Application of the Polynomial Function in the Analysis of Statistical Indicators of Risk and Safety in Shipping

Jasminka Bonato, Biserka Draščić Ban

University of Rijeka, Faculty of Maritime Studies, Studentska 2, 51 000 Rijeka, Croatia, e-mail: jasminka.bonato@uniri.hr

ABSTRACT

Navigation safety issues are of common interest to all maritime countries due to their actions in a complex and risky environment. The world's most serious ship accidents can cause major economic losses, loss of life, and severe environmental pollution in the sea.

The data included in trust and security databases are usually collected from different sources. The accuracy of the collected data may vary from source to source due to a greater number of external and internal factors such as different manufacturers, and working and environmental conditions. The quality of the data may vary due to the incompleteness of the details, subjectivity, and experience of the data collector.

Databases on the frequency of failures, the estimation of which can be based on recorded failure events, are difficult to access. Therefore, the aim of the research, based on the available statistical indicators of maritime risks and safety, is to indicate the application of mathematical statistics, analysis and assessment of unwanted events. The statistical methods proposed in the paper aim to point out the importance of prediction and the possibility of improving safety. In this sense, the polynomial function was applied. To visualize the results obtained during the research, the Excel program is used.

ARTICLE INFO

1 Introduction

The international shipping industry is responsible for transporting about 90% of world trade, so the safety of vessels is essential [1]. During the early 1990s, the global fleet lost more than 200 ships a year. South China, Indochina, Indonesia, and the Philippines are the main focuses of global losses. The Southeast Asian area also had its biggest losses in a decade, Figure 1.

During this research, the analysis of total losses in the period from 2015 to 2021 was made. The type of vessel was taken into consideration, as well as the cause of losses, within the period from 2011 to 2020. The malfunction of the device while sinking stands as one of the causes, but so do fire, explosion, and collision. People interact with different types of equipment to perform work tasks. Therefore, the risks of different types of violations when performing work tasks can occur. Regarding this notion, Allianz Global Corporate & Specialty (AGCS) offers maritime risk advice for loss prevention [2]. In this research, estimates based on statistical observations, aim to highlight the possibility of risk management [3, 4]. In case the system failure is the cause, this assessment is possible by calculating the time leading up to the failure, as well as the number of failures per system component [5, 6].

2 Research

The paper analyzed the available statistical data on total losses in the period from 2015 to 2021. The data are displayed in Table 1 and presented graphically in Excel chart. The average number of losses was determined, as well as the average rate of change, and the calculation of the trend forecast of the observed phenomenon in shipping. [7, 8]. The number of losses by vessel type and cause in the period from 2011 to 2020 was also analyzed, Table 3. In the analysis, the polynomial function was applied, because the data does not show constant monotonous growth or decline.

Figure 1 Numerical losses in international shipping [1]

3 Methods (statistical and computational, application of Excel)

Empirical methods evaluate the risk based on existing data on the frequency of the occurrence of an adverse event. The available data shows the total losses between 2015 and 2021, Table 1. For transparency, the data is presented graphically (Figure 2).

Monitoring of various phenomena (economic and other) is important, as is the statistical processing of such data. When there is a set of chronologically edited values of the phenomenon, i.e. time series, the average number of losses in the observed period can be calculated by the expression [7].

$$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{564}{7} = 80.57 \approx 81 \qquad (1)$$

where the *Y* code refers to the frequency, and *N* is the total number of time units of the observed period. i.e. the value of the observed variable (number of losses) between 2015 and 2021.

It can be concluded that the number of losses in 2021.54, represents a significant improvement over the seven-year average number of losses of 81.

