SYLLABUS

Scientific Thinking and Reasoning

ANTH S018 SUMMER 2021
Tuesdays/Thursdays 1 – 4:15 pm

Eduardo Fernandez-Duque
Professor
Anthropology and School of the Environment

"Why is he thinking? Did he lose his smartphone?"

I am writing this syllabus in mid-November 2020 when it is still uncertain what the situation will be like in Summer 2021. On normal semesters, the syllabus of my courses includes the following as the opening sentence: "I firmly believe that education should be an interactive process that motivates you to actively seek information and to evaluate that information critically. I am looking forward to working WITH you so that you learn the most from participating in this course."

I have learned from teaching this course this past Fall 2020, that I can make the sentence remain true even under the conditions imposed by the pandemic this semester. Summer session may bring new challenges if we find ourselves sharing our learning environment virtually. But challenges are what you decided to seek when you decided to join our Yale community! We will face those challenges together. I am firmly committed to making all I can, to remain true to my philosophy that learning must be an active process.

Instructor

Eduardo Fernandez-Duque
Office: 10 Sachem St., Room 314
Phone: 215-432-3086
E-mail: eduardo.fernandez-duque@yale.edu
Websites:
https://owlmonkeyproject.wordpress.com/
https://fernandezduque.wordpress.com/
https://www.nationalgeographic.org/find-explorers/?filter=&page=1&q=fernandez-duque

Office Hours:

I am always available to meet you before and after class. Otherwise, you can ask me for an appointment or sign up through CANVAS for an appointment.

Brief Course Description:

Students read, discuss and reflect on the paramount importance of science and quantitative reasoning in their lives through an exploration of the basic elements of a quantitative scientific process of inquiry. Part I focuses on reading about truth, facts and skepticism, causality, inference, deductive and inductive reasoning, research questions, and formulation of hypotheses and predictions. Part II considers aspects related to the actual development and implementation of a scientific study including considerations of types of study (e.g, observational, experimental), study feasibility, sample size, selection and validity of variables, power analysis, confounding factors. Part III considers the analyses, interpretation and presentation of results, offering introductory explanations of *a priori* statistical protocols; predictive and/or explanatory power and interpretation of both statistical significance and research relevance. The course is neither a lecture or seminar, but instead each meeting is a hybrid of both formats; a format where students are required to be active participants in the process of learning.

Goal of the course

Given the importance of science in everyone's life, from medical reasons to emotional to financial ones, it is paramount that we develop the skills to critically evaluate the existing information that is presented to us "based on science". The goal of the course is for you to be able to understand some of the basic components of a process of scientific inquiry, a process that includes a substantial amount of quantitative reasoning (QR). You will develop critical quantitative thinking skills, and improve your ability to either embark in your own scientific research or be better prepared to evaluate the science that is presented to you.

Main Learning Objectives (see specific learning objectives for each topic in the weekly plans at the end of syllabus)

1- You will understand some of the basic components of a process of scientific inquiry.
2- You will learn to recognize the different sources of evidence in the literature (primary, secondary) and the potential contributions each makes to advancing a scientific discipline.
3- You will become familiar with the latest findings in the scientific study of the course topics.
4- You will be able to explain, and illustrate with examples, the process of scientific inquiry, including the formulation of hypotheses and predictions and the evaluation and interpretation of data.
5- You will be able to quantitatively describe and interpret, using tables, graphs or other forms of data summary, scientific information presented in professional articles and popular outlets.
6- You will be able to comfortably use key concepts in data sciences and statistics.

Class Attendance

You are expected to attend all meetings. Please, consider carefully the information provided in the Yale Summer Session Faculty Handbook, page 5: "Absences during the summer cannot be tolerated. Students normally take no more than two courses at any one time during the summer. This is considered a full load. Faculty should be very tough on students' requests for absences and, in the case of unexcused absences, may take this into consideration for final grades. Students must be committed 100% to their summer studies. It's simply too hard to catch up. Of course, in the case of illness or special circumstances, arrangements can be made, but if a student misses too much of the course, withdrawal may be the only viable option. There are no dean's excuses in the summer."

