Pre-Lab Worksheet: Science Done Right.

Learning Goals:
Basics of scientific experimental design, statistical evaluation, peer-review.

What are some properties of science done right?

Think about sources you consider scientific or academic, what qualities do they have? Do they have any safeguards to ensure the quality of their research?

Think about sources you consider scientific or academic, what qualities do they have? Do they have any safeguards to ensure the quality of their research?

Have you ever encountered published papers or seemingly academic sources that seem to have suspicious findings? If so, what made you suspect?

Important Previous Concepts:

Hypothesis: A testable statement/prediction about the relationship between variables.

Experiments: Controlled investigations designed to test hypotheses.

Randomized Control Trials (6.1)
An attempt to identify causal relations by randomly assigning subjects into two groups (control and treatment) and then performing an experimental intervention on the subjects in one of the groups (treatment).

Control Group/Condition
A subset of the study sample, often half, treated the same as the rest except that the experimental intervention is withheld. This yields a baseline against which the part of the sample subjected to the experimental intervention (the "experimental condition") can be compared.

Trial/Experimental Intervention
The act of the experimenter changing a variable under study (the "independent variable") on a subset of the sample, to see if it influences a second variable (the "dependent variable").

Randomized Assignment (6.1)
Randomized assignment is one way to remove the systematic uncertainty by making sure that the intervention and control groups are not correlated with some other variable related to the method of assignment itself, e.g. a male vs. female group in a drug trial.

Null & Alternative Hypothesis (3.1)
Null Hypothesis (H_0): The default assumption that there is no effect or no difference between groups or variables. Alternative hypothesis (H_a): A statement that contradicts the null hypothesis, suggesting there is an effect or a difference, this is typically our experiment's "hypothesis" that there is some difference between our treatment and control group.

Statistical Significance (3.1)
Which typically measure statistical significance with a P-Value: which is how unlikely a given set of results would be if the null hypothesis were true (i.e. if the hypothesized effect did not actually exist). A common misconception is that P-values are the probability of the hypothesis being false. This is not quite the same thing. If the P-value

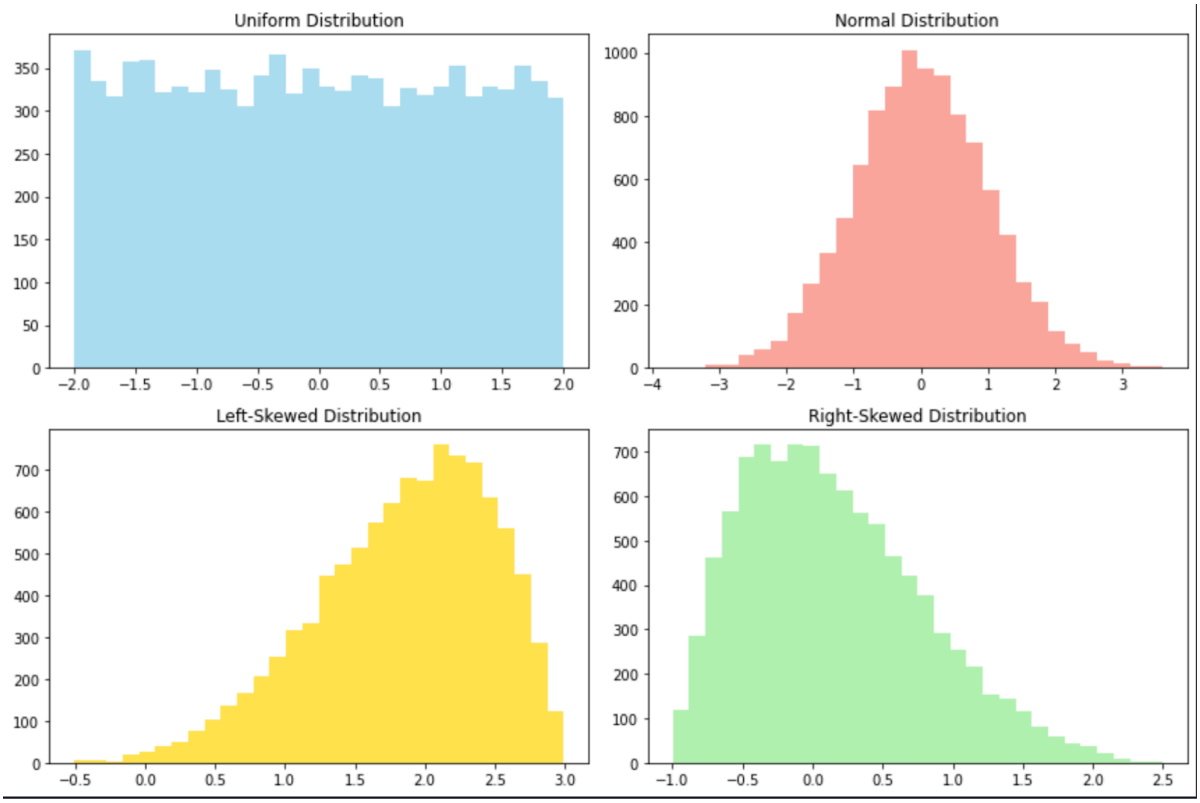
is small, that is evidence against the null hypothesis, typically P<0.05 is statistically significant.

New Concepts:

Hypothesis Testing: T-Test

Hypothesis Testing: a statistical method to decide if there exists statistically significant evidence to support a specific hypothesis given the data. The T-Test is a commonly used hypothesis test that compares the means of two groups to see if they are statistically significantly different from each other.

