- 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?

Key concepts underlying statistics and statistical thinking

Decision: What we decide to do based on the probability of rain (e.g., whether to bring an umbrella).

Uncertainty: How unsure we are about whether it will rain (e.g., a 60% probability of rain means that rain is possible, but not certain).

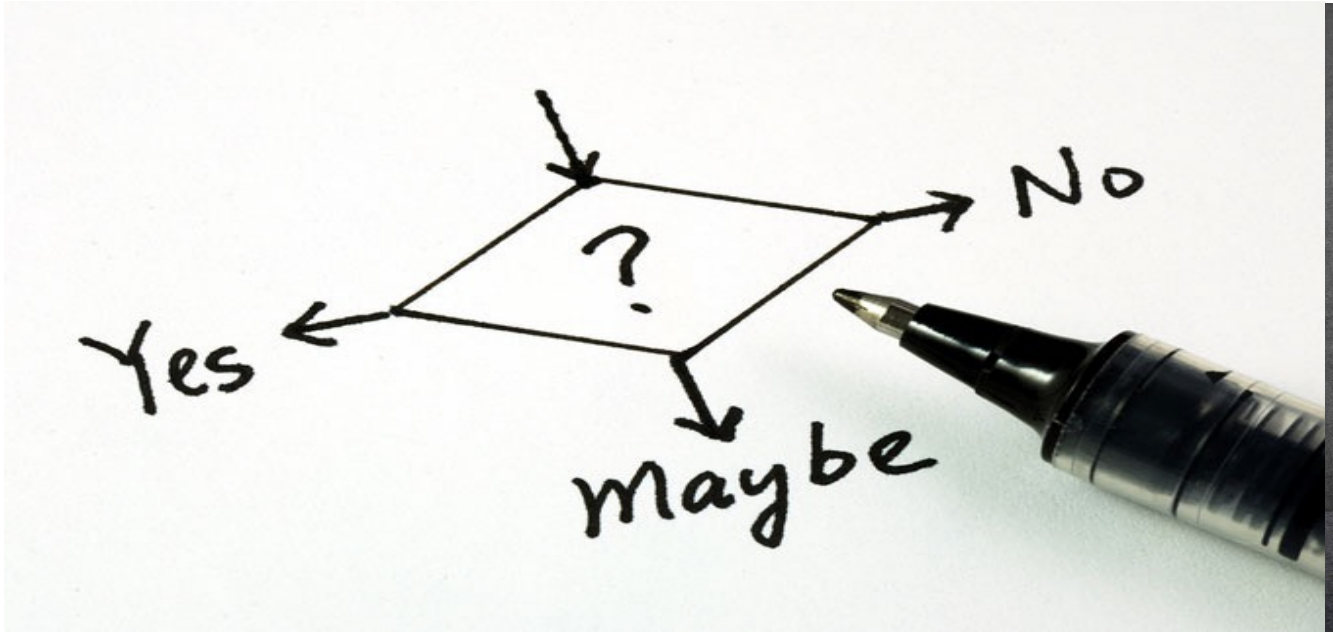
Risk of being wrong (error): The possibility that our prediction differs from what actually happens (e.g., predicting rain when it does not rain).

Variability: How much the estimated probability of rain can change across samples, observations, or forecasts (e.g., different data may produce somewhat different estimates of the probability of rain).

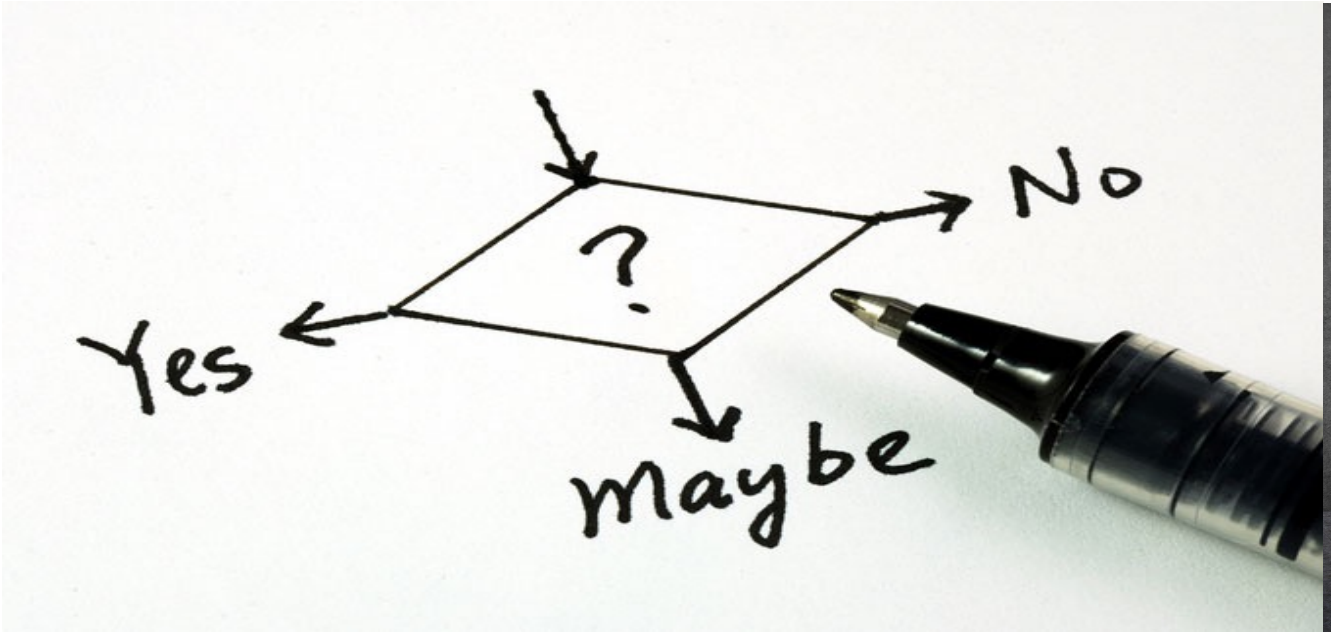
Accuracy : How closely our predictions correspond to what actually happens (e.g., how often predicted probabilities of rain agree with observed rainfall).

Knowledge: The accumulation of evidence about the processes that determine rainfall, allowing us to improve our estimates, predictions, and decisions over time.

The science of supporting decision-making when information is incomplete



Statistics is key to decision-making because most decisions must be made with incomplete knowledge and therefore involve some degree of uncertainty.



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:

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

The question is a decision: Is the evidence already strong enough to stop the trial early and adopt the new drug, or should the trial continue to collect more evidence?

Biostatistical decision-making without all information

The question is a *decision:* Is the evidence already strong enough to stop the trial early and adopt the new drug, or should the trial continue to collect more evidence?

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 of the new drug.

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!

“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.

Welcome to BIOL 322

Statistics for Biological Sciences (BioStatiscs)

Pedro Peres-Neto, PhD

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

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.

Statistics is key!

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

- Herbert George Wells

"efficient citizenship" means **being able to participate effectively and responsibly in society;** for example, by understanding evidence, evaluating claims, and making informed decisions about public issues.

