PREPARING STUDENTS TO SOLVE REAL PROBLEMS WITH STATISTICS

Volker Kraft and Curt Hinrichs
JMP Statistical Discovery LLC, United States
JMP Statistical Discovery LLC, Germany
volker.kraft@jmp.com

Students who master basic statistical methods are not automatically good problem solvers. This can be challenging when starting a career in industry. The journey from problem to solution requires a basic understanding of the typical workflow of data-driven decision making. A simple framework for the process of data discovery is a practical and useful addition to basic statistics education. This paper will mention some challenges students often experience when they move into industry along with a framework that connects the problem-solving elements from data access to data exploration and analysis to sharing results. It will also share freely accessible teaching resources such as the JMP Case Study Library that can help bring statistical problem solving into the classroom.

INTRODUCTION

The mission of JMP's academic program is to help bridge the gap between academia and industry and to build an awareness and an understanding of the importance of statistical thinking and data skills to better prepare students for careers in industry. In this context, "better preparation" means being able to solve real-world problems, such as those described in our Customer Stories (JMP, 2022c), by applying a sequence of analytical capabilities to transform raw data into shareable insights. Rather than simply applying a single statistical method to data of interest, we support a wider scope of learning objectives (Cao et al., 2013) to develop concepts and skills including:

- adopting an example driven, multistep solution path from data to insight;
- preparing and cleaning data and collecting new data from statistically designed experiments;
- exploring data through exploratory data analysis (Carver & Kraft, 2018);
- selecting the "best" method from many analysis options, e.g., for statistical modeling; and
- defining the problem and sharing solutions.

This paper presents the Case Study Library (JMP, 2022b), a main initiative of the JMP academic program, aimed at providing practical, real-world problem solving using JMP software (JMP, 2022d) in statistics classrooms. Driven by a real-world problem and dataset included in a chosen case study, the student should apply statistical thinking based on a flexible step-by-step analytic workflow (JMP, 2022a). The cases and the workflow support a wide range of courses, from introductory and second courses in statistics to more specific courses such as quality management, data science, design of experiments (DOE), or digital marketing. In the following sections we will present the case study library and analytic workflow in more detail, describe some sample cases, and discuss their impact in teaching.

METHOD

Starting with a problem at hand, students leverage a comprehensive set of analytical capabilities (Hinrichs et al., 2020). These capabilities are the core of the JMP Analytic Workflow (JMP, 2022a; see Figure 1) and can be seen as a "playboard" to move from raw, messy data to valuable insight. Before a solution is found, data must be accessed and imported, cleaned and organized, visualized and explored, and maybe passed through various machine learning algorithms. Learnings from each step must be interpreted and considered alongside practical knowledge. The flow is *not* static or predefined but dynamically chosen by the student, combining statistical thinking and pre-knowledge. Although a solution only needs a small subset of all available capabilities, the challenge is to apply the right methods in the right order. This is a perfect match to a real problem-solving scenario in industry and the opposite of practicing a single analytic method in isolation.
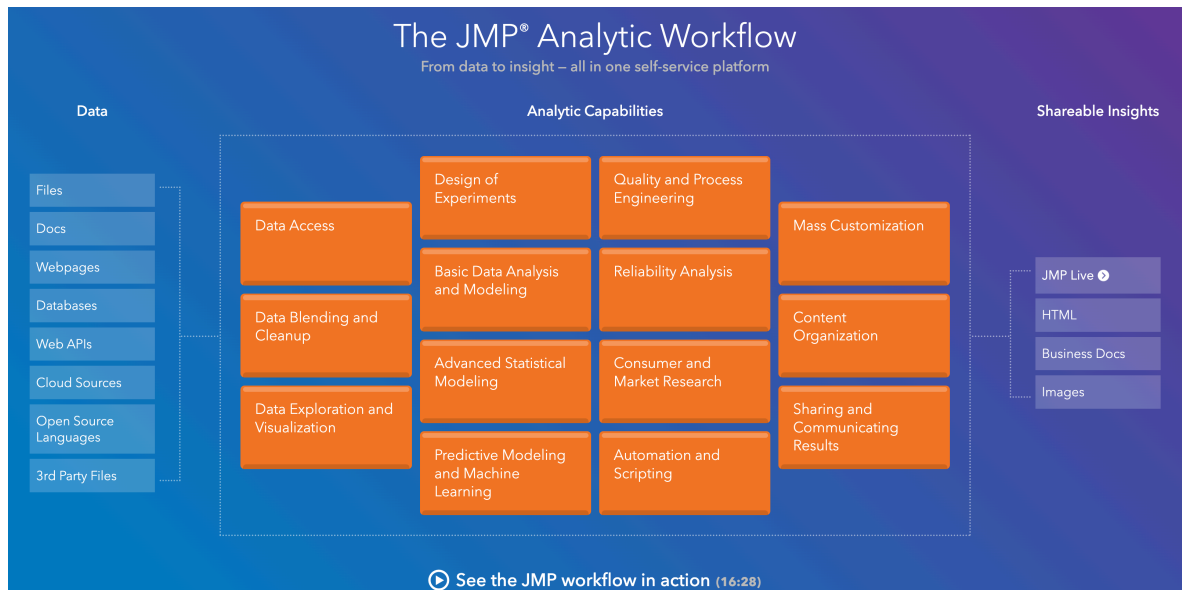
Figure 1. The JMP analytic workflow, a framework from data (left) to insight (right)

