PERIODICA POLYTECHNICA SER. CIV. ENG. VOL. 46, NO. 1, PP. 71–82 (2002)

PROBLEMS OF RELIABILITY AND EXACTITUDE OF TRAFFIC SURVEYS' DATA

Zsuzsanna TÓTH

Department of Highway and Railway Engineering
Budapest University of Technology and Economics
H–1521 Budapest, Hungary

Received: Oct. 10, 2001

Abstract

The paper dealt with the important problems of exactitude and reliability of composite traffic parameters. To collect traffic data, the representative sampling method is applied actually in traffic surveys in Hungary. Nearly all parameters characterising the traffic are determined as products of different components. Depending upon the methodology applied in the survey, each component has its own exactitude and reliability, An appropriate methodology is presented allowing to calculate for all traffic parameters the error of the estimation and the weight attributed to it, being proportional to reliability.

Keywords: exactitude, reliability, traffic parameters.

1. Introduction

Since the start of traffic surveys aiming at characterizing the road traffic conditions, identification and determination of appropriate parameters suitable for that purpose is considered to be a crucial problem.

While it is commonplace to characterize the traffic by its volume, this approach is deficient, because the exact number of vehicles passing at a given cross section of the road varies by the hour of the day and seasonally as well. Neither the method applied to determine traffic volumes nor the methodology's efficiency could be justified convincingly.

It is easy to understand that the simplest approach (i.e. counting all passing vehicles every day) is not affordable. It is better to determine an average value, approximating the actual value of the traffic volume. The determination of the following traffic characteristics, used in everyday practice is always based on the traffic volume, i.e. indirectly relates to the results of traffic counts:

- Average annual daily traffic at a given cross section of the road (AADT)
- Average traffic performance related to a given section of the road (Q_{road})
- The type of the traffic (Ψ, B_v)
- Design hour traffic volume (DHV)
- Multiplier of yearly traffic growth (d)

It is questionable, how accurate the result of a traffic count is (or what is its difference from the exact value) and how reliable that data is (i.e. how frequent the differences are). Meanwhile, this question can be reversed, asking for how and how long the traffic should be counted to obtain data with a required exactitude or reliability.

2. Theoretical Background: the Method of Representative Sampling

Generally, the traffic volume of a given road is determined using the method of representative sampling. Essentials of that methodology can be summarised as follows. Aiming at determining the looking for parameters of a whole set (the 'basic set' being under consideration), only thoroughly selected data, related to its smaller subset (called representative sample) are collected and analysed.

Applying the rules of representative sampling the reliability and exactitude of an estimation can be determined, i.e. it can be discovered, how close the value on the base of the sample is estimated to that related to the basic set. In this respect both following rules have crucial importance.

2.1. The Rule of Big Numbers

According to that rule in case of numerous observations the relative frequency of an event certainly falls close to the probability of that event (i.e. to the value, around which the relative frequency fluctuates). The advantage provided by the rule is that it allows to observe frequencies (and calculates relative frequency from them) instead of probabilities. The theory of representative sampling is based on the rules of the probability theory. Probabilistic calculations become applicable in practice, provided there is an appropriately high amount of data related to observed independent events, at our disposal.

According to Tchebisev's inequality formula, reflecting the rule described above, the probability of an outcome, that the difference between a probability variable x and the expected value will exceed an arbitrarily determined Δ value is lower or equal to

$$\frac{D^2(x)}{\Delta^2},$$

where $D^2(x)$ is the variation,
Δ is an arbitrary value.

2.2. The Rule of Central Partition of Distribution

According to that rule, the distribution of probability variables related to random mass events is normal. The normal distribution has a great advantage: knowing the value of the standard deviation, probabilities related to the distribution of variables' positions (calculated from the sample) around the expected value (i.e. the parameter calculated from the basic set) can be predicted.

Fig. 1 demonstrates that 68.3 per cent of arithmetical averages calculated from different samples does not differ from the expected value more than the value of the standard deviation, at 95.5% the difference is less than the double of the standard deviation, while at 99.7% it is less than the treble of the standard deviation.

