MIS770 Foundation Skills in Business Analysis

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

Tutorial Topic 2 (Solutions)

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/Answers/Readings

2.15 The values for a data set vary from 11.6 to 97.8

a. If these values are grouped into nine classes, indicate appropriate class boundaries
The class boundaries of the 9 classes can be '10 to less than 20', '20 to less than 30', '30 to less than 40', '40 to less than 50', '50 to less than 60', '60 to less than 70', '70 to less than 80', '80 to less than 90' and '90 to less than 100'

b. What class width did you choose?
The class-interval width is $= \frac{97.8 - 11.6}{9} = 9.58 \cong 10$

c. What are the corresponding class mid-points?
The nine class mid-points are: 15, 25, 35, 45, 55, 65, 75, 85 and 95

Reading: Berenson Ch. 2, Section 2.3

2.17 The following data represent the electricity cost (in dollars) during the month of July for a random sample of 50 two-bedroom apartments in a New Zealand city. [**Dataset: ELECTRICITY.XLSX**]

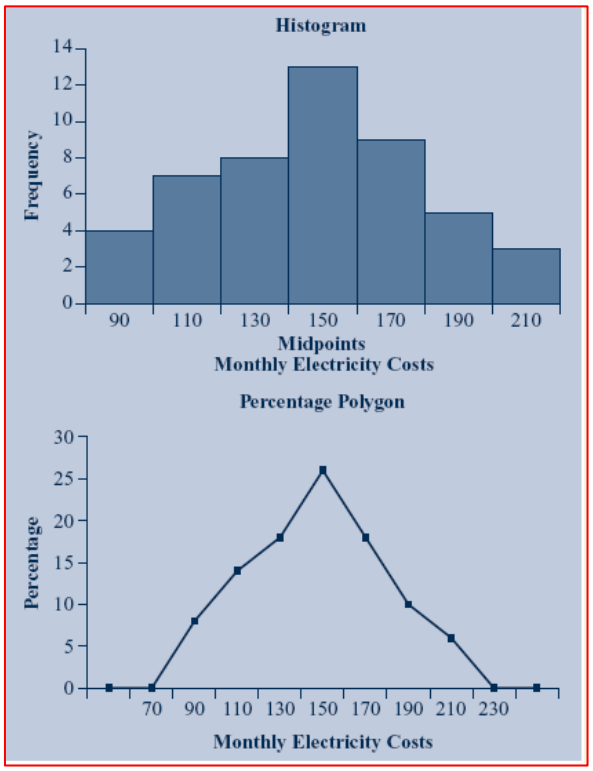
Electricity charge ($)									
96	171	202	178	147	102	153	197	127	82
157	185	90	116	172	111	148	213	130	165
141	149	206	175	123	128	144	168	109	167
95	163	150	154	130	143	187	166	139	149
108	119	183	151	114	135	191	137	129	158

a. Construct a frequency distribution and a percentage distribution with upper class boundaries of <$100, <$120, and so on. [Do manually or use Excel – Instructions in Textbook pg. 80 - Summarising Numerical Data]

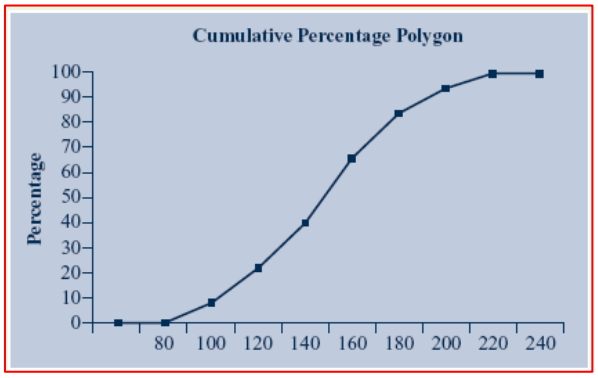
Note: For Excel, you need to manually setup Bins. Resultant Excel output - needs refining

Electricity costs	Frequency	Percentage	Cumulative percentage
$80 to < $100	4	8	8
$100 to < $120	7	14	22
$120 to < $140	9	18	40
$140 to < $160	13	26	66
$160 to < $180	9	18	84
$180 to < $200	5	10	94
$200 to < $220	3	6	100

b. Plot the corresponding histogram and percentage polygon.



c. Construct the corresponding cumulative percentage distribution and plot the corresponding ogive (cumulative percentage polygon).



d. Around what amount does the monthly electricity cost seem to be concentrated?

