Statistics: Principles, Methods and R (I)

Fengnan Gao[1,2]

2017.9.11

[1]School of Data Science, Fudan University

[2]Shanghai Center for Mathematical Sciences

fngao@fudan.edu.cn

Overview

- 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.
- It also covers simple linear regression analysis.
- 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/1718FSSDS.html`
- I am teaching another course *Statistical Learning* for Shanghai Center for Mathematical Sciences.

Contents i

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.

Contents ii

5. (1 lecture) Inequalities: probability inequalities, inequalities for expectations.
6. (2 lectures)
   Convergence of random variables and Monte Carlo Methods: types of convergence, law of large numbers (LLN), central limit theorem (CLT), Monte Carlo integrals, importance sampling.
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, parametric bootstrap, EM algorithms.

Contents iii

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 χ^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.

- 15 lectures planned in total.
- The lectures on October 2 and January 1 will not take place.
- The lecture next week will be taught by Wang Qingwen, who is teaching *Introduction to Stochastic Processes* this semester.

Relations to Stats II

- 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, and Stats II will be more advanced and state-of-art.

Course Homepapge

- Course Homepage: `http://www.sdspeople.fudan.edu.cn/gaofengnan/teaching/1718FSSDS.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.

An execellent English course?

- The course will be given (almost) fully in English.
- The course is under consideration as a candidate for *excelent English courses*, there might be other professors attending the lectures from time to time.
- There might be a recorded session in one lecture, someone (or someones) might come with a big video camera.
- However, the evaluation of the course are not about you (the students), but rather, about evaluating me (the teacher). Do NOT panic.

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.

Examinations 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, (estimated) on November 6
  - 50% — Final exam on January 8, 2018
- Please keep in mind that at most **30%** of all students get A or A-.

Homework Assignments

- Please meet the teaching assistant—He Siyuan (14110690008@fudan.edu.cn)
- There is homework after each lecture, please hand in the homework before the next lecture. For example, the homework after today (September 11) is due on September 18.
- The TA collects the homework in each lecture and grade them.
- Please hand in the solutions **in time**!
- 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.