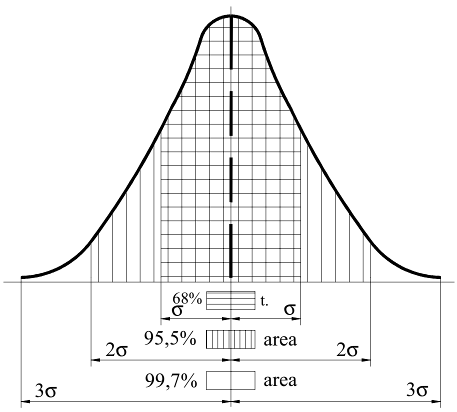

Fig. 1. The normal distribution

It is obvious that the representative sampling can be reliable only in case if the sample is selected to reflect properly the characteristics of the basic set. This can be achieved when all elements of the basic set have equal chance to be incorporated into the sample, i.e. basically random sampling is implemented.

3. Determining the n Number of the Elements of a Sample in case $\Delta = t\sigma_x^-$ is the Required Exactitude of the Estimation

To determine the n number of the elements of a sample, the following three equations should be used:

$$\sigma_x^- = \frac{S}{\sqrt{n}},$$

$$S = \sqrt{\frac{n}{n-1}} \cdot \sigma_m,$$

$$\Delta = t \cdot \sigma_x^-,$$

where σ is the standard deviation of the sample,
S is the experimentally corrected variance.

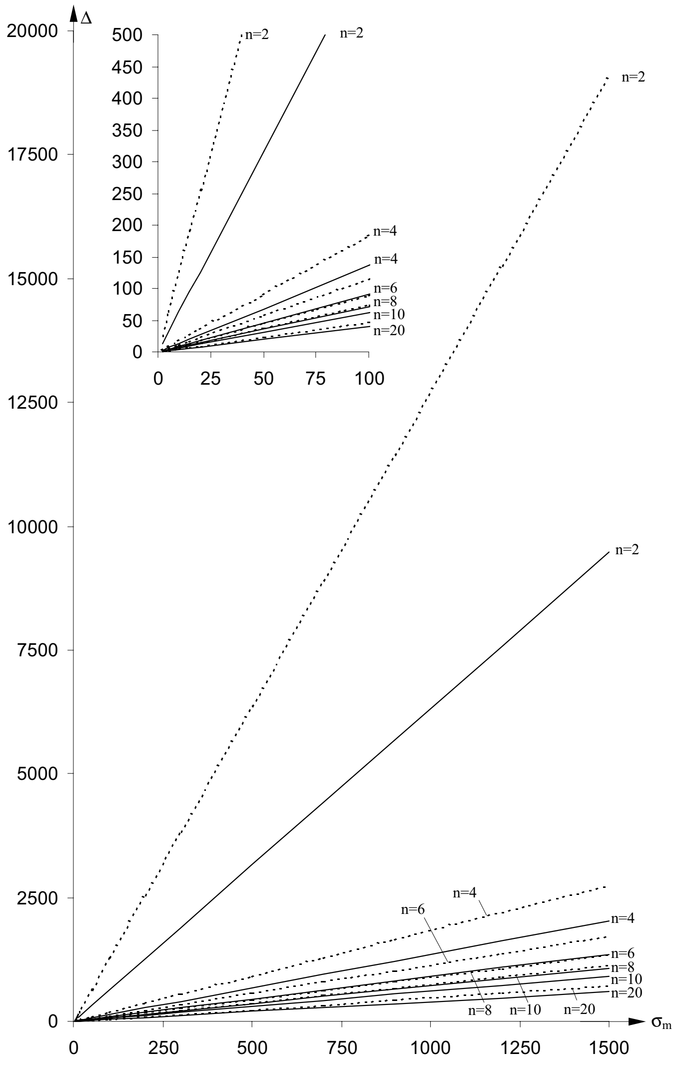

Fig. 2. Relationships between the exactitude of the estimation (Δ) and standard deviation of the sample (σ_m), in function of the number of the sample's elements (n) at 90% (uninterrupted line) and at 95% (dotted line) probability level

Using these equations the Δ (exactitude of the estimation) can be derived. Thus

$$\Delta = \sqrt{\frac{1}{n-1}} \cdot t \cdot \sigma_m .$$

Let us look for the relation between the exactitude of the estimation (Δ), the standard deviation of the sample (σ_m) and the number of the elements in the sample at probability levels of 90% and 95%.

