PSTAT 5A

# Introduction to Statistics

Instructor: Alexandra Chronopoulou

MWF 12.00-12.50pm

01/07/2013

# General Info

| | |
|---|---|
| *Instructor* | **Alexandra Chronopoulou**<br>chronopoulou@pstat.ucsb.edu<br>South Hall 5501 |
| *Office Hours* | `MW` 1:30am – 2:30pm |
| *Course Website* | `https://gauchospace.ucsb.edu/` |
| *TA* | Go to your section to meet your TA |

Waiting List online for add codes.

# Lectures

- A lectures plan with the corresponding sections from the book is posted on the course website.
- After each lecture, the slides will also be posted online.

i>clicker

* Buy an i>Clicker from UCen bookstore.
* Register your clicker @ `iclicker.com`.
* For any problems refer to the instructions at the course website.

i>clicker

- Clickers will be used to answer questions related to the material that is being discussed in the lecture.
  - You get 1 point by answering all but one question during a lecture (right or wrong).
  - Otherwise, you get 0.
  - You can drop the three lowest scores (that is up to three 0s).

# Discussion Sessions

| *Classroom Session* | *Lab Session* |
|---|---|
| Written assignments | Online applets |
| 5% of final grade | 5% of final grade |

## General Comments

1. Finish the assignment during the session and return it to your TA.
2. Grades will be posted on Gauchospace. You have **one week** to make sure that grades have been posted correctly. If you want to challenge a grade, talk to your TA.
3. There are no make-up discussion sessions.

# Quizzes & Exams

- Quizzes
  1. **Wednesday, January 23*rd***
  2. **Wednesday, February 6*th***
  3. **Monday, February 18*th***
  4. **Monday, March 4*th***
- Final
  Tuesday, March 19*th* (12.00-3.00pm)

## General comments

- Closed books/notes. Only non-graphing calculators are allowed.
- It is your responsibility to ensure you are available at the designated times.
- There are no make-up exams, unless you are on a sports team.

# Grading Scheme

| | Grade A | Grade B |
|---|---|---|
| Classroom & Labs | 10% | 10% |
| Quizzes | 60% (15% each) | 55% (lowest will count only 10%) |
| Participation | 0% | 5% |
| Final Exam | 30% | 30% |
| Total | 100% | 100% |

**Final Grade = Maximum (Grade A, Grade B)**

# Where do I get help?

1. Ask me any questions during office hours.
2. Ask your TAs during the weekly sessions.
3. Work in groups!
4. CLAS $\longrightarrow$ `http://www.clas.ucsb.edu/` is great!
5. Approved tutors

   **`http://www.pstat.ucsb.edu/projects/leadtatraining/tutoring.htm`**

# Academic Integrity

- You should not go to another classroom or lab section.
- All assignments and exams should be your own. Bring your student ID at every exam.
- Do not use anybody else's clicker. If you are caught having two clickers, you and your friend will loose all participation points and you will be reported for cheating.

Cheating is a very serious offense and at a *minimum* will result in your receiving an **F** in this class.

