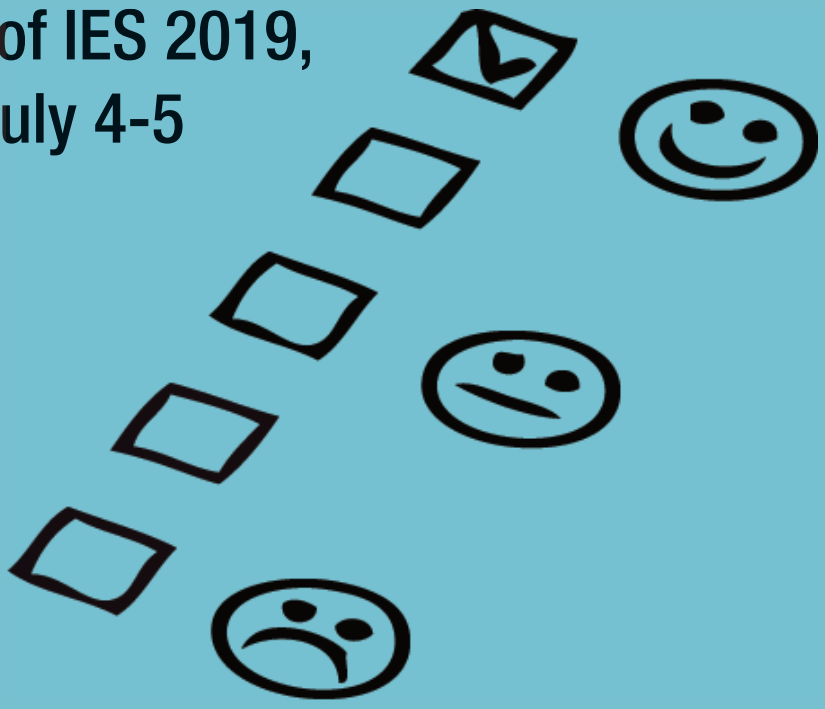
Luigi D'Ambra, Pietro Amenta, Antonio Lucadamo, Anna Crisci
EDITORS

Statistical Methods for Service Quality Evaluation

Proceedings of IES 2019, Rome, Italy, July 4-5



Pearson

Luigi D'Ambra, Pietro Amenta, Antonio Lucadamo, Anna Crisci
EDITORS

Statistical Methods for Service Quality Evaluation

Proceedings of IES 2019, Rome, Italy, July 4-5

9th International Conference **IES 2019** - Innovation & Society - **Statistical evaluation systems at 360°: techniques, technologies and new frontiers** organized by Statistics for the Evaluation and Quality in Services Group of the Italian Statistical Society and European University of Rome

Published by Pearson

www.pearson.com

Giugno 2020 ISBN 9788891921239

On the acceptable level of inconsistency of Pairwise Comparison Matrices: a proposal of thresholds based on the percentile rule....1
Livelli accettabili di inconsistenza per le matrici di confronto a coppie: una proposta di calcolo delle soglie basata sulla regola del percentile
Pietro Amenta, Antonio Lucadamo and Gabriella Marcarelli

Mines and quarries: An analysis of withdrawals determinants in Italy 9
Cave e Miniere: un'Analisi delle Determinanti dell'Estrazione in Italia
Sabrina Auci and Donatella Vignani

Evaluating people's behaviour towards risk: a multidimensional problem....15
Valutazione del comportamento individuale verso il rischio: un problema multidimensionale
Luigi Bollani, Guido Antonio Rossi and Ivan Sciascia

Inferential versus descriptive statistical approach in the analysis of Delphi performance: A case study. 23
Analisi delle performance del Delphi. Approccio inferenziale verso descrittivo: studio di un caso
Bolzan M., Auciello M., Pesarin F

Biplot and unfolding models for evaluation of teacher behaviour in the classroom....33
Utilizzo dei modelli biplot e unfolding per la valutazione dell'atteggiamento dell'insegnante in classe
Giuseppe Bove, Maria Gaetana Catalano, Paola Perucchini, Alessio Serafini, Giovanni Maria Vecchio

Student satisfaction for Teaching and its impact on drop-out rate....41
La soddisfazione degli studenti per la didattica e il suo legame con il tasso di abbandono
Barbara Cafarelli, Angela Maria D'Uggento and Alessandra Petrucci

Space-time nonparametric analysis for the Italian real estate market....49
Analisi non parametrica spazio-temporale per il mercato immobiliare italiano
Cappello C. and De Iaco S. and Palma M. and Pellegrino D. and Posa D.

