

# STA 305/1004 Winter 2021 - Assignment 2

Instructor: Shivon Sue-Chee

posted on Saturday, February 6, 2021

## General Instructions

- Due: Electronic submission into Quercus by **10pm on Sunday, February 14, 2020**.
- Late assignments will be subjected to a penalty of 1% per hour late. Submissions will not be accepted beyond 48 hours of the due date. Email submissions are not allowed.
- Students who would like additional accommodations should email the instructor at least 48 hours before the assignment is due.
- Use RMarkdown to write and show **all** of your codes and answers.
- Submit your knitted RMarkdown file in pdf or docx format.
- Use a benchmark significance level of 5%. Report  $p$ -values to 2 significant digits.

## Part I (3 marks)

Suppose that you are an engineer who plans to design a simulation study to compare the average lifetime of LED bulbs under two different temperature settings, namely,  $S^{\circ}C$  and  $T^{\circ}C$ . Suppose that the distribution of the bulb's lifetime is known to follow an exponential distribution with rate parameter  $\lambda$ ,  $\lambda > 0$ . The density of this distribution is

$$f(x) = \lambda \exp(-\lambda x), \quad x \geq 0.$$

The mean and standard deviation of the exponential distribution is  $1/\lambda$ . This parameterization is in units corresponding to the reciprocal of time.

You hypothesize that the expected lifetime of LED bulbs is 3 years under temperate  $S$  and 1 year under temperature,  $T$ . Randomly generate a sample of 18 data points to form the observations under two experimental designs: a completely randomized design and a randomized paired design, to compare the average lifetimes between the two groups -  $S$  and  $T$ , by carrying out the following steps:

1. Set the seed of your randomization to be your student number.
2. Randomly generate 9 observations from the  $Exp(\lambda = 1/3)$  distribution to correspond to treatment  $S$ . List the observed values, to 3 decimal places, and the order in which they appeared. (Hint: A random sample from an exponential distribution can be generated in R using the function `rexp(n = , rate =)`.)
3. Randomly generate 9 observations from the  $Exp(\lambda = 1)$  distribution to correspond to treatment  $T$ . List the observed values, to 3 decimal places, and the order in which they

appeared.

4. Use the order of the observations in 2) and 3) to form pairs of observations. Display the pairs of observations of treatment  $S$  and  $T$  for the randomized paired design.

## Part II (12 marks)

For both designs, based on the data simulated in part I, conduct a randomization test to compare the means of the two treatments.

- i. Describe the randomization distribution for this comparison by stating the number of values that this distribution contains and the probability of the observed treatment allocation?
- ii. Create a histogram of this randomization distribution; include vertical line(s) to mark the area(s) corresponding to the P-value.
- iii. Use the randomization test to determine if there is evidence of a difference in means between the two treatments. Explain your answer, including the P-value of your test.

## Part III (10 marks)

For both designs, based on the data simulated in part I, conduct a t-test to compare the means of the two treatments. Note: Assume that the population distributions and parameters are unknown.

- i. Explain your answer, including the P-value of your test.
- ii. Are the assumptions behind the  $t$ -test satisfied?
- iii. Do the results of the  $t$ -test agree with the results of the randomization test? Explain.

## Part IV (10 marks)

You realize that the data does not follow a normal distribution and you should use a non-parametric method called the Mann-Whitney Test (also called the Wilcoxon- Rank Sum test) to compare the average lifetimes under the two conditions for both designs. To implement this test in R, we use the function `wilcox.test()`. If 20 bulbs are tested under each temperature condition, find and compare the power of the tests to detect a difference in means at the 5% significance level for the following cases:

- i. Completely randomized design and  $t$ -test,
- ii. Randomized paired design and  $t$ -test,
- iii. Completely randomized design and Wilcoxon test,
- iv. Randomized paired design and Wilcoxon test

To simulate power and produce reproducible results, use your student number to set the seed of your randomization and make 1000 replications.

Explain which statistical test you would recommend for each of the two experimental designs and why.