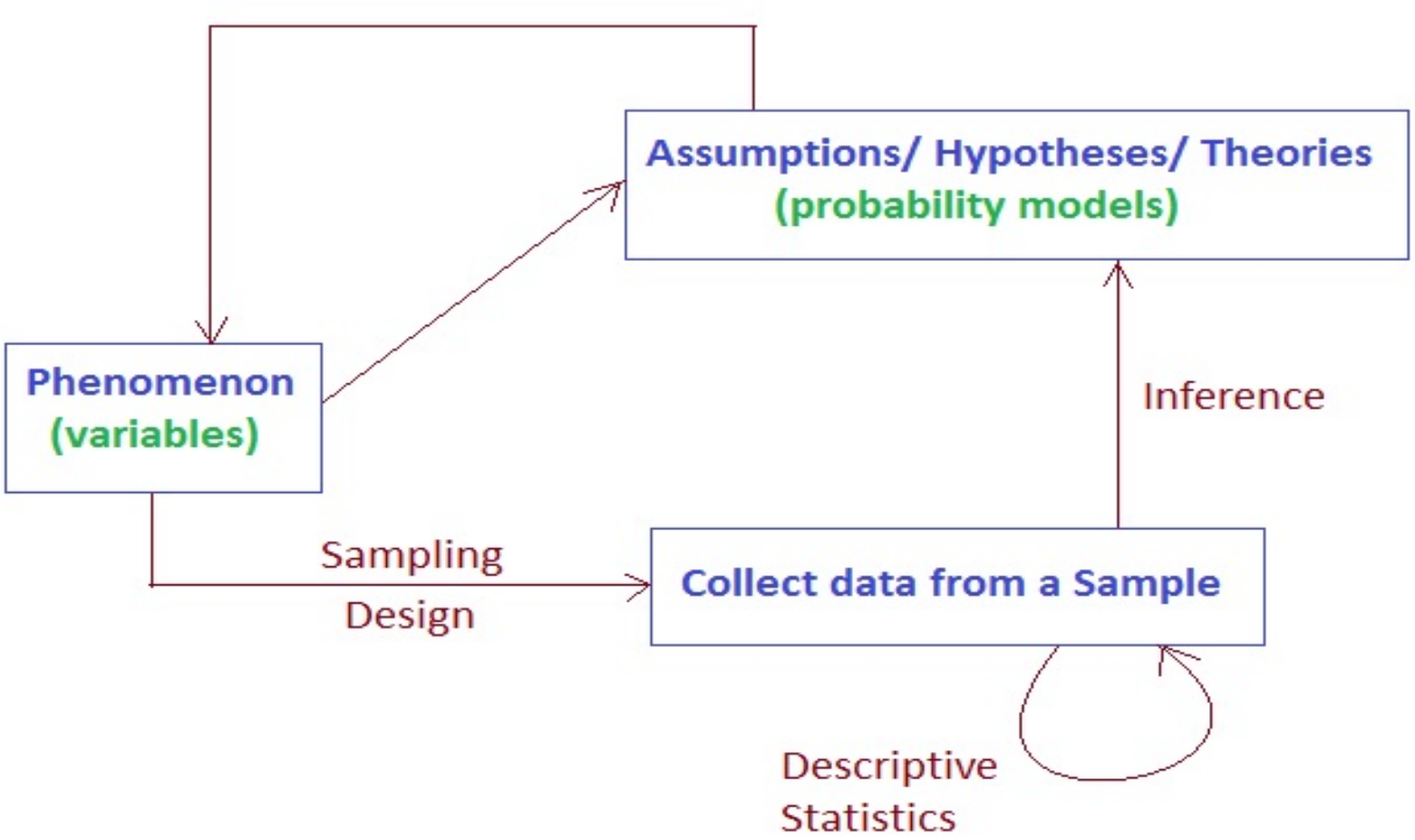
Any questions?

# Statistics

The material of this course consists of three parts:

1. Probability
   - We use probability in order to build models for the data.
2. Descriptive Statistics
   - We summarize and visualize a dataset.
3. Statistical inference
   - We use probability models in order to draw conclusions from datasets.

# Statistics


Assumptions/ Hypotheses/ Theories
(probability models)
Phenomenon
(variables)
Inference
Sampling
Design
Collect data from a Sample
Descriptive
Statistics

# The *Literary Digest* Poll

## We are in 1936...

- Franklin Roosevelt is the President of the U.S. and we are preparing to have elections.
- The country has just recovered from the Great Depression.
- The real income just started turning upwards and there are still 9 million unemployed.
- The Republican candidate is Alfred Landon, the governor of Kansas.
- Landon is campaigning on a program of economy in government, while Roosevelt is defensive about his deficit financing.

# The *Literary Digest* Poll

## The Poll

- The *Literary Digest* was a magazine that had an enormous prestige at the time.
- They had called the winner in every presidential election since 1916.
- They conducted a poll in 2.4 million individuals and predicted the victory of Landon:

| Landon | Roosevelt |
|---|---|
| 57% | 43% |

# The *Literary Digest* Poll

## What happened?

- Roosevelt won the 1936 elections by a landslide:

| Landon | Roosevelt |
|---|---|
| 38% | 62% |

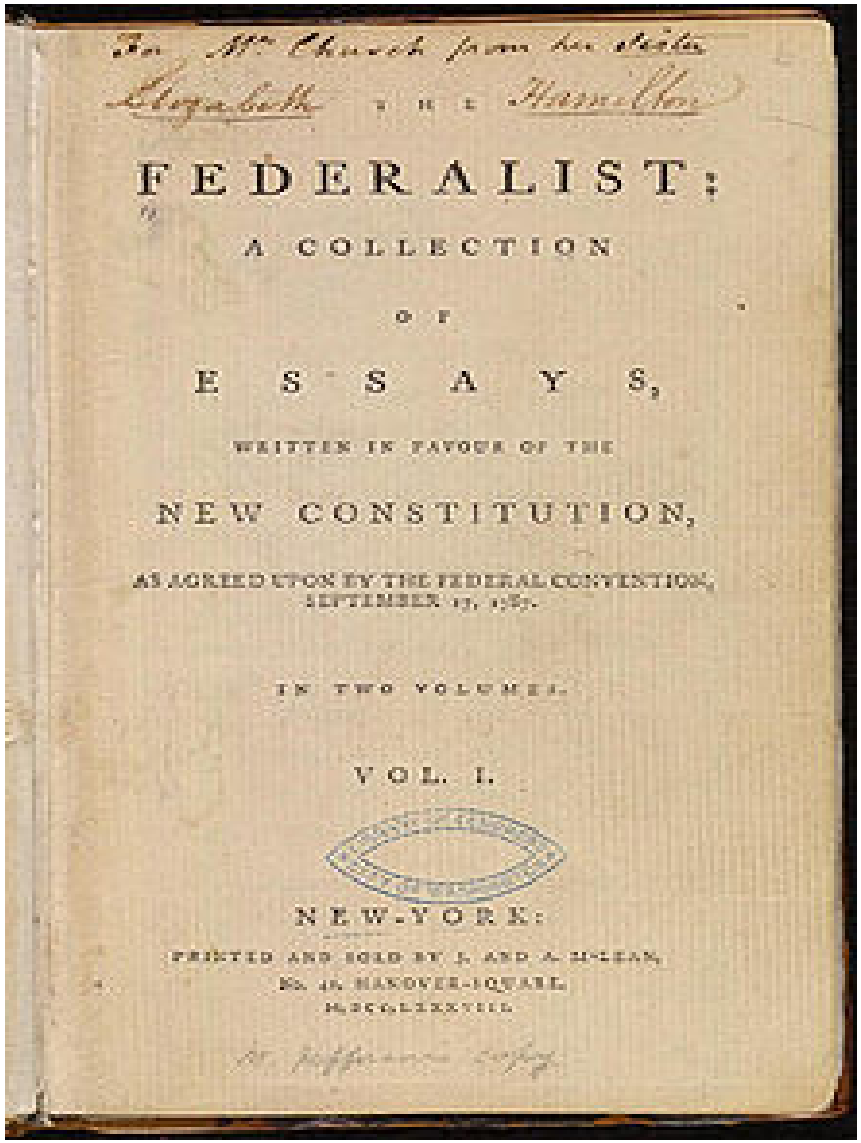
- The *Literary Digest* went bankrupt soon after.

## What did they do wrong?

# Who wrote this document?

* 85 essays written by
*James Madison*
*Alexander Hamilton* and
*John Jay*
under the pseudonym “Publius”.

