

Welcome to BIOL 422 & 680
(cross-listed)

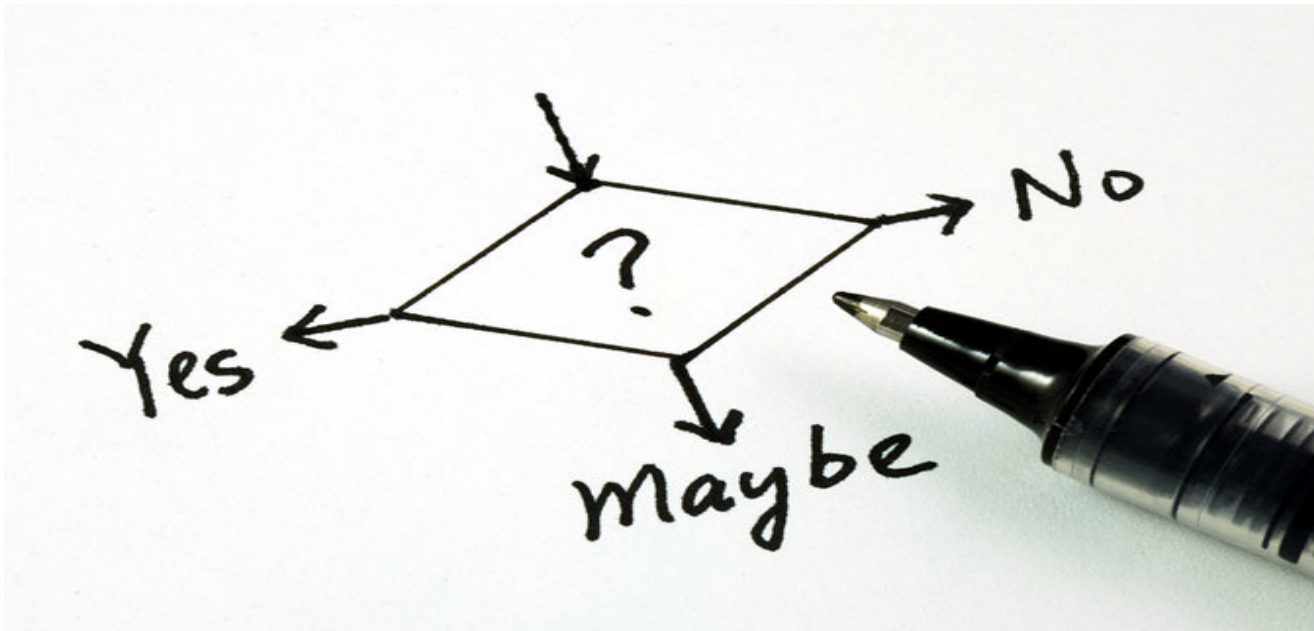
Advanced Statistics for Biological Sciences

Pedro Peres-Neto, PhD

Professor, Department of Biology,
Concordia University

& Canada Research Chair & Editor-in-Chief of *Oikos*

Welcome to the science of aiding decision-making with incomplete information (or without complete knowledge)



Statistics is key!

“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write”

- Herbert George Wells

Statistics is key!

*“Mathematics may rule the universe,
but statistics rule societies”*

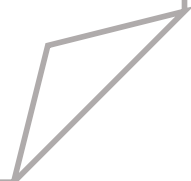
- Pedro in a very inspired moment during a BIOL322 lecture in 2018 😊

some initial thoughts

Learning is not a spectator sport. We don't learn much just sitting in classes listening to teachers, memorizing pre-packaged assignments, and spitting out answers.

We must talk about what we are learning, write about it, relate it to past experiences, and apply it to our daily lives. We must make what we learn part of ourselves.

- Chickering and Gamson



some initial thoughts about teaching philosophy



get cape.

wear cape.

fly.

information

practice

innovate

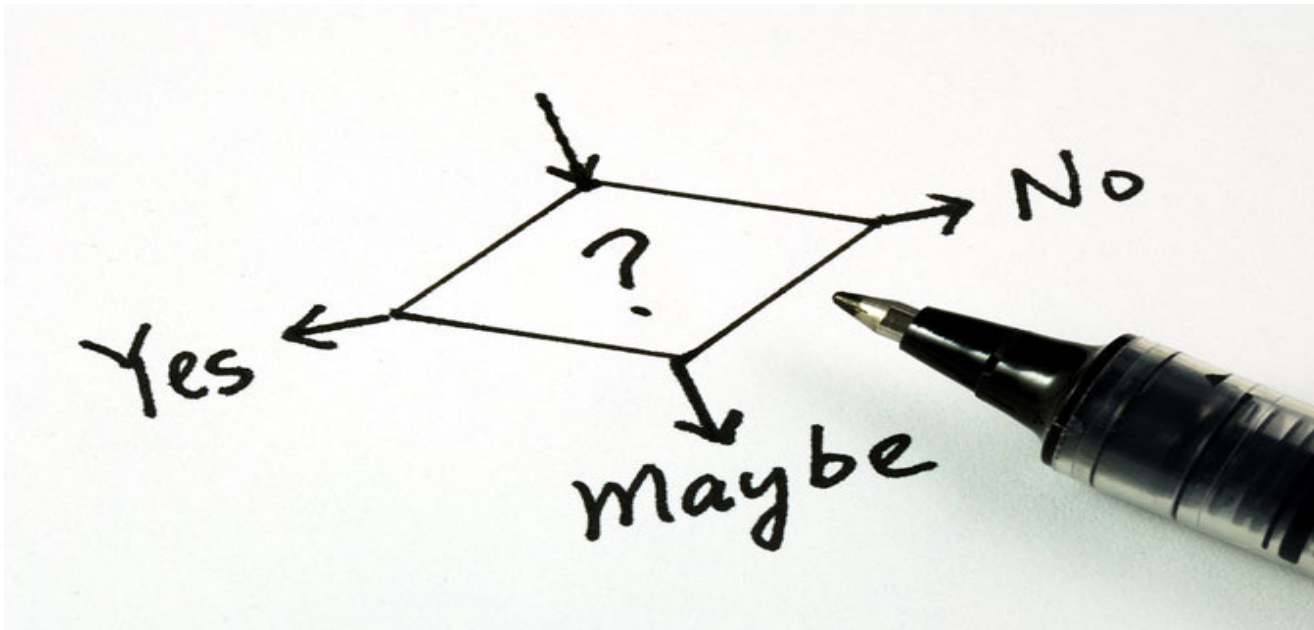
some initial thoughts

Learning is not a spectator sport.



EASIER SAID
Than Done

Statistics is key in decision-making processes because most decisions are made without complete knowledge (i.e., decisions always carry some level of uncertainty).



What is the difference between the two definitions?

“Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data.” Wikipedia

“Statistics is the science of learning from data, and of measuring, controlling and **communicating uncertainty.**” ASA

ASA includes critical thinking!

“Statistics” as defined by the American Statistical Association (ASA) “is the science of learning from data, and of measuring, controlling and **communicating uncertainty.**”

“The statements of science are not of what is true and what is not true, but statements of what is known with different degrees of certainty.” (Richard Feynman)

What is a statistical question?

- ✓ What is the average size of Canadians?
- ✓ Is 10 a number?

What is the difference between these two questions?

What is a statistical question?

- ✓ What is the average size of Canadians?
- ✓ Is 10 a number?

More information (data) changes (hopefully improving) the answer; i.e., one requires statistics and the other doesn't.

Do we need statistics to calculate the number of female candidates?

The image is a screenshot of a news article on the Global News website. At the top, there is a blue navigation bar with the Global News logo on the left, and links for 'National' (with a dropdown arrow and 'Change Location'), 'TV News Programs' (with a dropdown arrow and 'Newscasts & Videos'), 'Ad Choices', and a search icon. Below this is a secondary navigation bar with links for 'World', 'Canada', 'Local' (with a dropdown arrow), 'Politics', 'Smart Living', 'Money', 'Entertainment', 'Health', 'Commentary', 'Trending', 'Sports', 'Video', and 'More' (with a dropdown arrow). A 'Follow' button with a Twitter icon is also present. The main content area has a dark blue header with the text 'QUEBEC POLITICS' and sub-links for 'Liberal Party', 'Parti Québécois', 'Coalition Avenir Québec', and 'Québec Solidaire'. The article itself has a red 'POLITICS' tag, a timestamp 'October 16, 2017 6:26 pm', and an update time 'Updated: October 18, 2017 11:35 am'. The headline is 'Women make up 31% of candidates in Quebec municipal elections'. The author is 'By Raquel Fletcher', identified as 'Quebec City Correspondent Global News'. On the right side, there is a weather widget for 'Montre' showing a moon icon, a temperature of '26°', and the text 'Weather Clear'.

Global NEWS

National ▼
Change Location

TV News Programs ▼
Newscasts & Videos

Ad Choices

World Canada Local ▼ Politics Smart Living Money Entertainment Health Commentary Trending Sports Video More ▼

Follow

QUEBEC POLITICS Liberal Party Parti Québécois Coalition Avenir Québec Québec Solidaire

POLITICS October 16, 2017 6:26 pm Updated: October 18, 2017 11:35 am

Women make up 31% of candidates in Quebec municipal elections

By Raquel Fletcher
Quebec City Correspondent Global News

WEATHER Montre

 26°

Weather Clear

Do we need statistics to calculate the number of female candidates?

The screenshot shows a news article from Global News. The main headline is "Women make up 31% of candidates in Quebec municipal elections". The article is dated October 16, 2017, and was updated on October 18, 2017. The author is Raquel Fletcher, a Quebec City Correspondent for Global News. The page also features a navigation menu with categories like National, TV News Programs, World, Canada, Local, Politics, Smart Living, Money, Entertainment, Health, Commentary, Trending, Sports, Video, and More. A weather widget for Montreal shows a temperature of 26°C and a clear sky.

NO! This question would be answered by simply counting the number of female candidates over the total number of candidates. This question is not answered by collecting more data that may change the results.

NO! This question would be answered by simply counting the number of female candidates over the total number of candidates. This question is not answered by collecting more data that may **change the results**.



