Statistical Inference for Estimation in Data Science

DTSA 5002 offered on Coursera

by the University of Colorado, Boulder

Instructor: J.N. Corcoran

Welcome to Statistical Inference for Estimation in Data Science! This is course 2 of a 3 part specialization in Data Science Foundations. It is also part of a pathway into a Master of Science in Data Science (MS-DS) degree offered by the University of Colorado at Boulder on the Coursera platform. You can take take a non-credit version of this course to sharpen your data science skills and you can upgrade at any time to get help from course learning assistants and to receive academic credit at CU Boulder. We truly care about providing an exceptional learning experience for all. That said, please note that we are not able to actively monitor forums and discussion boards, which are designed for peer support, for the non-credit version of this course. To report problems, including suspected typos or issues with autograders, there is a forum entitled "Questions and Urgent Help" that is monitored by course assistants.

Thank you for choosing to learn with us. Let's do this!

What is Inference?

Statistical inference is the process of learning about characteristics of a population based on what is observed in a relatively small sample from that population. For example, suppose that we want to know about the true average weight of all adult Sockeye salmon in a particular region in the Pacific Northwest of the United States. We can take a "random sample" of these salmon and report the average weight from the sample. Assuming that we have somehow truly randomly sampled from the entire population do you think this sample average will be close to the true population average? There are many factors that go into answering this question including the size of our sample and how variable weights are in the overall population. While estimating a true population average using an average from a sample makes sense intuitively, there are many other types of "parameters" that control population characteristics that are not so easily interpreted and do not have obvious sample counterparts. In this course we will learn about what makes an estimator "good" and how to derive them in non-obvious scenarios!