HW1_ahcooper

Andrew Cooper

8/30/2020

Problem 2

Part A

1. I hope to really understand the different types and classes that exist in R, as those concepts have frequently confused me in the past.

2. I hope to pick up useful tools and tricks regarding R and R-studio that can help me in future research and projects.

3. I hope to gain enough confidence with basic git commands such that I can confidently utilize them without having to look them up.
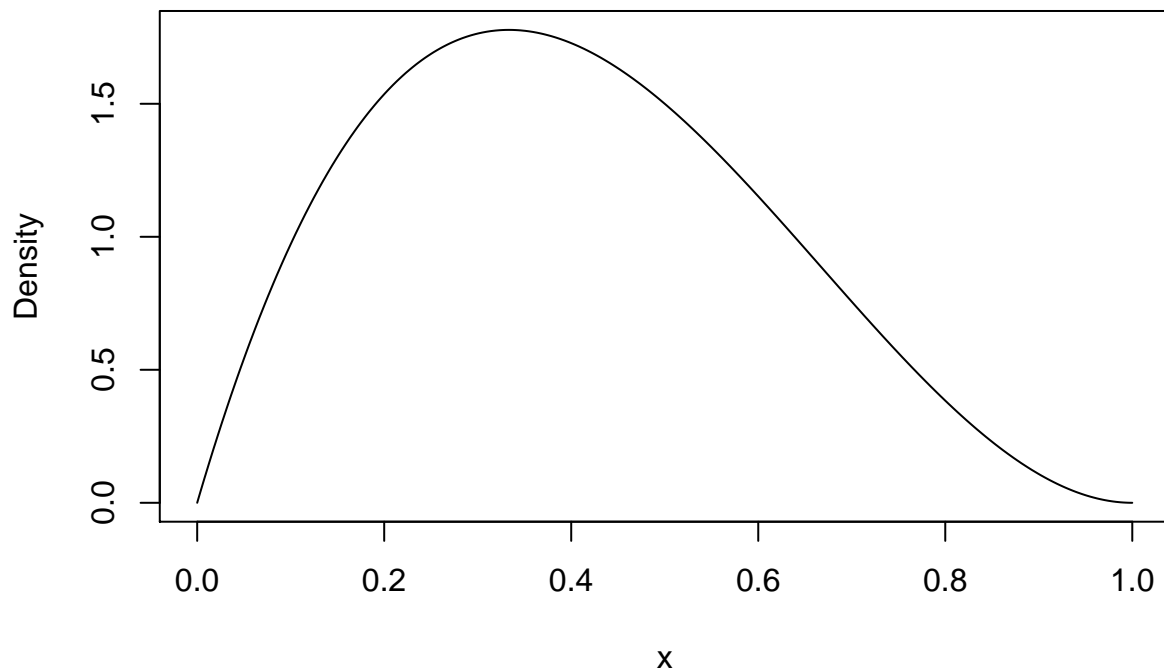
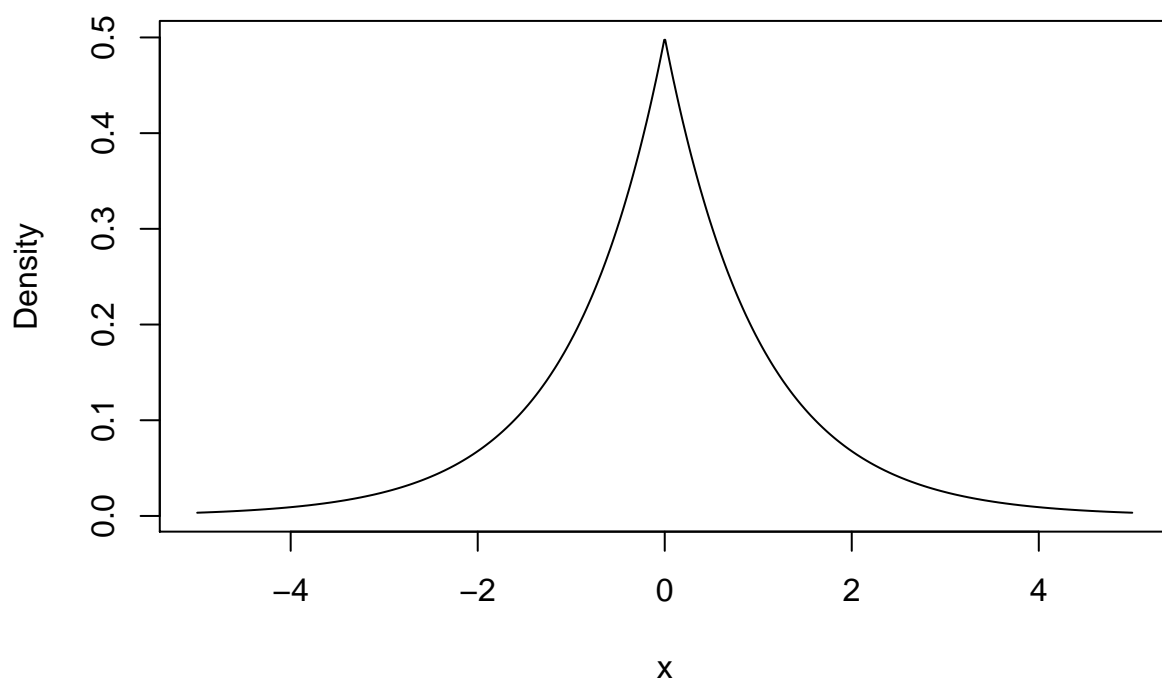### Part B

