Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (1) 5–17
http://journals.rudn.ru/miph

Research article

UDC 519.23
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-1-5-17 EDN: BAUTRQ

Statistical methods for estimating quartiles of scientific conferences

Anna M. Ermolayeva

RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

(received: January 27, 2024; revised: February 16, 2024; accepted: March 1, 2024)

Abstract. The article presents the results of the evaluation of quartiles of scientific conferences presented by leading rating agencies. The estimates are based on the use of three methods of multivariate statistical analysis: linear regression, discriminant analysis and neural networks. A training sample was used for evaluation, including the following factors: age and frequency of the conference, number of participants and number of reports, publication activity of the conference organizers, citation of reports. As a result of the study, the linear regression model confirmed the correctness of the quartiles exposed for 77% of conferences, while the methods of neural networks and discriminant analysis gave similar results, confirming the correctness of the quartiles exposed for 81 and 85% of conferences, respectively.

Key words and phrases: evaluation of quartiles of scientific conferences, discriminant analysis, neural networks, linear regression

1. Introduction

As it is known [1], quartile (quarter) is a category of scientific publications, which is determined by bibliometric indicators reflecting, first of all, the level of citation, that is, the relevance of the publication by the scientific community. And if the procedure for assigning quartiles to scientific journals has long been developed and successfully applied in practice [2–5]. In addition, many metrics have been introduced to assess the impact of journals, such as impact factor, 5-year impact factor, immediacy index, and impact factor without self cites, median impact factor, aggregate impact factor and others [6]. At the same time, this issue remains the subject of research for scientific conferences [7–11]. Some rating agencies have already begun to rank scientific conferences without disclosing the details of this procedure. For example, there is a CORE conference ranking [12], a CCF conference ranking [13], and a Microsoft Academic conference ranking (has been deleted) [14]. The disadvantages of the first two ratings are that they are expert, regional and do not fully disclose the procedure for ranking conferences. They also rank only computer science conferences.

Researchers use various methods to compile new conference rankings, such as correlation analysis [7, 15], statistical analysis [15, 16], calculation of indicators similar to journal ones [9], graph and tree analysis [8, 17], regression analysis [11, 16]. Many of these studies involved the use of several of the listed methods. There were also works devoted to the search for methods for predicting the rating of a conference or predicting the impact of works presented at a particular conference [18]. Machine learning was used for these purposes [19, 20]. Therefore, this study is devoted to comparing two popular methods for predicting conference rankings, and I also included in the study such a statistical method as discriminant analysis, which is essentially a mathematical prerequisite for machine learning.

We managed to find data on some conferences via the Internet, including their quartiles and a number of other indicators, which will be discussed below. As a result, we received a training sample from 23 conferences, on the basis of which we will try to assess the adequacy of the quartiles exposed using three methods of multidimensional statistical analysis: linear regression, discriminant analysis and neural networks.

© Ermolayeva A. M., 2024

This work is licensed under a Creative Commons Attribution 4.0 International License
https://creativecommons.org/licenses/by-nc/4.0/legalcode

2. Training sample

Let's introduce the notation:

- Y is a random variable (r.v.), taking the values 1, 2, 3 or 4 is the quartile of a scientific conference;
- X_1 is a non–negative r.v., which takes values from a set of real numbers is the average citation of conference materials (the number of citations per report over the 10 years from 2011 to 2020);
- X_2 is an integer positive r.v. is the number of conference participants;
- X_3 is an integer positive r.v. is the number of reports at the conference;
- X_4 is an integer positive r.v. is the number of participants who submitted more than one report;
- X_5 is an integer value that takes two values: 0 or 1 is an indicator of the publication activity of the conference organizers (1 — if the organizers submitted a report to the conference and 0 — otherwise);
- X_6 is a non–negative r.v., which takes values from a set of real numbers is an indicator of the publication activity of the conference organizers, equal to the average citation of scientific publications per conference organizer.

The table 1 shows a training sample of the values of r.v. Y, X_1–X_6, compiled from the materials of the websites [21–23].

3. Linear regression model

Based on the data presented in the table 1, we will build a linear regression model reflecting the dependence of Y on the factors listed above. We will carry out the construction using the SPSS statistical package.

