TRAFFIC ENGINEERING

SAB3843

STATISTICS & TRAFFIC DATA ANALYSIS

CHE ROS BIN ISMAIL

and

OTHMAN BIN CHE PUAN

Statistics

- Statistics is the branch of scientific method which deals with the data obtained by counting or measuring the properties of population of natural phenomena.
- This branch of study includes:
  - The process of collecting data
  - The study of manipulation and arrangement of figures using mathematical processes, and
  - Interpretation of the figures

Traffic Data Analysis

Understanding of traffic data – types of data, data presentation and description, validity, basic statistical distribution of the data, etc.

Sampling – Population vs. Sample

- ‘Population’ refers to all the measurements that could be made.
- A ‘sample’ is a subset of measurements selected from the population.
- Samples are tested in order to make inferences about the properties of the population. Therefore, it is vital to be clear which population is of interest.

  E.g., when considering the level of car ownership, are we interested in a particular group of people, or a population of an area. Hence, information in the sample is used to make an inference about the population.

Accuracy of Sampling

- Sampling is necessary because it is usually impracticable to test the entire population.
- Each and every sample must be selected in a random manner so that it is representative of the population from which it is drawn
- A value arrived at by sampling is absolutely accurate only for the sample itself. For the population which it represents, a sample can only give an estimate whose accuracy is expressed in terms of probability.
- Therefore, the greater the size of the sample (i.e. number of observations) the greater the confidence that can be placed on the estimate for the population.

Data Description

- Data is an information which in general has 2 main characteristics;
    a) Qualitative – involves non-numerical data, e.g. consider “YES” or “NO” as an answer to questionnaires
    b) Quantitative – involves numerical data

Quantitative Data

Two forms:

i. Discrete data – figures obtained from counting processes, usually in integer form.

ii. Continuous data – figures obtained from measurements, can be in any forms.

Two ways of describing data are:

- Numerically
- Graphically

Numerical Descriptive Measures for Describing Data

Two most common measures are:

1. Measures of Central Tendency
2. Measures of scatter

Measures of central tendency

1. Arithmetic Mean (or simply known as 'mean')
   - Mean of a set of measurements is the sum of the measurements divided by the total number of measurements:

$$\bar{X} = \frac{\sum_{i=1}^{n}(f_i X_i)}{\sum_{i=1}^{n} f_i}$$

where i = 1,2,3,…….. n

Measures of central tendency

2. Median

The median of a set of measurements is the middle value when the measurements are arranged in order of magnitude. It, therefore, divides a histogram and a frequency polygon into two equal areas.

E.g., consider this set of data: 1, 3, 4, **7**, 8, 9, 10

Median = 50th percentile = $\mathbf{P_{50}}$

3. Mode – is the measurement that occurs most often

Mode = Mean – 3(Mean – Median)

