MODELING OF 85TH PERCENTILE SPEED FOR RURAL HIGHWAYS FOR ENHANCED TRAFFIC SAFETY
ANNUAL REPORT FOR FY 2009

(ODOT SPR ITEM No. 2211)

Submitted to:

Ginger McGovern, P.E.
Planning and Research Division Engineer
Oklahoma Department of Transportation
200 N.E. 21st Street
Oklahoma City, Oklahoma 73105

Submitted by:

Dharamveer Singh
Musharraf M. Zaman
School of Civil Engineering and Environmental Science
The University of Oklahoma
Norman, Oklahoma, 73019

and

Luther White
Department of Mathematics
The University of Oklahoma
Norman, Oklahoma, 73019

October, 2009

TABLE OF CONTENTS

Pages

1. INTRODUCTION 4

2. BACKGROUND AND PREVIOUS WORKS 4

3. PROJECT PROGRESS 5

3.1 Task 1: Selection of Model Parameters 5

3.2 Task 2: Updating Previous Data 6

4. DATA ANALYSIS 7

4.1 Development of Histograms 7

4.2 Neural Network Modeling 8

5. PROPOSED ACTIVITIES FOR YEAR 2 8

6. REFERENCES 9

APPENDIX A 11

APPENDIX B 24

LIST OF FIGURES

Pages

Figure 1 Geographical Distribution of Control Sections 13
Figure 2 Distribution of Surface Width 14
Figure 3 Distribution of IRI 14
Figure 4 Distribution of ADT 15
Figure 5 Distribution of Shoulder Type 15
Figure 6 Distribution of Shoulder Width 16
Figure 7 Distribution of Skid Number 16
Figure 8 Distribution of Posted Speed 17
Figure 9 Distribution of No. Fatal Collision 17
Figure 10 Distribution of No. of Injury Collision 18
Figure 11 Distribution of No. of Property Damage 18
Figure 12 Distribution of No. of Total Collision 19
Figure 13 Distribution of Location Fatal Rate 19
Figure 14 Distribution of Location Injury Rate 20
Figure 15 Distribution of Location Overall Collision Rate 20
Figure 16 Distribution of Statewide Overall Collision Rate 21
Figure 17 Distribution of % Driver Unsafe Speed 21

LIST OF TABLES

Table 1 Maximum, minimum, and mean values of each variables 12
Table 2 The New Data Set 22

1. INTRODUCTION

Traffic operations on two-lane rural highways and setting realistic speed limits are some of the difficult tasks faced by the Oklahoma Department of Transportation (ODOT). For such highways, overtaking slower vehicles is possible only by the use of the opposing lane where sight distance and gap in the opposing traffic stream play a key role. While many states, including Oklahoma, use the 85th percentile speed as a major factor in determining posted speed for rural highways, other factors such as pavement width, type and width of shoulder, topography, weather, roadside development, and accident experience also play an important role in determining posted speeds. In recent years, neural network models have been used successfully for many engineering problems, including modeling of 85th percentile speeds in rural highways in Kansas, for example (Najjar, et al., 1999, 2002). Similar models are needed for Oklahoma for enhanced traffic safety on rural highways in the state. A neural network model based on appropriate pavement and traffic data can be an effective tool for ODOT to enhance traffic safety in the state. Research is needed to develop such a model to analyze its effectiveness and to assess data and information requirement for its implementation. In the present study such a model is being developed in close cooperation with ODOT. The current study is based on an earlier study conducted by the Principal Investigator on this topic about a decade ago (Issa, et al., 1998; Zaman, et al., 2000).

2. BACKGROUND AND PREVIOUS WORKS

In a previous project conducted by Zaman et al., (2000), a neural network (NN) model was developed to include highway parameters in predicting the 85th percentile speed on two-lane rural highways. Data from 121 sites, distributed throughout Oklahoma, were used in that study (Issa, et al., 1998). The following parameters were considered in developing the NN model: percentage passing zone (%passing), average daily traffic (ADT), international roughness index (IRI), pavement serviceability index (PSI), surface width (SW), shoulder type (ST), and shoulder width (SHW). Results from that project indicate that the developed Neural Network model might have suffered from over fitting. This phenomenon sometimes occurs as a result of inadequate data sets and the lack of sufficient distribution of data for the entire range of a model parameter. Nevertheless, the previous model developed by the University of Oklahoma provides an important first step in realizing the objective of developing a tool for the setting of 85th percentile speed on two-lane rural highways in Oklahoma.

3. PROJECT PROGRESS

Task 1: Selection of Model Parameters has been completed and significant progress has been made on Task 2: Updating Previous Data set. The background information for neural network modeling has been gathered and studied to use in Year 2. The work progress has been discussed in detail in the paragraphs below.

3.1 Task 1: Selection of Model Parameters

The various model parameters are selected in the initial stage of modeling to judge their suitability. The model parameters are divided in four different groups: physical characteristics of road, pavement condition, traffic parameter, and accident data. The parameters for each group are listed below.

(i) Physical Characteristics of Road

- Surface Width (SW)
- Shoulder Width (SHW)
- Shoulder Type (ST)

(ii) Pavement Condition

- Skid Number (SN)
- International Roughness Index (IRI)

(iii) Traffic Parameter

- Average Daily Traffic (ADT)
- Posted Speed (PS)

(iv) Accident Data

- No. of Fatal Collision
- No. of Injury Collision
- No. of Property Damage
- Total Collision
- Location Fatal Collision Rate
- Location Injury Collision Rate
- Location Overall Collision Rate
- Statewide Overall Collision Rate
- % Driver Unsafe Speed

The maximum, minimum and average values of each parameter are given in Table 1. The physical characteristics of road covers surface width, shoulder width, and shoulder type. The maximum and minimum surface widths for two-lane rural highways were found to be 24 and18 feet, respectively. The shoulder types included in the new data set are: Type 1 (Paved), Type 2 (Gravel), and Type 3 (Sod). However, It is decided to include other shoulder types: Type 4 (Curb both sides), Type 5 (Curb one side), and

Type 6 (Combination of paved and Sod) in the future work. Comparatively, the previous study included only three types (1, 2 and 3). The shoulder width ranges from 1 to 10 feet, with average value of 6 feet.

The pavement condition parameter includes SN and IRI. Since two directional skid numbers are available at particular control section, it was suggested by ODOT that primary direction skid number be used in modeling (primary direction being the direction along which a given section ends). In the old data set both PSI (Present Serviceability Index) and IRI were included as model parameters; while, in the new data set it was decided to drop PSI, and to use IRI only, as it is more reflective of ODOT's current practice. The maximum and minimum values of SN were 62 and 22, respectively, while the IRI values in the new dataset range from 46 to 121 in/mile.

The traffic parameter covers ADT and posted speed. The ADT for all the sub-sections were taken from the Need and Sufficiency Report, while the posted speed data were gathered from the Highway Performance monitoring System (HPMS) data file. The ADT values of two-lane rural highways range from 210 to 8,000, with an average of 3,099. The maximum and minimum posted speeds were found to be 70 mph and 35 mph, respectively.

The ODOT has provided the research team access to accident data by using a particular user and password. The accident data include, number of collision (fatal, injury and property damage), collision rate of the location, and state wide collision rate for the similar control section. It was discussed in semiannual meeting with ODOT to include the following parameters: total crashes, location crash rate and statewide crash rate. Truck and unsafe speed crash rate are not included in the data set because the percentages of trucks are insignificant on two-lane rural highways, as per discussions at the meeting.

3.2 Task 2: Updating Previous Data

Since the data from the 121 sites used in developing the previous model have changed over the past ten years, there is a need to update them. As an initial step, the research team talked to the Traffic Division and the Planning and Research Division about the sources of data we plan to use initially. A majority of the data (e.g., those available in the Need and Sufficiency Rating Manual (ODOT, 2007), accident data, Highway Performance Monitoring System (HPMS), Skid Resistance, etc.) are now available electronically. This makes the data collection process much easier.

