Course: Introduction to Linux and Shell scripting

# Target audience:

Undergraduate students, no prior working experience in Linux environment

# Teaching Objectives:

**COK**. have the profound knowledge of the fundamental mathematical logic and its application in programming

**COA**. have the capability to generalize the simple programming problem in Unix environment

**COC**. have the ability to recognize and solve the simple mathematical problems by using programming with Unix scripting

# Teaching Points

1. Introduction to Linux operating system

2. Basic command line interface

3. Linux utilities and text editor

4. Shell scripting (variable, array and expression)

5. Shell scripting (flow control and repetition)

6. Shell tricks for one-liner bioinformatics I

7. Shell tricks for one-liner bioinformatics II.

# Teaching Materials and Approaches

* Slides: Interactive, with step-by-step examples especially for scripting tasks.
* Exercises: Practical, based on real-world bioinformatics data.
* Tests: A mix of conceptual questions and practical scripting tasks.
* Assignments: Part 1 focuses on shell scripting, while Part 2 revolves around basic bioinformatics tasks, aligning with the students' background in Statistics and Bioinformatics.

This course design ensures a hands-on approach, allowing students to apply mathematical logic and general programming problem-solving in a Unix environment, particularly in the context of bioinformatics. The progressive complexity ensures that by the end of the course, students will be well-equipped to tackle basic bioinformatics tasks using Linux and Bash scripting.

# Short description of sessions

## Session 1: Introduction to Linux and Basic Commands

***Date: Friday 17/11.***

Time: 19 - 21 Indonesian time.

## Session 2: Introduction to Shell Scripting

***Date: 24/11***

Time:19 - 21 Indonesian time.

## Session 3: Advanced Shell Scripting and Bioinformatics Examples

***Date: Monday 27/11***

Time:19 - 21 Indonesian time.

## Session 4: Advanced Shell Scripting and Bioinformatics Examples

Date: **TBD**.

Time:19 - 21 Indonesian time.

# Extended version of session descriptions

## Session 1: Introduction to Linux and Basic Commands

Objective: Familiarize students with the Linux OS and basic command-line interface.

* ***Introduction to Linux OS (45 minutes)***
  + Overview of Linux in statistical and bioinformatics computing
  + Navigating the file system (ls, cd, mkdir, etc.)
* ***Basic Command Line Interface (45 minutes)***
  + Introduction to the command line
  + Basic file operations (copying, moving, deleting files)
* ***Linux Utilities and Text Editors (45 minutes)***
  + Using utilities like grep, find, and sort
  + Introduction to text editors (nano, vi)
* ***Hands-On Exercise (45 minutes)***
  + Basic command line exercises
  + Editing files using text editors
* ***Assignment: Shell Scripting Basics***
  + Create a basic shell script involving file operations

## Session 2: Introduction to Shell Scripting

Objective: Teach students shell scripting basics and their applications in problem-solving.

* ***Shell Scripting: Variables, Arrays, and Expressions (60 minutes)***
  + Understanding variables and arrays in Bash
  + Basic expressions and operations
* ***Flow Control and Repetition (60 minutes)***
  + Conditional statements (if-else, case)
  + Loops (for, while, until)
* ***Hands-On Exercise (60 minutes)***
  + Writing scripts using variables, arrays, and flow control
* ***Assignment: Basic Bioinformatics Workflow***
  + Write a script to perform simple file manipulations relevant to bioinformatics

## Session 3: Advanced Shell Scripting and Bioinformatics I

Objective: Delve into more complex scripting techniques with a focus on bioinformatics applications.

* ***Advanced Shell Scripting Techniques (60 minutes)***
  + Regular expressions and advanced text manipulation
  + Scripting for file and data processing
* ***Introduction to Bioinformatics Scripting I (60 minutes)***
  + Accessing online databases and retrieving datasets
  + Basic FASTQ file operations: quality control and checks
* ***Hands-On Exercise and Mini-Project (60 minutes)***
  + Practical scripting tasks in bioinformatics
* ***Assignment: Bioinformatics Scripting Task***
  + Automate a sequence of bioinformatics tasks (e.g., FASTQ file processing)

## Session 4: Advanced Shell Scripting and Bioinformatics II

Objective: Apply learned skills in advanced bioinformatics scripting and consolidate knowledge.

* ***Advanced Bioinformatics Scripting II (60 minutes)***
  + FASTQ file trimming and simple analysis
  + Integrating scripts with bioinformatics tools
* ***Project Work and Review (90 minutes)***
  + Students work on a comprehensive scripting project
  + Review and discuss various approaches
* ***Final Assessment and Course Wrap-up (30 minutes)***
  + Further Learning Topics: Discussion on additional topics such as Git, R package development, in-depth bioinformatics, and machine learning, to guide students on their learning path beyond the course.
  + Feedback and course summary

# Teaching Materials and Approaches

* Slides: Interactive, with step-by-step examples especially for scripting tasks.
* Exercises: Practical, based on real-world bioinformatics data.
* Tests: A mix of conceptual questions and practical scripting tasks.
* Assignments: Part 1 focuses on shell scripting, while Part 2 revolves around basic bioinformatics tasks, aligning with the students' background in Statistics and Bioinformatics.

This course design ensures a hands-on approach, allowing students to apply mathematical logic and general programming problem-solving in a Unix environment, particularly in the context of bioinformatics. The progressive complexity ensures that by the end of the course, students will be well-equipped to tackle basic bioinformatics tasks using Linux and Bash scripting.

Note:

Questions to clarify:

1. Do I need to show students how to install Linux?
   1. No. Linux is installed.
2. Do you have any common tasks of data analysis ?
   1. Bioinformatics in general.
   2. FASTQ file: Quality control, check, trimming.
   3. Get dataset from online databases -> QC and preprop -> Simple analysis.
3. Do you have any commonly used programs that you would like me to show you how to do in Linux?
   1. No preference. → Will use some qc program for fastq.
4. What is students’ major? Statistics or Mathematics, or Bioinformatics?
   1. Undergrad. Major: Statistics. Minor: Bioinformatics.