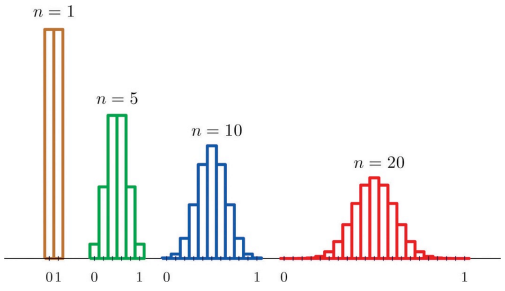
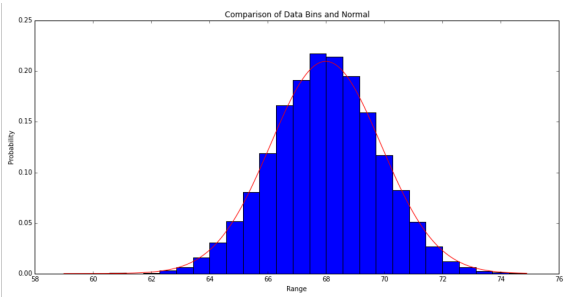
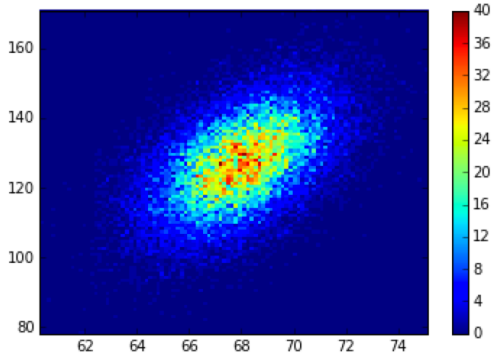
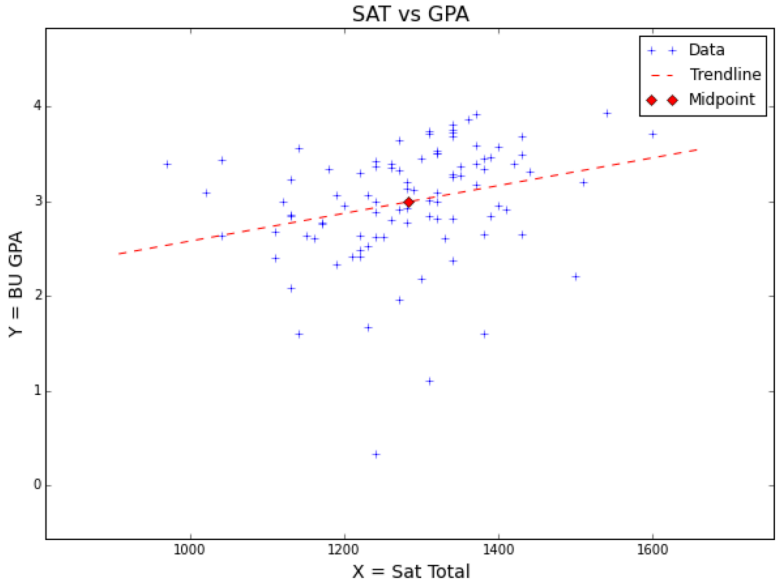
CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University



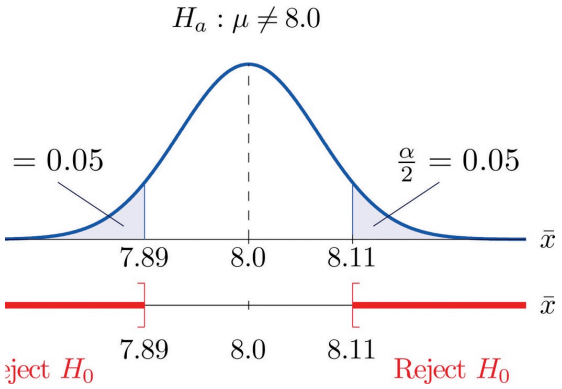







```
Var(x): 12702.5596      std(x): 112.7056
Var(y): 0.3757  std(y): 0.6129

rho(x,y):  0.269
```



Randomness: Life is Uncertain!

Informal Definition of Randomness:

"The apparent lack of pattern or predictability in events"

Information Theory Definition of Randomness:

"The maximum entropy" OR "the minimum of information"

Example: Flip a coin – before you look at it: Is it heads or tails?

Our experience is filled with random events! Your chance in 2021 of

- being audited by IRS: 1 in 175
- finding a pearl in an oyster: 1 in 12,000
- winning $1,000,000 in Powerball: 1 in 11.6 million
- being killed by
  - a champagne cork:
  - a shark
  - a vending machine
  - a cow
  - hot tap water

Which are most and least likely?

Randomness: Life is Uncertain!

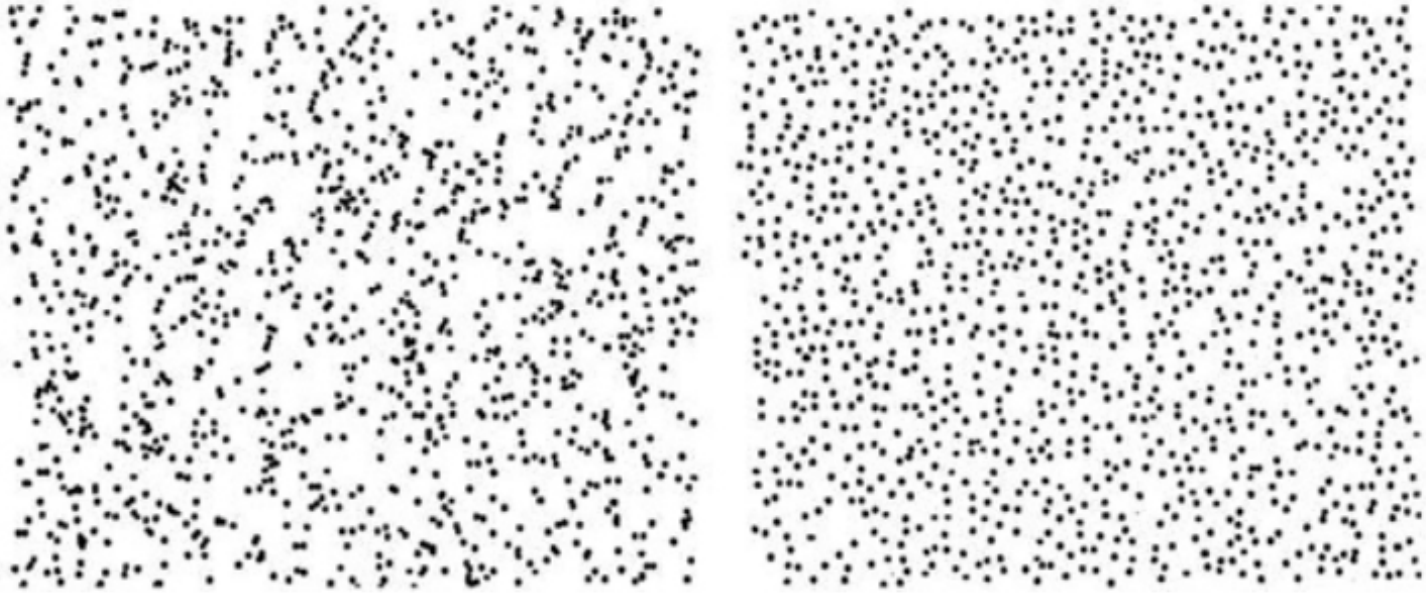
Our experience is filled with random events! Your chance in 2021 of

- being audited by IRS: 1 in 175
- finding a pearl in an oyster: 1 in 12,000
- winning $1,000,000 in Powerball: 1 in 11.6 million
- being killed by
  - hot tap water: 1 in 3 million
  - a champagne cork: 1 in 13 million
  - a cow: 1 in 16 million
  - a vending machine: 1 in 146 million
  - a shark: 1 in 319 million

Randomness: "It's Complicated"!

But what IS randomness?

What about a "random" arrangement of dots in 2D. One of these is randomly generated by computer, and one is natural (non-random); which is which?



Randomness: "It's Complicated"!

One more: Suppose I ask all of you to randomly flip a coin 100 times, and to write down the sequence of 100 H's and T's. But one slacker decides to save time by simply writing down the sequence "at random" but without actually flipping the coins. One of these sequences is truly random, and one is not. Which is which?

