2020 Fall "Biostats and Big Data 2" at SKKU GBME

- Lecturer: Choong-Wan Woo, Ph.D. Assistant professor (GBME).
- Office: N-center, 86335
- Web: Cocoan lab
- E-mail: choongwan.woo@gmail.com

- Class: Tue 1:30-2:45, Thu 12:00-1:15 at TBA (possibly, the computer room)
- Office hours: Wed 10:00-12:00, you can book a time in advance through https://choongwanwoo.youcanbook.me

Download

You can download the class materials using the following command line.

```
$ git clone https://github.com/wanirepo/Stats2_2020Fall
```

Once you clone the github repository, you can just type the following command to get the updated github repository.

```
$ git pull
```

Or you can download the repository as a zip file or you can also use GitHub Desktop. The class materials will be uploaded (e.g., lecture slides, assignments) before each class.

There is a good github tutorial: https://rogerdudler.github.io/git-guide/index.html

What are the aims of this course?

Data are everywhere. It is a big data era. Data science already became a key element in many research and industrial areas. The primary aim of this course is to learn basic concepts and skills for data analysis, including concepts of random variables, sampling distributions, hypothesis testing, linear modeling, data visualization, etc., preparing students for their lives

after the graduation in this data-are-everywhere age. This class, “Biostats and Big data 2”, is the advanced version of the “Biostats and Big data” course. You will also get some hands-on experience and in-depth study materials for statistics. If you did not finish the “Biostats and Big data”, I recommend you attending the class first.

Course format (flipped classroom) and expectations

This class uses a flipped classroom format, which is a new way of teaching and learning. Different from the traditional learning environment (passively listening to the lecture in the class and doing homework at home), in the flipped classroom, you will listen to the lecture at home and do homework and practices in the classroom. I personally experienced this format of learning during my PhD (for the Machine Learning class) and deeply enjoyed it. I found the flipped classroom helped students stay engaged and provided a good environment for hands-on experience. For these reasons, I have wanted to do the class with the flipped classroom format.

This semester is the first semester I open this class. Therefore, class materials are not going to be perfect, and I might not be able to provide videos for all weeks. Please consider you as a co-creator of this class. We will make and shape the class together.

Potential impacts in the course format due to the Covid-19

As you are all aware of, all the classes have been impacted by Covid-19. I will try to open this class as off-line, if possible, but it really depends on the number of students. I will keep you posted!

Textbooks

Main textbook:

> "Stats: Data and Models" by De Veaux, Velleman, and Bock
> "Statistical Thinking for the 21st Century" by Russell Poldrack Link

Softwares

I will use Matlab, R, and two free software packages for statistical analysis, JAMOVI and JASP. You can download Matlab through SKKU. R is an open-source programming

language.

TAs

- TBA

Evaluation

Absolute evaluation will be used for this course.

1. Attendance (40%)
2. Participation (including pop-questions) (35%)
3. Midterm exam (10%)
4. Final exam (15%)

Schedule (TBA)

Week	Video lectures	Class	Chapters
Week 1			
9/1		Overview	
9/3		Softwares and programming languages	
Week 2			
9/8		Descriptive stats - analysis and report	
9/10		Data visualization 1	
Week 3			
9/15		Describing distribution (Data visualization 2)	
9/17		Comparing distribution 1 - Histogram, box plot	
Week 4			
		Comparing distribution 2 - scatter plot, and KL	

9/22		divergence, Q-Q plot	
9/24		Sampling - standard error of the mean, confidence interval	
Week			
9/29		추석	
10/1		추석	
Week 5			
10/6		Resampling 1 - bootstrap test	
10/8		One sample t-test and NHST	
Week 6			
10/13		Resampling 2 - permutation test	
10/15		Paired t-test, two-sample t-test	
Week 7			
10/20		Chi-square test	
10/22		Correlation	
Week 8			
10/27		mid-term	
10/29		mid-term	
Week 9			
11/3		Multiple comparisons correction, FWER (bon ferroni), FDR	
11/5		Power calculation	
Week 10			

11/10		ANOVA - basics	
11/12		ANOVA - advanced	
Week 11			
11/17*		Multiple regression (basics) - basic stats, R-squared	
11/19		Multiple regression (advanced 1) - Dummy coding, etc.	
Week 12			
11/24		Multiple regression (advanced 2) - hierarchical, additional R-squared	
11/26		Cross-validation and independent model testing	
Week 13			
12/1		Multivariate methods 1 (PCA)	
12/3		Multivariate methods 2 (PLS)	
Week 14			
12/8		Multivariate methods 3 (or network modeling)	
12/10		Review	
Week 15			
12/15		final	
12/17		final	

Note.

* Potentially make-up due to a conference