CONFIDENCE INTERVAL DETERMINATION OF ON-TIME PERFORMANCE WITHIN MEASUREMENT OF THE TRANSIT TIME FOR SINGLE PIECE PRIORITY MAIL

Libor Švadlenka – Kateřina Pojkarová, Daniel Salava

Abstract

The paper is focused on usage of methods for confidence interval calculation by measurement of the transit time of end-to-end services for single piece priority mail. The results come out from modelling of measurement system on the basis of relevant field of study with particular discriminant characteristics. 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 probability distribution for modelling of on-time performance in simple random sample is based on relevant approximation of binomial distribution. Considered approximations are normal distribution, Agresti-Coull approximation and approximation by beta inverse function. The main objective of this paper is calculation of confidence interval by mentioned approximation methods, comparison and assessment of the results from viewpoint of next modelling. The results of on-time probability estimate accuracy are related to input assumptions, i.e. especially geographical coverage of postal services and geographical stratification on disjunctive set of postal areas. Calculation methods consider one-Operator field of study with domestic mail for various sample sizes.

Key words: transit time, priority mail, confidence interval, measurement system

JEL Code: C02, C13

Introduction

European Comission emphasizes the necessity to have common rules for development of postal services within Community and improvement of quality of service (QoS). Comission has defined requirements on postal measuring systems of QoS with possibility of independent end-to-end measurement. The aim of this measurement is to estimate QoS by transit time of end-to-end services for single-piece priority mail (SPPM) provided to customer by domestic mail in each European country and by cross-border mail among European countries. SPPM is collected, processed and delivered by postal operators and measuring process uses representative sample of end-to-end services for addressed mail with set level of service transit time. The total QoS level by transit time is expressed as percentage of end-to-end items distributed during $D+m$ days (see Chap. 1). Design of QoS measuring system includes selection and allocation of test items. These items are posted and received by selected panellists. Sample design includes specifications of panellists and test items, which must be representative in consideration of design basis. Design basis is the most appropriate structural information available for characterization of real mail distributed in particular field of study.

The paper is focused on usage of methods for confidence interval calculation by measurement of the transit time of end-to-end services for SPPM. 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 probability distribution for modelling of on-time performance in simple random sample is based on relevant approximation of binomial distribution.

Interval estimation of probability in a binomial distribution can bring some difficulties with validity of gained results. Some related recent articles point out the fact that coverage attributes of the standard interval do not have to be sufficient and need not to give acceptable and reliable results. Interval estimation of binomial proportion has been recently reviewed due to unsatisfactory results of Wald confidence interval of coverage probability. It appears in cases, when related probability is not close to limits, see Agresti and Coull (1998), Newcombe (1998), Brown, Cai, and DasGupta (2001), Reiczigel (2003), Vollset (1993), Blyth (1986), Blyth and Still (1983). Brown, Cai, and DasGupta (2001) has presented and analyzed various alternatives to the standard interval for a general confidence level and also have made recommendations for choice of specific interval for practical application separately for different intervals of sample sizes. Recommendation of Agresti and Coull (1998) made for the nominal 95 % case differs from conclusions of Brown, Cai, and DasGupta (2001) for small sample sizes (40 or less). Agresti and Caffo (2000) point out the unsatisfactory behavior of standard Wald confidence interval. They use and describe numerical methods and design improvement of new confidence interval by coverage probability, which improved confidence interval is known as Agresti-Coull confidence interval. Agresti and Coull (1998) improved confidence interval for the difference of two binomial proportions, which has given new confidence interval of coverage probability with outstanding behavior. Newcomb (1998) used lower and upper limits of score confidence interval and combined these limits for difference of two binomial proportions. This new confidence interval has been compared with others confidence intervals and this comparison points out good performance of coverage probability.

It will follow from this article, that sample sizes necessary for real concrete measuring of on-time performance by estimate of on-time probability will be sufficient for acceptance of used calculation methods. These methods will give reliable results of confidence intervals for measuring accuracy assessment.

1 Initial assumptions of measuring system

Transit time of postal item is measured in days and it is expressed as *D+m* days. Day of posting *D* is the date of next clearance after induction of item into postal network. The result of QoS is defined by on-time performance. This basic postal performance indicator means percentage of postal items delivered in defined service standard.

The result of on-time performance must be expressed as percentage of postal items delivered in transit time *D+m*, where *m* represents figure of days determining service standard. 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. It means, that at least 92 % of SPPM must have transit time *D+1*. Then service standard (set postal performance level) is at least 92 % of SPPM delivered in *D+1*. It is aim of transit time and naturally of on-time performance result as well.