Monthly electricity costs are concentrated between $140 and $160 a month, with more than 25% falling in that interval.

Reading: Berenson Ch. 2, Section 2.3

2.19 The ordered arrays in the table below give the life (in hours of usage) of samples of forty 15-watt CFL (Compact Fluorescent Lamp) energy-saving light bulbs produced by two manufacturers, A and B. **[Dataset: BULBS.XLSX]**

Manufacturer A									
5544	5814	6190	6307	6342	6423	6429	6485	6612	6667
6832	6868	6879	6930	6941	7007	7037	7043	7059	7136
7497	7645	7654	7773	7816	7838	7924	7999	8038	8067
8091	8119	8392	8416	8416	8514	8532	8542	8544	8731

Manufacturer B									
6701	6837	6961	7118	7133	7142	7156	7344	7493	7569
7607	7612	7651	7721	7754	7767	7806	7839	7888	7983
8298	8344	8535	8666	8792	8800	8856	8861	8993	9001
9036	9096	9262	9385	9460	9471	9521	9540	9693	9744

a. Construct a frequency distribution and percentage distribution for each manufacturer.

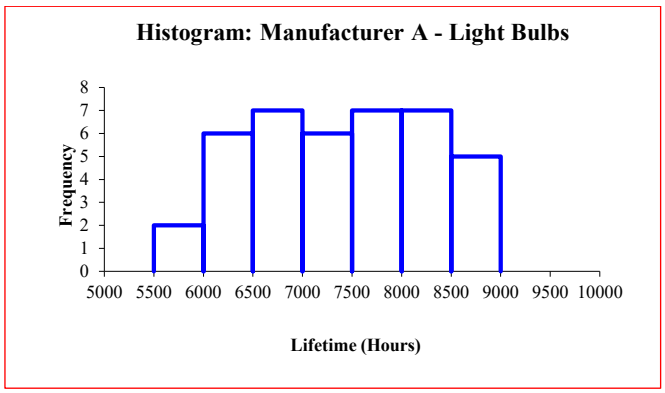
Manufacturer A

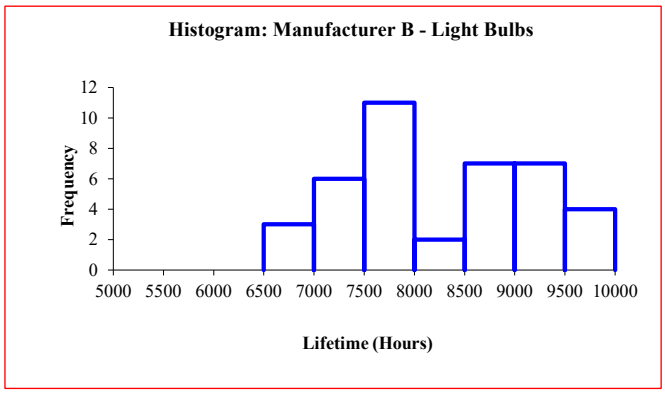
Life in hours	Frequency	Percentage %	Cumulative %
5500 to < 6000	2	5	5.00%
6000 to < 6500	6	15	20.00%
6500 to < 7000	7	17.5	37.50%
7000 to < 7500	6	15	52.50%
7500 to < 8000	7	17.5	70.00%
8000 to < 8500	7	17.5	87.50%
8500 to< 9000	5	12.5	100.00%
9000 to < 9500	0	0	100.00%
9500 to < 10000	0	0	100.00%
Total	**40**	**100**	

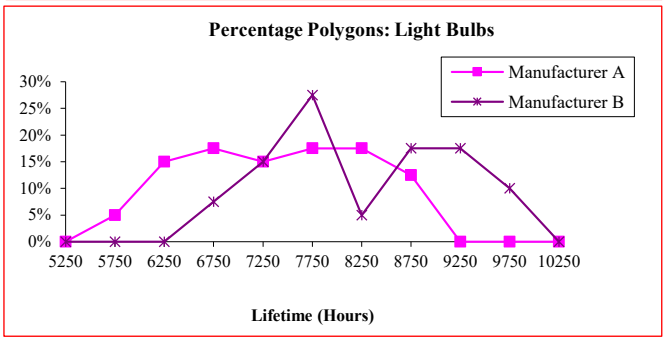
Manufacturer B

Life in hours	Frequency	Percentage	Cumulative %
5500 to < 6000	0	0	0.00%
6000 to < 6500	0	0	0.00%
6500 to < 7000	3	7.5	7.50%
7000 to < 7500	6	15	22.50%
7500 to < 8000	11	27.5	50.00%
8000 to < 8500	2	5	55.00%
8500 to< 9000	7	17.5	72.50%
9000 to < 9500	7	17.5	90.00%
9500 to < 10000	4	10	100.00%
Total	**40**	**100**	

