



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Sekolah Pendidikan
Profesional dan
Pendidikan
Berterusan
(SPACE)

**FINAL EXAMINATION / PEPERIKSAAN AKHIR
SEMESTER I – SESSION 2022 / 2023 / SEMESTER 1-SESI 2022/2023
PROGRAM KERJASAMA**

COURSE CODE : DDWD 3773
KOD KURSUS

COURSE NAME : DATA ANALYTICS / ANALISIS DATA
NAMA KURSUS

YEAR / PROGRAMME : 3 DWWD
TAHUN / PROGRAM

DURATION : 2 HOURS 30 MINUTES
TEMPOH : 2 JAM 30 MINIT

DATE : DECEMBER 2022 / JANUARY 2023
TARIKH : DISEMBER 2022 / JANUARI 2023

INSTRUCTION :
ARAHAN

Answer **ALL** questions in the spaces provided in this question paper.

Jawab **SEMUA** soalan di ruang yang disediakan dalam kertas soalan ini.

(You are required to write your name and your lecturer's name on your answer script)

(Pelajar dikehendaki tuliskan nama dan nama pensyarah pada skrip jawapan)

NAME / NAMA PELAJAR	:
I.C NO. / NO. K/PENGENALAN	:
YEAR / PROGRAMME TAHUN / PROGRAM	:
COLLEGE NAME NAMA KOLEJ	:

This examination paper consists of **16** pages including the cover
Kertas soalan ini mengandungi **16** muka surat termasuk kulit hadapan



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

School of
Professional and
Continuing
Education
(SPACE)

PUSAT PRGORAM KERJASAMA

PETIKAN DARIPADA PERATURAN AKADEMIK ARAHAN AM – PENYELEWENGAN AKADEMIK

1. SALAH LAKU SEMASA PEPERIKSAAN

1.1. Pelajar tidak boleh melakukan mana-mana salah laku peperiksaan seperti berikut :-

- 1.1.1. memberi dan/atau menerima dan/atau memiliki sebarang maklumat dalam bentuk elektronik, bercetak atau apa jua bentuk lain yang tidak dibenarkan semasa berlangsungnya peperiksaan sama ada di dalam atau di luar Dewan/Bilik Peperiksaan melainkan dengan kebenaran Ketua Pengawas; atau
- 1.1.2. menggunakan maklumat yang diperoleh seperti di atas bagi tujuan menjawab soalan peperiksaan; atau
- 1.1.3. menipu atau cuba untuk menipu atau berkelakuan mengikut cara yang boleh ditafsirkan sebagai menipu semasa berlangsungnya peperiksaan; atau
- 1.1.4. lain-lain salah laku yang ditetapkan oleh Universiti (seperti membuat bising, mengganggu pelajar lain, mengganggu Pengawas menjalankan tugasnya).

2. HUKUMAN SALAH LAKU PEPERIKSAAN

2.1. Sekiranya pelajar didapati telah melakukan pelanggaran mana-mana peraturan peperiksaan ini, setelah diperakukan oleh Jawatankuasa Peperiksaan Fakulti dan disabitkan kesalahannya, Senat boleh mengambil tindakan dari mana-mana satu yang berikut :-

- 2.1.1. memberi markah SIFAR (0) bagi keseluruhan keputusan peperiksaan kursus yang berkenaan (termasuk kerja kursus); atau
 - 2.1.2. memberi markah SIFAR (0) bagi semua kursus yang didaftarkan pada semester tersebut.
- 2.2. Jawatankuasa Akademik Fakulti boleh mencadangkan untuk diambil tindakan tatatertib mengikut peruntukan Akta Universiti dan Kolej Universiti, 1971, Kaedah-kaedah Universiti Teknologi Malaysia (Tatatertib Pelajar-pelajar), 1999 bergantung kepada tahap kesalahan yang dilakukan oleh pelajar.
- 2.3. Pelajar yang didapati melakukan kesalahan kali kedua akan diambil tindakan seperti di perkara dan dicadang untuk diambil tindakan tatatertib mengikut peruntukan Akta Universiti dan Kolej Universiti, 1971, Kaedah-kaedah Universiti Teknologi Malaysia (Tatatertib Pelajar-pelajar), 1999.