We should become comfortable with the idea that the most interesting and useful results may change if new information (data) is gathered



Statistics: “the science of assisting in decision making with incomplete knowledge”

What is statistics?

“Statistics is a science, not a branch of mathematics, but uses mathematical models as essential tools.”

- John Tukey

Statistical Thinking *versus* Mathematical Thinking

Mathematics is, by and large, a deterministic way of thinking and the way mathematics is taught in schools entrenches students into a deterministic way of viewing the quantitative world around them - What is the size of our planet?

Statistics is, by and large, a probabilistic or stochastic way of thinking (i.e., it considers uncertainty) - What is the probability that it will rain tomorrow?

Statistical Thinking *versus* Mathematical Thinking

Statistics is a separate discipline with its own unique ways of thinking and its own tools for approaching problems.

- J. Michael Shaughnessy, "Research on Students' Understanding of Some Big Concepts in Statistics" (2006)

Statistics *versus* Data Science

(demystifying a trend)

“For statisticians, the entire data science trend seems a bit patronizing. No matter what your exact definition of data science is, it’s going to sound pretty similar to the work that statisticians have been doing for decades.”

- Nate Silver

Let's take a small break – 1 minute

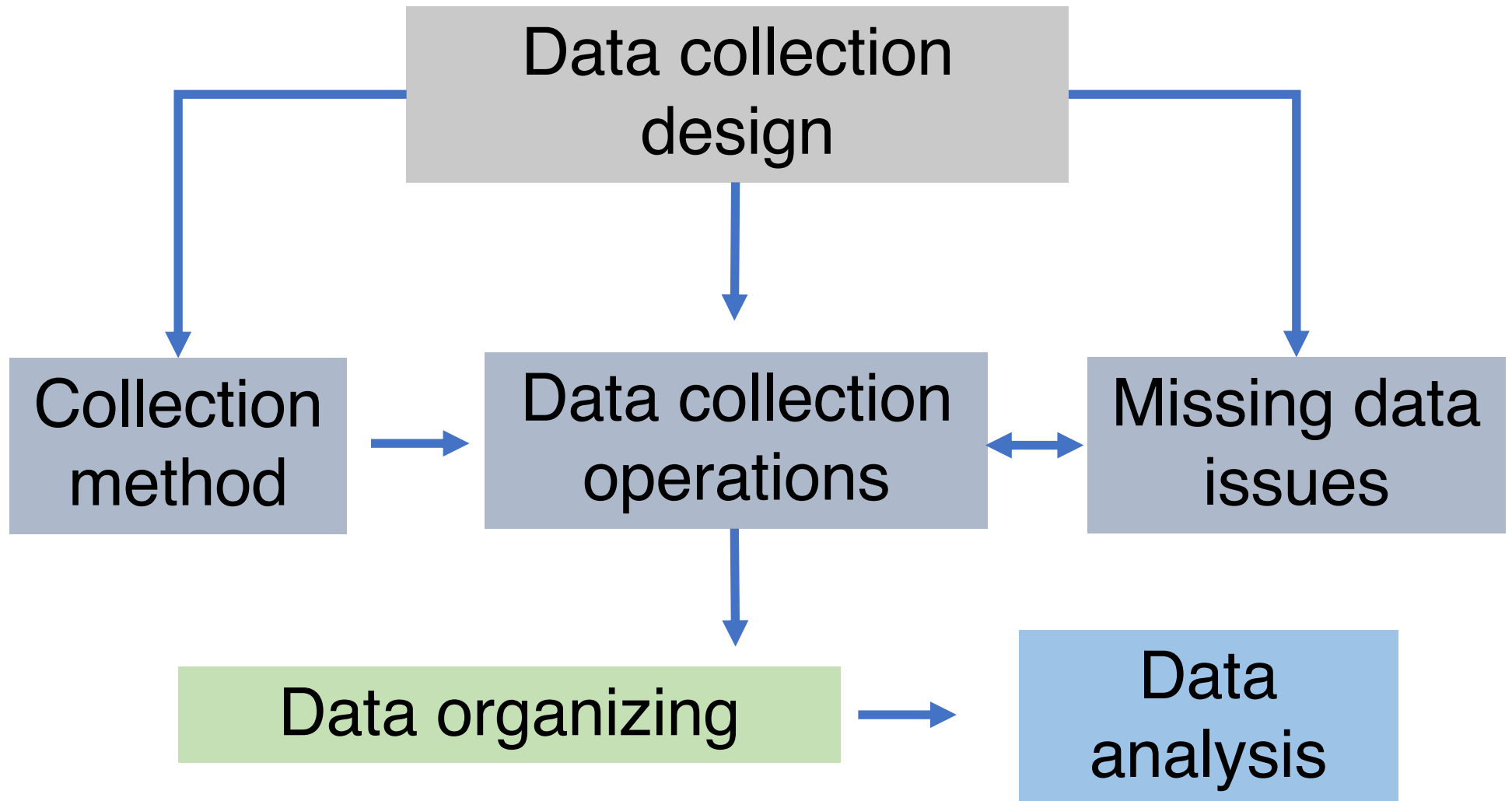


Roles of statistics

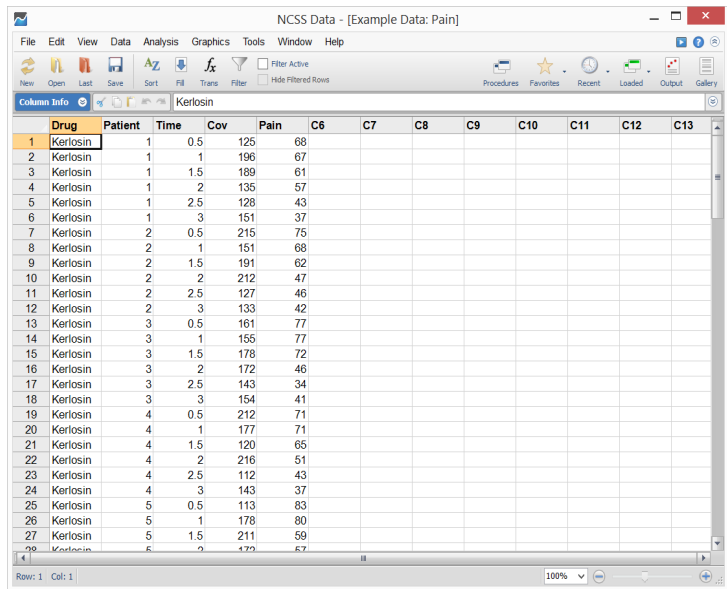
Statistics is a discipline that aims at:

- 1) Designing data collection protocols (observational and experimental).
- 2) Summarizing information to aid understanding.
- 3) Drawing conclusions from data.
- 4) Estimating the present or predict the future.
- 5) Communicate uncertainty.

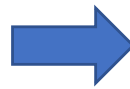
Roles of statistics (1) - Design data collection and protocols (experimental and observational)



Roles of statistics (2) – Summarize information to aid understanding



	Drug	Patient	Time	Cov	Pain	C6	C7	C8	C9	C10	C11	C12	C13
1	Kerlosin	1	0.5	125	68								
2	Kerlosin	1	1	196	67								
3	Kerlosin	1	1.5	189	61								
4	Kerlosin	1	2	135	57								
5	Kerlosin	1	2.5	128	43								
6	Kerlosin	1	3	151	37								
7	Kerlosin	2	0.5	215	75								
8	Kerlosin	2	1	151	68								
9	Kerlosin	2	1.5	191	62								
10	Kerlosin	2	2	212	47								
11	Kerlosin	2	2.5	127	46								
12	Kerlosin	2	3	133	42								
13	Kerlosin	3	0.5	161	77								
14	Kerlosin	3	1	155	77								
15	Kerlosin	3	1.5	178	72								
16	Kerlosin	3	2	172	46								
17	Kerlosin	3	2.5	143	34								
18	Kerlosin	3	3	154	41								
19	Kerlosin	4	0.5	212	71								
20	Kerlosin	4	1	177	71								
21	Kerlosin	4	1.5	120	65								
22	Kerlosin	4	2	216	51								
23	Kerlosin	4	2.5	112	43								
24	Kerlosin	4	3	143	37								
25	Kerlosin	5	0.5	113	83								
26	Kerlosin	5	1	178	80								
27	Kerlosin	5	1.5	211	59								
28	Kerlosin	5	2	179	67								



From raw (primary) data.....to summaries....
to information

Roles of statistics (3) – Produce estimates and draw conclusions from data based on samples

“There is a probability of 32% that a particular biological population will go extinct. The margin of error is 5%.”