At the beginning, we will estimate the degree of linear dependence of Y on factors from X_1 to X_6 by constructing a Pearson pair correlation matrix. The study showed that a significant relationship is observed between Y and factors X_1, X_2, X_3 (table 2). The X_4–X_6 factors have little effect on the Y values, so we will not take them into account in the future. At the same time, a strong relationship is observed between factors X_2 and X_3. To avoid the negative impact of multicollinearity, we excluded factor X_3 from consideration and construct a two-factor regression model $Y(X_1, X_2)$.

As a result, the equation is obtained (see the table 3):

$$Y = -0.049 \cdot X_1 + 0.012 \cdot X_2 + 2.237. \tag{1}$$

Note that table 3 shows not only the absolute values of the coefficients of the model, but also the results of checking their significance using the T-criterion. According to the data from the last column of the table, all coefficients are significant with a significance level not exceeding 10^{-3}. In the second column of the table, estimates of the standard deviation σ_j of the coefficients of the model are calculated, as well as their values after standardization. According to these data, a change in the j-th coefficient of the model by one σ_j entails a change in Y by approximately $0.62\sigma_j$ downward for the coefficient at X_1 and by $0.517\sigma_j$ upward for the coefficient at X_2.

According to the data presented in the table 4, the constructed model reflects by 85.1% the real dependence of the quartile on the citation of materials and the number of conference participants. At the same time, 72.4% of the Y variation in our model is due to the variability of factors X_1 and X_2. The model itself is significant at a significance level not exceeding 10^{-3} (see the table 5).

There i s no autocorrelation of residues in the constructed model, because the Durbin–Watson statistics, equal to 1.722 (see the table 4), falls into the interval $(d_u; 4 - d_u)$, where $d_u = 1.33$ (according to the table of critical values for the significance level $\alpha = 0.05$).

The absence of auto-correlation of the residuals in combination with the condition of independence of the observational results actually means that the conditions of the Gauss–Markov theorem are fulfilled, on the basis of which it is true.

Statement 1. Model (1) is a model with minimal variance among all linear models of a fixed level of significance α.

Let's determine the estimate of the variance of the errors of the model (1). To do this, first solve the question of the normality of the residuals. We will check the normality using the Frosini criterion [24].

Table 1

Training sample

Number	Y	X_1	X_2	X_3	X_4	X_5	X_6
1	1	55.20	55	146	2	1	36.17
2	1	37.85	58	126	8	0	32.50
3	1	25.62	79	153	3	1	9.33
4	2	18.93	74	139	9	0	16.21
5	2	16.38	48	132	0	1	7.17
6	2	14.39	95	143	6	1	12.21
7	2	7.06	31	48	5	0	8.83
8	2	7.03	30	87	0	1	15.83
9	2	6.87	59	105	8	0	17.32
10	2	6.46	33	94	9	1	16.58
11	2	6.04	30	78	0	0	4.83
12	2	5.95	31	66	1	0	23.21
13	3	5.34	26	33	0	1	8.50
14	2	3.69	25	48	0	1	10.51
15	3	3.42	17	34	0	1	10.67
16	3	3.39	13	26	2	0	18.67
17	4	3.20	95	255	9	1	22.37
18	2	3.07	52	100	5	1	25.87
19	4	2.48	110	301	5	0	14.83
20	4	2.42	157	345	0	1	20.17
21	4	2.04	99	282	8	0	16.71
22	4	1.89	135	380	7	1	25.60
23	4	1.76	169	382	2	1	10.50

Table 2

Pearson Pair Correlation Matrix

	Y	X_1	X_2	X_3
Y	1	−0.678	0.588	0.666
X_1	−0.678	1	−0.114	−0.141
X_2	0.588	−0.114	1	0.959
X_3	0.666	−0.141	0.959	1

Table 3

Coefficients

Model	Non-standard	Standard error	Standard	t	Signific.
Constant	2.237	0.246		9.082	0.000
X_1	−0.049	0.009	−0.62	−5.244	0.000
X_2	0.012	0.003	0.517	4.377	0.000

Table 4

Summary for the model

Model	R	R^2	Adjust. R^2	Standard estim. error	Durbin–Watson
1	0.851	0.724	0.697	0.572	1.722

Table 5

Analysis of variance

Model	Sum of squar	Degr. of freed.	Stand. deviat.	F	Signific.
Regression	17.196	2	8.598	26.283	0.000
Residual	6.543	20	0.327		
Total	23.739	22			

To do this, you need to calculate the statistics:

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \Phi(z_i) - \frac{i - 0.5}{n} \right|, \tag{2}$$

where $z_i = \dfrac{x_i - \overline{x}}{s}$; $\overline{x} = \dfrac{1}{n} \sum\limits_{i=1}^{n} x_i$; $s^2 = \dfrac{1}{n} \sum\limits_{i=1}^{n} (x_i - \overline{x})^2$; $\Phi(z_i)$ is the distribution function $N(0, 1)$.

