The processing of data for statistical tests of their basic parameters

ŠTEFAN BEREŽNÝ

ABSTRACT. This article focuses on the application of established data on flight delays at the airport Košice and its adaptation for further processing. These data were recorded for 36 months of 2006–2008. Modification of data is necessary for their further analysis in terms of time series analysis. Statistical tests of the basic parameters of the values of the flight delays at the airport Košice are made with their analysis.

1. THE DESCRIPTION OF PROCESSED DATA

The data, which we have obtained the Košice airport, is prepared in this article. These data cover the period of 36 months from January 2006 to December 2008. We have given large data sets, which have undergone initial processing of data. This initial processing of basic data and a formal modification of data to be able to continue working with them, is partly published in the works [1, 2, 3, 5]. In this work the adjustment process of the population was described to allow the stored data to obtain initial indicators delays at the airport Košice. Also from these data, their basic parameters of flights at Košice airport were published. This initial analysis also forms the basis of our research in this area. Assuming no fundamental processed and adjusted data and also from already published the basic parameters of these large data sets.

Based on this information, we had to decide which data will be analysed further. Thus, the selected data will be subjected to further processing treatment and subsequent analysis. This article will restrict itself only to the data selection, processing, presentation and analysis using basic statistical tests of hypotheses. This analysis tells us more about these data and because of their treatment will be possible to do a partial comparison between the years 2006 to 2008.

We build on the work [5] in which we have described the characteristics and parameters of the recorded data. There is described the whole process of modification of data and how to further processing. As regards the extensive information in our article we appointed only to those that we immediately need for our analysis.

Similarly, we will get the basic information from the articles [1, 2, 3], which are already published with the results of processing, which we refer in work [5]. These data are served as input for our decision on the selection of data for further processing and analysis. Also we can use them to compare with our obtained results.

In conclusion of this article we will attempt to indicate the direction of further work in this field. Space for research is relatively large because it is possible to examine whether the economic crisis has an impact on the development of monitored parameters and how it affected the predicted tendencies to the real trends of monitored parameters. This is

not only in preparation, since relevant data for 2009 are only to be prepared for further processing.

2. Used Methods

For data processing we use software: Matlab 2009b, MS Excel 2003 and Excel 2007, also qtOctave. Matlab and qtOctave are used mainly for analysis and testing of processed data and MS Excel, especially when the actual processing and selection of appropriate data for testing and analysis. Outputs, which are published in this article, are consistent with the format of these software packages.

Firstly, we try to describe the basic methods that we use in data processing and their next selection. For this type of calculations we preferred software package MS Office - Excel.

Processing and data selection

As the population we took the edited file, which is mentioned in [1, 2, 3, 5]. These data have already undergone by primary treatment and do not include dumb data or data that showed errors. During processing and selection of appropriate data we used filters that are available in MS Excel. Based on these initial filters, we found anomalies recorded data and we were trying to identify them. If it was able to be repaired immediately, it was corrected, if not, so we marked them and we later adjusted them to count with them.

Similarly, we looked at the data, which we kept in the various forms of contingency tables. These tables reveal to us about the structure of records and also suggested to suitable candidates for further processing and analysis. We also used following tables as a specific filter, respectively we used it for detecting the frequency of disagreement in each class and we looked for the causes of these disagreements and we tried to correct them. If found information could not be adjusted so that we can continue with it, so we have missed it in our group. The count of such cases was very few, given the range of records.

Data testing and analysis

In this section we take as a basic set of already modified data set using the methods described above. These sets are already in scope considerably smaller and more specified for further analysis and testing.

For analysis and testing, we downloaded data from MS Excel workbook into Matlab (or qtOctave) and there we used the statistical functions of Matlab toolbox for testing and verification of data loaded parameters. The results were then exported into text files, which are then processed in a table published in this article.

3. Basic Characteristics of Processed Data

We restrict our research only for the data of some airlines. These airlines are in the articles [1, 2, 3] that under the designation, which is clearly displayed in Table 1. In the first column of this table, we show the labels that we use in our article.