Updating of the old data set was done by matching the physical characteristics of the roads (surface width, shoulder width, and shoulder type).The Need and Sufficiency

Rating Manual was used to categorize the sub-control sections, having similar physical characteristics of the road as recorded in the initial study. For example, if a sub-control section of a county in the old data set has surface width 24 feet, shoulder type 1, and shoulder with 8 feet, similar features were searched for the selected county and the same control section to identify the desired sub-section. A total of 74 sub-sections have been updated, so far, out of the 121 sites in the previous study. Some of the sites from the previous study could not be updated because either they were nonexistent (e.g., a two-lane highway has become a multilane highway). Such sites are not included in the new data set. In this report, these 74 sub-sections are called the new data set. The new data set covers 33 counties; the geographical distribution of these sub-sections is shown in Figure1. It is evident from Figure 1 that the distribution of these sub-sections is not even and does not include many counties in the state. The work planned in Year 2 would remedy this shortcoming by including additional counties.

Once the sub-sections are updated, the next job was to select their starting and end points. The starting and end point mileage for all subsections in the new data set were taken from the HPMS data file. Two directional skid numbers are available at a particular sub-section. It was suggested at the 15th June meeting with the ODOT Research Panel that only skid numbers for the primary direction be included in the nural network model. The primary direction skid number is the average of the skid values for that sub-section. The posted speed of each sub-section was taken from HPMS data file.

The accident data in Table 2 (see Appendix A) include the following: number of collision rates (fatal, injury and property damage), collision rate (fatal and injury), and number of drivers with unsafe safe speed. The location collision rate and state wide collision rate are also reported.

4. DATA ANALYSIS

4.1 Development of Histograms

The histogram plots were made for the new data set. These plots were constructed for average daily traffic, shoulder width, shoulder type, surface width, international roughness index, skid value and posted speed, and accident data. These plots provide a visual distribution of data for each parameter and are found to be useful in identifying the nature of additional sub-sections to be included in the data set. Appendix-A includes the histogram of all model parameters.

4.2 Neural Network Modeling

The objective is to develop a neural network model that expresses the 85th percentile speed as a function of selected parameters. The following parameters, also called “independent variables” in this section, are considered in developing the neural network model.

- Surface width (SW)
- Shoulder type (ST)
- Shoulder width (SHW)
- International Roughness Index (IRI)
- Skid Number (SN)
- Average daily traffic (ADT)
- Posted speed (PS)
- Crash rate of all vehicles
- Unsafe speed % driver
- Total crashes
- Location overall collision rate

A brief description of the model is given in Appendix B, in a general form.

5. PROPOSED ACTIVITIES FOR YEAR 2

Year 2 of this project will include adding more sub-sections to the existing data base, and to analyze their suitability for the modeling. An overview of the tasks is given below.

Task 3: Scaling of Input Data

Scaling of input data has been found to significantly influence the predictive capability of a neural network model. Scaling of variables is done in such a way that the interval of all input parameters is constant (e.g., +1 to -1). This facilitates the analysis of the sensitivity of outputs to different factors and is an important improvement to the model. This task will be pursued in Year 2.

Task 4: Principal Component Analysis

Conducting a principal component analysis (PCA) of the data will help identify data redundancies. It will also help express data with respect to an orthogonal basis in the data space. Thus, a PCA in the preprocessing of data results in a more efficient representation of the input data, and will be pursued in Year 2.

Task 5: Improved NN Model Development

In the previous project the well-known back propagation algorithm was used to determine the NN weights. More recently the Levenberg Marquart method (Tarefder, et al., 2005) has been applied for more efficient training of NN models. Using a more efficient tool will allow us to develop a better model and provide us ability to analyze the network's sensitivity with respect to perturbations of parameters.

Task 6: Probabilistic Analysis of NN

In other problems we have found it useful to randomly generate initialization variables (for training the NN model) and view the resulting network outputs as random variables defined over the sample space of the initialization parameters. Thus, we may accumulate results from multiple runs to approximate the probabilities that are valuable in assessing system behavior and determining the value added from the inclusion of additional data.

Task 7: New Data and Model Refinement

Based on the aforementioned tasks, the neural network model will be refined using the revised data sets to enhance the model accuracy and robustness.

Task 8: Workshop

A workshop will be organized in collaboration with ODOT, as noted in the proposal.

6. REFERENCES

Issa, R., and Zaman, M., (1998), "Modeling the 85th Percentile Speed on Oklahoma Two-Lane Rural Highways: A Neural Network Approach," *Final Report* (Project No. 6007; ORA 125-5227) submitted to ODOT.

Najjar, Y., Stokes, R., and Russell, E. (1999), "Using Roadway Characteristics to Set Reasonable Speed Limits on Kansas Two-Lane Rural Roads," *Intelligent Engineering System through Artificial Neural Network,* Vol. 9, pp. 833-838.

Najjar, Y., Russell, E., Stokes, R., and Ghassan, A. L. (2002), "New Speed Limits on Kansas Highways: Impact on Crashes and Fatalities," *Journal of the Transportation Research Forum,* Vol. 56, No.4, pp.119-147.

Tarefder, R., White, L. and Zaman, M. (2005), “Neural Network Model for Asphalt Concrete Permeability,” *Journal of Materials in Civil Engineering,* ASCE, Vol. 17, No. 1, pp.19-27.

Zaman, M., Issa, R., and Najjar, Y., (2000),“Modeling the 85th Percentile Speed on Oklahoma Two-Lane Rural Highways via Neural Network Approach,” *Computational Intelligence Applications in Pavement and Geotechnical Systems*, Attoh-Okine (Editor), pp. 99-108

APPENDIX A

Table 1 Maximum, minimum, and mean values of each variables

Variable	Maximum Value	Minimum Value	Mean Value
Surface Width	24	18	24
IRI (in/mile)	279	46	116
Average Daily Traffic	8000	210	3099
Shoulder Type	6	1	2
Shoulder Width	10	1	6
Skid Number	62	22	43
No. of Fatal Collision	5	0	1
No. of Inury Collision	101	0	11
No. of Property Damage	132	0	11
Total Collision	236	0	23
Location Fatal Rate	58	0	5
Location Injury Rate	488	0	73
Location Overall Collision Rate	743	0	150
Statewide Overall Collision Rate	189	86	120
% Unsafe Speed Drivers	100	0	21
Posted Speed	70	35	54

