Dr. Tony Diana

DATA 602 Introduction to Machine Learning

Practice Exercises | Lecture Week 2

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Testing Hypothesis in Python

- Generate a normally distributed data, with mean=3 and standard deviation=2, for 10,000 observations:

import numpy as np

mu, sigma = 3, 2

s = np.random.normal(mu, sigma, 10000)

- Plot the normal distribution overlaid on top of a histogram in matplotlib so you get the following result:

A picture containing kite

Description automatically generated

- Create a sub-sample of this population with mean of 1.5, a sigma of 2.0, and 200 observations.

import numpy as np

sample\_mean, sample\_sigma = 1.5, 2

sample = np.random.normal(sample\_mean, sample\_sigma, 200)

-Use:

count, bins, ignored = plt.hist(s, 30, alpha=0.1, density=True)

-Compare both distributions on a chart. The chart should look like this:

A close up of a mans face

Description automatically generated

- Use red for the sample distribution and blue for the real population distribution.

- Use ‘from scipy.stats import norm’ to create 95% confidence intervals. Consider the interval as

plt.axvline(ci[0],color='g')

plt.axvline(ci[1],color='g')

-Create a graph that shows the null and hypotheses, the type 1 and 2 errors that looks like this:

A close up of a logo

Description automatically generated

-Compute the z-score, which is equal to (sample mean – mu)/sigma. What is the p-value and do you reject it?