UNIVERSITI TEKNOLOGI MALAYSIA

SCSI2143 / SCSI2143

Probability & Statistical Data Analysis

2019/2020 – Semester 2

ASSIGNMENT 3 (10%)

Due date – 20th April 2020

This is an individual assignment. Please submit this assignment in pdf format via your e-learning.

Question 1 (10 Marks)

Based on statistics given in Table 1, determine the following:

a) Determine the data type (discrete or continuous, qualitative or quantitative) and measuring scales for each variable. (7 marks)

b) Is the data in Table 1 considered as primary or secondary data source? Why? (1½ marks)

c) Determine appropriate graph or plot that can be used to present below variables: (1½ marks)

   i. Gender

   ii. Age

   iii. Date Discharged

Table 1: COVID-19 cases in Malaysia

(https://www.soyacincau.com/2020/03/15/covid-19-malaysia-records)

Kes COVID-19 di Malaysia: Pesakit yang sembuh & discaj

COVID-19 cases in Malaysia: Patients recovered & discharged

Jumlah pesakit sembuh *Total of cured patients* **42**

No kes/ Case No.	Jantina/ Gender	Umur/ Age	Warganegara/ Nationality	Tarikh disahkan positif/ Date confirmed	Tarikh discaj Date discharged	Hospital
6	Perempuan Female	4 Tahun 4 Years Old		28 Jan 2020	4 Feb 2020	Hospital Sultanah Maliha, Langkawi
4	Lelaki Male	40 Tahun 40 Years Old		25 Jan 2020	8 Feb 2020	Hospital Permai, Johor
10	Lelaki Male	63 Tahun 63 Years Old		3 Feb 2020	9 Feb 2020	Hospital Kuala Lumpur
1	Lelaki Male	11 Tahun 11 Years Old		24 Jan 2020	14 Feb 2020	Hospital Sungai Buloh, Selangor
2	Lelaki Male	2 Tahun 2 Years Old		24 Jan 2020	14 Feb 2020	Hospital Sungai Buloh, Selangor
3	Perempuan Female	65 Tahun 65 Years Old		24 Jan 2020	14 Feb 2020	Hospital Sungai Buloh, Selangor
5	Perempuan Female	36 Tahun 36 Years Old		28 Jan 2020	14 Feb 2020	Hospital Sungai Buloh, Selangor
15	Perempuan Female	59 Tahun 59 Years Old		7 Feb 2020	16 Feb 2020	Hospital Permai, Johor
9	Lelaki Male	41 Tahun 41 Years Old		4 Feb 2020	17 Feb 2020	Hospital Sungai Buloh, Selangor
11	Lelaki Male	45 Tahun 45 Years Old		5 Feb 2020	18 Feb 2020	Hospital Tuanku Jaafar, Seremban
12	Lelaki Male	9 Tahun 9 Years Old		5 Feb 2020	18 Feb 2020	Hospital Tuanku Jaafar, Seremban
7	Lelaki Male	52 Tahun 52 Years Old		28 Jan 2020	18 Feb 2020	Hospital Permai, Johor
8	Perempuan Female	49 Tahun 49 Years Old		30 Jan 2020	18 Feb 2020	Hospital Permai, Johor
17	Perempuan Female	65 Tahun 65 Years Old		9 Feb 2020	19 Feb 2020	Hospital Sungai Buloh, Selangor
18	Lelaki Male	31 Tahun 31 Years Old		9 Feb 2020	19 Feb 2020	Hospital Sungai Buloh, Selangor
20	Lelaki Male	27 Tahun 27 Years Old		14 Feb 2020	19 Feb 2020	Hospital Sultanah Bahiyah,Kedah

Question 2 (20 Marks)

a) The temperatures (in °C) measured from 30 COVID-19 patients on day 5 of quarantine in isolation ward are listed below:

