ORIGINAL ARTICLE
DOI: https://doi.org/10.30932/1992-3252-2022-20-2-8

World of Transport and Transportation, 2022,
Vol. 20, Iss. 2 (99), pp. 199–204

Mathematical Methods for Validation of Data on Reliability, Operation and Maintenance of Locomotives

Igor I. LAKIN

Viktor A. MELNIKOV

***Igor I. Lakin*[1], *Viktor A. Melnikov*[2]**

[1] JSC Transmashholding (TMH), Moscow, Russia.
[2] LLC Clover Group, Moscow, Russia.
✉ *[1] i.lakin@tmholding.ru.*
✉ *[2] Lordson@yandex.ru.*

ABSTRACT

The objective of the article is to describe the method proposed by the authors and intended for validity check of initial data for managing reliability of locomotives and the entire locomotive facilities. It is shown that probabilistic-statistical methods which constitute the basis of Edward Deming's theory of enterprise variability allowing to control the quality of enterprises' products, including their reliability.

The averaged statistical data used in practices of railway transport and its locomotive economy are not homogeneous, representing what is called in popular publications «average temperature of hospital patients». The homogeneity of data is determined by their unimodality, i.e., the presence of a single process in the sample. Unsuccessful sampling leads to its bimodality and even multimodality.

The research method proposed in the article is check for unimodality of the initial data based on the consequence of the law of large numbers, according to which, with an increase in the number of data, homogeneous samples tend to one of the laws of distribution of random variable: to normal, exponential, lognormal or other known law. Therefore, and since any unimodal sample must meet a goodness-of-fit test, the article proposes to apply Pearson's chi-square test (χ^2). The unimodality of the data is suggested to be estimated through the probability of compliance with the chosen law of distribution of a random variable considering the probability of more than 0,3 (30 %) to be sufficient.

Using the example of locomotive operation data and on-board microprocessor systems data, data are shown that cannot really be unimodal, as well as the data that require changing the sampling rules to achieve unimodality. For example, when considering the average daily runs of locomotives by series at specific home depots with participation in one type of traffic (main traffic, export, or shunting operations), their unimodality is achieved. An attempt to enlarge the data (to consider several series, several polygons, etc.) results in loss of unimodality.

The article considers the unimodality of data of MSU-TP on-board microprocessor control systems for diesel locomotives of 2TE116U series. The expected operating time for the positions of the driver's controller turned out to be multimodal data. Surprisingly, the current of the traction motors turned out to be unimodal, regardless of the driving position of the driver's controller.

Keywords: railway transport, locomotives, reliability, verification.

For citation: Lakin, I. I., Melnikov, V. A. Mathematical Methods for Validation of Data on Reliability, Operation and Maintenance of Locomotives. World of Transport and Transportation, 2022, Vol. 20, Iss. 2, (99), pp. 199–204. DOI: https://doi.org/10.30932/1992-3252-2022-20-2-8.

The text of the article originally written in Russian is published in the first part of the issue.
Текст статьи на русском языке публикуется в первой части данного выпуска.

INTRODUCTION

The task of scientific organisation of production, quality control of technological processes was first solved in the late 19^{th}–early 20^{th} centuries by Frederick Taylor [1]. At the beginning of 20^{th} century, the principles of product quality and cost management were scientifically and practically developed by Henry Gantt (the most famous is the «Gantt chart», although this is by no means the only one and not his main development). A qualitatively new stage in production management in the middle of 20^{th} century is associated with the names of Walter Shewhart and Edward Deming, who introduced the obligatory application of the principle of continuous improvement (PDCA cycle) into management. Taking as a basis the thesis of constant variability of the production situation, E. Deming developed the theory of enterprise variability, making the practice of applying probabilistic-statistical methods in management mandatory [1].

