The procedure of statistical evaluation of accident rate equation for roundabouts

A N Kloyan, A M Byrgonytdinov, O A Fedoseeva, G B Lyalkina and L V Yashmanova

Perm National Research Polytechnic University, 29, Komsomolsky ave., Perm, 614990, Russia

E-mail: Anyutka.sh@mail.ru

Abstract. Road Design Manuals propose the procedure for determining the accident rate at modern roundabouts by the probable number of major accident of types on specific sites of a roundabout per year. An assumption is made that the regression equation may not take into account the design features of modern roundabouts, climatic conditions, as well as driving style in the Russian Federation. The paper presents a step-by-step method of the regression equation validation to assess the accident rate at the approaches to the roundabouts on the basis of other statistical methods. The validation involves the statistical evaluation of the sample correlation coefficients between the accident equation and the factors affecting it. Within the work the values included in the equation were non-dimensionalized, necessary assumptions were introduced, and the features of performing correlation analysis in the particular case were analyzed. The procedure proposed enables to determine the reliability of the equation for determining the accident rate at the approaches to a modern roundabout in the conditions of the Russian Federation, taking into account climatic conditions, design features, the state of the existing road transport network, as well as driving style.

1. Introduction
The problem of reducing road accident rates is urgent worldwide [1–3]. According to the World Health Organization (WHO) report the road traffic accidents are the major cause of deaths for children and young adults aged 15–25 years. Improving road safety, aimed at protecting people's lives and health, is the main priority of the Russian Federation transport policy [4, 5].

The main percentage of traffic accidents occurs at road intersections, due to the concentration of the largest number of conflict points. While studying the world practice on ways of improving road safety at road intersections, the modern roundabouts were found to be among the most popular and safest types of road intersection.

The Russian Federation is only beginning to increase the number of modern roundabouts therefore the design standards are being updated¹. However, the adaptation of many aspects of design and reconstruction for the conditions of our country requires significant labor costs. For example, until now no one has studied in detail the methodology for calculating accidents at modern roundabouts.

To evaluate the road safety conditions¹ at modern roundabout, it is recommended to use the procedure designed by the Department of Main Roads of Australia in 2006. The procedure is given in detail in the road design manual and allows determining the accident rate for the major accident types on specific parts of a roundabout per year [6]. The regression equation proposed by the authors,
however, may not take into account the design features of modern roundabouts, climatic conditions, as well as driving style in the Russian Federation. Therefore, in order to specify the adequacy of the proposed procedure, the regression equation requires additional checking for reliability.

2. Characteristics of regression equation

In the Design Standards the intersection accident rate is characterized by the traffic safety indicator $K_s$, as the number of accidents per 10 million cars that passed through one intersection. According to the procedure [6], it is recommended to assess the safety of a modern roundabout considering the accident rate indicators at the approaches including entry and exit curves to the roundabout by studying the number of road accidents per year.

The procedure [6] gives the empirical equation for an estimation of accident rate on approaches to a roundabout expressed by the formula:

$$ A_i = 1.81 \times 10^{-18} \times N_r^{1.39} \times N_k^{0.65} \times V_a^{4.77} \times n^{2.31}, $$

where $N_r$ – a number of approaching vehicles in one lane, (veh/day); $N_k$ – the average annual daily traffic on the circulating carriageway adjacent to the approach (veh/d); $V_a$ – 85th percentile speed on the entry curve (km/h); $n$ is the number of lanes on the particular roundabout approach.

Equation (1) is a regression equation, where $A_i$ – the response function of the factors $N_r$, $N_k$, $V_a$ and $n$.

The work, however, does not give clear evidence [6] at what values of the significance level and at what confidence intervals this equation can be used for exponents. Concerning this in order to validate the proposed equation (1), it is necessary to verify its validity on the basis of other statistical methods [7, 8]. First of all, it is necessary to perform a statistical evaluation of the sample correlation coefficients between the accident rate equation and the factors affecting it.