For better imagination of the actual structure of the researched sets we created Table 2 and Table 3. These tables show the percentage of companies selected for the overall operation of the Košice airport and also a mutual interest in the arrivals and departures during the reporting period 2006–2008. We assume that in 2006 was carried out 7 166 arrivals and departures.

In 2007 it was made 9 329 arrivals and departures and in 2008 was carried out 10 340 arrivals and departures at the international airport Košice. The first column in each year represents the relative percentages (together give 100%) and the second column in each

TABLE 1. Equivalent marking of the same airlines.

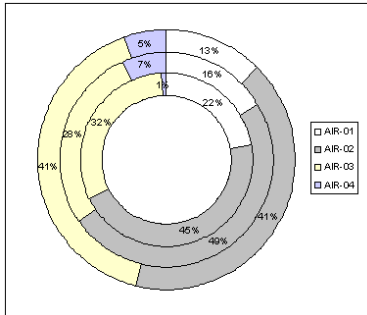|	YEAR		
	2006	2007	2008
AIR-01	AIR-02	AIR-02	AIR-02
AIR-02	AIR-05	AIR-05	AIR-06
AIR-03	AIR-06	AIR-06	AIR-08
AIR-04	AIR-12	AIR-11	AIR-15

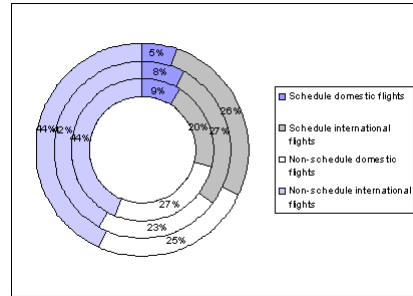FIGURE 1. Share of air companies on traffic at Koice airport.

FIGURE 2. Traffic at Koice airport with respect to the flight type.

TABLE 2. The ratio of airlines AIR-01 – AIR-04 on selected and total air traffic at the airport Košice.

	YEAR					
	2006		2007		2008	
AIR-01	22.26%	14.88%	16.37%	12.23%	12.62%	9.44%
AIR-02	45.21%	30.21%	48.81%	36.48%	41.14%	30.76%
AIR-03	31.74%	21.21%	28.08%	20.99%	41.02%	30.68%
AIR-04	0.79%	0.53%	6.74%	5.04%	5.21%	3.90%
		66.83%		74.73%		74.78%

TABLE 3. The ratio of types of flights on total air traffic at the airport Košice.

	2006	2007	2008
	Ratio in %	Ratio in %	Ratio in %
Schedule domestic flights	8.97%	7.84%	5.37%
Schedule international flights	20.00%	27.47%	26.47%
Non-schedule domestic flights	27.46%	22.79%	24.80%
Non-schedule international flights	43.57%	41.90%	43.37%

year is a percentage, given the above numbers. Graphic representation of data is on Figures 1 and 2.

TABLE 4. Basic characteristics of time deviations of arrivals at Košice Airport.

| | YEAR | | | | | |
| | 2006 | | 2007 | | 2008 | |
	Expected Value	Variance	Expected Value	Variance	Expected Value	Variance
AIR-01	2.75	7656.13	3.95	843.69	0.49	1892.62
AIR-02	12.92	6521.1	3.77	333.23	5.07	1133.18
AIR-03	-5.36	28557.22	10.51	1880	-1.96	1140.61
AIR-04	71.16	74034.87	16.17	1678.28	27.06	2784.36
Schedule flights	3.45	13753.95	5.84	891.48	2.27	891.32
Non-schedule flights	35.99	18858.89	13.21	1811.21	15.81	1232.23
Košice Airport	15.34	15864.97	8.1	1185.18	6.36	1032.84

TABLE 5. Basic characteristics of time deviations of departures at Košice Airport.