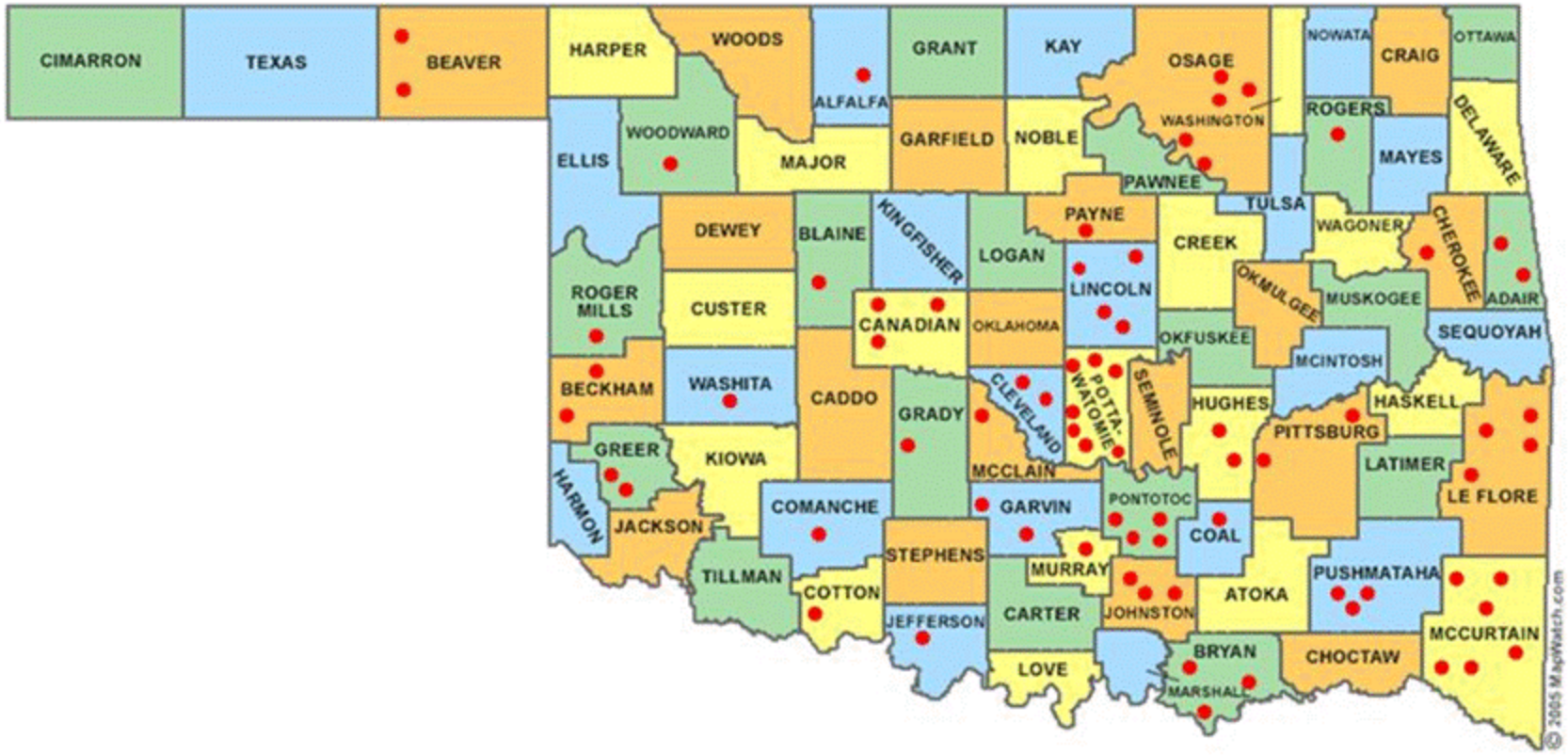

Figure 1 Geographical Distribution of Control Sections

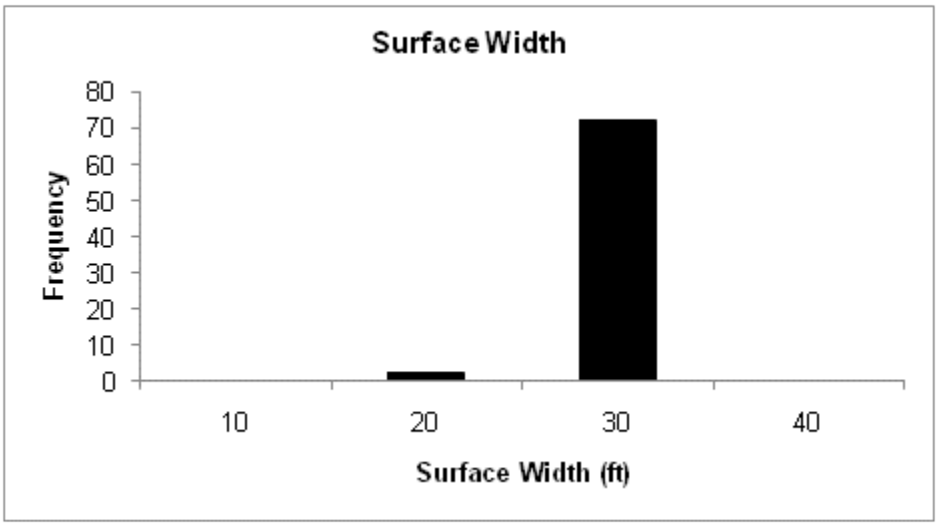

Figure 2 Distribution of Surface Width

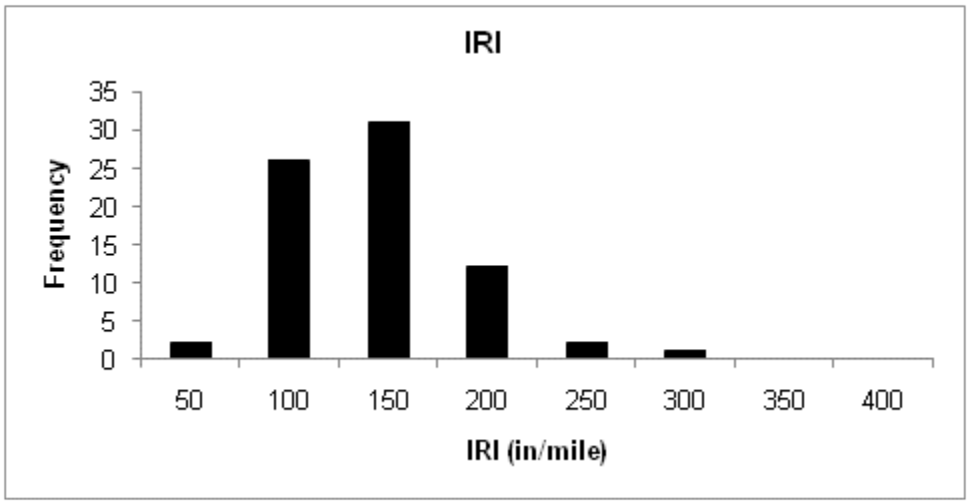

Figure 3 Distribution of IRI

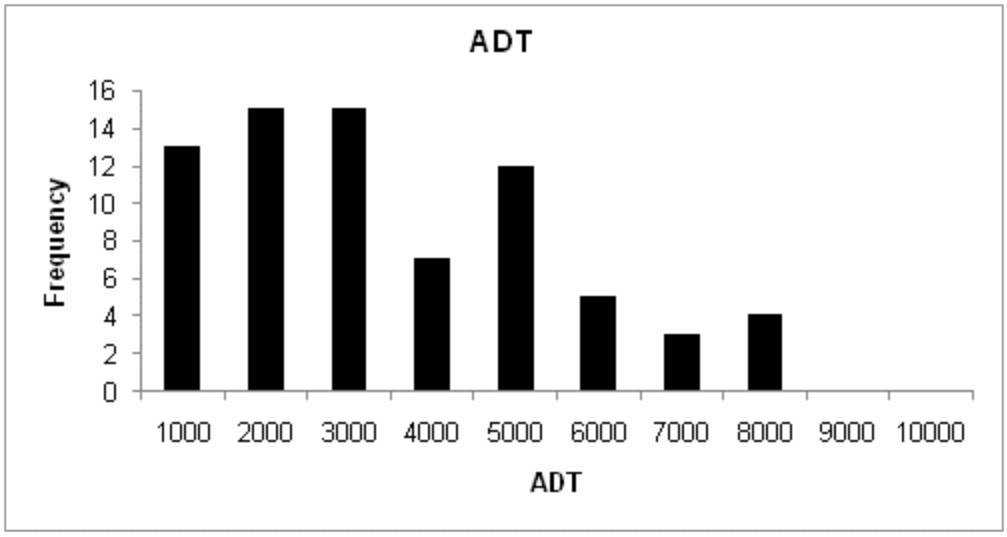

Figure 4 Distribution of ADT

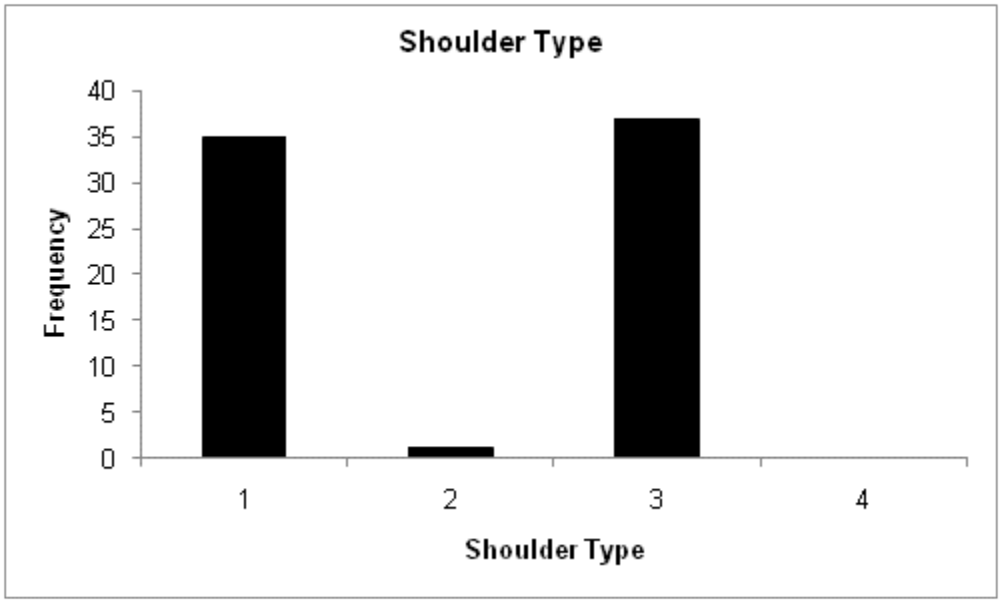

Figure 5 Distribution of Shoulder Type

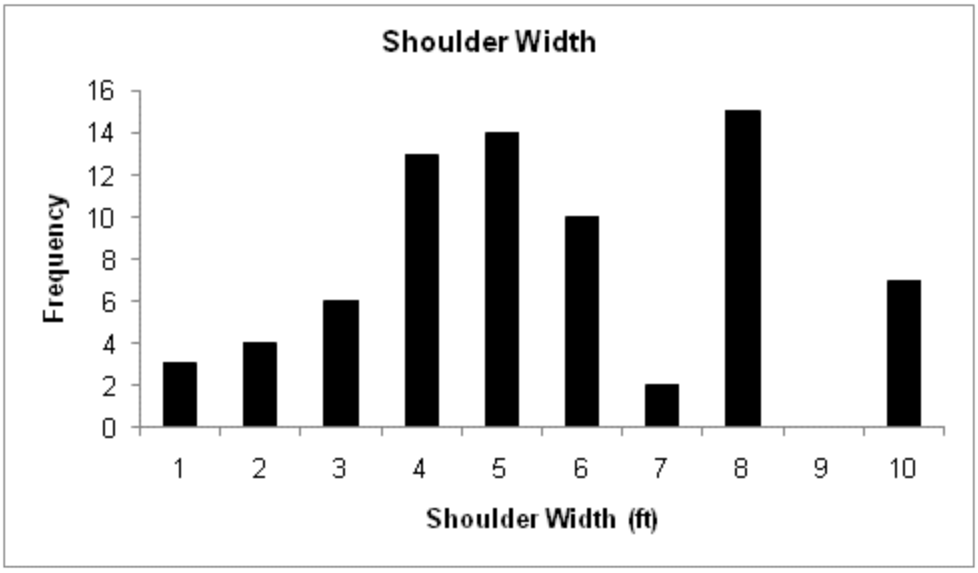

Figure 6 Distribution of Shoulder Width

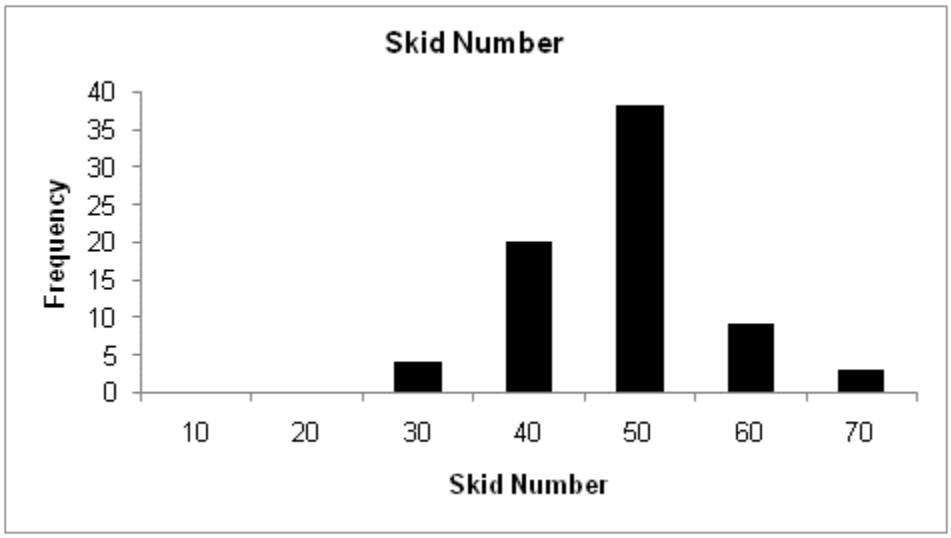

Figure 7 Distribution of Skid Number

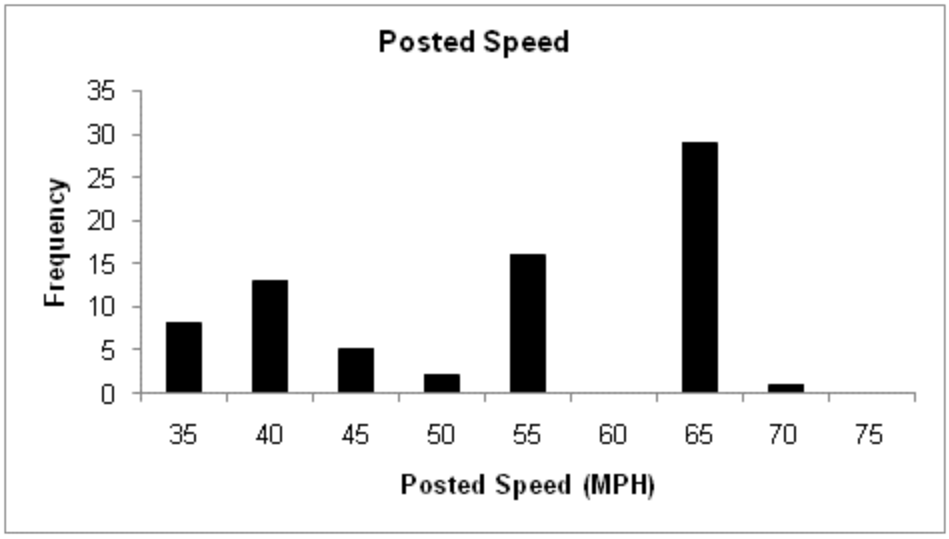

Figure 8 Distribution of Posted Speed

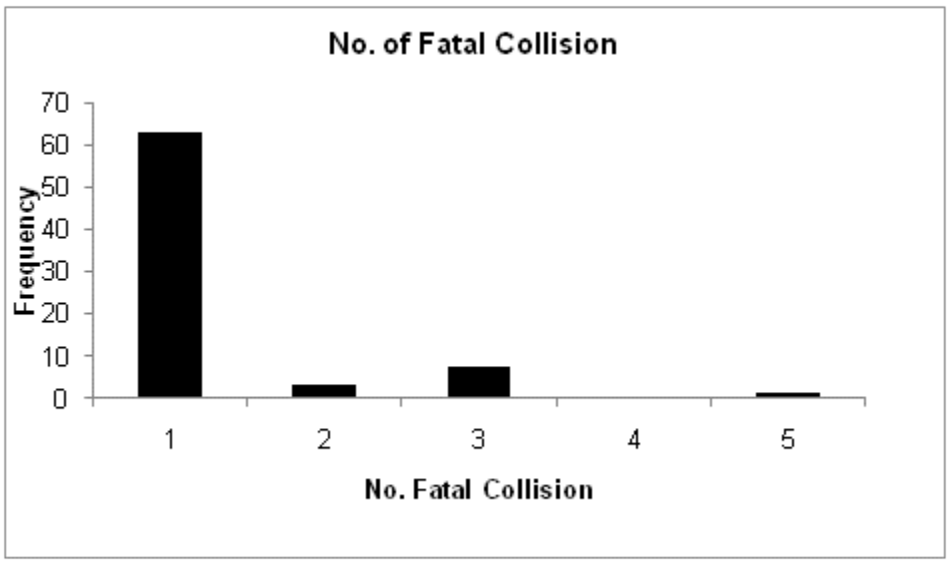

Figure 9 Distribution of No. Fatal Collision

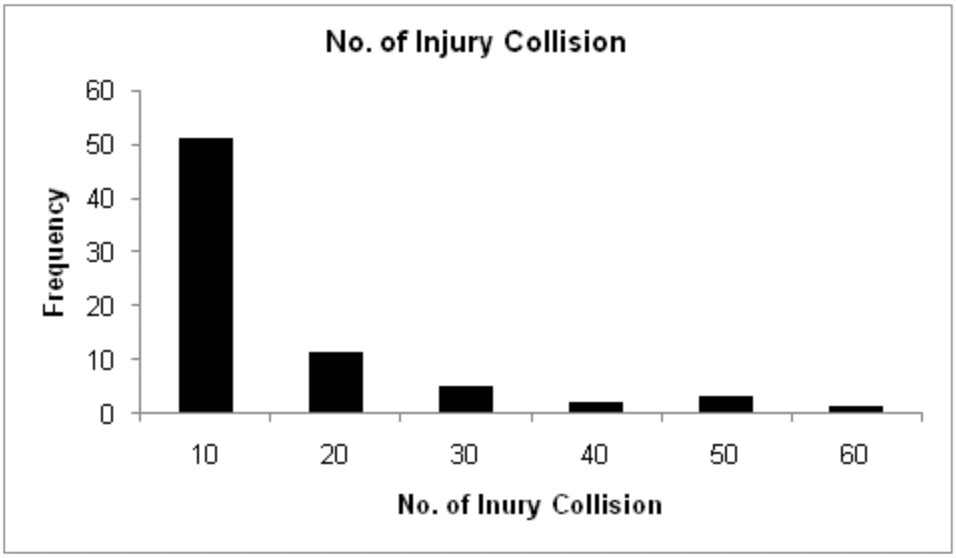

Figure 10 Distribution of No. of Injury Collision

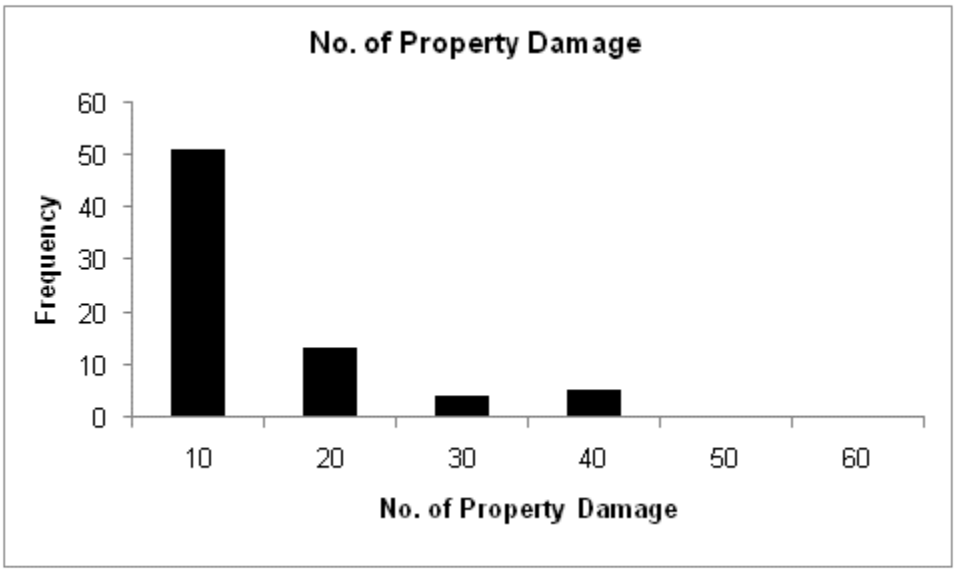

Figure 11 Distribution of No. of Property Damage

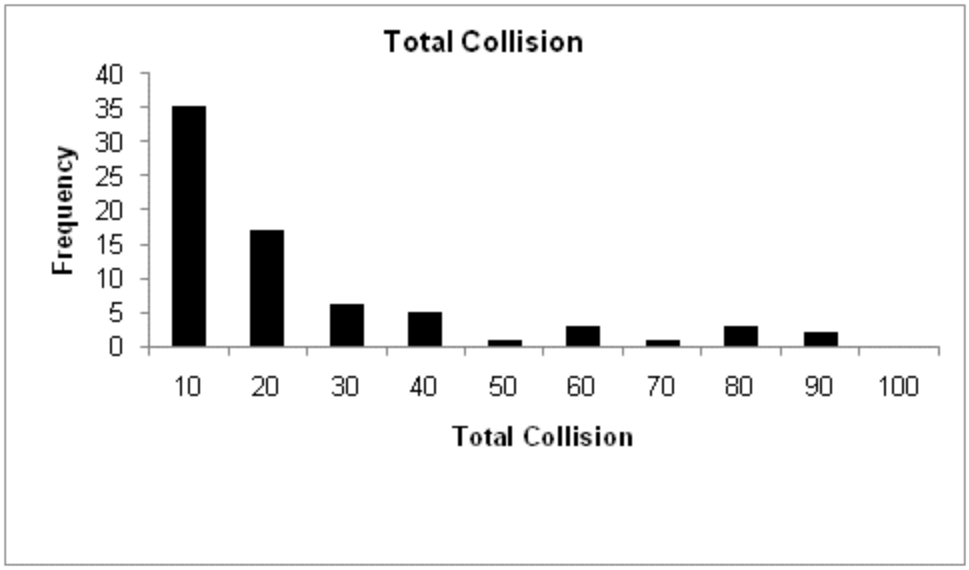

Figure 12 Distribution of No. of Total Collision

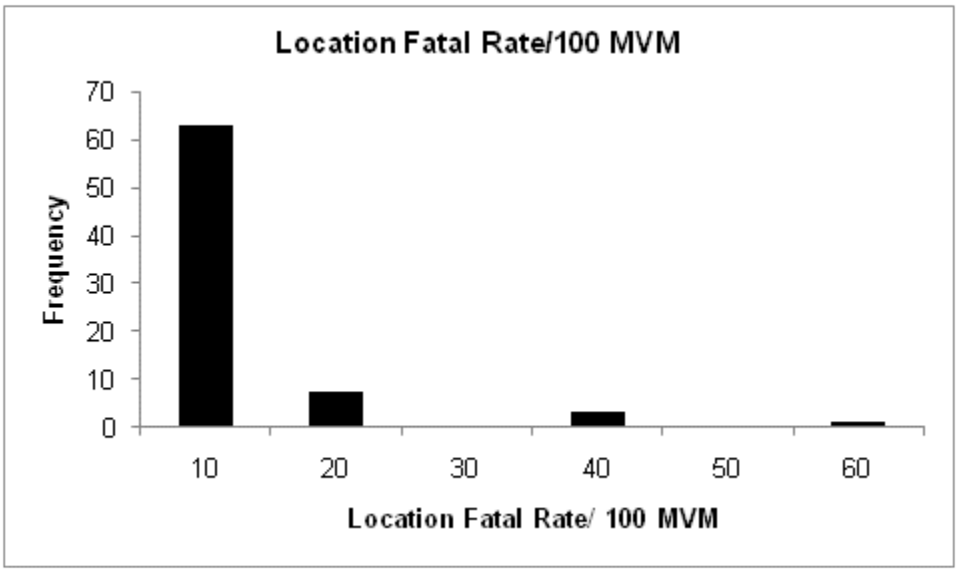

