Mia Monintja (am2366a)

Yashvi Malviya (ym7137a)

Kevin Norris (kn7295a)

30 November 2023

DATA 613 Data Science

Data Science: Final Project Status Update: Computing “n = 1 inference”

Wall, Boen and Tweedie (2001, hence WBT) offer an approach to calculating a finite confidence interval for the mean of a normal distribution when n = 1.

x ~ N(µ, σ2) represents a Normal distribution.

x = sample observation that follows this Normal distribution

µ (mu) = mean of the dataset

σ2 (sigma squared) = variance of the dataset (square of the standard deviation)

Here, the confidence interval is of the following form:

𝜇 ∈ 𝑥 ± δ‖𝑥‖, where δ (delta) represents a positive scalar.

For our final project, our team will create an R package for computing confidence intervals for means when the sample size is n = 1. We will focus on the following aspects when maximizing functionality:

1. We will need to demonstrate a couple of methods for deriving δ, either through:
   1. numerically solving an equation for δ
   2. closed-form approximations (basic operations – addition, subtraction, etc.)
2. Our user should have the ability to set the coverage probability. In this case, “coverage probability” represents the opposite of the confidence interval. For example, if our confidence interval is 95%, the coverage probability is 1 – 0.95, or 0.05, represented by α (alpha).
3. Our package should allow the user to shift x by some constant ‘a’ and obtain:
   1. a new Normal distribution (x – a) ~ N(µ – a, σ2), and
   2. a new confidence interval 𝜇 ∈ (𝑥 – a) ± δ‖(𝑥 – a)‖.
4. In our package, we will allow the user to obtain p-values against a null (this allows the user to test a null hypothesis against some alternative hypothesis; for example, H0: p-value < 0.05, HA: p-value > 0.05)
5. We also hope to extend our package to finding confidence intervals of a similar form for when n >1; however, WBT finds that results are not ideal when n > 2.

So far, we have successfully set up our git repository for us to share our work with each other. An important step in the process of collaborative learning and working together, our git repository will serve as the hub for all our ongoing contributions to the final project. We have also set up our package skeleton so that we can begin the process of writing our functions and putting our package together.

In the next few days, we need to begin writing our description file and doing research on the statistics behind our work so that we can put a package together that reflects the research and efforts of WBT. A large portion of our grade lies in detailed and ongoing updates to our package, as well as a strong description that explains all the functions at work. We also want to make sure that we include the statistical plotting that the project calls for, since we have graduate students in the project team.

We are excited to take our excitement about working with statistical procedures in other classes and apply them to the creation and implementation of an R package for Data Science. With a lot of hard work, we hope to produce a package that can stand with some of the packages already out there in the world of R programming.