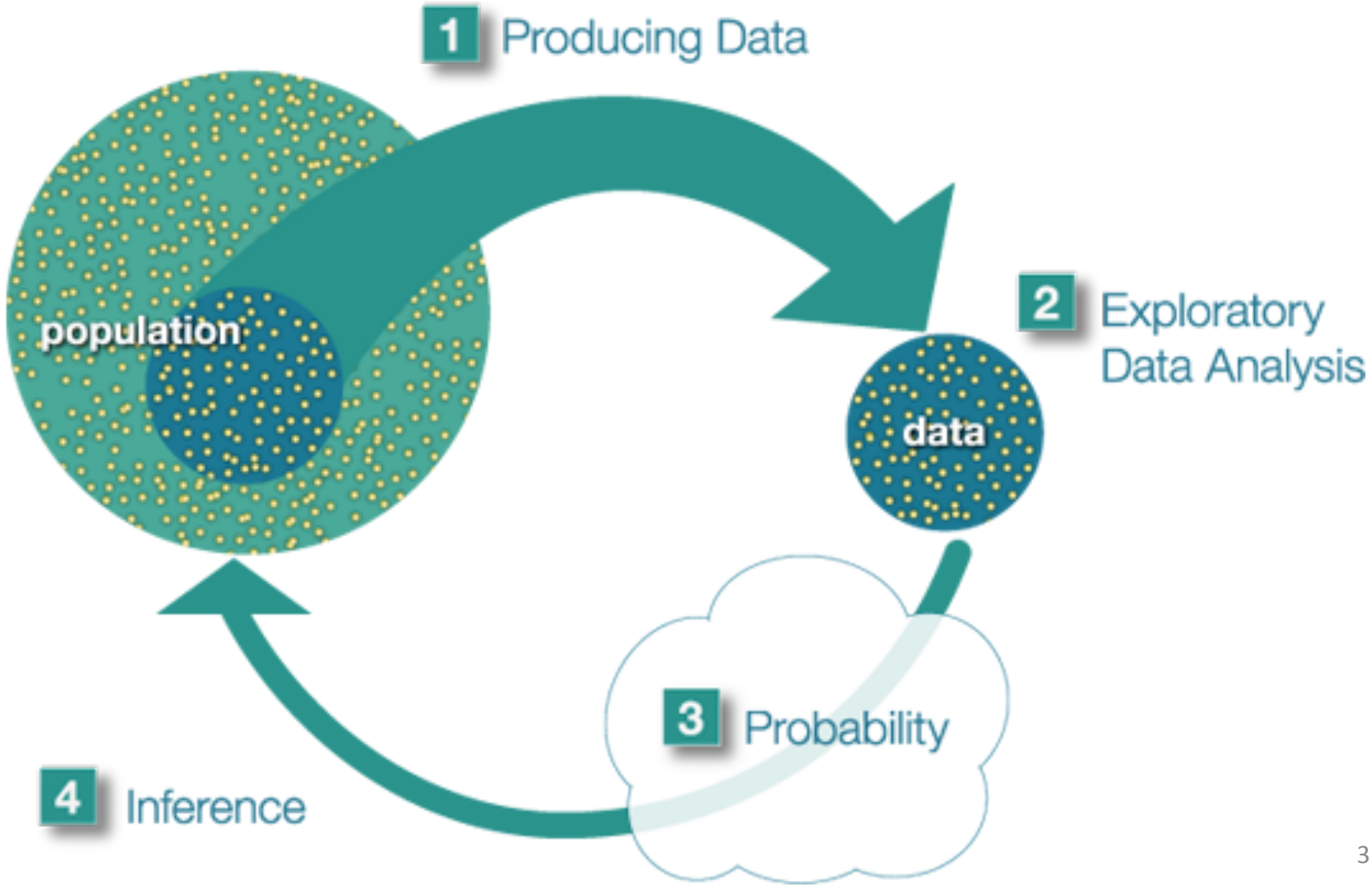
Introduction to Probability

Review: data and distributions

- Central tendency is important…
- …But in statistics, variance is more important
- Visualization of data can help us understand the distribution of a variable


1 Producing Data
population
2 Exploratory Data Analysis
data
3 Probability
4 Inference

Probability

- What is my chance of winning the lottery?
- Will I die in a plane crash?
- Will it rain today?
- Will I win this game of Chutes & Ladders?