Previously, before using the empirical equation for assessing the accident rate on the approaches to the roundabouts, it is necessary to carry out the procedure of non-dimensionalizing the values included in it.

It allows moving from a physical problem to an abstract mathematical problem. This technique enables to apply formal mathematical methods for solving equations. In addition, non-dimensionalizing the quantities provides a way for moving from physical quantities, which can have very large or very small values in modulus, to their dimensionless analogues, which will take values of the order of units. This, in turn, ensures a much slower error accumulation.

In formula (1) the value which requires the procedure of non-dimensionalizing is $V_a$ – the 85th percentile speed on the entry curve (km/h).

The most suitable value for its non-dimensionalizing is the traffic speed limits on the particular section $V_{orp}$. Dividing the value $V_a$ by $V_{orp}$, we obtain the value $V$, which is dimensionless and optimally determines the accident rate under consideration. Then formula (1) will be as follows:

$$ A_i = 1.81 \times 10^{-18} \times N_r^{1.39} \times N_k^{0.65} \times V^{4.77} \times n^{2.31}, $$

where $V = V_a / V_{orp}$.

Take the logarithm of equation 2, to get its linear form. The expression will be as follows:

$$ \log(A_i) = \log(1.18) - 18 + 1.39 \log(N_r) + 0.65 \log(N_k) + 4.77 \log(V) + 2.31 \log(n), \text{ i.e. equation} $$

$$ \log(A_i) = -17.93 + 1.39 \log(N_r) + 0.65 \log(N_k) + 4.77 \log(V) + 2.31 \log(n), $$

It is necessary to note that response function $Y = \log(A_i)$ has a linear form relative to factors $X^{(1)} = \log(N_r)$, $X^{(2)} = \log(N_k)$, $X^{(3)} = \log(V)$, $X^{(4)} = \log(n)$.

As a result, the equation linking the accident rate indicator $Y$ at the roundabout with explanatory variables $X^{(i)} (i = 1, 2, 3, 4)$ will take the following form:
\[ Y = -17.93 + 1.39X^1 + 0.65X^2 + 4.77X^3 + 2.31X^4. \] (4)

It is necessary to note that the analysis of multiple correlation relationships (statistical relationships between more than two variables) is associated with the need to measure the strength of relationship between the dependent variable \( Y \) and independent variables \([8]\), as well as with emerging difficulties in interpreting paired correlation coefficients \( r_{Y,X^i} \) between \( Y \) and \( X^i \), due to the possible indirect influence of others on this pairing (obviously not taken into account in the calculation of the independent variables). Therefore, in the general case, when selecting regression equations the number of independent variables may increase \( X^{(i)} \) \((i = 1, 2, 3,\ldots, k)\).

Multiple correlation analysis requires the introduction of such indicators of the statistical relationships that would give the correlation degree evaluation of each of the pairs \((Y, X^i)\) for the variables we are interested in. These indicators must be "cleared" of indirect influence of other variables, that is, they would be calculated provided that the values of the other variables are fixed.

The search for a multiple regression equation begins with the evaluation of paired correlation coefficients for many factors \((i = 1, 2, 3,\ldots, k)\) and the resulting response function \( Y, r_{Y,X^i} \).

3. Validation procedure of the reliability of the accident rate assessment

Assuming that the multidimensional distribution law of the random variables \( X^1, X^2,\ldots, X^k \) is normal, we denote by \( r_{Y,X^i} = r_{Y,X^i(X^2,\ldots,X^k)} \) a sample partial correlation coefficient between the response function \( Y \) and the variable \( X^i \) for constant values of the variables \( X^1, X^2,\ldots, X^k \). Similarly (if \( i > 1 \)) through \( r_{Y,X^i} \) we will define the sample value of the partial correlation coefficient between \( Y \) and another factor \( X^i \), at constant values of the remaining factors \( X^{i-1}, X^{i+1},\ldots, X^k \) when \( i \geq 2 \) \([8]\).