More than 50 workflow examples are shared for teaching use in the Case Study Library (JMP, 2022b; see Figure 2). Examples are listed by field, subject, statistical concepts, and complexity. Each case can be downloaded for free including dataset(s) (in .jmp format) and a PDF document with the full documentation. Many cases have been developed by practitioners in industry. Some companies are named in their cases, other cases are anonymized or just inspired by real-world examples.

Figure 2. Five examples from the Case Study Library (JMP, 2022b)

All case studies follow a consistent format and scenario. The case studies start with a description of the problem, followed by questions or tasks at hand, and then a description of the data. This is followed by a multistep illustration of the solution followed by a summary of the statistical and managerial insights. These are followed by exercises that expand or pose "what if" scenarios. Solutions are shared upon request.

Although educators can use the case studies for in-class demonstrations, the most effective uses are as homework assignments or group projects for students. The cases can stimulate discussions about which steps to take or comparisons of alternative solutions. Students can also be asked to present the learning outcome to "other decision makers" or to explain why and how certain methods have been applied.

EXAMPLES

All cases have in common that they are problem- or task-driven, and all solutions need a multistep workflow combining several analytical capabilities. This section summarizes three examples.

JMP020: Direct Mail (Regression and Forecasting)

Field: Marketing; Subject: Advertising; Complexity: Low

An antique dealer recently opened a shop. Direct mail of flyers and informational brochures is the primary advertising outlet. Direct mailings are sent out every Wednesday. The task is to investigate if sales are related to the direct mail campaign, and—if related—to determine the nature of the relationship.

Analyzing the data using simple linear regression and forecasting techniques shows that the return on direct mail is very high but with a time lag between mailings and sales. Students learn to take knowledge about the organizational behavior into account, which can be described here as a time-lagged relationship. They also learn how to use prediction intervals and confidence intervals to access the uncertainty of the predictive model.
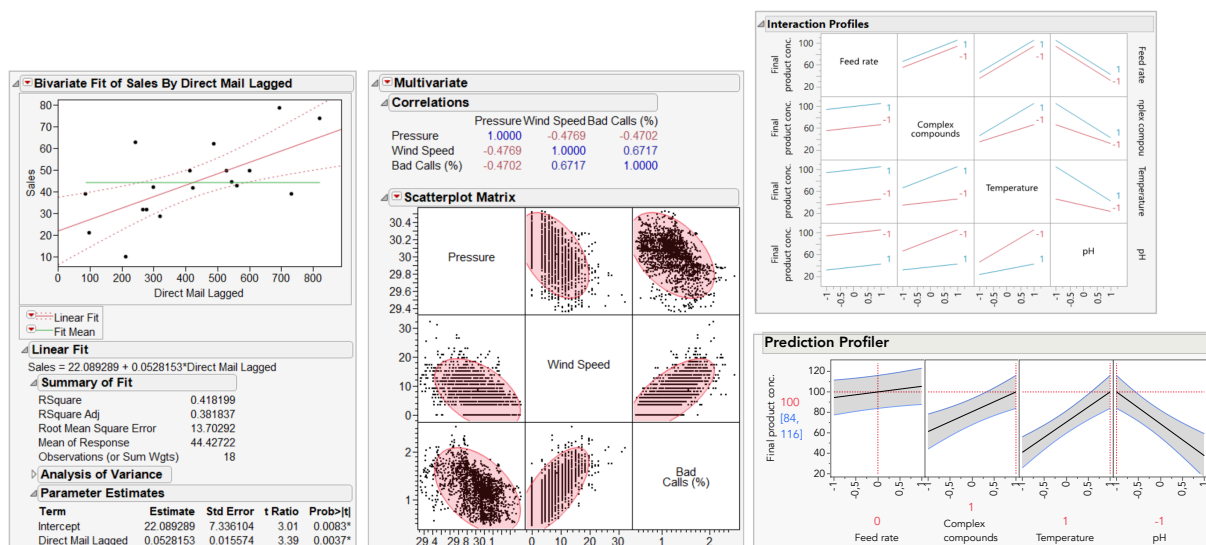

