1

- Will it rain tomorrow?
  (a deterministic question)
- What is the probability that it will rain tomorrow?
  (a question framed in terms of uncertainty)

How do these two questions differ, and which one is more in line with a statistical question?

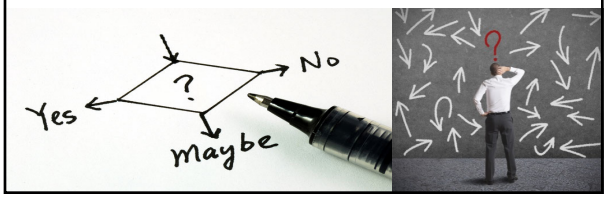
2

Key concepts underlying statistics and statistical thinking

- Decision (e.g., based on probability of raining).
- Uncertainty (unpredictability).
- Risk of being wrong (error).
- Variability - Answer (estimates of the probability of raining) may change with more data (preferably towards more accurate answers, i.e., probabilities).
- Accuracy (close to reality, i.e., yes/no rain; models predicts correctly).
- Knowledge (accumulation of evidence, i.e., that the model that we used to predict rain becomes more and more accurate).

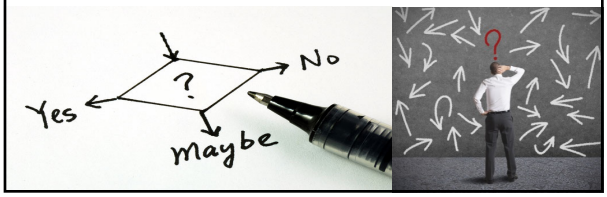
3

The science of supporting decision-making under incomplete information / knowledge



4

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



5

Biostatistical decision-making without all information (incomplete knowledge):

Deciding whether to stop a clinical trial early

A hospital is testing a new antibiotic to reduce post-surgery infections. After 120 patients (out of a planned 400), the interim results are:

New drug: **8 infections / 60 patients** (13.3%)
Standard care (current drug): **16 infections / 60 patients** (26.7%)

The question is a *decision:* **Should the trial be stopped early and the drug adopted, or should it continue?**

6

Biostatistical decision-making without all information

The question is a *decision:* **Should the trial be stopped early and the drug adopted, or should it continue?**

Aspects involved in the decision:

- Estimating *how big the effect might be (i.e., difference in % infection),*
- Showing *how uncertain we are* about the % difference
- Helping us *anticipate what might happen* if we collect more data (more patients),
- Clarifying trade-offs between acting now and waiting longer means weighing the risk of acting with limited evidence against the cost of delaying a potentially beneficial decision.

So the decision is made using probabilities and uncertainty because we don't (yet) know the true effect.

7

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!

8

"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)

Statement framed as absolute (*misleading*): "Species X will go extinct within the next 50 years."

Scientific statement (*what we actually know*): "Based on current population trends, habitat loss rates, and climate projections, Species X has a high probability (63%) of extinction within the next 50 years, with uncertainty depending on future environmental and conservation scenarios."

In biology, we rarely say what will happen; we say how confident we are that it might happen, given what we currently know.

9

Welcome to BIOL 322

Statistics for Biological Sciences (BioStatiscs)

Pedro Peres-Neto, PhD

Professor & Canada Research Chair,
Department of Biology,
Concordia University

10

Don't hesitate to raise your hand during lectures if you have any questions.

I also try to "read the room" and will step in when I sense that students may be unsure or have questions they haven't yet voiced.

11

Statistics is key!

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

- Herbert George Wells

12

Statistics is key!

"Mathematics may rule the universe; statistics rules society."

- An inspiring moment during a BIOL322 lecture in 2018

13

Statistics is key!

Biostatistics is arguably the most general and widely applicable course you will take during a biology degree.

Biostatistics doesn't just help you analyze data - it changes how you think about evidence.

It is critical to every discipline, biological OR NOT.

