The reliability of the urban road network: Accident forecast models

Sascia Canale
Dipartimento di Ingegneria Civile e Ambientale. Università degli Studi di Catania

Salvatore Leonardi
Dipartimento di Ingegneria Civile e Ambientale. Università degli Studi di Catania

Giuseppina Pappalardo
Dipartimento di Ingegneria Civile e Ambientale. Università degli Studi di Catania

SYNOPSIS

The fundamental characteristic of the Italian road accident is the highest rate of urban crashes: 73% approximately of the total of the accidents happens in urban contexts where, against the European average, every day are recorded more than 4,2 died; moreover, more than 50% of the totality of urban accidents, happens at the intersections, which represent dangerous points.

Concentrating the study to the urban intersections, it has been characterized the relations existing between the number of accidents and the relative characteristics to geometry, to traffic regulation systems and to traffic flows.

Starting from the theoretical support constituted by various international studies, as the ones executed by the Federal Highway American Administration (Statistical models of at-grade intersection accidents. Technical report. November 1996), a study campaign in Catania has been carried out.

In particular, the accidents in the space of three years, the traffic flows on the arteries situated inside the city area, the geometric characteristics and the traffic regulation systems of 400 intersections (the statistical sample) were collected.

The data processing have been conducted putting exclusively into account the accidents with injuries, happened on five types of urban nodes (four-leg uncontrolled, three-leg uncontrolled, four-leg STOP-controlled, three-leg STOP-controlled, four-leg signalized).

Two general types of statistical models were applied to the accident data in this study: (1) a lognormal regression model and (2) a loglinear regression model (Poisson regression).

The accident forecast models have an elevated degree of significance; moreover the contribution supplied from some variable relative to geometry and the traffic regulation systems is equal, and sometimes greater, that one supplied from the vehicular capacities.

The proposed methodology is a useful instrument for the effective understanding of the accident case in the urban intersections and for the eventual organization of improving intervention of the total reliability of the same road nodes.

The reliability of the urban road network: Accident forecast models

Research to develop accident predictive models published in recent literature has moved away from approaches based on multiple regression and begun to use underlying distributional assumptions other than normal. The Poisson distribution is appropriate for rare events like traffic accident counts where the number of events in a given time period is likely to be zero.
One of the basic assumptions when choosing a Poisson model is that the mean and the variance of the error distribution are equal: in many applications data exhibit extra variation (i.e. the variance is greater than the mean of the estimated Poisson model). This situation is referred to as overdispersion and an alternative model for addressing error structures with overdispersion like than often found in accident data is the binomial distribution.
The objective of this research is to develop statistical models of the relationship between traffic accidents and geometric design, traffic control and traffic volume.
Starting from the theoretical support constituted by various international studies, between which that perform from the Federal Highway American Administration, has been carried out a study campaign in Catania.
In particular, has been collected the accidents in the space of three years, the traffic volume on the arteries situated inside the city area, the geometric characteristics and the traffic control of 400 intersections (the statistical sample).
The data processing have been conducted putting exclusively into account the accidents with injuries, happened on five types of urban nodes (no control four-leg, no control three-leg, STOP-controlled four-leg, STOP-controlled three-leg, signalized four-leg).
Two general types of statistical models were applied to the accident data in this study: (1) a lognormal regression model and (2) a loglinear regression model (Poisson regression).
The accidental data have been adequately deals to obtain two categories of previsional models:

- previsional models of the total number of incidents in a period of reference;
- previsional models of the incidents classified in two various type (in a prefixed temporal interval): 1) accident for lateral collision; 2) accident for head-on collision + sideswipe collision + rear-end collision.

The accident forecast models have an elevated degree of significance; moreover the contribution supplied by some variable relative to geometry and the traffic control systems is equal, and sometimes greater, that one supplied from the vehicular capacities.
The proposed methodology is an useful instrument for the effective understanding of the accident case in the urban intersections and for the eventual organization of improving intervention of the total reliability of the same road nodes.

STATISTICAL MODELS OF AT-GRADE URBAN INTERSECTION ACCIDENTS

Data Preparation

The major activity in the research was to identify databases of geometric design, traffic control, traffic volume (figure 1, table 1) and accident data (figure 2) for at-grade intersections, that were suitable for testing the development of statistical models for accident prediction.
The intersections, pertaining to the study area, are approximately 2000; not being able to find them, all an equal random sample to 20% of the total is itself chosen; inside the champion are, therefore, included also the intersections where no accident occurred in the study period (figure 3).
The variables have been continuously divided, with quantitative values on a continuous scale, and categorical, with a finite number of discrete levels or categories. For each intersection, is filled in a form (table 2) and, they have been classified in five groups, on the basis of leg number and type of traffic control (table 3):

1. three - leg, no control intersections;
2. three – leg, STOP - controlled intersections;
3. four - leg, no control intersections;
4. four – leg, STOP - controlled intersections;
5. four – leg, signalized intersections.

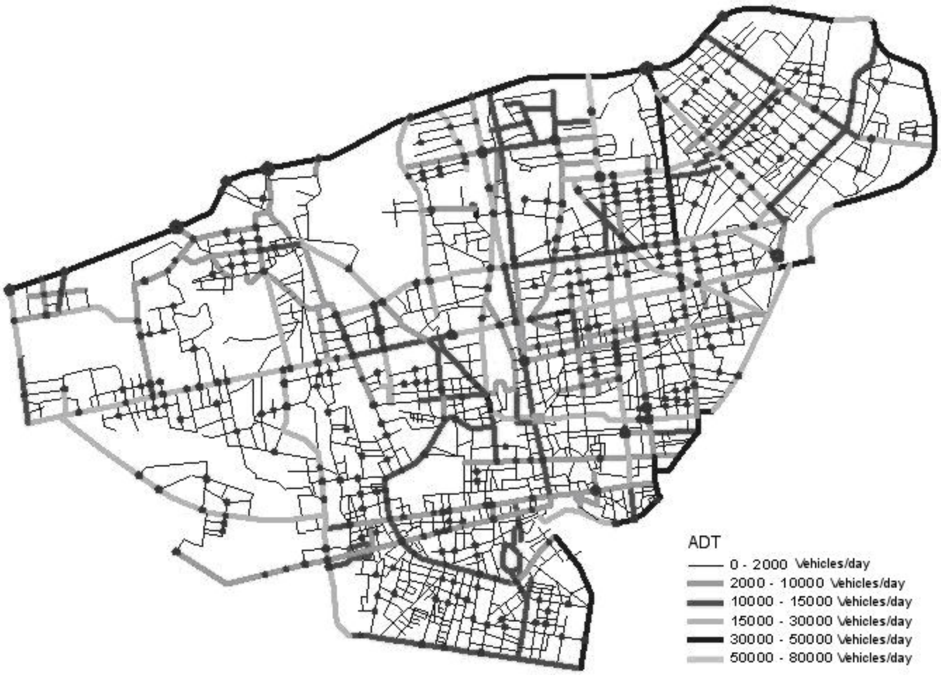

Figure 1: Average daily traffic (vehicles/day)

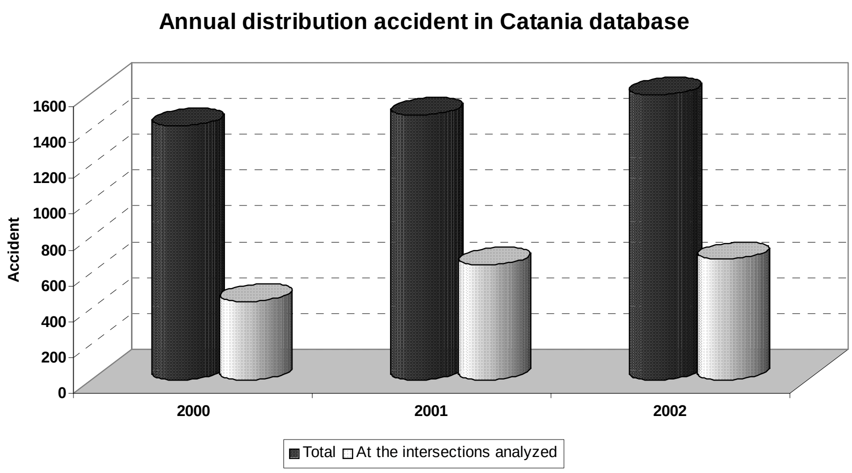

Figure 2: Annual accident distribution in Catania database

Figure 3: Intersection database

Table 1: Definitions of variables obtained in field study

Variable	Variable type	Range of levels
Geometric Design Features		
Intersection configuration	Categorical	Three – leg T intersections
		Three – leg Y intersections
		Four - leg intersections
		Four - leg offset intersections
		Multi – leg intersection
Number of lanes on major road	Continuous	
Number of lanes on crossroad	Continuous	
Presence of median on major road	Categorical	Absent
		Present
Type of left-turn treatment on major road	Categorical	No left-turn
		Without exclusive turn lane
		With exclusive turn lane
Type of left-turn treatment on crossroad	Categorical	No right-turn
		Without exclusive turn lane
		With exclusive turn lane
Type of right-turn treatment on major road	Categorical	No right-turn
		Without exclusive turn lane
		With exclusive turn lane
Type of right-turn treatment on crossroad	Categorical	No right-turn
		Without exclusive turn lane
		With exclusive turn lane
Average lane width on major road, computed as total traveled way width divided by total number of lanes (m)	Continuous	
Average lane width on crossroad (m)	Continuous	
Side-walk width on major road (m)	Continuous	
Side-walk width on crossroad (m)	Continuous	
Percent grade on major road	Categorical	Level (<3%)
		Moderate grade (3%-6%)
		Steep grade (>6%)
Percent grade on crossroad	Categorical	Level (<3%)
		Moderate grade (3%-6%)
		Steep grade (>6%)
Traffic Control Features		
Type of intersection control	Categorical	No control
		STOP control
		Signal control
One-way vs. two way operation on major road	Categorical	Two way operation
		One way operation
Signal phasing	Categorical	Two-phase
		Multiphase
Traffic Volume Data		
ADT of major road (veh/day)	Continuous	
ADT of crossroad (veh/day)	Continuous	

Table 2: Form for the field study

Intersection..	**Major road**			**Crossroad**	
1. Intersection configuration	3T❍	3Y❍	4S❍	4A❍	PR❍
2. Type of intersection control	No control ❍	STOP❍	Signal control❍		
3. Number of lanes on major road	....				
4. Number of lanes on crossroad	....				
5. Presence of median on major road	Absent❍			Present ❍	
6. Type of left-turn treatment on major road	No left-turn ❍	Without exclusive turn lane ❍	With exclusive turn lane ❍		
7. Type of left-turn treatment on crossroad	No left-turn ❍	Without exclusive turn lane ❍	With exclusive turn lane ❍		
8. Type of right-turn treatment on major road	No left-turn ❍	Without exclusive turn lane ❍	With exclusive turn lane ❍		
9. Type of right-turn treatment on crossroad	No left-turn ❍	Without exclusive turn lane ❍	With exclusive turn lane ❍		
10. One-way vs. two way operation on major road	One way operation ❍			Two way operation ❍	
11. One-way vs. two way operation on crossroad	One way operation ❍			Two way operation ❍	
12. Signal phasing	2❍			Multi❍	
13. Road markings	Present❍			Absent❍	
14. Percent grade on major road	Level❍	Moderate grade ❍	Steep grade❍		
15. Percent grade on crossroad	Level❍	Moderate grade ❍	Steep grade❍		
16. Average lane width on major road	....				
17. Average lane width on crossroad	....				
18. Side-walk width on major road	....				
19. Side-walk width on crossroad	....				

