

## CM1606: Computational Mathematics (Statistics Component)

### Tutorial No 01

(Covers data analysis: R Studio may be used)

- 1) Identify the data type for the following variables:
  - a) Body temperature ( $^{\circ}\text{C}$ )
  - b) Area of a plot of land ( $m$ )
  - c) The population of a country in a given year.
  - d) The blood type of a person ( $O, A, B, \text{ or } AB$ )
  - e) Housing type (*rented, leased, owned*)
  - f) shirt size ( $S, M, L, XL, XXL$ )
  - g) The number of blemishes on the hood of a car coming off a production line
  - h) A survey question that asks the participant to select from **Strongly agree, Agree, Neutral, Disagree, and Strongly disagree.**
  - i) The noise level (*in decibels*) at a concert
  - j) The noise level out of three possible choices: **high, medium, low.**
  - k) A choice of primary color.
  - l) The distance between a cat and a mouse.

- 2) Consider the following (artificial) dataset: (Employee salaries for a small firm)

<u>Position</u>	<u>Salary</u>	<u>Years of experience</u>
Labourer	10,000	01
Receptionist	15,000	01
Management Asst 1	25,000	01
Management Asst 2	30,000	02
Junior executive 1	40,000	02
Junior executive 2	50,000	03
Accountant	100,000	02
CEO	800,000	03

- a) What's the 'average' salary paid in the firm?
- b) What do you think is more reasonable, 'mean' or 'median' to represent the location of the data? (justify your answer)
- c) Calculate the weighted mean of the salary, taking years of experience as weight. Does it make any improvement?

- 3) Consider the data of Question 2. For the salary,
  - a) Calculate the “5-number summary”
  - b) Show the “5-number summary” in a Boxplot (you may divide by 10,000 and use the re-scaled values)
  - c) Calculate the variance, SD, IQR, and median absolute deviation from the median (MAD)
  - d) Briefly interpret your results.
  
- 4) Generate a sample of 100 values from a normal distribution with a mean of 50 and SD of 10.
  - a) Construct a histogram with 10 bins of same size.
  - b) Make a density plot of the data and show it over the histogram
  - c) Find the 25<sup>th</sup>, 40<sup>th</sup>, 60<sup>th</sup>, and 75<sup>th</sup> percentile points in the data.