**Crown Institute of Higher Education**

BUS104

Week 1 - workshop discussion/practice questions

Introduction

A recently released report suggests that mobile phone addiction is costing Australians $560 million a year in bill “blow outs”. The following data represents the “bill shocks” (correct to the nearest dollar) over a 12 month period of a random sample of 45 mobile phone users whose annual charges exceeded what they expected

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 482 | 608 | 428 | 619 | 328 |
| **695** | 368 | 594 | 299 | 547 |
| 137 | 311 | 222 | 447 | 88 |
| 533 | 692 | 101 | 358 | 239 |
| 58 | 114 | 561 | 265 | 166 |
| 691 | 99 | 264 | **19** | 468 |
| 526 | 347 | 98 | 193 | 544 |
| 253 | 333 | 276 | 406 | 659 |
| 575 | 23 | 287 | 218 | 661 |

1. Describe as precisely as possible the variable of interest represented by the above data. What units, if any, are applicable to this variable?

2. Are the data categorical or quantitative? If categorical, are the data nominal or ordinal? If quantitative, are the data continuous or discrete (**you should distinguish between “theoretically continuous but discrete in practice” and “simply discrete” and briefly explain the reasoning behind your answer**)? Further, if quantitative, are the data measured on a ratio scale or an interval scale? In relation to your answer to the last question (if appropriate) what are the practical implications of the data being measured on either a ratio or an interval scale (whichever you decided)? How can you most easily decide whether quantitative data are measured on a ratio or an interval scale?

3. The range of values covered by the above data is 695 – 19 = 676 ≈ 700 ($), suggesting that of the order of seven class intervals each of size $100 would be sufficient to encompass/cover the entire data set.

Perform a “tally count” of the data in each of the class intervals 0 – (<100), 100 – (<200), ...., 600 – (<700) and from this count construct a frequency (f) distribution/table for the “bill shock” data.

4. Add three more columns to the frequency distribution table and construct a cumulative frequency (cf) distribution, a relative frequency (rf(%)) distribution and a cumulative relative frequency (crf(%)) distribution (in the one table). Round all rf(%) and crf(%) values in the above table to one decimal place.

5. Perform the routine totalling checks of the frequency and relative frequency (%) columns of the table constructed in Q3 and 4 to obtain some confirmation that your calculations have been performed appropriately in each case. Explain any discrepancies from what you expected in each case in these two checks.

6. Perform a routine check of the last value of the cf and crf(%) columns of the table constructed in Q4 to obtain some confirmation that your calculations have been performed appropriately in each case. Explain any discrepancies from what you expected in each case in these two checks.

7. Use an appropriate distribution/table/column constructed in Q3 and 4 to determine:

(a) The number of mobile phone users in the sample who had “bill shocks” of between $400 and not quite $500.

(b) The number of mobile phone users in the sample who had “bill shocks” of less than $500.

(c) The percentage of mobile phone users in the sample who had “bill shocks” of between $400 and not quite $500.

(d) The proportion of mobile phone users in the sample who had “bill shocks” of less than $500.

8. State the distribution/table/column (f, cf, rf(%), crf(%)) used to obtain your answers to (a) – (d) in Q7. Make sure you check your answers to this question to ensure that you have used the most appropriate distribution/table/column to obtain your answers particularly for (c) and (d).

9. Comment on the tendency of the “bill shock” sample data to cluster/concentrate (or not) about a particular interval.

10. Use the frequency distribution/table constructed in Q3 to manually develop a graphical representation of the “bill shock” data in the form of a frequency histogram. In developing your graph make sure you clearly label both axes, you use a consistent scale on both axes and that the columns of your histogram do not have any gaps between them (unless the gap represents a column of zero frequency).

11. Use the relative frequency (%) distribution/table/column constructed in Q4 to manually develop a graphical representation of the “bill shock” data in the form of a relative frequency (%) histogram. In developing your graph make sure you clearly label both axes, you use a consistent scale on both axes and that the columns of your histogram do not have any gaps between them (unless the gap represents a column of zero relative frequency).

12. Compare the appearances of the two graphs constructed in Q10 and 11. How would you identify any tendency of the “bill shock” sample data to cluster/concentrate (or not) about a particular interval from either of these two graphs?

13. How might you attempt to describe the variability of the sample data from either of the two graphs constructed in Q10 and 11?