The results of the calculations are presented in the table 6.

Fixing the significance level $\alpha = 0.01$ and considering that $C_{\text{cr}}(0.01) = 0.341$ [24], we obtain:

$$B_n = 0.306 < C_{\text{cr}}(0.01) = 0.341. \tag{3}$$

Therefore, the residuals are distributed normally.

And finally, considering that the variance estimate σ^2 is determined by the formula:

$$\sigma^2 = \frac{1}{n - (p + 1)} (Y - Y^*)^T (Y - Y^*) \tag{4}$$

and is equal in our case to 0.327 (see the last column of the table 6), we come to the following result.

Statement 2. The model of linear regression of quartiles of scientific conferences, based on the data presented in the table 1, has the form:

$$Y = -0.049 \cdot X_1 + 0.012 \cdot X_2 + 2.237 + \epsilon, \tag{5}$$

where ϵ is the r.v. having a normal distribution with parameters $m = 0$ and $\sigma = 0.57$.

Table 6

Calculation of B_n statistics

Num.	Y	X_1	X_2	Y^*	$Y - Y^*$	$F(Z_i)$	$B(i)$	$(Y - Y^*)^{**}2$
1	1	25.620	79.000	1.930	−0.930	0.040	0.018	0.864
2	2	3.070	52.000	2.711	−0.711	0.090	0.025	0.505
3	2	14.390	95.000	2.672	−0.672	0.102	0.007	0.451
4	2	6.870	59.000	2.608	−0.608	0.125	0.027	0.370
5	2	3.69	25	2.356	−0.356	0.250	0.054	0.127
6	2	5.95	31	2.317	−0.317	0.274	0.035	0.101
7	2	6.46	33	2.316	−0.316	0.274	0.009	0.100
8	2	6.04	30	2.301	−0.301	0.284	0.042	0.091
9	2	7.06	31	2.263	−0.263	0.309	0.061	0.069
10	2	7.03	30	2.253	−0.253	0.316	0.097	0.064
11	2	18.93	74	2.197	−0.197	0.359	0.097	0.039
12	4	1.76	169	4.179	−0.179	0.370	0.130	0.032
13	1	37.85	58	1.078	−0.078	0.440	0.103	0.006
14	2	16.38	48	2.010	−0.010	0.494	0.093	0.000
15	4	2.42	157	4.002	−0.002	0.500	0.130	0.000
16	4	1.89	135	3.764	0.236	0.670	0.004	0.056
17	4	2.48	110	3.435	0.565	0.857	0.140	0.319
18	4	2.04	99	3.325	0.675	0.898	0.137	0.456
19	3	5.34	26	2.287	0.713	0.910	0.106	0.508
20	3	3.42	17	2.273	0.727	0.915	0.067	0.528
21	3	3.39	13	2.227	0.773	0.928	0.037	0.598
22	4	3.20	95	3.220	0.780	0.929	0.006	0.608
23	1	55.20	55	0.192	0.808	0.936	0.043	0.653
						$Sum =$	1,467	6,543
						$B(n) =$	0,306	0,327
						$C_{\text{cr}}(0,01)=$	0,341	

Further, based on the data for the three “new” conferences, we obtained the predicted values of their quartiles using the model 1 (see the table 7).

As we can see from the results presented in the table 7, conferences numbered 24 and 25 should be assigned the 1st and 2nd quartiles, respectively. With conference number 26, the picture is not so clear, because the predicted value of Y lies approximately in the middle between numbers 3 and 4, which suggests that this conference should be assigned the 4th quartile with a probability of 0.55 or the 3rd quartile with a probability of 0.45.

