

Health Information Management Systems

Data Science in Health

Hye-Chung Kum (kum@tamu.edu)
Associate Professor
Population Informatics Lab (<https://pinformatics.org/>)
Course URL: <http://pinformatics.org/phpm631>

License:
Health Information Technology by Hye-Chung Kum is licensed under a
[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/)



1

1

Outline: Health Care Data Use

- Operational vs. Decision Support Systems
- What is Data Science/Business Intelligence
 - What is Data Science?
 - What is Big Data?
 - Overview of Data Mining
- Understanding Data

2

2

Operational vs Decision Support Systems



- Operational Systems
 - Support day to day transactions
 - Contain current, "up to date" data
 - Examples: EMR, customer orders, inventory levels, payroll, bank account balances
- Decision Support Systems
 - Support strategic decision making
 - Contain historical, "summarized" data
 - Examples:
 - Clinical support: what treatment is best?
 - Population health
 - Management support: performance summary, market segmentation

3

Operational Application: EMR



The screenshot displays an EMR application interface. At the top, patient information is shown: "DOB: 04/26/2005 (14 years)", "Weight: 54.00 lb (24.49 kg)", "Allergies (7)", "Problems (8)", "Diagnoses (143)", "Medications (6)". The patient's name is "FCR Adams, Kerie PA" and the location is "Refering: Fenderberg, Colorado, Tennessee, MD".

The interface includes a "Configuration Navigation" sidebar on the left with options like "Home Page", "Framework Content", "Manage Template Links", "Office Diagnostics", "My Plans/Order Sets", "DE Picklist Setup", "Document Configuration", "Ficklist Setup", "Immunizations", "Clinical Guidelines", "Path Development Setup", "CE Education Setup", "Recommended Case", and "Dr Groups/Reference Systems".

The main area is titled "Table and Subnavigation" and contains a table of templates and sub-navigation options. The table has columns for "Caption", "Template", and "Update | Delete | Clear". The sub-navigation section below it has columns for "Caption" and "Template".

Caption	Template	Update Delete Clear
Home Page	Home Page	
Tab 1	Intake (card_intake)	
Tab 2	History (card_history)	
Tab 3	Procedures (card_procedure)	
Tab 4	SOAP (card_soap)	
Tab 5	Finalize (card_finalize)	
Tab 6	Checkout (ph_checkout)	
Tab 7	Template	

4

DSS (Reports)

- PricePoint
 - for consumers
- <http://www.txpricepoint.org/>

St Joseph Regional Health Center
 201 Franciscan Drive
 Bryan, TX 77802
 (979) 776-3777
 Additional Hospital Information

Oncology Section W/C CC/MCC
January 2013 – December 2013

	Selected Hospital	All Hospitals in this County	Hospitals with Similar Patient Care	All Texas Hospitals
● Number of Discharges:	264	725	80,374	88,597
● Average Length-of-Stay:	2.9 Day(s)	2.9 Day(s)	2.8 Day(s)	2.8 Day(s)
● Average Charge Per Day:	\$18,732	\$30,540	\$19,483	\$19,422
● Median Charge:	\$6,466	\$10,531	\$6,858	\$6,937
● Median Charge:	\$18,271	\$30,576	\$17,571	\$17,643

NR = 1-4 Discharges (Not Reported) ■ Notes About This Table
 * Show hospitals in that group ■ Understanding Hospital Charge Information
 ■ Why Charges May Differ Between Hospitals

ST JOSEPH REGIONAL HEALTH CENTER
CHARGE AND PAYMENT INFORMATION
MOST RECENT FISCAL YEAR - ALL SERVICES

What is the selected hospital's "payer mix"?
 A hospital's "payer mix" refers to the proportion of its total charges attributable to different types of insurance coverage.

How much do government programs pay compared to private insurance?
 In many cases, Medicare & Medicaid reimburse hospitals at rates that do not cover the costs they incur to provide care. Payments from privately insured patients generally subsidize the shortfalls created by Medicare and Medicaid and therefore represent a "hidden tax" on individuals and families not covered by government programs.

The graphs below represent all services provided by the hospital; they are not specific to the the selected service.

ALL PAYERS*	MEDICARE*	MEDICAID*
This hospital collects an average of 26% of its charges from all payers.	This hospital collects an average of 18% of its charges from Medicare.	This hospital collects an average of 12% of its charges from Medicaid.
□ \$912,174,398 Charges Not Paid ■ \$288,208,099 Charges Paid	□ \$392,733,282 Charges Not Paid ■ \$87,416,762 Charges Paid <small>*Payments determined by federal government</small>	□ \$98,135,695 Charges Not Paid ■ \$13,123,796 Charges Paid <small>*Includes Medicaid/Disproportionate Share and DFL payments</small>

The above information is for **all services** at the selected hospital. It is not specific to the service you selected or any other single service. Contact your insurer to determine the specific amount that will be paid under your policy for the selected service.