Measures of Scatter

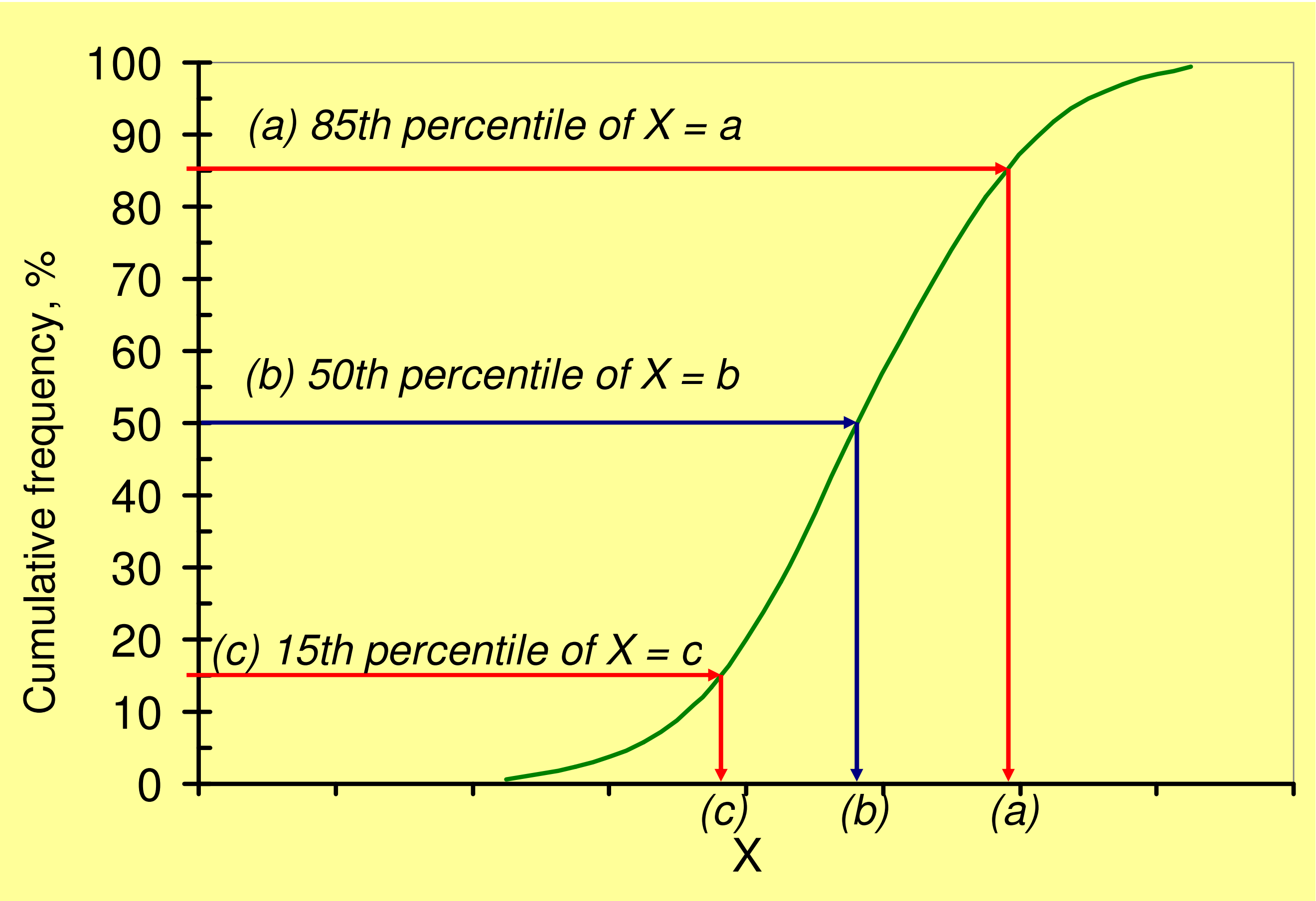
1) Range – the range of a set of measurements is defined to be the difference between the largest and the smallest measurements of the set.

eg. 15, 15, 20, 21, 30, 12, 11, 5, 40, 40, 26

Range = largest – smallest = 40 – 5 = 35

Measures of Scatter

2) Percentile – the r^{th} percentile of a set of n measurements arranged in order of magnitude is that value that has *r%* of the measurements below it.



Measures of Scatter

3) Variance (S^2)

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

or

$$S^2 = \frac{\sum_{i=1}^{n} fX_i^2}{\sum f} - \left(\frac{\sum_{i=1}^{n} fX_i}{\sum f} \right)^2$$

Measures of Scatter

4) Standard Deviation (SD)

SD is a measure of the average deviation of readings from their mean.

