Chapter 0: Welcome to Stat 218 - Applied Statistics for Life Science

> **Note**
>
> - **Why statistics?** To separate real patterns from coincidence and bias.
> - **What we use:** variables, observations, study designs, and models (often starting with a "just-by-chance" model).
> - **How we'll practice today:** two short investigations (one about an arson case, one about "Bouba/Kiki") to build intuition for variability, simulation, and evidence.
>
> In the next chapter, we'll pause to put names to these ideas (e.g., population, sample, variable, data type) so we can describe our investigations more precisely.

Example 0.1: Question the Arson

I am a detective in San Luis Obispo, California, and everyone in this room is a suspect in the three fires that have been set in the last 6 weeks.

SLO police arrest resident suspected of lighting 3 fires at Madonna Plaza

I have called all of you in for questioning. Given are 16 questions I want you to answer about these fires. Please answer all 16 questions.

(1)	One of the fires was started using gasoline.	T	F
(2)	Matches were used to set the fires.	T	F
(3)	The trash was on the curb in front of the first house that caught on fire.	T	F
(4)	There was a doghouse in the backyard at the scene of the second fire.	T	F
(5)	The residence of the third fire had off-street parking.	T	F
(6)	The second residence had a screened porch.	T	F
(7)	The suspected entered the first residence through the kitchen window.	T	F
(8)	The suspect left a flashlight behind at the scene of the second fire.	T	F
(9)	All three fires took place on Tuesdays.	T	F
(10)	The porch light was on at the residence where the first fire took place.	T	F
(11)	The third fire was started in the garage.	T	F
(12)	There was a dog barking at the house next door to the third fire.	T	F
(13)	The suspect wore work boots at the scene of the first fire.	T	F
(14)	The home security system was triggered when the suspect fled the scene of the third fire.	T	F
(15)	The lock on the back door at the first residence was broken.	T	F
(16)	Multiple fires were set at the second residence.	T	F

Before we look at results, let's discuss some predictions first! With your neighbors, discuss the following questions.

1. If you are not the criminal, how many questions would you expect to get correct?

Just guessing with two outcomes (Correct / Incorrect) and equal probability implies we would expect 50% correct. Thus, expect 8 out of 16 correct.

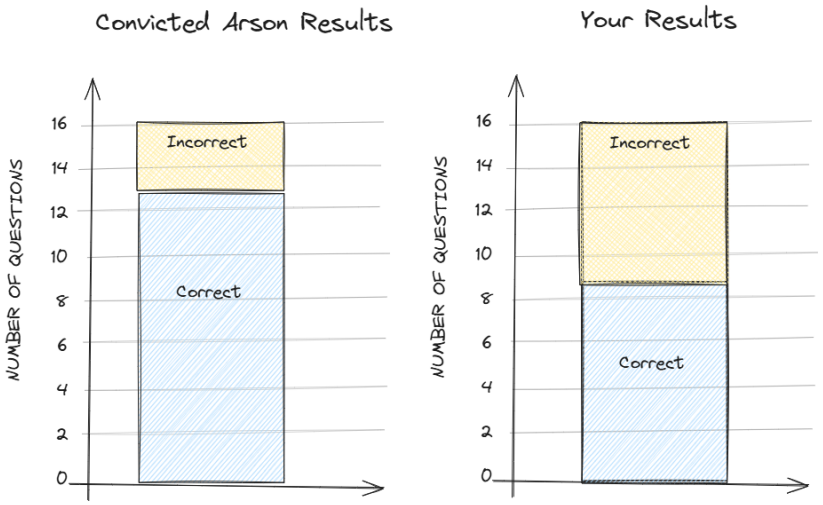
2. If you are the criminal, how many questions would you expect to get correct?

Assuming they are honest and a perfect memory, they might know something and we would expect 16/16 correct.

Let's see how everyone did!

The suspect convicted of the crime was found to have answered 13 of the 16 questions correct. The plot on the left displays the results from the convicted arson suspect. Sketch your results in the plot on the right.

This is a stacked bar plot where the height of the bar indicates how many questions were correct for the convicted suspect and for you. The colors indicate the response category – Correct or Incorrect.



3. How do your results compare to the convicted suspect's results? How about the other suspects (i.e., your neighbor's) results?