Table 1 Total Shipping Losses by Year [1]

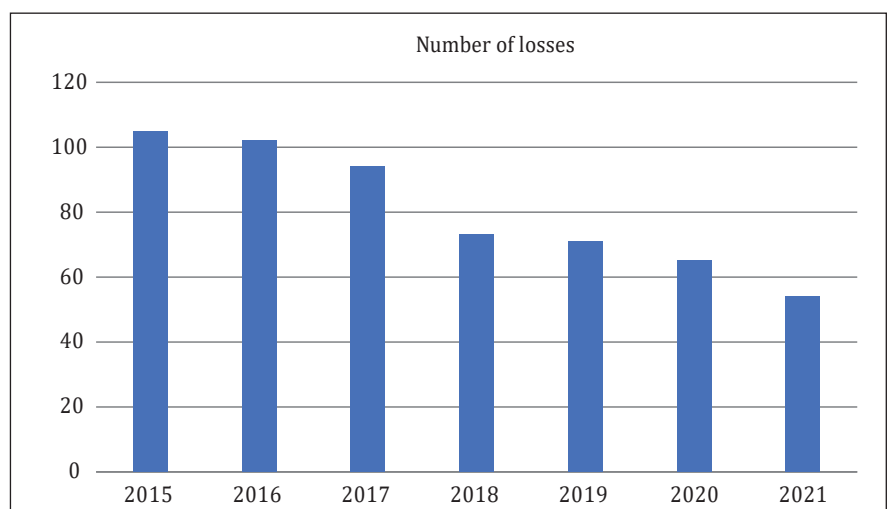| Year	Number of losses, Y_i
2015	105
2016	102
2017	94
2018	73
2019	71
2020	65
2021	54
Σ	564

Figure 2 Total shipping losses in the observed period

Source: Authors

Assuming that the values of total losses in shipping will continue in the future in the same way, the average rate of change can be calculated, and in the observed period, using the geometric mean, starting from the last element (Y_N) in the series, a prediction of its movement can also be made. The average annual rate of change means the average relative change in the value (in %) of a phenomenon during one time unit of the observed time period. It is calculated according to the expression:

$$\bar{S} = (G-1)\cdot 100, \tag{2}$$

where is

$$G = \sqrt[N-1]{\frac{Y_N}{Y_1}}, \tag{3}$$

Y_N is the last, Y_1 is the first element in the observed period.

Table 1 data account:

$$G = \sqrt[N-1]{\frac{Y_N}{Y_1}} = \sqrt[6]{\frac{54}{105}} = 0.859 \tag{4}$$

$$\bar{S} = (G-1)\cdot 100 = (0.8594-1)\cdot 100 =$$
$$= -0.1045 \cdot 100 = -10.45\% \tag{5}$$

The rate is negative, which means that in the observed seven-year period, the annual number of losses decreased by an average of 10,45% per year. Based on the calculated rate, assuming that the values of a phenomenon will continue to move in the future in the same way, that is, according to the calculated average rate of change as in the

observed period, a prediction of its movement can be made over the geometric mean, starting from the last element (Y_N) in the series:

$$\bar{Y}_{N+t} = Y_N \cdot G^t \tag{6}$$

$\bar{Y}_{N+t}$ – prediction value of the phenomenon assuming an unchanged G in the ($N+t$) period,

Y_N – the last value of occurrence in a series,

G – calculated or assumed geometric mean of the vertical indices,

t – the number of time periods after the last of the series for which the prediction is made.

Table 1 covers the period of 7 years, $N = 7$. If we are interested in the number of losses in 2030, i.e. the prediction for 9 years in advance, then we can note:

$$\bar{Y}_{N+t} = \bar{Y}_{7+9} = Y_7 \cdot G^9 = 54 \cdot 0.8594^9 = 13.8 \tag{7}$$

The southern China region remains the focal point of the biggest losses in the past decade, driven by factors including high levels of local and international trade, congested ports, older fleets, and extreme weather. We also analyze the number of losses by the type of vessel or cause in the period from 2011 to 2020, table 2.

As can be seen from the data, as well as from the graphical representation, the total shipping losses by vessel type do not show a steady monotonous increase or decrease in values. So, in this case, we cannot use the average rate to describe or predict the annual change in shipping losses. In such cases, the use of trend is a good basic option for such analysis. For this paper, we'll focus on the Fishery vessels.

