

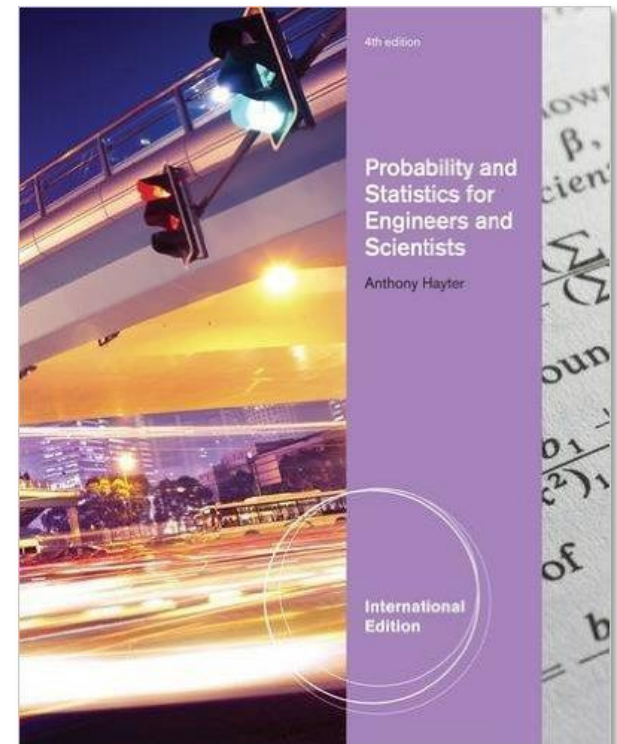
IE208: Applied Statistics

Course Overview

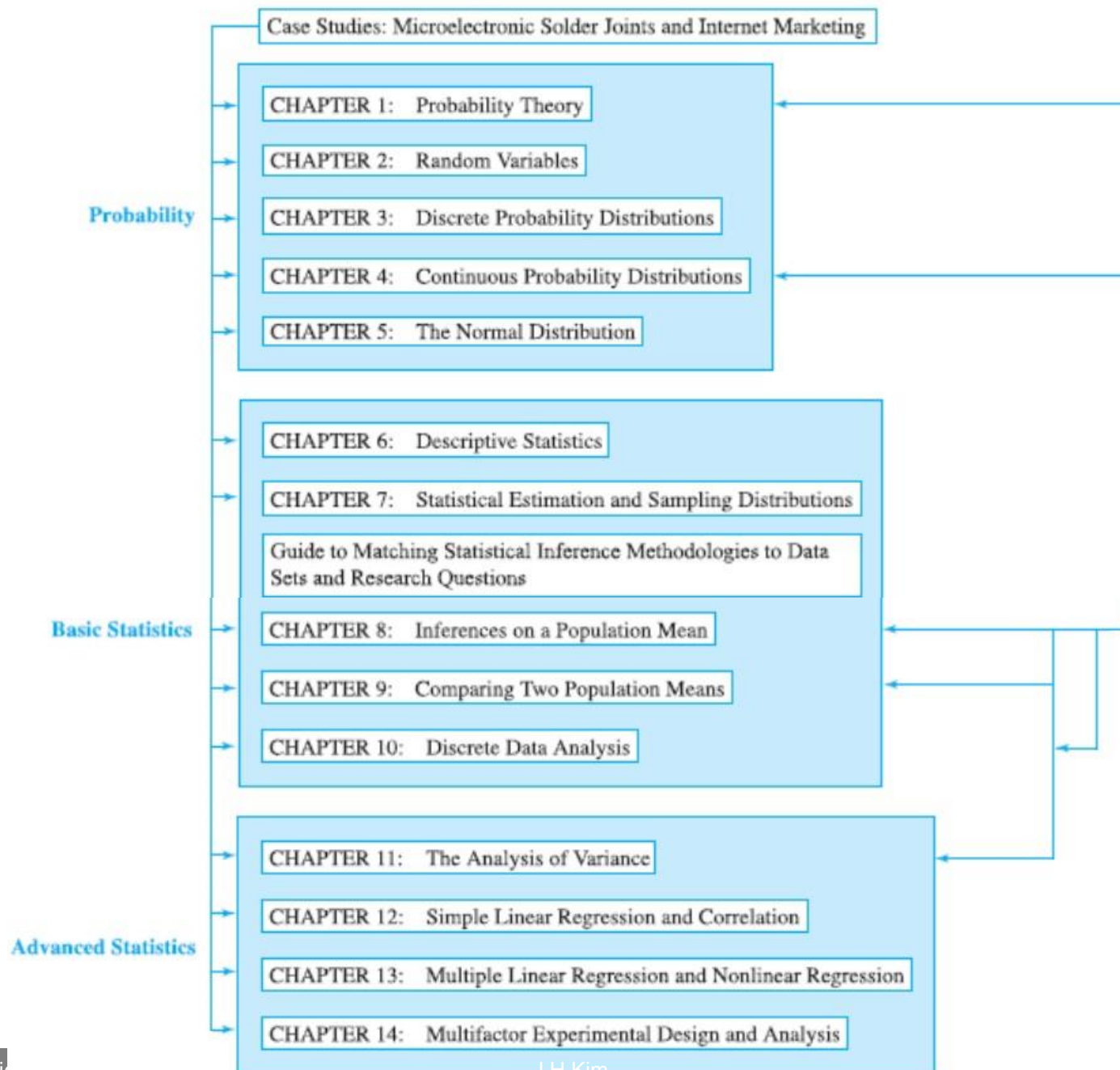
- IE208: Applied Statistics (응용통계학)
 - Sequel to IE207 (실험통계학)

