Week 4 Practice Exam

Name: George Smith

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?

*We would collect samples rather than collect data from whole populations when collecting data from an entire population would be expensive or time consuming. Statistical inference allows us to make inferences about a population from a random sample.*

1. 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?

*Mu (unobserved) is the population mean while x-bar (observed) is the sample mean. They are not typically related however not typically the same. A sample should be representative of a population so the sample mean should follow suit.*

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?  
    *Normal distribution*
2. 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*

* 1. Does this confidence interval contain the population mean difference?  
     *potentially, there is a 5 % chance that this interval does not contain the population mean*
  2. Which banner ad do you prefer (A or B) and why?
  3. *We would prefer B because we are about 95 % confident that B gets between 58 and 83 more clicks*
  4. 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?

*If we created 100 confidence intervals, we know that roughly 5 of them would not contain the true population mean click difference between A and B and we also know that roughly 95 of them would contain the true population mean click difference between A and B.*