Empirical Methods for Computer Science (CS 5970) Homework 1

September 10, 2008

This homework assignment is due on Thursday, September 18th at 5:00pm. Your work may be handed in electronically (use the **Homework 1** digital dropbox on D2L) or in hardcopy form.

This assignment must be done individually: do not share/discuss your answers with others or look at the answers of others.

All data sets are contained within the dat.mat file available on the main homework page.

Question 1

1. (10pts) Suppose we have an **independent** variable that can take on one of two values (A, B), and a **dependent** variable that can take on one of three values (x, y, z). The following table gives the number of occurrences of each combination:

	x	y	z
A	17	45	32
B	28	72	54

Compute χ^2 for the hypothesis that the distribution of the dependent variable is the same given the independent variable (we refer to this as the *null* hypothesis – more on this later). Show your work.

2. (10pts) Suppose we have an independent variable that can take on one of two values (C, D), and a dependent variable that can take on one of three values (x, y, z). The following table gives the number of occurrences of each combination:

	x	y	z
C	15	36	8
D	21	43	16

Compute χ^2 for the hypothesis that the distribution of the dependent variable is the same given C and D. Show your work.

3. (10pts) What can you conclude (relatively) about these two different hypotheses?

4. (10pts) "dat1" is a matrix containing a set of paired samples of an independent variable (column 1) and a dependent variable (column 2). What is $p(v_1)$? What is $p(v_2|v_1)$? According to χ^2 is there a relationship between these two variables?

Question 2

1. (10pts) "dat2" contains several samples of a continuous random variable. Describe (in brief) the distribution of the data. Is this distribution monotonic? Is it Gaussian? (this is not intended to be a long answer; you do not need to do any hypothesis testing)

2. (10pts) "dat3" contains another set of samples from a random variable. Describe (in brief) the distribution of the data. Is this distribution monotonic? Is it a Gaussian?

3. (10pts) Assume that dat2 and dat3 are paired tuples. Briefly describe the relationship between these two variables.

4. (10pts) “dat4” contains yet another set of samples from a random variable. Describe (in brief) the distribution of the data. Is this distribution monotonic? Is it a Gaussian?

Question 3

(20pts) “dat5” contains a set of 4-tuple observations (the data are represented as a single matrix). Describe the relationship (if any) between the four variables and show the process by which you came to these conclusions.

Question 4

(20pts) “dat6” contains two time series (represented as a single matrix). Describe the relationship between these variables and show the process by which you arrived at this conclusion.