Statistical Reasoning for Environmental Scientists
Understanding Data in the 21st Century

Dr. Noah Morris
3 weekly hours, 3 academic credits

> *Statistical thinking will one day be as necessary for efficient*
> *citizenship as the ability to read and write*
> *(attributed to H. G. Wells)*

Data and statistical analysis are being presented to us in ever increasing quantities. This course is based on the belief that statistical reasoning and an ability to understand data are essential tools for any well educated person and in particular for those involved in environmental science. The course does not aim to introduce complex mathematical methods of statistical analysis but it does aim to introduce a way of thinking about statistics and data. The course will succeed if the following aims are achieved:

1) Students are able to understand and explain the statistics presented in various academic papers.
2) Students are able to think critically about data which is presented.
3) Students have an understanding of the methods involved in data collection and the possible pitfalls which might be faced.
4) Students are able to present data clearly and choose which graphs are best suited to describe various data sets.
5) Students have an appreciation of the importance of making decisions in situations of uncertainty.
6) When planning research the students will have the ability to design what data should be collected and how to collect it.

After taking this course it is hoped that students will be better prepared to make rational decisions in situations of uncertainty about matters of social policy. They will be able to assess critically statistical claims that they encounter during discussions or when considering a news article or an academic paper. Statistical reasoning introduces students to the basic concepts and logic of statistical analysis and gives the students introductory-level practical ability to choose, generate, and properly interpret appropriate descriptive and inferential methods. In addition, the course will help students gain an appreciation for the diverse applications of statistics and its relevance to their lives and fields of study.
The course does not assume any prior knowledge in statistics nor will it involve any complicated mathematical formulae.

The course will include numerous examples of data concerning environmental issues and students will be expected to access and analyse data on a given environmental topic.

Description:

The course will include the following components:

1) **Data tells stories**. Students will be expected to look at various data-sets and explain what information they have learnt and what questions they want to ask about the data.

2) **Reading and understanding articles and graphs**. Each week one or two students will be expected to present the statistical results described in one of various articles (mainly from the environmental sciences). They will be asked to describe what they have read. This will include answering questions such as: What question was being asked? How was the data collected? Who paid for the research? What method of analysis was used? Were graphs used? What conclusions were made? Any criticism of the method, presentation and or conclusion?

3) **Don't believe everything you read.** Use will be made of resources such as the BBC podcasts "More or less" as well as newspaper articles and videos. Students will be taught to look critically at graphs and data as it is presented.

4) **Case Study – data collection and analysis**. The students will be expected to choose a research question of environmental interest, access and analyse the relevant data.
Each student will prepare a short (ten minute) presentation on the subject and submit a "newspaper article" (not more than 800 words) in which the data will be presented in an interesting and accurate way.

Marking System:

Presentation of graph or article	10%
Pop Quizzes	10%
Class Participation, Attendance and Punctuality	15%
Case Study (presentation and written report)	35%
Final Exam	30%

22nd February (week 1) - Data Tells Stories.
Introduction – what do we mean by statistical reasoning?
      How statistics can be useful and affect decisions
      How statistics can be misused
      Subjectivity in statistical analysis
Resources:
- Rosling H. *How not to be Ignorant.* (Ted talk)
- Paulsen R. (2019). *Why you shouldn't Listen to Self-Serving Optimists Like Hans Rosling and Steven Pinker*
- Blauw S. *The Number Bias – Foreword "Captivated by Numbers"*

(week 2)–How to read a graph. How to look at data.
What are the variables?
What are we supposed to learn from the graph?
Where does the data come from?
Examples of misleading graphs

Introduce the aims and structure of the case study including suggestions of possible subjects.
Resources:
- Reference to be added, Article about Covid 19.
- *Lea Gaslowitz. How to spot a misleading graph.*
  https://www.youtube.com/watch?v=E91bGT9BjYk
- *When Graphs Are a Matter of Life and Death.* The New Yorker, 21/6/2021

(week 3) – Reasoning in situations of uncertainty
An introduction to Bayesian Decision theory
What do we mean by uncertainty?
What is probability?
Subjective probabilities
Estimating and updating Probabilities
Resources:
- Bram U. *Thinking Statistically – Chapter One "Selection".*

(week 4) Collecting Data
Random sampling
How to construct a survey
How big should the sample be?
Common problems
What conclusions can we draw from a sample?
Confidence Intervals and p-values
What do we mean by "significance"?
Prepare a survey about household garbage in Kibbutz Ketura
Resources:
- Tim Harford: *Why pollsters so often get it wrong.* Financial Times 5/11/2020

By week four each student is expected to have chosen which data base they want to use for their case study. During the following week there will be individual meetings with each of the students to discuss what questions they want to ask.