Biology is now a data-driven science - from genomics to ecology, every field relies on statistical inference.

Biostatistics teaches you how to separate patterns from noise rather than trusting intuition or anecdotes.

You learn how to design better experiments - saving time, money, samples, and effort.

It makes you a critical consumer of research - you can spot weak claims, flawed methods, or overinterpretation.

Employers and graduate programs value it - quantitative and analytical skills consistently rank as top hiring criteria.

14

What is a statistical question?

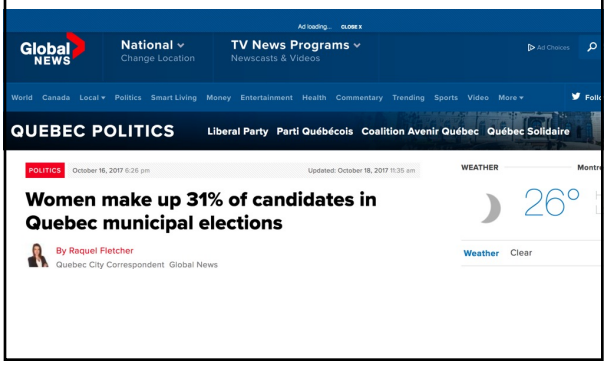
✓ What is the average size of Canadians?

✓ Is 10 a number?

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

15

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



16

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

NO! This question can be answered simply by counting the number of female candidates relative to the total number of candidates. It is not a question that would be resolved by collecting additional data; doing so would not change the result.

17

ARTICLE

Mapping tree density at a global scale

A study led by Yale University researchers has found that there are over 3 trillion trees on Earth - but they are disappearing at an alarming rate.

The study found that there are around 3.04 trillion trees on Earth, or around 422 for each person on the planet.

The number is a huge increase on the previous global estimate, which was just over 400 billion trees worldwide.

The study was based on on-the-ground data about the number of trees in more than 400,000 plots of forest from all continents except Antarctica.

Source - https://www.independent.co.uk/environment/how-many-trees-are-there-on-earth-10483553.html

18

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

We should become comfortable with the idea that the most interesting and useful results may evolve as new information (data) becomes available.

⬇

Statistics: "The science of supporting decision-making under incomplete information."

19

Which one is a statistical question?

What percentage of students in today's class checked their cell phone by the end of the lecture?

What percentage of students across Montreal universities check or use their cell phones during class?

20

Some initial thoughts on learning/teaching philosophy

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

21

Let's take a break – 1 minute

22

Statistics is a science in its own!

"Statistics is a science, not a branch of mathematics, but uses mathematical models as essential tools." John Tukey

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)

23

Statistics is a Mathematical Science

(not a branch of Mathematics)

- We use the singular *is* and not the plural *are* to emphasize that statistics is a field of study, not just a "bunch" of methods.
- We use *mathematical* as an adjective because although statistics certainly makes use of much mathematics (another discipline), it is a separate discipline and not a branch of mathematics.
- We use the noun *science* because statistics is the science of gaining insight from data.

- From "Some Important Comparisons between Statistics and Mathematics, and Why Teachers Should Care" by Rossman, Chance, and Medina (2006).

24

Statistical Thinking versus Mathematical Thinking

Mathematics is, by and large, a ***deterministic*** way of thinking. The way it is often taught in schools reinforces a deterministic view of the quantitative world; for example:

What is the size of our planet?

Statistics, by contrast, is fundamentally a ***probabilistic*** (or stochastic) way of thinking, explicitly accounting for uncertainty; for example:

What is the probability that it will rain tomorrow?