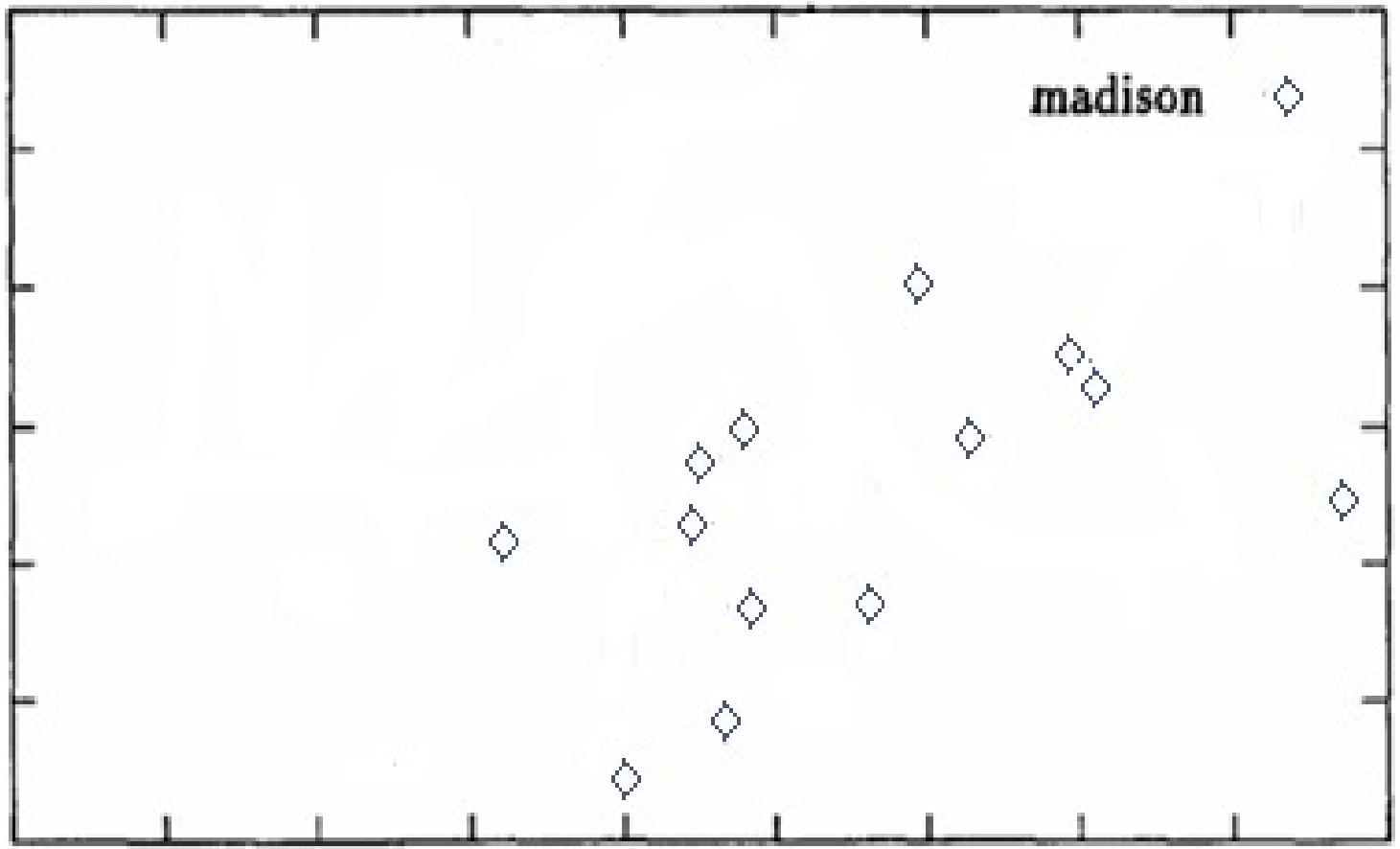
* Authorship of 11 has been disputed.