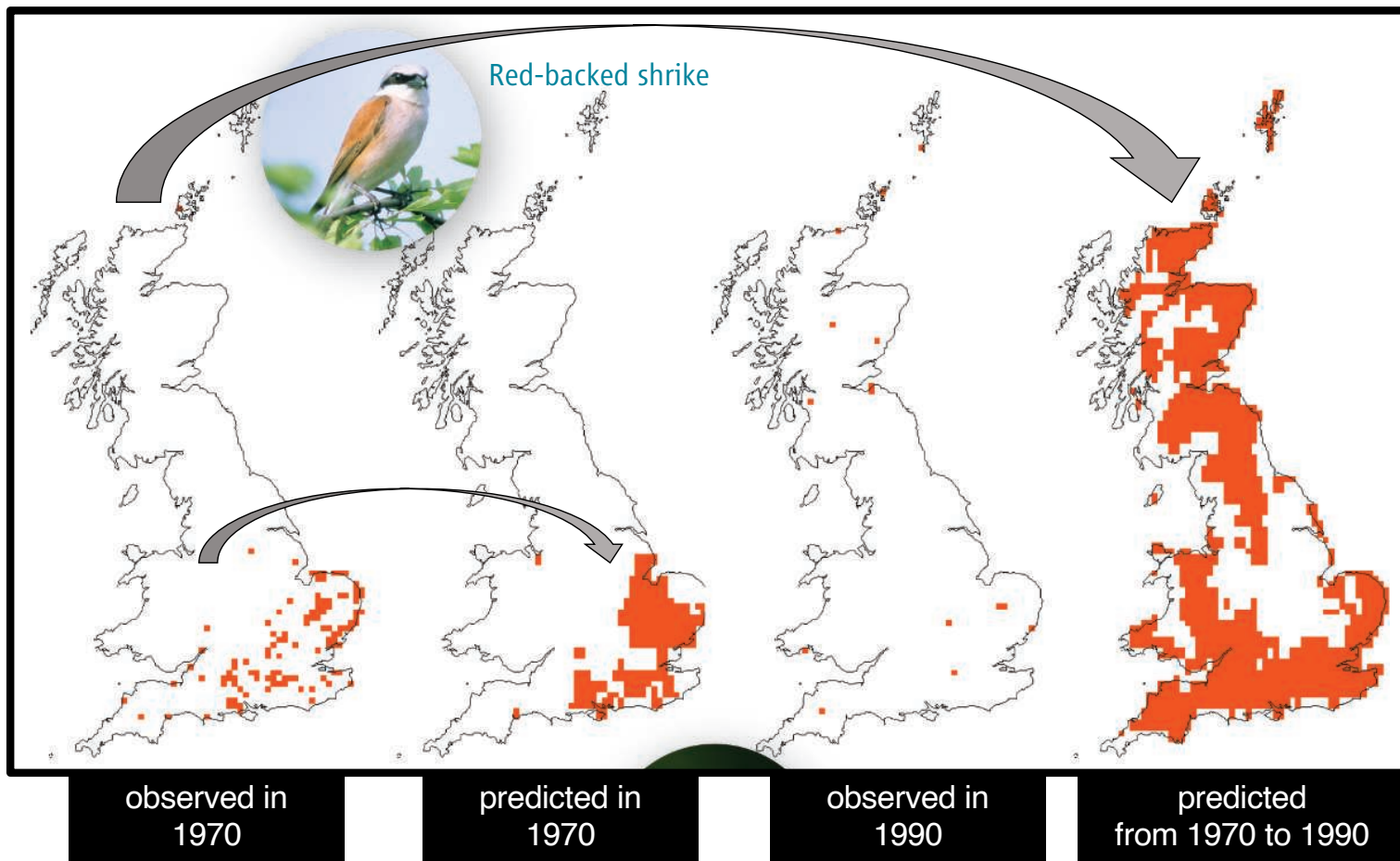
What does that mean?

Roles of statistics (3) – Produce estimates and draw conclusions from data based on samples

“There is a probability of 32% that a particular biological population will go extinct. The margin of error is 5%.”

What does that mean? (“we're pretty confident that the true probability is between $32 \pm 5\%$ or somewhere between 27% and 37%”)

Roles of statistics (4) - Predict the present or predict the future



Roles of statistics (4) - Predict the present or predict the future

BIODIVERSITY

Predicting Oblivion: Are Existing Models Up to the Task?

Huge numbers of species may be at risk of extinction from climate change, but coming up with precise estimates is proving tough

