# Class 5

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

# Agenda:

## Recap Chapter 3.2 - 3.3

## Lecture Chapter 4.1 - 4.2

# Recap Chapter 3.2 - 3.3

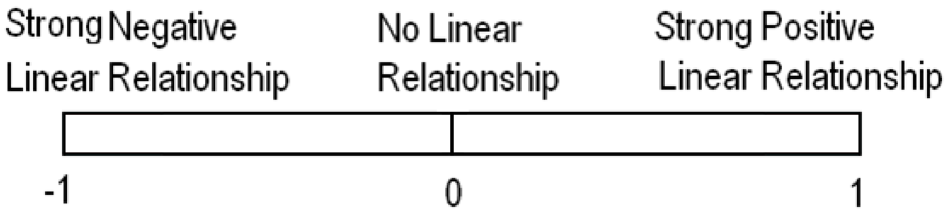
# 3: Descriptive Analysis and Bivariate Data

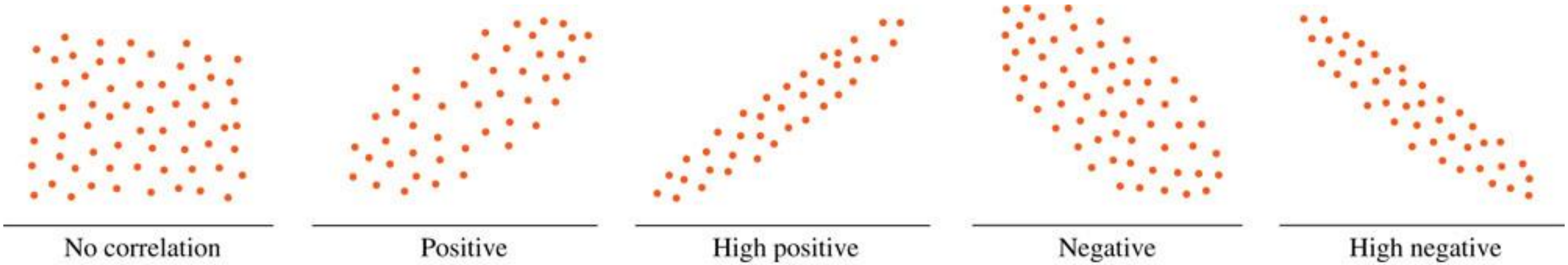
## 3.2 Linear Correlation

$$-1 \le r \le 1$$

**Linear Correlation**, $\boldsymbol{r}$, is a measure of the strength of a linear relationship between two variables $x$ and $y$.

**positive relationship:**
as $x$ increases so does $y$
**negative relationship:**
as $x$ increases $y$ decreases




$r \approx 0$ $\quad r \approx 0.5$ $\quad r \approx 0.8$ $\quad r \approx -0.5$ $\quad r \approx -0.8$

Figure from Johnson & Kuby, 2012.

# 2: Descriptive Analysis and Single Variable Data

## 2.4 Measures of Dispersion

**Sample Variance:** The mean of the squared deviations using $n$-1 as a divisor. p. 75

There are two equivalent formulas that can be used.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n}x_i^2 - \left[\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2\right]\right\} = \frac{SS(x)}{n-1}$$

where $x_i$ is $i^{\text{th}}$ data value, $\bar{x}$ is sample mean, $n$ is sample size.

# 3: Descriptive Analysis and Bivariate Data

## 3.2 Linear Correlation

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{n-1} = \frac{SS(x)}{n-1}$$

Sum of Squared $x$ Deviations

$$s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2}{n-1} = \frac{SS(y)}{n-1}$$

Sum of Squared $y$ Deviations

Sum of Squared $xy$ Deviations

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n-1} = \frac{SS(xy)}{n-1}$$

# 3: Descriptive Analysis and Bivariate Data

## 3.2 Linear Correlation

**Example:**

| Push-ups, $x$ | 27 | 22 | 15 | 35 | 30 | 52 | 35 | 55 | 40 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sit-ups, $y$ | 30 | 26 | 25 | 42 | 38 | 40 | 32 | 54 | 50 | 43 |

$$\sum_{i=1}^{n} x_i = 351 \qquad \sum_{i=1}^{n} x_i^2 = 13717 \qquad \sum_{i=1}^{n} y_i = 380 \qquad \sum_{i=1}^{n} y_i^2 = 15298 \qquad \sum_{i=1}^{n} x_i y_i = 14257$$

$$SS(x) = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 13717 - \frac{(351)^2}{10} = 1396.9$$

$$SS(y) = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 15298 - \frac{(380)^2}{10} = 858.0$$

$$SS(xy) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = 14257 - \frac{(351)(380)}{10} = 919.0$$

$$r = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}} = \frac{919.0}{\sqrt{(1396.9)(858.0)}} = 0.84$$

**Questions?**

Figure from Johnson & Kuby, 2012.

$$r = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}}$$

# 3: Descriptive Analysis and Bivariate Data

## 3.2 Linear Correlation

$n=10$

**Example:**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Push-ups, $x$ | 27 | 22 | 15 | 35 | 30 | 52 | 35 | 55 | 40 | 40 |
| Sit-ups, $y$ | 30 | 26 | 25 | 42 | 38 | 40 | 32 | 54 | 50 | 43 |

$$\sum_{i=1}^{n} x_i = 351$$

$$\sum_{i=1}^{n} x_i^2 = 13717$$

$$\sum_{i=1}^{n} y_i = 380$$

$$\sum_{i=1}^{n} y_i^2 = 15298$$

$$\sum_{i=1}^{n} x_i y_i = 14257$$

$$SS(x) = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 13717 - \frac{(351)^2}{10} = 1396.9$$

$$SS(y) = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 15298 - \frac{(380)^2}{10} = 858.0$$

$$SS(xy) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = 14257 - \frac{(351)(380)}{10} = 919.0$$

$$r = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}} = \frac{919.0}{\sqrt{(1396.9)(858.0)}} = 0.84$$