My neighbors and I were much closer to 50% correct (e.g., 6, 7, 8, 9 correct) than the convicted suspect.

4. Are you convinced the arson suspect was rightfully convicted?

Ehh, maybe. Nobody "guessed" **13 or more** correct.

5. How would your answer change if the convicted arson suspect answered…

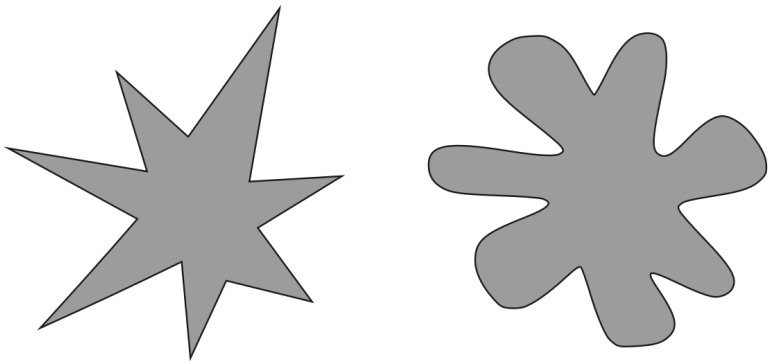
- 15 correct? Probably convinced. 15 correct is unusual!

- 10 correct? They could be just guessing. A handful of students "guessed" **10 or more** correct.

- 0 correct? This is suspicious!

Example 0.2: Can our class speak Martian?

When shown the two Martian letters, Kiki and Bouba, write down whether you think Bouba is on the left or on the right.



1. Were you correct or incorrect in identifying Bouba?

Bouba is on the right!

2. How many students are in class today?

n = 36 students in class.

3. How many students in the entire class were correct in identifying Bouba?(Sketch the *observed* class results in Q10)

There were 30 students correct in identify Bouba on the right.

4. If we really don't know Martian and are just guessing which is Bouba, how many students would you *expect* to choose Bouba correctly? Explain your reasoning.

We would expect 50% (two outcomes – correct/incorrect with equal probability) of 36, thus expect 18 students correct *if just guessing*

5. How could we use a coin to simulate each student "just guessing" which Martian letter is Bouba?

Flip the coin once (Heads = correctly identify Bouba; Tails = incorrectly identify Bouba)

6. How could we use coins to simulate the entire class "just guessing" which Martian letter is Bouba?

Everyone flip their coin and count up (i.e., aggregate) the results.

7. Each student will flip a coin one time to simulate your "guess" under the assumption that we can't read Martian. (Let Heads = correct, Tails = incorrect).

- What was the result of your one simulation?

If heads – correct!

- How many students "guessed" correctly in the first class simulation? (Sketch your results in Q10)

21/36 correct

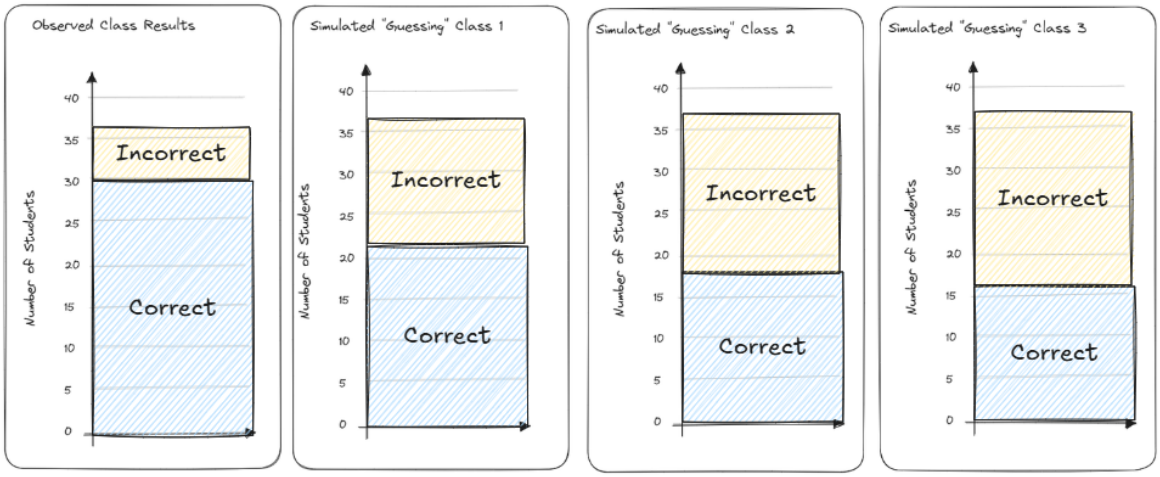
8. Each student should flip their coin again. How many students "guessed" correctly in the second class simulation? (Sketch your results in Q10)

