JAS 1101H: Introduction to Astrostatistics
Winter 2020

Final Project Description

Purpose

The purpose of the term project is to (1) give you some experience collaborating with someone from another discipline on an interdisciplinary project, (2) dive deeper into astronomy/statistics if you are in statistics/astronomy, (3) learn some techniques for exploratory data analysis, (4) understand how to apply a statistical method to real astronomical data, and (5) improve written and oral communication skills.

By the end of the project, participants should be able to

- describe an astrophysical system and identify a research question
- perform exploratory data analysis on relevant data
- understand the limitations and issues of the data
- apply a statistical method appropriate for the data in order to perform inference
- write a summary report on their findings
- describe their research and present the results to the class

Structure

The project is an application of statistical inference on a real astronomical data set in order to answer a specific science question about an astronomical object or system. The project may involve or be an extension of a statistics topic we are covering in class.

Students will work in groups of three for the project. Each group needs to have at least one student with a statistics and/or computer science background. Students are expected to collaborate, work together, and help each other understand concepts that are new to them. Within each group, all work should be split equally and fairly; everyone should perform some of the actual analysis, provide some interpretation of the results, present some of the research, and contribute to writing the final report. Any code you use must be supplied with the final report and/or correctly cited. Use of version control and collaboration software (e.g. Github, overleaf, etc) is highly encouraged! **Remember, this is a *collaborative* project. Learn from each other and work together.**

There are four stages to the project, and the requirements for each are described below. Additionally, on Feb 26, groups will exchange progress reports and provide each other with constructive peer feedback.

Requirements

1. Choose a project and develop a research question (5%)

Each group must choose a project idea from the list provided and expand upon it, or come up with a project idea together. Once you've decided on a topic, together write a 1 page, single-spaced "proposal" describing:

- what astronomical system, object, etc. you have chosen to work on
- what research question you are asking about this system, object, etc
- what data you will use to try to answer this question
- what kind(s) of statistical analysis/analyses you will apply to these data to perform inference and/or answer your scientific question
- any challenges you suspect might arise

2. Project Progress Report (10%)

The progress report is a single document written by the group, maximum 5 pages double spaced excluding figures, references, etc. The progress report should include the following:

- description of the research question
- description of the data and where it came from (include relevant information about how and when it was collected as you see fit)
- any plots, figures, tables, etc. that you have created up to this point, including appropriate labelling, captions, etc.
- description of any exploratory analysis data analysis you performed, and identification/understanding of potential outliers and/or issues with these data
- a short description of the statistical analysis you intend to perform, and how the output will answer your research questions

3. In-Class Presentation (10%)

Each group must give an in-class presentation on their project during the last week of class. The logistics for the presentations are:

- 30 minutes, split equally among the group members
- 7 minutes for questions
- 3 minutes for audience to fill out a question/answer form (to be handed in to the instructor)

4. Final Written Report (25%)

Each group must submit one written report about their final project. The report is a group effort, and everyone is expected to contribute equally. Each group member must also send me, via email, a short paragraph describing their contributions to the project by the end of term.

Expectations for the final written report:

- Includes at least the following sections: Introduction/Motivation, Data/Exploratory Analysis, Methods/Techniques, Results, and Conclusion/Discussion. Feel free to modify these section titles and add others as you see fit.
- Maximum of 20 pages, double-spaced, excluding figures, tables, references, etc.
- Figures should be publication quality (easy to read, uncluttered, appropriate captions, etc).
- Clearly written, displays critical thinking, and is free of most grammatical and spelling errors.
- Properly cite all references, data, software, etc.
- Submit any code you have written for the project (can be a link to e.g. GitHub)
- You may re-use material that was in your progress report (make sure to incorporate any peer feedback)
- The report doesn't have to be ground-breaking science, but it should display a strong understanding and solid application of the statistical analysis as it applies to the research question and data at hand.

Overall, the report should be well-polished and "publication ready". There are three people in each group – this is a lot of human power and you are all talented graduate students. I believe you will all do some really interesting astrostatistics research and write great reports!

Project Ideas List

Almost all of the following are taken from exercises listed in Appendix C of Feigelson & Babu's book. Feel free to start there and expand upon one of these ideas for your project.

1. Astronomy topic: Globular Cluster (GCs) populations around the Milky Way (MW) Galaxy and the Andromeda Galaxy (M31)
   Possible data: K-band magnitudes of GCs from both galaxies
   Statistics topics: univariate two-sample tests, parametric estimation, truncation of data

2. ~~Astronomy topic: Galaxy clustering on large scales~~
   Possible data: Shapley Concentration of galaxies (i.e. Drinkwater et al 2004 dataset)
   Statistics topics: multivariate clustering, measurement errors, maximum likelihood mixture models

3. ~~Astronomy topic: Moving groups of stars~~
   Possible data: Gaia DR2
   Statistics topics: multivariate analysis, mixture models, measurement errors, bootstrap techniques, Bayesian analysis

4. Astronomy topic: Individual Globular Cluster properties
   Possible data: Webbink (1985), or something more recent!
   Statistics topics: regression, multivariate analysis, measurement error

5. Astronomy topic: Quasars
   Possible data: Sloan Digital Sky Survey
   Statistics topics: regression, multivariate analysis, censored data

6. Astronomy topic: Elliptical Galaxies
   Possible data: Virgo Cluster (data set from Kormendy et al 2009)
   Statistics topics: nonlinear regression, model selection

7. ~~Astronomy topic: Sunspot numbers and the solar cycle~~
   Possible data: NASA's Solar Physics group (https://solarscience.msfc.nasa.gov/SunspotCycle.shtml)
   Statistics topics: time series analysis, spectral analysis (for time series)

8. ~~Astronomy topic: Exoplanets~~
   Possible data: Kepler Spacecraft time series data
   Statistics topics: time series analysis, autoregressive models, measurement error

9. ~~Astronomy topic: Gamma-ray light curves or Gamma-ray bursts~~
   Possible data: see Fermi Science Support Center website (light curves) or NASA's Swift Science Center for data products (bursts)
   Statistics topics: time series analysis, heteroscedastic measurement errors, modelling