Nonparametric Statistics Homework II.

Due date: Two weeks from today. Worth 5% of your final grade. Computer printout should be included.

1. The binary representation of integers $1, 2, 3, 4, \cdots$, is given by $1, 10, 11, 100, \cdots$. Is the binary sequence $11011100\cdots$ random? Test it based on the binary representation of 10 numbers between 1 and 10. Use the total number of runs as a test statistic. What is your P-value?

2. From Problem 1, use the length of the longest run as a test statistic. What is your P-value? (R or Splus may have a built-in function for the length of the longest run test. Feel free to use it. Matlab does not have a built-in function as far as I know).

3. Suppose that 20 out of 50 females and 10 out of 40 males are depressed in a sample. Determine whether the frequency of depression is related to sex.

4. We are interested in investigating the proportions of four blood types(ABO system). Somehow we believe that 34% of people have blood type A, 15% blood type B, 23% blood type AB, and 28% blood type O. We go out to campus and collect a sample of 100 students, and find the following: A: 12, B: 56, AB: 2, O: 30. What do you conclude? State the null hypothesis and compute P-value.

5. Morton counted the occurrence of various words in several works by Jane Austen: chapters 1 and 3 of Sense and Sensibility, chapters 1,2 and 3 of Emma, chapters 1 and 6 of Sanditon (written by Austen).

	a	an	this	that	with	without
Sense and Sensibility	147	25	32	94	59	18
Emma	186	26	39	105	74	10
Sanditon	147	25	32	94	59	18

   Examine the consistency of Austen's usage of words from book to book. State the null hypothesis. What is your P-value?

6. (Optional problem) You are not required to solve it but if you solve it, you get bonus points. From Problem 1, test the randomness based on the binary representation of 10,000 integers between 1 and 10,000. The test based on the total number of run would be easier. What is your P-value?