

RECENT TRENDS IN QUANTITATIVE RESEARCH IN THE FIELD OF TOURISM AND HOSPITALITY

Nemanja Stanišić

Singidunum University,
Belgrade, Serbia

Abstract:

Quantitative research in the field of tourism and hospitality is undergoing an important transformative phase, with crowdsourced data being the major driving force. The author discusses several key aspects of the research process likely to evolve during this phase and related implications for researchers.

Keywords:

tourism, hospitality, quantitative research, multilevel modelling, data mining.

INTRODUCTION

The last decade in the tourism industry is characterized by the emergence and rapid proliferation of globally accessible Internet services, which have redefined the way customers select properties, make reservations and payments, and share their impressions about the services received. The exponentially increasing amount of data available due to these services has also drawn the attention of the research community, which increasingly uses these data for research purposes. In the following sections, the author shall briefly discuss the implications of such development for several key aspects of quantitative research, namely: 1) data collection; 2) data characteristics; 3) statistical tools being employed; 4) type of research questions being addressed, and 5) research reproducibility and extensibility.

DATA COLLECTION

Whenever applicable, researchers are replacing the data sets collected through questionnaires and interviews by those collected from online services, with Booking.com and TripAdvisor.com being the most prominent sources. Besides being less costly and more readily available, these data sets are not nearly as limiting with regard to the obtainable sample size. Samples of thousands, or even millions, of observations are becoming commonplace in the literature. Such large samples endow researchers with enough statistical power to examine even the subtlest causal effects that may be of theoretical interest (Radojevic, Stanisić, Stanić, & Sarac, 2014). Still, researchers have to be mindful of the fact that these data (more often than not) have different statistical properties than those collected through traditional surveys.

Correspondence:

Nemanja Stanišić

e-mail:

nstanisic@singidunum.ac.rs



DATA CHARACTERISTICS

Owing to their typically wide spatial and temporal coverage, online data have characteristics of panel data. Specifically, customer reviews (or ratings or comments) are provided over a certain period of time, requiring that the temporal trend be accounted for. Also, the facts that the reviews pertain to more than one property and that each reviewer is allowed to provide more than one review impose the presence of property and reviewer-specific effects, which also need to be controlled for. It is also likely that samples include properties from multiple destinations and reviewers from different countries. If so, these classifications need to be accounted for as well, as they have been shown (Bulchand-Gidumal, Melián-González, & González Lopez-Valcarcel, 2013; Kozak, 2001; Liu, Teichert, Rossi, Li, & Hu, 2017) to be capable of exerting unique (although less prominent) effects on the valence of customer reviews. Moreover, as each property can be located in only one destination, and each reviewer can be of only one nationality, there is a hierarchical structure to these classifications. Technically, hotels are nested within the destinations while reviewers are nested within their nationalities. The presence of structures and hierarchical relationships in online data will ultimately force researchers to adopt statistical tools capable of accounting for these properties.

STATISTICAL TOOLS

The initial reaction to the abundance of data was the adoption of data mining and machine learning techniques, which resulted in numerous inspiring research papers (Phillips, Zigan, Santos Silva, & Schegg, 2015; Xiang, Schwartz, Gerdes, & Uysal, 2015). However, the capability of these tools to fully address complex research questions and contribute to the existing body of theoretical knowledge is limited. While being highly efficient in detecting patterns and selecting a subset of relevant explanatory factors, these tools typically (although certain exceptions do exist) do not prompt the researcher to either define the structures present in the data or to include any prior information that might be relevant for the research problem at any point of the research process. Disregarding these important pieces of information results in suboptimal performance of the effect estimation, including biased estimates. Furthermore, these tools typically stay opaque with regard to the underlying causal mechanisms (the “black box” effect). In this regard, a multilevel modeling framework seems to be the next step forward due to some important advantages:

- a) It accounts for the structures and hierarchical relationships present in the data.
- b) It allows partitioning of the total variance in the values of the dependent variable into between- and within-unit components, which may be particularly interesting from a theoretical perspective.
- c) It allows the inclusion of level-specific explanatory variables (such as destination- or nationality-specific variables) without committing ecological fallacy.
- d) It allows for the fact that the examined effects might not be identical across units (*e.g.* properties, destinations, reviewers, nationalities) and even possibly correlated with the unit-specific effects.

While most of the listed properties can be attained by modifying classical (single-level) statistical data mining or machine learning tools, within the multilevel modeling framework, they are implemented in an unprecedentedly systematic way. The application of mixed linear models on empirical data collected from TripAdvisor has been demonstrated by Radojevic, Stanisic, and Stanic (2015a, 2015b, 2017). Other equally prospective techniques belonging to the same multilevel family include multilevel structural equation modeling (Mehra, 2013) and Bayesian multilevel modeling (Bürkner, 2016). Clearly, the advantages of using multilevel frameworks are not limited to correcting for the bias that would otherwise be introduced, as they also allow researchers to tackle complex, multilevel questions that were previously beyond reach.

RESEARCH QUESTIONS

To promote the use of the multilevel framework, we offer a few examples of prospective research questions that could be easily addressed within the multilevel framework:

- a) Are there significant differences between the average rating scores across the destinations? (allow the intercepts to vary across the destinations)
- b) Does the importance (indicating the reliability) of star classification vary across the destinations? (allow the slopes for the star classification variable to vary across the destinations)
- c) Do the hotels located in destinations that have implemented more reliable star classification systems get higher ratings on average? (examine the correlation between the random intercepts and the random slopes)



- d) Can the differences in reliability of the star classification systems implemented in the destinations be explained by their respective cultural characteristics (e.g., uncertainty avoidance)? (cross-level interaction between the cultural dimension variables and star classification variable)

While the first question can be tackled within single-level statistical techniques (ANCOVA), the latter three can be tackled exclusively within the multilevel modeling framework. These and many more interesting research questions can be tackled by extending the multilevel research framework presented by Radojevic *et al.* (2017). This brings us to another major recent development, which is the increased implementation of the concept of reproducible and extensible research.