Academic Integrity

Please read the Yale College statement on academic honesty, available at: http://catalog.yale.edu/handbook-instructors-undergraduates-yale-college/teaching/academic-dishonesty/

Also, please use the link on that page to reach the relevant sections of the Undergraduate

Regulations, if you are not familiar with these, and use the link to the Yale College Writing Center's website to get information about citing sources. This course is an excellent example of what the authors of the intellectual honesty statement meant when they wrote "Discovering how to use others' work to advance your own is a key part of learning. Very few of us ever have completely original ideas, and even the greatest scholars build on their predecessors' achievements."

Classroom Decorum

As a matter of courtesy to me and to your fellow students, I expect you to arrive at class on time, prepared and ready to participate. If you know that you will need to leave a particular lecture early, please let me know beforehand. If it is necessary to leave or enter the room once class has begun, please do so quietly.

The use of cell phones is NOT allowed during class and exams. I ask that you turn them off and you put them away.

Laptops can be used during class as long as they are being used in relationship to the class and they do not become a source of distraction to your peers or to me.

If you have trouble hearing or concentrating on the meeting or media presentation because of distractions around you, quietly ask those responsible for the distraction to stop. If the distraction continues, please let me know.

No audio or video recording of my lecture is permitted without prior, written approval.

Course Requirements

At the time of submission of this syllabus, together with the new Summer Session course proposal, I base the proposed requirements on the assumption of 30-40 students and 1TF. They will be adjusted accordingly if the number of students is substantially smaller (e.g. 10-15).

Undergraduate Student Requirements

All assignments will be deducted 10% of the final grade for each day that they are late. I do not give assignments for extra credit, so please be sure to pay close attention to dates of exams and assignments. Makeup examinations will only be given with a Dean's excuse.

	Percentage of Final Grade	Date when is due
TQs (thoughts and questions)	15%	For 4 meetings, due 24 hs before the first weekly meeting
Midterm 1	30%	First half of meeting, June 11th
Midterm 2	35%	
Participation during class meetings	20%	Throughout the course
Total	**100%**	

TQs ("Thoughts and Questions", 15% final grade): On 4 occasions during the course you will need to respond in writing to two or three prompts about the assigned readings for the meetings that week. You will submit your "thoughts and questions" via the Assignments

page in Canvas at least a day before class so that I can use your responses to shape the upcoming class meetings. The forms will be due on xxxxx at xxxx hs. They will not be graded individually, the overall quality of your submissions during the course will be graded.

Two Midterms (30% and 35% each of final grade): They will consist of a combination of short and long essay questions. The second midterm will require to draw from material learned and understood during the first half of the course. Midterm 1 during the 5th week, Midterm 2 during the 10th week.

Participation (20% of final grade): Participation in class will be graded throughout the semestser.

Grading is NOT on the curve; it will be based on the following cut-offs:

93 - 100% = A	83 – 86% = B	73 – 76% = C	63 – 66% = D
90 – 92% = A-	80 – 82% = B-	70 – 72% = C-	60 – 62% = D-
87 – 89% = B+	77 – 79% = C+	67 – 69% = D+	<60% = F

<u>Readings</u>

There is no required textbook for the class. *Required* readings will be available in the weekly modules of the course site in Canvas. There may also be *Suggested* readings for those with a particular interest in the topic.

The internet is wonderful, but it is also a great source of misinformation. As you can imagine there is no shortage of websites focused on the topics of the course, but no one exercises quality control over most of them, and much of what they contain is outdated, misleading, and/or wrong. Feel free to consult with me when deciding on readings from the internet that you are not sure if they are an acceptable source of information for a scientific course like this one.

Note: Weekly Schedule begins in next page.

Weekly Schedule, Topics and Required Readings

Scientific Thinking and Reasoning, ANTH S018, Summer 2021

CLASS 1: Introduction: the cult of statistical significance, p-values and good statistical practice

> "Good statistical practice, as an essential component of good scientific practice, emphasizes principles of good study design and conduct, a variety of numerical and graphical summaries of data, understanding of the phenomenon under study, interpretation of results in context, complete reporting and proper logical and quantitative understanding of what data summaries mean. No single index should substitute for scientific reasoning."
>
> Wasserstein & N.A. Lazar (2016). The ASA's statement on p-values: context, process, and purpose, The American Statistician

Learning objectives for Class 1:

Understand the course goals and aims and know where to find course resources.
Be able to write a paragraph explaining the goals of the course, including a summary of the problems associated with the use and misuse of p-values in science.
Be able to define what a p-value is.