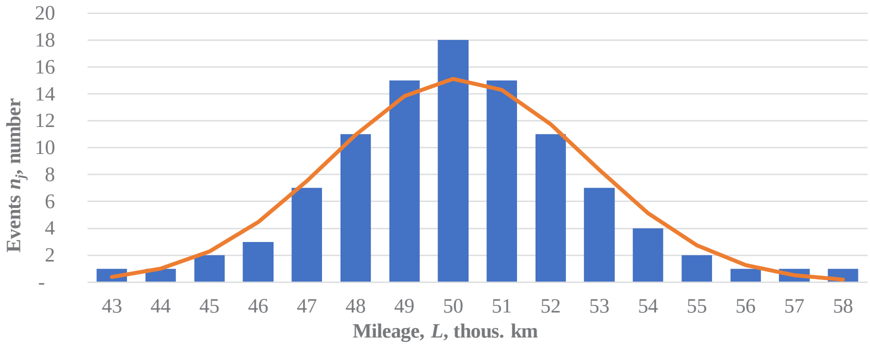
The development of the theory of E. Deming in relation to the domestic locomotive facilities in the second half of 20^{th} century was carried out by an outstanding scientist, D.Sc., Professor Igor P. Isaev, who not only adapted the mathematical apparatus to the transport environment, but also practically implemented system in the Rybnoye locomotive depot [2]. I. P. Isaev developed a theory of application of probabilistic-statistical methods in the study of reliability of locomotives, which was developed by his students and followers [3]. Probabilistic-statistical methods have become an obligatory part of the educational process [4].

The current stage of development of the theory of reliability of locomotives is associated with development of computer technology (IT-technologies). D.Sc. (Eng) Igor K. Lakin headed the development of an automated control system (ACS) for locomotive facilities (ACST) [5] which is the basis of the control system for the locomotive divisions of JSC Russian Railways. As a development of ACST, I. K. Lakin and his students created ACS for maintenance and repair (MR) of locomotives called ACS «Network Schedule» (ACS NS) to manage MR in 85 service locomotive depots of the LocoTech group of companies [6; 7]. ACS NS encapsulates the scientific provisions of quality management, Lean Production, TQM, etc. [13–15], as a result, five Ph.D. theses and one doctoral dissertation were prepared by the developers.

Since July 2014, JSC Russian Railways has switched to a service system for MR of locomotives, in which service companies (representatives of manufacturers) receive income (revenue) not for the volume of MR performed, but for the useful mileage of locomotives (rubles/km). A reduction factor is provided if the reliability level (SLA) specified in the service agreement is not met. All newly purchased locomotives by JSC Russian Railways are supplied under a Life Cycle Contract (LCC), which establishes more stringent requirements for reliability of locomotives. Estimating maintenance costs 40 years ahead is a difficult task. The transition from the budgetary form of payment for maintenance and repair to the life cycle requires the use of probabilistic and statistical methods: it is necessary to predict the volume of maintenance and repair both within the framework of a preventive maintenance system and for over-cycle unscheduled work due to failures, the need for modernisation, etc. The role of variable management methods is increasing.

During the operation of locomotives, the work performed by them (in t•km, kW•h, million km, etc.) is recorded in the information systems of JSC Russian Railways [5]. At the same time, the on-board microprocessor control systems (MCS) of the locomotive, according to information from the sensors, records information about the operating modes of the equipment, and technical diagnostics are carried out [6; 7]. This information is the starting point for organisation of MR as a scheduled preventive maintenance with elements of predictive repair (taking into account the actual technical condition of the locomotives and the work performed by them).

When a locomotive enters a service locomotive depot (SLD), MR analyses the information accumulated during its operation, performs additional diagnostics using portable and stationary systems (vibration diagnostics, rim profile measurement, insulation resistance measurement, etc.). The results of a visual inspection, the driver's comments from the logbook are also used. All information is collected in an electronic diagnostic card, which is the basis for appointment of over-cycle work to eliminate failures and pre-failure states. An electronic work order is issued for each scheduled and over-cycle work, which records labour costs, materials and spare parts received from the warehouse, and in the future, it will integrate the



Pic. 1. An example of distribution of overhaul mileage, thousand km (columns – actual data, graph – theoretical distribution).

recordings on the tools and equipment used during maintenance and repair. The demurrage of a locomotive is recorded as for various operations. As a result, during the operation, and maintenance and repair of locomotives, a comprehensive database is formed throughout the entire life cycle of locomotives using paperless technology. It is important that the data are the result of operation and maintenance processes, and not the information entered manually.

