Maximizing Data Science Success with Information Quality (InfoQ) and JMP

Ron S. Kenett and Chris Gotwalt

Learning goals

- Why is data driven decision making accelerating at pace?
  Statistical analysis is focused on deriving findings from data. Information quality is expanding this view by considering the derivation of information from data. The consideration of information quality links decision making to specific goals, thereby accelerating the building of knowledge and data driven decision making.
- How do we ensure our data, analytics, interpretations, conclusions are fit for the intended purpose?
  Information quality requires a proactive identification of goal and utility. These two information quality components are key to ensure our data, analytics, interpretations, conclusions are fit for the intended purpose.
- How do we scale the use of data analytics to ensure our colleagues who may not be formally trained statisticians or data scientists can get the best from their data analytic endeavours?
  The trick is to move the discussion from data and statistical analysis to information quality. This is not a statistical topic per se so that the "colleagues" you refer to have equal standing in the argumentation. This is the essence of democratisation of statistics, in a wide sense. The approach requires a spirit of teamwork. Another benefit of focusing on information quality.

Background

Data analysis, from designed experiments, generalized regression to machine learning, is being deployed at an accelerating rate. At the same time, concerns with reproducibility of findings and flawed p-value interpretation indicate that well intentioned statistical analyses can lead to mistaken conclusions and bad decisions. For a data analysis project to properly fulfill its goals, one must assess the scope and strength of the conclusions derived from the data and tools available. This focus on statistical strategy requires a framework that isolates the components of the project: the goals, data collection procedure, data properties, the analysis provided, etc. The InfoQ Framework provides structured procedures for making this assessment. Moreover, it is easy to operationalize InfoQ in JMP. In this presentation, we give an overview of InfoQ, and use case studies drawing from consumer research and pharmaceutical manufacturing to illustrate how JMP can be used to make an InfoQ assessment, highlighting situations of both high and low InfoQ. We also give tips showing how JMP can be used to increase information quality, by enhanced study design, without necessarily acquiring more data. The talk is aimed at statisticians, machine learning experts and data scientists whose job it is to turn numbers into information.