Data Distributions



We have three main types of commonly encountered distributions:

1. Uniform distribution: all outcomes are roughly equally likely (eg. rolling one die)
2. Normal (Gaussian): data is symmetrically distributed around the mean in a “bell curve”, with many centered around the mean, and less further away from the mean (eg. heights of people)
3. Skewed: Asymmetric distribution with a “tail” on one side, the side where the tail is is our “skew” so a right-tailed distribution would be right-skewed (eg. distribution of income).

Write a testable hypothesis about a topic of your choice.

Given your hypothesis, design an experiment. Consider how you would control your experiment, how would you assign participants.

Describe which type of distribution you would expect for the following scenarios and why:

a. Number of steps taken daily by a group of students.

b. Students grades in a large class with a high average (A-/B+).

c. Probability of hitting a H or T on a fair coin.

Post Lab Worksheet: Science Done Wrong.

Learning Goals:
Distinguishing pathological science, pseudo-science, fraudulent science, poorly-done science, good science that happens to get the wrong answer. What do practicing scientists do when they try to judge a paper in a field or subfield outside their immediate area of expertise?

Publishing & Peer Review (10.2)
New results are evaluated by other experts in the same field to determine whether they are valid. This only reduces confirmation bias insofar as reviewers don't share the same biases.

Detecting Bad Science
Sometimes, it's difficult to tell if studies are examples of bad science or not, just like in your lab. In fact, the lab was based off of a real paper published in a major peer-reviewed academic journal. So, how can we identify bad science from good?

First let's categorize the different types of "bad science"

Spectrum of Poor Research

I. **Pathological Science:** This is what happens when a researcher falls a little too "in love" with their ideas. It is indicated by Langmuir's criteria.

II. **Pseudo-Science:** Pseudo-science is characterized by using scientific vocabulary without aligning with the corresponding concepts or engaging in real scientific practices (i.e., science being "skin deep," not scientific below the surface).

III. **Fraudulent Science:** Research that involves intentional deception, such as deliberately fabricating data or deliberately deceiving the reader about the strength of evidence.

IV. **Poorly-done Science:** Science done honestly, but not done well.

V. **Good Science:** This includes well-done science that may occasionally yield incorrect results. Even with a 95% confidence interval, there's a 5% chance of obtaining results that do not reflect the true effect due to random variation. This is a normal and expected part of scientific research.

Publish or Perish: Replication Crisis

Read this article on the ongoing "replication" crisis in some fields of science. Why do you think replication is so important? What are some core issues that drive this crisis?

Why does bad science exist? Recall from the lab, Edward et. al.'s research on perverse incentives in academia and the paper on reproducibility.

TABLE 1. GROWING PERVERSE INCENTIVES IN ACADEMIA

Incentive	*Intended effect*	*Actual effect*
"Researchers rewarded for increased number of publications."	"Improve research productivity," provide a means of evaluating performance.	"Avalanche of" substandard, "incremental papers"; poor methods and increase in false discovery rates leading to a "natural selection of bad science" (Smaldino and Mcelreath, 2016); reduced quality of peer review
"Researchers rewarded for increased number of citations."	Reward quality work that influences others.	Extended reference lists to inflate citations; reviewers request citation of their work through peer review
"Researchers rewarded for increased grant funding."	"Ensure that research programs are funded, promote growth, generate overhead."	Increased time writing proposals and less time gathering and thinking about data. Overselling positive results and downplay of negative results.
Increase PhD student productivity	Higher school ranking and more prestige of program.	Lower standards and create oversupply of PhDs. Postdocs often required for entry-level academic positions, and PhDs hired for work MS students used to do.
Reduced teaching load for research-active faculty	Necessary to pursue additional competitive grants.	Increased demand for untenured, adjunct faculty to teach classes.
"Teachers rewarded for increased student evaluation scores."	"Improved accountability; ensure customer satisfaction."	Reduced course work, grade inflation.
"Teachers rewarded for increased student test scores."	"Improve teacher effectiveness."	"Teaching to the tests; emphasis on short-term learning."
"Departments rewarded for increasing U.S. News ranking."	"Stronger departments."	Extensive efforts to reverse engineer, game, and cheat rankings.
"Departments rewarded for increasing numbers of BS, MS, and PhD degrees granted."	"Promote efficiency; stop students from being trapped in degree programs; impress the state legislature."	"Class sizes increase; entrance requirements" decrease; reduce graduation requirements.
"Departments rewarded for increasing student credit/contact hours (SCH)."	"The university's teaching mission is fulfilled."	"SCH-maximization games are played": duplication of classes, competition for service courses.

Which of these incentives do you think have the greatest impact on the creation and publishing of bad science and why?

How are some ways we can identify bad science?
Pathological Science Indicators (Langmuir's Criteria):

A. The effect is produced by a barely detectable cause, and the magnitude of the effect is substantially independent of the intensity of the cause

B. The effect is barely detectable or has very low statistical significance

C. Claims of great accuracy

D. Involving fantastic theories contrary to experience

E. Criticisms are met with ad hoc excuses

F. Conclusion-motivated design & analysis

Read about the polywater studies. Was this study an example of pathological science? If so, which of Langmuir's criteria apply? If not, explain why not.

Now that you have some idea of what Langmuir's Criteria is, how many and which of these indicators apply to the study we explored with the odometer readings? Is this pathological science, fraudulent or poorly done science – or is it a mix of many?

We've learned in the lab that peer review is no panacea. What do you think some methods we can identify good vs suspect science – and what are some methods academia can implement to help reduce the amount of bad science published?