Table 7

The results of the calculation of the predicted quartile values

Num.	Quartile (forecast. significant.) Y	Citation X_1	Quantity participants X_2
24	1.0323	38.30	56
25	2.23101	5.51	22
26	3.55425	2.75	121

4. Discriminant analysis

Discriminant analysis is a classification method, the purpose of which is to divide the objects of observation into classes according to the values of the effective feature, depending on a number of controlled factors [24]. In our case, the effective feature is the quartile, and the controlled factors are the citation of the conference materials and the number of its participants. Our further goal is to classify new conferences using discriminant analysis, the data for which are presented in the table 7, based on the training sample presented in the table 1. To solve this problem, we still use the SPSS statistical package.

First of all, we pay attention to the data shown in the table 8. This table shows the results of checking the significance of differences in the average values of discriminant functions in data groups corresponding to factors X_1 and X_2 using the Wilkes' Lambda criterion. In our case, the significance levels for each factor do not exceed 0.05, which proves the existence of discriminating features of these factors and confirms the possibility of their use for discriminant analysis.

Table 8

The criterion of equality of group averages

Function	Wilkes' Lambda	F	Degr. of freedom 1	Degr. of freedom 2	Sgn.
X_1	0.190	26.972	3	19	0.000
X_2	0.232	20.913	3	19	0.000

According to the data presented in the table 9, the first discriminant function takes into account 67.5% of the variance of the effective feature, and the correlation between the training sample data and the data obtained by the model is 0.918, which is a fairly high indicator. For the second discriminant function, these indicators are 32.5% and 0.849, respectively. The significance of discriminant functions was assessed using the Wilkes' Lambda criterion. According to the results presented in the table 10, the significance of both discriminant functions does not exceed 0.05.

Table 9

Eigenvalues

Function	Proper. value	% of var. explained	Cumulative %	Canonical correl.
X_1	5.378^a	67.5	67.5	0.918
X_2	2.586^a	32.5	100.0	0.849

According to the table 10, we obtain the following expressions for discriminant functions:

$$D_1(X_1, X_2) = 0.141 \cdot X_1 - 0.028 \cdot X_2 + 0.352, \tag{6}$$

$$D_2(X_1, X_2) = 0.083 \cdot X_1 + 0.034 \cdot X_2 - 3.100. \tag{7}$$

Table 10

Non-normalized coefficients of canonical discriminant functions

Function	1	2
X_1	0.141	0.083
X_2	−0.028	0.034
(Constant)	0.352	−3.100

Statement 3. Discriminant functions (6) and (7) are significant at the significance level $\alpha = 0.01$.

Proof. We will evaluate the significance of discriminant functions using the Wilkes' Lambda criterion [25], according to which it is necessary to calculate statistics:

$$\chi_k^2(m_k) = -(n - ((p+g)/2 - 1)\ln \Lambda_k, \quad k = 1, 2, \dots, \tag{8}$$

where $\Lambda_1 = \frac{1}{1+\lambda_1} \cdot \frac{1}{1+\lambda_2}$, $\Lambda_2 = \frac{1}{1+\lambda_2}$, $p = 2$ is the number of discriminant features; $g = 4$ is number of groups $m_1 = p + g$; $m_2 = p$ is number of degrees of freedom.

The calculation results are presented in the table 11.

Table 11

Lambda – Wilkes Statistics

Function	Λ_k	χ_k^2	m_k
1	0.044	59.470	6
2	0.279	24.265	2

It is known [25] that statistics $\chi_k^2(m_k)$ have a χ^2 distribution with m_k degrees of freedom. Fixing $\alpha = 0.01$ and considering that $(1 - \alpha)$ are quantiles χ^2 are distributions with degrees of freedom $m_1 = 6$ and $m_2 = 2$ are 16.8 and 9.21, respectively, we arrive at the following result:

1) since $59.470 > 16.8$, the hypothesis of the significance of the discriminant function (6) is accepted;

2) since $24.265 > 9.21$, the hypothesis of the significance of the discriminant function (7) is accepted.

Thus, the statement is proved. □

The results of the analysis are presented in the table 12. As a result, quartiles 1, 2 and 4 were assigned to the "new" conferences, respectively. At the same time, quartiles were predicted for conferences numbered 24 and 26 with probabilities of 1 and 0.996. For conference number 25, the picture was not so unambiguous. It was predicted the second quartile with a probability of 0.673, or the 3*rd* quartile with a probability of 0.327.

In addition, the quartiles of conferences from the training sample were recalculated. As a result, conferences with numbers 3, 13, 15 and 16 received new quartile values. The quartiles of the remaining conferences, amounting to 82.6%, were found to be correct.