A Multidimensional Approach for Classifying Italian University Students by Mobility 57
Un approccio multidimensionale per la classificazione degli studenti universitari italiani secondo la mobilità
Sara Casacci

Robust analysis of the labor market 65
Analisi robusta del mercato del lavoro
Aldo Corbellini, Marco Magnani and Gianluca Morelli

Analysis of the financial performance in Italian football championship clubs via longitudinal count data 73
Analisi della performance finanziaria delle squadre di calcio Italiane attraverso longitudinal count data
Anna Crisci and Luigi D'Ambra

A Poset perspective for the evaluation of self-reported health of the elderly in Italy 81
Un metodo basato sul POSET per valutare lo stato di salute auto-percepita della popolazione anziana in Italia
E. Furfaro, L. Pagani and M. C. Zanarotti

Reduced K-means Principal Component Multinomial Regression with external information to evaluate consumer's preferences 89
Reduced K-means Principal Component Regression con informazioni esterne per la valutazione delle preferenze dei consumatori
Antonio Lucadamo and Pietro Amenta

A two-part finite mixture quantile regression model for semi-continuous longitudinal data ...97
Modello di regressione quantilica mistura a due parti per dati longitudinali
Maruotti Antonello, Merlo Luca and Petrella Lea

New developments in the evaluation of goodness of fit for multidimensional IRT models based on posterior predictive assessment: Results from the INVALSI data....103
Nuovi sviluppi nello studio della bontà di adattamento per i modelli di IRT multidimensionali basati sulla valutazione predittiva a posteriori: Alcuni risultati sui dati INVALSI
Mariagiulia Matteucci and Stefania Mignani

Models for measuring well-being. Reflective or Formative? 111
Modelli per misurare il benessere. Riflessivi o formativi?
Matteo Mazziotta and Adriano Pareto

Automated Content Analysis of Destination Image: a Case Study 119
Automated Content Analysis dell'immagine della destinazione: studio di un caso
Antonino Mario Oliveri and Gabriella Polizzi

Evaluating the determinants of environmentally significant behaviour in Europe.....127
Le determinanti dei comportamenti sostenibili in Europa
Gennaro Punzo, Demetrio Panarello, Margherita Maria Pagliuca, and Rosalia Castellano

Financial transaction data for early estimates of macroeconomic indicators for Services in Italy: value added and turnover index 135
I dati di pagamenti elettronici per le stime anticipate di indicatori macroeconomici dei Servizi in Italia: valore aggiunto e indice del fatturato
Alessandra Righi, Guerino Ardizzi, Alessandro Gambini, Filippo Moauro, Nazzareno Renzi

Education, Women and Empowerment: the case of India 143
Istruzione, donne ed empowerment: il caso indiano
Azzurra Rinaldi and Fabiana Sciarelli

LOCAL AUTHORITIES AND TOURIST USE OF THE TERRITORY.....151
ENTI LOCALI E USO TURISTICO DEL TERRITORIO
Sara Sergio

An index for crowdsourced data on multipoint scales in tourism services evaluation 159
Un indice per dati crowdsourced su scale multipoint nella valutazione dei servizi turistici
Venera Tomaselli and Giulio Giacomo Cantone

Testing consistency in preference orderings.....167
Un test sulla consistenza tra ordinamenti di preferenze
Amalia Vanacore, Maria Sole Pellegrino, Yariv N. Marmor and Emil Bashkansky

A simulation study to investigate the paradoxical behaviour of inter-rater agreement coefficients in non-asymptotic conditions 177
Studio in simulazione del comportamento paradossale dei coefficienti di accordo inter-valutatore in condizioni non asintotiche
Amalia Vanacore and Maria Sole Pellegrino

The effect of noise factors in experimental studies on aircraft comfort.....187
L'effetto dei fattori di disturbo sulla valutazione sperimentale del comfort aereo
Amalia Vanacore and Chiara Percuoco

Climate Change and Italian Cities: evidence from Meteo-climatic Statistics and Indices on Extreme Events.....197
Cambiamenti Climatici e Città Italiane: le Statistiche Meteoclimatiche e gli Indici di Estremi Climatici
Donatella Vignani, Francesca Budano, Claudia Busetti