HTTHTTHTHHTTHTHTHTTHHTHTT	*THHHTHTTTTHTTHTTTHHTHTTHT*
HTTHHHTTHTTHTHTHTHHTTHTTH	*HHHTHTHHTHTTHHTTTTHTTTHTH*
THTHTHTHHHTTHTHTHTHHTHTTT	*TTHHTTTTTTTTHTHHHHHTHTHTH*
HTHHTHTHTHTHHTTHTHTHTTHHT	*THTHTHHHHHTHHTTTTTHTTHHTH*

Randomness and Non-Randomness

But of course the universe is not completely uncertain, and in the last 350 years or so we have developed mathematical tools for understanding the difference:

> **"Probability Theory is the mathematical study of random phenomena."**
> (Encylopedia Britannica)

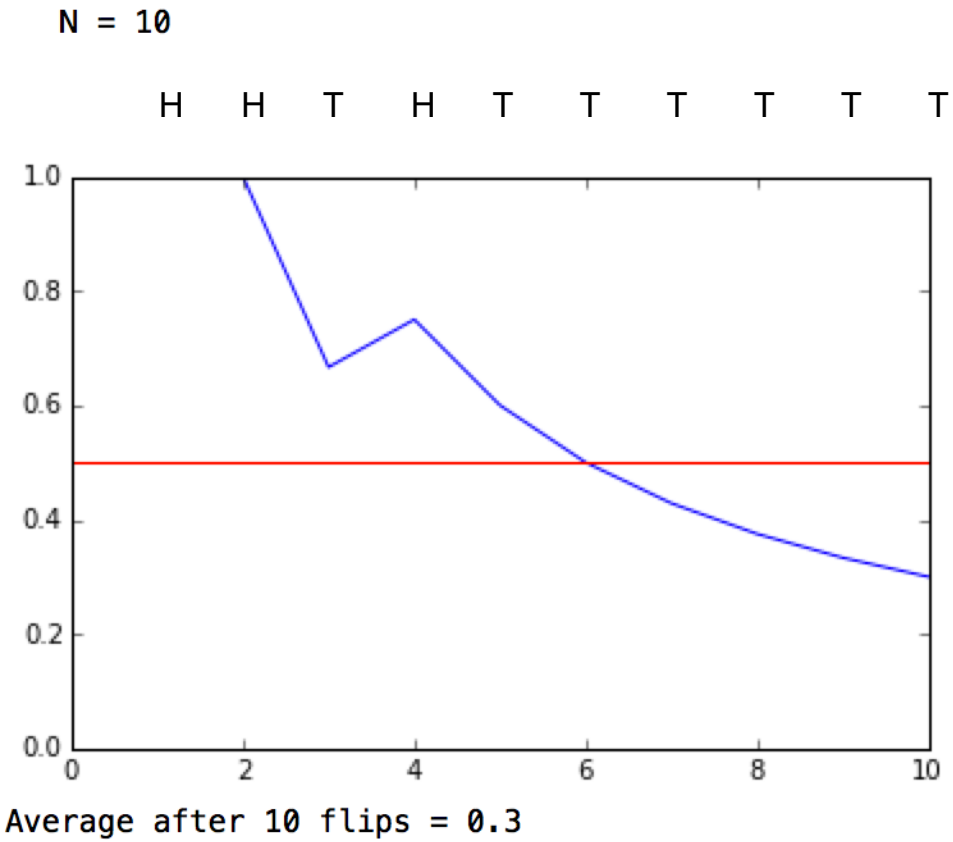
> **"Statistics is the science of learning from data, and of measuring, controlling, and communicating uncertainty…."**
> (American Statistical Association)

Much of this work has to do with **discovering structure** within a group or sequence of random events; many random phenomena behave in ways that are unpredictable in the short term, but have non-random characteristics when viewed as a whole – we seek to understand the difference!

Randomness and Non-Randomness

Examples of patterns within random events:

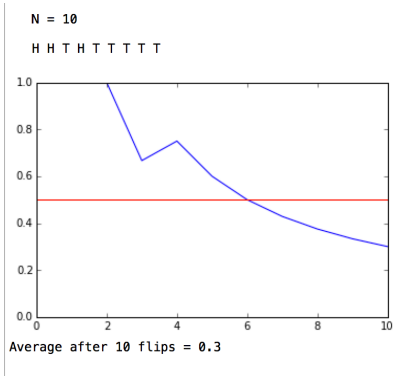
- **Example 1:** Flip a coin over and over; what is the average number of heads?

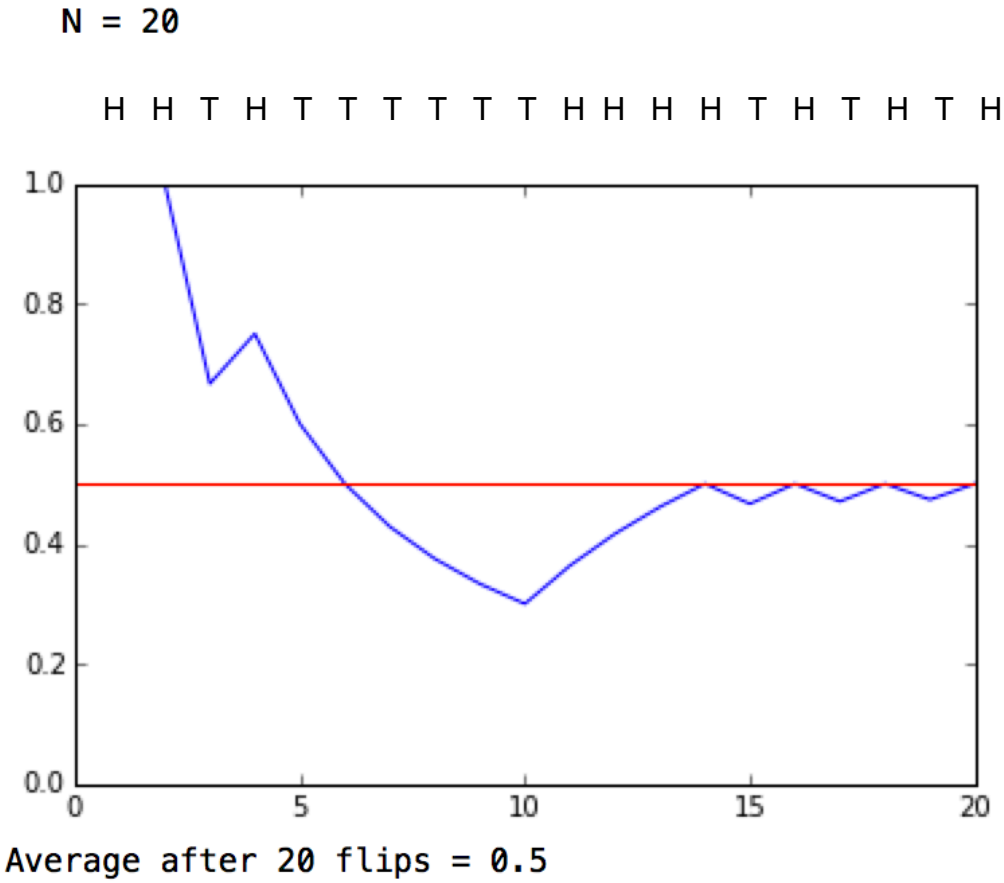


Randomness and Non-Randomness

Examples of patterns within random events:

- **Example 1:** Flip a coin over and over; what is the average number of heads?