Calculating Probability

$$\frac{\#\ \text{successes}}{\#\ \text{events}}$$

Calculating Probability

- If I flip a quarter, what is the probability that it will land on *Heads*?

$$\frac{\text{\# successes}}{\text{\# events}} \quad \frac{\text{1 Heads}}{\text{2 Sides}} = 50\%$$

Calculating Probability

- If a mother has three girls, what is the probability that the fourth child will be a girl?

Calculating Probability

Calculating Probability

Conditional Probability

- The probability that some event *A* happens, given (or "conditional upon") the occurrence of some other event *B*.

Conditional Probability

- What is the probability that I flip a Quarter two times and *both times* it lands on *Tails*?

$$\frac{\textbf{1 Tails}}{\textbf{2 Sides}} * \frac{\textbf{1 Tails}}{\textbf{2 Sides}} = \textbf{25\%}$$

Conditional Probability

- Four possible outcomes:

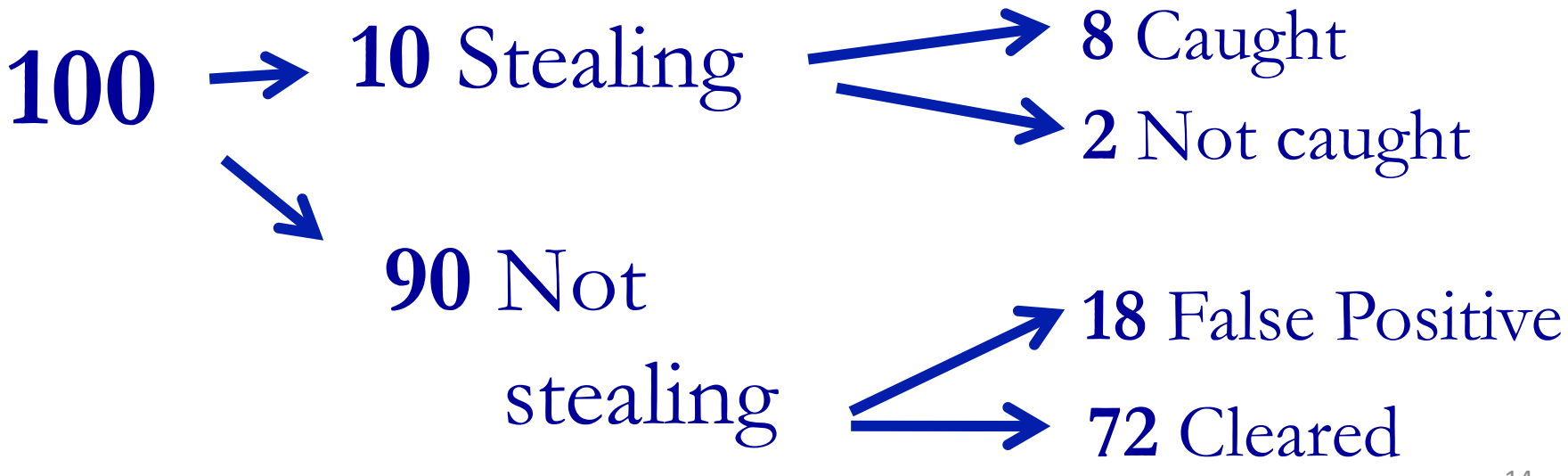
HH
HT
TH
TT

Conditional Probability

- 10% of employees are stealing (but we don't know which 10%)
- A lie-detector test is 80% accurate
- What is the probability that someone who is found to be stealing is actually doing so?

