Introduction to Statistics and Data Science using *eStat*

Chapter 1

Statistics and Data Science

Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

1.1 Statistics and Data Science

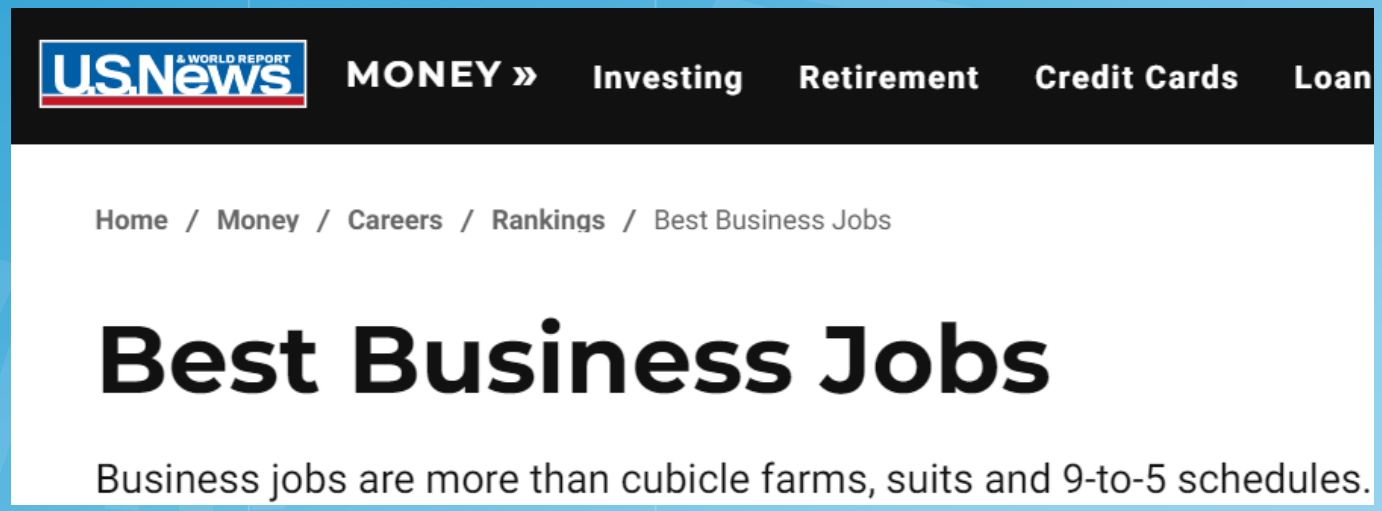
1.2 Population and Sample

1.3 Variables and Data

1.4 Software for Statistical Analysis

1.1 Statistics and Data Science

https://money.usnews.com/careers/best-jobs/rankings/best-business-jobs



In Google, type

'Best Business Jobs US News Careers'

Statistician

#1 in **Best Business Jobs**

Statistics is the science of using data to make decisions. This is relevant in almost all fields of work and there are many opportunities for employment.

1.1 Statistics and Data Science

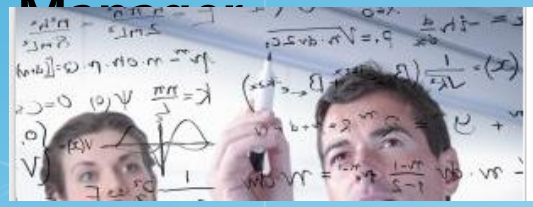
U.S.News & WORLD REPORT MONEY » Investing Retirement Credit Cards Loan

Home / Money / Careers / Rankings / Best Business Jobs

Best Business Jobs

Business jobs are more than cubicle farms, suits and 9-to-5 schedules.

