Final Project: Journal of Erroneous Results

AstroStats

2022-04-01

The details of this document may change slightly, to better fit class needs. If it does, I'll make an announcement in class. You can always find the current version on Canvas.

I

Contents

1 Introduction

Scientists communicate via the publication of peer-reviewed papers. We will be replicating this process in class, with a fun twist.

You will all be analyzing the same set of Moon crater data. This is the same data you've already looked at in one of your assignments. This set is "AstroStats_Robbins_Moon.csv" and is available in the Files portion of Canvas. This data is from "A New Global Database of Lunar Impact Craters >1–2 km: 1. Crater Locations and Sizes, Comparisons With Published Databases, and Global Analysis" by Stuart J. Robbins, JGR Planets, 2018. That paper is also on Canvas.

The original dataset has over two million craters. As much as I like giving you original data to work with, that much data is difficult for the Olin-issued computers to process in a timely fashion, and I didn't want you to spend all your time on this project waiting for the data to load. I've removed the smallest craters (anything smaller than 5km) to create a dataset with 83,000 craters.

You'll pick a research question that uses the dataset uses a statistical tool.

Example questions:

- Are there fewer craters on the poles than the equators?

- Are there significantly different distributions of craters on the maria vs the highlands?

- Is there really no dominant direction for elliptical crater orientation based on location, as the paper claims?

- Can you use Monte Carlo modeling to simulate the sizes of the impacting asteroids that created the lunar craters? How does that compare to the current population of near-Earth asteroids?

You will investigate your question and write a manuscript on your results. You'll then be a referee, where you review one of your classmate's manuscripts.

2 Timeline

This is the timeline of the project. Note that the work is not evenly distributed between deliverables– I expect that you'll be working on aspects of the project (data investigation, manuscript writing) well in advance of those deadlines.

Date	In Class	Deliverable (group deliverable, unless indicated)
Monday March 21	Introduction, activities to identify a partner	Partner Choice (individual, due at end of class)
Thursday March 24	Visit by Librarian Maggie Anderson	Workstyle Reflection(individual)
Monday March 28	Check-ins with teaching team - come prepared to discuss your research question.	Data Exploration (individual) and Question (group)
Thursday March 31	Open work time. You are encouraged to have paragraph drafts to share with instructors for feedback.	Research Log and Synthesis Matrix
Monday April 4	Check-ins with teaching team - come prepared to share your notebooks	Introduction paragraphs with references
Thursday April 7	Overview on best practices for figures	Reading and questions for Prof. Lawler (individual)
Monday April 11	Prof. Lawler virtual visit	Questions for Dr. Singer (individual)
Thursday April 14	Dr. Singer virtual visit	One nice figure
Thursday April 21	Check-ins with teaching team	Check-in questions 1 (due night before class)
Monday April 25	Overview of good/bad reviews. Class collaboration on how to complete review process.	Manuscript, Jupyter Notebook, and associated materials.
Thursday April 28	Check-ins with teaching team	Check-in questions 2 (due night before class)
Monday May 2	Finish drafting referee reports	
Tuesday May 3		Referee report
Friday May 6 (Final, 10 am)	Read reviews. Discussion.	Final Reflection (individual)

3   Main Project Deliverables

3.1   Partner Choice

3 points. Please give the name of the person you would like to work
with. I will only assign partners if both people indicate they want
the other for a partner. If you are having problems finding a partner,
you can let me know in your response. (though ideally you'd let me
know during class). You and your partner will get identical grades
for the project assignments, except the Final Reflection.

3.2   Workstyle Reflection

You will be working on this project together for many weeks. It is
important to establish good working parameters. Please complete
the three pages at the end of this document: Personal Workstyle
Reflection,Team Workstyle Reflection, and Goals and Expectations.

   If you have any teaming concerns throughout the semester, please
share them with Carrie. Your concerns will be kept confidential.

3.3   Data Exploration

You've already started this step in one of your assignments. I think
this step is important for everyone to do, so this is an individual
assignment. You may work together, for example by reading the
paper side by side and discussing. I would like everyone to read the
paper themselves.

   It is important to establish "ground truth" when working with
data like this. Anytime you are confused by an entry in the database,
take a look at the actual crater it is referring to. You can do this by
plugging in the latitude and longitude coordinates here. These are
the some of the same images Robbins used to make his database.

1. Take a look at one of the craters in the database using LROC. What
   was the latitude/longitude? What was the diameter? Could you
   clearly ID the crater?

2. Take a look at another one of the craters in the database using
   LROC. What was the latitude/longitude? What was the diameter?
   Could you clearly ID the crater?

3. Take a look at a third crater in the database using LROC. What
   was the latitude/longitude? What was the diameter? Could you
   clearly ID the crater?

