ST5201: Basic Statistical Theory
Chap0: An overview

Choi, Yunjin
stachoiy@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

August 22nd, 2017

Outline

- Module information
- Introduction of Statistics
- A general set-up of statistical problem

Module Information I

Aim and Objective

To provide an introduction to statistics for layman (with no background in statistics but some knowledge in elementary calculus) at beginning graduate level via going through fundamental concepts together with formal theory in statistics

Lecture/Tutorial Time

Every Tuesdays at 7:00-10:00 PM
One midterm in Week 7, Time & Location TBA

Module webpage

IVLE (https://ivle.nus.edu.sg/) for lecture notes, problem sheets, solutions

Module Information II

Compulsory Textbook

Mathematical Statistics and Data Analysis (3rd ed.)
written by John A. Rice
published by Cengage Learning

Syllabus

Selected topics from Chapter 1-9 and Chapter 11 of the textbook including probability, random variables, limit theorems, families of distributions, estimation, method of moments, maximum likelihood method, bootstrap method, classical confidence intervals, bootstrap confidence intervals, hypothesis testing, likelihood ratio tests, and two sample tests.

Module Information III

Assessment

- Homework: 10%
- Midterm (in class): 30%
- Final examination (in class): 60%
- Exams are CLOSED BOOK (ONE A4 papers are allowed)

Always keep in mind

- No late Homework
- No make-up final exam

Module Information IV

Advice

- Check IVLE frequently for lecture notes, homework sheets, and solutions
- Attend most, if not all, lectures & tutorials
- Seriously read through & understand the corresponding materials (especially the examples) in the compulsory textbook after lectures
- Attempt homework problem, & other problems from the textbook

Learning Outcomes

- Questions to Address:
  - What is uncertainty
  - What statisticians are doing
  - A general set-up of statistical problem
- Concept & Terminology: $\star$ uncertainty $\star$ population $\star$ parameter $\star$ sample $\star$ statistic

What is uncertainty, I

Questions:

1. How many people vote for Hillary Clinton in 2016 United States presidential election?
2. Will it rain tomorrow?
3. Will I win this dice game?
4. Will Singapore win a medal in Rio Olympics?
5. Is it better to work in Singapore or abroad?

No deterministic answer to these questions

What is uncertainty, II

Reasons:

1. Lack of information
   - Real constraints: cost, technology, ...
   - Theoretical problems: mathematics, quantification, ...
2. Randomness: unable to predict
3. Both

In real life, many problems have uncertainty.
How do we study these problems?

What people usually do

Utilize known information to predict:

- American presidential election poll[1]

Question: Can we claim Hillary will win?

[1] http://www.telegraph.co.uk/news/2016/05/26/us-presidential-elections-2016-poll-tracker/

What people usually do

Utilize known information to predict:

- American presidential election poll[1]

Question: Can we claim Hillary will win?

[1] http://www.telegraph.co.uk/news/2016/05/26/us-presidential-elections-2016-poll-tracker/

What statisticians do

Statisticians' concerns:

- learn from the data; measure, control, and communicate uncertainty
  - Is this poll reliable? How reliable is this poll?
  - What conclusions one can get from the data? Can we claim Hillary will win?
  - What is the probability for Hillary/Trump to win?
  - How to make this poll more reliable?

Statisticians not only do the math, but also interpretation

Some statistical jokes

Without statisticians, what will happen...

- It is proven that the celebration of birthdays is healthy. Studies show that those people who celebrate the most birthdays become the oldest.
- One out of every four people is suffering from some form of mental illness.
  Check three friends. If they're OK, then it's you

Just a joke? No!

Have you heard about this news?

In 2012, this news became worldwide. It was a misunderstanding of statistics in a study[2].

They need a statistician!

[2] Chocolate Consumption, Cognitive Function, and Nobel Laureates (2012)

Just a joke? No!

Have you heard about this news?

In 2012, this news became worldwide. It was a misunderstanding of statistics in a study[2].

They need a statistician!

