Original article
DOI: 10.21683/1729-2646-2018-18-2-38-41

Method of conversion of MTBF from cycles to kilometers travelled

Maria V. Belousova, *PF KMT, Saint Petersburg, Russia*
Vitaly V. Bulatov, *PF KMT, Saint Petersburg, Russia*

Maria V. Belousova

Vitaly V. Bulatov

Abstract. **Aim.** *At machine-building enterprises, dependability indicators are evaluated at the stage of product design and later based on operational data. At the design stage, automated software systems are widely used and employ a number of methods of dependability indicators calculation: fault trees, Markov chains, etc. The input data for such calculations are based on the analysis of a product design and properties of its units and elements.In operation, the dependability indicators are analyzed totally differently. Failure information processing involves deficiency reports filed by customers and operating organizations to the service departments of manufacturing enterprises. The total number of failures for all types of products must be evaluated by the dependability service/unit within a specified period of time. This failure data processing procedure is required for the calculation of reliability and maintainability indicators. The resulting numerical characteristics are compared with the standard values set forth in the technical documentation. Based on this comparison the conclusion is made regarding the compliance or non-compliance of a specific product with the specified dependability requirements. The values of the dependability indicators given in the technical documentation are based on the results of dependability testing of prototypes. However, due to the difference in the test conditions, results recording procedures and measurement units, the values of dependability indicators set forth in the technical documentation and collected in the course of operation are not comparable. In the wagon-building industry, the operation time of rolling stock is normally measured in kilometers travelled. However, the operation of a large number of wagon components is measured in cycles, hours, etc. In most cases these measurement units are used to express the values of dependability indicators obtained during prototype testing. In the process of reliability evaluation of plug doors installed in commuter trains, it became necessary to approximately convert the operation time expressed in opening/closing cycles into operation time expressed in kilometers travelled. In light of the emerged problem it was decided to construct a mathematical model that would best reflect the association between the two values. In most cases mathematical models are constructed and verified using the initial observations of the given indicator and the explanatory factors. In this case, the input data is one factor (opening/closing cycles) and one indicator (kilometers travelled), therefore the pair linear regression model can be used.* **Results.** *The correlation between the opening/closing cycle of plug doors and the kilometers travelled by a commuter train was analyzed. The model of pair linear regression was then generated. Verification was conducted, the outcome of which gives ground for the conclusion regarding the representativeness of the resulting data.* **Conclusions.** *The presented method of calculation of the generalizing controllable dependability indicator (mean time between failures) with the example of plug doors shows that the model of pair linear regression can be used for conversion of mean time between failures from cycles into kilometers travelled required for the evaluation of dependability indicators in operation.*

Keywords: *dependability, mean time between failures, rolling stock, plug door, pair linear regression*

For citation: *Belousova MV, Bulatov VV. Method of conversion of MTBF from cycles to kilometres travelled. Dependability 2018;2: 38-41. DOI: 10.21683/1729-2646-2018-18-2-38-41*

Introduction

In the wagon-building industry, the most important factor in the evaluation of dependability indicators is obtaining reliable information on the number and type of failures. Many units of passenger cars operate in cycles, e.g. stepboards, suspensions, doors, etc. The specifications for rolling stock components must contain the values of dependability indicators expressed in cycles, hours, kilometers travelled. The value of operation time in cycles is identified by calculation and confirmed through dependability tests. Evaluating the mean time between failures in cycles over the course of operation is complicated or in some cases impossible. For this reason calculations normally involve the time to failure expressed in kilometers travelled, as in the process of operational testing monitoring statistical data using this continuous value is most practical.

This poses the question of the need for a reliable procedure of operation time values conversion. In Russian literature this matter has not been considered in depth despite its high practical significance. Given the above, the need for the mentioned method is ever more relevant.

Input data

The aim of this research is to identify the type of dependence between the dependent variable y (kilometers travelled by train) and dependent variable x (opening/closing cycles of doors). In similar cases, in technical, socioeconomic and other research, regression analysis is used.

Let us consider the use of the pair linear regression model with the example of conversion of the values of operation time from cycles to kilometers for plug doors of commuter trains.

In accordance with the technical documentation, the controllable indicator of door dependability is the door's mean time to failure $\left(T_{\text{TD}}^{\text{KM}}\right)$, not less than 300 000 opening/closing cycles.

At the first stage, a sample was generated that covered 17 EMU depots (Figure 1) and 27 commuter lines.

The distance between the stations of each line of the sample was evaluated using open source information [2, 3].

As the subjects of research were chosen l_i, the number of opening/closing cycles, and S_i, the distance in kilometers travelled by the train within the number of cycles l_i in the i–th direction, $i = 1, 2, \dots, n$ (Table 1).