**Questions?**

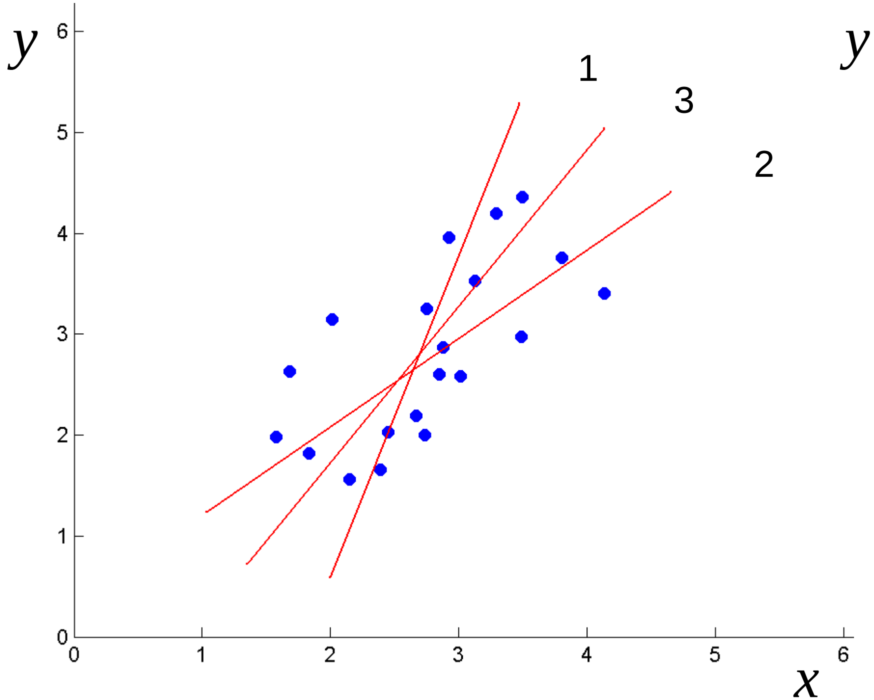
Figure from Johnson & Kuby, 2012.

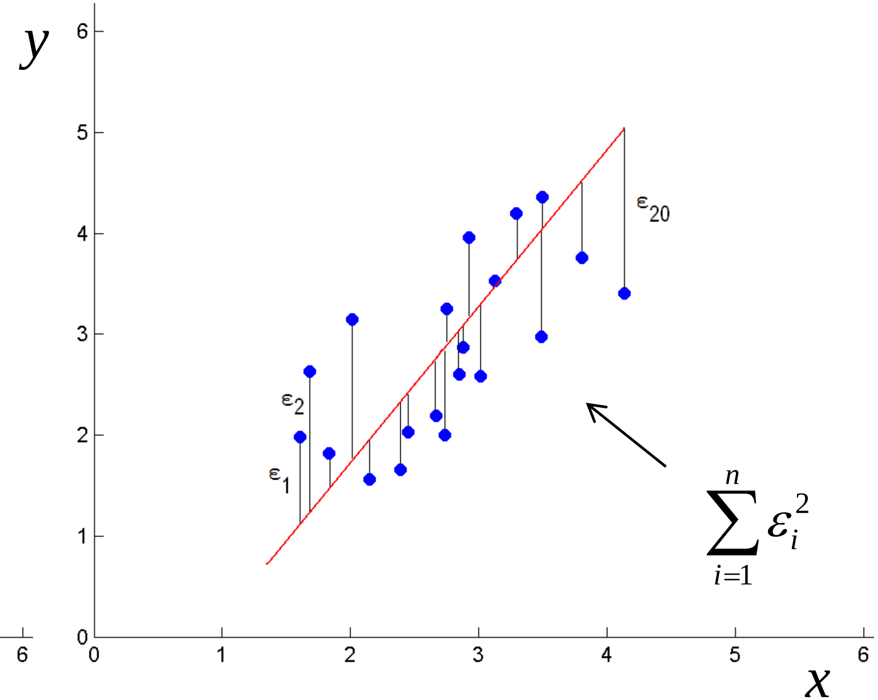
# 3: Descriptive Analysis and Bivariate Data

## 3.3 Linear Regression

$b_0$ is estimated y-intercept
$b_1$ is estimated slope.

We try different lines until we find the "best" one, $\hat{y} = b_0 + b_1 x$ .




Move line until sum of the squared residuals is a minimum.

# 3: Descriptive Analysis and Bivariate Data

## 3.3 Linear Regression

($x$,$y$) pairs: (1,1),(3,2),(2,3),(4,4)

Plotted points.