Natural Frequencies

Say you choose a base of 100…



Natural Frequencies

Say you choose a base of 100…

8 Caught

2 Not caught

18 False Positive

72 Cleared

$$\frac{\textbf{8}\text{ Caught}}{\textbf{8 + 18}\text{ All positives}} = \textbf{0.308}$$

Bayes' Theorem

$$\frac{P(\checkmark \text{Success})P(\text{Success})}{P(\checkmark \text{Success})P(\text{Success}) + P(\checkmark \text{Nonsuccess})P(\text{Nonsuccess})}$$

$$\frac{0.80 * 0.10}{(0.80 * 0.10) + (0.20 * 0.90)} = \mathbf{0.308}$$

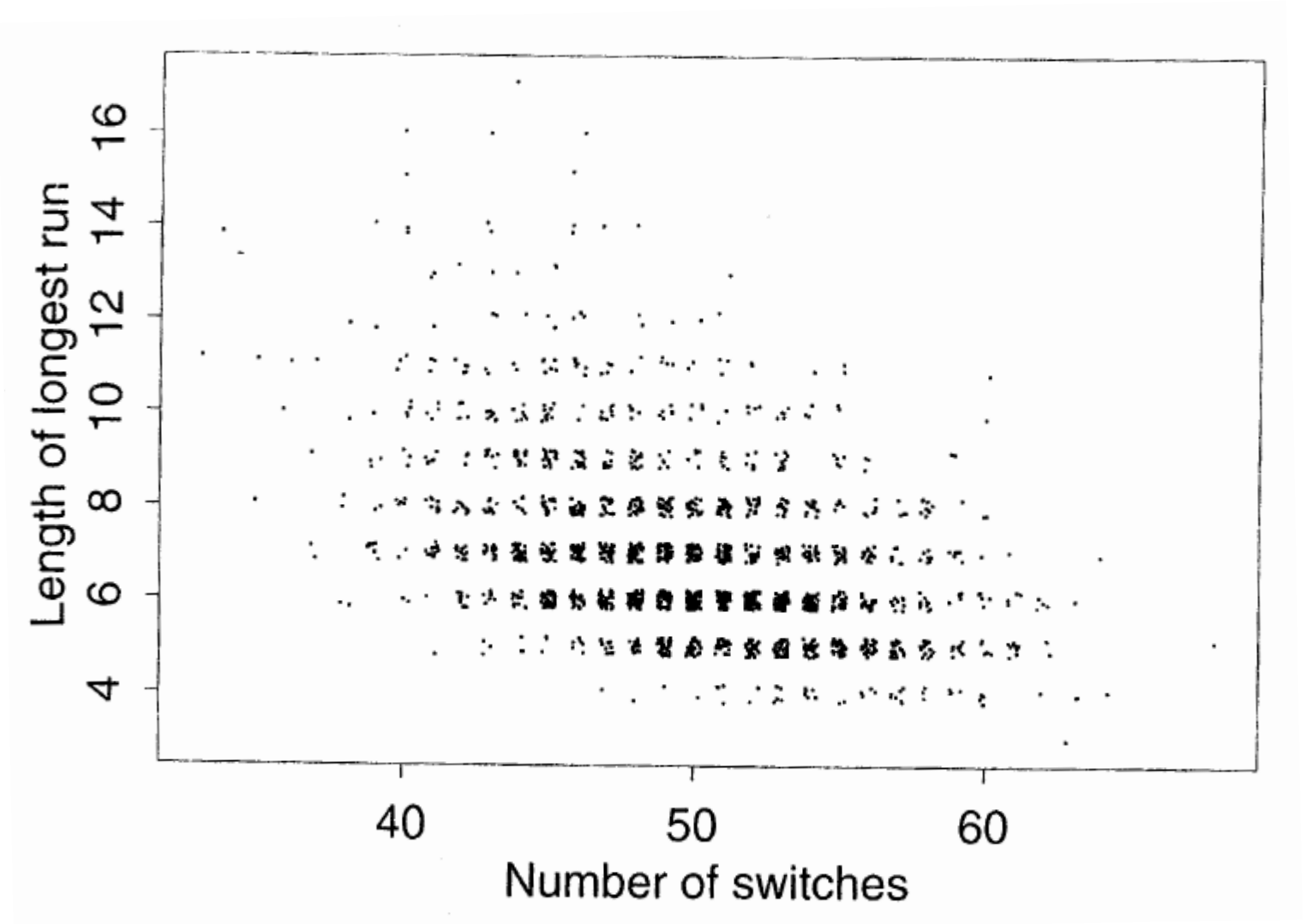
Conditional Probability

- Among women aged 40-50, with no symptoms or family history of breast cancer:
  - The probability that she has breast cancer is 0.8%
  - If a woman has breast cancer, the probability is 90% that she will have a positive mammogram
  - If a woman does not have breast cancer, the probability is 7% that she will still have a positive mammogram
- What is the probability that a woman who tests positive for breast cancer actually has breast cancer?