Making calculations using the following values:

- $2 < \sigma_m < 1500$ and
- $n = 2, 4, 6, 8$ and 10,

the result can be seen in *Fig. 2*. The difference between the exactitudes related to 95% and 90% reliability levels ($\Delta_{95\%} - \Delta_{90\%}$) has also been calculated and expressed as a percentage in function of the identical number of the sample's elements in the following way:

The results of these calculations are presented in *Table 1*, and in *Fig. 2*.

Table 1. The difference (in percentage) between values of exactitude at 95% and at 90% level of reliability ($\Delta_{95\%}$ and $\Delta_{90\%}$) in function of the (identical) number of the sample's elements (n)

n	Difference (%)
2	50.4
3	32.2
4	26.1
6	21.7
8	20.0
10	18.7
20	17.7

4. Problems of the Manual Counts' Supplementing Data Provided by Automatic Counters

Before installation, automatic counters have to be controlled and validated by manual counts at the beginning of all data collection cycles. This is naturally another sampling methodology.

The size of the sample (i.e. the number of manual counts) should be determined in the light of the required exactitude and reliability of the expected results. Essentially the following questions have to be answered:

The counted number of axles corresponds to 'how many vehicles?'

What is the composition of the traffic expressed by numbers and by percentage of vehicles in different categories?

How can the volume of traffic be transformed and expressed in passenger car units (PCU)?

The problem can be solved in two ways. In the first approach the size of the sample is determined in compliance with the required exactitude and reliability level. The following equation applies (see the explanation of symbols under point 1):

Under the second approach at a given sample size the exactitude and reliability level of the estimation is calculated from the results, using the following equation:

$$n = \sigma_m^2 \cdot \frac{t^2}{\Delta^2} + 1.$$

This method can be used for calculating the conditions of asphalt pavements in function of changing temperature (see Géza BARTHA, *Periodica Polytechnica* 2001).

$$\Delta = \sqrt{\frac{1}{n-1}} \cdot t \cdot \sigma_m.$$

Data collected by manual counts at a given traffic counting station can be used for transformation of data collected by automatic counters in each vehicle category, either these latter counts were accompanied by simultaneous manual counts or not.

5. Remarks Concerning the Possibility to Increase the *n* Number of the Elements in a Sample

According to the point 2 the exactitude and reliability of each parameter is highly dependent upon the number of elements in a sample (and its standard deviation). For calculating the exactitude and reliability, the number of elements of a sample may be increased in a way, that data collected at other, nearby traffic counting stations having similar traffic characteristics, would be incorporated into the data collected at the station under consideration. This approach is not suitable for determining q value, while this is the traffic volume at the secondary station, but can be applied for calculating the hourly variation parameter, reflecting the daily distribution of the traffic flow.

6. Determining the Exactitude and Reliability when Calculating Results of a Traffic Count

The main results of traffic counts are the following:

- The volume of average annual daily traffic at a given cross section of the road (AADT)
- The volume of average annual daily traffic on a road section or road network
- The type of the traffic
- The design hour traffic volume (DHV)
- The traffic growth multiplier

6.1. Determining AADT

Annual average daily traffic at a given cross section of the road is calculated by the following basic equation:

$$Q = q \cdot a \cdot b \cdot c \qquad \text{(vehicle/day)},$$

where Q is the average annual daily traffic (vehicle/day),
q is the result of an uninterrupted traffic count executed in a less than 24 hours long time period (vehicle/t), t being the counting time period in hours,
a is the daily time period variation parameter, used as a multiplier to get the entire daily traffic volume from the traffic counts' data,
b is the weekly variation parameter, used as a multiplier to get the weekly average of the daily traffic volume from the traffic counts' data collected at a given day of the week,
c is the monthly variation parameter, used as a multiplier to get the monthly average of the daily traffic from the counts' data collected in a given month of the year.