17 AUGUST 2007 VOL 317 SCIENCE

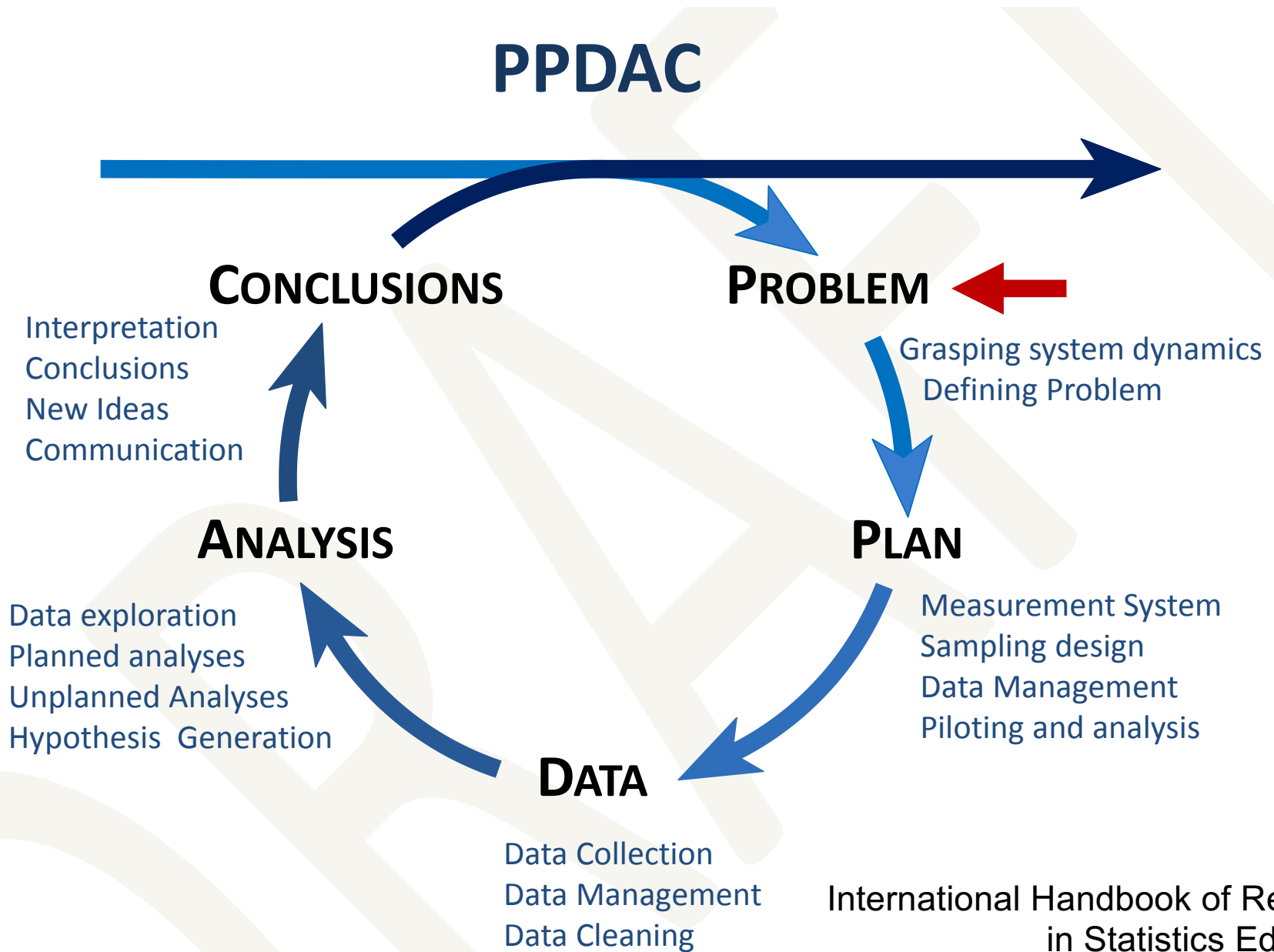
1970

1970

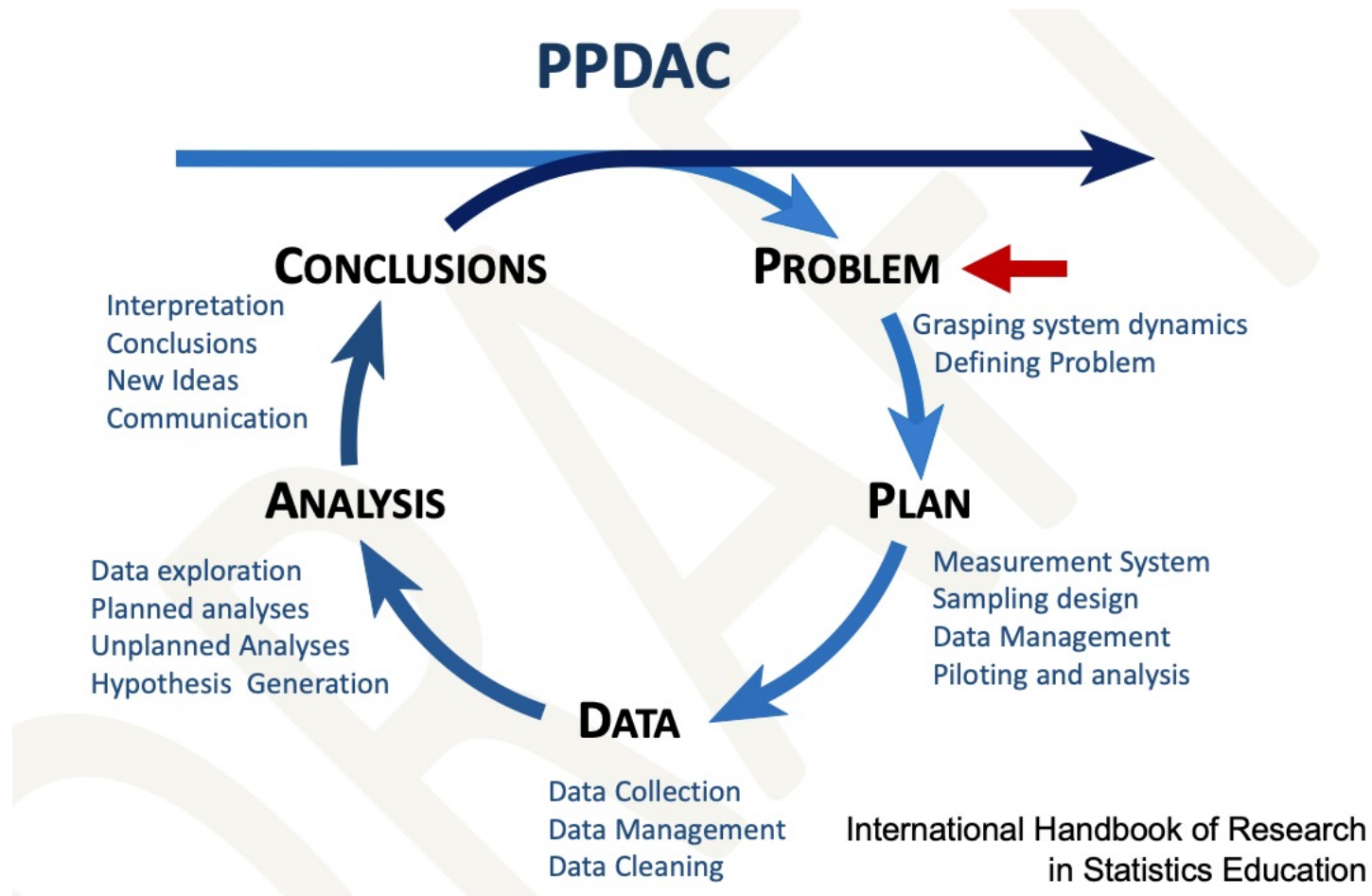
1990

from 1970 to 1990

The statistical inquiry cycle



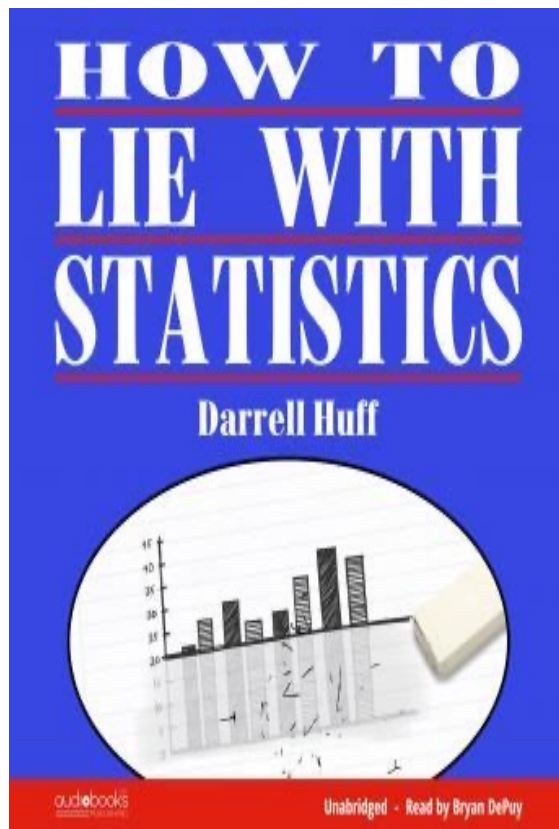
“To consult statisticians after an experiment is finished is often merely to ask them to conduct a *post mortem* examination. They can perhaps say what the experiment died of.” (paraphrased from Fisher, 1938).



We need to develop our ability to think critically and make decisions by estimating parameters and assessing probabilities.

Understanding and embracing uncertainty is crucial in accurately estimating these values.

What is the role of statistics? **Convince you and others!**



HOW TO LIE WITH STATISTICS

(Huff, D. 1954)

There are three kinds of lies: lies, damned lies, and statistics.
—Disraeli

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.
—H. G. Wells

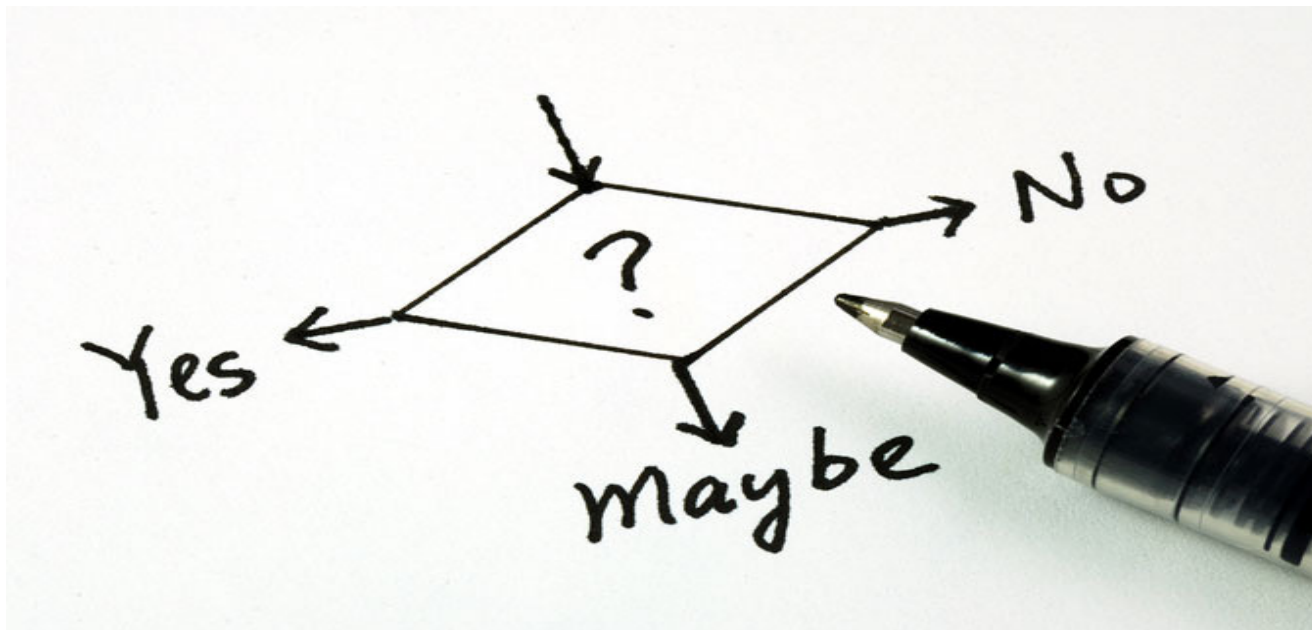
It ain't so much the things we don't know that get us in trouble.
It's the things we know that ain't so.
—Artemus Ward

Round numbers are always false.
—Samuel Johnson

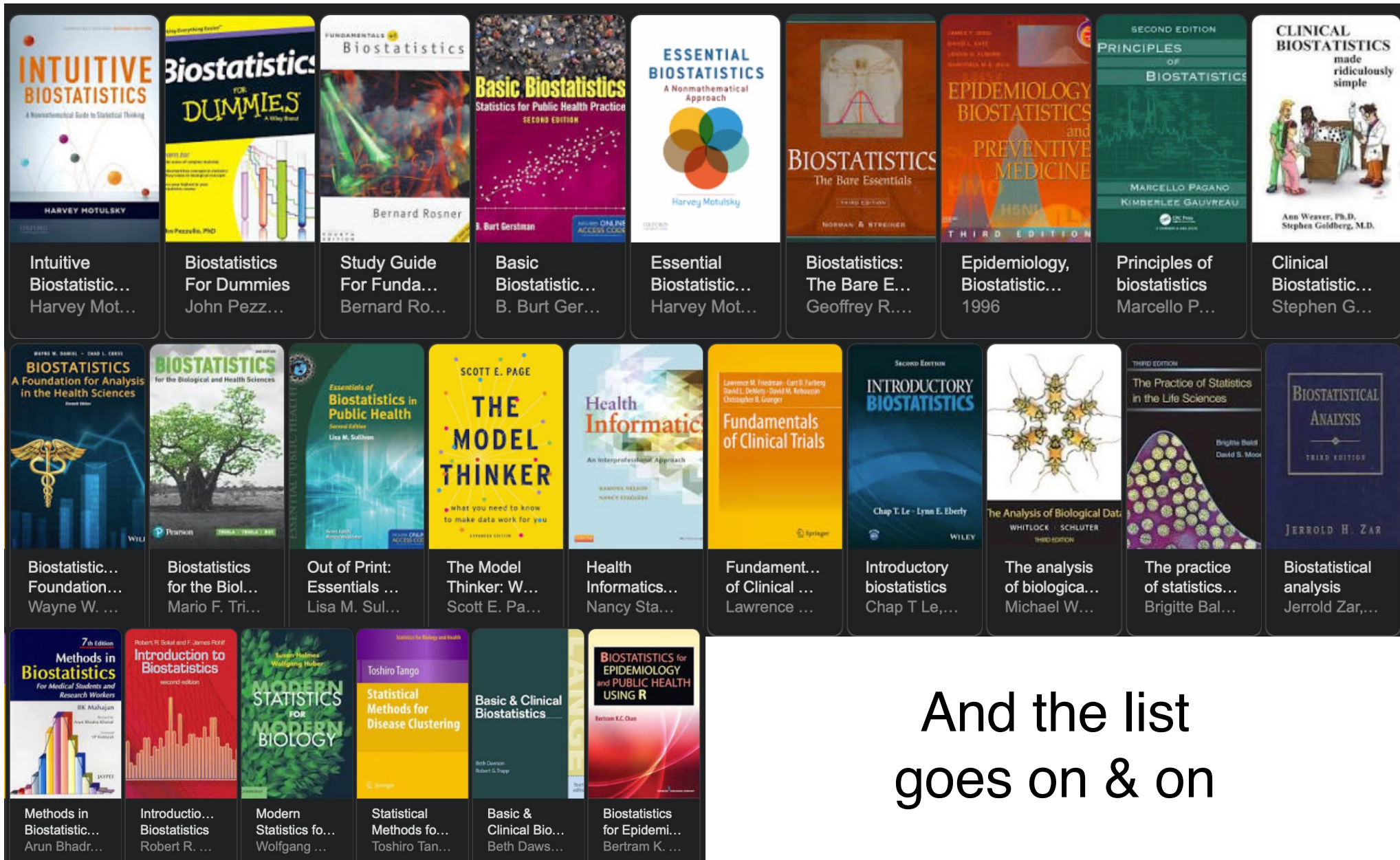
I have a great subject [statistics] to write upon, but feel keenly my literary incapacity to make it easily intelligible without sacrificing accuracy and thoroughness.
—Sir Francis Galton

What is the role of statistics? **Convince you and others (make decisions)!**

Statistics is part of a decision-making processes because most decisions are made without complete knowledge (i.e., they are based on samples and as such have some level of uncertainty).



