

Homework 5. Due by 5pm on Thursday 10/5.

Data and the reproducibility of research results

Read pages 8–11 of *On Being a Scientist* and the Wikipedia page on data sharing ([wikipedia.org/wiki/Data_sharing](https://en.wikipedia.org/wiki/Data_sharing)). Write brief answers to the following questions, by editing the tex file available at <https://github.com/ionides/810f19>, and submit the resulting pdf file via Canvas.

1. What are the roles of ‘data’ and ‘reproducibility’ in the scientific method?

YOUR ANSWER HERE

2. What are the federal requirements on sharing data? To what extent do you think these rules are enforced?

YOUR ANSWER HERE

3. Advanced statistical methods often require sophisticated computational implementations. Should statistical researchers be expected to share their computer code on request?

YOUR ANSWER HERE

4. What is the difference between data and a statistical model for the data? For example, comment on the assertion “Let y_1, \dots, y_n be independent identically distributed data.”

YOUR ANSWER HERE

The remaining questions consider the following hypothetical case study:

Ben is a Statistics PhD student who has written computer code for a simulation study to test a new statistical theory and methodology which he is developing. He plans to put the results in his thesis and to publish them in a journal paper. The results of the simulations are usually consistent with his theoretical analysis. However, sometimes the code crashes, particularly when investigating more extreme values of the parameter space. Ben has checked and rechecked the code very carefully, and cannot find any error. He decides that there must be some weird numerical effect, perhaps to do with occasional extremely large or small numbers. Ben decides to report the results only in the region of the parameter space where the code never crashed.

5. Is Ben’s course of action a reasonable balance between the necessity to make progress on his thesis and his desire to report correct results?

YOUR ANSWER HERE

6. What are the ‘data’ in this example? What is ‘reproducibility’ in this context?

YOUR ANSWER HERE

7. Ben asks your opinion on how to proceed. What is your advice?

YOUR ANSWER HERE