Table 1

i	l_i	S_i, km	i	l_i	S_i, km	i	l_i	S_i, km
1	11	20.853	10	19	40.0389	19	27	93.6574
2	12	36.5727	11	21	53.6669	20	28	62.5396
3	12	26.9275	12	21	61.2759	21	29	66.8559
4	13	34.0387	13	22	58.8737	22	30	62.365
5	13	35.8461	14	22	46.6185	23	34	64.4684
6	13	55.4732	15	23	50.6731	24	39	106.451
7	13	34.7382	16	23	42.7334	25	42	105.483
8	14	32.3032	17	24	59.7328	26	47	129.837
9	15	28.8071	18	24	43.8535	27	47	102.54

According to the primary premises of regression analysis, the number of observations must exceed the number of regression parameters included in the model, otherwise the regression parameters become statistically insignificant [5].

Model of pair linear regression

The empirical method of identifying the functional dependence comes down to evaluating unknown parameters using the least squares method. It is assumed that the factor and indicator are associated with the $y = \lambda + \beta x + \varepsilon$ dependence. First, function $\hat{f}(x)$ is chosen, the values of the parameters of which are identified in such a way as to minimize the sum of deviation squares of actual values of the attribute y_i from the expected value $\hat{y}_i = \hat{f}(x_i)$:

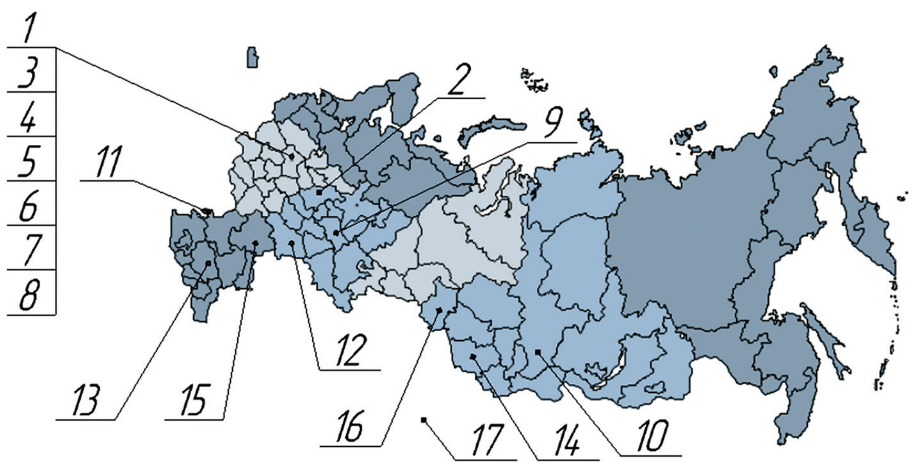

Figure 1.Multipleunitdepot: 1, Aeroxpress; 2, Gorky-Moskovsky; 3, Lobnia (TCh-14 MSK); 4, Aprelevka (TChPRIG-20); 5, Moskva 2, Yaroslavskaya; 6, Nakhabino (TCh-17 MSK); 7, Ramenskoye (TCh-7 MSK); 8, Zhelezhodorozhnaya; 9, Kazan (TChM-17); 10, Karsnoyarsk; 11, Rostov; 12, Anisovka (TChM-14); 13, MoneralnyeVody; 14, Altayskaya; 15, Volgograd; 16, Omsk; 17, Karaganda (Kazakhstan)

$$\min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

Based on the input data, the diagram was constructed of the dependence between the indicator (km) and the factor (cycles), then the a and b regression coefficients were calculated, as well as the values of $\hat{y}_i$, the regression line was plotted on the correlation field (Figure 2).

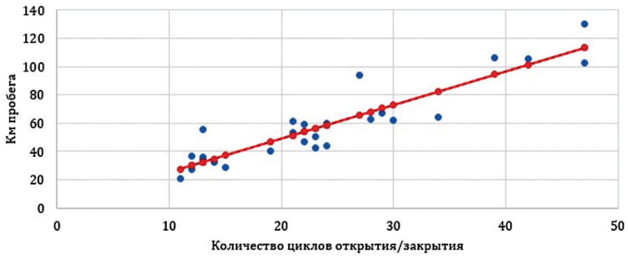

Figure 2. Diagram of dependence between indicator (km) and factor (cycles) and linear regression

Thus, the equation of pair linear regression is as follows:

$$\hat{y} = 2{,}38684x + 1{,}27483.$$

Verification of the model

In order to prove the correctness of the resulting equation of pair linear regression, let us use hypotheses of statistical significance of the obtained evaluations.

