

COMMENTARY

## Descriptive statistics are powerful tools for organizational research practitioners

Jennifer Diamond Acosta<sup>1\*</sup> and Scott Brooks<sup>2</sup>

<sup>1</sup>Allstate Insurance Company and <sup>2</sup>OrgVitality

\*Corresponding author. Email: [jdiamondacosta@gmail.com](mailto:jdiamondacosta@gmail.com)

Murphy (2021) argues that, in scholarly publications, increasing focus on presentations of descriptive statistics over complex statistics would increase the value and interpretability of research. As practitioners who conduct research for organizations, we believe the same could be said about the use of statistics in organizations. Like academic researchers, practitioners too are under pressure to increase our use of complex statistics. This can be exemplified by a popular analytics maturity model published by Gartner (2012), which implies that organizations should strive to move beyond descriptive statistics (i.e., low-maturity analytics) and invest more in predictive and prescriptive statistics (i.e., high-maturity analytics). A problem with this model is that it positions some statistics as inherently more desirable than others rather than advocating for the optimal match between the problem at hand and statistical approach. Moreover, it ignores that well-conducted research in organizations involves an interplay of statistical approaches as you move through the research process.

In this commentary, we discuss the types of statistics that are typically involved in each of four common steps in organizational research: (a) define the problem, (b) understand the data, (c) analyze the data, and (d) communicate results. As we move through the research process, we start with descriptive statistics to ensure we are focusing on a meaningful problem. We use descriptive statistics further to understand our data and inform more advanced analyses. We leverage more complex statistics, if appropriate, to address the problem and inform recommendations. Finally, we translate statistical findings and insights into simple representations in order to tell a clear, compelling, and actionable story to leaders.

### Define the problem

Problems that organizational research practitioners study are often driven by the concerns of organizational leaders who request that we conduct an analysis to better understand the problem and make recommendations. It is critical that we begin this type of engagement by validating that the concern is truly problematic. Descriptive statistics can be useful for pressure testing concerns before investing more heavily in analyzing and addressing them. For example, at one author's company, business leaders were concerned about an insufficient pipeline to replace retiring Baby Boomers. A simple descriptive analysis revealed that the number of employees nearing retirement in critical roles was significantly less than the number of successors deemed ready for the same roles; these findings put this concern to rest and prevented further investment into something that was not actually a problem.

Descriptive statistics can also be used to help us detect potential problems and deepen our understanding of them by shedding light on questions such as: How big is the problem? How long have we had it? Do other companies have it too? What factors might be related? This initial

work to clearly define and understand the problem plays an important role in informing the focus and approach of deeper or more advanced analytic efforts. For example, perhaps survey results suggest a problem with employee well-being, as a high percentage of employees have endorsed an item about feeling stressed at work. Several descriptive analyses could be done to learn more about this trend. A cost analysis could be done (e.g., stress-related healthcare costs to insurance, costs of stress induced leaves or time off) to quantify the magnitude of effect and determine how much to invest in addressing it. Historical trending could be referenced to understand whether stress is a new problem or something that has existed previously. Comparing employee segments could help pinpoint the groups that are most affected by stress (e.g., particular functions, levels, locations). Comparing against external benchmarks (i.e., indicators of how employees at other companies responded to the same item in the same year) would shed light on whether the trend around stress is similar to what is happening elsewhere or unique to the company. Bivariate correlations with other survey items can provide initial ideas about what factors might be related to employee stress. Findings from these descriptive analyses shape subsequent stages of analyses by highlighting important employee segments and potential causal factors to examine.

### Understand the data

From our practitioner perspective, we echo Murphy's (2021) points on the importance of examining descriptive statistics upfront to determine whether the data meet the necessary conditions for being analyzed with a more complex method. In contrast to academic research, where typically the researcher thoughtfully designs the data collection procedures to optimally support the study, organizational data used by practitioners are fraught with imperfections because they are sourced from imperfect processes and systems that were built for operational purposes rather than research. This means that organizational data are especially prone to limitations in accuracy, variance, or other problems. For example, a common challenge for practitioners is predicting performance, given that performance rating processes designed to support compensation and other administrative decisions generally produce ratings with limited variability and validity (Colquitt, 2017). Sourcing data from complex and unfamiliar human resource (HR) information systems creates challenges as well. With experience, practitioners learn how to properly clean and use the HR system fields, but often analysts don't have a high degree of familiarity with the nuances of the data, especially given the trend of HR departments outsourcing their analytics to centralized data science teams or vendors (Angrave et al., 2016). All of this poses an inherent risk for conducting complex analytics that fail or are invalid. Paying attention to descriptive statistics and following the recommendations of Bedian (2014) can alleviate that risk by ensuring that problematic data issues are detected proactively.

