UNITED STATES MILITARY ACADEMY
WEST POINT®

Educating Statistically Literate Army Leaders

LTC Nicholas Clark

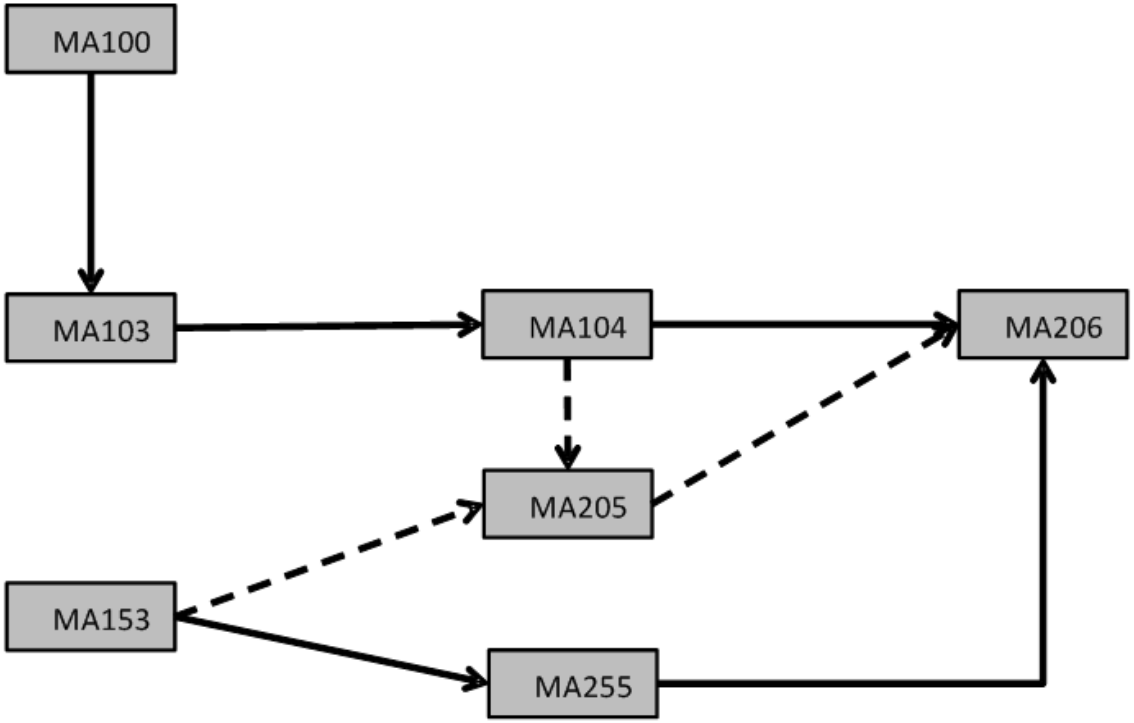
Nicholas.clark@westpoint.edu

UNITED STATES MILITARY ACADEMY
WEST POINT

West Point Academic Goals

- Communicate Effectively
- Think Critically and Creatively
- Demonstrate the Capability and Desire to Pursue Intellectual Development
- Apply Ethical Perspective and Concepts in Decision Making
- Apply STEM Concepts and Processes
- Apply Concepts from Humanities and Social Science
- Integrate and Apply Knowledge and Methodological Approaches Gained Through Academic Discipline

Core Overview



MA206: Probability and Statistics

- **Reasoning Under Uncertainty**
- **Fundamental Probability and Statistics Concepts**
- **Data Analysis, Probabilistic Models, Simulation, Random Variables**
- **Hypothesis Testing, Confidence Intervals, Linear Regression**
- **Uses R/Rstudio primarily**

Capstone Data Analysis Project

- Regression Focused, Students Choose Topics

Does Money Buy Happiness?