Figure 1: Beta Distribution (alpha = 2, beta = 3)

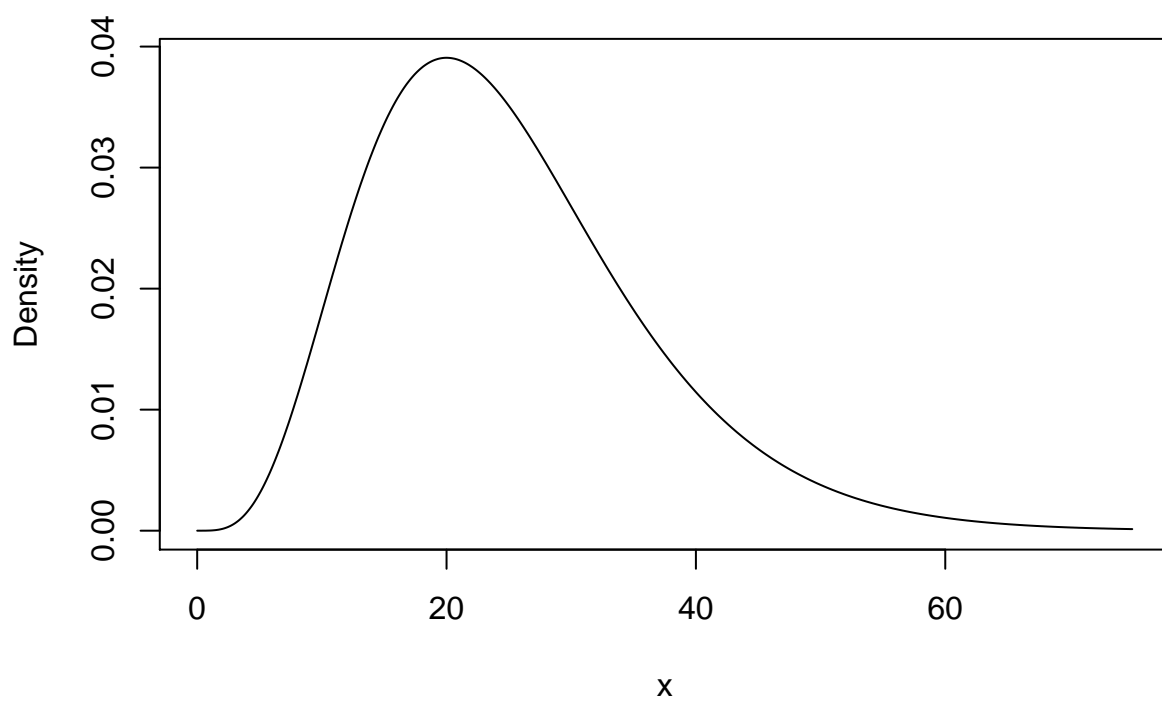Figure 2: Double-Exponential Distribution (mu = 0, sigma = 1)