SECTION A: OBJECTIVE [10 MARKS]

BAHAGIAN A: OBJEKTIF [10 MARKS]

Instruction: Circle the correct answer in this questions paper.

Arahan: Bulatkan jawapan yang tepat di dalam kertas soalan ini.

1. Below are correct tools for Data Sciences, **EXCEPT**:

*Di bawah adalah alat yang tepat untuk Data Sains, **KECUALI**:*

- A. Python
- B. Assembly language / *Bahasa himpunan*
- C. RapidMiner
- D. KNIME

2. Which one of the following is a qualitative data

Berikut yang mana satukah data kualitatif

- A. Picture size is 10cm x 10cm / *siaz gambar 10cm x 10cm*
- B. Cost of frame is RM15 / *harga bingkai adalah RM15*
- C. The frame color is a gold / *warna bingkai adalah emas*
- D. Weigh of frame is 8 pounds / *berat bingkai adalah 8 pounds*

3. A teacher notes down the weight of each student in the class. What level of measurement is being used in this case?

Seorang guru mencatat berat bagi setiap pelajar di dalam kelas. Apakah tahap pengukuran yang digunakan bagi situasi ini?

- A. Nominal / *Nominal*
- B. Ordinal / *Ordinal*
- C. Interval / *Selang*
- D. Ratio / *Nisbah*

4. A course is being rated by students on a scale of 1 star to 5 stars. What is the level of measurement being used in this case?

Satu kursus yang dinilai oleh pelajar adalah berskala 1 bintang sehingga 5 bintang. Apakah tahap pengukuran yang digunakan bagi situasi ini?

- A. Nominal / *Nominal*
- B. Ordinal / *Ordinal*
- C. Interval / *Selang*
- D. Ratio / *Nisbah*

5. Interval and Ratio types of data by nature are:

Selang dan Nisbah adalah jenis data bersifat:

- A. Qualitative data / *Data kualitatif*
- B. Quantitative data / *Data kuantitatif*
- C. Univariate data / *Data univariat*
- D. Multivariate data / *Data multivariat*

6. Choose the **CORRECT** descriptive statistics for nominal data.

Pilih statistic deskriptif yang **BETUL** bagi data nominal.

- A. Percentile / *Persentil*
- B. Mode / *Mod*
- C. Skewness / *Kepencongan*
- D. Kurtosis / *Kurtosis*

7. The appropriate graphs for ratio data type are:

Graf yang bersesuaian bagi data berjenis nisbah adalah:

- i. Histogram / *Histogram*
- ii. Boxplot / *plot kotak*
- iii. Stem-and-leaf / *batang dan daun*
- iv. Bar chart / *carta bar*

- A. i, and ii / *i dan ii*
- B. i, ii, and iii / *i, ii dan iii*
- C. ii and iv / *ii dan iv*
- D. iii ONLY / *iii SAHAJA*

8. Mode refers to the value within a series that occurs _____ number of times.

Mod merujuk kepada nilai dalam siri yang berlaku pada kadar bilangan _____.

- A. maximum / maksima
- B. minimum / minima
- C. infinite / tak terhingga
- D. zero / kosong

9. What will be the value of the following Python expression?

Apakah kemungkinan nilai bagi pernyataan Python berikut?

$4 + 3 \% 5$

- A. 7
- B. 2
- C. 4
- D. 1

10. What will be the output of the following Python function?

Apakah kemungkinan output bagi fungsi Python berikut:

$\min(\max(\text{False}, -3, -4), 2, 7)$

- A. -4
- B. -3
- C. 2
- D. False /Salah

SECTION B: STRUCTURE [60 MARKS]

SEKSYEN B: STRUKTUR [60 MARKAH]

Instruction: Write your answers in the spaces provided in this questions paper.

Arahan: Tuliskan jawapan di ruang yang disediakan di dalam kertas soalan ini.

