

What is Probability and Statistics and Why Should You Care?

CS 3130: Probability and Statistics for Engineers

August 26, 2014

What is Probability?

What is Probability?

Definition

Probability theory is the study of the mathematical rules that govern random events.

What is Probability?

Definition

Probability theory is the study of the mathematical rules that govern random events.

But what is randomness?

What is Probability?

Definition

Probability theory is the study of the mathematical rules that govern random events.

But what is randomness?

Informally, a **random event** is an event in which we do not know the outcome without observing it.

What is Probability?

Definition

Probability theory is the study of the mathematical rules that govern random events.

But what is randomness?

Informally, a **random event** is an event in which we do not know the outcome without observing it.

Probability tells us what we can say about such events, given our assumptions about the possible outcomes.

What is Statistics?

What is Statistics?

Definition

Statistics is the application of probability to the collection, analysis, and description of random data.

What is Statistics?

Definition

Statistics is the application of probability to the collection, analysis, and description of random data.

Statistics is used to:

- ▶ **Design** experiments

What is Statistics?

Definition

Statistics is the application of probability to the collection, analysis, and description of random data.

Statistics is used to:

- ▶ **Design** experiments
- ▶ **Summarize** data

What is Statistics?

Definition

Statistics is the application of probability to the collection, analysis, and description of random data.

Statistics is used to:

- ▶ **Design** experiments
- ▶ **Summarize** data
- ▶ **Make conclusions** about the world

What is Statistics?

Definition

Statistics is the application of probability to the collection, analysis, and description of random data.

Statistics is used to:

- ▶ **Design** experiments
- ▶ **Summarize** data
- ▶ **Make conclusions** about the world
- ▶ **Explore** complex data

Applications of Probability and Statistics

Computer Science:

Electrical Engineering:

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning

Electrical Engineering:

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining

Electrical Engineering:

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence

Electrical Engineering:

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation

Electrical Engineering:

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing

Electrical Engineering:

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing
- ▶ Computer Graphics

Electrical Engineering:

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing
- ▶ Computer Graphics
- ▶ Visualization

Electrical Engineering:

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing
- ▶ Computer Graphics
- ▶ Visualization
- ▶ Software Testing

Electrical Engineering:

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing
- ▶ Computer Graphics
- ▶ Visualization
- ▶ Software Testing
- ▶ Algorithms

Electrical Engineering:

- ▶ Signal Processing

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing
- ▶ Computer Graphics
- ▶ Visualization
- ▶ Software Testing
- ▶ Algorithms

Electrical Engineering:

- ▶ Signal Processing
- ▶ Telecommunications

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing
- ▶ Computer Graphics
- ▶ Visualization
- ▶ Software Testing
- ▶ Algorithms

Electrical Engineering:

- ▶ Signal Processing
- ▶ Telecommunications
- ▶ Information Theory

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing
- ▶ Computer Graphics
- ▶ Visualization
- ▶ Software Testing
- ▶ Algorithms

Electrical Engineering:

- ▶ Signal Processing
- ▶ Telecommunications
- ▶ Information Theory
- ▶ Control Theory

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing
- ▶ Computer Graphics
- ▶ Visualization
- ▶ Software Testing
- ▶ Algorithms

Electrical Engineering:

- ▶ Signal Processing
- ▶ Telecommunications
- ▶ Information Theory
- ▶ Control Theory
- ▶ Instrumentation, Sensors

Applications of Probability and Statistics

Computer Science:

- ▶ Machine Learning
- ▶ Data Mining
- ▶ Artificial Intelligence
- ▶ Simulation
- ▶ Image Processing
- ▶ Computer Graphics
- ▶ Visualization
- ▶ Software Testing
- ▶ Algorithms

Electrical Engineering:

- ▶ Signal Processing
- ▶ Telecommunications
- ▶ Information Theory
- ▶ Control Theory
- ▶ Instrumentation, Sensors
- ▶ Hardware/Electronics Testing

Applications of Probability and Statistics

General:

- ▶ Gambling

Applications of Probability and Statistics

General:

- ▶ Gambling (not recommended)

Applications of Probability and Statistics

General:

- ▶ Gambling (not recommended)
- ▶ Stock Market Analysis

Applications of Probability and Statistics

General:

- ▶ Gambling (not recommended)
- ▶ Stock Market Analysis
- ▶ Politics

Applications of Probability and Statistics

General:

- ▶ Gambling (not recommended)
- ▶ Stock Market Analysis
- ▶ Politics
- ▶ Sports

Applications of Probability and Statistics

General:

- ▶ Gambling (not recommended)
- ▶ Stock Market Analysis
- ▶ Politics
- ▶ Sports
- ▶ Demographics

Applications of Probability and Statistics

General:

- ▶ Gambling (not recommended)
- ▶ Stock Market Analysis
- ▶ Politics
- ▶ Sports
- ▶ Demographics
- ▶ Medicine

Applications of Probability and Statistics

General:

- ▶ Gambling (not recommended)
- ▶ Stock Market Analysis
- ▶ Politics
- ▶ Sports
- ▶ Demographics
- ▶ Medicine
- ▶ Economics

Applications of Probability and Statistics

General:

- ▶ Gambling (not recommended)
- ▶ Stock Market Analysis
- ▶ Politics
- ▶ Sports
- ▶ Demographics
- ▶ Medicine
- ▶ Economics
- ▶ All Sciences!!

Alan Turing: Connecting CS and Probability

- ▶ “Father of Computer Science”



Alan Turing: Connecting CS and Probability

- ▶ “Father of Computer Science”
- ▶ Most famous for:
 - ▶ Computability, Turing machine
 - ▶ Stored-program computer
 - ▶ Turing test
 - ▶ WWII cryptanalysis



Alan Turing: Connecting CS and Probability

- ▶ “Father of Computer Science”
- ▶ Most famous for:
 - ▶ Computability, Turing machine
 - ▶ Stored-program computer
 - ▶ Turing test
 - ▶ WWII cryptanalysis
- ▶ Wrote a dissertation on probability theory!



Alan Turing: Connecting CS and Probability

- ▶ “Father of Computer Science”
- ▶ Most famous for:
 - ▶ Computability, Turing machine
 - ▶ Stored-program computer
 - ▶ Turing test
 - ▶ WWII cryptanalysis
- ▶ Wrote a dissertation on probability theory!
- ▶ Turing used probability and statistics to crack Enigma



Application: Machine Learning

Machine Learning builds statistical models of data in order to recognize complex patterns and to make decisions based on these observations.

Application: Machine Learning

Machine Learning builds statistical models of data in order to recognize complex patterns and to make decisions based on these observations.

Examples:

- ▶ Classification (recognition of faces or handwriting)

Application: Machine Learning

Machine Learning builds statistical models of data in order to recognize complex patterns and to make decisions based on these observations.

Examples:

- ▶ Classification (recognition of faces or handwriting)
- ▶ Prediction (stock market, elections)

Application: Randomized Algorithms

Application: Randomized Algorithms

- ▶ Some algorithms benefit from using random steps rather than deterministic ones

Application: Randomized Algorithms

- ▶ Some algorithms benefit from using random steps rather than deterministic ones
- ▶ Example: primality testing
 - ▶ Testing for all possible divisors is slow for large numbers
 - ▶ Instead test a random selection of divisors
 - ▶ Can be confident of primality up to a certain degree

Application: Randomized Algorithms

- ▶ Some algorithms benefit from using random steps rather than deterministic ones
- ▶ Example: primality testing
 - ▶ Testing for all possible divisors is slow for large numbers
 - ▶ Instead test a random selection of divisors
 - ▶ Can be confident of primality up to a certain degree
- ▶ Example: stochastic optimization methods
 - ▶ Optimizations can get “stuck” in the wrong answer, depending on how they are initialized
 - ▶ Re-run the algorithm with several random initializations

Application: Computer Graphics

- ▶ Ray tracing models light photons bouncing around a scene
- ▶ Impossible to model *every* photon
- ▶ Monte Carlo ray tracing simulates a random selection of photons

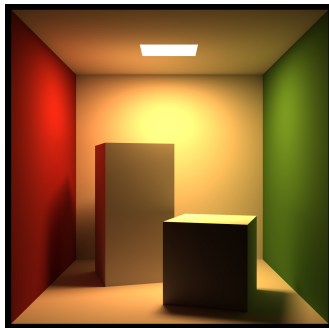
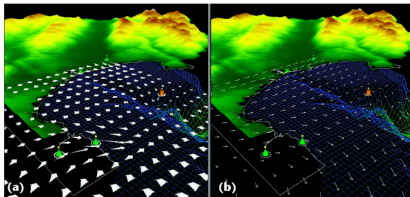


Image by Steve Parker (U of U)