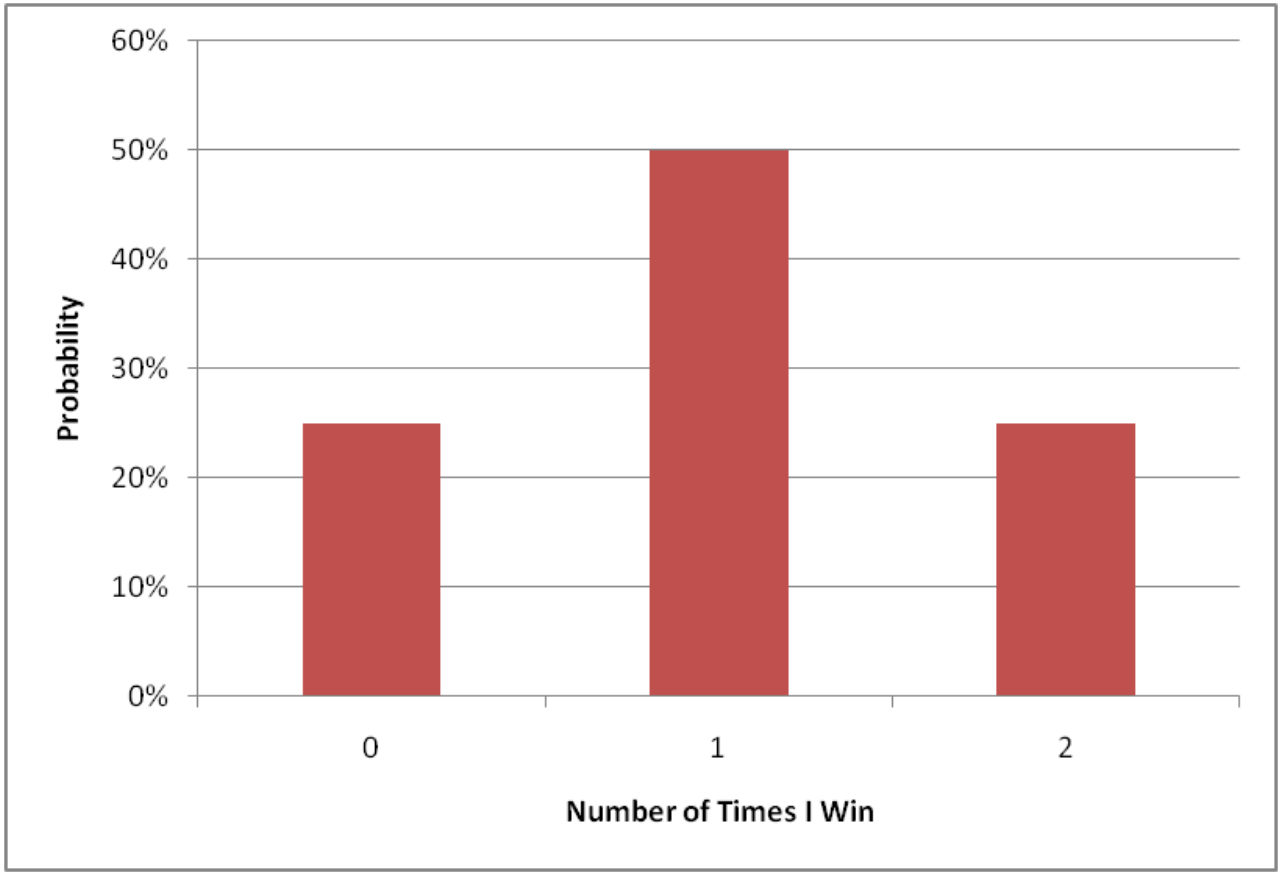
Conditional Probability

- What is the probability of this happening?