5

Outline: Health Care Data Use



- Operational vs. Decision Support Systems
- What is Data Science/Business Intelligence
 - What is Data Science ?
 - What is Big Data?
 - Overview of Data Mining
- Understanding Data

6

6

What is Data Science?


- Other words
 - Knowledge Discovery & Data mining (KDD)
 - Business Intelligence / Business Analytics
- **Collecting** and **refining** information from many sources
- **Analyzing** and **presenting** the information in useful ways
- So **people** can make better business **decisions**

7


7

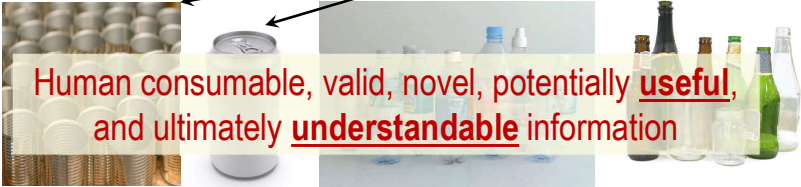
Data Science
Knowledge Discovery & Data mining (KDD)

Big Data : operational data



KDD: Clean, Merge, Reprocess





Human consumable, valid, novel, potentially **useful**,
and ultimately **understandable** information

8

8

NIST Big Data

POPULATION INFORMATICS

Data Science Definition (Big Data less consensus)

- **Data Science** is the extraction of actionable knowledge directly from data through a process of discovery, hypothesis, and analytical hypothesis analysis.
- A **Data Scientist** is a practitioner who has sufficient knowledge of the overlapping regimes of expertise in business needs, domain knowledge, analytical skills and programming expertise to manage the end-to-end scientific method process through each stage in the big data lifecycle.

Big Data refers to digital data volume, velocity and/or variety whose management requires scalability across coupled horizontal resources

9/29/13

IEEE BigData Overview October 9 2013

8

9

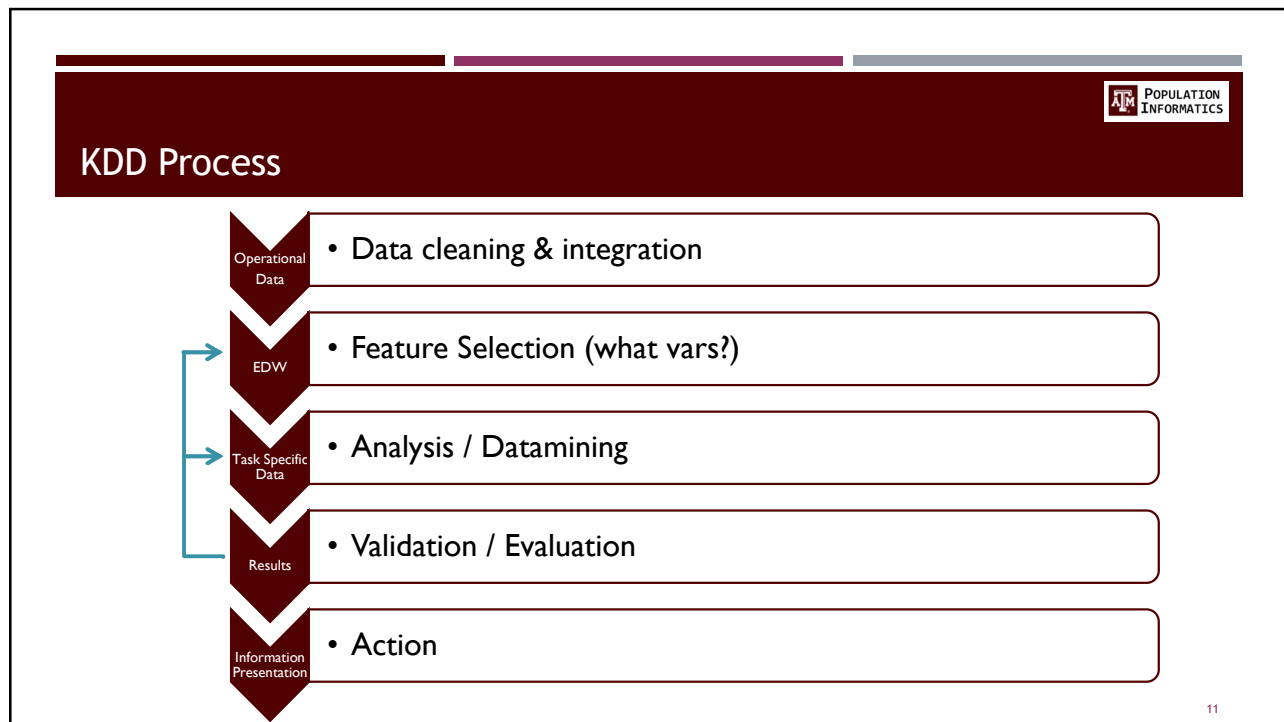
Data to Decision

POPULATION INFORMATICS

Hierarchy of Data Science & Different Types of Data Scientists
Hye-Chung Kum, Population Informatics Research Group