While studying the properties of the selective partial correlation coefficient \( r_{Y,X^i(X^1,\ldots,X^{i-1},X^{i+1},\ldots,X^k)} \) \(k\)-th order we assume, as usual \([8]\) that it is statistically distributed in exactly the same way as a normal (paired) sample correlation coefficient between the same variables \( Y \) and \( X^i \). Naturally, it is necessary to introduce a correction for the number of degrees of freedom of the sampling, reducing its volume in \( k \) units, i.e. to take the appropriate number of degrees of freedom equal to \( n_i - k \), where \( n_i \) – the size of each of the analyzed sample sets of values \( X^{(i)} \) \((i = 1, 2,\ldots, k)\).

To determine the accident rate, the peculiarities of its calculation on the approaches to the roundabout, as well as to validate the empirical relationships with necessary assumptions we propose the procedure enabling statistically reliable at a given \( \alpha \) to identify the strength of relationship between the dependent variable \( Y \) and multiple independent variables \( X^{(i)} \) \((i = 1, 2,\ldots, k)\).

The main steps of calculation include:
- the calculation of the sample correlation coefficients \( r_{Y,X^i(X^1,\ldots,X^{i-1},X^{i+1},\ldots,X^k)} \) for each of the \( k \) independent variables \( X^i \);
- analysis of the statistical properties of sample correlation coefficients \( r_{Y,X^i(X^1,\ldots,X^{i-1},X^{i+1},\ldots,X^k)} \);
- evaluation of multiple correlation coefficient \( R_{Y,X} \) between the dependent variable \( Y \) and the vector \( \overline{X} = \{\overline{X'}, \overline{X^2},\ldots, \overline{X^k}\} \).

3.1 Step 1

Calculation of the sample (empirical) value of partial correlation coefficients is performed on the basis of the recurrent ratio \([8]\):
In particular, equation 5 includes four dependent variables $X^1$, $X^2$, ..., $X^4$ therefore requires calculation of the four sample correlation coefficients.

### 3.2 Step 2

When studying the statistical properties of the sample partial correlation coefficient of order $k$ we assume that it is distributed in the same way as the usual (paired) sample correlation coefficient between the same variables. It is necessary to introduce an amendment, reducing the sample volume in $k$ units, i.e. accept it to be equal to $n - k$. For small values of $n$ and $r$, close to ±1, you should take into account that the value $\hat{r}$ is a biased evaluation of their theoretical $r$ values.

Knowledge of the probability distribution law $\hat{r}$ enables to determine the adequate value of the sample correlation coefficient for a statistically valid conclusion about the presence of correlation between the studied variables.

For small values of $|r|$ there is a relatively good degree of approximation of the normal distribution, as a result a simple criterion for testing hypotheses about the lack of correlation between the studied variables is accepted:

$$H_0: r = 0.$$  

The approximate distribution of statistics according to the Student's probability distribution law with $n - 2$ degrees of freedom is taken into account together.

It follows that the hypothesis $H_0$ is rejected, with probability of a mistake being equal to $\alpha$ following the inequality (6):

$$\left|\hat{r}\right| \left((n-2-k)^{1/2}\right)(1-\hat{r}^2)^{-1/2} > t_{n-2-k}(n-2-k),$$  

where $t_{n-2-k}$ is a $100q\%$ point $t$ distribution with freedom degrees $m$.

Using the Fisher transformation (7), confidence intervals are calculated for the true value of the correlation coefficient:

$$z_{1,2} = \arctanh(\hat{r}) \pm \left(n-k-3\right)^{-1/2} - \hat{r}(2(n-k-1))^{-1},$$  

it follows that the true value of the correlation coefficient $r$ with confidence probability $1-\alpha$ is concluded within:

$$\text{th} z_1 < r < \text{th} z_2,$$

where $z$ is the hyperbolic tangent of the argument $z$, calculated by the formula:

$$z = \frac{e^z - e^{-z}}{e^z + e^{-z}}.$$  

Finding $z$ is performed according to the table (Appendix 1 [8]), where the extreme columns show the values $|\hat{r}|$, and the corresponding values $|z|$ are between them. It should be remembered that the signs of the argument and the function are the same.