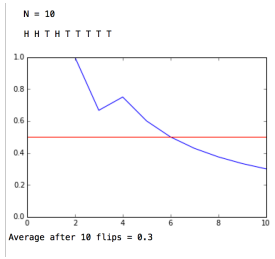


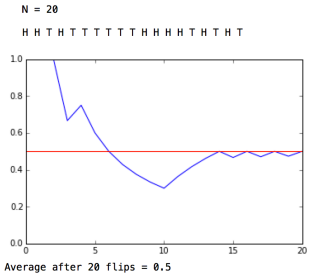


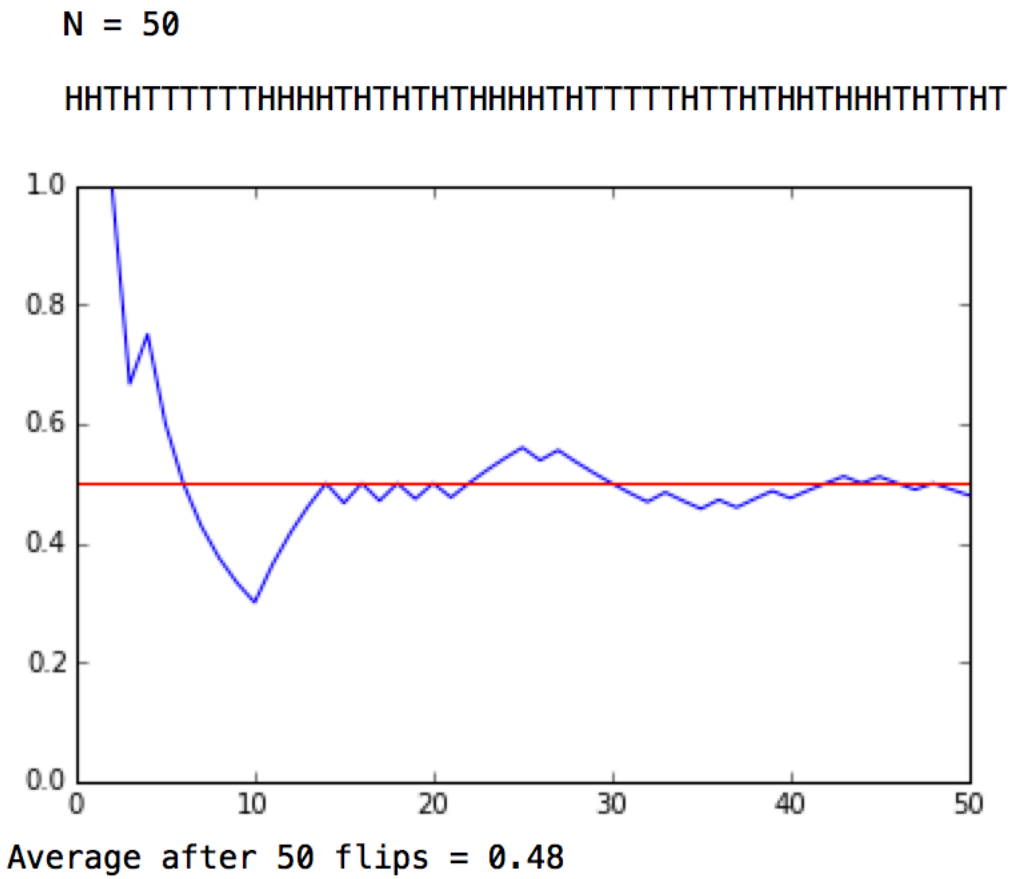
Randomness and Non-Randomness

Examples of patterns within random events:

- **Example 1:** Flip a coin over and over; what is the average number of heads?



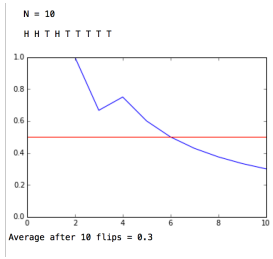


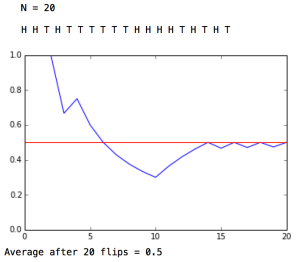


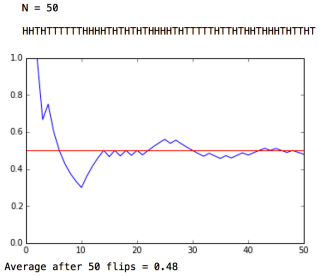
Randomness and Non-Randomness

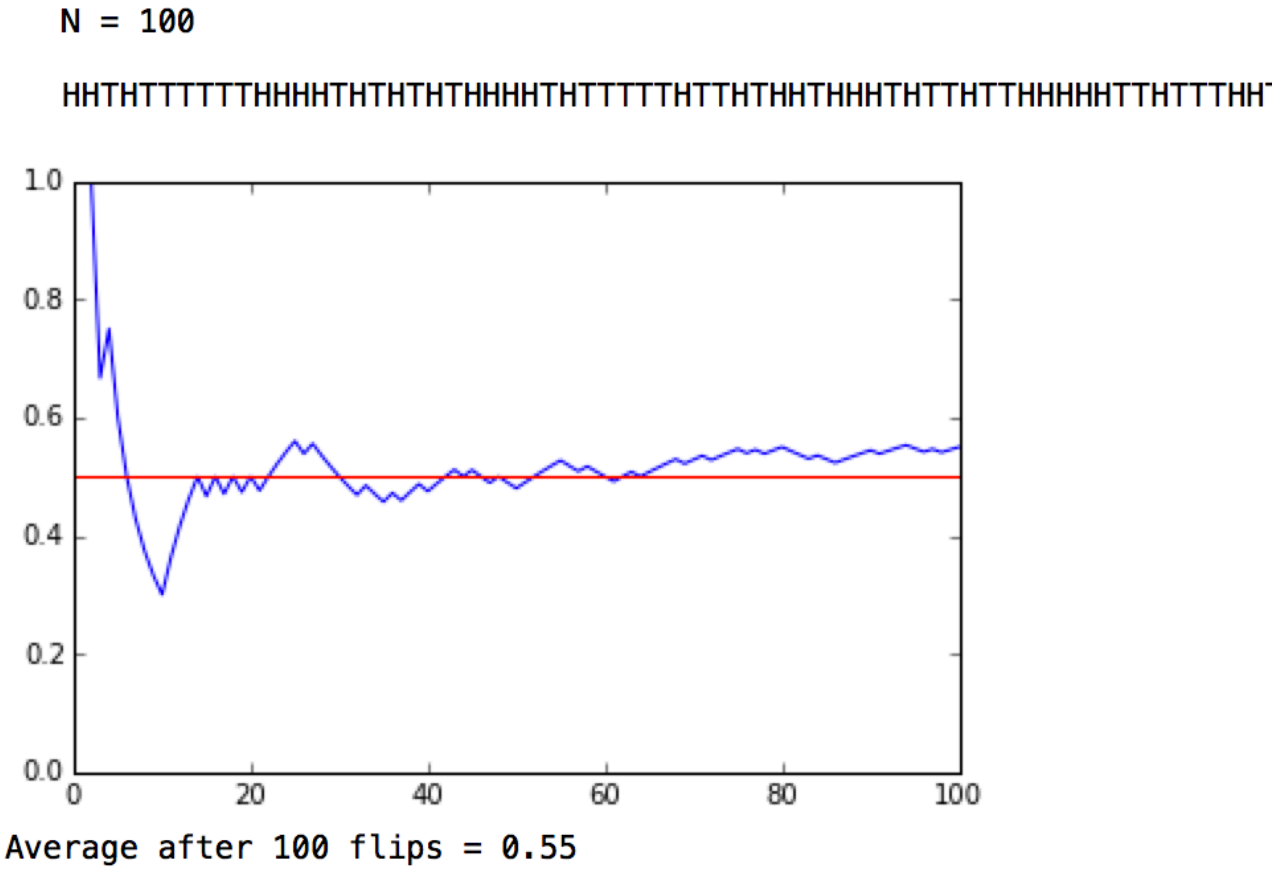
Examples of patterns within random events:

- **Example 1:** Flip a coin over and over; what is the average number of heads?