#2 Medical and Health Service Manager



Mathematician

#3 in Best Business Jobs

Operations Research Analyst

#4 in Best Business Jobs

From data mining to mathematical modeling, operations research analysts use advanced techniques to help businesses run in a more efficient and cost-

#5 Financial Manager

#6 Financial Advisor

#7 Accountant

#8 Market Research Analyst

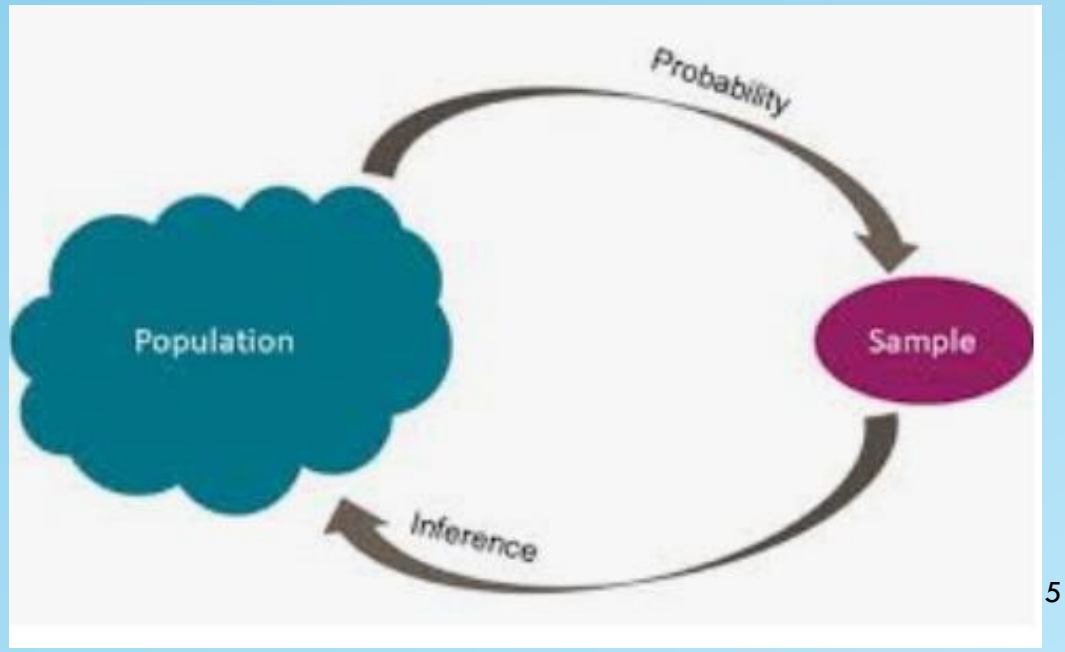
#9 Business Operation Manager

#10 Social and Community Service Manager

#11 Actuary

1.1 Statistics and Data Science

- **Statistics**



1.1 Statistics and Data Science

- **Statistics** = ‘State ‘ + ‘istics’

- **History** tells which country appeared where, when, how large its territory, how much population and how many households
  - In Egypt, Greece and Rome, population and farmland area were used for the management of their country.
- 8th to 13th century, **concept of probability and inference**
  Al-Khalil (717–786), Al-Kindi (801–873), Ibn Adlan (1187–1268)
- 17th to early 19th century, **mathematical foundations of statistics**
  Gerolamo Cardano, Blaise Pascal and Pierre de Fermat.
- late 19th century, Francis Galton and **Karl Pearson**
- early 20th century, **Ronald Fisher**

1.1 Statistics and Data Science

▪ Statistics

- **Statistical methods are applied in all fields nowadays.**
  => management, economics,
  politics, social science, education
  physics, chemistry, biology,
  computer science
  medical science, pharmacy,
  agricultural science
  electrical, electronical, chemical, civil engineering