Table 2 Total shipping losses by vessel type [9]

Vessel type/number of losses	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Σ
Cargo	37	61	40	31	40	34	54	17	16	18	348
Fishery	14	12	13	15	16	10	8	12	10	10	120
Bulk	14	11	15	5	13	5	7	2	2	2	76
Passenger	7	7	8	11	6	11	5	6	3	5	69
Tug	2	7	7	7	6	7	4	4	4	3	51
Chemical	4	8	10	2	3	7	4	4	4	3	40
RoRo	3	6	2	5	6	9	-	1	4	1	37
Container	3	7	4	4	5	5	3	2	1	1	35
Tanker	4	1	-	1	-	-	2	3	-	1	12
LPG	1	1	-	-	2	1	-	-	1	.	5

Source: Authors

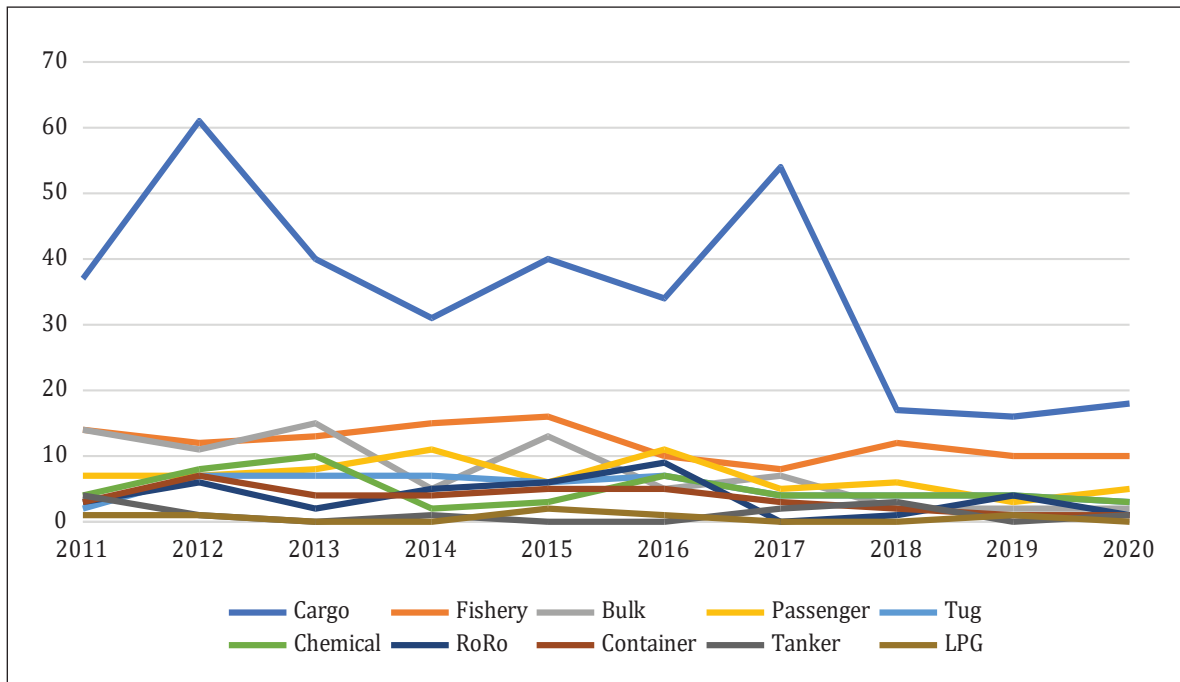**Figure 3** Total shipping losses by vessel type