Table 3: Intersection distribution by type of intersection

Configuration	Type of intersection control			Total
	No control	*STOP*	*Signal control*	
Three – leg T intersections	86	35	0	121
Three – leg Y intersections	2	1	0	3
Four - leg intersections	121	111	26	258
Four - leg offset intersections	3	4	4	11
Multi – leg intersection	2	2	3	7
				400

Regression Models

The objective of the statistical models is to provide a relationship between a function of the expected number of accident, E(Yi) = μ_i, at the ith intersection and the q intersection parameters, X_1, X_2, …, X_q, describing the geometric design, traffic control and traffic volume. This relationship can be formulated as:

$$Y_i = \exp(\beta_0 + \beta_1 X_{i1} + \ldots + \beta_q X_{iq}) \quad [1]$$

where the regression coefficients, β_i, are to estimated from the data.

The estimation procedure used to obtain the regression coefficients is dependent on the assumption made about the distribution of Y_i:

- Lognormal regression models;
- Loglinear Poisson models.

Lognormal regression models are based on the assumption that the natural logarithm of Y_i follows a normal distribution with mean μ_i and variance σ^2. In other words, it is assumed that Yi follows a lognormal distribution, a reasonable choice whenever the data are inherently non-negative, suggesting that a model with positive skewness is needed and the mean is relatively large. This model also ensures that μ_i, the expected number of accident, remains positive. In this case, the relationship between the expected number of accidents at the ith intersection and the q predictor variables, X_1, X_2, …, X_q, can be written as:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_q X_{iq} \quad [2]$$

or alternatively, in the moltiplicative form, as:

$$\mu_i = \exp(\beta_0) \cdot \exp(\beta_1 X_{i1}) \cdot \exp(\beta_2 X_{i2}) \ldots \cdot \exp(\beta_q X_{iq}) \quad [3]$$

where the log(μ_i) is assumed to follow a normal distribution with mean μ_i and variance σ^2. The coefficients β_0, β_1, …, β_q are the linear regression coefficients to be estimated by ordinary least-squares method. This is the classical case of a multiple linear relationship between the logarithm of the dependent variable and q independent predictor variables. For the lognormal regressions, the normal-theory tests of significance of the parameters and goodness-of-fit of the model measures apply.

When the average number of accidents at an intersection becomes small, the assumption of a lognormal distribution is no longer valid. The Poisson model becomes a natural choice as it models the occurrence of rare discrete events well. The relationship between the expected number of accidents occurring at the ith intersection and the q intersection parameters, X_1, X_2, …, X_q, is assumed to be of the form:

$$\log(\mu_i) = \beta_0 + \sum_{j=1}^{q} \beta_j X_{ij} \quad [4]$$

The assumption is made that the number of accident, Y_i, follows a Poisson distribution with mean μ_i. The probability that an intersection defined by a known set of predictor variables, X_1, X_2, …, X_q, experiences Y_i accidents can be expressed as:

$$P(Y_i = y_i; \mu_i) = \mu_i^{y_i} \frac{e^{-\mu_i}}{y_i!} \quad [5]$$

Note that the Poisson distribution has only one parameter, its mean, with the limitation that the variance equals the mean of the distribution. The regression coefficients, β_0, β_1, …, β_q are estimated by the maximum likelihood method. This function is the product of the terms in equation [2] over all n intersections in the class of interest. This function is viewed as a function of the parameters, μ_i e, and through them, the parameters β_i. The parameters are estimated by maximizing the likelihood, or more usually, by maximizing the logarithm of the likelihood, or minimizing the negative of the log likelihood:

$$\log(L) = \sum_{i=1}^{n} [y_i \log(\mu_i) - \mu_i - \log(y_i!)] \quad [6]$$

The maximum value possible for the likelihood for a given data set occurs if the model fits the data exactly. This occurs if the μ_i are replaced by Y_i in [3]. The difference between the log-likelihood functions for two models is a measure of how much one model improves the fit over the other.

A limitation of the Poisson distribution is that the mean equals the variance of the distribution; previous work has shown that this is not always the case. If the variance of the data exceeds the estimated mean of accident data distribution, the data are said to be overdispersed.

The negative binomial provides an alternative model to deal with overdispersion in count data. Unlike the Poisson distribution, the negative binomial distribution has two parameters. As for the Poisson model, the relationship between the expected number of accidents occurring at the i^{th} intersection and the q intersection parameters, X_{i1}, X_{i2}, …, X_{iq}, is taken to be:

$$f(\mu_i) = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_q X_{iq} \quad [8]$$

The assumption is made that the number of accidents, Y_i, follows a negative binomial distribution with parameters α and k (with $0 \le \alpha \le 1$ and $k \ge 0$). The probability that an intersection defined by a known set of predictor variables, X_{i1}, X_{i2}, …, X_{iq}, experiences $Y_i = y_i$ accidents can be expressed as:

$$\Pr(Y_i = y_i; \alpha, k) = \frac{(y_i + k - 1)!}{y_i!(k-1)} \frac{\alpha^{y_i}}{(1+\alpha)^{y_i+k}}; \quad y_i = 0, 1, 2, \ldots \quad [9]$$

The mean and variance can be expressed in terms of the parameters α e k as follows:

mean = E(Y) = μ_i = k α_i [10]

variance = Var (Y) = $k\alpha_i + k\alpha_i^2 = \mu_i + \frac{\mu_i^2}{k}$ [11]

The parameter k is not known a priori, but can be estimated so that the mean deviance become unity or the chi-square statistic equals its expectation (i.e., equals its degrees of freedom). As for the Poisson model, the model regression coefficients β_0, β_1, …, β_q are estimated by the method of maximum likelihood. The estimation of the model parameters can be done minimizing the negative of the log likelihood. For the negative binomial distribution, the log likelihood is given by the equation:

$$\log(L) = \sum_{i=1}^{n} y_i \log\left[\frac{\alpha_i}{1+\alpha_i}\right] - nk\log(1+\alpha_i) + f(y_i, k) \quad [12]$$

Substituting $\alpha_i = \frac{\mu_i}{k}$ into the term $\log\left[\frac{\alpha_i}{1+\alpha_i}\right]$ in the [12], gives the function:

$$f\left(\frac{\mu_i}{\mu_i + k}\right) = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_q X_{iq} \quad [13]$$

The parameters α_i e k of the negative binomial distribution can be indirectly estimated using a generalized linear model and, by the means of equations [8] e [13], the model regression coefficients β_0, β_1, , β_q are obtained.

In all models, the natural logarithm of the major road and crossroad ADT variables are used. Thus, in the lognormal and Poisson and negative binomial models described above, X_1 e X_2 generally represent $\log(ADT_{major\_road})$ and $\log(ADT_{crossroad})$, respectively. The multiplicative model can be rewritten as:

$$\mu_i = \exp(\beta_0) \cdot (ADT_{major\_road})^{\beta_1} \cdot (ADT_{crossroad})^{\beta_2} \cdot \exp(\beta_3 X_{i3}) \ldots \cdot \exp(\beta_q X_{iq}) \quad [14]$$

For every group they will come develop at least two models, one containing all variable saying *complete model* and one containing variable whose significance is not inferior to 90%. The choice of the appropriated model more will come carried out being based on the following goodness-of-fit measures:

- C_p of Mallow holds account of the ability that has the model to explain the phenomenon and of the number of variables used. It is calculated through the following relation:

  $$C_p = p + \frac{(n-p) \cdot (s_p^2 - \sigma^2)}{\sigma^2} \quad [15]$$

  where:
  n = number of observations;
  p = number of variables used in the model more one;
  s_p^2 = relative standard deviation;
  σ^2 = minimal standard deviation between the models.

  The better model is characterized from the lowest value than C_p. This index is applicable single in models developed through procedures of linear type like the lognormal regression; in association with the stepwise procedure, for the reduction of the number of variable, this index will be helped in the choice of the better model.
- Mean deviance is an index whose value is given from the relationship:

  $$\overline{D} = \frac{Deviance}{n - df} \quad [16]$$

  where the value of the deviance is calculated according to Nelder and Wedderburn [11] based on the hypotheses on the distribution like:

$$\text{Deviance} = \frac{\sum_{i=1}^{n}(\mu_i - y_i)^2}{\sigma^2} \text{ for the normal distribution} \quad [17]$$

$$\text{Deviance} = 2\left[\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\mu_i}\right) - \sum_{i=1}^{n}(y_i - \mu_i)\right] \text{ for the Poisson distribution} \quad [18]$$

The term *df* is called *degree of freedom* of the model and calculated as it follows: every continuous variable has a degree of freedom; categorical variables have many degrees of freedom how many are the levels less one; finally a degree of freedom for the present constant joins. The better models are those in which the mean deviance assumes values next to the unit.

- R^2: measure the fraction of variability explained from the model. It varies between 0 and 100 and is given from the relation:

$$R^2 = \frac{\sum_{i=1}^{n}(\mu_i^2) - \sum_{i=1}^{n}(\mu_i - y_i)^2}{\sum_{i=1}^{n}(\mu_i^2)} \quad [19]$$

where:

μ_i = expected number of accidents to i[th] intersection;

y_i = observed number of accidents to i[th] intersection.