In Wells's quote, *efficient* is being used in the older sense of **capable or effective**, not "fast" or "productive."

Statistics is key!

“Mathematics may rule the universe; statistics rules society.”

- An inspiring moment during a BIOL322 lecture in 2018

Statistics is key to describing, understanding, and making inferences from data!

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 (Information or data used to support or challenge a claim, conclusion, or decision).

It is critical to every discipline, biological OR NOT.

Biology is 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.

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 in an election?

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

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.

ARTICLE

doi:10.1038/nature14967

Mapping tree density at a global scale

T. W. Crowther[1], H. B. Glick[1], K. R. Covey[1], C. Bettigole[1], D. S. Maynard[1], S. M. Thomas[2], J. R. Smith[1], G. Hintler[1], M. C. Duguid[1], G. Amatulli[3], M.-N. Tuanmu[3], W. Jetz[1,3,4], C. Salas[5], C. Stam[6], D. Piotto[7], R. Tavani[8], S. Green[9,10], G. Bruce[9], S. J. Williams[11], S. K. Wiser[12], M. O. Huber[13], G. M. Hengeveld[14], G.-J. Nabuurs[14], E. Tikhonova[15], P. Borchardt[16], C.-F. Li[17], L. W. Powrie[18], M. Fischer[19,20], A. Hemp[21], J. Homeier[22], P. Cho[23], A. C. Vibrans[24], P. M. Umunay[1], S. L. Piao[25], C. W. Rowe[1], M. S. Ashton[1], P. R. Crane[1] & M. A. Bradford[1]

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

"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 when information is incomplete.

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?

Let’s take a break – 1 minute

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)

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).

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?

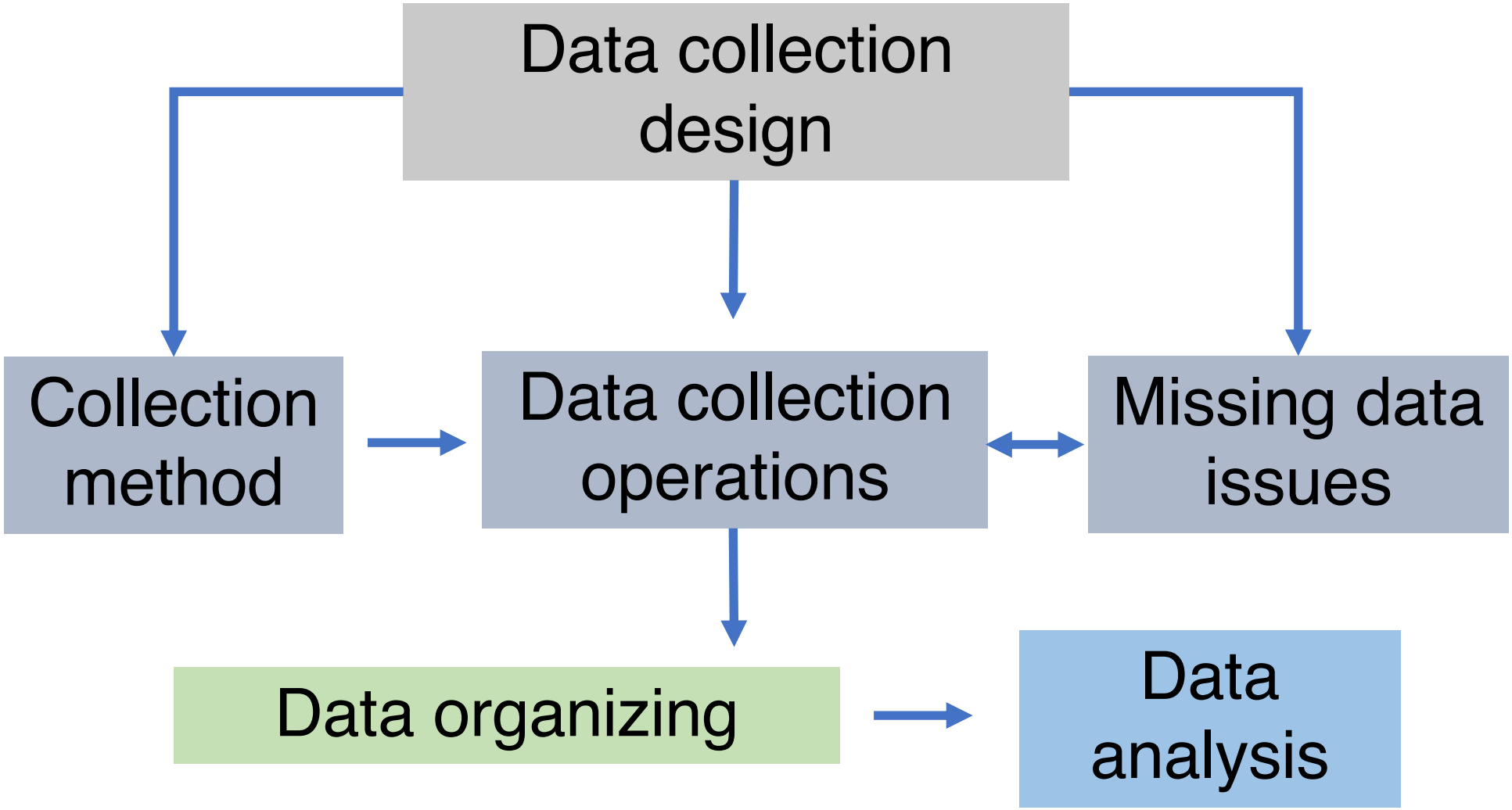
Roles of statistics

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/

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



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

NCSS Data - [Example Data: Pain]

File Edit View Data Analysis Graphics Tools Window Help

Column Info | Kerlosin

	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								

From raw (primary) data……to summaries

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?