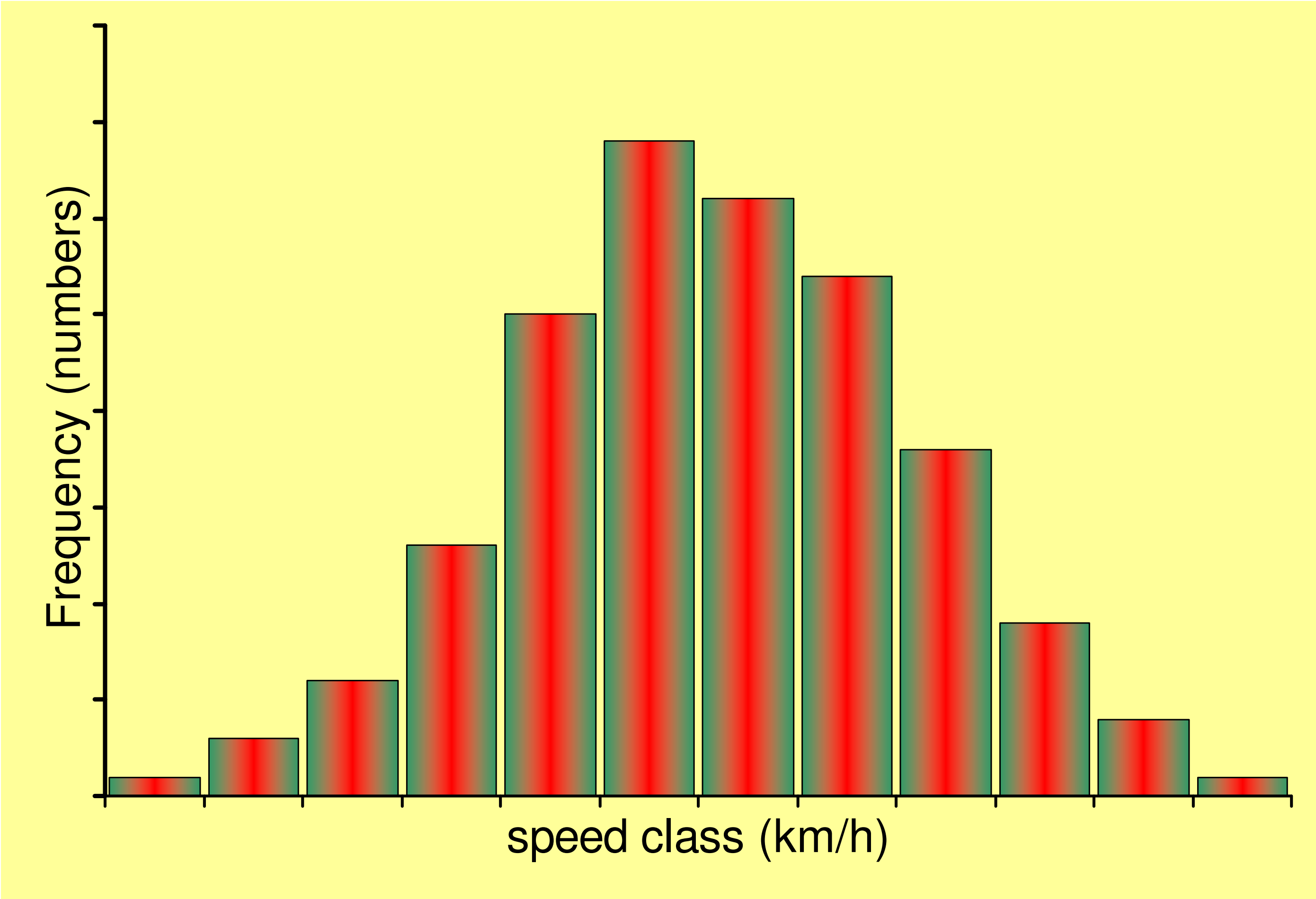
$$SD = \sqrt{\text{variance}}$$

5) Standard Error (SE)