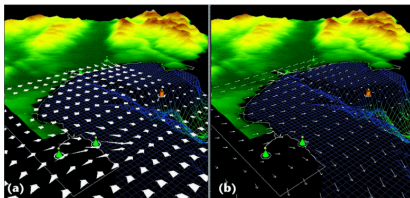
Application: Visualization

- ▶ Scientific data contains uncertainty



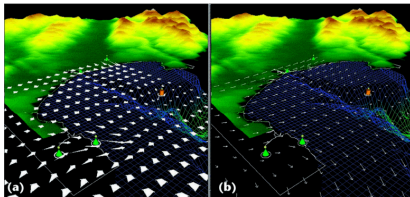
Application: Visualization

- ▶ Scientific data contains uncertainty
- ▶ Visualizations can be misleading as to “truth”



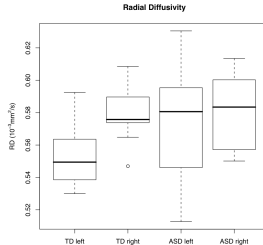
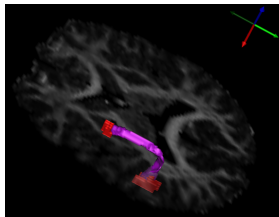
Application: Visualization

- ▶ Scientific data contains uncertainty
- ▶ Visualizations can be misleading as to “truth”
- ▶ Current research focuses on how to visualize uncertainty



Application: Medical Image Analysis

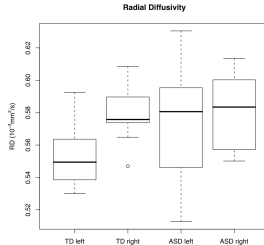
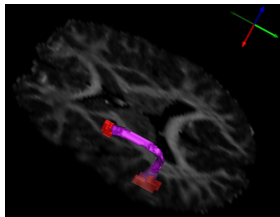
- ▶ Must deal with noisy image data



Fletcher et al, NeuroImage, 2010

Application: Medical Image Analysis

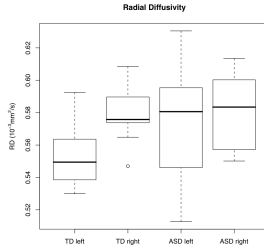
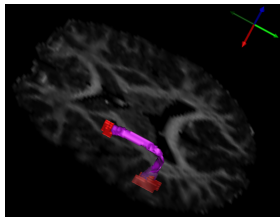
- ▶ Must deal with noisy image data
- ▶ Example: finding an anatomical structure in a 3D image



Fletcher et al, NeuroImage, 2010

Application: Medical Image Analysis

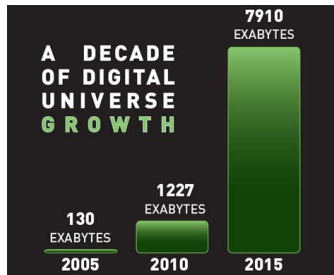
- ▶ Must deal with noisy image data
- ▶ Example: finding an anatomical structure in a 3D image
- ▶ Often includes statistical analysis of resulting data



Fletcher et al, NeuroImage, 2010

“Big Data” and “Analytics”

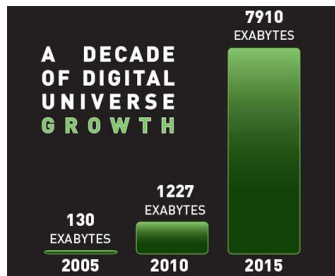
- ▶ The amount of digital data is exploding!



Source: IDC/EMC Digital Universe Study

“Big Data” and “Analytics”

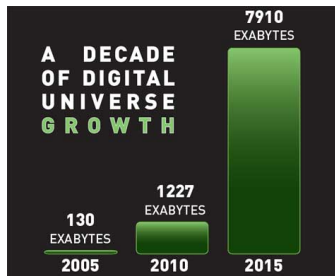
- ▶ The amount of digital data is exploding!
- ▶ Big data analysis is statistics + scalable CS.