   Please read Robbins 2018. While you are reading, please use and
add to a collaborative Glossary. I'm not grading the glossary, this

is just a way for everyone to help each other and make the reading easier to understand.

4. Did you read section 1. Introduction? (For this and the following questions, you can respond with a yes or no).

5. Did you read section 2. Previous Lunar Crater Databases?

6. Did you read section 3. Generating a New Global Lunar Crater Database, and Information Contained Within?

   You may skip section 4.

7. Did you read section 5. Database Properties and Implications?

8. Did you read section 6. Summary and Future Work?

9. Did you read Appendix A?

3.4 Question

Please submit the question you plan on investigating. Be clear which statistical method you will be using. This can be a few sentences. Please do not link to external sites.

3.5 Research Log and Synthesis Matrix

Maggie shared how to make a research log and synthesis matrix. Here's a link to the libguide she created for us. Please fill out a research log and synthesis matrix as you research your question and lunar craters generally.

Please upload the documents. Please do not link to an external site (like google sheets).

3.6 Introduction paragraphs with references

Your introduction section will have two paragraphs. The first talks about why this work is interesting. You will probably start with your science question in the first sentence or two. Then you will explain why that question is interesting.[1] Depending on your topic, you might address why scientists study craters on the Moon. You might also explain what your study will tell us about how the moon formed or evolved.

The second gives background to the work. It gives a quick overview about what scientists have learned about the moon and craters. This will include a summary of the Robbins 2018 paper. You want to explain the relevant parts of this paper to your reader, as it applies to

[1] By interesting, I mean why is it interesting in the conventions of science. You could just think it's totally awesome, but for science papers and proposals you generally show something is interesting by linking it to bigger questions like how did the solar system form and is there life elsewhere in the universe. Take a look at the papers you've found- you should see each of them do this.

your research. For example, if your research focused on the diameter of craters, you'd want to explain how those are measured in this database. This is also the place to summarize and reference the work other people have done related to your specific science question. Has anyone addressed it before? What did they find?

Please follow the AAS Style Guidelines and AAS Journal Reference Instructions.

Here's an example from one of my papers on asteroids.

Sizes and albedos of asteroids are basic quantities that can be used to answer a range of scientific questions. A significant number of diameter measurements produce a size–frequency distribution, which can constrain models of asteroid formation and evolution (Zellner 1979; Gradie & Tedesco 1982; Bus & Binzel 2002; Tedesco et al. 2002; Masiero et al. 2011). Asteroid albedos aid the identification of collisional family members (Carruba et al. 2013; Masiero et al. 2013, 2015; Walsh et al. 2013; Milani et al. 2014), and allow for basic characterization of asteroid composition (Mainzer et al. 2011c; Grav et al. 2012a; Masiero et al. 2014).

Most observations of asteroids are made in visible wavelengths, where flux is dependent on both size and albedo. Observations in other wavelengths, such as the infrared (e.g., Hansen 1976; Cruikshank 1977; Lebofsky et al. 1978; Morrison & Lebofsky 1979; Delbó et al. 2003, 2011; Wolters et al. 2005, 2008; Matter et al. 2011; Müller et al. 2012, 2013) or radio (e.g., Ostro et al. 2002; Benner et al. 2015), are needed to determine these quantities precisely. At present, well-determined diameters and albedos have been measured for less than a quarter of known asteroids.

You'll note this has a lot of references. It covers a lot of ground in a short time. It doesn't fully explain anything, but it tells you where to look if you want more information. This is all what I'll be looking for.

I advise you to run your paragraphs through the Hemingway Editor. The Editor will supply suggested changes, you do not have to follow all of them. It is a useful tool to help you stick to short sentences, and it helps you avoid words like "very" and "extremely" which are discouraged in scientific writing. One way the Hemingway Editor diverges from scientific writing is in the use of passive voice. Passive voice is encouraged in scientific writing, even though it's stylistically unpleasant.

You are encouraged to share your draft of these paragraphs, with references, with someone on the teaching team before you turn it in.

Rubric

1. (3 points) First paragraph

2. (3 points) Second paragraph

3. (3 pts) I would like you to have at least four references, in the AAS reference style. This means references in the paragraph text **and** a

bibliography at the end.

Please address any comments you are given on these paragraphs after they are graded. Then you can put the paragraphs and bibliography straight into your final manuscript.

3.7 *Reading and questions for Prof. Lawler*

(We're taking a brief break from the project to discuss an important ethics and values issue. I'm delighted to have a guest speaker on this emerging issue. Dr. Meredith Rawls had a schedule conflict, and we'll be instead hearing from Prof. Lawler of the University of Regina,)

In the last two years, people have begun to realize that large artificial satellite constellations may block access to the night sky. The night sky is important to astronomers, indigenous groups, astrophotographers, and some non-human animal species.