25

Roles of statistics

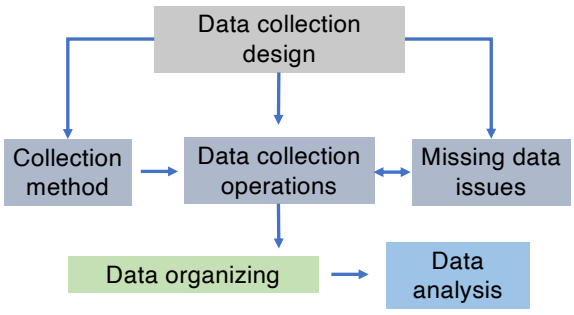
Statistics is a discipline that:

1) Designs data collection protocols (observational and experimental).
2) Summarizes information to aid understanding.
3) Draws conclusions from data.
4) Communicate uncertainty.
5) Estimates the present or predict the future.

- adapted from: http://www.scc.ms.unimelb.edu.au/whatisstatistics/

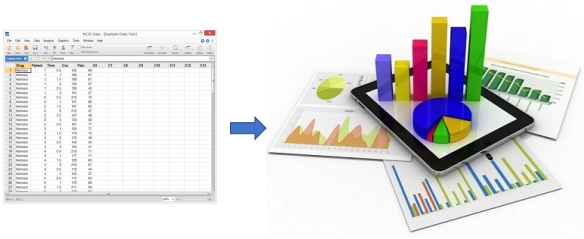
26

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



27

Roles of statistics: [2] - Summarize information to aid understanding



From raw (primary) data…….to summaries

28

Roles of statistics: [3] & [4] - Draw conclusions from data & communicate uncertainty (estimate error)

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

What does that mean?

- Source - http://www.scc.ms.unimelb.edu.au/whatisstatistics/ssize.html

29

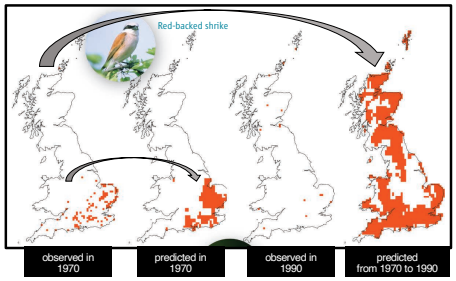
Roles of statistics: [3] & [4] - Draw conclusions from data & communicate uncertainty (estimate error)

"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 ± 5% or somewhere between 27% and 37%")

30

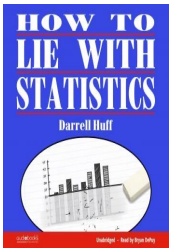
Roles of statistics: [5] - Estimate the present or predict the future



Araujo and Rahbeck (2006, Science)

31

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

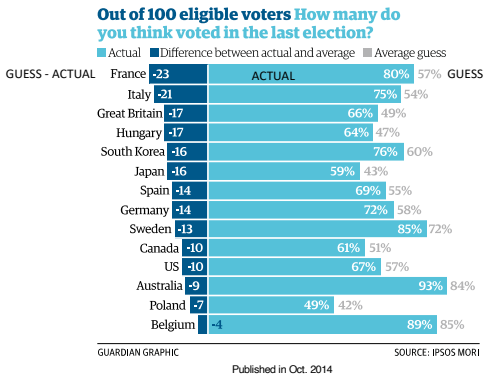
It ain't so much the things we don't know that get us in trouble. It's the things we know that ain't so.
—Artemus Ward

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

32

We often make decisions based on wrong guesses!

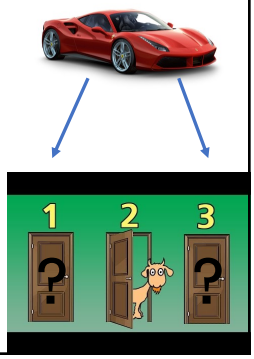


Published in Oct. 2014

33

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

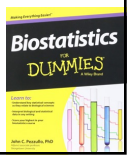
The Monty Hall Problem (from "Let's make a deal"): In search of a new car, you pick a door, say 1. The game host then opens one of the other doors, say 2, to reveal a goat and offers to let you pick door 3 instead of door 1 if you want to. ***Would you switch or keep the same door?***