3.1 Verification of the significance of the correlation coefficient

The measure of the strength of linear connection between two random variables is the Pearson linear correlation coefficient that is evaluated with the sample correlation coefficient r_{xy}. In this case r_{xy} = 92%. In order to verify the hypothesis $H_0$of statistical significance of coefficient r_{xy}statistic $t_r = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$ is calculated, which– if the alternative hypothesis H_1 is true – has the Student's distribution with the number of degrees of freedom n–2. We obtained t_r = 57.5, which is higher than t_{tabl} = 2.06, that is identified out of the Student's distribution table under n–2 degrees of freedom as the critical point that corresponds to the two-sided critical region with the level of significance of 5%. Therefore, coefficient r_{xy} can be deemed significant and, according to the Chaddock's scale, the strength of connection between the indicator and the factor is quite high.

3.2 Verification of the significance of the linear regression

Let us also verify the significance of the linear regression in general. To do that, let us calculate the determination coefficient R^2 and statistic F in formula:

$$F = \frac{R^2}{1-R^2} \cdot (n-2).$$

If the value of this statistic is higher than the critical value under the level of significance of 5%, then hypothesis H_0 on the insignificance of linear regression is discarded. By substituting input data we obtain F = 132.4, which is higher than the critical point per Fisher's distribution table F_{tabl} = 4.24 with (1, n–2) degrees of freedom, therefore the constructed regression equation is statistically significant.

3.3 Verification of homoscedasticity hypothesis

One of the primary assumptions of regression analysis is the homoscedasticity assumption that consists in the equality of dispersions of observations:

$$D(y_i) = \sigma^2, \quad i = 1, \ldots, n.$$

Non-fulfilment of this assumption deteriorates the quality of evaluation of unknown parameters. Homoscedasticity is identified using the Goldfeld-Quandt method [1]. For that purpose, m central observations are excluded from the sample and two independent regression models are constructed, for each of which residual sums of squares are calculated:

$$\tilde{S}_1^2 = \sum_{i=1}^{\frac{n-m}{2}} (y_{1_i} - \hat{y}_{1_i})^2,$$

$$\tilde{S}_2^2 = \sum_{i=\frac{n+m}{2}+1}^{n} (y_{2_i} - \hat{y}_{2_i})^2.$$

Then, statistic $\tilde{F} = \frac{\tilde{S}_2^2}{\tilde{S}_1^2}$ is calculated. If the hypothesis is correct, the F-statistic has Fisher's distribution with $\left(\frac{n-m}{2} - 2; \frac{n-m}{2} - 2\right)$ degrees of freedom. We obtained the value $\tilde{F} = 2{,}58$, while the critical value per Fisher's distribution table is F_{tabl} = 3.79. As $\tilde{F} < F_{\text{tabl}}$, the homoscedasticity hypothesis is accepted.

Conclusion

Let us construct a point prediction $\hat{y}_p = a + b \cdot x_p$ of indicator $y_p$for the cased when x_p = 300000 and find the confidence interval of the resulting prediction with the level of confidence of 0.95:

$$\hat{y}_p - m_y \cdot t_{tabl} < y_p < \hat{y}_p + m_y \cdot t_{tabl},$$

$$\text{where } m_y = S_{res} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}.$$

We will obtain the following prediction value: y_p = 1.274 83 + 2.38684·x_p = 716 053, for which the confidence interval is (58788; 844 217).

Thus, we obtain the value of the controllable dependability indicator (mean time between failures) $T_{\text{TD}}^{\text{KM}} = 716053 \approx 700000$ that was found using the pair linear regression equation.

The obtained results are to be used as the controllable dependability indicator in the evaluation of the reliability level of plug doors. This approach can be recommended for the evaluation of the mean time to failure of other components of ground passenger transport vehicles that operate in cycles.

References

[1]. Bure VM, Yevseev EA. Osnovyekonomentriki [The fundamentals of econometrics]. Saint Petersburg: IzdatelstvoSPbGU; 2004 [in Russian].

[2]. <https://www.google.ru/maps> (accessed 05.05.2017).

[3]. <https://yandex.ru/maps> (accessed 05.05.2017).

[4]. GOST 27.301-95. Industrial product dependability. Dependability calculation. General provisions [in Russian].

[5]. Förster E, Rönz B. Methods of correlation and regression analysis. Guidebook for economists. Moscow: Financy i statistika; 1983 [in Russian].

About the authors

Maria V. Belousova, Dependability Engineer, PF KMT, Saint Petersburg, Russia, e-mail: 27bmw1993@mail.ru

Vitaly B. Bulatov, Candidate of Engineering, Dependability Engineer, PF KMT, Saint Petersburg, Russia, e-mail: bulatov-vitaly@yandex.ru

Received on 06.11.2017