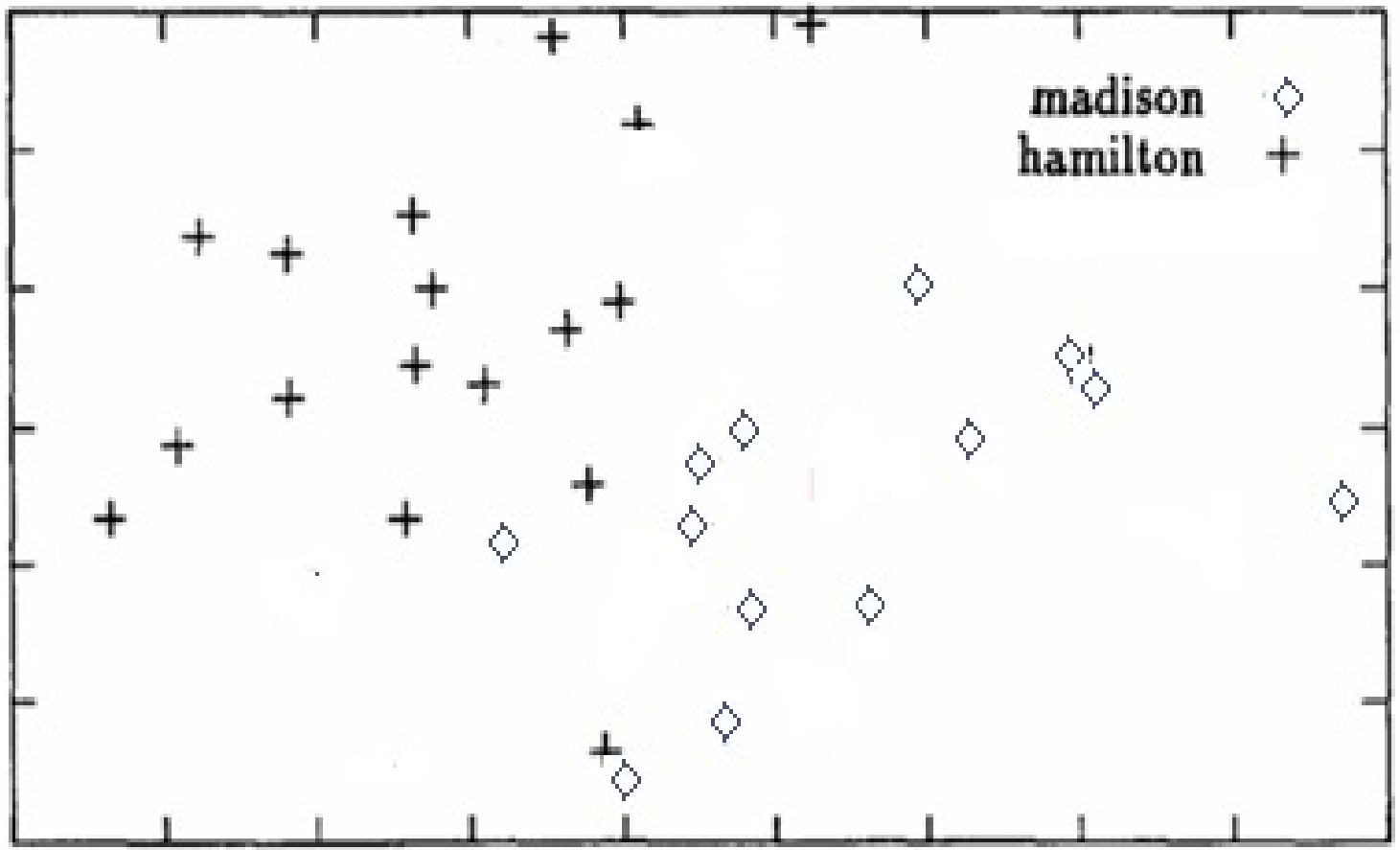
# Statistics can help!

- Hamilton uses the word ‘upon’ much more often than Madison.
- Hamilton uses ‘while’ in the sense of *‘at the same time as’*, but Madison uses ‘whilst’.

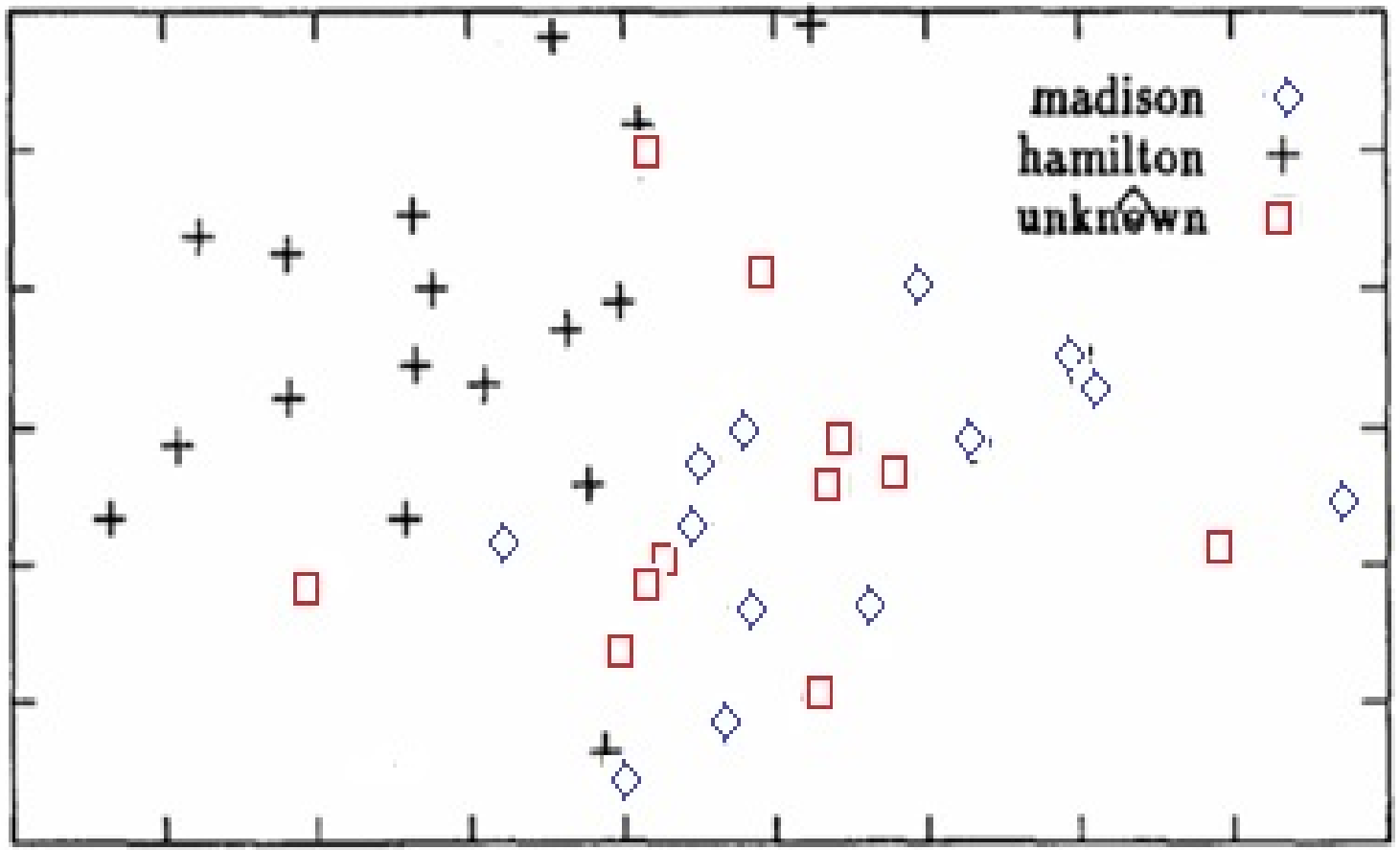
# Statistics can help!



