# Sensitivity of Confidence Interval to Changes in Sample Variance Parameters within Transit Time Assessment of Postal Transportation Flows


**L. Švadlenka*, D. Salava**, K. Pojkarová*****

**University of Pardubice, The Faculty of Transport Engineering, Studentská 95, 53210, Pardubice, Czech Republic, E-mail:* libor.svadlenka@upce.cz
***University of Pardubice, The Faculty of Transport Engineering, Studentská 95, 53210, Pardubice, Czech Republic, E-mail:* daniel.salava@upce.cz
****University of Pardubice, The Faculty of Transport Engineering, Studentská 95, 53210, Pardubice, Czech Republic, E-mail:* katerina.pojkarova@upce.cz



**Abstract**

The paper is focused on confidence interval sensitivity to changes of dispersion parameters within measurement of the transit time of postal transportation services, namely end-to-end services for single piece priority mail. The role of confidence interval is the expression of estimate accuracy for on-time probability. The accuracy is expressed as width of the confidence interval of on-time probability estimate. On-time probability is interpreted by variable called weighted estimate of on-time probability, which also includes distribution of mail flows within geographical stratification according to disjunctive set of defined postal areas. The rate of variability caused by used sample design is expressed by dispersion parameters of sample variance in various forms and by design factors regarding particular samples. The main objective of this paper is to research sensitivity of confidention interval to changes in sample variance parameters considering limit value of interval width related to acceptable accuracy level. Measurement system created by sample of test postal items flows is based on real traffic mail flows on fixed postal transportation network. The results will come out from modelling of this measurement system and evaluation of postal transport network efficiency by such system is an important managerial tool for quality of service. Analysis of mentioned sensitivity can help to find the most appropriate sample design.


**KEY WORDS:** *transit time, priority mail, traffic flows, sample variance, confidention interval, dispersion*


## 1. Introduction

The aim of transit time measurement system is to estimate quality of service by transit time of end-to-end services for single-piece priority mail (SPPM – in essence ordinary letters) provided to customers by domestic mail in each European country and by cross-border mail among European countries. European Comission has defined requirements on postal measuring systems of quality of service with possibility of independent end-to-end measurement and these requirements are defined in European norm EN 13850:2012.

Provision and operation of mentioned services is measured by on-time performance evaluation of test postal items flows with key measured parameter of transit time. Service standard for SPPM in Czech Republic is given by Regulation nr. 464/2012 Coll. Relevant part sets, that by measuring of transit times per calendar year there must be achieved result at least 92% of postal mail delivered on the first working day following day of its posting. Methodology of measuring must be based on test items volumes representative in relation to existing real mail flows for SPPM represented by real mail study.

Measurement is realized by independent performance monitoring organisation and field of this study in Czech Republic includes related services of national (public) postal operator within domestic mail provision. The course of measuring contained in report with concrete results must be audited. Auditor has the main task to assess measuring process from viewpoint of conformity with mentioned European norm and correctness and interpretation of measured and calculated results. Position of auditor is established by Czech national regulative authority for postal sector on the basis of public tender result.

Measurement can be realized within period approved by regulative authority (at least with annual results), e.g. each quarter of a year and after its closing assessment procedure must be made. In this case audit report for each quarterly measurement and summary audit report per year must be made. Mission following from position of auditor is to verify conformity of measurement process with mentioned European norm including proposal for corrections to eliminate found imperfections and steps to improvement of measuring process efficiency including data collection and processing.

## 2. Key Characteristics of Measuring Process and Analysis

Method of test items within measuring must be created by process where panellists work as senders and/or receivers. Senders induct test items to postal operator network and record date and time of posting and receivers record date of delivery. Senders and receivers must be distributed within the whole field of study and this panel size must respect and correspond especially to minimal number of covered postal areas, minimal sample size of test items, maximal workload of panellists and stratification and geographical coverage.

Test items must be made in the way meeting specifications of discriminant mail characteristics (DMC). Each other element on the list of potential candidates for DMC is reviewed by real mail study with result if it is discriminant or not. Elements with verified discriminant feature must be included into stratification. The most frequent possible discriminant characteristics (with impact on transit time result) with own specific modes are type of geographical area (urban / rural), type of payment (postal stamps / evidence / payment), type of induction (postbox / post office / sorting centre), time of posting, formats, weight degrees and method of addressing (written / printed) or weekday of induction (including Saturday and Sunday).

