Honors Fellow Lecture: Data Analysis (March 28, 2018)

Nanda Surendra, Associate Professor of Management Information Systems, College of Business & Economics
Nanda.Surendra@mail.wvu.edu

To install Python program files, download the Anaconda distribution: https://www.anaconda.com/download/
Lecture Files: https://github.com/Nanda-Surendra/honors-fellow-lecture

Useful resources
Nate Silver (2015) The Signal and the Noise: Why So Many Predictions Fail--but Some Don't
Nate Silver's Data Analysis site: http://fivethirtyeight.com/
Ellenberg, J. (2014) How Not to Be Wrong: The Power of Mathematical Thinking

Determining Data Relevance (Any missing data?)

Section of plane	Bullet holes per square foot
Engine	1.11
Fuselage	1.73
Fuel system	1.55
Rest of the plane	1.8

Making Data Connections
A man spots a bear. Afraid, he drops the black cap he is wearing and starts running. The bear chases him. He runs 1 mile south, 1 mile east, and 1 mile north when he finally escapes the bear (the GPS watch he is wearing tracks his run). He spots the black cap he dropped and realizes that he is back at the same place from where he started running.
What is the color of the bear?

Applying Bayesian Reasoning to Data Analysis
If H is the hypothesis and D is the data observed, Bayes Theorem states
P(H | D) = (p(H) * p(D | H)) / p(D)
A simple example (the fruits and baskets problem) helps illustrate Bayes' theorem. There are two baskets of fruits (Basket 1 and Basket 2). There are two types of fruits (apples and oranges). Basket 1 contains 15 oranges and 5 apples. Basket 2 contains 10 oranges and 10 applies. I choose one of the baskets at random and, since I like both oranges and apples, I choose a fruit at random without looking at the baskets. If the fruit I picked up is an orange, what is the probability that it came from the first basket (Basket 1)? Has the probability that I picked from Basket 1 changed (posterior versus prior)?

Cancer screening
A screening test for a certain cancer is 80 percent accurate; that is, if someone has the cancer, the test will be positive 80 percent of the time. If someone doesn't have the cancer, the test is positive 10 percent of the time. 1 percent actually have this type of cancer - about 1,000,000 screenings for this cancer are conducted every year. If someone takes the test and the doctor states that she tested positive, what is the probability that she is likely to have this type of cancer?

OJ Simpson Trial (Prior evidence showed that Nicole Brown Simpson had been abused by OJ Simpson)
One of his lawyers, F. Lee Bailey, used the following statistics to argue that Simpson was innocent:
There are approximately 4 million women abused each year. In 1992, approximately 1,500 were murdered by their abusers. In other words, only 1 in 2,500 abused women ended up being murdered by their abuser.
A total of about 4,500 women were murder victims that year.

Guns and homicides
"Gun ownership and safe societies go hand-in-hand across America. Considering that there are over 350 million guns in America, these shootings are extremely rare. Gun murders kill around 11,000 each year. So, chances of a gun being involved in a homicide are about 11 in 350,000." There are about 15,000 homicides per year, about 11,000 of them are gun-related.