**STAT 1061 Data Science Foundations**

**Fall 2022**

**Lecture Hours**: M/W/F 12:00 – 12:50 PM

**Meeting Location:** 232 Cathedral of Learning

**Instructor**: Junshu Bao

**Email**: jub69@pitt.edu

**Office:** 1806 Posvar Hall

**Office Hours**: T 1:00 – 2:00 PM and TH 11:00 AM – 12:00 PM

**TA:** Shuoran Li

Email: shl198@pitt.edu

Office Hours : W 1:00 – 3:00 PM

**Recitation:**  1020 TH 9:00– 9:50 AM

 1030 TH 10:00 – 10:50 AM

 1040 TH 1:00 – 1:50 PM

 1050 TH 2:00 – 2:50 PM

**Textbooks and Reference Books**:

1. Moore, D.S., G.P. McCabe and B.A. Craig, *Introduction to the Practice of Statistics,* 9th Ed. (ISBN: 9781319013387).
2. Peter Dalgaard, Introductory Statistics with R, Second Edition, Springer, New York, NY (ISBN: 978-0387790534).
3. Peter Bruce and Andrew Bruce, Practical Statistics for Data Scientists, 2nd Edition (ISBN: 978-1491952962)

**Course Description**This course is a general introduction to the rapidly growing field of data science. The topics covered include descriptive statistics, probability, sampling, hypothesis testing, and ANOVA. Broader topics such as regression analysis, model selection, model diagnostics, etc. will also be discussed if time permits. The course will be very hands-on with students actively carrying out the techniques/procedures to foster the ability to reason with data. As part of this process, students will learn the basics of programming in R, and these skills will be reinforced through weekly labs. In developing the core concepts, students will also be exposed to ancillary topics such as data ethics, simulation, and best practices in programming. No previous coding experience nor background in statistics will be assumed.

**Learning Objectives:**

* Understand basic concepts of statistics and probability which include descriptive statistics, elementary probability, random sampling, controlled experiments, inference procedures, regression and the analysis of variance
* Be proficient in the use of R and apply basic statistical techniques to a variety of subjects with the aid of R.
* Understand, analyze, and criticize quantitative arguments.

**Software**

Throughout the semester we will use the statistical software package R. R is a free, open-source software package/programming language for statistical computing and is available on the PCs at all campus computing labs. If in addition you would like to have R on your PC/Mac/Unix, it can be downloaded for free at http://www.r-project.org/.

**Canvas**

We will be using Canvas as the learning management system (LMS) for this course. You should visit Canvas (<http://canvas.pitt.edu>) at least once a week. Here you will find all relevant course material.

**Grades Calculation**

Homework assignments (8) 40%
Quizzes (4) 20%

Midterm 15%

Final exam 25%

Attendance bonus 3%

**Homework Assignments**

Homework assignment will be posted on Canvas. All assignments will be submitted electronically through Canvas. All assignments should be your individual work; otherwise, points will be deducted. Late homework will not be accepted.

**Examinations**

The two exams will be in-class. No make-up exams will be given. If you’ll miss an exam (an extremely rare event for an unavoidable emergency), please let me know as soon as possible.

**Quizzes**There will be a total of four quizzes given throughout the semester in recitation. You may bring a prepared cheat sheet. No make-up quizzes will be given.

**Labs**

During selected weeks, you will complete a lab during recitations. Almost all lab exercises are included in homework assignments, so attending each lab and receiving detailed instructions are crucial to complete homework assignments.

**Attendance/ Class Participation**

Attendance is not mandatory, although very strongly encouraged as the material tends to be difficult to learn on your own. Regular attendance and active participation will surely positively impact your performance in the course. A 3% attendance bonus will be granted to students who missed at most five lectures throughout the semester. It will be added directly to your overall grade. For example, 88% + 3% = 91% ( B+ 🡪 A-).

**Free Tutoring**

Beyond lecture, recitation, my office hours and your TA’s office hours, there is a free tutoring service available to Pitt students in the introductory statistics courses. Graduate statistics students who are available to help you on your homework assignments staff the Statistics Lab in WWPH 1201. Refer to

<http://www.stat.pitt.edu/resources/statistics-computing-lab> to see when the lab is staffed and when it is reserved for other recitations.

**Academic Integrity**
Students in this course are expected to comply with the University of Pittsburgh’s Academic Integrity policy, which can be found at <http://www.as.pitt.edu//fac/policies/academic-integrity>. Any student found blatantly copying another student’s homework assignment will receive a zero for that assignment. Any student allowing another student to copy his/her homework will also receive a zero for the assignment. Any student found cheating on an exam or assisting others in cheating on an exam will receive an F for the course and may be subject to further disciplinary action. Generally speaking, it is expected that you do not lie, cheat, or steal in your academic endeavors (As the great Mr. Feeney once said, *“Do good”*).

**Disabilities**
Students with documented disabilities are entitled to reasonable accommodations if necessary. If you have a disability that requires special accommodations, please contact Disability Resources and Services in 140 William Pitt Union no later than the second week of the semester. Their website is <http://www.drs.pitt.edu/> and their phone number is 412-648-7890. Accommodations will not be granted retrospectively. They will verify your disability and determine reasonable accommodations for this course.

**Course Grades:**

 A+ $\geq 97\%$

 A $93-<97\%$

 A- $90- <93\%$

 B+ $87-<90\%$

 B $83-<87\%$

 B- $80-<83\%$

etc…