Figure 13 Distribution of Location Fatal Rate

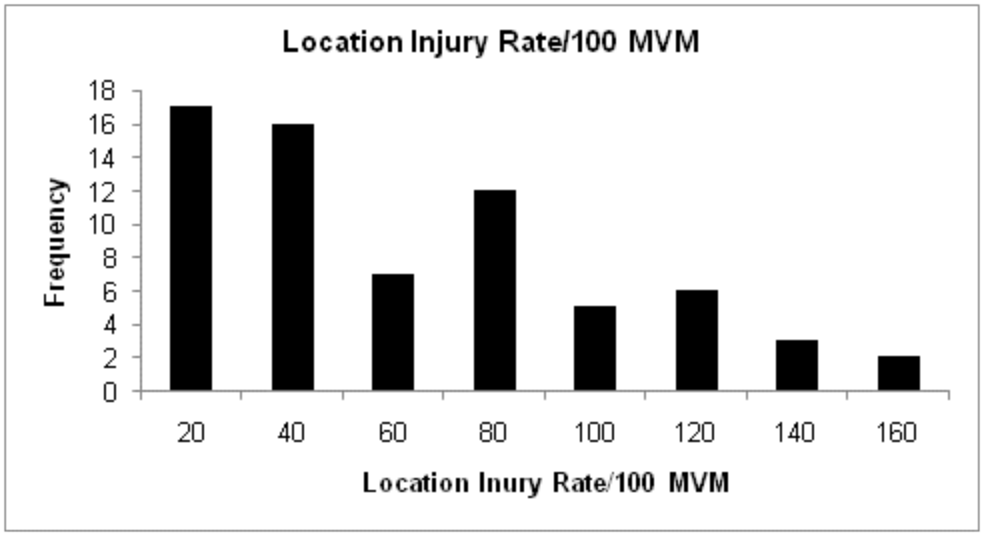

Figure 14 Distribution of Location Injury Rate

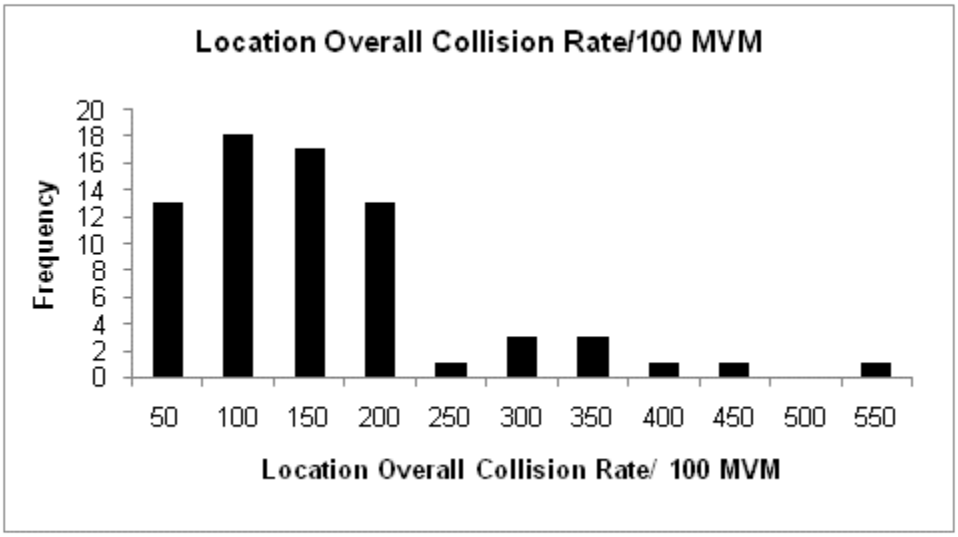

Figure 15 Distribution of Location Overall Collision Rate

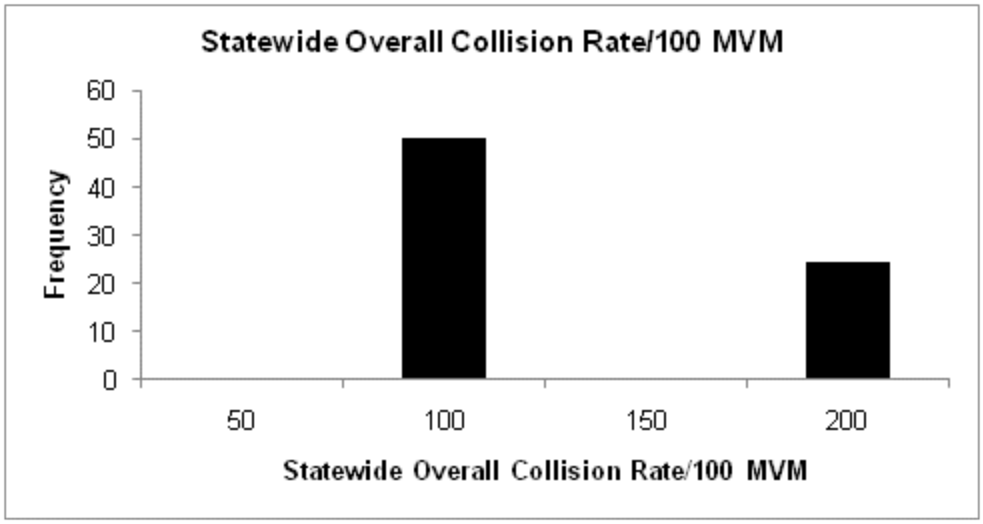

Figure 16 Distribution of Statewide Overall Collision Rate

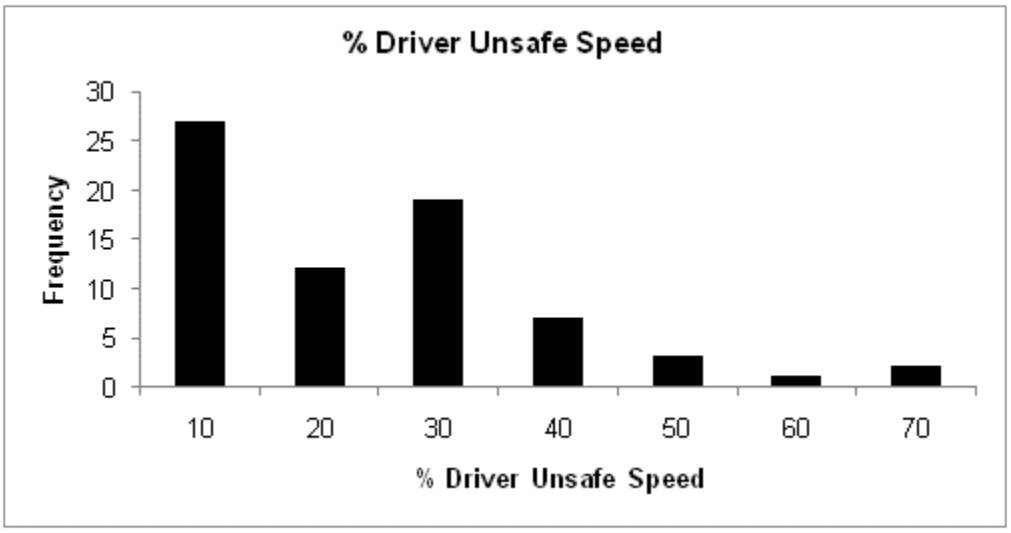

Figure 17 Distribution of % Driver Unsafe Speed

Table 2 The New Data Set

Number	County No.	Control Section	Start Mileage	End Mileage	SubSec. (Miles)	Primary Direction	Surface Width (ft)	IRI (in/mile)	Average Daily Traffic	Shoulder Type	Shoulder Width (ft)	Skid Number	Number of Collisions				Collision Rate (100 Million Vehicle Miles)				% # Unsafe Speed Drivers
													Fatal	Injury	Property damage	Total Collision	Location Fatal Rate	Location Injury Rate	Location Overall Collision Rate	Statewide Overall Collision Rate	
1	1	2	7.45	11.56	4.11	NB	24	121	4400	1	4	41.3	3	34	40	77	4.1	46.8	106.1	86.29	25.8