In Poisson regression, the relation for the calculation of R^2 is similar to [19], but reported to the portion of deviance explained from the model:

$$R^2 = 1 - \frac{\lambda(\text{complete model})}{\lambda(\text{model with one parameter})} \quad [20]$$

where λ represents the value of the deviance defined through the relationship of maximum likelihood.

- R^2_{FT}: measure the fraction of variability explained from the model, holding account of the number of independent variables. It varies between 0 and 100 and is given from the relation:

$$R^2_{FT} = 1 - \left(\frac{n-1}{n-p}\right) \cdot \left(1 - R^2\right) \quad [21]$$

In the Poisson regression, the relation for the determination of this index assumes the shape:

$$R^2_{FT} = 1 - \left(\frac{n-1}{n-p}\right) \cdot \frac{\lambda(\text{complete model})}{\lambda(\text{model with one parameter})} \quad [22]$$

- Quadratic medium error (MSE): represents the quadratic error on the forecast, mediated on the residual degrees of freedom. He is given from the relation:

$$MSE = \frac{\sum_{i=1}^{n}(\mu_i - y_i)^2}{n - df} \quad [23]$$

Smaller it is the error, greater is the reliability of the previewed values.

- Absolute medium error (MAE): represents the error on the forecast, in absolute value, mediated on the residual degrees of freedom. He is given from the relation:

$$MAE = \sum_{i=1}^{n}\left|\mu_i - y_i\right| \quad [24]$$

Smaller it is the error, greater is the reliability of the previewed values.

Three-leg, no-control intersections

The accident distribution (figure 3) tends to follow the shape of a Poisson distribution; consequently the model coefficients are estimated by maximum likelihood method. Minimum, mean, median, and maximum values are given for the continuous variables (table 4); for categorical variables (table 5), the percent of intersections within each level is given.

Using all the continuous and categorical variables, a Poisson regression model was fit to the data for fatal and injury vehicle accidents. This model is referred to as the full model because all the candidate independent variables are included in the model. Generally, the analysis results for the full model found some independent variables to be statistically significance at the 10 percent significance level and other variables to be not statistically significance. To obtain the best estimates of the regression coefficients for the independent variables that are statically significant and the best estimate of the goodness of fit of the model as a whole, the Poisson regression model was fit again, including only those independent variables that were found to be statistically significant in the full model. This model is referred to as the reduced model. The goodness-of-fit measures are shown in table 6. The values show a good dependency between the expected number of incidents and the variable used. Reducing the number of variable, they improve some measures: even if the deviance goes away from the unit, the error on the forecast is reduced and the index R^2_{FT} goes up to 57%. These results provide an indication that the choice of the reduced model appears appropriate.

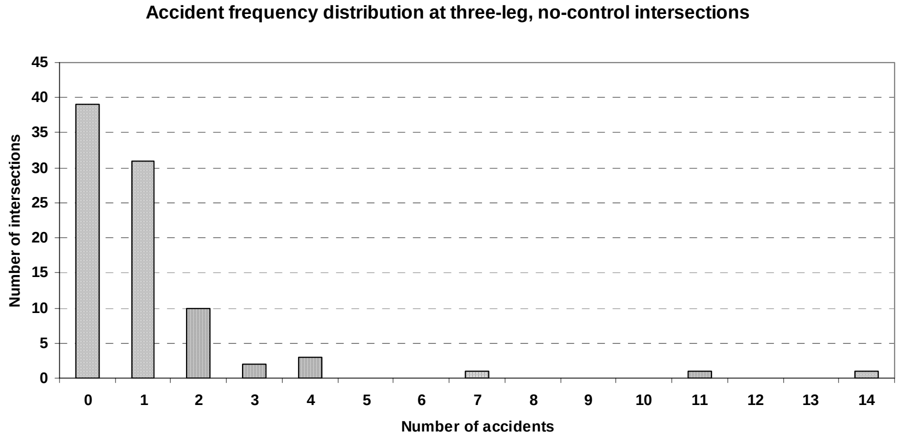

Figure 3: Accident frequency distribution at a sample of 88 three-leg, no-control, intersections

Table 4: Continuous variables - Three-leg, no-control, intersections

Variable	Minimum	Mean	Median	Maximum
Fatal and injury vehicle accidents (2000-2002)	0	1.15	1	14
Major-road ADT (veh/day)	2000	13119	9000	50000
Crossroad ADT (veh/day)	2000	3108	2000	50000
Number of lanes on major road	1	2.10	2	6
Number of lanes on crossroad	1	1.39	1	2
Average lane width on major road (m)	0.92	2.89	2.67	6.05
Average lane width on crossroad (m)	1.60	2.92	2.66	6.15
Side-walk width on major road (m)	0	2.04	1.98	5.2
Side-walk width on crossroad (m)	0	1.40	1.45	4.6

Table 5: Categorical variables - Three-leg, no-control, intersections

Variable	Level	% of intersections
Presence of median on major road	Absent	89.77
	Present	10.23
Type of left-turn treatment on major road	No left-turn	47.73
	Without/with exclusive turn lane	52.27
Type of left-turn treatment on crossroad	No left-turn	57.95
	Without/with exclusive turn lane	42.05
Type of right-turn treatment on major road	No right-turn	43.18
	Without/with exclusive turn lane	56.82
Type of right-turn treatment on crossroad	No right-turn	44.32
	Without/with exclusive turn lane	55.68
One-way vs. two way operation on major road	One-way operation	36.36
	Two-way operation	63.64
One-way vs. two way operation on crossroad	One-way operation	60.23
	Two-way operation	39.77
Road markings	Present	20.45
	Absent	79.55
Percent grade on major road	Level	47.73
	Moderate grade	36.36
	Steep grade	15.91
Percent grade on crossroad	Level	45.45
	Moderate grade	36.36
	Steep grade	18.18

Table 6: Model diagnostics for fatal and injury vehicle accidents - Three-leg, no-control, intersections

	Poisson regression	
	Full model	Reduced model (6 variables)
Number of intersections (n)	88	88
Number of parameters (p)	19	7
Degrees of freedom (df)	21	9
Deviance/(n-df)	0.56	0.46
MSE	1.17	1.10
MAE	0.77	0.67
R^2 (%)	67.82	65.90
R^2_{FT} (%)	46.85	56.91

The expected 3-years fatal and injury accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} -7.39574 + 0.731482 \cdot X_1 + 0.422077 \cdot X_2 + 0.703508 \cdot X_3 - 0.706929 \cdot X_4 \\ -0.745259 X_5 - 0.0288565 \cdot X_6 + 1.17284 \cdot X_7 + 0.301925 \cdot X_8 \end{pmatrix} \quad [25]$$

where:

Y = expected number of fatal and injury vehicle accidents in a 3-year period;
X_1 = ADT of the major road;
X_2 = 1 if permitted left-turn lane is present on the major road; 0 otherwise;
X_3 = 1 if two-way operation is present on major road; 0 otherwise;
X_4 = 1 if road markings is absent; 0 otherwise;
X_5 = 1 if the percent grade on major road is steep grade; 0 otherwise;
X_6 = 1 if the percent grade on major road is level; 0 otherwise;
X_7 = 1 if the percent grade on crossroad is steep grade; 0 otherwise;
X_8 = 1 if the percent grade on crossroad is level; 0 otherwise.

Three-leg, STOP-controlled intersections

The accident distribution (figure 4) don't suggest the better method to calculate the regression coefficients; so the model coefficients are estimated both by least-squares method and by maximum likelihood method. Minimum, mean, median, and maximum values are given for the continuous variables (table 7); for categorical variables (table 8), the percent of intersections within each level is given.

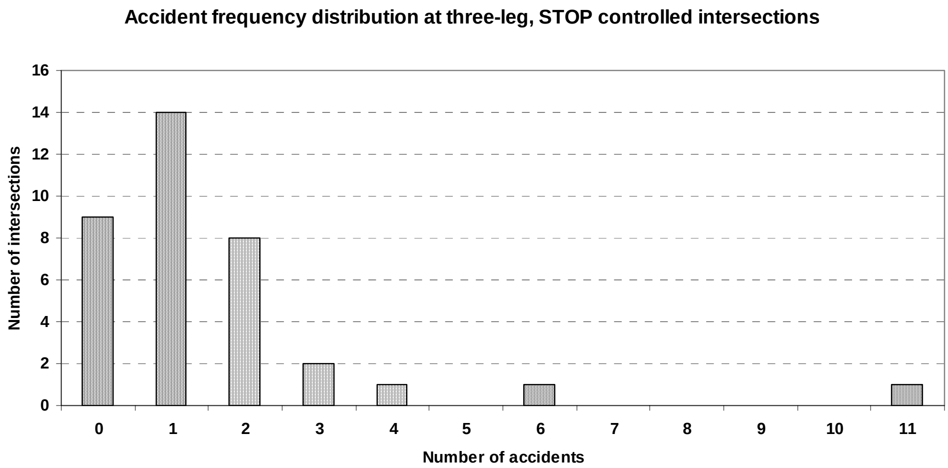

Figure 4: Accident frequency distribution at a sample of 36 three-leg, STOP-controlled intersections

Table 7: Continuous variables - Three-leg, STOP-controlled, intersections

Variable	Minimum	Mean	Median	Maximum
Fatal and injury vehicle accidents (2000-2002)	0	1.58	1	11
Major-road ADT (veh/day)	2000	24611	20000	50000
Crossroad ADT (veh/day)	2000	4611	2000	20000
Number of lanes on major road	1	3.08	2	8
Number of lanes on crossroad	1	1.81	2	4
Average lane width on major road (m)	1.75	3.33	3.10	6.55
Average lane width on crossroad (m)	2.10	3.48	3.10	7.10
Side-walk width on major road (m)	0.95	2.43	2.28	4.95
Side-walk width on crossroad (m)	0	1.18	1.40	2.10

The models through the least-squares method, assuming a lognormal distribution of the number of incidents, and the maximum likelihood method, assuming a Poisson distribution, will come elaborate. For both the procedures, a complete model, with 18 variables, and the reduced model, obtained through the stepwise methodology (table 9), will come elaborate Confronting the two models, we notice as the reduction has carried to an light increment of the error in the forecast; but the reduction of the number of variables involves a simplification of the model, confirmed also from the improvement of some indices like the mean deviance and R^2_{FT}. The hypothesis of a lognormal distribution gives instead the elaboration of three models (table 9): one complete with 18 variables, one reduced to 10 variables characterized from the maximum value of R^2_{FT}, but containing also variables not statistically meaningful to 90%, like "One-way vs. two way operation on major road" (P=23,79%), the percent grade on crossroad (P=20,47%), the number of lanes on crossroad (P=17,00%) and the crossroad ADT (P=16,76%), and one reduced to 6 variables all statistically meaningful to 90%. Between the three models, all the goodness-of-fit measures suggest the use of the model reduced to 6 variables: has a mean deviance near to the unit and is characterized from the lowest values than Cp and of error on the forecasts. The choice is reduced to the reduced models of Poisson and lognormal: we notice as the lognormal model has a value of mean deviance slightly more near to the unit. So the choice of the reduced model of Poisson appears appropriate.

