

M212B Performance Task

Predictive Modeling - UFOs

The Modeling Cycle

Since the beginning of recorded human history, people have reported seeing flying objects unlike any familiar animals or aircraft. Popularly known as Unidentified Flying Objects, UFOs could be secret military or civilian technology, a trick of the light, an outright fabrication...or perhaps evidence that alien life has arrived at Earth. In this task, you will take on the role of a data scientist to investigate a dataset with multiple variables describing UFO sightings over the past 100 years. As a data scientist, you will conduct an exploratory analysis of the data, look for emergent patterns in the data, and interpret what story about UFOs the data may be hiding.

Mathematical models are used to make sense of ideas and phenomena in our world. As illustrated in the image below, the modeling cycle is a process of collecting data about an idea or problem, creating models of that data, analyzing the story told by those models, and reporting the results to the appropriate audience. In this task, you will be engaging with this mathematical modeling cycle to identify patterns in how, where, and when people report seeing UFOs.

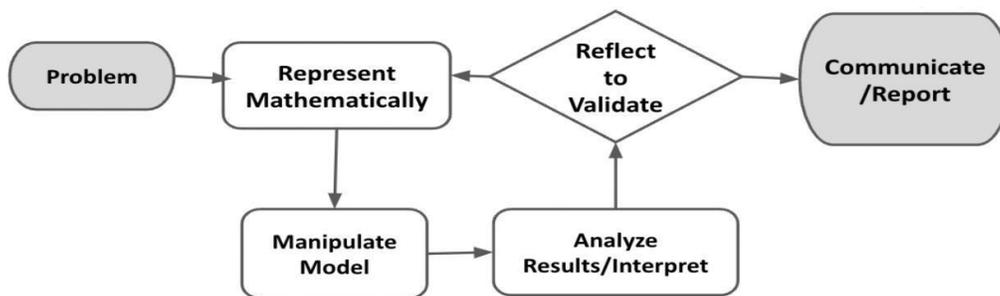


Figure 1. *Diagram of Modeling Process.* From “Progressions for the Common Core State Standards for Mathematics” [Draft manuscript], by Common Core Standards Writing Team, 2019, p. 6 (http://mathematicalmusings.org/wp-content/uploads/2022/05/modeling_02082019.pdf).

Introduction: What makes a UFO sighting, anyways?

By definition, an Unidentified Flying Object (UFO) must be in the air, and we must not have enough information to definitely identify it as a known aircraft, animal, or natural phenomena. This includes all sorts of mundane occurrences: if someone throws a tomato past your head and you don’t get a clear look at it, that tomato would be a UFO. In popular usage, a UFO is generally an object seen in the sky that looks or moves like a piloted vessel. Most sightings are of far away objects, but people have also reported seeing alien spacecraft up close, perhaps even being taken on board!

Read some of the text descriptions of sightings in the UFO dataset, then answer the questions below.

QUESTION 1



Imagine receiving a UFO report with a grainy photo of a faraway circular object. The reporter says they saw the object hovering in the sky for 30 seconds before it disappeared behind a cloud, and swears the photo is conclusive proof that we are not alone in the universe. Brainstorm as many possible explanations for this sighting as you can. Which explanations might you think are more likely than others, and why?

QUESTION 2



What features or evidence from a UFO sighting would make you more inclined to believe that it was an artificial aircraft of some sort? What features or evidence would convince you more that it was an *alien* aircraft?

Initial Data Exploration

QUESTION 3



Plot separate distributions of responses to three variables from the UFO dataset. You might consider histograms, violin plots, or box plots. What do you notice about the distribution of these variables?

Planning your analysis

This UFO dataset contains several variables, which may or may not be related to each other. We also lack ground truth data; since these are self-reported UFO sightings, we don't know for sure what each respondent actually saw or even if their report is accurate.

Of course, that doesn't mean we are helpless. Researchers work with imperfect datasets all the time. We will begin by brainstorming some hypotheses that you think this dataset could help you answer.

QUESTION 4



Brainstorm two testable hypotheses that you could analyze with this dataset. For each hypothesis, include:

- Your hypothesis written in plain English
- How the data might show your hypothesis is false
- What you think could be interesting if your hypothesis is true

For example,

Hypothesis: Certain UFO shapes are more commonly seen than others.

