Revision 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

1.4   For each of the following variables, determine whether the variable is categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the level of measurement.

   a.   Number of mobile phones per household

   b.   Length (in minutes) of the longest mobile call made per month

   c.   Whether all mobile phones in the household use the same telecommunications provider

   d.   Whether there is a landline telephone in the household

1.24   Suppose that 5,000 sales invoices are separated into four strata. Stratum 1 contains 50 electrical invoices, stratum 2 contains 500 paint invoices, stratum 3 contains 1,000 plumbing supplies invoices and stratum 4 contains 3,450 hardware invoices. A sample of 500 sales invoices is needed.

   a.   What type of sampling method should you use? Why?

   b.   Explain how you would carry out the sampling according to the method stated in (a).

   c.   Why is the sampling in (a) not simple random sampling?

1.30   The online restaurant search site Dimmi <www.dimmi.com.au> encourages diners to rate restaurants they have been to by giving them reward points which can be accumulated until a meal discount is available. A restaurant at The Rocks in Sydney has been rated as follows: Recommended 8.7; Food 8.5; Service 8.7; Value for money 7.8; Atmosphere 8.4. What differences could arise from this type of survey compared with ratings derived from a random sample of diners?

1.50   A new bus network is proposed for a north-eastern Sydney region. A survey is sent out to residents asking questions which relate to:

   1.   the resident's age
   2.   frequency of bus use
   3.   usual ticket type purchased
   4.   main purpose of using the bus

a. Describe the population.

b. Indicate whether each of the questions above is categorical or numerical.

c. Develop two more numerical questions and state whether the variables are discrete or continuous.

d. Develop two more categorical questions.

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