LIBERTY
IN GOD WE TRUST
1937
UNITED STATES OF AMERICA
QUARTER DOLLAR


Length of longest run
4
6
8
10
12
14
16
40
50
60
Number of switches

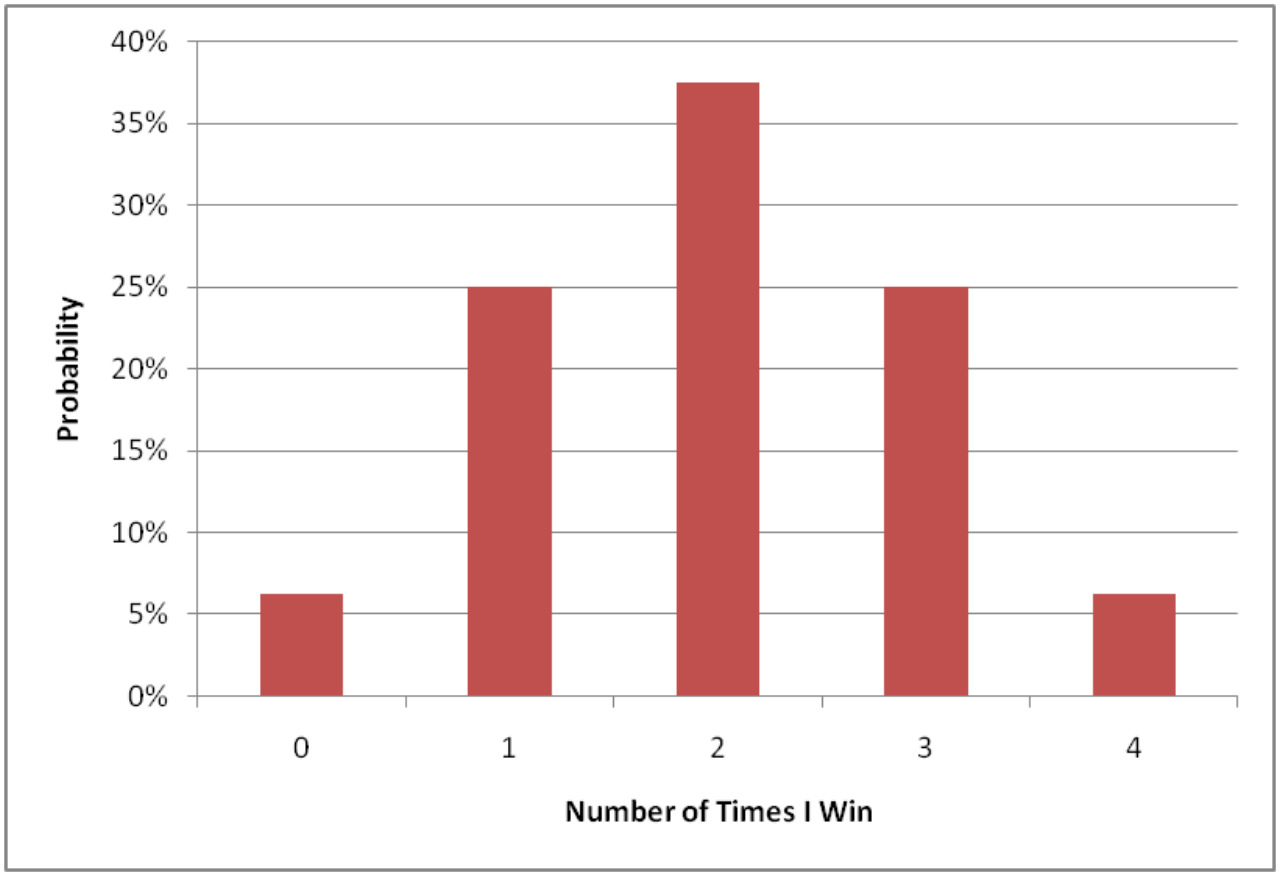
Two Games of Chutes & Ladders



Two Games of Chutes & Ladders

Number of Wins	Results	Number of Ways	Probability
0	L L	1	25%
1	L W W L	2	50%
2	W W	1	25%

Four Games of Chutes & Ladders



Four Games of Chutes & Ladders

Number of Wins	Results	Number of Ways	Probability
0	L L L L	1	6.25%
1	W L L L L W L L L L W L L L L W	4	25%
2	W W L L W L W L W L L W L L W W L W L W L W W L	6	37.5%
3	L W W W W L W W W W L W W W W L	4	25%
4	W W W W	1	6.25%

