Data analysis and error analysis
PHY 535 — Fall, 2011

`http://www.physics.nau.edu/~trilling/teaching/fall2011/phy535/`

Physical Sciences (Building 19) Room 218 MW 9:55 a.m. – 11:10 a.m.

Instructor: Professor David E. Trilling

Course description:
This graduate course is intended to teach a set of critical data analysis and error analysis tools for use in any scientific or applied science application. We will cover a range of topics related to data analysis and error analysis. Chances are good that you will use a number of these tools in your future work, whether in physics or astronomy or another field, and in any of academia, research, or industry. We will not cover any of these topics in great detail. Instead, the goal is to expose you to a range of fundamental tools that are necessary for rigorous analysis of data and measurements. You would be amazed how many of your peers will not have this set of analytical tools. These skills will serve as a critical backbone for your future endeavors.

Course structure and approach:
It's a graduate course. I'm not going to load you up with problem sets and exams. In fact, there will be no exams. Instead, there will be a semester project in which you will present some new work that you have done, based on and extending from material that we have covered in the course. More on this later in the semester. I will also give you some exercises to do – more like practice than like homeworks.

Overall, I hope the class will be more seminar-like and less class-like. Don't be too formal, and for god's sake, don't just sit there quietly – discussion is encouraged. Your attendance at every class meeting is expected.

Textbook and required materials:
Nothing required, but I strongly suggest that you acquire Bevington ("Data reducction and error analysis for the physical sciences"), third edition. You also might find Taylor ("An introduction to error analysis: The study of uncertainties in physical measurements") useful. Taylor's book is better written, but less rigorous; Bevington is more rigorous, but harder to understand. Other resources may be useful as the course goes on.

You also need good, reliable access to a computer. In the last part of the class, you'll need access to a Unix/Linux computer. We have machines in room 232 in this building that are just fine. People from outside the department, you might need to get an access code for that room.

The course web page:
The course web page is given at the very top of this syllabus. I'll post stuff there, including lecture notes, data sets, exercises, etc., throughout the course.

Office hours (how to find me and ask questions):
Just come find me. You know where my office is. Formally, my office hours will be Mondays and Wednesdays 11:30–1:30, but I'm not going to stick too strictly to that. Feel free to make an appointment with me, which is a good way to ensure that we can meet at a specific time.

Grading and assignments:
The final project will be worth 45% of your grade and the homework exercises will be worth 45% of your grade. Attendance, participation, and so on will be worth 10% of your grade.

A note about working together (statement on plagiarism and cheating):
Science works by sharing ideas. I encourage you to work together in this class. However, anything that has *only your name on it* should be written by you and you alone. Let me be clearer about the exercises: I encourage you to work together on these, by which I mean that you can discuss the best way to do the questions and compare answers. However, after working together, you should then write up your assignments by yourself. You should not have identical answers to anyone else in the class. If you do, you have cheated and perhaps plagiarized. This is absolutely not allowed in this class or this University, and I am very serious about this. Cheating pisses me off.

Here's the official statement: This course requires professional and ethical behavior. Plagiarism, or any form of cheating, violates this principle and will not be tolerated. The University regards acts of academic dishonesty as very serious offenses. Students charged with academic dishonesty are subject to the Arizona Board of Regents Code of Conduct and Procedures established by NAU.

Tentative course plan

- Unit 1: Introductory material (2 weeks)
  - Practicalities
  - What is data, what kinds of data sets are of interest
  - Example data sets – yours and mine
- Unit 2: Error analysis (2 weeks)
  - Measurements; precision and accuracy
  - Sampling, aliasing, binning, biases
  - Systematic, random, calibration errors
  - Error on mean, mean of errors, etc.
  - Propagation of errors, RSS, etc.
  - Homework exercise assigned
- Unit 3: Data analysis and properties of data (1–2 weeks)
  - Standard deviation, confidence intervals
  - Binomial, Gaussian, Poisson
  - Homework exercise assigned
- Unit 4: Data handling (2–3 weeks)
  - Significance
  - Best fit line, least squares, curve fitting
  - Correlations
  - χ^2, likelihoods, f test
  - t test, K-S test
  - Homework exercise assigned
- Unit 5: Modeling (1 week)
  - Monte Carlo

  - Bayesian analysis

- Unit 6: Computer skills (2 weeks)
  - Unix/Linux basics
  - Intro to programming – loops, variables, debugging, etc.
  - `awk` and `sed`
  - Python
  - Perl
  - MySQL
  - Small exercises assigned for each topic

- Your projects (3 weeks)

The class meets every Monday and Wednesday with the following exceptions: Sept 5 (Labor Day); Oct 10 and 12 (Trilling travel); and Nov 2 (Trilling travel).