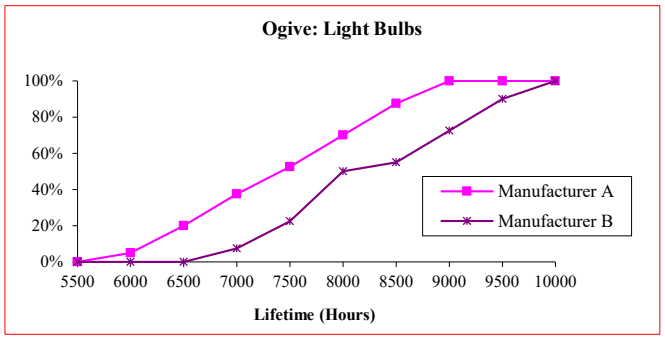
b. Plot the corresponding frequency histograms on separate graphs and the percentage polygons on the same graph.







c. Form the cumulative percentage distributions and plot the ogives on one graph.



d. Which manufacturer has bulbs with a longer life – Manufacturer A or Manufacturer B? Explain.

Manufacturer B produces bulbs with longer lives than manufacturer A's bulbs. The cumulative percentage for manufacturer B shows 50% of its bulbs lasted 8000 hours or less contrasted with 70% of manufacturer A's bulbs. None of manufacturer A's bulbs lasted more than 9000 hours, but 27.5% of manufacturer B's bulbs did. At the same time, 20% of manufacturer A's bulbs lasted less than 6500 hours, while all of manufacturer B's bulbs lasted at least 6500 hours.

Reading: Berenson Ch. 2, Section 2.3

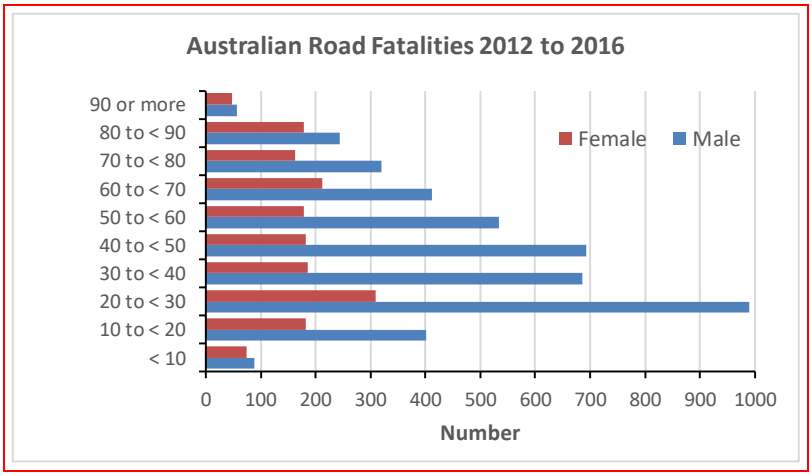
2.23 The following table classifies road fatalities in Australia from 2012 to 2016 (inclusive) by age and gender. **[Dataset: ROAD_FATALITIES_2012_2016.XLSX]**

	Gender			
Age	Male	Female	Unknown	Total
< 10	89	74	3	166
10 to < 20	402	182	0	584
20 to < 30	990	310	0	1,300
30 to < 40	686	185	0	871
40 to < 50	693	182	0	875
50 to < 60	534	179	0	713
60 to < 70	412	212	0	624
70 to < 80	319	162	0	481
80 to < 90	243	178	0	421
90 or more	56	47	0	103
Unknown	3	1	0	4
Total	**4,427**	**1,712**	**3**	**6,142**

Data obtained from the Australian Road Deaths Database <www.bitre.gov.au/statistics/safety/fatal_road_crash_database.aspx> accessed 18 March 2017

Ignore the unknown categories.

a. Investigate the relationship between age and gender by constructing a side-by-side bar chart to highlight the pattern of male and female road fatalities.



b. Discuss the pattern of male and female road fatalities for 2012 to 2016.

In all age groups there are more male fatalities than female. While female fatalities in most age groups are approximately 200 male fatalities vary from under 100 to approximately 1000. Most male fatalities are in the 20 to 60 age group. Both male and female fatalities peak in the 20 to 30 age group.

Reading: Berenson Ch. 2, Section 2.4

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