35.5	35.7	35.8	35.9	36.1	36.1	36.3	36.4	36.5	36.6
36.7	36.7	36.7	36.9	37.0	37.0	37.0	37.1	37.2	37.2
37.4	37.5	37.7	37.7	37.8	38.0	38.1	38.1	38.3	38.7

i. Draw a stem-and-leaf of the above data (3 marks)
ii. Calculate the 1st, 2nd and 3rd Quartile of the dataset (6 marks)
iii. Based on your answer in (i) and (ii), draw a boxplot of the data (3 marks)
iv. What can you say about the dataset by looking at the boxplot created in (iii)? (2 marks)

b) Table 2 shows the age distribution of females with positive COVID-19 in the red zone in Malaysia.

Table 2: Age of COVID-19 female patients

Age (years)	Frequency
30 - under 35	10
35 - under 40	27
40 - under 45	38
45 - under 50	47
50 - under 55	86
55 – under 60	102
60 – under 65	78
65 – under 70	56

i. Obtain the relative-frequency and cumulative relative-frequency table for the data. (4 marks)
ii. Based on your answer in (i), draw the relative ogive to represent the data. (2 marks)

Question 3 (20 Marks)

a) The table below shows the no. of positive case of COVID-19 in Malaysia according to age group provided by Malaysian Ministry of Health on 31st March 2020.

Table 3: Positive Cases versus Age Group

Age Group	No. of cases
1-5	27
6-10	23
11-15	43
16-20	78
21-25	123
26-30	180
31-35	144
36-40	136
41-45	118
46-50	133
51-55	143
56-60	182
61-65	137
66-70	90
71-75	46
76-80	20
81-85	14

Based on the data in **Table 3**, find:

i. mode (2 marks)

ii. median (2 marks)

iii. mean (2 marks)

b) **Table 4** shows the example temperature data taken from the workers who going to work at a premise of a grocery store in the morning.
**As suggested by World Health Organization (WHO) if the person develop a mild cough or low-grade fever (37.3c or more) need to stay home for self-isolation.*

Table 4: Temperatures of grocery workers in the morning

36.5	36.5	36.5	36.6	36.6
36.6	36.7	36.7	36.7	36.7
36.7	36.7	36.7	36.7	36.8
36.8	36.8	36.9	36.9	36.9

Based on the data in **Table 4,** find

i. mean temperature (2 marks)
ii. mode temperature (1 mark)
iii. median temperature (2 marks)

Ali as the supervisor of the grocery store, decided to have a close inspection of the staff body temperature throughout the day by taking another temperature when the store is closed in the evening. **Table 5** shows the temperature data at the evening.

Table 5: Temperatures of grocery workers in the evening

36.5	36.5	36.5	36.6	36.6
36.6	36.7	36.7	36.7	36.7
36.7	36.7	36.7	36.7	36.8
36.8	36.9	37.0	37.0	37.0

Based on the data in **Table 5,** answer below questions:

iv. As a precaution, if Ali decided to suggest the staff whose body temperature shows 36.9 and above to stay put at home. What is the percentile for 36.9? (2 marks)
v. What is the standard deviation of the data in Table 5? (2 marks)
vi. Discuss the data in terms of skewness. (3 marks)
vii. Calculate the kurtosis of the data and what does it mean? (2 marks)

Question 4 (15 Marks)

The data in **Table 6** shown below were extracted from a crowdsourced Malaysian website (https://www.outbreak.my/stats) which comprised of age on 160 COVID-19 patients.
You can use any software available to calculate the parameters for this question.