In other words, each test item is characteristic for its specific combination of DMC modes. Particular proportions of these modes are explored within real mail study as well. Real mail study is a part of our audit activity including audit report and proposal of appropriate corrections in methodology of measurement.

Measuring must be performed with sufficient accuracy defined by requirements of Norm. For appropriate and relevant proposals, relevant analyses are related to:

1. flows of test items divergent from planned proportional flows defined by real mail study and corresponding sample design,
2. optimal workload of panellists and their working attitude,
3. consistency of submited physical and electronic materials,
4. possible causes of imperfections,
5. statistical analyses related to required accuracy with key parameters of on-time probability estimation, sample dispersions, design factors, confidention intervals etc.
6. correctness of calculated results and their intepretation,
7. corrections of input data with impact to calculated parameters of performance and efficiency of measuring process,
8. other problems connected with possible distortion and transparency of measuring process.

In general – mentioned analyses reflect common problems which may occur by data collection, creation of test items database containing all their characteristics and successive calculating and assessment procedures.

Participation of auditor on measuring assessment in the course of measurement cycles should lead to positive changes due to apllication of audit recommendations. These changes should improve transparency, correctness and efficiency of measuring process.

Main result of measuring is on-time performance parameter. The result of on-time performance is represented by estimate of on-time probability, which can be defined as probability of case, when transit time of postal item does not exceed just one day. On-time probability is interpreted by variable called weighted estimate of on-time probability, which also includes distribution of mail flows within geographical stratification according to disjunctive set of defined postal areas. On-time performance estimate must be sufficiently accurate. The role of confidence interval is the expression of estimate accuracy for on-time probability. The accuracy is expressed as width of the confidence interval of on-time probability estimate. The rate of variability caused by used sample design is expressed by dispersion parameters of sample variance in various forms and by design factors regarding particular samples. Confidence interval, among others, depends on dispersion parameters and next part of paper reflects sensitivity of confidention interval to changes in sample variance parameters.

## 3. Characteristic of Dispersion Parameters

Design of measuring system should ensure representative sample of SPPM test items in the field of study. The most common way to reach representative sample would be design of simple random sample (*SRS*) comprising real mail and then monitoring of its transit time. Realization of such design with high accuracy of measuring would be unrealistic. Thus design works with prepared test items sent and received by selected panellists to ensure high quality measuring. This approach to design requires test items to integrate into existing real mail flows with no structural differences.

The rate of variation caused by used sample design is expressed by design factor from possible viewpoint of unstratified end-to-end measuring system (*EtE*), stratified random sample (*StrRS*) and stratified end-to-end sample (*StrEtE*). It is related to sample design and on-time probability estimate. Significant parameters of dispersion describing exploitation of measuring system are represented by sample variance from viewpoint of *EtE*, *StrRS* and *StrEtE* as well [1].

Dispersion parameter determinative for confidence interval calculation is sample variance of stratified end-to-end sample (*StrEtE*) formed as:

$$\hat{\text{var}}_{StrEtE}\left[\hat{p}_{weighted}\right] := \frac{\hat{\text{var}}_{EtE}\left[\hat{p}\right] * \hat{\text{var}}_{StrRS}\left[\hat{p}_{weighted}\right]}{\hat{\text{var}}_{SRS}\left[\hat{p}\right]} \qquad (1)$$

here $\hat{p}$ is value of on-time probability estimator, $\hat{p}_{weighted}$ is weighted estimation for on-time probability with respect to stratification, $\hat{\text{var}}_{SRS}\left[\hat{p}\right]$ is sample variance of simple random sample (*SRS)*, $\hat{\text{var}}_{EtE}\left[\hat{p}\right]$ is sample variance of on-time probability estimate in unstratified end-to-end measurement system (*EtE*), $\hat{\text{var}}_{StrRS}\left[\hat{p}_{weighted}\right]$ is sample variance of stratified random sample (*StrRS*).

Measuring system works with parameter of design factor to reflect covariance and weighing impacts more transparently. The rate for added variance is expressed by design factor *df*, which measures loss of accuracy. Design factor is always related to set sample design and probability estimator [1]. In basis it can be defined as ratio of sample variance of on-time probability estimator $\hat{p}$ in set sample design to sample variance of on-time probability estimator $\hat{p}$ in *SRS* of the same size.

