

# Statistics: Principles, Methods and R (I)

---

Fengnan Gao<sup>1,2</sup>

1-3-2019

<sup>1</sup>School of Data Science, Fudan University

<sup>2</sup>Shanghai Center for Mathematical Sciences

fngao@fudan.edu.cn

- The course is primarily modeled after **Qiwei Yao**'s course.
- The course will provide a **comprehensive** coverage on some fundamental aspects of probability and statistics methods and principles.
- Data illustration using statistical package **R** constitutes an integral part throughout the course, therefore provides the hands-on experience in simulation and data analysis.
- Bear in mind that this is **NOT** a course about **R**!
- Course Homepage: <http://www.sdspeople.fudan.edu.cn/gaofengnan/teaching/StatsI/1819SS.html>

1. (1 lecture) Introduction to R: What is R? Installing R, help and documentation, data objects, data import and export, basic data manipulation, computing with data, organising an analysis.
2. (2/3 lecture) Probability: sample space and events, probability, independent events, conditional probability, Bayes' formula.
3. (2/3 lecture) Random variables and distributions: distribution functions and probability functions, mean and variance, moment generating functions, discrete random variables, continuous random variables.
4. (2/3 lecture) Multivariate distributions: bivariate distributions, marginal distributions, independent random variables, conditional distributions, multivariate distributions, IID samples, transformations of random variables.

5. (1 lecture) Inequalities: probability inequalities, inequalities for expectations.
6. (2 lectures) Convergence of random variables: types of convergence, law of large numbers (LLN), central limit theorem (CLT).
7. (2 lectures) Introduction to Statistical Inference: what is statistics? parametric and nonparametric models, fundamental concepts in inference, empirical distributions.
8. (2 lectures) Point estimation: method of moments estimation, maximum likelihood estimation (MLE), properties of MLE, asymptotic efficiency.
9. (1.5 lectures) Hypothesis testing I: null and alternative hypotheses,  $p$ -values, two-types of errors, the Wald test,  $t$ -tests and  $t$ -intervals.

10. (1.5 lectures) Hypothesis testing II: likelihood ratio tests, Pearson's  $\chi^2$ -test, goodness-of-fit tests, permutation tests.
  11. (2 lectures) Regression models: simple linear regression, least squares estimation (LSE), prediction, multiple linear regression, model selection.
- 12 (16 – 3 – 1) lectures in total.
  - The lectures on April 5 (Qingming holiday), May 17 (University sports day), June 7 (Dragon boat festival) will not take place.
  - The lecture on June 14 will not take place either. We will figure out how to make up for the time.

- There will be the course **Statistics: Principles, Methods and R (II)** in the next semester.
- I will not teach that course.
- Stats I covers the **basic** aspects of statistics
- Stats II will be more advanced and state-of-art.

- Course Homepage: <http://www.sdspeople.fudan.edu.cn/gaofengnan/teaching/StatsI/1819SS.html>
- The course **slides** will be put on the the course homepage after each lecture.
- There will be **homework** assignments on the course homepage.
- If necessary, R programs and related data will be put there as well.

## References

- Wasserman, L. (2004). *All of Statistics*. Springer. (**Main textbook**)
- Casella, G.S. and Berger, R.L. (2002). *Statistical Inference* (2nd edition). Duxbury.
- Knight, K. (2000). *Mathematical Statistics*. Chapman & Hall.
- Pawitan, Y. (2001). *In All Likelihood*. Oxford University Press.
- Venables N. et. al. (2014). *An Introduction to R*. Available online at <http://cran.r-project.org/doc/manuals/R-intro.pdf>
- Zuur, A., Ieno, E. and Meesters, E. (2009). *A Beginner's Guide to R*. Springer. Available online from Fudan Library.
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer.



# Exams and Grading

- Exams
  - This is a basic course, not an applied one
  - Two exams: mid-term and final
  - Most exam problems are of theoretical nature—computations and proofs
  - Interpreting **R** code and results will also be in the exams!
- The final grades will be a weighted average of the following evaluations
  - 20% — Homework and participations
  - 30% — Mid-term exam, (approximately) on Friday, April 26
  - 50% — Final exam on Friday, June 14
- Please keep in mind that at most **30%** of all students get A or A-.

# Homework Assignments

- Please meet the teaching assistants— Zhang Yuanyuan(18210980079@fudan.edu.cn)
- There is homework after each lecture, please hand in the homework before the next lecture. For example, the homework after today is due before the next lecture, typically one week later.
- The TA collects the homework in each lecture and grade them.
- Please hand in the solutions **in time!** No late submission is accepted.
- Homework of programming on **R** should be handed in along with the actual code. Please do not handwrite the code. Hand them in either by email or by printouts.

**Proceed...**