One organization's modeling story illustrates the risks and consequences of skipping this step. A firm was hired to create a resume screening model using historical data from the company's applicant tracking system (i.e., the fields applicants complete when applying for a job) to predict early performance and retention. The firm produced a strong model, but prediction was heavily reliant on one variable: whether the applicant was internal or external. The company's internal industrial-organizational psychologist looked at the descriptive statistics and noted that the internal applicant rate was far too high to be realistic for an entry-level job. Further investigation revealed a misunderstanding: The field was not indicating the candidate's status at the time of application for the entry level role but rather at the time of the most recent application. The model was great at predicting retention largely because the internal candidate field was retrospectively flagging employees who stuck around long enough to apply for more roles in the company. This discovery rendered the model nearly useless, and the team had to redo the work. A lot of wasted effort could have been prevented by examining the descriptive statistics first to ensure the data was fully understood before running complex models.

## Analyze the data

To be sure, descriptive statistics are insufficient by themselves. Sophisticated analyses fill a critical role in the toolbox. We are not suggesting pulling away from those, but we advocate ensuring a prominent role for basic descriptions of a data set. As practitioners, we choose our statistical approaches based on what is most appropriate for the analysis at hand. Though we often leverage complex analytics in our analyses (e.g., tree-based modeling methods for predicting attrition, latent class analysis for understanding employee segments), there are times when descriptive statistics are the best option as well. Consider two very realistic, and somewhat opposing, scenarios where descriptive statistics are valuable:

1. *Early stages of a new development program.* Here, we have an insufficient sample size for significance testing, but program owners will be expected to report on early indicators of success. Our only reasonable option is to monitor the descriptive data (e.g., participation, favorability of survey responses) or even leverage qualitative data (e.g., testimonials).
2. *Census engagement survey analysis at a large company.* In the space of employee surveys, we often find ourselves with large data sets of tens of thousands of cases or more. In these contexts, any comparison with history or between two groups will be statistically significant. We need to identify findings that are *important* or practically significant based on the magnitude of an effect size that is illustrated through descriptive statistics such as a comparison of means.

## Communicate results

As practitioners, above all else, we seek to make a positive influence on the organizations we serve. Often, the weak point in organizational interventions is not the analysis but the follow-through. To make a meaningful difference, therefore, we need to create strong, data-driven recommendations for change, but more importantly, we need to persuade decision makers to act on them. This is not easy, given that taking action almost always means an investment of money and other resources.

Our typical audience lacks knowledge of complex statistics. Moreover, they tend to be senior leaders who have many priorities, already live in a data-saturated world, and have limited time and attention at any given moment. Thus, our communication needs to be simple, clear, compelling, and concise. As a result, we find the best approach is guided descriptive statistics. This means that we use appropriate complex statistics to yield important insights. Then we take only the most important of those insights and translate them back into descriptive statistics in an intentional manner that enables us to make a strong case for action.

There are a couple of ways that descriptive statistics can be useful in making a compelling case for action. We can use them to quantify value or effectiveness in business terms. For example, let's say we are presenting on a new assessment to be added to the hiring process. Rather than sharing a validity coefficient, we could estimate the percent increase in good hiring decisions that could be expected from implementing the assessment, and the cost savings that would be incurred due to making fewer bad decisions. We recommend Cascio and Boudreau's (2011) book for useful guidance on quantifying the business value of HR initiatives. When rational ROI-based cases cannot be easily made, descriptive statistics can also offer provocative illustrations to appeal to the emotional sides of an audience. This might take the form of illustrating a large deficiency in the company compared with a benchmark (e.g., employees here are 50% less engaged than at other companies) or large year-over-year change (e.g., favorable views of leadership went from 80% to 60%), or highlight very low incidence (e.g., board of directors has only one minority member). These types of depictions can ignite an emotional or visceral reaction that can inspire action.