Abstract

In this paper, we build a multiple regression model to investigate what factors influence national happiness. After finding significant variables such as money, health, inequality, and human freedom, we create an initial model and use a variant of purposeful selection to systematically trim insignificant interactions. The resulting model is not only significant but also predicts national happiness well, with an adjusted $R2$ value of .7891. It turns out money has a huge effect on national happiness.

$$\hat{y} = 13.21 + 1.13log(a) - .29b - .82c - .07d + .03bc + .004bd - .02cd$$

a = GDP per capita; b = Life expectancy; c = Human freedom; d = Inequality

At first glance, the model looks nonsensical since there are negative coefficients in front of life expectancy and human freedom, implying a reduction will increase happiness. This clearly contradicts our data. However, the model still makes sense when you take into account all of the interactions. Additionally, another confusing part of the model is its initial intercept of 13.21. Since the happiness index has a maximum value of 10, it is impossible for a country to have a happiness index of 13.21. Consequently, it seems as if we should be careful when applying the model at extreme prediction values.

UNITED STATES MILITARY ACADEMY
WEST POINT

MA206Y vs MA206

MA206Y
Introduction to Data Analysis and Statistics

1. Use the investigative process in statistics to answer research questions of interest and be able to communicate results to general audiences.
2. Produce and interpret graphical displays and numerical summaries of data.
3. Recognize and explain the central role of variability and randomness in designing studies and drawing conclusions.
4. Employ appropriate single and multivariable statistical models and interpret and draw conclusions from standard output of statistical software packages.
5. Understand and use basic ideas of statistical inference (hypothesis tests and confidence intervals) in a variety of settings.
6. Be critical consumers of statistically-based results and recognize whether reported results reasonably follow from the study and analysis conducted.

MA206
Introduction to Probability and Statistics

1. Apply the axioms, theorems, and basic properties of probability to quantify the likelihood of events.
2. Use discrete and continuous random variables to model phenomena and answer basic probability questions.
3. Produce and interpret graphical displays and numerical summaries of data.
4. Understand and use basic ideas of statistical inference (hypothesis tests and confidence intervals) in a variety of settings.
5. Employ linear regression models in the context of statistical analysis; interpret and draw conclusions from the standard output of statistical software packages.
6. Communicate results of statistical analyses to general audiences.
7. Apply probability and statistics concepts to a variety of engineering topics.

PURPLE – indicates course objectives not in the other course.
ORANGE – indicates common course objectives taught differently.

UNITED STATES MILITARY ACADEMY
WEST POINT

Applied Statistics and Data Science Major

-Applied statistics uses a variety of computational techniques, methods of visualizing and exploring data, methods of seeking and establishing structure and trends in data, and questioning and reasoning that *quantifies uncertainty*

-Data science encompass the *entire life cycle of data*, from its specification, gathering, and cleaning through its management and analysis, to its use in making decisions and setting policy

-The ASDS Major provides Cadets the opportunity to:

- -effectively explore structured and unstructured data
- -define answerable questions
- -perform statistical analysis
- -communicating results both written and orally

-The program provides an introduction to the underlying mathematics of Data Science and Applied Statistics while simultaneously offering an exposure to computation and optimization issues inherent in large and disparate data sets

UNITED STATES MILITARY ACADEMY
WEST POINT

ASDS Major - Objectives

- **SLO 1: Demonstrate competence in computational and statistical thinking**
  - Understand the basic statistical concepts of data analysis, data collection, modeling and inference
  - Formulate problems, plan data collection campaigns and analyze the data to provide insights
  - Demonstrate proficiency in foundational software skills and the associated algorithmic, computational problem solving strategies
- **SLO 2: Demonstrate competence in mathematical foundations**
  - Understand the underlying structure of common models used in statistical and machine learning as well as the issues of optimization and convergence of algorithms
- **SLO 3: Apply statistical model building and assessment techniques**
  - Be adapt at data visualization using visualization techniques to communicate with others and identify weaknesses in proposed models
  - Employ statistical inference and draw conclusions using formal modeling. Understand how data issues impact analysis and interpretation of statistical finding
- **SLO 4: Employ algorithmic problem-solving skills**
  - Define clear requirements to a problem, use efficient strategies to arrive at an algorithmic solution using a suitable high-level computer language
  - Leverage existing packages and tools to solve computational problems
- **SLO 5: Prepare and manage data through the entire problem solving process**
  - Work with a variety of sources and formats of data
  - Prepare the data for use with a variety of statistical methods and models
  - Ensure the integrity of the data throughout the entire analytical process
- **SLO 6: Transfer knowledge**
  - Communicate results both written and orally
  - Demonstrate understanding of ethical issues in reproducibility