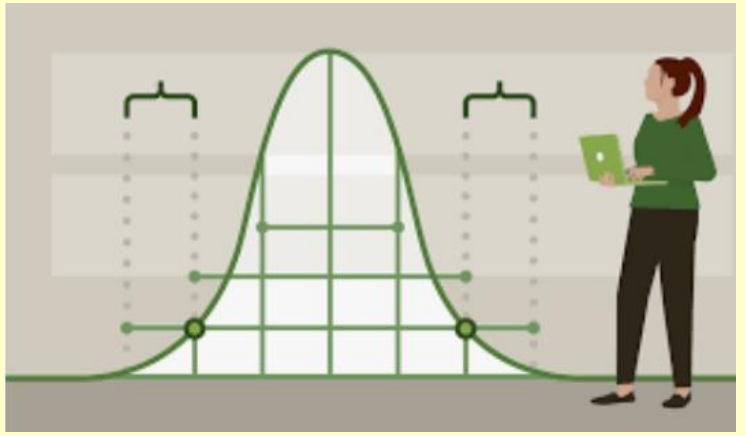
- **Modern computers has expedited to use statistics.**

1.1 Statistics and Data Science

▪ Statistics

- **Modern statistics is the discipline**
  => efficiently collect data, summarize data
  => analyze data to make scientific decisions using various probabilistic models in uncertain situations.

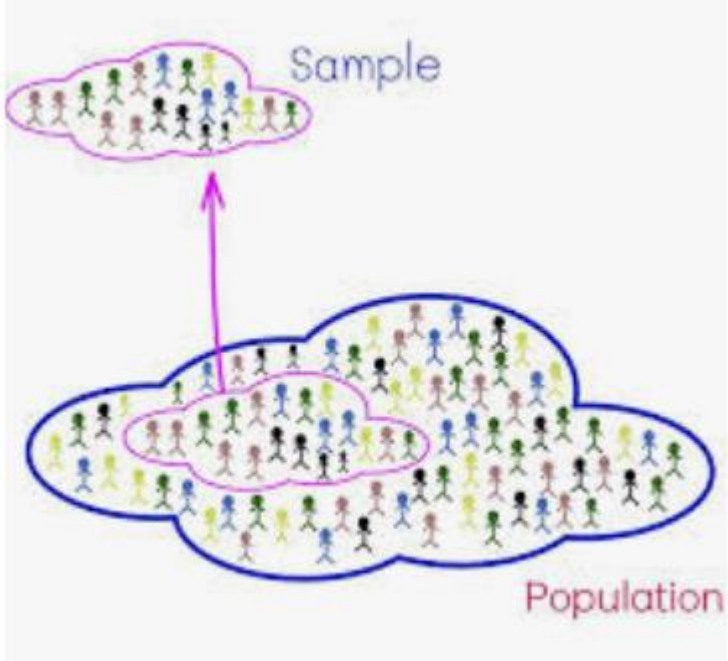
  - **Company predicts sales, government establish economic development plan**

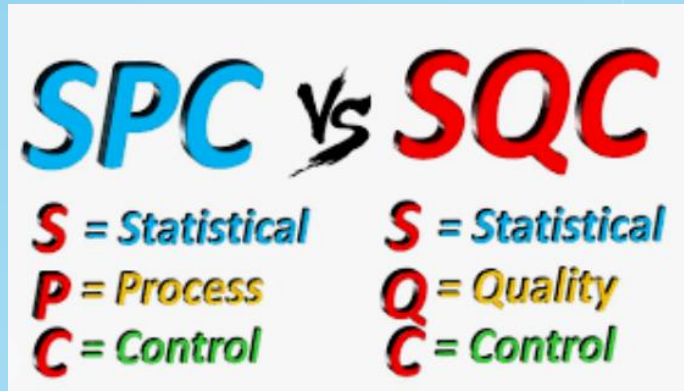


1.1 Statistics and Data Science

▪ Application of Statistics

- sample surveys to predict the winners of the election.
- Test new drug by a pharmaceutical company.
- Quality Control