<u>Activities</u>

Building our community of learners; present teaching strategy and philosophy on learning (active learning and expectations for student engagement)
Expectations - major learning objectives, guidelines for class discussions & participation
Course structure & logistics (Canvas, readings, timeline of the semester & assignments overview, etc.)
In preparation for first graded assignment (TQ#1, due xxx) go over Canvas submission process and review responses to TQ#0 non-graded.

<u>Required Readings</u>

(Bruni 2018). The most important reading for your first class of your first semester of your first year of college!
(Leek and Peng 2015). A roadmap to our course. P-values have been misused, but there are other stages of the scientific inquiry process that deserve more attention.
(Amhrein, Greenland et al. 2019). A more recent article in Nature, signed by 800 scientists asking to "retire" statistical significance.
(Nuzzo 2014). A commentary in Nature calling attention to the misuse of p-values.
(Wasserstein, Schirm et al. 2019). A challenging reading. The goal is that by the end of the semester, not on the first week! ☺, you will understand most of it,
(Wasserstein and Lazar 2016)
(Ziliak and McCloskey 2008). Preface and section titled "A Significant Problem", pages 1-22. Again, do a first read, you will understand some of it, it is about getting started with some jargon.

<u>Supplementary Readings</u>

(Wheelan 2013). "Introduction: Why I hated calculus but loved statistics."
(Vickers 2010). Some entertaining brief chapters that illustrate central topics of the course.

(Reinhart 2015), Preface, Introduction and Chapter 1.

Bibliography

Amhrein, V., S. Greenland and B. McShane (2019). "Retire Statistical Significance." Nature **567**: 305-307.
Bruni, F. (2018). How to get the most out of college. New York Times. New York City.
Leek, J. and R. Peng (2015). "P-values are just the tip of the iceberg." Nature **520**: 612.
Nuzzo, R. (2014). "Statistical Errors. P-values, the "gold standard" of statistical validity, are not as reliable as many scientists assume." Nature **506**: 151-152.
Reinhart, A. (2015). Statistics Done Wrong. The woefully complete guide. San Francisco, William Pollock.
Vickers, A. (2010). What is a p-value anyway? 34 Stories to Help You Actually Understand Statistics, Pearson Education, Inc.
Wasserstein, R. L. and N. A. Lazar (2016). "The ASA's Statement on p-Values: Context, Process, and Purpose." The American Statistician **70**(2): 129-133.
Wasserstein, R. L., A. L. Schirm and N. A. Lazar (2019). "Moving to a World Beyond "p < 0.05"." The American Statistician **73**: 1-19.
Wheelan, C. J. (2013). Naked Statistics. Stripping the Dread from the Data. New York, W.W. Norton & Company.
Ziliak, S. T. and D. N. McCloskey (2008). The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives. Ann Arbor, The University of Michigan Press.

CLASS 2: Scientific Skepticism, Facts and Truths

Why are we doing what we are doing? What are we trying to accomplish? Knowledge? What is it? How do we determine we have "learned"? Methods of inquiry: religion, art, philosophy, science. How we accept new ideas: tenacity, reason, common sense, science. Science: empiricism and rationalism. Deductive and inductive reasoning. Propositional logic, fallacies.

Learning objectives for Class 2

Write a paragraph explaining what "ATOM" means for Wasserstein et al. 2019
Define skepticism as proposed by Michael Shermer and explain, orally, what he means by it.
Explain what is meant by Ziliak and McCloskey when they say: "statistical significance is neither necessary nor sufficient for a scientific result" and "fit is not the same as importance". Be able to illustrate your explanation with an example.