2	1	8	11.17	12.14	0.97	EB	24	165	2000	1	8	53.3	0	2	7	9	0	25.7	115.5	86.29	7.1
3	2	34	0.27	2.81	2.54	NB	24	147	420	3	5	39.0	0	0	3	3	0	0.0	7	86.29	66.7
4	4	4	2.53	3.01	0.48	EB	24	83	3600	1	10	33.5	1	3	3	7	14.4	43.2	100.9	86.29	30
5	4	26	0	7.42	7.42	EB	24	103	2200	1	8	49.4	0	18	29	47	0	27.5	71.7	86.29	9.5
6	5	20	0.84	1.29	0.45	NB	24	159	820	3	3	37.9	0	4	7	11	0	270.0	742.5	189.05	12.5
7	5	42	0	5	5	EB	24	67	4200	1	8	37.7	3	16	15	34	3.6	19.0	40.3	86.29	9.6
8	6	4	1	2.56	1.56	EBOL	24	77	4600	1	8	40.9	1	9	7	17	3.5	31.2	59	86.29	35
9	7	10	19.56	19.99	0.43	EB	22	128	4500	1	8	46.1	0	2	1	3	0	25.7	38.6	189.05	0
10	7	25	9.76	10.45	0.69	EB	24	114	2700	6	4	46.0	0	4	4	8	0	53.5	107	189.05	25
11	7	25	10.45	10.7	0.25	EB	24	172	2700	1	4	49.0	1	3	8	12	36.9	110.7	442.8	189.05	43.8
12	9	8	2.37	4.56	2.19	NB	24	128	4800	3	2	42.4	1	7	8	16	2.4	16.6	37.9	189.05	4.2
13	9	14	0	2	2	EB	24	113	690	3	3	41.7	1	2	7	10	18	36.1	180.5	86.29	9.1
14	9	14	10.8	11.23	0.43	EB	24	141	1000	3	3	48.0	1	0	2	3	57.9	0.0	173.8	189.05	0
15	11	24	16.28	17.55	1.27	NB	24	74	3700	2	5	41.4	0	13	15	28	0	68.9	148.4	86.29	18.8