$$SE = \frac{SD}{\sqrt{n}}$$

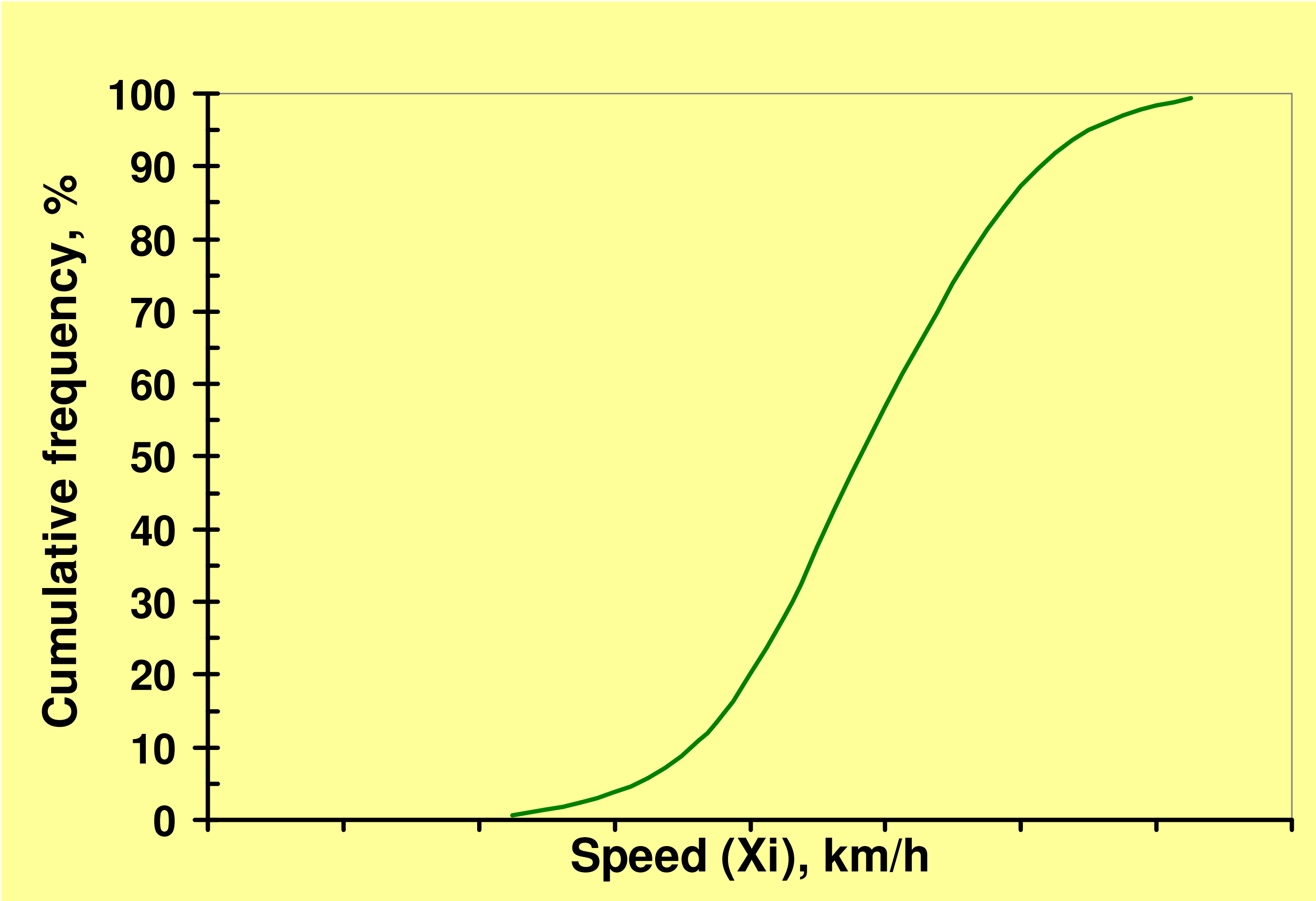
Graphical Method for Describing Data (typical diagrams)

1) Histogram



Graphical Method for Describing Data

2) Cumulative Frequency Curve



Example 1 – Spot speed analysis

Analyse the following spot speed data based on a sample of 172 vehicles traversing a section of sub-urban roadway.