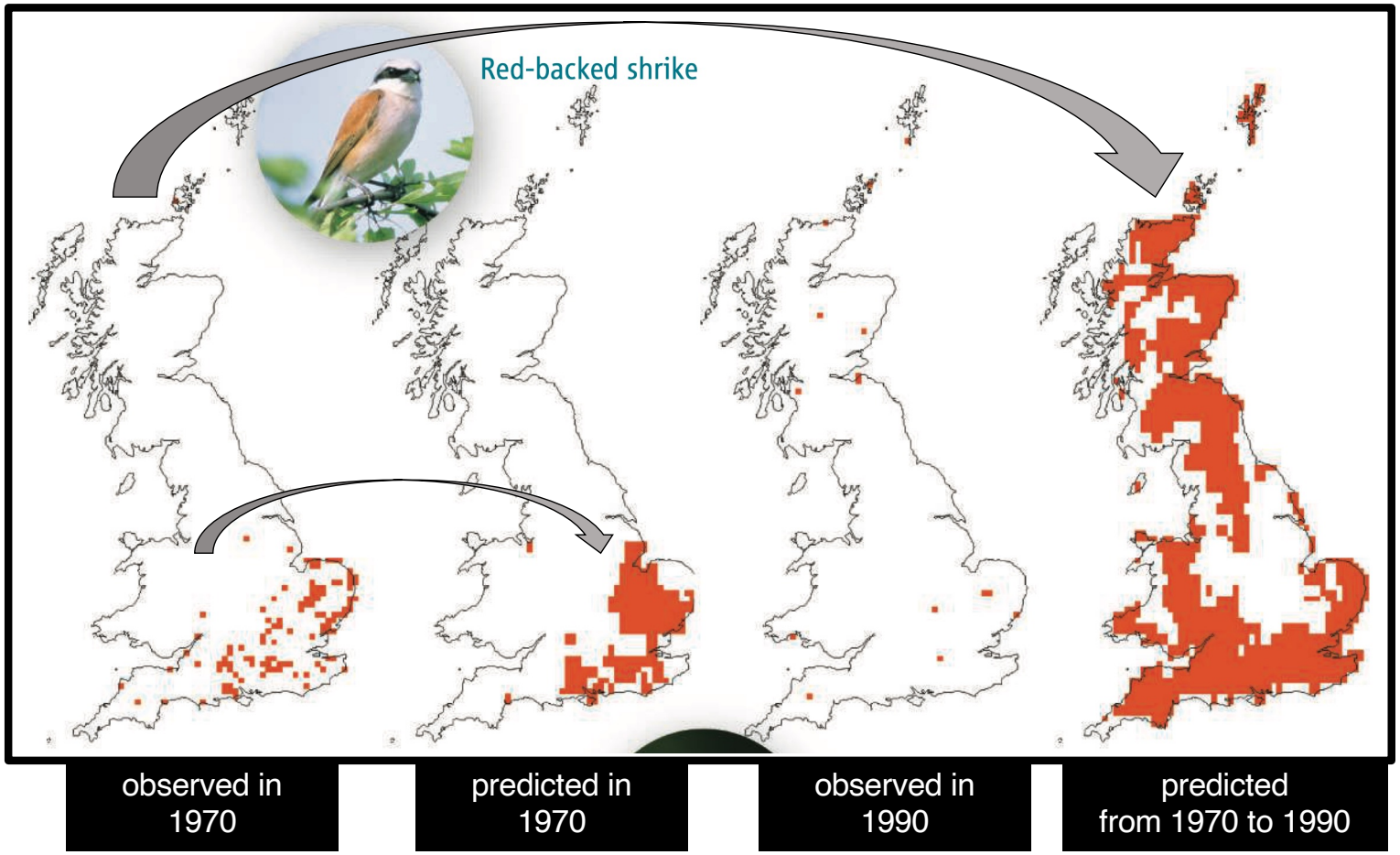
- Source - http://www.scc.ms.unimelb.edu.au/whatisstatistics/ssize.html

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%")

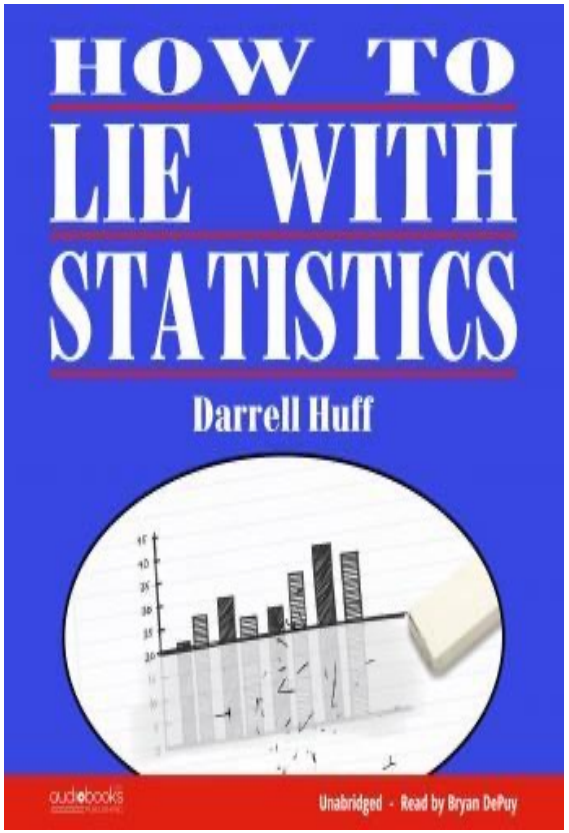
Roles of statistics: [5] - Estimate the present or predict the future



Araujo and Rahbeck (2006, Science)

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

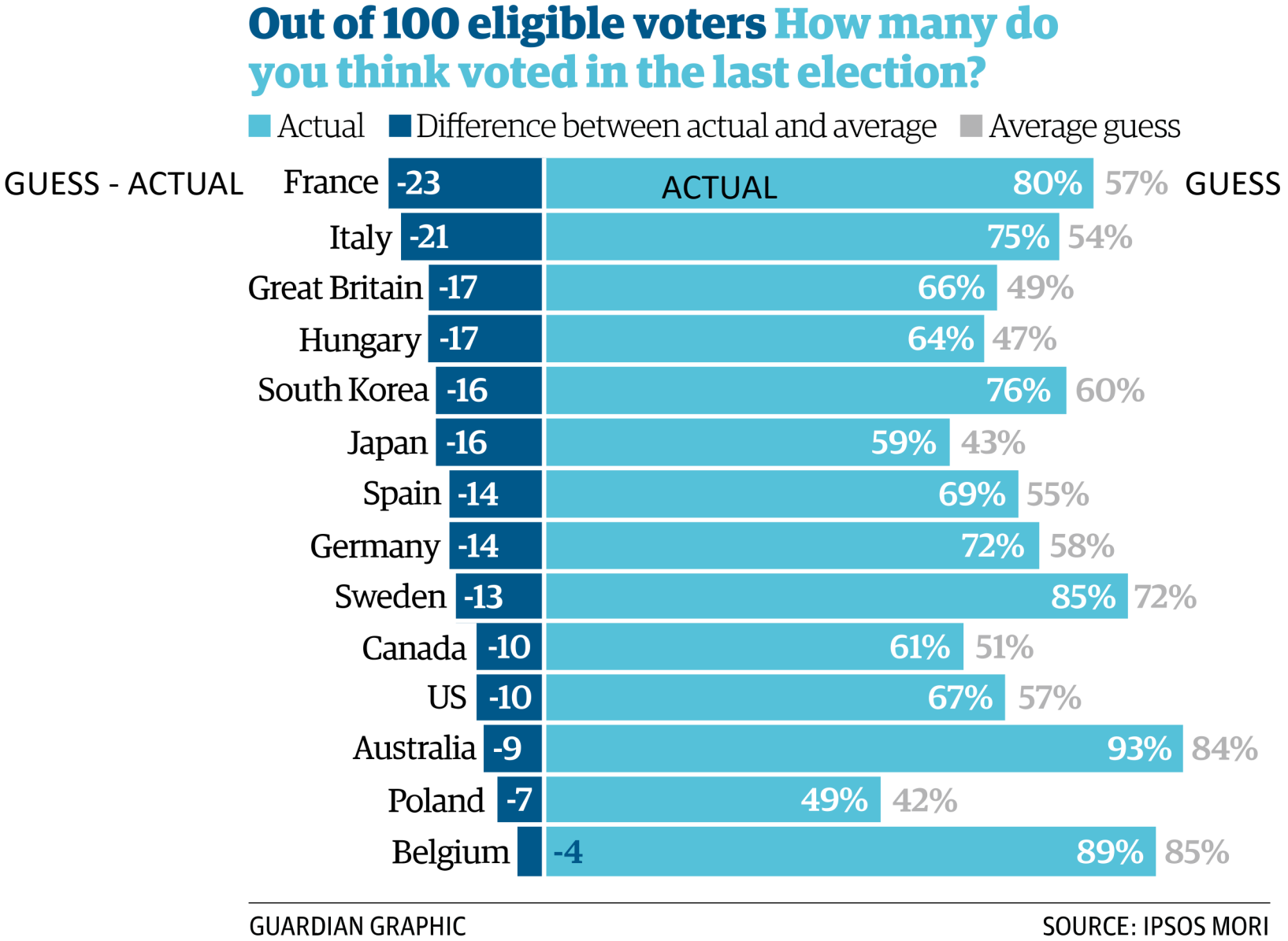
Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.
—H. G. Wells

