The Evolution of AP® Statistics: How Big Data and Machine Learning Are Changing the Course

2008-2019 Timeline

- **2008**
  - Typical Data File (9 observations, 4 variables)
  - Data & Technology Project
  - CATALYSTS FOR CHANGE
    - Workshops (Darren Starnes and Floyd Bullard)
    - Emphasized the important role of simulations to anchor key statistical concepts
  - CURRICULUM CONTENT
    - Introduced simulations into curriculum across curricular elements
  - PROJECT
    - 2008

- **2009**
  - Industry Initiatives for Math and Science Education (Summer Internship)
  - Developed unit to introduce students to JMP – Penn State dataset: 227 observations, 8 variables
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Introduced Year-long statistics project – Literature review and analysis activity in the fall; data gathering and data analysis in the spring
  - PROJECT
    - 2009

- **2010**
  - Sabbatical Research on Statistical Literacy
  - Big data, importance of statistical literacy, introduced JMP more formally to students
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Conducted a review of statistical software; R was widely used but had too steep a learning curve; Excel did not support the range of statistical analysis activities; JMP chosen as the best tool based on functionality, ease of use, and cost
  - PROJECT
    - 2010

- **2011**
  - Statistics Integration Grant
  - Outgrowth of sabbatical research, impact on student preparation for AP stats; increased knowledge of descriptive statistics in grades 6-10
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Revised teaching of Statistical inference; used a holistic approach and encouraged students to look at the similarities across both tests for proportions and tests for means; Student understanding of the tests improved by focusing on common elements and difference between tests
  - PROJECT
    - 2011

- **2012**
  - Introduce Census at School data to teach basic data analysis skills (838 observations, 60 variables)
  - Most projects were student surveys or experiments; a couple of student did analysis projects with outside organizations
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Less time on descriptive statistics – students were better prepared as a result of greater statistical literacy
  - PROJECT
    - 2012

- **2013**
  - Increasing availability of public data; Students (with support) can begin to combine data from different sources
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Provided additional scaffolding for project; reviewed drafts of most major elements before final submission; Not sustainable with larger student population
  - PROJECT
    - 2013

- **2014**
  - Research module on machine learning
  - Realized the central role of linear and logistic regression in machine learning
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Introduced FARS (Fatality Analysis and Reporting System) data for student analysis; Data on 5 years of traffic fatalities in the US; all students required to work with this dataset (378,000 observations, 24 variables); Students to determine own research question and use JMP to analyze; insufficient teacher support for JMP
  - PROJECT
    - 2014

- **2015**
  - Elective course on Artificial Intelligence
  - Many classifiers require knowledge of statistical techniques
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Added data analysis unit that focused on teaching students how to use JMP for data analysis; unit culminates with analysis of FARS data
  - PROJECT
    - 2015

- **2016**
  - Fatality Analysis: Number of Accidents Involving Drunk Driving per 100 accidents For Holidays and Non-Holidays Over Time
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Work with data in the cloud?
  - PROJECT
    - 2016

- **2017**
  - Big data and machine learning will influence future course content
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Continue work with large datasets (>10,000 observations) requires more critical analysis of results; introduce effect size calculation?
  - PROJECT
    - 2017

- **2018**
  - Focus on understanding the data
  - Data & Technology Project
  - CURRICULUM CONTENT
    - Projects promote authentic learning and provide critical analysis skills
  - PROJECT
    - 2018

- **2019**
  - Emphasize working with data subsets and validating results on different subsets
  - Data & Technology Project
  - CURRICULUM CONTENT
    - More future projects will be data analysis projects
  - PROJECT
    - 2019

FUTURE DIRECTIONS

- **CURRICULUM CONTENT**
  - Big data and machine learning will influence future course content
  - Supervised Learning classifiers are key – linear regression (multivariate) is common for prediction problems and logistic regression is common for classification problems; students should understand the basics of both of these approaches
  - Computers do the mechanics of all statistics tests; Should the future direction focus more on the value humans can add to the tests:
    - » Understand your data
    - » Constructing hypotheses
    - » Checking conditions
    - » Interpreting results
  - Inference with large datasets (>10,000 observations) requires more critical analysis of results; introduce effect size calculation?

- **TECHNOLOGY AND DATA**
  - Continue work with large datasets (50,000+ observations, many variables)
  - Focus on understanding the data
  - Emphasize working with data subsets and validating results on different subsets
  - Work with data in the cloud?

- **PROJECTS**
  - Projects promote authentic learning and provide critical analysis skills
  - Most future projects will be data analysis projects
  - Having a project stakeholder who cares about results is valuable; need a sustainable model for doing this
  - Students still want to take data at face value without considering how the data was produced or generated; how do we change this attitude?
  - Projects are time intensive to review; need to consider models that are scalable