### Case example of full process: Uncovering diversity, equity, and inclusion insights in survey results

Identifying meaningful demographic differences in survey responses is a common challenge. Organizations often want to know whether survey responses differ by gender or racial/ethnic groups to learn whether there is a problem to be addressed. One author on this paper undertook such an analysis, and it provides a great example of how both descriptive and advanced statistics are used in complementary ways through the research process.

1. *Define the problem:* The organization sought to analyze survey data by demographic segments to detect problems and learn more about them. Descriptive comparisons did not reveal differences, but there was general recognition that this “surface of the ocean” examination could easily miss important dynamics underneath. Fueled by a strong interest in diversity, equity, and inclusion, the organization wanted to deploy more complex methods to understand whether there were diversity, equity, and inclusion concerns buried within specific pockets of the organization.
2. *Understand the data:* Descriptive views of demographic representation rates were used to understand where women and minorities were represented within primary workforce segments such as function, level, location, or intersections of these. This informed decisions about which segmentation variables could be reasonably used in advanced analyses.
3. *Analyze the data:* Tree-based analyses (i.e., CHAID) were conducted with employee engagement as the dependent variable. Job level was the most potent demographic variable, with vice presidents being very favorable, directors and manager being next most favorable, and individual contributors being least favorable. CHAID provided additional insight in that it further broke down the vice presidents by gender, revealing a 10-point gap between male and female vice presidents. Executive women were still quite engaged, but executive men were extraordinarily so.
4. *Communicate results:* This finding was discovered with an assist from fairly involved, inferential statistics; however, it was easy to illustrate in a bar chart showing the percentage of favorable values by gender. The top executive found the 10-point gap highly distressing and contrary to his self-concept as an inclusive leader.

What happened? Heart-to-heart conversations were held, focus groups were conducted, and female leadership programs were launched or refined. The engagement survey was launched again a year later, and the researchers were asked to follow up on this issue. Unexpectedly, the results on the primary indicators revealed a widened gender gap on not only engagement but also career issues and confidence that workplace issues would be resolved appropriately. Notably, this coincided with the rising public awareness of #metoo. Undeterred, the organization used the new findings—expressed in simple descriptive statistics guided by more advanced analyses—to refine and address a clearer understanding of the underlying issues. As a result, the following year the gender gaps at the executive level were dramatically reduced. Throughout this case, the issues were complex, and some complex statistics were used to understand them, but the language of the story to executives was purely descriptive statistics.

### Conclusion

Though organizational research practitioners are often pressured to make use of sophisticated statistical techniques, it is important to recognize the critical role that descriptive statistics play in helping us drive change in organizations. We argue that descriptive statistics are even *more* valuable to practitioners in organizations than in the academic context about which Murphy (2021) writes. To make a difference, practitioners need important problems to solve and need

to convince leaders to take action on research-based recommendations for solving those problems. Descriptive statistics help us achieve these objectives. Although we may use sophisticated statistical approaches to define our positions and recommendations, we find descriptive statistics to be highly effective tools for depicting priorities, illustrating effect sizes, highlighting implications, and demonstrating progress.

## References

- Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal*, *26*, 1–11.
- Bedeian, A. G. (2014). More than meets the eye: A guide to interpreting the descriptive statistics and correlation matrices reported in management research. *Academy of Management Learning and Education*, *13*, 121–135.
- Cascio, W., & Boudreau, J. (2011). *Investing in people: Financial impact of human resource initiatives*. Pearson Education.
- Colquitt, A. (2017) *Next generation performance management: The triumph of science over myth and superstition*. Information Age Publishing.
- Gartner Inc. (2012) *Analytics maturity model*. <https://www.gartner.com>.
- Murphy, K. (2021). In praise of table 1: The importance of making better use of descriptive statistics. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *14*(4), 461–477.

---

**Cite this article:** Acosta, JD. and Brooks, S. (2021). Descriptive statistics are powerful tools for organizational research practitioners. *Industrial and Organizational Psychology* *14*, 481–485. <https://doi.org/10.1017/iop.2021.116>