Stat 427 Homework Assignment 4 (due Monday, May 8)
Topics: Goodness-of-Fit Tests, Tests of Independence and Homogeneity of Proportions

1. D&S, page 541, #2; page 542, #5a.

2. To investigate whether birthdays are equally likely to occur across the seven days of the week, I gathered data on the birthdays of 147 "notable personalities" listed in *The World Almanac and Book of Facts*. I found that 17 birthdays were on Monday, 26 on Tuesday, 22 on Wednesday, 23 on Thursday, 19 on Friday, 15 on Saturday, and 25 on Sunday.
a) Carry out a goodness-of-fit test to assess the hypothesis that the seven birthdays are equally likely. State the hypotheses, report the expected counts, calculate the test statistic, and determine the p-value. Also state your conclusion using the α=.05 significance level.
b) Now suppose that the sample size had been five times larger and that the proportions born on each day had been identical. Re-conduct your analysis from a) in this situation.
c) How do the test statistic and p-value in b) compare to those from a)? What does this say about the strength of evidence against the hypothesis of equal likeliness in the two situations? Explain.
d) Determine the smallest sample size that would produce a significant result at the α=.01 level, assuming that the sample proportions would remain the same.

3. The leading digit of a quantity is the first non-zero digit when reading from left-to-right. A result known as "Benford's Law" asserts that the leading digits of many quantities that follow a distribution for which the probability of the leading digit being k is $\log(1+1/k)$, for k = 1, 2, 3, 4, 5, 6, 7, 8, 9, where the log is taken base 10.
a) Use properties of logarithms to confirm that these Benford probabilities sum to one. [*Hint*: Do not convert the probabilities to decimals first.]
b) Now report these probabilities as decimals to three decimal places.
The following table reports the frequencies of leading digits in the populations of the 125 counties of Pennsylvania and California, as reported in *The 2000 World Almanac and Book of Facts*:

Digit	1	2	3	4	5	6	7	8	9
Freq.	38	17	21	16	8	8	6	6	5

Report the expected counts for testing whether the leading digits in this dataset appear to follow Benford's Law. Also show the calculation of the chi-square test statistic and its p-value. State the conclusion of your test, and explain (as if to an educated person with no knowledge of statistics) how it follows from your analysis.
c) The following table reports the frequencies of the leading digits in closing stock prices for stocks beginning with the letter A on April 27, 2001:

Digit	1	2	3	4	5	6	7	8	9
Freq.	45	52	28	29	17	5	9	10	5

Conduct a chi-square test of whether these data appear to follow Benford's Law. Show the details of your analyses, and explain your conclusions.

4. D&S, page 549, #6.

6. Reconsider the genetics example that we analyzed in class. The Japanese morning glory admits a number of recessive traits, two of which are crumpled leaves and variegated leaves.

The following table reports the results of cross-breeding hybrid morning glories bearing one dominant and one recessive gene in each trait:

Phenotype	Normal	Crumpled only	Variegated only	Crumpled and variegated
Frequency	1450	191	184	351

a) Determine the maximum likelihood estimate of the recombination parameter θ based on these data.
b) Conduct a goodness-of-fit test of whether these data appear to follow the recombination model. Show the details of your analysis, and explain your conclusions.

7. Recall the class example about surveying people's views about government spending. The same survey also asked for people's opinion about the federal government's spending on the environment. The sample results are presented in the following two-way contingency table:

	Liberal	Moderate	Conservative
Too little	292	271	239
Just right	85	142	133
Too much	11	27	64

Conduct a chi-square test of independence. Report the hypotheses, expected counts, test statistic, and p-value. Also state a conclusion in context and explain how it follows from your test result.

8. Consider a two-way contingency table on which a chi-square test of independence is conducted. Suppose that you construct a new contingency table by multiplying every cell in the old one by a positive integer constant *c*. (In other words, you increase the sample size by a factor of *c* and find the exact same proportion of observations in each cell of the table.) Determine how the chi-square test statistic for the new table relates to the chi-square test statistic from the old table. Prove that your conjecture is correct.