10

10



11

The slide is titled "Video" and lists three video recommendations. Each recommendation includes a title, a duration, and a YouTube link. The slide features a header with the text "Video" and the "ATM POPULATION INFORMATICS" logo. A small number "12" is located in the bottom right corner of the slide frame.

- TX Mental Health Landscape (2:46)
 - <https://www.youtube.com/watch?v=8dPqQt0yXJA>
- Wealth Inequality (1:30)
 - <https://www.youtube.com/watch?v=QPKKQnijnsM>
- Good managers know how to build the data story!

12




Outline: Health Care Data Use

- Operational vs. Decision Support Systems
- What is Data Science/Business Intelligence
 - What is Data Science?
 - What is Big Data?
 - Overview of Data Mining
- Understanding Data

13

13



Properties of BIG DATA : 4V

- Volume : constantly generating
- Velocity : constantly changing
- Variety : expressed in many ways
- Veracity : lots of errors
- (Value)

EXAMPLE: the INTERNET!
What do you do to find information/knowledge on the Internet?

14

14

POPULATION INFORMATICS

The Big Data Problem - Nutshelled

Michael Franklin (UC Berkley)

Something's gotta give:

15

15

POPULATION INFORMATICS

AMPLab: Integrating Three Key Resources

Algorithms

Machines

People

- Machine Learning, Statistical Methods
- Prediction, Business Intelligence

- Clusters and Clouds
- Warehouse Scale Computing

- Crowdsourcing, Human Computation
- Data Scientists, Analysts

16

16




Outline: Health Care Data Use

- Operational vs. Decision Support Systems
- What is Data Science/Business Intelligence
 - What is Data Science?
 - What is Big Data?
 - Overview of Data Mining
- Understanding Data

17

17




What is Data Mining?

- Using a combination of **artificial intelligence**, **machine learning**, and **statistical analysis** to analyze **data**
- and discover useful **patterns** that are “**hidden**” there

18

18




Business uses of data mining: Essentially five tasks

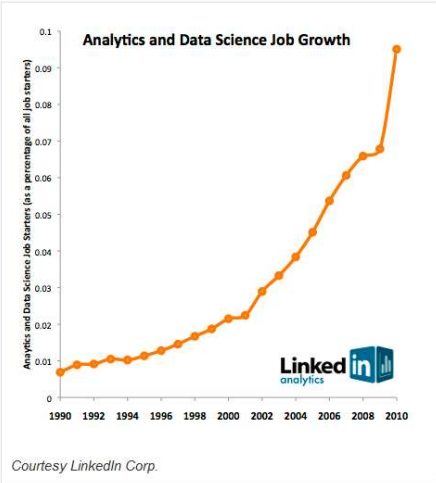
- **Classification:** Group data into predetermined categories
 - Classify credit applicants as low, medium, high risk
 - Classify insurance claims as normal, suspicious
- **Estimation:** Estimate probability of an event through models built from previous data
 - Estimate the probability of a direct mailing response
 - Estimate the potential cohort size for a clinical trial
- **Prediction:** Predict an outcome based on input based on models built from previous data
 - Predict which customers will leave within six months
 - Predict which patient will return to the ED
- **Affinity Grouping:** Group people based on similar characteristics
 - Find out what books to recommend to Amazon.com users
 - Find treatment regime that was successful for similar patient
- **Description**
 - Help understand large volumes of data by uncovering interesting, useful, and actionable patterns

19

19



Job market of data scientists




Courtesy LinkedIn Corp.

- statisticians will be the next sexy job
 - Google Chief Economist Hal Varian
- shortage of 190,000 data scientists by the year 2019
 - McKinsey Global Institute

20

20




Outline: Health Care Data Use

- Operational vs. Decision Support Systems
- What is Data Science/Business Intelligence
 - What is Data Science?
 - What is Big Data?
 - Overview of Data Mining
- Understanding Data

21

21



Applications in Health

- A March 2014 poll from MeriTalk and EMC found that **63 percent of healthcare executives** in the federal government believe that **big data will improve population health management**
- Examples
 - Manage population health
 - Accountable Care Organizations (ACO)
 - Clinical decision support
 - Cohort identification for clinical trials
 - Medical fraud detection

