

UNIVERSITI TEKNOLOGI MALAYSIA

SCSI2143 / SCSI2143

Probability & Statistical Data Analysis

2019/2020 – Semester 2

ASSIGNMENT 1 (5%)

Due date – 10 March 2020

Answer all questions below

Question 1 – 5M

Determine which of the four levels of measurement (nominal, ordinal, interval and ratio) is most appropriate.

- a) Heights of women basketball players in a tournament.
- b) Ratings of superior, above average, average, below average or poor for blind dates.
- c) Noon temperatures (in degrees Celsius) in a Johor Bahru this week.
- d) A movie critic's classification of "drama, comedy, adventure".
- e) Consumer Reports magazine ratings of "best buy, recommended, not recommended".
- f) Distances travelled by students who commute to faculty.
- g) The actual contents (in ml) of cola in Pepsi cans.
- h) IQ scores, where the score is considered to be a measure of intelligence.
- i) IQ scores, where the score is considered to be the number of points scored on the IQ test.
- j) The number of bugs made when a programmer develop a coding for a project.

Question 2 – 10M

- a) A survey will be given to 100 students randomly selected from the freshmen class at Lincoln High School.
 - i. What is the population?
 - ii. What is the sample?
- b) State the scale/level of measurement for the following:
 - i. City of birth
 - ii. A score on a 5-point quiz measuring knowledge of algebra.
 - iii. Comparing the number of girls to the number of boys in a classroom.
- c) A verbal MUET scores of 20 high school students are recorded 280, 340, 440, 490, 520, 540, 560, 560, 580, 580, 600, 610, 630, 650, 660, 680, 710, 730, 740, 740.
 - i. Draw the box-plot of these data.
 - ii. Show the calculations for quartiles and inter-quartile range.

Question 3 – 10M

Air pollutant index (API) is used as reference to measure the level of air pollution. Table 1 shows the reading of air pollutant index in 19 areas taken at 12 noon. The API readings were grouped into according index.

Table 1: Air pollutant index in 19 areas taken at 12 noon

API	Frequencies
0-50 (good)	1
50-100 (moderate)	2
100-200 (Unhealthy)	6
200-300 (Very Unhealthy)	7
300-500 (Hazardous)	3

- Sketch a histogram to represent the frequencies of API in table 1.
- In order to better understand the air pollutant situation, calculate relative frequencies (2 decimal places) and state the percentage of areas affected with $\text{API} \geq 100$.
- Assume that an API reading of 950 was taken at 12 noon at one of the areas. Does this value affect the mean value for API? Justify your answer.

Question 4 – 10M

Table 2 shows deaths caused by dengue fever in Malaysia from the year 2006 – 2016.

Table 2: Dengue fever deaths from year 2006 – 2016 in Malaysia

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Deaths	92	98	112	88	134	36	35	92	215	336	237

- Find mean, mode, median and standard deviation
- Number of deaths above 200 is considerably high. Determine the percentile of the deaths 215?
- Which statistic measure do you think is the most suitable to show the danger of dengue fever? Explain your answer.

Question 5 – 10M

In January of 2006, Adam family moved to a tropical climate country. For the following year (2007), Adam has recorded the number of rainy days that occurred each month. The data from January to December contained 14, 14, 10, 12, 11, 13, 11, 11, 14, 10, 13, 8.

- a) Find the mean, mode, median and range for the data set of rainy days.
- b) What can you tell about the distribution of the data based on the median and mean found in (a)?
- c) Find the standard deviation for the data set of rainy days above
- d) Suppose the number of rainy day in December has been recorded wrongly. The observed data is 7 not 8. Does this data can be considered an outlier? Support your answer

Question 6 – 5M

A civil engineer monitors water quality by measuring the amount of suspended solids in a sample of river water. Over 5 days, she observed suspended solids (parts per million) 14, 12, 21, 28, 30.

- a) Find the skewness value for the sample.
- b) Is the sample positively skewed or negatively skewed? Justify your answer.