| | YEAR | | | | | |
| | 2006 | | 2007 | | 2008 | |
	Expected Value	Variance	Expected Value	Variance	Expected Value	Variance
AIR-01	5.47	159.28	4.47	162.35	3.81	334.87
AIR-02	16.96	4252.16	2.95	247.21	1.3	811.16
AIR-03	17.58	1651.09	16.06	1397.24	9.38	1777.21
AIR-04	26.05	984.47	17.57	2056.78	38.09	3938.9
Schedule flights	14.21	2491.02	7.19	614.78	4.88	1179.89
Non-schedule flights	26.59	10524.71	15.35	1851.24	18.77	1666.23
Košice Airport	18.71	5445.18	9.69	1006.94	9.07	1367.16

Basic characteristics are related to deviations from the planned arrival times, respectively planned departure time. Later departure, respectively arrival time (called delay) is represented by positive values and the earlier arrival respectively departure is represented by negative values. The ideal situation is zero mean with minimal variance. Actual values in minutes are shown in Table 4 and Table 5. Table 4 describes the situation for arrivals and Table 5 departures for the Košice airport.

In the tables we can observe the temporal evolution of the expected value (mean) of time deviations for arrivals and departures at the Košice airport in 2006 and 2008. These values preceded of the global economic crisis. Interesting are the values calculated for schedule and non-schedule flights and their comparison with the total value of the Košice airport.

In Table 4 there are only 2 negative values in the third row in 2006 and 2008. In Table 5, which shows the values for the departure is not even a negative value, which corresponds to the fact that the aircraft should leave no earlier than planned to have the flight plan. It is also interesting evolution of the average temporal variations in different years, along with the evolution of variances.

More interested are the calculated time deviations of flights in case of the selection to national and international. The results are shown in Table 6 and Table 7.

TABLE 6. Basic characteristics of time deviations of arrivals at Košice Airport.

| | YEAR | | | | | |
| | 2006 | | 2007 | | 2008 | |
	Expected Value	Variance	Expected Value	Variance	Expected Value	Variance
Schedule domestic flights	-5.09	31340.53	7.85	1407.12	-4.29	804.99
Schedule international flights	7.33	5710.85	4.52	549.66	6.28	901.69
Non-schedule domestic flights	86.23	51060.28	10.75	1618.4	10.12	645.08
Non-schedule international flights	19.49	7175.79	14.08	1876.69	17	1346.74

TABLE 7. Basic characteristics of time deviations of departures at Košice Airport.

| | YEAR | | | | | |
| | 2006 | | 2007 | | 2008 | |
	Expected Value	Variance	Expected Value	Variance	Expected Value	Variance
Schedule domestic flights	25.38	6926.48	10.38	1046.62	3.29	1352.17
Schedule international flights	9.03	350.91	5.09	320.01	5.84	1072.21
Non-schedule domestic flights	70.82	34231.75	11.25	1067.83	16.72	1375.21
Non-schedule international flights	12.23	1981.77	16.72	2104.65	19.23	1730.27

TABLE 8. Two-Sample t-test of time deviation of sets of arrivals and departures.

| | YEAR | | |
	2006	2007	2008
2006	-	No	No
2007	No	-	No
2008	No	No	-

From the above results we can see quite a big difference between schedule and non-schedule flights at Košice airport. Similarly, we can observe the differences between arrivals and departures. The question is how to interpret these differences and where to find the causes of these variations.

4. TESTING OF PROCESSED DATA

In this article, we appoint to find out how these sets behave in the monitored period. For the beginning, we need to test the sets, if they fulfil the conditions of these tests. For used sets then we describe the statistical tests of hypotheses that will be used and their results will be summarized in convenient tables.

Among the various possible tests we carried out a series of tests for Two-Sample t-test. All these tests of statistical hypotheses are made on a significance level of 5 % (i.e. $= 0.05$). The calculations were made in program Matlab 2009b. All results are processed in the tables. In Table 8 we show the results of the Two-Sample t-tests of time of all flights (arrivals and departures) in each year.

In Table 9 we show the results of the Two-Sample t-tests of time of arrivals in each year.

Comparison between 2008 and 2007 became negative, but the significance level of 1% there would have been matched. In Table 10 we show the results of the Two-Sample t-tests of time of departures in each year.

In this case, it would be appropriate to carry out more tests in addition to schedule and non-schedule flights. It would be advisable to test the selected four airlines AIR-01–AIR-04. Given the number of values in Table 6 and Table 7 it can be estimated, where there is a

TABLE 9. Two-Sample t-test of time deviation of set of arrivals.