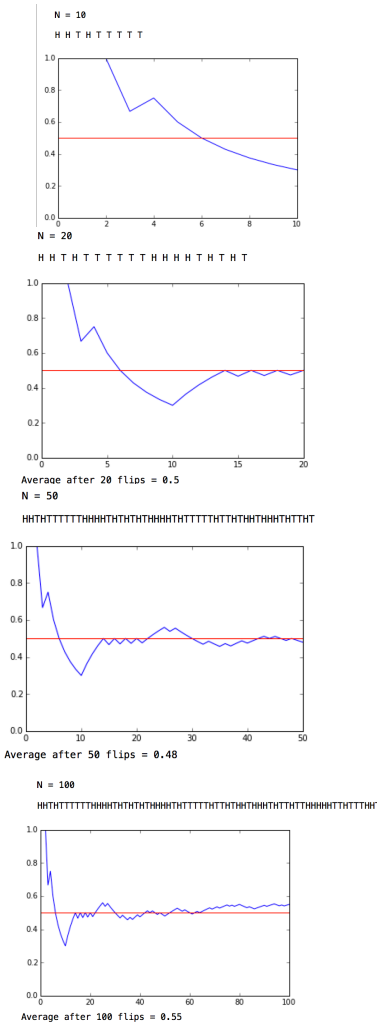


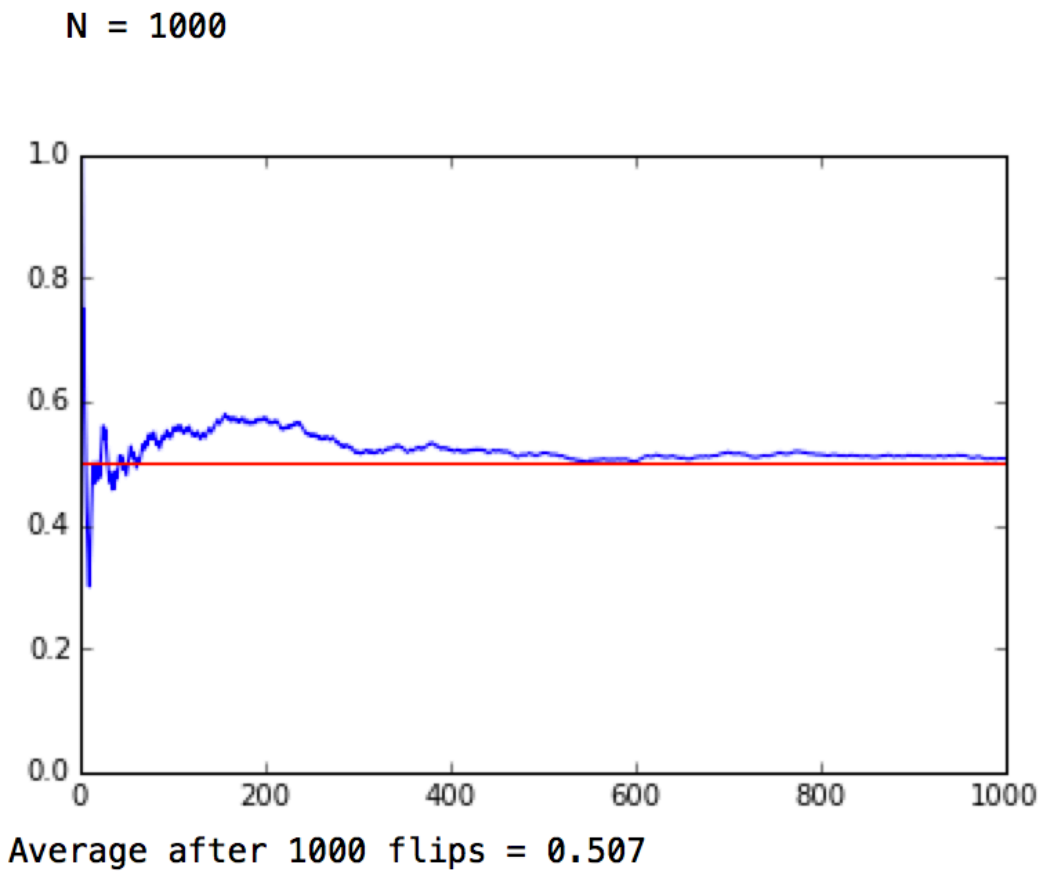


Randomness and Non-Randomness

Examples of patterns within random events:

- **Example 1:** Flip a coin over and over; what is the average number of heads?

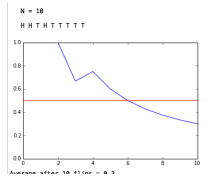


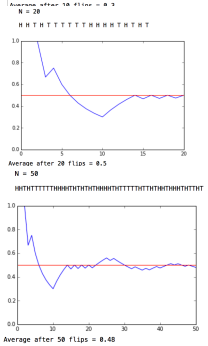


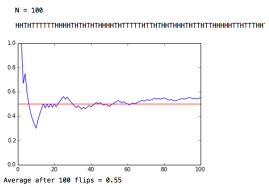
Randomness and Non-Randomness

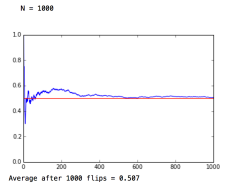
Examples of patterns within random events:

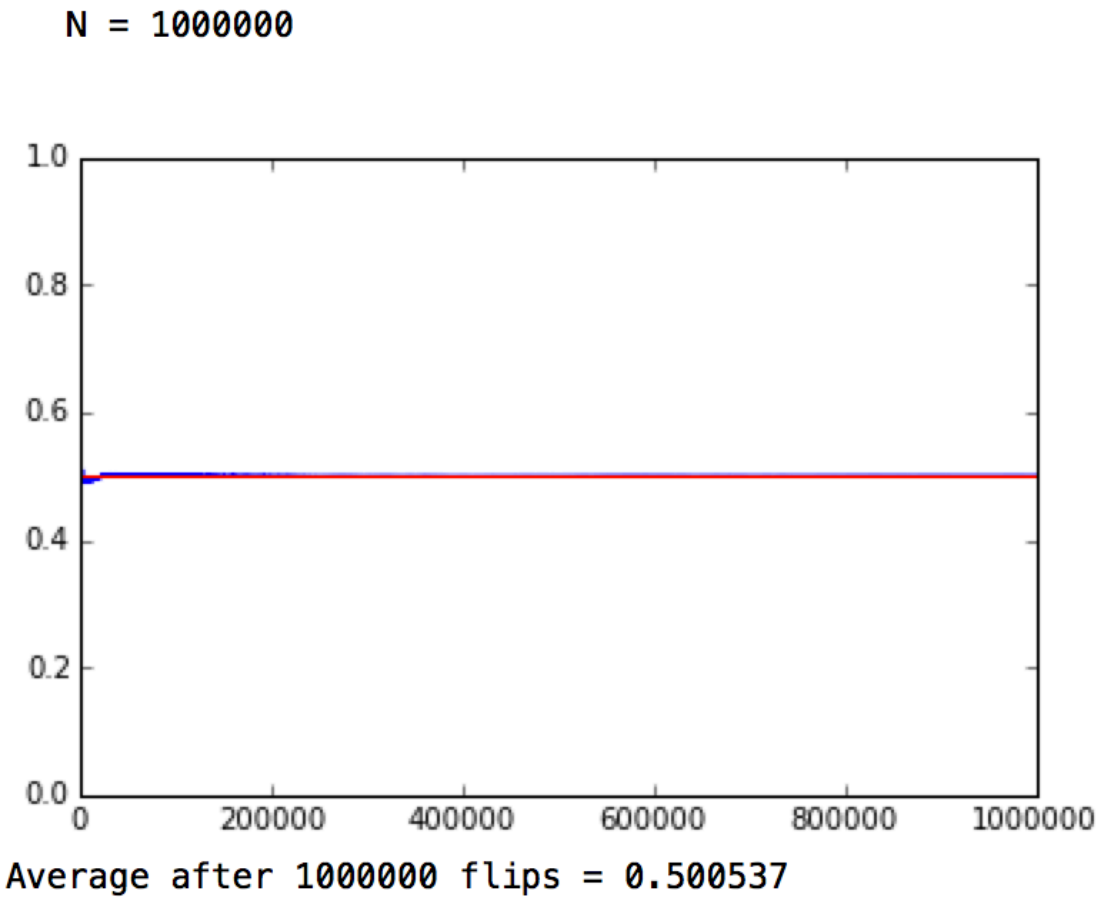
- **Example 1:** Flip a coin over and over; what is the average number of heads?







