

UNIVERSITY OF ESWATINI

FINAL EXAMINATION PAPER 2018/2019

TITLE OF PAPER : STATISTICAL DATA PROCESSING

COURSE CODE : STA206

TIME ALLOWED : TWO (2) HOURS

REQUIREMENTS : CALCULATOR

INSTRUCTIONS : THIS PAPER HAS FIVE (5) QUESTIONS. ANSWER ANY THREE (3) QUESTIONS.

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

(a) Consider the following objects that are created in R:

```
> x <- c(1,3,5,7,9,2)
> y <- c(seq(1,5,2),5:3)
> z <- rep(c(3,6), times=2, each=3)
> D <- c(TRUE,FALSE,TRUE)
```

Given the above assignments, depict the output of the following commands?

```
(i) > z
(ii) > z+x
(iii) > x*y
(iv) > c(FALSE,D,TRUE)
(v) > mode(D)
(vi) > array(y,c(2,3))
(vii) > matrix(x,nrow=3,byrow=T)
(viii) > month.abb[y]
(ix) > letters[x]
(x) > data.frame(x,y)
```

(b) Consider the following function:

```
testfn1 <-function(x, y=3, z=TRUE){
  for (i in 1:x){y <- y+1}
  if (z) {y <- y/2}
  y
}
```

Given the function definition above, what would be the output from the following commands:

```
(i) > testfn1(3)
(ii) > testfn1(2,4)
(iii) > testfn1(3,z=FALSE)
(iv) > testfn1(2,z=FALSE,y=1:3)
(v) > testfn1(1<2,2+2==5,0==1)
```

Question 2

[20 marks, 8+6+6]

An online picture sharing company uses a database with the following schema:

Users(uid, uname, city)

Picture(pid, author, size, pdf)

- Users stores all users; uid is the key.
 - Picture stores their pictures; pid is the key; author is the uid of the picture's author; size represents the size of the picture in bytes; pdf is the actual pdf content of the picture.
 - uid, pid, author, size are integers; unname, city, pdf are text.
- (a) Write the SQL statements to create the tables for this database.
- (b) Write a SQL query that returns all users that have posted both a picture larger than 1MB (size > 1000000) and a picture smaller than 1MB. Your query should return the users' uid and unname
- (c) Write a SQL query that retrieves all users who do not have any picture greater than 1MB (size > 1000000). Your query should return the users' uid and unname

Question 3

[20 marks, 8+4+6+2]

- (a) Explain what the term data validation means. Using your own examples, describe the various data validation techniques that may be embedded into a forms-based interface to a database - for example, ensuring that the correct type and range of data values are entered.
- (b) Describe, with an example, one type of problem that can occur in a multi-user environment when concurrent access to the database is allowed.
- (c) Backups of the database should be taken in order to protect data. Describe five measures that can be taken in order to ensure the security and effectiveness of database backups.
- (d) An important concept in the theory of relational databases is that of a *functional dependency*. Explain what is meant by a functional dependency and give an example.

Question 4

[20 marks, 4+8+8]

- (a) When should the raw data from a respondent be altered by a data editor?
- (b) What is the purpose of editing? Provide two examples of questions that might need editing.
- (c) Summarize procedures for detecting errors and omissions from survey data. How can these errors be corrected.

Question 5

[20 marks, 4+6+2+8]

Consider the following examples:

- (1) Opinion on the quality of a movie expressed freely by a group of teenagers.
- (2) Degree classification of graduating students from the Faculty of Social Sciences.
- (3) Brand of toothpaste.
- (4) Income of workers at a textile factory.

- (5) Amount of calcium deposits (in milligrams) in the organs of rats that have been subjected to different experimental treatments.
- (a) List the name of the variable you can extract from each of the examples.
 - (b) List some of all the possible values of the variable extracted from each of the examples.
 - (c) State the measurement scale to use for those variables.
 - (d) Prepare a code-book for the above 5 variables.