Week 4 Practice Exam

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Instructions: This is a “low stakes” (i.e., not graded) learning assessment of your comprehension of the first four weeks of this course*.* Compose brief answers to each of the following six questions, typing your response in *italics* below each question.

1. Why do we collect samples of data rather than collect data from whole populations?  
    *Answer: Populations are large and it’s generally hard (expensive, time-consuming etc.) to obtain all values or observations from it. Instead, we resort to sampling from it – meaning, obtaining values or observations from a smaller subset.*
2. Describe the conceptual connection between (“mu,” the population mean) and (“x-bar,” a sample mean). Are they always the same? Or are they always different? Or something else?

*Answer: Population mean is an absolute mean which tends to be unknown.*

*Sample mean is a known average of the sample from the population.*

*Relationship b/n population or sample mean: The sample mean is a good estimate of the population mean.*

*As the sample increases and gets closer to the number of observations in a population, the sample mean converges or gets closer to the population mean.*

1. A large retail franchise company conducted a study of new cashier productivity by examining item UPC code scanning data from a sample of new cashiers at each of 853 different locations. The mean time between item scans across all samples was 4.3 seconds. If you made a histogram of all 853 sample means what would that look like?  
    *Answer: Following the central limit theorem, this would tend to be a bell-shaped curve.*

1. Your boss at the social media marketing company asks you to conduct an A/B test on two different banner ad configurations. Each of the two banners is placed on 86 very popular web pages:   
     
   The A banner gets an average of 1323 clicks per hour.   
   The B banner gets an average of 1394 clicks per hour.   
     
   The 95% confidence interval is as follows:   
    -83 < (mean difference between A and B) < -58.   
     
   Answer the following questions about that confidence interval:   
   1. What is the center of the confidence interval – in other words what is the *point estimate* of the mean difference in clicks/hour between A and B?  
       *Point estimate is -70.5*
   2. Does this confidence interval contain the population mean difference?  
       *May be. We know that if this is a 95% confidence*

* 1. Which banner ad do you prefer (A or B) and why?  
      *B – since the difference mean constantly tends to show that banner B received more clicks.*

*Alternate:*

*B – we are 95% confident that B gets between 58 & 83 more clicks than A*

* 1. Your boss tells you to run the same experiment 99 more times, calculating a new confidence interval each time. Now you have a collection of 100 confidence intervals, each of which was constructed in the same way, but from new data samples: What can you say about this collection of confidence intervals?   
      *We can say is that 95% of these intervals contain the true population mean difference which implies that 95 out of 100 intervals.*