Probability for Multiple Combinations

$$\frac{n!}{2^n x!(n-x)!}$$

- **n** = number of events
- **x** = number of 'successes' within the event

Probability for Multiple Combinations

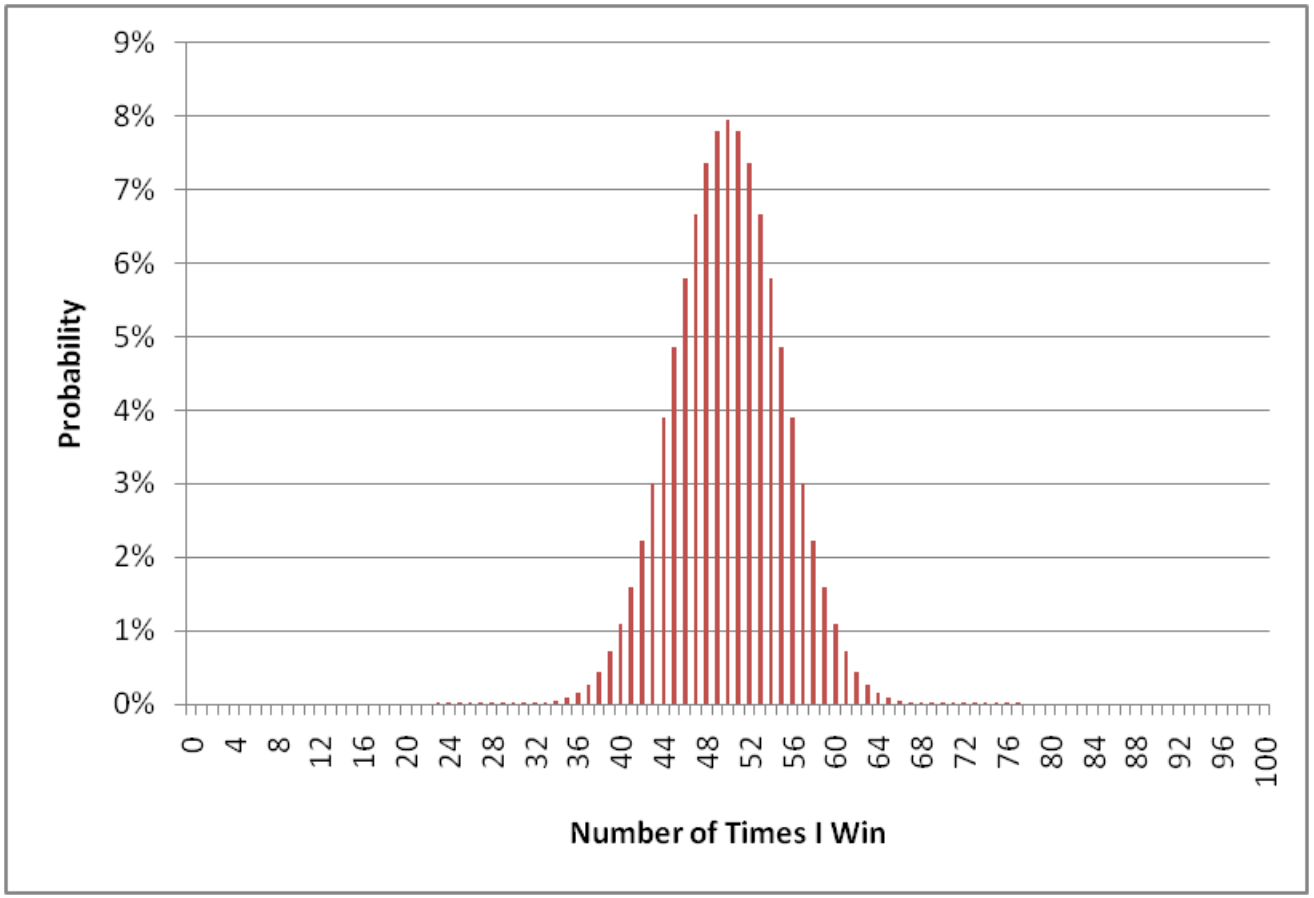
$$\frac{n!}{2^n x!(n-x)!}$$

