Week 4 Practice Exam

Name: \_\_Quinn Knudsen\_\_\_

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?  
    *It is often impossible to sample the entirety of a population. We when can then extrapolate to the broader population based upon that sample.*
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?  
    *X bar represents the mean of a subset of the population mean mu. Mu is comprised of many X bar (sample means). As the same size or means of means increases, it should approach the population mean, but they are not always identical due to random sampling. Smaller samples and less means of means are more susceptible to outliers and other ill effects of random sampling.*
3. 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?  
    *It would be a normal distribution with a peak at 4.3 seconds (likely the value for the mean, median and mode).*
4. 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?

*-71 is the point difference between A banner and B banner.*

* 1. Does this confidence interval contain the population mean difference?  
      *This confidence interval likely contains the population mean differences (with a 95% degree of certainty). There are non 0 (entirely negative) 95% confidence interval ranges.*
  2. Which banner ad do you prefer (A or B) and why?  
      *I prefer B banner due to the higher average clicks per hour.*
  3. 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?   
      *These confidence intervals will be more accurate than if I had just one confidence interval to base my results off of. With greater sample size I will minimize the ill effects of outliers and sampling deficiencies. Sample size in this instance is our friend.*