Figure 3: Gamma Distribution (alpha = 5, beta = 5)

Problem 3

1. Keep a regularly updated documentation of all the steps taken in your analysis. Oftentimes research tends to deviate from the outline planned at the beginning, either due to technical obstacles or changing research interests. It is important to account for this, and make sure the documentation aligns with the reality of the research process as closely as possible.
   - When the process of documentation is manually generated, you leave it open to human-error, where people fail to include important details about the research process, or forget to change earlier documentation that is outdated. It would be ideal to have the documentation automatically generated somehow.
2. Don't subjugate data modifications to human-error whenever possible. Make changes using an automated process that is more reliable, less error-prone, and easier to document and keep track of.
   - For some specific tasks it can be challenging and time-consuming to create an automated solution when a manual solution could be significantly quicker.
3. Keep track of the current versions of all software used.
   - Software could look completely different in the future due to changing technology. For instance, if we look back to papers from decades ago that use software that aren't used anymore, trying to find the exact version of the software used could be difficult or even impossible.
4. Keep track of the current versions of all scripts involved in analysis.
   - Keeping every version of a script, including logging every minute change and update made throughout the research process, can sometimes make sorting through the code difficult when wanting to reproduce the analysis. If the most update-to-date version isn't working for some reason having to dig through several previous versions to find the working version can be time-consuming.
5. Keep track of the state of analysis/results in between steps.
