**Health and Biomedical Data Analysis**

2025 Spring Semester

**Lectures**: Wednesday 9:00 –11:50, PB1007

**Instructor**: 김선영 (Sun-Young Kim), 김준태 (Juntae Kim)

* Email: [sykim@ncc.re.kr](mailto:sykim@ncc.re.kr)
* Phone: 031-920-2745
* Office: PB1006-3

**Class website**: GCSP Learning Management System

* Lecture notes and home assignments will be posted
* Questions & answers and discussion are available

**Lecture notes**

* Notes are posted on the class website

**Software**: R and Phyton

**Office hour**: Tuesday 10:30 – 12:30

**Grading**:

* Homework and presentation: 50 %
* Final project: 50%

**Class objectives**

This course will provide practical knowledge of different analysis approaches using real-world data. During the course, students will learn the application of statistical methods and machine learning algorithm to the same data. This application will provide the knowledge for data processing, analysis techniques, and interpretation of findings as well as the understanding of the differences between commonly-applied and different analysis approaches.

**Class schedule**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Week | Date | Topic | Lab | Homework |
| 1 | 09/05 (Tue) | Introduction |  |  |
| 2 | 09/22 (Fri) | Statistical methods | О |  |
| 3 | 09/27 (Wed) | Machine learning algorithm | О |  |
| 4 | 10/04 (Wed) | Overview of research question for real-world data 1  Review of previous studies (student presentation) |  | О |
| 5 | 10/11 (Wed) | Real-world data 1: statistical method (1) | О | О |
| 6 | 10/18 (Wed) | Real-world data 1: statistical method (2) | О | О |
| 7 | 10/25 (Wed) | Real-world data 1: machine learning algorithm (1) | О | О |
| 8 | 11/01 (Wed) | Real-world data 1: machine learning algorithm (2) | О | О |
| 9 | 11/08 (Wed) | Overview of research question for real-word data 2  Review of previous studies (student presentation) |  | О |
| 10 | 11/15 (Wed) | Real-world data 2: statistical method (1) | О | О |
| 11 | 11/22 (Wed) | Mid-term: final project proposal presentation |  |  |
| 12 | 11/29 (Wed) | Real-world data 2: statistical method (2) | О | О |
| 13 | 12/06 (Wed) | Real-world data 2: machine learning algorithm (1) | О | О |
| 14 | 12/13 (Wed) | Real-world data 2: machine learning algorithm (2) | О | О |
| 15 | 12/20 (Wed) | Final project presentation |  |  |

\* Real-world data 1: air pollution prediction

Real-world data2: cancer prediction

* Two sets of real-word data and analysis
  + Air pollution prediction
    - Data
      * Hourly measurements of fine particle at 300 monitoring sites in 2010, South Korea
      * > 300 geographic variables
    - Analysis approach
      * Statistical method: university kriging with partial least squares
      * Machine learning: ridge regression and lasso
    - Reference: Jin et al. in preparation
  + Cancer prognosis prediction
    - Data
      * Cancer genomics data in the Cancer Genome Atlas Program (TCGA)
      * Mortality
    - Analysis approach
      * Statistical method: survival analysis
      * Machine learning: ridge regression and lasso
    - Reference:
* Final project: data analysis using statistical and machine learning approaches using student’ own dataset