	YEAR		
	2006	2007	2008
2006	-	No	No
2007	No	-	No
2008	No	No	-

TABLE 10. Two-Sample t-test of time deviation of set of departures.

	YEAR		
	2006	2007	2008
2006	-	No	No
2007	No	-	Yes
2008	No	Yes	-

consensus and where consensus does not occur. Despite so narrow results it can be noted that between 2007 and 2008 the aircraft depart from Košice International Airport approximately with the same delay. Similarly, we could say that in these years, flights landed with the same delay. Other cases have varied significantly, which is essential information for studying the causes of these disagreements.

5. CONCLUSION

From the above results we can get an idea how the time variations were developed in the Košice airport in 2006 and 2008. Those tests confirmed that in most cases, compliance has not been confirmed in average delays. It is true that we have not done detailed analysis, but the general tests set up at least area for further progress in examining of this issue. It would be appropriate to examine separately the behaviour of the temporal variations for individual airlines. Subsequently, it would be appropriate to analyse the results of these published results.

These results can also be used in the analysis, as described in [6]. Similarly, we could look at these tests for schedule and non-schedule flights, which are designated as national and international. For such a finer partitioning the results can be more specific.

This analysis is essential for processing and testing time deviations from the time series, which may be represented by the months, weeks or days of the week. Such data processing and its results are ideal for time analysis, which is described in [4]. Time series analysis should be the target result of all these sub-studies, which are described in this article as well as in the above articles. The results of the analysis of time series are an important source of information for management of Košice International Airport. This model analysis is of course also applicable to other airports and it would be appropriate to compare these results among themselves.

Even more interesting thing is the analysis of data of the crisis extended year 2009. It will be interesting to follow the development time variations in 2009, compared with their non-crisis years 2006 to 2008 and on the basis of the results try to describe the reasons for any differences. On the other hand, the development of these time variations should theoretically not be affected by the crisis, but as weather, or unforeseen circumstances. On the other hand, however, we could see significant differences in the calculated values in

non-crisis years 2006 to 2008, suggesting that this endpoint is non-standard behaviour and it must be dealt with research on this issue even more research in this area could clarify hidden patterns of temporal anomalies, arrivals and departures at the airport, which could be effectively used in planning and managing the airport.

Further direction of research in this area, then we would split into two lines. One line addresses to expanding the population of other statistics of coming years, or expand the set of data from other airports and the second line should analyse these data in more detail and depth.

REFERENCES

[1] Berežný, Št. and Andrejkovič, M., *Statistical processing of flights on Košice airport*, Acta Avionica **11** (2009), No. 17, 9–14

[2] Berežný, Št., *Statistical processing of arrivals and departures on Košice airport in 2007*, Acta Avionica **12** (2010), No. 19, 42–48

[3] Berežný, Št., *Statistical processing of arrivals and departures on Košice airport in 2006*, Acta Avionica **12** (2010), No. 19, 35–41

[4] Kimáková, Z. and Andrejiová, M., *Analýza časového radu emisií sox pomocou programu r*, Forum statisticum slovacum. Nitra, (2009), 6/2009, 63–68

[5] Laczkó, R., *Štatistické spracovanie meškaní letov na Košickom letisku. Bakalárska práca*, Košice, Technická univerzita v Košiciach, Letecká fakulta, (2009), 48 pp.

[6] Tkáč, M., Andrejkovič, M. and Hajduová, Z., *Analysis of Sky Europe airlines flights on Koice international airport*, Acta Avionica, Košice, Technical University Košice, Faculty of aeronautics, **11** (2009), No. 17, 174–179

DEPARTMENT OF MATHEMATICS AND THEORETICAL INFORMATICS
TECHNICAL UNIVERSITY IN KOŠICE,
FACULTY OF ELECTRICAL ENGINEERING AND INFORMATICS
NĚMCOVEJ 32, 040 02 KOŠICE, SLOVAK REPUBLIC
E-mail address: Stefan.Berezny@tuke.sk