There is a lot of BioStatistics out there



And the list goes on & on

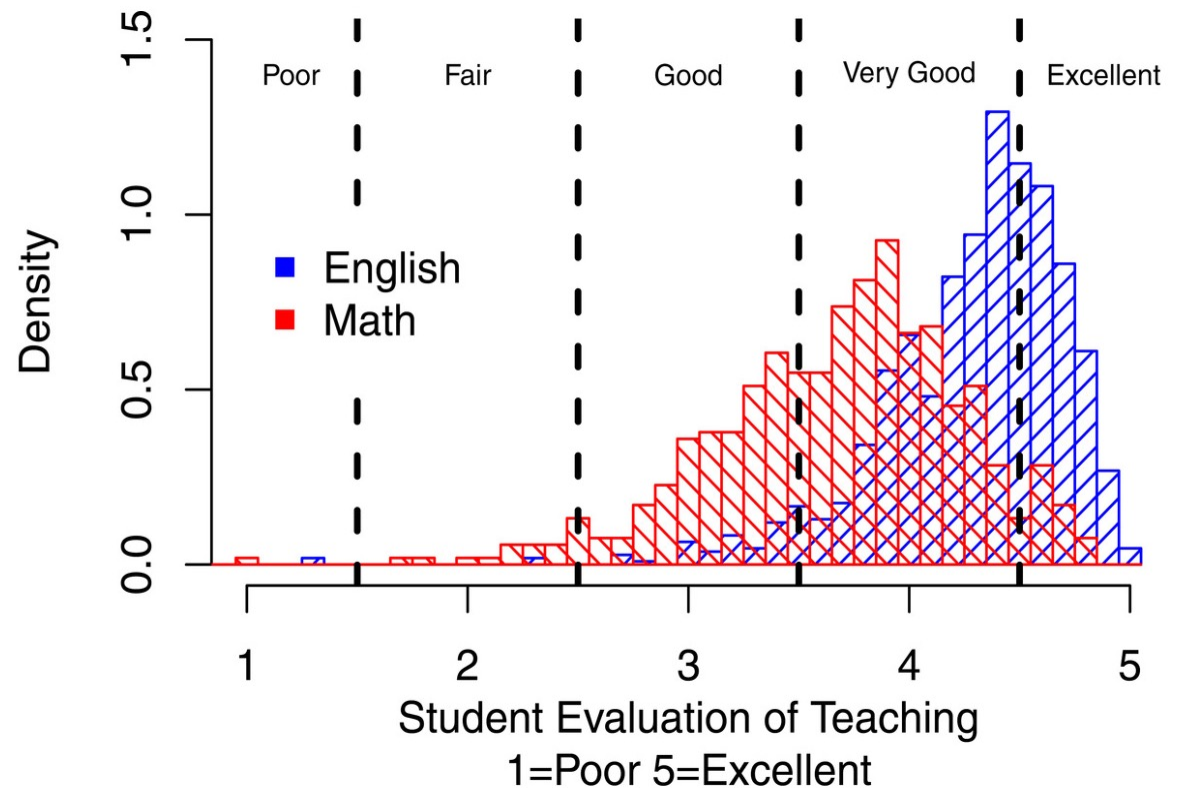
The “fear” of BioStatistics



THE ELEPHANT IN THE ROOM

Let's Talk About It

The fear of statistics



Source: <https://www.insidehighered.com/news/2017/05/10/study-student-ratings-instructors-dependent-discipline-quantitative-fields-are-most>

The fear of statistics



I Hate Statistics

Let us help you understand
the world through data

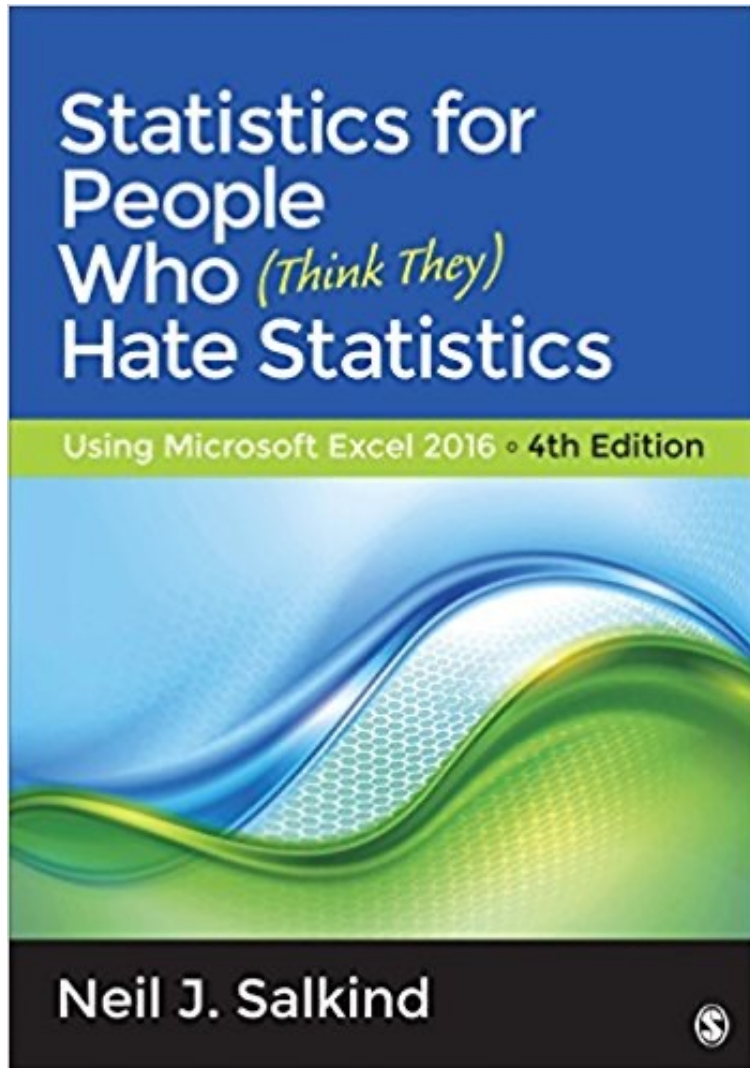
Why I Hate Statistics?

We figured we'd change the reputation of one of the world's most hated subjects: statistics! All of our friends said the same thing about going to university. They had a great time, but absolutely hated statistics. So we decided to combine the two things we like doing most: **teaching and changing the world for the better.**

We started our social enterprise to help everyone understand data and statistics. Because we believe that everyone can learn statistics. And we're here to make you have some fun along the way!

- <https://www.ihatestatistics.com/>

The fear of statistics



Bottom line:

There is fear of statistics...but *there is no need to!*

How to become interested in statistics?

Statistical concepts can promote interest, ability and intuition towards quantitative thinking & numeracy.

With time and experience, you will be able to choose or adapt available tools to particular needs....and perhaps even generate your own.

Let's take a break – 1 minute



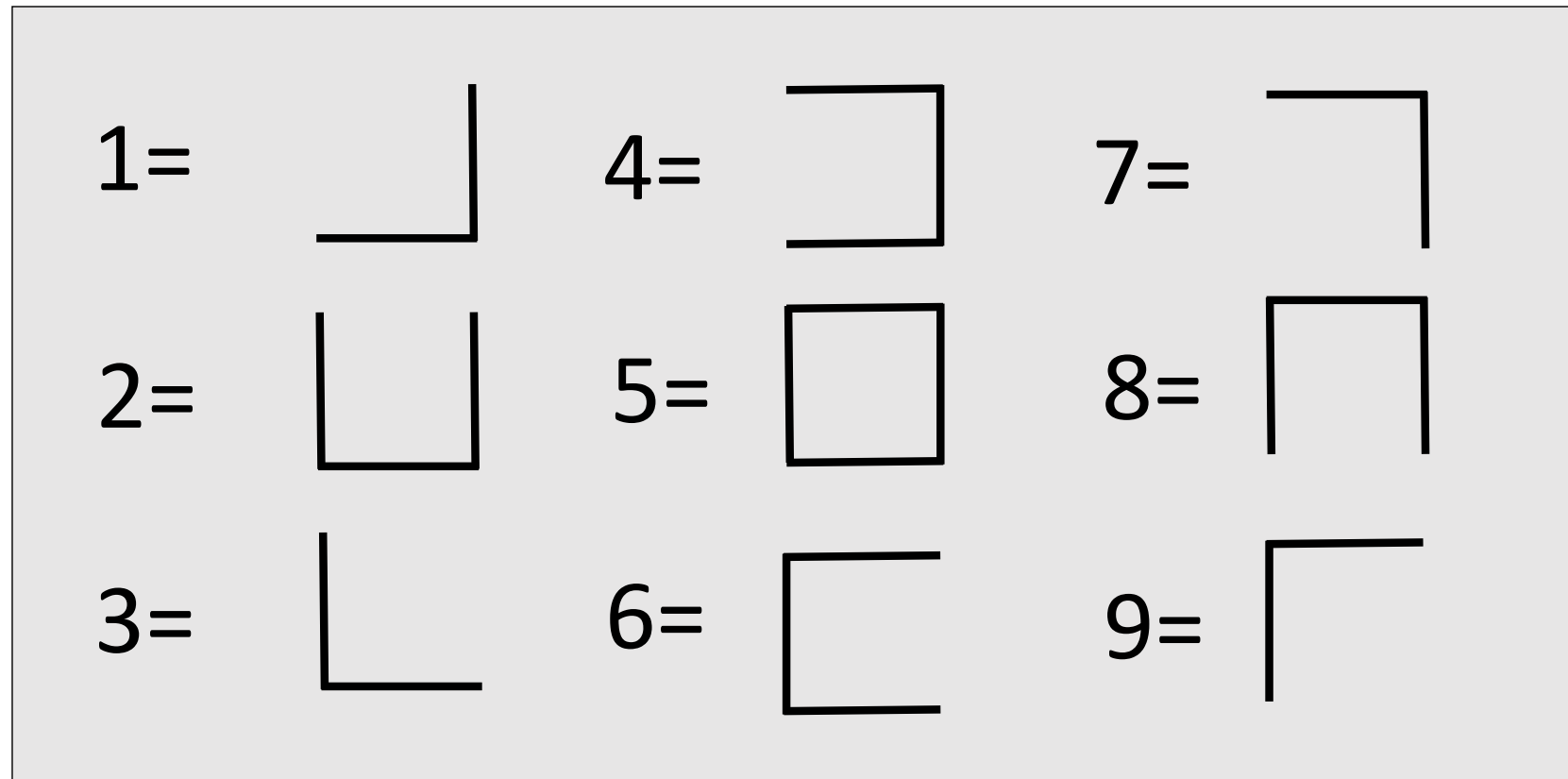
The challenges in understanding statistics

A word on intuition – two meanings:

- 1) Easy to use and understand (goal here).
- 2) “Instinctive” – act on what one feels to be true even without reason (ultimate goal of experience).