It follows that the values of correlation coefficients $r_{y,x^v}$ with probability $p = 1-\alpha$ will lie within the confidence interval:

$$\left(\text{th}(z-m_z); \text{th}(z+m_z)\right).$$

The value of the hyperbolic tangent from the argument is determined using statistical tables [8].

### 3.3 Step 3

The multiple correlation coefficient $R_{y,X}$ is used as a measure of the statistical relationship strength between the dependent variable $Y$ by the set of independent variables $X^{(i)}$ ($i = 1, 2, ..., k$).
The calculation of the multiple correlation coefficient $R_{Y,X}$ is performed using partial selective correlation coefficients $r_{Y',X''(x',...,x'')}^{i}$ according to formula (9):

$$R_{Y,X}^2 = 1 - (1 - r_{Y',X'}^{2})(1 - r_{Y',X''(x',...,x'')}^{2})\cdots(1 - r_{Y',X''(x',...,x'')}^{2})\cdots(1 - r_{Y',X''(x',...,x'')}^{2}) \cdot (9).$$

To assess the statistical significance of the difference in the value of the multiple correlation coefficient $R_{Y,X}$ from zero with a given significance level $\alpha$ we consider the following hypothesis $H_0$:

$$H_0 : R_{Y,X}^2 = 0.$$

Then the competing hypothesis $H_1$ has the form:

$$H_1 : R_{Y,X}^2 \neq 0.$$

In order to test the hypothesis $H_0$ about the insignificance of the multiple correlation coefficient $R_{Y,X}$ we use the Fisher criterion. The experimental value $F_{op}$ of the Fisher criterion is calculated by the formula:

$$F(\hat{R}) = \frac{R_{Y,X}^2}{(1 - R_{Y,X}^2)} \cdot \frac{(n - k - 1)k^{-1}}{(n-1)} \cdot (10).$$

and at a given level of significance, it is compared with its critical value $F_{kr} = F(k, n - k - 1)$ determined according to the statistical tables [7, 8].

4. Discussion of the results

Within the work, the analysis of the regression equation is performed, necessary assumptions are introduced and the features of the correlation analysis implementation in a particular case are analyzed.

On the basis of the conducted research using multiple and partial correlation coefficients the step-by-step procedure for the correlation analysis of the linear dependence of the accident rate at approaches to a modern roundabout ($A_i$) from the independent variables is developed and presented. The independent variables include a number of approaching vehicles in one lane ($N_r$); the average annual daily traffic on the circulating carriageway adjacent to the approach ($N_k$); 85th percentile speed on the entry curve ($V_\alpha$); the number of lanes on the particular roundabout approach.

Correlation analysis of the equation for finding the accident rate at the approaches to modern roundabouts according to the proposed method enables to determine the reliability of this equation in the conditions of the Russian Federation, taking into account climatic conditions, design features, the state of the existing road transport network, as well as driving style.

References

[1] Gates T J and Maki R E 2001 Converting old traffic circles to modern roundabouts (Michigan: Michigan State University)

[2] Laureshyn A, Svensson A and Hyden C 2010 Evaluation of traffic safety, based on micro-level behavioral data Accident Analysis and Prevention 42 1637–46

[3] Roundabouts 2010 An Information Guide 2rd ed (Washington: National Research Council)

[4] Burgenutdinov A M, Jushkov B S and Okuneva A G Organization and safety of traffic on roads (Perm: Perm National Research Polytechnic University)

[5] Lipnitsky A S 2009 Modern roundabouts (Irkutsk: Irkutsk national research technical university)

[6] Roundabouts 2006 Road planning and design manual ch 14 (Australia: Department of Main Roads) pp 67–87

[7] Lyalkina G B and Berdyshev O V 2014 Mathematical processing of the experimental results: training manual Sovremennye probl. nauki i obraz. 3 180

[8] Aivazian S A and Mkhitarian V S 2001 Probability theory and applied statistics vol 2 (Moscow: Unity)