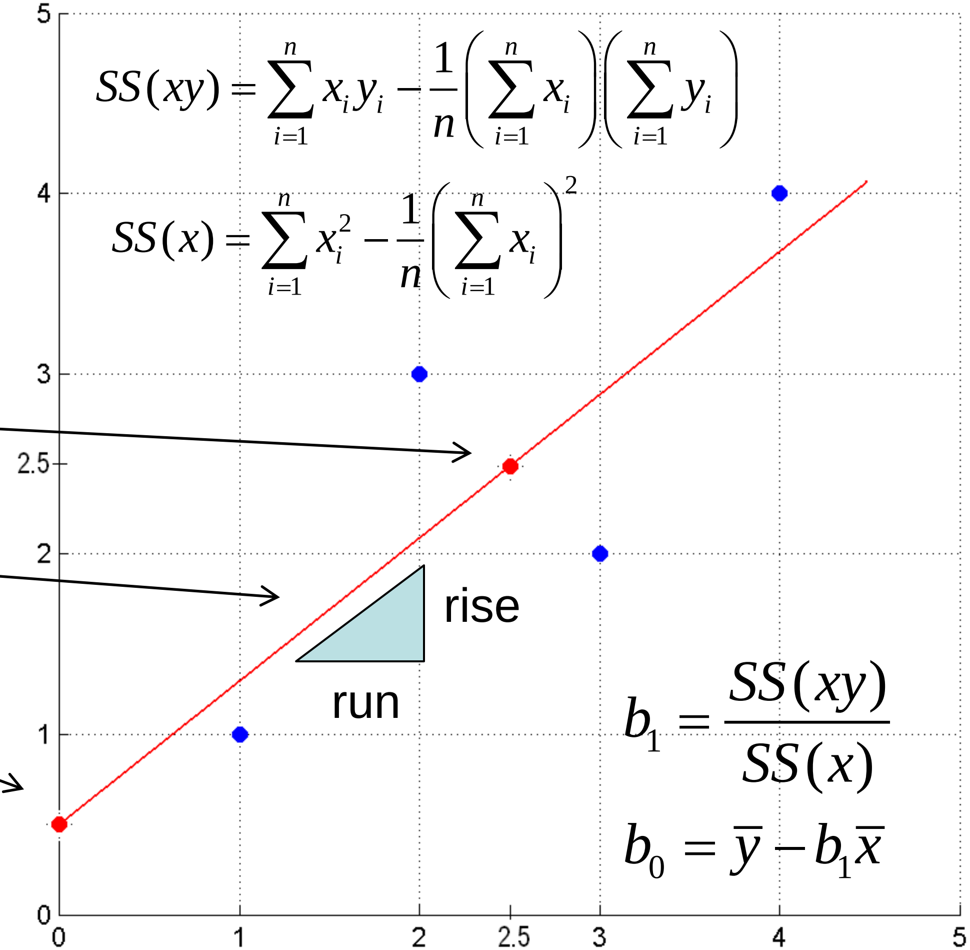
The line goes through $(\bar{x}, \bar{y})$ .

The slope is $b_1$=0.8.

The $y$ - intercept $b_0$=0.5 .

Two points (2.5,2.5) and (0,.5).

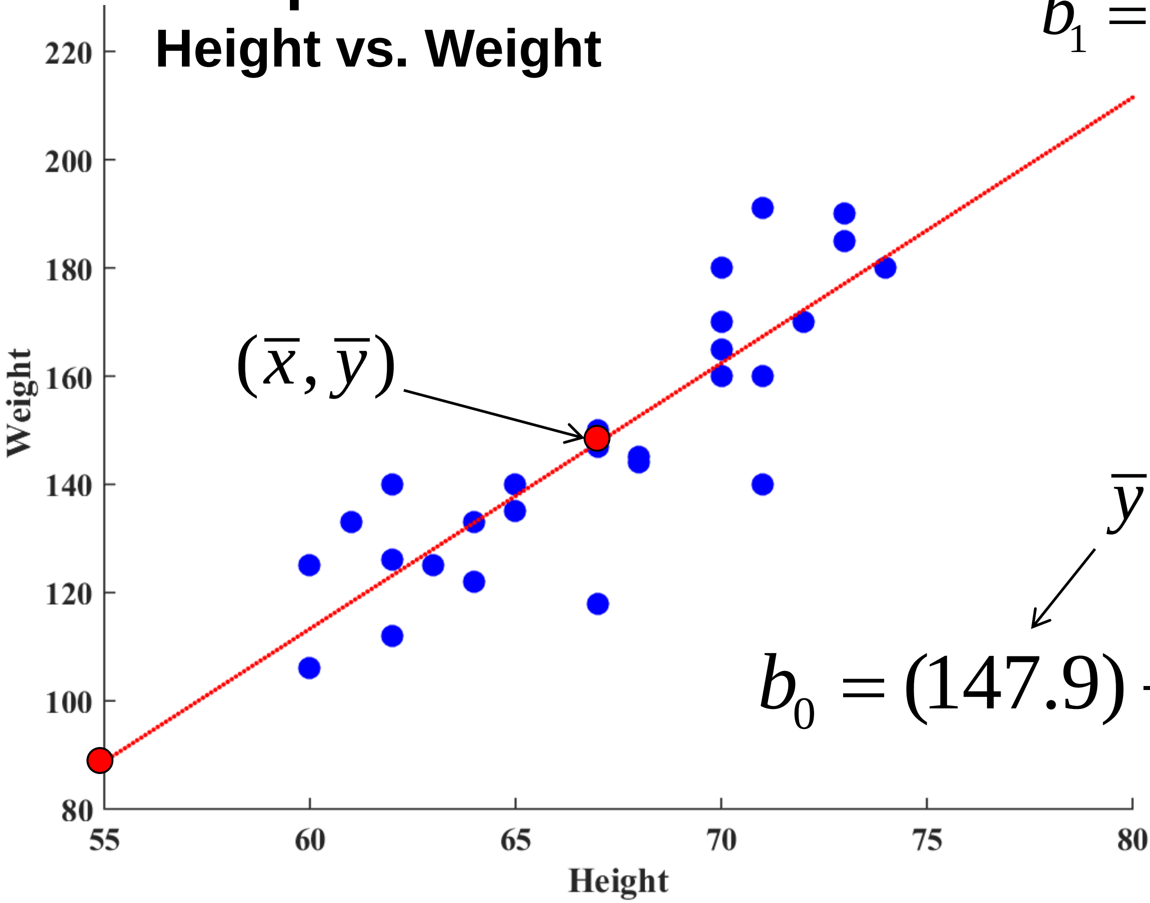
$\sum x = 10 \quad \sum y = 10 \quad \sum xy = 29$



# 3: Descriptive Analysis and Bivariate Data

## 3.3 Linear Regression

### Example: Previous class data!




$$b_1 = \frac{SS(xy)}{SS(x)} = \frac{2370.1}{483.0} = 4.9$$

units of lbs/in

$$b_0 = (147.9) - (4.9)(67.0) = -180.4$$

($\bar{y}$, $b_1$, $\bar{x}$)

point-slope formula

# Chapter 4: Probability

Daniel B. Rowe, Ph.D.