The challenges in understanding statistics

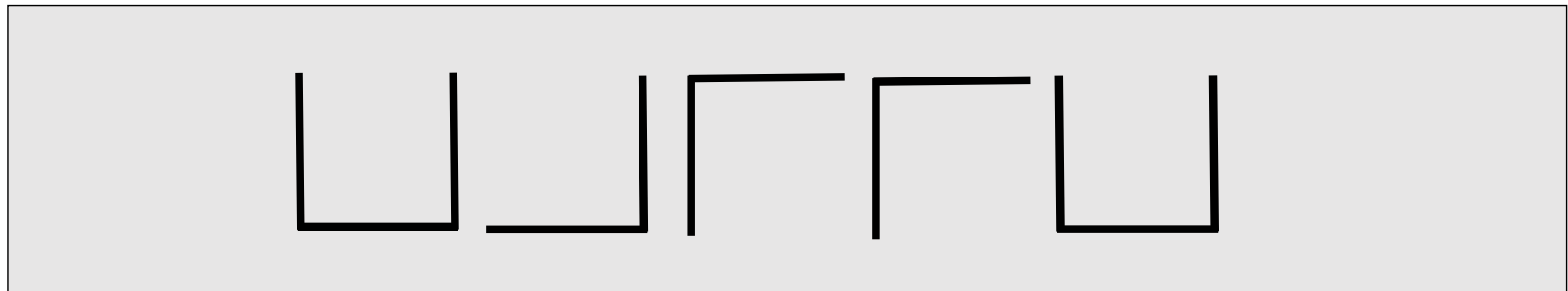
Let me describe my goal & teaching philosophy with a problem: “Here is a new number system that you will need to learn”



The challenges in understanding statistics

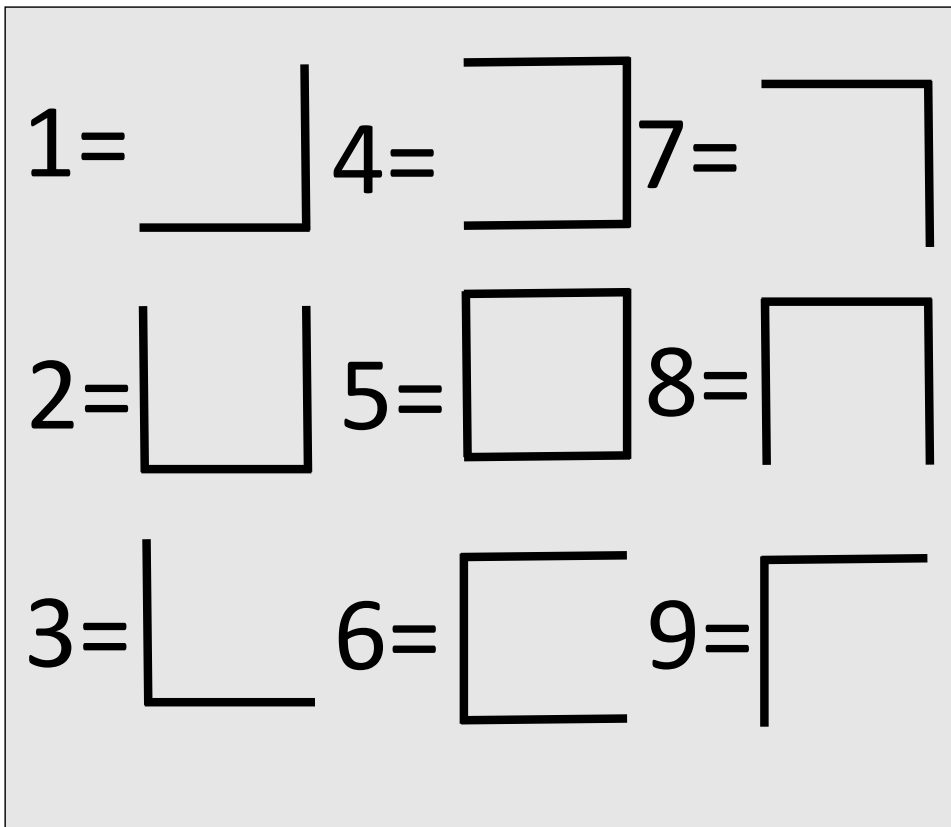
Let me describe my goal & teaching philosophy with a problem: “Here is a new number system that you will need to learn”

What is this number?



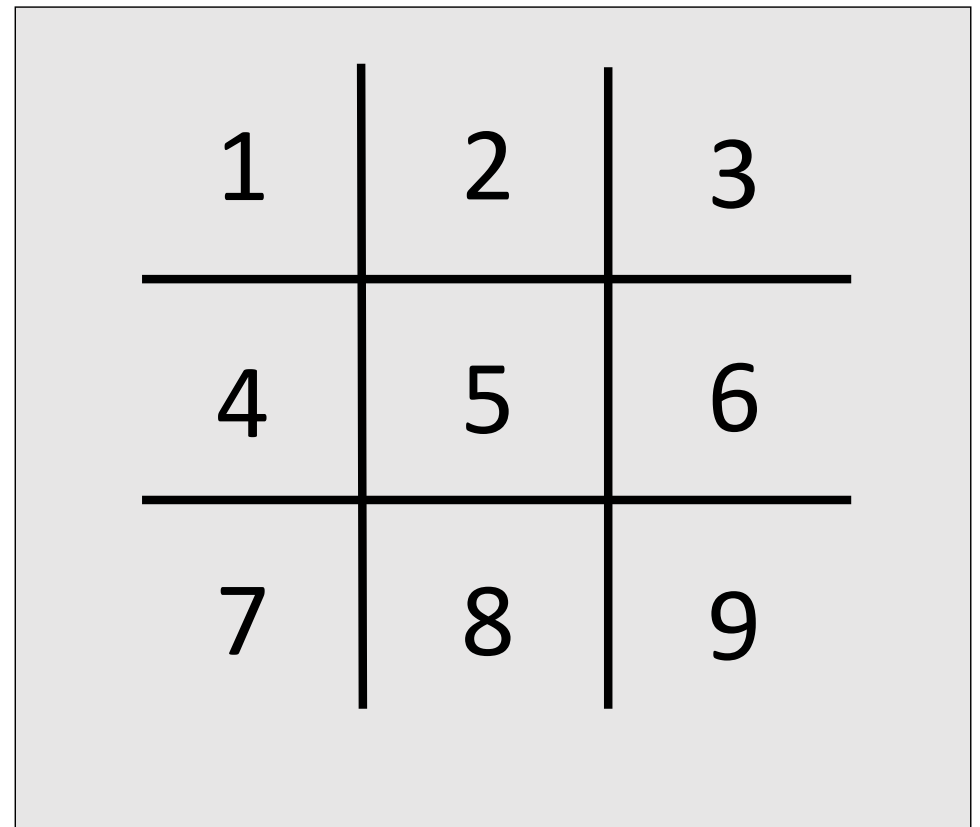
Teaching styles

Transmissionist



Unsupported content

Constructivist

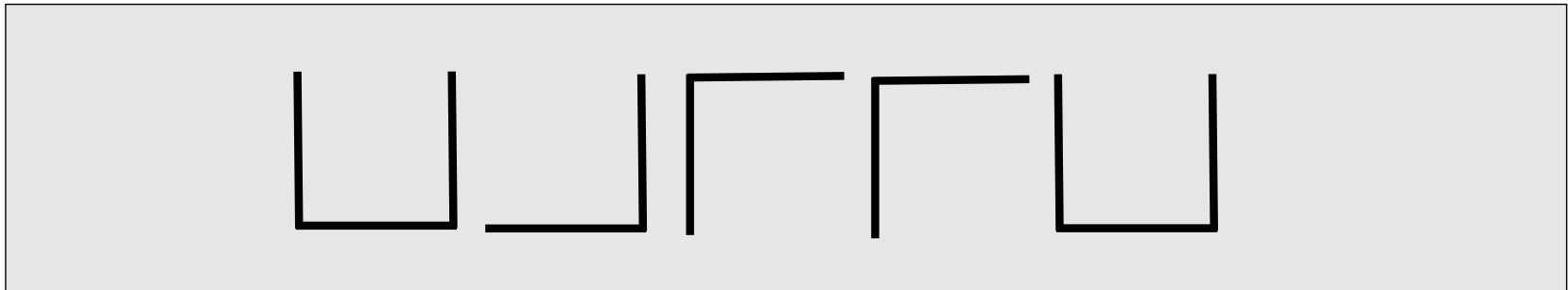


Built on familiar content

The challenges in understanding statistics

Let me describe my goal & teaching philosophy with a problem: “Here is a new number system that you will need to learn”

What is this number?