5. Neural network

To solve the classification problem, a neural network called a multilayer perceptron is best suited [26]. Typically, a network consists of one input layer, one or more hidden layers, and one output layer. Each layer consists of several neurons. The neuron processes its inputs and generates one output value, which is transmitted to the neurons in the subsequent layer. Each neuron in the input layer represents the values of one predictor from the vector $x = (x_1, x_2)$. In our case, x_1 and x_2 are the citation and the number of participants in the scientific conference.

Table 12

Classification results

Num.	Actual group	1st most likely predicted group	Group probability	2nd most likely predicted group	Group probability
1	1	1	1.000	2	0.000
2	1	1	1.000	2	0.000
3	1	2**	0.524	1	0.473
4	2	2	0.983	3	0.011
5	2	2	0.951	3	0.048
6	2	2	0.828	4	0.166
7	2	2	0.785	3	0.215
8	2	2	0.777	3	0.223
9	2	2	0.927	3	0.069
10	2	2	0.791	3	0.209
11	2	2	0.760	3	0.240
12	2	2	0.767	3	0.233
13	3	2**	0.710	3	0.290
14	2	2	0.667	3	0.333
15	3	2**	0.572	3	0.428
16	3	2**	0.524	3	0.476
17	4	4	0.786	2	0.210
18	2	2	0.868	3	0.128
19	4	4	0.981	2	0.019
20	4	4	1.000	2	0.000
21	4	4	0.905	2	0.094
22	4	4	1.000	2	0.000
23	4	4	1.000	2	0.000
24	not grouped	1	1.000	2	0.000
25	not grouped	2	0.673	3	0.327
26	not grouped	4	0.996	2	0.004

To build the network, we use the “neural networks” section of the SPSS package, in which we specify the quartile of the conference as the dependent variable, and the citation and number of participants as the covariant, and set the data division into three subsets: training, control and verification in a ratio of 20 : 3 : 3. We set the network architecture manually, fixing the presence of one hidden layer with four neurons. We select the sigmoid as the activation function for the hidden and output layers. Then we select the interactive type of training using the gradient descent method and set the time and the rule for stopping the learning process. The network parameters are shown in the figure 1, and its configuration is shown in the figure 3.

In the report presented in the figure 2, we pay attention to the lines “sum of squares error” and “relative error” in the section “test sample”. The error values turned out to be 0.016 and 0.045. These values are quite small, which indicates that the neural network is well trained.

Input layer	Covariates	1	X1
		2	X2
		Number of neurons[a]	2
		Scaling method for covariates	Standardized
Hidden layers		Number of hidden layers	1
		The number of neurons in hidden layer1[a]	4
		Activation function	Sigmoid
Output layer	Dependent variables	1	Y
		Number of neurons	1
		Method of changing the scale for quantitative dependent variables	Normalized
		Activation function	Sigmoid
		Error function	Sum of the squares

Figure 1. Network Parameters

Training sample	Sum of squares error	0,115
	Relative error	0,100
	Stop rule used	Number of consecutive steps without reducing the error: 1
	Training time	0:00:00.002
Verification sample	Sum of squares error	0,016
	Relative error	0,045

Figure 2. Summary for the model

The predicted quartile values for both “new” conferences and conferences from the training sample are contained in the fourth column of the table 13. Note that for “new” conferences, the quartiles obtained using a neural network coincide with the quartiles obtained using discriminant analysis.

Table 13

Quartile values

Num.	The actual quartile value	The value of the quartile according to the regression method	The quartile value obtained by discriminant analysis	The quartile value predicted by the neural network
1	1	1	1	1
2	1	1	1	2*
3	1	2*	2**	2*
4	2	2	2	2
5	2	2	2	2
6	2	3*	2	2
7	2	2	2	2
8	2	2	2	2
9	2	2	2	2
10	2	2	2	2*
11	2	2	2	2*
12	2	2	2	2*
13	3	2*	2**	3
14	2	2	2	2
15	3	3	2**	3
16	3	3	2**	3
17	4	4	4	4
18	2	3*	2	2
19	4	3*	4	4
20	4	4	4	4
21	4	3*	4	4
22	4	4	4	4
23	4	4	4	4
24	1	1	1	1
25	2	2	2	2
26	4	4	4	4
Num.	discrep.	6	4	5
of %	matches	76.92	84.61	80.77