Speed class (km/h)	Frequency f_i
20 - 25	1
25 - 30	3
30 - 35	6
35 - 40	13
40 - 45	25
45 - 50	34
50 - 55	31
55 - 60	27
60 - 65	18
65 - 70	9
70 - 75	4
75 - 80	1
Total	172

Solution 1 – tabulate data as follows

Speed class v (km/h)	Mid point v_i	Frequency F_i	Cum. Freq.	Cum. Freq. (%)	$F_i * v_i$	$F_i * v_i^2$
20 - 25	22.5	1	1	0.6	22.5	506.25
25 - 30	27.5	3	4	2.3	82.5	2268.75
30 - 35	32.5	6	10	5.8	195.0	6337.5
35 - 40	37.5	13	23	13.4	487.5	18281.25
40 - 45	42.5	25	48	27.9	1062.5	45156.25
45 - 50	47.5	34	82	47.7	1615.0	76712.5
50 - 55	52.5	31	113	65.7	1627.5	85443.75
55 - 60	57.5	27	140	81.4	1552.5	89268.75
60 - 65	62.5	18	158	91.9	1125.0	70312.5
65 - 70	67.5	9	167	97.1	607.5	41006.25
70 - 75	72.5	4	171	99.4	290.0	21025
75 - 80	77.5	1	172	100.0	77.5	6006.25
Total	600	172			8745	462325

Solution 1 – compute basic statistical facts for the data

- Mean speed = $$\bar{v} = \frac{\sum_{i=1}^{n}(f_i v_i)}{\sum_{i=1}^{n} f_i}$$ = 8745/172 = **50.84 km/h**

- Std deviation: $$SD = \sqrt{\left(\frac{\sum_{i=1}^{n} f_i v_i^2}{\sum f} - \left(\frac{\sum_{i=1}^{n} f_i v_i}{\sum f} \right)^2 \right)} = \sqrt{\left\{ \frac{462325}{172} - \left(\frac{8745}{172} \right)^2 \right\}}$$

  = **10.16 km/h**

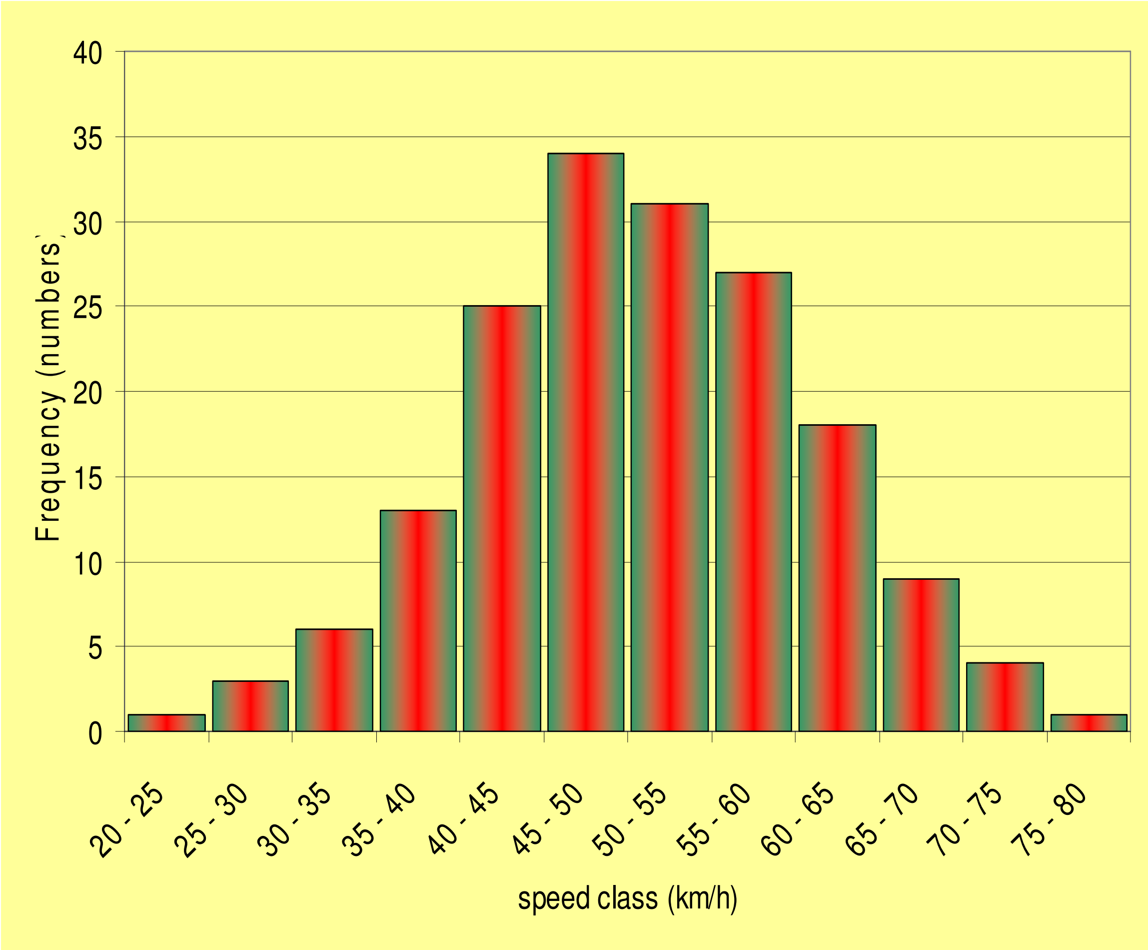
Can you compute the variance and standard error for the data?