22

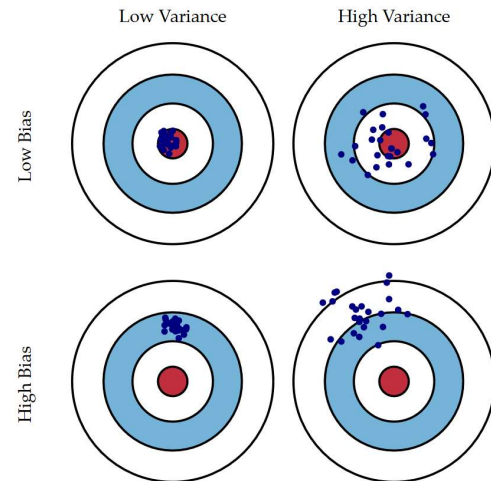
22

Bias and Variance



<http://scott.fortmann-roe.com/docs/BiasVariance.html>

- precise but not valid?
- What is real data like?
- Adjust for bias
- Take into account variance

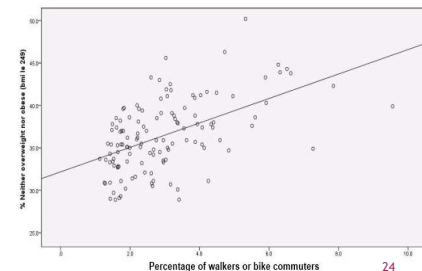
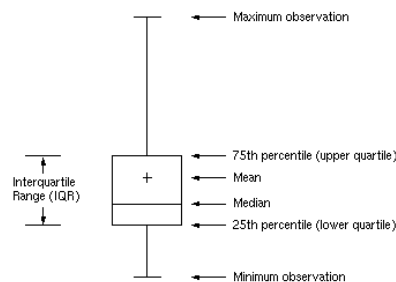
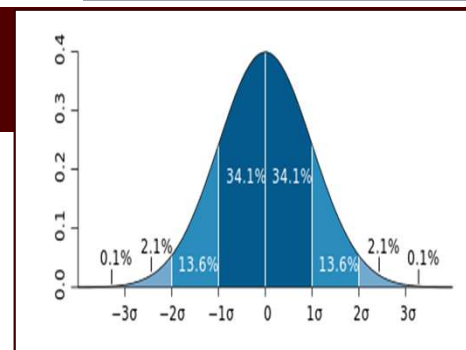


23

23

Numerical Data : distribution

- Mean
- Standard Deviation
 - How dispersed
- Range: Max/Min
- Median (percentile)
- Scatter Plot: 2 vars



24

24

Categorical Data

- Tabulation
- Cross tabulation
 - 2 variables
- GIS: maps

Frequency of visit by Age

Age Group	Daily	Twice a week	Weekly	Monthly
Under 18	9	11	16	23
18-24	5	8	16	22
25-34	5	8	16	13
35-44	0	7	16	13
45-54	0	2	4	16
55+	0	1	4	3

Frequency of visit

Frequency	Count	Percentage
Daily	19	21%
Twice a week	35	39%
Weekly	64	71%
Monthly	87	97%

25

Take Away I What is Data Science? KDD Process

Operational Data

EDW

Task Specific Data

Results

Information Presentation

- Data cleaning & integration
- Feature Selection (what vars?)
- Analysis / Datamining
- Validation / Evaluation
- Action

26

Take Away II What is Big Data ?



■ 4 Vs of Big Data

- Volume : lots of data
- Velocity : constantly generating & changing
- Variety : expressed in many ways
- Veracity : lots of errors
- (Value)



■ Big Data Problems

- Time
- Money
- Quality (Precision)

■ Three Resources: AMP

- Algorithm
- Machine
- People

27

27

Take Away III Business uses of data mining: Essentially five tasks



■ Classification

- Classify credit applicants as low, medium, high risk
- Classify insurance claims as normal, suspicious

■ Estimation

- Estimate the probability of a direct mailing response
- Estimate the potential cohort size for a clinical trial

■ Prediction

- Predict which customers will leave within six months
- Predict which patient will return to the ED

■ Affinity Grouping


- Find out what books to recommend to Amazon.com users
- Find treatment regime that was successful for similar patient

■ Description

- Help understand large volumes of data by uncovering interesting, useful, and actionable patterns



28

28




How do you get good with data ?

- Sorry, no short cuts. Build experience.
- In this course, start you out.
 - Tableau / Excel
 - SQL
 - Assignment 1
 - Labs



29

29



Reminder: due next two weeks

- Lab 1: most of you should be done during class
- Assignment 1: submit on E-campus day before class
 - Week one: progress report
 - Week two: Final Tutorial
- Readings: Chapters 1 & 2
- Quiz 1 (E-campus: posted on Tues)
 - Practice quiz
- Group presentation emails

30

30