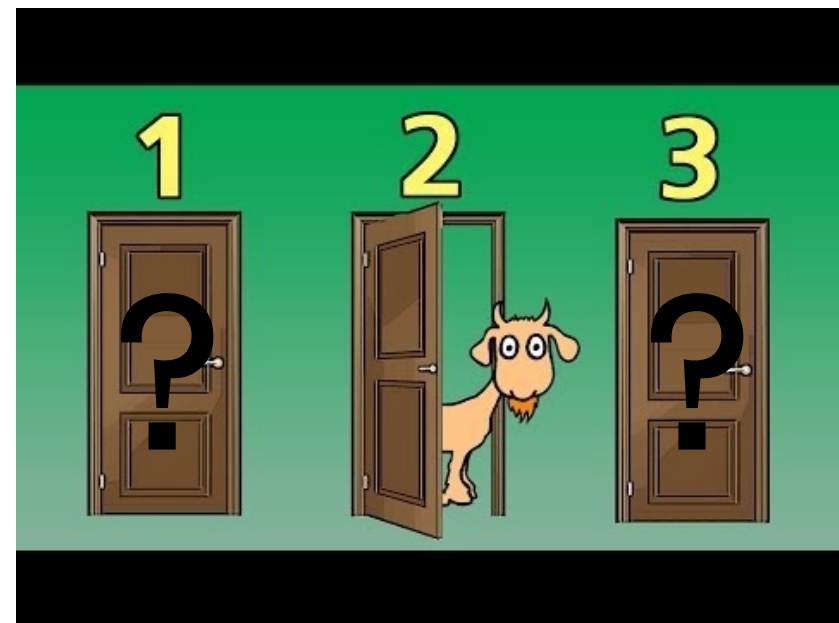


We are witnessing a pedagogical movement in statistics education aimed at shifting the focus of *instruction away from theory and recipes toward statistics thinking, genuine data, conceptual understanding, and active learning.*

Chance and Garfield, 2001

The challenges in understanding statistics - many problems are not intuitive at first

The Monty Hall Problem (from “Let’s make a deal”): In search of a new car, you pick a door, say 1. The game host then opens one of the other doors, say 2, to reveal a goat and offers to let you pick door 3 instead of door 1 if you want to. **Would you switch or keep the same door?**



Is it okay to learn statistics without memorizing the mathematical derivation of the formulas?
I like Peter Flom's answer!



Peter Flom, Independent statistical consultant for researchers in behavioral, social and medical sciences

Answered May 2, 2016

Yes, it's OK.

I've been a data analyst for 20+ years and I've never needed to know the derivation or proof of anything. In fact, you can mostly forget the formulas because they are easy to look up and, in any case, the computer does the calculating.

BUT: You do need to know a lot more than just the formulas. You need to know the assumptions, the things that can go wrong, the reasons to use or avoid a particular statistic, the alternatives that are available and so on.

Is it okay to learn statistics without memorizing the mathematical derivation of the formulas? I like Peter Flom's answer!



Peter Flom, Independent statistical consultant for researchers in behavioral, social and medical sciences

Answered May 2, 2016

Yes, it's OK.

I've been a data analyst for 20+ years and I've never needed to know the derivation or proof of anything. In fact, you can mostly forget the formulas because they are easy to look up and, in any case, the computer does the calculating.

BUT: You do need to know a lot more than just the formulas. You need to know the assumptions, the things that can go wrong, the reasons to use or avoid a particular statistic, the alternatives that are available and so on.

Of course, if you want to be a theoretical statistician and derive new statistics and stuff like that, forget all the above - you'll need to know a LOT of math.

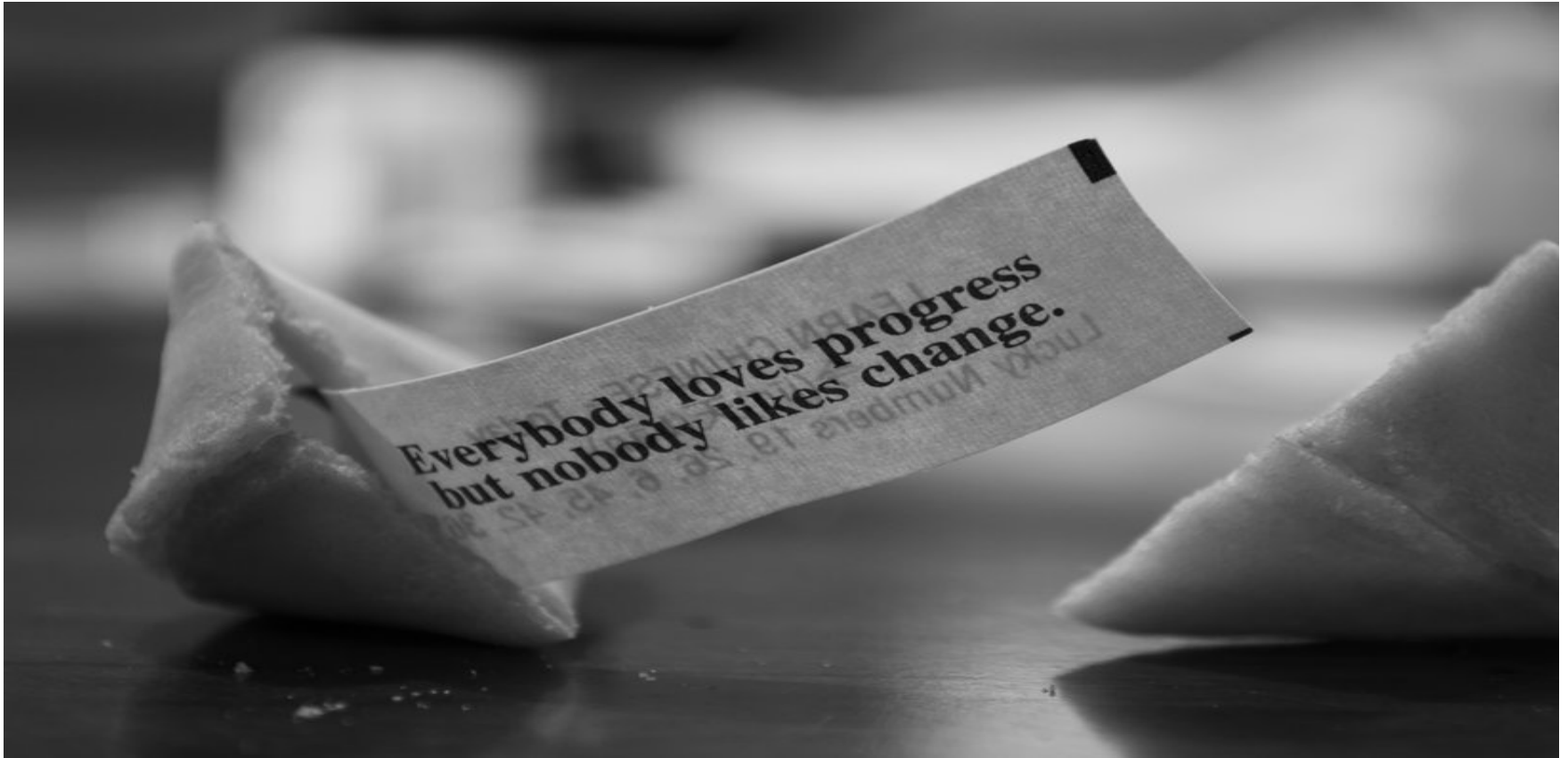
You may be able to use a car very well (or an environmental probe) even though you likely don't know how it really works!

Statistical inference and statistical hypothesis testing (not to be confused with hypothesis testing) are key statistical components that are general and can be understood without excruciating details. Understand the assumptions and limitations of methods are also analogous to driving without understanding the car's engine.



The challenges in understanding statistics

Imagination is our job



Let's connect ideas and concepts

The challenges in understanding statistics

Let's connect ideas and concepts



This is the role of tutorials & reports

Let's take a break – 1 minute

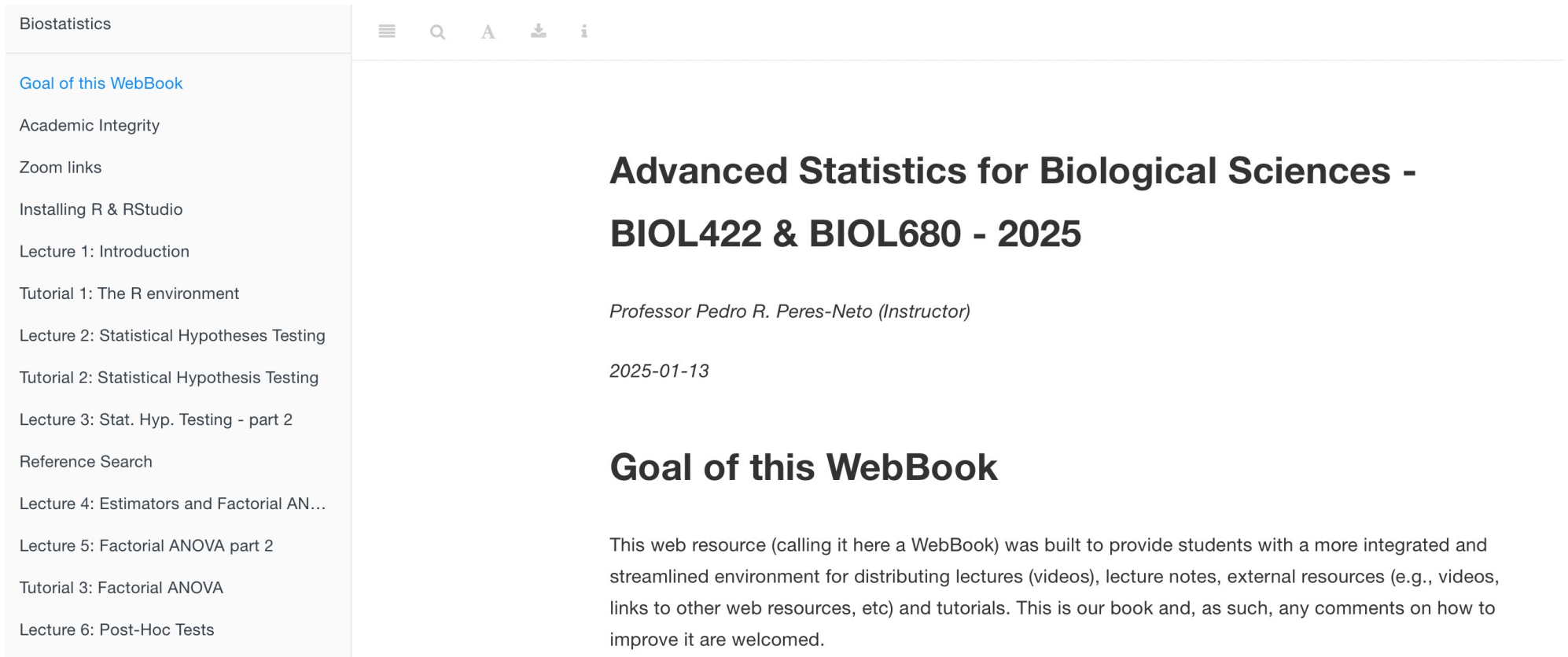




Please use our
Moodle FORUM:
the answer to your
question can help
everyone!



