Empirical Methods of Data Science

PROFESSOR MICHELLE LEVINE

WEEK 1: INTRODUCTION

1/23/19

Welcome!

Course enrollment:
- If currently enrolled, sign the attendance sheet
- If you are on the waitling list or want to be on it, please mark your name on the bottom of the second page

A bit about you, a bit about me

A bit about me…

◦ As an undergrad, studied psychology and computer science

◦ Decided to pursue a graduate degree in clinical psychology.

◦ First worked in clinical psychology,then went back to school to focus on cognitive psychology

◦ Worked at AT&T labs working on human factors experiments

◦ Barnard Department of Psychology, now at Columbia Department of Computer Science

Today

1. What is this course about?

2. Defining key terms for the course

3. Syllabus/Course requirements

COMS6998: Empirical Methods of Data Science

Main goal of the course is for students to understand how to conduct research.

Specifically, you will:
- Be familiar with a range of research methods and experimental designs
- Be able to critique prior research in order to build upon with further research
- Learn how to run common statistical measures such as correlations and ANOVAs
- Learn how to use R, a statistical tool for analyses and visualization
- Know how to write a research report and to give a scientific presentation

Experimental research

Why do we do research?

How do we conduct research?

What do we do with our findings?

The methods we will be studying, and we will be seeing in research, are based on methods rooted in psychological research.

Psychology

History of psychology = history of experimental psychology

The oldest of questions
being answered by the *newest of methods*

Scientific study of human mind (mental processes) and human behavior
◦ Scientific (for now) meaning the rigorous observation and measurement of behavior
◦ Studied from multiple levels/perspectives

What is the scientific method?

Rational empiricism

Empiricism: gaining knowledge through observation

◦ Empirical questions: questions that can be answered through systematic observations and experiences

Scientific method: rules and techniques of observation that minimize errors allowed by simple observation

Psychology subfields

Clinical

Social

Cognitive

Perception

Behavioral

Developmental

Neuropsychology

Educational

Evolutionary

Human Factors

Experimental

Psychology subfields

Clinical

Social

Cognitive

Perception

Behavioral

Developmental

Neuropsychology

Educational

Evolutionary

Human Factors

Experimental

Psycholinguistics
Computational cognitive science

Natural language processing

Summarization

Machine translation

Information extraction

Part of speech tagging

Sentiment analysis

Word sense disambiguation

Question answering

Spoken language processing

Speech recognition (ASR)

Speech synthesis (TTS)

Speech to speech translation

Speech search

Spoken dialogue systems

Paralinguistic detection (gender, age, language/dialect, emotion, personality, deception)

Data science & Big data

The emergence of data science and big data:

- Text
- Audio
- Video
- Social media

Syllabus & course overview

http://www.cs.columbia.edu/~mlevine/EMODS/

Important things to note

Schedule:
◦ subject to change
◦ announcements will be emailed through CourseWorks

Readings:
◦ Posted in CourseWorks under "Files"
◦ assigned for the week we will be discussing them
◦ ie Week 2 readings should be completed before the week 2 class meeting

Week 2 readings

Creswell, J.W. & Creswell, J.D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Chapter 1.

Johnston, D'Haro, Levine & Renger. (2007). A Multimodal Interface for Access to Content in the Home

Steele, K.M., Bass, K.E., & Crook, M.D. (1999). The Mystery of the Mozart Effect: Failure to Replicate.