16	14	4	7.19	9.24	2.05	NB	24	161	8000	1	4	23.2	1	48	38	87	1.5	72.9	132.1	86.29	21.9
17	14	22	2.44	5.76	3.32	EB	24	87	6300	3	5	40.6	1	27	26	54	1.2	32.2	64.3	86.29	18
18	15	6	4.89	5.68	0.79	EB	24	124	3800	1	10	55.5	0	8	8	16	0	66.4	132.7	86.29	0
19	16	5	0.98	1.24	0.26	EB	24	232	5800	1	10	44.3	1	6	7	14	16.5	99.1	231.2	86.29	22.7
20	17	2	1.34	1.55	0.21	EB	22	65	2100	3	7	47.1	0	0	0	0	0	0.0	0	189.05	0
21	25	4	5.6	6.2	0.6	EB	24	103	4100	1	8	43.5	1	7	5	13	10.1	70.9	131.6	86.29	21.1
22	25	36	9.66	10.76	1.1	NB	24	92	5200	1	8	37.6	0	21	38	59	0	91.4	256.9	189.05	2.6
23	26	28	7.56	7.93	0.37	EB	22	64	2500	3	5	33.2	0	8	5	13	0	215.4	350	189.05	17.6
24	28	12	0	0.64	0.64	EB	24	95	1500	3	6	29.2	0	4	3	7	0	103.8	181.6	86.29	33.3
25	28	12	6.34	6.79	0.45	EB	24	81	2900	3	6	31.3	0	0	0	0	0	0.0	0	86.29	0
26	32	16	6.61	7.61	1	EB	24	70	1900	1	8	47.6	0	3	5	8	0	39.3	104.9	189.05	21.4
27	32	16	7.61	7.95	0.34	EB	24	68	2000	1	8	50.3	1	1	0	2	36.6	36.6	73.3	86.29	0
28	34	10	6.83	10.53	3.7	EB	22	77	940	3	5	43.4	1	5	3	9	7.2	35.8	64.5	189.05	27.3