Table 8: Categorical variables - Three-leg, STOP-controlled, intersections

Variable	Level	% of intersections
Presence of median on major road	Absent	69.44
	Present	30.56
Type of left-turn treatment on major road	No left-turn	66.67
	Without/with exclusive turn lane	33.33
Type of left-turn treatment on crossroad	No left-turn	44.44
	Without/with exclusive turn lane	27.76
Type of right-turn treatment on major road	No right-turn	38.89
	Without/with exclusive turn lane	61.11
Type of right-turn treatment on crossroad	No right-turn	13.89
	Without/with exclusive turn lane	86.11
One-way vs. two way operation on major road	One-way operation	16.67
	Two-way operation	83.33
One-way vs. two way operation on crossroad	One-way operation	38.89
	Two-way operation	61.11
Road markings	Present	55.56
	Absent	44.44
Percent grade on major road	Level	33.33
	Moderate grade	47.22
	Steep grade	19.44
Percent grade on crossroad	Level	44.44
	Moderate grade	30.56
	Steep grade	25.00

Table 9: Model diagnostics for fatal and injury vehicle accidents - Three-leg, STOP-controlled, intersections

	Poisson regression		Lognormal regression		
	Full model	Reduced model (8 variables)	Full model	Reduced model (10 variables)	Reduced model (6 variables)
Number of intersections (n)	36	36	36	36	36
Number of parameters (p)	19	9	19	11	7
Degrees of freedom (df)	21	10	21	12	7
C_p	-	-	31.97	11	9.09
Deviance/(n-df)	0.52	0.54	9.95	1.52	1.26
MSE	0.64	1.49	16.68	3.98	3.64
MAE	1.00	1.03	2.46	1.32	1.18
R^2 (%)	79.09	65.20	88.71	73.09	58.78
R^2_{FT} (%)	17.61	35.92	51.08	53.36	46.41

The expected 3-years fatal and injury accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} -14.5577 + 0.810414 \cdot X_1 + 0.405196 \cdot X_2 + 0.339415 \cdot X_3 + 0.445391 \cdot X_4 \\ +1.20729259 X_5 + 0.825767 \cdot X_6 + 0.557943 \cdot X_7 + 0.152255 \cdot X_8 - 1.25338 \cdot X_9 \end{pmatrix} \quad [26]$$

dove:
Y = expected number of fatal and injury vehicle accidents in a 3-year period;
X_1 = ADT of the major road;
X_2 = ADT of the crossroad;
X_3 = average lane width on crossroad;
X_4 = side-walk width on major road;
X_5 = 1 if the median is absent; 0 otherwise;
X_6 = 1 if two-way operation is present on crossroad; 0 otherwise;
X_7 = 1 if road markings is absent; 0 otherwise;
X_8 = 1 if the percent grade on crossroad is steep grade; 0 otherwise;
X_9 = 1 if the percent grade on crossroad is level; 0 otherwise.

Four-leg, no - control intersections

The accident distribution (figure 5) tends to follow the shape of a Poisson distribution; consequently the model coefficients are estimated by maximum likelihood method. Minimum, mean, median, and maximum values are given for the continuous variables (table 10); for categorical variables (table 11), the percent of intersections within each level is given.

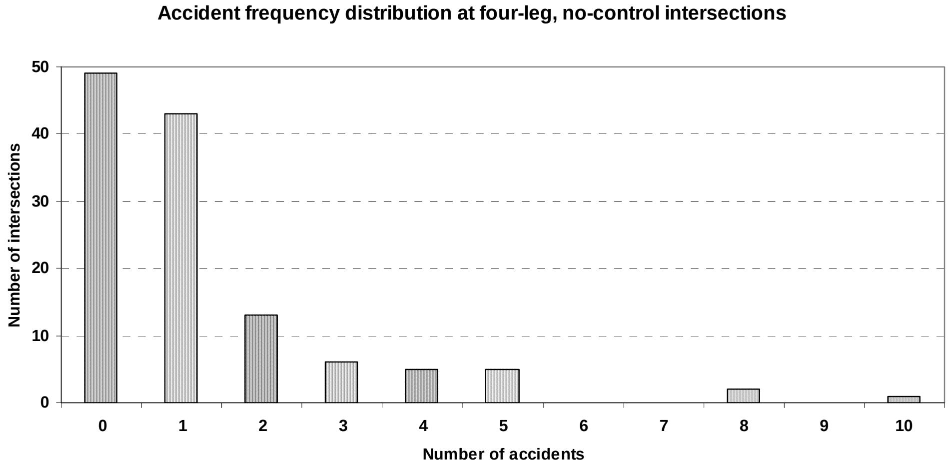

Figure 5: Accident frequency distribution at a sample of 124 four-leg, no-control intersections

Table 10: Continuous variables - Four-leg, no-control intersections

Variable	Minimum	Mean	Median	Maximum
Fatal and injury vehicle accidents (2000-2002)	0	1.27	1	10
Major-road ADT (veh/day)	2000	7399.2	2000	50000
Crossroad ADT (veh/day)	2000	2701.6	2000	15000
Number of lanes on major road	1	1.60	1.50	4
Number of lanes on crossroad	1	1.40	1.00	3
Average lane width on major road (m)	0.95	2.68	2.55	5.03
Average lane width on crossroad (m)	1.33	2.62	2.50	4.90
Side-walk width on major road (m)	0.00	1.80	1.78	4.90
Side-walk width on crossroad (m)	0.00	1.35	1.40	3.50

Table 11: Categorical variables - Four-leg, no-control intersections

Variable	Level	% of intersections
Type of left-turn treatment on major road	No left-turn	30.65
	Without/with exclusive turn lane	69.35
Type of left-turn treatment on crossroad	No left-turn	28.23
	Without/with exclusive turn lane	71.77
Type of right-turn treatment on major road	No right-turn	21.77
	Without/with exclusive turn lane	78.23
Type of right-turn treatment on crossroad	No right-turn	33.06
	Without/with exclusive turn lane	66.94
One-way vs. two way operation on major road	One-way operation	53.23
	Two-way operation	46.77
One-way vs. two way operation on crossroad	One-way operation	61.29
	Two-way operation	38.71
Road markings	Present	18.55
	Absent	81.45
Percent grade on major road	Level	57.26
	Moderate grade	34.68
	Steep grade	8.06
Percent grade on crossroad	Level	60.48
	Moderate grade	33.06
	Steep grade	6.45

Using all the continuous and categorical variables, a Poisson regression model was fit to the data for fatal and injury vehicle accidents. With the data coming from 124 intersections, have been elaborate a complete model, considering all the variables, and one reduced, obtained from the first one through the stepwise procedure (table 12).

The goodness-of-fit measures show a discreet dependency between the expected number of accidents and the used variables. Reducing the number of variables, a simplification of the model is had and even if the mean deviance goes away from the unit, the error on the forecast is reduced and the index R^2_{FT} goes to 25%. These results provide an indication that the choice of the reduced model appears appropriate.

The expected 3-years fatal and injury accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} -5.51515 + 0.250929 \cdot X_1 + 0.243562 \cdot X_2 + 0.316932 \cdot X_3 + 0.256321 \cdot X_4 \\ + 0.403429 X_5 - 0.522508 \cdot X_6 + 0.477602 \cdot X_7 + 0.549291 \cdot X_8 \end{pmatrix} \quad [27]$$

where:
Y = expected number of fatal and injury vehicle accidents in a 3-year period;
X_1 = ADT of the major road;
X_2 = ADT of the crossroad;
X_3 = average lane width on crossroad;
X_4 = side-walk width on crossroad;
X_5 = 1 if two-way operation is present on major road; 0 otherwise;
X_6 = 1 if two-way operation is present on crossroad; 0 otherwise;
X_7 = 1 if the percent grade on crossroad is steep grade; 0 otherwise;
X_8 = 1 if the percent grade on crossroad is level; 0 otherwise.

Table 12: Model diagnostics for fatal and injury vehicle accidents - Four-leg, no-control, intersections

	Poisson regression	
	Full model	**Reduced model (7 variables)**
Number of intersections (n)	124	124
Number of parameters (p)	18	8
Degrees of freedom (df)	20	9
Deviance/(n-df)	0.86	0.78
MSE	2.40	2.24
MAE	1.21	1.09
R^2 (%)	33.48	31.94
R^2_{FT} (%)	17.32	24.66

Four-leg, STOP-controlled intersections

The accident distribution (figure 6) don't suggest the better method to calculate the regression coefficients; so the model coefficients are estimated both by least-squares method and by maximum likelihood method.
Minimum, mean, median, and maximum values are given for the continuous variables (table 12); for categorical variables (table 13), the percent of intersections within each level is given.
Two models through the least-squares method, assuming a lognormal distribution of the number of incidents, and the maximum likelihood method, assuming a Poisson distribution will come elaborate. For both the procedures, a complete model, with 17 variables, and the reduced model, obtained through the stepwise methodology will come elaborate(table 15).