Textbook

- Anthony J. Hayter (2013). *Probability and Statistics for Engineers and Scientists* (Fourth Edition, International Edition)
 - IE207: 이공학도를 위한 확률과 통계 (한글)
 - Using the Korean version is allowed,
But assignment problems will be based
on the English version



Composition of the Book



CHAPTER 1: Probability Theory

CHAPTER 2: Random Variables

CHAPTER 3: Discrete Probability Distributions

CHAPTER 4: Continuous Probability Distributions

CHAPTER 5: The Normal Distribution

IE207

CHAPTER 6: Descriptive Statistics

CHAPTER 7: Statistical Estimation and Sampling Distributions

CHAPTER 8: Inferences on a Population Mean

CHAPTER 9: Comparing Two Population Means

CHAPTER 10: Discrete Data Analysis

CHAPTER 11: The Analysis of Variance

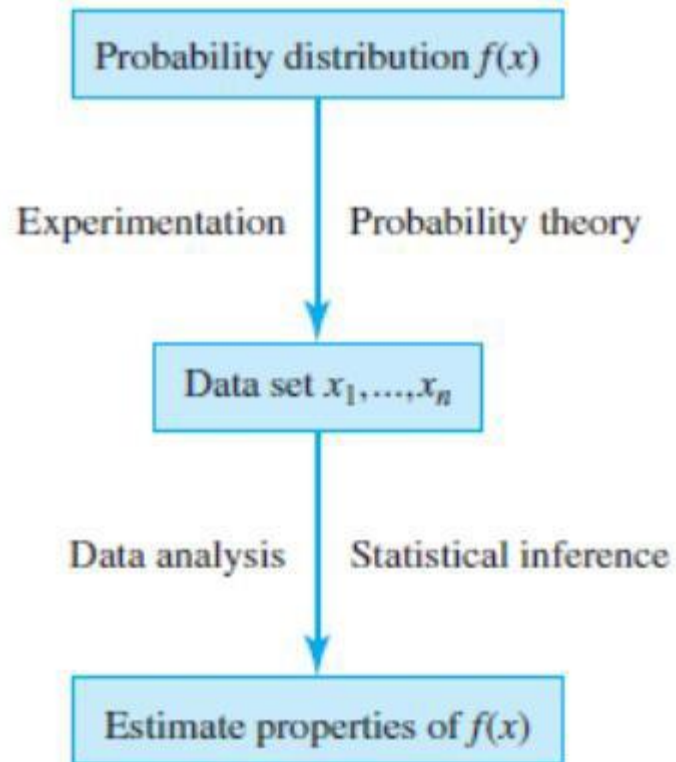
CHAPTER 12: Simple Linear Regression and Correlation

CHAPTER 13: Multiple Linear Regression and Nonlinear Regression

CHAPTER 14: Multifactor Experimental Design and Analysis

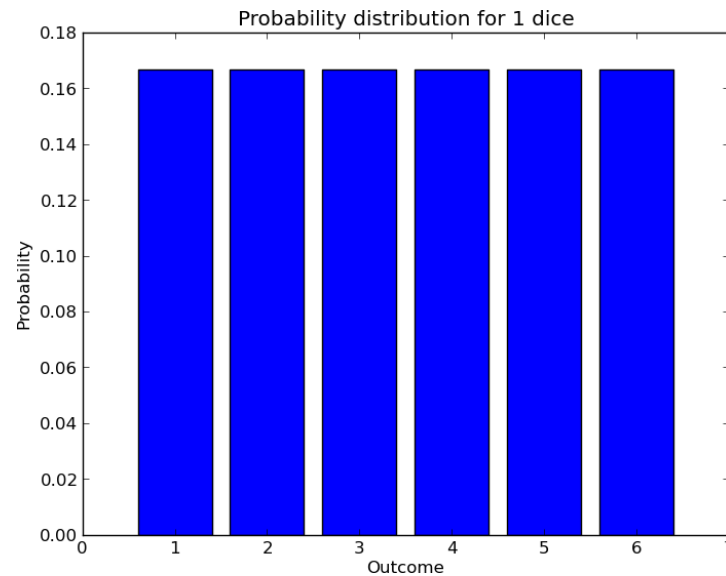
IE208

Statistics



Statistics

- *Example:* Outcome of rolling a fair die



We know exactly what could happen

Statistics

- *Example:* What about an unfair die?

How can we find its probability distribution?