29	35	8	2.28	2.66	0.38	NB	22	150	520	3	2	45.2	0	0	0	0	0	0.0	0	86.29	0
30	35	16	3.27	3.91	0.64	EB	24	97	4700	1	8	45.0	0	2	11	13	0	16.6	107.6	189.05	0
31	35	36	0	5.1	5.1	EB	24	101	680	1	6	49.6	0	11	10	21	0	79.0	150.8	86.29	50
32	40	4	1.71	3.76	2.05	NB	24	104	7800	1	10	41.0	2	8	7	17	3.1	12.5	26.5	86.29	44.8
33	40	40	0	8	8	EB	22	112	2000	3	3	50.2	3	33	25	61	4.7	51.4	95	86.29	12.1
34	40	42	6.71	11.36	4.65	EB	24	141	840	3	2	60.1	1	9	9	19	6.4	57.4	121.2	86.29	21.4
35	40	46	0.15	1.08	0.93	EB	24	122	2000	3	1	43.6	0	10	2	12	0	133.9	160.7	189.05	33.3
36	40	74	0	4.31	4.31	NB	22	164	910	3	2	56.9	2	18	9	29	12.7	114.3	184.2	86.29	34.4
37	41	6	3.94	4.25	0.31	EB	24	83	4200	3	5	42.5	0	1	0	1	0	19.1	19.1	86.29	0
38	41	6	7.92	9.78	1.86	EB	24	121	2800	3	5	30.6	1	13	16	30	4.8	62.2	143.5	189.05	17.1
39	41	8	3.78	4.2	0.42	EB	24	246	4300	1	6	39.0	0	10	3	13	0	137.9	179.3	86.29	30.4
40	41	26	0	0.55	0.55	NB	24	104	5600	3	5	44.2	0	2	0	2	0	16.2	16.2	86.29	0
41	44	45	7.76	10.76	3	NB	24	50	5900	3	5	21.9	3	101	132	236	4.2	142.1	332.1	189.05	9.9

