Stat 401 A XM Homework 1
On-Campus Due Date: September 5
Off-Campus Due Date: September 12

1. Suppose researchers collect a large sample of adults in the United States. They give the sampled people a survey to determine their eating habits over the past 20 years. They classify the people into two groups. One group contains those people who have been eating a diet consisting mainly of organically grown fruits and vegetables. The other groups consists of people who have been eating a more traditional diet. The researchers find that the rate of cancer is significantly lower cancer among people in the first group. The researchers conclude that a diet consisting of organically grown fruits and vegetables reduces the risk of cancer.

   (a) Is this an experiment or an observation study? Explain.

   (b) Comment on the validity of the researchers' conclusion. Can you think of any other reasons why the first group might have a significantly lower rate of cancer?

2. As evidence in support of natural selection in 1898, Hermon Bumpus presented measurements on house sparrows brought to the Anatomical Laboratory of Brown University after an uncommonly severe winter storm. The data below are 24 humerus lengths of a sample of adult male sparrows who died in the storm and 35 humerus lengths of a sample of adult male sparrows who survived. Compute five number summaries for each group. Construct side-by-side boxplots that allow you to visually compare the distribution of humerus lengths for the survivors and nonsurvivors.

Died	Survived	
659	687	749
689	703	751
702	709	752
703	715	752
709	721	755
713	723	756
720	723	766
720	726	767
726	728	769
726	728	770
729	728	780
731	729	
736	730	
737	730	
738	733	
738	733	
739	735	
743	736	
744	739	
745	741	
752	741	
752	741	
754	741	
765	743	

3. Some short problems on summary statistics:

   (a) Provide a set of numbers for which exactly 3 of the 5 numbers in the five-number summary are zero.

   (b) Provide a set of 5 numbers that has an average of 7 and a standard deviation 0.

   (c) For which data set is the standard deviation larger?

$$1, 1, 1, 1, 1, 10, 10, 10, 10, 10 \quad \text{or} \quad 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$$

4. Researchers examined the amount of a particular soybean protein produced the root tissue of each of 6 soybean plants. Prior to measuring the amount of the protein, 3 plants – randomly selected from the 6 – were infected with soybean cyst nematodes. A gel-like substance containing the nematodes was spread over the roots of these three plants to induce the infection. The roots of the other 3 plants were treated with a gel-like substance that contained no nematodes.

   (a) Is this an observational study or an experiment? Explain.

   (b) Examine Display 1.5 in your text. Into which of the 4 boxes might this investigation be classified?

   (c) Explain the purpose of treating the 3 plants with the gel-like substance that contained no nematodes.

   (d) Data from this investigation are provided below. Units on the protein amounts are not provided, but the important thing to know is that the larger numbers are associated with the presence of more protein. Compute a two-sided p-value that provides information about whether nematode infection affects the amount of the protein present in the roots.

		Protein Amounts		
Infected Plants	:	12	16	8
Uninfected Plants	:	7	2	6