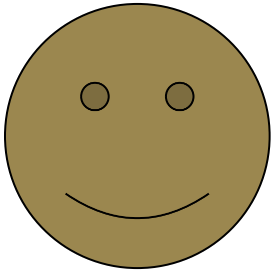
Department of Mathematical and Statistical Sciences

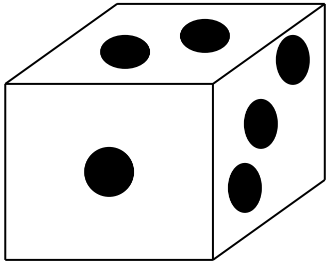
# 4: Probability
## 4.1 Probability of Events

Let's talk about **experiments**, **events**, and **probabilities**.

An **experiment** is a process by which a measurement is taken or observations is made.

i.e. *flip coin* or *roll die*





# 4: Probability

## 4.1 Probability of Events

An **outcome** is the result of an experiment. i.e. *Heads*, or *3*

Coin: $O_1=H,\ O_2=T$
Die: $O_1=1,\ O_2=2,\ O_3=3,\ O_4=4,\ O_5=5,\ O_6=6$

**Sample space** is a listing of possible outcomes.
$S=\{O_1,O_2\}$ or $S=\{O_1,O_2,O_3,O_4,O_5,O_6\}$

Coin: $S=\{H,T\}$
*Die:* $S=\{1,2,3,4,5,6\}$

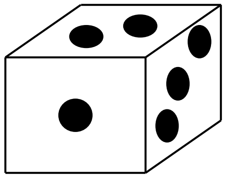
# 4: Probability

## 4.1 Probability of Events

An **event** $A$ is an outcome or a combination of outcomes.
i.e. $A$=*even number when rolling a die*=$\{2,4,6\}$

The probability of an event $A$ is written $P(A)$.

i.e. $P(A)=P($*even number when rolling a die)*
$=P(\{2,4,6\})$
=3/6



# 4: Probability

## 4.1 Probability of Events - Properties

**Property 1**

In words:

"A probability is always a numerical value between 0 and 1."

In algebra:

$A$ is an event

$$0 \le \text{each } P(A) \le 1$$

$P$(Heads on coin flip),
or $P$(*3* on roll of die)

If the event $A$ can never occur, then $P(A)$=0.
If the event $A$ is sure to occur, then $P(A)$=1.

# 4: Probability

## 4.1 Probability of Events - Properties

**Property 2**

In words:

"The sum of probabilities for all outcomes of an experiment is equal to exactly 1."

In algebra:

$O_i$ are nonoverlapping outcomes that include all possibilities

$$\sum_{i=1}^{n} P(O_i) = 1 \qquad i = 1, \ldots, n$$

The book uses $A$.

# 4: Probability

## 4.1 Probability of Events

Now that we talked about events and probabilities, how do we get probabilities of events?

**Probability of a event:** The relative frequency with which that event can be expected to occur.

There are three different approaches to probability.

(1) Empirical (AKA experimental)
(2) Theoretical (AKA classical or equally likely)
(3) Subjective (expression of belief, Bayesian, not discuss)

# 4: Probability
## 4.1 Probability of Events

**Empirical (Observed) Probability: $P'(A)$**

In words:

$$\text{empirical probability of } A = \frac{\text{number of times } A \text{ occured}}{\text{number of trials}}$$

In algebra:

$$P'(A) = \frac{n(A)}{n}$$

The " ' " in $P'(A)$ means an empirical probability.

# 4: Probability

## 4.1 Probability of Events – Law of large numbers

Had computer flip a coin 500 times.

Flip # on $x$ axis

$P'(H)$ on $y$ axis.

This shows convergence to true value of 1/2.

| $F$ | $O$ | $X$ | $P'$ |
|---|---|---|---|
| 1 | T | 0 | 0.0000 |
| 2 | T | 0 | 0.0000 |
| 3 | H | 1 | 0.3333 |
| 4 | T | 1 | 0.2500 |
| 5 | T | 1 | 0.2000 |
| 6 | H | 2 | 0.3333 |
| 7 | H | 3 | 0.4286 |
| 8 | T | 3 | 0.3750 |
| 9 | T | 3 | 0.3333 |
| 10 | T | 3 | 0.3000 |

