9.07
Introduction to Statistical Methods
Homework 4

Name: _____

1. **Estimating the population standard deviation and variance.**
Homework #2 contained a problem (#4) on estimating the population standard deviation. In that problem, you showed that the method of estimating the population standard deviation which has an N in the denominator (std(x,1)) was biased. In other words, this estimate of the population standard deviation was consistently low. You then looked at the estimate of the standard deviation which has an N-1 in the denominator (std(x)), and found that that estimate was better, but still biased – still consistently lower than the population standard deviation. What's going on here?

Well, really the estimate s2 = std(x) is intended to give you an unbiased estimator of the *variance*, not the standard deviation. When the estimate of the variance is unbiased, the estimate of the standard deviation can still be biased.

As in HW2, #4, generate at least 100000 examples of 5 samples from a normal distribution with mean 0 and standard deviation 3. Find s1 = std(x,1) and s2 = std(x). Confirm, as in the previous problem set, that these estimates are biased. Now compute v1 = std(x,1).^2 = var(x,1), and v2 = std(x).^2 = var(x). Comment on whether the estimates of the variance, v1 and v2, seem to be biased.

The correct correction for the standard deviation is s = [1 + 1/(4*(N-1))]*s2. Generate at least 100000 samples of s. What is the mean? Comment on whether this estimator seems biased.

9.07
Introduction to Statistical Methods
Homework 4

Name: _____

2. Kiko received a grade of 55 on a biology test (mean = 50), and a grade of 45 on a philosophy test (mean = 50). She is considering whether to ask her two professors to curve the grades using z-scores. (a) What other information should she consider before making her request? (b) Does she want the standard deviation to be large or small in biology? Why? (c) Does she want the standard deviation to be large or small in philosophy? Why?

3. A researcher measures the performance of students on a reaction time task after they have stayed up all night studying. This task takes an average of 10 seconds when students have not stayed up all night.

(a) The mean reaction time after staying up all night, for 30 students, is 15 seconds, with a standard deviation of 17 seconds. Find the 95% confidence interval for the mean reaction time. Does this interval include the with-sleep reaction time of 10 seconds?

(b) With 50 students, the researcher finds the same mean reaction time (15 seconds), and the same standard deviation. Now what is the 95% confidence interval for the mean reaction time? Does it include the with-sleep reaction time of 10 seconds?

9.07
Introduction to Statistical Methods
Homework 4

Name: _____

4. **Sampling distribution for the difference of two means.** Let x_i, $i=1:n_x$ be normally distributed random variables, with mean μ and standard deviation σ. Let y_i, $i=1:n_y$ be normally distributed random variables, also with mean μ and standard deviation σ. Using the facts we learned about E(), linear combinations of normal random variables, and the variance of a sum of random variables, find the following:
(a) What is the shape of z = mean(x) – mean(y)? (I.E. what kind of distribution is it?)
(b) What is mean(z)?
(c) What is the standard deviation of z (i.e. its "standard error")?

5. x is a random variable with mean 5 and standard deviation 4. I take 5 samples of this random variable, and compute the sample mean, m. Find the range (5-d, 5+d) such that $P(5-d < m < 5+d) = 99\%$. (Hint: for this small sample size, use the t-table.)

9.07
Introduction to Statistical Methods
Homework 4

Name: _____

6. You draw 50 balls, with replacement, from a bag containing only black and red balls. 40 out of the 50 balls you draw are red. With 95% confidence, what is the range of possible values for the actual fraction of red balls in the bag?

7. A candidate wants to conduct a poll to see what percentage of eligible voters favor him over his opponent. How many voters does he need to poll, in order to have the margin of error, for 95% confidence, be 2%? Assume (since you haven't yet done the poll) that the actual percent of voters that favor each candidate is 50%.

8. **Efficiency of the sample mean vs. median as an estimator for the mean.** In this problem, you will estimate the efficiency of the sample mean vs. the median, as estimators for the population mean. We will do this for the special case of 10 samples from a normal distribution with mean 3 and standard deviation 5.

(a) As in previous homework problems, generate 100000 trials of 10 samples from $N(3, 5)$. Take their mean. Take their median. Which estimator is more efficient? (Note: we mean efficiency in the sense of efficiency of estimators, as discussed in class. We do NOT mean computational efficiency, i.e. how quickly does MATLAB finish its computations of mean vs. median!)

(b) Do you think this is a fair comparison between the mean and the median? Why or why not?

9.07
Introduction to Statistical Methods
Homework 4

Name: _____

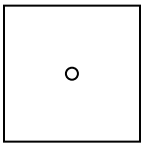
9. **A very strange game of archery.** In the archery example in class, the archer got an arrow in the central 10 cm bull's eye 95% of the time. The central region of the target was a circle. When we looked at the target from the back, and wanted to estimate where the center of the target was, we drew, around each arrow, a 10 cm radius circle, and said that we were 95% confident that the true target center lay within that circle.

What if the central region were a different shape, and the archer still got an arrow in that central region 95% of the time? For both of the following two cases, draw the 95% confidence region around the single arrow shown, i.e. given the location of the arrow, we are 95% confident that the center of the target is within what region? (Assume the very center of the target is as indicated by the small circle.)

Center of target
(front view)

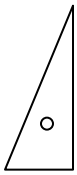
Back of target.
Arrow shown with
"x". Draw the 95%
confidence region.

(a)



x

(b)



x

9.07
Introduction to Statistical Methods
Homework 4

Name: _____

10. A pollster again wants to see what percent of voters favor candidate A over candidate B. 100 voters are polled, and 55% favor candidate A over B. (The rest favor B over A.) With what confidence is the % favoring A greater than 50%?