Activities

Discuss Robert M. Pirsig, Ch. 13, The Art of Motorcycle Maintenance.
Further conversation on the topics of WEEK 1: be prepared to discuss the points that Ziliak and McCloskey (2008) which was assigned for WEEK1 make about "sizeless" science, "oomph", "magnitude". There is an emphasis on paying attention to the magnitude of findings, their relevance. Focus on the section "But the point of counting…" (pages 5 to 13).
From Wasserstein et al. 2019, we will discuss what they mean by remember "ATOM".
The second half of the meeting we will use to have our first open discussion, among all of us, about skepticism.

Required Readings

(Pirsig 1974). This reading is assigned because of the pandemic we are living through and given how it has changed the way in which universities function. An appropriate reading for this time of virtual education.

(Sagan 1987). An enjoyable easy read.
(Shermer 1997, Shermer 1997). Chapters 1 and 3 are the ones to focus on.
(Shermer 2017). Michael Shermer founding publisher of Skeptic magazine
https://www.skeptic.com/
(Salovey 2016). President Salovey's speech to incoming students in Fall 2016.
(Wasserstein, Schirm et al. 2019). You read this one last week. Go back to it trying to find text that relates to skepticism. They do not use the word skepticism, but there is something in "ATOM" that is implying it.
(Gelman 2019). I like this brief essay because it connects to some of the ideas behind ATOM. How?
(Gelman 2018).

Non-Required Supplementary Readings
(Shermer 1997). Given that Chapters 1 and 3 are assigned, I want you to have Chapter 2 available in case you want to read it.
Watch https://www.ted.com/talks/michael_shermer_on_believing_strange_things

Bibliography
Gelman, A. (2018). Ethics in statistical practice and communication. Five recommendations. Significance. **October:** 40-43.
Gelman, A. (2019). "When we make recommendations for scientific practice, we are (at best) acting as social scientists." European Journal of Clinical Investigation **49**(10).
Pirsig, R. M. (1974). Zen and The Art of Motorcyle Maintenance. Chapter 13.
Sagan, C. (1987). "The Burden of Skepticism." Skeptical Inquirer **12**(Fall 1987).
Salovey, P. (2016). "Countering False Narratives." from https://president.yale.edu/speeches-writings/speeches/countering-false-narratives.
Shermer, M. (1997). How Thinking Goes Wrong, Chapter 3. Why People Believe Weird Things.
Shermer, M. (1997). Science vs Pseudoscience, Chapter 2. Why People Believe Weird Things.
Shermer, M. (1997). Skeptic Manifesto. Chapter 1. Why People Believe Weird Things? Pseudoscience, Superstition, and Other Confusions of Our Time.
Shermer, M. (2017). "Postmodernism vs. Science." Sci Am **317**(3): 90.
Wasserstein, R. L., A. L. Schirm and N. A. Lazar (2019). "Moving to a World Beyond "p < 0.05"." The American Statistician **73**: 1-19.

CLASS 3: Research Questions, Theories, Hypotheses and Predictions

Learning objectives for Class 3
Formulate a scientific research question and propose hypotheses to explain your observations.
Write a paragraph describing the basic types of research questions as described by Leek.
Explain why are multiple hypotheses necessary and relate this to the "fallacy of the transposed conditional" (Ziliak & McCloskey, 2008, p. 17)
Define primary and secondary sources of information. Explain "in-text" citation and bibliography
Scientific literature and use of the library

Activities
We will use Leek 2015 for then trying to understand the research questions and hypotheses in Wright et al. 2020
We will practice how to formulate hypotheses and predictions.

Required Readings
(Leek and Peng 2015). Read this one first.
Formulating a research question and hypotheses. This is an exercise that I want you to read before class so that you make sure you understand it. You will need to do the exercise later on.
(Wright, Linton et al. 2020). Our first true "scientific" reading. Read it trying to identify research questions, hypotheses and predictions. I encourage you to try writing the questions down.
(Strode 2011). A powerpoint presentation. Please look at it.
(Strode 2015). In this article Paul Strode explains the same ideas than in the powerpoint.
(Chamberlin 1965). Please notice that this is a version of a talk given by Thomas C. Chamberlain in 1890!