   - It is not always obvious where or how frequently one needs to provide these intermediate results. Keeping track of results to frequently or with too much detail can be time-consuming.
6. Make sure stochastic elements don't prevent reproducibility.
   - This may be related to rule 3, but the underlying pseudo-random structure used by software to mimic randomness could potentially change in an update. Perhaps it might be impossible to find or use the exact random seed used in a research paper, which would prevent results from being replicated.
7. Make the data behind each plot/visual accessible, without any changes or transformations made to make the visual.
   - The raw data made to create a plot could be very large and hard to store or include in a paper in its entirety, or might be too sensitive to include in a raw format.
8. Express results with varying levels of specificity to make sure results are clear but reproduced results can be compared to the original results in necessary detail.
   - What is considered a "detailed" report is subjective. What some researchers might feel is a clear explanation of results with sufficient detail could be ambiguous to other researchers trying to reproduce the study.
9. Make the connection between results and discussion clear and easily traceable.
   - Variation in formatting amoung different journals can make the connection between results and discussion difficult. For example, some journals move figures and plots to the end and change figure labels, making it harder as a reader to understand what exactly a written analysis is referring to.
10. Make public all related materials needed to accurately reproduce research.
    - If the nature of the research is sensitive, being able to make public all the materials might be difficult or impossible. In addition, regulations or companies that wish to keep proprietary analytical tools or data private might make this step difficult as well.
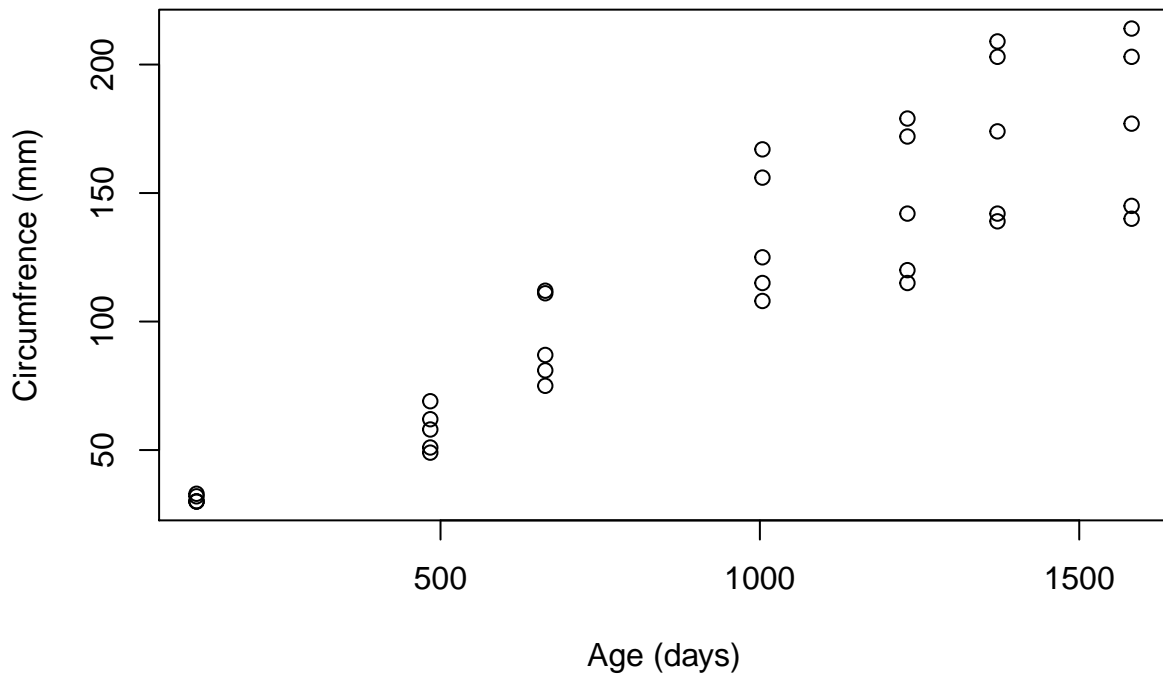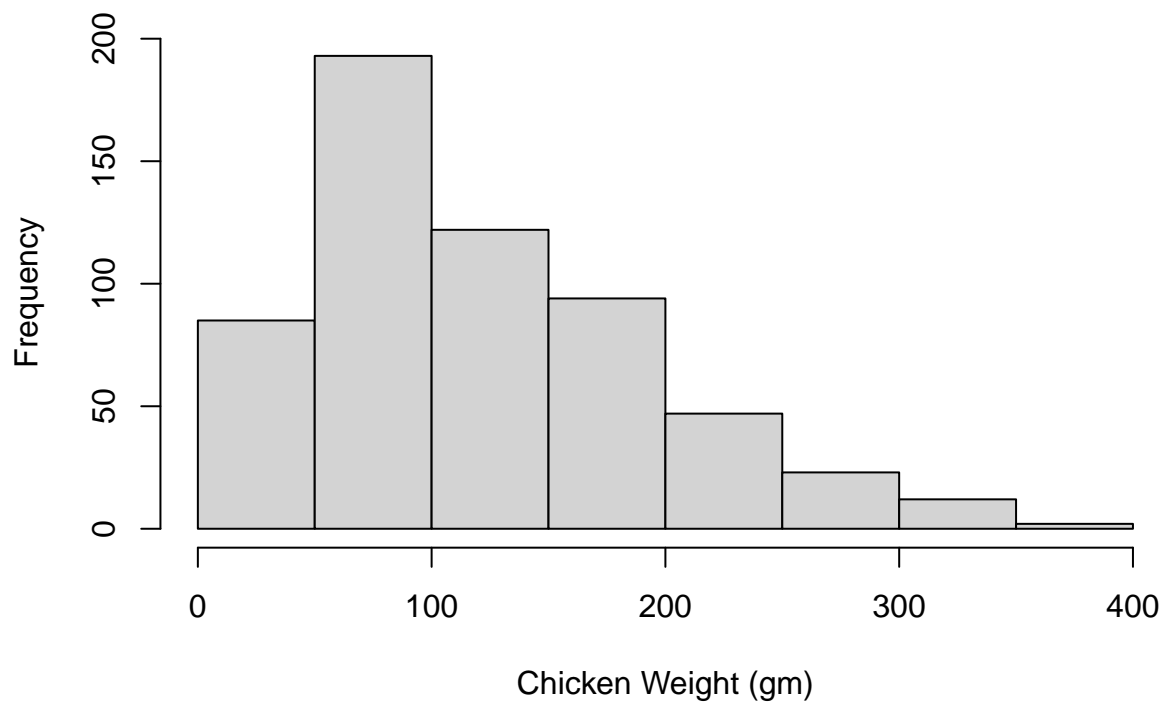
Problem 4

Figure 4: Orange Tree Size vs. Age

Figure 5: Weights of Chickens