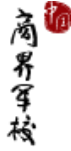




# PHBS

北京大学汇丰商学院



## Course Code Applied Data Analytics Module 1, 2021

### Course Information

---

**Instructor: Shubin Yu**

Office: PHBS Building, Room 631

Phone: 86-755-2603-5185

Email: Shubin.Yu@phbs.pku.edu.cn

Office Hour: 14:00-17:00 (Tue, Fri)

**Teaching Assistant: Rongxin Ouyang**

Phone: 86-132-6197-7913

Email: rey@pku.edu.cn

**Classes:**

Lectures: Tue & Fri 10:30-12:20

Venue: PHBS Building, Room

**Course Website:**

If any.

## 1. Course Description

### 1.1 Context

Course overview: Data is regarded as the new currency in the 21<sup>st</sup> Century. This course aims to help students use appropriate research methods and data science techniques to answer research questions and solve real-world problems. Going beyond theory, students will learn how to collect data through experiments and surveys. Students will also learn how to clean, prepare, and analyse the data using R studio. The last decade has seen a significant increase in the number of analysis approaches. This course will include new approaches such as Moderated Mediation Analysis, Structural Equation Modelling, Decision Trees, Random Forest, and Sentiment Analysis. This course will cover these approaches in the context of Communication Sciences: Social Media, Marketing Communications, and other important research and business decisions. Upon completion of this course, R studio will become your best partner in data analytics.

Prerequisites: No prior knowledge is required. However, students are advised to first complete the course: *Methodologies for Social Science*.

### 1.2 Textbooks and Reading Materials

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Second edition). New York: Guilford Press.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning*. Springer.

De Jonge, E., & Van Der Loo, M. (2013). *An introduction to data cleaning with R*. Statistics Netherlands Heerlen.

## 2. Learning Outcomes

### 2.1 Intended Learning Outcomes

Learning Goals	Objectives	Assessment (YES with details or NO)
1. Our graduates will be effective communicators.	1.1. Our students will produce quality business and research-oriented documents.	Yes, students are required to produce a final report.
	1.2. Students are able to professionally present their ideas and also logically explain and defend their argument.	Yes, students are expected to actively participate in the class discussion and present their group project.
2. Our graduates will be skilled in team work and leadership.	2.1. Students will be able to lead and participate in group for projects, discussion, and presentation.	Yes
	2.2. Students will be able to apply leadership theories and related skills.	Yes
3. Our graduates will be trained in ethics.	3.1. In a case setting, students will use appropriate techniques to analyze business problems and identify the ethical aspects, provide a solution and defend it.	Yes
	3.2. Our students will practice ethics in the duration of the program.	Yes
4. Our graduates will have a global perspective.	4.1. Students will have an international exposure.	Yes
5. Our graduates will be skilled in problem-solving and critical thinking.	5.1. Our students will have a good understanding of fundamental theories in their fields.	Yes
	5.2. Our students will be prepared to face problems in various business settings and find solutions.	Yes
	5.3. Our students will demonstrate competency in critical thinking.	Yes

### 2.2 Course specific objectives

Upon successful completion of this course, students will be able to:

1. Master appropriate practices to tackle a research question: experimental design, survey design, and hypothesis testing.
2. Apply proper data science techniques to the research question.
3. Deploy advanced algorithms to make predictions and improve decision making.
4. Write R-scripts.

### 2.3 Assessment/Grading Details

Subject	Percent of Grade
Attendance and participation	10%
Group project	40%
Assignments	50%

Students are required to read assigned materials before the class in order to actively participate in the class discussion. Attendance and participation in discussion account for 10% of the final grade.

For the group project, students will be randomly divided into several groups. Each group needs to make a detailed report for a research question or a business problem. The students should clearly show how they collect the data, prepare the data, analyse the data, and interpret the results. Students will be judged independently. An oral defence may take place during the exam period.

Students need to complete 3-4 assignments.

## **2.4 Academic Honesty and Plagiarism**

It is important for a student's effort and credit to be recognized through class assessment. Credits earned for a student work due to efforts done by others are clearly unfair. Deliberate dishonesty is considered academic misconducts, which include plagiarism; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; or altering, forging, or misusing a University academic record; or fabricating or falsifying of data, research procedures, or data analysis.

All assessments are subject to academic misconduct check. Misconduct check may include reproducing the assessment, providing a copy to another member of faculty, and/or communicate a copy of this assignment to the PHBS Discipline Committee. A suspected plagiarized document/assignment submitted to a plagiarism checking service may be kept in its database for future reference purpose.

Where violation is suspected, penalties will be implemented. The penalties for academic misconduct may include: deduction of honour points, a mark of zero on the assessment, a fail grade for the whole course, and reference of the matter to the Peking University Registrar.

For more information of plagiarism, please refer to *PHBS Student Handbook*.

## **3. Topics, Teaching and Assessment Schedule**

	Topic	Recommended reading	Tool
1	Introduction to applied data analytics		
2	Basic statistics	Chapter 2 [Discovering statistics using R]	R studio
3	Experimental Design: Online Experiment, Field Experiment, Quasi-experiment, Lab Experiment, Eye Tracking, Experimental Neuroscience	[Discovering statistics using R]	R studio
4	Analysis of Experiment Data I: T-Test, ANOVA	Chapter 9 [Discovering statistics using R]	R studio
5	Analysis of Experiment Data II: Moderation and Mediation Analysis	[Introduction to mediation, moderation, and conditional process analysis]	R studio
6	Analysis of Experiment Data III: Moderated Mediation Analysis	[Introduction to mediation, moderation, and conditional process analysis]	R studio

7	Survey Methodology		
8	Analysis of Survey Data I: Linear Regression, Logistic Regression	Chapter 7, 8 [Discovering statistics using R]	R studio
9	Analysis of Survey Data II: CB-SEM		R studio
10	Analysis of Survey Data III: CB-SEM		AMOS
11	Analysis of Survey Data IV: PLS-SEM		R studio
12	Analysis of Survey Data V: PLS-SEM		SmartPLS
13	Introduction to Statistical Learning	Chapter 2 [An Introduction to Statistical Learning]	R studio
14	Data Preparation: Missing Value, Exploring the Data, Feature Engineering, Training and Validation Sets	Chapter 2 [An Introduction to Statistical Learning]	R studio
15	Estimating Competing Models I: Linear Regression & Regression Tree	Chapter 3 [An Introduction to Statistical Learning]	R studio
16	Estimating Competing Models II: Logistic Regression, Decision Tree, and Random Forest	Chapter 4, 8 [An Introduction to Statistical Learning]	R studio
17	Estimating Competing Models III: Forecasting Air Pollution		R studio
18	Text Analysis and Sentiment Analysis		R studio

#### 4. Miscellaneous

n/a