Because measurement of transit time is realized by representative sample of test items, the result of on-time performance is represented by estimate of on-time probability. On-time probability in Czech conditions can be defined as probability of case, when transit time of item does not exceed 1 day (*D+1*). On-time probability is interpreted by variable called weighted estimate of on-time probability $\hat{p}_{weighted}$, which also includes distribution of mail within geographical stratification according to disjunctive set of defined postal areas. The rate of variation caused by used sample design is expressed by design factor of stratified end-to-end sample df_{StrEtE} (ČSN EN 13850, 2013). It is related to sample design and on-time probability estimate.

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.

2 Confidence interval calculation

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.

Smaller width of interval leads to higher accuracy of measurement. Maximum width 2ε of this confidence interval is defined on the basis of confidence level $(1-\alpha)$ = 95 %.

Whichever confidence interval is based on probability distribution. In dependence on chosen probability distribution, we can have symmetrical interval in the form (ČSN EN 13850, 2013):

$$[\hat{p}-\varepsilon;\hat{p}+\varepsilon] \qquad (1)$$

or asymmetrical interval in the form:

$$[\hat{p}-\varepsilon_{lower};\hat{p}+\varepsilon_{upper}], \text{ where } [\varepsilon_{lower} > \varepsilon_{upper}] \qquad (2)$$

with assumption of probability $\hat{p}$ > 50 %.

If confidence interval is not symmetrical, its maximum width 2ε is defined as:

$$[2\varepsilon = \varepsilon_{lower} + \varepsilon_{upper}] \qquad (3)$$

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.

2.1 Normal confidence interval

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 (ČSN EN 13850, 2013):

$$\left[\hat{p}_{weighted} \pm 1{,}95966 * \sqrt{\frac{\hat{p}_{weighted}\left(1-\hat{p}_{weighted}\right)}{n-1} * df_{StrEtE}}\right] \qquad (4)$$

$$2\varepsilon = 3{,}91992 * \sqrt{\mathrm{var}_{StrEtE}\left[\hat{p}_{weighted}\right]} \qquad (5)$$

This normal confidence interval is symmetrical and it is frequently apllied for assessment of measurement system accuracy. Minimum sample size n_{minSS} is calculated by ε and defined as:

$$n_{\min SS} := \min n^* \quad \text{with} \quad n^* \geq df_{StrEtE} * \left[3{,}84145 * \frac{\hat{p}_{weighted}\left(1-\hat{p}_{weighted}\right)}{\varepsilon^2}\right] + 1 \qquad (6)$$

In practice design factor can be applied directly on minimum simple random sample size for minimum design sample size. For constant value of design factor, it is necessary to increase sample size.

2.2 Applicability of normal confidence interval

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. In fact discrete form of binomial distribution and on-time performance level over 85 % causes frequently insufficiency of normal approximation by average sample sizes as well. Then the result is too liberal confidence interval. It means that real confidence level is often lower than set confidence level of 95 %.

Maximum tolerable deviation by confidence interval of normal distribution apllication does not have to exceed 4 % quota of delayed mail, which fails to meet the requirements of on-time performance. This condition will be accepted with evidence of minimum figure of delayed mail, which depends on on-time performance level. For example on-time performance 92 % requires at least 45 delayed sample items for normal distribution application with maximum deviation of 0,32 % (ČSN EN 13850, 2013). Less then 45 delayed items by on-time performance 92 % is not complying with normal distribution apllication and then other more appropriate calculation methods of confidence interval must be used. These methods result in asymmetrical confidence intervals with inconsiderable deviation – such as Agresti-Coull or inverse beta interval.

2.3 Agresti-Coull approximation

Improved calculation method is called Agresti-Coull interval (Agresti & Coull, 1998). It is based on normal distribution and can be used for all sample sizes with at least 40 items. Agresti-Coull interval calculation for simple random sample is in form as follows:

$$\left[\hat{p}_{ac} \pm 1{,}95966 * \sqrt{\frac{\hat{p}_{ac}(1-\hat{p}_{ac})}{n+3{,}84145}}\right] \text{ with } \hat{p}_{ac} := \frac{x+1{,}92072}{n+3{,}84145} \qquad (7)$$

This calculation method works with effective sample size (*ESS*) with inclusion of design factor. Effective sample size of design is size of simple random sample with same sample variance as the design has, it means total design sample size divided by design factor.