Statistics

- *Example:* What about an unfair die?
 - ➔ Throw the die many times and collect outcomes

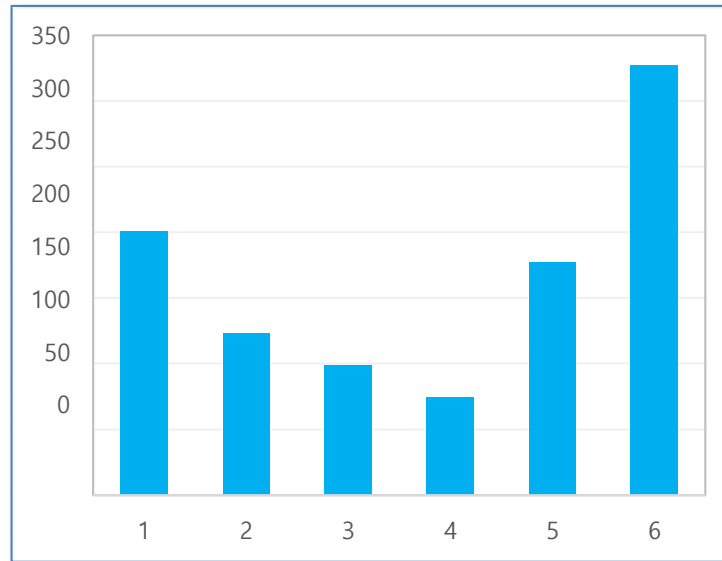
Suppose we threw 1,000 times

	Occurrences
1	201
2	123
3	98
4	74
5	177
6	327

Statistics

- *Example:* What about an unfair die?
➔ Throw the die many times and collect outcomes

Then we can guess the distribution to be the following



Statistics

- *Example:* What about an unfair die?

How can we find its probability distribution?



➔ *Throw the die many times to understand the possible outcomes
(we can only estimate what the actual distribution looks like)*

Statistics

- Why do we need to study statistics?
 (1) The study of statistics is used in our daily lives
 - *Example:* sports, weather, poll, etc.



Statistics

- Why do we need to study statistics?
 - (1) The study of statistics is used in our daily lives
 - *Example:* sports, weather, poll, etc.

“... 통계적으로는 유의미하지 않음을 알 수 있다....”

내일 선거일이라면 누구 찍겠나?

문재인 44.1% > 안철수 35.9%

최경호 기자 squeeze@joongang.co.kr



가상 양자대결에서는 安 46.3% vs 文 45.8% 오차범위 내 초접전...전문가들 “이대로가면 文 유리, 安 다
자대결시 필승카드 절실”

월간중앙이 여론조사 전문기관 타임리서치와 공동으로 제19대 대통령선거와 관련한 여론조사를 실시했다.
여론조사는 4월 12일 하루 동안 전국의 만 19세 이상 성인남녀 1002명을 대상으로 실시됐다. 조사는 임의
전화걸기(RDD)를 통한 자동응답(ARS) 방식을 택했으며, 100% 휴대전화로 표본을 추출했다. 95% 신뢰
수준에 표본오차는 $\pm 3.1\%$ 포인트다. 통계보정은 2016년 12월 말 행정자치부 발표 주민등록 인구를 기반으
로 성·연령·지역별 가중값을 부여했다.(※그 밖의 자세한 사항은 중앙선거여론조사 공정심의위원회 홈페이지
참조)

홍준표 후보가 보수진영 단일후보가 될 경우 유승민 후보 지지층은 30.0%가 지지 후보 선택을 포기한
가운데 51.0%가 안철수 후보로 지지를 이동하는 것으로 나타났다.

정현복 타임리서치 책임연구원은 “유승민 후보의 지지도는 2.6%에 불과한 만큼 통계적으로 유의미한
분석 결과라고 보기 어려울 수 있다”며 “보수표가 결집되기는 커녕 흩어지는 상황에서 보수 후보 단일화
가 얼마나 파괴력을 가질지 의문”이라고 설명했다.

코로나19 무증상환자, 유증상환자와 바이러스 배출양 차이 없어

이은정 순천향대서울병원 교수팀 303명 조사 "무증상 환자 관리 중요"

이병문 기자 | 입력 : 2020.08.09 14:08:18 수정 : 2020.08.09 14:11:05

코로나19(신종 코로나바이러스감염증)는 무증상 환자와 유증상 환자의 바이러스 배출량이 비슷해 무증상 환자도 철저한 관리가 필요한 것으로 나타났다.

연구팀은 총 303명의 코로나19 확진자를 유증상그룹(214명)과 무증상그룹(89명)으로 나눠 유전자증폭(RT-PCR)검사 결과를 통해 바이러스 배출 양을 비교했다. 중위 연령은 25세로 젊은 편이었고, 12명만 동반질환이 있었다. 303명 중 193명이 격리 시점부터 증상을 보였고, 입소시 110명이 무증상이었지만 입소 후 21명이 새롭게 증상이 나타났다. 새롭게 증상이 나타나는 기간은 평균 15일이었고, 짧게는 13일, 길게는 20일인 경우도 있었다. 89명은 퇴소까지 무증상이었다.

검사는 입소일로부터 8일, 9일, 그리고 15일, 16일째에 상기도 검체(비인두 및 구인두 표본) 및 하기도 검체(객담)에 대한 RT-PCR검사를 시행했다. 임상주의 판단에 따라 추가적으로 10, 17, 18, 19일째에도 검사를 시행하고 바이러스의 배출량을 측정하는 Ct(Cycle threshold) 값까지 확인했다.