A trend is a pattern found in time series datasets. It is typically described by linear regression or polynomial regression. It is used to describe if the **data is showing an up**ward or downward movement for the part, or all of, the time series [8], and a trendline is a line drawn on a chart highlighting an underlying pattern of individual values. The line itself can take on many forms depending on the shape of the data: straight, curved, etc. This is common practice when using statistical techniques to understand and forecast data [10, 11].
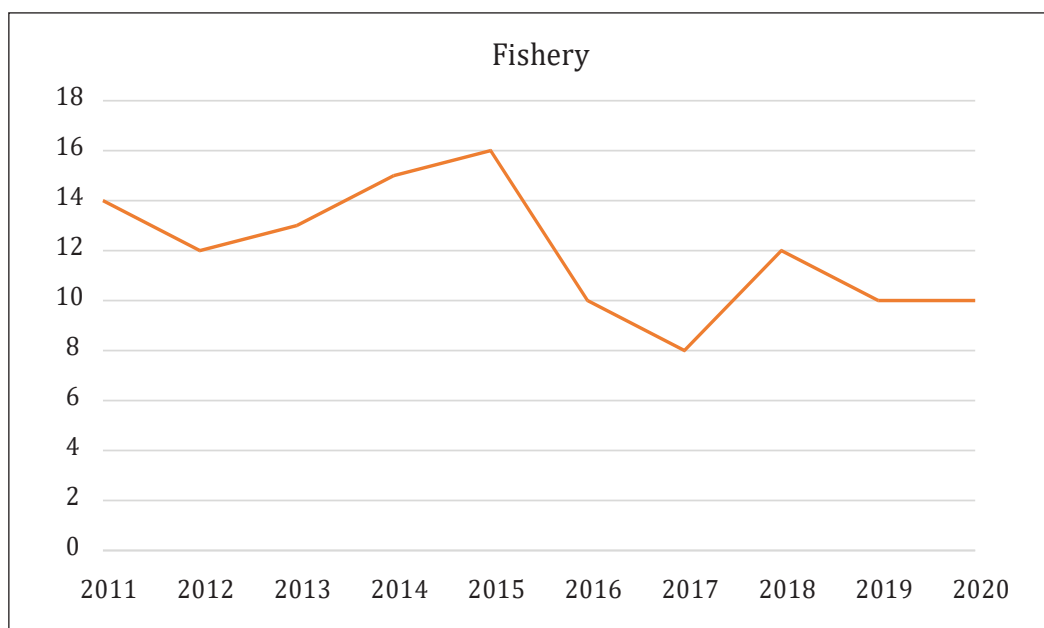

Figure 4 Total shipping losses of the Fishery vessels

If we take a closer look at the graph showing the total losses of the Fishery vessels, we can see that the graph resembles to some kind of polynomial function. So, our trend function, or polynomial regression function, will be:

$$f(x) = \hat{y} = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \qquad (8)$$

We can see from the graph above that there are five "extreme values", so it would be appropriate to choose a polynomial of at least 5th order.

$$f(x) = \hat{y} = a_5 x^5 + a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0 \qquad (9)$$

With the adequate choice of the origin of the coordinate system. It is common to place the origin (that is x = 0) in the beginning of the time interval, so our dataset now is

Table 3 Number of losses for fishery vessels, Y, of the coordinate system, X

X	0	1	2	3	4	5	6	7	8	9
Y	14	12	13	15	16	10	8	12	10	10

Writing down the equations

$$f(x_i) = y_i = a_5 x^5{}_i + a_4 x^4{}_i + a_3 x^3{}_i + a_2 x^2{}_i +$$
$$+ a_1 x_i + a_0, \quad i = 1, \dots 10 \qquad (10)$$

we get a system of linear equations.

The coefficients of the polynomial are calculated via the Least Square Method. Let us denote by $\hat{y}$ the estimated value of the number of loses calculated via the chosen polynomial function, and by y the actual measured number of loses.

We want to choose the coefficients of the polynomial function to be such that the total sum of the squared differences between the actual and the estimated value of losses in the observation points is minimal, that is

$$\min \sum_{i=1}^{10} (\hat{y_i} - y_i)^2 = \min \sum_{i=1}^{10} (a_5 x^5{}_i + a_4 x^4{}_i +$$
$$+ a_3 x^3{}_i + a_2 x^2{}_i + a_1 x_i + a_0 - y_i)^2 \qquad (11)$$