Please read The impact of satellite constellations on space as an ancestral global commons by Venkatesan et al., Nature Astronomy, 2020. This is a dense, well-referenced article that covers a range of space-ethics issues.

Please also listen to my interview with Dr. Rawls. (This will be published on March 20th, 2022)

Head over to Canvas to log how you engaged in these resources, answer a few quiz questions, and provide questions in advance of Prof. Lawler's visit.

3.8 *Questions for Dr. Singer*

(3 points) I'm delighted that Dr. Singer found time in her busy schedule to visit us. I would like us to use this time wisely! In about 150 words, please describe the question(s) you'd like to ask her.

If you are stuck, consider the following.

- Are there any puzzling craters you've discovered while data cleaning you'd like to ask about?

- Do you have any questions about the process of crater mapping?

- Could you use an additional reference on your specific investigation?

- Could you use her opinion on your science implications of your question?

If you do not have any crater questions, you are also welcome to submit a question about what it's like to propose a NASA planetary science mission.

3.9 One Nice Figure

In class, we will go over the elements of a good figure. Please submit your figure in this assignment.

Your figure will be evaluated using this rubric:

1. Is your figure visually engaging? Does it make me want to delve deeper? Are colors/ transparency/ point type/ point size used in a meaningful way?

2. Is it comprehensible to a colorblind person?

3. Does it have labeled axis, units, legend, as appropriate?

4. Is there a caption or title that explains the major message of the figure?

5. Does it tell the story of the data? Is the data presented in an objective way? Can the reader extract quantitative values from the figure?

Remember you can always ask for feedback on your figure before it is submitted. Please incorporate any changes requested after your figure is graded, and then you can directly add it to your manuscript.

3.10 Check in questions

Please answer the following questions so that the teaching team can be prepared to help you in class.

1. 1-3 sentence description of your project

2. Current detailed status of your project (what you have accomplished so far)

3. How do you view your progress so far on a scale of 1 to 7? 1 means you have made very little progress; 4 means you have made good progress; 7 means you have made amazing progress.

4. What would you like help with? What would you like feedback on? (Try to provide some detail here if possible so we can have a little time to contemplate your questions before we talk.)

5. You current work for our reference (manuscript and notebook). You can upload this to Canvas or submit links to shared document or github (just please make sure we have access). You don't have to include everything, but having access to this will help us understand your current status and give you better help.

3.11   Manuscript

Please follow the AAS Style Guidelines.

You can write the paper in a standard word processor or LaTeX. The final version needs to be submitted as a pdf. If you want to nerd out and have your article look fancy, you may use this AAS journal template. However getting to know and troubleshoot LaTeX is a significant amount of work, so something like Word is also a good choice.

The paper needs to have the following sections. It needs to be submitted as a pdf *without your name on the document.*

Introduction Unless I requested changes, this can be identical to the Introduction section you submitted earlier.

Methods This explains how you did the work. What statistical tools did you use? Did you do any data cleaning? How? What programming language did you use? The person reading the paper should be able to reproduce your work by reading this. They should have all the details they need.

Results What did you find? What did your statistics show? Your Results section must include a figure. (If a figure turns out to be a poor way to share your results, please contact Carrie to discuss an alternative.)

Conclusion What is the meaning of what you found? What does it tell you about the way the moon or craters work?

References References should be in AAS format (see link to Style Guidelines above).

THE TWIST I ask that you include *one significant mistake* in your paper. This will give your reviewer something fun to find. It can be in the Introduction, Methods, or Results section.

Here's examples:

- You cite a reference and say it says (thing) but it does not at all say (thing)

- You improperly deal with missing data

- You use the wrong distribution in your analysis

- Your conclusions do not make sense given your data

   Rubric
   Writing

1. Is your introduction paragraph clear, accurate, and engaging?

2. Is your second paragraph clear and accurate?

3. Is your methods section clear and accurate?

4. Is your results section clear and accurate?

5. Does your conclusion section link your results to the broader scientific picture? Is it clear?

6. Do you have at least four references, properly formatted?

Content (note: the following excludes your one, intentional error that you will describe to me elsewhere).