연구 기간 동안 무증상확진자의 Ct값이 증상 확진자의 Ct값과 비슷해 증상에 상관없이 바이러스가 배출되는 양의 차이가 없었음을 알 수 있었다. RT-PCR 검사가 양성에서 음성이 되는 음전 기간은 확진일로부터 무증상 그룹이 17일, 유증상 그룹이 19.5일로 두 그룹 간 통계적으로 유의미한 차이가 없었다.

Yet another study shows hydroxychloroquine doesn't work against Covid-19

By **Elizabeth Cohen**, Senior Medical Correspondent

🕒 Updated 2057 GMT (0457 HKT) May 11, 2020

(CNN) — A new study -- the largest of its kind -- shows that hydroxychloroquine, the drug touted by President Trump, does not work against Covid-19 and could cause heart problems.

The study was published Monday in the [Journal of the American Medical Association](#). It follows a study published Thursday in the New England Journal of Medicine that also showed the drug doesn't fight the virus.

Even before these reports were published, the [US Food and Drug Administration](#) and the [National Institutes of Health](#) issued warnings about using the drugs for coronavirus patients.

In this group, doctors prescribed both drugs to 735 patients, just hydroxychloroquine to 271 patients, just azithromycin to 211 patients and neither drug to 221 patients.

They found that those taking hydroxychloroquine, either alone or with the antibiotic, were sicker than other patients to begin with and as time went on had a higher death rate. However, once the researchers statistically adjusted for the fact that the patients who took the drugs were sicker to start with, there was **no statistical significance** between the two death rates.

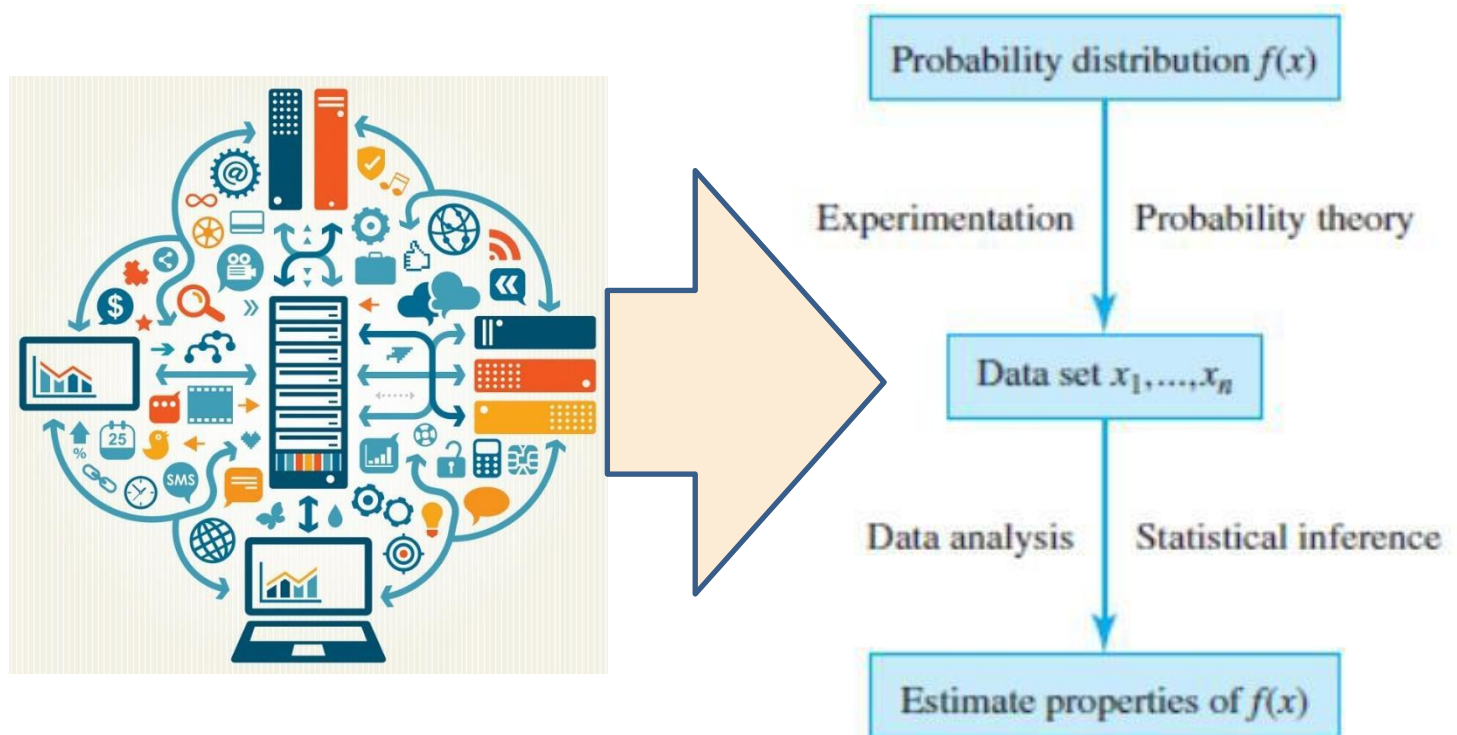
Statistics

- Why do we need to study statistics?
 - (2) With more *big data*, statistics become more important
 - *Example*: marketing, customer satisfaction, SNS, etc.
 - Data analysis is very important in all industries