Suggested Readings and Browsing
https://undsci.berkeley.edu/article/howscienceworks_01
https://mrdrscienceteacher.wordpress.com/2014/11/02/teaching-the-hypothesis/
https://undsci.berkeley.edu/search/lessonsummary.php?topic_id=&keywords=&type_id=9&discipline_id=&sort_by=audience_rank&Submit=%20%20%20%20Go%20%20%20%20&thisaudience=College&resource_id=221
https://www.biointeractive.org/classroom-resources
(Lipton 2005)
(Lipton 2005). A challenging reading.

Bibliography
Chamberlin, T. C. (1965). "The Method of Multiple Working Hypotheses: With this method the dangers of parental affection for a favorite theory can be circumvented." Science **148**(3671): 754-759.
Leek, J. T. and R. D. Peng (2015). "What is the question?" Science **347**(6228): 1314-1315.
Lipton, P. (2005). "Accommodation or prediction? Response." Science **308**(5727): 1411-1412.
Lipton, P. (2005). "Testing hypotheses: Prediction and prejudice." Science **307**(5707): 219-221.
Strode, P. K. (2011). The global epidemic of confusing hypotheses with predictions.
Strode, P. K. (2015). "Hypothesis Generation in Biology: A Science Teaching Challenge & Potential Solution." The American Biology Teacher **77**(7): 500-506.
Wright, K. P., Jr., S. K. Linton, D. Withrow, L. Casiraghi, S. M. Lanza, H. Iglesia, C. Vetter and C. M. Depner (2020). "Sleep in university students prior to and during COVID-19 Stay-at-Home orders." Curr Biol **30**(14): R797-R798.

CLASS 4: Study Design
This week we will begin discussing the next steps after we have identified an interesting questions, hypotheses and predictions. Namely, we need to start about collecting data. The readings that I have chosen still go over some of the topics we have already discussed, but then introduce issues of data type, data collection and study design. Following is a list of concepts that I want you to start noticing; by the end of the semester you are expected to know them all. List: non-experimental and experimental designs, Naturalistic observations, Archival research, Correlational studies, Single subject studies, randomized studies, experimental group vs control group, independent variable vs dependent variable, treatment effect, error, validity, reliability, selection bias, confounder.

Learning objectives for Class 4
Identify (e.g. mark them in a graph) the concepts in Figure 1.3 of H.J. Seltman, 2018, Ch. 1 that we have discussed in the course so far, define them and provide an example.

Explain the difference between a prediction and a "predilection"
Describe with text and a graph the Medawar zone. Write a paragraph explaining how it fits in the context of the goal and objectives of the course.
Formulate a "good" scientific question, explaining what elements of it make it "good". Both using an example from the readings and an example of your own.
Explain the concept of "validity". Illustrate with an example of your own.
Explain the two figures in Kardish et al. 2015.
Presented with a given variable, be able to identify its type.

Activities
The main activity will be to discuss the required readings so that we understand some of the basic concepts listed above.
An emphasis on discussing our second empirical journal article (Kardish) against the background of the other readings.

Required Readings
(chapter 5, Cohen and Medley 2005). I have uploaded the complete book as one single file, because the more I consult the book, the more I like it and the more I think it is absolutely a fantastic book for students. For this week, the emphasis is on chapter 5, but I encourage you to have a look at all the chapters before 5.
(chapter 3, Hartig 2015). A brief easy reading, but may want to browse chapter 1 which touches on things we have already discussed.
(chapter 1, Seltman 2018).
(chapter 2, Seltman 2018).
(Kardish, Mueller et al. 2015)

Supplementary Readings and Browsing
(Preamble, chapters 1 and 2 Cohen and Medley 2005)
https://undsci.berkeley.edu/teaching/misconceptions.php
Diaz's book, section 1.3.5
qualitative and quantitative
research. American Political Science Review 95: 529-546.

Bibliography
Cohen, J. and G. Medley (2005). Stop working & start thinking : a guide to becoming a scientist New York ; Abingdon, UK, Taylor & Franci.
Hartig, F. (2015). Research Skills. An Introduction to the Crafts of a Scientist.
Kardish, M. R., U. G. Mueller, S. Amador-Vargas, E. I. Dietrich, R. Ma, B. Barrett and C. C. Fang (2015). "Blind trust in unblinded observation in Ecology, Evolution, and Behavior." Frontiers in Ecology and Evolution **3**.
Seltman, H. J. (2018). Experimental Design and Analysis.