RESEARCH REPRODUCIBILITY AND EXTENSIBILITY

The concept of reproducible research requires that “researchers make code and data available to others so that the data can be analyzed in a similar manner as in the original publication” (Stodden, Leisch, & Peng, 2014). While the key benefit of this practice is verifiable and, hence, more credible research results, it also allows researchers to reuse data sets and program code, avoiding redundancy in efforts related to the collection and cleaning of data, and even some parts of statistical modeling. This is particularly important because the extensive geographic coverage of online data sets makes them interesting and relevant for researchers around the world. Our recent research (Radojevic *et al.*, 2017) has been made completely reproducible, and both the data set¹ and the complete R code used for the analysis will be made publicly available² as soon as the article is published. We encourage researchers to extend and build upon the existing multilevel platform or use the data set in even more creative ways.

CONCLUSIONS

The author discusses several key aspects of quantitative research that are likely to evolve in the near future as a result of the increased use of online data for re-

- 1 The data set includes 3,488,473 customer reviews provided by 2,233,671 unique, registered TripAdvisor users of 210 different nationalities related to 13,410 hotels located in 80 capital cities around the globe. It also includes a comprehensive list of attributes of all the reviewed hotels, each reviewer's TripAdvisor rank and nationality, the review-specific type of travel, and the date of the review (between August 31, 2002, and May 15, 2015)
- 2 Through the Mendeley Data service (Radojevic, Stanic, & Stanic, 2016)

search purposes. The main points stressed in the discussion are summarized in Table 1.

Table 1. Summary of the key differences between classic and modern approaches to quantitative analysis

	The past	The future
Sources of empirical data	Traditional surveys	Big data (crowdsourced)
Data characteristics	Small, simply structured samples	Large samples with complex structures (repeated observations and hierarchical structures)
Statistical tools	Single level models: OLS regression, (M) AN(C)OVA, structural equation modeling (SEM)	Multilevel models: multilevel regressions, Bayesian multilevel models, multilevel SEM
Type of research questions	Simple, single-level questions	Complex, multilevel questions
Research reproducibility	Research cannot be reproduced, modified, or built upon	Research can be easily reproduced, modified, and built upon

The trends that are likely to prevail in the near future will, besides the above described procedural aspects, affect the skills of quantitative researchers in the field of tourism and hospitality. In addition to advancing theoretical knowledge and developing creativity, researchers will unavoidably need to learn how to become data scientists in order to remain vanguard in the near future.

REFERENCES

- Bulchand-Gidumal, J., Melián-González, S., & González Lopez-Valcarcel, B. (2013). A social media analysis of the contribution of destinations to client satisfaction with hotels. *International Journal of Hospitality Management*, 35, 44-47. doi:10.1016/j.ijhm.2013.05.003
- Bürkner, P. C. (2016). brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software*, in press. Retrieved from <https://cran.r-project.org/web/packages/brms/vignettes/brms.pdf>



- Kozak, M. (2001). Comparative assessment of tourist satisfaction with destinations across two nationalities. *Tourism Management*, 22(4), 391-401. doi:10.1016/S0261-5177(00)00064-9
- Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, 59, 554-563. doi:10.1016/j.tourman.2016.08.012
- Mehta, P. D. (2013). *Applied Quantitative Analysis in Education and the Social Sciences*. In Y. Petscher, C. Schatschneider, & D. L. Compton (Eds.), *Applied Quantitative Analysis in Education and the Social Sciences* (p. 329). Routledge. Retrieved from <http://books.google.com/books?hl=en&lr=&id=ZzBYKYZPNBgC&pgis=1>
- Phillips, P., Zigan, K., Santos Silva, M. M., & Schegg, R. (2015). The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management*, 50, 130-141. doi:10.1016/j.tourman.2015.01.028
- Radojevic, T., Stanistic, N., & Stanic, N. (2015a). Ensuring positive feedback: Factors that influence customer satisfaction in the contemporary hospitality industry. *Tourism Management*, 51, 13-21. doi:10.1016/j.tourman.2015.04.002
- Radojevic, T., Stanistic, N., & Stanic, N. (2015b). Solo travellers assign higher ratings than families: Examining customer satisfaction by demographic group. *Tourism Management Perspectives*, 16, 247-258. doi:10.1016/j.tmp.2015.08.004
- Radojevic, T., Stanistic, N., & Stanic, N. (2016). Inside the Rating Scores: A Multilevel Analysis of the Factors Influencing Customer Satisfaction in the Hotel Industry. Data set, Mendeley Data, v1. doi:10.17632/kwsrxshf9x.1
- Radojevic, T., Stanistic, N., & Stanic, N. (2017). Inside the Rating Scores: A Multilevel Analysis of the Factors Influencing Customer Satisfaction in the Hotel Industry. *Cornell Hospitality Quarterly*, in press.
- Radojevic, T., Stanistic, N., Stanic, N., & Sarac, M. (2014). Measuring customer satisfaction in the hospitality industry: an empirical study of the hotels in the capital cities of Europe. In *Proceedings of the 1st International Scientific Conference - Sinteza 2014* (pp. 788-795). Belgrade, Serbia: Singidunum University. doi:10.15308/sinteza-2014-788-795
- Stodden, V., Leisch, F., & Peng, R. D. (2014). *Implementing reproducible research*. Boca Raton: Chapman and Hall/CRC.
- Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120-130. doi:10.1016/j.ijhm.2014.10.013