Design factor determinative for confidence interval calculation is design factor for end-to-end stratified measuring system (*StrEtE*):

$$df_{StrEtE} := df_{EtE} * df_{StrRS} \tag{2}$$

here $df_{EtE}$ is design factor for end-to-end unstratified measuring system and $df_{StrRS}$ is design factor for stratified random sample, *n* is volume of test items, both formulations follow [1]:

$$df_{EtE} := \frac{\hat{\text{var}}_{EtE}\left[\hat{p}\right]}{\hat{\text{var}}_{SRS}\left[\hat{p}\right]} \tag{3}$$

$$df_{StrRS} := \frac{\hat{\text{var}}_{StrRS}\left[\hat{p}_{weighted}\right] * (n-1)}{\hat{p}_{weighted} * \left(1-\hat{p}_{weighted}\right)} \tag{4}$$

## 4. Confidence Interval Characteristic

Key indicator for accuracy of measuring system is represented by width of the confidence interval of on-time probability estimate. In other words the role of confidence interval is the expression of estimate accuracy for on-time probability. Smaller width of interval leads to higher accuracy of measurement. Maximum width $2\varepsilon$ of this confidence interval is defined on the basis of confidence level $(1\text{-}\alpha) = 95\ \%$ [2].

Whichever confidence interval is based on probability distribution. Appropriate probability distribution for modelling of on-time estimate in simple random sample would be binomial distribution. Confidence intervals for this distribution are not calculated directly, but appropriate approximation of binomial distribution is recommended for assessment of measurement system accuracy. For majority of measurement system models, the normal distribution is appropriate approximation of binomial distribution. Confidence interval and accuracy for $\hat{p}_{weighted}$ are defined as follows [1]:

$$\left[\hat{p}_{weighted} \pm 1{,}95996 * \sqrt{\frac{\hat{p}_{weighted} * \left(1-\hat{p}_{weighted}\right)}{n-1} * df_{StrEtE}}\right] \tag{5}$$

$$2\varepsilon = 3{,}91992 * \sqrt{\hat{\text{var}}_{StrEtE}\left[\hat{p}_{weighted}\right]} \tag{6}$$

This normal confidence interval is symmetrical and it is frequently apllied for assessment of measurement system accuracy. Normal approximation will be applicable in case, when on-time performance does not reach *100 %.* Increase of on-time performance level will cause increase of approximation deviation. Confidence intervals can be calculated by asymptotic normality of maximum probability estimate for sample sizes of large scale and when probability estimate does not reach *0* or *1*.

## 5. Test Samples and Results of Dispersion and Accuracy Parameters

Test samples are based on parameters of geographical coverage by postal transportation services and stratification of measuring sample for four periods of 2015-2016. Modelling presumes one-Operator field of study with domestic SPPM for various sample sizes in proportion of design basis. Used sample sizes are necessary to cover all postal areas with concrete flows of test items, which must fulfil requirement of proportionality with design basis of real SPPM flows. Modelling samples consider half-month period ($1^{st}$ period), one-month period ($2^{nd}$ period), one-and-half-month period ($3^{rd}$ period) and two-months period ($4^{th}$ period).

Measuring system modelling for all periods is based on relevant indicators figures. These parameters comprised in Table 1 create database necessary for assessment of measuring process. Applying these variables, calculated results of dispersion parameters and accuracy using above-mentioned equations follow.