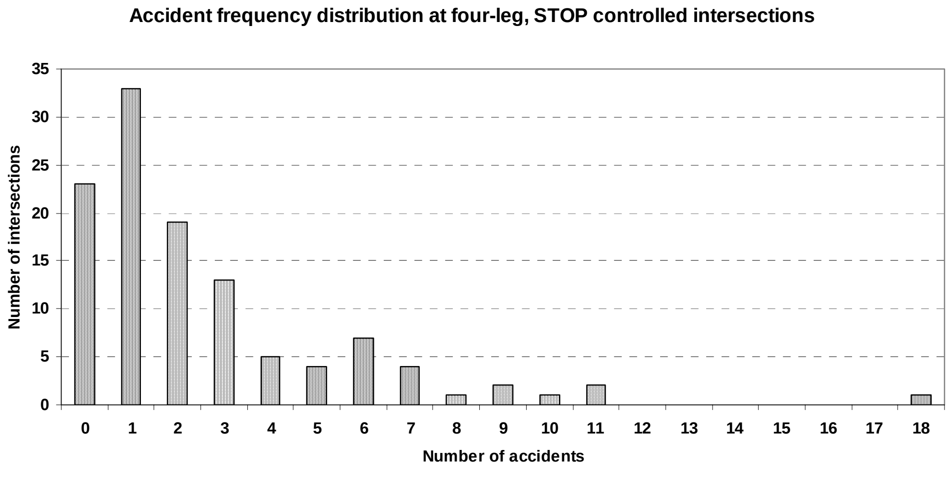

Figure 5: Accident frequency distribution at a sample of 115 four-leg, STOP-controlled intersections

Table 13: Continuous variables - Four-leg, STOP-controlled intersections

Variable	Minimum	Mean	Median	Maximum
Fatal and injury vehicle accidents (2000-2002)	0	2.57	2	18
Major-road ADT (veh/day)	2000	10634.8	9000	50000
Crossroad ADT (veh/day)	2000	2817.4	2000	17500
Number of lanes on major road	1	1.79	2.00	4
Number of lanes on crossroad	1	1.39	1.00	3
Average lane width on major road (m)	1.17	2.94	2.95	6.00
Average lane width on crossroad (m)	1.33	2.97	2.95	6.95
Side-walk width on major road (m)	0.00	1.90	1.75	4.90
Side-walk width on crossroad (m)	0.00	1.52	1.50	3.45

Table 14: Categorical variables - Four-leg, STOP-controlled intersections

Variable	Level	% of intersections
Type of left-turn treatment on major road	No left-turn	25.22
	Without/with exclusive turn lane	74.78
Type of left-turn treatment on crossroad	No left-turn	36.52
	Without/with exclusive turn lane	63.48
Type of right-turn treatment on major road	No right-turn	40.00
	Without/with exclusive turn lane	60.00
Type of right-turn treatment on crossroad	No right-turn	22.61
	Without/with exclusive turn lane	77.39
One-way vs. two way operation on major road	One-way operation	49.57
	Two-way operation	50.43
One-way vs. two way operation on crossroad	One-way operation	65.22
	Two-way operation	34.78
Road markings	Present	53.04
	Absent	46.96
Percent grade on major road	Level	35.65
	Moderate grade	49.57
	Steep grade	14.78
Percent grade on crossroad	Level	51.30
	Moderate grade	33.04
	Steep grade	15.65

Table 15: Model diagnostics for fatal and injury vehicle accidents - Four-leg, STOP-controlled, intersections

	Poisson regression		Lognormal regression		
	Full model	Reduced model (9 variables)	Full model	Reduced model (9 variables)	Reduced model (7 variables)
Number of intersections (n)	115	115	115	115	115
Number of parameters (p)	18	10	18	10	8
Degrees of freedom (df)	20	12	20	12	8
C_p	-	-	24.195	12.253	8.000
Deviance/(n-df)	1.36	1.31	0.92	0.81	0.78
MSE	6.61	6.15	7.94	7.03	6.75
MAE	2.12	1.96	2.19	2.00	1.97
R^2 (%)	34.11	32.59	37.99	35.33	33.70
R^2_{FT} (%)	21.83	25.23	21.63	26.44	26.42

Confronting the two models we notice as the reduction has carried to an improvement of all the goodness-of-fit measures: we have a reduction of the error in the forecast, an increment of R^2_{FT} , and the mean deviance is next to the unit. The hypothesis of a lognormal distribution gives instead the elaboration of three models (table 15): one complete with 17 variables, one reduced to 9 variables characterized from the maximum value of R^2_{FT} , but containing also variables not statistically meaningful to 90%, like the type of left-turn treatment on crossroad (P=22.08%), the “One-way vs. two way operation on major road” (P=29.46%), the percent grade on major road (P=10.91, and one reduced to 7 variables all statistically meaningful to 90%. Between the three models, all the statistical indices suggest the use of the model reduced to 7 variables.. The choice is reduced to the reduced models of Poisson and lognormal.
Observing the indices that characterize them, we notice as the lognormal model has a value of mean deviance slightly more near to the unit; with reference to the error on the forecast, it is the Poisson model to adapt itself better. So the choice of the reduced model of Poisson appears appropriate.
The expected 3-years fatal and injury accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} -5.39761 + 0.275383 \cdot X_1 + 0.444986 \cdot X_2 + 0.178572 \cdot X_3 + 0.818525 \cdot X_4 \\ -0.46902 X_5 + 0.885859 \cdot X_6 - 0.604733 \cdot X_7 - 0.673398 \cdot X_8 - 0.232165 \cdot X_9 \\ -0.604094 \cdot X_{10} - 0.15267 \cdot X_{11} \end{pmatrix} \quad [28]$$

where:
Y = expected number of fatal and injury vehicle accidents in a 3-year period;
X_1 = ADT of the major road;
X_2 = ADT of the crossroad;

X_4 = 1 if permitted left-turn lane is present on the major road; 0 otherwise;
X_5 = 1 if permitted left-turn lane is present on the crossroad; 0 otherwise;
X_6 = 1 if permitted right-turn lane is present on the major road; 0 otherwise;
X_7 = 1 if permitted right-turn lane is present on the crossroad; 0 otherwise;
X_8 = 1 if the percent grade on major road is steep grade; 0 otherwise;
X_9 = 1 if the percent grade on major road is level; 0 otherwise.
X_{10} = 1 if the percent grade on crossroad is steep grade; 0 otherwise;
X_{11} = 1 if the percent grade on crossroad is level; 0 otherwise.

Four-leg, signalized intersections

The accident distribution (figure 10) don't suggest the better method to calculate the regression coefficients; so the model coefficients are estimated both by least-squares method and by maximum likelihood method.
Minimum, mean, median, and maximum values are given for the continuous variables (table 15); for categorical variables (table 16), the percent of intersections within each level is given.
Two models through the least-squares method, assuming a lognormal distribution of the number of incidents, and the maximum likelihood method, assuming a Poisson distribution will come elaborate. For both the procedures, a complete model, with 19 variables, and the reduced model, obtained through the stepwise methodology, will come elaborate (table 17).

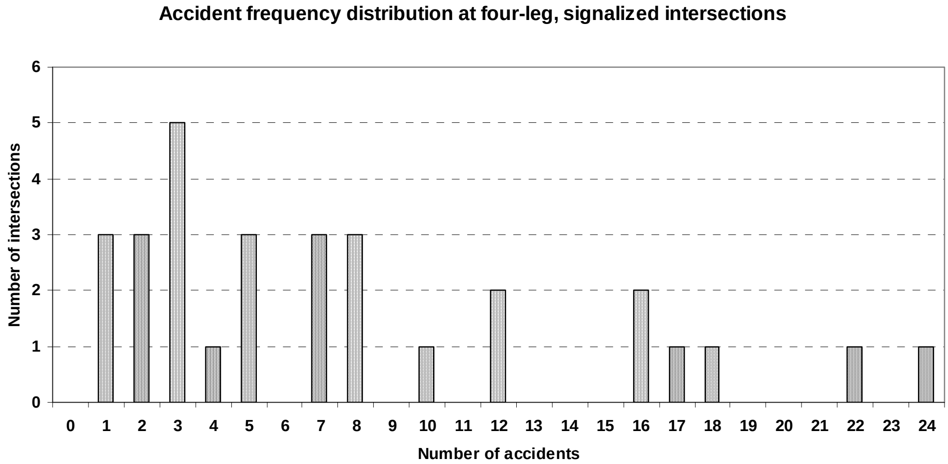

Figure 10: Accident frequency distribution at a sample of 30 four-leg, signalized intersections

Table 15: Continuous variables - Four-leg, signalized intersections

Variable	Minimum	Mean	Median	Maximum
Fatal and injury vehicle accidents (2000-2002)	1	7.83	6	24
Major-road ADT (veh/day)	8500	27433.3	22500	50000
Crossroad ADT (veh/day)	2000	13100	13250	37500
Number of lanes on major road	2	3.47	3	6
Number of lanes on crossroad	1	2.47	2	8
Average lane width on major road (m)	2.17	3.09	2.93	4.50
Average lane width on crossroad (m)	1.80	3.12	3.01	5.75
Side-walk width on major road (m)	1.20	2.79	2.53	6.00
Side-walk width on crossroad (m)	0.00	2.12	1.85	6.50

Confronting the two models we notice as the reduction has carried to an improvement of all the goodness-of-fit measures: we have a reduction of the absolute error in the forecast, an increment of R^2_{FT}, and the mean deviance is next to the unit. The hypothesis of a lognormal distribution gives instead the elaboration of three models (table 17): one complete with 19 variables, one reduced to 10 variables characterized from the maximum value of R^2_{FT}, but containing also variables not statistically meaningful to 90%, like the type of right-turn treatment on major road (P=15.13%), the percent grade on major road (P=27.17%), and one reduced to 7 variables all statistically meaningful to 90 So the choice of the reduced model of Poisson appears appropriate.

