

**FINAL
ALTERNATIVE ASSESSMENT**

(COVER PAGE)

Session : APRIL 2021

Programme : Diploma In Computer Science (DCS)

Course : DCS1104: INTRODUCTION TO STATISTICS AND DATA ANALYTICS

Date of Examination : July 27, 2021 (Tuesday)

Time : 8.00am – 10.30am Reading Time : Nil

Duration : 2 Hours 30 Minutes

Note: 30 minutes is added into the duration of the examination to factor in any connectivity matters and for you to scan and upload your scripts.

Special Instructions :

This paper consists of **FOUR (4)** questions. Answer **ALL** the questions. **Write ALL your answers** in the foolscap papers.

IMPORTANT NOTE : THIS PAPER SHOULD NOT BE TAKEN OUT OF THE EXAMINATION HALL

Materials permitted : Nil

Materials provided : Nil

Examiner(s) : Yogeswari Suppiah and Siti Syakirah

Chief Moderator : Ms Siti Hajar Khairuddin

This paper consists of 5 printed pages, including the cover page

DIPLOMA IN COMPUTER SCIENCE PROGRAMME (DCS)
DCS1104: INTRODUCTION TO STATISTICS AND DATA ANALYTICS
FINAL ALTERNATIVE ASSESSMENT: APRIL 2021 SESSION

Instruction: This paper consists of **FOUR (4)** questions. Answer **ALL** the questions in foolscap papers.

Question 1

- (a) List **THREE(3)** usage of data. (3 marks)
- (b) Describe 3Vs in big data. (6 marks)
- (c) Data Analytics are widely used to support decision making in variety of fields. List **FIVE(5)** field that uses data analytics extensively. (5 marks)
- (d) Explain what is meant by small data, and provide **TWO (2)** examples of small data. (3 marks)
- (e) What is the difference between small data and big data? (2 marks)
- (f) Information is normally stored in table, which consists of rows and columns. A column is referred to as an attribute (feature) whereas row is referred to as an instance (object). Describe the term attribute and instance. (4 marks)
- (g) List **TWO(2)** taxonomy that exists in data analytics. (2 marks)

[Total :25 marks]

Question 2

Table shows a dataset for employees in a department.

ID	HEIGHT (cm)	WEIGHT (kg)	GENDER	HEALTH
100001	150	80	M	SICK
100002	159	79	M	HEALTHY
100003	153	56	F	SICK
100004	169	120	F	HEALTHY
100005	184	65	M	HEALTHY
100006	190	71	F	SICK
100007	158	83	M	HEALTHY
100008	167	92	F	HEALTHY
100009	163	79	M	HEALTHY
100010	175	75	F	SICK
100011	162	78	F	HEALTHY
100012	182	98	M	HEALTHY
100013	161	107	M	SICK
100014	160	56	M	SICK
100015	167	63	F	HEALTHY

Based on the above table, calculate the following for **Weight** column:

- (a) Minimum (1 mark)
- (b) Maximum (1 mark)
- (c) Mode (1 mark)
- (d) Mean (2 marks)
- (e) Median (2 marks)
- (f) First Quartile (2 marks)

- (g) Third Quartile (2marks)
- (h) Interquartile Range (1 mark)
- (i) Amplitude (1 mark)
- (j) Mean Absolute Deviation (4 marks)
- (k) Variance (4 marks)
- (l) Standard Deviation (4 marks)

[Total :25 marks]

Question 3

Referring to the table in Question 2:

- (a) Draw a scatter plot chart for **weight**, based on **height**. (5 marks)
- (b) Based on plot chart from (a), modify the chart so that it is defined by **gender** in term of:
 - i. Color
 - ii. Shape(15 marks)
- (c) Draw a contingency table with absolute joint frequencies for “**Gender**” and “**Health**”. (5 marks)

[Total :25 marks]

Question 4

(a) In real-life applications, it is common that some of predictive attribute values for some records may be missing in the data set. List **FIVE (5)** causes of missing values.

(5 marks)

(b) Discuss the following Data Mining techniques:

- i. Clustering
- ii. Frequent Pattern Mining
- iii. Regression
- iv. Classification

20 marks)

[Total: 25 marks]

~ The End ~

DCS1104 (Final)/April2021