18/36 correct

9. Let's flip our coins one more time. How many students "guessed" correctly in the third class simulation? (Sketch your results in Q10)

17/36 correct

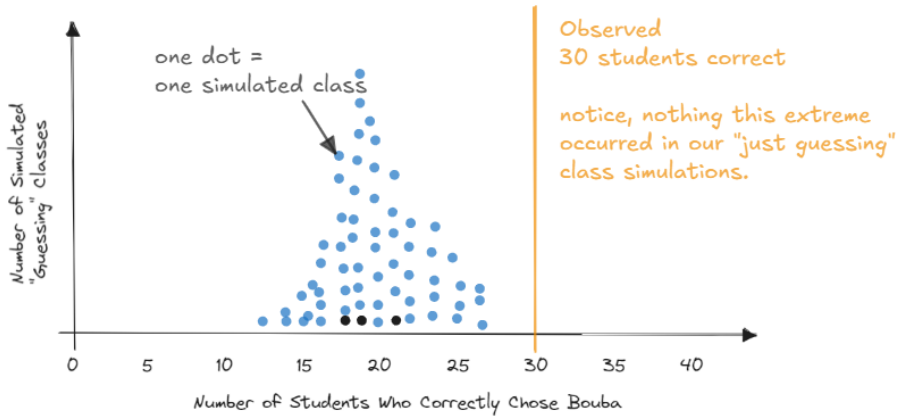
10. Sketch the observed class results and the three simulated "guessing" class results.



11. How do the **observed** class results compare to the simulated guessing classes results?

The observed class results have a lot more students who were correct in identifying Bouba than the simulated "just guessing" class results.

12. Let's display our results in a different type of graph. Create a *dot-plot* to display the results of the three simulations compared to the observed class results.



We still only have a few simulations to compare our class data to. It would be much better to be able to see how our class compared to hundreds or thousands of "just-guessing" classes. Since we don't want to flip coins all class period, we will use a computer simulation (`Canvas > Online Simulation Applets > One proportion inference`) to get 100 "random guessing" statistics.

13. Carry out the applet simulation. Note that you should consider the following questions when designing your simulation study:

- What are the two possible outcomes on each of the trials? Correct / Incorrect
- What is the probability that a Bouba is accurately identified under the assumption that we are "just-guessing"? Change your `Probability of heads` accordingly.

  Probability of heads: 0.5 -> 50%

- How many students were there in this study? Keep this value in mind when setting the `Number of tosses` value.

  Number of tosses: 36 correct

Carry out the simulation study 100 times overall, keeping track of the probability of employees chosen for management that were female on each of the simulated experiments. Sketch the 100 simulated "guessing" class results on the dot-plot in Q12.

Number of repetitions: 100 correct

14. Is your class particularly good or bad at Martian? Use the plot above to explain your answer.

We appear to be better than "just guessing" since our class observed results (30/36 correct) are noticably higher than the simulated class results. They are "unusual" compared to what we would *expect*.

15. Is it *possible* that we could see our class results just by chance if everyone was just guessing? Explain your reasoning.

I suppose it is *possible* to get 30/36 correct if "just guessing".

16. Is it *likely* that we could see our class results just by chance if everyone was just guessing? Explain your reasoning.

No, it is not "likely" to see 30/36 correct because we never saw results as extreme as those observed in our class simulations *under the assumption we were "just guessing"*.

17. Does this activity provide evidence that students were not just guessing at random? If so, what do you think is going on here? Can we as class read Martian?

Since our observed results are "unusual" compared to what we expected, we have evidence to conclude that students were not just guessing at random.

TED Talk: 3 Clues to Understanding Your Brain by Vilayanur Ramachandran (2007)

The synesthesia part begins at roughly 17:30 minutes: https://www.ted.com/talks/vs_ramachandran_3_clues_to_understanding_your_brain