MIS770 Foundation Skills in Business Analysis

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

Tutorial Topic 1 (Solutions)

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/Answers/Readings

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
Numerical, continuous, ratio scale

b. Number of textbooks purchased
Numerical, discrete, ratio scale

c. Name of degree
Categorical, nominal scale

d. Gender
Categorical, nominal scale

e. Level of service quality
Ordinal scale

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.

Reading: Berenson Ch. 1, Section 1.2

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.

f. 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.
As a complete list of voters exists, a simple random sample of 2,000 voters could be taken. If attitudes to urban consolidation randomly fluctuate across the alphabetical listing of voters, a systematic 1-in-20 sample could also be taken from the population frame. If attitudes may differ by gender and by ward of residence, a stratified sample using eight strata could be selected. If attitudes to urban consolidation are thought to fluctuate as much within clusters as between them, a cluster sample could be taken.

g. What is the advantage of selecting a simple random sample in (a)?
A simple random sample is one of the simplest to select. The population frame is the council's list of 40,000 voters' names.

h. What is the advantage of selecting a systematic sample in (a)?
A systematic sample is easier to select by hand from the council's records than a simple random sample, since an initial person at random is selected and then every 20th person thereafter would be sampled. The systematic sample would have the additional benefit that the alphabetical distribution of sampled voters' names would be more comparable to the alphabetic distribution of voters' names in the population. However, if certain ethnic groups were more likely to be concentrated alphabetically and they have particular attitudes to urban consolidation due to their cultural background, then this method may introduce bias.

i. 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.
If lists by gender and ward of residence are readily available, a stratified sample should be taken. Since attitudes to urban consolidation may indeed differ by gender and area of residence, the use of a stratified sampling design will ensure all strata are represented in the sample. It will also generate a more representative sample and produce estimates of the population parameter that have greater precision.

j. 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?
People who already live in high-rise apartments are likely to have different views about urban consolidation from those living in low-density areas. Systematic sampling by street is likely to lead to selection bias if a high proportion of streets in one ward is chosen.

Reading: Berenson Ch. 1, Section 1.4

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?
The SMS/phone/online voting system is a non-probability sampling method. It is open to bias as some viewers may choose not to vote, while supporters of a particular contestant can vote a large number of times and have the potential to influence the result. A random poll of viewers without replacement conducted by phone should produce a less-biased result, although it is possible some viewers without phones or with unlisted numbers might be excluded.

b. What are the costs and benefits for the owners of the show for each voting method?
A phone-in-voting system has cost benefits for the show's owners as they can earn revenue from premium rate calls. However, SMS polling may incur a set up fee. Social media polling can show results such as the number of 'likes' quickly and can be useful for advertising purposes but has the disadvantage that later voters could be influenced by earlier votes. There may also be better ratings for the show if the audience feels more involved. Phoning viewers using research staff would be more costly but should give more reliable data.

Reading: Berenson Ch. 1, Section 1.5

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
The population of interest was all households in Tasmania.

b. For each of the four items listed, indicate whether the variable is Categorical or Numerical. If Numerical, is it Discrete or Continuous?
Categorical variables (where and what flavour of milk is purchased) and numerical variables (number of people living in the household and number of millilitres drunk).

c. Develop five Categorical questions for the survey
Answers could vary.

d. Develop five Numerical questions for the survey
Answers could vary.

Reading: Berenson Ch. 1, Sections 1.1 to 1.6

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