

CS 485/ECE 440/CS 585 Fall 2013 Homework 4

Due 11:59pm on Friday, 20 September 2013

Homework 4 is worth 30 points, while normal homeworks are worth only 10. Homework 4 is a group assignment, I'll assign you all into groups of three in class. Please submit your answers for homework 4 in the body of an email with no attachments to "unmnetworkingclass@gmail.com". Do not submit your homework to any other address. You only need to submit homework 4 once per group, so one group member can send in the email and put all three names on it.

You may discuss this homework assignment with anyone, but each group is expected to do their own work. All group members are expected to contribute to all aspects of the assignment.

Below are 10 hypotheses about the data we gathered. You should discuss all ten and address each with a few sentences. Is there enough data to support the hypothesis? Is the data Gaussian? What kind of test could you use?

Each group must choose one of the ten hypotheses and actually carry out the statistical analysis to evaluate the hypothesis to determine whether or not it is true. You should not make any claims without being able to back them up with statistically significant data. I suggest choosing one of the hypotheses where this is actually possible. In other words, there's not enough data to support a hypothesis, it will be difficult to support or refute it. You may prove the converse of one of these hypotheses if you wish (I haven't had time to actually look at the data, so all these hypotheses are based on my own personal biases and might be bassackwards).

Your grade will be 10 points based on your three-sentence assessment of each of the 10 hypotheses, and then 20 points based on your statistical analysis of the one hypothesis that you chose to try to support.

Hypotheses:

1. In our class, international students know more about India than non-international students do, on average.
2. In our class, non-international students know more about football than international students do, on average.
3. People that are 25 years old or older did better at following directions for the data collection than younger students, on average.
4. Undergraduate students in our class know more about Star Wars than graduate students do, on average.
5. People that are 25 years old or older in our class know more about hippies than younger students, on average.

6. On average, people's middle fingers are longer than their pinky fingers.
7. In terms of people's commutes to this class in the morning, graduate students have to come further from home than undergraduate students, on average.
8. Graduate students in our class can throw paper airplanes further than the undergraduate students can, on average.
9. People in our class that followed directions for data collection can throw paper airplanes further than the students who didn't follow directions, on average.
10. People in our class that followed directions for data collection know more about Star Wars than the students who didn't follow directions, on average.

Some recommendations:

- Read up on t-tests. Consider doing a paired t-test, if possible.
- Don't assume things are Gaussian. Prove it.
- Start by getting the data loaded into a good environment where the code for all kinds of statistical tests is already written for you.
- Check out "R", which is basically Matlab for statistics and is free.
- You can save the OpenOffice spreadsheet with the data in it to a csv file and then parse it with script code, if you like.
- Check out Mann-Whitney tests if you can't assume a Gaussian distribution.
- Don't be afraid to attempt one of the more challenging hypotheses if you are so inclined, I'll be more forgiving of errors (and less likely to notice them) if you do more sophisticated statistics than just t-tests.
- Also don't be afraid to stick to t-tests, they're very useful and are a good starting point to learn statistical hypothesis testing.