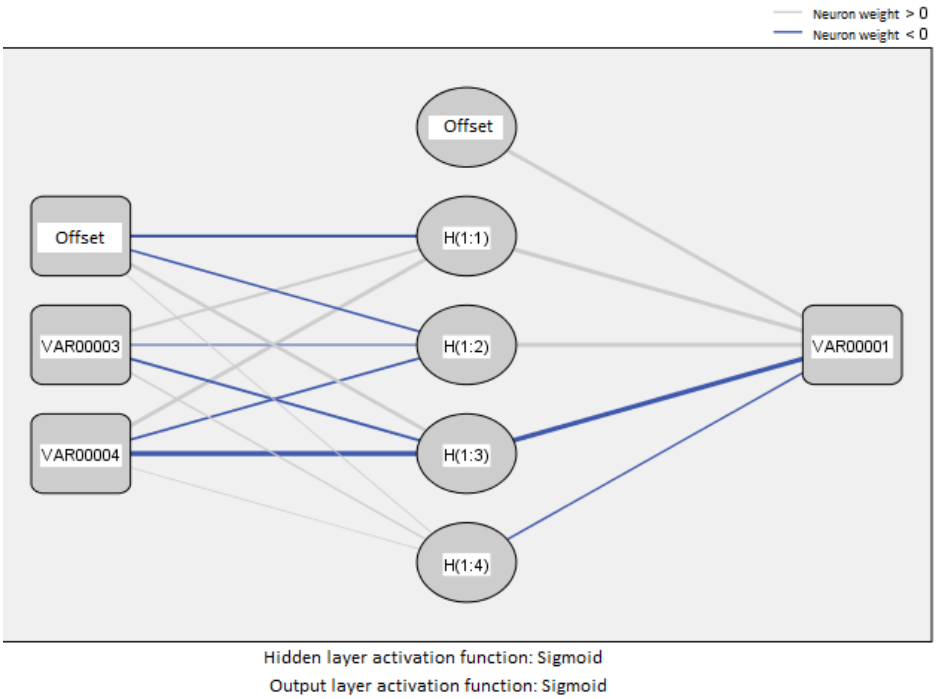

Figure 3. Neural Network configuration

6. Conclusion

As a result of the conducted research, we calculated quartiles of scientific conferences using three different methods. The results of the calculations are shown in the table 13. The quartiles marked with asterisks do not match those that were put up by rating agencies and which we called actual.

The last row of the table 13 shows the percentage of matches of the actual quantiles and quartiles calculated using the appropriate method. As we can see, the best indicator is for the discriminant analysis (4 discrepancies). In second place, with a difference of one conference, is the neural network. In third place is the linear regression method, which revealed 6 discrepancies.

Funding: The publication has been prepared with the support of the RUDN University Strategic Academic Leadership Program.

References

1. Prakash, B. *Quartiles of the journals and the secret of publishing* https://www.manuscriptedit.com/scholar-hangout/quartiles-of-the-journals-and-the-secret-of-publishing/.
2. Garfield, E. Citation indexes for science: A new dimension in documentation through association of ideas. *Science* **122,** 108–111 (1955).
3. Bergstrom, C. T., West, J. D. & Wiseman, M. A. The eigenfactor metrics. *Journal of neuroscience* **28,** 11433–11434 (2008).
4. Moed, H. F. Measuring contextual citation impact of scientific journals. *Journal of informetrics* **4,** 265–277. doi:10.1016/j.joi.2010.01.002 (2010).
5. González-Pereira, B., Guerrero-Bote, V. P. & Moya-Anegón, F. A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of informetrics* **4,** 379–391. doi:10.1016/j.joi.2010.03.002 (2010).
6. Kim, K. & Chung, Y. Overview of journal metrics. *Science Editing* **5,** 16–20 (2018).

