



(Knowledge for Development)
KIBABII UNIVERSITY

**UNIVERSITY EXAMINATIONS
2016/2017 ACADEMIC YEAR**

**SPECIAL/SUPPLEMENTARY EXAMINATIONS
YEAR ONE EXAMINATIONS**

**FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN INFORMATION
TECHNOLOGY**

COURSE CODE : PIT 914

**COURSE TITLE : ADVANCED DATA ANALYSIS
ELECTRONICS**

DATE: 30/09/2017 TIME: 2.00P.M. – 4.00.M.

INSTRUCTIONS TO CANDIDATES

ANSWER QUESTIONS ONE AND ANY OTHER TWO.

QUESTION ONE (30 MARKS)

- a) Discuss the relations among the following technologies. Data Mining, Business Intelligence, Data Analytics and Big Data. (5 Marks)
- b) Big Data has gained much attention from the academia and the IT industry. In the digital and computing world, information is generated and collected at a rate that rapidly exceeds the boundary range. However, the fast growth rate of such large data generates numerous challenges. Explain these challenges. (5 Marks)
- c) In reality, data analysis is a highly iterative and non-linear process, better reflected by a series of epicycles, in which information is learned at each step, which then informs whether (and how) to refine, and redo, the step that was just performed, or whether (and how) to proceed to the next step. Discuss the 5 core activities of data analysis as captured in the epicycle: (10 Marks)
- d) What are the goals of formal modeling as relates to data analysis? Distinguish between a primary and secondary models? (10 Marks)

QUESTION TWO (20 MARKS)

If we follow the variation of some quantity over time, we are dealing with a *time series*. Time series are incredibly common: examples range from stock market movements to the tiny icon that constantly displays the cpu utilization of your desktop computer for the previous 10 seconds. What makes time series so common and so important is that they allow us to see not only a single quantity by itself but at the same time give us the typical “context” for this quantity. Because we have not only a single value but a bit of history as well, we can recognize any changes from the typical behavior particularly easily.

On the face of it, time-series analysis is a bivariate problem. Discuss the various methods that have been developed to deal with this type of data.

QUESTION THREE (20 MARKS)

As soon as we are dealing with more than two variables simultaneously, things become much more complicated—in particular, graphical methods quickly become impractical.

Discuss the various graphical methods that can be applied to multivariate problems for which the number of variables is not *too* large

QUESTION FOUR (20 MARKS)

Write short notes on the following data analysis domains.

- Guestimations
- Probability models
- Data mining
- Big Data

e. Data Analytics

QUESTION FIVE (20 MARKS)

When dealing with univariate data, we are usually mostly concerned with the overall *shape* of the distribution. Discuss some of the initial questions we may ask.

When we are dealing with a data set that consists of *two* variables (that is, a *bivariate* data set), we are mostly interested in seeing whether some kind of relationship exists between the two variables and, if so, what kind of relationship this is. Plotting one variable against another is pretty straightforward.

Discuss some of the tools and transformations that can be applied to characterize the nature of the relationship between the two inputs.