Information about the life cycle of locomotives is needed not only to manage MR production processes, but also to analyse the effectiveness of organisation of operation and MR of locomotives. A modern interoperable system ACS for design, production, operation, and maintenance of locomotives generates a large amount of data (Big Data), the analysis of which is not possible without programmable analysis algorithms. But automated data processing has the risk of drawing false conclusions.

Mark Twain, in his book Chapters of My Biography, quoted the British Prime Minister «There are three kinds of lies: lies, damned lies and statistics». This well- known expression emphasises the presence of risks in formal mathematical processing of data. One of them is known as the «average temperature of hospital patients»: mixing dissimilar data. With an average temperature, the error is obvious. In the practice of locomotive divisions, averaged values are often used to assess the efficiency of operation, and maintenance and repair of locomotives: speed on a section, downtime for maintenance and repair and its waiting, failure rate by type of equipment, required frequency of maintenance and repair by type of equipment, etc. In this case, unimodality is not checked.

Therefore, the *objective* of the ongoing work study is to automate the process of checking statistical data for validity.

As a research *method*, verification of the distribution of the studied quantity for unimodality was chosen, which is possible due to the property of the law of large numbers [4]: with an increase in the sample, the probability of the occurrence of a particular value approaches a stable value.

RESULTS

The unimodal process takes the form of a standard distribution of a random variable: normal, exponential, lognormal, or others: unimodality can be estimated by the criterion of goodness-of-fit for testing hypotheses. Let us consider an example of the distribution of locomotives' mileage L between current repair-1 (Pic. 1 and Table 1).

Four samples were formed with the same distribution proportion, but with a different number of data N = {100, 1000, 10,000, 100,000}. The distributions have the same mathematical expectation L_n = 50,12 thousand km and almost the same standard deviation σ_n [4]:

$$L_n = \frac{1}{N_n}\sum_{i=1}^{N_n} L_{ni}\ , \qquad (1)$$

$$\sigma_n = \sqrt{\frac{1}{N_{n-1}}\sum_{i=1}^{N_n}(L_n - L_{ni})^2}\ , \qquad (2)$$

where L_n – mathematical expectation of the sample $n \in N$ = {100, 1000, 10 000, 100 000};

L_{ni} – i-th mileage of the sample $n \in N$ = {100, 1000, 10 000, 100 000};

N_n – maximum quantity of data in the sample $n \in N$.

According to Pearson's goodness-of-fit test, the chi-square χ^2_n is calculated for each sample $n \in N$:

Table 1

Examples of distributions of overhaul mileage [performed by the authors]

i	L_i, thous. km	Sample size actual N				Sample size theoretical N_t				
		100	1 000	10 000	100 000	f(Li)	100	1 000	10 000	100 000
1	43	1	10	100	1 000	0,0040	0	4	40	395
2	44	1	10	100	1 000	0,0102	1	10	102	1 024
3	45	2	20	200	2 000	0,0230	2	23	230	2 298
4	46	3	30	300	3 000	0,0446	4	45	446	4 465
5	47	7	70	700	7 000	0,0751	8	75	751	7 513
6	48	11	110	1 100	11 000	0,1095	11	110	1 095	10 951
7	49	15	150	1 500	15 000	0,1382	14	138	1 382	13 823
8	50	18	180	1 800	18 000	0,1511	15	151	1 511	15 111
9	51	15	150	1 500	15 000	0,1431	14	143	1 431	14 308
10	52	11	110	1 100	11 000	0,1173	12	117	1 173	11 733
11	53	7	70	700	7 000	0,0833	8	83	833	8 333
12	54	4	40	400	4 000	0,0513	5	51	513	5 126
13	55	2	20	200	2 000	0,0273	3	27	273	2 731
14	56	1	10	100	1 000	0,0126	1	13	126	1 260
15	57	1	10	100	1 000	0,0050	1	5	50	503
16	58	1	10	100	1 000	0,0017	0	2	17	174
	L_n	50,12	50,12	50,12	50,12					
	σ	2,64	2,63	2,62	2,6					