# Statistics can help!



# Statistics can help!



# The Statistician's Objectives

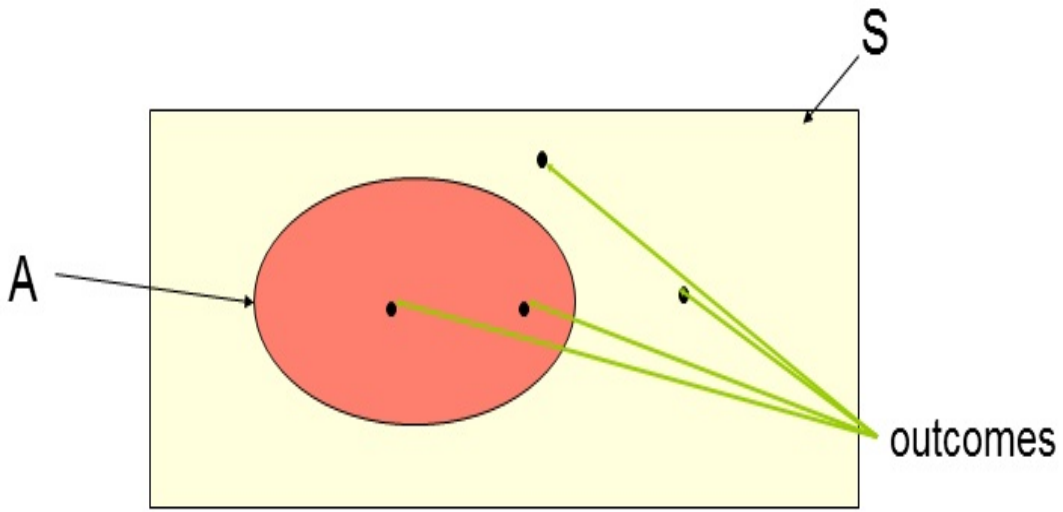
1. Ask the right questions
2. Collect useful data
3. Summarize the data
4. Make decisions and generalizations based on the data
5. Turn the data and decisions into new knowledge

# Introduction to Probability

# Modeling Random Phenomena

## Probability Models

- Outcomes: possible results of an experiment
- Sample space/ Universe/ Population: set of all possible outcomes (denoted by $S$ or $\Omega$).
- Event: set of outcomes (e.g. A in the picture)



