Applied and Mathematical Statistics
Homework Two

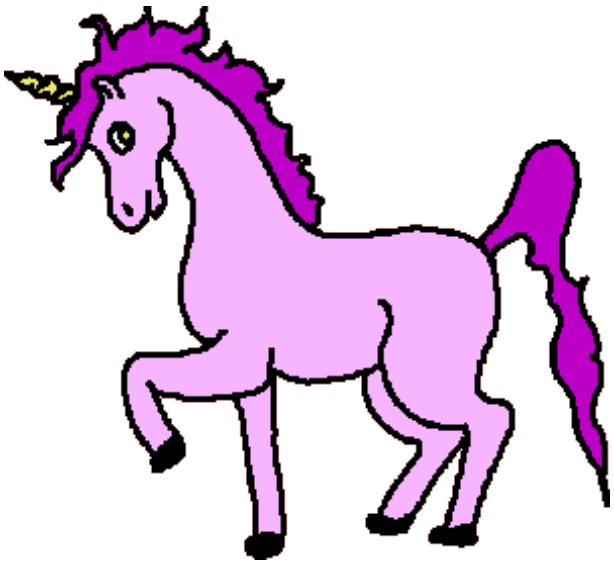

Figure 1: A unicorn. Figure © Lucy Learns Ltd., `www.lucylearns.com/unicorn-picture.html`.

Please do problems marked with ❀ before class on Thursday. The other problems you should aim to do by next Monday, April 11.

1. ❀ You have reason to believe that pink and purple unicorns are equally likely to occur in the wild. You sample 36 unicorns and find that 23 are purple. What do you conclude about this? What is your null hypothesis? What is your p-value? What does it tell you? What is the sample estimate for the fraction p of purple unicorns? What is the 95% confidence for your estimate of p? What does this confidence interval mean?

2. ❀ Repeat the above analysis but for the case where you find 69 purple unicorns out of a total of 108.

3. ❀ Suppose you sample five unicorns and find that none are purple. What can you conclude from this?

4. ❀ Suppose need to figure out the true fraction of purple unicorns and you want to be reasonably sure that your estimate is within 0.04 of the true value. How many unicorns would you need to sample in order to figure this out? (We haven't really talked about how to do this yet. Experiment and ponder, and we will discuss this on Thursday.)

5. ❀ You are trying to figure out what faction of Mainers approve of the job our governor is doing. You interview 136 people and 45 report that they approve of the governor and

everyone else says they disapprove. Assuming that you interviewed an unbiased sample,[1] what is the 95% confidence interval for the fraction of Mainers that approve of the governor?

6. Dalgaard, exercise 3.1, page 65.

7. Dalgaard, exercise 3.3, page 65.

8. Write a python program that calculates p-values for a binomial test. The program should take as input N, the number of trials, p the probability of success, and n, the number of successes observed. Then simulate a large number of trials. (Each trial is one experiment in which a coin is tossed N times.) Use this to estimate the one- and two-sided p-values.

Determine decimal values for all answers.

[1] which is actually very hard to do