Number	County No.	Control Section	Start Mileage	End Mileage	SubSec. (Miles)	Primary Direction	Surface Width (ft)	IRI (in/mile)	Average Daily Traffic	Shoulder Type	Shoulder Width (ft)	Skid Number	Number of Collisions				Collision Rate (100 Million Vehicle Miles)				% # Unsafe Speed Drivers
													Fatal	Injury	Property damage	Total Collision	Location Fatal Rate	Location Injury Rate	Location Overall Collision Rate	Statewide Overall Collision Rate	
42	45	16	8.23	8.83	0.6	NB	24	90	3600	1	6	36.2	1	3	2	6	11.5	34.6	69.2	86.29	0
43	45	16	12.33	16.83	4.5	NB	24	114	3300	1	6	47.6	1	22	14	37	1.7	36.9	62.1	86.29	32.6
44	45	18	13.1	13.2	0.1	NB	24	110	1800	1	6	33.3	0	3	1	4	0	415.1	553.5	189.05	0
45	45	48	0.56	0.87	0.31	EB	22	157	1200	3	1	40.2	0	0	1	1	0	0.0	67	189.05	0
46	45	48	2.75	4.32	1.57	EB	22	189	910	3	5	36.4	1	4	4	9	17.4	69.7	156.9	86.29	81.8
47	45	85	0	5.42	5.42	EB	20	155	210	3	3	62.3	0	5	4	9	0	109.4	196.9	86.29	20
48	50	20	8.98	11.21	2.23	NB	24	100	2500	3	4	41.5	1	14	20	35	4.5	62.5	156.4	86.29	19.6
49	57	10	0	1.26	1.26	EB	24	69	2900	1	8	37.1	0	3	12	15	0	20.4	102.2	189.05	24
50	57	10	5.32	5.91	0.59	EB	24	162	2300	3	6	37.3	0	5	2	7	0	91.8	128.5	86.29	0
51	57	10	5.91	10.03	4.12	EB	24	166	1600	3	4	37.2	0	7	11	18	0	26.4	68	86.29	16.7
52	57	12	0	0.26	0.26	NB	24	122	2500	3	6	42.4	1	3	3	7	38.3	115.0	268.2	86.29	25
53	57	12	5.17	5.43	0.26	NB	24	174	1900	1	5	46.3	0	1	0	1	0	50.4	50.4	189.05	0
54	60	6	12.5	12.93	0.43	EB	24	109	7800	1	10	39.1	0	2	3	5	0	14.9	37.1	86.29	14.3
55	61	10	2.7	4.7	2	EB	24	102	6500	1	6	58.2	3	43	33	79	5.7	82.4	151.4	86.29	26.7
56	61	12	0	0.97	0.97	EB	18	279	450	3	1	54.9	0	0	1	1	0	0.0	57.1	86.29	100
57	62	16	10.63	11.04	0.41	EB	24	109	7100	1	8	48.2	1	3	5	9	8.6	25.7	77	86.29	22.2
58	62	40	4.89	7.4	2.51	NB	24	125	1900	3	5	47.6	0	12	8	20	0	62.7	104.5	86.29	21.7
59	62	40	7.4	7.64	0.24	NB	22	81	2000	3	7	45.7	0	4	6	10	0	207.6	518.9	86.29	17.6
60	62	40	15.35	15.57	0.22	NB	24	158	2400	1	10	45.8	0	3	5	8	0	141.5	377.4	86.29	54.5
61	63	10	0	0.18	0.18	EB	24	117	5700	1	8	31.4	0	9	4	13	0	218.5	315.6	189.05	66.7
62	63	18	5.2	8.25	3.05	NB	22	66	3400	3	4	41.7	2	30	19	51	4.8	72.1	122.5	189.05	7.5
63	63	20	7.4	10.53	3.13	NB	24	97	6100	1	10	43.8	3	24	12	39	3.9	31.3	50.9	86.29	9.4
64	63	28	0	0.21	0.21	EB	24	78	2000	3	4	34.9	0	0	0	0	0	0.0	0	86.29	0
65	63	28	6.86	7.03	0.17	EB	24	87	1800	3	4	34.1	0	6	3	9	0	488.4	732.5	86.29	27.8
66	63	36	0	0.63	0.63	NB	24	104	4200	1	8	41.8	1	7	0	8	9.4	65.9	75.3	86.29	25
67	63	36	0.63	4.43	3.8	NB	24	84	4200	3	5	41.8	5	52	30	87	7.8	81.1	135.8	86.29	26.8
68	64	4	0.97	1.07	0.1	NB	24	126	3100	1	4	53.3	0	0	1	1	0	0.0	80.3	189.05	100
69	64	16	0	0.47	0.47	EB	24	114	4500	1	4	56.8	0	11	12	23	0	129.5	270.9	189.05	4.4
70	64	19	0	3.55	3.55	NB	24	71	1600	1	4	60.7	1	17	19	37	4.4	74.5	162.2	86.29	11.1
71	65	2	12.89	13.07	0.18	NB	22	68	2500	3	5	29.1	0	0	0	0	0	0.0	0	189.05	0
72	66	22	3.7	7	3.3	NB	22	111	3000	3	3	39.6	1	41	32	74	2.5	103.1	186.2	86.29	25.8
73	75	4	6.8	12.96	6.16	NB	24	46	2200	1	8	43.7	3	12	15	30	5.5	22.1	55.1	86.29	33.3
74	77	15	0	4.16	4.16	EB	22	133	510	3	4	47.8	0	5	3	8	0	58.7	93.9	86.29	22.2

APPENDIX B

Brief Description of the Neural Network Model

The input parameters in the neural network model are treated as a vector, p. The output t, 85th percentile speed, is expressed as a function. A 3 layered architecture consisting of an input layer (p), 1 hidden layer, and an output layer (t) is considered. The function to be determined can be expressed in terms of a composition of functions

$$t = f_2(W_2\, f_1\, (W_1\, p + b_1)+b_2)$$

The function is determined by specifying the matrices W_1 and W_2 along with the bias vectors b_1 and b_2 and the functions f_1 and f_2. The functions f_1 and f_2 are also specified. In the current model, f_1 consists of N_1 hyperbolic functions that map each of the N_1 components of the (W_1 p + b_1) to a number, and thus returning an N_1 vector. Since the result of the next multiplication is a number, the function f_2 is simply a function that takes a number and returns a number which, in this case, is the output t. A simple linear function is used for f_2. In examples it has been convenient to take N_1 to be 4. To "train" the neural network means to determine the 14 N_1 +1 parameters to match the (input, output) or (p_o,t_o) data pairs. Typically, data is presented as a collection of M data pairs. It is useful to preprocess the data normalizing so that means are zero and variances are unity. A principal component analysis is also useful to possibly reduce the number of variables, although in this application the number of variables is sufficiently small that it does not appear to be necessary. This collection is partitioned into subcollections for training, validation, and testing. The 14 N_1 +1 parameters are determined to minimize the squared error between model output and the observation

$$J(W_1,b_1,W_2,b_2) = \Sigma\, [f_2(W_2\, f_1\, (W_1\, p_o + b_1)+b_2) - t_o]^2$$

where the sum is over the set of training data. The training procedure amounts to a minimization algorithm that iteratively determines W_1, b_1, W_2, and b_2 by means of the backpropagation or the Levenburg-Marquardt algorithms. It is clear that the most important data in the procedure is the 85th percentile speed since that data provides the output without which we cannot train the neural network. The new data set, while it has skid test data, does not contain the 85th percentile speed data. When the network is trained, validation and testing is conducted using subcolletion of data that is not used in the training. It can occur that outputs depend on the initialization of the weights entering the minimization algorithms. Ideally, the network should give close to the same answer regardless of the initial value of the weight.

Another issue is to decide the value of additional data. Again, our previous work as set the stage to use the Akaike information content to analyze this problem. The approach will be to observe the behavior of the information content as new data is added. We may obtain a measure of the efficiency of the information by dividing the information content by the cost to obtain it. This gives information per unit dollar. This quantity should help decide the value of new information.