CLASS 5: The Replication Crisis

Learning objectives for Class 5
Explain the difference between reproducibility and replicability
Illustrate an essay describing the goals of the course with points made in the readings by Yanai and Lercher, Stark and Saltelli and Cohen & Medley.
Write a paragraph explaining the goal of the course. You have the goal in the syllabus, you should be able to expand on it.

Activities
MIDTERM 1
Discussion of assigned readings.

Required Readings
(chapter 3, Cohen and Medley 2005).
(McNutt 2014)
(Broman 2017)
(Stark and Saltelli 2018)
(Yanai and Lercher 2020)

Bibliography

Broman, K. e. a. (2017). "Recommendations to Funding Agencies for Supporting Reproducible Research " American Statistical Association.
Cohen, J. and G. Medley (2005). Stop working & start thinking : a guide to becoming a scientist New York ; Abingdon, UK, Taylor & Franci.
McNutt, M. (2014). "Journals unite for reproducibility." Science **346**(6210): 679-679.
Stark, P. B. and A. Saltelli (2018). "Cargo-cult statistics and scientific crisis." Significance: 40-43.
Yanai, I. and M. Lercher (2020). "A hypothesis is a liability." Genome Biology **21**(231-).

CLASS 6: How we make decisions in Science?

Learning objectives for Class 6

Explain the specifics of the Final Assignment.
Define causal inference.
Describe the difference between descriptive and inferential statistics and provide an example of each.
J. Pearl (The Book of Why) writes that "data are profoundly dumb." Write a paragraph explaining why.
In your opinion, what is the role of "principle" (or "principled argument") in statistics?

Activities:
You will need to write something on the assigned readings. Come prepared for some form of in-class TQ about the assigned readings.
First discussion of Final Essay plans.
Discussion of Midterm 1.

Required Readings
(Wheelan 2013). I think you will love this reading. Read the Introduction and Chapter 1. What's the Point?
(Abelson 1995). Robert P. Abelson taught an introductory statistics course in the Dept. Psychology at Yale. Read the Preface and pages 1-8.
(Pearl 2018). Read the Preface and Introduction "Mind over Data".
https://www.youtube.com/watch?v=PirFrDVRBo4
Gelman A 2019 *The Book of Why* by Pearl and Mackenzie. In: Statistical Modeling, Causal Inference, and Social Science (blog post) https://andrewgelman.com/2019/01/08/book-pearl-mackenzie/
(Pearl 2018). Read the Preface and Introduction "Mind over Data".
(Shapiro 2002). Wait on this one until I confirm if it is "required" or not.

Bibliography

Abelson, R. P. (1995). Statistics as Principled Argument. Hillsdale NJ, Lawrence Erlbaum Associates.
Pearl, J. (2018). The Book of Why. The New Science of Cause and Effect.
Shapiro, I. (2002). "Problems, methods, and theories in the study of politics, or what's wrong with political science and what to do about it." Political Theory **30**(4): 596-619.
Wheelan, C. J. (2013). Naked Statistics. Stripping the Dread from the Data. New York, W.W. Norton & Company.

CLASS 7: Planning Your Project

Learning objectives for Class 7

Conduct a bibliographic search using Endnote and organize the articles you have found in your Endnote Library. You will definitely need to be able to do this for your Final Assignment.
Write a paragraph explaining your INITIAL ideas on a possible Final Assignment
Write something of your choice from one of the chapters V to XI in Beveridge, 1957 The Art of Scientific Investigation.
Know how to use the Feedback function in Canvas
Describe the difference between experimental and observational studies and explain one benefit and one drawback of each.
Explain the role of construct, internal, external, and statistical validity in evaluating the quality of a research question.
Compare Erren's "Ten Simple Rules" with the ATOM paradigm from earlier this semester. Which do you like better? How are they similar or different?