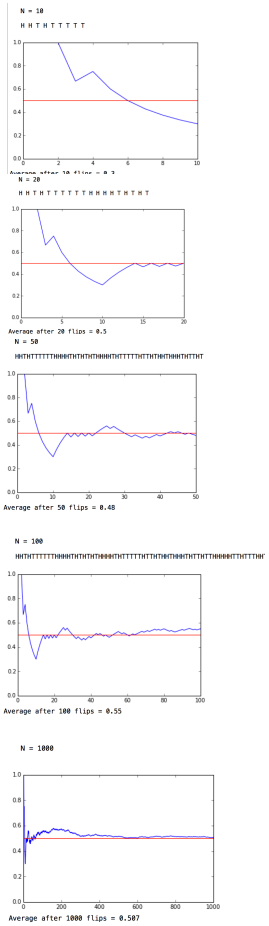


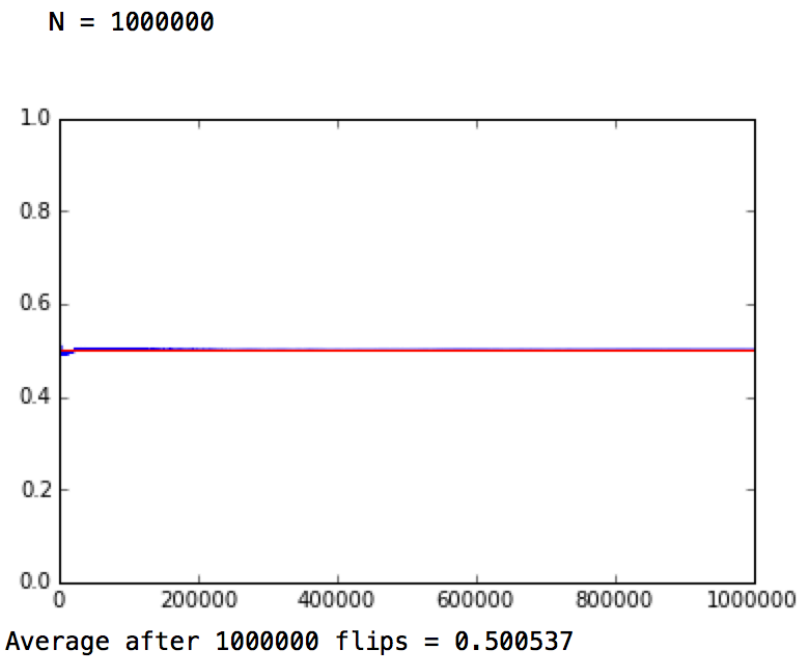


Randomness and Non-Randomness

Examples of patterns within random events:

- **Example 1:** Flip a coin over and over; what is the average number of heads?





The average number of heads ALWAYS approaches 0.5 as N gets larger!

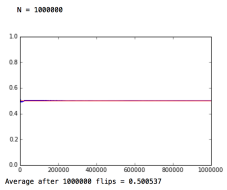
One coin flip is random, but many coin flips are non-random….

Patterns emerge when we repeat random experiments over and over….

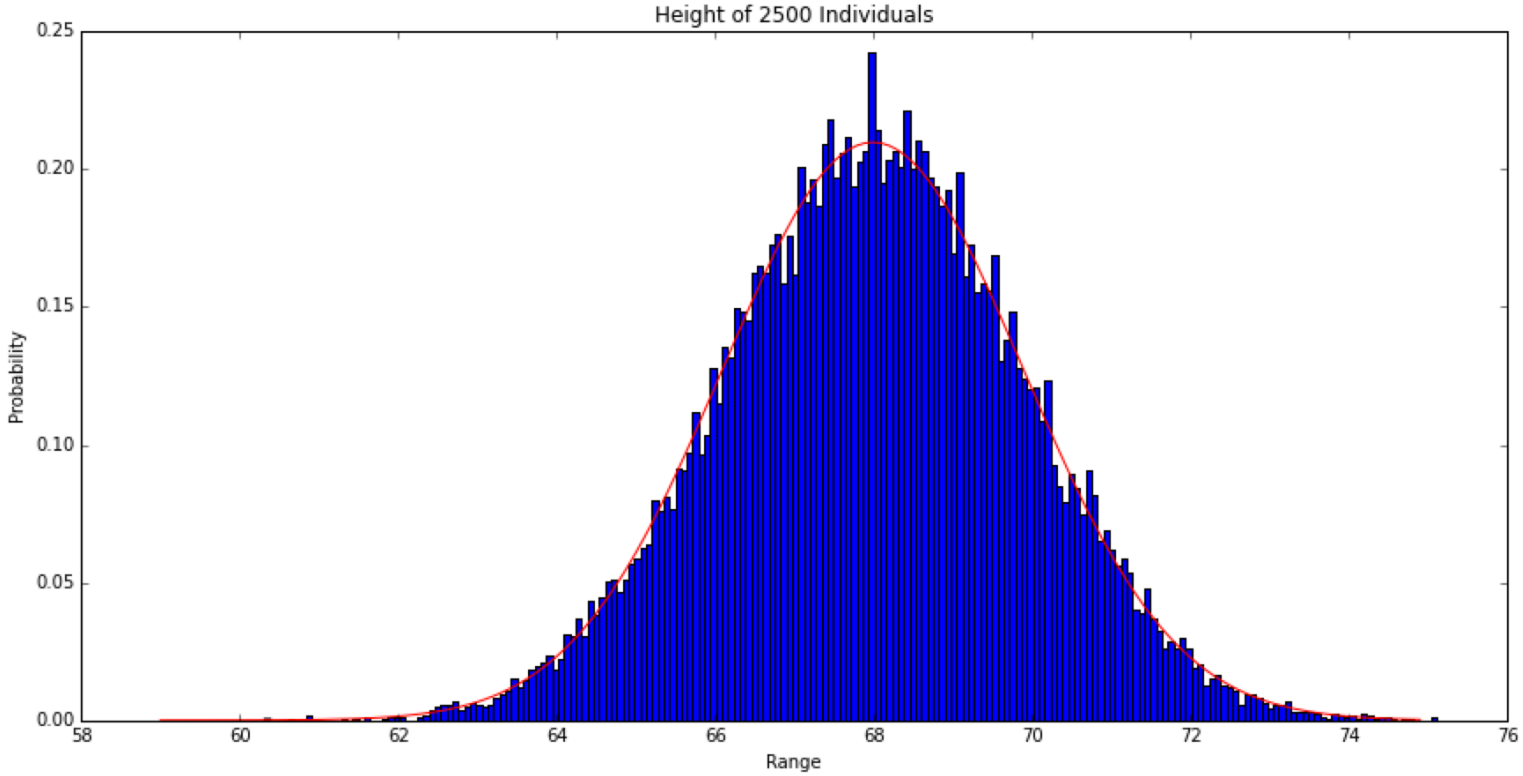
Randomness and Non-Randomness

Examples of patterns emerging when we "zoom out" and look at large numbers of seemingly random events:

- **Example 1:** Flip a coin over and over; what is the average number of heads?



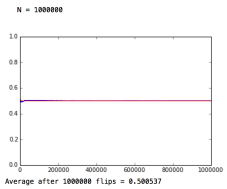
- **Example 2:** What is the height of a human being?



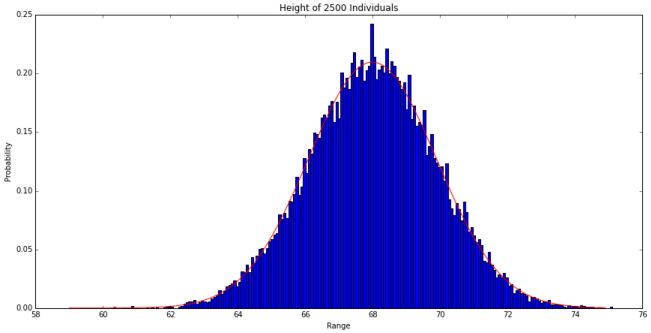
Randomness and Non-Randomness

Examples of patterns emerging when we "zoom out" and look at large numbers of seemingly random events:

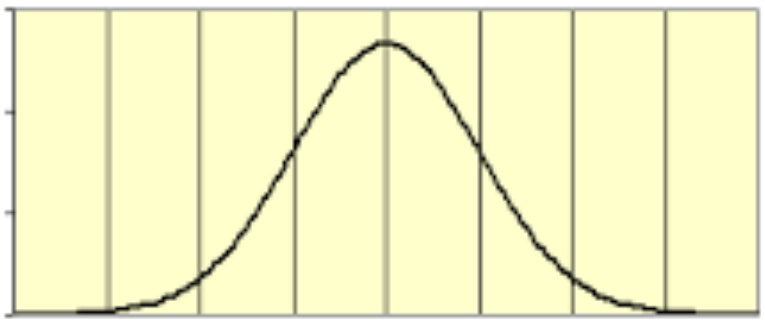
- **Example 1:** Flip a coin over and over; what is the average number of heads?



- **Example 2:** What is the height of a human being?



- **Example 3:** What is the IQ of a human being?

