

Introduction



Feng Li

feng.li@cufe.edu.cn

**School of Statistics and Mathematics
Central University of Finance and Economics**

February 24, 2014

General information

- **Lecturer** : Feng Li, Email: feng.li@cufe.edu.cn
- **Language**: The course is taught in Chinese. The course materials are in English.
- **Reception hours**: Questions concerned with this course can be asked after each lecture or via email.
- **Lecture notes**
Course materials and updated news are available at the course homepage <http://feng.li/teaching/stat-software/>
- **Working load**: Depending on your own situation and your ambition, you decide how much time you want to input. But there are compulsory assignments including presentation and discussions should be done in order to pass the exam.

Assignments and examinations

To pass the final exam you have to have at least 60p in the total score with the following tasks.

- **Lab assignments and presentation (50%)**
There will be a **mandatory** presentation for each of you.
- **Computer-based exam (50%)**

Don't worry and it's fun!

The modern statistical computing environment

- Based on Linux
- Open source
- Cheap to do parallel computing
- Flexible to handle big data
- Examples: R, hadoop...

The differences between Linux and Windows

- Linux distributions are built with free software.
 - Users are free to do whatever they want with it - this includes the freedom to redistribute the software free-of-charge, or to sell it (or related services such as support or warranty) for profit.
- Linux has the complete tool chain for software development.

Concepts under Linux

- **GUI**: graphical user interface
- **CLI**: command-line interface
- **root**: the user who has all rights or permissions (to all files and programs) in all modes (single- or multi-user)
- **console**: a text output device for system administration messages
- **terminal**: a hardware device for data entry and display in a computer system
- **environment variables**: a set of dynamic named values that can affect the way running processes will behave on a computer.
 - For example, a running process can query the value of the `TEMP` environment variable to discover a suitable location to store temporary files, or the `HOME` or `USERPROFILE` variable to find the directory structure owned by the user running the process.

The procedure of statistical analysis

- **Data collection**

- In the old days, we got the data from questionnaires.
- Nowadays, a lot data need to be retrieved from Internet, digital books, pdf files.

- **Model estimation**

- New statistical methods needs updated statistical software.
- Big Data needs efficient statistical software.

- **Reporting**

- A simple pie chat or histogram will not work if the data are complicated and high-dimensional.
- Visualization is also important