$$P'(H) = \frac{\text{\# of heads}}{\text{\# coin flips}}$$

# 4: Probability

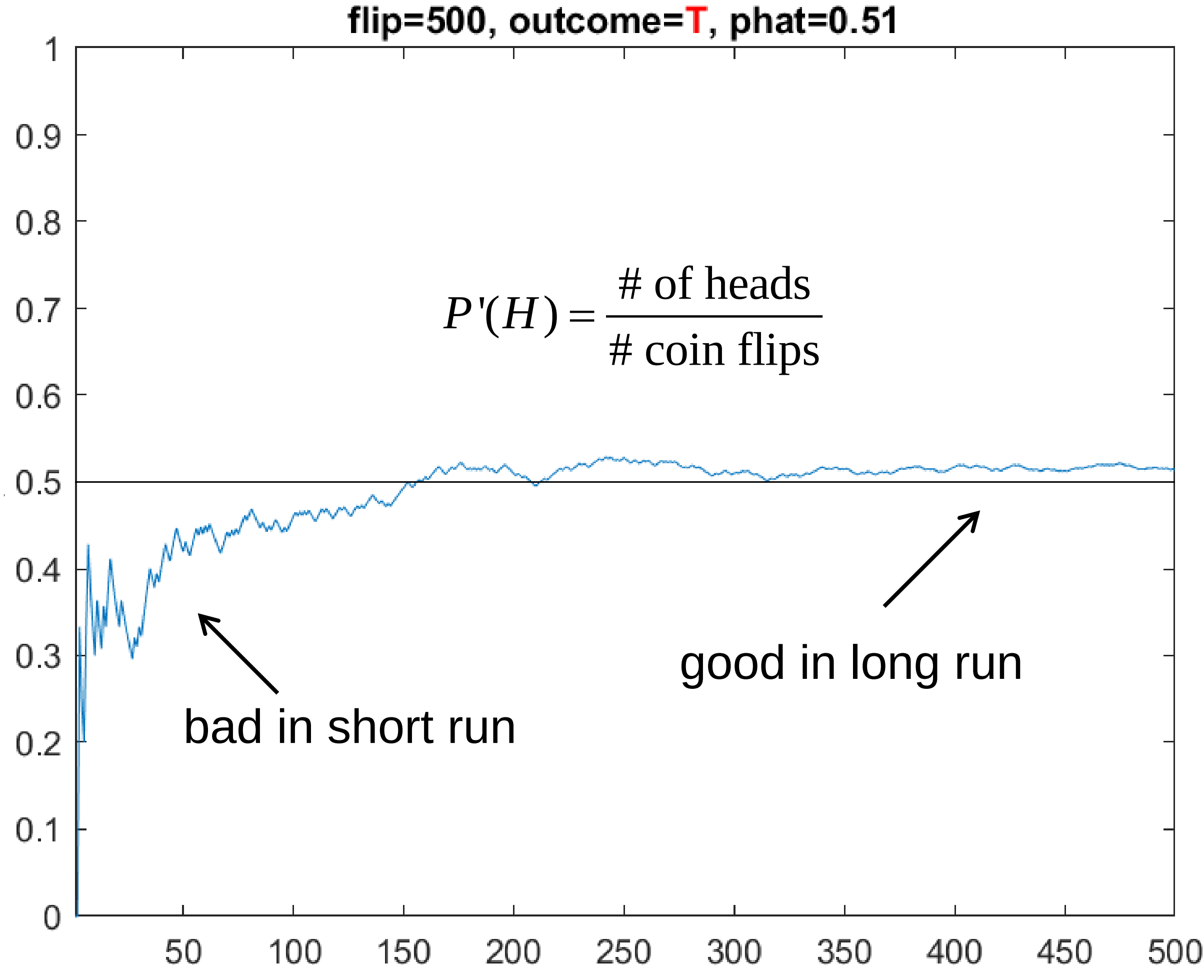
## 4.1 Probability of Events – Law of large numbers

Had computer flip a coin 500 times.

Flip # on $x$ axis

$P'(H)$ on $y$ axis.

This shows convergence to true value of 1/2.






Play Video

https://mssc.mu.edu/~daniel/math1700/fall2024/flipvid1.mp4

# 4: Probability

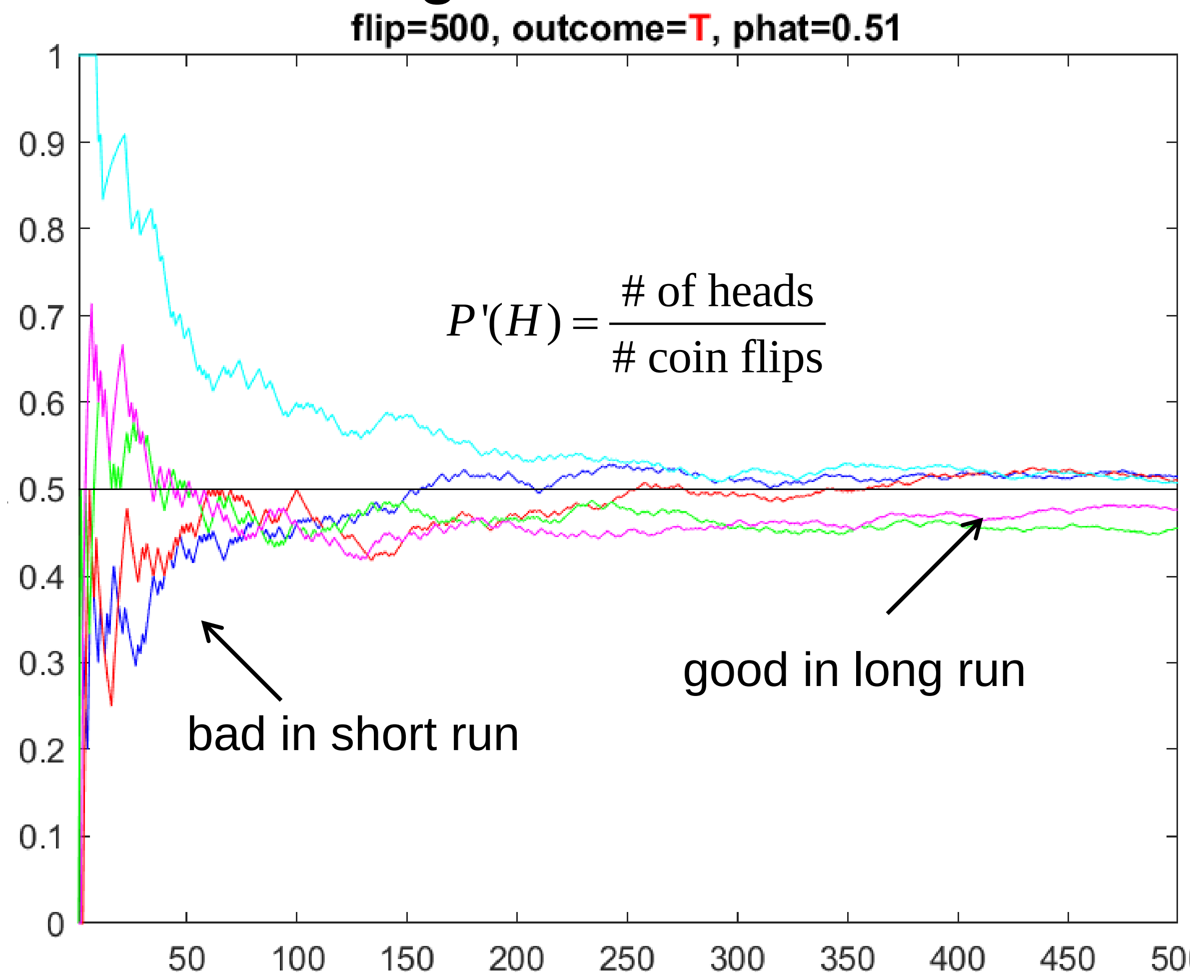
## 4.1 Probability of Events – Law of large numbers

