**STA 101: Data Analysis and Statistical Inference**

**Duke University, Summer I 2021**

*Instructor*: Fan Bu ([fan.bu1@duke.edu](mailto:fan.bu1@duke.edu))

*Course time*: Every weekday, 10:00-11:15 am (via Zoom)

*Course website*: access through <https://sakai.duke.edu/>

*Teaching Assistants*: TAs undecided

*Lab time*: TBD

*Office hours*:

* Fan Bu: TBD
* TAs: TBD

*Learning objectives*. By the end of this course, you will be able to

1. Comprehend the process of data collection, identify potential limitations in data collection methods and other sources of statistical bias, and determine how they affect the scope of inference.
2. Possess beginner-level experience with statistical software (e.g., R) to summarize data numerically and visually, and to conduct data analysis.
3. Understand basic concepts of statistical inference.
4. Apply estimation and testing methods to analyze single variables or the relationship between two variables, in order to investigate natural phenomena and make data-based decisions.
5. Model numerical response variables using a single explanatory variable or multiple explanatory variables in order to study relationships between variables.
6. Communicate modelling results correctly, effectively, in context and in a manner that is understandable to non-technical audiences.
7. Critique data-based claims and evaluate data-based decisions.
8. Carry out research projects with rigorous statistical data analysis.

*Grading policy*. Your final grade will consist of (100% in total):

* Class participation (5%). Answer or ask questions in class, OR discuss on Piazza.
* Assignments (40%). Includes problem sets, application exercises, and labs (submit via Gradescope).
* Mid-term exam (25%). Online test on Sakai. Time: flexible, June 2-6.
* Final exam (35%). Online test on Sakai. Time: flexible, June 21-23.

**Don’t be late!** 50% of the points will be deducted if you are late *no more than* 24 hours, and you will lose all credit if you are more than 24 hours late.

The lowest score of all your assignments will be dropped.

*Schedule*.

|  |  |  |
| --- | --- | --- |
| DATE | CONTENT | NOTE |
| May 12 | Start Chapter 1: Intro to data |  |
| May 13 | Application Exercise 1 posted | Due: 11:59pm, May 15 |
| May 17 | Start Chapter 2: Summarizing data |  |
| May 18 | Application Exercise 2.1 posted | Due: 11:59pm, May 20 |
| May 19 | Application Exercise 2.2 posted | Due: 11:59pm, May 21 |
|  | Problem Set 1 posted  **PS1: 1.5, 1.35, 1.37, 2.25, 2.27, 2.29** | Due: 11:59pm, May 24 |
| May 20 | Start Chapter 3: Probability |  |
| May 21 | Application Exercise 3.1 posted | Due: 11:59pm, May 23 |
| May 24 | Application Exercise 3.2 posted | Due: 11:59pm, May 26 |
| May 26 | Start Chapter 4: Distributions of random variables |  |
| May 27 | Application Exercise 4 posted | Due: 11:59pm, May 29 |
| May 28 | Problem Set 2 posted  **PS2: 3.5, 3.27, 3.31, 3.43, 4.13, 4.31** | Due: 11:59pm, June 3 |
| June 1 | Start Chapter 5: Foundation for inference |  |
| June 2 | Mid-term Exam open | Finish by 11:59pm, June 6 |
| June 3 | Application Exercise 5 posted | Due: 11:59pm, June 5 |
| June 3 | Start Chapter 6: Inference for categorical data |  |
| June 7 | Application Exercise 6 posted | Due: 11:59pm, June 9 |
|  | Problem Set 3 posted  **PS3: 5.29, 5.36, 6.19, 6.33, 6.41, 7.7, 7.31, 7.33** | Due: 11:59pm, June 14 |
| June 8 | Start Chapter 7: Inference for numerical data |  |
| June 10 | Application Exercise 7 posted | Due: 11:59pm, June 12 |
| June 11 | Start Chapter 8: Intro to linear regression |  |
|  | Application Exercise 8 posted | Due: 11:59pm, June 13 |
| June 15 | Problem Set 4 posted  **PS4: 8.21, 8.29, 8.35, 9.3, 9.9, 9.13** | Due: 11:59pm, June 21 |
|  |
| June 16 | Start Chapter 9: Multiple & logistic regression |  |
| June 18 | Application Exercise 9 posted | Due: 11:59pm, June 20 |
| June 21 | Final Exam open | Finish by 12pm(!), June 23 |
|  |  |  |

*Academic Honesty*. Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity. Cheating on exams and quizzes, plagiarism on homework assignments, projects, and code, lying about an illness or absence and other forms of academic dishonesty are a breach of trust with classmates and faculty, violate the Duke Community Standard, and will not be tolerated. Such incidences will result in a 0 grade for all parties involved as well as being reported to the University Judicial Board. Additionally, there may be penalties to your final class grade. Please review Duke’s Standards of Conduct. For more information on the Duke honor code (known as Duke Community Standard), please go to <https://integrity.duke.edu/>.

*Students with Disabilities*. Students who require special accommodations in class or during exams should follow the procedures outlined by the Disability Management Program <https://access.duke.edu/students>. Students with disabilities who believe they may need accommodations in this class are encouraged to contact the Student Disability Access Office at (919) 668-1267 as soon as possible to better ensure that such accommodations can be made.

*Privacy Policies*. Student records are confidential.