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

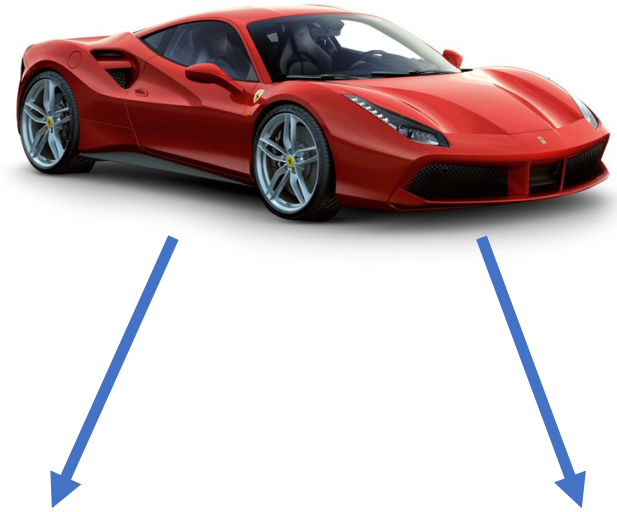
We often make decisions based on wrong guesses!



Published in Oct. 2014

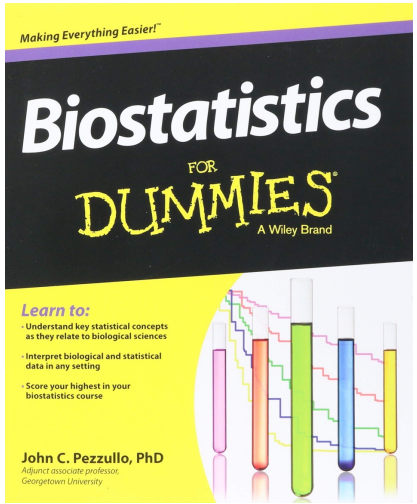
The challenges in understanding statistics (incomplete knowledge - many problems are not intuitive at first

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?***

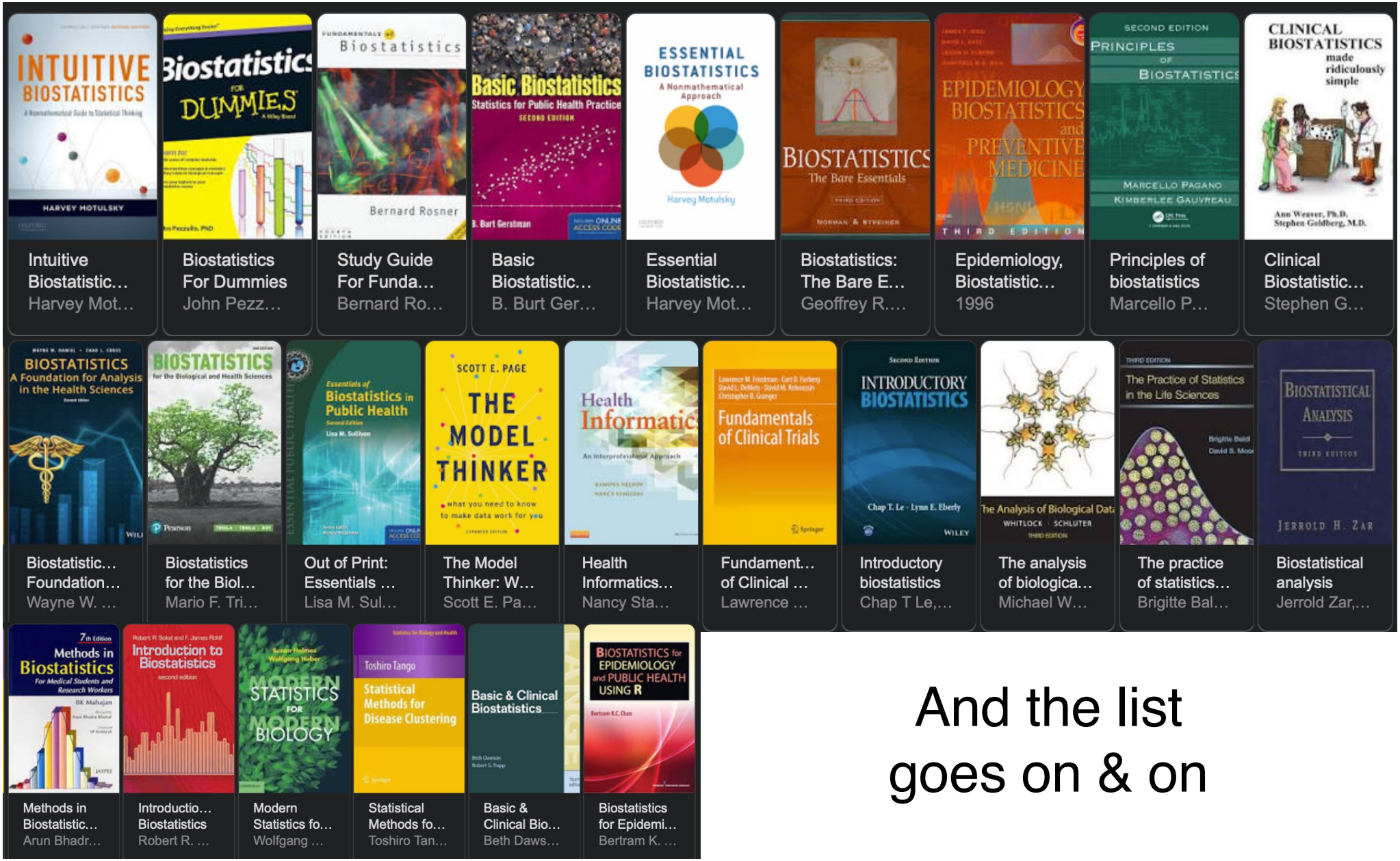


Biostatistics is the branch of statistics that focuses on analyzing data arising from biological systems.

