Title (Units): **COMP2865 Fundamental of Data Analysis and Management (3,2,1)**

Course Aims: This course introduces the basics of data and its analysis. In addition, the course offers an introduction to probabilities and statistics to enhance student capabilities to reason and report correct stories from data.

Prerequisite: General Education - Quantitative Reasoning (GFQR)

Anti-requisite: COMP2016 Database Management or MATH2206 Probability and Statistics or MATH2005 Calculus, Probability, and Statistics for Computer Science or MATH2006 Calculus, Probability, and Statistics for Science

Course Intended Learning Outcomes (CILOs):
Upon successful completion of this course, students should be able to:

No.	Course Intended Learning Outcomes (CILOs)
	Knowledge
1	Describe probabilities and statistics for data analysis
2	Describe fundamental of data management and analysis
	Professional Skill
3	Perform probabilistic and statistical analysis
4	Formulate SQL queries to a relational database
	Attitude
5	Report stories that are correctly derived from data

Calendar Description: This course introduces the basics of data and its analysis, and statistics that are useful to report stories from data. Topics include: characteristics of different types of data, data management, SQL, probabilities and statistics for data analysis.

Teaching and Learning Activities (TLAs):

CILOs	Type of TLA
1-2, 5	Students will learn the fundamental of data management, probabilities and statistical concepts via lectures, tutorials, and assignments.
3-5	Students will gain practical experience on a database management system via laboratory and tutorial session.

Assessment:

No.	Assessment Methods	Weighting	CILOs to be addressed	Description of Assessment Tasks
1	Assignments	35%	1 - 5	Assignments are designed to measure how well students have learned the fundamentals of data analysis and management and probabilities and statistical analysis.
2	Tutorial	10%	1 - 5	Students are required to submit tutorial worksheets during the lectures. The tutorial worksheets are designed to help students apply the knowledge to solve problems.
3	Laboratory exercise	15%	1 - 5	Three laboratory sessions are designed to provide students practical skills to perform statistical analysis and data management. Students are required to submit three exercises following the laboratory sessions.
4	Examination	40%	1 - 5	Final examination questions are designed to see how far students have achieved their intended learning outcomes. Questions will primarily be knowledge and skills to assess the student's ability

				in data analysis and management, and perform statistical analysis.

Assessment Rubrics:

Excellent (A)	• Achieve the five CILOs, demonstrating a mastery of both fundamental of data management, probabilities and statistical concepts for data analysis • Able to create, populate, update a relational database in SQL language and formulate database queries in SQL • Able to perform probabilities and statistical analysis
Good (B)	• Achieve the five CILOs, demonstrating a good understanding of both fundamental of data management, probabilities and statistical concepts for data analysis • Able to create, populate, update a relational database in SQL language and formulate database queries in SQL • Able to perform probabilities and statistical analysis
Satisfactory (C)	• Achieve the five CILOs, demonstrating a basic understanding of both fundamental of data management, probabilities and statistical concepts for data analysis • Able to create, populate, update a relational database in SQL language and formulate database queries in SQL for common applications • Able to perform probabilities and statistical analysis for common scenarios
Marginal Pass (D)	• Achieve the five CILOs, demonstrating a minimal understanding of both fundamental of data management, probabilities and statistical concepts for data analysis • Able to create, populate, update a relational database in SQL language and formulate database queries in SQL for simple applications • Able to perform probabilities and statistical analysis for limited scenarios
Fail (F)	• Achieve the five CILOs, demonstrating little understanding of both fundamental of data management, probabilities and statistical concepts for data analysis • Unable to create, populate, update a relational database in SQL language and formulate database queries in SQL • Unable to perform probabilities and statistical analysis

Course Content and CILOs Mapping:

Content		**CILO No.**
I	Introduction to probabilities	1,3,5
II	Introduction to statistical analysis	1,3,5
III	Fundamental of data management	2,4,5
IV	Applications of data management	2,4,5

References:

- Roxy Peck, Chris Olsen, and Jay L. Devore, Introduction to Statistics and Data Analysis, 5th Edition, Cengage Learning, 2014.
- Ramez Elmasri and Shamkant B. Navathe, Fundamentals of Database Systems, 7th Edition, Addison Wesley, 2015.
- David M. Levine, Kathryn A. Szabat, and David F. Stephan, Business Statistics: A First Course, 7th Edition, Pearson, 2015.
- Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying E. Ye, Probability and Statistics for Engineers and Scientists, 9th Edition, Pearson, 2011.
- Robert Kabacoff, R in Action: Data Analysis and Graphics with R, 2nd Edition, Manning Publications, 2015.

Course Content:

Topic

I. Introduction to probabilities
   A. Mathematical models and notations
   B. Conditional probabilities
   C. Bayes's theorem
   D. Random variables

II. Introduction to statistical analysis
   A. Parametric distributions
   B. Null hypothesis testing
   C. Significance testing
   D. Confidence interval

III. Fundamental of data management
   A. Database system concepts
   B. Relational data model
   C. Queries and updates in SQL
   D. Concepts of NoSQL database systems

IV. Applications of data management
   A. Identifying unreliable data and creating clean datasets
   B. Validating hypothesis testing result with SQL