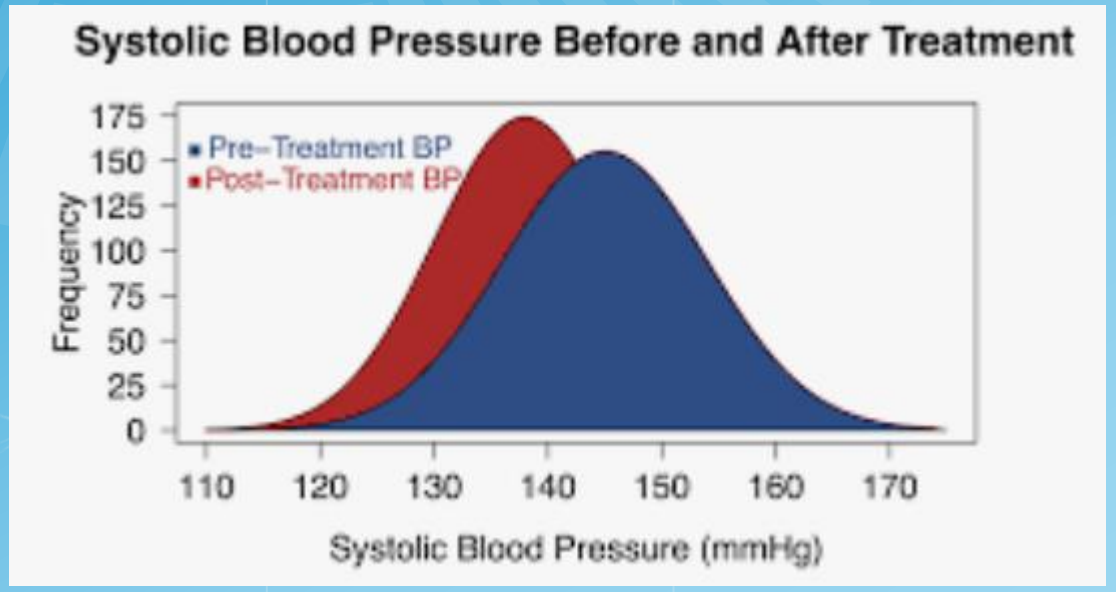




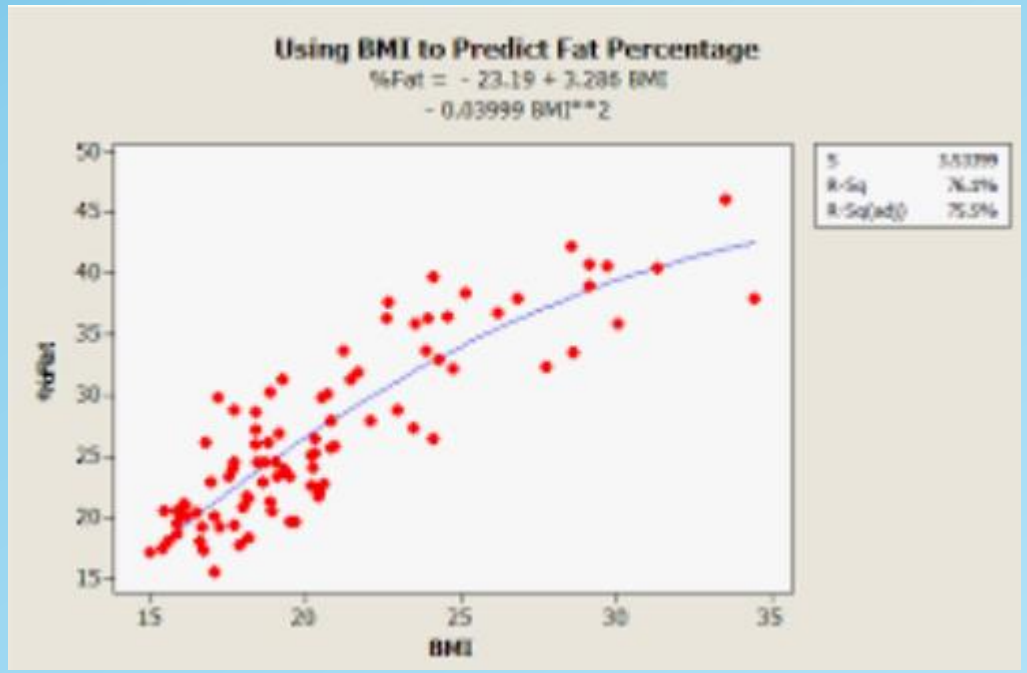
1.1 Statistics and Data Science

▪ Application of Statistics

- **Examine blood pressure before and after treatment**



- **Using BMI to predict FAT Percentage**



1.1 Statistics and Data Science

Data Science

Harvard Business Review

SEARCH

THE MAGAZINE BLOGS VIDEO BOOKS CASES WEBINARS COURSE

Guest Subscribe today and get access to all current articles and HBR online archi

THE MAGAZINE
October 2012

ARTICLE PREVIEW To read the full article, sign in or register. HBR subscribers, click here to register for FREE access »

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

Comments (97)

RELATED

Executive Summary

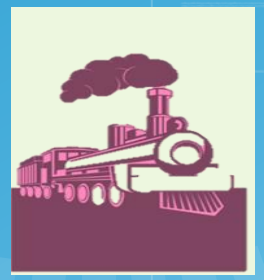
ALSO AVAILABLE
- Buy PDF

1.1 Statistics and Data Science

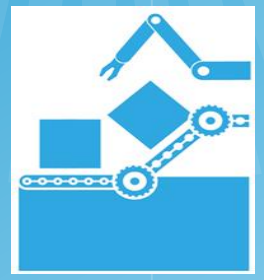
Data Science

1.1 Statistics and Data Science

▪ Evolving Industrial Revolution



18th Century
1st Industrial Revolution



19th –Early 20th Century
2nd Industrial Revolution

Late 20th Century -
3rd Industrial Revolution

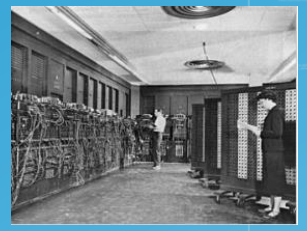
4th Industrial Revolution is under going by using Big Data

- Artificial Intelligence (AI)
- Internet of Things (IoT)
- Hyper-forecasting

Automatic driving car
3D printing
Virtual Reality
Alpha Go
....

1.1 Statistics and Data Science

▪ Evolving computer and tele-communication technology



- 1946 Modern Digital Computer(ENIAC) by Eckert and Mouchly of Univ of Pennsylvania

- 1981 IBM Personal Computer
- Microsoft Operating System by Bill Gates

The CERN data centre in 2010 housing some WWW servers

- 1990s Networking of computers in the world
- World Wide Web by Berners-Lee
- Google search engine
- Yahoo, MSN web portal

Two smartphones: a Samsung Galaxy J5 (left) and an iPhone 6S (right)

