Homework 6 - Computing Point and Interval Statistics

June 2019

A key part of statistical analysis is computing point and *interval* estimates of the mean. Computing the point estimate of the mean is the standard concept of computing $E[Sn]/n$ as an estimate. The concept of an interval is equally important because point estimates are assumed to be a stochastic estimate of a true mean μ.

In an interval estimate, we define upper and lower bounds $c_l \leq E[X] \leq c_u$ such that we estimate the probability of the true mean being within that interval as having some probability, related a *significance level*. There are several methods to provide interval estimates for the mean based on our assumptions of the underlying data and how much data we have. This assignment uses a large data set of traffic accidents in England to motivate the question of those intervals. Because we have a lot of data, we can use the Central Limit Theorem conclusion that sums of many random variables will tend to the standard normal distribution. If the sample size is small, we use the *Student-T* distribution rather than the standard normal. And, lastly, we can use a *bootstrap confidence interval* if we believe that the sample mean is skewed. This homework will use all of these methods. To complete the homework, you'll need to continue through the material on confidence estimation, bootstrap and the like, but you can start reviewing the data and getting started on the point estimation tasks needed. The homework includes a ZIP file containing HW6_CI.ipynb as a python notebook. It also contains a (rather large) data file of traffic accidents in England from 2005-2007 that we'll use in the first problem.