Table 16: Categorical variables - Four-leg, signalized intersections

Variable	Level	% of intersections
Presence of median on major road	Absent	76.67
	Present	23.33
Type of left-turn treatment on major road	No left-turn	16.67
	Without exclusive turn lane	53.33
	With exclusive turn lane	30.00
Type of left-turn treatment on crossroad	No left-turn	16.67
	Without exclusive turn lane	76.67
	With exclusive turn lane	6.66
Type of right-turn treatment on major road	No left-turn	16.67
	Without exclusive turn lane	46.67
	With exclusive turn lane	36.66
Type of right-turn treatment on crossroad	No left-turn	13.33
	Without exclusive turn lane	73.33
	With exclusive turn lane	13.34
One-way vs. two way operation on major road	One-way operation	13.33
	Two-way operation	86.67
One-way vs. two way operation on crossroad	One-way operation	23.33
	Two-way operation	76.67
Signal phasing	Two-phase	80.00
	Multiphase	20.00
Road markings	Present	66.67
	Absent	33.33
Percent grade on major road	Level	46.67
	Moderate or steep grade	53.33
Percent grade on crossroad	Level	43.33
	Moderate or steep grade	56.67

Table 17: Model diagnostics for fatal and injury vehicle accidents - Four-leg, signalized intersections

	Poisson regression		Lognormal regression		
	Full model	**Reduced model (9 variables)**	**Full model**	**Reduced model (9 variables)**	**Reduced model (7 variables)**
Number of intersections (n)	30	30	30	30	30
Number of parameters (p)	20	11	20	11	8
Degrees of freedom (df)	24	14	24	14	9
C_p	-	-	20	21.22	33.07
Deviance/(n-df)	1.40	0.97	0.81	0.47	0.49
MSE	6.94	7.43	34.11	19.67	20.85
MAE	4.47	2.77	8.75	3.88	3.66
R^2 (%)	94.28	89.41	91.71	89.06	84.90
R^2_{FT} (%)	61.51	70.29	59.91	80.23	79.14

The expected 3-years fatal and injury accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} -5.63006 + 1.03804 \cdot X_1 - 0.515058 \cdot X_2 - 0.293142 \cdot X_3 - 0.23727 \cdot X_4 \\ -1.4698 \cdot X_5 - 0.98365 \cdot X_6 + 0.737353 \cdot X_7 + 1.14668 \cdot X_8 + 2.06118 \cdot X_9 \\ + 0.341667 \cdot X_{10} + 1.89303 \cdot X_{11} - 1.66943 \cdot X_{12} - 0.528992 \cdot X_{13} \end{pmatrix} \quad [28]$$

where:
Y = expected number of fatal and injury vehicle accidents in a 3-year period;
X_1 = ADT of the crossroad;
X_2 = number of lanes on crossroad;
X_3 = average lane width on crossroad;
X_4 = side-walk width on major road;
X_5 = 1 if protected left-turn lane is present on the crossroad; 0 otherwise;
X_6 = 1 if permitted left-turn lane is present on the crossroad; 0 otherwise;
X_7 = 1 if protected right-turn lane is present on the major road; 0 otherwise;
X_8 = 1 if permitted right-turn lane is present on the major road; 0 otherwise;
X_9 = 1 if protected right-turn lane is present on the crossroad; 0 otherwise;
X_{10} = 1 if permitted right-turn lane is present on the crossroad; 0 otherwise;

X_{11} = 1 if two-way operation is present on major road; 0 otherwise;
X_{12} = 1 if signal phasing is two-phase; 0 otherwise;
X_{13} = 1 if the percent grade on major road is level; 0 otherwise.

STATISTICAL MODELS FOR ACCIDENT TYPOLOGY

Selection of analysis group for accident type

Now the objective of these statistical models is to provide a relationship between the number of accident divided for type and the intersection parameters (geometric design, traffic control, traffic volume). We have been characterized two analysis groups for the elaboration of provisional models (table18, table 19):
1° Group: lateral collision;
2° Group: head-on collision + sideswipe collision+ rear-end collision.

Table 18: Distribution by accident type

Type	Lateral	Head-on	Sideswipe	Rear-end
1	45,12%	7,32%	13,41%	14,63%
2	47,92%	2,08%	12,50%	14,58%
3	73,54%	2,12%	11,64%	6,88%
4	78,26%	1,34%	7,69%	6,69%
5	53,10%	4,42%	10,18%	20,35%

Table 19: Distribution by intersection type

TYPOLOGY	1° GROUP	2° GROUP
three - leg, no control intersections	45,12%	35,37%
three – leg, STOP - controlled intersections	47,92%	28,17%
four - leg, no control intersections	73,54%	20,63%
four – leg, STOP - controlled intersections	78,26%	15,72%
four – leg, signalized intersections	53,10%	34,96%

Three–leg, no-control intersection for 1° group accidents

The accident distribution (figure 11) tends to follow the shape of a Poisson distribution; consequently the model coefficients are estimated by maximum likelihood method.

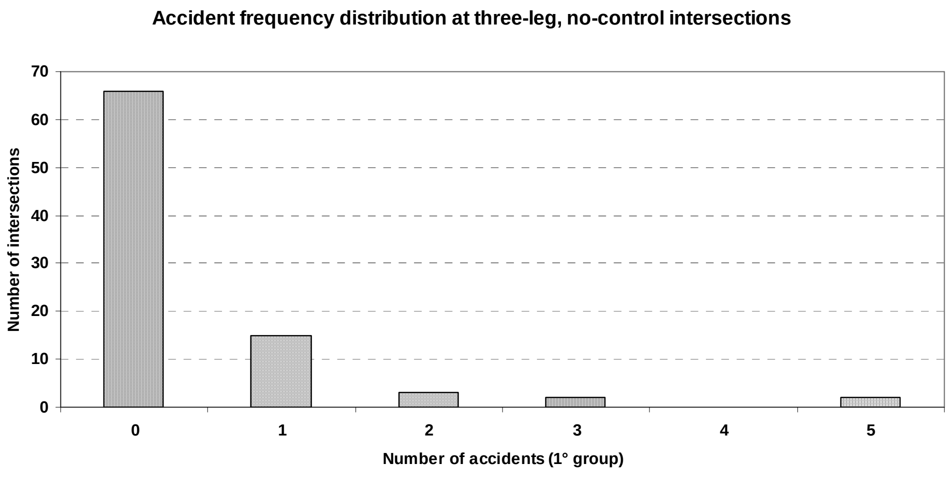

Figure 11: 1° group accident frequency distribution at three-leg, no-control intersections

With the data coming from intersections, a complete model have been elaborate, considering all the variables, and one reduced, obtained from the first one through the stepwise procedure (table 20).
The goodness-of-fit measures show a discreet dependency between the expected number of accidents and the used variables. Reducing the number of variables, a simplification of the model is had and even if the mean deviance goes away from the unit, the index R^2_{FT} goes to 37%. These results provide an indication that the choice of the reduced model appears appropriate.

Table 20: Model diagnostics for 1° group accidents - Three-leg, no-control, intersections

	Poisson regression	
	Full model	Reduced model (7 variables)
Number of intersections (n)	88	88
Number of parameters (p)	19	8
Degrees of freedom (df)	21	10
Deviance/(n-df)	0.29	0.26
MSE	0.40	0.52
MAE	0.42	0.43
R^2 (%)	61.44	54.05
R^2_{FT} (%)	25.80	37.08

The expected 3-years fatal and injury accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix}-4.06796 + 0.739975 \cdot X_1 - 0.620585 \cdot X_2 + 0.765237 \cdot X_3 + 1.01829 \cdot X_4 \\ -0.977124 \cdot X_5 - 15.3041 \cdot X_6 + 0.128352 \cdot X_7 + 1.49253 \cdot X_8 + 1.04362 \cdot X_9\end{pmatrix}$$

where:
Y = expected number of 1°group accidents in a 3-year period;
X_1 = ADT of the major road;
X_2 = ADT of the crossroad;
X_3 = 1 if permitted right-turn lane is present on the crossroad; 0 otherwise;
X_4 = 1 if permitted left-turn lane is present on the crossroad; 0 otherwise;
X_5 = 1 if the road marking is absent; 0 otherwise;
X_6 = 1 if the percent grade on major road is steep grade; 0 otherwise;
X_7 = 1 if the percent grade on major road is level; 0 otherwise;
X_8 = 1 if the percent grade on crossroad is steep grade; 0 otherwise;
X_9 = 1 if the percent grade on crossroad is level; 0 otherwise.

Three–leg, no-control intersection for 2° group accidents

The accident distribution (figure 12) tends to follow the shape of a Poisson distribution; consequently the model coefficients are estimated by maximum likelihood method.

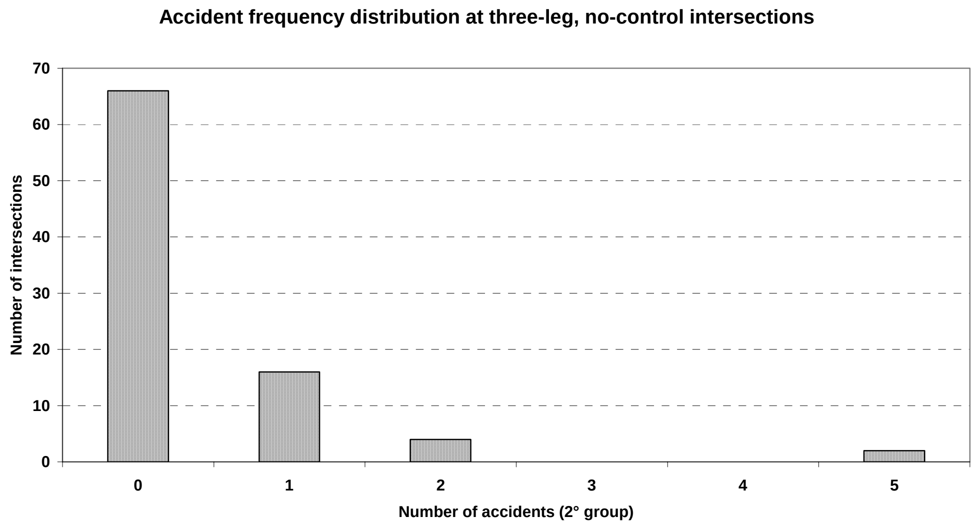

Figure 12: 2° group accident frequency distribution at three-leg, no-control intersections

Table 21: Model diagnostics for 2° group accidents - Three-leg, no-control, intersections

	Poisson regression	
	Full model	Reduced model (7 variables)
Number of intersections (n)	88	88
Number of parameters (p)	19	8
Degrees of freedom (df)	21	10
Deviance/(n-df)	0.10	0.14
MSE	0.38	0.33
MAE	0.37	0.36
R^2 (%)	49.01	46.23
R^2_{FT} (%)	15.26	28.46

With the data coming from intersections, a complete model have been elaborate, considering all the variables, and one reduced, obtained from the first one through the stepwise procedure (table 21).
The goodness-of-fit measures show a discreet dependency between the expected number of accidents and the used variables. Reducing the number of variables, a simplification of the model is had; the mean deviance goes next the unit, the error of forecast is reduced and the index R^2_{FT} goes to 28.46%. These results provide an indication that the choice of the reduced model appears appropriate.
The expected 3-years fatal and injury accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} -7.19026 + 0.609891 \cdot X_1 - 0.61826 \cdot X_2 - 1.24818 \cdot X_3 - 0.724402 \cdot X_4 \\ -1.00929 \cdot X_5 + 0.4906 \cdot X_6 + 1.20703 \cdot X_7 + 1.51759 \cdot X_8 + 0.316618 \cdot X_9 \end{pmatrix}$$

where:
Y = expected number of 1°group accidents in a 3-year period;
X_1 = ADT of the major road;
X_2 = side-walk width on major road;
X_3 = 1 if two-way operation is present on major road; 0 otherwise;
X_4 = 1 if permitted right-turn lane is present on the major road; 0 otherwise;
X_5 = 1 if the road marking is absent; 0 otherwise;
X_6 = 1 if the percent grade on major road is steep grade; 0 otherwise;
X_7 = 1 if the percent grade on major road is level; 0 otherwise;
X_8 = 1 if the percent grade on crossroad is steep grade; 0 otherwise;
X_9 = 1 if the percent grade on crossroad is level; 0 otherwise.

