

Introduction to Statistics (통계란 무엇인가?)

Chanseok Park (박찬석)

Applied Statistics Laboratory
Department of Industrial Engineering
Pusan National University

July 13, 2018

Hosted by SEC



부산대학교
PUSAN NATIONAL UNIVERSITY

Overview

- 1 Definition of Statistics (통계의 정의)
- 2 Quotes about Statistics (통계학에 대한 말 말 말)
- 3 통계학은 거짓인가?
- 4 통계학 성공예?
- 5 Applications of Statistics (통계학 응용)
- 6 Software for Statistics (통계학 소프트웨어)

Overview

- 1 Definition of Statistics (통계의 정의)
- 2 Quotes about Statistics (통계학에 대한 말 말 말)
- 3 통계학은 거짓인가?
- 4 통계학 성공예?
- 5 Applications of Statistics (통계학 응용)
- 6 Software for Statistics (통계학 소프트웨어)

Overview

- 1 Definition of Statistics (통계의 정의)
- 2 Quotes about Statistics (통계학에 대한 말 말 말)
- 3 통계학은 거짓인가?
- 4 통계학 성공예?
- 5 Applications of Statistics (통계학 응용)
- 6 Software for Statistics (통계학 소프트웨어)

Overview

- 1 Definition of Statistics (통계의 정의)
- 2 Quotes about Statistics (통계학에 대한 말 말 말)
- 3 통계학은 거짓인가?
- 4 통계학 성공예?
- 5 Applications of Statistics (통계학 응용)
- 6 Software for Statistics (통계학 소프트웨어)

Overview

- 1 Definition of Statistics (통계의 정의)
- 2 Quotes about Statistics (통계학에 대한 말 말 말)
- 3 통계학은 거짓인가?
- 4 통계학 성공예?
- 5 Applications of Statistics (통계학 응용)
- 6 Software for Statistics (통계학 소프트웨어)

Overview

- 1 Definition of Statistics (통계의 정의)
- 2 Quotes about Statistics (통계학에 대한 말 말 말)
- 3 통계학은 거짓인가?
- 4 통계학 성공예?
- 5 Applications of Statistics (통계학 응용)
- 6 Software for Statistics (통계학 소프트웨어)

Definition of Statistics (통계학/統計學의 정의)

Google

수학의 한 부문으로, 사회 현상을 통계에 의하여 관찰·연구하는 학문.

집단에 관한 자료를 정리하여 그 특징을 나타내는 여러 가지 수치(數値)를 산출하고 그 자료가 가리키는 것을 알려고 하는 기술(記述) 통계학과 집단의 상태를 그로부터 추출(抽出)한 표본에서 수리적(數理的)으로 추측하는 추측 통계학으로 나뉨.

My Definition

Data Reduction with information.

7/15/2018

통계학 - 위키백과, 우리 모두의 백과사전

WIKIPEDIA

통계학

위키백과, 우리 모두의 백과사전.

통계학(統計學, 영어: statistics)은 수량적 비교를 기초로 하여, 많은 사실을 통계적으로 관찰하고 처리하는 방법을 연구하는 학문이다. 근대 과학으로서의 통계학은 19세기 중반 벨기에의 케틀레가 독일의 "국상학(國狀學, Staatenkunde, 넓은 의미의 국가학)"과 영국의 "정치 산술(Political Arithmetic, 정치 사회에 대한 수량적 연구 방법)"을 자연과학의 "확률 이론"과 결합하여, 수립한 학문에서 발전되었다.^{[1][2]}

Quotes about Statistics (통계학에 대한 말 말 말)

- Statistics is the grammar of science. *Karl Pearson*
- All statistics have outliers. *Nenia Campbell, Terrorscape*
- There are three types of lies – lies, damn lies, and statistics. *Benjamin Disraeli*
- There are lies, damned lies and statistics. *Mark Twain*
- Facts are stubborn things, but statistics are pliable. *Mark Twain.*
- Statistics, likelihoods, and probabilities mean everything to men, nothing to God. *Richelle E. Goodrich*
- God does not play dice. *Albert Einstein*
- Democracy is an abuse of statistics. *Jorge Luis Borges*
- 통계학은 과학의 매춘부. *김용운*

통계학은 거짓인가?

- U.S Presidential Election (Hillary vs. Trump).
Hillary가 Trump를 90% 이상 앞서고 있었다는 보도. 특히 CNN...
실제는 전문 여론 조사 기관(방송사가 아님)에서는 트럼프의 승리
조심스럽게 예상.
Band-wagon effect를 노린 일부 언론의 장난.
- Brexit.
한 명의 사망. 그 후 조사. Brexit반대 여론 순간적을 급증. 그 후, 여론
조사 중지. 급작스런 여론 변화후 서서히 최초 여론으로 되돌아 가고
있었으나 여론 조사 중지로 이를 알아채지 못했음.
- Dewey Defeats Truman (1948, US election).
https://en.wikipedia.org/wiki/Dewey_Defeats_Truman
이 문제는 위의 예와는 달리 통계기술의 부족으로 생긴 오류. 통계
교과서에 자주 인용되는 예.
- Trump와 Brexit의 예가 통계 교과서에서 기술적인 문제에 대한
것으로 나온다면, 통계학은 끝.

통계학 성공예?

- Alpha-go
Brutal Force, Monte Carlo simulation, 모든 경우의 수, 승리 확률, 등등.
통계학의 원리 상당히 이용.
- 원주율 π 의 계산.
한 변의 길이가 2cm인 정사각형 안에 무작위로 점을 찍는다고 하자.
이 직사각형 안에 내접하는 반경 1cm인 원이 있다고 하면, 이 원안에 점을 찍을 확률을 구하면:

$$\pi/4.$$

위와 비슷하게, 한 변의 길이가 2cm인 정육면체 안에 무작위로 점을 찍는다고 하자. 이 정육면체 안에 내접하는 반경 1cm인 구가 있다고 하면, 이 구안에 점이 있을 확률을 구하면:

$$\frac{(4/3)\pi}{2^3} = \frac{\pi}{6}.$$

Software for Statistics (통계학 소프트웨어)

- SAS, SPSS, Minitab, Matlab, BMDP, S, Splus, etc.
- R language: 5th place in the world.
<http://blog.revolutionanalytics.com/2016/07/r-moves-up-to-5th-place-in-ieee-language-rankings.html>

R language (<https://www.r-project.org/>) 실습.

- R manuals, see contributed manuals.
- `help("name")`
- `help.start()`
- Basic functions: mean, sd, var, hist, stem
- Study some theories using simulation: X and Y are uniform.
Then what is $X + Y$?
- π 값 구하기.

R Example:

<https://github.com/AppliedStat/seminar/tree/master/2018/R>