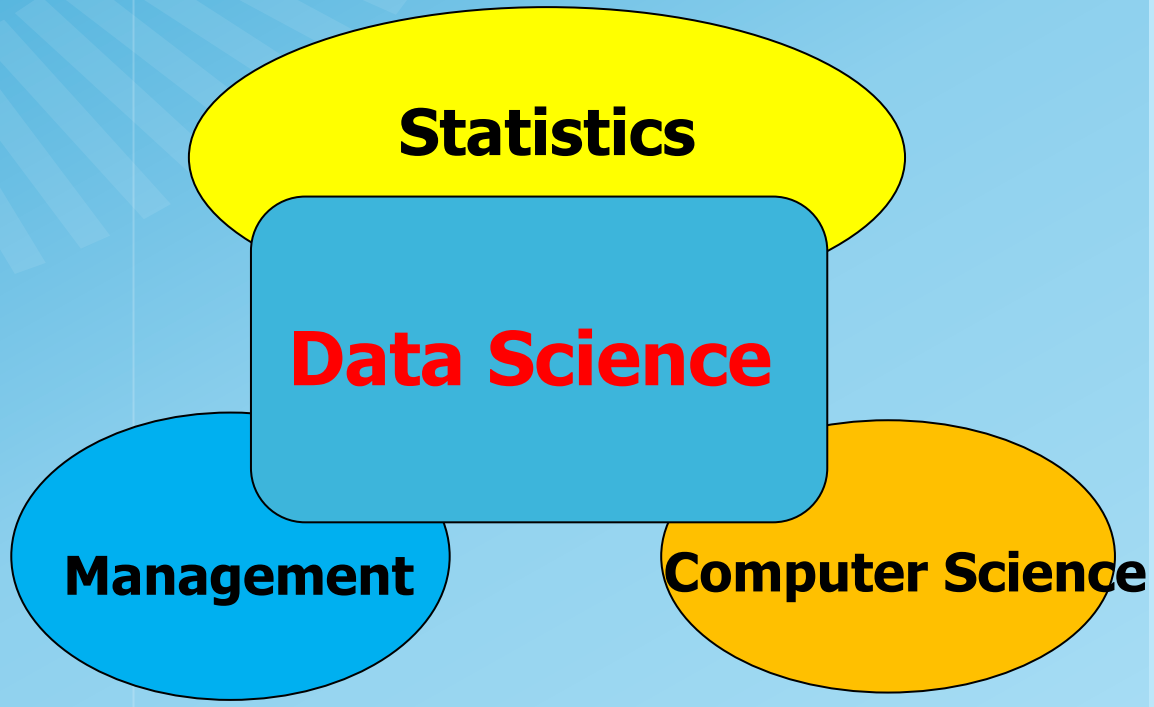
- 2000s Smartphone = PC + Phone
- www + wireless connection of Smartphone
- YouTube, Facebook, Twitter, LinkedIn

Big Data

- SNS Data
- web log data
- Bank data
- credit card data
- Health data

1.1 Statistics and Data Science

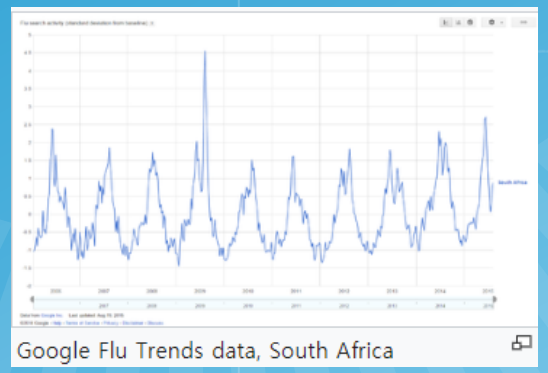
- Data Science is to collect big data, analyze and apply it in real life
  ⇨ Data Science is a fusion of several science



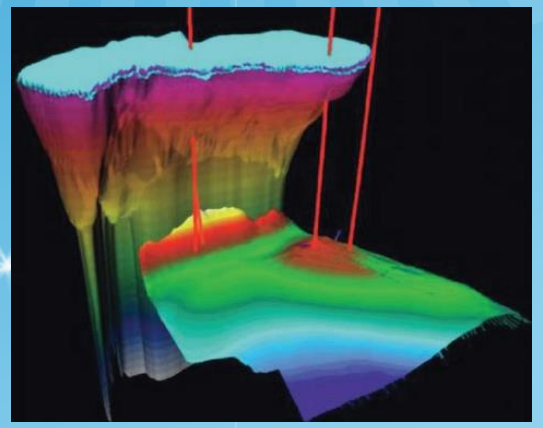
- Probability
- Estimation
- Testing
- Sampling
- Multivariate Stat Anal
- Database
- Information Retrieval
- Distributed Computing
- Artificial Intellignece
- Pattern Recognition
- Machine Learning
- Optimization
- MIS
- Marketing

1.1 Statistics and Data Science

❖ Example of Data Science



- Google Flu Trend to estimate influenza activity



- Crude oil exploration

- Market basket analysis

- Car insurance fraud detection

=> This book introduces Statistics toward Data Science

Thank you