So, to minimize the function

$$\sum_{i=1}^{10} (a_5 x^5{}_i + a_4 x^4{}_i + a_3 x^3{}_i + a_2 x^2{}_i + a_1 x_i +$$
$$+ a_0 - y_i)^2 = g(a_1, a_2, a_3, a_4, a_5) \qquad (12)$$

we need to calculate

$$\frac{\partial g(a_1, a_2, a_3, a_4, a_5)}{\partial a_i} = 0, \quad i = 1, \dots, 5, \qquad (13)$$

which gives us a system of linear equations from which we obtain the coefficients a_1, a_2, a_3, a_4, a_5. Applying all this to our dataset, we get the following polynomial equation

$$f(x) = \hat{y} = -0.0141x^5 + 0.3966x^4 -$$
$$- 4.0583x^3 + 18.232x^2 - 34.451x + 34 \qquad (14)$$

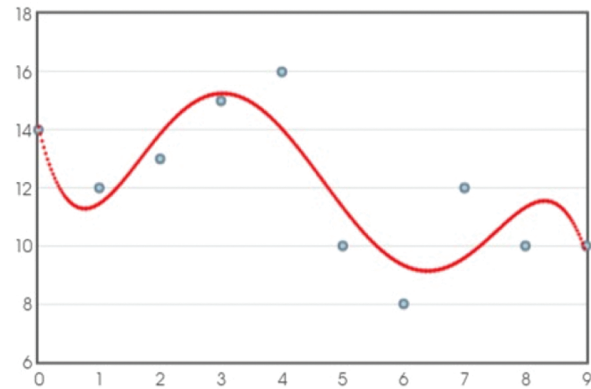and the fit of the trendline to the data is shown in Figure 5 below.

Figure 5 The fit of the trendline to the Fishery vessel data (n=5)

Source: Authors

The calculated trend polynomial can now easily be used to predict the possible shipping losses in the upcoming years for the Fishery vessels. For example, if we want to calculate the expected number of losses in 2021, then, since the year 2011 has the value of x=0, the value of x=10. So, we calculate

$$f(10) = \hat{y} = -0.0141 \cdot 10^5 + 0.3966 \cdot 10^4 -$$
$$- 4.0583 \cdot 10^3 + 18.232 \cdot 10^2 - 34.451 \cdot 10 + 34 \qquad (15)$$

$$f(10) = 10.39$$

So, the expected number of losses is about 10 to 11 vessels.

The natural question that arises is: how convenient is such an estimation? As we can see in Figure 5, some original values (blue points) are quite distant from the trendline. But are they too distant so that we can say that this simple model is a good representative of the original data? The answer to this question is given by calculating the R-squared:

$$R^2 = \frac{SRR}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}}, \qquad (16)$$

where

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad (17)$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y})^2 \qquad (18)$$

$$SS_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad (19)$$

The closer the value of R-squared is to 1, the better the model is. So, the calculation of R-squared gives

$$R^2 = 0.7216. \tag{20}$$

So maybe we can find a better fit. If we increase the degree of the polynomial to 8, that is

$$f(x) = \hat{y} = a_8 x^8 + a_7 x^7 + a_6 x^6 + a_5 x^5 + a_4 x^4 + \\ + a_3 x^3 + a_2 x^2 + a_1 x + a_0 \tag{21}$$

Again, calculating the values of the coefficients via Least Square Method, the polynomial trend is as follows

$$f(x) = \hat{y} = 0.0011 x^8 - 0.0367 x^7 + 0.4807 x^6 - \\ - 3.2167 x^5 + 11.7424 x^4 - 23.5277 x^3 + \\ + 25.2973 x^2 - 12.7013 x + 13.9961 \tag{22}$$

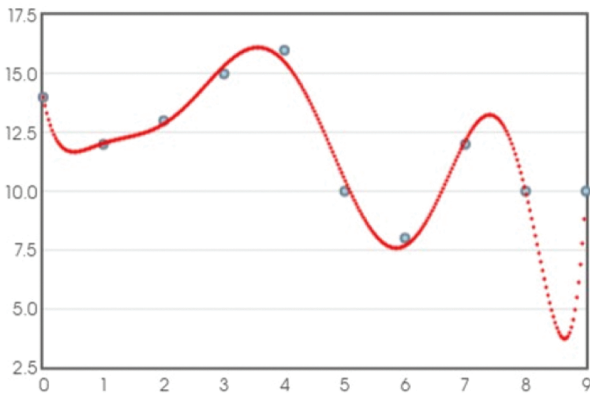The fit of the trendline is now shown in Figure 6.