Had computer flip
a coin 500 times.

Flip # on $x$ axis

$P'(H)$ on $y$ axis.

This shows convergence
to true value of 1/2.



# 4: Probability

## 4.1 Probability of Events

In the empirical method you actually have to perform the experiment of flipping the coin.

The empirical approach may be off in the short run.

Suppose you get on a streak and out of 10 flips all 10 are heads?

By the empirical method we would say that $P'(H) = 1$.

# 4: Probability

## 4.1 Probability of Events

**Theoretical (Expected) Probability: $P(A)$**

In words:

$$\text{theoretical probability of } A = \frac{\text{number of times } A \text{ occus in sample space}}{\text{number of elements in the sample space}}$$

In algebra:

$$P(A) = \frac{n(A)}{n(S)}$$

# 4: Probability

## 4.1 Probability of Events

So let's examine what could potentially happen when we flip a coin twice.

Before flip.

# 4: Probability

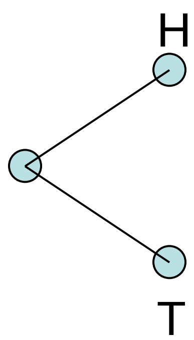
## 4.1 Probability of Events

So let's flip a coin twice.

**Sample space:**
listing of outcomes
for 1 flip

Flip once.

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




$$P(H) = \frac{\text{\# times H occurs in } S}{\text{\# elements in } S}$$

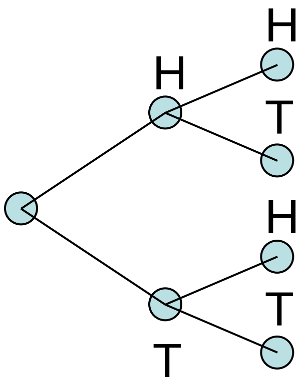
# 4: Probability

## 4.1 Probability of Events

So let's flip a coin twice.

**Sample space:** listing of outcomes for 2 flips

Flip twice.




$$S = \{HH, HT, TH, TT\}$$

$$P(HH) = \frac{\#\text{ times } HH \text{ occurs in } S}{\#\text{ elements in } S}$$

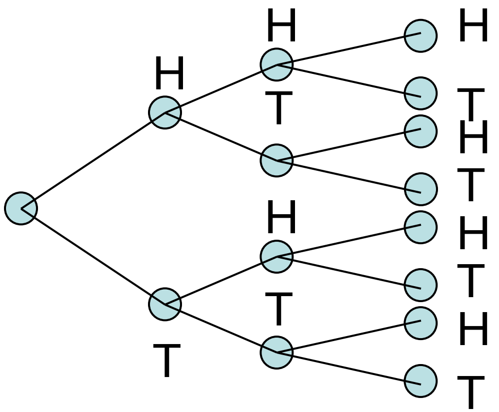
# 4: Probability

## 4.1 Probability of Events

So let's flip a coin three times.

**Sample space:** listing of outcomes for 3 flips

Can flip three times.




$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

$$P(HHH) = \frac{\text{\# times } HHH \text{ occurs in } S}{\text{\# elements in } S}$$

# 4: Probability
## 4.1 Probability of Events

In the theoretical method you do not have to perform the experiment of flipping the coin.

If each of the events are equally likely, then the theoretical approach is correct from the start.

If the events are not equally likely, then the theoretical method is not correct and we should use a different approach.

# 4: Probability

## 4.1 Probability of Events – Probabilities as odds

If the odds in favor of an event $A$ **are a to b (**or **a:b),** then

1. The odds against event $A$ are **b to a** (or **b:a**).
2. The probability of event $A$ is $P(A) = \frac{a}{a+b}$.
3. The probability that event $A$ will not occur is

$$P(\text{not } A) = \frac{b}{a+b}$$

# 4: Probability

## 4.1 Probability of Events

The odds against event $A$ are $b$ to a (or b:a).
The probability of event $A$ is $P(A)=\frac{a}{a+b}$ .

Therefore, if we are at the race track and we define

$A$ = our horse wins the race.

If the odds against $A$ are 100 to 1 (100:1),

then the probability of $A$ is $P(A)=\frac{1}{100+1}=\frac{1}{101}$

# 4: Probability
## 4.2 Conditional Probability of Events

We use conditional probability in our daily lives and sometimes do not realize it.

What is the probability that the Professor will put an exam question on topic $x$?

What is the probability that the Professor will put an exam question on topic $x$ given that he covered topic $x$ in class?

# 4: Probability

## 4.2 Conditional Probability of Events

What is the probability that the Professor will put an exam question on topic $x$?

What is the probability that the Professor will put an exam question on topic $x$ given that he covered topic $x$ in class?

Let $A$ = Professor will put an exam question on topic $x$
$B$ = he covered topic $x$ in class

$P(A)$ *vs.* $P(A|B)$

# 4: Probability

## 4.2 Conditional Probability of Events

**Conditional probability an event will occur:** A conditional probability is the relative frequency with which an event can be expected to occur under the condition that that additional preexisting information is known about some other event.

