BULLET BACKGROUND PAPER

CORE STATISTICS AT USAFA

This point paper summarizes curricular design information and ideas for core statistics at USAFA. It proposes courses of action (COA) for DFMS.

- Background
    -- Core statistics refers to Math 300 as it is currently described in the curriculum handbook. Core substitutes Math 356 and Math 377 are not addressed in this paper.
    -- The current core curriculum design goal is to produce "consumers of data" and not "producers". This may have been appropriate 20 years ago, but it is antiquated and out of line with current Air Force and Space Force needs. We need better data acumen in all our graduates.
    -- Math 300 is typically taught and directed by junior and experienced instructors where some do not have a great deal of experience teaching statistics. The course must be designed for them to be successful in teaching the material.
    -- Math 300 has a natural stable state of being a traditional small sample normal-based inference class with a higher emphasis on probability and inference. Excel has been the software of choice for a number of years. This is also out of date as we need to move towards modern software, multivariate problems, and computational versus mathematical engines for problem solutions. This needs to become the new stable state.
    -- Our current statistics curriculum relies too heavily on mathematics. From George Cobb[1] "For the last half-century, mathematics has been hovering like a helicopter parent, making it hard for our Stat 101 course to go out and play with its curricular friends. In our current day-care center the central limit theorem is bully. The goal is to get students to tests and intervals based on the normal approximation. Moreover, the object of those inferences is always the center of some probability distribution, either the mean or a binomial proportion."

- Data Science in core statistics
    -- Mathematics no longer needs to be the main engine for the statistical analysis cycle. Computation and data should be at the heart of our core statistics course.
    -- Students should have a capability to interact and understand our multivariate work space.
    -- Students should be practitioners of the complete data analysis cycle.
    -- Students should be immersed in a modern open-source computational platform.

- Learning Outcome

    Students will be able to complete the entire data analysis cycle to include data collection, data wrangling, data visualization, data modeling, inference, prediction, interpretation of results, and communication using reproducible research.

- Recommendations
    -- Core calculus should be linked with and feed into core statistics.
    -- Use R or python as the computing environment. (Suggestion: Use R because 141/142 uses it)
    -- Use one open source textbook and follow it. (Suggestion: Use the book Modern Dive – Statistical Inference via Data Science)

[1] Cobb, G. *The American Statistician.* 69(4):266-282. 2015.

Professor Bradley Warner/USAFA/DFMS/333-1935/baw/18 May 22

-- Keep the number of hypothesis tests and probability distributions to a minimum.
-- Less lecture and more scaffolded work in class using R.
-- Instead of one big project, use problem sets to get feedback to students faster, provide faculty with assessment of student progress, and reduce end of the semester grading load.
-- Maybe have a case study at the end of the semester to tie all the ideas together.
-- Change curriculum handbook course description from:

Math 300. Introduction to Statistics. 3(1). An introduction in probability and statistics for decision-makers. Topics include basic probability, statistical inference, prediction, data visualization, and data management. This course emphasizes critical thinking among decision-makers, preparing future officers to be critical consumers of data. Math 300 is designed primarily for majors in the Social Sciences and Humanities.

to:

Math 300. Introduction to Statistics. 3(1). An introduction to data analysis for decision-makers. Students will be able to complete the complete data analysis cycle to include data collection, data wrangling, data visualization, data modeling, inference, prediction, interpretation of results, and communication using reproducible research. Math 300 is designed primarily for majors in the Social Sciences and Humanities.

- COAs

1. Starting in the fall of 2022, we use the book *Modern Dive – Statistical Inference via Data Science*. A draft syllabus is attached. Homework is only suggested for the first block to give an idea of how it would be implemented. Professor Kaplan, Coors Endowed Chair, will work on revising the book to add more elements that an AF/SF decision maker would need to know. His proposal will be complete in early October when a decision by the Academic Staff of DFMS will be made on using it in the spring of 2023.
2. Adopt the Modern Dive book and use it all AY 2022-2023. Professor Kaplan will develop and design a course based on the Modern Dive book to start using in the fall of 2023.
3. Professor Kaplan will develop the materials to start using in the fall of 2022. He would need to be done by the end of June.
4. Continue with the current offering of Math 300 and have an action item for the Core Directorate to have a proposal on the way ahead with Math 300 in place by May 2023.

In all COAs, Professor Kaplan is to develop a proposal on how all three core math courses, 141/142/300, complement and build on each other.

MATH 300 - Introduction to Statistics

Lesson	Topic	Reading	Homework	Graded Events
Data with Tidyverse				
1	Data with R	1	LC 1.1 - 1.7	
2	Scatterplots	2 - 2.3	LC 2.1 - 2.8	
3	Linegraphs, Histograms, and Facets	2.4 - 2.6	LC 2.9 -2.21	
4	Boxplots and Barcharts	2.7 - 2.9	LC 2.22 - 2.37	
5	filter and summarize	3 - 3.3	LC 3.1 - 3.4	PS 1
6	group_by, mutate, and arrange	3.4 - 3.6	LC 3.5 - 3.12	
7	join, select, rename, and top_n	3.7 - 3.9	LC 3.13 - 3.20	
8	Importing Data	4 - 4.2	LC 4.1 - 4.3	
9	Case Study/ Review	4.3 - 4.5	LC 4.4 - 4.5	PS 2
10	GR 1 (Chapters 1-4)			GR 1
Regression				
11	SLR - Continuous x	5 - 5.1	LC 5.1 - 5.3	
12	SLR - Discrete x	5.2	LC 5.4 - 5.7	
13	SLR Summary	5.3 - 5.4	LC 5.8	
14	Multiple Regrssion	6 - 6.1	LC 6.1	PS 3
15	Multiple Regression - Interaction	6.2	LC 6.2 - 6.3	
16	Multiple Regression - Two numerical	6.3		
17	Multiple Regression Conclusion/Review	6.4		PS 4
18	GR 2 (Chapters 5-6)			GR 2
Statistical Inference - Confidence Intervals				
19	Sampling	7 - 7.2		
20	Case Study	7.3 - 7.4		
21	CLT, Normal Distribution	A.2, 7.5 - 7.6		
22	Bootstrap Intro	8 - 8.2		PS 5
23	Understanding CI	8.3		
24	Copnstructing CI	8.4		
25	Interpreting CI	8.5		
26	Case Study	8.6		PS 6
27	GR 3 (Chapters 7-8)			GR 3
Statistical Inference - Hypothesis Testing				
28	Permutation Tests	9 - 9.1		
29	Conducting Hypothesis Tests	9.2 - 9.3		
30	Interpreting Hypothesis Tests	9.4		
31	Case Study	9.5		
32	More on Hypothesis Tests	9.6		
33	Interpreting Regression Tables	10 - 10.2		PS 7
34	Regression Assumptions	10.3		
35	Regression Inference via Simulation	10.4		
36	Summary	10.5		PS 8
37	GR 4 (Chapters 9-10)			GR 4
38	Intro to more Inference Examples	B		
39	Case Study	11.2		
40	Case Study	11.3		