Preface

This Book of Proceedings includes a selection of 25 peer-reviewed papers submitted to the **Innovation & Society 2019** (IES 2019) - *Statistical evaluation systems at* 360°: *techniques, technologies and new frontiers* held at the European University of Roma from July 4^{th} to 5^{th} , 2019. With 80 contributions organized in solicited and contributed sessions, two plenary talks and 220 authors, underlying a very strong interest around the evaluation topics, IES2019 has been the 9^{th} meeting of a two-year initiative proposed by the permanent group Statistics for the Evaluation and Quality in Services (SVQS).

SVQS group was born in 2004 thanks to the collaborative work of some members of the Italian Statistical Society (SIS) with a focus on national research programs and applied research activities, on statistical methods and methodologies for the evaluation of the quality of services in the public and private fields.

SVQS organizes IES conference every two years. Recent debate, indeed, has shown that public and private service quality measurement is the basic prerequisite for quality planning and strongly improve public policies that highly impact our societies. This topic is also strongly interconnected with the satisfaction measurement, which is one of the most important tools for firms and public institutions to fully capture consumers and citizens needs. Recently, new challenges have emerged in particular from study designs, large and heterogeneous data sources availability and complex treatment assignment mechanisms. Big data also provide a complement to traditional data sources, such as survey and census data, to create a complete analysis of a service process. In this field, numerical taxonomy, classification, multidimensional scaling and other ordination techniques, clustering, tree structures and other network models, as well as other statistical models (e.g., multilevel or latent variable models) for the analysis of data of a different nature (e.g., ranking or ordinal categorical data) play a crucial role together with related inferential methods that may depart from traditional methods (e.g., methods based on composite likelihood or generalized estimating equations).

IES2019 has been sponsored by the Italian National Institute of Statistics (ISTAT), the European Network for Business and Industrial Statistics (ENBIS), and by two groups of the Italian Statistical Society: Statistics and Data Science (SDS) and Enhancement of Public Statistics (VSP).

This conference aimed at

- shedding light on the main statistical approaches and methodologies for evaluation, currently in use in different contexts, of public utility services;
- fostering advanced methodological research supporting the dissemination of ideas related to several fields of interest;
- contributing to the discussion on the innovative statistical evaluation systems impact of services, involving several economic and social policies actors;
- being a platform where statisticians, data analysts, machine learning researchers meet to understand and analyze service phenomena with data.

Previous editions of IES conference were:

- **IES2009** held at University of Brescia (June 24-26, 2009) with selected papers published in special issues of *Electronic Journal of Applied Statistical Analysis* (*EJASA*) and *Statistica & Applicazioni*;
- **IES2011** held at University of Florence (May 30-June 1, 2011) with selected papers published in a special issue of the *Journal of Applied Quantitative Methods* (ADSE);
- **IES2013** held at University of Milan "Bicocca" (December 9-13, 2013) with selected papers published in the *Procedia Economics & Finance* (Elsevier);
- **IES2015** held at University of Bari "Aldo Moro" (June 8-9, 2015) with selected papers published in a special issue of *Quality & Quantity* (Springer);
- **IES2017** held at University of Naples "Federico II" (September 6-7, 2017) with selected papers published in special issues of *Social Indicator Research* (Springer), *Quality & Quantity* (Springer) and *EJASA* (ESE).

Next IES conference *Innovation and Society 5.0: Statistical and Economic Models and Techniques for Quality Assessment* (**IES2021**) will take place from July 1^{th} to 2^{th} 2021 at the Department of Economics of the University of Campania "Luigi Vanvitelli" (Capua, Italy).

Moreover, special issues of the international journals *Socio-Economic Planning Science* (Elsevier) and *Metron* (Springer) published a selection of full papers. This is a strategic way to disseminate recent developments and critical discussion in statistics to a wider community who did not participate directly to the event.

The Scientific Program Committee, the Chair (Prof. Matilde Bini) and the Local Organizing Committee, with the support of the European University of Rome, have all contributed to a productive and stimulating IES2019 conference. We acknowledge their precious work.