Evidence of falsity: The reported proportions for each UFO shape are similar (i.e., all ~10%).

Potential interesting findings: If there are one or two shapes that are much more common, and those shapes happen to be the same shapes we see on popular media (generally ovals or circles), that could show the media affect our perception and attention in the wider world to the point of causing people to see aliens where they might not exist.

QUESTION 5



Choose **one** of your hypotheses from the previous question and explain how you would test the dataset to see if your hypothesis is false. This is likely to be a statistical test of some kind. However, if you have a very interesting question that doesn't seem to fit neatly into a statistical test on *this* dataset, you may propose a different analysis or visualization along with how you will use the data to support or refute your hypothesis.

Because this dataset is relatively messy, you may not be able to prove your hypothesis true or false. That is fine, as long as your analysis can partly support or refute your hypothesis.

For example,

Hypothesis: Certain UFO shapes are more commonly seen than others.

Test: χ^2 goodness of fit test on UFO shape proportions against the uniform distribution. Our hypothesis is supported if the test shows that the distribution of shape proportions is significantly different from uniform.

Or another example,

Hypothesis: People are mistaking man-made objects for UFOs.

Test: Create a heatmap of sightings based on the geolocated data and compare it to an online map of major world airports. Even better, compare it against an online map of the average number of man-made objects in the sky, if such a map exists. Our hypothesis is supported if our UFO heatmap shows sightings are significantly concentrated around major airports or places where many man-made objects are in the sky.

Initial Data Exploration and Planning Your Analysis Checklist

Requirement	Fulfilled?
I plotted separate distributions of responses to three variables from the UFO dataset described what I noticed about the distribution of those variables.	<input type="checkbox"/>
I proposed two testable hypotheses for the dataset and reported on evidence of falsity and potential interesting findings.	<input type="checkbox"/>
I chose one of my hypotheses and tested the data to either support or refute that hypothesis.	<input type="checkbox"/>
Teacher Consultation:	_____

Open Modeling Challenge

QUESTION 6



Attempt to support or refute your hypothesis using your suggested analysis from the previous section. Make sure to include a table, graph, or some other visualization that would help a reader interpret your results.

Interpretation and Evaluation

QUESTION 7



Explain whether or not your analysis supported or refuted your hypothesis. If you feel the result was ambiguous or inconclusive, explain why and what you could have done differently to get a more conclusive result. Explain your reasoning as if you were giving a presentation to a government official who only knows a little statistics.

QUESTION 8



One of the marks of a principled approach to scientific analysis is to consider counterarguments. For your analyses, present at least one counterargument to your findings and what you would need to do to evaluate the counterargument.

For example, let's say we discovered that the proportion of UFO shapes are *not* significantly different from the uniform distribution. In other words, in this dataset each shape is reported a similar amount of the time. A counterargument to this finding could be that in fact some shapes are more common, but the dataset is polluted by random and false reports because it was drawn from online surveys with minimal verification. To check whether this is true, we could do a similar analysis but only on sightings with a "description" field containing more than 5 words and written in relatively coherent English, because random reporters would likely not take the time to fully describe what they saw.

Reflection

QUESTION 9



Congratulations on completing an open-ended data analysis task!

- Say you were on a mission to figure out once and for all whether UFO sightings are glimpses of top-secret military aircraft or even alien spaceships, rather than people misinterpreting a cloud or normal aircraft for a UFO or even making up reports. Did your analysis help you decide whether people are truly seeing something “secret?”
- If you could access more data than the dataset provides, what other existing data would you like to include in your analysis to help you find patterns or even figure out what people are really seeing? Why do you think this new data could help you?

Open Modeling Challenge Checklist

Requirement	Fulfilled?
I included a visualization to help support or refute my analysis.	<input type="checkbox"/>
I explained my reasoning as if I were giving a presentation to a government official who only knows a little statistics.	<input type="checkbox"/>
I presented a counterargument to my findings as well as what I would need to do to evaluate the counterargument.	<input type="checkbox"/>
I suggested additional data I would like to include in my analysis to help find patterns or figure out what people are really seeing.	<input type="checkbox"/>
Teacher Consultation:	_____