NjP *Journal of Organizational Behavior Education 14: 111-128.*

Navigating a Sea of Data: An Experiential Exercise to Teach Data Analysis Using Secondary Data

Kenneth M. York
School of Business Administration, Oakland University, USA

Abstract. In the era of "Big Data" organizations have large amounts of data about their customers and their employees. Increasingly, all organizations must deal with large amounts of data as part of their daily business tasks. A critical skill for business students is using data to aid decision making. Without data analysis skills to make sense of the data and make management decisions based on the data, it is all just numbers. In this experiential exercise, students learn how to analyze data by being given management questions and an online data set, then get feedback on their answers. The exercise includes five questions with feedback to give to students on the results they should have gotten and the conclusions they should have made, plus additional learning points and variations on the exercise. The exercise can be used in any organizational behavior course to build the data analysis skills of students.

Keywords: experiential exercise, data analysis, statistics, secondary data.

1. Introduction

We are in the era of "Big Data". Organizations now have access to large amounts of data about their customers and their own employees. It was not always this way. When Cramer (1977) examined whether "clutch hitters" really existed, the analysis was based on only two years of Major League Baseball (MLB) batting average data. In a second study of the same research question, Grabiner (2006) used ten seasons of data. For a typical MLB game today, about seven terabytes of data are generated (Statcast 2020). The game data are available to all MLB teams, but how a team makes use of that sea of data is key. Increasingly, both large and small organizations must deal with large amounts of data as part of their daily business tasks, and effectively making use of that data is an organizational challenge.

The classic example of making a business decision without using data is the decision of Coca-Cola to replace the traditional recipe for company's flagship brand with "New Coke" (Schindler 1992). Using data from focus groups on how

many people preferred the original recipe and how strongly they felt about it (rather than trying to find a recipe that would beat Pepsi), would have given the company data on how their customers might react to the change. Netflix has extensive data on what its customers stream, and makes film production and distribution decisions based on user data (Hadida, Lampel, Walls, & Joshi 2020).

A critical skill for business students to learn is how to use data to answer questions and make management decisions based on the data (Ashraf 2020). It does not matter how much data there are, without data analysis skills to make sense of the data, it is all just numbers. When Akhtar, Frynas, Mellahi, and Ullah (2019) examined the links between big data skills, big data-driven actions, and business performance, they found that effective teams depended on multidisciplinary skills (e.g., computing, mathematics, statistics, machine learning, and business domain knowledge), which lead to data-driven insights and actions that enhanced business performance. They concluded that for an organization to make use of big data requires effective talent management (recruiting, selecting, developing, motivating, and retaining employees who have data analysis skills) and creating an organizational culture that encourages making decisions based on data. Tabesh, Mousavidin, and Hasani (2019) describe the two challenges to be overcome for successful implementation of big data strategies as Technological (i.e., data analysis knowledge and skills) and Cultural (i.e., leadership and an organizational culture that supports data-based rather than intuition-based decision-making approaches).

Organizational behavior students need to learn which statistics can be used to answer different types of research questions. If the research question to be tested is about a difference between groups, a statistical test of difference is needed, such as t-Test or ANOVA. If the research question to be tested is about the relationship between variables, a statistical test of association is needed, such as correlation or regression. If the research question to be tested is about the association of a set of independent variables with a dependent variable, multiple regression is needed (rather than separate correlations of each independent variable with the dependent variable).

Illustrating the use of big data to answer organizational behavior questions, Wang and Cotton (2018) used a dataset of 15,837 MLB players from all 30 teams to predict team performance from players' strategic or support roles. They used MLB data because it is an industry similar to broadcasting, advertising, accounting, law, public relations, consulting, and software development in having selection practices primarily based on credentials and expertise, development primarily based on on-the-job learning, high turnover, and cross-employer career paths.

Baseball is big business, too; the MLB grossed $10.7 billion in revenue for the 2019 season (Brown 2019). MLB is in the entertainment business, producing a product for fans to watch in the stadium or on TV (broadcast or streaming). Coaches must make the same type of business decisions as managers in other

businesses, about how to manage their team and motivate players in an effort to win the championship, or at least put on a show that fans want to watch in numbers large enough to generate income by ticket sales, luxury boxes, TV contracts, and merchandise sales.