Luigi D'Ambra, Pietro Amenta, Antonio Lucadamo, Anna Crisci

Editors

IES2019 Scientific and Program Committee

Giorgio Alleva (University of Rome ”La Sapienza”)
Pietro Amenta (University of Sannio)
Francesco Bartolucci (University of Perugia)
Eric Beh (University of Newcastle, Australia)
Wicher Bergsma (The London School of Economics and Political Science, UK)
Matilde Bini (European University of Rome) - **Chair**
Mario Bolzan (University of Padova)
Eugenio Brentari (University of Brescia)
Maurizio Carpita (University of Brescia)
Rosalia Castellano (University of Naples ”Parthenope”)
Paola Cerchiello (University of Pavia)
Vartan Choulakian (University of Moncton, Canada)
Corrado Crocetta (University of Foggia)
Antonello D’Ambra (University of Campania ”L.Vanvitelli”)
Luigi DAmbra (University of Naples ”Federico II”)
Tonio Di Battista (University of Chieti ”G.DAnnunzio”)
Michele Gallo (University of Naples ”Orientale”)
Anna Giraldo (University of Padova)
Luca Greco (University of Sannio)
P.M. Kroonenberg (University of Leiden, Netherlands)
Michele La Rocca (University of Salerno)
Rosaria Lombardo (University of Campania ”L.Vanvitelli”)
Filomena Maggino (University di Florence)
Paolo Mariani (University of Milan ”Bicocca”)
Lucio Masserini (University of Pisa)
Stefania Mignani (University of Bologna)
Francesco Palumbo (University of Naples ”Federico II”)
Lea Petrella (University of Rome ”Sapienza”)
Alessandra Petrucci (University of Florence)
Jean Michel Poggi (Paris-Descartes University & Lab. Maths Orsay, France)
Emilio Porcu (University of Newcastle, UK)
Donato Posa (University of Salento)
Gennaro Punzo (University of Naples ”Parthenope”)
Carla Rampichini (University of Florence)
Alessandra Righi (ISTAT)
Pasquale Sarnacchiaro (University of Rome ”Unitelma Sapienza”)
Nicola Tedesco (University of Cagliari)
Amalia Vanacore (University of Naples ”Federico II”)
Grazia Vicario (Polytechnic University of Turin)
Maurizio Vichi (University of Rome ”Sapienza”)

IES2019 Organising Committee

Antonio Lucadamo (University of Sannio)
Pietro Amenta (University of Sannio)
Matilde Bini (European University of Rome)
Lucio Masserini (University of Pisa)
Pasquale Sarnacchiaro (University of Rome ”Unitelma Sapienza”)
Margherita Velucchi (University of Rome ”Unitelma Sapienza”)

An index for crowdsourced data on multipoint scales in tourism services evaluation

Un indice per dati crowdsourced su scale multipoint nella valutazione dei servizi turistici

Venera Tomaselli and Giulio Giacomo Cantone

Abstract When statistical approaches for customer satisfaction are employed in larger digital applications, longitudinal structured Big Data are produced. Result of this 'crowd rating' is not independent from interaction between users and online platforms. Given an empirical case study, we propose our interpretation on employment of parametric and not parametric indexes for rankings' construction from data collected from rating online platforms.

Abstract *Quando i metodi statistici della customer satisfaction sono impiegati in larghi contesti digitali si producono Big Data longitudinali strutturati. Il risultato di questo "crowd rating" non è indipendente dalle interazioni tra consumatori e piattaforma. Alla luce di un caso studio, proponiamo le nostre interpretazioni sull'impiego di indici parametrici e non parametrici per la costruzione di ranking di punteggio per mezzo di dati ottenuti da piattaforme di rating.*

Key words: tourism evaluation**,** customer satisfaction, crowd rating, ranking.

1 Evaluation from crowd rating

Crowd rating is a data gathering process to collect opinions on a topic. A common application of crowd rating in tourism is for evaluation of perceived quality:

[1] Venera Tomaselli, Department of Political and Social Sciences, University of Catania, IT; email: tomavene@unict.it.
Giulio Giacomo Cantone, email: prgcan@gmail.com.

platforms and businesses relying on recommender systems (*Facebook, Amazon Group, TripAdvisor*, etc.) are common examples.

A contributing factor of enthusiasm for crowd rating is the generally low cost to achieve acquisition of large structured datasets [7]. We found four relevant reasons to adopt crowd rating for organization of people's opinion:

- to build trust in digital communities (e.g., *eBay*)
- to display to the public a massive flux of information (e.g., to rearrange Big Data into sorted rankings)
- to develop matching algorithms for recommender systems
- to lock-in and select users, as after they 'scored' a desirable reputation in a platform, it's less likely that they will leave the platform for a competitor, so as not to lose their previous 'score' [4].

