

Advanced Data Analysis Techniques for Large Datasets
Phys 434, Autumn 2020
Instructor: Miguel Morales

An introduction to advanced data analysis techniques through a set of computational laboratories. Topics include non-Gaussian statistics, determining the significance of a detection, identifying and mitigating systematic effects in large data sets, data visualization, and collaborative software development. Final two laboratories apply what we have learned to real world data sets. Taught as a senior capstone, and assumes fluency in either python or MATLAB. Prerequisite: Phys 334.

Topics:

- Welcome & basic statistics
- git & GitHub
- Introduction to non-Gaussian statistics
- AND, OR, and convolutions
- Asking a statistical question
- Trials factor
- More statistical examples, Feldman-Cousins
- Finding the background
- Using metadata
- Worries, data exploration, and finding systematics
- The Road Ahead
- Introduction to the LHC & HERA data sets
- Parameters & Confidence Intervals
- Jackknife tests
- Selection Bias
- LHC data analysis primer
- 21 cm Cosmology & putting it all together