Applied Stats and Data Science Curriculum

AST1 REQUIRED COURSES (7 of 7)

MA 376	Applied Statistics
MA 371	Linear Algebra
MA 476	Mathematical Statistics
MA 477	Theory and Application of Data Science
MA 478	Generalized Linear Models
MA 486	Mathematical Computation
MA 491	Research Seminar in Applied Math

COMPLIMENTARY SUPPORT COURSES (3 of 3)

CS393	Database Systems
CY355/305	Cyber Foundations or Cyber Foundations-Computing
SE 370	Computer Aided Systems Engineering

ELECTIVES (2 of 14)

MA 372	Introduction to Discrete Math
MA 383	Foundations of Math
MA 386	Introduction to Numerical Analysis
MA 387	Mathematical Analysis 1
MA 388	Sabermetrics
MA 391	Mathematical Modeling
MA 394	Fundamentals of Network Science
MA 461	Graph Theory and Networks
MA 462	Combinatorics
MA 488	Special Topics in Mathematics
MA 489	Advanced Individual Study in Math
CS 473	Computer Graphics
CS 486	Artificial Intelligence
IT 383	User Interface Development
XE 402	Integrative System Design II
XE 492	Disruptive Innovations

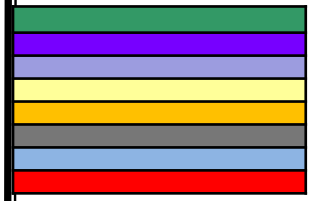
ENGINEERING SEQUENCE (3 of 3)

CY300	Programming Fundamentals
CY350	Network Engineering and Management
CY450	Cyber Security Engineering

SCIENCE DEPTH (1 of 3)

CH102	Chemistry II
CH275	Biology
PH206	Physics II

Fourth Class		Third Class		Second Class		First Class	
1st Term	2nd Term	1st Term	2nd Term	1st Term	2nd Term	1st Term	2nd Term
MA103 Math Modeling / Intro to Calc	MA104 Calculus I	MA205 Calculus II	MA206 Prob/Stats	MA376 Applied Statistics	MA476 Mathematical Statistics	ASDS Elective	MA490 Integrative Experience
CH101 Chem	PH205 Physics	Science Depth	EV203 Phys Geog	CY300 Programming Fundamental	CY350 Network Engineering and Management	CY450 Cyber Security Engineering	ASDS Capstone I
PL100 Psych	IT105 Info Tech I	MA371 Linear Algebra	PY201 Philosophy	CY355 Cyber Foundations-Computing	SE370 Computer Aided SE	MX400 Officership	LW403 Law
HI105 History	HI108 History	SS201 Econ	SS202 Pol Sci	HI302 Mil Art	PL300 Leadership	SS307 Int'l Rel	ASDS Elective
EN101 English Comp	EN102 Literature	LXXXX For Lang	LXXXX For Lang	MA486 Mathematical Computation	MA 477 Theory and Application of Data Science	CS393 Database Systems	MA 478 Generalized Linear Models



Core Program – STEM
STEM Depth
Science Depth
Major Required Courses
New Courses
Complementary Support Courses
Core Program – Humanities
Engineering Sequence

Observations/Conclusions

- Statistical thinking/reasoning is not natural to Cadets
- Struggle with distilling the “so what” from the output of a statistical model
- Courses involve more programming then “when I was a Cadet” – some are good at it, some are not
- In 400 level courses, challenge is mathematical maturity to string together topics from previous courses