We noticed an affinity with established practices in customer satisfaction [15], but we relate this methodology to historical Galton's experiment [6]. The British polymath showed that, challenged 787 totally unknown people to estimate the weight of an ox, the difference between the median of crowd's opinions and the exact value was lesser than 1%. In particular, we will develop this intuition about the employment of median for measuring people's opinion.

Table 1: Differences between experimental design of research and implemented rating systems

Controlled research design	***Implemented design on websites***
An exact value exists and it's approximate by a metric	Supplies the lack of unit of measures for features like 'taste'
The experiment has a fixed end, and until then, other's people opinion is secret	Public crowd rating websites run with no end times and no secrecy of what is trending
No competition among subjects of measurement	Enables a competition to get better positions in future rankings, or to influence recommender systems

The following features highlight structural complexity in data production in crowd rating:

- lack of exact measure: while Galton asked people to estimate *weights*, crowd rating often aims to estimate latent features like *quality* or *satisfaction*. An established method to evaluate an inter-subjective value in perceptions employs ordinal multipoint scales. We commonly observe this method in rating systems

- secrecy of opinions: While open platforms vigorously enforce secrecy on how their algorithms are 'hardcoded', their business model is still based on making public the monitored data that includes reviews and ratings
- competition: when a new technology enables to rank products under a common criterion of interrogation ('query'), to keep a ranking position online becomes a primary target for any of those products, and in particular to be in the first visualized webpage of any related *query* on search engines [17].

Adopting definition of this new frontier of tourism as a micro-social system in Jeacle and Carter [8], a quantitative evaluation of satisfaction should not ignore the following statistical biases commonly associated to non-experimental studies on public opinions:

- non-independency of observations: earlier ratings influence late ratings. Experimental studies [16] and empirical findings [9] on crowd rating suggest that, in the absence of secrecy of trends, judgements over products converge towards a strong modal class of answers ('herding'). Research on platforms *Amazon* and *Yelp* [2] confirmed the hypothesis of the existence of a social mechanism of 'herding' that ensures that earlier ratings are more likely to influence future 'popularity' of products than later ones
- survivorship bias: competition of subjects reflects competition for survival in a market [5]. By this struggle for survival, some subjects disappear from the market others show up. Not only subjects in the same *query* or list have different lifespans, but their data can be retroactively censored by platforms too. Platforms do not desire to host an inactive subject in their online rating service. This could be a misleading factor in analysis because it censors those subjects where it is more likely that 'unpopularity' and *weaknesses* will be observed. More generally, it skews the distribution of ratings into higher numerical values [13]
- frauds and optimization strategies: platforms monitor data which are voluntarily submitted. Sometimes they lack clear procedures to confirm the general *sincerity* of the submitted data. While technologies to improve *fake detection* are constantly in development[2], frauds are usually a consistent factor of skewness in reviews [14, 12] A further reflection is necessary: while a subject who actually manipulates a ranking by the submission of *fakes* may be held responsible of crime under a variety of legislation, *TripAdvisor* states that 'optimization' and anything that does not involve a 'payment' to fake a review is not against its Terms of Service[3]. We could conclude that 'asking gently' to submit a max-scored rating should be considered a legitimate strategy of optimization of reputation and awareness, but it is made clear that material incentives in exchange for max-scored ratings is inadmissible behaviour under ToS[4]. Thus, those ratings will be

[2] https://www.tripadvisor.com/TripAdvisorInsights/w3703
[3] *ibidem.*
[4] https://www.tripadvisor.com/TripAdvisorInsights/w591

subjected to censor, introducing another bias in observed results between ratings already revised and those not.

Inequalities and biases can be extremely relevant for an effective rating. In presence of these, even with high number of reviews, many standard assumptions are not a suggestible approach of time series data analysis.

2 Web-scraped Data

We sampled a list of 60 web pages of active restaurants on *TripAdvisor.com*, the first operating since October 10th, 2009. We define 'time-point' an interval of time, which marks and gives an information about the time when the recorded review was submitted When is not explicated otherwise, we adopt 'day' as standard time-point.

