מכון הערבה
Arava Institute
معهد وادي عربة

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

Instructor: Dr. Noah Morris
12 weeks 3 hours,

*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 have an ability to present data clearly and choose which graphs are best suited to describe various data sets.
5) 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 issue.

**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.

   [Resources – United Nations World Population Prospects, The IPCC report on climate change, Gapminder, PISA data, OECD statistics and more].

2) **Reading and understanding articles**. 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.
   a. Choose a data set to be collected and what questions are of interest.
   b. Decide how to collect the data.
   c. Collect the data.
   d. What stories do the data tell?
   e. What conclusions can be drawn
   f. Further questions.

**Week1 - Data Tells Stories.**

Introduction – what do we mean by statistical reasoning?
How can a mass of data be effectively summarised to give useful information and affect decisions?
Examples of data and what we can learn from them.
Exercise in pairs –
Look at different presentations of various data sets.
What can we learn and what questions should we ask?

References:


- 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*


**Week 2 – Reasoning in situations of uncertainty**

What do we mean by uncertainty?
What is probability?
How can data help to understand situations of uncertainty?
Exercise – Estimating probabilities of various events

- Estimate probabilities by calculating frequencies in games (dice, loaded dice, spinner etc)

References:

- Reference to be added, Article about estimating the probability of dyeing from Corvin 19.

**Week 3 – Collecting Data**

Random sampling
How to construct a survey.
Examples of good and problematic sample surveys.
Exercise – prepare a random sample of the local population and a list of quantitative questions (Height, Arm span, Age, Years of education, Country of Birth, Sex, etc). During the following week the students will collect the data from those selected in the random sample. (The data collected by the students and the resulting data base will be used in further exercises throughout the course).
***Introduce the aims and structure of the case study including suggestions of possible subjects***.
References:

- Bram U. *Thinking Statistically – Chapter One "Selection*".

**Week 4 – Central Values.**

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.
Example – using the data base collected in week three to find examples of mean and median and discuss which measure represents the population better.
***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.***
References:

- Levitan D. A Field Guide to Lies and Statistics. Chapter 2, pg 11-25.
- To be added – a relevant article estimating central values.

**Week 5 – 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.
Calculating spread for the data base collected during week three.
Examples of research using measures of spread.
The importance of measure of spread when comparing two populations.
References: To be added

Week **6 - 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.
References:

I plan to find two articles connected with environmental science which use significance tests and p-vales to evaluated the difference between two groups

**Week 7 - Covariation**
Correlation is not causation.
How can we find evidence for causation as opposed to correlation?
Exercise – testing DeVinci's hypothesis that arm span = height.
Correlation can be non-linear.
Scattergrams
Exercise – Look at various examples of correlation and answer the question – is this evidence for causation?
***By week seven 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 what questions they want to ask.***
References:
Smeets I. The danger of Mixing up Causality and Correlation (Ted talk)
Garfield G. and Ben-Zvi D. (2008). *Developing Students' Statistical Reasoning*, Chapter 14. Learning to Reason about Covariation

**Week 8 – Graphs and Predictions**
Which graphs are appropriate in which situations?
How to spot a misleading graph.
Predictions and uncertainty.
Exercise: In group look at examples of graphs from the media:
What can we learn from the graph
Any criticisms of the graph and suggested improvements
References:
Huff D. *How to lie with statistics pages* 58-63
Gaslowitz L. *How to Spot a Misleading Graph*

**Week 9 – Sampling and Sample Size**
What conclusions can we draw from a sample?
When is a sample large enough?
Confidence Intervals and p-values
What do we mean by "significance"?
Exercise: Look at various articles and explain in words what can be learnt from the reference to "significance", "confidence interval" and "p-value".
References:
United Nations Report (2017). *World Population* Prospects (to look at the use of confidence intervals)

**Week 10 – Decisions in Situations of Uncertainty**
An Introduction to Bayesian decision theory.
Loss functions
A polemic against being dogmatic.
Exercise: Questions concerning decisions in situations of uncertainty.
References:
Uri Bram – *Thinking Statistically* 75 – 84.

**Week 11 -**

Presentation of case studies.

**Week 12 – Summary**

What have we learnt about statistical reasoning?
Review of definitions.
Ethical questions - honesty with statistics – let the numbers talk.
Subjective (a-priori) opinions are legitimate but a rational person should allow new data to affect their opinions.

Possible Marking System:

| | |
|---|---|
| Class Participation | 10% |
| Analysis of article (class presentation) | 10% |
| Analysis of article (written summary) | 10% |
| Pop Quizzes | 10% |
| Case study (presentation and written report) | 30% |
| Final Exam | 30% |

References:

Reports:

IPCC (2018). *Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming.*

United Nations Report (2017). World Population Prospects, Key Findings and Advance Tables

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

Books and Articles:

Bram U. (2012). *Thinking Statistically*. Capara Books.

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

Garfield G. and Ben-Zvi D. (2008). *Developing Students' Statistical Reasoning.* Springer Science+Bussiness Media B.V.

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

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