Welcome to

APSTA-GE 2003: Intermediate Quantitative Methods Lab Section 003

Fall 2020

NYU

Hello
my name is

Tong Jin

I'm a second year A3SR student

- 2 lab sections: 002 and 003
  - Both will focus on R programming
  - Also cover math and simulation
- Lab Resources
  - Available before each lab
  - In R script format
  - Available for download on NYU Classes
    - under "Resources - Tong's Lab (003)"

Before we get started ...

How to increase programming skills?

Cutting corners to meet arbitrary management deadlines

Essential

Copying and Pasting from Stack Overflow

O'REILLY®

The Practical Developer
@ThePracticalDev

Software can be chaotic, but we make it work

Expert

Trying Stuff Until it Works

O RLY?

The Practical Developer
@ThePracticalDev

How to actually learn any new programming concept

Essential

Changing Stuff and Seeing What Happens

O RLY?

@ThePracticalDev

The internet will make those bad words go away

Essential

Googling the Error Message

O RLY?

The Practical Developer
@ThePracticalDev

Before we get started …

- [Syllabus](#)
- About [Welcome Packet](#)
  - [Assignment 0](#)
    - Submission is optional
    - You can turn in and get feedback
- [Install R and RStudio](#) and setup development environment
  - R: software environment for statistical computing and data visualization
  - **RStudio**: R-based [IDE](#)
- Create a local folder for this course and download materials on NYU Classes
  - [Lab1_Script_ClassNotes_090820.R](#)
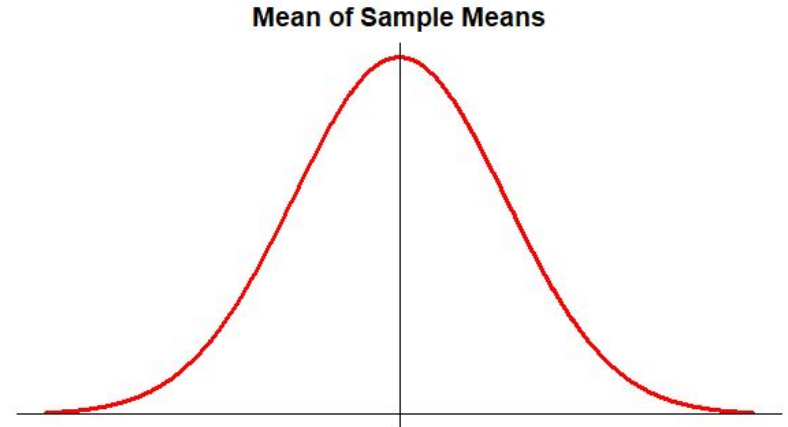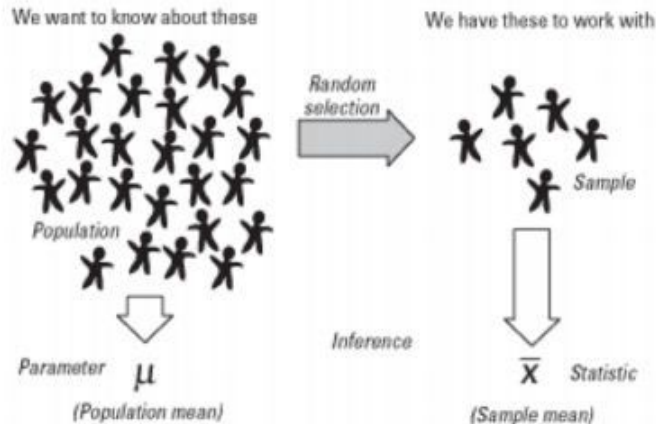 - [marathon.csv](#)

Notes for Prof. Lu's Lecture

- **Average** vs. Individual

It's important to remember that we're measuring the relationship **on average**.

On average, for two groups of runner whose average training experience differs by one year, we expect the more experienced group to have finish time that is 5 minutes faster.