Table 2

Probability of compliance with the normal law in χ^2 [performed by the authors]

	Sample size actual N			
N_n	100	1 000	10 000	100 000
χ^2	7	73	732	7 325
P_n	0,3	0	0	0

$$\chi_n^2 = \sum_{i=1}^{N_n} \frac{(L_{ni} - L_{theori})^2}{L_{theori}}, \tag{3}$$

where L_{theori} – theoretical number of hits into the range for normal distribution of a random variable, calculated by the distribution density function *fx* [4] (see Table 1):

$$L_{theor} = \int_{xn1}^{xn2} f(x)dx, \tag{4}$$

The probability of compliance with the normal law P_n is determined from the Pearson's table. Table 2 shows the calculation results: only for $N = 100$ there is a significant probability of compliance with the normal law $P_{100} = 0{,}3$.

Four identical distributions have completely different probability of fitting the law: the probability of fitting is extremely sensitive to sample size. In the practice of the locomotive economy, samples will always be of different sizes and noisy even with large samples. Orientation to the minimum of them will reduce the amount of useful information. The Pearson's criterion is effective and sensitive for strictly unimodal samples and is not suitable for encapsulation in ACS.

The considered example was checked with Kolmogorov–Smirnov criterion, according to which it is necessary to find the maximum difference between the actual L_{ni} and theoretical L_{theori} distributions:

$$\Delta_{max} = Max(f_{ai} - f_{ti}), \tag{5}$$

where $f_{ai} = N_i / N;$

$f_{ti} = N_{ti} / N_t;$

N_i – number of actual hitting in the range *i*;

N – empirical sample size;

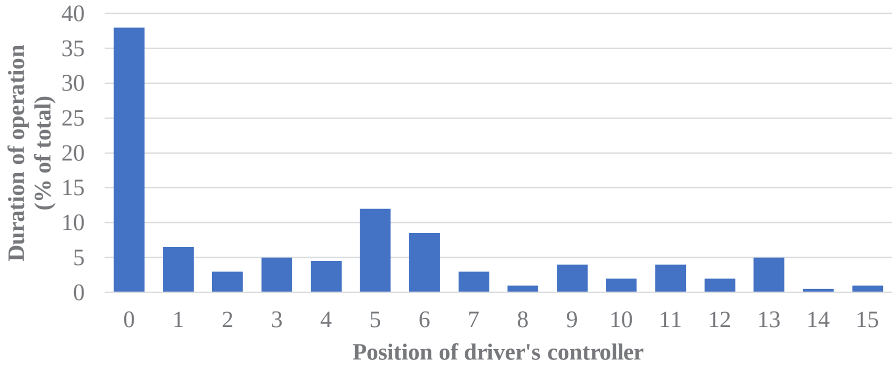
N_{ti} – number of theoretical hitting in the range *i*;

N_t – control sample size ($N_t = N$).

Next, it is necessary to calculate the indicator *D*, which for the normal law is:

$$D_\varphi = \Delta_{\text{max}} \bullet (N^{0,5} - 0{,}01 + 0{,}85 / N^{0,5}), \tag{6}$$

Table 3 shows the results of the calculation: with an increase in the sample, the probability of compliance decreases more slowly than according to the Pearson's goodness-of-fit criterion: with



Pic. 2. Distribution of operating time of diesel locomotive's DGUs by positions [performed by the authors].

Table 3

Probability of compliance with the normal law according to D_φ [performed by the authors]

N	100	1000	10000	100000
Δ_{max}	0,029	0,029	0,029	0,029
D	0,2910	0,9139	2,8885	9,1341
P_n	0,9999	0,3791	0,00000018	0

