Week 4 Practice Exam

Name: *Kyle Walter*

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. Try to complete the whole exam in an hour. When you are done, check your answers against the key that the instructor will distribute. To make best use of the test, it is important to complete the whole test before consulting the key. If you have unanswered questions about your responses after examining the key, feel free to submit your exam to the LMS for feedback.

1. Why do we collect samples of data rather than collect data from whole populations?  
    *Samples are used because the size of a population will usually cause the ability to collect data to be to expensive, time consuming, etc.*
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?  
    *mu and x-bar are almost always different. However, by using sample means, ie many x-bars we can usually get close to the population mean. An example is 1 sample of 9-coin flips average for heads might be no where near 50% but taken 100 samples we’ll see the number start to converge near 50%*
3. A large retail chain conducted a study of new cashier productivity by examining item UPC code scanning data from samples of new cashiers at each of 853 different locations. The mean time between item scans across all collected data was 2.3 seconds. What would a histogram of the raw data look like? If you calculated a mean for each of the locations and plotted a histogram of those means what would that look like? Would the two distributions look the same?
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 deployed on 98 highly popular web pages during a one-hour test period:   
     
   A banner: mean of 13.23 clicks (per 1000 impressions) across n=98 pages.   
   B banner: mean of 13.94 clicks (per 1000 impressions) across n=98 pages.   
     
   The 95% confidence interval for the mean difference is as follows:   
    *-0.83 < (mean difference, A - B) < -0.58*.   
     
   Answer the following questions about that confidence interval:   
   1. What is value that is at the **center** of the confidence interval – in other words what is the *point estimate* of the mean difference in clicks (per 1000 impressions) between A and B?  
       *The Point Estimate of the confidence interval is -0.705*
   2. Does this confidence interval contain the population mean difference somewhere within its span?  
       *We are 95% certain that the population mean falls somewhere in this confidence interval*
   3. Which banner ad do you prefer (A or B) and why?  
       *based on the samples B would be the better banner having on average .7 more clicks than A*
   4. Write a brief paragraph that provides an interpretation of the point estimate and the confidence interval for your boss. Your boss is an expert at marketing, but knows very little about statistics. Make sure that you clarify the connections between (“mu,” the population mean), (“x-bar,” a sample mean), and the confidence interval.

*We have collected 1000 samples of both Banner A and Banner B clicks and calculated that Banner B is statistically more likely to attract users to click on it on it by .7 more clicks per 1000 people. Because this is a sample, we can say with 95% confidence that the population is likely to click on Banner B between .58 and .83 more times per 1000 people.*

* 1. Your boss tells you to run the same experiment 99 more times next week, calculating a new confidence interval each time. After completing this project, you now have a collection of 100 different confidence intervals, each of which was constructed in the same fashion and with the same sample sizes, but each from new data samples: What, if anything, can you say about this collection of confidence intervals?   
      *There are likely 5 confidence intervals in the this set that do not contain the population mean.*