Figure 6 The fit of the trendline to the Fishery vessel data (n=8)

Source: Authors

Now it is clear that the fit on the trendline is much better than the previous one. So, let's check the R-squared for the polynomial of the 8th degree. The calculation now gives

$$R^2 = 0.9872,$$

which is very close to 1, so our easy polynomial estimation is good.

4 Results and Discussion

Cargo ships accounted for more than a third of all vessels lost in 2020. According to the data presented this amount is about 44%. The most common cause of loss was sinking and the majority of cargo ships were lost in southern East Asian waters. From January 1, 2020 to December 31, 2020, contributing factors included bad weather, poor visibility leading to flooding and water intrusion, sinking,

and machine failure [7]. Damage or machine failure is one of the causes of maritime incidents. Failure due to wear and tear is a common factor in the failure of mechanisms and devices, especially in the conditions of the marine environment and vessel operation. For example, excessive strain on mooring lines contributes to failures of this equipment. There are still container ship fires and incidents involving RoRo vessels. Some safety studies [3] highlighted the routine and monotony of daily tasks, resulting in a reduced concentration of crew members, which contributes to ship collisions. It is difficult to obtain data on the causes of machine failures, but predictive maintenance [9] and probabilistic assessments could reduce the frequency of, for example, human errors, which lead to damage or failure. Graphical presentation of the average number of losses provides a clearer and more transparent picture of the movement of the observed phenomenon in a certain period of time.

It is shown that using some basic statistical methods we can obtain good mathematical models that can be used for forecasting future vessel losses. Excel and statistical data analysis were used in the work.

5 Conclusion

In specific statistical research, it is preferable to analyze longer periods of time for which more relevant conclusions can be drawn about the characteristics of the observed phenomenon's values. The quality of available data may vary due to incompleteness of details, subjectivity and the experience of the person collecting the data.

Probabilistic assessment of technical risks, as well as safety, includes procedures for identifying and analyzing hazards, determining the consequences of random, unwanted events, and determining the probability of consequences caused by such events. Risk management includes the process of identification, assessment, selection and application of methods aimed at reducing the risks to which human health and the environment are exposed. It is important to check what undesirable things can happen, what the probability is, and what the consequences are.

For further research, it is important to develop the possibilities of predictive methods, as well as simulation methods in the maritime field, to improve safety. Probabilistic assessments are analytical, extensive, demanding and assess the risk of unlikely events, but with potentially large consequences.

Author Contributions: data collection, research, concept, statistical analysis, editing, writing, review, checking JB.; statistical analysis, checking, editing, writing and review the article, BDB.

References

[1] *Available from* https://www.agcs.allianz.com/news-and-insights/reports/shipping-safety.html.

[2] *Available from* https://www.agcs.allianz.com/services/risk-consulting/marine-risk-consulting.html.

[3] *Accident Investigation Publications,* available from https://www.emsa.europa.eu/accident-investigation-publications/annual-overview.html.

[4] Bonato, J., Draščić Ban, B., Kraš, A. (2020). Statistical Methods in Theory of Technical Systems // Maritime proceedings, Spec. edition 3; 373.

[5] Dhillon, B. S. (2005). "Reliability, Quality, and Safety for Engineers", CRC PRESS, Florida.

[6] Ebeling, E. C. (2009). "An Introduction to Reliability and Maintainability Engineering", McGraw-Hill.

[7] Pivac, S., Šego, B. (2014). *Statistika*, Alka script d.o.o., Zagreb

[8] *Polynomial Regression Calculator*, available from https://stats.blue/Stats_Suite/polynomial_regression_calculator.html.

[9] Safety and Shipping Review (2021). *Available from* https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/AGCS-Safety-Shipping-Review-2021.pdf.

[10] Thoughts on trendlines. *Available from* https://www.storytellingwithdata.com/blog/2020/10/20/thoughts-on-trendlines.

[11] Zenzerović, Z. (2004). "Statistical handbook", Pomorski fakultet u Rijeci.