7. Is your science question clearly stated?

8. Does the data manipulation show a significant amount of effort?

9. Is the data treated properly? In other words, are you using the data in an appropriate way?

10. Is there a significant error analysis/statistical component?

11. Is the error analysis/statistical component appropriate for this type of work?

12. Is the error analysis/statistical component explained and implemented correctly?

13. Are the uncertainties in your results clearly communicated? Can I tell how you treated missing data, for example?

14. Does the manuscript demonstrate a significant amount of work?

15. Is there a figure? Does that figure meet the criteria discussed in the "one nice figure" section?

3.12 Jupyter Notebook

You will need to submit a Jupyter Notebook that shows your work. It should be easy to read and well-explained. Remember, you reviewer will not have access to your notebook, just your paper.

This should be a ipynb file, not a link.

Rubric

1. Can I run your notebook without errors?

2. Is it formatted and organized nicely?

3. Does it include a data-cleaning (or, at least, data-checking) step?

4. Does it include code that does your statistical analysis?

5. Is your statistical analysis correct? (If your intentional error is in the statistical analysis, please show the correct way to do things alongside the erroneous way)

6. Does it include code that produces your plot?

3.13 Associated materials

You will also need to include a reflection document that answers the following questions. I'm looking for about 600 words total.

1. What was the significant mistake you included in your paper?

2. If you had done things correctly (not included the mistake) what would you have done? In other words, what was the correct way to handle the data/do the statistical analysis/ interpret the results, etc.

3. You choose not to submit to a free journal, knowing that they burden libraries with outrageous journal subscription costs. (For example, the publisher Elsevier charged the University of California system $10 million *a year* for access.) Calculate the cost of submitting this article to the Astronomical Journal, following this article charge guide. Please show your work. In other words, include your number of words, number of figures, number of tables, etc. Then include the math of how you calculated the article charges.

4 Peer Review

As Editor of the Journal of Erroneous Results, I will assign your manuscript to a referee. This will be "anonymous"– your name will not be attached at this time. You might be able to guess whose paper you got though, if you heard them talking about it. This is exactly how the real scientific review process works.

As a reviewer, you will evaluate the manuscript for accuracy. You will not have access to the code the authors used. You will produce a written report, called a "Referee Report."

Your referee comments will be kept anonymous, though again, someone might be able to figure out who wrote them. You can reveal your identity if you wish.

Although referee comments in the broader world are sometimes rude, your comments for this class should be polite, kind, and helpful.

4.1 Referee Report

This will cover all aspects of the paper, including the Introduction, Methods, and Results. You will investigate it for accuracy.

Your referee report will be due as a pdf. *Make sure your name is not on the document.* The referee report should be 1-2 pages[2], and answer the following questions.

[2] If you believe you are reviewing an extraordinarily good manuscript and don't have much to say, please contact Carrie to discuss.

1. (18 points) Start the report by describing the good things about the paper.

2. (18 points) Then say what could be improved. This should be the majority of the document. This hopefully includes the significant error you found.

3. (3 points) Please make a recommendation to the editors, in the form of a single concluding sentence, in this form: "Recommendation: (your recommendation here)". Your recommendation options are 'accept without revision,' 'accept with minor revisions,' 'revise and resubmit,' or 'reject.'

5   Final Reflection

(12 Points) You will also need to include a reflection document that answers the following questions. I'm looking for about 600 words total.

This is an individual assignment. Even if you had a partner, I'd like you to do this one individually.

1. Do you think you found the significant mistake in the manuscript? What was it? What other suggestions did you give the author(s)?

2. What took the most time? Was there effort you undertook that was not captured in your Referee Report?

3. What do you think about the peer review process? Do you think it is effective? How could it be improved?

Personal Workstyle Reflection

When do you typically get work done on a project? (e.g. deadline driven: right before a deadline vs liking to finish work early or working on a specific schedule). Do you multitask, or work in a dedicated fashion on a single project to completion?

What is your usual mode of communicating with colleagues? (e.g. tend to over-communication vs prefer minimal, highly directed communication. Hate email, prefer face-to-face, etc.)

How do you organize your work? (e.g. computer files, handwritten notes/planner, vs "organization, what's that?"). Preferred document creation tools? Shared document platform?

What do you feel are your greatest strengths in work style and communication?

What are things in workstyle and communication that you feel are a weakness, or that you would like to change/improve?

What helps you work effectively when working on a team? What drives you crazy?

Team Workstyle Reflection

Where are there synergies in workstyle among your team?

What are possible sources of conflict in workstyle among your team?

How can you structure your team logistics, workflow, etc to leverage individual workstyles to best effect?

What mechanisms/agreements can your team put in place to try to head off workstyle issues, mitigate them, and/or help everyone meet their goals for improvement?

Goals and Expectations

What are your personal goals on this project? (What do you want to learn? Improve at?)

What are the team goals for this project?

Are there situations where your personal goals and the team goals may come into conflict?

What are the team expectations? Be specific. Consider communications, conflict, work styles, timeliness, etc.