7. Freyne, J., Coyle, L., Smyth, B. & Cunningham, P. Relative status of journal and conference publications in computer science. *Communications of the ACM* **53,** 124–132. doi:10.1145/1839676.1839701 (2010).
8. Jahja, I., Effendy, S. & Yap, R. H. Experiments on rating conferences with CORE and DBLP. *D-Lib Magazine* **20.** doi:10.1045/november14-jahja (2014).
9. Meho, L. I. Using Scopus's CiteScore for assessing the quality of computer science conferences. *Journal of Informetrics* **13,** 419–433. doi:10.1016/j.joi.2019.02.006 (2019).
10. Effendy, S. & Yap, R. H. C. *Investigations on rating computer sciences conferences: an experiment with the Microsoft Academic Graph Dataset* Apr. 2016. doi:10.1145/2872518.2890525.
11. Lee, D. H. Predictive power of conference-related factors on citation rates of conference papers. *Scientometrics* **118,** 281–304. doi:10.1007/s11192-018-2943-z (2019).
12. *Core conference ranking* http://portal.core.edu.au/conf-ranks/.
13. *CCF conference ranking* https://www.ccf.org.cn/en/.
14. *Microsoft Academic's field ratings for conferences* https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-analytics/.
15. Vrettas, G. & Sanderson, M. Conferences versus journals in computer science. *Journal of the Association for Information Science and Technology* **66,** 2674–2684 (2015).
16. Li, X., Rong, W., Shi, H., Tang, J. & Xiong, Z. The impact of conference ranking systems in computer science: A comparative regression analysis. *Scientometrics* **116,** 879–907. doi:10.1007/s11192-018-2763-1 (2018).
17. Küngas, P., Karus, S., Vakulenko, S., Dumas, M., Parra, C. & Casati, F. Reverse-engineering conference rankings: what does it take to make a reputable conference? *Scientometrics* **96,** 651–665. doi:10.1007/s11192-012-0938-8 (2013).
18. Steck, H. *Evaluation of recommendations: rating-prediction and ranking* in *Proceedings of the 7th ACM conference on Recommender systems* (2013), 213–220. doi:10.1145/2507157.2507160.
19. Chowdhury, G. R., Al Abid, F. B., Rahman, M. A., Masum, A. K. M. & Hassan, M. M. *Prediction of upcoming conferences ranking in Bangladesh based on analytic network process and machine learning* in *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)* (2018), 463–467. doi:10.1109/ICISET.2018.8745590.
20. Udupi, P. K., Dattana, V., Netravathi, P. & Pandey, J. *Predicting global ranking of universities across the world using machine learning regression technique* in *SHS Web of Conferences* **156** (2023), 04001.
21. *Scopus* https://www.scopus.com.
22. *DBLP* https://dblp.org/.
23. *Google Scholar* https://scholar.google.com/.
24. Kobzar, A. I. *Applied mathematical statistics* (Fizmatlit, 2006).
25. Orlova, I. V., Kontsevaya, N. V., Turundaevsky, V. B., Urodovskikh, V. N. & Filonova, E. S. Multidimensional statistical analysis in economic problems: computer modeling in SPSS (textbook). *International Journal of Applied and Fundamental Research,* 248–250 (2014).
26. Gafarov, F. M., Galimyanov, A. F., *et al. Artificial neural networks and applications* (Kazan Publishing House University, Kazan, 2018).

To cite: Ermolayeva A. M., Statistical methods for estimating quartiles of scientific conferences, Discrete and Continuous Models and Applied Computational Science 32 (1)(2024)5–17.DOI: 10.22363/2658-4670-2024-32-1-5-17.

Information about the authors

Ermolayeva, Anna M.—Assistant of Probability Theory and Cyber Security of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (e-mail: ermolaeva-am@rudn.ru, ORCID: https://orcid.org/0000-0001-6107-6461)

УДК 519.23
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-1-5-17 EDN: BAUTRQ

Статистические методы оценки квартилей научных конференций

А. М. Ермолаева

Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

Аннотация. В статье представлены результаты оценки квартилей научных конференций, выставленных ведущими рейтинговыми агентствами. Оценки получены на основе применения трёх методов многомерного статистического анализа: линейной регрессии, дискриминантного анализа и нейронных сетей. Для оценки использовалась обучающая выборка, включающая следующие факторы: возраст и периодичность конференции, количество участников и количество докладов, публикационная активность организаторов конференции, цитируемость докладов. В результате проведённого исследования линейная регрессионная модель подтвердила верность выставленных квартилей для 77% конференций, в то время как методы нейронных сетей и дискриминантного анализа дали близкие результаты, подтвердив верность выставленных квартилей для 81 и 85% конференций соответственно.

Ключевые слова: оценка квартилей научных конференций, дискриминантный анализ, нейронные сети, линейная регрессия