A restaurant with at least one recorded review at a time-point considered an 'active subject since that time-point' ('active subjects') and we expect that the restaurant was operating at least since that time-point. Restaurants are considered 'inactive subjects' until the time-point they receive the first review.

For all the restaurants, the sampling criteria were:

- addressed in the tourism city of Catania, IT
- not less than 20 reviews at August 5th, 2018, from a total of 3204 days of activity
- 'pizza' in the menu

With a web scraping script in *R* framework, we collected metadata from the reviews in the sample (N=26.888), in particular we recorded only the following variables of metadata from reviews:

- day of submission, in the range of 3204 days, as *t* timing
- uniquely associated ID of subject restaurant on *TripAdvisor*
- recorded class of review scores, within the ordinal scale of 1 through 5.

While the number of active subjects grows linearly, the number of collected reviews does not. Even taking into account survivorship bias, which obscures data from subjects active in the past but inactive at August 5th, 2018, this does not explain the difference between the two growth ratios. The maximal divergence between the two growth ratios of (i) active subjects and (ii) collected reviews is reached on Day 1513th (December 18th, 2013) of 3204 (47%), when 34 of 60 subjects (57%) were already active.

Daily relative cumulative (until the last day) frequencies of classes of scores $F(x) = n_x / N$ were stable most of the time. The modal class, indeed, was always $x = 5$, floating around a median frequency of .441. Data are consistent with results from previous studies on Italian cities on *TripAdvisor* [1]. Frequencies were stable and $x = 5$ scored almost half of total reviews, hence we supposed that weekly $x = 5$ should have been distributed around a central value of .445 (therefore $x \neq 5$ around .555).

After we aggregated daily data-points into weekly data-points by summing all the reviews with 7 days between a Sunday and its subsequent Saturday, starting from August 7th, 2011 and ending August 4th, 2018, for a total of 365 weeks in 7 years, we found the aforementioned hypothesis to be coherent with our data: ***weekly***

$x = 5$ had a geometric mean of .4372, an average of .4487, a median of .4487, with a standard deviation of .0983, confirming the stable value.

We framed the time series from the starting week (August 7th, 2011) because this is the first week that satisfies this condition: every subsequent week had at least one $x = 5$ and one $x \neq 5$ reviews. Another noteworthy property of this starting week is that at least 10 subjects were already active on that day. By doing this frame, we excluded .005 of total recorded reviews ('N') and .202 of total recorded time-points ('days').

3 Ratings estimation in a ranking system

Although the debate between mean and median as estimators of the central value of records from multipoint scales is an open controversy [11, 18], we will argue that for low amounts of classes of score, the mean must be adopted.

We noticed that, in ordinal scales, the robustness towards the extreme values of the median as estimator of central value is of no utility because the values are enclosed in a finite domain. Its lack of sensitivity towards small differences, on the other hand, is a disadvantage in cases where these differences, even the smallest, are decisive in sorting. In particular, when the estimated parameter is argument of a rank-function for a benchmark. This incongruence is exacerbated in a longitudinal context: the median as an argument of a rank-function is sensible to factors such as skewness in frequencies of classes of values, as in our case. In particular, we observe that a minimal increase of the median of scores of an item after t causes a big 'jump' or permutation of rank [3] of the item in the ranking towards the first positions. This property seems undesirable because, under ideal conditions, every permutation of ranks after t should be imputable much to a mutation of the measured performance, less to random or structural error in the model.

More specifically, the sum of amounts of $x = 4$ and $x = 5$ was always over .7 of the total, both in our data and in another study [1]. Hence, to rank subjects by the median always produces a binary classification, which is of no use for ranking purposes for the aforementioned reasons. To estimate rating of subjects, we came to conclusion that a normalized average:

$$\frac{\frac{\sum n_{x*x}}{N} - \min(\xi)}{\max(x) - \min(\xi)} \tag{1}$$

may be the viable solution when subjects are sorted in a ranking.

For those situations where we can be confident to detect strong skewedness toward the highest ("max") class of scores, we noticed that the simpler non parametric ratio:

$$f(x = \max) = \frac{n_{max(x)}}{N} \tag{2}$$

will lead into a more stable over time, ranking. We don't assume that the feature of lesser variability over time of a ranking is valuable by default in statistical analysis, but it can be in some cases.