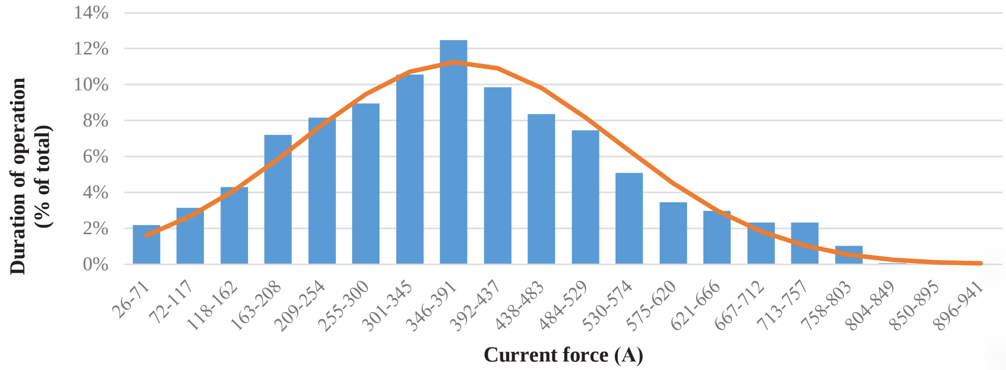
samples less than 1000, the unimodality check by D_φ is effective.

Checking the unimodality of data is implemented by the authors in their practices. Let us give examples of real data.

One of the important indicators of operation of locomotives is the average average speed when it is running through the section. Analysis of data from onboard microprocessor control systems (MCS) showed the presence of at least two peaks. For example, diesel locomotives of 2TE116U series on Oktyabrskaya railway showed there are three peaks – in the ranges of 18–25 km/h, 34–41 km/h and 50–57 km/h, on Privolzhskaya – two peaks: 42–49 km/h and 58–65 km/h, on Sverdlovskaya railway: 26–31 km/h, 68–73 km/h.

Interesting data were obtained by the authors on operation of the diesel generator set of diesel locomotives of 2TE116U series for 3164 days of operation (120 sections of locomotives were considered at three network segments). As expected, distribution of operating time by positions turned out to be multimodal (Pic. 2), but the current of traction motors turned out to be unimodal (Pic. 3).

In conclusion, it should be noted that the analysis of publications world-wide shows that in statistical materials on railway transport, insufficient attention is paid to checking unimodality of data, that affects the reliability [8–12]. However, in general, a few publications were devoted to the problem of checking data validity (e.g., [13–15]).



Pic. 3. TED current distribution for diesel locomotives of 2TE116U series [performed by the authors].

CONCLUSIONS

1. Automation of the process of managing the life cycle of locomotives in general and of their maintenance and repair (MR) in particular makes it possible to encapsulate mathematical control methods, including probabilistic-statistical ones, in software according to Edward Deming's theory of variability.

2. In probabilistic-statistical processing of data on operation of locomotives and their maintenance and repair, before making control decisions, it is necessary to check data for unimodality as belonging to a single process. In a bimodal or multimodal process, data should be divided into groups, thereby achieving unimodality.

3. It is proposed to check for unimodality in automated control systems (ACS) by checking the compliance of the hypothesis of unimodal distribution as of the presence of a significant probability that distribution of the considered parameter corresponds to one of the distribution laws of a random variable. Kolmogorov–Smirnov agreement criterion is preferable.

4. A comprehensive analysis of data on operation and maintenance and repair of parameters of diesel generator sets and traction motors of locomotives showed that data of one series of locomotives of one home depot with participation in one type of traffic (main traffic, export or shunting operations) have unimodality: local operating conditions, weight norms, profile and speed mode lead to loss of unimodality when trying to generalise data.

5. The revealed unimodality of the current of traction electric drives (TED) of diesel locomotives in traction mode and according to the totality of work at all traction positions of the diesel generator set makes it possible to take this fact into account when designing TED.

REFERENCES

1. Lakin, I. K., Suprun, V. N. The history of creation of quality management systems (QMS) and the features of their implementation in railway transport [*Istoriya sozdaniya sistem menedzhmenta kachestva (SMK) i osobennosti ikh vnedreniya na zheleznodorozhnom transporte*]. Krasnoyarsk, KF IrGUPS, 2006, 91 p.

2. Strelnikov, V. T., Isaev, I. P. Integrated management of quality of maintenance and repair of electric locomotives [*Kompleksnoe upravlenie kachestvom tekhnicheskogo obsluzhivaniya i remonta elektrovozov*]. Moscow, Transport publ., 1980, 207 p.