Determination of Q is made by vehicle categories:

$$Q_i = q_i \cdot a_i \cdot b_i \cdot c_i,$$

where i means the ith category of vehicles. For determining AADT, traffic counts are generally executed throughout several (n) days. The value of q has to be determined for each day. Thus values of $Q_{i1}, Q_{i2}, \ldots Q_{im}$ are generated. The arithmetical average of these values gives the value of $Q_{i\ av}$:

$$Q_{i\ av} = \frac{Q_{i1} + Q_{i2} + Q_{i3} + \ldots + Q_{im}}{m} \qquad \text{(vehicle/day)}.$$

The $Q_{i\ av}$ values related to each category of vehicles should be aggregated to get the AADT value as required:

$$Q = Q_{i\ av} + Q_{j\ av} + Q_{k\ av} + \ldots + Q_{m\ av}.$$

6.2. Exactitude of the Estimation of AADT at a Given Cross Section of the Road and the Number Reflecting Reliability

As it was discussed under 6.2, for determining Q the q result of a traffic count has to be multiplied by a, b and c parameters (each of which contains some errors), then the Q values related to each vehicle category should be aggregated. Obviously, the Q received as a final value contains an error linked to a reliability level. What kind of equations can be used to calculate the exactitude of Q, and what recommendation can be made to assess the reliability level linked to it? According to our best information, the theorem of mean error and weight of a function's value can be used, which states the following. Let a function be given as follows:

$$\Omega = f(u_1, u_2, \dots, u_n).$$

Carry out measurements to get values of $u_1, u_2, \dots, u_n$. The results of these measurements are $l_1, l_2, \dots, l_n$. The weights allocated to these measurements' results are $p_1, p_2, \dots p_n$, while their mean errors are $\mu_1, \mu_2, \dots \mu_n$.

Introducing the measurements; results into the functional equation, not the exact Ω value, but a different ω is received:

$$\omega = f(l_1, l_2, \dots, l_n).$$

The mean error and the weight of each measurement's result is known. Using these how can be determined the μ_ω mean error and the ρ_ω weight of the function value?

The theorem of the mean error and the weight of the function value states:

$$\mu_\omega = \sqrt{f_1^2 \cdot \mu_1^2 + f_2^2 \cdot \mu_2^2 + \dots + f_n^2 \cdot \mu_n^2},$$
$$\frac{1}{p_\omega} = f_1^2 \cdot \frac{1}{p_1} + f_2^2 \cdot \frac{1}{p_2} + \dots + f_n^2 \cdot \frac{1}{p_n},$$

where the interpretation of $f_1, f_2 \dots f_n$ parameters is the following:

$$f_i = \left(\frac{\partial f}{\partial U}\right)_{\substack{U_1 = l_1 \\ U_2 = l_2 \\ \vdots \\ U_n = l_n}}.$$

These equations are valid under the following conditions:

- The f function is continuous and can be derived.
- Measurements' results contain only irregular errors (i.e. average value equals 0).
- Errors of the measurements' results are as small that their squares and products can be neglected (i. e. measurements were made thoroughly).
- Errors of the measurements' results are independent from each other.

The theorem of the mean error and weight of a function's value referred to above can be used in the practice of traffic counts as well. When determining AADT (Q), first a product's mean error and weight have to be calculated. After that comes the calculation of the aggregated value's mean error and weight, twice: for determining $Q_{i\ av}$ and Q. The results of that calculation can be shown in *Table 2*.

The value received in that way, however, is a relative one, related to the weight only, therefore it is unsuitable for comparisons. The result associated to a bigger weight is considered as a better one. According to the methodology presented, in determining AADT the correctness of the parameters impacts first of all its value's exactitude and only in a less extent the reliability of the final result.

