MIS770 Foundation Skills in Business Analysis

DEPARTMENT OF INFORMATION SYSTEMS AND BUSINESS ANALYTICS
DEAKIN BUSINESS SCHOOL
FACULTY OF BUSINESS AND LAW

Tutorial Topic 1

Defining and Collecting Data

Introduction

When we undertake Data Analysis, our aim is draw conclusions from **Sample Data** that is representative of our **Population Data**. Therefore, in order for us to **Present** and **Interpret** this sample data effectively, we must gather our data in a valid manner. If the sample data isn't representative, we cannot be confident that the conclusions we drawn from the data are valid. Hence, the saying, "garbage in, garbage out". If my sample data is flawed (or garbage), then my conclusions must also be flawed (or garbage).

Therefore, the aims of this tutorial are to:

- distinguish between a population and a sample
- understand the importance of sampling
- recognise and avoid errors in sampling
- understand both probabilistic and non-probabilistic methods of sampling

Textbook Questions

1.5 The following information is collected from students as they leave the campus bookshop during the first week of classes:

a. Amount of time spent shopping in the bookshop

b. Number of textbooks purchased

c. Name of degree

d. Gender

e. Level of service quality

Classify each of these variables as Categorical or Numerical. If the variable is Numerical, determine whether the variable is Discrete or Continuous. In addition, determine the Level of Measurement.

1.23 The town planning department of a Sydney council with a population of N = 40,000 registered voters is asked by the mayor to conduct a survey to measure community attitudes to urban consolidation. The table following contains a breakdown of the 40,000 registered voters by gender and ward of residence.

	Ward of Residence				
Gender	North	South	East	West	Total
Female	7,000	5,200	5,000	4,800	**22,000**
Male	5,600	4,600	4,000	3,800	**18,000**
Total	**12,600**	**9,800**	**9,000**	**8,600**	**40,000**

The planning department intends to take a probability sample of n = 2,000 voters and project the results from the sample to the entire population of voters.

a. If the frame available from the council files is an alphabetical listing of the names of all *N* = 40,000 registered voters, what type of sample could you take? Discuss.

b. What is the advantage of selecting a simple random sample in (a)?

c. What is the advantage of selecting a systematic sample in (a)?

d. If the frame available from the council's files is a listing of the names and addresses of all $N = 40{,}000$ registered voters, compiled from eight separate alphabetical lists based on the gender and address breakdowns shown in the ward-of-residence table, what type of sample should you take? Discuss.

e. At present East Ward has many high-rise apartments, West Ward and South Ward have single dwellings only and North Ward has a mixture of low- and medium-density housing. What would be the danger in randomly choosing 40 street names and systematically sampling 50 of the residents of those streets?

1.29 Reality TV shows have incorporated surveys of audience opinion into their formats. In Australia several shows have allowed the audience to vote on whether contestants should remain on the show or be excluded. Consider a show where voting is by SMS, premium rate phone call, Facebook or another online site, and viewers are limited to 10 votes using each method. Compare this type of survey with a random poll of viewers without replacement conducted by phone for the TV show.

a. How might the results differ?

b. What are the costs and benefits for the owners of the show for each voting method?

1.49 A manufacturer of flavoured milk is planning to survey households in Tasmania to determine the purchasing habits of consumers. Among the questions to be included are those that relate to:

1. where flavoured milk is primarily purchased
2. what flavour of milk is purchased most often
3. how many people living in the household drink flavoured milk
4. the total number of millilitres of flavoured milk drunk in the past week by members of the household

a. Describe the Population

b. For each of the four items listed, indicate whether the variable is Categorical or Numerical. If Numerical, is it Discrete or Continuous?

c. Develop five Categorical questions for the survey

d. Develop five Numerical questions for the survey

TEXTBOOK REFERENCE:

Basic Business Statistics: Concepts and Applications. *Berenson, M.L. Levine, D.M. Szabat, K.A. O'Brien, M. Jayne, N. Watson, J.* 5th edition. 2019. Pearson Australia Group Pty Ltd. ISBN 9781488617249. Chapter 1, sections 1.1 to 1.6