Topic: Exploratory Data Analysis (EDA)

Presentation of Univariate Data

Part A: In Tables

School of Mathematics and Applied Statistics

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Where in the Statistical Process

- Ethics
- Nature of the question to be answered
- Context/Expertise
- Design:
  - Experiment vs. observational study
  - Sampling
  - Measurement
- **Description and analysis
  - Reporting**
- Conclusions and decision making

VARIATION

Presentation of Data

The aim is to turn data into meaningful information
covering all major aspects of the data, with precision
AND to communicate that information (paragraphs).

Data can be presented

- as a list   as *raw data*
- in a table   in *tabular form*
- in a graph   *graphically*

Appropriate presentation is important.

Presentation of Data in a Table

A frequency table:

- is a useful way to present categorical or discrete data

- lists all possible values along with number of observations (frequency or count) for each value

- Relative frequency = frequency / total is often included (possibly as %)

- Cumulative Frequency

Frequency Table: Qualitative Data

Categorical data: Travelling to School

	Tally	Frequency	Relative Frequency	
Bike	|||| ||||	9	9/30=0.30	30%
Bus	|||| |||	8	8/30=0.27	27%
Car	|||| ||||	10	10/30= 0.33	33%
Walked	|||	3	3/30 =0.10	10%
Total		30	1.0	

Frequency Table: Quantitative Discrete Data

Discrete data: Number of Siblings for 65 students:

No. of Siblings	Frequency	Relative Frequency	Cumulative Frequency	
0	5	5/65 = 0.077	5	
1	28	28/65 = 0.431	33	
2	17	17/65 = 0.262	50	→ 50/65 = 77%
3	9	9/65 = 0.138	59	
4	3	3/65 = 0.046	62	
5	3	3/65 = 0.046	65 ✓	
Total	65	1.000	✗	

Grouping Quantitative Data

For discrete (with many different values) or continuous data, it is often necessary to group the observations into classes.

The larger the chosen class width, the smaller the number of classes.

Eg. Marks in a mathematics test

Marks in test	Frequency	Relative Frequency	Cumulative Frequency
30 up to 39	1	1/50= 0.02	1
40 up to 49	7	7/50= 0.14	8 ←
50 up to 59	10	10/50= 0.20	18
60 up to 69	14	14/50= 0.28	32
70 up to 79	10	10/50= 0.20	42
80 up to 89	6	6/50= 0.12	48
90 up to 100	2	2/50= 0.04	50
Total	50	1.00	

Note: once the data have been grouped, the raw data are no longer visible.