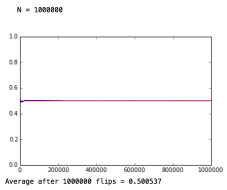


Standard Deviations	-4	-3	-2	-1	0	1	2	3	4
Wechsler IQ	40	55	70	85	100	115	130	145	160
Stanford-Binet IQ	36	52	68	84	100	116	132	148	164
Cumulative %	0.003	0.135	2.275	15.866	50.000	84.134	97.725	99.865	99.997

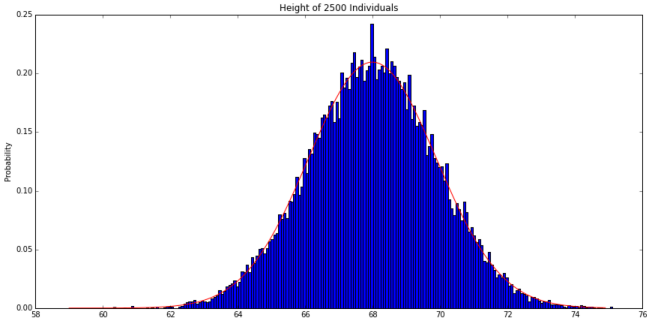
Randomness and Non-Randomness

Examples of patterns emerging when we "zoom out" and look at large numbers of seemingly random events:

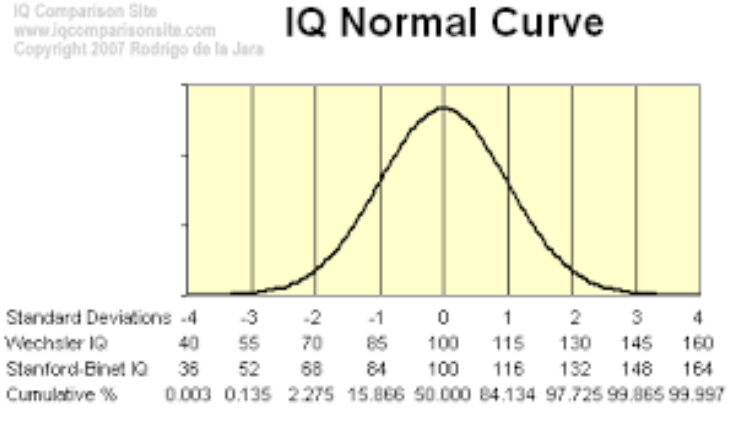
- **Example 1:** Flip a coin over and over; what is the average number of heads?



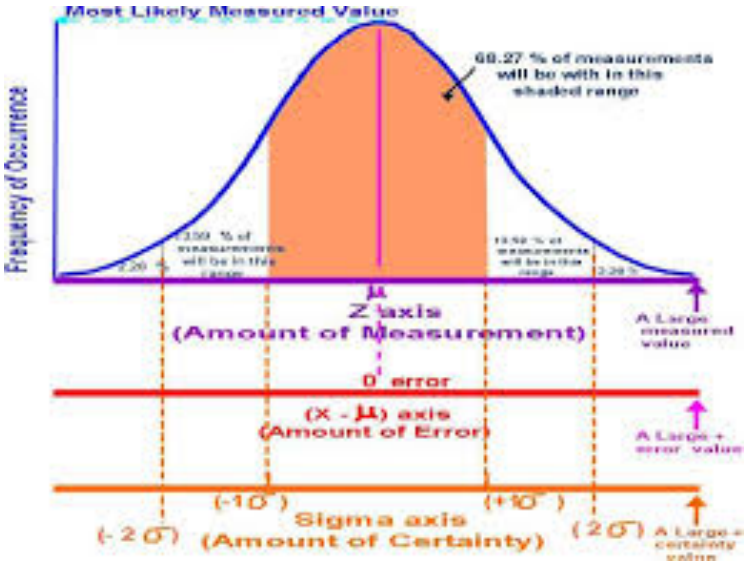
- **Example 2:** What is the height of a human being?



- **Example 3:** What is the IQ of a human being?



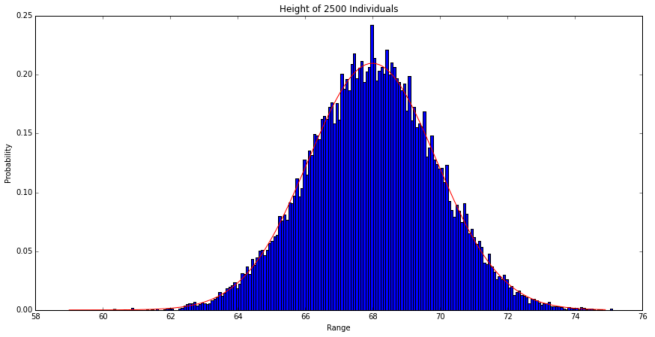
- **Example 4:** What is the distribution of measurement errors?

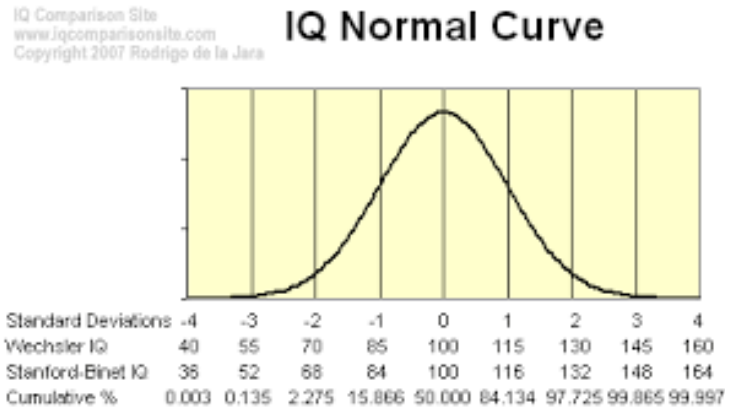


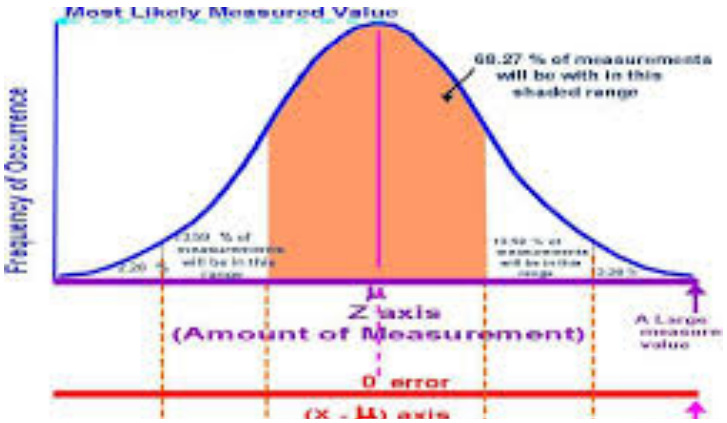
Randomness and Non-Randomness

Examples of patterns emerging when we "zoom out" and look at large numbers of seemingly random events:

- **Example 1:** Flip a coin over and over; what is the average nu
- **Example 2:** What is the height of a human being?
- **Example 3:** What is the IQ of a human being?
- **Example 4:** What is the distribution of measurement errors?
- **Example 5:** What is the expected return on a stock?