(week 5) – How to summarise data: Central Values and Measures of Spread.
Can one number represent a whole population?
Mean versus Median – when is one more appropriate than the other?
Examples of research comparing the mean of two populations.
Potential misleading use of central values
Measures of Spread
Motivation – why do we need to measure the spread of a distribution?
Possible measures – Standard Deviation, Inter Quartile Range, Mean absolute Deviation.
Use of box plots
Examples of research using measures of spread.

The importance of measure of spread when comparing two populations.
Measure of spread for our garbage survey

Resources:
- Levitan D. *A Field Guide to Lies and Statistics. Chapter 2, pg 11-25*.
- Podcast – *More or Less (Extinction Rebellion*)
  https://www.bbc.co.uk/sounds/play/p07mnsmh

(week 6) - Ehical Issues
Kantian or Utilitarian approach?
1) Data collection:
   Using data from immoral studies
   What to do if the placebo group is deprived of a good treatment?
2) **Reporting Results:**
   Privacy
   Cherry picking
   Reporting negative results
3) Potential misuse of results

Resources:
- Book Review - *Hans Eysenck's controversial career* The Lancet 7/8/10
- Amodia D et al - *Neurocognitive correlates of liberalism and conservatism*. Nature, Neuroscience September 2007
- Lawrence M. Lesser &Erik Nordenhaug. Ethical Statistics and Statistical Ethics: Making an Interdisciplinary Module. *Journal of Statistics Education* Volume 12, 2004 - Issue 3
- Gelman A. *Ethics in statistical practice and communication: Five recommendations.* Royal Statistical Society 2018 https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.17409713.2018.01193.x.

By week six each student is expected to submit an outline of their case study including a brief summary of the data and what questions they intend to analyse. During the following week there will be individual meetings with each of the students to discuss how to present their findings.

(week 7) - Comparing Different Groups
Using a graph to compare two groups.
Comparing the mean of two groups.
What do we mean by a significant result?
How to understand p-values.

Resources:
- Hanna-Attisha M. (2018). *What the Eyes Don't See*

(week 8) - Covariation
Correlation is not causation.
How can we find evidence for causation as opposed to correlation?
Correlation can be non-linear.
Scattergrams

Resources:
- *Understanding Causation.*
  https://www.youtube.com/watch?v=VMUQSMFGBDo
- *The danger of mixing up causality and correlation: Ionica Smeets at TEDxDelft.* https://www.youtube.com/watch?v=8B271L3NtAw

(week 9) – Graphs, Predictions and Time Series
Which graphs are appropriate in which situations?
How to spot a misleading graph.
Predictions and uncertainty.
Trend, seasonality and noise
Resources:
- Gaslowitz L. *How to Spot a Misleading Graph.* www.youtube.com/watch?v=E91bGT9BjYk
- *Misleading Graphs – Real Life Examples. https://www.youtube.com/watch?v=1F7gm_BG0iQ*

(week 10) - Presentation of case studies

(week 11) – More presentations and a summary of the course
What have we learnt:
    Statistical reasoning?
    Decisions in situations of uncertainty
    Critical reading of results
    Ethical questions

(week 12) Final Exam

References:
Books and Articles:
    Bram U. (2012). *Thinking Statistically*. Capara Books.

    Blauw S. (2020). *The Number Bias – How Numbers Lead and Mislead us*. London: Hodder and Stroughton Ltd.

    Francois K., Monteiro C. and Allo P. (2020). Big-Data Literacy as a new Vocation for Statistical Literacy. *Statistics Educational Research Journal*, March 2020.

    Hanna-Attisha M. (2018). *What the Eyes Don't See: A Story of Crisis, Resistance, and Hope in an American City.*New York: Penguin Random House

    Harford T (2021). *How to Make the World Add Up. London*: The Bridge Street Press

    Huff D. (1983). *How to Lie with Statistics*. Penguin Books.

    Lawrence M. Lesser &Erik Nordenhaug. Ethical Statistics and Statistical Ethics: Making an Interdisciplinary Module. *Journal of Statistics Education* Volume 12, 2004 - Issue 3

    Levitan D. (2016). A Field Guide to Lies and Statistics. Penguin Books.

    Paulsen R. (2019). *Why you shouldn't Listen to Self-Serving Optimists Like Hans Rosling and Steven Pinker. http*://www.inthesetimes.com

Reports:
    IPCC (2021). *Climate Change 2021, The Physical Science Basis, Summary for Policymakers.*
    https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM_final.pdf

United Nations Report (2019). *World Population Prospects 2019, Highlights*
Tableshttps://reliefweb.int/sites/reliefweb.int/files/resources/WPP2019_Highlights.p
df

Videos:

Gaslowitz L. *How to Spot a Misleading Graph*. www.youtube.com/watch?v=E91bGT9BjYk

Misleading Graphs – Real Life Examples.
https://www.youtube.com/watch?v=1F7gm_BG0iQ

Rosling H. *How Not to be Ignorant.* (Ted talk)

Smeets I. The danger of Mixing up Causality and Correlation.
https://www.youtube.com/watch?v=8B271L3NtAw