Our WebBook in a glance



Biostatistics

☰ 🔍 A ⬇️ i

[Goal of this WebBook](#)

Academic Integrity

Zoom links

Installing R & RStudio

Lecture 1: Introduction

Tutorial 1: The R environment

Lecture 2: Statistical Hypotheses Testing

Tutorial 2: Statistical Hypothesis Testing

Lecture 3: Stat. Hyp. Testing - part 2

Reference Search

Lecture 4: Estimators and Factorial AN...

Lecture 5: Factorial ANOVA part 2

Tutorial 3: Factorial ANOVA

Lecture 6: Post-Hoc Tests

Advanced Statistics for Biological Sciences - BIOL422 & BIOL680 - 2025

Professor Pedro R. Peres-Neto (Instructor)

2025-01-13

Goal of this WebBook

This web resource (calling it here a WebBook) was built to provide students with a more integrated and streamlined environment for distributing lectures (videos), lecture notes, external resources (e.g., videos, links to other web resources, etc) and tutorials. This is our book and, as such, any comments on how to improve it are welcomed.

http://mypage.concordia.ca/faculty/pperesne/BIOL_422_680/

Let's be personal!

Instructors enjoy being greeted cordially; for example:

Hello Pedro (I'm very informal)

Hello Dr. Peres-Neto; or Hello Professor Peres-Neto

Hello could be replaced by Hi or Dear depending on the occasion.

Perhaps avoid being impersonal:

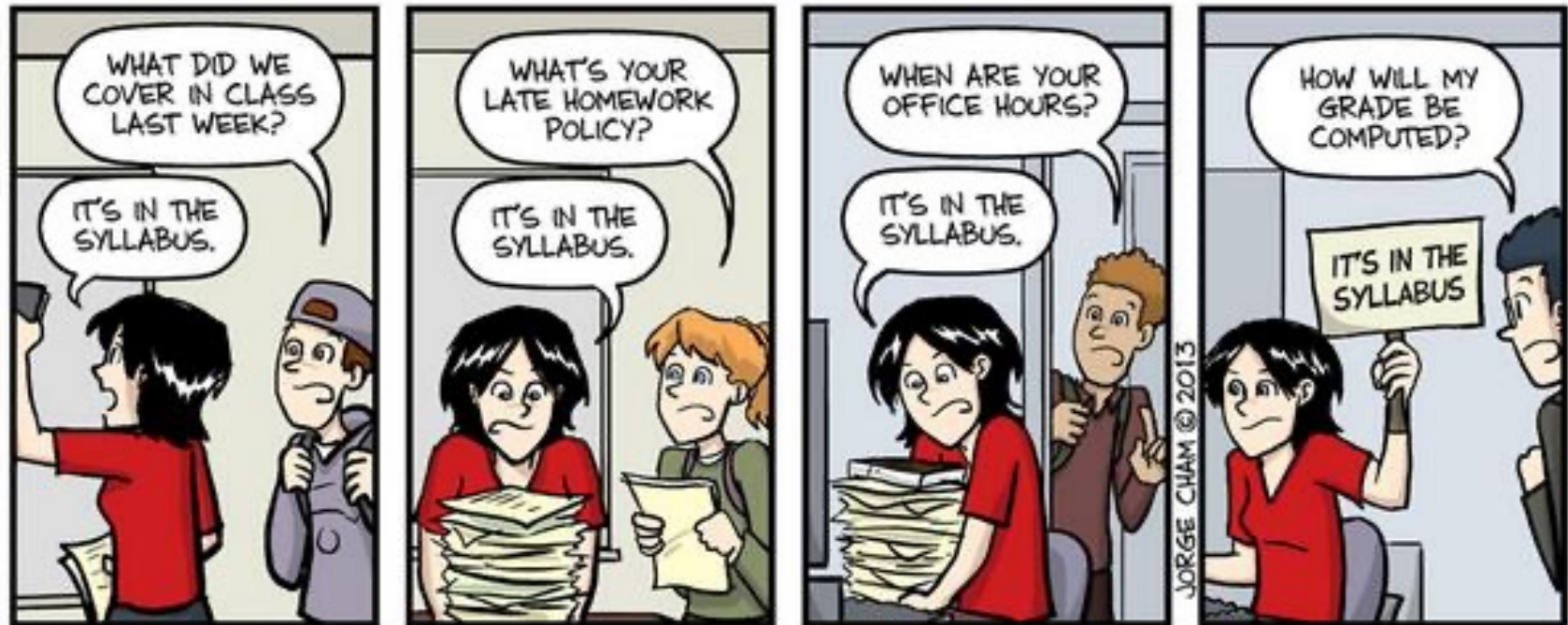
Hello,

Hi,

Hello sir "If you forgot your instructor's name, please look over their course syllabus." I know is a sign of respect, but I do prefer to be referred by my name.

Thank you 😊 - I may send a reminder about this from time to time!

Read the Syllabus



IT'S IN THE SYLLABUS

This message brought to you by every instructor that ever lived.

WWW.PHDCOMICS.COM

But if it's not there...let me know

Academic Integrity

Sharing any assignment (e.g., exams, quizzes, reports, etc) and course material/content in public sites and with other students from Concordia or other institutions goes against academic integrity and can lead to sanctions and infraction. Please visit the Concordia website on academic integrity and contact me if you have any questions: <https://www.concordia.ca/conduct/academic-integrity>

PLEASE RESPECT THE WORK OF YOUR INSTRUCTORS

Sharing course material and assignments also demonstrates a lack of respect for the work of your instructors that put a lot of their time into your education.

WHAT YOU CAN and CAN'T DO in ASSIGNMENTS and EXAMS?



PLAGIARISM

The most common offense under the Academic Code of Conduct is plagiarism which the Code defines as “the presentation of the work of another person as one’s own or without proper acknowledgement.” This could be material copied word for word from books, journals, internet sites, professor’s course notes, etc. It could be material that is paraphrased but closely resembles the original source. It could be the work of a fellow student, for example, an answer on a quiz, data for a lab report, a paper or assignment completed by another student. It might be a paper purchased through one of the many available sources. Plagiarism does not refer to words alone - it can also refer to copying images, graphs, tables, and ideas. “Presentation” is not limited to written work. It also includes oral presentations, computer assignments and artistic works. Finally, if you translate the work of another person into French or English and do not cite the source, this is also plagiarism.

In simple words: DO NOT COPY, PARAPHRASE OR TRANSLATE ANYTHING FROM ANYWHERE WITHOUT SAYING FROM WHERE YOU OBTAINED IT! Source: <https://www.concordia.ca/students/academic-integrity.html>

Academic Integrity

Sharing any assignment (e.g., exams, quizzes, reports, etc) and course material/content in public sites and with other students from Concordia or other institutions goes against academic integrity and can lead to sanctions and infraction. It also demonstrates a complete lack of respect for the work of professors and instructors.

We faculty put 100s of hours every semester for each course and many of these hours are taken away from our personal time, including time we don't spend with our families. Please respect us and the time we dedicate to your training.



Sharing any material is a high academic infraction that can have consequences to your records even if found out after your graduation.

Please visit the Concordia website on academic integrity and contact me if you have any questions:

<https://www.concordia.ca/conduct/academic-integrity>

PLEASE RESPECT THE WORK OF YOUR INSTRUCTORS

Sharing course material and assignments also demonstrates a lack of respect for the work of your instructors that put a lot of their time into your education.

The use of ChatGPT and other AI systems for natural and computer languages

The use of ChatGPT and other AI systems is prohibited unless clearly stated in the assignment instructions.

Any submission that is found to have used ChatGPT or other AI systems without explicit permission will be considered a violation of the academic integrity policy. It is the responsibility of the student to ensure that all work submitted is their own original creation, and to seek clarification from the instructor if there is any doubt about the use of ChatGPT and other AI systems. We will have open discussions about these tools and the ethical implications of these to learning and assessment.



Subjects covered in the course



Biological problems and associated data.

R for statistical computing.

Data structure and types of statistical variables.

Field versus laboratory studies, experimental versus observational studies.

The concepts of probability, parameters and maximum likelihood; revisiting inferential statistics and statistical hypothesis testing.

Revisiting Analysis of Variance (ANOVA) – parametric and non-parametric.

Advanced Multiple testing and post-hoc analysis.

Multifactorial Analysis of Variance.

Analysis of Covariance (ANCOVA).

Fixed versus random factors: mixed model ANOVA.

Multiple regression and variation partitioning.

Generalized Linear Models (GLMs); spatial and phylogenetic autocorrelation: generalized least square solutions.

Multivariate analyses: introduction and the concept of latent variables and processes.

Multivariate inference: Multivariate Analysis of Variance (MANOVA) and Discriminant Function Analysis (DFA).

Multivariate analyses: Principal Component Analysis (PCA), Principal Coordinate Analysis (PCoA) and Correspondence Analysis (CA).

Multi -response multiple regression: Redundancy Analysis (RDA), relating species characteristics to their environments.

Cluster Analysis, Machine learning, Classification and Regression Tree (CART), and K means.

Advanced non-parametric inference: Monte Carlo testing and bootstrap.



Moodle will be used for sending announcement,
Forums and posting assignments



Tutorial 1: The R environment

January 10, 2023 (1st week of classes)

CRITICAL: Tutorials have no reports or grades attached to them. Note though that reports and midterm 2 are heavily based on tutorials and your knowledge of R. Therefore you should attend tutorials and/or work on them weekly on your own time.

Your TAs

Section 101 or 201 (14:00h-16:00h): Aliénor Stahl (a.stahl67@gmail.com)

Section 102 or 202 (16:15h-18:15h): Gabrielle Rimok (gabrielle.rimok@mail.concordia.ca)

General Information

This tutorial is meant to help you get acquainted (or reacquainted) with the R environment for statistical computing and its basic commands, ways to handle data and plot graphs. If you are familiar with it, you can skip it.