Table 1

Dispersion and accuracy parameters

| Parameter | 1st period | 2nd period | 3rd period | 4th period |
|---|---|---|---|---|
| $n$ | 1086 | 2096 | 3157 | 3906 |
| $\hat{p}_{weighted}$ | 0,9280 | 0,9395 | 0,9444 | 0,9414 |
| $\hat{\text{var}}_{EtE}\left[\hat{p}\right]$ | 1,165E-04 | 5,829E-05 | 4,221E-05 | 4,303E-05 |
| $\hat{\text{var}}_{SRS}\left[\hat{p}\right]$ | 6,217E-05 | 2,717E-05 | 1,659E-05 | 1,413E-05 |
| $\hat{\text{var}}_{StrRS}\left[\hat{p}_{weighted}\right]$ | 6,095E-05 | 2,723E-05 | 1,666E-05 | 1,413E-05 |
| $\hat{\text{var}}_{StrEtE}\left[\hat{p}_{weighted}\right]$ | **1,142E-04** | **5,841E-05** | **4,237E-05** | **4,303E-05** |
| $df_{EtE}$ | 1,874 | 2,145 | 2,544 | 3,045 |
| $df_{StrRS}$ | 0,990 | 1,004 | 1,001 | 0,999 |
| $df_{StrEtE}$ | **1,856** | **2,154** | **2,546** | **3,044** |
| **$2\varepsilon$** | **0,0419** | **0,0300** | **0,0255** | **0,0257** |

Acceptable margin for measured result of on-time probability estimate within accuracy demands is set on value $\varepsilon$ equal *1%* for domestic mail and so *2%* for $2\varepsilon$ [1]. This minimum accuracy limit is connected with volume of test items and corresponds with set level of on-time performance represented by service standard. More robust sample could bring more accurate result, but there are more influencing factors follows from formal definition of accuracy by confidence interval as design factors, on-time probability estimator or sample variances. These are complex interconnected indicators. Exceeding of demanded maximum width of confidence interval $2\varepsilon$ means lowered accuracy of measuring because real on-time performance indicator is covered by wider interval, with probability corresponding to given confidence level. Calculated value of this parameter estimator on the basis of measuring means for real value of on-time performance the most probable position in interval. Narrower confidence interval creates smaller scope for fluctuation of real on-time performance indicator and that is why its result is precised. Non-fulfilment of set confidence interval width would be arguable in case of its multiple exceeding which significantly devalues result of observed on-time performance indicator due to low level of accuracy given wide possible scope of real on-time performance indicator fluctuation.

Calculated values of $2\varepsilon$ exceed set level of *2%*, the are getting closer to this level with lower values of sample variance parameters. Changes of design factors have inverse trend, their position is connected with on-time probability estimators, sample size and also follow from dispersions influenced by stratification and distribution of end-to-end points. More robust samples have not to lead to lower design factors due to smaller group of panellists and lower figure of end-to-end points. On the opposite, correlation influences due to this points can be reflected by relation of stratified dispersion parameters and end-to-end dispersion parameters. Volume of test items itself does not ensure optimal results. Experimental modelling with sample sizes up to *15* thousand has led to confidence interval width below *2%*, but there were arranged some modifications in distribution of induction and delivery points and stratification. Design factors of stratified end-to-end measuring system $df_{StrEtE}$ have ranged to the contrary from cca *3,5* to *6,5*, but there were sufficiently lower values of stratified end-to-end dispersions. Weighted estimate for on-time probability has moved on comparable level. Considering construction forms of confidence interval and dispersion parameters following from on-time performance parameters there is not easy to reach reliable and accurate results without complex approach for measuring system design. Sensitivity of accuracy to dispersion has more complex character.

## 5. Conclusions

Important categories for sufficient accuracy demands are sample dispersion parameters and design factors with characteristics of apllied stratification and set of end-to-end points. These indicators create basis for determination of confidence interval reflecting accuracy of on-time performance result by width of this interval. Narrower confidence interval creates smaller scope for fluctuation of real on-time performance indicator and that is why its result is precised. Due to complexity of measuring parameters it is not easy to point out unequivocally any indicator with determinative influence on result accuracy. It is important to consider factors of organisation of measuring system with possible compromise situations between expensiveness and complicatedness of system and its reliability and accuracy.

## References


1. ČSN EN 13850: Postal Services – Quality of Services – Measurement of the transit time of end-to-end services for single piece priority mail and first class mail. 2013. Prague: Czech Office for Standards, Metrology And Testing, 104 p.
2. **Bickel, P.J.; Lehmann, E.L.** 1976. Descriptive Statistics for Nonparametric Models (III. Dispersion), Annals of Statistics**,** Mathematical Statistics 4(6): 1139–1158.
3. Reg. no. 464/2012 Coll. on Establishing of Specifications for Basic Services and on Basic Quality Requirements for Their Provision [online cit.: 2017-04-12]. Available from: http://www.mpo.cz/dokument142643.html