34

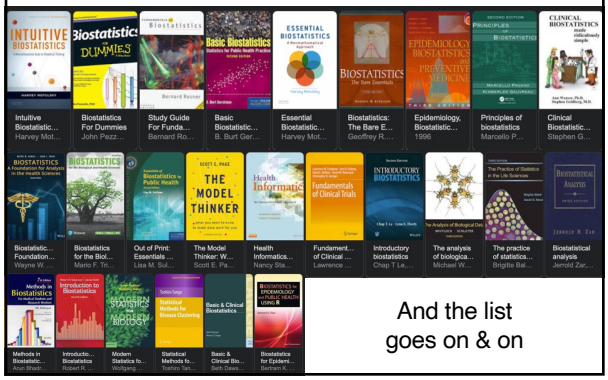
Biostatistics is the branch of statistics that focuses on analyzing data arising from biological systems.

Biostatistics is an extremely active field of statistics



35

There is a lot of BioStatistics out there



And the list goes on & on

36

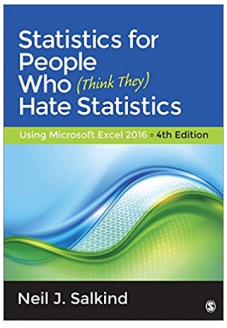
Let's take a small break – 1 minute

37

The fear of statistics (bio or not)

38

The fear of statistics



Bottom line:

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

39

The challenges in understanding numbers and numerical thinking

"Statistics is not primarily a matter of plugging numbers into formulas and performing rote computations.

It is a way of questioning and thinking that may be unfamiliar to many of us but is available to almost all of us."

- Adapted from John Allen Paulos, A Mathematician Reads the Newspaper.

40

The challenges in understanding statistics

My goal & teaching philosophy with a single problem: "Pretend that you need to learn a new number system"

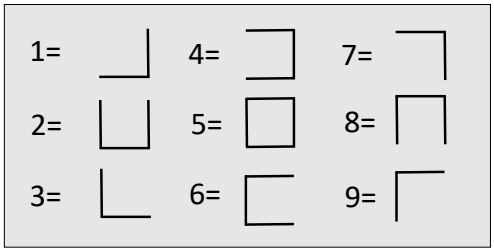


Image by P. Newbury

41

The challenges in understanding statistics

Let me describe my goal & teaching philosophy with a single problem: "Pretend that you need to learn a new number system"

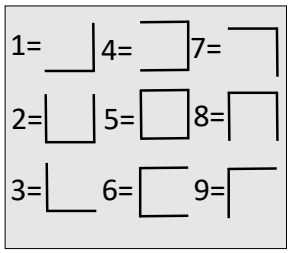
What is this number?

Image by P. Newbury

42

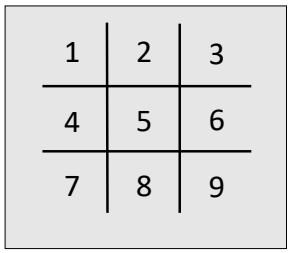
Teaching styles

Transmissionist



Unsupported content

Constructivist



Built on familiar content

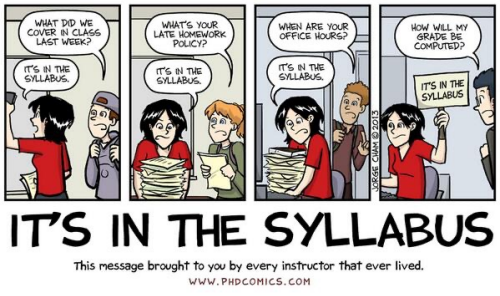
43

Let's work together!