6.3. *Exactitude of the Calculation of AADT on a Road Section or Road Network and the Number Reflecting the Exactitude and Reliability of the Estimation*

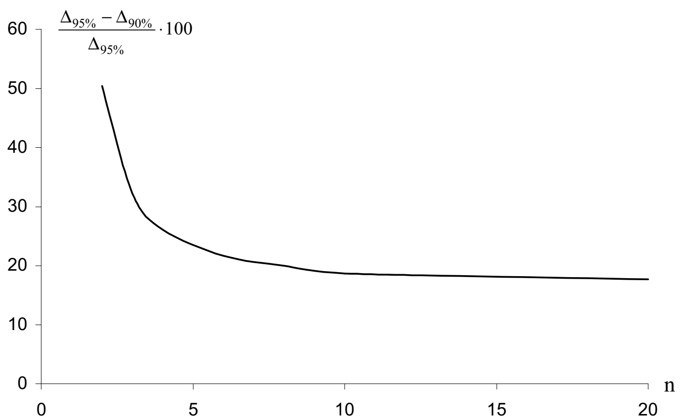

Fig. 3. The difference (in percentage) between values of exactitude at 95% and at 90% level of reliability ($\Delta_{95\%}$ és $\Delta_{90\%}$) in function of the (identical) number of the sample's elements (n)

Using the AADT values determined by traffic counts and the appropriate section lengths related to them, the average AADT on a given road section can be determined. This average should be calculated by multiplying the appropriate AADT values by the section lengths, then aggregating these products and weighting them with the section lengths as follows:

$$Q_{\text{section}} = \frac{L_1 \cdot Q_1 + L_2 \cdot Q_2 + L_3 \cdot Q_3 + \ldots + L_m \cdot Q_n}{L_1 + L_2 + L_3 + \ldots + L_m},$$

where Q_{section} is the average AADT on the road section under consideration
$L_1, L_2 \ldots L_m$ are section lengths
$Q_1, Q_2 \ldots Q_m$ the AADT on each section

Average AADT on a road network can be determined in a similar way.

In the equation discussed under point 5.2. $Q_1, Q_2, \ldots Q_n$ average annual daily traffic volumes contain $\Delta = \mu$ error ($\Delta_1 = \mu_1, \Delta_2 = \mu_2, \ldots, \Delta_m = \mu_m$) and a well defined reliability level is associated to them. The calculation of the mean error and weight of the sum can be made by applying the methodology presented under point 5.2. Results of the calculation are presented in *Table2*.

Table 2. Calculation of traffic parameter's estimation error and the weight proportional to their reliability