QUESTION 1 / SOALAN 1

- a) Tick the **CORRECT** answer for the level of data measurement below: [4 Marks]

Tandakan jawapan yang **BETUL** bagi tahap pengukuran data di bawah:

Data <i>Data</i>	Nominal <i>Nominal</i>	Ordinal <i>Ordinal</i>	Interval <i>Selang</i>	Ratio <i>Nisbah</i>
Eg: Hair color <i>Cth: Warna Rambut</i>	/			
Shoes Size <i>Saiz Kasut</i>				
Temperature in Celsius <i>Suhu dalam Celsius</i>				
Team Number <i>Nombor Kumpulan</i>				
Height <i>Ketinggian</i>				

- b) Draw a box plot for the following of set of data: [5 Marks]

Lukis sebuah kotak plot bagi data set berikut:

4.7, 3.8, 3.9, 3.9, 4.6, 4.5, 5

- c) Provide a steam-and-leaf plot for the following numbers: [5 Marks]

Hasilkan sebuah plot batang dan daun bagi nombor-nombor berikut:

2.1, 2.0, 3.2, 3.7, 1.4, 1.5, 2.6

d)

Travel survey:

92 people were asked how they got to work:

92 orang telah disoal bagaimana mereka pergi bekerja:

- 35 used a car / *35 menggunakan kereta*
- 42 took public transport / *42 menaiki pengangkutan awam*
- 8 rode a bicycle / *8 menunggang basikal*
- 7 walked / *7 berjalan kaki*

Based on the travel survey above. Answer the following questions:

Berdasarkan tinjauan perjalanan di atas. Jawab soalan yang berikut:

- i. Calculate the relative frequency [4 Marks]
Kira kekerapan relatif
- ii. Draw an appropriate chart [4 Marks]
Lukis sebuah carta yang bersesuaian

Answer / Jawapan:

QUESTION 2 / SOALAN 2

Class Interval <i>Selang Kelas</i>	Frequency <i>Kekerapan</i>
40.5 - 45.5	7
45.5 - 50.5	10
45.5 - 55.5	15
55.5 - 60.5	2
60.5 - 65.5	6
TOTAL	40

a) Based on the grouped data above. Calculate:

Berdasarkan kumpulan data di atas. Kira:

- i. The mean for the following data [14 Marks]
Min bagi data tersebut
- ii. The median for the following data [14 Marks]
Median data tersebut

Answer / Jawapan:

b) Based on the following data below, use the 18 sorted (left to right) weights of can drinks.

Berdasarkan data di bawah berikut, gunakan sebanyak 18 berat tin minuman yang telah disusun (kiri ke kanan).

0.7901	0.8044	0.8062	0.8073	0.8079	0.8110
0.8126	0.8128	0.8143	0.8150	0.8150	0.8152
0.8152	0.8161	0.8161	0.8163	0.8165	0.8170

Find the percentile correspondent to the given value:

Cari koresponden persentil terhadap nilai yang diberikan:

i. 0.8062

[2 Marks]

ii. 0.8143

[2 Marks]

Find the indicated percentile and quartile.

Cari persentil dan kuartil yang ditunjukkan.

iii. P_{60}

[2 Marks]

iv. Q_3

[2 Marks]

v. P_{40}

[2 Marks]

SECTION C: PYTHON PROGRAMMING [30 MARKS]
SEKSYEN C: PENGATURCARAAN PYTHON [30 MARKAH]

Instruction: Write your answers in the spaces provided in this questions paper.

Arahan: Tuliskan jawapan di ruang yang disediakan di dalam kertas soalan ini.

QUESTION 1/ SOALAN 1

- a) Give an appropriate data type and declare the following variables, which have the following values:
Berikan jenis data yang bersesuaian dan isytiharkan pembolehubah berikut, yang mempunyai nilai-nilai berikut:

[4Marks]

	Variable/ <i>Pembolehubah</i>	Value / <i>Nilai</i>	Data Type / <i>Jenis Data</i>	Variable Declaration / <i>Pengisytiharan Pembolehubah</i>
i.	StudID	ADW2015		
ii.	win_rate	2.5		
iii.	NumOfBooks	200		
iv.	amsRate	False		

- b) Refer to the variables in Q1(a), evaluate each expression in the following table. [3 marks]
Rujuk pembolehubah di Q1(a), nilaikan setiap pernyataan di dalam jadual berikut:

	Expression / <i>Penyataan</i>	Value / <i>Nilai</i>
i.	NumOfBooks * 1.2	
ii.	int(StudID[3:]) // 5	
iii.	int(not amsRate) + 7	