What can you say about this result?

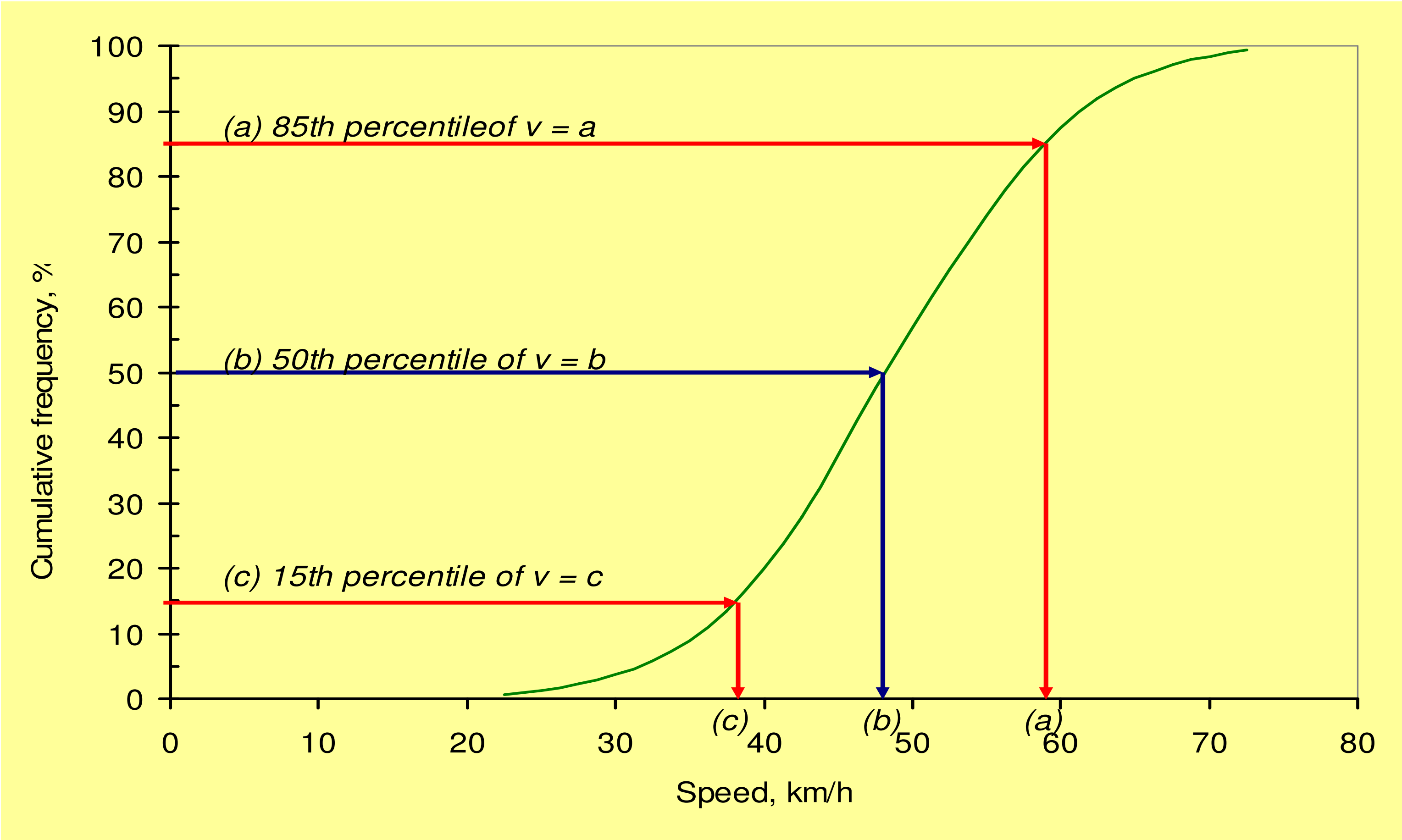
Solution 1 – plot the histogram for the data