Basic formula	Calculation of estimation error (μ) and weight proportional to reliability ($1/p$)
$Q_i = q_i \cdot a_i \cdot b_i \cdot c_i$	$\mu_{\omega Q_i} = \sqrt{a_i^2 \cdot b_i^2 \cdot c_1^2 \cdot \mu_{qi}^2 + q_i^2 \cdot b_i^2 \cdot c_1^2 \cdot \mu_{ai}^2 + q_i^2 \cdot a_i^2 \cdot c_1^2 \cdot \mu_{bi}^2 + q_i^2 \cdot a_i^2 \cdot b_1^2 \cdot \mu_{ci}^2}$ $\frac{1}{p_\omega} = \frac{a_i^2 \cdot b_i^2 \cdot c_1^2}{p_1} + \frac{q_i^2 \cdot b_i^2 \cdot c_1^2}{p_2} + \frac{q_i^2 \cdot a_i^2 \cdot c_1^2}{p_3} + \frac{q_i^2 \cdot a_i^2 \cdot b_1^2}{p_4}$
$Q_{i\ av} = \frac{Q_{i1}}{m} + \frac{Q_{i2}}{m} + \frac{Q_{i3}}{m} + \ldots + \frac{Q_{in}}{m}$	$\mu_{\omega Q av} = \sqrt{\frac{1}{m^2} \cdot \mu_{Q_{i1}}^2 + \frac{1}{m^2} \cdot \mu_{Q_{i2}}^2 + \frac{1}{m^2} \cdot \mu_{Q_{i3}}^2 + \ldots + \frac{1}{m^2} \cdot \mu_{Q_{im}}^2}$ $\frac{1}{p_\omega} = \frac{1}{m^2} \cdot \frac{1}{n_{Q_{i1}}} + \frac{1}{m^2} \cdot \frac{1}{n_{Q_{i2}}} + \frac{1}{m^2} \cdot \frac{1}{n_{Q_{i3}}} + \ldots + \frac{1}{m^2} \cdot \frac{1}{n_{Q_{in}}}$
$Q = Q_{i\ av} + Q_{j\ av} + Q_{k\ av} + \ldots + Q_{m\ av}$	$\mu_{\omega Q} = \sqrt{\mu_{Q_{i\ av}}^2 + \mu_{Q_{j\ av}}^2 + \mu_{Q_{k\ av}}^2 + \ldots + \mu_{Q_{m\ av}}^2}$ $\frac{1}{p_\omega} = f_1^2 \cdot \frac{1}{p_1} + f_2^2 \cdot \frac{1}{p_2} + f_3^2 \cdot \frac{1}{p_3} + \ldots + f_m^2 \cdot \frac{1}{p_m}$
$Q_{\text{road}} = \frac{L_1 \cdot Q_1 + L_2 \cdot Q_2 + L_3 \cdot Q_3 + \ldots + L_m \cdot Q_n}{L_1 + L_2 + L_3 + \ldots + L_m}$	$\mu_{\omega Q_{\text{road}}} = \sqrt{\left(\frac{L_1}{L}\right)^2 \cdot \mu_{Q_1}^2 + \left(\frac{L_2}{L}\right)^2 \cdot \mu_{Q_2}^2 + \left(\frac{L_3}{L}\right)^2 \cdot \mu_{Q_3}^2 + \ldots + \left(\frac{L_m}{L}\right)^2 \cdot \mu_{Q_m}^2}$ $\frac{1}{p_\omega} = \left(\frac{L_1}{L}\right)^2 \cdot \frac{1}{n_{Q1}} + \left(\frac{L_2}{L}\right)^2 \cdot \frac{1}{n_{Q2}} + \left(\frac{L_3}{L}\right)^2 \cdot \frac{1}{n_{Q3}} + \ldots + \left(\frac{L_m}{L}\right)^2 \cdot \frac{1}{n_{Qm}}$
$\Psi = \frac{F_{\text{July}}^{\text{Weekday}} + F_{\text{August}}^{\text{Weekday}}}{F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}}}$ $B_v = \frac{F_{\text{July}}^{\text{Sunday}} + F_{\text{August}}^{\text{Sunday}}}{F_{\text{July}}^{\text{Weekday}} + F_{\text{August}}^{\text{Weekday}}}$	$\mu_{\omega\Psi} = \sqrt{\left(\frac{1}{F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}}}\right)^2 \cdot \left(\mu_{\text{July}}^{\text{Weekday}}\right)^2 + \left(\frac{1}{F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}}}\right)^2 \cdot \left(\mu_{\text{August}}^{\text{Weekday}}\right)^2 + \left(-\frac{F_{\text{July}}^{\text{Weekday}} + F_{\text{August}}^{\text{Weekday}}}{(F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}})^2}\right)^2 \cdot \left(\mu_{\text{April}}^{\text{Weekday}}\right)^2 + \left(-\frac{F_{\text{July}}^{\text{Weekday}} + F_{\text{August}}^{\text{Weekday}}}{(F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}})^2}\right)^2 \cdot \left(\mu_{\text{May}}^{\text{Weekday}}\right)^2}$ $\frac{1}{p_{\omega\Psi}} = \left(\frac{1}{F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}}}\right)^2 \cdot \frac{1}{n_1} + \left(\frac{1}{F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}}}\right)^2 \cdot \frac{1}{n_2} + \left(-\frac{F_{\text{July}}^{\text{Weekday}} + F_{\text{August}}^{\text{Weekday}}}{(F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}})^2}\right)^2 \cdot \frac{1}{n_3} + \left(-\frac{F_{\text{July}}^{\text{Weekday}} + F_{\text{August}}^{\text{Weekday}}}{(F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}})^2}\right)^2 \cdot \frac{1}{n_4}$
$\text{DHV} = \omega \cdot \text{AADT}$	$\mu_{\omega\ \text{DHV}} = \sqrt{\text{AADT}^2 \cdot \mu_\omega^2 + \omega^2 \cdot \mu_{\text{AADT}}^2}$ $\frac{1}{p_{\omega\ \text{DHV}}} = f_1^2 \cdot \frac{1}{p_1} + f_2^2 \cdot \frac{1}{p_2} = \text{AADT}^2 \cdot \frac{1}{n_\omega} + \omega^2 \cdot \frac{1}{n_{\text{AADT}}}$
$d = \frac{\text{AADT}_{\text{future}}}{\text{AADT}_{\text{basic}}}$	$\mu_{\omega d} = \sqrt{\left(\frac{1}{\text{AADT}_{\text{basic}}}\right)^2 \cdot \mu_{\text{AADT}_{\text{future}}} + \left(-\frac{\text{AADT}_{\text{future}}}{\text{AADT}_{\text{basic}}^2}\right)^2 \cdot \mu_{\text{AADT}_{\text{future}}}}$ $\frac{1}{p_{\omega d}} = \left(\frac{1}{\text{AADT}_{\text{basic}}}\right) \cdot \frac{1}{n_{\text{AADT}_{\text{basic}}}} + \left(-\frac{\text{AADT}_{\text{future}}}{\text{AADT}_{\text{basic}}^2}\right)^2 \cdot \frac{1}{n_{\text{AADT}_{\text{future}}}}$