Biostatistics is an extremely active field of statistics



There is a lot of BioStatistics out there

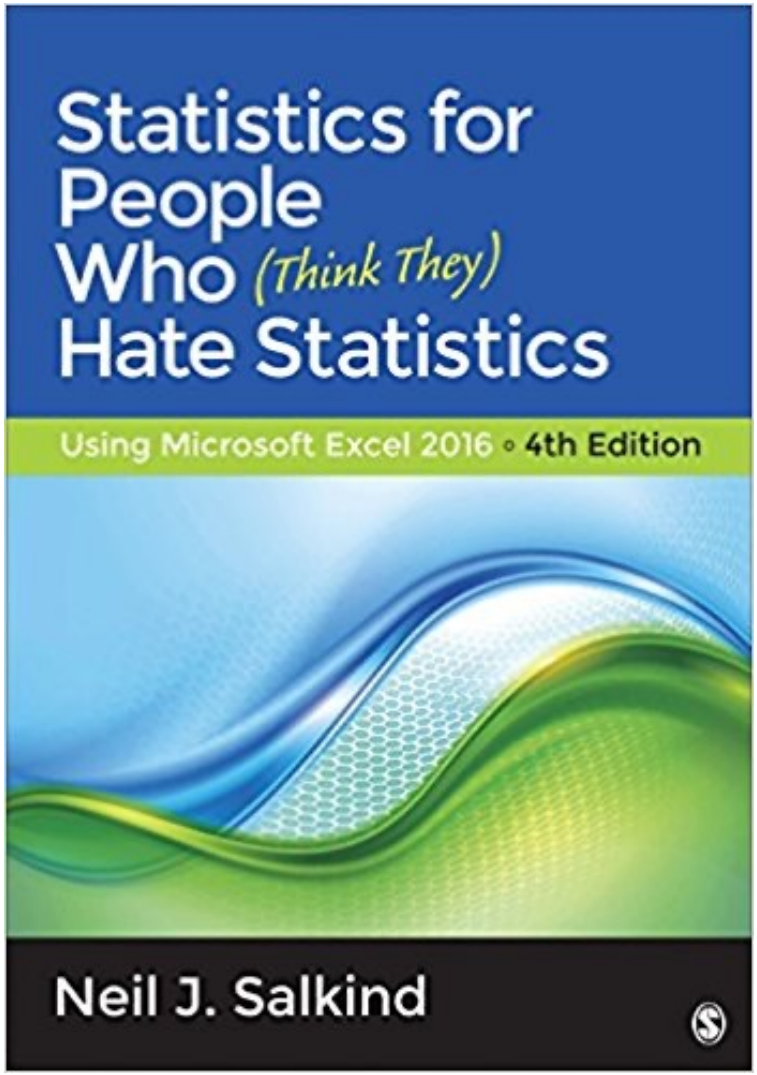


And the list goes on & on

Let’s take a small break – 1 minute

The fear of statistics (bio or not)

The fear of statistics



Bottom line:

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

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.

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

The challenges in understanding statistics

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

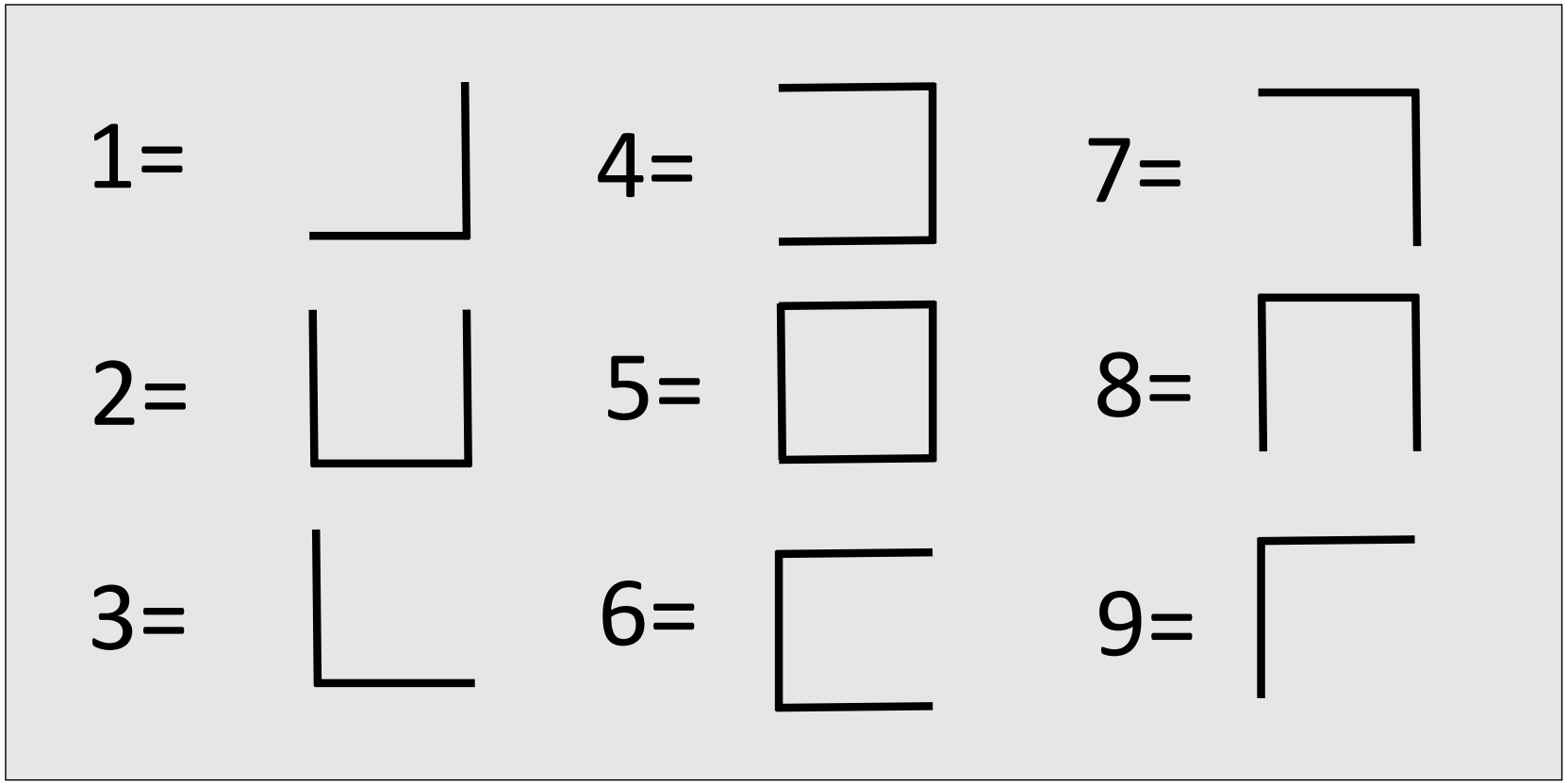


Image by P. Newbury

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?

