SAMPLING TECHNIQUES

A Comprehensive Overview

SHEFALI NAYAK

Preface

The book is intended to give you a comprehensive understanding of the Sampling Techniques. At the end of this book, you should be able to define the data collection process and choose the sampling technique that works best for your data.

The book is designed in a student-friendly manner and explains the core concepts and terminologies using real-world examples. Remember the time we took our tests and jotted down a list of important points along with some pictorial representations to retain a concept, term or jargon... well we just did that with this book! This is your go-to book for understanding difficult concepts in Sampling Techniques in a lucid language.

Complimentary chapters included – Types of data collection and Probability for a holistic view of the Sampling Techniques.

Keep on Learning! 💛

Contents, Disclaimer and Rights

The book cover has been designed by the author. The examples have been created using dummy data and the graphical and image representation in the book has been illustrated by the author.

This book is not a piece of research and is intended to explain the established concepts of statistics in a simple manner.

The design and contents of the book may not be copied, published or circulated without the consent of the author.

Acknowledgments

I am extremely thankful and grateful to my parents, sibling and spouse for their constant encouragement and indispensable advice for the improvement of the content to shape it out in its present form.

What are we learning today?

1. Importance of Sampling 8
2. Data Collection 11
   - 2.1. Primary data collection 12
     - 2.1.1. Interviews 12
     - 2.1.2. Checklist 13
     - 2.1.3. Monitor or Observations 14
     - 2.1.4. Survey or Questionnaire 15
     - 2.1.5. Case Study 15
   - 2.2. Secondary data collection 16
   - 2.3. Quiz 1: Know your data 17
3. Population and Sample: Basic Concepts 19
   - 3.1. Statistical Population 19
   - 3.2. Sampling Frame 19
   - 3.3. Statistical Sample 20
   - 3.4. Sampling Fraction 20
   - 3.5. Sampling Bias 20

4. Why infer from a sample? 11
5. Characteristics of good sample design 22
6. Probability 24
6.1. Complementary Events 25
6.2. Certain and Impossible Events 26
6.3. Dependent Events 27
6.4. Independent Events 29
7. Random Sample 31
8. Types of Sampling Techniques 33
9. Random Sampling Techniques 12
9.1. Simple Random Sampling 12
9.2. Systematic Sampling 37
9.3. Stratified Sampling 40
9.4. Cluster Sampling 43
9.4.1. Single-Stage Clustering 45
9.4.2. Two-Stage Clustering 46
9.4.3. Multi-Stage Clustering 47

9.5. Cluster Vs Stratified Sampling 48
10. Random Sampling Techniques - Advantages and Disadvantages 50
11. Non-Random Sampling Techniques 51
11.1. Purposive Sampling 51
11.2. Snowball Sampling 52
11.3. Quota Sampling 53
11.4. Convenience Sampling 54
12. Random Vs Non-Random Sampling 55
13. Potential sources of error in Sampling 56
13.1. Sampling Error 57
13.2. Non-Sampling Error 58
14. Reference 60
14.1. Reference 1: Images, Illustrations 60
14.2. Reference 2: Tables 61
14.3. Reference 3: Quiz 61
About the Author and Editor 62

1. Importance of Sampling

The data selected for the research and analysis plays a crucial role and varied samples can result in different conclusions for the same experiment conducted.

In a research and analysis project majority of the time is spent in getting the data right before testing the hypotheses, seeking to infer about the population, processing the data for descriptive statistics, predictive models or prescriptive measures. The conclusion derived from the data is as good as the data itself.

Example: Let us compare two uses of sampling and the various types of data collection.

i. Understanding reviews on two new movie releases and providing your recommendation to a friend on which movie to watch.
ii. Drawing a conclusion on which of the two medicines are effective in treating a disease.

A sampling technique that is less tedious and cost-efficient may be considered in the first analysis as the consequences of a wrong review may not be as drastic as compared to a wrong conclusion from the second analysis.

Also, relying on newspaper reviews, casting and the director's previous film knowledge rather than watching the movie yourself before recommending one of the two movies to your friend may not sound as bad in the first example.

However, in the second example, one cannot depend on hearsay with regards to a person’s medical state post taking the medication and the researcher would require to look at medical reports to come to a definite conclusion.

The data collection process and sampling techniques rely on the objective of the project and the accuracy required in the conclusion.

While gathering data for the analysis, the researcher requires to address points such as-

1. What constitutes the elements of the sample?
2. How is the sample being collected?
3. What is the cost associated with sampling?
4. The timeline of the project?
5. The consequences of making the wrong decision.

