Undergraduate Data Science Statistics Pathways: What is Needed for Entry to the Major?

Background
The Dana Center Mathematics Pathways initiative ensures that all students in our nation have equitable access to, and the opportunity for success in, rigorous statistics and mathematics pathways that are aligned and relevant to their future aspirations, propelling them to upward economic and social mobility.

The explosion of data science programs and the demand for data science professionals has created opportunities for students at all levels (from 2-year colleges to graduate programs). In response, the Dana Center is exploring possible data science pathways.

What might a pathway for data science look like at Two-Year Colleges?
We see three options:

Option 1: Two-year degree designed for those who plan to enter the workforce
Option 2: Two-year degree designed for those who plan to transfer to a four-year data science program
Option 3: Certificate programs designed for those already in the workforce or for those with degrees in other areas that want to move in a data science direction

In this session, we are focusing on two-year to work programs.

Mathematics Foundations: Recommendations for Small Group Discussion 1

Recommendations from the NAS Report:
- Set theory and basic logic
- Multivariate thinking via functions and graphical displays
- Basic probability theory and randomness
- Matrices and basic linear algebra
- Networks and graph theory
- Optimization

Recommendations from the AMS Notices paper:
- **Mathematical Foundations I (discrete)**
  - Linear algebra
  - Counting principles
  - Graph theory
  - Simulation
- **Mathematical Foundations II (continuous)**
  - Functions and basic mathematical logic
  - Enough calculus to understand the ideas of partial derivatives
  - Taylor expansion method of approximating functions
  - Probability as area/integration
  - Multivariate thinking (functions, optimization, integration)
Two Examples for Small Group Discussion 2:

Duke University: Data Science Math Skills
- Set theory, including Venn diagrams
- Properties of the real number line
- Interval notation and algebra with inequalities
- Uses for summation and sigma notation
- Math on the Cartesian (x, y) plane, slope and distance formulas
- Graphing and describing functions and their inverses on the x-y plane
- Concept of instantaneous rate of change and tangent lines to a curve
- Exponents, logarithms, and the natural log function
- Probability theory, including Bayes’ theorem

James Cook University: Essential Maths for Data Science
- Logic
- Sets
- Proof
- Functions
- Relations
- Graphs and trees
- Linear systems and matrix formulation
- Vector spaces
- Eigenvalues/eigenvectors
- Singular value decomposition
- Optimization and numerical methods.

Resources for those who are planning to implement a data science pathway:
http://nas.edu/envisioningds (Data Science for Undergraduates: Opportunities and Options consensus study: The National Academies of Sciences, Engineering, and Medicine set forth a vision for the emerging discipline of data science at the undergraduate level)
https://www.amstat.org/ASA/Education/Two-Year-College-Data-Science-Summit.aspx (A diverse group of participants met during the summer of 2018 to make recommendations for two-year college data science programs, keeping in mind the needs of each of three student populations: 1) those seeking employment following an associate’s degree, 2) those seeking transfer to four-year programs, and 3) those seeking certificate programs and college-level courses in data science for professional development)
http://www.utdanacenter.org/higher-education/dcmp (Dana Center Mathematics Pathways (DCMP) ensure that all students in our nation have equitable access to, and the opportunity for success in, rigorous statistics and mathematics pathways that are aligned and relevant to their future aspirations, propelling them to upward economic and social mobility)

Contact Information
Rebecca Hartzler (rebeccahartzler@austin.utexas.edu)
Roxy Peck (rpeck@calpoly.edu)
Data Science Statistics Pathways: What is Needed for Entry into the Major?

Rebecca Hartzler, rebeccahartzler@austin.utexas.edu

Roxy Peck, rpeck@calpoly.edu

Friday November 16, 2018
Three Potential Types of Data Science Programs at Two-Year Colleges

1. Two-year degree designed for those who plan to enter the workforce
2. Two-year degree designed for those who plan to transfer to a four-year data science program
3. Certificate programs designed for those already in the workforce or for those with degrees in other areas that want to move in a data science direction
Certificate Programs

**Diverse in content and focus**

- Data management, cyber security, business analytics are just a few examples.
- Usually tailored to local and regional workforce needs.
Two-Year to Work Data Science Programs

• Goal is to produce “middle-skills” data practitioners. See Oceans of Data Project http://oceansofdata.org/sites/oceansofdata.org/files/FINAL%20Data%20Practitioner%20Profile-September%202016.pdf

• Part of a data science team, capable of data acquisition, data cleaning, data base management, combining data from different sources, producing simple data visualizations, basic data analysis

• Foundation that would allow for future career development in data science
Many challenges

- No consensus among four-year programs on curriculum content
- No consensus on mix and balance of foundations in mathematics, statistics, computer science
- Articulation can be difficult

- Recommend co-development with partner four-year university if possible
Focus of this Session

**Two-year to work programs**

- What are the mathematics and statistics foundations?
- What might a Mathematics for Data Science course look like?
- How to apply the learnings of comprehensive pathway design
  - Placement
  - Support for underprepared students
  - Broadening participation
Mathematics and Statistics Foundations

• *Data Science for Undergraduates: Opportunities and Options*, The National Academies of Science

• Report from the Two-Year College Data Science Summit
  Coming soon to [www.amstat.org/ASA/Education/Two-Year-College-Data-Science-Summit.aspx](www.amstat.org/ASA/Education/Two-Year-College-Data-Science-Summit.aspx)

• *Ensuring that Mathematics is Relevant in a World of Data Science*, Notices of the AMS, 2017
Primarily addresses four-year programs, but does recognize the need for programs at two-year institutions.

