Empirical Methods for Computer Science
(CS 5970)
Homework 2

October 28, 2008

This homework assignment is due on Tuesday, October 21st at 5:00pm. Your work may be handed in electronically (use the Homework 2 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 hw2.mat file available on the main homework page.

Question 1

1. (10pts) Suppose that we are performing a robot navigation experiment in a busy building. The foot traffic level is a potential factor that could influence our performance metric. Should we consider this factor as an extraneous or noise variable? Explain in detail.

2. (10pts) What is the relationship between a hypothesis and extraneous variables?

3. (10pts) Explain the process of controlling a noise variable.
4. (10pts) Suppose that in comparing the performance of algorithms A and B, we find that we have a ceiling effect. What could you change about the experimental design to solve this problem?

5. (10pts) Define censoring. Why might we choose to use this technique? Can we always design an experiment to avoid the need for censoring data?

6. (10pts) Define sampling bias.

7. (10pts) True/False and explain: In a factorial experiment we must collect samples for all possible combinations of extraneous variables.

**Question 2**

1. (10pts) Define a sampling distribution.

2. (10pts) Under what conditions is it appropriate to use the Z-test?

3. (10pts) What are the key differences between the Z-test and the t-test?
4. (10pts) Which is more sensitive to differences between distributions: a paired t-test or a two-sample t-test? Under what conditions can you use the more sensitive one?

5. (10pts) In many situations, with enough samples, we can show a statistically significant difference between two distributions. Explain in detail why we may not want to use this capability.

**Question 3**

The matlab variable “dat3” contains a set of 3-tuple discrete observations (the data are represented as a single matrix). Columns 1 and 2 are independent variables; column 3 is a dependent variable.

(10pts) Describe the relationship between variables 1 and 3 in detail.

(10pts) Describe the relationship between variables 2 and 3 in detail.

(10pts) Is there an interaction effect of variables 1 and 2 on their influence of variable 3? One way to ask this question is whether variable 1 changes the effect that variable 2 has on variable 3.