4. Why infer from a sample?

Counting the actual number of people for a large population can be very *time-consuming.* It will also involve a lot of resources to fill in surveys and will be *very expensive*. The process will further *require trained resources* who will manage to get the data without missing values and completely filled out questionnaire. Sometimes, in addition, it is a *not feasible* approach to collect population data. Examples:

i. *Time Consuming, Expensive*: In understanding the shoe size and design most commonly sold, the researcher may have to rely on the data of shoe sales in the past than ask everyone in the region to participate in a survey.

ii. *Not Feasible:* Estimating the duration of a Light-Bulb - One cannot use all the light-bulbs produced to make the claim of the average number of years the light-bulb will work.

9. Random Sampling Techniques

Let us now look at the various random sampling techniques available to us to arrive at a statistical sample that is representative of the universe.

9.1. Simple Random Sampling

- This method involves the task of selecting the sampling units randomly out of the population. For a population size N; the Probability of each item getting selected into the statistical sample is 1/N.
- Examples: Lottery method, the random number generated from Uniform distribution.

- Simple Random Sampling is bifurcated into two types

i. SRSWR: Simple Random Sample With Replacement

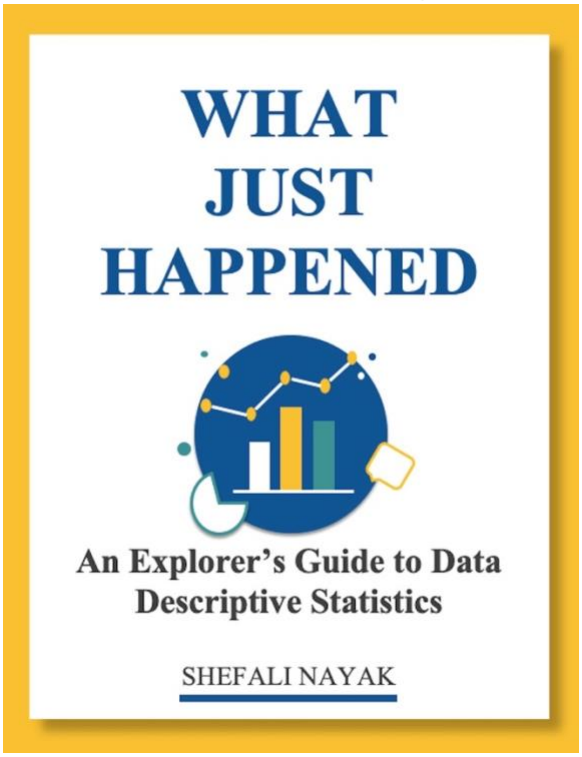
ii. SRSWOR: Simple Random Sample Without Replacement

- As the names of the terms suggest, the researcher may allow the observation or individual that forms the population to be selected again or may deliberately avoid selecting the same element more than once.
- Simple Random Sample Without Replacement is dependent on the remaining items in the population and does not appear to be truly random; but since the Joint Probability distribution of the observations or individuals in the Sampling Frame does not change, the results still hold and the sampling is random in nature. Also, for a small sample from a large population, SRSWOR is approximately the same as sampling with replacement, since the probability of

choosing the same observation or individual more than once is low.

- The number of possible samples of sample size n from a population size N:
  i. The number of SRSWR samples = N^n
  ii. The number of SRSWR samples = $\frac{N!}{n!(N-n)!}$
- The major advantage of this sampling technique is that it requires very little knowledge of the population unlike Stratified Random Sampling; where the researcher needs to have prior knowledge of the population.
- It however may not be able to capture specific unique features in the population and may not be efficient (in case of rework to arrive at a true representative sample).

Other books by the author



Descriptive Statistics

The book is intended to give you a comprehensive understanding of Descriptive Statistics and help you figure out "What just Happened?" through your data.

Complimentary chapters – Types of data, Normal distribution and data visualization for a holistic view of the Descriptive Statistics.

Learn more about Descriptive Statistics in lucid language.

Big Data Analytics

The book is designed to walk you through the various stages of a data analytics project, concepts and possible avenues when dealing with huge and overwhelming amounts of structured and unstructured data. It is crisp and concise roadmap on a Big Data project. Complimentary chapters included – Description on the various AI/ML Models, data visualizations and storyboarding techniques to empower you make relevant decisions related to the analytics approach.

Follow the author

Share your review and feedback with the author about this book on social media. Let us connect, learn and collaborate on our common passion for DATA!

Instagram: keep_on_learning_