- **n** = number of events
- **x** = number of 'successes' within the event

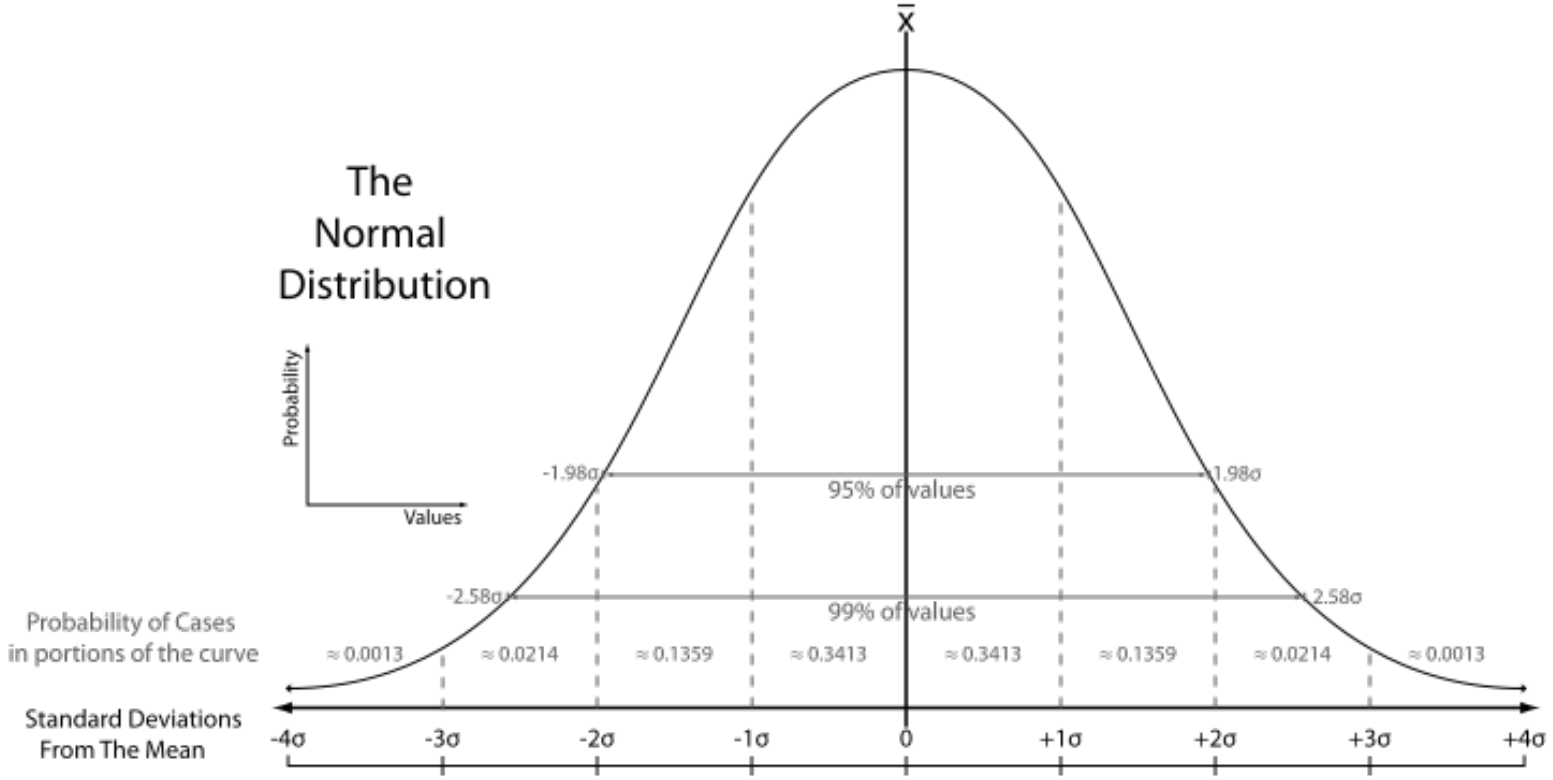
If we play 4 games (**n=4**), the probability of any combination of 2 wins (**x=2**) would be calculated as:

$4! / (2^4 * 2! * (4 - 2)!)$
$= 4! / (16 * 2 * 2)$
$= 24 / 64$
$= 0.375$ or 37.5%

100 Games of Chutes & Ladders



The Normal Distribution



Probability & Statistics

- In statistics, we estimate the probability that the data would be at least as extreme as those observed if there were no effect in the population.

Next Class

- Sampling Distribution

Review

- Probability estimates the likelihood of some random occurrence
- In practice, we use probability *distributions*

a)

T	H	H	H	H	T	T	T	T	H
T	T	T	T	T	T	T	T	T	T
T	T	H	T	T	H	T	H	T	H
H	T	T	T	T	T	T	H	H	H
H	H	H	T	H	T	H	T	T	H
T	T	H	H	H	H	H	T	H	T
T	T	T	H	H	H	T	H	T	H
H	H	T	H	T	T	T	T	T	H
H	T	T	T	H	H	T	H	H	T
T	H	T	H	H	T	H	T	T	T

b)

H	T	H	T	H	T	T	H	T	H
H	T	T	H	H	T	H	T	H	T
H	H	H	T	H	T	H	H	T	H
H	T	T	T	H	H	H	T	T	H
H	H	T	T	T	H	T	T	T	T
T	H	H	H	H	T	T	T	H	H
H	T	T	H	T	H	T	T	H	H
T	T	H	H	H	T	T	T	H	T
H	T	T	T	H	H	H	T	T	H
H	T	H	T	H	T	T	H	T	H

c)

T	T	H	T	H	T	H	H	H	H
T	H	T	H	H	T	T	H	T	H
H	H	H	H	T	H	T	H	T	T
T	T	T	T	H	T	H	T	H	T
H	H	H	T	T	H	T	H	H	T
H	H	T	H	T	T	H	T	H	H
T	T	T	T	H	T	T	T	T	T
H	T	T	H	H	T	H	H	T	H
H	T	H	H	H	H	T	T	H	T
T	T	H	H	T	H	H	T	T	T

d)

H	T	H	H	T	T	T	H	T	T
H	H	H	T	H	T	T	T	T	H
T	H	T	T	T	T	T	H	T	T
H	H	T	T	H	T	H	H	H	H
T	H	T	H	T	H	H	T	T	H
H	T	H	H	H	T	H	H	T	T
H	H	T	T	H	T	H	T	H	H
H	T	T	T	H	H	H	T	T	H
T	H	T	T	H	H	T	H	T	H
H	T	T	T	T	T	T	T	H	T