Figure 3. Sample JMP output for case studies #20 (left), #23 (middle) and #52 (right)

JMP023: Cell Phone Service (Multiple Regression—Two Predictors)

Field: Operations; Subject: Service Reliability; Complexity: Medium

Reliable Wireless, a U.S. mobile phone company, wants to improve the quality of their cell phone service. Knowledge of the impact of wind-induced fading on call performance should factor into overall decisions about capital expenditures on equipment and technology. Although they have no control over climatic conditions, students can make decisions about things such as data call capacity, number and configuration of base stations, band spectrums, and choice of wireless technologies.

Analysis of the data suggests that interrupted cell phone service is related, in part, to climate factors. Two predictors, wind speed and barometric pressure, are analyzed separately and both together

in regression models. Exploratory data analysis helps to understand distributions and the correlation between the two predictors, and relationships are explained based on simple and multiple regression models.

JMP052: Optimization of Microbial Cultivation Process (Design of Experiments and Predictive Modeling)

Field: Engineering; Subject: Biotech Manufacturing; Complexity: Medium

This case study, using Design of Experiments to optimize a microbial cultivation process, has been shared by Lonza, a Swiss multinational biotech and pharmaceutical company headquartered in Basel. A manufacturing team was tasked to speed development of a novel biomanufacturing process to help products reach market in less time. The team focused on increasing process yield because a poor yield translates to higher manufacturing costs.

By doing smarter experimentation using flexible and efficient designs, Lonza saved time and was able to extract maximum information from minimum resources. Finding the optimal settings helped to increase the yield and better meet the time-to-market expectations.

This case study addresses how to understand the causes for low yield and how to optimize the process by utilizing an optimal design for efficient data collection and informative analysis results. Students learn how to create and evaluate different designs, how to analyze experimental data by fitting a model and how to test for and interpret interaction effects. Model accuracy and diagnostics are analyzed, and the optimal process setting are determined.

DISCUSSION

JMP continues to work with leading practitioners in industry and educators from around the world to add more cases from other fields and to expand on other concepts in the future. Each new case adds another problem with an analytic workflow example to the library, readily available for teaching statistical thinking and problem solving—without any line of code.

Additional engaging case studies can be found in Cox et al. (2016). In addition to our case study library, educators also get access to other types of free course material: One-page guides from the JMP Learning Library (JMP, 2022e) provide quick-start info to use a specific method; teaching modules in JMP allow students to explore specific core concepts interactively; and the STIPS (JMP, 2022f) MOOC teaches the most relevant data skills in seven online modules for self-paced learning. Sharing real-world problems and solutions from industry for teaching use in higher education is just one way to stimulate collaboration in the JMP Academic-Industry Network.

REFERENCES

Cao, J., Kraft, V., & Stephens, M. (2013). *Statistical education standards and the role of technology* [PowerPoint slides]. Joint IASE/IAOS Satellite Conference, Macao. https://iase-web.org/documents/papers/sat2013/IASE_IAOS_2013_Paper_3.2.4_Kraft_et_al_ppt.pdf?1402525001

Carver, R., & Kraft, V. (2018). Keeping it real with data visualization. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)*. ISI/IASE. https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_3C2.pdf?1531364257

Cox, I., Gaudard, M.A., & Stephens, M.L. (2016). *Visual six sigma: Making data analysis lean* (2nd ed.). Wiley & Sons.

Hinrichs, C., Boiler, C., & Walsh, S. (2020). *JMP essentials: An illustrated guide for new users* (3rd ed.). SAS Institute.

JMP Statistical Discovery. (2022a). *JMP analytic workflow*. http://www.jmp.com/workflow

JMP Statistical Discovery. (2022b). *Case study library*. http://www.jmp.com/cases

JMP Statistical Discovery. (2022c). *Customer stories*. http://www.jmp.com/success

JMP Statistical Discovery. (2022d). *JMP data analysis software* (Version 16.2) [Computer software]. http://www.jmp.com/software

JMP Statistical Discovery. (2022e). *JMP learning library*. http://www.jmp.com/learn

JMP Statistical Discovery. (2022f). *Statistical Thinking for Industrial Problem Solving (STIPS)*. https://www.jmp.com/statisticalthinking