CS 4630/5630 Kresman Homework 7

Statistics and STEM: Statistics—the science of learning from data—is one of the fastest-growing STEM disciplines. It is also an area of science that is heavily involved in the data sciences pipeline. Your boss wants you to not miss out on the application of statistics before the internship winds down.

**Problem 1**: Jet ski is one f the many ways boaters in Alaska fish (is it for real?), recreate, and access remote lakes, rivers, and coastline in the summer. Jet skis are propelled as the engine-powered impeller sucks in water from beneath the watercraft and shoots a high-powered jet of water behind it. They are equipped with highly efficient, fuel injected 4-stroke engines (sorry for this crap before getting to the meat😊)

A national jet ski manufacturer claims that their model XYZ 4-stroke jet skis last an average of 300 hours under normal use. A researcher randomly selected 20 XYZ jet skis from the manufacturer’s production line and found their mean life span to be 270 hours with a standard deviation of 50 hours. Do we have enough evidence to reject the vendor’s claim?

**Problem 2**: Researchers have had only minimal success in identifying the neurochemical markers for evaluating life satisfaction or ‘happiness.’ In fact, the most common way that researchers assess happiness is through self-reports. And, here is one your boss wants you to dig into.

A research study was conducted to examine the differences between older and younger adults on perceived life satisfaction (Figure 1). Ten older adults (age > 70) and ten younger adults (age 20 to 30) were given a life satisfaction test (known to have high reliability and validity; well, I don’t buy it 😊). Scores on the measure range from 0 to 60 – higher scores mean more satisfied. Do the appropriate t-test to determine if we have enough evidence to suggest that the difference in the two means is significant.

**Problem 3**: Your boss’s nephrologist and cardiologist seem to agree on the connection b/w their disciplines - heart can affect the health of kidneys, and kidneys can affect the health of heart. Your boss, being a hamster lover, is much more curious about the connection b/w hamsters’ heart and kidney size! [Your boss is not off the wall as hamsters are a popular model for many human heart diseases and used to test a variety of therapies.]

Problem 3 Input.txt: Vital stats (heart and kidney size) of a sample of 73 hamsters. Do a linear regression fit of heart size as a function of kidney size. Test the null hypothesis that the population slope is 0 (i.e., no relationship b/w heart and kidney size)

**Notes**

* CS 4630: Do any 2 problems. CS 5630 do all three.
* Check out <https://docs.scipy.org/doc/scipy/reference/tutorial/stats.html>. It can also help compute t statistic, p, and (see stats.t.isf) the T table critical values.
* In class, we did a one sample test, where the null hypothesis was no difference, or mu0 = population mean. For two sample test, null hypothesis is no difference, i.e., mu0: sample 1 mean = sample 2 mean.
* Problem 3: stats.linregress can help compute parameters of interest. It appears that there is no direct way to get the t statistic from python, but the expression for t stat is simple and is easily computed as slopeOfTheRegressionLine / standardError
* State the null hypothesis, alternate hypothesis, alpaha (= 0.05). Compute the t statistic, critical value, and p all in the program and print them (including the null and alternate hypothesis). Finally, print yourconclusion … some like: The null hypothesis is (not) rejected at confidence level \_\_\_, because \_\_\_\_\_\_\_\_\_\_
* If R or MATLAB is in your comfort zone, feel free to use it instead of Python😊 If so, document run, platform & OS instructions please!

**Canvas** submissions (do not zip): (each problem in its own cell) lastnameHw7.ipynb (or MATLAB/R code), a word doc of run snapshot

|  |
| --- |
| Figure 1 |
|  |