- c) Write a code below / *Tulis kod di bawah:*
- i. Create a NumPy array / *Cipta satu tatasusunan NumPy* [1 Mark]

- ii. Print the even elements in the given array. / Cetak elemen genap dalam tatasusunan yang diberikan. [1 Mark]

```
arr = np.array([1, 2, 3, 4, 5, 6, 7, 8])
```

- iii. Assign the value 20 for the second element in the given array: [1 Mark]

Berikan nilai 20 bagi elemen kedua di dalam tatasusunan yang diberikan:

```
a = np.array([10, 2, 30, 40, 50])
```

QUESTION 2/ SOALAN 2

Question 2(a) –(d) are based on *titanic.csv* file.

Soalan 2(a)-(d) adalah berdasarkan fail titanic.csv

- a) Write a statement to read in titanic data (titanic.csv) into a pandas Data Frame called *titanic*.

Tulis sebuah pernyataan untuk membaca data titanic (titanic.csv) ke dalam Bingkai Data pandas yang bernama titanic. [4 marks]

- b) Print out the first few rows.

Cetak keluar beberapa baris awal. [2 mark]

- c) How to display the data type for each column, by calling 'dtypes' on the entire DataFrame?

Bagaimana untuk papar jenis data bagi setiap lajur dengan memanggil 'dtypes' pada seluruh DataFrame? [2 mark]

- d) How to provide a statistical summary of each column such as such as count, column mean value and column standard deviation?

Bagaimana untuk menyediakan ringakasan statistic bagi setiap lajur seperti kira, lajur nilai min, dan lajur stsisihan piawai? [2 mark]

QUESTION 3/ SOALAN 3

The Woodmill Company makes windows and door trim products. Currently, the company uses the manual process in which an experienced operator quickly looks at a board and determines what rip widths to use. In addition, another company has developed an optical scanner that can be used to determine the rip widths. The scanner is programmed to recognize defects and to determine the rip widths that will optimize the value of the board. A test run of 100 boards was put through the scanner and the rip widths are identified. Next, the same 100 boards were manually ripped using the normal process. The grader then determined the value for each board after the manual rip process was completed. The resulting data, in the file WOODMILL.xlsx, consist of manual rip values and scanner rip values for each of the 100 boards. Compute appropriate descriptive statistics for both manual and scanner values.

Syarikat Woodmill membuat produk hiasan tingkap dan pintu. Pada masa ini, syarikat menggunakan kaedah manual di mana pengendali yang berpengalaman melihat papan dan menentukan berapakah koyakan lebar yang diguna dengan pantas. Disamping itu, syarikat lain telah membangunkan pengimbas optik yang boleh digunakan untuk menentukan lebar koyakan. Pengimbas diprogramkan untuk mengenalpasti kecacatan dan menentukan lebar koyakan yang akan mengoptimumkan nilai papan. Ujian saringan 100 papan telah dimasukkan melalui pengimbas dan lebar koyakan dikenalpasti. Kemudian, 100 papan yang sama, diuji menggunakan kaedah manual. Penggred kemudian menentukan nilai untuk setiap papan selepas proses mengoyak manual selesai. Data yang terhasil, dalam fail WOODMILL.xlsx, terdiri daripada nilai koyakan manual dan nilai koyakan pengimbas untuk setiap 100 papan. Kira deskriptif statistik yang sesuai untuk kedua-dua nilai manual dan pengimbas.

Write a python code based on the following comments below:

Tulis kod python berdasarkan komen di bawah:

- a) Import python library for data frame / *Import perpustakaan python ke dalam bingkai data*

[1 Mark]

- b) Read excel file / *Baca fail excel*

[2 Marks]

- c) Showing the header of data file / *Tunjukkan kepala fail data*

[1 Mark]

- d) Display descriptive statistic information / *Papar maklumat deskriptif statistik*

[1 Mark]

e) Print out the median / *Cetak keluar median* [1 Mark]

f) Calculate coefficient of variation (CV) for scanner
Kira pekali variasi bagi pengimbas [2 Marks]

g) Calculate coefficient of variation (CV) for manual
Kira pekali variasi bagi manual [2 Marks]

END OF QUESTIONS / SOALAN TAMAT