MASENO UNIVERSITY

UNIVERSITY EXAMINATIONS 2013/2014

SECOND YEAR SECOND SEMESTER EXAMINATIONS FOR THE DEGREE OF BACHELOR OF SCIENCE WITH INFORMATION TECHNOLOGY

(MAIN CAMPUS)

MIT 202: DATA MANAGEMENT

Date: *9th April, 2014* ***Time:*** *2.45 – 5.00pm*

INSTRUCTIONS:

- Attempt QUESTION ONE and ANY OTHER TWO questions.

Instructions: Attempt Question ONE and any other TWO Questions.

QUESTION 1 [30 Marks] – Compulsory

a) Suppose you have just been employed as a statistician/data manager in a research organisation. From your transcripts they notice the course on data management and ask you to write a brief report explaining the data flow concept and illustrating how it might apply to research projects. In your report you should describe the concept, illustrate the steps in a diagram, explain where quality control fits and relate theory to practical application in a research project. [10 Marks]

b) Describe data management and give two reasons on why it is so important to plan for data management even before data is collected. [4 Marks]

c) What is metadata? What is its role in data management? [4 Marks]

d) It is advisable to keep a master copy of data apart from a working copy. Explain what is meant by a 'master copy' and give the reason for creating it. [4 Marks]

e) Explain the meaning of a 'data entry system' and discuss at what stage of the data process in a study it would be ideal to have it ready. [4 Marks]

f) There are numerous softwares that can be used for data entry as discussed in this course. Pick one of your choices and discuss its pros and cons for data entry tasks. [4 Marks]

QUESTION 2 [20 Marks] - Optional

Suppose you are part of a team involved in a study which needs to be completed as quickly as possible. The study is being designed and then a pilot is being conducted before the main team goes out to collect the data.

a) At what stage should you start thinking about the data entry? [3 Marks]

b) What are the different options that you should consider? [3 Marks]

c) Under what circumstances might you choose each of the different data entry options? [3 Marks]

d) Discuss any 3 questions that are important to ask yourself when planning for data entry? [6 Marks]

e) Choose one of the data entry options and describe what you can do to maintain high quality through the data entry process. Give specific software specific tips that you would recommend. [5 Marks]

QUESTION 3 [20 Marks] - Optional

The task of preparing data for analysis can often require the transfer of data between packages.

a) What shape data does a statistics package expect? Why does a statistics package expect data in this shape? [5 Marks]

b) To the right is some data in Excel which has been prepared for analysis using the Excel one way ANOVA. The data represents sales figure for a book in various shops where the book was either displayed at the front back or in the middle of the shop.

One-Way ANOVA		
Front	Back	Middle
7	12	10
10	13	11
8	15	12
9	16	
11		

i. What form would the data need to be in for an ANOVA in a statistics package? [2 Marks]

ii. State two advantages and two disadvantages of the form that you have described above. [4 Marks]

iii. How could you go about transforming the data from the form given above to the one usually used in a statistics package? [2 Marks]

c) Part of preparing data for analysis is checking for errors. Give examples of common errors and how you can go about checking for them and resolving them. Explain how this might relate to the transfer of data to a statistics package. [7 Marks]

QUESTION 4 [20 Marks] - Optional

Suppose you are working on a research project where the lead research is a senior professor at a university who overseas a team of research assistants who are collecting the data. The project is jointly funded by the government and a donor.

a) Discuss potential data ownership issues that might affect such a project. [4 Marks]

b) At what stage should such issues be resolved? [2 Marks]

c) The research assistants all contribute some analysis towards a joint paper with the lead researcher. At the end of the study the journal is demanding that the lead researcher make the data available to reviewers so that results can be verified. Discuss potential difficulties that could arise due to poor planning and good practice guidelines that could help avoid them. [5 Marks]

d) At the end of this project all parties agree in principle to make the data publicly available. You have been given the task of suggesting how this might be practically achieved. Include details of any software/sites that you might use along with the information you will be requesting from all parties. [9 Marks]

QUESTION 5 [20 Marks] - Optional

One of the aims of the course was to provide you with some of the skills needed to take a data set however badly organised and salvage it as much as possible. This should also provide you with the motivation to learn how to manage data well so as to avoid having to deal with messy data.

a) In the course and your previous experience you have had example of difficulties caused by data entered in ways that are not conducive to analysis. List some of the key problems along with the skills needed to resolve them. [7 Marks]

b) Consider the statement:

'It is more useful to become master of a single software package, irrespective of which package it is than being a user of many.'

If you are often presented with messy data explain ways in which the above statement is true. [7 Marks]

c) You are part of a large research team which has just been awarded a grant for a research project. As a data manager you have convinced the team of the importance of avoiding messy data. You have been asked to write a brief document for the whole team that will provide guidelines which if followed will make it easy to manage the projects data. [6 Marks]