Includes a section titled “Two-Year Degrees and Certificates”

Although curriculum recommendations are for four-year programs, they can also be of help in the development of four-year programs.

Recommendations for mathematics and statistics foundations.
Mathematical Foundations
  o Set theory and basic logic
  o Multivariate thinking via functions and graphical displays
  o Basic probability theory and randomness
  o Matrices and basic linear algebra
  o Networks and graph theory
  o Optimization

  • Recommends development of “math for data science” course rather than requiring multiple semesters of course work.
Statistical Foundations

- Variability, uncertainty, sampling error and inference
- Multivariate thinking
- Nonsampling error, experimental design, bias, confounding
- Exploratory data analysis
- Statistical modeling and model assessment
- Simulations and experiments
Includes learning outcomes for each of the three types of programs, including two-year to work

Statistics learning outcomes include:

- Determine if conclusions are appropriate for a study based on study design (observational or controlled experiment), including identifying potential confounding factors and appropriate controls.
- Produce and interpret data visualizations, including dashboards, graphs and charts to describe and explore data and communicate findings.
- Produce and interpret numerical summaries to describe and explore data.
- Produce and interpret confidence intervals.
- Formulate statistical claims in terms of null and alternative hypotheses, carry out and interpret basic hypothesis tests.
- Investigate and explore relationships between more than two variables.
- Use statistical models to describe relationships between variables.
- Fit, interpret and evaluate basic statistical models (e.g. linear, logistic, exponential).
- Explain the purpose of cross-validation
Helpful in clarifying statistical foundations, not as helpful for the mathematics foundations.

“Programs offering an Associates degree may be required to include mathematics required to transfer to a four-year program, regardless of whether that is the primary purpose of the program. While mathematical preparation for a generic STEM major may suffice, such preparation might also be more than is necessary for data science, as noted in the PCMI report. Instead, mathematics courses for data science programs should emphasize applications and take special care to make explicit the ties between theory and application.”
## Report from the Two-Year College Data Science Summit

<table>
<thead>
<tr>
<th>Topic</th>
<th>To Work</th>
<th>Transfer</th>
<th>Certificate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Calculus</td>
<td>not required</td>
<td>working knowledge/mastery</td>
<td>not required</td>
</tr>
<tr>
<td>B. Matrices and basic linear algebra</td>
<td>exposure</td>
<td>exposure</td>
<td>not required</td>
</tr>
<tr>
<td>C. Basic probability</td>
<td>working knowledge</td>
<td>working knowledge</td>
<td>not required</td>
</tr>
</tbody>
</table>
Hardin and Horton propose two courses in their AMS Notices paper:

- Mathematical Foundations I: Discrete Mathematics
- Mathematical Foundations II: Continuous Mathematics

Note: This recommendation is for four-year programs.
Ensuring that Mathematics is Relevant in a World of Data Science, Notices of the AMS (2017)

Mathematical Foundations I (discrete)
- Linear algebra
- Counting principles
- Graph theory
- Simulation

Mathematical Foundations II (continuous)
- Functions and basic mathematical logic
- Enough calculus to understand the ideas of partial derivatives
- Taylor expansion method of approximating functions
- Probability as area/integration
- Multivariate thinking (functions, optimization, integration)
Small group discussion

- Envision a one-semester course titled Mathematics for Data Science that would be appropriate for a two-year to work data science program at a community college.
- Based on the NAS recommendations and the two courses proposed by Hardin and Horton, what topics would you include in your course?
- What topics would definitely be left out?

So what do you think??
Two Examples of Math for Data Science Courses

1. Duke University (online through Coursera)
   Data Science Math Skills
   https://www.coursera.org/learn/datasciencemathskills#syllabus

2. James Cook University
   Essential Maths for Data Science
Course description

- Set theory, including Venn diagrams
- Properties of the real number line
- Interval notation and algebra with inequalities
- Uses for summation and sigma notation
- Math on the Cartesian (x, y) plane, slope and distance formulas
- Graphing and describing functions and their inverses on the x-y plane
- Concept of instantaneous rate of change and tangent lines to a curve
- Exponents, logarithms, and the natural log function
- Probability theory, including Bayes’ theorem
Course description

Data science is grounded in mathematics. This subject will provide you with the essential elements of mathematics required for data scientists. It includes elements of discrete mathematics including logics, sets, proof, functions, relations, graphs and trees; as well as elements of linear algebra including linear systems and matrix formulation, vector spaces, eigenvalues/eigenvectors, singular value decomposition, optimization and numerical methods.
What do you think??

**Small group discussion**

These two courses were not necessarily designed for a two-year to work program.

