UNIVERSITY OF SWAZILAND

FINAL EXAMINATION PAPER 2017/8

TITLE OF PAPER : STATISTICAL DATA PROCESSING

COURSE CODE : STA206

TIME ALLOWED : TWO (2) HOURS

REQUIREMENTS : CALCULATOR

INSTRUCTIONS : ANSWER ANY THREE (3) QUESTIONS.

Question 1   [20 marks, 3+3+3+3+5+3]

(a) Explain the role of a database administrator

(b) What are the disadvantages of the file system?

(c) What is data independence?

(d) Explain the concept of functional dependency

(e) What are the major components used in E-R diagram design?

(f) What is the role of a data dictionary?

Question 2   [20 marks, 8+4+8]

Refer to the following tables then answer the questions that follow.

Table 1: Publisher

pub_ID	pub_name	address	city	state
736	New Age Books	4 1st Ave.	Boston	MA
877	Binnet&Barney	23 34rd St.	Washington	DC
1389	AlgoDumini	47 9th Ave.	Berkeley	CA

Table 2: Titles

title_ID	title	cat	pub_ID	price
2091	is anger the enemy?	psychology	0736	10.00
2106	life without fear	psychology	0736	7.00
1035	alien	sci-fi	1389	19.99

(a) Show the data output produced when each of the following SQL queries are executed.

- SQL query 1:

```
SELECT pub_ID,pub_name
FROM publisher
WHERE state NOT IN ('MA','IL')
AND pub_name NOT LIKE 'A%'
```

- SQL query 2:

```
SELECT cat ,
AVG(price) AS AveragePrice,
SUM(price + price*0.2) AS TotalPrice
FROM titles
GROUP BY cat;
```

(b) What is the purpose of the HAVING keyword? Use a HAVING clause in query 2 above to show how it can be used to limit the rows in the output.

(c) Discuss the advantages of using stored procedures as a way of implementing database queries.

2

Question 3 [20 marks, 1+1+1+1+1+1+1+1+2+2+2+2+2+2]

(a) We create the following objects in R:

```
> alpha <- c(1,0,2,1,0,1)
> beta <- c(TRUE,FALSE,TRUE)
> gamma <- c(rep(3,times=2),rep(c(2,1),each=2))
> delta <- c("T","F","T")
```

Given the assignments above, what is the output of the following commands?

   (i) `> gamma`

  (ii) `gamma*alpha`

 (iii) `alpha[1:3]+alpha[3:5]`

  (iv) `length(delta):length(gamma)`

   (v) `> as.logical(alpha)==bet`

  (vi) `> delta[gamma]`

 (vii) `> delta[!beta]`

(viii) `> beta|!c(TRUE,TRUE,FALSE)`

  (ix) `> gamma[seq(length(beta),by=-2)] + gamma[-c(3,6)]`

   (x) `> !(gamma<3 & delta=="F")|beta[2:3]`

(b) We have collected data on sea surface temperature and number of hurricanes for the past 30 years (1988 to 2017 inclusive) and stored these data in vectors in R named SST and NOH. Give the R commands to do the following:

   (i) Create a dataframe in R called WarmingDF containing sea surface temperature and number of hurricanes data. Set row names to be the years in which data were collected. Remove the original vectors from the workspace.

  (ii) Make the variables encapsulated in WarmingDF visible from the command line. Calculate the coefficient of variation (mean divided by standard deviation) for sea surface temperature.

 (iii) Generate a scatter plot of hurricane numbers against sea surface temperature with the values for the last ten years marked in a different colour from the other points.

  (iv) Find out if the year with the highest sea surface temperature is the same as the year with the largest number of hurricanes.

Question 4   [20 marks, 12+8]

(a) In the past almost all aspects of data processing were done manually, but in recent times computers have made it possible to process data in an automated manner. With the aid of examples discuss the benefits and drawbacks of automating data processing.

(b) What is coding? Create a coding scheme for the following questionnaire.

SAMPLE QUESTIONS:

Q2.1 Are you an employee or self-employed? (Employed [go to Q2.5] or Self-employed)

Q2.2 Do you have an incorporated business? (Yes or No)

Q2.3 Do you have employees? (Yes or No)

Q2.4 What is the name of your business? [go to Q2.6]

Q2.5 For who do you work?

Q2.6 What kind of business, industry or service is this? (e.g., road maintenance, primary school, rice farm, retail shoe store, garage)

Q2.7 What is your occupation? (e.g., legal secretary, plumber, fishing guide, school teacher)

Q2.8 In this work, what are your main activities or duties?