Source: IDC/EMC Digital Universe Study

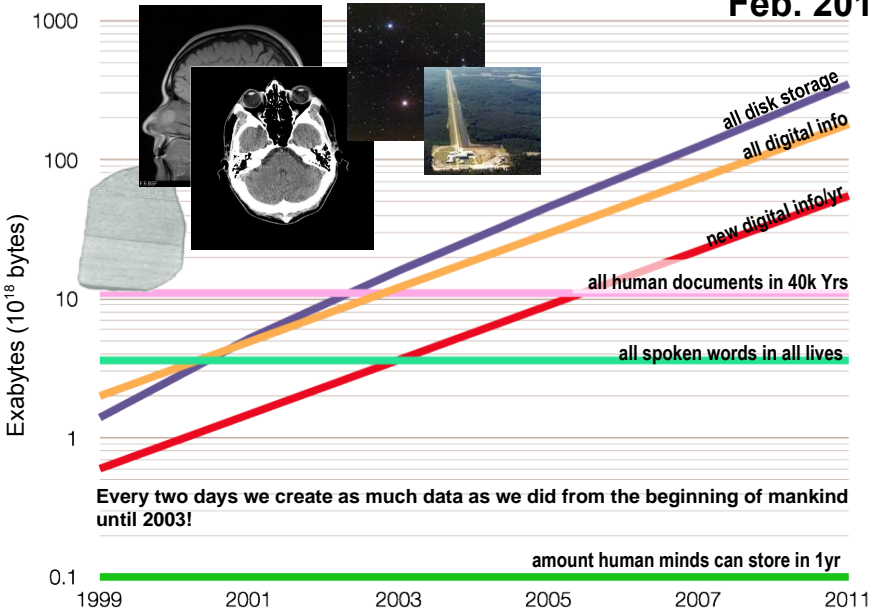
“Big Data” and “Analytics”

- ▶ The amount of digital data is exploding!
- ▶ Big data analysis is statistics + scalable CS.
- ▶ Examples: social media, internet purchases, news articles, scientific data, medical data



Source: IDC/EMC Digital Universe Study

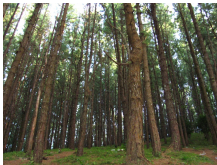
Feb. 2011



Every two days we create as much data as we did from the beginning of mankind until 2003!

Sources: Lesk, Berkeley SIMS, Landauer, EMC, TechCrunch, Smart Planet (slide by Chris Johnson)

How Much is an Exabyte?



How many trees does it take to print out an Exabyte?

1 Exabyte = 1000 Petabytes = could hold approximately
500,000,000,000,000 pages of standard printed text

It takes one tree to produce **94,200** pages of a book

Thus it will take **530,785,562,327** trees to store an Exabyte of data

In 2005, there were **400,246,300,201** trees on Earth

We can store **.75** Exabytes of data using all the trees on the entire planet.

Sources: <http://www.whatsabyte.com/> and <http://wiki.answers.com>
(slide by Chris Johnson)



car insurance



Web Maps News Images Shopping More Search tools

About 203,000,000 results (0.46 seconds)

Allstate® Car Insurance - allstate.com

Ad www.allstate.com/car-insurance/ (888) 458-6954

Drivers Who Switched Saved \$498/yr. Get a Car Insurance Quote Today!

Ratings: Ease of purchase 9.5/10 - Policy selection 8.5/10 - Service 7.5/10

Allstate Insurance has 21,903 followers on Google+

Progressive Car Insurance - Progressive.com

Ad www.progressive.com/

3.9 ★★★★★ rating for progressive.com

Named #1 Car Insurance Website. Get A Free Online Quote Now.

Ratings: Ease of purchase 9.5/10 - Policy selection 9/10

\$29 Cheap Car Insurance - Cheapest Utah Car Insurance

Ad utah.usautoinsurance.com/

Lowest Rates From \$28.99 / Month!

High Value Insurance - Minimum or Full Coverage - All Driving Records

Auto Insurance – State Farm® - State Farm Insurance

<https://www.statefarm.com/insurance/auto> State Farm Insurance

Choosing the right auto insurance can be tough. State Farm® makes it easier with a fast online quote that helps you get car insurance discounts as you go.

Coverage Options - Discounts - Safe Driver Program - Resources

Auto Insurance - Progressive

www.progressive.com/auto/ Progressive Corporation

Get the coverage you need, for the price you want. Start a car insurance quote today with the details you know offhand.

Comparison Rates - Coverages - Cheap Car Insurance - Why Progressive

Auto Insurance & Car Insurance Quotes—Allstate

www.allstate.com/auto-insurance.aspx Allstate

Customized car insurance coverage options, discounts galore and amazing features. Get a free online quote and speak to a dedicated Allstate agent.

①



Map for car insurance

Ad ①

GEICO Car Insurance

www.geico.com/

4.0 ★★★★★ rating for geico.com

You could save over \$500.