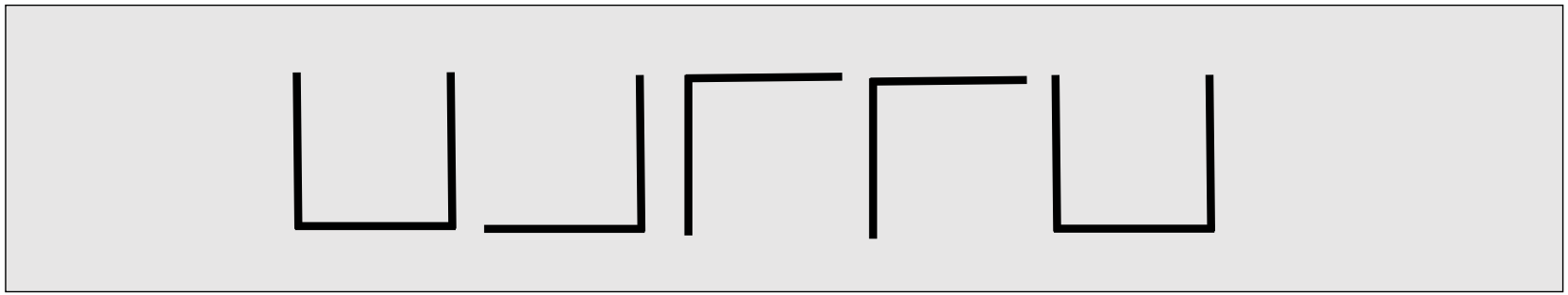
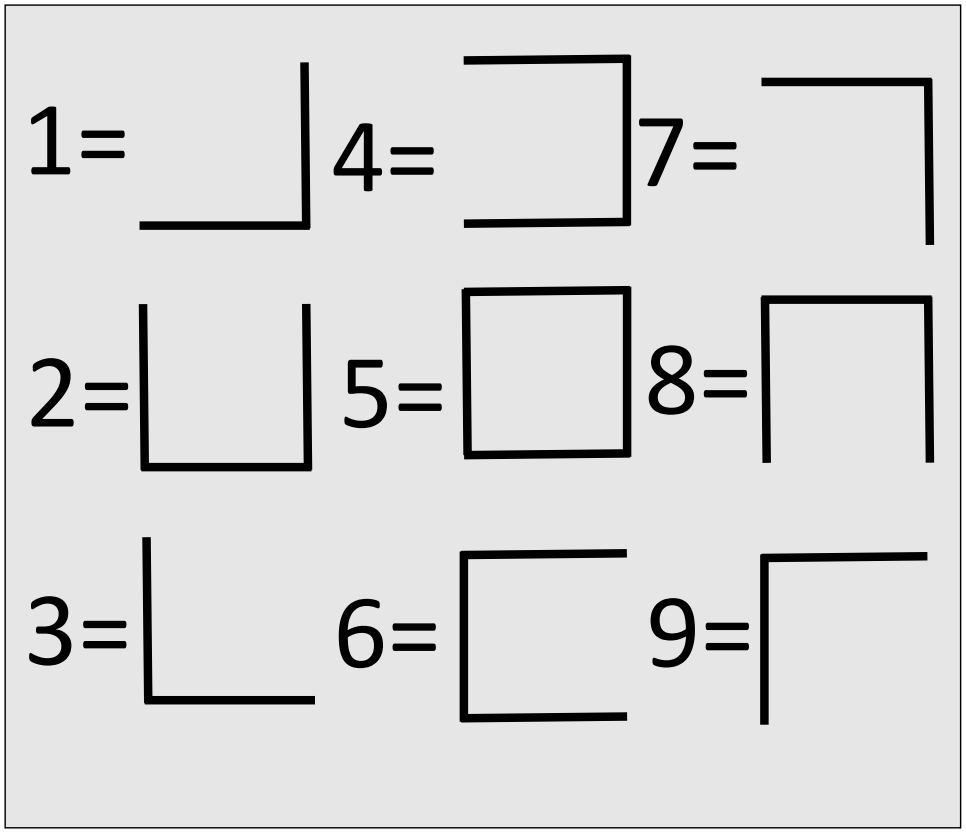