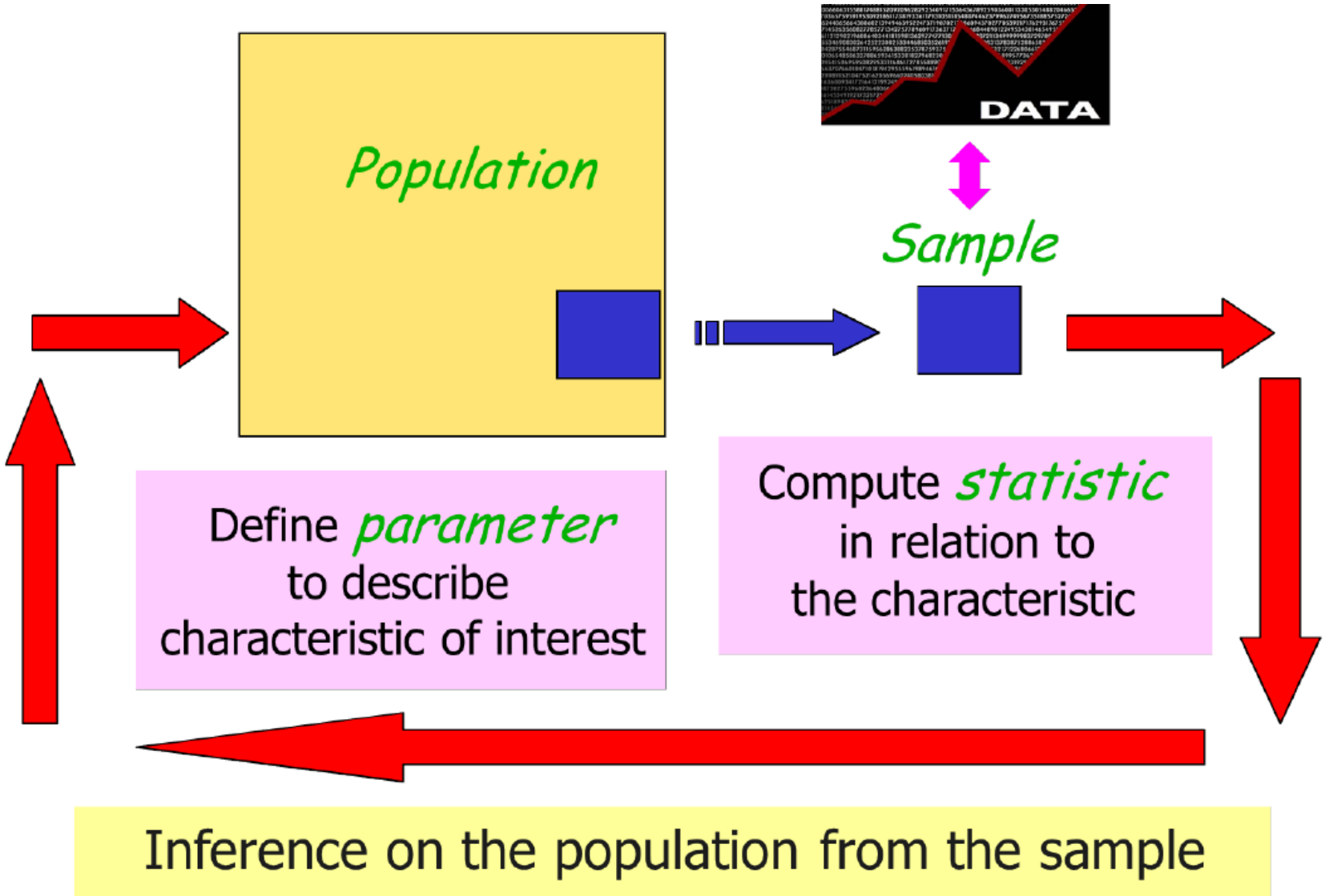
[2] Chocolate Consumption, Cognitive Function, and Nobel Laureates (2012)

How statisticians work?

A general set-up of statistical problems:

- The problem of interest, e.g. ratio of people vote for Hillary Clinton
- Population: all the voters in United States
- Parameter: ratio of people vote for Hillary Clinton
- Data/Sample: the subset (or part) of the designated population, for which the information is available
- Statistics: computed from the sample
- Infer/Guess the unknown parameter based on some statistics

A General Setup of Statistical Problems



The discipline of statistics is concerned with

Collection of data

Observational studies versus Experimental designs

Analysis of data

Selection of appropriate & well-defined statistical methodologies

Interpretation of data

Retrieval of relevant information to address questions in mind

Effective communication & presentation of results relying on data

Drawing meaningful conclusions in laymans language

Example: election political poll

Pre-election polling provides some information of people's engagement in election and opinions about candidates.

With results of these polls, candidates can be clear of their current stages, and adjust their running strategies to win more votes.

However, it costs too much for a nation-wide investigation (300 million population). Most newspaper offices choose to investigate about 10,000 people only.

Research centres can only rely on the limited information about these 10,000 people to predict the election results and provide suggestions for candidates.

Example: election political poll (Cont'd)

Problem: Suppose we are interested in the performance of Hillary Clinton

Population: All qualified voters in United States, about 300 million.

Parameter: The ratio of all these qualified voters who will vote for Hillary Clinton in the election

Sample: Voters enrolled in the poll

Statistic: Ratio of the voters enrolled in the poll who claimed to vote for Hillary Clinton during the election

Example: election political poll (Cont'd)

Considering questions:

Collection of data: How to sample? Is there any bias?

Analysis of data: Will the people lie?

Interpretation of data: Does this result support Hillary? Who supports her?

Communication & Presentation of results: What Hillary Clinton should do next?