Cannlytics

Cannabis Data Science

Saturday Morning Statistics #11

February 5th, 2022

Sample Statistics

Given datasets, X, Y, or Z, we can quickly create a sample, n.

Question of the Day

- Can we use a **sample of sale items** to estimate statistics for the **population of sale items**?

Methods of Sampling

- **Complete Sampling** – A specific sample is selected from the dataset(s) by specific criterion.
- **Unbiased (Representative) Sampling** – A sampling process that does not depend on the underlying data.
- **Non-random Sampling** – A sampling process that depends on the underlying data.

Non-Random Sampling

- **Convenience sampling** – Based on ease.
- **Judgment sample** – Based on researcher's opinions.
- **Quota sampling** – a non-probabilistic version of stratified sampling.
- **Snowball sampling** – sampling through referrals.
- **Quasi-Monte Carlo method** – a numerical process for sampling.

Representative Sampling

- **Simple random sample** – a subset of individuals (k) are chosen randomly, all with the same probability $p \sim U(0, 1)$.
- **Systematic sampling** – ordered sampling; start by selecting an element from the a population, N, at random and then every k elements in the population, where k is a function of the desired sample size, n:

$$k = \frac{N}{n}$$

- **Stratified sampling** – a method of sampling from a population which can be partitioned into subpopulations.
- **Cluster sampling** – a sampling process used when mutually homogeneous yet internally heterogeneous groupings are evident in a statistical population.

Thank you for coming.

Lesson of the Day

- Sampling large datasets makes estimating crude statistics quick work, which can be quite helpful.