3. Gorsky, A. V., Vorobyov, A. A. Reliability of electric rolling stock [*Nadezhnost elektropodvizhnogo sostava*]. Moscow, Marshrut publ., 2005, 303 p.

4. Ventzel, E. S., Ovcharov, L. A. Probability theory and its engineering applications [*Teoriya veroyatnostei i ee inzhenernie prilozheniya*]. Moscow, Academiya publ., 2000, 464 p.

5. Lakin, I. K., Smirnov, Yu. V., Timchenko, A. Yu. [*et al*]. Automated control system for locomotive facilities. ACST [*Avtomatizirovannaya Sistema upravleniya lokomotivnym khozyaistvom. ASUT*]. Moscow, Industry center for introduction of new equipment and technologies, 2002, 516 p.

6. Lipa, K. V., Pustovoi, V. N., Lyangasov, S. L., Lakin, I. K., Abolmasov, A. A., Lakin, I. I., Melnikov, V. A. [*et al*]. Monitoring of technical condition and modes operation of locomotives. Theory and practice [*Monitoring tekhnicheskogo sostoyaniya i rezhimov ekspluatatsii lokomotivov. Teoriya i praktika*]. Moscow, LocoTech publ., 2015, 212 p.

7. Lakin, I. K., Pustovoi, V. N., Abolmasov, A. A. Automated control system for operational life cycle of locomotives [*Avtomatizirovannaya Sistema upravleniya ekspluatatsionnym zhiznennym tsiklom lokomotivov*]. *Proceedings of All-Russian scientific and practical conference with international participation «Operation and maintenance of electronic and microprocessor equipment of traction rolling stock»*. Krasnoyarsk, DCV of Krasnoyarsk railway, 2020, pp. 223−242.

8. Bonadero, A. Statistical methods and dynamic railway. *Ingegneria Ferroviaria,* December 2011, Vol. 66 (12), pp. 1087–1107.

9. Ghosh, Saptarshi; Banerjee, Avishek; Sharma, Naveen; Ganguly, Niloy; Bhattacharya, Saurav; Mukherjee, Animesh. Statistical Analysis of the Indian Railway Network: a Complex Network Approach. *Acta Physica Polonica B Proceedings Supplement,* 2011, Vol. 2, pp. 123–137. DOI: 10.5506/APhysPolBSupp.4.123.

10. Railway Statistics. Indian Railways. 108 p. [Electronic resource]: http://122.252.243.98/ZRTI1/statistics.pdf. Last accessed 24.03.2022.

11. Reference Manual in Rail transport statistics. Eurostat. January 2021. [Electronic resource]: https://ec.europa.eu/eurostat/documents/29567/3217334/Reference_manual_for_rail_transport_statistics%E2%80%93version_10.4_%282021_edition%29.pdf/b152c85c-11d6-cff4-a1a8-d34bd4cac839?t=1610708187248. Last accessed 24.03.2022.

12. Rail transportation occurrences in 2020. Transportation Safety Board of Canada. [Electronic resource]: https://www.bst-tsb.gc.ca/eng/stats/rail/2020/sser-ssro-2020.html. Last accessed 24.03.2022.

13. Yi, Mike. A Complete Guide to Histograms. Chartio. [Electronic resource]: https://chartio.com/learn/charts/histogram-complete-guide/. Last accessed 24.03.2022.

14. Unimodal Distribution: Overview. Statistics How To. [Electronic resource]: https://www.statisticshowto.com/unimodal-distribution-2/. Last accessed 24.03.2022.

15. What Is Statistical Process Control? Infinity QS. [Electronic resource]: https://www.infinityqs.com/resources/what-is-spc. Last accessed 24.03.2022. ●

Information about the authors:

Lakin, Igor I., *Ph.D. (Eng), Head of the Direction for Technical Management of JSC Transmashholding (TMH), Moscow, Russia, i.lakin@tmholding.ru.*

Melnikov, Viktor A., *specialist, Senior Expert of LLC Clover Group, Moscow, Russia, melnikov.viktor@corp.itclover.ru.*

Article received 01.03.2022, approved 24.03.2022, accepted 31.03.2022.