Image by P. Newbury

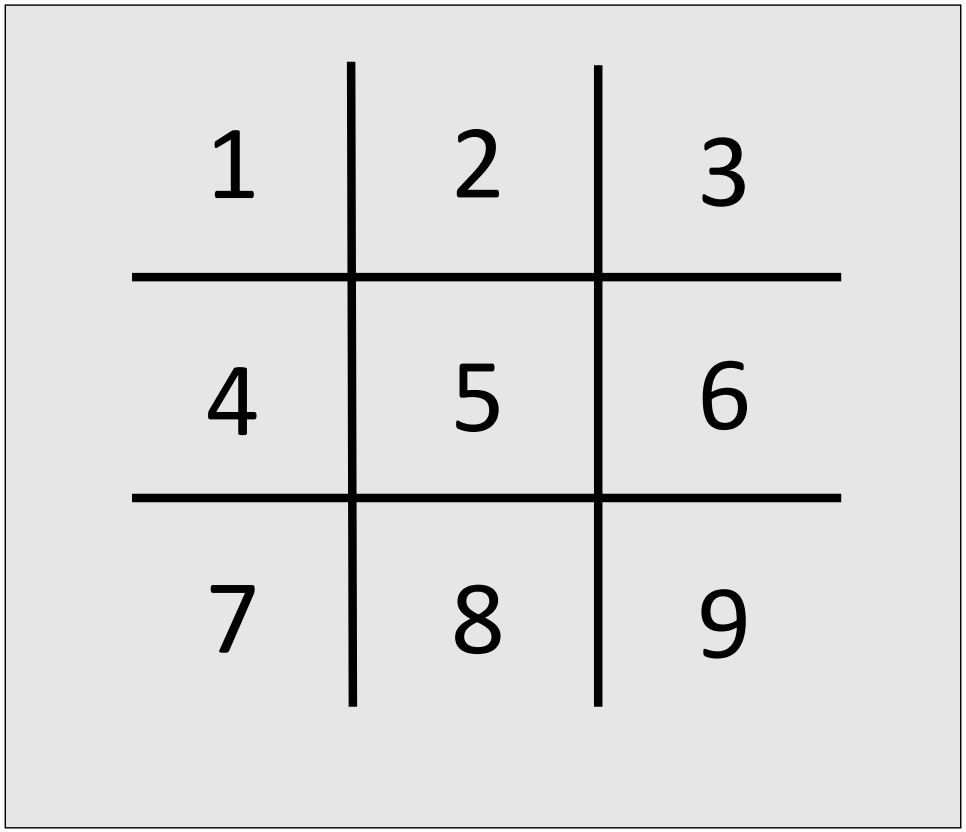
Teaching styles

Transmissionist



Unsupported content

Constructivist

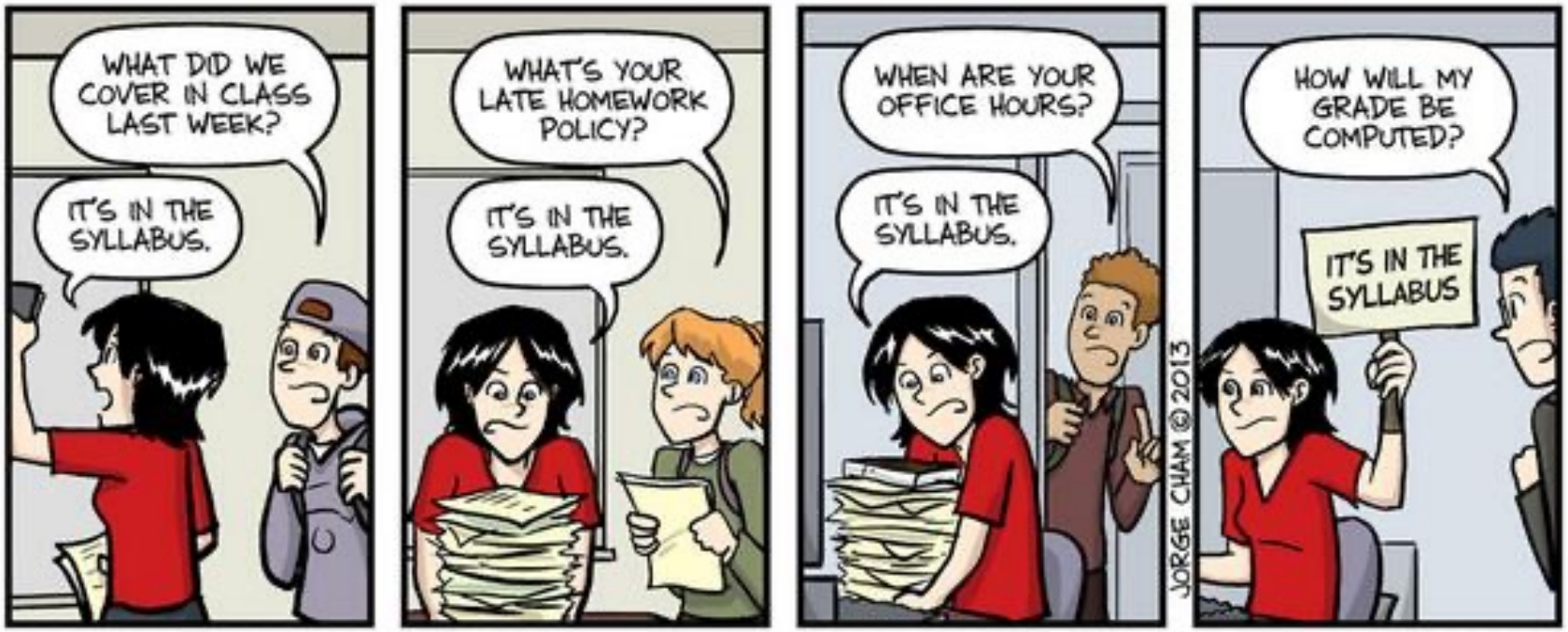


Built on familiar content

Let’s work together!