Communication, communication & communication

44

PLEASE READ THE SYLLABUS



But if it's not in there…let me know

45

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

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

46

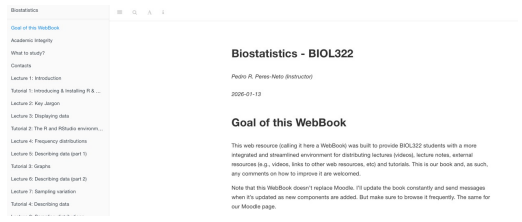
Moodle will be used for sending announcements, Forums and posting assignments

- Announcements
- Syllabus — PDF document
- WebBook
- FORUM for Content and Lectures only (not tutorials, not reports)
- Forum for Tutorials
- Forum for Exams (Midterms and Final)
- Forum for Reports 1, 2 & 3
- FORUM for Quizzes - only discuss closed (not ongoing) quizzes
- FORUM - Varia - share anything

47

WebBook - This web resource was designed to provide BIOL322 students with an integrated and streamlined platform for accessing lecture videos, lecture notes, tutorials, and external resources (e.g., videos and links).

This is *our* book, and your feedback and suggestions for improvement are always welcome.



48

Sharing any course material is an academic infraction

Sharing course materials, including but not limited to slides, tutorials, reports, assignments, recordings, and any other educational content provided in this course, **is strictly prohibited and considered an academic infraction**. **See Syllabus and material posted in Moodle & the WebBook** for more information.

Generative AI like ChatGPT - policy for BIOL322 Biostatistics

AI tools are NOT allowed for your assignments unless explicitly permitted; if so, it will be indicated in the assignment. If they are used, unless permitted in the assignment, it will be considered as an **academic misconduct.**

Why shouldn't you use generative AI for your assignments?
See Syllabus and material posted in Moodle & the WebBook

49

What to study?

1) What do you need to study for assignments and exams?

Answer: You are responsible for all material listed under the lecture chapters in the WebBook. This includes lecture slides (which can be downloaded from the WebBook), written explanations, figures, and videos.

2) Do exams require detailed knowledge of equations?

Answer: No. You will not be expected to memorize equations in detail. However, exams and assignments will include some calculations that can be done by hand. Examples of the types of calculations you should be comfortable with will be provided in quizzes and practice questions.

3) Do exams cover detailed knowledge of R? NO!

4) Do lectures cover material that is not included in the lecture slides?

Answer: Yes. Lectures and their accompanying tutorials are designed to work together to deepen understanding and reinforce key concepts, going beyond what is explicitly presented on the slides. While the slides are comprehensive, they are not intended to serve as a standalone textbook. In short, active participation in both lectures and tutorials is essential for success in this course.

50

Why drop-in hours?
(applies to any course)

Seek clarification on topics that were unclear in lectures or readings.

Ask questions about assignments to ensure understanding.

Share and exchange ideas on how to improve the course experience for yourself and others.

Build a rapport with the instructor, which can help in finding an honors thesis advisor and obtaining strong letters of recommendation.

51

How to Say Hello (It Matters!)

Instructors enjoy to be greeted cordially; for example:

Hello Pedro
Hello Dr. Peres-Neto; or Hello Prof. Peres-Neto
Hello could be replaced by Hi or Dear depending on the occasion.

Try to avoid being impersonal in messages:

Hello,
Hi,
Hello sir/Hi sir - "If you forgot your instructor's name, then please look over the course syllabus."

Thank you ☺

52

Some closing thoughts for today

"Understand. Don't memorize. Learn principles, not formulas."

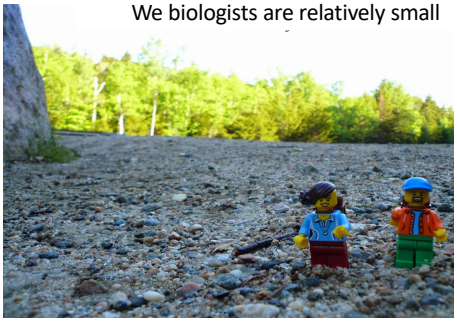
"Learning is underrated. Grades are overrated"

Richard Feynman

53

Next class - Sampling

We biologists are relatively small



From Chris Lortie

54