

9.07

Introduction to Statistical Methods

Homework 2

Name: _____

1. Fun with Venn diagrams.

A polling agency takes a poll of 200 voters. There are 3 candidates, A, B, and C, running for 3 different offices. The results of the poll show that:

28 voters are in favor of both A and B, 98 voters are in favor of A or B, but not C,

42 voters are in favor of B, but not A or C, 122 voters are in favor of B or C, but not A

64 voters are in favor of C, but not A or B, 14 voters are in favor of A and C, but not B

How many, of the 200 voters, favor (a) all candidates? (b) A (regardless of what they think of B or C)? (c) B (regardless of what they think of A or C)? (d) C, (regardless of what they think of A or B)? (e) A and B, but not C? (f) only one of the candidates?

9.07

Introduction to Statistical Methods

Homework 2

Name: _____

- d. Do you see any patterns? What relationship does the mean of $X+Y$ have compared to the means of X and Y ? What relationship does the standard deviation of $X+Y$ have with the standard deviations of X and Y ? What relationship does the mean of $X+Y$ have to the standard deviations of X and Y ? What relationship does the standard deviation of $X+Y$ have to the means of X and Y ? Generate more distributions as necessary, and see if you can come up with a hypothesis as to how the distributions of X and Y related to the distribution of $X+Y$.

3. The normal approximation.

A researcher measures reaction time (RT) on a particular task, for 100 subjects. The researcher has reason to believe that this reaction time is approximately normally distributed. The sample mean is 10 seconds, which she takes as an estimate of the population mean. She computes, from the sample, an estimate of the population standard deviation (using the formula with the $N-1$ in the denominator). This estimate of the population standard deviation is 2 seconds.

The researcher wants to estimate the probability that someone performing this RT task would do the task in under 2 seconds. Estimate this two ways:

- a. Use MATLAB to generate a normal distribution with the estimated mean and standard deviation. What fraction of the reaction times in this distribution fall below 2 seconds?

Name: _____

- b. Use the z-tables in the back of your book to estimate the probability that someone would do the task in under 2 seconds. How good was your estimate from MATLAB?

4. Estimating the population standard deviation.

In this problem, we'll use MATLAB to look at the bias in estimates of the population standard deviation, for small sample sizes. An estimator is biased when the mean value of the estimate differs from the true parameter being estimated. (Of course, due to chance our estimates will always be off by some amount, but some estimates are better than others...)

To look at bias in the estimate of the standard deviation of a normal distribution, we need to take a small number of samples from this distribution, and compute the standard deviation. We'll do this a LOT of times (100,000), and look at the distribution of estimates of the standard deviation. This is called the "sampling distribution" of the standard deviation.

Generate 100,000 examples of 5 samples from a normal distribution with mean 0 and standard deviation 3. (Note you can generate 100,000 examples of 5 samples from a normal distribution with mean 0 and standard deviation 1 using "x = randn(5,100000);". Note also that this gives you an array with 5 rows and 100000 columns – that semi-colon at the end of the command keeps you from having to look at 500000 numbers! Also note that mean(x) or std(x) will take the mean or standard deviation of each column, i.e. they will return 100000 means or standard deviations.)

Compute the mean, m, for each set of 5 samples (each column). Compute the standard deviation, s1, for each column, using the method where there's an N in the denominator. Compute the standard deviation, s2, using the method where there's an N-1 in the denominator.

Plot histograms for m, s1, and s2. What is the mean of m? Of s1? Of s2? Does the estimate of the mean seem biased or unbiased? How about the estimate of s1? If there's a bias, does s1 tend to estimate low or high? For the estimate of s2, we supposedly have N-1 in the denominator to correct for bias in the s1 estimate – did it work? Is s2 unbiased? Is it an improvement over s1?

9.07

Introduction to Statistical Methods

Homework 2

Name: _____

In addition, do the following problems from the book:

p. 93, #1. p. 235, #4 (see pp. 225-227 for drawing cards without replacement), #5, #7, #8,
#11. p. 252, #5.