Communication, communication & communication

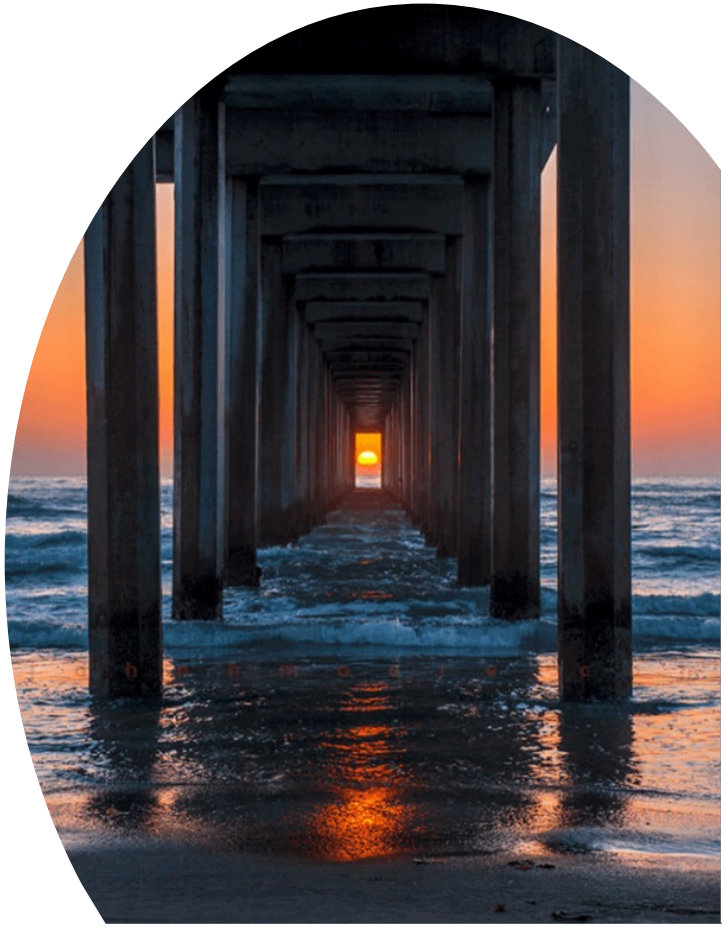
PLEASE READ THE SYLLABUS



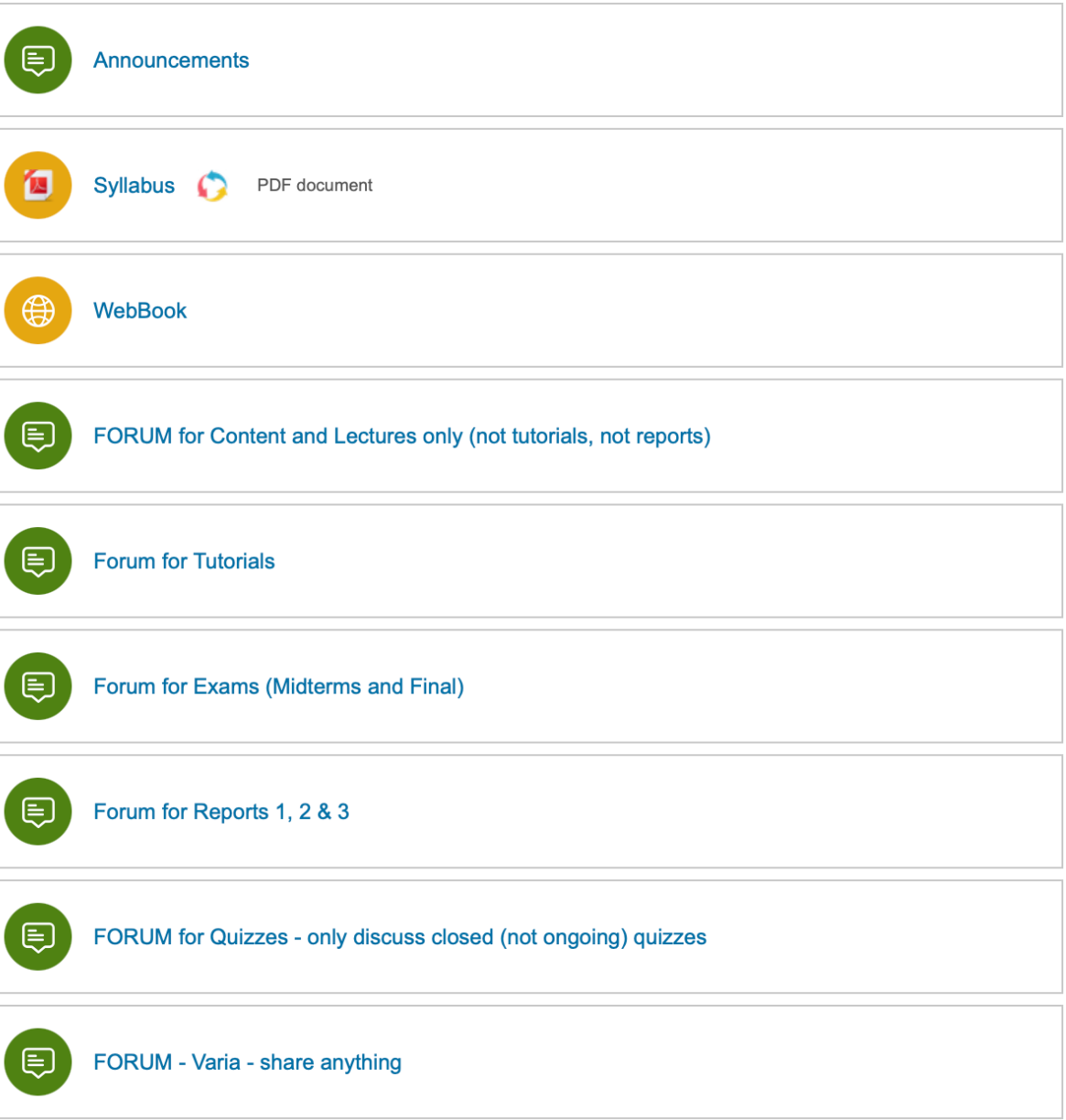
But if it's not in there…let me know

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

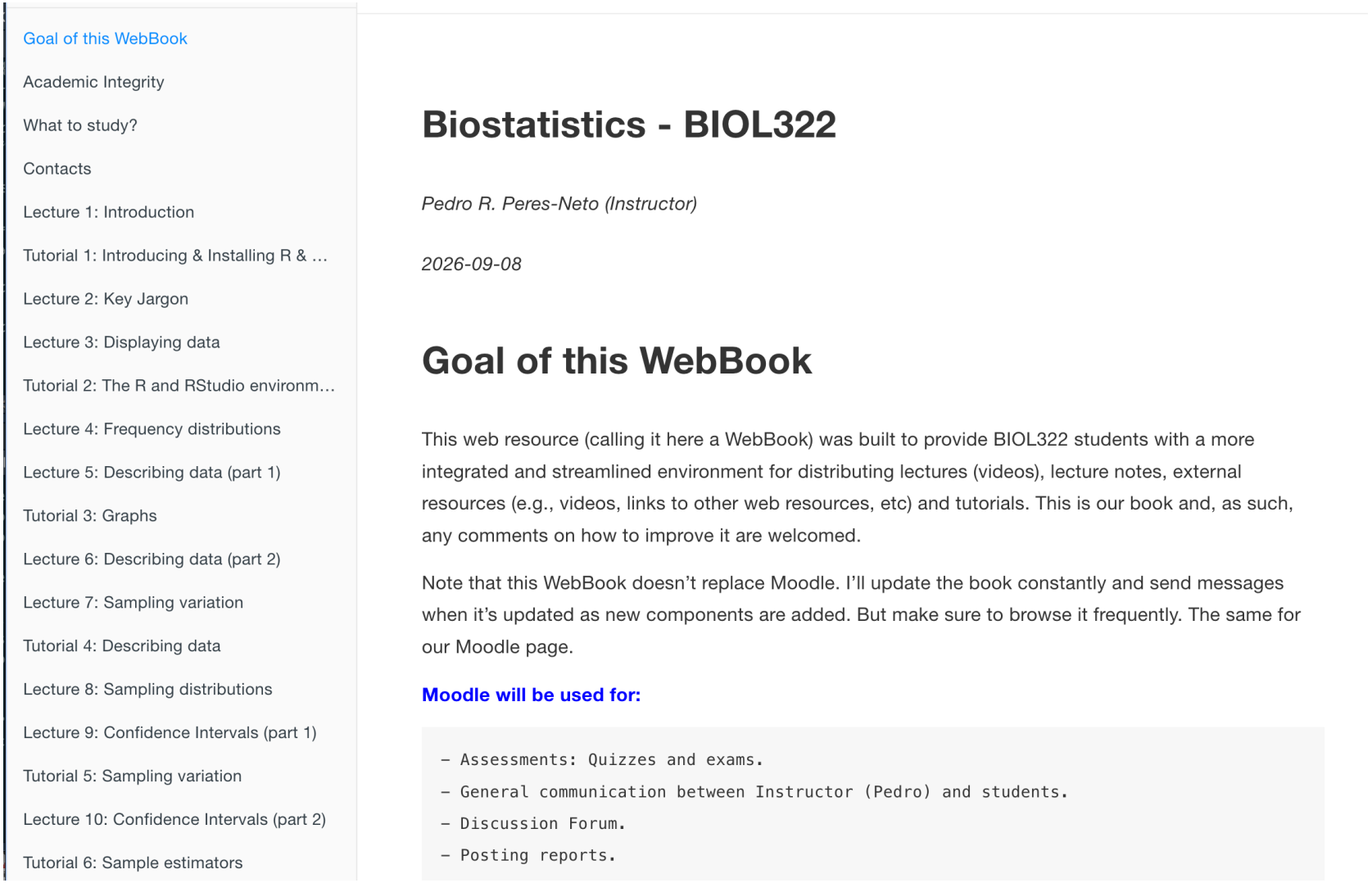
Please use our
Moodle FORUM:
the answer to your
question can help
everyone!



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



WebBook - This web resource was designed to provide BIOL322 students with an integrated and streamlined platform for accessing lecture videos, notes, tutorials, and additional resources, including external videos and links. *Your feedback and suggestions for improving this resource are welcome.*



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

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.

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 when seeking an Honours thesis or BIOL 490 supervisor, obtaining strong letters of recommendation, and pursuing other academic and research opportunities.

Some closing thoughts for today

“Understand. Don't memorize. Learn principles, not formulas.”

“Learning is underrated. Grades are overrated”

Richard Feynman

Next lecture - Sampling

We biologists are relatively small compared with the world we want to understand.

From Chris Lortie