- Statistical Inference
  - Random sampling for many times
  - Record the mean of each sample
  - The distribution of sample means follows normal (bell-shaped)
    - Central Limit Theorem
  - The mean of the sample distribution = population mean
  - Standard deviation of sample distribution (Standard Error)
    - SE = SD / sqrt(sample_size)

We want to know about these

We have these to work with

Random selection

Population

Sample

Parameter μ

(Population mean)

Inference

X̄ Statistic

(Sample mean)

**Mean of Sample Means**

- Hypothesis Testing
  - One sample
    - Make an assumption on population mean
    - Randomly draw a sample
    - Test if sample mean is **statistically different** from population mean
  - Two sample
    - Test if two populations have equal mean value

- Hypotheses
  - Null hypothesis - difference is not [significant](#)
    - One sample
      - sample mean = population mean (w. confidence)
    - [Two samples](#)
      - mean of pop 1 = mean of pop 2 (w. confidence)
  - Alternative hypothesis - difference is significant
    - One sample
      - sample mean ≠ population mean (w. confidence)
    - Two samples
      - mean of pop 1 ≠ mean of pop 2 (w. confidence)

- [T-test](#)
  - T statistic
    - One sample
      - Mean difference standardized by sample standard error
    - Two sample
      - Mean difference standardized by **standard error**
      - **Standard error** assumptions:
        - Unequal variance
          - Standardized ($SE\_1 + SE\_2$)
        - **Equal variance**
          - Standardized (addition of two variances)
          - [Degrees of freedom](#): $n\_1 + n\_2 - 2$
    - [Levene's Test](#)
      - Assess the equality of variances

- p-value
  - Assume null is correct (=)
  - The probability of obtaining test results at least as extreme as the results actually observed
    - The smaller, the less likely to observe again
    - The area under the curve

Mean of Sample Means
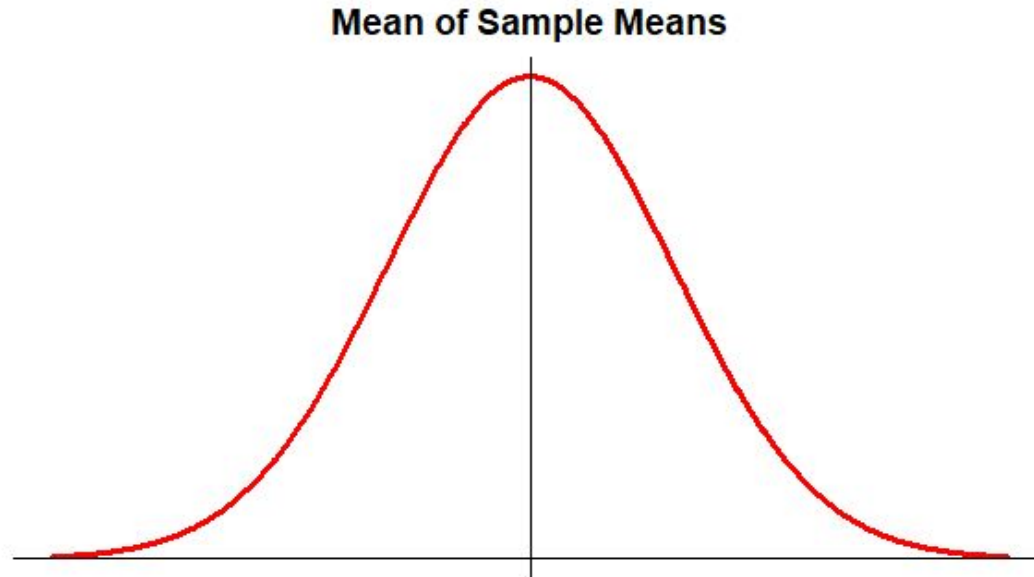
RStudio Basics

- Panes
  - Source - editing and executing code
  - Console - results
  - Environment - **global** and package-based **environment** (variables, data sets, …)
  - History - code execution history
- Comment/Uncomment
  - Add "#" at the beginning to comment a line out
  - `Ctrl/⌘ + Shift + C`
- File Types
  - **R Scripts**
  - R Markdown
  - R Project
  - …

R Basics

- R Console
  - Edit and run code
  - Return script results
- R Editor
  - "File - New Script" (`Ctrl/⌘ + N`)
  - To execute code, highlight or select lines and click "Run line or selection" (`Ctrl/⌘ + R`)
- Comment/Uncomment
  - Any line starting with "#" will be skipped
  - Add "#" at the beginning to comment a line out

Contact

- Email: tj1061@nyu.edu
- Office hours
    - Monday 9 - 10am
    - Wednesday 12:30 - 1:30pm

Note: 09/09/2020 - Monday schedule

Office hours: 9 - 10am (EST)