# Day 18 of #100daysofmathandstats: Data sampling Concepts(Contd...)

By Harsh Kathiriya

# Outline

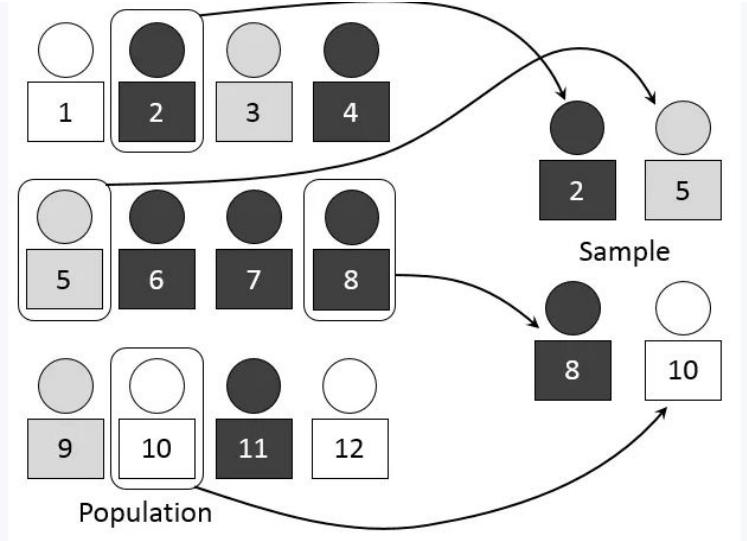
- Sample statistic
- Data distribution
- Sampling distribution
- Central limit theorem
- Standard error

# Sample statitic

- A metric calculated for a sample of data drawn from a larger population.
- A sample is ideally an accurate representation of the entire population, and a sample statistic would ideally be an accurate summary of the entire population.
- Thus, a sample is taken to represent the population and present the ability to manage and work with the data. Then, a sample statistic is taken to run an analysis or conduct research.

# Usage of sample statistic

- The main uses for sample statistics are for quantitative research and analysis. Sample statistics are often used in regression models to predict variables. However, a sample statistic is a very broad term.



# Data distribution vs Sampling distribution

- Data distribution:
  - The frequency distribution of individual values in a data set.
  - distribution of the individual data points
- Sampling distribution:
  - The frequency distribution of a sample statistic over many samples or resamples.
  - distribution of a sample statistic

# Central limit theorem

- The tendency of the sampling distribution to take on a normal shape as sample size rises.
- The central limit theorem in statistics states that, given a sufficiently large sample size, the sampling distribution of the mean for a variable will approximate a normal distribution regardless of that variable's distribution in the population.
- Three types of central limit theorem
  a. Central Limit Theorem with a Normal Population
  b. Central Limit Theorem with a Dichotomous Outcome
  c. Central Limit Theorem with a Skewed Distribution

# Standard error

- The variability (standard deviation) of a sample statistic over many samples (not to be confused with standard deviation, which by itself, refers to variability of individual data values).
- Examples of standard error:
  a. Sampling from populations with percent-in-favor close to 50% have wider sampling distributions than populations with percentages closer to 0% or 100%.
  b. Larger sample sizes have narrower sampling distributions.

- Want to Know more?
  a. https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/what-is-the-standard-error-of-a-sample/

# Thank you

**Github Link:** https://github.com/harsh9898/100daysofstatandmath

Don’t forget to post your queries or feedbacks on the post.

Share or like for the benefit of others.

Source for some of the content: https://www.openintro.org/book/os/