SECI2143: PROBABILITY & STATISTICAL DATA ANALYSIS

2021/2022 - SEMESTER 2

ASSIGNMENT 1

INSTRUCTION:

1. This is an assignment for a **GROUP of 3-4**. Please clearly write the group members name & matric number in the front page of the submission.
2. This assignment has 4 questions (100 marks), which will give 5% of overall course marks.
3. Only **HANDWRITTEN** submission is accepted:
    a. Submission using any reporting or statistical tools (e.g.: MS Word, MS Excel, etc,) will be **REJECTED**.
    b. Make sure the submission is neatly written. Any submission with handwriting that is unreadable, will be **REJECTED**.
    c. For answer that need to draw graphs, using graph paper is optional. You can use plain paper.
    d. Round your answers to **TWO** decimal places.
    e. Please scan/snapshot your work and save as **PDF** file.
4. Due date: 21st April 2022, 11.59 pm (via eLearning) – only **ONE** group member needs to submit on behalf of the group.

QUESTION 1 (20 MARKS)

A lecturer asked her students to conduct a study on "Internet Usage among Computer Science Students in Malaysia". She outlined the expected output of the study, in the form of various graphical formats as shown in Figure 1 – 5.

As a teaching assistant, you are required to assist the student to:

1.1 Explain descriptive and Inferential Stat related to the study. (2 marks)

1.2 Describe population and sample of the study. (2 marks)

1.3 Outline data analysis process regarding the study. (6 marks)

1.4 Suggest variables of type Nominal, Ordinal, Interval and Ratio that can be used in the study. (other variables suggested by the lecturer) (4 marks)

1.5 Prepare a google form with suitable questions and corresponding possible answers that is able to collect data and generate output as outlined in Figure (1) to (5). Provide a link of the google form in your answer sheet. (6 marks)

