Bayesian vs. Frequentist Statistics
Michael Katz
Your name, Editor

Purpose
Understanding statistics is a fundamental skill necessary for a large variety of real world applications. This lesson will introduce students to the main schools of thought in relation to statistics: Bayesian and Frequentist. The students will learn he main aspects of each school, as well as their unique advantages and disadvantages, as well as how they compare when addressing certain problems in statistics. Additionally, the students will be using the Jupyter Notebook programming environment to illustrate these topics, giving the students exposure to these types of interfacing tools used often in physics, astronomy, statistics, etc.

Overview
The instructor will walk the students through the Jupyter Notebook discussing each each topic and examining the example of testing if a coin is fair. The students should work in groups. The notebook will give a brief introduction to each school of thought with this example in mind. Throughout this procedure, the instructor will ask the students to pass around a coin. Each student will flip the coin a certain number of times and record their results of the number of heads and tails. The students will record their results in their group and show how the answer to the question "is the coin fair" changes as more results come in.

Student Outcomes
• Students will gain exposure to Jupyter Notebooks and their utility.
• Students will gain an understanding of how to examine statistical question from different points of view.

Standards Addressed
SEPs, DCIs, and CCCs addressed.
List the performance expectation by DCI code (ie MS-PS4-d) and include the summary sentence.
(http://www.nextgenscience.org/search-standards-dci)

Time
45 minutes

Level
9th - 12th Grade Astronomy, Physics, Mathematics, Statistics

Materials and Tools
 Computers with Jupyter Notebook installed: link to notebook
 A projector and projector screen
 A coin or multiple coins
 Supplementary handout

Preparation
Install Jupyter Notebooks and download materials.
Test program.

Prerequisites
Students should be familiar with the basics of conditional probability (Probability of A given B) and a normal distribution (just mean and standard deviation).

Background
No programming skills are necessary. Unless the instructor wants to personalize the lesson, everything is setup to require minimal interaction with actual code. The students should understand basics of probability, like conditional probability. The students will use the notebook to answer questions about their specific experiments, as well as Bayesian and Frequentist statistics in general.

Teaching Notes
1) Instructors should begin by ensuring all groups have a coin and a computer with the Jupyter notebook opened and ready to run.
2) Instructor should give a short introduction (~5-10min) about Bayesian and Frequentist statistics that mirrors the introduction in the notebook. They should also explain how to run the notebook (explain you click the cell and push either the play button or shift+return).
3) Students will discuss and work in groups through the notebook while flipping their coins many times and recording their results. The notebook will explain and show examples of the concepts.
4) Students will answer supplementary questions verifying they are learning from the activity. Some questions will be answerable from the notebook. Others will be designed to create discussion after the groups have finished their worksheets.
5) The instructor should gather the class and have the students share their coin findings as well as answers to the discussion related questions.

Assessment

Walking around and ensuring the students are participating.

Collecting the worksheets after the lesson.

Additional Information

Instructions for running the notebook are at the link above.

QUESTIONS

1) What is the Frequentist error in a measurement with n=2, n=100, n=1000 (assume Z=2.5759)?

2) If we assume a uniform prior (meaning we are not including any prior information), what is the probability the coin is fair if we have 2 Heads and 3 Tails? 20 Heads and 30 Tails? 200 Heads and 300 Tails?

3) Under Bayesian statistics, how important is the prior information?
   A. What happens if we assume a completely wrong prior mean in the last example above? Try setting the prior_mean to 0.2 and the prior_standard_deviation to 0.1. Describe the posterior distribution. What is the probability the coin is fair if we assume 400 Heads and 390 Tails?

B. Now assume the prior_mean is 0.5, which is what is expected, and the prior_standard_deviation is 0.01. What is probability the coin is fair with 400 Heads and 390 Tails? Now with the prior_standard_deviation set to 0.25, what is the probability the coin is fair?

4) With a low number of trials (1-5), can you really make statistical statements with Frequentist statistics? What about with Bayesian statistics? With this in mind, what real world and/or scientific problems benefit from proper choice of Bayesian vs. Frequentist?