Hands-on Project for ACMS 20340 Statistics for Life Sciences — Spring 2011

I want you all to do a hands-on statistical project, which will be worth 50 points. The project will involve four stages:

1. Formulating a question

2. Gathering data

3. Statistical analysis

4. Conclusion

You should do the project in groups of size 1, 2, or 3.

Here are some more details for the stages:

1. **Formulating a question**

   You should begin by deciding on an interesting or amusing question that you might like to answer, and that the methods of statistics are useful for answering. The question should involve a hypothesis to be tested, concerning some parameter associated with a random variable. For example:

   - Is the North Dining Hall more popular among male or female students?
   - Are NFL players on average taller than the general population?

   Associated with your question, you'll have to formulate a null hypothesis and an alternative hypothesis. For example:

   - Null hypothesis: No difference in popularity of North Dining Hall between men and women; Alternative hypothesis: there is a statistically significant difference in popularity.
   - Null hypothesis: NFL players are on average the same height as typical American males; Alternative hypothesis: NFL players are on average taller.

   You may have to decide whether to use a two-sided alternative hypothesis (as in the first example) or a one-sided alternative hypothesis (as in the second example).

2. **Gathering data**

   Once you have your question, you will have to collect data. First, you will have to decide on a scheme which selects a random sample from your population. This scheme should be related to the sample space of the random variable. Is the sample space the population of a residence hall, the population of a campus, or a larger (or other) population? For example:

   - If sample space is ND students, choose names randomly from the ND student directory and make calls.

1

- If sample space is football players, choose names randomly from online rosters; for information about heights of the general population, you may have to rely on previously published studies.

Then, you will have to implement the scheme and collect the data.

3. **Statistical analysis**

Once you have your data, you will have to decide on an appropriate statistical test to run to test your hypothesis. For example: Is it a one sided or two sided test? Are you testing a hypothesis about the mean of a population or about the parameter of a binomial population?

4. **Conclusion**

Having run a statistical test on your data, you will finally have to draw a conclusion about your hypothesis. Do you accept or reject the alternative hypothesis, and what does that say about your original question? At what significance level is your conclusion drawn?

You should type up a paper describing the whole process. The paper should begin with a statement of the question you will be addressing, including a clear indication of what is/are the relevant population/populations. You should then explain how you chose to collect data, with a justification of why you feel your method leads to a good random sample, and an indication of any potential problems that might be associated with your data-collection method. Next, you should present the data you have collected and explain what test you are using on the data and why. Then, you should present your conclusion. The paper should end with a discussion of the process, addressing the following questions: Is the conclusion a reasonable one? What factors in the process might have influenced the conclusion and affected its validity? If you were repeating the whole process, are there things you might do differently?

The paper should have a cover page with the course number and name, date of submission, and names of team members; only one paper per team needs to be turned in.

Here is a timeline for the project:

- Friday, March 25: Email me with the names on your team (1-3 people).

- Friday, April 1: Email me a very brief description of your proposed question. I will ok it by April 4.

- Friday, April 15: Data collection should be completed by this date.

- Monday, May 2: The paper is due by this date.

  If at any time during the process you have any questions, please contact me.