Practice Exam 2

Name: ____

You may use a calculator and two 8.5” by 11” sheets of notes (front and back). Please plan on bringing your own calculator.

Please show all your work, including all calculations, and explain your answers. Whenever needed, please round numbers (including intermediate calculations) to the nearest 0.001.

Note: This practice test will give you a general sense of the set up for exam 2. However, this practice test is not as similar to the real test as our practice quizzes have been to the real quizzes. I will ask some types of questions on the real exam that do not appear on this exam. You should also review our old labs and homework assignments 6 and 7.

I Conceptual Questions

Please answer the following in a few sentences each. (You don't need to write multiple paragraphs.)

1. Suppose I conduct a hypothesis test about the average amount of tea in a Twinings tea bag. The null hypothesis is that the population mean amount of tea is (less than or) equal to 2.5 grams, and the alternative hypothesis is that the population mean amount of tea is larger than 2.5 grams. In this context, what would a Type I error be? If I change the significance level of the test, α, from 0.05 to 0.01, how does that affect the probability of making a Type I error in this test?

2. A study found that the use of bed nets was associated with a lower prevalence of malarial infections in the Gambia, relative to the prevalence of infections with commonly used mosquito control practices there. A report of the study states that the significance is $p < 0.001$. Explain what this means in a way that could be understood by someone who has not studied statistics. You may use the word ”probability”, but you can’t use any formulas or other statistical jargon like ”statistically significant”, ”null hypothesis”, or similar.

3. Suppose I'm studying cedar waxwings (a type of bird common throughout North America), and I want to estimate how much weight the birds lose during migration. I capture 1000 cedar waxwing birds in Mexico during the winter months, measure their weights, and band them with a unique identifying tag. Then, at the beginning of the summer months I attempt to recapture those same birds in Canada and measure their weights again as they are finishing their migration. I plan to treat the data as paired, and I will construct a 95% confidence interval for the population mean difference in weights before and after migration using the weight differences for the sample of birds I am able to recapture. Are the following statements true or false?

   (a) If I am only able to recapture 10 of the birds again in Canada (so that $n = 10$), my 95% confidence interval will be **wider** than it would have been if I had been able to recapture all 1000 birds (so that $n = 1000$).

   True False

   (b) If I am only able to recapture 10 of the birds again in Canada (so that $n = 10$), my 95% confidence interval will be **less likely to contain the population mean weight difference** than it would have been if I had been able to recapture all 1000 birds (so that $n = 1000$).

   True False

4. In a t test about a population mean, what is the test statistic? Address the following points:

   (a) Write down the formula for the test statistic

   (b) Write a sentence or two describing what the statistic measures

   (c) If the test statistic has a very large value, is that strong evidence against the null hypothesis or weak evidence against the null hypothesis? Explain why, with reference to your answer to (b).

II Applied Problems

1. We are interested in estimating the proportion of graduates at a mid-sized university who found a job within one year of completing their undergraduate degree. Suppose we conduct a survey and find out that 348 of the 400 randomly sampled graduates found jobs. The graduating class under consideration included over 4500 students.

   (a) Describe the population parameter of interest and the sample statistic. What symbol is commonly used for these?

   (b) Check whether the conditions for constructing a confidence interval and conducting a hypothesis test based on these data are met. (For each condition, write a complete sentence answering whether or not that condition is met and why.)

(c) I ran some R code and found that a 95% confidence interval for the population parameter was [0.833, 0.901]. Interpret the interval **in the context of this problem**. As part of your answer, describe what the phrase "95% confident" means.

(d) According to the National Center for Education Statistics, the proportion of all 20-to-24 year olds with college degrees who were employed as of 2015 is 0.89. Write down a statement of the null and alternative hypotheses for a test that the employment rate for the current graduating class is different from the national average among 20 to 24 year olds.

(e) I used R to compute a p-value for this test, and I got a p-value of 0.2. **In the context of this problem**, what is your conclusion? Use a significance level of $\alpha = 0.05$.

(f) How could you have conducted this test using the confidence interval from part c?

2. (16 points) Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these officers had an average lead concentration of 124.32 micrograms/liter and a SD of 37.74 micrograms/liter; a previous study of a large number of individuals (not all police officers) from a nearby suburb, with no history of exposure, found an average blood level concentration of 35 micrograms/liter.

   (a) What is the population parameter of interest? Describe it in a sentence and state what symbol is commonly used for this parameter.

   (b) Write down the hypotheses that would be appropriate for testing if the police officers appear to have been exposed to a higher concentration of lead than their neighbors in the suburbs. You may treat the value of 35 micrograms/liter from the suburban study as a fixed, known constant (i.e., we are not structuring this as a test to compare the means of two groups, but rather as a test about the mean value for the group of urban police officers).

(c) Explicitly state and check all conditions necessary for inference on these data. If you don't have enough information, say what you would want to know. If you would want to look at any plots, describe what plot(s) you would want to make and what you'd be looking for.

(d) Calculate a p-value for the test that the downtown police officers have a higher lead exposure than the group in the previous study. **For full credit, you must show all of your work!** Points are assigned to correct set-up. In doing this calculation, you may use the following facts (you just need one of these numbers - which one?):

- If $T \sim t_{51}$, then $P(T > 23.754) < 10^{-16}$
- If $T \sim t_{51}$, then $P(T > 17.067) < 10^{-16}$
- If $T \sim t_{51}$, then $P(T > 3.294) = 0.001$
- If $T \sim t_{51}$, then $P(T > 2.367) = 0.011$

(e) Draw a picture of a relevant t distribution. Label it with the value of the test statistic you found in part (c) and shade in the region corresponding to the p-value.

(f) Interpret the results of your test in context. What is your conclusion?

(g) Suppose that you rejected the null hypothesis (which you may or may not have actually done based on the data above - this is a hypothetical question). Would this prove that there was a causal relationship between exposure to car exhaust and increased concentrations of lead in the blood? Explain why or why not.

(h) Find a 99% confidence interval for the population parameter, and interpret it in context (including a description of what the phrase "99% confident" means.) You may use the following R output to help in your calculation (you only need one of these numbers - which one?)

```
> qt(0.995, df = 51)
[1] 2.676
> qt(0.995, df = 52)
[1] 2.674
> qt(0.99, df = 51)
[1] 2.402
> qt(0.99, df = 52)
[1] 2.4
```