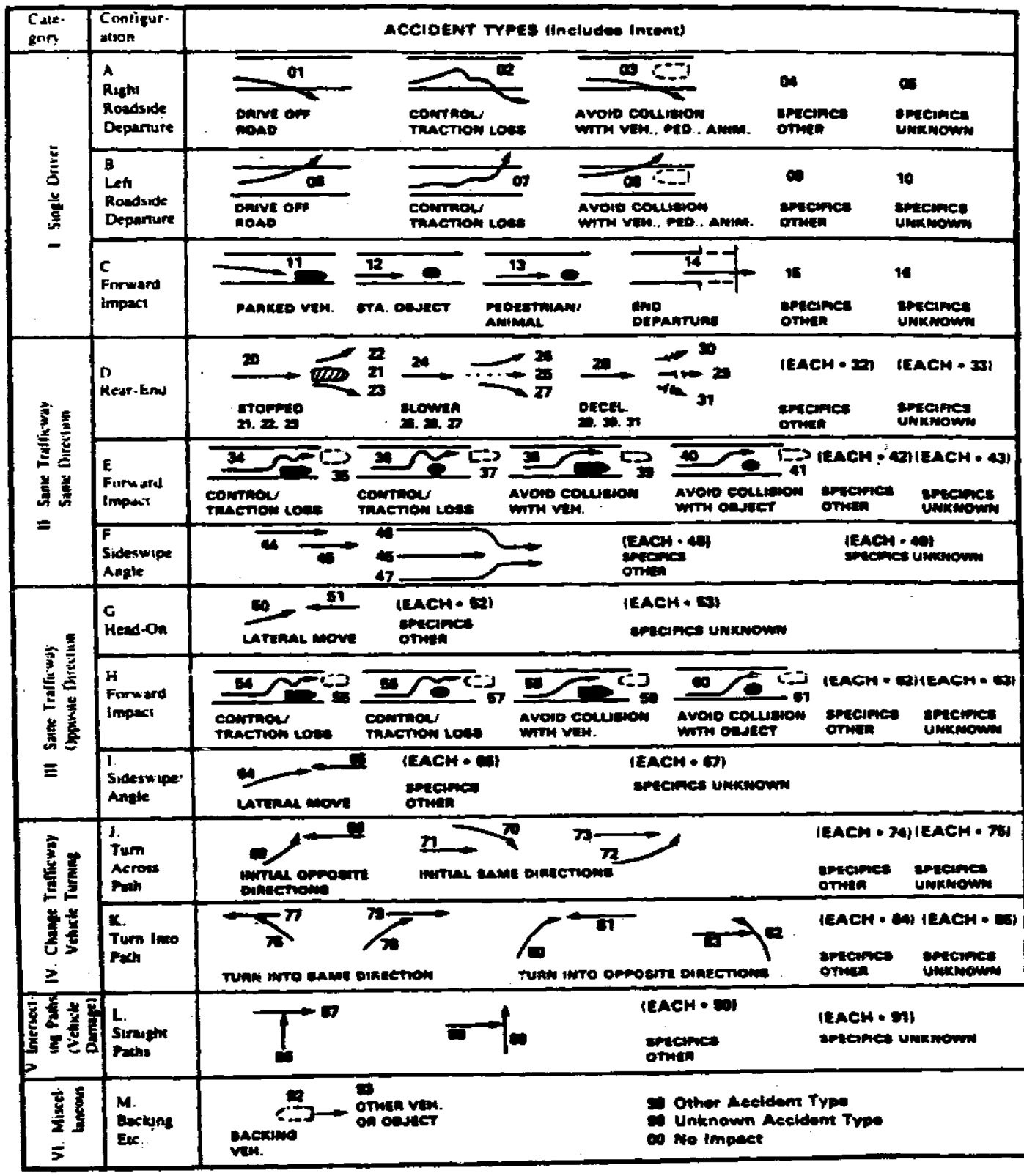
70= Motorcycle
71= Moped (motorized bicycle)
72= Three wheeled motorcycle or moped
78= Other motored cycle type (minibike, motorscooter)
79= Unknown motored cycle type

Other Vehicles

80= ATV (all terrain vehicle including dune/swamp buggy) and ATC (all terrain cycle)
81= Snowmobile
82= Farm equipment other than trucks
83= Construction equipment other than trucks (includes graders)
88= Other type vehicle (includes go-cart, fork lift, city street sweeper)
89= Unknown other vehicle

99= Unknown body type

TABLE C3: V23 ACCIDENT TYPE



*U.S. G.P.O.:1993-343-273:80182