For stratified end-to-end measurement system we can calculate Agresti-Coull interval for $\hat{p}_{weighted}$ in following form (ČSN EN 13850, 2013):

$$\left[\hat{p}_{acESS} \pm 1{,}95966 * \sqrt{\frac{\hat{p}_{acESS}(1-\hat{p}_{acESS})}{n_{ESS}+3{,}84145}}\right] \text{ with } \hat{p}_{acESS} := \frac{x_{ESS}+1{,}92072}{n_{ESS}+3{,}84145}, \quad (8)$$

where:

$$n_{ESS} := \max n^{*} \text{ with } n^{*} \le \frac{n}{df_{StrEtE}} \quad (9)$$

$$x_{ESS} := \max x^{*} \text{ with } x^{*} \le \hat{p}_{weighted} * n_{ESS} \quad (10)$$

Then we can set width of interval with lower and upper limits:

$$2\varepsilon = 3{,}91992 * \sqrt{\frac{\hat{p}_{acESS}(1-\hat{p}_{acESS})}{n_{ESS}+3{,}84145}} \quad (11)$$

$$\varepsilon_{lower} := \left(\hat{p}_{weighted} - \hat{p}_{acESS}\right) + 1{,}95966 * \sqrt{\frac{\hat{p}_{acESS}(1-\hat{p}_{acESS})}{n_{ESS}+3{,}84145}} \quad (12)$$

$$\varepsilon_{upper} := \left(\hat{p}_{acESS} - \hat{p}_{wieghted}\right) + 1{,}95966 * \sqrt{\frac{\hat{p}_{acESS}(1-\hat{p}_{acESS})}{n_{ESS}+3{,}84145}} \quad (13)$$

This adjusted confidence interval of normal distribution is asymmetrical and can be apllied for assessment of measurement system accuracy.

2.4 Inverse Beta approximation

The next improved calculation method uses inverse beta function (*BetaLnv*). It is based on beta distribution, which is continuous form of binomial distribution. Inverse beta function rule directly calculates confidence interval without previous calculation of sample variance and uses effective sample size as well.

Inverse beta interval calculation for simple random sample is in form as follows (Blyth, 1986; Blyth & Still, 1983):

$$[BetaLnv[0{,}025; x; n-x+1]; BetaLnv[0{,}975; x+1; n-x]] \quad (14)$$

For stratified end-to-end measurement system, we can calculate inverse interval for $\hat{p}_{weighted}$ in following form (ČSN EN 13850, 2013):

$$[BetaLnv[0{,}025; x_{ESS}; n_{ESS}-x_{ESS}+1]; BetaLnv[0{,}975; x_{ESS}+1; n_{ESS}-x_{ESS}]] \quad (15)$$

where:

$$n_{ESS} := \max n^{*} \text{ with } n^{*} \le \frac{n}{df_{StrEtE}} \quad (16)$$

$$x_{ESS} := \max x^{*} \quad \text{with} \quad x^{*} \le \hat{p}_{weighted} * n_{ESS} \qquad (17)$$

Then we can set width of interval with lower and upper limits:

$$2\varepsilon = BetaLnv[0,975; x_{ESS}+1; n_{ESS}-x_{ESS}] - BetaLnv[0,025; x_{ESS}; n_{ESS}-x_{ESS}+1] \qquad (18)$$

$$\varepsilon_{lower} := \hat{p}_{weighted} - BetaLnv[0,025; x_{ESS}; n_{ESS}-x_{ESS}+1] \qquad (19)$$

$$\varepsilon_{upper} := BetaLnv[0,975; x_{ESS}+1; n_{ESS}-x_{ESS}] - \hat{p}_{weighted} \qquad (20)$$

This confidention interval of beta distribution is asymmetrical and can be apllied for assessment of measurement system accuracy.

3 Modelling of test items sample and results of interval calculations

Modelling of test items sample is based on parameters of geographical coverage by postal services and stratification of measuring sample for two periods of 2014. 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.

Measuring system modelling for the first period is based on relevant indicators figures, which are input data for interval calculation methods. Calculated results using Eq. (4)-(20) are contained in Tab. 1. and input indicators as follows:

n = 14529 test items - sample size,

x = 13691 test items - amount of on-time mail,

$\hat{p}_{weighted}$ = 0,942212469 - weighted estimate for on-time probability,

df_{StrEtE} = 3,363586 - design factor for stratified end-to-end sample.

Tab. 1: The results of interval calculations of the 1st period

	Normal distribution	Agresti-Coull interval	Inverse beta interval
Lower limit of interval	0,935253626	0,934843	0,934834
Upper limit of interval	0,949171312	0,948796	0,948986
Width of interval	0,01391769	0,013953	0,014152

Source: authors

Measuring system modelling of the second period includes following figures of indicators used for interval calculation methods. Calculated results using Eq. (4)-(20) are contained in Tab. 2 and input indicators as follows:

n = 14966 test items - sample size,

x = 13948 test items - amount of on-time mail,

$\hat{p}_{weighted}$ = 0,935566174 - weighted estimate for on-time probability,

df_{StrEtE} = 6,768318 - design factor for stratified end-to-end sample.