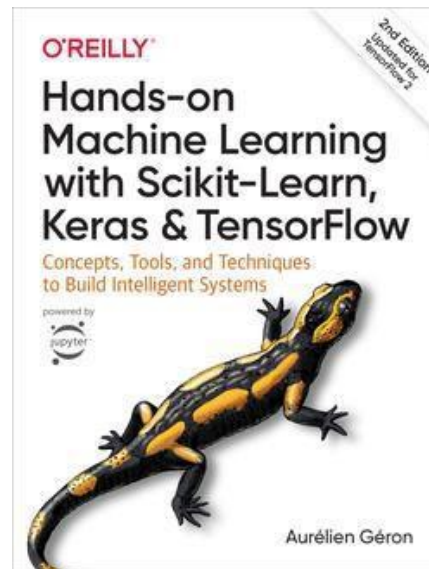
Statistics

- Why do we need to study statistics?
- (2) With more *big data*, statistics become more important



Statistics

- Why do we need to study statistics?
 - (3) Advanced topics in industrial engineering require statistics
 - Example: Machine learning



4. Training Models

Linear Regression

The Normal Equation

Computational Complexity

Gradient Descent

Batch Gradient Descent

Stochastic Gradient Descent

Mini-batch Gradient Descent

Polynomial Regression

Learning Curves

Regularized Linear Models

Ridge Regression

Lasso Regression

Elastic Net

Early Stopping

Logistic Regression

Estimating Probabilities

Training and Cost Function

Decision Boundaries

Softmax Regression

Exercises

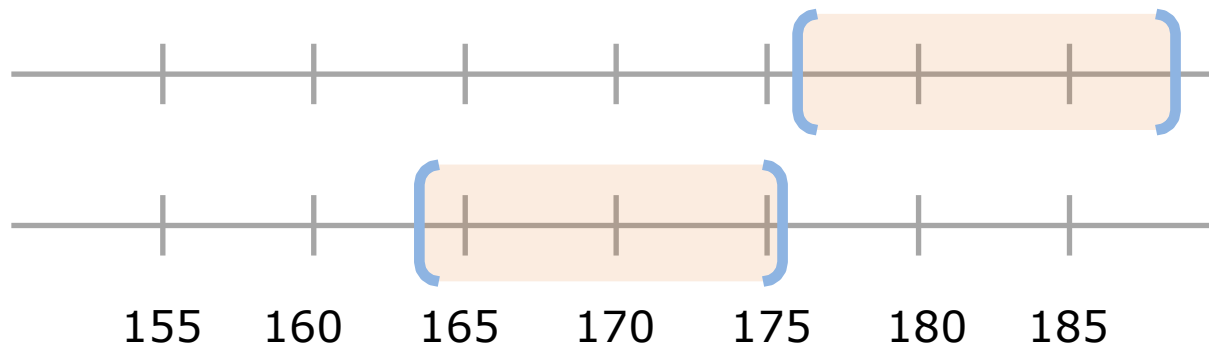
Course Material

Main Topics

- **Confidence intervals**

- *Example:* Suppose we want to know the average height of college students. Instead of attempting to find the actual value, we only want to estimate an interval.

Which one seems more plausible?



Confidence interval: 신뢰구간

Main Topics

- **Confidence intervals**

- *Example:* Suppose we want to know the average height of KHU students. Instead of attempting to find the actual value, we only want to estimate an interval.

How do we calculate a reasonable interval?

We collect the height of 60 students

$$(height_1, height_2, \dots, height_{60}) = (160, 176, \dots, 182)$$

and calculate an interval based on the sample data

➔ *“The average will be between 164.3cm and 171.0cm with 90% confidence”*

Main Topics

- **Hypothesis testing**

- *Example:* Suppose we want to know the average height of KHU students.

Bobby thinks that the average height is less than 163 cm, but you disagree. How can you show that Bobby is wrong?

Again, we first need to collect data

$$(height_1, height_2, \dots, height_{60}) = (160, 176, \dots, 182)$$

Hypothesis test: 가설검정

Main Topics

- **Hypothesis testing**

- *Example:* Suppose we want to know the average height of KHU students.

Bobby thinks that the average height is less than 163 cm, but you disagree. How can you show that Bobby is wrong?

Then perform one-sided hypothesis test based on data:

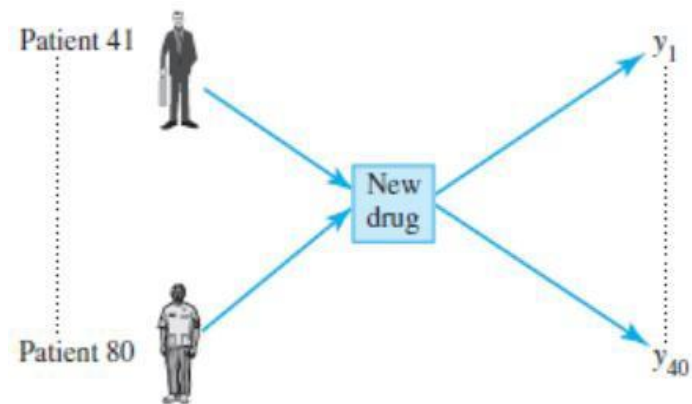
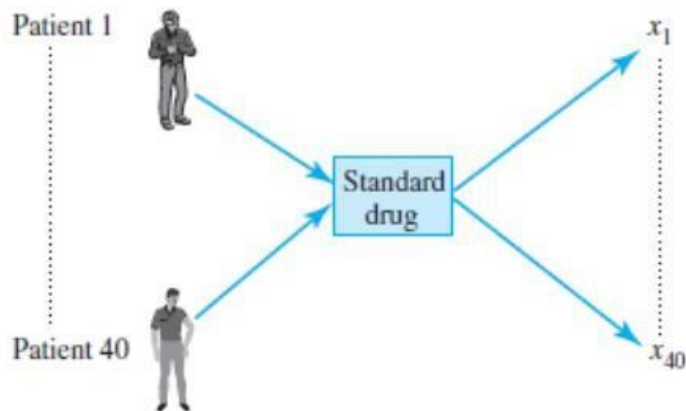
$$\mu \leq 163 \quad \text{versus} \quad \mu > 163$$

➔ Measure the plausibility of $\mu \leq 163$

➔ If the plausibility is very small, we can reject Bobby's claim

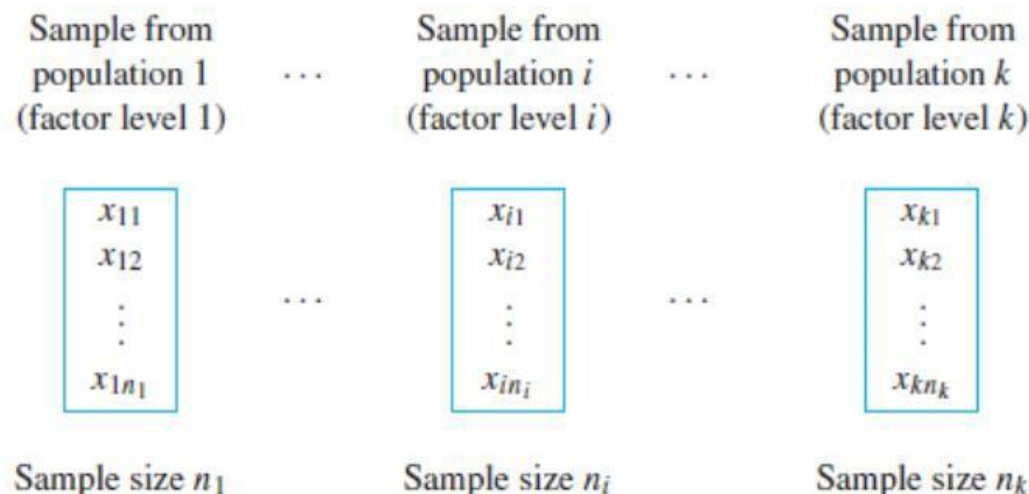
Main Topics

- But sometimes, we consider more than one set of data
 - **Comparing two populations means**
 - *Example:* Is a new drug better than the previous standard drug?
- We can begin by collecting samples from each of the two groups



Main Topics

- But sometimes, we consider more than one set of data
- **Comparing three or more populations means**
 - *Example:* If there are many drugs that claim to treat high fever, can we check if one is better than the rest?



Main Topics

- What if data are not continuous?

- **Comparing discrete data**

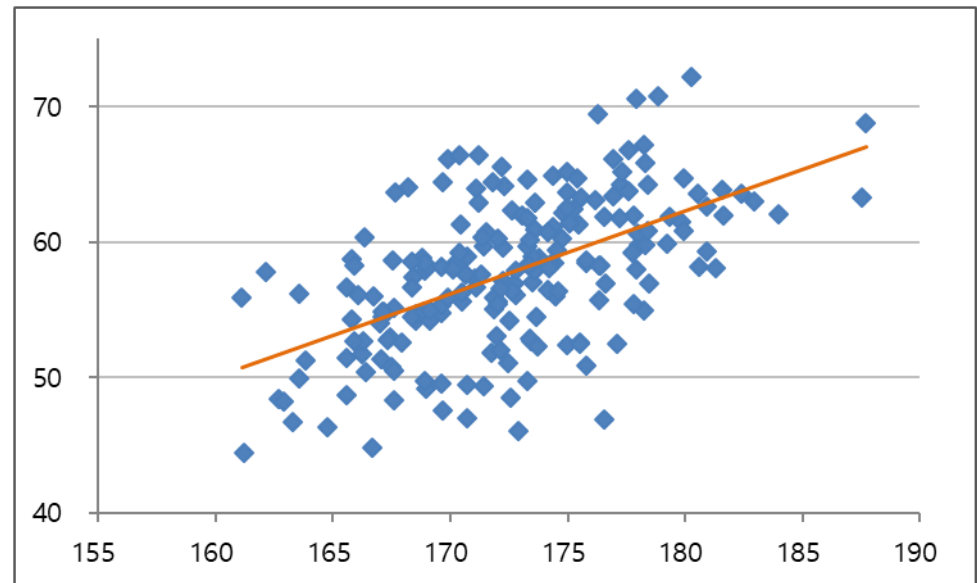
Example: (Dart practice)

- Use a simple dartboard and record each throw as ‘Center’, ‘Target’, or ‘Miss’
- Is there strong evidence that the thrower hits the ‘Center’ over 80% of the time?