Descriptive Statistics

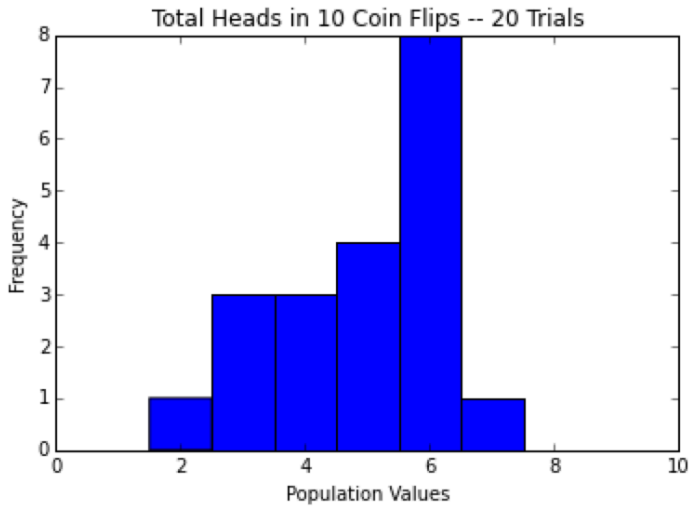
The **Distribution** of a data set is a graph of values vs. frequency in the population:

- **Example 1:** Toss a single die 20 times and record the number of dots each time:

```
Population: [4, 4, 4, 3, 2, 4, 6, 2, 1, 4, 1, 4, 4, 3, 6, 3, 6, 1, 2, 2]
```

- **Example 2:** Flip a coin 10 times and count the number of heads; repeat 20 times and record the number of heads each time:

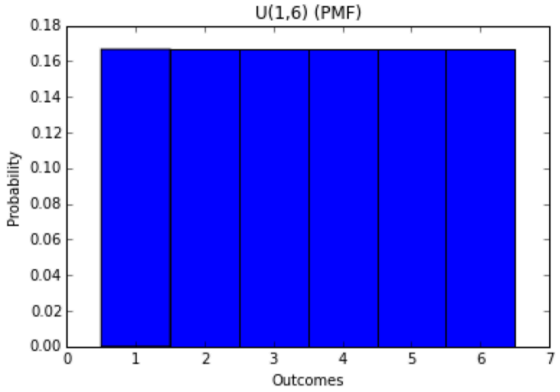
```
Population: [4, 6, 5, 6, 4, 6, 5, 4, 6, 3, 6, 5, 2, 3, 6, 6, 3, 7, 5, 6]
```



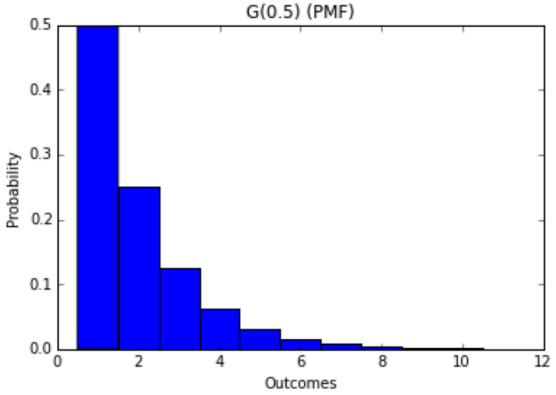
Descriptive Statistics

There is a large number of **Theoretical Distributions** to describe different kinds of populations; these abstractions aid in the analysis of real data sets:

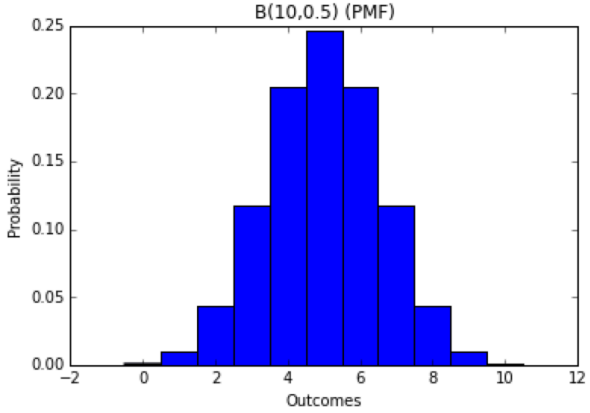
Uniform Distribution



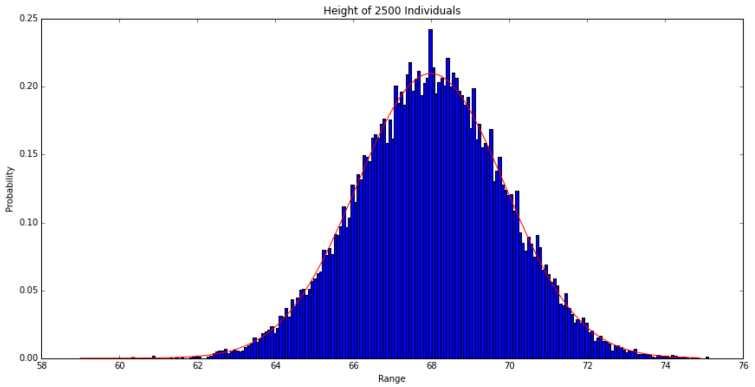
Geometric Distribution



Binomial Distribution



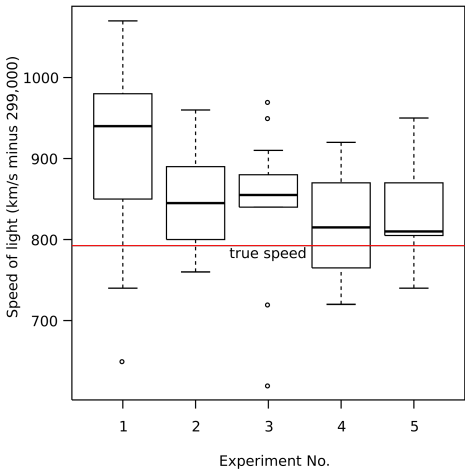
Normal Distribution

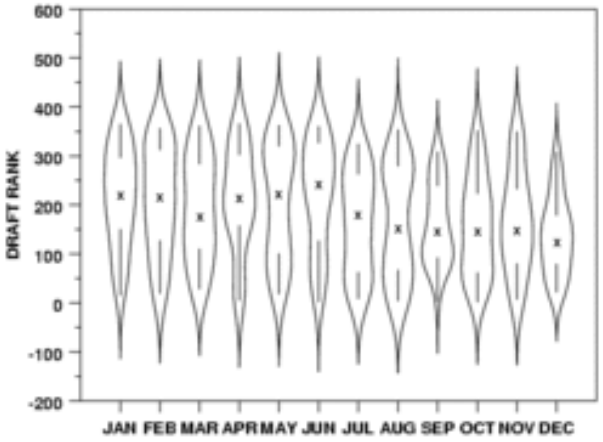


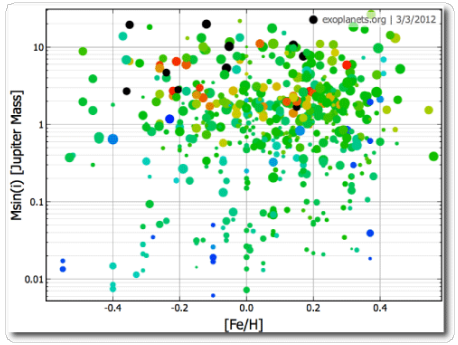
Descriptive Statistics

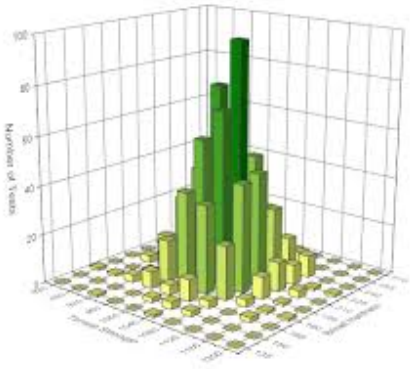
Data Graphics

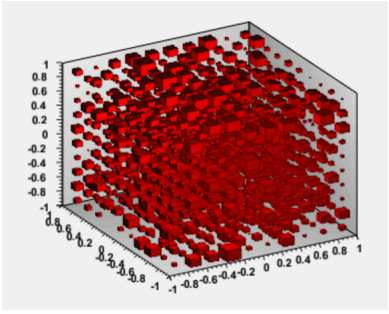
- Box plots, violin plots
- Histograms 2D and 3D
- Scatterplots 2D and 3D
- Special purpose graphics

