

HW1

Steven Barnett

8/29/2020

Problem 2

Part A

I graduated in April of 2016 from Brigham Young University with a bachelor's degree in computer science. I have worked as a software engineer at Qualtrics for the past four years. Because of this experience, I am confident in my coding ability in several different languages and frameworks. However, I have limited expertise in statistical analysis and the required computational tools. My goal this semester is to expand my knowledge of how to perform statistical analysis and master the tools used to perform and display such analysis.

My learning objectives for STAT 5014 are as follows:

1. Learn and be comfortable developing solutions in the R language.
2. Be able to communicate data analysis effectively through the use of R Markdown.
3. Become familiar with Advanced Research Computing (ARC) for graduate research.

Part B

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0 \quad (1)$$

$$f(x|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \frac{1}{(1 + (\frac{x^2}{\nu}))^{(\nu+1)/2}}, \quad -\infty < x < \infty, \quad \nu = 1, \dots \quad (2)$$

$$f(x|p) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}; \quad 0 \leq x < \infty; \quad p = 1, 2, \dots \quad (3)$$

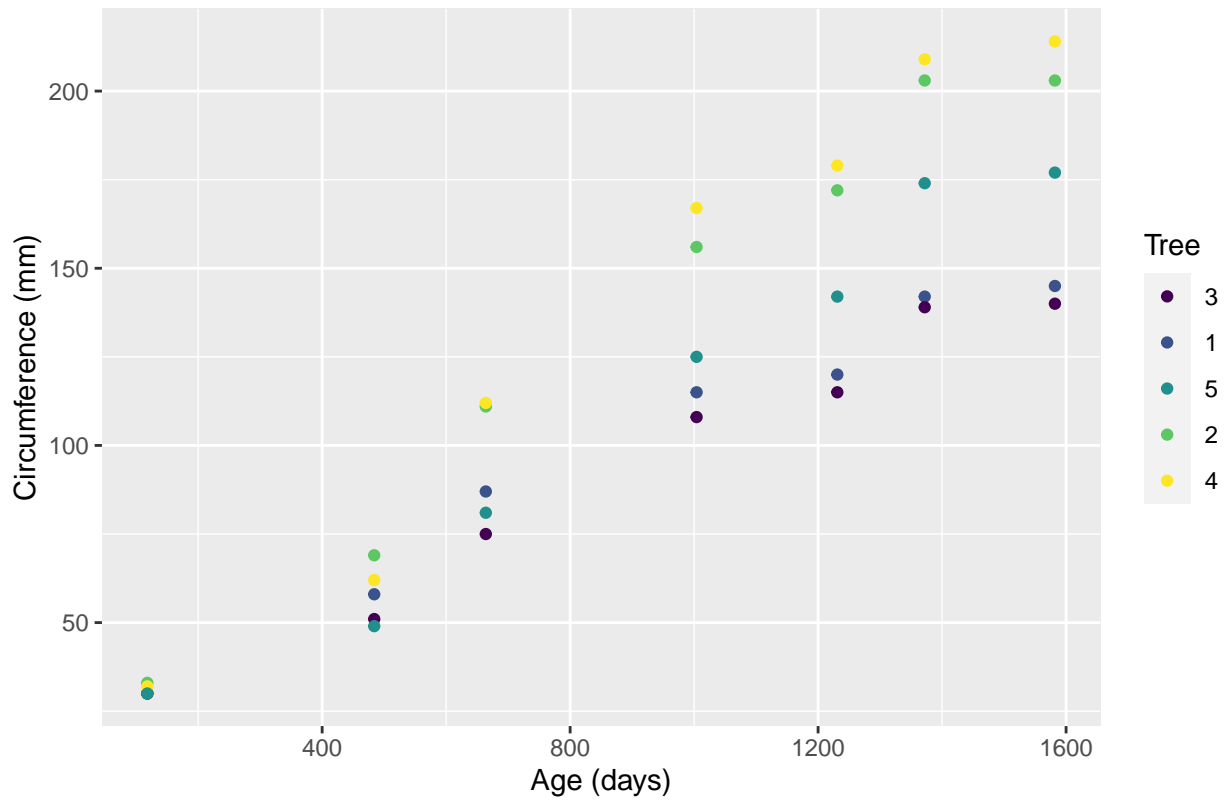
Problem 3

1. Record results and the steps taken to achieve them in order to reproduce those results at a later date.
 - It can be difficult to have the motivation to document each step when trying to produce the desired results. Typically someone is trying multiple different ways and documentation can slow down the process.
2. Convert manual steps to manipulate data into a reproducible list of commands or a custom script.
 - It can be easier in the short-term to manually manipulate data. Articulating the process and putting it in code can take longer, but will save time in the future.
3. Write down the names and versions of programs and frameworks used to produce results.
 - It is easy to document the names and versions of programs used. However, if the version is too old when the result needs to be reproduced, it can be difficult to access the software necessary.

4. Use a version control system when developing custom scripts and other code.
 - Fear of and unfamiliarity with version control systems can be deterrents for executing this step. Additionally, being entrenched in a manual way of version control can prevent someone from adopting an external system.
5. Document intermediate results in a format that is easy to understand, rely on, and track.
 - It is hard to produce a system that is easy to understand and step through when there isn't a set process in place or if processes could change in the future.
6. If a process relies on randomness, log the random seeds that are used when results are generated.
 - Documenting the random seeds is something that is easy to forget. We can easily be caught up in the generation of results and forgot what needs to be written down.
7. Save the raw data that is used to produce plots showing results. Raw data can be used later to modify the plots when revisions are suggested.
 - There can be so much data behind a simple high-level result. Storing all that data can be cumbersome, especially if we don't have a system in place to organize it.
8. Generate output in a way that allows one to inspect different layers of detail that support the main result.
 - Researchers are concerned with an end result. Once that end result is achieved, producing different layers of detail behind the result can seem unnecessary.
9. Enable easy connection of summaries and text analyses of data with the data itself.
 - Data and text are not usually stored in the same location. It takes an extra step to make sure they are connected. That extra step can be easily forgotten.
10. Allow the public to see results and any code or documentation on how the results were achieved.
 - It takes courage and being vulnerable to let the world see all of your work and not just the end result. However, posting all this information creates more trust and robustness to one's work.

Problem 4

Growth of orange trees with respect to age and type



Circumference of Orange Trees

