

# EMTH1019 Linear Algebra and Statistics for Engineers

## Week 1

**Learning outcomes for this session.** You should be able to

1. Present data using various graphical means.
2. Summarise data using descriptive statistics.

### Exercises

1. International Communications Research (ICR) conducted the 2004 National Spring Cleaning Survey for The Soap and Detergent Association. ICR questioned 1000 American male and female heads of household regarding their house cleaning attitudes. The survey has a margin of error of plus or minus 5%.

- (a) What is the population?
- (b) How many people were polled?
- (c) What information was obtained from each person?

#### THOSE HARD TO CLEAN PLACES

Cleaning windows is rated the most difficult household task by more than one third of adults.

Behind the TV	24%
Don't know	8%
Tops of Shelves	16%
Under couch	12%
Venetian	35%
Wooden floors	5%

- (d) Using the information given, estimate the number of surveyed adults who think cleaning under the couch is the most difficult cleaning job.
  - (e) What do you think the **margin of error of plus or minus 5%** means?
  - (f) How would you use the **margin of error** in estimating the percentage of all adults who think that the Venetian blinds are the hardest to clean?
2. A working knowledge of statistics is very helpful when you want to understand the statistics reported in the news. The news media and the government often make statements like, "Crime rate jumps 50% in your city."
- (a) Does an increase in the crime rate from 4% to 6% represent an increase of 50%?

- (b) Why would anybody report an increase from 4% to 6% as a "50%" rate jump?
3. A quality-control technician selects assembled parts from an assembly line and records the following information concerning each part:
- A: defective or non-defective  
 B: the employee number of the individual who assembled the part C:  
 the weight of the part
- (a) What is the population?  
 (b) Is the population finite or infinite?  
 (c) What is the sample?  
 (d) Classify the three variables.
4. The number of children,  $x$ , belonging to each of eight families registering for swimming was 1,2,1,3,2,1,5,3. Find the mean,  $\bar{x}$ .
5. Find the median rate paid at Jim's Burgers (in the US) if the workers' hourly rates are \$4.25, \$4.15, \$4.90, \$4.25, \$4.60, \$4.50, \$4.60, \$4.75.
6. A random sample of 10 of the 2005 Nextel Cup NASCAR drivers produced the following ages: 33, 48, 41, 29, 40, 48, 44, 42, 49, 28.
- (a) Find the mean age for the 10 NASCAR drivers of the 2005 Nextel Cup.  
 (b) Find the median age for the 10 NASCAR drivers of the 2005 Nextel Cup.  
 (c) Find the mode for age for the 10 NASCAR drivers of the 2005 Nextel Cup.
7. The summation  $\sum(x - \bar{x})$  is always zero. Why? Think back to the definition of the mean and see if you can justify your answer.
8. Consider the sample 2, 4, 7, 8, 9. Find the following:
- (a) Range  
 (b) Variance,  $s^2$   
 (c) Standard deviation,  $s$
9. Recruits for a police academy were required to undergo a test that measures their exercise capacity. The exercise capacity (in minutes) was obtained for each of 20 recruits:
- |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 25 | 27 | 30 | 33 | 30 | 32 | 30 | 34 | 30 | 27 |
| 26 | 25 | 29 | 31 | 31 | 32 | 34 | 32 | 33 | 30 |
- (a) Draw a dot plot of the data.  
 (b) Find the mean.  
 (c) Find the range.  
 (d) Find the variance.  
 (e) Find the standard deviation.

- (f) Using the dot plot from part (a), draw a line starting at the mean with a length that represents the value of the standard deviation.
- (g) Describe how the distribution of data, the range and the standard deviation are related.

10. The following data are the yields (in pounds) of hops:

3.9	3.4	5.1	2.7	4.4	7.0	5.6	2.6	4.8	5.6
7.0	4.8	5.0	6.8	4.8	3.7	5.8	3.6	4.0	5.6

- (a) Find the first and third quartiles of the yields.
- (b) Find the second quartile.
- (c) Find and explain the percentiles  $P_{15}$ ,  $P_{33}$ ,  $P_{90}$ .

11. The U.S. Geological Survey collected atmospheric deposition data in the Rocky Mountains. Part of the sampling process was to determine the concentration of ammonium ions (in percentages). Here are the results from the 52 samples. Use Excel to determine:

2.9	4.1	2.7	3.5	1.4	5.6	13.3	3.9	4.0
2.9	7.0	4.2	4.9	4.6	3.5	3.7	3.3	5.7
3.2	4.2	4.4	6.5	3.1	5.2	2.6	2.4	5.2
4.8	4.8	3.9	3.7	2.8	4.8	2.7	4.2	2.9
2.8	3.4	4.0	4.6	3.0	2.3	4.4	3.1	5.5
4.1	4.5	4.6	4.7	3.6	2.6	4.0		

- (a) Find  $Q_1$ .
- (b) Find  $Q_2$ . (c) Find  $Q_3$ .
- (d) Find  $P_{30}$ .
- (e) Find the 5-number summary.
- (f) Draw the box-and-whiskers plot.

12. Given a sample of measured values:

6	10	13	11	12	8	8	11	4
---	----	----	----	----	---	---	----	---

- (a) Calculate the Sample Mean, Variance and Standard Deviation.
- (b) Calculate the Five Number Summary, Range and IQR (by pen and paper). You will have to show your steps on paper.
- (c) Make a Stem-and-leaf plot from the data.
- (d) Check if there are outliers using the  $1.5 \times IQR$  rules. You will have to show your steps on paper. Construct a Boxplot from the 5 Number Summary.
- (e) If we multiple each of the original data by 10 then subtract 3, work out the new Sample Mean, Variance, Standard Deviation, Median, Range and IQR. Try to work out a general rule for doing this. These are referred to as affine transformations.