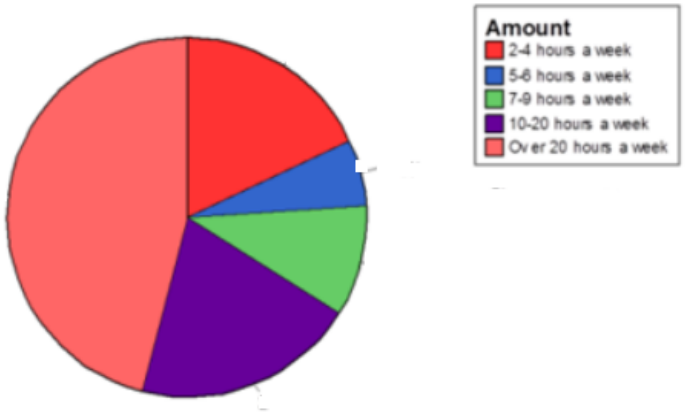

Figure 1. Time Spent on the Internet

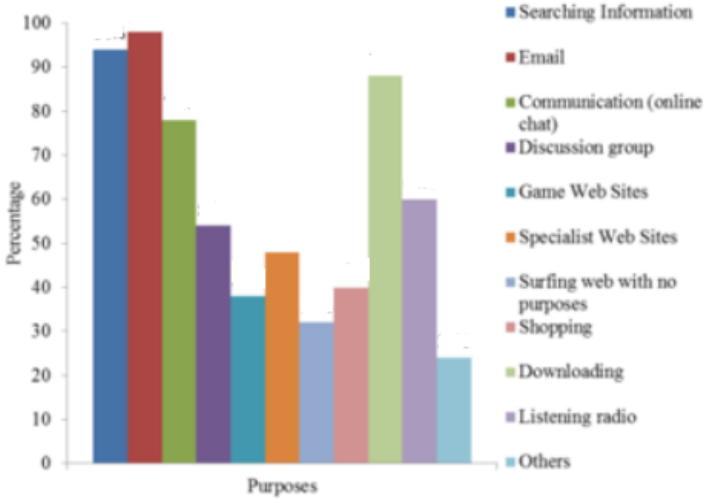

Figure 2. Purposes of using Internet

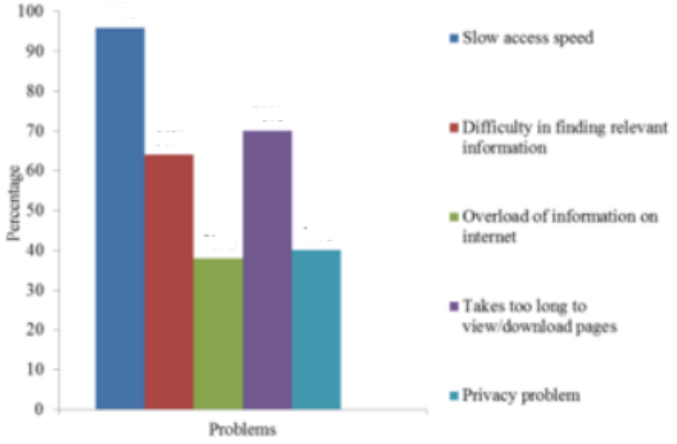

Figure 4. Problems when using internet

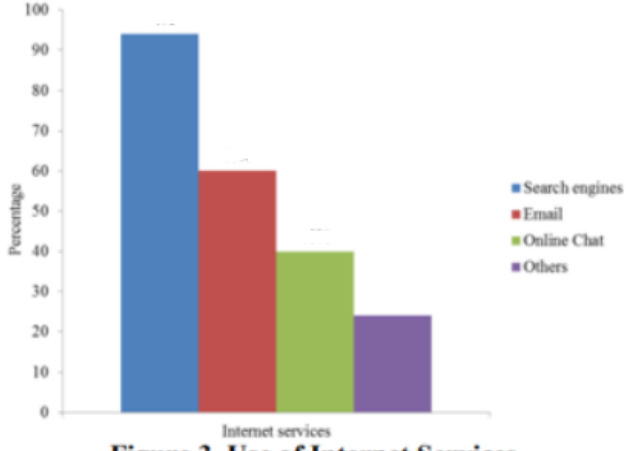

Figure 3. Use of Internet Services

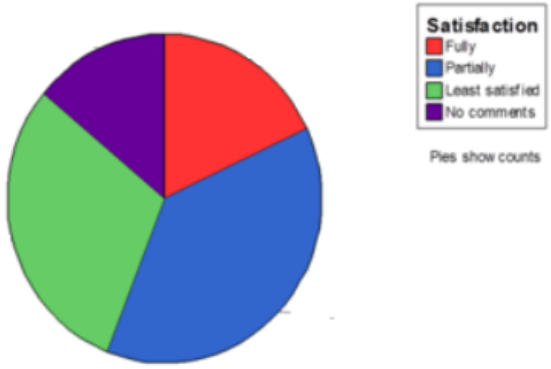

Figure 5. Satisfaction with Internet Facilities

QUESTION 2 (10 MARKS)

The following are data regarding the time taken by a group of 1st year students who scored 75 and above in an Online Computer Skills Assessment. The students are from UNIVERSITI CANGGIH MALAYSIA.

35	33	38	10	52	48	34	44	41	56
53	41	39	32	37	60	34	45	33	57
38	48	40	32	55	52	35	47	44	15

2.1) Represent the data in form of:

a) Dot Plot (3 marks)

b) A Stem-and-Leaf plot (3 marks)

2.2) If you are the second fastest student in answering the assessment, what percentile are you at? (4 marks)

QUESTION 3 (20 MARKS)

The same Online Computer Skills Assessment were taken by a group of FINAL year students from the same university. The time taken to answer the assessment for the students who scored 75 and above were recorded and presented in the following table. (**Note:** Lowest limit of the class start at 0)

Mid-value	5	15	25	35	45	55
Frequency	7	10	23	51	6	3

3.1) Complete the above frequency table to enable you to answer the following questions. (2 marks)

3.2) Represent the grouped data in form of

a) Histogram (3 marks)

b) Polygon (3 marks)

c) Ogive graph (3 marks)

d) Using ogive graph, estimate what percentile of students answer the assessment in more than 40 minutes? (3 marks)

3.3) Calculate mean, median and mode of the grouped data. (6 marks)

QUESTION 4 (50 MARKS)

A study is then conducted to compare the answering speed among the first year students from two different universities. The answering speed represents time taken to answer Online Computer Skills Assessment for those who scored 75% and above. The first sample of data is as stated in Question 2 and the second sample of data by students from UNIVERSITI MUHIBBAH MALAYSIA) are shown below:

56	59	72	70	51	57	55	57	54	57
49	73	65	60	52	48	54	55	63	53
50	50	90	52	49	54	49	48	59	50

4.1) For both datasets, produce summary statistics which comprise of sample

a) range (1 mark)
b) mean, (1 mark)
c) median, (2 marks)
d) mode, (1 mark)
d) first Quartile, (2 marks)
e) third Quartile, (2 marks)
f) interquartile range, (1 mark)
g) variance, (2 marks)
h) standard deviation, (2 marks)
i) skewness, (2 marks)
j) Kurtosis

4.2) Is there any outlier in both datasets? Justify your answers. (3 marks)

4.3) Draw a normal/modified box plot for both datasets. (4 marks)

4.4) Based on the summary statistics that you have produced in (4.1), comment on the differences between the two datasets in terms of

a) central tendency (4 marks)
b) spread/dispersion and (4 marks)
c) shape (4 marks)

of the box-plots for both datasets.

4.5) Apart from summary statistics that you have produced in (4.1),

a) Suggest a graphical format that can be used to describe the spread and shape of both datasets. (1 mark)
b) Represent each dataset in form of graphical format suggested in 4.5 (a). (6 marks)
d) Comment on the differences between the two datasets in terms of shape. (Give your comment based on graphical format drawn in 4.5 (b)) (4 marks)

4.6) Do the answering speed of students from UNIVERSITI CANGGIH look significantly different from those students from UNIVERSITI MUHIBBAH MALAYSIA? Justify your answers based on your drawings in 4.3 and comments made in 4.4. (4 marks)