6.4. Determining the Type of the Traffic and the Number Reflecting the Exactitude and Reliability of the Estimation

The equations used to determine the type of the traffic are the following:

$$\Psi = \frac{F_{\text{July}}^{\text{Weekday}} + F_{\text{August}}^{\text{Weekday}}}{F_{\text{April}}^{\text{Weekday}} + F_{\text{May}}^{\text{Weekday}}}, \qquad B_v = \frac{F_{\text{July}}^{\text{Sunday}} + F_{\text{August}}^{\text{Sunday}}}{F_{\text{July}}^{\text{Weekday}} + F_{\text{August}}^{\text{Weekday}}},$$

where F is the volume of the car traffic (car/day) at the given day in the given month. The methodology to be used to calculate the mean error of the function's value and the reliability of the estimation is described in *Table*2.

6.5. Determining Design Hour Traffic and the Number Reflecting the Exactitude and Reliability of the Estimation

The design hour traffic volume (DHV) is calculated as the product of AADT and ω peak hour parameter:

$$\text{DHV} = \omega \cdot \text{AADT}.$$

The results of the calculation obtained in compliance with the methodology presented above are presented in *Table* 2.

6.6. The Calculation of the Traffic Growth Multiplier and the Number Reflecting the Exactitude and Reliability

The d traffic growth multiplier is interpreted as the rate of the future and basic years' average annual daily traffic (AADT):

$$d = \frac{\text{AADT}_{\text{future}}}{\text{AADT}_{\text{basic}}}.$$

The calculation is made in compliance with the methodology presented above. The results are shown in *Table* 2.

References

[1] BÉNYEI, A. – TÓTH, ZS. – BACHMANN, D. – GULYÁS, A. – THURZÓ, G., A Sampling módszer szerint történő forgalomszámlálás és adatfeldoldolgozás eljárása, Összefoglaló tanulmány, Budapest, 1999 október, pp. 1–42.

[2] FI, I., Forgalmi Tervezés Technika Menedzsment, Egyetemi tankönyv, Egyetemi kiadó, Budapest, 1997.

[3] BACSÓ, A. – BERÉNYI, J. – CHOLNOKYNÉ, FERENCZI É. – CSORJA, ZS. – MOGYORÓSI, I. – REINISCH, E. – TAKÁCS, F. – TÓTH, J., Közutak tervezési szabályzata, tervezet, Magyar Útügyi Társaság, Budapest, 2000, pp. 1–134.

[4] BÉNYEI, A. – GOLARITS, P., Measurements to Define Relationship between Traffic Volume and Traffic Conditions in Hungary *Periodica Polytechnica Ser. Civil Eng.*, **46** pp. 83–94, 2002.

[5] BARTHA, G., Temperature Change in Asphalt Pavements in Summer, *Periodica Polytechnica Ser. Civil Eng.*, **46** pp. 43–52, 2002.