Table 6: Observed age of COVID-19 patients from a Malaysian database

11	45	32	50	27	9	58	56
2	9	83	43	38	42	50	44
65	40	53	58	59	50	8	53
40	37	41	40	62	46	28	35
36	59	54	51	30	36	52	21
4	67	52	49	41	50	61	24
52	64	20	63	62	16	33	57
49	32	45	60	55	29	36	47
41	39	35	45	51	61	32	32
63	7	38	49	56	63	51	36
51	80	63	69	70	61	52	85
53	46	38	64	70	63	64	69
53	73	60	77	50	56	85	69
12	60	61	66	68	69	61	81
60	60	72	48	37	48	48	73
68	68	74	49	66	62	91	56
26	16	57	79	61	47	76	37
66	58	60	40	40	57	27	78
58	53	51	62	53	83	46	71
34	62	76	35	75	73	84	55

a) Based on the data, find the range, variance, and standard deviation of age distribution for the Malaysia COVID-19 patients. (5 marks)

b) The following Figure 1 shows the age distribution of COVID-19 cases based on a crowdsourced China database.

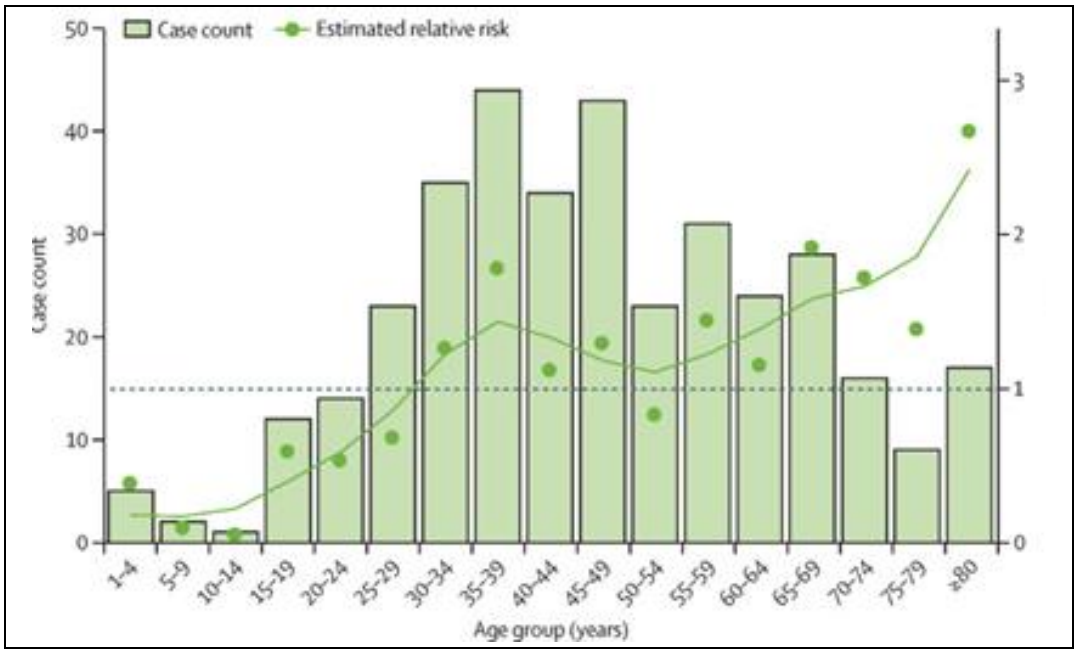

Figure 1: Age Distribution of COVID-19 Patients in China

i. Using the same scale of age group (*x*-axis) as shown in Figure 1, generate an "Age Distribution of patients with COVID19" based on the data in **Table 6** above (You can ignore the line chart for estimated relative risk). (3 marks)

ii. Calculate the skewness and kurtosis of the generated distribution in (i). (4 marks)

iii. Using the generated graph in (i), provide comparisons and discussions of the two distributions in terms of skewness and kurtosis. (3 marks)

Question 5 (10 Marks)

In the recent COVID-19 outbreak, 58% of the suspected cases who goes for screening test get positive result. It is known that 28% of the confirmed positive cases has history of traveling to the country with outbreak, 50% has close contact with confirmed positive person and another 22% is from unknown cluster.