Three-leg, STOP controlled intersections

The study carried out on the basis of a statistic analysis, for this group of intersection, don't justify the elaboration of a model. In fact only 23 incidents have happened for lateral collision; moreover also the number of intersections characterized from this type of incidents are evidently limited. Analogous to the previous case, the number of incidents, for head-on+sideswipe+rear-end collision, to analyze is equal to 14, not enough for the elaboration of a model.

Four-leg, no - control intersections for 1° group accidents

The accident distribution (figure 13) tends to follow the shape of a Poisson distribution; consequently the model coefficients are estimated by maximum likelihood method.

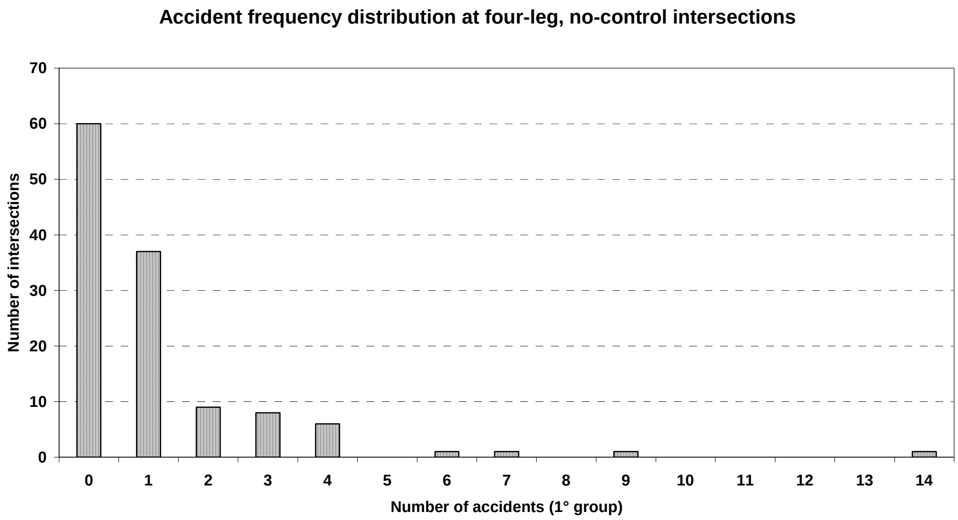

Figure 12: 1° group accident frequency distribution at four-leg, no-control intersections

With the data coming from intersections, a complete model have been elaborate, considering all the variables, and one reduced, obtained from the first one through the stepwise procedure (table 21).
The goodness-of-fit measures show a discreet dependency between the expected number of accidents and the used variables. Reducing the number of variables, a simplification of the model is had; the mean deviance goes away from the unit, the error of forecast is reduced and the index R^2_{FT} is increased These results provide an indication that the choice of the reduced model appears appropriate.
The expected 3-years fatal and injury accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} -5.61551 + 0.490974 \cdot X_1 + 0.51329 \cdot X_2 + 0.24207 \cdot X_3 + 0.388853 \cdot X_4 \\ -0.53136 \cdot X_5 - 0.510431 \cdot X_6 - 0.0248862 \cdot X_7 + 0.580791 \cdot X_8 \end{pmatrix}$$

where:
Y = expected number of 1°group accidents in a 3-year period;

X_1 = ADT of the crossroad;
X_2 = average lane width on crossroad;
X_3 = side-walk width on major road;
X_4 = 1 if two-way operation is present on major road; 0 otherwise;
X_5 = 1 if two-way operation is present on crossroad; 0 otherwise;
X_6 = 1 if permitted left-turn lane is present on the major road; 0 otherwise;
X_7 = 1 if the percent grade on crossroad is steep grade; 0 otherwise;
X_8 = 1 if the percent grade on crossroad is level; 0 otherwise.

Table 22: Model diagnostics for 1° group accidents - Four-leg, no-control, intersections

	Poisson regression	
	Full model	**Reduced model (7 variables)**
Number of intersections (n)	124	124
Number of parameters (p)	18	8
Degrees of freedom (df)	20	9
Deviance/(n-df)	0.77	0.71
MSE	2.62	2.45
MAE	1.18	1.10
R^2 (%)	33.36	30.42
R^2_{FT} (%)	18.79	23.87

Four-leg, no - control intersections for 2° group accidents

The study carried out on the basis of a statistic analysis, for this group of intersection, don't justify the elaboration of a model. In fact only 47 incidents have happened for head-on+sideswipe+rear-end collision; moreover also the number of intersections characterized from this type of incidents are evidently limited.

Four-leg, STOP controlled intersections for 1° group accidents

The accident distribution (figure 13) tends to follow the shape of a Poisson distribution; consequently the model coefficients are estimated by maximum likelihood method.

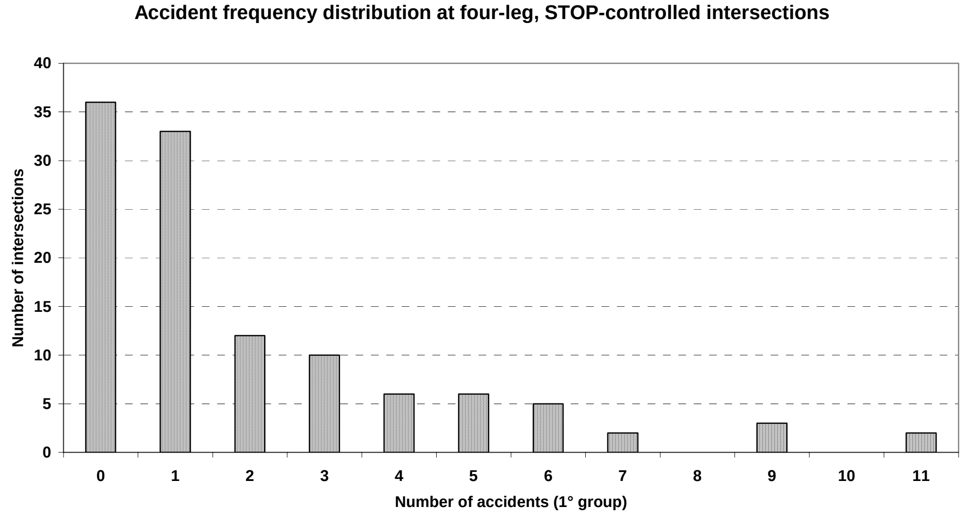

Figure 13: 1° group accident frequency distribution at four-leg, STOP-controlled intersections

Table 22: Model diagnostics for 1° group accidents - Four-leg, stop-controlled intersections

	Poisson regression	
	Full model	**Reduced model (7 variables)**
Number of intersections (n)	115	115
Number of parameters (p)	18	10
Degrees of freedom (df)	20	12
Deviance/(n-df)	2.70	2.48
MSE	9.43	8.47
MAE	2.55	2.28
R^2 (%)	16.42	14.57
R^2_{FT} (%)	5.04	7.74

With the data coming from intersections, a complete model have been elaborate, considering all the variables, and one reduced, obtained from the first one through the stepwise procedure (table 22). The goodness-of-fit measures show a discreet dependency between the expected number of accidents and the used variables. Reducing the number of variables, a simplification of the model is had; the mean deviance goes next the unit, the error of forecast is reduced and the index R^2_{FT} goes to 7.75%. These results provide an indication that the choice of the reduced model appears appropriate.
The expected 3-years fatal and injury accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} 0.703053 + 0.240907 \cdot X_1 - 0.356302 \cdot X_2 + 0.594643 \cdot X_3 + 0.459024 \cdot X_4 \\ -0.375682 \cdot X_5 - 0.502751 \cdot X_6 - 0.300335 \cdot X_7 - 0.12291 \cdot X_8 + 0.388077 \cdot X_9 + 0.39133 \cdot X_{10} + 0.431922 \cdot X_{11} \end{pmatrix}$$

where:
Y = expected number of 1°group accidents in a 3-year period;
X_1 = ADT of the major road;
X_2 = ADT of the crossroad;
X_3 = 1 if permitted right-turn lane is present on the major road; 0 otherwise;
X_4 = 1 if permitted right-turn lane is present on the crossroad; 0 otherwise;
X_5 = 1 if permitted left-turn lane is present on the crossroad; 0 otherwise;
X_6 = 1 if two-way operation is present on crossroad; 0 otherwise;
X_7 = 1 if road marking is absent; 0 otherwise;
X_8 = 1 if the percent grade on major road is steep grade; 0 otherwise.
X_9 = 1 if the percent grade on major road is level; 0 otherwise.
X_{10} = 1 if the percent grade on crossroad is steep grade; 0 otherwise;
X_{11} = 1 if the percent grade on crossroad is level; 0 otherwise.

Four-leg, STOP controlled intersections for 2° group accidents

The study carried out on the basis of a statistic analysis, for this group of intersection, don't justify the elaboration of a model. In fact only 47 incidents have happened for head-on+sideswipe+rear-end collision; moreover also the number of intersections characterized from this type of incidents are evidently limited.

Four-leg, signalized intersections for 1° group accidents

The accident distribution (figure 14) don't suggest the better method to calculate the regression coefficients; so the model coefficients are estimated both by least-squares method and by maximum likelihood method.