$P(A|B)$, the "|" is spoken as "given" or "knowing"

# 4: Probability

## 4.2 Conditional Probability of Events

$$P(A) = \frac{n(A)}{n(S)}$$

$$P(B) = \frac{n(B)}{n(S)}$$

**Example:** Roll two die.

Let $A$ be that 10 is the sum of the two die.

$P(A)$=

Let $B$ that the first die is a 4.

$P(B)$=

What is P($A|B$)?.

$P(A|B)$=

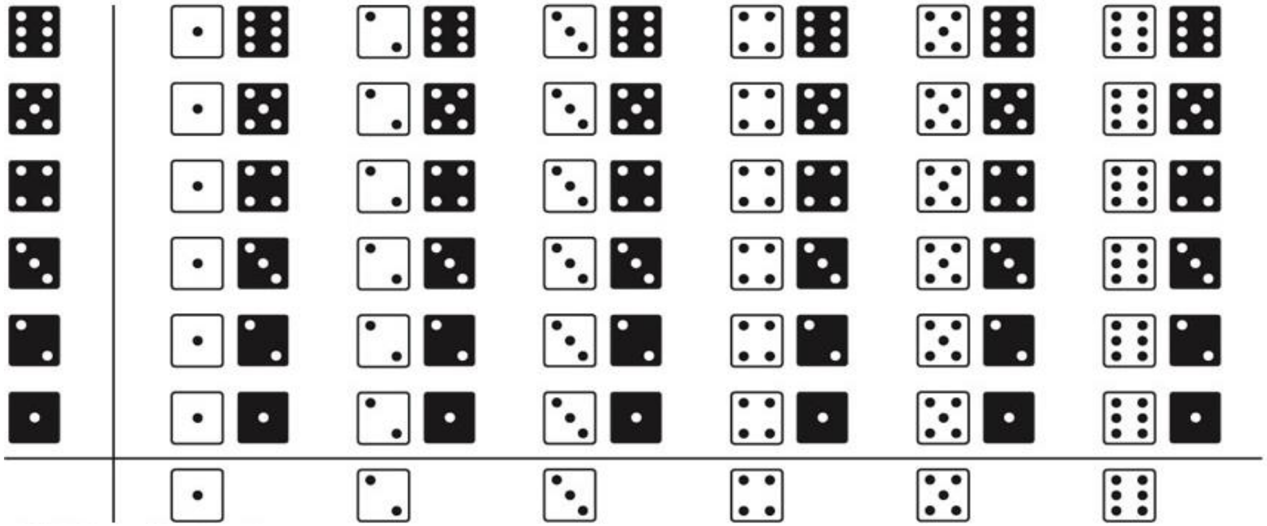


Figure from Johnson & Kuby, 2012.

# 4: Probability

## 4.2 Conditional Probability of Events

$$P(A) = \frac{n(A)}{n(S)}$$

$$P(B) = \frac{n(B)}{n(S)}$$

**Example:** Roll two die.

Let $A$ be that 10 is the sum of the two die.

$P(A)$=3/36

Let $B$ that the first die is a 4.

$P(B)$=

What is P($A|B$)?.

$P(A|B)$=

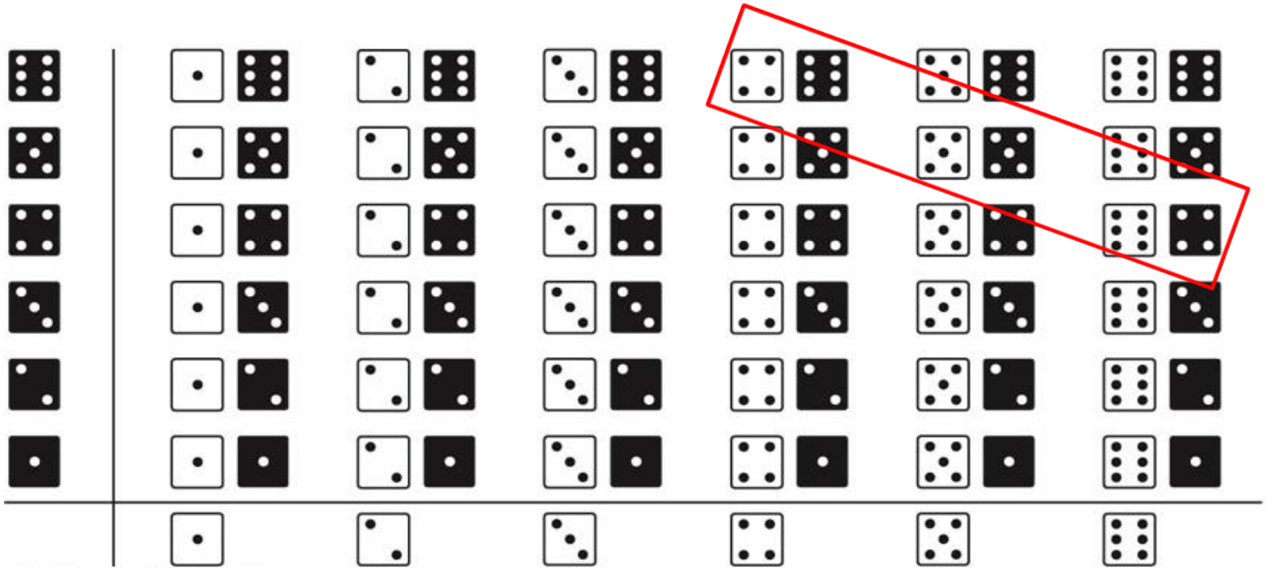


Figure from Johnson & Kuby, 2012.

# 4: Probability

## 4.2 Conditional Probability of Events

$$P(A) = \frac{n(A)}{n(S)}$$

$$P(B) = \frac{n(B)}{n(S)}$$

**Example:** Roll two die.

Let $A$ be that 10 is the sum of the two die.

