# STATISTICAL DATA MANAGEMENT

**UNIT CODE:** MATH/CU/AS/CC/05/6/A

**Relationship to Occupational Standards**

**This unit addresses the unit of competency:** Manage Statistical Data

**Duration of Unit: 140** hours

**Unit Description:**

This unit specifies competencies required to manage database system. They include data management using excel, R, SPSS and Python.

**Summary of Learning Outcomes:**

1. Data management using excel
2. Data management using R
3. Data management using SPSS
4. Manage statistical data on Python

**Learning Outcomes, Content and Suggested Assessment Methods**

| Learning Outcome | Content | Suggested Assessment Method |
|---|---|---|
| 1. Manage statistical data on excel spreadsheet | • Excel Environment<br>• Worksheets<br>• Workbooks<br>• Data labelling, coding and entry<br>• validation<br>• Multiple-key sorting<br>• Sorting of data based on custom lists<br>• creating single- and multi-level subtotals<br>• Filtering of data using text, numeric, date<br>• Filtering of tables using slicers<br>• Advanced Filter<br>• eliminating duplicate | • Practical exercises<br>• Oral questioning<br>• Written test<br>• Learner portfolio of evidence. |

| | | |
|---|---|---|
| | • Use of SUMIF and related functions for quick data analysis<br>• of Index & Match<br>• Conditional Formatting<br>• Filtering & Sorting<br>• Find & Replace<br><br>• Data Analysis in Excel<br>  • Descriptive statistics<br>  • Correlation & Covariance<br>  • ANOVA<br>  • Regression<br>  • T-test & Z-test<br>• Random numbers<br>• Data Presentation<br>  • Pivot Table & Charts<br>• CSV conversion | |
| 2. Manage statistical data on R | • Installing R and R studio<br>• Getting started with R<br>• Data structures in R<br>• Data entry in R<br>  • Arrays<br>  • Data frames<br>  • Lists<br>  • Vectors<br>  • Matrices<br>• Creating R projects<br>• Importing data into R<br>• Installing R packages<br>• Data manipulation in R<br>  • Sorting<br>  • Merging<br>  • Aggregating<br>  • Creating new variables<br>  • Indexing<br>  • Sub setting<br>  • Exporting<br>• Exploratory data analysis<br>  • Scatter plot<br>  • Line graphs | • Practical exercises<br>• Oral questioning<br>• Written test<br>• Learner portfolio of evidence. |

| | Histogram | |
|---|---|---|
| | • Density plot | |
| | • Pie charts | |
| | • Bar charts | |
| | • Box plot etc. | |
| | • Descriptive statistics | |
| |    • Mean | |
| |    • Mode | |
| |    • Median | |
| |    • Dispersion | |
| | • Statistical inference | |
| |    • Regression analysis | |
| | • Time series analysis in R | |
| | • Probability distribution in R | |
| | • Random numbers | |
| | • R commander | |
| | • Built-in functions in R | |
| 3. Manage statistical data on SPSS | • Installing SPSS<br>• SPSS Environment<br>  • Data views<br>  • Variable views<br>  • Output Window<br>• Data transformations<br>• Creation of variable & data coding<br>• Data entry<br>• SPSS syntax<br><br>• Data Analysis in SPSS<br>  • Descriptive statistics<br>     • Mean<br>     • Frequencies<br>     • Cumulative Frequencies<br>  • Pearson Correlation & Covariance<br>  • ANOVA<br>  • Regression<br>  • T-test & Z-test<br>• Random numbers | • Practical exercises<br>• Oral questioning<br>• Written test<br>• Learner portfolio of evidence. |

| | • Data Presentation <br>     • Table & Charts | |
|---|---|---|
| 4. Manage statistical data on Python | • Python Basics <br>     • Running Python <br>     • Literals <br>     • Python Comments <br>     • Data Types <br>     • Variables <br>     • Writing a Python Module <br>     • print () Function <br>     • Named Arguments <br>     • Collecting User Input <br>     • Getting Help <br> • Functions and Modules <br>     • Defining Functions <br>     • Variable Scope <br>     • Global Variables <br>     • Function Parameters <br>     • Returning Values <br>     • Importing Modules <br> • Math <br>     • Arithmetic Operators <br>     • Modulus and Floor Division <br>     • Assignment Operators <br>     • Built-in Math Functions <br>     • The math Module <br>     • The random Module <br>     • Seeding <br> • Python Strings <br>     • Quotation Marks and Special Characters <br>     • String Indexing <br>     • Slicing Strings <br>     • Concatenation and Repetition <br>     • Common String Methods <br>     • String Formatting <br>     • Built-in String Functions | • Practical exercises <br> • Oral questioning <br> • Written test <br> • Learner portfolio of evidence. |

|  |  |  |
|---|---|---|
|  | <ul><li>Sequences, Dictionaries, and Sets<ul><li>Definitions</li><li>Sequences</li><li>Unpacking Sequences</li><li>Dictionaries</li><li>The Len () Function</li><li>Sets</li><li>*args and **kwargs</li></ul></li><li>Flow Control<ul><li>Conditional Statements</li><li>The is and is not Operators</li><li>Python's Ternary Operator</li><li>Loops in Python</li><li>The enumerate() Function</li><li>Generators</li><li>List Comprehensions</li></ul></li><li>File Processing<ul><li>Opening Files</li><li>The os and os.path Modules</li></ul></li><li>Exception Handling<ul><li>Wildcard except Clauses</li><li>Getting Information on Exceptions</li><li>The else Clause</li><li>The finally Clause</li><li>Using Exceptions for Flow Control</li><li>Exception Hierarchy</li></ul></li><li>Dates and Times<ul><li>Understanding Time</li><li>The time Module</li><li>The date-time Module</li></ul></li><li>Running Python Scripts from the Command Line<ul><li>The sys Module</li><li>sys.argv</li></ul></li></ul> |  |

**Suggested Methods of Instructions**
- Projects
- Demonstration by trainer
- Practice by the trainee
- Discussions
- Direct instruction

**Recommended Resources and equipment**
- Computer
- Internet connection
- Stationary
- Printer
- Internet
- Notes
- Data sets
- SPSS
- R
- Python
- Projector