Instructor: Dr. Akira Horiguchi (he/him)
TA: Zhentao Li
TA: Zijie Tian
TA: Lingyou Pang (he/him)

STA 141A – Fundamentals of Statistical Data Science

(Syllabus – as of Mar 27, 2025)

Instructor: Akira Horiguchi (▦ MSB 1227, ✉ ahoriguchi@ucdavis.edu) – contact for lecture related and general questions

- ⏱ Office hours: M 09:00 – 09:50 AM (▦ MSB 1143)

TAs: Zhentao Li (✉ ztlli@ucdavis.edu) – contact for general questions and homework grading

- ⏱ Office hours: TBD

Zijie Tian (✉ zijtian@ucdavis.edu) – contact for general questions and homework grading

- ⏱ Office hours: TBD

Lingyou Pang (✉ lyopang@ucdavis.edu) – contact for discussion related and general questions and homework grading

- ⏱ Office hours: TBD

Presence times

- Lecture: MWF 1:10 – 2:00 PM (▦ Young 198)

- Discussions: W 3:10 – 4:00 PM, W 4:10 – 5:00 PM, and W 5:10 – 6:00 PM (▦ Wellman 230)

Prerequisites: STA 10 or STA 13 or STA 32 or STA 100; STA 106 or STA 108

Textbooks

- G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning with Applications in R, 2nd ed.* Springer Texts in Statistics. Springer-Verlag New York, 2021. ✇ statlearning.com (The website also contains R code for the "lab" sections of each textbook chapter.)

- W. N. Venables, D. M. Smith, and the R Core Team. *An Introduction to R.* 2020.
  ✇ cran.r-project.org/doc/manuals/r-release/R-intro.pdf

- H. Wickham and G. Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* O'Reilly Media, Inc., 2017. ✇ r4ds.had.co.nz

Computing: The freely available programming language R can be downloaded at ✇ r-project.org. It is very useful with the integrated development environment RStudio, downloadable at ✇ rstudio.com. Both resources are used in this course. RStudio can also be accessed with mobile devices using jupyter, hosted by UC Davis ✇ jupyterhub.ucdavis.edu/hub/spawn. Further useful links are the following:

- Stackexchange: ✇ stats.stackexchange.com

- R bloggers: ✇ r-bloggers.com

- CRAN vignettes: ✇ search.r-project.org

- R-help Mailing List: ✇ stat.ethz.ch/pipermail/r-help/

Grading

- Students get an ...

  - ... **A+** for scoring $\geq 95.00\%$,
  - ... **A** for scoring $\geq 85.00\%$,
  - ... **A-** for scoring $\geq 80.00\%$,
  - ... **B+** for scoring $\geq 75.00\%$,
  - ... **B** for scoring $\geq 65.00\%$,
  - ... **B-** for scoring $\geq 60.00\%$,
  - ... **C+** for scoring $\geq 55.00\%$,
  - ... **C** for scoring $\geq 45.00\%$,
  - ... **C-** for scoring $\geq 40.00\%$,
  - ... **D+** for scoring $\geq 35.00\%$,
  - ... **D** for scoring $\geq 25.00\%$,
  - ... **D-** for scoring $\geq 20.00\%$.

- If needed, grades will be curved as follows: top 25% of the students will be guaranteed at least A-, the next 25% will be guaranteed at least B-, and the next 25% will be guaranteed at least C-. The relative grading will be non-penalizing.

- <u>Grading scheme</u>

  - Homework (graded for completion): 5%
  - Two midterm exams: 2*30% = 60% (written in class).
  - Final project: 35%

- To be fair to everyone in the class, single student's grades cannot be raised. Please avoid emailing the instructor or TAs about raising scores.

Notes about lecture

- Unannotated slides are uploaded before the lecture.

- Annotated slides and R code are uploaded up to one day after the lecture.

- Lectures will not be recorded.

Homework

- The purpose of the homework is for you to learn to apply the material from class. The homework assignments will be graded for completion, which means it will be your responsibility to check the correctness of your own homework. Solutions for each homework assignment will be uploaded after the assignment's deadline.

