STAT/CS 94 Fall 2015 Adhikari HW07, Due: 10/14/15

NAME: SID:

As usual: Please write your answers on a printed copy of this assignment, in the space provided. If you print this assignment yourself, please print exactly this document *double-sided* on a single sheet of paper. (There will be a small penalty for not following this instruction; it makes the graders' jobs more difficult.)

Suggestion: It's hard to write code without a computer, but it's a useful skill, and it will be necessary to write some code on the exams for this class. We recommend you try the coding questions in this assignment that way first for practice, then check your answer on a computer (you can use a cell in a blank notebook or a copy of a lab at `ds8.berkeley.edu`).

Problem 1 *Punnett Plants?*

A species of plant comes in four different varieties. According to a genetics model, the varieties appear like random draws with replacement from a population in which there is a 9:3:3:1 ratio of varieties A, B, C, and D respectively.

A lab has n plants of this species. You can assume that n is large.

(a) True or false (explain): To test whether the model is good, we should use a permutation test.

(b) Fill in the blank with a fraction, and provide your statistical reasoning: If the model were good, the number of plants of variety A would be around _____ of n.

(c) A researcher tests the null hypothesis that the model is good, and gets a P-value of 0.5% (that's one-half of 1%). Circle all that are true among the statements below, and explain.

(i) There is only a 0.5% chance that the model is good.

(ii) There is a 99.5% chance that the model is bad.

(iii) The P-value of 0.5% was computed assuming the model is good.

(iv) The result of the test is highly statistically significant, and the data support the hypothesis that the model is not good.

Problem 2 *Heads Helper*

Define a function `heads` that takes a positive integer `n` as its argument and returns a (random) simulated value of the number of heads in `n` tosses of a fair coin.

Problem 3 *Soda Scramble*

In "blind taste tests," participants are given unmarked containers of two different beverages that look identical and taste similar, and are asked to identify which is which by drinking from the containers. The Pepsi Challenge of the 1970's compared Pepsi and Coke in this way.

Suppose n people take a blind taste test, and k of them make the correct identification. How would you test whether or not the result was just due to chance? Answer the question in the following steps.

(a) State the null hypothesis as a clearly defined set of assumptions.

(b) State an alternative hypothesis. Please be consistent with the question posed in the statement of the problem.

(c) Pick a test statistic and justify your choice. You're welcome to assume that n is even.

(d) Write code to perform the test by repeated sampling. Use r repetitions of the sampling process. You can use any function that you have defined in this homework, but please do not simply call functions defined in class. Your code should return an empirical P-value and a conclusion. Use 3% as your cutoff for "small" P.