Activities
Discuss your answers to TQ#2
Rationale for Study, Prospectus Plans

Required Readings and Browsing
(chapter 2, Hartig 2015). We read chapters 3 and 1 of F. Hartig's book before. Basic, simple chapter relevant for starting to think about your Final Assignment.
touches on things we have already discussed.
(Paydarfar 2001)
(Erren, Slanger et al. 2015)
(Erren, Cullen et al. 2007)
(chapter 4, Hartig 2015). A brief easy reading, but may want to browse chapter 1 which
(Beveridge 1957)

Bibliography

Beveridge, W. I. B. (1957). The Art of Scientific Investigation.
Erren, T. C., P. Cullen, M. Erren and P. E. Bourne (2007). "Ten simple rules for doing your best research, according to Hamming." PLoS Comput Biol **3**(10): 1839-1840.
Erren, T. C., T. E. Slanger, J. V. Gross, P. E. Bourne and P. Cullen (2015). "Ten Simple Rules for Lifelong Learning, According to Hamming." Plos Computational Biology **11**(2).
Hartig, F. (2015). Research Skills. An Introduction to the Crafts of a Scientist.
Paydarfar (2001). "Algorithm for discovery." Science.

CLASS 8: Descriptive Statistics and Exploratory Data Analyses (EDA)

One of the most valuable pieces of advice I can offer you is to become mindfully familiar with their dataset. For example, we should all really understand the how the data are distributed, what their central tendency is, their dispersion patterns, outliers, whether each variable is normally distributed and what does that mean for the analytical tests we may be applying as we try to make "statistical inferences" about the data. It is also extremely useful to visualize the data set through the use of graphs. The more we understand the "shape" of our dataset BEFORE attempting any statistical inferences, the more we will be in a position to evaluate which statistical tools are more appropriate. Data exploration generally includes, among other tasks, data cleaning, numeric summaries, visualization tools, identification of outliers, missing data, what to do with lots of zeros, normality of distribution of data.

Learning objectives for Class 8

Explain the role of EDA in science, especially in the social sciences.
Define and explain concepts in **bold** in Seltman 2018
Write a paragraph explaining a histogram, boxplot, quantile-normal plots, and Likert plot graphs. Describe their functions and the characteristics of data that each type of graph is best-equipped to visually demonstrate.
When looking at a set of data, or a graph of them, explain the extent to which the data "look" normally distributed and the apparent extent of skewness and kurtosis.
Use Excel, both workbooks and csv (comma separated value) files. Identify the different types of variables in a "cortisol" owl monkey data set and graph some of the data using Excel or any other program of your choice.

Activities

You need to come prepared to look at the real data on "cortisol" values as estimated from fecal samples of owl monkeys that is available in the file "Cort.data.19Nov19.csv".
And to discuss the manuscript we are working on in our group

Required Readings and Browsing

Corley et al in preparation ANTH 018
Cort.data.19Nov19.csv
(chapter 7, Wheelan)
(Seltman 2018) Chapter 4 (pp. 61-98) on Exploratory Data Analysis
(Ch 3 Exploratory Data Analysis, Cox 2017). Excellent book very clear, simple.
(Ch 4 Descriptive Statistics, Cox 2017). Excellent book very clear, simple.

Supplementary Readings

(Zuur, Ieno et al. 2010) An excellent "recipe" for data exploration, with clear examples
(Tabachnick and Fidell 2013) A crucial book for multivariate statistic. For EDA, check Chapter 4.

Bibliography

Cox, V. (2017). Translating Statistics to Make Decisions: A Guide for the Non-Statistician. Salisbury, UK, APress.
Seltman, H. J. (2018). Experimental Design and Analysis.
Tabachnick, B. and L. Fidell (2013). Chapter 4 - Cleaning up your act. Using Multivariate Statistics
60-115.
Wheelan, C. J. (2013). Naked Statistics. Stripping the Dread from the Data. New York, W.W. Norton & Company.
Zuur, A. F., E. N. Ieno and C. S. Elphick (2010). "A protocol for data exploration to avoid

common statistical problems." Methods in Ecology and Evolution **1**(1): 3-14.

