BST 140.652 Midterm Exam
Notes:

- You may use your one 8.5 by 11 formula sheet.
- Please use only the basic mathematical functions on your calculator.
- Show your work on all questions. Simple "yes" or "no" answers will be graded as if blank.
- Please be neat and write legibly. Use the back of the pages if necessary.
- Good luck!

_____
signature and **printed name**

1. Two drugs, A and B , are being investigated in a randomized trial with the data are given below. Investigators would like to know if the Drug A has a greater probability of side effects than drug B.

	None	Side effects	N
Drug A	2	4	6
Drug B	5	1	6

   a. State relevant null and alternative hypotheses clearly defining any notation that you use

   b. Perform the relevant test and briefly interpret the results in the language of the problem.

   c. Describe how you could use the Chi-squared statistic to perform an exact two-sided test.

2. Researchers are interested in estimating the natural log of the proportion of people in the population with hypertension. In a random sample of n subjects, let X be the number with hypertension.

   a. Derive a confidence interval for the natural log of the proportion of people with hypertension. Assume that n is large.

3. A researcher is studying migration patterns. She collected the location of the current and previous homes for subjects who moved across regions. She recorded the following:

	Previous home		
Current home	Northeast	Southeast	West
Northeast	-	267	255
Southeast	135	-	139
West	240	234	-

Here the diagonals are not included since she only studied subjects who moved between regions. She would like to know if the probability of moving from region a to b is the same as the probability of moving from region b to a for all regions a and b.

a. Mathematically state her null and alternative hypotheses defining any notation you use.

b. Calculate the expected counts under the null hypothesis.

c. Perform the Chi-squared test and state your conclusions in the language of the problem. (Hint the df is 3.)

4. Benford's law states that the leading digit $d = 1, 2, 3, 4, 5, 6, 7, 8, 9$ occurs with probability $\log_{10}(d+1) - \log_{10}(d)$, yielding the probabilities:

Digit	1	2	3	4	5	6	7	8	9	Total
Probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046	1

The following table was obtained for the leading digits from addresses randomly sampled from the phone book.

Digit	1	2	3	4	5	6	7	8	9	Total
Count	275	183	133	111	76	66	66	44	46	1000

a. Mathematically state the relevant hypothesis for testing whether or not these data follow Benford's law clearly defining any notation that you use.

b. Perform the relevant test and state your conclusions in the language of the problem.