a) Represent the probability of the COVID-19 outbreak described above using tree diagram.
You can see examples of tree diagram in https://www.mathsisfun.com/data/probability-tree-diagrams.html (3 marks)

b) What is the probability of suspected cases with negative screening test result? (1 mark)

c) If a total of 72 person-under-investigation (PUI) was suspected to be exposed to this virus and are required to take swab test. How many people will get negative result? (1 mark)

d) Find the probability of positive cases having travelling history to the country with outbreak or close contact with person infected? (2 marks)

e) Death rate of this pandemic is 1.6% while the recovery rate is 34%. If the total infected person is 2766 people,
   i. What is the probability of infected person under treatment? (2 marks)
   ii. How many are under treatment? (1 mark)

Question 6 (25 Marks)

a) On 17th March 2020, Malaysia recorded its first death of COVID-19. Since then, the number of death is increasing. Table 7 shows the probability distribution of number of deaths in a day for 22 days' period since 17th March 2020.

Table 7: Probability distribution of number of deaths is a day

0	1	2	3	4	5	6	7	8
0.091	0.182	0.182	0.227	0.137	0.091	0.045	0.000	0.045

   i. Find the probability that at most 2 deaths occur in a day. (1 mark)
   ii. Find the probability that at least 5 deaths occur in a day. (1 mark)
   iii. Find the probability that at least 3 deaths but not more than 5 deaths occur in a day. (1 mark)
   iv. Based on Table 7 data for 22 days, find mean and variance of the number of deaths occur in a day. (2 marks)

b) According to World Health Organization (WHO), the risk of dying if infected by COVID-19 for a patient with a given pre-existing condition is shown in Table 8.

Table 8: Probability of dying for a patient with a given pre-existing condition

Pre-existing Condition	Probability
Cardiovascular Disease	0.105
Diabetes	0.073
Chronis Respiratory Disease	0.063
Hypertension	0.060
Cancer	0.056
No Pre-existing Conditions	0.009

Suppose that among 56 COVID-10 patients in a hospital, there are 8 patients with cardiovascular disease, 4 patients with diabetes, 10 patients with chronic respiratory disease, 8 patients with hypertension and 5 patients with cancer.

i. Suppose that patients with cardiovascular disease are selected at random, what is the probability that at most two patients will die? (2 marks)

ii. Suppose that patients with diabetes are selected at random, what is the probability that at least two patients will die? (2 marks)

iii. Suppose that patients with chronic respiratory disease are selected at random, what is the probability that at most one patients will recover? (1½ marks)

iv. Suppose that patients with cancer are selected at random, what is the probability that at least two but no more than four patients will recover? (1½ marks)

v. Suppose that patients with no pre-existing conditions are selected at random, what is the probability that all patients will recover? (1 mark)

c) The death rate (probability of dying if infected with COVID-19) for a male patient is 0.047 while for a female patient is 0.028. Suppose that 7 male cases and 5 female cases are randomly selected.

i. Assuming independence trials, what is variable *X* and how *X* is distributed? (1 mark)

ii. What is the probability that the first death among female patients occurs in the 3rd cases? (1 mark)

iii. An experimental drug will be introduced to COVID-19 patients. What is the probability that among the infected male patients, the first patient recover after taking the drug is found at the fifth trial? (1 mark)

d) A team of scientists are collaborating to identify how quickly COVID-19 can spread from person to person. The team produces a serial interval of COVID-19 which defined as the time duration between a primary case (infector) developing symptoms and secondary case (infectee) developing symptoms. To obtain reliable estimates of the serial interval, they obtained data on 468 COVID-19 transmission events reported in mainland China outside of Hubei Province between 21st January 2020 and 8th February 2020. The team of scientists find that COVID-19 serial intervals better resemble a normal distribution than other more commonly distributions. They found that the distribution is having a mean of 4 days and standard deviation of 5 days.

What is the probability that the time duration between infector and infectee is

i. more than 6 days? (2 marks)
ii. between 2 to 4 days? (5 marks)
iii. less than 0 days? It means that there is a possibility of asymptomatic transmission (transmission of the virus from an infector who does not develop any symptoms) (2 marks)