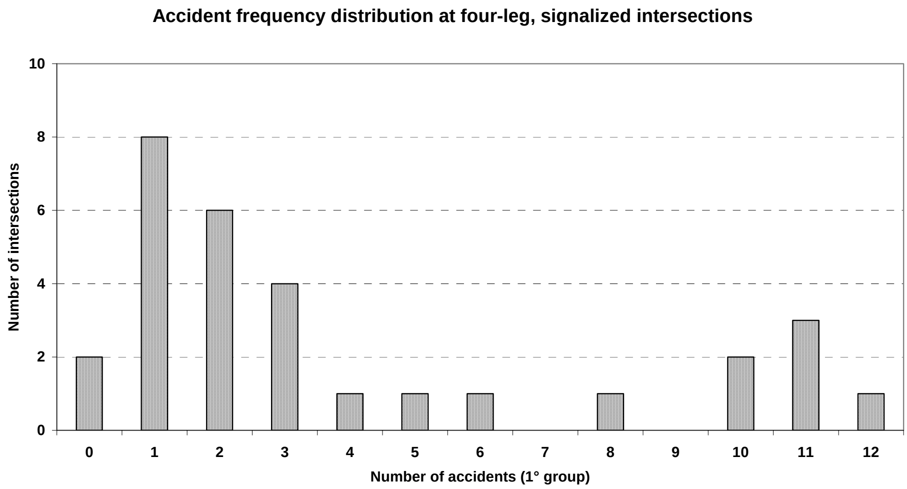

Figure 14: 1° group accident frequency distribution at four-leg, signalized intersections

Two models through the least-squares method, assuming a lognormal distribution of the number of incidents, and the maximum likelihood method, assuming a Poisson distribution will come elaborate. For both the procedures, a complete model, with 19 variables, and the reduced model, obtained through the stepwise methodology, will come elaborate(table 23).
Confronting the two models we notice as the reduction has carried to an improvement of all the goodness-of-fit measures: we have a reduction of the absolute error in the forecast, an increment of R^2_{FT}, and the mean deviance is next to the unit. The hypothesis of a lognormal distribution gives instead the elaboration of three models (table 23): one complete with 19 variables, one reduced to 12 variables characterized from the maximum value of R^2_{FT}, but containing also variables not statistically meaningful to 90%, and one reduced to 11 variables all statistically meaningful to 90%. So the choice of the reduced model of Poisson appears appropriate.

Table 23: Model diagnostics for 1° group accidents - Four-leg, signalized intersections

	Poisson regression		Lognormal regression		
	Full model	**Reduced model (9 variables)**	**Full model**	**Reduced model (9 variables)**	**Reduced model (7 variables)**
Number of intersections (n)	30	30	30	30	30
Number of parameters (p)	20	12	20	13	12
Degrees of freedom (df)	24	14	24	16	15
C_p	-	-	20.00	27.68	30.31
Deviance/(n-df)	2.74	1.14	0.35	0.28	0.62
MSE	9.34	4.13	5.30	4.23	9.34
MAE	5.29	2.29	3.72	1.98	2.68
R^2 (%)	77.20	73.05	91.70	89.09	84.89
R^2_{FT} (%)	30.25	45.67	59.91	80.23	79.14

The expected 3-years accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} -3.70536 + 0.708817 \cdot X_1 - 0.627562 \cdot X_2 - 0.342992 \cdot X_3 + 0.309049 \cdot X_4 + 0.03711 \cdot X_5 \\ +1.43171 \cdot X_6 + 1.25578 \cdot X_7 - 0.99734 \cdot X_8 - 2.72587 \cdot X_9 + 0.697059 \cdot X_{10} + 0.545366 \cdot X_{11} \\ +2.08928 \cdot X_{12} - 0.804072 \cdot X_{13} \end{pmatrix}$$

where:
Y = expected number of 1°group accidents in a 3-year period;
X_1 = ADT of the crossroad;
X_2 = number of lanes on crossroad;
X_3 = side-walk width on major road;
X_4 = side-walk width on crossroad;
X_5 = 1 if protected right-turn lane is present on the major road; 0 otherwise;
X_6 = 1 if permitted right-turn lane is present on the major road; 0 otherwise;
X_7 = 1 if protected left-turn lane is present on the major road; 0 otherwise;
X_8 = 1 if permitted left-turn lane is present on the major road; 0 otherwise;
X_9 = 1 if signal phasing is two-phase; 0 otherwise;
X_{10} = 1 if median is absent; 0 otherwise;
X_{11} = 1 if road marking is absent; 0 otherwise;
X_{12} = 1 if two-way operation is present on major road; 0 otherwise;
X_{13} = 1 if the percent grade on major road is level; 0 otherwise.

Four-leg, signalized intersections for 2° group accidents

The accident distribution (figure 15) tends to follow the shape of a Poisson distribution; consequently the model coefficients are estimated by maximum likelihood method.
With the data coming from intersections, a complete model have been elaborate, considering all the variables, and one reduced, obtained from the first one through the stepwise procedure (table 24).
The goodness-of-fit measures show a discreet dependency between the expected number of accidents and the used variables. Reducing the number of variables, a simplification of the model is had; the mean deviance goes next the unit, the error of forecast is reduced and the index R^2_{FT} is increased. These results provide an indication that the choice of the reduced model appears appropriate.

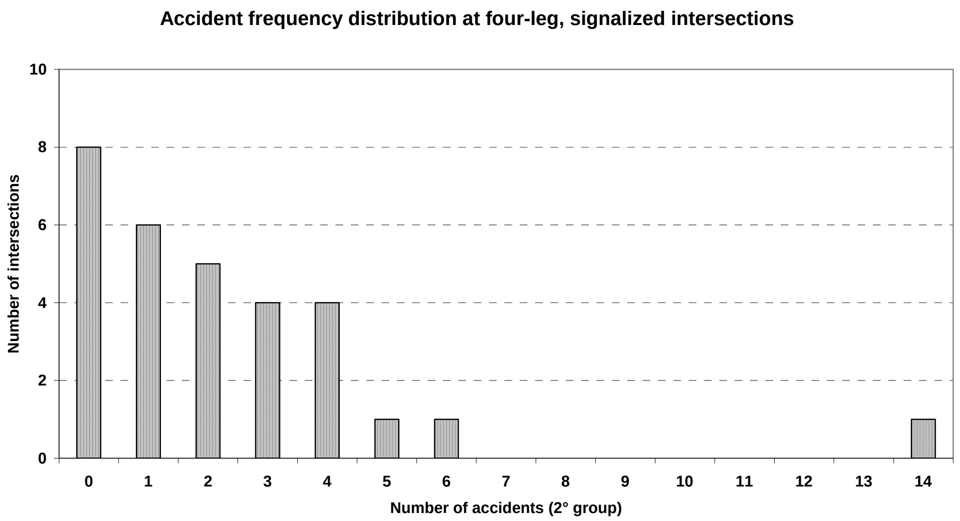

Figure 14: 2° group accident frequency distribution at four-leg, signalized intersections

Table 24: Model diagnostics for 2° group accidents - Four-leg, signalized intersections

	Poisson regression	
	Full model	**Reduced model (7 variables)**
Number of intersections (n)	30	30
Number of parameters (p)	18	12
Degrees of freedom (df)	24	16
Deviance/(n-df)	0.19	0.20
MSE	0.57	0.61
MAE	1.18	0.82
R^2 (%)	96.75	92.20
R^2_{FT} (%)	34.53	50.72

The expected 3-years accident frequency can be estimated using the model as:

$$Y = \exp\begin{pmatrix} -6.5373 + 0.896734 \cdot X_1 + 0.659043 \cdot X_2 - 1.35579 \cdot X_3 - 0.828998 \cdot X_4 - 0.785289 \cdot X_5 \\ + 8.99581 \cdot X_6 + 10.414 \cdot X_7 + 5.11448 \cdot X_8 + 7.64187 \cdot X_9 - 1.50268 \cdot X_{10} - 3.43997 \cdot X_{11} \\ - 10.7409 \cdot X_{12} - 8.53979 \cdot X_{13} + 0.922822 \cdot X_{14} + 1.02627 \cdot X_{15} \end{pmatrix}$$

where:
Y = expected number of 2°group accidents in a 3-year period;
X_1 = ADT of the crossroad;
X_2 = number of lanes on crossroad;
X_3 =average lane width on major road;
X_4 = average lane width on crossroad;
X_5 = side-walk width on crossroad;
X_6 = 1 if protected right-turn lane is present on the major road; 0 otherwise;
X_7 = 1 if permitted right-turn lane is present on the major road; 0 otherwise;
X_8 = 1 if protected right-turn lane is present on the crossroad; 0 otherwise;
X_9 = 1 if permitted right-turn lane is present on the crossroad; 0 otherwise;
X_{10} = 1 if protected left-turn lane is present on the major road; 0 otherwise;
X_{11} = 1 if permitted left-turn lane is present on the major road; 0 otherwise;
X_{12} = 1 if protected left-turn lane is present on the crossroad; 0 otherwise;
X_{13} =.1 if permitted left-turn lane is present on the crossroad; 0 otherwise;
X_{14} = 1 if road marking is absent; 0 otherwise;
X_{15} = 1 if the percent grade on crossroad is level; 0 otherwise.

CONCLUSIONS

With the present study we reach to the formulation of analytical models for the forecast of the urban accidents with reference to the various typologies of road intersections. In synthesis, it can be asserted that: 1) the Poisson reduced models are those that better describe the accidentality level of the urban intersections; 2) the variables existing in every model are not always the same ones (they are distinguished in relation to the junction type); 3) the variable "traffic" is not always predominant; 4) the previsional models,, separated for accident type, have a inferior reliability level regarding the models that estimate the total number of accidents. We suppose that the proposed relations can constitute a valid support for the characterization of the riskiness level of the urban junctions, especially in: 1) verification of the really conditions of risk in the intersections (that can happen through the comparison between the expected accidents and happened accidents); 2) optimization of the design elements and regulation of the new intersections (already in design phase can be studied the more opportune solutions in order to reduce the risk level of a urban intersection).

REFERENCES

AA.VV. (1992), "Relationships between road accidents and traffic flows in an urban network", Traffic Engineering Control.

BAUER, K. M., HARWOOD, D. W. (1996), "Statistical Models of At – grade Intersection Accidents", U.S. Department of Transportation, Federal Highway Administration, FHWA-RD-96-125.

BRUDE, U., HEDMAN, K., LARSSON, J. (1998), "Design of major urban junction accident prediction and empirical comparisons", VTI EC Research 3.

BRUDE, U., HEDMAN, K., LARSSON, J., THURESSON, L. (1998), "Design of major urban junction comprehensive report", VTI EC Research 2.

FRIDSTRØM, L., IFVER, J., INGEBRIGTSEN, S., KULMALA, R., THOMSEN, K.L. (1995), "Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts", *Accident Analysis and Prevention*, Vol.27, N°1.

LAYFIELD, R.E., SUMMERSGILL, I., HALL, R.D., CHATTERJEE, K. (1996), "Accidents at urban priority crossroads and staggered junctions", TRL Report 185.

LEONARDI, S., PAPPALARDO, G. (2003), "Un Sistema Informativo per l'analisi di sicurezza delle intersezioni stradali urbane", Le Strade, N° 6.