How much could you save?

Amica® Auto Insurance

www.amicacoverage.com/Car+Insurance

Official Car Insurance Quotes From

Amica®. Available 24/7-Quote Today!

Nationwide® Car Insurance

www.nationwide.com/JoinTheNation

Get a Free Online Quote Today!

Join the Nation and Drive with Us.

21st Century Insurance

www.21st.com/

Free Car Insurance Price Quote.

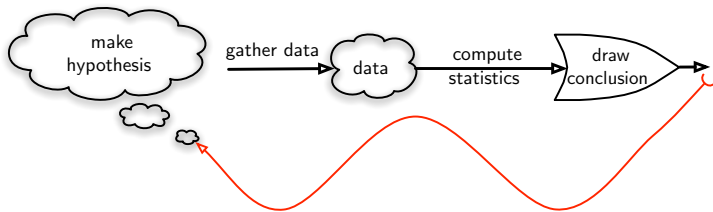
See If You Could Save Hundreds.

Esurance Car Insurance

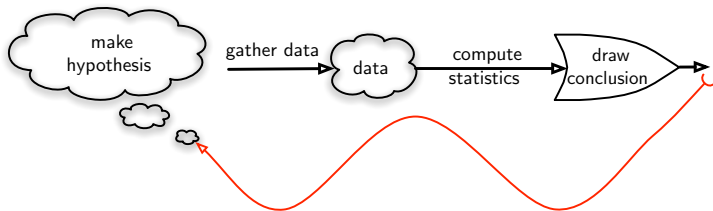
www.esurance.com/Utah

Same Big Discounts & Great Service.

The Scientific Method

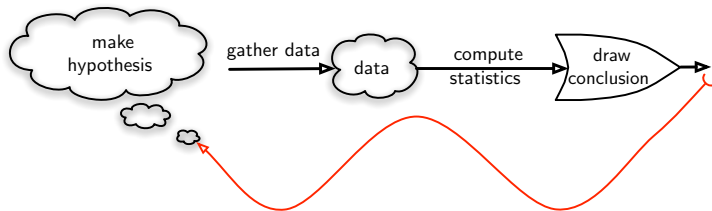


The Scientific Method



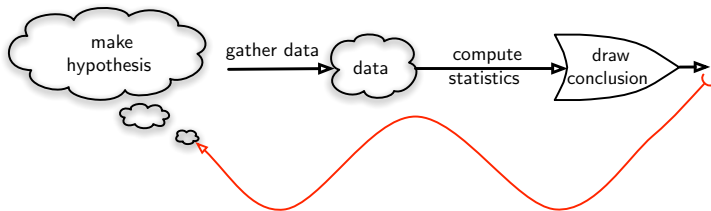
1. Define the question

The Scientific Method



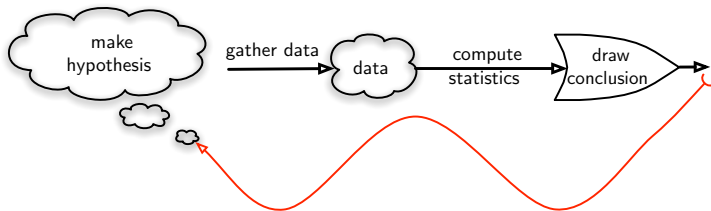
1. Define the question
2. Background research, observation

The Scientific Method



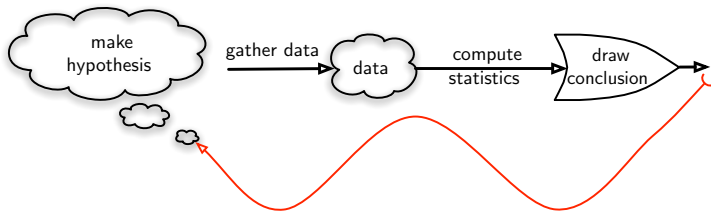
1. Define the question
2. Background research, observation
3. Formulate a hypothesis

The Scientific Method



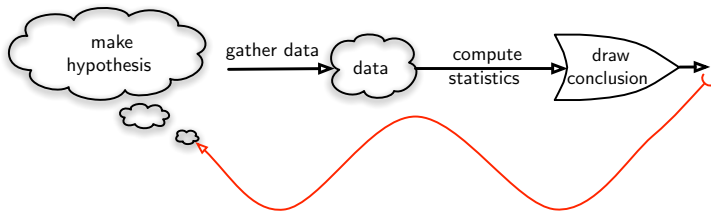
1. Define the question
2. Background research, observation
3. Formulate a hypothesis
4. Design and run an experiment