Main Topics

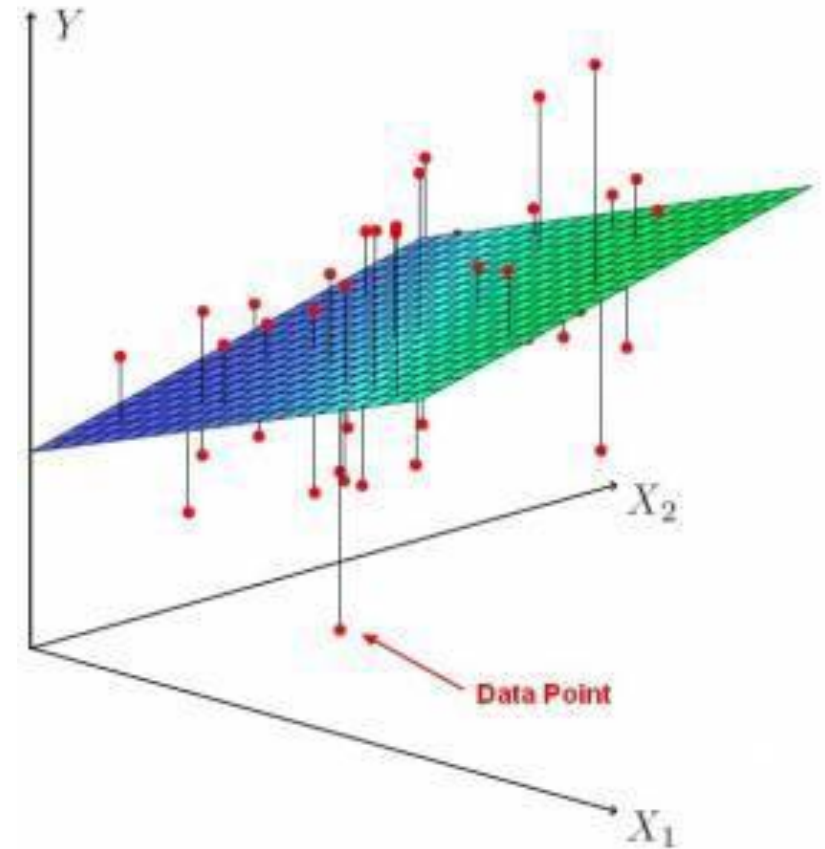
- **Linear regression**
 - Simple linear regression
- *Example: Height and weight*
 - x : Height (cm)
 - y : Weight (kg)



Regression analysis: 회귀분석

Main Topics

- **Linear regression**
 - Multiple linear regression
 - $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$



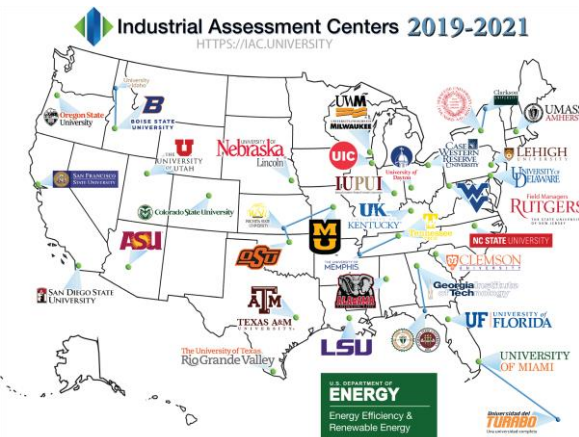
About Dr. Jeon

Brief Biography

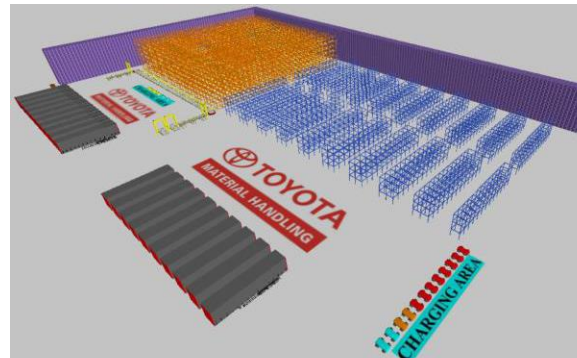
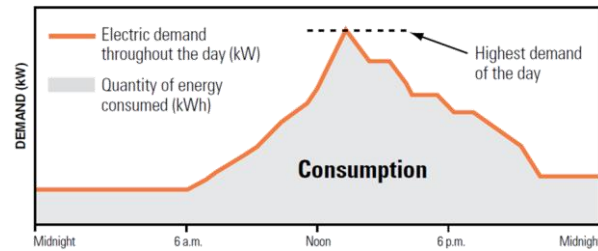
- **B.E. from Korea University (2001) – IE Major**
- **Samsung Networks (2001 – 2009) – Computer Network Engineer**
- **M.S. & Ph.D. from Penn State University (2013, 2015) – IE Major**
- **Louisiana State University (2015 – 2021) – Assistant Professor in IE**
- **Kyung Hee University (2021 – Present) – Assistant Professor in IMSE**
- **Research Area & Methodologies**
 - Manufacturing System & Energy
 - Simulation, Applied Probability, Machine Learning