Experiential exercises are useful ways to teach students and have a long history. As long ago as 1976, Certo (1976, p. 113) described the use of experiential exercises as "a growing trend in management education," where individuals learn-by-doing, and are "actively engaged in performance of a goal oriented task and learn both from the performance itself and from evaluating that performance." More recently, Kolb and Kolb (2005, p. 194) suggested that to improve learning in higher education, the primary focus should be on "engaging students in a process that best enhances their learning – a process that includes feedback of the effectiveness of their learning efforts."

Experiential learning has been used for a wide variety of purposes, teaching organization behavior concepts such as group problem solving (Hedges, & Pedigo, 2002), diversity (Berger 2001; Blandon, & Barbuto 2005; Merta, Stringham & Ponterotto 1988), attribution theory (Paglis 2008), leadership (Reilly & Ehlinger 2007; Sronce & Arendt 2009), managing intergroup behavior (Hunsaker 2004), team building (Yanson & Mann 2013), change management (McDonald & Mansour-Cole 2000), conflict management (Anakwe & Purohit 2006), face-to-face versus computer-mediated communication (Mehra 2010), and critical thought in qualitative research (Hopkinson & Hogg 2004).

Experiential exercises are both effective ways for students to learn because it prepares them for tasks they will do, and effective ways for instructors to teach because they are an additional way to evaluate student learning. A student may be able to answer test questions about which statistical analysis should be used to answer a given research question or interpret the output of a statistical test, but with an experiential exercise the instructor can assess whether the student can answer a research question given only the research question and the data. On Bloom's Taxonomy of Learning Objectives (Bloom, Engelhart, Furst, Hill, & Krathwohl 1956), test questions allow the instructor to assess Knowledge (recognizing or remembering facts, terms, basic concepts, or answers without necessarily understanding what they mean) or Comprehension (demonstrating an understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating the main ideas). An experiential exercise allows the instructor to assess Application (using acquired knowledge – solving problems in new situations by applying acquired knowledge, facts, techniques, and rules) or Analysis (examining and breaking information into component parts, determining how the parts relate to one another, identifying motives or causes, making inferences, and finding evidence to support generalizations).

In this experiential exercise, a secondary data set is used which has the following characteristics: 1) Available online, so that both students and

instructors can easily access it, 2) At least 50 variables, so that students can be given different research questions to answer, 3) A mix of variable types, some continuous and some categorical, so that both tests for difference and tests of association can be done, and 4) Updated at least annually, so that different students can be given different data sets to analyze each year.

Although the exercise is designed to help students learn about using data to make management decisions, the instructor can use the exercise as the basis of further discussion about leadership, change management, team building, organizational culture, and other organizational behavior concepts.

2. The Exercise: Using Data to Make Management Decisions

The exercise gives students (or student teams) various scenarios and tasks them with making management decisions using Major League Baseball (MLB) data. Seven scenarios that can be given to students, and answers and feedback that can be given, are described below. For each scenario, there is the Excel output from the data analysis, conclusions that should be made in interpreting the output, and additional learning points. The advantage of using a secondary data set is that the instructor does not have to first collect the data and then enter the data into a file for students to access. The advantage of using sports data is that there is a new data set each year, and therefore different answers to the questions in the scenarios each year. Instructors may choose to create their own scenarios and questions; there are many variables in the data set. Suggestions for other scenarios and questions, and for variations in the exercise are given below.

The MLB data comes from Baseball Reference.com (Baseball Reference 2021), which describes itself as "The complete source for current and historical baseball players, teams, scores and leaders." It is a secondary data set of MLB data for multiple seasons. Under the link "Seasons" there are tables of data for Team Standard Batting (including Batting Average, Runs Scored, Hits, On Base Percentage, Slugging Percentage, and many more), Team Standard Pitching (including Earned Run Average, Strikouts, Runners Left on Base, and many more), Team Fielding (Errors, Double Plays Turned, Fielding Percentage, Defensive Efficiency, and many more), and Postseason (i.e., Playoff) data. Some of the variables are continuous or ratios (e.g., Batting Average, Earned Run Average), and some are counts (e.g., Hits, Runs). Category variables can be created from which Division (East or Central or West) teams were in, which League (American or National), or which teams played in the playoffs (Wild Card, Division Series, League Championship Series, World Series). There are separate data tables for American League, National League, and both leagues. For each data table there is a "Glossary" (i.e., a codebook) for each variable in that data table, defining the variable and how it is calculated. For example, in "Team Fielding" the glossary entry for the variable "Fld%" is: Fielding Percentage, (Putouts + Assists) / (Putouts + Assists + Errors). The data can be downloaded