A variant that helped further comparison of ratings between (1) and (2) for our case study reckons on adding the 4th class of score (the second highest) to the numerator of the ratio

$$f(x = 5) + f(x = 4) \tag{3}$$

4 Conclusions

One of the results of the normalization in (1) is to enclose estimations in the dominion (0,1). This is valuable for practical uses because allows to make further reflections and compute systemic differences (i.e. stability) between rankings sorted by parametric and non-parametric estimators of rating, i.e. for items rated with different scales of scores.

The issue of developing a tool to evaluate different estimators of ratings for ranking is still open. Our suggestion is to take in consideration the rigid mathematical structure of a ranking, which is a succession of natural number, where the distance among ranks is linear. Therefore, the more the estimated ratings associated to ranks fit the assumption of linearity, proceeding from $rank_{min} -> rating = 1$ through $rank_{max} -> rating = 0$, the more that estimator fits the purpose of ranking the empirical sample and, we assume, the target population.

References

1. Baccianella, S., Esuli, A., Sebastiani, F.: Multi-facet Rating of Product Reviews. In: Proceedings of 31th European Conference on IR Research on Advances in Information Retrieval, pp. 461-472. (2009).
2. Bai, T., Zhao, X., He, Y., Nie, J. Y., Wen, J. R.: Characterizing and Predicting Early Reviewers for Effective Product Marketing on E-Commerce Websites. IEEE Transactions on Knowledge and Data Engineering, 30(12), 1-14 (2018).
3. Corain, L., Arboretti, R., Bonnini, S.: Ranking of multivariate populations: a permutation approach with applications, CRC Press, Boca Raton, FL (2016).
4. Dellarocas C.: Designing Reputation Systems for the Social Web. In: Masum, H., Tovey, M. (eds) The Reputation Society, pp. 3-12. MIT Press, Cambridge, MA (2011).
5. Farmer R.: Web Reputation Systems and the Real world. In: Masum, H., Tovey, M. (eds) The Reputation Society, pp. 13-24. MIT Press, Cambridge, MA (2011).
6. Galton, F.: Vox Populi, Nature, **75**, 450-451 (1907).
7. Geiger, D., Schader, M., Rosemann, M., Fielt, E.: Crowdsourcing information systems - definition, typology, and design. In: Proceeding of International Conference on Information Systems, pp. 9-11 (2012).
8. Jeacle, I., Carter, C.: In TripAdvisor we trust: Rankings, calculative regimes and abstract systems, Accounting, Organizations and Society, **36(**4/5), 293-309 (2011).

9. Lee, Y.J., Hosanagar, K., Tan, Y.: Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings. Management Science, **61**(9), 2241-2258 (2015). doi:10.1287/mnsc.2014.2082
10. Lewis, J. R.: Multipoint scales: mean and median differences and observed significance levels. Int. J. Hum.-Comput. Interact. **5**, 382–392 (1993).
11. Lewis, J.R., Sauro, J.: Quantifying the user experience: Practical statistics for user research, : Morgan Kaufmann , Cambridge, MA (2016).
12. Li, J., Ott, M., Cardie, C., Hovy, E.: Towards a General Rule for Identifying Deceptive Opinion Spam. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics pp. 1566–1576, Baltimore, MD (2014).
13. Mangel, M., Samaniego, F.: Abraham Wald's work on aircraft survivability. Journal of the American Statistical Association. **79,** 259–267 (1984).
14. Ott, M., Cardie, C., Hancock, J.: Estimating the prevalence of deception in online review communities, Proceedings of the 21st international conference on World Wide Web, 201-210 (2012).
15. Pizam, A., Shapoval, V., Ellis, T.: Customer satisfaction and its measurement in hospitality enterprises: a revisit and update. International Journal of Contemporary Hospitality Management, **28**(1), 2–35 (2016).
16. Salganik, M. J., Dodds, P. S., Watts, D. J.: Experimental study of inequality and unpredictability in an artificial cultural market. Science, **311**, 854–856 (2006).
17. Varian, H. R.: The economics of Internet search. In: Bauer, J., Latzer, (eds) Handbook on the Economics of the Internet (pp. 385–394), Edward Elgar Publishing, Cheltenham UK (2016).
18. Velleman, P. F., Wilkinson, L.: Nominal, ordinal, interval, and ratio typologies are misleading. American Statistician, **47**(1), 65–72 (1993).