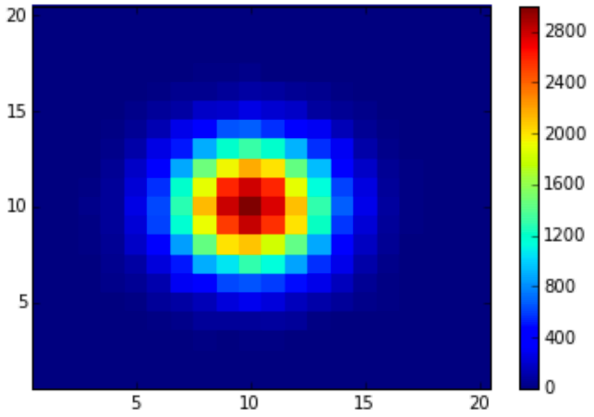


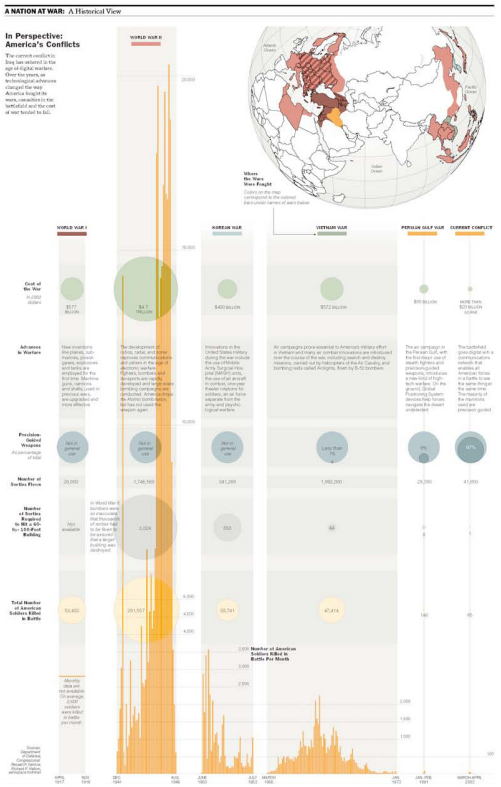












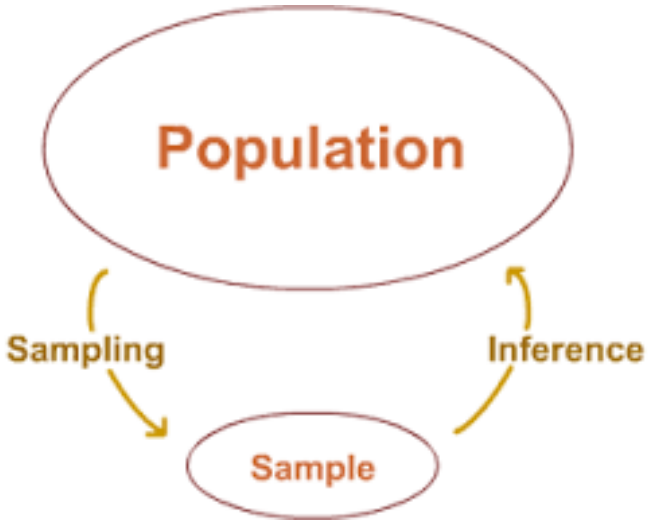
M. Ericson, I

Inferential Statistics

…. is the process of deducing properties of a population by examining samples; properties of interest include:

- the underlying theoretical distribution;
- summary statistics such as the mean;
- tests of hypotheses;
- correlation and regression (for multi-dimensional data).

We conduct statistical experiments involving a (relatively small) random sample from a larger population,which can not be examined as a whole; properties of the sample are used to infer properties of the population. Often the estimate comes with a "confidence interval" telling us how certain we are of our result.

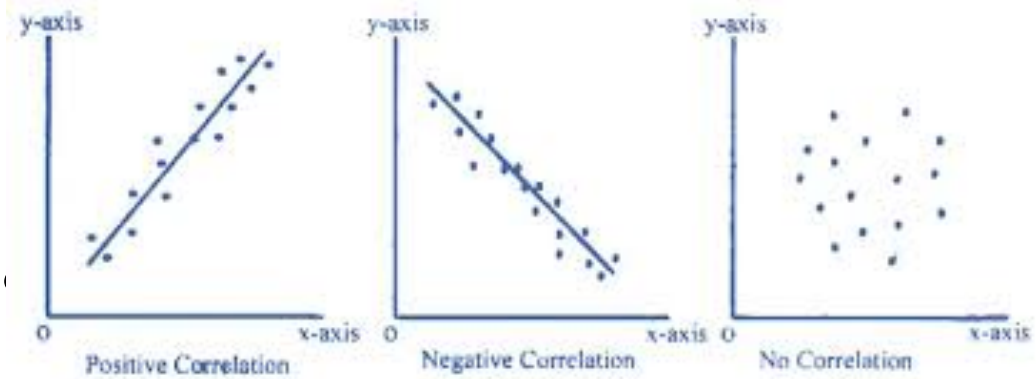


Think Polling!

Review of CS 237

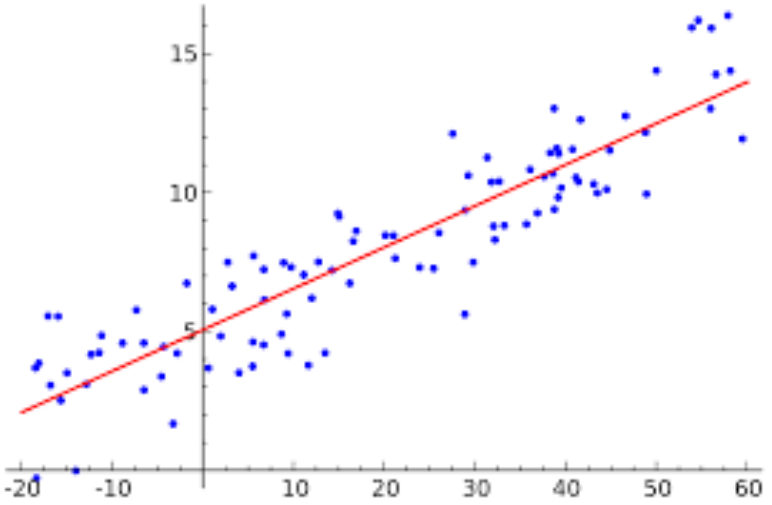
Descriptive Statistics:

- What does the data look like?
- How can we summarize it?
- How can we display it graphically?
  - Diagrams, bar charts, scatter plots,



Inferential Statistics:

- Sampling theory and statistical experiments
- Reasoning about sample statistics
- Standard Statistical Procedures:
  - Estimating means with Confidence Intervals
  - Hypothesis Testing
  - 2D Data: Correlation and Regression



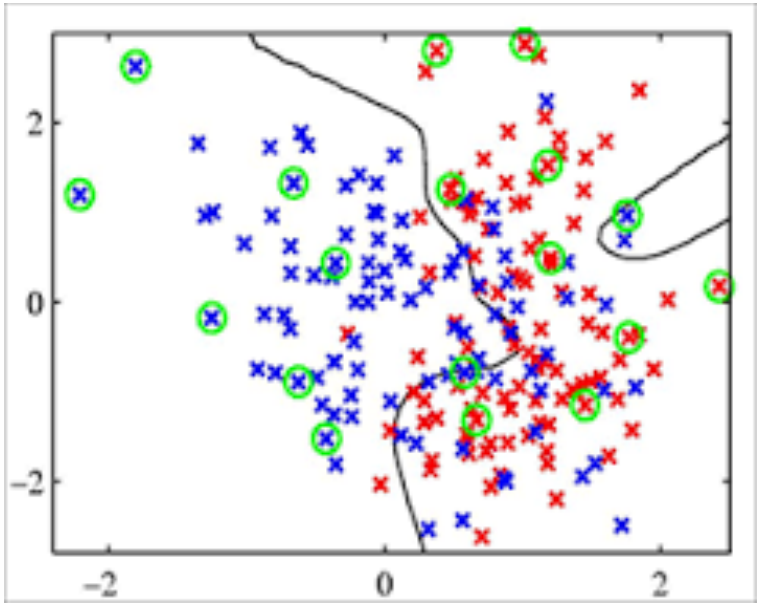
Review of CS 237

Descriptive Statistics:

- What does the data look like?
- How can we summarize it?
- How can we display it graphically?
  - Diagrams, bar charts, scatter plots, etc.

Inferential Statistics:

- Sampling theory and statistical experiments
- Reasoning about sample statistics
- Standard Statistical Procedures:
  - Estimating means with Confidence Intervals
  - Hypothesis Testing
  - 2D Data: Correlation and Regression
  - Pattern Recognition/Machine Learning



Overview of CS 237

Probabilistic Algorithms: Using random processes to compute

- **Monte Carlo Algorithms:** Result may be incorrect with a small random error (e.g., calculation of π with a dart board)
- **Las Vegas Algorithms:** Result will be correct but running time is random (e.g. testing a program for correctness by randomly generating inputs)
- **Probabilistic Data Structures:** Randomness in storing data (e.g., hash tables, randomized BSTs, Bloom Filters)
- **Discrete Event Simulation:** Analyze a computer system or network by generating random workflows and measuring the performance using statistical tools (e.g., simulating an MM1 queueing system)