Examine the plot & answer these:

- Do you think that the general shape of the plot is a typical of a normal distribution data?
- Do you think that the data should follow a normal distribution curve?
- Is the expected mean lies somewhere in the middle of the plot?
- What if it is not so?



Solution 1 – plot cumulative curve



- Compare the calculated mean & median, should both values are equal? Which value to report?
- Establish the required speeds at various percentiles, in what way these values will be used?

Example 2

Evaluate the following traffic data obtained for 7 consecutive days on a stretch of road section.

Day	Traffic volume (veh/day)
Monday	3231
Tuesday	3011
Wednesday	3137
Thursday	3247
Friday	3065
Saturday	3240
Sunday	1530

Solution 2

Compute the average traffic volume per day:

Day	Traffic volume (veh/day)
Monday	3231
Tuesday	3011
Wednesday	3137
Thursday	3247
Friday	3065
Saturday	3240
Sunday	1530

Average = total traffic/7	= 20461/7
	= **2923 veh/day**

By definition, the average volume of 2923 veh/day can be reported as the PLH or Purata Lalu Lintas Harian for the road.

But, certain traffic analyst may remove the data taken on Sunday because we have 6 data points with more than 3000 & only 1 data is much lesser. PLH is not representative.

The reported PLH would be

= 18931/6 = 3155 veh/day

Example 3

Two series of one–week traffic counts were carried out on a stretch of rural road and the data obtained are as follows:

Day	*March 2006* (veh/day)	*October 2006* (veh/day)
Monday	12500	10300
Tuesday	10500	12000
Wednesday	15200	13000
Thursday	13400	14500
Friday	16000	15200
Saturday	10500	8500
Sunday	8000	10200

(a) Determine the ADT and AADT on that particular road section.

(b) State the AADT in PCU/day if the average composition is 45% cars, 20% medium lorries, 10% buses, 7% heavy lorries & 18% motorcycles.

References

1. Garber, N.J., Hoel, L.A., TRAFFIC AND HIGHWAY ENGINEERING,4th Edition, SI Version., Cengage Learning (2010).
2. Currin, T. R., INTRODUCTION TO TRAFFIC ENGINEERING – A Manual for Data Collection and Analysis, Brooks/Cole (2001).
3. Kadiyali, L.R., TRAFFIC ENGINEERING AND TRANSPORT PLANNING, Khanna Publishers (1987) .
4. Othman Che Puan. Modul Kuliah Kejuruteraan Lalu Lintas. Published for Internal Circulation. (2004).
5. Dorina Astana, Othman Che Puan, Che Ros Ismail, TRAFFIC ENGINEERING NOTES, Published for Internal Circulation. (2011)