The Scientific Method



1. Define the question
2. Background research, observation
3. Formulate a hypothesis
4. Design and run an experiment
5. Analyze the results

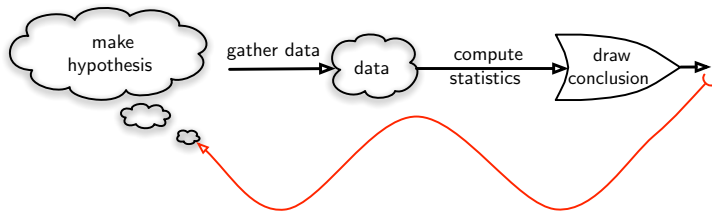
The Scientific Method



1. Define the question
2. Background research, observation
3. Formulate a hypothesis
4. Design and run an experiment
5. Analyze the results

Experimental measurements are noisy (randomness).

The Scientific Method

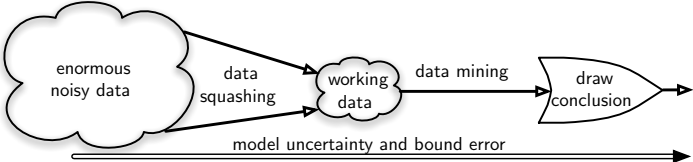


1. Define the question
2. Background research, observation
3. Formulate a hypothesis
4. Design and run an experiment
5. Analyze the results

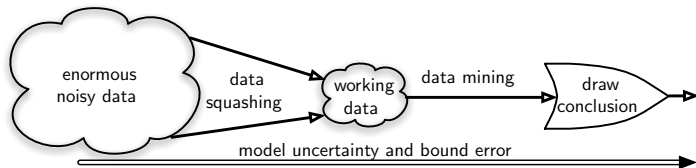
Experimental measurements are noisy (randomness).

Statistics is critical in the last *two* steps!

Data Science

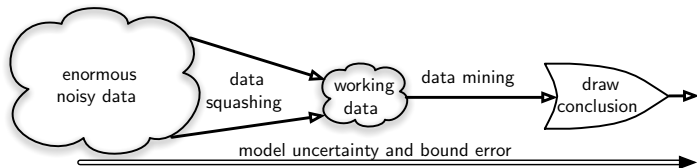


Data Science



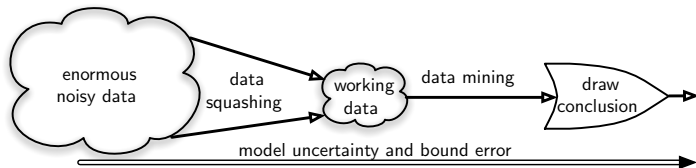
1. Process/Squash enormous available data

Data Science



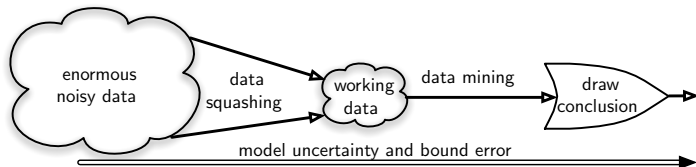
1. Process/Squash enormous available data
2. Mine working data (calculate many statistics)

Data Science



1. Process/Squash enormous available data
2. Mine working data (calculate many statistics)
3. Analyze the results / Draw conclusions

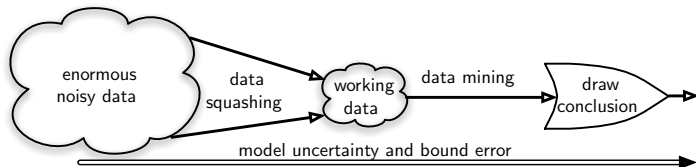
Data Science



1. Process/Squash enormous available data
2. Mine working data (calculate many statistics)
3. Analyze the results / Draw conclusions

Every step is subject to noise and involves statistics.

Data Science



1. Process/Squash enormous available data
2. Mine working data (calculate many statistics)
3. Analyze the results / Draw conclusions

Every step is subject to noise and involves statistics.

What statistics can and *cannot* do!

What You Should Do Now

1. Check out the class web page: `www.cs.utah.edu/~jeffp/teaching/cs3130.html`
2. Download the book
(start reading Ch 1 & 2)
3. Download and install R on your machine
(take a look at R tutorial)