CLASS 9: Causation and Causal Inference
Learning objectives for Class 9
-Define the concept of conceptual frameworks and conceptual diagrams.
-List the basic elements in the diagram and describe what they mean.
-Define the concepts of mediators, moderators, confounders and colliders.
-Provide examples of descriptive and causal diagrams, including the narrative that goes with them.
-Work out a conceptual diagram from a research question you want to investigate.
-Explain the "Ladder of Causation" with reference to the three rungs.
-Write a paragraph on the possible role of "imagination" in human evolution.

Monday, October 26th
Activities
Presentation by Dr. Claudia Valeggia

Required Readings
(Pearl 2018). The Book of Why. Read Chapter 1, The Ladder of Causation.(Paradies and Stevens 2005) and REREAD Introduction (you have read it already for Week 6).
(Paradies and Stevens 2005). A 2-page article that explains Conceptual Diagrams. Have a quick look at it before reading Pearl.
Watch video on Conceptual Diagrams:
https://www.youtube.com/watch?v=MnfRdTCUIsc&feature=youtu.be
(Barrowman 2014). A magazine article, a good easy read to find those terms that may be challenging and try clarify what they mean.

Suggested Readings
Baron RM, Kenny DA 1986 The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of Personal and Social Psychology 51: 1173-1182.

Bibliography
Barrowman, N. (2014). Correlation, Causation and Confusion. The New Atlantis. **43:** 23-44.
Paradies, Y. and M. Stevens (2005). "Conceptual diagrams in public health research." J Epidemiol Community Health **59**(12): 1012-1013.
Pearl, J. (2018). The Book of Why. The New Science of Cause and Effect.

CLASS 10: Observation and Experiment
Learning objectives for Class 10
-Explain counterfactuals, confounders/confounding, selection bias, operationalization, and provide examples to illustrate.
-List Hill's criteria and provide examples for 2 of them.
-Draw a diagram that illustrates some of the thinking around the issue of smoking and cancer.
-Explain qualitative research. Illustrate the value of it in the context of quantitative research
-Write an essay on the history of the smoking-cancer research. In your writing, show understanding of some of the scientific technical concepts/ideas discussed by J. Pearl (Ch. 5).
-Explain the following types of data/studies: time-series, case-control, dose-effect.
-Explain the "fundamental problem of causal inference."

-Review the historical instances of discovery outlined in Freedman and identify three or more unifying characteristics of scientific innovation.
-explain the low birth weight paradox

Activities
We will discuss the readings.
Further discussion of the low birth weight-smoking paradox and other paradoxes from the required reading.
Watching video of DAGs Course
Midterm 2

Required Readings
(Smith 2019). A biological anthropology colleague; he has published extensively on the limitations of statistical significance. This article discusses the challenges common to our discipline, biological anthropology, which relies much on "observations" (observational data).
(Pearl 2018). Read Chapter 5, a historical account of identifying the causality between smoking and health outcomes.
(Freedman 2008). The first reading on "qualitative" research. Focus on introduction and conclusions and read the example that catches your attention.
(Hernán, Hsu et al. 2019). Read it first.
(Glass, Goodman et al. 2013). This article discusses again smoking and Hill's criteria
J. Pearl, Chapter 6. Paradoxes Galore!
Hernandez-Diaz et al. Birth Weight Paradox

Suggested Reading
(Holland 1986). This is the author who identified "The fundamental problem of causal inference". Section 5 has some comments on philosophers who have thought about causes. Very interesting.

Bibliography
Freedman, D. A. (2008). On types of scientific inquiry: the role of qualitative reasoning. The Oxford Handbook of Political Methodology. J. M. Box-Steffensmeier, H. E. Brady and D. Collier.
Glass, T. A., S. N. Goodman, M. A. Hernan and J. M. Samet (2013). "Causal inference in public health." Annu Rev Public Health **34**: 61-75.
Hernán, M. A., J. Hsu and B. Healey (2019). "A second chance to get causal inference right: a classification of data science tasks." Chance **32**(1): 42-49.
Holland, P. W. (1986). "Statistics and Causal Inference - Rejoinder." Journal of the American Statistical Association **81**(396): 968-970.
Pearl, J. (2018). The Book of Why. The New Science of Cause and Effect.
Smith, R. J. (2019). "Living with observational data in biological anthropology." American Journal of Physical Anthropology **169**(4): 591-598.