Midterm Exam – Intro. to Computational Statistics

Unless otherwise specified, all calculations should be done by hand (without R), and all work shown using math in well-formatted latex. P-values, percentiles, and threshold values / rejection regions should of course always be calculated in R.

1. Using R, write a program to calculate all the prime numbers less than 100. A prime number is a positive integer greater than 1 that is divisible (without remainder) only by 1 and itself. Create the program by testing each number from 2 to 100 against all integers less than it using `%%`. Your function should return a vector of all the primes < 100. (15 pt)

2. Using R, create a histogram of the result from 1 using ggplot. Be sure to nicely label your axes and title the graph. (5pt)

3. You flip a coin five times.

a. What's the chance of getting three or more heads in a row? (5 pt)
b. What's the chance of getting three or more heads in a row conditional on knowing the first flip was a heads? (5 pt)

4. NASA has declared that the Earth is likely to be hit by an asteroid this year based on an astronomical observation it has made. These things are hard to judge for certain, but it is known that the test NASA used is pretty good – it has a sensitivity of 99% and a false positive rate of only 1%. It is further known that the general probability of an asteroid hitting earth in any given year is 1 in 100,000. What is the probability we will actually be hit by an asteroid this year given NASA's test? (10 pt)

5. The average number of snow days in Boston in a winter month is 1. Assuming these events follow a poisson distribution, calculate (using R) the probability of getting 5 or more snow days in a month. (5 pt)

6. You want to know how many hours of sleep the average college student gets. You start out with a preliminary survey of 10 people, and get the following data (in hours): 7,6,5,8,6,6,4,5,8,7. You hypothesize that despite what the doctors say, the average college student does not get 7 hours of sleep a night. What does your survey say? State your null hypothesis, research hypothesis (two tailed), and calculate your threshold value, test statistic, and p value. Do you reject the null or not? (10 pt)

7. Despite the disappointing results in 6, you are confident in your hypothesis. Assuming your sample standard deviation and mean do not change and you want to survey as few people as possible, how many additional people would you have to survey to reject the null at the 0.05 level? (5 pt)

8. You survey the same 10 people during finals period, and get the following hours: 5,4,5,7,5,4,5,4,6,5 . Do college students get significantly less sleep than usual during finals? (10 pt)

9. You are a very bad gardener, and hypothesize that feeding houseplants vodka might help them relax and grow better. You perform an experiment to test your hypothesis, giving 15 houseplants water spiked with vodka, and 15 houseplants water alone. These are your results:

condition	live	die
treatment	4	11
control	8	7

This looks pretty bad for the treatment, but being as good at statistics as you are bad at gardening, you test it using the chi-square test. What are your results? (15 pt)

10. Perhaps you got things backwards, and plants need more stimulation to thrive. So you adjust your experiment into three treatment groups: water, vodka, and coffee. These are your results:

condition	mean days alive	sd	n
water	50	10	20
vodka	45	7	10
coffee	55	4	10

The overall mean is 50 days (as we said, you're a bad gardener). Use an F test to determine if there is any significant difference among these three groups. (15 pt)