Homework 2, STAT 252, AY 22-23, SS
Course instructor: Michael Daniel Lucagbo
Due date: May 22, 2023, before midnight

Homework instructions

- Your homework must be submitted with a careful and concise write-up of the algorithms and the results. Email me your answers as portable document format (PDF) files. Any necessary codes should also be included in the file/s that you submit. However, a solution to a problem that consists only of software code and output will receive no credit.
- For the algorithms you write, use clear notations. Define the notations that you use.
- You may work individually or in groups of at most three members. Discussions with your classmates is encouraged. However, copying someone else's work or some other group's work is NOT allowed. If this happens, both parties will get a 0 on the assignment, and further proper disciplinary action will be taken.
- You are encouraged (but not required) to use R Markdown. For an introduction to R Markdown, you may watch this video: https://www.youtube.com/watch?v=DNS7i2m4sB0

---

1. As a small example, the survival times of 9 rats were 10, 27, 30, 40, 46, 51, 52, 104, and 146 days. Because of the skewness in the data, consider estimating the population median survival time θ through the sample median.
   (a) Compute a 95% CI for θ using the bootstrap-t method. Use $B_1 = 1000$ first-level bootstrap samples and $B_2 = 50$ second level bootstrap samples (to estimate the standard error). Interpret the CI.
   (b) Compute a 95% CI for θ using the bootstrap percentile CI with $B = 1000$ bootstrap samples. Interpret the CI.
   (c) Compute a 95% confidence interval for the mean time between failures μ using the basic bootstrap method with $B = 1000$ bootstrap samples. Interpret the CI.
   (d) Plot a density estimate of the data. In R, you can do this through the **density** function. Compare the results in parts (a), (b), and (c). If there is a difference in the results, does it have to do with the shape of the data?
2. Consider the spatial test data from Table 14.1 of Efron and Tibshirani (1993) shown below. From the table's description, it is clear that the measurements A and B are paired. Suppose the data consist of a random sample from an unknown joint distribution of A and B. Whenever ratios are scientifically or statistically preferred to differences, we gain stability by considering the logarithm of the ratios. Let $\theta_1 = \log E\,(A_i/B_i)$, $\theta_2 = E\,(\log A_i/B_i)$ for all i. In answering (a) to (d) below, exclude observation #14 because the logarithm of its ratio is undefined. Use 2000 bootstrap samples.

Table 14.1. *Spatial Test Data; $n = 26$ children have each taken two tests of spatial ability, called A and B.*

	1	2	3	4	5	6	7	8	9	10	11	12	13
A	48	36	20	29	42	42	20	42	22	41	45	14	6
B	42	33	16	39	38	36	15	33	20	43	34	22	7

	14	15	16	17	18	19	20	21	22	23	24	25	26
A	0	33	28	34	4	32	24	47	41	24	26	30	41
B	15	34	29	41	13	38	25	27	41	28	14	28	40

(a) Compute a bootstrap percentile confidence interval for θ_1. Interpret the CI.

(b) Compute a BC_a confidence interval for θ_1. Interpret the CI.

(c) Compute a BC_a confidence interval for θ_2. Interpret the CI.

(d) Compare your CIs in (a) and (b). How different are the two CIs?