Discussion

- The discussion session is exclusively there to deepen the lecture content. There will be no further assessment in the discussion session. Mainly the contents of the exercises and lectures are discussed, but also other topics that can be important as a basis or also as an outlook on further lecture content.

- We are interested in your ability to accurately communicate mathematical and statistical content. So please take the opportunity to communicate your own ideas and questions and also be prepared to present your solutions to the group.

Questions

- Please ask all questions related to the lecture content that you think are not to be assigned to the discussion session during or after the lecture. If you have questions that could not be answered during the lecture, visit my office hours (preferred).

- All questions concerning the exercise sheets and content discussed in the exercises are to be asked in the discussion session if possible.

- This course will use Piazza. Please try to ask your questions there before you e-mail the Instructor or TA.

- For further questions regarding the discussion content which could not be clarified during the discussion session, with other students, or in Piazza, get in touch the TA Lingyou Pang.

- For unanswered questions or content covered in Piazza, firstly get in touch with the TAs.

- For general homework questions that could not be solved in Piazza, the discussion sessions or in the office hours, get in touch with the TAs.

Midterm exams

- For each midterm exam, except for a (black or dark grey) pen/pencil, a non-graphing calculator and one hand-written sheet of paper (both-sided), no other materials are necessary.

- Further information about the materials that can be taken for the midterm exams and the final exam, as well as other information concerning these exams will be given in the lecture and also in the announcement section in canvas.

- Material for each quiz is cumulative, but likely will cover what was on the most recent homework.

Final project

- The final project consists of a case study which showcases the methods learned in this course.

- Each project report has to be compiled by groups of five students. The groups will be randomly assigned at the end of the 3rd week.

- Contribution of each group member must be clearly indicated on the first page of the report. Contact the TAs regarding any issues with the group.

- The grade for the project report will consider the individuals contribution, the overall presentation and meaningfulness of the study. A detailed rubric will be published later.

- On the date specified in the course website, each group must submit a non-graded project proposal report, consisting of no more than two pages. The proposal should outline the research idea, methods intended to be used and responsibilities.

Code of conduct

- Students must adhere to the UC Davis code of conduct (⅛ ossja.ucdavis.edu/code-academic-conduct). Violations of the code of conduct include (but are not limited to) communicating and collaborating during midterm exams, copying, attempting to copy and letting someone copy a graded assignment, doing someone else's homework/exam, having someone else doing one's homework/exam, sharing homework/exams, and submitting work that is not yours. The fact that the violation did not benefit you directly, does not mean that it is less important.

- One of the responsibilities delegated to faculty by the UC Davis Code of Academic Conduct is to report suspected academic misconduct to the Office of Student Support and Judicial Affairs (OSSJA). Thus, any violation will be reported, and students found guilty will get an F, no matter the extend and type of their violation. Please, do not do it.

Course materials: My lectures and course materials, including lecture notes and discussions etc., are protected by U.S. copyright law and by University policy. I and the TAs are the exclusive owner of the copyright in those materials created. You may take notes, make copies of course materials for your own use, and share those materials with other students enrolled in or auditing this course. However, you may not reproduce, distribute or display (post/upload) lecture notes or course materials

in any other way – whether or not a fee is charged – without my express prior written consent. You also may not allow others to do so. If you do so, you may be subject to student conduct proceedings under the UC Davis Code of Academic Conduct. Similarly, you own the copyright in your original papers and exam essays. If I am interested in posting your answers or papers on the course web site, I will ask for your permission.

Changes to syllabus: The syllabus may be updated based on the instructor's discretion. Changes may, for instance, result due to students' progress in the course or changes in instructional priorities.

Acknowledgment: The course material is developed based on the recommended textbook, on online resources as cited in the relevant materials, and on the content taught in STA141A past quarters by Akira Horiguchi and Sebastian Kuehnert, where ideas from Emanuela Furfaro (University of Washington, Seattle) and Debashis Paul (University of California, Davis) were incorporated.