Revision Tutorial Topic 2

Organising and Visualising Data

Introduction

Our underlying goal here is to faithfully represent our raw Sample Data visually, such that the **visual** story we're telling is accurate and not misleading.

So, specifically we will see how to construct various types of tables, histograms, frequency distributions, frequency polygons and Ogives. Then we will further illustrate our data using various charts.

Therefore, the aims of this tutorial are to:

- describe the distribution of a single categorical variable using tables and charts
- describe the distribution of a single numerical variable using tables and graphs
- describe the relationship between two categorical variables using contingency tables
- describe the relationship between two numerical variables using scatter diagrams and time-series plots
- correctly present data in graphs

Textbook Questions

2.4 The following table gives the top 10 websites ranked by estimated number of unique monthly visitors in March 2017.

Website	Unique monthly visitors (millions)
Google	1,600
Facebook	1,100
YouTube	1,100
Yahoo!	750
Amazon	500
Wikipedia	475
Twitter	290
Bing	285
eBay	285
MSN	280

Source: Data obtained from eBusMBA Guide, Top 15 Most Popular Websites March 2017, at <www.ebizmba.com/articles/most-popular-websites> accessed 13 March 2017

a. Construct bar and pie charts.

b. Which graphical method do you think best portrays these data?

c. What conclusions can you reach concerning the number of unique visitors?

2.14 Low-fat foods are not necessarily low calorie, as many are high in sugar. The following data give calories per 250 ml cup of a random sample of brands of fresh cow's milk for sale in Australia. [**Dataset: FRESH_MILK.XLS**]

Full cream milk
155 188 160 155 160 163 170 185 135 160 165 160 163
Low- or reduced-fat milk
120 133 133 125 118 113 140 110 128 115
No-fat or skim milk
133 90 90 98 88 85 115 108 88 90 90 98
Source: Data obtained from Calorie King Australia <www.calorieking.com.au> accessed 22 December 2013

For each category of milk:

a. Display the data in ordered arrays.

b. Construct stem-and-leaf displays for the data.

c. Which arrangement provides more information? Discuss.

d. Compare the items in terms of calories. What conclusions can you make?

2.18 To investigate the variation in fuel prices in New South Wales on a day in March 2017, a random sample of 45 petrol stations, each in a different location, was selected. The price per litre of both unleaded petrol and diesel is recorded in **[Dataset: FUEL_2017.XLSX]**

Using the New South Wales data:

a. Construct frequency, percentage and cumulative distributions for the price of petrol and diesel.

b. As separate graphs, plot frequency histograms for the price of petrol and diesel.

c. On the same set of axes plot percentage polygons for the price of petrol and diesel.

d. On the same set of axes plot cumulative percentage polygons for the price of petrol and diesel.

e. What can you conclude about the variation in the fuel prices in New South Wales at the time the data were collected?

2.22 The Living in Australia Study gives information on the study mode (full or part time) of students studying for a post-school qualification, as well as their employment status.

Percentage of students enrolled in post-school education			
Employment status	Studying full-time	Studying part-time	All students
Employed full-time	6.4	37.7	44.1
Employed part-time	18.1	12.2	30.3
Not employed	17.4	8.2	25.6
All students	41.9	58.1	100.0
Source: Data obtained from the Household, Income and Labour Dynamics in Australia (HILDA) Survey, 2001–2005 (also known as the Living in Australia Study), The University of Melbourne 1994–2011			

a. Construct cross-classification tables based on row percentages and column percentages.

b. Construct a side-by-side bar chart for employment status and study mode.

c. What conclusions do you draw from these analyses?

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 2, sections 2 to 2.7