$P(A)$=3/36

Let $B$ that the first die is a 4.

$P(B)$=6/36

What is P($A$|$B$)?.

$P(A|B)$=

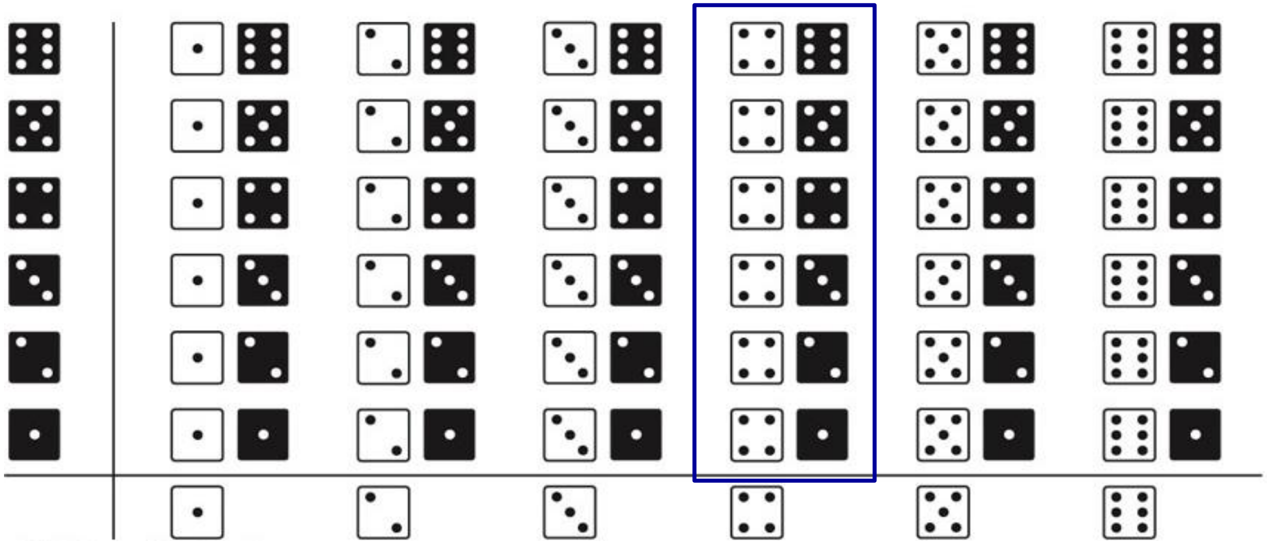


Figure from Johnson & Kuby, 2012.

# 4: Probability

## 4.2 Conditional Probability of Events

$$P(A)=\frac{n(A)}{n(S)}$$

$$P(B)=\frac{n(B)}{n(S)}$$

**Example:** Roll two die.

Let $A$ be that 10 is the sum of the two die.

$P(A)$=3/36

Let $B$ that the first die is a 4.

$P(B)$=6/36

What is P($A|B$)?.

$P(A|B)$=1/6

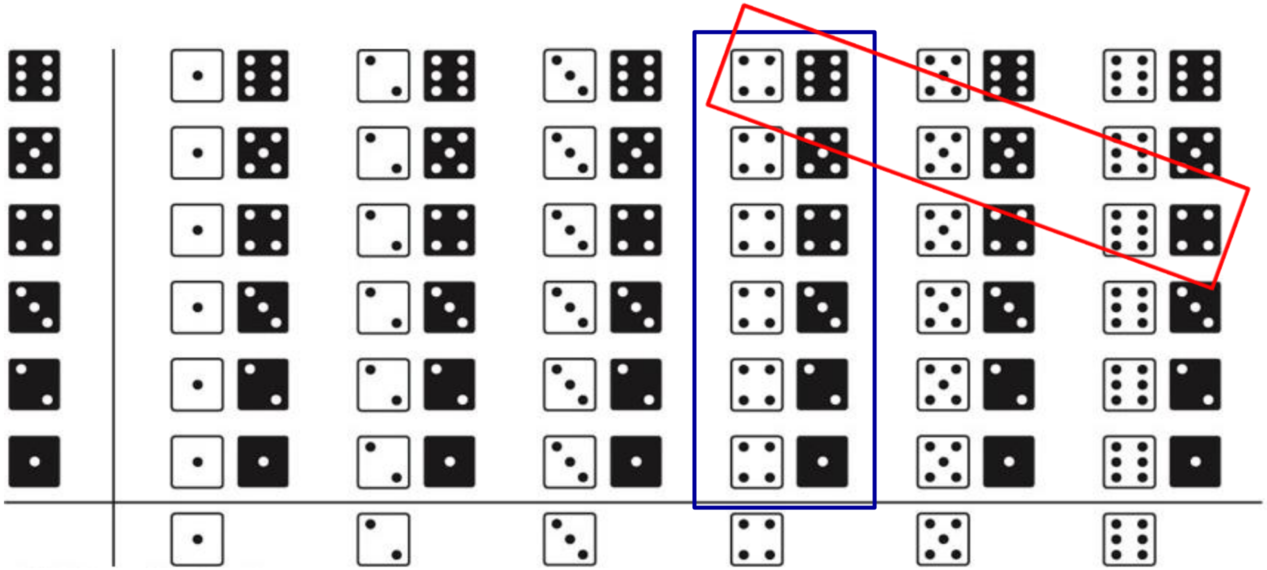


Figure from Johnson & Kuby, 2012.

# 4: Probability

Questions?

Homework: Finish Reading Chapter 4
WebAssign
Chapter 4 # 3,11,12,13, 31, 51, 57

Roll a pair of die 100 times.
Let $A$ be that 7 is the sum of the two die.
Calculate $P(A)$ using the empirical approach.