# Example: *Rolling a die*

- Sample space:

$$\Omega = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$$

- Examples of events:

$$\begin{aligned} O &= \text{“}\textit{rolling an odd number}\text{”} = \{1,3,5\} \\ E &= \text{“}\textit{rolling an even number}\text{”} = \{2,4,6\} \\ A &= \text{“}\textit{rolling a 5 and a 6}\text{”} = \{5,6\} \\ F &= \text{“}\textit{rolling a 5 }\text{”} = \{5\}. \end{aligned}$$

# Example: *Fair coin tossing*

- Sample space:

$$\Omega = \{H, T\}$$

- If the coin is *fair*, then the **chances of it landing in Tails or Heads are the same**, i.e.

$$\text{Probability (Heads)} = \text{Probability (Tails)},$$

or in other words

$$\mathbb{P}(H) = \mathbb{P}(T).$$

# Probability

## Assigning probabilities

- If all all outcomes are *equally-likely*, i.e. have the same probability to occur

$$\mathbb{P}(\text{Event}) = \frac{\#\text{ outcomes in the event}}{\text{Total \# of outcomes}}.$$

# Example: *Rolling a die*

1. What is the probability that we get a 5?
2. What is the probability that the die lands on an odd number?

- We are able to use the *classical definition of probability*, because all numbers of a die are equally likely to occur, since this is a fair die.

# Example: *Rolling a die*

(i) Determine the sample space, which are the possible outcomes of a die:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

(ii) Determine the events that we are interested in:

$$F = \{ \text{ get a } 5\} = \{5\}.$$

$$O = \{ \text{ get an odd number } \} = \{1, 3, 5\}.$$

(iii) Compute the probability using the definition:

$$\mathbb{P}(F) = \frac{\text{\# of outcomes in } F}{\text{Total \# of possible outcomes}} = \frac{1}{6}.$$

$$\mathbb{P}(O) = \frac{\text{\# of outcomes in O}}{\text{Total \# of possible outcomes}} = \frac{3}{6} = \frac{1}{2}.$$

# Example: *Flip a Coin Three Times*

- Outcomes

| | | | |
|---|---|---|---|
| HHH | HHT | HTH | HTT |
| THH | THT | TTH | TTT |

1. Probability to get 3 Heads:

$$\mathbb{P}(HHH) = \frac{1}{8} = 0.125$$

# Example: *Flip a Coin Three Times*

- Outcomes

| | | | |
|---|---|---|---|
| HHH | HHT | HTH | HTT |
| THH | THT | TTH | TTT |

1. Probability to get 3 Heads:

$$\mathbb{P}(HHH) = \frac{1}{8} = 0.125$$

2. Probability to get 2 Heads:

$$\mathbb{P}(\text{Two Heads}) = \frac{3}{8} = 0.375$$

# Example: *Flip a Coin Three Times*

- Outcomes

| | | | |
|---|---|---|---|
| HHH | HHT | HTH | HTT |
| THH | THT | TTH | TTT |

1. Probability to get 3 Heads:

$$\mathbb{P}(HHH) = \frac{1}{8} = 0.125$$

2. Probability to get 2 Heads:

$$\mathbb{P}(\text{Two Heads}) = \frac{3}{8} = 0.375$$

3. Probability to get at least 2 Heads:

$$\mathbb{P}(\text{at least two Heads}) = \frac{4}{8} = 0.5$$

# Properties of probability

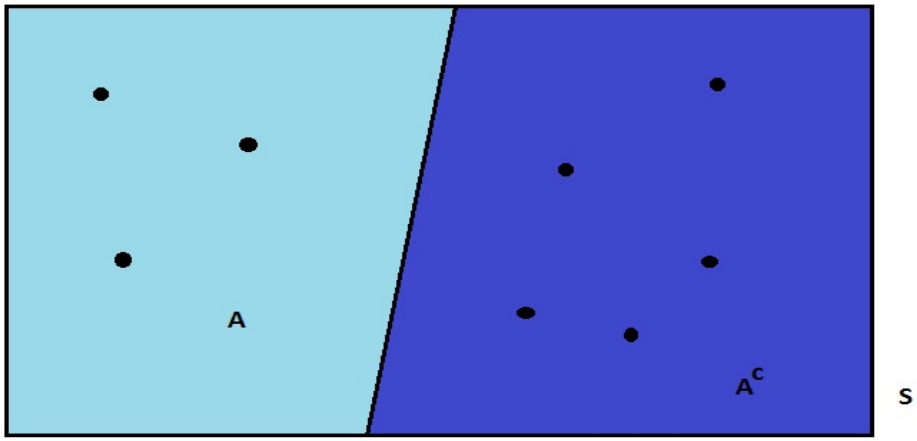
1. $0 \leq \mathbb{P}(A) \leq 1,$ for any event $A$.