Tab. 2: The results of interval calculations of the 2nd period

	Normal distribution	Agresti-Coull interval	Inverse beta interval
Lower limit of interval	0,925332	0,924532	0,924517
Upper limit of interval	0,9458	0,94509	0,945443
Width of interval	0,020468	0,020558	0,020927

Source: authors

Considering input data for interval calculations, all used methods of interval calculation are applicable and give comparable results. High level of accuracy has been reached by normal distribution with the smallest width of calculated interval, on the opossite the widest interval is beta inverse interval. Using normal confidence interval, values of $\hat{p}_{weighted}$ within cca 93-94 % bring maximum deviation of 0,24-0,28 % with requirement of increasing minimum delayed mail figure within 46-47 items (ČSN EN 13850, 2013). Lower value of weighted estimate for on-time probability and higher value of design factor has in this case lower level of accuracy with wider calculated interval.

Sample sizes necessary for real concrete measuring of on-time performance by estimate of on-time probability are sufficiently large for acceptance of used calculation methods. These methods give reliable and comparable results of confidence intervals for measuring accuracy assessment and variance of their results is very small.

Apparently the normal distribution is sufficiently acceptable approximation method also from viewpoint of next modelling of mentioned measurement system by creation of sample design and measuring of transit time with subsequent assessment of measuring accuracy. Although this approximation brings some small deviation, necessary sample sizes of measurement system and applied values of on-time performance enable to gain reliable results. Agresti-Coull and inverse beta intervals bring insignificant deviation and enable to gain reliable results as well. Applying data used by modelling, width of these intervals is very slighlty larger than width of normal confidence interval.

Conclusion

Generally the most appropriate approximation method has the smallest deviation. Method selected for measurement system accuracy should have the smallest width of calculated confidence interval.

Inverse beta approximation gives the smallest deviation, on the other hand – as the most conservative method – leads to the widest confidence interval. It uses inverse beta function. Agresti-Coull approximation gives higher deviation than in previous case and it is less conservative and transparent. Both approximations lead to asymmetrical confidence interval. Normal approximation is generally widespread and frequently used. It is symmetrical and has a generally known and understood structure. On the other hand there exist some cases, when it is not applicable.

Concrete measurement system for SPPM transit times applied in domestic conditions admits usage of normal confidention interval. By modelling of proportional sample design, this interval has given reliable results from viewpoint of measuring accuracy definition with the smallest width, but generally shows some small deviation. Agresti-Coull and inverse beta intervals apllied on the same sample design model have given accurate results as well. They are slightly wider, but have insignificant deviation.

References

Agresti, A., & Caffo, B. (2000). Simple Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures. *The American Statistician*, *54*(4), 280–288.

Agresti, A., & Coull, B. (1998). Approximate is Better than “Exact” for Interval Estimation of Binomial Proportions. *The American Statistician, 52*(2), 119-126.

Blyth, C. (1986). Approximate Binomial Confidence Limits. *Journal of the American Statistical Association, 81*(395), 843-855.

Blyth, C., & Still, H. (1983). Binomial Confidence Intervals. *Journal of the American Statistical Association, 78*(381), 108-116.

Brown, L., Cai, T., & DasGupta, A. (2001). Interval Estimation for a Binomial Proportion. *Statistical Science, 16*(2), 101-117.

Brown, L., Cai, T., & DasGupta, A. (2002). Confidence Intervals for a Binomial Proportion and Asymptotic Expansions. *Annals of Statistics, 30*(1), 160–201.

Brown, L., & Li, X. (2005). Confidence Intervals for Two Sample Binomial Distribution. *Journal of Statistical Planning and Inference, 130*(1-2), 359–375.

Newcombe, R. (1998). Interval Estimation for the Difference between Independent Proportions: Comparison of Eleven Methods. *Statistics in Medicine*, *17*(8), 873–890.

Reiczigel, J. (2003). Confidence Intervals for the Binomial Parameter: Some New Considerations. *Statistics in Medicine*, *22*(4), 611-621.

Vollset, S. E. (1993). Confidence Intervals for a Binomial Proportion. *Statistics in Medicine*, *12*(9), 809-824.

ČSN EN 13850: Poštovní služby - Kvalita služby - Měření přepravní doby služeb mezi koncovými body pro jednotlivé prioritní zásilky a zásilky první třídy (2013). Úřad pro technickou normalizaci, metrologii a státní zkušebnictví.

Vyhláška č. 464/2012 Sb., o stanovení specifikace jednotlivých základních služeb a základních kvalitativních požadavků na jejich poskytování. Retrieved from http://www.mpo.cz/dokument142643.html

Contact

Doc. Ing. Libor Švadlenka, Ph.D.

University of Pardubice – Jan Perner Transport Faculty

Studentská 95, 532 10 Pardubice

Libor.svadlenka@upce.cz

Ing. Kateřina Pojkarová, Ph.D., Ing. Daniel Salava, Ph.D.

University of Pardubice – Jan Perner Transport Faculty

Studentská 95, 532 10 Pardubice

Katerina.pojkarova@upce.cz, daniel.salava@upce.cz