- What do you think of these two courses?
- Is one more appropriate for a two-year to work program than the other?
- Given your earlier discussion of what a math for data science course might look like and these two sample course descriptions, are there topics that you would now want to add to your course? That you would now want to leave out?
Considerations for Pathways for Data Science

Dana Center Mathematics Pathways: ensures that all students in our nation have equitable access to, and the opportunity for success in, **rigorous** statistics and mathematics pathways that are aligned and relevant to their future aspirations, propelling them to upward economic and social mobility.

How do we create entry pathways for Data Science and structure them for student success?
Comprehensive Design of Math Pathways

Core elements:

• Aligned math pathways with default or recommended math requirements
• Meta-majors with default or recommended math requirements
• Multiple measures placement
• Enhanced advising for those students still deemed underprepared
• Co-requisite supports for underprepared students
• Too many students are placed in developmental education who could thrive at college level
  o Better assessment systems are needed
• Research shows that HS GPA is the best predictor of success in college math & English
• Other measures may be needed to better place students from underrepresented groups
• Several models of MMA are being studied including algorithm models, decision rule models, etc
Percentage of Students Placed in College-Level Courses, by Placement System, 2017 Pilot at Ten Colleges

<table>
<thead>
<tr>
<th>Course</th>
<th>Test only</th>
<th>Multiple measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>College-level math</td>
<td>29%</td>
<td>56%</td>
</tr>
<tr>
<td>College-level English</td>
<td>57%</td>
<td>74%</td>
</tr>
</tbody>
</table>

SOURCES: Placement data collected from the ten Minnesota and Wisconsin colleges participating in the pilot.

NOTES: Most colleges ran their pilots for a week or so in June or July, with the exception of one Wisconsin college that continued its pilot for several months, yielding much higher sample sizes. Colleges are equally weighted in the averages displayed.
What is Co-Requisite Instruction?

• Co-requisite instruction allows underprepared students to enroll in a college-level course with added support designed to help them succeed.
• There are several successful design structures and modes of delivery
• Resource: *Key Considerations in Designing Co-Requisite Supports*, Emerging Issues in Mathematics Pathways (2018)
Multiple Statewide Examples

• **Georgia**
  - Traditional: 20% success rate in two years
  - Co-requisite: 63% success rates

• **West Virginia**
  - Traditional: 14% success rate in two years
  - Co-requisite: 62% success rates

• **Indiana**
  - Co-requisite courses in math, along with new quantitative reasoning and technical math pathways
  - Traditional: 29% success rate in three years
  - With changes: 64% success rates

http://completecollege.org/spanningthedivide/#the-bridge-builders
Tennessee Community Colleges
Gateway Math Success in One Year

Tennessee Board of Regents Brief #3: Co-requisite Remediation Full Implementation 2015-16

ACT Math
- Prerequisite Model 2012-13 Cohort
- Co-requisite Full Implementation AY 2015-16

Tennessee Board of Regents Brief #3: Co-requisite Remediation Full Implementation 2015-16
Equity

Results of TBR Co-requisite Mathematics Full Implementation - Minority Students

- Pre-requisite Model AY 2012-13
- Co-requisite Pilots: 2014-15 Cohort
- Full Implementation: Fall 2015

Tennessee Board of Regents Brief #3: Co-requisite Remediation Full Implementation 2015-16
City University of New York (CUNY) Study

Randomized Controlled Trial

Enrollment Status After 3 Years

<table>
<thead>
<tr>
<th></th>
<th>Not Enrolled</th>
<th>Enrolled</th>
<th>Graduated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elem Alg N=297</td>
<td>17.1%</td>
<td>30.0%</td>
<td>52.9%</td>
</tr>
<tr>
<td>Elem Alg w/WS N=313</td>
<td>19.4%</td>
<td>25.0%</td>
<td>55.6%</td>
</tr>
<tr>
<td>Stat w/ WS N=297</td>
<td>25.2%</td>
<td>26.7%</td>
<td>48.1%</td>
</tr>
</tbody>
</table>

Logue, Watanabe-Rose, & Douglas, randomized control trial conducted Fall 2013
Finding 4.1: The nature of data science is such that it offers multiple pathways for students of different backgrounds to engage at levels ranging from basic to expert.

Finding 4.2: Data science would particularly benefit from broad participation by underrepresented minorities because of the many applications to problems of interest to diverse populations.

Recommendation 4.1: As data science programs develop, they should focus on attracting students with varied backgrounds and degrees of preparation and preparing them for success in a variety of careers.
What do you think?

Small Group Discussion

Consider the access and diversity goals at your institution – In what ways might a Data Science Pathway have the potential to broaden participation?
About the Dana Center

The Charles A. Dana Center at The University of Texas at Austin works with our nation’s education systems to ensure that every student leaves school prepared for success in postsecondary education and the contemporary workplace.

Our work, based on research and two decades of experience, focuses on K–16 mathematics and science education with an emphasis on strategies for improving student engagement, motivation, persistence, and achievement.

We develop innovative curricula, tools, protocols, and instructional supports and deliver powerful instructional and leadership development.