2. $\mathbb{P}(\Omega) = 1$.
   An event that has probability 1 is called *certain*.

3. $\mathbb{P}(A) = 0$.
   An event that has probability 0 is called *impossible*.

# Complement/Opposite Events

$$\mathbb{P}(A) = 1 - \mathbb{P}(\text{opposite of} A) = 1 - \mathbb{P}(A^c)$$




## Example: *Rolling a Die*

$$A = \{\text{roll an odd number}\} = (1, 3, 5)$$

$$A^c = \{\text{NOT roll an odd number}\} = \{\text{roll an even number}\} = (2, 4, 6).$$

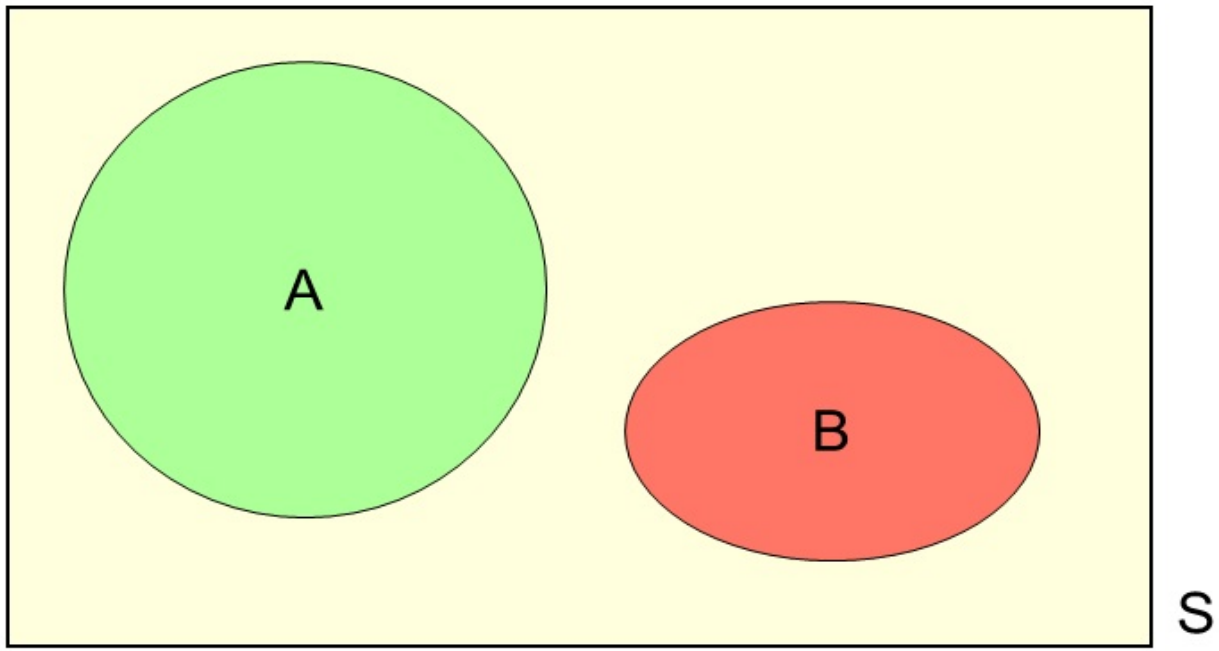
$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

$$\mathbb{P}(\text{odd number}) = 1 - \mathbb{P}(\text{even number}).$$

# OR Rule

- If $A$ and $B$ cannot both happen

$$\mathbb{P}(A \text{ OR } B) = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$




(*) $A$ and $B$ have no elements in common.

# Mutually Exclusive Events

- When $A$ and $B$ (denoted by $A \cap B$) cannot both happen are called *Mutually Exclusive*.

## Examples

1. Elections:

$$A = \text{“Obama elected president”}$$

$$B = \text{“ McCain elected president”}$$

2. Toss a coin

$$A = \{ \text{ get Heads } \} = \{H\}.$$

$$B = \{ \text{ get Tails } \} = \{T\}.$$

See you on Wednesday!