Coursera Stat Notes – Chapter 4 Foundations for Inference

*Introduction*

1. Inference’s main concern is the quality of the pramater estimates in statistics
   1. such as, how close is the sample mean to the true population mean *µ*?

*Variability in estimates*

1. Population VS Sample Mean
   1. Population mean *µ* can be approximated by taking the mean of the sample and inferring the population mean. The sample must be selected randomly for this approach to be acceptable.
   2. Simplest way to find the population mean: just taking the mean of the whole population, if this is possible.
   3. **Point Estimate**: the best guess of the true variable.
      1. taking several estimates from the population and comparing them to each other will show **sampling variation**. If they are good point estimates, the sampling variation will typically converge.
      2. Point estimates of **population parameters** can be approximated too. For instance, the median and standard deviation can be estimated.
2. Point Estimates are not exact
   1. **Running Mean:** the sequence of means where each mean uses one more observation in its calculation than the mean directly before it in the sequence.
      1. RunningMeans tend to converge on the true mean as more data becomes available.
3. Standard Error of Mean
   1. **Sampling Distribution**: taking repeated samples of a population and calculating the same variable.
      1. Formal Definition: The Sampling Distribution represents the distribution of the point estimates based on a fixed size from a certain population. It is useful to think of a particular point estimate as being drawn from such a distribution. Understanding the concept of a sampling distribution is central to understanding statistical inference.