Research Topics

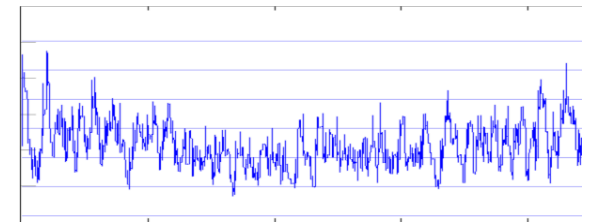
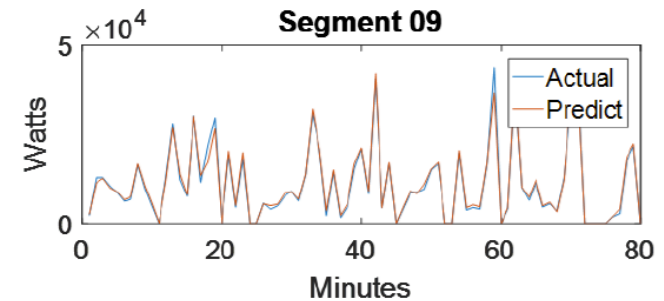
Industrial Assessment Center



Manufacturing Energy

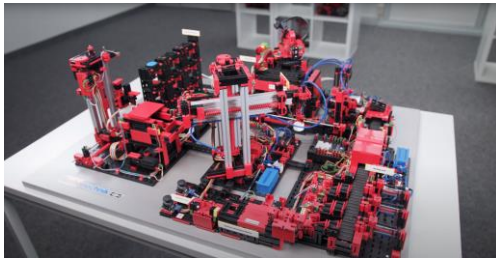


Machine Learning for Energy Cost Reduction



Research Topics

Machine Learning with on a Physical Manufacturing System



Interested? I'm looking for students to work together!!

Let's work together!! Please email me at hwjeon@khu.ac.kr

Syllabus

IE208: Schedule

Week 1 Introduction / Review of Chapters 1-7

Week 2	8. Inferences on a Population Mean
--------	------------------------------------

Week 3 8. Inferences on a Population Mean

No lecture on 9/21 (추석 Holiday)

Week 4	9. Comparing Two Population Means
--------	-----------------------------------

Week 5 9. Comparing Two Population Means

Week 6	10. Discrete Data Analysis
--------	----------------------------

Week 7 10. Discrete Data Analysis

Week 8	10. Discrete Data Analysis
--------	----------------------------

*** Midterm Exam (10/20 – 10/26)**

Week 9 11. The Analysis of Variance

Week 10	11. The Analysis of Variance
---------	------------------------------

Week 11 12. Simple Linear Regression

Week 12	12. Simple Linear Regression
---------	------------------------------

Week 13 12. Correlation

Week 14	13. Multiple Linear Regression
---------	--------------------------------

Week 15 13. Multiple Linear Regression


Week 16	* Final Exam (12/15 – 12/21)
---------	-------------------------------------

*** Exact date will be announced later**

Lecture

- IE208: Tue/Thu 16:30-17:45 (Zoom Online) ~~공학관 148~~
- Lecture material will be posted on e-campus (<https://e-campus.khu.ac.kr>)
- Partial English Class
 - English goes first, and Korean translation comes next.
 - You may ask questions in Korean during lectures and in emails.

Grading

- Midterm exam: 35% (*online/offline*)
 - Final exam: 50% (*online/offline*)
- 
- Dates will be announced later*
- Assignment: 5%
 - Exercise problems from the textbook
 - 2 assignments
 - Attendance: 10%
 - E-campus will be used
 - Random authentication numbers (인증번호) will be used
 - During lecture time

Attendance

- No grade penalty up to 4 absences
- 1% deduction in final grade for each additional absence
- Must attend quizzes and exams
- (5번째 결석부터 1% 감점. 시험 때 결석 시 0점 처리)
- 2 latenesses is counted as one absence

Absolute Grading:

- Absolute grading (not relative grading / no “grade on a curve”)

Details on Online Lectures

Lecture

- Lecture for this course will be conducted with Zoom online.
- Zoom URL will be the same for the entire semester.
 - <https://zoom.us/j/96807562146?pwd=clk5ZDhVcjFuTmQ2MGdmcmJEMkd2dz09>
 - Meeting ID: 968 0756 2146
 - Password: hb4LrE

If you have questions...

- If you have questions on the lecture material, then you can do one of the following,
 - ① Send me an email
 - ② Use e-campus (문의게시판) to ask questions, or
 - ③ Ask questions after each lecture
- Also, there are almost 100 students taking IE208, so individual email response can only be limited or delayed.

e-campus

- **At the beginning of each lecture time (Tue and Thu), please check announcements on e-campus**
 - There may be special announcements
 - There may be corrections to errors in lecture files
 - .. etc.
- Since all lecture are online, it is very important that everyone regularly checks announcements on e-campus
 - For example, date for midterm and final exams will be posted on e-campus

Finally...

- Any questions or concerns?
- Email: hwjeon@khu.ac.kr